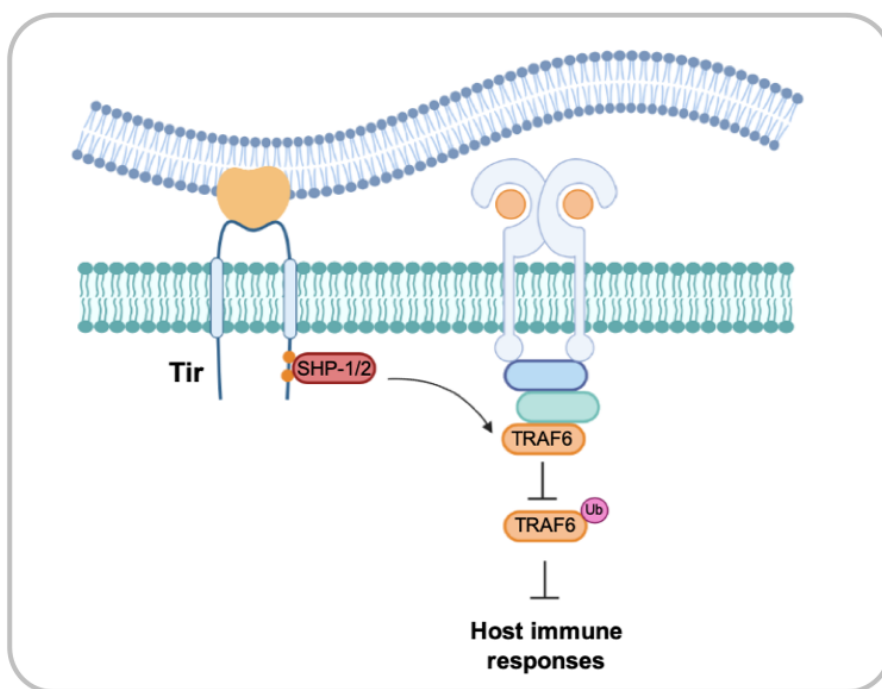


Unraveling structural dynamics of Tir-Host interactions blocking immune response

Marta Filipa Mota Vieira



Dissertation presented to obtain the **Ph.D. degree in Molecular Biosciences**
MolBioS PhD Programme

Oeiras, July, 2022

Unraveling structural dynamics of Tir-Host interactions blocking immune response

Marta Filipa Mota Vieira

Dissertation presented to obtain the Ph.D. degree in Molecular
Biosciences

Instituto de Tecnologia Química e Biológica António Xavier | Universidade
Nova de Lisboa

Supervisor: Dr. Tiago Neto Cordeiro

Oeiras, July 2022



First Edition, July 2022

Dynamic Structural Biology Laboratory
Instituto de Tecnologia Química e Biológica António Xavier
Universidade Nova de Lisboa
Av. da República, Oeiras

*“The important thing is to never stop
questioning.”*

Albert Einstein

Acknowledgments

The work achieved in this dissertation was only possible due to the contribution of many people to whom I would like to acknowledge and express my gratitude.

To Fundação para a Ciência e Tecnologia (FCT), for the financial support of my Ph.D., and to ITQB NOVA for providing an excellent environment to carry out my scientific work.

To my supervisor, Dr. Tiago Neto Cordeiro, I express my gratitude for the guidance and the opportunity to open horizons of knowledge that made me grow professionally and personally.

To the thesis committee members, Dr. Mariana Pinho and Dr. Ricardo Louro, for supervising my work, encouraging it, and for all the helpful discussions. A special acknowledgment to Dr. Ricardo Louro for his constant availability, for supporting my work, and for encouraging my growth as a scientist over these years.

To all the members of the Dynamic Structural Biology group, that contributed to a good and collaborative working environment. A special thanks to my “administrative cost” Tiago Veloso and Guillem Hernandez for your patience, availability, and help, without which this work would have been more difficult, as well as for the good talks, support, and funny moments that helped me overcome some more difficult times.

To my “neighbors” of the Inorganic Biochemistry and NMR group, for the availability and help always demonstrated, and for the countless times in which they let me use their equipment and reagents.

To all colleagues and ITQB technicians who contributed to my learning over these years, for their friendship and help in the most varied situations. A special acknowledgment to Dr. Cristina Timóteo for all the help given in the purification of proteins, but mostly, for the constant support and concern, and for the sweet treats that brightened my days.

To the group of friends formed in these last years: Anaísa Coelho, Diogo Silva, Rodrigo Nobre, Jéssica Soares and Carolina Rodrigues. Thank you for the company, for the scientific discussions, and for the sharing of knowledge that contributed to the development of this work. Thank you for the conversations, good moments and laughs, support, and most importantly for your friendship.

A very special acknowledgment to Anaísa, one of the first people I met at ITQB and without a doubt the best person this journey has given me. Thank you for your friendship and unconditional support. For sharing with me the “roller coaster” of these last years. For your contagious optimism and for making me want to be a better person.

To my parents, Gabriela and Luís, I express immeasurable gratitude for your infinite dedication, for the unconditional support you always gave me, and for your love, patience, and motivation. For making me the person I am today.

To my brother Tiago and Carolina for their love and care, unconditional support, and for believing in me.

To André for all the love, patience, and dedication, for being by my side on this journey from day one. For always encouraging me and helping to keep my sanity. I cannot express how fortunate I am to go hand in hand with you.

To my faithful and unconditional four-legged companions, Tita and Lucas, for showing me that even on the days that go less well, there is always a reason to smile.

This thesis is dedicated to the little stars that always light my way, my grandparents Maria do Bom Sucesso, Elias, Brás, and Eulália. I hope you are proud.

Thesis Publications

Vieira, M. F. M., Hernandez, G., Zhong, Q., Veloso, T., Monteiro, H., Arbesú, M., Frankel, G., Zanzoni, A., Cordeiro T.N. The pathogen-encoded signaling receptor Tir exploits host-like intrinsic disorder for infection. eLife (RC refereed preprint @ bioRxiv 2021.04.22.440577) (*under revision*).

Vieira, M. F. M., Veloso, T., Cordeiro, T. N. Into the molecular basis of host TLR signaling suppression by the bacterial effector Tir. (*in preparation*).

Abbreviations, symbols and constants

A/E	Attaching and effacing
ATP	Adenosine triphosphate
BMRB	Biological Magnetic Resonance Data Bank
CD	Circular dichroism
CR	<i>Citrobacter rodentium</i>
D₂O	Deuterium oxide
DHPC	1,2-diheptanoyl-sn-glycero-3-phosphocholine
D_{max}	Maximum dimension
DMEM	Dulbecco's Modified Eagle Medium
DMPC	1,2-Dimyristoyl-sn-glycero-3-phosphocholine
DMPG	1,2-Dimyristoyl-sn-glycero-3-phosphorylglycerol sodium salt
DNA	Deoxyribonucleic acid
DSS	Sodium trimethylsilylpropanesulfonate
DTT	1,4-dithiothreitol
EDTA	Ethylenediamine tetra acetic acid
EHEC	Enterohemorrhagic <i>Escherichia coli</i>
ELM	Eukaryotic linear motif
EPEC	Enteropathogenic <i>Escherichia coli</i>
ESRF	European Synchrotron Radiation Facility
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRV-3C	Recombinant type 14 3C protease from human rhinovirus
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
IRTKS	Insulin receptor tyrosine kinase substrate
ITIM	Immunoreceptor tyrosine-based inhibition motif
K_D	Equilibrium dissociation constant

kDa	Kilo Dalton
LB	Luria-Bertani
LEE	Locus of enterocyte effacement
LIC	Ligase Independent Cloning
LPS	Lipopolysaccharide
MSTL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OD₆₀₀	Optical density at 600 nm
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
Ppm	Parts per million
PRE	Paramagnetic relaxation enhancement
<i>PreSMos</i>	Pre-Structured <i>Motifs</i>
Psi	Pound force per square inch
Rg	Radius of gyration
Rpm	Rotations per minute
RT-NMR	Real time-Nuclear Magnetic Resonance
SASBDB	Small Angle Scattering Biological Data Bank
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH2	Src homology 2
SHP-1	Src homology region 2 domain-containing phosphatase-1
SHP-2	Src homology region 2 domain-containing phosphatase-2
T3SS	Type III secretion system
Tir	Translocated intimin receptor
TLR	Toll-like receptor
TOCSY	Total Correlation Spectroscopy
TRAF6	TNF Receptor Associated Factor 6

Tris	Tris – tris(hydroxymethyl)aminomethane
Trx	Thioredoxin
δ	Chemical shift
$\Delta\delta$	Chemical shift deviation

List of Aminoacids

Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartate	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamate	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

Abstract

Innate immunity is the first line of defense against pathogenic microorganisms. However, pathogens have evolved sophisticated mechanisms to evade immune responses during infection, often by delivering effectors/virulence factors into host cells.

The translocated intimin receptor (Tir) is a central effector of Attaching and Effacing (A/E) pathogens responsible for worldwide foodborne disease cases. Upon delivery into host cells, Tir migrates to the plasma membrane to act as a cell-surface receptor for intimin, anchoring A/E pathogens to targeted host cells and suppressing host autophagy and immune response processes to assist infection. Despite the considerable evidence that Tir is an essential effector for hijacking innate inflammatory responses, structural and quantitative information about protein interactions that underlie the Tir-mediated rewiring of host signaling remains largely unexplored. Thus, the work presented in this thesis sought to understand and elucidate the basis for Tir binding diversity in signaling by providing a molecular and atomic description of Tir and by studying the mechanisms of how this bacterial "weapon" recruits host targets to disarm defense mechanisms and ensure infection.

From disorder prediction and short linear motif analysis, we found that this bacterial-encoded transmembrane protein comprises highly disordered intracellular domains bearing host-like motifs that bind host proteins, a trait that we unexpectedly found to be prevalent in several other effectors secreted by A/E bacteria.

We assessed Tir's intrinsic disorder by an integrative structural biophysics approach, using a combination of NMR, SAXS, and CD spectroscopies. Our results unveiled that Tir comprises a disordered intracellular C-terminal domain (C-Tir) with features of an intrinsically disordered region (IDR). NMR was further employed to dissect the subtle structural features of C-Tir at the atomic level. To do so, we assigned the backbone resonances of C-Tir and analyzed its secondary structural

propensities using secondary chemical shifts. Paramagnetic relaxation enhancement (PRE) experiments were also carried out for several C-Tir variants (V405C, S428C, S463C, S509C, and S536C) to map the presence of long-range contacts. PRE-data was further integrated with SAXS to determine a conformational ensemble of C-Tir. Our results revealed that C-Tir has non-random structural preferences at phosphorylation sites, including host-like immunoreceptor tyrosine-based inhibitory motifs (ITIMs). Using NMR, we also verified that C-Tir's host-like ITIM motifs are capable of engaging versatile interactions with host SH2 domain-containing phosphatases SHP-1 and SHP-2 pre-phosphorylation, even though the intracellular signaling associated with these motifs relies on the phosphorylation of the central tyrosine by Src family protein tyrosine kinases.

To fully assess C-Tir's ability to recruit SH2 domains, C-Tir's Tyr-phosphorylated state (pC-Tir) was reconstructed by reassigning the backbone NMR resonances. We also monitored the kinetics and the extension of the phosphorylation reaction of the four tyrosine residues in the ITIM-like regions by NMR. In doing so, we found that the four positions are simultaneously phosphorylated yet with different reaction rates, with Y454 and Y483 being the fastest and slowest, respectively. Furthermore, we show that multi-site tyrosine phosphorylation enables C-Tir to engage SH2 domains of host phosphatases SHP-1 and SHP-2 in a dynamic and multivalent manner via its ITIM-like sequences, reinforcing the role of Tir as scaffold/hub interacting with host proteins. To understand the contribution of individual SH2 domains in Tir-mediated recruitment of SHP-2, we resorted to multiple truncated versions of SHP-2 and studied their interaction with pC-Tir using NMR titrations. By estimating the local binding affinities of the four binding sites in pC-Tir for all the different SH2 domains of SHP-2 we found that C-Tir's ITIM-like sequences comprising Y483 and Y511 are the preferential ones for SH2-domain binding, likely in an antiparallel fashion in which N-SH2 and C-SH2 bind to Y511 and Y483, respectively. However, interaction via the Y454- and Y474-containing motifs also occurs.

By NMR, we also found that C-Tir's host-like ITIM motifs bind lipid bilayers, and that membrane affinity is residue-specific and modulated by lipid composition, highlighting a previously unrecognized interaction with the host. To investigate whether the lipid-binding residues affected the known biological functions of Tir during the infection we also mutated key residues in C-Tir (R388, R389, L513 and A541), expressed plasmid-encoded wild-type and mutant Tir, and used these strains to infect the model intestinal epithelial cell line SNU-C5. Using immunofluorescence and propidium iodide uptake assays we demonstrate that C-Tir lipid-binding affects Tir-induced host cell death and actin polymerization. In addition to lipid bilayers, we uncovered a novel interaction of Tir with host intracellular targets. Using NMR, we demonstrate that C-Tir can directly recruit and bind to the C-terminal domain of TRAF6, likely acting as a negative regulator of TLR signaling. In C-Tir, the binding region comprises a conserved motif including Y474. Our results revealed that binding to TRAF6 is restricted by phosphorylation, and that Y474 appears to be absolutely required for the interaction with TRAF6. Direct interaction between SHP-1/2 and TRAF6 prevents the ubiquitination of TRAF6 and inhibits the expression of proinflammatory cytokines. Our analysis suggests that C-Tir might be a molecular scaffold bringing SHP-1/2 and TRAF6 to suppress immune responses.

In conclusion, our findings collectively provide an updated picture of Tir's intracellular side, highlighting its ability to exploit host-like protein disorder versatility as a molecular strategy for host evasion.

Resumo

A imunidade inata é a primeira linha de defesa contra microrganismos patogênicos. No entanto, os agentes patogênicos evoluíram no sentido de desenvolver mecanismos sofisticados para evitar respostas imunes durante a infecção, muitas vezes conseguidos através da entrega de efetores/fatores de virulência nas células hospedeiras.

O recetor de intimina translocado (Tir, do inglês *translocated intimin receptor*) é um efector central de agentes patogênicos de fixação e eliminação (A/E, do inglês *Attaching and Effacing*) responsáveis por infecções transmitidas através de fontes alimentares em todo o mundo. Uma vez dentro das células hospedeiras, Tir migra para a membrana plasmática atuando como recetor da proteína bacteriana intimina, de forma a fixar agentes patogênicos A/E às células hospedeiras alvo e suprimir processos de resposta imune e autofagia do hospedeiro para auxiliar o processo de infecção. Embora exista informação considerável que Tir é um efector essencial para a subversão de respostas inflamatórias inatas, continua a haver uma lacuna relativa a informação estrutural e quantitativa sobre as interações mediadas por este efector na subversão de vias de sinalização do hospedeiro. Posto isto, na presente dissertação pretendeu-se compreender e elucidar os princípios da diversidade de interações de Tir em processos de sinalização ao fornecer uma descrição atômica e molecular deste efector e ao estudar os mecanismos pelos quais esta “arma” bacteriana consegue recrutar alvos do hospedeiro para desarmar mecanismos de defesa e garantir infecção.

Com base em previsões de desordem e análise de pequenos motivos lineares, observou-se que esta proteína transmembranar, codificada por bactérias, possui domínios intracelulares altamente desordenados com motivos semelhantes aos do hospedeiro que interagem com proteínas do mesmo. Inesperadamente, esta característica parece ser prevalente em vários outros efetores secretados por bactérias A/E.

A desordem intrínseca de Tir foi avaliada utilizando uma abordagem biofísica estrutural integrativa, através da combinação de espectroscopias de

Ressonância Magnética Nuclear (RMN), SAXS (do inglês *Small-angle X-ray scattering*) e dicroísmo circular (CD, do inglês *circular dichroism*). Os resultados demonstraram que Tir contém um domínio intracelular C-terminal (C-Tir) com características de uma região intrinsecamente desordenada. A espectroscopia de RMN foi ainda utilizada para dissecar as características estruturais subtis de C-Tir a nível atómico. Para tal, atribuíram-se as ressonâncias da cadeia principal de C-Tir e analisou-se a propensão de estrutura secundária através de desvios químicos secundários. Foram também realizadas experiências de aumento de relaxação paramagnética (PRE, do inglês *Paramagnetic Relaxation Enhancement*) recorrendo a vários mutantes de C-Tir (V405C, S428C, S463C, S509C e S536C) para mapear a presença de contactos de longo alcance. Os dados de PRE foram ainda integrados com dados de SAXS para determinar um conjunto de conformações de C-Tir. Os resultados obtidos demonstraram que C-Tir possui preferências estruturais não aleatórias em locais de fosforilação, incluindo recetores que contêm motivos de inibição de imunorecetores baseados em tirosinas (ITIMs, do inglês *immunoreceptors tyrosine-based inhibitory motifs*) semelhantes aos do hospedeiro. Utilizando RMN, também se verificou que os ITIMs de C-Tir conseguem participar em interações versáteis com os domínios SH2 das fosfatases do hospedeiro SHP-1 e SHP-2 previamente à sua fosforilação, embora a sinalização intracelular associada a estes motivos dependa da fosforilação de um resíduo de tirosina por proteínas cinases da família Src.

De forma a avaliar a capacidade de C-Tir em recrutar domínios SH2, reconstruiu-se o seu estado fosforilado (pC-Tir) através da reatribuição das ressonâncias da cadeia principal. A cinética e extensão da reação de fosforilação nos quatro resíduos de tirosina dos motivos ITIM de C-Tir foram também monitorizadas através de RMN. Desta forma, observou-se que as quatro posições são fosforiladas simultaneamente, mas com diferentes velocidades de reação, sendo os resíduos Y454 e Y483 o mais rápido e mais lento, respetivamente. Além disso, demonstrou-se que a fosforilação de múltiplas tirosinas permite que C-Tir interaja com os domínios SH2 de SHP-

1 e SHP-2 de maneira dinâmica e multivalente através das suas sequências ITIM, reforçando o papel de Tir como *scaffold/hub* ao interagir com as proteínas do hospedeiro. Para compreender a contribuição individual dos domínios SH2 no recrutamento da SHP-2 mediado por Tir, recorreu-se a várias versões truncadas de SHP-2 e estudou-se a interação das mesmas com pC-Tir através de titulações de RMN. Ao estimar a afinidade dos quatro locais de ligação em pC-Tir para todos os domínios SH2 de SHP-2, observou-se uma ligação preferencial das sequências ITIM que contêm os resíduos Y483 e Y511, sugerindo uma interação antiparalela em que os domínios N-SH2 e C-SH2 interagem com as tirosinas Y511 e Y483, respetivamente. No entanto, também foi observada interação através dos motivos que contêm os resíduos Y454 e Y474.

Através do uso de RMN, também se observou que os motivos ITIM de C-Tir interagem com membranas, e que a afinidade para as mesmas é específica para determinados resíduos e mediada pela composição de lípidos, destacando uma interação de Tir com o hospedeiro previamente desconhecida. De forma a investigar se os resíduos que interagem com membranas afetavam as funções biológicas de Tir durante a infeção, mutaram-se vários resíduos em C-Tir (R388, R389, L513 e A541). A estirpe nativa (do inglês *wild type*) e as estirpes mutantes foram expressas e utilizadas para infetar células SNU-C5, uma linha modelo de células epiteliais do intestino. Através de imunofluorescência e ensaios de iodeto de propídio (PI, do inglês *propidium iodide*) demonstrou-se que a ligação de C-Tir a lípidos afeta os processos de morte celular do hospedeiro e polimerização de actina mediados por Tir.

Adicionalmente, descobriu-se uma nova interação de Tir com alvos intracelulares do hospedeiro. Com a utilização de RMN, demonstrou-se que C-Tir pode recrutar e interagir diretamente com o domínio C-terminal de TRAF6, atuando como um regulador negativo da sinalização de recetores Toll-Like (TLR, do inglês *Toll Like Receptors*). Em C-Tir, a região de interação compreende um motivo conservado que inclui o resíduo Y474. Os nossos resultados demonstraram que a interação com TRAF6 é restringida por

fosforilação e sugerem que Y474 é necessária para a interação. A interação direta entre SHP-1/2 e TRAF6 previne a ubiquitinação de TRAF6 e inibe a expressão de citocinas pró-inflamatórias. A nossa análise sugere que C-Tir pode atuar como *scaffold* molecular ao fortalecer o contacto com SHP-1/2 e TRAF6 para suprimir respostas imunes.

Em suma, os resultados obtidos fornecem uma imagem atualizada da região intracelular de Tir, destacando a sua capacidade em explorar a versatilidade da desordem presente em proteínas do hospedeiro como uma estratégia molecular de forma a assegurar a evasão do mesmo.

Table of Contents

Acknowledgements	vii
Thesis Publications	xi
Abbreviations, symbols and constants	xiii
Abstract	xvii
Resumo	xxi
Table of contents	xxv
Index of Figures	xxix
Index of Tables	xxxv
Preface	xxxvii
Chapter 1 – General Introduction	1
Host and pathogen interplay	3
Host defense mechanisms.....	3
Pathogen mechanisms to escape host recognition.....	5
Enteropathogenic <i>E. coli</i> (EPEC).....	5
Enterohemorrhagic <i>E. coli</i> (EHEC).....	7
Attaching and effacing bacterial pathogens.....	8
The prominent translocated intimin receptor.....	11
Cellular processes mediated by Tir.....	14
Actin pedestal formation.....	14
Membrane phosphoinositide signaling.....	17
Suppression of host immune responses.....	17
Regulation of Rac1 GTPase function.....	19
Down-regulation of Map-dependent filopodia formation.....	19
Induced pyroptosis.....	19
Intrinsically Disordered Proteins	20
Sequence-encoded signature of disorder.....	21
Roles of disorder in IDPs functions.....	22
IDPs in Bacteria.....	23
NMR Spectroscopy in characterization of IDPs	27
Principles of NMR spectroscopy.....	27

Challenges of IDPs studies by NMR spectroscopy.....	31
NMR applications to characterize IDPs.....	32
Resonance assignment.....	32
Mapping transient structural elements and long-range contacts....	33
Monitoring post-translational modifications.....	36
Monitoring protein-protein interactions with binding partners.....	41
Other methods to study IDPs.....	46
CD spectroscopy.....	46
Small angle X-ray scattering (SAXS).....	47
References.....	51
Objectives.....	65
Chapter 2 – Materials and Methods.....	67
Protein samples production.....	69
C-Terminal region of Tir.....	69
Construct design and cloning.....	69
Site-directed mutagenesis.....	71
Heterologous expression.....	73
Unlabeled protein expression.....	74
Uniformly labeled protein expression.....	74
Selectively unlabeled protein expression.....	74
Cell harvesting, disruption and extraction from inclusion bodies.....	75
Protein purification.....	76
Protein tyrosine phosphatase SHP-1.....	77
Constructs design and cloning.....	77
Heterologous expression.....	77
Protein purification.....	78
Protein tyrosine phosphatase SHP-2.....	78
Constructs.....	78
Site-directed mutagenesis.....	79
Heterologous expression.....	81
Protein purification.....	81
C-Terminal domain of TRAF6.....	82

Construct design and cloning.....	82
Heterologous expression.....	84
Protein purification.....	84
NMR spectroscopy	86
Backbone assignment.....	86
Secondary chemical shifts.....	89
Paramagnetic Relaxation Enhancement experiments.....	89
2,2,2-Trifluoroethanol assays.....	90
Lipid binding assays.....	91
<i>In vitro</i> phosphorylation.....	91
Quenched phosphorylation reactions.....	92
Protein-protein interactions.....	94
Other techniques	96
Disorder prediction and short linear motif analysis.....	96
Tir sequences alignment.....	98
CD spectroscopy.....	99
SAXS.....	99
Structural ensemble generation.....	100
Tir-TRAF6 complex modeling with AlphaFold2.....	100
Lipid-binding site mutations.....	101
EPEC infection.....	103
Immunofluorescence imaging.....	103
Propidium iodide uptake assays.....	103
References.....	104
Chapter 3 – Results	109
The pathogen-encoded signaling receptor Tir exploits host-like intrinsic disorder for infection	111
Structural disorder and short linear motifs are common features of A/E pathogen effectors.....	111
Tir bears an intrinsically disordered C-terminal tail.....	114
C-Tir displays non-random structural preferences at phosphorylation sites.....	118

C-Tir can interact with the host SHP-1 prior to phosphorylation.....	123
Multi-phosphorylated C-Tir is a disordered multivalent cytosolic tail.....	127
Disordered C-Tir binds lipids.....	133
Tir-induced actin polymerization and cell death depend on C-Tir lipid binding.....	136
Into the molecular basis of host TLR signaling suppression by the bacterial effector Tir.....	140
Unphosphorylated C-Tir exploits differential SHP-1/SHP-2 binding.....	140
Phosphorylated C-Tir interacts with SHP-2 forming fuzzy complexes...	142
Phosphorylated C-Tir prefers to recruit SHP-2 in an anti-parallel fashion.....	146
C-Tir can recruit TRAF6 without binding to host phosphatases.....	150
The C-Tir tail harbors an overlapping phospho-regulated TRAF6 binding motif.....	157
Tyrosine 474 appears to be crucial for C-Tir interaction with TRAF6-C.....	159
References.....	162
Chapter 4 – Discussion.....	167
References.....	177
Concluding remarks and future perspectives.....	181
Supplementary Information.....	185
Supplementary Figures.....	187
Supplementary Tables.....	208
References.....	212

Index of Figures

Chapter 1 – General Introduction

Figure 1.1. Schematic representation of TLR signaling pathway.....	4
Figure 1.2. Genetic structure of the LEE pathogenicity island of <i>Escherichia coli</i> O127:H6 strain E2348/69 (EPEC).....	8
Figure 1.3. EPEC and EHEC molecular pathogenicity mechanisms.....	10
Figure 1.4. Sequence alignment of Tir from <i>E. coli</i> O127:H6 strain E2348/69 (EPEC) and <i>E. coli</i> O157:H7 (EHEC).....	12
Figure 1.5. Tir A/E effector.....	13
Figure 1.6. EPEC mechanisms of Tir-mediated signaling activation to form actin pedestals.....	15
Figure 1.7. EHEC mechanisms of Tir-mediated signaling activation to form actin pedestals.....	16
Figure 1.8. Tir-mediated suppression of host immune responses.....	18
Figure 1.9. Schematic representation of intrinsic and emergent sticker multivalence.....	24
Figure 1.10. Schematic representation of magnetic dipole orientation and precession for a nuclear spin $\frac{1}{2}$ in the presence of an external magnetic field B_0	28
Figure 1.11. Schematic representation of magnetization transfer.....	29
Figure 1.12. Paramagnetic probes for detection of long-range interactions.....	35
Figure 1.13. Phosphorylation of Serine, Threonine and Tyrosine residues.....	37
Figure 1.14. Schematic overview of NMR procedures for monitoring real-time protein phosphorylation reactions <i>in vitro</i>	39
Figure 1.15. Schematic overview for monitoring discontinuous NMR measurements of quenched phosphorylation reactions.....	40
Figure 1.16. Schematic representation for the titration of a labelled protein A with increasing concentration of unlabeled protein B in slow, fast and intermediate exchange regimes.....	44

Figure 1.17. Experimental setup of a SAXS experiment.....	48
Figure 1.18. SAXS profiles for globular, partially unfolded and fully disordered proteins.....	49

Chapter 3 – Results

Figure 3.1. Distribution of disorder fraction in A/E pathogens.....	111
Figure 3.2. Accumulated fractions of the structural categories in terms of structural disorder.....	112
Figure 3.3. Fraction of ELMs vs. disorder fraction in A/E effectors.....	113
Figure 3.4. C-Tir displays an aberrant migration profile during SDS-PAGE.....	115
Figure 3.5. Kratky representation and pair distance distribution function derived from the SAXS experimental profiles measured for C-Tir.....	116
Figure 3.6. The Tir C-terminus is an intrinsically disordered domain.....	117
Figure 3.7. C-Tir has partially structured elements.....	119
Figure 3.8. C-Tir disordered chain makes long range contacts.....	120
Figure 3.9. C-Tir is disordered with residual structure at sites of host interaction.....	122
Figure 3.10. C-Tir can bind host C-SH2 domain of SHP-1 pre-phosphorylation.....	123
Figure 3.11. Interaction of C-Tir with the C-SH2 domain of SHP-1.....	124
Figure 3.12. Impact of TFE on C-Tir.....	125
Figure 3.13. (A) NMR [¹ H- ¹⁵ N]-HSQC spectra of C-Tir in the presence of increasing amounts of TFE. (B) Residue-resolution mapping of the effect of TFE on C-Tir.....	126
Figure 3.14. Tyrosine phosphorylation of C-Tir.....	127
Figure 3.15. Secondary structure propensity of C-Tir and pC-Tir.....	128
Figure 3.16. C-Tir NMR measurements of quenched phosphorylation reactions.....	129
Figure 3.17. C-Tir contains multiple phosphorylated docking sites.....	131

Figure 3.18. NMR quantification of the binding of C-SH2 to individual pY-sites.....	132
Figure 3.19. Lipid-binding by disordered C-Tir.....	134
Figure 3.20. Effect of Tir C-terminus mutations on Tir-induced actin polymerization.....	137
Figure 3.21. Effect of Tir C-terminus mutations on Tir-induced cell death.....	138
Figure 3.22. C-Tir can bind host SH2 domains of SHP-1 and SHP-2 prior to phosphorylation.....	141
Figure 3.23. Overview of NMR titrations to study the interaction of pC-Tir with SH2 domains of SHP-2.....	143
Figure 3.24. Interaction of pC-Tir with the C-SH2 domain of SHP-2.....	144
Figure 3.25. Interaction of pC-Tir with the N-SH2 domain of SHP-2.....	145
Figure 3.26. Interaction of pC-Tir with the NC-SH2 domain of SHP-2....	146
Figure 3.27. Interaction of pC-Tir with the NC-SH2 C-dead domain of SHP-2.....	147
Figure 3.28. Interaction of pC-Tir with the NC-SH2 N-dead domain of SHP-2.....	148
Figure 3.29. Cartoon representation depicting the relative binding affinities of SHP-2 individual SH2 to C-Tir's ITIM-like sequences.....	149
Figure 3.30. C-Tir possess a putative site for TRAF6-C domain binding.....	151
Figure 3.31. Overlay of $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of ^{15}N -C-Tir in the presence of TRAF6-C domain.....	152
Figure 3.32. C-Tir can directly recruit TRAF6-C.....	153
Figure 3.33. Predicted complexes of C-Tir peptides and TRAF6-C binding interface obtained with AlphaFold2.....	156
Figure 3.34. Binding to TRAF6-C is hindered upon C-Tir phosphorylation by host kinases.....	158
Figure 3.35. Interaction of pC-Tir_Y454F+Y474F with NC-SH2 domain of SHP-2.....	160
Figure 3.36. C-Tir Y474 is required for interaction with TRAF6-C.....	161

Chapter 4 – Discussion

Figure 4.1. Proposed model of the mechanisms employed by Tir to hijack intracellular signaling.	176
---	-----

Supplementary Information

Figure SI.1. Schematic representation of the constructs used in this work.....	187
Figure SI.2. Overview of C-Tir purification.....	188
Figure SI.3. Overview of SHP-1 C-SH2 domain purification.....	189
Figure SI.4. Overview of SHP-2 C-SH2 domain purification.....	190
Figure SI.5. Overview of SHP-2 N-SH2 domain purification.....	191
Figure SI.6. Overview of SHP-2 NC-SH2 domain purification.....	192
Figure SI.7. Overview of TRAF6-C domain purification.....	193
Figure SI.8. Predicted structural disorder in A/E pathogens with IUPred.....	194
Figure SI.9. A/E disordered effectors.....	195
Figure SI.10. Multiple sequence alignment of Tir sequences.....	196
Figure SI.11. SEC-SAXS chromatogram of C-Tir.....	197
Figure SI.12. Comparative SAXS: C-Tir versus globular Myoglobin (Myb).....	197
Figure SI.13. Reverse labelling of C-Tir.....	198
Figure SI.14. Time course PI uptake of EPEC-infected cells.....	199
Figure SI.15. Global fittings obtained for the interaction of unphosphorylated C-Tir with C-SH2 domain of SHP-2.....	199
Figure SI.16. Global fittings obtained for the interaction of unphosphorylated C-Tir with NC-SH2 domain of SHP-2.....	200
Figure SI.17. Global fittings obtained for the interaction of pC-Tir with C-SH2 domain of SHP-2.....	200
Figure SI.18. Global fittings obtained for the interaction of pC-Tir with N-SH2 domain of SHP-2.....	201
Figure SI.19. Global fittings obtained for the interaction of pC-Tir with NC-SH2 domain of SHP-2.....	202

Figure SI.20. Global fittings obtained for the interaction of pC-Tir with NC-SH2 C-dead domain of SHP-2.....	203
Figure SI.21. Global fittings obtained for the interaction of pC-Tir with NC-SH2 N-dead domain of SHP-2.....	204
Figure SI.22. Multiple sequence alignment of SHP-1 and SHP-2 sequences.....	205
Figure SI.23. pLDDT score as a function of C-Tir residue number obtained for the predictions of the binding interface between TRAF6-C and C-Tir using AlphaFold2.....	206
Figure SI.24. Cartoon representation of the TRAF6-CD40 complex.....	206
Figure SI.25. Overlay of [¹ H- ¹⁵ N]-HSQC spectra of pC-Tir in the presence of TRAF6-C domain.....	207
Figure SI.26. Global fittings obtained for the interaction of pC-Tir _{Y454F+Y474F} with NC-SH2 domain of SHP-2.....	207

Index of Tables

Chapter 2 – Materials and Methods

Table 2.1. Oligonucleotide sequences (primers) used for amplification of C-Tir.....	70
Table 2.2. Oligonucleotide sequences (primers) used for site-directed mutagenesis of C-Tir.....	72
Table 2.3. Oligonucleotide sequences (primers) used for amplification of SHP-1 C-SH2 domain.....	77
Table 2.4. Oligonucleotide sequences (primers) used for site-directed mutagenesis of SHP-2 SH2 domains.....	80
Table 2.5. Oligonucleotide sequences (primers) used for amplification of TRAF6-C and pCoofy16 linearization.....	83
Table 2.6. Acquisition details of NMR experiments for C-Tir and pC-Tir backbone assignment.....	88
Table 2.7. Overview of the interaction studies carried out between C-Tir and SHP-1, SHP-2 or TRAF6 human proteins.....	94
Table 2.8. Reference proteomes by specie name and taxon.....	97
Table 2.9. Structural categories by disorder content.....	98
Table 2.10. Oligonucleotide sequences (primers) used for site-directed mutagenesis of full length Tir.....	101
Table 2.11. Plasmids used for Tir lipid binding site mutations.....	102

Supplementary Information

Table SI.1. Disorder fractions (DISOPRED), predicted motif densities and structural categories for CR, EHEC and EPEC effectors.....	208
Table SI.2. C-Tir SAXS data collection and analysis information.....	211
Table SI.3. Input sequences employed in the prediction of the 3D structure of the complex formed between Tir and TRAF6 using AlphaFold2.....	212

Preface

This thesis was developed at the Dynamic Structural Biology group from the Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, under the supervision of Dr. Tiago Neto Cordeiro.

The document is organized as follows:

- A general introduction organized in three main sections. The first section describes the biological system under study, mainly focusing on the translocated intimin receptor (Tir). The second chapter is focused on intrinsically disordered proteins. The third chapter depicts the main techniques employed in this work, particularly NMR spectroscopy.
- A small section where the objectives of this dissertation are defined.
- A Materials and Methods section describing the experimental procedures and protocols used in this work, with detailed information on experimental aspects and data treatment.
- A section with experimental results obtained during this work, divided into two main sections containing an exposition and brief discussion of the results.
- A general discussion of all the work performed in this thesis, followed by a small section containing the main conclusions and some future perspectives.
- A supplementary information section with supporting figures and tables.

Chapter 1

General Introduction

This chapter describes the main concepts relevant under the scope of this work, and it is divided into three main sections. The first section is dedicated to mechanisms at the host-pathogen interface, particularly involving the pathogen-encoded translocated intimin receptor (Tir), the main subject of this thesis. A second section presents the topic of intrinsically disordered proteins (IDPs), relevant for the study of Tir which includes disordered regions. Finally, the third section focuses on biophysical methods to characterize IDPs, particularly Nuclear Magnetic Resonance (NMR) spectroscopy. This was the primary technique employed to scrutinize the disordered C-terminal domain of Tir.

Host and pathogen interplay

In an environment constantly threatened by potentially harmful microorganisms, host survival relies on effectively mounted defense responses. This way, hosts have evolved immune response systems to eliminate pathogens. However, in this arms race of host-microbe co-evolution, many microbial pathogens have successfully evolved clever mechanisms to evade host immune responses and undergo highly dynamic interactions with the host in all stages of pathogenic infection [1].

Host defense mechanisms

Host immune systems are composed of innate and acquired immunity mechanisms. The innate immune system is the first line of defense against pathogenic microorganisms. Innate immunity is mediated by several defense cells, including macrophages, neutrophils, and dendritic cells, that detect invading pathogens through a number of germline-encoded receptors designated as pattern recognition receptors (PPRs). PPRs recognize conserved molecular components essential for microorganism survival, known as pathogen-associated molecular patterns (PAMPs). Different PPRs recognize specific PAMPs, leading to the activation of specific intracellular

signaling pathways and consequently to the incitement of distinct responses [1,2]. Gram-negative bacteria [(e.g. *Escherichia coli* (*E. coli*)] recognize toll-like receptors (TLRs) using lipopolysaccharide (LPS) molecules as PAMPs. Upon ligand binding, TLRs dimerize and undergo conformational changes which allow the recruitment and interaction with several cytoplasmic adaptor proteins, including the myeloid differentiation factor 88 (MyD88) and the TIR-domain-containing adaptor protein-inducing IFN- β (TRIF) [2].

The activation of differential responses relies on the selective use of these adaptor molecules, as MyD88 and TRIF recruitment leads to the activation of distinct signaling pathways [3,4]. MyD88 recruitment induces the expression of several cytokine-encoding genes involved in inflammatory responses, and activation of TRIF-dependent pathway triggers the expression of several IFN-inducible genes [5] (Figure 1.1).

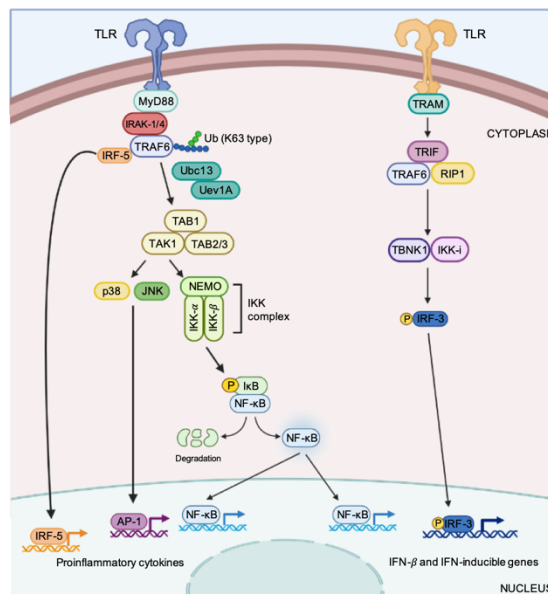


Figure 1.1. Schematic representation of TLR signaling pathway (Adapted from [1]).

In contrast, to avoid harmful and inappropriate inflammatory responses, TLR signaling is negatively controlled by host cellular immunoreceptor tyrosine-based inhibition motifs (ITIMs). Phosphorylation of ITIMs tyrosine residues enables the binding of the Src-homology 2 domain (SH2)-containing protein tyrosine phosphatases SHP-1 and SHP-2, which then decrease the phosphorylation of several host targets (e.g. Syk, γ -chain of the CD16 complex) or directly dephosphorylate NF- κ B and MAP kinases to inhibit activation of host immune responses [6,7].

Pathogen mechanisms to escape host recognition

At the same time that hosts have evolved defense mechanisms against invading microorganisms, pathogens have also evolved multiple strategies to evade the host immune system recognition.

Some bacterial pathogens have evolved modified forms of the regular TLR ligands, such as the LPSs from *Helicobacter pylori* and *Legionella pneumophila* [8,9]. Others evade host recognition by changing and adapting TLR signaling pathways for their benefit. For example, *Mycobacterium tuberculosis* avoids being killed by macrophages by inhibiting IFN- γ -mediated signaling mediated by its lipoprotein [10]. Another example, particularly relevant within the scope of this work, is employed by Enteropathogenic and Enterohaemorrhagic *E. coli* pathogenic strains. These pathogens can recruit and hack signaling components of immune response pathways to limit host-inflammatory response and ensure infection by secreting an effector protein, the translocated intimin receptor, that shares sequence similarities with host proteins [11].

Enteropathogenic *E. coli* (EPEC). EPEC is one of the most predominant pathogens infecting children in developing countries and one of the leading causes of persistent diarrhea, often associated with high mortality rates (12,337 total deaths in 2017, from which 9,459 were children under 5

years old) because of their high prevalence in both community and hospital setting [12,13]. It affects mainly infants under 11 months old, causing persistent moderate to severe diarrhea accompanied by mild fever and vomiting [12,14,15].

EPEC strains are classified as diarrhea-causing, non-Shiga toxin-producing *E. coli*. Additionally, EPEC strains are further classified into typical or atypical subtypes based on the presence or absence of an EPEC adherence factor (EAF) plasmid, which encodes the type IV bundle-forming pilus (BFP). Therefore, *E. coli* strains *eae*⁺ *bfpA*⁺ *stx*⁻ are classified as typical EPEC (tEPEC). Most of these strains belong to classic O:H serotypes and produce localized adherence (LA) phenotype associated with the production of BFP. On the other hand, *E. coli* strains that are *eae*⁺ *bfpA*⁻ *stx*⁻ are classified as atypical EPEC (aEPEC) [12].

Transmission to the host occurs via the fecal-oral route through contaminated surfaces, weaning fluids or human carriers, as it happens for other diarrheagenic *E. coli*. Colonization may result from an interplay of multiple factors, including host susceptibility (related to age, breastfeeding, nutritional and immunological status), bacterial factors (different virulence genes) and environmental factors (poor hygiene and high fecal contamination) [15,16].

The prevalence of EPEC infection varies between epidemiological studies based on differences in study populations, age distributions, methods used for detection and diagnosis, as well as geographic region and socioeconomic class [15]. According to studies performed by the Global Enteric Multicenter Study (GEMS), tEPEC was significantly associated with moderate to severe diarrhea in children under 2 years old of age in developing countries. In contrast, aEPEC was not related to this type of diarrhea at any of the GEMS sites (Gambia, Mali, Mozambique, Kenya, India, Bangladesh and Pakistan) [17]. However, among children with persistent diarrhea from countries in the developed world, aEPEC are among the most common pathogens isolated. In aEPEC cases, persistent diarrhea is the most frequent clinical manifestation accounting for more than half of the episodes,

suggesting that these strains may have an innate propensity to remain and persist longer in the intestine than other diarrheagenic *E. coli* [12,18]. Overall, aEPEC isolates are more commonly found among persistent cases of diarrhea than tEPEC isolates, yet, severe disease outcomes associated with tEPEC infections have been re-emerging in the last years [12,17].

Enterohemorrhagic *E. coli* (EHEC). EHEC is a zoonotic pathogen frequently associated with foodborne outbreaks in the developed world. EHEC is a subset of Shiga toxin producing *E. coli* (STEC) that is predominantly distinguished from EPEC strains by the presence of Shiga toxin (*stx*) genes. EHEC strains are further classified as sorbitol-fermenting SF-EHEC and nonfermenting or typical strains [15, 19-21]. Strains of EHEC are commensal in many ruminants, making cattle the primary reservoir hosts. Transmission of EHEC from animal reservoirs to humans occurs predominantly through food vehicles (e.g., fecal contamination of food) or water [22].

Infection can lead to bloody diarrhea, acute gastroenteritis, and haemorrhagic colitis. Furthermore, severe cases of EHEC infection can cause Hemolytic Uremic Syndrome (HUS) due to the translocation of Shiga toxins Stx1 and Stx2 across the gut, leading to kidney failure, chronic post-infection, or even death [23,24]. Infection is most prevalent in the United States (US) and Europe, with predominant serotypes varying with geographic localization. EHEC strain O157:H7 has caused numerous outbreaks in the US, accounting for more than 73,000 cases of disease [25]. In Europe, in 2011, EHEC strain O104:H4 was responsible for the deadliest foodborne outbreak ever recorded causing 46 deaths out of more than 3,900 cases of infection (~3.3% mortality rate) with striking high rates of hemorrhagic colitis (78%) and HUS (22%) [26].

A hallmark phenotype of both EPEC and EHEC strains is the ability to produce attaching and effacing (A/E) lesions in the small intestine, further described in the following section.

Attaching and effacing bacterial pathogens

Attaching and effacing (A/E) bacteria are a family of gastrointestinal pathogens able to subvert the intestinal epithelial cell function to induce a peculiar phenotype on host cells called the A/E lesion [27,28]. These lesions are characterized by the effacement of epithelial microvilli caused by intimate bacterial attachment to the host-cell membrane and the subsequent formation of protruding actin-rich cytoskeletal pedestal-like structures right beneath the adherent bacteria, to which they remain intimately attached [11,29,30].

The ability to induce A/E lesion is conferred by the locus of enterocyte effacement (LEE), a chromosomally located pathogenicity island (PAI). In EPEC and EHEC pathogens, the LEE is highly conserved (94% average identity at nucleotide level) [31,32]. The LEE is ~35 kb long and contains 41 genes (Figure 1.2), organized into seven distinct operons (LEE1 to LEE7) and four individual transcriptional units, which encode all the structural components of the type III secretion system (T3SS), seven effectors and their cognate chaperones and several proteins that participate in the intimate bacterial adherence or are involved in secretion regulation and LEE expression [33].

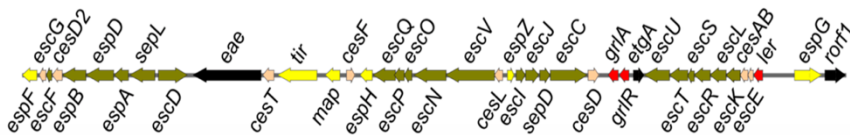


Figure 1.2. Genetic structure of the LEE pathogenicity island of *Escherichia coli* O127:H6 strain E2348/69 (EPEC). Genes are depicted as filled arrows colored according to their functional category: green for secretion apparatus components, red for transcriptional regulators, yellow for effectors, light orange for chaperones and black for other genes (Adapted from [34]).

The development of A/E lesion comprises three main stages: the initial adherence followed by signal transduction and lastly the bacterial intimate attachment [11]. Initial attachment to the surface of the host intestinal epithelium is mediated by type IV pilus (T4P) producing LA patterns. In EPEC, the T4P involved are the BFP which are expressed from 14 genes of a virulence-associated plasmid (EAF) [35]. BFP are not only important for the initial adherence to host cells but also to promote bacterial-bacterial interactions creating a network of filaments and leading to the formation of three-dimensional micro-colonies [22]. Regarding EHEC, the specific long polar fimbriae (LPF) is required for adherence and micro-colony formation, whereas a T4P named hemorrhagic coli pilus (HCP) promotes biofilm formation [36,37].

Upon initial adherence, bacteria employ a highly conserved protein-secretion apparatus called the type III secretion system (T3SS) to facilitate the translocation of several effector proteins (23 effectors in EPEC and up to 50 in some EHEC strains) into the eukaryotic cell cytoplasm which are then targeted to different subcellular compartments where they subvert actin dynamics and hamper other host signalling transduction pathways [30,38]. EPEC and EHEC T3SS starts by delivering Tir, a LEE-encoded effector [39]. The intimate attachment of the bacteria with the host cell is mediated by the interaction between intimin, an outer-membrane adhesin encoded on the *eae* gene, and the extracellular loop of Tir, which is inserted in the host cell plasma membrane in a hairpin topology [11,40]. Aside from its function as a receptor for intimin, Tir is also an essential mediator of protein signaling within epithelial cells. Despite evolving different mechanisms in EPEC and EHEC strains, Tir signaling culminates in actin accumulation and polymerization to produce pedestals in the interface between the bacteria and the host cell (Figure 1.3. A).

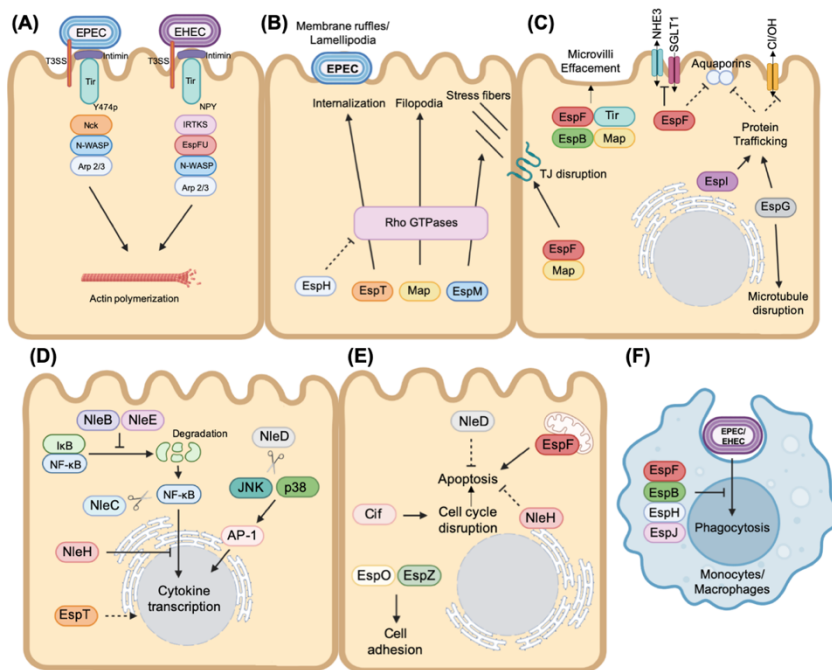


Figure 1.3. EPEC and EHEC molecular pathogenicity mechanisms. (A) Intimate adherence; **(B)** Actin remodeling; **(C)** Disruption of gut integrity; **(D)** Manipulation of immune responses; **(E)** Balance between apoptosis and survival; **(F)** Phagocytosis inhibition (Adapted from [11]).

Once the intimate bacterial attachment has occurred, several LEE-encoded effectors modulate Rho GTPase signaling to subvert host cell actin dynamics (Figure 1.3. B). Effectors of this family mimic Rho GTPase activity and function as guanine exchange factors (GEFs) by binding to their respective Rho GTPases, inducing a conformational change that allows GTP binding and results in activation [41]. In EPEC and EHEC, the main LEE-encoded effector described as modulator of Rho GTPase activity is Map [42,43].

Disruption of gut integrity is another strategy employed by EPEC and EHEC to ensure colonization and induce diarrhea (Figure 1.3. C). Changing the absorptive surface area of the gut, modulating junctional complexes, or

altering selective channels are among the mechanisms used to affect the integrity of the intestinal epithelial layer. Together with intimin, the LEE-encoded effectors Map, Tir, EspB and EspF are involved in effacement of the microvilli, which results in a reduced gut surface area, and subsequently, in the impairment of normal absorptive processes [44-47].

Manipulation of host immune responses is another essential requirement for successful colonization and dissemination. In EPEC and EHEC infections, inflammatory responses are usually initiated by recognizing the bacterial flagella [48]. Subsequently, increased bacterial loads cause a reduced production of pro-inflammatory cytokines as well as interleukins, with several effectors acting together in order to illicit this effect (Figure 1.3. D) [49]. Subversion of host responses can also be achieved by inhibition of phagocytosis [50,51]. Both EPEC and EHEC have developed mechanisms to inhibit internalization of unopsonized (cis-phagocytosis) and opsonized bacteria (trans-phagocytosis) through several effectors (Figure 1.3 F).

Once the infection has been established, EPEC and EHEC can further manipulate host cell death by inducing apoptosis in a regulated manner by balancing pro and anti-apoptotic effectors, allowing these pathogens to maintain an infective niche within the host (Figure 1.3. E) [52-55].

The prominent translocated intimin receptor

Among a wide variety of effector proteins that exert distinct pathophysiological effects, we decided to focus our attention on Tir in this work. Tir is a unique LEE-encoded effector imperative for host cell attachment and signaling in EPEC and EHEC pathogenesis. Although EPEC and EHEC have evolved divergent signaling mechanisms, Tir effectors are structurally homologous, sharing about 60% sequence identity (Figure 1.4).

General Introduction

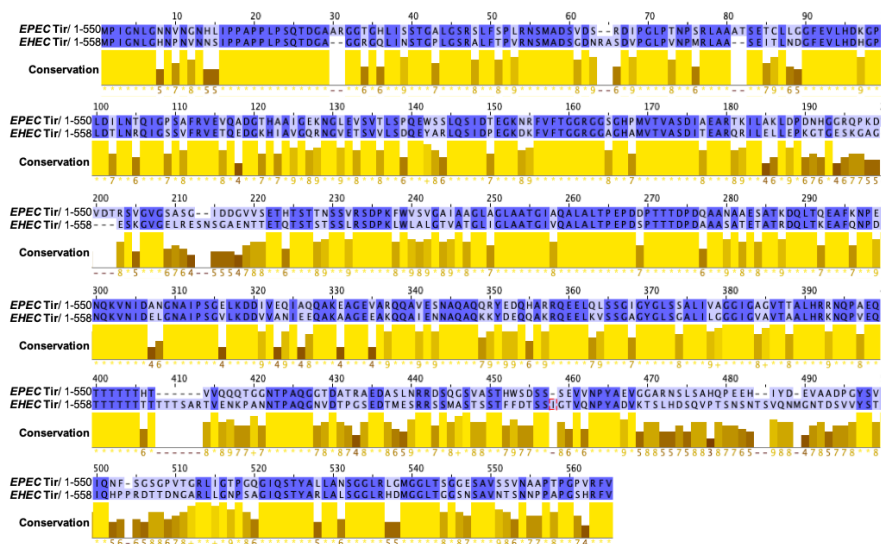


Figure 1.4. Sequence alignment of Tir from *E. coli* O127:H6 strain E2348/69 (EPEC) and *E. coli* O157:H7 (EHEC). Uniprot accession numbers for EPEC and EHEC Tir sequences are B7UM99 and Q7DB77, respectively. Sequence alignment was generated using the Clustal Omega server [56]. Alignment positions conserved in the two sequences are shaded in lavender. The conservation of physicochemical properties in each alignment position is reported in the corresponding barplot below the alignment.

Tir is comprised of two predicted helical transmembrane domains (residues 234–259 and 353–382 in EPEC) connected by a central Intimin-Binding Domain (IBD) (residues 260–352 in EPEC) and flanked by two cytosolic N- and C-terminal tail domains (residues 1–233 and 383–550 in EPEC, respectively).

Upon T3SS-mediated translocation into host cells, Tir is inserted into the apical host membrane adopting a hairpin topology. The helical transmembrane domains (TMD1 and TMD2) traverse the host membrane and the N- and C-terminal portions are located on the cytosol interacting with multiple intracellular eukaryotic proteins. In contrast, the extracellular IBD

domain mediates the tightly binding to the bacterial adhesin intimin at bacteria outer membrane [57] (Figure 1.5).

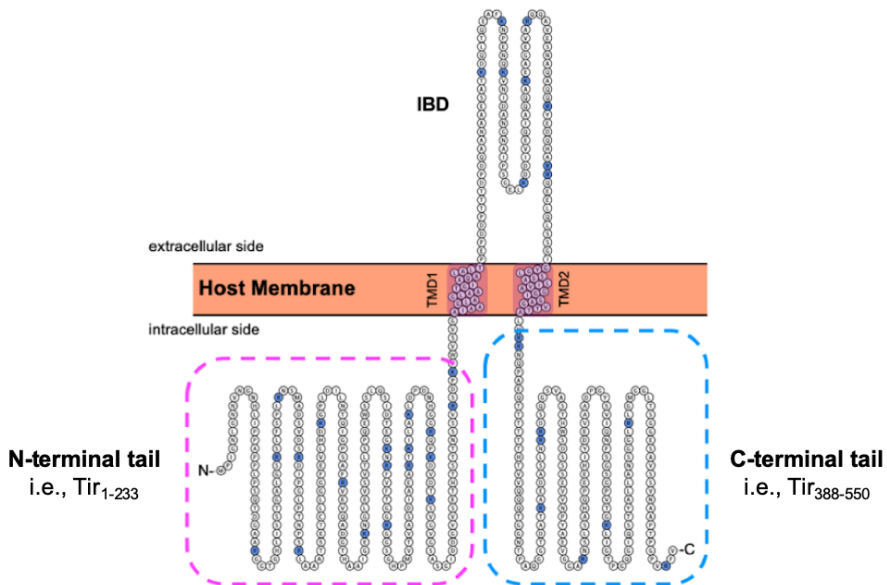


Figure 1.5. Tir A/E effector. Model of the 56 kDa protein Tir in host membranes adopting a hairpin topology with an extracellular intimin binding domain (IBD) flanked by two transmembrane domains (TMD) and termini dangling inside the host cytosol. This scheme displays the sequence of Tir from EPEC O127:H6 with its intracellular domains inside dashed boxes.

The IBD domain is composed of two helical motifs separated by a flexible extracellular loop that interacts with the C-terminal domain of intimin [58]. The high binding affinity between intimin and Tir, reported to have equilibrium dissociation constant (K_D) values of nearly 10 nM for EPEC and less than 1 nM for EHEC, is fundamental to ensure the tight attachment of these pathogens to the host cell [59]. The N-terminal tail Tir domain, also defined as the α -actinin-binding domain, is functionally interchangeable between EPEC and EHEC. Moreover, it can bind to several focal adhesion proteins and thereby tie together Tir to the host cytoskeleton, although in an insufficient way to trigger actin pedestal formation [60,61]. In turn, the C-

terminal tail domain is the most divergent portion of Tir between EPEC and EHEC, and it is absolutely required for pedestal formation, as it has been linked to dysfunction in actin nucleation and alteration of cellular profiles [60,61].

Cellular processes mediated by Tir

Actin pedestal formation. Post-translational modification of the C-terminal domain of Tir by the host components is essential for its activity. Host-mediated phosphorylation at multiple tyrosine, threonine, and serine residues in Tir's C-terminal region is crucial for Tir's virulent function. While phosphorylation of serine and threonine residues appear to play a role in Tir insertion into the host apical membranes and other processes such as cell motility and elongation [63], tyrosine phosphorylation by host Src-protein kinases is mandatory for proper and efficient actin pedestal formation [63,64].

EHEC and EPEC strains employ distinct Tir-mediated mechanisms to form actin pedestals [65]. The differences between these mechanisms appear to be related to the divergence of the C-terminal tail domain of Tir between EPEC and EHEC, sharing only 41 % of sequence identity [62]. The major pathway by which EPEC strains trigger actin pedestal formation (Figure 1.6, pathway 1) relies on the phosphorylation of the tyrosine residue 474 (Y474), the most critical C-terminal tyrosine residue for focal actin assembly [66].

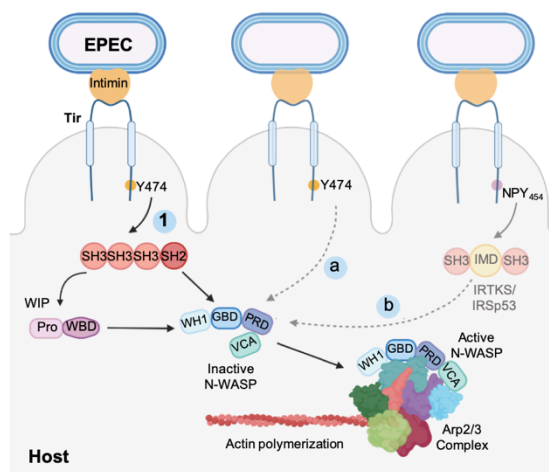


Figure 1.6. EPEC mechanisms of Tir-mediated signaling activation to form actin pedestals. The major actin assembly pathway is identified as pathway 1, and the minor pathways are represented as **a** and **b** (Adapted from [67]).

Tir phosphorylation promotes the interaction with the adaptor protein Nck through an SH2 domain that binds to a 12-residue region of Tir (EEHIYDEVAADP) where the tyrosine residue 474 is located [68,69]. In turn, Nck promotes the indirect activation of the neural Wiskott-Aldrich syndrome protein (N-WASP), through the intermediary adaptor WASP Interacting Protein (WIP). Consecutively, N-WASP activation allows the interaction with the actin-related protein 2/3 (Arp2/3) complex triggering actin polymerization [70]. Although the Nck-dependent pathway is the primary mechanism employed by EPEC to trigger actin polymerization, Tir is also able to promote actin assembly and pedestal formation in cells that lack the Nck adaptor protein, albeit at a much lower frequency (Figure 1.6, pathways a and b). In these pathways, phosphorylation of Tir on the tyrosine residues 454 and 474 allows the recruitment of N-WASP and Arp2/3 in the absence of Nck [71,72].

Interestingly, this peripheral mechanism in EPEC is highly related to the primary pathway of actin pedestal formation in EHEC strains, mediated by the NPY₄₅₈ tripeptide of Tir (Figure 1.7, pathway 1).

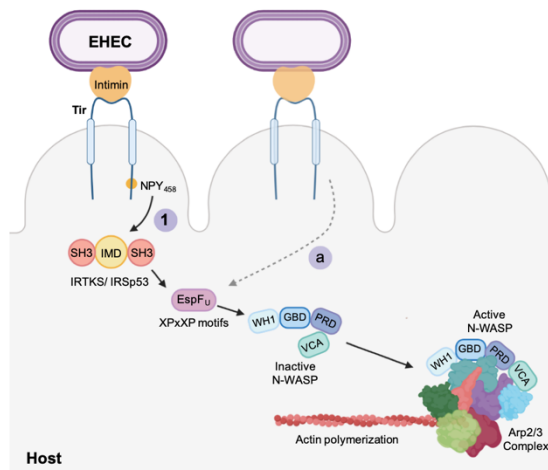


Figure 1.7. EHEC mechanisms of Tir-mediated signaling activation to form actin pedestals. The major and minor actin assembly pathways are identified as pathway 1 and a, respectively (Adapted from [67]).

In EHEC, the residues NPY₄₅₈ of Tir interact with the N-terminal IMD/I-BAR domains of the insulin receptor tyrosine kinase substrate (IRTKS) and with the IRSp53 adapter protein, two membrane remodeling proteins [72]. The Tir NPY-mediated pedestal formation pathway requires the additional Tir-cytoskeleton coupling protein (TccP) also known as EspFu (*E. coli* secreted protein F in prophage U), a non-LEE encoded effector [73,74]. EspFu/TccP is composed by an N-terminal secretion signal sequence, followed by a variable number of repeating units, varying from three to eight, each of which contains two structurally and functionally distinct domains, a helical domain N-terminal and a C-terminal proline-rich domain [74]. The C-terminal domain interacts with the SH3 domains in IRTKS and IRSp53, through multiple tandem PxxP motifs [75-77]. The N-terminal helical domain binds to and activates N-WASP, allowing the interaction with the Arp2/3 complex to trigger actin pedestal formation [78]. An alternative pathway that does not require the activation of N-WASP has also been found in EHEC (Figure 1.7, pathway a), although the involved elements remain unidentified to date.

Membrane phosphoinositide signaling. To influence the morphology of the actin pedestals, Tir is also responsible for changes in the composition of the plasma membrane by mediating membrane phosphoinositide signaling processes. In EPEC, Tir phosphotyrosine Y454 recruits and binds to the SH2 domain of host phosphoinositide 3-kinase (PI3K) to convert phosphatidylinositol 4,5-bisphosphate [PI(4,5)P₂] into phosphatidylinositol 3,4,5-trisphosphate [PI(3,4,5)P₃] in the membrane [79]. The conversion from PI(4,5)P₂ to PI(3,4,5)P₃ aids the downregulation of signaling pathways and generates a burst of actin polymerization responsible for pedestal reorganization into more compact structures that, ultimately, allow EPEC to firmly slide along the cell [79,80].

Tyrosine residues Y483 and Y511, located on the C-terminal domain of Tir, are also involved in regulating the actin-pedestal formation. These tyrosine residues are positioned within a putative tandem ITIM-like sequence (TILS) and therefore mimic eukaryotic ITIM-containing receptors by recruiting Src-homology 2 (SH2)-domain containing phosphatases, which dephosphorylate target proteins or lipids to restrict further downstream signaling pathways [81]. In EPEC, phosphorylation of Y483 and Y511 allows the recruitment of the cellular SH2-containing inositol 5-phosphatase (SHIP2) to dephosphorylate PI(3,4,5)P₃ and produce PI(4,5)P₂. The PI(3,4,5)P₃ intermediate is generated by PI3K early during infection, when SHIP2 recruitment is prevented, leading to the formation of multiple elongated pedestals, and then further converted to PI(3,4)P₂ by Tir-mediated recruitment of SHIP2, suggesting a regulated process of initial PI(3,4,5)P₃-enhanced actin polymerization followed by signaling down-regulation by conversion into PI(4,5)P₂ [80,82].

Suppression of host immune responses. Moreover, phosphorylation of Y483 and Y511 by host kinases has been shown to enable Tir's direct association with SHP-1 phosphatase [83]. Tir directly associates with SHP-1, promoting the recruitment of SHP-1 to the tumor necrosis factor (TNF) receptor-associated factor 6 (TRAF6). Consecutively, direct interaction between SHP-1 and TRAF6 prevents the ubiquitination of TRAF6 and inhibits

the activation of NF- κ B and MAP kinases as well as the expression of proinflammatory cytokines (Figure 1.8).

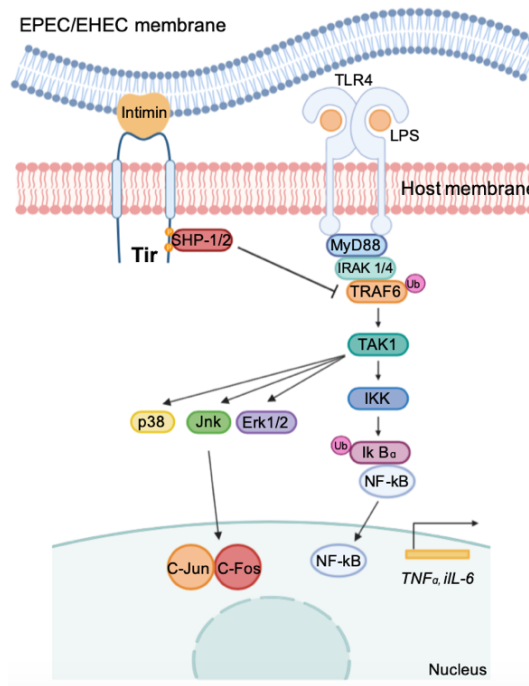


Figure 1.8. Tir-mediated suppression of host immune responses. Upon recognition of EPEC by macrophages, LPS triggers the activation of MAPK NF- κ B pathways, leading to the translocation of transcription factors into nucleus and transcription of cytokines. Once injected into host cells, Tir is phosphorylated at tyrosine residues within its ITIM-like motifs, leading to the recruitment of SHP-1 and SHP-2 and subsequently to the inhibition of TRAF6 autoubiquitination and downstream signaling pathways (Adapted from [84]).

The mechanism by which Tir promotes the association between SHP-1 and TRAF6 remains unclear. However, it is thought that the interaction between Tir and SHP-1 may promote a conformational change in SHP-1, making it more accessible to bind TRAF6 [83]. In many cases, SHP-2 plays the same role as SHP-1 regarding the control of host immune responses. Not surprisingly, SHP-2 has been shown to interact with Tir and promote

suppression of host cytokine production in the same way as it happens with SHP-1 [84]. However, the mechanism by which the complex of SHP-1/ SHP-2 and Tir is organized to negatively regulate host cytokine production remains unclarified. Despite it all, Tir has been considered a negative regulator of TLR signaling pathways by enhancing the recruitment of SHP-1 and SHP-2 to TRAF6 [83,84].

Regulation of Rac1 GTPase function. Aside the multiple mechanisms mediated by tyrosine phosphorylation, Tir is also serine phosphorylated by the host protein kinase A (PKA) as an additional strategy to communicate with host target cells, emphasizing the versatility of this bacterial effector. During infection, EPEC stimulates cellular PKA activation for Tir serine phosphorylation at positions 434 and 463, which is crucial for the correct insertion and assembly of Tir in the host cell membrane and interaction with intimin [85,86]. Activated PKA directly regulates the activity of Rho GTPase member Rac1, leading to reduced GTP-load and inhibition of cell elongation and thus contributing to the formation of A/E lesions [85,86].

Down-regulation of Map-dependent filopodia formation. Filopodia formation is mediated by Map, which in turn is down-regulated by Tir. The C-terminal of Tir domain contains a conserved putative arginine finger motif (RRXS) shown to interact with the host cellular target Cdc42. As Map signaling, mediated by Cdc42, has been shown to inhibit Tir–intimin-triggered pedestal formation, EPEC has evolved a mechanism by which Tir regulates Map activity and ensure appropriate levels of Cdc42 for filopodia formation prior to down-regulating Map-mediated Cdc42 signaling to trigger pedestal formation [87,88].

Induced pyroptosis. Tir-mediated actin polymerization stimulates the activation of the non-canonical NOD leucine-rich repeat and Pyrin domain-containing protein 3 (NLRP3) inflammasome, leading to macrophage pyroptosis [89]. Moreover, Tir has been recently shown to trigger pyroptotic cell death in intestinal epithelial cells in a mechanism dependent on the endogenous levels of gasdermin D (GSDMD), an executor of pyroptosis [90].

By promoting rapid Ca^{2+} influx, Tir clustering induces LPS internalization and promotes activation of caspase-4 causing pyroptosis [91].

In addition to its role in the aforementioned mechanisms and involved interaction partners, Tir has also been reported to bind to numerous other host cellular targets including cytokeratin 18, 14-3-3tau, α -actinin, cortactin, IQGAP1, talin, vinculin, and annexin A2, to regulate intestinal ion channels, and promote reorganization of the host cytoskeleton [92].

Our working hypothesis is that the versatility of Tir to mediate and participate in several distinct cellular processes relies on the presence of functionally relevant disordered regions, a trait found in a substantial fraction of the human proteome and other bacterial effectors. Thus, the following section aims to introduce the characteristics of disordered proteins and regions and the roles of disorder in biological functions, particularly in bacterial proteins.

Intrinsically Disordered Proteins

Proteins compose the dominant components of living cells in all kingdoms of life. These fundamental molecular machines engage almost all biological reactions highlighting their pivotal role in maintaining life itself.

For a long time, it has been thought that the unique three-dimensional structure of a specific protein predetermined its specific functionality [93]. According to this concept, the protein folding reaction can be described as a reversible transition between order and disorder. A given protein acquires its biological function when it folds up until reaching the structured state, representing the most favorable energetic conformation. On the other hand, the denatured state, in which a protein loses its function, would only be associated with the loss of its three-dimensional structure due to high temperature or pressure, extreme variations in pH value, presence of detergents, or chaotropic agents [94].

Over the last decades, however, several individual proteins lacking a well-defined 3D structure under physiological conditions but still playing important biological roles have been discovered [95,96]. Nowadays, such proteins are widely known as intrinsically disordered proteins (IDPs) [97].

Sequence-encoded signature of disorder

The ability of proteins to fold into a unique stable 3D structure or remain intrinsically disordered is encoded in its amino acid sequence. Although both ordered and disordered proteins resort to the same arsenal of 20 amino acid residues as major building blocks, there are significant differences between ordered proteins and IDPs at the level of their amino acid composition [95,98].

IDPs are characterized by low sequential complexity and a low number of “order-promoting” residues (cysteine, valine, leucine, isoleucine, methionine, phenylalanine, tryptophan, histidine, asparagine and tyrosine), which are mainly bulky hydrophobic residues normally present in the core of ordered proteins. In contrast, they are generally enriched in polar and charged residues, so-called “disorder-promoting” residues, (glutamine, serine, proline, glutamic acid, lysine, and occasionally glycine and alanine), which are typically found at the surface of ordered proteins [94,97]. The enhanced proportion in polar and charged residues favors electrostatic interactions at lipid bilayer surfaces [99]. As a result, intrinsic disorder is a common feature of membrane signaling receptors, particularly in their intracellular domains [100,101], with lipid-binding proteins accounting for 15% of all disordered proteins [102].

Besides having different amino acid compositions, ordered proteins and IDPs also distinctively use their aminoacids. For example, in ordered proteins, glutamic acid is critical for protein solubility and formation and stabilization of multiple structures such as α -helices and electrostatic valves in pores of ion channels. In turn, in IDPs, the over-abundance of glutamic acid residues is related to extended conformations of native coils and pre-molten globules [97].

Roles of disorder in IDPs functions

To understand the roles of disorder, it is important to consider the advantages that conformational flexibility offers to proteins specific functions [97]. IDPs not only tolerate this disordered structural state, but it is this disorder that predisposes them to unique functional modes complementary to globular proteins.

Structural flexibility and plasticity represent the primary functional advantages of IDPs. This flexibility and the ability to fluctuate through a set of structural states, due to lack of a stable 3D structure, allows them to engage in transient interactions with multiple partners, or various interactions with a single partner, presenting what is called binding promiscuity [103]. The binding selectivity and promiscuous multivalence are often conferred by modular recognition elements designated as Short Linear Motifs (SLiMs) [104], Eukaryotic Linear Motifs (ELMs) [105], or Molecular Recognition Features (MoRFs) [106], depending on the definition criteria.

Furthermore, IDPs often undergo structural changes by binding to specific partners allowing them to perform their biological functions. In such cases, the interaction involves a transition at a structural level from a disordered conformation to an ordered or partially ordered conformation. These transitions have an entropy cost, implying low to moderate affinities [107]. Although many IDPs undergo folding upon binding to their partners, others remain significantly disordered when bound, resulting in heterogeneous “fuzzy” complexes [108]. This outstanding structural flexibility allows the combination of low affinity and high specificity for binding-induced folding IDPs, as well as a higher binding promiscuity [109-111].

As sites within their disordered polypeptide chains are highly accessible, IDPs are preferential substrates for additional regulatory processes governed by post-translational modifications (PTMs), such as phosphorylation, acetylation, and/or ubiquitination. In these cases, flexibility allows the sites for PTM to be readily accessible, providing a transient but specific interaction with the enzyme. Furthermore, short linear motifs

employed to interact with binding partners can be fine-tuned by these modifications, and IDPs activity can be adjusted by selectively, permanently, or transiently, suppressing or enhancing the role of a single or several motifs [97,110,112].

The large hydrodynamic radius of IDPs, compared to folded proteins, provides an additional advantage to establish contacts with their multiple binding partners in a putative mechanism called “fly-casting” [113].

All these traits are particularly relevant in the context of cell signaling. IDPs often act as hub proteins at the center of several protein interaction networks. They are also commonly involved in the regulation of several intracellular processes such as regulation of cell division, transcription and translation, signal transduction, and regulation of large multiprotein complexes [111,114-117]. In the case of pathogenic IDPs, the ability to engage multiple partners can also be advantageous for interfering with host cell defense mechanisms, as it happens with Tir.

Not surprisingly, the role of IDPs in cellular regulation has further implications in disease [118]. As signaling pathways are tightly controlled at all levels, a proper activity requires high spatiotemporal precision and fast responsiveness to a large number of possible inputs. In this way, and as IDPs are consistently present as hub proteins of large interaction networks or regulators of multiple signaling pathways, deregulation of cascades, mutations or changes in cellular abundance of IDPs are often associated with the development of several diseases such as cancer and neurodegenerative disorders [119,120]. Given the ubiquitous role of IDPs in eukaryotic systems, protein disorder is also exploited by bacterial pathogens for infecting eukaryotic cells, as highlighted bellow.

IDPs in Bacteria

IDPs compose 30-50% of eukaryotic proteomes, corroborating their crucial role in cell signaling regulation [121,122]. In contrast, they are

remarkably less abundant in prokaryotes, making up less than 2-5 % of bacterial proteomes, suggesting that these proteins are less relevant for bacteria [123].

However, over the past decade, multiple studies have demonstrated that bacterial IDPs or intrinsically disordered regions (IDRs) are key players in regulating main cellular processes, including cell division, DNA and RNA metabolism, macromolecular assembly, and bacterial warfare [124].

Cellular processes such as transcription, cell signaling, and protein quality control are tightly controlled at all levels, requiring spatial and temporal regulation. In many bacterial systems, this is achieved by reversible formation of nonstoichiometric assemblies of multivalent macromolecules, called biomolecular condensates, which provide an effective spatiotemporal control by organizing biochemical reactions, allowing quality control and concentrating cellular matter [125,126].

Multivalency is achieved by the number of interaction sites designated as stickers, which can be hot spots located at the surface of folded protein domains or SLiMs in IDPs or IDRs [126], as illustrated in Figure 1.9.

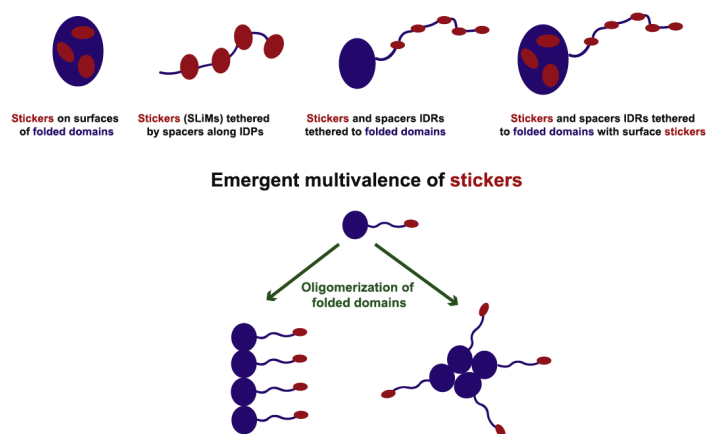


Figure 1.9. Schematic representation of intrinsic and emergent sticker multivalence. Intrinsic multivalence of stickers (top) can be present in different architectures. When a protein with a single sticker assembles into a larger complex, forming oligomers, emergent multivalence (bottom) occurs (Adapted from [124]).

In this way, IDPs and IDRs can influence several processes and molecular functions in bacterial systems, including cell division, DNA replication, transcription, post-transcriptional regulation, and stress responses. Well-known examples include the bacterial proteins FtsZ [127] and the single-stranded DNA binding proteins (SSBs) [128], both containing hypervariable disordered linkers shown to be crucial for cell division and regulation of binding cooperativity to single-stranded DNA, respectively. Both cases highlight the relevance of sequence hypervariability of IDPs /IDRs in bacteria as an evolutionary strategy for tuning condensate formation and regulating molecular functions.

Other cases include the disordered C-terminal domain of RNaseE, a critical driver of the formation of the RNA degradosome, and the disordered polar organizing protein (Pop)Z, a major player in differentiation, from *Caulobacter crescentus*, highlighting the relevance of disordered proteins and regions in the regulation of cellular processes in bacteria [129,130].

Several bacterial pathogens seem to exploit the lack of structure and flexibility of IDPs and IDRs to bypass direct recognition by the host and trigger the initiation of effector-triggered immunity mechanisms [131].

Structural flexibility can be advantageous during secretion as it reduces the requirement of effectors to be actively unfolded before delivery by the secretion systems. Moreover, it has been observed that several effectors undergo partial folding upon binding their chaperones to promote the translocation through the secretion system. An example can be given by the YopE effector, from *Yersinia*, that undergoes disorder-to-order transitions by binding to its chaperone SycE, forming a targeting signal that allows its translocation through the T3SS [132].

In addition, to facilitate secretion, other disordered regions in bacterial effectors have evolved to enhance the subversion of host signaling mechanisms. In such cases, the structural flexibility provides an additional advantage, as it allows IDPs and IDRs to adopt multiple conformations and therefore interact with structurally divergent host targets [133]. IDPs and IDRs

have a more exposed surface area, and thus can expose more binding sites at one time and a lower cell energetic cost when compared to ordered proteins, essentially getting more for less [116]. This phenomenon can be illustrated by the disordered C-terminal region of Tir that can interact with multiple host targets by containing multiple ITIM-like sequences. A similar example is the oncogenic effector CagA from *Helicobacter pylori* that operates as scaffold, providing a platform for multiple interactions with host proteins, through its Glu-Pro-Ile-Tyr-Ala (EPIYA) and multimerization (CM) short linear motifs located within the disordered C-terminal region [134].

Structural flexibility not only allows IDPs to interact with multiple host proteins as it can make them more accessible to undergo post-translational modifications to further regulate multiple protein-protein interactions and therefore subvert the host machinery. Both ITIM-like sequences of Tir and EPIYA motif of CagA are phosphorylated by host kinases, promoting their interaction with Src homology 2 (SH2) domain-containing proteins, such as the tyrosine phosphatase SHP-2. These interactions unlock the autoinhibitory conformation of SHP-2 by inducing a conformational change. In EPEC and EHEC, the interaction of Tir and SHP-2 facilitates the recruitment of TRAF6, inhibiting its ubiquitination and the subsequent activation of several intracellular signaling cascades such as the mitogen-activated protein kinase pathway, as well as the expression of proinflammatory cytokines [84]. On the other hand, the interaction between CagA and SHP-2 leads to aberrant phosphatase activity, promoting the activation of pro-oncogenic mitogenic Ras-ERK signaling [135].

Enrichment of aromatic residues in linear motifs or short molecular recognition elements of IDPs serves an additional way to mediate protein-protein interactions. For example, the intrinsically disordered effector EspF_U, from EHEC, interacts with high affinity to both N-WASP and IRTKS, through its XPxXP motifs, forming a tri-molecular complex. Curiously, it has been observed that in the linker between the two XPxXP motifs in EspF_U there is a tryptophan residue that is absent in known host SH3-domain containing binding partners of IRTKS, suggesting that EHEC may have evolved by switching this

residue in EspFu to outcompete host cellular targets by promoting a higher binding affinity [136].

Ironically, the structural heterogeneity and conspicuous dynamics that predispose IDPs/IDRs to sustain special functional modes, are also the main hurdles for its biophysical characterization. NMR spectroscopy is arguably the best method to study IDPs, because of the versatility of this technique towards highly flexible and dynamic systems. The following section details the principles of the method, the associated difficulties in characterizing IDPs, and several applications employed in this work.

NMR Spectroscopy in characterization of IDPs

NMR spectroscopy offers a wide range of applications in biology and chemistry, ranging from small organic and inorganic compounds to polymers and biomolecules. NMR allows the structural characterization of biological macromolecules at atomic resolution along with X-ray crystallography and electron microscopy. However, an important advantage of NMR is that biological samples can be studied in the liquid state enabling structural studies of dynamic systems as IDPs, often of challenging crystallization.

Principles of NMR spectroscopy

NMR spectroscopy relies on the interaction between active nuclear spins, subject to an external magnetic field, with electromagnetic waves in the radio frequency range. This section will introduce the fundamentals of this technique following the vector geometry approach, treating atoms as magnets inside an external magnetic field.

The nuclear magnetic dipole moment $\vec{\mu}$, comes from the spin angular moment $\vec{S}$ of the nucleus, described in equation 1.1:

$$\vec{S} = \hbar\sqrt{I(I+1)} \quad (1.1)$$

where $\hbar$ is the reduced Planck constant and I is the spin quantum number. In order to be observable by NMR spectroscopy, an atomic nucleus has to possess a non-zero spin quantum number I . For common nuclei in biomolecular NMR, such as ^1H , ^{13}C and ^{15}N , spin quantum number I has two allowed states, $I = \frac{1}{2}$ and $I = -\frac{1}{2}$, with high and low energies, that can be described as magnetic poles of a magnet spinning around the z -axis. Its z components are determined by the magnetic quantum number m , which can assume values of $-I, -I+1, \dots, +I-1, +I$, in a total of $2I+1$ energy states possible. The quotient of the nuclear magnetic moment and the spin angular momentum is defined as the nuclear gyromagnetic ratio, γ , and it is intrinsic of each nucleus. Therefore, the nuclear magnetic momentum $\vec{\mu}$, can be described by:

$$\vec{\mu} = \gamma \vec{I} \quad (1.2)$$

When nuclei are placed in an external magnetic field, that is by convention aligned along the z -axis, the nuclear magnetic moment $\vec{\mu}$ interacts with the magnetic field $\vec{B}_0$, leading to the precession of the nucleus around the magnetic field and splitting of states (Figure 1.10).

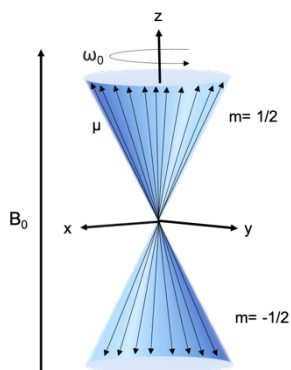


Figure 1.10. Schematic representation of magnetic dipole orientation and precession for a nuclear spin $\frac{1}{2}$ in the presence of an external magnetic field B_0 . m is the magnetic quantum number, μ is the magnetic dipole moment and ω_0 is the Larmor frequency.

The splitting of the system into $2I+1$ energy levels, denominated Zeeman splitting, is described by equation 1.3.

$$\Delta E = \hbar \gamma \vec{B}_0 \quad (1.3)$$

As seen in Figure 1.10, the precession frequency for each state, called Larmor frequency, ω_0 , is the resonance frequency of each nucleus. The absorption frequency for each transition depends on both strength of the magnetic field and the gyromagnetic ratio of the nuclei, and therefore the Larmor frequency can be described by the equation 1.4.

$$\omega_0 = \gamma \vec{B}_0 \quad (1.4)$$

To record an NMR signal, magnetization must be transferred from the z-axis to the x-y plane. In practice, this is achieved by applying a radio frequency pulse along an axis in order to shift the magnetization, splitting the energy states, as illustrated in Figure 1.11.

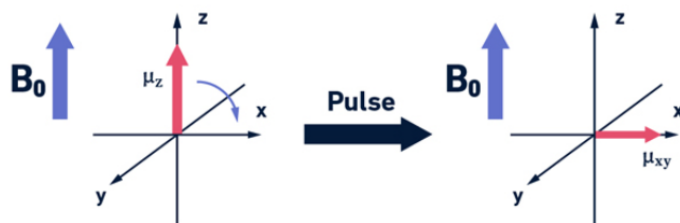


Figure 1.11. Schematic representation of magnetization transfer to the x-y plane upon application of a magnetic pulse.

As a result, a new magnetic local field is created with the duration of the pulse applied. Once the pulse terminates, magnetization returns to equilibrium, precessing about the z-axis. Recording of magnetization decay over time produces the free induction decay (FID). Finally, by mathematical

application of a Fourier transform (FT), this time-domain information is converted into frequency-domain information and a spectrum is obtained.

Peak positions in NMR spectra are dependent on the resonance frequencies of the observed nuclei. In addition to energy splitting, nuclei experience several other interactions that are dependent on the surrounding chemical environment, meaning that nuclei resonate at distinct frequencies if they sense different environments within the molecule. This occurs as a consequence of the shielding effect of the electronic cloud surrounding each nucleus, causing the local magnetic field B to slightly differ from the applied magnetic field B_0 . As so, the local magnetic field B can be written as:

$$B = (1 - \sigma)B_0 \quad (1.5)$$

where σ is the anisotropic shielding factor, describing the electron density around the nucleus.

The observed frequency, influenced by the local magnetic field B , is called the chemical shift. Chemical shifts are characteristic of each nucleus chemical surrounding, thus often providing the first useful information obtained from the NMR spectrum. Chemical shifts are measured against a reference compound and expressed in parts per million (ppm), and are defined by the following equation:

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 \quad (1.6)$$

where ω_0 is the magnet operating frequency in MHz, Ω is the offset of the nucleus of interest in Hz, and Ω_{ref} is the offset of the reference compound in Hz.

Challenges of IDPs studies by NMR spectroscopy

IDPs inherent flexibility can provide both advantages and disadvantages when it comes to its study using NMR spectroscopy. A major challenge is imposed by their reduced spectral resolution, as IDPs display low frequency dispersion in the ^1H dimension. Because of the lack of structure, proton chemical shift values are similar to those in random coil systems. Furthermore, low complexity in IDPs sequences, makes nuclei resonances nearly identical, leading to extensive spectral signal overlap [137,138].

IDPs with random-coil like structures lacking transient structure have large ^{15}N T_2 relaxation time constants. As a result, there is an increased sensitivity and sharp lines in the amide proton-nitrogen correlation spectra are observed. On the other hand, in regions with extensive transient structures, T_2 relaxation time constants can be significantly shorter than those in other regions, leading to a large dynamic range of peak intensities [137,138].

Aggregation can also lead to reduce sensitivity. Some IDPs, containing exposed hydrophobic regions, such as α -synuclein [139], tend to aggregate, thus restricting NMR studies as aggregation limits sample concentration in solution, and high-order oligomers can become undetectable at least by liquid-state NMR. The proteolytic degradation susceptibility of IDPs is another challenge. In some cases, this can be overcome with heat denaturation by performing sample boiling to deactivate proteases, as performed in this work. However, if there are exposed hydrophobic regions, thermal denaturation can cause irreversible changes in residual structure and even more lead to precipitation, meaning that, for these situations, NMR data needs to be hastily acquired before sample degradation arises [137,138].

IDPs' high sensitivity to measurement conditions, such as pH and temperature, can also be a hurdle for NMR spectroscopy, particularly when physiological conditions want to be mimicked or when enzymatic reactions, such as post-translational modifications, are studied. These situations often require elevated pH and temperature conditions, leading to fast exchange of amide protons with the solvent water protons and, consequently, to

broadening of amide resonances and to a dynamic range of peak intensities in the amide proton nitrogen correlation spectrum [137,138].

NMR applications to characterize IDPs

NMR spectroscopy plays a vital role in the characterization of IDPs, as it is the only spectroscopic technique capable of providing structural and dynamic information at the atomic level of highly dynamic and disordered systems.

Resonance assignment. The first step required to obtain structural information by NMR is the assignment of resonances. In the case of IDPs this process can be particularly challenging. IDPs display low chemical shift dispersion of the amide proton signals, leading to poor spectral dispersion. Due to its dynamic and flexible nature, their chemical shifts collapse in region of random-coil conformations (^1H range: 7.5-8.5 ppm), causing significant resonance overlap [137,140]. Additionally, as IDPs are highly dynamic, they are remarkably exposed to solvent compared to proteins with a three-dimensional structure. This exposure leads to extensive exchange processes between the amide and water protons, resulting in substantial broadening and loss of intensity of the amide proton signals, and consequently in decreased sensitivity and resolution of the spectra [140].

Fortunately, nowadays, these problems can be mitigated through the application of ultra-high field spectrometers, employment of non-uniform sampling methodologies [141], acquisition of higher-dimensional spectra [142] or use of ^{13}C direct-detect NMR experiments [143] which have been emerging as a leading alternative to proton-detection for IDPs, owing to the excellent chemical shift dispersion and narrow linewidths. Additionally, as most IDPs are stable over a wide pH range, working at slightly acidic pH can be used to overcome some proton exchange processes [144]. Spectra acquisition at lower temperature is other strategy employed to minimize proton chemical exchange and acquire spectra with higher resolution [145]. Alternatively, in the case of higher molecular weight proteins, a so-called “divide and conquer”

approach, through use of smaller truncated versions of a protein, is often used to reduce spectral complexity, as it has been done in the present work.

Mapping transient structural elements and long-range contacts.

Secondary structure in IDPs is often transient and located in short segments that may contain binding motifs for biological partners or post-translational modification sites with functional relevance. As an atomic resolution technique, NMR spectroscopy can be employed for the detection of secondary structural propensities. Because of their high sensitivity to structure, chemical shifts are typically the elected probes for detection of secondary structure propensities [146]. Chemical shifts are known to be exquisite reporters of backbone conformation as they are influenced by the local environment of a nucleus, and therefore directly depend on the primary, secondary and tertiary structure of a given protein [147,148]. As so, chemical shifts can be employed as conformational restrains for structure determination as well as for defining structural ensembles of IDPs [149].

As in solution IDPs are present as interchanging conformers, the observed chemical shifts can be described as a populated-weighted average of these interconverting conformers over a time scale.

Secondary chemical shifts ($\Delta\delta$) are calculated as deviations from random coil determined values and are particularly useful to describe local geometries in IDPs and quantify local secondary structure elements or secondary structural propensities. Generally, positive values of secondary chemical shifts from $^{13}\text{C}\alpha$ ($\Delta\delta\text{C}\alpha$) indicate a propensity for helical structures, while negative ones are indicative of a preference for β -sheets. As for the values of secondary chemical shifts of $^{13}\text{C}\beta$ ($\Delta\delta\text{C}\beta$) the effect is the opposite. Thus, it is convenient to monitor the difference between $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ secondary chemical shifts ($\Delta\delta\text{C}\alpha - \Delta\delta\text{C}\beta$). Overall, positive values of $\Delta\delta\text{C}\alpha - \Delta\delta\text{C}\beta$ support a tendency for α -helical elements of secondary structure, while negative values suggest a preponderance for β -sheets [150].

However, as the result of secondary structural predictions is reliant on the need of reference random coil chemical shifts, systematic errors can be

introduced if the referencing is incorrect. This aspect becomes even more crucial for IDPs, as the experimentally measured chemical shifts are generally close to the corresponding random coil chemical shift values, making the discrimination between the contribution of primary and secondary/tertiary structures to the chemical shift of particular importance. Thus, secondary chemical shifts should be calculated taking into account the contribution of neighboring amino acids, in order to improve the reliability of structural propensity predictions [147,148].

Aside from chemical shifts, frequently employed to detect secondary structure propensities, paramagnetic relaxation enhancement (PRE) is the most frequently employed NMR experiment to detect long-range interactions in IDPs. PRE experiments are based on the dipolar interaction between nuclear spins and the spin of an unpaired electron from a paramagnetic component which induces increased relaxation rates. As the PRE is dependent on the correlation time and on the distance between the paramagnetic probe and the NMR active nucleus, it can provide distance information with an upper limit of 25 Å [151,152].

In ideal random coils, the nuclei of residues in the vicinity of the residue carrying the paramagnetic probe experience a strong PRE and as result its resonances are broadened. In turn, signals from residues located upstream and downstream of the tag will experience less or even no broadening as the PRE effect decreases with increasing distance from the paramagnetic tag. However, if long-range contacts are present, the nuclei of residues sequentially distant can be spatially close to the paramagnetic tag and will experience the PRE effect more frequently than would be expected from ideal random coil behavior (Figure 1.12).

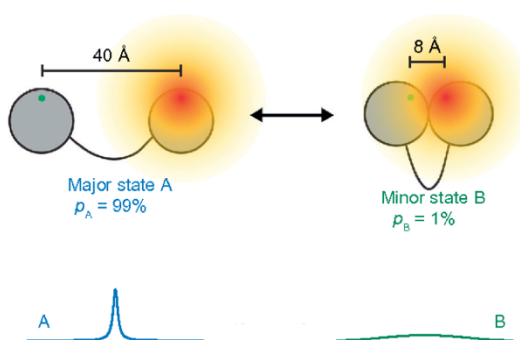


Figure 1.12. Paramagnetic probes for detection of long-range interactions. In the presence of long-range contacts, nuclei of residues distant in the primary sequence can be spatially close to the tag and experience the PRE effect. The unpaired electron of the paramagnetic probe and a proton of interest are depicted as green and red dots, respectively (Adapted from [153]).

For PRE studies of IDPs, a paramagnetic tag is usually attached to a single amino acid in order to be used as a probe. This is generally achieved by conjugation of a thiol-reactive nitroxide compound, most commonly the thiol-reactive methanethiosulfonate spin label (MTSL), to a single cysteine residue previously introduced by site-directed mutagenesis. If other cysteine residues are present in the protein of study they must be formerly removed prior to the conjugation of the paramagnetic tag. Once the paramagnetic probe has been attached in the desired position, the spectrum of the paramagnetic sample is acquired. After this, MTSL is reduced with ascorbate, without cleavage of the disulfide bond and consequent label detachment, and a spectrum of the diamagnetic reference sample is acquired. PREs are then extracted from a single pair of ^1H - ^{15}N -HSQC spectra acquired in the presence and absence of a paramagnetic probe [140,152].

The most common approach to determine PREs in proteins relies on the measurement of ^1H transverse relaxation rates, R_2 , for the amide H-N pairs in a uniformly ^{15}N -isotopic labeled sample. In the paramagnetic sample, the observed values for R_2 relaxation rates contain information of the native

diamagnetic term plus and additional paramagnetic contribution. This way, by subtracting the diamagnetic term, upon acquisition of a reference spectrum, the paramagnetic contribution can be obtained [154].

However, in IDPs studies, the spatial averaging needs to account for the variability of the unpaired electron position as well as the dynamics of coupled nuclei, which is far more complex [155]. Furthermore, as in IDPs relaxation properties of spins are variable and dependent on the extent of transient structure in different parts of the molecule, peak intensities are usually heterogeneous. Intensity variations can also arise from multiple additional factors as differential solvent exchange rates, residue type, temperature, pH and the extent of solvent-exposition of the amide proton. Therefore, in these cases, a simplified and semi-quantitative approach based on 2D ^1H - ^{15}N correlation spectra is commonly employed. As R_2 is directly related to linewidth, and thus, inversely related with signal peak intensity, the ratio between peak intensities in both paramagnetic and diamagnetic samples ($I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$) can be used to report long-range contacts [156].

Monitoring post-translational modifications. Post-translational modifications (PTMs) are essential for proteins to undergo rapid and reversible changes in structure and function in response to internal and external stimuli, playing a vital role in regulating of multiple cellular processes, including transcription, translation, cell cycle, and immune response.

As IDPs' conformational heterogeneity and structural plasticity provide greater access to modifying enzymes and can facilitate interactions with multiple binding partners, PTMs are critical regulators of IDP function [112]. PTMs can affect binding due to different mechanisms, including changes in secondary structural propensities, dynamics, or charge-charge interactions [157]. Although the effects of PTMs on IDPs can be varied, the covalent modification in the amino acid side chains frequently generates perturbations to the nearby residues, making NMR one of the main techniques employed to study these modifications. Phosphorylation is one of the most well-studied post-translational modifications as it represents a crucial regulatory mechanism of eukaryotic cell signaling pathways. This mechanism, carried

out by protein kinases, consists of adding of a phosphate group (PO_4) on the side chains of serine, threonine, and tyrosine residues. As a result, polarity increases, often allowing a protein to interact with other molecules.

Standard 2D ^1H , ^{15}N -HSQC correlation spectra are particularly useful for identifying and assigning phosphorylated residues, particularly Ser and Thr, as adding a phosphate group strongly perturbs nearby ^1HN resonances. These significant chemical shift changes are primarily caused by the hydrogen bond formed between the protons of the backbone amide group and the phosphate group added in the phosphorylation reaction (Figure 1.13). Thus, identifying protein phosphorylation sites can be achieved by chemical shift difference mapping of unmodified and phosphorylated substrate residues [158,159].

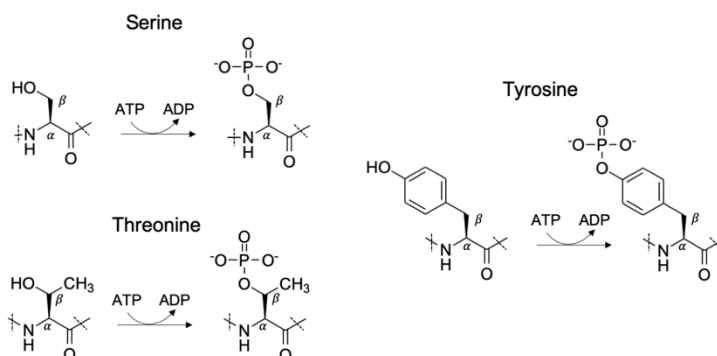


Figure 1.13. Phosphorylation of Serine, Threonine and Tyrosine residues.

A phosphate group (PO_4) is added to each residue by protein kinases during the phosphorylation reaction.

Phosphorylation on tyrosine residues does not induce such significant changes on the backbone amide chemical shift, due to the most distal position of the hydroxyl group compared to serine and threonine residues, rendering it to be harder to identify in ^1H , ^{15}N -HSQC spectra. As an alternative, mapping of CH_ϵ resonances can be achieved as these are substantially affected, working

as unequivocal indicators of the presence of phosphorylated tyrosine residues [158,159].

Multi-site phosphorylation is particularly relevant and prevalent in IDPs involved in signaling pathways. However, these multiple events often lead to increasing complexity in spectra and require additional phospho-detection methods to analyze the modified amino acid resonances, such as real-time NMR (RT-NMR) [138,159]. In order to follow phosphorylation reactions *in vitro*, ^{15}N isotope-label protein substrates are added to a reaction mixture containing the unlabeled recombinant kinases, ATP, Mg^{2+} and other additional cofactors needed for kinase activity. Consecutive sets of time-resolved NMR experiments are then employed to directly detect phosphorylation of individual substrate sites as it leads to time-dependent evolution of characteristic chemical shift changes. Phosphorylation-induced changes of NMR signals can then be followed given that unmodified and phosphorylated residues are detected side-by-side. Finally, by extracting signal the intensities of both unmodified and modified substrate residues, site-specific information about individual modification levels and evolution over-time can be retrieved (Figure 1.14) [160].

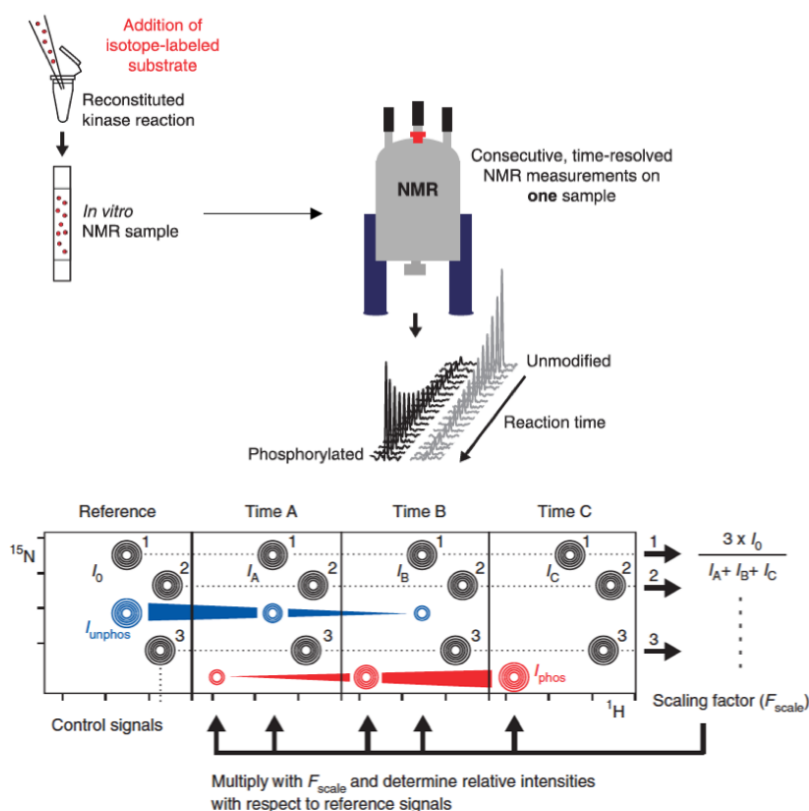


Figure 1.14. Schematic overview of NMR procedures for monitoring real-time protein phosphorylation reactions *in vitro*. Upon sample preparation multiple consecutive time-resolved NMR measurements are acquired. Phosphorylation-induced changes of NMR signals (reaction time points A, B, C) can be followed as the disappearance of NMR signals of unmodified substrate sites (blue) coincides with the appearance of NMR signals of the phosphorylated residues (red) (Adapted from [160]).

The major disadvantage of NMR spectroscopy in continuous phospho-detection studies of protein phosphorylation reactions is the inherently poor sensitivity of the technique, requiring substrate concentrations in the μM range. When working at this concentration range, it takes approximately 4 minutes to acquire a linearly sampled 2D spectrum, meaning that individual measurement time points are set 4 minutes apart. This can be particularly inconvenient if rapid phosphorylation reactions are under study.

There are, however, several approaches to overcome these limitations, such as using smaller amounts of recombinant kinase in the reaction to slow down the phosphorylation reaction or changing temperature and/or pH to specifically decrease kinase activity [160,161]. Another alternative is the use of quenched phosphorylation reactions and discontinuous NMR measurements, which allow extracting quantitative information about individual phosphorylation states by using shorter time intervals as well as lower substrate concentrations. Unlike RT-NMR experiments, NMR measurements of quenched discontinuous phosphorylation reactions require multiple *in vitro* kinase reaction mixtures, to which equal amounts of ^{15}N isotope-labeled substrates are added for different time periods (Figure 1.15).

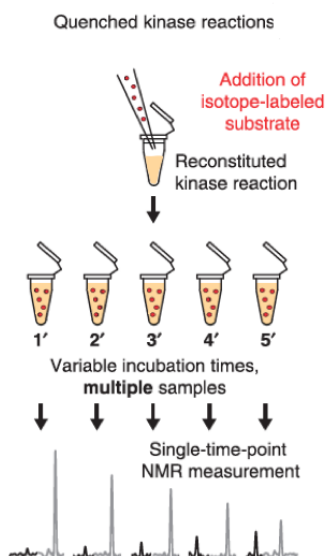


Figure 1.15. Schematic overview for monitoring discontinuous NMR measurements of quenched phosphorylation reactions. A sample of ^{15}N isotope-labeled substrate lacking the recombinant kinase is initially prepared to be used as reference. Multiple *in vitro* kinase reaction mixtures are then prepared, to which equal amounts of ^{15}N isotope-labeled substrates are added at different reaction time points. At the desired time points, the phosphorylation reactions are stopped by heat inactivation and single 2D ^1H , ^{15}N -HSQC spectra are acquired for each sample using the same acquisition parameters as for the reference non-phosphorylated sample (Adapted from [160]).

At the desired time points, phosphorylation reactions are stopped by heat inactivation of the kinase, and single 2D ^1H - ^{15}N spectra are acquired for each time point allowing further quantification of early protein phosphorylation levels [159]. Because phosphorylation reactions are stopped by heat inactivation, which usually denatures folded proteins, this approach is particularly suited for studies with IDPs or peptides.

As these NMR experiments disclose single-time phosphorylation states, instead of progressive modifications throughout the phosphorylation reaction as in continuous NMR studies, substrate concentrations below 25 μM may be used, and 2D ^1H - ^{15}N NMR spectra should be acquired with a high number of scans to accurately measure low signal intensities of early phosphorylated substrate species [159].

Monitoring protein-protein interactions with binding partners.

NMR spectroscopy is especially well suited to deliver the quantitative information needed to study interactions and complexes formation by IDPs and better understand its molecular mechanisms, as it can provide residue-specific structural information concerning the binding interface, as well as give reasonable estimates of binding constants [162,163].

Characterization of protein-protein interaction complexes by NMR spectroscopy is typically achieved by titration. A solution containing an ^{15}N or ^{13}C isotope-labeled protein (protein A) is titrated with increasing amounts of an unlabeled protein (protein B). Upon each addition of titrant, a heteronuclear correlation spectrum, usually ^1H - ^{15}N HSQC is acquired. In this way, it is possible to monitor transitions between free state protein A and bound state to form the complex AB [162,164].

As chemical shifts are very sensitive to changes in the chemical environment, working as powerful probes of binding interactions, it is possible to detect protein-protein interactions even when they are weak. Considering the aforementioned system, as protein B is titrated over the isotope-labeled protein A the observable chemical shift changes in the ^1H - ^{15}N HSQC spectrum will arise from changes in chemical environment of nuclei affected by the

interaction between these two proteins. Therefore, by monitoring the chemical shift changes at each titration point it is possible to retrieve information about the residues located at the interaction interface, as the chemical shift values of these residues will be the most affected [162].

As an example, a simple two-state reaction system, between two proteins A and B that form a complex AB, taken as an example, is describe in equation 1.7,



where K_{on} and K_{off} describe the on and off rates of the complex AB, respectively.

Considering this system, protein A will give rise to two species with different resonance frequencies, ω_A and ω_{AB} , and, consequently, to a resonance difference frequency defined by $\Delta\omega = \omega_A - \omega_{AB}$. However, appearance of different species in NMR spectra varies and it is strongly dependent on three factors: concentration of the species, dissociation constant, K_D , and exchange rate of reaction, k_{ex} , defined by:

$$K_D = \frac{[A][B]}{[AB]} \quad , \quad k_{ex} = k_{on}[B] + k_{off} \quad (1.8, 1.9)$$

The rate constants also determine the populations of the free protein, A, and bound protein, AB, which at equilibrium are given by:

$$P_A = \frac{k_{off}}{k_{on}[B] + k_{off}} \quad , \quad P_{AB} = \frac{k_{on}[B]}{k_{on}[B] + k_{off}} \quad (1.10, 1.11)$$

Considering a diffusion-controlled interaction, in which the interaction site is stereochemically accessible, the values of K_{on} are constant and in the order of $10^9 \text{ M}^{-1}\text{s}^{-1}$, existing an approximate relationship between K_D and K_{on} , where $K_{off} \approx 10^9 K_D$. Such considerations are relevant in understanding and distinguishing different protein-protein interaction chemical change regimes. Chemical exchange regimes are typically classified into three categories: slow, intermediate and fast exchange. This classification relies on the magnitude of k_{ex} relative to the size of the chemical shift (given by the difference in resonance frequency) for a given nucleus in A and AB states [162,165].

Slow chemical exchange regimes occur when the exchange rate is much slower than the observed difference in resonance frequency ($k_{ex} \ll \Delta\omega$). The intensity of a subset of peaks for the free state (protein A) progressively diminish with increasing amounts of titrant (protein B), and, at the same time, cross-peaks corresponding to the bound state of protein (AB) start to appear and grow in intensity. As a result, each state and the corresponding resonance frequencies are individually observed (Figure 1.16 A).

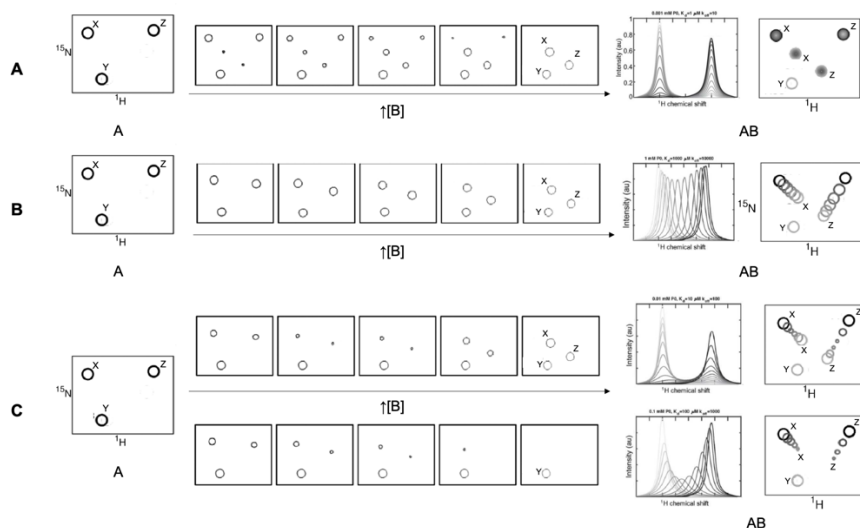


Figure 1.16. Schematic representation for the titration of a labelled protein A with increasing concentration of unlabeled protein B in (A) slow, (B) fast and (C) intermediate (top) and fast/intermediate (bottom) exchange regimes. At left, the spectra of the ligand-free protein A are depicted containing three cross-peaks representing the backbone NH signals for fictional residues X, Y and Z. At the right the successive steps of each titration are superposed in single1D and 2D panels. The spectra recorded with increasing molar concentrations of protein B is depicted on a grey scale: with black and light corresponding to free labeled protein and excess ligand B, respectively (Adapted from [162,165]).

As in slow exchange regimes, relative population changes do not cause peaks to move in the spectrum, peak position itself does not provide a quantitative measure of the binding event. Nevertheless, binding events can be directly monitored by the changes in intensities observed for the bound state, and apparent dissociation constants can be extracted by fitting these changes in intensity of the free- and bound-state cross-peaks to a saturation binding curve. Furthermore, peak intensities can be employed to determine relative population sizes of the free and bound states [162,165].

On the other hand, when the exchange rate is fast in the NMR time scale ($k_{\text{ex}} \gg \Delta\omega$), a single peak is observed at the population-weighted

average position of their respective resonance frequencies (Figure 1.16 B). The peaks perturbed migrate systematically in a linear fashion towards a new position in the spectrum, and most times little, if any, changes in peak height are observed. Thus, binding events can be monitored by plotting the change in chemical shift as a function of the concentration of protein B (ligand) added, obtaining a rectangular hyperbolic transition from the free to bound state, which can be fitted to a standard saturation binding curve. In contrast, and in opposition to a slow exchange situation, the peak intensities will not correctly provide a quantitative measure of the binding event [162,165].

Nevertheless, in both slow and fast exchange regimes, plotting the intensities or changes in chemical shift, respectively, as a function of the molar ratio between protein and ligand ($[A]/[B]$) allows the estimate of K_D and, thus, the quantification of binding affinities [165]. While typically fast exchange occurs where the off-rate (K_{off}) for protein A from the complex is relatively rapid, slow exchange arise when the K_{off} is relatively slow. Thus, usually, fast exchange is associated with weak binding events whereas slow exchange occurs for strong intermolecular interactions [162,165].

By last, when K_{off} is of the order of the chemical shift differences involved ($(k_{ex} \approx \Delta\omega)$), intermediate exchange conditions prevail, and a single peak, markedly broadened by the exchange process, is observed (Figure 1.16 C). In this scenario, a progressive change in the chemical shifts of perturbed peaks is observed, accompanied by initial line broadening and decrease, followed by recovery of peak intensity in the course of the titration. However, as in this situation both peak positions and intensities are eminently uncertain, quantitative interpretation is fairly complex, problematic, and often not feasible [162,165].

Overall, the schematic representations of chemical exchange regimes presented above are overly simplified. First, complex formation itself could generate an overall broadening of all peaks in the spectrum due to an increase in the overall rate of transverse magnetization decay associated with the slower tumbling time of the complex compared to the isolated protein. Furthermore, as chemical exchange regimes rely on the magnitude of

chemical shift changes, which are ligand-dependent, the three different behaviors (i.e., fast, intermediate and slow) can be observed in the same experiment. Moreover, the outcome of these experiments depends on the spectrometer used as the absolute frequency differences between the signals from free and bound states are dependent on the magnetic field strength [162,165].

Other methods to study IDPs

In addition to NMR spectroscopy, other methods can be used for the experimental characterization of disordered proteins, providing valuable insights [166]. In this thesis we also employed Small-angle X-ray scattering (SAXS) and Circular dichroism (CD) to complement the NMR data on C-Tir. Below, both methods are briefly described.

CD spectroscopy. CD is a straightforward and powerful technique that allows obtaining macromolecular structural information by assessing secondary conformational properties of proteins. It relies on the measurement absorbance of right and left circularly polarized light in the UV frequency range.

The far-ultraviolet (UV) CD spectrum of a protein is the result of the combination of spectral bands corresponding to the different secondary structure elements that compose it. Therefore, CD is widely used to assess the secondary structure content of a protein, and is particularly useful to analyze IDPs because of its sensitivity for reporting different conformational transitions involving changes in secondary structure. IDPs lack many attributes present in globular proteins, such as stable canonical secondary structure. However, these proteins display particular conformational properties such as an extended and dynamic conformation, and the presence of significant secondary structure content in certain regions, including a high content of polyproline type II structure [167,168].

Although CD is not useful for a detailed structural characterization, as it is not an atomic resolution technique, it can be extremely useful to monitor changes of secondary structure in response to multiple factors such as changing temperature, pH, ionic strength or presence of chaotropic agents [145].

Small angle X-ray scattering (SAXS). SAXS is a powerful method for the analysis and low-resolution (10 - 20 Å) structural characterization of biological macromolecules in solution [169]. Furthermore, SAXS can provide unique insights into biomolecular behaviours that can only be detected in solution, such as large conformational changes and transient protein-protein interactions allowing the quantitative analysis of flexible systems [169,170].

Under physiological conditions, IDPs constantly fluctuate between different structural states and undertake different interactions, resulting in a dynamic and heterogeneous mixture of conformations [170]. SAXS is particularly suited to study IDPs as it is one of the few methods that quantitatively characterize conformational polydispersity, often providing unique information on the structural properties of flexible macromolecules [170,171].

Experimentally, a SAXS experiment involves exposing a solution of particles and its corresponding solvent to a monochromatic X-ray beam where the scattering intensity of radiation is detected and recorded by an X-ray detector as a function of scattering angles (Figure 1.17).

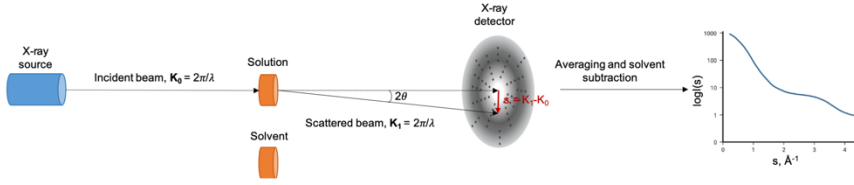


Figure 1.17. Experimental setup of a SAXS experiment. In a SAXS experiment, both solution and corresponding solvent are measured. After averaging, for both measurements, curves are subtracted to obtain a SAXS profile, where intensity of scattered waves $I(s)$ is plotted against the scattering vector s (Adapted from [170]).

The amplitude of the scattering vector s is given by equation 1.12:

$$s = \frac{4\pi \sin \theta}{\lambda} \quad (1.12)$$

where, θ is the angle enclosed between the incident and scattered beams, K_0 and K_1 , respectively, and λ is the wavelength of X-ray radiation.

The scattering intensity of the particles under study is obtained by subtracting the scattering intensity of the pure solvent from the scattering intensity of the sample solution. Intensity of scattered radiation is given by equation 1.13.

$$I(s) = \langle I(s) \rangle_{\Omega} \quad (1.13)$$

where $\langle I(s) \rangle_{\Omega}$ is the spherical average of the scattering intensity. In polydisperse systems such as IDPs, the scattering intensity is given by the sum of scattering arising from individual components, according to equation 1.14:

$$I(s) = \frac{1}{N} \sum_{n=1}^N I_n(s) \quad (1.14)$$

where, N represents the total number of conformers present in solution, and $I_n(s)$ is the scattering intensity given by the n -th conformer.

The scattering patterns are then plotted as radially averaged one-dimensional curves (Figure 1.18 A), from which specific parameters can be directly obtained, including molecular weight, maximum dimension (D_{max}), and the radius of gyration (R_g), which allow determining the size and general shape (envelope) of the protein [169,170].

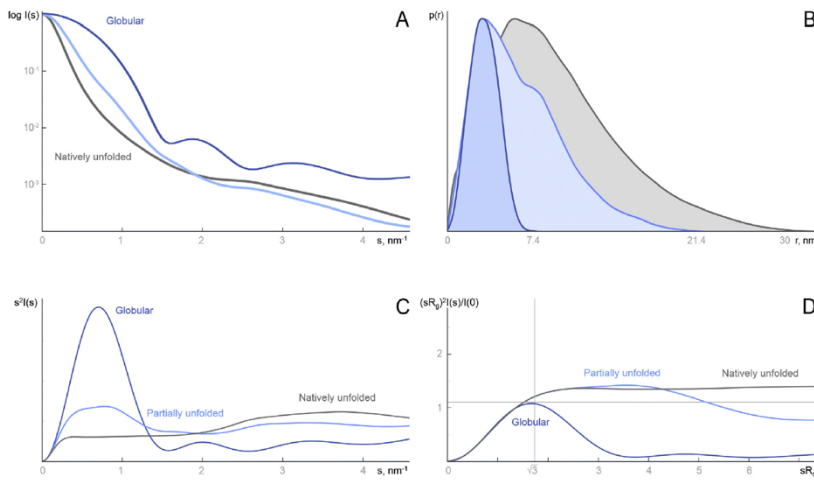


Figure 1.18. SAXS profiles for globular (dark blue), partially unfolded (light blue) and fully disordered (grey) proteins. (A) Logarithmic plot of the scattering intensity; **(B)** Distance distribution functions $p(r)$; **(C)** Kratky plot and **(D)** normalized (dimensionless) Kratky plot (Adapted from [172]).

R_g is the average root-mean-square distanced to the centre of density in the molecule weighted by the scattering length density and provides a measure of the overall size of the macromolecule [169]. It can be estimated using the Guinier approximation, which states that for very small angles ($s < 1/R_g$) the intensity depends only on two parameters according to the equation 1.15.

$$I(s) = I(0)\exp\left(-\frac{1}{3}R_g^2 s^2\right) \quad (1.15)$$

The particle sample can be described in real space by the distance distribution function $p(r)$, which corresponds to a histogram of distances between all possible pairs of atoms within a particle [169]. The scattering pattern of a particle $I(s)$ is a Fourier transform of its $p(r)$ function and vice versa (equation 1.16).

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty \frac{s^2 I(s) \sin(sr)}{sr} ds \quad (1.16)$$

In some cases, it is more intuitive to interpret the structural properties by analysing the $p(r)$ function rather than the scattering data itself since compact globular particles have a bell-shaped symmetric $p(r)$. In contrast, unfolded particles have an extended tail (Figure 1.18 B) [170]. The radius of gyration (R_g) can also be calculated from the $p(r)$ function (equation 1.17), often giving a more reliable estimate compared to the Guinier approximation, in particular for polydisperse systems, such as IDPs.

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 p(r) dr}{2 \int_0^{D_{max}} p(r) dr} \quad (1.17)$$

In the study of unstructured and polydisperse systems, R_g can also be estimated using Flory's equation (1.18) which correlates the length of the protein with the R_g .

$$R_g = R_0 N^\nu \quad (1.18)$$

N is the number of amino acids and for intrinsically disordered proteins R_0 is 2.54 ± 0.01 and ν is 0.522 ± 0.01 .

In addition to these characteristic parameters, the SAXS data also provides several indicators of the presence of protein flexibility. The Kratky plot (Figure 1.18 C) is often employed to qualitatively identify disordered states and distinguish them from globular proteins given that unstructured proteins, due to the presence of highly extended conformations, are characterized by large average sizes compared to globular proteins with a tightly packed core. The dimensionless Kratky plot (Figure 1.18 D), obtained by data normalization (assuming $I(0)=1$ and multiplying s by the R_g before plotting) is also valuable for comparing folding states of different proteins, as information about the shape is kept while size information is removed [170,173].

References

- [1] Akira, S., Uematsu, S., Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124** (4), 783-801(2006).
- [2] Meylan, E., Tschopp, J., Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442**, 39-44 (2006).
- [3] Akira, S., Takeda, K. Toll-like receptor signalling. *Nature Rev Immunol.* **4**, 499—511 (2004).
- [4] Beutler, B. Inferences, questions and possibilities in Toll-like receptor signalling. *Nature* **430**, 257–263 (2004).
- [5] Hajishengallis, G., Lambris, J.D. Microbial manipulation of receptor crosstalk in innate immunity. *Nat Rev Immunol* **11**, 187-200 (2011).
- [6] Lorenz, U. SHP-1 and SHP-2 in T cells: two phosphatases functioning at many levels. *Immunol Rev* **228**, 342-359 (2009).
- [7] You, R., Chu, C. SHP-1 (PTPN6) keeps the inflammation at bay: limiting IL-1 α -mediated neutrophilic dermatoses by preventing Syk kinase activation. *Cell Mol Immunol* **14**, 881-883 (2017).
- [8] Asrat, S., Dugan, A.S., Isberg, R.R. The frustrated host response to *Legionella pneumophila* is bypassed by MyD88-dependent translation of pro-inflammatory cytokines. *PLoS Pathog.* **10**, e1004229 (2014).
- [9] Lina, T.T., Alzahrani, S., Gonzalez, J., Pinchuk, I.V., Beswick, E.J., Reyes, V.E. Immune evasion strategies used by *Helicobacter pylori*. *World J Gastroenterol.* **20**, 12753-66 (2014).
- [10] Fortune, S.M., Solache, A., Jaeger, A., Hill, P.J., Belisle, J.T., Bloom, B.R., Rubin, E.J., Ernst, J.D. Mycobacterium tuberculosis inhibits macrophage responses to IFN-gamma through myeloid differentiation factor 88-dependent and -independent mechanisms. *J Immunol.* **172**, 6272-80 (2004).

General Introduction

- [11] Clements, A., Young, J.C., Constantinou, N., Frankel, G. Infection strategies of enteric pathogenic *Escherichia coli*. *Gut Microbes*. **3**, 71-87 (2012).
- [12] Ochoa, T.J., Contreras, C.A. Enteropathogenic *escherichia coli* infection in children. *Curr Opin Infect Dis*. **24**, 478-83 (2011).
- [13] GBD 2016 Diarrhoeal Disease Collaborators. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*. **18**, 1211-1228 (2018).
- [14] Abba, K., Sinfield, R., Hart, C.A., Garner, P. Pathogens associated with persistent diarrhea in children in low and middle-income countries: systematic review. *BMC Infect Dis* **10**, 9-88 (2009).
- [15] Croxen, M.A., Law, R.J., Scholz, R., Keeney, K.M., Wlodarska, M., Finlay, B.B. Recent Advances in Understanding Enteric Pathogenic *Escherichia coli*. *Clin Microbiol Rev* **26**, 822-880 (2013).
- [16] Levine, M.M., Edelman, R. Enteropathogenic *Escherichia coli* of classic serotypes associated with infant diarrhea: epidemiology and pathogenesis. *Epidemiol Rev* **6**, 31–51 (1984).
- [17] Kotloff, K.L., Nataro, J.P., Blackwelder, W.C., Nasrin, D., Farag, T.H., Panchalligam, S., Wu, Y., Sow, S.O., Sur, D., Brejman, R.F., Faruque, A.S., *et al*. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209-222 (2013).
- [18] Afset, J.E., Bevanger, L., Romundstad, P., Bergh, K. Association of atypical enteropathogenic *Escherichia coli* (EPEC) with prolonged diarrhoea. *J Med Microbiol* **53**, 1137-1144 (2004).
- [19] Croxen, M.A., Finlay, B.B. Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* **8**, 26-38 (2010).
- [20] Schmidt, M.A. LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* **12**, 1544-1552 (2010).
- [21] Nguyen, Y., Sperandio, N. Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front Cell Infect Microbiol* **2**, 90 (2012).
- [22] Spears, K.J., Roe, A.J., Gally, D.L. A comparison of enteropathogenic and enterohaemorrhagic *Escherichia coli* pathogenesis. *FEMS Microbiol Lett* **255**, 187-202 (2006).
- [23] Tarr, P.I., Gordon, C.A., Chandler, W.L. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. *Lancet* **365**, 1073-1086 (2005).
- [24] Spinale, J.M., Ruebner, R.L., Copelovitch, L., Kaplan, B.S. Long-term outcomes of Shiga toxin hemolytic uremic syndrome. *Pediatr Nephrol* **28**, 2097-2105 (2013).
- [25] Eppinger, M., Mammel, M.K., Leclerc, J.E., Ravel, J., Cebula, T.A. Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc Natl Acad Sci USA* **108**, 20142–20147 (2011).

- [26] Radosavljevic, V., Finke, E.J., Belojevic, G. *Escherichia coli* O104:H4 outbreak in Germany – clarification of the origin of the epidemic. *Eur J Public Health* **25**, 125-129 (2015).
- [27] Golan, L., Gonen, F., Yagel, S., Rosenshine, I., Shpigel, N.Y. Enterohemorrhagic *Escherichia coli* induce attaching and effacing lesions and hemorrhagic colitis in human and bovine intestinal xenograft models. *Dis Model Mech* **4**, 86-94 (2011).
- [28] Cepeda-Molero, M., Berger, C.N., Walsham, A.D.S., Ellis, S.J., Wemyss-Holden, S., Schüller, S., Frankel, G., Fernández, L.A. Attaching and effacing (A/E) lesion formation by enteropathogenic *E. coli* on human intestinal mucosa is dependent on non-LEE effectors. *PLoS Pathog* **13**, e1006706 (2017).
- [29] Nataro, J.P., Kaper, J. Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**, 142-201 (1998).
- [30] Gaytán, M.O., Martínez-Santos, V.I., Soto, E., González-Pedraio, B. Type Three Secretion System in Attaching and Effacing Pathogens. *Front Cell Infect Microbiol* **6**, 129 (2016).
- [31] Perna, N.T., Mayhew, G.F., Pósfai, G., Elliot, S., Sonnenberg, M.S., Kaper, J.B., Blattner, F.R. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **66**, 3810-3817 (1998).
- [32] Muller, D., Benz, I., Liebchen, A., Gallitz, I., Karch, H., Schmidt, M.A. Comparative analysis of the locus of enterocyte effacement and its flanking regions. *Infect Immun* **77**, 3501-3513 (2009).
- [33] Jarvis, K.G., Giron, J.A., Jerse, A.E., McDaniel, T.K., Sonnenberg, M.S., Kaper, J.B. Enteropathogenic *Escherichia coli* contains a putative type 3 secretion system necessary for the export of proteins involved in attaching and effacing lesion formation. *Proc Natl Acad Sci USA* **92**, 7996–8000 (1995).
- [34] Araujo-Garrido, J.L., Bernal-Bavars, J., Ramos-Morales, F. Type III Secretion Effectors with Arginine N-Glycosyltransferase Activity. *Microorganisms* **8**, 357-380 (2020).
- [35] Ramboarina, S., Fernandes, P.J., Daniell, S., Islam, S., Simpson, P., Frankel, G., Booy, F., Sonnenberg, M.S., Matthews, S. Structure of the bundle-forming pilus from enteropathogenic *Escherichia coli*. *J Biol Chem* **280**: 40252-40260 (2005).
- [36] Torres, A.G., Kanack, K.J., Tutt, C.B., Popov, V., Kaper, J.B. Characterization of the second long polar (LP) fimbriae of *Escherichia coli* O157:H7 and distribution of LP fimbriae in other pathogenic *E. coli* strains. *FEMS Microbiol Lett* **238**, 333–344 (2004).
- [37] Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldana, Z., Ledesma, M.A., Puente, J.L., Girón, J.A. The type 4 pili of enterohemorrhagic *Escherichia coli* O157:H7 are multipurpose structures with pathogenic attributes. *J Bacteriol* **191**, 411-421 (2009).

[38] Slater, S.L., S  gfors, A.M., Pollard, D.J., Ruano-Gallego, D., Frankel, G. The Type III Secretion System of Pathogenic *Escherichia coli*. *Curr Top Microbiol Immunol* **416**, 51-72 (2018).

[39] Mills, E., Baruch, K., Charpentier, X., Kobi, S., Rosenshine, I. Real-time analysis of effector translocation by the type III secretion system of enteropathogenic *Escherichia coli*. *Cell Host Microbe* **3**, 104-113 (2008).

[40] Touz  , T., Hayward, R.D., Eswaran, J., Leong, J.M., Koronakis, V. Self-association of EPEC intimin mediated by the beta-barrel-containing anchor domain: a role in clustering of the Tir receptor. *Mol Microbiol* **51**, 73-87 (2004).

[41] Alto, N.M., Shao, F., Lazar, C.S., Brost, R.L., Chua, G., Mattoo, S., McMahon, S.A., Ghosh, P., Hughes, T.R., Boone, C., Dixon, J.E. Identification of a bacterial type III effector family with G protein mimicry functions. *Cell* **124**, 133-145 (2006).

[42] Berger, C.N., Crepin, V.F., Jepson, M.A., Arbeloa, A., Frankel, G. The mechanisms used by enteropathogenic *Escherichia coli* to control filopodia dynamics. *Cell Microbiol* **11**, 309-322 (2009).

[43] Orchard, R.C., Kittisopikul, M., Altschuler, S.J., Wu, L.F., Suel, G.M., Alto, N.M. Identification of F-actin as the dynamic hub in a microbial-induced GTPase polarity circuit. *Cell* **148**, 803-815 (2012).

[44] Shaw, R.K., Cleary, J., Murphy, M.S., Frankel, G., Knutton, S. Interaction of Enteropathogenic *Escherichia coli* with Human Intestinal Mucosa: Role of Effector Proteins in Brush Border Remodeling and Formation of Attaching and Effacing Lesions. *Infect Immun* **73**, 1243-1251 (2005).

[45] Dean, P., Maresca, M., Schuller, S., Phillips, A.D., Kenny, B. Potent diarrheagenic mechanism mediated by the cooperative action of three enteropathogenic *Escherichia coli*-injected effector proteins. *Proc Natl Acad Sci USA* **103**, 1876-1881 (2006).

[46] Iizumi, Y., Sagara, H., Kabe, Y., Azuma, M., Kume, K., Ogawa, M., Nagai, T., Gillespie, P.G., Sasakawa, C., Handa, H. The enteropathogenic *E. coli* effector EspB facilitates microvillus effacing and antiphagocytosis by inhibiting myosin function. *Cell Host Microbe* **2**, 383-392 (2007).

[47] Dean, P., Young, L., Quidard, S., Kenny, B. Insights into the pathogenesis of enteropathogenic *E. coli* using an improved intestinal enterocyte model. *PLoS One* **8**, e55284 (2013).

[48] Sch  ller, S., Lucas, M., Kaper, J.B., Gir  n, J.A., Phillips, A.D. The ex vivo response of human intestinal mucosa to enteropathogenic *Escherichia coli* infection. *Cell Microbiol* **11**, 521-530 (2009).

[49] Ruchaud-Sparagano, M.H., Maresca, M., Kenny, B. Enteropathogenic *Escherichia coli* (EPEC) inactivate innate immune responses prior to compromising epithelial barrier function. *Cell Microbiol* **9**, 1909-1921 (2007).

[50] Goosney, D.L., Celli, J., Kenny, B., Finlay, B.B. Enteropathogenic *Escherichia coli* inhibits phagocytosis. *Infect Immun.* **67**, 490-5 (1999).

- [51] Marchès, O., Covarelli, V., Dahan, S., Cougoule, C., Bhatta, P., Frankel, G., Caron, E. EspJ of enteropathogenic and enterohaemorrhagic *Escherichia coli* inhibits opsono-phagocytosis. *Cell Microbiol.* **10**, 1104-15 (2008).
- [52] Nougayrede, J.P., Sonnenberg, M.S. Enteropathogenic *Escherichia coli* EspF is targeted to mitochondria and is required to initiate the mitochondrial death pathway. *Cell Microbiol* **6**, 1097-1111 (2004).
- [53] Shames, S.R., Deng, W., Guttman, J.A., de Hoog, C.L., Li, Y., Hardwidge, P.R., Sham, H.P., Vallance, B.A., Foster, L.J., Finlay, B.B. The pathogenic *E. coli* type III effector EspZ interacts with host CD98 and facilitates host cell prosurvival signalling. *Cell Microbiol* **12**, 1322-1339 (2010).
- [54] Hemrajani, C., Berger, C.N., Robinson, K.S., Marchès, O., Mousnier, A., Frankel, G. NleH effectors interact with Bax inhibitor-1 to block apoptosis during enteropathogenic *Escherichia coli* infection. *Proc Natl Acad Sci USA* **107**, 3129-3134 (2010).
- [55] Baruch, K., Gur-Arie, L., Nadler, C., Koby, S., Yerushalmi, G., Ben-Neriah, Y., Yogev, O., Shaulian, E., Guttman, C., Zarivach, R., Rosenshine, I. Metalloprotease type III effectors that specifically cleave JNK and NF- κ B. *EMBO J* **30**, 221-231 (2011).
- [56] Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7**, 539 (2011).
- [57] Munera, D., Martinez, E., Varyukhina, S., Mahajan, A., Ayala-Sanmartin, J., Frankel, G. Recruitment and membrane interactions of host cell proteins during attachment of enteropathogenic and enterohaemorrhagic *Escherichia coli*. *Biochem J* **445**, 383-392 (2012).
- [58] Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., Finlay, B.B., Strynadka, N.C. Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex. *Nature* **405**, 1073-1077 (2000).
- [59] Ross, N., Miller, B. Characterization of the binding surface of the translocated intimin receptor, an essential protein for EPEC and EHEC cell adhesion. *Protein Sci* **16**, 2677-2683 (2007).
- [60] Goosney, D.L., DeVinney, R., Pfuetzner, R.A., Frey, E.A., Strynadka, N.C., Finlay, B.B. Enteropathogenic *E. coli* translocated intimin receptor, Tir, interacts directly with alpha-actinin. *Curr Biol* **10**, 735-738 (2000).
- [61] Kenny, B. The enterohaemorrhagic *Escherichia coli* (serotype O157:H7) Tir molecule is not functionally interchangeable for its enteropathogenic *E. coli* (serotype O127:H6) homologue. *Cell Microbiol* **3**, 499-510 (2001).
- [62] DeVinney, R., Puente, J.L., Gauthier, A., Goosney, D., Finlay, B.B. Enterohaemorrhagic and enteropathogenic *Escherichia coli* use a different Tir-based mechanism for pedestal formation. *Mol Microbiol* **41**, 1445-1458 (2001).
- [63] Rosenshine, I., Sonnenberg, M., Kaper, J.B., Finlay, B.B. Signal transduction between enteropathogenic *Escherichia coli* (EPEC) and epithelial cells:

General Introduction

EPEC induces tyrosine phosphorylation of host cell proteins to initiate cytoskeletal rearrangement and bacterial uptake. *EMBO J* **11**, 3551-3560 (1992).

[64] Warawa, J., Kenny, B. Phosphoserine modification of the enteropathogenic *Escherichia coli* Tir molecule is required to trigger conformational changes in Tir and efficient pedestal elongation. *Mol Microbiol* **42**, 1269-1280 (2001).

[65] DeVinney, R., Stein, M., Reinscheid, D., Abe, A., Ruschkowski, S., Finlay, B.B. Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. *Infect Immun* **67**, 2389-2398 (1999).

[66] Kenny, B. Phosphorylation of tyrosine 474 of the enteropathogenic *Escherichia coli* (EPEC) Tir receptor molecule is essential for actin nucleating activity and is preceded by additional host modifications. *Mol Microbiol* **31**, 1229-1241 (1999).

[67] Lai, Y., Rosenshine, I., Leong, J.M., Frankel, G. Intimate host attachment: enteropathogenic and enterohaemorrhagic *Escherichia coli*. *Cell Microbiol* **15**, 1796-1808 (2013).

[68] Gruenheid, S., DeVinney, R., Bladt, F., Goosney, D., Gelkop, S., Gish, G.D., Pawson, T., Finlay, B.B. Enteropathogenic *E. coli* Tir binds Nck to initiate actin pedestal formation in host cells. *Nat Cell Biol* **3**, 856-859 (2001).

[69] Campellone, K.G., Giese, A., Tipper, D.J., Leong, J.M. A tyrosine-phosphorylated 12-amino-acid sequence of enteropathogenic *Escherichia coli* Tir binds the host adaptor protein Nck and is required for Nck localization to actin pedestals. *Mol Microbiol* **43**, 1227-1241 (2002).

[70] Kalman, D., Weiner, O.D., Goosney, D.L., Sedat, J.W., Finlay, B.B., Abo, A., Bishop, J.M. Enteropathogenic *E. coli* acts through WASP and Arp2/3 complex to form actin pedestals. *Nat Cell Biol* **1**, 389-391 (1999).

[71] Campellone, K.G., Leong, J.M. Nck-independent actin assembly is mediated by two phosphorylated tyrosines within enteropathogenic *Escherichia coli* Tir. *Mol Microbiol* **56**, 416-432 (2005).

[72] Brady, M.J., Campellone, K.G., Ghildiyal, M., Leong, J.M. Enterohaemorrhagic and enteropathogenic *Escherichia coli* Tir proteins trigger a common Nck-independent actin assembly pathway. *Cell Microbiol* **9**, 2242-2253 (2007).

[73] Campellone, K.G., Robbins, D., Leong, J.M. Esp_{FU} is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Dev Cell* **7**, 217-228 (2004).

[74] Garmendia, J., Phillips, A.D., Carlier, M.-F., Chong, Y., Schüller, S., Marches, O., Dahan, S., Oswald, E., Shaw, R.K., Knutton, S., Frankel, G. TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol* **6**, 1167-1183 (2004).

[75] Garmendia, J., Carlier, M.F., Egile, C., Didry, D., Frankel, G. Characterization of TccP-mediated N-WASP activation during enterohaemorrhagic *Escherichia coli* infection. *Cell Microbiol* **8**, 1444-1455 (2006).

- [76] Vingadassalom, D., Campellone, K.G., Brady, M.J., Skehan, B., Battle, S.E., Robbins, D., Kapoor, A., Hecht, G., Snapper, S.B., Leong, J.M. Enterohemorrhagic *E. coli* requires N-WASP for efficient type III translocation but not for EspFU-mediated actin pedestal formation. *PLoS Pathog* **6**, e1001056 (2010).
- [77] Aitio, O., Hellman, M., Kazlauskas, A., Vingadassalom, D.F., Leong, J.M., Saksela, K. Permi, P. Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly. *Proc Natl Acad Sci USA* **107**, 21743-21748 (2010).
- [78] Cheng, H.C., Skehan, B.M., Campellone, K.G., Leong, J.M., Rosen, M.K. Structural mechanism of WASP activation by the enterohaemorrhagic *E. coli* effector EspF(U). *Nature* **454**, 1009-1013 (2008).
- [79] Sason, H., Milgrom, M., Weiss, A.M., Melamed-Book, N., Balla, T., Grinstein, S., Backert, S., Rosenshine, I., Aroeti, B. Enteropathogenic *Escherichia coli* Subverts Phosphatidylinositol 4,5-Bisphosphate and Phosphatidylinositol 3,4,5-Trisphosphate upon Epithelial Cell Infection. *Mol Biol Cell* **20**, 544-555 (2009).
- [80] Campellone, K.G. Phosphoisotides influence pathogen surfing: EPEC rights the SHIP. *Cell Host Microbe* **7**, 1-2 (2010).
- [81] Barrow, A.D., Trowsdale, J. You say ITAM and I say ITIM, let's call the whole thing off: the ambiguity of immunoreceptor signalling. *Eur J Immunol* **36**, 1646-1653 (2006).
- [82] Ooms, L.M., Horan, K.A., Rahman, P., Seaton, G., Gurung, R., Kethesparan, D.S., Mitchell, C.A. The role of the inositol polyphosphate 5-phosphatases in cellular function and human disease. *Biochem J* **419**, 29-49 (2009).
- [83] Yan, D., Wang, X., Luo, L., Cao, X. and Ge, B. Inhibition of TLR signaling by a bacterial protein containing immunoreceptor tyrosine-based inhibitory motifs. *Nat Immunol* **13**, 1063-1071 (2012).
- [84] Yan, D., Quan, H., Wang, L., Liu, F., Liu, H., Chen, J., Cao, X., Ge, B. Enteropathogenic *Escherichia coli* Tir recruits cellular SHP-2 through ITIM motifs to suppress host immune response. *Cell Signal* **25**, 1887-1894 (2013).
- [85] Brandt, S., Kenny, B., Rohde, M., Martinez-Quiles, N., Backert, S. Dual infection system identifies a crucial role for PKA-mediated serine phosphorylation of the EPEC-Tir-injected effector protein in regulating Rac1 function. *Cell Microbiol* **11**, 1254-1271 (2009).
- [86] Backert, S., Kenny, B., Gerhard, R., Tegtmeyer, N., Brandt, S. PKA-mediated phosphorylation of EPEC-Tir at serine residues 434 and 463. A novel pathway in regulating Rac1 GTPase function. *Gut Microbes* **1**, 94-99 (2010).
- [87] Erickson, J. W., Cerione, R. A. Multiple roles for Cdc42 in cell regulation. *Curr Opin Cell Biol* **13**, 153-157 (2001).
- [88] Kenny, B., Ellis, S., Leard, A.D., Warawa, J., Mellor, H., Jepson, M.A. Co-ordinate regulation of distinct host cell signalling pathways by multifunctional enteropathogenic *Escherichia coli* effector molecules. *Mol Microbiol* **44**, 1095-1107 (2002).

- [89] Goddard, P.J., Sanchez-Garrido, J., Slater, S.L., Kalyan, M., Ruano-Gallego, D., Marchès, O., Fernández, L.Á., Frankel, G., Shenoy, A.R. Enteropathogenic *Escherichia coli* Stimulates Effector-Driven Rapid Caspase-4 Activation in Human Macrophages. *Cell Rep.* **27**, 1008-1017 (2019).
- [90] He, W.T., Wan, H., Hu, L., Chen, P., Wang, X., Huang, Z., Yang, Z.H., Zhong, C.Q., Han, J. Gasdermin D is an executor of pyroptosis and required for interleukin-1 β secretion. *Cell Res.* **25**, 1285-98 (2015).
- [91] Zhong, Q., Roumeliotis, T.I., Kozik, Z., Cepeda-Molero, M., Fernández, L.Á., Shenoy, A.R., Bakal, C., Frankel, G., Choudhary, J.S. Clustering of Tir during enteropathogenic *E. coli* infection triggers calcium influx-dependent pyroptosis in intestinal epithelial cells. *PLoS Biol.* **18**, e3000986 (2020).
- [92] Blasche, S., Arens, S., Ceol, A., Siszler, G., Schmidt, M.A., Häuser, R., Schwarz, F., Wuchty, S., Aloy, P., Uetz, P., Stradal, T., Koegl, M. The EHEC-host interactome reveals novel targets for the translocated intimin receptor. *Sci Rep.* **4**, 7531 (2014).
- [93] Lemieux, R.U., Spohr, U. How Emil Fischer was led to the lock and key concept for enzyme specificity. *Adv Carbohydr Chem Biochem.* **50**, 1-20 (1994).
- [94] Fleming, P.J., Rose, G.D. Conformational Properties of Unfolded Proteins. *Protein Folding Handbook*. Part I. 710-36 (2008).
- [95] Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., Villafranca, J.E. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput.* 473-84 (1998).
- [96] Dunker, A.K., Silman, I., Uversky, V.N., Sussman, J.L. Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology* **18**, 756–764 (2008).
- [97] Uversky, V.N. The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol* **2010**, 568068 (2010).
- [98] Uversky, V.N., Gillespie, J.R., Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins.* **41**, 415 – 27 (2000).
- [99] Cornish, J., Chamberlain, S.G., Owen, D., Mott, H.R. Intrinsically disordered proteins and membranes: a marriage of convenience for cell signalling? *Biochem Soc Trans.* **48**, 2669-2689 (2020).
- [100] Bürgi, J., Xue, B., Uversky, V.N., van der Goot, F.G. Intrinsic Disorder in Transmembrane Proteins: Roles in Signaling and Topology Prediction. *PLoS One.* **11**, e0158594 (2016).
- [101] Kjaergaard, M., Kragelund, B.B. Functions of intrinsic disorder in transmembrane proteins. *Cell Mol Life Sci.* **74**, 3205-3224 (2017).
- [102] Deryusheva, E., Nemashkalova, E., Galloux, M., Richard, C.A., Eléouët, J.F., Kovacs, D., Van Belle, K., Tompa, P., Uversky, V., Permyakov, S. Does Intrinsic Disorder in Proteins Favor Their Interaction with Lipids? *Proteomics.* **19**, e1800098 (2019).

- [103] Tompa, P., Szász, C., Buday, L. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* **30**, 484–489 (2005).
- [104] Van Roey, K., Uyar, B., Weatheritt, R.J., Dinkel, H., Seiler, M., Budd, A., Gibson, T.J., Davey, N.E. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* **114**, 6733–78 (2014).
- [105] Tompa, P., Davey, N.E., Gibson, T.J., Babu, M.M. A million peptide motifs for the molecular biologist. *Mol Cell.* **55**, 161–9 (2014).
- [106] Kotta-Loizou, I., Tsaousis, G.N., Hamodrakas, S.J. Analysis of Molecular Recognition Features (MoRFs) in membrane proteins. *Biochim. Biophys. Acta - Proteins Proteomics*, **1834**, 798–807 (2013).
- [107] Delaforge, E., Cordeiro, T. N., Bernadó, P. and Sibille, N. Conformational characterization of intrinsically disordered proteins and its biological significance. *Modern Magnetic Resonance* 381–399 (Springer International Publishing, 2018).
- [108] Tompa, P., Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2–8 (2008).
- [109] Dunker, A.K., Brown, C.J., Obradović, Z. Identification and functions of usefully disordered proteins. *Advances in Protein Chemistry* **62**, 25–49 (2002).
- [110] Iakoucheva, L.M., Radivojac, P., Brown, C.J., *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **32**, 1037–1049 (2004).
- [111] Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradović, Z., Uversky, V.N., Dunker, A.K., Intrinsic disorder and functional proteomics. *Biophysical Journal* **92**, 1439–1456 (2007).
- [112] Bah, A., Forman-Kay, J.D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J Biol Chem* **29**, 6696–6705 (2016).
- [113] Huang, Y., Liu, Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J Mol Biol.* **393**, 1143–59 (2009).
- [114] Tompa, P., The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters* **579**, 3346–3354 (2005).
- [115] Uversky, V.N., Oldfield, C.J., Dunker, A.K., Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *Journal of Molecular Recognition* **18**, 343–384 (2005).
- [116] Cortese, M.S., Uversky, V.N., Dunker, A.K. Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol.* **98**, 85–106 (2008).
- [117] Wright, P.E., Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
- [118] Uversky, V.N., Oldfield, C.J., Dunker, A.K., Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annual Review of Biophysics* **37**, 215–246 (2008).
- [119] Martinelli, A.H.S., Lopes, F.C., John, E.B.O., Carlini, C.R. and Ligabue-Braun, R. Modulation of Disordered Proteins with a Focus on Neurodegenerative Diseases and Other Pathologies. *Int J Mol Sci* **20**, 1322 (2019).

[120] Santofimia-Castaño, P., Rizzuti, B., Xia, Y. *et al.* Targeting intrinsically disordered proteins involved in cancer. *Cell Mol Life Sci* **77**, 1695–1707 (2020).

[121] Xue, B., Dunker, A.K., Uversky, V.N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*. **30**, 137-49 (2012).

[122] Basile, W., Salvatore, M., Bassot, C., Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput Biol*. **15**, e1007186 (2019).

[123] van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., Kim, P.M., Kriwacki, R.W., Oldfield, C.J., Pappu, R.V., Tompa, P., Uversky, V.N., Wright, P.E., Babu, M.M. Classification of intrinsically disordered regions and proteins. *Chem Rev*. **114**, 6589-631 (2014).

[124] Cohan, M.C., Pappu, R.V. Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria. *Trends Biochem Sci*. **45**, 668-680 (2020).

[125] Shin, Y., Brangwynne, C.P. Liquid phase condensation in cell physiology and disease. *Science*. **357**, eaaf4382 (2017).

[126] Holehouse, A.S., Pappu, R.V. Functional Implications of Intracellular Phase Transitions. *Biochemistry*. **57**, 2415-2423 (2018).

[127] Cohan, M.C., Eddelbuettel, A.M.P., Levin, P.A., Pappu, R.V. Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of *Bacillus subtilis* FtsZ. *J Mol Biol*. **432**, 3205-3221 (2020).

[128] Shinn, M.K., Kozlov, A.G., Nguyen, B., Bujalowski, W.M., Lohman, T.M. Are the intrinsically disordered linkers involved in SSB binding to accessory proteins? *Nucleic Acids Res*. **47**, 8581-8594 (2019).

[129] Ait-Bara, S., Carpousis, A.J., Quentin, Y. RNase E in the γ -Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains. *Mol Genet Genomics*. **290**, 847-62 (2015).

[130] Holmes, J.A., Follett, S.E., Wang, H., Meadows, C.P., Varga, K., Bowman, G.R. *Caulobacter* PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc Natl Acad Sci U S A*. **113**, 12490-12495 (2016).

[131] Marin, M., Uversky, V.N., Ott, T. Intrinsic disorder in pathogen effectors: protein flexibility as an evolutionary hallmark in a molecular arms race. *Plant Cell*. **25**, 3153-7 (2013).

[132] Rodgers, L., Gamez, A., Riek, R., Ghosh, P. The Type III Secretion Chaperone SycE Promotes a Localized Disorder-to-Order Transition in the Natively Unfolded Effector YopE. *J. Biol. Chem*. **283**, 20857–20863 (2008).

[133] Dyson, H.J., Wright, P.E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol*. **6**, 197–208 (2005).

[134] Nishikawa, H., Hayashi, T., Arisaka, F., Senda, T., Hatakeyama, M. Impact of structural polymorphism for the *Helicobacter pylori* CagA oncoprotein on binding to polarity-regulating kinase PAR1b. *Sci Rep*. **6**, 30031 (2016).

- [135] Hatakeyama, M. Structure and Function of Helicobacter Pylori Caga, the First-Identified Bacterial Protein Involved in Human Cancer. *Proc. Jpn. Acad. Ser. B, Phys. Biol. Sci.* **93**, 196–219 (2017).
- [136] Aitio, O., Hellman, M., Skehan, B., Kesti, T., Leong, J.M., Saksela, K., Permi, P. Enterohaemorrhagic *Escherichia coli* exploits a tryptophan switch to hijack host f-actin assembly. *Structure*. **20**, 1692-703 (2012).
- [137] Brutscher, B., Felli, I.C., Gil-Caballer, S., Hošek, T., Kümmerle, R., Piai, A., Pierattelli, R., Sólyó, Z. NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv Exp Med Biol* **870**, 49-122 (2015).
- [138] Gibbs, E.B., Cook, E.C., Showalter, S.A. Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys* **628**, 57-70 (2017).
- [139] Hijaz, B.A., Volpicelli-Daley, L.A. Initiation and propagation of α -synuclein aggregation in the nervous system. *Mol Neurodegener.* **15**, 19 (2020).
- [140] Kosol, S., Contreras-Martos, S., Cedeño, C., Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **18**, 10802-10828 (2013).
- [141] Gołowicz, D., Kasprzak, P., Orekhov, V., Kazimierczuk, K. Fast time-resolved NMR with non-uniform sampling. *Prog Nucl Magn Reson Spectrosc* **116**, 40-55 (2020).
- [142] Zawadzka-Kazimierczuk, A., Kozminski, W., Sanderova, H., Krasny, L. High dimensional and high-resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. *J Biomol NMR* **52**, 329–337 (2012).
- [143] Felli, I.C., Pierattelli, R. Novel methods based on ^{13}C detection to study intrinsically disordered proteins. *J Magn Reson.* **241**, 115-125 (2014).
- [144] Payliss, B.J., Vogel, J., Mittermaier, A.K. Side chain electrostatic interactions and pH-dependent expansion of the intrinsically disordered, highly acidic carboxyl-terminus of γ -tubulin. *Protein Sci* **28**, 1095-1105 (2019).
- [145] Kjaergaard, M., Nørholm, A-B., Hendus-Altenburger, R., Pedersen, S.F., Poulsen, F.M. and Kragelund, B.B. Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α -helices or loss of polyproline II? *Protein Sci* **19**, 1555–1564 (2010).
- [146] Camilloni, C., De Simone, A., Vranken, W.F., Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry.* **51**, 2224-31 (2012).
- [147] De Simone, A., Cavalli, A., Hsu, S.T., Vranken, W., Vendruscolo, M. Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *JACS* **131**, 16332–16333 (2009).
- [148] Tamiola, K., Acar, B., Mulder, F.A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *JACS* **132**, 18000–18003 (2010).

[149] Shen, Y., Delaglio, F., Cornilescu, G., Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. **44**, 213-23 (2009).

[150] Wishart, D.S., Sykes, B.D. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol. NMR* **4**, 171–180 (1994).

[151] Gillespie, J.R., Shortle, D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* **268**, 158–169 (1997).

[152] Clore, G.M., Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108-39 (2009).

[153] Clore, G.M. Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules. *Methods Enzymol* **564**, 485-97 (2015).

[154] Iwahara, J., Tang, C., Marius Clore, G. Practical aspects of ^1H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184**, 185–195 (2007).

[155] Ganguly, D., Chen, J. Structural Interpretation of Paramagnetic Relaxation Enhancement-Derived Distances for Disordered Protein States. *J. Mol. Biol.* **390**, 467–477 (2009).

[156] Battiste, J.L., Wagner, G. Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data. *Biochemistry* **39**, 5355–5365 (2000).

[157] Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., et al., Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519**, 106-109 (2015).

[158] Du, J.T., Li, Y.M., Wei, W., et al. Low-barrier hydrogen bond between phosphate and the amide group in phosphopeptide. *J Am Chem Soc* **127**, 16350-16351 (2005).

[159] Theillet, F.X., Smet-Nocca, C., Liokatis, S., Thongwichian, R., Kosten, J., Yoon, M.K., Kriwacki, R.W., Landrieu, I., Lippens, G., Selenko, P. Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR* **54**, 217-36 (2012).

[160] Theillet, F.X., Rose, H.M., Liokatis, S., Binolfi, A., Thongwichian, R., Stuiver, M., Selenko, P. Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nat Protoc.* **8**, 1416-32 (2013).

[161] Selenko, P., Frueh, D.P., Elsaesser, S.J., Haas, W., Gygi, S.P., Wagner, G. In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat Struct Mol Biol* **15**, 321-9 (2008).

- [162] Teilum, K., Kunze, M.B.A., Erlendsson, S., Kragelund, B.B. (S)Pinning down protein interactions by NMR. *Protein Sci.* **26**, 436–451 (2017).
- [163] Purslow, J.A., Khatiwada, B., Bayro, M.J., Venditti, V. NMR Methods for Structural Characterization of Protein-Protein Complexes. *Front Mol Biosci* **7** (2020).
- [164] Mollica, L., Bessa, L.M., Hanouille, X., Jensen, M.R., Blackledge, M., Schneider, R. Binding mechanisms of intrinsically disordered proteins: Theory, simulation, and experiment. *Front Mol Biosci.* **3**, 1-18 (2016).
- [165] Driscoll, P.C. Macromolecular Complexes. *Protein NMR Spectrosc Pract Tech Appl.* 269-317 (2011).
- [166] Receveur-Bréchet, V., Bourhis, J.M., Uversky, V.N., Canard, B., Longhi, S. Assessing protein disorder and induced folding. *Proteins.* **62**, 24-45 (2006).
- [167] Chemes, L.B., Alonso, L.G., Noval, M.G., de Prat-Gay G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol Biol* **895**, 387-404 (2012).
- [168] Contreras-Martos S, Nguyen HH, Nguyen PN, Hristozova N, Macossay-Castillo M, Kovacs D, Bekesi A, Oemig JS, Maes D, Pauwels K, Tompa P, Lebrun P. Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Front Mol Biosci* **5**, 83 (2018).
- [169] Mertens, H.D., Svergun, D.I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128-41 (2010).
- [170] Bernardó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656-64 (2007).
- [171] Cordeiro, T.N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N. & Bernardó, P. Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr. Opin. Struct. Biol.* **42**, 15-23 (2017).
- [172] Svergun, D.I., Koch, M.H. Advances in structure analysis using small-angle scattering in solution. *Curr. Opin. Struct. Biol.* **12**, 654-60 (2002).
- [173] Kikhney, A.G., Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570-7 (2015).

Objectives

Tir is recognized as an essential effector for A/E pathogens virulence. Nonetheless, the molecular mechanisms by which Tir interacts with multiple host proteins to rewire intracellular signaling and assist infection remain largely unexplored.

Thus, the main goal of this work is to elucidate the fundamental principles for Tir binding multiteity in signaling regulation, by targeting the following specific objectives:

- **Provide a structural biophysical description of the intracellular C-terminal domain of Tir (C-Tir).** On this topic, a combination of techniques including NMR and SAXS will be employed to deliver the structural and dynamical features underpinning the hostile-mode of action of this bacterial "weapon". To further explore the structural propensity of C-Tir, computational modeling will be used to integrate the experimental data and determine a conformational ensemble of C-Tir.
- **Understand and explore the impact of multi-site phosphorylation in signaling.** To do so, we will rely on NMR to quantitatively monitor phosphorylation, study the reaction kinetics and extension, and explore the local structural implications in C-Tir.
- **Characterize host proteins- and membrane-binding by C-Tir.** NMR will be exquisitely used to probe at atomic level Tir interacting assembly with multiple host targets to tackle this goal. We will use recombinantly produced host protein domains and membrane mimics to explore the details of Tir-mediated biomolecular associations. Validation of the role of such interactions in a cellular context will be delegated to expert collaborators using Immunofluorescence and cell-based assays to investigate the effect of Tir binding in its biological functions.

Chapter 2

Materials and Methods

This Chapter describes the general protocols used under the scope of this thesis, and it is divided into three main sections. The first section describes the protocols employed to produce the proteins studied. The second section describes the specific experimental protocols used in studying each protein or set of proteins by NMR spectroscopy. The third section includes the protocols employed to investigate each protein or set of proteins using other techniques such as CD spectroscopy, SAXS, and computational modeling.

Protein samples production

This section highlights the construction of the expression vectors, heterologous expression, and purification processes of the proteins studied during this work.

C-Terminal region of Tir

Construct design and cloning. Tir is a 56 KDa protein comprised of two predicted transmembrane helices (TM-Tir) connected by a flexible extracellular loop (IBD) that binds intimin, and flanked by intracellular N- and C-terminal disordered tails (N-Tir and C-Tir, respectively). Several constructs were designed based on this molecular architecture: Tir, encoding the residues 1 to 550; N-Tir composed by residues 1 to 233; TM-Tir encoding the residues 229 to 394 and C-terminal Tir comprising the residues 388 to 550. As this thesis then focused on the study of C-Tir disordered tail, only the protocols employed to obtain this construction will be presented.

The expression vector containing the gene sequence for Tir protein was synthesized by *ATG: biosynthetics*, and then used as the template for cloning the gene encoding for the C-Tir sequence. The DNA sequence for C-Tir was cloned in the pHTP8 expression vector (NZYTech) using the Ligase Independent Cloning (LIC) technique. The gene was amplified by Polymerase Chain Reaction (PCR) with primers containing both non-

priming overhang and gene-specific sequences, a recognition sequence for HRV-3C protease, to remove the N-terminal fusion tag, and a Strep-tag sequence on the C-terminus, which allow subsequent recombinant protein purification by affinity chromatography (Table 2.1) (Supplementary Figure SI.1). Fragment size (522 bp) was confirmed by 1.0% (w/v) agarose gel electrophoresis.

Table 2.1. Oligonucleotide sequences (primers) used for amplification of C-Tir. Underlined nucleotides correspond to the HRV-3C protease specific sequence. The DNA sequence coding for the Strep-tag sequence is highlighted in bold. fw and rv indicate forward and reverse primers, respectively.

Primer	Sequence (5'→ 3')
C-Tir_fw	TCAGCAAGGGCTGAGGCTGGAAGTTCTGTTCCAGGGGCC CGTCGCAATCAGCCGG
C-Tir_rv	TCAGCGGAAGCTGAGGCTACTTCTCGAACTGCGGATGGCT CCACACGAAACGCACCGGG

Purified PCR amplified fragments were mixed with the vector by preparing a reaction mixture containing 5 µL purified DNA fragment, 1 µL linearized pHTP8 vector, 1 µL 10X Reaction Buffer, 0.5 µL NZYEasy enzyme mix and nuclease-free water to a final volume of 10 µL. The cloning reaction was performed in T100™ thermocycler (Bio-Rad) including an initial step at 37 °C for 1 hour, followed by 10 minutes at 80 °C, then 10 minutes at 30 °C and a final step at 4°C. DH5α *Escherichia coli* competent cells were then transformed with the ligation products from the cloning reaction. Briefly, the mixtures were incubated on ice for 30 minutes, followed by a heat shock at 42 °C for 40 seconds. Then, after 2 minutes on ice, 900 µL of pre-warmed SOC broth (20 g/L tryptone, 5 g/L yeast extract, 3.6 g/L glucose, 0.186 g/L KCl, 0.5 g/L NaCl, 0.96 g/L MgCl₂) was added to the mixture. The cells were grown in an incubator shaker (AccuTherm, Labnet), at 37°C for 60 min at 200 rpm. The cell suspension was centrifuged at 5000 rpm for 1 minute, from which 900 µL of supernatant were removed.

Cells were gently re-suspended and plated for selection onto LB agar supplemented with 50 µg/mL kanamycin.

The resulting colonies were screened by colony PCR with pHTP primers (*pHTP_fw*: 5'- GTTCAAAAACGGTGAAGTGGCGGC -3' and *pHTP_rv*: 5'- GGTTATGCTAGTTATTGCTCAG CG -3'). The colonies with a PCR product of the correct size were grown aerobically overnight, at 37 °C and 180 rpm, in LB broth (5 g/L yeast extract, 10 g/L tryptone, and 10 g/L NaCl) supplemented with 50 µg/mL kanamycin for plasmid extraction using the NZYMiniprep Kit (NZYTech) and DNA sequencing (GATC Biotech).

Site-directed mutagenesis. All C-Tir mutants designed for this work were prepared using the NZYMutagenesis kit (NZYTech), following the manufacturer's instructions. For the PRE NMR experiments, C-Tir mutants V405C, S428C, S463C, S509C, and S536C were prepared for further conjugation with the thiol-reactive methanethiosulfonate spin-label (MTSL). For protein interaction studies, C-Tir double mutant Y454F_Y474F was prepared by inserting one mutation at a time. The oligonucleotide sequences used for amplification are listed in Table 2.2.

Table 2.2. Oligonucleotide sequences (primers) used for site-directed mutagenesis of C-Tir. Underlined nucleotides correspond to the sequence encoding the desired mutation. fw and rv indicate forward and reverse primers, respectively.

C-Tir mutant	Primer	Sequence (5'→3')
V405C	V405C_fw	ACAACCACCCACACTGTGT <u>GCC</u> CAGCAGCAAACCG GCGGC
	V405_rv	GCCGCCGGTTTGCTGCT <u>GGC</u> CACACAGTGTGGGTG GTTGT
S428C	S428C_fw	CGCGTGCAGAAGACGCGT <u>GCT</u> TGAACCGTCGTGA CTCACAAGGC
	S428C_rv	GCCTTGTGAGTCACGACGGTTCAAG <u>CAC</u> CGCGTCT TCTGCACGCG
Y454F	Y454F_fw	GAAGTAGTAAATCCCTTTGCGGAAGTAGGGGGTG CG
	Y454F_rv	CGCACCCCCTACTTCCGCAAAAGGGATTTACTACTT C
S463C	S463C_fw	GAAGTAGGGGGTGCGCGTAATT <u>GCT</u> CTGAGCGCAC ATCAGCCCCG
	S463_rv	CGGGCTGATGTGCGCTCAGGCAATTACGCGCACC CCCTACTTC
Y474F	Y474F_fw	CAGCCCGAAGAACACATCTT <u>CG</u> ATGAAGTAGCGG CC
	Y474F_rv	GGCCGCTACTTCATC <u>GAA</u> GATGTGTTCTTCGGGC TG
S509C	S509C_fw	CCAGGCCAGGGCATTCAGT <u>GCA</u> CTACGCCCTGC TCGCG
	S509C_rv	CGCGACCAGGGCGTAGGT <u>GCA</u> CTGAATGCCCTG GCCTGG
S536C	S536C_fw	CGGCGGTGAGAGCGCGGTTT <u>GCT</u> CTGTCAACGC GGCGC
	S536C_rv	GCGCCGCGTTGACAGAGCAAACCGCGCTCTCACC GCCG

Oligonucleotide sequences containing the desired mutation for each mutant (Table 2.2) were synthesized, and the pHTP8 expression vector harboring the *c-tir* gene was used as template for amplification by Polymerase Chain Reaction (PCR). Upon amplification, 5 μ L of *Dpn I* were added to the reactions and incubated at 37 °C for 1 hour to digest the parental methylated and hemi-methylated DNA. After that, 100 μ L of chemically competent DH5 α *E. coli* cells were transformed with 10 μ L of the *Dpn I* treated DNA following the established protocol.

The resulting colonies were screened by colony PCR with pHTP primers. Colonies giving a PCR product of the correct size were grown aerobically overnight, at 37 °C and 180 rpm, in LB broth supplemented with 50 μ g/mL kanamycin. The plasmid was extracted using NZYMiniprep Kit (NZYTech) and subsequently sequenced (GATC Biotech).

Heterologous Expression. The heterologous expression of C-Tir sequence-verified clones was performed using *E. coli* chemically competent BL21 Star (DE3) pLysS cells. Following the established protocol, 50 μ L competent cells were transformed with 1 μ L of plasmid containing the gene sequence encoding for C-Tir. The mixture was incubated on ice for 30 minutes, followed by a heat shock at 42 °C for 40 seconds. After 5 minutes on ice, 500 μ L of pre-warmed LB broth was added to the mixture and the cells were grown in an incubator shaker (AccuTherm, Labnet), at 37 °C for 60 min at 180 rpm. Lastly, 100 μ L cell suspension were plated onto LB agar plates supplemented with 50 μ g/mL kanamycin and 34 μ g/mL chloramphenicol and incubated overnight at 37 °C.

Transformed *E. coli* single cell colonies were aerobically cultured in 60 mL of LB broth supplemented with 50 μ g/mL kanamycin (NZYTech) and 34 μ g/mL chloramphenicol (NZYTech). Cultures were grown aerobically at 37 °C and 180 rpm, overnight. After this period, cells were transferred to 500 mL LB broth supplemented with the antibiotic concentrations above mentioned, ensuring an initial OD_{600nm} of 0.05.

The steps above were the same employed to produce unlabeled or labeled C-Tir. From this step forward, two different processes, described in the following sections, were carried out to produce unlabeled, selectively unlabeled, or uniformly labeled C-Tir.

Unlabeled protein expression. Cell cultures were grown aerobically in LB broth at 37°C and 165 rpm until $OD_{600nm} \sim 0.6-0.7$. Protein expression was induced by adding sterile isopropyl β -D-1-thiogalactoside (IPTG) to a final concentration of 1 mM. After adding IPTG, the bacterial growths were incubated for 24 hours at 20 °C with orbital rotation at 110 rpm.

Uniformly labeled protein expression. Cell cultures were grown aerobically in LB broth at 37°C and 165 rpm until $OD_{600nm} \sim 1.0$. After reaching the desired OD_{600nm} , cells were harvested by centrifugation in a Beckman centrifuge, model Avanti J-26 (rotor JA-10) at 12000 xg for 15 minutes at 4 °C. The cell pellets were resuspended in 500 mL of Minimal media containing 33.7 mM Na_2HPO_4 , (Fisher Bioreagents), 22 mM KH_2PO_4 (Fisher Bioreagents), 8.55 mM NaCl (PanReac), 1 mM $MgSO_4$ (PanReac), 0.3 mM $CaCl_2$ (Fisher Scientific), 1 μ g Biotin (Sigma), 1 μ g Thiamin (Acros Organics) and 1x Trace Elements solution (134 μ M EDTA, 31 μ M $FeCl_3 \cdot 6H_2O$, 6.2 μ M $ZnCl_2$, 760 nM $CuCl_2 \cdot 2H_2O$, 420 nM $CoCl_2 \cdot 2H_2O$, 1.62 μ M H_3BO_3 , 81 nM $MnCl_2 \cdot 4H_2O$), 1 g/L $^{15}NH_4Cl$ (Novagen) as nitrogen source and 2 g/L ^{13}C -Glucose as carbon source (Cambridge Isotope Laboratories). The minimal media was also supplemented with 50 μ g/mL kanamycin and 34 μ g/mL chloramphenicol. Expression of uniformly labeled C-Tir was induced with 1 mM IPTG for 24 hours at 20 °C with orbital rotation at 110 rpm.

Selectively unlabeled protein expression. Selective unlabeled C-Tir with L-Arginine, L-Asparagine, L-Histidine or L-Glutamine (PanReac) was carried out by aerobically growing the cell cultures in LB broth at 37°C and 165 rpm until $OD_{600nm} \sim 1.0$. In the same way as for uniformly labeled protein expression, after reaching the desired OD_{600nm} , cells were harvested by centrifugation, and the cell pellets were

resuspended in 500 mL of Minimal media with the same composition above mentioned except for the ^{13}C -Glucose replacement with D(+)Glucose as carbon source. One hour before induction and prior to pellet resuspension, the minimal media was supplemented with 1 g/L of the desired amino acid (L-Arginine, L-Asparagine, L-Histidine or L-Glutamine). The media was also supplemented with 50 $\mu\text{g/mL}$ kanamycin and 34 $\mu\text{g/mL}$ chloramphenicol. Expression of selectively unlabeled C-Tir was induced with 1 mM IPTG. After induction, the cell cultures were incubated for 24 hours at 20 °C with orbital rotation at 110 rpm.

Cell harvesting, disruption and extraction from inclusion bodies. Regardless of the growth protocol employed, upon 24 hours of induction, cells were harvested in a Beckman centrifuge, model Avanti J-26 (rotor JA-10) for 15 minutes at 12000 xg and 4 °C. The cell pellet was gently resuspended in Cell Lysis Buffer (50 mM Tris-HCl (Fisher Bioreagents), pH 7.5, 150 mM NaCl (PanReac), 5 mM EDTA (Fisher Bioreagents), 1mM PMSF (NZYTech) and 15 mM β -Mercaptoethanol (Fisher Bioreagents)) containing 0.5 mg/mL lysozyme (Fisher Bioreagents), 0.5 mg/mL DNase (NZYTech) and 50 μL of protease inhibitor cocktail (Sigma-Aldrich) per gram of cell culture.

Cell lysis was achieved by boiling the cell extract at 90 °C for 30 minutes, gently shaking it every 5 minutes in a water bath (Labnet). The cellular extract was clarified by ultracentrifugation in a Beckman Coulter Optima LE-80K (rotor 70 Ti) for 40 minutes, 42000 rpm and 4 °C. The clear supernatant was supplemented with a protease inhibitor cocktail (Sigma-Aldrich) and stored at 4 °C. The cell pellet was washed and resuspended in Wash Buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 1 mM PMSF, 15 mM β -Mercaptoethanol, 1% Triton X-100, 0.5 M Urea, pH 7.5). Cell lysate was centrifuged for 25 minutes, 42000 rpm and 4 °C. Once again, the supernatant was supplemented with a protease inhibitor cocktail (Sigma-Aldrich) and stored at 4 °C, and the cell pellet was washed and resuspended in Cell Lysis Buffer. The cell lysate was centrifuged for 40 minutes, at 42000 rpm and 4 °C. Clarified cell lysates collected upon

ultracentrifugation were gathered to perform the below-described purification steps.

Protein purification. Clarified cell lysate was loaded onto a 5 mL Strep-Tactin[®] XT column (IBA Lifesciences) equilibrated in 3C Cleavage Buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1mM DTT, pH 7.5). Upon complete sample loading and column washing with 5 column volumes (CV) of 3C Cleavage Buffer, on-column cleavage with HRV 3C protease was performed overnight at 4 °C by loading 50-200 μ L HRV 3C protease (1 mg HRV-3C : 100 mg target protein ratio) diluted in 4.5 mL of 3C Cleavage Buffer. On the following day, unbound proteins were washed out with 3C Cleavage Buffer and elution of C-Tir was achieved using BXT Buffer (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 50 mM Biotin, pH 8.0). Upon elution, the column was washed with 15 CV of freshly prepared 10 mM NaOH, equilibrated with 8 CV Storage Buffer (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) and stored at 4 °C. Protein purity was evaluated by SDS-PAGE using a Mini-PROTEAN[®] Electrophoresis System (Bio-Rad) (Supplementary Figure SI.2). Briefly, 5 μ L 5 $\times$ SDS-PAGE Sample Loading Buffer (NZYTech) were added to 20 μ L protein samples. Samples containing loading buffer were boiled for 3 minutes and loaded on an SDS-PAGE gel with 5% and 12% acrylamide stacking and resolving gel, respectively. After 1 hour at 180 V, proteins were detected with BlueSafe (NZYTech).

Fractions containing the pure protein were pooled together and concentrated in NMR Buffers (20 mM Sodium Phosphate, 150 mM NaCl, 1 mM EDTA, pH 6.5; 20 mM HEPES, 150 mM NaCl, 12 mM MgCl₂, 1 mM DTT, pH 6.8; or 20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 6.8) or CD Buffer (20 mM Sodium Phosphate, 150 mM NaF, 1 mM EDTA, pH 6.5) using an Amicon Ultra-15 centrifugal filter unit with 3 KDa cutoff (Merck). Protein concentration was determined on Nanodrop ND-2000C, assuming a theoretical extinction coefficient of 16960 M⁻¹ cm⁻¹ for wild type C-Tir and PRE-mutants (V405C, S428C, S463C, S509C and S536C), and of 13980 M⁻¹ cm⁻¹ for the Y454F_Y474F double mutant.

Protein tyrosine phosphatase SHP-1

Constructs design and cloning. SHP-1 is an ~67 kDa protein tyrosine phosphatase containing two tandem Src homology 2 (SH2) domains followed by a catalytic domain and an inhibitory C-terminal tail comprising two sites for tyrosine phosphorylation.

The human SHP-1 C-SH2 domain, comprising residues 101 to 217, was PCR-amplified from a vector with the cDNA encoding the tandem SH2 domains of SHP-1 acquired from Addgene (pGEX SHP-1(NC)-SH2; Plasmid #46496), using the primers listed in Table 2.3. The purified PCR-product was then inserted into the pHTP8 vector (NZYTech), fused to an N-terminal Trx-His₆-tag and C-terminal Strep-tag (Supplementary Figure SI.1), using the same methodology employed for C-Tir.

Table 2.3. Oligonucleotide sequences (primers) used for amplification of SHP-1 C-SH2 domain. Underlined nucleotides correspond to the HRV-3C protease specific sequence. The DNA sequence coding for the Strep-tag sequence is highlighted in bold. fw and rv indicate forward and reverse primers, respectively.

Primer	Sequence (5'→ 3')
SHP-1_C-SH2_fw	TCAGCAAGGGCTGAGGCTCGAGGTA <u>CTCTTTCAAGGA</u> <u>CCGCTGGGATCCGGGCGAGCTTCCGATCCCACTAGTG</u> AGAGG
SHP-1_C-SH2_rv	TCAGCGGAAGCTGAGGCTACTTCTCGAACTGCGGATG GCTCCACCTCGTGGCATAGTACGGCTG

Heterologous expression. Chemically competent *E. coli* BL21 Star (DE3) pLysS were transformed with the plasmid encoding the SHP-1 C-SH2 domain, following the established protocol. Transformed single colonies were aerobically cultured overnight at 37 °C and 180 rpm in 60 mL of LB broth supplemented with 50 µg/mL kanamycin and 34 µg/mL chloramphenicol. The next day, cells were transferred to 500 mL LB broth supplemented with antibiotics and grown aerobically at 37°C and 180 rpm

Materials and Methods

until OD_{600nm} ~ 0.6. Protein expression was induced with 0.1 mM IPTG for 24 hours at 20 °C with orbital rotation at 120 rpm. Upon 24 hours of induction, cells were harvested in a Beckman centrifuge, model Avanti J-26 (rotor JA-10) for 15 minutes at 12000 xg and 4 °C.

Protein purification. Cell pellets were resuspended in Lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5), supplemented with EDTA-free protease inhibitor cocktail. Cell disruption was then achieved using a French® Press high-pressure cell (Thermo Electron). Clarification of the cellular extract was done by ultracentrifugation in a Beckman Coulter Optima LE-80K (rotor 70 Ti) for 1 hour, at 42000 rpm and 4 °C. Clarified cell lysate was loaded onto a 5 mL Strep-Tactin® XT column (IBA Lifesciences) using the same protocol described for C-Tir purification. Upon separation of the fusion tags, protein purity was evaluated by SDS-PAGE (Supplementary Figure SI.3), and the fractions containing the protein of interest were concentrated and loaded onto a Superdex™ 200 Increase 10/300 GL column (Sigma Aldrich) for a final step of purification using size exclusion chromatography. The fractions containing pure protein were pooled together and concentrated in NMR Buffers (20 mM HEPES, 150 mM NaCl, 12 mM MgCl₂, 1 mM DTT, pH 6.8; or 20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 6.8) using an Amicon Ultra-15 centrifugal filter unit with 10 KDa cutoff (Merck). Protein concentration was determined on Nanodrop ND-2000C, assuming an extinction coefficient of 18450 M⁻¹ cm⁻¹.

Protein tyrosine phosphatase SHP-2

Constructs. SHP-2 is an ~68 kDa protein tyrosine phosphatase comprised of two SH2 domains (N-terminal and C-terminal) and a catalytic protein tyrosine phosphatase (PTP) domain. The plasmids encoding the human N-terminal SH2 domain (N-SH2, residues 1-104), C-terminal SH2 domain (C-SH2, residues 101-222) and tandem SH2 domains (NC-SH2, residues 1-222) of SHP-2 were acquired from Addgene (pGEX SHP-2(N)-

SH2 Plasmid #46498, pGEX SHP-2(C)-SH2 Plasmid #46497, pGEX SHP-2(NC)-SH2 Plasmid #46499). All plasmids contained an N-terminal Glutathione S-transferase (GST) tag followed by an HRV-3C recognition sequence (Supplementary Figure SI.1), allowing recombinant protein purification by affinity chromatography and separating the tags from the protein of interest.

Site-directed mutagenesis. A mutant for the SHP-2 N-SH2 domain was designed to improve stability and avoid aggregation. In this mutant, the cysteine residue on position 104 was replaced by a serine residue (C104S). The last non-natural occurring 10 residues on the C-terminal (PGRLERPHRD), present in the original plasmid acquired from Addgene were also removed by insertion of a stop codon.

Point mutations were also introduced in the SHP-2 NC-SH2 domain. Arginine and Histidine in positions 32 and 53 (Arg32 and His53), respectively, were substituted by alanine residues to create the NC-SH2 N-SH2 dead mutant, and Arginine and Histidine in positions 138 and 169 (Arg138 and His169), respectively, were replaced by alanine residues for the NC-SH2 C-SH2 dead mutant. The oligonucleotide sequences used for amplification are listed in Table 2.4.

Table 2.4. Oligonucleotide sequences (primers) used for site-directed mutagenesis of SHP-2 SH2 domains. Underlined nucleotides correspond to the sequence encoding the desired mutation. fw and rv indicate forward and reverse primers, respectively.

SHP-2 mutant	Primer	Sequence (5' → 3')
N-SH2 stop codon	Stop_fw	TGTTTAATTAACGAATTCT <u>AGGG</u> TCGACTCGAGCG GCCG
	Stop_rv	CGGCCGCTCGAGTCGACC <u>CTAG</u> AATTCGTTAATTA AACA
N-SH2 C104S	C104S_fw	GCTTAAATATCCTCTGAACT <u>CGTT</u> AATTAACGAAT TCTAGGGTCG
	C104S_rv	CGACCCTAGAATTCGTTAATTAAC <u>GAG</u> TTTCAGAG GATATTTAAGC
NC-SH2 N-dead	R23A_fw	GGCAGTTTTTTGGCAG <u>CGC</u> CTAGTAAAAGTAACC CTGG
	R32A_rv	CCAGGGTACTTTTACTAGG <u>CGCT</u> GCCAAAAAAC TGCC
	H53A_fw	GGAGCTGTCACC <u>GCC</u> ATCAAGATTCAGAACTG G
	H53A_rv	CCAGTGTTCTGAATCTTGAT <u>GCG</u> GTGACAGCTC C
NC-SH2 N-dead	R138A_fw	GGTAGTTTTCTTGTA <u>GCA</u> GAGAGCCAGAGCCACC C
	R138A_rv	GGGTGGCTCTGGCTCTCT <u>GCT</u> ACAAGAAAACTAC C
	H169A_fw	GGCAAGTCTAAAGTGACC <u>GCT</u> GTTATGATTGCT GTCAGG
	H169A_rv	CCTGACAGCGAATCATAAC <u>AGCG</u> GTCACTTTAGA CTTGCC

All SHP-2 mutants used in this work were prepared using the NZYMutagenesis kit (NZYTech), following the established protocol.

Heterologous expression. Chemically competent *E. coli* BL21-Codon Plus were transformed with the plasmids encoding the different SH2 domains of SHP-2, following the established protocol. Transformed single colonies were aerobically cultured overnight at 37 °C and 180 rpm, in 60 mL of LB broth supplemented with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. The next day, cells were transferred to 500 mL LB broth supplemented with antibiotics, and grown aerobically at 37°C and 150 rpm until OD_{600nm} ~ 0.6. Protein expression was induced with 0.2 mM IPTG for 4 hours at 37 °C with orbital rotation at 150 rpm. After induction, cells were harvested in a Beckman centrifuge, model Avanti J-26 (rotor JA-10) for 15 minutes at 12000 xg and 4 °C. Cell pellets were stored at – 20 °C until further use.

Protein purification. Cell pellets were resuspended in Lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 10 mM β-Mercaptoethanol, pH 7.5), supplemented with 0.5 mg/mL lysozyme, 0.5 mg/mL DNase and 50 µL of protease inhibitor cocktail, per gram of cell culture. Cell disruption was achieved using a French® Press high pressure cell (Thermo Electron), performing 3 cycles at 1000 psi. Clarification of the cellular extract was done by ultracentrifugation in a Beckman Coulter Optima LE-80K (rotor 70 Ti) for 1 hour and 15 minutes, at 42000 rpm and 4 °C. Clarified cell lysate was loaded onto two coupled 5 mL GSTrap™ affinity columns (Cytiva) equilibrated in 3C Cleavage Buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA and 1mM DTT). Upon complete sample loading and column washing with 5 CV of 3C Cleavage Buffer, protein elution was achieved using 5 CV of Elution Buffer (50mM Tris-HCl, 10 mM reduced glutathione, 1 mM DTT, pH 8.0). GST cleavage was carried out at 4°C by overnight incubation with HRV 3C protease (1 mg HRV-3C : 100 mg target protein ratio) in a 3.5K MWCO dialysis membrane (Thermo Fisher) against 2L of 3C Cleavage Buffer. The next day, the dialyzed protein sample was loaded onto two coupled 5 mL GSTrap™ columns equilibrated in 3C Cleavage Buffer. Proteins of interest were recovered by columns washing with 5 CV of 3C Cleavage Buffer, and GST was eluted using 5 CV of Elution Buffer.

Protein purity was evaluated by SDS-PAGE (Supplementary Figures SI.4-SI.6), and the fractions containing the proteins of interest were concentrated for a final purification step using size exclusion chromatography using a Superdex™ 75 Increase 10/300 GL column (Sigma Aldrich). The fractions containing pure protein were pooled together and concentrated in NMR Buffers (20 mM HEPES, 150 mM NaCl, 12 mM MgCl₂, 1 mM DTT, pH 6.8; or 20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 6.8) using an Amicon Ultra-15 centrifugal filter unit with 10 KDa cutoff (Merck). Protein concentration was determined on Nanodrop ND-2000C, assuming extinction coefficients of 14440 M⁻¹ cm⁻¹, 8480 M⁻¹ cm⁻¹, and 22920 M⁻¹ cm⁻¹ for N-SH2, C-SH2 and NC-SH2 domains, respectively.

C-Terminal domain of TRAF6

Construct design and cloning. Human TRAF6 protein (~60 kDa) harbors an N-terminal RING domain followed by a series of five Zinc finger domains and a C-terminal region that comprises a coiled-coil TRAF6-N domain and a TRAF6-C domain.

The bacterial expression vector containing the gene sequence for human TRAF6 was synthesized by *ATG: biosynthetics*, and used as template for cloning the TRAF6-C domain, comprising residues 350 to 501, into the pCoofy16 expression vector, which contains an N-terminal His₁₀-NusA purification and solubility enhancing tag.

The gene encoding for the TRAF6-C domain was amplified by Polymerase Chain Reaction (PCR) with a set of primers containing complementary primer extensions for the gene of interest, an HRV 3C recognition sequence used for tag removal and the 3' end of the toxic *ccdB* gene used to suppress growth in non-resistant cells at 100% efficiency. Additionally, a DNA sequence encoding for a linker of 7 residues (GNGNGNG) was artificially cloned between the HRV 3C recognition sequence and the gene of interest to ensure proper tags cleavage

(Supplementary Figure SI.1). Briefly, 50 ng of template DNA, 50 pmol of each primer (Table 2.5; TRAF6-C_fw and TRAF6-C_rv primers), 0.4 mM dNTP Mix, 1 Unit of Phusion Polymerase and 1x Phusion Polymerase buffer (New England Biolabs) were used in a PCR reaction.

At the same time, a set of primers was used to linearize pCoofy16 (Table 2.5; LP1_fw_vector primer and LP2_rv_vector primer). For this, a PCR reaction was carried out with a 50 μ L reaction mix containing 25 ng vector DNA, 50 pmol of each primer (LP1_fw and LP2_rv), 0.4 mM dNTP Mix, 1 Unit of Phusion Polymerase and 1x Phusion Polymerase buffer (New England Biolabs).

Table 2.5. Oligonucleotide sequences (primers) used for amplification of TRAF6-C and pCoofy16 linearization. Underlined nucleotides correspond to the GNGNGNG linker encoding sequence and gene specific sequences are highlighted in bold. fw and rv indicate forward and reverse primers, respectively.

Primer	Sequence (5'→ 3')
TRAF6-C_fw	AAGTTCTGTTCCAGGGGCCCGGTAACGGCAACGGGAA CGGAAACGGGTATTTACATCTGGAAAATTGG
TRAF6-C_rv	CCCCAGAACATCAGGTTAATGGCGCTAGGTCGAGACC TCGCAAC
LP1_fw_vector primer	GGGCCCCTGGAACAGAACTTCCAG
LP2_rv_vector primer	CGCCATTAACTGATGTTCTGGGG

Vector linearization and gene fragment size (543 bp) were confirmed by 1.0% (w/v) agarose gel electrophoresis. PCR products were purified using the NZYGelPure kit (NZYTech), following the manufacturer's instructions.

Materials and Methods

For the Sequencing and Ligation Independent Cloning (SLIC) reaction, 100 ng of linearized pCoofy16 vector were mixed with the insert, at a molar ratio of 1:3, 1 μ L recA enzyme (2 μ g/mL stock) and 1 μ L recA buffer (10x) (NEB, Germany) to a final reaction volume of 10 μ L. The reaction mixture was incubated for 1 hour at 37°C followed by transformation into chemically competent *Escherichia coli* Top10 cells, following the established protocol, with a slightly extended heat shock for 1 min at 42 °C. Selected DNA-sequenced clones were aerobically grown overnight at 37 °C and 180 rpm, in LB broth supplemented with 50 μ g/mL kanamycin, for plasmid extraction using the NZYMiniprep Kit (NZYTech) and DNA sequencing (GATC Biotech).

Heterologous expression. The heterologous expression of TRAF6-C sequence-verified clones was performed using *E. coli* chemically competent BL21-Codon Plus cells, following the established transformation protocol. Transformed *E. coli* single cell colonies were aerobically cultured in 60 mL of LB broth supplemented with 50 μ g/mL kanamycin and 34 μ g/mL chloramphenicol, for an overnight period at 37 °C and 180 rpm. After this period, cells were transferred to 500 mL LB broth supplemented with the antibiotic concentrations above mentioned, and grown at 37 °C and 150 rpm until reaching an $OD_{600nm} \sim 0.6$. Protein expression was induced with 0.1 mM IPTG for 16 hours at 18 °C with orbital rotation at 120 rpm. After induction, cells were harvested in a Beckman centrifuge, model Avanti J-26 (rotor JA-10) for 15 minutes at 12000 xg and 4 °C. Cell pellets were stored at - 20 °C until further use.

Protein purification. Cell pellets were resuspended in Lysis buffer (50 mM HEPES, 500 mM NaCl, 5 % (v/v) glycerol, 1 mM DTT, 20 mM imidazole, pH 7.5), supplemented with 0.5 mg/mL lysozyme, 0.5 mg/mL DNase and 50 μ L of EDTA-free protease inhibitor cocktail, per gram of cell culture. Cell disruption was then achieved using a French® Press high pressure cell (Thermo Electron), performing 3 cycles at 1000 psi. Clarification of the cellular extract was done by ultracentrifugation in a

Beckman Coulter Optima LE-80K (rotor 70 Ti) for 2 hours, at 42000 rpm and 4 °C.

Clarified cell lysate was loaded onto two coupled 5 mL HisTrap™ affinity columns (Cytiva) equilibrated in Binding Buffer (50 mM HEPES, 500 mM NaCl, 5 % (v/v) glycerol, 1 mM DTT, 20 mM imidazole, pH 7.5). Upon complete sample loading and column washing with 5 CV of Binding Buffer, protein elution was achieved using 5 CV of Elution Buffer (50mM HEPES, 500 mM NaCl, 5 % (v/v) glycerol, 1 mM DTT, 300 mM imidazole, pH 7.5). Cleavage of N-terminal His₁₀-NusA tags was carried out at 4°C by overnight incubation with HRV 3C protease (1 mg HRV-3C : 100 mg target protein ratio) in a 3.5K MWCO dialysis membrane (Thermo Fisher) against 2L of 3C Cleavage Buffer (50 mM HEPES, 300 mM NaCl, 1 mM DTT, 1 mM DTT). The next day, the dialyzed protein sample was loaded onto two coupled 5 mL HisTrap™ columns equilibrated in Binding Buffer. TRAF6-C domain protein samples were recovered during columns washing with 5 CV of 3C Binding Buffer, and His₁₀-NusA tag was eluted using 5 CV of Elution Buffer. Protein purity was evaluated by SDS-PAGE (Supplementary Figure SI.7), and the fractions containing the protein of interest were concentrated for a final purification step using size exclusion chromatography using a Superdex™ 200 Increase 10/300 GL column (Sigma Aldrich). The fractions containing pure protein were pooled together and concentrated in NMR Buffers (20 mM HEPES, 150 mM NaCl, 12 mM MgCl₂, 1 mM DTT, pH 6.8; or 20 mM HEPES, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 6.8) using an Amicon Ultra-15 centrifugal filter unit with 10 KDa cutoff (Merck). Protein concentration was determined on Nanodrop ND-2000C, assuming a theoretical extinction coefficient of 19440 M⁻¹ cm⁻¹.

NMR spectroscopy

This section describes specific experimental protocols used in the study of each protein or set of proteins studied in this work using NMR spectroscopy. All NMR experiments were acquired at 283 K on a Bruker Avance 800 II+ spectrometer operating at 18.8T (800.1 MHz ^1H Larmor frequency) equipped with a 5 mm TCI triple resonance cryoprobe. All spectra were processed using Bruker TopSpin software (Bruker Biospin, Karlsruhe, Germany) and analyzed using CARA [1].

Backbone assignment. Regarding the backbone assignment of C-Tir, U- $^{13}\text{C}/^{15}\text{N}$ -labeled protein samples (250-950 μM) were prepared in 20 mM Phosphate pH 6.5, 150 mM NaCl, 1mM EDTA, 20 μM 3-trimethylsilyl-1-propanesulfonic acid sodium salt (DSS) and 8% D_2O . To enable backbone assignment, phase-sensitive gradient-enhanced 2D $^1\text{H},^{15}\text{N}$ -HSQC spectra were acquired and preservation of equivalent pathways (PEP) methodology was employed to acquire triple resonance HNCO, HNCA, HN(CA)CO, HN(CO)CA, HNCACB and HN(CO)CACB experiments. 3D $^1\text{H},^{15}\text{N}$ TOCSY-HSQC and $^1\text{H},^{15}\text{N}$ NOESY-HSQC experiments were also acquired using a 480 μM U- ^{15}N -labeled sample of C-Tir.

Assignment of phosphorylated C-Tir (pC-Tir) was achieved with HNCACB and HNCO experiments. To this end, a U- $^{13}\text{C}/^{15}\text{N}$ -labeled pC-Tir protein sample (1 mM) was prepared in 20 mM HEPES pH 6.8, 150 mM NaCl, 12 mM MgCl_2 , 5.5 mM ATP, 20 μM DSS and 8% D_2O .

^1H chemical shifts were referenced directly, and ^{15}N and ^{13}C chemical shifts indirectly using DSS (methyl ^1H signal at 0.00 ppm). Relevant acquisition parameters are summarized in Table 2.6. Sequential backbone assignment of C-Tir was achieved starting by picking peaks in 2D $^1\text{H},^{15}\text{N}$ -HSQC root spectrum, followed by pick peaking in the 3D spectra, in order to combine the peaks into spin systems associated with the root peaks on $^1\text{H},^{15}\text{N}$ -HSQC spectrum. The assignment was then achieved by selecting

spin systems with characteristic chemical shift values and identifying spin systems sequential to it. When unique, these sequential connectivities were extended to generate a fragment of uniquely connected spin systems which were then assigned on the basis of chemical shifts. This procedure was repeated until all possible spin systems were assigned.

The NMR chemical shifts of C-Tir and pC-Tir are available at the Biological Magnetic Resonance Bank (BMRB) [2] under the entries 50758 and 50759, respectively.

Table 2.6. Acquisition details of NMR experiments for C-Tir and pC-Tir backbone assignment. The number of scans used in each experiment is represented by ns.* In 3D ^1H , ^{15}N TOCSY-HSQC and ^1H , ^{15}N NOESY-HSQC the ^{13}C dimension is replaced by ^1H .

Sample	NMR experiment	Bruker pulse program	Complex Points			Spectral width (Hz/ ppm)			Frequency offset (Hz/ ppm)			ns
			^1H	^{15}N	^{13}C *	^1H	^{15}N	^{13}C *	^1H	^{15}N	^{13}C *	
C-Tir	2D ^1H , ^{15}N -HSQC	hsqcetf3psi	2048	256	—	11160.71/ 13.945	1906.0/ 23.500	—	3753.55/ 4.690	9650.55/ 119.000	—	10
	3D HNCO	hncogp3d	2048	64	64	8802.81/ 10.998	1885.72/ 22.250	2314.70/ 11.500	3753.55/ 4.690	9609.96/ 118.500	35279.92/ 175.300	16
	3D HNCA	hncagp3d	2048	64	64	8802.81/ 10.998	1885.72/ 22.250	5031.34/ 25.000	3753.55/ 4.690	9609.96/ 118.500	10464.64/ 52.000	16
	3D HN(CA)CO	hncacogp3d	2048	64	72	8802.81/ 10.998	1885.72/ 22.250	2314.70/ 11.500	3753.55/ 4.690	9609.96/ 118.500	35277.92/ 175.300	24
	3D HN(CO)CA	hncocagp3d	2048	64	64	8802.81/ 10.998	1885.72/ 22.250	5031.34/ 25.000	3753.55/ 4.690	9609.96/ 118.500	10464.64/ 52.000	16
	3D HNCACB	hncacbgp3d	2048	56	96	8802.81/ 10.998	1702.23/ 21.000	11068.82/ 55.000	3753.55/ 4.690	9528.87/	8150.35/ 40.500	24
	3D HN(CO)CACB	hncocacbgp3d	2048	56	96	8802.81/ 10.998	1702.23/ 21.000	11068.82/ 55.000	3753.55/ 4.690	9528.87/	8150.35/ 40.500	24
	3D ^1H , ^{15}N TOCSY-HSQC *	mlevhshqcetf3gp3d	2048	48	128	8802.81/ 10.998	1885.72/ 22.250	8803.67/ 11.000	3753.55/ 4.690	9609.96/ 118.500	3753.55/ 4.690	8
	3D ^1H , ^{15}N NOESY-HSQC *	noesyshqcetf3gp3d	2048	48	128	9615.39/ 12.014	1906.0/ 23.500	9604.01/ 12.000	3753.55/ 4.690	9650.55/ 119.000	3753.55/ 4.690	8
pC-Tir	2D ^1H , ^{15}N -HSQC	hsqcetf3psi	2048	256	—	11160.71/ 13.945	1906.0/ 23.500	—	3753.55/ 4.690	9650.55/ 119.000	—	10
	3D HNCO	hncogp3d	2048	64	64	8802.81/ 10.998	1885.72/ 22.250	2314.70/ 11.500	3753.55/ 4.690	9609.96/ 118.500	35279.92/ 175.300	16
	3D HNCACB	hncacbgp3d	2048	56	96	8802.81/ 10.998	1702.23/ 21.000	11068.82/ 55.000	3753.55/ 4.690	9528.87/ 117.500	8150.35/ 40.500	72

Secondary chemical shifts. Secondary chemical shift values ($\Delta\delta\alpha-\Delta\delta\beta$) were calculated as the difference between the experimental chemical shifts and their amino-acid specific random-coil values of intrinsically disordered proteins [3]. To provide a more robust local structure prediction, all the assigned backbone resonances were combined into a neighbor-corrected sequence structural propensity calculator (nsSPC) score [4].

Paramagnetic Relaxation Enhancement experiments. PRE experiments are based on the site-specific attachment of a paramagnetic spin label to a biomolecule [5]. Therefore, C-Tir mutants V405C, S428C, S463C, S509C and S536C were prepared for further conjugation with the thiol-reactive methanethiosulfonate spin label (MTSL). Heterologous expression and purification of all [$U-^{15}N$]-labeled mutants was achieved as described previously for wild type C-Tir, with the exception of adding reducing agents in the buffers composition to prevent formation of inter-molecular disulfide bonds.

To attach the MTSL probe to the $U-^{15}N$ -labeled purified cysteine mutants of C-Tir, buffer exchange to Labeling Buffer (50 mM Tris-HCl pH 7.5 and 50 mM NaCl) was performed. Briefly, after elution of C-Tir variants in BXT Buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 50 mM Biotin and 1 mM DTT), fractions containing the protein were pooled together and concentrated to a final volume of 2.5 mL. The protein fraction was then loaded on a PD-10 desalting column (GE Healthcare) placed above a 50 mL conical tube with 100 μ L of labeling solution containing Labeling Buffer and MTSL in the amount needed to achieve a final concentration of MTSL of 10x the molar amount of protein to be labelled. The protein was then eluted with 2.4 mL of Labeling Buffer flowing through the column and into the labeling solution, while DTT present in the protein solution remained on the column. The conical tube containing the modification reaction was covered with foil and incubated overnight at 4 °C. On the following day, the MTSL-labeled protein sample was buffer exchanged into NMR Buffer (20 mM Sodium Phosphate pH 6.5, 150 mM NaCl and 1 mM EDTA) using an Amicon Ultra-15

centrifugal filter unit with 3 KDa cutoff (Merck), concentrating the protein sample while removing any excess of unbound MTSL.

Prior to NMR data acquisition, two protein samples (250-300 μ M) for each variant were prepared in 20 mM Phosphate pH 6.5, 150 mM NaCl, 1mM EDTA, 20 μ M DSS and 8% D₂O. One sample containing the MTSL-labeled protein, corresponding to the paramagnetic state, and another containing MTSL-labeled protein and sodium ascorbate in 5x the molar amount of protein. The sample containing sodium ascorbate was incubated for 4 hours prior to NMR data acquisition for reduction of MTSL unpaired electron to form the diamagnetic state. An 2D ¹H,¹⁵N-HSQC pulse sequence was first employed to acquire a spectrum of the freshly MTSL-labeled C-Tir variant, corresponding to the sample in the paramagnetic state. After the 2D ¹H,¹⁵N-HSQC spectrum has been acquired of the paramagnetic protein, a 2D ¹H,¹⁵N-HSQC spectrum of the reduced diamagnetic MTSL protein sample was acquired using identical acquisition parameters to those used to acquire the paramagnetic spectrum. Diamagnetic spectrum was analyzed first as it was the most closely related to the assigned spectrum of native C-Tir. Cross-peaks in the diamagnetic spectrum were assigned using the known chemical shifts of the native protein by overlapping the 2D ¹H,¹⁵N-HSQC spectra of native and labeled proteins and transferring the assignment. Then, intensity of cross peaks was measured in the diamagnetic assigned spectrum. After that, the cross peaks in the paramagnetic spectrum were labeled using the curated chemical shift list employed to analyze the diamagnetic spectrum and peak intensities were measured following the same procedure as for the diamagnetic spectrum. The cross-peak intensity data was finally used to calculate the intensity ratio in the paramagnetic (I_{ox}) and diamagnetic states (I_{red}) for each residue.

2,2,2-Trifluoroethanol assays. U-¹⁵N-labeled purified samples of C-Tir (100 μ M) were prepared in 20 mM Phosphate pH 6.5, 150 mM NaCl, 1mM EDTA, 20 μ M DSS and 8% D₂O and mixed with increasing amounts of TFE: 1.0, 2.0, 5.0, 10.0 and 12.5% (v/v)). A 2D ¹H,¹⁵N-HSQC spectrum was acquired for each sample and the TFE-induced effects were determined as

the ratio of ^1H , ^{15}N -HSQC cross peak intensity of the samples in the absence and presence of TFE.

Lipid binding assays. Lipid bicelles were prepared based on the protocols described in Maffei *et al.*, 2013 [6] and Pérez *et al.*, 2013 [7]. Briefly, 12.5% (w/w) bicellar dispersions were prepared by mixing long-chain (1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) or 1,2-Dimyristoyl-sn-glycero-3-phosphorylglycerol sodium salt (DMPG)) and short-chain (1,2-diheptanoyl-sn-glycero-3-phosphocholine (DHPC)) phospholipids (Avanti Polar Lipids) in chloroform or chloroform: methanol (65: 35) solutions. Solvent evaporation was achieved under a nitrogen flow and the resulting lipid films were rehydrated in 20 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA pH 6.5. The mixtures were subjected to 35 freeze-thaw cycles with vortexing until lipid suspensions were clear and transparent. Bicelle size dispersion was checked by dynamic light scattering. The molar ratio of lipids in bicelle samples was 1.0 DHPC: 0.8 DMPG ($q = 0.8$, 44.4% DMPG molar ratio) and 1.0 DHPC: 0.8 DMPC ($q = 0.8$, 44.4% DMPC molar ratio). The q -ratio (DMPC or DMPG)/DHPC) of 0.8 provides isotropic fast tumbling particles.

The lipid bicelles were then added to U- ^{15}N -labeled purified samples of C-Tir and pC-Tir in 20 mM Phosphate pH 6.5, 150 mM NaCl, 1mM EDTA, 20 μM DSS and 8% D_2O , for a final concentration of C-Tir of 115 μM and 6% (w/v) of lipids bicelles (DMPC/ DHPC or DMPG/ DHPC). 2D ^1H , ^{15}N -HSQC spectra were acquired for each sample, as well as for a reference C-Tir samples in the absence of lipid bicelles. Lipid-binding effects were determined as the ratio of ^1H , ^{15}N -HSQC cross peak intensity of the samples in the absence and presence of lipid bicelles.

***In vitro* phosphorylation.** The phosphorylation of C-Tir tyrosine residues Y454, Y474, Y483 and Y511, located on its ITIM-like sequences, was carried out *in vitro* using the Fyn recombinant human protein kinase (Thermo Fisher Scientific). In brief, a U- $^{13}\text{C}/^{15}\text{N}$ -labeled C-Tir protein sample (1 mM) in 20 mM HEPES pH 7.5, 150 mM NaCl, 12 mM MgCl_2 and 5.5 mM ATP was incubated with Fyn (1.93 μM) for 16 hours at 25 °C. The amount of Fyn used was calculated taking into account the enzyme's optimal activity at pH = 7.5

and 25 °C, as well as its ability to completely phosphorylate four substrate sites. Upon overnight phosphorylation, the sample was buffer exchanged to 20 mM HEPES pH 6.8, 150 mM NaCl, 12 mM MgCl₂ 5.5 mM ATP, concentrated and further supplemented with 20 µM DSS and 8% D₂O to acquire a 2D ¹H,¹⁵N-HSQC spectrum. Phosphorylation of C-Tir was assessed by comparison of the 2D ¹H,¹⁵N-HSQC spectra of pC-Tir and the non-phosphorylated protein (969 µM) in the same buffer. Phosphorylation caused subtle changes in backbone resonances, and thus, re-assignment of pC-Tir was achieved by acquiring 3D HNCACB and HNCO experiments (Table 2.6).

Quenched phosphorylation reactions. Quenched phosphorylation reactions were carried out to determine early time points of C-Tir phosphorylation reaction and extract quantitative information about individual phosphorylation states. To this end, a U-¹⁵N-labeled C-Tir protein sample in 20 mM HEPES pH 7.5, 150 mM NaCl, 12 mM MgCl₂ and 5.5 mM ATP was incubated with Fyn following the same conditions described above. At selected time points upon addition of Fyn (2, 4, 8, 16, 24, 32, 48, and 64 minutes), aliquots of pC-Tir (final concentration of 50 µM) were collected, placed at 95 °C for 5 minutes to stop the phosphorylation reaction, flash-freezed in liquid nitrogen and stored at -80 °C. The samples were then thawed and supplemented with 20 µM DSS and 8% D₂O to be analyzed by NMR spectroscopy. A single 2D ¹H,¹⁵N-HSQC spectrum was acquired for each sample using the same spectrometer settings as for the non-phosphorylated C-Tir reference sample, used as a control for the reaction.

To extract quantitative information about site-specific phosphorylation levels, an arbitrary number of NMR signals that were not affected by the phosphorylation reactions was initially selected. The peak intensities of these signals were measured in all spectra ensuring that the values did not change substantially over the reaction time course. For each selected control signal, the intensity ratios in each time-course spectra (*I*) with respect to their intensities in the reference NMR spectrum (*I*₀) were calculated. Control signal ratios were then averaged, yielding a global scaling factor *F*_{scale} that reports

the differences in experimental conditions between reference and reaction samples.

To correct changes in experimental conditions, the intensities of progressively disappearing signals of the unmodified substrate site residues (I_{unphos}) were multiplied by F_{scale} in all NMR spectra corresponding to different reaction time courses. Then, the relative intensities of these signals in relation to the absolute intensity (I_{unphos}) in the reference data set ($t = 0$) were calculated according to equation 2.1.

$$\frac{F_{scale} \cdot I_{unphos}(t)}{I_{unphos}(t=0)} \quad (2.1)$$

On the basis of the law of mass conservation, the individual phosphorylation levels for all disappearing residues (j) was calculated according to the equation 2.2.

$$\% \text{ Phosphorylation } (t) = 1 - \frac{F_{scale} \cdot I_{unphos}^j(t)}{I_{unphos}^j(t=0)} \quad (2.2)$$

In order to draw conclusions about the individual phosphorylation behaviors, the individual phosphorylation levels of each tyrosine residue (% Phosphorylation) throughout the reaction were fitted to a Michaelis-Menten curve by adjusting K_m and V_{max} parameters, following equation 2.3.

$$\frac{[pTyr](t)}{[Tyr](t=0)} = 1 - \left(\frac{K_m}{[Tyr](t=0)} \cdot \ln \left(1 + \left(e^{\frac{[Tyr](t=0)}{K_m}} - 1 \right) \cdot \exp \left(-\frac{V_{max}}{K_m} \cdot t \right) \right) \right) \quad (2.3)$$

It is, however, important to point out that these adjusted parameters only provide semi-quantitative information about the individual phosphorylation behaviors and must be analyzed in relative terms. To obtain the exact properties and kinetic constants of these phosphorylation reactions, measurements at different substrate concentrations should be performed in the future.

Protein-protein interactions. All interaction studies were carried out with NMR titrations of the multiple unlabeled domains of SHP-1, SHP-2 and TRAF6-C over U-¹⁵N-labeled samples of C-Tir, pC-Tir or pC-Tir Y454F_Y474F. 2D ¹H, ¹⁵N-HSQC spectra were acquired for each titration with increasing molar equivalents on unlabeled titrants. Further experimental details are listed on Table 2.7.

Table 2.7. Overview of the interaction studies carried out between C-Tir and SHP-1, SHP-2 or TRAF6 human proteins.

Labeled Sample	Unlabeled titrant	Molar equivalents of titrant
C-Tir	SHP-1 C-SH2	1.0, 2.0, 4.0, 6.0, 8.0
	SHP-2 C-SH2	1.0, 2.0, 4.0, 8.0
	SHP-2 NC-SH2	1.0, 2.0, 4.0, 8.0
	TRAF6-C	0.25, 0.50, 0.75, 1.0, 1.5, 2.0
pC-Tir	SHP-1 C-SH2	0.25, 0.50, 0.75, 1.0, 2.0, 3.0, 4.0, 6.0
	SHP-2 N-SH2	0.23, 0.47, 0.94, 1.4, 1.9, 3.7, 7.5, 11.2
	SHP-2 C-SH2	0.35, 0.70, 1.4, 2.8, 5.6, 11.2, 16.7
	SHP-2 NC-SH2	0.23, 0.46, 0.91, 1.4, 1.8, 3.7, 5.5, 7.3, 10.9
	SHP-2 NC-SH2 N-dead	0.35, 0.70, 1.4, 2.8, 5.6, 11.2
	SHP-2 NC-SH2 C-dead	0.35, 0.70, 1.4, 2.8, 5.6, 11.2
	TRAF6-C	0.25, 0.50, 0.75, 1.0, 1.5, 2.0
pC-Tir Y454F_Y474F	SHP-2 NC-SH2	0.48, 0.96, 1.4, 1.9, 3.8, 7.7, 11.5
	TRAF6-C	1.0

Intensity ratio perturbation (I_{obs}^i) of the NMR signals from C-Tir were plotted against the molar ratio of titrant/C-Tir for all titrations. As peaks adjacent to each phosphorylation site exhibited a similar intensity drop, they were clustered by site and globally fitted into a 1:1 binding model to estimate the dissociation constant (K_D) for each interaction.

Assuming a simple 1:1 binding model, between proteins A and B ($A + B \rightleftharpoons AB$), the dissociation constant is given by equation 2.4.

$$K_D = \frac{[A][B]}{[AB]} \Leftrightarrow K_D \cdot [AB] = [A][B] \quad (2.4)$$

During the titration, the total concentration of proteins A and B, designed as A_0 and B_0 , is given by the sum of free and bound states:

$$A_0 = [A] + [AB]; B_0 = [B] + [AB]$$

Given that:

$$\begin{aligned} AB &= f^i \cdot A_0 \\ A &= (1 - f^i) \cdot A_0 \\ B &= B_0 - f^i \cdot A_0 \end{aligned}$$

f^i , representing the fraction of bounded protein for a residue i , can be given by:

$$\begin{aligned} (f^i \cdot A_0) \cdot K_D &= [(1 - f^i) \cdot A_0] \cdot (B_0 - f^i \cdot A_0) \Leftrightarrow \\ \Leftrightarrow f^i \cdot K_D - (B_0 - f^i \cdot A_0 - B_0 \cdot f^i + f^{i2} \cdot A_0) &= 0 \Leftrightarrow \\ \Leftrightarrow f^i &= \frac{(K_D + B_0 + A_0) - \sqrt{(K_D + B_0 + A_0)^2 - (4 \cdot B_0 \cdot A_0)}}{2 \cdot A_0} \end{aligned}$$

As $f^i = \frac{I_{max}^i - I_{obs}^i}{I_{max}^i - I_{min}^i}$, equation 2.5 can be further employed to estimate the K_D value:

$$I_{obs}^i = I_{max}^i - (I_{max}^i - I_{min}^i) \cdot \frac{(K_D + B_0 + A_0) - \sqrt{(K_D + B_0 + A_0)^2 - (4 \cdot B_0 \cdot A_0)}}{2 \cdot A_0} \quad (2.5)$$

where I_{max}^i is the maximal relative intensity of protein in free-state and I_{min}^i is the minimal relative intensity of bound protein under saturation conditions for residue i . I_{obs}^i is the relative intensity of residue i , corrected by a F_{scale} value (equation 2.1). A_0 is the total protein concentration of protein A, in this case, C-Tir, and B_0 is the concentration of protein B, in this case the unlabeled titrant. The fitting was achieved by defining I_{min}^i and K_D as adjustable parameters.

Other techniques

This section includes the protocols employed for C-Tir characterization using CD spectroscopy and SAXS, as well as the methods employed for disorder prediction analysis and ensemble generation. The bioinformatic analysis of search for occurrences of putative ELMs was performed by Dr. Miguel Arbesú and Dr. Andreas Zanzoni under the scope of a collaboration. This section also describes the methodologies employed in mutagenesis of Tir, immunofluorescence, Propidium Iodide (PI) uptake assays and EPEC infection studies performed by Qiyun Zhong under the supervision of Professor Gad Frankel at the Centre for Molecular Bacteriology and Infection, Imperial College London, under the scope of a collaboration established in this work.

Disorder prediction and short linear motif analysis. The sequence of A/E pathogen effectors and corresponding reference proteomes were collected from the UniprotKB database (UniProt Consortium, 2021). The initial effector collections fetched from UniprotKB contained 28 *Citrobacter rodentium* (CR), 39 EHEC and 24 EPEC effectors, respectively. A single UniProt reference proteome was found to represent most effectors for each species (Table 2.8): UP000001889 (CR, including 24 of the 28 effectors), UP000000558 (EHEC, all effectors are present), and UP000008205 (EPEC including 20 out of 24 effectors). For the comparison of the disorder fraction between effector and the reference proteomes, only these subsets of effectors

with reference proteomes were considered. For the remaining analyses, all effectors were used.

Table 2.8. Reference proteomes by specie name and taxon.

Specie	A/E type	Taxon	Reference Uniprot proteome
<i>Homo sapiens</i>	-	9606	UP000005640 (Swiss-Prot reviewed)
<i>E. coli</i> O157:H7	EHEC	83334	UP000000558 (Full)
<i>E. coli</i> O127:H6	EPEC	574521	UP000008205 (Full)
<i>C. rodentium</i> (strain ICC168)	CR	637910	UP000001889 (Full)

Three effector sets were assembled for the following strains: EHEC O157:H7, EPEC H127:H6, and CR. The representative set of 20365 human proteins was extracted from UniprotKB human reference proteome. To compute disorder propensity at the residue level, DISOPRED 3 [8] and IUPred 1.0 ("long" mode prediction) [9] structural disorder predictors were employed. Disordered residues were defined as those with a propensity score equal to or above 0.5. This metric was used to calculate the fraction of disordered residues for each protein. The one-sided Mann-WhitneyU-test [10] was employed to compare disorder fraction distribution of the effector collections and their corresponding proteomes. Both per residue scores and aggregated disorder fractions were used to classify each protein according to the five structural categories adapted from [11] (IDP: Intrinsically disordered proteins; PDR: Proteins with intrinsically disordered regions; FRAG: Proteins with fragmented-disorder; NDR: Not disordered proteins; ORD: Ordered Proteins). The categories were based on average disorder content and presence of significant disordered stretches. Based on the shorter average length of prokaryotic proteins with respect to eukaryotic ones ($320/472=0.68$) [12] and the largest window size of the disorder predictors (21 amino acids in IUPred), the threshold for disordered stretches was changed from the original 30 amino

acids to 22. The corresponding class descriptions and criteria are summarized in Table 2.9.

Table 2.9. Structural categories by disorder content (Adapted from [11]).

Class	Description	Average disorder content	≥ 22 consecutive disordered residues
IDP	Intrinsically disordered proteins	≥ 30 %	Yes
PDR	Proteins with intrinsically disordered regions	< 30 %	Yes
FRAG	Proteins with fragmented disorder	≥ 30 %	No
NDR	Not disordered proteins	$10 \% \leq x < 30$ %	No
ORD	Ordered proteins	< 10 %	No

For short linear motif analysis, the list of eukaryotic short linear motifs (ELMs) was downloaded from the ELM database [13] (version 1.4, May 2017). A search for occurrences of putative ELMs in effector sequences was carried out using the ANCHOR tool [14]. For each effector sequence, the motif density was calculated as the fraction between the amino acids belonging to putative ELMs predicted as disordered and the total number of amino acids predicted as disordered.

Tir sequences alignment. A set of 153 protein sequences belonging to the "Tir receptor family" was downloaded from UniprotKB database [15]. Sequence redundancy was reduced using the CD-HIT algorithm [16] and a set of 114 non-identical sequences was compiled. Multiple sequence alignment (MSA) was generated using the Clustal Omega program [17] with default parameters and JalView [18] was used for MSA visualization and figure generation. Pairwise alignments among EPEC, EHEC and CR Tir

sequences were performed using the Smith-Waterman algorithm as implemented in the EMBOSS software suite [19].

CD spectroscopy. For Circular Dichroism measurements, C-Tir samples (11.9 μM) were prepared in CD Buffer (2 mM phosphate, 15 mM NaF and 0.1 mM EDTA, pH 6.5) supplemented with 0, 1, 2, 5, 10, 25 and 50 % TFE (v/v). The CD spectra were recorded in-house with a J-815 spectrometer (Jasco) at 10°C with 10 accumulations collected in the 190 nm to 260 nm wavelength range using a 0.1 nm data pitch and 2 nm bandwidth at a scan speed of 50 nm/min. All spectra were averaged over 10 scans and converted to mean residue ellipticity $[\theta]$.

SAXS. C-Tir protein samples were prepared in 20 mM Phosphate pH 6.5, 150 mM NaCl and 1 mM EDTA at 1.0, 2.5, 5.0, 7.5 and 10.0 mg/mL. As a standard, BSA (Sigma Aldrich) protein samples were also prepared at the same concentrations as the C-Tir protein samples. SAXS measurements were performed in BM29 beamline at ESRF (Grenoble, France). Data was collected in 1-second frames at a wavelength of 0.9919 Å using a Pilatus 1M (DECTRIS) detector and a sample-detector distance of 2.867 m, covering a momentum transfer (s) range of 0.00359-0.48896 Å⁻¹. For SEC-SAXS measurements, a 50 μL sample at 10.0 mg/mL was injected into a 2.4 mL Superdex™ 200 Increase 2.2/300 size exclusion column (GE Healthcare) connected to a Shimadzu HPLC System. The scattering intensities from the monomeric C-Tir single-peak region were integrated and buffer subtracted using ScÅtter [20]. A series of 50 μL C-Tir samples at 1.0, 2.5, 5.0, 7.5, and 10.0 mg/mL were also run in batch mode. To account for concentration effects, such as interparticle interference, 7.5 and 10mg/mL curves (10 frames each) were integrated, buffer subtracted and merged using PRIMUS [21].

Data was processed using ATSAS program package [22]. PRIMUS was employed to calculate the radius of gyration, R_g , using the Guinier approximation. The distance distribution function, $P(r)$, and the maximum particle size, D_{max} , were determined using GNOM [23]. The SAXS profiles have been deposited at the repository for small-angle scattering data

SASBDB [24] under the project "SAXS studies on the intracellular region of the translocated intimin receptor" with the accession code SASDKH8.

Structural ensemble generation. A random coil ensemble containing 50.000 conformations was generated for C-Tir using Flexible-Meccano (FM) [25] where torsion angle pairs were randomly selected from a database of amino acid specific conformations in loop regions of high-resolution X-ray structures. After that, side-chain modeling [26], energy-refinement and *in-silico* incorporation of multiple spin-label dispositions at single-cysteines to interpret the PRE data [27,28] were carried out.

For each conformer, a SAXS pattern simulation was performed using CRY SOL [29] and theoretical intramolecular ^1H - ^{15}N PRE-PRE rates were determined assuming the Solomon-Bloembergen approximation [28]. To account for the flexibility of the paramagnetic MTSL moiety in the analysis of the measured PREs, multiple states derived from a 100 ns Molecular Dynamics (MD) simulation were used, as recently described for the PROXYL spin-label [27]. Force field AMBER99sb-ILDN [30] was used with ions added, and TIP3P for the water model [31]. Frames collected every 100 ps from the production MD-run were then extracted to create a library of 1000 different MTSL conformations. From this library, engineered single-cysteine residues were *in-silico* labeled with multiple sterically allowed spin-label dispositions. This MTSL ensemble representation enabled the estimation of the order parameters associated with the dipolar proton-electron interaction vector motion [32]. With this strategy, the PRE rates for V409C, S428C, S463C, S509C, and S536C were independently calculated for each C-Tir conformer. Lastly, a searched for subsets of 200 models that collectively best reproduce the five PRE rates datasets and the experimental SAXS profile was carried out using a genetic algorithm [33].

Tir-TRAF6 complex modeling with AlphaFold2. AlphaFold-multimer [34], a refined version of AlphaFold2 [35], was used to predict binding interfaces in the Tir-TRAF6 complex. In a first stage, the sequences of C-Tir and TRAF6-C were used as input to predict the 3D structure of Tir bound to TRAF6.

Next, we run three independent predictions replacing the C-Tir sequence by the peptide regions engaged in binding as seen by NMR. We did not use template structures for the predictions iterated for up to 48 recycles, followed by energy refinement with AMBER using default settings implemented in LocalColabFold [36] and using MMseqs2 for creating multiple sequence alignments [37]. Model confidence was assessed by the predicted Local Distance Difference Test (pLDDT) and inter-complex predicted alignment error (PAE), i.e., the uncertainty about the interface. Regions with pLDDT > 90 are expected to have high accuracy.

Tir lipid-binding site mutations. The pSA10-Tir_{WT} plasmid was used as a template to insert point mutations in full length Tir. The primer sequences used and resulting plasmids are listed in Tables 2.10 and 2.11, respectively.

Table 2.10. Oligonucleotide sequences (primers) used for site-directed mutagenesis of full length Tir. Underlined nucleotides correspond to the sequence encoding the desired mutation. fw and rv indicate forward and reverse primers, respectively.

Primer	Sequence (5'→ 3')
Tir_388E389E_Fw	TGCGCTCCAT <u>gaagaa</u> AATCAGCCGGC
Tir_388E389E_Rv	GTCGTTACACCAGCACCA
Tir_513E_Fw	TACTTATGCG <u>gag</u> CTGGCAAACAGCGG
Tir_513E_Rv	CTTTGGATACCTTGCCCTG
Tir_541*_Fw	TGTAAATGCC <u>taa</u> CCAACGCCGG
Tir_541*_Rv	GAACCTACTGCGCTCTCG
Tir_CT_seq	CGATATGAGGATCAGCATGCC

Table 2.11. Plasmids used for Tir lipid binding site mutations.

Plasmid name	Plasmid description
pSA10-Tir _{WT}	pSA10 plasmid encoding wild-type Tir from EPEC E2348/69
pSA10-Tir _{388E389E}	pSA10 plasmid encoding EPEC Tir with R388E and R389E mutations
pSA10-Tir _{513E}	pSA10 plasmid encoding EPEC Tir with L513E mutation
pSA10-Tir _{541*}	pSA10 plasmid encoding EPEC Tir with a STOP codon replacing A541
pSA10-Tir _{388E389E513E}	pSA10 plasmid encoding EPEC Tir with R388E, R389E and L513E mutations
pSA10-Tir _{388E389E513E541*}	pSA10 plasmid encoding EPEC Tir with R388E, R389E and L513E mutations and a STOP codon replacing A541

Residues R388 and R389 were mutated to glutamic acid using primers Tir_388E389E_Fw and Tir_388E389E_Rv. L513 were mutated to glutamic acid using primers Tir_513E_Fw and Tir_513E_Rv. A STOP codon was introduced at A541 using the primers Tir_541*_Fw and Tir_541*_Rv. The resulting plasmids were transformed into *E. coli* DH5 α , isolated and sequenced by Eurofins using primer Tir_CT_seq before the next round of mutation. To create pSA10-Tir_{388E389E513E}, a second round of mutation was performed using primers Tir_513E_Fw and Tir_513E_Rv to introduce the 513E mutation into pSA10-Tir_{388E389E}. To create pSA10-Tir_{388E389E513E541*}, a third round of mutation was performed using primers Tir_541*_Fw and Tir_541*_Rv to introduce the 541* STOP codon into pSA10-Tir_{388E389E513E}. The resulting plasmids were also transformed into *E. coli* DH5 α , isolated and sequenced using primer Tir_CT_seq. All sequencing-verified plasmids (Table 2.11) were electroporated into EPEC-0.

EPEC infection. Model intestinal epithelial SNU-C5 cells were seeded at 110^5 on glass coverslips in each well of a 24-well plate (immunofluorescence) or 510^4 in each well of a black clear bottom 96-well plate (PI uptake) in Park Memorial Institute (RPMI) medium (Sigma-Aldrich) with 10% (v/v) foetal bovine serum (FBS) (Gibco), 2 mM Glutamax (Gibco), 1 mM sodium pyruvate (Sigma-Aldrich), 10 mM HEPES (Sigma-Aldrich), and 2.5 µg/mL glucose (Sigma-Aldrich). SNU-C5 cells were primed by IFN γ at 10 µg/mL 24 h before infection. EPEC E2348/69 strains were cultured at 37 °C with 180 rpm shaking overnight in Luria Bertani (LB) medium and sub-cultured 1:50 in unsupplemented Dulbecco's Modified Eagle Medium (DMEM) (low glucose) for 3 hours at 37 °C with 5% CO $_2$ static before infection. IPTG (Sigma-Aldrich) at 0.5 mM was added to the bacterial culture to induce protein expression from the pSA10 plasmid 30 minutes before infection. Infection was performed at a multiplicity of infection of 50:1. Infected cells were centrifuged at 700 g for 10 minutes and incubated at 37 °C with 5% CO $_2$ static for required time.

Immunofluorescence imaging. At 2 hours post infection, infected cells were fixed by 4% paraformaldehyde for 15 minutes, washed by PBS for 3 times, permeabilized by 0.2% Triton X-100 (Sigma Aldrich) in PBS for 4 minutes, washed again by PBS and blocked by 1% bovine serum albumin (BSA) (Sigma-Aldrich) for 10 minutes. Samples were incubated with rabbit anti-EPEC serum (VLA) (1:1000) in 1% BSA for 45 minutes, washed by PBS, re-blocked and incubated with the donkey anti-rabbit IgG-Alexa 488 antibody (Jackson ImmunoResearch) (1:200), phalloidin-TRITC (Sigma-Aldrich) (1:200) and DAPI (Sigma) (1:1000) in 1% BSA for another 30 minutes and washed by PBS afterwards. Coverslips were mounted on glass slides (VWR) with Gold-Pro-Long-Anti-fade (Invitrogen). Imaging was performed using a Zeiss AxioImager Z1 microscope (Carl Zeiss) equipped with 100 lenses.

Propidium iodide uptake assay. Cells were incubated with 5 µg/mL propidium iodide (PI) (Invitrogen) from 1 hour before infection in phenol red free serum-free RPMI medium. Positive control was performed by treating uninfected cells incubated in 0.05% Triton X-100 in PI-supplemented RPMI

medium. Uninfected cells were used as negative control. No cell medium-only wells were used as blank. Time-course measurements were carried out in a FLUOstar Omega Microplate Reader (BMG Labtech) from 10 min to 8 hours post infection with 10-minutes intervals, measuring 620-nm emission with 520-nm excitation. After 2 hours post infection, 200 µg/mL gentamicin (Sigma-Aldrich) was added to each well. All PI uptake data were blank-corrected and normalized to negative control. Increase in percentage PI uptake was then calculated using the normalized PI uptake value of the infected sample divided by the normalized PI uptake value of the positive control. Statistical analysis of PI uptake assay was performed using GraphPad Prism 5.1.

References

- [1] Keller, R.L.J. Computer aided resonance assignment tutorial. Cantina, 2004.
- [2] Romero, P.R., Kobayashi, N., Wedell, J.R., Baskaran, K., Iwata, T., Yokochi, M., Maziuk, D., Yao, H., Fujiwara, T., Kurusu, G., Ulrich, E.L., Hoch, J.C., Markley, J.L. BioMagResBank (BMRB) as a Resource for Structural Biology. *Methods Mol Biol.* **2112**, 187-218 (2020).
- [3] Tamiola, K., Acar, B., Mulder, F.A.A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 18000–18003 (2010).
- [4] Tamiola, K., Mulder, F.A.A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40**, 1014–1020 (2012).
- [5] Sjødt, M., Clubb, R.T. Nitroxide Labeling of Proteins and the Determination of Paramagnetic Relaxation Derived Distance Restraints for NMR Studies. *Bio Protoc.* **7**, e2207 (2017).
- [6] Maffei, M., Perez, Y., Amata, I., Arbesú, M., Pons, M. Lipid binding by disordered proteins. *Protoc exch.* (2013).
- [7] Pérez, Y., Maffei, M., Igea, A., Amata, I., Gairí, M., Nebreda, A.R., Bernadó, P., Pons, M. Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci Rep.* **3**, 1295 (2013).
- [8] Jones, D.T., Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics.* **31**, 857-63 (2015).
- [9] Dosztányi, Z., Csizmok, V., Tompa, P., Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* **21**, 3433-4 (2005).

- [10] Mann, H.B, Whitney, D.R. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *Ann. Math. Statist.* **18**, 50–60 (1947).
- [11] Giansanti, A., Deiana, A., Forcelloni, S., Porrello, A. New classification of intrinsic disorder in the human proteome. *BioRxiv* (2018) <https://doi.org/10.1101/446351>.
- [12] Tiessen, A., Pérez-Rodríguez, P., Delaye-Arredondo, L.J. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res Notes.* **5**, 85. (2012).
- [13] Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., *et al.* ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* **44**, D294-300 (2016).
- [14] Mészáros, B., Simon, I., Dosztányi, Z. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol.* **5**, e1000376 (2009).
- [15] UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49** (D1), D480-D489 (2021).
- [16] Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* **28**, 3150-2 (2012).
- [17] Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7**, 539 (2011).
- [18] Procter, J.B., Carstairs, G.M., Soares, B., Mourão, K., Ofoegbu, T.C., Barton, D., Lui, L., Menard, A., Sherstnev, N., Roldan-Martinez, D., Duce, S., Martin, D.M.A., Barton, G.J. Alignment of Biological Sequences with Jalview. *Methods Mol Biol.* **2231**, 203-224 (2021).
- [19] Rice, P., Longden, I., Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276-7 (2000).
- [20] Förster, S., Apostol, L., Bras, W. *Scatter: Software for the analysis of nano- and mesoscale small-angle scattering.* *J. Appl. Cryst.* **43**, 639–646 (2010).
- [21] Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst.* **36**, 1277–1282 (2003).
- [22] Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhnev, A.G., Gaida, M., Gorba, C., Mertens, H.D., Konarev, P.V., Svergun, D.I. New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Cryst.* **45**, 342–350 (2012).
- [23] Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* **25**, 495–503 (1992).

- [24] Valentini, E., Kikhney, A.G., Previtali, G., Jeffries, C.M., Svergun, D.I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357-63 (2015).
- [25] Ozenne, V., Bauer, F., Salmon, L., Huang, J.R., Jensen, M.R., Sefard, S., Bernardó, P., Charavav, C., Blackledge, M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463-1470 (2012).
- [26] Eyal, E., Najmanovich, R., McConkey, B.J., Edelman, M., Sobolev, V. Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J Comput Chem.* **25**, 712-24 (2004).
- [27] Cordeiro, T.N., Sibille, N., Germain, P., Barthe, P., Boulahtouf, A., Allemand, F., Bailly, R., Vivat, V., Ebel, C., Barducci, A., Bourquet, W., le Maire, A., Bernardó, P. Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression. *Structure* **8**, 1270-1285 (2019).
- [28] Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M.R., Zweckstetter, M., Blackledge, M. NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc.* **132**, 8407-18 (2010).
- [29] Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., Svergun, D.I. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Cryst.* **50**, 1212-1225 (2017).
- [30] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).
- [31] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-935 (1983).
- [32] Chen, Y., Campbell, S.L., Dokholyan, N.V. Deciphering protein dynamics from NMR data using explicit structure sampling and selection. *Biophys J.* **93**, 2300-6 (2007).
- [33] Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., Svergun, D.I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc.* **129**, 5656-64 (2007).
- [34] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2021).
- [35] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583-589 (2021).
- [36] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M. ColabFold - Making protein folding accessible to all. *bioRxiv* 2021.08.15.456425 (2021).

[37] Steinegger, M., Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol.* **35**, 1026-1028 (2017).

Chapter 3

Results

Results

This chapter is devoted to the exposition and brief discussion of the results obtained in this thesis. It is divided into two main sections/messages entitled “The pathogen-encoded signaling receptor Tir exploits host-like intrinsic disorder for infection” and “Into the molecular basis of host TLR signaling suppression by the bacterial effector Tir”.

The pathogen-encoded signaling receptor Tir exploits host-like intrinsic disorder for infection

Structural disorder and short linear motifs are common features of A/E pathogen effectors

EPEC and EHEC belong to the family of attaching and effacing (A/E) pathogens. This family also includes the mouse-specific pathogen *Citrobacter rodentium* (CR). To quantify the prevalence of disorder propensity among A/E pathogens, the disorder content of their effectors was estimated using DISOPRED3 [1] and IUPred 1.0 [2] (Supplementary Table SI.1). As a result, we observed that A/E effector proteins have a higher disorder propensity than their proteome counterparts in EHEC O157:H7, the closely related EPEC H127:H6, and CR (Figure 3.1; Supplementary Figure SI.8).

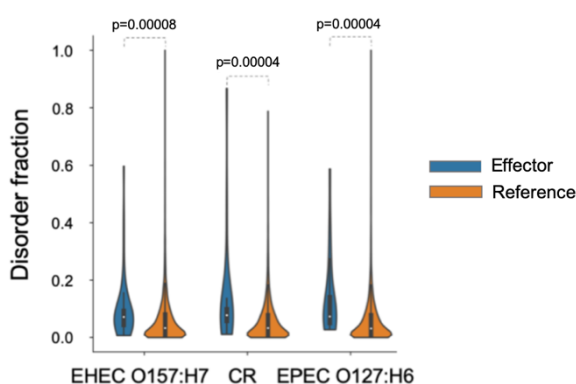


Figure 3.1. Distribution of disorder fraction in A/E pathogens. Effectors (blue) and full-proteomes (red) of A/E pathogens are displayed as blue and orange violin plots, respectively.

Results

Subsequently, we classified these effectors into five structural categories based on disorder content (Materials and Methods, Table 2.9). Our analysis shows that A/E pathogens have a structurally diverse repertoire of effectors, ranging from fully unstructured to ordered proteins (Figure 3.2).

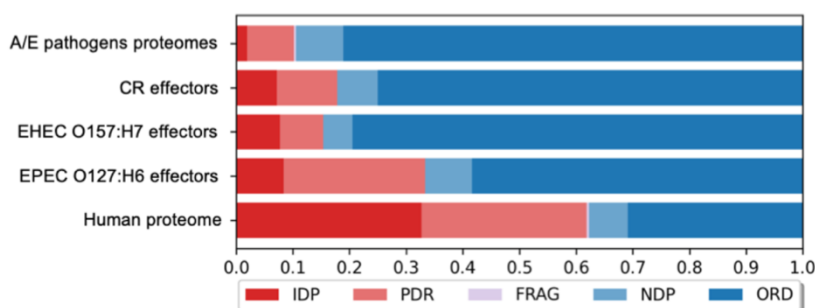


Figure 3.2. Accumulated fractions of the structural categories in terms of structural disorder. IDP: Intrinsically disordered proteins; PDR: Proteins with intrinsically disordered regions; FRAG: Proteins with fragmented-disorder; NDP: Not disordered proteins; ORD: Ordered Proteins.

While most proteins are folded in all prokaryotic collections (Supplementary Figure SI.9 A), fully disordered effectors are predicted to be two to four-fold more frequent than the whole bacterial proteome (1.9% vs. 7.7 % on average). Furthermore, Partially Disordered Proteins (PDR) with long disordered regions occur in EPEC effectors similar to the human proteome. For instance, we classified as PDR the EspB effector involved in cytoskeleton rearrangement and previously identified as an inherently less-ordered effector [3] (Figure 3.3).

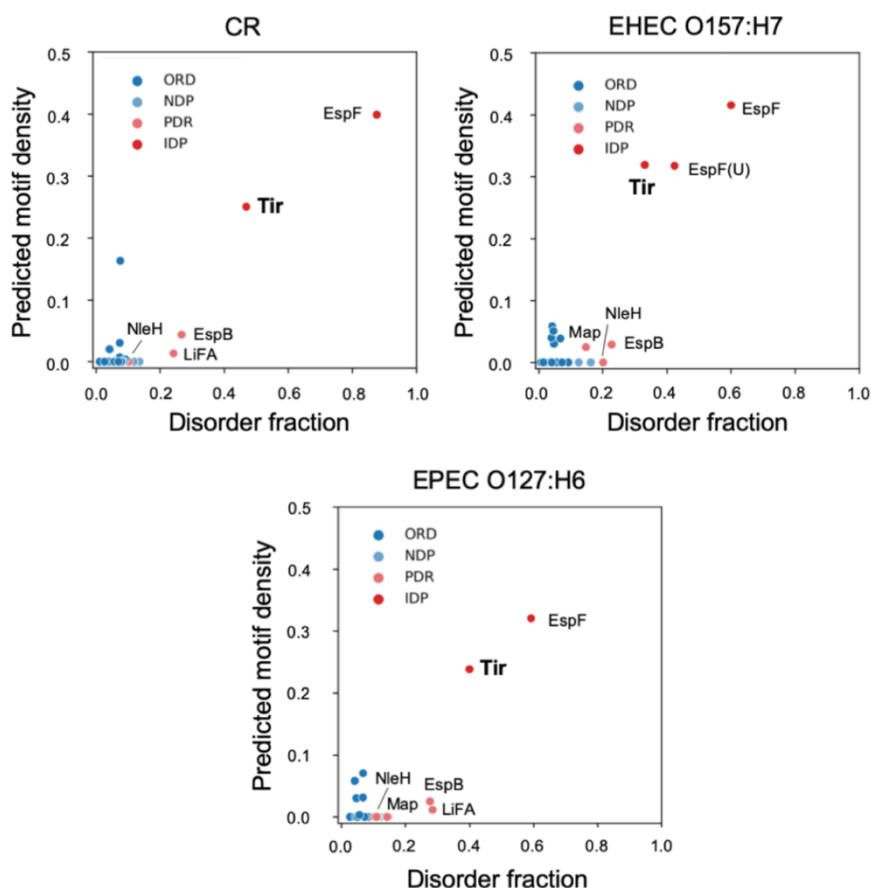


Figure 3.3. Fraction of ELMs vs. disorder fraction in A/E effectors. Tir, EspF, and EspF(U) display a high motif and disorder content. IDP (red) and PDR (pink) effectors are labeled.

An additional example of PDR is the NleH effector, a protein kinase with a disordered N-terminal domain that binds the ribosomal protein 3 (RPS3) to manipulate the NF- κ B pathway [4]. In the order-to disorder continuum, the effector EspF is akin to EspF(U) as a fully disordered protein (Figure 3.3). Both effectors contain multiple consecutive repeats of linear motif pairs able to bind critical host components for actin assembly, with EspF having three to five repetitions depending on the strain [5] and EspFu, only secreted by EHEC, five-and-a-half repeats [6] (Supplementary Figure SI.9 B). Our analysis further

Results

highlighted Tir as a disordered effector (Figure 3.3). Tir displays a high disorder propensity within its IBD and intracellular regions (Supplementary Figure SI.9 B).

Pathogens can exploit host-like motifs as a strategy to interact with host components to subvert host cellular functions [7,8]. On this basis, in collaboration with Dr. Miguel Arbesú and Dr. Andreas Zanzoni, we searched the A/E effectors for eukaryotic linear motifs (ELMs) and found that aside from EspF and EspF(U), other A/E effectors bear multiple putative ELM instances (Figure 3.3). Tir emerged again as having several predicted ELMs in its intracellular regions, including experimentally verified SH3- and SH2-domain binding motifs [9]. Structural disorder and high ELM density support Tir's ability to interact with several host proteins [10]. There is a correlation between disorder content and ELM density, being Tir, EspF(U), and EspF clustered with the highest disorder fraction and motif overall density, similarly to eukaryotic IDPs [11,12]. This correlation highlights ELMs within disordered segments of bacterial effectors as a strategy to disrupt host networks.

Tir bears an intrinsically disordered C-terminal tail

Tir secreted by A/E pathogens share substantial sequence-conservation (Supplementary Figure SI.10) and high frequency of disorder-promoting residues (Supplementary Figure SI.9 B), particularly at the C-terminal region (C-Tir) localized to the host intracellular side.

We found that purified C-Tir exhibited an aberrant migration pattern during SDS-PAGE, a trait commonly displayed by IDPs/IDRs [13-15]. C-Tir migrates as a polypeptide with *ca.* 21 kDa (Figure 3.4), exhibiting a ~ 4 kDa shift over the molecular weight of 17.8 kDa predicted from the amino acid sequence, revealing a larger hydrodynamic radius than that of a globular protein of similar length in the presence of denaturant agents, concurrent with protein flexibility.

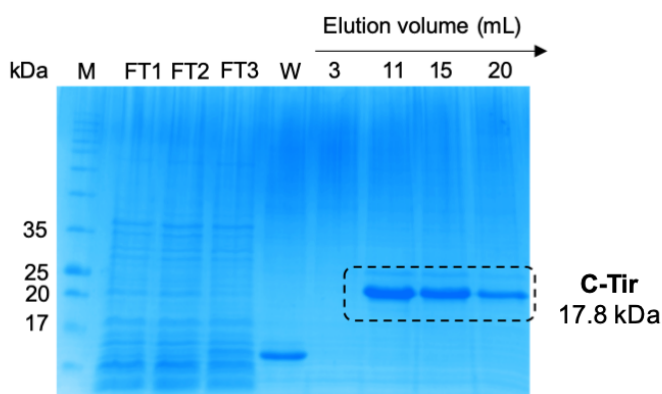


Figure 3.4. C-Tir displays an aberrant migration profile during SDS-PAGE.

Reducing SDS-PAGE analysis of C-Tir purification from a cleared lysate loaded onto a StrepTactin affinity column. Lane M corresponds to the Protein Marker II molecular weight marker (NZYTech). Lanes FT1, 2, and 3 correspond to column flow-through fractions. Following on-column cleavage, tags were washed from the column (Lane W) before elution of C-Tir.

To assess the structural disorder propensity of EPEC C-Tir, multiple biophysical methods were employed, including small-angle X-ray scattering (SAXS), circular dichroism (CD), and nuclear magnetic resonance (NMR). For this integrative study, a protein construct encompassing the residues 388-550 from the full length EPEC Tir, encoding for the C-terminal cytosolic region, was used.

Analytical size-exclusion chromatography (SEC) in combination with SAXS indicated a monomeric state for C-Tir in solution with a radius of gyration (R_g) of 38.8 ± 0.2 Å and a maximum distance ($D_{\max}$) of 128.0 ± 5.0 Å (Supplementary Figure SI.11; Table SI.2). As the R_g -values of IDPs approximately follow a power law of $R_g = 2.54N^{0.522}$ as a function of sequence length (N) [16,17], this relationship was employed, obtaining a predicted R_g value of 37.4 Å for the 173-residue C-Tir construct cloned with an additional C-Terminal Strep Tag (8 extra residues). The similarity between the experimental and predicted values suggests that C-Tir might adopt IDP-like structures in solution. The SAXS-derived Kratky plot, often employed to

Results

qualitatively identify disordered states, of C-Tir also shows characteristics of a disordered or unfolded protein, monotonically increasing without a well-defined maximum (Figure 3.5 A). This feature is absent in globular proteins and implies conformational heterogeneity and flexibility for C-Tir.

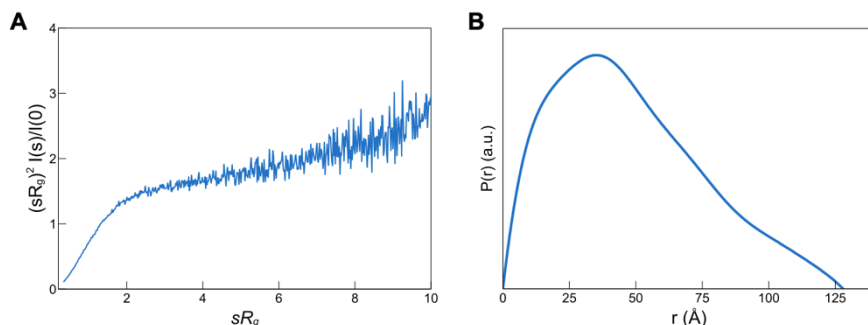


Figure 3.5. Kratky representation **(A)** and pair distance distribution function **(B)** derived from the SAXS experimental profiles measured for C-Tir, indicating its lack of compactness.

Likewise, the asymmetric pair distance distribution function, $P(r)$, obtained from the scattering data (Figure 3.5 B), is compatible with a highly flexible protein sampling pairwise distances far exceeding those expected for a globular protein of the same molecular weight [17] (Supplementary Figure SI.12).

The intrinsic disorder of C-Tir is also reflected in its CD and $[^1\text{H}-^{15}\text{N}]$ -HSQC NMR spectra. C-Tir's CD profile has negative ellipticity at 200 nm and a shallow band in the 210–230 nm range (Figure 3.6 A), indicating a high random coil content with minimal ordered structural elements [18].

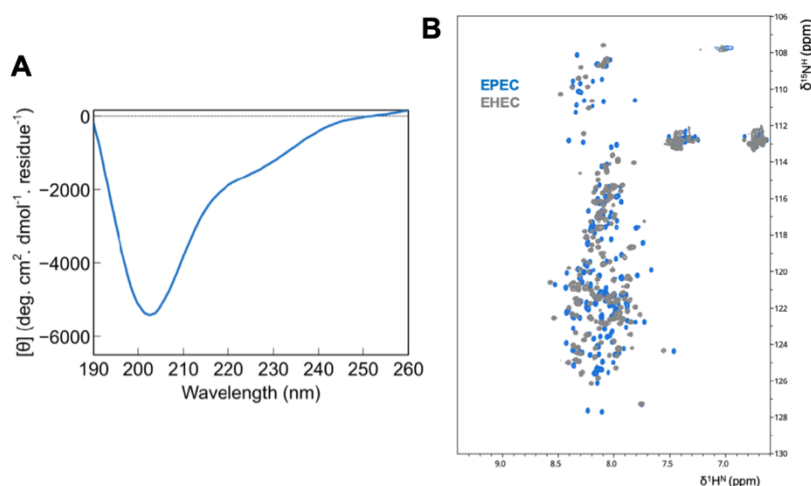


Figure 3.6. The Tir C-terminus is an intrinsically disordered domain. (A) Far-UV circular dichroism spectrum of C-Tir, revealing the absence of high-populated structural secondary elements. **(B)** [¹H-¹⁵N]-HSQC spectra of C-Tir EPEC (light blue) and C-Tir EHEC (grey), revealing the narrow ¹H chemical shift dispersion characteristic of IDPs with ¹H amide backbone resonances clustering between 7.7 and 8.5 ppm.

Regarding NMR, the [¹H-¹⁵N]-HSQC NMR spectrum for this construct displays a typical IDP fingerprint with low chemical shift dispersion [19] (Figure 3.6 B). We observed a similar spectrum for the equivalent region of Tir from EHEC (Figure 3.6 B). Although all A/E pathogens produce Tir, actin polymerization by EHEC O157:H7 Tir C-terminus requires Esp(U)/TccP [20], which is unneeded in EPEC O127:H6 to initiate actin pedestal formation in host cells [21]. Yet, C-Tir is disordered in both strains, highlighting its structural disorder's conservation and functional importance for A/E virulence.

Overall, the different biophysical measurements are consistent with a highly dynamic protein, unambiguously showing that C-Tir is an IDR that lacks a well-folded structure. Such inherent flexibility may enable Tir for multiple binding while acting as a pathogenic scaffold/hub of intracellular host proteins.

C-Tir displays non-random structural preferences at phosphorylation sites

To further dissect the subtle structural features of C-Tir, NMR spectroscopy was employed. The NMR chemical shift is the most readily accessible observable in NMR and it is sensitive to, even partial, secondary structural elements present in IDPs [19]. To gain access to this information, we assigned the backbone resonances of C-Tir nuclei using standard triple-resonance spectra at high-field, together with reverse labeling (Supplementary Figure SI.13). Reverse labeling of selected amino acids [22] proved to be a cost-efficient approach to aid the assignment of this disordered protein by reducing the spectral ambiguity, and should be readily applied to other IDPs.

Provided with the assignment, we calculated the NMR chemical shift deviations ($\Delta\delta$) to random coil values of intrinsically disordered proteins [23]. Although C-Tir has properties of a random coil, these so-called NMR secondary shifts allowed us to identify two regions with non-random structural preferences. In IDPs, transient secondary structural elements are often crucial in acting as molecular recognition units, playing significant roles in binding [24,25].

In Figure 3.7, the consecutive positive $\Delta\delta\text{C}\alpha$ – $\Delta\delta\text{C}\beta$ values indicate a tendency for α -helical structure in two distal segments of C-Tir: (I) D420-Q435; II) I507-A515. Both stretches include residues that are phosphorylated by host kinases, S434 [26] and Y511 [27], respectively.

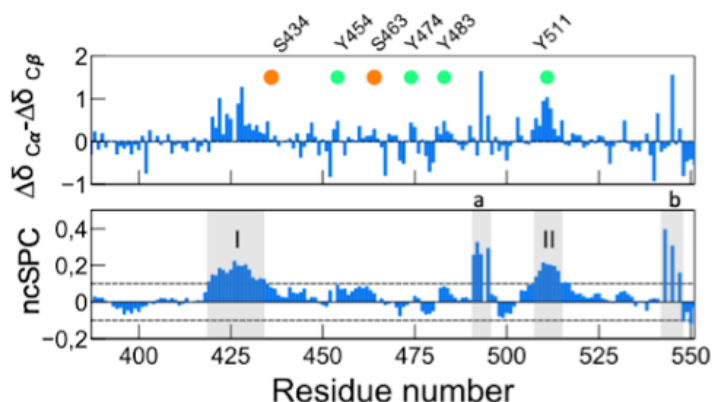


Figure 3.7. C-Tir has partially structured elements. Secondary chemical shifts ($\Delta\delta_{C\alpha} - \Delta\delta_{C\beta}$) (top) and structural propensity plot (bottom) for C-Tir where the dashed-lines depict the random-coil threshold. Values above or below this threshold have α -helix or β -sheet propensities, respectively. Positive values (grey regions) show an increased α -helix tendency at segments D420-Q435 (I) and I507-A515 (II), as well as F489-V495 (a) and T543-V547 (b).

By considering all the assigned backbone resonances into a neighbor-corrected sequence structural propensity calculator (nsSPC) score [28], these apparent local α -helical regions became even more evident. New stretches also emerged with a putative α -helical propensity, including a sequence contiguous to Y511 and downstream tyrosine 483 and a few C-terminal residues (Figure 3.7, bottom). Interestingly, Y483 and Y511 are core residues of Tir's host-like Immunoreceptor tyrosine-based inhibitory motifs (ITIMs) that recruit tyrosine phosphatases SHP-1/2 in an ITIM phosphorylation-dependent manner to inhibit host innate immune responses [27,29].

We further probed C-Tir's non-random coil preferences using paramagnetic relaxation enhancement (PRE) NMR experiments. To this end, five residues scattered along C-Tir's sequence were single-point mutated to cysteine to attach a thiol-reactive nitroxide radical probe (i.e., MTSL). The relaxation induced by this paramagnetic tag resulted in substantial peak broadening.

Results

In ideal random coils, the broadening effect of the paramagnetic tag is observed at the residues in its vicinity, but other residues experience less or no broadening. However, if long-range contacts are present within the disordered chain, residues far away in the sequence can be spatially close to the tag and experience the PRE effect.

On this basis, we calculated the deviations between the experimental PRE and those predicted from a random-coil model (i.e., ΔPRE) to detect non-random coil structural preferences. ΔPRE -values were measured for each single-cysteine mutant. Figure 3.8 shows, in a grayscale, the ΔPRE -induced by MTSL at positions 405, 428, 463, 509, and 536, all consistent with the presence of compact conformations in the disordered C-Tir.

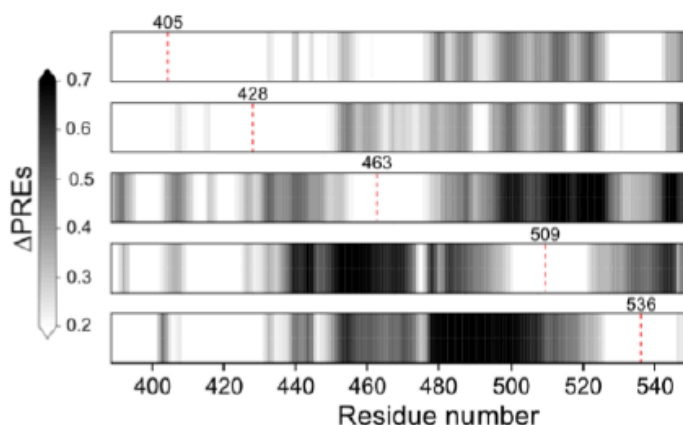


Figure 3.8. C-Tir disordered chain makes long range contacts. ΔPRE -values measured on V405C, S428C, S463C, S509C, and S536C C-Tir mutants as gray-scale heatmaps, from top to bottom, respectively. Color intensity reflects the relative deviations between the experimental PRE and those predicted from a random-coil model ($\Delta\text{PRE} = [I_{\text{para}}/I_{\text{dia}}]_{\text{exp}} - [I_{\text{para}}/I_{\text{dia}}]_{\text{RC}}$). Dashed red lines mark the position of MTSL tags.

The paramagnetic probe in positions 405 and 428 showed noticeable long-range interactions ($\Delta\text{PRE} > 0.2$) with distant downstream residues. The MTSL at S463C caused significant PRE ($\Delta\text{PRE} > 0.6$) in the residues 497–526, and when placed in position 509, it induced broadening in the region 438–479, including position 463. These complementary long-range interactions observed between residues around positions 463 and 509 are mutually self-consistent with compact conformations. Remarkably, both areas include tyrosine-based ELMs that are involved in assembling host components and intracellular signaling.

To better describe and visualize the non-random preferences of C-Tir, we integrated the PRE data with SAXS data to determine the conformational ensemble of C-Tir (Figure 3.9 A).

The established sub-ensembles agree with the experimental data (Figure 3.9 B-C), revealing a bias for expanded and compact structures. Furthermore, the ensemble-averaged distances between the C α atoms for the experimentally selected conformations ($\langle r_{ij} \rangle_{\text{ens}}$), normalized by the respective random coil (RC) ensemble ($\langle r_{ij} \rangle_{\text{RC}}$), show an N-terminal region more extended relative to the RC. At the same time, the C-terminal half is more compact (Figure 3.9 D). Notably, compaction is clustered around the host-like ITIMs implicated in interaction with host proteins.

Results

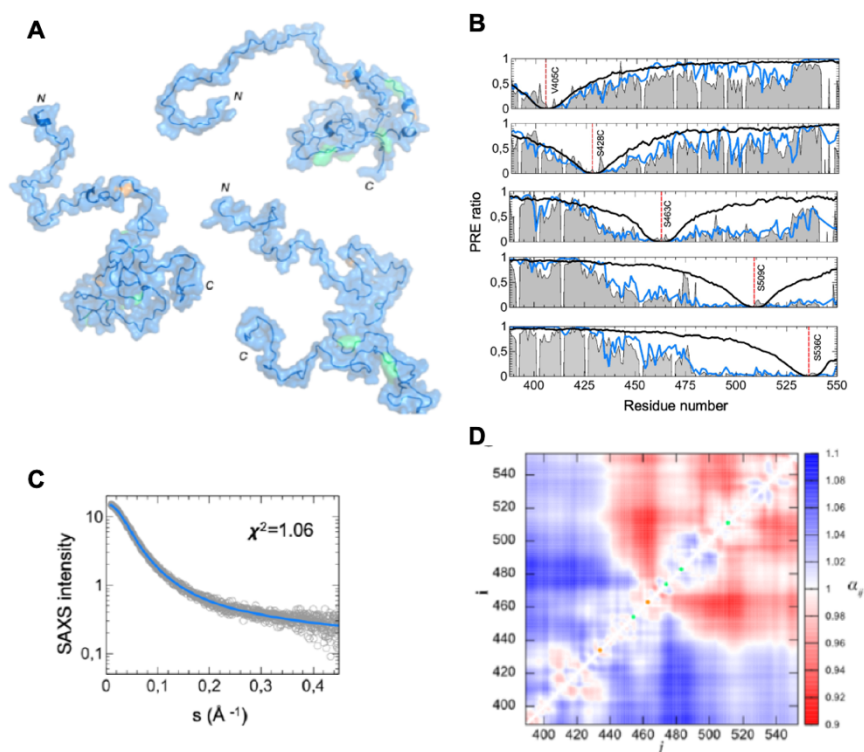


Figure 3.9. C-Tir is disordered with residual structure at sites of host interaction. **(A)** Cartoon representations of 3 representative conformers of the C-Tir ensemble refined with PRE+SAXS data with serine- and tyrosine-phosphorylation sites colored in orange and green, respectively. **(B)** ^1H - ^{15}N C-Tir PRE experimental data (grey) from five single cysteine C-Tir mutants with an MTSL at positions V405C, S428C, S463C, S509C, and S536C (top to bottom, respectively). Back-calculated PRE ratios from an ideal random coil ensemble and from a set of conformers selected using EOM are represented by solid black and blue lines, respectively. **(C)** Logarithmic-scale representation of scattering intensity, $I(s)$, as a function of the momentum transfer, s , measured for C-Tir (empty gray circles). The solid line is the averaged back-calculated curve derived from PRE+SAXS-based ensemble of C-Tir (blue). **(D)** Ensemble-averaged contact map normalized to random-coil distances ($\alpha_{ij} = \langle r_{ij} \rangle_{\text{ens}} / \langle r_{ij} \rangle_{\text{RC}}$). Green and orange circles mark phosphorylation sites on C-Tir for tyrosine and serine residues, respectively.

C-Tir can interact with the host SHP-1 prior to phosphorylation

The intracellular signaling associated with Tir ITIM-like sequences relies on the phosphorylation of the central tyrosine by Src family kinases (SFKs), creating a binding site for SH2-containing proteins, which become activated upon recruitment. Yet, we found that unphosphorylated C-Tir can recognize host SH2-domains with low-intermediate affinity. Using the C-terminal SH2 domain (C-SH2) of phosphatase SHP-1, we identified, by NMR, that residues surrounding the unphosphorylated Y511 ITIM-like motif are involved in binding SH2 domains (Figure 3.10).

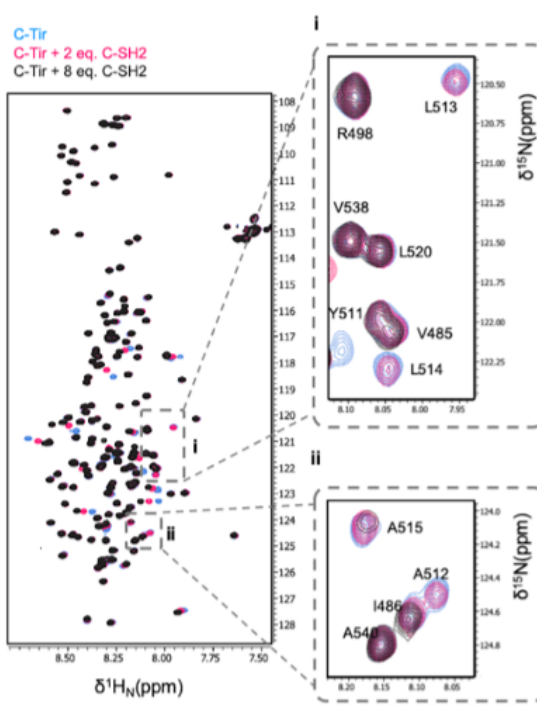


Figure 3.10. C-Tir can bind host C-SH2 domain of SHP-1 pre-phosphorylation. Overlay of $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of C-Tir in the absence (blue) and the presence of C-SH2 at 2.0 (pink) and 8.0 (dark grey) equivalents. Expansions (right panels) show the NMR signal of residues A₅₁₂LLA₅₁₅ with a significant intensity drop due to C-SH2 binding.

Results

Binding to C-Tir caused a selective loss of intensities for the HSQC peaks, mainly from residues A₅₁₂LLA₅₁₅ (Figure 3.11 A). Visible NMR signals retained low-dispersion, indicating that the protein remains mostly disordered and flexible in the complex. We estimated an apparent K_D of ca. 68.5 μ M for this interaction (Figure 3.11 B).

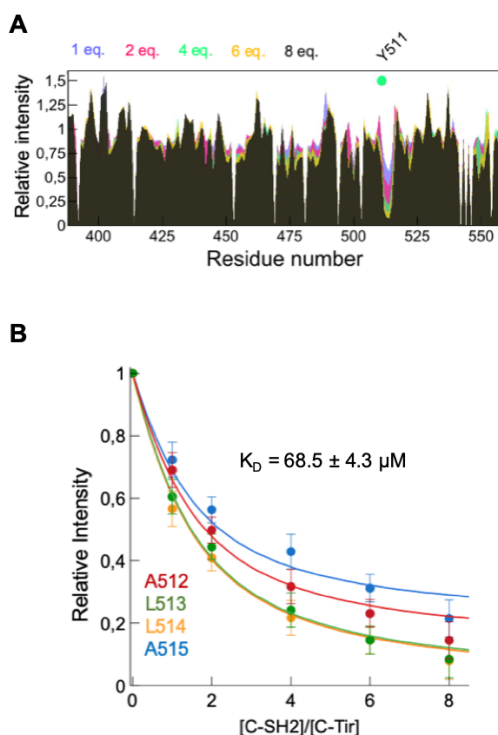


Figure 3.11. Interaction of C-Tir with the C-SH2 domain of SHP-1. (A) Relative [^1H - ^{15}N]-peak intensities from the titration analysis. The green circle marks the tyrosine Y511 position. **(B)** Global fitting NMR quantification of the binding of C-SH2 to unphosphorylated C-Tir.

Thus, besides encoding for α -helical structures, the sequence surrounding Y511 (i+4) appears to have a chemical signature for SH2-domain binding pre-phosphorylation. This observation highlights that those residues adjacent to Y511 can define a binding motif for C-SH2, likely serving as a precursor to a tighter binding upon phosphorylation. The presence of a

transient pre-formed helix with long-range interactions at this region suggests a possible role for this structure as molecular recognition elements for C-Tir.

Remarkably, Tir ITIM-like motifs and cytoplasmic tyrosine-based sequence motifs of host receptors share the ability to adopt α -helical conformations. Indeed, host ITIMs and Immunoreceptor tyrosine-based activation motifs (ITAMs) can form dynamic/transient α -helices, whether in free [30,31] or membrane bound state [32]. Their dynamic binding to membranes is postulated to regulate tyrosine sites' access to phosphorylation, providing a switch between functional and non-functional conformations [33]. Moreover, the helical conformation in ITIMs/ITAMs has been shown to be stabilized *in vitro* in the presence of 2,2,2- trifluoroethanol (TFE) or through binding to detergent micelles and lipids [31,32,34]. In the presence of membrane-mimetic solvent TFE, we found that C-Tir exhibited a substantial increase in secondary structure, with the CD signature of a partial helical protein (Figure 3.12).

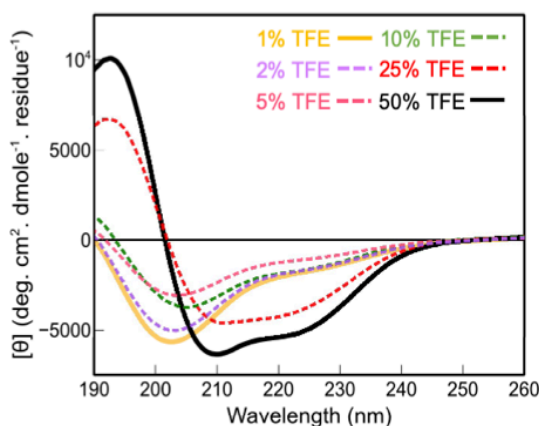


Figure 3.12. Impact of TFE on C-Tir. Far-UV CD spectra of C-Tir in the presence of increasing amounts of TFE, showing C-Tir gradually acquiring structure.

This conformational change is followed by chemical shift differences (Figure 3.13 A-B) and a differential decrease in peak intensities of the C-Tir amide proton resonances in the presence of increasing amounts of TFE,

Results

mainly around Y511 and Y483 and not at S434 (Figure 3.13 B, bottom). This effect supports an enhanced α -helical conformation propensity around the two tyrosine residues, and a putative role in membrane-binding, as postulated for some host tyrosine-based motifs that undergo TFE induced structural changes [31].

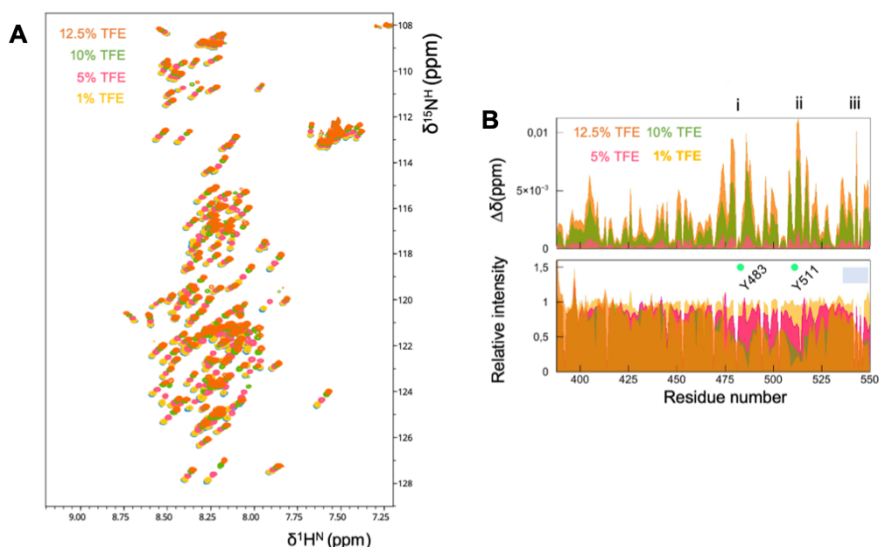


Figure 3.13. (A) NMR [^1H - ^{15}N]-HSQC spectra of C-Tir in the presence of increasing amounts of TFE. Most cross-peaks in the $^1\text{H}^{\text{N}}$ - ^{15}N plane move upfield in both dimensions, even after referencing all chemical shifts to DSS. **(B)** Residue-resolution mapping of the effect of TFE on C-Tir. TFE-induced chemical shift perturbations ($\Delta\delta$, top) as a function of residue number, with 3-regions displaying $\Delta\delta$ higher than the mean plus one standard deviation, including Y483, Y511, and C-Tir C-terminal residues, respectively. TFE-induced NMR attenuation profiles (bottom), i.e., the ratio of peak intensity in the presence and absence of TFE, are plotted along the sequence. Green circles mark the position of Y483 and Y511. The grey bar highlights the C-terminal residues with NMR-signal attenuation.

Thus, our NMR results show that unphosphorylated C-Tir host-like ITIM motifs can adopt pre-formed structures potentially relevant to membrane binding and molecular recognition host of signaling proteins.

Multi-phosphorylated C-Tir is a disordered multivalent cytosolic tail

To fully assess Tir ability to recruit SH2 domains, we reconstructed C-Tir's Tyr-phosphorylated state and evaluated its binding to C-SH2. To this end, we incubated C-Tir with the Src family PTK Fyn, and quantitatively monitored its phosphorylation by NMR [35]. The sensitivity of chemical shifts to changes caused by phosphorylation allowed us to identify four phosphorylated tyrosine sites along the disordered C-Tir modified by Fyn (i.e., pY454, pY474, pY483, pY511). Phosphorylation caused small but noticeable changes in the chemical shifts of the tyrosine residues and adjacent residues. The $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQC spectrum displayed low amide proton dispersion, a diagnostic that the C-Tir remains disordered upon multisite phosphorylation (Figure 3.14).

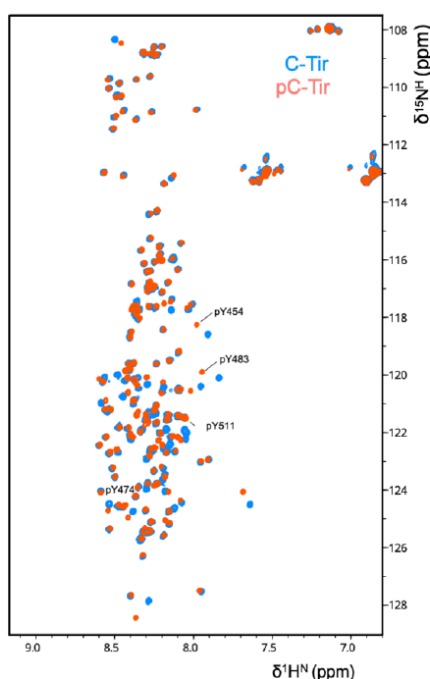


Figure 3.14. Tyrosine phosphorylation of C-Tir. Overlay of $[^1\text{H}\text{-}^{15}\text{N}]$ -HSQC spectra of unphosphorylated (blue) and phosphorylated (orange) C-Tir samples.

Results

Moreover, the secondary chemical shifts of this 4-fold phosphorylated state (pC-Tir) did not reveal substantial changes in local structure propensity due to phosphorylation (Figure 3.15), as similarly reported for other disordered proteins [37,38].

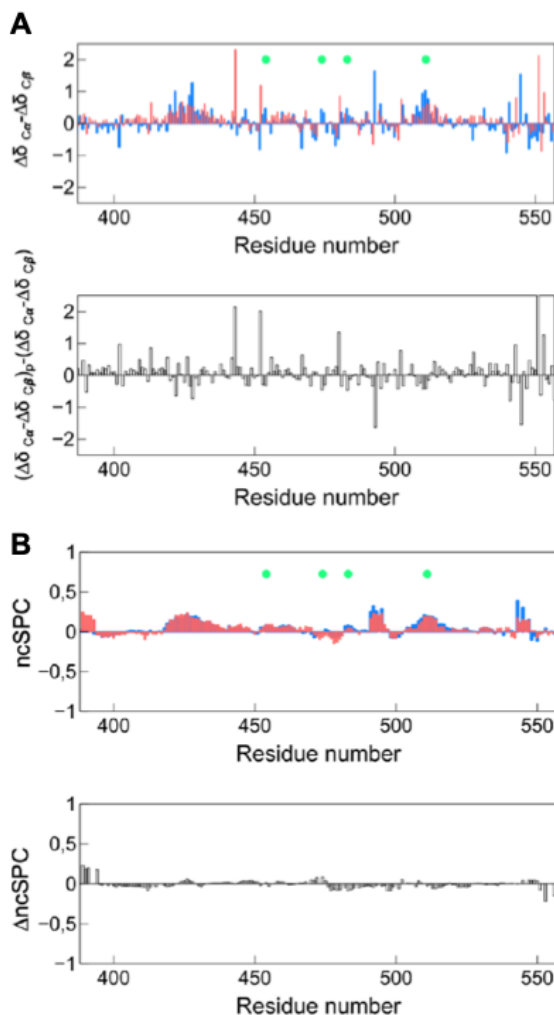


Figure 3.15. Secondary structure propensity of C-Tir and pC-Tir. (A) Secondary chemical shifts ($\Delta\delta_{C\alpha}-\Delta\delta_{C\beta}$) (top) for C-Tir (blue) and pC-Tir (orange). Changes in secondary chemical shift upon phosphorylation (bottom). **(B)** Neighbor-corrected sequence structural propensity score (top) of C-Tir (blue) and pC-Tir (orange). Minor differences in ncSCP were observed upon phosphorylation (bottom). Green circles mark the positions of pY454, pY474, pY483 and pY511.

NMR measurements of quenched phosphorylation reactions were also carried out to further understand the dynamics of individual phosphorylation events [35] (Figure 3.16). By calculating the individual phosphorylation levels for each phospho-tyrosine, we noticed that Y454 is the first tyrosine to be phosphorylated, reaching almost 50% phosphorylation at 4 minutes upon addition of Fyn (Figure 3.16 B).

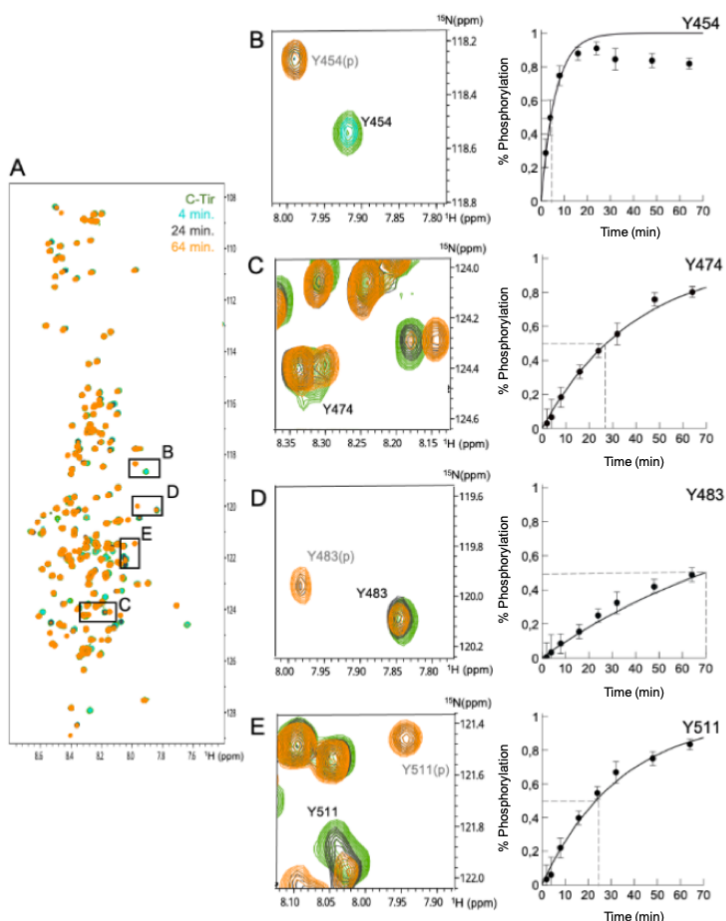


Figure 3.16. C-Tir NMR measurements of quenched phosphorylation reactions. (A) Overlay of $[\text{}^1\text{H}\text{-}^{15}\text{N}]$ -HSQC NMR spectra of non-phosphorylated C-Tir (green) and 4 (blue), 24 (grey) and 64 minutes (orange) upon addition of Fyn. (B – E) Left panels: NMR signal of tyrosine residues (Y454, Y474, Y483 and Y511) at different reaction times (4, 24 and 64 minutes). Right panels: Phosphorylation curves for each individual tyrosine residue.

Results

In contrast, the phosphorylation kinetics of Y483 appears to be the slowest, reaching only ca. 60% phosphorylation at the last time point of reaction, even though the remaining tyrosine residues are 80-90% phosphorylated at the same time point (Figure 3.16 C-E). This way, we can classify the rate of individual tyrosine phosphorylation reactions in the following order: Y454> Y511> Y474> Y483.

Our results suggest that Fyn selectivity and affinity towards the tyrosine residues in C-Tir is affected by its the sequence location, as the individual phosphorylation patterns are distinct, even though all tyrosines are phosphorylated at the same time. In a sequential phosphorylation system, the order in which the phosphorylation sites are activated by the action of the kinase determines the phosphorylation states. However, C-Tir tyrosine phosphorylation does not appear to be a sequential process (i.e., conditional phosphorylation), but rather a preferential event since phosphorylation on the four tyrosine residues occurs simultaneously but at different rate. Nevertheless, further studies are needed to corroborate this observation.

To evaluate the interaction of phosphorylated C-Tir (pC-Tir) with an SH2 domain, we titrated unlabeled C-SH2 domain of SHP-1 into a ^{15}N -labeled pC-Tir solution and monitored the binding interaction broadening of signals in NMR [^1H - ^{15}N]-HSQC spectra (Figure 3.17 A).

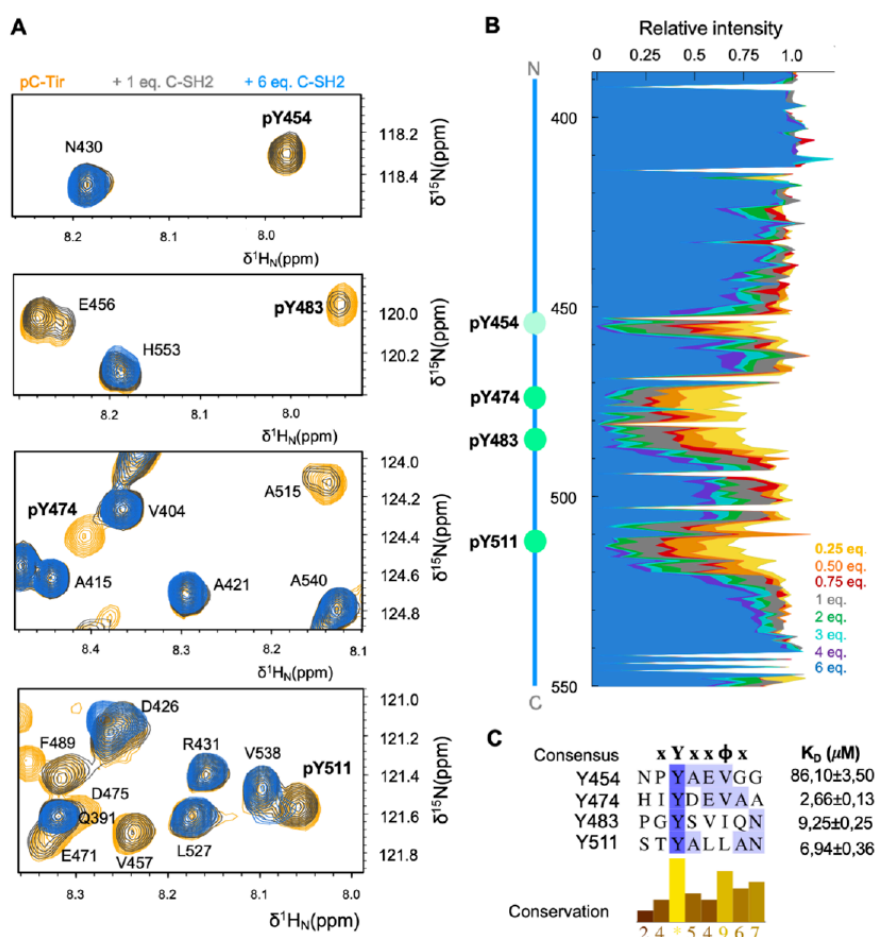


Figure 3.17. C-Tir contains multiple phosphorylated docking sites. (A) NMR-detected binding of C-SH2 to phosphotyrosines pY454, pY474, pY483, pY511 before (orange) and after adding 1.0 and 6.0 equivalents of C-SH2 (blue and grey, respectively). **(B)** Site-specific NMR signal attenuation on pC-Tir due to C-SH2 binding, with local intensity drop around each phosphorylation site. **(C)** Sequence alignment of xYx(x) ϕ motifs of C-Tir interacting with C-SH2 and their respective apparent K_D values obtained by global fitting 1:1 model the intensities drop around each phosphorylation site. Darker shades of blue refer to higher sequence identity. The conservation of physicochemical properties in each alignment position is reported in the corresponding bar-plot below the alignment. Green circles denote phosphorylation tyrosine-sites with color intensity reflecting the apparent relative affinity.

Results

With the resonance re-assignment of pC-Tir, we identified that upon phosphorylation, all tyrosine sites interact with C-SH2, and not exclusively the Y511-based motif as observed for unphosphorylated C-Tir. The interaction caused a drop in peak intensities around each site (i.e., i= pY454, pY474, pY483, pY511), including roughly residues i+4 and i-1, even at sub-stoichiometric conditions (Figure 3.17 B). This suggests a “fuzzy” multivalent binding of C-SH2 to four xYx(x) ϕ motifs, where ϕ is hydrophobic (V/L/I) (Figure 3.17 C), with multiple pYs on C-Tir interacting with C-SH2 in dynamic equilibrium [36,37]. Except for these residues, backbone chemical shift changes were relatively small, with spectra displaying sharp line widths and low dispersion, both characteristics of a disordered protein. Thus, pC-Tir binds C-SH2 retaining its high level of disorder.

As the relative signal intensities at each phospho-tyrosine (pY) and surrounding residues approximately reflect the fraction of unbound sites, we were able to estimate their apparent local binding affinity (Figure 3.18).

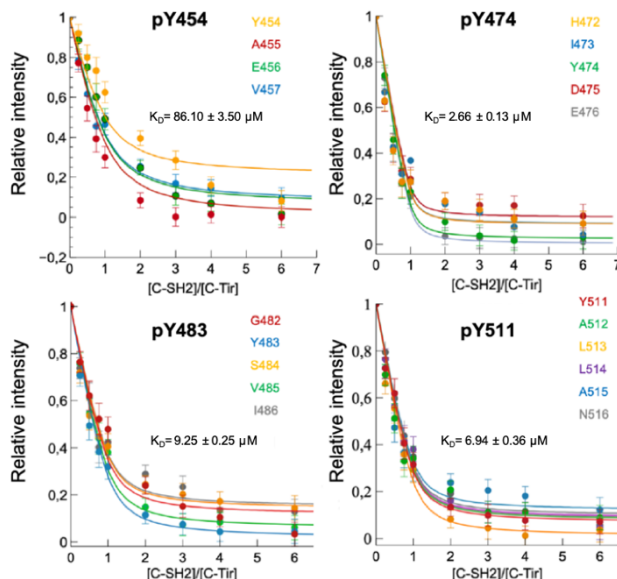


Figure 3.18. NMR quantification of the binding of C-SH2 to individual pY-sites. Relative apparent K_D values were obtained by global fitting a 1:1 model to the relative local intensity drop around each phosphorylation site.

Phosphorylation enhanced the binding to Y511 ITIM by ~10 fold and enabled other tyrosine-based motifs to engage C-SH2. Among the four pYs, the resonances around pY454 were less broadened during the titration, suggesting lower local binding. The remaining pYs bind C-SH2 with similar strength (~3-9 μ M). Overall, phosphorylation at multiple tyrosine sites provides various SH2 docking sites and reinforces the role of Tir as a scaffolding hub.

Disordered C-Tir binds lipids

The ability to form α -helices stabilized by TFE prompts us to further investigate lipid binding by C-Tir. To do so, we incubated ^{15} N-labelled C-Tir with different bicelles, which are small planar bilayers of long-chain lipids closed by curved micelle-like walls of short-chain lipids. These disk-like structures are membrane models extensively used to explore protein-membrane interactions including those involving disordered proteins [38]. As long-chain lipids, we used 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) or the acidic lipid 1,2-dimyristoyl-sn-glycero-3-phosphorylglycerol (DMPG). The long-chain lipids were mixed with the short-chain lipid 1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC), offering the possibility to create bicelles with different lipid charge density.

In the presence of bicelles containing DMPC lipids (Figure 3.19 A-B), we observed a significant residue-specific decrease in the ratio of NMR signals around Y511; yet more subtle, also around Y483, as previously induced with TFE (Figure 3.13 B). This finding shows that C-Tir does interact with non-charged lipid bilayers predominantly via its Y511-motif, including mostly hydrophobic residues, such as A₅₁₂LLA₅₁₅.

Results

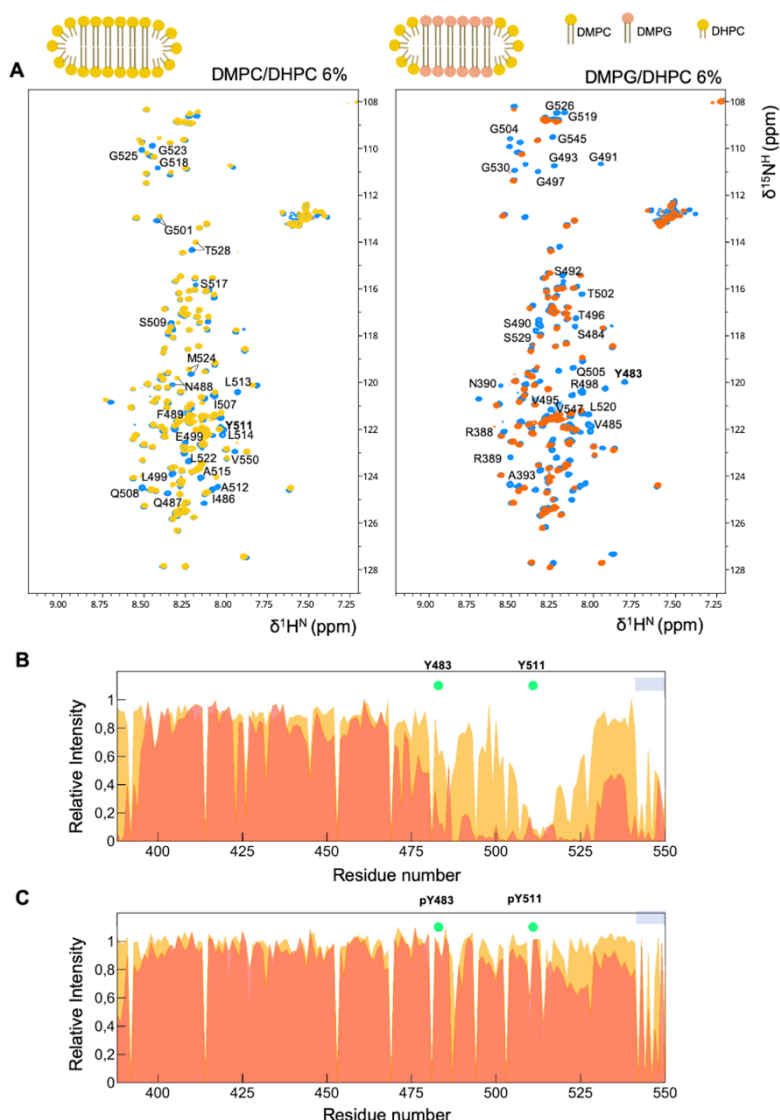


Figure 3.19. Lipid-binding by disordered C-Tir. (A) NMR [^1H - ^{15}N]-HSQC spectra of C-Tir in the absence (blue) and the presence of DMPC/DHPC (6% w/v, yellow) or DMPG/DHPC (6% w/v, orange) bicelles. Residues with NMR signal attenuation upon the addition of DMPC/DHPC bicelles are indicated in the spectra (left). The additional residues affected by DMPG/DHPC are also shown (right). Lipid-induced NMR attenuation profiles, plotted against C-Tir primary sequence, for (B) C-Tir and (C) pC-Tir in the presence of DMPC/DHPC (yellow) and DMPG/DHPC (orange) bicelles. The grey bar highlights the C-terminal residues affected by lipids.

We further tested the influence of increasing negatively charged lipid head groups on C-Tir membrane association with bicelles containing DMPG lipids. Our NMR data show a more extended NMR signal attenuation in the presence of DMPG/DHPC bicelles, even including the first stretch of residues of this construct, which display some positively charged residues (e.g., R388 and K389) (Figure 3.19 A-B). In the context of the full-length protein, these residues define the membrane-proximal region. Given its proximity to the cell plasma membrane, it is reasonable to assume its involvement in membrane binding.

Aside from Y511 surrounding residues, the sequence around Y483 is also notably affected by anionic lipid content, indicating that lipid-charge content appears to modulate C-Tir membrane binding modes. Contrary, the residues 395-475 did not show any membrane interaction. This region has a net negative charge ($pI \sim 4.5$), whereas the sequences displaying membrane ability are positive ($pI \sim 9.9$), thus in line with an electrostatic model. In this residue range, C-Tir bears two experimental confirmed host protein binding sites: a) NCK Src Homology 2 (SH2) domain binding motif [21], and b) NPY motif that interacts with the I-BAR domain of host IRSp53/IRTKS [39].

In the presence of both lipid bicelles, the last C-terminal residues of C-Tir also show NMR-signal attenuation indicative of membrane-binding (Figure 3.19 A-B). The residues A₅₄₁PTPGPVRFV₅₅₀ emerge as part of a potential new lipid-binding region, also affected by TFE (Figure 3.13 B) and showing ncSPC values indicating α -helical propensity (Figure 3.7 A) and pre-existing compaction (Figure 3.8).

Overall, our data thus suggest that C-Tir can undergo multivalent electrostatic interactions with lipid bilayers via pre-structured elements, suggesting that membrane-protein interplay at the intracellular side might control the activity and interactions of Tir in host cells.

We also tested if phosphorylation of Y483 and Y511 could affect lipid-binding. To this end, we mutated Y454 and Y474 to phenylalanine residues and phosphorylated the remaining two tyrosine residues. Next, we incubated

Results

the phosphorylated and ^{15}N -labeled mutant (pC-Tir_{454_F474F}) with both types of bicelles to probe lipid binding using NMR. Our data reveal that the phosphorylation of Y483 and Y511 drastically hampers membrane association of those sites (Figure 3.19 C) while partially unaffected the ability to interact with anion bicelles of the first and last stretch of residues.

Tir-induced actin polymerization and cell death depend on C-Tir lipid-binding

The results presented in this sub-section were obtained by Qiyun Zhong under the supervision of Professor Gad Frankel at the Centre for Molecular Bacteriology and Infection, Imperial College London, under the scope of a collaboration established in this work.

Tir promotes actin polymerization at the bacterial attachment site, leading to pedestal formation [40]; moreover, Tir activates pyroptotic cell death in intestinal epithelial cells [41]. To investigate whether the lipid-binding residues affect the known biological functions of Tir during the infection, we mutated key residues in the C-terminus of Tir (R388, R389, and L513). Plasmid-encoded wild-type and mutant Tir were expressed in EPEC-0, which encodes the T3SS and intimin but is missing all the known effector genes, including *tir* [42]. These strains were then used to infect the model intestinal epithelial cell line SNU-C5 [42].

We first mutated the positively charged residues R388 and R389, as well as L513 located in the Y511 ITIM motif to the negatively charged glutamic acid. Similar to EPEC-0-Tir_{WT}, EPEC-0-Tir_{388E389E} and EPEC-0-Tir_{513E} triggered the formation of actin pedestals (Figure 3.20) and propidium iodide (PI) uptake as a marker of lytic cell death (Figure 3.21).

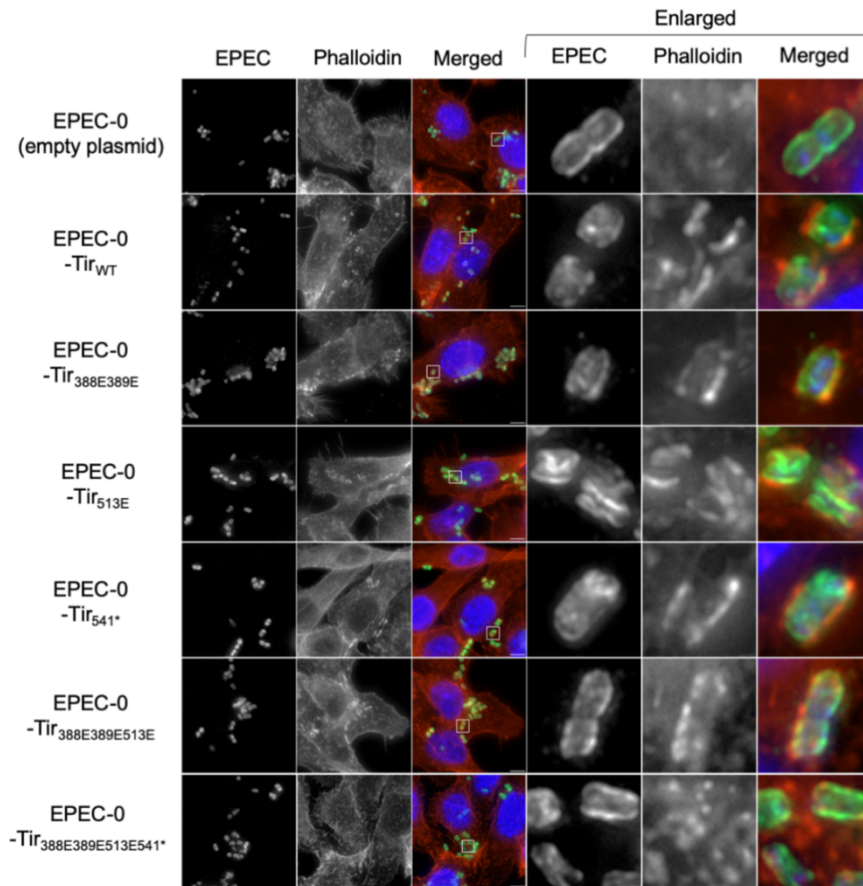


Figure 3.20. Effect of Tir C-terminus mutations on Tir-induced actin polymerization. Immunofluorescence images of SNU-C5 cells infected with EPEC-0 (empty plasmid), EPEC-0-Tir_{WT}, EPEC-0-Tir_{388E389E}, EPEC-0-Tir_{513E}, EPEC-0-Tir_{541*}, EPEC-0-Tir_{388E389E513E} or EPEC-0-Tir_{388E389E513E541*} at 2 h post-infection. Enlarged cell-attached EPEC are seen in the white box (right-hand side panels). Shown are representative images from 3 independent biological repeats. Scale bar: 5 μ m. Blue: DAPI (DNA); Green: EPEC; Red: phalloidin (actin).

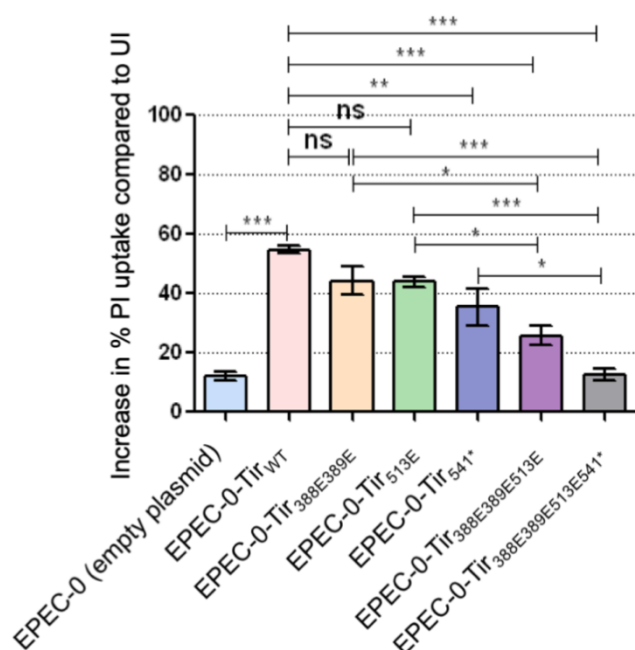


Figure 3.21. Effect of Tir C-terminus mutations on Tir-induced cell death.

PI uptake into SNU-C5 cells infected with EPEC-0 (empty plasmid), EPEC-0-Tir_{WT}, EPEC-0-Tir_{388E389E}, EPEC-0-Tir_{513E}, EPEC-0-Tir_{541*}, EPEC-0-Tir_{388E389E513E}, or EPEC-0-Tir_{388E389E513E541*} at 8 h post-infection. Data shown are mean±SEM from 3 (EPEC-0-Tir_{388E389E513E541*}) or 6 (all other strains) independent biological repeats. Statistical analysis was performed using 1-way ANOVA with Tukey's post-test. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ns: non-significant.

We then introduced a STOP codon at residue A541, removing the ten residues of the C-terminus tail (A₅₄₁PTGPGVRFV₅₅₀), which are also implicated in lipid-binding, creating EPEC-0-Tir_{541*}. While pedestal formation was unaffected in EPEC-0-Tir_{541*}-infected cells (Figure 3.20), we detected a mild yet significant reduction in PI uptake compared to EPEC-0-Tir_{WT} (Figure 3.21) that became more evident over time (Supplementary Figure S1.14). Accordingly, mutations in all three sites individually did not affect pedestal formation and only resulted in a limited reduction in Tir-induced cell death.

Next, we generated a triple R388E / R389E / L513E mutant. The resulting strain EPEC-0-Tir_{388E389E513E} triggered typical actin pedestals (Figure 3.20); however, we observed significantly reduced cell death compared to both EPEC-0-Tir_{388E389E} and EPEC-0-Tir_{513E} (Figure. 3.21). Finally, we mutated all four sites simultaneously, including point mutations and the C-terminus ten residue truncation, creating EPEC-0-Tir_{388E389E513E541*}. EPEC-0-Tir_{388E389E513E541*} was impaired in actin pedestal formation (Figure 3.20) and was not cytotoxic, similar to the negative control EPEC-0 with an empty plasmid (Figure 3.21).

Therefore, while the lipid-binding sites at the C-terminus of Tir appear to be essential for its biological functions, they might be partly redundant as only the quadruple mutant affected Tir-induced cell death and pedestal formation.

Into the molecular basis of host TLR signaling suppression by the bacterial effector Tir

Unphosphorylated C-Tir exploits differential SHP-1/SHP-2 binding

EPEC Tir has been shown to directly associate with the host phosphatases SHP-1 and SHP-2 [27, 29]. Although these interactions have been proposed to rely on tyrosine phosphorylation of Tir's ITIM-like sequences by host Src family kinases, we have previously provided experimental evidence that C-Tir is able to interact with the C-SH2 domain of SHP-1 regardless of being phosphorylated. Indeed, without phosphorylation, C-Tir can recognize the C-SH2 of SHP-1 through a single ITIM-like sequence comprising Y511 (Figure 3.22 A).

Here, we assessed whether C-Tir also recognizes SHP-2 SH2 domains pre-phosphorylation. To do so, we have titrated unlabeled samples of the C-terminal SH2 (C-SH2) and tandem SH2 (NC-SH2) domains of SHP-2 against uniformly ^{15}N -labeled sample of C-Tir and monitored the interactions using NMR.

With increasing amounts of SHP-2 C-SH2 domain, we observed a tiny decrease in the ratio of NMR signals around Y511 (Figure 3.22 B), far less accentuated than the observed for the homologous domain of SHP-1 (Figure 3.22 A). We estimated an apparent K_D of ca. $190.41\ \mu\text{M}$ for this interaction (Supplementary Figure SI.15), about ~3-fold weaker than the same observed for the C-SH2 domain of SHP-1. We also detected a not significant loss of intensity of NMR signals in the residues around Y483, not alluding to a tight binding either ($K_D = 129.89 \pm 3.15\ \mu\text{M}$) (Supplementary Figure SI.15). Overall, these observations highlight a differential feature for SHP-2 binding by Tir when compared to SHP-1.

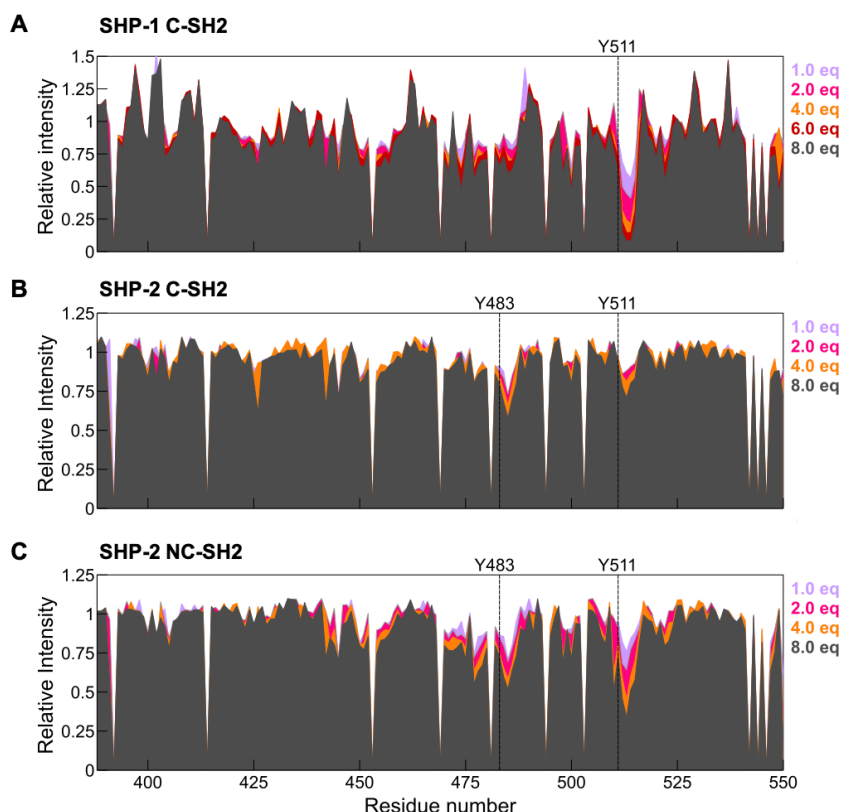


Figure 3.22. C-Tir can bind host SH2 domains of SHP-1 and SHP-2 prior to phosphorylation. Relative [^1H - ^{15}N]-peak intensities from the titrations of C-Tir with (A) C-SH2 domain of SHP-1, and (B) C-SH2 and (C) NC-SH2 domains of SHP-2. Dashed lines mark the positions of Y483 and Y511.

In the presence of the tandem NC-SH2 of SHP-2, comprising both N- and C-SH2 domains, we have noticed a slightly more pronounced selective loss of intensities in the regions surrounding Y483 and Y511 (Figure 3.22 C). This can be related to the presence of a second SH2 domain, instead of a single one, reinforcing binding. Nevertheless, the association to the tandem NC-SH2 of SHP-2 is weaker than the C-SH2 SHP-1 interaction pre-phosphorylation. Interestingly, in this scenario, the region comprising Y511 is now being more affected ($K_D = 95.86 \pm 3.07 \mu\text{M}$), alluding that the C-SH2 domain alone has more affinity to the ITIM-like sequence comprising Y483 ($K_D = 114.30 \pm 5.81 \mu\text{M}$) (Supplementary Figure SI.16). Moreover, it suggests a

Results

preferential binding through the Y511 surrounding residues, likely via the N-SH2.

Overall, our results reinforce that the sequences surrounding Y483 and Y511 play an important role in defining pre-phosphorylation motifs for SH2-domain binding, and shows that C-Tir encodes differential features to bind SHP-1 and SHP-2 host phosphatases prior to phosphorylation.

Phosphorylated C-Tir interacts with SHP-2 forming fuzzy complexes

In the host, the long-term Tir signaling, mediated by the recruitment SH2-domain containing phosphatases, which dephosphorylate target proteins to restrict further downstream signaling pathways, is maintained by the action of Src-family kinase c-Fyn and other protein kinases that ensure a persistent tyrosine phosphorylated state [9, 43, 44].

There so, we appraised the ability of C-Tir to recruit the SH2 domains of SHP-2 upon *in vitro* phosphorylation by Fyn human kinase, using NMR spectroscopy, as previously performed for the C-SH2 domain of SHP-1. To further understand the determinants underlying the interactions between phosphorylated C-Tir and SHP-2, we resort to multiple truncated versions of SHP-2 including the N-SH2, C-SH2 and NC-SH2 domains.

By monitoring the interaction through broadening of the NMR signals in [¹H-¹⁵N]-HSQC spectra acquired with increasing amounts of SH2 domains we started by noticing that pC-Tir engages transient interactions with all domains of SHP-2. As observed for SHP-1, Fyn-mediated phosphorylation promotes C-Tir binding to N-SH2, C-SH2, and NC-SH2 of SHP-2, with all constructs engaging pC-Tir at all phosphorylated sites (Figure 3.23).

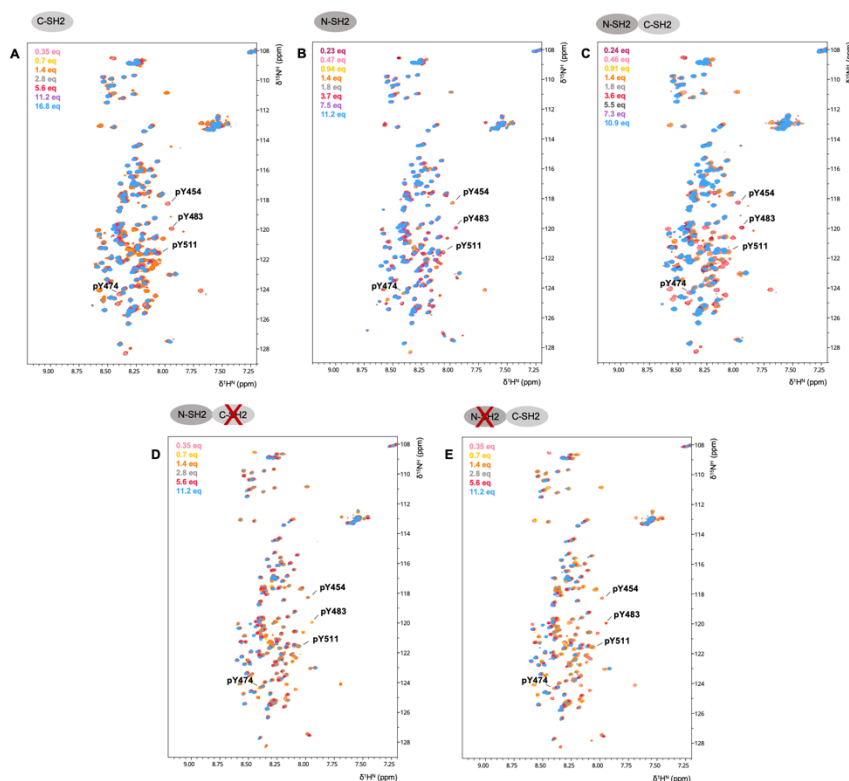


Figure 3.23. Overview of NMR titrations to study the interaction of pC-Tir with SH2 domains of SHP-2. Overlay of $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of pC-Tir in the presence of (A) C-SH2, (B) N-SH2, (C) NC-SH2, (D) NC-SH2 C-dead, (E) NC-SH2 N-dead domains of SHP-2.

The transient interactions of pC-Tir, via several adjacently positioned short linear motifs (SLiMs), with the different folded domains of SHP-2, are accompanied by retained disorder of pC-Tir (Figure 3.23), indicating the presence of fuzzy complexes [45].

To gain insights into the mechanism underlying Tir-mediated activation of SHP-2, we estimated the local binding affinities of the four binding sites in pC-Tir for all the individual and tandem SH2 domains of SHP-2. As peaks adjacent to each phosphorylation site exhibited a similar drop in signal intensity, we clustered them by site and globally fitted a 1:1 binding model (Supplementary Figures SI.17-SI.21).

Results

Interaction with the C-SH2 domain revealed a similar binding affinity of ca. 0.3 μM for all phosphotyrosine residues in C-Tir ITIM-like sequences (Figure 3.24). We also noticed an overall increased binding, compared to the C-SH2 domain of SHP-1, particularly for the Y454-motif, reinforcing that C-Tir may carry differential features to bind SHP-1 and SHP-2 host phosphatases or vice-versa. Although SHP-1 and SHP-2 share significant overall sequence identity (55%; Supplementary Figure SI.22) and are regulated in identical ways, they often perform opposite functions. Indeed, albeit both SHP-1 and SHP-2 are considered negative regulators of TLR signaling, SHP-2 often acts as a positive signal transducer of cell signaling [46].

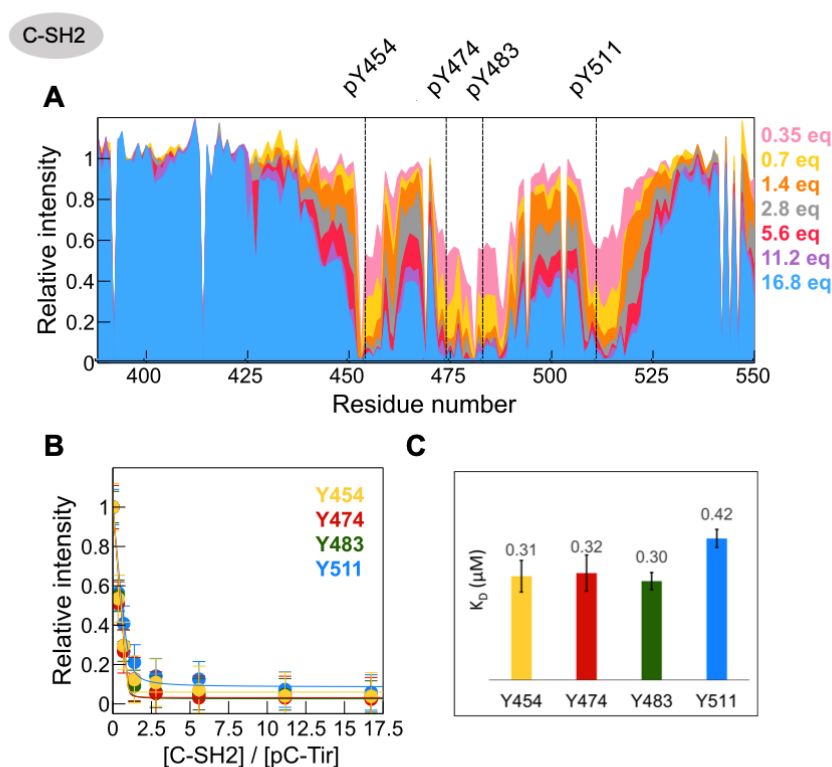


Figure 3.24. Interaction of pC-Tir with the C-SH2 domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of pY454, pY474, pY483 and pY511 is marked by dashed lines. **(B)** Global fitting curves and **(C)** histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

Binding to N-SH2 domain is still promiscuous but with an overall lower affinity than the C-SH2 domain (Figure 3.25).

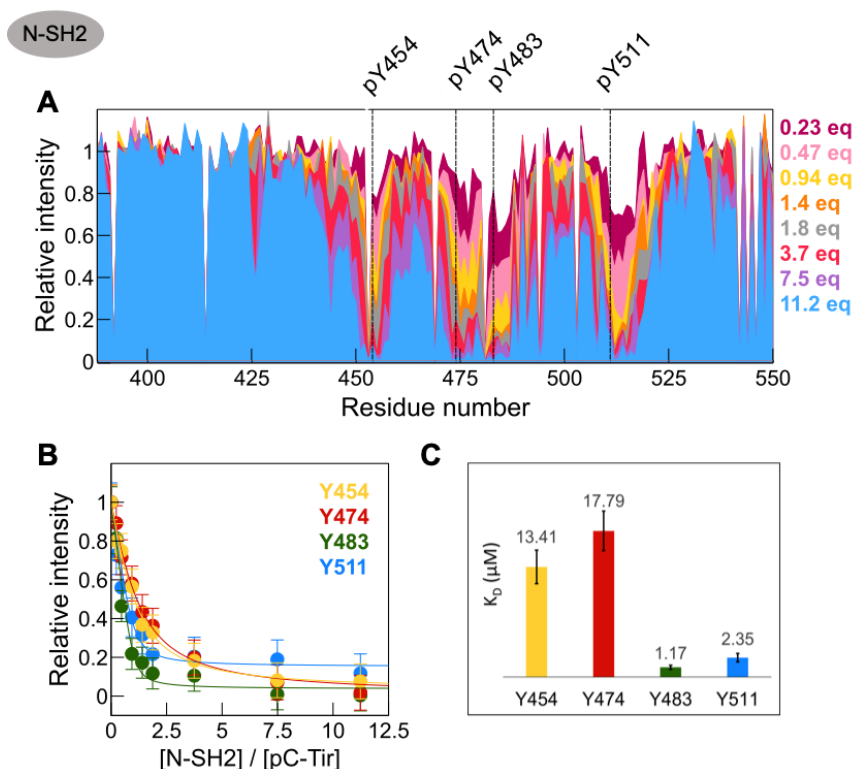


Figure 3.25. Interaction of pC-Tir with the N-SH2 domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of pY454, pY474, pY483 and pY511 is marked by dashed lines. (B) Global fitting curves and (C) histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

Furthermore, a higher binding affinity for Y483 ($K_D = 1.17 \pm 0.26 \mu\text{M}$) and Y511 ($K_D = 2.35 \pm 0.52 \mu\text{M}$) and a lower affinity for Y454 ($K_D = 13.41 \pm 2.04 \mu\text{M}$) and Y474 ($K_D = 17.79 \pm 2.40 \mu\text{M}$) were observed, revealing a binding preference for the ITIM-like sequences comprising Y483 and Y511.

When both SH2 domains are present (NC-SH2), we noticed an overall affinity enhancement, one to two orders of magnitude higher when compared to the single N-SH2. All tyrosine-containing motifs in pC-Tir appear to bind

Results

NC-SH2 with similar strength (0.1-0.3 μM), with a slight but not substantial apparent preference for Y454 and Y483 (Figure 3.26).

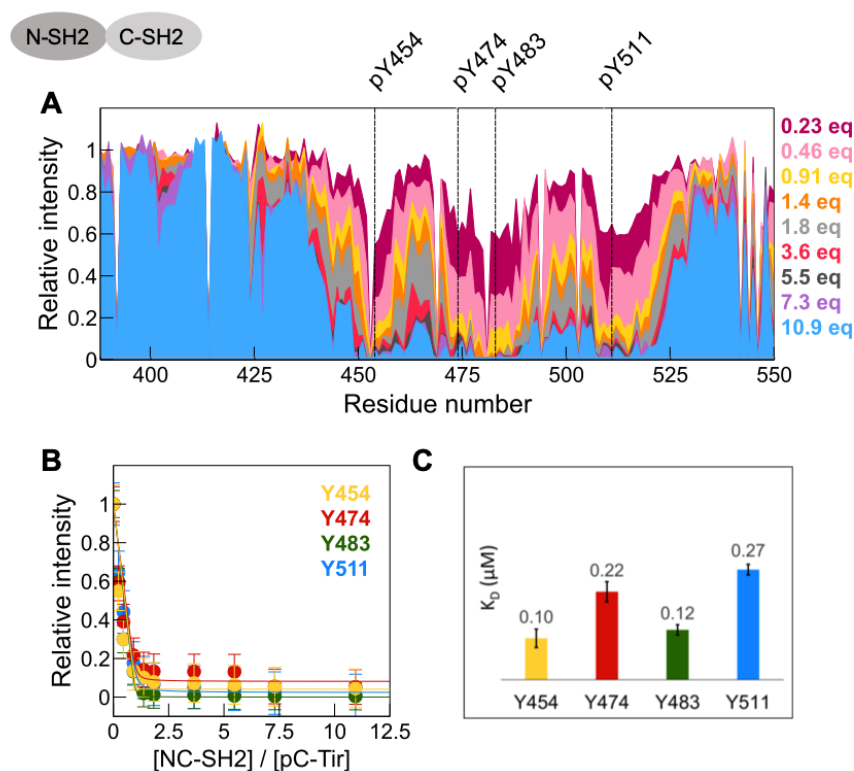


Figure 3.26. Interaction of pC-Tir with the NC-SH2 domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of pY454, pY474, pY483 and pY511 is marked by dashed lines. **(B)** Global fitting curves and **(C)** histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

Phosphorylated C-Tir prefers to recruit SHP-2 in an anti-parallel fashion

Interactions of phosphorylated C-Tir with NC-SH2 mutants in which one of the two SH2 domains was functionally inactivated (designated N-dead and C-dead) were also carried out to understand the contribution of individual SH2 domains in Tir-mediated activation of SHP-2. Inactivation of the C-SH2

domain in NC-SH2 C-dead mutant decreased binding affinity by one to two orders of magnitude compared the NC-SH2. Notably, it resulted in a favored binding to the Y511 containing motif (Figure 3.27). The presence of an extra domain doesn't abolish the observed promiscuous binding but improves the relative specificity of N-SH2 towards Y511.

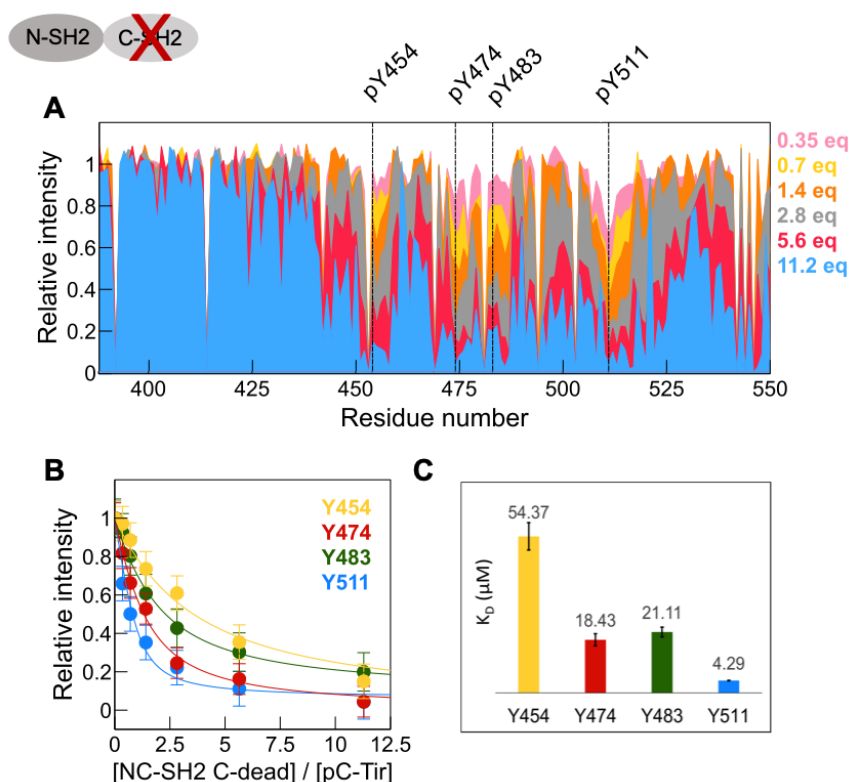


Figure 3.27. Interaction of pC-Tir with the NC-SH2 C-dead domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of pY454, pY474, pY483 and pY511 is marked by dashed lines. (B) Global fitting curves and (C) histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

Together with the results obtained with the single N-SH2 domain, it appears that the N-SH2 domain preferentially binds to pC-Tir by the ITIM-like sequence comprising Y511.

Results

In its turn, when N-SH2 was functionally inactivated (Figure 3.28), we noticed a decrease in binding strength of about 1-fold for Y454 and Y474 ($K_D = 1.85 \pm 0.31$ and $1.09 \pm 0.11 \mu\text{M}$, respectively) but binding of the same order of magnitude for Y483 and Y511 ($K_D = 0.29 \pm 0.02$ and $0.47 \pm 0.05 \mu\text{M}$, respectively), compared to the NC-SH2 domain.

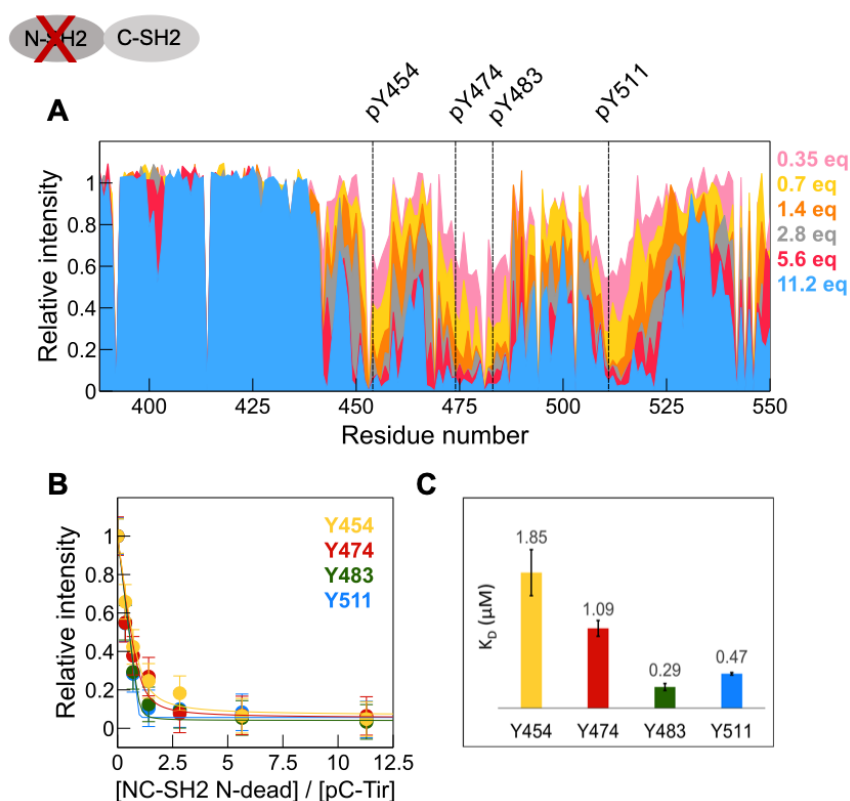


Figure 3.28. Interaction of pC-Tir with the NC-SH2 N-dead domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of pY454, pY474, pY483 and pY511 is marked by dashed lines. (B) Global fitting curves and (C) histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

Although we observed identical binding for all phosphotyrosine residues upon interaction with the single C-SH2 domain, our results suggest

that, in the context of the tandem, C-SH2 preferentially interacts with pC-Tir by the motif containing Y483.

Overall, our results support that the large flexibility and conformational heterogeneity of C-Tir allows multiple transient and dynamic interactions. Moreover, it highlights the relevance of Tir ITIM-like sequences to exploit host subversion, as phosphorylation of these sequences appears to function as a regulatory mechanism to enable and enhance multivalent binding to multiple SH2 domains. Our interaction studies with SHP-2 suggest that the ITIM-like sequences comprising Y483 and Y511 are preferential for binding, and likely through an anti-parallel association in which N-SH2 and C-SH2 bind to pY511 and pY483, respectively (Figure 3.29). Moreover, yet binding to pY454 and pY474 also occurs, protein hopping and binding promiscuity appear to be reduced by an additional SH2 domain. As a whole, this emphasize that Tir act as a signaling platform and function as a docking hub by establishing multivalent interactions with distinct and diverse host targets.

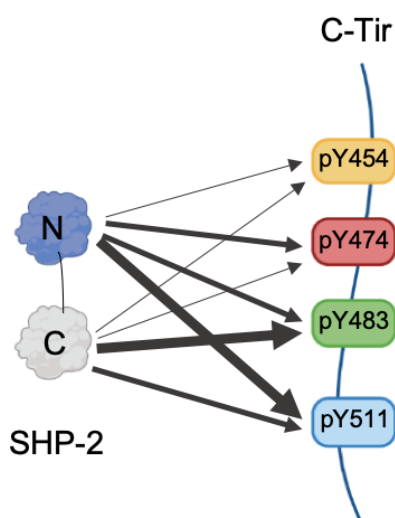


Figure 3.29. Cartoon representation depicting the relative binding affinities of SHP-2 individual SH2 to C-Tir's ITIM-like sequences. The thickness of the arrows matches the relative apparent affinities calculated from the NMR data.

C-Tir can recruit TRAF6 without binding to host phosphatases

In the host, Tir enhances the recruitment of host phosphatases SHP-1 and SHP-2 to TRAF6, acting as a negative regulator of Toll-Like Receptor (TLR) signaling pathways. In TLR-signaling, TRAF6 functions as an adaptor and E3 ubiquitin ligase to activate NF- κ B-dependent pro-inflammatory cytokines expression [47-49]. Consecutively, direct interaction between SHP-1/2 and TRAF6 has been shown to prevent the ubiquitination of TRAF6, inhibiting the activation of NF- κ B and MAP kinases and consequently suppressing the expression of several genes and pro-inflammatory cytokines involved in inflammatory responses [29].

The mechanism by which Tir promotes the association between SHP-1/2 and TRAF6 remains unclear. Still, in our search to shed light on this mechanism, we have found that C-Tir can directly interact with the C-terminal domain of TRAF6 (TRAF6-C) without prior binding to host phosphatases.

By sequence analysis using the ELM database [50], we found that C-Tir bears a TRAF6 binding motif [i.e., Pro-X-Glu-X-X-(Ar/Ac), where Ar and Ac correspond to aromatic and acidic aminoacids, respectively]. This well-characterized ELM is involved in TRAF6-C recognition by host receptors [51] (Figure 3.30), often engaged in immune response signaling pathways. One example is the tumor necrosis factor receptor superfamily member 5 (TNFRSF5, also designated as CD40) that interacts with TRAF6, triggering the initiation of signal transduction and resulting in the expression of inflammatory mediators [52]. The binding of the receptor activator of nuclear factor-kappaB (RANK) to TRAF6 has been reported to be required for the proper formation of cytoskeletal structures and functional activity of osteoclasts in bone resorption [53]. The mitochondrial antiviral-signaling protein (MAVS) has been shown to recruit TRAF6 to activate an antiviral response [54]. Association of TRAF6 with ubiquitinated interleukin-1 receptor-associated kinase-1 (IRAK-1) has been shown to have a negative regulatory action on TLR signaling as it leads to proteasome degradation of TRAF6 [55].

Identifying this conserved TRAF6-binding motif often present in host proteins involved in immune responses prompted us to inspect the Tir sequences from EPEC, EHEC and CR. Notably, from sequence alignment, we have found out that both EPEC and CR Tir sequences possess a putative binding site for TRAF6-C (Figure 3.30).

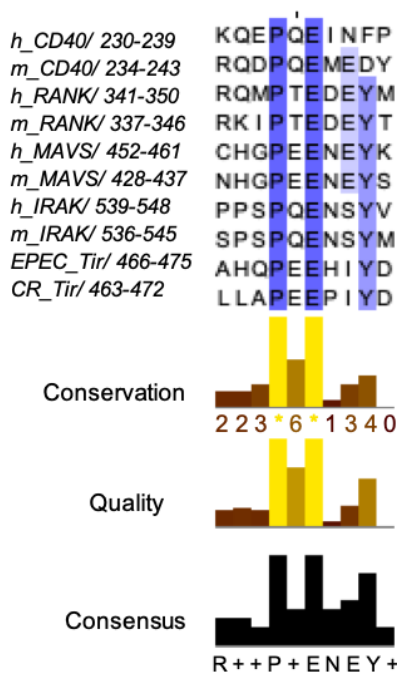


Figure 3.30. C-Tir possess a putative site for TRAF6-C domain binding. Multiple sequence alignment of Pro-X-Glu-X-X-(Ar/Ac) motifs in CD40, RANK, MAVS, IRAK-1 and Tir proteins. *h* and *m* stand for human and mouse, respectively. Alignment positions conserved in all sequences are shaded in lavender. The conservation of physicochemical properties in each alignment position is reported in the corresponding bar-plot below the alignment.

In EPEC and CR Tir, the TRAF6-binding motif Pro-Glu-Glu-His-Ile-Tyr (residues 469-474 and 466-471, respectively) is conserved suggesting a similar binding mechanism for both organisms. Interestingly, the TRAF6-binding motif is not present in EHEC Tir, proposing that in EHEC the recruitment of TRAF6-C may occur by an alternative mechanism. The

Results

differences between the mechanisms used to recruit TRAF6-C may be related to the larger sequence divergency between EPEC and EHEC C-terminus of Tir (41 % of sequence identity) [56].

The simultaneous presence of ITIMs and a TRAF6-binding motif within C-Tir sparked the working hypothesis that C-Tir works as a molecular scaffold tethering together TRAF6 and SHPs host proteins. With that in mind, we decided to validate whether C-Tir can bind TRAF6 directly. NMR spectroscopy was then employed to access the interaction between C-Tir and TRAF6-C domain at the residue level. To this end, a uniformly ^{15}N -labeled C-Tir sample was titrated with increasing molar equivalents of unlabeled TRAF6-C. First, we noticed that low amide proton dispersion was retained in the $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra, acquired for each titration point (Figure 3.31), indicating that C-Tir retains its disorder upon binding to TRAF6-C, as verified for the interactions with SHP-1 and SHP-2.

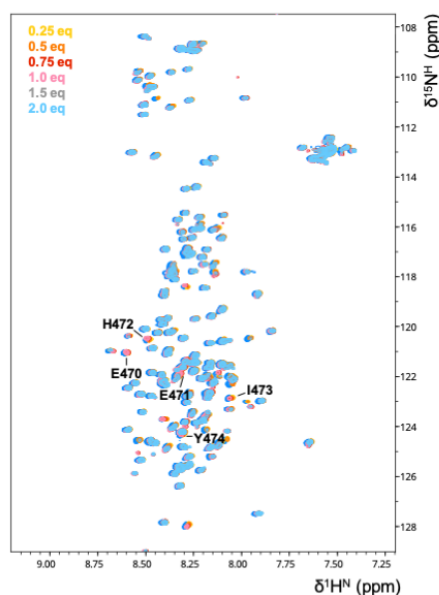


Figure 3.31. Overlay of $[\text{}^1\text{H}-^{15}\text{N}]$ -HSQC spectra of ^{15}N -C-Tir in the presence of 0.25 (yellow), 0.5 (orange), 0.75 (red), 1.0 (pink), 1.5 (grey) and 2.0 (light blue) molar equivalents of TRAF6-C domain.

Moreover, upon binding, we observed a selective loss of intensities for the HSQC peaks of C-Tir spectra, mainly from residues E₄₇₀EHIY₄₇₄ (Figure 3.32 A), where the predicted binding site to TRAF6-C is located.

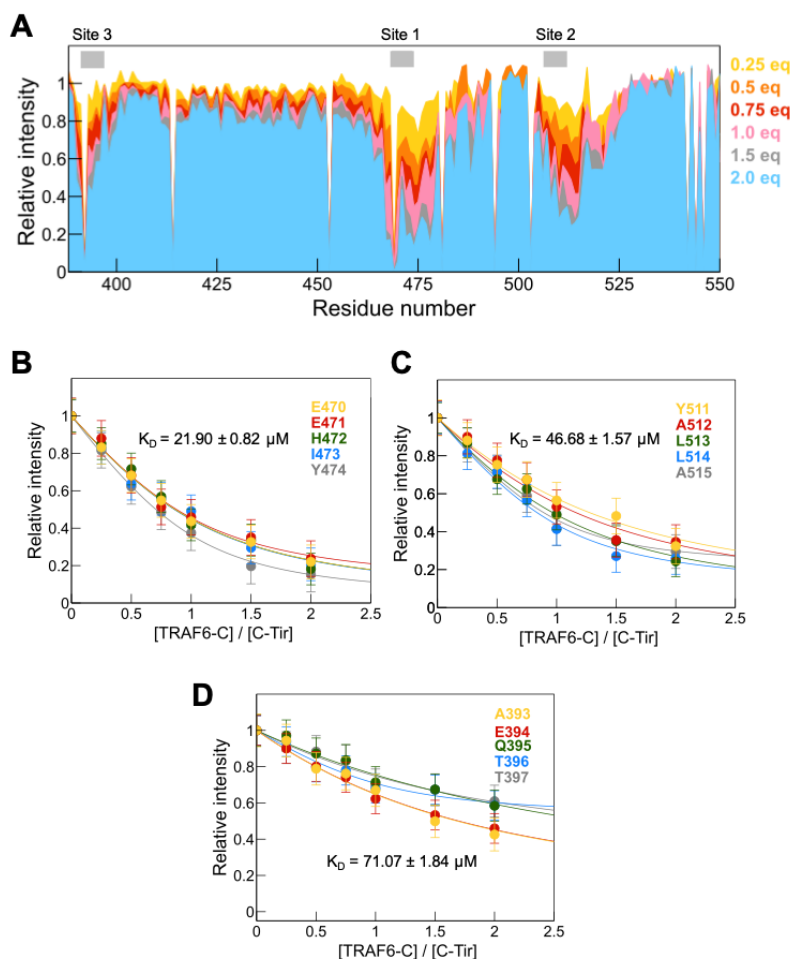


Figure 3.32. C-Tir can directly recruit TRAF6-C. (A) Relative [¹H-¹⁵N]-peak intensities from the titration analysis. The grey boxes mark the position of the three binding sites. (B), (C) and (D) NMR quantification of the binding of TRAF6-C domain to C-Tir regions 469-474, 511-515 and 393-397 (sites 1,2 and 3 in panel A), respectively. Relative apparent K_D values were obtained by global fitting a 1:1 model to the relative local intensity drop around each binding site.

Results

The residues in the binding motif were clustered and globally fitted into a 1:1 binding model equation, allowing us to estimate an apparent K_D of ca. 21.9 μM for this interaction (Figure 3.32 B). Our results confirm that C-Tir can directly recruit TRAF6, unraveling a previously unrecognized target for Tir.

Surprisingly enough, in addition to the region encompassing the binding motif, we also observed a subtler attenuation of NMR signals near Y511 (Figure 3.32 A). We estimated an apparent K_D of ca. 46.7 μM for secondary binding site (Figure 3.32 C). As previously shown, Y511 engages binding to SH2 domains through a $\text{xYx(x)\phi}$ motif, where ϕ is a hydrophobic (V/L/I) residue. Furthermore, the residues preceding Y511 (Gly-Ile-Gln-Ser-Thr) do not meet the criteria defined for TRAF6-binding motif (Pro-X-Glu-X-X-(Ar/Ac)). That being said, the observed decreased intensity of NMR signals in the region surrounding Y511 may arise from the primary interaction of C-Tir with TRAF6-C on the predicted binding motif, or could represent an additional and distinct binding mode of C-Tir to TRAF6-C. Recall that MTSL at position 509 (Figure 3.8) induced broadening in the region 438-479, consistent with compact conformations involving both sites. The binding to the TRAF6-binding motif may alter the observed low-range interaction, manifested in chemical shift perturbations.

Also, subtle but noticeable is the drop in intensity observed around residues 392-397 of C-Tir (Figure 3.32 A). Notably, this region (Pro-Ala-Glu-Gln-Thr-Thr) is in agreement with the defined TRAF6-binding motif features, suggesting a third binding site for TRAF6-C with an estimated apparent K_D of ca. 71.1 μM (Figure 3.32 D). In light of this, a construct of TRAF6 (residues 333-508) containing part of the coiled-coil domain and the whole TRAF6-C domain produced soluble protein that existed predominantly as trimers in solution [51]. Nevertheless, future work will be needed to reconcile and reinforce these observations.

To further detail this interaction, we employed AlphaFold2-multimer [57] to predict the binding interface between TRAF6-C and C-Tir. This extension of AlphaFold2 [58] can work with multiple chains and be trained to drive the structure of protein complexes of known stoichiometry. From NMR, we

identified two to three potential binding sites on C-Tir with differential affinity. With that in mind, we tested whether AlphaFold2 could find those sites within the C-Tir sequence. The AlphaFold models displayed a structural undefined C-Tir with general low structure confidence (low pLDDT) as expected for a disordered region. Still, three areas emerged with a slightly higher score for structure and in contact with the folded TRAF6-C domain (Supplementary Figures SI.23), coinciding with those observed by NMR to be engaged or affected by binding. There so, we then launched three independent structure predictions for each binding site within C-Tir identified by NMR (Supplementary Table SI.3). AlphaFold2 computed five models for each prediction, and the top-ranked model (rank 1) obtained for each complex are displayed in Figure 3.33.

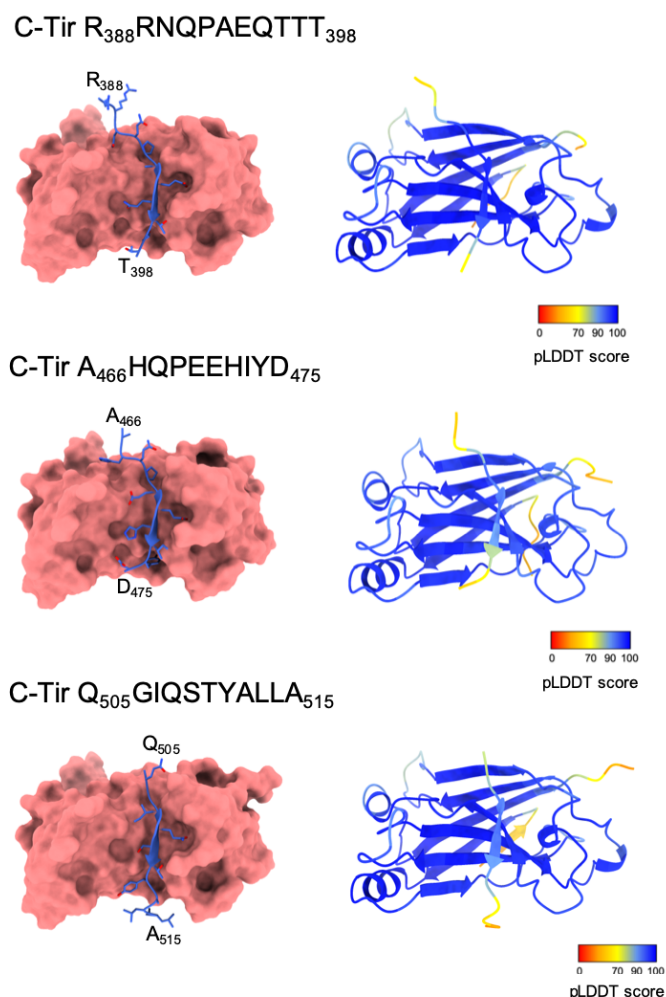


Figure 3.33. Predicted complexes of C-Tir peptides and TRAF6-C binding interface obtained with AlphaFold2. Models are represented by cartoon representation (left) and by pLDDT values per residue (right). The peptide sequences of C-Tir employed for each prediction (R₃₈₈RNQPAEQTTT₃₉₈, A₄₆₆HQPEEHIYD₄₇₅, and Q₅₀₅GIQSTYALLA₅₁₅) are described above each model. In the cartoon representations (left), the C-Tir peptides and TRAF6-C are depicted in blue and salmon, respectively. In the representation by pLDDT values per residue (right), residues are colored according to the pLDDT scale depicted below each model.

In all predicted complexes, the short peptide regions of C-Tir (R₃₈₈-T₃₉₆, A₄₆₆-D₄₇₅ or Q₅₀₅-A₅₁₅) bind the same groove on the TRAF6-C surface. The bound peptides and folded host domain have high certainty standards (pLDDT values per residue) (Figure 3.33, right), smaller inter-complex predicted alignment errors (PAE), and plausible contacts reminiscent of the complex established between TRAF6-C and CD40 (Supplementary Figure SI.24).

Overall, our findings are a milestone towards the structural understanding of how Tir hijacks host TLR signaling, by revealing that C-Tir can directly interact with TRAF6-C, opening up the possibility that Tir may act as a scaffold to bring together TRAF6 and SHP-1/2 to enhance suppression of inflammatory responses. Even so, a next future step will be to probe this interaction with SAXS.

The C-Tir tail harbors an overlapping phospho-regulated TRAF6 binding motif

To further perceive the determinants underlying the interaction between Tir and TRAF6 we studied the effect of C-Tir *in vitro* phosphorylation on binding to TRAF6-C domain. As observed for unphosphorylated C-Tir, upon phosphorylation C-Tir binds TRAF6-C while remaining disorder (Supplementary Figure SI.25). Unexpectedly, attenuation of NMR signals in residues E₄₇₀EHLY₄₇₄, defining the binding motif to TRAF6-C, was observed, although at a lesser extent than the observed in the absence of phosphorylation (Figure 3.34 A).

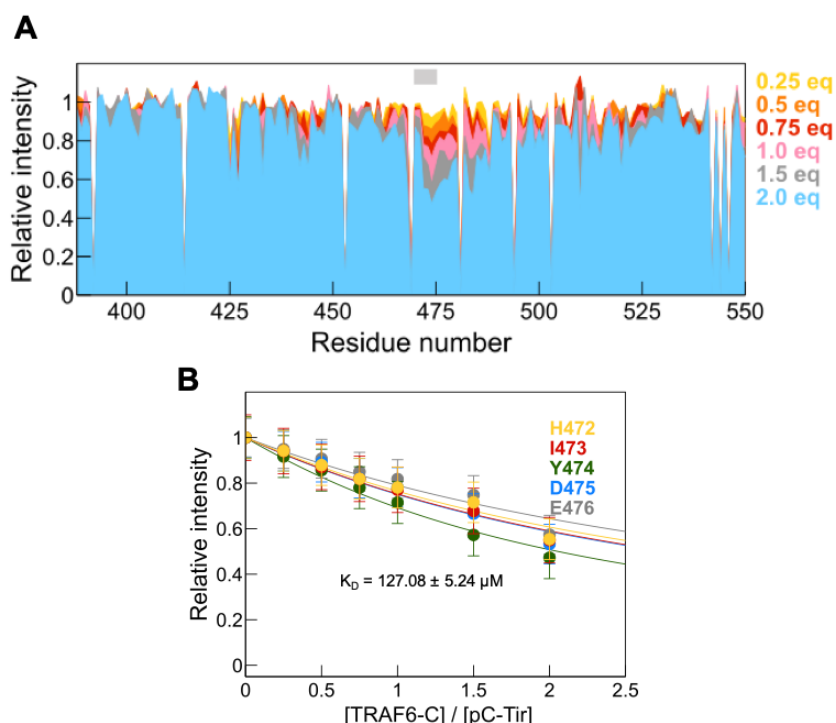


Figure 3.34. Binding to TRAF6-C is hindered upon C-Tir phosphorylation by host kinases. (A) Relative [^1H - ^{15}N]-peak intensities from the titration analysis. The grey box marks the position of Pro-Glu-Glu-His-Ile-Tyr binding motif (residues 469-474). **(B)** NMR quantification of the binding of TRAF6-C domain to pC-Tir. Relative apparent K_D value was obtained by global fitting a 1:1 model to the relative local intensity drop around the binding site.

Phosphorylation of C-Tir also rendered a reduced binding affinity for TRAF6-C, with an estimated apparent K_D of ca. 127.1 μM for this interaction (Figure 3.34 B), one order less of magnitude than the one observed for the interaction of C-Tir and TRAF6-C in an unphosphorylated setting. Moreover, the regions comprising the other binding sites appear to be no longer perturbed upon phosphorylation. Hence, our results show that phosphorylation of C-Tir restricts the interaction with TRAF6-C domain, decreasing the binding strength, and suggest that Y474 may have a pivotal role to enable the interaction with TRAF6 via the Pro-Glu-Glu-His-Ile-Tyr motif.

Interestingly, not only Y474 appears to be relevant for TRAF6-binding, as it has been shown to be essential for focal actin assembly and proper pedestal formation through interaction with the host-cell adaptor protein Nck [43,59,60]. As phosphorylation of Y474 promotes the interaction of Tir with the SH2 domain of Nck [61], it could act as a phospho-switch mechanism to allow the SH2-binding to Y474 to trigger the formation of actin pedestals.

Tyrosine 474 appears to be crucial for C-Tir interaction with TRAF6-C

Upon dissecting the determinants underlying the binary interactions of Tir-SHP-2 and Tir TRAF6-C, we aimed to assemble and deliver the molecular architecture of the ternary complex.

To do so, we resort to NMR, and acquired [^1H - ^{15}N]-HSQC spectra of uniformly ^{15}N -labeled phosphorylated C-Tir samples, where Y454 and Y474 had been previously replaced by phenylalanine residues, in the absence and presence of 1.0 molar equivalents of the NC-SH2 domain of SHP-2 or TRAF6-C. Taking into account the observed binding promiscuity of C-Tir for the SH2 domains of SHP-2, we used the double mutant to restrict the binding of SHP-2 to the Y483- and Y511-containing ITIM-like motifs upon phosphorylation, and maintain the residues 470-474 available for binding to TRAF6-C.

As expected, replacement of Y454 and Y474 by phenylalanine residues drastically restrained binding of these regions to the tandem SH2 and enhanced binding of Y483 ($K_D = 0.06 \pm 0.002 \mu\text{M}$) and Y511 ($K_D = 0.13 \pm 0.04 \mu\text{M}$) (Figure 3.35; Supplementary Figure SI.26).

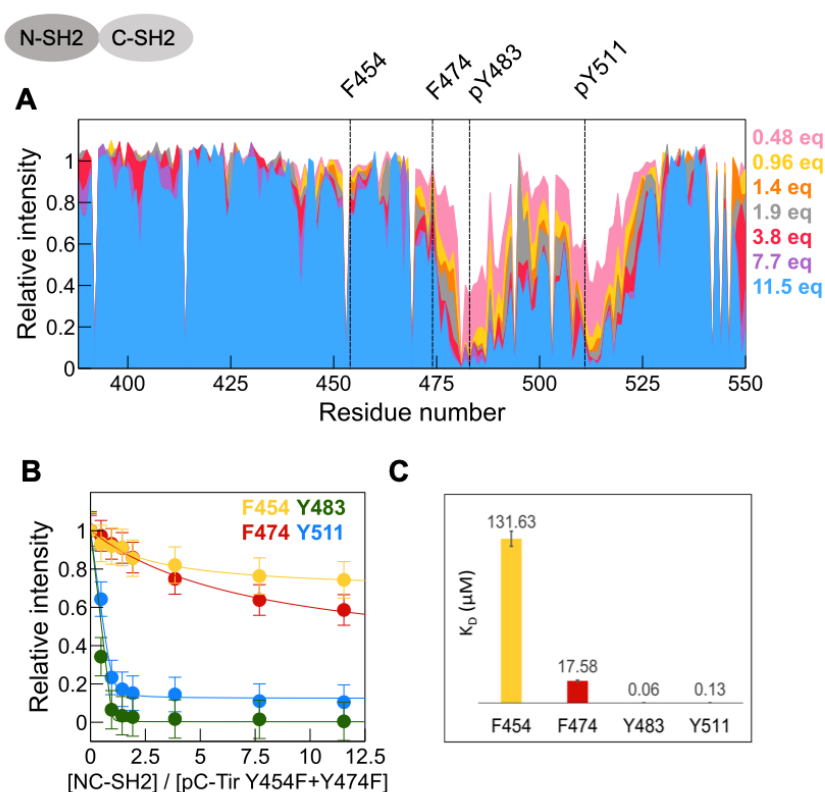


Figure 3.35. Interaction of pC-Tir_Y454F+Y474F with NC-SH2 domain of SHP-2. (A) Site-specific NMR attenuation profiles, where the position of F454, F474, pY483 and pY511 is marked by dashed lines. (B) Global fitting curves and (C) histogram containing the estimated K_D values for the different ITIM-like sequences of pC-Tir.

However, in the interaction between the phosphorylated mutant of C-Tir and TRAF6-C we surprisingly noticed an unexpected effect. Apart from some relatively small changes in backbone chemical shifts (Figure 3.36 A), no residue-specific attenuation in the ratio of NMR signals around residues 470-474 was registered (Figure 3.36 B), indicating that the replacement of Y474 by a phenylalanine residue abolishes the interaction of C-Tir with TRAF6-C and suggesting a specificity for a tyrosine in this position. Nevertheless, the

interaction of TRAF6-C with the unphosphorylated C-Tir mutant must be studied to validate this observation.

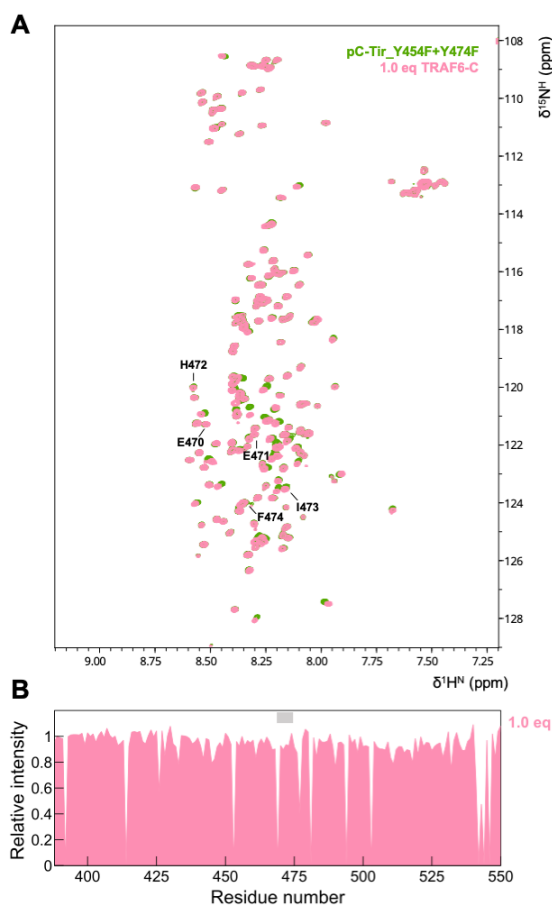


Figure 3.36. C-Tir Y474 is required for interaction with TRAF6-C. (A) NMR [^1H - ^{15}N]-HSQC spectra of pC-Tir_Y454F+Y474F in the absence (green) and presence of 1.0 molar equivalent of TRAF6-C domain (pink). (B) NMR attenuation profile plotted against C-Tir primary sequence.

In order to deliver the assembly of the ternary complex, in the future, Y474 should be replaced by a glutamate residue, expecting to retain the motif and at the same time, avoid binding of SHP-2.

References

- [1] Jones, D.T., Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. **31**, 857-63 (2015).
- [2] Dosztányi, Z., Csizmek, V., Tompa, P., Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. **21**, 3433-4 (2005).
- [3] Hamada, D., Hamaguchi, M., Suzuki, K.N., Sakata, I., Yanagihara, I. Cytoskeleton-modulating effectors of enteropathogenic and enterohemorrhagic *Escherichia coli*: a case for EspB as an intrinsically less-ordered effector. *FEBS J.* **277**, 2409-15 (2010).
- [4] Grishin, A.M., Cherney, M., Anderson, D.H., Phanse, S., Babu, M., Cygler, M. NleH defines a new family of bacterial effector kinases. *Structure*. **22**, 250-9 (2014).
- [5] Holmes, A., Mühlen, S., Roe, A.J., Dean, P. The EspF effector, a bacterial pathogen's Swiss army knife. *Infect Immun*. **78**, 4445-53 (2010).
- [6] Sallee, N.A., Rivera, G.M., Dueber, J.E., Vasilescu, D., Mullins, R.D., Mayer, B.J., Lim, W.A. The pathogen protein EspF(U) hijacks actin polymerization using mimicry and multivalency. *Nature*. **454**, 1005-8 (2008).
- [7] Hagai, T., Azia, A., Babu, M.M., Andino, R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.* **7**, 1729-1739 (2014).
- [8] Via, A., Uyar, B., Brun, C., Zanzoni, A. How pathogens use linear motifs to perturb host cell networks. *Trends Biochem Sci.* **40**, 36-48 (2015).
- [9] Bommarius, B., Maxwell, D., Swimm, A., Leung, S., Corbett, A., Bornmann, W., Kalman, D. Enteropathogenic *Escherichia coli* Tir is an SH2/3 ligand that recruits and activates tyrosine kinases required for pedestal formation. *Mol Microbiol.* **63**, 1748-68 (2007).
- [10] Blasche, S., Arens, S., Ceol, A., Siszler, G., Schmidt, M.A., Häuser, R., Schwarz, F., Wuchty, S., Aloy, P., Uetz, P., Stradal, T., Koegl, M. The EHEC-host interactome reveals novel targets for the translocated intimin receptor. *Sci Rep.* **4**, 7531 (2014).
- [11] Van Roey, K., Uyar, B., Weatheritt, R.J., Dinkel, H., Seiler, M., Budd, A., Gibson, T.J., Davey, N.E. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* **114**, 6733-78 (2014).
- [12] Tompa, P., Davey, N.E., Gibson, T.J., Babu, M.M. A million peptide motifs for the molecular biologist. *Mol Cell.* **55**, 161-9 (2014).
- [13] Popelka, H., Uversky, V.N., Klionsky, D.J. Identification of Atg3 as an intrinsically disordered polypeptide yields insights into the molecular dynamics of autophagy-related proteins in yeast. *Autophagy*. **10**, 1093-104 (2014).
- [14] Danis, C., Despres, C., Bessa, L.M., Malki, I., Merzougui, H., Huvent, I., Qi, H., Lippens, G., Cantrelle, F.X., Schneider, R., Hanouille, X., Smet-Nocca, C., Landrieu, I. Nuclear Magnetic Resonance Spectroscopy for the Identification of

Multiple Phosphorylations of Intrinsically Disordered Proteins. *J Vis Exp.* **118**, 55001 (2016).

[15] Pesce, G., Gondelaud, F., Ptchelkine, D., Nilsson, J.F., Bignon, C., Cartalas, J., Fourquet, P., Longhi, S. Experimental Evidence of Intrinsic Disorder and Amyloid Formation by the *Henipavirus* W Proteins. *Int J Mol Sci.* **23**, 923 (2022).

[16] Bernadó, P., Blackledge, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J.* **97**, 2839-45 (2009).

[17] Cordeiro, T.N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., Bernadó, P. Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. *Adv Exp Med Biol.* **1009**, 107-129 (2017).

[18] Chemes, L.B., Alonso, L.G., Noval, M.G., de Prat-Gay, G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol Biol.* **895**, 387-404 (2012).

[19] Delaforge, E., Cordeiro, T.N., Bernadó, P., Sibille, N. "Conformational characterization of intrinsically disordered proteins and its biological significance" in *Modern Magnetic Resonance*, G.A. Webb, Ed. (Springer International Publishing, 2018), 381–399.

[20] Garmendia, J., Phillips, A.D., Carlier, M.F., Chong, Y., Schüller, S., Marches, O., Dahan, S., Oswald, E., Shaw, R.K., Knutton, S., Frankel, G. TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol.* **6**, 1167-83 (2004).

[21] Gruenheid, S., DeVinney, R., Bladt, F., Goosney, D., Gelkop, S., Gish, G.D., Pawson, T., Finlay, B.B. Enteropathogenic *E. coli* Tir binds Nck to initiate actin pedestal formation in host cells. *Nat Cell Biol.* **3**, 856-9 (2001).

[22] Prasanna, C., Dubey, A., Atreya, H.S. Amino acid selective unlabeled in protein NMR spectroscopy. *Methods Enzymol.* **565**, 167-89 (2015).

[23] Tamiola, K., Acar, B., Mulder, F.A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc.* **132**, 18000-3 (2010).

[24] Fuxreiter, M., Simon, I., Friedrich, P., Tompa, P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol.* **338**, 1015-26 (2004).

[25] De Biasio, A., Ibáñez de Opakua, A., Cordeiro, T.N., Villate, M., Merino, N., Sibille, N., Lelli, M., Diercks, T., Bernadó, P., Blanco, F.J. p15PAF is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophys J.* **106**, 865-74 (2014).

[26] Backert, S., Kenny, B., Gerhard, R., Tegtmeyer, N., Brandt, S. PKA-mediated phosphorylation of EPEC-Tir at serine residues 434 and 463: A novel pathway in regulating Rac1 GTPase function. *Gut Microbes.* **1**, 94-99 (2010).

[27] Yan, D., Wang, X., Luo, L., Cao, X., Ge, B. Inhibition of TLR signaling by a bacterial protein containing immunoreceptor tyrosine-based inhibitory motifs. *Nat Immunol.* **13**, 1063-71 (2012).

Results

- [28] Tamiola, K., Mulder, F.A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem Soc Trans.* **40**, 1014-20 (2012).
- [29] Yan, D., Quan, H., Wang, L., Liu, F., Liu, H., Chen, J., Cao, X., Ge, B. Enteropathogenic *Escherichia coli* Tir recruits cellular SHP-2 through ITIM motifs to suppress host immune response. *Cell Signal.* **25**, 1887-94 (2013).
- [30] Rosenl w, J., Isaksson, L., Mayzel, M., Lengqvist, J., Orekhov, V.Y. Tyrosine phosphorylation within the intrinsically disordered cytosolic domains of the B-cell receptor: an NMR-based structural analysis. *PLoS One.* **9**, e96199 (2014).
- [31] Cheng, H., Schwell, V., Curtis, B.R., Fazlieva, R., Roder, H., Campbell, K.S. Conformational Changes in the Cytoplasmic Region of KIR3DL1 upon Interaction with SHP-2. *Structure.* **27**, 639-650 (2019).
- [32] Xu, C., Gagnon, E., Call, M.E., Schnell, J.R., Schwieters, C.D., Carman, C.V., Chou, J.J., Wucherpfennig, K.W. Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif. *Cell.* **135**, 702-13 (2008).
- [33] Xu, X., Li, H., Xu, C. Structural understanding of T cell receptor triggering. *Cell Mol Immunol.* **17**, 193-202 (2020).
- [34] Duchardt, E., Sigalov, A.B., Aivazian, D., Stern, L.J., Schwalbe, H. Structure induction of the T-cell receptor zeta-chain upon lipid binding investigated by NMR spectroscopy. *ChemBiochem.* **8**, 820-7 (2007).
- [35] Theillet, F.X., Rose, H.M., Liokatis, S., Binolfi, A., Thongwichian, R., Stuiver, M., Selenko, P. Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nat Protoc.* **8**, 1416-32 (2013).
- [36] Mittag, T., Orlicky, S., Choy, W.Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M., Forman-Kay, J.D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A.* **105**, 17772-7 (2008).
- [37] Turner, A.L., Watson, M., Wilkins, O.G., Cato, L., Travers, A., Thomas, J.O., Stott, K. Highly disordered histone H1-DNA model complexes and their condensates. *Proc Natl Acad Sci U S A.* **115**, 11964-11969 (2018).
- [38] P rez, Y., Maffei, M., Igea, A., Amata, I., Gair , M., Nebreda, A.R., Bernad , P., Pons, M. Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci Rep.* **3**, 1295 (2013).
- [39] de Groot, J.C., Schl ter, K., Carius, Y., Quedenau, C., Vingadassalom, D., Faix, J., Weiss, S.M., Reichelt, J., Standfuss-Gabisch, C., Lesser, C.F., Leong, J.M., Heinz, D.W., B ssow, K., Stradal, T.E. Structural basis for complex formation between human IRSp53 and the translocated intimin receptor Tir of enterohemorrhagic *E. coli*. *Structure.* **19**, 1294-306 (2011).
- [40] Kalman, D., Weiner, O.D., Goosney, D.L., Sedat, J.W., Finlay, B.B., Abo, A., Bishop, J.M. Enteropathogenic *E. coli* acts through WASP and Arp2/3 complex to form actin pedestals. *Nat Cell Biol.* **1**, 389-91 (1999).

[41] Zhong, Q., Roumeliotis, T.I., Kozik, Z., Cepeda-Molero, M., Fernández, L.Á., Shenoy, A.R., Bakal, C., Frankel, G., Choudhary, J.S. Clustering of Tir during enteropathogenic *E. coli* infection triggers calcium influx-dependent pyroptosis in intestinal epithelial cells. *PLoS Biol.* **18**, e3000986 (2020).

[42] Cepeda-Molero, M., Berger, C.N., Walsham, A.D.S., Ellis, S.J., Wemyss-Holden, S., Schüller, S., Frankel, G., Fernández, L.Á. Attaching and effacing (A/E) lesion formation by enteropathogenic *E. coli* on human intestinal mucosa is dependent on non-LEE effectors. *PLoS Pathog.* **13**, e1006706 (2017).

[43] Kenny, B. Phosphorylation of tyrosine 474 of the enteropathogenic *Escherichia coli* (EPEC) Tir receptor molecule is essential for actin nucleating activity and is preceded by additional host modifications. *Mol Microbiol.* **31**, 1229-1241 (1999).

[44] Swimm, A., Bommarius, B., Li, Y., Cheng, D., Reeves, P., Sherman, M., Veach, D., Bornmann, W., Kalman, D. Enteropathogenic *Escherichia coli* use redundant tyrosine kinases to form actin pedestals. *Mol Biol Cell.* **15**, 3520-9 (2004).

[45] Tompa, P., Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci.* **33**, 2-8 (2008).

[46] Xu, F., Xu, M.J., Zhao, R., Guerrah, A., Zeng, F., Zhao, Z.J. Tyrosine phosphatases SHP-1 and SHP-2 are associated with distinct tyrosine-phosphorylated proteins. *Exp Cell Res.* **272**, 75-83 (2002).

[47] Wu, H., Arron, J.R. TRAF6, a molecular bridge spanning adaptive immunity, innate immunity and osteoimmunology. *Bioessays.* **25**, 1096-105 (2003).

[48] Bhoj, V.G., Chen, Z.J. Ubiquitylation in innate and adaptive immunity. *Nature.* **458**, 430-7 (2009).

[49] Kondo, T., Kawai, T., Akira, S. Dissecting negative regulation of Toll-like receptor signaling. *Trends Immunol.* **33**, 449-58 (2012).

[50] Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N.E., Chemes, L.B., Gibson, T.J. ELM-the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* **48** (D1), D296-D306 (2020).

[51] Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaya, M., Yim, M., Du, K., Singh, S., Pike, J.W., Darnay, B.G., Choi, Y., Wu, H. Distinct molecular mechanism for initiating TRAF6 signalling. *Nature.* **418**, 443-7 (2002).

[52] Engel, D., Seijkens, T., Poggi, M., Sanati, M., Thevissen, L., Beckers, L., Wijnands, E., Lievens, D., Lutgens, E. The immunobiology of CD154-CD40-TRAF interactions in atherosclerosis. *Semin Immunol.* **21**, 308-12 (2009).

[53] Armstrong, A.P., Tometsko, M.E., Glaccum, M., Sutherland, C.L., Cosman, D., Dougall, W.C. A RANK/TRAF6-dependent signal transduction pathway is essential for osteoclast cytoskeletal organization and resorptive function. *J Biol Chem.* **277**, 44347-56 (2002).

[54] Shi, Z., Zhang, Z., Zhang, Z., Wang, Y., Li, C., Wang, X., He, F., Sun, L., Jiao, S., Shi, W., Zhou, Z. Structural Insights into mitochondrial antiviral signaling

Results

protein (MAVS)-tumor necrosis factor receptor-associated factor 6 (TRAF6) signaling. *J Biol Chem.* **290**, 26811-20 (2015).

[55] Muroi, M., Tanamoto, K. IRAK-1-mediated negative regulation of Toll-like receptor signaling through proteasome-dependent downregulation of TRAF6. *Biochim Biophys Acta.* **1823**, 255-63 (2012).

[56] DeVinney, R., Stein, M., Reinscheid, D., Abe, A., Ruschkowski, S., Finlay, B.B. Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. *Infect Immun* **67**, 2389-2398 (1999).

[57] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv 2021.10.04.463034; doi.org/10.1101/2021.10.04.463034

[58] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**, 583-589 (2021).

[59] Gruenheid, S., DeVinney, R., Bladt, F., Goosney, D., Gelkop, S., Gish, G.D., Pawson, T., Finlay, B.B. Enteropathogenic *E. coli* Tir binds Nck to initiate actin pedestal formation in host cells. *Nat Cell Biol* **3**, 856-859 (2001).

[60] Campellone, K.G., Giese, A., Tipper, D.J., Leong, J.M. A tyrosine-phosphorylated 12-amino-acid sequence of enteropathogenic *Escherichia coli* Tir binds the host adaptor protein Nck and is required for Nck localization to actin pedestals. *Mol Microbiol* **43**, 1227-1241 (2002).

[61] Campellone, K.G., Rankin, S., Pawson, T., Kirschner, M.W., Tipper, D.J., Leong, J.M. Clustering of Nck by a 12-residue Tir phosphopeptide is sufficient to trigger localized actin assembly. *J Cell Biol.* **164**, 407-16 (2004).

Chapter 4

Discussion

Discussion

Intrinsically disordered proteins and regions (IDPs and IDRs, respectively) are widespread across all kingdoms of life [1]. They do not adopt well-defined structures but exist as dynamic conformational ensembles that display unique properties complementary to globular proteins [2,3].

As disordered proteins make up 30-50% of eukaryotic proteomes [4,5], protein disorder is dominant in protein hubs controlling the complex intracellular networks that influence physiological responses. IDPs/IDRs play essential roles in signaling pathways, including cell cycle [6], circadian circuits [7], post-transcriptional regulation [8], and protein degradation [9]. Given their ubiquitous relevance in cellular signaling, the onset of several human diseases, including cancer and neurodegeneration, are linked to dysfunctional disordered proteins [10].

In contrast, they are remarkably less abundant in prokaryotes. However, increasing experimental evidence in the literature shows that several effector proteins secreted by pathogenic bacteria contain functionally relevant IDRs [11]. Well-known cases include the oncogenic *Helicobacter pylori* effector CagA [12], and the effector CyaA of the causative agent of whooping cough *Bordetella pertussis* [13]. Other less-ordered effectors are also secreted by Enteropathogenic and Enterohemorrhagic *E. coli* strains (EPEC and EHEC, respectively), such as EspF(U)/TccP [14] and EspB [15].

In this work, we sought to identify additional disordered prokaryotic effectors, focusing our attention on A/E pathogens. This family of pathogens, including EPEC and EHEC strains and CR mucosal bacterium, is known to attach to the gut mucosa, causing a histopathological condition known as attaching and effacing (A/E) lesions characterized by actin-rich pedestal formation and microvilli degeneration.

Among the enteropathogenic *E. coli* strains, EPEC O127:H6 (E2348/69) is commonly used as a prototype strain to study EPEC biology, genetics, and virulence [16]. Here, we have determined the structural disorder propensity of EPEC O127:H6 sequences and two additional representatives of A/E bacteria: EHEC O157:H7 and CR ICC168. We also searched for

Discussion

eukaryotic linear motifs (ELMs), which are highly occurrent in IDPs/IDRs [17]. ELMs are stretches of 3-10 contiguous residues mediating transient protein-protein interactions (e.g., post-translational modifications sites, targeting signals) [18]. The prevalence of such functional elements in IDRs provides versatility to cell interaction hubs by enabling flexibility and adaptability to multiple interaction interfaces [19,20].

The three studied prototype strains of A/E pathogens have a structurally diverse repertoire of protein effectors enriched in disorder-prone residues. Some effectors are fully disordered proteins with a high density of host-like short linear motifs, challenging the general trend that prokaryotic proteins are less disordered.

Among such disordered effectors emerged the prominent translocated intimin receptor (Tir), with a relatively high density of host-like ELMs, particularly in its intracellular C-terminal domain. Tir is a 56 kDa cell surface receptor able to reshape host cellular behavior during A/E bacterial infections.

During infection, Tir is the first effector secreted [21]. Once inside the host, it migrates to the plasma membrane to act as a receptor for intimin presented at the bacterial surface [22], anchoring A/E pathogens to targeted host cells. Moreover, Tir promotes actin polymerization [23], suppresses autophagy [24] and immune response [25] to allow bacterial survival on the host cell's surface. It also triggers pyroptotic cell death [26,27]. As a double-pass membrane receptor, Tir adopts a hairpin topology with the external intimin-binding domain (IBD) [28] connecting two transmembrane domains, with both N- and C-terminus (N-Tir and C-Tir, respectively) in the host cytoplasm. Those intracellular regions enable Tir to interact with at least 25 host intracellular targets [29], such as host tyrosine phosphatases SHP-1/2 to suppress pro-inflammatory cytokines signaling [25,30] and TAK1-mediated immune responses [31].

The mechanisms employed by Tir to hijack intracellular signaling by interacting with multiple host proteins to assist infection are poorly understood. Thus, to gain insights into Tir's detailed biophysics and biological function, we

devised a structural biophysical study to assess its intrinsic disorder and interactions with host proteins. We focused on stoichiometry, accessibility to multi-site phosphorylation, host protein- and membrane-binding of Tir C-terminal intracellular region.

In the present work, we examined the structure of Tir from EPEC O127:H6, finding that its intracellular C-terminal domain (C-Tir) is flexible with IDR features similar to those found in the cytoplasmic domains of host transmembrane proteins. A combination of biophysical methods (SAXS, NMR, and CD spectroscopies) demonstrated that C-Tir lacks a well-defined 3D-structure and has characteristics of an IDR. Further NMR analysis showed that C-Tir is not exclusively disordered but can adopt partial secondary structural elements and non-random long-range contacts around phosphorylation sites and residues involved in biomolecular interactions (i.e., host protein and lipid binding). This observation supports the idea of transient pre-structured motifs (PreSMOs) as potential pre-existing signatures for binding and function within target-free Tir, and reinforces the role of these local secondary structural elements as the critical determinants for target binding, as increasingly observed for numerous IDPs/IDRs over the last decades [32]. However, we cannot rule out that C-Tir also undergoes structural rearrangement upon binding via induced fit [33].

As in disordered cytosolic tails of host receptors, C-Tir displays tyrosine-based motifs, which are phosphorylated by Src family PTKs enabling the interaction with SH2-domain containing proteins. Interestingly, in C-Tir, we found that two of those motifs, involving Y483 and Y511, bind SH2 domains of SHP-1 and SHP-2 host phosphatases with low affinity prior to phosphorylation (Figure 4.1). NMR shows that the residues adjacent to Y511 define a binding site for the C-SH2 domain of SHP-1, most likely serving as a precursor for a tighter binding upon phosphorylation. NMR also revealed that this same binding site to the C-SH2 domain of SHP-1 is significantly less pronounced for homologous domain of SHP-2, suggesting that differential features are employed by Tir to recruit SH2 domains of SHP-1 and SHP-2 prior to phosphorylation. Moreover, we observed a predisposition in C-Tir to

Discussion

bind the C-SH2 domain of SHP-2 through the residues surrounding Y483, which became more evident with the tandem SH2 domain, suggesting the presence of an additional binding site to SH2 domain-containing proteins pre-phosphorylation.

Another notable feature of these motifs is their α -helical propensity, particularly observed for the region comprising Y511. Although we cannot tell whether a helix is a prerequisite for SH2 domains binding, our results suggest an alternative interaction mode to the "two-pronged plug two-hole socket" model [34,35]. This atypical SH2 binding may offer an additional mechanism for engaging with host SH2-containing proteins pre-phosphorylation.

Remarkably, the Y483- and Y511-containing motifs also interact with lipid bilayers, as observed for other disordered membrane signaling receptors [36-38]. Some T-cell receptor cytoplasmic disordered domains associate with the plasma membrane via tyrosine-based motifs with a helical propensity, and such association regulates phosphorylation and downstream signaling [39]. Like the CD3 ϵ cytoplasmic domain of the T-cell receptor [40], C-Tir also has two tyrosine residues (Y483 and Y511) involved in lipid-binding and helical propensity. Our results suggest that the C-terminal half Tir can mimic this regulatory stratagem to fine-tune its ability to interact with human cell components, thereby interfering with normal cellular functions. Further studies are needed to corroborate this initial observation. Nevertheless, we show that the C-Tir can interact with membranes via two ITIM-like sequences and its last C-terminal residues (Figure 4.1). Importantly, our results establish that this membrane affinity is residue-specific and modulated by lipid composition in a quantitative and site-resolved way, thus suggesting the existence of a regulation layer based on lipid composition. Our results also support that transiently structured lipid-binding regions might bury tyrosine residues rendering them inaccessible, as their phosphorylation perturbs membrane anchoring [41] (Figure 4.1). Importantly, we uncover that lipid binding by Tir C-terminus is essential for Tir-induced host cell death and pedestal formation (Figure 4.1).

The flexibility and conformational plasticity of C-Tir make it readily accessible for phosphorylation. We found in C-Tir four *bona fide* tyrosine phosphorylation sites modified by Fyn that render the ability for multivalent/promiscuous binding, supporting the idea of Tir acting as a signaling hub. Even though we only used protein fragments and explored one phosphorylation state, our results support the presence of four phosphotyrosine residues (pY) in the C-Tir sequence coexisting simultaneously.

Furthermore, subsequent multi-site phosphorylation enabled the transition from binary to multivalent fuzzy complexes with SH2 domains of both SHP-1 and SHP-2 (Figure 4.1). All pY-containing motifs can bind SH2 domains of SHP-1 and SHP-2 in the micromolar range, as reported for other proteins [42]. Sub-stoichiometric perturbation of NMR signals suggests that SH2 domains can undergo microscopic dissociation (hopping) from one phosphorylated site and “jump” to another. This hopping mechanism will offer a means to sample the available phosphorylation-sites as the protein jumps from one segment to another, allowing other proteins to interact with the released sites.

Interestingly, while in pre-phosphorylation the binding of C-Tir to C-SH2 domain of SHP-1 through residues surrounding Y511 appears to be preferential over SHP-2, we observed that upon phosphorylation C-Tir binds to the C-SH2 domain of SHP-2 with a binding strength of one or two (in pY454 site) orders of magnitude superior to that observed for the homologous domain of SHP-1.

Our results also suggest a general preferential binding of SH2 domains of SHP-2 through the ITIM-like sequences comprising pY483 and pY511, likely in an anti-parallel association in which N-SH2 and C-SH2 bind to pY511 and pY483, respectively. The diminished affinity for pY454, also observed for SHP-1, and pY474 sites may be associated with the role of both tyrosine residues in actin polymerization, essential for pedestals formation. Phosphorylation of Y474 is required for the interaction with the adaptor protein Nck, which in turn promotes activation of N-WASP and recruitment of the Arp2/3 complex to trigger actin polymerization [43]. In turn, the Y454 site is

part of a conserved NPY binding motif for the I-BAR domain of host IRSp53/IRTKS, linking Tir to the actin polymerization machinery in a phosphorylation-independent manner [44]. So, Y454 and Y474 host-like ITIM sequences might act as binding sites for distinct and competing host proteins depending on the stage of infection or on its phosphorylation state. The superposition of interplaying functional elements observed in C-Tir mimics the mechanics of human intrinsically disordered domains. Nevertheless, as the presence of four tyrosine phosphorylation sites, which allow simultaneous interactions with multiple SH2 domains, considerably increases the complexity in analysis of the presented results, future work using high-resolution structural studies targeting the interaction with multiple partners and different C-Tir's phosphorylation states must be employed to provide additional insights into Tir's binding specificity and action mode.

These observations combined highlight a functional diversity for C-Tir and a possible cooperation between host proteins, phosphorylation, and the membrane in signaling processes inside the host [45].

In the host, Tir acts as a negative regulator of TLR signaling pathways by enhancing the recruitment of SHP-1/2 to the adaptor protein and E3 ubiquitin ligase TRAF6 [30]. Yet, we uncover that Tir can directly recruit and bind to the C-terminal domain of TRAF6 (TRAF6-C) through a conserved binding motif (Pro-X-Glu-X-X-(Ar/Ac)) present in several host proteins involved in regulation of immune responses and reported to interact with TRAF6, such as CD40 [46] and IRAK-1 [47]. Our finding reinforces the idea that disordered bacterial effectors resource to host-like protein features as an effective approach to undermine host eukaryotic systems.

Curiously, the TRAF6 binding motif is present in the EPEC Tir sequence but absent in EHEC Tir, suggesting that if hijacked during infection, different mechanisms are employed to recruit TRAF6 in the other *E. coli* strains, as it happens in actin pedestal formation pathways [48].

We found that EPEC C-Tir binds to TRAF6-C domain with intermediate affinity through the region containing the residues Pro₄₆₉-Glu-Glu-His-Ile-

Tyr₄₇₄ (Figure 4.1). NMR analysis also showed that phosphorylation diminishes C-Tir's affinity to bind TRAF6 by one order of magnitude, reinforcing the idea of a cooperative mechanism between interaction with host proteins and phosphorylation in signaling processes inside the host (Figure 4.1). Importantly, we disclose that Y474 is required to interact with TRAF6. Our findings unravel a novel host intracellular target of Tir, and propose that Tir might enhance the suppression of host immune responses by acting as a scaffold to bring TRAF6 and SHP-1/2 together. However, further studies are required to better understand the determinants underlying these interactions.

Overall, we established an updated picture of Tir's intracellular side (Figure 4.1), laying the path towards illustrating how Tir hijacks host signaling. Notably, yet less abundant in prokaryotes, the structural disorder of Tir and several other A/E effectors reinforce the idea of a positive evolutionary selection towards disordered proteins in the pool of secreted effectors by bacterial pathogens to target host cellular machinery [11]. Given the ubiquitous presence of IDPs as transcriptional factors [49] and, more generally, as hubs in host networks [50], bacterial effectors with host-like disordered protein features are an efficient way to subvert host eukaryotic systems and promote infection.

Figure 4.1. Proposed model of the mechanisms employed by Tir to hijack intracellular signaling. Schematics of Tir in the host plasma membrane in a hairpin-topology while bound to intimin at the bacterial outer membrane. Our data support that the C-terminal intracellular region of Tir (blue) membrane receptor is disordered with transient helical structural elements involved in protein (SHP-1/2 and TRAF6) and lipid interactions, both pre-phosphorylation (orange left panel) and post-phosphorylation (pink right panel). Phosphorylation acts as a regulatory mechanism in signaling processes inside the host; while it enhances binding to SH2 domains of host phosphatases SHP-1 and SHP-2, enabling the formation of multivalent fuzzy complexes, it reduces C-Tir's affinity to lipid membranes and impairs binding to TRAF6. Relevant residues for interactions with membranes (top panel), SHP-1/2 (middle panel) and TRAF6 (down panel) are indicated.

References

- [1] Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., Hu, G., Uversky, V.N., Kurgan, L. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci.* **72**, 137-51 (2015).
- [2] Babu, M.M., Kriwacki, R.W., Pappu, R.V. Structural biology. Versatility from protein disorder. *Science.* **337**, 1460-1 (2012).
- [3] Uversky, V.N. Intrinsically disordered proteins and their “mysterious” (meta)physics. *Front. Phys.* **7** (2019).
- [4] Basile, W., Salvatore, M., Bassot, C., Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput Biol.* **15**, e1007186 (2019).
- [5] Xue, B., Dunker, A.K., Uversky, V.N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn.* **30**, 137-49 (2012).
- [6] Galea, C.A., Wang, Y., Sivakolundu, S.G., Kriwacki, R.W. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry.* **47**, 7598-609 (2008).
- [7] Pelham, J.F., Dunlap, J.C., Hurley, J.M. Intrinsic disorder is an essential characteristic of components in the conserved circadian circuit. *Cell Commun Signal.* **18**, 181 (2020).
- [8] Bah, A., Forman-Kay, J.D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J Biol Chem.* **291**, 6696-705 (2016).
- [9] van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sánchez, de Groot. N., Huynen, M.A., Matouschek, A., Fuxreiter, M., Babu, M.M. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep.* **8**, 1832-1844 (2014).
- [10] Uversky, V.N. Wrecked regulation of intrinsically disordered proteins in diseases: pathogenicity of deregulated regulators. *Front Mol Biosci.* **1**, 6 (2014).
- [11] Marín, M., Uversky, V.N., Ott, T. Intrinsic disorder in pathogen effectors: protein flexibility as an evolutionary hallmark in a molecular arms race. *Plant Cell.* **25**, 3153-7 (2013).
- [12] Backert, S., Blaser, M.J. The Role of CagA in the Gastric Biology of *Helicobacter pylori*. *Cancer Res.* **76**, 4028-31 (2016).
- [13] O'Brien, D.P., Durand, D., Voegelé, A., Hourdel, V., Davi, M., Chamot-Rooke, J., Vachette, P., Brier, S., Ladant, D., Chenal, A. Calmodulin fishing with a structurally disordered bait triggers CyaA catalysis. *PLoS Biol.* **15**, e2004486 (2017).
- [14] Aitio, O., Hellman, M., Skehan, B., Kesti, T., Leong, J.M., Saksela, K., Permi, P. Enterohaemorrhagic *Escherichia coli* exploits a tryptophan switch to hijack host f-actin assembly. *Structure.* **20**, 1692-703 (2012).

Discussion

- [15] Hamada, D., Hamaguchi, M., Suzuki, K.N., Sakata, I., Yanagihara, I. Cytoskeleton-modulating effectors of enteropathogenic and enterohemorrhagic *Escherichia coli*: a case for EspB as an intrinsically less-ordered effector. *FEBS J.* **277**, 2409-15 (2010).
- [16] Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., *et al.* Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. *J Bacteriol.* **191**, 347-54 (2009).
- [17] Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., Gibson, T.J. Attributes of short linear motifs. *Mol Biosyst.* **8**, 268-81 (2012).
- [18] Van Roey, K., Uyar, B., Weatheritt, R.J., Dinkel, H., Seiler, M., Budd, A., Gibson, T.J., Davey, N.E. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* **114**, 6733-78 (2014).
- [19] Csizmok, V., Follis, A.V., Kriwacki, R.W., Forman-Kay, J.D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chem Rev.* **116**, 6424-62 (2016).
- [20] Jespersen, N., Barbar, E. Emerging Features of Linear Motif-Binding Hub Proteins. *Trends Biochem Sci.* **45**, 375-384 (2020).
- [21] Mills, E., Baruch, K., Charpentier, X., Kobi, S., Rosenshine, I. Real-time analysis of effector translocation by the type III secretion system of enteropathogenic *Escherichia coli*. *Cell Host Microbe.* **3**, 104-13 (2008).
- [22] Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., Finlay, B.B. Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell.* **91**, 511-20 (1997).
- [23] Gruenheid, S., DeVinney, R., Bladt, F., Goosney, D., Gelkop, S., Gish, G.D., Pawson, T., Finlay, B.B. Enteropathogenic *E. coli* Tir binds Nck to initiate actin pedestal formation in host cells. *Nat Cell Biol.* **3**, 856-9 (2001).
- [24] Xue, Y., Du, M., Sheng, H., Hovde, C.J., Zhu, M.J. *Escherichia coli* O157:H7 suppresses host autophagy and promotes epithelial adhesion via Tir-mediated and cAMP-independent activation of protein kinase A. *Cell Death Discov.* **3**, 17055 (2017).
- [25] Yan, D., Wang, X., Luo, L., Cao, X., Ge, B. Inhibition of TLR signaling by a bacterial protein containing immunoreceptor tyrosine-based inhibitory motifs. *Nat Immunol.* **13**, 1063-71 (2012).
- [26] Goddard, P.J., Sanchez-Garrido, J., Slater, S.L., Kalyan, M., Ruano-Gallego, D., Marchès, O., Fernández, L.Á., Frankel, G., Shenoy, A.R. Enteropathogenic *Escherichia coli* Stimulates Effector-Driven Rapid Caspase-4 Activation in Human Macrophages. *Cell Rep.* **27**, 1008-1017 (2019).
- [27] Zhong, Q., Roumeliotis, T.I., Kozik, Z., Cepeda-Molero, M., Fernández, L.Á., Shenoy, A.R., Bakal, C., Frankel, G., Choudhary, J.S. Clustering of Tir during

enteropathogenic *E. coli* infection triggers calcium influx-dependent pyroptosis in intestinal epithelial cells. *PLoS Biol.* **18**, e3000986 (2020).

[28] Luo, Y., Frey, E.A., Pfuetzner, R.A., Creagh, A.L., Knoechel, D.G., Haynes, C.A., Finlay, B.B., Strynadka, N.C. Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature.* **405**, 1073-7 (2000).

[29] Blasche, S., Arens, S., Ceol, A., Siszler, G., Schmidt, M.A., Häuser, R., Schwarz, F., Wuchty, S., Aloy, P., Uetz, P., Stradal, T., Koegl, M. The EHEC-host interactome reveals novel targets for the translocated intimin receptor. *Sci Rep.* **4**, 7531 (2014).

[30] Yan, D., Quan, H., Wang, L., Liu, F., Liu, H., Chen, J., Cao, X., Ge, B. Enteropathogenic *Escherichia coli* Tir recruits cellular SHP-2 through ITIM motifs to suppress host immune response. *Cell Signal.* **25**, 1887-94 (2013).

[31] Zhou, R., Chen, Z., Hao, D., Wang, Y., Zhang, Y., Yi, X., Lyu, L.D., Liu, H., Zou, Q., Chu, Y., Ge, B., Yan, D. Enterohemorrhagic *Escherichia coli* Tir inhibits TAK1 activation and mediates immune evasion. *Emerg Microbes Infect.* **8**, 734-748 (2019).

[32] Kim, D.H., Han, K.H. PreSMo Target-Binding Signatures in Intrinsically Disordered Proteins. *Mol Cells.* **41**, 889-899 (2018).

[33] Mollica, L., Bessa, L.M., Hanouille, X., Jensen, M.R., Blackledge, M., Schneider, R. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front Mol Biosci.* **3**, 52 (2016).

[34] Nachman, J., Gish, G., Virag, C., Pawson, T., Pomès, R., Pai, E. Conformational determinants of phosphotyrosine peptides complexed with the Src SH2 domain. *PLoS One.* **5**, e11215 (2010).

[35] Nieto, L., Tharun, I.M., Balk, M., Wienk, H., Boelens, R., Ottmann, C., Milroy, L.G., Brunsveld, L. Estrogen Receptor Folding Modulates cSrc Kinase SH2 Interaction via a Helical Binding Mode. *ACS Chem Biol.* **10**, 2624-32 (2015).

[36] Deryusheva, E., Nemashkalova, E., Galloux, M., Richard, C.A., Eléouët, J.F., Kovacs, D., Van Belle, K., Tompa, P., Uversky, V., Permyakov, S. Does Intrinsic Disorder in Proteins Favor Their Interaction with Lipids? *Proteomics.* **19**, e1800098 (2019).

[37] Bürgi, J., Xue, B., Uversky, V.N., van der Goot, F.G. Intrinsic Disorder in Transmembrane Proteins: Roles in Signaling and Topology Prediction. *PLoS One.* **11**, e0158594 (2016).

[38] Kjaergaard, M., Kragelund, B.B. Functions of intrinsic disorder in transmembrane proteins. *Cell Mol Life Sci.* **74**, 3205-3224 (2017).

[39] Xu, X., Li, H., Xu, C. Structural understanding of T cell receptor triggering. *Cell Mol Immunol.* **17**, 193-202 (2020).

[40] Xu, C., Gagnon, E., Call, M.E., Schnell, J.R., Schwieters, C.D., Carman, C.V., Chou, J.J., Wucherpfennig, K.W. Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif. *Cell.* **135**, 702-13 (2008).

Discussion

- [41] Arbesú, M., Iruela, G., Fuentes, H., Teixeira, J.M.C., Pons, M. Intramolecular Fuzzy Interactions Involving Intrinsically Disordered Domains. *Front Mol Biosci.* **5**, 39 (2018).
- [42] Wagner, M.J., Stacey, M.M., Liu, B.A., Pawson, T. Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harb Perspect Biol.* **5**, a008987 (2013).
- [43] Campellone, K.G., Giese, A., Tipper, D.J., Leong, J.M. A tyrosine-phosphorylated 12-amino-acid sequence of enteropathogenic *Escherichia coli* Tir binds the host adaptor protein Nck and is required for Nck localization to actin pedestals. *Mol Microbiol* **43**, 1227-1241 (2002).
- [44] de Groot, J.C., Schlüter, K., Carius, Y., Quedenau, C., Vingadassalom, D., Faix, J., Weiss, S.M., Reichelt, J., Standfuss-Gabisch, C., Lesser, C.F., Leong, J.M., Heinz, D.W., Büssow, K., Stradal, T.E. Structural basis for complex formation between human IRSp53 and the translocated intimin receptor Tir of enterohemorrhagic *E. coli*. *Structure.* **19**, 1294-306 (2011).
- [45] Cornish, J., Chamberlain, S.G., Owen, D., Mott, H.R. Intrinsically disordered proteins and membranes: a marriage of convenience for cell signalling? *Biochem Soc Trans.* **48**, 2669-2689 (2020).
- [46] Engel, D., Seijkens, T., Poggi, M., Sanati, M., Thevissen, L., Beckers, L., Wijnands, E., Lievens, D., Lutgens, E. The immunobiology of CD154-CD40-TRAF interactions in atherosclerosis. *Semin Immunol.* **21**, 308-12 (2009).
- [47] Muroi, M., Tanamoto, K. IRAK-1-mediated negative regulation of Toll-like receptor signaling through proteasome-dependent downregulation of TRAF6. *Biochim Biophys Acta.* **1823**, 255-63 (2012).
- [48] DeVinney, R., Puente, J.L., Gauthier, A., Goosney, D., Finlay, B.B. Enterohaemorrhagic and enteropathogenic *Escherichia coli* use a different Tir-based mechanism for pedestal formation. *Mol Microbiol.* **41**, 1445-58 (2001).
- [49] Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., Skriver, K. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem J.* **474**, 2509-2532 (2017).
- [50] Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., Iakoucheva, L.M. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol.* **2**, e100 (2006)

Concluding remarks and future perspectives

In conclusion, this dissertation provided an updated understanding of Tir's intracellular side and molecular mechanisms exploited by this effector to subvert host immune system responses.

The results herein presented show that:

- **C-Tir is disordered.** The C-terminal intracellular domain of Tir (C-Tir) is flexible with IDR features, but it can adopt partial secondary structural elements and non-random long-range contacts around phosphorylation sites and motifs that are functionally relevant in diversified interactions;
- **The disordered C-Tir binds lipids.** C-Tir can undergo multivalent electrostatic interactions with lipid bilayers essential for Tir-induced host cell death and actin pedestal formation. Membrane affinity is residue-specific, modulated by lipid composition, and regulated by phosphorylation.
- **C-Tir is a multivalent disordered tail.** C-Tir resources to its host-like ITIMs to recruit SH2 domains of host phosphatases SHP-1 and SHP-2 in a mechanism related to tyrosine phosphorylation, enabling the formation of multivalent fuzzy complexes.
- **Tir prefers to bind SHP-2 in an anti-parallel fashion.** Yet in a dynamic association, C-Tir interacts preferably with SHP-2 N-SH2 and C-SH2 via pY511 and pY483, respectively.
- **C-Tir binds SHP-2 and TRAF6.** C-Tir can directly recruit the TRAF6-C domain through a conserved binding motif, found in host proteins, encompassing Y474 which is absolutely required for the interaction. This thesis supports the view that EPEC creates the conditions to assemble TRAF6 and SHP-2 in a complex to suppress TLR signaling by the bacterial effector scaffold Tir.

Altogether, this thesis brings a unique quantitative framework to understand Tir-mediated virulence and host-defense evasion and opens the way for new studies that aim the understanding of virulence mechanisms with implications for health and disease.

Nevertheless, as with any research project, there is still a lot of future work and goals to be achieved, including:

- **Extract quantitative information about site-specific phosphorylation events in C-Tir.** Mutations in the tyrosine residues (Y454, Y474, Y483 and Y511) must be carried out to clarify the determinants underlying the phosphorylation events in C-Tir, particularly to infer about their potential cross-talk. Moreover, measurements of quenched phosphorylation reactions at different substrate concentrations should be performed to obtain the exact properties and kinetic constants of C-Tir's phosphorylation reactions.
- **Assess if phosphorylation depends on lipid-binding.** We showed that phosphorylation hinders C-Tir's membrane association. We wonder if the membrane association could, in turn, affect residue-specific phosphorylation extension and kinetics.
- **Probe the interactions of C-Tir with host targets (SHP-2 and TRAF6) using SAXS.** To reinforce the observations driven from this work, interactions of C-Tir with SH2 domains of SHP-2 and TRAF6-C must be further explored using SAXS to probe the results obtained by NMR, bringing new insights about these transient protein-protein interactions and obtaining additional quantitative information.
- **Deliver the structural biology of the Tir –SHP-2–TRAF6 ternary complex.** All of the structural characterization from free-state Tir as well as from bound to host components (TRAF6 and SHP-2), obtained in this thesis will be integrated and complemented with NMR and SAXS experiments in the presence of the three proteins (Tir, TRAF6 and SHP-2). In this topic, we aim to start by replacing

Y474 with a glutamate residue, expecting to retain the TRAF6-binding motif in C-Tir and at the same time avoid binding of SHP-2 in this region. ITC experiments will also be important to gain insights about the directionality of the binding needed for the formation of this complex.

Supplementary Information

Supplementary Figures

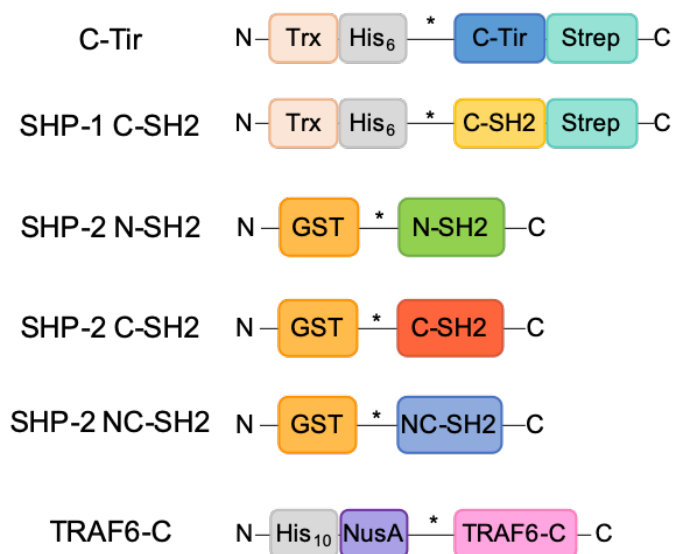


Figure SI.1. Schematic representation of the constructs used in this work for the production of C-Tir, C-SH2 domain of SHP-1, C-, N- and NC-SH2 domains of SHP-2 and TRAF6-C. The * marks the position of the HRV-3C recognition sequence.

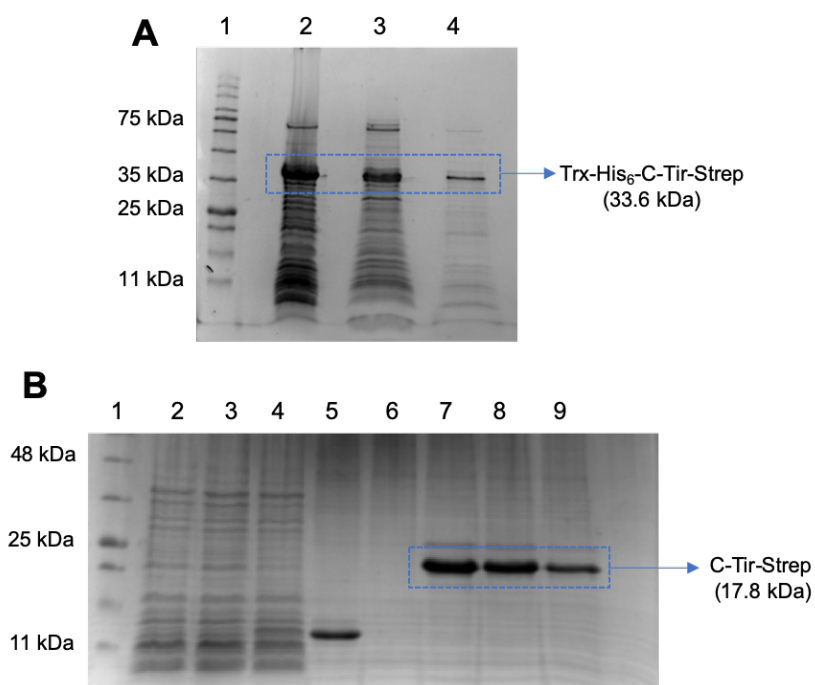


Figure SI.2. Overview of C-Tir purification. **(A)** SDS-PAGE analysis of the clarified cell lysates collected upon ultracentrifugation. Lane 1- NZYColour Protein Marker II; Lane 2- soluble Fraction; Lanes 3 and 4- fractions collected after first and second pellet washes, respectively. **(B)** SDS-PAGE analysis of C-Tir purification by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lanes 2 to 4- column flow-through fractions; Lane 5- column washing fraction; Lanes 6 to 9- protein elution fractions.

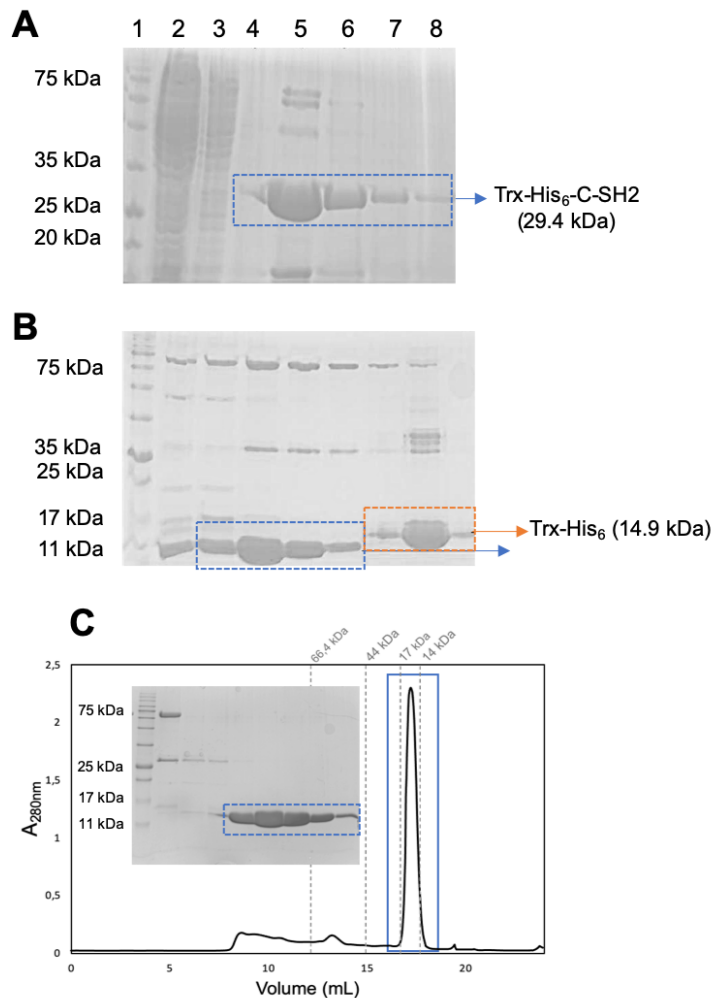


Figure SI.3. Overview of SHP-1 C-SH2 domain purification. **(A)** SDS-PAGE analysis of the first purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lanes 2 and 3- column flow-through fractions; Lane 4- column washing fraction; Lanes 5 to 8- protein elution fractions. **(B)** SDS-PAGE analysis of the second purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2- column flow-through fraction; Lanes 3 to 6- column washing fractions; Lanes 7 to 9- elution fraction. **(C)** SDS-PAGE analysis and chromatogram of the purification step by size exclusion chromatography using a Superdex™ 200 Increase 10/300 GL column. The fractions containing pure protein are highlighted inside blue boxes.

Supplementary Information

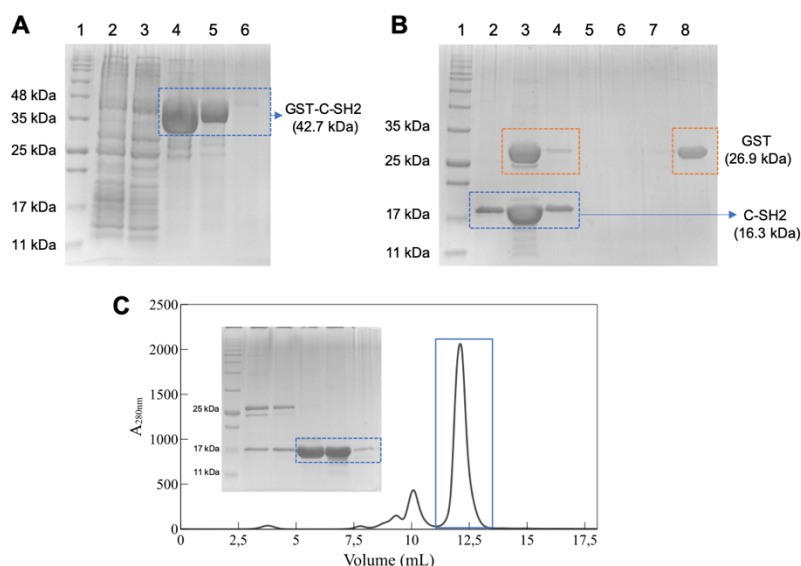


Figure SI.4. Overview of SHP-2 C-SH2 domain purification. **(A)** SDS-PAGE analysis of the first purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2- column flow-through fraction; Lane 3- column washing fraction; Lanes 4 to 6- protein elution fractions. **(B)** SDS-PAGE analysis of the second purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2- column flow-through fraction; Lanes 3 to 6- column washing fractions; Lane 7- elution fraction. **(C)** SDS-PAGE analysis and chromatogram of the last purification step by size exclusion chromatography using a Superdex™ 75 Increase 10/300 GL column. The fractions containing pure protein are highlighted inside blue boxes.

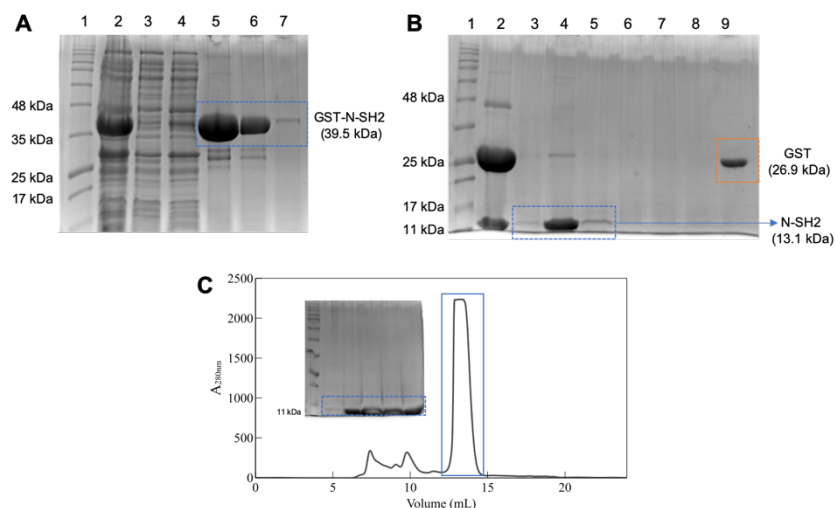


Figure SI.5. Overview of SHP-2 N-SH2 domain purification. **(A)** SDS-PAGE analysis of the first purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lanes 2 and 3- column flow-through fractions; Lane 4- column washing fraction; Lanes 5 to 7- protein elution fractions. **(B)** SDS-PAGE analysis of the second purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2- dialyzed fraction; Lane 3- column flow-through fraction; Lanes 4 to 8- column washing fractions; Lane 9- elution fraction. **(C)** SDS-PAGE analysis and chromatogram of the last purification step by size exclusion chromatography using a Superdex™ 75 Increase 10/300 GL column. The fractions containing pure protein are highlighted inside blue boxes.

Supplementary Information

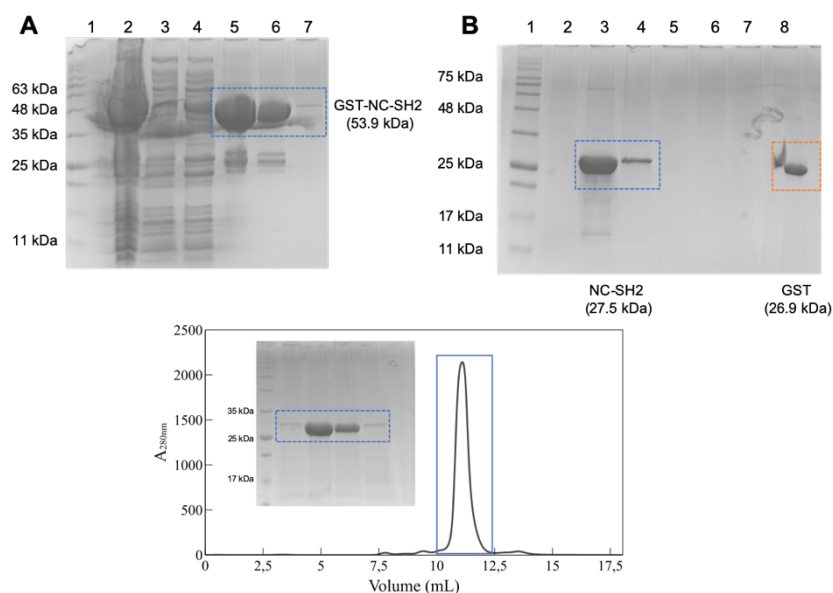


Figure SI.6. Overview of SHP-2 NC-SH2 domain purification. (A) SDS-PAGE analysis of the first purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lanes 2 and 3- column flow-through fraction; Lane 4- column washing fraction; Lanes 5 to 7- protein elution fractions. **(B)** SDS-PAGE analysis of the second purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2- column flow-through fraction; Lanes 3 to 7- column washing fractions; Lane 8- elution fraction. **(C)** SDS-PAGE analysis and chromatogram of the last purification step by size exclusion chromatography using a Superdex™ 75 Increase 10/300 GL column. The fractions containing pure protein are highlighted inside blue boxes.

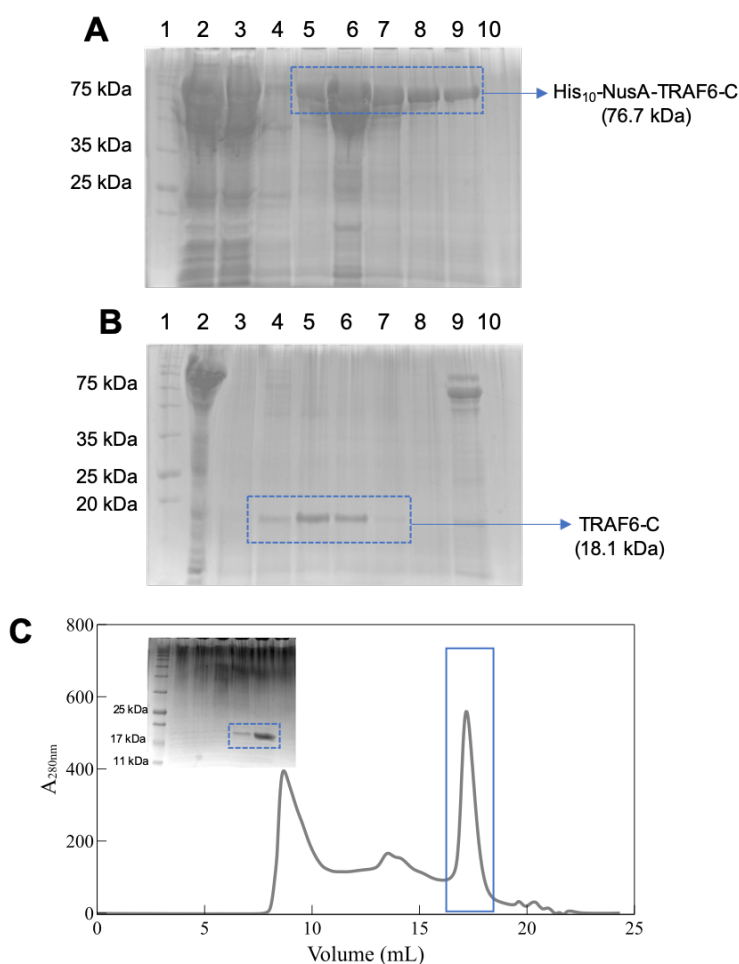


Figure SI.7. Overview of TRAF6-C domain purification. (A) SDS-PAGE analysis of the first purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lanes 2 and 3- column flow-through fraction; Lane 4- column washing fraction; Lanes 5 to 9- protein elution fractions. **(B)** SDS-PAGE analysis of the second purification step by affinity chromatography. Lane 1- NZYColour Protein Marker II; Lane 2-dialyzed fraction; Lane 3- column flow-through fraction; Lanes 4 to 8- column washing fractions; Lane 9- elution fraction. **(C)** SDS-PAGE analysis and chromatogram of the last purification step by size exclusion chromatography using a Superdex™ 200 Increase 10/300 GL column. The fractions containing pure protein are highlighted inside blue boxes.

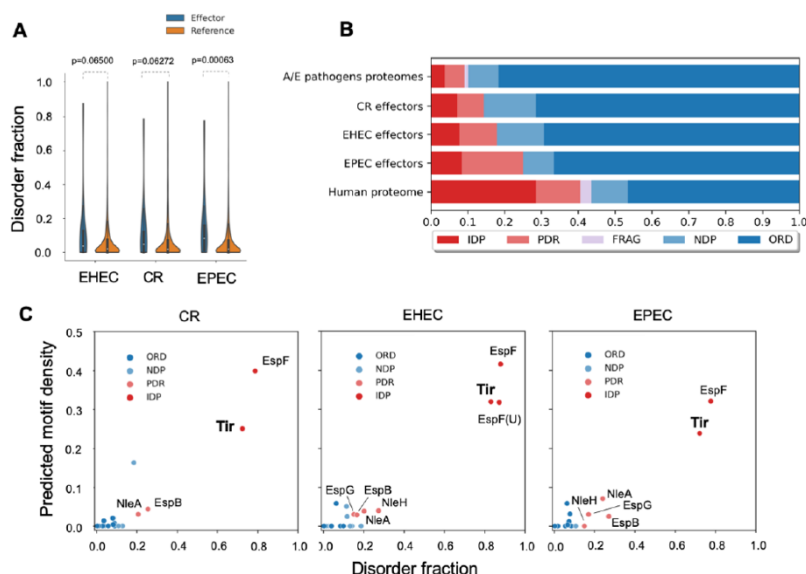


Figure SI.8. Predicted structural disorder in A/E pathogens with IUPred.

(A) Distribution of IUPred based disorder fraction displayed as violin plots of effectors (blue) and full-proteomes (red) of A/E pathogens. **(B)** Accumulated fractions of the structural categories in terms of sequence disorder based on IUPred disorder-score (long mode) as described for Figure 4.1. **(C)** Fraction of ELMs vs. disorder fraction in A/E effectors. Tir, EspF, and EspF(U) display a high motif content and disorder fraction. IDP-like (red dots) and PDR (orange) effectors are labeled.

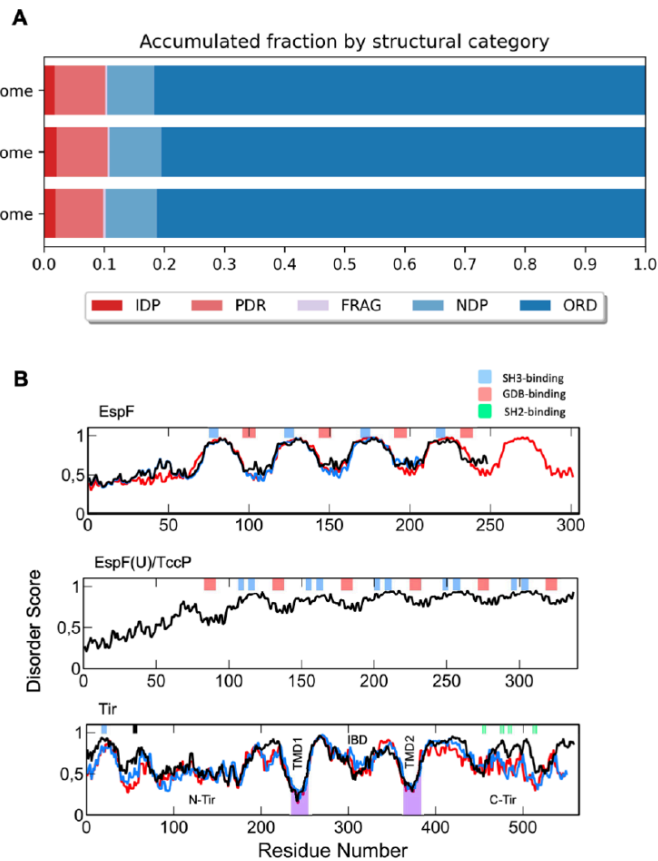


Figure SI.9. A/E disordered effectors. (A) Accumulated fractions of the structural categories in terms of sequence disorder based on Disopred for the individual A/E pathogen reference proteomes (see Figure 4.1 and Materials and Methods Table 3.8). **(B)** Disorder propensity of EspF (top), Espf(U)/Tccp (middle), and Tir (bottom) from EHEC (black line), EPEC (blue line), and CR (red line). Tir's disorder propensity is conserved among A/E pathogens. Experimental verified GBD-, SH3- and SH2-binding motifs are indicated in light red, blue, and green bars, respectively. The 14-3-3 motif within N-Tir first residues is indicated by a black bar. The transmembrane domains of Tir are highlighted in purple.

Supplementary Information

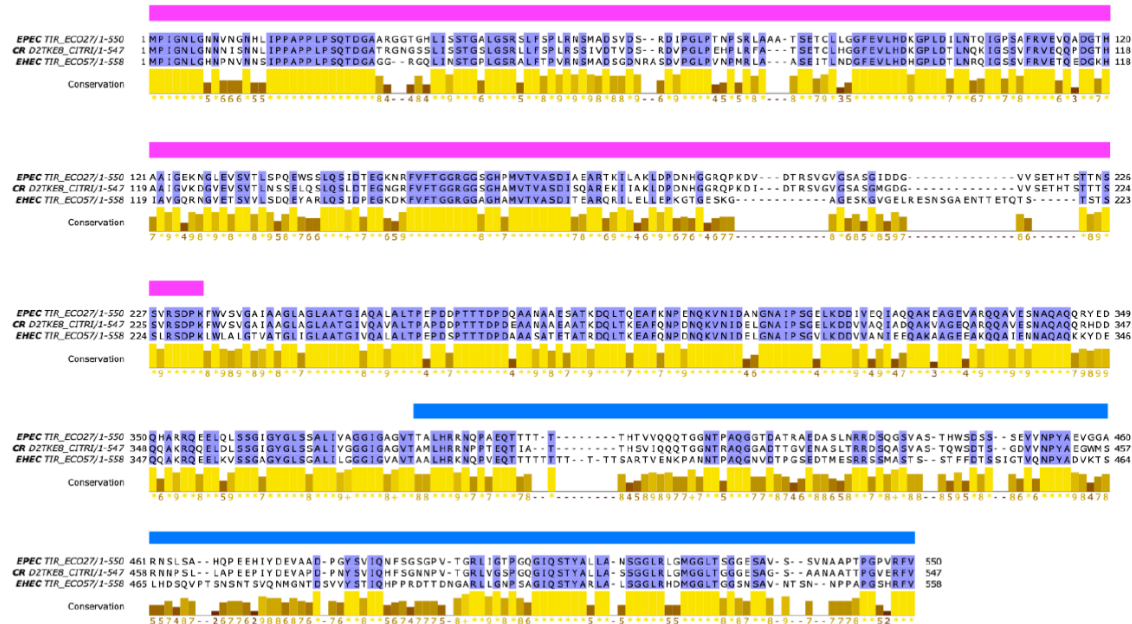


Figure SI.10. Multiple sequence alignment of Tir sequences generated from a set of 114 non-identical sequences belonging to the Tir receptor family from the Uniprot database. For convenience, only Tir sequences from EPEC, CR and EHEC are shown. Alignment positions conserved in all three sequences are shaded in lavender. The conservation of physicochemical properties in each position of the alignment is reported in the corresponding bar plot below the alignment. The intracellular domains of EPEC Tir are delimited by magenta (N-Tir) and light-blue (C-Tir) horizontal boxes, respectively. The multiple sequence alignment (MSA) was generated using Clustal Omega [1] with default parameters and JalView [2] was used for MSA visualization and figure generation. Pairwise alignments among EPEC, EHEC and CR Tir sequences were performed using the Smith-Waterman algorithm as implemented in the EMBOSS software suite [3].

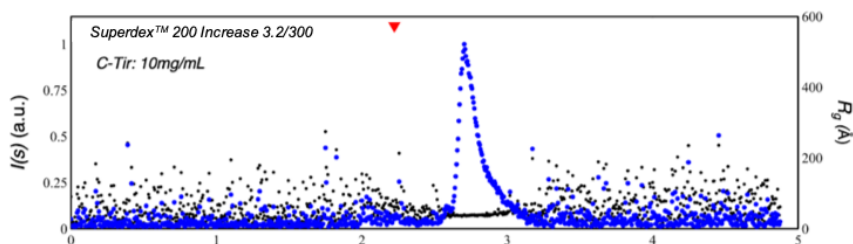


Figure SI.11. SEC-SAXS chromatogram of C-Tir. The panel shows normalized total scattering intensity $I(s)$, over the entire s range, from each frame acquired along elution volume and respective R_g -value (black circles). The flat variation of R_g values reflects a pure monodisperse sample. The column type for high-resolution size exclusion chromatography and sample concentrations are on the top left of the panel. For reference, the retention volume for monomeric BSA (66.4 kDa) is displayed by a red triangle.

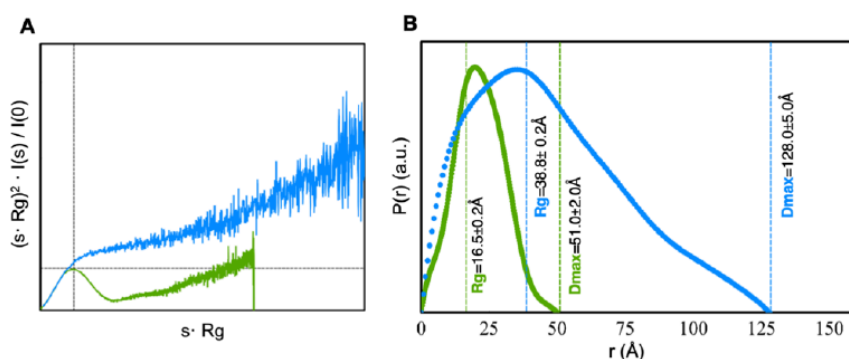


Figure SI.12. Comparative SAXS: C-Tir versus globular Myoglobin (Myb). Myb is a monomeric globular protein with a molecular mass similar to Strep-tagged C-Tir (ca 17 kDa). **(A)** Kratky representations of the SAXS patterns of C-Tir (blue) and Myb (green) are compared next to their respective $P(r)$ versus r profiles and plotted using the same color code. The absence of a clear peak maximum at $sR_g = \sqrt{3}$ (dashed cross) indicates that C-Tir is conformationally flexible and nonglobular. In contrast, the Kratky plot of Myb is bell-shaped as expected for well-folded globular proteins. The SAXS experimental set for Myb was obtained from the curated repository of scattering data SASDB [4], with the entry code SASDAH2. **(B)** Normalized pairwise distance distribution, $P(r)$, computed from experimental SAXS curves of C-Tir (blue) and Myb (green). Dashed lines indicate the derived R_g and D_{max} values. C-Tir's internal distances are more significant than in Myb, as expected for a more extended and nonglobular protein.

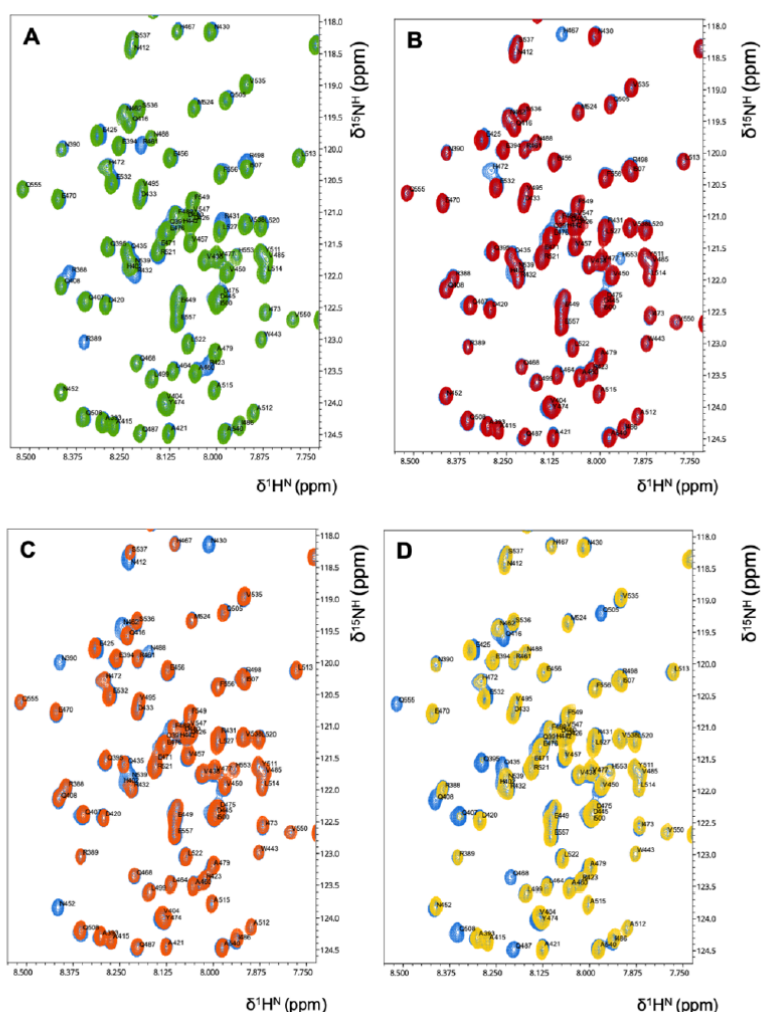


Figure SI.13. Reverse labelling of C-Tir. The $[^1\text{H}-^{15}\text{N}]$ -HSQC NMR spectrum of C-Tir exhibits a characteristic spectral crowding in the random-coil region around 8.0 ppm. To alleviate the ambiguity caused by substantial resonances overlapping, we selectively unlabeled (A) asparagine, (B) histidine, (C) lysine and (D) glutamine residues against a uniformly ^{15}N -labelled C-Tir (blue). Selectively unlabeled residues became absent in the HSQC spectra due to reverse labeling, allowing pinpoint the amino-acid type. We used this prior-knowledge to aid and validate the assignment.

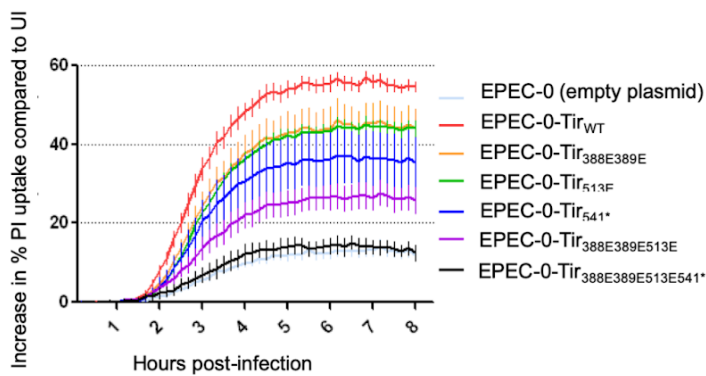


Figure SI.14. Time course PI uptake of EPEC-infected cells. PI uptake into SNU-C5 cells infected with EPEC-0 (empty plasmid), EPEC-0-Tir_{WT}, EPEC-0-Tir_{388E389E}, EPEC-0-Tir_{513E}, EPEC-0-Tir_{541*}, EPEC-0-Tir_{388E389E513E} or EPEC-0-Tir_{388E389E513E541*} from 10 min to 8 h post-infection with 10 min intervals. Data shown are mean $\pm$ SEM from 3 (EPEC-0-Tir_{388E389E513E541*}) or 6 (all other strains) independent biological repeats.

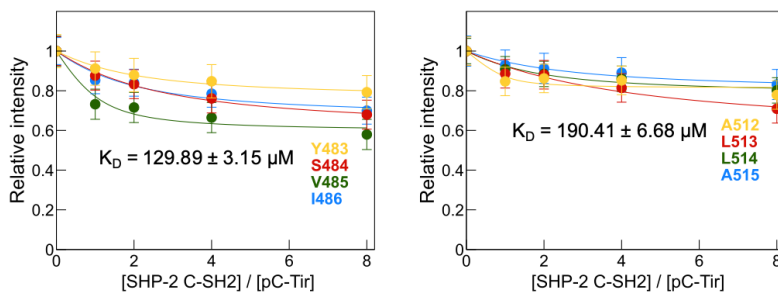


Figure SI.15. Global fittings obtained for the interaction of unphosphorylated C-Tir with C-SH2 domain of SHP-2. The relative local intensity drops around each site were employed to fit a 1:1 binding model.

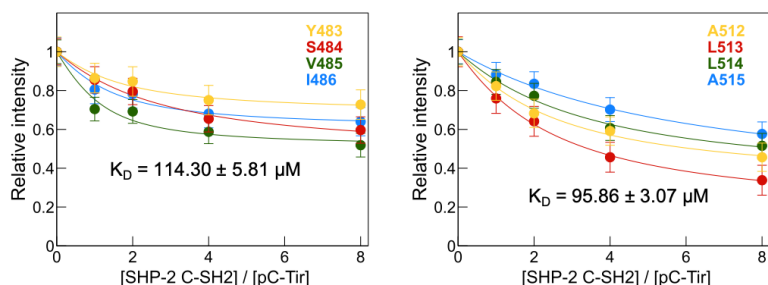


Figure SI.16. Global fittings obtained for the interaction of unphosphorylated C-Tir with NC-SH2 domain of SHP-2. The relative local intensity drops around each site were employed to fit a 1:1 binding model.

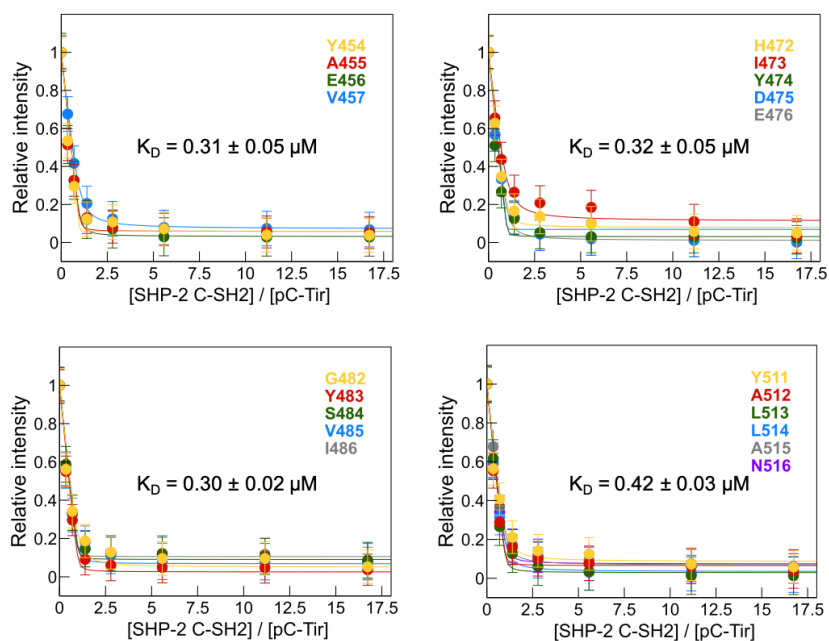


Figure SI.17. Global fittings obtained for the interaction of pC-Tir with C-SH2 domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

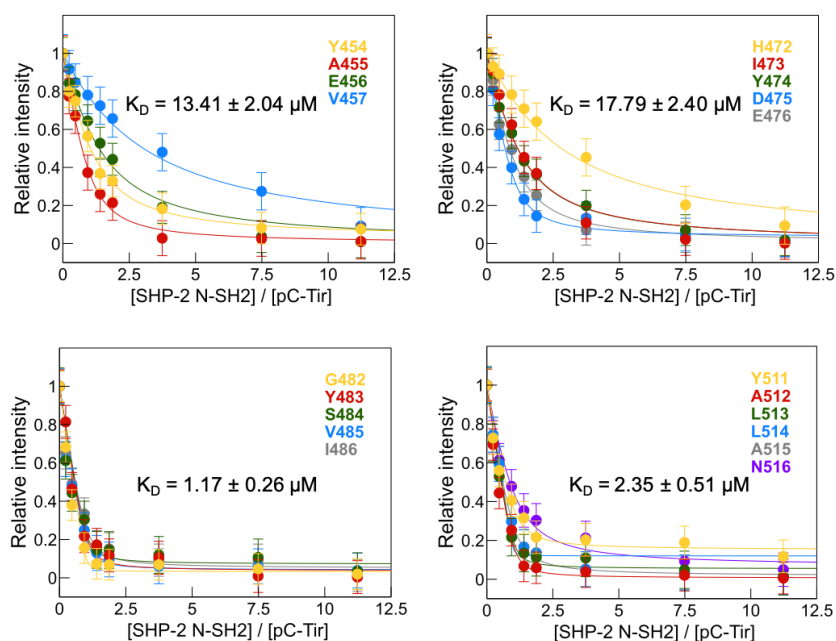


Figure SI.18. Global fittings obtained for the interaction of pC-Tir with N-SH2 domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

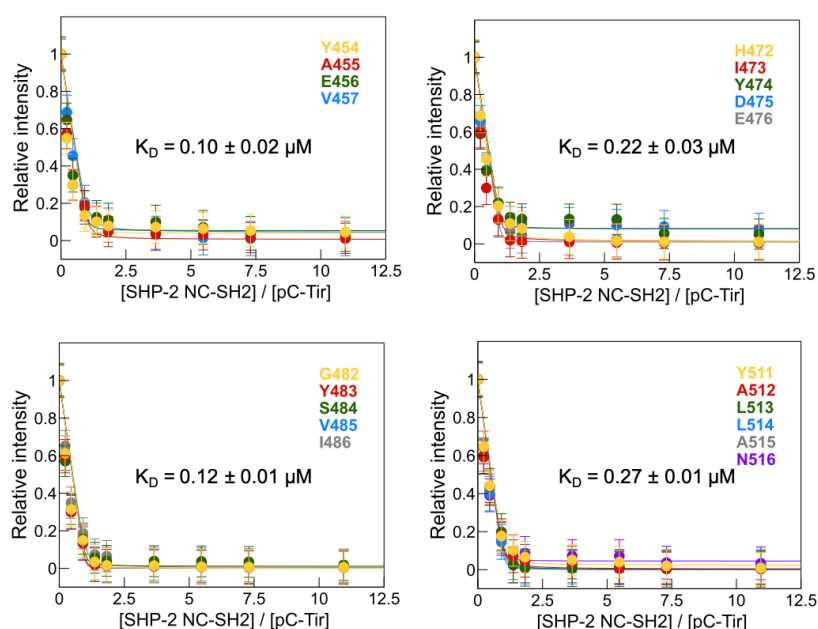


Figure S1.19. Global fittings obtained for the interaction of pC-Tir with NC-SH2 domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

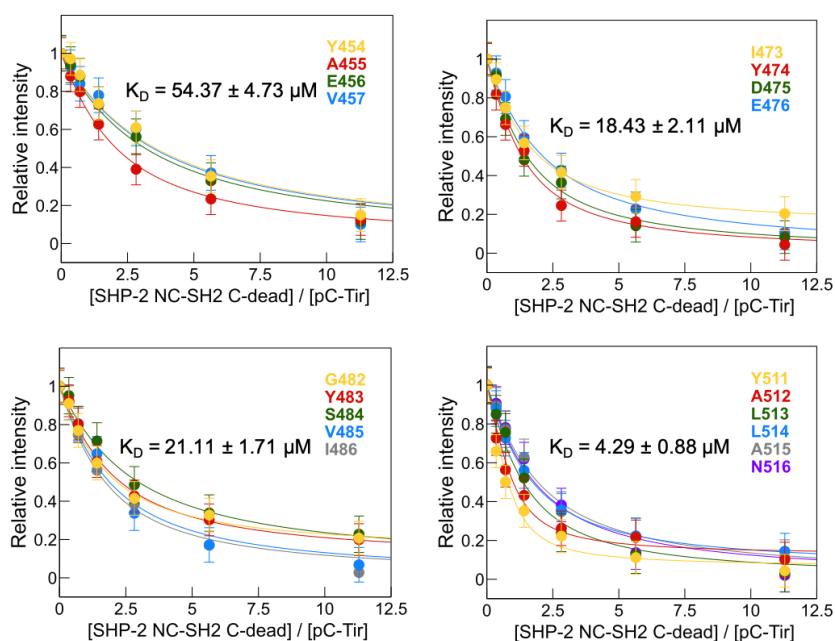


Figure SI.120. Global fittings obtained for the interaction of pC-Tir with NC-SH2 C-dead domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

Supplementary Information

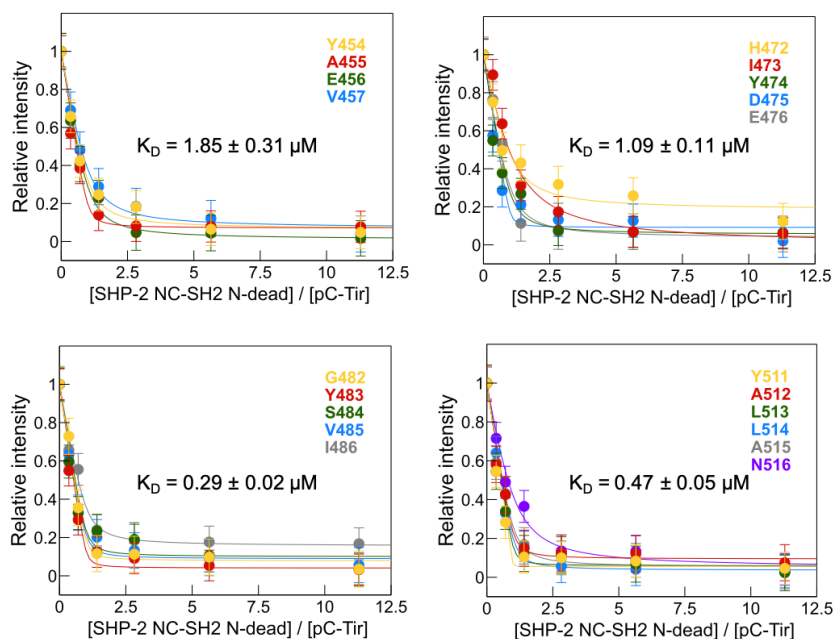


Figure SI.21. Global fittings obtained for the interaction of pC-Tir with NC-SH2 N-dead domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

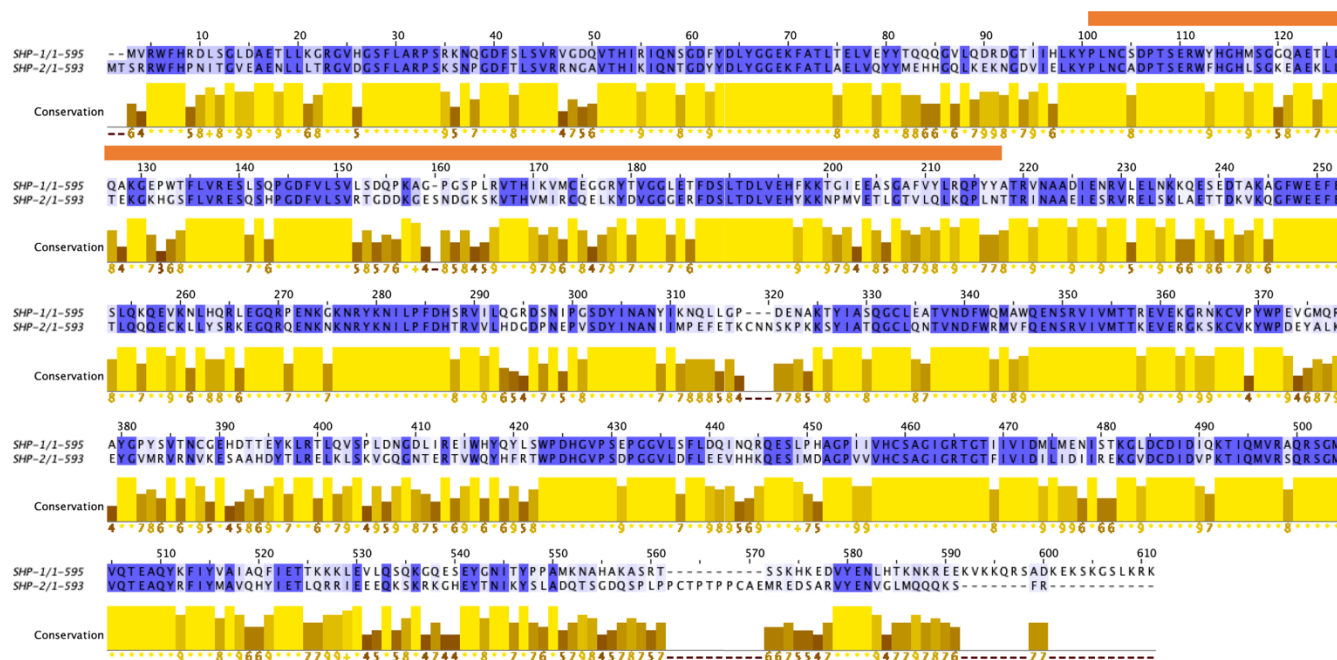


Figure SI.22. Multiple sequence alignment of SHP-1 and SHP-2 sequences (Uniprot accession codes P29350 and Q06124, respectively). Alignment positions conserved in the two sequences are shaded in lavender. The conservation of physicochemical properties in each position of the alignment is reported in the corresponding bar plot below the alignment. The C-SH2 domain is delimited by the orange horizontal box. The multiple sequence alignment (MSA) was generated using Clustal Omega [1] with default parameters and JalView [2] was used for MSA visualization and figure generation.

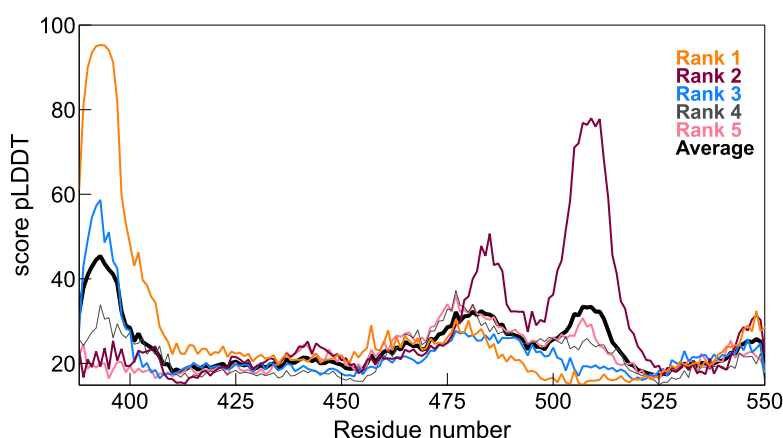


Figure SI.23. pLDDT score as function of C-Tir residue number obtained for the predictions of the binding interface between TRAF6-C and C-Tir using AlphaFold2. Rank 1 to Rank 5 correspond to the five models generated in the structural prediction. pLDDT average score is depicted in black.



Figure SI.24. Cartoon representation of the TRAF6-CD40 complex (PDB accession code: 1LB6). TRAF6 (residues 333-508) and the CD40 antigen are depicted in green and orange, respectively.

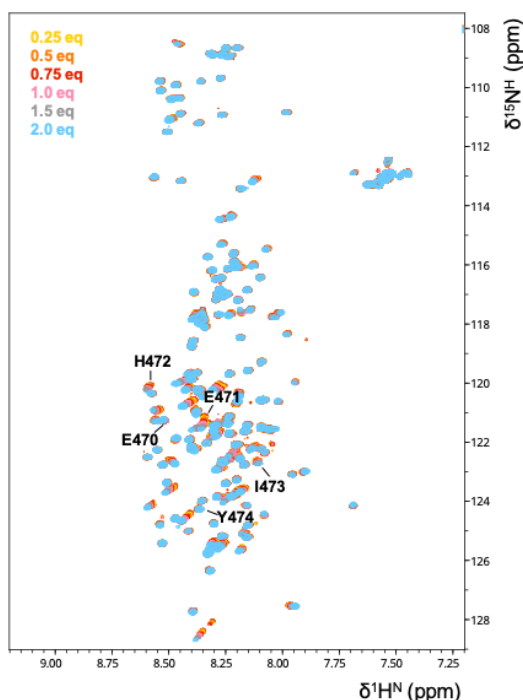


Figure SI.25. Overlay of $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra of pC-Tir in the presence of 0.25 (yellow), 0.5 (orange), 0.75 (red), 1.0 (pink), 1.5 (grey) and 2.0 (light blue) molar equivalents of TRAF6-C domain.

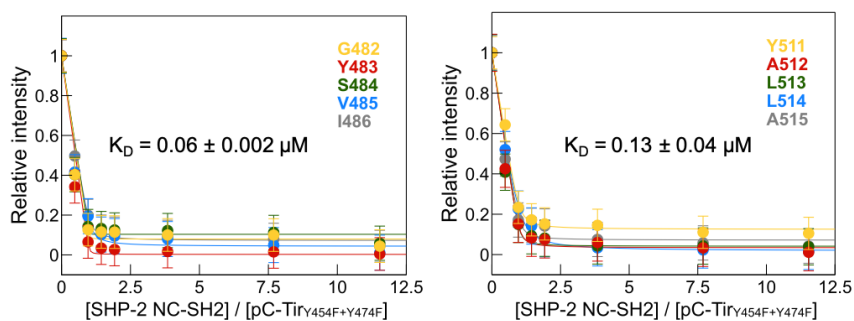


Figure SI.26. Global fittings obtained for the interaction of pC-Tir_{Y454F+Y474F} with NC-SH2 domain of SHP-2. The relative local intensity drops around each phosphorylation site (i.e., each phospho-tyrosine and neighboring residues) were employed to fit a 1:1 binding model.

Supplementary Tables

Table SI.1. Disorder fractions (DISOPRED), predicted motif densities and structural categories for CR, EHEC and EPEC effectors.

protein_ac	disorder_ fraction	overall_motif _density	disorder_ category	effector_ collection
A0A0H3JC80	0.007	0.000	ORD	EHEC
A0A0H3JD33	0.073	0.000	ORD	EHEC
A0A0H3JDH6	0.011	0.000	ORD	EHEC
A0A0H3JDV8	0.032	0.000	ORD	EHEC
A0A0H3JE38	0.091	0.000	ORD	EHEC
A0A0H3JFN8	0.077	0.000	ORD	EHEC
A0A0H3JGR6	0.095	0.000	ORD	EHEC
A0A0H3JHA4	0.163	0.000	NDP	EHEC
A0A0H3JJJ4	0.007	0.000	ORD	EHEC
A0A0H3JP21	0.020	0.000	ORD	EHEC
P0DJ88	0.424	0.318	IDP	EHEC
Q7DB50	0.048	0.030	ORD	EHEC
Q7DB68	0.051	0.000	ORD	EHEC
Q7DB74	0.125	0.000	NDP	EHEC
Q7DB76	0.148	0.025	PDR	EHEC
Q7DB77	0.332	0.319	IDP	EHEC
Q7DB85	0.601	0.415	IDP	EHEC
Q7DBA6	0.045	0.000	ORD	EHEC
Q8X4Q6	0.061	0.000	ORD	EHEC
Q8X4W3	0.077	0.000	ORD	EHEC
Q8X4X1	0.068	0.000	ORD	EHEC
Q8X4X2	0.072	0.000	ORD	EHEC
Q8X4X3	0.094	0.000	ORD	EHEC
Q8X507	0.089	0.000	ORD	EHEC
Q8X509	0.068	0.000	ORD	EHEC
Q8X782	0.075	0.000	ORD	EHEC
Q8X831	0.201	0.000	PDR	EHEC
Q8X834	0.052	0.000	ORD	EHEC
Q8X9A5	0.014	0.000	ORD	EHEC
Q8X9A7	0.093	0.000	ORD	EHEC
Q8XA11	0.056	0.000	ORD	EHEC
Q8XAJ5	0.068	0.039	ORD	EHEC

Q8XAL6	0.040	0.040	ORD	EHEC
Q8XAL7	0.042	0.058	ORD	EHEC
Q8XAN6	0.093	0.000	ORD	EHEC
Q8XB17	0.041	0.000	ORD	EHEC
Q8XB62	0.046	0.051	ORD	EHEC
Q8XBX8	0.073	0.000	ORD	EHEC
Q8XC86	0.228	0.029	PDR	EHEC
A0A1A9NR20	0.074	0.000	ORD	EPEC
A0A1E5MD86	0.121	0.000	NDP	EPEC
B7UH72	0.074	0.000	ORD	EPEC
B7UI20	0.027	0.000	ORD	EPEC
B7UI21	0.073	0.000	ORD	EPEC
B7UI22	0.045	0.000	ORD	EPEC
B7UI23	0.285	0.011	PDR	EPEC
B7ULW4	0.147	0.000	PDR	EPEC
B7ULW8	0.083	0.000	ORD	EPEC
B7UM88	0.592	0.320	IDP	EPEC
B7UM99	0.400	0.238	IDP	EPEC
B7UMA0	0.143	0.000	PDR	EPEC
B7UMA2	0.113	0.000	NDP	EPEC
B7UMA9	0.051	0.000	ORD	EPEC
B7UMC8	0.048	0.030	ORD	EPEC
B7UNX2	0.068	0.031	ORD	EPEC
B7UNX4	0.052	0.000	ORD	EPEC
B7UNX6	0.103	0.000	PDR	EPEC
B7UR60	0.068	0.070	ORD	EPEC
B7UR63	0.042	0.058	ORD	EPEC
Q05129	0.277	0.025	PDR	EPEC
Q7WRZ5	0.110	0.000	PDR	EPEC
Q8VLH6	0.073	0.000	ORD	EPEC
Q9EZE7	0.057	0.003	ORD	EPEC
A0A2X2U9X2	0.074	0.007	ORD	CR
A0A482PJC5	0.033	0.000	ORD	CR
D2TI12	0.076	0.163	ORD	CR
D2TI20	0.136	0.000	NDP	CR
D2TI21	0.119	0.000	NDP	CR
D2TI55	0.011	0.000	ORD	CR
D2TJZ4	0.074	0.030	ORD	CR
D2TK70	0.058	0.000	ORD	CR

Supplementary Information

D2TK72	0.098	0.000	ORD	CR
D2TKD5	0.043	0.020	ORD	CR
D2TKD7	0.877	0.399	IDP	CR
D2TKE1	0.268	0.044	PDR	CR
D2TKE8	0.470	0.250	IDP	CR
D2TKF1	0.078	0.000	ORD	CR
D2TKF8	0.050	0.000	ORD	CR
D2TM85	0.094	0.000	ORD	CR
D2TML3	0.102	0.000	PDR	CR
D2TQZ7	0.092	0.004	ORD	CR
D2TRX7	0.065	0.000	ORD	CR
D2TRX8	0.042	0.000	ORD	CR
D2TRY0	0.057	0.000	ORD	CR
D2TRY1	0.088	0.000	ORD	CR
D2TT36	0.027	0.000	ORD	CR
D2TT37	0.079	0.000	ORD	CR
D2TT38	0.083	0.000	ORD	CR
D2TTX8	0.079	0.000	ORD	CR
Q5DKN7	0.243	0.013	PDR	CR
Q5XMK8	0.070	0.000	ORD	CR

Table SI.2. C-Tir SAXS data collection and analysis information.

Acquisition	
Beamline – Facility	BM29 – ESRF
Wavelength (Å)	0.992
Sample-to-detector distance (m)	2.867
s range (Å ⁻¹)	0.00359 – 0.48896
Sample concentration (mg.mL ⁻¹)	10.0
HPLC system / SEC column	Shimadzu HPLC System / Superdex™ 200 Increase 3.2/300
Detector	Pilatus 1M
Temperature (K)	298.15
Overall parameters	
R _g (Å) (from P(r))	38.72 ± 0.42
R _g (Å) (from Guinier)	38.81 ± 0.40
D _{max} (Å)	128 ± 5
Software	
SEC-SAXS data integration	ScÅtter3
P(r)	GNOM 5.0
SASBDB accession code	SASDKH8

Table SI.3. Input sequences employed in the prediction of the 3D structure of the complex formed between Tir and TRAF6 using AlphaFold2.

Protein	Uniprot		Sequence
	accession code	Residues	
Tir	B7UM99	388-550	RRNQPAEQTTTTTHTVVQQQTGG NPAQGGTDATRAEDASLNRRDSQG SVASTHWSDSSEVVNPYAEVGGA RNSLSAHQPEEHIYDEVAADPGYSVI QNFSGSGPVTGRLIGTPGQGIQSTY ALLANSGLRLGMGGLTSGGESAVS SVNAAPTPGPVRFV
		388-398	RRNQPAEQTTT
		466-475	AHQPEEHIYD
		505-515	QGIQSTYALLA
TRAF6	Q9Y4K3	346-504	AQQCNGIYIWKIGNFGMHLKCQEEE KPVVIHSPGFYTGKPGYKLCMRLHL QLPTAQRCANYISLFVHTMQGEYDS HLPWPFQGTIRLTILDQSEAPVRQNH EEIMDAKPELLAFQRPTIPRNPKGFG YVTFMHLEALRQRTFIKDDTLLVRCE VSTRFDL

References

- [1] Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **7**, 539 (2011).
- [2] Procter, J.B., Carstairs, G.M., Soares, B., Mourão, K., Ofoegbu, T.C., Barton, D., Lui, L., Menard, A., Sherstnev, N., Roldan-Martinez, D., Duce, S., Martin, D.M.A., Barton, G.J. Alignment of Biological Sequences with Jalview. *Methods Mol Biol.* **2231**, 203-224 (2021).
- [3] Rice, P., Longden, I., Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276-7 (2000).
- [4] Valentini, E., Kikhney, A.G., Previtali, G., Jeffries, C.M., Svergun, D.I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43** (Database issue), D357-63 (2015).

