The Molecular Basis of Calcineurin Inhibition by Protein Regulators

By

Simina Grigoriu

B.S., Tufts University, 2007

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the

Degree of Doctor of Philosophy in the

Division of Biology and Medicine at Brown University

Providence, Rhode Island

May 2013

This dissertation by Simina Grigoriu is accepted in its present form

by the Division of Biology and Medicine as satisfying the

dissertation requirement for the degree of Doctor of Philosophy.

Date_____ _____

Wolfgang Peti, Ph.D., Advisor

Recommended to the Graduate Council

Date_____ _____

Sarah Delaney, Ph.D., Reader

Date_____ _____

William G. Fairbrother, Ph.D., Reader

Date_____ _____

Gerwald Jogl, Ph.D., Reader

Date_____ _____

Jonathan A. Ellman, Ph.D., Reader
Yale University

Approved by the Graduate Council

Date_____ _____

Peter M. Weber, PhD
Dean of the Graduate School

SIMINA GRIGORIU

Curriculum Vitae

EDUCATION

09/2007-05/2013 **Brown University,** Providence, RI
PhD in Structural Biology
Department of Molecular Biology, Cell Biology and Biochemistry
Advisor: Dr. Wolfgang Peti; Thesis: "The Molecular Basis of Calcineurin Inhibition by Protein Regulators"

09/2003-05/2007 **Tufts University,** Medford, MA
BS in Biochemistry, Biomedical Engineering, *cum laude*
Advisor: Dr. Marc d'Alarcao*;* Thesis with Highest Honors: "Synthesis of a quinone-based electrochemically degradable polymer cross-linking agent"

HONORS AND AWARDS

2013 Article featured on the cover of *PLoS Biology (*February 2013)

2010 Oliver Cromwell Gorton Arnold Biological Graduate Fellow

2007-09 NIH Predoctoral Training Grant in Molecular and Cell Biology and Biochemistry, two-time recipient

2009 National Science Foundation Graduate Research Fellowship Program Honorable Mention

2005-07 Dean's List, Tufts University

2006 Summer Scholars Program Research Apprenticeship Recipient, Tufts University

2005 Summer Undergraduate Research Experience Program, University of Massachusetts Medical School

2003 James A. Burley Scholarship Essay Contest, First Place Winner

PUBLICATIONS

1. **Grigoriu S**, Bond R, Cossio P, Chen JA, Ly N, Hummer G, Page R, Cyert MS, Peti W. (2013) The molecular mechanism of substrate engagement and immunosuppressant inhibition of Calcineurin. *PLoS Biology* 11(2): e1001492
2. Brown BL, Lord DM, **Grigoriu S**, Peti W, Page R. (2013) The E. coli toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter. *Journal of Biological Chemistry* 288(2): 1286-94
3. Eibl C, **Grigoriu S**, Hessenberger M, Wenger J, Puehringer S, Pinheiro AS, Wagner RN, Proell M, Reed JC, Page R, Diederichs K, Peti W. (2012) Structural and functional analysis of the NLRP4 pyrin domain. *Biochemistry* 18;51(37):7330-41

4. Kim Y, Wang X, Zhang XS, **Grigoriu S**, Page R, Peti W, Wood TK. (2010) Escherichia coli toxin/antitoxin pair MqsR/MqsA regulate toxin CspD. *Environmental Microbiology* 12(5):1105-21
5. Brown BL*, **Grigoriu S***, Kim Y, Arruda JM, Davenport A, Wood TK, Peti W, Page R. (2009) Three Dimensional Structure of the MqsR:MqsA Complex: A Novel TA Pair Comprised of a Toxin Homologous to RelE and an Antitoxin with Unique Properties. *PLoS Pathogens* 5(12): e1000706. **[*co-first authors]**
6. Zaidi SK, Pande S, Pratap J, Gaur T, **Grigoriu S**, Ali SA, Stein JL, Lian JB, van Wijnen AJ, Stein GS (2007). Runx2 deficiency and defective subnuclear targeting bypass senescence to promote immortalization and tumorigenic potential. *PNAS* 11;104(50):19861-6

PUBLIC DATA FOR DISTRIBUTION

1. Protein Data Base (PDB): 3HI2, 4F0Z
2. Biological Magnetic Resonance Data Bank (BMRB): 17009

PRESENTATIONS

09/2012 International Conference on Frontiers in Pharmaceutical Sciences
University of Rhode Island, Kingston, RI
Poster: "Structural Basis for the Design of New Immunosuppressants"

07/2012 American Crystallographic Association
Boston, MA
Poster: "Structural Basis for the Design of New Immunosuppressants"

01/2012 New York Structural Biology Discussion Group Winter Meeting
Manhattan, NY
Poster: "Structural Basis for Calcineurin as an Immunosuppressant Target"

02/2011 MCB Department Data Club Series
Brown University, Providence, RI
*Oral Presentation. "*Signaling in the Innate Immune System"

01/2011 Keystone Symposia: Frontiers of NMR in Biology
Big Sky, Montana
Poster: "Signaling in the Innate Immune System"

11/2009 MCB Department Data Club Series
Brown University, Providence, RI.
Oral Presentation: "The Structural Basis of Toxin-Antitoxin Interactions in *E. coli* Biofilms"

04/2007 Tufts Undergraduate Research & Scholarship Symposium
Tufts University, Medford, MA
Poster: "Organic Synthesis of a Quinone-Based Electrochemically Degradable Polymer Cross Linking Agent"

08/2005 Summer Undergraduate Research Experience Program Poster Session
University of Massachusetts Medical School, Worcester, MA
Poster: "The Role of the Transcription Factor Runx2 in the DNA Damage Response of Osteoblasts"

PROFESSIONAL MEMBERSHIPS

2012-present	American Crystallographic Association, member
2010-present	AAAS, member
2010-present	Faculty of 1000, associate member
2006-2007	American Chemical Society, *class representative*, Tufts University
2004-2007	Tufts University Biomedical Engineering Club, member

RESEARCH EXPERIENCE

Graduate research:

2008-present **Dissertation research**
Brown University, Providence, RI

Laboratory of Dr. Wolfgang Peti (structural biology)

- Determined the molecular basis for substrate recognition by the phosphatase calcineurin, using a combination of X-ray crystallography, protein NMR spectroscopy, Isothermal Titration Calorimetry and biochemical analysis
- Determined the structure of the novel MqsR:MqsA *E. coli* toxin:antitoxin protein complex using X-ray crystallography, and studied its role in bacterial persistence using bacterial cell growth assays and biochemical experiments
- Investigated the role of the Nod-Like Receptors (NLRs) NRLP1, NLRP3, NLRP4, NLRC5, and AIM2, and of the adaptor protein ASC, in inflammasome formation using NMR spectroscopy and biochemical and biophysical methods

Undergraduate research:

09/2005-08/2007 **Honors Thesis research**
Tufts University, Medford, MA

Laboratory of Dr. Marc d'Alarcao (medicinal chemistry)

- Independently designed and executed the organic synthesis of an electrochemically degradable polymer cross-linking agent with drug delivery applications, using organic chemistry techniques and instrumentation, including small molecule NMR spectroscopy and mass spectroscopy

06/2005-08/2005 **Undergraduate Summer Research**
University of Massachusetts Medical School, Worcester, MA

Laboratory of Dr. Gary Stein (cell biology)

- Examined the role of Runx2 transcription factor in the DNA damage response of osteoblasts using cell biology methods, including *in situ* fluorescence spectroscopy, mammalian cell culture, and molecular biology approaches

TEACHING AND MENTORING

2008-present **Mentor for undergraduate students**
Brown University, Providence, RI

- Supervised two undergraduate researchers: taught laboratory techniques and experimental design
- Guided Honors thesis research and writing of one student

09/2008-12/2008 **Teaching Assistant**
Brown University, Providence, RI

- Taught and supervised Advanced Biochemistry laboratory course (BIOL 1270)
- Introduced undergraduate juniors and seniors to molecular biology techniques, protein expression and purification, and biochemical enzyme activity assays

PROFICIENCIES: TECHNIQUES, INSTRUMENTATION, AND SOFTWARE

Protein production and optimization:
Techniques: molecular cloning; bacterial protein expression; labeled protein expression (^{13}C,^{15}N and selenomethionine), small-scale cell-free protein expression; metal-affinity, ion exchange and size exclusion chromatography; protein denaturation and refolding; preparation of protein-protein complexes; optimization of protein samples for crystallography and NMR. Instrumentation: GE AKTA Protein Purification Systems.

Crystallography:
Techniques: crystallization construct optimization using NMR techniques; crystallization methods including vapor and microbatch crystallization, fine/additive screening, microseeding, macroseeding; protein-protein complex crystallization; crystal harvesting and cryoprotection; data collection using Rigaku X-ray systems; data collection using synchrotron radiation including microfocus and undulator beam lines (National Synchrotron Light Source, Brookhaven National Laboratories; beam lines X6A, X25 and X29); home source (ACTOR) and synchrotron crystal mounting robotics for screening and data collection (ALS puck system). Instrumentation: Phoenix Liquid Handling System and Cryscam Digital Microscope (Art Robbins); *Rigaku system 1:* FRE+ Superbright microfocus X-ray generator/Varimax-HF confocal optics/Saturn 944HG CCD detector/ACTOR crystal mounting robot; *Rigaku system 2:* Micromax-007 X-ray generator/R-AXIS IV++ imaging plate detector/X-stream 2000 low temperature system. Software: JDirector (home source screening and data collection); HKL2000 (data processing); Phenix crystallography suite (phase determination, model refinement, and analysis); CCP4 software suite (model refinement and analysis); Coot (model building); Pymol (model visualization and figure generation).

NMR spectroscopy:
Techniques: Sequence-specific backbone assignment; protein-protein interaction studies. Instrumentation: Bruker AVANCEII 500 MHz and Avance IIIHD 850 MHz spectrometers. Software: Topspin 2.1/3.1 (data acquisition and processing), CARA (sequence-specific backbone assignment), Bruker Dynamics Center 2.0.

Protein biophysics:
Isothermal Titration Calorimetry (Microcal VP-ITC); Circular Dichroism Spectropolarimetry (Jasco J-815); Dynamic Light Scattering (Viscotek 802); Electrophoretic Mobility Shift Assays; Western blotting.

Molecular Biology
Cloning and expression of protein targets using multiple vector systems; polymerase chain reaction (PCR); site-directed mutagenesis.

Acknowledgements

I owe my deepest gratitude to my advisor, Dr. Wolfgang Peti, who has been committed to helping me succeed since I first joined his lab. He has always held his students up to the highest possible standards, and though meeting his expectations has been challenging at times, I look back on my experience knowing that I was fortunate to have had such a dedicated mentor. I am thankful for the countless hours that Wolfgang has devoted toward my development as a scientist, writer, speaker, and individual. I also want to express my warmest gratitude to Dr. Rebecca Page, who has taught me everything I know about X-ray crystallography. She has been a second mentor to me throughout my graduate education, and her passion and drive for excellence have helped me develop into a stronger, more confident scientist.

I would also like to thank the members of my dissertation committee, Dr. Gerwald Jogl, Dr. Will Fairbrother, and Dr. Sarah Delaney, for taking the time to read my research proposal and dissertation, and for providing important input along the way. I am also grateful to Dr. Jonathan Ellman for agreeing to serve as my external reader.

I am indebted to all current and past members of the Peti and Page labs for providing a fun, intellectually stimulating work environment. Many colleagues I have had over the years have become good friends. I especially want to thank Drs. Mike Ragusa, Barbara “Violet” Dancheck, and Anderson Pinheiro, for the kind advice and encouragement they have provided over the years. I will also forever share a bond with Dorothy Koveal and Dana Francis, who have gone

through this experience side by side with me since the first day. Finally, I must thank Dr. Nichole O'Connell, who is like a big sister, for her genuine kindness and friendship and for always reminding me that I should be proud of my achievements and knowledge.

I am indebted to our collaborators, Dr. Martha Cyert and Rachel Bond, without whom the calcineurin-A238L project would not have existed, and whom I have come to know quite well after hours of stimulating scientific discussions and the exchange of countless emails. I am also grateful to Dan Miller, Antoinette Oot, and Antoni Tortajada of the Peti and Page labs for their help with the calcineurin-RCAN1 project.

Lastly, but perhaps most importantly, I am deeply grateful to have a family who has supported and loved me unconditionally. I cannot imagine what life would be like without my parents, the two people I know I can count on even if everything else falls apart. Thank you to Mike, who is also my family, and who helps me see the good in everything and everyone, including myself.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AKAP79	A-kinase anchor protein 5
ATP	adenosine triphosphate
CaM	calmodulin
CBD	calcineurin-binding domain (of A238L)
CN	calcineurin
CNA	calcineurin A subunit
CNB	calcineurin B subunit
CRAC	Ca^{2+} release-activated Ca^{2+} channel
CSA	cyclosporin A
DFT	density functional theory
DSCR1	Down syndrome critical region 1
E	enzyme
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ES	enzyme-substrate complex
FKBP	FK506 binding protein
IP_3	inositol1,4,5-trisphosphate
IPTG	isopropylthio-β-D-galactoside
ITC	isothermal titration calorimetry
K_a	association constant
K_{cat}	catalytic rate constant
k_{cat}/K_M	enzyme catalytic efficiency

K_{eq}	equilibrium constant
K_M	Michaelis constant
KSR2	kinase suppressor of Ras 2
M	macromolecule
MAPK	mitogen activated protein kinase
MD	molecular dynamics
MX	macromolecule-ligand complex
n	stoichiometry
NCX	Na/Ca exchanger
NFAT	nuclear factor of activated T cells
NMR	nuclear magnetic resonance
P	product
PKA	protein kinase A
PMCA	plasma membrane Ca^{2+} ATPase
pNPP	para-nitrophenyl phosphate
PP1	protein phosphatase 1
PP2A	protein phosphatase 2A
PP2B	protein phosphatase 2B
PP3	protein phosphatase 3
PPP	phosphoprotein phosphatase
pSer	phospho-Serine
PSP	protein Ser/Thr protein phosphatase
RCAN1	regulator of calcineurin 1

RCAN1	regulator of calcineurin 1
RIA	radioimmunoassay
RII	RII regulatory subunit of PKA
RRM	RNA recognition motif
S	substrate
Ser/Thr	Serine/Threonine
SERCA	smooth endoplasmic reticular Ca^{2+} ATPase
SIM ITC	single injection method isothermal titration calorimetry
SOCE	store-operated Ca^{2+} entry
SPR	surface plasmon resonance
v_0	reaction rate
X	ligand
ΔC_p	change in specific heat of binding
ΔG	change in Gibbs free energy
ΔH	change in enthalpy
ΔS	change in entropy

Abstract

Ser/Thr phosphatases dephosphorylate their targets with high specificity, yet the structural determinants of substrate recognition and dephosphorylation are poorly understood. Calcineurin (CN) is a conserved Ca^{2+}/calmodulin-dependent Ser/Thr phosphatase that regulates many cellular functions and is the direct target of the immunosuppressant drugs FK506 and cyclosporin A (CSA). CN dephosphorylates the NFAT family of transcription factors, which then activate the transcription of genes required for various cellular functions. This dissertation describes the structural basis for CN-substrate recognition and provides insight into the molecular mechanisms used by CN inhibitors.

To investigate CN-substrate recognition we used X-ray crystallography, biochemistry, molecular dynamics simulations and *in vivo* experiments to study A238L, a viral protein inhibitor of CN and a substrate mimic. The 1.7 Å structure of the A238L-CN complex provides the first structural description of how CN recognizes its substrates via the conserved "LxVP" motif, and enabled molecular dynamics simulations that provided the first description of substrate dephosphorylation by a Ser/Thr phosphatase. The structure also revealed that A238L competitively inhibits CN by occupying the PxIxIT and LxVP substrate recognition sites to prevent substrate docking, leaving the catalytic center fully accessible. This study establishes that FK506 and CSA also bind to the CN LxVP site to prevent LxVP-mediated substrate recognition. This highlights the importance of this interaction for substrate dephosphorylation and provides a structural basis for the development of new CN inhibitors.

To gain further insight into CN regulation, we used NMR spectroscopy and isothermal titration calorimetry to characterize CN inhibition by the endogenous regulator RCAN1. NMR studies of free RCAN1 showed that it belongs to a family of proteins known as intrinsically disordered proteins, which are characterized by their lack of tertiary structure. However, we identified two partially populated α-helices in RCAN1, which are likely important for the ability of RCAN1 to act as both an inhibitor and an enhancer of CN activity. RCAN1 is overexpressed in individuals with Down syndrome, and its inhibition of CN directly correlated with symptoms of this disease. Therefore, these findings significantly advance our understanding of the underlying mechanism of Down syndrome.

Chapter 1. Introduction

1.1. Perspectives and overview

Calcineurin (CN), also known as protein phosphatase 2B (PP2B) and protein phosphatase 3 (PP3), is a calcium (Ca^{2+}) - activated phosphatase that translates Ca^{2+} input signals into a large variety of cellular responses, including development, apoptosis, memory formation, transcriptional regulation, cardiac function and activation of the immune response. Aberrant CN function is associated with a host of diseases, including Alzheimer's disease, Down syndrome, and autoinflammatory disorders [15]. Despite its biological importance, the details driving CN-substrate specificity are not fully understood. CN belongs to the phosphoprotein phosphatase (PPP) subfamily of protein Ser/Thr protein phosphatases (PSPs), whose members also include protein phosphatase 1 (PP1) and protein phosphatase 2A (PP2A). The catalytic domains of PPPs share high structure and sequence conservation, and many PPPs achieve substrate specificity by interacting with a variety regulatory proteins or subunits to form multi-subunit holoenzymes. The CN catalytic subunit (CNA), however, interacts with only one regulatory subunit that is invariable [16]. How this single CN holoenzyme discriminates between its various substrates remains poorly understood, even after decades of CN research.

CN is a prominent medical target, as the immunosuppressant drugs FK506 [1] and cyclosporin A (CSA) [17, 18] function by directly inhibiting CN activity. Not only have these drugs revolutionized transplantation medicine over the past three decades, but they also enabled functional studies investigating the biology of CN [19-21]. Over the past two decades, only limited information

describing CN-substrate interactions has emerged, and additional *in vitro* and *in vivo* studies are necessary to fully understand the molecular and cellular determinants of CN-substrate interactions.

Understanding CN function and the mechanism by which immunosuppressants inhibit CN is crucial for the development of new drugs. Current CN inhibitors have severe side effects, including neurotoxicity, kidney damage, anemia, and the onset of diabetes, and are only used in life-or-death situations when no other options are available [22-24]. The structure-based design of more specific CN inhibitors is a critical medical need which can only be addressed once an atomic-level structural description of CN-substrate binding and regulation has been achieved.

1.2. Cellular signaling

All biological processes are regulated by external stimuli. Signal transduction is used by all cells to transmit information and activate specific cellular processes such as apoptosis, cell differentiation, or the secretion of proteins and chemokines, ultimately giving rise to more global biological events like immune activation or organ development [25]. In most cases, an external signal, such as antigen binding to a T-cell receptor, gives rise to a cascade of events that ends with the transcription of a particular set of genes. Most signaling pathways are either activated by external factors like chemokines, growth factors that bind to extracellular receptors, ions such as Ca^{2+}, K^{+}, Mg^{2+} and Cl^{-} that enter the cell through membrane channels and pumps, or even by a combination of multiple external stimuli [26]. Further transmission of a signal typically relies on highly

controlled phosphorylation cascades (a series of sequential phosphorylation events) that control and establish the transfer of information [27-29]. CN is at the cusp of the two most prevalent signal transduction methods: Ca^{2+} and phosphate signaling. It is activated by increases in cytosolic Ca^{2+} concentrations, and propagates a particular message by dephosphorylating its substrates [30].

1.2.A. Calcium signaling

Ca^{2+} is a prevalent and versatile intracellular signal in all eukaryotic cells. It is a secondary messenger that plays a prominent role in nearly all biological functions, ranging from immune system activation to development. High cytoplasmic Ca^{2+} concentrations are toxic because Ca^{2+} binds water only weakly, and as a result precipitates cellular phosphate [26]. Increases in intracellular Ca^{2+} concentration associated with signaling events must therefore be transient and tightly controlled. To this end, various membrane-associated pumps and transporters, such as the plasma membrane Ca^{2+} ATPase (PMCA), smooth endoplasmic reticular Ca^{2+} ATPase (SERCA), and Na/Ca exchanger (NCX), among others, constantly work to export cytosolic Ca^{2+} to extracellular space or to internal storage compartment like the endoplasmic reticulum (ER) [26]. In resting cells, cytoplasmic intracellular Ca^{2+} levels are kept at ~50-100 nM, while Ca^{2+} concentrations in extracellular space or intracellular storage compartments can be up to 20,000-fold higher (extracellular ~2 mM, ER ~500 μM) [31], creating an enormous concentration gradient. Various types of membrane Ca^{2+} channels and transporters have evolved to move Ca^{2+} ions across plasma membranes. The mechanism of Ca^{2+} cytosolic influx varies across different cell and tissue

types, and different mechanisms of Ca^{2+} transport are activated in response to specific signals. Signals that trigger cytosolic Ca^{2+} influx vary and range from changes in membrane potential to the binding of various ligands, such as antigen molecules and chemical messengers, to extracellular receptors or to the Ca^{2+} channels themselves [26, 31].

An example of the importance of Ca^{2+} signaling is seen during the process of T-cell activation. Here, Ca^{2+} signaling is activated by a mechanism called

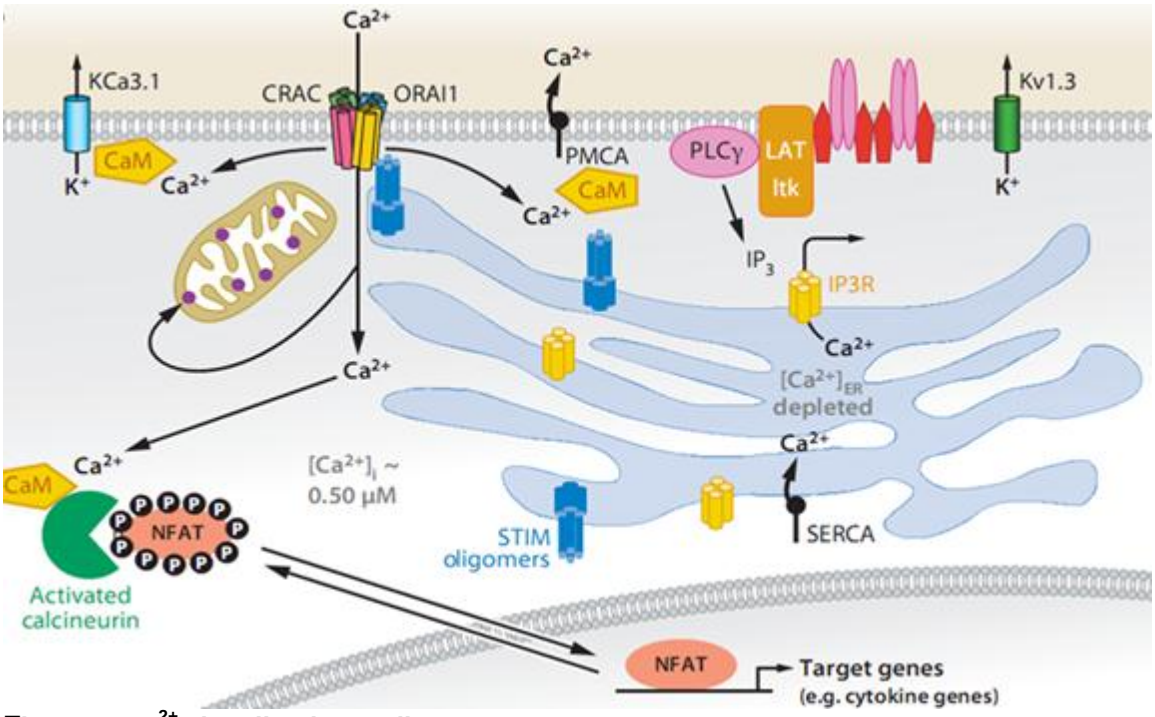


Figure 1: Ca^{2+} signaling in T-cells.
Na^{+}/Ca^{2+} and K^{+}/Ca^{2+} pumps and transporters on the plasma and ER membranes, including PMCA transporter on the plasma membrane and SERCA pump on the ER membrane, maintain homeostasis by constantly exporting Ca^{2+} to extracellular space and intracellular storage compartments like the ER. Antigen binding to the T-cell receptor (pink) causes assembly of a scaffolding complex and release of IP_3 protein from the plasma membrane. IP_3 binds to the IP_3 receptor (IP3R) (yellow, on the ER membrane) and Ca^{2+} is released from the ER via the IP_3 receptor and pumped out of the cell by PMCA. Ca^{2+} depletion from ER stores causes the ER membrane protein STIM1 (blue) to multimerize and recruit ORAI proteins to CRAC channels on the plasma membrane, which then open to release Ca^{2+} into the cell. Intracellular Ca^{2+} activates a host of enzymes, including CN, and additional downstream signaling leads to transcription of immune response genes like cytokines. SOCE, store-operated Ca^{2+} entry; PMCA, plasma membrane Ca^{2+} ATPase; SERCA, sarco-endoplasmic reticulum Ca^{2+}-ATPase; ER, endoplasmic reticulum; IP3, inositol (1,4,5) trisphosphate; ER, endoplasmic reticulum; CN, calcineurin. Adapted from Hogan et al (2010) Annu. Rev. Immunol.

store-operated Ca^{2+} entry (SOCE) (Figure 1). First, antigen binding to the T-cell receptor triggers the release of the protein inositol 1,4,5-trisphosphate (IP_3) from the plasma membrane, and IP_3 binds to the IP_3 receptor to release Ca^{2+} from internal endoplasmic reticulum (ER) stores. Released Ca^{2+} is pumped out of the cell by plasma membrane Ca^{2+} ATPase (PMCA) transporters, and the depletion of ER Ca^{2+} stores triggers Ca^{2+} cellular entry through Ca^{2+} release-activated Ca^{2+} (CRAC) channels. CRAC channels maintain a sustained Ca^{2+} influx for several hours, during which they activate a host of Ca^{2+}-dependent enzymes, including CN, and upregulate the transcription of immune response genes in the case of lymphocytes [31].

Eukaryotes have evolved hundreds of Ca^{2+}-activated proteins that bind Ca^{2+} with affinities ranging from the nM to the mM range. Most Ca^{2+} sensors bind Ca^{2+} via a structural motif called the EF-hand. EF-hands typically occur in pairs and have a wide range of affinities for Ca^{2+}. The EF-hand motif adopts a helix-loop-helix topology resembling a thumb and forefinger, with a Ca^{2+}-binding loop in the middle (Figure 2A). A typical EF-hand Ca^{2+}-binding loop consists of 12 residues arranged in the pattern X•Y•Z•-Y•-X••-Z, where X, Y, Z, -Y, -X, and –Z represent amino acid residues (and sometimes water molecules) coordinating the Ca^{2+} atom, and dots represent intervening residues. Each Ca^{2+}-binding loop coordinates one Ca^{2+} atom via electrostatic interactions with seven oxygen atoms arranged in pentagonal bipyramidal geometry around the Ca^{2+} atom (Figure 2B). Calmodulin (CaM), which is the archetypal Ca^{2+}-binding protein, has four EF-hands. CaM is a highly conserved 17 kDa ubiquitous Ca^{2+} sensor that

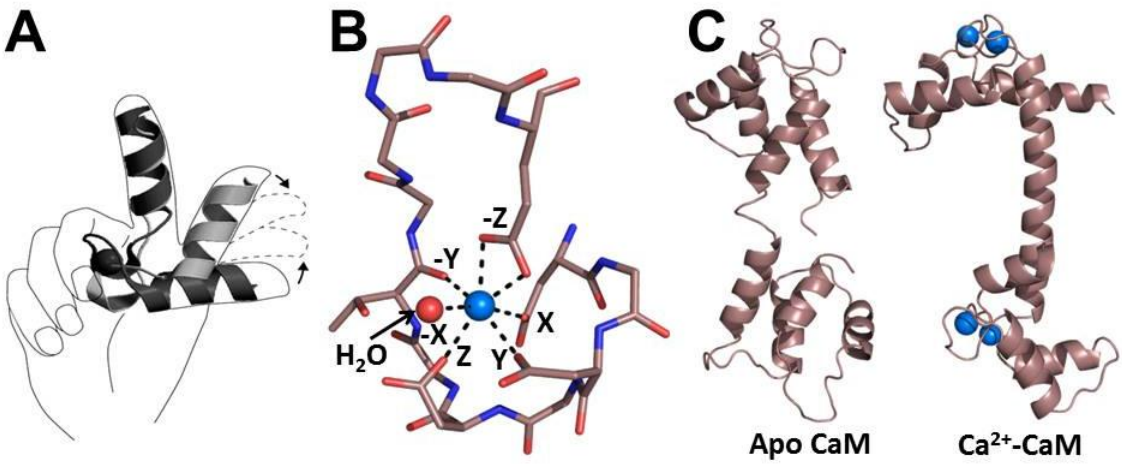


Figure 2: The EF-hand motif and calmodulin (CaM).
A) The EF-hand motif consists of two α-helices joined by a loop, and resembles a hand with the thumb swinging outward upon Ca^{2+} binding to the loop. B) Detailed view of Ca^{2+} (blue sphere) coordination by a typical EF-hand loop, using an EF-hand from CaM as an example. One Ca^{2+} atom is coordinated via seven electrostatic interactions with sidechain and backbone oxygen atoms and one water molecule (red sphere). C) Apo-CaM [7] and Ca^{2+}-bound CaM [11]. CaM binds four Ca^{2+} ions via its four EF-hands, and the ensuing conformational changes promote its binding to other proteins. Adapted in part from Lewit-Bentley et al (2000) Curr. Op. Str. Biol.

acts as an adaptor protein for a vast number of enzymes during Ca^{2+} signaling. Its four EF-hands are organized into two globular domains separated by a flexible linker. The X•Y•Z•-Y•-X••-Z pattern of each CaM EF-hand is Asp•Asp•Asp•Thr•H_2O••Glu, with oxygens from the Asp and Glu sidechains, Thr backbone, and from one water molecule coordinating one Ca^{2+} atom [26] (Figure 2B). Ca^{2+} binding to all four EF-hands induces large structural rearrangements in CaM, which exposes hydrophobic regions within CaM and promotes its binding to and activation of Ca^{2+}-activated proteins (Figure 2C). The two EF-hands located at the C-terminal lobe of CaM bind Ca^{2+} at least 10-fold more tightly than the N-terminal lobe EF-hands, with reported dissociation constants of approximately 2 μM and 0.2 μM respectively [32-34].

1.2.B. Phosphorylation

Protein phosphorylation is a widespread post-translational modification that regulates virtually all biological processes. During a phosphorylation reaction, a protein kinase catalyzes the removal of a phosphate ion from an ATP molecule and covalently attaches it to a Ser, Thr or, more rarely, Tyr sidechain (86.4%, 11.8%, 1.8% relative phosphorylation of Ser, Thr and Tyr residues) [25]. Phosphorylation alters protein charge and shape and thus influences protein function. Protein phosphorylation is an efficient and dynamic method of signal transduction, because it is a rapid and reversible process. Thousands of proteins (~17,000) [35] undergo phosphorylation, and it is therefore not surprising that kinases account for 2% of all cellular proteins. Most signaling pathways, like the mitogen activated protein kinase (MAPK) pathway, which is activated when a growth factor molecule binds to an external cellular receptor, rely on sequential phosphorylation events called phosphorylation cascades to modulate vital cellular processes like cell cycle progression and transcription (Figure 3). Phosphate hydrolysis is catalyzed by phosphatases, and is crucial for maintaining cellular homeostasis following signaling events. In fact, the maintenance of proper protein phosphorylation levels is so crucial for cellular function that its dysregulation is linked to a multitude of disease phenotypes, including cancer, diabetes, and neurodegenerative and inflammatory disorders [27, 36].

Of the 518 kinases encoded by the human genome [37], at least 428 are Ser/Thr protein kinases [36, 38], making this one of the largest known protein

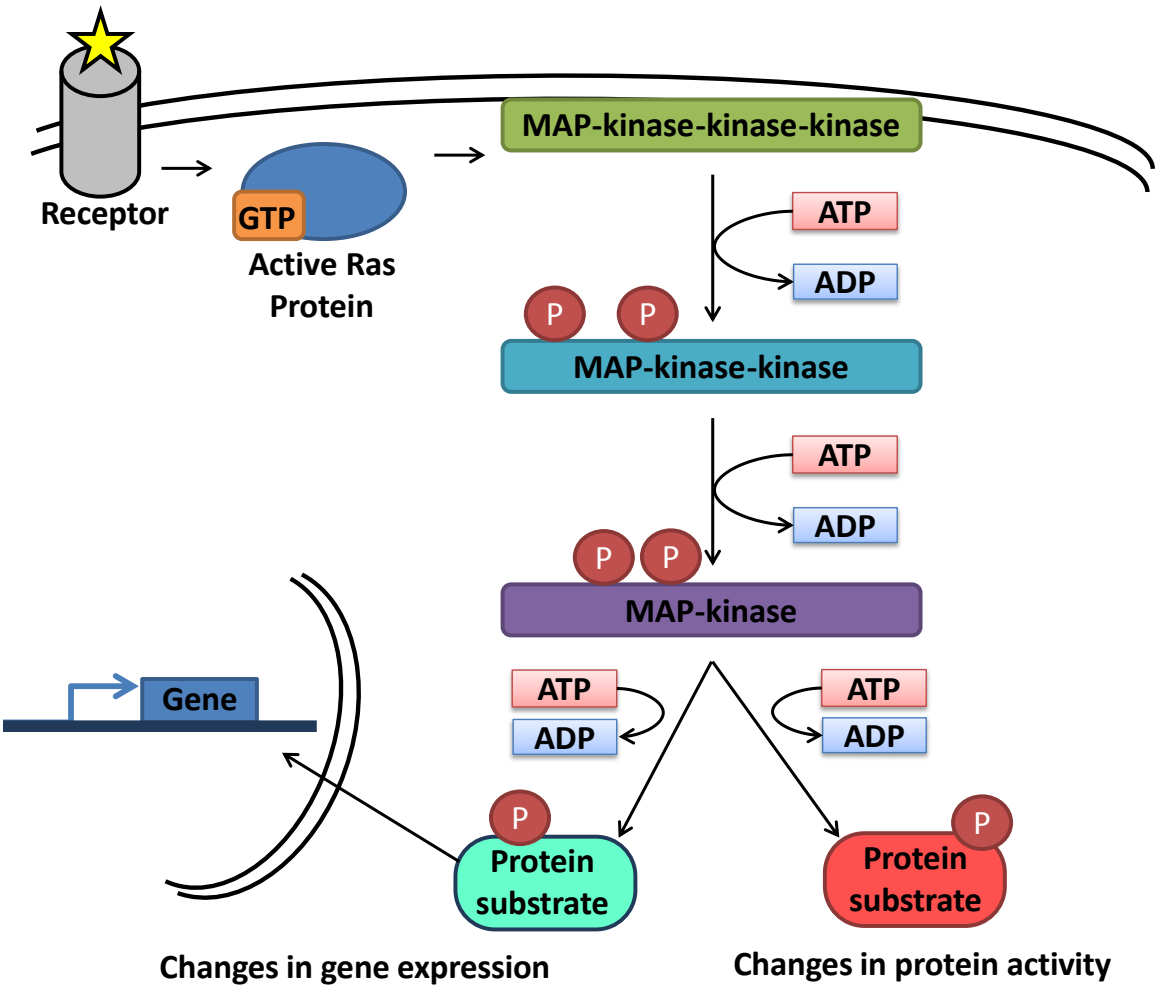


Figure 3: The mitogen activated protein kinase (MAPK) signaling pathway.
A series of phosphorylation events result in changes in gene expression or protein activity associated with specific external stimuli.

families. Surprisingly, the protein Ser/Thr phosphatase (PSP) family is smaller by an entire order of magnitude, with only ~30 members [38]. Each PSP, therefore, must discriminate between 10-fold more substrates at any given time than its opposing kinase. As a result, most PSPs have evolved to function as holoenzymes, forming complexes between their catalytic subunit and an array of regulatory subunits or targeting proteins [39, 40] that direct substrate specificity.

For example, PP1, which itself consists of only a catalytic subunit, forms holoenzyme complexes with more than 200 known regulatory proteins [41].

PP2A also forms holoenzymes, always consisting of a heterodimeric core enzyme (the scaffolding subunit A and the catalytic subunit C) that associates with a number of variable regulatory subunits (subunit B) (Figure 4). Interestingly, despite having a large array of substrates, CN only forms a single holoenzyme that must dephosphorylate all substrates. Although CN has been studied for decades, the current understanding of how it achieves substrate specificity remains incomplete.

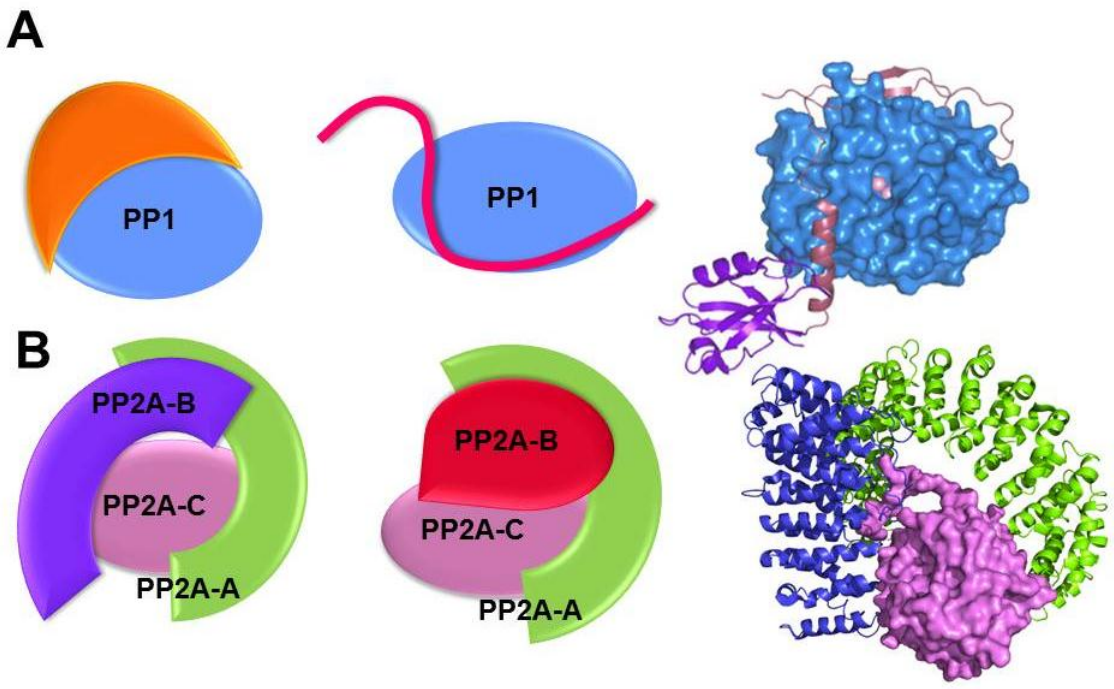


Figure 4: The various holoenzymes of PP1 and PP2A.
A) Protein phosphatase 1 (PP1) catalytic subunit (blue) interacts with various targeting proteins to form holoenzymes. On the right is the PP1:spinophilin crystal structure [5]. B) Protein phosphatase 2A (PP2A) catalytic subunit C (pink) and scaffolding subunit A (green), shown with variable regulatory B subunits. On the right is the PP2A holoenzyme with the B56 regulatory subunit [13].

1.3. Calcineurin

1.3.A. Domain architecture and 3D structure of calcineurin

Biochemical and structural studies have led to the detailed characterization of many physical properties of calcineurin (CN). In the resting state, CN is a heterodimer comprised of two subunits: a catalytic subunit (calcineurin subunit A; CNA) that is ~60 kDa in size, and a "regulatory" subunit (calcineurin subunit B; CNB) of ~19 kDa. CNB, unlike the highly variable B subunit of PP2A, is invariable and is referred to as a "regulatory" domain not because it targets CN to specific substrates but because it plays an important role in Ca^{2+}-sensing and calcineurin activation. In the presence of increasing intracellular Ca^{2+} levels, CN forms a holoenzyme consisting of the CNA/CNB heterodimer bound to four Ca^{2+} ions and to the Ca^{2+}-binding protein calmodulin (CaM).

The CN catalytic subunit, CNA, is comprised of an N-terminal catalytic domain (residues 1-342) followed by a regulatory region at its C-terminus (residues 343-521) (Figure 5). This regulatory region contains a CNB-binding

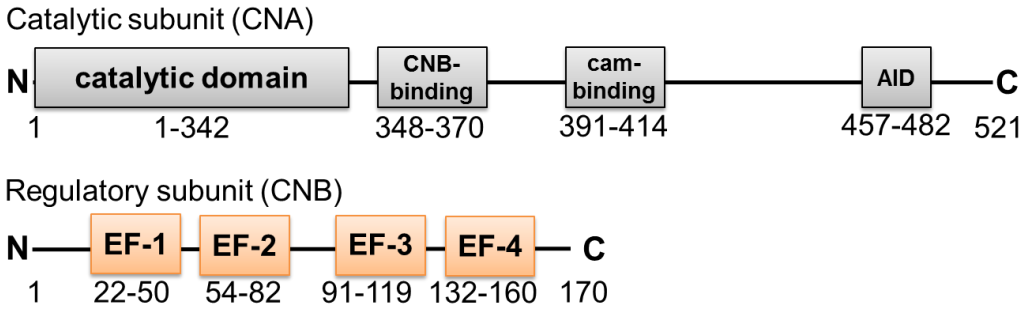


Figure 5: Domain architecture of calcineurin.
The A subunit of CN (CNA) has a catalytic domain and three regulatory elements: the B subunit (CNB) -binding domain, the calmodulin (CaM) -binding domain, and the autoinhibitory domain (AID). CNB has two domains, each composed of two Ca^{2+}-binding EF-hand structural motifs. Each EF-hand can bind one Ca^{2+} atom.

region (residues 348-370), a CaM-binding domain (residues 391-414), and a C-terminal autoinhibitory domain (AID) (residues 457-482) that blocks enzymatic activity by occluding the active site [1, 42]. CNB shares 35% sequence identity with calmodulin (CaM) and, like CaM, contains four EF-hand motifs [1, 42] (Figure 5).

The crystal structure of CN (Figure 6) reveals a globular catalytic domain. An extended α-helix (α-helix 14) called the CNB-binding helix (BBH) serves as the docking site for CNB, and interacts with CNB via a network of hydrophobic contacts. The AID forms a short α-helix that sits over the active site of CN, and is held in place by electrostatic interactions with residues surrounding the CNA active site. Electron density is missing for the CaM-binding region located between the BBH and the AID, indicating that when CN is in the resting state this region is likely disordered [1]. In the presence of excess cellular Ca^{2+}, all four EF-hands of CNB are loaded with Ca^{2+}. The two Ca^{2+} ions bound to the C-terminal

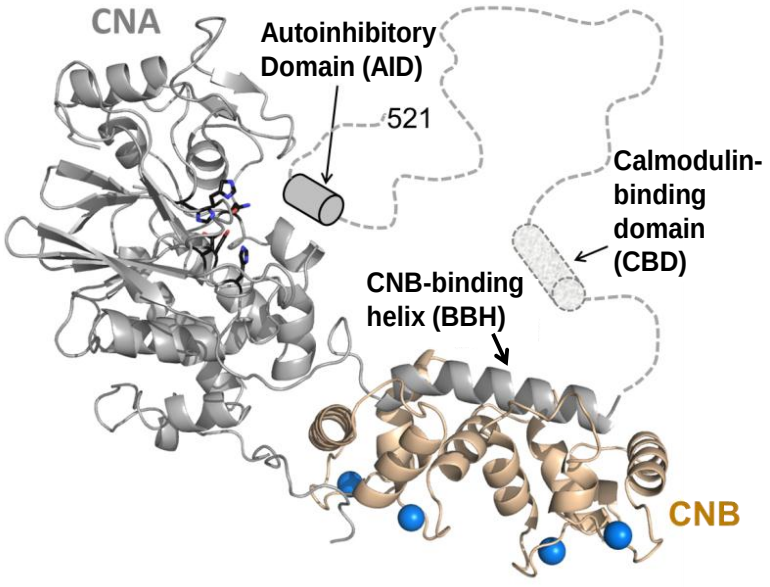


Figure 6: Structure of CN. CNA is shown in grey and CNB is shown in beige. Active site residues are shown as black sticks, and Ca^{2+} ions bound to the four EF-hand motifs of CNB are shown as blue spheres. The catalytic domain, CNB-binding helix (BBH), calmodulin-binding domain (CBD), and autoinhibitory domain (AID) are illustrated. The region between the BBH and AID, which includes the calmodulin binding domain, lacks electron density in the crystal structure [1] and is believed to be mostly disordered.

lobe (EF-hands 3 and 4) bind with higher affinity than the Ca^{2+} bound at the N-terminal lobe (nM vs. μM K_d range), and are therefore thought to be structurally important and remain bound even when CN is in the resting state [43, 44].

The CNA catalytic domain adopts a β-sandwich fold, with two β-sheets located at its center and a combination of α-helices and β-strands surrounding the β-sandwich (Figure 7). At the active site, several catalytic residues coordinate two metal ions: Fe^{3+} is coordinated by Asp90, His92, and Asp118 of CNA, and Zn^{2+} is coordinated by Asp118, Asn150, His199, and His 281 of CNA. During the dephosphorylation reaction, the active site metals activate a water molecule for

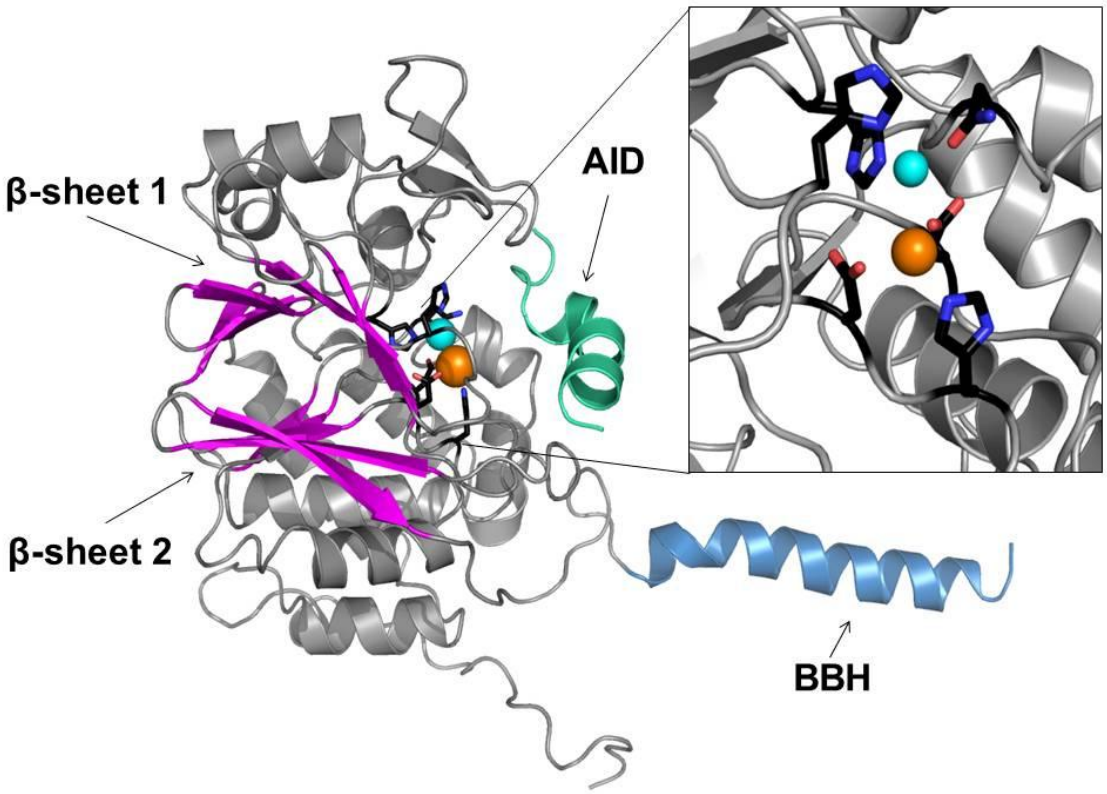


Figure 7: The CN catalytic subunit (CNA).
The CNA subunit contains β-sheets 1 and 2 (pink), a CNB-binding helix (BBH) (blue), active site residues (black sticks), and the autoinhibitory domain (AID) (teal). The AID can be seen capping the active site to block substrate access. A close-up view of the active site is shown to the right. Two metals are coordinated by catalytic residues at the active site: Fe^{3+} (orange) and Zn^{2+} (cyan) [1].

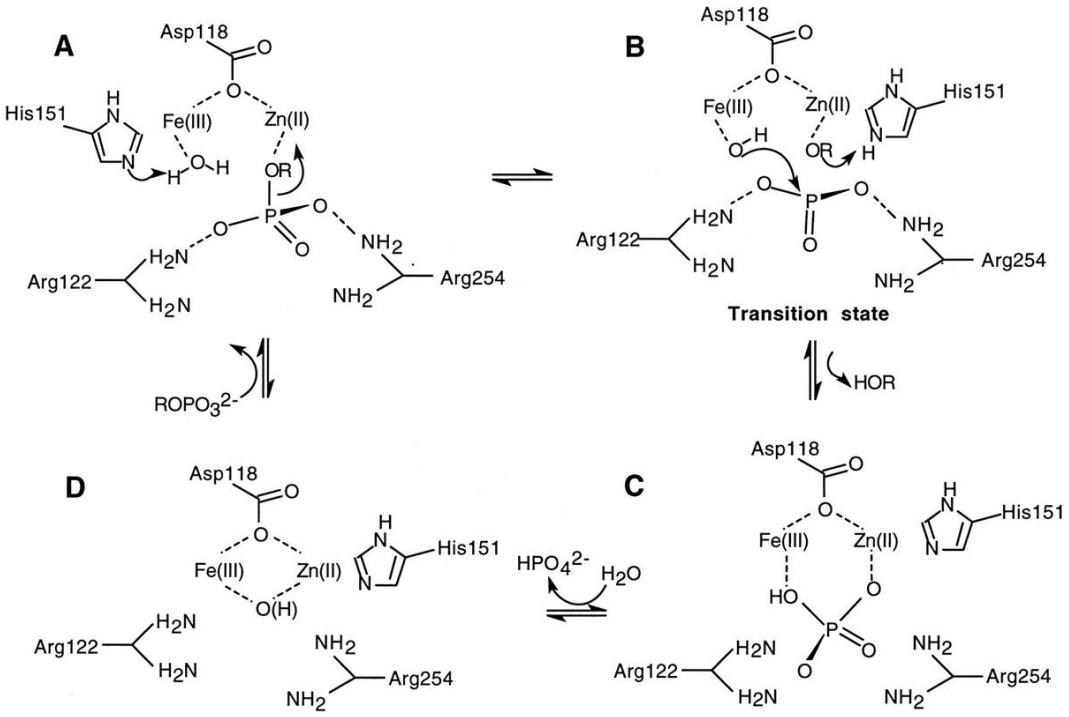


Figure 8: Proposed catalytic mechanism of CN.
His151 functions as a general base and a general acid. A) His151 acts as a general base to deprotonate a water molecule that is coordinated by Fe^{3+} to form hydroxide. The phosphate group is coordinated by Arg122 and Arg254 and bond cleavage occurs at the leaving group. B) In the transition state, the hydroxide (activated water molecule) forms a bond with the cleaved phosphate molecule. The leaving group is protonated by His151, which now acts as a general acid, and leaves the active site. C) During the product-inhibited state, the phosphate molecule is coordinated by the two metal ions. D) The catalytic center is regenerated as a water molecule displaces the pohsphate. Adapted from Rusnak and Mertz (2000) Physiol. Rev.

nucleophilic attack on the phosphorous atom of the substrate (Figure 8) [1, 45]. The catalytic domain of CNA closely resembles that of other PPP family members, especially PP1 and PP2A (Figure 9), as expected given the high sequence conservation of this family (39% amino acid identity is shared between CN and PP1). All three PPP family members share the same overall active site fold and organization, illustrating the need for tight regulation of substrate interactions.

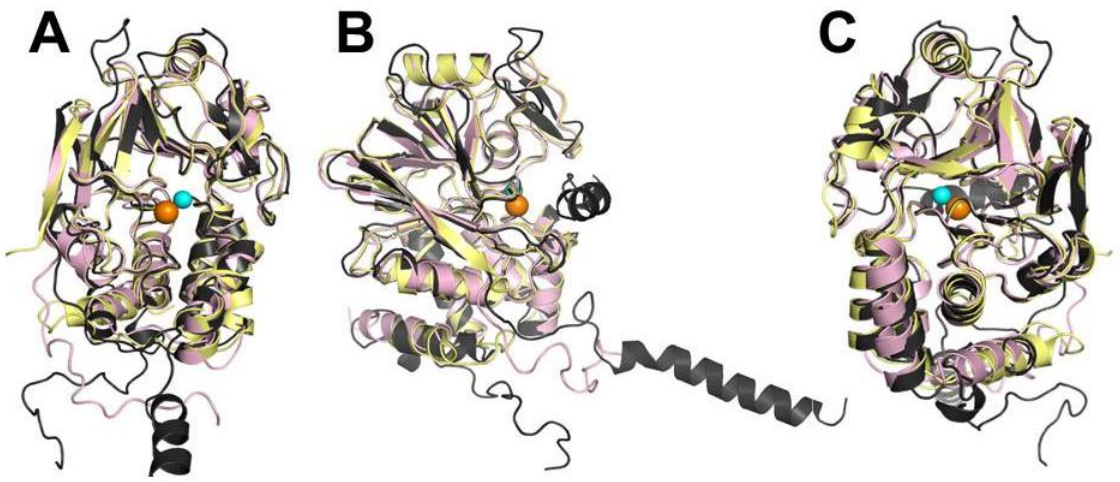


Figure 9: Catalytic domains of CN, PP1, and PP2A.
Overlay of the catalytic domains of CN (black), PP1 (yellow) and PP2A (pink). Metals (Zn^{2+} in cyan, Fe^{3+} in orange) are from the CN structure. A) Front view, looking into the active site. B) Side view, with AID of CN sitting over the active site. The CNB binding helix extends away from the catalytic core. C) View from the back.

Three events must take place for CN activation: all four Ca^{2+} binding sites in CNB must become populated, the AID must be displaced from the active site to provide access for the substrate, and Ca^{2+}-loaded CaM must bind to the CaM-binding region of CNA to form the fully active holoenzyme CNA/CNB/Ca^{2+}/CaM. The sequence of these events has not yet been determined, but a mechanism involving sequential changes in the conformation of CN has been proposed based on the cumulative biochemical and structural data available (Figure 10) [42]. In this proposed mechanism, inactive calcineurin at basal Ca^{2+} levels is in a "closed" conformation, with the AID bound to the active site and only the two high affinity "structural" Ca^{2+} binding sites populated in CNB. In this closed state, the CaM-binding region of CNA is believed to form intramolecular interactions with the BBH of CNA. As the intracellular Ca^{2+} concentration rises, CN is proposed to transition from the "closed" conformation to an "open" active confirmation via the following sequence of events: first, all four EF-hands of CNB become loaded as

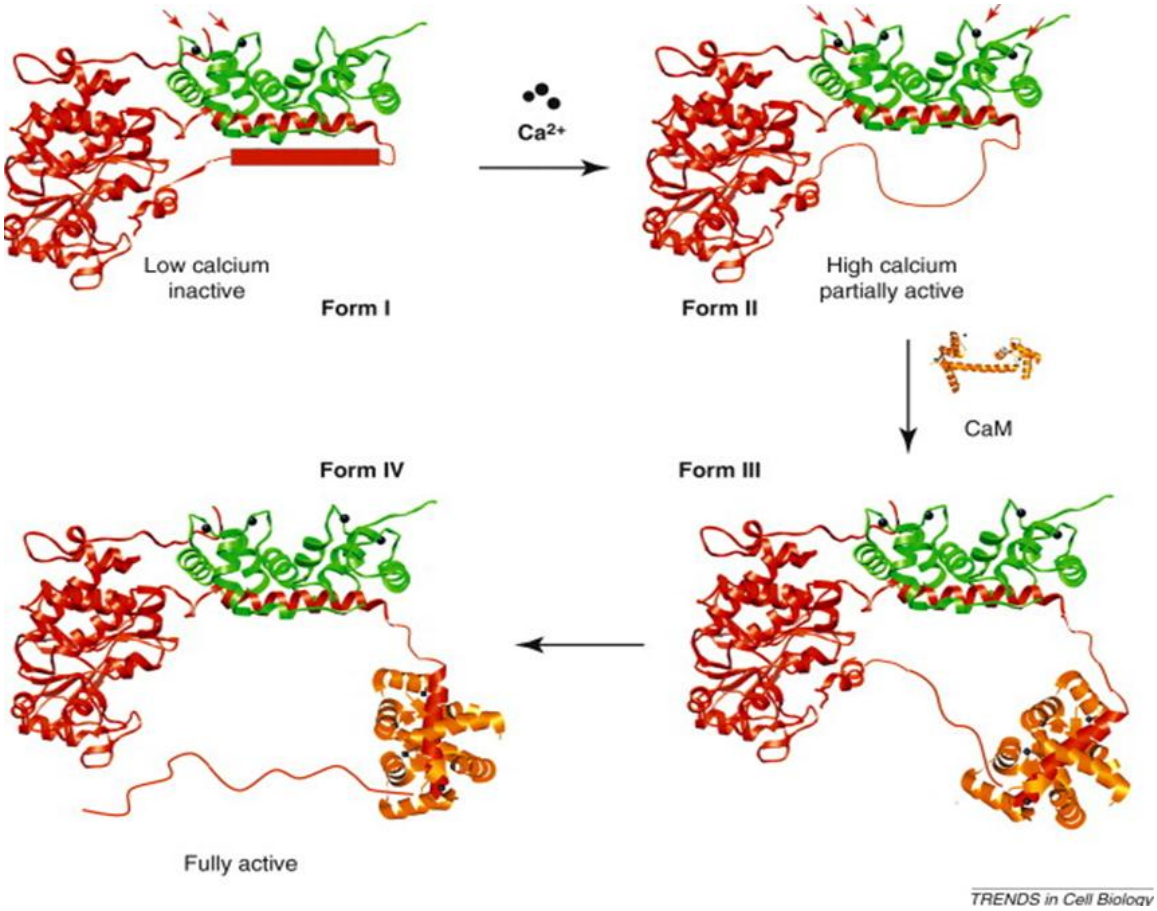


Figure 10: Proposed mechanism for CN holoenzyme activation by Ca^{2+}.
Form I to Form II transition: binding of Ca^{2+} to CNB (green) frees the CaM binding domain of CNA (red). Form III: CaM binds to CN. Form IV: The autoinhibitory domain to be released from the active site and frees CN for substrate binding. Adapted from Li et al (2011) TIBS.

Ca^{2+} populates the two low affinity EF-hands. This causes a rearrangement in the relative orientation of all four CNB EF-hands and leads to the release of the CaM-binding region of CNA from the BBH. CaM then binds to the exposed CaM binding region of CNA, and this in turn causes the displacement of the AID from the active site as the holoenzyme adopts an "open" conformation where the active site is accessible to substrates [42, 43, 46-48].

1.3.B. CN-substrate interactions

CN controls numerous biological processes, such as neurotransmitter release, neuronal structure and development, insulin production, cell cycle progression,

and renal growth and development, among many others. It is therefore not surprising that CN is found in a large number of organs, including the brain (where it is most abundant and makes up ~1% of the total brain protein), heart, pancreas, skin, kidney and lymphocytes. Dysregulation of Ca^{2+}-CN signaling leads to several clinical conditions including neurodegenerative diseases such as Down syndrome and Alzheimer's disease, as well as cancer, inflammatory skin diseases, and cardiac hypertrophy [45], [15].

The primary substrates of calcineurin are the NFAT (nuclear factor of activated T-cells) family of transcription factors (NFATc1-c4). NFATs are expressed in virtually all vertebrate cell types, and they are immensely important for the control of Ca^{2+}-CN-mediated immune cell activation, as well as for the development of bone, heart, muscle, brain, and vascular tissues [29]. In resting cells, NFATs reside in the cytosol and are heavily phosphorylated at 12 conserved Ser residues that are located within a conserved ~300 amino acid regulatory region (Figure 11). CN dephosphorylation of all twelve phospho-Ser residues exposes an NFAT nuclear localization signal and promotes NFAT nuclear translocation. Nuclear NFATs can upregulate the transcription of a large number of genes, whose identity varies according to tissue type. In lymphocytes, antigen binding to the T-cell receptor triggers a cellular Ca^{2+} influx and activates CN dephosphorylation of NFATs. During T-cell activation, for example, nuclear NFATs activate immune response genes such interleukin-2, interleukin-4, and interferon γ [49] (Figure 1 and Figure 11). The action of CN on NFATs is opposed by at least three constitutive kinases (DYRK, CK1 and GSK3) that

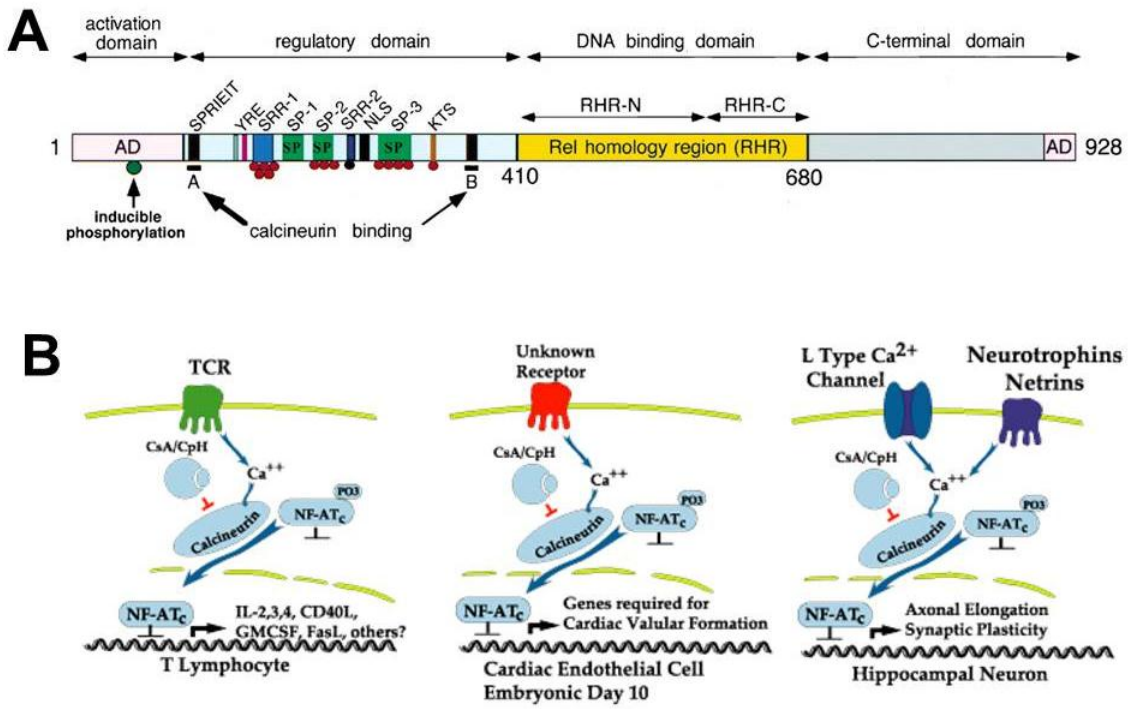


Figure 11: NFAT domain architecture and function.
A) The NFAT (shown is murine NFATc2) regulatory region has two CN binding sites labeled A and B (whose consensus sequences are "PxIxIT" and "LxVP," respectively), and 12 phospho-Ser sites targeted by CN (red dots). Dephosphorylation of the phospho-Ser regions exposes the NLS and triggers NFAT nuclear translocation. B) Ca^{2+}-CN-NFAT signaling in various cell types (left to right: immune, cardiovascular and nervous system cells). Different extracellular stimuli trigger NFAT upregulation of specific genes. Adapted from Hogan et al (2003) Genes Dev.

phosphorylate nuclear NFAT, causing its nuclear export and keeping cytosolic NFAT phosphorylated in the absence of Ca^{2+} signaling [50].

The DNA-binding regions of NFATs are ~300 amino acids in length, and are composed of two distinct DNA binding domains that are joined by a flexible linker. The presence of this unstructured linker allows the two DNA-binding domains to adopt a variety of orientations relative to one another and as a result NFAT monomers can bind DNA in a number of different conformations, depending on the gene being transcribed [51]. Furthermore, NFATs can bind DNA as dimers or by forming complexes with other transcription factors, like AP-1 and FOXP, which themselves are under the control of other intracellular

signaling pathways [52, 53]. The requirement for cooperation with other classes of transcription factors could in part explain how NFATs are able to be involved in such a wide array of cellular processes, as the integration of multiple signaling pathways adds an additional layer of control to the activation of specific genes [29].

It is interesting to note that NFAT signaling occurs only in vertebrates, whereas most other signal transduction pathways are conserved among eukaryotes. This raises the possibility that the NFAT family arose during evolution specifically to suit the needs of vertebrates, whose larger bodies require complex adaptive responses. This is in agreement with the fact that many of the downstream effects of NFATs involve cell-to-cell interactions, and that NFAT signaling is necessary for the development of complex organ systems like the vascular system and certain aspects of the nervous system [54].

The molecular basis for substrate recognition by CN is an active area of investigation. Unlike the related phosphatases PP1 and PP2A, which can form up to hundreds of holoenzymes by swapping regulatory domains, CN has one invariable regulatory domain (CNB) and can only form one holoenzyme that must discriminate between all CN substrates. It remains unclear how CN achieves substrate specificity without the help of a variable regulatory subunit or targeting proteins.

NFATs, and almost all other CN substrates and interaction partners, contain two conserved amino acid sequence motifs that mediate docking to CN: PxIxIT and LxVP, where “x” can be any amino acid (Figure 12, and regions “A”

and “B” in Figure 11A). The PxIxIT site was the first to be identified and has been characterized extensively [55-61], whereas the contribution of the LxVP site to CN-substrate binding [62-65] remains relatively poorly understood. The PxIxIT consensus sequence was first identified in NFATs following many landmark biochemical and structural studies [55-61], and its presence was also confirmed in other CN substrates [4, 64, 66]. In NFATs, the PxIxIT sequence, which is located in a region distinct from the sites dephosphorylated by CN, was shown to be essential for substrate dephosphorylation: PxIxIT peptides at high concentrations inhibit the dephosphorylation of NFATs both *in vivo* and *in vitro* [58]. Interestingly, there is a high degree of variability in the PxIxIT sequences of various CN substrates, both at conserved and non-conserved positions (Figure 12). Fluorescence polarization studies investigating the binding affinities between CN and peptides encoding PxIxIT sequences from different substrates revealed a broad range of affinities, with dissociation constant (K_d) values ranging from 0.5 to 250 μM. For example, the PxIxIT of NFATc1 (“SPRIEIT”) interacts relatively weakly with CN (K_d ~20 μM),

CN Substrate	Biological Function	P	x	I	x	I	T
PVIVIT	synthetic peptide	P	V	I	V	I	T
NFATc1	transcription regulation	P	R	I	E	I	T
NFATc2	transcription regulation	P	R	I	E	I	T
NFATc3	transcription regulation	P	S	I	Q	I	T
NFATc4	transcription regulation	P	S	I	R	I	T
TRESK	receptors/ion channels	P	Q	I	I	I	S
RCAN1	regulatory	P	S	V	V	V	H
Crz1	transcription regulation	P	I	I	S	I	Q
AKAP79	targeting	P	I	A	I	I	I	T

CN Substrate	Biological Function	L	x	V	P
NFATc1	transcription regulation	Y	L	A	V	P
NFATc2	transcription regulation	I	L	L	V	P
NFATc3	transcription regulation	F	L	S	V	P
NFATc4	transcription regulation	Y	L	A	V	P
RCAN1	regulatory	H	L	A	P	P
KSR2	scaffold	T	L	S	V	P
STEP	tyrosine phosphatase	S	L	T	V	K

Figure 12: Consensus PxIxIT and LxVP sequences of CN-binding proteins.
All sequences are from human proteins, with the exception of Crz1 which is a yeast substrate.

suggesting that the NFATc1-CN cellular complex is transient [58, 67]. The PxIxIT motif of the membrane scaffolding protein AKAP79 (“PIAIIIT”), however, forms a tighter, more stable complex (K_d ~0.5 μM) [10, 42]. These differences in affinity reflect the importance that the intensity and duration of a Ca^{2+} pulse has for cellular signaling, as CN dephosphorylation of lower affinity substrates requires Ca^{2+} pulses to be sustained for longer periods of time. Indeed, NFAT signaling in lymphocytes requires several hours of sustained Ca^{2+} influx to achieve full T-cell activation [26].

The LxVP site was also shown to be important for CN substrate dephosphorylation. In enzymatic assays, an LxVP peptide derived from NFATc1 inhibits dephosphorylation of LxVP-containing phosphoprotein and phosphopeptide substrates that lack a PxIxIT sequence. To date no K_d values for LxVP sequences have been reported, but peptides containing LxVP motifs from NFATc1-c4 (see Figure 12 for consensus LxVP sequences) have been indirectly shown to bind to CN with significantly different relative strengths *in vivo* and *in vitro*. In fact, both the LxVP and PxIxIT motifs of NFATs show widely differing relative affinities for CN. For example, the NFATc2 LxVP has almost no affinity for CN, suggesting that NFATc2 relies on its PxIxIT sequence for CN docking. Despite this however, the PxIxIT of NFATc2 still binds to CN more weakly than the LxVP of NFATc1 [42, 65]. These findings further support the conclusion that the fine-tuning of CN-substrate affinity plays a role in mediating CN substrate specificity and the outcome of CN-dependent Ca^{2+} signaling.

Structural data have recently become available for CN in complex with two separate peptides that encode two different PxIxIT sequences: PVIVIT [4, 68] and PIAIIIT [10], both of which bind to CN tightly (K_d~0.5 μM) (Figure 13). The PVIVIT peptide is synthetic, and was chosen for its high affinity for CN from a combinatorial library of peptides with different PxIxIT consensus sequences [69]. The peptide containing the PIAIIIT sequence is a segment of AKAP79, a docking/scaffolding protein that targets CN to the plasma membrane. The CN-PxIxIT interface is almost perfectly conserved in both complexes. In both structures, the PxIxIT forms a β-strand and sits in a hydrophobic groove on the surface of the catalytic domain of CN, extending one of the two β-sheets in the CNA central β-sandwich. Both PxIxITs are aligned parallel to β-14 of CNA, and the peptides are anchored to CNA via the conserved "I" residues of the PxIxIT (PvIvIT and PiaIiIT), which are buried inside the surface hydrophobic pockets

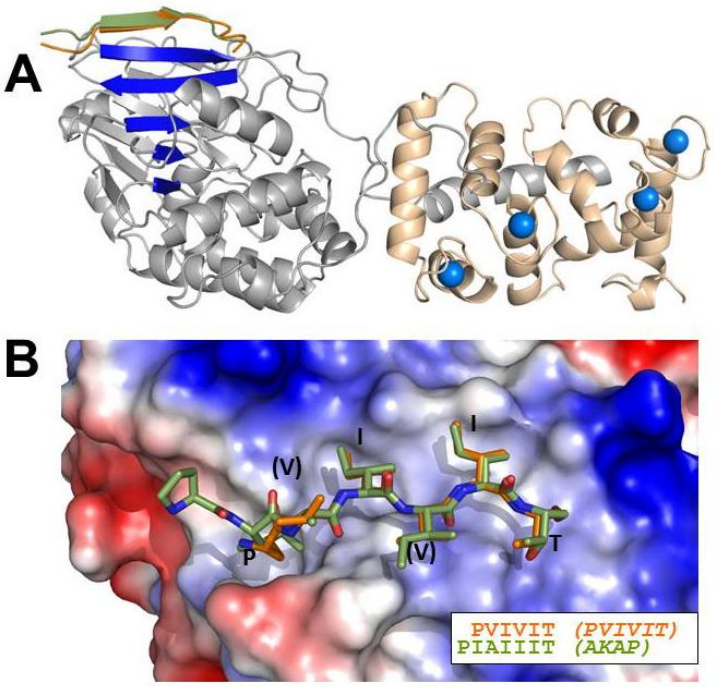


Figure 13: 3D structure of the CN-PxIxIT interface. Overlaid X-ray crystallography structures of the PVIVIT synthetic peptide [4] and the AKAP79 PIAIIIT peptide [10] bound to CN. The overlay was generated using CNA from each complex structure. A) Both peptides form β-strands that extend a central β-sheet in the CNA catalytic domain (PVIVIT, orange; AKAP, green; CNA, grey; CNB, beige; Ca^{2+}, blue spheres). B) Stick representation of only the "PxIxIT" portion of each peptide, emphasizing hydrophobic nature of the PxIxIT binding pocket (surface electrostatics representation of CN).

formed by the edges of the two β-sheets of CNA. The “P” of PVIVIT is also buried in a hydrophobic pocket, and the “T” forms hydrogen bonds with CNA. In AKAP79, the “P” is shifted N-terminally by one position, and “P” is replaced with “PI,” resulting in a “(PI)xIxIT” instead of “PxIxIT” sequence. Despite this, the overall interaction is perfectly conserved and the first “I” in AKAP79 participates in the same hydrophobic interactions as the “P” in “PVIVIT” [4, 10].

The CN-PxIxIT structures provide a structural basis for understanding how the wide range of CN affinities of various natural PxIxIT sequences can lead to a graded variation in substrate-CN binding affinities. Data show that each residue of the PxIxIT consensus sequence makes an individual contribution to the overall binding affinity of the site. Individual residue contributions to the overall free energy of a binding motif are rarely encountered in protein-protein interactions, and the PxIxIT motif likely derives this property from a random-to-order transition it undergoes during CN binding (i.e. transition from a free disordered state to a constrained β-strand bound state). This is consistent with the fact that many CN-interacting proteins contain long disordered regions [4, 42], and suggests that they could at least partially adopt secondary or tertiary structure upon CN binding.

In summary, currently available data indicate that CN-substrate interactions are regulated at the molecular level by the affinity of the PxIxIT and LxVP sites for CN, and that these affinities are dictated by the amino acid sequence of each site. Prior to the work presented in this thesis, the LxVP docking site on CN had not yet been identified. Structural data describing the

CN-LxVP interaction provides important insight into how the two consensus sequences on CN substrates together contribute to the fine-tuning of CN-substrate binding.

Additional *in vivo* and *in vitro* studies are also necessary to fully understand how CN is targeted to specific substrates at the right times, but existing data suggest that many layers of regulation are involved, including the need for NFATs to interact with additional transcription factors, and differences in the length, intensity, and spatiotemporal distribution of Ca^{2+} signals.

1.3.C. Calcineurin regulation by small molecules

As mentioned earlier, CN activity is inhibited by direct binding of the fungal-derived immunosuppressant drugs CSA and FK506 (Figure 14). The crystal structures of CN in complex with each of these drugs unexpectedly showed that each CN-drug complex also contains an additional immunophilin protein (Figure 15) [1, 17, 18]. Immunophilins are ubiquitous proteins that function as chaperones and as peptidyl-prolyl isomerases [70]. The X-ray crystallography structures of the CN-drug-immunophilin complexes CN-CSA-cyclophilin and CN-FK506-FKBP reveal that, unlike the CN AID, these drug-immunophilin complexes do not target the active site. Most of the interactions with CN are made by the small

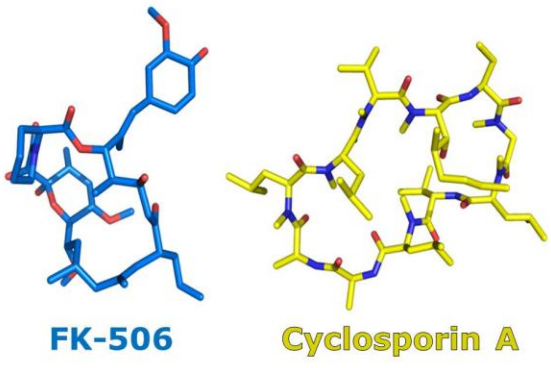


Figure 14: Structures of the immunosuppressant drugs FK506 and Cyclosporin A.

molecules CSA and FK506, which bind over 30 Å away from the CN active site at the CNA/CNB interface [1, 17, 18]. Despite their larger size, the immunophilins cyclophilin and FKBP make only limited contacts with compared to CSA and FK-506. For example, while seven CSA residues form hydrophobic contacts with ten CN residues at the CNA/CNB interface, cyclophilin only forms two hydrogen bonds and one hydrophobic contact with CN at that same interface. At a secondary interface located on the globular catalytic domain of CNA, CSA makes two hydrophobic contacts with CN while cyclophilin makes only one [18].

Interestingly, while CN phosphoprotein and phosphopeptide substrates cannot be dephosphorylated by CN-drug-immunophilin complexes, the small

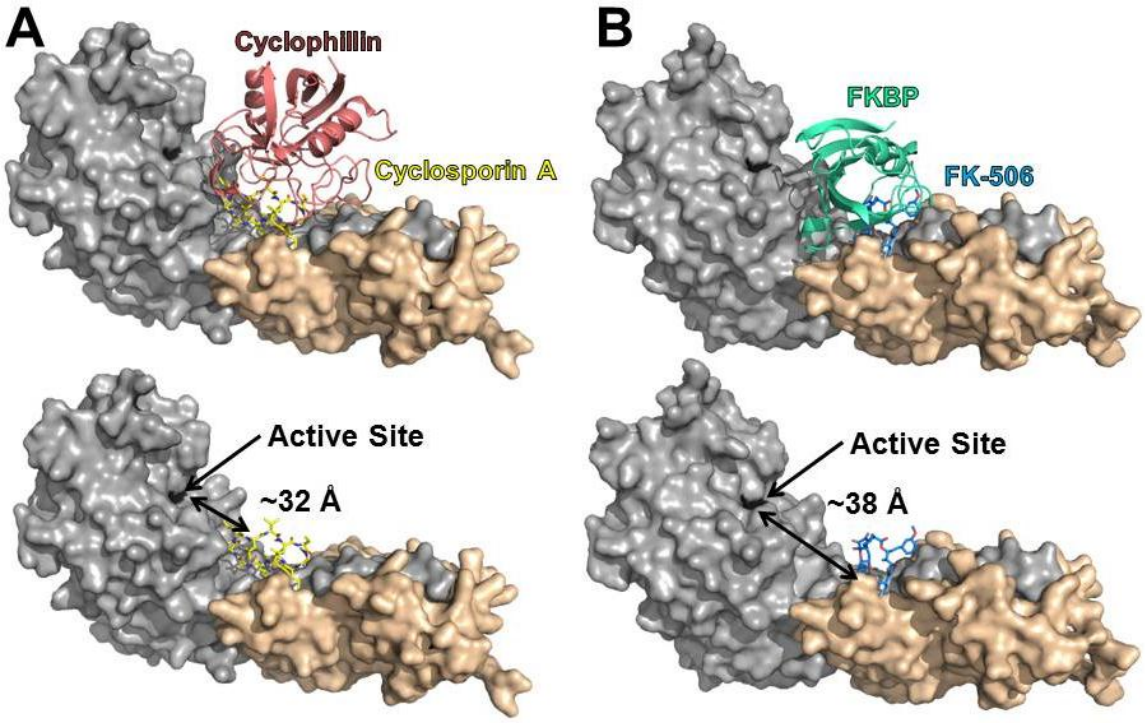


Figure 15: 3D structure of CN-immunosuppressant-immunophillin complexes.
A) The structure of CN (shown in surface representation with CNA in grey and CNB in beige) bound to cyclosporin A (CSA, yellow) and cyclophillin (red). Cyclophillin is removed in the bottom pannel for clarity. B) CN bound to FK-506 (blue) and FK-binding protein (FKBP, cyan), with FKBP removed in the bottom pannel for clarity. Both CSA and FK-506 bind over 30 Å away from the active site, while the immunophillins bind only ~10-20 Å away from the active site.

molecule para-Nitrophenyl Phosphate (pNPP) (Figure 16), which is a nonspecific phospho-Tyrosine mimic, is readily dephosphorylated [71]. This suggests that the presence of the drug-immunophilin complexes impedes substrate-CN binding by an as yet unknown mechanism that does not alter the active site and thus activity of CN. It has been suggested that substrate access to the active site is simply blocked by the large immunophilin proteins, which prevent access by steric inhibition. However, whether the immunophilins are necessary for CN inhibition by CSA and FK506 is currently an area of debate. A second possibility is that drug-immunophilin complexes impede CN binding by substrates via the LxVP motif, but without knowledge of where the LxVP docking site is located on the surface of CN, this hypothesis has not been experimentally confirmed. It is possible that the immunophilin proteins simply serve to prime CSA and FK506 for binding to CN, with no additional role in CN inhibition, as free CSA and FK506 exhibit low CN affinity [72].

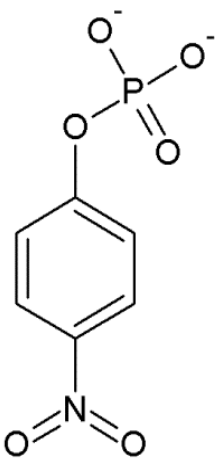


Figure 16: pNPP (p-Nitrophenyl Phosphate).

1.3.D. Calcineurin regulation by proteins

A238L, a protein encoded by the African Swine Fever virus, is an extremely potent inhibitor of CN. African Swine Fever is a double-stranded DNA virus that infects domestic pigs and causes hemorrhagic fever, resulting in death [73]. The C-terminal domain of A238L is a CN-binding domain (CBD) containing both PxIxIT and LxVP consensus sequences (Figure 17A), and this region was demonstrated to be necessary and sufficient for inhibition of CN activity in a

yeast reporter system [74]. Given the presence of PxIxIT and LxVP consensus sequences in A238L, this viral protein could inhibit CN either by blocking the active site, or by competing with substrates like NFAT for CN binding.

Several known endogenous regulator proteins of CN are known, including Regulator of Calcineurin 1 (RCAN1) and the scaffolding proteins KSR2 and AKAP79. RCAN1 is quickly emerging as a key regulator of CN activity. It was first identified as part of a cluster of genes on chromosome 21 and is linked to Down syndrome, hence its original name Down Syndrome Critical gene Region 1 (DSCR1) [75]. Dysregulation of RCAN1 protein levels has also been linked to other Ca^{2+}-CN diseases, including several neurodegenerative conditions, cardiac hypertrophy, and most recently Fragile X syndrome [76-78]. RCAN1 is a

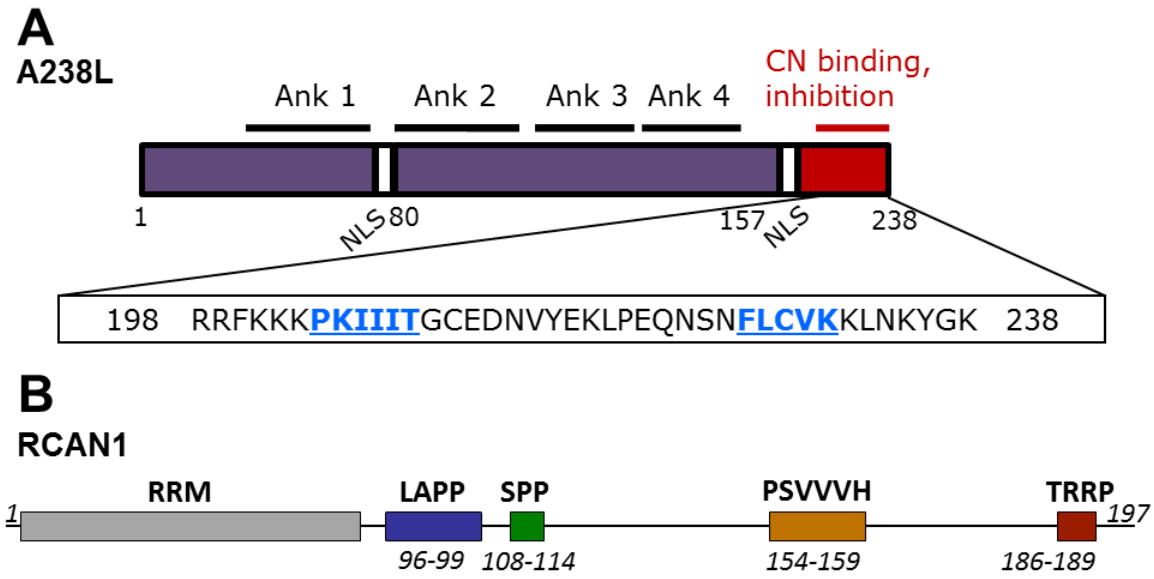


Figure 17: Two protein regulators of CN.
A) Domain architecture of the viral protein A238L. A238L consists of four Ankryn repeats (Ank) and a C-terminal CN-binding domain (CBD) that encodes PxIxIT (PKIIIT) and LxVP (FLCVK) motifs. B) Domain architecture of the endogenous CN regulator RCAN1. RCAN1 contains LxVP (LAPP, blue) and PxIxIT (PSVVVH, orange) CN-binding motifs. A Ser-rich SPP motif (green) contains Ser108, which is dephosphorylated by CN. The TRRP (red) motif inhibits CN phosphatase activity. The N-terminal RNA recognition motif (RRM) is not implicated in CN regulation.

relatively small protein (~23 kDa) and contains an RNA-Recognition Motif (RRM) and a CN-binding region that encodes PxIxIT and LxVP sites (Figure 17B) [66]. It was initially demonstrated to be an inhibitor of CN, but emerging data suggests that under certain conditions it can also upregulate CN activity. It is not yet understood what determines the effect of RCAN1 on CN, but the phosphorylation state of RCAN1 and its local cellular concentration levels may play a role [66].

The scaffolding protein AKAP79 targets CN to Ca^{2+} channels in the plasma membrane. It was recently shown that CN activity against an NFAT reporter gene is increased following mutations that augment the affinity of AKAP79 PxIxIT for CN. However, additional mutations that increased affinity even further caused a decrease in CN activity. A possible explanation is that tight CN-AKAP79 binding increases the local CN concentration near the Ca^{2+} channels, exposing CN to more Ca^{2+} and enhancing its activity. An interaction that is too tight, however, sequesters too much CN to the plasma membrane, preventing its interaction with cytosolic substrates and decreasing its cellular effects [10].

1.4. Contribution of this dissertation to the calcineurin field

Structural studies of CN-substrate complexes will be invaluable for understanding of how CN-substrate regulation is achieved. Prior to this work, the site of the LxVP binding interaction on CN was still unknown, and this information is essential for understanding how differences in the LxVP sequence affect CN binding. Furthermore, most substrates have both an LxVP and a PxIxIT site, and it is important to understand how these two sites come together to coordinate

substrate binding to CN. NFATs are large proteins, measuring almost 1000 amino acids in length, and their LxVP and PxIxIT motifs are separated by hundreds of amino acids, which include the 12 phospho-Ser residues targeted by CN. No explanation is available for how NFATs interact with the CN active site to achieve dephosphorylation of all 12 phospho-Ser residues, but understanding how the LxVP and PxIxIT substrate motifs work together to mediate substrate binding may shed light on substrate recognition by CN. Structural biology studies of CN-NFAT complexes are extremely difficult: these complexes are too large for structure determination by NMR spectroscopy, and the long (~300 amino acids), disordered intervening region between the NFAT LxVP and PxIxIT motifs makes crystallization of these complexes nearly impossible.

This dissertation provides the structural basis of CN-substrate binding by presenting the X-ray crystallography structure of CN bound to the C-terminal CN-binding domain (CBD) of the viral protein A238L. The A238L CBD mimics CN substrates and provides an ideal system for the study of CN-substrate complexes. The LxVP and PxIxIT sites of A238L bind to CN with high affinity and are separated by fewer than 20 amino acids, unlike the hundreds of amino acids separating the PxIxIT and LxVP of NFATs. These key features make A238L-CN amenable to crystallographic studies.

The crystal structure of CN-A238L provides the first understanding of CN-substrate binding in the context of both the PxIxIT and LxVP sites. Importantly, it provides the first description of the CN-LxVP interaction and shows that the LxVP docking site on CN is also the binding site of the immunosuppressant drugs CSA

and FK506. The A238L-CN structure identifies the LxVP as being critical for immune activation, and provides the first molecular basis for understanding the mechanism used by immunosuppressants to inhibit CN. CSA and FK506 have serious side effects, as they inhibit several functions of CN. The atomic level description of the interactions at the LxVP site will allow for the design of drugs with much more targeted effects on CN-substrate interactions.

Finally, RCAN1 is rapidly emerging as a key regulator of CN activity. In addition to PxIxIT and LxVP motifs, the CN-binding domain of RCAN1 contains additional regions that are believed to interact with CN. Structural, biophysical and biochemical studies of the CN-RCAN1 interaction will be invaluable for elucidating the molecular determinants of CN regulation by RCAN1. In this dissertation, the complex between CN and RCAN1 is investigated using NMR spectroscopy and isothermal titration calorimetry (ITC), and the contributions of conserved sequence motifs and residues of RCAN1 to CN binding are determined. Because protein structure is an indicator of protein function, the structure of the CN-RCAN1 complex is pursued in order to gain an understanding of how RCAN1 binding inhibits CN activity, and how the structure of the complex would have to change for RCAN1 to become an enhancer of CN activity.

Chapter 2. Isothermal Titration Calorimetry (ITC)

2.1. Introduction and Overview

Proteins carry out their biological functions by interacting with other biomolecules (proteins, nucleic acids, lipids, etc.), and the formation and dissociation of these complexes is a dynamic process that is critical for normal cellular function. One way to study these interactions is through structural biology, which describes macromolecular complexes in detail. In particular, structural data provides an atomic level description of the specific interactions taking place between the individual components of a complex. An additional, complementary approach to the study of biological interactions is isothermal titration calorimetry (ITC), which provides valuable information about the thermodynamic parameters that govern these interactions. During the past twenty years, ITC has become increasingly widespread and is now commonly used in conjunction with structural biology and other modern spectroscopy techniques to study the binding events taking place between biological molecules, thus providing valuable insight into biological processes [79].

ITC functions by measuring small changes in the heat associated with binding interactions, conformational changes, or chemical and biochemical reactions taking place in solution. ITC is capable of directly determining the thermodynamic parameters associated with binding interactions between biological molecules, and of characterizing the kinetics of enzyme-catalyzed reactions (Figure 18). Specifically, ITC is the only technique available that, in one single experiment, provides all thermodynamic parameters associated with an interaction: the binding constant (K_a), enthalpy (ΔH), entropy (ΔS), stoichiometry

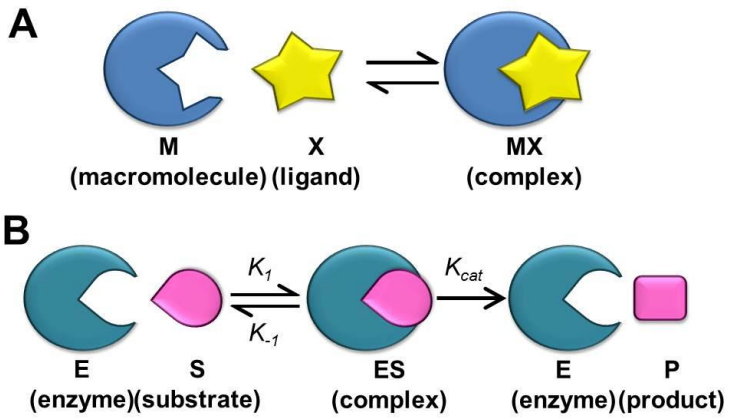


Figure 18: ITC can be used to characterize binding interactions and enzymatic reactions.
A) Representation of a binding interaction. During an ITC experiment, a macromolecule and its ligand are gradually mixed together, and the heat associated with complex formation is measured. B) Simplified model of an enzymatic reaction. ITC can determine kinetic parameters by gradually mixing a substrate and enzyme.

(n) of binding, and Gibbs free energy (ΔG) of a reaction. In addition, the specific heat of binding (ΔC_p), whose value reflects changes in the amount of solvent exposed hydrophobic surfaces during an interaction, can be determined by carrying out a series of ITC experiments at different temperatures. ITC can also be used to study enzyme kinetics, as the ITC output signal directly measures reaction rate (v_0) and ΔH of a biochemical reaction, and Michaelis-Menten relationships allow for calculation of the catalytic rate constant (K_{cat}) and the Michaelis constant (K_M) [80].

The success of ITC largely stems from the fact that it is a versatile technique with numerous applications. In addition to its pure research applications aimed at better understanding biological processes, ITC It is now regularly used by pharmaceutical companies as part of the drug discovery pipeline, where it plays a prominent role in lead selection and optimization, and where it has been adapted to function in a high throughput, automated manner [81]. Additional advantages of ITC include the ability to perform experiments in biologically relevant conditions (pH, salt, temperature, etc.), the use of

spectroscopically silent samples (no requirement for chromophore or fluorophore tags), and the ability to function as a universal enzymatic assay even when reaction solutions are turbid or opaque.

The current section describes the theory behind ITC and explains how the technique can be used to determine specific thermodynamic and kinetic parameters. Finally, important factors that must be considered for optimal acquisition of high quality data, as well as data analysis and interpretation are discussed.

2.2. ITC instrumentation and theory

2.2.A. Instrumental design

The calorimeter, which measures the heat produced or absorbed during physical processes or chemical reactions, was one of the earliest scientific instruments to be reported in the literature [82]. In 1780, Antoine-Laurent de Lavoisier designed the first calorimeter, which used the melting of ice as a measure of the metabolic heat generated by a guinea pig enclosed in a central chamber (Figure 19). Lavoisier's instrument is an example of a heat conduction calorimeter, and some modern calorimeters still operate based on this principle. In modern heat conduction calorimeters, the measurement cell is connected to a heat sink whose temperature is actively kept constant. Heat associated with the mixing of the two reactants in the measurement cell generates a small ΔT between the measurement cell and the heat sink. The temperature difference causes the transfer of heat between the sample cell and heat sink, passively maintaining a constant temperature in the measurement cell. The measurement cell and heat

sink are connected by heat flow sensors, and the experimental output is a small voltage that is proportional to the ΔT developed across the heat sensors and to the heat signal [80, 83]. Another type of calorimeter still used today is the temperature change calorimeter, which functions by simply measuring the temperature change in the sample cell after mixing two reactants. In this case, the instrument's output is the temperature of the sample chamber expressed as a function of time, and can be converted to a heat by multiplying the observed ΔT by an energy equivalent constant, ε_c (cal/°C), that is specific to the calorimeter [80].

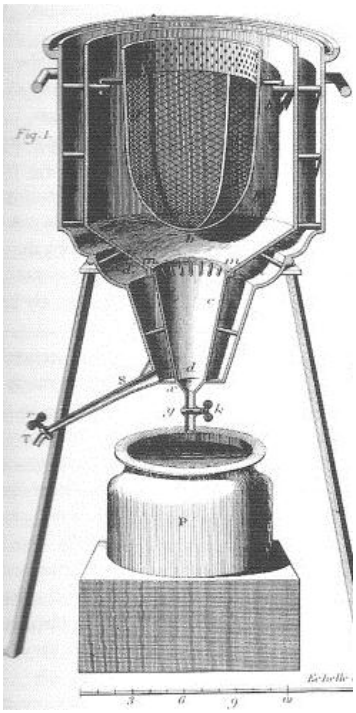


Figure 19: Lavoisier's ice calorimeter.
The original ice calorimeter used by Lavoisier and Laplace was used to measure metabolic heat released by a guinea pig. The animal was confined to an inner chamber made of mesh material. A surrounding middle chamber was filled with ice that acted as a heat sensing material. The outer chamber, also filled with ice, was enclosed in wood and acted as an adiabatic shield. The ice in the middle chamber melted in response to heat generated by the guinea pig in the inner chamber. The water produced in the middle chamber was collected and measured, and used to calculate the metabolic heat produced by the guinea pig. Adapted from Lavoisier and Laplace (1780) *Mem. Acad. Roy. Sci.*

Most modern calorimeters, however, operate based on the power compensation principle. The experimental output of power compensation instruments is the amount of power generated by the instrument in order to maintain the sample cell temperature constant following the mixing of two reactants (hence, "isothermal"). A simplified representation of a modern power compensation ITC instrument and the raw data obtained from a typical

experiment are shown in Figure 20. The instrument consists of a sample cell which, for binding experiments of the type shown in Figure 18A, initially contains only the macromolecule. Typically, the sample cell volume is ~1.4 ml, but can be as low as 200 µl for some instruments. The ligand (a small molecule, peptide, protein, nucleic acid, etc.) is initially stored in a syringe (typical volume is ~300 µl)

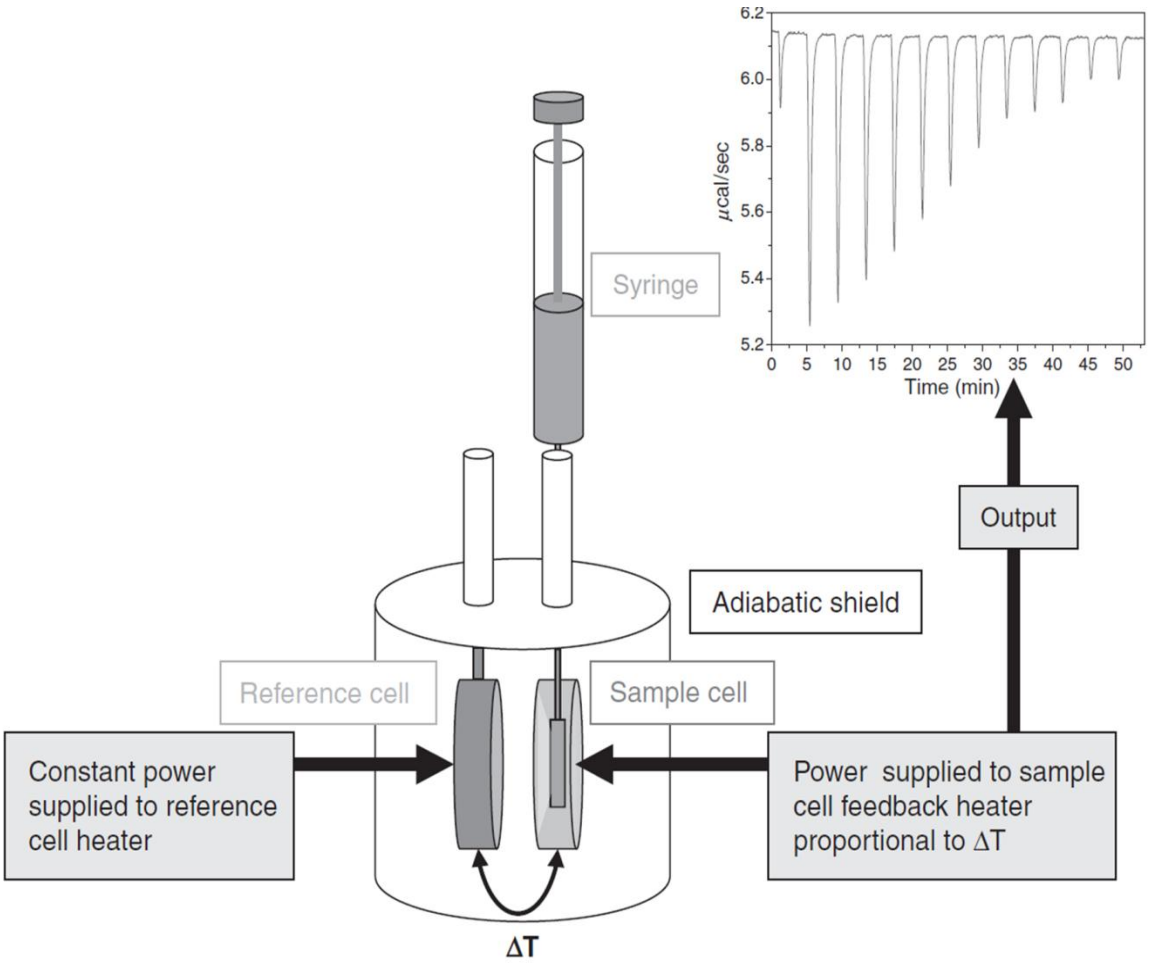


Figure 20: Power compensation ITC instrument.
Major components of the instrument include: sample and reference cells that are kept at constant temperature, a syringe for adding ligand and stirring the reaction, and an adiabatic shield. The heat associated with each injection of titrant into the sample cell is initially detected as a change in temperature (ΔT) between the sample and reference cells. The signal detected during the ITC experiment is the power (µcal/sec) applied by the instrument to the sample cell after each injection in order to keep the sample cell temperature constant relative to the reference cell temperature. The raw output, shown in the inset at the top right, is a plot of power as a function of time. The heat exchanged is calculated by integrating the power peak associated with each injection over the time that is required for the applied power to return to its baseline value. Adapted from Freyer and Lewis (2008), *Meth. Cell Bio.*

and added to the sample cell via a series of injections (titrations). The syringe is shaped like a paddle and also serves to stir the contents of the sample cell, ensuring adequate mixing of the reactants. The instrument also contains a reference cell that is filled with water and kept at constant temperature, and an adiabatic shield. The sample cell is kept at constant temperature relative to the reference cell throughout the experiment duration (most instruments can collect data at temperatures between 2-80 °C). During each injection, heat is generated or absorbed and the ITC instrument must exert a certain amount of power (measured in µcal/sec or µJ/sec) in order to counter the effect of the reaction heat and keep the sample cell temperature constant. The power generated by the instrument is continuously plotted as a function of time over the duration of the experiment, giving rise to a plot (Figure 20 inset) with one distinct peak for each injection of ligand. As the ITC experiment progresses, the amount of free macromolecule available for complex decreases and the heat associated with each injection becomes increasingly smaller, until ligand saturation is reached and the signal plateaus [80].

2.2.B. Determination of thermodynamic parameters

ITC is the most quantitative method for determining thermodynamic parameters available to date. Other methods, such as enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and surface plasmon resonance (SPR) are also used for the study of biologically relevant interactions, but use indirect measurements to provide thermodynamic parameters [84]. The thermodynamic binding constant (K_a), the binding enthalpy (ΔH), and the binding stoichiometry

(n) can be directly obtained from a single ITC experiment, and are used to calculate the Gibbs free energy (ΔG) and entropy (ΔS) of binding. In addition, the heat capacity of binding (ΔC_p) can be determined by measuring ΔH at several temperatures. The raw ITC signal has units of power (μcal/sec), and integration of the power signal peaks provides the heat associated with the signal. Equation 1 describes the relationship between power and heat:

$$Power = \delta Q/\delta t \quad (1)$$

where Q is the measured heat and t is time. Q is the total apparent heat that is observed, and reflects contributions from mechanical effects and from the ligand and macromolecule heats of dilution in addition to the heat contribution of the binding interaction [80]. Control experiments and corrections applied to this raw data (described later in this chapter) can remove most extraneous heat contributions, leaving only the heat of binding contribution.

Once Q is determined, the binding enthalpy (ΔH) of the reaction, in units of cal/mol of ligand), can be obtained from:

$$Q = V \bullet \Delta H \bullet \Delta X_b \quad (2)$$

where V is the volume of the reaction (i.e. volume of the sample cell) and ΔX_b is the concentration of bound ligand [85]. Equation 2 establishes that the total cumulative heat that is released or absorbed during a binding interaction is proportional to the total amount of bound ligand. Since $[X]_b$ is not an experimentally known value, Equation 2 must be rewritten in terms of known quantities (i.e. the total ligand and macromolecule concentrations $[X]_t$ and $[M]_t$) [85, 86]. The specific mathematical expression used depends on the properties

of the specific system under investigation (i.e. the number of binding sites, etc.), and a mathematical model describing the system must be carefully selected by the experimenter. (Model selection and data processing are addressed in Section 2.4.B.) ΔH, and K_a are directly obtained from the raw ITC data, and Equations 3-5 provide ΔG and K_d [80, 84, 86]:

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

$$\Delta G = -RT\ln K_a \quad (4)$$

$$K_d = 1/K_a \quad (5)$$

R is the gas constant (8.314 J K^{-1} mol^{-1}) and T is the temperature in Kelvin. K_d is a useful term, and in practice is more widely used than K_a to describe binding affinity because its value is equal to the ligand concentration at which half of the protein molecules are bound to ligand. Finally, ΔH is temperature-dependent, and its measurement over a range of temperatures allows for calculation of the binding heat capacity (ΔC_p) using Equation 6 [86]:

$$\Delta C_p = \delta\Delta H/\delta T \quad (6)$$

ΔCp can be used as an indicator of changes in the amount of hydrophobic surface area that is solvent-exposed during protein folding or ligand-binding processes. Negative values of ΔCp typically indicate the burial of hydrophobic regions (i.e. a decrease in the hydrophobic solvent-exposed surface area due to protein folding or complex formation), while positive ΔCp values are attributed to an increase in the solvent exposed hydrophobic surface area [87].

2.2.C.Interpretation of thermodynamic parameters

ITC is the only technique that, in addition to measuring K_a, can also determine the magnitudes of the two terms that define K_a: ΔH and ΔS. K_a is dictated by the Gibbs free energy of binding, ΔG [86]:

$$K_a = e^{-\frac{\Delta G}{RT}} \quad (7)$$

Since ΔG can be expressed in terms of ΔH and ΔS according to Equation 3, the effect of ΔH and ΔS on K_a it is readily apparent: in order to increase K_a, ΔH must be made more negative and ΔS more positive (Figure 21). ΔH reflects the strength of ligand interactions with the protein target (mainly due to van der Waals forces, hydrogen bonds, and salt bridges) relative to those between ligand and solvent, while ΔS reflects changes in the solvation entropy and in the conformational entropy that occur during binding [86].

The ability of ITC to dissect the enthalpic and entropic components of binding is easily understood in the context of targeted drug development, though the same principles can be applied to ITC results describing macromolecular interactions (i.e. protein-protein interactions). ITC is particularly useful during the lead optimization stage of drug development, which is centered around increasing the binding affinity between a lead compound (potential drug candidate) and its protein target. Equations 3 and 4 clearly show that many combinations of ΔH and ΔS values can give rise to the same K_a, and efficient lead optimization requires an in-depth understanding of

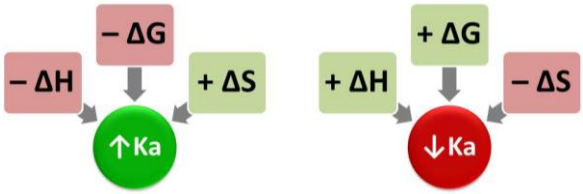


Figure 21: Effect of ΔG, ΔH, and ΔS on K_a. Large negative (-) values of ΔH and ΔG and large positive (+) values of ΔS increase the binding affinity (left); Large positive (+) values of ΔH and ΔG and large negative (-) values of ΔS decrease the binding affinity (right).

the forces driving the binding interaction [86]. Binding affinity can be optimized by selectively controlling ΔH and ΔS, thus influencing whether the binding interaction is enthalpically or entropically driven. This is achieved by careful modifications to the ligand structure and therefore directly impacts the ligand's properties. Some approaches for optimizing K_a by selectively controlling ΔH and ΔS [88] are shown in Figure 22 and include: 1) pre-shaping the drug molecule to fit the geometry of the target binding site with high specificity (Figure 22A). Such structurally optimized ligands usually have a high amount of rigidity. In this case, complex formation is entropically favored because structurally constrained ligands experience minimal conformational entropy decrease upon binding; 2) introducing changes that favor additional specific ligand-target interactions (Figure 22B). This is enthalpically favorable

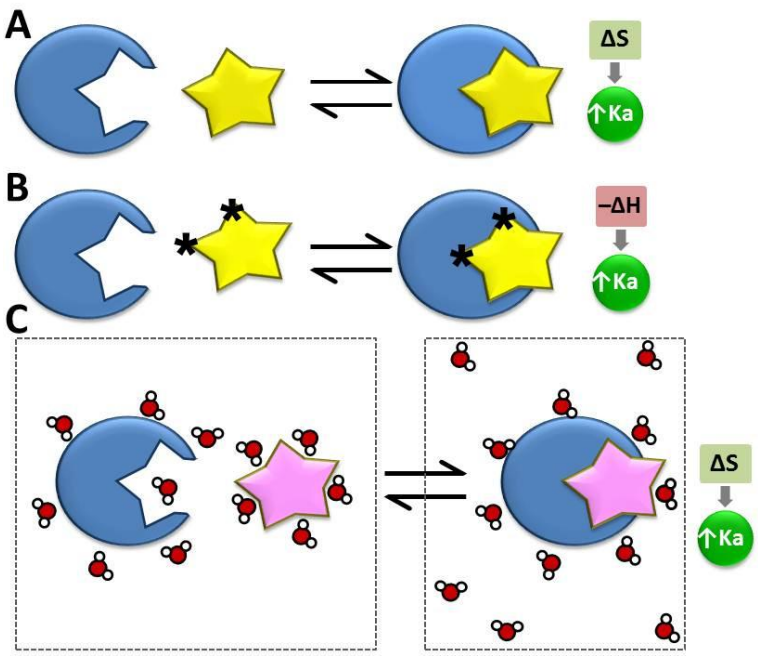


Figure 22: Optimization of binding affinity by selective control of ΔH and ΔS.
A) Pre-shaping the ligand to fit a specific surface on the target protein entropically drives K_a: the entropic cost of binding is small, as the ligand does not lose much conformational freedom (large disordered-to-ordered transitions are entropically unfavorable). B) Favoring ligand-protein interactions over ligand-solvent interactions (additional interactions compared to A shown as black asterisks) enthalpically favors K_a, generating a large negative ΔH upon binding. C) Making the ligand more hydrophobic (pink) entropically drives K_a. Solvent molecules form a motionally constrained "cage" around exposed hydrophobic objects. Complex formation decreases the exposed hydrophobic surface area frees solvent molecules, thus increasing solvent entropy $\Delta S_{solvent}$.

as the ΔH of binding gains a large negative value; and 3) making the ligand more hydrophobic (Figure 22C), which entropically drives affinity. This is explained by the Hydrophobic Effect, which states that water molecules in the solvent form a “cage” around a hydrophobic object, constraining their motion and lowering solvent entropy. Complex formation between hydrophobic objects is entropically favored because it decreases the exposed hydrophobic surface area, releasing the constrained water molecules and thus increasing the solvent entropy [86, 88].

In the case of protein-protein interactions, large positive ΔS values can indicate that the interface of the two proteins is largely hydrophobic and is formed between rigid surfaces with low flexibility, while a large negative ΔH can indicate that the interface between the two proteins is held together by several specific, complementary interactions. However, ΔS and ΔH are closely linked and the interpretation of their values in the context of macromolecular complexes is not always straightforward, unless structural information is also available.

2.2.D. Determination of kinetic parameters

The study of enzyme function and activity is essential for biochemical studies and for drug discovery and development. Enzymes control all biochemical pathways, and account for more than half of current drug targets. Traditional enzymatic assays monitor product formation or substrate depletion, either continuously or at fixed time points. Colorimetric reactions and reactions where the substrate or product is tagged with a chromophore or fluorophore can be monitored using spectrophotometry. Otherwise, analytical methods such as mass spectrometry, chromatography, or gel electrophoresis can be used, though they are less

sensitive, often indirect, and prone to higher error. It is not always possible to monitor product or substrate concentration for a variety of reasons, including the interference of opaque solutions with spectrophotometric detection, or enzyme activity that is below the assay detection limit. Since all enzymatic reactions give off heat, ITC can be used as a universal assay for enzyme activity, overcoming many of the obstacles listed above.

The use of ITC to study reaction kinetics was initially brought forth decades ago [89-91], but only became practical for most scientists once power compensation calorimeters became commercially available [92]. ITC is relatively rapid and precise, and can yield a complete set of kinetic parameters for almost any enzyme-substrate system (as long as the reaction rate is slower than the ~15 second response time of a typical ITC instrument). Alternatively, ITC can be used for activity screening when testing enzyme inhibitors. Because the observable event of ITC is the power generated by the instrument in response to heat given off during the reaction, measurements can be obtained for opaque or turbid solutions that are not suitable for spectrophotometric assays [80, 93].

During an enzymatic reaction, the enzyme (E) and substrate (S) form a complex (ES). The ES complex is converted to an enzyme-product (EP) complex and dissociates into free enzyme and product (E+P). In biological systems, where the substrate concentration far exceeds the enzyme concentration, a typical enzymatic reaction can be simplified and summarized as follows (also see Figure 18B):

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_{cat}}{\rightarrow} E + P \quad (8)$$

where E, S, and P refer to the enzyme substrate and product, k_1 and k_{-1} are the rate constants for the formation and dissociation of the enzyme-substrate complex ES, respectively, and k_{cat} is the catalytic rate for the conversion of substrate to product (it is also known as the turnover number).

Simple reactions of the type shown in Equation 8 can be analyzed by Michaelis-Menten kinetics using the following relationships: [94]:

$$K_M = \frac{(k_{-1} + k_{cat})}{k_1} \quad (9)$$

$$v_0 = \frac{V_{max}[S]}{K_M + [S]} \quad (10)$$

$$v_0 = \frac{k_{cat}[E]_t[S]}{K_M + [S]} \quad (11)$$

K_M is the Michaelis constant and is a measure of the affinity between enzyme and substrate, v_0 is the initial velocity (or the rate of the enzyme-catalyzed reaction), [S] is the substrate concentration, $[E]_t$ is the total enzyme concentration, and V_{max} is the maximal velocity (the reaction rate when [ES] = $[E]_t$). Equation 10 is known as the Michaelis-Menten Equation, and is the basic equation used to describe enzyme kinetics. Using Equation 10 to plot v_0 as a function of [S] produces a rectangular hyperbola, where [S] = K_M at v_0 = ½ V_{max} (Figure 23). The catalytic efficiency of an enzyme is expressed in terms of k_{cat}/K_M, which reflects both the rate of catalysis (k_{cat}) and the affinity of the enzyme for its substrate (K_M) [94].

The reaction rate (v_0 or V_{max}) is expressed in units of moles of product formed over time, or moles of substrate consumed over time. As described in

Equation 1, the raw ITC power signal is equal to the heat given off or absorbed as a function of time: *Power* = *δQ/δt*. ITC measures the ΔH value associated with the substrate (S) → product (P) conversion in the same way that it detects thermodynamic heats of binding, with the difference that the heat of enzyme-catalyzed reactions is much larger than the heat associated with noncovalent binding interactions. The amount of heat associated with converting *n* moles of substrate to product is:

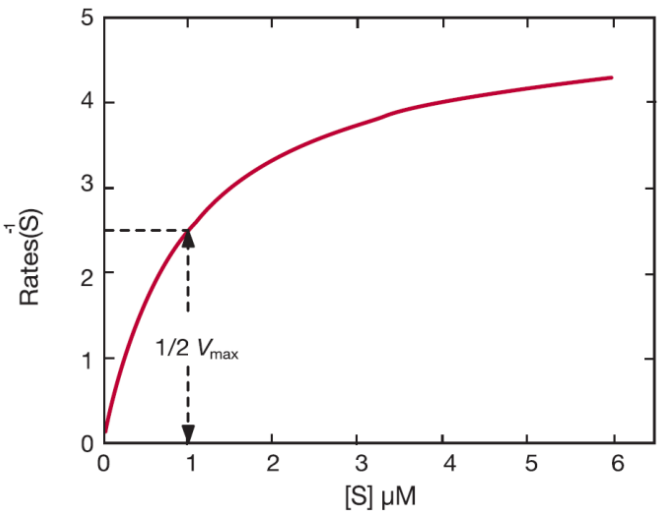


Figure 23: Simulated Michaelis-Menten curve. Rate (v_0) is plotted as a function of substrate concentration [S] using Equation 10. Adapted from Haq (2002) MicroCal, LLC.

$$Q = n\Delta H_{app} = [P]_t V \Delta H_{app} \quad (12)$$

where ΔH_{app} is the experimentally determined molar enthalpy of the reaction, $[P]_t$ is the total concentration of product, and V is the reaction volume. Given that *Power = δQ/δt*, it follows that the power generated by the ITC instrument is:

$$Power = \frac{\delta Q}{\delta t} = \frac{\delta [P]_t}{\delta t} V \Delta H_{app} \quad (13)$$

where $\frac{\delta [P]_t}{\delta t}$ is the rate of product formation, or the reaction rate v_0. Therefore, Equation 13 can be rewritten in terms of v_0, showing that the raw calorimetric power signal is a direct measure of reaction rate [92]:

$$v_0 = \frac{\delta[P]_t}{\delta t} = \frac{1}{V\Delta H_{app}} \cdot \frac{\delta Q}{\delta t} \quad (14)$$

The reaction rate v_0 can be directly determined, as all values in Equation 14 are either known or can be experimentally determined: V is the known sample cell volume, $\frac{\delta Q}{\delta t}$ is directly determined from the ITC power output, and ΔH_{app} can be determined experimentally under the right set of conditions [92] (described in greater detail section 2.3.B). v_0, $[E]_t$, and [S] are fit to Equation 11 ($[E]_t$ is known and [S] can be determined for any time point during the ITC experiment from the known total substrate concentration and the ITC injection volumes) by the ITC software using nonlinear least squares analysis to determine the K_M and k_{cat} values that best fit the data, thus providing the measure of the catalytic efficiency, k_{cat}/K_M [93].

ITC is a convenient method for the acquisition of kinetic data. Kinetic studies of newly discovered enzymes by ITC require minimal method development, and experiments are short (~2 hours) and easy to carry out. Mechanized titration replaces the tedious pipetting steps associated with traditional enzymatic assays, thus increasing precision and reducing experimental error. Importantly, a single ITC experiment provides measurements for an entire range of substrate concentrations, whereas traditional assays require repeating the assay at multiple substrate concentrations. Finally, ITC experiments can be performed in the presence of inhibitors and changes in K_M can easily be monitored, making this technique ideal for drug development research [93].

2.3. Experimental design

ITC is a powerful technique that provides valuable thermodynamic and kinetic information, but accurate measurements are extremely dependent on proper experimental planning and setup. The acquisition of high quality data requires careful planning to achieve optimal experimental conditions. Important considerations include: sample preparation, macromolecule and ligand concentrations, the number and volume of titrant (ligand) injections, the choice of buffer, and the acquisition of the proper control ("blank") experiments. In some cases, data acquisition over a range of different temperatures and sample concentrations, or even reverse titrations (using the macromolecule as the titrant) may be necessary for a full thermodynamic profile of the system under study. Finally, the single injection mode (SIM), which involves the continuous addition of ligand over the time course of the experiment, can provide certain advantages over the traditional multiple injection method, particularly when used for kinetic studies and in high throughput screening applications. These considerations and the steps for setting up a typical ITC experiment are explored in detail below.

2.3.A. Optimization of sample concentrations

During a typical ITC experiment, the sample cell is filled with a macromolecule solution, and the ligand is added to the macromolecule by a series of injections (titrations) over the course of the experiment. With each injection, ligand molecules (X) bind to free macromolecule (M) in the sample cell to form a complex (XM), and heat associated with the binding reaction (ΔH) causes a change in the raw power signal generated by the instrument. The raw ITC signal

is displayed in units of power as a function of time, and each injection produces a distinct peak (Figure 24A). As the experiment progresses, the amount of free macromolecule available for complex formation decreases, and peaks become progressively smaller until ligand saturating conditions are reached and the signal plateaus. Integration of the peaks in the raw ITC signal produces an isotherm that displays the heat given off with each injection as a function of molar ratio (Figure 24B).

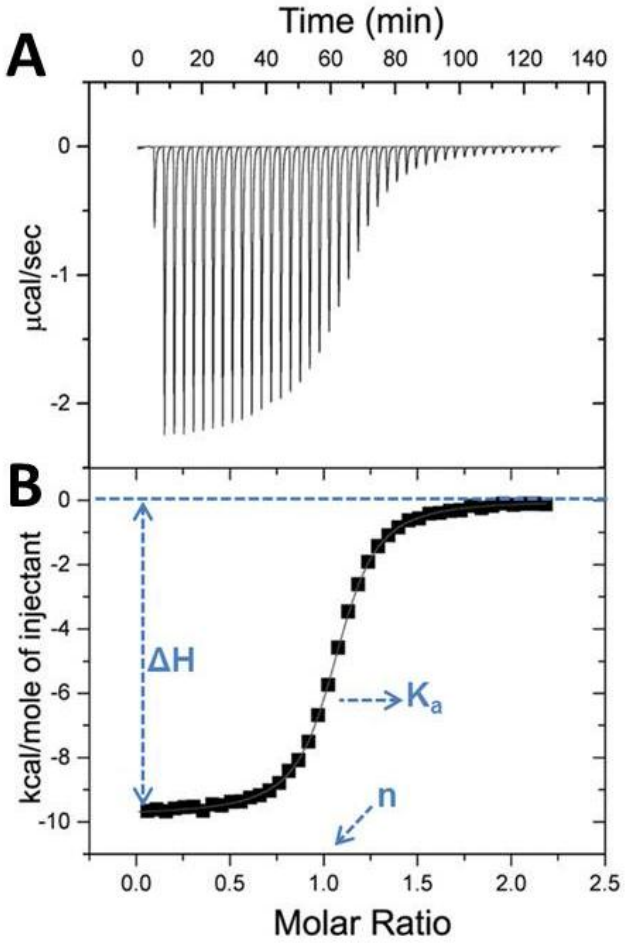


Figure 24: Data from a typical ITC experiment. A) Raw output of the instrument (power as a function of time), with each peak representing one injection. B) Isotherm generated by integrating the area under each peak in "A". Heat (ΔH) is plotted as a function of molar ratio (ligand:macromolecule). Each point corresponds to one injection in "A"; the solid line going through the data points is the binding model, fit to the data by nonlinear regression analysis. The thermodynamic parameters (ΔH, K_a and n) obtained directly from this fit are indicated in blue. Adapted from Petit et al. (2009) *PNAS.*

Before setting up an ITC experiment, it is critical to carefully consider macromolecule and ligand concentrations. Typical macromolecule and ligand concentrations range from ~1-50 µM and ~10-500 µM, respectively. The lower limit of the concentration is dictated by the sensitivity limit of the instrument and the strength of the interaction under study. A typical

ITC instrument has a sensitivity limit of ~0.1 µcal. To ensure precise measurements, each injection should absorb or give off a heat (ΔH) of at least 3-5 µcal. A measurable heat change can be obtained from a macromolecule sample with a concentration as low as 1 µM, provided that its interaction with the ligand is strong.

Another important consideration in determining the macromolecule concentration is the effect of the concentration on the curvature of the isotherm. The shape of the isotherm is determined by the unitless "c" parameter (sometimes referred to as the critical parameter), which is defined by the binding stoichiometry for the reaction (n), the total macromolecule concentration [M_{tot}], and the association constant K_a:

$$c = n \bullet [M_{tot}] \bullet K_a \qquad (15)$$

Figure 25 illustrates the effect of c on the shape of the isotherm. To obtain an optimal isotherm, the value of c should ideally be $10 \leq c \leq 100$, but values in the range of $10 \leq c \leq 1000$ can also be acceptable. An optimal isotherm should meet the following set of criteria to the extent that the sample allows it: 1) a steady initial baseline, established by the first few injections. This is the part of the experiment where the largest signal is generated, because all injected ligand binds to free macromolecule; 2) a well-defined equivalence region, which covers the transition from excess macromolecule to excess ligand. The slope of this region is equal to the association constant K_a, and the equivalence point located at the center of this region defines the stoichiometry of the reaction; 3) a well-established baseline at the end of the titration. This results from the final set of

injections, performed in the presence of excess ligand [84]. These guidelines reduce the error rate associated with the nonlinear regression analysis of the data and increase the accuracy of the results [95]. This is especially important for very strong interactions, where the equivalence region is short and contains very few points or no points at all. Lowering the macromolecule concentration to decrease the value of c adds additional data points to the transition region, and additional points improve the fit of a mathematical model to the data.

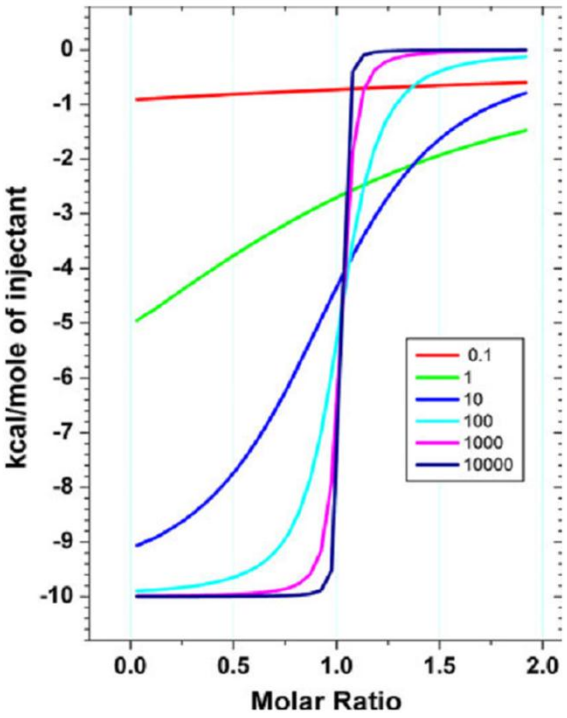


Figure 25: The effect of c on isotherm shape.
(c = n • [Mtot] • Ka). Ideal values of c range from 10-100. Isotherms with high c values have poorly defined equivalence regions, where the slope is too steep and the transition does not contain enough points to accurately determine K_a. Isotherms with low c values lack well-established baselines at the beginning and end of the titration, resulting in poor nonlinear regression analysis of the data. Adapted from Grossoehme et al. (2010) J. Biol. Inorg. Chem.

The effect of K_a and macromolecular concentration on the c parameter is illustrated in Figure 26. Panels A and B show similar curvature with c = 10, though the K_a value of the experiment represented in panel B is 1000-fold higher than the K_a in panel A. The macromolecular concentration in panel B must therefore be 1000-fold lower than in A to achieve the same c value, making the sample in B is too dilute to produce a signal that falls within the detection limit of the instrument. Panels C and D show a 10-fold increase in the c parameter when

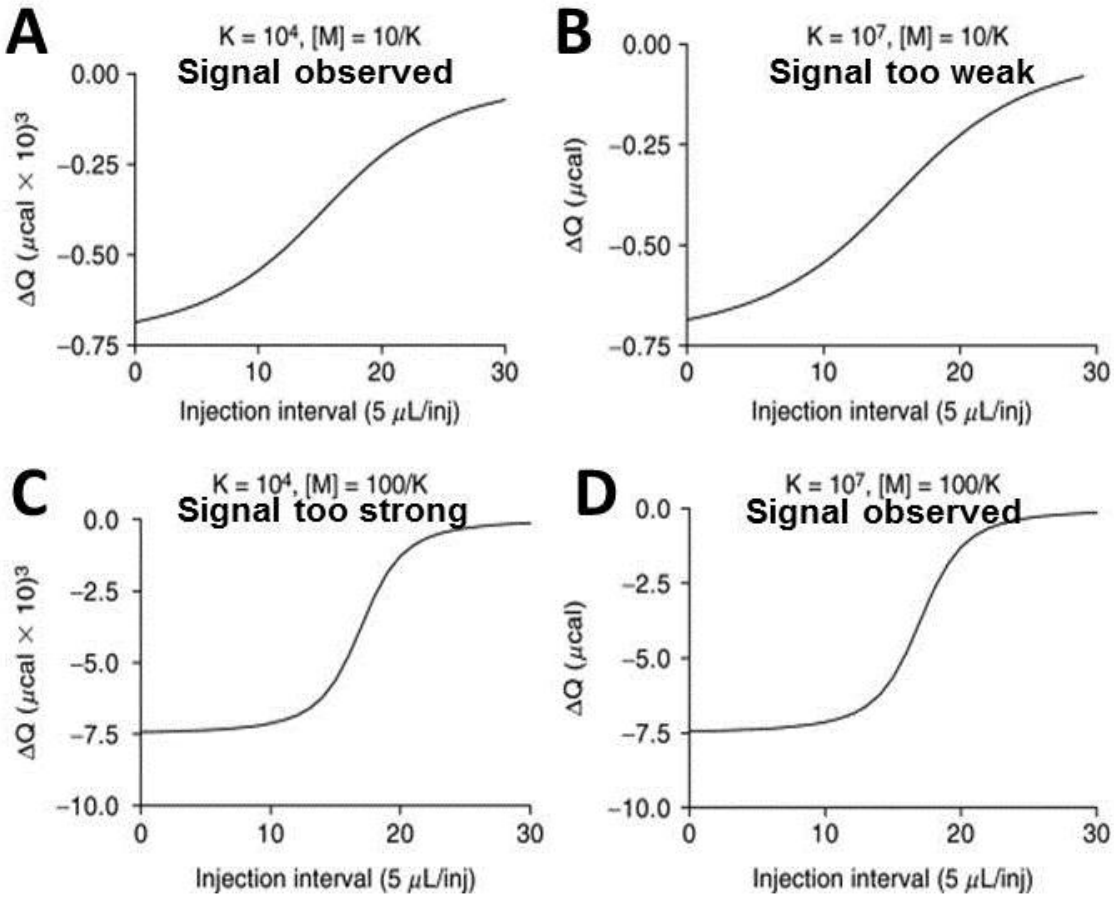


Figure 26: The relationship between isotherm curvature and experimental concentrations. All curves are simulated. In A and B, c = 10; in C and D, C = 100; K_a values are indicated, and are equal between A and C, and between B and D. The experiment depicted in B does not produce enough heat: the high value of K_a must be compensated by a low $[M]_{tot}$ in order to satisfy the c = 10 requirement, resulting in a signal that is too weak and lies below the instrument detection limit. The experiment depicted in C produces too much heat: the low Ka value must be compensated by a high $[M]_{tot}$, resulting in a signal that is too strong and lies above the instrument's detection limit. Adapted from Freyer and Lewis (2008) Methods in Cell Biol.

compared to panels A and B, which can be achieved experimentally by increasing macromolecular concentrations 10-fold. At these increased concentrations, the signal in panel D falls within the detection range of the instrument, while the signal in panel C becomes too strong and falls above the instrument's detection limit [80].

The ligand concentration is also an important consideration. Though it does not influence the c parameter, ligand concentration determines how quickly

the equivalence point is reached during the titration. A ligand concentration that is too high compared to the macromolecule concentration causes the entire isotherm ot shift to the left, resulting in a curve that lacks an initial baseline but has a very long final baseline. The opposite is true for ligand concentrations that are too low and cause the isotherm to shift to the right. Though it is common to start with a 10-fold higher concentration of ligand (as compared to the macromolecule concentration), this parameter often requires adjustment. In most cases the adjustment is minor and the final ligand concentration remains approximately 5-fold to 15-fold higher than the macromolecule concentration. In the case of very weak interactions (i.e. $K_d \geq$ mM range), however, the ligand concentration can reach values several hundred-fold higher than the macromolecule concentration. In cases where the experimenter has an *a priori* estimate of the K_a value of the system in question, the steps outlined in this section can be of help in guiding initial sample concentrations. However, the initial titration experiment usually plays an exploratory role, and concentration adjustments are often necessary in order to obtain an optimal isotherm. In the event that the researcher has no *a priori* knowledge of the K_a value, a good place to start is 10 μM of macromolecule and a ten-fold excess (100 μM) of ligand. The experimenter can then make the necessary adjustments based on this preliminary run.

Since the final results of the ITC experiment are highly dependent on the exact knowledge of the ligand and macromolecule concentrations, accurate concentration measurements are critical. The bicinchoninic acid (BCA) assay [96]

is one of the most accurate and consistent methods for measuring protein concentration, because it relies on interactions of the BCA reagent with the peptide backbone and is not affected by amino acid sequence. Additional methods, including the Pierce 660 nm protein assay, the Bradford protein assay [97], and measurement of absorption at 280 nm are also widely accepted, though results are more dependent on the amino acid sequence or tertiary structure of the protein sample and therefore less accurate. Accurate concentration measurements also require that both the ligand and macromolecule be highly pure. Finally, care should be taken to avoid sample aggregation, as this affects the molar concentration of the sample. For this reason, it is often best to use freshly prepared samples and perform ITC experiments as rapidly as possible.

Finally, it is imperative that the macromolecule and ligand be dialyzed together in the same exact buffer solution (i.e. in the same beaker). Using mismatched buffers (especially with vastly different components, salt concentrations, pH, etc.) can lead to contributions from heats of mixing and dilution that are great enough to mask the heat of binding. Even using different batches of the “same” buffer can affect the accuracy of the final results, though small effects can be subtracted from the raw signal by performing control titrations of buffer into buffer [80, 87].

2.3.B. Optimization of titration parameters

Optimization of the number and volume of injections to be performed over the course of the ITC experiment is important for maximizing the amount of information obtained and the fit of the mathematical binding model to the data. A

typical syringe is capable of storing 300 µl of ligand, which it injects into the 1.4 ml sample cell over the course of the ITC experiment. The number of injections typically ranges from 10-30, and injection volume is usually 3-15 µl, though the number of injections can be as high as 60 or more, and the injection volume can be as small as 1 µl. The use of a larger number of injections with smaller volumes is ideal, as this generates more data points and allows for improved fitting of the data. In some cases, however, the injection volume is dictated by the strength of the signal [80]. The stimulated experiment in Figure 26B (discussed in the previous section) uses 5 µl injections and does not produce enough heat to be detected by the calorimeter. A potential solution to this problem is to increase the injection volume to 15 µl (and decrease the number injections), thus increasing the amount of heat (Q) generated by each injection and producing larger raw signal peaks. Finally, the number and volume of injections can be set to vary over the duration of the ITC experiment. For example, smaller injections spaced more closely together can be used to cover the steep equivalence region of a high affinity complex, thus generating more data points to better define this region [87].

If macromolecule-ligand association occurs relatively rapidly and the signal is not very large, the baseline is recovered rapidly after each injection and four or five minute intervals between injections are sufficient. However, some processes, such as enzymatic reactions that involve the formation or destruction of covalent bonds, generate much larger heat signals and are also slower to

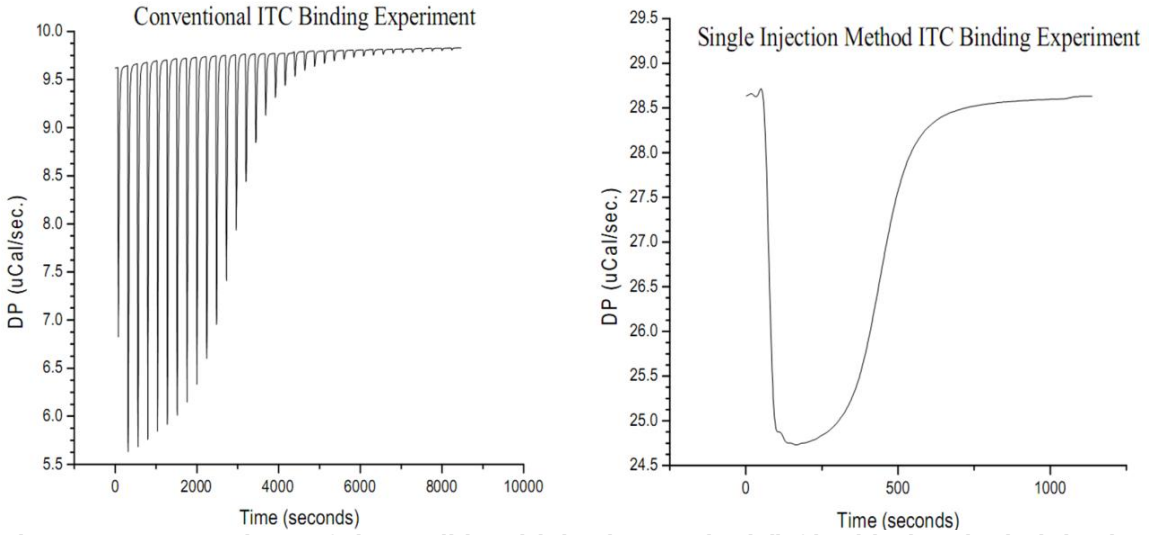


Figure 27: Comparison of the traditional injection method (left) with the single injection method (SIM, right).
The experiment on the left requires one hour or longer to complete, because baseline re-equilibration must be achieved after each injection. The SIM experiment is completed in 20 minutes. Adapted from MicroCal, *VP-ITC User's Manual.*

reach equilibrium. In this case, the interval between injections may require 10 minutes or even longer [87, 93].

In some cases, it is more advantageous to use the single injection method (SIM) instead of the traditional multi-injection approach. The SIM uses one single long injection, whose volume can be as large as the syringe volume, to continuously introduce ligand into the sample cell over the duration of the entire experiment (usually ~20 minutes) (Figure 27). Once the ligand becomes saturating, the signal returns to the baseline value and the total ΔH is calculated by integrating the resulting peak. This method is advantageous because experimental time is typically reduced by ~1/2 to 1/3 when compared with traditional ITC, and the number of data points is increased ~30-fold, allowing for a more accurate determination of K_a in tightly binding complexes [98]. Currently the method is optimized for samples exhibiting linear relationships (i.e. 1:1, 2:2, etc.), and has a higher error rate than traditional ITC when applied to more

complicated systems due to noisier baselines. In addition, SIM experiments require higher macromolecule concentrations, as the technique works best when $c \geq 100$. SIM ITC is especially useful for drug development studies where, in combination with automated data acquisition and analysis, it can be used in a high throughput manner for lead screening.

SIM ITC is also used in kinetic studies to determine ΔH. The traditional method for conducting kinetic studies by ITC requires two separate experiments in order to generate a Michaelis-Menten curve. The first experiment is a multi-titration experiment conducted in the presence of saturating substrate concentration, and is used to determine the reaction rate v_0 as described in Section 2.2.D. In this experiment, injections are spaced such that no more than 5% of the substrate is converted to product between injections (steady state conditions). ΔH, however, must be determined from a second experiment where substrate concentration is not saturating and the reaction can proceed to completion (non-steady state conditions). This experiment is usually a short single injection experiment (25 µl). ΔH is obtained as described for binding experiments, by integration of the single injection peak [80, 92, 93].

A second approach to kinetics experiments makes use of one single SIM ITC experiment to obtain v_0 and ΔH at once, eliminating the need for two separate experiments. This experiment looks very similar to the binding SIM experiment shown in Figure 27, and involves continuous rate measurements. The substrate concentration can be calculated for any given time, and the rate can be plotted as a function of substrate concentration to give a continuous

kinetic curve. Because the experiment is allowed to proceed until the substrate is depleted, the signal returns to the pre-injection baseline and ΔH can be obtained by integrating the resulting peak. The advantage of using the traditional method over the SIM method to calculate the reaction rate is that discrete v_0 values can be directly calculated for well-determined substrate concentration values at each injection, thus introducing less statistical error during data analysis [80].

2.4. Data analysis

2.4.A. Correction of the raw data for extraneous heat contributions

Provided that the ligand and macromolecule are dialyzed into the same buffer solution, the apparent enthalpy (ΔH_{app}) measured during the experiment is a result of the following contributions:

$$\Delta H_{app} = \Delta H_{bind} + \Delta H_{X\,dil.} + \Delta H_{mech.} + n_H \bullet \Delta H_{ion} \quad (16)$$

where ΔH_{bind}, the enthalpy of binding, is usually the only value of interest. $\Delta H_{X\,dil}$ is the heat of dilution of the ligand as it is titrated into the macromolecule, $\Delta H_{M\,dil}$ is the heat of dilution of the macromolecule, ΔH_{mech} is the heat given off by mechanical processes associated with the experiment, such as stirring of the solution, ΔH_{ion} is the enthalpy of ionization of the buffer for reactions that involve the release or uptake of protons during binding, and n_H is the number of protons that are absorbed (or released, if n_H is negative) by the macromolecule-ligand complex upon binding. In practice, $\Delta H_{M\,dil}$ is usually negligible, and contributions made by $\Delta H_{X\,dil.} + \Delta H_{mech}$ can easily be determined by performing a control experiment where ligand is titrated into buffer, and the value subtracted from ΔH_{app}.

Correcting ΔH_{app} for $n_H \bullet \Delta H_{ion}$ is more complicated and requires several steps. Since n_H is usually unknown but ΔH_{ion} values for several buffers have been experimentally determined, ΔH_{app} can be corrected for $n_H \bullet \Delta H_{ion}$ by performing the titration in two different buffers with different ΔH_{ion} values. Alternatively, a buffer with $\Delta H_{ion} \approx 0$ can be used, samples permitting, thus avoiding the problem altogether. It is useful to note that in the case of very weak signals, a buffer with a high ΔH_{ion} value can be exploited to enhance the observed signal [80].

2.4.B. Fitting of thermodynamic data to a binding model

In Section 2.2.B., the following equation was introduced, relating the observed heat (Q) obtained from integration of the raw ITC signal to the volume of the reaction cell (V) and the change in the concentration of bound ligand:

$$Q = V \bullet \Delta H \bullet \Delta X_b \qquad (2)$$

ΔX_b is not an experimentally known value (only the total ligand concentration $[X]_t$ is known), and Equation 2 must be rewritten in terms of experimentally measurable parameters. As mentioned in Section 2.2.B., the final mathematical expression depends on the particular system under study. The presence of complicating factors including multiple binding sites, cooperativity between multiple binding sites, and the presence of several species in competing equilibria all complicate the mathematical model used to describe the binding interaction [85].

For the simplest case, where one ligand binds to one macromolecule via one binding site ($M + X \rightleftharpoons MX$), Equation 2 is rewritten as follows:

$$Q = \frac{n[M]_t V \Delta H}{2}\left[1 + \frac{[X]_t}{n[M]_t} + \frac{1}{nK_a[M]_t} - \sqrt{\left(1 + \frac{[X]_t}{n[M]_t} + \frac{1}{nK_a[M]_t}\right)^2 - \frac{4[X]_t}{n[M]_t}}\right] \quad (17)$$

where $[M]_t$ and $[X]_t$ are the total macromolecule and ligand concentrations, V is the reaction volume (sample cell volume). ΔH is the enthalpy of binding, K_a is the association constant, and n is the binding stoichiometry. $[M]_t$, $[X]_t$, and V are known values, and ΔH, K_a, and n are independent variables obtained directly from the data fitting process. During data fitting, the mathematical model is fit to the integrated ITC data (see Figure 24) using nonlinear least squares analysis, and the values of ΔH and K_a that best fit the data are reported. Though n can be used as a floating parameter during data fitting, more accurate results are obtained if n is known ahead of time. It is desirable to minimize the number of variables in the binding model, because this reduces error associated with data fitting [86, 99].

Data analysis software is provided with all commercial ITC instruments, and the fitting of simple binding models to straightforward data can be easily achieved by any user. Software packages are pre-loaded with binding models for the analysis of systems with: 1) one set of binding sites (Figure 28A), which include the ”simplest case” 1:1 interaction described above; 2) multiple sets of independent binding sites (Figure 28B); and 3) sequential binding (Figure 28C), which refers to cooperative binding. In this type of system, the binding of a first ligand molecule to the first site directly affects the affinity of a second ligand molecule for a second binding site on the same macromolecule.

As the biological system under study increases in complexity, the mathematical binding model also becomes increasingly complicated and

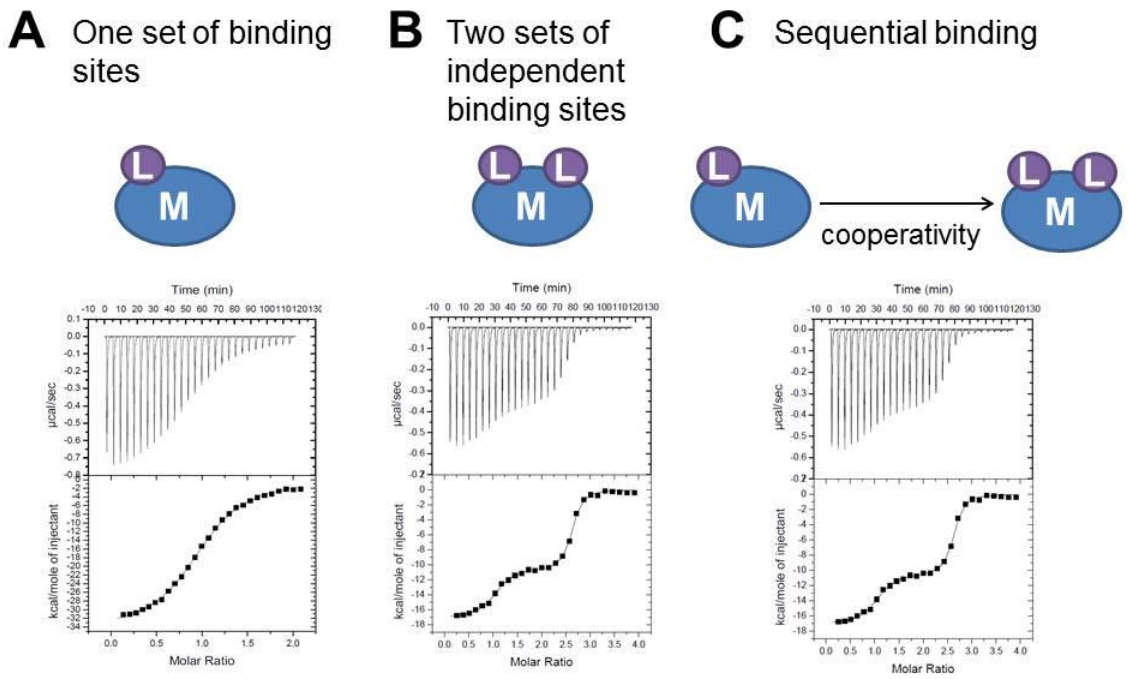


Figure 28: Commonly encountered binding models used for fitting ITC data.
A) The simplest binding model, "one set of binding sites," is used when one ligand, or one set of ligands, binds to one molecule and produces a single phase isotherm. This mathematical model uses Equation 17 to fit the data. B) The "two sets of independent binding sites" model describes a system where two ligands, or two sets of ligands, bind to two binding sites on one macromolecule independently of each other. This type of interaction produces a bimodal isotherm. C) "Sequential binding" describes a system where two ligands bind to two binding sites on one macromolecule in a cooperative fashion, where the binding of the first ligand molecule influences the binding of the second. This interaction also produces a bimodal isotherm.

Equation 17 becomes longer. As the number of binding sites increases, so does the number of partially formed complexes that are present at equilibrium. These partial complexes must all be represented in the mathematical model used to describe the interaction. Cooperative binding is even more difficult to describe mathematically, as the number of possible species present at equilibrium is even higher.

In cases where one ligand binds one macromolecule via multiple binding sites, a single phase isotherm is expected. Figure 29A shows the single phase isotherm observed for the titration between the viral CN inhibitor A238L and CN. A238L binds to CN via two recognition sites: a PxIxIT and an LxVP motif. The

multiple site binding models described above (Figure 28B and C) cannot be used to extract information about the individual binding sites in a system like A238L-CN, as they are designed for the analysis of systems where multiple ligand molecules bind to one macromolecule. Analysis of such 1:1 ligand:macromolecule complexes is limited, and only the overall, or "global," binding information can be obtained. The single phase isotherm arises because the two binding sites on the ligand are tethered to each other and bind in unison (Figure 29B). The first binding event immediately brings the second binding site of the ligand in close proximity to the second binding site of the macromolecule,

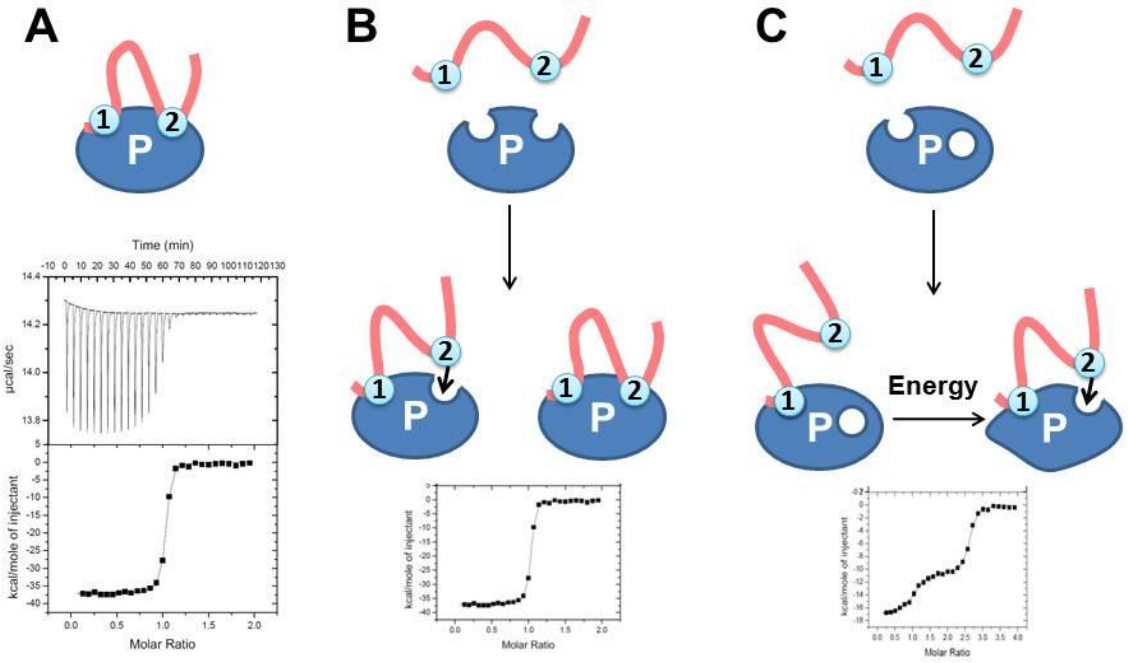


Figure 29: ITC analysis of 1:1 ligand:protein complexes with multiple binding sites.
A) One ligand molecule interacting with one protein molecule via two recognition sites produces a single phase isotherm. Shown is the isotherm obtained by titrating the viral CN inhibitor A238L into CN. A238L has two CN recognition motifs: LxVP and PxIxIT. B) A description of the binding events that give rise to the isotherm in A. The two binding events in this 1:1 system cannot be distinguished by ITC because the two binding sites are linked together and behave as one. C) A 1:1 complex would produce a bimodal curve only when the first binding event must overcome a large activation energy barrier to expose the second binding site on the protein. The change induced by the first binding event must involve a high amount of energy, and could include the breaking or formation of bonds, major structural rearrangements, or the ejection of active site metals. This type of system requires that the first binding site be fully saturated before the second binding event can take place.

increasing its local concentration. Therefore, both the first and second binding events occur nearly simultaneously despite differences in the individual affinities of each interaction, giving rise to the single phase isotherm. It is rare that a bimodal isotherm arises from a 1:1 complex (Figure 29C). In such a case, the first binding at the first site must overcome a high activation energy barrier to induce a major change in the macromolecule and expose the second binding site.

2.4.C. Experimental approaches for simplifying complicated data

Proteins have large recognition interfaces with complex binding interactions. It is therefore often necessary to simplify these interfaces in order to extract useful binding information from ITC. The simplest way to achieve this is to take a "piece-by-piece" approach, performing ITC titrations with several truncated ligand or protein constructs. Alternatively, mutagenesis of conserved sequence motifs known to participate in binding can be used to eliminate a specific interaction and simplify the binding isotherm. This approach is used in Chapter 4 of this thesis to study the interaction between CN and its regulator RCAN1. Primary sequence analysis of RCAN1 reveals at least four conserved CN-binding motifs. ITC titrations were carried out between CN and several RCAN1 constructs that were either truncated or mutated to systematically remove individual CN-binding motifs. These experiments provided information that was crucial for determining the mechanism of complex formation.

Another possible approach to simplifying the analysis of ITC data is to perform a "reverse" titration, where the macromolecule is added to the ligand.

This is not always possible, as many proteins are not stable at the higher concentrations required for this experiment (the protein in the syringe must be several-fold more concentrated than the ligand in the sample cell). This “reverse” titration can be used to accurately determine the stoichiometry of a reaction, or to verify parameters obtained from the “forward” titration.

2.5. Concluding remarks

ITC is a powerful technique that has gained popularity with the commercialization of user-friendly instruments and software. Over the past few years, highly sensitive “nano ITC” instruments with sample cells as small as 300 µl (compared to the 1.4 ml sample cells in traditional instruments) have become available, and they are rapidly gaining in popularity. ITC and structural biology techniques are complementary. Structural data of protein-protein interfaces can serve as the basis for the design of ITC experiments that describe the mechanism of complex formation. For example, ITC experiments can detect cooperative binding, or binding-induced conformational changes that may not otherwise be apparent from structural data alone. ITC can also be used to guide the optimization of crystallographic constructs. For example, many protein complexes contain flexible regions that are not involved in complex formation and impede crystallization. ITC experiments with various constructs can guide the design of truncated ligand or protein constructs that lack such unbound flexible regions.

In the following chapter, ITC is used in conjunction with crystallographic data to determine the individual contributions of the PxIxIT and LxVP CN-binding sites in the viral CN inhibitor A238L. In Chapter 4, ITC and NMR are combined to

determine the mechanism by which the CN regulator RCAN1 recognizes CN and modulates its activity. In both cases, ITC provides valuable information that cannot be obtained from structural data.

Chapter 3. The molecular mechanism of substrate engagement and immunosuppressant inhibition of Calcineurin

Reproduced from the journal PLoS Biology, Volume 11, Number 12, e1001492

Supplementary material that accompanies this paper is located in **Appendix A**

Author Contributions:

Conceived and designed the experiments: SG RB PC GH RP MSC WP. Performed the experiments: SG RB NL JC PC RP WP. Analyzed the data: SG RB PC GH RP MSC WP. Wrote the paper: SG RB PC GH RP MSC WP.

Simina Grigoriu[1,2], Rachel Bond[3], Pilar Cossio[4], Jennifer A. Chen[3], Nina Ly[3], Gerhard Hummer[4], Rebecca Page[2], Martha S. Cyert[3*] and Wolfgang Peti[1,5*]

[1]Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI, 02912; [2]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, 02912; [3]Department of Biology, Stanford University, Stanford, CA, 94305; [4]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, MD 20892, USA; [5]Department of Chemistry, Brown University, Providence, RI, 02912.

Running Title. Structure of the calcineurin-A238L complex.

Abbreviations

AID, auto-inhibitory domain; CN, Calcineurin; CNA, calcineurin catalytic A subunit; CNB, calcineurin regulatory B subunit; CSA, cyclosporin A; DFT, density functional theory; ITC, isothermal titration calorimetry; MD, Molecular Dynamics; NFAT, nuclear factor of activated T-cells; pNPP, p-nitrophenyl phosphate; PP1, Protein Phosphatase 1; PP2B, Protein Phosphatase 2B; PP3, Protein Phosphatase 3; PPP, phosphoprotein phosphatase; RII, RII regulatory subunit of PKA

Accession Numbers

The structure factors and coordinates for the CN-A238L complex have been deposited with the Protein Databank with accession number 4F0Z.

Acknowledgements

The authors thank Dr. M. Allaire (National Synchrotron Light Source, NSLS) for his support at NSLS beamline X25. Use of the NSLS at Brookhaven National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-98CH10886. MD calculations were performed on the Biowulf computing cluster at NIH. The authors thank Dr. Linda Dixon for providing information and materials relating to A238L, and Dan Herschlag, Jagoree Roy and Evan Guiney for helpful discussion and reading of the manuscript. We thank Dara Dowlatshahi for technical advice

and Dr. V. Kaila (National Institutes of Health) for his support with the quantum mechanical calculations of the model substrate.

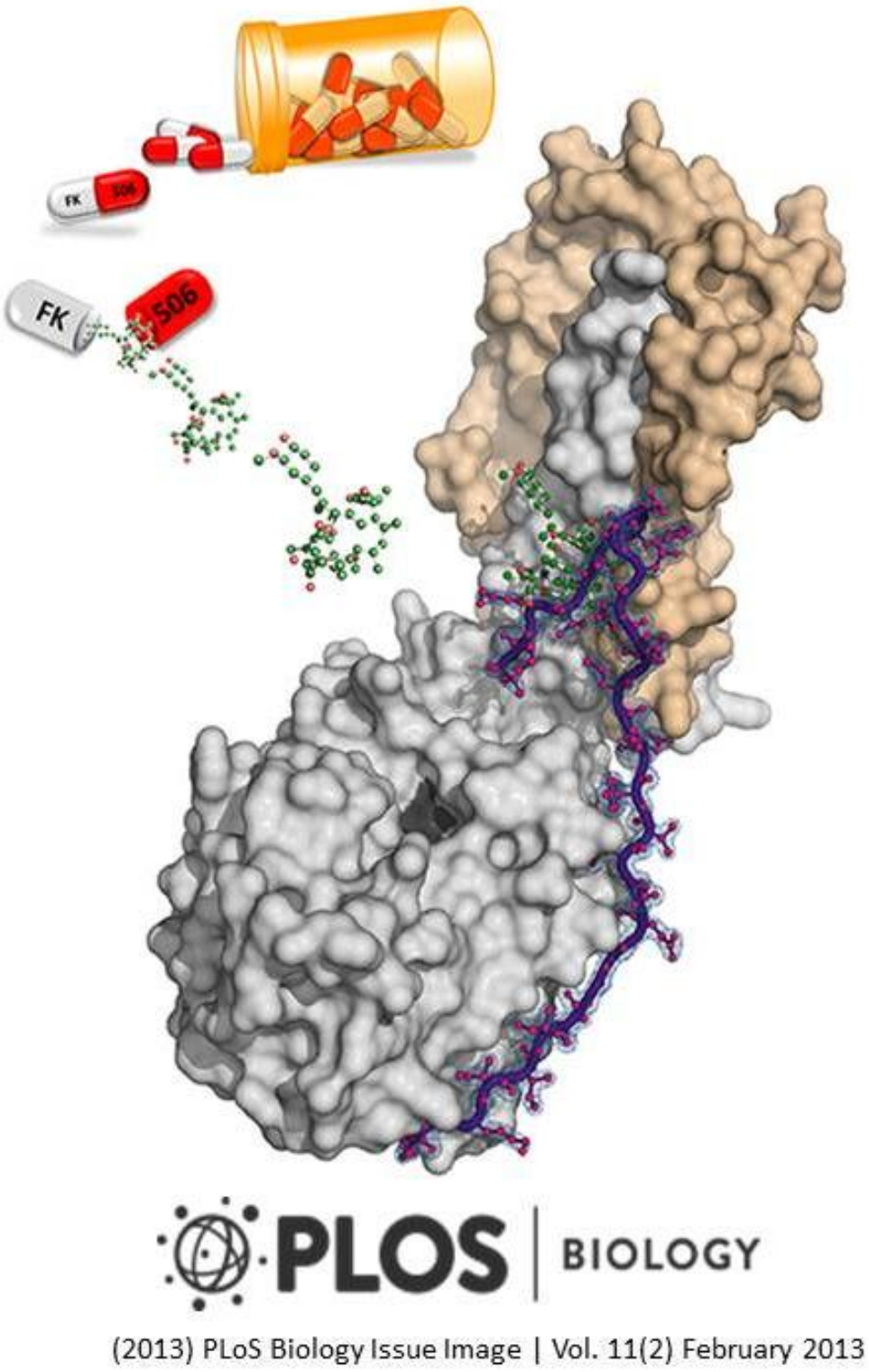


Figure 30: PLoS Biology Feb. 2013 Issue Cover Image.

3.1. Author Summary

Transplantation medicine was revolutionized by the introduction of the immunosuppressant drugs cyclosporin A and FK506 that prevent rejection of transplanted organs by the recipient's immune system. These drugs work by inhibiting calcineurin, a conserved protein phosphatase. Calcineurin regulates the immune response by dephosphorylating and activating the members of the NFAT family of transcription factors, which in turn activate genes required for the antigen-dependent stimulation of T-cells. Despite its biological and clinical importance, we have only a limited understanding of how calcineurin and other protein phosphatases interact with their substrates and target specific phosphorylated residues for dephosphorylation. Here, we determined the structure of calcineurin in complex with A238L, a viral peptide inhibitor of its function. This study shows that the viral peptide inhibits calcineurin not by targeting its active site but rather by occupying two critical substrate-binding regions of calcineurin (distant from each other and from the active site), thereby preventing its interaction with protein substrates. These findings allow us to present the first computational model of calcineurin bound to a phospho-substrate at its active site. Furthermore, by elucidating the structural basis for one particular mode of substrate–calcineurin interaction, this study reveals that both this viral peptide and immunosuppressant drugs inhibit calcineurin by blocking substrate access to a single critical region of the enzyme.

3.2. Abstract

Ser/Thr phosphatases dephosphorylate their targets with high specificity, yet the structural and sequence determinants of phosphosite recognition are poorly understood. Calcineurin (CN) is a conserved Ca^{2+}/calmodulin-dependent Ser/Thr phosphatase and the target of immunosuppressants, FK506 and cyclosporin A (CSA). To investigate CN substrate recognition we used X-ray crystallography, biochemistry, modeling, and in vivo experiments to study A238L, a viral protein inhibitor of CN. We show that A238L competitively inhibits CN by occupying a critical substrate recognition site, while leaving the catalytic center fully accessible. Critically, the 1.7 Å structure of the A238L-CN complex reveals how CN recognizes residues in A238L that are analogous to a substrate motif, "LxVP." The structure enabled modeling of a peptide substrate bound to CN, which predicts substrate interactions beyond the catalytic center. Finally, this study establishes that "LxVP" sequences and immunosuppressants bind to the identical site on CN. Thus, FK506, CSA, and A238L all prevent "LxVP"-mediated substrate recognition by CN, highlighting the importance of this interaction for substrate dephosphorylation. Collectively, this work presents the first integrated structural model for substrate selection and dephosphorylation by CN and lays the groundwork for structure-based development of new CN inhibitors.

3.3. Introduction

The vast majority of eukaryotic proteins are phosphorylated, and this modification rapidly and reversibly modulates protein dynamics, interactions, activities, localization and/or stability [25]. This essential regulation is carried out by the opposing activities of a large array of protein kinases, and a surprisingly small cadre of phosphoprotein phosphatases. Despite decades of investigation, basic questions about how these phosphatases act on phosphosites that share little similarity in primary sequence remain unanswered [40]. Here, we unveil a key mechanism of substrate recognition by calcineurin (CN) [30], the highly conserved Ca^{2+}/calmodulin activated Ser/Thr phosphatase (also called Protein Phosphatase 2B (PP2B) or Protein Phosphatase 3 (PP3)), establish that structurally unrelated inhibitors of CN specifically disrupt this interaction, and show that substrates engaged at this site have additional interactions to orient the phosphosite towards the catalytic center of the enzyme.

Calcineurin is ubiquitously expressed, and is particularly abundant in the brain. By dephosphorylating a variety of protein substrates in response to Ca^{2+} signals, CN regulates development, learning and memory, cardiac function and the immune response [30]. One of the best-studied activities of CN is its dephosphorylation of the nuclear factor of activated T-cells family of transcription factors (NFATc1-c4), which allows NFAT to translocate to the nucleus where it induces the expression of genes required for T-cell activation [100]. Because inhibition of NFAT signaling suppresses T cell activation, natural products that

specifically inhibit CN, cyclosporin A (CSA) and FK506, are widely prescribed as immunosuppressants to prevent post-transplant organ rejection [101].

CN is a member of the PPP family of protein phosphatases, which also includes PP1 and PP2A, among others. These enzymes contain structurally related catalytic domains, which rely on coordinated metal ions to directly bind phosphate and hydrolyze phospho-Ser/phospho-Thr [40]. Despite their similarities, substrate recognition by these phosphatases is distinct, and natural products selectively inhibit each enzyme. In contrast to PP1 and PP2A, whose catalytic subunits combine with different regulatory subunits to create a suite of distinct holoenzymes [40], CN is always composed of a catalytic A subunit (CNA), bound to a regulatory B subunit (CNB) that binds four Ca^{2+} ions [1]. The C-terminus of CNA also contains a calmodulin-binding domain and an auto-inhibitory domain (AID), which regulate CN activity. Under basal Ca^{2+} conditions, the AID interacts with the catalytic center and prevents dephosphorylation. During signaling, increased Ca^{2+} levels cause Ca^{2+}-loaded calmodulin to bind CNA, which displaces the AID from the catalytic site and stimulates CN phosphatase activity [46].

Despite the critical biological importance of CN, the molecular mechanisms that allow CN to recognize and dephosphorylate specific protein substrates are still not well understood. Some, but not all substrates contain a short CN-binding motif, termed “PxIxIT” for its consensus sequence in NFATc1-c4 [102]. This sequence, which also occurs in scaffold proteins, i.e. AKAP79 [10], binds to CNA at a groove distal from the catalytic center. PxIxIT peptides bind

with equal affinities to the inactive or active form of CN [59]. Because they do not occlude or alter the CN active site, they fail to inhibit dephosphorylation of either a model phosphopeptide substrate or a small molecule, p-nitrophenyl phosphate (pNPP) [64, 67]. Thus, while this interaction can improve dephosphorylation efficiency by tethering a substrate to CN, it is not essential to the mechanism of dephosphorylation.

There has been limited insight into which substrate features do influence catalysis. Short peptides are not efficiently acted upon by CN; however, phosphopeptides that contain a basic residue at the -3 position relative to the phosphosite show 4-fold better dephosphorylation [103]. Examination of one protein substrate, the RII regulatory subunit of PKA, defined a 19-mer as the smallest peptide that was robustly dephosphorylated (K_m= 26 µM; V_{max}= 1.7 µmol min^{-1} mg^{-1}), and showed that N-terminal residues (DLDV) lying ten amino acids upstream of the phosphosite, were critical for substrate recognition [104]. Subsequent studies of the NFAT family (NFATc1-c4) similarly identified a conserved CN-interaction site in these proteins, "ΦLxVP", which interacts with Ca^{2+}/calmodulin-activated CN, and contributes to efficient dephosphorylation [102]. However, this sequence is significantly displaced from and C-terminal to CN-regulated phosphosites. Thus, the molecular details of the "ΦLxVP"-CN interaction and its role in dephosphorylation are still unclear.

Finally, CN is inhibited by the fungal-derived immunosuppressant drugs CSA and FK506, which bind the immunophilin proteins cyclophilin and FKBP, respectively, and engage CN as drug-immunophilin complexes. Structural

analyses revealed that, unlike the AID, these drug-immunophilin complexes do not target the active site, but instead bind in a pocket ~30 Å away at the interface of the CNA/CNB subunits [1, 17, 18]. However, while pNPP can be readily dephosphorylated by CN:drug-immunophilin complexes, CN phosphoprotein and phosphopeptide substrates cannot[71], suggesting that the presence of the drug-immunophilin complexes impedes substrate/CN binding [64]. The detailed mechanism by which immunosuppressants achieve CN inhibition is still unclear [105, 106].

Here, we describe the first high resolution structure of CN bound to a physiological binding partner: the protein inhibitor A238L from African swine fever virus (ASFV), a highly virulent double-stranded DNA virus that infects domestic pigs in Africa and Europe. Upon infection, A238L suppresses the host immune response by inhibiting both NFkB and CN. While A238L contains a “PxIxIT”-type anchoring sequence, the molecular mechanism by which A238L inhibits the enzyme is unknown [73]. Our studies demonstrate that A238L competitively inhibits CN and that residues in A238L (“FLCV”) directly compete with substrates for binding to a substrate-recognition cleft, the ΦLxVP cleft, in activated CN. Our 1.7 Å crystal structure of the CN-A238L complex reveals the molecular interactions that mediate this key element of substrate recognition, and surprisingly shows that A238L effectively inhibits CN not by blocking and occluding the active site, which is fully accessible in the complex, but instead by binding and blocking the ΦLxVP substrate recognition groove. This structure enabled molecular dynamics (MD) modeling of a minimal substrate, the RII

peptide, bound to CN. The model reveals that interactions at the hydrophobic substrate-recognition groove are augmented by charged interactions at position -3 upstream of the phosphosite [103]. These studies provide the first structural insights into active site substrate engagement for any Ser/Thr phosphatase. Thus, this work elucidates a key mechanism by which CN recognizes substrates and provides structural insights into the presentation of phosphosites to the active site during dephosphorylation. Furthermore, the CN-A238L structure also unequivocally shows that the LxVP sequence and immunosuppressants bind to the identical site on CN. Thus, these studies finally establish the mechanism of action of these drugs and lay the foundation for renewed efforts in the structure-based targeted design of novel CN inhibitors.

3.4. Materials and Methods

3.4.A. Proteins

CNA (residues 1-370 or 1-391), CNB (residues 1-170) and wt- and mutant A238L (residues 200-239) were subcloned into vectors containing either an N-terminal His_6- or GST-tags and expressed in *E. coli*. To form the CN-A238L (CNA, 1-370; CNB, 1-170; A238L, 200-239) complex used for X-ray crystallography, His_6-tagged CN was purified over a HisTrap HP column and directly eluted into an at least two-fold molar excess of previously purified A238L. His_6-tag cleavage and subtraction purification were performed and the complex was further purified by SEC (Superdex 200 26/60) in 20 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM TCEP, 1.5 mM $CaCl_2$. Additional cloning, protein expression and

protein purification procedures are provided in the Supporting Material (Appendix A).

3.4.B. Crystallization and structure determination

The CN-A238L complex (~7 mg/ml) formed thin plate crystals in 0.16 M ammonium citrate and 20% (w/v) PEG3350 at 22°C. Crystals were obtained using the sitting drop vapor diffusion method (3-well Intelliplate, Art Robbins), with 0.6 µl drops containing a 1:2 ratio of precipitant solution to protein and 50 µl of precipitant solution in the reservoir. Crystals were cryo-protected using a 10 min soak in mother liquor supplemented with 30% glycerol and immediately flash frozen in liquid N_2. Data were collected at the NSLS X25 beamline at Brookhaven National Laboratory. Crystals of CN-A23L formed in space group $P2_1$, with unit cell dimensions *a*=72.69 Å, *b*=48.98 Å, *c*=82.44 Å, and *β*=104.4°. Data were indexed, integrated and scaled with *DENZO* and *SCALEPACK* as part of *HKL2000* [107]. The structure of CN-A238L was determined to 1.7 Å by molecular replacement using a CN heterodimer molecule (CNA/B; PDBID 1AUI [1]; the auto-inhibitory domain was omitted) as the search model. The final model of the CN-A238L complex was obtained using iterative rounds of refinement in Phenix [108] and model building in Coot [109], with TLS used in the final round of refinement. The asymmetric unit contains one copy of the CN-A238L heterodimer/inhibitor complex. The final structure refined to a final R factor of 15.8% (R_{free} = 17.8%). CN-A238L crystals formed at pH 5, leading to protonation of active site residues and thus displacement of the positively charged active site metal ions, as previously seen by Jin & Harrison who crystallized CN at pH

4.6[18]. 100% of all residues are in the allowed region of the Ramachandran diagram. Structure validation and stereochemistry analysis was performed with Molprobity [110]. Details and statistics of data analysis and model building are provided in Table S1 and Supporting Methods. A stereoview of the A238L electron density is in Figure S5.

3.4.C. Molecular dynamics simulations

The CN-A238L structure was used as the starting structure for the MD simulations. The phospho-Ser (pSer) was placed in the active according to the PDB structure 1TCO, together with the expected OH ligand [111]. We used density functional theory (DFT) with the B3LYP functional [112], as implemented in TURBOMOLE [113], to obtain an energy minimized structure of the CN active site including the partial charges of the active site metal ions and their ligands. During the MD simulations, the active site region was restrained to the DFT-optimized configuration. Initial structures of RII bound to CN were built according to the backbone coordinates of A238L at the LxVP site, and the last 8 residues of the auto-inhibitory domain in PDB structure 1AUI [1]. Ten different starting models for the intervening sequence (^{86}IPGRFD^{91}) were constructed using the ROSETTA loop model tool [114]. The resulting ten CN-RII complexes were inserted into rectangular boxes (8x8x12 nm^3) with 26190 TIP3P water molecules [115] and 8 Na^+ ions each for electro-neutrality. MD simulations were performed using GROMACSv4.5.3 [116], the Amber ff99SB-ILDN force field [117], particle mesh Ewald summation [118] using a 0.12 nm grid spacing, a time step of 2 fs and a 0.9 nm real-space cutoff at a constant temperature [119] of 300 K and at a

pressure [120, 121] of 1 bar. After 2 ns of equilibration of each of the 10 starting models, bias-exchange metadynamics [122] using the PLUMED1.3.0 plugin [123] was used to accelerate the conformational sampling. In 9 of the 10 replicas, the ψ dihedral angles of RII residues Ile86, Pro87, Arg89, Phe90, Asp91, Arg92, Arg93, Cys97 and Ala98, respectively, were biased, while replica 10 was kept unbiased. To ensure RII stayed bound to CN during the conformational sampling, the Fe^{3+}-pSer and LxVP interactions were harmonically restrained. Convergence was reached after 1.15 μs (115 ns/replica), as judged by the free energy profiles of the biased ψ. The RII structures of the unbiased replicas were clustered according to a 1.2 Å backbone root-mean-square-distance (RMSD) threshold. The most populated cluster contains 58% of all structures, from which a representative was selected with the lowest Coulomb energy. A different clustering technique using reweighting of the biased replicas [124] led to similar results.

3.4.D. Isothermal Titration Calorimetry

ITC experiments were performed at 25°C using a VP-ITC microcalorimeter (GE Healthcare). All protein samples were equilibrated in ITC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1.5 mM $CaCl_2$, 0.5 mM TCEP). Wild-type or mutated A238L was titrated into $CN_{A1\text{-}391/B1\text{-}170}$. Titrant (10 µL per injection) was injected into the sample cell over a period of 20 seconds with a 250 second interval between titrations to allow for complete equilibration and baseline recovery. 28 injections were delivered during each experiment, and the solution in the sample cell was stirred at 307 rpm to ensure rapid mixing. Data were analyzed with one set of

sites binding model, based on the 1:1 stoichiometry observed in the crystal structure, using Origin 7.0 (OriginLab).

3.4.E. CN activity assays with RII

The rate of RII phosphopeptide dephosphorylation by calcineurin was determined by measuring the total phosphate released over four time points (total 5-20 min). Reaction rates were linear over this time period and constituted less than 1% of product formation. 50 µl reactions contained assay buffer (50 mM Tris pH 7.4, 100 mM NaCl, 6 mM $MgCl_2$, 0.5 mM $CaCl_2$, 0.1% PEG 3250, 0.5 mM DTT), 10 nM calcineurin, wt- or mutant A238L (0-10 µM) and 50-1000 µM RII phosphopeptide. Reactions were performed at 37°C and were initiated by the addition of 10x RII and terminated using 100 µl Biomol Green Reagent (Enzo Life Sciences). After color development for 20 min, absorbance was measured at 595 nm and compared to phosphate standards of known concentration to determine the amount of phosphate released. Kinetic constants were determined with GraphPad Prism by fitting the points to the Morrison model for tight-binding inhibitors using nonlinear regression analysis[23]. Competitive and noncompetitive models were compared using the corrected Akaike information criterion, AICc [125]. Phosphorylated RII was synthesized by the Tufts University Core Facility.

3.4.F. CN activity assays with pNPP

The rate of calcineurin hydrolysis of pNPP was measured in a continuous assay by monitoring the production of pNP at 415 nm. Reaction rates were linear for at least one hour and reaction progress was < 0.1%. 100 µl reactions contained assay buffer (100 mM Tris pH 7.5, 100 mM NaCl, 0.4 mM $CaCl_2$, 100 µg/ml BSA,

1 mM $MnCl_2$, 0.5 mM DTT), 10-20 nM truncated calcineurin, wt or mutant A238L (0-8 µM) and 10 mM pNPP. Experiments were performed at room temperature. Standards of known pNP concentration were used to convert absorbance units to pNP concentration.

3.4.G. GST pull-down and competition assays

Cells extracts were prepared by resuspending cell pellets in lysis buffer (50 mM Tris, pH 7.4, 100 mM NaCl, 2 mM EDTA, 2 mM EGTA, 5 mM DTT, 1 mM PMSF, 5 µg/ml each pepstatin, leupeptin, aprotinin, and benzamidine) and lysed by sonication. Cell debris was pelleted by centrifugation (20,000x*g*, 20 min) and clarified lysate was brought to 0.1% Tween-20 and stored in aliquots at -80°C. For GST-peptide fusion experiments, 50-200 µg cell extracts containing GST or GST fusion proteins were bound to glutathione sepharose 4B beads (GE Healthcare), washed 3x with wash buffer (10 mM Tris, pH 8.0, 110 mM KOAc, 2 mM MgOAc, 0.1% Tween-20), incubated with 200 µg $calcineurin_{A1\text{-}391/B1\text{-}370}$ cell extract with or without 200 µM competing peptide, and finally washed 3x with wash buffer (containing competing peptides if present in the previous step). For GST-CNA1 experiments involving S-$A238L_{157\text{-}239}$, 20 µg cell extracts containing co-expressed GST-CNA1 and CNB1 were bound to beads as described above, then incubated with 350 ng purified S-$A238L_{157\text{-}239}$ in the presence or absence of competing peptide. In both experiments, bound proteins were eluted by boiling samples in Laemmli buffer. Peptides for competition assays were synthesized by the Tufts University Core Facility. The amino acid sequences were: LxVPc1,

DQYLAVPQHPYQWAK; LxVPmut, DQYAAAAQHPYQWAK; PVIVIT, GPHPVIVITGPHEE; PVIVITscrambled, GPIVPIHVTHPGEE.

3.5. Results

3.5.A. A238L engages CN via both an LxVP and a PxIxIT docking motif

To understand how substrates engage CN, we investigated A238L (aa 157-239, Malawi LIL20-1 strain), a protein inhibitor from the African swine fever virus (ASFV) that suppresses the host immune response by inhibiting both NFkB and CN [73]. Two substrate motifs, PxIxIT and LxVP, have been shown to bind to CN [102]. While it was known that a C-terminal fragment of A238L contains a PxIxIT motif that anchors the protein to CN [74], we showed that excess amounts of a peptide that encodes the LxVP docking motif from NFATc1, but not a mutant (LxVPc1mut LAVP → AAAA) also disrupted A238L-CN binding (Figure S1A). This establishes that A238L also contains an LxVP motif. The only sequence in $A238L_{157-239}$ that is similar to the LxVP motif is LCVK (Figure 31A). To verify that this sequence binds CN, we generated a GST fusion protein containing the LCVK sequence from A238L (but lacking the PKIIIT sequence) and showed that it forms a complex with recombinant CN that is disrupted by the addition of a peptide encoding the LxVP motif from NFATc1, but not a mutant version of this peptide (Figure 31B). In contrast, a GST-fusion protein with a mutated sequence (FLCVK → AACAA) did not bind CN (Figure 31B). Next, we expressed full length and truncated forms of A238L in yeast as GST fusions and measured their ability to inhibit expression of a CN-dependent reporter gene, CDRE-lacZ [3]. This analysis showed that a shorter A238L fragment, $A238L_{200-239}$, which includes

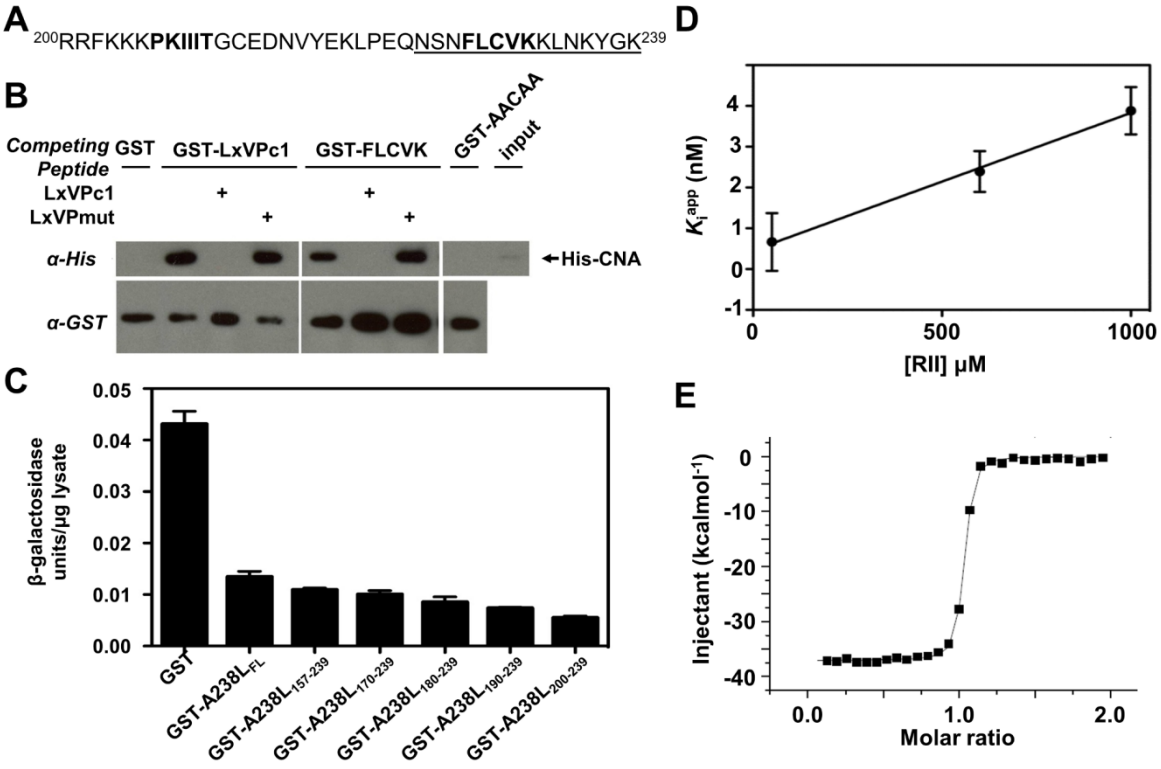


Figure 31. A238L interacts with CN via an LxVP and a PxIxIT motif.
A) C-terminal residues (200-239) of A238L showing putative docking site FLCVK (aa 228-232). Underlined residues were fused to GST. B) Recombinant CN was incubated with GST fused to 15 amino acids encoding the LxVP motif of NFATc1 or the FLCVK sequence in A238L. CN co-purifies with both motifs; this interaction is disrupted by incubation with excess peptide LxVPc1 encoding the LxVP motif from NFATc1, but not LxVP mut. CN fails to co-purify with GST fused to mutated FLCVK sequence (FLCVK mutated to AACAA). C) β-galactosidase activity of extracts from yeast strains that harbor 2xCDRE-lacZ, a CN-dependent reporter gene, and expression of GST or GST-A238L truncations are shown. 50 mM $CaCl_2$ was added to the cell culture 2.5 hrs before harvesting to induce CN-dependent activation of the Crz1 transcription factor ([3]; see also supplemental methods). Error bars indicate ± sd from three independent experiments. D) Secondary plot of K_i^{app} as a function of [RII] for $A238L_{200-239}$ inhibition of CN. Data show a linear dependence characteristic of competitive inhibition, with K_i = 0.37 nM. K_i^{app} values were obtained from the nonlinear fit of Figure S1B. Points represent averages ± sem. E) Isothermal titration calorimetry confirming that purified $A238L_{200-239}$ binds to CN.

both CN binding motifs (^{206}PKIIIT^{211} and ^{229}LCVK^{232}), is sufficient to inhibit CN (Figure 31C). Moreover, this same fragment competitively inhibits dephosphorylation of an RII phosphopeptide substrate by CN with a K_i of 0.37 nM (Figure 31D, Figure S1B; RII contains an LxVP motif) [104]. $A238L_{200-239}$ also forms a very tight complex with CN, with a dissociation constant (K_D) of 4 nM as determined using isothermal titration calorimetry (ITC) measurements (Figure

31E, Figure S2, Table 1). In agreement with this result, the CN-A238L complex is a stable trimer ($CN_{A1-370,B1-170}$-$A238L_{200-239}$; hereafter referred to as CN-A238L), as evidenced by the elution of the complex in a single peak during size exclusion chromatography at the expected elution volume for a 67.2 kDa trimeric complex. Taken together, these observations demonstrate that the competitive protein inhibitor $A238L_{200-239,}$ hereafter referred to as A238L, binds tightly to CN using both a PxIxIT-type anchoring sequence and an LxVP motif.

Table 1: Thermodynamic parameters and dissociation (K_D) and inhibition (K_i) constants for $CN_{A1-391/B1-170}$ with $A238L_{200-239}$
Results are shown for wild-type A238L, and A238L PxIxIT and LxVP mutants, derived from ITC experiments at 25°C or enzyme assays performed with RII at 37°C.

Complex	K_D (nM)	ΔH (kcal·mol^{-1})	-TΔS (kcal·mol^{-1})	ΔG (kcal·mol^{-1})	K_i (nM)
CN:$A238L_{WT}$	4±1	-40.3 ± 4.9	28.7 ± 4.9	-11.5 ± 0.2	0.37 ± 0.03
CN:$A238L_{PKIIITmut.}$	624 ± 26	-34.3 ± 0.9	25.9 ± 0.9	-8.5 ± 0.0	15 ± 1
CN:$A238L_{FLCVKmut.}$	803 ± 26	-14.0 ±0.1	5.7 ± 0.1	-8.3 ± 0.0	7700 ± 3000

Thermodynamic and dissociation constant data represent mean values ± one s.d. for triplicate measurements except $A238L_{WT}$, which was performed 5 times. Inhibition constants are mean values ± one s.e.m. from three independent experiments.

3.5.B. Structure of the CN-A238L heterodimer/inhibitor complex

To elucidate the molecular mechanism of A238L binding and inhibition, we determined the 1.7 Å crystal structure of the CN-A238L heterodimer/inhibitor complex (CNA/CNB/A238L). The structure of the CNA/B heterodimer in complex with A238L is virtually identical to CNA/B heterodimer structures from previous

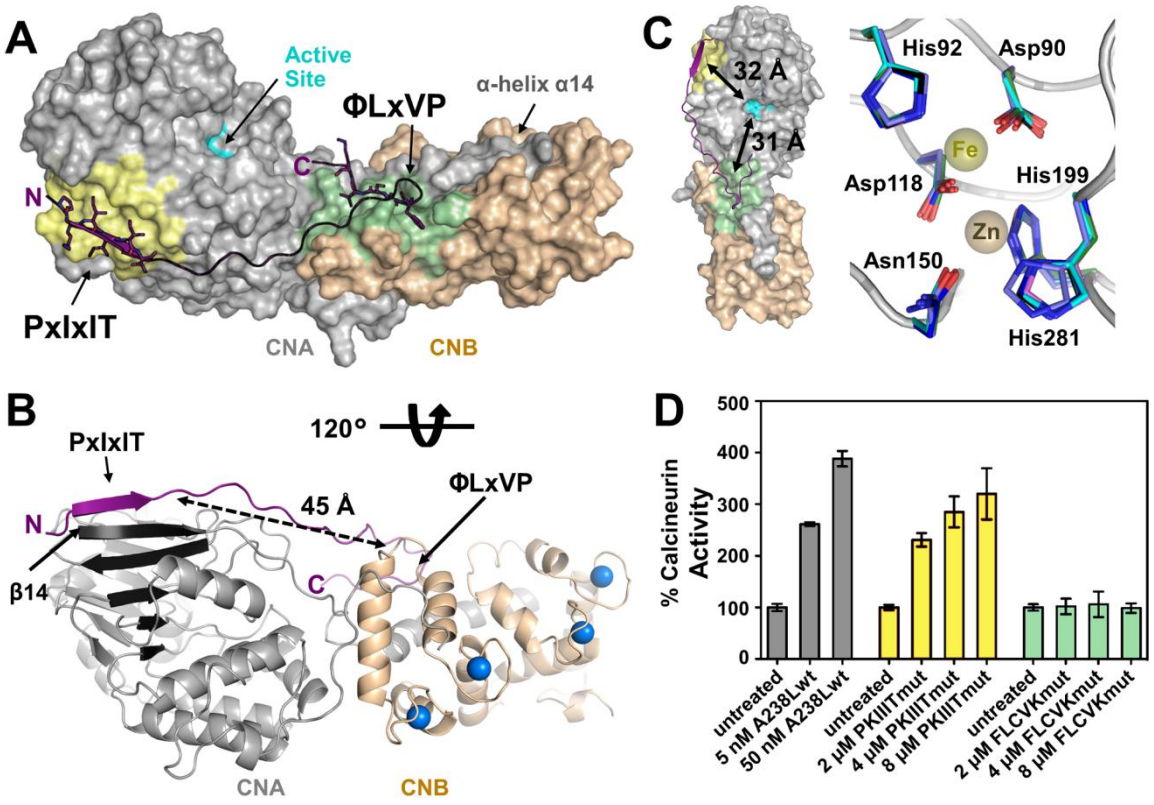


Figure 32 The CN active site is fully accessible in the CN-A238L complex.
A) Overview of the CN-A238L complex. CNA (gray, surface representation) interacts via helix α14 with CNB (beige, surface representation). The CNA active site is highlighted in cyan, with PxIxIT and LxVP binding pockets shown in yellow and green respectively. A238L is shown as a cartoon representation (purple), with the PxIxIT (PKIIIT) and LxVP (LCVK) substrate binding motifs highlighted as sticks. B) Cartoon representation of the CN-A238L complex. Structure is rotated 120° about the x-axis relative to A; all colors as in A; blue spheres representing four Ca^{2+} atoms in CNB. CNA β-sheet1 is shown in dark grey. The N-terminal β-strand formed by the A238L PxIxIT motif complements β-strand14 of CNA and thus extends β-sheet1. C) Left panel, CN-A238L illustrated as in A, with the distances between the CN active site and the PxIxIT and LxVP binding grooves indicated by arrows (~32 Å and ~31 Å, respectively). Right panel, overlap of the catalytic residues from the CN-A238L complex (black), apo-CN (cyan, PDBID 1AUI), CN-AKAPpeptide (dark blue, PDBID 3LL8), CN-PVIVITpeptide (green, PDBID 2P6B), CN-Cyclosporin (light blue, PDBID 1M63) and CN-FK506 (pink, PDBID 1TCO). Catalytic residues of CNA are shown as sticks and labeled (Asp90, His92, Asp118, Asn150, His199, His281). Fe^{3+} and Zn^{2+} ions from 1AUI are shown as spheres. D) The rate of pNPP hydrolysis was measured in the presence of A238L (grey), A238LPKIIITmut (yellow), or A238LFLCVKmut (green). Assays were performed in triplicate using 10 mM pNPP, and error bars indicate one s.d.

reports [1, 4, 10, 17, 18, 68, 111]. CNA residues 1-13, CNB residues 1-5/161-170 and A238L residues 200-204/235-239 were not visible in the electron density map and thus were not modeled. The absence of electron density for A238L residues 200-204 (N-terminal to the PxIxIT motif) and A238L residues 235-239

(C-terminal to the LxVP motif) suggests that these regions remain flexible upon complex formation and do not contribute to CN binding. A238L binds CN in a largely extended conformation, stretching from the PxIxIT binding site to the CNA/CNB interface and then looping back along the CNA/B interface to occupy the newly identified LxVP binding pocket (Figure 32A, B). As a consequence, the CN-A238L interaction buries 3084 $Å^2$ of solvent accessible surface area (SASA). Because A238L potently inhibits CN, it was predicted that A238L would bind and occlude the CN active site. However, the structure reveals that the CN active site is fully accessible in the CN-A238L complex (Figure 32C, left). Moreover, the key catalytic residues are structurally invariant when compared with those in previously determined CN structures [1, 4, 18], suggesting that CN in the CN-A238L complex is catalytically active (Figure 32C, right). The ability of CN to dephosphorylate small molecule substrates was confirmed using p-nitrophenyl phosphate (pNPP) dephosphorylation assays (Figure 32D). In fact, A238L and $A238L_{PKIIITmut}$ both increased the rate of pNPP hydrolysis compared to untreated CN, consistent with previously reported rate increases by both a peptide containing the LxVP sequence from NFATc1, as well as by the immunosuppressants FK506 and CSA [64, 71]. Thus, the CN-A238L complex retains full catalytic activity. Together, these data show that A238L does not inhibit CN by blocking the catalytic center, but instead utilizes an alternative mechanism.

3.5.C. The CN-A238L PxIxIT interaction

The interaction of A238L with CNA at the PxIxIT binding pocket buries 886 $Å^2$ of

SASA and is mediated by A238L residues 206-211 (PKIIIT, the PxIxIT motif from A238L), which form a short β-strand that hydrogen bonds with β14 of CNA to extend one of its central β-sheets (Figure 33A-C). As observed for other CN-$PxIxIT_{peptide}$ complexes, hydrophobic contacts are the dominant determinants of specificity in the PKIIIT interaction between A238L and CNA (Figure 33B). The PxIxIT interaction in the CN-A238L complex is nearly identical to those observed in the CN-$PVIVIT_{peptide}$ and the CN-$AKAP_{peptide}$ complexes [4, 10], demonstrating that PxIxIT sequences, contained in many CN-interacting proteins, likely bind CN in a similar manner (Figure 33C, D). The CN-A238L complex also allows for the comparison of interactions outside the PxIxIT binding groove (Figure 33C). As expected, these interactions are more variable. For example, while the C-terminal residues of both peptides angle down away from CNB, the A238L chain continues upwards towards the CNA/B interface. Furthermore, both $Pro13_{PVIVIT}$ and $Cys213_{A238L}$, but not $Thr345_{AKAP,}$ bind in a shallow hydrophobic pocket comprised of the methyl groups of $Lys318_{CNA}$, $Gln333_{CNA}$ and $Asn335_{CNA,}$ suggesting that this pocket might be engaged by other endogenous substrates to enhance binding (Figure 33C). These results illustrate how sequence variations not only in the PxIxIT sequence itself but also in flanking residues may fine-tune the affinity of PxIxIT sequences for CN [126].

3.5.D. The CN-A238L LxVP interaction

A short motif, LxVP, mediates interaction of several substrates, as well as A238L, with Ca^{2+}/calmodulin-activated CN [64]. However, until now, the molecular interactions that mediate LxVP recognition by CN have remained unknown

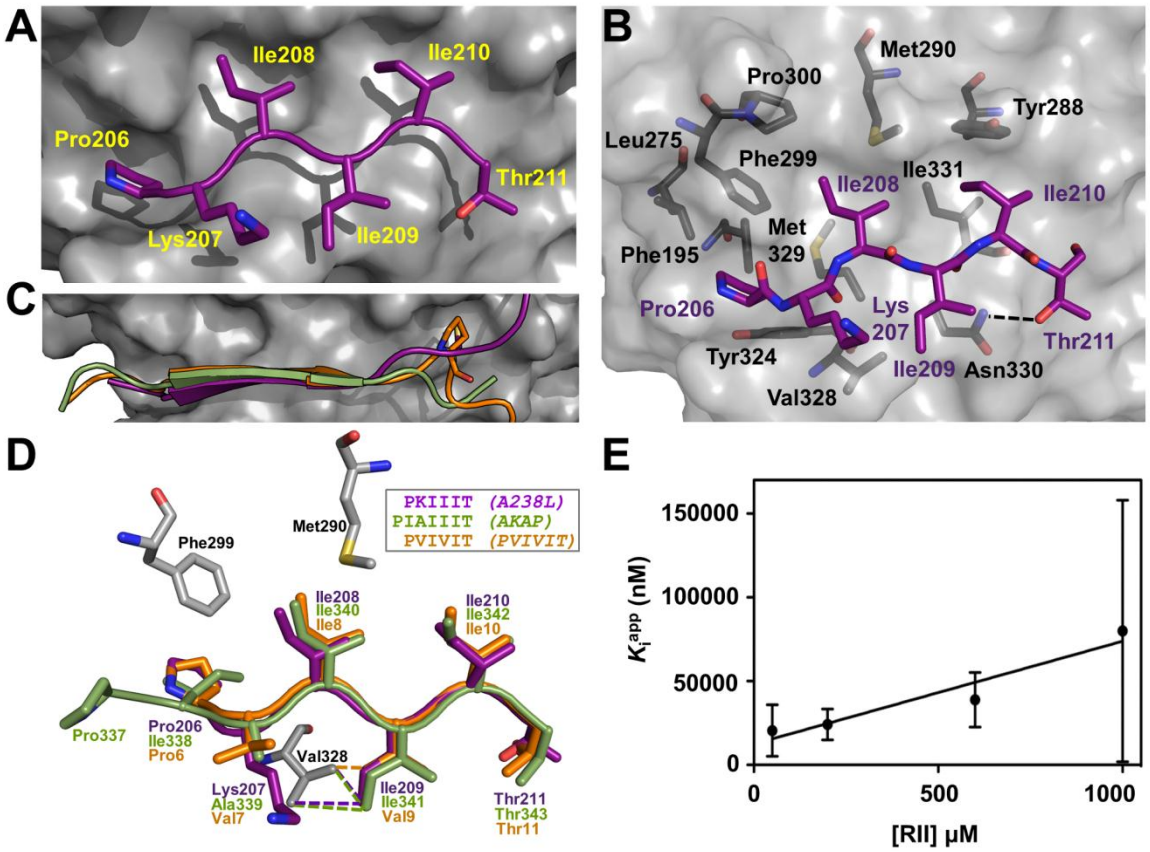


Figure 33. CN-A238L interactions: the PxIxIT substrate binding site.
A) Close-up view of the CN PxIxIT substrate binding site. A238L PKIIIT is shown as magenta sticks and labeled; CNA is shown as a grey surface. B) Same view as A, with a transparent CNA surface. Individual CNA residues that participate in the interaction with $PKIIIT_{A238L}$ are shown as grey sticks. C) Superposition of the $PKIIIT_{A238L}$ motif (purple) with a synthetic PVIVIT peptide (orange) and the IAIIIT CN docking site of AKAP79 (green) bound to CN. $Pro13_{PVIVIT}$ and $Cys213_{A238L}$ are shown as sticks. D) Overlay as in C but illustrated as sticks. Corresponding "variable" residues $Ile209_{A238L}$, $Val9_{PVIVIT}$ and $Ile341_{AKAP79}$ (the second "x" in PxIxIT) participate in the same hydrophobic interaction with $Val328_{CNA}$ (dotted lines). E) Secondary plot of K_i^{app} as a function of [RII] for inhibition of CN by $A238L_{FLCVKmut}$, which retains the PKIIIT site. Data show a linear dependence characteristic of competitive inhibition, with K_i = 7700 ± 3 nM. K_i^{app} values were obtained from the nonlinear fit of Figure S4A. Points represent averages ± sem.

(Figure 34). In the CN-A238L complex, the LxVP motif from A238L, ^{229}LCVK^{232}, is bound to CN (Figure 34A), revealing the very hydrophobic LxVP binding pocket at the CNA:CNB interface in CN (Figure 34B). When bound to CN, the A238L LCVK residues are extended, burying 728 Å^2 of SASA. Unlike the CN PxIxIT binding pocket, which is comprised of residues only from CNA, the CN LxVP binding pocket contains residues from both CNA and CNB. $Leu229_{A238L}$, which

becomes 92% buried upon complex formation, makes hydrophobic contacts with residues from both CNA and CNB subunits: $Trp352_{CNA}$ and $Phe356_{CNA}$, which, when mutated, alter binding to the LxVP sequence from NFATc1 [64], as well as $Leu115_{CNB}$, $Met118_{CNB}$, and $Val119_{CNB}$ (Figure 34C, left panel). Similarly, $Val231_{A238L}$ becomes 97% buried when bound to CN, forming hydrophobic interactions with $Tyr341_{CNA}$, $Leu343_{CNA}$, $Pro344_{CNA}$, $Trp352_{CNA}$ and $Leu123_{CNB}$ (Figure 34C, right panel). In addition, $Cys230_{A238L}$ (the "x" in LxVP) forms hydrogen bonds with $Trp352_{CNA}$ and $Asn122_{CNB}$ (Figure 34D, S3A). Finally, $Lys232_{A238L}$, which is a non-canonical residue in the LxVP motif as it has a "K" instead of the expected "P", does not interact with CN and thus is not important for CN recognition by A238L. Consequently, this work shows that residues $Leu229_{A238L}$ and $Val231_{A238L}$, which make significant interactions with both CNA and CNB, are the key residues that mediate binding to the LxVP interaction pocket. Notably, $Phe228_{A238L}$, which is immediately N-terminal to the LxVP binding motif, also contributes to CN binding (Figure 34A, B), fitting into a deep pocket formed by the loops connecting EF-hands 1 and 2 and EF-hands 3 and 4 of CNB. This results in a 25% increase in the SASA buried at this site (FLCVK buries 912 Å SASA). Critically, multiple LxVP sites are immediately preceded by an aromatic residue (Phe or Tyr), which may act as a binding strength enhancer, as mutation of this aromatic residue weakens the LxVP interaction [64]. Thus, a subset of LxVP sites, including the one in A238L, is best described as ΦLxVP.

3.5.E. A238L residues also contribute to the LxVP binding pocket

The ΦLxVP binding groove is comprised of residues from both CNA and CNB, in

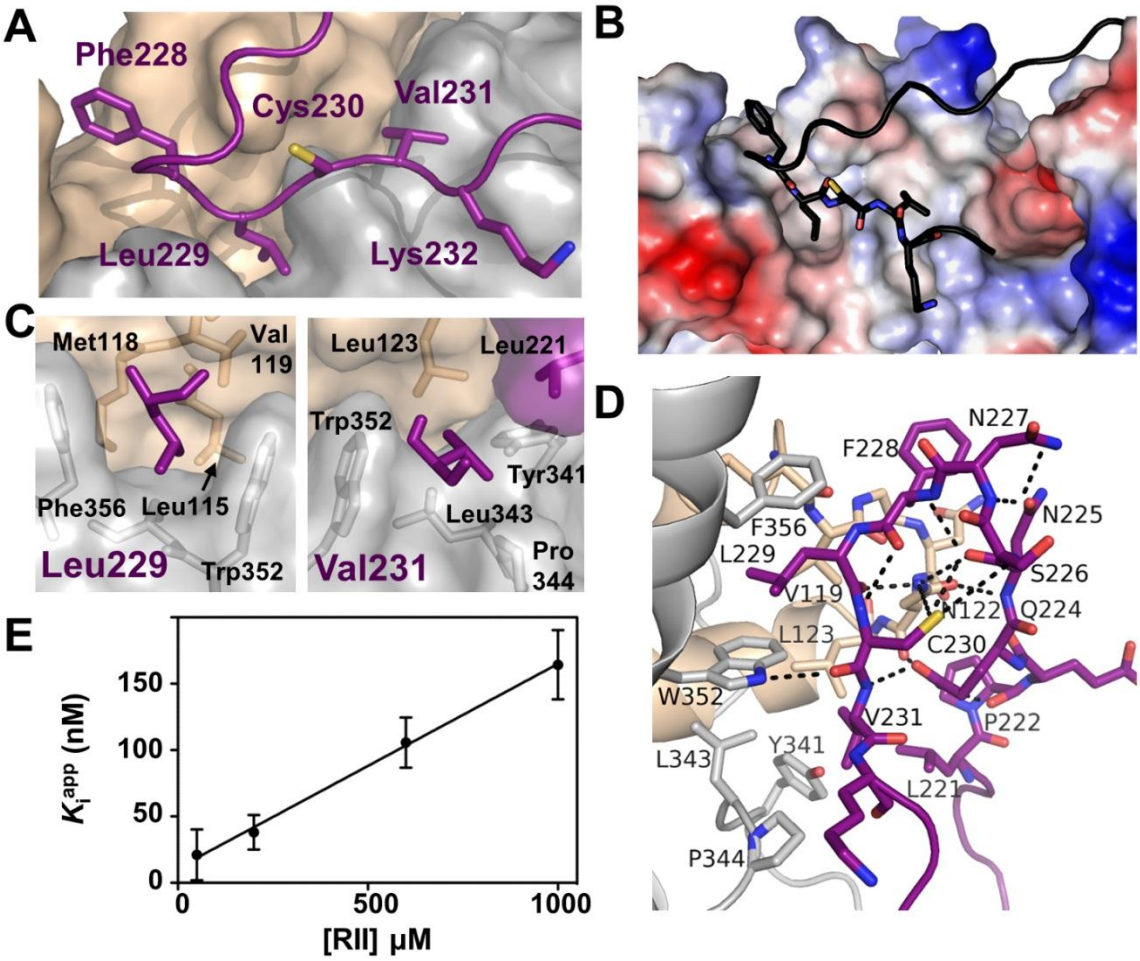


Figure 34. CN-A238L interactions: the LxVP substrate binding site.
(A) Close-up view of the CN ΦLxVP substrate binding site. (B) Electrostatic surface potential of the A238L-CN complex, highlighting the hydrophobic nature of the ΦLxVP binding groove. (C) CN residues that make up the Leu229 hydrophobic binding pocket (left) and the Val231 hydrophobic binding pocket (right). (D) Electrostatic intermolecular interactions between CN and A238L and intramolecular interactions at the A238L kink, which coordinate A238L for ΦLxVP binding. (E) Secondary plot of K_i^{app} as a function of [RII] for inhibition of CN by $A238L_{PKIIITmut}$, which retains the FLCVK site. Data show a linear dependence characteristic of competitive inhibition, with K_i = 15 nM. Points represent averages ± s.e.m. from three independent experiments.

the CN-A238L complex. Unexpectedly, however, a few A238L residues also contribute to the ΦLxVP binding pocket. Specifically, A238L, which projects upwards from the PxIxIT site towards the CNB interface, forms a tight 180° turn at $Asn227_{A238L}$ (Figure 34D, S3A). This kink is essential for ΦLxVP binding, as it redirects the ΦLxVP sequence (FLCVK) back toward its docking site on CN at

the CNA/B interface. This enables the A238L residues that immediately precede the ΦLxVP to interact directly with the CN-bound ΦLxVP residues and contribute to the ΦLxVP binding site. Thus, in addition to the multiple interactions observed with residues from CNA and CNB, $Val231_{A238L}$ (the 'V' in the LxVP motif) also makes intramolecular hydrophobic contacts with $Leu221_{A238L}$. In addition, the side chain $Gln224_{A238L}$ hydrogen bonds with the backbone amide of $Val231_{A238L}$ (Figure 34D, S3A). Therefore, residues from all three proteins—CNA, CNB and A238L—function to keep $Val231_{A238L}$ occluded from solvent in the bound conformation (92% buried). In addition, although $Cys230_{A238L}$ is the 'x' in the LxVP motif, this residue also makes multiple intramolecular interactions that help stabilize the A238L bound conformation. Specifically, the amide nitrogen of $Cys230_{A238L}$ hydrogen bonds with the carbonyl oxygen of $Phe228_{A238L}$. $Cys230_{A238L}$ also forms thiol hydrogen bonds with the amide nitrogen of $S226_{A238L}$ and the carbonyl of $N225_{A238L}$. These intramolecular interactions explain why $Cys230_{A238L}$ is still nearly completely buried (70%) in the CN-A238L complex even though its side chain points away from the LxVP docking groove. Thus, although it is not yet known how similar the CN- LxVP interaction of CN-A238L is with that of other LxVP docking motifs from substrates, our structure suggests that residues flanking the LxVP sequence may also modulate the affinity of this motif for CN.

Although the most extensive interactions between A238L and CN occur at the PxIxIT and ΦLxVP binding grooves, additional, largely polar intra- and intermolecular interactions outside of these docking sites also contribute to

A238L binding. For example, the interactions that stabilize the A238L kink (the 180° tight turn at $Asn227_{A238L}$, which enables the rest of A238L to point back towards LxVP binding pocket) are mediated by an extensive network of more than 10 hydrogen bonds, the center of which is $Asn122_{CNB}$ (Figure 34D, 3SA). The side chain amide nitrogen and carbonyl of $Asn122_{CNB}$ form hydrogen bonds with the backbone carbonyl and backbone amide nitrogen of two residues that border the kink, $L229_{A238L}$ and $Asn225_{A238L}$, respectively. Similarly, A238L residues C-terminal to the PxIxIT motif (^{212}GCEDNVY^{219}) also interact with CNA through main chain/side chain side chain/side chain hydrogen bonds (Figure S3B). Although the two dominant A238L:CN PxIxIT and LxVP interfaces together bury 1798 $Å^2$ SASA, these additional mostly polar interactions also contribute to the CN-A238L interface, burying 1286 $Å^2$ SASA.

3.5.F. A238L inhibits CN by interfering with substrate recognition

We next sought to determine the relative importance of the PxIxIT and LxVP sites for CN binding. Using ITC and two A238L mutants (a PxIxIT motif mutant, with PKIIIT mutated to AKAIAA and a ΦLxVP motif mutant, with FLCVK mutated to AACAA; $A238L_{PKIIITmut}$ and $A238L_{FLCVKmut}$, respectively), we found that both the PxIxIT and ΦLxVP sequences are important for CN binding, as the PxIxIT and ΦLxVP mutations increase the K_D similarly by ~150-fold and ~200-fold, respectively (Table 1 and Figure S2). Next, we investigated the contribution of the individual A238L motifs toward the inhibition of CN. $A238L_{PKIIITmut}$, which still contains the FLCVK sequence, competitively inhibits dephosphorylation of RII by CN with K_i = 15 nM (Table 1 and Figure 34E) [104]. In contrast, $A238L_{FLCVKmut}$,

which still contains the PKIIIT sequence, inhibited CN very poorly, with K_i = 7700 nM, a 20,000-fold decrease in inhibitor efficacy compared to wild-type A238L (Table 1, Figure 33E, S4). In addition, when expressed in HEK293T cells in which NFAT signaling has been stimulated, both A238L mutants reduced the activity of an NFAT-dependent reporter gene. At roughly equal levels of protein expression, neither $A238L_{FLCVKmut}$ nor $A238L_{PKIIITmut}$ were as effective as wt-A238L in NFAT inhibition (Figure 35A, B). Taken together, these results show that in A238L, FLCVK, but not PKIIIT, is required to competitively inhibit dephosphorylation of the RII phosphopeptide, whereas both sequences contribute to reducing CN-mediated dephosphorylation of substrates, such as NFATs, that contain an LxVP substrate recognition motif, and also require PxIxIT-mediated anchoring to CN. Therefore, A238L does not directly inhibit CN activity by blocking its

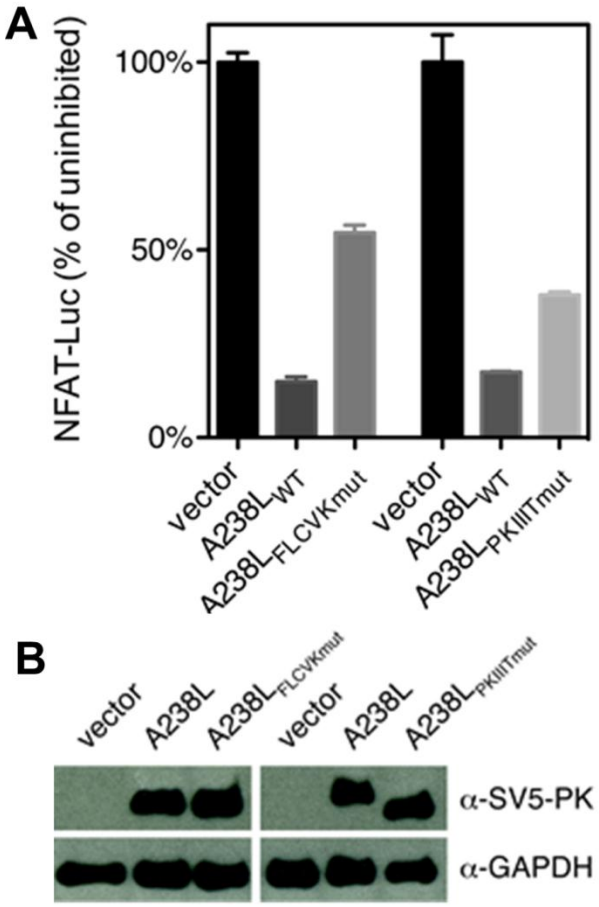


Figure 35: A238L inhibits CN by interfering with substrate docking.
A) NFAT-dependent transcription was measured in stimulated HEK293T cells co-transfected with an NFAT-luciferase reporter plasmid and a plasmid expressing SV5-PK-tagged A238L wild-type or motif mutant proteins, or vector alone. Data are means ± s.e.m. from three independent experiments. B) Transiently transfected SV5-PK-tagged A238L protein expression levels were analyzed by immunoblotting. The increased electrophoretic mobility of $A238L_{PKIIITmut}$ relative to wt-A238L may be due to changes in a posttranslational modification which is known to affect the electrophoretic mobility of A238L [2].

active site, but instead inhibits CN via a model of steric occlusion where it blocks the access of substrates to key binding sites on CN, similar to the steric occlusion of substrate selection previously reported for Protein Phosphatase 1 [5].

3.5.G. Model of CN-substrate dephosphorylation

The CN-A238L structure revealed the molecular interactions that mediate CN recognition of the LxVP motif, and we showed that the RII phosphopeptide requires this interaction for its dephosphorylation; therefore, we used bias-exchange metadynamics MD [122] to generate the first detailed model of a substrate, the RII phosphopeptide, bound to activated CN and poised for dephosphorylation (Figure 36A). The LxVP sequence of the RII peptide (^{81}D**L**D**VP**IPGRFDRRV***pS***VCAE^{99}) is separated from the phospho-Ser95 residue by nine amino acids, which, due to the distance between the LxVP-binding site and the catalytic center, limits the potential conformations this peptide can adopt during dephosphorylation. Our model of the CN:RII peptide complex shows that the interaction of RII (^{82}LDVP^{85}) with the CN LxVP site is very similar to that observed in the CN-A238L complex. It also shows that additional electrostatic and hydrophobic interactions stabilize interactions between CN and the intervening RII residues, which function to guide pSer95 to the CN active site. In particular, RII residues Asp91 and Arg92 form a number of key side chain hydrogen bonds and salt bridge interactions with CNA (Figure 36B). This is consistent with previous work suggesting that Arg in the -3 position is a strong positive determinant of dephosphorylation efficiency by CN [103]. The model also

shows that the nine intervening residues between the LxVP and phospho-Ser95 are essential for spanning the distance between the CN LxVP binding and active sites. This demonstrates that substrates with phosphorylated residues that are close to the LxVP site (e.g., within six or fewer residues) are unlikely to be dephosphorylated by CN as they will be unable to reach the active site.

3.6. Discussion

Our biochemical and structural studies reveal the mechanism by which A238L, a protein made by African swine fever virus that dampens the host immune response during infection, potently inhibits CN. The structure of CN (CNA_{1-370} and full-length CNB-subunit) in complex with A238L shows that 30 amino acids of A238L form a tight complex with CN. We conclude that an equivalent complex forms with full-length CN, as neither the presence of the AID nor additionally described inhibitory sequences [127] interfere with the inhibition of CN by A238L in vivo. As anticipated, A238L interacts with the PxIxIT binding groove on CNA. However, we discovered that A238L also binds CN in additional surface grooves, the most important of which is the LxVP-binding pocket. Remarkably, while A238L is a potent inhibitor of CN, the CN catalytic center is fully accessible and active, as confirmed by our biochemical findings that show that A238L rapidly dephosphorylates the small molecule substrate pNPP. In contrast, the CN-A238L complex is unable to dephosphorylate a peptide substrate derived from the RII subunit of protein kinase A, which contains an LxVP sequence, or a protein substrate, NFAT, which contains an LxVP sequence and a PxIxIT

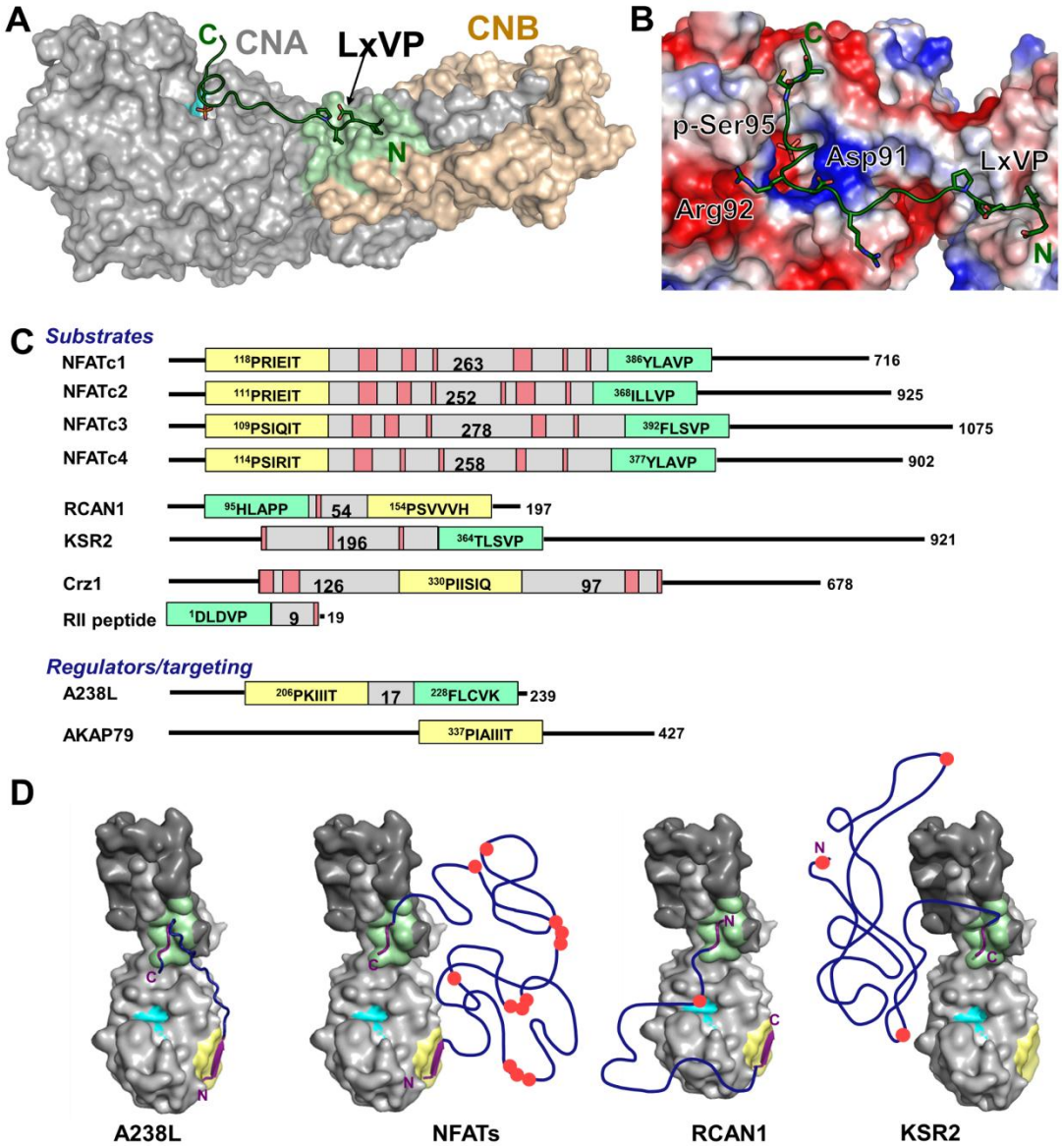


Figure 36: Potential interaction modes of CN substrates/regulators with CN.
(A) The CN-RII peptide complex obtained by MD. Colors as in Figure 31a,c. CN is shown in surface representation and the RII peptide in dark green with the LxVP motif (LDVP) and phospho-Ser95 as green sticks. LDVP is bound to the LxVP binding pocket (light green) and phospho-Ser95 is bound in the CN active site (cyan). (B) Electrostatic interactions between CN and the RII peptide. The CN electrostatic surface has positively and negatively charged areas colored blue and red, respectively. The LxVP motif and residues in RII that participate in polar interactions with CN are shown as green sticks. (C) Features of selected CN substrates and regulators, including substrates tested in this work (NFAT, Crz1 and the RII peptide). PxIxIT and LxVP motifs are highlighted in yellow and green, respectively with intervening residues in grey. Regions containing S-T residues that are dephosphorylated by CN are pink. (D) Potential modes of interaction of CN with various binding partners. CN is shown in grey, with the active site in cyan, the PxIxIT docking site in yellow and the LxVP docking site in green. CN binding partners are shown in blue, with PxIxIT and LxVP motifs in purple and phosphorylated regions shown as red circles. The residues between the two CN docking motifs, or between one docking motif and regions dephosphorylated by CN, are represented as coils, as they are predicted to be unstructured in solution. A238L is the CN-A238L crystal structure.

anchoring sequence. Collectively, these structural and biochemical data show that A238L inhibits CN function through a model of ‘steric inhibition’, in which A238L effectively inhibits CN function, not by directly blocking the active site of the phosphatase, but instead by occupying two critical docking grooves and sterically occluding CN from interacting with substrates. To our best knowledge, this mechanism of altered substrate dephosphorylation by steric occlusion of substrate binding sites has only been directly observed in one other Ser/Thr protein phosphatase, that of PP1, which, like CN, is one of the key members of the PPP family [5]. Thus our results establish that this mechanism, where by enzyme activity is modulated via substrate access rather than through active site inhibition or allostery, is likely utilized by the entire PPP family.

Although the structure of CN bound to a bona fide substrate has yet to be determined, the CN-A238L structure and our CN-RII substrate model significantly advance our understanding of how CN uses two complementary strategies to recognize substrates. First, the PxIxIT anchoring motif is used by protein scaffolds, inhibitors, and some substrates to form a stable interaction with CN by docking to a site that is available regardless of the activation state of the enzyme. For PxIxIT-containing substrates, the strength of this anchoring modulates the Ca^{2+}-concentration dependence of their dephosphorylation, but does not directly contribute to recognition of phosphosites during the dephosphorylation reaction [126, 128]. Second, CN overcomes the limited specificity of its catalytic site by recognizing specific residues, such as ‘ΦLxVP’, which are distal to the phosphosite. This interaction is likely required for the dephosphorylation of

multiple substrates, as it is this site that is targeted by multiple inhibitors, including A238L and the immunosuppressants CSA and FK506 (Figure 36C).

This work also reveals interactions that contribute to phosphosite selection by CN. The 'ΦLxVP' binding site is now molecularly defined as a hydrophobic binding surface composed of residues from the CNA and CNB subunits which becomes accessible after Ca^{2+}/Calmodulin binding displaces the C-terminal AID domain from the active site [64]. Our structure and the RII model show that the substrate residue dephosphorylated by CN (hereafter referred to as pS/pT) must be a minimum of 9-15 residues away from the ΦLxVP sequence, and be in an extended conformation in order for the pS/pT residue to reach the catalytic site. In fact, the pS/pT residues in the RII peptide and the substrate RCAN1 are only 9 and 10 residues C-terminal to the 'P' in their LxVP motifs, respectively (Figure 36A, B, D) [104, 129]. Our model also shows that additional hydrostatic interactions, especially at basic residue at position -3 helps orient the phosphosite in RII toward the catalytic center. In addition, the C-terminus of the RII peptide lies in a groove identified in other PPP's, namely PP1, to be a substrate recognition groove (hydrophobic groove in PP1), suggesting that this groove functions in a similar manner in CN and thus likely for the entire family of PPPs [39, 40].

While our model provides fundamental insights into phosphosite selection by CN for those substrates in which the phosphosite is 9-10 residues C-terminal to LxVP sequence, other surface grooves near the catalytic center are likely also important for substrate recognition. This is because the spacing between

experimentally determined ‘LxVP’ and phosphosites substrates is quite variable. For example, substrates such as NFAT and KSR2 have pS/pT sites that are 10s-100s of residues away from the known LxVP sequences [130-134]. In these cases, one or several pS/pT sites are found in long, extended sequences that are predicted to be unstructured, suggesting that these regions are dynamic when bound to CN and that interactions at the LxVP and PxIxIT sites tether the substrate to CN allowing the phosphosite(s) to encounter the catalytic site and be dephosphorylated (Figure 36D). A detailed understanding of how phosphosites that are distal from the LxVP and PxIxIT motifs engage the CN catalytic center is an active area of investigation.

Finally, these studies also provide critical new insights into the mechanism by which the immunosuppressants CSA and FK506 inhibit CN. Our structure reveals that the location of the LxVP substrate-binding site on CN is identical to that of the CSA/FK506 binding site (Figure 37A, B). Furthermore, we identified several contacts common to all of these interactions. In particular, multiple residues from the cyclic immunosuppressant molecules overlap nearly perfectly with key residues of the A238L LxVP sequence, LCVK, especially at $Leu229_{A238L}$ and $Val231_{A238L}$, and occupy hydrophobic surfaces formed by Trp352 and Phe356 of CNA and Met118 and Val119 of CNB ((Figure 37C). These findings indicate that disrupting the interaction between CN and LxVP motifs found within substrates is sufficient for inhibiting the dephosphorylation of LxVP-containing substrates, and explains the molecular mode of action of these drugs. Furthermore, the ability of FK506 and CSA to antagonize all known functions of

CN suggests that every substrate contains at least one LxVP motif, and that interaction of this motif with CN is essential for substrate dephosphorylation.

Consequently, our high-resolution 3-dimensional structure of the CN-A238L complex opens up a new avenue for the development of a novel class of powerful CN inhibitors that selectively and potently bind the CN LxVP binding site. Both CN substrate-binding pockets, but especially the LxVP pocket, present excellent targets for the development of immunosuppressive drugs. Clearly, such drugs will also be an extremely powerful tool for investigating the many regulatory processes, including immune activation and cardiac hypertrophy, that are driven by CN.

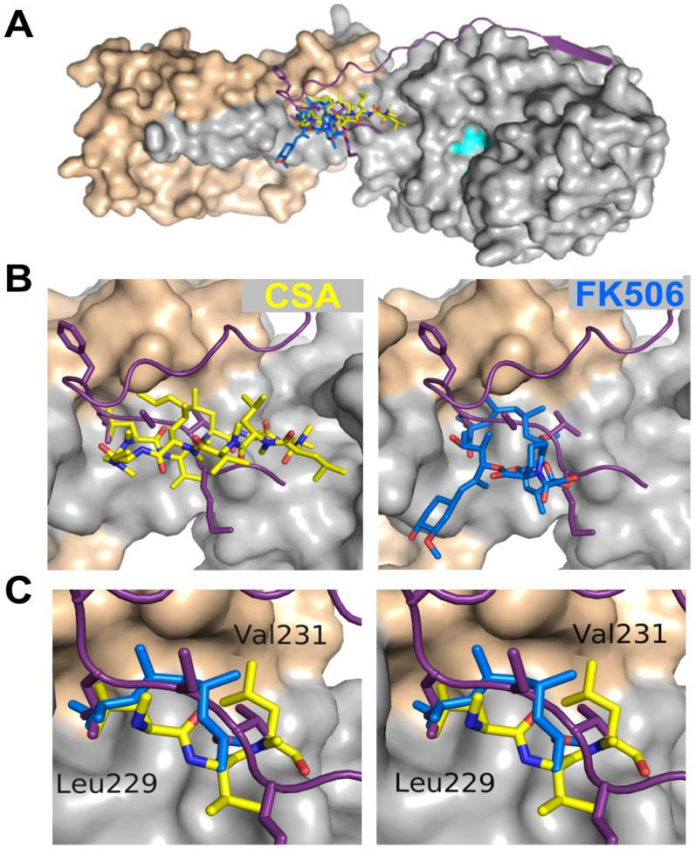


Figure 37: The immunosuppressive drugs FK506 and CSA inhibit CN by occupying the LxVP substrate recognition site.
(A) Overlay of CN-bound A238L (purple), CSA (yellow), and FK506 (blue). (B, left) Overlay of A238L with CSA at the LxVP substrate binding groove. (B, right) Overlay of A238L with FK506. (C) Stereoview of the Leu229 and Val231 binding pockets in the LxVP substrate binding groove with A238L, CSA, and FK506. Some atoms of CSA and FK506 removed for clarity

Chapter 4. Regulation of calcineurin by RCAN1

4.1. Introduction

Regulator of Calcineurin 1 (RCAN1) is an endogenous protein that regulates the activity of calcineurin (CN) against many of its substrates. The human RCAN family consists of three members: RCAN1, RCAN2 and RCAN3 [135]. RCAN1-3 are highly conserved in animals, and an RCAN homolog is also found in invertebrates and fungi [136]. The function of RCAN1 as a CN inhibitor has been well-established through biochemical and *in vivo* studies, which showed that RCAN1 directly inhibits CN phosphatase activity *in vitro*, and that it interferes with the dephosphorylation of the NFAT family of transcription regulators by CN and their nuclear translocation [75, 129, 137-141]. However, there is evidence that RCAN1 can also function as an enhancer of CN activity [136, 142, 143]. These studies show that disruption of RCAN1 expression in mice and yeast decreased NFAT nuclear localization and the transcription of genes under NFAT control. Current research efforts aim to understand how RCAN1 can act as both an inhibitor and enhancer of CN activity. Significant progress has been made toward elucidating the mechanism of CN inhibition by RCAN1, though its mode of function as an enhancer of CN activity is poorly understood.

RCAN1 is a 197 amino acid long protein [77] (Figure 38). The N-terminal domain of RCAN1 (residues 1-88) is a structured RNA Recognition Motif (RRM) that has not been implicated in CN regulation. The C-terminal half of RCAN1 (residues 89-197) contains several conserved CN-binding motifs, including an LxVP motif (^{96}LAPP^{99}) and a PxIxIT site (^{154}PSVVVH^{159}), which are used by CN substrates and regulators for CN docking [55-65, 144]. The LxVP motif of

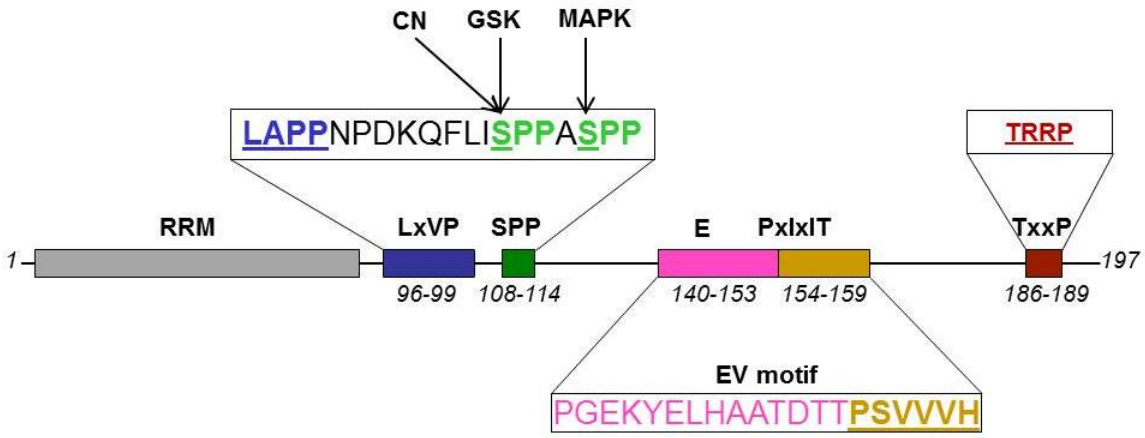


Figure 38: Domain architecture of RCAN1 and CN-interacting regions.
Regions of RCAN1 experimentally shown to interact with CN are highlighted. They include the LxVP sequence (blue), the SP motif (green), the E motif (pink, named for the ELHA sequence it contains), the PxIxIT sequence (orange), and the TxxP motif (red). The RR domain, which does not interact with CN, is also shown (grey). Ser112 phosphorylation primes Ser108 for phosphorylation by GSK-3. Ser108 is dephosphorylated by CN.

RCAN1, located immediately C-terminal to the RRM, lacks the conserved "V" residue ("V" is replaced by a "P" to give LAPP) and is immediately preceded by a His residue (^{95}HLAPP^{99}). The RCAN1 PxIxIT motif, ^{154}PSVVVH^{159}, deviates significantly from the canonical consensus sequence (only the "P" matches the consensus) but is still able to bind to the PxIxIT-binding pocket of CN with specificity [140]. Immediately C-terminal to the PxIxIT is a 14 amino acid sequence called the E-motif (residues 140-153), named for the ^{145}ELHA^{148} sequence it contains. The 20 amino acid region that encodes the E-motif and the V-rich PxIxIT (residues 140-159) is called the EV-motif and is highly conserved among all three RCANs and across eukaryotes. The EV motif is necessary and sufficient for binding to CN and inhibiting its activity against NFAT substrates *in vivo* by preventing substrate docking at the PxIxIT binding pocket of CN. RCAN1 also contains an SP phosphorylation motif (^{108}SPPASPP^{114}) located seven amino acids C-terminal to the LxVP site. Ser112 in the SP motif can be phosphorylated

by MAPK [129], BMK1 [145], NIK [146], and DYRK1 [147] (Figure 38 and Figure 39). This primes Ser108 for phosphorylation by GSK-3 [129]. Ser108 is the only RCAN1 residue known so far to be dephosphorylated by CN. The SP phosphorylation state could play a role in determining whether RCAN1 functions as an inhibitor or activator of CN activity [129]. Finally, RCAN1 contains a C-terminal conserved TxxP motif (^{186}TRRP^{189}) (Figure 38) that is necessary for RCAN1 inhibition of CN phosphatase activity *in vitro*, against the small nonspecific phospho-Tyr mimic pNPP. The TxxP motif is therefore believed to occupy the CN active site. The importance of the TxxP is not yet clear since the EV alone can prevent CN dephosphorylation of NFAT without the need to occupy the CN active site, but one possibility is that the TRRP functions to increase the inhibitory effect of RCAN1 on CN against NFAT and other substrates [140, 148].

Though the dual function of RCAN1 as an inhibitor and enhancer of CN activity is still somewhat debated, it is generally agreed that the phosphorylation state of RCAN1 plays a role in determining whether RCAN1 inhibits or enhances CN activity [143, 145, 149-151]. In addition to the phosphorylation events at Ser112 and Ser108 mentioned above, RCAN1 is also phosphorylated at Ser94 and Ser136 by TAK1 [152], and at Thr192 by DYRK1 [147] (Figure 39). Thr192 phosphorylation was shown to enhance the inhibitory effect of RCAN1 on CN phosphatase activity against pNPP, while phosphorylation of Ser94, Ser136, Ser112, and Ser108 have been shown to turn RCAN1 into an activator of CN. It has also been suggested that the phosphorylation state of RCAN1 influences its cellular stability and localization, which also determine its effect on CN function.

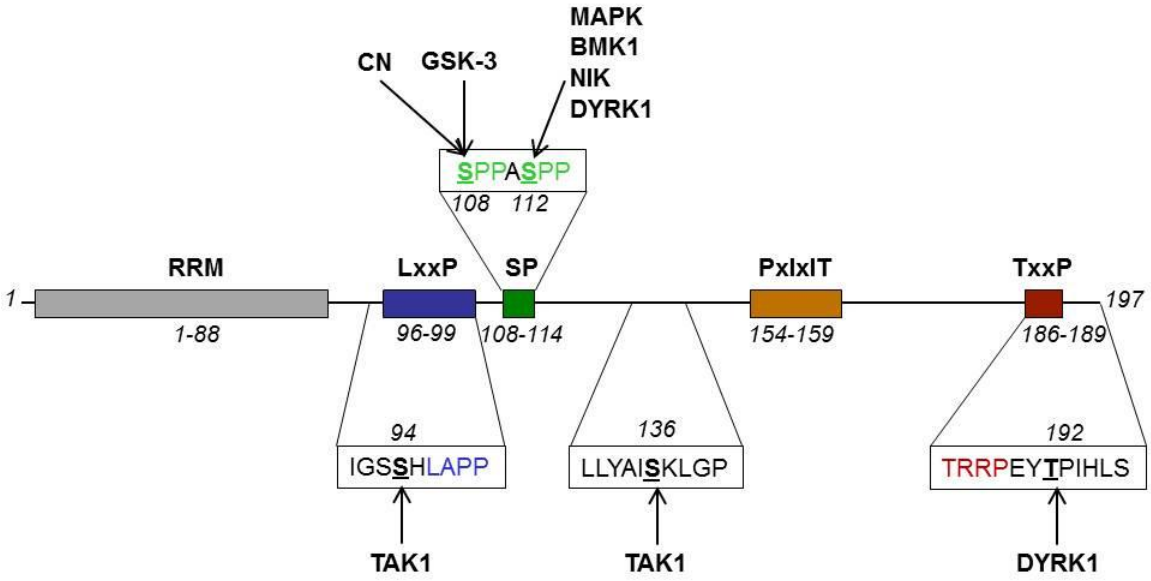


Figure 39: Phosphorylated residues on RCAN1 and their kinases.
All known phosphorylation events on RCAN1 are shown. Phosphorylation of Ser192 enhances the ability of RCAN1 to inhibit CN phosphatase activity. Phosphorylation of Ser108, Ser112, Ser94, and Ser136 have been reported to change the function of RCAN1 from an inhibitor to an activator of CN.

The gene encoding RCAN1 was initially discovered as part of a group of genes encoded in the Down Syndrome Critical Region of human chromosome 21 [153]. Down syndrome is caused by trisomy 21 (the presence of three copies of chromosome 21), and is largely characterized by developmental defects. Overexpression of RCAN1 in Down syndrome individuals is believed to interfere with the function of the NFAT family of transcription factors during development. Cytosolic NFATs must be dephosphorylated by CN in order to translocate to the nucleus, where they can activate the transcription of target genes. RCAN1 inhibits CN dephosphorylation of NFATs, and prevents their nuclear translocation. RCAN1 is believed to act synergistically with the kinase DYRK1 (also encoded on chromosome 21 and overexpressed) to keep cellular NFATs phosphorylated and inactive in Down syndrome patients [154].

RCAN1 has received much attention in recent years, as it has become increasingly clear that it is a key regulator of many CN functions. Because RCAN1 indirectly affects the activity of CN against many important substrates, including the NFAT transcription factors, it is a promising drug target. However, much remains unclear about the RCAN1-CN interaction, and especially about the factors that control whether RCAN1 functions as an enhancer or repressor of CN activity.

Here we provide the detailed biophysical characterization of the RCAN1-CN complex using NMR and ITC data. Using NMR, we show that RCAN1 is an intrinsically disordered protein (IDP) with regions of transient secondary structure. We use NMR to define the exact regions of RCAN1 that participate in CN binding, and determine the relative binding strengths of each CN binding region by ITC. Combined, these results offer novel insight into the mechanism used by RCAN1 to bind and regulate CN, and provide valuable information that will assist in the crystallization of the RCAN1-CN complex.

4.2. Methods

4.2.A. Expression and Purification of RCAN1 constructs

Full-length human RCAN1 (residues 1-197), the RCAN1 C-terminal CN binding domain (residues 89–197), and shorter constructs from the CN binding domain (89-180, 89-164, 151-192, 128-192, 128-164, 179-192) were subcloned into the pET-based bacterial expression vectors THMT-MBP (Cornell Protein Facility) or RP1B [155] using NdeI/XhoI restriction sites. All constructs expressed in THMT-MBP encode an N-terminal His_6 purification tag followed by a solubility-

enhancing MBP (maltose binding protein) tag and a TEV protease cleavage site. Constructs were also cloned into expressed in RP1B are identical except for the missing MBP tag.

All RCAN1 constructs were expressed using the following protocol: the expression plasmid was transformed into BL21-(DE3) RIL *E. coli* cells (Agilent) and cultures were grown in LB containing chloramphenicol (34 µg/ml) and kanamycin (50 µg/ml) to an optical density (OD_{600}) of 0.7 at 37 °C with shaking at 250 rpm. Protein expression was induced by the addition of 1 mM isopropylthio-β-D-galactoside (IPTG) and the cultures were grown for an additional 18-20 hours at 18 °C with shaking at 250 rpm. To express uniformly $^{13}C/^{15}N$-labeled and ^{15}N-labeled $RCAN1_{89-197}$, bacterial cultures were grown in M9 minimal medium containing selective antibiotics and 4 g/l [^{13}C]-D-glucose and/or 1 g/l $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) as the sole carbon and nitrogen sources, respectively. For all samples, cells were harvested by centrifugation (6,000 *xg*, 15 min, 4 °C) and stored at -80 °C.

For all RCAN1 purifications, *E. coli* pellets were resuspended in ice-cold lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100) containing EDTA-free protease inhibitor tablets (Roche), lysed by high-pressure cell homogenization (Avestin C3 Emulsiflex), and the bacterial lysate was clarified by centrifugation (45,447 *xg*, 45 min, 4 °C). Filtered supernatant was loaded onto Ni-NTA resin (Qiagen) equilibrated in His-tag buffer A (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole), the resin was resuspended, and the column was rocked at 4 °C for 1 h. Flow-through was collected on ice, the

column was thoroughly washed with His-tag buffer A, and the His_6-MBP-tagged protein was eluted with His-tag buffer B (50 mM Tris pH 8.0, 500 mM NaCl, 400 mM imidazole). TEV (His_6-tagged; produced in-house) was added to the eluted RCAN1 and the sample was dialyzed against 20 mM Tris pH 8, 500 mM NaCl at 4 °C for 16 h.

The enzymatically-cleaved His_6-MBP-tag and TEV protease were removed by heat purification of the dialysate (85 °C, 20 min) (this step was not performed for the RCAN1 full-length construct). Remaining His_6-MBP was removed by an additional Ni^{2+}-affinity purification and RCAN1 was further purified by size exclusion chromatography (SEC; Superdex 75 26/60 [GE Healthcare]) equilibrated in RCAN1 buffer (20 mM Tris pH 7.5, 500 mM NaCl, 0.5 mM TCEP) for complex formation with Calcineurin, in NMR buffer (20 mM sodium phosphate pH 6.2, 50 mM NaCl, 0.5 mM TCEP), or in ITC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 1.5 mM $CaCl_2$). Fractions containing RCAN1, as determined by SDS-PAGE, were pooled and RCAN1 was concentrated up to 1 mM using Amicon Ultra concentrators (3 kDa MW cutoff, Millipore). All SEC-purified RCAN1 constructs (except full-length RCAN1) were heat purified one more time to remove any remaining MBP and to denature trace proteases that may have been introduced while handling the sample. Purified RCAN1 was immediately used for ITC and NMR experiments. RCAN1 constructs used for complex formation with CN were either used immediately or frozen in liquid nitrogen and stored at -80 °C until needed. The 14-mer peptide $RCAN1_{179\text{-}192}$ was synthesized at CS Bio Co. (Menlo Park, CA).

4.2.B. Expression and Purification of calcineurin (CN)

The A (CNA, residues 1-370 or 1-391) and B (CNB, residues 1-170) subunits of human CN were expressed and purified as described previously [144].

4.2.C. CN-RCAN1 complex purification

Complexes between $CN_{A1-370,B1-170}$ and RCAN1 constructs (full-length, 89-197, 89-180, 89-164, 151-192, 128-192, 128-164, 179-192) were formed as previously described for CN-A238L [144]. The complex used for NMR contained ^{15}N-labeled $RCAN1_{89-197}$ and unlabeled $CN_{A1-370,B1-170}$ in NMR buffer (20 mM HEPES pH 6.7, 150 mM NaCl, 0.5 mM TCEP, 1.5 mM $CaCl_2$).

4.2.D. NMR spectroscopy

All NMR measurements were performed at 298 K. $RCAN1_{89-197}$ and $RCAN1_{89-197}$-$CN_{A1-370,\ B1-170}$ NMR samples were prepared in the NMR buffers with 10% (v/v) D_2O. The sequence-specific backbone assignment of $RCAN1_{89-197}$ was obtained from the following experiments performed on Bruker Avance 500 and Avance IIIHD 850 MHz spectrometers equipped with TCI HCN *Z*-gradient cryoprobes: 2D [^{1}H,^{15}N] HSQC, 3D ^{15}N-resolved [^{1}H,^{1}H] NOESY (τ_m= 120 ms), 3D HNCA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HN(CA)CO, 3D (H)CC(CO)NH (τm= 12 ms), and 3D selective I-, LA-, QN-, E-, S-, P-, FHY-, TAVI-, W-, A- and TA-HSQC ([^{1}H,^{15}N] plane). The NMR spectra were processed with Topspin 3.1 (Bruker, Billerica, MA) and analyzed using the CARA software package (http://www.cara.nmr.ch). ^{15}N [^{1}H]-NOE (hetNOE) measurements were acquired using sensitivity enhanced pulse sequences. ^{15}N [^{1}H]-NOEs were measured from

a pair of spectra acquired with and without presaturation recorded in an interleaved manner. A recycle delay of 5 s between scans was used for the heteronuclear NOE experiments. ^{15}N [^{1}H]-NOEs were calculated using the Bruker Dynamics Center 2.0 software. Secondary Structure Propensity (SSP) calculations using the $RCAN_{189-197}$ Cα and Cβ chemical shifts were performed [156] using the random coil database RefDB [157].

4.2.E. Isothermal Titration Calorimetry

ITC experiments were performed at 25 °C using a VP-ITC microcalorimeter (GE Healthcare). All protein samples were equilibrated in ITC buffer. RCAN1 was titrated into $CN_{A1-391/B1-170}$. $RCAN1_{179-192}$ was titrated into both CN alone and into previously formed CN/$RCAN1_{128-164}$ complex. Titrant (10 µL per injection) was injected into the sample cell over a period of 20 seconds with a 250 second interval between titrations to allow for complete equilibration and baseline recovery. 28 injections were delivered during each experiment, and the solution in the sample cell was stirred at 307 rpm to ensure rapid mixing. All ITC experiments were performed at least in triplicate, with the exception of $RCAN1_{179-192}$ titrated into CN (performed in duplicate) and $RCAN1_{179-192}$ (single experiment). “Global” K_d values were approximated in Origin 7.0 (OriginLab) using a ‘one set of sites’ binding model, based on the 1:1 stoichiometry observed by SEC.

4.3. Results

4.3.A. The RCAN1 CN binding domain is intrinsically disordered

His_6-tagged RCAN1 constructs showed high overexpression, but were insoluble. In contrast, all His_6-MBP-tagged RCAN1 constructs showed soluble expression in *E. coli*. Bioinformatics analysis using IUPred [158] predicts the CN binding region of RCAN1 (amino acids 89-197) to be intrinsically disordered, based on the presence of disorder inducing amino acids in the primary sequence. $RCAN1_{89-197}$ is heat stable and remains soluble at temperatures up to 95 °C, which is consistent with the prediction that $RCAN1_{89-197}$ is an IDP. In addition, the SEC elution profile of $RCAN1_{89-197}$ reveals an apparent hydrodynamic radius larger than expected for a globular protein the size of $RCAN1_{89-}$

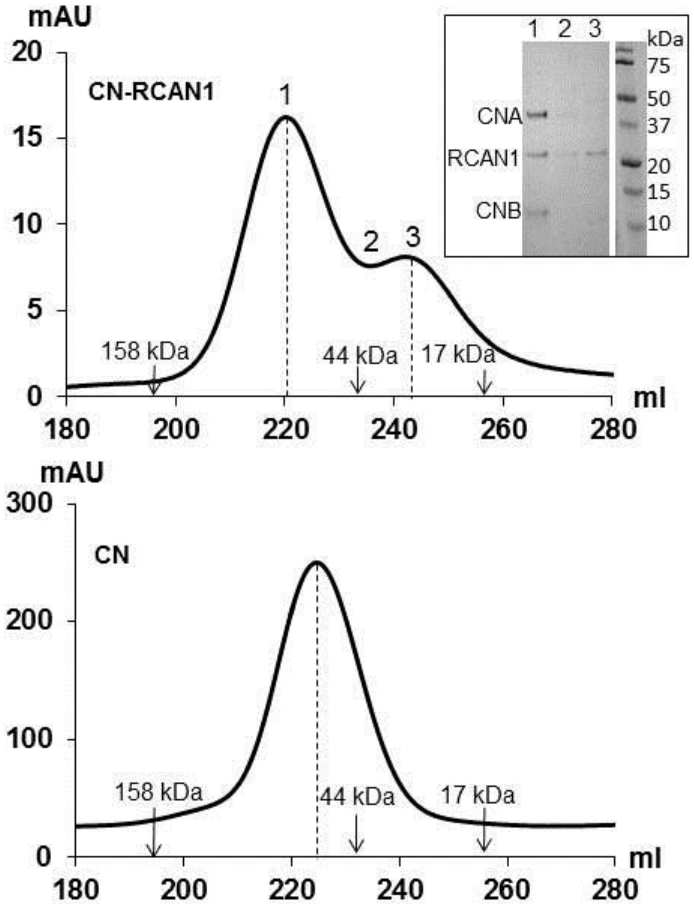


Figure 40: Size exclusion chromatography of RCAN1 and CN.
A) SEC of the $CN_{A1-370,B1-170}$:$RCAN1_{full-length}$ complex (excess RCAN1 was used to form the complex). The first peak at 220 ml contains CN:RCAN1(85 kDa) and has an elution volume consistent with 1:1 stoichiometry. The second peak at 241 ml is excess RCAN1 (23 kDa), and its elution volume shows it is monomeric. The SDS-PAGE inset confirms the contents of each peak. B) SEC of $CN_{A1-391,B1-170}$. CN (64.1) elutes at 224 ml in the absence of RCAN1.

$_{197}$ (12.3 kDa), also a property characteristic of IDPs. SEC elution profiles obtained for the various RCAN1 constructs show that RCAN1 is a monomer in solution, and that the RCAN1-CN complex has 1:1 stoichiometry (Figure 40)

4.3.B. NMR analysis of RCAN1

NMR analysis of RCAN1$_{89-197}$ further confirmed that it is an IDP: the 2D [^{1}H,^{15}N] HSQC spectrum of RCAN1$_{89-197}$ (Figure 41) has the narrow chemical shift dispersion in the ^{1}H dimension that is characteristic of IDPs. The RCAN1$_{89-197}$ sequence is 109 amino acids in length and contains 14 Pro residues, making the number of expected peaks for RCAN1$_{89-197}$ 95 (109 total residues – 14 Pro).

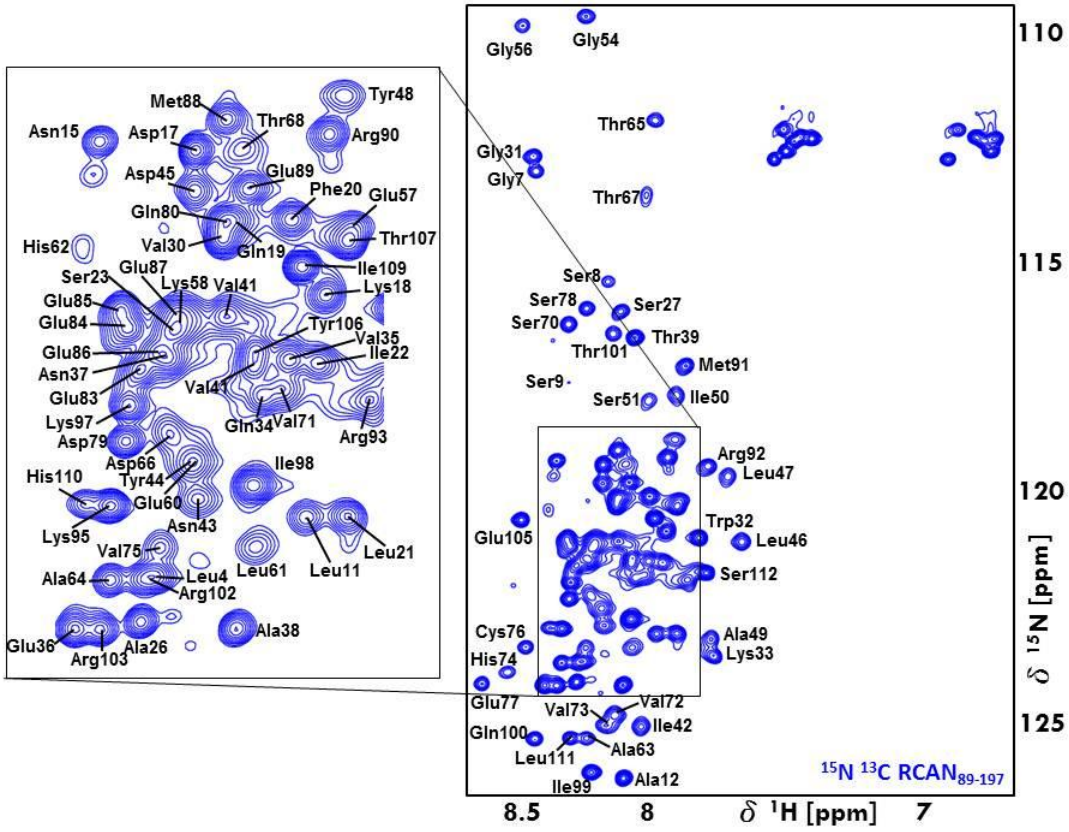


Figure 41: 2D [^{1}H,^{15}N] HSQC spectrum of RCAN1$_{89-197}$.
The sequence-specific backbone assignment is shown. Peaks in the upper right of the spectrum correspond to sidechain N-H pairs. The narrow ~1 ppm chemical shift dispersion in the ^{1}H dimension confirms that RCAN1$_{89-197}$ is an IDP.

Sequence-specific backbone assignments were obtained for 85 of these 95 residues (89% completion). At least 20 additional peaks can be detected in the $RCAN1_{89-197}$ 2D [^{1}H,^{15}N] HSQC spectrum, indicating slow conformational exchange, possibly due Pro cis/trans isomerization which is often seen in IDPs. Conversely, 10 residues in the $RCAN1_{89-197}$ sequence were not assigned (His90, His95, Ala134, Lys137, Leu138, Lys143, Tyr144, Glu166, and Lys167), either due to overlap or line-width broadening beyond detection caused by intermediate conformational exchange.

We calculated the secondary structure propensity (SSP) of $RCAN1_{89-197}$ (Figure 42A) in order to identify any transient secondary structure elements, which may be important for the function of the protein. Based on this analysis, $RCAN1_{89-197}$ is mostly disordered with the exception of two α-helical regions, α1 and α2 (Figure 42A), which are ~40 and ~50% populated, respectively. Interestingly, helices α1 and α2 form outside any of the canonical CN recognition motifs, and α1 is overall highly hydrophobic (amino acid sequence ^{129}YDLLYAISKL^{138}) while α2 is overall highly positively charged (amino acid sequence ^{168}EEEEEMERMR^{177}). The N- and C-termini of $RCAN1_{89-197}$ are disordered, as is the EV-motif. Importantly, the β-strand that is characteristic of PxIxIT motifs, as seen in several PxIxIT-CN complex structures [4, 10, 68, 144], is absent in unbound RCAN1. Heteronuclear Nuclear Overhauser Effect (hetNOE) analysis of $RCAN1_{89-197}$ shows, in agreement with SSP data, that most

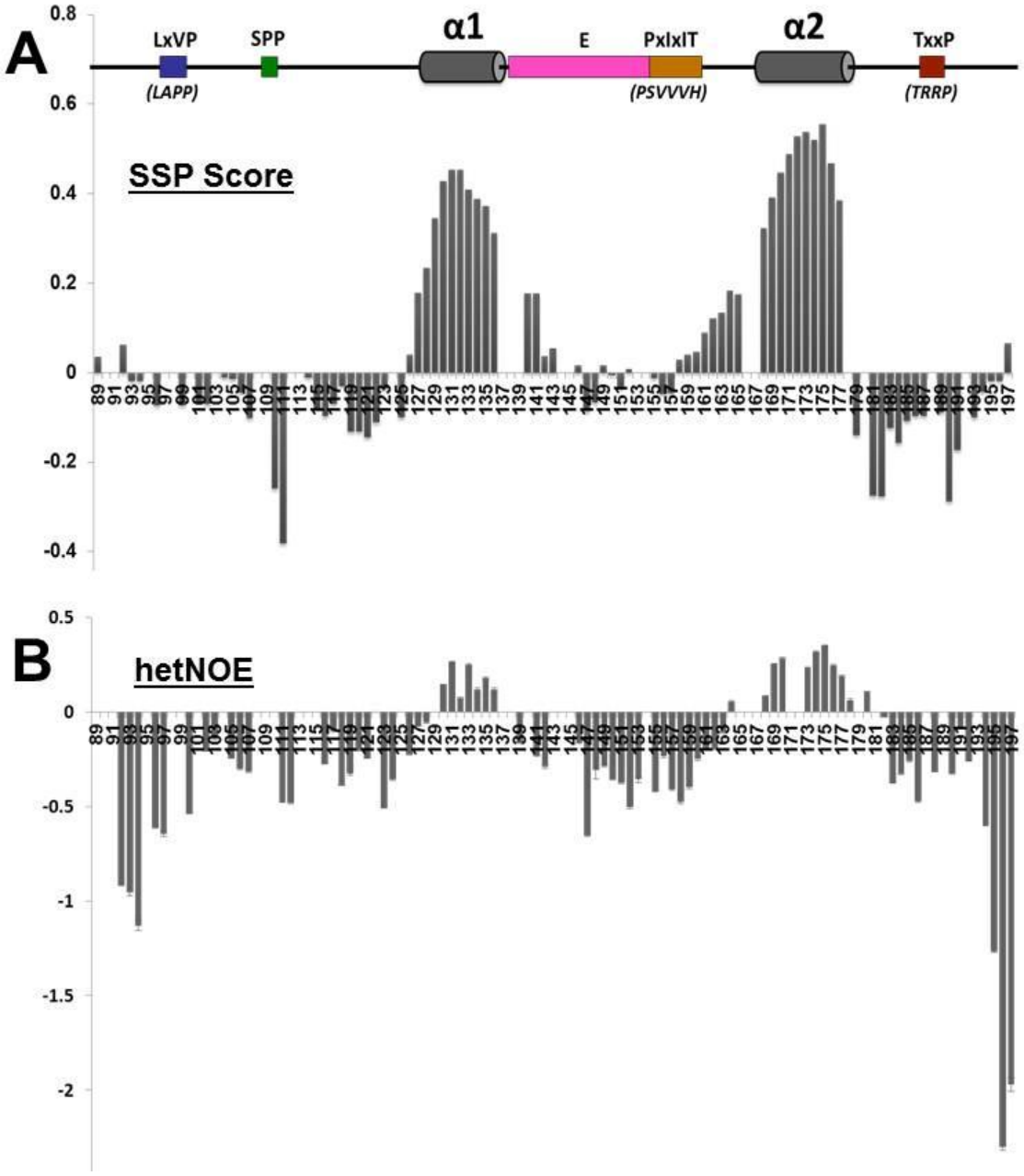


Figure 42: Secondary Structure Propensity (SSP) and hetNOE analysis of RCAN1$_{89-197}$. A) SSP: two regions of RCAN1 form α helices that are populated ~40-60% of the time. The rest of RCAN1 lacks significant secondary structure, with SSP scores falling within the -0.4 to 0.4 range. The conserved CN-binding LxVP, SPP, EV (including PxIxIT), and TxxP motifs show no secondary structure. B) hetNOE analysis is in agreement with SSP scores. RCAN1$_{89-197}$ is highly flexible, except at the regions shown by SSP analysis to form helices α1 and α2. The N- and C-termini show increased flexibility, as would be expected for these less restricted regions.

of RCAN1$_{89-197}$ is highly flexible (small or negative hetNOE values), with the exception of the regions forming the transient α1 and α2 helices (Figure 42B).

4.3.C. NMR analysis of the RCAN1-CN complex

NMR spectroscopy was used to investigate the interaction between RCAN1 and CN. We recorded the 2D [^{1}H,^{15}N] HSQC spectrum of a complex between ^{15}N-labeled RCAN$_{89-197}$ and unlabeled CN$_{A1-370,B1-170}$ (Figure 43). Only residues of ^{15}N-labeled RCAN1 that do not bind to CN are detected in the 2D [^{1}H,^{15}N] HSQC spectrum (red in Figure 43), while amino acids that interact with CN are not visible. Regions of ^{15}N-labeled RCAN that directly interact with CN (62 kDa) behave as though they were part of a larger protein and experience severe peak

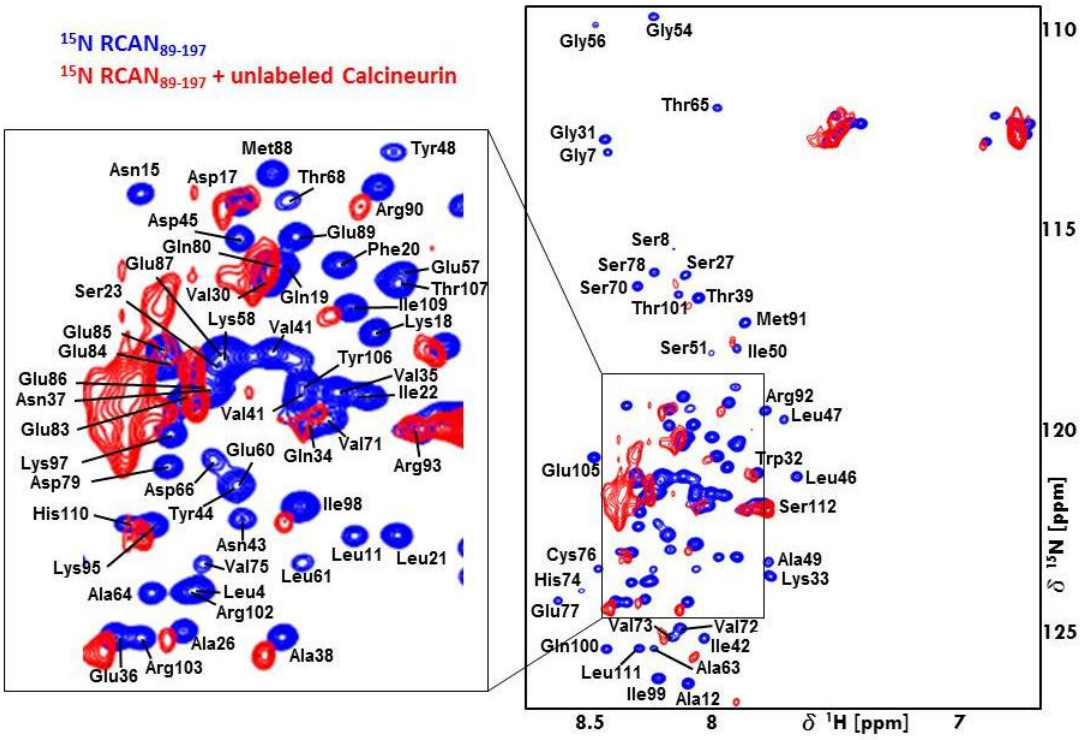


Figure 43: NMR of the RCAN1-CN complex.
2D [^{1}H,^{15}N] HSQC of ^{15}N-RCAN$_{89-197}$ in complex with unlabeled CN$_{A1-370,\ B1-170}$ (red) overlaid with 2D [^{1}H,^{15}N] HSQC of apo-^{15}N-labeled RCAN1$_{89-197}$ (blue). The peaks observed in red correspond to RCAN1 residues that do not interact with CN.

broadening caused by the rapid decay of the NMR signal before it can be detected.

Direct comparison of the 2D [^{1}H,^{15}N] HSQC spectra of unbound and CN-bound RCAN1 (Figure 43), readily identifies residues of RCAN1 that do not bind to CN (regions represented as red dotted lines in Figure 44). Interestingly, the non-interacting RCAN1 residues are restricted to two major regions: the disordered stretch of amino acids between the SPP motif and helix α1, and helix α2. While helix α1 directly interacts with CN, helix α2 does not bind to CN and thus is likely important for the recruitment of additional proteins to the RCAN1-

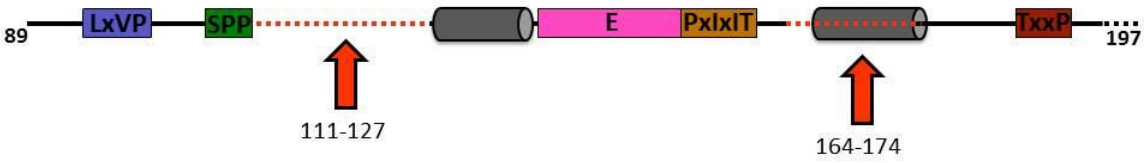


Figure 44: Identification of the CN-interacting interface of RCAN1.
Dotted red lines represent regions of RCAN1 that do not participate in CN binding directly. The amino acids making up these regions give rise to the only visible peaks in the RCAN1-CN 2D [^{1}H,^{15}N] HSQC. One region immediately precedes α1, while the other overlaps with α2.

CN complex, potentially helping RCAN1 to function as a scaffolding or targeting protein. The RCAN1 TxxP motif is not visible in the complex 2D [^{1}H,^{15}N] HSQC spectrum, indicating either that: 1) under these conditions the TxxP motif binds to CN, presumably at the active site, or 2) the TxxP motif is simply in the proximity of the CN active site (though not necessarily bound), and loss of signal is due to the paramagnetic effect of the active site metals. We could not determine with certainty whether the Pro-rich SPP motif is bound to the CN active site, because Pro residues lack backbone N-H pairs and thus cannot be detected in a 2D [^{1}H,^{15}N] HSQC spectrum. It is interesting to note, however, that

the SPP motif is located between a bound segment of RCAN1 (immediately N-terminal to the SPP) and an unbound region (immediately C-terminal to the SPP). Given its location, it is possible that the SPP is displaced from the active site by the TxxP motif, though nearby LxVP motif clearly remains bound and anchors the N-terminus of RCAN1 to CN.

4.3.D. ITC studies of RCAN1 and CN

To further characterize the RCAN1:CN interaction, we used Isothermal titration calorimetry (ITC). Surprisingly, titration of $RCAN1_{89\text{-}197}$ into CN produced a bimodal isotherm (Figure 45A), which is typically indicative of multiple ligands binding to one protein (i.e. 2:1 stoichiometry). However, SEC shows that the RCAN1-CN complex has 1:1 stoichiometry, and there have been no literature reports of multiple RCAN1 molecules interacting with one CNA/CNB heterodimer. As explained in Chapter 2, a 1:1 interaction between one ligand (RCAN1) and one protein (CN) is expected to produce an isotherm with a single equivalence region even when multiple binding sites are involved. Because two-site or sequential binding models cannot accurately describe this mode of binding, we used a one-site binding model to evaluate the ITC data (Figure 45B).

To extract specific information about the individual contribution of each CN-binding region of RCAN1, we carried out ITC experiments with a large number of truncated RCAN1 constructs. The constructs were designed to contain different combinations of CN-binding motifs (i.e. E, PxIxIT, LxVP, TxxP), based on consensus sequences and the binding information obtained from NMR analysis. These ITC results are summarized in Figure 46. It is important to

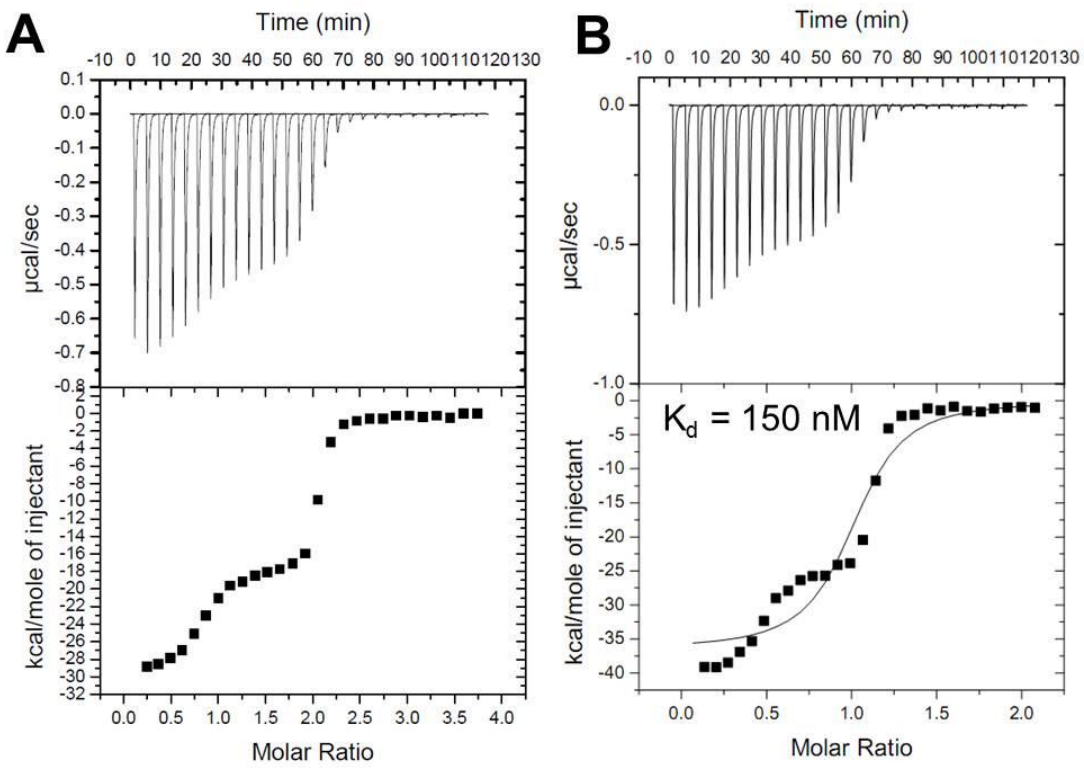


Figure 45: Isotherms obtained for the titration of RCAN1$_{89-197}$ with CN.
A) The isotherm is bimodal, showing two transitions. B) The same data as in A, fit with a one-site binding model. The K_d value shown is approximate, and represents all binding events.

emphasize that the values of the thermodynamic parameters calculated for these interactions are approximations, as we are currently using a one-site binding model to evaluate these complex isotherms due to the lack of a structural model. All datasets were processed in an identical manner, to allow for direct comparisons between these different ITC experiments. This analysis led to four key conclusions that provided novel insight into the binding of RCAN1 to CN.

First, comparison of RCAN1$_{151-192}$ and RCAN1$_{128-192}$ ITC data revealed that the region containing α1 and the E motif makes a key contribution to the overall binding of RCAN1 to CN, as it increases the overall K_d of RCAN1 ~30-fold (120 nM to 3,500 nM) (Figure 47A). Second, helix α2 of RCAN1 does not

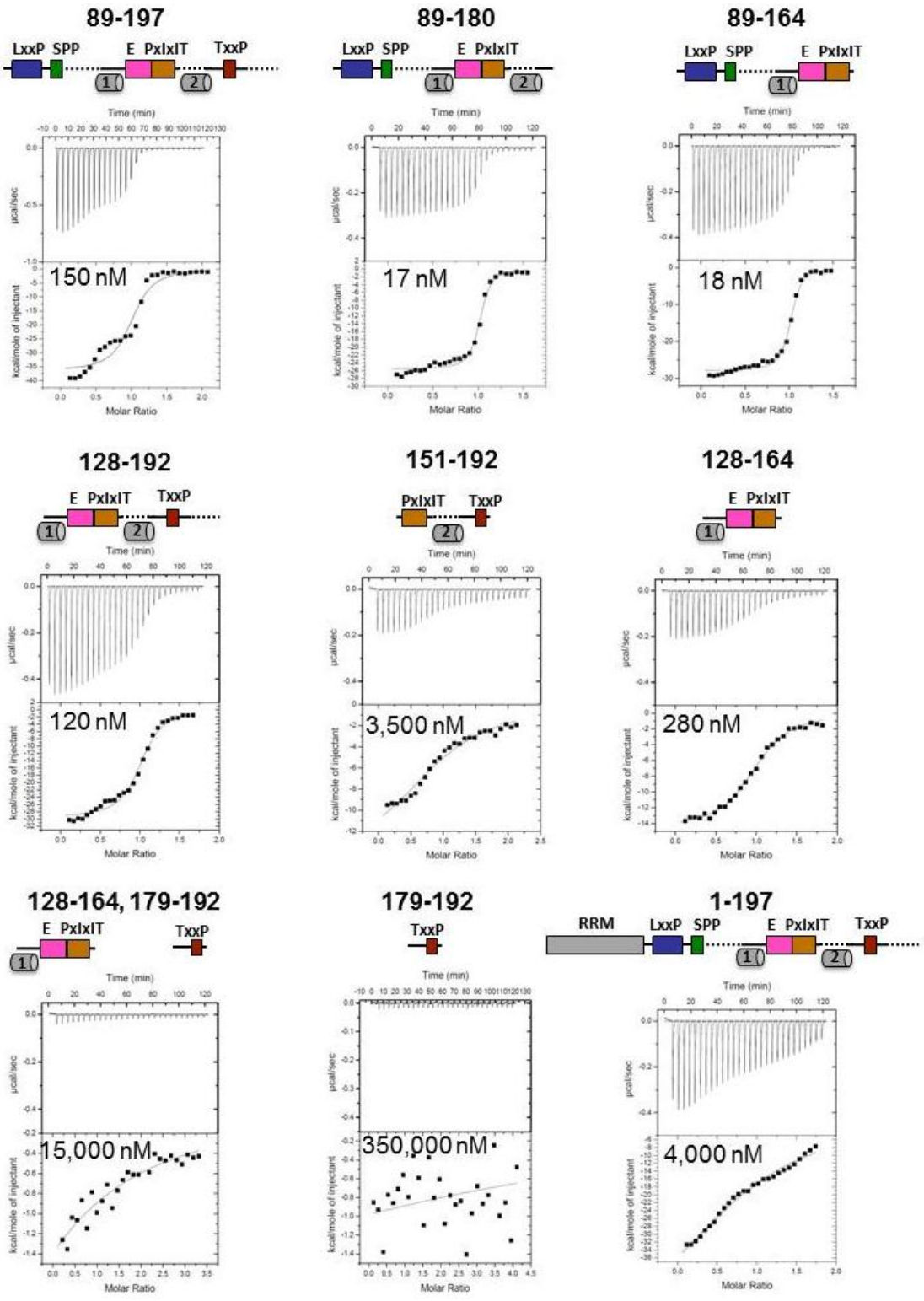


Figure 46: Overview of ITC experiments.
The amino acid residues and a representation of each RCAN1 construct is shown above each set of ITC results. For each experiment, the raw data (µcal/sec) and binding curve (kcal/mole of injectant) are shown.

contribute to RCAN1-CN binding, confirming results obtained by NMR: RCAN189-$_{180}$ (+ helix α2) and RCAN1$_{89-164}$ (- helix α2) have identical K_d values (17 nM and 18 nM, respectively) (Figure 47B) as well as isotherm shapes (Figure 46). *Third*, and critical, the TxxP motif does not make a strong contribution to the overall RCAN1:CN interaction (Figure 47C). Nevertheless, it is essential for the bimodal form if the ITC isotherm, indicating it plays an important role in CN regulation. Deletion of TxxP increases the K_d ~8-fold (RCAN1$_{89-197}$ vs. RCAN1$_{89-180}$), while deletion of the α2-TxxP region increases the K_d ~2-fold (RCAN1$_{128-192}$

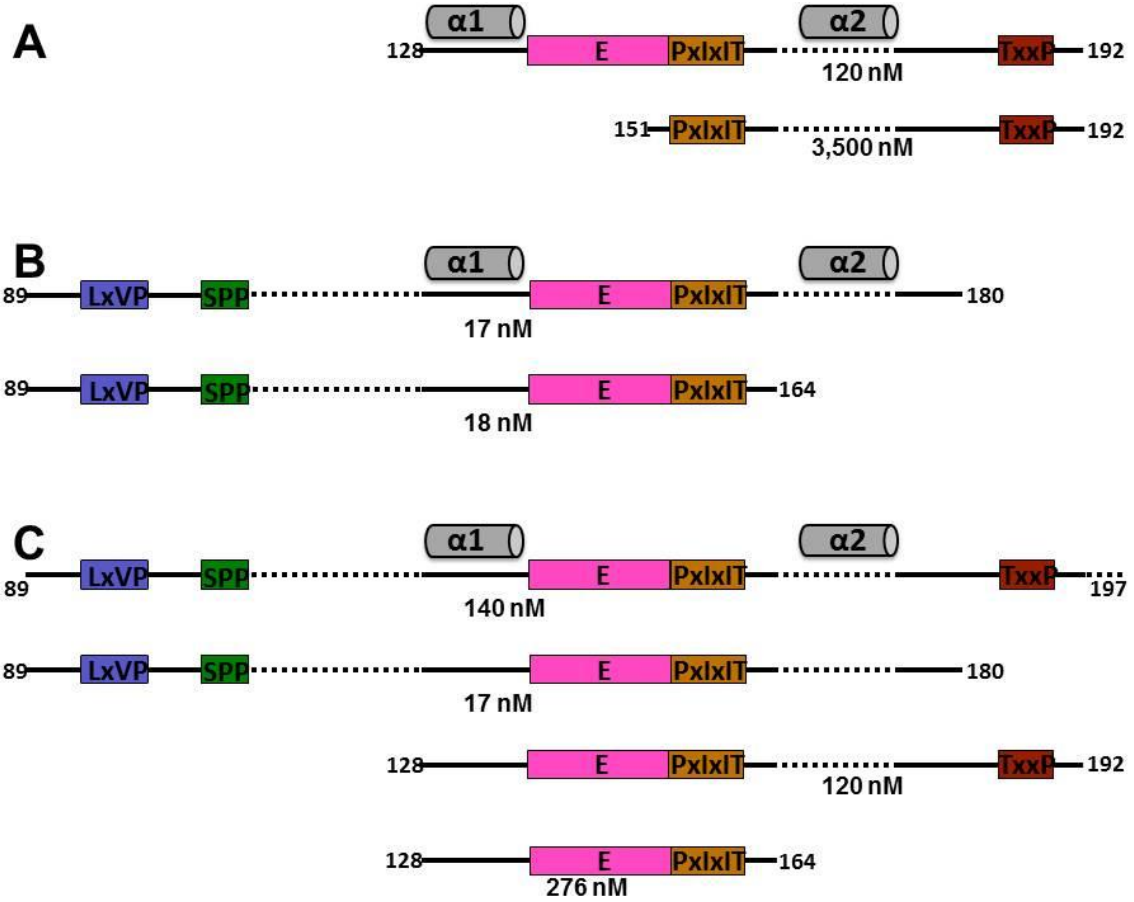


Figure 47: Comparison of ITC results to determine binding contributions of specific RCAN1 regions.
A) α1 and the E motif make a significant contribution to binding, and K_d increases ~30-fold in the absence of this region. B) α2 does not contribute to binding, and the K_d does not change when it is deleted. C) The TxxP makes a much smaller contribution to binding than the α1-E motif region. Deletion of the α2-E motif region increases the K_d ~8-fold, and deletion of α2 and the TxxP together increases the K_d by ~2-fold.

vs. RCAN1$_{128-164}$) (Figure 47C). Since we have shown that RCAN1 helix α2 does not bind to CN, the difference between these two K_d values reflects the error associated with fitting the bimodal ITC data using a one-site binding model. *Fourth*, the absence of the LxVP and SPP motifs did not significantly alter the isotherm or the K_d value (compare RCAN1 constructs 89-197 and 128-192 in Figure 46 and Figure 47).

To determine the relative binding contributions of the PxIxIT and LxVP motifs, we performed ITC using two mutants of RCAN1$_{89-197}$: a "PxIxIT mutant" (PSVVVH→ASAVAA) and an "LxVP mutant" (HLAPP→AAAAA), where conserved residues were mutated to Alanine residues (Figure 48). The "H" preceding the LxVP was also mutated since many LxVP sites are immediately preceded by an aromatic residue (often Tyr or Phe). The LxVP mutant, which

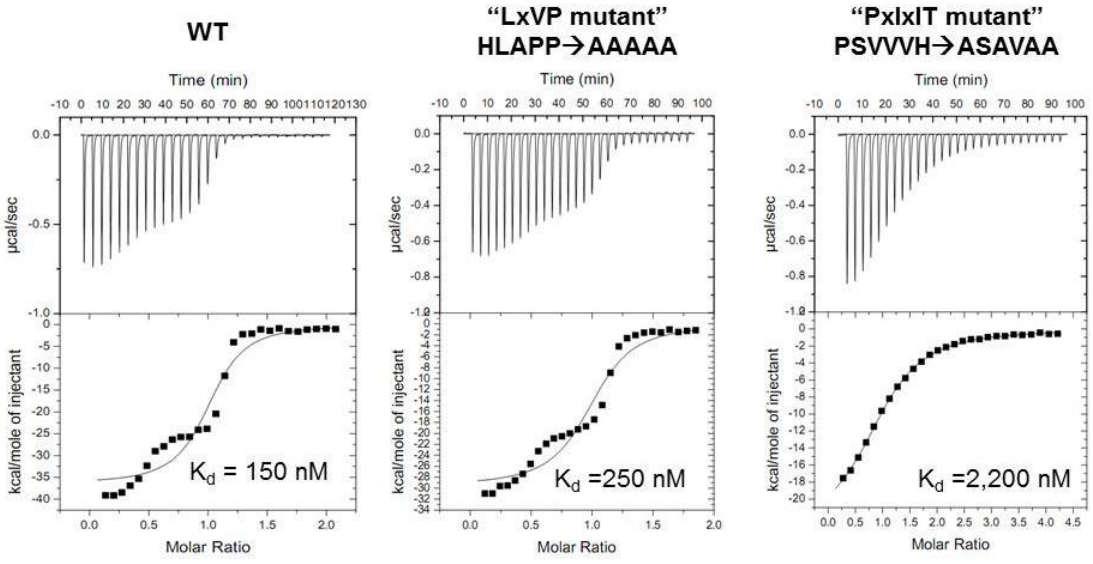


Figure 48: PxIxIT makes a larger contribution to binding than LxVP.
Constructs shown are RCAN1$_{89-197}$. The LxVP mutant has an intact PxIxIT. It closely resembles wild-type RCAN1 in the shape of the isotherm, and has a K_d that is very close to that of the wild-type. The PxIxIT mutant, on the other hand, produces a very differently shaped isotherm, and has a K_d that is ~15-fold higher than wild-type RCAN1.

contained an intact PxIxIT motif produced a bimodal isotherm identical in shape with wild-type $RCAN1_{89-197}$, with a K_d value (250 nM) ~1.5-fold larger than wild-type (150 nM). The PxIxIT mutant, on the other hand, showed a dramatic difference in the shape of the isotherm, and a ~15-fold increase in K_d. Therefore, this analysis showed that the PxIxIT motif makes a more significant contribution to CN binding than the LxVP, and that the PxIxIT is likely necessary for the interaction of α1-E with CN.

Bimodal curves are only observed when both the TxxP motif region (residues 179-192) (Figure 49A) and residues 128-164 of RCAN1 (α1 + E-motif + PxIxIT) (Figure 48 and Figure 49A) are 1) present and 2) part of the same construct. This clearly suggests cooperativity between these regions, with the first binding event likely causing a major molecular change that enables the second binding event to occur. A short peptide encoding the TxxP ($RCAN1_{179-192}$) shows almost no intrinsic affinity for CN (Figure 47C). However, titration of this peptide into a pre-formed CN-$RCAN1_{128-164}$ complex increases the K_d ~20-fold. Therefore, binding of RCAN1 residues 128-164 must induce a change in the CN conformation that enables TxxP cooperative binding. Binding affinity is further increased when the TxxP and residues 128-164 of RCAN1 are present on the same construct, likely due to the closer proximity of the TxxP to its binding site after the first binding event.

Additional cooperativity takes place between the PxIxIT motif and the α1-E-motif region. The PxIxIT motif is necessary but not sufficient for inducing cooperative binding of the TxxP motif. Mutating the PxIxIT motif of $RCAN_{89-197}$

produced a single phase isotherm, suggesting that this motif alone is necessary for TxxP cooperative binding (Figure 48). However, ITC with RCAN1$_{151-192}$ (PxIxIT-α2-TxxP) shows that the PxIxIT sequence is not sufficient for inducing TxxP binding in the absence of α1 and the E-motif. Similarly, the α1-E-motif region is necessary (Figure 49B) but not sufficient (Figure 48) to induce TxxP cooperative binding. Since PxIxIT binding has no effect on CN structure [1, 4, 144], the RCAN1 PxIxIT motif likely serves to target (or increase the affinity of) the E motif and α1 to CN. Therefore, Helix α1, the E-motif, and the PxIxIT motif cooperative induce a conformational change in CN that enables binding of the TxxP motif. Cooperativity between these elements is less apparent and would not have been detected were it not for its impact on TxxP binding.

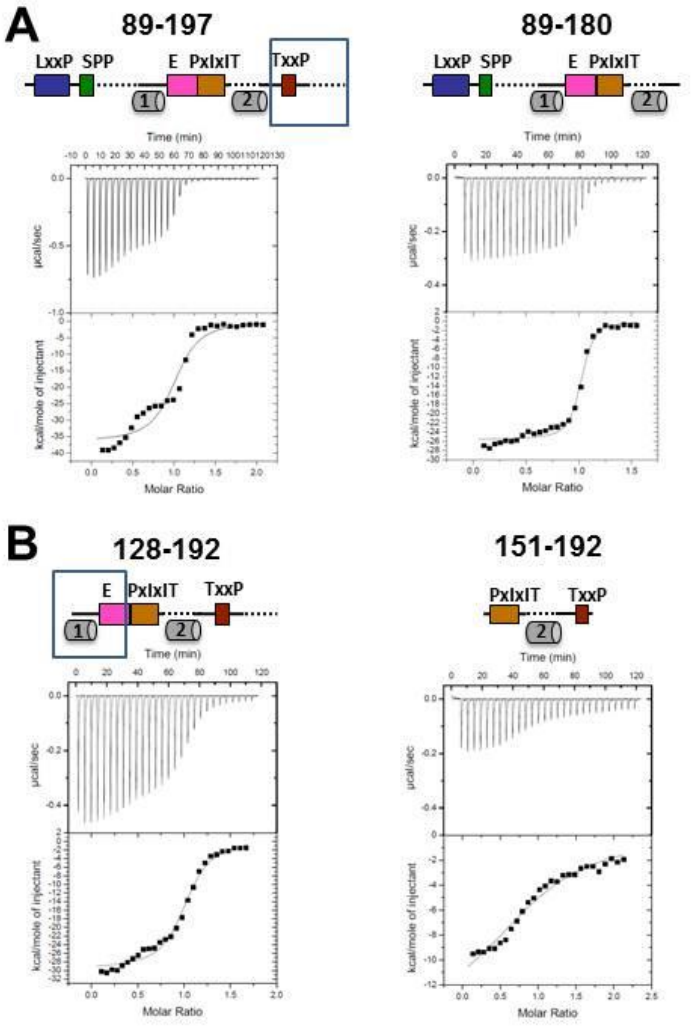


Figure 49: Bimodality requires the presence of the α1-E motif region and TxxP.
A) Deletion of the TxxP causes an isotherm to more closely resemble a one-site binding interaction. B) Deletion of the α1-E also causes the curve to become less bimodal.

4.4. Discussion

This work provides the most detailed characterization of RCAN1 and its interaction with CN to date. We showed that the RCAN1 CN-binding domain (residues 89-197) is intrinsically disordered, with the exception of two α-helices that flank the EV-motif and have distinctly different characteristics: helix α1 is predominantly hydrophobic and binds to CN, while α2 is almost entirely charged and does not interact with CN. Interestingly, phosphorylation of Ser136, a conserved residue at the C-terminus of α1, was previously shown to decrease RCAN1 inhibition of CN [152], possibly by disrupting α1 secondary structure and/or preventing α1 docking to CN. The full α2 amino acid sequence is found only in RCAN1, which indicates it serves a function specific to RCAN1. A possible function for α2 is the recruitment of additional proteins to the RCAN1-CN complex. For example, a recent study showed that RCAN1 interacts with FMRP (Fragile-X Mental Retardation Protein) and promotes its dephosphorylation by CN [159]. FMRP could bind to α2 of RCAN1, with RCAN1 serving as a scaffold to bring FMRP and CN together.

We showed by ITC that a short peptide (residues 179-192) encoding only the TxxP motif of RCAN1 was unable to bind CN. However, the presence of separate $RCAN1_{128-164}$ (α1 + E-motif + PxIxIT) induced weak TxxP binding. The only possible interpretation is that $RCAN1_{128-164}$ induces a conformational change in CN that exposes the TxxP binding site. Several structures of CN bound to PxIxIT motifs show that PxIxIT binding induces absolutely no change in the structure of the active site. Therefore, the allosteric change must be induced by

α1 or the E-motif. Martinez-Martinez et al showed that a peptide encoding the TxxP motif together (i.e. in *cis*) with the E- and PxIxIT motifs could not inhibit CN phosphatase activity against pNPP or RII, but could bind CN and inhibit NFAT dephosphorylation *in vivo* [140]. The E- and PxIxIT motifs are therefore not sufficient to induce a conformational change in CN and expose the TxxP binding site. Combined with our results, this suggests that α1 is responsible for the conformational change that allows TxxP binding to CN. Though the PxIxIT motif almost certainly serves as a targeting sequence for α1, the E-motif could either serve to further increase α1's affinity for CN, or it could directly contribute to the allosteric change induced by α1.

Based on our NMR and ITC results, we propose a sequence of steps that mediate the RCAN1-CN interaction (Figure 50). In the first binding event, nearly the entire CN-binding domain of RCAN1, (α1, the E-motif, PxIxIT, LxVP, and SPP) binds to CN, giving rise to the first transition in the ITC isotherm. This binding induces a conformational change in CN, possibly at the active site that enables the final binding of the TxxP motif. For example, even a slight change in the active site conformation could cause the expulsion of the CN catalytic metals and enable TxxP binding to the active site. The large activation energy barrier associated with the ejection of active site metals would require saturation of the first binding event and generate a bimodal isotherm. Given this model, the TxxP could bind somewhere other than the active site. For instance, it could bind near the active site in a way that helps stabilize the "new" metal-free conformation and still inhibit CN catalytic activity.

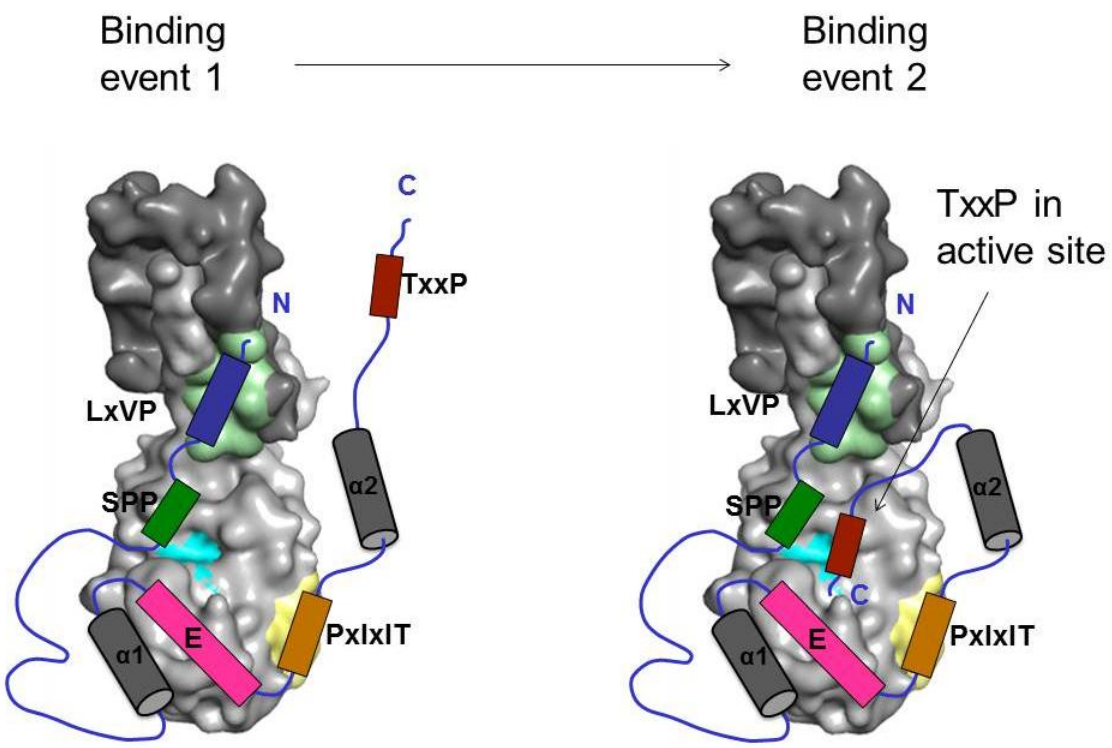


Figure 50: Model for RCAN1 binding to CN, based on NMR and ITC data.
During the first binding event, the LxVP, SPP, α1, E-motif, and PxIxIT bind to CN. The binding sites of the individual motifs, especially α1 and the E-motif, on CN is not known, and represented for illustrtaive purposes here. The first binding event induces a change in CN that enables the second bindnig event: TxxP binding at or near the active site.

Like the viral inhibitor A238L [144], RCAN1 can inhibit CN by interfering with substrate recognition in the absence of the TxxP motif. This was demonstrated against NFAT *in vitro* [140] and is consistent with our results that show strong CN binding independent of the TxxP motif. The TxxP motif could provide an additional layer of inhibition to compensate for the low affinity of its LxVP site. It is also interesting to note that the phosphorylation sites that reportedly influence whether RCAN1 inhibits or activates CN are located in CN-binding regions (see Figure 39): adjacent to the LxVP, on α1, and adjacent to the

TxxP. Thr192 phosphorylation increases RCAN1's inhibition of CN, possibly by enhancing the affinity of the TxxP motif for CN.

Here we analyzed the entire CN-binding domain of RCAN1 and described the involvement of each region in CN binding. Our NMR analysis shows a high amount of flexibility present even in the CN-bound form of RCAN1, which explains why crystallographic studies have so far been unsuccessful.

Chapter 5. Outlook

5.1. Conclusions and significance of this work

This dissertation work has made several significant contributions to our understanding of calcineurin (CN) function and regulation. First, the CN-A238L structure revealed the location of the LxVP-binding pocket on the surface of CN, providing a long-awaited atomic-level description of the CN-substrate LxVP interface. The structure of the viral protein A238L bound to calcineurin (CN) showed that the “L” and “V” of the LxVP occupy deep hydrophobic pockets at the interface of the CNA and CNB subunits, making them critical for LxVP binding to CN. The LxVP structure also showed that a Phe (“F”) residue immediately N-terminal to the A238L LxVP sequence (FLcVK) occupies a hydrophobic pocket on the CNB surface, supporting the idea that aromatic residues at this position serve to enhance binding strength in a subset of CN substrates.

Secondly, we provide an understanding of the structural basis for CN inhibition by the small molecule immunosuppressants cyclosporin A (CSA) and FK506. Crystal structures of CN in complex with CSA and FK506 show that these inhibitors do not bind to the active site. Consequently, the mechanism by which CSA and FK506 inhibited CN activity has remained unclear. Here, we showed that these immunosuppressant drugs bind to the LxVP binding pocket of CN, thereby preventing the interaction between CN and LxVP-containing substrates.

Finally, we used a combination of NMR spectroscopy and isothermal titration calorimetry (ITC) to achieve novel insight into the mechanism employed by RCAN1 to play a dual role in both the inhibition and enhancement of substrate

dephosphorylation by CN. We identified two distinct α-helices in the CN-binding domain of RCAN1: α1, which is hydrophobic and binds to CN, and α2, which is highly charged and does not interact with CN. Our data support a model in which α1, in concert with additional CN-binding elements, induces an inhibitory conformational change in RCAN1. Helix α2, on the other hand, may recruit additional proteins to the RCAN1-CN complex to facilitate their dephosphorylation by CN. Whereas previous studies focused mostly on identifying conserved CN-binding motifs and testing their effect on CN activity using biochemical and cell biology assays, we begin to provide an understanding of *how* RCAN1 affects CN function.

5.2. Future directions

5.2.A. Structural and biophysical characterization of CN-NFAT interactions

Understanding how the PxIxIT and LxVP substrate motifs individually recognize CN was an important milestone in the characterization of CN substrate recognition. However, a description of how the docking of NFATs to CN via their PxIxIT and LxVP motifs results in the dephosphorylation of 12 Ser residues is still lacking. CN regulates a variety of cellular processes by deposphorylating NFATs, and understanding this complex process will not only broaden our understanding of substrate dephosphorylation by Ser/Thr phosphatases, but will also aid in the development additional CN inhibitors with even higher specificity. Despite the rapidly increasing amount of biochemical and cellular data describing CN-NFAT interactions, structural and biophysical descriptions of CN-NFAT complexes are

missing. Such studies are necessary to understand how all 12 NFAT phospho-residues interact with the active site of CN.

Structural studies of CN-NFAT complexes are extremely challenging. The 12 phospho-residues in each NFAT are located within a ~300 amino acid long regulatory region that is flanked by the PxIxIT and LxVP motifs. This regulatory region is predicted to be unstructured, making X-ray crystallography studies of CN-NFAT complexes nearly impossible (Figure 51). In addition, NFATs are difficult to study using NMR spectroscopy due to their large size. Small angle X-ray scattering (SAXS) of CN-NFAT complexes, however, can be used to obtain molecular envelopes like that shown in Figure 52, and could be useful in determining the level of disorder in NFATs. SAXS could also be used to detect any preferred conformations adopted by NFATs at a global level. For example, SAXS experiments of CN in complex with NFAT constructs that contain non-hydrolyzable phosphomimetic mutations (i.e. Ser to Glu) may be particularly useful for trapping distinct conformations of CN-bound NFATs and study their distinct

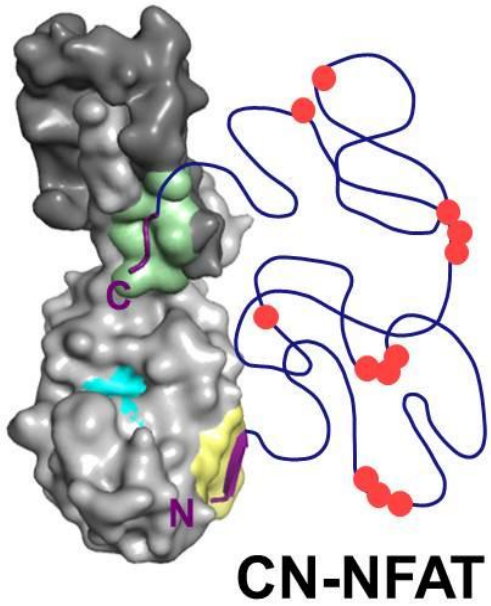


Figure 51: Representation of a CN-NFAT complex.
CN is shown in grey with the active site in cyan, LxVP binding site in green, and PxIxIT binding site in yellow. The regulatory region of NFAT is shown in blue with the LxVP and PxIxIT sites that flank it colored in purple and bound to the green and yellow surfaces of CN, respectively. Phosphorylated Ser residues are shown in red. The regulatory region of NFATs is highly disordered and therefore not amenable to crystallization.

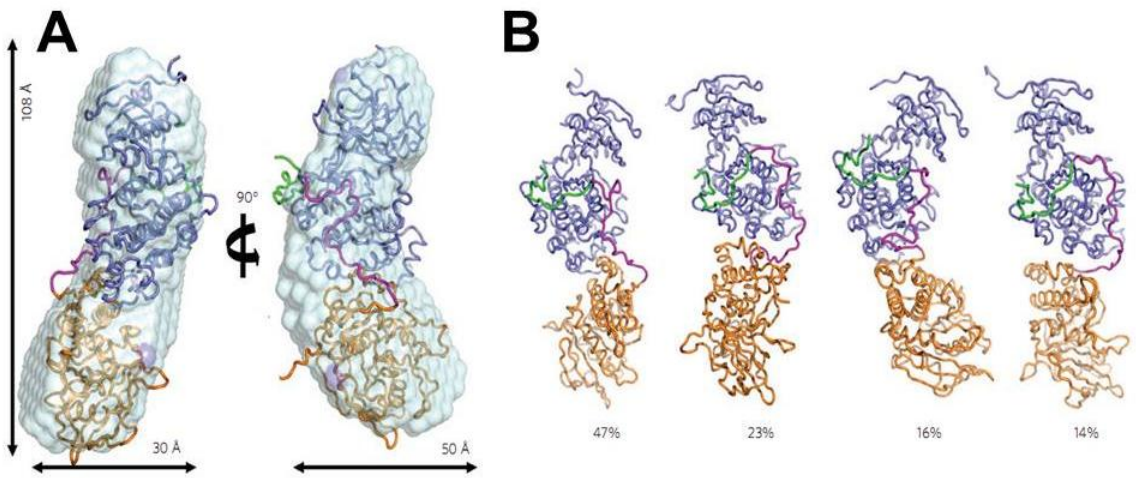


Figure 52: Small angle X-ray scattering (SAXS).
A) SAXS provides information about the global structure of a molecule or complex in solution. The final result consists of a molecular envelope into which existing structures of the individual components can be modeled. B) The shape of the molecular envelope reflects an ensemble of different conformations, and the percentage of time each conformation is occupied can be calculated. Adapted from Francis et al (2011) Nat. Chem. Biol.

modes of binding to CN. In order to trap individual NFAT conformations, these experiments would be performed using a series of NFAT constructs where either single Ser residues or clusters of adjacent Ser residues are systematically mutated to Glu. Additional SAXS and ITC experiments of CN in complex with truncated NFAT samples that contain shortened regulatory regions flanked by the PxIxIT and LxVP can provide further information about how each phospho-Ser (or cluster of neighboring phospho-Ser residues) is targeted to the active site of CN. Such complexes between CN and truncated NFAT constructs may also be amenable to X-ray crystallography studies, due to the removal of long unstructured regions.

One possible explanation for how all 12 NFAT Ser residues sample the CN active site is the presence of several LxVP sites within the NFAT regulatory region. Each LxVP site would favor the placement of specific phospho-Ser residues or clusters into the CN active site. LxVP sequences are difficult to

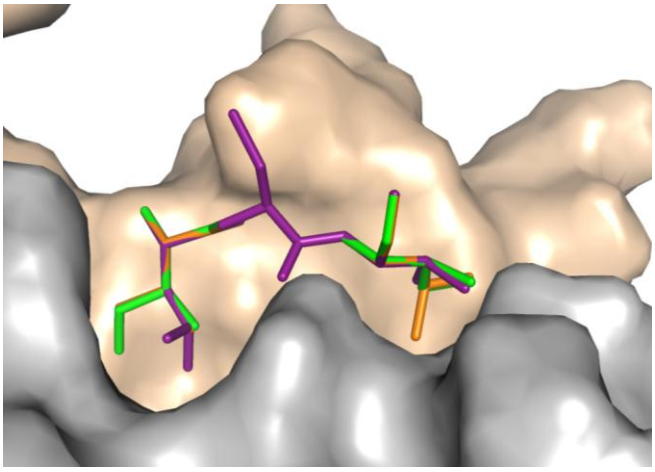


Figure 53: Potential variations in the LxVP motif. The residues LcV (purple) from the A238L LxVP motif shown bound to the CNA-CNB interface. Mutations of the L and V residues in LcV were simulated using the program Pymol: LcV→VcL (orange) and LcV→IcI (green) mutations show that the "L" and "V"-binding pockets of CN may each be able to accomodate a Leu, Val, or Ile residue. Therefore, the LxVP motif could be as broad as [L/V/I]x[L/V/I].

identify, because any "LxV" sequence in NFAT could qualify to be an LxVP motif (the preceding aromatic residue and the "P" at the end are not necessary for binding). Furthermore, the "LxV" motif could be as varied as [L/V/I]x[L/V/I], since the "L" and "V" binding pockets on the CN surface could each potentially accomodate Leu, Val, or Ile residues (Figure 53). Analysis of the NFATc1-4 primary sequences using the 3of5 pattern matching server [160] found 8, 4, 5, and 4 [L/V/I]x[L/V/I] sites between the PxIxIT and LxVP motifs of NFATc1-4, respectively. The ability of these motifs to bind CN can be confirmed experimentally by ITC and NMR experiments that use truncated NFAT constructs.

5.2.B. Development of new immunosuppressant drugs

FK506 and CSA bind to CN in the presence of immunophilin proteins, forming CN-drug-immunophilin complexes. Most drug design efforts over the past decades have focused on mimicking or stabilizing the immunophilin-CN interaction, which many people believed to be ultimately responsible for CN inhibition. However, comparison of CN-drug-immunophilin complex structures

with our CN-A238L structure shows that FK506 and CSA closely imitate the specific interactions made by the “L” and “V” of the LxVP with the deep hydrophobic pockets on the CN surface, and confirm that it is FK506 and CSA that are important for CN inhibition, not the immunophilins.

Importantly, FK506 and CSA are natural fungal-derived products that have not been optimized for CN binding. Both are cyclic molecules that bind CN using one side of the ring and an immunophilin via the other (Figure 54) ([17]. In the case of CSA, the immunophilin induces a conformational change in the drug, allowing it to bind to CN in a lock-and-key fashion. A new drug could be designed

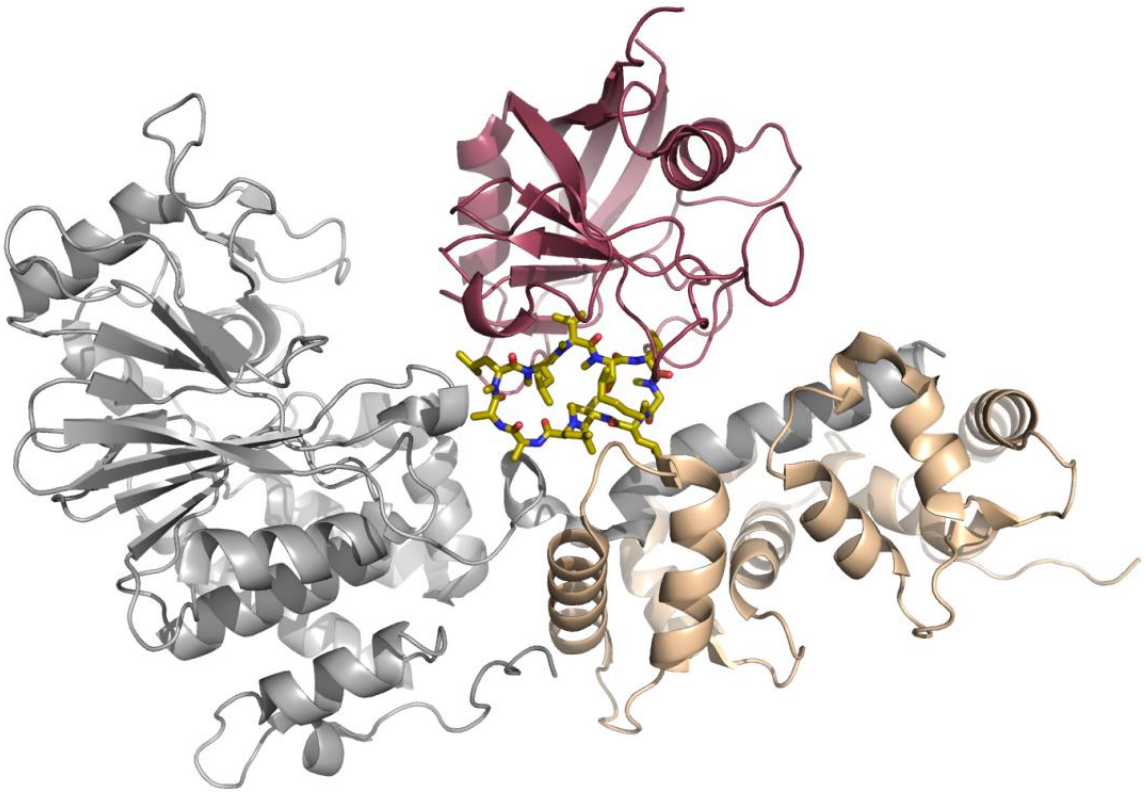


Figure 54: Structure of the CN-CSA-Cyclophilin complex.
Cyclosporin A (CSA) is a ring. It binds to the immunophilin protein cyclophilin (pink) using one side of the ring and to CN (grey and beige) using the opposite side. The side of the ring that binds to cyclophilin does not interact with CN.

to be linear instead of cyclic, eliminating the region that binds to immunophilins altogether. A new “linear” drug that is capable of binding to CN tightly without help from an immunophilin cold lead to decreased side effects compared with currently used immunosuppressants. The severe side effects in patients caused by FK506 and CSA could be due in part to the sequestration of immunophilins by CN-drug-immunophilin complexes, which could prevent immunophilins from carrying out their cellular functions as cis/trans isomerases.

The structural description of the LxVP site can be used as a tool in the design of new immunosuppressant drugs. Drugs with higher affinity for CN can be synthesized by modifying FK506 and CSA to more closely mimic LxVP binding. In fact, a short peptide encoding an LxVP motif (i.e. FLCVP) with high affinity for CN could be used as the starting point for the design of new CN inhibitors. Information about the interactions of A238L residues outside of the LxVP motif could be used to engineer additional contacts outside of the “L” and “V” binding pockets to further increase affinity.

5.2.C. Structural characterization of RCAN1

Our combined NMR and ITC studies have provided an excellent overview of the RCAN1-CN interaction. However, a detailed structural description of the RCAN1-CN complex will be invaluable for understanding the mechanism of CN regulation by RCAN1. The dynamic nature of this complex, together with the many flexible regions of RCAN1, make the RCAN1-CN complex a poor crystallographic target. We have so far been unsuccessful in obtaining RCAN1-CN crystals that could be used for structure determination, despite using truncated RCAN1 constructs that

lacked the flexible regions detected by NMR. Future crystal optimization strategies include mutagenesis of the RCAN1 PxIxIT (PSVVVH) and LxVP (HLAPP) motifs to sequences that have higher affinity for CN, such as PVIVIT [4] (the PxIxIT sequence of the high affinity synthetic peptide) and FLCVP [144] (the LxVP sequence of A238L, except with the last “K” residue replaced with a more highly conserved “P”).

Interestingly, many phosphorylation sites that regulate RCAN1 function are located near CN-binding elements, and their phosphorylation could have a stabilizing effect on RCAN1 and its CN-bound form, thus promoting crystal formation. For example, phospho-Thr192 is separated from the TxxP by two amino acids. Phosphorylation of RCAN1 at Thr92 increased its inhibition of CN catalytic activity against pNPP and could enhance TxxP affinity for the CN active site.

Finally, it will be interesting to determine whether RCAN1 can act as a scaffold to bring CN close to certain substrates. Currently, the only studies showing that RCAN1 can function as an enhancer of CN activity are performed *in vivo,* where it is often difficult to control for the interference of additional factors. The Fragile-X syndrome protein FMRP is dephosphorylated by CN and is reported to interact with RCAN1 [159]. A triple complex between FMRP-RCAN1-CN may be more amenable to crystallization, with FMRP having a stabilizing effect on RCAN1.

Appendix A: Supplementary information for "The molecular mechanism of substrate engagement and immunosuppressant inhibition of Calcineurin"

Simina Grigoriu[2], Rachel Bond, Pilar Cossio, Jennifer A. Chen, Nina Ly, Gerhard Hummer, Rebecca Page, Martha S. Cyert* and Wolfgang Peti*

Reproduced from the journal PLoS Biology, Volume 11, Number 12, e1001492

A.1. Supplementary Protocols

A.1.a. Cloning

Cloning was carried out using standard protocols. The $A238L_{200-239}$ sequence from the Malawi LIL20-1 strain was PCR-amplified and cloned into the bacterial expression vector RP1B [155], which encodes an N-terminal $Thio_6$- and His_6-tag followed by a tobacco etch virus (TEV) protease cleavage site. The $CN_{A1-391/B1-170}$ and the $CN_{A1-370/B1-170}$ sequences were PCR-amplified and the open reading frames of both CNA and CNB subunits were cloned into the pET15b-derived p11 bicistronic bacterial expression vector (PSI:Biology-Materials Repository) as a single cassette, which contains an N-terminal His_6-tag followed by a TEV protease cleavage site. $A238L_{200-239}$ mutants were generated by direct cloning of annealed double-stranded overhang oligonucleotides into the digested $A238L_{200-239}$ expression plasmid. GST-peptide fusion expression vectors were generated by the same method using pGEX4T3 (GE Life Sciences) as the parent vector. The A238L C-terminal domain truncation series was generated by PCR amplification of desired sequences and cloning into the yeast expression vector pRD56. A pET32a vector containing S-tagged $A238L_{157-239}$ was generously provided by Linda Dixon (Pirbright Laboratory, UK). GST-tagged yeast $CNA1_{1-417}$ (BJP3003) and CNB1-pET9a have been previously described [126]. The mammalian expression vector pCDNASV5A238L encoding full-length A238L was provided by Linda Dixon. Docking motif mutants $A238L_{PKIIITmut}$ and $A238L_{FLCVKmut}$ were generated by site-directed mutagenesis.

A.1.b. Protein expression and purification

RP1B expression plasmid containing either wild-type or mutated (PKIIIT mutated to AKAIAA or FLCVK mutated to AACAA) $A238L_{200-239}$ was transformed into BL21-(DE3) RIL *E. coli* cells (Agilent) and expression was carried out in Luria broth medium containing chloramphenicol (34 µg/ml) and kanamycin (50 µg/ml). Cell cultures were grown at 37°C under vigorous shaking (250 rpm) to an OD_{600} of 0.7. Cells were cooled at 4°C for one hour, while the shaker temperature was lowered to 18°C. Expression of wild-type or mutant $A238L_{200-239}$ was induced by addition of 1 mM IPTG, and the cultures were grown for an additional 18 h at 18°C (250 rpm). The cells were harvested by centrifugation (6000x*g*, 15 min, 4°C) and stored at -80°C. The GST-peptide fusion plasmids and the p11 expression plasmid containing CN was transformed into BL21-(DE3) RIL *E. coli* cells (Agilent) and expression was carried out as described above for the $A238L_{200-239}$ samples. For GST-$CNA1_{1-417}$, CNB1, and S-$A238L_{157-239}$, plasmids were transformed into BL21-(DE3) *E. coli* cells and expressed as described above.

Wild-type and mutant (PKIIIT→AKAIAA and FLCVK→AACAA) $A238L_{200-239}$ were purified as follows. *E. coli* cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100, EDTA-free protease inhibitor tablets [Roche]) and lysed by high-pressure cell homogenization (Avestin C3 Emulsiflex). The cell debris was removed by centrifugation (35000x*g*, 50 min, 4°C) and the supernatant was filtered and loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 50 mM Tris

pH 8.0, 500 mM NaCl, 5 mM imidazole. His_6-tagged $A238L_{200-239}$ was eluted from the column with a 5-500 mM imidazole gradient over 60 minutes. Fractions containing purified $A238L_{200-239}$ were pooled, TEV (His_6-tagged; produced in-house) was added for His_6-tag cleavage, and samples were dialyzed against protein buffer (20 mM Tris pH 8, 500 mM NaCl) at 4°C for 16 h. The enzymatically-cleaved His_6-tag and TEV protease were removed by Ni^{2+}-affinity subtraction purification. $A238L_{200-239}$ used for crystallographic complex formation was further purified by size exclusion chromatography (SEC; Superdex 75 26/60 [GE Healthcare]) equilibrated in 20 mM Tris pH 7.5, 500 mM NaCl, 0.5 mM TCEP. $A238L_{200-239}$ was concentrated to ~1 mM and frozen at -80°C for later use. Finally, wild-type or mutant $A238L_{200-239}$ used for ITC was concentrated to >1 mM immediately following the subtraction purification step, and stored at -80°C. Immediately prior to ITC experiments, $A238L_{200-239}$ samples were thawed and SEC-purified in ITC buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM TCEP, 1.5 mM $CaCl_2$). $A238L_{157-239}$ was purified as described above up to and including the Ni^{2+}- affinity purification step, and stored at -80°C.

His_6-tagged $CN_{A1-391/B1-170}$ used for ITC was purified over a HisTrap HP affinity column as described above for $A238L_{200-239}$. Following removal of the His_6-tag and TEV by subtraction purification, $CN_{A1-391/B1-170}$ was further purified for ITC by SEC (Superdex 200 26/60 [GE Healthcare]) equilibrated in ITC buffer.

A.1.c. CN-dependent gene expression assays in yeast

GST or GST fused to full length or C-terminal truncations of A238L were expressed under control of the GAL1 promoter in *S. cerevisiae* strain SGY104

(*MATa lys2-801 ade2-101 trp1-Δ63 his3−Δ200 leu2-Δ1 ura3-52::TRP1*-2X CDRE *lacZ*), which harbors the CN-dependent reporter gene, 2X CDRE-lacZ [3]. Cells were grown to mid-log phase with 4% raffinose as the carbon source, and expression was induced with 2% galactose and supplemented with 50 mM $CaCl_2$ for 2.5 hrs to activate CN. β-galactosidase activity is reported as maximum rate OD_{415} change/minute/mg of protein and was determined as described in [161] using 300 µg protein in each assay. Values represent an average of three independent extracts, each measured in triplicate. Error bars indicate the standard deviation.

A.1.d. Luciferase reporter assays.

HEK293T cells were plated at 5 x 10^5 cell/well in 24-well plates and co-transfected the next day with indicated amount of A238L, NFAT-Luc [162], and Renilla-Luc[28] DNA in 2 µL of Lipofectamine2000 (Invitrogen) in 100 µL of OptiMEM (Invitrogen). The ratio of A238L:NFAT-Luc:Renilla-Luc was as follows: 1: 2: 0.5, with the exception of $A238L_{PKIIITmut}$ which was as follows: 5:2:0.5 to compensate for low expression. Fourteen hours after transfection, the cells were treated with 1 µM phorbol 12,13-dibutyrate (PDBu) (Sigma-Aldrich) to activate AP-1, and 1 µM ionomycin (Calbiochem). Luciferase levels were assayed 6 hours later using a homemade substitute [163] and a Veritas 96-well luminometer (Turner Biosystems) according to the instructions of the manufacturer. The luminescence of each sample was measured by integrating for 2 s after injection of the luciferase substrate. NFAT-Luc activity was normalized to Renilla-Luc activity for analysis. Protein expression levels were verified by western blot with

monoclonal anti-simian virus 5 (SV5) PK tag (Serotec) and monoclonal anti-GAPDH (Santa Cruz Biotechnology).

A.2. Supplementary Figures S1-S5

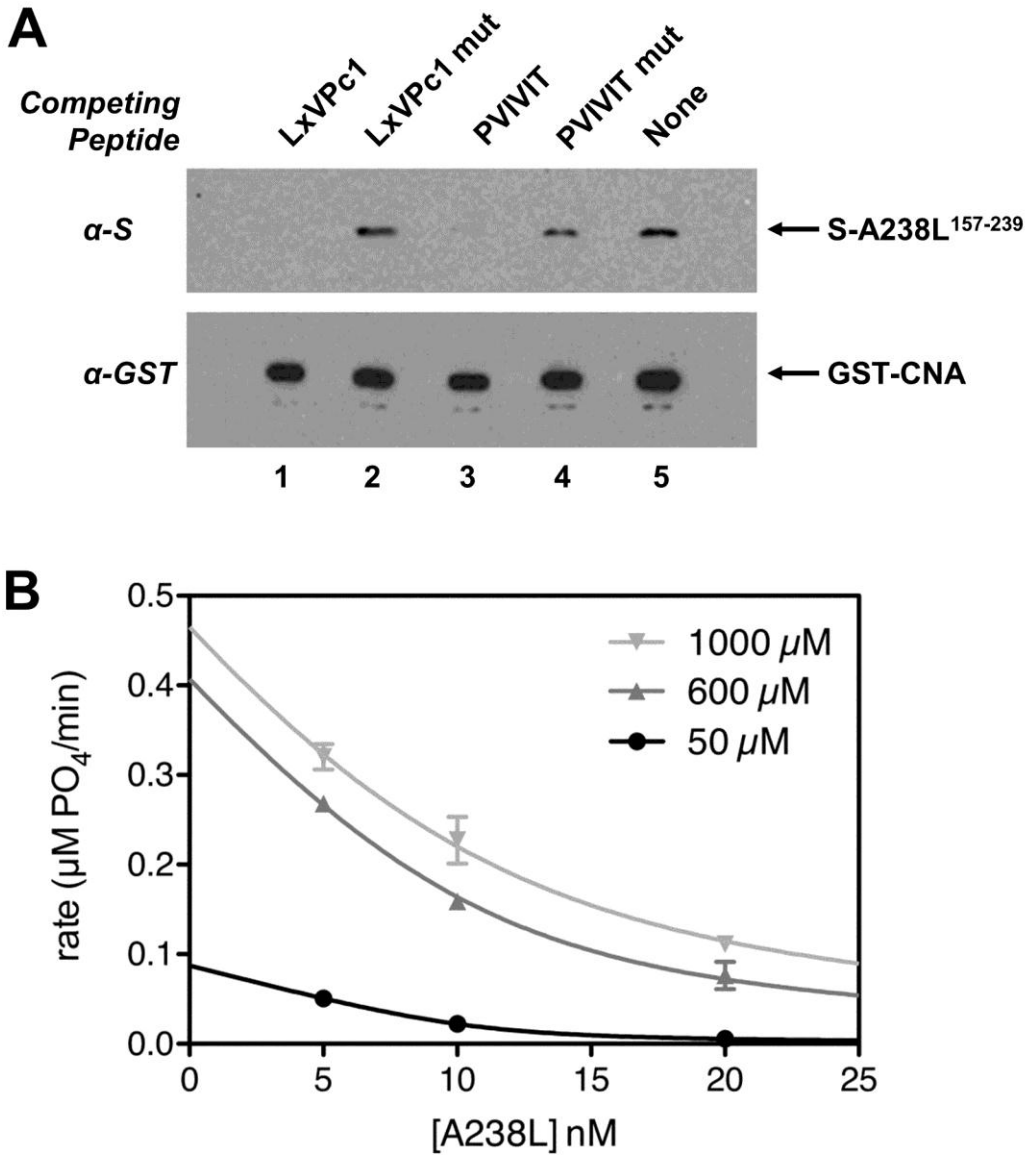


Figure 55: Figure S1, Grigoriu, PLoS Biol. A238L binds CN via a PxIxIT and LxVP motif
A) Recombinant S-tagged $A238L_{157-239}$ incubated with GST-CNA and CNB co-purifies with GST-CNA (lane 5). Incubation with excess peptides encoding the LxVP site from NFATc1 (lane 1) or the high affinity PxIxIT peptide PVIVIT (lane 3) interferes with A238L-CNA binding. Control peptides (lanes 2 and 4) do not interfere with binding. B) Plot of CN rate as a function of $A238L_{200-239}$ concentration at different RII concentrations ranging from 50-1000 µM. Curve fit obtained by nonlinear regression using the Morrison equation to account for tight binding inhibition. Error bars indicate one s.d. from three independent experiments.

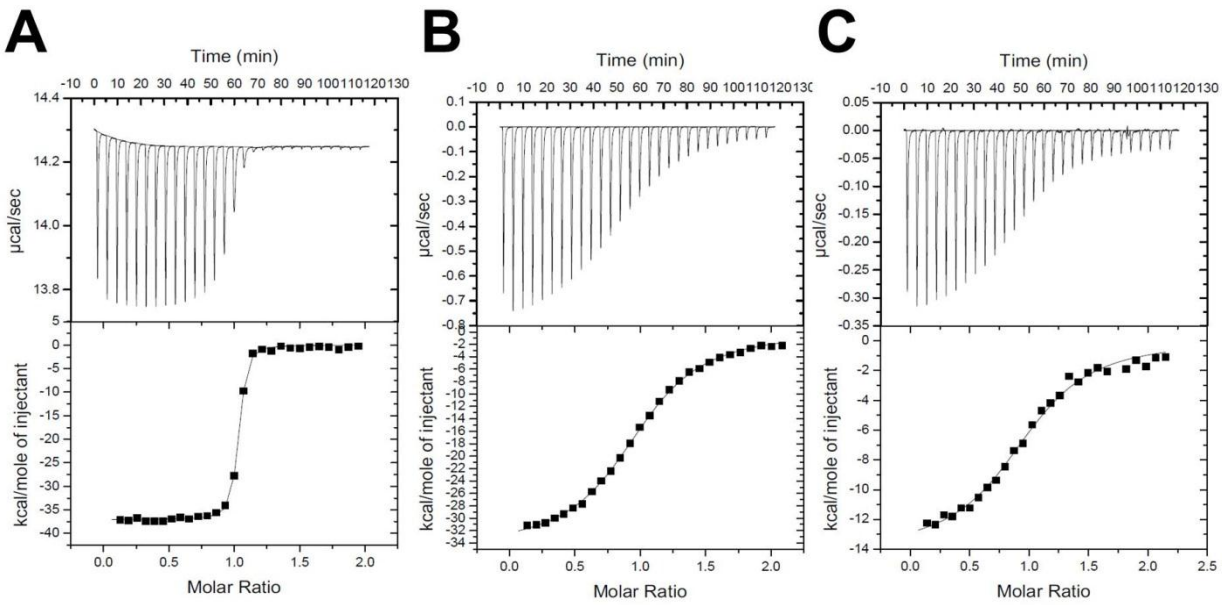


Figure 56: Figure S2, Grigoriu, PLoS Biol. Role of the PxIxIT and LxVP sites in the CN-A238L interaction.

Raw isothermal titration calorimetry data (upper panels) and derived binding isotherm plotted vs. the molar ratio of titrant fit using a one-site model (lower panels) for $CN_{A1\text{-}391/B1\text{-}170}$ titrated with: A) WT A238L, B) A238L PxIxIT mutant (PKIIIT mutated to AKAIAA) and C) A238L LxVP mutant (FLCVK mutated to AACAA). Thermodynamic data and K_D values are summarized in Table 1.

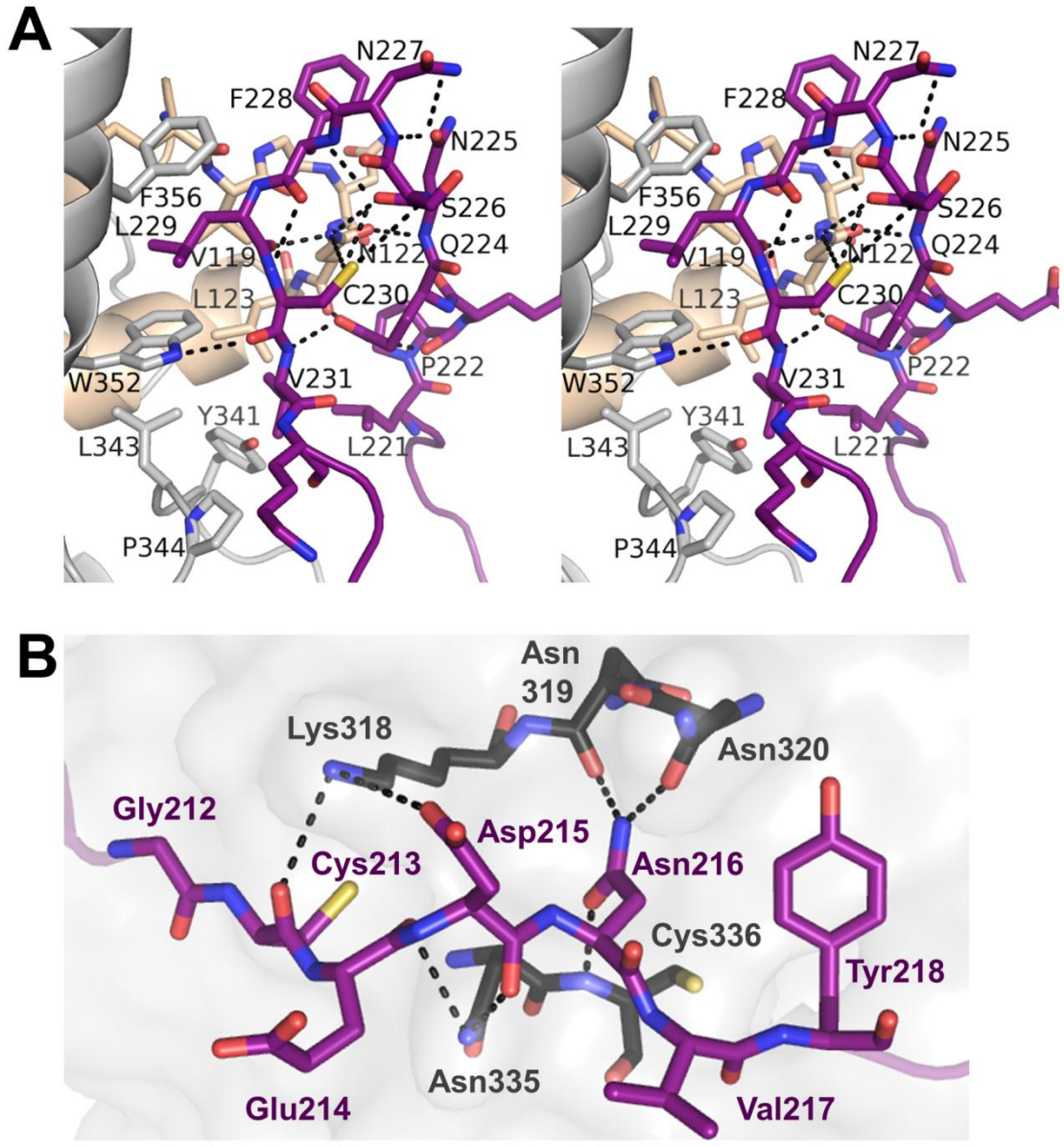


Figure 57: Figure S3, Grigoriu, PLoS Biol. A238L-CN polar interactions.
A) Stereo-view of the FLCVK$_{A238L}$ interface. CN residues participating in the interaction are shown as grey (CNA) or beige (CNB) sticks, with A238L residues shown in magenta. The multiple intra- and intermolecular hydrogen bonds that stabilize the A238L kink are shown as black dotted lines. B) A238L residues immediately C-terminal to the ^{206}PKIIIT^{211} motif, ^{212}GCEDNVY^{218}, are illustrated as sticks and labeled. CNA residues that interact with these A238L residues are also shown as sticks (black). Hydrogen bonds/salt bridge interactions are indicated by black dashed lines.

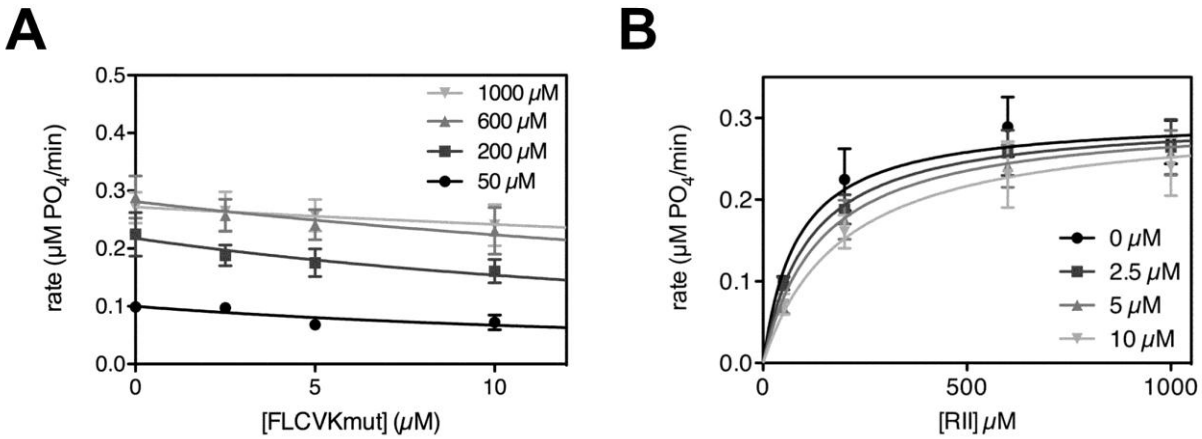


Figure 58: Figure S4, Grigoriu, PLoS Biol. A238L LxVP motif mutant weakly inhibits RII dephosphorylation.

A) Dose-response plot of CN rate as a function of $A238L_{FLCVKmut}$ at different RII concentrations ranging from 50-1000 μM. Curves were fit by nonlinear regression using the Morrison equation. Error bars indicate one s.d. from three independent experiments. B) Plot of CN rate as a function of [RII]. Data fit the Michaelis-Menten model for competitive inhibition. Points represent averages ± s.d. from three independent experiments. Concentrations of $A238L_{FLCVKmut}$ are indicated.

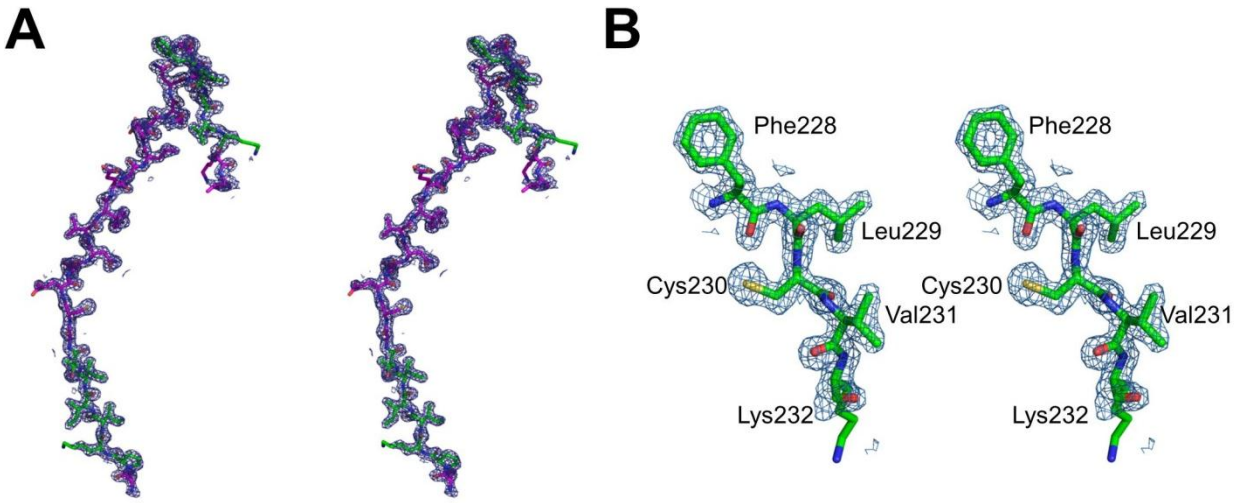


Figure 59: Figure S5, Grigoriu, PLoS Biol. Stereo view of the A238L electron density.
A) Sigma 2mF_o-DF_c electron density map of A238L contoured at 1σ to 1.70 Å (blue mesh). A238L shown as magenta sticks with the PxIxIT and LxVP motifs in green. B) Close-up stereoview of the A238L LxVP motif, with LxVP residues labeled.

A.3. Supplementary tables

Table 2: Table S1. Data collection and refinement statistics

	CN-A238L*
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	72.69, 48.98, 82.44
β (°)	104.4
Resolution (Å)	50.0-1.7 (1.73 – 1.70)**
R_{sym} or R_{merge}	6.8 (29.9)
I / σI	20.4 (3.0)
Completeness (%)	97.4 (87.5)
Redundancy	3.0 (2.3)
Refinement	
Resolution (Å)	47.3 – 1.7
No. reflections	59661
R_{work} / R_{free}	15.8/17.8
No. atoms	
Protein	5025
Ligand/ion	52
Water	470
B-factors	
Protein	20.6
Ligand/ion	36.0
Water	32.6
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.016

*One crystal. **Values in parentheses are for highest-resolution shell.

Appendix B: "Structural and functional analysis of the NLRP4 pyrin domain" and supplementary material

Reproduced from the journal Biochemistry, Volume 51, Number 37, p. 7330-41

I performed all NMR experiments described in this manuscript and in the supplementary information. I also optimized the ASC construct for NMR studies at physiological pH.

Supplementary material is located in **section B.6** of this appendix.

Clarissa Eibl[1], Simina Grigoriu[2], Manuel Hessenberger[1], Julia Wenger[1], Sandra Puehringer[3, 4], Anderson S. Pinheiro[2], Roland N. Wagner[5], Martina Proell[5], John C. Reed[5], Rebecca Page[6], Kay Diederichs[7,*] & Wolfgang Peti[2,8,*]

[1]Department of Molecular Biology, University of Salzburg, 5020 Salzburg, Austria; [2]Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI 02903, USA; [3]Institute for Soft Matter and Functional Materials, Macromolecular Crystallography, Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, 12489, Germany; [4]Department of Biology and Chemistry, Freie Universität Berlin, 14195 Berlin, Germany; [5]Sanford-Burnham Medical Research Institute, La Jolla, CA, 92037, USA; [6]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02903, USA; [7]Department of Biology, University of Konstanz, 78457 Konstanz, Germany and [8]Department Chemistry, Brown University, Providence, RI 02912, USA.

Key words:

NLR proteins, Death Domain, Pyrin Domain, NLRP4, ASC, Innate immune system, inflammasome, apoptosis, X-ray crystallography, protein-protein interaction

Abbreviations:

rmsd, root-mean-square deviation; NLR, Nucleotide-binding domain and leucine-rich repeat containing proteins; LRR, leucine-rich repeats; DD, Death Domain; PYD, pyrin domain; CARD, caspase activation and recruitment domain; ASC, apoptosis-associated speck-like protein containing a CARD; ESP, electrostatic surface potential; NMR, nuclear magnetic resonance.

Acknowledgements

The authors thank Dr. Eva de Alba, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain for sharing the ASC plasmid.

Protein Data Bank accession code

Coordinates and structure factors were deposited in the RCSB Protein Data Bank under accession code 4EWI.

B.1. Abstract

NLRP4 is a member of the nucleotide-binding and leucine-rich repeat receptor (NLR) family of cytosolic receptors and a member of an inflammation signaling cascade. Here, we present the crystal structure of the NLRP4 pyrin domain (PYD) at 2.3 Å resolution. The NLRP4 PYD is a member of the Death Domain (DD) superfamily and adopts a DD fold consisting of six α-helices tightly packed around a hydrophobic core, with a highly charged surface that is typical of PYDs. Importantly, however, we identified several differences between the NLRP4 PYD crystal structure and other PYD structures that are significant enough to affect NLRP4 function and its interactions with binding partners. Notably, the length of helix α3 and the α2-α3 connecting loop in NLRP4 PYD are unique among PYDs. The apoptosis-associated speck-like protein containing a CARD (ASC) is an adaptor protein whose interactions with a number of distinct PYDs is believed to be critical for activation of the inflammatory response. Here, we use co-immunoprecipitation, yeast two-hybrid and NMR chemical shift perturbation analysis to demonstrate that, despite being important for activation of the inflammatory response and sharing several similarities to other known ASC-interacting PYDs (i.e. ASC2), NLRP4 does not interact with the adaptor protein ASC. Thus, we propose that the factors governing homotypic PYD interactions are more complex than the currently accepted model, which states that complementary charged surfaces are the main determinants of PYD-PYD interaction specificity.

B.2. Introduction

Multicellular organisms have evolved two distinct types of immune responses against foreign invaders: the faster innate immune system and the delayed adaptive immune system. Innate immunity forms the first line of defense against invading pathogens and harmful environmental factors by detecting danger- and pathogen-associated molecular patterns (DAMPs and PAMPs) via diverse pattern recognition receptors (PRRs). Whereas the membrane-anchored Toll-like receptors represent the classical PRRs, the cytosolic nucleotide-binding and leucine-rich repeat-containing receptors (NLRs, alternatively named NBD-LRR or CATERPILLER) are essential for the recognition of cytosolic PAMPs not intercepted by membrane-bound receptors (for review see [164-166]).

NLR genes encode large multi-domain proteins with a conserved tripartite domain architecture characterized by: 1) C-terminal leucine-rich repeats, LRRs, necessary for ligand binding; 2) a central nucleotide-binding and oligomerization domain termed the NACHT domain; and 3) an N-terminal effector domain most commonly in the form of either a pyrin domain, PYD, a caspase activation and recruitment domain, CARD, or a baculovirus inhibitor repeat domain, BIR. These effector domains bind downstream signaling molecules, ultimately leading to activation of various protein kinases, transcription factors, proteases, and other components of host defense and inflammatory responses [167-169].

NLRs are divided into three subfamilies according to the identity of their N-terminal effector domains. NLRs containing a PYD (NLRPs) form the largest subfamily, with fourteen members in total (NLRP1-NLRP14) in humans [169].

NLRP1 and NLRP3, the best-characterized NLRPs, activate the inflammatory immune response by detecting cytosolic PAMPs via their C-terminal LRR domains. Once active, NLRP1 and NLRP3 oligomerize via their central NACHT domains. Finally, the N-terminal PYDs recruit the adaptor protein ASC via homotypic PYD-PYD interactions, resulting in recruitment of pro-caspase 1 molecules via homotypic CARD-CARD interactions [167]. The ensuing multi-component signaling platform, termed the inflammasome, is analogous to the well-studied Apaf-1 apoptosome [170]. Inflammasome formation results in the autoproteolytic activation of caspase 1 which, in turn, cleaves proinflammatory cytokines such as pro-IL-1β and pro-IL-18 into their mature forms [167, 168, 171]. Polymorphisms in NLRP1 and NLRP3 are associated with susceptibility to chronic inflammatory disorders, such as Crohn's disease, Muckle Wells syndrome and familial cold autoimmune syndrome, highlighting the importance of these receptors in innate immunity [172, 173]. A detailed understanding of the molecular mechanisms underlying NLR signaling is of great biological and medical importance, being critical for more efficient treatment of these diseases.

Recent reports have identified novel, significantly divergent functions for NLRPs, in addition to their already well-established proinflammatory functions [174]. This is especially the case for NLRP4 (113 kDa; also known as PAN2, Nalp4 or PYPAF4), which is highly expressed in several diverse tissues including placenta, testis, oocytes, spleen, pancreas, lung, liver, kidney and thymus [175-177]. NLRP4 has recently been reported as a negative regulator of autophagy and type I interferon signaling, resulting from the interaction of its NACHT domain

with Beclin1 and TBK1, respectively [175, 178]. Furthermore, NLRP4 was identified as an inhibitor of TNF-α- and IL-1β-mediated NF-κB activation, which is achieved through an interaction with IKKα. The PYD of NLRP4 is necessary for this inhibitory effect on NF-κB, underscoring its importance as a critical regulator of inflammation signaling pathways [175, 176, 178].

Several PYD structures have been determined over the past few years. These structures showed a high degree of conservation of their 3-dimensional organization and surface charge distribution. Thus, the molecular basis for the specificity of their differential interactions remains poorly understood [179-183]. Here, we report the crystal structure of human NLRP4 PYD at 2.3 Å resolution, alongside interaction studies with the adaptor protein ASC, and provide novel insights into the complexity of PYD interactions and the factors dictating their specificity.

B.3. Experimental Procedures

B.3.a. Protein Expression and Purification

The coding region of NLRP4 PYD, (residues Met1–Leu110), was PCR-amplified using human NLRP4 cDNA (GenBank accession No. BC050326) as a DNA template and cloned into the pET28a stop plasmid (Invitrogen), with an N-terminal His_6-tag followed by a thrombin cleavage site. All NLRP4 PYD mutants were generated using the “Round-the-horn” site-directed mutagenesis approach. *E. coli* BL21 Star (DE3) (Invitrogen) cells were transformed with the sequence-verified NLRP4 PYD plasmid (Eurofins MWG Operon, Ebersberg, Germany) and grown in 600 ml LB medium to an OD_{600} of 0.8, at which point protein expression

was induced with 420 μM IPTG at 25°C for 5 h. Harvested cells were resuspended in lysis buffer I (50 mM KH_2PO_4 pH 7.8, 300 mM NaCl, 5 mM imidazole, and 0.1 g/ml lysozyme) and stored at -20°C until lysed by sonication. The lysate containing His_6-tagged NLRP4 PYD was clarified by centrifugation and loaded on 800 μl pre-equilibrated Ni^{2+}-NTA resin (Qiagen) for immobilized metal affinity chromatography. The beads were washed with 15 ml wash buffer I (20 mM Tris pH 8.0, 1 M NaCl, 20 mM imidazole), followed by 15 ml wash buffer II (20 mM Tris pH 8.0, 1 M NaCl, 40 mM imidazole). NLRP4 was eluted by thrombin digestion (Sigma Aldrich). Untagged NLRP4 PYD was further purified using a Superdex 75 10/300 size exclusion column (GE Healthcare) equilibrated in size exclusion buffer I (20 mM Tris pH 8.0, 150 mM NaCl, 5 mM TCEP). Fractions containing NLRP4 PYD were combined and concentrated by evaporation to 1.5 mg/ml. Selenomethionine (SeMet)-labeled protein was expressed in minimal medium (Molecular Dimensions) and purified under identical conditions. ^{15}N-labeled wt and mutant NLRP4 PYD were expressed in M9 minimal medium containing 4 g/l [^{12}C]-D-glucose and 1 g/l $^{15}NH_4Cl$ (Cambridge Isotope Laboratories, Cambridge, MA) as the sole carbon and nitrogen sources, respectively, and purified under identical conditions.

The ASC PYD (residues 1-94), containing the solubility-enhancing L25A mutation, was PCR-amplified from template DNA containing full-length ASC L25A PYD (hereafter referred to as ASC PYD). PCR-amplified ASC PYD was subcloned into the pET28a stop plasmid, containing an N-terminal His_6-tag as well as into the pETRP1B-GB1 bacterial expression vector (in-house), which

includes an N-terminal His_6-affinity tag followed by a TEV (Tobacco Etch Virus) protease cleavage site and a non-cleavable GB1 solubility tag. His_6-tagged ASC PYD was expressed and purified as previously described [183]. For large-scale expression of GB1-ASC L25A PYD (hereafter referred to as GB1-ASC PYD), the sequence-verified (Beckman Genomics) GB1-ASC PYD construct was transformed into BL21-CodonPlus (DE3)-RIL cells (Agilent) and 1 L cultures were grown in LB medium to OD_{600} of 0.7. Cultures were chilled at 4°C for 1 h and protein expression was induced with 1 mM IPTG for 18 h with vigorous shaking (250 rpm) at 18°C. Cell pellets were harvested (6000x*g*, 15 min, 4°C) and cell paste was stored at -80°C. ^{15}N-labeled ASC PYD and GB1-ASC PYD were expressed as described for ^{15}N-labeled NLRP4 PYD. GB1-ASC PYD (^{15}N-labeled and unlabeled) was purified as follows: cell pellet was thawed and resuspended in lysis buffer II (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.1% Triton X-100, EDTA-free protease inhibitor tablets [Roche]) and lysed using high-pressure cell homogenization (Avestin C3 Emulsiflex). Clarified cell lysate (35000x*g*, 50 min, 4°C) was loaded onto a HisTrap HP column (GE Healthcare) equilibrated with His Tag Buffer A (50 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole) and the protein was eluted with a 5-500 mM imidazole gradient over 60 min. Fractions containing the His_6-tagged GB1-ASC PYD, as determined by SDS-PAGE, were pooled, His_6-tagged TEV was added for His_6-tag cleavage, and dialysis was carried out simultaneously against TEV cleavage buffer (50 mM Tris pH 8, 500 mM NaCl) at 4°C for 16 h. His_6-tag cleavage was verified by SDS-PAGE and His_6-tagged TEV, the cleaved His_6-tag, and any

residual uncleaved GB1-ASC PYD were removed via a subtraction purification step over a gravity flow Ni^{2+}-NTA column (Qiagen). Cleaved GB1-ASC PYD was further purified by size exclusion chromatography using a Superdex 75 26/60 column (GE Healthcare) equilibrated in protein buffer II (20 mM sodium phosphate pH 7.2, 100 mM NaCl, 0.5 mM TCEP). The purified protein was concentrated to 100 µM using centrifugal ultrafiltration (Amicon), flash-frozen and stored at -80°C.

B.3.b. Crystallization and Data Collection

Native and SeMet crystals were grown at 20°C using vapor diffusion in sitting drops containing equal volumes of protein solution (1.0 µl purified NLRP4 PYD at 1.5 mg/ml in concentrated size exclusion buffer I) and precipitant solution (0.1 µl of 0.2 M Disodium tartrate, 2.2 M Ammonium sulfate). Crystals formed after four days and were harvested, cryoprotected in paraffin oil, then flash-frozen and stored in liquid nitrogen. Diffraction data were collected at beamline BL14.1 at the electron storage ring BESSY II (Berlin-Adlershof, Germany) operated by the Helmholtz-Zentrum Berlin (HZB) [184] and processed using XDS [185]. One SeMet and one native crystal were used to determine the structure.

B.3.c. Phasing and Refinement

Data were phased in SHELX [186, 187] using ten sites from the 3.2 Å SeMet dataset yielded an interpretable electron density map and the model refined against a 2.3 Å native dataset. Iterative cycles of refinement and model building were carried out in Coot [109] and Phenix [108]. The final model was refined to

an R/R_{free} of 0.207/0.245 at 2.28 Å resolution (R_{free} was calculated using 5% of reflections randomly omitted from the refinement). Data collection, model and refinement statistics are summarized in Table 1.

Table 3: Table 1, Eibl, Biochemistry. Summary of Crystal Parameters, Data Collection and Refinement Statistics

Data collection				
Space group	$P3_221$			
Unit cell parameters	a= 62.1Å, b= 62.1Å, c= 124.9Å , α=β= 90°, γ=120°			
	λ1 MADSe	**λ2 MADSe**	**λ3 MADSe**	**λ4 native**
Wavelength [Å]	0.97958	0.97973	0.9719	0.9184
Resolution [Å]	32.82–2.82	32.86-2.76	32.89–2.88	40.76-2.28
Highest resolution shell [Å]	2.89–2.82	2.83–2.76	2.96–2.88	2.42-2.28
Mosaicity [°]	0.241	0.235	0.241	0.136
[a]Number of reflections	67335 (3276)	76639 (5174)	69396 (5428)	102943 (16562)
[a]Unique Reflections	[b]12689 (938)	[b]13680 (1076)	[b]12091 (930)	13300 (2097)
[a]Completeness [%]	99.4 (99.6)	99.7 (100)	99.7 (100)	99.9 (99.9)
[a]Multiplicity	5.3 (3.5)	5.6 (4.8)	5.7 (5.8)	9.9 (5.2)
[a]Mean I/σ(I)	18.0 (2.1)	18.6 (1.8)	20.2 (2.2)	20.8 (3.0)
[a,c]R_{meas}[%]	6.5 (65.6)	6.5 (91.4)	6.4 (77.3)	6.3 (87.9)
Model and refinement statistics				
Resolution range [Å]	40.76-2.28			
[d]Number of reflections (total)	13245			
Number of reflections (test)	623			
Completeness (% total)	99.6			
[e]R_{cryst}	0.207			
[f]R_{free}	0.246			
Stereochemical Parameters				
Bond length	0.015 Å			
Bond angle	1.626°			
Wilson- B-value [$Å^2$]	48.72			
[g]ESU based on [f]R_{free} value	0.25 Å			
Protein residues/atoms	191/1610			
Solvent molecules	40			
Ligand molecules	3 (2x Cl^-, 1x SO_4)			

λ1-λ3 were processed by XDS (version May 1, 2010); λ4 was processed by XDS (version March 15, 2012)

[a]Values in parentheses are for the highest resolution bin.

[b]members of a Friedel pair were counted separately

[c]Rmeas = Σ[N / (N-1]1/2 Σ|Ii(hkl) - I(hkl)| ΣΣIi(hkl) where Ii is the scaled intensity of the ith measurement and |Ii| is the mean intensity for that reflection (62).

[d]Typically, the number of unique reflections used in refinement is slightly less than the total number that were integrated and scaled. Reflections are excluded due to systematic absences, negative intensities, and rounding errors in the resolution limits and cell parameters.

[e]Rcryst = Σ||Fobs| - |Fcalc||/Σ| Fobs where Fcalc and Fobs are the calculated and observed structure factor amplitudes, respectively.

[f]Rfree = as for Rcryst, but for 5.0% of the total reflections chosen at random and omitted from refinement.

[g]ESU = Estimated overall coordinate error (63, 64)

B.3.d. NMR spectroscopy

The interaction between NLRP4 PYD (residues 1-110) and ASC PYD (residues 1-94) was tested by NMR titration experiments. 2D [^{1}H,^{15}N] HSQC spectra of ^{15}N NLRP4 PYD were acquired at 298 K on a Bruker AvanceII 500 MHz spectrometer equipped with a TCI HCN *z*-gradient cryoprobe and processed with Topspin 2.1 (Bruker). Data was analyzed using Cara (http://cara.nmr.ch). 10% D_2O was added to all samples. An reference 2D [^{1}H,^{15}N] HSQC spectrum of ^{15}N-labeled NLRP4 PYD was recorded at a concentration of 50 μM (20 mM sodium phosphate pH 7.2, 100 mM NaCl, 0.5 mM TCEP). Identical conditions were used for the three electrostatic surface NLRP4 PYD mutants. Titration experiments were performed with a 2-fold or 3-fold molar excess of unlabeled GB1-ASC L25A PYD (25 μM ^{15}N NLRP4 PYD:50 μM GB1-ASC-PYD or 16 μM ^{15}N NLRP4 PYD:48 μM GB1-ASC PYD, respectively).

B.3.e. Bioinformatics Analysis

The sequence alignment of NLRP4 PYD with the PYDs of NLRP1, 3, 7, 12, ASC2 and ASC was generated in Multalin [188] and secondary structure comparison was completed using the online servers Dali [189] and TopMatch-web [190-192]. Protein-protein interactions between the two monomers making up the asymmetric unit were evaluated using ProtorP [193]. Structure visualization and analysis were performed in Pymol [194]. Electrostatic surface potentials were calculated using APBS [195] as part of the PIPSA (Protein Interaction Property Similarity Analysis) server [196] to generate quantitative comparisons between PYD electrostatic surface potentials.

B.3.f. Co-immunoprecipitation

HEK293FT cells were seeded at 60% confluency in 6-well tissue culture plates (Greiner). The following day, cells were co-transfected with 2 μg of plasmids encoding full-length Myc-tagged ASC and 1 of 3 truncated Flag-tagged NLRP4 constructs: NLRP4 PYD (residues 1-110), NLRP4 PYD+Linker (residues 1-162) and NLRP4 ΔLRR (residues 1-324) using Transfectin (BioRad) according to manufacturer's protocol. Flag-tagged Aim2 PYD (residues1-96) was co-transfected as positive control. On the day following transfection, cells were harvested and lysed in 200 μl of IP lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.5% Nonidet P-40) supplemented with protease inhibitor (Sigma Aldrich). After centrifugation, clarified lysate from each sample was subjected to overnight immunoprecipitation at 4°C with anti-Myc antibodies (AbD SeroTec) immobilized on G-Sepharose (Invitrogen). The immunoprecipitated complexes were washed four times with IP lysis buffer and, along with a 5% total cell extract control, were resolved on a 12% SDS-PAGE, transferred to nitrocellulose membranes (Whatman), and analyzed by immunoblotting using anti-Myc (AbD SeroTec) or anti-Flag (Sigma) primary antibodies and AP-labeled secondary antibodies (Santa Cruz Biotechnology). Nitro-blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (Roche) were used to augment the signal for protein detection.

B.3.g. Yeast two-hybrid analysis

The two-hybrid experiments were performed with the MATCHMAKER GAL4 Two-Hybrid System 3 (Clonetech) following the manufacturer's instructions. In

summary, NLRP4 PYD, NLRP4 PYD+Linker and ASC PYD were each inserted into the pGBKT7 and pGADT7 fusion vectors encoding a reporter GAL-4 DNA-Binding Domain (BD) and an Activation Domain (AD), respectively, whose proximity is necessary for producing a signal. The *Saccharomyces cerevisiae* reporter strain Y2HGold (Clonetech) was co-transformed with either pGADT7-NLRP4 PYD (bait protein) or pGADT7-NLRP4 PYD+Linker and pGBKT7-ASC PYD (library protein) or pGADT7-ASC PYD (bait protein) and pGBKT7-NLRP4 PYD or pGBKT7-NLRP4 PYD+Linker (library protein) using a small-scale lithium acetate/single-stranded carrier DNA/polyethylene glycol (LiAc/ss-DNA/PEG) transformation protocol [197]. Transformed yeast cells were resuspended in sterile water and spotted onto SD/-Leu/-Trp dropout medium to assess transformation efficiency and onto SD/-Ade/-His/-Leu/-Trp selection medium to test for potential interactions. Plates were incubated for 4-5 days at 30 °C. To rule out self-activation of reporter genes, all fusion constructs were tested for their ability to activate transcription of the GAL-4 reporter or to associate with either the GAL-4 BD or GAL-4 AD. Co-transformation of pGBKT7-ASC PYD and pGADT7-Aim2 PYD (residue 1-96) served as positive interaction control [198].

B.4. Results

B.4.a. Bioinformatics analysis of NLRP4 PYD

The human genome contains 14 NLRP proteins (NLRP1-NLRP14; HUGO Gene Nomenclature Committee [HGNC, www.genenames.org]). Structural information is available for the PYDs of human NLRP1, NLRP3, NLRP7, NLRP12, ASC and ASC2, and murine NLRP10 (PDB entry 2DO9) [179-183, 199]. A sequence

homology search using the FFAS server [200] shows high conservation between NLRPs, with the highest sequence identity (50%) observed between NLRP4 PYD and NLRP14 PYD (NCBI accession: NP_789792.1). In addition, NLRP4 PYD shares 40% sequence identity with NLRP12 PYD (PDB entry 2L6A) [182], 34% with NLRP3 PYD (PDB entry 3QF2) [179], 33% with NLRP1 PYD (PDB entry 1PN5) [180] and 33% with NLRP7 PYD (PDB entry 2KM6) [183] (Table 2). Furthermore, NLRP4 PYD shares 22% sequence identity with the adaptor protein ASC PYD (PDB entry 1UCP) [181] and 23% with ASC2 (also known as POP1) (PDB entry 2HM2) [199], the dominant-negative modulator of ASC [201].

Table 4: Table 2, Eibl, Biochemistry. Summary of structural and electrostatic similarity statistics

PYDs	PDB entry	[a]Z-score	[b,c]RMSD (Å)	[b,d]Is (%)	[b,e]L	[b,f]Qc (%)	[g]ED	References
NLRP1	1PN5	10.1	1.78	33	75	82	1.60	[180]
NLRP3	3QF2_A	12.9	1.79	34	83	91	1.44	[179]
NLRP7	2KM6	10.5	1.87	33	80	88	1.64	[183]
NLRP12	2L6A	10.7	1.82	40	81	89	1.52	[182]
ASC	1UCP	11.5	2.11	22	83	91	1.12	[181]
ASC2	2HM2	12.0	1.86	23	81	89	1.02	[199]

[a]Structural similarity search with the structure of NLRP4 PYD using the DALI server[189]
[b]Values were calculated by TopMatch-web [190, 191].
[c]Values for **RMSD** were calculated by using all structurally equivalent Cα atoms.
[d]**Is** gives the sequence identity of query and target in equivalent regions.
[e]**L** (Length of alignment) gives the number of residue pairs that are structurally equivalent.
[f]**Qc** represents the relative cover of the query with respect to the target. Qc= 100 x L/Q_n, where Q_n is the length of the query sequence)
[e]**ED** (electrostatic distance) calculated by PIPSA [196].

B.4.b. NLRP4 PYD is a monomer in solution

The NLRP4 PYD (residues Met1-Leu110) was expressed in *Escherichia coli* and purified to homogeneity. As verified by size exclusion chromatography, NLRP4 PYD exists as a monomer in solution under highly reducing conditions (5

mM TCEP). To further characterize NLRP4 PYD multimerization in solution, we placed ^{15}N-labeled NLRP4 PYD in a low-reducing buffer (0.5 mM TCEP; NLRP4 PYD has one cysteine, Cys85), waited several days to allow for any dimerization to occur, and recorded a 2D [^{1}H,^{15}N] HSQC NMR spectrum. The outstanding quality of the spectrum (Supporting Figure S1) is typical for a small 13.9 kDa protein, indicating that ^{15}N-labeled NLRP4 PYD exists as a monomer after several days in low reducing conditions.

NLRP4 PYD in NMR buffer (20 mM sodium phosphate pH 7.2, 100 mM NaCl, 0.5 mM TCEP) cannot be concentrated using Amicon Ultra centrifugal filters (Millipore), as it readily precipitates. This phenomenon has been observed in other NLR effector domains (NLRP1 [180], NLRC5 (unpublished data)) and is a common feature of this subfamily. We found concentration by evaporation, which increases the concentration of all sample components (i.e., protein, salt, TCEP), to be more efficient, allowing NLRP4 PYD to reach a concentration of 3 mg/ml. Crystallization was carried out using the sitting drop vapor diffusion method. Crystals appeared within four days, diffracted to 2.3 Å and indexed in space group $P3_221$, with two molecules in the asymmetric unit. Due to heavy twinning (45%) (phenix.xtriage) [108], it was impossible to determine the structure using the initial SeMet datasets. However, several rounds of crystallization optimization combined with diffraction screening of 50 SeMet and 15 native crystals allowed one SeMet and one native crystal with less than 3% twinning to be identified that yielded quality data suitable for structure determination.

B.4.c. Crystal structure of NLRP4 PYD

The final model includes two NLRP4 PYD monomers (residues Phe5-Tyr95 in chain A, Phe5-Gly94 in chain B), one sulfate ion, two chloride ions and 40 water molecules in the asymmetric unit. No electron density was observed for the N- and C-termini (residues Met1-Ser4 and Thr96-Leu110 in chain A; Met1-Ser4 and Tyr95-Leu110 in chain B). The Matthews coefficient (V_m)[20] for NLRP4 PYD is 2.50 $Å^3$/Da and the estimated solvent content is 51%. The Ramachandran plot shows that 99.4% of the residues are in the allowed regions.

The crystal structure of NLRP4 PYD adopts a typical death domain fold (Figure 1). Six α-helices (α1-α6) are tightly packed around a central hydrophobic core (formed by Leu11, Tyr14, Leu15, Leu18, Phe23, Phe26, Leu30, Leu55, Leu59, Trp69, Ile71, Leu73, Ile75, Phe76, Met79, Leu84 and Val88; Figure 1A,C,D) and form an anti-parallel α-helical bundle. This well-defined hydrophobic core, which defines the death domain fold, is conserved throughout the entire DD superfamily [182, 202]. The 6 ordered α-helices of NLRP4 PYD are connected by loops ranging from 1 to 5 amino acids in length (Figure 1B,C). The loops connecting helices α1 (Gly10-Glu16) to α2 (Lys20-Gln36), α3 (Trp44-Lys49) to α4 (Arg52-Tyr63) and α5 (Glu65-Lys78) to α6 (Lys82-Thr93) are each three amino acids in length, while the α2-α3 loop is composed of five amino acids that are disordered. Helices α4 and α5 are linked by only a single residue, Glu64, making this linker structurally constrained, which accounts for its outlier position in the Ramachandran plot.

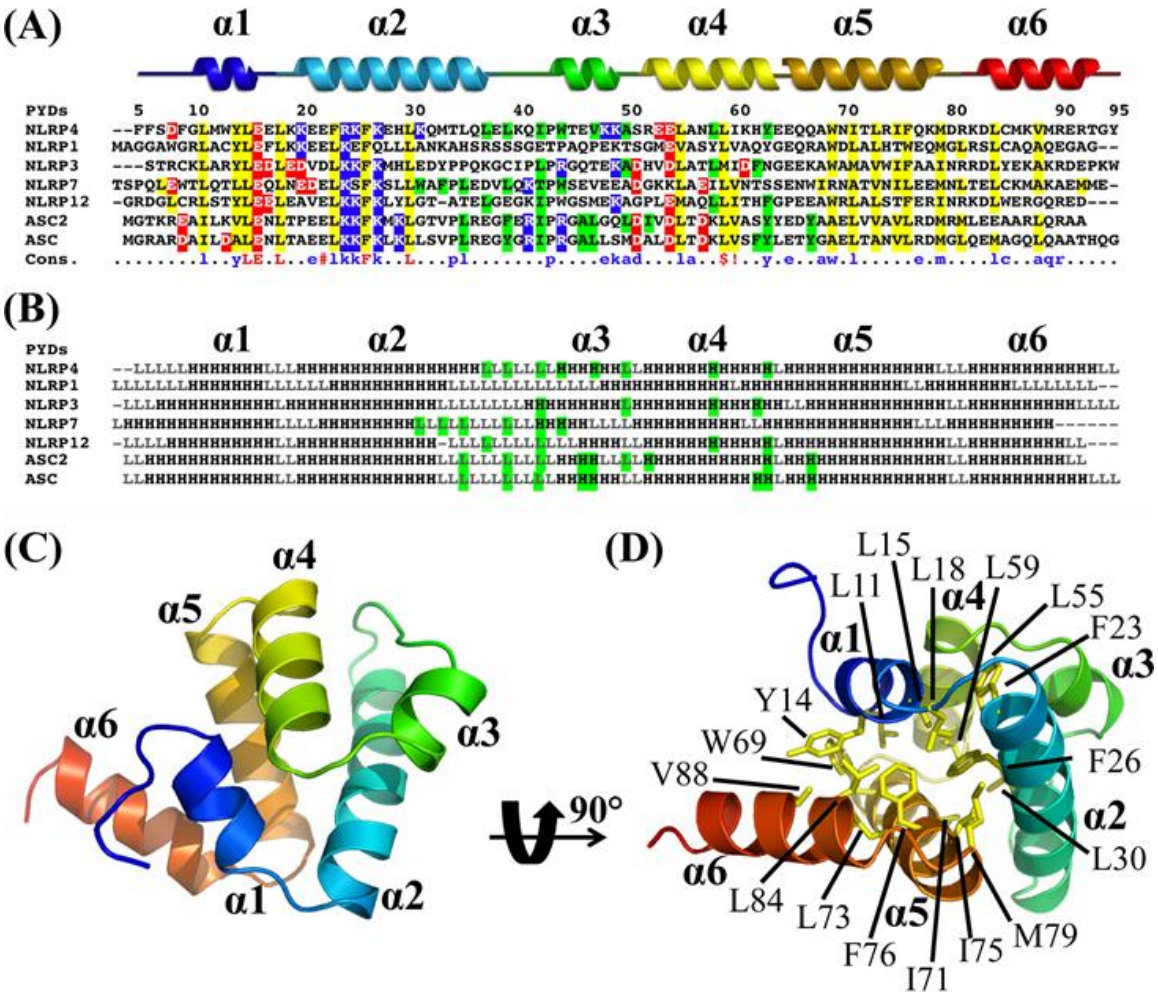


Figure 60: Figure 1, Eibl, Biochemistry. Structural characterization of the NLRP4 PYD.
(A) Primary sequence alignment of the NLRP4 PYD and the PYDs of NLRP1, NLRP3, NLRP7, NLRP12, ASC2, and ASC. The six NLRP4 PYD α-helices, illustrated above the sequence, are color-coded from the N-terminus (blue) to the C-terminus (red) and labeled from α1 to α6. Hydrophobic core-forming residues are highlighted in yellow; residues forming the second hydrophobic cluster to stabilize helix α3 are highlighted in green. Positively and negatively charged residues responsible for PYD surface charges are colored blue and red, respectively. Consensus indicates residues that are >90% identical (red) and 50% identical (blue) (! is either Ile or Val; $ is either Leu or Met; # is either Asp or Glu). (B) Comparison of the secondary structural elements of the NLRP4 PYD and the PYDs of NLRP1, NLRP3, NLRP7, NLRP12, ASC2, and ASC (H, helix; L, loop). Helices are labeled from α1 to α6. (C) Cartoon representation of the NLRP4 PYD with the six α-helices colored from the N-terminus (blue) to the C-terminus (red) and labeled from α1 to α6. (D) The NLRP4 PYD (rotated by 90° along the *x*-axis relative to the view in panel (C) is stabilized by an extensive (hydrophobic core residues are labeled and shown as yellow sticks).

B.4.d. Homo-dimeric interface at the NLRP4 PYD crystallographic dimer

Two NLRP4 PYD monomers (chains A and B) form a dimer in the asymmetric unit, which are related by a pseudo two-fold axis at the dimerization interface. The interface buries 2252 $Å^2$ of solvent accessible surface area SASA, which corresponds to 17.83% burial of the SASA of each monomer [193] and is in the range of biologically relevant protein-protein interfaces. The NLRP4 dimer interface is comprised of both electrostatic and hydrophobic interactions. Contacts are formed by the sidechains of Phe9_A (chain A) with Leu73_B (chain B), Ile60_A with Ile60_B, Q66_A with the main-chain of Phe9_B, and Lys87_A with the main-chain of E91_B (Supporting Figure S2). Due to the pseudo two-fold axis, the reverse interactions (i.e. Phe9_B with Leu73_A, etc) are also present.

While a dimer was observed in the crystal, NMR data shows that NLRP4 PYD is monomeric in solution. Interestingly, a similar behavior has been observed for the CARD of NLRC1 (also referred as CARD1 or Nod1) [203]. Specifically, two NMR structures clearly show that NLRC1 CARD forms a monomer in solution ([204] and PDB entry 2DBD), while two crystal structures [203, 205, 206] show that NLRC1 CARD forms disulfide-linked dimers. Similarly, the NLRP3 PYD crystal structure shows a disulfide bond between Cys8 in helix α1 and Cys108 located in the linker region connecting the PYD with the C-terminal NACHT domain [179]. Interestingly, NLRP4 PYD dimerization does not appear to be the result of a disulfide bond between Cys85 from each monomer, as Cys85 is distal from the dimerization interface, with the corresponding sulfur atoms more than 26 Å apart. Taken together, our data and other available PYD

and CARD data suggest that dimer formation (either via disulfide bonds, or, in our case, other protein-protein interactions) is induced by crystal packing.

However, PYD or CARD dimerization may still be necessary for some aspects of PYD or CARD function. Homotypic interactions between PYDs (in the case of inflammasomes) and CARDs (in apoptosomes and some inflammasomes) are complex, and even dynamic in nature. The ability of PYDs and CARDs to easily transition between the monomeric and dimeric states might be important for their function, such as stabilization of the inflammasome and apoptosome signaling platforms, respectively, thereby improving the efficacy of downstream signaling [207]. It should be noted that Cys85 in helix α6 of NLRP4 PYD is conserved in NLRP1, 7, 8, 11, 13 and 14 and, although no cysteine-mutation with a discernible phenotype has been reported so far, future studies are necessary to investigate the physiological relevance, if any, of disulfide linked PYDs.

B.4.e. Variation among NLR PYDs

The death domain superfamily, which includes the death domain (DD), the death effector domain (DED), the CARD, and the PYD subfamilies is characterized by six antiparallel α-helices packed in a Greek key topology [208]. Recently, we used the structures of the PYDs of NLRP 7 and 12 to identify the key conserved residues necessary for forming the conserved hydrophobic core of DD family members [182]. Based on the aforementioned structural comparisons, we expected the hydrophobic core of NLRP4 PYD to be formed by Leu11, Tyr14, Leu15, Leu18, Phe23, Phe26, Leu30, Leu55, Leu59, Trp69, Leu73, Phe76,

Met79 and Leu84, which are in perfect agreement with the core residues seen in our crystal structure (yellow highlighted residues in Figure 1A) [182].

NLRP4 PYD was superimposed onto the structures of several PYDs using the TopMatch-web server [190-192] (Figure 2), which revealed high levels of structural similarity between all analyzed PYDs (pairwise RMSDs of ~1.8 Å; Table 2). However, NLRP4 PYD, as well as NLRP1 PYD, showed relatively subtle but significant differences from the other PYDs. Namely, both NLRP4 PYD and NLRP1 PYD possess the shortest α1 helix as well as the longest α2 and α6 helices (Figure 1B, 2A). Secondary structure analysis also revealed many similarities among the investigated PYDs, with NLRP4 PYD being most similar to NLRP3 PYD, followed by NLRP12 PYD, NLRP7 PYD and NLRP1 PYD in decreasing order. Given the high conservation of tertiary and secondary structural elements of PYDs, their evolutionary relationship was investigated using TraceSuite II [209]. The phylogenetic tree is shown in Supporting Figure S3 and assigns two large branches. One branch consists of the adaptor ASC and its inhibitor ASC2, while the second branch groups the receptor PYDs of NLRP1, 3, 4, 7 and 12.

The most striking difference seen in NLRP4 PYD, when compared to all other PYDs, is the length of helix α3 and the α2-α3 connecting loop (Figure 2A,B). In NLRP4 PYD, helix α3 is composed of two turns and is stabilized by Leu37, Leu39 and Ile42 from the α2-α3 connecting loop, Trp44 and

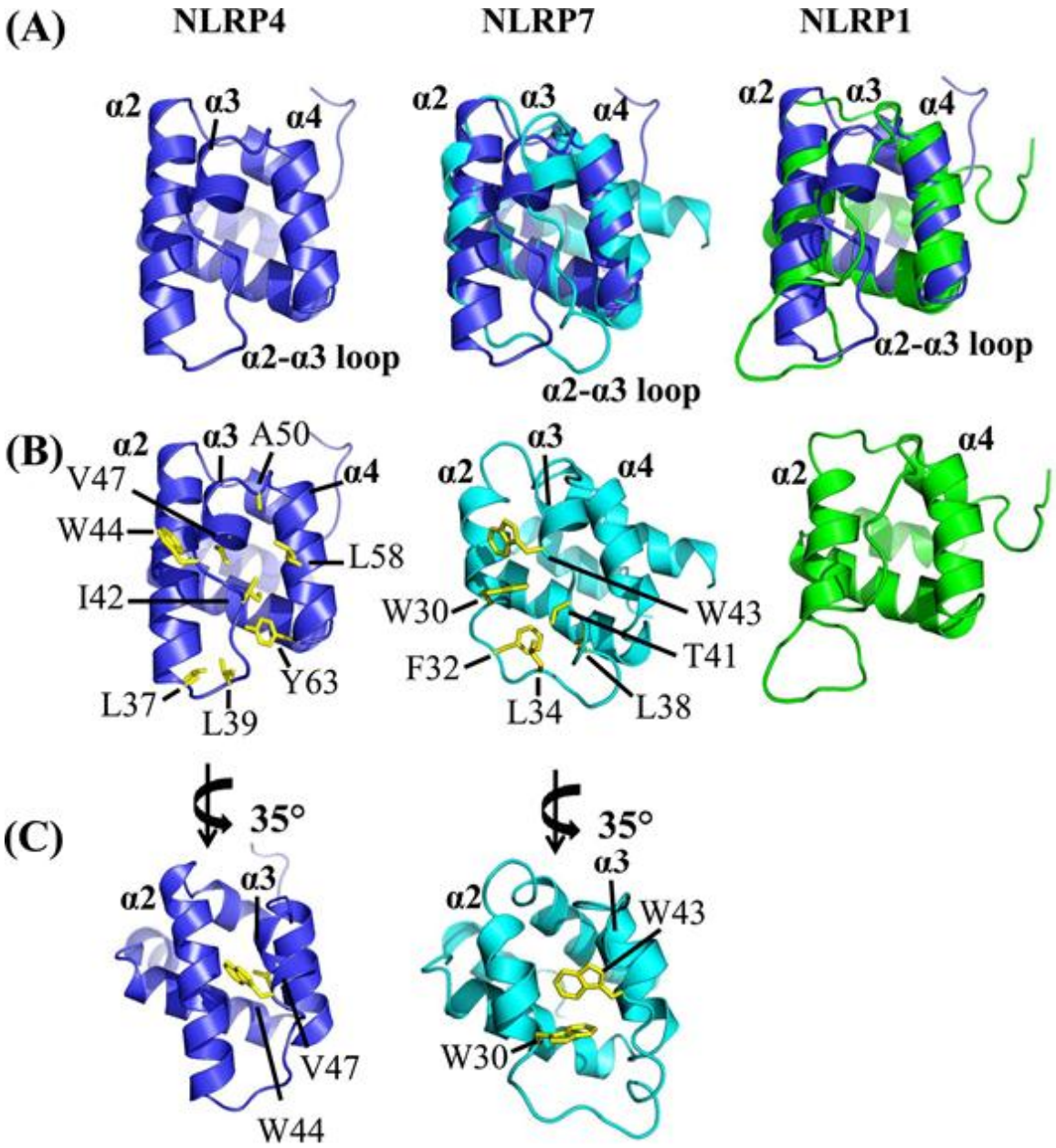


Figure 61: Figure 2, Eibl, Biochemistry. Structural comparison of the NLRP4 PYD with the PYDs of NLRP7 and NLRP1.

(A) Superposition of the NLRP4 PYD (blue) on the NLRP7 PYD (cyan) and the NLRP1 PYD (green). (B) Residues that stabilize helix α3 in the NLRP4 PYD and the NLRP7 PYD are labeled and shown as yellow sticks. Corresponding hydrophobic residues are missing in the NLRP1 PYD. (C) Trp44 in the NLRP4 PYD is structurally conserved in the NLRP7 PYD.

Val47 from helix α3, Ala50 from the α3-α4 connecting loop, and Leu58 and Tyr63 form helix α4 (Figure 1A,B, 2B). This hydrophobic cluster, which is centered on helix α3 and the α2-α3 connecting loop, is conserved in NLRP3 PYD, NLRP7 PYD and ASC PYD and, to a lesser extent, in ASC2 (where helix α3 is only a single turn helix). In contrast, NLRP1 PYD completely lacks hydrophobic residues in α-helix3 and the α2-α3 connecting loop that are required for stabilization of these structural elements. Instead, NLRP1 PYD forms a long unstructured linker between α-helices2 and 4 [180]. Lastly, residue Trp44 in NLRP4 PYD is structurally conserved in NLRP7 and NLRP12 PYDs. Both Trp43 (NLRP7 PYD) and Trp44 (NLRP4 PYD) form a hydrophobic stacking interaction which stabilizes helix α3 (Figure 2C). In contrast, Trp45 in NLRP12 PYD is surface exposed and mediates a protein-protein interaction with FAF-1 UBA domain [182].

The difference in helix α3 between the various PYD structures is important, as it has been speculated that a disordered helix α3 that only becomes fully folded upon binding to downstream partners is a common feature in PYDs and thus was considered to have a crucial role in the protein-protein interactions required for inflammasome signaling [180, 183, 210]. In fact, this highly variable region, along with the variable α2-α3 connecting loop, was shown to play a key role in guiding homotypic, as well as heterotypic PYD binding interactions [182, 211]. While it is extremely likely that all other PYDs with unknown structures contain the well-established hydrophobic core formed by helices α1, α2, α4, α5 and α6, we can also expect to see increasing diversity in the conformations of

helix α3 and the α2-α3 connecting loop, like those seen in NLRP4 PYD, as more PYD structures are determined.

B.4.f. Comparison of the PYD family reveals distinct electrostatic surfaces

It has been well established that oppositely charged electrostatic surface potentials (ESPs) have a vital role in the formation of homotypic death domain interactions. Recently reported structures of the PIDDosome, the FAS-FADD complex and the Apaf-1-Caspase-9 complex highlighted the importance of charged surfaces in homotypic DD and CARD interactions [212-214]. A similar electrostatic interaction mechanism was described for the homotypic PYD interaction of ASC PYD and ASC2 PYD. Specifically, a large positively charged surface on ASC2 PYD (formed by Lys 21, Lys22, Lys26 and Arg41 from helices α2 and α3) interacts with a complementary negatively charged surface on ASC PYD (formed by residues Asp6, Asp10, Glu13, Asp48, Asp51, and Asp54 from helices α1 and α4; blue and red highlighted residues in Figure 1A and Figure 3) [211].

Thus, we analyzed the ESP of NLRP4 PYD, which showed that NLRP4 PYD also possesses two distinctly charged surfaces. NLRP4 PYD has a prominent positively charged region spanning helices α2 and α3, as well as a weaker negatively charged surface consisting of Asp8, Glu16 from helix α1 and Glu53 and Glu54 from helix α4, which is especially similar to ASC2. Figure 3 illustrates the electrostatic surfaces of NLRP4 PYD and the PYDs of NLRP1, 3, 7 and 12 as well as ASC and ASC2. Electrostatic similarity calculations performed by the PIPSA server [196] showed that NLRP4 PYD is most similar to ASC2 PYD

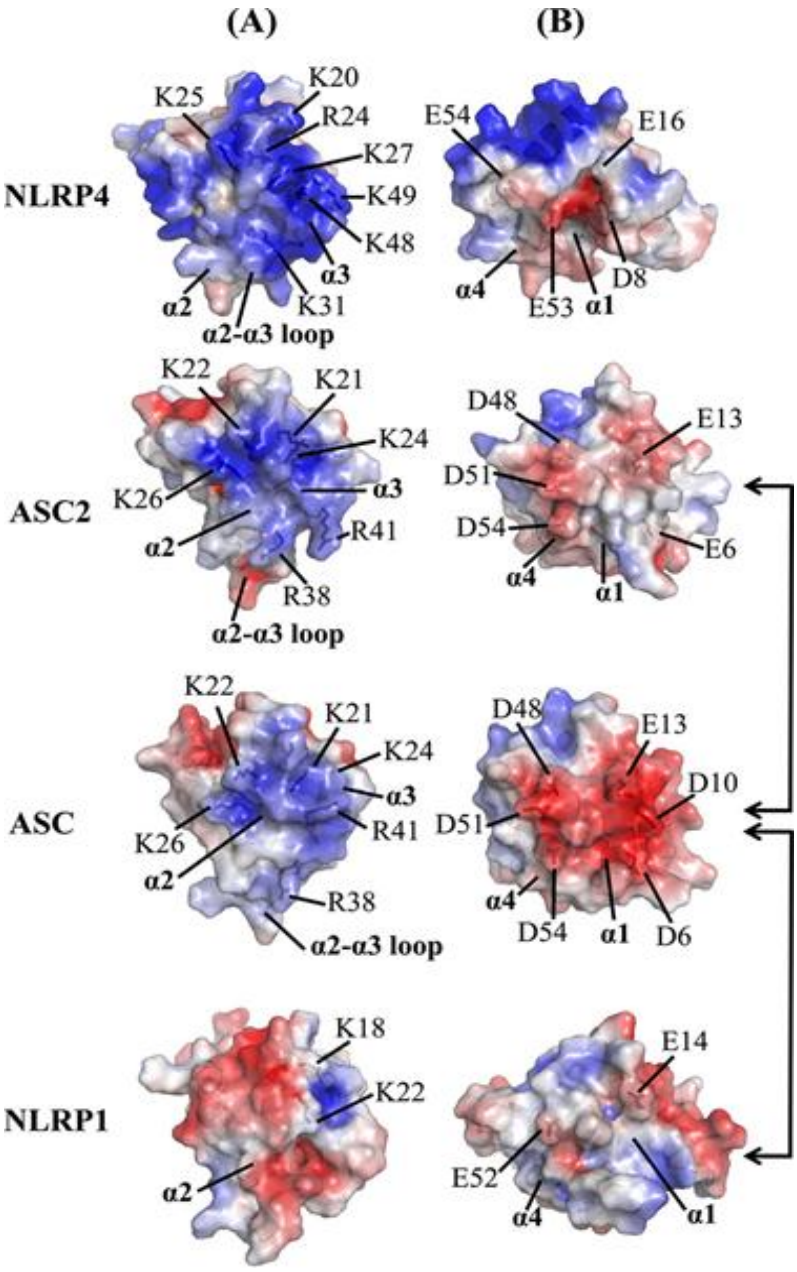


Figure 62: Figure 3, Eibl, Biochemistry. Electrostatic surface potentials of the NLRP4 PYD compared to ASC2 and the PYDs of ASC and NLRP1.

The electrostatic potentials (ESPs) of the two interaction surfaces formed by helices α2 and α3 and helices α1 and α4 are mapped on the solvent-accessible surface of the identically oriented PYDs from NLRP4, NLRP1, ASC, and ASC2. The color scale varies from red (negative ESP) to blue (positive ESP). Residues responsible for these charges are labeled on each structure. Arrows indicate previously experimentally detected direct homtoypic PYD interactions.

(electrostatic distance 1.02; a distance of 0 indicates the electrostatic surface potentials are identical while a score of 2 indicates they are completely opposite) followed by ASC PYD (electrostatic distance 1.12), NLRP3 PYD (electrostatic distance 1.44), NLRP12 PYD (electrostatic distance 1.52), NLRP1 PYD (electrostatic distance 1.60) and NLRP7 PYD (electrostatic distance 1.64).

The significant electrostatic similarity of NLRP4 and ASC PYD is based on a highly positively charged ESP encompassing Lys20, Arg24, Lys25, Lys27 and Lys31 on helix α2 and Lys48 and Lys49 on helix α3 (Figure 3). Currently, no other NLRP PYD has been show to possess a similarly extensive positively charged surface. In other NLRP PYD structures, this surface is either mostly negatively charged (NLRP1 PYD), slightly positively charged with a prominent hydrophobic patch (NLRP12 PYD), or slightly negatively charged with a significant hydrophobic patch (NLRP7 and NLRP3 PYD).

The inflammasome receptors NLRP1 and NLRP3 were shown to interact with ASC, leading to the formation of the inflammasome signaling platform [182, 183, 215, 216]. Interestingly, the ESPs of NLRP1 and NLRP3 are different, highlighting that the electrostatic nature of these protein-protein interactions might have been over-emphasized. Specifically, NLRP3 PYD is mainly negatively charged, and thus was proposed to interact with the positively charged surface of ASC PYD, while not competing with ASC2 for the negatively charge surface of ASC PYD [211]. The electrostatic surfaces of NLRP1 PYD also differ from those of NLRP3 or ASC2 PYD. Taken together, this indicates that homotypic PYD:PYD interactions, which were previously thought to be due to the attraction between

oppositely charged surfaces, are likely much more complex. This is further confirmed by our observation that both NLRP7 PYD and NLRP12 PYD have negatively charged surfaces that are comparable to NLRP3, but, unlike NLRP3, they are unable to interact with ASC [182, 183].

B.4.g. NLRP4 PYD does not interact directly with ASC PYD

Interestingly, NLRP4 PYD has a positively charged surface that is very similar to that of ASC2. In ASC2, this positively charged surface is responsible for its interaction with ASC PYD (Figure 3) [211]. Furthermore, NLRP4 was shown to form heterotypic oligomers with the inflammasome-forming receptors NLRP3 and NLRC4 via NACHT domain interactions [178, 217]. This raises the question of whether NLRP4 PYD also interacts with ASC PYD, thus leading to caspase 1 and IL-1β activation.

Based on these observations, we tested for an interaction between NLRP4 PYD and ASC PYD via NMR chemical shift perturbation (CSP) analysis. Specifically, we collected a 2D heteronuclear [^{1}H,^{15}N] HSQC spectrum of ^{15}N-labeled NLRP4 PYD and then added unlabeled ASC PYD in a 2-fold and 3-fold molar excess to the NLRP4 PYD sample (the limited solubility of these proteins did not allow for higher ratios to be tested). No CSPs were detected in the titration experiments, indicating that there is no direct interaction between the PYDs of NLRP4 and ASC (Supporting Figure S4). It should be noted that NMR titration experiments were conducted at pH 7.2 to more closely resemble the cellular environment, as previous reports indicated that death domain interactions are affected by low pH [213] (previous *in vitro* interaction experiments involving

ASC PYD were conducted at pH 4.5, due to the insolubility of ASC PYD at physiological pH). As recombinant ASC PYD is expressed insolubly in bacterial cells, purification requires denaturation of inclusion bodies and protein refolding. Using a GB1-solubility tag, we succeeded in expressing and purifying soluble, monomeric ASC PYD (as tested by SEC). The increased solubility of the GB1-tagged construct allowed for the acquisition of NMR data at physiological pH. To ensure that the GB1-solubility tag does not affect the 3-dimensional structure of ASC, the 2D [^{1}H,^{15}N] HSQC spectra of untagged and GB1-tagged ^{15}N-labeled ASC PYD were superimposed. Supporting Figure S5 shows that the resonances from ASC PYD overlap very well in both spectra, confirming that the GB1-taged ASC PYD adopts an identical structure to untagged ASC PYD. In summary, we developed a new approach that uses GB1-ASC PYD to test homo- and heterotypic PYD interactions at physiological pH, and used it to show that there is no interaction between NLRP4 PYD and ASC PYD.

In an additional experiment, we generated three NLRP4 PYD triple mutants that were identified by the Surface Entropy Reduction prediction (SERp) Server [218]. In each mutant, we replaced three residues in order to alter the electrostatic surface of NLRP4 PYD (NLRP4 PYD mutant 1: K19S K20S E21Q; NLRP4 PYD mutant 2: E64Q E65Q Q66E; NLRP4 PYD mutant 3: E38Q K40S Q41E). Using ^{15}N-labeled proteins, we collected a 2D [^{1}H,^{15}N] HSQC spectrum of each NLRP4 PYD mutant, which showed that all mutants are well-folded and the fold is not disrupted. (Supporting Figure S6). Furthermore, we repeated the NMR CSP studies between these mutant proteins and ASC PYD to test if changes of

electrostatics might lead to NLRP4 PYD binding to ASC PYD (Supporting Figure S6). Again, no interaction was detected with any of the mutants, highlighting that electrostatics likely play a much more insignificant role than previously expected in homotypic PYD interactions.

We also tested for an interaction between NLRP4 and ASC PYD using cellular assays. First co-immunoprecipitation assays were performed in HEK 293FT cells with Myc-tagged ASC and several truncated Flag-tagged NLRP4 constructs (Supporting Figure S7A). Since the ligand for NLRP4 activation is unknown, truncated constructs including NLRP4ΔLRR (NLRP4 lacking the LRR domain), NLRP4 PYD+Linker and NLRP4 PYD were used to avoid a misleading result from an auto-inhibited receptor that is unable to bind the downstream signaling partner. However, in agreement with the NMR studies, we were not able to detect an interaction between NLRP4 and ASC by co-immunoprecipitation (Supporting Figure S7A). The previously reported Aim2 PYD:ASC PYD interaction was used as a positive control. Lastly, we performed a yeast two-hybrid analysis of NLRP4 PYD and NLRP4 PYD+Linker with ASC PYD, which again confirmed the NMR and the co-IP results (Supporting Figure S7B) [198].

Thus, although the NLRP4 PYD has distinct and oppositely charged surface patches that were originally thought to be necessary for homotypic PYD interactions, we were not able to detect a direct interaction between NLRP4 PYD and ASC PYD. These findings are in agreement with the observation that NLRP4 has no effect on IL-1β activation [177, 178, 219] and show that other factors must

play a vital role in forming these interactions. Furthermore, they are more in line with the previously reported importance of hydrophobic residues in CARD:CARD homotypic interactions [220] as well as a PYD heterotypic interaction recently reported between NLRP12 PYD and FAF-1 [182].

B.5. Discussion

PYDs are the most abundant members in the DD superfamily and it has become apparent that they have an essential role in the control of signaling in immune and apoptotic pathways [164, 208]. During the last 5 years, numerous PYD structures have been reported, which have resulted in detailed analyses of both the structural, and, in the case of NMR structures, dynamic parameters that control PYD activity and specificity [179-183, 199]. While the differences between PYDs, in both structure and dynamics, are more numerous than the similarities, work on the signaling adaptor protein ASC and its inhibitor ASC2 showed the importance of electrostatic complementary between PYD domains for their interactions [211]. Similar results were reported for CARD:CARD and DD:DD interactions [212, 213, 220].

Here, we report the 3-dimensional crystal structure of the NLRP4 PYD at 2.3 Å resolution. NLRP4 PYD adopts a typical death domain fold with six α-helices that are arranged in an anti-parallel helical bundle around a highly conserved hydrophobic core. This hydrophobic core was previously established based on our work on multiple NLRP PYD domains [182, 183]. Hydrophobic residues also play a decisive role in stabilizing the conformation of helix α3 and the α2-α3 connecting loop relative to the rest of the domain, as readily seen in

the structures of PYDs of NLRP3, NLRP7, NLRP12 and now NLRP4. The only outlier is NLRP1 PYD, in which the helix α3 does not exist and the α2-α3 connecting loop is exceedingly dynamic, as probed by fast timescale auto-correlated relaxation measurements [180]. This is important, as the surface formed by helix α3 and the α2-α3 connecting loop has been identified as an important protein-protein interaction site for PYDs and thus differences in this region are likely the driving force for differential binding and selectivity among PYDs.

Interestingly, in NLRP4 PYD, the surface electrostatics of helix α3 and the α2-α3 loop region are highly similar to those of ASC PYD and ASC2 PYD. The 3-dimensional structure of the ASC PYD and ASC2 complex showed that the positive ESP of ASC2 interacts with the negative ESP of ASC PYD [211]. Thus, we tested for an interaction of NLRP4 PYD with ASC PYD using a variety of *in vitro* and *in vivo* tools. Significantly, to perform *in vitro* studies with ASC PYD under biologically relevant conditions, it was necessary to develop a novel construct of ASC PYD that was stable at physiological pH. This is because, until this work, heterologously expressed ASC PYD had only been isolated and refolded from inclusion bodies and was only stable at acidic pH values. By fusing the ASC PYD to GB1, a small highly soluble protein domain, we were able to express soluble ASC PYD which, once purified, is stable at a physiological pH values (i.e., 7.2). This is critical, as previous reports showed that homotypic PYD:PYD interactions are very sensitive to pH, with optimal interaction

conditions occurring at physiological pH [221]. Thus, the GB1-ASC PYD construct is an exciting new tool for research in the NLRP field.

Interestingly, using both *in vitro* CSP measurements and *in vivo* pull-down studies, we were not able to detect a direct interaction between NLRP4 PYD and ASC PYD. Overlap of the 3-dimensional structure of NLRP4 PYD and ASC2 and ASC PYD allows for structural comparisons of the Arg and Lys residues that are responsible for the formation of the positively charged surface. The superimposition of NLRP4 PYD with ASC2 (Figure 4) shows that the Arg and Lys residues of NLRP4 PYD do not align with those in ASC2, indicating that the local

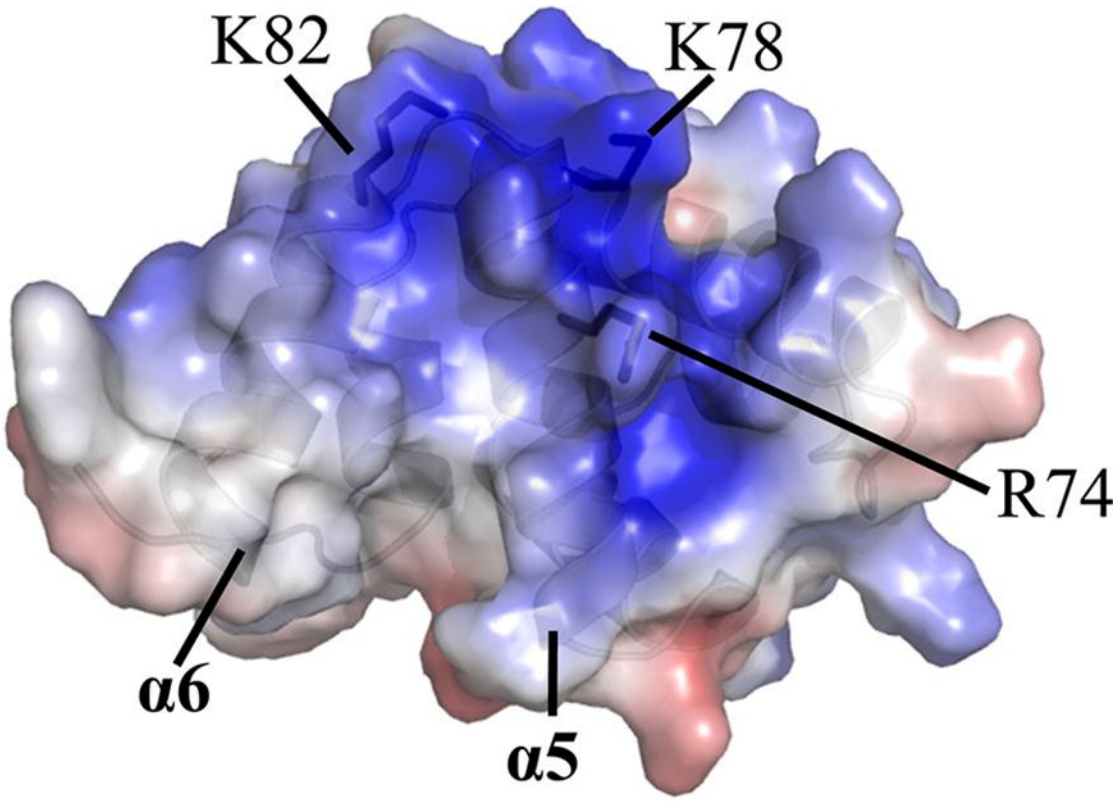


Figure 63: Figure 4, Eibl, Biochemistry. NLRP4 PYD showing an additional positive patch.
The electrostatic potentials of helices α5 and α6 are mapped onto the solvent-accessible surface of the NLRP4 PYD. Residues Arg74, Lys78, and Lys82, form this positively charged surface.

conformation and orientation of these residues might be important for the formation of these complexes. Notably, NLRP4 PYD also has a novel second positively charged surface, which is formed by residues Arg74 and Lys78 on helix α5 and Lys82 on helix α6 (Figure 4). This is interesting, as the α5-α6 loop and helix α6 are proposed to engage in homotypic type II interactions that occur within platforms, as was described for the homotypic DD interaction within the PIDDosome [202, 208, 212].

Our structure of the NLRP4 PYD, especially in the context of the PYD family, further illustrates that electrostatic surface distribution varies widely among PYDs. Thus, it is increasingly apparent that individual PYDs most likely have distinct interaction partners [182, 222]. Furthermore, we have also shown that NLRP4 differs from ASC-binding PYDs in several important aspects, and thus likely functions in an ASC-independent manner. Our analysis and mutagenesis CSP experiments also nicely show that electrostatic interactions might not be as important as previously assumed for death domain fold protein interactions. In agreement with our data, previous reports highlighted the importance of hydrophobic interactions for homotypic death domain. Indeed, as shown in Figure 5, it is possible to structurally align the CARD domains of Apaf-1 and caspase 9 with NLRP4 PYD. This alignment enabled us to identify hydrophobic residues that were important for the CARD:CARD interaction between Apaf-1 and caspase 9 show structural conservation with hydrophobic residues in NLRP4 PYD.

Taken together, our earlier findings describing the Pyrin domain fold and stability, combined with our current work, have significantly contributed to the characterization of PYDs and their function. Our current findings suggest that the generally accepted electrostatic model of PYD-PYD interaction specificity is incomplete and that additional factors, such as dynamics and hydrophobic interactions, also likely play an important role in these interactions.

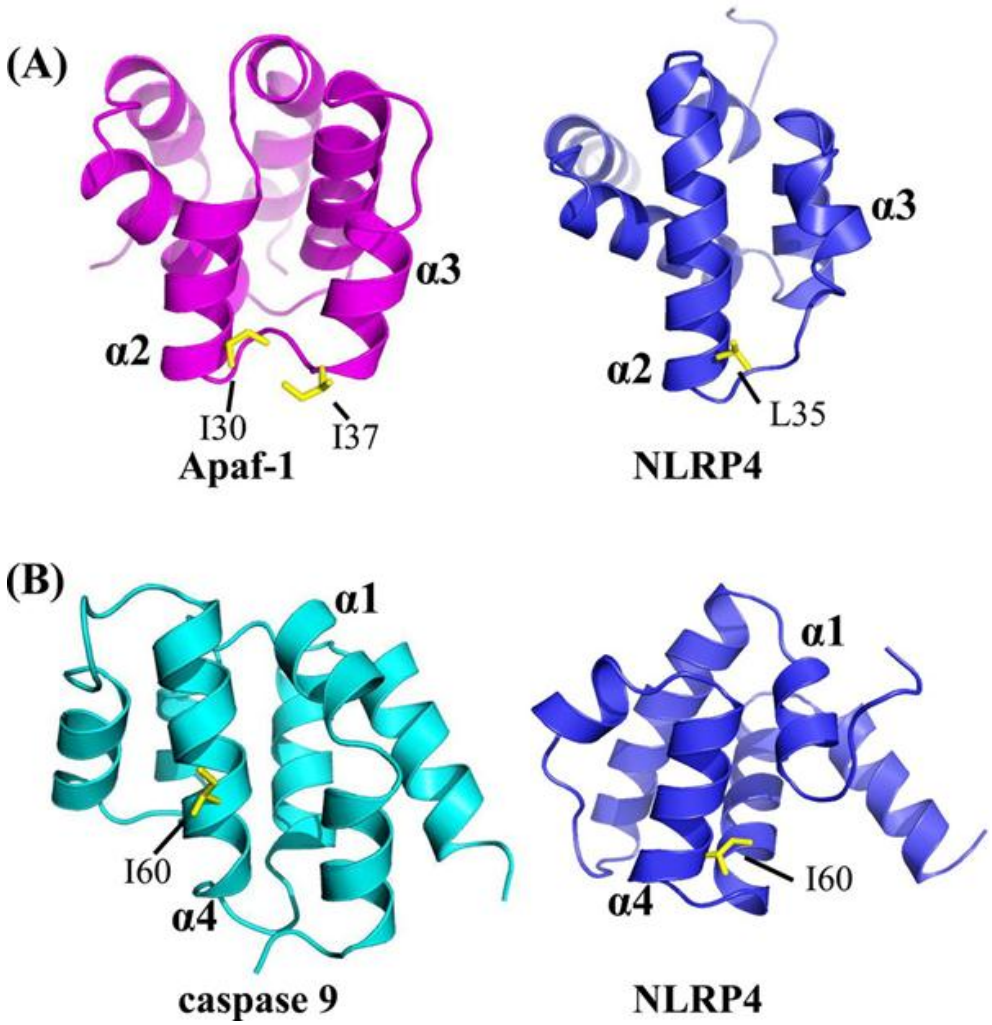


Figure 64: Figure 5, Eibl, Biochemistry. Comparison of the NLRP4 PYD with the Apaf-1 CARD and the caspase 9 CARD.
Residues (yellow sticks) that mediate a hydrophobic interaction between the Apaf-1 CARD (magenta) and the caspase 9 CARD (cyan) are also surface-exposed in the NLRP4 PYD (blue).

B.6. Supplementary figures

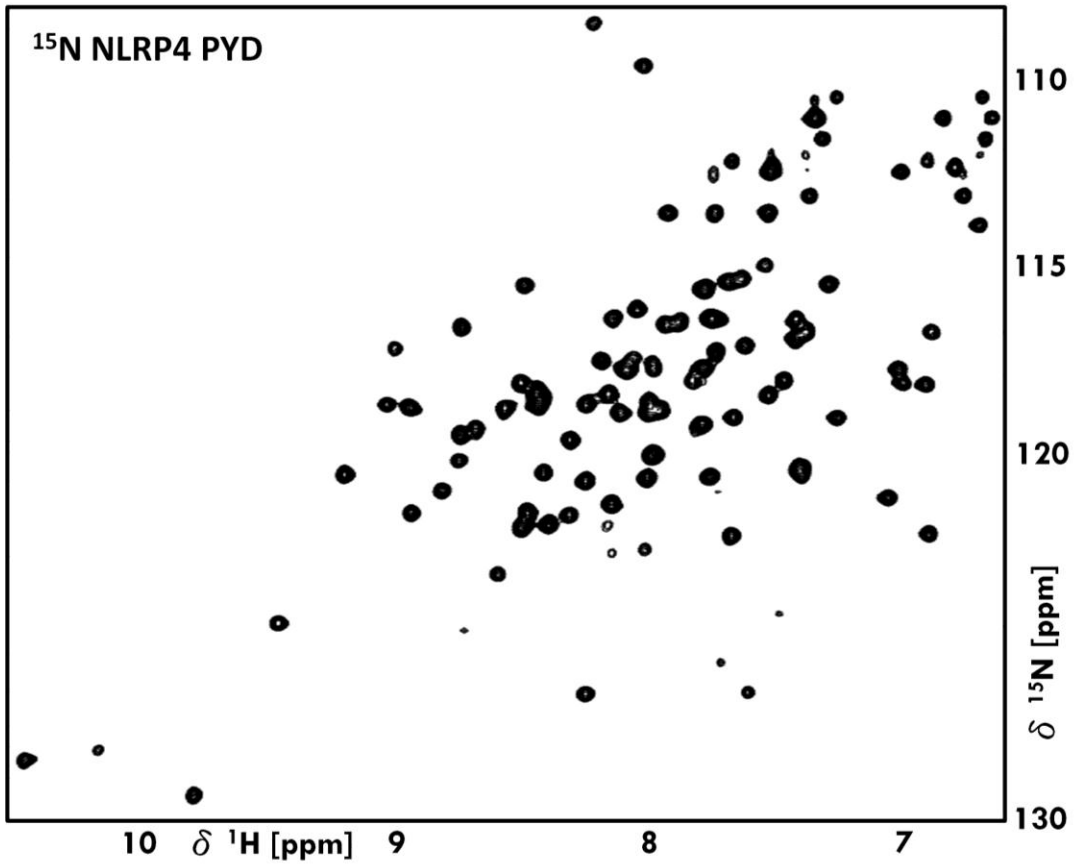


Figure 65: Figure S1, Eibl, Biochemistry. NMR analysis of NLRP4 PYD.
2D [^{1}H,^{15}N] HSQC spectrum of ^{15}N-labeled NLRP4 PYD (20 mM sodium phosphate, pH 7.2, 100 mM NaCl, 0.5 mM TCEP).

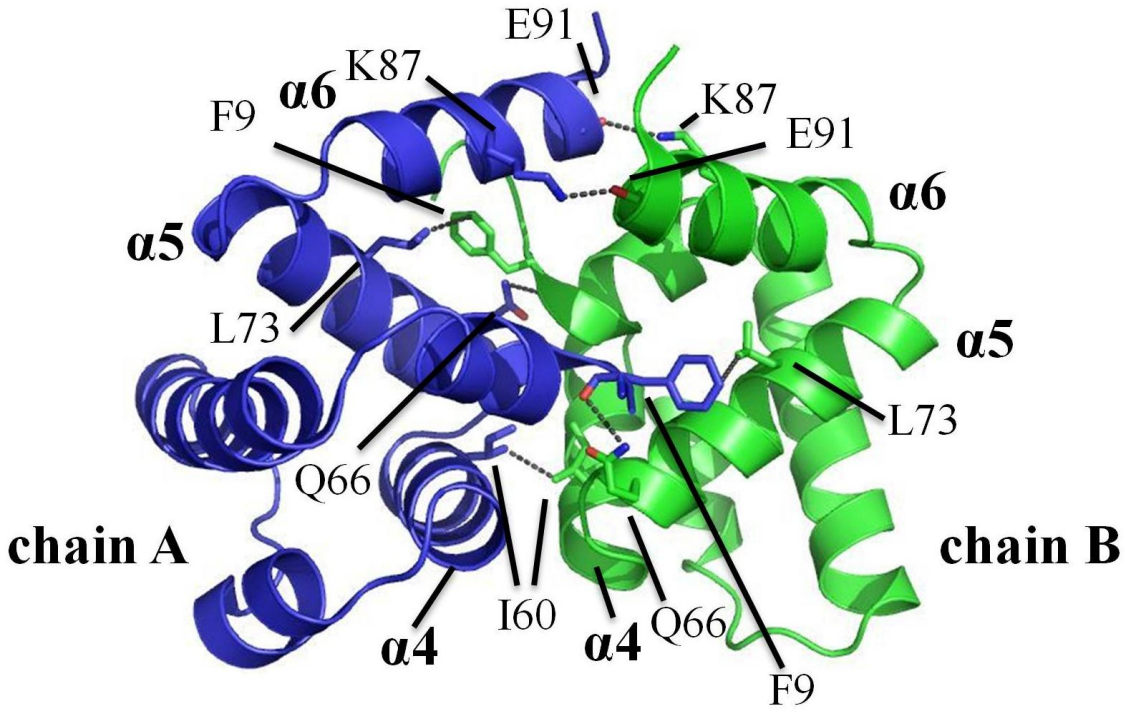


Figure 66: Figure S2, Eibl, Biochemistry. Structural characterization of NLRP4 PYD – Crystallographic dimer formation.
The NLRP4 PYD crystallographic dimer (dimer is not observed in solution). Chain A is colored blue and chain B is colored green. The residues mediating dimerization are shown in stick representation and labeled. Dashed lines highlight interacting residues.

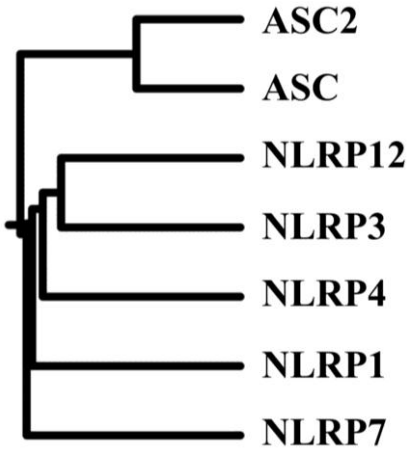


Figure 67: Figure S3, Eibl, Biochemistry. Phylogenetic tree of the PYDs based on their structural similarity.

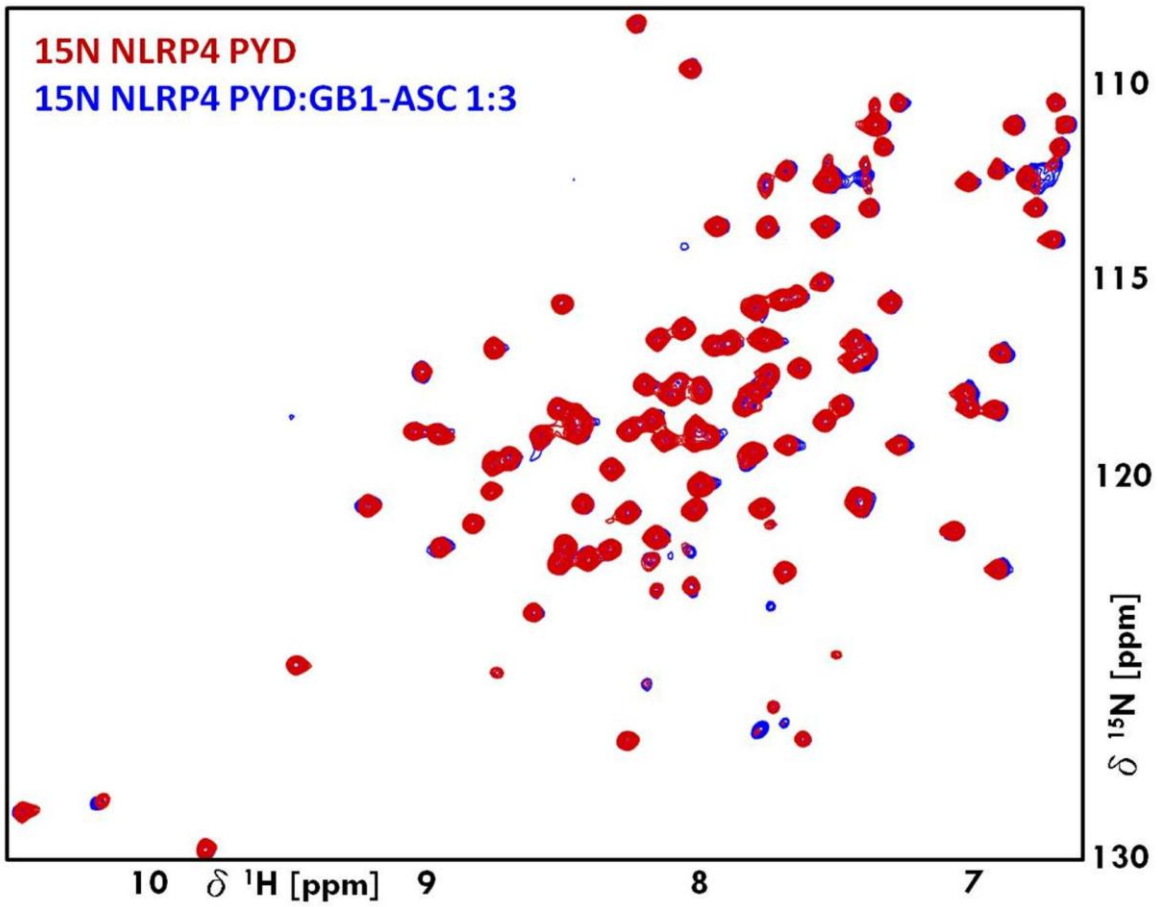


Figure 68: Figure S4, Eibl, Biochemistry. Chemical Shift Perturbation Study of NLRP4 PYD and ASC PYD.
Overlay of the 2D [^{1}H,^{15}N] HSQC spectra of ^{15}N-labeled NLRP4 PYD (red) ^{15}N-labeled NLRP4 PYD:GB1-ASC PYD 1:3 (blue). Spectra overlap perfectly, showing that NLRP4 PYD and GB1-ASC PYD do not interact with one another. Both spectra were acquired in 20 mM sodium phosphate, pH 7.2, 100 mM NaCl, 0.5 mM TCEP. Reference and titration spectra were recorded using ^{15}N-labeled NLRP4 PYD at 50 µM and 16 µM respectively.

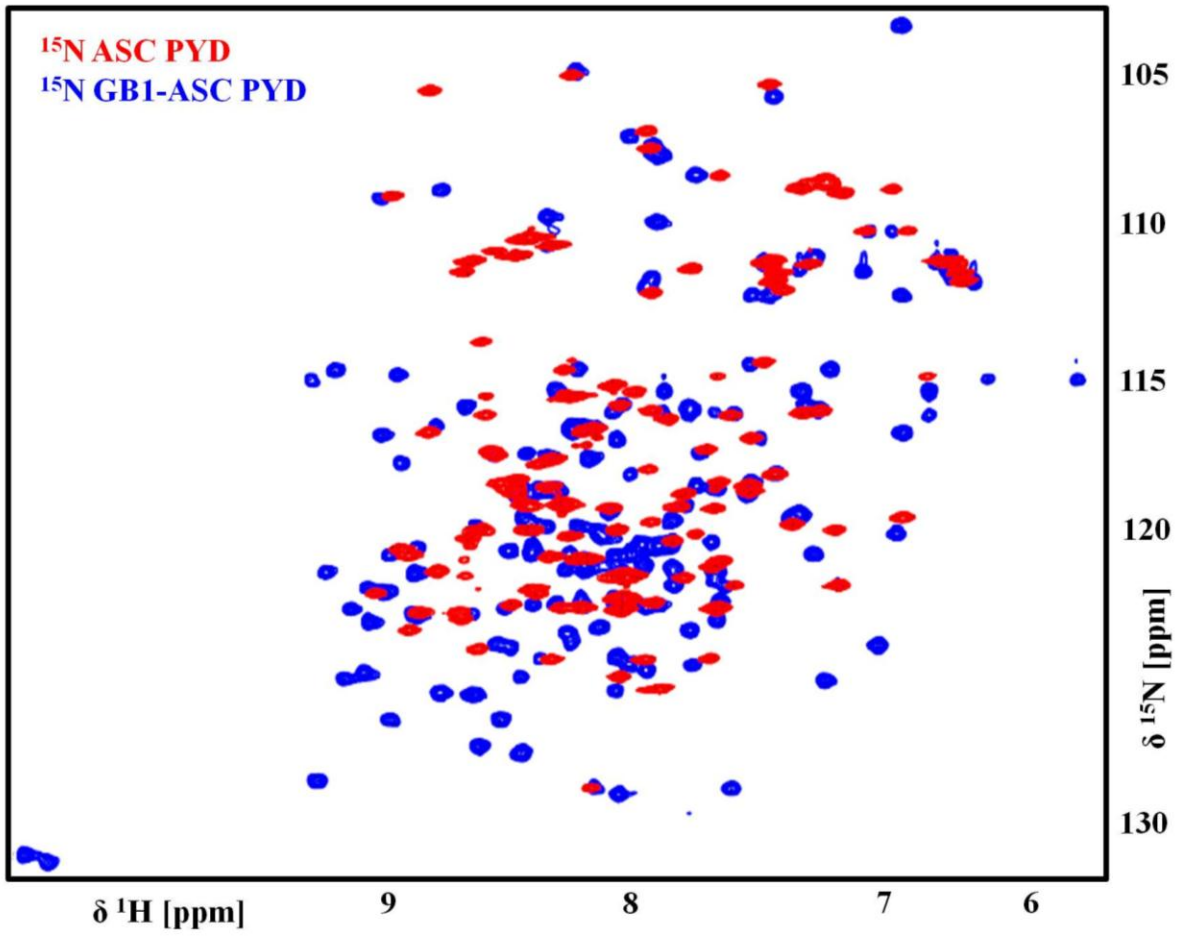


Figure 69: Figure S5, Eibl, Biochemistry.
Overlay of the 2D [^{1}H,^{15}N] HSQC spectra of untagged, refolded ^{15}N-labeled ASC PYD (L25A; red) and ^{15}N-labeled GB1-ASC (L25A; blue). Spectra overlap very well, confirming that GB1-ASC PYD adopts the identical structure as ASC PYD itself. Additional peaks in the 2D [^{1}H,^{15}N] HSQC spectrum of GB1-ASC belong to GB1. The chemical shift overlap is not perfect due to the significant difference in pH, as the 2D [^{1}H,^{15}N] HSQC spectrum of untagged ASC PYD (40 µM – at higher concentration the protein aggregates rapidly) was acquired in 40 mM sodium acetate pH 4.5 (low pH is required), 100 mM NaCl, 20 mM DTT and the 2D [^{1}H,^{15}N] HSQC spectrum of GB1-tagged ASC PYD (540 µM) was acquired in 20 mM sodium phosphate pH 7.2, 100 mM NaCl, 0.5 mM TCEP, 50 mM L-Arginine, 50 mM L-Glutamic acid.

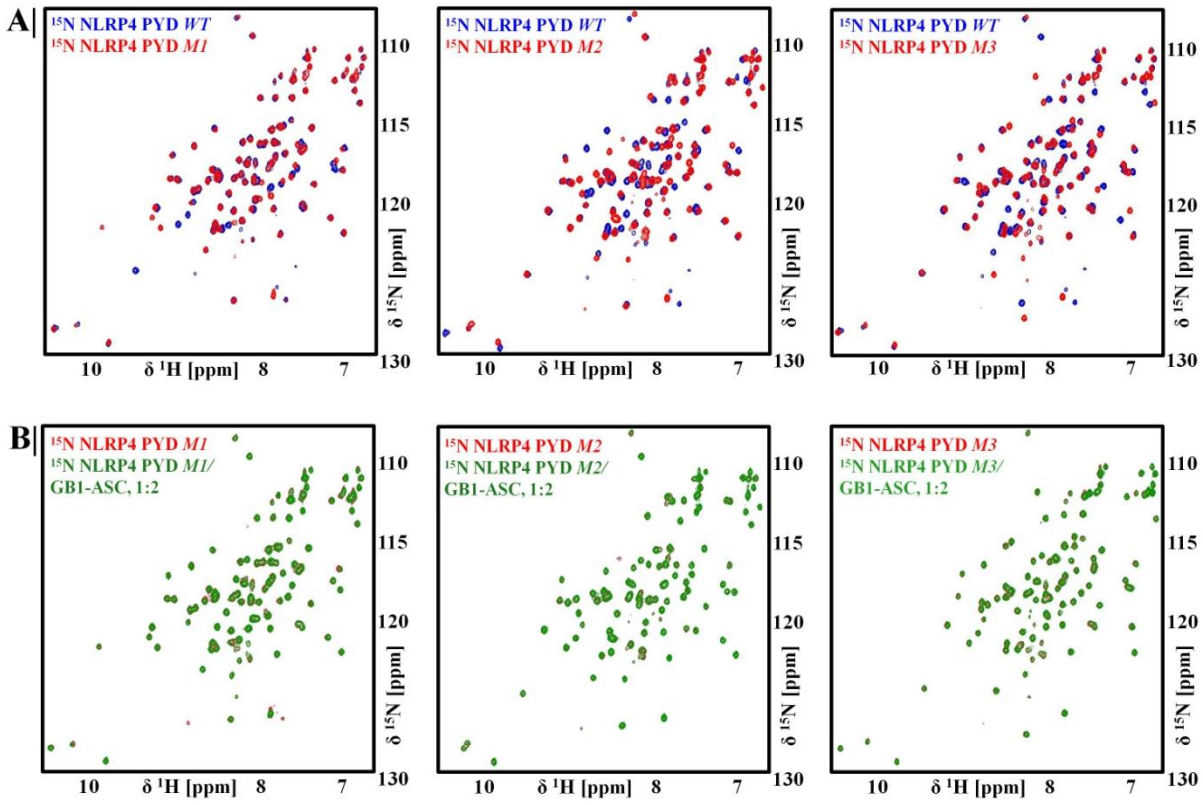


Figure 70: Figure S6, Eibl, Biochemistry.
(A) Overlay of the 2D [^{1}H,^{15}N] HSQC spectra of wt-NLRP4 PYD with 3 NLRP4 PYD mutants, which were designed to change the electrostatic surface of NLRP4 PYD: NLRP4 PYD mutant 1 (M1): K19S K20S E21Q; NLRP4 PYD mutant 2 (M2): E64Q E65Q Q66E; NLRP4 PYD mutant 3 (M3): E38Q K40S Q41E. The 2D [^{1}H,^{15}N] HSQC spectrum for each NLRP4 PYD mutant show that all mutants are well folded and thus amenable for functional protein:protein interaction studies. (B) Overlay of the 2D [^{1}H,^{15}N] HSQC spectra of ^{15}N-labeled NLRP4 PYD (M1, M2, M3) (red) and ^{15}N-labeled NLRP4 PYD (M1, M2, M3):GB1-ASC PYD 1:2 (green). Spectra overlap perfectly, showing that NLRP4 PYD electrostatic surface mutants and GB1-ASC PYD do not interact with one another.

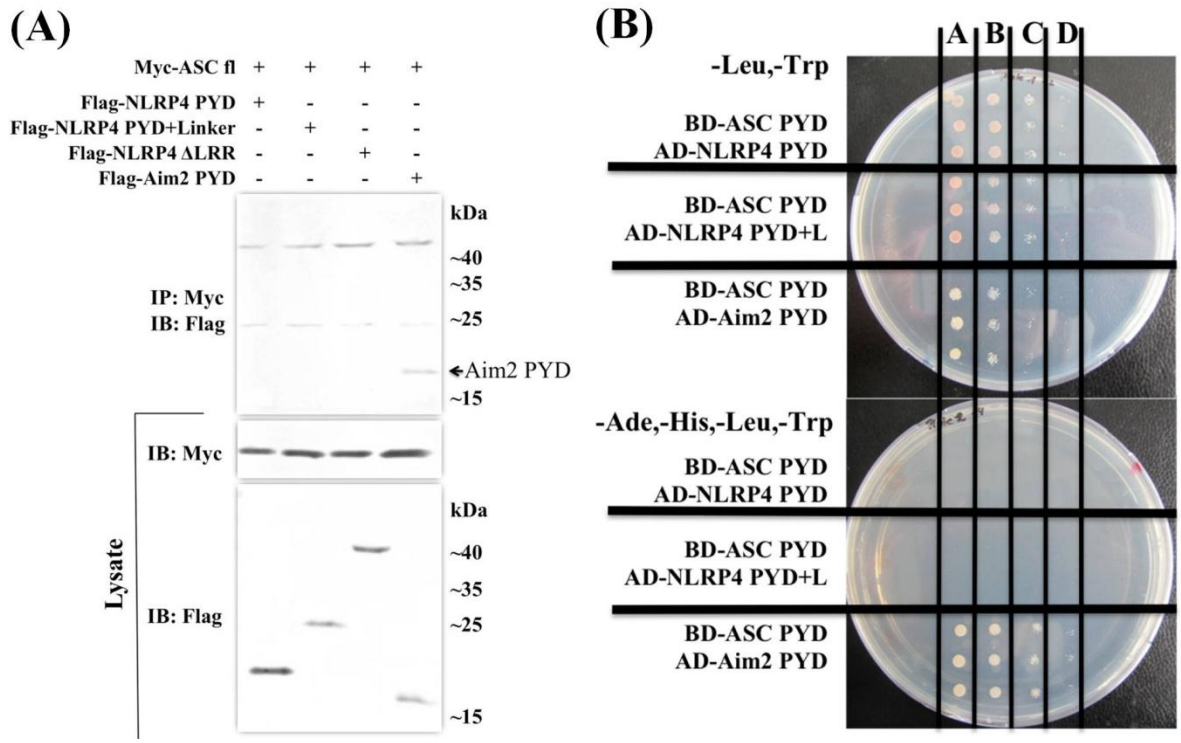


Figure 71: Figure S7, Eibl, Biochemistry. NLRP4 PYD does not interact with ASC PYD.
(A) Co-Immunoprecipitation analysis was performed in HEK293FT cells. Myc-ASC was co-transfected with indicated Flag-NLRP4 constructs as well as with Flag-Aim2 PYD as positive control. Lysate (5% of total cell extract) demonstrate the expression of the proteins. The IP panel indicates that only Aim2 PYD was immunoprecipitated by Myc-ASC, which was immobilized on G-Sepharose beads. IB stands for Immunoblot whereas IP stands for Immunoprecipitation. (**B)** Yeast two-hybrid analysis was performed in Y2HGold *Saccharomyces cerevisiae* strain. ASC PYD fused to the DNA-Binding Domain (BD) was either co-transformed with Activation Domain (AD) fused NLRP4 PYD, NLRP4 PYD+Linker or Aim2 PYD and spotted on SD/–Leu/-Trp dropout medium to assess transformation efficiency and onto SD/-Ade/-His/-Leu/-Trp selection medium to test for potential interactions. Three clones from the SD/-Leu/-Trp plates from each transformation were diluted 1:1 (Lane A), 1:10 (Lane B), 1:100 (Lane C) and 1:1000 (Lane D) and again spotted onto SD/–Leu/-Trp and onto SD/-Ade/-His/-Leu/-Trp selection medium and incubated for 4-5 days on 30°C. Whereas the SD/–Leu/-Trp plate shows that the transformation worked for all combinations, the SD/-Ade/-His/-Leu/-Trp plate indicates that neither NLRP4 PYD nor NLRP4 PYD+Linker interacts with ASC PYD. However, Aim2 PYD does interact with ASC PYD, conforming that the experimental setup worked.

Appendix C: "Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties" and supplementary information

Reproduced from the journal PLoS Pathogens, Volume 5, Number 12, e1000706

I am co-first author of this manuscript together with Breann L. Brown. I performed all experiments related to the structure determination of MqsR:MqsA, as well as *E. coli* growth curves to assess toxicity of various MqsR:MqsA complexes. I also performed structural analysis of MqsR:MqsA and MqsR.

Supplementary protocols, figures, and tables for this manuscript are located in **Sections C.7-C.9 of this appendix.**

Author Contributions:

Conceived and designed the experiments: TKW WP RP. Performed the experiments: BLB SG YK JMA AD. Analyzed the data: BLB SG YK JMA AD TKW WP RP. Wrote the paper: TKW WP RP. Commented on the manuscript: BLB SG YK JMA AD.

Breann L. Brown[1,6], Simina Grigoriu[2,6], Younghoon Kim[3], Jennifer M. Arruda[2], Andrew Davenport[1], Thomas K. Wood[3,4,5], Wolfgang Peti[1] and Rebecca Page[2,*]

[1]Department of Molecular Pharmacology, Physiology and Biotechnology Brown University, Box GE-3, Providence, RI, United States of America, [2]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Box GE-4, Providence, RI, United States of America, [3]Artie McFerrin Department of Chemical Engineering, Texas A & M University, College Station, Texas, United States of America, [4]Department of Biology, Texas A & M University, College Station, Texas, United States of America, [5]Zachry Department of Civil Engineering, Texas A & M University, College Station, Texas, United States of America, [6]***These authors contributed equally to this work.***

Accession Numbers

The structure factors and coordinates for MqsA-F and the MqsR:MqsA-N complex have been deposited with the Protein Databank with accession numbers 3GN5 and 3HI2, respectively.

Acknowledgments

Crystallographic data was collected at the Brookhaven National Laboratory (BNL) National Synchrotron Light Source (NSLS) X6A beamline. We thank the beamline staff for help with X-ray fluorescence scans. WP is the Manning Assistant Professor for Medical Science at Brown University. TKW is the T. Michael O'Connor II Chair and Professor at Texas A & M University.

C.1. Author Summary

Most bacteria live in biofilms, microbial communities that cause more than 80% of human infections. Biofilms have a genetically identical sub-population of dormant cells, named persister cells, which are the well-recognized source of antibiotic resistance. Recently, it was demonstrated that toxins are highly upregulated in persisters and have therefore been postulated to play a role in the persister state. Using an inter-disciplinary approach, we reveal how *mqsR*, the gene most highly upregulated in persisters, together with *mqsA*, function: they are the founding members of a new family of toxin:antitoxin (TA) systems. Unexpectedly, the structure of MqsR reveals that it is a ribonuclease, a protein that controls the production of other essential proteins. Moreover, we identified multiple features of this TA system that are so unique that each is a starting point for drug development. Unlike other antitoxins, MqsA is structured throughout its entire sequence, its structure is unchanged between the free and toxin-bound states and it binds zinc. It also binds DNA via its C- and not N-terminal domain. Finally, MqsA binds both its own promoter and additional genes important for ***E. coli*** physiology. Taken together, our data provide fundamental new insights into the role of MqsR and MqsA in bacterial persistence and biofilms.

C.2. Abstract

One mechanism by which bacteria survive environmental stress is through the formation of bacterial persisters, a sub-population of genetically identical quiescent cells that exhibit multidrug tolerance and are highly enriched in bacterial toxins. Recently, the *Escherichia coli* gene *mqsR* **(***b3022***)** was identified as the gene most highly upregulated in persisters. Here, we report multiple individual and complex three-dimensional structures of MqsR and its antitoxin MqsA (B3021), which reveal that MqsR:MqsA form a novel toxin:antitoxin (TA) pair. MqsR adopts an α/β fold that is homologous with the RelE/YoeB family of bacterial ribonuclease toxins. MqsA is an elongated dimer that neutralizes MqsR toxicity. As expected for a TA pair, MqsA binds its own promoter. Unexpectedly, it also binds the promoters of genes important for *E. coli* physiology (e.g., *mcbR***,** *spy*). Unlike canonical antitoxins, MqsA is also structured throughout its entire sequence, binds zinc and coordinates DNA via its C- and not N-terminal domain. These studies reveal that TA systems, especially the antitoxins, are significantly more diverse than previously recognized and provide new insights into the role of toxins in maintaining the persister state.

C.3. Introduction

The emergence of increasing numbers of bacteria that are resistant to antibiotics portends a major public health crisis. One well-recognized but poorly understood mechanism used by bacteria to survive environmental stress is through the formation of persisters, a subpopulation of cells that survive prolonged exposure to antibiotics [223] and exhibit multidrug tolerance [224]. Persisters are not antibiotic-resistant mutants. Instead, they are phenotypic variants that pre-exist in bacterial populations. The dormant, non-dividing persister cells [223-225] allow bacteria to survive until the environmental stress is relieved, after which the persisters spontaneously revert to the non-persistent state and repopulate the original culture. Critically, the detailed molecular events that lead to and propagate the persister phenotype are still elusive, as persisters typically represent only a small fraction of the bacterial population. In wild-type *E. coli*, the frequency of persisters in planktonic cultures is only about one in a million [226]. However, in biofilms, complex multicellular bacterial communities that are highly resistant to antibiotics and that are responsible for more than 80% of human infections, this frequency increases substantially, up to one in a hundred [227]. The increased incidence of persister cells in biofilms, and their role in human bacterial infections, has stimulated renewed efforts to understand the molecular mechanism(s) that underlies the persister phenotype.

Recent studies have demonstrated that the persister state is correlated with the increased expression of chromosomal toxins from toxin:antitoxin (TA) genes [224, 228]. TA pairs [229, 230], also known as plasmid addiction systems,

are highly abundant on bacterial plasmids [231, 232] and chromosomes [233-237]. They are composed of two genes organized in an operon that encode an unstable antitoxin and a stable toxin, respectively. Critical to their function, the protein products of TA pairs have considerable differences in lifetimes [238], with the antitoxin being highly susceptible to degradation by cellular proteases and the toxin comparatively stable. Under normal conditions, the toxin and antitoxin associate to form a tight, non-toxic complex. However, under conditions of stress, the antitoxins are degraded by either the ATP-dependent protease (Lon [238, 239]) or the bacterial protease systems (ClpXP [240]; ClpAP [241]). This leads to a dramatic reduction of both translation and replication rates and, in turn, the cessation of cell growth due to the cellular effects of the toxin. Toxin activities are diverse, and include inhibiting replication by blocking DNA gyrase [242, 243], halting translation via mRNA cleavage [244, 245], or inactivating EF-Tu by phosphorylation [246], among others. TA complexes also typically function as transcriptional repressor:corepressors, where the antitoxin binds to the promoter DNA within the TA operon and the toxin enhances DNA binding [247-249].

To date, more than ten TA loci have been identified in *E. coli* [229], including *relBE* [234, 250], *mazEF* [233, 251], *dinJ-yafQ* [252], *hipBA* [246, 253], *hicAB* [254] and *yefM-yoeB* [6, 239]. Gene expression profiling experiments have shown that multiple toxins are highly upregulated in persister cells, especially *relE*, *mazF* and *yoeB* [224, 228]. The activities of these toxins lead to a rapid cessation of cell growth, and have been postulated to play a role in the persistence phenotype [228]. Unexpectedly, the gene most highly upregulated in

persisters is *mqsR* (*ygiU*/*b3022*), a gene originally identified as a motility quorum sensing regulator sensitive to autoinducer-2 that controls biofilm formation [255, 256], but which had not, until recently, been shown to be a bacterial toxin. Because the sequence of MqsR is not similar to that of any other known toxin, its molecular function is unknown.

Deletion of *mqsA* (*ygiT/b3021*), the second gene in the two-gene *mqsRA* operon, is lethal [228, 257]. This led to the postulation that *mqsRA* constitutes a novel TA module [228, 237], with MqsR as the toxin and MqsA as its cognate antitoxin. However, *mqsRA* has many characteristics that differ from canonical TA systems. First, in the *mqsRA* operon, *mqsR* precedes, instead of follows, *mqsA*. This unusual genetic organization has only been observed in two other recently characterized TA systems, that of *higBA* [258] and *hicAB* [254]. Second, their isoelectric points are nearly identical (8.8, MqsR; 9.1, MqsA) rather than being basic and acidic for the toxin and antitoxin, respectively. Third, the MqsA protein is larger, instead of smaller, than MqsR; the only other TA system with an antitoxin larger than its cognate toxin is that of *hicAB* [254]. Finally, their sequences are not homologous to any member of a recognized TA system.

In this paper, we employed a combination of biochemical and structural studies to show that MqsR, along with MqsA, are a *bona fide* TA pair that, because of the unique features of MqsA, define a novel family of TA modules. We show that MqsR is toxic and forms a tight complex with its antitoxin, MqsA, an interaction that mitigates MqsR toxicity. MqsA and the MqsR:MqsA complex also bind the promoters of the *mqsRA* operon and, unexpectedly, genes critical

for *E. coli* physiology, including *mcbR* and *spy*. To the best of our knowledge, this is the first time a TA pair has been shown to bind and regulate promoters other than its own. The structure of MqsR reveals that it is a member of the RelE/YoeB family of bacterial RNase toxins. Based on its similarity with RelE, MqsR likely functions as a ribosome-dependent RNase. This suggests that MqsR is important for bacterial persistence via its ability to inhibit translation and, in turn, cell growth. MqsA itself is a two-domain protein with a novel fold that, unlike every other antitoxin, is well-ordered throughout its entire sequence and whose structure does not change upon toxin binding. It is also the first antitoxin known that binds metal, in this case zinc. These studies reveal the molecular mechanisms by which MqsR and MqsA mediate the cessation of cell growth and provide novel targets for the development of a new class of antibiotics that target TA pairs.

C.4. Materials and Methods

C.4.a. BW25113 colony formation

The bacterial strains and plasmids used in this study are listed in Table S1. Growth experiments with *E. coli* strain BW25113 were conducted in LB medium at 37°C. Growth experiments were monitored using pBS(Kan)-based plasmids [259]. To construct pBS(Kan)-based plasmids for producing MqsR, MqsA-F, and MqsR-MqsA-F from a *lac* promoter, the fragments from genomic DNA were amplified by PCR (Table S2) and directionally cloned into pBS(Kan). The toxicity of selected proteins was investigated using pBS(Kan) plasmids with 1 mM IPTG added upon inoculation. Cell viability (CFU) measured by diluting cells from 10^2

to 10^7 via 10-fold serial dilution steps into 0.85% NaCl solution and applying them as 10 µL drops on LB agar with kanamycin or chloramphenicol [260].

C.4.b. Cloning, Expression and Purification

Full-length MqsA (MqsA-F, residues 1-131), the MqsA N-terminal domain (MqsA-N, residues 1-76) and the MqsA C-terminal domain (MqsA-C, residues 62-131) were subcloned into a modified pET28a vector which contained an N-terminal his_6-tag with a TEV cleavage site [261]. Proteins were expressed in One Shot BL21 (DE3) cells (Invitrogen) and purified by sequential his_6-tag, TEV cleavage, second subtraction his_6-tag and size-exclusion chromatography (SEC). For the MqsR:MqsA-N complex, MqsR was subcloned into the pET28a vector (Novagen) which contains an N-terminal his_6-tag with a thrombin cleavage site. MqsA-N was subcloned into the untagged pCA21a vector (Expression Technologies, Inc.). The complex was co-expressed and purified by sequential his_6-tag, thrombin cleavage, second subtraction his_6-tag and SEC.

C.4.c. Crystallization

Crystals of MqsA-F were obtained using sitting drop vapor diffusion method at 4°C. Crystals were grown by mixing 0.2 µL of MqsA-F protein (3.5 mg/ml; 10 mM Tris pH 7.0, 50 mM NaCl, 0.5 mM TCEP, 5% (v/v) ethanol) with 0.2 µL of precipitant (75 mM Bis-Tris, pH 5.5, 150 mM $MgCl_2$, and 19% (w/v) PEG3350). Crystals of MqsR:MqsA-N were obtained using sitting drop vapor diffusion method at 4°C. Crystals were grown by mixing 0.2 µL MqsR:MqsA-N protein complex (14 mg/ml; 10 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM TCEP) with 0.4 µL of precipitant (0.1 M Bis-Tris pH 5.5 and 25% (w/v) PEG3350). MqsA-F and

MqsR:MqsA-N crystals were cryoprotected in precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection.

C.4.d. Data Collection and Processing

Data for both MqsA-F and the MqsR:MqsA-N complex were collected at the National Synchrotron Light Source, beamline X6A at Brookhaven National Laboratory. Native data for MqsA-F was collected from a single crystal at a single wavelength. Anomalous data for the MqsR:MqsA-N complex was collected from a single crystal; a three-wavelength MAD experiment was performed collecting data at the experimentally determined Zn/K absorption edge, the inflection point and a high energy remote wavelength. Data were processed and scaled using HKL2000 [262].

C.4.e. Model building and refinement

MqsA-F: Molecular replacement (Phaser [263]) using the structures of the MqsA individual domains (MqsA-N and MqsA-C; structure determination of MqsA-N and MqsA-C described in Supporting Protocols, Figs. S5A, S5B and Tables S3, S4) as search models resulted in a single solution, with two molecules of each domain present in the asymmetric unit. The resulting electron density maps were readily interpretable. Model building was performed in Coot [264] followed by restrained refinement in REFMAC 5.2.0019 [265]. As observed for the structure of MqsA C-terminal domain alone (Fig. S3B), Gln108 of monomer B is methylated at N5. *MqsR:MqsA-N complex*: SOLVE was used to determine the positions of two expected zinc atoms. Initial phases to 2.0 Å were improved with density modification in RESOLVE. An initial model built by ARP/wARP served as

the basis for subsequent manual model building and refinement. Iterative cycles of refinement in REFMAC against the high energy remote data were used to complete and refine the model. Data collection, model and refinement statistics for both structures are reported in Table 1. Structure validation and stereochemistry analysis was performed with Molprobity [110] and SFCHECK [266].

C.4.f. Electrophoretic mobility shift assays (EMSA)

The targeted promoter regions (150–250 bp upstream of the start codon using primers reported in Table S2) were amplified, purified and labeled with biotin using the Biotin 3'-end DNA Labeling Kit (Pierce Biotechnology). After binding the proteins (200 ng; MqsA-F, MqsA-N, MqsA-C, MqsR:MqsA-N and MqsR:MqsA-F) with biotin-labeled target promoters (5 ng), electrophoresis was conducted at 100 V at 4^{o}C using a 6% DNA retardation gel (Invitrogen). The binding mixtures were transferred to a nylon membrane (Roche Diagnostics GmbH) using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad), and 3'-biotin-labeled DNA was detected with the Light-Shift Chemiluminescent EMSA kit (Pierce Biotechnology).

C.4.g. MqsR toxicity assays

Mutagenesis of MqsR was carried out using the Quikchange mutagenesis kit (Stratagene) and sequence verified. Growth experiments were conducted in LB at 18°C and all mutants were tested in parallel. Briefly, 10 ml of an overnight culture was used to inoculate 1 L of LB and then incubated at 37°C with rigorous

Table 5: Table 1, Brown and Grigoriu, PLoS Pathog. Data collection and refinement statistics for MqsA-F and MqsR:Mqs-N

	MqsA-F	MqsR:MqsA-N (peak)	MqsR:MqsA-N (inflection)	MqsR:MqsA-N (remote)
Data Collection[1]				
Space Group	$P2_1$		$P4_1\,2_1\,2$	
Unit Cell (Å)	62.1, 31.0, 75.5		63.2, 63.2, 194.9	
Unit Cell (°)	90.0, 106.6, 90.0		90.0, 90.0, 90.0	
Wavelength (Å)	1.0	1.2815	1.283	1.0
Resolution (Å)	50-2.15 (2.19-2.15)	50.0-2.30 (2.38-2.30)	50.0-2.30 (2.38-2.30)	50.0-2.00 (2.03-2.00)
R_{sym} (%)	7.4 (27.1)	4.7 (12.9)	4.8 (13.7)	5.8 (36.5)
<I/σI>	12.9 (4.85)	65.5 (15.1)	66.7 (28.9)	36.0 (8.6)
Completeness (%)	99.4 (99.7)	97.0 (98.1)	96.9 (98.9)	95.8 (98.0)
Redundancy	3.2 (3.1)	6.9 (2.4)	7.6 (7.2)	12.5 (11.6)
Refinement Statistics				
Resolution (Å)	50.00-2.15			40.64-2.00
R_{cryst} (%)	18.2			20.5
R_{free} (%)	23.3			24.8
Protein atoms	2052			2495
Waters	159			207
Ions	2			2
r.m.s.d bond length (Å)	0.011			0.013
r.m.s.d bond angle (°)	1.225			1.338
Average B factors (Å^2)				
Protein	27.31			35.29
Water	29.53			23.10
Ions	23.61			15.79
Ramachandran Plot				
Favored (%)	98.8			99.0
Allowed (%)	1.2			1.0
Disallowed (%)	0.0			0.0
PDB Code	3GN5			3HI2

[1] Highest-resolution shell data are shown in parentheses.

shaking. Once a mutant culture reached an OD_{600} of 0.4, it was cooled on ice. After all the mutants reached an OD_{600} of 0.4, the cultures were cooled for an additional hour, induced with 0.5 mM IPTG and incubated at 18°C with vigorous shaking. OD_{600} measurements were taken every hour for 20 hours. Growth assays for MqsR:MqsA-N and MqsR:MqsA-C were carried out similarly, except expression was induced at an OD_{600} of 0.7.

C.4.h. MqsR:MqsA$_2$:MqsR:DNA complex modeling

The P22 C2 repressor protein (PDBID 2R1J) [9] was identified as the closest structural homolog to the MqsA C-terminal domain (DALI Z-score = 6.0), whose DNA-bound structure was known. The MqsA:DNA complex was modeled by superimposing the conserved HTH-XRE helices from both proteins (MqsA residues 98-105; P22R residues 34-41) and mapping the rotated DNA coordinates onto one MqsA monomer. This MqsA monomer:DNA complex was then superimposed onto the symmetry-related MqsA monomer to obtain the rotated DNA coordinates for the second monomer.

C.4.i. Genes, proteins and structures discussed in the text.

The genes/proteins mentioned in the text include (UniProtKB/Swiss-prot ID unless stated otherwise): MqsR/YgiU (Q46865), MqsA/YgiT (Q46864), McbR (P76114), Spy (P77754), RelE (P0C077), RelB (P0C079), MazE (P0AE72), MazF (P0AE70), DinJ (Q8X7Q6), YaFQ (Q47149), HipA (P23874), HipB (P23873), HicA (P76106), HicB (P67697), YefM (P69346), YoeB (P69348), HigA (Q9KMG4/Q9KMA5), HigB (Q9KMG5/Q9KMA6). The PDB files mentioned in the

text are RelE (PDBID 2KC8), YoeB (PDBID 2A6Q) and RNase Sa (PDBID 1RSN).

C.5. Results

C.5.a. *mqsRA* is a bona fide TA locus

The ability of MqsR to arrest cell growth was examined by measuring its effect on colony formation (CFU/ml) and cell viability. Expression of MqsR alone leads to cell growth arrest in multiple bacterial strains (BW25113 and MG1655), while co-expression of MqsR with full-length MqsA (referred to hereafter as MqsA-F) rescues the cell growth arrest phenotype (**Figs. 1A, S1A**-E and [228, 267]). In addition, MqsR and MqsA-F form a tight oligomeric complex, as MqsA-F (untagged) co-purifies with MqsR (his-tagged) and forms a dimer of dimers, composed of two copies of MqsR and two copies of MqsA-F (hereafter referred to as MqsR:$MqsA_2$:MqsR), as determined using size exclusion chromatography (**Fig. 1B)** and confirmed using dynamic light scattering. Furthermore, deletion of *mqsA* is lethal [228, 257]; similar results have been found with other antitoxins, such as HigA of *Vibrio cholerae* [268]. MqsA-F is also sensitive to proteolysis (Fig. S2). Using electrophoretic mobility shift assays (EMSA), we demonstrate that both the MqsR:$MqsA_2$:MqsR complex and MqsA-F bind specifically to the *mqsR* promoter (P*mqsR*; Figs. 1C, 1D). It was also recently shown that MqsA binds two distinct palindromic sequences within P*mqsR* and that MqsA binding is enhanced in the presence of MqsR [267]. Finally, MqsR and MqsA are conserved, both in sequence and in gene structure, throughout the *gamma-delta-* and *epsilon proteobacterial* classes (Fig. S3). Taken together, these

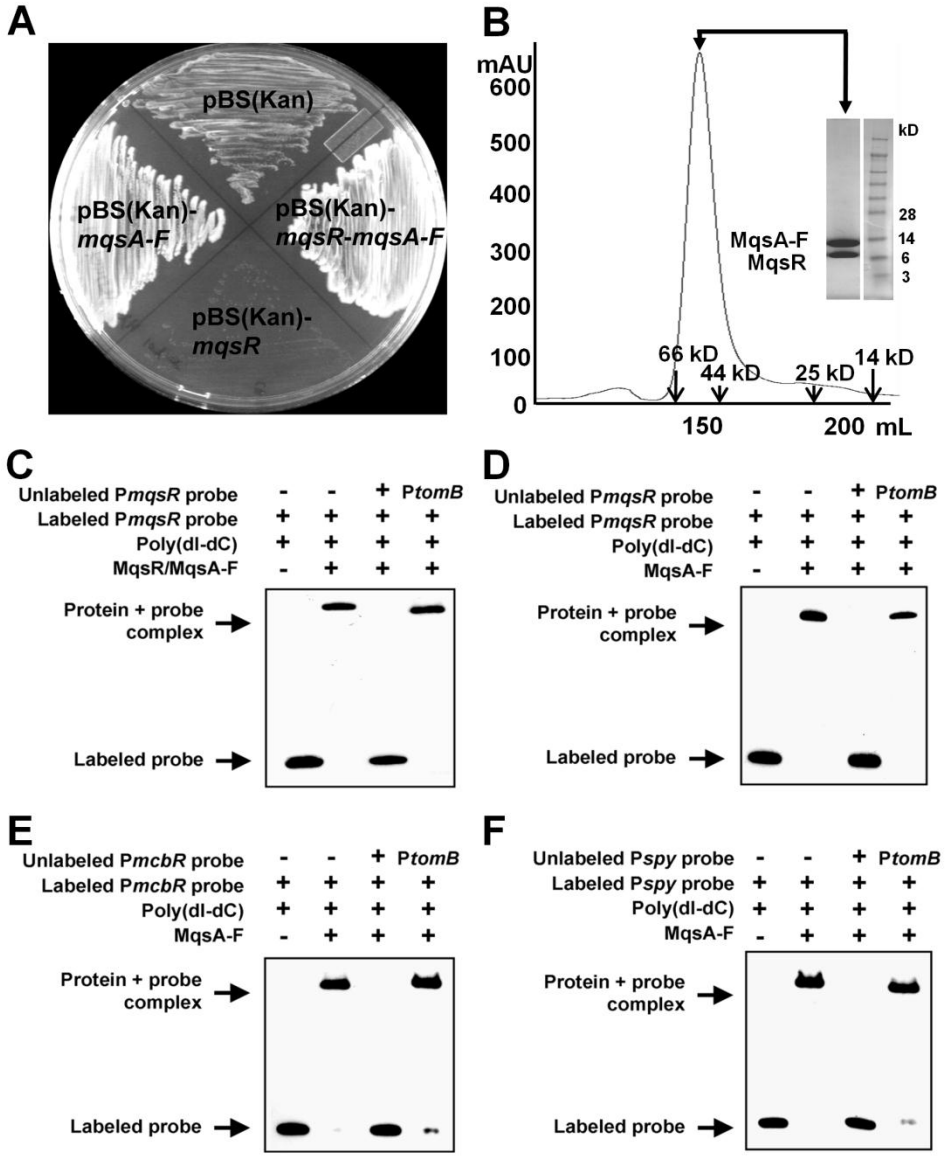


Figure 1

Figure 72: Figure 1, Brown and Grigoriu, PLoS Pathog. MqsR:MqsA are a *bona fide* toxin:antitoxin pair.

The effect of *mqsR*, *mqsA-F* and *mqsR-mqsA-F* on colony formation (A) for *E. coli* strain BW25113 containing pBS(Kan; empty plasmid), pBS(Kan)-*mqsR*, pBS(Kan)-*mqsA-F,* and pBS(Kan)-*mqsR-mqsA-F*. (B) Size exclusion chromatogram and SDS-PAGE gel of the MqsR:MqsA-F complex (Superdex 75 26/60 column; MqsA-F, 14.7 kD; MqsR, 11.5 kD). Elution positions of molecular weight standards are indicated by arrows. The complex elutes as a single peak at a position consistent with the formation of a dimer of dimers (two copies of MqsR and two copies of MqsA-F; expected MW of the complex = 52.4 kD). Electrophoretic mobility shift assays: both the MqsR:MqsA-F complex (C) and MqsA-F alone (D) bind the *mqsR* promoter; MqsA-F also binds the *mcbR* promoter (E) and the *spy* promoter (F). P*tomB* [8] is a control DNA fragment that we show in Figures S1F, S1G is not bound by either the MqsR:MqsA-F complex of MqsA-F.

results show that MqsR is a *bona fide* toxin and MqsA is the proteolytically sensitive antitoxin that blocks MqsR toxicity. Because MqsR:$MqsA_2$:MqsR was also recently identified to regulate the expression of a number of *E. coli* genes including one that encodes the colonic acid regulator McbR [255, 269], we reasoned that this regulation may be mediated by the direct binding of MqsA-F and/or the MqsR:$MqsA_2$:MqsR complex to the promoters of these genes. EMSA was used to show that MqsA-F binds specifically to the promoters of *mcbR* and *spy* (Figs. 1E, 1F, S1F, S1G).

C.5.b. Structure determination

The structure of full-length MqsA (MqsR-F; residues 1-131, the constructs used in this study are shown in Fig. S4) is shown in Figure 2A. MqsA-F was determined to a resolution of 2.15 Å by molecular replacement using the structures of the MqsA N-terminal domain (residues 1-76, referred to hereafter as MqsA-N; Fig. S5A) and the MqsA C-terminal domain (residues 62-131; referred to hereafter as MqsA-C; Fig. S5B), as search models. The space group of the MqsA-F crystal is $P2_1$, with two molecules in the asymmetric unit. The two monomers are identical in terms of the overall fold with well-ordered electron density throughout both chains. The final model contains 262 residues, and includes all 131 residues for each MqsA-F molecule.

The structure of the MqsR:MqsA-N complex (Fig. 2B) was solved to a resolution of 2.0 Å using the multiple wavelength anomalous dispersion method. The space group of the crystal is $P4_12_12$, with two crystallographically independent complexes in the asymmetric unit. The two complexes are identical

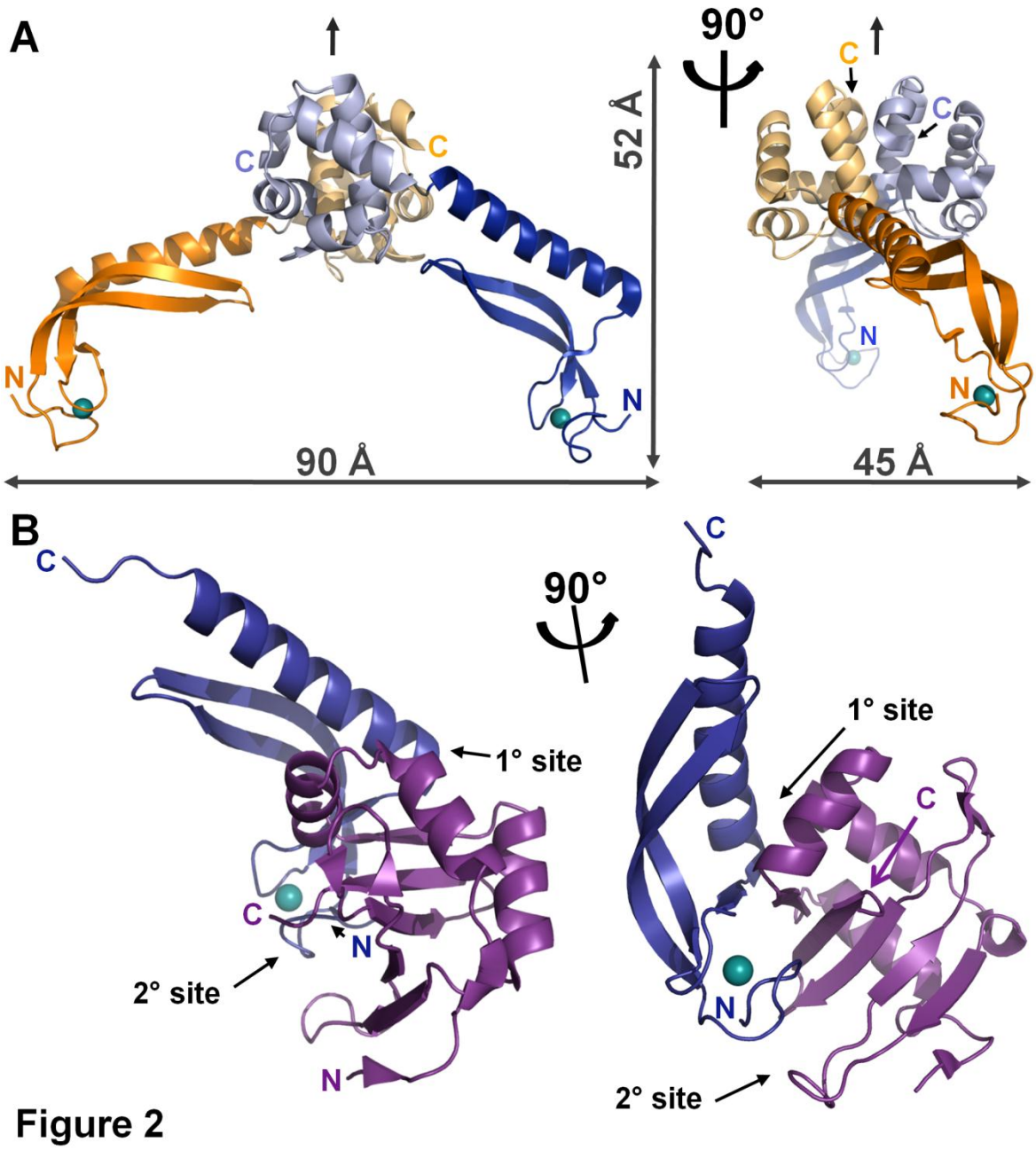


Figure 73: Figure 2, Brown and Grigoriu, PLoS Pathog. The structures of MqsA and MqsR. (A) Ribbon representation of MqsA-F dimer. The MqsA-F dimer is composed of two chains (blue and orange), each of which have two distinct N- and C-terminal domains (illustrated in dark and light shades, respectively). MqsA-F dimerization is mediated by the C-terminal HTH-XRE domains (MqsA-C; light blue and light orange), with the N-terminal zinc binding domains (MqsA-N; dark blue and dark orange) projecting away from the dimer interface; the coordinated zinc ion is shown as a teal sphere. The local two-fold axis relating the MqsA-F monomers to one another is indicated by an arrow. The dimensions of the dimer (45 x 52 x 90 Å) are shown. (B) Ribbon representation of the complex between MqsR (magenta) and MqsA-N (dark blue). The MqsR:MqsA-N interface is composed of two sites, the primary site (1°) and the secondary site (2°). In (A) and (B), the panel on the right is rotated by 90° with respect to that on the left about the axis shown.

in terms of the overall fold of each protein, with well-ordered electron density observed throughout the first complex and throughout the majority of the second complex. The final model contains 70 residues (1-70) of the 76 residues for MqsA-N and 97 residues (1-97) of the 98 residues for MqsR in the first complex and 59 residues (1-20, 27-65) of MqsA-N and 90 residues (2-60, 64-95) for MqsR in the second complex. X-ray diffraction data quality and refinement statistics are reported in Table 1.

C.5.c. MqsA is a structured antitoxin

The structure of the MqsA-F dimer is shown in Figure 2A. The two monomers are related to one another by local two-fold symmetry (Fig. 2A, arrow). The MqsA monomer is composed of two structurally distinct domains connected by a flexible linker and resembles a human leg (Fig. 3A): the MqsA-N, composed of residues 1-67, is the 'foot' and 'calf' and the MqsA-C, composed of residues 69-131, is the 'thigh'. The 'knee'-like linker connecting the domains, centered on residue 68, is flexible, allowing the N- and C-terminal domains to rotate independently of one another as rigid bodies (superposition of the N- and C-terminal domains from chains A and B give RMSD values of 0.35 Å and 0.58 Å, respectively). Superposition of the C-terminal domains from both monomers shows that the corresponding N-terminal domains are rotated by ~25° with respect to one another (Fig. S5C).

MqsA-N binds zinc and adopts a novel, elongated fold (Figs. 3A-3C). It is composed of one long five-turn α-helix, a twisted β-sheet, loops that connect the secondary structural elements and a coordinated zinc ion. The zinc, which serves

a structural and not catalytic role [270, 271], is coordinated by four cysteines (Cys3, Cys6, Cys37, Cys40) with an average sulfur-zinc distance of 2.35 Å (Figs. 3B, S6A). In spite of the low sequence identity (12% identity, 18% similarity) in the MqsA-N domain among different species, these zinc-coordinated cysteines are perfectly conserved (Fig. S3B), suggesting that all bacterial MqsA proteins adopt a similar fold.

The interaction between the two, long twisted β-strands and the five turn α-helix of the MqsA zinc binding domain is stabilized by an extended hydrophobic core composed of 11 residues: Ile18, Tyr20, Phe22, Leu29, Ile32, Tyr36, Met45, Phe53, Val57, Phe60 and Val64 (Fig. 3C**;** residues in beige). Although none of these residues are identical among the MqsA bacterial homologs (Fig. S3B), they are highly similar. A second cluster of hydrophobic residues is found near the zinc binding pocket and includes Met1, Met11, Leu35 and Ile44 (Fig. 3C; residues in orange). The DALI server [272] was used to identify proteins with similar folds. Although more than 269 hits were obtained (Z-scores from 3.2 to 2.0), they aligned to only the two long β-strands and the five turn α-helix; none had the combination of strands, helix and the zinc binding pocket observed in MqsA. Thus, to the best of our knowledge, MqsA-N is a zinc binding domain with a novel fold, hereafter referred to as the MqsA fold.

MqsA-C, the helix-turn-helix (HTH) domain, is composed of five tightly packed α-helices (Figs. 3A, 3D-3E), which bury a central hydrophobic core. This core is composed of eight residues, including Val69, Val77, Leu83, Phe92, Phe99, Tyr102, Pro109, and Leu117. Of these, all but one (Val69) are perfectly

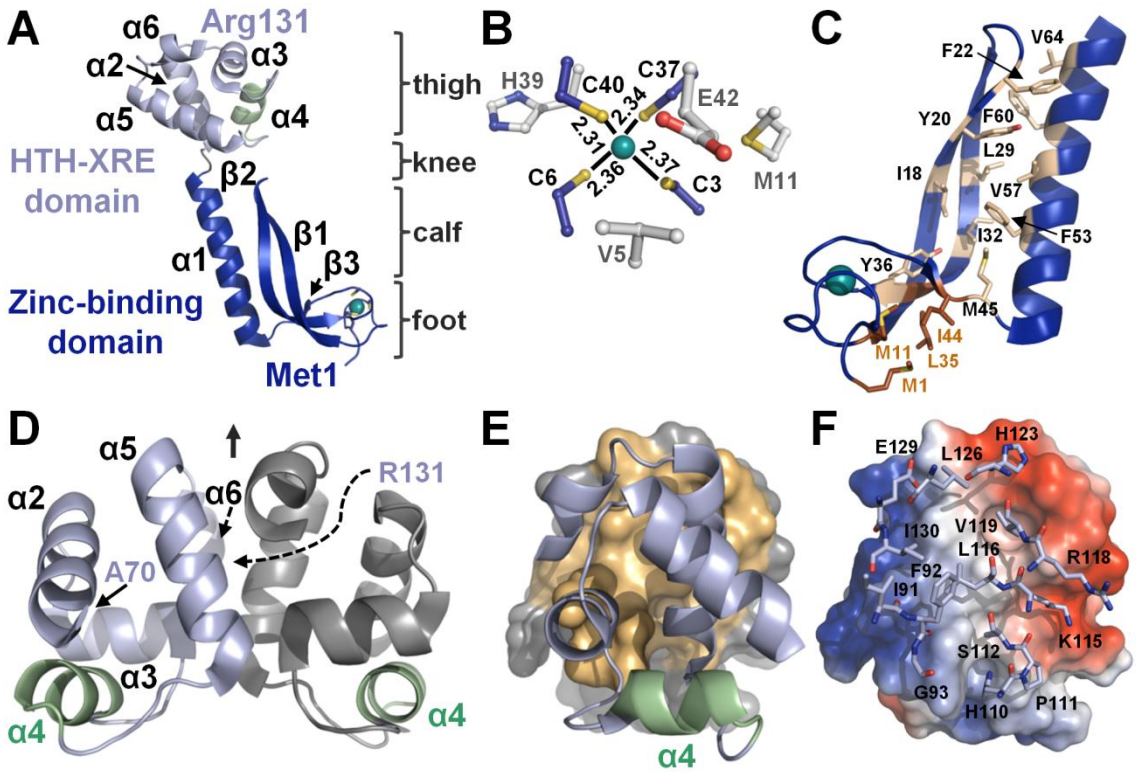


Figure 3

Figure 74: Figure 3, Brown and Grigoriu, PLoS Pathog. MqsA is a structured antitoxin. (A) The MqsA-F monomer can be visualized as a 'leg': the MqsA-C (the HTH-XRE domain; light blue) is the thigh, the flexible linker (centered on T68) is the knee and the MqsA-N (the zinc binding domain; dark blue) is the 'calf' and 'foot'; the zinc (teal sphere) is bound by the 'toes'. (B) The zinc is coordinated by Cys3, Cys6, Cys37 and Cys40 (blue/yellow sticks). Sulfur-zinc distances are commensurate with values typical for structural zinc binding sites. Residues near the zinc binding pocket are also shown (grey sticks). (C) The MqsA zinc binding domain adopts a novel fold characterized by a long, twisted β-sheet buttressed by a five-turn α-helix with the loops coordinated by zinc. It is stabilized by two hydrophobic cores (core 1, sidechains shown as sticks in beige; core 2, sidechains shown as sticks in orange). (D) The MqsA-C dimerization interface is composed of the MqsA-C α-helices 3, 5, and 6, with the two monomers (chain A, grey; chain B, light blue) related by local two-fold symmetry. The sequence most highly conserved among MqsA proteins (residues 98-105, green) is predicted to bind DNA (see text). (E) The MqsA-C dimerization interface rotated by 90° with one monomer represented in surface representation (1600 $Å^2$ of buried SASA illustrated in beige) and the second in ribbon representation. (F) Residues buried at the MqsA-C dimer interface shown as sticks (chain B) and the surface charge distribution (chain A; positive charge, blue; negative charge, red).

conserved among the MqsA family (Fig. S3B). Structural similarity searches demonstrated that MqsA-C shows significant homology to the bacteriophage 434 Cro repressor, the P22 C2 repressor and HigA antitoxin, placing it in HTH-XRE family of DNA binding proteins [273].

MqsA dimerization is mediated by the MqsA-C HTH-XRE domain. Residues from α-helices 3, 5, and 6 participate in the dimerization interface (Fig. 3D), which buries 1600 $Å^2$ of solvent accessible surface area (SASA; this constitutes 19.9% of the total SASA; Fig. 3E) and has a surface complementarity of 0.69, both of which are within the ranges expected for biologically relevant protein:protein interactions. Six residues are buried upon dimer formation, including Ser112, Lys115, Leu116, Val119, Leu126 and Ile130 (Fig. 3F). Five of these six residues are either identical or highly similar among the MqsA family (Fig. S3B). Eight additional residues (Ile91, Phe92, Gly93, His110, Pro111, Arg118, His123 and Glu129) participate, but are not fully buried, in the dimerization interface. The residues from α-helix 5 (Pro111, Leu116, Val119 and the aliphatic portions of the Lys115 and Arg118 sidechains) form a long hydrophobic pocket down the center of the face of the monomer, which is bordered on one side by negatively charged residues and on the other by positively charged residues (ig. 3F).

C.5.d. The MqsA antitoxin binds the MqsR toxin and DNA via distinct domains

In order to determine which domain of MqsA (MqsA-N, MqsA-C or both) binds MqsR, we carried out two co-expression toxicity experiments using the same

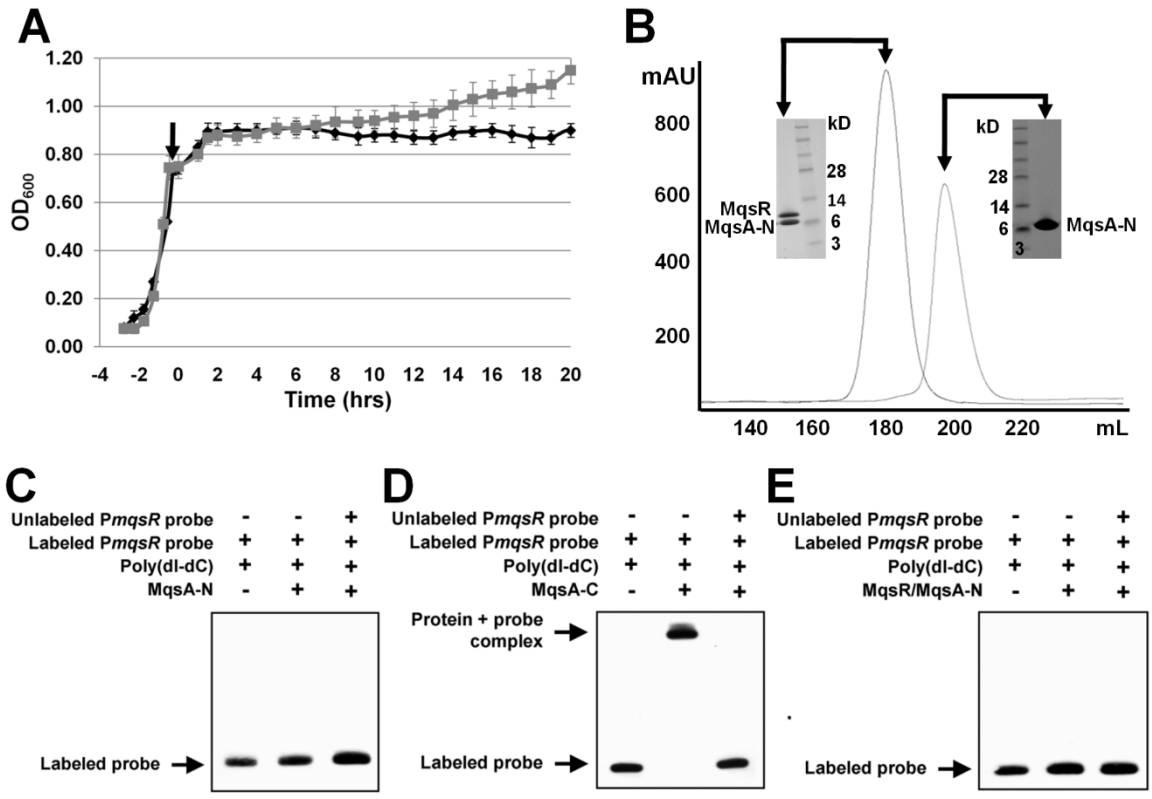


Figure 4

Figure 75: Figure 4, Brown and Grigoriu, PLoS Pathog. MqsA binds MqsR via its N-terminal zinc binding domain and DNA via its C-terminal HTH domain.
(A) Growth curves of the co-expression of MqsR with the N-terminal domain of MqsA (MqsR:MqsA-N, ■) and the C-terminal domain of MqsA (MqsR:MqsA-C, ♦) in BL21 (DE3) cells. Induction of protein expression using 0.5 mM IPTG corresponds to t=0 (arrow; expression carried out at 18°C, following the same protocol used to produce the proteins for structural studies). Co-expression of MqsR with MqsA-C leads to growth arrest while co-expression with MqsA-N results in robust growth. Data shown represents the average of the two measurements with the standard deviation shown as error bars. (B) Size exclusion chromatogram (SEC) of the co-expressed MqsR:MqsA-N complex (dark grey, left) and the SDS-PAGE gel of the pooled peak. As can be seen, both MqsR (11.5 kD) and MqsA-N (8.5 kD) are present. The MqsR:MqsA-N SEC, with corresponding SDS-PAGE gel, is overlaid with that of MqsA-N alone (light grey, right), illustrating the shift in the elution position of MqsA-N upon complex formation. (C-E) EMSAs using only the MqsA-N (C), only the MqsA-C (D) or the MqsR:MqsA-N complex (E) with the *mqsR* promoter.

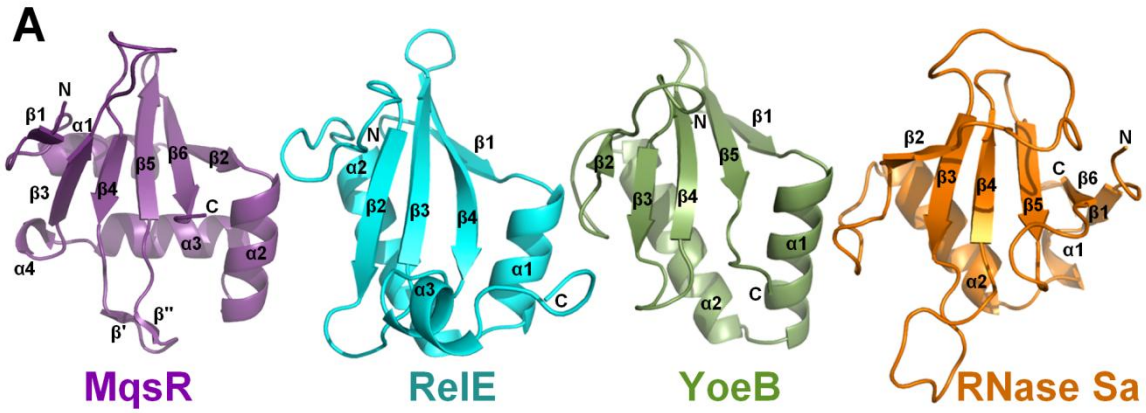
protocol as that used to produce the proteins for structural studies: 1) MqsR co-expressed with MqsA-N and 2) MqsR co-expressed with MqsA-C. When MqsR and MqsA-C are co-expressed, bacterial cell growth is arrested (Figs. 4A, S7A, ♦). In contrast, when MqsR and MqsA-N are co-expressed, cell growth is robust

and both MqsR and MqsA-N express to high levels (Figs. 4A, S7A, ■). Moreover, following co-expression, the MqsR:MqsA-N complex is readily purified and forms a heterodimer (one copy of MqsR and one copy of MqsA-N), as determined using size exclusion chromatography (Fig. 4B), and confirmed using dynamic light scattering.

EMSA was used to determine which domain of MqsA binds DNA. As shown in Figures 4C-E, the MqsA C-terminal domain is necessary and sufficient for binding the *mqsR* promoter (P*mqsR*), as incubation of P*mqsR* with MqsA-C results in a shift in the electrophoretic mobility of the DNA (Fig. 4D). In contrast, no shift is observed when the DNA is incubated with either MqsA-N alone or the MqsR:MqsA-N complex (Figs. 4C, 4E). These results show that DNA binding is mediated exclusively by MqsA-C.

C.5.e. MqsR adopts an RNase fold characteristic of the YoeB, RelE toxin family

MqsR is a small, globular protein, consisting of a central six-stranded β-sheet (β1-β3-β4-β5-β6-β2) and three α-helices, with α-helix 2 adjacent to and α-helices 1 and 3 abutting the backside of the β-sheet (Fig. 5A, left; magenta). A three-dimensional structure alignment revealed that MqsR is most similar to the bacterial toxins YoeB [6] and RelE/aRelE [12, 250] (DALI Z-scores = 5.1 and 4.0/5.3 respectively [272]), both of which are ribonucleases (RNases) and adopt a microbial RNase fold, the RelE-like fold [274] (Fig. 5A; PDBIDs: RelE, 2KC8; YoeB, 2A6Q; RNase Sa, 1RSN). The sequence identity of MqsR with YoeB and RelE/aRelE is extremely low, only 11% and 13%, respectively, and demonstrates



B

Protein	Secondary Structural Elements														
MqsR	--	β1	α1	β2	α2	--	α3	β3	β4	β'	β''	β5	β6	--	
RelE	--	--	--	β1	α1	α2	--	--	β2	--	--	β3	β4	α3	
YoeB	--	--	--	β1	α1	α2	--	β2	β3	--	--	β4	β5	--	
RNase Sa	β1	--	--		--	α2	--	β2	β3	--	--	β4	β5	--	β6

Figure 5

Figure 76: Figure 5, Brown and Grigoriu, PLoS Pathog. MqsR shares structural homology with the RelE and YoeB family of *E. coli* RNase toxins.
(A) MqsR (magenta) with labeled secondary structural elements, and those of its closest structural homologs YoeB (green; DALI Z-score=5.1; PDBID 2A6Q [6]) and RelE (cyan, DALI Z-score=4.0; PDBID 2KC8 [12]). RNase Sa (orange; PDBID 1RSN [14]), a canonical bacterial RNase, is also shown. The bacterial RNase fold is characterized by a central, twisted β-sheet and adjacent α–helix and is conserved among all four proteins. (B) Corresponding secondary structural elements of each protein; elements highlighted in green are structurally conserved between the proteins.

why structure determination was essential to identify MqsR as a member of this family. Critically, our finding that MqsR functions as a bacterial ribonuclease toxin was recently confirmed [267].

The RelE-like fold is characterized by a central, antiparallel β-sheet and adjacent α-helix, which is conserved among the MqsR, YoeB and RelE bacterial toxins (Fig. 5B; conserved secondary structural elements shaded in green). However, there are a few key differences. First, the β-sheet in MqsR is extended by one β-strand (β1) because MqsR has a longer N-terminus. Second, in contrast to YoeB, RelE and RNase Sa, which each have one α-helix that folds across the back of the central β-sheet at a 45° angle, MqsR has two α-helices,

both of which are perpendicular to the β-strands. This allows the long loop in MqsR that connects β-strands 4 and 5 to extend towards and interact with α-helix 2, an interaction that is sterically prohibited in YoeB and RelE.

C.5.f. MqsR potential functional site

In order to test which residues play a role in MqsR-mediated toxicity, we used alanine-scanning mutagenesis of evolutionarily and structurally conserved residues. Toxicity was measured by monitoring MqsR-mediated growth arrest and protein expression levels. The MqsR mutants that exhibited the most robust growth (i.e., those with the least toxicity) were K56A, Q68A, Y81A and K96A (Figs. 6A, S7B). The mutant proteins also expressed and could be detected by Western Blot using an antibody directed to the MqsR his_6-tag (expression of wild-type, WT, MqsR arrests cell growth so rapidly that free MqsR is not detectable, even by Western Blot; Fig 6B). In contrast, growth curves for MqsR mutants Y55A, M58A and R72A were similar to WT and, like WT, did not express to detectable levels. This was also observed for MqsR mutants H7A, H64A and H88A (not shown). Thus, these results show that MqsR residues K56, Q68, Y81 and K96 play key roles in MqsR-mediated toxicity while residues H7, Y55, M58, H64, R72 and H88 do not.

C.5.g. MqsA recognition and neutralization of the MqsR toxin

MqsA recognition of MqsR occurs at two distinct interfaces. The primary interface buries 1537 $Å^2$ of SASA while the secondary interface buries 479 $Å^2$ of SASA. A total of 2016 $Å^2$ of SASA is buried upon complex formation, or 17.7% of the total SASA, well within the range expected for biological interfaces. The primary

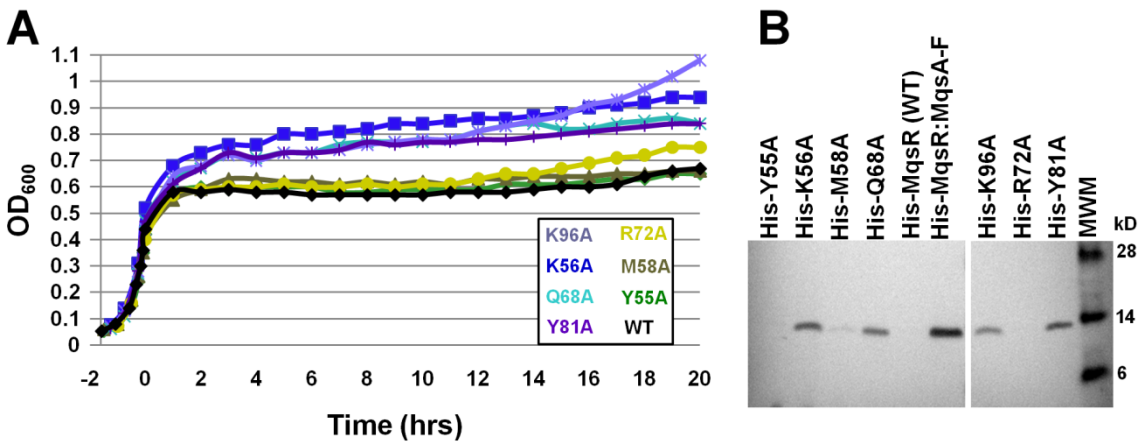


Figure 6

Figure 77: Figure 6, Brown and Grigoriu, PLoS Pathog. Identification of the MqsR functional site.
(A) Growth curves of cells over-expressing WT MqsR and seven MqsR mutants. MqsR mutants K56A, Q68A, Y81A and K96A show decreased toxicity compared to WT MqsR, as evidenced by their ability to grow following induction with 0.5 mM IPTG at t=0 (arrow; expression carried out at 18°C, following the same protocol used to produce the proteins for structural studies). In contrast, induction of MqsR mutants Y55A, M58A and R72A, like WT, lead to growth arrest. (B) Western blots showing soluble expression of MqsR mutants K56A, Q68A, Y81A and K96A from the cultures used to generate the growth curves in (A) (WT MqsR arrests cell growth so rapidly that free MqsR is not detectable by Western Blot). All MqsR constructs include an N-terminal his_6-tag and protein expression was detected using a polyhistidine polyclonal antibody. Like WT, expressed protein for MqsR mutants Y55A, M58A and R72A was not detected.

interface is centered on MqsA β-strand 3 (Ser43-Met45), which interacts with MqsR β-strand 2 (Val22-Thr25) to form a single continuous β-sheet ($\beta1_R$-$\beta3_R$-$\beta4_R$-$\beta5_R$-$\beta6_R$-$\beta2_R$-$\beta3_A$-$\beta2_A$-$\beta1_A$; Figs. 2B, 7A) throughout the complex. Sixteen residues from MqsA-N and 12 residues from MqsR contribute to the primary interface. The interaction is hydrophobic, with Pro4, Ser43, Ile44, Met45, Ser50, Phe53, and Met54 from MqsA (light blue sticks, Fig. 7A) and Thr24, Thr25, Arg26, Leu29, Phe39 and Ile92 from MqsR (light pink sticks, Fig. 7A) becoming completely buried upon complex formation. In addition, three key electrostatic interactions are located at the periphery of the interface: (1) Arg26 from MqsR

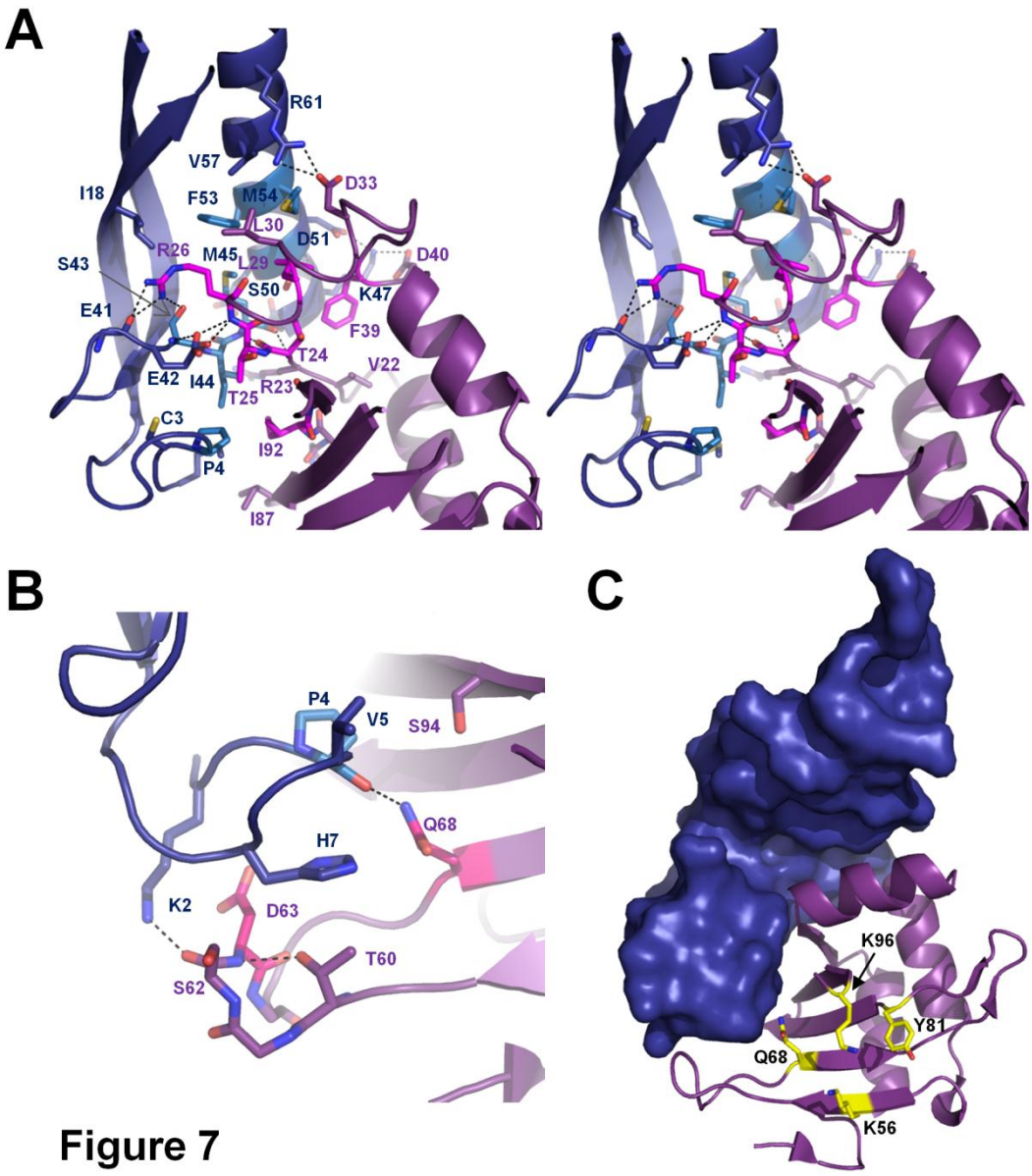


Figure 7

Figure 78: Figure 7, Brown and Grigoriu, PLoS Pathog. MqsA recognition of MqsR.
(A) Stereoimage of the 1° interaction site between MqsR (magenta) and MqsA-N (dark blue; see also Fig. 2B), which buries 1537 Å^2 of solvent accessible surface area (SASA). The residues that participate in the interaction are shown as sticks. Those that are completely buried upon complex formation (>80% loss of SASA) are shown in pink (MqsR) and light blue (MqsA-N). The interface is centered on MqsA-N residue M45 and is predominantly hydrophobic. Electrostatic interactions, including the hydrogen bonds formed between MqsA-N β-strand 3 (S43-M45) and MqsR β-strand 2 (V22-T25) that function to extend the MqsR β-sheet with that of MqsA-N ($\beta1_R$-$\beta3_R$-$\beta4_R$-$\beta5_R$-$\beta6_R$-$\beta2_R$-$\beta3_A$-$\beta2_A$-$\beta1_A$), are indicated by dashed lines. (B) The 2° interaction site buries 479 Å^2 of SASA; colors and electrostatic interactions shown as in (A). (C) The MqsR:MqsA-N complex, with MqsA-N in surface representation (dark blue) and MqsR as a ribbon diagram (magenta). The residues that play a role in MqsR-mediated toxicity (Fig. 6) are shown as sticks in yellow. With the exception of Q68, the residues that play a role in toxicity are accessible when MqsR is bound to MqsA-N.

forms hydrogen bonds with the hydroxyl sidechain of Ser43 and the carbonyl of Glu41 from MqsA-N, (2) Asp33 from MqsR forms a salt bridge with Arg61 from MqsA-N and (3) Asp40 from MqsR forms a salt bridge with Lys47 from MqsA-N. The secondary interface is much smaller, being formed by four residues from MqsA-N and five residues from MqsR (Figs. 7B, S6B). The interface is centered on His7 (MqsA-N), which interacts with Thr60, Ser62, Asp63 and Gln68 from MqsR. Additional interactions are observed between Lys2, Val5 and the carbonyl of Pro4 from MqsA-N and Asp63, Gln68 and Ser94 of MqsR. MqsR residues Asp63 and Gln68 are the only residues that become buried in the secondary interface (light pink, Fig. 7B). Finally, MqsA-N does not block the MqsR active site (Fig. 7C). Instead, nearly all of the residues predicted to be important for MqsR activity are accessible in the MqsR:MqsA-N complex. This accessibility of the active site has also been observed for the RelBE system [250]. The only exception is Gln68, which is part of the secondary interface. This suggests neutralization is achieved either through cellular localization (i.e., towards DNA via MqsA-C) or through steric occlusion by blocking the interaction of MqsR with other biomolecules, such as the ribosome.

C.6. Discussion

Our biochemical and structural analysis of the *E. coli* MqsA antitoxin and the MqsR:MqsA-N complex provide a detailed 3-dimensional structural view of the free antitoxin and the TA complex. These structures combined with our biochemical data unequivocally demonstrate that MqsR, a protein previously shown to be critical for biofilm formation [255, 256] and the most highly

upregulated gene in persister cells [228] in *E. coli*, and MqsA form a novel bacterial TA system. Our data show that, as expected for a TA system, the expression of the MqsR toxin leads to growth arrest, while co-expression with its antitoxin, MqsA, rescues the growth arrest phenotype. In addition, MqsR associates with MqsA to form a tight, non-toxic complex and both MqsA alone and the MqsR:$MqsA_2$:MqsR complex bind and regulate the *mqsR* promoter. Finally, the structure of MqsR reveals that it is a member of the RelE/YoeB family of bacterial RNases, which are structurally and functionally characterized bacterial toxins.

Comparison of the microbial RNase active sites between MqsR, RelE, YoeB and RNase Sa demonstrates that MqsR is most similar to RelE (Fig. 8A). In microbial RNases, such as RNase Sa, RNA binding is mediated by polar (Q38) and aromatic (Y86) residues while RNA cleavage is catalyzed by a histidine (H85) and glutamic acid (E54) residue, which function as a general acid and base, respectively [14, 275]. While the catalytic histidine and glutamic acid are conserved in YoeB (H83, E46), which has intrinsic endoribonuclease activity [6], they are not found in RelE, which functions as a ribosome-dependent RNase [12, 245]. These catalytic residues are also not present at the MqsR functional site (Figs. 7C, 8B, 8C). This is further supported by our demonstration that single deletion mutants of the three histidines in MqsR (H7, H64, H88; none of which overlap with the positions of H83 in YoeB and/or H85 in RNase Sa) do not attenuate MqsR-mediated toxicity. Instead, the MqsR residues that play a role in MqsR-mediated toxicity (K56, Q68, Y81 and K96) overlap best with those that

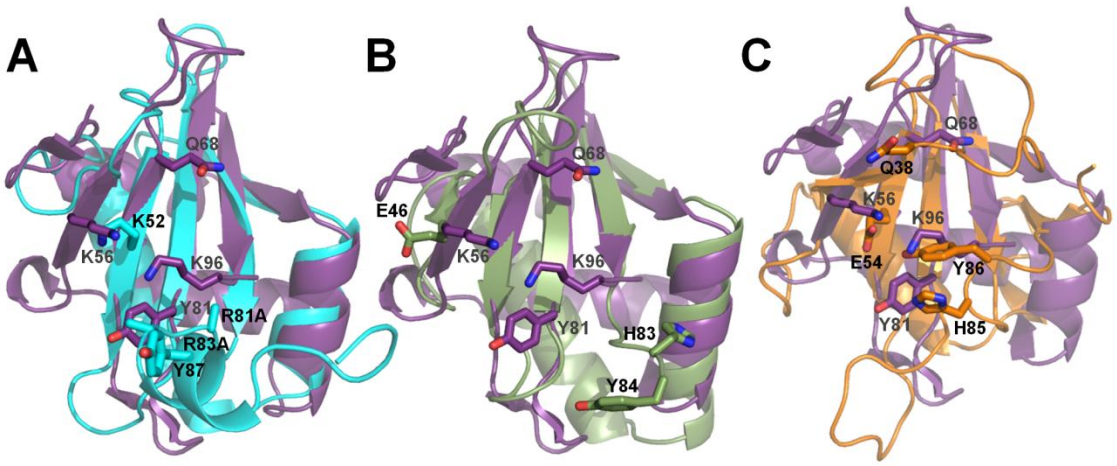


Figure 8

Figure 79: Figure 8, Brown and Griogriu, PLoS Pathog. MqsR is most similar to the bacterial toxin RelE.
Superposition of MqsR (magenta) with RelE (cyan; A), YoeB (green; B) and RNase Sa (orange; C) using corresponding β-strands (MqsR, β4/β5; RelE, β2/β3; YoeB, β3/β4; RNase Sa, β3/β4). Residues in RelE, YoeB and RNase SA that play a role in toxicity are shown as sticks with black labels while those in MqsR are shown as sticks with grey labels. MqsR lacks the histidine and glutamic acid residues that function as the catalytic acid and base in YoeB (E46, H83) and RNase Sa (H85, E54).

are important for RelE-mediated toxicity (see K52/K56, R81/K96, R83/K96, Y87/Y81 between RelE/MqsR, respectively; Fig. 8A) [12]. This strongly suggests that MqsR, like RelE, is a ribosome-dependent RNase. This structural and functional data contradict recent results that show MqsR cleaves mRNA in the absence of the ribosome [267] and demonstrates that the precise catalytic mechanism by which mRNA is cleaved by MqsR remains to be elucidated.

However, in sharp contrast to established TA pairs such as RelE:RelB, YoeB:YefM and MazF:MazE, the MqsR:MqsA TA pair has many unique characteristics that are not observed in canonical TA systems, and thus represents the founding member of a new family of TA systems. In typical TA pairs, the antitoxin gene precedes that of the toxin. In contrast, *mqsR* precedes *mqsA*. To date, this genetic organization has only been observed in two other

recently characterized TA systems, that of *higBA* [258] and *hicAB* [254]. In addition, in canonical TA systems, the toxin is larger than the antitoxin (with the exception of HicB [254]) and the toxin is basic while the antitoxin is acidic. In the MqsR:MqsA TA system, MqsA is larger than MqsR, 14.7 kD and 11.2 kD, respectively, and both proteins are basic.

Critically, one of the most significant differences between MqsR:MqsA and typical TA systems is the nature of the MqsA antitoxin itself. All other canonical antitoxins whose structures are known, including HipB [246], RelB [250], YefM [6, 276] and MazE [251], have at least one dynamic, flexible domain and all of these, with the exception of HipB, either change conformation or become ordered upon toxin binding. In contrast, MqsA is well-ordered throughout its entire sequence and its structure does not change when bound to MqsR. This explains why MqsA must be larger than MqsR; MqsA binds MqsR via a longer, folded domain, while other antitoxins bind their corresponding toxins via shorter, unstructured peptides. In addition, all other functionally characterized antitoxins bind DNA via their N-terminal domains. MqsA is the first antitoxin that has been shown experimentally to bind DNA via its C-terminal domain. The only other antitoxin predicted to bind DNA via its C- and not N-terminal domain is HicB [254, 277]. MqsA is also the first antitoxin described that requires a metal, zinc, for structural stability. Moreover, unlike most other TA inhibition mechanisms, MqsA (like its RelB homolog [250]), does not occlude the toxin active site. Finally, in addition to binding its own promoter, MqsA and the MqsR:$MqsA_2$:MqsR complex also bind and regulate the promoters of genes that play roles in *E. coli* physiology,

including *mcbR* and *spy*. To the best of our knowledge, this is the first time a TA system has been shown to bind the promoters of genes other than its own.

The MqsR:$MqsA_2$:MqsR complex is oligomeric, and forms a dimer of dimers (MqsR:$MqsA_2$:MqsR). Since the structure of the MqsA-N is invariant among all structures determined (MqsA-F; MqsA-N and the MqsR:MqsA-N complex), we used the MqsA-F and MqsR:MqsA-N structures to generate an accurate model of the full MqsR:$MqsA_2$:MqsR complex. As can be seen in Figure 9A, the MqsR:$MqsA_2$:MqsR complex forms a highly extended structure with the MqsA-C domains located at the center of the complex (MqsA monomers in light blue and light orange; α-helix 4 in green) and the MqsR toxins (magenta) bound at the periphery. In overall shape, the MqsR:$MqsA_2$:MqsR complex is most similar to the MazF:MazE complex [251], which is also highly extended, although the structures of the individual proteins and oligomerization states of the complexes between the two families differ dramatically.

We and others [267] used EMSA to show that MqsA-F, MqsA-C and the MqsR:$MqsA_2$:MqsR complex bind directly to the *mqsR* promoter (Figs. 1C, 1D, 4D). Therefore, MqsA likely is a transcriptional regulator of its own promoter via direct binding interactions through MqsA-C. The electrostatic surface of the MqsR:$MqsA_2$:MqsR complex is shown in Figure 9B. As can be readily observed, the bottom of the MqsA-C and by direct extension, the top of the MqsA-N are nearly exclusively positively charged, typical for a strong DNA binding site. As we showed experimentally, DNA binding is localized to MqsA-

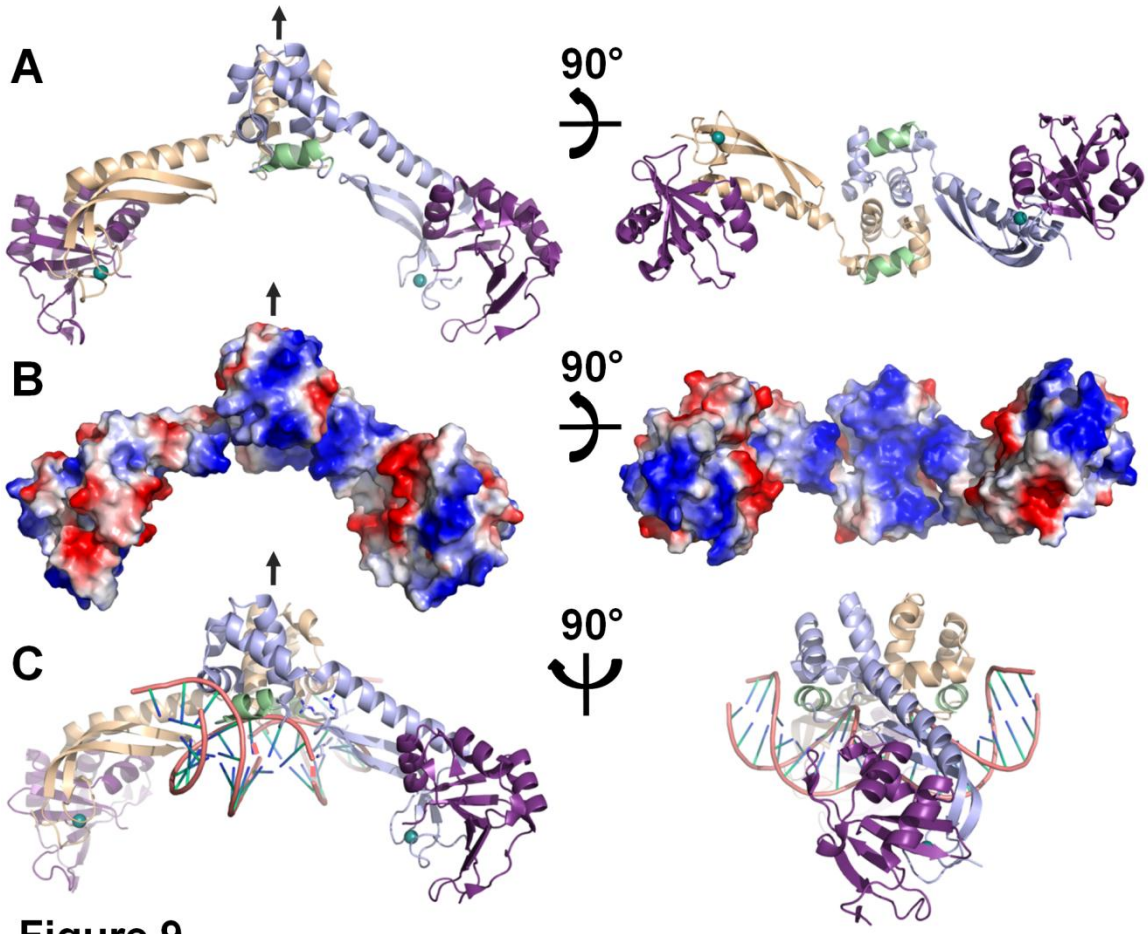


Figure 9

Figure 80: Figure 9, Brown and Grigoriu, PLoS Pathog. Model of the MqsR:MqsA2:MqsR:DNA complex.

(A) The MqsA-F and MqsR:MqsA-N structures were used to generate the full MqsR:MqsA2:MqsR complex. The MqsA monomers are shown in light blue and light orange, with conserved MqsA-F α–helix 4 in green. MqsR is in magenta. The local two-fold symmetry axis is indicated by a black arrow. Right and left panels are rotated by 90° about the horizontal axis. (B) The electrostatic surface of the MqsR:MqsA2:MqsR complex (positive charges, blue; negative charges red). The right panel is rotated about the horizontal axis, illustrating the extensive positively charged surface at the bottom of the C-terminal HTH-XRE domain. (C) Model of the MqsR:MqsA2:MqsR:DNA complex, which is based on the observation that all HTH-XRE DNA binding domains bind DNA using the same helix. In MqsA, this corresponds to α-helix 4. Model generation described in the methods. Because the dimerization interface differs between MqsA and the P22 C2 repressor [9], the DNA bound by MqsA must bend (illustrated by dashed lines between the two DNA segments). Two views of the MqsR:MqsA2:MqsR:DNA complex are shown. The left panel is the same orientation as that shown in the left panel of A. The image is rotated about the vertical axis to illustrate that very little of the MqsA-N zinc binding domain, which binds MqsR, interacts with the bound DNA. The pseudo two-fold symmetry axis is indicated by an arrow.

C. This enabled us to exploit the similarity of the MqsA-C HTH-XRE domain to other experimentally determined HTH-XRE protein structures to generate a model for DNA binding by the MqsR:$MqsA_2$:MqsR complex. All HTH-XRE DNA binding proteins whose structures have been determined in complex with DNA and which are structurally homologous to MqsA (16 protein:DNA complexes with DALI Z-scores ranging from 6.0 to 2.5), bind DNA via the same helix, the HTH-XRE DNA binding helix. In MqsA, the HTH-XRE DNA binding helix corresponds to α-helix 4 (Fig. 9A, α-helix 4 highlighted in green). Importantly, in the MqsA-F dimer, this α-helix is accessible to solvent and thus, by extension, to DNA. It is also highly positively charged, decisively contributing to the positively charged surface of the MqsA-C HTH-XRE domain dimer. Moreover, MqsA α-helix 4 is the only stretch of residues that is perfectly conserved among all MqsA proteins (Fig. S3B). Taken together, these data provide strong evidence that MqsA binds DNA via α-helix 4.

We generated a model of the MqsR:$MqsA_2$:MqsR:DNA complex using the coordinates of the P22 C2 repressor bound to DNA (PDBID 2R1J) [9], the HTH-XRE protein identified to be most similar to MqsA, and the MqsR:$MqsA_2$:MqsR complex (Fig. 9C). As can be seen, the DNA binds between the MqsA-N domains and makes extensive interactions with the bottom surface of the MqsA-C dimer, with α-helix 4 of both monomers projecting into the major groove of DNA. In contrast, there is minimal interaction between the DNA and MqsA-N and no interaction between the DNA and MqsR. This is perfectly consistent with the EMSA results, which show that neither MqsA-N nor MqsR:MqsA-N complex bind

the *mqsR* promoter (Figs. 4C, 4E). Notably, the MqsR:$MqsA_2$:MqsR:DNA complex reveals that residues at the turn connecting β-strands 1 and 2 in the MqsA N-terminal domain may interact with DNA. In support of this hypothesis, these residues, ^{23}RGRK^{26}, are highly basic and thus are capable of forming ionic interactions with the phosphate backbone. Thus, although MqsA-N is not capable of binding DNA alone, it may contribute to DNA binding via the interactions of these residues. This also suggests that the ability of the MqsA N- and C-terminal domains to rotate independently of one another (Fig. S5C) may help facilitate DNA binding as it would enable the MqsA-N domains to reposition themselves so residues ^{23}RGRK^{26} may optimally coordinate the phosphate backbone. Since the MqsR:$MqsA_2$:MqsR complex has been shown to bind DNA more tightly than MqsA alone [267], the interaction of MqsR with MqsA-N may also affect this repositioning. Finally, the model also predicts that MqsA binding induces a 30°-45° bend in the DNA. Although α-helix 4 is connected to α-helix 3 by a glycine-rich turn (^{93}GGG^{95}), and thus may be able to adopt a slightly different conformation when bound to DNA, it is highly unlikely that it could shift sufficiently to permit binding to unbent DNA.

Taken together, the crystal structures of full-length MqsA and the MqsR:MqsA-N complex reveal that the MqsR:MqsA TA complex is the founding member of a novel family of TA pairs. It forms a dimer of dimers in which DNA binding and MqsR recognition by MqsA are mediated via distinct, structured domains. Because MqsR is the gene most highly upregulated in *E. coli* persister cells [228] and because it also plays an essential role in biofilm regulation and

cell signaling [255, 256], these structures provide fundamental new insights into how this novel TA system participates in bacterial persistence, biofilm formation and multidrug tolerance in the *gamma*- *delta*- and *epsilon proteobacterial* classes. This work has provided the first steps for understanding this novel, unique TA system at a molecular level and provides a basis for developing novel antibacterial therapies that target TA pairs.

C.7. Supplementary Protocols

C.7.a. BW25113/MG1655 growth assays

The bacterial strains and plasmids used in this study are listed in Table S1. Growth experiments with *E. coli* strain BW25113 and MG1655 were conducted in LB medium at 37°C. Growth experiments were monitored using pBS(Kan)-based plasmids [259]. To construct pBS(Kan)-based plasmids for producing MqsR, MqsA-F, and MqsR-MqsA-F from a *lac* promoter, the fragments from genomic DNA were amplified by PCR (Table S2) and directionally cloned into pBS(Kan). The toxicity of selected proteins was investigated using pBS(Kan) plasmids with 1 mM IPTG added upon inoculation. Growth measured using culture optical density at 600 nm (OD_{600}). Cell viability (CFU) measured by diluting cells from 10^2 to 10^7 via 10-fold serial dilution steps into 0.85% NaCl solution and applying them as 10 µL drops on LB agar with kanamycin or chloramphenicol [260]. Two independent cultures were used for each strain. Data shown (Figs. 1, S1) represents the average of the two measurements with the standard deviation shown as error bars.

C.7.b. Proteolytic Digestion

MqsA-F was incubated at 30°C with trypsin, chymostrypsin, or proteinase K at a 1000:1 ratio (100 µg of protein:100 ng protease) for 5, 15, 30, or 60 min. The digestion reactions were stopped by heating the samples at 100°C and cleavage was identified using SDS-PAGE analysis. Proteolytic digestion of a stable, folded domain (SPAR PDZ domain [278]) and an intrinsically unstructured protein (DARPP-32 [279]) were carried out similarly. Cleavage products at 6 and 7 kDa were excised from the SDS-PAGE gel, prepared according to standard protocols and analyzed using LC-MS/MS (LTQ mass spectrometer; Ottawa Institute of Systems Biology Proteomics Facility) coupled with data analysis using Mascot.

C.7.c. Mutagenesis of MqsA-C

Mutagenesis of MqsA-C L81M was carried out using the Quikchange Mutagenesis Kit (Stratagene) using manufacture protocols. Protein expression in selenomethionine medium was carried out as previously described [280].

Crystallization of MqsA-N (residues 1-76) and MqsA-C (residues 62-131)

MqsA-N: Crystals were obtained using two-sequential rounds of seeding. Initially, MqsA-N formed thin needle-like crystals in 0.1 M CHES pH 9.5, 20% (w/v) PEG8000 using the sitting drop vapor diffusion method. These crystals were crushed, diluted and used as crystallization microseed solution (seed bead; Hampton Research) for subsequent crystallization trials. Crystals were obtained in a new condition, which contained 0.1 M Tris pH 8.5, 0.2 M Li_2SO_4, 30% (w/v) PEG4000 (0.3 µL protein, 0.2 µL crystallization condition, and 0.1 µL seed solution; sitting drop vapor diffusion). These crystals, which also formed thin

needles, were then crushed, diluted and used as crystallization microseed solution for a second round of seeding which produced the MqsA-N crystals used for data collection. Final crystals were grown in 0.1 M Tris pH 8.1, 0.2 M Na_2SeO_4, 30% (w/v) PEG4000 using the microbatch method at 22°C. Crystals were cryoprotected by a 30 second soak in mother liquor supplemented with 25% glycerol and immediately flash frozen in liquid N_2. *MqsA-C*: Crystals were obtained at 4°C in 0.1 M Tris pH 8.5, 5% (w/v) PEG 8000, 10% (v/v) glycerol, 20% (v/v) PEG300 using sitting drop vapor diffusion with 2:1 v/v ratio of protein to crystallization condition. MqsA-C L81M SeMet-derivative protein crystallized in 0.1 M HEPES pH 7.5, 0.2 M NaCl, 40% (w/v) PEG300 at 4°C. Both crystallization conditions for native and SeMet crystals were sufficient for cryoprotection and crystals were frozen by direct transfer to liquid nitrogen.

C.7.d. Data Collection and Structure Determination of MqsA-N and MqsA-C

All data were collected at the National Synchrotron Light Source Beamline X6A at 100 K using an ADSC Q270 (MqsA-N) or Q210 (MqsA-C) CCD detector and processed using HKL2000 [262]. *MqsA-N*: The crystal structure of MqsA-N was solved using single wavelength anomalous dispersion (SAD). Experimental phases for MqsA-N were determined using HKL2MAP with SHELXC/D/E [281]. One heavy atom was identified, which, after automated model building with the program ARP/wARP [282, 283], was determined to be a metal coordinated by the four N-terminal domain cysteines. X-ray wavelength scans at the K-α_1 absorption edges of zinc, nickel and cobalt confirmed that the coordinated metal was zinc. The MqsA-N was refined against the data collected at the high energy

remote wavelength. The model was completed with iterative rounds of model building using Coot [264] followed by restrained refinement in REFMAC 5.2.0019 [265] with a TLS model determined using the TLSMD server [284]. Residues and sidechains for which no density was observed were not modeled. Data collection, model and refinement statistics are reported in Table S3 and the structure is illustrated in Figure S5A. The final MqsA-N structure was refined to 1.7 Å with an R-factor of 16.5% (R_{free} = 18.6%) and contains MqsA residues 1-67, one zinc ion, and 97 water molecules; residues 68-76 were not observed in the electron density. The Matthews coefficient (V_m) for MqsA-N is 2.48 $Å^3$/Da. and the estimated solvent content is 50.4%. One molecule of MqsA-N was present in the asymmetric unit. Analysis of the crystal contacts was carried out using AREAIMOL (which measures the buried solvent accessible surface area) [285] and SC (which measures the surface complementarity of protein:protein interfaces) [286], and confirms that the oligomerization state of MqsA-N is monomeric. Structure validation and stereochemistry analysis was performed with Molprobity [110] and SFCHECK [266].

MqsA-C: The structure of MqsA-C was solved using the multi-wavelength anomalous dispersion (MAD) method. Data collected at the peak, inflection, and remote wavelengths were indexed, integrated, and scaled using HKL2000 [262]. Experimental phases were determined using SOLVE/RESOLVE [287]. Automated model building was performed with ARP/wARP [282, 283] with 10 cycles of autobuilding and REFMAC [265] refinement. The model was refined against native data. The model was completed with iterative rounds of model

building using Coot [264] followed by anisotropic restrained refinement in REFMAC 5.2.0019 [265]. Residues and sidechains for which no density was observed were not modeled. Data collection, model and refinement statistics are reported in Table S4 and the structure is illustrated in Figure S5B. The final structure of MqsA-C was refined to 1.4 Å with an R-factor of 15.5% (R_{free} = 18.2%). The model includes MqsA residues 66-131 and 80 water molecules; residues 62-65 were not observed in the electron density map. Strong difference density (>7σ) was observed suggesting Gln108 is methylated, which we confirmed using LC-MS/MS (Thermo LTQ MS; Brown University Proteomics facility). The Matthews coefficient (V_m) for MqsA-C is 2.18 $Å^3$/Da and the estimated solvent content is 43.5%. One molecule of MqsA-C was present in the asymmetric unit. Analysis of the crystal contacts using AREAIMOL [285] and SC [286] confirmed that the oligomerization state of MqsA-C is dimeric, with the monomers of the dimer related by crystallographic 2-fold symmetry. Structure validation and stereochemistry analysis was performed with Molprobity [110] and SFCHECK [266].

C.8. Supplementary Figures S1-S7

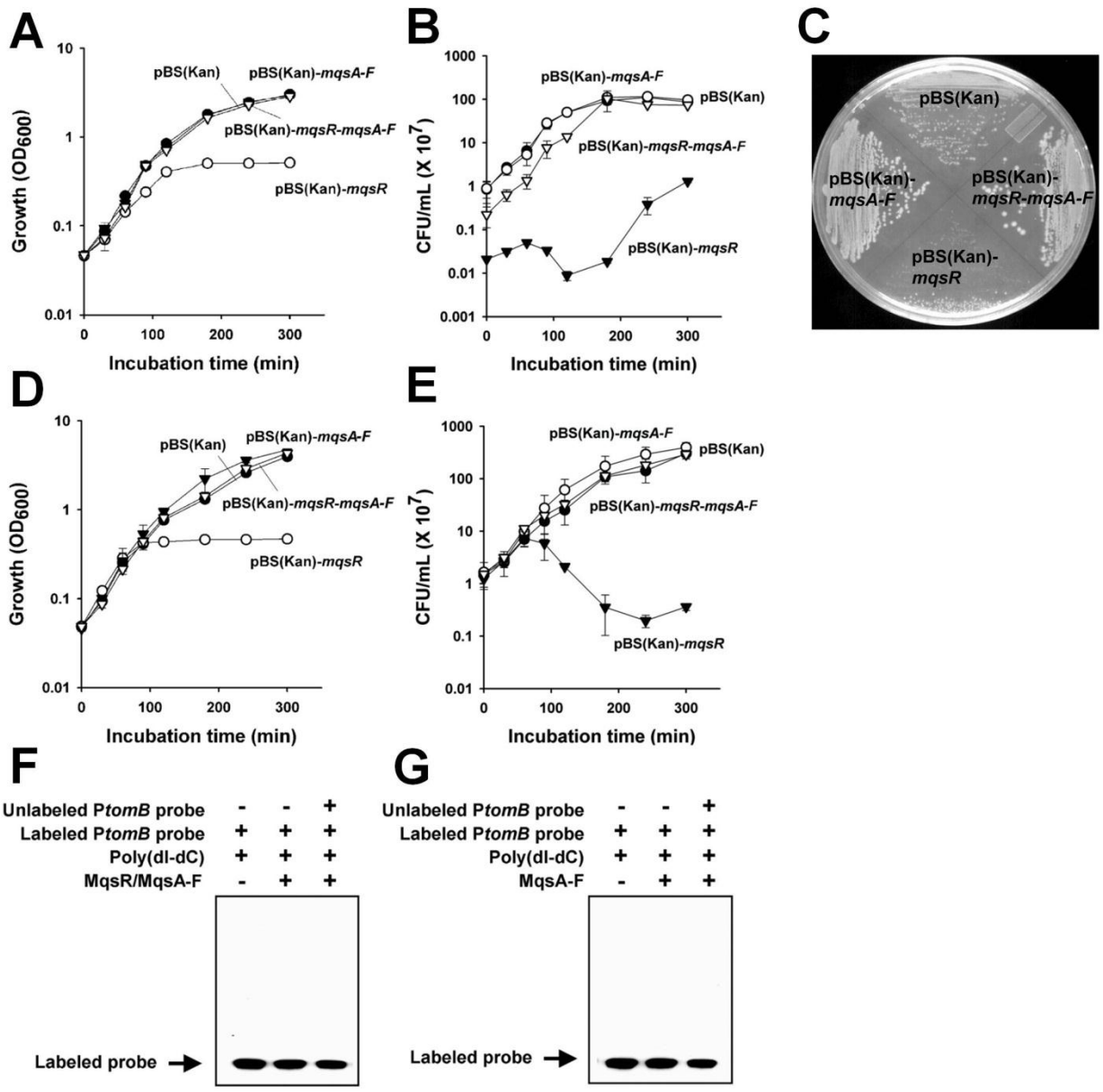


Figure 81: Figure S1, Brown and Grigoriu, PLoS Pathog. MqsR and MqsA are a Toxin:Antitoxin Pair.

The effect of MqsR, MqsA-F and the MqsR:MqsA-F complex on cell growth (A, D), cell viability (CFU/ml) (B, E) and colony formation (C) for *E. coli* strain MG1655 (A-C) and BW25113 (D,E) containing pBS(Kan) (●,empty plasmid), pBS(Kan)-*mqsA-F* (○), pBS(Kan)-*mqsR* (▼) and pBS(Kan)-*mqsR-mqsA-F* (▽) at 37° C in LB with 1 mM IPTG induction upon inoculation (incubation time = 0 min). Electrophoretic mobility shift assay controls that show both MqsR:MqsA-F (F) and MqsA-F (G) do not bind the P*tomB* DNA probe, which was used as a non-specific competitive control for the EMSA assays shown in Figure 1.

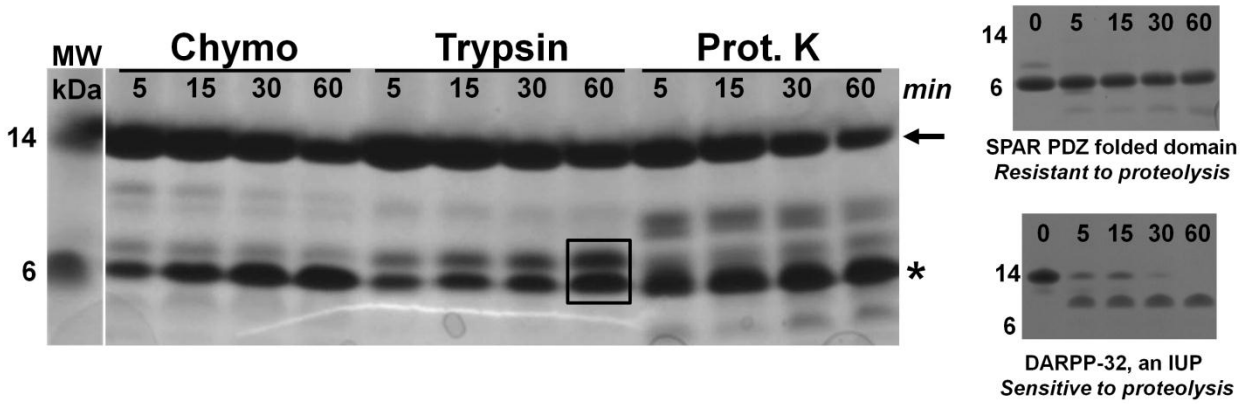


Figure 82: Figure S2, Brown and Grigoriu, PLoS Pathog. MqsA is Susceptible to Proteolytic Degradation.

SDS-PAGE analysis of MqsA-F incubated with chymotrypsin, trypsin or proteinase K for 5, 15, 30 and 60 minutes at 30 °C. The upper band indicated with an arrow corresponds to undigested $MqsA_{1-131}$ (14.9 kDa), while the proteolytic fragments are denoted with *. The trypsin digested samples analyzed by MALDI-TOF MS and LC-MS/MS are boxed. Proteolytic digestion of a stable, folded domain (SPAR PDZ domain) and an intrinsically unstructured protein (DARPP-32) are shown on the right for comparison.

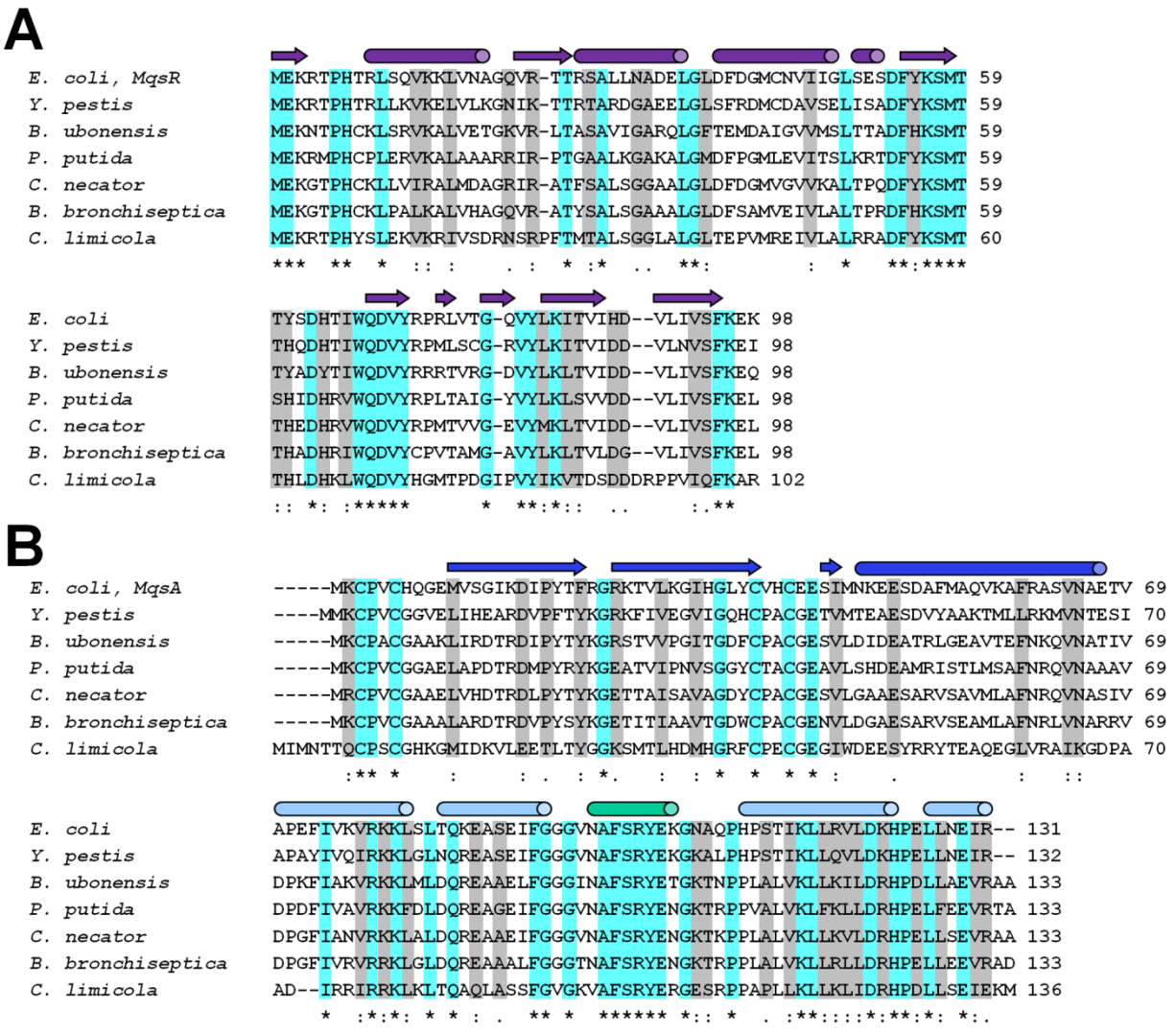


Figure 83: Figure S3, Brown and Grigoriu, PLoS Pathog. MqsR and MqsA are conserved among multiple bacterial species.

Sequence alignment of MqsR (A) and MqsA (B). Secondary structural elements for MqsR and MqsA are represented as cylinders (α-helices) or arrows (β-strands) and shaded according to the protein/domain to which they belong (MqsR, magenta; MqsA-N, MqsA N-terminal zinc binding domain, dark blue; MqsA-C, MqsA C-terminal HTH-XRE domain, light blue, with the exception of the HTH-XRE conserved helix, which is green). Identical residues are highlighted in cyan whereas similar residues are shaded in gray; symbols below the sequence: '*' identical residues; ':' highly similar residues; '.' similar residues.

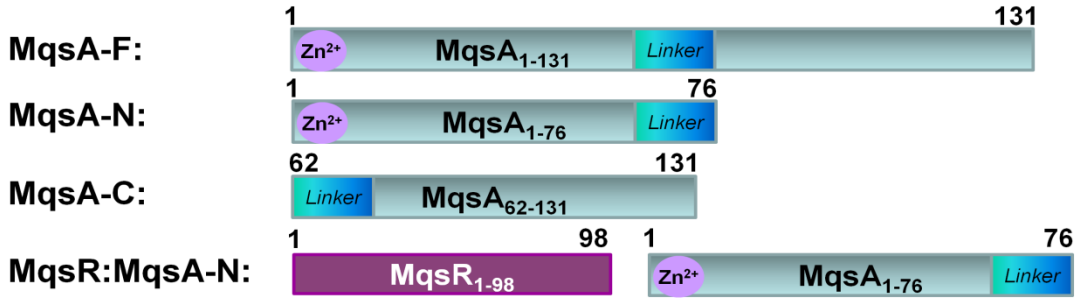


Figure 84: Figure S4, Brown and Grigoriu, PLoS Pathog. Constructs used in this study. MqsA residues 62-76 represent the 'linker' between the two domains of MqsA. Subsequent structure determination demonstrates that the physical linker between the MqsA N- and C-terminal domains is short, centered on residue T68.

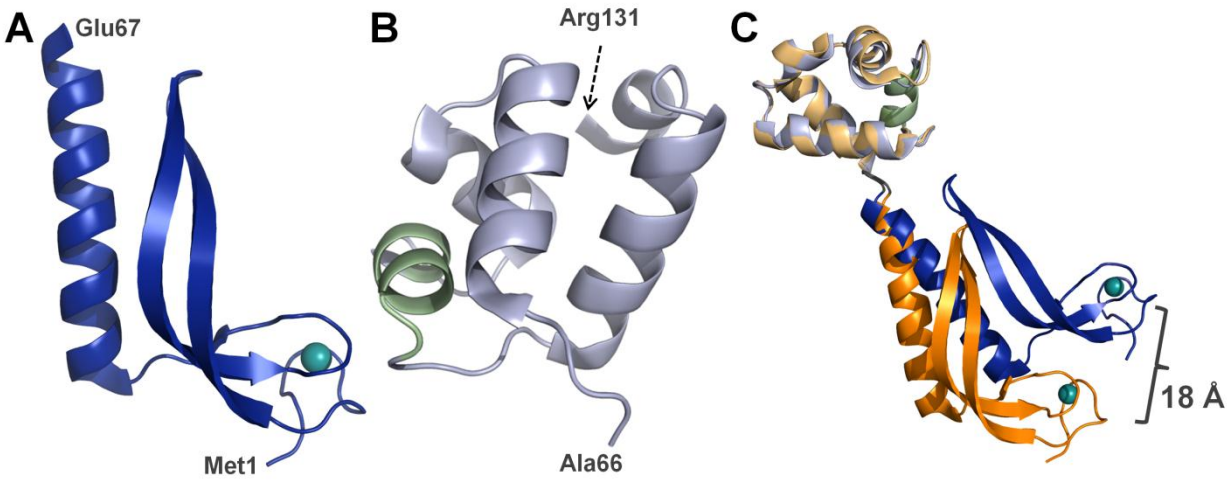


Figure 85: Figure S5, Brown and Grigoriu, PLoS Pathog. Crystal Structures of the Individual MqsA Domains and their Independent Rotation in the MqsA dimer.
(A) Ribbon model of MqsA-N. Bound zinc is illustrated as a teal sphere. (B) Ribbon model of MqsA-C. α–helix 4, which is predicted to mediate DNA binding, is shaded in green. (C) The N- and C-terminal domains of MqsA rotate independently via the short flexible linker. Both monomers of MqsA were superimposed on the MqsA-C domains (light orange, light blue) to illustrate the relative rotation of the MqsA-N domains (dark orange, dark blue), resulting in a shift of the bound zinc ions by 18 Å.

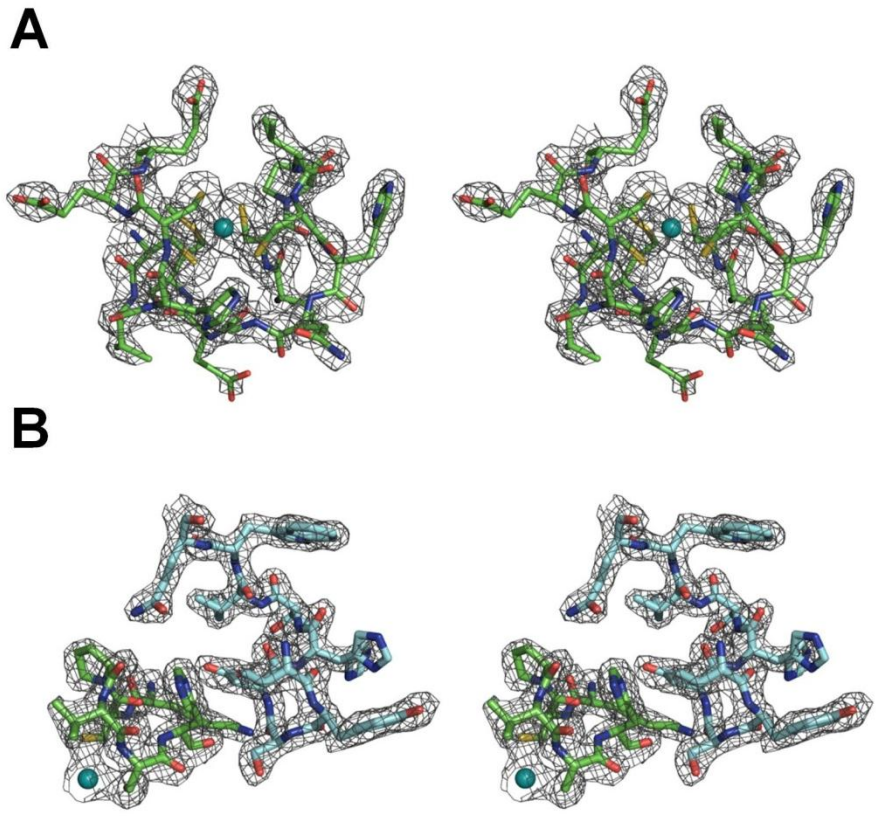


Figure 86: Figure S6, Brown and Grigoriu, PLoS Pathog. Stereoimages of MqsA and MqsR.
(A) Stick representation of the MqsA zinc binding pocket (residues K2-M11 and C37-E42) with σA-weighted 2mF_o-dF_c map at 1.5 σ. MqsA in green, zinc ion in teal. (B) Stick representation of a portion of the MqsR:MqsA secondary interface (MqsR residues T59-Q68 in blue and MqsA residues K2-H7 in green) with σA-weighted 2mF_o-dF_c map at 1.5 σ.

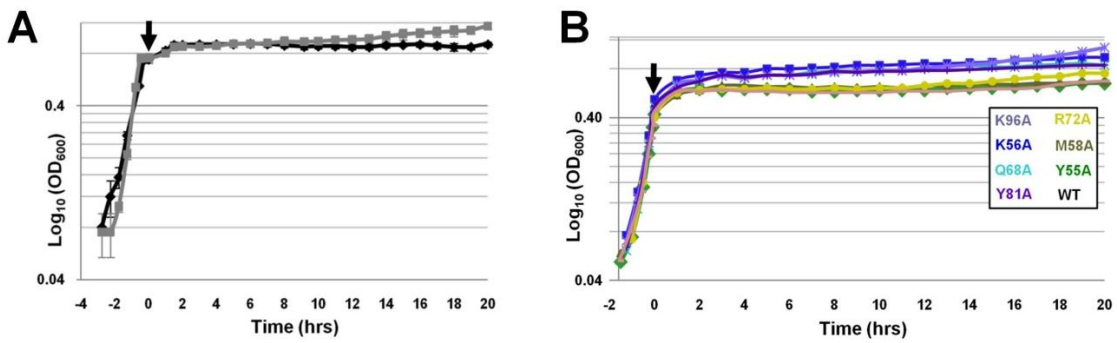


Figure 87: Figure S7, Brown and Grigoriu, PLoS Pathog. Semi-log plots of MqsR toxicity assays illustrated in Figures 4A and 6A.
(A) Semi-log plot of the growth curves of the co-expression of MqsR with the N-terminal domain of MqsA (MqsR:MqsA-N, ■) and the C-terminal domain of MqsA (MqsR:MqsA-C, ♦) in BL21 (DE3) cells. Induction of protein expression using 0.5 mM IPTG corresponds to t=0 (arrow; expression carried out at 18°C, following the same protocol used to produce the proteins for structural studies). Co-expression of MqsR with MqsA-C leads to growth arrest while co-expression with MqsA-N results in robust growth. (B) Semi-log plot of the growth curves of cells over-expressing WT MqsR and seven MqsR mutants. MqsR mutants K56A, Q68A, Y81A and K96A show decreased toxicity compared to WT MqsR, as evidenced by their ability to grow following induction with 0.5 mM IPTG at t=0 (arrow; expression carried out at 18°C, following the same protocol used to produce the proteins for structural studies). In contrast, induction of MqsR mutants Y55A, M58A and R72A, like WT, lead to growth arrest.

C.9. Supplementary Tables

Table 6: Table S1, Brown and Grigoriu, PLoS Pathog. Bacterial strains and plasmids used in this study

Strains and plasmids	Genotype/relevant characteristics	Source
***E. coli* K-12 strains**		
MG1655	F^- λ^- *ilvG rfb*-50 *rph*-1	[288]
BW25113	*lacI*^{q} *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR*514 Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$	[257]
BW25113 *mqsR*	K-12 BW25113 Δ*mqsR* Ω Km^R	[257]
Plasmids		
pBS(Kan)	Km^R; cloning vector	[259]
pBS(Kan)-*mqsR*	Km^R; pBS(Kan) P_{lac}::*mqsR*$^+$	this study
pBS(Kan)-*mqsA-F*	Km^R; pBS(Kan) P_{lac}::*mqsA*$^+$	this study
pBS(Kan)-*mqsR-mqsA-F*	Km^R; pBS(Kan) P_{lac}::*mqsR-mqsA*$^+$	this study
pET28a(+)-*mqsR*	Km^R; pET28a(+) P_{T7}::*mqsR*$^+$	this study
pET28a(+)-*mqsA-F*	Km^R; pET28a(+) P_{T7}::*mqsA*$^+$	this study
pET28a(+)-*mqsA*$_{1-76}$	Km^R; pET28a(+) P_{T7}::*mqsA*$_{1-76}^+$	this study
pET28a(+)-*mqsA*$_{62-131}$	Km^R; pET28a(+) P_{T7}::*mqsA*$_{62-131}^+$	this study
pET28a(+)-*mqsA*$_{62-131}$*L81M*	Km^R; pET28a(+) P_{T7}::*mqsA*$_{62-131}$*L81M*$^+$	this study
pCA21a-*mqsA-F*	Cm^R; pCA21a P_{T7}::*mqsA*$^+$	this study
pCA21a-*mqsA-N*	Cm^R; pCA21a P_{T7}::*mqsA-N*$^+$	this study
pCA21a-*mqsA-C*	Cm^R; pCA21a P_{T7}::*mqsA-C*$^+$	this study

Km^R and Cm^R are kanamycin and chloramphenicol resistance, respectively.

Table 7: Table S2, Brown and Grigoriu, PLoS Pathog. Oligonucleotides used for this study. f indicates forward primer and r indicates reverse primer.

Name	Sequence
Construction of plasmids	
pBS(Kan)-*mqsR*	f: 5'-GCGCGCGATCGGATCCACTAAAGTAACAGGGAGGCGGGGGTTATG-3' r: 5'-GCGCGCCTAGTCTAGATGGCAAACCGGACATTTCATATTACTTC-3'
pBS(Kan)-*mqsA-F*	f: 5'-GCGCGCGATCGGATCCAGGCCAGGTTTATCTTAA AATTACGGTA-3' r: 5'-GCGCGCCTAGTCTAGACCCGCTTTTCCATTAATTAACGGATTTC-3'
pBS(Kan)-*mqsR-mqsA-F*	f: 5'-GCGCGCGATCGGATCCACTAAAGTAACAGGGAGGCGGGGGTTATG-3' r: 5'-GCGCGCCTAGTCTAGACCCGCTTTTCCATTAATTAACGGATTTC-3'
EMSA	
P*mqsR*	f: 5'-GTGATGCCTGACTCCAGCTT-3' r: 5'-CGTGTATGTGGTGTGCGTTT-3'
P*mcbR*	f: 5'-GCAAAGTGGTGATCCGCG-3' r: 5'-CCTGCAGAGTCAAACTGA-3'
P*spy*	f: 5'-CAGTCATCCGGTATAGTT-3' r: 5'-GGCAACAAACAGTGCAGT-3'
P*tomB*	f: 5'-CGGGTTAGTGCTAGTATGAAAAAGT-3' r: 5'-ACTTAAGCTGTGCGATATCATGTCT-3'

Table 8: Table S3, Brown and Grigoriu, PLoS Pathog. Data Collection and Refinement Statistics for MqsA-N

	MqsA-N (peak)	MqsA-N (remote)
Data Collection[1]		
Space Group	P 2_1 2_1 2_1	
Unit cell (Å)	30.8, 52.1, 53.8	
Wavelength (Å)	0.9787	0.9321
Resolution (Å)	50.0-1.70 (1.73-1.70)	50.0-1.70 (1.73-1.70)
R_{sym} (%)	6.4 (31.4)	5.7 (26.2)
<I/σI>	27.25 (7.98)	21.10 (6.62)
Completeness (%)	97.5 (97.0)	97.4 (96.7)
Redundancy	8.8 (9.1)	5.0 (5.2)
Refinement Statistics		
Resolution (Å)		37.42-1.70
R_{cryst}		16.5
R_{free}[c]		18.6
Protein atoms		534
Waters		97
Ligand/Ion atoms		11
r.m.s.d bond length (Å)		0.012
r.m.s.d bond angle (°)		1.283
Average B factors (Å^2)		
Protein		15.80
Water		34.73
Ligand/ions		32.44
Ramachandran Plot		
Favored (%)		98.6
Allowed (%)		1.4
Disallowed (%)		0.0
PDB Code		3GA8

[1] Highest-resolution shell data are shown in parentheses

Table 9: Table S4, Brown and Grigoriu, PLoS Pathog. Data Collection and Refinement Statistics for MqsA-C

	MqsA-C High Resolution	**MqsA-C (peak)**	**MqsA-C (inflection)**	**MqsA-C (remote)**
Data Collection[1]				
Space Group	P 3_1 2 1		P 3_1 2 1	
Unit Cell (Å)	39.6, 39.6, 78.3		39.35 39.35 78.3	
Wavelength (Å)	1.0	0.9788	0.9794	0.9322
Resolution (Å)[a]	50.0-1.40 (1.42-1.40)	50.0-2.0 (2.07-2.0)	50.0-2.0 (2.07-2.0)	50.0-2.0 (2.07-2.0)
R_{sym} (%)	3.2 (23.2)	4.3 (6.8)	4.0 (6.5)	4.1 (7.5)
<I/σI>	31.5 (6.19)	77.1 (50.6)	53.8 (34.7)	28.8 (18.5)
Completeness (%)	99.6 (99.2)	99.7 (97.0)	99.6 (96.4)	99.9 (99.6)
Redundancy	5.7 (4.5)	8.5 (7.3)	4.2 (3.7)	4.3 (4.3)
Refinement Statistics				
Resolution (Å)	20.77-1.40			
R_{cryst} (%)	15.5			
R_{free} (%)	18.2			
Number of atoms				
Protein atoms	571			
Waters	80			
r.m.s.d bond length (Å)	0.011			
r.m.s.d bond angle (°)	1.413			
Average B factors (Å^2)				
Protein	15.01			
Water	29.33			
Ligand/ion	N/A			
Ramachandran Plot				
Favored (%)	100.0			
PDB Code	3FMY			

[1] Highest-resolution shell data are shown in parentheses

Appendix D: "The E. coli toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter"

Reproduced from the Journal of Biological Chemistry, Volume 288, Number 2, p. 1286-94

My contribution to this manuscript was initial work with MqsR, and the initial development of a denaturing and refolding protocol for obtaining free MqsR.

Supplementary, figures and tables for this manuscript are located in **Sections D.6 and D.7 of this appendix.**

Breann L. Brown[1], Dana M. Lord[1], Simina Grigoriu[2], Wolfgang Peti[1,3], Rebecca Page[2]

From the [1]Department of Molecular Pharmacology, Physiology and Biotechnology, [2]Department of Molecular Biology, Cell Biology, and Biochemistry, [3]Department of Chemistry, Brown University, Providence, RI 02912, USA

Acknowledgments

Crystallographic data was collected at the Brookhaven National Laboratory (BNL) National Synchrotron Light Source (NSLS) X6A beamline. We thank the beamline staff for help with X-ray fluorescence scans. WP is the Manning Assistant Professor for Medical Science at Brown University. TKW is the T. Michael O'Connor II Chair and Professor at Texas A & M University.

Keywords

Toxin/antitoxin, persister cells, endoribonuclease

Background

MqsR, an endoribonuclease, and MqsA, a transcriptional regulator, form a unique toxin-antitoxin (TA) pair.

Results

The high-affinity, stable MqsR/MqsA complex is unable to bind DNA.

Conclusion

MqsR is the only toxin shown to disrupt the antitoxin/DNA complex, which would promote transcription.

Significance

The MqsR toxin may promote multidrug tolerance in *E. coli* by disrupting MqsA-mediated transcriptional repression of several genes related to persistence.

Running Title: *MqsR destabilizes the MqsA/DNA complex*

D.1. Abstract

Bacterial biofilms are complex communities of cells containing an increased prevalence of dormant cells known as persisters, which are characterized by an upregulation of genes known as toxin/antitoxin (TA) modules. The association of toxins with their cognate antitoxins neutralizes toxin activity allowing for normal cell growth. Additionally, protein antitoxins bind their own promoters and repress transcription, while the toxins serve as co-repressors. Recently, TA pairs have been shown to regulate their own transcription through a phenomenon known as conditional cooperativity, where the TA complexes bind operator DNA and repress transcription only when present in the proper stoichiometric amounts. The most differentially upregulated gene in persister cells is *mqsR*, a gene that, with the antitoxin *mqsA*, constitutes a TA module. Here, we reveal that, unlike other TA systems, MqsR is not a transcription co-repressor but instead functions to destabilize the MqsA/DNA complex. We further show that DNA binding is not regulated by conditional cooperativity. Finally, using biophysical studies, we show that complex formation between MqsR and MqsA results in an exceptionally stable interaction, resulting in a sub-nanomolar dissociation constant that is similar to that observed between MqsA and DNA. In combination with crystallographic studies, this work reveals that MqsA binding to DNA and MqsR is mutually exclusive. To our knowledge, this is the first TA system in which the toxin does not function as a transcriptional co-repressor, but instead functions to destabilize the antitoxin/operator complex under all conditions, and thus defines another unique feature of the *mqsRA* TA module.

D.2. Introduction

Bacterial biofilms are communities of bacterial cells that adhere to an inert surface and are enclosed in a self-produced extracellular polymeric matrix [289]. Notably, biofilms are present in at least 65% of all bacterial infections [290]. Their presence is particularly challenging because biofilms are extremely recalcitrant to antibiotic treatment [291]. The cellular population of biofilms differs from that of normal planktonic cultures in that they display an increased prevalence of a subpopulation of cells known as persister cells [227]. First identified in 1944 [223], persister cells are phenotypic variants of wild type cells that exhibit multidrug tolerance [224, 225, 292]. While much remains unknown about the persister state and how it is regulated *in vivo*, it is becoming widely evident that a class of genes known as toxin/antitoxin (TA) modules play a significant role in the persister phenotype [224, 228]. For instance, deletion of certain TA operons has been shown to decrease the levels of persistence [293, 294]. More importantly, it has also been shown that many TA genes, especially toxins, are preferentially upregulated in persister versus non-persister cells [228] and it is the overexpression of toxins that leads to increased antibiotic tolerance [295, 296].

There are currently five known types of TA systems [297-299]. While the toxin genes always code for proteins, the antitoxins are either protein (types II, IV, V) [229, 298, 299], or RNA (types I and III) [300, 301]. Type II TA systems are the most well-characterized and are comprised of a two-gene operon that encodes a labile protein antitoxin and a stable toxin [229]. Under normal conditions, the protein products associate to form a non-toxic complex. However,

under conditions of environmental stress, such as nutrient starvation, oxidative stress, or antibiotic challenge, the antitoxins are rapidly degraded by cellular proteases, mainly Lon or ClpXP [233, 238, 239, 302], and the toxins are free to exert their cellular effects.

Previously, gene expression profiling revealed that the gene most highly upregulated in *Escherichia coli* persister cells is the toxin *mqsR* [228]. The *mqsR* gene encodes a small 98 amino acid protein which functions as a sequence-specific mRNA endoribonuclease belonging to the RelE family of bacterial toxins [267, 303, 304]. Its cognate antitoxin *mqsA*, which is located immediately downstream of *mqsR* in the same operon, encodes a 131 amino acid protein that functions to neutralize MqsR toxicity. There is a growing body of evidence that the *mqsRA* TA system plays a significant role in multiple cellular processes in *E. coli*. For instance, MqsR has been shown to facilitate biofilm formation by affecting cellular motility as a result of autoinducer-2 signaling [255]. Also, deletion of *mqsR* and the *mqsRA* operon leads to decreased persister cell formation, the only type II TA system known to do so, while overexpression of MqsR increases persister cell formation [293]. Finally, the antitoxin MqsA has been shown to serve as a transcriptional regulator not only of the *mqsRA* operon, but also of several other *E. coli* genes, including *mcbR*, *cspD*, and the general stress response regulator *rpoS* [302, 303, 305].

We previously showed that the antitoxin MqsA is unique among all antitoxins studied to date. First, MqsA is completely structured throughout its entire sequence both in the free and DNA- or toxin-bound states [303, 306], while

other antitoxins are either partially or completely unstructured in the free state [307-311]. Second, antitoxins typically bind DNA and their cognate toxins via their N- and C-terminal domains, respectively [251, 308, 309, 312]. Conversely, MqsA binds its toxin, MqsR, via its N-terminal domain and DNA via its C-terminal domain. Third, MqsA is the only antitoxin that binds metal, zinc, at its N-terminal toxin-binding domain for structural stability [303].

However, while the unique features of the MqsA antitoxin are well-established, it is only now becoming evident that the MqsR toxin is also atypical when compared to other *E. coli* toxins. Instead of binding MqsA and enhancing transcriptional repression, as has been observed in other TA systems [234, 248, 313-315], here we show that MqsR serves solely to destabilize the complex formed between MqsA and DNA. This occurs because both the MqsA/DNA and MqsR/MqsA complexes have comparably strong affinities and because the binding sites of DNA and MqsR on MqsA partially overlap, making it impossible for MqsA to bind DNA and MqsR simultaneously without conformational changes in either the MqsA antitoxin, the MqsR toxin or both. Thus, these data provide further evidence that both the MqsA antitoxin and MqsR toxin have multiple unique characteristics that are not observed in canonical TA systems, and thus *mqsRA* is the founding member of a unique family of TA pairs.

D.3. Experimental procedures

D.3.a. Protein expression and purification

Free MqsR was obtained by co-expressing pET30a-MqsR, which included an N-terminal hexahistidine (His_6) tag, with untagged pCA21a-MqsA-N I44A (MqsA-N

is the N-terminal domain of MqsA and includes residues 1-76) at 18°C in BL21(DE3) *E. coli* competent cells. After lysis, the His_6-MqsR/MqsA-N complex was bound to nickel metal affinity resin and denatured with 6M guanidine hydrochloride (GuHCl). His_6-MqsR (hereafter referred to as MqsR) was eluted with 500 mM imidazole and refolded by stepwise dialysis in buffers containing decreasing amounts of GuHCl followed by a final preparative gel filtration step (Superdex 75 16/60, GE Healthcare). The untagged MqsR/MqsA co-expressed and co-purified complex was prepared as described previously [303]. The His_6-MqsR/MqsA co-expressed and co-purified complex (CoExp-MqsR/MqsA) was formed by co-expressing pET30a-His_6-MqsR with untagged full-length pCA21a-MqsA (MqsA-F, residues 1-131) at 18°C in BL21(DE3) *E. coli* competent cells. The complex was co-purified using nickel affinity chromatography and eluted with 50 mM Tris pH 8.0, 500 mM NaCl, 500 mM imidazole. The pooled protein was concentrated and purified by a final preparative gel filtration step (Superdex 75 26/60 or 16/60, GE Healthcare). The His_6-MqsR/MqsA reconstituted complex (Rec-MqsR/MqsA) was prepared by incubating equimolar amounts of purified His_6-MqsR and MqsA for 30 min on ice and purified by a final preparative gel filtration step (Superdex 75 16/60, GE Healthcare). Mutagenesis of *mqsA* was carried out using the QuickChange Mutagenesis Kit (Agilent Technologies) using the manufacturer's protocols; all constructs were verified by sequencing. Finally, full-length MqsA constructs (WT and each mutant) were expressed and purified as described previously [303]. All protein samples were stored in 10 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM TCEP.

D.3.b. Circular dichroism

All circular dichroism (CD) measurements were performed with a Jasco J-815 CD spectropolarimeter. Dilute samples in 10 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM TCEP were placed in a 0.2 cm glass cuvette. Wavelength scans were performed in the far-UV region from 260-195 nm at 25°C at a speed of 20 nm/min. Thermal denaturation scans were performed at 208 nm from 25-100°C for MqsA (2.5 μM) and MqsR (5 μM) or 25-110°C for MqsR/MqsA, MqsR/MqsA R61A, and MqsR/MqsA R61D (2.5 μM) at a scan speed of 1°C/min. Raw data were processed using the denatured protein analysis implemented in SepctraAnalysis software and converted to mean residue molar ellipticity using the following equation: $[\theta]=\theta/(10xCxlxn)$, where θ is ellipticity, C is the molar concentration, l is the cell path length in cm and n is the number of residues. Plots are the average of three replicate scans.

D.3.c. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC (GE Healthcare) at 25°C. All proteins were exchanged into the same protein buffer using size exclusion chromatography (Superdex 75 16/60) and then concentrated immediately prior to the ITC experiments. MqsA-F (30 μM; syringe) was titrated into MqsR (2.8 μM; sample cell) with constant stirring. Due to the extremely tight binding between MqsR and MqsA, the injection volume was decreased from 10 to 5 μL in the middle of the titrations (injections #8-27 out of 38 total injections) and then increased back to their initial value. This allowed more data points to be obtained at the isotherm transition which, in turn, allowed

the binding transition to be fit with higher accuracy. Binding isotherms were fit using a one-site binding model and association constant (K_A) values calculated based on the known binding stoichiometry (n) of one [303] using Origin.

D.3.d. mRNA cleavage assay

DNA containing a T7 RNA polymerase promoter sequence was obtained using the primers shown in Supplemental Table S1 and purified using the PureLink PCR purification kit (Invitrogen). The PCR products were used as templates for the *in vitro* RNA synthesis reaction with the AmpliScribe T7-Flash transcription kit (Epicentre). The MqsR cleavage assay contained 2 µg RNA and 15 ng protein. For the RNA only control reaction, protein was replaced by buffer (10 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM TCEP). The reactions were incubated at 37°C for 15 min and quenched by the addition of an equal volume of 2X NOVEX TBE-urea sample loading buffer (Invitrogen). The reaction products were resolved by electrophoresis with RNA denaturing gels (15% polyacrylamide with 7 M urea, Invitrogen) and stained with 0.5 µg/ml ethidium bromide.

D.3.e. Electrophoretic mobility shift assay

The *mqsRA* operon promoter (P*mqsRA*, Table 1) was PCR amplified using primers shown in Supplemental Table S2 from BW25113 WT genomic DNA as previously described [303]. The DNA was then labeled using the Biotin 3' End DNA Labeling Kit (Pierce). Palindromes 1 and 2 (each 26 bp, Table 1) were obtained by annealing. Specifically, following synthesis of the individual oligonucleotides (IDT Technologies, each oligonucleotide included a 3′ biotin label; Supplemental Table S2), the complementary oligonucleotides were

Table 10: TABLE 1, Brown, JBC. DNA constructs used in EMSA experiments

DNA Construct	base pairs	Palindrome 1	Palindrome 2
P*mqsRA*	234	X	X
Pal1	26	X	
Pal2	26		X

combined, heated to 95°C and then cooled by 1°C per min to a final temperature of 25°C. For titration experiments, increasing concentrations (0.2, 1, 2, 5 or 10 pmol) of MqsA-F, MqsA R61A, MqsA R61D, MqsR, or the MqsR/MqsA-F complex were added to a constant amount (100 fmol) of labeled DNA and all reactions were carried out in binding buffer (10 mM Tris pH 7.5, 50 mM KCl, 1 mM DTT) in the presence of poly (dI-dC) DNA probe (50 ng/µL) to prevent non-specific binding. The binding reactions were incubated at room temperature for 20 min. Samples were then loaded onto a 6% DNA retardation gel (Invitrogen) and subjected to electrophoresis at 4°C for either 75 min (Pal1 and Pal2) or 90 min (P*mqsRA*) at 100 V in 0.5X TBE buffer (45 mM Tris pH 8.3, 45 mM Boric acid, 1 mM EDTA). The DNA was transferred to a nylon membrane at 390 mA for 30 min and subsequently UV crosslinked at 302 nm by placing the membrane face-down on a UV illuminator for 15 min. Chemiluminescence was performed with the LightShift Chemiluminescent EMSA Kit (Pierce) and the samples were detected using a CCD imager (Typhoon 9410 Imager).

For experiments altering the relative amount of MqsR, a constant amount of MqsA-F (400 fmol) was incubated with increasing amounts of MqsR (either 200, 300, 400, or 800 fmol for experiment 1 or 50, 100, 150, 200, or 400 fmol for experiment 2). The MqsR/MqsA-F complex was allowed to incubate at room temperature for 10 min. Next, 50 fmol of labeled P*mqsRA* DNA was added to the pre-formed MqsR/MqsA-F complex and the ternary binding reaction allowed to

incubate an additional 20 min at room temperature. For the second experiment, the reverse reaction was also prepared in which a constant amount of MqsA was incubated with a constant amount of labeled P*mqsRA* DNA followed by addition of increasing amounts of MqsR. The electrophoresis, transfer and chemiluminescent detection were carried out as described above.

D.4. Results

D.4.a. Free MqsR readily degrades its own mRNA

Free MqsR, an endoribonuclease toxin, was isolated by purifying and denaturing the MqsR/MqsA-N complex (MqsA-N is the N-terminal domain of MqsA, residues 1-76) and then isolating and refolding MqsR by step-wise dialysis followed by size exclusion chromatography (Fig. 1A). Circular dichroism (CD) polarimetry (Fig. 1B) was used to verify MqsR was folded. Refolded MqsR also formed a complex with full-length MqsA (reconstituted MqsR/MqsA, or Rec-MqsR/MqsA) that was indistinguishable from co-expressed and co-purified MqsR/MqsA (CoExp-MqsR/MqsA), as determined by size exclusion chromatography experiments (Fig. 1A). We used electrophoretic mobility shift assays (EMSA) to demonstrate that, in contrast to MqsA, MqsR does not bind the full *mqsRA* promoter (Fig. 1C; the specificity of MqsA for P*mqsRA* is shown in Supplementary Fig. S1 and in [306] for an individual *mqsRA* palindrome). To verify that the refolded MqsR was active, we performed mRNA cleavage assays with conditions similar to those previously reported [267] and using mRNA transcripts of *mqsR* and *mqsA* as substrates. MqsR was previously shown to specifically cleave mRNA at GCU and, to a lesser extent, GCA sequences [267,

303, 304]; *mqsR* and *mqsA* transcripts contain 1 and 3 GCU motifs, respectively. As can be seen in Figure 1D, incubation of refolded MqsR with *in vitro* transcribed *mqsR* and *mqsA* mRNA (amplified using primers in Supplemental Table S1) resulted in the cleavage of both mRNAs. Significantly more cleavage products were observed for *mqsA* than *mqsR* mRNA, consistent with the presence of 3 versus 1 GCU cleavage sites in the *mqsA* versus *mqsR* transcript, respectively. These data also showed that MqsR activity is potently inhibited when bound to MqsA, independent of whether the MqsR/MqsA complex was co-expressed and co-purified or the MqsA and MqsR proteins were purified individually and the complex reconstituted *in vitro*. Collectively, these data show that refolded MqsR behaves identically to WT MqsR and that it is one of only a few examples in which a bacterial toxin has been shown to cleave its own mRNA transcript, an activity that may serve as a mechanism of auto-regulation to alleviate the persister state [316].

D.4.b. Unlike most antitoxins, MqsA binds MqsR and its operator DNA with similar sub-nanomolar affinities

Several toxins associate with their cognate antitoxins with extremely high affinities, likely to allow for proper cell growth under non-stressful environmental conditions [317-319]. To determine the affinity of the MqsR/MqsA complex, we used isothermal titration calorimetry (ITC). ITC measurements using MqsA and MqsR reported a dissociation constant (K_D) of 0.08 ± 0.07 nM (a representative of three replicates is shown in Fig. 2A). This K_D is similar to those reported for

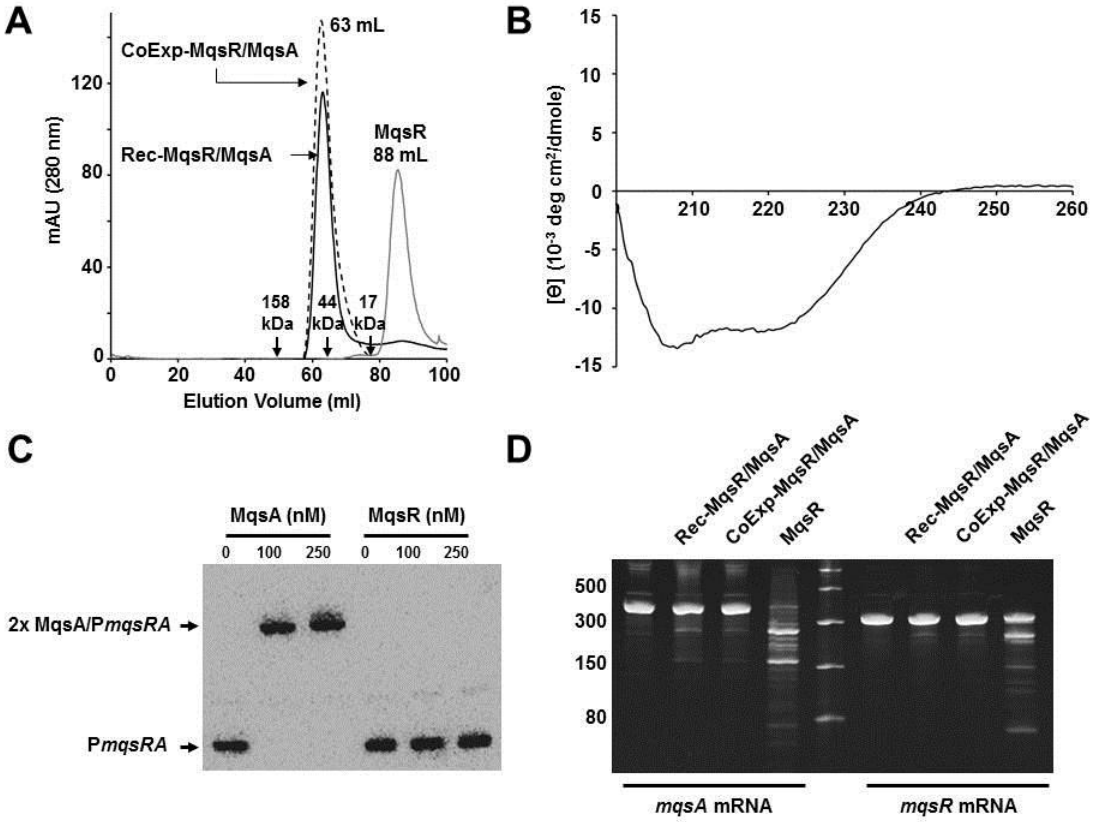


Figure 88: FIGURE 1, Brown, JBC. MqsR readily cleaves *mqsRA* mRNA.
A, Size exclusion chromatogram of free MqsR, the MqsR/MqsA complex which was co-expressed and co-purified (CoExp-MqsR/MqsA) and the MqsR/MqsA complex which was produced by incubating free MqsA with refolded MqsR prior to SEC (Rec-MqsR/MqsA). *B*, CD wavelength scan of refolded MqsR. *C*, EMSA with increasing amounts of either MqsA alone (lanes 2,3) or MqsR alone (lanes 5,6) with biotin-labeled P*mqsRA*. *D*, The MqsR toxin readily degraded both *mqsA* and *mqsR* transcripts (lanes 4 and 9). MqsR activity was inhibited when bound to MqsA, independent of whether the MqsR/MqsA complex was formed *in vitro* (lanes 2 and 7) or co-expressed and co-purified (lanes 3 and 8). Full-length *mqsA* mRNA (415 nucleotides, lane 1) and full-length *mqsR* mRNA (316 nucleotides, lane 6) were incubated with buffer under the same reaction conditions as a negative control. Lane 5 contains a single-stranded RNA ladder. All reactions contained 2 µg RNA, 15 ng of protein (or equal volume of buffer for the negative control) and brought up to volume (10 µl) in RNase-free water.

other TA systems, including CcdA/CcdB (0.03 nM) [317], RelB/RelE (0.33 nM) [318] and PezA/PezT (< 1 nM) [319].

However, the K_D of MqsA for MqsR (0.8 nM) is nearly identical to that between MqsA and its DNA operator sequence, which we measured previously using quantitative EMSA [306]. This similarity in affinities of an antitoxin for DNA and for its cognate toxin is not typical of most TA systems. Instead, in other cases, the binding affinity between the antitoxin and DNA is much lower than that between the antitoxin and toxin. For example, RelB binds its DNA operator with a K_D of 200-300 μM, which is nearly 1000-fold higher (weaker affinity) than that between RelB and RelE [311]. Similarly, the K_D of CcdA and its DNA operator is 2-90 μM, or 100-1000 fold higher (weaker affinity) than that between CcdA and CcdB [308].

D.4.c. The MqsR/MqsA complex is exceptionally stable

To compare the stabilities of free MqsR, free MqsA and the MqsR/MqsA complex, we monitored protein unfolding with increasing temperature using CD polarimetry. Both MqsA and MqsR exhibited a two-state melting curve, with melting temperatures (T_m) of 61.1 ± 3.3°C and 48.1 ± 0.3°C respectively (Fig. 2B, Table 2). The MqsR/MqsA complex also exhibited a two-state melting curve, demonstrating that the complex unfolds cooperatively and behaves like a single folded protein. However, the T_m of the MqsR/MqsA complex was 83.4 ± 0.3°C (Fig. 2B), more than 21°C and 35°C higher than for free MqsA and free MqsR, respectively. This shows that complex formation significantly enhances the stability of both proteins, resulting in a complex that is exceptionally stable.

Table 11: TABLE 2, Brown, JBC. Thermal melting transition temperatures determined by CD (n≥2)

Protein	T_m (°C)
MqsR	48.1 ± 0.3
MqsA-F	61.1 ± 3.3
MqsR/MqsA-F	83.4 ± 0.3
MqsR/MqsA-F *R61A*	75.2 ± 0.4
MqsR/MqsA-F *R61D*	70.5 ± 1.0

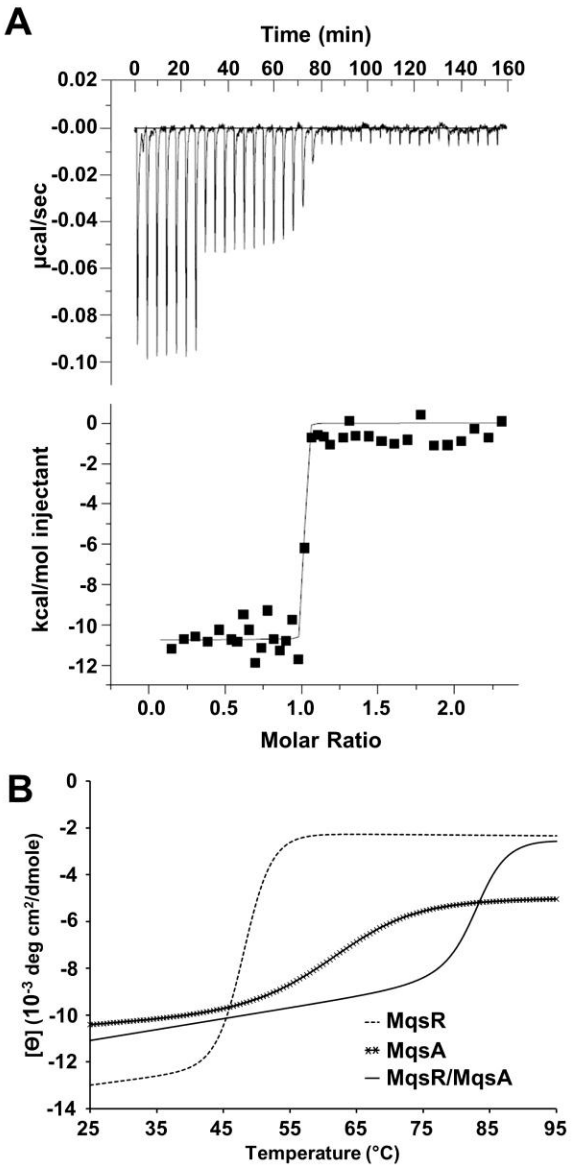


Figure 89: FIGURE 2, Brown, JBC. MqsR forms an extremely stable complex with MqsA.
A, Raw isothermal titration calorimetry data (upper panel) and derived binding isotherm plotted versus the molar ratio of titrant which was fit using a one-site model (lower panel) for MqsA (titrant) into MqsR (sample). Due to the high affinity, the volume of injections 8-27 (of 38 total) was decreased from 10 to 5 µL in order to better monitor the binding transition. The solid line in the lower panel is the best fit to the data using the nonlinear least-squares regression algorithm (ORIGIN) and the calculated K_D is 0.08 ± 0.07 nM. *B*, Thermal denaturation at 208 nm of MqsR/MqsA (2.5 µM, solid line). The complex unfolds with a single-state transition and a T_m of 83.4 ± 0.3°C. The T_m for MqsR (5 µM, dashed line) and MqsA (2.5 µM, crossed line) are 48.1 ± 0.3°C and 61.1 ± 3.3°, respectively.

D.4.d. MqsR destabilizes the MqsA/DNA transcriptional repressor complex

In most type II TA systems, both the antitoxin and the TA complex are able to bind DNA and regulate the transcription of the TA operon. Moreover, in many cases, the purified TA complex binds DNA with higher affinity. The *mqsRA* promoter contains two palindromic MqsA binding sites (Fig. 3A, Table 1) [267, 306]. Using EMSAs and X-ray crystallography, we have previously shown that MqsA binds robustly and with high affinity to both palindromic binding sites [306, 320]. However, detailed analysis of the structures of the MqsA/DNA [306] and MqsR/MqsA-N complexes [303] also revealed that, if the MqsR/MqsA complex was able to bind DNA, either the MqsA/DNA or the MqsA/MqsR interface would have to change in order for both DNA and MqsR to bind MqsA simultaneously. This suggested that the MqsR/MqsA complex might interact with DNA in a completely different manner than MqsA alone.

To test this, we examined the ability of the MqsR/MqsA complex to bind either the first (Fig. 3B) or second palindrome (Fig. 3C) in P*mqsRA* using EMSAs; free MqsA was tested in parallel for comparison. As expected, free MqsA robustly binds to both DNA palindromes at all concentrations tested (10-500 nM, Fig. 3B-C lanes 2-6). Unexpectedly however, at the same concentrations, the MqsR/MqsA complex shows practically no DNA binding to either palindrome, even up to 500 nM (Fig. 3B-C lanes 8-12). Notably, these protein concentrations represent a 2, 10, 20, 50, and 100-fold excess of protein versus DNA. The inability of the MqsR/MqsA complex to bind DNA is contrary to

other TA systems, in which the toxin serves as a transcriptional co-repressor and enhances the effective DNA binding affinity of their cognate antitoxins. This result is also in contrast to previous reports, which concluded that the MqsR/MqsA complex does bind the *mqsRA* promoter [267, 303]. To further investigate this difference, we combined increasing amounts of both the MqsA antitoxin and the MqsR/MqsA complex with the full P*mqsRA* promoter DNA (this DNA contains both palindromes) and monitored binding. Critically, in these EMSA experiments, both free MqsA and the MqsR/MqsA complex reactions were run on the same gel to allow, for the first time, a side-by-side comparison of the observed shift with MqsA and the MqsR/MqsA complex. As can be seen in Figure 3D, our data show that the shift in DNA migration that occurs in the presence of the MqsR/MqsA complex (54.2 kDa) is identical to that which occurs in the presence of MqsA alone (29.8 kDa). Thus, the slower migration (higher position on the gel) expected with the MqsR/MqsA complex due to its nearly two-fold higher molecular weight is not observed. This result shows that at the high molar excess of protein to DNA tested here, the MqsR/MqsA and MqsA/DNA interactions, which have similar K_D values, result in a fraction of MqsA dissociating from MqsR and binding the DNA. Thus, because previous EMSA experiments with MqsA and MqsR/MqsA were never performed on the same gel, the observed shift for the MqsR/MqsA complex was attributed to the complex. However, these results unequivocally show that the shift is not due to the MqsR/MqsA complex binding DNA, but instead due to a small fraction of free MqsA that binds DNA. Critically, this binding is observed only when the concentration of MqsR/MqsA significantly

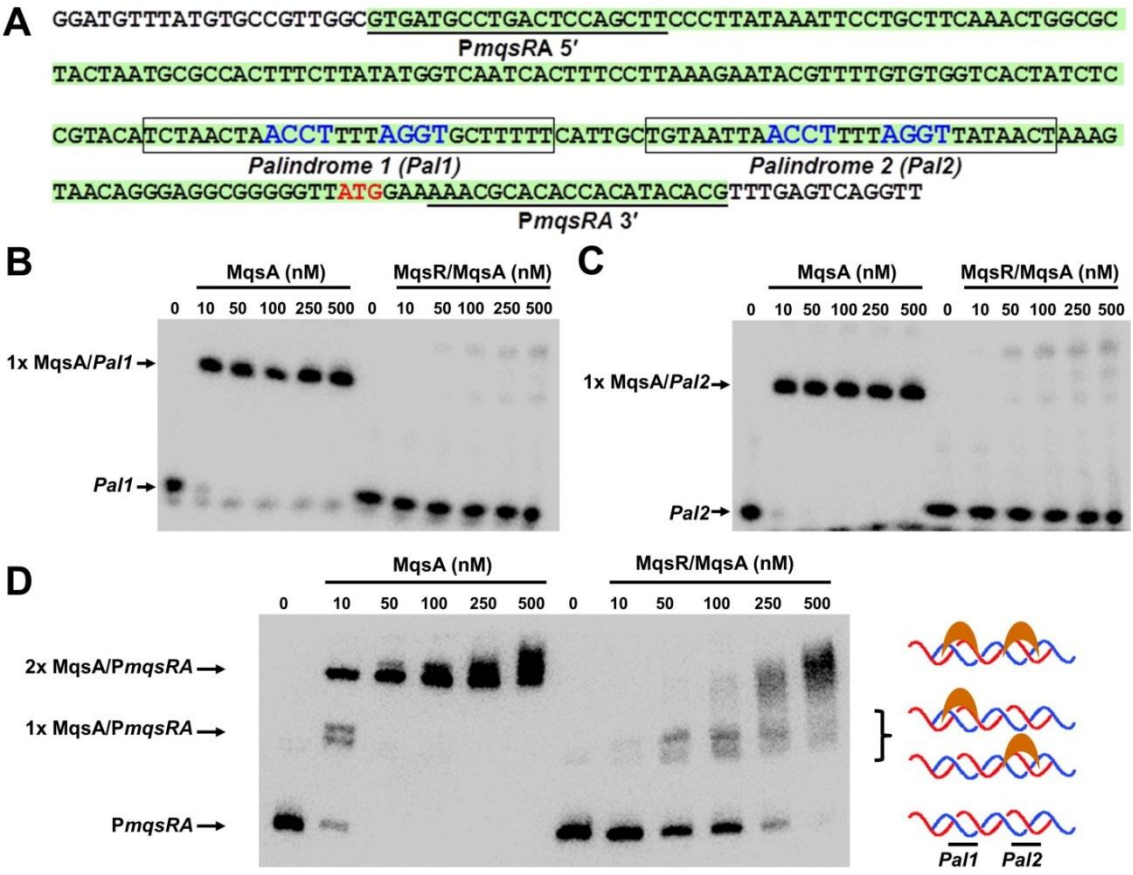


Figure 90: FIGURE 3, Brown, JBC. The MqsR/MqsA complex does not bind to the *mqsRA* promoter *in vitro*.

A, Sequence of the *mqsRA* promoter (P*mqsRA*). DNA constructs used in this study are illustrated: Palindromes 1 and 2 are boxed (nucleotides that interact specifically with MqsA, as determined from the MqsA/DNA crystal structure, are in blue); the P*mqsRA* DNA is highlighted in green and the primers used to amplify it are underlined. *B,* EMSA with increasing amounts of either MqsA alone (lanes 2-6) or the untagged MqsR/MqsA complex (lanes 8-12) incubated with biotin-labeled Palindrome 1 (Pal1) of P*mqsRA*. *C*, Same as (*B*), except proteins were incubated with biotin-labeled Palindrome 2 (Pal2). *D*, Same as (*B*), except proteins were incubated with biotin-labeled P*mqsRA* promoter DNA. The two shifted bands represent protein bound to either one (middle arrow; 1x MqsA/P*mqsRA*) or both palindromes (top arrow; 2x MqsA/P*mqsRA*). DNA binding in the presence of the MqsR/MqsA complex is due to trace amounts of free MqsA, as the observed migration positions are identical to that seen with MqsA alone. For all gels, the negative control (100 fmol labeled DNA lacking protein) is shown in lanes indicated as '0'.

exceeds that of the DNA. Collectively, these data show for the first time that, unlike most other TA systems, the MqsR/MqsA complex does not bind DNA *in vitro*.

D.4.e. mqsRA is also unique in that it does not exhibit conditional cooperativity

An emerging feature of type II TA systems is their ability to auto-regulate transcription by a mechanism known as conditional cooperativity [321-324]. In this mechanism, the relative ratio of toxin and antitoxin present in the cell dictates the repression state of the operon. For canonical TA systems, when the amount of antitoxin is in excess or equal to that of the toxin, transcription is repressed and the toxin serves as a co-repressor by increasing DNA binding affinity. However, when this ratio is reversed and the amount of toxin exceeds that of the antitoxin, the antitoxin dissociates from the DNA, allowing transcription to proceed.

We have already shown, using the co-purified complex, that when MqsR and MqsA are present in equimolar ratios, the complex does not bind DNA (Fig. 3). In order to determine if the *mqsRA* TA system exhibits conditional cooperativity, we tested the ability of MqsA to bind DNA in the presence of sub-equimolar ratios of MqsR (Fig. 4A). In the absence of MqsR, 20 nM MqsA is able to simultaneously bind both palindromic sites and completely shift the P*mqsRA* DNA (Fig. 4A, lane 2). However, upon the addition of MqsR toxin, DNA binding is immediately destabilized, as can be seen by a decrease in intensity of the band corresponding to MqsA bound to both palindromes and a gradual increase

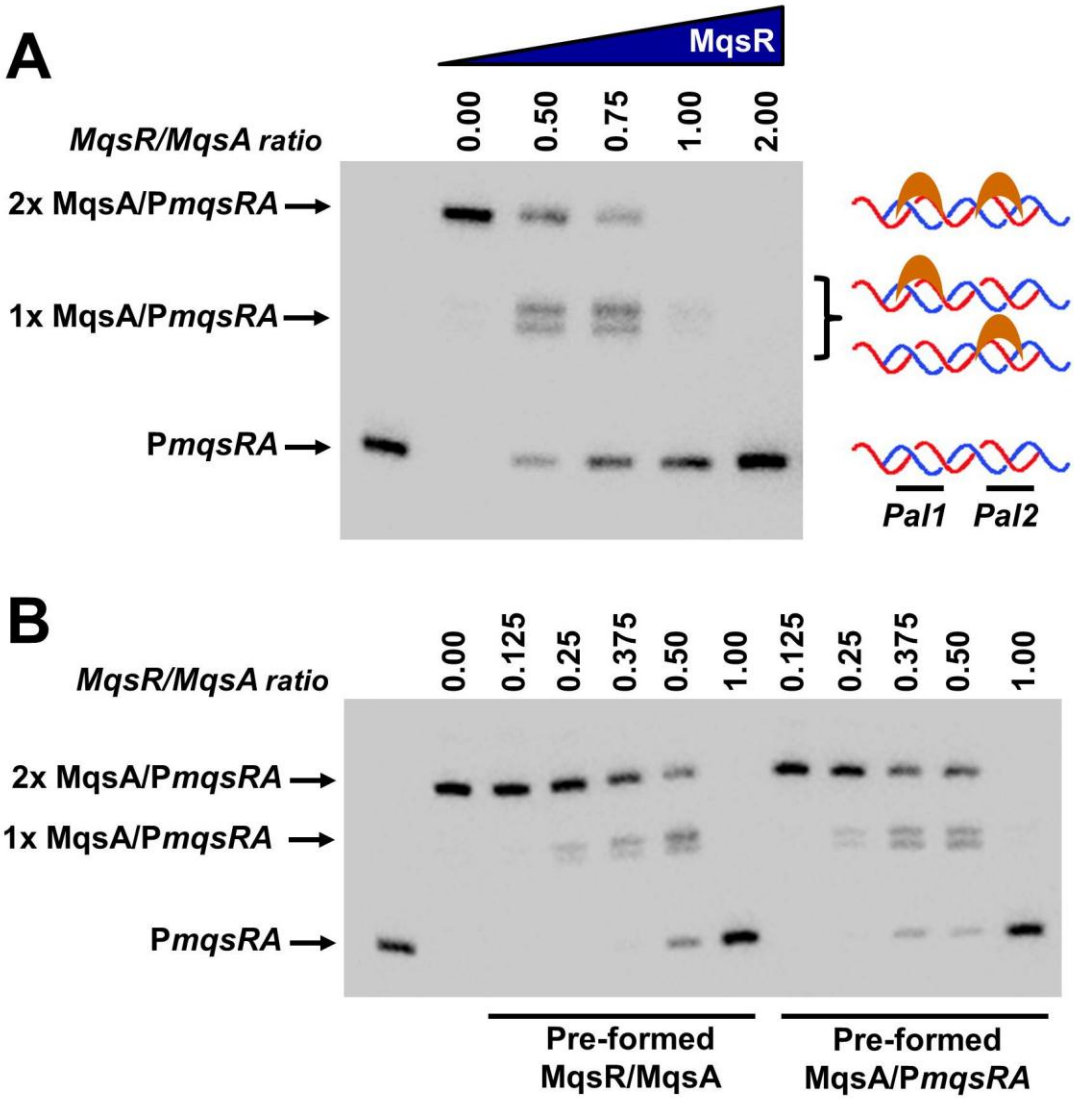


Figure 91: FIGURE 4, Brown, JBC. MqsR destabilizes the interaction of MqsA with DNA. *A*, An EMSA of MqsA binding to P*mqsRA* DNA in the presence of increasing amounts of MqsR. The two shifted bands represent protein bound to either one palindrome (middle arrow; 1x MqsA/P*mqsRA*) or both palindromes (top arrow; 2x MqsA/P*mqsRA*) simultaneously. Lanes 2-6 contain a constant amount of MqsA (20 nM) while lanes 3-6 contain either 10, 15, 20, or 40 nM MqsR. Biotin-labeled P*mqsRA* in the absence of protein (50 fmol, lane 1) was used as a negative control. *B*, *Pre-formed MqsR/MqsA samples*: MqsA was incubated with MqsR at the indicated ratios for 10 min at room temperature. Biotin-labeled P*mqsRA* DNA was then added to the MqsR/MqsA complex and the entire reaction incubated at room temperature for an additional 20 min followed by electrophoresis. *Pre-formed MqsA/*P*mqsRA samples*: A constant amount of MqsA (400 fmol) was first incubated with P*mqsRA* DNA (50 fmol; 10 min, RT) then increasing amounts of MqsR added at the indicated ratios. MqsR amounts were 0, 50, 100, 150, 200, or 400 fmol. The control reaction (50 fmol biotin-labeled P*mqsRA* only) is shown in the first lane.

intensity of the band corresponding MqsA bound to only a single palindrome (Fig. 4A, lanes 3-4). At a 1:1 ratio of MqsR to MqsA, which is the same ratio of the complex in solution, practically all DNA binding has been abolished. Similar results were obtained regardless of whether MqsR was added to the pre-formed MqsA/DNA complex or whether DNA was added to the pre-formed MqsR/MqsA complex (Fig. 4B). Thus, these experiments show that at all concentrations, MqsR serves simply to dissociate the MqsA/DNA complex *in vitro* and never functions to enhance binding of MqsA to the operon promoter.

D.5. Discussion

The central finding of our work is that, unlike other type II toxins, MqsR does not enhance the binding of MqsA to the *mqsRA* promoter. Instead, under all circumstances, we show that MqsR functions solely to disrupt the MqsA/DNA interaction (Figs. 3, 4), a function due in part to the tight binding (K_D of 0.08 ± 0.07 nM; Fig. 2A) and exceptional stability of the MqsR/MqsA complex (Fig. 2B). This behavior is in stark contrast to most other TA systems, in which the toxin functions as a transcriptional co-repressor until a certain threshold is reached, above which it then functions as a de-repressor resulting in robust transcription (known as conditional cooperativity).

There are two proposed mechanisms that underlie conditional cooperativity which rely on interactions made either by the toxin monomer (mechanism 1) or the toxin dimer (mechanism 2). In the first mechanism, shown for the PhD/Doc [322] and RelB/RelE [323] TA systems, antitoxins interact with toxins at two separate interfaces present on the antitoxin, a high affinity site and

a low affinity site. At low levels of toxin, both sites are populated and the toxin molecules serve to bridge antitoxin dimers onto multiple DNA binding sites, which leads to increased affinity for DNA [322, 323]. However, when the toxin levels increase to exceed that of the antitoxin, only the high affinity binding sites on the antitoxin are populated. This alters the toxin/antitoxin complex stoichiometry, which in turn destabilizes DNA binding. In the second mechanism, which was recently reported for the VapB/VapC system [324], the VapC toxin dimerizes and these toxin dimers serve to bridge antitoxin molecules bound to multiple DNA sites to increase antitoxin avidity for DNA. However, when there is excess toxin, the toxin dimer interactions are broken which subsequently destabilizes antitoxin DNA binding.

We believe that MqsR cannot utilize either mechanism to serve as a transcriptional co-repressor for two reasons. First, there is no evidence for multiple binding interfaces between MqsR and MqsA. Our analysis of the MqsR/MqsA complex in solution has not revealed the presence of any oligomeric species larger than the MqsR-$MqsA_2$-MqsR hetero-tetramer, as determined using size exclusion chromatography (Fig. 1A). Second, MqsR exists as a monomer in solution and has not been demonstrated to dimerize, which we also showed using size exclusion chromatography (Fig. 1A). Therefore, MqsR cannot serve as a co-repressor either alone (mechanism 1) or as a dimer (mechanism 2). Instead, and in contrast to all other characterized TA systems, even sub-equimolar quantities of MqsR toxin serve to destabilize the MqsA/DNA interaction (Fig. 4). In fact, at a 1:1 ratio of MqsR toxin to MqsA antitoxin, DNA binding by MqsA is

essentially ablated. This behavior is unique to *mqsRA* because in all other TA systems, transcriptional de-repression occurs only when toxin concentrations exceed their normal stoichiometric levels.

A detailed analysis of the X-ray crystal structures of the various states of MqsR and MqsA provides insight as to why MqsR destabilizes the interaction between MqsA and the *mqsRA* promoter. A model of the hypothetical MqsR-MqsA-DNA ternary complex based on our crystal structures of MqsA bound to DNA (PDB 3O9X) and MqsR bound to the N-terminal domain of MqsA (PDB 3HI2) revealed the potential for clashes between the DNA and MqsR when both are bound to MqsA (Fig. 5A). Moreover, the binding sites on MqsA for both MqsR and DNA partially overlap as MqsA Arg61 forms interactions with both binding partners in the respective crystal structures (Fig. 5B-C). In support of the role of Arg61 in DNA and MqsR binding, mutation of this residue to either alanine (R61A) or aspartic acid (R16D) leads to both a decrease in DNA binding, as determined using EMSAs (Figure 5), and a decrease in the thermal stability of the MqsR/MqsA*mut* complexes, as determined using CD thermal denaturation (Table 2; Supplemental Fig. S2). These observations suggested two possibilities: (1) that a conformational rearrangement in one or more of the macromolecules is necessary for simultaneous binding, or (2) that all three macromolecules are not capable of forming a single complex.

Here, we show that the MqsR-MqsA-DNA ternary complex does not exist *in vitro*. Instead, MqsR, which binds the MqsA N-terminal domain with a 0.08 nM K_D, prevents this domain from interacting with the DNA due to steric hindrance.

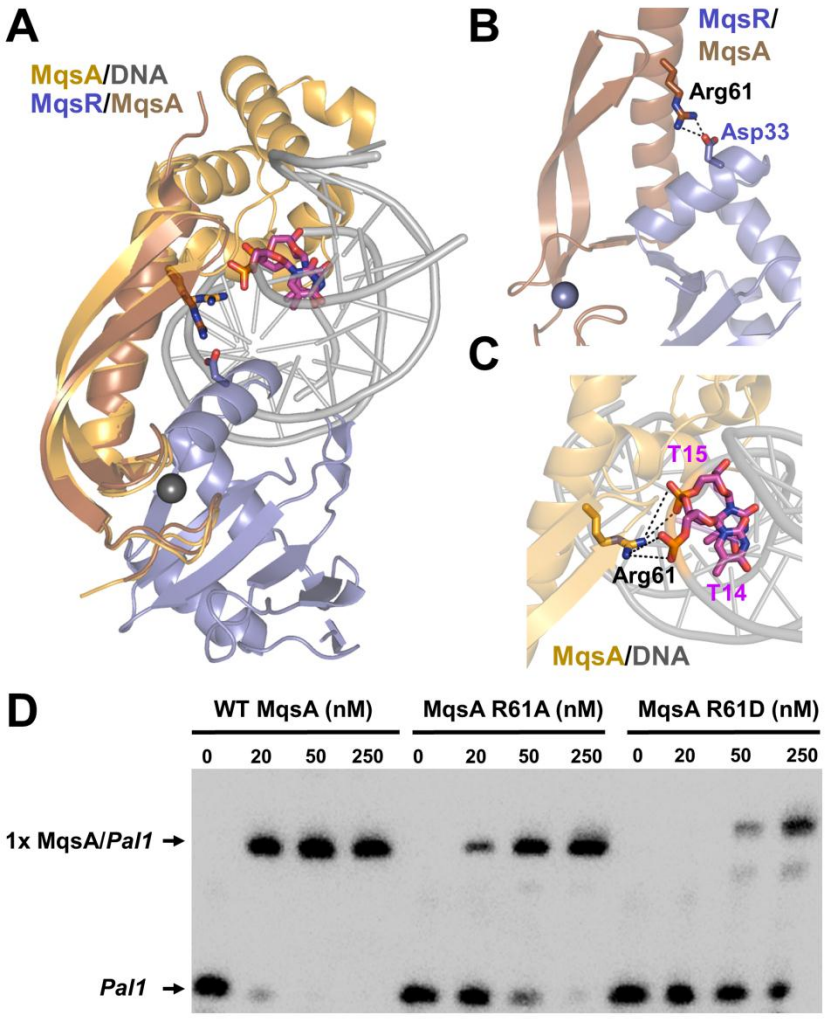


Figure 92: FIGURE 5, Brown, JBC. MqsR and DNA have overlapping binding sites on MqsA.

A, *Left panel*: Model of the hypothetical MqsR-MqsA-P*mqsRA* ternary complex. The model was generated by superimposing the MqsR/MqsA-N crystal structure (PDB 3HI2; colored blue/brown, respectively) with the MqsA/DNA crystal structure (PDB 3O9X; colored orange/gray, respectively) using the respective MqsA N-terminal domains; for clarity, only one monomer of MqsA from the MqsA/DNA crystal structure is shown. As modeled, there are structural clashes between MqsR (blue surface) and the DNA (grey sticks). The location of the clash is highlighted by a blue dashed line. *Right panel*: Model of the hypothetical MqsR-MqsA-P*mqsRA* ternary complex created as in left panel, but with MqsR shown in cartoon representation. The MqsA binding sites for MqsR and the P*mqsRA* DNA partially overlap as MqsA residue Arg61 provides electrostatic interactions in both the toxin- and DNA-bound states. MqsR-bound MqsA is colored in brown and DNA-bound MqsA is colored in orange with Arg61 from both structures shows as sticks. Asp33 (MqsR, blue) and the phosphate backbone from nucleotides Thymine 14 and 15 (pink) are also shown; nitrogen atoms in dark blue, oxygen atoms in red. *B*, Zoom in view of overlay in (*A*) of just the MqsR/MqsA-N structure. MqsA residue Arg61 forms a salt bridge with MqsR Asp33. *C,* Zoom in view of overlay in (*A*) of just the MqsA/DNA structure. MqsA residue Arg61 forms a salt bridge with the phosphate backbone of nucleotides Thymine 14 and 15 from one strand. Electrostatic interactions in (*B*) and (*C*) are indicated by black dashed lines. *D*, An EMSA with increasing amounts of WT MqsA (lanes 1-4), MqsA R61A (lanes 5-8), or MqsA R61D (lanes 9-12) incubated with biotin-labeled Palindrome 1 (Pal1) of P*mqsRA.*

Although, in isolation, the MqsA N-terminal domains do not bind DNA, the interactions they provide within the context of full-length MqsA are important, as we previously showed that the binding affinity of MqsA for DNA decreases approximately 50-fold in the absence of the N-terminal domain interactions, from a K_D of 0.8 nM for full-length MqsA to 41.5 nM for the MqsA C-terminal domain [306]. Consequently, MqsR binding results in a decrease in the effective binding affinity of MqsA for DNA and, at equimolar concentrations of MqsR, results in the disruption of the MqsA/DNA complex and the formation only of the MqsR/MqsA complex. Thus, unlike most toxins, MqsR does not act as a transcriptional co-repressor but instead a transcriptional de-repressor. Critically, this newly-identified function of MqsR may also explain previous results which showed that overexpression of MqsR led to an increase in the transcription of genes known to be repressed by MqsA, including *cspD*, *rpoS*, and *mqsA* itself [305]. Therefore, both *in vitro* (these studies) and *in vivo* experiments [305] show that MqsR also functions as a transcriptional de-repressor.

We previously determined that the *mqsRA* toxin/antitoxin module is unique among *E. coli* TA modules based upon the novel features of the MqsA antitoxin. However, it is now becoming evident that the MqsR toxin also displays unusual toxin characteristics which, together with the unique features of the MqsA antitoxin, may also lead to differences in transcriptional regulation of the *mqsRA* operon. We report that the MqsR/MqsA complex does not bind its own operon promoter, which would lead to transcriptional repression, as is common for several other TA systems. Instead, in all instances, the MqsR toxin serves solely

to disrupt the MqsA/DNA complex *in vitro* and thus results in transcriptional activation. In addition, we show that MqsR is capable of cleaving mRNA corresponding not only to *mqsA*, but to *mqsR* as well. This suggests that MqsR may auto-regulate its own activity, which could allow for rapid exit from the persister state upon removal of environmental stress. Expression of MqsR was previously shown to play a role in persistence as well as biofilm formation through positively regulating expression of various genes including the toxin *cspD* and the two-component motility regulatory system *qseBC* [255, 305]. Since MqsA regulates multiple genes that play a vital role in *E. coli* physiology, including activation of the general stress response [302], the function of MqsR in alleviating MqsA-mediated repression would further elucidate the mechanism by which MqsR exerts some of its cellular effects. Thus, MqsR may be a central regulator of bacterial persistence not only through its function as an endoribonuclease but also through its role in indirectly activating transcription of other genes that lead to cell dormancy.

D.6. Supplementary figures

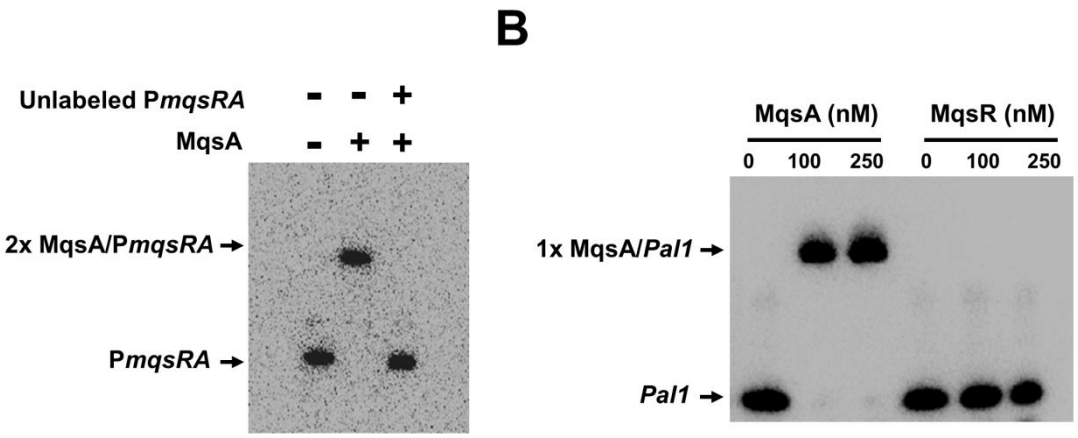


Figure 93: Figure S1, Brown, JBC. MqsA binds P*mqsRA* specifically.
(A) EMSA in the presence and absence of unlabeled P*mqsRA* DNA. In lane 2, MqsA was incubated with 100 fmol of biotin-labeled P*mqsRA* DNA. Lane 3 is the same as lane 2 except MqsA was incubated first with a 2-fold excess of unlabeled P*mqsRA* DNA. (B) EMSA with increasing amounts of MqsA (lanes 2-3) or MqsR (5-6) incubated with 100 fmol biotin-labeled Palindrome 1 (Pal1).

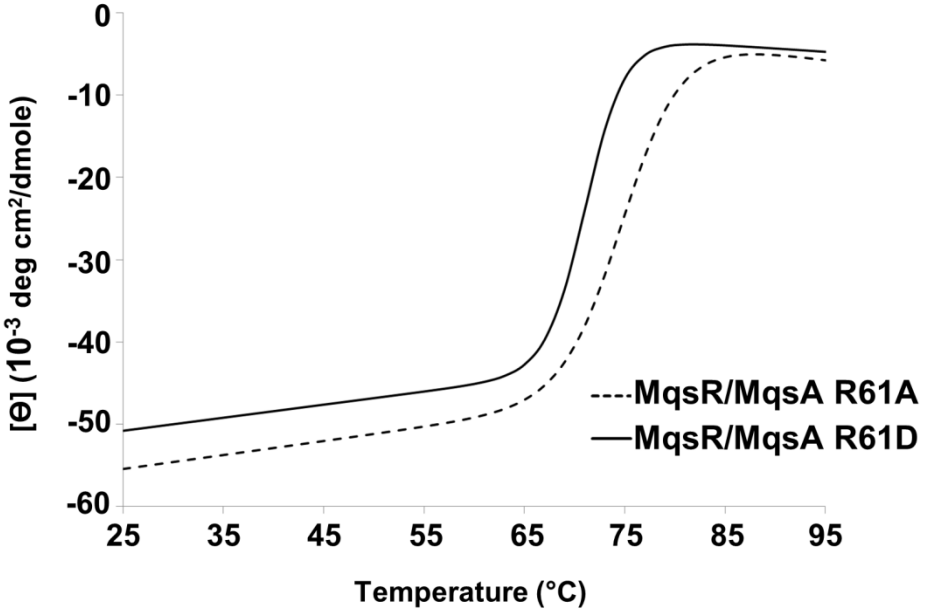


Figure 94: Figure S2, Brown, JBC. Thermal denaturation at 208 nm of the MqsR/MqsA R61 mutants.
The T_m for MqsR/MqsA R61A (5 μM, dashed line) and MqsR/MqsA R61D (5 μM, solid line) are 75.2 ± 0.4°C and 70.5 ± 1.0°, respectively.

D.7. Supplementary Tables

Table 12: TABLE S1, Brown, JBC. PCR primers used to generate templates for T7 RNA polymerase *in vitro* transcription.
F indicates forward primer and *R* indicates reverse primer. The T7 promoter sequence is underlined and the bases incorporated into mRNA during transcription are in bold.

RNA	Transcript size (nt)	Primer Name	Sequence (5' - 3')
mqsA	415	PT7-*mqsRA-F*	TAATACGACTCACTATA**GGGAGAAGGAG ATATACAT**
		PT7-*mqsA-R*	TTAACGGATTTCATTCAATAGTTCTGG
mqsR	315	PT7-*mqsRA-F*	TAATACGACTCACTATA**GGGAGAAGGAG ATATACAT**
		PT7-*mqsR-R*	TTACTTCTCCTTAAACGAGACGAT

Table 13: TABLE S2, Brown, JBC. PCR primers used to generate DNA for EMSA experiments.
F indicates forward primer and *R* indicates reverse primer.

DNA	Transcript size (nt)	Primer Name	Sequence (5' - 3')
P*mqsRA*	234	P*mqsRA-F*	GTGATGCCTGACTCCAGCTT
		P*mqsRA-R*	CGTGTATGTGGTGTGCGTTT
Pal1	26	Pal1-*F*	TCTAACTAACCTTTTAGGTGCTTTTT
		Pal1-*R*	AAAAAGCACCTAAAAGGTTAGTTAGA
Pal2	26	Pal2-*F*	TGTAATTAACCTTTTAGGTTATAACT
		Pal2-*R*	AGTTATAACCTAAAAGGTTAATTACA

List of References

1. Kissinger, C.R., et al., *Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex.* Nature, 1995. **378**(6557): p. 641-4.
2. Tait, S.W., et al., *Mechanism of inactivation of NF-kappa B by a viral homologue of I kappa b alpha. Signal-induced release of i kappa b alpha results in binding of the viral homologue to NF-kappa B.* J Biol Chem, 2000. **275**(44): p. 34656-64.
3. Stathopoulos, A.M. and M.S. Cyert, *Calcineurin acts through the CRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast.* Genes Dev, 1997. **11**(24): p. 3432-44.
4. Li, H., et al., *Structure of calcineurin in complex with PVIVIT peptide: portrait of a low-affinity signalling interaction.* J Mol Biol, 2007. **369**(5): p. 1296-306.
5. Ragusa, M.J., et al., *Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites.* Nat Struct Mol Biol, 2010. **17**(4): p. 459-64.
6. Kamada, K. and F. Hanaoka, *Conformational change in the catalytic site of the ribonuclease YoeB toxin by YefM antitoxin.* Mol. Cell., 2005. **19**(4): p. 497-509.
7. Kuboniwa, H., et al., *Solution structure of calcium-free calmodulin.* Nat Struct Biol, 1995. **2**(9): p. 768-76.
8. Garcia-Contreras, R., et al., *Protein translation and cell death: the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes.* PLoS One, 2008. **3**(6): p. e2394.
9. Watkins, D., et al., *P22 c2 repressor-operator complex: mechanisms of direct and indirect readout.* Biochemistry, 2008. **47**(8): p. 2325-2338.
10. Li, H., et al., *Balanced interactions of calcineurin with AKAP79 regulate Ca(2+)-calcineurin-NFAT signaling.* Nat Struct Mol Biol, 2012. **19**(3): p. 337-45.
11. Chattopadhyaya, R., et al., *Calmodulin structure refined at 1.7 Å resolution.* Journal of Molecular Biology, 1992. **228**(4): p. 1177-1192.
12. Li, G.Y., et al., *Inhibitory mechanism of Escherichia coli RelE-RelB toxin-antitoxin module involves a helix displacement near an mRNA interferase active site.* J. Biol. Chem., 2009. **284**(21): p. 14628-14636.
13. Cho, U.S. and W. Xu, *Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme.* Nature, 2007. **445**(7123): p. 53-7.
14. Sevcik, J., et al., *Complex of ribonuclease Sa with a cyclic nucleotide and a proposed model for the reaction intermediate.* Eur. J. Biochem., 1993. **216**(1): p. 301-305.
15. Musson, R.E., C.M. Cobbaert, and N.P. Smit, *Molecular diagnostics of calcineurin-related pathologies.* Clin Chem, 2012. **58**(3): p. 511-22.
16. Pereira, S.R., V.M. Vasconcelos, and A. Antunes, *The phosphoprotein phosphatase family of Ser/Thr phosphatases as principal targets of naturally occurring toxins.* Crit Rev Toxicol, 2011. **41**(2): p. 83-110.

17. Huai, Q., et al., *Crystal structure of calcineurin-cyclophilin-cyclosporin shows common but distinct recognition of immunophilin-drug complexes.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12037-42.
18. Jin, L. and S.C. Harrison, *Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13522-6.
19. Shaw, K.T., et al., *Immunosuppressive drugs prevent a rapid dephosphorylation of transcription factor NFAT1 in stimulated immune cells.* Proc Natl Acad Sci U S A, 1995. **92**(24): p. 11205-9.
20. Jenkins, M.K., R.H. Schwartz, and D.M. Pardoll, *Effects of cyclosporine A on T cell development and clonal deletion.* Science, 1988. **241**(4873): p. 1655-8.
21. Gao, E.K., et al., *Abnormal differentiation of thymocytes in mice treated with cyclosporin A.* Nature, 1988. **336**(6195): p. 176-9.
22. Textor, S.C., et al., *Systemic and renal hemodynamic differences between FK506 and cyclosporine in liver transplant recipients.* Transplantation, 1993. **55**(6): p. 1332-9.
23. Abu-Elmagd, K., et al., *The effect of graft function on FK506 plasma levels, dosages, and renal function, with particular reference to the liver.* Transplantation, 1991. **52**(1): p. 71-7.
24. Pham, S.M., et al., *A prospective trial of tacrolimus (FK 506) in clinical heart transplantation: intermediate-term results.* J Thorac Cardiovasc Surg, 1996. **111**(4): p. 764-72.
25. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
26. Clapham, D.E., *Calcium signaling.* Cell, 2007. **131**(6): p. 1047-58.
27. Graves, J.D. and E.G. Krebs, *Protein Phosphorylation and Signal Transduction.* Pharmacology & Therapeutics, 1999. **82**(2–3): p. 111-121.
28. Moorhead, G.B.G., L. Trinkle-Mulcahy, and A. Ulke-Lemee, *Emerging roles of nuclear protein phosphatases.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 234-244.
29. Feske, S., A. Rao, and P.G. Hogan, *The Ca2+–calcineurin–NFAT signalling pathway*, in *New Comprehensive Biochemistry*, K. Joachim and M. Marek, Editors. 2007, Elsevier. p. 365-401.
30. Aramburu, J., A. Rao, and C.B. Klee, *Calcineurin: from structure to function.* Curr Top Cell Regul, 2000. **36**: p. 237-95.
31. Hogan, P.G., R.S. Lewis, and A. Rao, *Molecular basis of calcium signaling in lymphocytes: STIM and ORAI.* Annu Rev Immunol, 2010. **28**: p. 491-533.
32. Ogawa, Y. and M. Tanokura, *Calcium binding to calmodulin: effects of ionic strength, Mg2+, pH and temperature.* J Biochem, 1984. **95**(1): p. 19-28.
33. Potter, J.D., et al., *Ca2+ binding to calmodulin.* Methods Enzymol, 1983. **102**: p. 135-43.
34. Haiech, J., C.B. Klee, and J.G. Demaille, *Effects of cations on affinity of calmodulin for calcium: ordered binding of calcium ions allows the specific*

activation of calmodulin-stimulated enzymes. Biochemistry, 1981. **20**(13): p. 3890-7.
35. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.
36. Sacco, F., et al., *The human phosphatase interactome: An intricate family portrait.* FEBS Letters, 2012. **586**(17): p. 2732-2739.
37. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.
38. Alonso, A., et al., *Protein tyrosine phosphatases in the human genome.* Cell, 2004. **117**(6): p. 699-711.
39. Peti, W., A.C. Nairn, and R. Page, *Structural basis for protein phosphatase 1 regulation and specificity.* FEBS J, 2012.
40. Shi, Y., *Serine/threonine phosphatases: mechanism through structure.* Cell, 2009. **139**(3): p. 468-84.
41. Choy, M.S., R. Page, and W. Peti, *Regulation of protein phosphatase 1 by intrinsically disordered proteins.* Biochem Soc Trans, 2012. **40**(5): p. 969-74.
42. Li, H., A. Rao, and P.G. Hogan, *Interaction of calcineurin with substrates and targeting proteins.* Trends Cell Biol, 2011. **21**(2): p. 91-103.
43. Stemmer, P.M. and C.B. Klee, *Dual calcium ion regulation of calcineurin by calmodulin and calcineurin B.* Biochemistry, 1994. **33**(22): p. 6859-66.
44. Feng, B. and P.M. Stemmer, *Interactions of calcineurin A, calcineurin B, and Ca2+.* Biochemistry, 1999. **38**(38): p. 12481-9.
45. Rusnak, F. and P. Mertz, *Calcineurin: form and function.* Physiol Rev, 2000. **80**(4): p. 1483-521.
46. Yang, S.A. and C.B. Klee, *Low affinity Ca2+-binding sites of calcineurin B mediate conformational changes in calcineurin A.* Biochemistry, 2000. **39**(51): p. 16147-54.
47. Shen, X., et al., *The secondary structure of calcineurin regulatory region and conformational change induced by calcium/calmodulin binding.* J Biol Chem, 2008. **283**(17): p. 11407-13.
48. Wang, H., et al., *A renewed model of CNA regulation involving its C-terminal regulatory domain and CaM.* Biochemistry, 2008. **47**(15): p. 4461-8.
49. Rao, A., C. Luo, and P.G. Hogan, *Transcription factors of the NFAT family: regulation and function.* Annu Rev Immunol, 1997. **15**: p. 707-47.
50. Feske, S., *Calcium signalling in lymphocyte activation and disease.* Nat Rev Immunol, 2007. **7**(9): p. 690-702.
51. Stroud, J.C. and L. Chen, *Structure of NFAT Bound to DNA as a Monomer.* Journal of Molecular Biology, 2003. **334**(5): p. 1009-1022.
52. Jin, L., et al., *An asymmetric NFAT1 dimer on a pseudo-palindromic kappa B-like DNA site.* Nat Struct Biol, 2003. **10**(10): p. 807-11.
53. Giffin, M.J., et al., *Structure of NFAT1 bound as a dimer to the HIV-1 LTR [kappa]B element.* Nat Struct Mol Biol, 2003. **10**(10): p. 800-806.

54. Crabtree, G.R. and E.N. Olson, *NFAT signaling: choreographing the social lives of cells.* Cell, 2002. **109 Suppl**: p. S67-79.
55. Loh, C., et al., *Calcineurin binds the transcription factor NFAT1 and reversibly regulates its activity.* J Biol Chem, 1996. **271**(18): p. 10884-91.
56. Masuda, E.S., et al., *Control of NFATx1 nuclear translocation by a calcineurin-regulated inhibitory domain.* Mol Cell Biol, 1997. **17**(4): p. 2066-75.
57. Wesselborg, S., et al., *Identification of a physical interaction between calcineurin and nuclear factor of activated T cells (NFATp).* J Biol Chem, 1996. **271**(3): p. 1274-7.
58. Aramburu, J., et al., *Selective inhibition of NFAT activation by a peptide spanning the calcineurin targeting site of NFAT.* Mol Cell, 1998. **1**(5): p. 627-37.
59. Garcia-Cozar, F.J., et al., *Two-site interaction of nuclear factor of activated T cells with activated calcineurin.* J Biol Chem, 1998. **273**(37): p. 23877-83.
60. Chow, C.W., M. Rincon, and R.J. Davis, *Requirement for transcription factor NFAT in interleukin-2 expression.* Mol Cell Biol, 1999. **19**(3): p. 2300-7.
61. Zhu, J., et al., *Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEKK1.* Cell, 1998. **93**(5): p. 851-61.
62. Liu, J., et al., *Two independent calcineurin-binding regions in the N-terminal domain of murine NF-ATx1 recruit calcineurin to murine NF-ATx1.* J Immunol, 1999. **162**(8): p. 4755-61.
63. Park, S., M. Uesugi, and G.L. Verdine, *A second calcineurin binding site on the NFAT regulatory domain.* Proc Natl Acad Sci U S A, 2000. **97**(13): p. 7130-5.
64. Rodriguez, A., et al., *A conserved docking surface on calcineurin mediates interaction with substrates and immunosuppressants.* Mol Cell, 2009. **33**(5): p. 616-26.
65. Martinez-Martinez, S., et al., *Blockade of NFAT activation by the second calcineurin binding site.* J Biol Chem, 2006. **281**(10): p. 6227-35.
66. Mehta, S., et al., *Domain architecture of the regulators of calcineurin (RCANs) and identification of a divergent RCAN in yeast.* Mol Cell Biol, 2009. **29**(10): p. 2777-93.
67. Aramburu, J., et al., *Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A.* Science, 1999. **285**(5436): p. 2129-33.
68. Takeuchi, K., et al., *Structure of the calcineurin-NFAT complex: defining a T cell activation switch using solution NMR and crystal coordinates.* Structure, 2007. **15**(5): p. 587-97.
69. Aramburu, J., et al., *Affinity-Driven Peptide Selection of an NFAT Inhibitor More Selective Than Cyclosporin A.* Science, 1999. **285**(5436): p. 2129-2133.
70. Barik, S., *Immunophilins: for the love of proteins.* Cellular and Molecular Life Sciences CMLS, 2006. **63**(24): p. 2889-2900.

71. Liu, J., et al., *Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes.* Cell, 1991. **66**(4): p. 807-15.
72. Briesewitz, R., et al., *Affinity modulation of small-molecule ligands by borrowing endogenous protein surfaces.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1953-8.
73. Dixon, L.K., et al., *African swine fever virus proteins involved in evading host defence systems.* Vet Immunol Immunopathol, 2004. **100**(3-4): p. 117-34.
74. Abrams, C.C., et al., *Domains involved in calcineurin phosphatase inhibition and nuclear localisation in the African swine fever virus A238L protein.* Virology, 2008. **374**(2): p. 477-86.
75. Fuentes, J.J., et al., *DSCR1, overexpressed in Down syndrome, is an inhibitor of calcineurin-mediated signaling pathways.* Human Molecular Genetics, 2000. **9**(11): p. 1681-1690.
76. Ermak, G., T.E. Morgan, and K.J. Davies, *Chronic overexpression of the calcineurin inhibitory gene DSCR1 (Adapt78) is associated with Alzheimer's disease.* J Biol Chem, 2001. **276**(42): p. 38787-94.
77. Harris, C.D., G. Ermak, and K.J. Davies, *Multiple roles of the DSCR1 (Adapt78 or RCAN1) gene and its protein product calcipressin 1 (or RCAN1) in disease.* Cell Mol Life Sci, 2005. **62**(21): p. 2477-86.
78. Wang, W., et al., *DSCR1 interacts with FMRP and is required for spine morphogenesis and local protein synthesis.* EMBO J, 2012. **31**(18): p. 3655-66.
79. Ladbury, J.E., *Application of isothermal titration calorimetry in the biological sciences: things are heating up!* Biotechniques, 2004. **37**(6): p. 885-7.
80. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions.* Methods Cell Biol, 2008. **84**: p. 79-113.
81. Chaires, J.B., *Calorimetry and thermodynamics in drug design.* Annu Rev Biophys, 2008. **37**: p. 135-51.
82. Lavoisier, A.d.L., PS, *Memoire sur la chaleur.* Memoires de l'Academie des sciences, 1780: p. 355.
83. Frankenfield, D.C., *On heat, respiration, and calorimetry.* Nutrition, 2010. **26**(10): p. 939-950.
84. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal titration calorimetry of protein-protein interactions.* Methods, 1999. **19**(2): p. 213-21.
85. Freire, E.M., O.L.; Straume, M., *Isothermal titration calorimetry.* Anal Chem, 1990. **62**(18): p. 950-959.
86. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry.* Curr Opin Struct Biol, 2001. **11**(5): p. 560-6.
87. García-Fuentes, L., et al., *Thermodynamics of Molecular Recognition by Calorimetry*. Thermodynamics - Physical Chemistry of Aqueous Systems. 2011.

88. Velazquez-Campoy, A., I. Luque, and E. Freire, *The application of thermodynamic methods in drug design.* Thermochimica Acta, 2001. **380**(2): p. 217-227.
89. Morin, P.E. and E. Freire, *Direct calorimetric analysis of the enzymatic activity of yeast cytochrome c oxidase.* Biochemistry, 1991. **30**(34): p. 8494-500.
90. Watt, G.D., *A microcalorimetric procedure for evaluating the kinetic parameters of enzyme-catalyzed reactions: kinetic measurements of the nitrogenase system.* Anal Biochem, 1990. **187**(1): p. 141-6.
91. Williams, B.A. and E.J. Toone, *Calorimetric evaluation of enzyme kinetic parameters.* The Journal of Organic Chemistry, 1993. **58**(13): p. 3507-3510.
92. Todd, M.J. and J. Gomez, *Enzyme Kinetics Determined Using Calorimetry: A General Assay for Enzyme Activity?* Anal Biochem, 2001. **296**(2): p. 179-187.
93. Haq, I., *Calorimetry in the fast lane: the use of ITC for obtaining enzyme kinetic constants*, M. GE Healthcare Life Sciences, Editor 2005: www.microcalorimetry.com. p. 1-9.
94. Voet, D. and J.G. Voet, *Biochemistry*. 2004: John Wiley & Sons.
95. Grossoehme, N., A. Spuches, and D. Wilcox, *Application of isothermal titration calorimetry in bioinorganic chemistry.* JBIC Journal of Biological Inorganic Chemistry, 2010. **15**(8): p. 1183-1191.
96. Smith, P.K., et al., *Measurement of protein using bicinchoninic acid.* Anal Biochem, 1985. **150**(1): p. 76-85.
97. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**(1–2): p. 248-254.
98. Markova, N. and D. Hallén, *The development of a continuous isothermal titration calorimetric method for equilibrium studies.* Anal Biochem, 2004. **331**(1): p. 77-88.
99. Brown, A., *Analysis of cooperativity by isothermal titration calorimetry.* Int J Mol Sci, 2009. **10**(8): p. 3457-77.
100. Crabtree, G.R. and S.L. Schreiber, *SnapShot: Ca2+-calcineurin-NFAT signaling.* Cell, 2009. **138**(1): p. 210, 210 e1.
101. Musson, R.E. and N.P. Smit, *Regulatory mechanisms of calcineurin phosphatase activity.* Curr Med Chem, 2011. **18**(2): p. 301-15.
102. Roy, J. and M.S. Cyert, *Cracking the phosphatase code: docking interactions determine substrate specificity.* Sci Signal, 2009. **2**(100): p. re9.
103. Donella-Deana, A., et al., *Dephosphorylation of phosphopeptides by calcineurin (protein phosphatase 2B).* Eur J Biochem, 1994. **219**(1-2): p. 109-17.
104. Blumenthal, D.K., et al., *Dephosphorylation of cAMP-dependent protein kinase regulatory subunit (type II) by calmodulin-dependent protein phosphatase. Determinants of substrate specificity.* J Biol Chem, 1986. **261**(18): p. 8140-5.

105. Etzkorn, F.A., et al., *Cyclophilin residues that affect noncompetitive inhibition of the protein serine phosphatase activity of calcineurin by the cyclophilin.cyclosporin A complex.* Biochemistry, 1994. **33**(9): p. 2380-8.
106. Salowe, S.P. and J.D. Hermes, *Competitive and slow-binding inhibition of calcineurin by drug x immunophilin complexes.* Arch Biochem Biophys, 1998. **355**(2): p. 165-74.
107. Otwinowski, Z. and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode.* Methods in Enzym (part A), 1997. **276**: p. 307-326.
108. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
109. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
110. Lovell, S.C., et al., *Structure validation by Calpha geometry: phi,psi and Cbeta deviation.* Proteins, 2003. **50**(3): p. 437-50.
111. Griffith, J.P., et al., *X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex.* Cell, 1995. **82**(3): p. 507-22.
112. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* J Chem Phys, 1993. **98**: p. 5648-5652.
113. Ahlrichs, R., et al., *Electronic-Structure calculations on workstation computers: The program system TURBOMOLE.* Chem Phys Lett, 1989. **162**(3): p. 165-169.
114. Rohl, C.A., et al., *Modeling structurally variable regions in homologous proteins with rosetta.* Proteins, 2004. **55**(3): p. 656-77.
115. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulation liquit water.* J Chem Phys, 1983. **79**: p. 926-935.
116. Hess, B., et al., *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.* J Chem Theory Comput, 2008. **4**(3): p. 435-447.
117. Lindorff-Larsen, K., et al., *Improved side-chain torsion potentials for the Amber ff99SB protein force field.* Proteins, 2010. **78**(8): p. 1950-8.
118. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems.* J Chem Phys, 1993. **98**: p. 10089-10092.
119. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions.* Phys Rev A, 1985. **31**(3): p. 1695-1697.
120. Andersen, H.C., *Molecular-Dynamics simulations at constant pressure and/or temperature.* J Chem Phys, 1980. **72**: p. 2384-2393.
121. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method.* J Appl Phys, 1981. **52**(12): p. 7182-7190.
122. Piana, S. and A. Laio, *A bias-exchange approach to protein folding.* J Phys Chem B, 2007. **111**(17): p. 4553-9.

123. Bonomi, M., et al., *PLUMED:A portable plugin for free-energy calculations with molecular dynamics.* Comput Phys Commun, 2009. **180**(10): p. 1961-1972.
124. Marinelli, F., et al., *A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations.* PLoS Comput Biol, 2009. **5**(8): p. e1000452.
125. Akaike, H., *A new look at the statistical model identification.* IEEE Transactions on Automatic Control, 1974. **19**: p. 716-723.
126. Roy, J., et al., *A conserved docking site modulates substrate affinity for calcineurin, signaling output, and in vivo function.* Mol Cell, 2007. **25**(6): p. 889-901.
127. Perrino, B.A., *Regulation of calcineurin phosphatase activity by its autoinhibitory domain.* Arch Biochem Biophys, 1999. **372**(1): p. 159-65.
128. Muller, M.R., et al., *Requirement for balanced Ca/NFAT signaling in hematopoietic and embryonic development.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7034-9.
129. Vega, R.B., et al., *Multiple domains of MCIP1 contribute to inhibition of calcineurin activity.* J Biol Chem, 2002. **277**(33): p. 30401-7.
130. Dougherty, M.K., et al., *KSR2 is a calcineurin substrate that promotes ERK cascade activation in response to calcium signals.* Mol Cell, 2009. **34**(6): p. 652-62.
131. Kafadar, K.A. and M.S. Cyert, *Integration of stress responses: modulation of calcineurin signaling in Saccharomyces cerevisiae by protein kinase A.* Eukaryot Cell, 2004. **3**(5): p. 1147-53.
132. Kafadar, K.A., et al., *Negative regulation of calcineurin signaling by Hrr25p, a yeast homolog of casein kinase I.* Genes Dev, 2003. **17**(21): p. 2698-708.
133. Okamura, H., et al., *Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity.* Mol Cell, 2000. **6**(3): p. 539-50.
134. Sopko, R., et al., *Mapping pathways and phenotypes by systematic gene overexpression.* Mol Cell, 2006. **21**(3): p. 319-30.
135. Porta, S., et al., *Differential expression of members of the RCAN family of calcineurin regulators suggests selective functions for these proteins in the brain.* Eur J Neurosci, 2007. **26**(5): p. 1213-26.
136. Kingsbury, T.J. and K.W. Cunningham, *A conserved family of calcineurin regulators.* Genes Dev, 2000. **14**(13): p. 1595-604.
137. Rothermel, B., et al., *A Protein Encoded within the Down Syndrome Critical Region Is Enriched in Striated Muscles and Inhibits Calcineurin Signaling.* Journal of Biological Chemistry, 2000. **275**(12): p. 8719-8725.
138. Rothermel, B.A., et al., *Myocyte-enriched calcineurin-interacting protein, MCIP1, inhibits cardiac hypertrophy in vivo.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3328-33.
139. Ryeom, S., et al., *The threshold pattern of calcineurin-dependent gene expression is altered by loss of the endogenous inhibitor calcipressin.* Nat Immunol, 2003. **4**(9): p. 874-81.

140. Martinez-Martinez, S., et al., *The RCAN carboxyl end mediates calcineurin docking-dependent inhibition via a site that dictates binding to substrates and regulators.* Proc Natl Acad Sci U S A, 2009. **106**(15): p. 6117-22.
141. Chan, B., et al., *Identification of a peptide fragment of DSCR1 that competitively inhibits calcineurin activity in vitro and in vivo.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13075-80.
142. Vega, R.B., et al., *Dual roles of modulatory calcineurin-interacting protein 1 in cardiac hypertrophy.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 669-74.
143. Sanna, B., et al., *Modulatory calcineurin-interacting proteins 1 and 2 function as calcineurin facilitators in vivo.* Proc Natl Acad Sci U S A, 2006. **103**(19): p. 7327-32.
144. Grigoriu, S., et al., *The molecular mechanism of substrate engagement and immunosuppressant inhibition of calcineurin.* PLoS Biol, 2013. **11**(2): p. e1001492.
145. Abbasi, S., et al., *Protein kinase-mediated regulation of calcineurin through the phosphorylation of modulatory calcineurin-interacting protein 1.* J Biol Chem, 2006. **281**(12): p. 7717-26.
146. Lee, E.J., et al., *NF-kappaB-inducing kinase phosphorylates and blocks the degradation of Down syndrome candidate region 1.* J Biol Chem, 2008. **283**(6): p. 3392-400.
147. Jung, M.S., et al., *Regulation of RCAN1 protein activity by Dyrk1A protein-mediated phosphorylation.* J Biol Chem, 2011. **286**(46): p. 40401-12.
148. Carme Mulero, M., et al., *A fluorescent polarization-based assay for the identification of disruptors of the RCAN1-calcineurin A protein complex.* Anal Biochem, 2010. **398**(1): p. 99-103.
149. Hilioti, Z., et al., *GSK-3 kinases enhance calcineurin signaling by phosphorylation of RCNs.* Genes Dev, 2004. **18**(1): p. 35-47.
150. Pfister, S.C., et al., *Mutational analyses of the signals involved in the subcellular location of DSCR1.* BMC Cell Biol, 2002. **3**: p. 24.
151. Cook, C.N., et al., *Expression of calcipressin1, an inhibitor of the phosphatase calcineurin, is altered with aging and Alzheimer's disease.* J Alzheimers Dis, 2005. **8**(1): p. 63-73.
152. Liu, Q., J.C. Busby, and J.D. Molkentin, *Interaction between TAK1-TAB1-TAB2 and RCAN1-calcineurin defines a signalling nodal control point.* Nat Cell Biol, 2009. **11**(2): p. 154-161.
153. Fuentes, J.-J., et al., *A new human gene from the Down syndrome critical region encodes a proline-rich protein highly expressed in fetal brain and heart.* Human Molecular Genetics, 1995. **4**(10): p. 1935-1944.
154. Arron, J.R., et al., *NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21.* Nature, 2006. **441**(7093): p. 595-600.
155. Peti, W. and R. Page, *Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost.* Protein Expr Purif, 2007. **51**(1): p. 1-10.

156. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Sci, 2006. **15**(12): p. 2795-804.
157. Zhang, H., S. Neal, and D.S. Wishart, *RefDB: a database of uniformly referenced protein chemical shifts.* J Biomol NMR, 2003. **25**(3): p. 173-95.
158. Dosztanyi, Z., et al., *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, 2005. **21**(16): p. 3433-4.
159. Wang, W., et al., *DSCR1 interacts with FMRP and is required for spine morphogenesis and local protein synthesis.* EMBO J, 2012. **31**(18): p. 3655-3666.
160. Seiler, M., et al., *The 3of5 web application for complex and comprehensive pattern matching in protein sequences.* BMC Bioinformatics, 2006. **7**: p. 144.
161. Bultynck, G., et al., *Slm1 and slm2 are novel substrates of the calcineurin phosphatase required for heat stress-induced endocytosis of the yeast uracil permease.* Mol Cell Biol, 2006. **26**(12): p. 4729-45.
162. Park, C.Y., et al., *STIM1 clusters and activates CRAC channels via direct binding of a cytosolic domain to Orai1.* Cell, 2009. **136**(5): p. 876-90.
163. Dyer, B.W., et al., *A noncommercial dual luciferase enzyme assay system for reporter gene analysis.* Anal Biochem, 2000. **282**(1): p. 158-61.
164. Kersse, K., et al., *NOD-like receptors and the innate immune system: coping with danger, damage and death.* Cytokine Growth Factor Rev, 2011. **22**(5-6): p. 257-76.
165. Martinon, F. and J. Tschopp, *NLRs join TLRs as innate sensors of pathogens.* Trends Immunol, 2005. **26**(8): p. 447-54.
166. Fritz, J.H., et al., *Nod-like proteins in immunity, inflammation and disease.* Nat Immunol, 2006. **7**(12): p. 1250-7.
167. Martinon, F., K. Burns, and J. Tschopp, *The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta.* Mol Cell, 2002. **10**(2): p. 417-26.
168. Petrilli, V., et al., *The inflammasome: a danger sensing complex triggering innate immunity.* Curr Opin Immunol, 2007. **19**(6): p. 615-22.
169. Ting, J.P., et al., *The NLR gene family: a standard nomenclature.* Immunity, 2008. **28**(3): p. 285-7.
170. Bao, Q. and Y. Shi, *Apoptosome: a platform for the activation of initiator caspases.* Cell Death Differ, 2007. **14**(1): p. 56-65.
171. Martinon, F., A. Mayor, and J. Tschopp, *The inflammasomes: guardians of the body.* Annu Rev Immunol, 2009. **27**: p. 229-65.
172. Hoffman, H.M., et al., *Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome.* Nat Genet, 2001. **29**(3): p. 301-5.
173. Villani, A.C., et al., *Common variants in the NLRP3 region contribute to Crohn's disease susceptibility.* Nat Genet, 2009. **41**(1): p. 71-6.

174. Grenier, J.M., et al., *Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-kappaB and caspase-1.* FEBS Lett, 2002. **530**(1-3): p. 73-8.
175. Cui, J., et al., *NLRP4 negatively regulates type I interferon signaling by targeting the kinase TBK1 for degradation via the ubiquitin ligase DTX4.* Nat Immunol, 2012.
176. Fiorentino, L., et al., *A novel PAAD-containing protein that modulates NF-kappa B induction by cytokines tumor necrosis factor-alpha and interleukin-1beta.* J Biol Chem, 2002. **277**(38): p. 35333-40.
177. Zhang, P., et al., *Expression analysis of the NLRP gene family suggests a role in human preimplantation development.* PLoS One, 2008. **3**(7): p. e2755.
178. Jounai, N., et al., *NLRP4 negatively regulates autophagic processes through an association with beclin1.* J Immunol, 2011. **186**(3): p. 1646-55.
179. Bae, J.Y. and H.H. Park, *Crystal structure of NALP3 PYD domain and its implications in inflammasome assembly.* J Biol Chem, 2011.
180. Hiller, S., et al., *NMR structure of the apoptosis- and inflammation-related NALP1 pyrin domain.* Structure, 2003. **11**(10): p. 1199-205.
181. Liepinsh, E., et al., *The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition.* J Mol Biol, 2003. **332**(5): p. 1155-63.
182. Pinheiro, A.S., et al., *The NLRP12 Pyrin Domain: Structure, Dynamics, and Functional Insights.* J Mol Biol, 2011. **413**(4): p. 790-803.
183. Pinheiro, A.S., et al., *Three-dimensional structure of the NLRP7 pyrin domain: insight into pyrin-pyrin-mediated effector domain signaling in innate immunity.* J Biol Chem, 2010. **285**(35): p. 27402-10.
184. Mueller, U., et al., *Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin.* J Synchrotron Radiat, 2012. **19**(Pt 3): p. 442-9.
185. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
186. Sheldrick, G.M., *A short history of SHELX.* Acta Crystallogr A, 2008. **64**(Pt 1): p. 112-22.
187. Sheldrick, G.M., *Experimental phasing with SHELXC/D/E: combining chain tracing with density modification.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 479-85.
188. Corpet, F., *Multiple sequence alignment with hierarchical clustering.* Nucleic Acids Res, 1988. **16**(22): p. 10881-90.
189. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.
190. Sippl, M.J., *On distance and similarity in fold space.* Bioinformatics, 2008. **24**(6): p. 872-3.
191. Sippl, M.J. and M. Wiederstein, *A note on difficult structure alignment problems.* Bioinformatics, 2008. **24**(3): p. 426-7.
192. Sippl, M.J. and M. Wiederstein, *Detection of spatial correlations in protein structures and molecular complexes.* Structure, 2012. **20**(4): p. 718-28.

193. Reynolds, C., D. Damerell, and S. Jones, *ProtorP: a protein-protein interaction analysis server.* Bioinformatics, 2009. **25**(3): p. 413-4.
194. Schrödinger, L., *The PyMOL Molecular Graphics System, Version 1.2r3pre.*
195. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
196. Richter, S., et al., *webPIPSA: a web server for the comparison of protein interaction properties.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W276-80.
197. Gietz, R.D. and R.A. Woods, *Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.* Methods Enzymol, 2002. **350**: p. 87-96.
198. Wagner, R.N., et al., *Evaluation of Nod-like receptor (NLR) effector domain interactions.* PLoS One, 2009. **4**(4): p. e4931.
199. Natarajan, A., R. Ghose, and J.M. Hill, *Structure and dynamics of ASC2, a pyrin domain-only protein that regulates inflammatory signaling.* J Biol Chem, 2006. **281**(42): p. 31863-75.
200. Jaroszewski, L., et al., *FFAS03: a server for profile--profile sequence alignments.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W284-8.
201. Stehlik, C., et al., *The PAAD/PYRIN-only protein POP1/ASC2 is a modulator of ASC-mediated nuclear-factor-kappa B and pro-caspase-1 regulation.* Biochem J, 2003. **373**(Pt 1): p. 101-13.
202. Weber, C.H. and C. Vincenz, *The death domain superfamily: a tale of two interfaces?* Trends Biochem Sci, 2001. **26**(8): p. 475-81.
203. Srimathi, T., et al., *Monomer/dimer transition of the caspase-recruitment domain of human Nod1.* Biochemistry, 2008. **47**(5): p. 1319-25.
204. Manon, F., et al., *Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK.* J Mol Biol, 2007. **365**(1): p. 160-74.
205. Srimathi, T., et al., *Purification, crystallization and preliminary crystallographic characterization of the caspase-recruitment domain of human Nod1.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2007. **63**(Pt 1): p. 21-3.
206. Coussens, N.P., et al., *Crystal structure of the Nod1 caspase activation and recruitment domain.* Biochem Biophys Res Commun, 2007. **353**(1): p. 1-5.
207. Tattoli, I., et al., *The Nodosome: Nod1 and Nod2 control bacterial infections and inflammation.* Semin Immunopathol, 2007. **29**(3): p. 289-301.
208. Kersse, K., et al., *The death-fold superfamily of homotypic interaction motifs.* Trends Biochem Sci, 2011. **36**(10): p. 541-52.
209. Innis, C.A., J. Shi, and T.L. Blundell, *Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis.* Protein Eng, 2000. **13**(12): p. 839-47.

210. Liu, T., et al., *Homology modeling provides insights into the binding mode of the PAAD/DAPIN/pyrin domain, a fourth member of the CARD/DD/DED domain family.* Protein Sci, 2003. **12**(9): p. 1872-81.
211. Srimathi, T., et al., *Mapping of POP1-binding site on pyrin domain of ASC.* J Biol Chem, 2008. **283**(22): p. 15390-8.
212. Park, H.H., et al., *Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex.* Cell, 2007. **128**(3): p. 533-46.
213. Scott, F.L., et al., *The Fas-FADD death domain complex structure unravels signalling by receptor clustering.* Nature, 2009. **457**(7232): p. 1019-22.
214. Zhou, P., et al., *Solution structure of Apaf-1 CARD and its interaction with caspase-9 CARD: a structural basis for specific adaptor/caspase interaction.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11265-70.
215. Manji, G.A., et al., *PYPAF1, a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF-kappa B.* J Biol Chem, 2002. **277**(13): p. 11570-5.
216. Martinon, F., K. Hofmann, and J. Tschopp, *The pyrin domain: a possible member of the death domain-fold family implicated in apoptosis and inflammation.* Curr Biol, 2001. **11**(4): p. R118-20.
217. Damiano, J.S., et al., *Heterotypic interactions among NACHT domains: implications for regulation of innate immune responses.* Biochem J, 2004. **381**(Pt 1): p. 213-9.
218. Goldschmidt, L., et al., *Toward rational protein crystallization: A Web server for the design of crystallizable protein variants.* Protein Sci, 2007. **16**(8): p. 1569-76.
219. Tian, X., G. Pascal, and P. Monget, *Evolution and functional divergence of NLRP genes in mammalian reproductive systems.* BMC Evol Biol, 2009. **9**: p. 202.
220. Qin, H., et al., *Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1.* Nature, 1999. **399**(6736): p. 549-57.
221. Esposito, D., et al., *Solution NMR investigation of the CD95/FADD homotypic death domain complex suggests lack of engagement of the CD95 C terminus.* Structure, 2010. **18**(10): p. 1378-90.
222. Park, H.H., *Structural analyses of death domains and their interactions.* Apoptosis, 2011. **16**(3): p. 209-20.
223. Bigger, J.W., *Treatment of staphylococcal infections with penicillin.* Lancet, 1944: p. 497-500.
224. Keren, I., et al., *Specialized persister cells and the mechanism of multidrug tolerance in Escherichia coli.* J. Bacteriol., 2004. **186**(24): p. 8172-8180.
225. Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch.* Science, 2004. **305**(5690): p. 1622-1625.
226. Black, D.S., et al., *Structure and organization of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis.* J. Bacteriol., 1991. **173**(18): p. 5732-5739.

227. Spoering, A.L. and K. Lewis, *Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials.* J. Bacteriol. , 2001. **183**(23): p. 6746-6751.
228. Shah, D., et al., *Persisters: a distinct physiological state of E. coli.* BMC Microbiol. , 2006. **6**: p. 53.
229. Gerdes, K., S.K. Christensen, and A. Lobner-Olesen, *Prokaryotic toxin-antitoxin stress response loci.* Nat. Rev. Microbiol., 2005. **3**(5): p. 371-382.
230. Magnuson, R.D., *Hypothetical functions of toxin-antitoxin systems.* J. Bacteriol. , 2007. **189**(17): p. 6089-6092.
231. Gerdes, K., P.B. Rasmussen, and S. Molin, *Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells.* Proc. Natl. Acad. Sci. U. S. A., 1986. **83**(10): p. 3116-3120.
232. Ogura, T. and S. Hiraga, *Mini-F plasmid genes that couple host cell division to plasmid proliferation.* Proc. Natl. Acad. Sci. U. S. A. , 1983. **80**(15): p. 4784-4788.
233. Aizenman, E., H. Engelberg-Kulka, and G. Glaser, *An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death.* Proc. Natl. Acad. Sci. U. S. A., 1996. **93**(12): p. 6059-6063.
234. Gotfredsen, M. and K. Gerdes, *The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family.* Mol. Microbiol., 1998. **29**(4): p. 1065-1076.
235. Masuda, Y., et al., *chpA and chpB, Escherichia coli chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100.* J. Bacteriol. , 1993. **175**(21): p. 6850-6856.
236. Grady, R. and F. Hayes, *Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of Enterococcus faecium.* Mol. Microbiol., 2003. **47**(5): p. 1419-1432.
237. Sevin, E.W. and F. Barloy-Hubler, *RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes.* Genome. Biol. , 2007. **8**(8): p. R155.
238. Van Melderen, L., P. Bernard, and M. Couturier, *Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria.* Mol. Microbiol. , 1994. **11**(6): p. 1151-1157.
239. Christensen, S.K., et al., *Overproduction of the Lon protease triggers inhibition of translation in Escherichia coli: involvement of the yefM-yoeB toxin-antitoxin system.* Mol. Microbiol., 2004. **51**(6): p. 1705-1717.
240. Lehnherr, H. and M.B. Yarmolinsky, *Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of Escherichia coli.* Proc. Natl. Acad. Sci. U. S. A., 1995. **92**(8): p. 3274-3277.
241. Hiraga, S., et al., *F plasmid ccd mechanism in Escherichia coli.* J. Bacteriol., 1986. **166**(1): p. 100-104.
242. Jiang, Y., et al., *ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase.* Mol. Microbiol., 2002. **44**(4): p. 971-979.

243. Maki, S., et al., *Modulation of DNA supercoiling activity of Escherichia coli DNA gyrase by F plasmid proteins. Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins.* J. Biol. Chem. , 1992. **267**(17): p. 12244-12251.
244. Christensen, S.K., et al., *Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA.* J. Mol. Biol., 2003. **332**(4): p. 809-819.
245. Pedersen, K., et al., *The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site.* Cell, 2003. **112**(1): p. 131-140.
246. Schumacher, M.A., et al., *Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB.* Science, 2009. **323**(5912): p. 396-401.
247. de Feyter, R., C. Wallace, and D. Lane, *Autoregulation of the ccd operon in the F plasmid.* Mol. Gen. Genet., 1989. **218**(3): p. 481-486.
248. Magnuson, R. and M.B. Yarmolinsky, *Corepression of the P1 addiction operon by Phd and Doc.* J. Bacteriol. , 1998. **180**(23): p. 6342-6351.
249. Tam, J.E. and B.C. Kline, *The F plasmid ccd autorepressor is a complex of CcdA and CcdB proteins.* Mol. Gen. Genet. , 1989. **219**(1-2): p. 26-32.
250. Takagi, H., et al., *Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects.* Nat. Struct. Mol. Biol. , 2005. **12**(4): p. 327-331.
251. Kamada, K., F. Hanaoka, and S.K. Burley, *Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition.* Mol. Cell., 2003. **11**(4): p. 875-884.
252. Motiejunaite, R., et al., *Escherichia coli dinJ-yafQ genes act as a toxin-antitoxin module.* FEMS Microbiol. Lett. , 2007. **268**(1): p. 112-119.
253. Black, D.S., B. Irwin, and H.S. Moyed, *Autoregulation of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis.* J. Bacteriol., 1994. **176**(13): p. 4081-4091.
254. Jorgensen, M.G., et al., *HicA of Escherichia coli defines a novel family of translation-independent mRNA interferases in bacteria and archaea.* J Bacteriol, 2009. **191**(4): p. 1191-9.
255. Gonzalez Barrios, A.F., et al., *Autoinducer 2 controls biofilm formation in Escherichia coli through a novel motility quorum-sensing regulator (MqsR, B3022).* J. Bacteriol., 2006. **188**(1): p. 305-316.
256. Ren, D., et al., *Gene expression in Escherichia coli biofilms.* Appl. Microbiol. Biotechnol. , 2004. **64**(4): p. 515-524.
257. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol. Syst. Biol., 2006. **2**: p. 2006.0008.
258. Christensen-Dalsgaard, M. and K. Gerdes, *Two higBA loci in the Vibrio cholerae superintegron encode mRNA cleaving enzymes and can stabilize plasmids.* Mol. Microbiol., 2006. **62**(2): p. 397-411.
259. Canada, K.A., et al., *Directed evolution of toluene ortho-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation.* J. Bacteriol., 2002. **184**(2): p. 344-349.

260. Donegan, K., et al., *Evaluation of Methods for Sampling, Recovery, and Enumeration of Bacteria Applied to the Phylloplane.* Appl. Environ. Microbiol., 1991. **57**(1): p. 51-56.
261. Peti, W. and R. Page, *Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost.* Protein Expr. Purif. , 2007. **51**(1): p. 1-10.
262. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
263. Mccoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**: p. 658-674.
264. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr. D Biol. Crystallogr., 2004. **60**(Pt 12 Pt 1): p. 2126-2132.
265. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr. D Biol. Crystallogr. , 1997. **53**(Pt 3): p. 240-255.
266. Vaguine, A.A., J. Richelle, and S.J. Wodak, *SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model.* Acta Crystallogr. D Biol. Crystallogr. , 1999. **55**(Pt 1): p. 191-205.
267. Yamaguchi, Y., J.H. Park, and M. Inouye, *MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in Escherichia coli.* J Biol Chem, 2009.
268. Budde, P.P., et al., *Characterization of a higBA toxin-antitoxin locus in Vibrio cholerae.* J. Bacteriol., 2007. **189**(2): p. 491-500.
269. Zhang, X.S., R. Garcia-Contreras, and T.K. Wood, *Escherichia coli transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA).* ISME J. , 2008. **2**(6): p. 615-631.
270. Alberts, I.L., K. Nadassy, and S.J. Wodak, *Analysis of zinc binding sites in protein crystal structures.* Protein Sci. , 1998. **7**(8): p. 1700-1716.
271. Karlin, S. and Z.Y. Zhu, *Classification of mononuclear zinc metal sites in protein structures.* Proc. Natl. Acad. Sci. U. S. A., 1997. **94**(26): p. 14231-14236.
272. Holm, L., et al., *Searching protein structure databases with DaliLite v.3.* Bioinformatics, 2008. **24**(23): p. 2780-2781.
273. Luscombe, N.M., et al., *An overview of the structures of protein-DNA complexes.* Genome Biol. , 2000. **1**(1): p. REVIEWS001.
274. Murzin, A.G., et al., *SCOP: a structural classification of proteins database for the investigation of sequences and structures.* J. Mol. Biol. , 1995. **247**(4): p. 536-540.
275. Yakovlev, G.I., et al., *Contribution of active site residues to the activity and thermal stability of ribonuclease Sa.* Protein Sci. , 2003. **12**(10): p. 2367-2373.

276. Kumar, P., et al., *Crystal structure of Mycobacterium tuberculosis YefM antitoxin reveals that it is not an intrinsically unstructured protein.* J. Mol. Biol., 2008. **383**(3): p. 482-493.
277. Makarova, K.S., N.V. Grishin, and E.V. Koonin, *The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria.* Bioinformatics, 2006. **22**(21): p. 2581-4.
278. Brown, B.L., M. Hadley, and R. Page, *Heterologous high-level E. coli expression, purification and biophysical characterization of the spine-associated RapGAP (SPAR) PDZ domain.* Protein Expr Purif, 2008. **62**(1): p. 9-14.
279. Dancheck, B., A.C. Nairn, and W. Peti, *Detailed structural characterization of unbound protein phosphatase 1 inhibitors.* Biochemistry, 2008. **47**(47): p. 12346-56.
280. Mustelin, T., L. Tautz, and R. Page, *Structure of the hematopoietic tyrosine phosphatase (HePTP) catalytic domain: structure of a KIM phosphatase with phosphate bound at the active site.* J. Mol. Biol. , 2005. **354**(1): p. 150-163.
281. Pape, T. and T.R. Schneider, *HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs.* Journal of Applied Crystallography, 2004. **37**: p. 843-844.
282. Lamzin, V.S. and K.S. Wilson, *Automated Refinement of Protein Models.* Acta Crystallographica Section D-Biological Crystallography, 1993. **49**: p. 129-147.
283. Perrakis, A., et al., *wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models.* Acta Crystallographica Section D-Biological Crystallography, 1997. **53**: p. 448-455.
284. Painter, J. and E.A. Merritt, *Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.* Acta Crystallogr. D Biol. Crystallogr. , 2006. **62**(Pt 4): p. 439-450.
285. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J. Mol. Biol., 1971. **55**(3): p. 379-400.
286. Lawrence, M.C. and P.M. Colman, *Shape complementarity at protein/protein interfaces.* J. Mol. Biol., 1993. **234**(4): p. 946-950.
287. Terwilliger, T.C. and J. Berendzen, *Automated MAD and MIR structure solution.* Acta Crystallogr. D Biol. Crystallogr. , 1999. **55**(Pt 4): p. 849-861.
288. Blattner, F.R., et al., *The complete genome sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-1474.
289. Lopez, D., H. Vlamakis, and R. Kolter, *Biofilms.* Cold Spring Harb Perspect Biol, 2010.
290. Cvitkovitch, D.G., Y.H. Li, and R.P. Ellen, *Quorum sensing and biofilm formation in Streptococcal infections.* J Clin Invest, 2003. **112**(11): p. 1626-32.
291. Costerton, J.W., P.S. Stewart, and E.P. Greenberg, *Bacterial biofilms: a common cause of persistent infections.* Science, 1999. **284**(5418): p. 1318-22.

292. Keren, I., et al., *Persister cells and tolerance to antimicrobials.* FEMS Microbiol Lett, 2004. **230**(1): p. 13-8.
293. Kim, Y. and T.K. Wood, *Toxins Hha and CspD and small RNA regulator Hfq are involved in persister cell formation through MqsR in Escherichia coli.* Biochem Biophys Res Commun, 2010. **391**(1): p. 209-13.
294. Dorr, T., M. Vulic, and K. Lewis, *Ciprofloxacin causes persister formation by inducing the TisB toxin in Escherichia coli.* PLoS Biol, 2010. **8**(2): p. e1000317.
295. Falla, T.J. and I. Chopra, *Joint tolerance to beta-lactam and fluoroquinolone antibiotics in Escherichia coli results from overexpression of hipA.* Antimicrob Agents Chemother, 1998. **42**(12): p. 3282-4.
296. Correia, F.F., et al., *Kinase activity of overexpressed HipA is required for growth arrest and multidrug tolerance in Escherichia coli.* J Bacteriol, 2006. **188**(24): p. 8360-7.
297. Blower, T.R., G.P. Salmond, and B.F. Luisi, *Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners.* Curr Opin Struct Biol, 2011. **21**(1): p. 109-18.
298. Masuda, H., et al., *YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli.* Mol Microbiol, 2012. **84**(5): p. 979-89.
299. Wang, X., et al., *A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS.* Nat Chem Biol, 2012.
300. Fozo, E.M., M.R. Hemm, and G. Storz, *Small toxic proteins and the antisense RNAs that repress them.* Microbiol Mol Biol Rev, 2008. **72**(4): p. 579-89.
301. Blower, T.R., et al., *Mutagenesis and functional characterization of the RNA and protein components of the toxIN abortive infection and toxin-antitoxin locus of Erwinia.* J Bacteriol, 2009. **191**(19): p. 6029-39.
302. Wang, X., et al., *Antitoxin MqsA helps mediate the bacterial general stress response.* Nat Chem Biol, 2011. **7**(6): p. 359-66.
303. Brown, B.L., et al., *Three dimensional structure of the MqsR:MqsA complex: a novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties.* PLoS Pathog, 2009. **5**(12): p. e1000706.
304. Christensen-Dalsgaard, M., M.G. Jorgensen, and K. Gerdes, *Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses.* Mol Microbiol, 2010. **75**(2): p. 333-48.
305. Kim, Y., et al., *Escherichia coli toxin/antitoxin pair MqsR/MqsA regulate toxin CspD.* Environ Microbiol, 2010. **12**: p. 1105-1121.
306. Brown, B.L., et al., *Structure of the Escherichia coli antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation.* J Biol Chem, 2011. **286**(3): p. 2285-96.
307. Loris, R., et al., *Crystal structure of the intrinsically flexible addiction antidote MazE.* J Biol Chem, 2003. **278**(30): p. 28252-7.
308. Madl, T., et al., *Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA.* J Mol Biol, 2006. **364**(2): p. 170-85.

309. Oberer, M., et al., *The solution structure of ParD, the antidote of the ParDE toxin antitoxin module, provides the structural basis for DNA and toxin binding.* Protein Sci, 2007. **16**(8): p. 1676-88.
310. Cherny, I. and E. Gazit, *The YefM antitoxin defines a family of natively unfolded proteins: implications as a novel antibacterial target.* J Biol Chem, 2004. **279**(9): p. 8252-61.
311. Li, G.Y., et al., *Structural mechanism of transcriptional autorepression of the Escherichia coli RelB/RelE antitoxin/toxin module.* J. Mol. Biol., 2008. **380**(1): p. 107-119.
312. Mattison, K., et al., *Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs.* J Biol Chem, 2006. **281**(49): p. 37942-51.
313. Marianovsky, I., et al., *The regulation of the Escherichia coli mazEF promoter involves an unusual alternating palindrome.* J Biol Chem, 2001. **276**(8): p. 5975-84.
314. Tam, J.E. and B.C. Kline, *The F plasmid ccd autorepressor is a complex of CcdA and CcdB proteins.* Mol Gen Genet, 1989. **219**(1-2): p. 26-32.
315. Wilbur, J.S., et al., *Neisseria gonorrhoeae FitA interacts with FitB to bind DNA through its ribbon-helix-helix motif.* Biochemistry, 2005. **44**(37): p. 12515-24.
316. Zhang, Y., et al., *MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli.* Mol Cell, 2003. **12**(4): p. 913-23.
317. Dao-Thi, M.H., et al., *Molecular basis of gyrase poisoning by the addiction toxin CcdB.* J Mol Biol, 2005. **348**(5): p. 1091-102.
318. Overgaard, M., J. Borch, and K. Gerdes, *RelB and RelE of Escherichia coli form a tight complex that represses transcription via the ribbon-helix-helix motif in RelB.* J Mol Biol, 2009. **394**(2): p. 183-96.
319. Khoo, S.K., et al., *Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen Streptococcus pneumoniae.* J Biol Chem, 2007. **282**(27): p. 19606-18.
320. Brown, B.L. and R. Page, *Preliminary crystallographic analysis of the Escherichia coli antitoxin MqsA (YgiT/b3021) in complex with mqsRA promoter DNA.* Acta Crystallogr F 2010. **66**(Pt 9): p. 1060-3.
321. Afif, H., et al., *The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system.* Mol Microbiol, 2001. **41**(1): p. 73-82.
322. Garcia-Pino, A., et al., *Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity.* Cell, 2010. **142**(1): p. 101-11.
323. Overgaard, M., et al., *Messenger RNA interferase RelE controls relBE transcription by conditional cooperativity.* Mol Microbiol, 2008. **69**(4): p. 841-57.
324. Winther, K.S. and K. Gerdes, *Regulation of Enteric vapBC Transcription: Induction by VapC Toxin Dimer-Breaking.* Nucleic Acids Res, 2012.