Abstract

System-level analysis of early signalling in T cells

Jiandong Huo, Wolfson College

Thesis submitted for the degree of Doctor of Philosophy

Trinity term, 2012

The prevailing view of signal transduction is that it proceeds through the linear relay of information via sequential bimolecular interactions, involving, for example, Src homology (SH) 2 domains. It has been assumed that such interactions are highly selective, i.e. that the affinities of these interactions are several orders of magnitude higher than that for non-specific interactions. However, recent studies have suggested that the difference in affinities between so-called specific and non-specific interactions is not sufficient to support such a proposal. This therefore raises the question of how signalling pathway specificity is generated at all. To address this, we have taken a systems approach by expressing and purifying >90% of the SH2 domains identified in a T cell line using a next-generation sequencing-based transcriptomic analysis, and performed a systematic survey of the interaction of these SH2 domains with a set of potential phosphorylated peptides derived from the key signalling receptors of the T cell (including CD28, CTLA-4, PD-1, ICOS, BTLA, LAT and the CD3 subunits of the TCR complex), using surface plasmon resonance-based binding assays. Our results show that, instead of being highly selective for certain SH2 domains, the T cell-expressed receptors are very cross-reactive, such that each receptor is found to interact with ~50 different SH2 domains on average. *In silico* analysis based on these results confirms the expectation that affinity itself is not the sole determining factor for receptor specificity. Further exploration of the system using *in silico* simulations incorporating the absolute concentrations of SH2 domain-containing proteins measured in T cells using a proteomics-based approach, suggests instead that the specificity of SH2 domain recruitment by T-cell receptors is the result of systems effects, with expression levels of the signalling proteins being a major factor. Surprisingly, LCK, the most highly expressed SH2 domain in resting Jurkat, is predicted to dominate the binding of most receptors, suggesting a novel mechanism of Src kinase activation and function.

Acknowledgements

First and foremost, I owe a great debt of gratitude to my Gurus, the Buddha and Bodhisattvas, whose blessing and teachings have been vital to keep me moving and lead a meaningful life.

To my supervisor, Simon Davis, who has provided me with the wonderful opportunity to study at Oxford. Many thanks for his time and effort spent on me, as well as all the patience and support to me. In particular, he is always very kind to push me to achieve the higher standards at every stage throughout my D. Phil, so that I can well explore my potentials and do better and better – I really appreciate it. Also, to Ray Owens, the head of the OPPF, who has been extremely kind and supportive during my time doing the protein work.

This work would not have been able to accomplish without the involvement of all my wonderful colleagues and collaborators. I firstly need to thank Louis Bird, who has supervised me to express and purify the large amount of different proteins, which has laid the material basis of this project. In addition, Mai Vuong has provided substantial help in performing the SPR study and making more proteins for the next stage of research during my thesis write-up. To Anil Verma and Veronica Chang, who offered me a lot of support in settling down at OPPF and the T cell biology group, respectively, as well as have been helpful all the time. I am also greatly indebted to Oliver Lui, who has provided enormous amount of help in bioinformatics analysis and general help. Additionally, I would like to thank Omer Dushek and Benedikt Kessler, who have introduced me to the amazing world of mathematic and mass spectrometry. The opportunity of working in two different groups has allowed me to meet a lot of nice persons and make friends, and thanks to all their support and encouragements.

Finally, to my parents, the two people I owe most. To Liu Lei, a most intimate friend of mine, who has always been very supportive and considerate. In particular, she had helped me in running the charity and taking care of a friend of mine who got injured, during the most critical time when I was writing up my thesis, so that I was able to fully concentrate on the writing; I feel deeply indebted to her.

What I can do in return is to attain enlightenment and make achievements at every stage of my life for the benefit of all sentient beings.

Abbreviations

APC	antigen presenting cell
BTLA	B and T lymphocyte attenuator
CBL-b	Casitas B-lineage lymphoma proto-oncogene b
CD	cluster of differentiation
CDR	complementarity determining region
Cn	calcineurin
cSMAC	central supramolecular activation cluster
cTEC	cortical thymic epithelial cell
CTL	cytotoxic T lymphocyte
CTLA-4	cytotoxic T lymphocyte-associated antigen-4
CVID	common variable immunodeficiency
Da	Dalton
DAG	diacylglycerol
dSMAC	distal supramolecular activation cluster
E. coli	*Escherichia coli*
EAE	experimental allergic encephalomyelitis
EGFR	epidermal growth factor receptor
emPAI	exponentially modified protein abundance index
ER	endoplasmic reticulum
ErbB	epidermal growth factor receptor
ERK	extracellular signal-regulated kinase
GADS	Grb2-related adaptor downstream of Shc
GEMs	glycol-lipid enriched micro-domains
GnHCl	guanidine hydrochloride
GPCR	G protein coupled receptor
GRB2	Growth factor receptor-bound protein 2

GST	glutathione-S-transferase
GVHD	graft-vs-host disease
HEV	high endothelial venule
His(6)	hexahistitine
HPLC	high-perfomance liquid chromatography
HVEM	herpesvirus entry mediator
iBAQ	intensity based absolute quantification
ICAM	intercellular adhesion molecule
ICOS	inducible costimulator
IEF	isoelectric focusing
IFN-γ	interferon-gamma
IgSF	immunoglobulin superfamily
IL	interleukin
IP3	inositol 1,4,5-trisphosphate
IS	immunological synapse
ITAM	immunoreceptor tyrosine-based activation motif
ITIM	immunoreceptor tyrosine-based inhibitory motif
ITK	interleukin-2 tyrosine kinase
iTreg	induced regulatory T cell
ITSM	immunoreceptor tyrosine-based switch motif
JAK	Janus kinase
K_A	association constant
K_D	dissociation constant
KS model	kinetic-segregation model
LAT	linker for activation of T cells
LCK	lymphocyte-specific tyrosine kinase
LDC	ligation-dependent cloning
LFA	lymphocyte function-associated antigen
LIC	ligation-independent clonin

MAPK	mitogen-activated protein kinase
MBP	maltose binding protein
MHC	major histocompatibility complex
MS	mass spectrometry
mTEC	medullary thymic epithelial cell
NFAT	nuclear factor of activated T cells
NGS	next-generation sequencing
Ni-NTA	nickel-nitrilotriacetic acid
NKT	natural killer T cell
NusA	N-utilisation substance A
PD-1	programmed cell death protein 1
PDGFR	platelet derived growth factor receptor
PDK1	PtdIns 3-Phosphoinositide-dependent protein kinase-1
PH domain	pleckstrin-homology domain
PIK3	phosphoinositide-3-kinase
PIK3R	phosphoinositide-3-kinase, regulatory subunit
PIP2	phosphatidylinositol-4,5-bisphosphate
PKC	protein kinase C
pMHC	peptide-complexed major histocompatibility complex
PP2A	protein phosphatase 2A
pSMAC	peripheral supramolecular activation cluster
PTB domain	phosphotyrosine-binding domain
PTK	protein tyrosine kinase
PTP	protein tyrosine phosphatases
pY	Phosphotyrosine
RTP	receptor tyrosine phosphatases
SAGE	serial analysis of gene expression
Ser/Thr	serine/threonine
SH2	Src Homology Domain 2

SHC	SHC-transforming protein
SHP1	SH2 domain-containing tyrosine phosphatase 1
SHP2	SH2 domain-containing tyrosine phosphatase 2
SILAC	stable isotope labelling by amino acids in cell culture
siRNA	small interfering RNA
SLAM	signaling lymphocytic activation molecule
SLAP	Src-like-adapter protein
SLP-76	SH2 domain-containing leukocyte protein of 76 kDa
SMR	selected reaction monitoring
SOCS	suppressor of cytokine signaling
SPR	surface plasmon resonance
SPR	surface plasmon resonance
STAT	signal transducer and transcription activator
SUMO	small ubiquitin-related modifier
Syk	spleen tyrosine kinase
TCR	T cell antigen receptor
TF	trigger factor
TFH	follicular B helper T cell
TGF	transforming growth factor
Th cell	T helper cells
TNF	tumor necrosis factor
TNFR	tumor necrosis factor receptor
TRX	thioredoxin
ZAP-70	ζ-chain associated protein of 70kDa

Table of Contents

ABSTRACT 1

ACKNOWLEDGEMENTS 2

ABBREVIATIONS 3

CHAPTER 1. INTRODUCTION 14

1.1 OVERVIEW 14
1.2 ADAPTIVE IMMUNITY 14
1.2.1 T cells 15
1.2.2 CD4$^+$ T cells 16
1.2.3 Cytotoxic T lymphocytes 18
1.2.4 Memory T cells 19
1.2.5 Regulatory T cells 19
1.2.6 Natural killer T cells 20
1.3 SURFACE MOLECULES INVOLVED IN THE ACTIVATION OF T CELLS 20
1.3.1 Molecules involved in cell adhesion 20
1.3.2 Proteins involved in antigen presentation and recognition 22
1.3.3 Co-receptor molecules 23
1.4 MODELS OF T CELL ACTIVATION 24
1.5 THE IMMUNOLOGICAL SYNAPSE 28
1.6 THE KEY SIGNALLING RECEPTORS IN T CELLS 30
1.6.1 TCR 31
1.6.2 LAT 32
1.6.3 Co-stimulatory/co-inhibitory receptor molecules 34
1.7 PHOSPHOTYROSINE SIGNALLING IN METAZOANS 47
1.7.1 Evolution of phosphotyrosine-based signalling 47
1.7.2 Phosphotyrosine binding domains 48
1.8 UNTANGLING SIGNAL TRANSDUCTION – A ROLE FOR SYSTEMS BIOLOGY 50
1.8.1 The definition of systems biology 50
1.8.2 Two approaches to systems biology 51
1.8.3 Two views of signal transduction 52
1.8.4 Is specificity system- rather than molecule-intrinsic? 55
1.8.5 Obtaining system insights from transcriptomic analysis 56
1.8.6 Mass spectrometry-based proteomics 58
1.9 AIMS OF THIS THESIS 60

CHAPTER 2. MATERIALS AND METHODS 62

2.1 POLYMERASE CHAIN REACTION 62
2.2 DPNI DIGESTION OF PCR REACTIONS 64
2.3 AGAROSE GEL ELECTROPHORESIS 64
2.4 LIGATION OF A GENE OF INTEREST INTO A PLASMID USING INFUSION™ 65
2.5 AMPLIFICATION OF PLASMID DNA AND VERIFICATION OF INSERT 65
2.6 GLYCEROL STOCKS OF TRANSFORMED BACTERIA 66
2.7 PREPARATION OF COMPETENT *E. COLI* EXPRESSION STRAINS USING THE RBCL METHOD 68
2.8 PROTEIN EXPRESSION SCREENING IN *E. COLI* 70

2.9 LARGE-SCALE PRODUCTION OF SOLUBLE PROTEINS 73
2.10 PROTEIN REFOLDING 74
2.10.1 Preparation of denatured inclusion bodies 74
2.10.2 The refolding screen 75
2.10.3 Refolding by rapid dilution 78
2.10.4 Surface plasmon resonance-based experiments 78

CHAPTER 3. TRANSCRIPTOME- AND PROTEOME-BASED ANALYSIS OF SH2 DOMAIN EXPRESSION 81

3.1 INTRODUCTION 81
3.1.1 Transcriptomic approaches 82
3.1.2 SAGE technology 82
3.1.3 Next-generation sequencing 83
3.1.4 Mass spectrometry-based proteomic analysis 84
3.1.5 The use of the Jurkat T-cell line in signal transduction research 85
3.2 RESULTS 86
3.2.1 Comparisons of SH2 domain-containing protein transcription between different cell types using SAGE 88
3.2.2 Proteomic analysis of SH2 protein expression 102
3.3 DISCUSSION 108

CHAPTER 4. EXPRESSION AND PURIFICATION OF RECOMBINANT SH2 DOMAINS 112

4.1 INTRODUCTION 112
4.1.1 The system for expressing recombinant SH2 domains 112
4.1.2 Overview of the expression strategies 115
4.2 RESULTS 121
4.2.1 Expression with a His-tag 122
4.2.2 Expression with GST and MBP fusion tags 124
4.2.3 Expression with Trx/SUMO/TF/MsyB/NusA fusion tag 130
4.2.4 Redesign of difficult constructs 132
4.2.5 Protein refolding 134
4.3 DISCUSSION 136

CHAPTER 5. AFFINITIES OF SH2 DOMAINS FOR THE PHOSPHOTYROSINE RECEPTOR MOTIFS OF T CELL 141

5.1 INTRODUCTION 141
5.1.1 The nature of molecular interactions involved in the initial step of intracellular signal transduction 141
5.1.2 Surface plasmon resonance 142
5.1.3 Implementation of SPR in the BIAcore instrument 144
5.1.4 Immobilisation of ligands to the sensor chip surface 144
5.1.5 Protein interaction affinity analysis 146
5.2 RESULTS 147
5.2.1 Investigation of the effect of peptide length on affinity measurements 147
5.2.2 Investigation of the effect of choice of control peptide for the SPR experiments 150
5.2.3 Systems views of the SH2/tyrosyl phosphopeptides interactions ... 151

5.2.4 The interactions between SH2 domains and the individual tyrosyl phosphopeptides 162
5.2.5 Comparisons of the binding profiles for the ITIM and ITSM motifs . 169
5.2.6 Comparison of the SH2 domain binding profiles of different receptors with similar functions 172
5.3 DISCUSSION 177

CHAPTER 6. ANALYSIS OF SPECIFICITY IN A SYSTEM-BASED MODEL OF RECEPTOR-LIGAND INTERACTIONS 181

6.1 INTRODUCTION 181
6.2 RESULTS 183
6.2.1 Analysis of the binding specificity of SH2 domains for the receptor motifs 183
6.2.2 Development of the model system 185
6.2.3 Effects of affinity and concentration on the probabilities of SH2 domain binding to phosphorylated tyrosine motifs 189
6.2.4 Comparisons of the SH2 binding probabilities for related receptor motifs in Models 1 and 2..... 196
6.2.5 The analysis of receptor specificity in the absence of LCK 205
6.3 DISCUSSION 210

CHAPTER 7. GENERAL DISCUSSION 213

7.1 RESULTS FROM THE THESIS 214
7.1.1 Transcriptomic and proteomic analysis of SH2 protein expression . 214
7.1.2 Production of SH2 domains for system-level binding analysis 217
7.1.3 Interactions of SH2 domains with the phosphotyrosine-containing motifs of T cell signalling receptors 218
7.1.4 Analysis of specificity in a system-based model of receptor-ligand interactions..... 221
7.1.5 Signalling implications of the new findings 223
7.1.6 Verification of the simulation model 225
7.1.7 Towards a more complete simulation model..... 227
7.2 FINAL REMARKS 228

BIBLIOGRAPHY 230

List of Figures

FIGURE 1.1. OVERALL SCHEME OF T-CELL DEVELOPMENT IN THE THYMUS. 17
FIGURE 1.2. THE KINETIC-SEGREGATION MODEL..... 27
FIGURE 3.1. THE AVERAGE NUMBER OF SH2 PROTEINS IDENTIFIED IN IMMUNE CELLS VERSUS NON-IMMUNE CELLS..... 89
FIGURE 3.2 (A). THE MRNA EXPRESSION FOR SH2 DOMAIN-CONTAINING IN 46 DIFFERENT SAGE LIBRARIES.. 90
FIGURE 3.2 (B). THE MRNA EXPRESSION FOR PTB DOMAIN-CONTAINING PROTEINS IN 46 DIFFERENT SAGE LIBRARIES.. 91
FIGURE 3.2 (C). THE MRNA EXPRESSION FOR TYROSINE MOTIF-CONTAINING RECEPTORS IN 46 DIFFERENT SAGE LIBRARIES.. 92
FIGURE 3.3.SAGE ANALYSIS OF SH2 DOMAIN-ENCODING MRNA EXPRESSION IN A CD4+ T CELL CLONE (CLONE 29).....97
FIGURE 3.4. FOLD CHANGES IN EXPRESSION OF (A) SH2 DOMAIN-ENCODING TRANSCRIPTS AND (B) PTB-DOMAIN-CNCODING TRANSCRIPTS BETWEEN ACTIVATED AND RESTING CELLS.....98
FIGURE 3.5. COMPARISON OF SH2 PROTEIN-ENCODING MRNA EXPRESSION BETWEEN CD4+ T CELLS (CLONE 29) AND CEREBELLUM AS WELL AS CD8+ T CELLS (CLONE 32)..... 101
FIGURE 3.6. COMPARISON OF PROTEIN-ENCODING MRNA EXPRESSION BETWEEN CD4+ T CELLS (CLONE 29) AND JURKAT T CELLS AS WELL AS 293T CELLS.. 101
FIGURE 3.7. SH2 DOMAIN-ENCODING MRNA EXPRESSION IN CLONE 29 AND JURKAT T CELLS.. 102
FIGURE 3.8. DETERMINATION OF A CALIBRATION CURVE THAT RELATES THE PROTEIN ABUNDANCE (EMPAI) MEASURED BY MASS SPECTROMETRY AND THE AMOUNT OF PROTEIN ANALYSED USING A UPS-2 PROTEIN STANDARD MIX..... 104
FIGURE 3.9. CALIBRATION CURVE RELATING THE PROTEIN ABUNDANCE (IBAQ) MEASURED BY MASS SPECTROMETRY AND THE CONCENTRATION OF THE PROTEIN ANALYSED..... 105
FIGURE 3.10. THE CONCENTRATION OF SH2 PROTEINS IN A JURKAT T-CELL LINE.. 107
FIGURE 3.11. A VENN DIAGRAM SHOWING THE NUMBER OF SH2 PROTEINS IDENTIFIED IN THE TRANSCRIPTOMIC AND PROTEOMIC DATA..... 108
FIGURE 3.12. COMPARISON OF THE SH2 PROTEIN CONCENTRATIONS AND MRNA TRANSCRIPTION IN JURKAT T CELLS. 109
FIGURE 4.1. THE CLONING VECTOR AND METHOD OF CLONING.. 119
FIGURE 4.2. REPRESENTATIVE PCR VERIFICATION GEL FOR SEVEN OF THE TARGETS IN THE FUSION VECTORS.. 120
FIGURE 4.3. REPRESENTATIVE EXPRESSION SCREEN RESULTS.. 120
FIGURE 4.4. REPRESENTATIVE RESULTS FOR EXPRESSION AND PURIFICATION OF SH2 DOMAINS FROM A HIS-TAGGED PROTEIN.. 123
FIGURE 4.5. REPRESENTATIVE RESULTS FOR EXPRESSION AND PURIFICATION OF SH2 DOMAINS FROM A FUSION PROTEIN.. 128
FIGURE 4.6. PURIFICATION OF THE SH2D3A SH2 DOMAIN USING AN MBP FUSION TAG.... 129
FIGURE 4.7. THE CLEAVED PRODUCTS OF SLAP-TF AFTER PURIFICATION BY SIZE-EXCLUSION CHROMATOGRAPHY.. 131
FIGURE 4.8. PURIFICATION OF STAT1. EACH LANE CORRESPONDS TO A FRACTION COLLECTED FROM THE HPLC PURIFICATION. 134
FIGURE 4.9. SEQUENCE ALIGNMENT BETWEEN THE HUMAN AND MOUSE SHC1 PROTEIN..... 134

FIGURE 4.10. SUMMARY SDS-PAGE GEL SHOWING THE SH2 PROTEINS EXPRESSED AND PURIFIED IN THIS STUDY. 139
FIGURE 5.1. A REPRESENTATIVE EXAMPLE OF EQUILIBRIUM BINDING ANALYSES FOR THE LCP2 SH2 DOMAIN WITH THE CD6 (662) LONG (PEPTIDE-P; LEFT) AND SHORT (PEPTIDE-T; RIGHT) PEPTIDES..... 149
FIGURE 5.2. COMPARISON OF THE AFFINITIES OF THE SH2 DOMAINS AND THE PD-1_ITIM MOTIF DERIVED FROM DATA USING TWO DIFFERENT PEPTIDES AS THE NEGATIVE CONTROL. THE AFFINITIES ARE PRESENTED AS K_A VALUES (ASSOCIATION CONSTANT).. 151
FIGURE 5.3. THE K_A VALUES DERIVED FROM DATA USING THE NON-PHOSPHORYLATED PEPTIDE AS THE NEGATIVE CONTROL WERE PLOTTED AGAINST THOSE DERIVED USING THE GS-CONTROL PEPTIDE AS THE NEGATIVE CONTROL. 152
FIGURE 5.4. A COMPARISON OF THE EFFECTS OF USING FOUR AND THREE DATA POINTS FOR GENERATING THE SCATCHARD PLOTS IN TWO REPLICATES. 155
FIGURE 5.5. A SUMMARY OF INTERACTIONS BETWEEN SH2 DOMAINS AND THE TYROSYL PHOSPHOPEPTIDES OF T CELL SURFACE SIGNALLING RECEPTORS. 157
FIGURE 5.6. A SUMMARY OF THE NUMBER OF SH2 DOMAINS FOUND TO INTERACT WITH THE TYROSYL PHOSPHOPEPTIDES. 159
FIGURE 5.7. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO DIFFERENT TYROSYL PHOSPHOPEPTIDES AT DIFFERENT AFFINITY RANGES..... 159
FIGURE 5.8. A SUMMARY OF THE INTERACTIONS BETWEEN THE SH2 DOMAINS OF LCK, FYN AND YES1 WITH DIFFERENT TYROSYL PHOSPHOPEPTIDES AT DIFFERENT AFFINITY RANGES. .. 161
FIGURE 5.9. A SUMMARY OF THE NUMBER OF SH2 DOMAINS FOUND TO INTERACT WITH THE T CELL SURFACE RECEPTORS INVESTIGATED IN THIS STUDY. 161
FIGURE 5.10. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO DIFFERENT T CELL SURFACE RECEPTORS AT DIFFERENT AFFINITY RANGES. 162
FIGURE 5.11. A SUMMARY OF THE AFFINITIES (PRESENTED AS K_A VALUES) OF THE SH2 DOMAINS FOR THE Y(-P)MNM AND Y(-P)APP MOTIFS OF CD28..... 163
FIGURE 5.12. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE Y(-P)VKM AND Y(-P)FIP MOTIFS OF CTLA-4..... 164
FIGURE 5.13. A SUMMARY OF THE SH2 DOMAINS FOR THE ITIM AND ITSM MOTIFS OF PD-1. 165
FIGURE 5.14. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE Y(-P)DND MOTIF, AS WELL AS THE ITIM AND ITSM MOTIFS OF BTLA. 166
FIGURE 5.15. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE Y(-P)MFM MOTIF OF ICOS. 167
FIGURE 5.16. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE ITAM MOTIFS OF THE CD3 SUBUNITS OF THE TCR COMPLEX. 167
FIGURE 5.17. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE YENE (Y110), YVNVP (Y171), YVNVS (Y191) AND YENL (Y226) MOTIFS OF LAT..... 169
FIGURE 5.18. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO THE ITIM MOTIFS OF PD-1 AND BTLA AT DIFFERENT AFFINITY RANGES. 170
FIGURE 5.19. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE ITIM MOTIF OF PD-1 AND BTLA RECEPTORS. 170
FIGURE 5.20. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO THE ITSM MOTIF OF THE PD-1 AND BTLA RECEPTORS AT DIFFERENT AFFINITY RANGES..... 171
FIGURE 5.21. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAINS FOR THE ITSM MOTIF OF THE PD-1 AND BTLA RECEPTORS. 172
FIGURE 5.22. A COMPARISON OF THE BINDING OF THE SH2 DOMAINS TO THE CD28 AND ICOS RECEPTORS..... 173

FIGURE 5.23. A COMPARISON OF THE BINDING OF THE SH2 DOMAINS TO THE CD28 RECEPTOR AND TCR. 174
FIGURE 5.24. A COMPARISON OF THE BINDING OF THE SH2 DOMAINS TO THE CD28 AND CTLA-4 RECEPTORS 175
FIGURE 5.25. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO THE CD28 AND LAT ADAPTORS AT DIFFERENT AFFINITY RANGES. 175
FIGURE 5.26. A COMPARISON OF THE BINDING OF THE SH2 DOMAINS TO THE CD28 AND LAT ADAPTORS. 176
FIGURE 5.27. A SUMMARY OF THE NUMBER OF SH2 DOMAINS BINDING TO THE CTLA-4, PD-1 AND BTLA RECEPTORS AT DIFFERENT AFFINITY RANGES. 177
FIGURE 5.28. A COMPARISON OF THE BINDING OF THE SH2 DOMAINS TO THE CTLA-4, PD-1 AND BTLA RECEPTORS. 177
FIGURE 5.29. A SUMMARY OF THE FOLD-DIFFERENCE BETWEEN THE TWO REPLICATES OF BIOPHYSICAL MEASUREMENTS FOR DIFFERENT TYROSYL PHOSPHOPEPTIDES 179
FIGURE 6.1. A SUMMARY OF GIBBS FREE ENERGY (ΔG) BETWEEN SH2 DOMAINS AND THE TYROSYL PHOSPHOPEPTIDES OF T CELL SURFACE SIGNALLING RECEPTORS.. 184
FIGURE 6.2. THE DISTRIBUTION OF THE GIBBS FREE ENERGY (ΔG) FOR THE 974 MEASURED INTERACTIONS OF THE SH2 DOMAINS WITH THE TYROSYL PHOSPHOPEPTIDES. 185
FIGURE 6.3. A SUMMARY OF THE SH2/RECEPTOR MOTIF BINDING PROBABILITIES ($P_{i,j}$) WHEN ONLY AFFINITY IS CONSIDERED (MODEL 1). 190
FIGURE 6.4. A SUMMARY OF THE SH2/RECEPTOR MOTIF BINDING PROBABILITIES ($P_{i,j}$) WHEN THE DIFFERENTIAL CONCENTRATIONS OF THE SH2 DOMAINS ARE CONSIDERED IN ADDITION TO AFFINITY (MODEL 2).. 191
FIGURE 6.5. A COMPARISON OF THE NUMBER OF INTERACTIONS BETWEEN SH2 DOMAINS AND THE 20 DIFFERENT RECEPTOR MOTIFS IN DIFFERENT BINDING PROBABILITY RANGES BETWEEN MODEL 1 AND MODEL 2 193
FIGURE 6.6. PROBABILITY SPECTRA FOR SH2 DOMAIN BINDING TO EXAMPLE RECEPTOR MOTIFS ACCORDING TO MODELS 1 (LEFT PANEL) AND 2 (RIGHT PANEL) 194
FIGURE 6.7. PROBABILITY SPECTRA FOR SH2 DOMAIN BINDING TO THE CD3 ITAM MOTIFS FOR MODELS 1 (LEFT PANEL) AND 2 (RIGHT PANEL). 197
FIGURE 6.8. PROBABILITY SPECTRA FOR SH2 DOMAIN BINDING TO THE TYROSYL PHOSPHOPEPTIDES OF LAT, ICOS AND CD28 IN MODELS 1 (LEFT PANEL) AND 2 (RIGHT PANEL) 199
FIGURE 6.9. PROBABILITY SPECTRA FOR SH2 DOMAIN BINDING TO THE TYROSYL PHOSPHOPEPTIDES OF CTLA, PD-1 AND BTLA IN MODELS 1 (LEFT PANEL) AND 2 (RIGHT PANEL) 202
FIGURE 6.10. RECEPTOR SPECIFICITY MEASURED USING THE REDUNDANCY MEASURE OF ENTROPY (Δ). 204
FIGURE 6.11. A COMPARISON OF THE AFFINITIES OF THE DIFFERENT PHOSPHORYLATED RECEPTOR MOTIFS AND THE LCK TYR^{505} MOTIF FOR THE LCK SH2 DOMAIN.. 206
FIGURE 6.12. A COMPARISON OF RECEPTOR SPECIFICITIES FOR SYSTEMS WITH AND WITHOUT LCK 208
FIGURE 6.13. PERCENT CHANGE IN RECEPTOR SPECIFICITY FOR THE SYSTEM WITH THE SRC-KINASES VERSUS THAT WITHOUT THE KINASES. 209
FIGURE 6.14. COMPARISON OF THE PROBABILITY SPECTRUM FOR THE BINDING OF THE SH2 DOMAINS TO THE PD-1_ITSM AND BTLA_ITSM MOTIFS IN MODEL 2 IN THE PRESENCE AND ABSENCE OF LCK. 210

List of Tables

TABLE 2.1. KOD XTREME™ DNA POLYMERASE REACTION MIX 62
TABLE 2.2. KOD XTREME™ DNA POLYMERASE PCR CONDITIONS 63
TABLE 2.3. PHUSION™ FLASH HIGH-FIDELITY PCR MASTER MIX REACTION MIX 63
TABLE 2.4. PHUSION™ FLASH HIGH-FIDELITY PCR MASTER MIX PCR CONDITIONS 64
TABLE 2.5. KOD HI-FI™ DNA POLYMERASE REACTION MIX 67
TABLE 2.6. KOD HI-FI™ DNA POLYMERASE PCR CONDITIONS 67
TABLE 2.7. COMPONENTS OF TRANSFORMATION BUFFER I (TFB I) 69
TABLE 2.8. COMPONENTS OF TRANSFORMATION BUFFER II (TFB II) 69
TABLE 2.9. COMPONENTS OF BUFFERS USED IN THE EXPRESSION SCREEN 72
TABLE 2.10. THE COMPONENTS OF THE REFOLDING SCREEN BUFFER 76
TABLE 2.11. THE CONDITIONS OF REFOLDING SCREEN BUFFER 77
TABLE 3.1. A LIST OF SAGE LIBRARIES AVAILABLE FOR THIS STUDY. 87
TABLE 3.2. LIST OF SH2 PROTEINS WHOSE ENCODING MRNA EXPRESSION IS SIGNIFICANTLY HIGHER IN IMMUNE VERSUS NON-IMMUNE CELL LIBRARIES. 94
TABLE 3.3. LIST OF SH2 PROTEINS WHOSE ENCODING MRNA EXPRESSION IS SIGNIFICANTLY HIGHER IN T-CELL VERSUS "NON-T" IMMUNE CELL LIBRARIES. 96
TABLE 3.4. A SUMMARY OF THE MASS SPECTROMETRY-BASED MEASUREMENTS FOR A SET OF PROTEINS SHARED BY JURKAT SAMPLES FROM THE KESSLER GROUP AND THE MANN GROUP PROTEINS. 105
TABLE 4.1. THE PROPERTIES OF THE VECTORS. 116
TABLE 4.2. SMALL-SCALE EXPRESSION-SCREEN RESULTS. 125
TABLE 4.3. THE REFOLDING SCREENING RESULTS FOR CHN1. 137
TABLE 4.4. THE FINAL LIST OF SH2 DOMAINS EXPRESSED FOR THIS STUDY. 138
TABLE 5.1. A SUMMARY OF THE AFFINITIES OF THE SH2 DOMAIN OF LCP2 FOR CD6 (662) PEPTIDE-P AND PEPTIDE-T. 149
TABLE 5.2. THE RESPONSES RESULTING FROM THE BINDING OF TNS2 SH2 DOMAIN TO PD-1_ITIM MOTIF–CONTAINING PEPTIDE. 152
TABLE 5.3. THE TYROSYL PHOSPHOPEPTIDES USED TO REPRESENT DIFFERENT RECEPTOR MOTIFS 154

Chapter 1. Introduction

1.1 Overview

The immune system functions to provide protection against disease by detecting disease-causing agents and eliciting relevant responses. Higher organisms possess innate immunity, as well as the more sophisticated adaptive immunity. As part of adaptive immunity, T cells have evolved highly efficient mechanisms not only for detecting and discriminating between foreign antigens, but also for integrating multiple environmental signals to generate appropriate responses. Whilst the majority of the surface receptors are characterised and the recognition mechanisms have been relatively well explored, how signal transduction occurs inside T cells is less clear; particularly, how signalling pathway specificity is actually generated in T cells, or in any eukaryotic cell for that matter, is poorly understood.

The recruitment of SH2 domains to the cytoplasmic motifs of triggered tyrosine phosphorylated surface receptors constitutes a key step of intracellular signal transduction upon activation of T cells. This thesis explores, at the system level, the properties of this recruitment step and attempts to determine the origins of signalling specificity. In this introductory chapter, in addition to adaptive immunity, the current understanding of T cell activation and the relevant signalling pathways will be reviewed. An overview of recent developments in systems biology relevant to the present study is also given.

1.2 Adaptive immunity

Adaptive immunity is characterised by highly specific responses and the requirement of prior exposure to antigen in order to trigger productive responses.

Adaptive immunity is mediated by lymphocytes expressing highly specific receptors, which are capable of recognising a variety of antigens due to their non-germline encoded nature. Upon successful protection against the pathogen, a fraction of the effector cells differentiate into memory cells, which offer accelerated responses upon re-exposure. However, self-antigens can potentially become the targets of adaptive immunity, whose highly specific responses can result in destructive effects on the host.

There are two major cell types that are involved in adaptive immunity – T cells and B cells. Both cell types are derived from early lymphoid progenitors in the bone marrow. Whilst some of these cells stay in the bone marrow to develop into B cells, others migrate to the thymus and develop into T cells [1]. The receptors of these cells are characterised by their assembly through recombination of the large number of variable (V), diversity (D), and joining (J) gene segments, which results in generation of significant diversity for antigen recognition [2].

1.2.1 T cells

Within the thymus, the development of double negative ($CD8^-CD4^-$) thymocytes results from β-selection. At this selection checkpoint, thymocytes are selected for successful rearrangement and pairing of the β chain to the pre-TCRα chain to form a pre-TCR. Effective signalling through the functional pre-TCR complex allows cellular proliferation and expression of CD4 and CD8 co-receptors [3]. Thus, thymocytes passing through this checkpoint become double positive ($CD8^+CD4^+$). Rearrangement of the TCRα chain at this stage leads to formation of the mature TCR complex. In the cortex of thymus, cells are exposed to the major histocompatibility complex (MHC)/self-peptide complex expressed by cortical thymic epithelial cells (cTECs). Cells that fail to interact with the MHC/peptide

complex undergo apoptosis (programmed cell death), while those that successfully interact receive a survival signal (positive selection). Selected cells are committed to single positive ($CD8^+CD4^-$ or $CD8^-CD4^+$) lineages; cells that interact with the MHC class I complex differentiate into CD8 single positive, while those interact with the MHC class II complex differentiate into CD4 single positive cells. These cells then migrate to the medulla, where they are subject to another selection checkpoint called negative selection. During this process, those with high affinity for the MHC/self-peptide complex presented by medullary dendritic cells and TECs (mTECs) are eliminated. Negative selection is crucial in maintaining tolerance to self-peptides and for preventing auto-immune diseases [4, 5]. | An Overall scheme of T-cell development in the thymus is shown in **Figure 1.1**.

1.2.2 $CD4^+$ T cells

$CD4^+$ T cells are also referred to T helper (Th) cells, which play critical roles in mediating adaptive immunity. Upon antigenic stimulation through peptide/MHC class II complexes, naïve CD4 T cells can expand and differentiate into several different lineages including Th1, Th2, Th17 and induced (adaptive) regulatory T cells (iTreg). The cytokine environment largely influences the differentiation of these cells. It was shown that Th1 differentiation is induced by Interleukin-12 (IL-12) and interferon-gamma (IFN-γ), Th2 cells by IL-4 and IL-2, Th17 cells by transforming growth factor beta (TGF-β) and IL-6, while iTreg cells are induced by TGF-β and IL-2 [6, 7]. Each Th cell lineage has "signature" cytokine production - IFN-γ and tumor necrosis factor-beta (TNF-β; also called lymphotoxin) by Th1; IL-4, IL-5 and IL-13 by Th2; IL-17A, IL-7F and IL-22 by Th17; TGF-β and IL-10 by iTreg [6, 8]. Production of these signature cytokines has conferred each cell type with distinct functions. Th1 cells play a crucial role in cell-mediated immune responses through mediating activation of macrophages and cytotoxic $CD8^+$ T

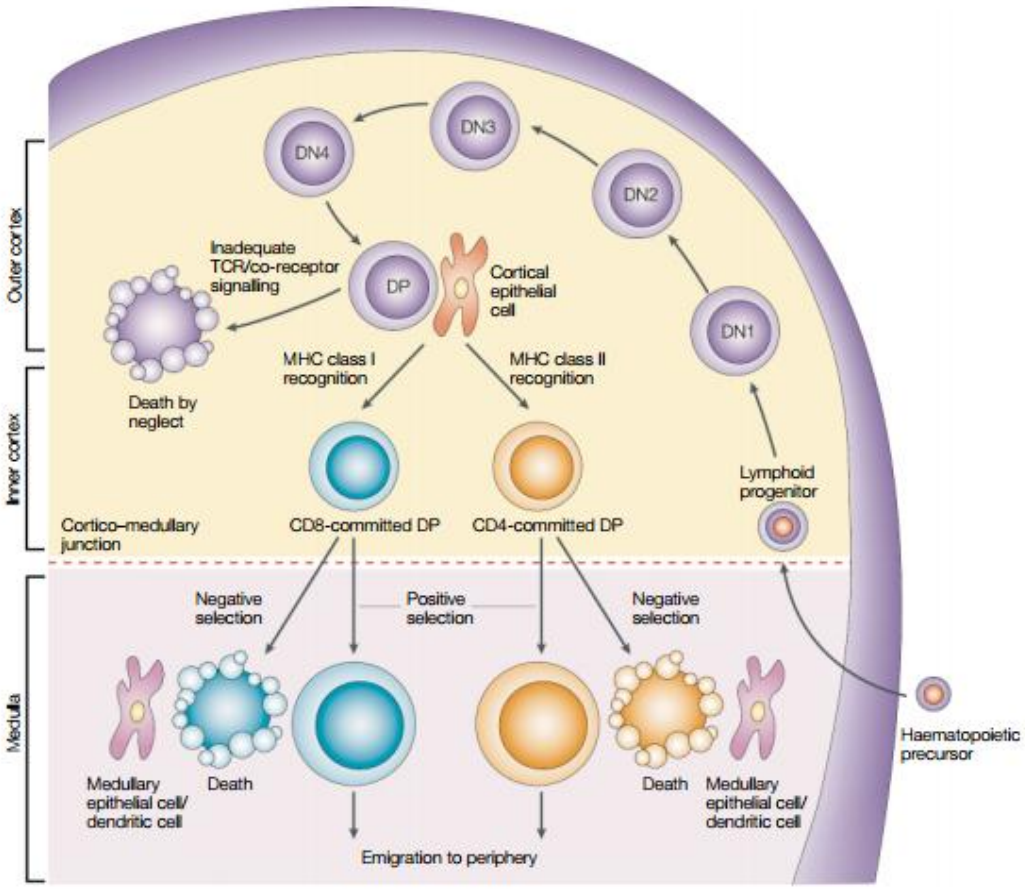


Figure 1.1. Overall scheme of T-cell development in the thymus (adapted from [9]). Committed lymphoid progenitors arise in the bone marrow and migrate to the thymus. Early committed T cells lack expression of T-cell receptor (TCR), CD4 and CD8, and are termed double-negative (DN; $CD4^-CD8^-$) thymocytes. DN thymocytes can be further subdivided into four stages of differentiation (DN1-DN4). As cells progress through the DN2 to DN4 stages, they express the preTCR, which is composed of the non-rearranging pre-Tα chain and a rearranged TCR β-chain. Successful pre-TCR expression leads to substantial cell proliferation during the DN4 to $CD4^+CD8^+$ double positive (DP) transition and replacement of the pre-TCR α-chain with a newly rearranged TCR α-chain, which yields a complete αβ TCR. The DP thymocytes then interact with cortical epithelial cells that express a high density of MHC class I and class II molecules associated with self-peptides. The fate of the DP thymocytes depends on signalling that is mediated by interaction of the TCR with these self-peptide–MHC ligands. Too little signalling results in delayed apoptosis (death by neglect). Too much signalling can promote acute apoptosis (negative selection); this is most common in the medulla on encounter with strongly activating self-ligands on haematopoietic cells, particularly dendritic cells. The appropriate, intermediate level of TCR signalling initiates effective maturation (positive selection). Thymocytes that express TCRs that bind selfpeptide–MHC-class-I complexes become $CD8^+$ T cells, whereas those that express TCRs that bind self-peptide–MHC-class-II ligands become $CD4^+$T cells; these cells are then ready for export from the medulla to peripheral lymphoid sites.

cells. They are responsible for fighting against viruses and intracellular pathogens. In contrast, Th2 cells are involved in humoral immunity. They mediate the defence against extracellular pathogens via inducing the secretion of various immunoglobulins by B cells. Also, they are capable of activating eosinophils and mast cells [10-13]. Whilst earlier studies have shown that Th17 cells play an important role in defence against extracellular pathogens (e.g. bacteria and fungi), emerging data suggests that Th17 cells are highly pro-inflammatory, which raises the possibility of their inducing severe auto-immune diseases if they happen to recognise self-antigens [14]. Unlike the helper T cells discussed above, iTreg cells down-regulate the immune system and prevent auto-immunity by inhibiting the initial priming and proliferation of auto-reactive T cells [15].

1.2.3 Cytotoxic T lymphocytes

Cytotoxic T lymphocytes (CTLs; cytotoxic $CD8^+$ T cells) are characterised by their ability to mediate cytotoxicity. Upon antigenic stimulation through peptide/MHC class I complexes, they can cause death of target cells via secretion of a number of granule-contained mediators, including perforin and granzymes, or cell signalling. Perforin can insert itself into the target cell membrane and polymerise to form a pore, which results in cell lysis. It was thought that this allows intake of serine proteases belonging to the granzyme family into the target cells, which induces apoptosis via caspase-independent mechanisms. However, recent studies have shown that perforin is not able to form pores large enough to allow entry of granzymes. Instead, it is now known that intake of granzymes is achieved through Rab5-dependent endocytosis of both perforin and granzymes inside the target cell. CTLs can also trigger apoptosis of the target cell by the binding of Fas ligand (FasL) to the Fas molecule expressed on target cells [16, 17].

1.2.4 Memory T cells

Immunological memory is a key feature of adaptive immunity. After the resolution of a primary immune response, clonal expansion of lymphocytes ceases and most of the effector cells commit to apoptosis (the contraction phase). A small fraction of effector cells become memory cells with longer life spans, reduced activation thresholds and enhanced effector functions, and a low level of T-cell proliferation is maintained by IL-7 and IL-15. Upon secondary exposure to the same pathogen, memory T cells are reactivated and mount much faster and stronger responses [18, 19].

1.2.5 Regulatory T cells

Naturally occurring regulatory T cells (Tregs) are commonly defined by their expression of the transcription factor forkhead box P3 (FOXP3) and are characterised as $CD4^+CD25^+FOXP3^+$ T cells due to expression of another signature marker CD25 (the IL-2 receptor α-chain). Although only constituting approximately 5% of $CD4^+$ T cells, they play crucial roles in the maintenance of self-tolerance, prevention of auto-immunity, as well as regulation of immune responses to pathogens. They are capable of suppressing the activation, proliferation and effector functions of a wide range of immune cells (including T cells, B cells, NK cells and antigen-presenting cells). While several putative suppressive mechanisms have been revealed (including secretion of immune-suppressive cytokines (e.g. TGF- β and IL-10), induction of cytolysis and inhibiting the activating capacity of DCs; reviewed in [20, 21]), further studies are required to fully understand how Tregs mediate suppression.

1.2.6 Natural killer T cells

Natural killer T (NKT) cells are a distinct subset of T cells that also express the NK1.1 molecule, which is a marker previously considered to be uniquely expressed by NK cells. NKT cells express T cell receptors that are restricted by CD1d, which is an MHC class I-related antigen-presenting molecule expressed by a wide variety of cell types (e.g. hepatocytes, B cells, macrophages and DCs). There are two subtypes of NKT cells: the (type I) invariant NKT (iNKT) cells that possess a semi-restricted TCR repertoire, and the (type II) non-invariant NKT (non-iNKT) cells that express a more diverse TCR repertoire. While iNKT cells recognise only glycolipid antigens, non-iNKT cells also recognise non-lipid small molecules. Once activated through TCR engagement, they are capable of rapidly producing a wide range of immunoregulatory cytokines, including those signature cytokines produced by Th1, Th2 and Th17 cells.

1.3 Surface molecules involved in the activation of T cells

1.3.1 Molecules involved in cell adhesion

In order to fulfil its roles in immune surveillance and host defence, T cells need to migrate from the blood stream into peripheral tissues. However, the shear force of blood flow imposes a principle challenge to these cells as they establish adhesion to the vascular endothelium as well as subsequent migration into the underlying tissues.

The initial contact with the endothelium is mediated by selectins, which are type I membrane glycoproteins. The most important one expressed by T cells is L-selectin, which binds with low affinity to a group of mucins expressed on endothelial cells of high endothelial venules (HEVs) in lymph nodes, activated

endothelial cells and other leukocytes. These mucins include CD34, glycosylated cell adhesion molecule-1 (GlyCAM-1), podocalyxin and mucosal vascular addressin cell adhesion molecule-1 (MAdCAM-1). Low-affinity interactions as such result in rolling of T cells along the endothelium [22, 23].

Thereafter, the arrest and migration phase is mediated by integrin adhesion. Lymphocyte function-associated antigen-1 (LFA-1) is the major integrin involved in such events, with its ligand being intercellular adhesion molecules (ICAMs), primarily ICAM-1. Like other leukocyte integrins, LFA-1 maintains a low-affinity state on circulating cells. Upon chemokine stimulations through G protein coupled receptors (GPCRs), integrins are activated by inside-out signals, which is characterised by conformational changes in the integrin. For LFA-1, a bent conformation is converted into extended ones with intermediate or high affinity, depending on the angle between the hybrid and β-like I domain [24]. Once strong adhesion is established by the activated integrins, the cells undergo a morphological change called polarisation, followed by migration along the apical surface and transmigration through interendothelial junctions [25].

In order to initiate immune responses, T cells need to scan for specific antigens on the antigen presenting cells (APC) surface. To this end, antigen-independent adhesive interactions are required. One of the key surface proteins responsible for such interactions is ICAM-3, which is a member of the immunoglobulin superfamily (IgSF). While ICAM-3 can form low-affinity interactions with LFA-1, it can form high-affinity interaction with the dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN). Once a T cell–APC interface called the immunological synapse is formed, ICAM-3 molecules are redistributed to the outer zone of the contact region [26]. It has been suggested that exclusion of the ICAM-3/DC-SIGN complex is required for the close contact between the TCR and the MHC-peptide complex (TCR-pMHC), due to the relatively large size of the interaction pair (ICAM-3 ~ 15nm [27], DC-

SIGN ~ 33nm [28]) compared to that of the TCR/MHC-peptide complex (15nm [29]). After the initial contact, closer membrane apposition is achieved by the CD2/CD58 (LFA-3) interaction, which brings the cell membranes together at a distance similar to that of the TCR-pMHC (~ 15nm) [30]. CD2 belongs to the IgSF and it forms loose associations with the TCR/CD3 complex. Although the affinity between CD2 and CD58 is very low (10-20 μM), the high abundance and exceptionally fast dissociation rate constant ($k_{off} > 4s^{-1}$) of such molecules allows hundreds or thousands of weak interactions to bring the cells into close contact [31].

1.3.2 Proteins involved in antigen presentation and recognition

Activation of αβ T cells is initiated by the interaction of the TCR complex with the MHC-peptide complex. There are two types of MHC molecules – class I and class II – which present endogenous (derived from molecules inside the cell) and exogenous (derived from molecules outside the cell) antigens, respectively. However, cross-presentation can occur in special cases when exogenous antigens can be presented by MHC class I molecules [32]. The MHC class I molecule is a heterodimer consisting of α heavy chain and a β2 microglobulin (β2m) domain. The α1 and α2 domains of the heavy chain fold to form a membrane distal polymorphic peptide-binding groove and the α3 constant domain is membrane proximal and non-covalently interacts with the β2m domain. MHC class II molecules are also heterodimers comprising an α chain and a β chain. The α1 and β1 domain fold to form a membrane distal polymorphic peptide-binding groove. The processing of antigens and their presentation by MHC molecules is reviewed in [33].

The TCR of αβ T cells is a heterodimer consisting of an α chain and a β chain linked by disulphide bonds. Each chain comprises a membrane-distal

variable domain and a membrane-proximal constant domain. The variable domains form the highly variable complementarity determining regions (CDRs) that interact with the pMHC complex. The α and β chains have short cytoplasmic domains only and have no signalling capacity. Signalling by the TCR is initiated by phosphorylation of the immunoreceptor tyrosine based activation motifs (ITAMs) of the CD3 subunits (CD3ζ, ε, δ, γ). An ITAM is defined by a consensus sequence of [D/E]-x-x-(Y)-x-x-[I/L]-x(6–8)-(Y)-x-x-[I/L] (where "x" is any amino acid, and residues in square brackets are alternatives). The CD3 molecules all together form a complex with TCR αβ chains and, when phosphorylated, the complex recruits a number of intracellular signalling molecules with SH2 domains [34, 35].

1.3.3 Co-receptor molecules

CD8 and CD4 molecules are co-receptors expressed by cytotoxic T cells and helper T cells: CD8 binds to MHC class I molecules while CD4 binds to MHC class II molecules. CD8 is a dimeric molecule that either exists as an αα homodimer or αβ heterodimer, with each chain consisting of a single Ig-like V domain. In contrast, CD4 is an extended molecule with four Ig-like domains. The TCR and CD8/CD4 recognise different regions of the same pMHC. Compared with TCR-pMHC binding, biophysical studies by surface plasmon resonance (SPR; using a BIAcore instrument) showed that both CD8 and CD4 bind pMHC with a much lower affinity (200 μM for CD8, undetectable for CD4, versus 1-90 μM for TCR) and faster kinetics, ensuring that binding is dominated by the TCR and that T cell-APC interaction cannot be initiated by the co-receptor interactions alone. However, Wooldbridge et al. have proposed that the CD8/pMHC interaction is capable of stabilising the TCR-pMHC interaction. The most important function of the co-receptors results from the association of their cytoplasmic regions with LCK, which is a key molecule responsible for phosphorylation of ITAM of CD3 molecules [36, 37]. It is, however, suggested that CD8 and CD4 might be recruited to the

TCR-pMHC complex through interactions between CD3 ITAMs and the SH2 domain of LCK, in which case the recruitment of LCK might amplify incipient TCR signalling [38].

1.4 Models of T cell activation

One of the most important but controversial topics in T cell biology is how engagement between TCR and pMHC triggers signal transduction across the membrane, resulting in TCR phosphorylation. A number of models have been proposed, and they can be grouped into three categories, namely aggregation models, conformational change models and the segregation models, according to their underlying mechanisms.

There are two principal versions of the aggregation model, namely "heterodimerization" and ""pseudodimerisation" model. The "heterodimerization model" proposes that TCR triggering requires binding of CD4 or CD8 co-receptors to the same pMHC, which brings the associated SRC kinases into close proximity with the TCR. However, this is challenged by the observation that T cells remain functional in mice lacking co-receptors [39, 40]. Based on the observation that soluble covalently linked agonist/self-pMHC heterodimers are capable of triggering TCR activation, a modified model called "pseudodimerisation" postulates that the recruitment of a second coreceptor-associated TCR to a TCR/agonist pMHC complex, via binding of the coreceptor to the agonist pMHC, can lead to formation of a pseudodimer complex as the second TCR binds to the redundant self-pMHC [41]. However, this model again relies on the presence of coreceptor and it is challenged by the observation that TCR triggering can occur in the absence of self-MHC [42].

Early proposals of TCR triggering suggested that conformational changes may be involved as is the case for the activation of G-protein coupled receptors

and nicotinic acetylcholine receptors [43, 44]. However, unlike those receptors, the αβ chains of TCR contain very short intracellular region without any signalling motifs or structures with enzymatic activity. Also, ligand-independent TCR activation by soluble CD3 antibodies indicates that the αβ chains of the TCR are not essential for the triggering mechanism. Whilst it was suspected that conformational changes may take place in the CD3 chains, no convincing evidence for this has been shown. Although a conformational change of CD3ε has been claimed [45], this proposal has been brought into question by Fernandes et al. [46].

The kinetic-segregation (KS) model was proposed by Davis and van der Merwe (**Figure 1.1**) [47], following the initial observation by Springer that formation of TCR-pMHC complex requires segregation of TCR from the so-called glycocalyx components [48]. These glycocalyx elements include receptor tyrosine phosphatases (RTPs) CD45 and CD148, as well as cellular adhesion molecules such as LFA-1 and ICAM-1, which contain large glycosylated ectodomains and have a significantly larger size compared to the signalling complexes. Size-exclusion of the RTP from the signalling molecules was expected to have effects on signalling [48]. A second observation forming the foundation of the KS model is that treating resting T cells with pervanadate (a tyrosine phosphatase inhibitor), results in very strong tyrosine phosphorylation [49]. This suggests that the TCR is constitutively phosphorylated and dephosphorylated to maintain a low phosphorylation level, and that this equilibrium can be disrupted in favour of phosphorylation by kinases through blocking the phosphatases.

In the KS model, it is suggested that all surface proteins associate transiently and randomly in a resting T cell by diffusion on the membrane, which allows phosphorylation of the TCR by membrane- or coreceptor–associated SRC kinases, as well as dephosphorylation of phosphotyrosines by RTPs. However, net phosphorylation is kept at a too low level for activation but perhaps allows

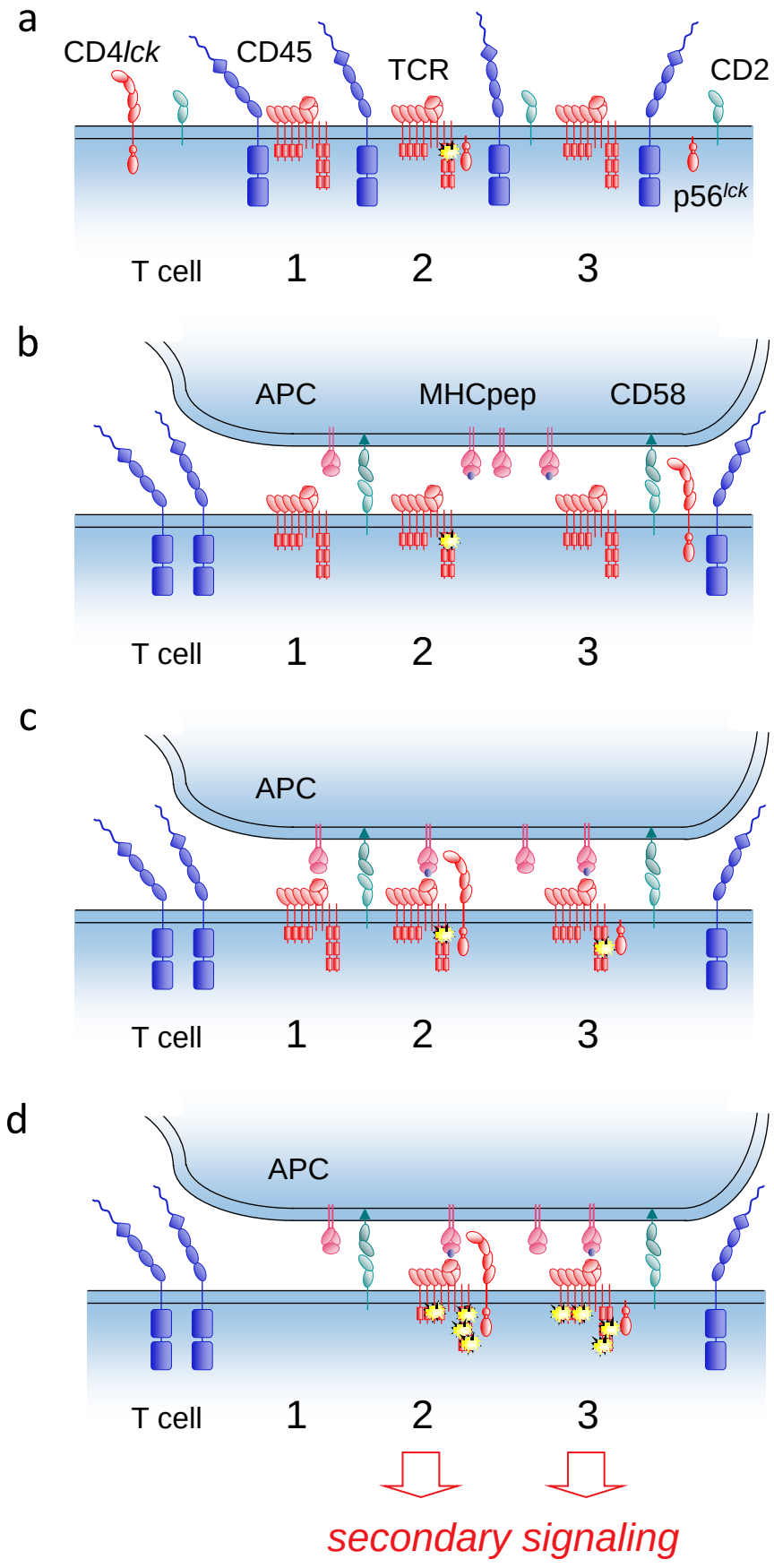

a
CD4lck
CD45
TCR
CD2
p56lck
T cell
1
2
3
b
APC
MHCpep
CD58
T cell
1
2
3
c
APC
T cell
1
2
3
d
APC
T cell
1
2
3
secondary signaling

Figure 1.1. The kinetic-segregation model (adapted from [47]). In a resting T cell (a), random protein interactions in the membrane lead to phosphorylation and dephosphorylation of molecules with tyrosine-phosphorylation motifs by SRC kinases and tyrosine phosphatases. Triggering occurs as the local balance of those constitutive processes is altered by the formation of close-contact zones (b–d). This process is 'nucleated' by small proteins such as CD2 and results in the local size-dependent exclusion of large proteins such as CD45 from the close-contact zone. The 'fates' of the three TCRs here (1, 2, 3) are presented according to the KS model. In the absence of peptide-MHC contact, TCR 1 diffuses from the contact zone. If TCR 1 had been phosphorylated in the close-contact zone or before its formation, rapid dephosphorylation outside the close-contact zone would prevent signalling. TCR 2, phosphorylated before close-contact zone formation, binds cognate MHC-peptide and is thereby 'held' in the close-contact zone long enough for 'downstream' events to occur. TCR 3, phosphorylated after close-contact zone formation, is phosphorylated by free LCK only, accounting for coreceptor-independent triggering. Relatively large amounts of phosphorylation of TCR 2 and TCR 3 lead to 'downstream' signalling after ZAP70 recruitment.

survival. Upon engagement of TCR and the pMHC, aided by apposition by CD2-CD58 interaction, a close contact zone is formed and RTPs are excluded by size. Within the close contact zone, unphosphorylated TCRs are as likely to be phosphorylated as those outside the contact zone, but those phosphorylated are now less likely to be dephosphorylated. This therefore increases the net phosphorylation in the contact zone, which allows T cell activation. TCRs failing to engage with the proper pMHC will diffuse out of the contact zone and will be incapable of contributing to downstream signalling, and so the half-lives of the TCR phosphorylation are dependent on the half-lives of the TCR-pMHC interaction.

The KS model has been supported by a number of experiments. Firstly, evidence of the role of size exclusion of RTP has been provided by Irles et al. and Lin et al., who have shown that TCR signalling is inhibited when a chimeric CD45 containing a small ectodomain is introduced into CD45-deficient T cells [50] or when CD148 is truncated [51], respectively. Interestingly, Irles et al. showed that a CD45 chimera containing the large ectodomain of CD43 is capable of restoring TCR activation [50]. Choudhuri et al. tested the model from another angle by elongating the pMHC ectodomain, thus increasing the inter-membrane distance upon TCR/pMHC engagement, which significantly reduces TCR triggering without affecting TCR/pMHC ligation [52]. However, further investigation is required to show whether segregation alone is sufficient for receptor triggering.

1.5 The immunological synapse

The contact interface between a T cell and an APC is called the immunological synapse (IS). This structure is generally characterised by three concentric rings, forming a bull's eye pattern. The inner region is called the central supramolecular activation cluster (cSMAC), which is further refined into the cSMACa and cSMACb. While cSMACa contains the TCR and is devoid of signalling molecules, CD28 and

PKCθ and LCK (intracellular signalling molecules) are found in cSMACb which locates at the periphery of cSMACa. A ring of integrins (including LFA-1 and talin) termed the peripheral supramolecular activation cluster (pSMAC) surrounds the cSMAC. The region outside the pSMAC is called the distal supramolecular activation cluster (dSMAC), which is characterised by a ring of molecules with large ectodomains such as CD45 [53].

However, interestingly, a bull's eye pattern is rarely observed at the contact site between T cells and dendritic cells, which are professional APCs capable of providing strong activation to T cells. Instead, "multifocal ISs" with multiple sub-micron sized contact spots are observed at the T-DC synapse, without large-scale segregation of TCR and integrins. This observation implies that the formation of the cSMAC with focal concentration of TCR is not crucial for providing sustained TCR signalling [54]. Recent studies have shown that, following TCR triggering, TCR microclusters form at the periphery of the contact site, which contain signalling molecules such as ZAP-70 and SLP-76 occurs. These clusters serve as sites that sustain signalling but are later transported centripetally towards the cSMAC through an actin cytoskeleton–driven process, during which dissociation of ZAP70 and SLP-76 occur, resulting in very weak or no signalling in the cSMAC [55, 56]. Mossman et al. have shown that restricting the movement of TCR-containing microclusters towards the cSMAC results in prolonged signalling [57]. All this argues against the idea that the IS is responsible for initiating or sustaining TCR signalling. It has been proposed, alternatively, that the cSMAC might play a role in targeting engaged TCRs for endocytosis and degradation. This is supported by the evidence that high concentrations of ubiquitin and the E3 ubiquitin ligase Cbl-b are observed within the cSMAC [58]. However, it is also reported that the cSMAC mediates down-regulation of signalling upon interaction with a strong agonist pMHC, while it can mediate enhancement of signalling when the agonist is weak [59].

It is also suggested that the immunological synapse mediates the site-directed secretion of T cell effectors molecules upon activation. For example, Sabatos et al. have observed secretion of IL-2 through the synaptic region, which aids in cell activation [60]. The IS structure is also observed between CTLs and their target cells. However, inside the pSMAC of these cells are two separate zones: one is the cSMAC concentrated with TCR while the other serves as a secretory domain that allows delivery of lytic granules (containing perforins and granzymes) into the contact zone. The intergrin-rich pSMAC is proposed to serve as a sealing ring that prevents diffusion of the granular contents beyond the IS and thus maximises cytotoxic activities [61, 62].

Finally, Maldonado et al. [63] have proposed that the IS may be involved in T cell differentiation. As previously discussed in **Section 1.2.2**, IFN-γ and IL-4 are cytokine receptors that play an important role in commitment of naïve T cells into Th1 and Th2 cells, respectively. Upon TCR engagement, IFN-γ receptors rapidly co-localise with the TCR in the IS. However, in the presence of IL-4, this co-localisation is prevented, thus favouring, it is suggested, differentiation of T cells into Th2 cells.

1.6 The key signalling receptors in T cells

In T cells, there are a number of receptors that have been under intensive investigation and characterised as being the most crucial for the control of cell proliferation and activation. The absence or mutation of these receptors can usually impair survival of the cell or cellular activity, as well as induction of auto-immune diseases in the host. Signalling by these receptors involves tyrosine phosphorylation in each case.

1.6.1 TCR

As previously discussed, the TCR exists as a complex consisting of a variable αβ heterodimer associated non-covalently with a copy of each of the CD3δε, CD3γε and CD3ζ-ζ invariant dimers. It is believed that while the αβ chains are responsible for ligand recognition, the CD3 dimers are responsible for downstream signal transduction through their cytoplasmic ITAM motifs [immunoreceptor tyrosine-based activation motif, previously known as antigen recognition activation motif (ARAM) motifs] [35]. An ITAM is defined by a consensus sequence of [D/E]-x-x-(Y)-x-x-[I/L]-$x_{(6-8)}$-(Y)-x-x-[I/L] (where "x" is any amino acid, and residues in square brackets are alternatives). While each CD3 δ, ε and γ subunit has one ITAM only, each CD3ζ contains three ITAMs. There are 10 of these motifs present in the TCR complex [CD3 –ζ (x6), -ε (x2), -γ, and –δ].

Although it is still controversial how TCR/pMHC engagement triggers TCR activation, and how the extracellular signal is transduced across the plasma membrane, it is generally accepted that intracellular signal transduction is initiated by phosphorylation of ITAM tyrosine residues by SRC family kinases, such as LCK and Fyn [35, 64]. ITAMs on ε- and ζ-chains have been shown to function as docking sites for the tandem SH2 domains of ZAP-70, which is an intracellular protein tyrosine kinase [65]. Thereafter, ZAP-70 is activated by LCK-mediated phosphorylation at Tyr-493 in its activation loop [66, 67], and the active ZAP-70 subsequently phosphorylates at least two critically important adaptor proteins called linker for activation of T cells (LAT) and SRC homology 2 (SH2) domain-containing leukocyte protein of 76 kDa (SLP-76) [68]. Phosphorylated LAT in turn serves as a docking site for recruitment of other signalling molecules to form signalling complexes, which initiate downstream signalling pathways [69, 70]. Defects in ZAP70 result in severe combined immunodeficiency (SCID), characterised by the absence of peripheral $CD8^+$ T

cells and the presence of peripheral $CD4^+$ T cells resistant to TCR-mediated activation, which indicates a crucial role of ZAP70 in T cell development and function [71].

1.6.2 LAT

LAT is one of the most important members of the family of transmembrane adaptor proteins (TRAPs), which usually contain multiple tyrosine-containing motifs. LAT is heavily phosphorylated after TCR/pMHC engagement, and thereafter serves as a scaffold to recruit a wide range of signalling molecules, leading to formation of a large signalling complex that links TCR activation to downstream signalling pathways. Functional analysis showed that LAT plays an important role in Ca^{2+} flux, Erk activation, CD69 up-regulation, as well as activation of the transcription factors activating protein 1 (AP-1) and nuclear factor of activated T cells (NFAT)-mediated gene transcription [72].

LAT contains a short extracellular region (4 amino acids), a transmembrane domain, and a large cytoplasmic domain (235 amino acids) containing ten tyrosine residues, and it lacks any intrinsic enzymatic activity. LAT contains two conserved cycteine residues, Cys26 and Cys29, in the juxtamembrane region. These two residues are essential for palmitoylation, which targets the protein to putative sub-domains of the plasma membrane called glycol-lipid enriched micro-domains (GEMs). Poor tyrosine phosphorylation was observed in LAT mutants that could not be positioned at the GEMs [73], although only defects in Cys26 actually impair LAT function and signalling through the TCR. As the SRC kinase LCK is also palmitoylated and localized at the GEMs, a role for LCK in phosphorylating LAT tyrosines is plausible [74].

Studies have shown that, among the ten tyrosines, only the five membrane-distal tyrosines [YHNP (Y127), YLVV (Y132), YVNVP (Y171), YVNVS

(Y191) and YENL (Y226)] are phosphorylated [75, 76]. The importance of LAT-mediated PLCG1 signalling is reflected by defects in Ca^{2+} flux when the four PLCG1binding sites (Y132, Y171, Y191 and Y226) are mutated.

Mutations in Y171 and Y191 can result in deleterious effects on Ca^{2+} flux by abolishing the binding of GRB2-related adapter protein (GADS) to LAT. Upon TCR activation, phosphorylated SLP-76 is recruited to LAT indirectly through interaction between the proline-rich region of SLP-76 and the SH3 domain of GADS [75]. Interleukin-2 tyrosine kinase (ITK), a Tec family tyrosine kinase, is then recruited to the LAT complex via binding of its SH2 domain to the phosphotyrosine of SLP-76. ITK in turn phosphorylates and activates PLCG-1, leading to production of two second messengers, inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG), through hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP_2) [77, 78]. IP_3 then triggers release of Ca^{2+} from the endoplasmic reticulum (ER), increasing intracellular [Ca^{2+}]. Binding of Ca^{2+}-bound calmodulin activates a phosphatase called calcineurin (Cn), which in turn activates the nuclear factor of activated T cells (NFAT)–mediated transcription of IL-2 and other genes that are required for T cell proliferation or development. [79]. On the other hand, the lipid messenger DAG recruits members of the protein kinase C (PKC) family, notably PKCθ, to the IS and mediates their activation [80]. Stimulation of JNK and IKKβ by activated PKCθ results in activation of AP-1 and NF-κB, respectively. This eventually leads to expression of IL-2 [81].

GRB2 is another key molecule recruited to LAT [75]. The importance of GRB2 is indicated by the findings that (1) LCK activity and TCR-mediated signalling are significantly attenuated in $GRB2^{-/-}$ T cells, and (2) thymic positive and negative selection of T cells are impaired in $GRB2^{-/-}$ mice, which die during early embryogenesis [82]. GRB2 is an adaptor protein consisting of a central SH2 domain flanked by two SH3 domains. LAT has three sites for the binding of GRB2

(Y171, Y191 and Y226), each of which contains the GRB2- and GADS- binding consensus sequence of YXNX (where X is any amino acid) [83].

After binding to LAT, GRB2 interacts through its SH3 domain with SLP-76 as well as Sos1, which is a guanosine triphosphate (GTP)-guanosine diphosphate (GDP) exchange factor for Ras. Recruitment of Sos1 eventually results in activation of extracellular signal-regulated kinase 1 and 2 (Erk1/2) through sequential activation of Raf kinase and mitogen-activated protein kinase (MAPK) kinase MEK. Erk1/2 has been shown to play an important role in diverse biological processes, including mRNA translation, cell proliferation and survival, through targeting a large number of substrates that include nuclear transcription factors, cytoskeletal proteins, signalling proteins and receptors [84].

The interaction between GRB2 and Sos1 not only initiates the ERK/MAPK signalling pathway (as discussed above), but also mediates LAT oligomerisation upon TCR activation through formation of a 2:1 GRB2-Sos1 complex. Disruption of the 2:1 association by overexpressing the C-terminal proline-rich domain of Sos1 in Jurkat E6.1 cells results in reduced levels of Ca^{2+} flux after TCR activation, which suggests that LAT clustering could be important for TCR-mediated signalling *in vivo* [85].

1.6.3 Co-stimulatory/co-inhibitory receptor molecules

While TCR triggering initiates T cell activation, a 'two signal' hypothesis proposed that this is not sufficient to result in optimal T cell activation, and a second set of stimuli, either provided by soluble molecules or membrane-bound receptors, is required [86]. So far, a number of receptor molecules have been identified that mediate stimulatory or inhibitory signalling and are thus called co-stimulatory/co-inhibitory receptors, and that affect the overall outcome of T cell responses. All of these receptors are type-I transmembrane proteins with an extracellular V-set

(variable domain–like) IgSF domain. Because none of these receptors possess any intrinsic enzymatic activity within the cytoplasmic tail, it is believed that signal transduction is mainly mediated through recruitment of SH2 domain- or SH3 domain-containing proteins to the phospho-tyrosine motifs or proline-rich motifs in their cytoplasmic tails. Our laboratory favours the hypothesis that phosphorylation of these receptors is also mediated by large phosphatase exclusion.

1.6.3.1 CD28

CD28 was the first identified co-stimulatory receptor in T cells, and is the most important one needed for T cell activation, proliferation and survival. Instead of providing a distinct second signal that induces another transcriptional profile, CD28 signalling amplifies the transcriptional effect of TCR signalling [87]. CD28 signalling not only lowers the activation threshold for T cell activation to allow triggering of the TCR by low-affinity ligands, but also allows reduction of the number of TCRs to be triggered for T cell activation. A lack of signalling through CD28 results in a non-responsive, anergic state or cell death [88]. Studies showed that T cells derived from CD28-deficient mice have impaired responses to lectins and the activity of helper T cells was attenuated [89]. However, the absence of CD28 does not lead to complete blockade of T cell function, suggesting that the role of CD28 can be partially compensated for by alternative co-stimulatory pathways. CD28 is constitutively expressed at the cell surface as a disulphide-linked covalent homodimer. Each CD28 chain has a single V-set IgSF domain as well as a cytoplasmic tail bearing multiple tyrosine residues [90]. Signalling is mainly mediated through recruitment of SH2 and SH3 domain-containing signalling molecules to the phosphorylated tyrosine motifs and the proline-rich motifs in its cytoplasmic tail, respectively [91].

CD28 and another co-stimulatory receptor CTLA-4 (discussed in **Section 1.6.3.2**) bind the same ligands expressed on APCs, i.e. CD80 (B7-1) and CD86 (B7-2), through a conserved transmembrane proximal hexapeptide (MYPPPY) binding motif [92]. CD80 and CD86 each consist of two IgSF domains, i.e. a receptor binding V-set domain and a membrane-proximal C1-set (constant domain-like) domain, connected to a single transmembrane domain and a cytoplasmic domain [93]. However, while CD80 is present as a mixed population of non-covalent dimers and monomers on the cell surface, CD86 is only present as monomers. Previous biophysical studies have shown that the affinity of CD80 for CTLA-4 is 20 times stronger than that for CD28 (0.2 μM vs 4 μM) and that the affinity of CD86 for CTLA-4 is 8 times stronger than that for CD28 (2.6 μM vs 20 μM), and (2) that, while CD28 binds monovalently to CD80 and CD86, CTLA-4 binds bivalently. In other work it has been shown that while CD28 is expressed constitutively, expression of CTLA-4 is often delayed by 24 hours. Computer simulations incorporating all these findings predicted that whereas CD86 primarily binds CD28, CD80 primarily binds to CTLA-4 [90, 94, 95].

As previously mentioned, CD28 possesses two important signalling motifs within the cytoplasmic tail: the membrane proximal YMNM motif and the membrane distal PYAPP motif. The importance of YMNM motif is highlighted by the two consensus sequences contained within the motif, i.e. the phosphoinositide 3-kinase (PIK3) binding-motif YXXM [96] as well as the GRB2/GADS-binding motif YXNX [97]. These interactions have subsequently been verified [98, 99]. As signalling through GRB2 has been summarized in **Section 1.6.2**, only PIK3 signalling is discussed here.

PIK3 is a heterodimer consisting of a p110 catalytic subunit and an associated adaptor subunit (p85) with two SH2 domains. PIK3 generates D3 phospholipids by phosphorylating the D-3 position of the inositol ring of phosphoinositides-4-phosphate (PI4P) and $PI(4,5)P_2$, generating $PI(3,4)P_2$ and

$PI(3,4,5)P_3$, respectively. These D3 lipids are anchored to the inner surface of plasma membrane, which allows them to recruit proteins with pleckstrin-homology (PH) domains, such as PtdIns 3-Phosphoinositide-dependent protein kinase-1 (PDK1) and protein kinase B (PKB; also known as Akt). While activation of PDK1 is mediated by auto-phosphorylation at a conserved serine residue (S241) in the activation loop [100], binding of D3 lipids relieves structural auto-inhibition and exposes the Akt activation loop to PDK1. Activation of these two molecules initiates multiple pathways that regulate diverse cellular functions, including protein synthesis (through activation of mTOR and S6 kinase), cell survival (by upregulating the prosurvival factors Bcl-2 and Bcl-X_L), proliferation (through activation of PKCθ) and glucose metabolism (by activating glycogen synthase kinase 3 β/ GSK3β). Regulation of PIK3 signalling is mediated by phosphatase and tensin homolog deleted on chromosome ten (PTEN), which dephosphorylates the D3 position of inositol and phosphatidylinositol substrates. [101-104].

However, despite the association of important signalling molecules with the YMNM motif, a recent study showed that this motif has a redundant role in CD28 signalling. By generating gene-targeted knock-in mice expressing a Y>F mutant for the YMNM motif, Dodson et al. [105] showed that, although binding of PIK3 and the CD28-dependent serine phosphorylation of PKB/Akt are impaired, no defects in IL-2 expression, Bcl-X_L expression or T cell functions were observed. In contrast, mutation of the distal proline-based motif results in markedly impaired IL-2 production, as well as attenuated phosphorylation of PKCθ and GSK3β. Moreover, humoral responses were profoundly impaired in knock-in mice expressing this mutant, and less severe disease manifestations in experimental allergic encephalomyelitis (EAE) were observed. These results suggest that the distal proline-based motif, instead of the proximal YMNM motif dominates CD28-mediated signalling. Unaffected PKCθ and GSK3β phosphorylation in the YMNM

mutant may be due to an alternative mode of signalling mediated by SRC kinases recruited to the proline-based motif [106, 107].

1.6.3.2 CTLA-4

Cytotoxic T lymphocyte-associated antigen-4 (CTLA-4), also known as CD152, is a disulphide-linked covalent homodimer. Like CD28, each CTLA-4 chain is composed of a single V-set IgSF domain as well as a cytoplasmic tail that contains two tyrosine-containing motifs as well as a proline-rich region [108]. However, while CD28 is constitutively expressed on the cell surface, CTLA-4 molecules are primarily localized in intracellular compartments and only small amounts are transported to the cell surface following T cell activation [109]. While CTLA-4 shares the same ligands as CD28, CTLA-4 is a co-stimulatory molecule that plays an inhibitory role in down-regulating T cell activities. This is revealed by a study in which CTLA-4 knock-out mice rapidly develop lympho-proliferative disease with massive tissue infiltration and multi-organ destruction, severe myocarditis and pancreatitis, and die by 3-4 weeks of age [110]. *In vitro* studies showed that CTLA-4 signalling inhibits T cell proliferation by interfering with the ERK and JNK signalling pathways, thereby attenuating activation of AP-1 and NF-κB and impairing IL-2 production [111, 112]. Several mechanisms have been proposed to explain the inhibitory function of CTLA-4.

As CTLA-4 possesses two tyrosine residues within the cytoplasmic tail, initial studies focused on indentifying interactions between SH2 domain-containing proteins and CTLA-4. In this approach, PIK3 [113] and SH2 domain-containing tyrosine phosphatase-2 (SHP2; also known as SYP) [114] were shown to associate with the YVKM motif. Moreover, blockade of CTLA-4 inhibits the activation of double-positive thymocytes, but enhances the activation of single-positive thymocytes, implying a positive role of CTLA-4 in thymocyte development

at the double-positive stage [115]. Association of SHP2 with CTLA-4 has remained controversial, as SHP2 has been reported to bind to a consensus I/VxYxxI/V/L motif, and no direct association between the phosphorylated YVKM and YFIP motifs of CTLA-4 and SHP2 has been observed. This implies that association between CTLA-4 and SHP2 occurs in an indirect manner [116]. In addition, as in the case of PIK3, how SHP2 might contribute to the inhibitory function of CTLA-4 has been obscure, as SHP2 has been shown to positively regulate the TCR-induced ERK/MAPK signalling pathway [117]. In fact, studies have shown that phosphotyrosine is not required for the inhibitory function of CTLA-4, as Y>F mutations in YVKM and YFIP motifs do not affect T cell inhibition, and the inhibitory function of CTLA-4 is preserved in LCK- or ZAP-70-deficient T cells [118].

While SHP2 does not seem to account for the phosphatase activity associated with the cytoplasmic tail of CTLA-4 and its inhibitory function, emerging evidence support such a role for protein phosphatase 2A (PP2A) since its interaction with CTLA-4 was identified [119]. PP2A comprises a family of serine/threonine phosphatases. They are holoenzyme complexes composed of a regulatory subunit $PP2A_B$ and a dimeric core enzyme, which consists of a catalytic subunit $PP2A_C$ and a scaffolding subunit $PP2A_A$ (acting as a scaffold for $PP2A_B$ binding) [120]. PP2A interacts with CTLA-4 through binding of the regulatory subunits to a conserved three-lysine motif in the juxtamembrane region within the cytoplasmic tail, and the catalytic subunit binds to the YVKM motif. But how does PP2A mediate the inhibitory function of CTLA-4? It is proposed that CTLA-4 molecules are associated with an inactive form of PP2A inside the cell, and upon TCR activation and engagement of CTLA-4, phosphorylation of $PP2A_A$ releases PP2A from CTLA-4. The dissociated PP2A becomes active and can inactivate PKB/Akt [119, 121].

Another suggestion is that CTLA-4 may partially function by outcompeting CD28 for their shared ligands. Biophysical studies have shown that the affinity of CD80 and CD86 for CTLA-4 is 20-fold and 8-fold higher than that of CD28, respectively [90], so surface expression and engagement of CTLA-4 on the cell surface could impair CD28-dependent stimulatory signals. This is supported by the findings that expression of CTLA-4 lacking a cytoplasmic domain rescues mice from massive tissue infiltration and multi-organ destruction, and that the mice instead develop lymphoadenopathy [122]. However, this also indicates that complete functioning of CTLA-4 requires both the ectodomain and the cytoplasmic domain.

On the other hand, Schneider et al. proposed a "reverse stop-signal model" in which, whereas TCR engagement provides a "stop signal" and reduces T cell mobility, CTLA-4 ligation reverses this "stop signal" and increases their mobility [123]. This means T cells cannot form long-term interactions with pMHC complexes, limiting their productive signalling and activation. Consistent with this finding, they observed that co-ligation of TCR and CTLA-4 results in impaired formation of ZAP-70-containing microclusters as well as a major reduction in Ca^{2+} mobilisation [124].

Finally, Qureshi et al. showed that CTLA-4 can act as an effector molecule that inhibits CD28 costimulation by the cell-extrinsic depletion of the shared ligands (CD80 and CD86). After these costimulatory ligands are captured from the interacting cells by trans-endocytosis, they are degraded inside the CTLA-4-expressing cells [125].

1.6.3.3 PD-1

Programmed death-1 (PD-1) is a cell surface monomer consisting of a single V-set IgSF domain as well as a cytoplasmic tail that contains two tyrosine residue

based signalling motifs. The N-terminal sequence VDYGEL and the C terminal sequence TEYATIV belong to the families of immunoreceptor tyrosine-based inhibitory motifs (ITIMs; with a consensus sequence of [I/L/V]-x-(Y)-x-x-[L/V] [126]) and immunoreceptor tyrosine-based switch motifs (ITSMs; with a consensus sequence of [T]-x-(Y)-x-x-[V/I] [127]), respectively. Like CTLA-4, PD-1 is not constitutively expressed, and expression is induced upon TCR activation.

PD-1 is another inhibitory co-receptor with a critical role in autoimmunity. This was firstly revealed when aged C57BL/6(B6)-PD-$1^{(-/-)}$ congenic mice developed lupus-like proliferative arthritis and glomerulonephritis, and when PD-$1^{-/-}$ H-$2^{b/d}$ mice (with an autoimmune-prone background) developed chronic and systemic graft-versus-host-like disease [128]. A role for PD-1 in the induction and/or maintenance of tolerance is further supported by the finding that Pdcd$1^{-/-}$ $Bal^{b/c}$ mice develop dilated cardiomyopathy due to the production of an autoantibody against cardiac troponin [129, 130]. Although PD-1 and CTLA-4 signalling pathways are critical for the induction of tolerance, PD-1 is responsible for its long-term maintenance [131].

The ligands for PD-1, which are PD-L1 (B7-H1, CD274) and PD-L2 (B7-DC, CD273), have different expression patterns. Whereas PD-L1 is widely and constitutively expressed on T cells, B cells, dendritic cells, macrophages, mesenchymal stem cells and bone marrow-derived mast cells, PD-L2 is only inducibly expressed on dendritic cells, macrophages and bone marrow-derived mast cells [132]. Studies have shown that interaction between PD-1 and PD-L1, but not PD-L2, seem to be predominantly responsible for regulating the severity of diseases. Murine experimental autoimmune encephalomyelitis (EAE) is a model for a T cell–dependent autoimmune disease called multiple sclerosis, in which Th1 cells mediate inflammatory damage in the central nervous system (CNS). Alan et al. have shown that blockade of PD-1 results in augmented antigen-specific T cell

expansion, activation, and cytokine production. Also, disease augmentation is observed when PD-L2, but not PD-L1, is blocked [133].

The outcome of PD-1-mediated signalling depends on the strength of TCR, CD28 and IL-2 signalling. *In vitro* experiments have shown that strong CD28 signalling is required to overcome the PD-1 inhibitory effect at low levels of TCR stimulation, whereas, under saturating conditions of TCR activation, inhibition is impaired even in the absence of CD28 signalling [134]. In contrast, exogenous IL-2 is able to overcome the PD-1-mediated inhibitory effect even at low concentrations, indicating that the PD-1 inhibitory pathway does not interfere with IL-2 responsiveness [135].

The two cytoplasmic tyrosine-based signalling motifs, i.e. the membrane-proximal ITIM and membrane-distal ITSM, are both phosphorylated upon ligand engagement. Surprisingly, mutation of the ITIM (inhibitory motif) had little effect on PD-1 signalling or functional activity. In contrast, mutation of the ITSM leads to impairment of PD-1-mediated inhibition of cytokine expression and proliferation, indicating the dominant role of the ITSM in this signalling pathway [136]. Consistent with the inhibitory function of PD-1, two phosphatases, namely the SH2-domain containing tyrosine phosphatase 1 (SHP1) and SH2-domain containing tyrosine phosphatase 2 (SHP2), have been shown to bind phosphorylated PD-1. Whereas both SHP1 and SHP2 bind the ITSM, the ITIM binds only the SHP2 [136]. Studies have shown that PD-1 signalling results in inhibition of TCR-mediated phosphorylation of the ZAP70/CD3ζ signalosome, PKCθ as well as ERK, which explains why IL-2 production is impaired upon TCR/PD-1 co-ligation [126]. The concentration of PD-1 at the IS [137] also implies that PD-1 associated phosphatases may directly dephosphorylate signalling molecules involved in TCR- and CD28-mediated signalling pathways.

In addition to the phosphatase-mediated inhibition discussed above, the PD-1/PD-L1 interaction has also been shown to impair productive stimulatory signalling by reversing the TCR-induced stop signal and increasing T cell mobility [138], a mechanism previously described for CTLA-4/B7 interactions [123]. However, interestingly, this study reported that blockade of CTLA-4 does not alter T cell mobility or abrogate tolerance. One major difference is that while naïve T cells were used in the previous CTLA-4 study, anergic T cells were used in the PD-1 study, which implies a critical role of CTLA-4 in regulating naïve T cell activation.

1.6.3.4 BTLA

B and T lymphocyte attenuator (BTLA; CD272) is a monomeric transmembrane receptor that consists of a single V-set domain and a cytoplasmic tail containing three conserved tyrosines. The membrane-proximal tyrosine motif at Y226 has a YXNX consensus sequence, which is a potential binding site for GRB2 [97], whilst the distal tyrosines at Y257 and Y282 belong to the ITIM and ITSM motif families, respectively [139]. While it was initially thought that BTLA is not expressed by naïve T cells and only induced after T cell activation [139], recent studies have shown that BTLA is expressed by both naïve and activated T cells, although expression is highly up-regulated in the latter [140, 141].

Consistent with the presence of ITIM and ITSM motifs as in PD-1, BTLA has been shown to be a third inhibitory receptor in T cells. This is evidenced by the fact that BTLA-deficient T cells showed increased proliferation and that BTLA-deficient mice have increased susceptibility to EAE [139]. Although deficiency of BTLA does not seem to trigger spontaneous autoimmunity, development of an autoimmune hepatitis-like disease has been reported in BTLA-deficient mice with a specific genetic background [142]. It is suggested that, like PD-1, BTLA plays a role in maintaining peripheral tolerance and in fine-tuning effector responses.

BTLA has a unique ligand called herpesvirus entry mediator (HVEM), which is a tumor necrosis factor receptor (TNFR) family member [143]. HVEM is a costimulatory molecule, which interacts with four other molecules [CD160, glycoprotein D, LIGHT (lymphotoxin-like, inducible expression, competes with herpes simplex virus glycoprotein D for HVEM, a receptor expressed by T lymphocytes) and lymphotoxin α] in addition to BTLA [144]. The interaction between BTLA and HVEM therefore allows for bidirectional signalling: an inhibitory signal is induced in T cells, whilst stimulatory signals are induced in the APCs. In addition to a *trans* interaction between BTLA and HVEM, which may exist as a ternary complex, a *cis* interaction might also occur in T and B cells where both BTLA and HVEM are expressed on the same cell [145]. A dynamic switch between the two forms of interaction may control inhibitory signalling by BTLA.

Like PD-1, the ITIM and ITSM motifs in the cytoplasmic tail of BTLA recruit both SHP1 and SHP2 [146]. However, whilst binding of SHPs to PD-1 can occur in the presence of only ITIM or ITSM [126], mutation of either of them in BTLA impairs the interactions, indicating that association of the SHPs to BTLA can only occur in the presence of both motifs [146]. BTLA has been reported to interact with the TCR complex (through CD3ζ) within the immunological synapse, and co-ligation of BTLA with the TCR complex results in a significant reduction in CD3ζ phosphorylation. This suggests that BTLA may mediate inhibitory functions through inhibition of the TCR-mediated signalling pathway [147]. Whereas CTLA-4 and PD-1 have been shown to exert their inhibitory functions at least partially through reversing the TCR-mediated stop signal [123, 138], this has not been reported for BTLA as yet.

In addition to the inhibitory function of BTLA, a role in sustaining the survival of T cells has been also reported. In support of this, Hurchla et al. reported that, in the non-irradiated parental-into-F1 model of acute graft-vs-host disease (GVHD), reduced survival of donor T cells was observed for BTLA-

deficient cells or when the mice were treated with anti-BTLA antibodies. This indicates that BTLA may have certain pro-survival effects under conditions of chronic stimulation [148].

1.6.3.5 ICOS

Inducible costimulator (ICOS, CD278) is, like CD28 and CTLA-4, a disulphide-linked homodimeric transmembrane protein, consisting of an extracellular V-set domain and a cytoplasmic tail containing two tyrosine-containing motifs. ICOS has got a unique ligand called LICOS [homolog of B7 (B7h), B7-related protein-1 (B7RP-1), CD275] [149, 150], which is expressed on a wide range of cells, including resting B cells, dendritic cells and activated monocytes [151]. Comparison in amino-acid sequence showed that ICOS shares 39% similarity with CD28 and it has been found to enhance T cell responses in the manner of CD28. However, unlike CD28, which is constitutively expressed on the cell surface, ICOS is expressed at very low levels on resting T cells but rapidly induced following T cell activation [152]. Expression of ICOS is regulated by TCR/CD28-mediated signalling, including the FYN-calcineurin-NFATc2 and MEK2-ERK1/2 pathways [153]. The importance of ICOS has been highlighted by the findings that ICOS-deficient T cells have impaired activation and proliferation capacity, and that ICOS-deficient mice are defective in germinal centre formation and antibody class switching and production, and become severely susceptible to EAE [154-156]. Moreover, ICOS-deficiency has also been found to be one of the causes of common variable immunodeficiency (CVID) in humans, which is characterized by hypogammaglobulinemia [157]. These all indicate that ICOS plays a crucial role in humoral immune responses as well as the control of immune tolerance.

Whether signalling through ICOS affects IL-2 production in the way CD28 does has been under debate: while early studies reported that engagement of

ICOS does not affect IL-2 production [149, 158], it was later found that ICOS-deficient cells are defective in IL-2 production [154]. Whereas ICOS plays an important role in the effector and regulatory activities of Th1, Th2 and Treg cells by inducing secretion of a number of cytokines, including IFN-γ, IL-4, IL-5, IL-13, and IL-10 [159], in the early stages of T cell priming, ICOS signalling seems to preferentially drive the differentiation of Th2 cells [160]. This is consistent with the findings that expression of NFATc1 [161] and GATA-3[162], which mediates production of IL-4 and IL-4 receptor, respectively, are selectively upregulated at the T cell differentiation stage. On the other hand, expression of c-Maf, which is the first Th2-specific transcriptional factor controlling IL-4 production, seems to be regulated by ICOS at the effector phase [161].

Signalling of ICOS is mediated mainly by the membrane-distal YMFM motif, which is highly similar to the PIK3-binding YMNM motif in CD28. It has been shown that IL-4 production, follicular B helper T cells (T_{FH}) generation, germinal centre formation, antibody class switching, and antibody affinity maturation are impaired in knock-in mice expressing a Y>F mutant form of ICOS [163]. This indicates that recruitment of PIK3 to the YMNM motif is crucial for ICOS signalling. Whereas both CD28 and ICOS are capable of recruiting PIK3, ICOS has been shown to recruit a specific form of PIK3, which contains a p50α regulatory subunit, instead of the p85α and p55α subunits recruited by CD28. Although both p50α and p85α are expressed on resting T cells, only p85α binds to CD28 upon receptor ligation. Compared to the p85α-containing PIK3, the p50α-containing PIK3 is shown to have more potent lipid kinase activity, leading to stronger induction of Akt phosphorylation, and accumulation of this PIK3 isoform at the immunological synapse is only seen in activated and not in resting T cells [164]. Interestingly, it has been shown that constitutive expression of ICOS in CD28-deficient T cells or mice is capable of rescuing most of CD28-dependent responses

[165]. These altogether suggest that ICOS can largely substitute or bypass the requirement of CD28, but is indispensible in activated T cell responses.

1.7 Phosphotyrosine signalling in metazoans

Since 1979, when Hunter and co-workers identified phosphotyrosine as a new type of post-translational modification mediated by protein kinases [166], tyrosine phosphorylation has emerged as a fundamentally important mechanism in metazoans to convey signals in response to mostly external cues. Signalling through this mechanism plays a central role in many key cellular processes, including cell proliferation, differentiation and development, transcriptional activation, cell cycle progression and metabolic homeostasis [167]. Phosphotyrosine-mediated signalling involves protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs) that phosphorylate and dephosphorylate substrates, respectively, as well as the modular protein domains that recognize the phosphorylated ligand and thereby recruitment of proteins which contain these domains to specify downstream signalling events [168]. SH2 domains form parts of a wide range of proteins, including kinases, phosphatases, adaptors, scaffolds, as well as proteins that are involved in signal regulation, ubiquitination, cytoskeletal regulation, phospholipid second messenger signalling, chromatin remodelling and small GTPase signalling [169], underlining the importance of this signalling system.

1.7.1 Evolution of phosphotyrosine-based signalling

The phosphotyrosine-based signalling system is estimated to have only appeared ~600 million years ago, just prior to the emergence of metazoans [170]. The presence of a few PTPs in many unicellular eukaryotic genomes, and absence of

PTKs [171] indicates that PTPs most likely evolved before PTKs. The reason why PTPs arose may be because some serine/threonine (Ser/Thr) kinases (which are more ancient than PTKs) may carry out sporadic but functionally important phosphorylation of tyrosines, and PTPs may have been needed to down-regulate these events. For example, activation of mitogen-activated protein kinases (MAPKs) requires phosphorylation of a threonine and tyrosine residue within its activation loop [170]. Although a proto-SH2 protein (SPT6) with an SH2-like sequence is found in fungi, it only binds to the Ser/Thr phosphorylated C-terminal tail of RNA polymerase II and does not bind to phosphotyrosine [172].

Aside from eukaryotes, the canonical tyrosine kinases have only been observed in the choanoflagellate *Monosiga brevicollis*, which is the closest known relative of eukaryotes [173]. Surprisingly, in contrast to other non-metazoan eukaryotes, which have no PTKs and very few PTPs and SH2 domains, the choanoflagellate *M. brevicollis*, has a comparable number of phosphotyrosine signalling proteins (43 PTKs, 34 PTPs, and 100 SH2 proteins) to complex multicellular metazoans such as humans (85 PTKs, 40 PTPs, and 110 SH2 proteins) [171]. Such an all-or-none pattern suggests that emergence of PTKs has driven rapid expansion of the phosphotyrosine signalling machinery. The increasing complexity and robustness in such an innovative signalling network may have facilitated the evolution of metazoans.

1.7.2 Phosphotyrosine binding domains

There are two major groups of modular domains involved in such phosphotyrosine-dependent interactions: the SRC homology 2 (SH2) domains [174] and phosphotyrosine-binding (PTB) domains [175]. Among these, SH2 domains are the predominant domains dedicated to recognition of phosphotyrosine-containing substrates, with 120 SH2 domains being present in

the human genome, forming part of 110 distinct proteins [169]. By contrast, there are only 50 PTB domains (in 46 proteins) encoded by the human genome, which are capable of recognising both phosphorylated and non-phosphorylated tyrosines in a consensus NPXY motif (where "x" is any amino acid; [176]) and phosphoinositides along with Pleckstrin homology (PH) domains [177].

SH2 domains are protein modules of about 100 amino acids, which fold into a structure consisting of a central antiparallel β-sheet flanked by two α-helices [178]. The binding sites of SH2 domains have two deep pockets, one that allows insertion of the phosphotyrosine (pY) residue of the binding partner, and another that accommodates a residue close to the phosphotyrosine, which is variable and important for binding specificity [65, 178, 179], although a few other exceptions have been found [180]. However, it was shown that changes of amino acid at the pY + 3 position only changes the binding affinity rather than resulting in elimination of binding [181]. Overall, the binding of an SH2 domain to the phosphotyrosine pTyr residue is responsible for about 50% of the total free energy of interaction, and the contribution of other amino acids close to the pY is often relatively small [182]. This implies that the extent to which the domain can be selective is limited, and that cross-reactivity is likely.

Structural studies have revealed that PTB domains have a common core fold consisting of a β sandwich with 2 perpendicular anti-parallel β sheets capped on one edge by a C-terminal alpha helix. In contrast to SH2 domains whose binding to tyrosyl phosphopeptide is primarily determined by the presence of phosphotyrosine and only secondarily by the surrounding amino acid sequence, binding of PTB domains to their targets is achieved through docking of the peptide ligand antiparallel to the β5 stand in the elongated cleft located between strand β5 and the C-terminal α helix. Besides, whereas SH2 recognition is partially dependent on the residues downstream of the pTyr residue, PTB recognition is highly dependent on the N-terminal region of the peptide backbone

[176]. The nature of PTB/ligand binding therefore also allows some promiscuous interactions to occur.

1.8 Untangling signal transduction – a role for systems biology

1.8.1 The definition of systems biology

The modern understanding of biology is, to a large extent, based on molecular biology. Whereas molecular approaches have allowed us to characterise a large fraction of individual molecular components of many organisms, they are not able to provide us a full understanding of an organism as an integrated whole. To address these problems and establish the links between molecules and physiology, a systems approach is needed [183].

Systems biology is not only concerned with system-level (genome-scale, transcriptome-scale or proteome-scale) measurements. In addition to an understanding of the system's structure, including molecular interactions, knowing how the behaviour of a system changes over time under various conditions (the system's dynamics) is another key aspect of understanding the system [184]. Moreover, whilst traditional molecular biology does not require a quantitative treatment of the data, a comprehensive set of quantitative data is required for a system-level analysis [183]. Overall, systems biology is a framework for modelling biological systems from systematic measurements, and for performing hypothesis-driven studies [185].

However, systems biology has lagged somewhat behind molecular biology, and systems studies have only become part of mainstream Biology in recent years. This is partly because quantitative treatment of the molecular data was not thought to be necessary in the past, leading to a lack of quantitative data, but

also because the technologies required for producing this type of data at a systems level did not become available until recently. For example, advances in bioinformatics, genome sequencing and a variety of high-throughput technologies have allowed production of comprehensive datasets at the molecular level, which has laid a solid foundation for the development of systems biology. Moreover, recent advances in software tools have enabled exploration of system-wide models and the formulation of new hypotheses. Together this has allowed a shift in emphasis of research from molecules to networks. Besides, although systems modelling not only relies on biological insights, but also on mathematical techniques, relatively few molecular bioscientists or mathematicians have a good understanding of both fields, although this is changing and is being circumvented by collaboration [183]. Although the development of systems biology is still at an early stage, notable successes have been seen in several applications, including metabolic control analysis, stem cell system biology, identification of diagnostic biomarkers and genetic-interaction-based therapy (reviewed in [185]).

1.8.2 Two approaches to systems biology

Two approaches have been applied in systems biology: the "top-down" and the "bottom-up" methodologies. Before the era of "omics", the bottom-up approach had played a dominant role in biological studies. This involves investigation of the known components of a manageable part of a system and their mechanistic details, followed by deduction of potential functional properties from such a subsystem. This is exemplified by early studies of signalling networks downstream of the EGFR [186]. These studies involve, for example, measurements of biophysical properties (e.g. binding data, kinetic parameters, rate constants) of the components, and responses of the subsystem to perturbations. The data are then incorporated into computational models, which allow predictions of system behaviour and testing of explicit hypotheses. The

ultimate goal is that, by combining the models of subsystems/pathways, it will be possible to develop a model for the whole system. One major limitation of this approach is the insufficient completeness of the picture, and the difficulty in obtaining precise data in certain cases, for example measuring the kinetic parameters for signalling and expression events *in vivo*, which means parameter estimation has got to be used [183].

In contrast, with the development of "omics", the "top-down" approach starts from generating an overview of the whole by measuring system-wide experimental data and investigating system behaviour. Subsequent modelling of the system gives rise to formulation of hypotheses and discovery of behavioural patterns. The predictions are then subjected to testing by biochemical analyses, which allows characterisation of the molecular constituents and pathway mechanisms. Although "top-down" systems biology can be potentially complete, the challenge is that it is often difficult to integrate the large amount of perturbations (e.g. mutations, overexpression or changes in environmental factors) into the modelling. In addition, whereas the models used in "bottom-up" systems biology are mechanism-based, those in "top-down" systems biology are phenomenological, which means that they are not always based on deep mechanistic insights, or they do not integrate knowledge about relationships between the molecular components. As a result, while the model can reveal potential mechanisms, these do not necessarily reveal the reality [183].

1.8.3 Two views of signal transduction

One of the main impacts of systems biology may end up being on how we consider intracellular signal transduction occurs. The prevailing view of signal transduction is that it proceeds through a linear relay of information via sequential bimolecular interactions. The emergence of this dogma was largely

dependent on the methods used to study these pathways, among which immune-precipitation or yeast two hybrid screens are the most widely applied. However, these techniques will under most conditions only be capable of identifying the binding partner with the highest affinity. Therefore, the assumption behind using these methods is that these interactions are highly selective. However, for a linear signalling pathway, it would be essential that the relevant affinity of a given SH2 domain and a specific cognate ligand is several orders of magnitude higher than that for non-specific interactions [187]. But this is not the case according to previous quantitative studies, where it was shown that the difference between so-called specific and nonspecific interactions is less than two orders of magnitude in affinity [188].

Using bottom-up systems approaches involving quantitative measurements, it has been recently shown that, instead of being a compartmentalised linear pathway consisting of only a few components, signal transduction likely involves a large number of previously unidentified components. One of these examples is epidermal growth factor receptor (EGFR, ErbB)-mediated signalling pathways, which happens to be one of the three main tumorigenesis pathways [189]. The ErbB family consists of four receptor proteins (ErbB 1-4), whose triggering is responsible for a wide range of cellular responses, including migration, proliferation, differentiation and apoptosis. Upon binding of a variety of growth factor ligands to the receptors, receptor dimerisation occurs. This is followed by phosphorylation of the tyrosine-based motifs, which then form docking sites for SH2-/PTB-containing proteins that mediate downstream signal transduction [190].

By using a microarray approach, Richard *et al.* established a quantitative signalling network involving over 50000 interactions between the 106 SH2 domains/41 PTB domains and 61 tyrosine–phosphorylated sites of ErbB1-4, including a sizeable number of previously unidentified interactions. In contrast to

the previous view that only very few proteins interact with the same target, this quantitative study has revealed that such an interpretation has to be affinity-threshold dependent: a large number of proteins in fact bind to the same target with different affinities. Among these, some are more promiscuous than others. For example, PIK3, SHC1, Crk, Syk and PLC-γ, have been found to bind at least three receptors (ErbB1-3). Surprisingly, Abl1 and Abl2, which were previously not considered to be involved in ErbB signalling, have been found to bind all four receptors with high affinity *in vitro*. In contrast to the other three receptors which possess many highly promiscuous binding sites and mediate high-affinity interactions with many target proteins, ErbB4 is markedly more selective with only a few interactions, indicating a specialised function of this receptor. An important implication arising from this study is that ErbB1 and ErbB2, which are overexpressed in many human cancers, become significantly more promiscuous as the affinity threshold is lowered. This implies these receptors may contribute to oncogenesis by turning on different signalling pathways when they are over-expressed, leading to changes of the cellular phenotypes [191]. In reality, the extensive cross-interaction identified by the systems analysis is not surprising considering the structural basis of the interaction between the SH2 domains/PTB domains and the tyrosyl phosphopeptide (**Section 1.7.2**).

Whereas the recruitment of SH2- and PTB-domain containing proteins to surface receptors represents the initial stage of intracellular signal transduction, protein kinases and phosphatases, which mediate phosphorylation and dephosphorylation of their substrates, respectively, are perhaps the crucial information carriers. Using a mass spectrometric analysis of protein complexes with the application of a computational data-filtering scheme, Breitkreutz *et al.* [192] identified a kinase and phosphatase interaction network of 1844 interactions in budding yeast. One important finding of this study is that interactions between kinases are significantly enriched compared to those with

other substrates, which means kinases frequently phosphorylate other kinases, thus allowing extensive cross-talk between signalling pathways. Overall, the extensive cross interactions between kinases and phosphatases suggest that they may form a cooperative network mediating cellular responses. However, further understanding of this signalling network requires a quantitative study of the interaction strength as well as exploration of the system dynamics with time-dependent experiments and computational modelling.

1.8.4 Is specificity system- rather than molecule-intrinsic?

Apart from biophysical data, the role of the numbers of molecules in a system is becoming increasingly obvious. A previous study on the T cell costimulatory receptors has revealed its importance. As previously discussed, CD28 and CTLA-4 have two shared ligand called B7-1 and B7-2. Whilst B7-1 is bivalent, B7-2 is monovalent but its expression is significantly higher than that of B7-1. On the other hand, while CD28 is monovalent and is expressed constitutively, CTLA-4 is bivalent but expressed at significantly lower levels than CD28, and only induced over the course of an immune response [95]. Biophysical studies using the surface plasmon resonance (SPR) technology showed that B7-1 and B7-2 bind CTLA-4 20 times and 8 times more strongly than CD28, respectively, which also means that the CTLA-4/CD28 engagement ratio (or "bias") of B7-1 is 2.5-fold higher than that of B7-2 [90]. Although affinity data alone suggest that CTLA-4 can out-compete CD28 binding to both B7-1 and B7-2, this changes when other factors are taken into account. By using mean-field, ordinary differential equations as well as a number of physiological parameters, a computational framework was established to simulate the interactions of these molecules at the immunological synapse. This showed that the overall outcome of complex formation can be profoundly influenced by the combination of stoichiometric properties, expression levels and competition effects [95]. According to the

modelling, when stoichiometric differences are allowed for, the CTLA-4/CD28 bias of B7-1 is 18-fold higher than that for B7-2, but both ligands "prefer" CTLA-4. On the other hand, when expression levels and competition effects are taken into account, the binding specificities of B7-1 and B7-2 polarize, with B7-2 binding 33-fold more CD28 than CTLA-4, and B7-1 binding 7-fold more CTLA-4 than CD28. This study suggests that receptor specificity might be system- rather than protein-intrinsic.

1.8.5 Obtaining system insights from transcriptomic analysis

Since genomic information is generally conserved in any cells within an organism, studying the genome alone generally yields few insights into gene expression in a given cell type. Transcriptomics is the study of gene expression profiles and the transcriptome corresponds to the complete set of transcripts (including mRNA, rRNA, tRNA, and other non-coding RNA) in a cell, with their quantity, for a specific developmental stage or physiological condition [193]. In particular, the abundance of mRNA provides crucial insights into systems biology by acting as an evidence for the presence of a protein in a specific system, whether or not that protein is actually expressed at levels sufficiently high to impact on the biology of the system [194]. Despite the attractions of obtaining this type of information, advances in transcriptomic studies had until recently been hampered by the availability of specific technologies and various technical issues.

Gene expression technologies using the hybridization-based approaches are known as "closed" technologies. This usually involves custom-made microarrays or synthesis of a series of high-density oligonucleotide tiling arrays representing a subset or the entire genome and hybridisation of fluorescently-labelled (reverse-transcribed) cDNA. Whilst genomic tiling microarrays have revealed enormous previously un-described transcripts [195], this technology is

not only limited by the high background levels due to cross-hybridisation between closely related sequences and problems of detecting and/or quantifying low abundance transcripts, but also requirement of prior knowledge of the genes likely to be of interest. This means not all transcripts can be confidently sampled due to the current uncertainty about the actual number of genes in the human genome and the fact that genomic information is not available for every organism. Although the whole genome could, in principle, be covered using a microarray approach, non-overlapping coverage of non-repetitive genomic DNA requires construction of a total of 51,874,388 36-nucleotide probes on 134 arrays [195].

In contrast, the so-called "open" gene expression profiling allows characterisation and discovery-driven analysis of transcriptomes without reliance upon the existing knowledge about genes. Technologies of this type include serial analysis of gene expression (SAGE) [196] and more recently next-generation sequencing (NGS) technologies (Reviewed in [197]). SAGE technology involves the generation, amplification and sequencing of short (14bp for the "ShortSAGE" method [196]; 21bp for "LongSAGE" [198]; 26bp for "SuperSAGE"[199]) tags derived from the 3' ends of transcripts. Whereas the sensitivity of cDNA or long oligonucleotide arrays is estimated to vary from 50 to 400 transcripts per million, species with at an average count as low as 5 transcripts per million could be reliably identified as being differentially expressed in large-scale tag-sequencing experiments [200]. Nonetheless, SAGE relies on the expensive Sanger sequencing technology and a large number of tags still cannot be uniquely mapped to the reference genome.

Recently, the approach has been superseded by the advent of NGS technologies, which allow low-cost and high-quality mapping and quantification of transcriptomes at great depth. The key advantage of NGS is the considerably sequencing depth (the total number of all the sequences reads generated) compared to other technologies, which means a much greater coverage

(percentage of transcripts) can be achieved. For example, compared to the 28,000 transcript tags generated by a typical SAGE experiment [198], a depth of 122 million reads from six different growth/developmental conditions has been reported in a study of the fission yeast Schizosaccharomyces pombe, which had covered >90% of the genome, including 453 novel transcripts that were not robustly transcribed or were rapidly degraded [201].

1.8.6 Mass spectrometry-based proteomics

As previously discussed, concentration of proteins can be a determining factor in systems biology modelling. However, obtaining reliable quantitative data for proteins has been a challenge. Whereas mRNA concentrations have been commonly used as proxies for protein concentrations, in which case transcript abundances are assumed to be the main determinant of protein abundances, emerging evidence shows that this is not the case. The correlation between mRNA and protein concentrations is usually defined by the squared Pearson's correlation coefficient (R^2), which ranges from 0.2 to 0.47 in bacteria, 0.34 to 0.87 in yeast, and 0.29-0.46 in human (reviewed in [202]). This implies that more than half of the variation in protein concentrations cannot be explained by variation in mRNA concentrations. A recent global quantitative study in mammalian cells showed that variation in protein expression levels are primarily determined by regulation of translation [203], although protein degradation also has an impact [204]. Overall, the discrepancy means that while mRNA expression indicates presence of a protein, it is rarely a reliable guide to protein abundance.

Proteomics is the study of the proteins in an organism or a cell population at a system level. Identification and quantification of proteins used to rely on affinity-based methods (e.g. western blotting or ELISA approaches aided with highly specific antibodies), which tend to be expensive and laborious. These

methods are not only time-consuming, but also impractical for large-scale analysis. However, a systematic effort to generate validated affinity reagents has been initiated [205]. To address these limitations, the two-dimensional (2D) gel electrophoresis approach, which separates proteins according to their charges and sizes through a combination of isoelectric focusing (IEF) and SDS-PAGE, was developed to aim for large-scale identification and quantification of proteins. However, this technique is limited by the fact that proteins with low abundance and those with extreme charges and sizes are often very difficult to analyse, such that only a few hundred proteins in the proteome are covered [206, 207]. This is obviously inadequate compared to the number (>22,000) of proteins in a human proteome [208]. Over the past few years, the development of systems biology has increased the demand for the high throughput, large-scale detection and quantification of protein sets. This has driven the improvement of proteomic technologies, as exemplified by the development of mass spectrometry (MS) proteomics. The MS approach not only allows identification of proteins but also, in principle, quantitates the amount of proteins at a systems level, as shown in studies of yeast [209, 210] and human pathogens [211].

Two main approaches have been employed for quantitative analysis: the stable isotope-based and labelling-free methods. The stable isotope-based method involves comparison of two samples, one of which is labelled with heavy isotopes either through the incorporation of isotopes into the proteins of cells growing in media enriched in stable isotopes (e.g. ^{2}H, ^{13}C and ^{15}N) [212]. An advanced technology for quantitative proteomic comparisons, called stable isotope labelling by amino acids in cell culture (SILAC) [213], involves growing one of the two cell samples in a medium that contains an isotope-labelled amino acid (e.g. leucine). The two (labelled and un-labelled) samples are then combined for processing and MS analysis, and relative amounts of the proteins are measured via the ratios of the two species identified by isotopic mass differences

[214]. Compared to the isotope-based methods, label-free quantification is a lot simpler but less accurate. It involves counting and comparing spectral signals of the same peptide from different experiments [206]. Whereas only "relative" quantification could be achieved in the past [215, 216], "absolute" quantification is now also achievable to allow determination of the average number of protein copies per cell in a cell population [217, 218].

Whereas these 'shotgun'-based strategies tend to favour proteins of higher abundance, a targeted proteomics approach called selected reaction monitoring (SMR) has recently emerged to allow quantification of proteins with a concentration below 50 copies per cell [210]. This approach involves selection of precursor peptides for a predetermined set of proteins and the use of triple quadrupole (QQQ) mass spectrometers to monitor the peptide-specific fragment ions, which gives highly consistent and sensitive data. Whereas this approach was initially only applied to a limited number of proteins, a high-throughput assay has recently been reported [219]. By determining the absolute copy numbers of selected heavy stable isotope-labelled reference peptides corresponding to proteins expressed at different abundance levels using the SRM approach, these data can then act as calibration points for translating the relative protein abundance into absolute protein abundance. By using this approach, Malstrom *et al.* have determined the copy numbers for 83% of the MS-detectable proteome of the human pathogen *Leptospira interrogans* [211].

1.9 Aims of this thesis

The aims of this thesis are as follows:

1. To explore the number, and expression patterns of SH2 domain-containing proteins expressed in different cell types and tissues, including a variety of immune and non-immune cells, at the transcript and protein expression level, as

the basis of choosing a model system in which to study tyrosine phosphorylation-based signal transduction. This is achieved via analysis of, and comparisons between, a number of published and unpublished transcriptomic libraries generated by serial analysis of gene expression and by next-generation sequencing, and newly released proteomic data.

2. To perform a system-level investigation of the affinities of SH2 domains for phosphotyrosine-containing receptor motifs derived from all of the most important T cell signalling surface receptors, including the TCR, LAT, CD28, ICOS, CTLA-4, BTLA and PD-1. This involves expression and purification of large quantities of soluble, properly folded recombinant SH2 domains and measurement of their affinities for phosphor-tyrosyl peptides representing the corresponding receptor motifs, using surface plasmon resonance (SPR)-based technology.

3. To explore how specificity is generated in signal transduction, via the development of a fully parameterized systems model of the recruitment of SH2 domains to phosphotyrosine-containing receptor motifs. The mathematical simulations are based on the rigorous affinity data (from the SPR-based binding assays) and expression data (from quantitative proteome analyses).

Chapter 2. Materials and methods

2.1 Polymerase chain reaction

The KOD Xtreme™ DNA Polymerase kit (Novagen) and the Phusion™ Flash High-Fidelity PCR Master Mix (Thermoscientific) were used for standard gene amplification as detailed in **Tables 2.1-2.4**. These systems include a high fidelity KOD DNA polymerase and Phusion Flash DNA Polymerase, respectively, and both of which permit hot start thermocycling, which ensures efficient amplification of the correct sequence and ease of use.

Table 2.1. KOD Xtreme™ DNA polymerase reaction mix

	Company	**Concentration**	**Volume**
Polymerase Buffer	Novagen	2X	25 µl
dNTP	Novagen	2mM	10 µl
Primer 1	Integrated DNA Technologies (IDT)	10 µM	3 µl
Primer 2	IDT	10 µM	3 µl
Template	Genescript	100 ng/µl	1 µl
DNA polymerase	Novagen	1 U/µl	1 µl
dH_2O	-	-	7 µl

Table 2.2. KOD Xtreme™ DNA polymerase PCR conditions

	Step	Temp (°C)	Time
1	Polymerase activation	94	2 min
2	Denaturation	98	10 s
3	Annealing	60	30 s
4	Extension	68	2.5 min
	Repeat 2 → 4	X30 cycles	
5	Final Extension	68	3 min
6	Cooling	4	∞

Table 2.3. Phusion™ Flash High-Fidelity PCR Master Mix reaction mix

Reagents	Company	Concentration	Volume
Polymerase Buffer	Thermoscientific	2X	25 µl
Primer 1	IDT	10 µM	3 µl
Primer 2	IDT	10 µM	3 µl
Template	Genescript	100 ng/µl	1 µl
dH_2O	-	-	18 µl

Table 2.4. Phusion™ Flash High-Fidelity PCR Master Mix PCR conditions

	Step	Temp (°C)	Time
1	Initial Denaturation	98	10 s
2	Denaturation	98	1 s
3	Annealing	60	5 s
4	Extension	72	1 min
	Repeat 2 → 4	X30 cycles	
5	Final Extension	72	1 min
6	Cooling	4	∞

2.2 DpnI digestion of PCR reactions

In order to remove the (wild type) DNA template, the PCR products were treated by adding 20 units of DpnI (NEB; a mixture of 1 ul DpnI, 1 ul 10x NEB-4 buffer and 3 ul H_2O) to each PCR reaction and incubating at 37 °C for 1 h, followed by purification using AMPure magnetic beads (Beckman-Coulter) according to the manufacturer's protocols.

2.3 Agarose gel electrophoresis

Agarose powder (Sigma) was added to Tris-borate (EDTA) (TBE) buffer (Sigma) gels at 1.25% w/v. The mixture is then heated to dissolve using a microwave, followed by cooling prior to addition of SybrSafe™ (InVitrogen) DNA Gel Stain, which is a safe reagent in substitution of the carcinogenic ethidium bromide, for

visualization of DNA (which is carcinogenic). The gel was poured and the appropriate combs were added prior to setting. After set, the combs were removed and the electrodes and gel were bathed in TBE buffered. DNA loading buffer (0.05% w/v bromophenol blue, 30% glycerol) was added to the DNA product prior to loading to the gel, and molecular marker (Sigma) was included to allow estimation of DNA band size. Electrophoresis was carried out at 100 mA using an ABgene system until the fragments were well separated. Gels were examined and photographed under UV light using (SYNgene Bio Imaging System).

2.4 Ligation of a gene of interest into a plasmid using Infusion™

1ul 33 ng/ul desired pOPIN vector was added to each well of a PCR plate. 5.5 ul H_2O and 1.5 ul PCR product were added to the Infusion™ dried-down cloning mix to make up a reaction mix, 2.3 ul of which was added to each well of the PCR plate. The reaction solution was incubated for 30 min at 42 °C. Immediately after the reaction finishes, 13.3 ul TE buffer (Invitrogen) was added into each Infusion™ reaction to terminate the reaction.

2.5 Amplification of plasmid DNA and verification of insert

Infusion™ reaction was transformed into the One Shot® OmniMAX™ II T1-phage-resistant competent cells (Invitrogen) according to the manufacturer's protocols. Briefly, the aliquoted competent cells were thawed on ice. 3 µl of Infusion™ reaction was used to transform a competent cell aliquot, followed by incubation on ice for 30 min. The cells were then treated with heat-shock for 30 s at 42 °C by placing the whole rack in the water bath, with the competent cells in the tubes

under the water level. Thereafter, the cells were placed back to ice for 2 min, followed by addition of SOC™ medium (Invitrogen) into each tube. The cells were incubated for 1 hour at 37 °C in an incubator. 60 µl of transformed cells per tube was transferred onto the LB agar plate, with 1ml of LB agar supplemented with 50 ug/ml Carbenicillin (Cb; Melford Laboratories Ltd.), 0.02% w/v X-Gal and 1 mM IPTG in each well of the 24-well tissue culture plate. The cells were spread over the well by shaking vigorously by hand in a horizontal plane, and the plates were allowed to dry off at 37 °C before turning upside-down. The plates were incubated overnight at 37 °C. Two white colonies were picked for each construct and each of them were inoculated in 1ml Luria Broth (LB) (Invitrogen) supplemented with 50 ug/ml Cb in deep-well, 96-well plates. The cultures were grown overnight at 37 °C, with shaking at 200 r.p.m., before harvesting by centrifugation at 5000g for 10min at 4 °C. Pelleted cells were then used for plasmid preparation, using a Qiagen BioRobot 8000 and QIAgen Turboprep kits, or a QiaVAC96 manifold (Qiagen), according to the manufacturer's instructions (Qiagen). The concentration of the protein was estimated by NanoDrop (Thermo Scientific) at A260. The KOD Hi-Fi™ DNA Polymerase kit [Novagen; details for reaction mix and PCR conditions are shown in **Table 2.5** and **Table 2.6**] was used to verify presence of insert according to the manufacturer's instructions.

2.6 Glycerol stocks of transformed bacteria

100 ul of the overnight culture for transformed bacteria from **Section 2.5** was mixed 100 ul of 30% Glycerol (diluted in LB) in a 96-well plate, and the cells were stored at −80 °C. To culture from frozen stocks, a pipette tip was used to dip into the corresponding well to collect a fraction of cells, followed by inoculation of the cells into 5ml of LB medium supplemented with 50 µg/ml carbenicillin and cultured overnight at 37 °C.

Table 2.5. KOD Hi-Fi™ DNA Polymerase reaction mix

Reagents	**Company**	**Concentration**	**Volume**
Polymerase Buffer	Novagen	2X	25 µl
T7-F primer	IDT	100 µM	0.15 µl
Reverse primer of gene insert	IDT	10 µM	1.5 µl
$MgCl_2$	Novagen	25 mM	1 µl
dNTP	Novagen	2 mM	2.5 µl
Template	-	-	1 µl
DNA polymerase	Novagen	10x	0.2 µl
dH_2O	-	-	18 µl

Table 2.6. KOD Hi-Fi™ DNA Polymerase PCR conditions

	Step	**Temp (°C)**	**Time**
1	Polymerase activation	95	2 min
2	Denaturation	95	30 s
3	Annealing	55	30 s
4	Extension	68	2:30 min
	Repeat 2 → 4	X30 cycles	
5	Final Extension	68	5 min
6	Cooling	4	∞

2.7 Preparation of competent *E. coli* expression strains using the RbCl method

Two *E. coli* expression strains were used in this study, which are B834 (DE3) and Rosetta (DE3) LysS (Novagen). The cells were inoculated in 5 ml LB [35 µg/ml chloramphenicol (Duchefa) was required to maintain the pRareLysS plasmid for Rosetta cells) and the culture was grown overnight at 37 °C with shaking.

The overnight culture was diluted 1 in 100 by adding 2.5 ml LB into 250ml pre-warmed LB supplemented with the appropriate antibiotic in flask. The culture was grown at 37°C with shaking for 2-3 hours, until OD_{600} is 0.25-0.3. The cells were then incubated on ice for 5 min, followed by spinning at 4000g for 5 min at 4°C.

The supernatant was discarded and the cell pellet was resuspended in 50 ml Tfb I [transformation buffer I; details are summarised in **Table 2.7**], with pipetting kept to the minimum. The cell suspension was then incubated on ice for 5 min, prior to spinning at 4000 g for 5 min at 4 °C. The supernatant was discarded and the cells were resuspended very gently in 5 ml Tfb II (transformation buffer II; details are summarised in **Table 2.8**], with pipetting kept to the minimum.

The cell suspension was incubated on ice for 15 min. 50 µl of cells was aliquoted into each 1.1 ml microtubes (Bioquote) tube in a 96-hole rack on dry ice, and the cells were stored at -80 °C. The protocol described above was sufficient to make one 96-well rack of 50 µl aliquots.

Table 2.7. Components of transformation buffer I (Tfb I)

Reagents	**Company**	**Final Concentration**	**Per 1 litre**
KOAc	Sigma	30 mM	2.94 g
RbCl	Sigma	100 mM	12.10 g
$CaCl_2$	Sigma	10 mM	1.47 g
$MnCl_2$	Alfa Aesar	50 mM	9.90 g
Hexamine cobalt Cl	Sigma	3 mM	0.80 g
Glycerol	Sigma	15% (v/v)	300 ml of 50% stock

The pH was adjusted to 5.8 with 0.2 M acetic acid and sterilised by filtration.

Table 2.8. Components of transformation buffer II (Tfb II)

Reagents	**Company**	**Final Concentration**	**Per 200 ml**
MOPS	Sigma	10 mM	419 mg
RbCl	Sigma	10 mM	2.418 g
$CaCl_2$	Sigma	75 mM	2.205 g
Glycerol	Sigma	15% (v/v)	60 ml of 50% stock

The pH was adjusted to 6.5 with KOH and sterilised by filtration.

2.8 Protein expression screening in *E. coli*

The sequencing-verified expression plasmids were used to transform the competent cells, either B834 (DE3) or Rosetta (DE3) LysS *E. coli* (Novagen), according to the manufacturer's instructions. Briefly, the aliquoted competent cells were thawed on ice. 3 µl of plasmid (the concentration of the DNA is typically around 60 ng/µl) was used to transform a competent cell aliquot, followed by incubation on ice for 30min. The cells were then treated with heat-shock for 30 seconds at 42 °C by placing the whole rack in the water bath, with the competent cells in the tubes under the water level. Thereafter, the cells were placed back to ice for 2 min, followed by addition of 450 µl Power Broth™ (PB) (Molecular Dimensions) into each tube. The cells were incubated for 1 hour at 37 °C in an incubator. 60 µl of transformed cells per tube was transferred onto the LB agar plate, with 1ml of LB agar supplemented with the appropriate antibiotic(s) in each well of the 24-well tissue culture plate. Matrix IMPACT pipettors were used to transfer between 96 and 24-well plate formats here and other procedures. The cells were spread over the well by shaking vigorously by hand in a horizontal plane, and the plates were allowed to dry off at 37 °C before turning upside-down.

All plates and subsequent media used for the culture of the Rosetta (DE3) LysS cells were essentially identical to those described for B834 (DE3) cells. While B834 (DE3) cells were grown in 50 µg/ml carbenicillin, additional 35 µg/ml chloramphenicol was required to maintain the pRareLysS plasmid for Rosetta cells. Plates were incubated for overnight at 37 °C before individual colonies were used to inoculate 1 ml PB medium supplemented with the appropriate antibiotics in 96-well deep-well plates. The plates were sealed with gas-permeable adhesive seals and shaken at 225r.p.m. for overnight at 37 °C. For IPTG induction of expression, 120 µl of each overnight culture was used to inoculate (in four 24-well deep-well plates) 2.5 ml of LB medium supplemented with appropriate antibiotics. The

diluted cultures were grown at 37 °C with shaking at 225r.p.m. for 5 h (the average OD_{600} is about 0.5) before reducing the temperature to 20°C and addition of ITPG (Sigma) to a final concentration of 1.0 mM. The plates were then grown for a further 22 hours with shaking at 225 r.p.m. For auto-induction of expression, 120 µl of each overnight culture was used to inoculate (in four 24-well deep-well plates) 2.5ml of TB Overnight Express™ media (TBONEX; Novagen) supplemented with appropriate antibiotics and glycerol (to a final concentration of 10%). The diluted cultures were grown at 37°C with shaking at 225r.p.m. for 5h before reducing the temperature to 25°C and shaking for a further 22 hours.

A 1.0 ml aliquot of culture from each well was transferred to a 2 ml 96-well deep-well plate, and harvested by centrifugation at 6000g for 10min at 4 °C before decanting and draining of the waste media. The Pelleted cells were frozen and stored at −80 °C for at least 30min prior to the protein expression screening.

The expression screening was carried out using the Qiagen BioRobot 8000 with standard Qiagen Ni-NTA magnetic bead protocols, according to manufacturer's instructions. Briefly, the defrosted cells were resuspended in 210 µl of Lysis Buffer (NPI-10-Tween; **Table 2.9**) supplemented with a final concentration of 1 mg/ml Lysozyme (Sigma) and 400 units/ml of DNAse Type I (Sigma). The plate was shaken with an orbital microtiter plate shaker at 1000 r.p.m. for 30 min, followed by centrifugation at 6000g for 30 min at 4 °C. In the meantime, 20 µl of the Ni–NTA magnetic bead suspension was dispensed into each well of a 96-well microtiter plate. The clear lysate was transferred to each well of the microtiter plate containing the Ni–NTA magnetic bead. The microtiter plate was then shaken at 600 r.p.m. for 30 min before being transferred to a 96-Well Magnet (QIAgen) for 1 min incubation. The supernatant was then removed carefully from the beads and 200 µl of Wash Buffer (NPI-20; **Table 2.9**) was added to each well. The plate was then removed from magnet and allowed to shake for 5 min. The place was transferred back to the 96-Well Magnet for 1 min

incubation and the buffer was removed. The washing step was repeated twice, followed by addition of 50 µl Elution Buffer (NPI-250; **Table 2.9**) to each well. The plate was shaken for 1 min before being transferred to the 96-Well Magnet for 1 min incubation. Finally, the supernatant (eluate) was transferred to a fresh microtiter plate for analysis on SDS–PAGE gels (InVitrogen NuPAGE™ Novex 10% Bis-Tris Midi gels with MES buffer) using the XCell4 Surelock™ Midi-Cell system (Invitrogen) and visualized with InstantBlue™ (Expedeon). 10 µl of the eluate was mixed with 10 µl of protein loading dye [100 mM Tris-HCl pH6.8, 4% SDS (Sigma), 0.2% w/v bromophenol blue, 20% glycerol], followed by 10-min incubation at 95 °C, before 10 µl of sample was loaded into each well. SigmaMarker™ Low Range and High Range are included on each end of the gel for estimation of the molecular weight of the samples.

Table 2.9. Components of buffers used in the expression screen

Buffer	**Contents**
Lysis buffer (Buffer NPI-10-Tween)	50 mM NaH2PO4 (Sigma), 300 mM NaCl, 10 mM imidazole (Sigma), 1% v/v Tween 20 (Sigma).
Wash buffer (Buffer NPI-20-Tween)	50 mM NaH2PO4, 300 mM NaCl, 20 mM imidazole, 0.05% v/v Tween 20.
Elution buffer (Buffer NPI-250-Tween)	50 mM NaH2PO4, 300 mM NaCl, 250 mM Imidazole, 0.05% Tween.

The pH of all three buffers was adjusted to 8.0 using NaOH and sterilised by filtration.

The buffers are stored at 4 °C.

2.9 Large-scale production of soluble proteins

Individual colonies were used to inoculate 20 ml PB medium supplemented with the appropriate antibiotics in a 50 ml Falcon™ conical centrifuge tube. The tubes were shaken at 225 r.p.m. for overnight at 37 °C.

For IPTG induction of expression, 5ml of each overnight culture was used to inoculate 500ml of LB medium supplemented with appropriate antibiotics. The diluted cultures were grown at 37 °C with shaking at 225 r.p.m. to an approximate OD_{600} of 0.6 before addition of ITPG to a final concentration of 1.0 mM and reducing the temperature to 20 °C. The plates were then grown for a further 22 h with shaking at 225 r.p.m. For auto-induction of expression, 5ml of each overnight culture was used to inoculate 500ml of TBONEX supplemented with appropriate antibiotics. The diluted cultures were grown at 37 °C with shaking at 225 r.p.m. for 5h before reducing the temperature to 25 °C and shaking for a further 22 h.

The cultures were collected in an ultracentrifuge bottle (Beckman Coulter) and harvested by centrifugation at 6000g for 15 min, and pelleted cells were frozen and stored at −80 °C. The cells were defrosted and resuspended in binding buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 0.2% (v/v) Tween 20 and an EDTA-free protease-inhibitor cocktail tablet (Roche), followed by lysis of the cells with a Basic Z Cell Disruptor (Constant Systems). After centrifugation at 24000g for 30 min, the cleared lysate was applied at 2 ml/min onto a 1 ml HisTrap FF column attached to an ÄKTA Express protein-purification system (GE Healthcare), which operates at 4 °C. The column was washed to a UV (280 nm) baseline with binding buffer and the protein was eluted at 1 ml/min with elution buffer (50 mM Tris pH 7.5, 500 mM NaCl, 500 mM imidazole). The eluate was automatically loaded onto a 16/60 HiLoad Superdex 75 column (GE Healthcare) pre-equilibrated in gel filtration

buffer (20 mM Tris pH 7.5, 500 mM NaCl). 0.5 ml fractions were collected in a 96-well deep-well plate following separation by size exclusion in the HPLC column, and those potentially containing proteins were analysed on SDS-PAGE gels and visualized with Instant Blue™.

Protein fractions of interest were collected based on the and concentrated with Amicon® Centriprep, Centricon and Microcon centrifugal concentrators (Fisher Scientific Ltd.) of appropriate molecular weight cut-off values, according to the manufacturer's instructions, until the desired level of concentration was achieved. Prior to use, the preservative in membrane of the concentrator was removed by passing gel filtration buffer through the membrane by centrifugation (4 °C, 2 min). The concentration of the protein was estimated by NanoDrop (Thermo Scientific) at A280, and The concentrated protein aliquots were frozen in liquid nitrogen and stored at −80 °C. The N-terminal fusion tag, where applicable, was removed by overnight incubation of the protein with His-tagged 3C protease, supplemented with 1 mM TCEP (Soltec Ventures), at 4°C. The 3C protease and any uncleaved protein were removed by passing the sample through a 1 ml HisTrap FF column, after which a secondary HPLC purification was performed to separate the SH2 protein from others by size exclusion.

2.10 Protein refolding

2.10.1 Preparation of denatured inclusion bodies

The cell pellet from 1 litre of culture was re-suspended in binding buffer (50 mM Tris-Hcl pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 0.2% (v/v) Tween 20, followed by lysis of the cells with a Basic Z Cell Disruptor (Constant Systems). Crude inclusion bodies were then pelleted at 12000g for 30 min at 4°C and the supernatant was discarded. The pellet was re-suspended in 20 ml Wash

Buffer A (50mM Tris-HCl pH8, 100mM NaCl, 0.1% Ssodium azide, 0.5% Triton X100) by grinding with a 5 ml pipette, pipetting up-and-down and votexing, and the inclusion bodies were then pelleted at 25000g for 10 min at 4 °C. The sample was processed in ice and this washing step was repeated for another 3 times or until the pellet becomes white. The pellet was finally resuspended in 20ml Wash Buffer B (50 mM Tris-HCl pH8, 100 mM NaCl), followed by spinning at 25000g for 10 min at 4°C, which removes any Triton residues.

The purified inclusion bodies are then dissolved in 20 ml of Denaturing Buffer [50mM Tris-HCl pH8, 100mM NaCl, 6M guanidine hydrochloride (GnHCl; Melford), 10 mM EDTA, 10 mM DTT (Sigma)]. The sample was left on ice on a shaker for at least an hour to allow the pellet to be fully dissolved. The sample was then 25000g for 20 min at 4 °C and the supernatant was carefully harvested by decanting in order to avoid tipping the debris out as the colour of the pellet is very similar to the supernatant. The concentration of the inclusion body solution was estimated at A280 and the samples were stored as 10mg / aliquot at −20°C.

2.10.2 The refolding screen

The components selected for the preparation of refolding screening buffers are shown in **Table 2.10** and **Table 2.11** and the detailed compositions of the screening buffers are shown in **Appendix 1**. The different screening buffers were prepared and 200 µl of each was dispensed into the corresponding well of a microtiter plate. The screening plates were stored at −20 °C and defrosted prior to use. 20 µl of the 10 mg/ml inclusion body solution was added to the screening plate shaken with an orbital microtiter plate shaker at 350 r.p.m. using a multi-channel pipette. After completion of adding the inclusion body samples, the plate was shaken for another 5 min before evaluation of buffer condition by measuring the light absorbance at 350 nm.

Table 2.10. The components of the refolding screen buffer

Components	Notes	Ref.
Buffer pH	pH should be more than 1-2 units away from the isoelectric point (IP) of the target protein to avoid precipitation	[220]
Ionic strength	Solubility of the protein of interest can increase (salting in) or decrease (salting out) as a function of ionic strength	[220]
Polyethylene glycol & Brij58	Prevents re-aggregation	[221]
L-arginine & Glucose	Prevents re-aggregation	[221]
Non-detergent sulfobetaines (NDSB)	stabilizing agents that can facilitate protein renaturation	[222]
Tris 2-carboxyethyl phosphine (TCEP)	Maintains a reducing environment and may have a positive synergistic interaction with NDSB	[223]
EDTA	Metal-chelating reagent to prevent metal-catalysed air oxidation of cysteine and inhibit the activity of some proteases	[220]
Divalent cations	May act as potential cofactors for refolding	[221]

Table 2.11. The conditions of refolding screen buffer

Buffer pH (50 mM)	**Ionic Strength**	**Amphiphilic Components**	**Surfactants (100mM)**	**Reducing Agent (10 mM)**	**Additives**
MES-pH 6	Null	Null	Null	Null	Null
Tris-pH 7	NaCl 200mM	Glycerol 20% (v/v)	NDSB-195	TCEP	Arginine 800mM
Tris-pH 8	NaCl 350mM	PEG-200 0.05% (w/v)	NDSB-201		Glucose 500mM
†BTP-pH 9	NaCl 500mM	PEG-2000 0.05% (w/v)	NDSB-221		EDTA 10mM
		PEG-4000 0.05% (w/v)	NDSB-256		Cocktail‡
		PEG-10000 0.05% (w/v)	Brij58 (0.5%)		

†BTP, Bis–Tris Propane. ‡Cocktail containing 2 mM of the following: $CaCl_2$, $MgCl_2$, $ZnCl_2$, $NiCl_2$, $CuSO_4$, $CoSO_4$ (Sigma).

2.10.3 Refolding by rapid dilution

Prior to use, an aliquot of denatured protein was added with DTT to a final concentration of 10 mM and left on ice on a shaker for an hour to ensure all the disulphides are broken, and the selected refolding buffer was pre-chilled. 20mg of the inclusion body was added to 200ml of the refolding buffer in a 500ml flask using a P200 pipette over a period of about 30 mins. Sufficient time was left between additions of the inclusion body so that the inclusion body was well dissolved in the buffer before addition of another drop. The refolding buffer was kept on ice on a magnetic stirrer, with the flea spinning at a high speed.

Upon completion of the sample addition, the solution was transferred to a magnetic stirrer, with the flea spinning at a lower speed, and kept in a cold cabinet at 4°C overnight.

The refolding mix was then centrifuged at 15000g for 20 mins at 4 °C, after which the supernatant was harvested and concentrated to 7 ml (the maximum volume can be loaded into the ÄKTA per time) using Amicon® Centriprep. The sample was then purified by size exclusion in the HPLC column and processed as described in **Section 2.9**.

2.10.4 Surface plasmon resonance-based experiments

2.10.4.1 Ligand immobilisation on CM5 sensor chip

Immobilisation of proteins onto the surface of individual flowcells of CM5 sensor chip (GE Healthcare) was performed under conditions of 10 µl/min flow rate and 37 °C and using HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% (v/v) Surfactant P20; GE Healthcare) as the running buffer. After priming of the system, 100 µl 0.05 M 1:1 solution of EDC [1-ethyl-3-(3-dimethylaminopropyl)

carbodiimide] / NHS (N-Hydroxysuccinimide) was injected to activate the dextran matrix of the sensor chip surface. For the streptavidin immobilization step, the streptavidin solution was prepared by dissolving the streptavidin powder (Sigma) in 10 mM sodium acetate buffer of pH 5.0. For the preliminary experiments, 100 µl of 0.5 mg/ml streptavidin was injected to achieve an immobilization level of 22000 – 26000 RU; for the large-scale binding analyses, injection of 0.25 mg/ml streptavidin was manually performed to achieve an immobilization level of approximately 5000 RU. Thereafter, 150 µl of 1 M ethanolamine hydrochloride was injected to deactivate the unreacted carboxylmethyl groups on the dextran matrix. Finally, 30 µl of Glycine pH 2.5 was injected to remove any non-covalently bound proteins.

2.10.4.2 Measurement of the binding affinity of the interaction between LCP2 and CD6 (662Y)

The affinity analyses of the interaction between LCP2 and CD6 (662Y) were performed using the BIAcore3000, under conditions of 20 µl/min flow rate and 37 °C and using HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% (v/v) Surfactant P20; GE Healthcare) as the running buffer. Prior to experiment, the LCP2 SH2 proteins were defrosted from the -80°C stock and purified by HPLC. A 2-fold serial dilution of totally nine concentrations was used to pass through the immobilised peptides. The data were analysed using Microcal Origin version 7.5 (Microcal Software Limited). The data were fitted using the Langmuir (1:1) model, and the following equation was used for a non-linear curve fitting to the data:

$$B = B_{\max}\left(\frac{[C^A]}{[C^A]+K_D}\right)$$

Where B indicates the binding response (RUs); "B_{Max}" is the maximum response (RUs); [L] indicates the free concentration of the analyte injected (µM); K_D

indicates the equilibrium binding constant (μM). The data were also used to produce Scatchard plots of Bound/C^A against Bound which provide a second K_D value.

2.10.4.3 System-level investigation of the affinities of SH2 domains binding to phosphotyrosine-containing receptor motifs

The affinity analyses of the interaction between the SH2 domain repertoire and the selected peptides were performed using the BIAcore3000, under conditions of 20 μl/min flow rate and 37°C and using HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% v/v Surfactant P20; GE Healthcare) as the running buffer. The tyrosyl phosphopeptides were custom-made by either the in-house peptide facility of WIMM or the Peptide Protein Research (PPR Ltd.). A 2-fold serial dilution of totally five concentrations was used to pass through the immobilised peptides. Washes of needles and IFC were performed between each injection. 10 μl of 25 mM NaOH / HBS-EP buffer (GE Healthcare) was injected between injections of the same set of analyte, and 30 μl of 25 mM NaOH (GE Healthcare) was injected at the end of the 5-injection cycle for regeneration. The data were analysed using the Macros in Excel. The data were used to produce Scatchard plots of Bound/C^A against Bound, from which the K_d values for the interactions were obtained.

Chapter 3. Transcriptome- and proteome-based analysis of SH2 domain expression[1]

3.1 Introduction

SH2 domains are protein modules that recognize phosphorylated tyrosine motifs and play an important role in intracellular signal transduction [224]. These domains are incorporated into proteins with a wide range of biochemical functions, including kinases, phosphatases, small GTPase signalling, phospholipid second messenger signalling, ubiquitination, transcription, adaptors and scaffolds [169]. The recruitment of SH2 domain-containing proteins (SH2 proteins) to the phosphorylated cytoplasmic tyrosine motifs of surface receptors represents the initial step of intracellular signal transduction [225]. To understand how these modules operate in different cellular systems, one would need to know first of all whether and how much each of the SH2 domains are expressed.

Whereas the proteomic-based approach would be ideal to tackle this problem, it has until very recently been a challenge to identify and obtain reliable quantitative data at great depth for most of the proteins expressed by a given cell. For example, only 2221 proteins of the human pathogen *Leptospira interrogans* (61% of the open-reading frames) have been absolutely quantified [211]. State-of-the-art mass spectrometry instruments, implementing the shotgun approach, are now starting to provide deep proteomic analysis of over 10,000 proteins [226]. However, since more than 22,000 proteins are predicted to be expressed in the human proteome [208], it seems that current proteomic approaches are still not capable of characterising complete sets of cellular proteins.

[1] The statistical analyses described in this chapter were done with the assistance of Dr. Oliver Lui, T-cell Biology Group, University of Oxford. The proteomic analyses were done in collaboration with Dr. Benedikt Kessler, Centre for Cellular and Molecular Physiology, University of Oxford.

In contrast, the transcriptomic approach, implemented in the form of large-scale gene expression analysis, is currently complementary to the proteomics method insofar as it provides useful information on the complete sets of proteins expressed by a given cell or tissue. For example, detection of transcripts corresponding to >90% of the genome of fission yeast *Schizosaccharomyces pombe* has been achieved by the RNAseq technology [201]. On the other hand, although modern transcriptomic analysis can allow quantification of protein-encoding transcripts this information does not necessarily reflect the actual protein expression levels [202]. It seems likely, however, that as the technology advances, proteomics will supersede the transcriptomic approach in the future.

3.1.1 Transcriptomic approaches

There are two gene expression profiling methods available nowadays: the "closed" technologies, such as microarrays, and the "open" technologies, such as the serial analysis of gene expression (SAGE) technology [196] and the more recently next-generation sequencing (NGS) technologies (Reviewed in [197]). Because the "open" technology does not, unlike the "closed" technology, rely on the prior knowledge of genes and allows identification of all possible transcripts, including those uncharacterised ones, it has been widely used and under extensive development. In the following section, the SAGE and NGE technologies will be briefly described.

3.1.2 SAGE technology

SAGE technology involves generation of short, transcript-specific tags that can be amplified and sequenced in a high-throughput manner. At the core of this

technology is the use of type IIS restriction enzymes, which have the special capability of cutting at a fixed distance downstream of the recognition sites (15bp for *BsmFI*; 21bp for *MmeI),* and thus preserving a short piece of sequence unique to the transcript of origin. Prior to generation of the SAGE tags, a library of cDNA attached to biotinylated magnetic beads is generated, followed by digestion by a frequently-cutting restriction enzyme (the anchoring enzyme, e.g. *NlaIII*) to create an "anchoring site". An adaptor encoding primer-binding sites for PCR amplification and the recognition sequence for a type IIS restriction enzyme (the tagging enzyme) is then ligated to the anchoring site. After the tagged-sequences are digested with the tagging enzyme, they can be ligated to form ditags that are available for PCR amplification. The amplified ditags can then be cleaved with the anchoring enzyme to remove the linkers, followed by ligation to form concatemers. These concatemers are then cloned and sequenced by Sanger's method [196, 198]. The accuracy in mirroring transcript abundance of SAGE has been proven by real-time PCR analysis [227]. The method has been used to characterise the gene expression patterns in a wide range of leukocyte populations, including $CD4^+$ [228] and $CD8^+$ [229] T cells subsets, intra-epithelial lymphocytes [227], and dendritic cells [230].

3.1.3 Next-generation sequencing

Compared to the tag-based methods such as SAGE, several next-generation sequencing (NGS) technologies have been able to offer the advantages of unprecedented depth and sensitivity of analysis of parallel direct sequencing. Among the commercially available NGS technologies, the most widely used is the Solexa technology owned by Illumina. This involves the generation of large number of single stranded DNA templates and a sequencing-by-synthesis approach with a four-colour cyclic reversible termination procedure.

During the solid-phase Bridge amplification step, a library of single-stranded cDNA fragments, with adaptors attached at two ends, are randomly bound to the inside surface of flow cells, where there is a dense lawn of covalently attached forward and reverse primers. After the completion of amplification, several million DNA clusters are generated. The DNA molecules are denatured into single stranded templates and each DNA stand has a free-terminus for subsequent hybridisation of a universal sequencing primer. The subsequent sequencing and imaging then employs cyclic reversible termination, which involves incorporation of four types of 3'-azidomethyl-dNTPs, each labelled with a different fluorescent dye. DNA synthesis is terminated following the addition of a single nucleotide, and the remaining un-incorporated nucleotides are washed away before imaging detection. This is then followed by a cleavage step that removes the terminating group and the fluorescent dye attached to the nucleotide. Thereafter, an additional washing step is performed before the next incorporation step is repeated, which eventually generates a large number of short sequences [197].

3.1.4 Mass spectrometry-based proteomic analysis

The 'shotgun' approach is the most widely used strategy in mass spectrometry-based proteomics and has shown great potential [231]. This involves digestion of proteins into peptides (usually by a sequence-specific enzyme such as trypsin), fractionation of the resulting peptide samples, separation by nanoflow high-perfomance liquid chromatography (HPLC), and the subsequent analysis by high-resolution tandem mass spectrometer (MS/MS). The resulting fragment ion MS/MS spectra are then assigned to sequences in a peptide database for identification of the corresponding proteins. In the case of label-free quantification of the proteome, the abundance of protein can be estimated using, for example, intensity based absolute quantification (iBAQ) or exponentially modified protein abundance index (emPAI) values [232].

3.1.5 The use of the Jurkat T-cell line in signal transduction research

Over the last two decades, intensive investigations have unveiled the signal transduction mechanisms of T cells. Many of the fundamental insights came from studies using transformed T-cell lines as model systems, with the Jurkat T-cell line being the best known one. The Jurkat T-cell line was initially screened and selected as a leukaemic cell line capable of robust production of IL-2, which is an essential growth factor for maintenance of cultured human T cells [233]. Since then, the Jurkat T cell has been used widely to investigate TCR signalling mechanisms. One of the earliest series of studies involved generation of a TCR-negative Jurkat subline, and revealed that TCR ligation triggers intracellular Ca^{2+} mobilisation (reviewed in [234]). The number of publications that list the term 'Jurkat' as a key word increased dramatically from a few per year in early 1980s to more than 500 per year by the early 2000s. However, the emergence and development of in vivo experiments using transgenic and knockout mouse models have led some to question the physiological relevance of in vitro experiments carried out with Jurkat and other transformed T-cell lines. Particularly, Jurkat cells were shown to be defective in the lipid phosphatase PTEN, which leads to constitutive activation of the PIK3-signalling pathway [235], although the importance of PIK3 in the early stages of signal transduction in activated T cells is still unclear [236].

Despite of these issues, Jurkat cells still have a number of notable advantages compared with primary human T cells and the in vivo model systems. Firstly, maintenance and genetic manipulations, such as somatic cell knockouts and knockdowns [via small interfering RNA (siRNA) technology; [237]], of T-cell lines are much easier than in cultured human T cells. Secondly, production of transgenic and knockout mice requires considerable time and expense. Thirdly, although it is assumed that the physiology in the mouse model system parallels

that of humans, this is not always the case. For example, whereas peripheral $CD4^+$ T cells are absent in ZAP70-deficient mice [238], they are abundant in ZAP70-deficient humans [239]. Thirdly, somatic mutations of genes for important signalling proteins can result in developmental blocks, and it would be difficult or impossible to investigate the function of these proteins in cultured human T cells. Furthermore, functional defects arising in mature T cells with a knockout genetic background may not reflect a direct role of the targeted protein in normal T cells but, instead, result from compensatory responses triggered by survival pressure. As a consequence, T-cell lines remain an invaluable complement to experiments carried out with cultured human T cells and mouse models. Last but not least, in the present study, cell lines such as Jurkat will be used as a model to study how signal transduction is configured in any eukaryotic cell, rather than just in T cells.

3.2 Results

Since the launching of SAGE technology, a number of transcriptomes have been released publicly, such that a collection of 46 SAGE libraries are now accessible [200, 229, 230, 240-243], including 22 immune cell libraries, among which there are 10 T cell-derived libraries, as well as 24 non-immune cell libraries (**Table 3.1**). With these relatively deep transcriptomes, it has become possible to explore the number of SH2 proteins transcribed in a cell and compare the SH2 protein expressions in different cell types at the mRNA level. In addition, the availability of very large SAGE libraries prepared in this laboratory from a CD4+ T cell clone (clone 29) in both resting and activated states (~500k tags/each library) has allowed us to investigate how the transcription of SH2 proteins changes upon activation. In addition, a deep transcriptome for Jurkat T cells was generated in our laboratory using the Solexa technology, allowing comparisons between a normal T cell clone and Jurkats (unpublished). Finally, a published deep proteomic analysis

Table 3.1. A list of SAGE libraries available for this study.

Immune cells				Non-immune cells
	T cells		**Non-T immune cells**	
1)	Resting $CD4^+$ cells (Clone 29)	1)	Oxford NK cells	1) Cerebellum CGAP
2)	Activated $CD4^+$ cells (Clone 29)	2)	Tokyo NK cells	2) Cerebellum Duke
3)	Resting Th1 $CD4^+$cells	3)	Resting B cells	3) Thalamus
4)	Resting Th2 $CD4^+$ cells	4)	Activated B cells	4) White matter
5)	Activated Th1 $CD4^+$ cells	5)	Granulocyte	5) Astrocytes
6)	Activated Th2 $CD4^+$ cells	6)	Monocyte	6) Colon 1
7)	Resting $CD8^+$ cells (Clone 32)	7)	LPS treated Monocyte	7) Colon 2
8)	Activated $CD8^+$ cells	8)	MCSF induced Macrophage	8) Stomach epithelium
9)	$CD8^+$ cells treated with anti-CD8 Ab	9)	GMCSF induced Macrophage	9) Peritoneum mesothelium
10)	ex-vivo CTL cells	10)	Langerhans-like cells	10) Heart
		11)	Immature DC	11) Kidney
		12)	Mature DC	12) Liver
				13) Lung
				14) Muscle from young person
				15) Muscle from older person
				16) Retina epithelium
				17) Retina macula
				18) Retina peripheral
				19) Treated retina peripheral
				20) Prostate
				21) Cervix
				22) Ovary epithelium
				23) Breast epithelium
				24) Vascular endothelium

of Jurkat T cells [226], wherein the proteins were semi-quantified, allows comparison of SH2 protein transcription in these cells at the mRNA and protein levels.

3.2.1 Comparisons of SH2 domain-containing protein transcription between different cell types using SAGE

In this section, a number of transcriptome-based comparisons are made between different cell types/tissues in order to characterise the transcription patterns of SH2 domain-containing proteins (SH2 proteins).

3.2.1.1 Immune and non-immune cells

A comparison between the immune and non-immune group SAGE libraries shows that the numbers of SH2 proteins expressed detectably in both groups is comparable (**Figure 3.1**) [47.8±0.6 (immune cells) versus 44.3±0.4 (non-immune cells)]. However, for abundance thresholds greater than 2 tags/10^5 tags and 8 tags/10^5 tags, the number of distinct transcripts identified in the non-immune cells is substantially lower than that in the immune libraries, i.e. less than half that for the most abundant transcript class. The larger number of highly transcribed genes encoding SH2 domain-containing proteins suggests that SH2 domains may have more important roles in immune cells than in non-immune cells.

Further comparisons of the transcription of SH2 proteins in immune cells and non-immune cells showed that there is a significant difference in the average levels of SH2 domain-encoding mRNA transcription [$p<0.02$; **Figure3.2 (A)**]: whereas there are 3.27 (±0.15) tags/10^5 for each SH2 protein on average in the immune cell libraries, the SH2 proteins in the non-immune cell group are

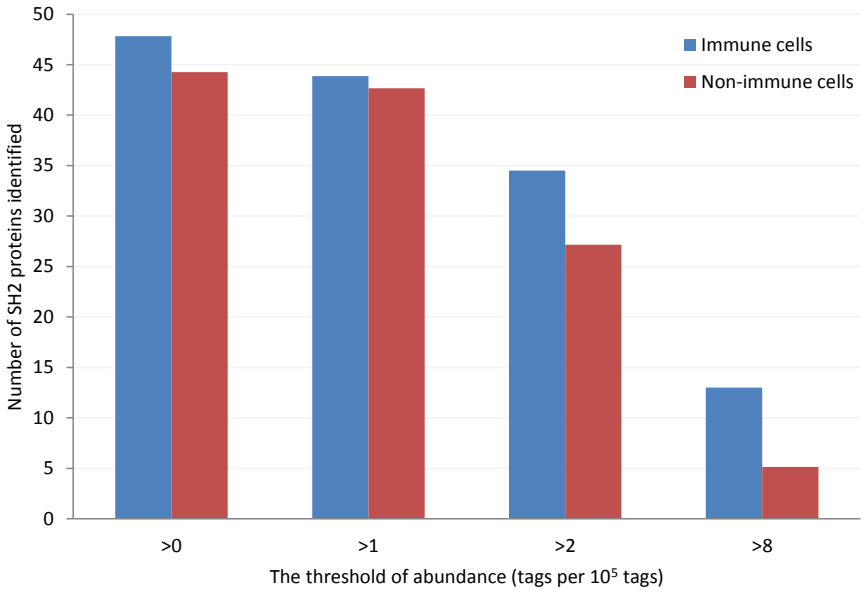


Figure 3.1. The average number of SH2 proteins identified in immune cells versus non-immune cells. Blue and red bars indicate data from immune and non-immune libraries, respectively. Four different thresholds for transcript abundance were set, which are above 0, 1, 2 and 8 tags per 10^5 tags.

expressed on average at 2.20 (±0.42) tags/10^5. In contrast to SH2 proteins, the mRNA expression of PTB domain containing proteins (another modular protein family involved in signal transduction) is significantly higher in non-immune cells compared to immune cells [**Figure 3.2 (B)**; $p<0.02$, the average number of tags/10^5 tags for each PTB protein is 2.24±0.26 (immune) versus 3.07±0.23 (non-immune)]. This implies that these two families of protein are involved in differential signalling mechanisms. Interestingly, the bias of SH2 transcription in immune cells correlates with the significantly higher transcription of tyrosine-containing surface receptors in immune cells [**Figure 3.2 (C)**; $p<0.01$, the average number of tags/10^5 for tyrosine motif-containing receptors is 2.98±0.25 (immune) versus 0.98±0.09 (non-immune)]. This suggests that the phosphotyrosine-dependent signalling system plays a particularly important role in immune cells.

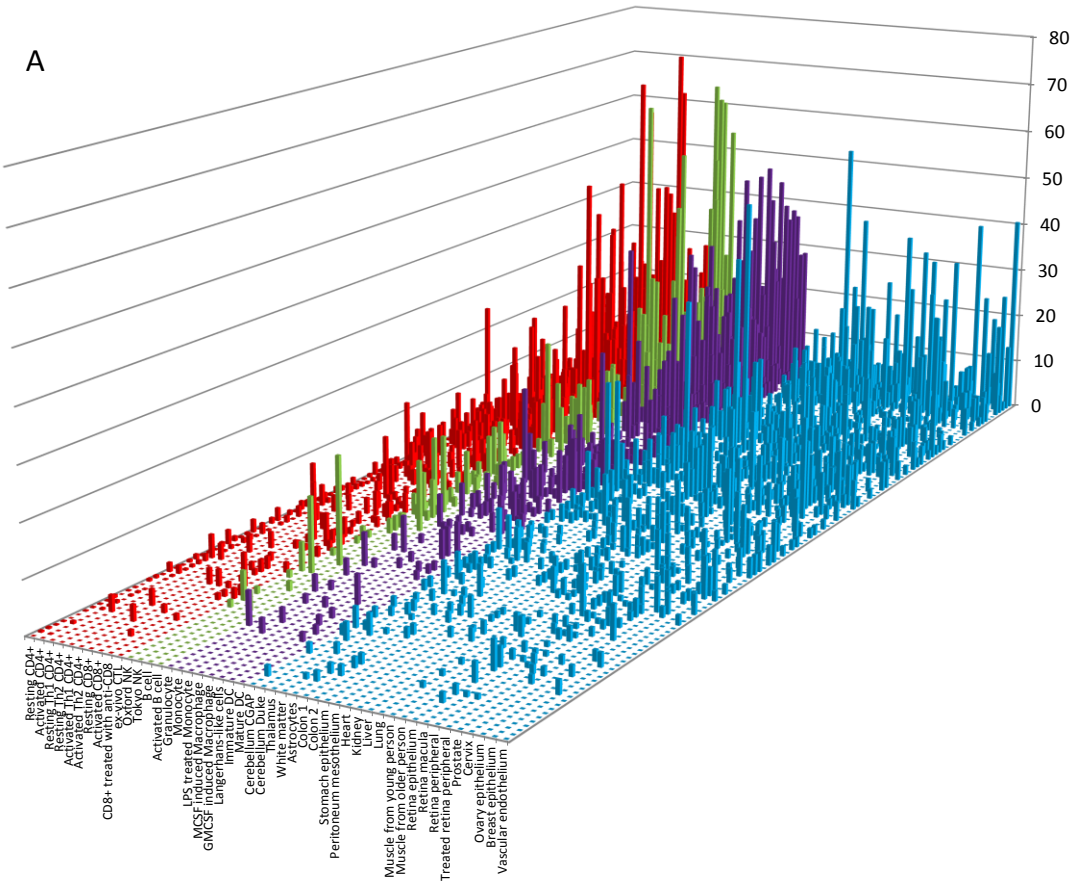


Figure 3.2 (A). The mRNA expression for SH2 domain-containing proteins in 46 different SAGE libraries. Transcription levels for the SH2 proteins in T cells (red), NK and B cells (green), myeloid cells (purple) and non-immune tissues (blue) are expressed as tags/10^5 tags.

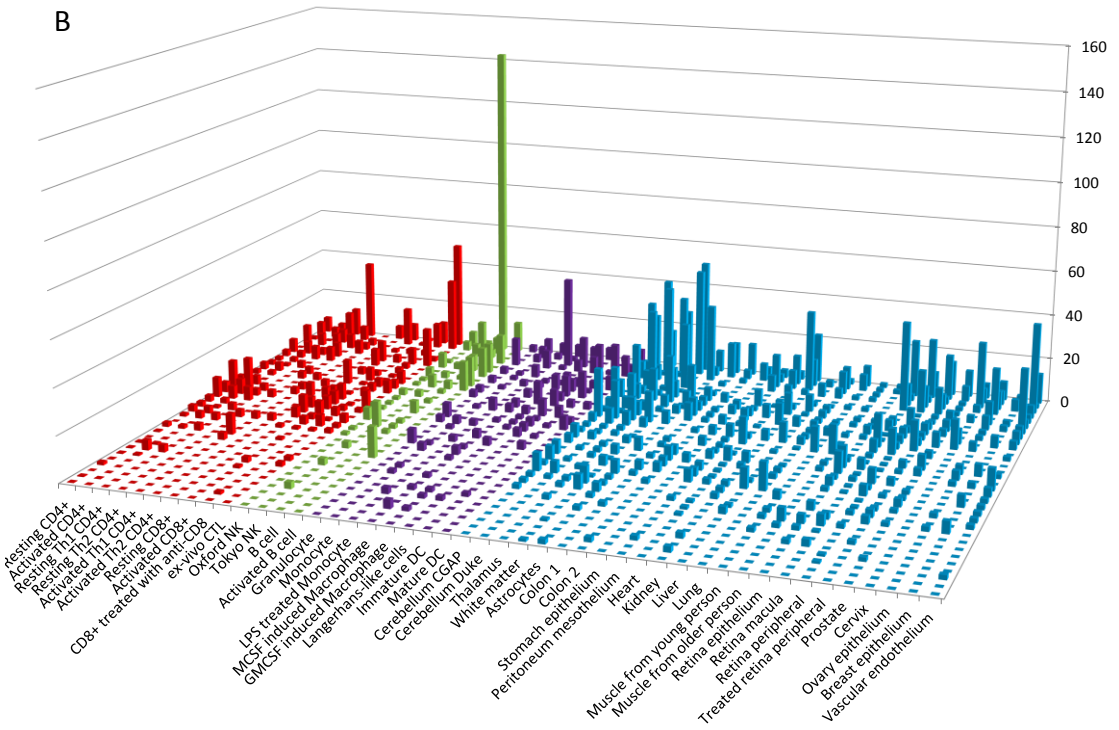


Figure 3.3 (B). The mRNA expression for PTB domain-containing proteins in 46 different SAGE libraries. Transcription levels for the SH2 proteins in T cells (red), NK and B cells (green), myeloid cells (purple) and non-immune tissues (blue) are expressed as tags/10^5 tags.

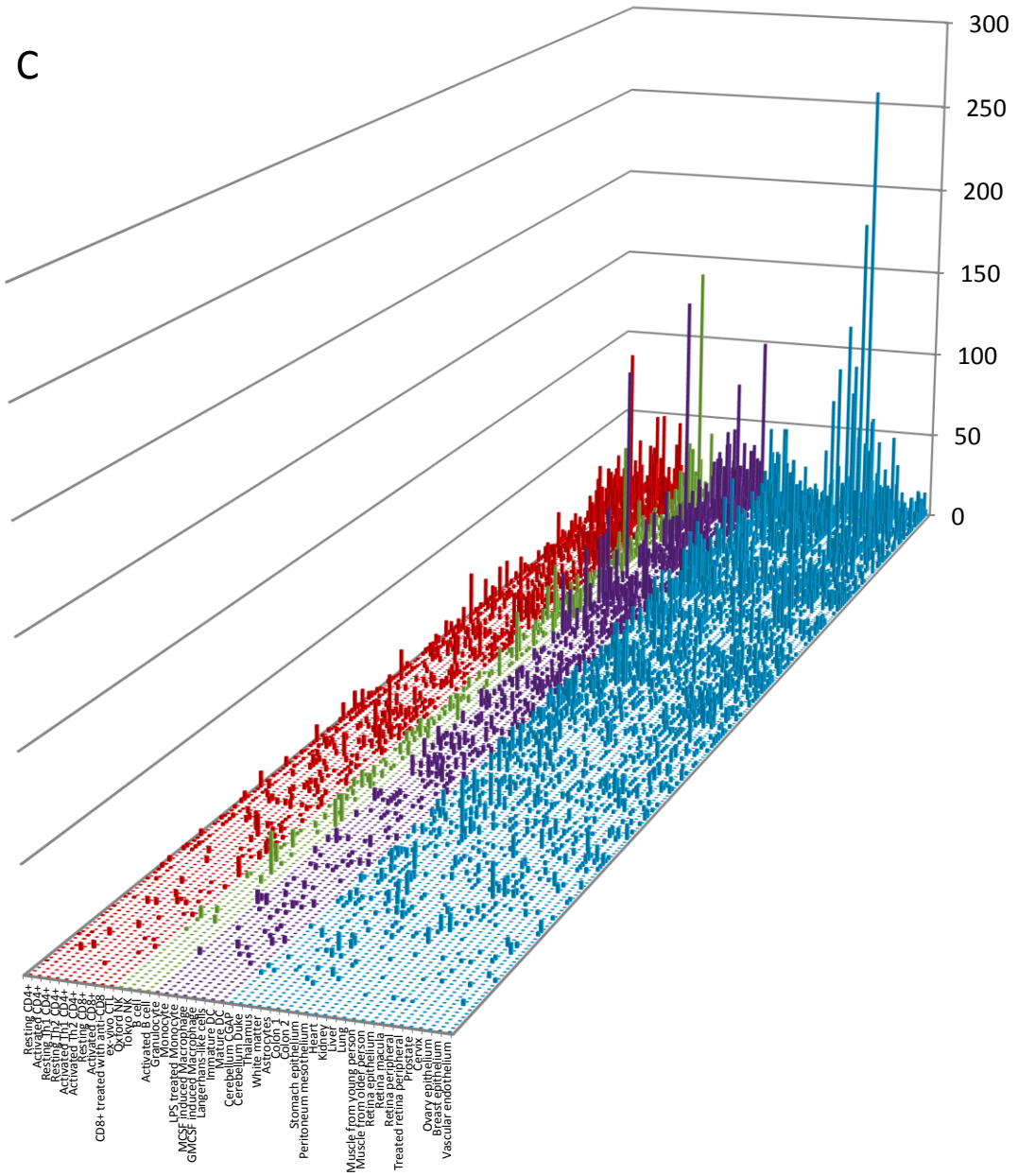


Figure 3.4 (C). The mRNA expression for tyrosine motif-containing receptors in 46 different SAGE libraries. Transcription levels for the SH2 proteins in T cells (red), NK and B cells (green), myeloid cells (purple) and non-immune tissues (blue) are expressed as tags/10^5 tags.

By comparing the transcription of individual SH2 proteins in these two groups, the transcription of 26 was found to be significantly over-transcribed in immune cells ($p < 0.05$, t-test; **Table 3.2**); 24 of these have been previously reported to be involved in immune cell signalling (references are listed in **Table 3.2**). Whereas a number of well-known signalling molecules, including ZAP70, LCK and SHP1, are included, it was surprising to see RIN3 ranked at the top of the list since this protein is not reported to participate in any specific signalling pathways in immune cells. Intriguingly, a number of signalling molecules that are important in immune cells, such as SHC1 [244], SHP2 [245] and TEC [246], are not on the list. This possibly suggests that the signalling mechanisms they may participate in maybe part of 'house-keeping' processes.

3.2.1.2 T cells and non-immune cells

The difference between immune and non-immune cells in SH2 protein transcription led us to consider how T cells differ from non-immune cells. As expected, comparison between the two groups shows that they are significantly different from each other ($p < 0.05$, t-test), although the significance is lower than that for immune cells as a whole [$p_{(immune/non\text{-}immune)} = 0.018$; $p_{(T\text{-}cell/non\text{-}immune)} = 0.042$)]. This is due to the slightly lower abundance of SH2 mRNA for the T cell group compared to the whole immune cell group (3.15±0.21 versus 3.27±0.15 tags/10^5 per SH2 protein). When individual genes were examined, there are 21 genes whose transcription is biased toward T cells ($p < 0.05$, t-test), all of which are included in the list of SH2 proteins whose transcription is immune cell-biased (**Table 3.2**), except CBLB. The SH2 proteins biased toward the immune group but not T cells include SHP1, FGR, SYK, FES, HCK and SH3BP2. Overall the mRNA transcription of SH2 proteins in T cell is comparable to that of all immune cells.

Table 3.2. List of SH2 proteins whose encoding mRNA expression is significantly higher in immune versus non-immune cell libraries[2]. The p-values given are derived from t-tests.

	Protein	P-value (Immune/ Non-immune)	Ref		Protein	P-value (Immune / Non-immune)	Ref
1	RIN3*	0.000007	N/A	14	CHK	0.007017	[247]
2	GRB2	0.000030	[99]	15	SLA2	0.007161	[248]
3	CSK	0.000222	[249]	16	SYK	0.007248	[250]
4	SHIP1	0.000241	[251]	17	FES	0.008015	[252]
5	STAT6	0.000852	[253]	18	LCK	0.008446	[254]
6	SLAP	0.001090	[255]	19	ZAP70	0.008886	[256]
7	JAK3	0.001604	[257]	20	STAT5B	0.008933	[258]
8	SHP1	0.001846	[259]	21	LCP2	0.009091	[260]
9	SH2D2A	0.002217	[261]	22	HCK	0.011339	[262]
10	FGR	0.002386	[263]	23	STAT5A	0.012055	[258]
11	GRB7	0.002734	[264]	24	SPT6*	0.012400	N/A
12	STAT4	0.002767	[265]	25	SAP	0.029867	[266]
13	VAV1	0.004689	[267]	26	SH3BP2	0.046233	[268]

[2] SH2 proteins labelled with (*) have not been reported to be involved in immune cell signalling according to the literature.

3.2.1.3 T- versus non-T immune cells

When the mRNA expression of SH2 proteins was compared between T cells and other immune cells (excluding T cells), no significant difference was found ($p>0.05$, t test). However, 9 SH2 proteins had significantly higher transcription in T cells, 7 of which have been reported to play a role in T cell signal transduction (**Table 3.3**). Whereas LCK was shown to rank at the top, ZAP70, another crucial signalling molecule involved in the earliest stage of T cell activation [269], is to our surprise not in the list. This is due to the high transcription of ZAP70 in NK cell libraries. The transcription of CHN2 and SHC4 is significantly biased toward T cells, mainly due to the absence of transcription in other immune cell types, but their transcriptions are actually very low in T cells. Interestingly, although SHP1 [259], CSK [249] and GRB2 [99] are reported to be involved in important signalling mechanisms in T cells, their expressions are biased toward non-T immune cells (data not shown).

3.2.1.4 The $CD4^+$ T cell clone

Clone 29 is a $CD4^+$ T cell clone derived from peripheral blood mononuclear cells of a HIV vaccinee. The SAGE data derived from both activated and resting states of this clone thus provides insights into the SH2 protein expression of a normal human T cell. The SAGE data reveals that there were 100 SH2 proteins transcribed in the T cell in the resting state, and 98 in the activated state, giving a total of 105 proteins in either cellular state (**Figure 3.3**). The larger number of proteins identified in this clone [versus approximately 50 proteins found in the non-immune cell comparisons (**Figure 3.1**)] results from the greater depth of the analysis. A total of 53 proteins were upregulated upon activation and 50 were down regulated, and only 2 proteins were unchanged (Error! Reference source not found.). Eight proteins

Table 3.3. List of SH2 proteins whose encoding mRNA expression is significantly higher in T-cell versus "non-T" immune cell libraries. The p-values given are derived from t-tests.[3]

	Protein	P-value (T/NonT-immune)	Ref
1	LCK	0.005581	[254]
2	CBLB	0.018297	[270]
3	STAT4	0.022529	[265]
4	FYN	0.027083	[271]
5	CHN2*	0.036923	N/A
6	STAT5B	0.041719	[258]
7	SHC4*	0.042010	N/A
8	SAP	0.043383	[272]
9	SH2D2A	0.043609	[261]

[3] SH2 proteins labelled with (*) have not been reported to be involved in immune cell signalling according to the literature.

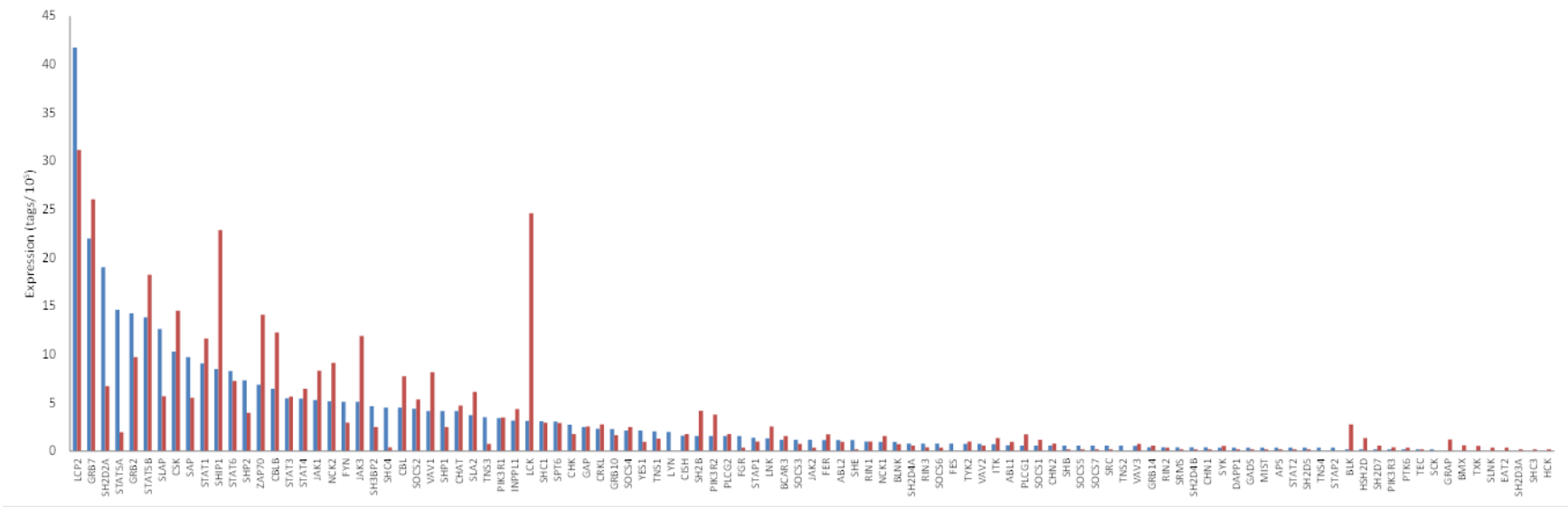


Figure 3.3. SAGE analysis of SH2 domain-encoding mRNA expression in a CD4⁺ T cell clone (Clone 29). Blue and red bars correspond to transcription in activated and resting T cells, respectively.

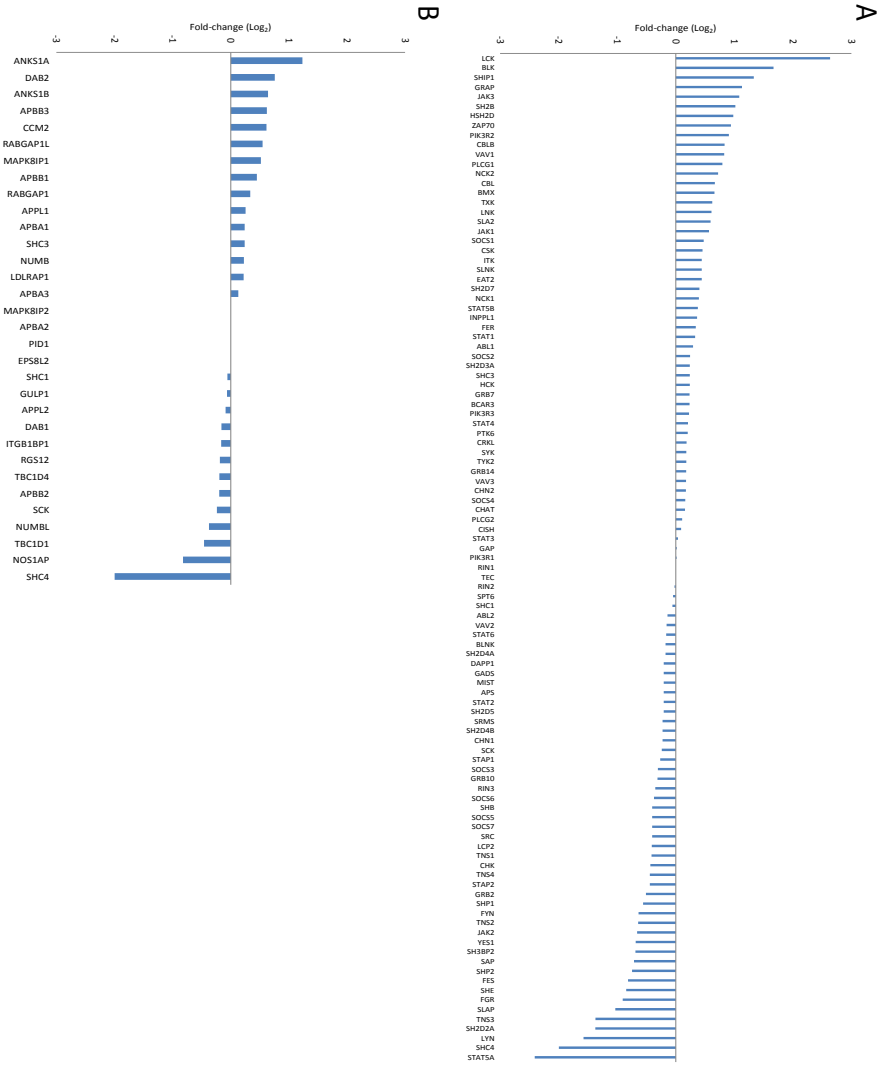


Figure 3.4. Fold changes in transcription of (A) SH2 domain-encoding transcripts and (B) PTB-domain-encoding transcripts between activated and resting cells. The fold-changes are presented as a scale of log2.

were exclusively transcribed in resting cells (i.e. GRAP, BMX, TXK, EAT2, SLNK, HCK, SH2D3A and SHC3) and six exclusively in activated cells (i.e. FES, LYN, SCK, STAP2, TNS2 and TNS4). The expression of SH2 proteins is therefore highly regulated. A number of the T cell-specific SH2 domains (**Table 3.2**), such as LCP2, SH2D2A, SLAP and SAP, are among the most abundant after the cells are

activated, indicating their important roles in mediating signal transduction following activation. Surprisingly, whilst LCK is the most highly transcribed SH2 domains followed by LCP2 and GRB7 in resting T cells, its RNA transcription reduced by 8-fold upon activation, implicating a highly dynamic change of expression of this important Src kinase.

An additional surprising finding is that the magnitude of change in SH2 protein mRNA expression is relatively high: only 65% of the transcript expressions have a less than 1.5-fold change, and 11% of the transcriptions have a more than 2-fold change. This is in contrast to, for example, the mRNA expression of PTB domain-containing proteins, which is relatively stable, with virtually all varying within a 2-fold change only **(**Error! Reference source not found.). The fact that there is nearly a 4-fold change in transcription of SHC4 (a PTB domain-containing protein) might be due to this protein containing an SH2 domain. The differential transcription of the SH2 proteins further suggests that they are playing a crucial role in signal transduction.

3.2.1.5 $CD4^+$ T cells and cerebellum

To further explore the discrepancy in SH2 protein transcription between T cells and un-related cells, SAGE libraries of Clone 29 $CD4^+$ T cells and cerebellum were compared with respect to transcripts encoding SH2 domains. Spearman's rank correlation calculations give a value of 0.052 and 0.142 for the resting and activated T cells, respectively, against the cerebellum cells. In marked contrast, comparison between the $CD4^+$ T cells (clone 29) and $CD8^+$ T cells (clone 32) yields a much higher correlation coefficient (0.662). The scatter plot in [**Figure 3.5 (A)**] shows that a number of SH2 proteins are transcribed in T cells only, including ZAP70, CSK, SAP, LCP2, SHP1, SHP2, SHIP1, and SHIP2, whereas transcription of CHN2 is markedly higher in brain than T cells. In contrast to the

cerebellum where the expression of SH2 proteins is generally not correlated to that in the $CD4^+$ T cells, $CD8^+$ T cells have much higher degree of correlation with the $CD4^+$ T cells [**Figure 3.5 (B)**].

3.2.1.6 $CD4^+$ T cells and Jurkat

Concerns over whether Jurkat T cell lines are representative of normal human T cells prompted an investigation of the similarity in SH2 transcription profiles between these two cells. Comparisons between resting clone 29 T cells (using SAGE data) and Jurkat (using Solexa data) yields a Spearman's rank correlation value of 0.58, suggesting that the SH2 transcription profiles in Jurkat T cells and normal human T cells are similar [**Figure 3.6 (A)**]. In contrast, when clone 29 T cells were compared with 293T cells (using Solexa data), which are of kidney origin [273], a much lower correlation value (0.31) was obtained [**Figure 3.6 (B)**].

When examining the number of SH2 proteins transcribed in clone 29 and Jurkats (**Figure 3.7**), it was found that a slightly larger number of SH2 proteins was transcribed in normal T cells (106 versus 85). By sorting the mRNA expression from high to low, a gradual decrease in abundance can be seen for Clone 29. In contrast, there is very high transcription of ZAP70 and LCK in Jurkat T cells, which are about 2-fold higher than the third most abundant protein (SLAP).

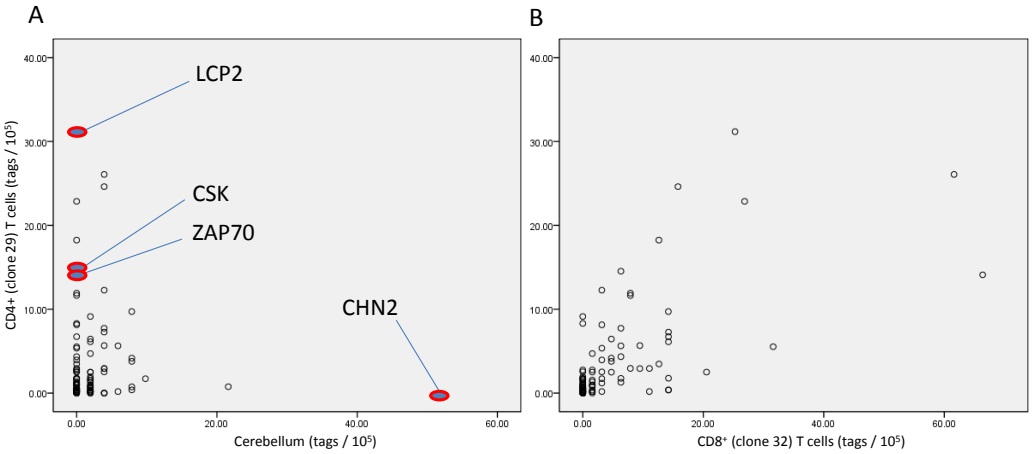


Figure 3.5. Comparison of SH2 protein-encoding mRNA expression between $CD4^+$ T cells (clone 29) and cerebellum as well as $CD8^+$ T cells (clone 32). The abundance of each transcript (tags per 10^5 tags) in the resting $CD4^+$ T cell (clone 29) library is plotted against those in libraries from (A) cerebellum and (B) resting $CD8^+$ T cells.

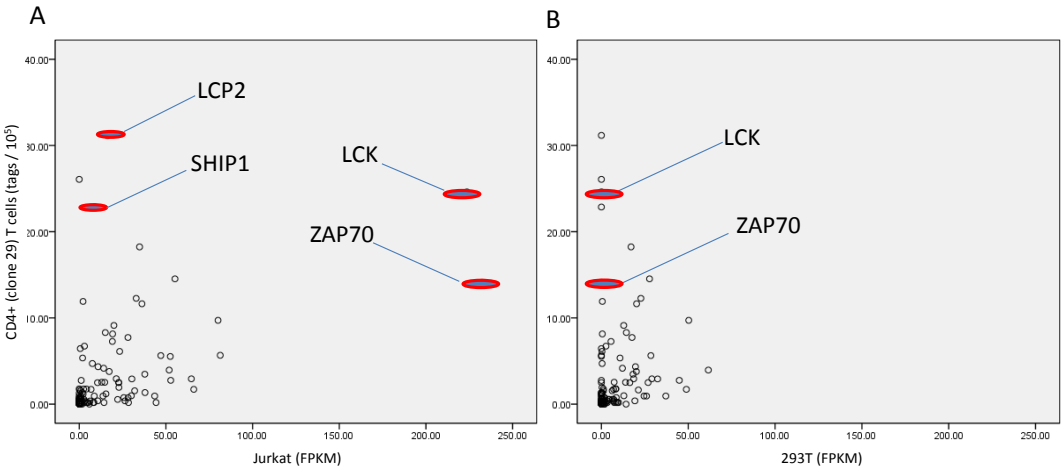


Figure 3.6. Comparison of protein-encoding mRNA expression between $CD4^+$ T cells (clone 29) and Jurkat T cells as well as 293T cells. The abundance of each transcript (tags per 10^5 SAGE tags) in the resting $CD4^+$ T cell (clone 29) library is plotted against the abundance (FPKM, fragments per kilo-base of exon per million fragments mapped) in (A) Jurkat and (B) 293T cells.

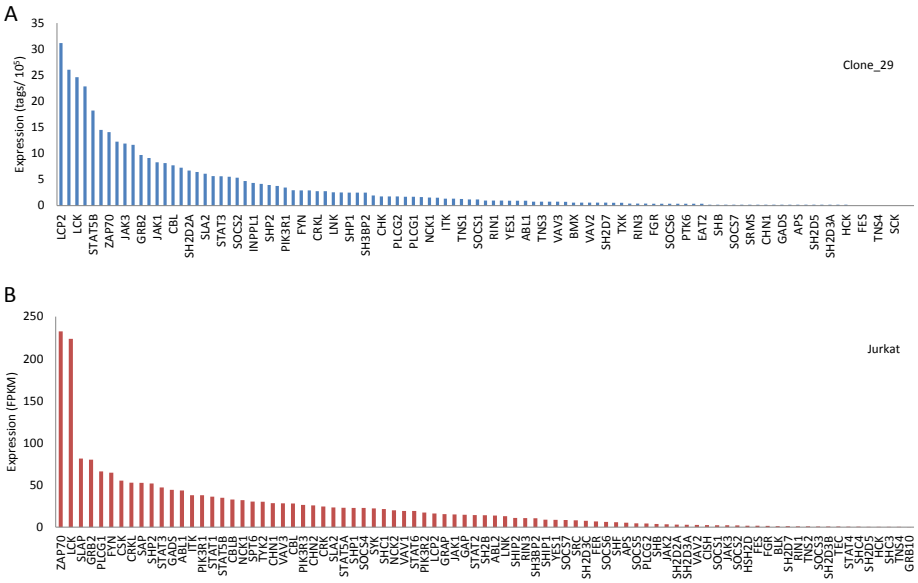


Figure 3.7. SH2 domain-encoding mRNA expression in clone 29 and Jurkat T cells. (A) The mRNA expression in clone 29 is derived from a SAGE-based transcriptome, whose abundance is expressed as numbers of tags per 10^5 SAGE tags. (B) The mRNA expression in Jurkat T cells is derived from a Solexa-based transcriptome, whose abundance is expressed as FPKM.

Interestingly, whereas LCP2 and SHIP1 are highly transcribed in Clone 29, their transcription was low in Jurkats. Overall, however, this analysis suggests that Jurkat T cells, to a large extent, resemble the $CD4^+$ T cell clone.

3.2.2 Proteomic analysis of SH2 protein expression

Recently, Mann and co-workers have published a deep proteomic analysis of 11 cell lines including Jurkat T cells that involves identification and semi-quantification of more than 10,000 proteins [226]. Using a label-free approach, proteins identified in their study were quantified by the iBAQ (intensity based absolute quantification) algorithm, which essentially normalises the summed peptide intensities by the number of theoretically observed peptides of the protein.

In order to have a better idea of the actual SH2 expression profile, the iBAQ values of the protein were converted into absolute concentrations, as described in the next section.

3.2.2.1 Obtaining absolute protein concentration values

Transformation of the (semi-quantified) iBAQ values into absolute concentrations requires a calibration curve that relates the two parameters [211]. This can be achieved by measuring the iBAQ values of a set of anchor proteins with known concentrations, which can then act as calibration points for translating the relative abundance (iBAQ) into absolute concentration. To this end, a commercially produced Universal Proteomics Standard-2 (UPS-2) Set was used for mass spectrometry-based quantification to obtain the calibration curve.

The UPS-2 proteomics standard set includes a mixture of 48 individual human source or human sequence recombinant proteins to present a dynamic range over 5 orders of magnitude (from 500 fmoles to 50 amoles). Using an LTQ Orbitrap Velos Mass spectrometer coupled to a nano-UPLC system, 22 components of this standard set were identified and quantified as emPAI (exponentially modified protein abundance index) values via the Mascot search engine [274]. A calibration curve generated according to the amount of the UPS-2 components analysed by the mass spectrometer as well as their corresponding emPAI values is shown in **Figure 3.8**. It was shown that the plots were well-correlated (R^2 = 0.642), and that the correlation is compromised by samples with an emPAI value lower than or around 0.1. This means the actual amount of sample determined from an emPAI value lower than this threshold is less reliable.

Because the UPS-2 standard was quantified (in the Kessler group) using a different instrument from that used to quantify the proteome of the Jurkat T cells

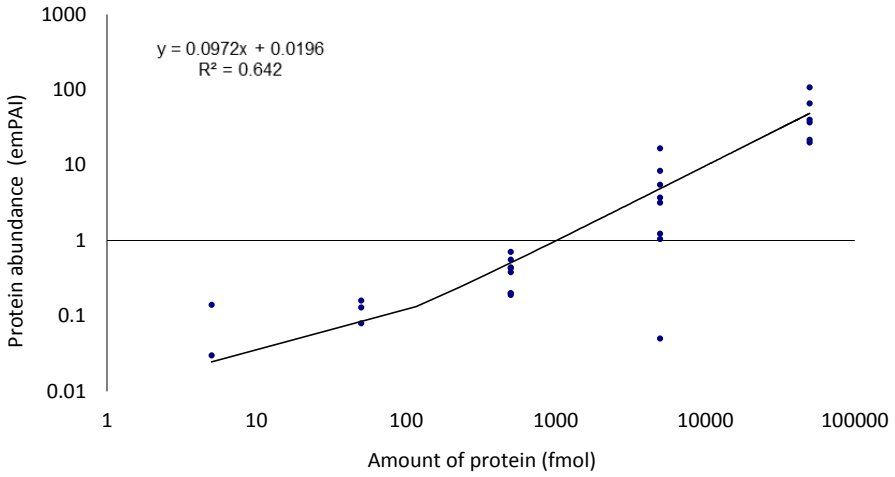


Figure 3.8. Determination of a calibration curve that relates the protein abundance (emPAI) measured by mass spectrometry and the amount of protein analysed using a UPS-2 protein standard mix.

by Mann and co-workers, it was not possible to perform the transformation directly. However, because a Jurkat T cell sample has been previously subjected to mass spectrometric analysis and quantified (in collaboration with the Kessler group) this can be used to link the calibration generated by the UPS-2 standard set and the iBAQ data set generated by Mann and co-workers, even though the depth of the Kessler group library is relatively low (with only approximately 2500 proteins identified).

To this end, a set of proteins shared by the two Jurkat proteomes but whose concentrations span a wide range were selected to determine the amount of protein analysed by the mass spectrometer (**Table 3.4**). Assuming that the concentration of the corresponding proteins in the Jurkat samples from the Kessler group and the Mann's group is the same, a calibration curve relating the iBAQ value and the amount of protein analysed by mass spectrometry was be generated (**Figure 3.9**). The very high correlation coefficient (R^2=0.9) for the

Table 3.4. A summary of the mass spectrometry-based measurements for a set of proteins shared by Jurkat samples from the Kessler group and the Mann group [226] proteins. The amount of protein (fmol) is calculated from the emPAI values of the proteins according to the calibration formula generated from the UPS-2 proteomic standard set. The iBAQ values are obtained from the mean of triplicates.

Protein	**MS abundance/ emPAI (From Kessler group)**	**Concentration of protein (fmol/µg)**	**MS abundance/ iBAQ (From Mann group)**
HSP60	24.71	254	8.17024
PGK1	13.78	141	8.24927
VCP p97	1.72	17.49	7.65579
HSPA5	2.46	25.11	7.67628
TXNDC17	0.24	2.27	7.18491
FKBP1A	0.31	2.99	7.59018
LPXN	0.07	0.52	6.65522

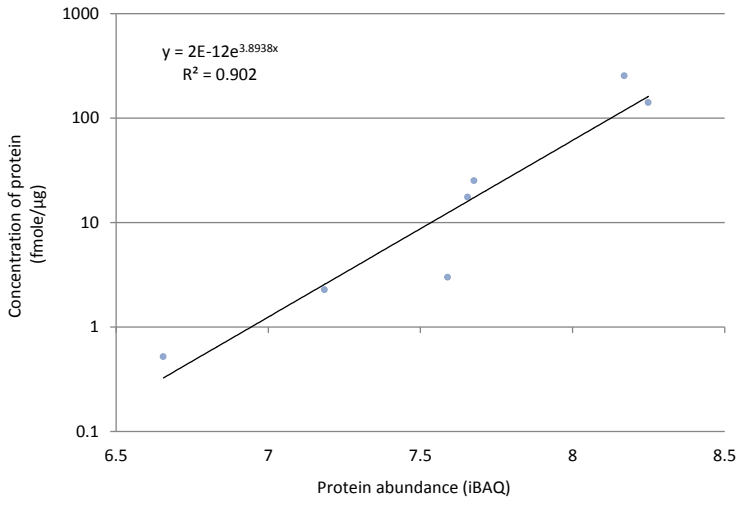


Figure 3.9. Calibration curve relating the protein abundance (iBAQ) measured by mass spectrometry and the concentration of the protein analysed.

plot indicates that the transformation from iBAQ values into absolute amount is likely valid. Assuming that 1 µg of protein mixture was analysed by the mass spectrometer, based on the limit of the UPLC column capacity, the amount of protein calculated from the calibration curve can be estimated to be the amount of protein present in 1 µg of whole-cell proteins in Jurkat T cells (fmol/µg).

3.2.2.2 The concentrations of SH2 proteins in a Jurkat T cell line

Using the calibration curve derived as described above, the concentration of a total of 49 SH2 proteins identified in the Jurkat proteome (by Mann and co-workers) was determined. It was found that their concentrations vary from 66 to 3.8 $x10^{-6}$ fmol/µg [**Figure 3.10 (A)**], exhibiting an approximately S-shaped distribution over seven orders of magnitude [**Figure 3.10 (B)**]. This is in great contrast to the approximately 500-fold range in expression seen in the transcriptomic data (for the corresponding set of proteins). It turns out that the concentration of most proteins is very low, with only 13 having concentrations higher than 1 fmol/µg [**Figure 3.10 (C)**].

Comparing this data with the mRNA expression data for Jurkats shows that these 49 proteins are all present in the transcriptomic set, whereas 36 are expressed at the mRNA level but not detected by proteomic analysis (**Figure 3.11**), including SLAP, which has a relatively high abundance at the transcript level (81 FPKM). Otherwise, proteins not included in the proteomic data set consistently have an mRNA abundance lower than 25 FPKM. Several proteins, including FYN and PIK3R1, are among the most abundant at the transcript level (>30 FPKM), but their concentrations are found to be very low in the proteome (<0.1 fmol/ug).

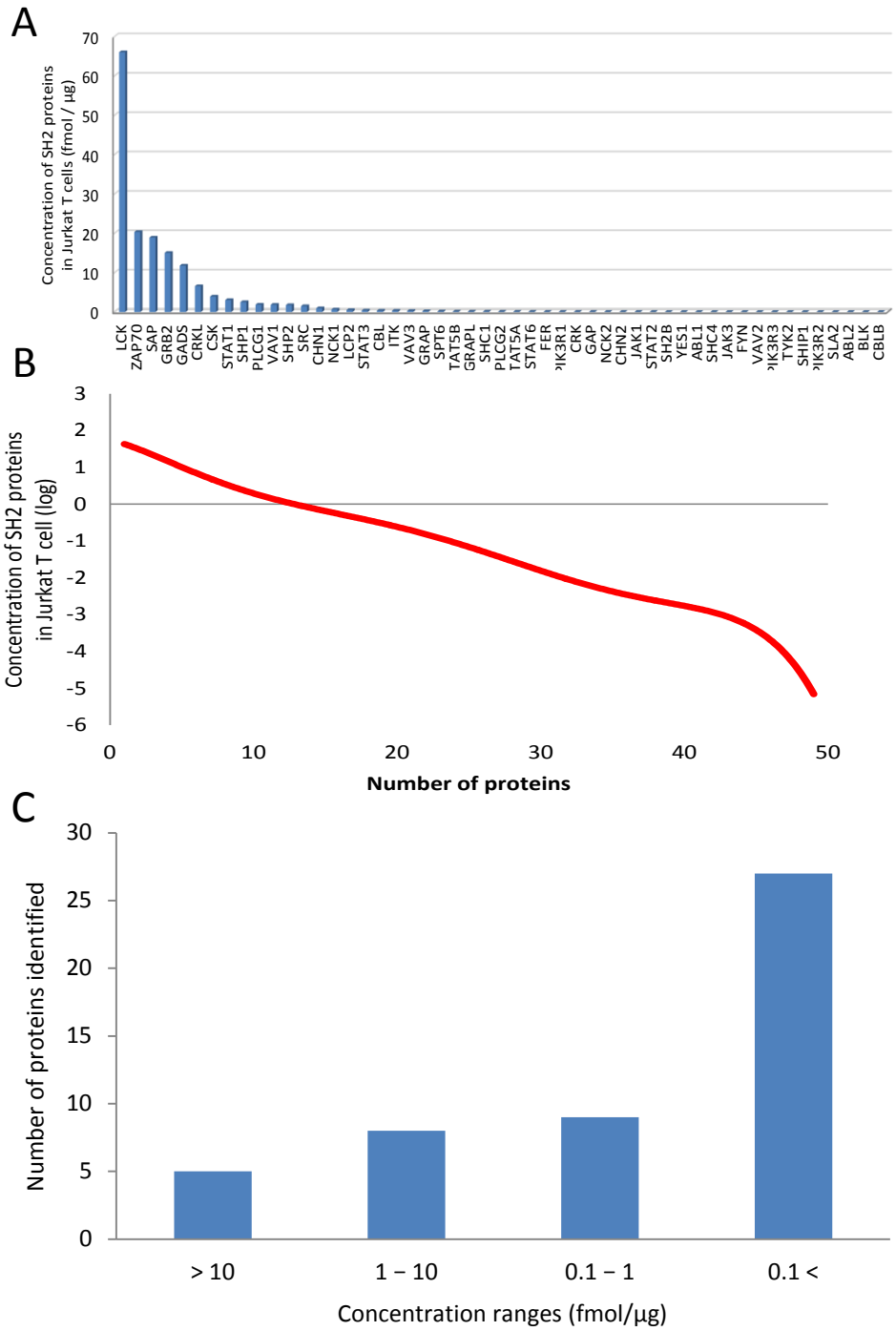


Figure 3.10. The concentration of SH2 proteins in a Jurkat T-cell line. (A) The concentrations in fmol/ug are shown sorted from high to low on a linear scale. (B) The concentration expressed in $\log_2$. (C) The numbers of SH2 proteins at different concentration ranges.

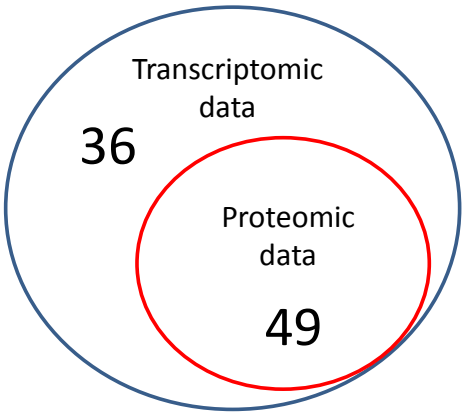


Figure 3.11. A Venn diagram showing the number of SH2 proteins identified in the transcriptomic and proteomic data. A total of 49 SH2 proteins were identified using the proteomic approach, which are all included in the total set of 85 SH2 proteins identified using transcriptomics.

Despite these differences the Spearman's rank correlation between these two sets of data was relatively high (0.56), which indicates that they are significantly correlated (**Figure 3.12)**. Overall, the expression of some of the most abundant proteins is well correlated in both data sets, most notably LCK and ZAP70.

3.3 Discussion

The emergence of "open" gene expression profiling methods and the increasing number of publicly released transcriptomes have allowed us to identify which proteins are expressed in a cell and the expression levels of transcripts encoding these proteins at great depth. Analyses in this chapter have revealed that, although SH2 proteins are widely expressed over different types of tissues, the overall transcription pattern is cell-type specific and considerable differences exist between immune cells and non-immune cells. Although the transcription patterns are similar between sub-groups within, for example, the immune cells, the transcription profile of certain SH2 proteins can be significantly different.

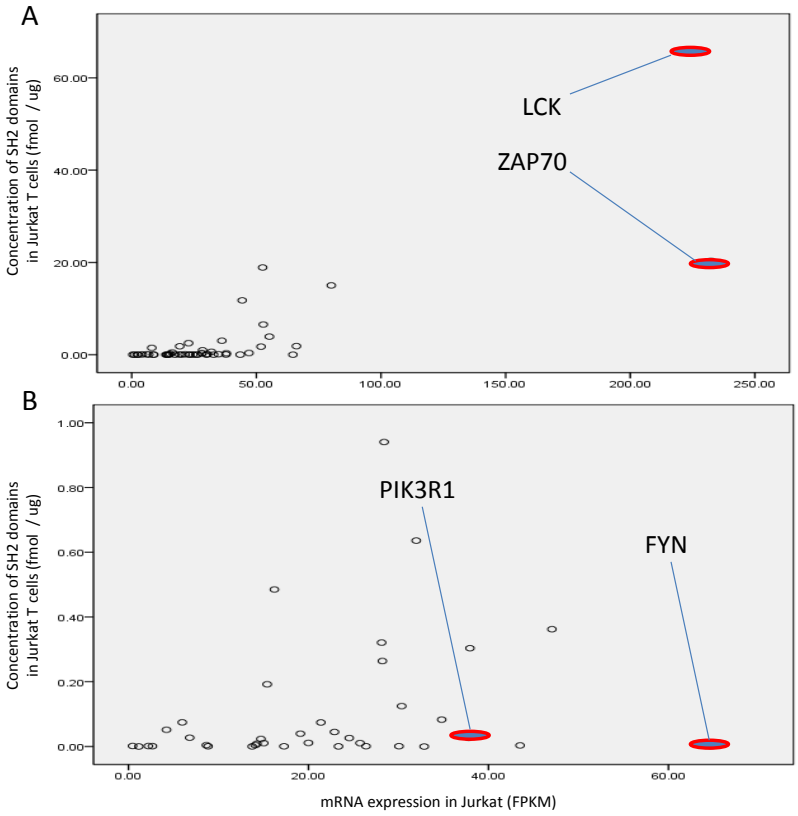


Figure 3.12. Comparison of the SH2 protein concentrations and mRNA transcription in Jurkat T cells. (A) Comparison between all the SH2 proteins. (B) Comparison between SH2 proteins with concentrations below 1 fmol/ug of whole cell proteins.

These differential transcription patterns are somewhat surprising, as one would expect the signalling machinery to be largely conserved between different cell types and signalling is to a large extent determined by the surface receptors. However, it seems that the SH2 protein network inside the cell in effect might have signature patterns so that the different tissues can respond to the extracellular stimuli from the specific environment that they are in, as well as to exert their specific functions. For the same reason, the transcription of intracellular signalling molecules appears to be under dynamic control, as shown by the marked changes in SH2 protein transcription before and after the cells are activated.

Whereas the use and value of transformed human cell lines is debated, comparison of the SH2 protein transcription profile between Jurkat T cells and normal human T cells (clone 29) shows that they are in effect similar to a relatively high degree. In contrast, cell lines with a different origin cell type are significantly different from a T cell. Although there is some uncertainty in this comparison since the transcriptome of clone 29 was determined by SAGE, whilst those for Jurkat and 293T were done by the next-generation sequencing (Solexa) technology, it would appear that the human cell lines have the majority of the signalling characteristics of their parental cells. Combined with several advantages of these cell lines, such as the ease with which they can be grown and manipulated, they can safely be regarded as good experimental tools for signal transduction research.

Compared to the development of transcriptomic technology, major instrumental and technical limitations have constrained the identification and quantification of proteins at the proteomic level in great depth for some time. This situation has however changed significantly over the last few years and absolute quantification of the majority of proteins in a cell is becoming possible. The absolute quantification of Jurkat T cells, as partially done in this study shows that, unlike the relatively comparable expression at mRNA level (within 500-fold difference), SH2 protein expression exhibits a seven-order dynamic range, with the concentrations of most SH2 proteins being very low to extremely low or undetectable. Nonetheless, statistical analysis shows that this proteomic data for Jurkat is overall, in terms of ranking of abundance, positively correlated to the transcriptomic data.

Despite the advances in proteomic technology, protein quantification remains challenging. For example, owing to the limitations in the measurement by the Orbitrap mass spectrometer, it is estimated that concentration values below 0.1 fmol/μg are imprecise, which means only the concentration of 22 SH2

proteins can be confidently estimated in this study. Another issue bearing on the accuracy of the absolute quantification is that the Universal Proteomics Standard and the Jurkat proteome were produced in different laboratories and in separate experiments, and the correlation of the two is achieved through an intermediate set of proteomic data. The absolute concentrations of the SH2 proteins determined in this study need, therefore, to be used cautiously.

Chapter 4. Expression and purification of recombinant SH2 domains[2]

4.1 Introduction

On the surface of T cells, there exist a number of receptors which play crucial roles in, for example, cell proliferation and survival. Signal transduction through these receptors involves phosphorylation of tyrosine-containing motifs in their cytoplasmic regions which then act as docking sites for recruitment of SH2 proteins. In Chapter 3, it was shown that a large number (~50) of SH2 proteins are transcribed in a given mammalian cell (either immune or non-immune). The question therefore arises: how can such large numbers of signalling molecules coordinate specific signalling inside the cell. To perform a systematic study of the interactions between SH2 domains and the tyrosine-containing motifs of some of the most important surface receptors, large quantities of soluble, properly folded recombinant SH2 proteins were required. Although the expression of recombinant proteins has been widely reported, they are often conducted on a small-scale in expression screens only [275, 276], and large-scale, systematic protein expression and purification is rarely performed. In the following sections, the various approaches and strategies for tackling this problem are considered.

4.1.1 The system for expressing recombinant SH2 domains

Expression of a recombinant protein involves cloning of a gene of interest into an expression vector, followed by amplification of the protein in an expression system. There is a wide variety of expression systems available, including

[2] The protein expression and purification work described in this chapter was done under the supervision of Dr. Louise Bird, Oxford Protein Production Facility (OPPF) UK, University of Oxford, and performed at OPPF UK.

bacteria (e.g. *Escherichia coli*; *E. coli*), baculovirus-infected insect cells (e.g. Sf9), yeast (e.g. Pichia pastoris), mammalian cell cultures (e.g. human embryonic kidney cells for transient transfection and Chinese hamster ovary cells for stable transfection), among which *E. coli* is one of the earliest and most widely used systems. *E. coli* offers a number of advantages against other systems, including ease of culture, rapid growth (owing to short doubling time), rapid expression, low costs (inexpensive media and equipment required), and high product yields [277]. As a result, *E. coli* has been playing a crucial role in, e.g., structural biology, which requires large amount of proteins. For example, approximately 90% of the structures in the Protein Data Bank (PDB) are derived from proteins produced in *E. coli* [278]. Whereas one of the key limitations of protein expression by bacteria is that it only facilitates expression of unglycosylated proteins, this is not a problem for SH2 domains as they are carbohydrate-free intracellular molecules. Over the last couple of decades, *E. coli* has been the dominant expression host for generating large quantity of high-quality SH2 domain proteins for structural analysis [83, 178, 279-281]. The successful expression of more than 100 SH2 domains by *E. coli* has been recently reported [191].

Despite a number of advantages, protein expression in *E. coli* is challenged by the formation of inactive/partially active mis-folded aggregates, either soluble or insoluble (i.e. inclusion body). To maximise the formation of soluble and properly folded proteins, several approaches have been used. The most straightforward strategy involves lowering the growth temperature for expression [282]. Other more complicated methods include co-expression of chaperones, which may promote proper folding [283], and recovery of active proteins by refolding [284]. However, a number of chaperone-linked negative side effects have been observed [285], and the refolding approach is limited by being technically demanding and by a low success rate. The most effective way of

improving the solubility can be addition of a fusion protein tag to the target protein. A number of fusion tags have been available for academic use, and a number of comparative studies have been reported to assess the properties of these fusion tags. One of the first large scale studies was conducted by Hammarström et al. [276], in which a test set of 27 target proteins were linked to six different fusion tags, including glutathione-S-transferase (GST; [286]), maltose binding protein (MBP; [287]), thioredoxin (TRX; [288]), Z-domain from protein A [289], N-utilisation substance A (NusA; [290]) and Gb1-domain from protein G (Gb1; [291]). This study showed that solubility of the target proteins was increased by expressing them with a fusion partner, and the most significant improvement was seen using thioredoxin, followed by MBP. Owing to the effectiveness of this approach, the repertoire of fusion tags has been expanded with regular addition of new members, including SUMO [292], MsyB [293, 294], Trigger factor [295, 296] and HaloTag7 [297].

Although some fusion tags seem to generally function better in improving solubility, there is no "universal" fusion partner and individual differences occur. This means screening of multiple constructs with different fusion tags is required to identify soluble versions of the target protein, which drives emergence of highly efficient cloning and screening methods. Two options are available for making the constructs for protein expression: the ligation-dependent and ligation-independent (only the In-fusion^{TM} technology is discussed here) cloning methods. The former involves digestion using restriction enzymes in combination with ligation using DNA ligase, while the latter involves the so-called cre-lox recombination (Clontech) between the overhangs in a PCR product and the complementary sequences in the recipient vector [298]. Ligation-independent cloning (LIC) technology has several advantages when compared to conventional ligation-dependent cloning (LDC) [298]: (1) it is sequence-independent, so there is no such problem that restriction sites could be present in the insert; (2) there

is no addition of amino acids to the expressed protein to meet the restriction site requirement; (3) only one-step processing of the PCR products is required, as no digestion of the PCR products is required; (4) it is capable of cloning a gene into multiple vectors with different tags; (5) it is amenable to a high-throughput setting. In combination with a high-throughput small-scale expression-screening assay, this would allow rapid assessment of the optimal construct.

4.1.2 Overview of the expression strategies

The sequence information for the SH2 domains was obtained from Uniprot [299]. Since disordered regions of proteins are known to adversely affect their solubility [300], the SH2 domain constructs were designed to have no disordered regions at either terminus. As there is currently no single disorder prediction technique that is entirely accurate, the design of the SH2 constructs was assisted by using two different methods for disorder detection - the PONDR® (Predictor of Naturally Disordered Regions), which is the best-known, and most intensively developed, disorder prediction tool, as well as RONN (regional order neural network), which is a recently developed programme shown to have the best performance compared to a number of widely used programmes [300]. The sequences of the SH2 constructs made in this study are shown in **Appendix 2**.

In the light of the large set of proteins to be expressed, the high-throughput ligation independent cloning strategy developed by Berrow et al. [298] was used for construction of expression vectors in this study. A range of fusion vector backbones, including pOPINF (His-tag), pOPINS3C (SUMO-tag), pOPINTRX (TRX-tag), pOPINMSYB (MsyB-tag), pOPINJ (GST-tag), pOPINHALO7 (HaloTag7), pOPINM (MBP-tag), pOPINTF (TF-tag) and pOPINNUSA (NusA-tag), were supplied by the Oxford Protein Production Facility (OPPF). The properties of these vectors are summarised in (**Table 4.1**). All fusion tags were combined with a N-terminal

Table 4.1. The properties of the vectors are shown in the table. The single letter codes designate different amino acids. The italics designate the fusion tag. ⇓ indicates the position of cleavage in the Rhinovirus 3C protease site (LEVLQFGP). POI indicates the protein of interest.

Vector	Fusion Protein	MW of tag/Da	The number of bases
pOPINF	MAHis$_6$-SSGLEVLQF⇓GP-POI	2158	57
pOPINS3C	MAHis$_6$-***SUMO***-SSGLEVLQF⇓GP-POI	13,213	345
pOPINTRX	MAHis$_6$-***thioredoxin***-SSGLEVLQF⇓GP-POI	13,815	381
pOPINMSYB	MAHis$_6$-**MsyB**-SSGLEVLQF⇓GP-POI	16,268	426
pOPINJ	MAHis$_6$-***glutathione-S-transferase***-SSGLEVLQF⇓GP-POI	27,954	720
pOPINHALO7	MAHis$_6$-***HaloTag7***-SSGLEVLQF⇓GP-POI	35,362	939
pOPINTF	MAHis$_6$-***Trigger Factor***-SSGLEVLQF⇓GP-POI	50,201	1350
pOPINNUSA	MAHis$_6$-***N-utilisation substance A***-SSGLEVLQF⇓GP-POI	56,910	1539

hexa-histidine (His_6) sequence as an affinity tag. Because of its simplicity in purification and the fact that the tag itself does not normally pose a problem in crystallization trials, addition of a His-tag is by far the most popular method for purification of recombinant proteins. In fact, the His-tag has been used in about 60% crystallographic studies, many of which left the His-tag intact [301]. Purification is based on the use of chelated metal ions (for example, nickel/Ni) as affinity ligands, and the metal ion is complexed with an immobilised chelating agent (for example, nitrilotriacetic acid). By passing the cell lysate containing the His-tagged protein through a nickel-nitrilotriacetic acid (Ni-NTA) matrix (or mixing the cell lysate with the nickel beads), the histidine residues will be bound to the

Ni-NTA matrix (or nickel beads) via an electrostatic coordination interaction between the nickel ion and the histidine ring. In order to increase the avidity between the recombinant protein and the nickel, a number of repeating histidine residues (mostly six repeats) are used. Following the initial binding event, the beads are usually washed with a buffer solution containing low concentrations of imidazole, which competes for binding of the nickel with non-specifically bound materials. Thereafter, the His-tagged proteins can be eluted with higher concentrations of imidazole [302].

For all other pOPIN vectors, His_6 tags were incorporated at the N-terminus to allow a standardised assay and convenience of purification. In addition, a Rhinovirus 3C protease site was included between the fusion protein and the gene insert to allow removal of the fusion tags. There are two main reasons why it is important to cleave the fusion partner and obtain tag-free SH2 proteins: firstly, this ensures that the SH2 domain retains its native structure (and activity) once released from the fusion tag. Secondly, cleavage eliminates the possibly non-specific interaction between the fusion tag and the SH2 binding target, ensuring that the K_D values obtained from the future binding experiments are most relevant.

The PCR product of selected target protein sequence was generated by a forward primer including an extension of 20bp region (AAGTTCTGTTTCAGGGCCCG) homologous to the 5' 3C protease site and a reverse primer including an extension of 18bp (ATGGTCTAGAAAGCTTTA) region homologous to the 3' infusion region and a stop codon. The DNA template was removed by addition of DpnI restriction enzyme to the PCR reaction and incubation at 37^{o}C for 1 hour, followed by purification using AMPure magnetic beads (Beckman-Coulter) according to the manufacturer's protocols.

The PCR products were then infused into the fusion vectors by the In-fusion^{TM} enzyme, in which process the KpnI restriction site on the vector will be

destroyed and a full 3C protease site is constituted immediately 5′ of the gene of interest (**Figure 4.1**). The In-fusion plasmid products were transformed into One Shot® OmniMAX™ competent cells (Invitrogen) for DNA amplification. The resulting plasmid DNA was firstly subjected to PCR verification (for the successful insertion of gene of interest) and subsequently sequencing verification. A representative verification gel is shown in **Figure 4.2**. The verified plasmid DNA was then transformed into two *E. coli* expression strains [B834 (DE3) or Rosetta (DE3) LysS *E. coli* competent cells (Novagen)] for small-scale expression screening using two expression induction regimes – LB with IPTG or TB overnight expression media.

The level of soluble protein expression was scored by visual inspection of SDS-PAGE gels of purified proteins. The scale utilised (0, 1, 2 and 3) corresponds to the strength of the bands on the SDS-PAGE gels: "0" = absence of a band; "1" = presence of a faint band; "2" = presence of a moderately strong band; "3" = presence of a strong band. Scale-up experiments are considered on the basis of the condition that gives rise to the strongest band. To allow consistent evaluation of the scales, expression screening was done under a standard protocol and the same amount of sample was loaded. Representative SDS-PAGE gels are shown in **Figure 4.3**.

Constructs showing positive results in the expression screening, were scaled up in 1-litre cultures and purified through the ÄKTA™ Express protein-purification system (GE Healthcare). For proteins expressed with a fusion partner, the tags were cleaved by incubation with 3C protease, followed by a secondary purification to obtain a tag-free version of protein. If the screened constructs gave rise to soluble proteins and eventually a tag-free soluble fusion protein, which can be due to formation of soluble aggregates or precipitation upon cleavage of the fusion tag, the construct was expressed using a different expression strain.

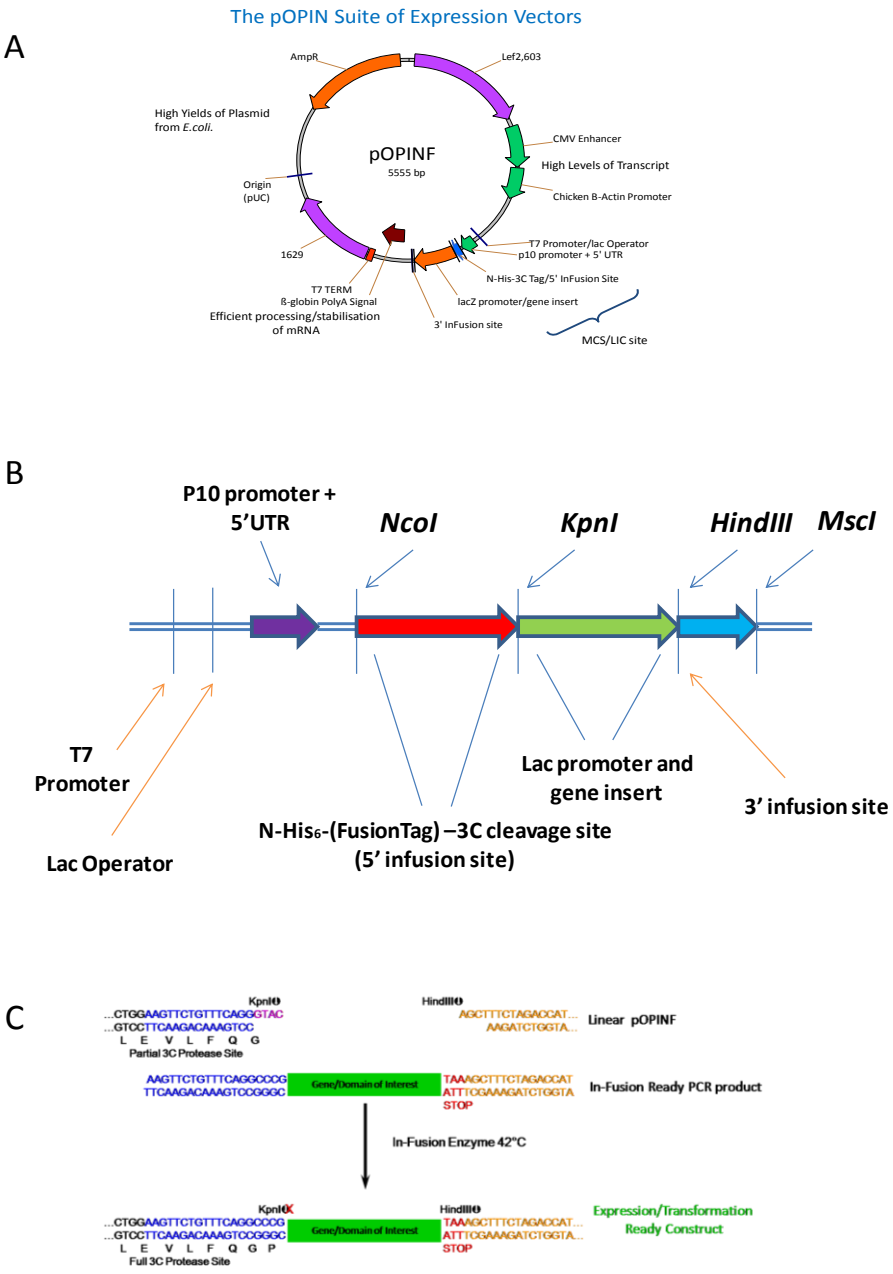


Figure 4.1. The cloning vector and method of cloning. (A) The pOPINF vector used in this study for transformation of *E. coli* cells [298]. Other pOPIN family vectors are derived from this. (B) The composition of the multiple cloning site (MCS) in detail. (C) The mechanism of the In-fusion reaction (adapted from L. Bird, UK Oxford Protein Production Facility). A PCR product for a protein of interest (POI, green) with the 5'extension (blue) and 3' (orange) extension was Infused into KpnI HinDIII restricted pOPINF vector. The Infusion™ process destroys the KpnI site and reconstitutes the Rhinovirus 3C protease site, and a 3' stop codon is added as part of the 3'Infusion extension. The same PCR product was also cloned into other members of the pOPINF family to make N-terminal hexa-histidine (pale blue) tagged fusion protein (dark blue) with the POI (green).

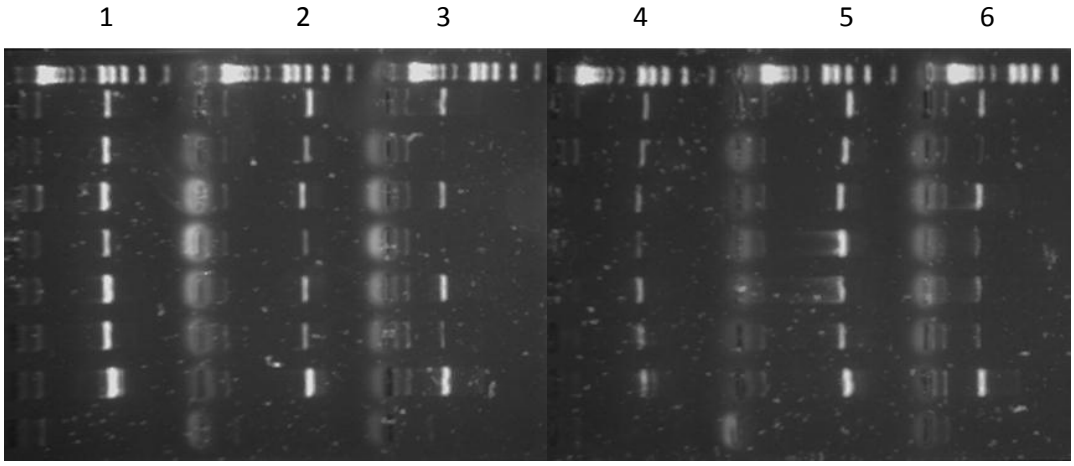


Figure 4.2. Representative PCR verification gel for seven of the targets in the fusion vectors. The image shows the verification for, from column 1 to column 3, pOPINTRX, S3C, and TF fusion proteins. Two clones were picked; clone 1 is in the left hand three columns and clone 2 in the right hand three columns.

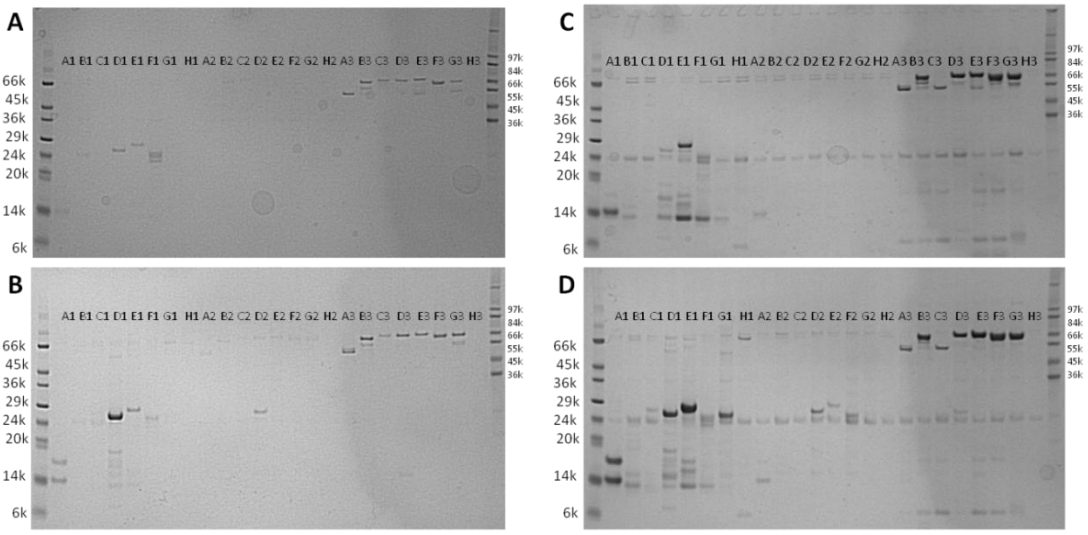


Figure 4.3. Representative expression screen results. Figures A-D show the expression screening results using different combinations of expression strains and growth media. (A) Rossetta cells in LB with IPTG induction; (B) Rossetta cells in TB Overnight Express medium; (C) B834 cells in LB with IPTG induction; (D) B834 cells in TB Overnight Express media. Lanes A-H indicates SHC1, SLAP, CBLB, not give SH2B, SPT6, RIN3, LNK and CHN1. A1-H1 are expressed with the Thioredoxin (TRX) fusion tag; A2-H2 are expressed with the SUMO (S3C) fusion tag; A3-H3 are expressed with the Trigger factor (TF) fusion tag.

The first choice strain in this study was Rosetta LysS, which is designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* and selected by two antibiotics: carbenicillin, and chloramphenicol. An alternative option was the B834 strain, which is a methionine auxotrophs that allows high specific-activity labelling of target proteins with ^{35}S-methionine and selenomethionine for crystallography, and selected only by carbenicillin. Alternatively, the gene of interest was cloned into new expression vectors with other fusion tags for another round of expression trials. In addition, the refolding approach or modification of the construct, involving changes of amino acid or extension of the N- and C-terminal regions, was attempted.

4.2 Results

In this study, Jurkat T cells have been employed as the target model for a system-level investigation of signal transduction principally owing to its ease of manipulation in downstream hypothesis testing experiments. Whilst the deep transcriptome of Jurkat T cell reveals that 85 SH2 proteins are expressed, only 67 of these have abundance above 2 FPKM (fragments per kilobase of exon per million fragments mapped). It was considered that the very low expression of the remaining 18 proteins didn't warrant their production (although the list is presently being extended). In the following sections, the use of a number of approaches, including high-through-put cloning of the genes encoding the SH2 domains into a variety of vectors for expression with different solubilising fusion tags and for refolding, is described.

4.2.1 Expression with a His-tag

There is a total of 87 SH2 domains derived from 67 SH2 domain-containing proteins (10 of them contain two SH2 domains, which will be individually expressed in addition to the double domain) to be expressed in this study. All SH2 proteins were firstly cloned into pOINF, i.e. to be expressed as His-tagged proteins. Using this approach, 50 of the 87 target SH2 domains were successfully expressed and purified when scaled up to 1-liter cultures. Our results showed that scores of 2 and 3 at the screening stage are reliable predictors of successful expression, whereas a score of 1, to a large extent, indicates that it is unlikely to be possible to purify the protein on scale up. However, exceptions were seen for PIK3R3_C and CHAT which were successfully purified to a yield of about 7.5mg on a 1-litre scale up, although they were found to only weakly express in the expression screen. For those successfully purified as a His-tagged protein, over 95% of them achieved high expression levels (> 5mg/L). A typical purification result is shown in **Figure 4.4**. In the ÄKTA^{TM} Express protein-purification system, individual SH2 domains are usually eluted as a distinct elution peak at about 80-85 ml (through a Sephadex S75 column). Occasionally, a shallow elution peak indicating a small fraction of oligomer or aggregates can be observed between 60-70 ml.

Combined with a bioinformatics analysis of family relationships, our results suggest that SH2 domain solubility seems to be largely protein dependent as evidenced by the observation that SH2 domains belonging to certain families are not soluble as His-tagged proteins. According to the sequence alignment, over 60% (23 out of 37) insoluble proteins can be attributed to the following families: (1) CBL and CBLB; (2) LNK, SH2B and SH2B2; (3) CHN1 and CHN2; (4) JAK1, JAK2, JAK3 and TYK2; (5) STAT1, STAT2, STAT3, STAT5A, STAT5B and STAT6; (6) SOCS1, SOCS4, SOCS5, SOCS6, SOCS7, CISH. It is likely that the SH2 domains belonging to these families may have certain structural patterns that

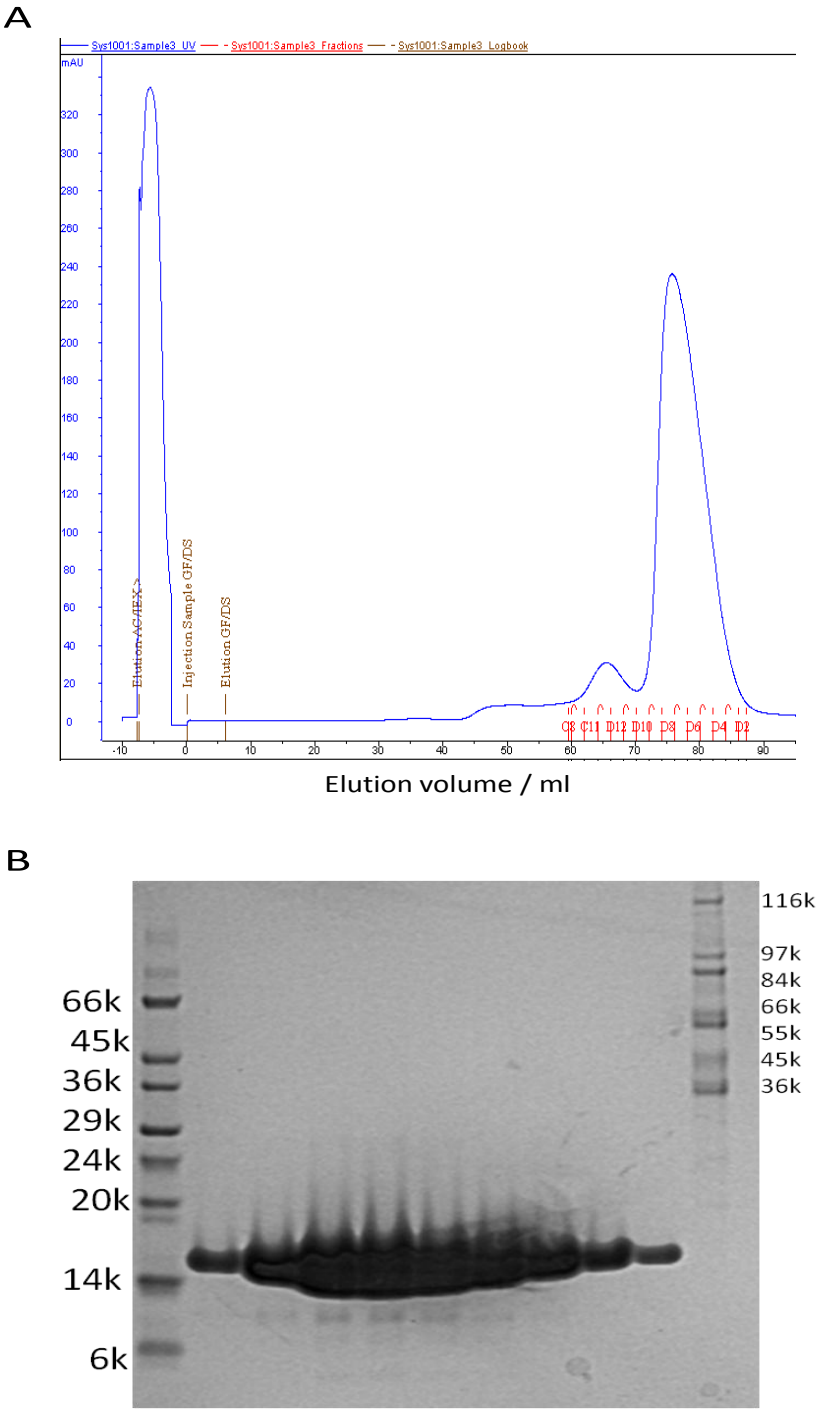


Figure 4.4. Representative results for expression and purification of SH2 domains from a His-tagged protein. (A) The HPLC profile for purification of LCK(-His) on a S75 column. (B) The SDS-PAGE gel showing the corresponding fractions (D10-D2) collected following HPLC purification.

prevent them from expressing. Interestingly, while the N+C double domains and the C-terminal SH2 domains of SHP1, GAP, SYK and PLCG2 were well expressed as His-tagged proteins, the N-terminal SH2 domains were not soluble. Whilst PIK3R1 and PIK3R3 were soluble as single or double domains, the solubility of single N-terminal or C-terminal SH2 domain of PIK3R2 was poor. SH2 domains that failed to express as soluble proteins were cloned into other pOPIN vectors for expression with different solubility-enhancing fusion partners. A summary of the expression screen for the insoluble SH2 domains expressed with different fusion tags is shown in **Table 4.2**. Because of the differential availability of the different expression vectors and limitations of time, not all target proteins were cloned into all the vector types.

4.2.2 Expression with GST and MBP fusion tags

pOPINF, pOPINJ and pOPINM vectors are the three first-generation vectors of the pOPINF family generated by OPPF. In the first round of re-expression, the SH2 domains were cloned into pOPINJ and pOPINM (thus expressed as GST- and MBP-tagged proteins), respectively. GST has been used as a soluble partner for over 10 years, and has permitted the expression of a number of insoluble proteins [303]. It has been shown that GST is able to protect against intracellular protease cleavage and stabilise the recombinant protein [304]. MBP is one of the most well-studied solubility factors and has been shown to be one of the most efficient solubilising partners [305-308]. It has been suggested that MBP may function as a molecular chaperone and interact with its fusion partner via the maltose-binding site. This interaction may prevent the aggregation of folding intermediates in which the tag domain is properly folded but the passenger domain is not) and eventually steers them towards the native conformation [307].

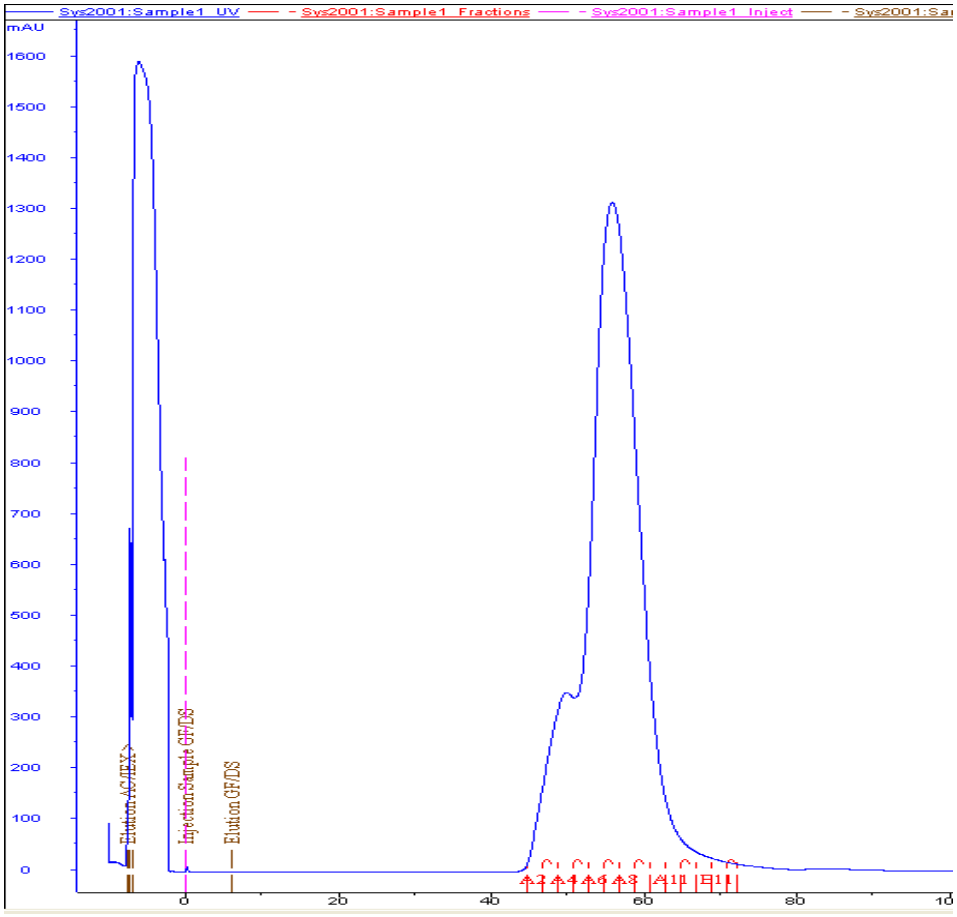
Table 4.2. Small-scale expression-screen results are scored 0–3. "0" = absence of a band; "1" = presence of a faint band; "2" = presence of a moderately strong band; "3" = presence of a strong band. The score in the table is the highest level of expression observed in two induction conditions. Soluble expression scored 1 is coloured yellow, scores of soluble expression 2 and 3 are coloured green, and no expression (0) is coloured red. N/D indicates that the construct was not made.

	Vector name (Fusion Partner)								
	F	J	M	TF	TRX	MSYB	NusA	3C	HALO7
Target protein	(N-His)	(N-His-GST)	(N-His-MBP)	(N-His-TF)	(N-His-TRX)	(N-His-MSYB)	(N-His-NUSA)	(N-His-SUMO)	(N-His-HaloTag7)
PLCG2_N	0	0	N/D	3	1	2	N/D	N/D	3
PLCG2_N+C	1	2	N/D	3	2	3	N/D	N/D	3
SLAP	0	0	2	2	0	2	N/D	0	2
CHN2	1	1	3	1	0	2	N/D	0	3
RIN3	1	0	1	3	1	2	N/D	0	2
SH2D3A	0	1	3	1	1	2	N/D	0	3
LNK	0	0	1	1	0	1	N/D	0	N/D
SHP1_N	0	1	2	2	1	N/D	N/D	0	N/D
SH2B	0	1	3	2	1	N/D	N/D	0	N/D
GAP_N	1	2	3	2	0	N/D	N/D	0	N/D
SPT6	1	3	2	2	2	N/D	N/D	0	N/D
PIK3R2_C	1	0	2	0	2	N/D	N/D	0	N/D
CBL	0	1	1	2	0	N/D	N/D	0	N/D
CBLB	0	N/D	N/D	1	0	N/D	N/D	0	N/D
CHN1	0	N/D	1	0	0	N/D	N/D	0	N/D
SHC1	0	0	0	0	0	N/D	1	0	N/D
PIK3R2_N	1	3	2	N/D	N/D	N/D	N/D	N/D	N/D
SH2B2	2	3	N/D	N/D	N/D	N/D	N/D	N/D	N/D
LCP2	1	3	N/D	N/D	N/D	N/D	N/D	N/D	N/D
SYK_N	1	3	N/D	N/D	N/D	N/D	N/D	N/D	N/D
FER	0	1	N/D	N/D	N/D	N/D	N/D	N/D	N/D
JAK1	1	1	3	3	N/D	1	1	N/D	N/D
JAK2	1	1	3	N/D	N/D	N/D	N/D	N/D	N/D
JAK3	1	1	3	3	N/D	1	1	N/D	N/D
TYK2	1	1	3	N/D	N/D	N/D	N/D	N/D	N/D
STAT1	1	1	2	N/D	N/D	N/D	N/D	N/D	N/D
STAT2	1	1	N/D	N/D	N/D	N/D	N/D	N/D	N/D
STAT3	1	1	3	N/D	N/D	N/D	N/D	N/D	N/D
STAT5A	0	3	3	N/D	N/D	N/D	N/D	N/D	N/D
STAT5B	1	1	N/D	N/D	N/D	N/D	N/D	N/D	N/D
STAT6	0	0	3	N/D	N/D	N/D	N/D	N/D	N/D
SOCS1	1	0	N/D	2	N/D	1	1	N/D	N/D
SOCS4	1	1	N/D	1	N/D	1	1	N/D	N/D
SOCS5	0	1	N/D	1	N/D	1	1	N/D	N/D
SOCS6	0	2	N/D	2	N/D	3	1	N/D	N/D

Our expression screening assay showed that the solubilising capability of the GST-tag is low: only 9/33 proteins showed soluble expression in the screen with a score of at least 2 justifying scale up. Among these, only 4, i.e. SH2B2, PIK3R2_N, LCP2, and SYK_N, were successfully purified as tag-free SH2 domains upon scale up and cleavage of the fusion partner. A representative purification result is shown in **Figure 4.5**. The observation of a small fraction of GST proteins found in the HPLC purification indicates that passing the 3C-protease-treated protein mix through the HisTrap FF column was not sufficient to remove all the tagged proteins, which means a secondary HPLC purification was required to ensure pure SH2 proteins were obtained. Whereas a number of other constructs with a score of at least 2 were able to be expressed as a fusion protein, they however failed to cleave and release tag-free SH2 (data not shown). Although a significant amount of fusion proteins were expressed upon scale up, the yield of tag-free SH2 proteins was low: only 1-3mg of protein was purified after cleavage from a 1-litre culture. This can be due to a number of reasons: (1) precipitation occurred upon addition of the 3C protease, indicating that a fraction of the SH2 proteins are not properly folded, and they are only held in solution by the fusion partner; (2) protein loss occurred over the process of re-purification through the HisTrap FF column and HPLC.

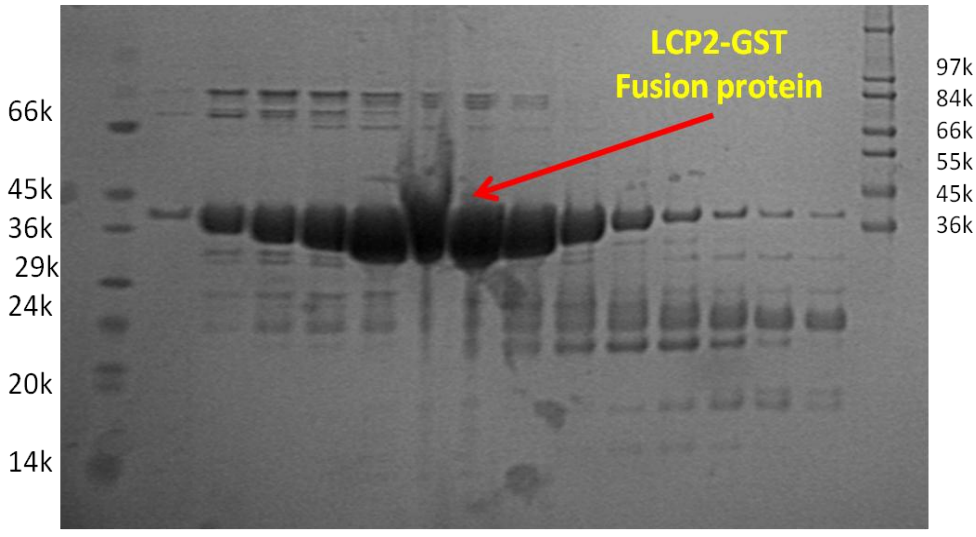
Striking results were seen for the expression screen with the MBP fusion tag: 80% (17/22) of the target proteins were shown to have a score of at least 2. However, when four of the targets – STAT1, SLAP, CHN2 and SH2D3A – were scaled up to 1-litre cultures, no purified protein was obtainable after cleavage. A remarkable feature of expression of the SH2/MBP fusion proteins is that significant proteolysis was observed when compared to expression of SH2/GST fusion proteins. Although a small quantity of cleaved SH2D3A SH2 domain was observed, it co-eluted with the MBP-tag after HPLC purification (**Figure 4.6)** It is possible that MBP might not be a real solubilising factor that can render the SH2

A

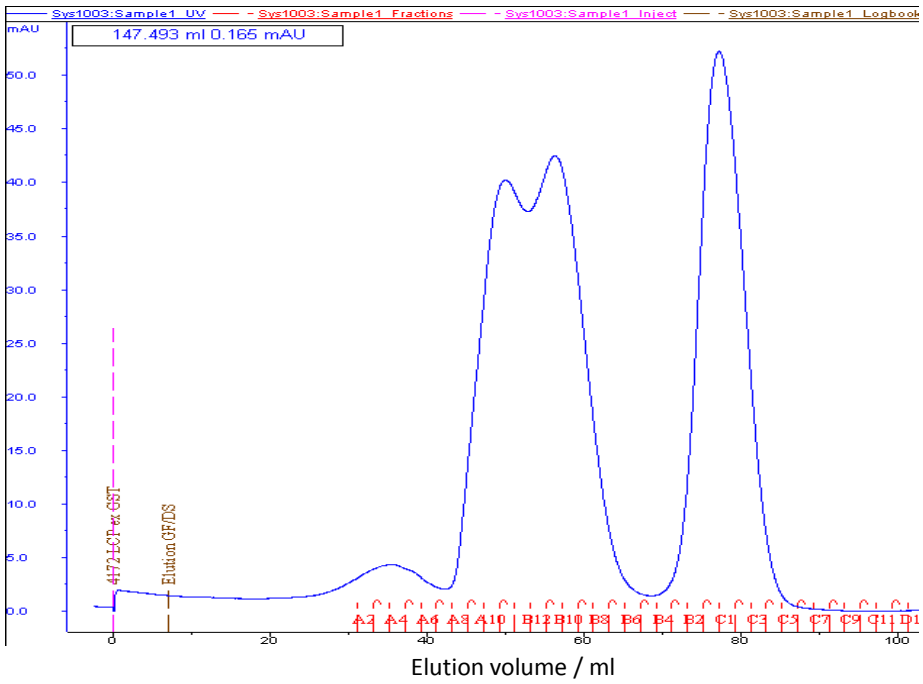


Elution volume / ml

B



A



B

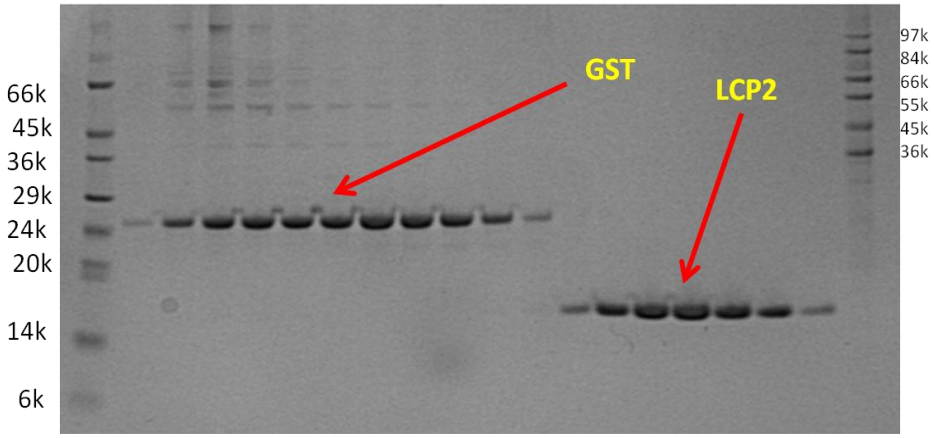


Figure 4.5. Representative results for expression and purification of SH2 domains from a fusion protein. (A) The HPLC profile for purification of LCP2-GST fusion protein (on a S75 column). (B) The SDS-PAGE gel showing the corresponding fractions (A2-B10) collected following HPLC purification. The expression of fusion protein is indicated by bands with a molecular size between 36kd and 45kd. (C) The HPLC profile for purification of LCP2 protein (on a S75 column) after cleavage of the fusion protein by 3C protease. (D) The SDS-PAGE gel showing the corresponding fractions collected from the HPLC purification. Fractions A8-B7 (lane 1-11 of the gel) indicate the GST fusion tag; fractions B3-C4 (lane 12-18 of the gel) indicate the LCP2 SH2 domain protein.

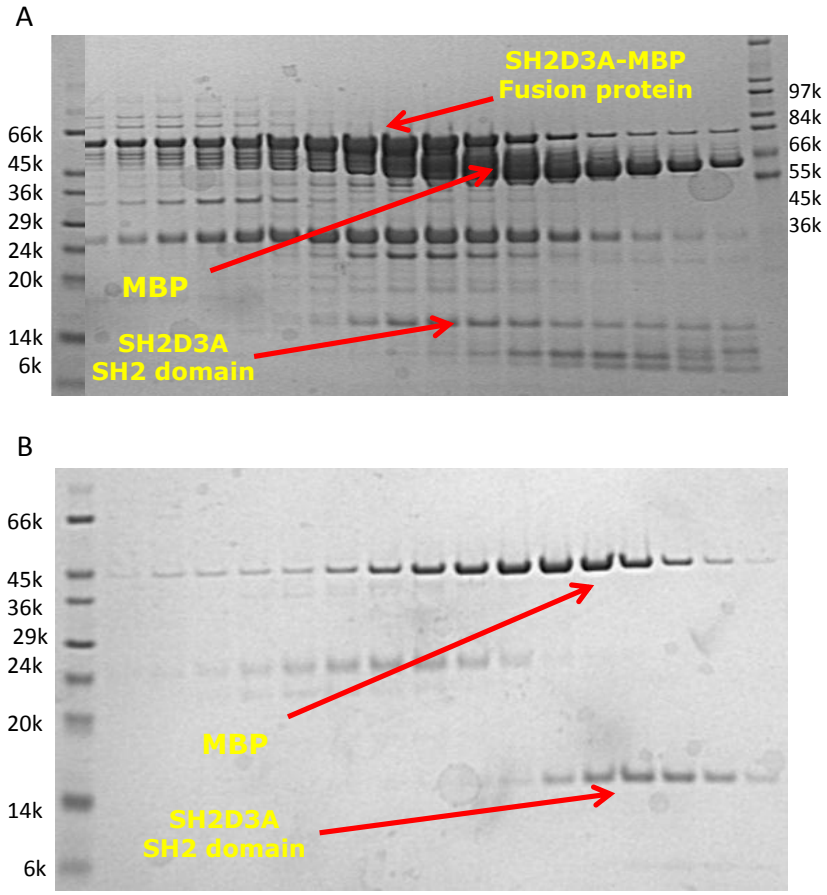


Figure 4.6. Purification of the SH2D3A SH2 domain using an MBP fusion tag. (A) SDS-PAGE gel showing the SH2D3A-MBP fusion protein following HPLC purification. High degrees of proteolysis were observed; the MBP protein and SH2D3A SH2 domain co-eluted together with the SH2D3A-MBP fusion protein. (B) SDS-PAGE gel showing the purified products following cleavage of the SH2D3A-MBP fusion protein. Elution of the SH2D3A SH2 domain overlapped with that of the MBP fusion tag, but were separated on the basis of size.

domain in a native and soluble form. Instead, it may hold the target proteins that are not properly folded in solution via direct interactions. Since the MBP-tagged proteins were not promising with regard to purification of tag-free proteins, even though they may have high scores in the expression screen, it was decided that no further attempts would be made for other constructs fused with the MBP tag.

4.2.3 Expression with Trx/SUMO/TF/MsyB/NusA fusion tag

In the light of the unsatisfactory performance of GST and MBP tags in large-scale purification, an additional six vectors were subsequently made, including Trx, SUMO, TF, MsyB, NusA and HaloTag7. Whereas Trx and NusA have been well studied and shown to be an excellent tool to enhance the solubility of recombinant proteins [305, 308, 309], SUMO, MsyB, TF and HaloTag7 have only been recently developed. How these fusion tags function to increase solubility of the target protein is still not clear, although it has been reported that TF (Trigger Factor) is one of the three chaperone systems that coordinates the folding of cytosolic proteins by maintaining the newly translated polypeptides in a folding-competent state in eubacteria [296]. HaloTag7 is the latest fusion tag, developed by Promega. The halo protein is a modified enzyme designed to bind rapidly and covalently with different ligands, which can, for example, be used to facilitate cellular imaging or purification as an affinity tag. Halotag7 is a specially engineered version of the Halo protein for enhancement of soluble expression particularly in *E. coli*. It was shown that Halotag7 has better performance in mediating soluble expression of the tagged proteins compared to MBP and GST [297].

Among the new fusion tags, TF was found to have excellent performance in expression screen, with 60% of the fusion proteins gaining a score of at least 2 in the expression screen. However, only four of them were successfully cleaved to allow purification of tag-free SH2 proteins. Intriguingly, whereas SHP1_N was not cleavable from the TF when expressed in Rosetta cells, success was achieved when using B834 as the expression strain. Although high levels of expression have been observed for a number of fusion proteins, no tag-free protein was successfully obtained after cleavage. In the case of SLAP-TF purification, even though some material separated from the tag, it was however heavily contaminated by E. coli proteins (**Figure 4.7**).

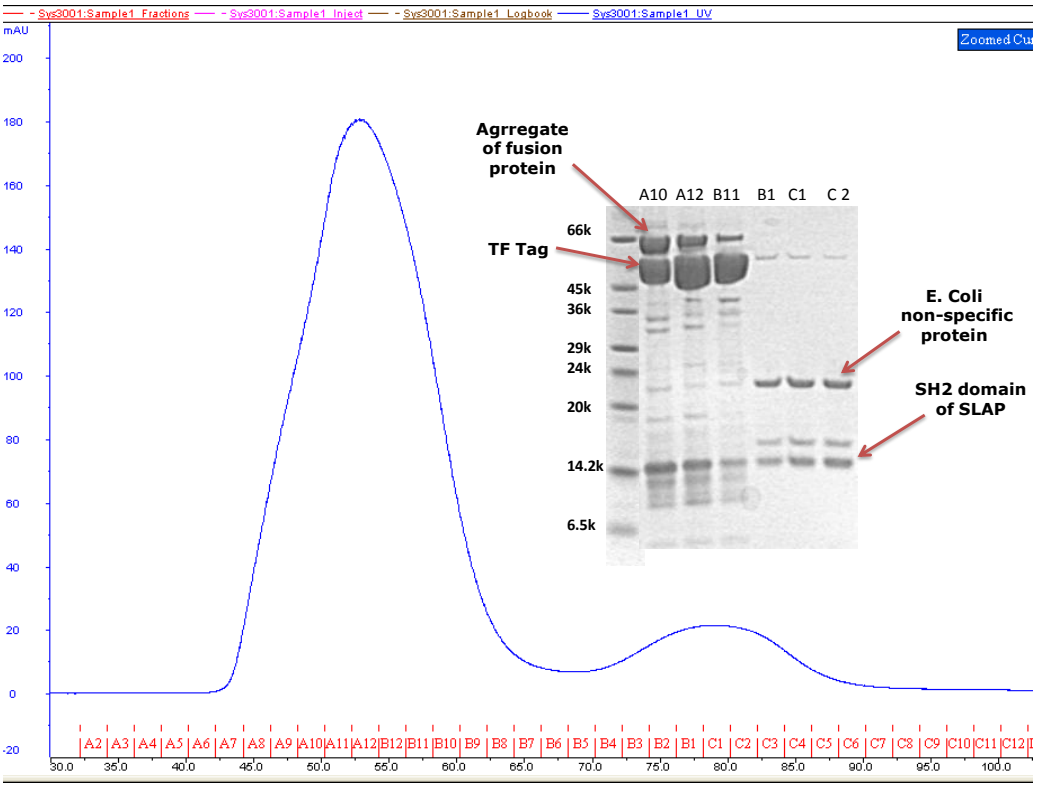


Figure 4.7. The cleaved products of SLAP-TF after purification by size-exclusion chromatography. The fractions corresponding to the bands are labelled on the top of the lanes. The SH2 domain of SLAP either co-eluted with the aggregate proteins as well as the cleaved tag, or was heavily contaminated with *E. coli* proteins.

Consistent with the results reported by Zou et al. [294], MsyB appeared to be a very good fusion partner for enhancing solubility of the target protein, as evidenced by the observation that over half of the fusion proteins gave promising expression, which was superior to other fusion tags like thioredoxin, SUMO and NuSA. However, since the molecular weights of the MsyB fusion protein and the SH2 proteins are too close, it was not possible to separate the two proteins by HPLC after cleavage of the fusion protein. This means it is difficult to evaluate whether MsyB is a genuinely good fusion partner that can mediate proper folding of the target protein or, like MBP, it is particularly good at holding mis-folded protein in solution. In contrast to the high performance of thioredoxin as a

solubilising fusion partner reported by Hammarström et ac. [276], only 3 out of the 16 fusion proteins were found to have a score of 2 in the expression was used as the fusion tag.

Although SUMO and NusA have previously been reported to be good solubility-enhancers [292, 294, 310], they however performed poorly in our expression-screen analysis: all the NusA-tagged proteins were only found to have an expression score of 1, and were therefore expected not to give rise to production of purified proteins on scale up, and the SUMO-tagged protein gave no soluble protein at all. The result obtained for the NusA tag is perhaps because the target proteins they fused with were mainly the Jak and SOCS family members, which seem very difficult to solubilise [191]; also, SHC1 was shown to be insoluble in all other fusion tags.

One promising fusion tag based on our study is Halotag7, which had a 100% success rate in giving fusion proteins scores >2 in the expression screen. Although the sample size is small, with only six constructs, our results suggest that it may be an excellent solubiliser of expression, since whereas CHN2 and SH2D3A are not soluble with TF, a fusion tag with very good performance, they were highly expressed with Halotag7. The success of purifying PLCG2_N as a tag-free SH2 protein from the Halotag7 fusion protein suggests that this fusion tag can potentially mediate proper folding of the target protein rather than simply making it soluble in the presence of the tag, which probably happens in the case of using MBP as a fusion partner.

4.2.4 Redesign of difficult constructs

4.2.4.1 Expression of the STAT family members

Another way to obtain purified proteins was to express the SH2 domain as an extended construct which includes other regions surrounding the SH2 domain. For

example, it was found by others that the 'core region' of STAT1, which contains the coiled coil domain, DNA binding domain, linker domain and SH2 domain, can be successfully expressed as a His-tagged protein and crystallised [311]. Since the STAT family members (including STAT1 - 6) are closely related [169], the core regions of STAT family members were determined by sequence alignment and expressed as His-tagged proteins. Using this strategy, all STATs were successfully expressed and purified, although the products of the first purification were heavily contaminated, presumably due to partial cleavage of the recombinant protein or non-specific interaction with *E. coli* proteins. However, after two further steps of washing and HPLC purifications, clean products were obtained (**Figure 4.8**), although the yields were very low - about 0.5 mg protein purified from one litre of bacterial culture).

4.2.4.2 Expression of CBL and CBLB

The SH2 domains of CBL and CBLB were cloned as an extended construct according to the crystal structure available for CBL and its alignment with CBLB [312], allowing these two proteins to be successfully expressed and purified. Interestingly, unlike the STAT family, they were less contaminated and only one extra round of His-tag extraction and size-exclusion chromatography purification was required to obtain pure products.

While there was no promising solution for expressing SHC1, which is known to be important in T cell signalling [244], the expression of a murine homologue was attempted, since the difference between the murine and human version is only 2 amino acids (E>D and L>V) (**Figure 4.9**). The change of these amino acids should not affect the binding affinity of the domain as they are not part of the binding pocket (for the ζ chain of the T cell receptor) according to structural data [313]. Interestingly, the murine version of SHC1 was successfully expressed as a His-tagged protein with high yield.

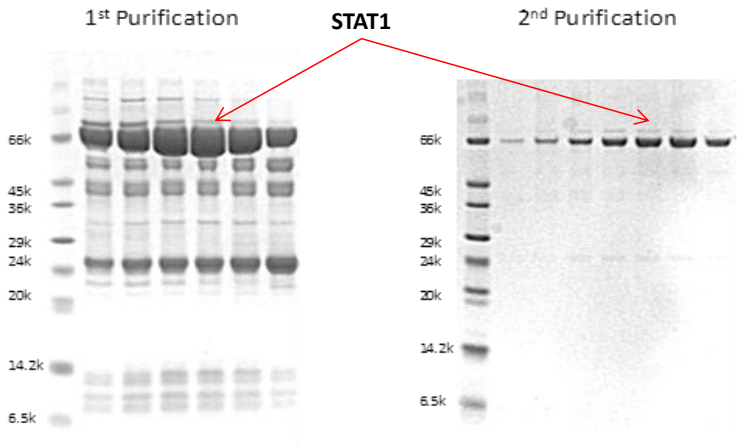


Figure 4.8. Purification of STAT1. Each lane corresponds to a fraction collected from the HPLC purification. The product after 1st purification was heavily contaminated. However, a reasonably clean product was obtained after a second round of purification.

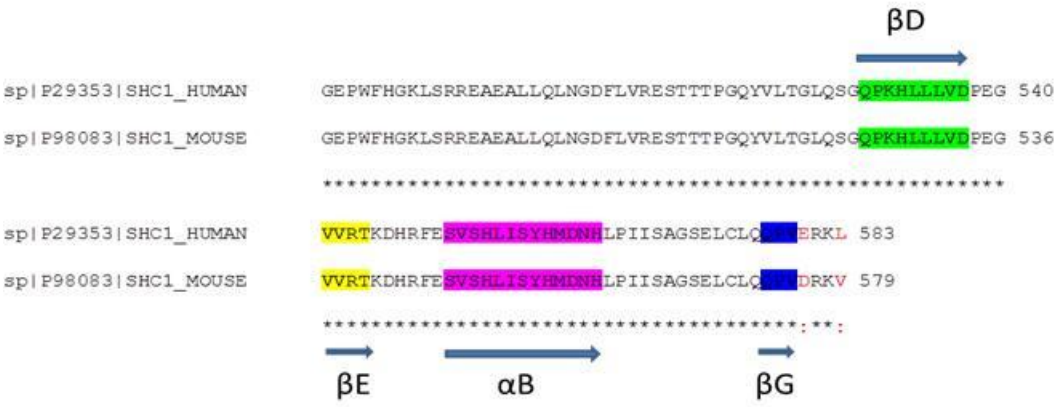


Figure 4.9. Sequence alignment between the human and mouse SHC1 protein. Letters in red mark discrepancies in amino acids. Coloured sequences are structures important in ligand binding.

4.2.5 Protein refolding

A refolding strategy has been applied to make soluble proteins from inclusion bodies (insoluble aggregates). There are two ways available for refolding – the

traditional 'solution refolding' and the recently developed 'high pressure' refolding method [314]. Since the latter strategy requires special equipment, solution refolding was used in this study. In this method, inclusion bodies are prepared and dissolved in a strong chemical chaotrope/denaturant (6M GdHCl-Tris buffer). Thereafter, refolding is achieved by removing the chaotrope via buffer exchange. In the light of the fact that there is no 'universal' refolding buffer, i.e. the composition of the necessary refolding buffer is strongly protein dependent, a refolding screening consisting of many different conditions is required [220]. In order to test several components and compositions simultaneously and generate an optimal set of experiments wherein the opportunity for exploring variables is maximised, a fractional factorial approach, using software called SAmBA, was applied to generate a screen with 96 different buffer conditions in microtiter plate format (**Appendix 1**) [315].

The inclusion bodies from a number of His-tagged SH2 proteins were prepared, including JAK2, JAK3, SOCS3, SOCS6, CHN1, CHN2, SH2D3A and SLAP. The inclusion body solution was mixed with the screening buffer in the microtiter plate on a shaker and the result was revealed by the turbidity of the solution, which was measured in terms of light absorbance at 350 nm. Conditions that were potentially suitable for refolding were characterised by giving an $OD<0.05$ [221]. Although RIN3 and LNK were not expressed as soluble proteins, no inclusion body was successfully obtained either. Although most of the refolding using buffers containing glycerol, Brij58, L-arginine or glucose were found to be promising, they did not give rise to isolated soluble protein in practice (data not shown). This indicated that soluble aggregates were formed in these buffers. In consequence, conditions containing these components were excluded from the set of candidate buffer conditions for refolding.

Among the 8 targets chosen for the refolding trial, only CHN1 and JAK2 were successfully refolded and purified. In the case of CHN1, while a large

number of conditions were shown to be suitable, only a few of them were free of the anti-aggregation reagents discussed above **(Table 4.3)**. When buffer F6, which showed the smallest OD, was used to refold CHN1 on a large scale, the protein was successfully refolded with a 10% refolding recovery (2mg of refolded protein from 20mg of denatured protein). By using similar strategies, JAK2 was also successfully refolded, with a 5% refolding efficiency.

4.3 Discussion

Using a number of strategies, we have successfully expressed ~90% of the SH2 domains chosen for analysis, i.e. the SH2 domains expressed in Jurkat T cells with an mRNA abundance of more than 2 FPKM, excluding SH2 domains belonging to the JAK, STAT and SOCS families. The final list of SH2 domains for this project is presented in **Table 4.4**, and the proteins purified are shown in **Figure 4.10**. It proved to be extremely difficult to express and purify the JAK and SOCS family proteins at all, and the yield for the full-length protein for the STATs was very low. The decision to exclude these proteins from the set also rested on the fact that they are well-characterised signalling molecules apparently confined to cytokine receptor signalling pathways [316].

Of the unsuccessful expression targets in our study, SLAP, LNK and PI3KR2 have been shown to play an important role in the earliest stages of TCR signalling [255, 317, 318], and therefore extra efforts will be made in the future to purify them. This will likely involve expressing them with Halotag7, redesigning the constructs and probably refolding. On the other hand, while we will try to express and purify CHN2, RIN3, SH2D3A and FER, they do not have high priority as SH2D3A [319] and FER [320] are involved in growth factor receptor–signalling pathways, CHN2 [321] is in the insulin-signalling pathway, and RIN3 [322] is involved in endocytosis.

Table 4.3. The refolding screening results for CHN1. The highlighted cells indicated conditions with an OD<0.05, while without glycerol, Brij58, L-arginine or glucose. Please refer to **Appendix 1** for the conditions wells A1-H12.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	0.033	0.003	0.061	-0.004	0.042	0.037	0.038	0.028	0.041	0.065	-0.02	0.056
B	0.015	0.025	0.036	0	0.07	0.025	0.003	0.043	0.075	0.057	0.053	0.061
C	0.014	0.024	0.004	0.012	0.026	0.115	0.01	0.043	0.074	0.049	0.006	0.063
D	0.014	0.069	0.076	0.013	0.119	0.067	0.017	0.054	0.056	0.014	0.016	0.071
E	0.005	0.078	0.122	0.008	0.038	0.065	0.009	0.066	0.015	0.06	0.013	-0.007
F	0.005	0.07	0.12	0.003	0.006	0.007	-0.03	0.067	0.068	0.029	0.001	0.001
G	0.028	0.069	-0.01	0.021	0.001	0.077	-0.01	0.083	0.012	0.019	0.064	0.062
H	0.049	0.088	0.125	0.025	0.062	0.003	0.004	0.007	0.051	0.062	0.005	-0.004

Table 4.4. The final list of SH2 domains expressed for this study. Light green: expressed as a His fusion protein; dark blue: expressed as a His fusion protein ('core region'); pale pink: expressed as a His fusion protein (murine); dark green: expressed as a TF fusion protein; yellow: expressed as a GST fusion protein; light blue: expressed as a Trx fusion protein; bright green: expressed as a Halotag7 fusion protein; orange: produced by refolding from inclusion bodies; bright pink: essential non-expressible targets subject to future attempts; purple: non-essential non-expressible targets subject to future attempts; dark red: non-expressible targets excluded from our target list.

1	VAV1	22	SHIP2	43	PIK3R2_N+C	64	PLCG2_N+C	84	SOCS5		
2	VAV2	23	SLA2	44	PIK3R3_N+C	65	PIK3R2_C	85	SOCS6		
3	LCK	24	VAV3	45	PLCG1_N+C	66	SLAP	86	SOCS7		
4	SH3BP2	25	PIK3R3_N	46	GAP_N+C	67	LNK	87	CISH		
5	SHP1_C	26	PIK3R3_C	47	ZAP70_N+C	68	CHN2				
6	SHP2_N	27	ABL2	48	SHP2_N+C	69	RIN3				
7	SHP2_C	28	CRK	49	SHP1_N+C	70	SH2D3A				
8	SHIP1	29	CRKL	50	SYK_N+C	71	FER/ TYK3				
9	NCK2	30	SH2D3C	51	CBL						
10	PLCG1_N	31	YES1	52	CBLB	72	JAK1				
11	PLCG1_C	32	GAP_C	53	SHC1	73	JAK2				His
12	CSK	33	SHB	54	SHP1_N	74	JAK3				Extended
13	FYN	34	ABL1	55	SH2B	75	TYK2				Murine
14	GADS	35	PIK3R1_N	56	GAP_N	76	STAT1				TF
15	SHF	36	PIK3R1_C	57	SH2B2	77	STAT2				GST
16	SAP	37	NCK1	58	PIK3R2_N	78	STAT3				TRX
17	GRB2	38	GRAP	59	LCP2	79	STAT5A				HALO7
18	ZAP70_N	39	SYK_C	60	SYK_N	80	STAT5B				Refolding
19	ZAP70_C	40	SRC	61	SPT6	81	STAT6				Future†
20	SH2D2A	41	PLCG2_C	62	PLCG2_N	82	SOCS1				Future*
21	ITK	42	PIK3R1_N+C	63	CHN1	83	SOCS4				Excluded

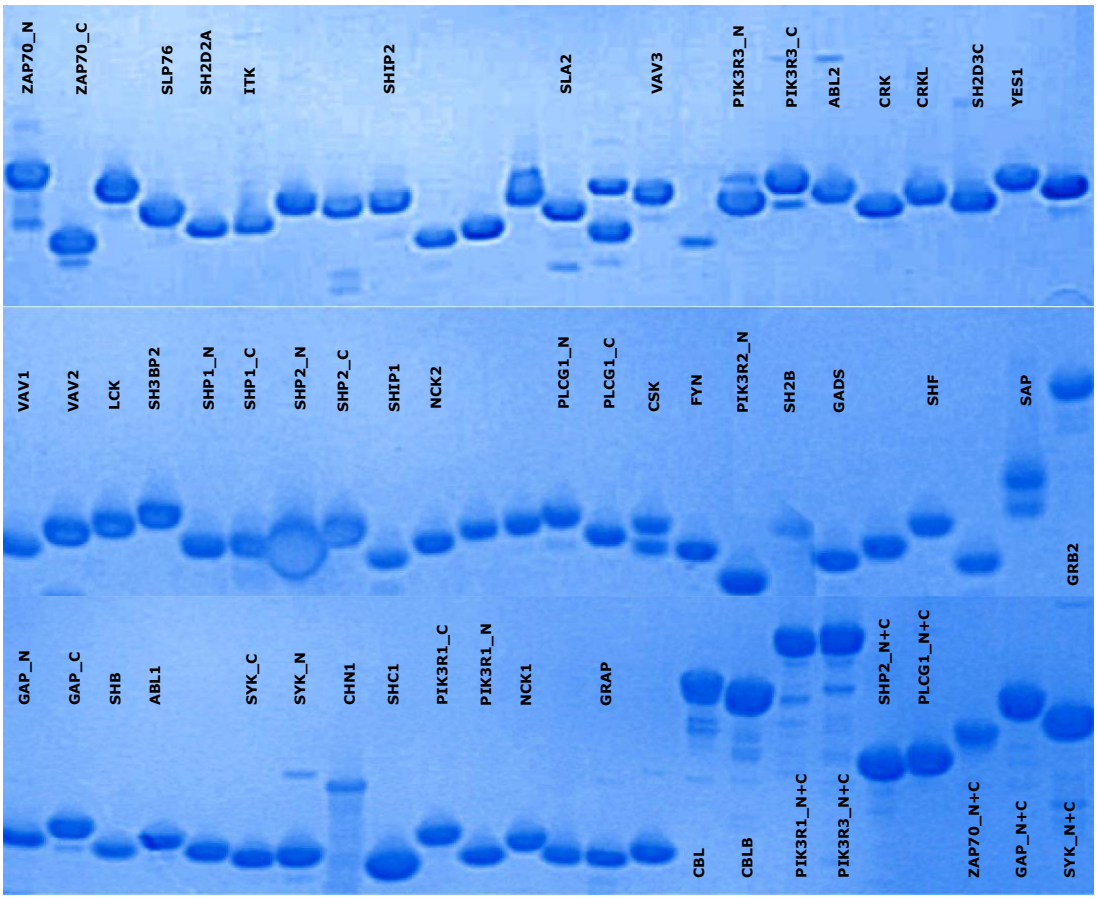


Figure 4.10. Summary SDS-PAGE gel showing the SH2 proteins expressed and purified in this study.

One important issue arisen from this study is that although a soluble expression screens are widely used as the gold standard for evaluating the effectiveness of the solubilising fusion tag, our results show that positive results in the expression screen cannot guarantee successful translation into scale up expression and, importantly, expression of properly folded proteins. For some of the good solubiliser used in a number of small-scale expression screen-based studies, no tag-free protein could be purified in our study although some gave good expression of fusion protein in the scale up. This also raises the question of whether it is appropriate to use the whole fusion protein as the analyte in biophysical studies as done in a previous large-scale interaction network

identification assay [191], and we suggest that it is of great importance to cleave the fusion partner and purify the tag-free target protein.

As proper folding of recombinant proteins is generally a prerequisite for reliable biophysical assay such as surface plasmon resonance experiments, all the proteins expressed in this study were purified by HPLC, and proper folding of the protein was indicated by the correct elusion volume to the corresponding size. In order to further confirm the folding status of the expressed protein, circular dichroism (CD) spectroscopy [323] and thermal denaturation will likely be performed in the future.

Chapter 5. Affinities of SH2 domains for the phosphotyrosine receptor motifs of T cell[3]

5.1 Introduction

5.1.1 The nature of molecular interactions involved in the initial step of intracellular signal transduction

After activation of T cells, the cytoplasmic tyrosine motifs of the surface receptors will become phosphorylated and these serve as docking sites for recruitment of SH2 domain–containing proteins. This comprises the initial step of intracellular signal transduction. It has been thought that the binding of SH2 proteins to the tyrosyl phosphopeptides is highly selective, and that signal transduction proceeds through a linear reply of information via sequential biomolecular protein interactions. However, this view could reflect the limitations of the methodologies used to investigate these interactions. The most widely applied methods, i.e. immune-precipitation or yeast two hybrid screens will under most conditions only be capable of identifying single binding partners of the highest affinity at a time, whereas those with lower binding affinity unable to be detected in this way can possibly bind to the receptors and have a large impact on the signalling outcome. On the other hand, studies have shown that the interactions of an SH2 domain with the pTyr residue is responsible for a large fraction of the binding energy of interaction, leaving the contribution of other amino acids close to the pY being relatively small [182]. This implies that the extent to which the domain can be selective may be limited, and that cross-reactivity is likely to occur.

[3] The SPR-based study described in this chapter was done with the assistance of Ms. Mai Vuong, T-cell Biology Group, University of Oxford.

In effect, a recent report showed that, by using a protein microarray approach, a large number of SH2 proteins can bind to each of the tyrosyl phosphopeptides of ErbB receptors with a wide range of affinity. The authors suggested that this promiscuity may contribute to the oncogenic potential of receptor tyrosine kinases and possibly other signalling proteins [191]. Another study by Breitkreutz *et al.* also showed that extensive cross reactions involving kinases and phosphatases, and they argued that although the main response may be derived from the canonical pathway mediated by some of the "specific" interactions, the overall outcome of signalling will likely be influenced by many less specific interactions and complementary responses [192].

Although signal transduction inside T cells has been under intensive investigation for several decades, understanding the nature of signalling has been limited by the way in which the important interactions can be investigated as discussed above, so that, for example, only SHP2 and PIK3 were thought to interact with the tyrosyl phosphopeptides of CTLA-4 [113, 114], albeit with the binding of SHP2 being controversial [116]. However, in the light of the large number of SH2 proteins present inside a T cell and the nature of molecular interactions between these types of proteins and their ligands, it seemed possible that such limited interactions would not be sufficient to account for the complicated responses of these cells to environmental changes and extracellular signals. A quantitative biophysical analysis of protein recruitment, performed on a transcriptome-wide scale, was therefore expected to provide valuable insight into the nature of intracellular T cell signalling at a system level.

5.1.2 Surface plasmon resonance

The Surface Plasmon Resonance (SPR) technology using the BIAcore instrument has been proven to be a method that is particularly valuable for monitoring low-

affinity biomolecular interactions. Although a number of techniques are available for studying molecular interactions (reviewed in [324]), they are superseded by SPR for a number of reasons. These include the capability of performing (fluorescence or radioisotope) label-free detection of biomolecules that interact in real time (thus allowing the measuring of binding kinetics), the much greater sensitivity of the method and the need for relatively small amounts of protein. Moreover, a wide range of analytes, including proteins, complex conjugates and small chemicals, can be analysed using this technology.

SPR is an optical phenomenon occurring at the interface of two transparent media of different refractive indices, one of which is adjacent to a metal film. When a beam of polarised light passes from a material with a high refractive index (e.g. glass) into material with a lower refractive index (e.g. liquid), the light can be completely reflected (total internal reflection) when the angle of incidence is greater than the critical angle. However, when the surface of the glass is coated with a semi-transparent noble metal film a fraction of incident light photons are absorbed by free electrons, leading to the oscillation of the electron plasma at the surface of the metal film and the generation of an evanescent field. These oscillating electron waves are called surface plasmons, which will propagate a short distance into the media of lower refractive index. When the wave vector of the incident light matches the wavelength of the surface plasmons, resonance of the electrons occurs (i.e. surface plasmon resonance). When the incident light is shone at a specific angle, called the surface plasmon resonance angle (θspr), maximum energy loss and minimum intensity of reflected light result from surface plasmon resonance. The SPR angle is highly sensitive to changes in the refractive index at the metal surface, which can be caused by, for example, the adsorption of protein onto the surface. An increase in the surface concentration can lead to an increase in the SPR angle, and thus a change in the intensity of the

reflected light, which will be detected by a photo-detector. The change is measured as response units (RUs), with 1 RU equivalent to a shift of 10^{-4} degrees.

5.1.3 Implementation of SPR in the BIAcore instrument

Biomolecular interactions are monitored on the surface of a sensor chip in the BIAcore instrument. The sensor surface comprises a glass surface with a thin gold film that is covalently coated with a carboxymethylated dextran matrix, where proteins are immobilised. The change in SPR angle that occurs when a ligand binds the immobilised protein is recorded as a sensorgram, where a signal of 1 RU corresponds to a change in surface protein concentration of approximately 1 pg/mm^{-2}. Combining this with the approximate thickness of 100nm of the dextran matrix, the molar concentration (mol/L) of the binding analyte can be calculated by RU / (100 x M_r), where M_r is the molecular mass of the analyte [325-328]. By combining the data obtained from the BIAcore instrument with the accurate concentration of injected ligand will allow measurement of binding affinity.

5.1.4 Immobilisation of ligands to the sensor chip surface

CM5 sensor chips were used in the BIAcore experiments in this study. Each sensor chip provides four individual flow-cells, which allows immobilisation of four different proteins and simultaneous monitoring of four different binding events. Protein immobilisation is achieved through two-step amine coupling. The first step is to activate the carboxymethyl groups with a 1:1 solution of EDC [1-ethyl-3-(3-dimethylaminopropyl) carbodiimide] / NHS (N-Hydroxysuccinimide). EDC functions to form a highly unstable activated acid intermediate and NHS reacts with the carboxymethyl group of the acid to form a highly reactive succinimide ester. The second step is to pass a protein through the flow-cell in a buffer that

produces a positive charge on the protein (i.e. a pH below the isoelectric point of the protein), so that it will be attracted to the negatively charged matrix, allowing formation of a secondary ester between the NHS-ester and the primary amine group ε-amino groups of lysine residues of the protein. The unreacted carboxylmethyl groups are then deactivated by passing through high concentrations of ethanolamine hydrochloride, which also helps to remove any non-covalently bound proteins.

The strategy described above is called direct immobilisation. Although this method produces a very stable level of immobilisation, it is non-directional such that not all of the ligands are orientated towards the analyte in solution. Importantly, this may interfere with the binding sites of the immobilised protein. Also, some proteins are unstable at pHs below their isoelectric points. To address this, the indirect ligand immobilisation technique has been established and widely use, in which an intermediate molecule is immobilised using the EDC/NHS chemistry and the protein of interest is immobilised onto the matrix via such an intermediate [329]. This allows immobilisation of the ligand through a defined site and for the protein to project away from the sensor chip surface. However, a disadvantage of such a strategy is the instability of the interaction between the intermediate and the protein of interest, particularly in the case of using antibody as an intermediate.

An indirect ligand immobilisation approach, using streptavidin as the intermediate, has been employed in the present study. Immobilisation of the ligand of interest onto the chip surface is through interaction between the streptavidin and the biotin group incorporated at the N-terminus of the peptide. The binding of the biotin to streptavidin represents one of the strongest non-covalent interactions in nature, with a K_D value on the order of 10^{-14} M, which allows stable immobilisation of the ligand on the chip surface.

5.1.5 Protein interaction affinity analysis

The affinity between an analyte and the ligand is usually measured as the dissociation constant (K_D). To obtain this data, a serial dilution of the analyte, ideally from 0.01x to 100x K_D, is prepared and passed through the sensor chip immobilised with the ligand. To obtain accurate data, adequate duration of time would be given for each injection such that the equilibrium can be reached and maintained. In addition, sufficient time gaps between injections should be given to allow all the analyte to dissociate. Since aggregated materials in the sample will tend not to dissociate from the immobilised ligand and affect the net responses for the subsequent injections, the sample is usually subjected to gel filtration to ensure that only a monomeric form of the analyte is used. Besides, as a key assumption in the affinity measurements is that the level of active immobilised ligand remains constant, the experiment is usually designed in such a way that the samples are injected in order of increasing and then decreasing concentration. Ideally, the response level should be comparable, and the same affinity should be obtained in both cases.

To obtain an affinity constant in this study, a Langmuir model ($A + L \Leftrightarrow AL$; A indicates analyte; L indicates ligand) is usually applied. When a monovalent interaction occurs between the analyte and the ligand, the data should conform to the Langmuir binding isotherm,

$$Bound = \frac{C^A \times Max}{C^A + K_D}$$

where "Bound" is the net binding responses, which is obtained by subtracting a non-specific control response from the binding response measured in RUs and "Max" is the maximum response. C^A is the concentration of the analyte injected.

The Bound values are then plotted against the corresponding C^A, and the data are fitted with a non-linear curve fitting to obtain the K_D value. In addition, a

Scatchard plot by plotting the Bound / C^A (Free) against Bound is usually generated, and a linear fit of the data will not only give the K_D value, but also provide a useful reference for the extent to which the data fit with the Langmuir model. If the interaction is not monovalent, for example, due to the bivalence of the analyte (exists in a dimeric form) or the immobilised ligand (e.g. possessing of two binding sites with different properties), the Scatchard plots derived from data of this type will be non-linear. If this is the case, fitting the data with a two-step binding model will result in a good fit and produce two K_D values [328, 330].

5.2 Results

To explore the complexity of recruitment of SH2 domains to triggered receptors expressed by T cells, a system-level investigation of the affinities of SH2 domains for phosphotyrosine-containing motifs derived from the most important T cell surface receptors was performed, which involved 1180 independent biophysical measurements in total. In the light of the large input of resources and effort, preliminary experiments was undertaken to optimise the conditions of assay.

5.2.1 Investigation of the effect of peptide length on affinity measurements

To optimise the experimental conditions for the SPR-based assay, a previous study of the interaction of the SH2 domain of LCP2 (also called SLP-76) with CD6, a costimulatory molecule required for optimal T cell activation [331, 332], was repeated. In that study, it was shown that the costimulatory effect of CD6 is mediated through binding of the SH2 domain of LCP2 (an adaptor protein) to a specific (phosphorylated) tyrosine residue, 662Y, in its cytoplasmic region [332].

The first question was whether the length of the peptide would have an impact on the interactions. While most SH2 domains have been reported to select for specific residues P – 2 [333] and P – 1 [334] N-terminal of the phospho-Tyr, and P + 1 [335], P + 2 [334], P + 3 [335], P + 4 [334] and P + 5 [336] C-terminal of the phospho-Tyr, so that residues outside the potential SH2-binding pockets between P – 2 and P + 5 should not affect the SH2 / peptide binding, a long CD6 (662Y) peptide of 15 amino acids (peptide-P: PDSTDNDDYDDISAA) was used in the previous study. As a result, we designed a short CD6 (662Y) peptide of 11 amino acids (peptide-T: TDNDDYDDISA) and compared the binding affinity between the long and short peptides with the LCP2 SH2 domain. Pairs of peptides (produced by Dr Kati DeGleria of the WIMM, Oxford University) were prepared phosphorylated or non-phosphorylated at the tyrosine residue, with biotinylation at the N terminus. The peptides were immobilised on a CM5 sensor chip *via* binding of the biotin to streptavidin, and all experiments were performed at 37°C.

Analysis of the data for equilibrium binding showed that the affinities of the LCP2 SH2 domain for the phosphorylated CD6 (662Y) peptide-P and peptide-T were 1.1 μM and 0.9 μM, respectively (**Figure 5.1** and **Table 5.1**), which were comparable with a K_D value of 0.5 μM reported previously [332].

Although the affinities of the LCP2 SH2 domain to the peptides of two different lengths were highly comparable, the data derived from the long peptide generated better results. This was indicated insofar as the Scatchard plots generated from peptide-P binding data fitted a straight line, whereas those from peptide-T showed some degree of deviation from linearity, suggesting that the 1:1 Langmuir binding isotherm did not perfectly fit the data in the latter case. While this is normally due to partial aggregation of the protein injected, it is

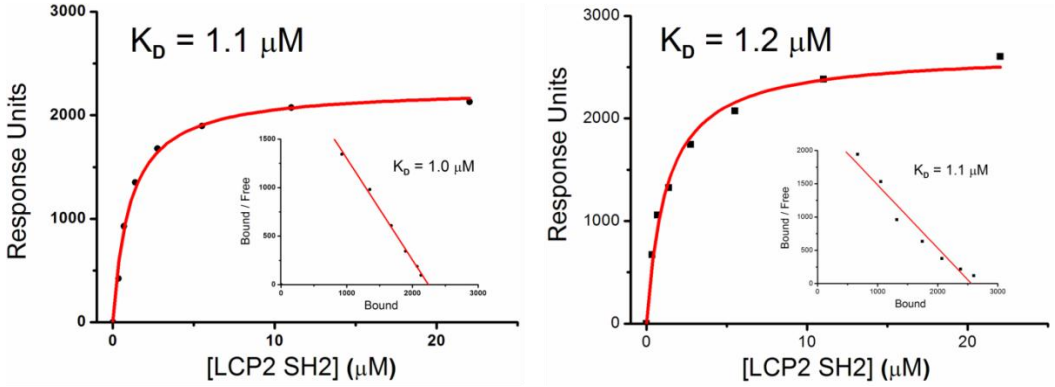


Figure 5.1. A representative example of equilibrium binding analyses for the LCP2 SH2 domain with the CD6 (662) long (peptide-P; left) and short (peptide-T; right) peptides. Approximately 400 RU of ligand was immobilised. Soluble proteins at a range of concentrations were injected at 20μl/min. All experiments were performed at 37°C.

Table 5.1. A summary of the affinities of the SH2 domain of LCP2 for CD6 (662) peptide-P and peptide-T.

Ligand immobilised	**CD6 (662) peptide-P (PDSTDNDDYDDISAA)**	**CD6 (662) peptide-T (TDNDDYDDISA)**
Analyte	LCP2 SH2 domain	LCP2 SH2 domain
Replicates	2	3
K_D (mean ± SD) / μM	1.1 ± 0.2	0.9± 0.2

unlikely in this case as the same batch of protein was injected into flow cells immobilised with both long peptides and short peptides in the same experiment. Since peptides of longer length seemed to generate data of higher quality, although the exact reason for this was not clear, 15-residue peptides, with eight residues N-terminal of the Tyr and six residues C-terminal of the Tyr, were used in the large-scale binding analyses in this chapter.

5.2.2 Investigation of the effect of choice of control peptide for the SPR experiments

The second important question was whether a universal control peptide, rather than a paired non-phosphorylated peptide, could be used as the negative control. This was important for three main reasons. Firstly, the use of a universal control peptide can greatly reduce the cost and increases the throughput. This is because one sensor chip can only accommodate four different peptides, which means the use of a universal control peptide can allow simultaneous analysis of three peptides of interest, while only two can be monitored if the non-phosphorylated peptide were used. Secondly, this can simplify and accelerate the analysing procedure as the same background response can be subtracted. Thirdly, the use of the same background makes the data derived from different peptides more comparable. To this end, the interactions between a set of SH2 domains that we have expressed and purified and the PD-1_ITIM motif-containing peptide were investigated. The universal control peptide used in this study, called the GS-control peptide, is a 15-residue neutral peptide with the sequence of GSGSGSGSGSGSGSG (G = Glycine, S = Serine). The binding affinities are presented as K_A (association constant; the inverse of the K_D) in all the figures comparing the affinities between different SH2 domains throughout this study as this can allow interactions with high affinity (high K_A values) ranking at the top.

It was shown that affinity data obtained using the GS-control peptide and the non-phosphorylated peptide are almost identical (**Figure 5.2**). The Pearson correlation coefficient derived from the comparison of the two sets of K_A values is 0.996, which indicates that these two sets of data are effectively identical (**Figure 5.3**). The corresponding responses seen in the flow-cell immobilised with the GS-control peptide and the non-phosphorylated peptide are also highly similar, particularly when low concentrations of analyte were injected (**Table 5.2**), indicating that any differences are likely to result from mechanical noise only.

Since there is essentially no difference between using the GS-control peptide and the non-phosphorylated peptide as the negative control, use of the GS-control peptide was adopted throughout the study.

5.2.3 Systems views of the SH2/tyrosyl phosphopeptides interactions

A large-scale quantitative study was performed to investigate the interactions between 59 SH2 domains expressed in Jurkat T-cells and the 20 tyrosyl phosphopeptides derived from some of the most important receptors. These receptors include TCR, LAT, CD28, ICOS, CTLA-4, PD-1 and BTLA,

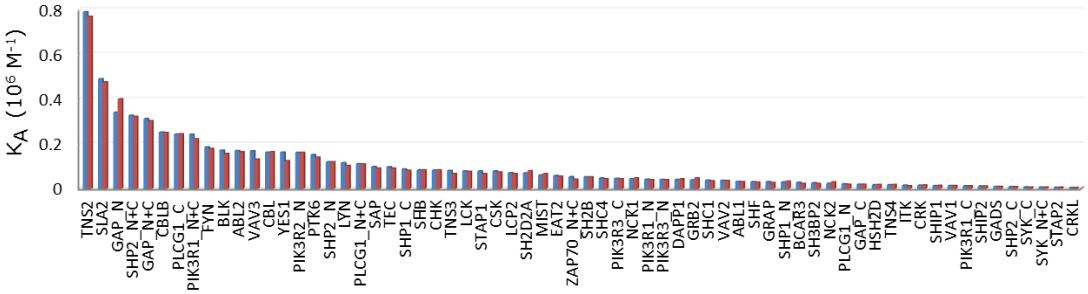


Figure 5.2. Comparison of the affinities of the SH2 domains and the PD-1_ITIM motif derived from data using two different peptides as the negative control. The affinities are presented as K_A values (association constant). The K_A values are sorted according to the affinity derived from data using the non-phosphorylated peptide as the negative control, from largest K_A value (highest affinity) to the smallest (lowest affinity; blue bars). The red bars represent affinities derived from data using the GS-control peptide as the negative control.

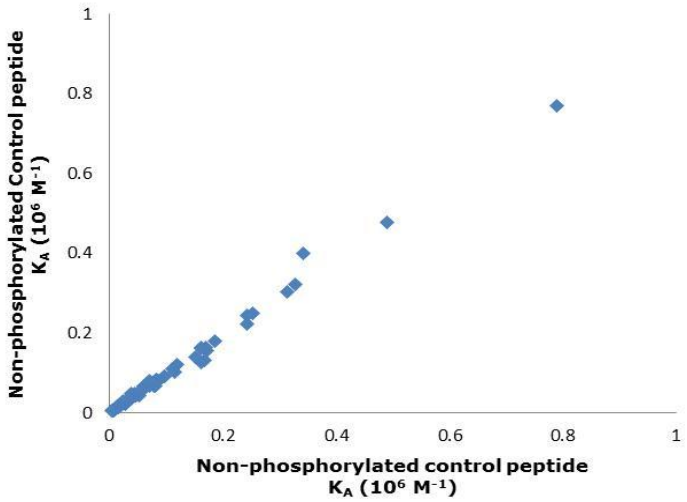


Figure 5.3. The K_A values derived from data using the non-phosphorylated peptide as the negative control were plotted against those derived using the GS-control peptide as the negative control. The Pearson correlation coefficient is 0.996.

Table 5.2. The responses resulting from the binding of TNS2 SH2 domain to PD-1_ITIM motif–containing peptide. A series of five concentrations were injected. The differences in RU between the two control peptides were calculated by subtracting the data derived from the non-phosphorylated peptide from those derived from the GS-control peptide.

Concentration of analyte (TNS2) / μM	Phosphorylated PD-1_ITIM peptide / RU	GS-control peptide / RU	Non-phosphorylated PD-1_ITIM peptide / RU	Difference between the two controls / RU
85.5	2687.1	2000.1	1842.3	157.8
42.75	1964.1	1105.9	1114.4	-8.5
21.375	1472.9	636.2	604.9	31.3
10.6875	1187.8	402.5	394.1	8.4
5.34375	942.9	229.8	222.3	7.5

which are crucial in controlling cell proliferation and activation and whose signalling is known to involve tyrosine phosphorylation. The tyrosyl phosphopeptides used in this study are listed in **Table 5.3**.

In the binding assays, a series of five concentrations, starting from ~100 μM, were injected for each SH2 protein, followed by a repeat of the experiment. This involved 11,800 injections in total. The decision to inject five concentrations was based on logistical considerations and on the expectation that this concentration range (6.25–100 μM) would provide reasonable data for affinities between 1 to 100 μM [326]. The affinities (K_D values) for the corresponding interactions were determined from fitting Scatchard plots and the affinities from two replicate experiments were averaged in all cases. When the affinity for an interaction was too high or too low, noisy data was collected so that not all five data points were usable for the affinity calculations. Whereas at least four points could be fitted with an R^2 (coefficient of variation) of >85% in most cases, fitting of three points with an R^2 >98% was occasionally adopted if the other two points could not be used. It was found that fitting of three data points with a high value of R^2 gave comparable results to those derived from using four points (**Figure 5.4**).

By using the SPR-based assays implemented by the BIAcore™, it was possible to determine 974 measurable affinities from the 1180 potential interactions between the 59 SH2 domains and the 20 tyrosyl phosphopeptides of T cell surface signalling receptors.. A representative set of Scatchard plots for the binding of SH2 domains with the PD-1_ITSM peptide is shown in **Appendix 3**. A quantitative graphical representation of the binding properties of SH2 domains to the tyrosyl phosphopeptides was constructed based on the biophysical data derived from this large-scale analysis (**Figure 5.5**).

Table 5.3. The tyrosyl phosphopeptides used to represent different receptor motifs. The single letter amino acid code is used. Y(-p) indicates the phosphorylated tyrosine residue

Receptor motif	Tyrosyl phosphopeptides representative of the receptor motifs
CD3δ_ITAM	ALLRNDQVY(-p)QPLRDRDDAQY(-p)SHLGGN
CD3ε_ITAM	PPPVPNPDY(-p)EPIRKGQRDLY(-p)SGLNQR
CD3γ_ITAM	TLLPNDQLY(-p)QPLKDREDDQY(-p)SHLQGN
CD3ζ_ITAM _1	QQGQNQLY(-p)NELNLGRREEY(-p)DVLDKR
CD3ζ_ITAM _2	RKNPQEGLY(-p)NELQKDKMAEAY(-p)SEIGMK
CD3ζ_ITAM _3	RGKGHDGLY(-p)QGLSTATKDTY(-p)DALHMQ
LAT_YENE	DGANSVASY(-p)ENEGAS
LAT_YVNVP	FSMESIDDY(-p)VNVPES
LAT_YVNVS	ASLDGSREY(-p)VNVSQE
LAT_YENL	VEEEGAPDY(-P)ENLQEL
CD28_YMNM	RSRLLHSDY(-p)MNMTPR
CD28_YAPP	RPGPTRKHYQPY(-p)APPRDF
CTLA-4_YVKM	RSPLTTGVY(-p)VKMPPT
CTLA-4_YFIP	ECEKQFQPY(-p)FIPIN
PD-1_ITIM	AVPVFSVDY(-p)GELDFQ
PD-1_ITSM	PCVPEQTEY(-p)ATIVFP
BTLA_YDND	VLLSETGIY(-p)DNDPDL
BTLA_ITIM	EENKPGIVY(-p)ASLNHS
BTLA_ITSM	NVKEAPTEY(-p)ASICVR
ICOS_YMFM	SVHDPNGEY(-p)MFMRAV

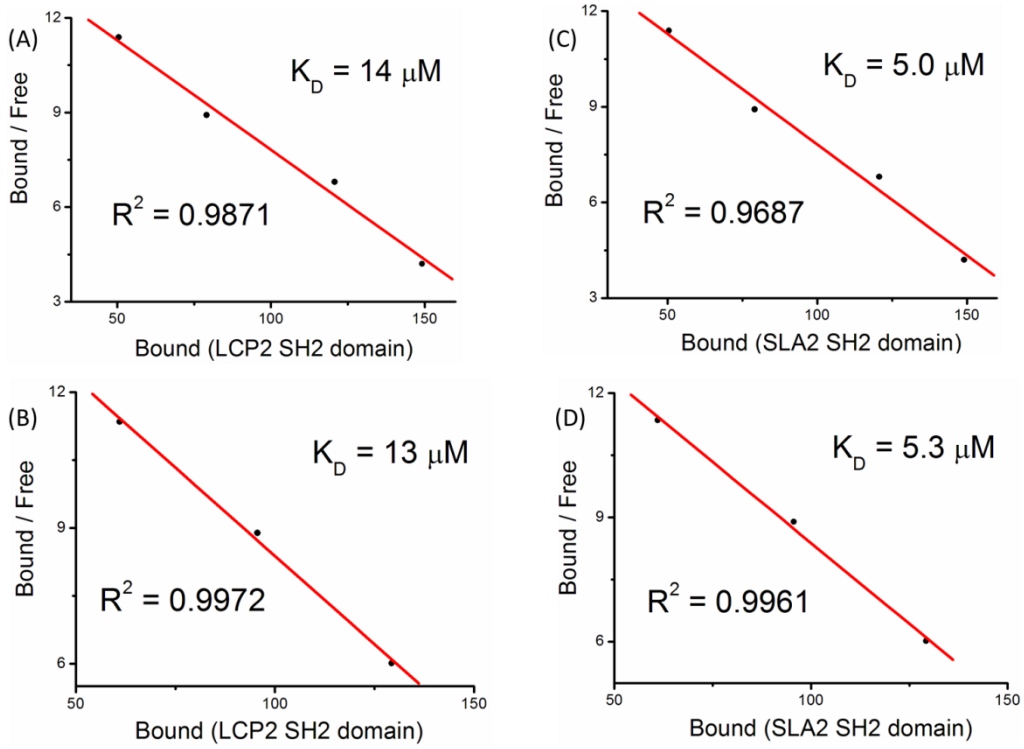


Figure 5.4. A comparison of the effects of using four and three data points for generating the Scatchard plots in two replicates. (A) and (B) Binding of the LCP2 SH2 domains to the ITIM motif of PD-1 in two replicate experiments. (C) and (D) Binding of the SLA2 SH2 domain to the ITIM motif of PD-1 in two replicate experiments. While four good data points were used for one replicate (A) and (C), only three were available for the other (B) and (D). The use of three points for fitting yielded comparable results when the fitting had a high coefficient of variation (R^2).

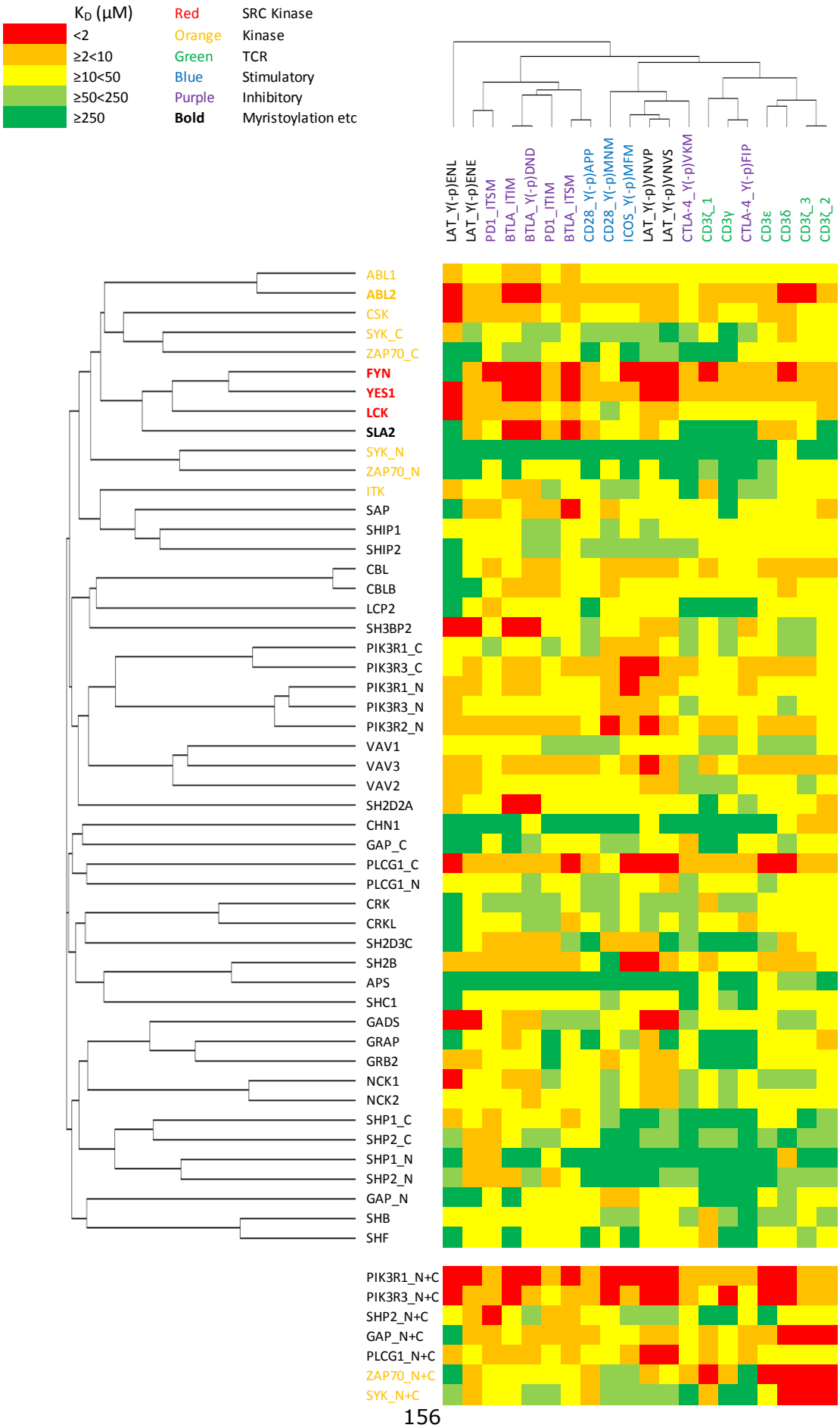

KD (μM)
<2
≥2<10
≥10<50
≥50<250
≥250
Red SRC Kinase
Orange Kinase
Green TCR
Blue Stimulatory
Purple Inhibitory
Bold Myristoylation etc
LAT_Y(-p)ENL
LAT_Y(-p)ENE
PD1_ITSM
BTLA_ITIM
BTLA_Y(-p)DND
PD1_ITIM
BTLA_ITSM
CD28_Y(-p)APP
CD28_Y(-p)MNM
ICOS_Y(-p)MFM
LAT_Y(-p)VNVP
LAT_Y(-p)VNVS
CTLA-4_Y(-p)VKM
CD3ζ_1
CD3γ
CTLA-4_Y(-p)FIP
CD3ε
CD3δ
CD3ζ_3
CD3ζ_2
ABL1
ABL2
CSK
SYK_C
ZAP70_C
FYN
YES1
LCK
SLA2
SYK_N
ZAP70_N
ITK
SAP
SHIP1
SHIP2
CBL
CBLB
LCP2
SH3BP2
PIK3R1_C
PIK3R3_C
PIK3R1_N
PIK3R3_N
PIK3R2_N
VAV1
VAV3
VAV2
SH2D2A
CHN1
GAP_C
PLCG1_C
PLCG1_N
CRK
CRKL
SH2D3C
SH2B
APS
SHC1
GADS
GRAP
GRB2
NCK1
NCK2
SHP1_C
SHP2_C
SHP1_N
SHP2_N
GAP_N
SHB
SHF
PIK3R1_N+C
PIK3R3_N+C
SHP2_N+C
GAP_N+C
PLCG1_N+C
ZAP70_N+C
SYK_N+C

Figure 5.5. A summary of interactions between SH2 domains and the tyrosyl phosphopeptides of T cell surface signalling receptors. The SH2 domains are clustered according to their sequence similarity, and the receptor motifs are clustered according to their SH2 binding patterns. SRC kinases are coloured red, other proteins possessing kinase activities are coloured yellow, and the protein (SLA2), which is palmitoylated, is shown in bold. The receptor motifs with stimulatory functions are coloured blue, and those with inhibitory functions are coloured purple. Where there is an SH2 domain-containing protein possessing two SH2 domains, "_N" indicates the N-terminal SH2 domain, "_C" indicates the C-terminal SH2 domain, and "_N+C" indicates the dual/paired SH2 domain as a whole. The affinity of the SH2 domains for the tyrosyl phosphopeptides are colour-coded according to decreasing affinity from red (K_D < 2 μM; strong affinity) to orange (2 μM ≤ K_D < 10 μM; upper-intermediate affinity), to yellow (10 μM ≤ K_D < 50 μM; intermediate affinity), to light green (50 μM ≤ K_D < 250 μM lower-intermediate affinity), and finally to dark green (K_D ≥ 250 μM; low affinity or undetectable).

The binding analysis shows that, instead of having highly selective interactions, high levels of cross-reactivity, with differential affinities, are seen for all the tyrosyl phosphopeptides in the study. **Figure 5.6** shows that these tyrosyl phosphopeptides bind to almost all of the SH2 domains, although a slightly smaller number of SH2 domains were shown to bind to CTLA-4_Y(-p)FIP, CD3_γ, CD3ζ_1 and LAT_Y(-p)ENL.

Whereas the binding profiles are comparable when the total number of SH2 domains is considered, differences emerge when the binding profiles are grouped into different ranges according to their affinities (**Figure 5.7**). Here, five affinity categories were used to dissect these affinity, with K_D <2 μM being strong affinity, 2 μM ≤ K_D <10 μM being upper-intermediate affinity, 10 μM ≤ K_D <50 μM being intermediate affinity, 50 μM ≤ K_D <250 being lower-intermediate affinity, and K_D ≥ 50 μM being low affinity. These designations will be applied throughout the rest of this thesis.

Most of the interactions are within the intermediate affinity range, with only a small number binding with extreme affinities. In addition, the difference in the affinity for most of the interactions is within two orders of magnitude. Three membrane-distal tyrosine-containing motifs of the LAT adaptor (LAT_YVNVP, LAT_YVNVS, LAT_YENL) were found to engage in a lot more high-affinity interactions compared to most other motifs. No strong interactions were observed between the SH2 domains and the tyrosyl phosphopeptides derived from the inhibitory receptor CTLA-4 and the ITIM motif of PD-1, and the ITSM motif of PD-1 only engages in two interactions with an affinity of K_D < 2μM. Interestingly, whereas BTLA was shown to be an inhibitory receptor, like CTLA-4 and PD-1 [337], a total of 21 high-affinity interactions was found for the three tyrosine-containing motifs of BTLA. In addition, the ITAM motif of CD3 δ was shown to interact with 8 SH2 domains with high affinities, whereas only an average of three such interactions were found for other ITAM motifs of the TCR.

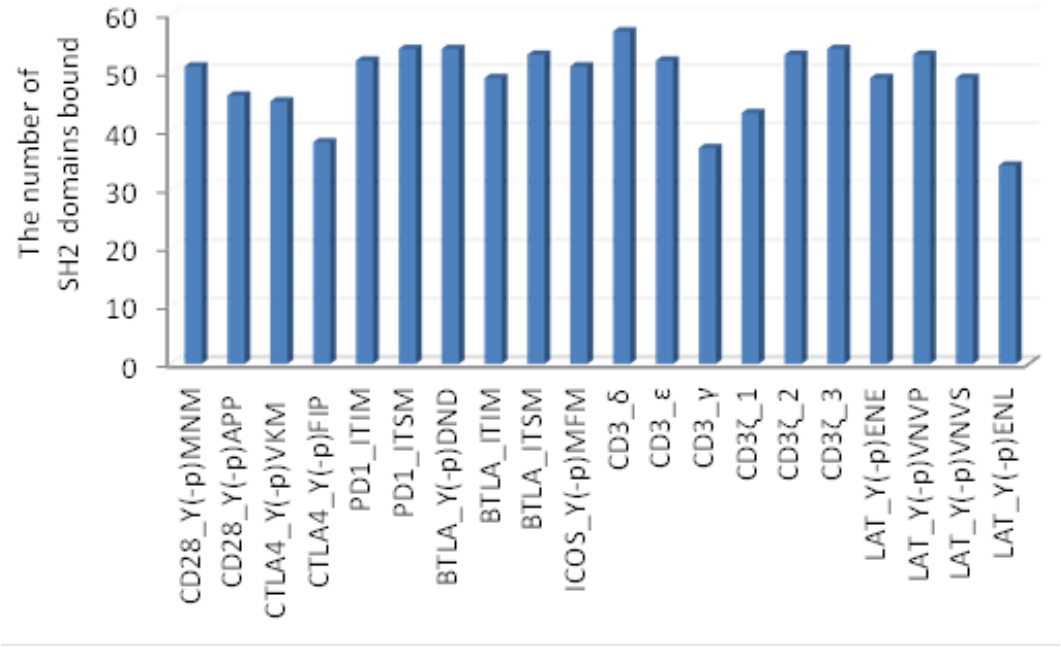


Figure 5.6. A summary of the number of SH2 domains found to interact with the tyrosyl phosphopeptides.

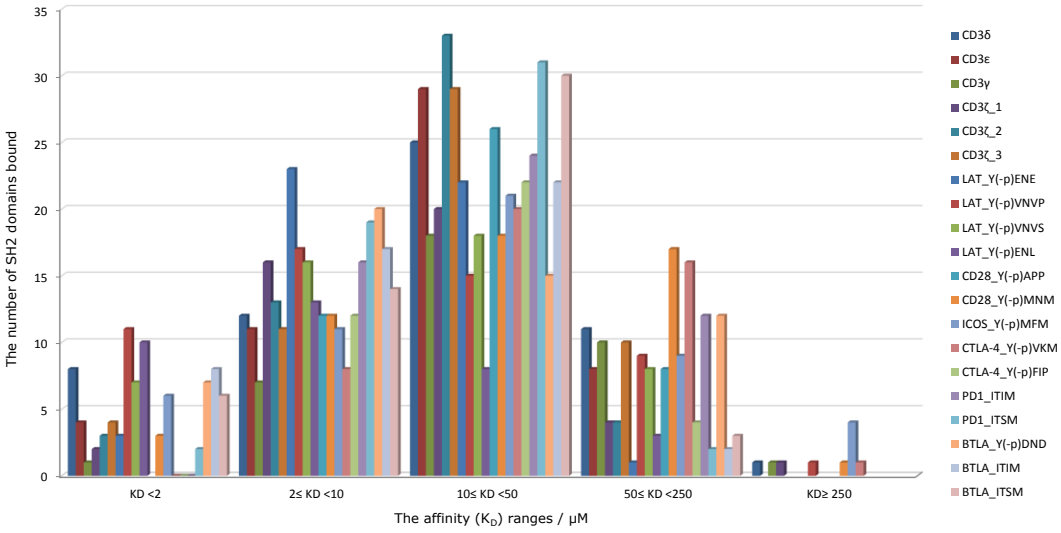


Figure 5.7. A summary of the number of SH2 domains binding to different tyrosyl phosphopeptides at different affinity ranges.). Five affinity categories were used: K_D <2 μM indicates strong affinity, 2 μM ≤ K_D <10 μM indicates upper-intermediate affinity, 10 μM ≤ K_D <50 μM indicates intermediate affinity, 50 μM ≤ K_D <250 indicates lower-intermediate affinity, and K_D ≥ 250 μM indicates low affinity.

One notable feature of the overall binding profile is that proteins belonging to the SRC kinase family, including LCK, FYN and YES1, are found to bind with relatively high affinities to most of the tyrosyl phosphopeptides. Almost all the interactions for FYN and YES1, and 50% for LCK, have at least an upper-intermediate affinity (K_D < 10 μM) with the tyrosyl phosphopeptides. Most notably, 50% of the binding for FYN have a K_D < 2 μM (**Figure 5.8**). In contrast, binding of proteins with phosphatase activities, i.e. SHP1 and SHP2, are found to be restricted to PD-1 and BTLA, which are inhibitory receptors. Whereas high-affinity cross-reactions have been observed for the dual SH2 domains of PIK3, and to a lesser extent for PLCG and GAP, the binding of ZAP70 and Syk paired SH2 domains are mainly restricted to the ITAMs of the TCR complex (**Figure 5.5**).

In order to obtain an overview of the SH2 binding profiles of each receptor as a whole, the strongest-binding SH2 domain for any of the constituent tyrosyl phosphopeptide was assigned to that receptor (**Figure 5.9**). From this qualitative view of the binding profile, all the receptors bound to almost all the SH2 domains; the TCR is the most promiscuous receptor, while CTLA-4 is the least.

By grouping the interactions according to the different affinity ranges that they fall into (**Figure 5.10**), it was found that the LAT adaptor has an exceptional capability of mediating high-affinity recruitments, with about 1/4 of SH2 domains binding to this receptor with an affinity of K_D < 2 μM, followed by BTLA and TCR. Here, the number of interactions within the upper-intermediate range is comparable with that within the intermediate range, and the number of interactions within the lower-intermediate range is relatively small, making most of the interactions fall into the upper-intermediate and intermediate ranges. This is in contrast to the situation when individual tyrosyl phosphopeptides were considered, where most of the interactions are within the intermediate range, and comparable levels of interactions are within the upper- and lower- intermediate ranges.

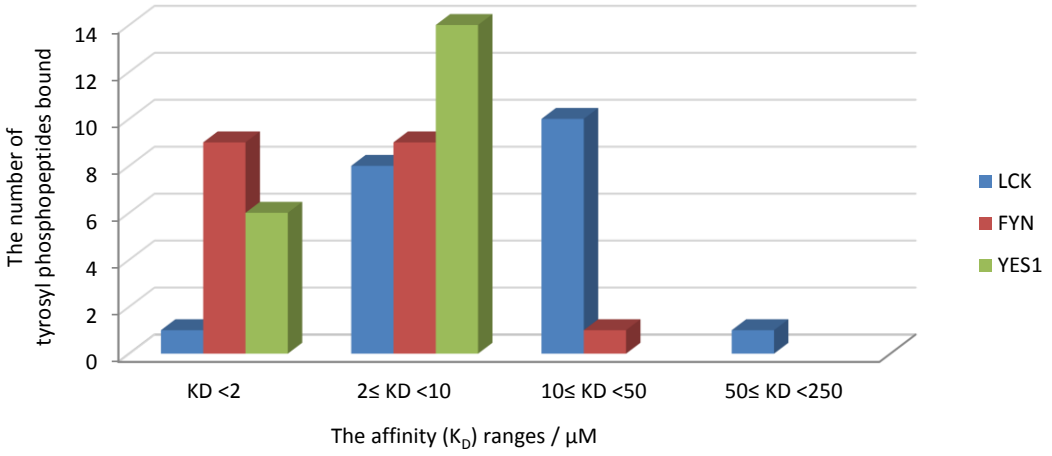


Figure 5.8. A summary of the interactions between the SH2 domains of LCK, FYN and YES1 with different tyrosyl phosphopeptides at different affinity ranges.

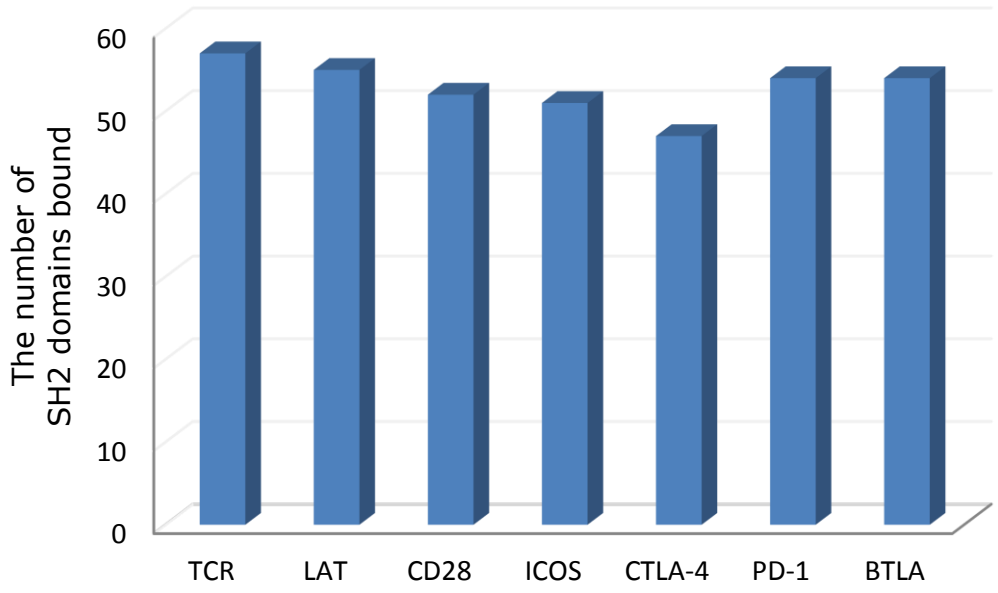


Figure 5.9. A summary of the number of SH2 domains found to interact with the T cell surface receptors investigated in this study.

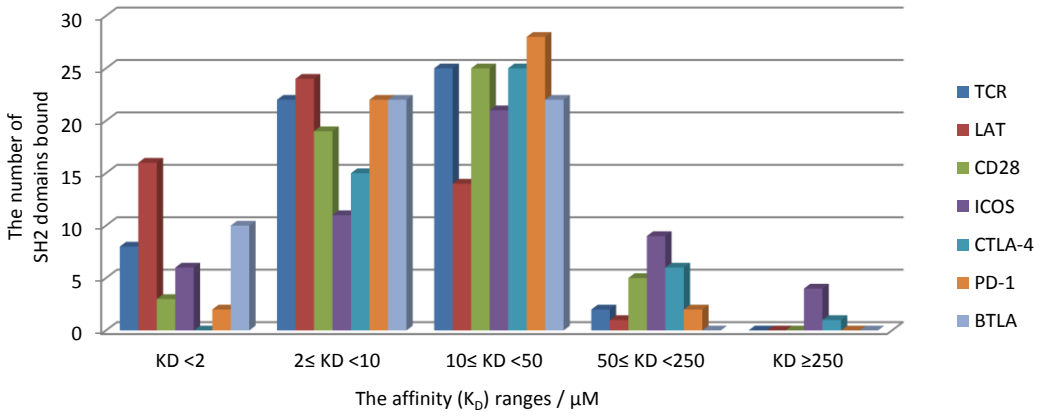


Figure 5.10. A summary of the number of SH2 domains binding to different T cell surface receptors at different affinity ranges.

5.2.4 The interactions between SH2 domains and the individual tyrosyl phosphopeptides

5.2.4.1 CD28

In good accord with previous immunoprecipitation-based studies in which association between CD28 and PIK3 was observed [98], the binding analyses showed that the PIK3 variants have much stronger affinities for the YMNM motif compared with other SH2 domains, with PIK3R3 being strongest (**Figure 5.11**). Whereas GRB2 interacts with CD28 with an approximately 10-fold lower affinity that of PIK3R3 (K_D = 7.1 μM vs 0.61 μM) in the present study, as reported previously [99], interactions between the YMNM motif and proteins possessing SH2 domains with higher affinity than GRB2, including ABL2, SH2D3C and YES1, were not identified in previous studies. Other GRB2-related SH2 domains which are supposed to bind to the same YXNX consensus sequence [97], including GADS and GRAP, were found to have comparable affinity for GRB2.

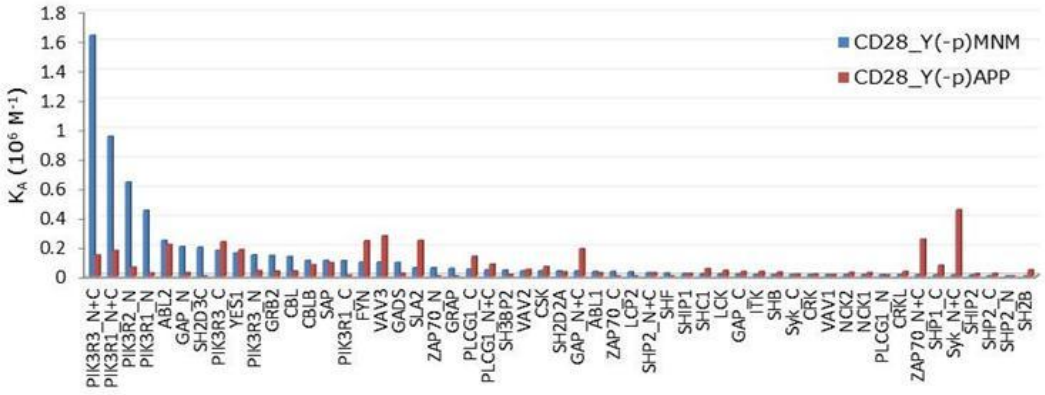


Figure 5.11. A summary of the affinities (presented as K_A values) of the SH2 domains for the Y(-p)MNM and Y(-p)APP motifs of CD28. The K_A values are sorted according to the affinity of the SH2 domains for the Y(-p)MNM motif, from largest K_A value (highest affinity) to the smallest (lowest affinity). The blue bars represent the affinity of the SH2 domains for the Y(-p)MNM motif, and the red bars represent affinity between SH2 domains and the Y(-p)APP motif.

Compared to the YMNM motif, the PIK3 variants bind much more weakly to the YAPP motif, with only about one tenth of the affinity of the YMNM motif. Whilst the affinities of SH2 domains to the YAPP motif are all fairly similar, Syk and ZAP70 bind particularly strongly to this motif (K_D = 2.2 μM and 4.0 μM, respectively).

5.2.4.2 CTLA-4

In the binding study PIK3 SH2 domains bound strongly to both the Y(-p)VKM motif and Y(-p)FIP motif, while the PLCG1 has slightly higher affinity than the PIK3 variants to the Y(-p)FIP motif (**Figure 5.12**). Following these are the two SRC kinases, YES1 and FYN, which bind strongly to both motifs. Whilst other SH2 domains have relatively weak affinity for the Y(-p)VKM motif compared to PIK3, a number of SH2 domains, including ABL2, GAP, SH3BP2, VAV3 and CRKL, can also

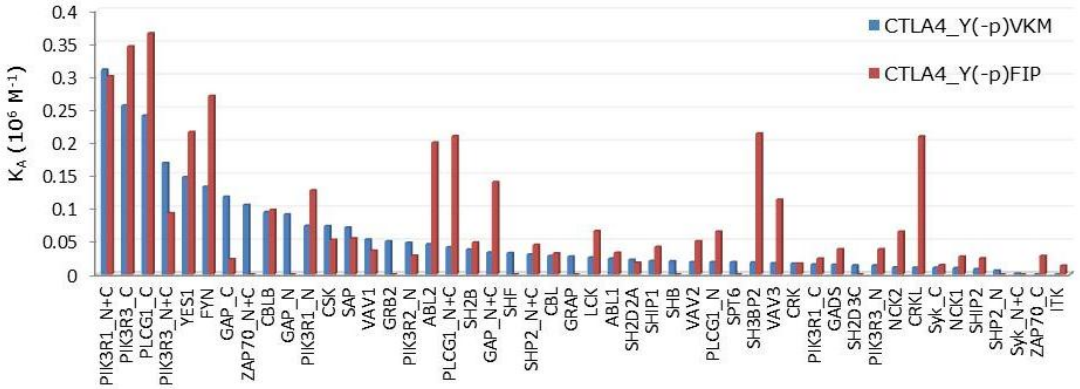


Figure 5.12. A summary of the affinities of the SH2 domains for the Y(-p)VKM and Y(-p)FIP motifs of CTLA-4.

bind strongly to the Y(-p)FIP motif. Surprisingly, whereas CTLA-4 has been shown to be a receptor with inhibitory function, no proteins with phosphatase activity bind CTLA-4 with high affinity, including SHP2, which is known to be controversial [116], although two molecules involved in the inhibition of SRC kinase activities, CBLB [338] and CSK [339], do bind to CTLA-4 with moderate affinity.

5.2.4.3 PD-1

Consistent with previous findings [126], SHP2 is among the strongest ligands for the ITSM motif of PD-1, and to a less extent to the ITIM motif (**Figure 5.13**). Strong interactions were seen between the two SRC kinases FYN and YES1 with PD-1: FYN and SHP2 are the strongest binders to the ITSM motif, whilst FYN and YES1 have the highest affinities for the ITIM motif. Whereas SHP1 was reported to bind to the ITSM [126], it was only found to have moderate affinity for this motif. In addition to SHP2 and FYN, a number of SH2 domains bound the ITSM motif with relatively high affinity. These include two other SRC kinases, LCK and YES1, as well as PIK3, PLCG1, GAP, whose affinities are only about half of those of SHP2 and FYN.

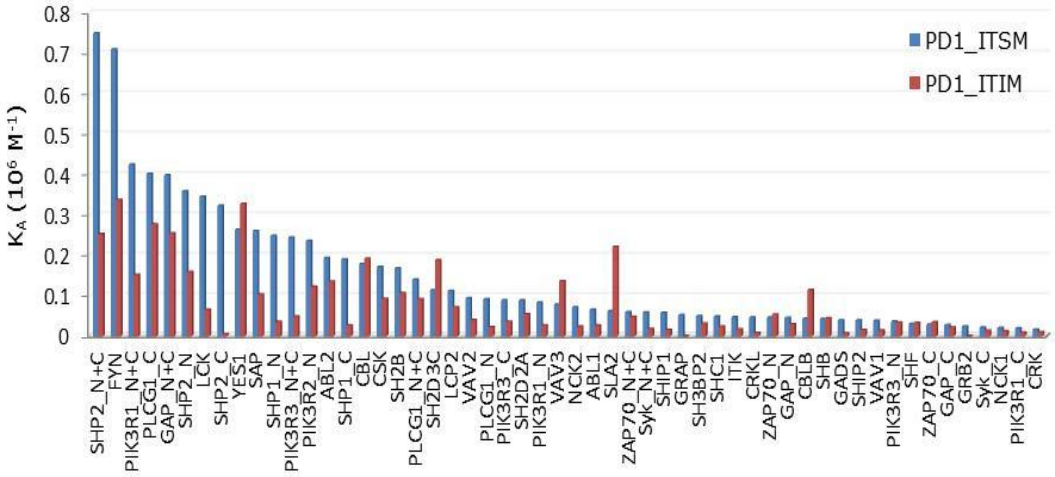


Figure 5.13. A summary of the SH2 domains for the ITIM and ITSM motifs of PD-1.

5.2.4.4 BTLA

The binding profiles of SH2 domains with the Y(-p)DND motif and ITIM are very similar, although the affinities of SH2 domains are generally higher to the Y(-p)DND motif (**Figure 5.14**). SH3BP2, SH2D2A, SLA2, FYN, YES1, ABL2 and PIK3 have comparable affinity and ranked at the top for both motifs. The very strong binding of two SRC kinases, FYN and YES1, to the ITSM is also notable, followed by PLCG1, SAP, SLA2, ABL2 and PIK3. Whereas SH3BP2 has the highest affinity for both the Y(-p)DND motif and the ITIM, its affinity for the ITSM is among the weakest. Surprisingly, whilst both SHP1 and SHP2 were previously reported to associate with both the ITIM and ITSM [146], the affinities of these proteins for the two motifs were found to be very low. In addition, while the Y(-p)DND motif was predicted to be a potential GRB2-binding YXNX consensus sequence [340], the binding of GRB2 to this motif is among the lowest compared with other SH2 domains.

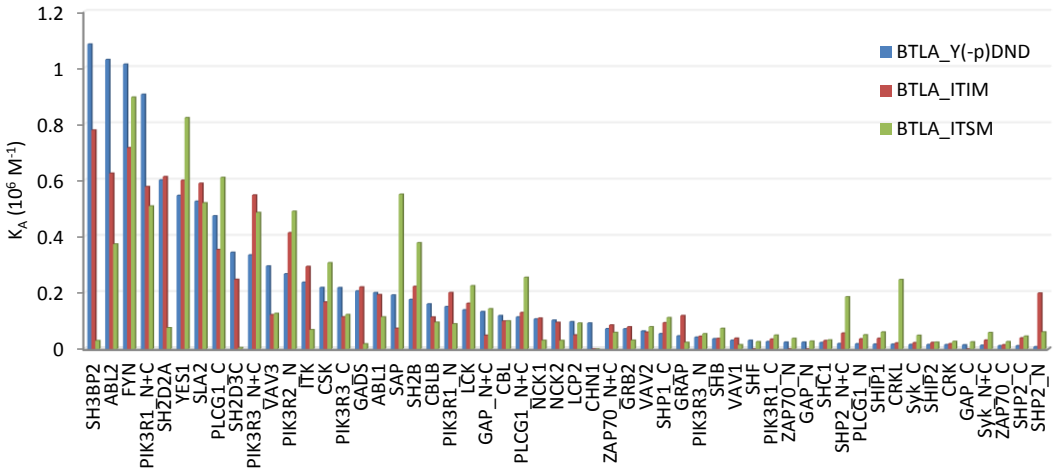


Figure 5.14. A summary of the affinities of the SH2 domains for the Y(-p)DND motif, as well as the ITIM and ITSM motifs of BTLA.

5.2.4.5 ICOS

In accord with the crucial role of PIK3 binding implicated with this receptor [163], extremely high affinities were seen for these proteins (approximately K_D = 0.5 μM) (**Figure 5.15**). Although the interactions of SH2B, FYN and PLCG1, YES1 and GAP for this motif seem to be a lot weaker than PIK3, their affinities are in effect very high, with a K_D value between 1.5 μM and 3 μM.

5.2.4.6 TCR

Although the binding analysis revealed broad interactions occur between SH2 domains and the CD3 ITAMs, interactions with very high affinity also exist (**Figure 5.16**). Interestingly, the PIK3 SH2 domains were found to have dominant affinities for the ITAMs of CD3δ, CD3ε and CD3γ, with a K_D value of approximately 0.4 μM to all three ITAMs. Although ZAP70 is ranked second after PIK3, its affinity for CD3δ and CD3ε is only about half, and only about 1/7 for

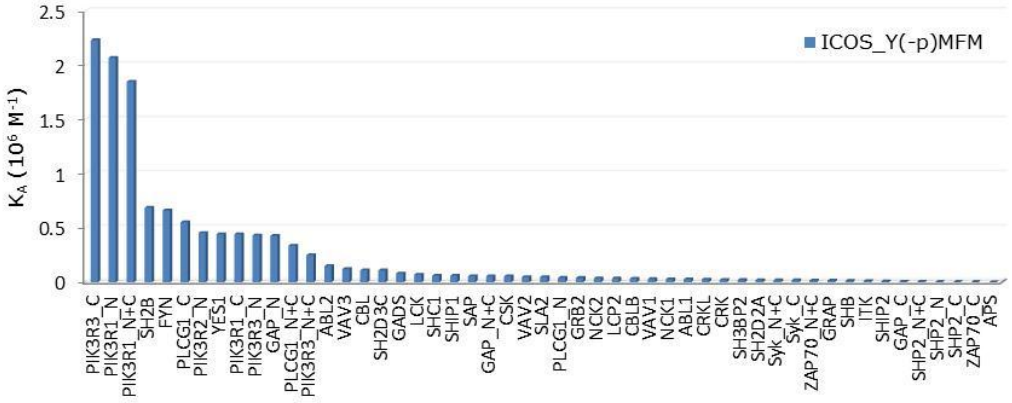


Figure 5.15. A summary of the affinities of the SH2 domains for the Y(-p)MFM motif of ICOS.

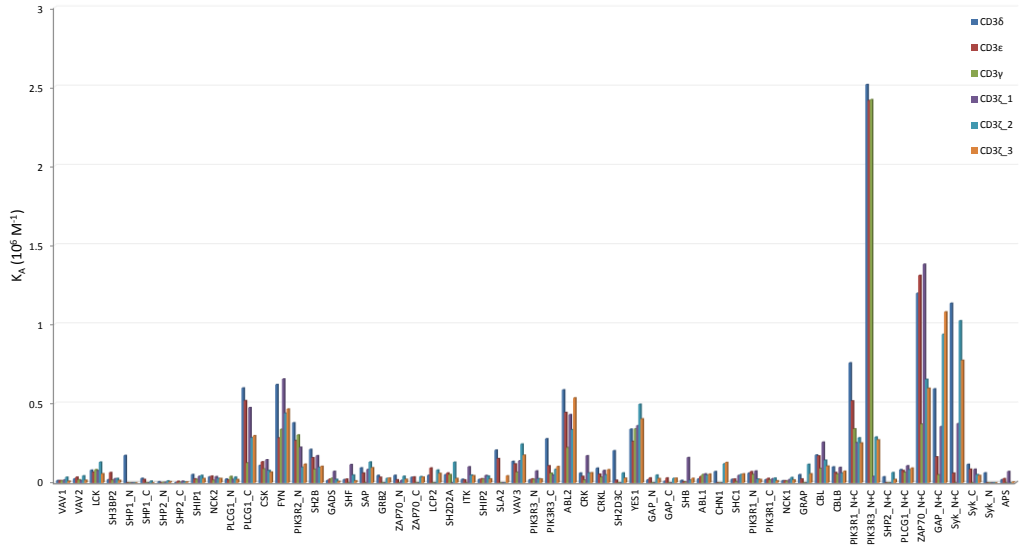


Figure 5.16. A summary of the affinities of the SH2 domains for the ITAM motifs of the CD3 subunits of the TCR complex.

CD3γ. Whilst Syk and GAP have comparable affinity to ZAP70 in the case of binding to CD3δ, their affinities to CD3ε and CD3γ are very low. This results in a potential dominant effect of PIK3 in the CD3γ signalling pathway.

In contrast, while PIK3 have been previously reported to associate with the CD3ζ [341], this is only a moderate affinity (approximately K_D = 4 μM) for all three CD3ζ ITAMs. ZAP70, Syk and GAP have the highest affinity for CD3ζ_1, CD3ζ_2 and CD3ζ_3, respectively. Whereas their affinities for CD3ζ_2 and CD3ζ_3 are comparable, the affinity of ZAP70 to CD3ζ_1 is about 4-fold higher than the other two.

5.2.4.7 LAT

The analysis of LAT revealed that a very strong ligand is PLCG1, which has outstanding binding affinities for YVNVP (Y171), YVNVS (Y191) and YENL (Y226) (KD = 0.24 μM, 0.56 μM and 0.23 μM, respectively) (**Figure 5.17**). Surprisingly, FYN is found here to have a much higher affinity for YVNVP (Y171, KD = 0.08 μM) than PLCG1 and slightly higher affinity for YVNVS (Y191, KD = 0.43 μM), although its affinity for YENL (Y226) is minimal. In addition, the PIK3 and YES1 bind with moderate affinity to YVNVP (Y171) and YVNVS (Y191); LCK, SH3BP2, CSK, GADS and PIK3 have comparable affinity for YENL (Y226). In contrast to the three well-characterised motifs, the Y(-p)ENE motif only binds weakly to PLCG1 (K_D =3.6 μM). This motif seems to be promiscuous, with SH3BP2 having the highest affinity and a number of others having comparable affinities. Since all four motifs contain the potential GRB2/GADS-binding YXNX consensus sequence, relatively strong binding of GADS to these motifs is seen (with a K_D ~ 0.65 – 1.86 μM), although the affinity of GRB2 to these motifs is only moderate (with a K_D ~ 2.4 – 5 μM).

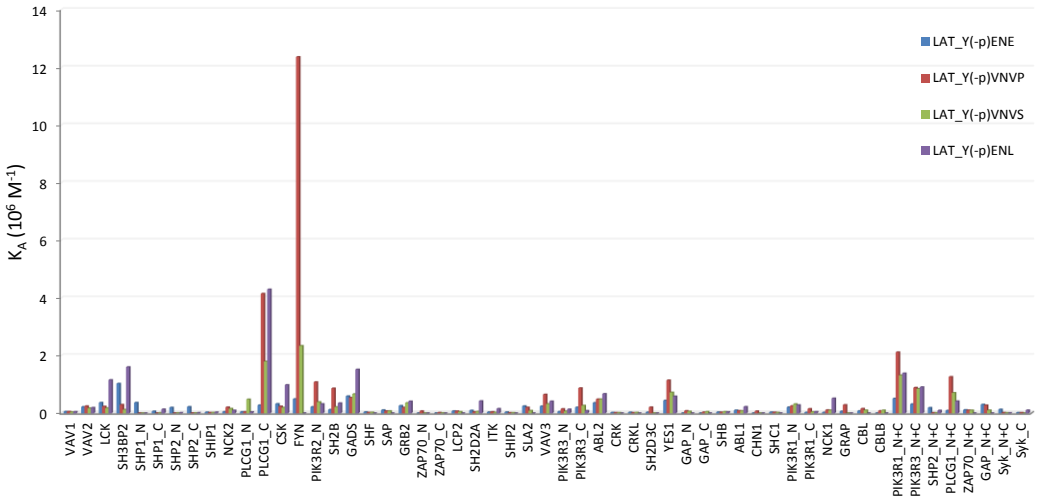


Figure 5.17. A summary of the affinities of the SH2 domains for the YENE (Y110), YVNVP (Y171), YVNVS (Y191) and YENL (Y226) motifs of LAT.

5.2.5 Comparisons of the binding profiles for the ITIM and ITSM motifs

5.2.5.1 The ITIM motifs of BTLA and PD-1

The ITIM (immunoreceptor tyrosine-based inhibitory motif) has the consensus sequence of [I/L/V]-x-(Y)-x-x-[L/V] [126]. The only conserved residue for the ITIM of PD-1 (VDYGEL) and BTLA (IVYASL) is the Y + 3 leucine C-terminal of the tyrosine. **Figure 5.18** shows that the BTLA ITIM can mediate interaction with several SH2 domains with high affinities, whereas no such interactions are seen for PD-1. However, the ITIM of PD-1 does interact with a larger number of SH2 domains at the lower-intermediate range. Apart from the difference at the two extreme ranges, similar numbers of SH2 domains bind to the ITIM of both receptors at the intermediate and upper-intermediate ranges, respectively. Overall, however, BTLA binds much more strongly than PD-1 to the SH2 domains. However, whilst SHP2, GAP and CBL are among the strongest binders to the ITIM

of PD-1, they only have moderate affinities for that of BTLA. Interestingly, the affinities of the SRC kinases FYN and YES1 to the ITIM are ranked at the top among the interactions of both inhibitory receptors (**Figure 5.19**).

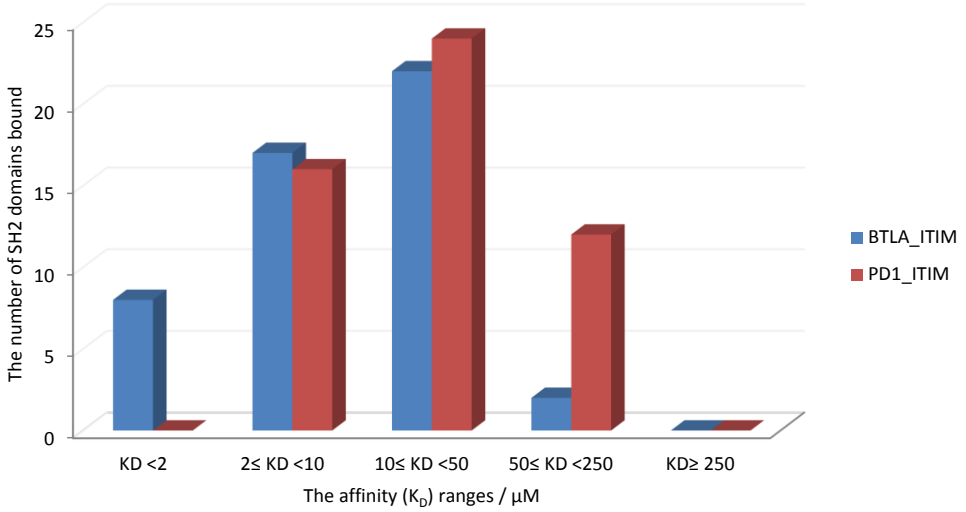


Figure 5.18. A summary of the number of SH2 domains binding to the ITIM motifs of PD-1 and BTLA at different affinity ranges.

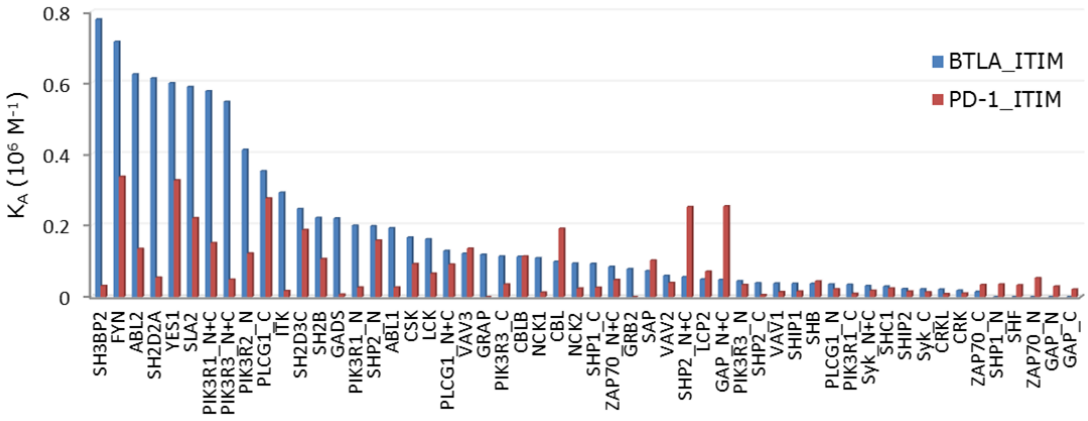


Figure 5.19. A summary of the affinities of the SH2 domains for the ITIM motif of PD-1 and BTLA receptors.

5.2.5.2 The ITSM motifs of BTLA and PD-1

The ITSM (immunoreceptor tyrosine-based switch motif) has the consensus sequence of [T]-x-(Y)-x-x-[V/I] [127]. The ITSM motif of PD-1 (TEYATI) and BTLA (TEYASI) are very similar, with only the Y+2 residue C-terminal of the tyrosine being different. The overall SH2 domain binding profile of the ITSM is similar between the two receptors, albeit the ITSM of BTLA is capable of recruiting a few more SH2 domains with high affinities compared to that of PD-1, while the latter can bind to more SH2 domains with an upper-intermediate affinity (**Figure 5.20**). By comparing the interactions in detail (**Figure 5.21**), it is found that FYN, YES1, PLCG1, SAP, and PIK3 all bind strongly to the ITSM of both receptors and rank at the top among all the interactions, although their affinities to BTLA are higher than for PD-1, as in the case of the ITIMs. However, one notable difference is that the strong interactions between SHP2, followed by GAP and SHP1, and PD-1, are not seen for BTLA.

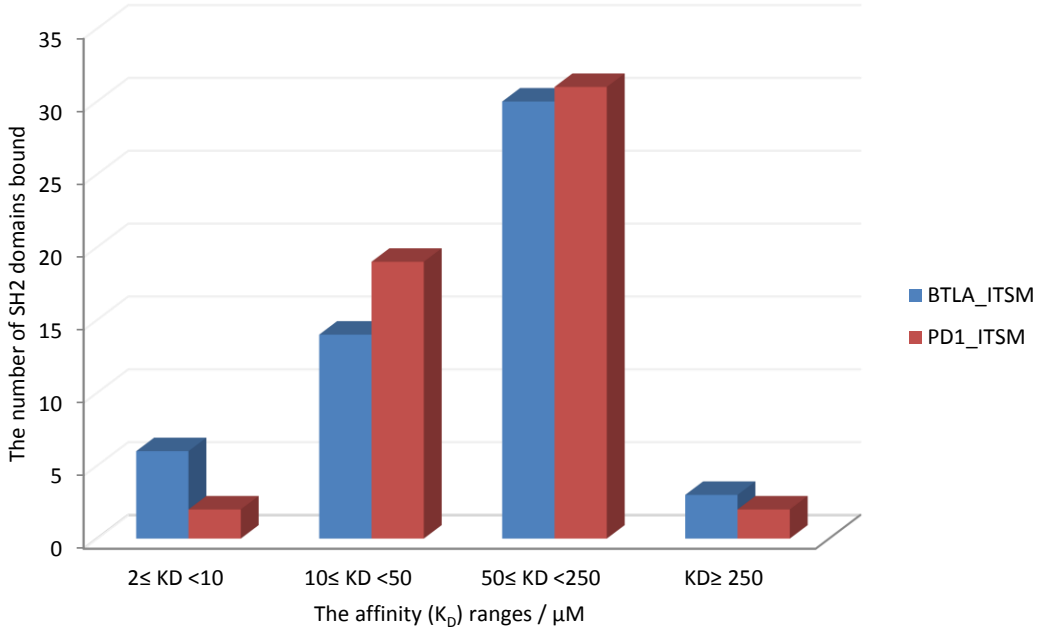


Figure 5.20. A summary of the number of SH2 domains binding to the ITSM motif of the PD-1 and BTLA receptors at different affinity ranges.

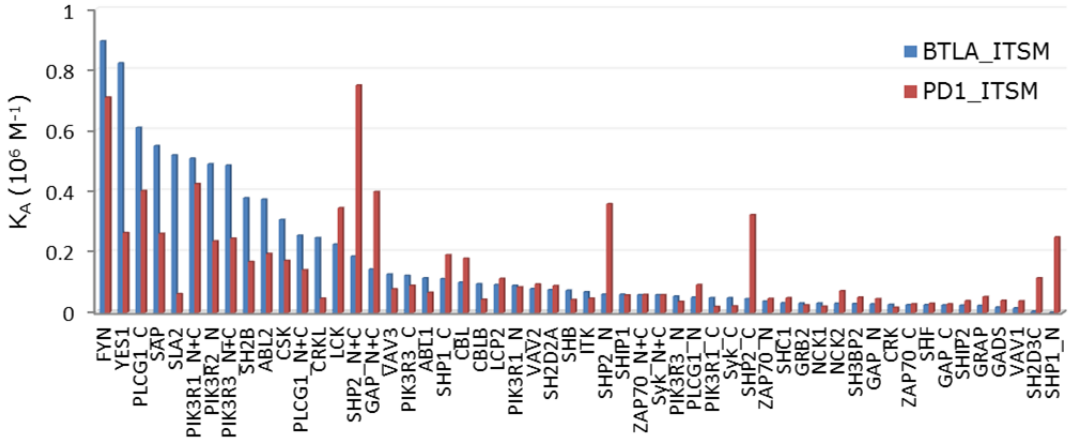


Figure 5.21. A summary of the affinities of the SH2 domains for the ITSM motif of the PD-1 and BTLA receptors.

5.2.6 Comparison of the SH2 domain binding profiles of different receptors with similar functions

In this section, SH2 domains are considered to be recruited to the receptor instead of individual tyrosyl phosphopeptides. To this end, the affinity set for each receptor is generated by taking the highest affinity for each SH2 domain of the individual tyrosyl phosphopeptides of the receptor.

5.2.6.1 ICOS and CD28

ICOS and CD28 are the most important co-stimulatory receptors, and promote activation, proliferation and survival of T cells [154, 342]. Both receptors contain the YMXM motif, which is considered to be the consensus PIK3 binding site and crucial for signalling. The binding profiles of the SH2 domains for these two motifs are very similar, including the scale of affinity, with the PIK3 having much higher affinity versus the other SH2 domains (Figure 5.22). However, whilst ICOS is

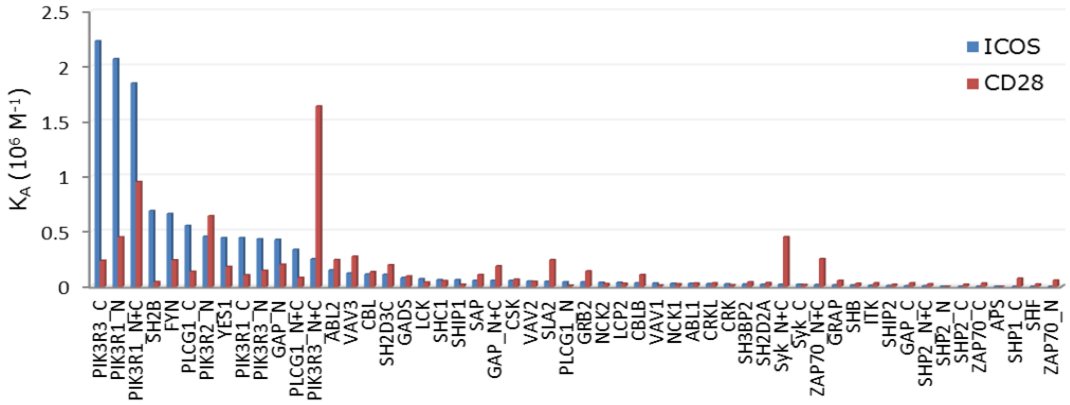


Figure 5.22. A comparison of the binding of the SH2 domains to the CD28 and ICOS receptors.

bound strongly by both PIK3R1 and PIK3R3, CD28 is bound most strongly by PIK3R3 and to a lesser extent by PIK3R1. Whereas several other SH2 domains have moderate affinities to ICOS, the binding to CD28 is to a large extent dominated by the binding of PIK3.

5.2.6.2 The TCR and CD28

While TCR triggering initiates activation of the T cells, ligation of CD28 amplifies the TCR-mediated signalling. Both receptors are found here to be bound most strongly by the PIK3, although CD28 seems to be specifically targeted by PIK3, while the TCR can also interact with a number of SH2 domains with fairly strong affinities: ZAP70, SYK and GAP have affinities approximately half of that of PIK3 for the TCR (**Figure 5.23**).

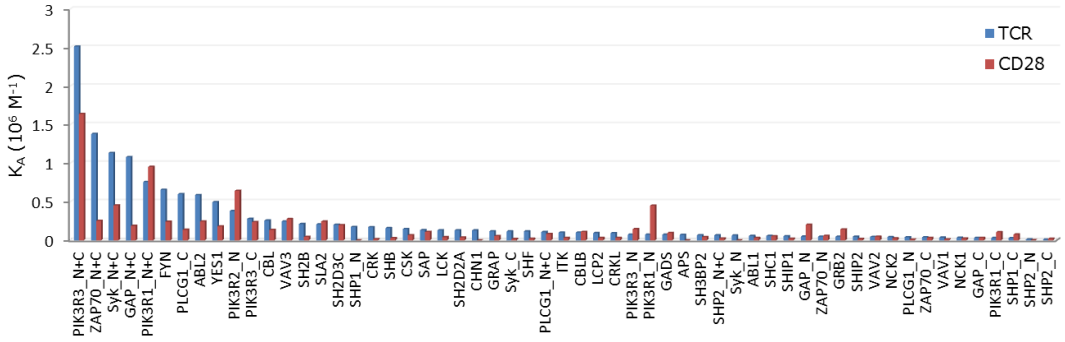


Figure 5.23. A comparison of the binding of the SH2 domains to the CD28 receptor and TCR.

5.2.6.3 CD28 and CTLA-4

Although CD28 is a stimulatory receptor and CTLA-4 is an inhibitory receptor, these two receptors bind most strongly to PIK3, albeit that the affinity for CD28 is approximately 5-fold higher than that of CTLA-4 (**Figure 5.24**). However, whereas PIK3 has highest affinity for CD28, this dominant role is not seen in the interactions with CTLA-4, since FYN, YES1 and PLCG1 also bind to CTLA-4 with comparable affinities to PIK3. Like CD28, no proteins implicated with inhibitory activities, such as SHP1/2 phosphatases, CSK [343], CBL and CBLB [344], bind to CTLA-4 with high affinity.

5.2.6.4 LAT and CD28

Distinctly different SH2 domain-binding profiles are seen for LAT and CD28 (**Figure 5.25** and **Figure 5.26**), which are adaptor and stimulatory receptors, respectively, that amplify TCR-mediated signalling. Whereas LAT is very promiscuous and capable of recruiting a large number of different SH2 domains with high affinity, the interactions of CD28 are largely limited to PIK3. Besides,

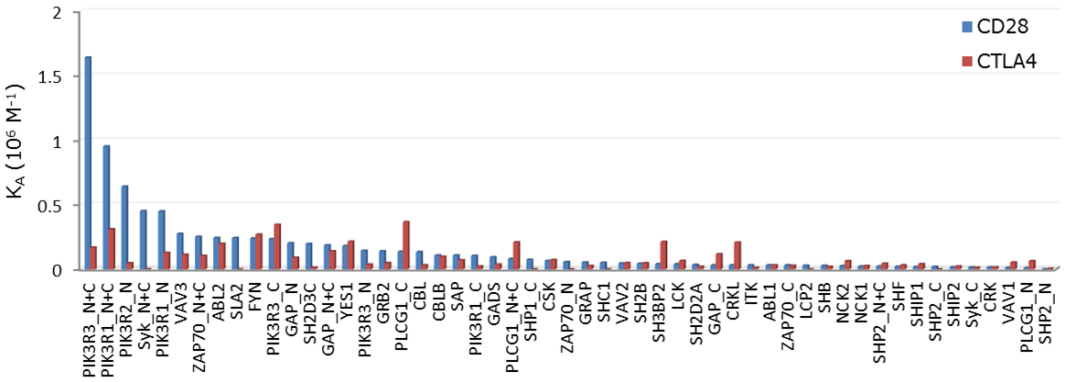


Figure 5.24. A comparison of the binding of the SH2 domains to the CD28 and CTLA-4 receptors.

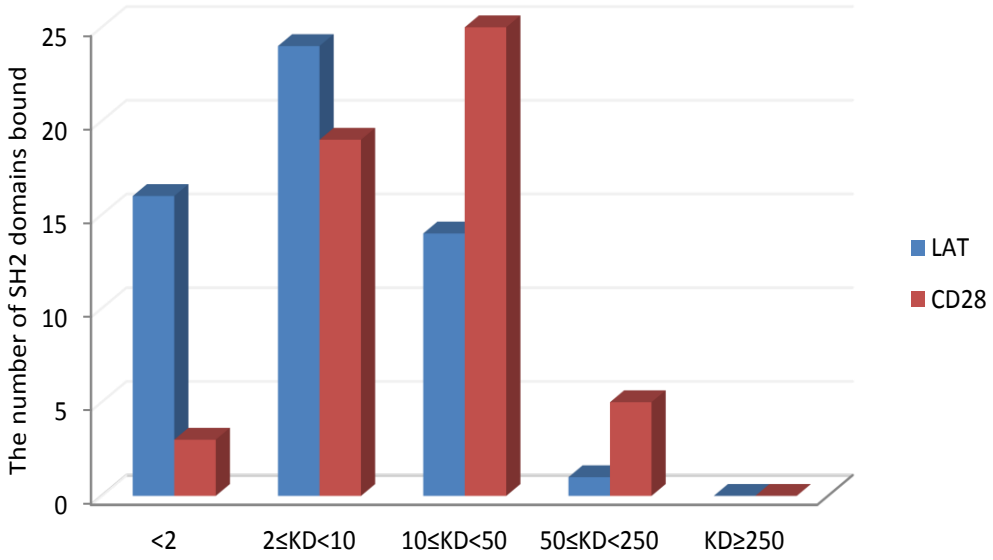


Figure 5.25. A summary of the number of SH2 domains binding to the CD28 and LAT adaptors at different affinity ranges.

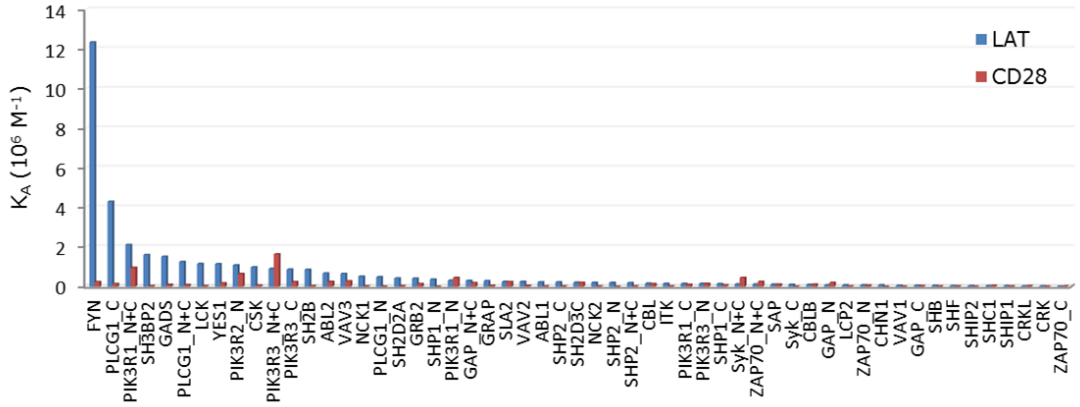


Figure 5.26. A comparison of the binding of the SH2 domains to the CD28 and LAT adaptors.

whilst PIK3 seems to have a dominant role in CD28 binding, a number of other SH2 domains have comparable affinities to PIK3 in the case of interactions with LAT. In addition, FYN and PLCG1, as well as other strong binders to LAT, have low affinities for CD28.

5.2.6.5 CTLA-4, PD-1 and BTLA

Whilst CTLA-4, PD-1 and BTLA are considered to be the three most important inhibitory receptors [345], very distinct SH2 domain binding profiles were observed in this study (**Figure 5.27** and **Figure 5.28**). Whereas BTLA is capable of mediating high-affinity interactions with a fair number of SH2 domains, only a small number is observed for PD-1, and none for CTLA-4. Both BTLA and CTLA-4 are bound strongly by PIK3 as well as PLCG1, and no proteins with phosphatase activity bind to these receptors with high affinity. In contrast, the SHP2 phosphatase, has the highest affinity for PD-1. Interestingly, all three receptors are found to be bound strongly by the two SRC kinases, FYN and YES1.

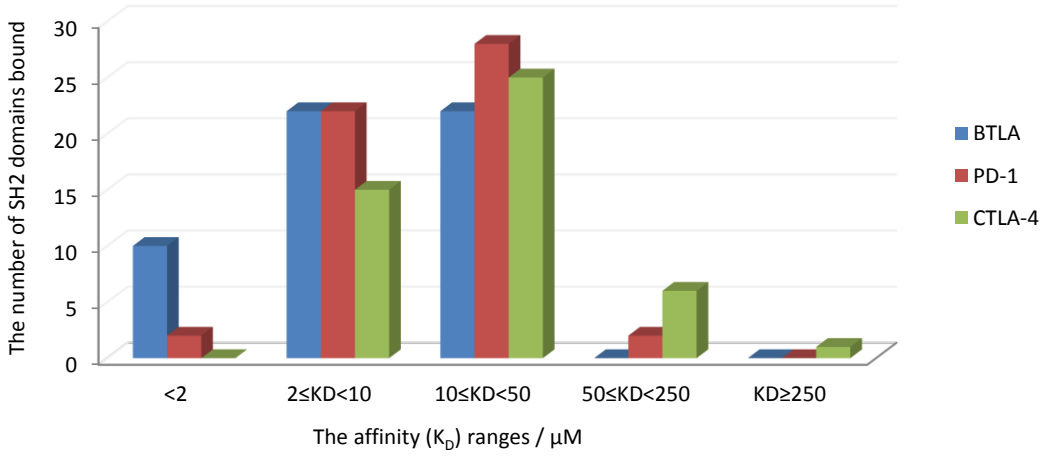


Figure 5.27. A summary of the number of SH2 domains binding to the CTLA-4, PD-1 and BTLA receptors at different affinity ranges.

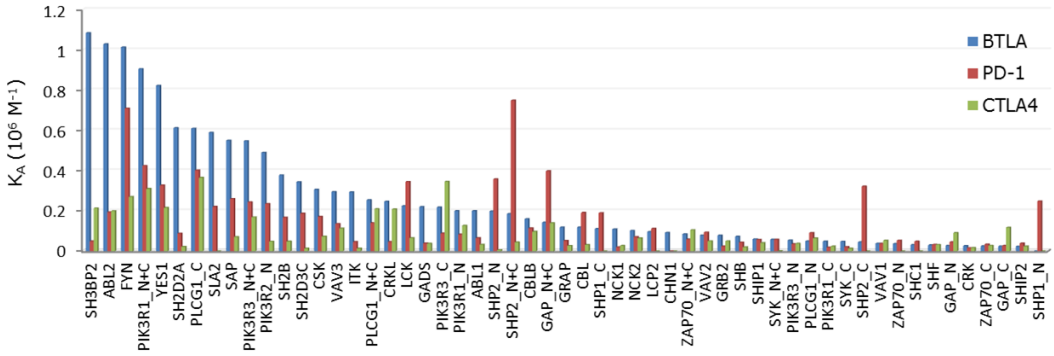


Figure 5.28. A comparison of the binding of the SH2 domains to the CTLA-4, PD-1 and BTLA receptors.

5.3 Discussion

The SPR-based assay represents, for the first time, a system-level investigation of the affinities of SH2 domains for phosphotyrosine-containing receptor motifs derived from all of the most important signalling receptors expressed by T cells.

To our surprise, we detected 974 measureable interactions between 59 SH2 domains and 20 tyrosyl phosphopeptides derived from the selected tyrosine-receptor motifs, i.e. 83% of the total number of possible interactions. Each receptor motif, on average, binds to 48.7 (±0.2) different SH2 domains. Whereas, for example, only a few SH2 proteins, including PIK3 and GRB2, were shown previously to interact with CD28, 52 potential interactions were discovered in this study. Together with the observation that these affinities vary by only two-orders of magnitude, it seems that the SH2 domains are incapable of mediating highly selective interactions.

Although SPR-based methods are considered to be among the best ways of quantifying low-affinity biomolecular interactions due to their high sensitivity, certain variation has been observed. For example, while a K_D value of 0.5 μM was previously reported for the interaction between LCP2 and the CD6 (662Y) [332], a K_D value of about 1.0 μM was determined in this study, giving a 2-fold difference. A preliminary experiment performed under the same conditions (including using the same batch of protein analyte and similar immobilisation levels of the target peptide), but using two different sensor chips, can in fact result in affinities with 3-fold differences, although the results from the same chip are highly comparable and the variation seen was analyte-dependent (data not shown). Based on this, the calculation of average affinities using replicates differing up to 3-fold is considered to be valid.

The binding assays for each tyrosyl phosphopeptides have thus far only been undertaken twice in this study, leaving a number of K_D values determined here that have more than 3-fold difference between the two replicates (**Figure 5.29**).

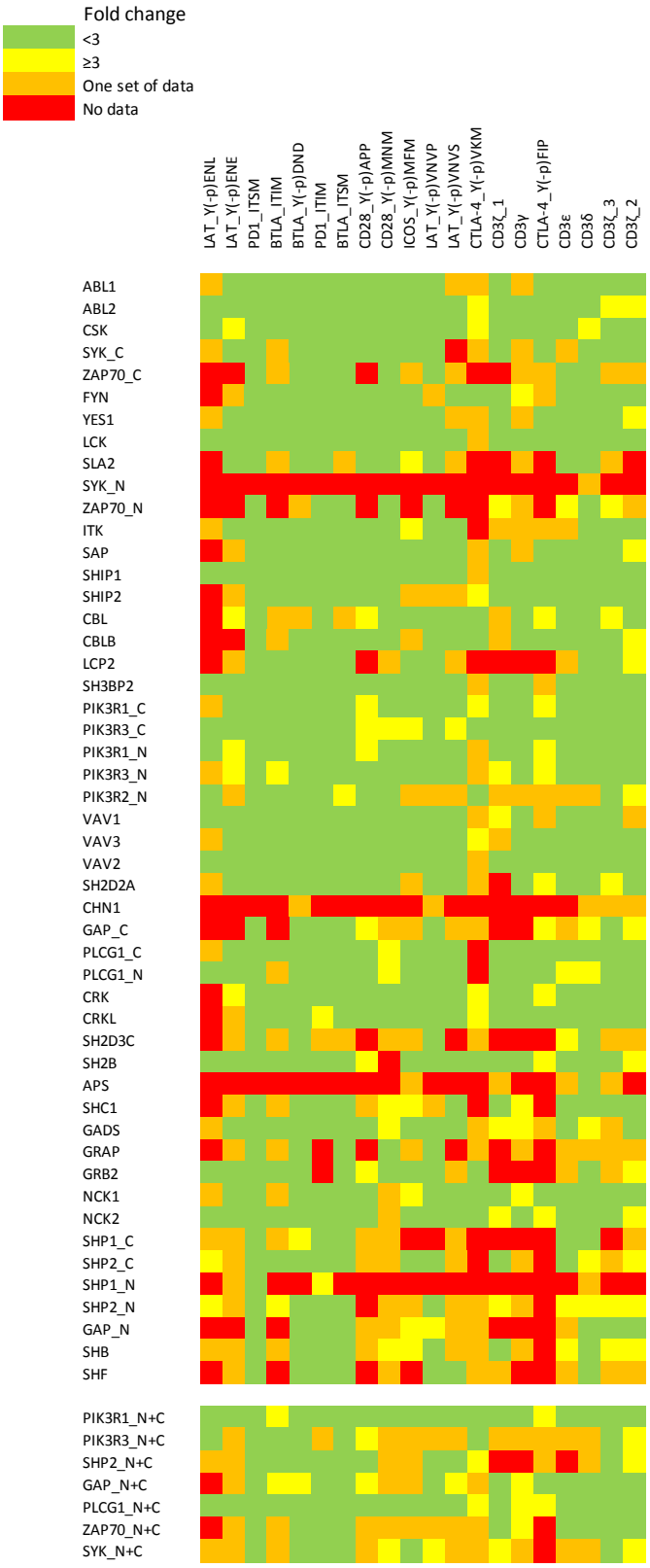


Figure 5.29. A summary of the fold-difference between the two replicates of biophysical measurements for different tyrosyl phosphopeptides. Green indicates the replicated data are within a 3-fold difference; Yellow indicates a difference of more than 3-fold; Orange indicates only one of the replicates provides useful data; Red indicates no data available.

As noticed, some of the affinity data were not available since the data was considered to be unreliable. This can be due to the extremely low affinity between the SH2 domain and the peptide, so that the data obtained resulted from noise or occasional emergence of abnormal data for unknown reasons. As a consequence, interactions with a K_D value beyond a 3-fold difference or those with missing affinity data will be repeated with a wider range of concentrations of analytes in order to validate the results.

One major issue in this study is the stability of the biotinylated tyrosyl phosphopeptides. Immobilisation of these peptides onto the sensor chips is achieved using an indirect immobilisation approach, in which the peptides were bound onto the streptavidin molecules covalently immobilised onto the sensor chip. Although the biotin-streptavidin interaction is very strong, certain fractions of biotinylated peptides dissociated over time upon regeneration, which may, to some extent, affect the affinity data obtained. In particular, several peptides failed to immobilise onto the streptavidin-coupled sensor chip, and these peptides will need to be re-made for future trials. Regeneration with NaOH/HBS-EP buffer is however inevitable since some of the SH2 proteins were found to bind irreversibly onto the peptide/sensor chip, which required the regeneration process to return background levels to zero.

Chapter 6. Analysis of specificity in a system-based model of receptor-ligand interactions[4]

6.1 Introduction

Eukaryotic cells sense extracellular signals through their surface receptors, in which case a receptor-binding event can initiate signal transduction inside the cell, eventually leading to a variety of cellular responses. To achieve the appropriate cellular response from stimulation of a given type of receptor, these signalling pathways have to be specific. The original dogma suggested that signal transduction proceeds through a linear relay of information via sequential bimolecular interactions, but support for this idea requires that the affinity for the specific interaction is at least several orders of magnitude higher than that for less specific interactions, so as to avoid cross reactions [187]. Emerging evidence has suggested that this is not the case. For example, the earliest binding event after tyrosine phosphorylation of a surface receptor upon triggering of the cell often involves interaction of SH2 domains with a phosphorylated tyrosine site of the receptor. It was shown that whereas the SH2 domain of SRC kinase has highest affinities with a tyrosyl phosphopeptide ligand containing the Y(-p)EEI motif [where Y, E and I designate different amino acids and (-p) designates phosphorylation], the affinities for these so-called "specific" motifs are only one to two orders of magnitude higher than the interactions with "non-specific" sites such as the Y(-p)VPM motif derived from the platelet-derived growth factor (PDGF) receptor and a very different motif such as Y(-p)QPG [182, 188]. In addition, a systems level investigation of interactions between SH2 domains and the tyrosyl phosphopeptides of ErbB receptors revealed that a large number of potential

[4] The mathematical modelling work described in this chapter was done in collaboration with Dr. Omer Dushek, Dunn School, University of Oxford.

cross interactions may potentially exist, with most of the affinities being within only one order of magnitude difference [191]. This suggests that the intrinsic binding properties of a single SH2 domain itself is not sufficient to support selective recognition of tyrosyl phosphopeptides, and therefore raises the question of how any signalling pathway specificity is generated at all.

There are a few potential factors that can contribute to generation of specificity in addition to affinity, including concentration and localisation of the molecules, stoichiometric properties of binding, interactions through other domains (e.g. PTB domains, SH3 domains and PH domains) as well as the assembly of multi-protein complexes. A previous study of the interactions between the T cell surface receptors CD28 and CTLA-4 with their common ligands revealed that competition effects resulting from differential expression levels can heavily impact on the binding bias of the receptor to their ligands [95]. In addition, localisation of signalling proteins close to the cell membrane can result in an increase in the effective concentration of the signalling proteins within the small volume of the area underneath the plasma membrane [346], and interactions between surface receptors and membrane-anchored signalling proteins are significantly enhanced compared to those with cytosolic proteins [347].

It is proposed that signalling molecules recruited to tyrosine-phosphorylated receptors can assemble into multi-protein complexes in a well-defined order of interaction [187]. Assembly of such complexes may also involve interactions via different modular domains, e.g. SH2 domains, PTB domains, SH3 domains and PH domains. The mutual dependency for a steric fit of the complex components thus provides high levels of specificity. Potential early signalling complexes as such have been proposed for the TCR-signalling pathway [348] and the IFN-α receptor-signalling pathway [349].

6.2 Results

With the affinity data of SH2 domains for the receptor motifs that have been previously determined (**Section 5.2.3**), as well as the concentration of SH2 domains calculated (**Section 3.2.2**) based on the very deep analysis of resting Jurkat T-cell proteome recently published [226], mathematical models will be attempted to develop in this chapter and discuss how differential SH2 domain concentrations can possibly affect the binding outcomes and specificity of the receptor.

6.2.1 Analysis of the binding specificity of SH2 domains for the receptor motifs

Affinity is defined as the free energy difference between associated and disassociated states. In order to determine whether the SH2/peptide interactions are specific or not, the Gibbs free energy [$\Delta G = -RT\ln(K_D)$; where R is the gas constant, 1.986×10^{-3} kcal K^{-1} mol^{-1}; T is the experimental temperature (37°C, 310K)] was calculated for all of the interactions whose affinities were determined in SPR-based assays, as described in the previous chapter. A summary of the Gibbs free energy calculated from each of the interactions (**Figure 6.1**) shows that only relatively few interactions have very high (≥8 kcal/mol) or very low (<6 kcal/mol) ΔG values, and most of them are within the intermediate range (between 6-8 kcal/mol). Comparing this data with the affinity data set (**Figure 5.5**) reveals good correlation between high ΔG value and strong affinity.

The distribution of ΔG values is shown in **Figure 6.2**. Statistical analysis of the distribution shows that the mean of the distribution is $< \Delta G > = 6.88$ kcal/mol, which is approximately equivalent to 11 RT (RT is the thermal energy which is the amount of energy available from the environment of the proteins] and the standard deviation is $\sigma_G = 0.83$ kcal/mol. This suggests that the

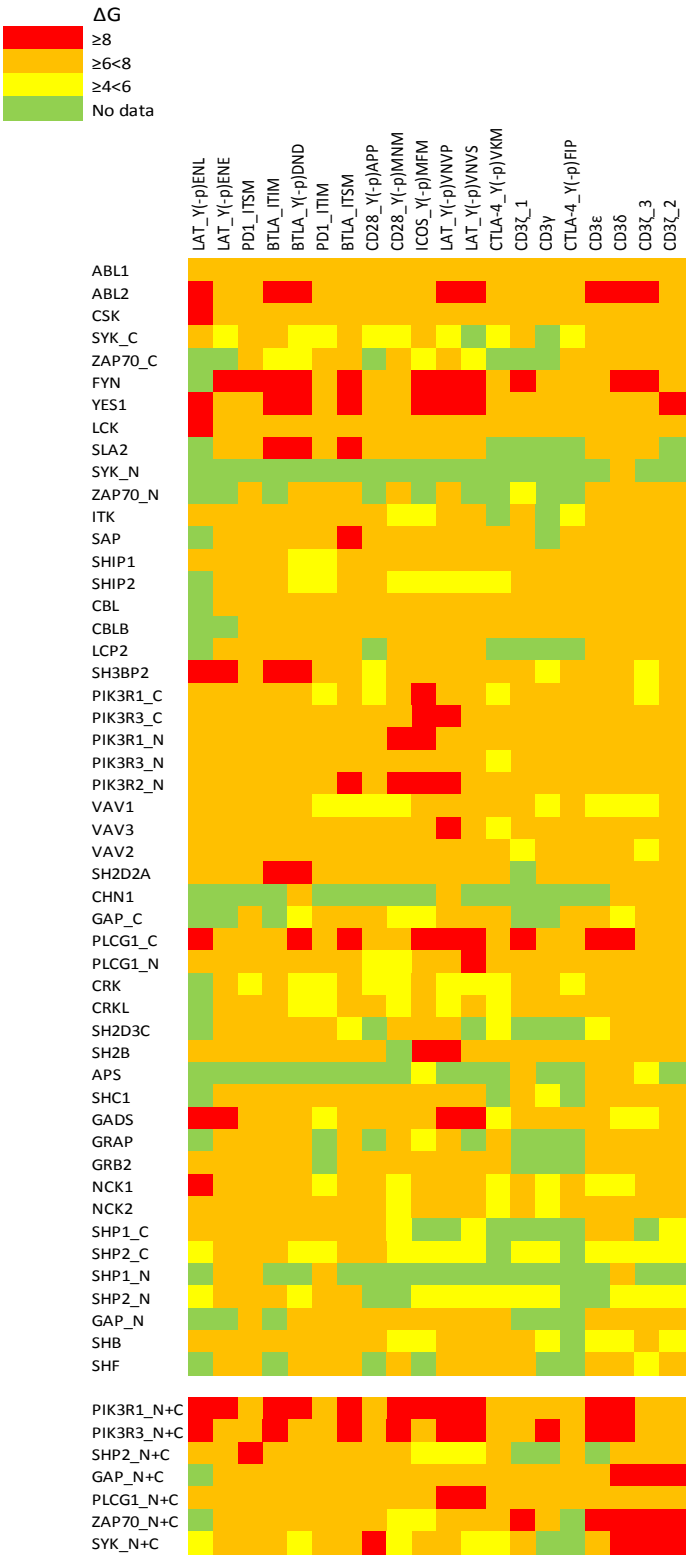


Figure 6.1. A summary of Gibbs free energy (ΔG) between SH2 domains and the tyrosyl phosphopeptides of T cell surface signalling receptors. The ΔG of the SH2 domains for the tyrosyl phosphopeptides are colour-coded according to decreasing affinity from red (ΔG ≥ 8 kcal/mol) to orange (6 kcal/mol ≤ ΔG < 8 kcal/mol), to yellow (4 kcal/mol ≤ ΔG < 6 kcal/mol). Green indicates the data for that interaction is not available.

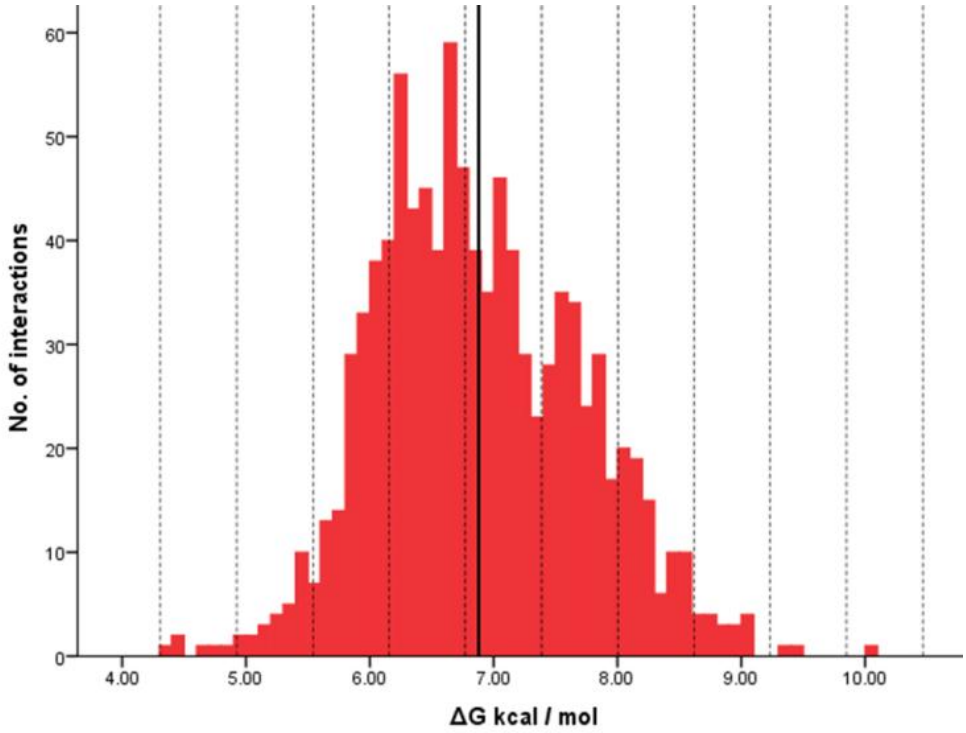


Figure 6.2. The distribution of the Gibbs free energy (ΔG) for the 974 measured interactions of the SH2 domains with the tyrosyl phosphopeptides. The vertical solid line indicates the distribution mean (6.88 kcal/mol; ~ 11RT) and the light dotted lines indicate multiples of RT ranging from 7RT to 16RT. Absolute values of ΔG are used.

measured interactions are specific and not due to background. Additionally, the distribution spans several multiples of the thermal energy scale, from 7RT to 16RT, suggesting that, whilst specific, the interactions span a large range of binding strengths.

6.2.2 Development of the model system

To access how binding affinity and ligand (SH2 domain) concentrations can affect receptor specificity, two system-based models have been developed in collaboration with Dr Omer Dushek (Pathology, Oxford). It was possible to

examine the effect of concentration because a very deep analysis of the proteome of the resting Jurkat T-cell was published by Mann and co-workers [226]. The published study was semi-quantitative and so it was necessary to transform the relative expression levels of the SH2 domains into absolute concentrations, as described in Chapter 3.

Model 1

The system contains *N* receptors (R_i , where *i* = 1…*N*) and *M* ligands (L_j , where *j* = 1…*M*) that react according to the following scheme,

$$R_i + L_j \overset{K_D^{i,j}}{\Leftrightarrow} C_{i,j} \tag{1}$$

where $C_{i,j}$ is the complex formed by the *i th* peptide and the *j th* ligand. The interaction is governed by the experimentally measured dissociation constant ($K_D^{i,j}$).

At equilibrium the following equation can be defined,

$$R_i L_j = K_D^{i,j} C_{i,j} \tag{2}$$

where R_i and L_j are the concentrations of the free receptor and ligand, respectively.

Assuming both the total amount of receptors and ligands are conserved, the following equations can be defined,

$$R_T^i A = R_i\, A + \sum_{j=1}^{j=M} C_{i,j} A \tag{3}$$

$$L_T^i V = L_i V + \sum_{i=1}^{i=N} C_{i,j} A \quad (4)$$

where R_T^i is the total concentration of receptor (*i)*, L_T^j is the total concentration of ligand (*j)*, *A* is the cell surface area, and *V* is the cell volume.

Given $C_{i,j}$, the probability that ligand (*j)* will be bound to receptor (*i)* can be defined as follows,

$$P_{i,j} = \frac{C_{i,j}}{R_T^i - R_i} \quad (5)$$

In Model 1, it is assumed that (1) all ligands have identical concentrations, (2) the ligand concentrations are in excess of the receptor concentrations so that there is no competition between receptors for ligands, and (3) all ligands are cytosolic. This means that the affinity of ligand for the receptors is the only determining factor for the binding outcome in this case. Therefore, an explicit expression for the probability ($P_{i,j}$) that ligand (*j*) will be bound to receptor (*i*) can be defined as follows,

$$P_{i,j} = \frac{K_A^{i,j}}{\sum_{j=1}^{j=M} K_A^{i,j}} \quad (6)$$

Where $K_A^{i,j}$ is the association constant ($K_A^{i,j}$ =1/ $K_D^{i,j}$). The binding probability spectrum in Model 1 should therefore resemble that of the binding affinity spectrum.

Since there is no universally accepted measure of specificity, a novel measure of specificity based on the $P_{i,j}$ spectrum was developed (by Dr Dushek) and applied in this study, which involves measuring specificity in terms of entropy (δ).

The term "specificity" is used loosely to refer to the phenomenon that some proteins exhibit restricted binding to a few substrates (high specificity) whereas others exhibit more promiscuous binding to many substrates (low specificity). In contrast to other well-established concepts (e.g. affinity), there is presently no mathematical definition of specificity. The majority of published studies aimed at studying specificity have been qualitative [350] with one exception [351]. Moreover, all studies only considered binding affinity as a contributor to specificity. Through an iterative approach Dr Dushek has proposed that the 'redundancy measure of entropy' is a mathematical formula that captures our intuition of the concept of specificity that is sufficiently general to incorporate any contribution to specificity (Dushek at al, unpublished). We therefore now define specificity as

$$\delta_i = 1 + \frac{\left(\sum_{j=1}^{j=M} P_{i,j} \log_2(P_{i,j})\right)}{\log_2(M)} \tag{7}$$

which is derived from

$$\delta_i = 1 - \left(- \frac{\left(\sum_{j=1}^{j=M} P_{i,j} \log_2(P_{i,j})\right)}{\log_2(M)} \right)$$

Here, $\log_2(M)$ functions as a normalising factor and divides the entropy value by $\log_2(M)$, normalizing entropy into a value within the range between 0 and 1. This is because when the ligands are binding to a given receptor with equal probabilities, where P_i = 1/M, maximum entropy $\log_2(M)$ can be reached. However, an entropy value of 1 would imply that there is no specificity (the binding probabilities of all the ligands are equal). To allow high specificity to be expressed as a high value (close to 1), and vice versa, the entropy value is therefore subtracted from 1.

Model 2

In Model 2, the differential concentrations of SH2 domains expressed in Jurkat were taken into account (therefore the ligand concentrations are no not assumed to be identical here in this model), keeping unchanged the assumptions that the ligand concentrations are in excess of the receptor concentrations and all ligands are cytosolic. The probabilities for this model can be defined as follows,

$$P_{i,j} = \frac{K_A^j \, L_T^j}{\sum_{j=1}^{j=M} K_A^j L_T^j} \tag{9}$$

where the $P_{i,j}$ is now determined by the receptor-ligand affinity, and also by the ligand concentrations. The specificity for receptor (*i*), defined entropically, in this model can be determined using equation (8).

6.2.3 Effects of affinity and concentration on the probabilities of SH2 domain binding to phosphorylated tyrosine motifs

Using the mathematics described above, and the affinities determined in the previous chapter, the probabilities for the SH2 domains binding to the selected tyrosyl phosphopeptides have been calculated for both Model 1 (**Figure 6.3**) and Model 2 (**Figure 6.4**). Comparison between the two figures shows that whereas a number of SH2 domains, mainly the double domains, have binding probability higher than 0.1 when only binding affinities were considered (Model 1), only certain SH2 domains (LCK, GADS, SAP, GRB2, CRKL and ZAP70) have such high binding probabilities when differential ligand concentrations were considered (Model 2). Most notably, LCK has high binding probabilities for all the receptor motifs. In addition, taking the concentration effect into account markedly reduces the number of interactions with an intermediate binding probability (between

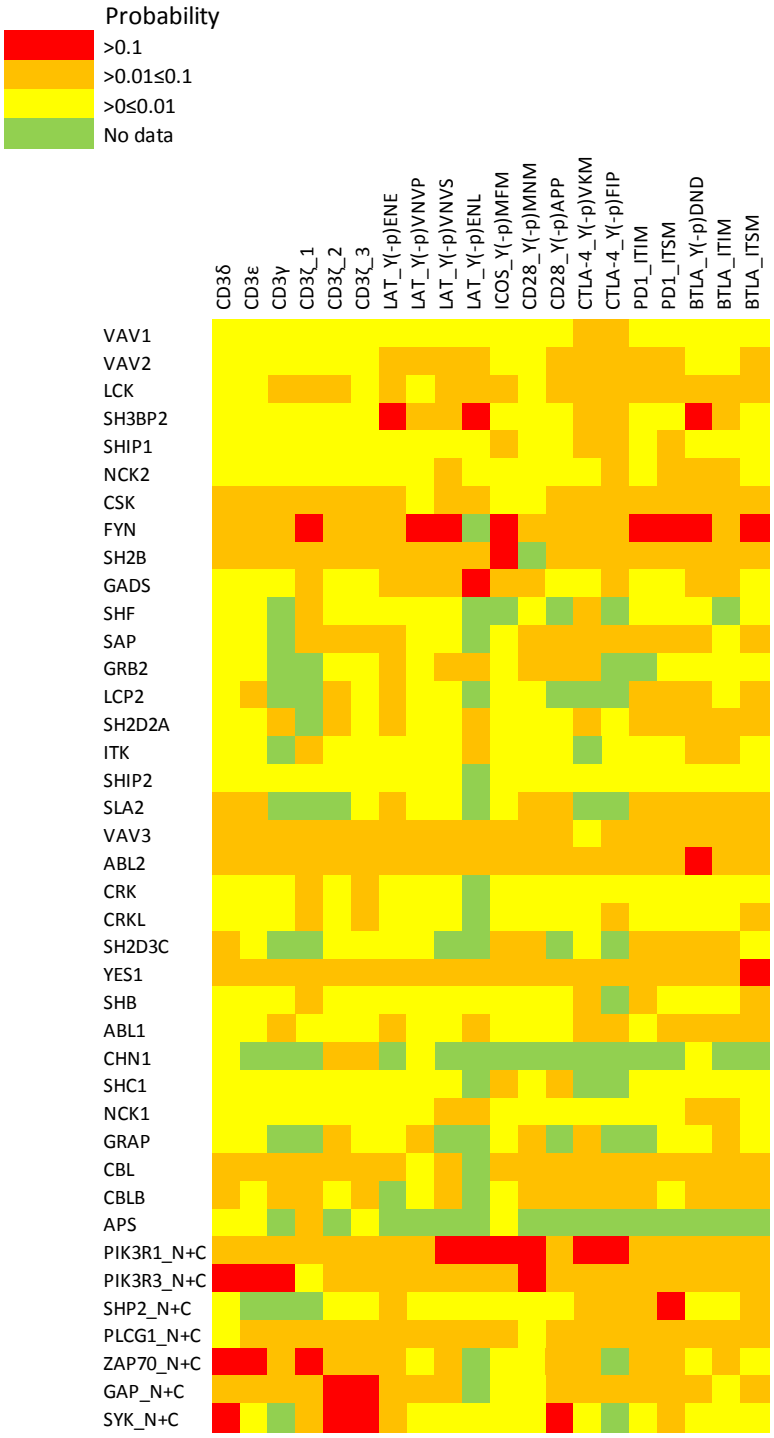


Figure 6.3. A summary of the SH2/receptor motif binding probabilities ($P_{i,j}$) when only affinity is considered (Model 1). The different colours indicate different ranges of binding probabilities: red: $P_{i,j}$ >0.1; orange: 0.01< $P_{i,j}$ ≤0.1; yellow: 0< $P_{i,j}$ ≤0.01; green: no data available.

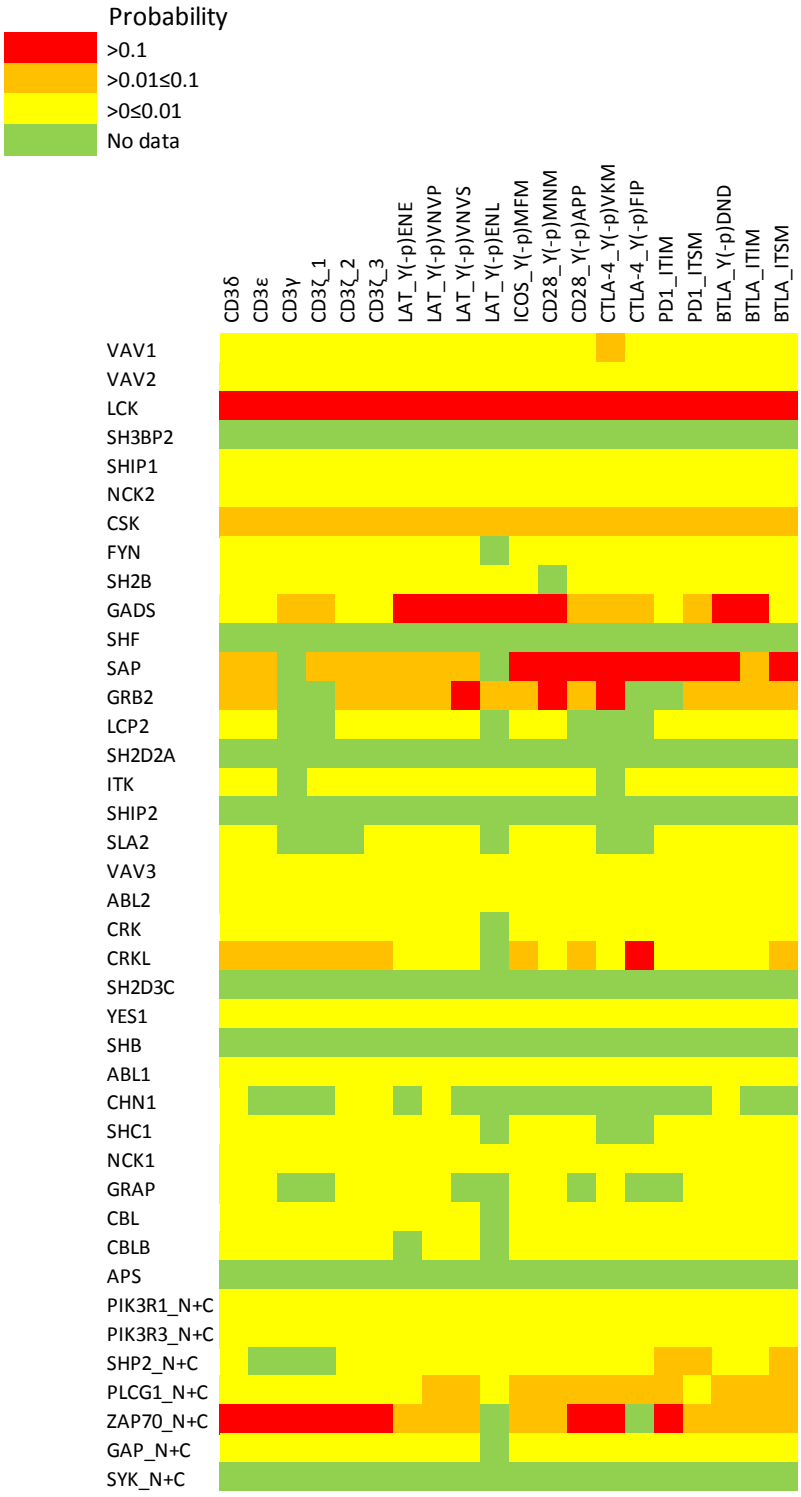


Figure 6.4. A summary of the SH2/receptor motif binding probabilities ($P_{i,j}$) when the differential concentrations of the SH2 domains are considered in addition to affinity (Model 2). The different colours indicate different ranges of binding probabilities: red: $P_{i,j}$ >0.1; orange: 0.01< $P_{i,j}$ ≤0.1; yellow: 0< $P_{i,j}$ ≤0.01; green: no data available. No data is available for SH3BP2, SHF, SH2D2A, SHIP2, SH2D3C, SHB, APS and SYK owing to the absence of proteomic data for these proteins.

0.01 and 0.1), and instead with more interactions with low binding probability (≤0.01).

Comparison of the number of interactions in different ranges of binding probability ($P_{i,j}$; with a scale of 0 to 1) between Model 1 (when affinity alone was considered) and Model 2 (when concentrations were also considered) (**Figure 6.5**) shows that adding the concentrations results in increased number of interactions with higher binding probabilities, but fewer interactions with intermediate $P_{i,j}$ values overall. Model 2 has a larger number of interactions with a binding probability greater than 0.1 (50 versus 28) and smaller than 0.01 (450 versus 232), whereas Model 1 has a greater number of interactions of intermediate probability between 0.01 and 0.1. Also, whilst only two interactions were found to have a binding probability higher than 0.4 in Model 1, 17 such interactions are present in Model 2, with the highest one being 0.78. Moreover, whilst interactions in Model 1 are all with a binding probability higher than 0.0001, 228 interactions were found to below this value in Model 2. These observations suggest that concentration tends to polarise the binding probabilities, presumably because certain SH2 domains out compete others, increasing receptor specificity.

Examples of the probability spectrum for the SH2 domain/tyrosyl phosphopeptides interactions are shown in **Figure 6.6**. Surprisingly, certain SH2 domains, notably LCK and ZAP70, and to a less extent SAP, have become the dominant binders and the binding probabilities for most other candidates were significantly reduced when ligand concentrations were taken into account, reflecting the particularly high abundance of these molecules [**Figure 3.10 (A)**]. The very high expression of LCK is consistent with this molecule dominating the binding of four of the seven receptor motifs [LAT_Y(-p)ENE, ICOS_Y(-p)MFM, PD-1_ITSM and BTLA_ITSM].

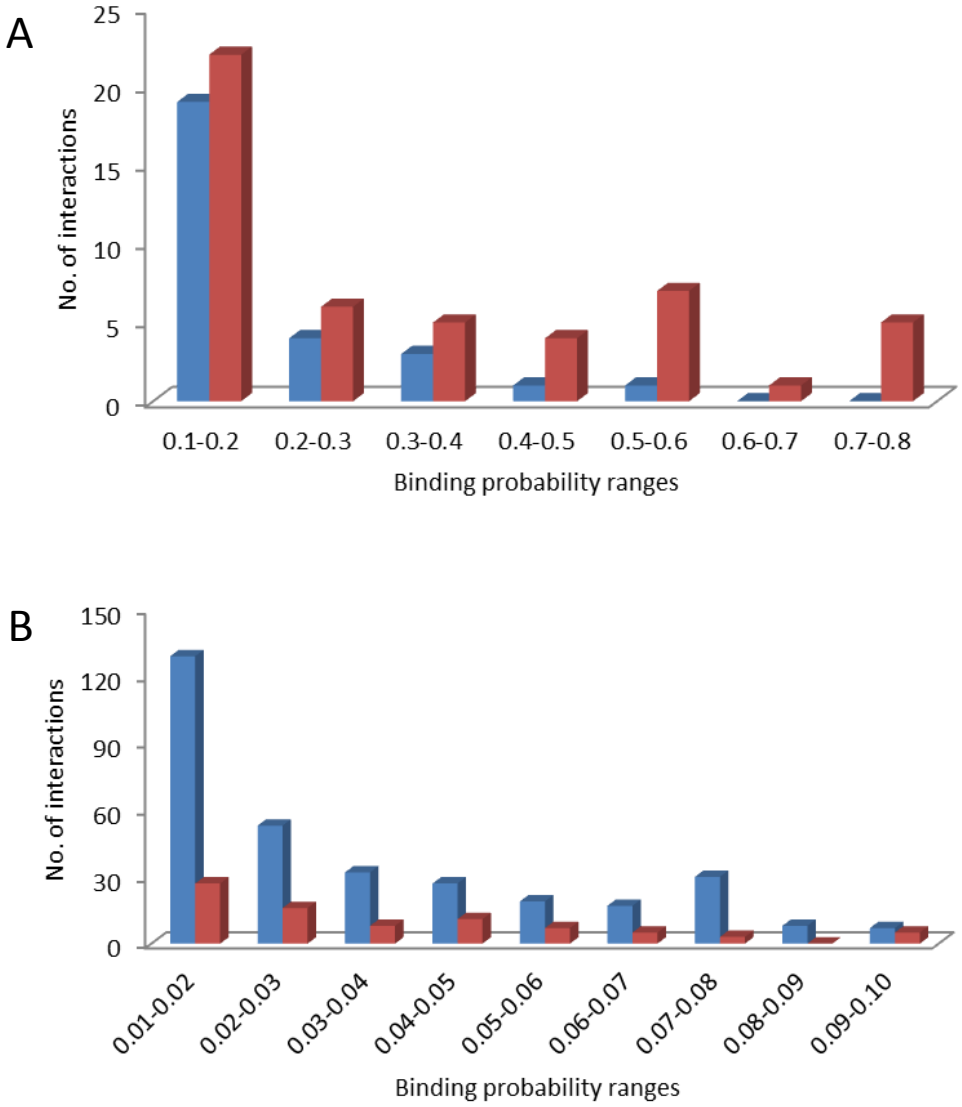


Figure 6.5. A comparison of the number of interactions between SH2 domains and the 20 different receptor motifs in different binding probability ranges between Model 1 and Model 2. The blue and red bars represent the numbers in Model 1 (when affinity alone is considered) and Model 2 (when differential ligand concentrations are also taken into account), respectively.

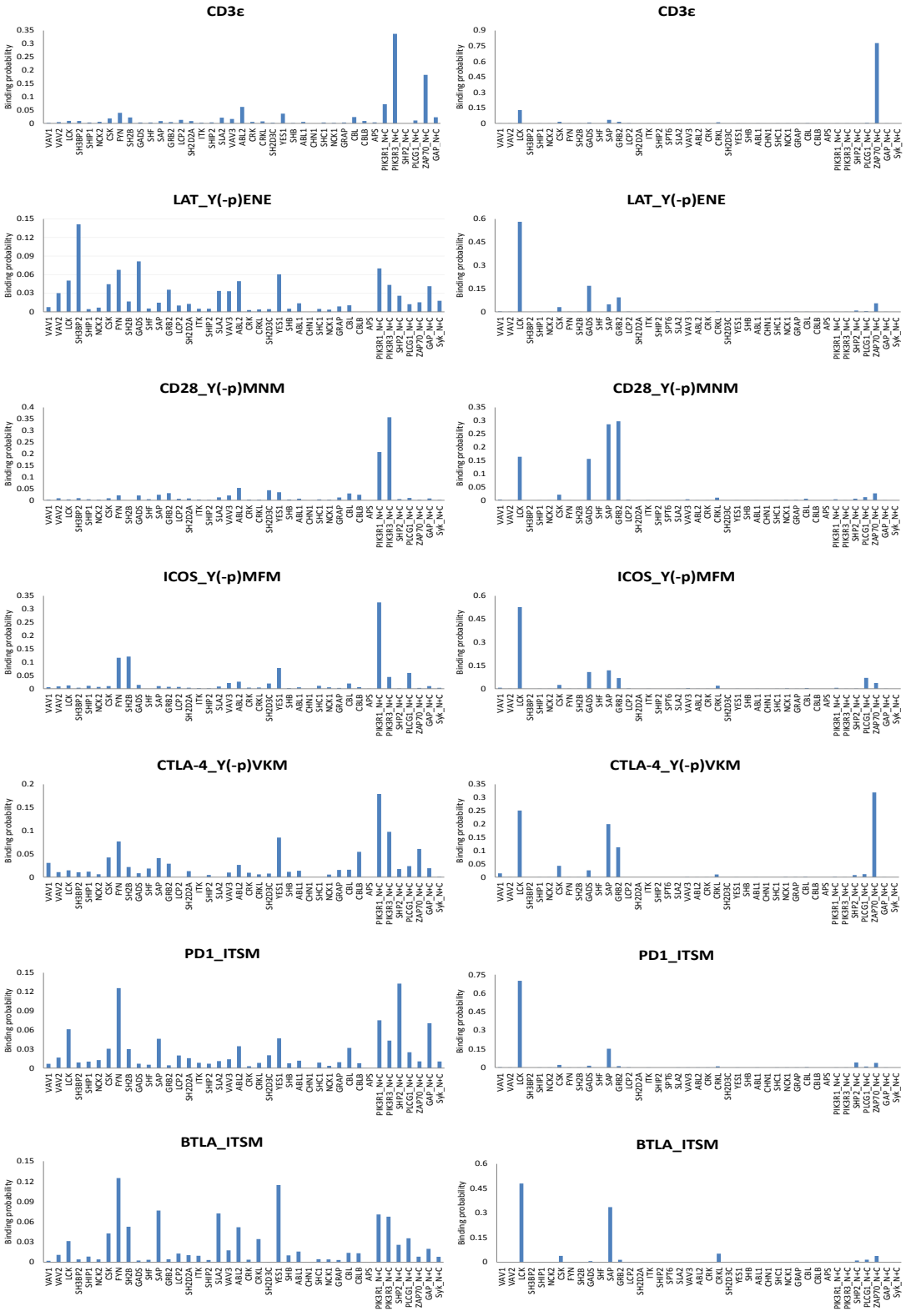


Figure 6.6. Probability spectra for SH2 domain binding to example receptor motifs according to Models 1 (left panel) and 2 (right panel). In both models, the ligand concentrations are assumed to be in excess of the receptor so that there is no competition between receptors. Note that for each receptor motif the bar graphs add up to 1.0 and that the y-axis scales differ.

When only affinity is considered, binding probabilities are generally similar for each receptor motif in most cases, and hardly any extremely high probabilities are seen. However, the promiscuity of these receptor motifs, notably LAT_Y(-p)ENE, CLTA4_Y(-p)VKM, PD-1_ITSM and BTLA_ITSM, are diminished when ligand concentrations were considered. Taking the binding to the PD-1_ITSM motif as an example, all the binding probabilities for all the interactions are below 0.15 when only binding affinity was considered (Model 1), whereas LCK was found to have a distinct dominance with a binding probability of 0.7 when ligand concentrations were considered (Model 2). In addition, whilst the binding probability of the top 23 binders (which accounts for 60% of the total amount of interactions) to this motif are within a 10-fold difference in Model 1, the difference between LCK and SHP2 (the third strongest binder) in Model 2 is about 17-fold.

In effect, LCK has a 330-fold higher binding probability than VAV1, which ranks as the 10^{th} strongest binder to the PD-1_ITSM in Model 2, when ligand concentrations were considered. This is owing to the fact that LCK not only has 33-fold higher concentration than VAV1, but also a 10-fold higher affinity.

Either LCK or ZAP70 dominates the binding in most cases, the only exception being CD28_Y(-p)MNM, in which case SAP and GRB2 dominate ($P_{i,j}$ ~ 0.3), although LCK and GADS have similar but somewhat lower binding probabilities ($P_{i,j}$ ~ 0.15) when concentration was taken into account. This is due to the fact that the concentration of LCK is about 3 to 6-fold higher than SAP, GRB2 and GADS, whilst its affinity to this receptor motif is ~1/5 of these molecules, conferring upon them similar receptor-binding probabilities. Although the PIK3 SH2 domains seem to be the exclusive dominant binders when only affinity is considered, the binding probabilities of PIK3 SH2 domains become negligible due to their extremely low concentrations in the cell, which is

approximately 80000-fold lower than LCK. Whilst their affinities are only about 100-fold higher than LCK, this still leads to an 800-fold lower binding probability for the PIK3 SH2 domains compared to LCK. In effect, whilst the PIK3 domains are generally observed to play a dominant role in binding to all the receptor domains in Model 1, their effects are very limited in Model 2 where concentration is taken into account.

6.2.4 Comparisons of the SH2 binding probabilities for related receptor motifs in Models 1 and 2

As discussed in Chapter 5, when the binding affinities measured in SPR-based assays were compared, notable differences in binding profile were observed between receptor motifs derived from different receptors, or even from the same receptor. Comparisons of the probability spectrums for binding of SH2 domains to some of the representative motifs from different receptors now reveals that simulating the concentration effect dramatically changes the binding profiles by greatly reducing the promiscuity of binding and reducing the differences between motifs and receptors, due to the dominance of certain SH2 domains such as LCK and ZAP70. In the following section, the binding probabilities of SH2 domains for phosphotyrosine-containing motifs derived from related receptors were compared.

6.2.4.1 The CD3 ITAM motifs of the TCR

Whereas the ITAM motif of CD3ζ_1 is the only one that recruits ZAP70 with exceptionally high probability when only affinity was considered (Model 1), ZAP70 dominates the binding to all ITAM motifs when SH2 domain concentration was taken into account (Model 2), with a binding probability over 0.5 in all cases, and over 0.7 to CD3δ, CD3ε and CD3ζ_1 (**Figure 6.7**). Although PIK3, GAP and SYK

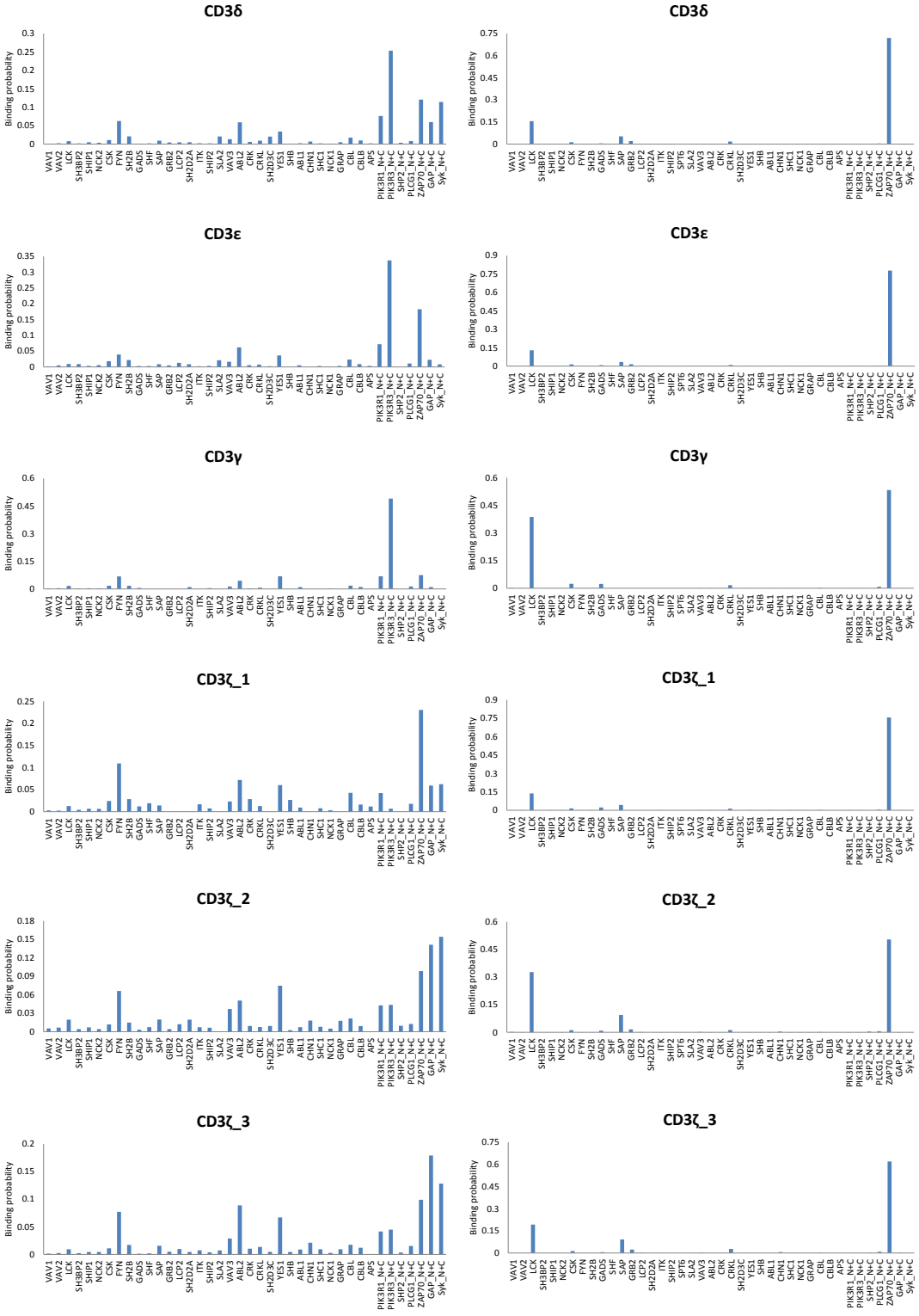


Figure 6.7. Probability spectra for SH2 domain binding to the CD3 ITAM motifs for Models 1 (left panel) and 2 (right panel).

are among the strongest binders to the CD3 ITAMs in Model 1, they are out-competed by ZAP70, followed by LCK, in Model2. The reason why these molecules have become insignificant in the probability spectrum is because of their extremely low abundance, with PIK3 and GAP being about 20000-fold lower than ZAP70, while SYK is undetectable. The comparable binding probabilities between LCK and ZAP70 in the case of binding to the CD3γ ($P_{i,j\,\mathrm{LCK}}$ ~ 0.4; $P_{i,j\,\mathrm{ZAP70}}$ ~ 0.5) and CD3ζ_2 ($P_{i,j\,\mathrm{LCK}}$ ~ 0.3; $P_{i,j\,\mathrm{ZAP70}}$ ~ 0.5) arise because the affinities of LCK to these motifs are approximately 5-fold lower than ZAP70, but its protein expression is about 3-fold higher. In other cases (binding of CD3δ and ε, as well as CD3ζ_1 and CD3ζ_3), where the affinities of ZAP70 are more than 10-fold higher than LCK, larger difference in binding probabilities are observed. Overall, the fact that ZAP70 is among the top three binders to the ITAM motifs in terms of affinity, and that it has the second highest protein concentration, allows this molecule to dominate the binding to the CD3 ITAM motifs.

6.2.4.2 Activating receptors: LAT, ICOS and CD28

The common feature of the probability spectrum for the LAT adaptor motifs, when differential ligand concentrations were considered (Model 2), is the dominant binding of LCK and relatively high binding probabilities for GADS and GRB2 (**Figure 6.8**). Although LCK does not have the highest affinity, its exceptionally high protein concentration allows it to out-compete others. For example, although the affinity of LCK to these receptor motifs can be up to approximately 50-fold weaker than FYN, the concentration of LCK is however about 60000-fold higher, thus making LCK 1200 times more competitive than FYN. The particularly high binding probability of LCK to LAT_Y(-p)ENE and LAT_Y(-p)ENL is due to the fact that LCK is among the strongest binders to these receptor motifs. Whereas LCK is also the one with highest binding probability for

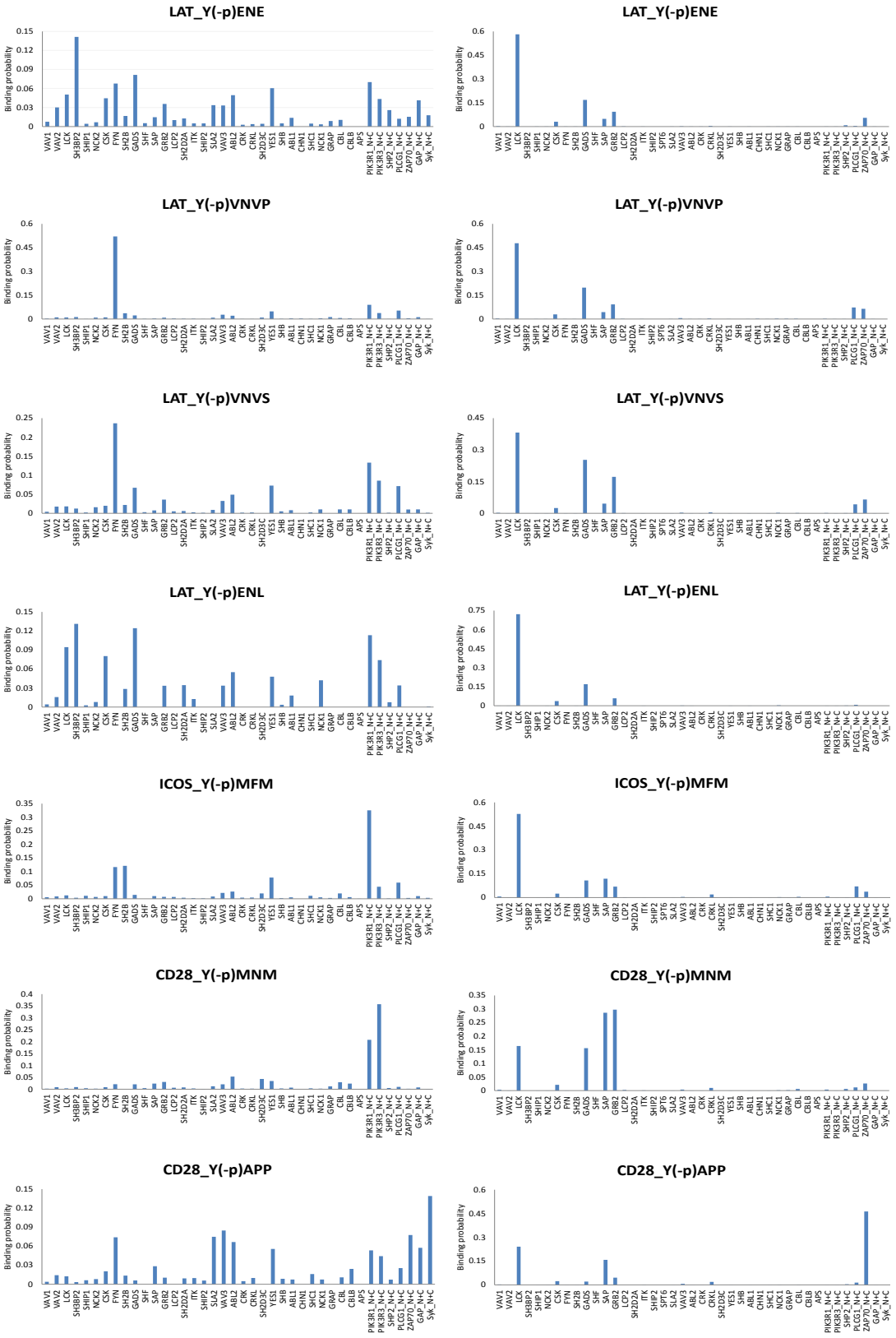


Figure 6.8. Probability spectra for SH2 domain binding to the tyrosyl phosphopeptides of LAT, ICOS and CD28 in Models 1 (left panel) and 2 (right panel).

LAT_Y(-p)VNVP and LAT_Y(-p)VNVS, its effect is less profound compared to the binding to the other two LAT motifs, due to the high binding probabilities of PLCG1, GADS and GRB2 whose affinities are similar to or higher that that of LCK.

Interestingly, although ICOS_Y(-p)MFM has a very different binding profile from the LAT motifs when only affinity is considered, they are highly similar when SH2 domain concentration is allowed for, with LCK being the dominant binder, followed by PLCG1, GADS and GRB2. This is because, although binding is dominated by PIK3 SH2 domains followed by FYN, SH2B, YES1 and GAP in Model 1, the extremely low concentrations of these proteins (>2000-fold lower than LCK) make their binding become trivial in Model 2. On the other hand, although molecules such as LCK, PLCG1, GADS, SAP and GRB2 only have moderate binding affinity, their very high protein concentrations allow them to dominate binding.

Although the Model 1 binding profiles of CD28_Y(-p)MNM and ICOS_Y(-p)MFM are highly similar, for example in terms of their binding to the PIK3 SH2 domains, notable differences are seen in Model 2. Whereas LCK is the dominant binder to ICOS_Y(-p)MFM, it is replaced by SAP and GRB2 in the case of binding to CD28_Y(-p)MNM. This is due to the 2-3 fold higher affinities of these molecules for CD28_Y(-p)MNM compared to ICOS_Y(-p)MFM, and because LCK binds 4-fold more weakly to CD28_Y(-p)MNM than to ICOS_Y(-p)MFM .

The binding profile to the CD28_Y(-p)APP in Model 1 is more similar to the LAT motifs, in particular the LAT_Y(-p)ENE motif due to the promiscuity of these motifs and the absence of dominance by the PIK3 SH2 domains. However, when ligand concentrations are taken into account, whereas the LAT adaptor motifs bind mostly to LCK, ZAP70 has the highest binding probability for CD28_Y(-p)APP. This is because the affinity of ZAP70 is particularly high to this motif (6.2-fold higher than LCK). One feature that makes the CD28_Y(-p)APP similar to the CD28_Y(-p)MNM motif is the relatively high binding probability of SAP, owing to

the merely 3-fold lower affinity of this molecule compared to ZAP70, and similar protein concentrations.

6.2.4.3 Inhibitory receptors: CTLA-4, PD-1 and BTLA

The common feature of the binding profiles for the inhibitory receptor motifs, when only affinity was considered (Model 1), is the prevalence of a high-degree promiscuity and the absence of high probability binding (**Figure 6.9**). However, once again this promiscuity is diminished when ligand concentrations were considered (Model 2), and LCK has become the dominant binder almost in all scenarios. However, this dominance is less profound for CTLA-4_Y(-p)VKM and BTLA_ITSM, where SAP has comparable binding probabilities to LCK to these twov receptor motifs in Model 2 (CTLA-4_Y(-p)VKM: $P_{i,j\ \mathrm{LCK}}$ = 0.25, $P_{i,j\ \mathrm{SAP}}$ = 0.2; BTLA_ITSM: $P_{i,j\ \mathrm{LCK}}$ = 0.48; $P_{i,j\ \mathrm{SAP}}$ = 0.34). Interestingly, the binding profile of the CTLA-4_Y(-p)VKM motif differs from other inhibitory receptor motifs, but resembles that of the CD28_Y(-p)APP motif, as characterised by the promiscuity to SH2 domains when only affinity was considered and outstanding dominance of binding by ZAP70, LCK, SAP and GRB2 when ligand concentrations were taken into account. One notable observation is that whilst the high binding probabilities of FYN and SHP2 to both ITIM and ITSM motifs of PD-1 are among the highest in Model 1, they are greatly reduced in Model 2. This is because, in the case of SHP2 for example, its affinity is only 2 to 4-fold higher than the dominant binder LCK, whilst its concentration is 38-fold lower, resulting in an approximately 10-20 fold lower binding probability overall.

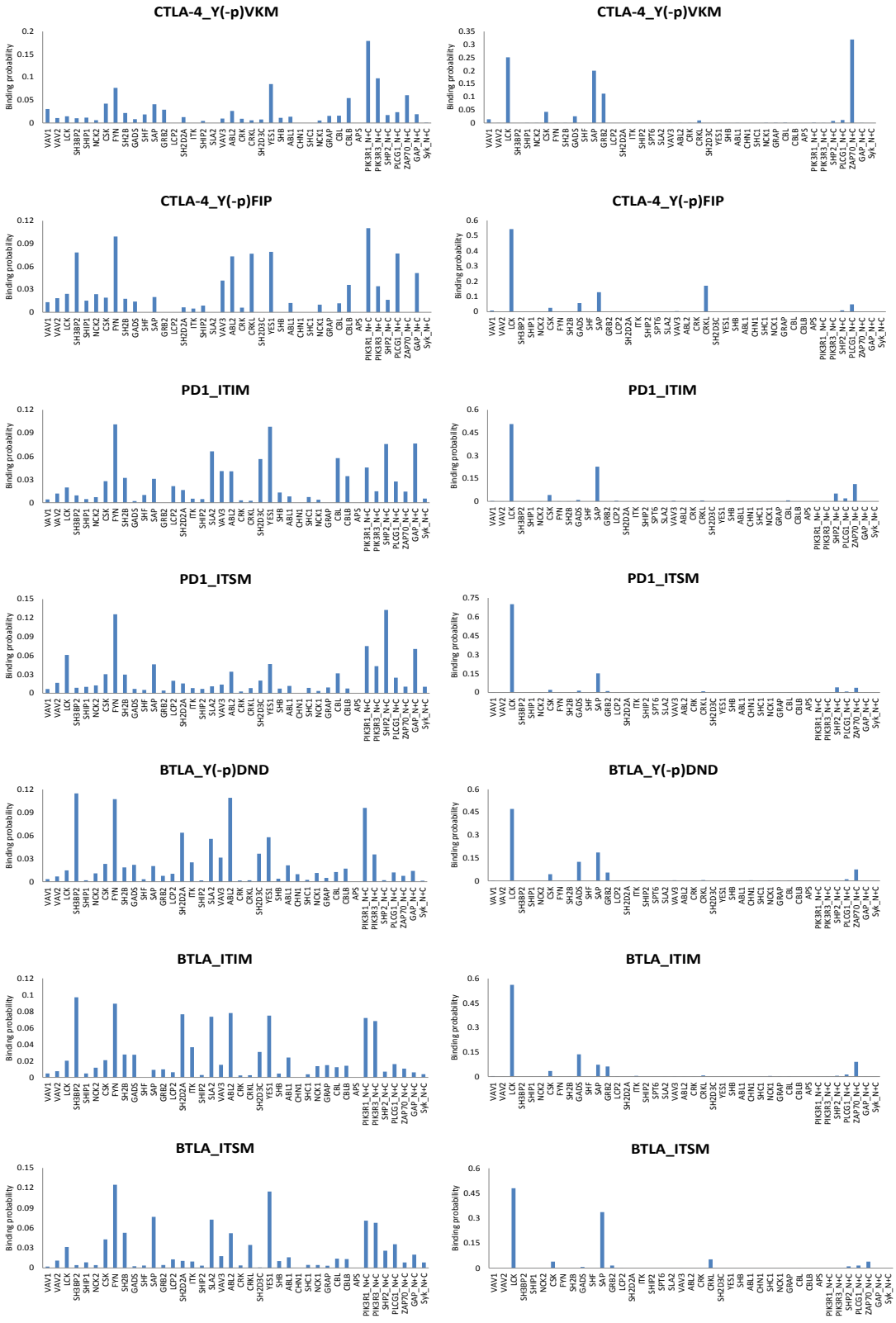


Figure 6.9. Probability spectra for SH2 domain binding to the tyrosyl phosphopeptides of CTLA, PD-1 and BTLA in Models 1 (left panel) and 2 (right panel).

6.2.4.4 The analysis of receptor specificity

Whereas the above section reveals that simulating concentration effects reduces apparent promiscuity and that certain SH2 domains, notably LCK and ZAP70, have greater binding probabilities compared to most others, this does not provide a quantitative measure of overall receptor specificity. Since there is no universally accepted measure of specificity, a novel application of the concept of entropy (δ) was developed for this purpose by Dr. Omer Dushek. Simply stated, higher specificity is observed for more "ordered" receptors that bind one or a few ligands with high binding probabilities whereas less ordered (more "random") receptors that bind many ligands with similar probabilities have low specificity. It was found that, using this new definition, specificity derived from affinity itself (Model 1) is generally low [**Figure 6.10 (A)**], with over half of the receptor motifs having a specificity below 0.2 (on a scale of between 0 and 1). LAT_Y(-p)VNVP has the highest specificity (δ=0.43), followed by the ITAM motifs of CD3 δ, ε, γ, as well as the stimulatory receptor motifs CD28_Y(-p)MNM and ICOS_Y(-p)MFM. The seemingly high specificity of LAT_Y(-p)VNVP is consistent with the particularly high binding probability of FYN to this receptor. The specificity for the inhibitory receptor motifs derived from CTLA-4, PD-1 and BTLA are all below 0.2, indicating that these receptors are likely to be non-specific in the case of there being equal concentrations of the ligands.

When differential ligand concentrations were simulated in the systems model (Model 2), increased specificity as defined by the entropic measure is observed for each receptor motif without exception [**Figure 6.10 (B and C)**]. Apart from the CD3γ ITAM motif, LAT_Y(-p)VNVP, CD28_Y(-p)MNM and ICOS_Y(-p)MFM motifs, which have relatively high specificity even when affinity alone is considered, more than 2-fold increases were observed for all other receptor motifs. The most significant increases, of ~5-fold, were seen for LAT_Y(-p)ENL, CTLA-4_Y(-p)FIP and PD-1_ITSM, all of which had the lowest specificity in

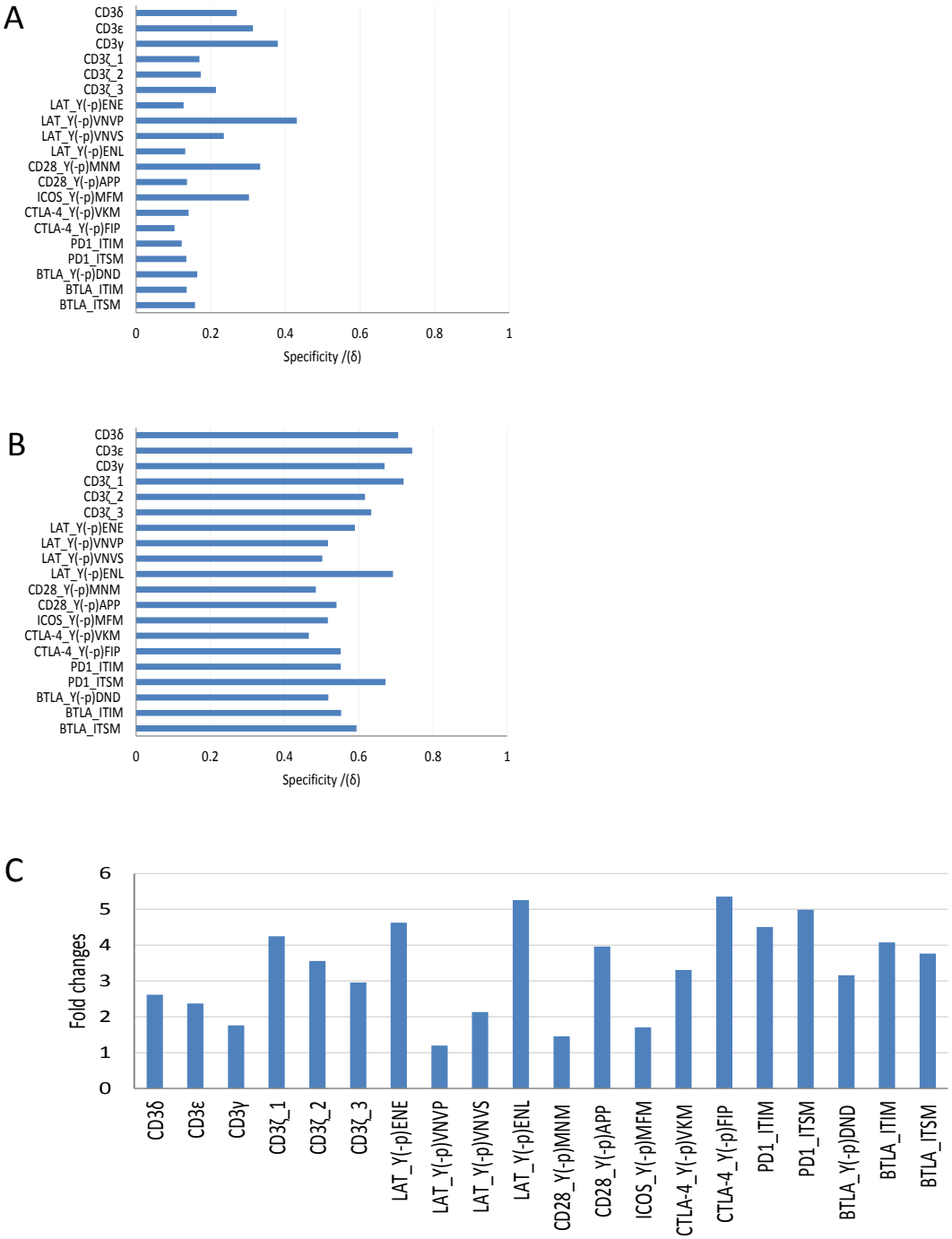


Figure 6.10. Receptor specificity measured using the redundancy measure of entropy (δ). (A) Receptor specificity according to Model 1. (B) Receptor specificity for Model 2. (C) The fold change in receptor specificity between Model 1 and 2.

Model 1. There is a >3-fold increase in specificity for all the inhibitory receptor motifs derived from CTLA-4, PD-1 and BTLA. These increases are consistent with the marked increase in binding probabilities as a result of the outstanding difference in concentrations between LCK/ZAP70 and the other SH2 domains.

6.2.5 The analysis of receptor specificity in the absence of LCK

Initial simulations including affinity and differential ligand concentrations suggest that, in these models, the SH2 domain of LCK dominates owing to its promiscuity and very high concentration. However, important questions are: what is the actual concentration of free LCK SH2 domain, and what would happen if most of the LCK is incapable of binding the receptors because it is auto-inhibited?

SRC kinases, such as LCK, possess a C-terminal tyrosine residue (Tyr^{505} for LCK) which, when phosphorylated by CSK, can form an intra-molecular interaction with the SH2 domain of itself, causing the protein to adopt an inactive conformation. This is stabilised by the additional interactions between the SH3 domain and a proline-rich sequence in the SH2-kinase linker region [352]. Since dephosphorylation of Tyr^{505} requires receptor protein tyrosine phosphatases, such as CD45, which may be excluded from the synapse during TCR triggering [353], it is likely that a large proportion of the LCK molecules inside the synapse may be retained in a non-functional state.

It is possible, however, that the phosphorylated receptor motifs can compete with phosphorylated Tyr^{505}. To investigate the likelihood of this, an SPR-based experiment was conducted to determine the affinity of the LCK SH2 domain for the Tyr^{505} residue for comparison with the previously determined SH2 domain affinities for the receptor motifs (**Figure 6.11**). The K_D value obtained for the LCK SH2 Tyr^{505} motif is 7.0 μM, which is somewhat lower (i.e. higher affinity) than most of the receptor motifs studied, including all the TCR ITAM motifs. Although

ITIM and ITSM motifs of BTLA, the ITSM motif of PD-1, as well as the LAT tyrosine-containing motifs have higher affinity than the Tyr^{505} motif for the LCK SH2 domain, the differences are only within 3-fold. The only real exception is the LAT_Y(-p)ENL, whose affinity for the LCK SH2 is 0.87 μM, almost 10-fold higher than the Tyr^{505} motif. The comparable affinities between the Tyr^{505} motif and most of the receptor motifs for the LCK SH2 domain suggests that the phosphorylated receptor motifs might be able to compete with the SH2 domain from the phosphorylated Tyr^{505} motif and render the LCK in an active conformation. However, it is also necessary to bear in mind that the intramolecular interaction may have entropic advantages over the intermolecular ones, although this is far from certain.

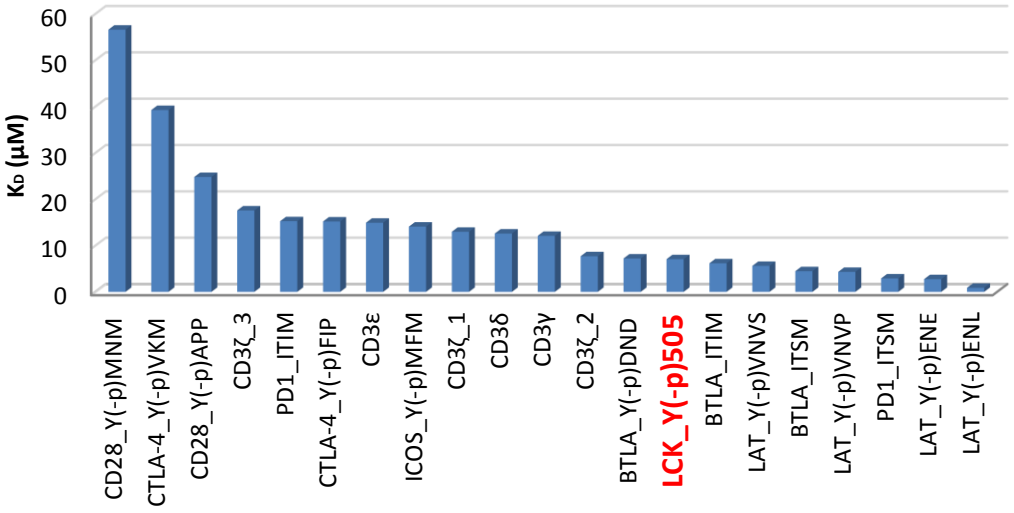


Figure 6.11. A comparison of the affinities of the different phosphorylated receptor motifs and the LCK Tyr^{505} motif for the LCK SH2 domain. The affinities are presented as K_D values (μM) and sorted from highest (lowest affinity) to lowest.

In the light of these considerations and the significant role of LCK in binding to most of the receptor motifs in Model 2, a further receptor specificity analysis was conducted, this time in the absence of this SRC kinase [**Figure 6.12 (B and D)**]. This showed that the absence of LCK only makes a marginal difference when only binding affinity is considered [**Figure 6.12 (A and C)** and **Figure 6.13 (A)**]. This is reflected by the fact that the affinity of LCK for the receptor motifs is relatively low compared to the strongest binders (**Figure 5.5**). However, notable differences are observed when LCK was removed and differential ligand concentrations were taken into account [**Figure 6.1.3 (B)**]. Moderate increases (10%-20%) in specificity were seen for all the CD3 ITAM motifs, as well as the CD28 motifs and the CTLA-4_Y(-p)VKM motif (about 10%), whose binding are largely dominated by ZAP70 in most cases, with CD28_Y(-p)MNM being the exception (dominated by SAP and GRB2). Almost all the inhibitory receptor motifs derived from CTLA-4, PD-1 and BTLA have reduced specificity in the absence of LCK, and the PD-1_ITSM motif is the one most heavily influenced by the removal of LCK from the system, experiencing a >20% reduction in specificity. This decrease is consistent with the previous dominance of LCK for these motifs (**Figure 6.9**). The exceptions are CTLA-4_Y(-p)VKM and BTLA_ITSM, in which cases an increase in specificity results (**Figure 6.14**).

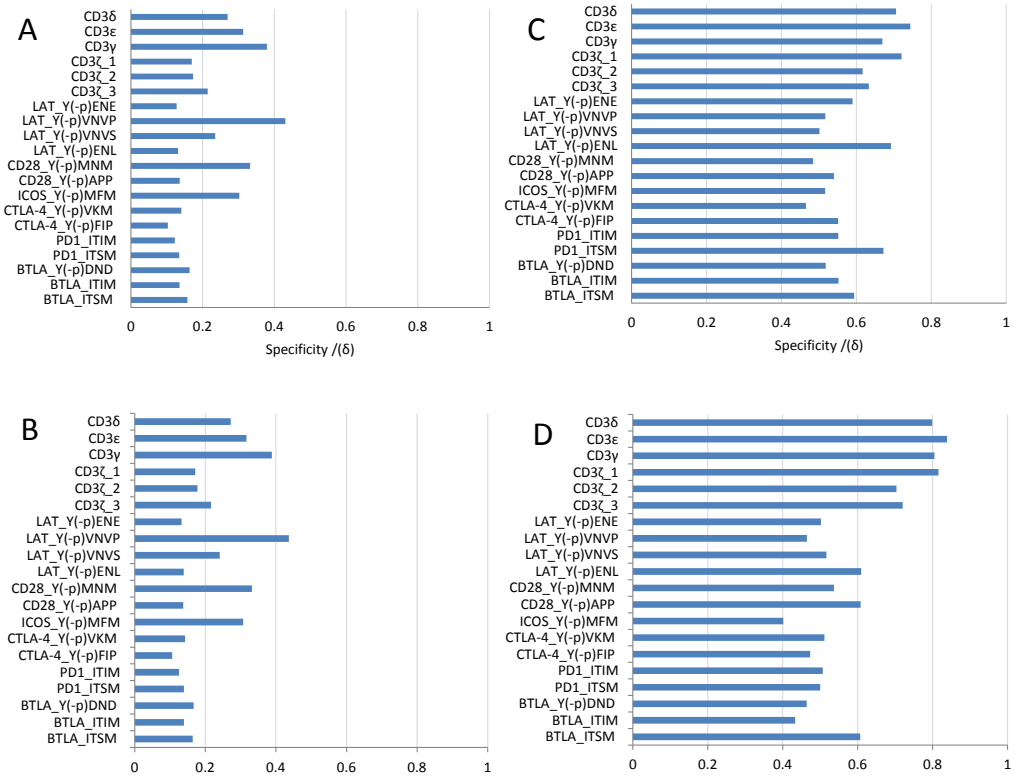


Figure 6.12. A comparison of receptor specificities for systems with and without LCK. (A) Specificity in the presence of LCK in Model 1. (B) Specificity in the absence of LCK in Model 1. (C) Specificity in the presence of LCK in Model 2. (D) Specificity in the absence of LCK in Model 2.

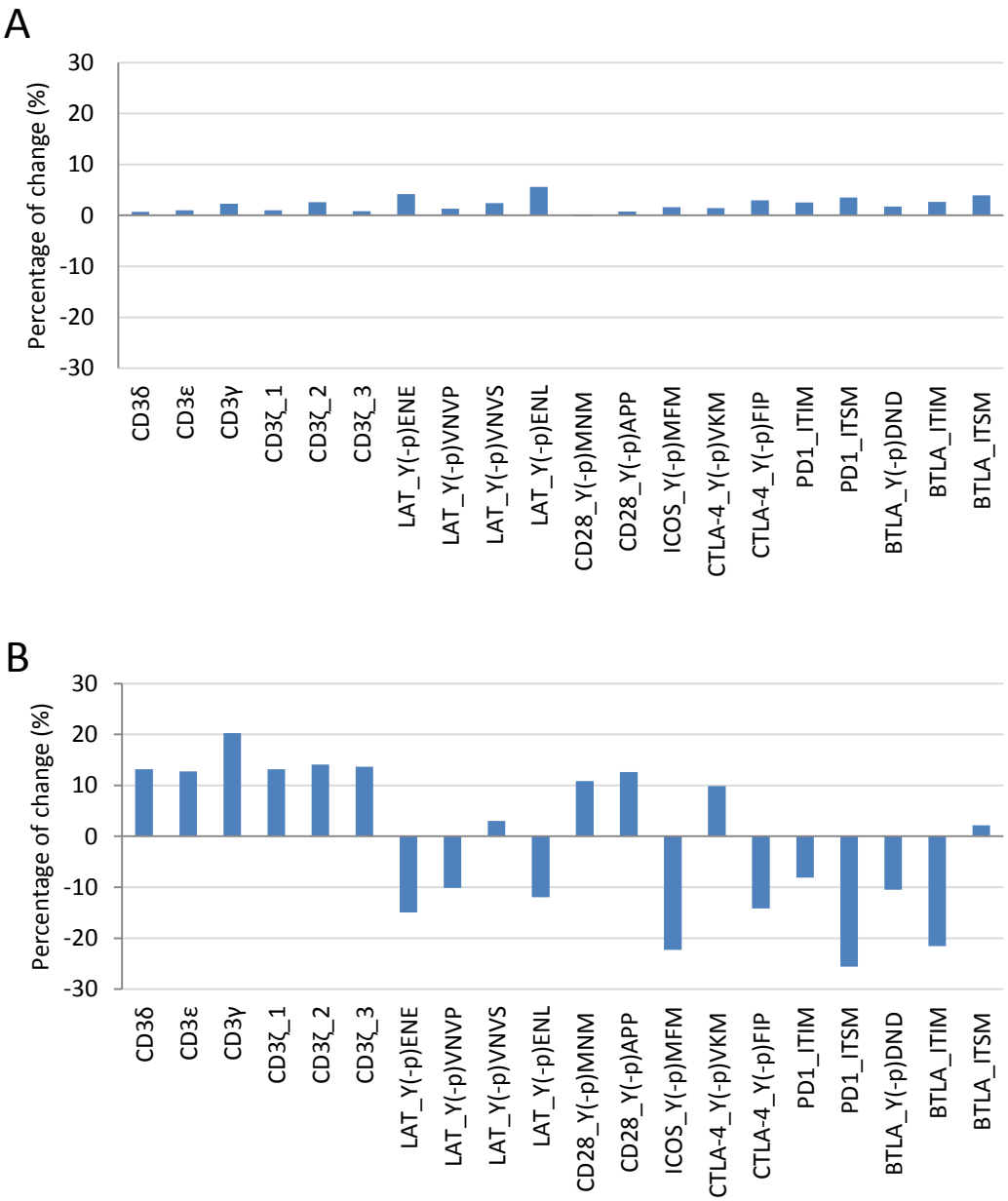


Figure 6.13. Percent change in receptor specificity for the system with the SRC-kinases versus that without the kinases. (A) Comparison in Model 1 (B) Comparison in Model 2.

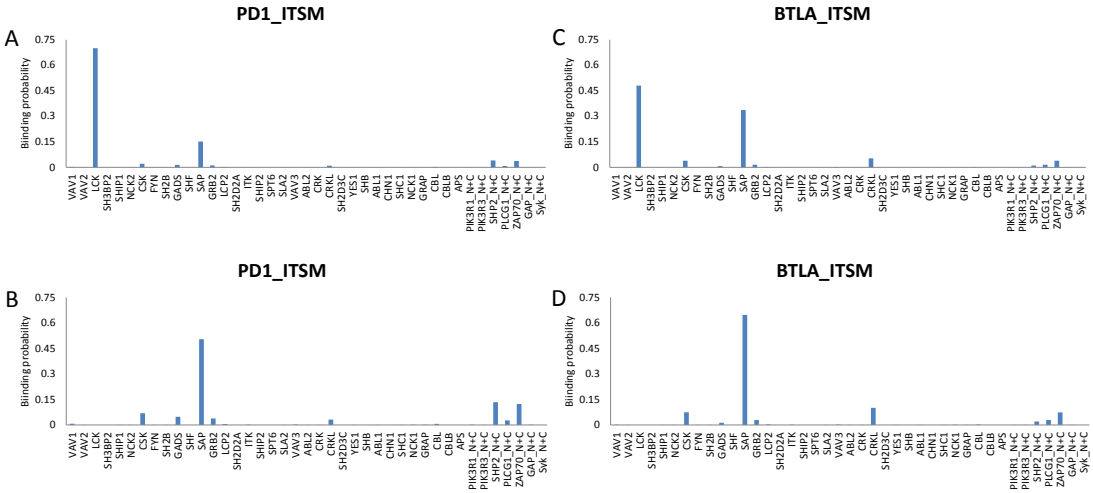


Figure 6.14. Comparison of the probability spectrum for the binding of the SH2 domains to the PD-1_ITSM and BTLA_ITSM motifs in Model 2 in the presence and absence of LCK. (A) and (C) Modelling in the presence of LCK; (B) and (D) Modelling in the absence of LCK. In the absence of LCK, SAP dominates the binding to both the PD-1_ITSM and BTLA_ITSM motifs.

6.3 Discussion

Signal transduction inside a cell has been previously described as consisting of a linear cascade of sequential biomolecular interactions. This idea relies on the assumption that these interactions are highly selective, and specificity is determined by the intrinsic property of the signalling molecules. However, analysis of the binding probability of SH2 domains (expressed in Jurkat T cells) to a series of key T cell receptor motifs in this study reveals that most receptor motifs can be occupied by many SH2 domains with comparable probabilities (i.e. the binding probability for more than 20 SH2 domains can be within a 10-fold range), when only binding affinity is considered. In order to give this a more appropriate physiological context, concentrations of the SH2 domain were simulated along with their biophysical properties. Strikingly, the high probabilities of binding are restricted to certain SH2 domains, notably LCK and ZAP70, and the

probabilities for most others are significantly reduced and almost become negligible. By using a novel measure of specificity, it was shown that receptor specificity resulted from binding affinity itself is relatively low, and adding ligand concentrations as a contributing factor can generally lead to notable increase in receptor specificity. These changes are largely, but not exclusively, a result of the exceptionally high concentrations of these dominant proteins, and it implies that receptor specificity is system- rather than molecule-intrinsic.

It should be acknowledged that this model does not reflect the real situation under physiological conditions. Although ligand concentrations are considered here, one major assumption here is that the receptor concentrations are in excess of the ligands, and only one receptor motif is considered at a time in the system, which means that no competition occurs between receptors for ligands. If competition exits, molecules which play a dominant role in binding, such as ZAP70 and LCK, may be sequestered by certain receptors and change the overall binding outcome. In addition, the proteomic data used in this study is from resting cells, whilst receptors such as ICOS, CTLA-4, PD-1 and BTLA are only up-regulated in activated cells whereupon the types of SH2 domains that are expressed is likely to change. For example, while only 2 proteins have unchanged mRNA expression level after a T cell clone (Clone 29) was activated, a total of 53 proteins were upregulated, and 50 were down-regulated (**Section 3.2.1.4**). In particular, it was shown that the mRNA expression of LCK was reduced approximately 10-fold after activation of the cell, which would affect its dominance in binding, if this change in expression were translated into protein levels. This therefore adds some uncertainties to the modelling of the receptors expressed upon activation. In addition, other factors such as the stoichiometry of binding and formation of defined signalling complexes (i.e. "signalosomes") may also affect specificity under physiology conditions.

One important issue in the modelling is the extent of inclusion of LCK in the system, since the amount of LCK in its active conformation inside the synapse is still unclear, while modelling in this study suggests that this factor can significantly affect the receptor specificity, in particular for the inhibitory receptors PD-1 and BTLA. As the SRC kinases, such as LCK, are anchored to the inner leaflet of the membrane via myristoylation and palmitoylation, they would be given even greater priority to interact with the surface receptors [347] and thus magnify the dominance of LCK and the increase in the receptor specificity, provided that a significant portion of these molecules are in their active states.

Chapter 7. General Discussion

"Dealing with these system properties, which ultimately must underlie our understanding of all cellular behaviour, will require more abstract conceptualization than biologists have been used to in the past… A useful metaphor could be the changes in thinking associated with physics at the beginning of this century… We might need to move into a strange more abstract world, more readily analysable in terms of mathematics than our present imaginings of cells operating as a microcosm of our everyday world." Paul Nurse, writing in 2000 [354].

For many years, the prevailing view of signal transduction has been that it proceeds through a linear relay of information via highly selective sequential bimolecular interactions. However, the impression has been that this is likely to have over-simplified what really happens inside a cell, considering the complexity of the responses needed to cope with their environments and to make decisions about survival, growth, proliferation, differentiation or migration, and the need for the system to be robust enough to allow proper cell function in the face of stochastic variation in gene expression.

One of the classic examples of signal transduction nevertheless proposed to adopt this linear relay of information with highly selective interactions is the signalling pathway through surface receptors containing cytoplasmic tyrosine motifs which, when the cells are activated, become phosphorylated thereby creating a platform for recruitment of SH2 domain-containing proteins [355]. This dogma has probably been strengthened by the use of techniques such as immune-precipitation or yeast two hybrid screens, wherein only limited numbers of SH2 proteins are found to interact with the receptor of interest, since only the

binding partners with the highest affinity will be identified under these conditions. However, the affinity for specific interactions are proposed to have to be at least several orders of magnitude higher than that for non-specific interactions, so as to avoid cross reactions [187], but emerging evidence suggests that such large differences in affinity may not exist [182, 188, 191].

In this study, we set out to measure, for the first time, the interactions of most of the SH2 domains expressed by a T cell with all the important phosphotyrosine-dependent receptors expressed by these cells, in order to understand the hierarchy of SH2/receptor interactions in T cells and to try to understand how signalling specificity might be achieved. By developing a mathematical model incorporating the affinities for the interactions and the SH2 protein concentrations, it was found that specificity perhaps is not predominantly determined at the level of the interactions of the receptors, but at the level of the system, with the largest changes in the system of interactions thusfar resulting from the regulation of expression.

7.1 Results from the thesis

7.1.1 Transcriptomic and proteomic analysis of SH2 protein expression

In Chapter 3, the transcriptomics data of SH2 protein expression in a wide range of immune and non-immune cells was examined. It was found that large numbers of SH2 proteins are transcribed, with an average of ~50 of the 110 genomically encoded SH2 proteins being transcribed in each of the cells examined. However, this is likely an underestimate, reflecting the relatively shallow sampling of

conventional SAGE experiments. Indeed, in the case of the T-cell clone 29, the expression of a total of 105 different SH2 domain-containing proteins was detected at the transcript level when 10-fold deeper SAGE sampling was undertaken. The levels of expression of each of the SH2 domain-encoding transcripts varied significantly between the different cell types. Whereas T cells (and immune cells as a group) have significantly higher expression of SH2 proteins compared to non-immune cells, the difference between T cells and other immune cells is marginal. This differential transcription between immune and non-immune cells correlates very well with the significant bias in tyrosine-motif containing receptor expression that is also observed in immune cells (Y. Lui et al, unpublished). We conclude from this that tyrosine phosphorylation was particularly exploited during the evolution of the immune system (as also discussed below).

By way of contrast, the mRNA expression of PTB domain-containing proteins was significantly biased toward non-immune cells. In addition, whereas SH2 protein transcription is under dynamic control in T cells, the transcription of PTB protein was significantly more stable. This might be due to the fact that PTB domains do not only recognise tyrosine-containing motifs after they are tyrosine phosphorylated [176], with the non-specific binding capability perhaps making them less well-suited to the regulation of signalling in immune cells, where the tyrosine-dependent signalling system (involving tyrosine-motif containing receptors and SH2 domains) confers upon signalling a high degree of sensitivity.

One surprising finding in the comparisons is that among the SH2 proteins found to have significantly higher mRNA expression in immune cells (and T cells) compared to non-immune cells, RIN3 was found to be the most significantly biased toward immune cells. The literature suggests that an immune-specific role for this protein is yet to be discovered. Since there is a good correlation between SH2 protein and mRNA expression for abundant transcripts significantly, one

would expect the levels of this protein to be high enough to participate in some immune cell-specific functions. It is known that RIN3 is a small GTPase Rab5-GEF (guanine-nucleotide exchange factor) involved in the early endocytic pathway, and that tyrosine-phosphorylation in the cell results in activation and translocation of RIN3 to Rab5-positive vesicles [322]. This suggests that this protein might have an important role in endocytic recycling of phosphotyrosine motif containing-surface receptors such as CTLA-4 [356], and the proposed CTLA-4 mediated trans-endocytosis of CD80 and CD86 [125].

Comparisons of the deep transcriptomes of a human T cell clone (clone 29) derived from peripheral blood mononuclear cells and Jurkat T cells, revealed that SH2 protein expression in these two cells are actually well correlated to each other (R^2=0.58), which supports the argument that the Jurkat T-cell is a good signalling model not only because of the ease of proliferation and genetic manipulation [234]), but most importantly because its signalling machinery indeed resembles that of a normal human T cell.

Previous use of Jurkat cells as a model signalling system has led to a number of important discoveries with respect to TCR signalling. These include the induction of calcium mobilisation by TCR triggering [357, 358], and also the induction of tyrosine phosphorylation [359], which was later shown to be a crucial step in TCR mediated signal transduction [360]. This was followed by the identification of ZAP70 [361], which is a kinase that plays a vital role in TCR signalling by initiating the phosphorylation and activation of a number of important intracellular signalling molecules, including LAT [362] and SLP-76[68]. In addition, the observation that tyrosine phosphorylation and activation of PLCG1, a key mediator of calcium mobilisation, is triggered by TCR signaling, was also made using Jurkat T cells. A search of the MEDLINE database (http://www.ncbi.nlm.nih.gov/pubmed) reveals that the terms “Jurkat” and

"signalling" are used in the titles or abstracts of more than 2500 publications, confirming the wide use of this T-cell signalling model.

7.1.2 Production of SH2 domains for system-level binding analysis

With the results from Chapter 3 indicating that a large number of SH2 proteins are expressed in a cell, the question arises of how such large numbers of signalling molecules can mediate specific signalling inside the cell. Although the numbers of proteins is relatively large, they fell within the range for which there was some expectation that the complete set could be produced by systematic binding studies. The advantages of the Jurkat T cells as a model, including in particular the ease with which it can be manipulated in experiments for testing hypotheses, drove us to employ it as the target model in this study.

To perform a systematic study of the interactions between SH2 domains and the tyrosine-containing motifs of some of the most important surface receptors, large quantities of soluble, properly folded recombinant SH2 proteins were required. As described in Chapter 4, the approach of using bacterial expression systems to express the SH2 domains with a variety of fusion tags, allowed us to prepare 63 properly folded, purified SH2 domains (approximately 90% of the total targets needed for a system-level analysis). The STAT, SOCS and JAK family proteins were eventually excluded from our list since they could not be purified and in view of their well-characterised role apparently confined to cytokine signalling [316].

On a technical level, this large-scale protein expression project revealed that it is often not helpful to use small-scale expression screens as the sole means of evaluating the effectiveness of a particular solubilising tag and thus the

success of making soluble recombinant proteins, as is presently the mainstream approach [275, 276]. For example, although fusion tags such as MBP and NuSA were found to be highly capable of solubilising recombinant proteins [305, 310], they can in fact make the target protein "falsely soluble", such that it either cannot be isolated from the fusion tag or is prone to precipitate upon cleavage. As a result, the validity of using un-cleaved fusion proteins to perform binding experiments as has been done previously in large-scale bottom-up systems biology-style experiments [191] is questionable as the protein may not be properly folded and functional.

7.1.3 Interactions of SH2 domains with the phosphotyrosine-containing motifs of T cell signalling receptors

With the large set of high-quality SH2 domains acquired from the large-scale protein expression and purification work described in Chapter 4, it was possible to investigate the interactions between these SH2 domains and the phosphotyrosine-containing motifs of T cells signalling receptors by SPR-based assays. All of the major tyrosine-phosphorylation dependent signalling receptors in T cells were chosen for this initial study, including TCR, LAT, CD28, ICOS, CTLA-4, PD-1 and BTLA.

In this large-scale assay, a number of well-known interactions were re-confirmed insofar as the strongest affinities found were for previously reported interacting motifs and SH2 domains, including those between ZAP70 with the membrane-proximal ITAM motif of CD3ζ [361], PLCG1 with the LAT_Y(-p)ENL motif [75], PIK3 with the CD28_Y(-p)MNM [98], CTLA-4_Y(-p)VKM [113] and ICOS_Y(-p)MFM [363] motifs, and SHP2 with the PD-1_ITSM motif [136]. In addition, this study offers biophysical support for the suggestion that there is unlikely to be a significant association between CTLA-4 and SHP2 [$K_{D(CTLA\text{-}4/SHP2)}$ =

33 μM versus $K_{D(PD\text{-}1_ITSM/SHP2)}$ = 1.3 μM] [114], as argued previously [116]. Whilst SHP2 was reported to be associated with the ITIM and ITSM motifs of BTLA [146], their affinities for these motifs were found to be only moderate in this study [$K_{D(BTLA_ITIM/SHP2)}$ = 5.4 μM; $K_{D(BTLA_ITIM/SHP2)}$ = 18 μM], which may reflect the finding that association between SHPs with BTLA can only occur in the presence of both the ITIM and ITSM [146]. This could be tested in the future by using a BTLA_ITIM/ITSM tandem motif as the ligand or coating sensor chips with both (BTLA_ITIM and ITSM) ligands.

Nonetheless, it was a considerable surprise to acquire measureable binding affinities for 974 interactions between 59 SH2 domains and 20 tyrosyl phosphopeptides derived from the selected tyrosine-receptor motifs (i.e. 83% of the 1180 interactions tested), and to find that these affinities only span a two-order magnitude range of differences. Each receptor motif can, on average, bind to 48.7 (±0.2) different SH2 domains, and even the least promiscuous motif, i.e. the ITAM motif of CD3_γ binds to 37 SH2 domains. In addition, a large number (365 in total, accounting for 37% of the total number of interactions) were found to have fairly high affinities (K_D < 10 μM). LAT is the most promiscuous receptor, which can accommodate 100 strong interactions (K_D < 10 μM), and the LAT_Y(-p)VNVP motif in particular can bind strongly (K_D < 10 μM) to 28 SH2 domains.

A previous system-level study of the interactions between SH2 domains and ErbB receptors [191], which also contain phosphotyrosine motifs, also revealed high degree of cross-reactivity in this signalling system, with the difference in binding affinities also being mostly within two-orders of magnitude. For example, the sites on ErbB2 can interact strongly with over 17 different proteins (SH2/PTB) on average (K_D < 2 μM), and the pY1139 motif of ErbB2 interacts with 24 different proteins with high affinity (K_D < 2 μM). However, this might have been expected given the structural similarities and evolutionary

relatedness of ErbB receptors, whereas the set of receptors studied here have profoundly different functions.

The data presented herein indicates that, instead of having highly selective interactions, high levels of cross-reactivity exist between the tyrosine motif-containing receptors and the SH2 proteins (based only on the intrinsic properties of the SH2 domains). This therefore supports the argument that affinity itself does not determine the specificity of signalling reactions [187]. Previous studies have, however, suggested the likelihood of high levels of cross-reactivity as the binding of an SH2 domain to the phosphotyrosine residue is responsible for about 50% of the total free energy of interaction, and the contribution of other amino acids is relatively small [182].

There are two groups of SH2 domains which are highly promiscuous: the SRC kinases (LCK, FYN and YES) and the PIK3 SH2 domains, although the affinities of PIK3 SH2 domains to most receptor motifs are much higher than the SH2 domains of SRC kinases. Whereas PIK3 is the SH2 domain with highest affinity for the CD28_Y(-p)MNM (K_D = 0.61 μM), CTLA-4_Y(-p)VKM (K_D = 3.2 μM) and ICOS_Y(-p)MFM (K_D = 0.54 μM) motifs, consistent with previous findings [98, 113, 363], this study also shows that PIK3 is potentially the highest affinity ligand for the ITAM motifs of CD3δ, CD3ε and CD3γ, which has not been previously reported. However, since PIK3 is known to be important for CD28 signalling [342, 364], which provides a potent co-stimulatory signal to the cell upon TCR engagement to rescue T cells from anergy or apoptosis and promote T cell clonal expansion [342], it is unclear why CD3δ, CD3ε and CD3γ are not able to replace CD28 with respect to such role since they can also bind strongly to PIK3 and their affinities for PIK3 are indeed higher than that for the CD28 phosphotyrosine motifs. Additionally, since PIK3 also has high affinity for CTLA-4 (which provides inhibitory signals to the cell [110-112]), the association of the two seems

somewhat incompatible with the widely accepted role of PIK3 involved in CD28 signalling.

One explanation for these apparent discrepancies is that although PIK3 has very high affinities for these receptors, it may not bind to one or the other receptors inside the cell. Instead, other SH2 domains with lower affinities to the receptor motifs may out-compete PIK3 for their binding sites. These associations may not necessarily be identifiable by approaches such as immunoprecipitation as this is somewhat biased toward proteins of interest or "expected" to interact or the affinity is too low to be "precipitated" in procedures involving several wash steps. Overall, whereas some receptors have only one or two binders with distinctly high affinities, others like the ITIM motifs of PD-1 and BTLA instead have a number of strong binders with almost identical affinities. It would therefore be difficult to generate any specificity at all if affinity was the sole driver of binding. As a consequence, it does not seem like affinity is the sole determining factor for receptor specificity.

7.1.4 Analysis of specificity in a system-based model of receptor-ligand interactions

Mathematical models are becoming indispensable tools for studying systems biology with the aim of understanding complex cellular signalling networks [365]. Large systems biology models contain many chemical species and as a consequence, many reaction rates and chemical concentrations. The large number of unknown parameters must be specified in order to make predictions, and investigators often can only estimate these as best they can. However, it is well known that model predictions are only valid for certain parameter choices and therefore any inaccurate parameter estimate may produce an inaccurate model prediction.

In Chapter 6, with the affinities determined for the receptor-ligand interactions and the concentration of SH2 proteins in Jurkat T cells estimated, we then developed a fully parameterized mathematical model to investigate how other factors, such as ligand concentrations, can affect receptor specificity. Incorporating the ligand (SH2 domain) concentrations into the model led to significant changes in the binding outcome (measured as binding probabilities, $P_{i,j}$, with a scale of 0 to 1). This is characterised by the disappearance of promiscuity and the emergence of interactions (mediated by certain SH2 domains, notably LCK and ZAP70) with distinctly high binding probabilities ($P_{i,j}$ >0.4). Whilst there is a moderate increase in the proportion of interactions with high binding probabilities ($P_{i,j}$ >0.1; from 5% to 9% of the total number of interactions), there is a dramatic decrease in intermediate binding probability (0.01< $P_{i,j}$ <0.1; from 55% to 14%. This is accompanied by an increase in interactions with a low binding probability ($P_{i,j}$ <0.01; from 40% to 77%), after ligand concentrations were taken into account. Overall, there is an overall sharp distinction in the binding probabilities between the strongest ligands and the weaker ones, signifying an increase in specificity.

Whereas ZAP70 was dominant in binding to the CD3 ITAM motifs, LCK dominated binding in most other cases. In contrast, the probability for other SH2 domain to interact with the receptor motifs was markedly reduced and PIK3 seems unlikely to bind to these receptors inside the cell due to its extremely low cellular abundance, according to the present study. By using a novel measure of specificity (with entropy), it was shown that taking ligand concentration into account allows for greatly enhanced receptor specificity, which is largely, but not exclusively, a result of the exceptionally high concentrations of several proteins such as LCK and ZAP70. Paradoxically, however, this in turn results in a reduction

of specificity, as only very few SH2 proteins are likely to bind to most of the receptors.

7.1.5 Signalling implications of the new findings

Although the dominance of ZAP70 in binding to the TCR is consistent with well-established previous findings, the dominance of LCK in binding to 95% and 45% of receptor motifs with K_D<50μM and K_D<10μM, respectively, is somewhat surprising since its principal function is believed to be simply to phosphorylate tyrosine motifs or residues [366]. The question is: why would LCK bind to these motifs so promiscuously? In fact, LCK may be given even higher priority access to the phosphotyrosine motifs compared to other cytosolic SH2 proteins because of its membrane localisation, as it was shown that pairs of proteins constrained to two dimensions at the cell surface have markedly enhanced random interactions compared with proteins localized on the membrane and in the cytoplasm [347].

However, a very important issue that may prevent LCK binding to the phosphorylated receptors is that phosphorylation of the C-terminal inhibitory Tyr^{505} residue of LCK by CSK, and subsequent binding of Tyr^{505} to the SH2 domain may blocks its ability to bind other receptors, whilst also conferring upon the kinase protein an inactivated conformation [352]. Based on the levels of phosphorylation of Tyr^{505} and Tyr^{394} (the activating site of the kinase which, upon *trans* auto-phosphorylation, promotes substrate binding and enzymatic activity [367]), it has been recently estimated that as much as 40% of total LCK is constitutively active in a resting cell [368]. Because there is a lack of receptor protein tyrosine phosphatases inside the immunological synapse [53] that might dephosphorylate Tyr^{505}, it then raises the question of how many free LCK SH2 domains are available for receptor binding. Importantly, since phosphorylation of the TCR (and other tyrosine-containing receptors) by LCK is the key step for

initiation of intracellular signalling [353], the fraction of active LCK inside the immunological synapse would therefore determine receptor phosphorylation levels, and thus signalling productivity. Inappropriate regulation of, for example, TCR phosphorylation can result in autoimmune disorders [369]. So, given all the possible types of interactions, how can LCK activity be regulated during signalling?

One possibility is that the binding of the SRC kinase SH2 domain to their substrates, which is itself phosphorylation dependent, would stabilise the kinase in an active state, favouring multi-site phosphorylation of receptors with more than one tyrosine motif. "Feed-forward" regulation of this type has been observed for c-SRC, in which case Sin (SRC interacting or signal integrating protein) acts as both an activator and a substrate for c-SRC, and phosphorylated Sin may serve as a signaling effector molecule for SRC by binding to multiple cellular proteins [370]. In this scenario, relatively small amounts of active LCK would only be needed in the resting cell, obviating the risk associated with having large levels of active kinase.

To explore the possibility that the LCK/receptor interactions identified here could compete for phosphorylated Tyr^{505}, we performed SPR-based experiments to examine the affinity of LCK SH2 domain for the phosphorylated Tyr^{505} motif and found that this affinity is in fact lower than that of LCK for tyrosine receptor motifs derived from LAT and the inhibitory receptors BTLA and PD-1. Although the affinities of CD3 ITAM motifs for LCK were lower than that of the phosphorylated Tyr^{505} motif, Fernandes et al. (unpublished) have found that phosphorylated CD3 ITAM-derived peptides strongly activate LCK activity measured with enolase as substrate. These findings imply that there could be relatively large numbers of tyrosine binding sites that can potentially compete for LCK SH2 domain against the phosphorylated Tyr^{505} residue, thus activating the LCK kinase. The extent to which this occurs *in vivo* would likely depend on: (1) the competing activities of CSK and other kinases for Tyr^{505} on LCK; (2) the degree to which binding is

favoured following encounter complex formation between LCK and the multi-tyrosine-containing receptors [371] and (3) the amount of time these receptors and LCK spend in the regions of contact between T-cells and APCs, where there is likely to be no receptor phosphatase (e.g. CD45) activities [47].

Although LCK or ZAP70 seemed to dominate binding in our model, it should be borne in mind that this only represents what may happen at the very early stages of cell activation, since only a resting cell proteome was used for modelling. The dynamic changes in the mRNA expression for SH2 proteins (a total of 53 proteins were upregulated upon activation and 50 were down regulated in the T cell clone, clone 29), and the sharp (8-fold) decrease in LCK (mRNA) expression after the cells were activated suggests that the outcome of the simulations would be different if proteomes at different cellular stages were used. However, it was shown that, when LCK is excluded from the system, only moderate decreases in receptor specificity were observed for receptor motifs whose binding was dominated by LCK when LCK was present (e.g. the LAT, ICOS, PD-1 and BTLA receptor motifs). In the absence of LCK, other SH2 domains, such as SAP, dominated receptor binding. This underlines the potential importance of changes in signalling protein expression, in regulating signalling outcomes.

7.1.6 Verification of the simulation model

The *in silico* simulations developed in this study have produced a number of surprising predictions that remain to be verified. The most surprising finding is that LCK has been shown to be the dominant binder of virtually all of the receptors studied in this model, whereas the main function of this molecule is supposed to be that it mediates phosphorylation of the receptor rather than mediating any signalling function itself, although it has been reported to provide an alternative mode of signalling by regulating the activities of PKCθ and PDK1

[106, 107]. A second prediction is that ZAP70 has a relatively high probability of binding to CD28 and CTLA-4, whereas this molecule was previously found to be the crucial binder for the ITAM motifs of TCR subunits [256]. The fact that ZAP70 is found to have relatively high affinity (mostly with a K_D between 4 and 20μM) to receptor motifs with only one phosphorylated tyrosine residue can perhaps be at least partly explained by binding of the adjacent receptor peptides on the sensor chip in the SPR-based binding experiments. Whereas GADS and GRB2 were found to be important mediators of the LAT and CD28 signalling pathways (discussed in **Section 1.6.2 and 1.6.3.1**), and therefore likely to deliver positive signals as expected, it is surprising that they may also interact with the inhibitory receptors BTLA and CTLA-4, respectively. This is reminiscent of the observation that PIK3 interacts with CTLA-4 in addition to CD28, which may function to deliver a survival signal to the cell when other inhibitory mechanisms are operating. Lastly, SAP is predicted by the simulations to interact with CD28 and all three inhibitory receptors (CTLA-4, PD-1 and BTLA). SAP was previously found to be a key adaptor molecule that interacts with SLAM [266], and was implicated in differentiation of Th2 cells [372]. The function of SAP may have to be reconsidered if the new interactions are shown to occur. That many of these "new" interactions have dissociation constants >1μM, both explains why they haven't been seen previously using the immune-precipitation method (owing to the need for wash steps), and suggests the requirement for new ways to identify significant interactions within the 2D environment of the membrane between proteins that have relatively low solution affinities. One approach to identifying these interactions would be to produce fluorescently tagged forms of the proteins, and to follow their co-ordinated movement following receptor triggering since this would imply induced co-association. These sorts of approaches have been used previously to investigate the stoichiometry of receptors such as the TCR [373, 374] and could perhaps be adapted to the identification of transiently interacting proteins.

7.1.7 Towards a more complete simulation model

This study represents, at best, only a first approximation of how tyrosine phosphorylation might be configured inside mammalian cells. Whereas the SPR-based binding study has revealed the remarkable levels of cross-reactions that could occur in principle, the simulations suggest that ligand expression effects resolve this. However, there are several additional levels of complexity that need to be built into the simulations. In addition to the ligand concentrations, the receptor concentrations might also substantially change the outcome, as there are 148 different tyrosine motif-containing receptors present on the cell surface (Y. Lui, unpublished) that can compete for the SH2 proteins. In addition, other factors such as the stoichiometry of binding [95], interactions through other modular domains forming part of the protein (e.g. PTB domains, SH3 domains and PH domains), as well as the assembly of multi-protein complexes will also complicate the outcome [348].

For example, it has been shown that GRB2 can interact with CD28 via its SH3 domains [375] and that these mediate complex formation. Since the systems model developed in this study has shown that the relatively high concentration of GRB2 confers upon it one of the highest binding probabilities next to LCK and ZAP70, the additional contribution to binding of its SH3 domain could possibly allow it to be responsible for initiating CD28 signalling inside the T cell. It has also been proposed that GRB2 can mediate complex formation with LAT via both of its SH2 and SH3 domains in collaboration with other adapter proteins such as LCP2 (SLP-76) [376]. Similarly, the SH3 domain of LCK will likely assist it with binding to receptor motifs and its activation. The obvious next step allowing the development of a model that would better mimic the *in vivo* situation would therefore involve acquiring data for the binding of paired SH2/SH3 domains to phosphorylated tyrosine receptor motifs. A second important matter is to properly incorporate protein localization effects in the model.

7.2 Final remarks

The coupling of an N-terminal SH2 domain with a kinase domain emerged early in evolution apparently coincidently with the occurrence of specific tyrosine phosphorylation [377] and has been highly conserved ever since: this arrangement has been reported to be present in the unicellular choanoflagellate *Monosiga brevicollis*, a primitive common ancestor of all metazoans [378]. Such an arrangement may have initially provided a basis for targeting kinases to their substrates [379], and later developed into an allosteric mechanism for controlling their local activity [380] in multicellular organisms where increased signalling complexity is required. The biased expression of SRC kinases and tyrosine motif-containing receptors towards leukocytes (Y. Lui, unpublished) suggests that the expansion of this family of kinases might have evolved to help solve a specialised triggering problem for this particular group of cells, one that involved the transient contacts of functionally interactive cells.

How SH2 domain function fits within the larger context of signalling events in vivo is a multi-dimensional puzzle of great complexity (involving space, time, cell type and cellular states etc; [381]). As Paul Nurse suggested in 2000 [354], the true nature of signalling might only be revealed by very large computational simulations of the manifold interactions occurring in the cell interior. The present work shows that it is possible to establish a fully parameterized systems model of a very complex system of potential interactions (>1000) that does not have even a single unknown parameter, i.e. to generate a model with zero degrees of freedom. The new model already suggests that signalling specificity is an emergent systems-level property rather than being protein-intrinsic, suggesting a change to our traditional view of how signal transduction proceeds. Based on these observations, unlike Professor Nurse, we are not so certain that understanding signal transduction will require a transformation in biological

thinking comparable to that which occurred in physics at the beginning of the 20th century.

Bibliography

1. Min, H., E. Montecino-Rodriguez, and K. Dorshkind, *Effects of aging on early B- and T-cell development.* Immunol Rev, 2005. **205**: p. 7-17.
2. Cobb, R.M., et al., *Accessibility control of V(D)J recombination.* Adv Immunol, 2006. **91**: p. 45-109.
3. von Boehmer, H., et al., *Pleiotropic changes controlled by the pre-T-cell receptor.* Curr Opin Immunol, 1999. **11**(2): p. 135-42.
4. Klein, L., et al., *Antigen presentation in the thymus for positive selection and central tolerance induction.* Nat Rev Immunol, 2009. **9**(12): p. 833-44.
5. Miosge, L. and R. Zamoyska, *Signalling in T-cell development: is it all location, location, location?* Curr Opin Immunol, 2007. **19**(2): p. 194-9.
6. Zhu, J., H. Yamane, and W.E. Paul, *Differentiation of effector CD4 T cell populations (*).* Annu Rev Immunol, 2010. **28**: p. 445-89.
7. Curotto de Lafaille, M.A. and J.J. Lafaille, *Natural and adaptive foxp3+ regulatory T cells: more of the same or a division of labor?* Immunity, 2009. **30**(5): p. 626-35.
8. Horwitz, D.A., et al., *Critical role of IL-2 and TGF-beta in generation, function and stabilization of Foxp3+CD4+ Treg.* Eur J Immunol, 2008. **38**(4): p. 912-5.
9. Germain, R.N., *T-cell development and the CD4-CD8 lineage decision.* Nat Rev Immunol, 2002. **2**(5): p. 309-22.
10. Romagnani, S., *Th1/Th2 cells.* Inflamm Bowel Dis, 1999. **5**(4): p. 285-94.
11. Vesosky, B., D.K. Flaherty, and J. Turner, *Th1 cytokines facilitate CD8-T-cell-mediated early resistance to infection with Mycobacterium tuberculosis in old mice.* Infect Immun, 2006. **74**(6): p. 3314-24.
12. Kidd, P., *Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease.* Altern Med Rev, 2003. **8**(3): p. 223-46.
13. Romagnani, S., *T-cell subsets (Th1 versus Th2).* Ann Allergy Asthma Immunol, 2000. **85**(1): p. 9-18; quiz 18, 21.
14. Bettelli, E., T. Korn, and V.K. Kuchroo, *Th17: the third member of the effector T cell trilogy.* Curr Opin Immunol, 2007. **19**(6): p. 652-7.
15. DiPaolo, R.J., et al., *Autoantigen-specific TGFbeta-induced Foxp3+ regulatory T cells prevent autoimmunity by inhibiting dendritic cells from activating autoreactive T cells.* J Immunol, 2007. **179**(7): p. 4685-93.
16. Podack, E.R., *Functional significance of two cytolytic pathways of cytotoxic T lymphocytes.* J Leukoc Biol, 1995. **57**(4): p. 548-52.
17. Gordy, C. and Y.W. He, *Endocytosis by target cells: an essential means for perforin- and granzyme-mediated killing.* Cell Mol Immunol, 2012. **9**(1): p. 5-6.
18. Weng, N.P., Y. Araki, and K. Subedi, *The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation.* Nat Rev Immunol, 2012.
19. Schluns, K.S. and L. Lefrancois, *Cytokine control of memory T-cell development and survival.* Nat Rev Immunol, 2003. **3**(4): p. 269-79.
20. Sakaguchi, S., et al., *FOXP3+ regulatory T cells in the human immune system.* Nat Rev Immunol, 2010. **10**(7): p. 490-500.
21. Sojka, D.K., Y.H. Huang, and D.J. Fowell, *Mechanisms of regulatory T-cell suppression - a diverse arsenal for a moving target.* Immunology, 2008. **124**(1): p. 13-22.
22. McEver, R.P., *Selectins: lectins that initiate cell adhesion under flow.* Curr Opin Cell Biol, 2002. **14**(5): p. 581-6.
23. Berg, E.L., et al., *L-selectin-mediated lymphocyte rolling on MAdCAM-1.* Nature, 1993. **366**(6456): p. 695-8.

24. Smith, A., et al., *The role of the integrin LFA-1 in T-lymphocyte migration.* Immunol Rev, 2007. **218**: p. 135-46.
25. Rose, D.M., R. Alon, and M.H. Ginsberg, *Integrin modulation and signaling in leukocyte adhesion and migration.* Immunol Rev, 2007. **218**: p. 126-34.
26. Montoya, M.C., et al., *Role of ICAM-3 in the initial interaction of T lymphocytes and APCs.* Nat Immunol, 2002. **3**(2): p. 159-68.
27. Sadhu, C., et al., *LFA-1 binding site in ICAM-3 contains a conserved motif and non-contiguous amino acids.* Cell Adhes Commun, 1994. **2**(5): p. 429-40.
28. Menon, S., et al., *Binding-site geometry and flexibility in DC-SIGN demonstrated with surface force measurements.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11524-9.
29. Qi, S.Y., J.T. Groves, and A.K. Chakraborty, *Synaptic pattern formation during cellular recognition.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6548-53.
30. Dustin, M.L. and J.A. Cooper, *The immunological synapse and the actin cytoskeleton: molecular hardware for T cell signaling.* Nat Immunol, 2000. **1**(1): p. 23-9.
31. Davis, S.J. and P.A. van der Merwe, *The structure and ligand interactions of CD2: implications for T-cell function.* Immunol Today, 1996. **17**(4): p. 177-87.
32. Kurts, C., B.W. Robinson, and P.A. Knolle, *Cross-priming in health and disease.* Nat Rev Immunol, 2010. **10**(6): p. 403-14.
33. Neefjes, J., et al., *Towards a systems understanding of MHC class I and MHC class II antigen presentation.* Nat Rev Immunol, 2011. **11**(12): p. 823-36.
34. Kane, L.P., J. Lin, and A. Weiss, *Signal transduction by the TCR for antigen.* Curr Opin Immunol, 2000. **12**(3): p. 242-9.
35. Pitcher, L.A. and N.S. van Oers, *T-cell receptor signal transmission: who gives an ITAM?* Trends Immunol, 2003. **24**(10): p. 554-60.
36. Barber, E.K., et al., *The CD4 and CD8 antigens are coupled to a protein-tyrosine kinase (p56lck) that phosphorylates the CD3 complex.* Proc Natl Acad Sci U S A, 1989. **86**(9): p. 3277-81.
37. Veillette, A., et al., *The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck.* Cell, 1988. **55**(2): p. 301-8.
38. Gao, G.F., Z. Rao, and J.I. Bell, *Molecular coordination of alphabeta T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands.* Trends Immunol, 2002. **23**(8): p. 408-13.
39. Locksley, R.M., et al., *Helper T cells without CD4: control of leishmaniasis in CD4-deficient mice.* Science, 1993. **261**(5127): p. 1448-51.
40. Schilham, M.W., et al., *Alloreactive cytotoxic T cells can develop and function in mice lacking both CD4 and CD8.* Eur J Immunol, 1993. **23**(6): p. 1299-304.
41. Krogsgaard, M., et al., *Agonist/endogenous peptide-MHC heterodimers drive T cell activation and sensitivity.* Nature, 2005. **434**(7030): p. 238-43.
42. Sporri, R. and C. Reis e Sousa, *Self peptide/MHC class I complexes have a negligible effect on the response of some CD8+ T cells to foreign antigen.* Eur J Immunol, 2002. **32**(11): p. 3161-70.
43. Bockaert, J. and J.P. Pin, *Molecular tinkering of G protein-coupled receptors: an evolutionary success.* EMBO J, 1999. **18**(7): p. 1723-9.
44. Unwin, N., et al., *Activation of the nicotinic acetylcholine receptor involves a switch in conformation of the alpha subunits.* J Mol Biol, 2002. **319**(5): p. 1165-76.
45. Gil, D., et al., *Recruitment of Nck by CD3 epsilon reveals a ligand-induced conformational change essential for T cell receptor signaling and synapse formation.* Cell, 2002. **109**(7): p. 901-12.
46. Fernandes, R.A., et al., *T Cell Receptors are Structures Capable of Initiating Signaling in the Absence of Large Conformational Rearrangements.* J Biol Chem, 2012. **287**(16): p. 13324-35.

47. Davis, S.J. and P.A. van der Merwe, *The kinetic-segregation model: TCR triggering and beyond.* Nat Immunol, 2006. **7**(8): p. 803-9.
48. Springer, T.A., *Adhesion receptors of the immune system.* Nature, 1990. **346**(6283): p. 425-34.
49. Secrist, J.P., et al., *Stimulatory effects of the protein tyrosine phosphatase inhibitor, pervanadate, on T-cell activation events.* J Biol Chem, 1993. **268**(8): p. 5886-93.
50. Irles, C., et al., *CD45 ectodomain controls interaction with GEMs and Lck activity for optimal TCR signaling.* Nat Immunol, 2003. **4**(2): p. 189-97.
51. Lin, J. and A. Weiss, *The tyrosine phosphatase CD148 is excluded from the immunologic synapse and down-regulates prolonged T cell signaling.* J Cell Biol, 2003. **162**(4): p. 673-82.
52. Choudhuri, K., et al., *T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand.* Nature, 2005. **436**(7050): p. 578-82.
53. Alarcon, B., D. Mestre, and N. Martinez-Martin, *The immunological synapse: a cause or consequence of T-cell receptor triggering?* Immunology, 2011. **133**(4): p. 420-5.
54. Brossard, C., et al., *Multifocal structure of the T cell - dendritic cell synapse.* Eur J Immunol, 2005. **35**(6): p. 1741-53.
55. Yokosuka, T., et al., *Newly generated T cell receptor microclusters initiate and sustain T cell activation by recruitment of Zap70 and SLP-76.* Nat Immunol, 2005. **6**(12): p. 1253-62.
56. Varma, R., et al., *T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster.* Immunity, 2006. **25**(1): p. 117-27.
57. Mossman, K.D., et al., *Altered TCR signaling from geometrically repatterned immunological synapses.* Science, 2005. **310**(5751): p. 1191-3.
58. Wiedemann, A., et al., *T-cell activation is accompanied by an ubiquitination process occurring at the immunological synapse.* Immunol Lett, 2005. **98**(1): p. 57-61.
59. Cemerski, S., et al., *The balance between T cell receptor signaling and degradation at the center of the immunological synapse is determined by antigen quality.* Immunity, 2008. **29**(3): p. 414-22.
60. Sabatos, C.A., et al., *A synaptic basis for paracrine interleukin-2 signaling during homotypic T cell interaction.* Immunity, 2008. **29**(2): p. 238-48.
61. Stinchcombe, J.C., et al., *The immunological synapse of CTL contains a secretory domain and membrane bridges.* Immunity, 2001. **15**(5): p. 751-61.
62. Stinchcombe, J.C., et al., *Centrosome polarization delivers secretory granules to the immunological synapse.* Nature, 2006. **443**(7110): p. 462-5.
63. Maldonado, R.A., et al., *A role for the immunological synapse in lineage commitment of CD4 lymphocytes.* Nature, 2004. **431**(7008): p. 527-32.
64. Weiss, A., *T cell antigen receptor signal transduction: a tale of tails and cytoplasmic protein-tyrosine kinases.* Cell, 1993. **73**(2): p. 209-12.
65. Hatada, M.H., et al., *Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor.* Nature, 1995. **377**(6544): p. 32-8.
66. Chan, A.C., et al., *Activation of ZAP-70 kinase activity by phosphorylation of tyrosine 493 is required for lymphocyte antigen receptor function.* EMBO J, 1995. **14**(11): p. 2499-508.
67. Wange, R.L., et al., *Activating and inhibitory mutations in adjacent tyrosines in the kinase domain of ZAP-70.* J Biol Chem, 1995. **270**(32): p. 18730-3.
68. Bubeck Wardenburg, J., et al., *Phosphorylation of SLP-76 by the ZAP-70 protein-tyrosine kinase is required for T-cell receptor function.* J Biol Chem, 1996. **271**(33): p. 19641-4.

69. Au-Yeung, B.B., et al., *The structure, regulation, and function of ZAP-70.* Immunol Rev, 2009. **228**(1): p. 41-57.
70. Smith-Garvin, J.E., G.A. Koretzky, and M.S. Jordan, *T cell activation.* Annu Rev Immunol, 2009. **27**: p. 591-619.
71. Elder, M.E., et al., *Human severe combined immunodeficiency due to a defect in ZAP-70, a T cell tyrosine kinase.* Science, 1994. **264**(5165): p. 1596-9.
72. Wange, R.L., *LAT, the linker for activation of T cells: a bridge between T cell-specific and general signaling pathways.* Sci STKE, 2000. **2000**(63): p. re1.
73. Zhang, W., R.P. Trible, and L.E. Samelson, *LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation.* Immunity, 1998. **9**(2): p. 239-46.
74. Kabouridis, P.S., A.I. Magee, and S.C. Ley, *S-acylation of LCK protein tyrosine kinase is essential for its signalling function in T lymphocytes.* EMBO J, 1997. **16**(16): p. 4983-98.
75. Zhang, W., et al., *Association of Grb2, Gads, and phospholipase C-gamma 1 with phosphorylated LAT tyrosine residues. Effect of LAT tyrosine mutations on T cell angigen receptor-mediated signaling.* J Biol Chem, 2000. **275**(30): p. 23355-61.
76. Zhu, M., E. Janssen, and W. Zhang, *Minimal requirement of tyrosine residues of linker for activation of T cells in TCR signaling and thymocyte development.* J Immunol, 2003. **170**(1): p. 325-33.
77. Bunnell, S.C., et al., *Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade.* J Biol Chem, 2000. **275**(3): p. 2219-30.
78. Qi, Q. and A. August, *Keeping the (kinase) party going: SLP-76 and ITK dance to the beat.* Sci STKE, 2007. **2007**(396): p. pe39.
79. Torgersen, K.M., et al., *Molecular mechanisms for protein kinase A-mediated modulation of immune function.* Cell Signal, 2002. **14**(1): p. 1-9.
80. Quann, E.J., et al., *A cascade of protein kinase C isozymes promotes cytoskeletal polarization in T cells.* Nat Immunol, 2011. **12**(7): p. 647-54.
81. Altman, A., N. Isakov, and G. Baier, *Protein kinase Ctheta: a new essential superstar on the T-cell stage.* Immunol Today, 2000. **21**(11): p. 567-73.
82. Jang, I.K., J. Zhang, and H. Gu, *Grb2, a simple adapter with complex roles in lymphocyte development, function, and signaling.* Immunol Rev, 2009. **232**(1): p. 150-9.
83. Songyang, Z., et al., *SH2 domains recognize specific phosphopeptide sequences.* Cell, 1993. **72**(5): p. 767-78.
84. Kolch, W., *Coordinating ERK/MAPK signalling through scaffolds and inhibitors.* Nat Rev Mol Cell Biol, 2005. **6**(11): p. 827-37.
85. Houtman, J.C., et al., *Oligomerization of signaling complexes by the multipoint binding of GRB2 to both LAT and SOS1.* Nat Struct Mol Biol, 2006. **13**(9): p. 798-805.
86. Lafferty, K.J., I.S. Misko, and M.A. Cooley, *Allogeneic stimulation modulates the in vitro response of T cells to transplantation antigen.* Nature, 1974. **249**(454): p. 275-6.
87. Riley, J.L., et al., *Modulation of TCR-induced transcriptional profiles by ligation of CD28, ICOS, and CTLA-4 receptors.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11790-5.
88. Bachmann, M.F., et al., *Distinct roles for LFA-1 and CD28 during activation of naive T cells: adhesion versus costimulation.* Immunity, 1997. **7**(4): p. 549-57.
89. Shahinian, A., et al., *Differential T cell costimulatory requirements in CD28-deficient mice.* Science, 1993. **261**(5121): p. 609-12.
90. Collins, A.V., et al., *The interaction properties of costimulatory molecules revisited.* Immunity, 2002. **17**(2): p. 201-10.
91. Evans, E.J., et al., *Crystal structure of a soluble CD28-Fab complex.* Nat Immunol, 2005. **6**(3): p. 271-9.

92. Peach, R.J., et al., *Complementarity determining region 1 (CDR1)- and CDR3-analogous regions in CTLA-4 and CD28 determine the binding to B7-1.* J Exp Med, 1994. **180**(6): p. 2049-58.
93. Ikemizu, S., et al., *Structure and dimerization of a soluble form of B7-1.* Immunity, 2000. **12**(1): p. 51-60.
94. Bhatia, S., et al., *Dynamic equilibrium of B7-1 dimers and monomers differentially affects immunological synapse formation and T cell activation in response to TCR/CD28 stimulation.* J Immunol, 2010. **184**(4): p. 1821-8.
95. Jansson, A., et al., *A theoretical framework for quantitative analysis of the molecular basis of costimulation.* J Immunol, 2005. **175**(3): p. 1575-85.
96. Wu, H., et al., *YXXM motifs in the PDGF-beta receptor serve dual roles as phosphoinositide 3-kinase binding motifs and tyrosine-based endocytic sorting signals.* J Biol Chem, 2003. **278**(42): p. 40425-8.
97. den Hertog, J., S. Tracy, and T. Hunter, *Phosphorylation of receptor protein-tyrosine phosphatase alpha on Tyr789, a binding site for the SH3-SH2-SH3 adaptor protein GRB-2 in vivo.* EMBO J, 1994. **13**(13): p. 3020-32.
98. Prasad, K.V., et al., *T-cell antigen CD28 interacts with the lipid kinase phosphatidylinositol 3-kinase by a cytoplasmic Tyr(P)-Met-Xaa-Met motif.* Proc Natl Acad Sci U S A, 1994. **91**(7): p. 2834-8.
99. Schneider, H., et al., *T cell antigen CD28 binds to the GRB-2/SOS complex, regulators of p21ras.* Eur J Immunol, 1995. **25**(4): p. 1044-50.
100. Wick, M.J., et al., *Mouse 3-phosphoinositide-dependent protein kinase-1 undergoes dimerization and trans-phosphorylation in the activation loop.* J Biol Chem, 2003. **278**(44): p. 42913-9.
101. Belham, C., S. Wu, and J. Avruch, *Intracellular signalling: PDK1--a kinase at the hub of things.* Curr Biol, 1999. **9**(3): p. R93-6.
102. Vanhaesebroeck, B., L. Stephens, and P. Hawkins, *PI3K signalling: the path to discovery and understanding.* Nat Rev Mol Cell Biol, 2012. **13**(3): p. 195-203.
103. Rudd, C.E. and H. Schneider, *Unifying concepts in CD28, ICOS and CTLA4 co-receptor signalling.* Nat Rev Immunol, 2003. **3**(7): p. 544-56.
104. Vazquez, F. and W.R. Sellers, *The PTEN tumor suppressor protein: an antagonist of phosphoinositide 3-kinase signaling.* Biochim Biophys Acta, 2000. **1470**(1): p. M21-35.
105. Dodson, L.F., et al., *Targeted knock-in mice expressing mutations of CD28 reveal an essential pathway for costimulation.* Mol Cell Biol, 2009. **29**(13): p. 3710-21.
106. Park, J., et al., *Identification of tyrosine phosphorylation sites on 3-phosphoinositide-dependent protein kinase-1 and their role in regulating kinase activity.* J Biol Chem, 2001. **276**(40): p. 37459-71.
107. Liu, Y., et al., *Regulation of protein kinase Ctheta function during T cell activation by Lck-mediated tyrosine phosphorylation.* J Biol Chem, 2000. **275**(5): p. 3603-9.
108. Lindsten, T., et al., *Characterization of CTLA-4 structure and expression on human T cells.* J Immunol, 1993. **151**(7): p. 3489-99.
109. Valk, E., C.E. Rudd, and H. Schneider, *CTLA-4 trafficking and surface expression.* Trends Immunol, 2008. **29**(6): p. 272-9.
110. Tivol, E.A., et al., *Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4.* Immunity, 1995. **3**(5): p. 541-7.
111. Calvo, C.R., D. Amsen, and A.M. Kruisbeek, *Cytotoxic T lymphocyte antigen 4 (CTLA-4) interferes with extracellular signal-regulated kinase (ERK) and Jun NH2-terminal kinase (JNK) activation, but does not affect phosphorylation of T cell receptor zeta and ZAP70.* J Exp Med, 1997. **186**(10): p. 1645-53.

112. Olsson, C., et al., *CTLA-4 ligation suppresses CD28-induced NF-kappaB and AP-1 activity in mouse T cell blasts.* J Biol Chem, 1999. **274**(20): p. 14400-5.
113. Schneider, H., et al., *CTLA-4 binding to the lipid kinase phosphatidylinositol 3-kinase in T cells.* J Exp Med, 1995. **181**(1): p. 351-5.
114. Marengere, L.E., et al., *Regulation of T cell receptor signaling by tyrosine phosphatase SYP association with CTLA-4.* Science, 1996. **272**(5265): p. 1170-3.
115. Kwon, H., et al., *Role of CTLA-4 in the activation of single- and double-positive thymocytes.* J Immunol, 2004. **173**(11): p. 6645-53.
116. Schneider, H. and C.E. Rudd, *Tyrosine phosphatase SHP-2 binding to CTLA-4: absence of direct YVKM/YFIP motif recognition.* Biochem Biophys Res Commun, 2000. **269**(1): p. 279-83.
117. Frearson, J.A. and D.R. Alexander, *The phosphotyrosine phosphatase SHP-2 participates in a multimeric signaling complex and regulates T cell receptor (TCR) coupling to the Ras/mitogen-activated protein kinase (MAPK) pathway in Jurkat T cells.* J Exp Med, 1998. **187**(9): p. 1417-26.
118. Baroja, M.L., et al., *The inhibitory function of CTLA-4 does not require its tyrosine phosphorylation.* J Immunol, 2000. **164**(1): p. 49-55.
119. Chuang, E., et al., *The CD28 and CTLA-4 receptors associate with the serine/threonine phosphatase PP2A.* Immunity, 2000. **13**(3): p. 313-22.
120. Janssens, V. and J. Goris, *Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling.* Biochem J, 2001. **353**(Pt 3): p. 417-39.
121. Baroja, M.L., et al., *Inhibition of CTLA-4 function by the regulatory subunit of serine/threonine phosphatase 2A.* J Immunol, 2002. **168**(10): p. 5070-8.
122. Masteller, E.L., et al., *Structural analysis of CTLA-4 function in vivo.* J Immunol, 2000. **164**(10): p. 5319-27.
123. Schneider, H., et al., *Reversal of the TCR stop signal by CTLA-4.* Science, 2006. **313**(5795): p. 1972-5.
124. Schneider, H., et al., *CTLA-4 disrupts ZAP70 microcluster formation with reduced T cell/APC dwell times and calcium mobilization.* Eur J Immunol, 2008. **38**(1): p. 40-7.
125. Qureshi, O.S., et al., *Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4.* Science, 2011. **332**(6029): p. 600-3.
126. Sheppard, K.A., et al., *PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3zeta signalosome and downstream signaling to PKCtheta.* FEBS Lett, 2004. **574**(1-3): p. 37-41.
127. Shlapatska, L.M., et al., *CD150 association with either the SH2-containing inositol phosphatase or the SH2-containing protein tyrosine phosphatase is regulated by the adaptor protein SH2D1A.* J Immunol, 2001. **166**(9): p. 5480-7.
128. Nishimura, H., et al., *Development of lupus-like autoimmune diseases by disruption of the PD-1 gene encoding an ITIM motif-carrying immunoreceptor.* Immunity, 1999. **11**(2): p. 141-51.
129. Nishimura, H., et al., *Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice.* Science, 2001. **291**(5502): p. 319-22.
130. Okazaki, T., et al., *Autoantibodies against cardiac troponin I are responsible for dilated cardiomyopathy in PD-1-deficient mice.* Nat Med, 2003. **9**(12): p. 1477-83.
131. Fife, B.T., et al., *Insulin-induced remission in new-onset NOD mice is maintained by the PD-1-PD-L1 pathway.* J Exp Med, 2006. **203**(12): p. 2737-47.
132. Sharpe, A.H., et al., *The function of programmed cell death 1 and its ligands in regulating autoimmunity and infection.* Nat Immunol, 2007. **8**(3): p. 239-45.

133. Salama, A.D., et al., *Critical role of the programmed death-1 (PD-1) pathway in regulation of experimental autoimmune encephalomyelitis.* J Exp Med, 2003. **198**(1): p. 71-8.
134. Freeman, G.J., et al., *Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation.* J Exp Med, 2000. **192**(7): p. 1027-34.
135. Carter, L., et al., *PD-1:PD-L inhibitory pathway affects both CD4(+) and CD8(+) T cells and is overcome by IL-2.* Eur J Immunol, 2002. **32**(3): p. 634-43.
136. Chemnitz, J.M., et al., *SHP-1 and SHP-2 associate with immunoreceptor tyrosine-based switch motif of programmed death 1 upon primary human T cell stimulation, but only receptor ligation prevents T cell activation.* J Immunol, 2004. **173**(2): p. 945-54.
137. Pentcheva-Hoang, T., et al., *Programmed death-1 concentration at the immunological synapse is determined by ligand affinity and availability.* Proc Natl Acad Sci U S A, 2007. **104**(45): p. 17765-70.
138. Fife, B.T., et al., *Interactions between PD-1 and PD-L1 promote tolerance by blocking the TCR-induced stop signal.* Nat Immunol, 2009. **10**(11): p. 1185-92.
139. Watanabe, N., et al., *BTLA is a lymphocyte inhibitory receptor with similarities to CTLA-4 and PD-1.* Nat Immunol, 2003. **4**(7): p. 670-9.
140. Hurchla, M.A., et al., *B and T lymphocyte attenuator exhibits structural and expression polymorphisms and is highly Induced in anergic CD4+ T cells.* J Immunol, 2005. **174**(6): p. 3377-85.
141. Steinberg, M.W., et al., *A crucial role for HVEM and BTLA in preventing intestinal inflammation.* J Exp Med, 2008. **205**(6): p. 1463-76.
142. Oya, Y., et al., *Development of autoimmune hepatitis-like disease and production of autoantibodies to nuclear antigens in mice lacking B and T lymphocyte attenuator.* Arthritis Rheum, 2008. **58**(8): p. 2498-510.
143. Sedy, J.R., et al., *B and T lymphocyte attenuator regulates T cell activation through interaction with herpesvirus entry mediator.* Nat Immunol, 2005. **6**(1): p. 90-8.
144. Steinberg, M.W., T.C. Cheung, and C.F. Ware, *The signaling networks of the herpesvirus entry mediator (TNFRSF14) in immune regulation.* Immunol Rev. **244**(1): p. 169-87.
145. Gonzalez, L.C., et al., *A coreceptor interaction between the CD28 and TNF receptor family members B and T lymphocyte attenuator and herpesvirus entry mediator.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1116-21.
146. Gavrieli, M., et al., *Characterization of phosphotyrosine binding motifs in the cytoplasmic domain of B and T lymphocyte attenuator required for association with protein tyrosine phosphatases SHP-1 and SHP-2.* Biochem Biophys Res Commun, 2003. **312**(4): p. 1236-43.
147. Wu, T.H., et al., *B and T lymphocyte attenuator interacts with CD3zeta and inhibits tyrosine phosphorylation of TCRzeta complex during T-cell activation.* Immunol Cell Biol, 2007. **85**(8): p. 590-5.
148. Hurchla, M.A., J.R. Sedy, and K.M. Murphy, *Unexpected role of B and T lymphocyte attenuator in sustaining cell survival during chronic allostimulation.* J Immunol, 2007. **178**(10): p. 6073-82.
149. Yoshinaga, S.K., et al., *T-cell co-stimulation through B7RP-1 and ICOS.* Nature, 1999. **402**(6763): p. 827-32.
150. Swallow, M.M., J.J. Wallin, and W.C. Sha, *B7h, a novel costimulatory homolog of B7.1 and B7.2, is induced by TNFalpha.* Immunity, 1999. **11**(4): p. 423-32.
151. Mahajan, S., et al., *The role of ICOS in the development of CD4 T cell help and the reactivation of memory T cells.* Eur J Immunol, 2007. **37**(7): p. 1796-808.

152. McAdam, A.J., et al., *Mouse inducible costimulatory molecule (ICOS) expression is enhanced by CD28 costimulation and regulates differentiation of CD4+ T cells.* J Immunol, 2000. **165**(9): p. 5035-40.
153. Tan, A.H., S.C. Wong, and K.P. Lam, *Regulation of mouse inducible costimulator (ICOS) expression by Fyn-NFATc2 and ERK signaling in T cells.* J Biol Chem, 2006. **281**(39): p. 28666-78.
154. Dong, C., et al., *ICOS co-stimulatory receptor is essential for T-cell activation and function.* Nature, 2001. **409**(6816): p. 97-101.
155. Tafuri, A., et al., *ICOS is essential for effective T-helper-cell responses.* Nature, 2001. **409**(6816): p. 105-9.
156. McAdam, A.J., et al., *ICOS is critical for CD40-mediated antibody class switching.* Nature, 2001. **409**(6816): p. 102-5.
157. Grimbacher, B., et al., *Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency.* Nat Immunol, 2003. **4**(3): p. 261-8.
158. Hutloff, A., et al., *ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28.* Nature, 1999. **397**(6716): p. 263-6.
159. Vieira, P.L., et al., *ICOS-mediated signaling regulates cytokine production by human T cells and provides a unique signal to selectively control the clonal expansion of Th2 helper cells.* Eur J Immunol, 2004. **34**(5): p. 1282-90.
160. Arimura, Y., et al., *A co-stimulatory molecule on activated T cells, H4/ICOS, delivers specific signals in T(h) cells and regulates their responses.* Int Immunol, 2002. **14**(6): p. 555-66.
161. Nurieva, R.I., et al., *Transcriptional regulation of th2 differentiation by inducible costimulator.* Immunity, 2003. **18**(6): p. 801-11.
162. Watanabe, M., et al., *ICOS-mediated costimulation on Th2 differentiation is achieved by the enhancement of IL-4 receptor-mediated signaling.* J Immunol, 2005. **174**(4): p. 1989-96.
163. Gigoux, M., et al., *Inducible costimulator promotes helper T-cell differentiation through phosphoinositide 3-kinase.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20371-6.
164. Fos, C., et al., *ICOS ligation recruits the p50alpha PI3K regulatory subunit to the immunological synapse.* J Immunol, 2008. **181**(3): p. 1969-77.
165. Linterman, M.A., et al., *Roquin differentiates the specialized functions of duplicated T cell costimulatory receptor genes CD28 and ICOS.* Immunity, 2009. **30**(2): p. 228-41.
166. Eckhart, W., M.A. Hutchinson, and T. Hunter, *An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates.* Cell, 1979. **18**(4): p. 925-33.
167. Hunter, T., *Tyrosine phosphorylation: thirty years and counting.* Curr Opin Cell Biol, 2009. **21**(2): p. 140-6.
168. Liu, B.A., et al., *The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes.* Sci Signal, 2011. **4**(202): p. ra83.
169. Liu, B.A., et al., *The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling.* Mol Cell, 2006. **22**(6): p. 851-68.
170. Lim, W.A. and T. Pawson, *Phosphotyrosine signaling: evolving a new cellular communication system.* Cell, 2010. **142**(5): p. 661-7.
171. Pincus, D., et al., *Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages.* Proc Natl Acad Sci U S A, 2008. **105**(28): p. 9680-4.
172. Dengl, S., et al., *Structure and in vivo requirement of the yeast Spt6 SH2 domain.* J Mol Biol, 2009. **389**(1): p. 211-25.

173. King, N., et al., *The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans.* Nature, 2008. **451**(7180): p. 783-8.
174. Pawson, T., *Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems.* Cell, 2004. **116**(2): p. 191-203.
175. Smith, M.J., et al., *Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays.* Mol Cell Biol, 2006. **26**(22): p. 8461-74.
176. DiNitto, J.P. and D.G. Lambright, *Membrane and juxtamembrane targeting by PH and PTB domains.* Biochim Biophys Acta, 2006. **1761**(8): p. 850-67.
177. Uhlik, M.T., et al., *Structural and evolutionary division of phosphotyrosine binding (PTB) domains.* J Mol Biol, 2005. **345**(1): p. 1-20.
178. Waksman, G., et al., *Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides.* Nature, 1992. **358**(6388): p. 646-53.
179. Waksman, G., et al., *Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms.* Cell, 1993. **72**(5): p. 779-90.
180. Machida, K. and B.J. Mayer, *The SH2 domain: versatile signaling module and pharmaceutical target.* Biochim Biophys Acta, 2005. **1747**(1): p. 1-25.
181. Henriques, D.A. and J.E. Ladbury, *Inhibitors to the Src SH2 domain: a lesson in structure--thermodynamic correlation in drug design.* Arch Biochem Biophys, 2001. **390**(2): p. 158-68.
182. Bradshaw, J.M., V. Mitaxov, and G. Waksman, *Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase.* J Mol Biol, 1999. **293**(4): p. 971-85.
183. Bruggeman, F.J. and H.V. Westerhoff, *The nature of systems biology.* Trends Microbiol, 2007. **15**(1): p. 45-50.
184. Kitano, H., *Systems biology: a brief overview.* Science, 2002. **295**(5560): p. 1662-4.
185. Chuang, H.Y., M. Hofree, and T. Ideker, *A decade of systems biology.* Annu Rev Cell Dev Biol. **26**: p. 721-44.
186. Kholodenko, B.N., et al., *Quantification of short term signaling by the epidermal growth factor receptor.* J Biol Chem, 1999. **274**(42): p. 30169-81.
187. O'Rourke, L. and J.E. Ladbury, *Specificity is complex and time consuming: mutual exclusivity in tyrosine kinase-mediated signaling.* Acc Chem Res, 2003. **36**(6): p. 410-6.
188. Ladbury, J.E. and S. Arold, *Searching for specificity in SH domains.* Chem Biol, 2000. **7**(1): p. R3-8.
189. Kolch, W. and A. Pitt, *Functional proteomics to dissect tyrosine kinase signalling pathways in cancer.* Nat Rev Cancer. **10**(9): p. 618-29.
190. Citri, A. and Y. Yarden, *EGF-ERBB signalling: towards the systems level.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 505-16.
191. Jones, R.B., et al., *A quantitative protein interaction network for the ErbB receptors using protein microarrays.* Nature, 2006. **439**(7073): p. 168-74.
192. Breitkreutz, A., et al., *A global protein kinase and phosphatase interaction network in yeast.* Science, 2010 **328**(5981): p. 1043-6.
193. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
194. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet. **13**(4): p. 227-32.
195. Bertone, P., et al., *Global identification of human transcribed sequences with genome tiling arrays.* Science, 2004. **306**(5705): p. 2242-6.
196. Velculescu, V.E., et al., *Serial analysis of gene expression.* Science, 1995. **270**(5235): p. 484-7.

197. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet. **11**(1): p. 31-46.
198. Saha, S., et al., *Using the transcriptome to annotate the genome.* Nat Biotechnol, 2002. **20**(5): p. 508-12.
199. Matsumura, H., et al., *SuperSAGE array: the direct use of 26-base-pair transcript tags in oligonucleotide arrays.* Nat Methods, 2006. **3**(6): p. 469-74.
200. Hene, L., et al., *Deep analysis of cellular transcriptomes - LongSAGE versus classic MPSS.* BMC Genomics, 2007. **8**: p. 333.
201. Wilhelm, B.T., et al., *Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution.* Nature, 2008. **453**(7199): p. 1239-43.
202. de Sousa Abreu, R., et al., *Global signatures of protein and mRNA expression levels.* Mol Biosyst, 2009. **5**(12): p. 1512-26.
203. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature. **473**(7347): p. 337-42.
204. Vogel, C., et al., *Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line.* Mol Syst Biol. **6**: p. 400.
205. Uhlen, M. and S. Hober, *Generation and validation of affinity reagents on a proteome-wide level.* J Mol Recognit, 2009. **22**(2): p. 57-64.
206. Cox, J. and M. Mann, *Quantitative, high-resolution proteomics for data-driven systems biology.* Annu Rev Biochem. **80**: p. 273-99.
207. Maier, T., M. Guell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Lett, 2009. **583**(24): p. 3966-73.
208. Pertea, M. and S.L. Salzberg, *Between a chicken and a grape: estimating the number of human genes.* Genome Biol, 2010. **11**(5): p. 206.
209. de Godoy, L.M., et al., *Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.* Nature, 2008. **455**(7217): p. 1251-4.
210. Picotti, P., et al., *Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics.* Cell, 2009. **138**(4): p. 795-806.
211. Malmstrom, J., et al., *Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans.* Nature, 2009. **460**(7256): p. 762-5.
212. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
213. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
214. Mann, M., *Functional and quantitative proteomics using SILAC.* Nat Rev Mol Cell Biol, 2006. **7**(12): p. 952-8.
215. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
216. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
217. Steen, H., et al., *Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 3948-53.
218. Singh, S., et al., *FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides.* J Proteome Res, 2009. **8**(5): p. 2201-10.
219. Picotti, P., et al., *High-throughput generation of selected reaction-monitoring assays for proteins and proteomes.* Nat Methods. **7**(1): p. 43-6.

220. Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expr Purif, 2007. **55**(2): p. 209-24.
221. Vincentelli, R., et al., *High-throughput automated refolding screening of inclusion bodies.* Protein Sci, 2004. **13**(10): p. 2782-92.
222. Vuillard, L., T. Rabilloud, and M.E. Goldberg, *Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation.* Eur J Biochem, 1998. **256**(1): p. 128-35.
223. Willis, M.S., et al., *Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions.* Protein Sci, 2005. **14**(7): p. 1818-26.
224. Pawson, T., G.D. Gish, and P. Nash, *SH2 domains, interaction modules and cellular wiring.* Trends Cell Biol, 2001. **11**(12): p. 504-11.
225. Yaffe, M.B., *Phosphotyrosine-binding domains in signal transduction.* Nat Rev Mol Cell Biol, 2002. **3**(3): p. 177-86.
226. Geiger, T., et al., *Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins.* Mol Cell Proteomics, 2012. **11**(3): p. M111 014050.
227. Shires, J., E. Theodoridis, and A.C. Hayday, *Biological insights into TCRgammadelta+ and TCRalphabeta+ intraepithelial lymphocytes provided by serial analysis of gene expression (SAGE).* Immunity, 2001. **15**(3): p. 419-34.
228. Nagai, S., et al., *Comprehensive gene expression profile of human activated T(h)1- and T(h)2-polarized cells.* Int Immunol, 2001. **13**(3): p. 367-76.
229. Evans, E.J., et al., *The T cell surface--how well do we know it?* Immunity, 2003. **19**(2): p. 213-23.
230. Hashimoto, S., et al., *Serial analysis of gene expression in human monocyte-derived dendritic cells.* Blood, 1999. **94**(3): p. 845-52.
231. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-7.
232. Arike, L., et al., *Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli.* J Proteomics, 2012. **75**(17): p. 5437-48.
233. Gillis, S. and J. Watson, *Biochemical and biological characterization of lymphocyte regulatory molecules. V. Identification of an interleukin 2-producing human leukemia T cell line.* J Exp Med, 1980. **152**(6): p. 1709-19.
234. Abraham, R.T. and A. Weiss, *Jurkat T cells and development of the T-cell receptor signalling paradigm.* Nat Rev Immunol, 2004. **4**(4): p. 301-8.
235. Seminario, M.C. and R.L. Wange, *Signaling pathways of D3-phosphoinositide-binding kinases in T cells and their regulation by PTEN.* Semin Immunol, 2002. **14**(1): p. 27-36.
236. Costello, P.S., M. Gallagher, and D.A. Cantrell, *Sustained and dynamic inositol lipid metabolism inside and outside the immunological synapse.* Nat Immunol, 2002. **3**(11): p. 1082-9.
237. Elbashir, S.M., et al., *Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.* Nature, 2001. **411**(6836): p. 494-8.
238. Otsu, M., et al., *Reconstitution of lymphoid development and function in ZAP-70-deficient mice following gene transfer into bone marrow cells.* Blood, 2002. **100**(4): p. 1248-56.
239. Chan, A.C., et al., *ZAP-70 deficiency in an autosomal recessive form of severe combined immunodeficiency.* Science, 1994. **264**(5165): p. 1599-601.
240. Abidi, S.H., et al., *Differential remodeling of a T-cell transcriptome following CD8- versus CD3-induced signaling.* Cell Res, 2008. **18**(6): p. 641-8.

241. Hashimoto, S., et al., *Gene expression profile in human leukocytes.* Blood, 2003. **101**(9): p. 3509-13.
242. Lal, A., et al., *A public database for gene expression in human cancers.* Cancer Res, 1999. **59**(21): p. 5403-7.
243. Lash, A.E., et al., *SAGEmap: a public gene expression resource.* Genome Res, 2000. **10**(7): p. 1051-60.
244. Fukushima, A., et al., *Lck couples Shc to TCR signaling.* Cell Signal, 2006. **18**(8): p. 1182-9.
245. Nguyen, T.V., et al., *Conditional deletion of Shp2 tyrosine phosphatase in thymocytes suppresses both pre-TCR and TCR signals.* J Immunol, 2006. **177**(9): p. 5990-6.
246. Tomlinson, M.G., et al., *Expression and function of Tec, Itk, and Btk in lymphocytes: evidence for a unique role for Tec.* Mol Cell Biol, 2004. **24**(6): p. 2455-66.
247. Hirao, A., et al., *Translocation of the Csk homologous kinase (Chk/Hyl) controls activity of CD36-anchored Lyn tyrosine kinase in thrombin-stimulated platelets.* EMBO J, 1997. **16**(9): p. 2342-51.
248. Holland, S.J., et al., *Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling.* J Exp Med, 2001. **194**(9): p. 1263-76.
249. Chow, L.M., et al., *Negative regulation of T-cell receptor signalling by tyrosine protein kinase p50csk.* Nature, 1993. **365**(6442): p. 156-60.
250. Tang, B., et al., *T cell receptor signaling induced by an analog peptide of type II collagen requires activation of Syk.* Clin Immunol, 2009. **133**(1): p. 145-53.
251. Freeburn, R.W., et al., *Evidence that SHIP-1 contributes to phosphatidylinositol 3,4,5-trisphosphate metabolism in T lymphocytes and can regulate novel phosphoinositide 3-kinase effectors.* J Immunol, 2002. **169**(10): p. 5441-50.
252. Kwok, E., et al., *FES Kinase Promotes Mast Cell Recruitment to Mammary Tumors via the Stem Cell Factor/KIT Receptor Signaling Axis.* Mol Cancer Res, 2012. **10**(7): p. 881-91.
253. Jin, D., et al., *STAT6 signalling is important in CD8 T-cell activation and defence against Toxoplasma gondii infection in the brain.* Immunology, 2009. **127**(2): p. 187-95.
254. Marth, J.D., et al., *Regulation of pp56lck during T-cell activation: functional implications for the src-like protein tyrosine kinases.* EMBO J, 1987. **6**(9): p. 2727-34.
255. Dragone, L.L., et al., *SLAP, a regulator of immunoreceptor ubiquitination, signaling, and trafficking.* Immunol Rev, 2009. **232**(1): p. 218-28.
256. Isakov, N., et al., *ZAP-70 binding specificity to T cell receptor tyrosine-based activation motifs: the tandem SH2 domains of ZAP-70 bind distinct tyrosine-based activation motifs with varying affinity.* J Exp Med, 1995. **181**(1): p. 375-80.
257. Lindholm, C.K., *IL-2 receptor signaling through the Shb adapter protein in T and NK cells.* Biochem Biophys Res Commun, 2002. **296**(4): p. 929-36.
258. Lin, J.X., et al., *Critical Role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function.* Immunity, 2012. **36**(4): p. 586-99.
259. Brockdorff, J., et al., *Dephosphorylation of ZAP-70 and inhibition of T cell activation by activated SHP1.* Eur J Immunol, 1999. **29**(8): p. 2539-50.
260. Sela, M., et al., *Sequential phosphorylation of SLP-76 at tyrosine 173 is required for activation of T and mast cells.* EMBO J, 2011. **30**(15): p. 3160-72.
261. Dai, K.Z., et al., *The T cell regulator gene SH2D2A contributes to the genetic susceptibility of multiple sclerosis.* Genes Immun, 2001. **2**(5): p. 263-8.
262. Guiet, R., et al., *Hematopoietic cell kinase (Hck) isoforms and phagocyte duties - from signaling and actin reorganization to migration and phagocytosis.* Eur J Cell Biol, 2008. **87**(8-9): p. 527-42.

263. Hamada, F., et al., *Association of immunoglobulin G Fc receptor II with Src-like protein-tyrosine kinase Fgr in neutrophils.* Proc Natl Acad Sci U S A, 1993. **90**(13): p. 6305-9.
264. Chu, P., et al., *Systematic identification of regulatory proteins critical for T-cell activation.* J Biol, 2003. **2**(3): p. 21.
265. Chang, H.C., et al., *Impaired development of human Th1 cells in patients with deficient expression of STAT4.* Blood, 2009. **113**(23): p. 5887-90.
266. Chan, B., et al., *SAP couples Fyn to SLAM immune receptors.* Nat Cell Biol, 2003. **5**(2): p. 155-60.
267. Tybulewicz, V.L., *Vav-family proteins in T-cell signalling.* Curr Opin Immunol, 2005. **17**(3): p. 267-74.
268. Qu, X., et al., *Tyrosine phosphorylation of adaptor protein 3BP2 induces T cell receptor-mediated activation of transcription factor.* Biochemistry, 2005. **44**(10): p. 3891-8.
269. Ashe, J.M., et al., *ZAP-70 protein promotes tyrosine phosphorylation of T cell receptor signaling motifs (ITAMs) in immature CD4(+)8(+) thymocytes with limiting p56(lck).* J Exp Med, 1999. **189**(7): p. 1163-8.
270. Paolino, M., et al., *Essential role of E3 ubiquitin ligase activity in Cbl-b-regulated T cell functions.* J Immunol, 2011. **186**(4): p. 2138-47.
271. Laird, R.M. and S.M. Hayes, *Roles of the Src tyrosine kinases Lck and Fyn in regulating gammadeltaTCR signal strength.* PLoS One, 2010. **5**(1): p. e8899.
272. Proust, R., J. Bertoglio, and F. Gesbert, *The Adaptor Protein SAP Directly Associates with CD3zeta Chain and Regulates T Cell Receptor Signaling.* PLoS One, 2012. **7**(8): p. e43200.
273. Debeb, B.G., et al., *Characterizing cancer cells with cancer stem cell-like features in 293T human embryonic kidney cells.* Mol Cancer, 2010. **9**: p. 180.
274. Trudgian, D.C., et al., *Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline.* Proteomics, 2011. **11**(14): p. 2790-7.
275. Kohl, T., et al., *Automated production of recombinant human proteins as resource for proteome research.* Proteome Sci, 2008. **6**: p. 4.
276. Hammarstrom, M., et al., *Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli.* Protein Sci, 2002. **11**(2): p. 313-21.
277. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnol Adv, 2009. **27**(3): p. 297-306.
278. Nettleship, J.E., et al., *Recent advances in the production of proteins in insect and mammalian cells for structural biology.* J Struct Biol. **172**(1): p. 55-65.
279. Kaneko, T., et al., *Loops govern SH2 domain specificity by controlling access to binding pockets.* Sci Signal. **3**(120): p. ra34.
280. Bae, J.H., et al., *The selectivity of receptor tyrosine kinase signaling is controlled by a secondary SH2 domain binding site.* Cell, 2009. **138**(3): p. 514-24.
281. Donaldson, L.W., et al., *Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14053-8.
282. Vera, A., et al., *The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures.* Biotechnol Bioeng, 2007. **96**(6): p. 1101-6.
283. Martinez-Alonso, M., A. Vera, and A. Villaverde, *Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells.* FEMS Microbiol Lett, 2007. **273**(2): p. 187-95.

284. Middelberg, A.P., *Preparative protein refolding.* Trends Biotechnol, 2002. **20**(10): p. 437-43.
285. Martinez-Alonso, M., et al., *Side effects of chaperone gene co-expression in recombinant protein production.* Microb Cell Fact. **9**: p. 64.
286. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
287. di Guan, C., et al., *Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein.* Gene, 1988. **67**(1): p. 21-30.
288. LaVallie, E.R., et al., *A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm.* Biotechnology (N Y), 1993. **11**(2): p. 187-93.
289. Nilsson, B., et al., *A synthetic IgG-binding domain based on staphylococcal protein A.* Protein Eng, 1987. **1**(2): p. 107-13.
290. Davis, G.D., et al., *New fusion protein systems designed to give soluble expression in Escherichia coli.* Biotechnol Bioeng, 1999. **65**(4): p. 382-8.
291. Huth, J.R., et al., *Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR.* Protein Sci, 1997. **6**(11): p. 2359-64.
292. Marblestone, J.G., et al., *Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO.* Protein Sci, 2006. **15**(1): p. 182-9.
293. Su, Y., et al., *The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in Escherichia coli.* J Biotechnol, 2007. **129**(3): p. 373-82.
294. Zou, Z., et al., *Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in Escherichia coli.* J Biotechnol, 2008. **135**(4): p. 333-9.
295. Nishihara, K., et al., *Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli.* Appl Environ Microbiol, 2000. **66**(3): p. 884-9.
296. Robin, S., et al., *Trigger factor from the psychrophilic bacterium Psychrobacter frigidicola is a monomeric chaperone.* J Bacteriol, 2009. **191**(4): p. 1162-8.
297. Ohana, R.F., et al., *HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification.* Protein Expr Purif, 2009. **68**(1): p. 110-20.
298. Berrow, N.S., et al., *A versatile ligation-independent cloning method suitable for high-throughput expression screening applications.* Nucleic Acids Res, 2007. **35**(6): p. e45.
299. *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Res. **40**(Database issue): p. D71-5.
300. Yang, Z.R., et al., *RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.* Bioinformatics, 2005. **21**(16): p. 3369-76.
301. Derewenda, Z.S., *The use of recombinant methods and molecular engineering in protein crystallization.* Methods, 2004. **34**(3): p. 354-63.
302. Crowe, J., et al., *6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification.* Methods Mol Biol, 1994. **31**: p. 371-87.
303. Nygren, P.A., S. Stahl, and M. Uhlen, *Engineering proteins to facilitate bioprocessing.* Trends Biotechnol, 1994. **12**(5): p. 184-8.

304. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.
305. Dyson, M.R., et al., *Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression.* BMC Biotechnol, 2004. **4**: p. 32.
306. Kataeva, I., et al., *Improving solubility of Shewanella oneidensis MR-1 and Clostridium thermocellum JW-20 proteins expressed into Esherichia coli.* J Proteome Res, 2005. **4**(6): p. 1942-51.
307. Kapust, R.B. and D.S. Waugh, *Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.* Protein Sci, 1999. **8**(8): p. 1668-74.
308. Dummler, A., A.M. Lawrence, and A. de Marco, *Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors.* Microb Cell Fact, 2005. **4**: p. 34.
309. Turner, P., O. Holst, and E.N. Karlsson, *Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration.* Protein Expr Purif, 2005. **39**(1): p. 54-60.
310. Korf, U., et al., *Large-scale protein expression for proteome research.* Proteomics, 2005. **5**(14): p. 3571-80.
311. Chen, X., et al., *Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA.* Cell, 1998. **93**(5): p. 827-39.
312. Meng, W., et al., *Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase.* Nature, 1999. **398**(6722): p. 84-90.
313. Zhou, M.M., et al., *Solution structure of the Shc SH2 domain complexed with a tyrosine-phosphorylated peptide from the T-cell receptor.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7784-8.
314. Kim, Y.S., et al., *High-pressure studies on protein aggregates and amyloid fibrils.* Methods Enzymol, 2006. **413**: p. 237-53.
315. Audic, S., et al., *SAmBA: an interactive software for optimizing the design of biological macromolecules crystallization experiments.* Proteins, 1997. **29**(2): p. 252-7.
316. Kubo, M., T. Hanada, and A. Yoshimura, *Suppressors of cytokine signaling and immunity.* Nat Immunol, 2003. **4**(12): p. 1169-76.
317. Li, Y., et al., *Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation.* J Immunol, 2000. **164**(10): p. 5199-206.
318. Okkenhaug, K. and B. Vanhaesebroeck, *PI3K in lymphocyte development, differentiation and activation.* Nat Rev Immunol, 2003. **3**(4): p. 317-30.
319. Vervoort, V.S., et al., *Splice variants and expression patterns of SHEP1, BCAR3 and NSP1, a gene family involved in integrin and receptor tyrosine kinase signaling.* Gene, 2007. **391**(1-2): p. 161-70.
320. Kim, L. and T.W. Wong, *The cytoplasmic tyrosine kinase FER is associated with the catenin-like substrate pp120 and is activated by growth factors.* Mol Cell Biol, 1995. **15**(8): p. 4553-61.
321. Suliman, S.G., et al., *Severe insulin resistance and intrauterine growth deficiency associated with haploinsufficiency for INSR and CHN2: new insights into synergistic pathways involved in growth and metabolism.* Diabetes, 2009. **58**(12): p. 2954-61.

322. Yoshikawa, M., et al., *Tyr-phosphorylation signals translocate RIN3, the small GTPase Rab5-GEF, to early endocytic vesicles.* Biochem Biophys Res Commun, 2008. **372**(1): p. 168-72.
323. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.
324. van der Merwe, P.A. and A.N. Barclay, *Analysis of cell-adhesion molecule interactions using surface plasmon resonance.* Curr Opin Immunol, 1996. **8**(2): p. 257-61.
325. Richens, J.L., et al., *Systems biology coupled with label-free high-throughput detection as a novel approach for diagnosis of chronic obstructive pulmonary disease.* Respir Res, 2009. **10**: p. 29.
326. PA, v.d.M., *Surface Plasmon Resonance in Protein-Ligand interactions: hydrodynamics and calorimetry.* Oxford University Press, 2001: p. 137-170.
327. Esa Stenberg, B.P., Håkan Roos, Csaba Urbaniczky, *Quantitativedetermination of surfaceconcentration of protein with surfaceplasmonresonance using radiolabeledproteins.* Journal of Colloid and Interface Science, 1991. **143**(2): p. 513–526.
328. Muller, K.M., K.M. Arndt, and A. Pluckthun, *Model and simulation of multivalent binding to fixed ligands.* Anal Biochem, 1998. **261**(2): p. 149-58.
329. Catimel, B., et al., *Kinetic analysis of the interaction between the monoclonal antibody A33 and its colonic epithelial antigen by the use of an optical biosensor. A comparison of immobilisation strategies.* J Chromatogr A, 1997. **776**(1): p. 15-30.
330. Pabbisetty, K.B., et al., *Kinetic analysis of the binding of monomeric and dimeric ephrins to Eph receptors: correlation to function in a growth cone collapse assay.* Protein Sci, 2007. **16**(3): p. 355-61.
331. Gimferrer, I., et al., *Relevance of CD6-mediated interactions in T cell activation and proliferation.* J Immunol, 2004. **173**(4): p. 2262-70.
332. Hassan, N.J., et al., *CD6 regulates T-cell responses through activation-dependent recruitment of the positive regulator SLP-76.* Mol Cell Biol, 2006. **26**(17): p. 6727-38.
333. Li, S.C., et al., *Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A.* Curr Biol, 1999. **9**(23): p. 1355-62.
334. Huang, H., et al., *Defining the specificity space of the human SRC homology 2 domain.* Mol Cell Proteomics, 2008. **7**(4): p. 768-84.
335. Songyang, Z., et al., *Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav.* Mol Cell Biol, 1994. **14**(4): p. 2777-85.
336. Kay, L.E., et al., *Correlation between binding and dynamics at SH2 domain interfaces.* Nat Struct Biol, 1998. **5**(2): p. 156-63.
337. Murphy, T.L. and K.M. Murphy, *Slow down and survive: Enigmatic immunoregulation by BTLA and HVEM.* Annu Rev Immunol, 2010. **28**: p. 389-411.
338. Caligiuri, M.A., et al., *Novel c-CBL and CBL-b ubiquitin ligase mutations in human acute myeloid leukemia.* Blood, 2007. **110**(3): p. 1022-4.
339. Superti-Furga, G., et al., *Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src.* EMBO J, 1993. **12**(7): p. 2625-34.
340. Muller, K., et al., *Rapid identification of phosphopeptide ligands for SH2 domains. Screening of peptide libraries by fluorescence-activated bead sorting.* J Biol Chem, 1996. **271**(28): p. 16500-5.
341. Exley, M., et al., *Association of phosphatidylinositol 3-kinase with a specific sequence of the T cell receptor zeta chain is dependent on T cell activation.* J Biol Chem, 1994. **269**(21): p. 15140-6.
342. Rudd, C.E., *Upstream-downstream: CD28 cosignaling pathways and T cell function.* Immunity, 1996. **4**(6): p. 527-34.

343. Bergman, M., et al., *The human p50csk tyrosine kinase phosphorylates p56lck at Tyr-505 and down regulates its catalytic activity.* EMBO J, 1992. **11**(8): p. 2919-24.
344. Ettenberg, S.A., et al., *cbl-b inhibits EGF-receptor-induced apoptosis by enhancing ubiquitination and degradation of activated receptors.* Mol Cell Biol Res Commun, 1999. **2**(2): p. 111-8.
345. Greenwald, R.J., G.J. Freeman, and A.H. Sharpe, *The B7 family revisited.* Annu Rev Immunol, 2005. **23**: p. 515-48.
346. Kholodenko, B.N., J.B. Hoek, and H.V. Westerhoff, *Why cytoplasmic signalling proteins should be recruited to cell membranes.* Trends Cell Biol, 2000. **10**(5): p. 173-8.
347. James, J.R., et al., *The T cell receptor triggering apparatus is composed of monovalent or monomeric proteins.* J Biol Chem, 2011. **286**(37): p. 31993-2001.
348. Pacini, S., et al., *Temporally regulated assembly of a dynamic signaling complex associated with the activated TCR.* Eur J Immunol, 2000. **30**(9): p. 2620-31.
349. Petricoin, E.F., 3rd, et al., *Antiproliferative action of interferon-alpha requires components of T-cell-receptor signalling.* Nature, 1997. **390**(6660): p. 629-32.
350. Liu, B.A., et al., *SH2 domains recognize contextual peptide sequence information to determine selectivity.* Mol Cell Proteomics, 2010. **9**(11): p. 2391-404.
351. Wang, J., et al., *Quantifying intrinsic specificity: a potential complement to affinity in drug screening.* Phys Rev Lett, 2007. **99**(19): p. 198101.
352. Boggon, T.J. and M.J. Eck, *Structure and regulation of Src family kinases.* Oncogene, 2004. **23**(48): p. 7918-27.
353. Davis, S.J. and P.A. van der Merwe, *Lck and the nature of the T cell receptor trigger.* Trends Immunol, 2011. **32**(1): p. 1-5.
354. Nurse, P., *A long twentieth century of the cell cycle and beyond.* Cell, 2000. **100**(1): p. 71-8.
355. Alegre, M.L., K.A. Frauwirth, and C.B. Thompson, *T-cell regulation by CD28 and CTLA-4.* Nat Rev Immunol, 2001. **1**(3): p. 220-8.
356. Qureshi, O.S., et al., *Constitutive clathrin-mediated endocytosis of CTLA-4 persists during T cell activation.* J Biol Chem, 2012. **287**(12): p. 9429-40.
357. Weiss, A., et al., *Role of T3 surface molecules in human T-cell activation: T3-dependent activation results in an increase in cytoplasmic free calcium.* Proc Natl Acad Sci U S A, 1984. **81**(13): p. 4169-73.
358. Imboden, J.B. and J.D. Stobo, *Transmembrane signalling by the T cell antigen receptor. Perturbation of the T3-antigen receptor complex generates inositol phosphates and releases calcium ions from intracellular stores.* J Exp Med, 1985. **161**(3): p. 446-56.
359. June, C.H., et al., *Increases in tyrosine phosphorylation are detectable before phospholipase C activation after T cell receptor stimulation.* J Immunol, 1990. **144**(5): p. 1591-9.
360. June, C.H., et al., *Inhibition of tyrosine phosphorylation prevents T-cell receptor-mediated signal transduction.* Proc Natl Acad Sci U S A, 1990. **87**(19): p. 7722-6.
361. Chan, A.C., et al., *ZAP-70: a 70 kd protein-tyrosine kinase that associates with the TCR zeta chain.* Cell, 1992. **71**(4): p. 649-62.
362. Zhang, W., et al., *LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation.* Cell, 1998. **92**(1): p. 83-92.
363. Coyle, A.J., et al., *The CD28-related molecule ICOS is required for effective T cell-dependent immune responses.* Immunity, 2000. **13**(1): p. 95-105.
364. Stein, P.H., J.D. Fraser, and A. Weiss, *The cytoplasmic domain of CD28 is both necessary and sufficient for costimulation of interleukin-2 secretion and association with phosphatidylinositol 3'-kinase.* Mol Cell Biol, 1994. **14**(5): p. 3392-402.

365. Stelling, J., *Mathematical models in microbial systems biology.* Curr Opin Microbiol, 2004. **7**(5): p. 513-8.
366. Palacios, E.H. and A. Weiss, *Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation.* Oncogene, 2004. **23**(48): p. 7990-8000.
367. Veillette, A. and M. Fournel, *The CD4 associated tyrosine protein kinase p56lck is positively regulated through its site of autophosphorylation.* Oncogene, 1990. **5**(10): p. 1455-62.
368. Nika, K., et al., *Constitutively active Lck kinase in T cells drives antigen receptor signal transduction.* Immunity, 2010. **32**(6): p. 766-77.
369. Guy, C.S. and D.A. Vignali, *Organization of proximal signal initiation at the TCR:CD3 complex.* Immunol Rev, 2009. **232**(1): p. 7-21.
370. Alexandropoulos, K. and D. Baltimore, *Coordinate activation of c-Src by SH3- and SH2-binding sites on a novel p130Cas-related protein, Sin.* Genes Dev, 1996. **10**(11): p. 1341-55.
371. Dushek, O., P.A. van der Merwe, and V. Shahrezaei, *Ultrasensitivity in multisite phosphorylation of membrane-anchored proteins.* Biophys J, 2011. **100**(5): p. 1189-97.
372. Wu, C., et al., *SAP controls T cell responses to virus and terminal differentiation of TH2 cells.* Nat Immunol, 2001. **2**(5): p. 410-4.
373. Dunne, P.D., et al., *DySCo: quantitating associations of membrane proteins using two-color single-molecule tracking.* Biophys J, 2009. **97**(4): p. L5-7.
374. James, J.R., et al., *Single-molecule level analysis of the subunit composition of the T cell receptor on live T cells.* Proc Natl Acad Sci U S A, 2007. **104**(45): p. 17662-7.
375. Okkenhaug, K. and R. Rottapel, *Grb2 forms an inducible protein complex with CD28 through a Src homology 3 domain-proline interaction.* J Biol Chem, 1998. **273**(33): p. 21194-202.
376. Balagopalan, L., et al., *The LAT story: a tale of cooperativity, coordination, and choreography.* Cold Spring Harb Perspect Biol, 2010. **2**(8): p. a005512.
377. Filippakopoulos, P., S. Muller, and S. Knapp, *SH2 domains: modulators of nonreceptor tyrosine kinase activity.* Curr Opin Struct Biol, 2009. **19**(6): p. 643-9.
378. Li, W., et al., *Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation.* J Biol Chem, 2008. **283**(22): p. 15491-501.
379. Mayer, B.J., H. Hirai, and R. Sakai, *Evidence that SH2 domains promote processive phosphorylation by protein-tyrosine kinases.* Curr Biol, 1995. **5**(3): p. 296-305.
380. Kuriyan, J. and D. Eisenberg, *The origin of protein interactions and allostery in colocalization.* Nature, 2007. **450**(7172): p. 983-90.
381. Lewitzky, M., P.C. Simister, and S.M. Feller, *Beyond 'furballs' and 'dumpling soups' - towards a molecular architecture of signaling complexes and networks.* FEBS Lett, 2012. **586**(17): p. 2740-50.