1-1-2017

Structural Basis Of Membrane Protein Scaffolding And Signaling In Human Disease

Joshua Holcomb
Wayne State University,

STRUCTURAL BASIS OF MEMBRANE PROTEIN SCAFFOLDING AND SIGNALING IN HUMAN DISEASE

by

JOSHUA HOLCOMB

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Approved By:

_____

Advisor Date

_____

_____

_____

ACKNOWLEDGEMENTS

The dissertation encompasses the influence of many individuals in which I would like to express my deepest gratitude for their time and insurmountable support over the past few years. I would first like to express my utmost appreciation to my advisor, Dr. Zhe Yang. It was by his motivation and passion for science that has inspired me to be able to achieve what I have thus far. Dr. Yang possesses a remarkable level knowledge when it comes to science and always provides his students with support and guidance while treating them with insuperable respect. He is an outstanding role model and has been remarkable mentor to me during my time here at Wayne State University. I would also like to express gratefulness to my committee members: Dr. David Evans, Ladislau Kovari, and Fei Sun, along with my co-advisor Dr. Jian-Ping Jin. Each of them are incredible scientists and have consistently provided me with valuable advice and insight.

I would next like to specifically acknowledge the current Ph.D. candidate in the laboratory, Nicholas Spellmon. Nick has been extraordinarily helpful over the past years by assisting my research along with data collection and management of the laboratory. He has been a good friend and influential in the success I have been able to achieve thus far. I would also like to thank our current laboratory members including masters' students: Stephanie Hayden, Junmei Wan, Maysaa Doughan, and Yingxue Zhang for their help and support.

Finally, I would like to thank my family and friends including my wife and parents who have provided with an insurmountable amount of encouragement and motivation. I would also like to recognize grandfather Daniel Doyle which for personal reasons inspired me to pursue my Ph.D. in the basic medical sciences.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART I

Structural Basis for NHERF1-PDZ2 Dimer Mediated Protein Scaffolding and NHERF2-LPA2 Complex Assembly

CHAPTER 1 GENERAL INTRODUCTION

1.1 Protein Scaffolding

The term protein scaffolding can be defined directly as the coming together of two or more proteins in which facilitates a stable conformational state and is often required for specific biological function[6]. There are numerous examples of such proteins identified in the literature in which play critical roles in many physiological signaling events and are achieved by variety of molecular mechanisms. Some key cellular functions of protein scaffolding have been identified to include the sorting, recycling, localization and stabilization of cell surface receptors in addition to performing critical roles in cell polarity and cell to cell adhesion[6-10]. Scaffolding proteins often exhibit distinct specificity for their binding partners in which lead to particular cellular events. However, several instances have also been observed in which certain scaffolding proteins have been found to bind to a variety of targets exhibiting specific yet promiscuous binding activity[2,6]. Regardless, these interactions are often stable but dynamic furthering the complexity of the mechanistic basis for which these proteins function and regulate various cellular pathways.

Scaffolding Proteins and Their Functions

Scaffolding proteins critical for the formation macromolecular complexes required for cellular signaling can be divided into three general categories: arrestins, postsynaptic density 95/disc large/zona occludens-1 (PDZ) domain-containing proteins, and non-PDZ proteins[11]. Arrestins encompass a small family of proteins important for regulation at G protein coupled receptors (GPCRs) with only four members found in mammals. However, even with a lack in

recognizable sequence for binding its targets, this family of scaffolding proteins has been shown to bind to several hundred GPCRs in order to promote cellular signaling[11-14]. Arrestins function by binding to the cytoplasmic face of GPCRs, preventing its activation or linking the receptor to internalization machinery for degradation[15]. Arrestin structure is highly conserved whereas their mechanistic binding to their substrates possess some noticeable differences. In the case of arrestin-1, it structure is composed of N- and C- terminal β-sheet domains with a series of buried polar residues termed the "polar core" in which the N-terminal domain is stabilized by a C-terminal tail securing the molecule into an inactive state[16,17]. However, Arrestin-4 exhibits variable conformation in the surfaces of the β-sheets involved in receptor recognition and the loop between β-strands 1 and 2[16,18]. This comparison exemplifies aforementioned statement that scaffolding proteins are able to achieve distinct specificity for their substrates in order to promote explicit signaling events critical to a variety of cellular functions.

In addition to the arrestin family, a number of other proteins exist in which contribute to macromolecular complexes necessary for specific signaling events. Examples of such would include non-PDZ proteins such as A-kinase anchor proteins (AKAPs), Janus kinase (Jak)-2, and 14-3-3 proteins[11]. Both AKAP79 and AKAP250 have been shown to interact with the β-2 adrenergic (β_2AR) receptor in which have been shown to promote mitogenic signaling or desensitization of β_2AR respectively whereas interaction of Jak-2 with AT_1R has been shown mediate STAT1 complexing with AT_1R and promote STAT1 mediated gene transcription [19,20]. Additionally, scaffolding of the 14-3-3 proteins such as with the FSH receptor has been shown to result in a modest increase in cellular accumulatio of follitropin among other effects[11,21]. However, even with the number of events facilitated by non-PDZ domain proteins, the vast number of

implicative actions arbitrated by those containing PDZ domains raise particular interest when it comes to the mediation of biological functions.

PDZ Proteins

1.2.1 Overview

PDZ proteins are a family of proteins involved in the regulation of protein scaffolding and cellular signaling in which contain one or more PDZ domains. PDZ is an acronym for the three proteins in which were originally identified to possess such a domain which includes post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1)[22] which have also referred to as DHR (Discs large homology repeat) domains after the highly conserved four-residue GLGF sequence found in these domains[23]. These domains have two primary functions as protein-protein interaction modules including the localization of cellular signaling events and regulation of different physiological pathways[23]. Interaction with their protein targets are mediated by recognition of a C-terminal consensus peptide motif most often on the cytoplasmic tail of transmembrane receptors and channels[23,24]. PDZ containing proteins are highly abundant in a variety of species with over 900 domains identified in more than 300 proteins in the mouse genome[25]. And due to their prevalence and diversity, it is not difficult to understand how this family of proteins could have a significant impact on a wide range of biological functions.

1.2.2 PDZ Function

As previously stated, PDZ domains are responsible for protein-protein scaffolding which lead diverse signaling cascades regulating cellular and physiological function. Due to the vast array of proteins containing such domains, there has been continual investigation into a number of these proteins and how they function in cellular regulation. Some of the most common and well-studied

PDZ proteins include PSD-95, GRIP, HOMER, and the NHERF family. PSD-95 is a synaptic protein containing three PDZ domains found to bind and scaffold a variety of membrane proteins via its two N-terminal PDZ domains[26]. These two domains were first discovered to bind to the C-termini of Shaker-type K^+ subunits and NDMA receptor NR2 subunits resulting in the scaffolding of these membrane proteins[26,27]. Glutamate receptor interacting protein (GRIP) is another post synaptic protein containing seven PDZ domains which have been shown to be responsible in binding AMPA receptor in a similar fashion as PSD-95 to NDMA[28]. This interaction has shown to cluster AMPA receptors to other proteins at excitatory synapses in brain which is thought to be critical in neuroplasticity, learning, and memory storage[28,29]. HOMER is also a neuronal protein encoded by the *HOMER1* gene but with different function to that of either PSD-95 or GRIP per its involved in metabotropic glutamate signaling[30,31]. HOMER contains only one PDZ domain in which it binds type 5 metabotropic glutamate receptor (mGluR5) and is expressed at high levels during embryonic stages in which suggests it may play an important role in neuronal development[32]. Lastly, the PDZ containing NHERF (Na^+/H^+ exchanger regulatory factor) family have been shown to be involved in wide variety of scaffolding interactions, the implication and mechanistic basis of which will be discussed successively.

1.2.3 The NHERF Protein Family

The NHERF protein family consists of four members which have been found to bind to a variety of membrane transporters and receptors and have been shown to modulate protein expression, mobility, protein-protein interactions and the formation of signaling complexes[33]. Each of the members are differentially expressed in mammalian tissues with NHERF-1 and NHERF-2 being the most similar and well-studied. Both NHERF-1 and NHERF-2 contain two PDZ domains and have been found to be essential regulators of hormonal signaling by inhibition

of NHE3 (Na^+/H^+ exchanger isoform 3) activity in the renal and gastrointestinal epithelial cells[34-36]. NHERF-1 and NHERF-2 have also been shown to associate with the C-terminus of β_2AR, ezrin, PKA, and CFTR, allowing a variety of agonist signaling with mechanistic specificity[33,37]. Specifically, absence or reduction of NHERF-1 has been shown to display defects in intestinal salt absorption and defective intestinal anion secretion[33,38,39]. Furthermore, NHERF-2 knockout was determined to increase CFTR dependent anion secretion but decrease intestinal salt absorption however, fluid absorption in in the jejunum of NHERF-2 knockout mice *in vivo* was shown to increase when compared to the wild-type[33,40,41]. Both PDZ domains of NHERF-1 have been shown to bind to the chemokine G protein coupled receptor CXCR2 in addition to the enzyme PLCβ2 creating a signal transduction scaffold in which chemokine stimulation of CXCR2 results PLCβ2 ligand induced calcium signaling[2,42,43]. This chemokine stimulated PLCβ2 signaling has been shown to promote neutrophil migration from the bone marrow to sites of inflammation and infiltration into the surrounding tissue exacerbating the inflammatory response in a variety of diseases including rheumatoid arthritis, psoriasis, emphysema, and others [2,42,43]. The PDZ domains of NHERF-2 have been determined to facilitate scaffolding of the cystic fibrosis transmembrane conductance regulator (CFTR) with Lysophosphatidic acid receptor 2 (LPA2). This has shown to play an inhibitory role on CFTR activity while disruption of this complex has been shown to augment CFTR Cl^- transport[1,44,45]. Therefore, due to the influential roles in which the NHERF family PDZ domains exhibit in cellular signaling and protein function, elucidation of the structural basis in which these interactions occur may provide for greater understanding of their involvement in a variety of human disease.

CHAPTER 2 CRYSTAL STRUCTURE OF THE NHERF1 PDZ2 DOMAIN IN COMPLEX WITH THE CHEMOKINE RECEPTOR CXCR2 REVEALS PROBABLE MODES OF PDZ2 DIMERIZATION

*Published in Biochem Biophys Res Commun. 2014 May 30;448(2):169-74. doi: 10.1016/j.bbrc.2014.04.085. Epub 2014 Apr 24. All authors agreed with including their work in this dissertation.

Abstract

The formation of CXCR2–NHERF1–PLCb2 macromolecular complex in neutrophils regulates CXCR2 signaling and plays a key role in neutrophil chemotaxis and transepithelial neutrophilic migration. However, NHERF1 by itself, with only two PDZ domains, has a limited capacity in scaffolding the multiprotein-complex formation. Here we report the crystal structure of the NHERF1 PDZ2 domain in complex with the C-terminal CXCR2 sequence. The structure reveals that the PDZ2–CXCR2 binding specificity is achieved by numerous hydrogen bonds and hydrophobic contacts with the last four CXCR2 residues contributing to specific interactions. The structure also reveals two probable modes of PDZ2 dimerization where the two canonical ligand-binding pockets are well separated and orientated in a unique parallel fashion. This study provides not only the structural basis for the PDZ-mediated NHERF1–CXCR2 interaction, but also an additional example of how PDZ domains may dimerize, which both could prove valuable in understanding NHERF1 complex-scaffolding function in neutrophils.

Introduction

The chemokine receptor CXCR2 is a G-protein coupled receptor that plays important roles in regulating neutrophil chemotaxis and directing neutrophils to sites of inflammation[46,47]. Defective regulation of CXCR2 can cause excessive release of neutrophils from bone marrow and has been implicated in a variety of inflammatory diseases, including rheumatoid arthritis, psoriasis,

inflammatory bowel disease, acute respiratory distress syndrome, septic shock, pulmonary emphysema, and chronic obstructive pulmonary disease[48]. Growing evidence suggests that CXCR2 interacts directly or indirectly with other receptors, ion channels, transporters, scaffolding proteins, effectors, and cytoskeletal elements to form macromolecular complexes at specialized subcellular domains[42,49]. These dynamic protein–protein interactions regulate CXCR2 signaling function as well as its localization and processing within cells[50,51]. We have shown that CXCR2, phospholipase C-b2 (PLCβ2), and Na+/H+ exchanger regulatory factor-1 (NHERF1) form macromolecular complexes at the plasma membrane of bone marrow neutrophil, which functionally couple chemokine signaling to PLCβ2-mediated signaling cascade[42]. We also showed that disruption of this NHERF1-bridged interaction abolishes CXCR2 signaling and inhibits neutrophil chemotaxis and transepithelial neutrophilic migration in vitro[42]. These findings imply that targeting the NHERF1-scaffolded CXCR2–PLCβ2 signaling cascade could provide new strategies for therapeutic interventions of CXCR2-related diseases.

The ability of NHERF1 to scaffold the formation of a multiprotein complex depends on its two PDZ (PSD-95/Discs-large/ZO-1) domains, PDZ1 and PDZ2. In general, PDZ domains mediate protein interactions by recognizing the C-terminal sequence of target proteins and binding to the targets through a canonically and structurally conserved PDZ peptide-binding pocket[23]. The specificity of the interactions is determined mainly by the residues at positions 0 and _2 of the peptides (position 0 referring to the C-terminal residue), whereas other residues do not significantly contribute to the interaction[23]. This has led to the classification of PDZ domains into two major specificity classes: type I, (S/T)X(V/I/L) (X denoting any amino acid); type II, (F/Y)X(F/V/A)[3,52]. The PDZ domains of NHERF1 are the type I PDZ domains that have been shown to be capable of interacting with many different proteins and robust in scaffolding

multiprotein complex formation. In addition to CXCR2 and PLCβ2, the NHERF1 PDZ domains are able to interact with a range of other membrane and signaling proteins, such as cystic fibrosis transmembrane conductance regulator (CFTR), β2-adrenergic receptor (β2AR), platelet-derived growth factor receptor (PDGFR), and parathyroid hormone receptor (PTHR)[53]. Through these protein interactions, NHERF1 plays central roles in signaling complex assembly and receptor recycling as well as in establishing cell polarity and directing protein trafficking[23]. However, NHERF1 by itself, with only two PDZ domains, has a limited capacity to form multiprotein arrays or scaffold interactive proteins within membrane microdomains[54].

Notably, some PDZ domains, including the NHERF1 PDZ domains, can increase their scaffolding capacity through PDZ oligomerization with the same or different PDZ-containing proteins. Oligomerization of InaD through the third and fourth PDZ domains of separate InaD proteins has been shown to amplify the capacity and complexity of InaD-sequestered proteins in the Drosophila phototransduction networks[54]. Dimerization of ZO PDZ2 domains has been proposed to play a pivotal role in initiating claudin polymerization and directing tight junction strands for correct localization[55]. Analysis of over 150 PDZ domains in mouse genome revealed that 30% of mouse PDZ domains participate in PDZ–PDZ interactions, suggesting that many PDZ domains have evolved as a dual binding module in facilitating the formation of multiprotein complexes[56]. For NHERF1, both PDZ1 and PDZ2 can dimerize with homologous PDZ–PDZ interactions being more prominent than heterologous interactions[54]. The homodimerization of NHERF1 PDZs has been suggested to provide a mechanism for NHERF1 to expand its capacity in multiprotein complex assembly and regulating the activity of interacting proteins within membrane microdomains[54]. It should be noted that the PDZ1 and PDZ2 domains of NHERF1 exhibit distinct binding specificity: for example, both CXCR2 and PLCβ2 prefer to bind to

PDZ2[42,57]. It is conceivable that the assembly of the CXCR2– NHERF1–PLCβ2 signaling complex in neutrophils may not only require the PDZ canonical ligand-binding ability, but also the ability of NHERF1 to engage in separate PDZ–PDZ interactions. Consequently, such non-canonical PDZ-binding mode could have important implications in CXCR2–PLCβ2 coupling and in CXCR2-mediated neutrophil chemotaxis. In this context, we here present the crystal structure of NHERF1 PDZ2 in complex with the CXCR2 C-terminal peptide TSTTL. The structure reveals the PDZ2–CXCR2 interaction specificity and two probable modes of PDZ2 dimerization. The structure also suggests a model that the PDZ2 dimerization utilizes a distinct interface that functions together with two well-separated canonical ligand-binding pockets in scaffolding multiprotein complex formation.

Materials and Methods

Protein Expression and Purification

A DNA fragment encoding the human NHERF1 PDZ2 (residues 155–234) was amplified by PCR using the full-length human NHERF1 cDNA as a template. The C-terminal extension TSTTL that corresponds to residues 356–360 of human CXCR2 was created by inclusion of 15 extra bases in the reverse primer. The PCR products were cloned in the pSUMO vector containing an N-terminal His6-SUMO tag. The resulting clone was transformed into Escherichia coli BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD600 (optical density at 600 nm) of 0.4 at 37 _C in LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside at 15 °C overnight. The cells were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by the cleavage of the His6-SUMO tag with yeast SUMO Protease 1. PDZ2 proteins were separated from the cleaved tag by a second Ni^{2+} affinity chromatography and further

purified by size-exclusion chromatography. Finally, the proteins were concentrated to 10–20 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM b-mercaptoethanol, and 5% glycerol.

Crystallization, Data Collection and Structure Determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the protein (20 mg/ml) with an equal volume of a reservoir solution containing 100 mM Tris–HCl, pH 8.5, 8% PEG8000 at 20 °C. Crystals typically appeared overnight and continued to grow to their full size in 2–3 days. Prior to X-ray diffraction data collection, crystals were cryoprotected in a solution containing the mother liquor and 25% glycerol and flash cooled in liquid nitrogen. The data were collected at 100 K at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and processed and scaled using the program XDS[58]. Crystals belong to the space group P21 with two molecules in the asymmetric unit (Table 1). The structure was solved by the molecular replacement method with the program PHASER[59] using the PDZ2–TETSL structure (PDB code: 2OZF) as a search model. Structure modeling was carried out in COOT[60], and refinement was performed with PHENIX[61]. To reduce the effects of model bias, iterative-build OMIT maps were used during model building and structure refinement. The final models were analyzed and validated with Molprobity[62]. All figures of 3D representations of the PDZ2–CXCR2 structure were made with PyMOL (www.pymol.org).

Protein Data Bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4Q3H.

Results and Discussion

Structural Basis of the PDZ2-CXCR2 Interaction

Table 1. Crystallographic data and refinement statistics

Data	
Space group	$P2_1$
Cell parameters	
a, b, c (Å)	32.8, 55.3, 54.2
β (°)	90.7
Wavelength (Å)	0.97856
Resolution (Å)	32.8-1.44 (1.49-1.44)
R_{merge}[a]	0.042 (0.369)[b]
Redundancy	3.5 (2.7)
Unique reflections	28,239
Completeness (%)	96.8 (78.2)
$\langle I/\sigma \rangle$	17.8 (2.3)
Refinement	
Resolution (Å)	32.8-1.44 (1.49-1.44)
Molecules/AU	2
R_{work}[c]	0.166 (0.254)
R_{free}[d]	0.197 (0.304)
Ramachandran plot	
Residues in favored	98.3%
Residues in allowed	1.7%
RMSD	
Bond lengths (Å)	0.007
Bond angels (°)	1.12
No. of atoms	
Protein	1299
Peptide	72
Water	249
B-factor (Å^2)	
Protein	24.1
Peptide	20.2
Water	33.4

[a]Numbers in parentheses refer to the highest resolution shell.

[b]$R_{merge} = \Sigma|I-\langle I\rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.

[c]$R_{work} = \Sigma|F_o-F_c| / \Sigma|F_o|$, where F_o is the observed structure factor, F_c is the calculated structure factor.

[d]R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

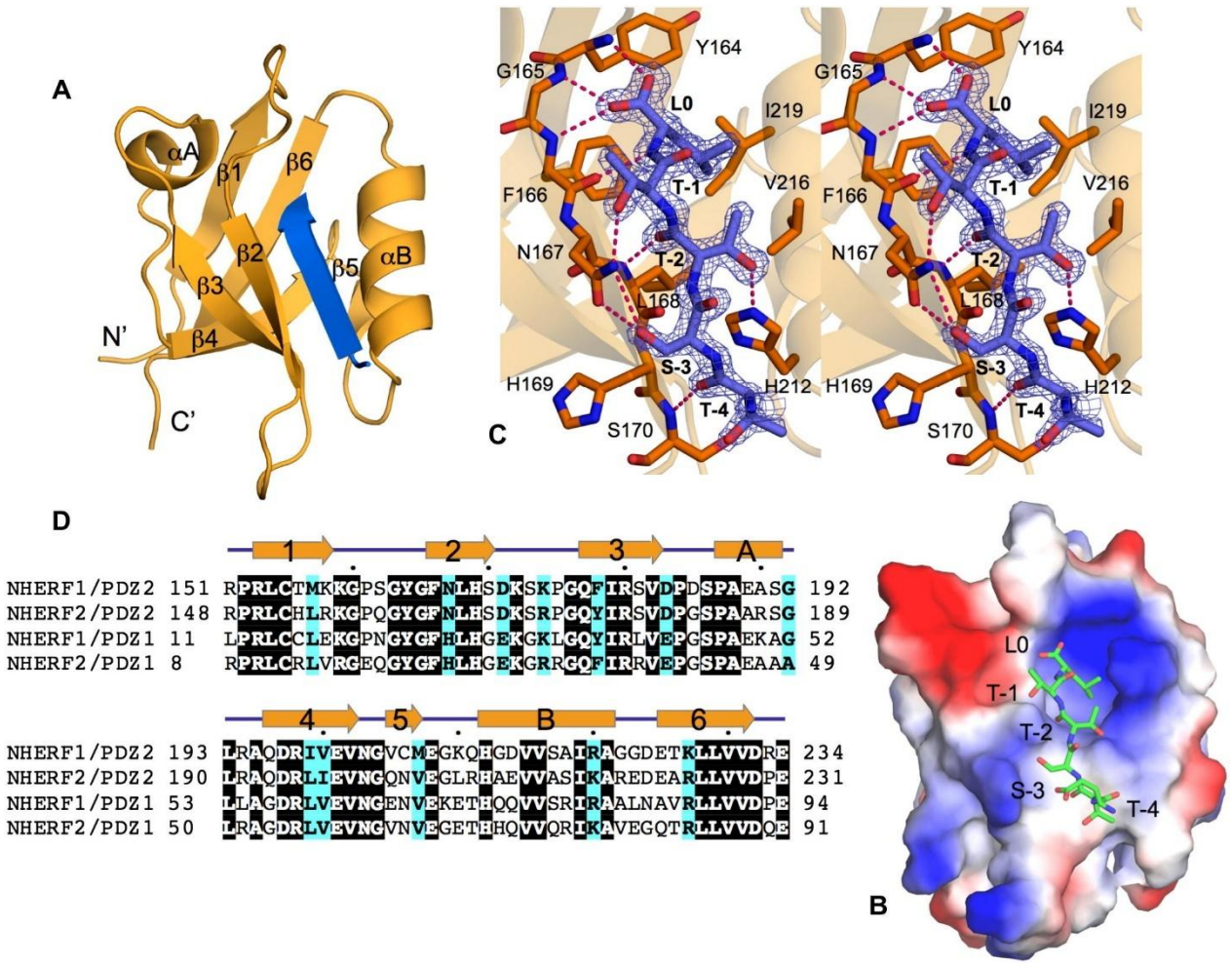

Figure 1. Structure of NHERF1 PDZ2 in complex with the CXCR2 C-terminal sequence TSTTL. (A) Ribbon diagram of the PDZ2–CXCR2 structure. PDZ2 is shown in orange and the CXCR2 peptide is shown in blue. Secondary structures of PDZ2, a-helices, and b-strands are labeled and numbered according to their position in the sequence. (B) Surface representation of the PDZ2–CXCR2 structure. Surface coloring is according to the electrostatic potential: red, white, and blue correspond to negative, neutral, and positive potential, respectively. The vacuum electrostatics/protein contact potential was generated by PyMOL. The CXCR2 peptide is depicted by sticks. (C) Stereo view of the PDZ2 ligand-binding site bound to the CXCR2 C-terminal peptide. PDZ2 residues are represented by sticks with their carbon atoms colored in orange. The CXCR2 peptide is depicted by sticks overlaid with 2Fo _ Fc omit map calculated at 1.44 Å and contoured at 1.8 r. Hydrogen bonds are illustrated as red broken lines. (D) Sequence alignment of selected PDZ domains. The alignment was performed by ClustalW[5], including human NHERF1 and NHERF2. Identical residues are shown as white on black, and similar residues appear shaded in cyan. Secondary structure elements are displayed above the sequences and labeled according to the scheme in A. Sequence numbering is displayed to the left of the sequences, with every 10th residue marked by a dot shown above the alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The overall structure of NHERF1 PDZ2 is similar to other PDZ domains[3,63], consisting of six b strands (β1–β6) and two α-helices (αA and αB) (Fig. 1A and B). The CXCR2 peptide binds in the cleft between β2 and αB, burying a total solvent-accessible surface area of 649 Å2. The binding specificity of the PDZ2–CXCR2 interaction is achieved through networks of hydrogen bonds and hydrophobic interactions (Fig. 1C). At the ligand position 0, the side chain of Leu0 is nestled in a deep hydrophobic pocket formed by invariant residues Tyr164, Phe166, and Leu168 from β2 and Val216 and Ile219 from αB (Fig. 1D). In the pocket, the position of Leu0 is further secured by both a hydrogen bond from its amide nitrogen to the Phe166 carbonyl oxygen and triplet hydrogen bonding between the Leu0 carboxylate and the amides of Tyr164, Gly165, and Phe166. Similar interactions have been observed in several other PDZ-mediated complexes[3,63], which represent the most-conserved binding mode for terminal Leu recognition. Residues at other peptide positions also contribute to the PDZ2– CXCR2 complex formation (Fig. 1C). At position –1, the side chain hydroxyl of Thr-1 forms a hydrogen bond with the Nδ2 atom of the Asn167 side chain. At position –2, Thr-2 makes one hydrogen bond to the His212 imidazole group and two hydrogen bonds to the highly conserved residue Leu168. At the ligand position –3, the interactions with Ser-3 include one hydrogen bond from its side chain hydroxyl to the Oδ1 atom of Asn167, and another hydrogen bond to the Nδ1 atom of Asn167. The latter two interactions represent PDZ2 specific interactions, as Ser-3 recognition in PDZ1–CXCR2 is mediated by a His residue at His169 position[2]. Finally, the peptide residue Thr-4 engages in a main-chain contact with Ser170, but does not participate in any specific side-chain interactions. These observations indicate that the last four residues of CXCR2 contribute to the binding specificity in the PDZ2–CXCR2 complex formation.

Differential CXCR2 Interaction with PDZ1 and PDZ2

To gain further insight into PDZ2 binding specificity, we compared the PDZ2–CXCR2 structure to the structure of the NHERF1 PDZ1–CXCR2 complex. The overall structures of the two liganded PDZs are very similar, with a root mean square difference (RMSD) of 1.02 Å for 83 Ca atoms (Fig. 2A). The main chains of the bound peptides superimpose well (RMSD of 0.15 Å), as do their relative spatial positions to the conserved PDZ motifs. In addition, the ligand recognition modes at the peptide positions 0 and –2 are virtually indistinguishable, characterized by structurally similar binding sites composed with completely conserved residues (Figs. 2B and 1D). However, substantial differences are observed in the ligand recognition at positions –1 and –3. Note these differences are associated directly with the only residue difference in their peptide-binding pockets. In PDZ1, the residue His at position 27 is replaced by Asn at the matching position (167) in PDZ2. As a result of this difference, Thr-1 does not hydrogen bond to the His27 side chain in PDZ1, instead of forming a general, rather ligand indiscriminative Van der Waals interaction (Fig. 2B). At position –3, no direct contacts are observed between Ser-3 and His27 in DZ1–CXCR2. This contrasts to the PDZ2–CXCR2 complex where the side chain of Asn167 adopts a different rotamer that allows direct hydrogen bonding with the Ser-3 hydroxyl (Fig. 1C). Specific interaction of Ser-3 with Asn167 in PDZ2–CXCR2, but not with His27 in PDZ1–CXCR2, is consistent with previous affinity selection experiments in which PDZ2 almost exclusively selected ligands with Ser at position -3 from random peptides, whereas PDZ1 showed no apparent preference for this position[64].

Probable Modes of PDZ2 Dimerization

Both NHERF1 PDZ1 and PDZ2 can dimerize, and their dimerization has been suggested to facilitate the formation of multiprotein complexes and contributes to NHERF1-mediated intracellular signaling[54]. However, the molecular details of NHERF1 PDZ dimerization are

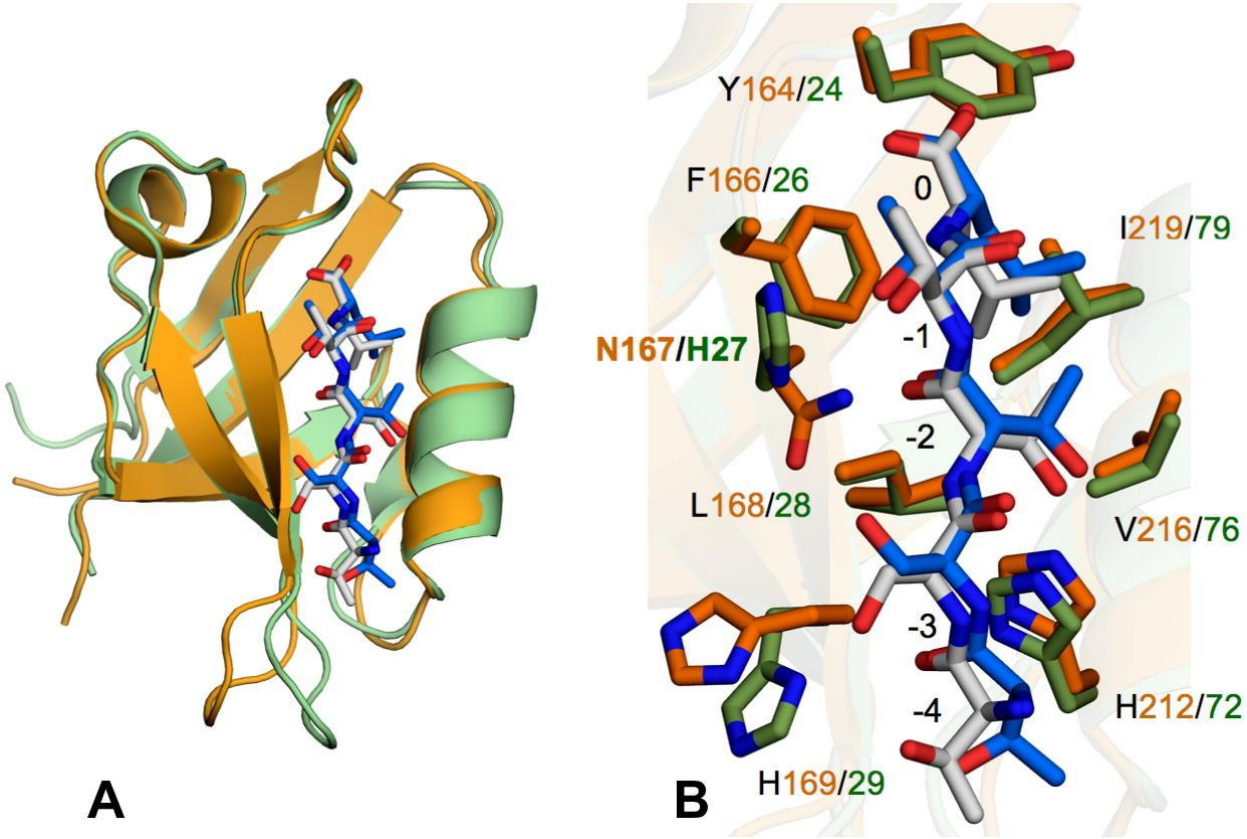

Figure 2. Structural comparison of NHERF1 PDZ2 and PDZ1. (A) Superposition of the structures of PDZ2–CXCR2 (orange–blue; PDB code: 4Q3H) and PDZ1–CXCR2 (green–gray; PDB code: 4JL7). PDZ domains are represented by ribbons. Residues in the ligands are displayed as sticks. (B) Superposition of the PDZ ligand-binding pockets. Both PDZ and ligand residues are depicted by sticks and colored according to the scheme in A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

unknown. Examination of the crystal packing reveals two probable modes of PDZ2 dimerization. As shown in Fig. 3A and B, both dimers are generated by a twofold symmetry. Dimer 1 is generated by the parallel stacking of helix αB from both monomers, while dimer 2 is generated by the interactions between two copies of strands β1, β4, and β6. In dimer 1, the buried surface area at the dimer interface is 497 Å2, which is comparable to 592 Å2 calculated for dimer 2. However, residues that form the dimer interface are only conserved in dimer 2 but not in dimer 1 (Figs. 3 and 1D). For dimer 1, the interactions at the interface include a salt-bridge between Lys210 and Asp214, two water-mediated hydrogen bonds between Asp214 and Ser217, and three stacking

interactions involving Asp214, Ser217, Arg220, and Ala221 (Fig. 3C). Because this interface is based on the twofold symmetry, the intermolecular interactions are contributed in the same way by both monomers in stabilizing the dimer structure. For dimer 2, the interactions at the interface include a salt bridge between Arg151 and Glu201 and the hydrophobic interactions between Pro152, Leu154, and Leu229 from both monomers (Fig. 3D). Although Arg153 is not directly involved in this interface, its side chain guanidinium makes three intramolecular hydrogen bonds to Asp197 and Gly192, which stabilize the conformation of the interface-forming b1 strand in dimer 2. It is of particular interest to note that NHERF1 R153Q mutation was recently found in patients who have impaired renal phosphate reabsorption[65]. Subsequent studies demonstrated that this disease mutation abolished NHERF1 dimerization but did not disrupt its interaction with the parathyroid hormone 1 receptor[66]. While future studies are required to demonstrate the biological significance of the crystallographically observed dimerization, the high degree of conservation and the close proximity of the disease-related mutation suggest that the interface observed in dimer 2 might have some physiological relevance. In this regard, only dimer-2 model is considered further.

In dimer 2, the canonical target-binding pockets are located at the distal sides of the dimer interface. They are arranged in a parallel fashion and related by the same twofold symmetry present at the dimer interface (Fig. 3B). In the dimer, the CXCR2 peptides are clearly resolved in both pockets, with the binding conformations highly similar to those observed in the monomeric PDZ–ligand complexes (Fig. 2A). This indicates that the dimerization does not affect the PDZ2–peptide binding in the crystal, or the binding sites in the dimer are open to allow the interaction with the target proteins. This observation is also consistent with the previously proposed model that PDZ dimerization utilizes a distinct interface that functions together with two well-separated canonical

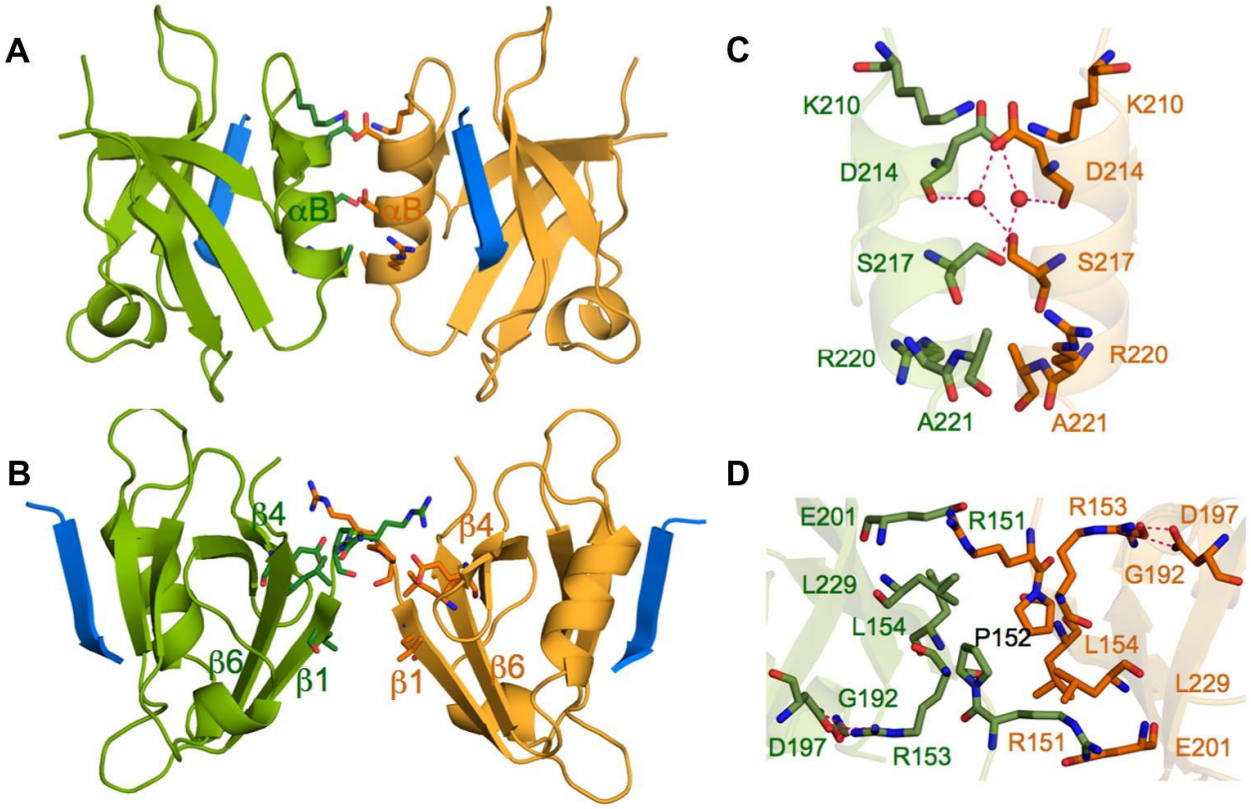

Figure 3. Two probable modes of PDZ2 dimerization. (A) Overall view of dimer 1 and (B) dimer 2. (C) Close-up view of the interactions in dimer 1 and (D) dimer 2. PDZ backbones are represented by ribbons with monomers colored green and orange. CXCR2 peptides are depicted by ribbons and shown in blue. Interacting residues at the dimer interface are depicted by sticks. Secondary structural elements are marked as in Fig. 1A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ligand-binding pockets in scaffolding multiprotein complex formation[67].

Structural Diversity in PDZ Dimerization

In addition to the NHERF1 PDZ domains, other PDZ domains capable of dimerization include GRIP (glutamate receptor interacting protein) PDZ6, SHANK (SH3 and multiple ankyrin repeat domains protein) PDZ, and ZO (zonula occludens protein) PDZ2[68-70]. Like NHERF1 oligomerization, the ability of these PDZ domains to dimerize has been suggested to expand PDZ scaffolding capacity and facilitate the formation of multiprotein complexes[67]. In line with this suggestion, recent family-wide proteomic investigation demonstrated that PDZ dimerization is surprisingly abundant, with ~30% mouse PDZ domains engaging in PDZ–PDZ interactions[56].

Interestingly, PDZ–PDZ interactions appear to be substantially more selective than interactions between PDZ domains and the C termini of their target proteins[56]. Modeling studies based on large-scale binding data predicted that one PDZ domain, on average, interacts with 245 different proteins via C-termini, contrasting only 1.7 PDZ domains via dimerization[56]. This suggests that the non-canonical binding mode contributes more to defining the precise composition of protein complexes than does the canonical binding mode[56].

Structurally, the high selectivity of PDZ–PDZ interactions is manifested by highly distinct dimerization modes and substantially different dimer interfaces. For example, in GRIP PDZ6, the dimer interface is formed primarily by the N-terminal β1 strand and parts of the β4–αA loop, whereas in ZO PDZ2 domain, the dimerization occurs via a domain-swapping mechanism, with β1 and β2 protruding out and exchanged between two symmetry related monomers (Fig. 4). In

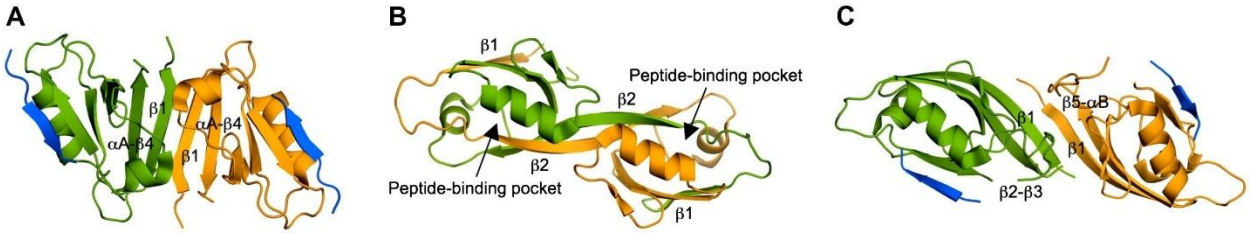

Figure 4. Structural diversity in PDZ dimerization. Ribbon representation of the (A) GRIP PDZ6 dimer (PDB code: 1N7F), (B) ZO PDZ2 dimer (PDB code: 2RCZ), and (C) SHANK PDZ dimer (PDB code: 1Q3P). PDZ monomers are shown in green and orange. Target peptides are shown in blue. Structural elements involved in PDZ dimerization are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contrast, the dimer interface in SHANK PDZ involves both β1 and two distinct loops of β2–β3 and β5–αB, which effectively define a new mode of PDZ dimerization. It is also interesting to note that the orientation of two canonical ligand-binding pockets in the NHERF1 PDZ2 dimer is unique among the known PDZ structures. In NHERF1 PDZ2, the ligands are orientated parallelly, whereas in other PDZ dimers the two ligands are positioned in an antiparallel fashion (Figs. 3B

and 4). Therefore, these observed differences demonstrate the structural diversity in PDZ dimerization, consistent with prior suggestion that diverse PDZ–PDZ interactions have been optimized as a mechanism in scaffolding the formation of distinct multiprotein complexes[56]. This also highlights the potential importance of the present study, as it provides an additional example of how PDZ domains may dimerize and could be valuable in understanding NHERF1 complex-scaffolding function in neutrophils and also in CXCR2-related diseases.

CHAPTER 3 STRUCTURAL INSIGHTS INTO PDZ-MEDIATED INTERACTION OF NHERF2 AND LPA2, A CELLULAR EVENT IMPLICATED IN CFTR CHANNEL REGULATION

*****Published in Biochem Biophys Res Commun. 2014 March 28; 446(1): 399–403. doi:10.1016/j.bbrc.2014.02.128. All authors agreed with including their work in this dissertation

Abstract

The formation of CFTR-NHERF2-LPA2 macromolecular complex in airway epithelia regulates CFTR channel function and plays an important role in compartmentalized cAMP signaling. We previously have shown that disruption of the PDZ-mediated NHERF2-LPA2 interaction abolishes the LPA inhibitory effect and augments CFTR Cl^- channel activity in vitro and in vivo. Here we report the first crystal structure of the NHERF2 PDZ1 domain in complex with the C-terminal LPA2 sequence. The structure reveals that the PDZ1-LPA2 binding specificity is achieved by numerous hydrogen bonds and hydrophobic contacts with the last four LPA2 residues contributing to specific interactions. Comparison of the PDZ1-LPA2 structure to the structure of PDZ1 in complex with a different peptide provides insights into the diverse nature of PDZ1 substrate recognition and suggests that the conformational flexibility in the ligand binding pocket is involved in determining the broad substrate specificity of PDZ1. In addition, the structure reveals a small surface pocket adjacent to the ligand-binding site, which may have therapeutic implications. This study provides an understanding of the structural basis for the PDZ-mediated NHERF2-LPA2 interaction that could prove valuable in selective drug design against CFTRrelated human diseases.

Introduction

Cystic fibrosis transmembrane conductance regulator (CFTR) is a cAMP-regulated chloride (Cl^-) channel primarily localized at the apical surfaces of epithelial cells lining the airway, gut and exocrine glands[71,72]. CFTR is responsible for transepithelial salt and water transport and plays critical roles in maintaining fluid homoeostasis, airway fluid clearance, and airway submucosal glands secretion in both healthy and disease phenotypes[73,74]. Growing evidence suggests that CFTR interacts directly or indirectly with other ion channels, transporters, scaffolding proteins, protein kinases, effectors, and cytoskeletal elements to form macromolecular complexes at specialized subcellular domains[44,75]. These dynamic protein-protein interactions regulate CFTR channel function as well as its localization and processing within cells[37,76]. We have shown that CFTR, lysophosphatidic acid receptor 2 (LPA2), and Na+/H+ exchanger regulatory factor-2 (NHERF2) form macromolecular complexes at the plasma membrane of gut epithelia, which functionally couple LPA2 signaling to CFTR-mediated Cl^- transport[41]. LPA2 is a G protein-coupled receptor that binds the lipid signaling molecule LPA and mediates diverse cellular responses such as cell proliferation and platelet aggregation[77]. NHERF2 is a PDZ domain-containing protein that typically functions as a scaffold to cluster transporters, receptors, and signaling molecules into supramolecular complexes[78]. We have demonstrated that LPA inhibits both CFTR-mediated Cl^- transport through the LPA2-mediated Gi pathway in a compartmentalized manner in cells and CFTR-dependent cholera toxin-induced mouse intestinal-fluid secretion in vivo[41]. We also demonstrated that disruption of the PDZ-mediated NHERF2-LPA2 interaction abolishes the LPA inhibitory effect and augments CFTR Cl^- channel activity in Calu-3 cells and also in fluid secretion from pig tracheal submucosal glands[79]. These findings imply that targeting the PDZ-mediated NHERF2-LPA2 interaction could provide new strategies for therapeutic interventions of CFTR-associated diseases[76,79].

In general, PDZ domains mediate protein interactions by recognizing the C-terminal sequence of target proteins and binding to the targets through a canonically and structurally conserved PDZ peptide-binding pocket[23]. Based on the residues at positions 0 and -2 of the peptides (position 0 referring to the C-terminal residue), early studies have grouped PDZ domains into two major specificity classes: class I, (S/T)X(V/I/L) (X denoting any amino acid); class II, (F/Y)X(F/V/A)[3,52,80]. However, more recent mounting evidence indicates that PDZ specificity is unexpectedly complex and diverse, with the PDZ domain family recognizing up to seven C-terminal ligand residues and forming at least 16 unique specificity classes[81]. The complexity of PDZ-peptide interactions is further exemplified by the facts that many PDZ domains can bind to multiple ligands of different peptide classes and that single peptides are capable of binding to distinct PDZ domains[81]. This complex picture of PDZ-peptide interactions raises a challenging problem regarding how PDZ domains, structurally simple protein-interaction modules, achieve binding promiscuity and specificity concomitantly, the nature of which remains obscure. In this context, we present the crystal structure of NHERF2 PDZ1 in complex with the LPA2 C-terminal peptide MDSTL. The structure reveals that the LPA2 peptide binds to PDZ1 in an extended conformation with the last four residues making specific side chain contacts. Comparison of the PDZ1-LPA2 structure to the structure of PDZ1 in complex with a different peptide suggests that the binding diversity of PDZ1 is facilitated by the conformational flexibility in the peptide-binding pocket. This study provides the structural basis of the PDZ-mediated NHERF2-LPA2 interaction and could be valuable in the development of novel therapeutic strategies against CFTR-related human diseases.

Materials and Methods

Protein Expression and Purification

A DNA fragment encoding the human NHERF2 PDZ1 (residues 9–90) was amplified by PCR using the full-length human NHERF2 cDNA as a template. The C-terminal extension MDSTL that corresponds to residues 347–351 of human LPA2 was created by inclusion of 15 extra bases in the reverse primer. The PCR products were cloned in the pSUMO vector containing an N-terminal His6-SUMO tag. The resulting clone was transformed into *Escherichia coli* BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD600 (optical density at 600 nm) of 0.4 at 37 °C in LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside at 15 °C overnight. The cells were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by the cleavage of the His6-SUMO tag with yeast SUMO Protease 1. PDZ1 proteins were separated from the cleaved tag by a second Ni^{2+} affinity chromatography and further purified by size-exclusion chromatography. Finally, the proteins were concentrated to 20–30 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol (BME), and 5% glycerol.

Crystallization, Data Collection and Structure Determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the protein (~20 mg/ml) with an equal volume of a reservoir solution containing 100 mM HEPES, pH 7.0, 0.2 M potassium thiocyanate (KSCN), 25% PEG3350 at 20 °C. Crystals typically appeared overnight and continued to grow to their full size in 2–3 days. Prior to X-ray diffraction data collection, crystals were cryoprotected in a solution containing the mother liquor and 25% glycerol and flash cooled in liquid nitrogen. The data were collected at 100 K at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and processed and scaled using the program XDS[58]. Crystals belong to the space group *P*21 with unit cell dimensions a = 26.4 Å, b = 40.3 Å, c = 37.1 Å, β = 107.4°,

and one molecule in the asymmetric unit (Table 2). The structure was solved by the molecular replacement method with the program PHASER[59] using the PDZ1-EDTSV structure (PDB code: 2OCS) as a search model. Structure modeling was carried out in COOT[60], and refinement was performed with PHENIX[61]. To reduce the effects of model bias, iterative-build OMIT maps were used during model building and structure refinement. The final models were analyzed and validated with Molprobity[62]. All figures of 3D representations of the PDZ1-LPA2 structure were made with PyMOL (www.pymol.org).

Protein Data Bank Accession Number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4POC.

Results and Discussion

Specificity Determinants of NHERF2-LPA2 Interaction

The overall structure of NHERF2 PDZ1 is similar to other PDZ domains[3,63], consisting of six β-strands (β1–β6) and two α-helices (αA and αB) (Fig. 5A and B). The LPA2 peptide binds in the cleft between β2 and αB, burying a total solvent-accessible surface area of 472 Å2. The binding specificity of the PDZ1-LPA2 interaction is achieved through networks of hydrogen bonds and hydrophobic interactions (Fig. 5C). At the ligand position 0, the side chain of Leu0 is nestled in a deep hydrophobic pocket formed by invariant residues Tyr21, Phe23, and Leu25 from β2 and Val73 and Ile76 from αB (Fig. 5D). In the pocket, the position of Leu0 is further secured by both a hydrogen bond from its amide nitrogen to the Phe23 carbonyl oxygen and triplet hydrogen bonding between the Leu0 carboxylate and the amides of Tyr21, Gly22, and Phe23. Similar interactions have been observed in several other PDZ-mediated complexes[3,63], which represent the most-conserved binding mode for terminal Leu recognition. Residues at other peptide positions

Table 2. Crystallographic data and refinement statistics

Data	
Space group	$P2_1$
Cell parameters	
a, b, c (Å)	26.4, 40.3, 37.1
Wavelength (Å)	1.2719
Resolution (Å)	24.2-1.34 (1.37-1.34)
R_{merge}[a]	0.039 (0.250)[b]
Redundancy	4.1 (4.0)
Unique reflections	17966
Completeness (%)	99.8 (99.6)
$\langle I/\sigma \rangle$	15.3 (3.0)
Refinement	
Resolution (Å)	24.2-1.34 (1.37-1.34)
Molecules/AU	1
R_{work}[c]	0.145 (0.268)
R_{free}[d]	0.177 (0.275)
Ramachandran plot	
Residues in favored	97.9%
Residues in allowed	2.1%
RMSD	
Bond lengths (Å)	0.007
Bond angels (°)	1.2
No. of atoms	
Protein	1347
Peptide	73
Water	143
Chloride	2
B-factor (Å^2)	
Protein	17.2
Peptide	17.7
Water	33.4
Chloride	20.2
SCN	12.4

[a]Numbers in parentheses refer to the highest resolution shell.

[b]$R_{merge} = \Sigma|I-\langle I\rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.

[c]$R_{work} = \Sigma|F_o-F_c| / \Sigma|F_o|$, where F_o is the observed structure factor, F_c is the calculated structure factor.

[d]R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

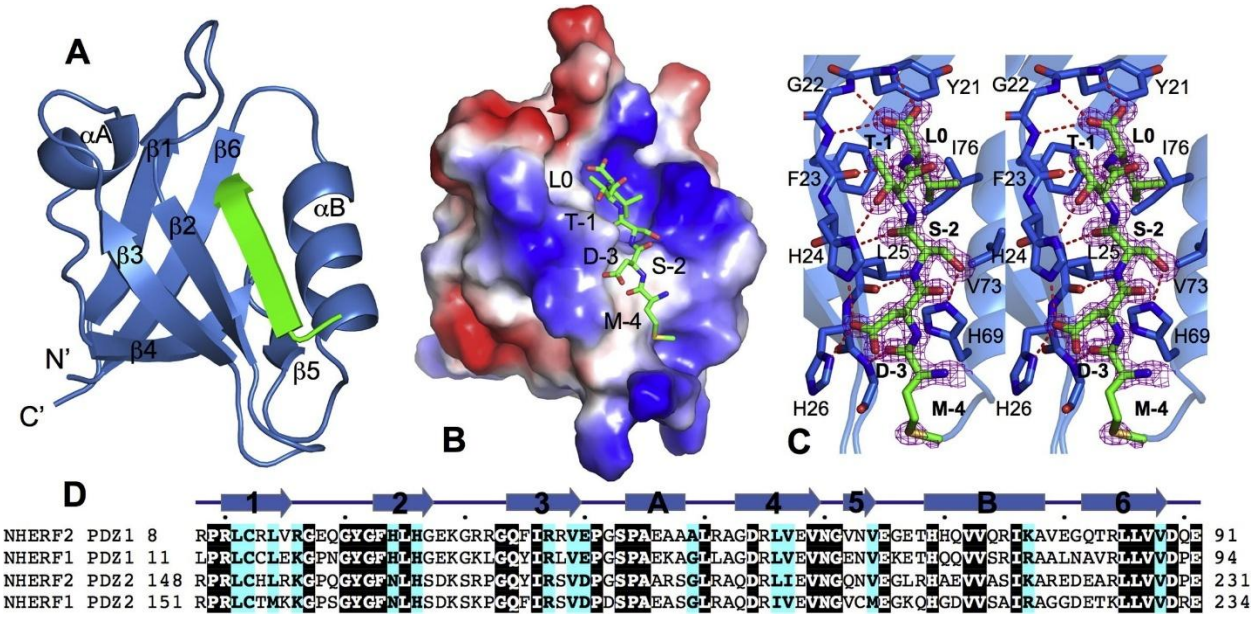

Figure 5. Structure of NHERF2 PDZ1 in complex with the LPA2 C-terminal sequence MDSTL. (A) Ribbon diagram of the PDZ1-LPA2 structure. PDZ1 is shown in light blue and the LPA2 peptide is shown in green. Secondary structures of PDZ1, α-helices, and β-strands are labeled and numbered according to their position in the sequence. (B) Surface representation of the PDZ1-LPA2 structure. Surface coloring is according to the electrostatic potential: red, white, and blue correspond to negative, neutral, and positive potential, respectively. The vacuum electrostatics/protein contact potential was generated by PyMOL. The LPA2 peptide is depicted by sticks. (C) Stereo view of the PDZ1 ligand-binding site bound to the LPA2 Cterminal peptide. PDZ1 residues are represented by sticks with their carbon atoms colored in light blue. The LPA2 peptide is depicted by sticks overlaid with 2Fo − Fc omit map calculated at 1.34 Å and contoured at 1.8 σ. Hydrogen bonds are illustrated as red broken lines. (D) Sequence alignment of selected PDZ domains. The alignment was performed by ClustalW[5], including human NHERF1 and NHERF2. Identical residues are shown as white on black, and similar residues appear shaded in cyan. Secondary structure elements are displayed above the sequences and labeled according to the scheme in Figure 1A. Sequence numbering is displayed to the left of the sequences, with every 10th residue marked by a dot shown above the alignment.

also contribute to the PDZ1-LPA2 complex formation (Fig. 5C). At position -1, the side chain hydroxyl of Thr-1 forms a hydrogen bond with the Nδ1 atom of the His24 imidazole ring. At position -2, Ser-2 makes one hydrogen bond to the His69 imidazole group and two hydrogen bonds to the highly conserved residue Leu25. At the ligand position -3, the interactions with Asp-3 include one hydrogen bond from its side chain carboxylate to the side chain of His24 and another hydrogen bond to the Nδ1 atom of His26. The latter interaction represents an unusual variation in

this structure, since the negatively charged Asp-3 is generally recognized by an Arg residue at Arg37 position in other PDZ complexes[4]. Finally, the peptide residue Met-4 engages in a main-chain contact with Gly27, but does not participate in any specific side-chain interactions. These observations indicate that the last four residues of LPA2 contribute to the binding specificity in the PDZ1-LPA2 complex formation.

Structural Basis for Broad PDZ1 Binding Specificity

To gain further insights into PDZ1 binding specificity, we compared the PDZ1-LPA2 structure to the structure of PDZ1 in complex with a different peptide (EDTSV) (Fig. 6). The overall structures of the two liganded PDZs are very similar, with a root mean square difference (RMSD) of 0.20 Å for 85 Cα atoms (Fig. 6A). The main chains of the bound peptides superimpose well (RMSD of 0.16 Å), as do their relative spatial positions to the conserved PDZ motifs. These observations indicate that binding of different peptides has little effect on the PDZ1 overall fold, consistent with previous evidence that the localized changes at a few key positions within HtrA1 PDZ are responsible for dramatically altered PDZ binding specificity[63]. Close examination of the structural alignment reveals some similarities, but also substantial differences, in the peptide- Val0 follow a similar path entering a pocket that is virtually identical between the two PDZ1 structures. At the ligand position -2, recognition of Ser-2 and Thr-2 is achieved by similar mechanisms where there is a direct hydrogen bond to the side chain of the structurally conserved His69 in both cases. However, large differences are observed around the residues at the -1 position of the ligands (Fig. 6C). In PDZ1-LPA2, His24 adopts a single conformation that simultaneously binds to Thr-1 and Asp-3. In PDZ1-EDTSV, His24 has a double conformation in which binding pockets (Fig. 6B). At the ligand position 0, the side chains of LPA2 Leu0 and EDTSV conformation 1 is similar to the one observed for PDZ1-LPA2, but conformation 2 represents a new conformer with the side

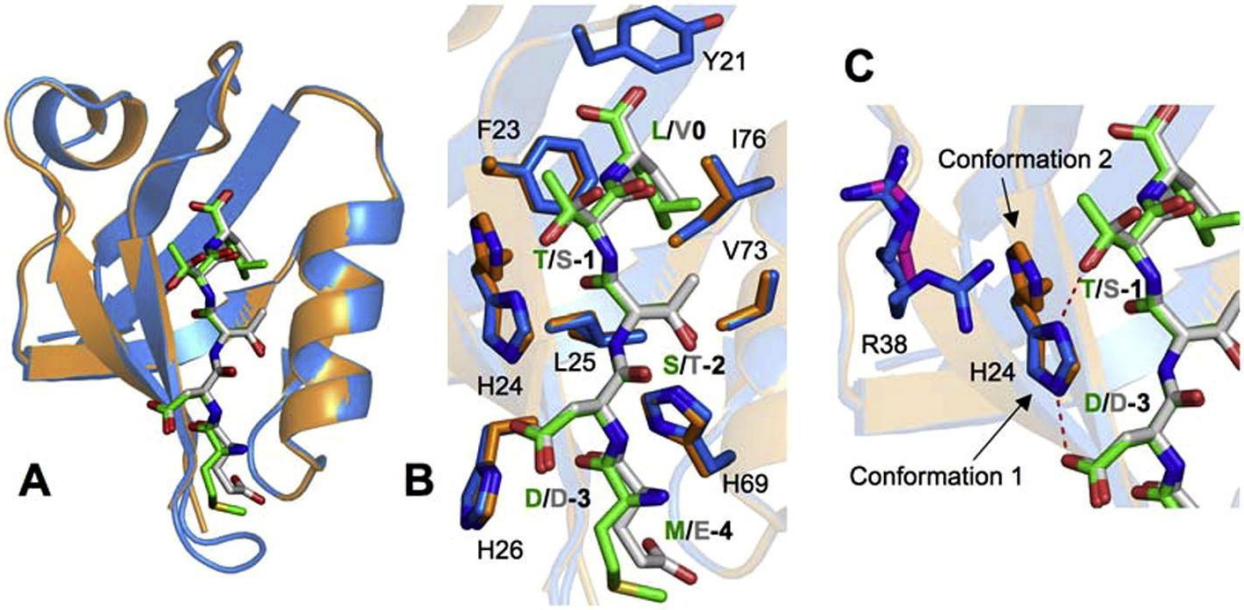

Figure 6. Structural comparison of NHERF2 PDZ1 domains. (A) Superposition of the structures of PDZ1–LPA_2 (light blue; PDB code: 4P0C) and PDZ1-EDTSV (orange; PDB code: 2OCS). PDZ domains are represented by ribbons. Residues in the ligands are displayed as sticks. Carbon atoms are shown in green for LPA_2 and in gray for EDTSV. (B) Superposition of the PDZ1 ligand-binding pockets. Both PDZ1 and ligand residues are depicted by sticks and colored according to the scheme in Fig. 2A. (C) Close-up view of structural differences of His24 and Arg38. Red broken lines represent hydrogen bonds between His24 and LPA_2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

chain packing against the hydroxyl of Ser-1. This conformational change is accompanied by large alteration in the Arg38 rotameric states. In PDZ1-EDTSV, the side chain of Arg38 points away from the bound peptide, whereas in PDZ1-LPA2 it adopts a double conformation with one conformation oriented toward the Thr-1 residue. These observed differences suggest that the conformational changes of His24 and Arg38 underlie the NHERF2 PDZ1 flexibility to accommodate ligands with different -1 side chains and denote a structural explanation for diverse peptide recognition.

Drug Design Perspective and Novel SCN Binding Site

CFTR protein is the product of the *CFTR* gene mutated in patients with CF, which is a lethal autosomal-recessive genetic disease that is most common among Caucasians[82]. We previously have suggested that targeting the NHERF2-LPA2 interaction may have a therapeutic potential in CF treatment, as inhibition of this interaction has been found to be sufficient to enhance CFTR channel activity both in vitro and in vivo[76,79]. These previous findings highlight the significance of our present structure studies and also imply that the structural details of the NHERF2-LPA2 interaction may be valuable in developing new methods and strategies for selective drug design. For instance, this information can be used to create new NHERF2 inhibitors that are potent and specific to block the NHERF2-LPA2 interaction. Such inhibitors have the potential to rescue epithelial cell function in the human CF airway by restoring or increasing CFTR channel activity. However, it should be noted that NHERF2 is capable of binding to a multitude of ligands, through which it regulates many cellular processes essential to normal physiological functions, such as testicular differentiation, signal transduction, endosomal recycling, membrane targeting, and hormone receptor desensitization[83-86]. It is therefore conceivable that the engagement of an inhibitor with the ligand-binding site would interfere with the full spectrum of NHERF2 PDZ-target interactions and could lead to considerable risks with a diverse range of unwanted physiological and hormonal abnormalities. In order to achieve NHERF2 inhibitor selectivity, one possible solution to this challenge is designing partially competitive inhibitors that only affect ligand-specific interactions and bind to a site other than the ligand-binding pocket. In this regard, it is interesting to note that the present structure reveals a small surface pocket adjacent to the ligand binding site (Fig. 7). This pocket is identified based on an extra electron density observed at the surface of the PDZ1 structure (Fig. 7A). Based on the components in the crystal

condition and the shape of the density, the density was assigned as a thiocyanate molecule (SCN). Residues contributing to SCN binding include His26 from β2, Gly53 from β4, and Phe35 and Arg37 from β3 (Fig. 7B). Note that residues His26 and Arg37 shared by the ligand binding site

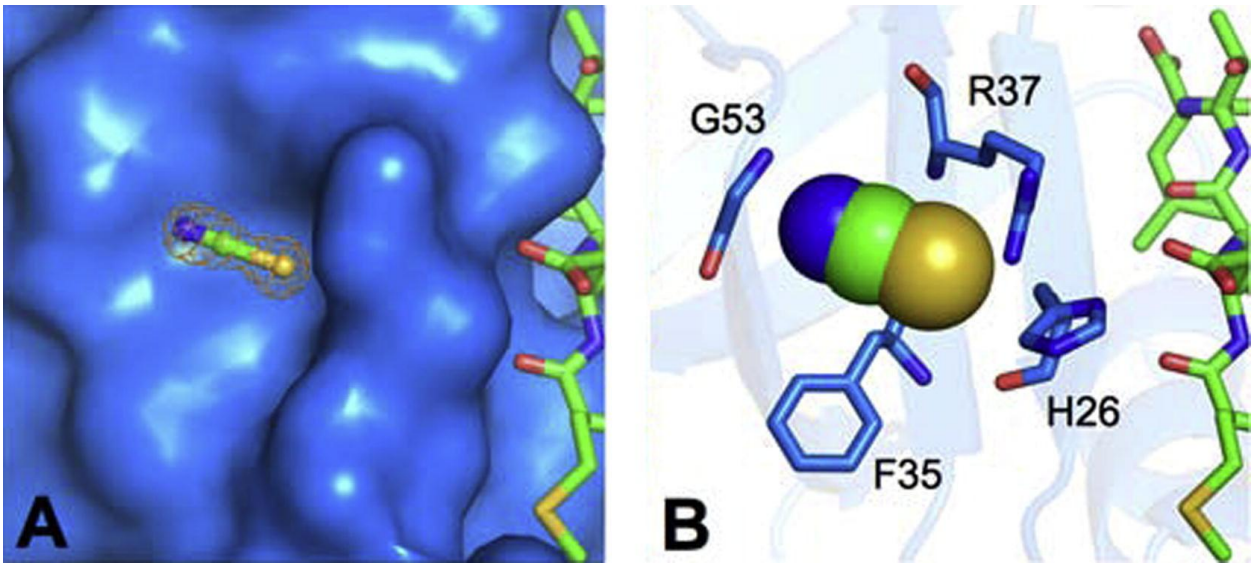

Figure 7. SCN binding pocket. (A) Surface representation of the SCN binding site. PDZ1 surface is colored in light blue. LPA_2 residues are represented by sticks with their carbon atoms colored in green. SCN is depicted by balls-and-sticks overlaid with $2F_o - F_c$ omit map calculated at 1.34 Å and contoured at 1.8 σ. (B) Putative SCN-interacting residues. PDZ1 residues are shown in light blue and LPA_2 residues are shown in green. SCN is depicted by spheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and the SCN binding site are highly conserved (Fig. 5D), and these sites have been implicated in ligand-specific interactions in other PDZ domains[2,87]. Therefore, strategies aiming at exploiting the novel SCN binding site may represent a promising approach to achieve NHERF2-inhibitor selectivity that would allow the differentiation among a wide range of NHERF2-mediatedinteractions. Such strategy should have important implications in specific NHERF2 scaffolding regulation and also in many CFTR-dependent human diseases.

CHAPTER 4 PDZ STRUCTURE AND IMPLICATION IN SELECTIVE DRUG DESIGN AGAINST CYSTIC FIBROSIS

*Published in Curr Drug Targets. 2015;16(9):945-50. All authors agreed with including their work in this dissertation.

Abstract

PDZ domains play an essential role in a number of cellular processes by facilitating protein scaffolding and assembly of protein complexes. These domains consist of 80 to 90 amino acids and are found to recognize short C-terminal sequences of target proteins. Protein complex formation between PDZ target molecules can lead to a number of signaling and regulatory cascades that may either promote or inhibit the activation of certain proteins. It has been shown that the interaction of the PDZ domains of NHERF2 with LPA2 plays an inhibitory role on the cystic fibrosis transmembrane conductance regulator (CFTR) by promoting the assembly of a CFTR–NHERF2–LPA2 complex. CFTR regulates chloride ion transport across the epithelial plasma membrane, and individuals possessing CFTR mutations show decreased protein function and consequently, viscous mucus accumulation due to improper fluid transport. This type of ailment is termed cystic fibrosis. Thus, insight to the structure of PDZ domains and how they function to form macromolecular complexes could be therapeutically important in augmenting CFTR channel activity in cystic fibrosis patients. Here we review the PDZ domain family while dissecting their structure, function and implications in CFTR regulation and cystic fibrosis.

CFTR and Cystic Fibrosis

Cystic fibrosis transmembrane conductance regulator (CFTR) is a glycoprotein encoded by the CFTR gene that is 230 kb in length coding for 1480 amino acids consisting of 6 domains: two transmembrane domains (MSD), two nucleotide binding domains (NBD), regulatory “R”

region, and a PDZ interacting motif[88,89]. CFTR is a member of the ABC transporter ATPase family and allows for the transmembrane flow of chloride ions (Cl^-) down the electrochemical gradient[88]. Such anion flow is critical for proper water balance between the extracellular and intracellular space of epithelial cells in the gastrointestinal tract, pulmonary tract, and ducts lining organs such as the pancreas, testes, and sweat glands[71-74]. Without proper water flow, mucus lining of these tracts becomes dehydrated and accumulate in these areas clogging ducts and giving rise to bacterial infections especially in the lungs[90,91]. Improper function of this channel can be caused by a variety of mutations with F508del being the most common[92,93]. This deletion results in the improper folding of the protein and consequently its degradation before it reaches the plasma membrane of epithelial cells[94-97]. Other mutations may result in a protein that is translocated to the membrane but in reduced number or functionality[98,99]. Such mutation resulting in these complications is termed cystic fibrosis.

Cystic fibrosis is perceived commonly by its associated symptoms such as difficulty breathing, respiratory infections, poor growth, and infertility[100]. Much is known about what causes the disease including allelic specific mutations and the mechanism in which chloride ion transport is affected[100,101]. Yet, even with the wide understanding of how the disease is manifested among humans, there is currently no cure for the illness and only limited therapeutics when it comes to treatment. Additionally, those diagnosed are only predicted to have an average life expectancy of 37 to 40 years[102]. Some of the current treatment options for chronic infections resulting from the disease include mucolytics, antibiotics, and potential lung transplantation as the disease worsens[103-107]. However, emerging studies suggest possible roles of other proteins in CFTR regulation. Interaction with transporters, ion channels, and scaffolding proteins regulates CFTR function at specialized subcellular domains[37,76,108,109]. The formation of a macromolecular complex

between CFTR, Na+/H+ exchanger regulatory factor-2 (NHERF2), and lysophosphatidic acid receptor 2 (LPA2) links CFTR-mediated Cl− transport to LPA2 signaling at the gut epithelial membrane[41]. It has been shown that the NHERF2–LPA2 interaction is mediated by PDZ domains and inhibition of this interaction abolishes the inhibitory effect of LPA resulting in a significant increase in CFTR channel activity. This inhibition was also found to stimulate fluid secretion in pig submucosal glands in the trachea[79]. Such promising findings not only offer insight into potential therapeutic targets for this challenging disease but additionally, provide a foundation for rational drug design in order to more effectively treat cystic fibrosis.

NHERF Function and Involvement in CFTR Regulation

Na^+/H^+ exchanger regulatory factor (NHERF) proteins are protein adaptors that have been shown to be highly expressed in epithelial tissue[78]. The adaptor function of these proteins is due to the presence of two PDZ domains at the N-terminus and a C-terminal ezrin-radixin-moesin (ERM) binding domain that binds cytoskeletal proteins[34]. The functional consensus of NHERF family proteins is that they are able to mediate the formation of multiprotein complexes involving membrane and non-membrane proteins. The NHERF-scaffolded protein complexes are important in cell metabolism and growth[33,110,111]. One of these membrane proteins targeted by NHERF, commonly expressed in epithelial tissue, is CFTR. NHERF proteins have been demonstrated to be essential for CFTR function because they have been shown to stabilize CFTR in the plasma membrane and prevent endocytotic-targeted degradation by CFTR-associated ligand (CAL)[112]. Thus, proper NHERF–CFTR interaction is essential in promoting correct CFTR function and proper anion transport across epithelial membranes. However, PDZ-mediated NHERF interaction has also been shown to be a negative regulator in some biological processes and produce inhibitory effects on certain proteins including CFTR[41,44,75,113]. This dual regulatory role presents a degree of

difficulty when therapeutically targeting such proteins. Therefore, a structural understanding of NHERF-PDZ binding specificity may allow for a more selective targeting and a better prediction of how certain therapeutics could affect specific cellular processes.

The NHERF family consists of four members with NHERF1 and NHERF2 being the most similar and well known[114]. NHERF1 has been shown to be particularly important in cellular signaling by promoting interleukin 8 receptor, beta (CXCR2)-mediated neutrophil migration[2]. NHERF1 was also found to regulate CXCR2 function during spinal cord development and is essential for controlling positions of oligodendrocyte precursor cells[2,47]. CXCR2 is a G-protein coupled receptor whose signaling has been well established[46,115]. In neutrophils, CXCR2 signaling requires the complex formation between CXCR2 and phospholipase C (PLC)-β2, a downstream effector of intracellular signaling[2]. The formation of this complex has been shown to be dependent on the PDZ domains of NHERF1 which bridge the two molecules by interacting with the PDZ-binding motif at their C-terminus[2]. The same type of interaction was observed for CXCR2–PLCβ3, which is required for tumor proliferation in pancreatic cancer cells[116]. However, scaffolding mediated by the PDZ domains of NHERF2 in the CFTR–LPA2 complex formation has been shown to inhibit CFTR function, and disrupting this complex using a cell-permeable LPA2-specific peptide was able to restore CFTR-dependent anion efflux[41]. Therefore, examination of the structural properties of PDZ domain-containing proteins, especially NHERF2, could provide new therapeutic avenues for the treatment of CFTR-associated diseases including cystic fibrosis.

PDZ Domains

PDZ domains are evolutionarily conserved domains present in numerous signaling proteins. These domains are found not only in animals and plants but also yeast and bacteria[117]. The acronym PDZ is derived from the first three proteins originally found to contain such a

domain. These three proteins are known as Postsynaptic density protein 95kD (PSD95), Drosophila disc large tumor suppressor (Dig1), and Zonula occludens-1 protein (ZO-1)[22]. In general, PDZ domains recognize and mediate protein interactions through small motifs localized at the C-terminus of target proteins[23]. Protein recognition by the PDZ domain is important for assembling signal complexes, recycle receptors, protein trafficking, and instituting cell polarity[23,52]. To date, two types of PDZ domains are well studied, class I and class II. Peptides with a consensus sequence (S/T)X(V/I/L) (X denoting any amino acid) and (F/Y)X(F/V/A) form the class I and class II groups respectively[52,80]. Structural insight into PDZ domains revealed an overall and conserved fold of six strands (β1–β6) and two α-helices (αA and αB)[1,2,43,52,80,87,116]. In addition, a similar peptide recognition mode is shared among all the PDZ domains. The 0 peptide residue (C-terminal residue) resides inside a hydrophobic pocket and the −2 peptide residue forms a specific hydrogen bond with a conserved residue[1,52,80]. Despite the similarity of peptide recognition, recent data indicate that PDZ binding is more complex than previously thought. Altogether, the PDZ domain family can recognize and bind to seven C-terminal ligand residues and form sixteen distinctive specificity classes[81]. Additionally, the ability of PDZ domains to bind to different classes of ligands and one ligand capable of binding to different PDZ domains highlight the complexity of PDZ-mediated interactions[81]. Although PDZ domains are structurally simple protein interaction modules, PDZ–peptide interactions create a challenging question to how binding promiscuity and specificity are concurrently achieved. The mechanism to address this problem is still not well understood.

PDZ Structure

PDZ1 and PDZ2 are each found in NHERF1 and NHERF2. Both domains share multiple conserved residues with a great level of sequence similarity[2]. The overall folds of these domains

are similar with two α-helices (αA and αB) and six β-strands (β1–β6) forming an evolutionally conserved fold (Fig. 8A)[1]. Crystal structures of PDZ1– CXCR2, PDZ1–CFTR, PDZ1–PDGFR, and PDZ1–β2AR are well superimposed. Pairwise RMSDs (root-mean-square differences) range from 0.33 to 0.64 Å for entire Cα atoms (Fig. 8B)[2]. In all structures, the peptide ligands bind inside the β2 and αB cleft. The backbones of the bound peptides and their relative spatial positions to the PDZ motifs are highly superimposed with an RMSD range of 0.09 to 0.15 Å. This indicates that the overall PDZ fold has little change upon peptide binding. Consistently, previous results show that localized changes in a few key areas of the PDZ fold were sufficient to alter specificity in PDZ binding[2,63]. The PDZ1 structural comparison provides additional supportive evidence for this observation. Only the peptide-binding pocket shows significant changes, particularly for PDZ residues interacting with peptide ligands at −1 and −3 positions. As shown in Figure (8C), the side chain of Asp22 has two different orientations, and the side chain of Glu43 has four different orientations. This demonstrates the conformational flexibility in the PDZ peptide-binding pocket, and more importantly, this flexibility explains how the PDZ domain binds the −1 ligand residue with different polarity and hydrophobicity[2,4,87,116].

In order to advance our understanding into NHERF2 PDZ1 binding specificity, the structure of PDZ1–LPA2 and PDZ1 bound to a different peptide (EDTSV) was compared[1]. Structural comparison of these domains revealed some similarities and considerable differences in the peptide-binding pockets (Fig. 9A). The side chains of LPA2 Leu0 and EDTSV Val0 follow a similar path at the ligand position 0 entering a pocket that is virtually identical between the two PDZ1 structures. At the −2 ligand position, Ser-2 and Thr-2 are recognized by similar mechanisms where a direct hydrogen bond forms to structurally conserved His69 side chain in both cases. However, residues around the −1 ligand position have noticeable differences. (Fig. 9B). In PDZ1–

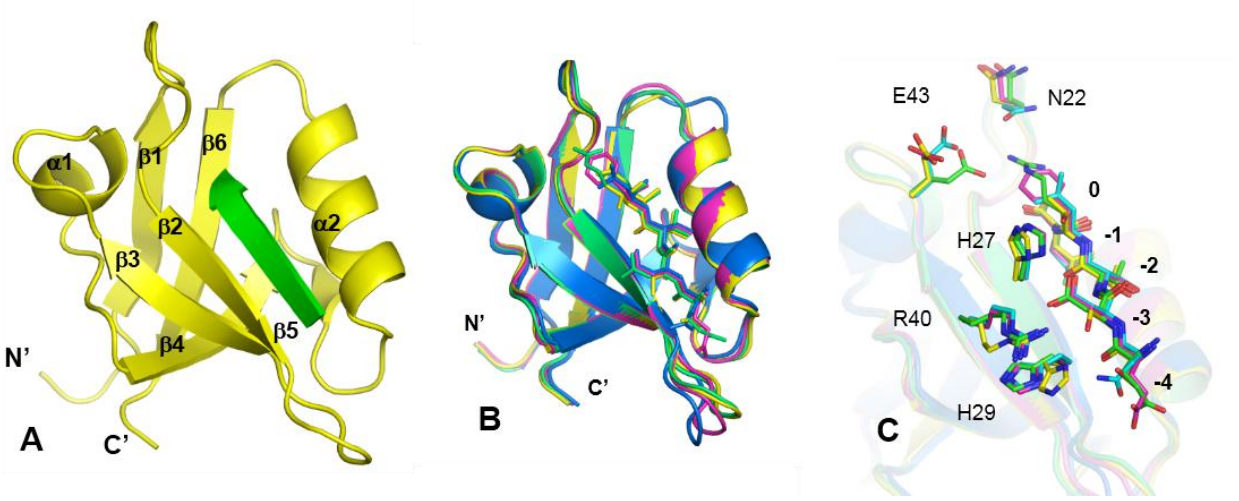

Figure 8. Structures of NHERF1 PDZ1 domain. (**A**) Ribbon illustration of PDZ1–CXCR2. PDZ1 and the CXCR2 peptide are represented in yellow and green, respectively. PDZ secondary structures are labeled based on their location in the sequence. (**B**) Structural alignment of PDZ1–CXCR2 (yellow; PDB code: 4JL7)[2], PDZ1–CFTR (green; PDB code: 1I92)[3], PDZ1–β2AR (blue; PDB code: 1GQ4)[4], and PDZ1–PDGFR (magenta; PDB code: 1GQ5)[4]. PDZ domains are depicted by ribbons, while ligand residues are represented as sticks. (**C**) Structural alignment of PDZ1 pockets important for ligand binding. Ligand and PDZ residues are represented by sticks and colored according to the scheme in B.

LPA2, His24 simultaneously interacts with Thr-1 and Asp-3 under a single conformation, whereas in PDZ1–EDTSV, His24 has two conformations in which the first conformation is observed in PDZ1–LPA2 and the second conformation presents the side chain packing against the hydroxyl of Ser-1. Additionally, a large change in the Arg38 rotameric state is coupled to the second conformation. In PDZ1–EDTSV, the Arg38 side chain points away from the bound peptide, whereas in PDZ1–LPA2, Arg38 adopts two conformations with one conformation oriented near the Thr-1 residue. These observed differences suggest that the conformational changes of His24 and Arg38 may contribute to NHERF2-PDZ1 binding promiscuity and allow for recognition of peptides with different −1 side chains[1].

Drug Design Perspective: Cystic Fibrosis

As demonstrated for NHERF1 bridging of CXCR2 and PLCβ along with NHERF2 bridging of CFTR and LPA2[1,49,79], specific interruption of certain PDZ-mediated interactions

could prove to be therapeutic in augmenting the normal function of certain proteins. Since previous studies have found such interruption to be sufficient in promoting CFTR channel activity[41,44,75], the targeting of PDZ in NHERF2 by pharmacological approaches might provide optimistic treatment for patients suffering from cystic fibrosis. Inhibition of the NHERF2–LPA2 interaction by the small molecule CO-068 elevates cAMP levels in proximity to CFTR and upregulates CFTR channel activity[79]. This proof-of-concept study demonstrates that CFTR Cl^- channel function can be fine-tuned by pharmacological approaches through modulating PDZ-mediated interactions within the CFTR-containing protein complexes. While the cellular effects of CO-068 has been thoroughly addressed[82], current findings underline the structural significance of PDZ domains. Better understanding the NHERF2–LPA2 interface may prove worthwhile in rational drug design. For instance, structural data can promote the development of potent NHERF2 inhibitors specifically designed to interrupt the NHERF2–LPA2 interaction. Inhibitors of the NHERF2–LPA2 interface may possess potential to rescue epithelial cell function in the human CF airway by increasing or restoring CFTR channel activity. However, it should be noted that NHERF2 is capable of binding to numerous ligands that are involved in several physiological processes such as signal transduction, testicular differentiation, membrane targeting, endosomal recycling, and hormone receptor desensitization[83-85]. The use of non-selective inhibitors to disrupt NHERF2–LPA2 interface will likely interrupt the NHERF2 interaction network setting the stage for physiological complications at considerable risk. Thus, alternative targeting strategies may be necessary in order to produce beneficial therapeutic outcomes without a vast array of side effects.

In order to achieve selectivity in NHERF2 inhibition, one practical option is the use of partially competitive inhibitors. Ideally, such inhibitors must not directly target the PDZ binding site but should be able to engage ligand-specific binding residues through a separate nearby pocket.

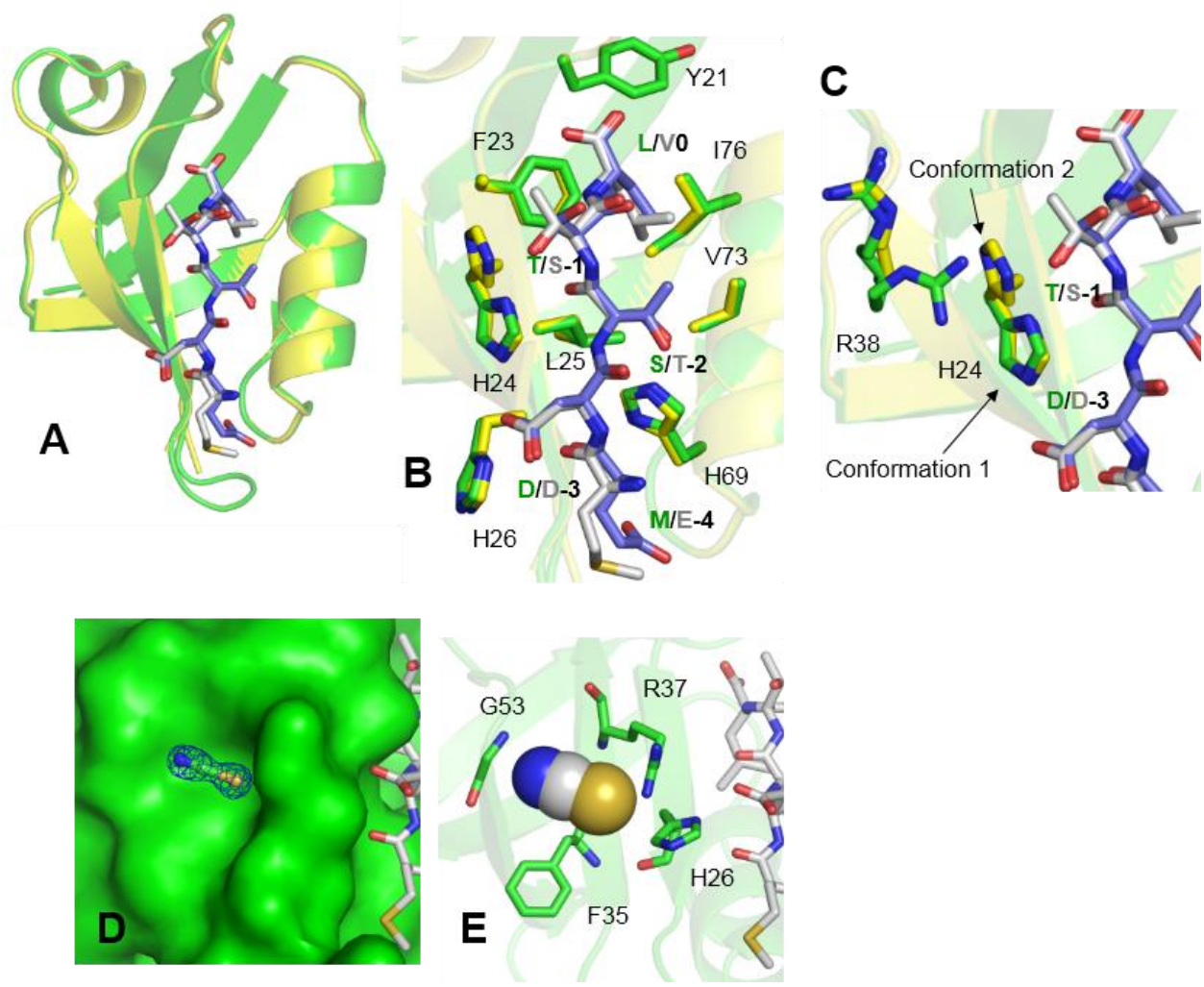

Figure 9. Crystal structure of NHERF2 PDZ1–LPA2 complex. (**A**) Structural alignment of PDZ1–LPA2 (green; PDB code: 4POC)[1] and PDZ1–EDTSV (yellow; PDB code: 2OCS). Ribbons depict the PDZ domain, and ligand residues are exhibited as sticks. Carbon atoms are shown in grey for LPA2 and in blue for EDTSV. (**B**) Structural alignment of the PDZ1 ligand-binding pockets. PDZ1 and ligand residues are colored according to the scheme in A and represented as sticks. (**C**) View of His24 and Arg38 structural differences. Red broken lines represent hydrogen bonds between His24 and LPA2. (**D**) Surface representation of the SCN binding site. PDZ1 surface is colored in green. LPA2 residues are colored in grey and represented as sticks. 2*Fo* − *Fc* omit map (calculated at 1.34 Å and contoured at 1.8 σ) overlays SCN depicted as balls-and-sticks. (**E**) Putative SCN-interacting residues. PDZ1 residues are represented in green and LPA2 residues are represented in grey. SCN is depicted by spheres. *(The color version of the figure is available in the electronic copy of the article).*

The structural analysis of the NHERF2 PDZ1–LPA2 complex suggests a potential site for such targeting strategy. In the PDZ1–LPA2 structure, a small surface pocket was found in close proximity to the ligand-binding site (Fig. 9C)[1]. This pocket is occupied by a thiocyanate molecule

(SCN) from crystallization conditions. Residues contributing to SCN binding include His26 from β2, Gly53 from β4, and Phe35 and Arg37 from β3 (Fig. 9D). The residues His26 and Arg37 are shared by both the ligand-binding site and SCN binding site. Interestingly, these two residues have been implicated in ligand-specific interactions in other PDZ domains[2,116]. Therefore, strategies aiming at exploiting the SCN binding site may represent a promising approach to achieve NHERF2–inhibitor selectivity and allow for ligand differentiation among a wide range of NHERF2-mediated interactions.

Concluding Remarks

From the promotion of CXCR2-mediated neutrophil migration to stabilization of CFTR in the cytoplasmic membrane of epithelial cells, the PDZ domains of NHERF proteins are imperative for promoting the proper functioning of membrane proteins in various cellular processes. However, the fact that NHERF2 bridging of LPA2 and CFTR has a negative impact on CFTR channel activity suggests a complex role of NHERF2 in CFTR regulation. While targeting the NHERF2-mediated LPA2–CFTR interaction is an attractive therapeutic option for cystic fibrosis, PDZ binding promiscuity makes such a targeting strategy seem intangible, as the substrate-binding site of PDZ domains may not be feasible for selective drug design. Nevertheless, high-resolution structures of PDZ domains are now leading insight in alternative strategies for such proteins. In the case of PDZ1 of NHERF2, small molecule binding motifs adjacent to the peptide recognition sequence may serve as a viable target for small molecule inhibitors. This not only would be beneficial for augmenting CFTR activity in cystic fibrosis patients but also potentially provide insight into the treatment of other diseases involving PDZ-containing proteins.

PART II

Insight into Structural Properties of the Nogo-b Receptor and Eluding the Bottleneck of Crystallography

CHAPTER 5 SOLUTION STRUCTURE OF THE TRANSMEMBRANE NOGO-B RECEPTOR AND INSIGHT INTO ITS TOPOLOGICAL ORIENTATIONS BY SMALL ANGLE X-RAY SCATTERING ANALYSIS

Abstract

The Nogo-B Receptor (NgBR) is newly identified class I transmembrane receptor known for its role in binding Nogo-B and promoting angiogenesis by chemotaxis *in vivo*. Most recently we discovered that NgBR shows binding specificity for the farnesylated oncoprotein Ras and facilitates Ras activation by recruitment to the plasma membrane. To understand the potential mode of interaction with farnesylated Ras, we performed Small Angle X-ray Scattering (SAXS) analysis of the NgBR construct 79–293 which includes the currently accepted cytosolic and transmembrane domains with a portion of the extracellular region. SAXS analysis reveals the radius of gyration (Rg) for our NgBR construct to be 18.2 Å with a maximum end to end distance (Dmax) of 61 Å. Ab initio modeling returns a globular molecular envelope with an estimated molecular weight of 23 kD closely correlated with the calculated molecular weight of our NgBR construct. These data indicate that our NgBR construct is well folded which exists as a globular domain in solution with no apparent intrinsic disordered regions. However, this result is conflicting with the currently accepted topological orientation of NgBR which would separate our NgBR construct into three distinct domains: extracellular, transmembrane and cytosolic domains. However, comparison of the molecular envelope structure of our NgBR construct with the structure of Undecaprenyl Pyrophosphate Synthase (UPPs) which shares similar sequence homology to NgBR, further supports our conclusion that the regions included in our construct are

folded into a single globular domain, as the two structures are highly superimposed. This led us to propose the idea that NgBR may exist in more than one topological orientation and that the currently accepted transmembrane domain may internalize and contribute to the overall fold of NgBR cytosolically. Altogether, our SAXS analysis provides the first molecular envelope structure of the Nogo-B Receptor and provides evidence for a topological orientation of NgBR not previously identified.

Introduction

The Nogo-b Receptor (NgBR) is a transmembrane receptor protein that has been found to localize to the plasma membrane as well as the endoplasmic reticulum (ER). Since its identification in 2006, there has been increasing investigation of NgBR due its medical relevance in cholesterol trafficking, cardiovascular disease, cancer, and Ras signaling. NgBR was first identified due to its interaction with Nogo-B, a protein that plays a major role in the development of vasculature of endothelial cells[118]. NgBR has also shown to play an influential role in embryogenesis. Experiments performed by Park and colleagues demonstrated that the knockout of NgBR leads to lethality in mouse embryos due to free intracellular cholesterol accumulation and defective regulation of protein glycosylation[119]. Due to its role in vasculogenesis, it has also been shown to be critically involved in tumor growth and cancer metastasis. NgBR has been specifically demonstrated to show increased expression in breast cancer which has been proven essential for EGF signaling in breast cancer cells contributing to both tumor growth and size[120]. Furthermore, we have recently proven that NgBR is able to bind the farnesylated oncoprotein Ras and recruit it to the plasma membrane for activation[120]. As is generally understood, the farnesylation of the CAAX box of Ras is required for its subsequent localization to the plasma membrane leading to Ras activation[120-123]. Furthermore, we showed that the farnesylation of Ras is essential for NgBR

binding to Ras and NgBR does not bind to the unfarnesylated Ras protein[120]. Interestingly, NgBR shares significant sequence identity to the cis-isoprenyl transferase family and maintains conserved residues responsible in the binding of farnesyl diphosphate (FPP)[120,124-126]. Thus, it is possible that NgBR and cis-isoprenyl transferase family may share a similar structural mechanism in binding to farnesyl groups.

The Nogo-B Receptor is a 293 residue type 1 receptor protein with three primary domains: a cytoplasmic domain consisting of residues ~140-293, a single transmembrane region with residues ~120-139, and an extracellular domain containing residues ~47-119 with a putative 46 residue signaling sequence on the N-terminus[127]. NgBR has also been found to adopt alternate topological conformations with its C-terminal domain existing in either the cytosol or lumen of the ER. In the ER, NgBR has been found to bind and stabilize Nieman-Pick type C2 protein (NPC2) consequently facilitating cholesterol homeostasis[128]. However, its major topological orientation appears to consist of its C-terminal domain existing cytosolically. In the cytosol, it has been shown to promote human cis-isoprenyl transferase (hCITase) activity therefore facilitating dolicol synthesis[129]. NgBR's C-terminal domain has been shown to share significant sequence similarity to the cis-IPTase family with upward of 44% with cis-IPTase of *Micrococcus luteus*[129]. This protein family is found to bind isoprenyl lipids and catalyze condensation of isopentenyl diphosphate to farnesyl diphosphate (FPP). However, even with high sequence similarity, NgBR does not possess such activity[118]. NgBR has also been shown to contain conserved residues with an additional cis-IPTase, undecaprenyl pyrophosphate synthase (UPPs). UPPs is a lipid carrier for peptidoglycan that synthesizes UPP and is responsible for synthesis of the bacterial cell wall[126]. UPPs has also been shown to bind and catalyze the condensation of isopentenyl diphosphate to FPP[126]. The binding of UPPs to its substrates is mediated the presence of a hydrophobic pocket at

the core of the protein. In case of its interaction with FPP, binding is manifested by hydrophobic interactions occurring between FPP farnesyl group and residues L85 and L88 of the α-3 helix. Interestingly, these farnesyl interacting residues are conserved with the two residues (I117 and L120) of NgBR required for H-Ras binding. Therefore, if NgBR was to fold similar to UPPs, it could be possible that NgBR shares a similar structural mechanism to that of UPPs for farnesyl binding.

The structure of NgBR is currently unknown and is likely due to its lack of solubility or its disordered nature[127]. Both the full-length recombinant protein as well as its individual domains have been shown to possess limited solubility upon expression. Previous studies using bioinformatics, CD spectroscopy, and NMR have been used to characterize NgBR's individual domains[127]. Such data suggests that NgBR's cytosolic domain is only partially folded and its extracellular region intrinsically disordered, although no structural information has yet been determined. In this study, we circumvent this issue by the development of a new expression construct for NgBR. In our construct, we include residues corresponding to the entire cytosolic domain of NgBR with the inclusion of its transmembrane domain and some extracellular components. This is rationalized based on its sequence similarity to the UPPs and the hypothesis that NgBR may exhibit a similar structural mechanism as UPPs for farnesyl binding. The resulting expression of the recombinant protein demonstrates that NgBR exhibits good solubility and is primarily monomeric based on size exclusion chromatography (SEC). Additionally, small angle x-ray scattering (SAXS) analysis reveals the recombinant protein is well folded with a similar shape and size to that of UPPs and is verified by the calculated radius of gyration (R_g), Parod volume, and D_{max} of the SAXS data. Furthermore, SAXS analysis demonstrates that the recombinant monomeric NgBR protein consisting of residues 79-293 is able fold as a single

globular domain in solution. However, this result conflicts with the currently accepted topology of NgBR in the plasma membrance which would separate our construct into three domains. Because NgBR has been shown to adopt alternate topological conformations in ER membrane, it is possible a similar topological change occurs at the plasma membrane allowing residues in our construct to fold into a single domain. Therefore, this would implicate the need for NgBR to adopt a topological orientation not yet identified.

Materials and Methods

Sequence Alignment and Modeling

Identification of proteins possessing structural data with sequence identity similar to NgBR was carried out using Basic Local Alignment Search (BLAST) of the Protein Data Bank (PDB). Manual alignment of NgBR with UPPs was then performed to identify conserved residues in which may be critical to NgBR's structure. Conserved residues exhibiting potential structural implications were then determined by based on the structure of UPPs (PDB code 1X08). Residues identified as potentially critical to NgBR's overall fold based on UPPs structural analysis were then included in the NgBR expression construct.

Protein Expression and Purification

The open reading frame corresponding to human NgBR (79-293) was amplified and cloned into a pCDF-SUMO vector containing an His6-SUMO tag. The vector was transformed into BL21(DE3) cells for recombinant protein expression. Cells were inoculated into LB media supplemented with 34 μg/mL streptomycin and grew to an optical density of 0.5. Induction growth was carried out in 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 15°C overnight. Harvested cells were lysed using a French press, spun down, and the supernatant was pooled for purification. The supernatant was passed through a HisTrap column (GE Healthcare), and eluted

out. Yeast SUMO protease (ULP1) was added to pooled fractions containing His6-SUMO-NgBR (79-293) and incubated overnight to remove His6-SUMO tag. The pooled fractions passed through HisTrap column again, and the native NgBR (79-293) was collected from the flowthrough and wash. NgBR (79-293) was further purified by HiLoad 16/60 Superdex 200 column in 20 mM Tris pH 7.5, 150 mM NaCl and 5% glycerol.

Small Angle X-Ray Scattering Data Collection and Analysis

Small Angle X-ray Scattering (SAXS) data was collected at Argonne National Laboratory beamline ID18. Prior to data collection, samples were dialyzed with buffer [20mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 1mM TCEP] overnight, and centrifuged at 13,000g for 10 minutes at 4°C. 20 mg/mL NgBR was injected into S200 column [GE Heathcare]. Scattering frames corresponding to monomer NgBR were scaled, averaged and subtracted from averaged buffer frames. The NgBR (79-293) monomer scattering curve was analyzed using the ATSAS suite. PRIMUS and GNOM were used to measure the radius of gyration and maximum radius (Dmax), respectfully. DAMFILT was used to generate a dummy atom model of NgBR (79-293) monomer.

Results

BLAST Analysis and Expression Construct Identification

BLAST analysis of the Protein Data Bank was carried out using the amino acid sequence corresponding to human NgBR. The top result returned was that of the *E. coli cis*-IPTase Undecaprenyl Pyrophosphate Synthase (UPPs) containing 24% sequence identity and 39% sequence similarity. Structural analysis of UPPs (PDB code 1X08) reveals an overall composition consisting of ten α-helices and six β-strands (Fig. 10B). Further analysis of the structure reveals the compound farnesyl pyrophosphate (FPP) secured deep in a hydrophobic tunnel formed by

helices α-1 and α-3 along with the β-1 strand of UPPs (Fig. 10B). FPP is further secured by hydrophobic interactions by leucine 85 and leucine 88 of the α-3 helix (side chains depicted in yellow). Upon manual alignment of NgBR with UPPs, NgBR is found to possess conserved residues corresponding to all α-helices and all β-strands apart from the first β-strand (β0) found in UPPs (Fig. 10A). Additionally, NgBR residues I117 and L120 are conserved with two FPP interacting residues (L85 and L88) of the UPPs crystal structure (Fig. 10A and 10B). Interestingly, residues conserved by NgBR with the UPPs α1, α2, α3, and β1 regions are found to compose NgBR's transmembrane domain and part of its extracellular domain. This presents problem if NgBR was to fold similar to that of UPPs. If NgBR was to exhibit similar structure to UPPs, these regions would need to be present in the cytosol and does not agree with the current topological oreintations of NgBR [127,129]. However, if the conserved residues were removed from the UPPs structure (Fig 1B, segments in red), it is unlikely that NgBR would be able to fold correctly based on the UPPs model (Fig. 10C). Thus, due to these observations, an expression construct (residues 79-293) was created for NgBR encompassing all conserved residues with that of UPPs.

Expression and Purification of NgBR

Residues 79-293 of NgBR were cloned and expressed as described in the methods. As can be observed, NgBR achieves good expression as shown by SDS-PAGE with a solid band at ~36.5kDa in the induced cell culture corresponding to the expected molecular weight of the construct plus the His6-SUMO tag (Fig. 11A). Additionally, over 50% of total recombinant NgBR present in the cell lysis appears to be recovered in the supernatant fraction (Fig 11A). Comparatively, we tested an additional construct in which included residues corresponding to NgBR's cytosolic region (137-293). As can be observed, NgBR achieves good expression with the solid band observed at ~32.5 kDa in the induced cell culture corresponding to NgBR with

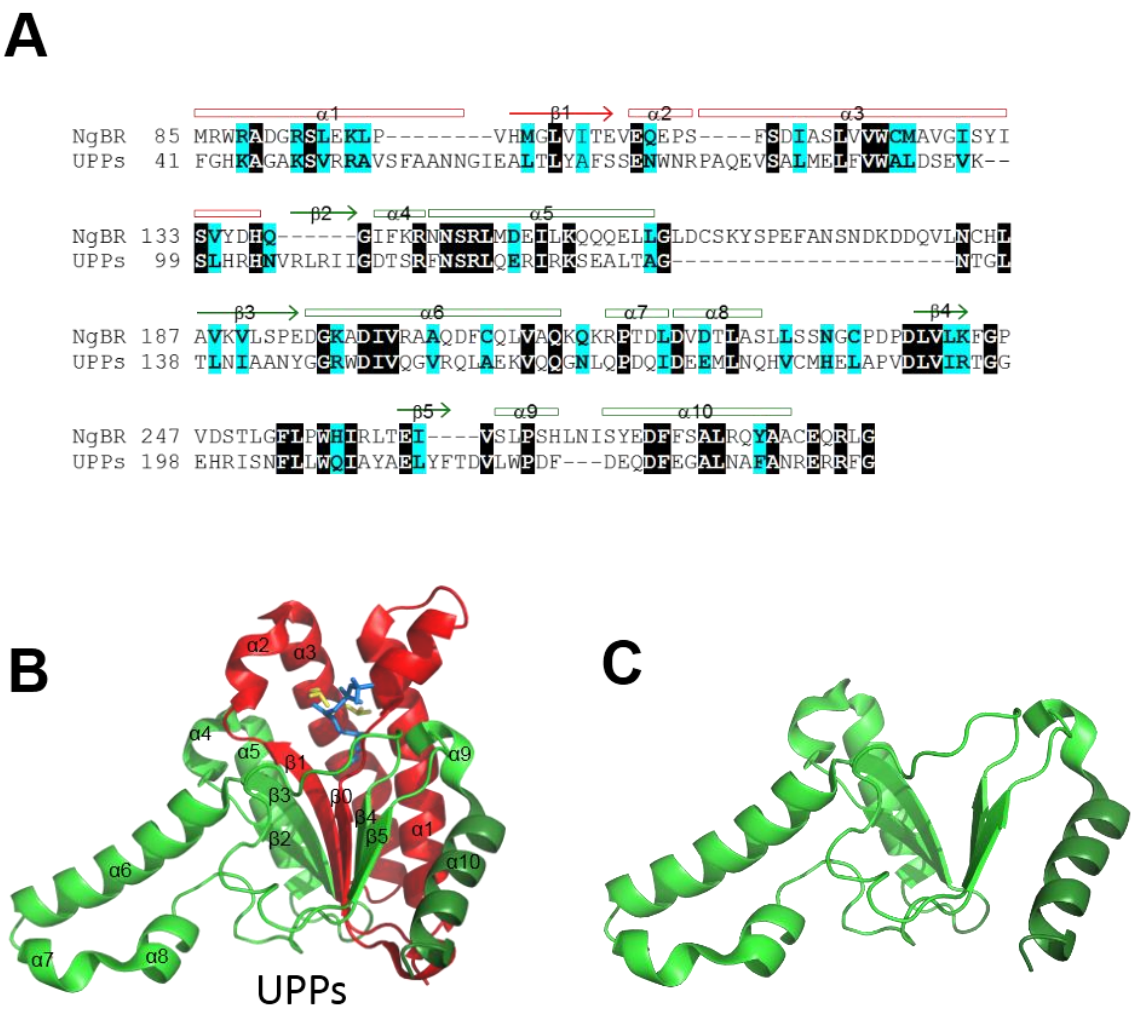

Figure 10. Modeling the NgBR–HRas interaction. (A) Sequence alignment of NgBR and Undecaprenyl pyrophosphate synthase (UPPs). NgBR and UPPs share 24% sequence identity and 39% sequence similarity. Such a level of sequence similarity suggests that the overall fold of NgBR may be similar to that of UPPs. Identical residues are shown as white on black, and similar residues appear shaded in cyan. Secondary structure elements in UPPs are displayed above the sequences and numbered according to their position in the sequence. Sequence numbering is displayed to the left of the sequences. (B) Crystal structure of UPPs (PDB code: 1X08) in complex with a farnesyl diphosphate (FPP) (blue). The secondary structure elements involved in FPP binding are colored in red. (C) UPPs structure with residues of UPPs corresponding to NgBR's transmembrane region with part of its extracellular domain removed resulting in eradication of the FPP binding site and likely reduction of UPPs structural integrity.

attached His6-SUMO tag (Fig 11B). However, when assessing the latter constructs solubility, NgBR is virtually non-existent in the supernatant fraction with comparison to the total cell lysate

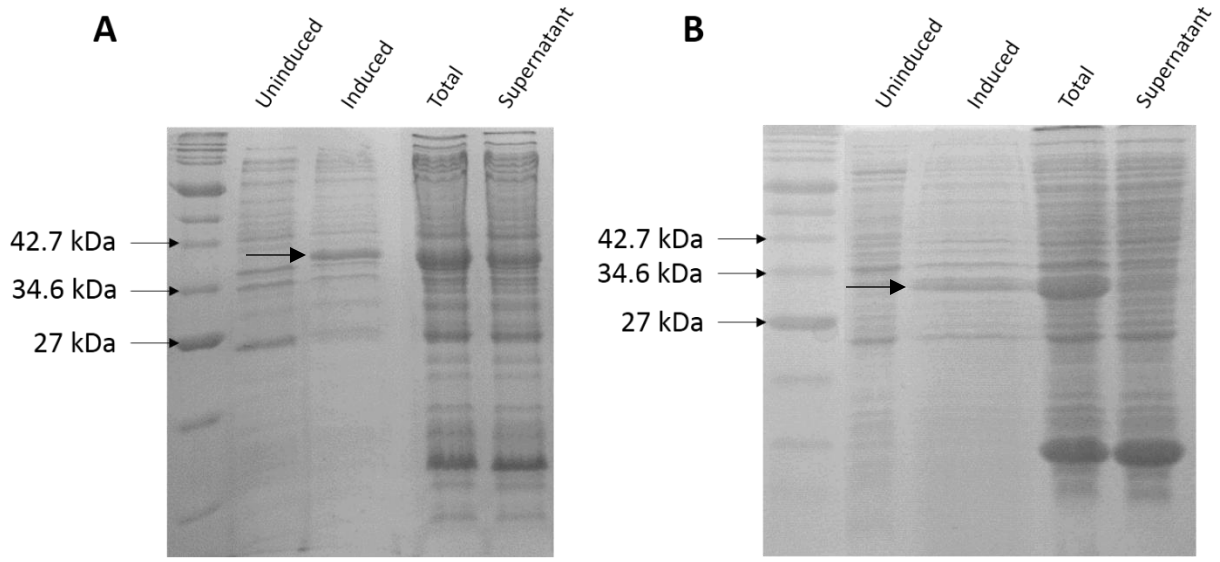

Figure 11. Expression test for NgBR constructs. (A) NgBR expression construct 79-293 which included NgBR's cytosolic domain along with its transmembrane and some of its extracellular region. Expression is noted in the induced cell culture with the band corresponding to the NgBR construct + SUMO tag at ~36.5 kDa. Following cell lysis NgBR is observed in the total cell lysate fraction composed of both soluble and insoluble protein fractions. Upon centrifugation, NgBR (~60% as compared to the total cell lysate) is observed in the supernatant suggesting that over half of the expressed protein demonstrates good solubility. (B) Expression and solubility analysis of NgBR residues corresponding to its cytosolic domain (137-293). Expression is noted for NgBR + SUMO tag with protein band observed at ~32.5 kDa. Following cell lysis, NgBR is observed in the total cell lysate fraction but upon centrifugation, little to no soluble protein is recovered in the supernatant fraction suggesting that residues present in the 79-293 construct aid in the solubility of NgBR's cytosolic domain.

(Fig. 11B). This suggests that inclusion of the additional residues in the prior construct appears to enhance the recombinant protein solubility.

Following expression, the NgBR construct (79-293) was purified as described in the methods. Upon subjection to Ni^{2+} affinity purification, NgBR appears to be eluted around 38% elution buffer (Fig. 12A) which is verified by SDS-PAGE exhibiting distinct bands at ~36.5 kDa corresponding to the expected molecular weight of NgBR with the His6-SUMO tag (Fig. 12B). Following the first nickel column, the protein fractions were pooled and His6-SUMO tag was then removed by protease digestion with yeast ULP1. The cleaved protein mixture was then subjected

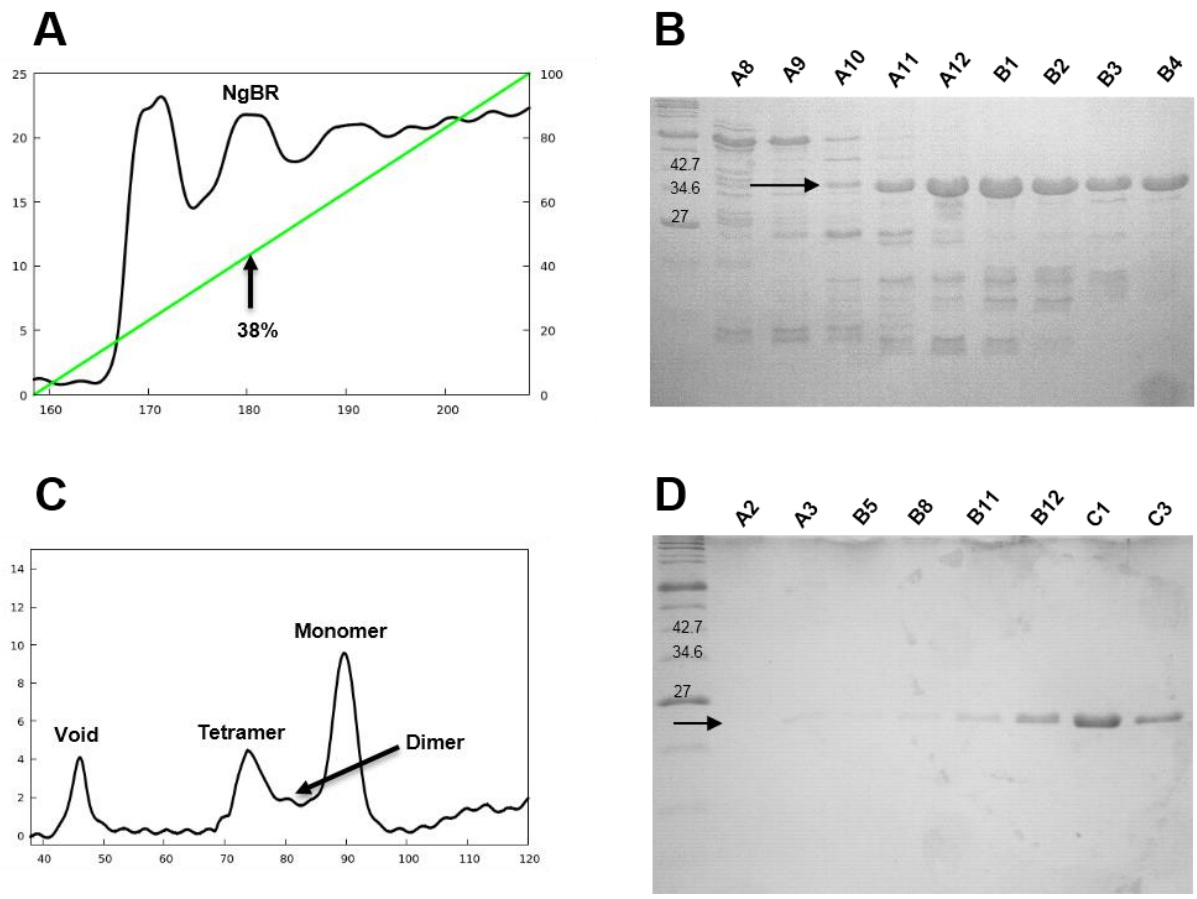

Figure 12. Purification of the NgBR (79-293) construct. (A) Chromatogram of the first Nickel affinity column. NgBR is eluted around 38% elution buffer (%B, green). (B) SDS-PAGE fractions from the first Nickel affinity purification. As can be observed, the protein band corresponding to NgBR + His6-SUMO tag at ~36.5 kDa is present in fractions A10-B4. Fractions A11-B4 were then pooled and subjected to protease cleavage of the His6-SUMO, followed by a second round of Nickel purification in order to remove any additional protein contaminants. (C) Further purification of NgBR by size exclusion chromatography. As can be seen by the chromatogram, NgBR exists primarily as a monomer based on its elution peak at ~ 90 ml. Tetrameric NgBR appears to be the second most prevalent species based on the observed peak around 76 ml. Some NgBR dimer is observed at ~80 ml but appears to be the least prevalent species in solution. Some NgBR protein is noted in the SEC void volume at 46 ml which suggests that NgBR displays some non-specific aggregation in solution.

to a second round of nickel purification to remove the His6-SUMO tag. Purified NgBR was then subjected to size exclusion chromatography (SEC) in the last step of purification. As can be observed by the SEC chromatogram, four distinct peaks are noted by UV detection (Fig. 12C). Molecular weight estimations by elution volume suggests the largest peak at ~90 ml corresponds

to NgBR in the monomeric state. Based on UV absorbance, this appears to be the majority of the species in solution. The next largest peak is observed at ~76 ml and consistent with the molecular weight corresponding to NgBR in the tetrameric state. Additionally, a small population of dimeric NgBR is observed at ~80 ml but appears to be the population minority. Furthermore, final peak at 46 ml corresponds to the void volume of the column in which is typically composed of protein or other large molecular aggregates. Some NgBR is found to be present in the void suggesting that the recombinant NgBR protein exhibits a certain degree of non-specific aggregation. However, SDS-PAGE of SEC fraction (Fig. 12D) demonstrates that NgBR is able to be purified with virtually no additional protein contamination suitable for successive structural analysis.

Small Angle X-ray Scattering Analysis

Following purification, NgBR's structural properties were elucidated using biophysical characterization by Small Angle X-ray Scattering (SAXS). In order to ensure monodispersity of the sample, an inline SAXS data collection approach was implemented. Concentrated NgBR (20 mg/ml) consisting of monomer, dimer, and tetrameric states were subjected to size exclusion chromatography with tandem protein diffraction upon elution from the column. Of the data collected, SAXS analysis was conducted for NgBR's monomeric state. Correction of background scattering of the SEC buffer was performed by subtraction in which resulted in the scattering curve for NgBR as seen in blue (Fig. 13A). For comparison, the CRYOSOL-generated scattering of UPPs (solid red line) was overlaid with scattering data of monomeric NgBR (Fig. 13A). As can be observed, NgBR's subtraction curve follows the theoretical scattering of UPPs very well at low q range suggesting that NgBR may exhibit an overall structure similar to that of UPPs. At increasing q values, slight variation is noted between NgBR scattering and UPPs. Such an observation may implicate that NgBR exhibits some higher resolution structural differences with UPPs. The Kratky

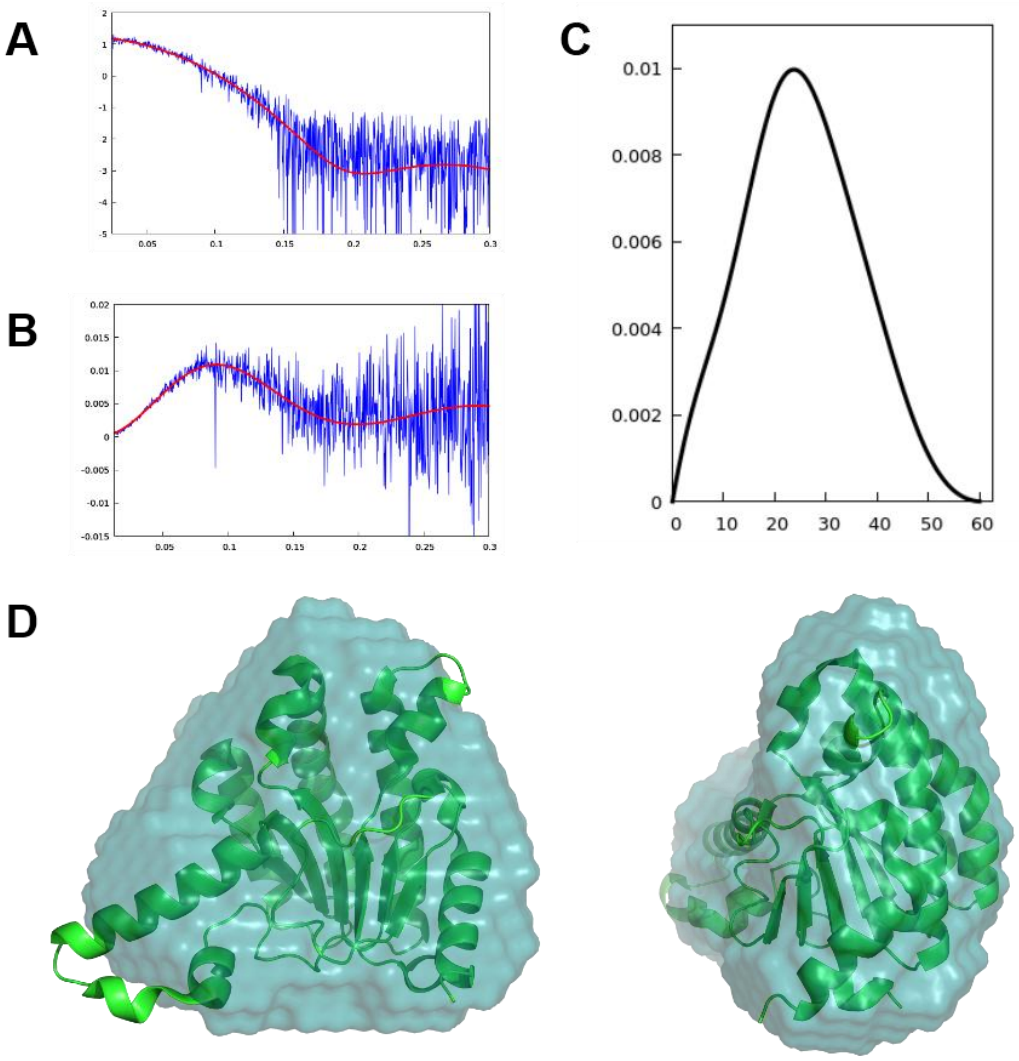

Figure 13. Small angle X-ray scattering analysis. (A) Subtraction scattering curve of NgBR (blue) overlay with theoretical scattering of UPPs (solid red line). NgBR scattering appears to follow closely to that of UPPs at low q values with slight variation at higher q suggesting that NgBR may share a similar fold to that of UPPs with potential high resolution structural differences. (B) Kratky plot representation of NgBR scattering. NgBR scattering (blue) exhibits a smooth bell shaped curve with higher q values averaging near base line characteristic of a well folded macromolecule in solution. (C) Pair distribution curve (P(r)) for NgBR. NgBR exhibits a smooth bell shape curve which is characteristic of a globular macromolecule in solution. (D) DAMFILT *ab initio* dummy atom model of NgBR. Modeling of SAXS data for NgBR using the program DAMFILT reveals a molecular envelope solution structure of NgBR (blue). Superposition of UPPs structure (PDB code 1X08) (green) demonstrates that UPPs superimposes well suggesting that the NgBR (79-293) expression construct folds similar to that of UPPs as a globular molecule in solution.

plot of NgBR's subtraction curve (Fig. 13B), demonstrates a similar pattern with lower q values of NgBR following closely to UPPs theoretical scattering. Again, overlay of UPPs scattering with

NgBR at high q implicates some possible variation. However, scattering as depicted by NgBR's Kratky plot exhibits an overall curve representative of a well folded macromolecule in solution. This is determined by the smooth bell shaped scattering curve of NgBR with higher q values averaging near baseline, characteristic of a well folded protein in solution[130].

The experimental Radius of gyration (R_g) value for NgBR is calculated to be 18.2 which closely coincides with the R_g of UPPS (18.6). Such a correlation implies that the recombinant NgBR protein exhibits a similar overall size to UPPs. This is further verified by the analysis of NgBR's pair distribution function. The maximum end to end distance (D_{max}) of NgBR is calculated to be 61 similar to UPPs D_{max} of 58.61. NgBR's similarity in size to UPPs is more so confirmed by analysis of the Parod volume calculated to be 7.22 and 6.76 for NgBR and UPPs respectively. Additionally, NgBR exhibits a smooth bell shaped p(r) curve which is characteristic of a globular macromolecule in solution (Fig. 13C). Finally, *ab initio* modeling of the SAXS data as performed by the program DAMFILT reveals the first in solution molecular envelope structure of NgBR (Fig. 13D) with calculated molecular weight (MW) to be 22.96 kDa. This is consistent with the estimated MW for the NgBR construct of 24.13 kDa. Upon superposition of the UPPs crystal structure with the NgBR molecular envelope, it is observed that UPPs superimposes very well with some possible differences in the α-6, α-7, and α-8 segments of UPPs (Fig. 10B and 13D). However, though is likely that NgBR and UPPs do exhibit some structural differences, SAXS analysis verifies that the recombinant monomeric NgBR protein consisting of residues 79-293 is able fold as a single globular domain in solution.

Discussion

Recently, there has been growing interest in the nogo-b receptor due to its therapeutic implications in a variety of human diseases. However, its limited structural information due to

complications in solubility has been the bottleneck in its biophysical characterization. Additionally, the ability of NgBR to adopt varying topological conformations further complicates the development of an appropriate expression construct. The current study provides evidence for potential circumvention of this issue. By inclusion of residues sharing conservation with UPPs, our data suggests that incorporation of NgBR's transmembrane domain with part of its extracellular region increases recombinant protein solubility. This observation may in turn open potential opportunities for future high resolution structural studies of the nogo-b receptor. Additionally, SAXS analysis confirms that our protein construct is able to fold as a single globular domain in solution. However, these data do not agree with the topological orientations of NgBR noted in the literature which would separate our construct into three domains: cytosolic, transmembrane, and extracellular. This may implicate that NgBR is able to adopt a topological conformation not yet identified. Upon some physiological stimulus, NgBR could experience a dynamic structural change where residues included in our construct internalize and fold with its cytosolic domain. Similar instances of domain transition have been noted in the literature. Multiple cases have been observed where membrane proteins were found to adopt varying topological conformations due to alteration in the phospholipid environment[131-134]. It has been specifically demonstrated in the case of *E.* coli LacY where deficiency of phosphatidylethanolamine (PE) results in the inversion of its N-terminal six transmembrane domain α-helical bundle. Consequently, this inversion results in the translocation of its seventh transmembrane domain to the extracellular space[131,132]. A similar topological transition is observed in the case of phenylalanine permease (PheP) upon PE reduction. In this instance, PheP is able to undergo complete inversion of its N-terminal first helical hairpin[135]. With evidence that such domain transitions have been observed for other membrane proteins, it may be possible that NgBR

experiences a similar change. Stimulated internalization of NgBR's transmembrane domain with additional N-terminal extracellular residues could result in stabilization of NgBR's cytosolic domain. This may also lend explanation to why NgBR shares sequence identity to UPPs in residues other than its cytosolic region. Future experimental approaches including alteration of cellular membrane lipid content or ligand induced receptor stimulation, could be used to confirm if NgBR is able to exist in such a topological orientation. Altogether, our study opens the way for additional exploration into the structural properties of the nogo-b receptor and lends insight into a potential molecular basis of how it facilitates farnesylated Ras recruitment and promotes Ras activation.

CHAPTER 6 PROTEIN CRYSTALLIZATION: ELUDING THE BOTTLENECK OF X-RAY CRYSTALLOGRAPHY

Abstract

To date, X-ray crystallography remains the gold standard for the determination of macromolecular structure and protein substrate interactions. However, the unpredictability of obtaining a protein crystal remains the limiting factor and continues to be the bottleneck in determining protein structures. A vast amount of research has been conducted in order to circumvent this issue with limited success. No single method has proven to guarantee the crystallization of all proteins. However, techniques using antibody fragments, lipids, carrier proteins, and even mutagenesis of crystal contacts have been implemented to increase the odds of obtaining a crystal with adequate diffraction. In addition, we review a new technique using the scaffolding ability of PDZ domains to facilitate nucleation and crystal lattice formation. Although in its infancy, such technology may be a valuable asset and another method in the crystallography toolbox to further the chances of crystallizing problematic proteins.

A Brief History

Protein crystallization was observed more than 170 years ago by Friedrich Ludwig Hünefeld with the unintended crystallization of hemoglobin from earth worm blood. This accidental finding was described in his book *Der Chemismus in der thierischen Organisation* (Chemical Properties in the Animal Organization) in 1840 [136-138]. However, it was not until the late 19th century that scientists began to replicate the crystallization of proteins. Early protein crystallization attempts aimed for purification of proteins. Scientists such as Funke in 1851 purified hemoglobin from red blood cells by dilution of red blood cells with solvents followed by slow evaporation to produce hemoglobin crystals [137-139]. Sequentially, botanists such as Ritthausen

and Osborn implemented similar techniques in the 1880s through the 1890s to purify a series of plant seed proteins [137,138,140,141]. What was not realized at the time is that this accidental discovery would lend far more than the ability to isolate proteins from a sample but would become the foundation for the elucidation of high-resolution protein structure.

The investigation of molecular crystal structure dates as far back as 1611 when Johannes Kepler hypothesized the hexagonal crystal packing of snow in his work *Strena seu de Nive Sexangula* (A New Year's Gift of Hexagonal Snow) [142]. However, it was not until the X-ray was discovered by Wilhelm Röntgen in 1895 that would make it possible to validate any proposed crystal models. In 1912, Max von Laue discovered the diffraction of X-rays by crystals. During the period of 1912-1913 William Laurence Bragg developed Braggs Law which describes the angles for coherent and incoherent scattering from a crystal lattice [143]. It was soon after that Bragg reported the first X-ray crystal structure of sodium chloride.

With X-ray diffraction in its infancy, the initial pioneers of protein crystallography focused on highly abundant proteins that could be produced and purified easily. The first protein structure to be solved was that of myoglobin from the sperm whale in 1958 followed by lysozyme from chicken egg whites in 1965 [144,145]. However, as the field progressed, scientists began to direct their efforts to objective-oriented projects involving proteins with different molecular weights and from different sources. It was then realized that the bottleneck of protein structure determination is the production of protein crystals suitable for X-ray diffraction.

The Premise of Protein Crystallization

Protein crystallization today is achieved by the same basic principle as was discovered over 170 years ago. Supersaturation of a protein in solution is the basis behind the crystallization. At the supersaturated state, the amount of proteins in solution exceed their solubility limit. Under this

non-equilibrium state, the proteins are being pushed out of the solution undergoing a first ordered phase transition known as nucleation. Supersaturation of a protein in solution can be achieved by several different methods. Usually, a chemical known as precipitant is used to reduce protein solubility and create the supersaturation state. The phase diagram (Fig. 14A) demonstrates the dependence of increasing protein and precipitant concentration on the saturated state. At both low protein concentration and precipitant concentration, the protein remains in the stable, undersaturated state. As either protein or precipitant concentration is increased in solution, the protein can undergo a transition to either the metastable, labile, or precipitation phase [137,138,146]. In the metastable phase, nuclei may form, which are stable compared to the parent liquid phase and metastable compared to the crystalline phase of the protein [147]. The labile phase is where both nucleation and crystal growth may occur [148]. The precipitation phase is where the highest degree of supersaturation exists, in which ordered nucleation does not occur and there is no crystal growth. Thus, crystallization is dependent on the magnitude and rate at which supersaturation is achieved.

There are several methods that can achieve the ideal supersaturation state for nucleation and crystal growth. Most methods can be placed into one of three categories: vapor diffusion (VD), batch crystallization, or liquid-liquid diffusion [138,148]. Vapor diffusion is the most extensively used method that includes two different techniques: hanging-drop vapor diffusion (HD-VD) or sitting-drop vapor diffusion (SD-VD). In both techniques, the protein and precipitant are equilibrated against the crystallization reservoir solution separated by an air gap. Their difference is simply as each name implies. In the SD-VD method, the protein/precipitant mixture resides in a well sitting above the reservoir solution. Whereas in the HD-VD method, the protein/precipitant mixture is hanging over the reservoir solution from an inverted glass slide. In each setting, water vapor

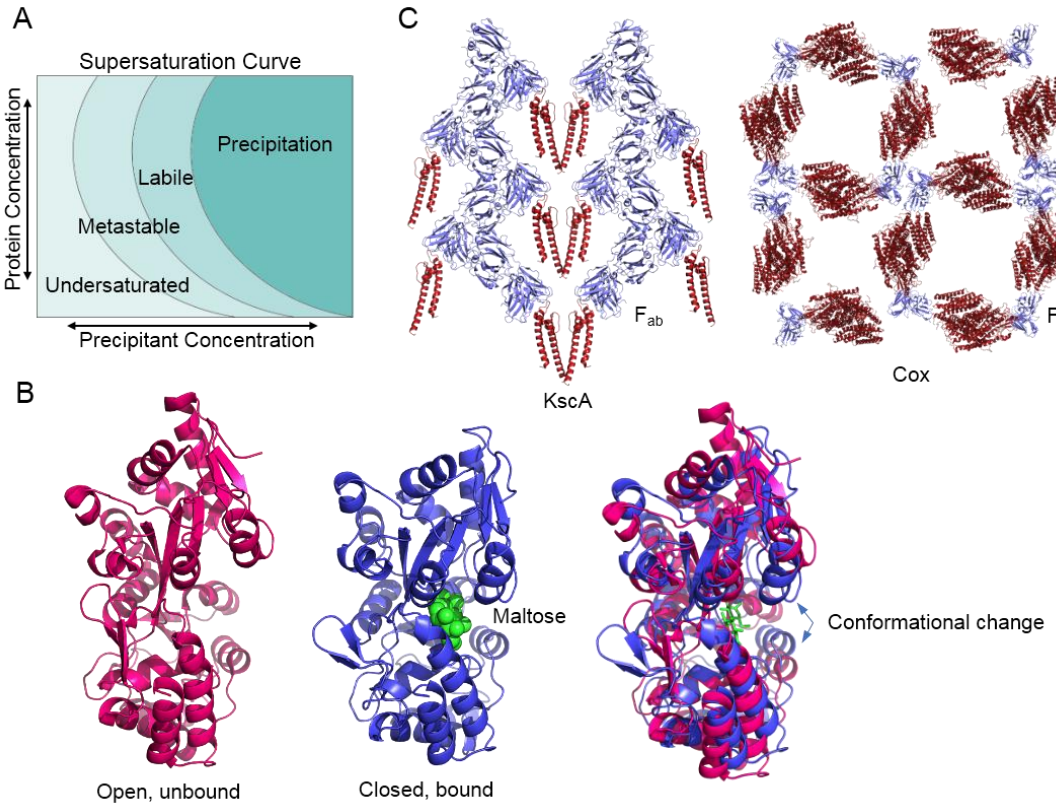

Figure 14. Carrier mediated protein crystallization. (A) The phase diagram for protein crystallization. (B) Conformational change of MBP upon binding of maltose. Unbound (left), bound (middle), and superposition of the unbound and bound forms (right). (C) Crystal lattice formation as mediated by antibody fragments. Left, the crystal lattice of KscA K_+ channel mediated by an F_{ab} fragment (PDB code 1K4C); right, the crystal lattice of COX mediated by a recombinant F_v fragment (PDB code 1QLE).

diffuses from the drop into the reservoir solution slowly concentrating the protein/precipitant mixture, promoting supersaturation and ideally nucleation of the protein.

Batch crystallization is a method in which both concentrated protein and precipitant are mixed together and covered by a layer of paraffin oil [137,138,146,149]. This technique can be used for very small volumes, often referred to microbatch crystallization with droplets as small as 1µl [137,150]. In batch crystallization, the crystallization conditions can be finely controlled due to the inability

of air to penetrate the oil layer. Airborne contamination and other variables are blocked from contacting with the sample reducing interference in protein crystallization [137,148].

Liquid-liquid diffusion (also known as counter diffusion) is a technique in which the protein and precipitant are injected on each side of a closed channel and gradually mixed through diffusion [138,148,151,152]. At the beginning of mixing, the two solutions come into contact at their maximum concentrations in a reagent chamber, resulting in supersaturation and promoting spontaneous nucleation. As the mixing proceeds, the mixture reaches equilibrium and the level of supersaturation is decreased, consequently favoring crystal growth. This method can be performed in a variety of configurations; for example, microfluidic devices have been developed using capillaries and microchips which now allow for *in situ* X-ray data collection [148,151].

Screening and Additives

In a typical crystallization experiment, thousands of conditions are often tested for a single protein in order to acquire a crystal suitable for X-ray diffraction. Variables that may affect crystallization include pH, temperature, and precipitant concentration. The pH is typically controlled by introducing a buffering agent into the crystallization condition. Buffering agents that are commonly used include Tris hydrochloride, HEPES, sodium cacodylate, and sodium acetate. Precipitants are among the most variable factors and can be divided into four different categories based on their properties: salts, organic solvents, long chain polymers, and low-molecular-weight polymers and nonvolatile organic compounds [138]. Common salts include ammonium sulfate or sodium chloride whereas common organic solvents include ethanol and isopropanol. The polyethylene glycol family (PEG) such as PEG3350 is representative of the third category whereas PEG1000 or less along with compounds such as methylpentanediol (MPD) are representative of the latter [138].

Additives can be classified as any foreign molecule introduced into the crystallization condition other than the aforementioned components. The purpose of adding additives is to facilitate or enhance crystal formation or growth. Examples may include small molecules, detergents, metal ions, or other various compounds. Additives do not necessarily promote supersaturation of the protein in solution but are intended to contribute to protein solubility or structural rigidity. These compounds can often perturb sample to sample and solvent to solvent interactions influencing the behavior of protein crystallization. There are several reports in which inclusion of additives resulted in improvement of both crystal size and quality [153-155]. However, screening numerous random molecules is tedious, and success is often limited. Thus, a more rational approach is the introduction of natural additives or compounds already found to interact with the protein of interest. These types of additives might include cofactors or ligands required for the biological activity of the protein. Such molecules not only facilitate successful crystallization but also provide functional insight into the protein by revealing the substrate or cofactor binding site [156-158].

Construct Optimization

Although supersaturation is the premise behind protein nucleation and crystallization, the protein itself can be a critical variable for the formation of a crystal and subsequent growth. It has been argued that the protein, rather than the crystallization condition, may be the most important variable in the crystallization process. Solubility and monodispersion of the protein is often necessary in successful crystallization experiments. Non-specific aggregation by hydrophobic amino acids or flexible protein regions can interfere with directional nucleation and overall crystal lattice formation. Therefore, protein construct optimization is often implemented in protein crystallography.

During the molecular biology boom of the 1980s and 1990s, proteins that had been previously understudied due to their low abundance in the cell could now be cloned, expressed, and purified in milligram quantities using bacterial expression systems [137,138]. However, the technology of molecular cloning would not only pave the way for the study of previously unobtainable proteins, but also would allow for manipulation of protein constructs to enhance X-ray crystallographic studies. Standard polymerase chain reaction (PCR) and recombinant DNA technology now allow for the deletion of protein regions that may interfere with crystallization. It is common practice in construct development for protein structural analysis to remove flexible amino acid sequences. These regions can be identified by a variety of techniques such as limited proteolytic cleavage followed by fragment analysis, orthological structure comparison, and multiple sequence alignment [159-161]. Removal of the flexible regions can reduce conformational heterogeneity of the protein and enhance ordered formation of the crystal lattice.

Surface Residue Modification

Besides removal of problematic amino acid sequences from the protein, mutagenesis of surface residues may also be implemented to enhance the formation of crystal contacts. One of the first successful examples of this strategy was that of human ferritin by Lawson in 1991 in which some surface residues were mutated to promote the crystal contacts analogous to the structure of the rat isoform [162,163]. Subsequent studies by other groups such as McElroy in 1992 with thymidylate synthase, Zhang in 1995 with T4 lysozyme, and Zhang in 1997 with leptin showed that mutagenesis of surface residues can greatly impact the formation of the crystal lattice [164-166].

Chemical modification of surface residues can also facilitate the formation of crystal contacts by reducing surface entropy of the protein. The most common approach has been the reductive methylation of primary amine groups by dimethylamine-borane in the presence of

formaldehyde [167]. Residues that are subject to such methylation include exposed lysine or arginine side chains and the N-terminal primary amine. This strategy offers some advantages over mutagenesis by eliminating the time-consuming process of protein production. Additionally, methylation is performed on the intact protein which prevents mutagenesis-induced improper folding of the nascent polypeptide [167,168]. Furthermore, only residues exposed on the surface of the protein will be modified, and those buried in the core or residues responsible for strong protein-protein interfaces are not affected. However, this method of non-specific surface modification may eliminate residues critical for substrate binding or other biologically relevant interactions.

Fusion Tags

Unfortunately, even with direct construct optimization and surface modification of the target protein, crystallization success is no guarantee. Optimized constructs may still experience solubility or aggregation issues due to the improper folding of the target protein with bacterial expression systems. To circumvent these issues, molecular cloning strategies are often used to attach a solubility tag to the target protein to promote protein folding and stability. This is often accomplished by cloning the target protein into a vector that contains a protein tag which is known to fold well and exhibit substantial expression and solubility. The most common solubility tags used in crystallography experiments include Small Ubiquitin-like Modifier (SUMO), Glutathione S-transferase (GST), Thioredoxin (TRX), avidin/streptavidin tags, and Maltose Binding Protein (MBP) [160]. Classically, once purified, these fusion tags are removed prior to crystallization using an engineered protease site in the linker region between the target protein and tag. In a sequential purification step, the tag and protease are separated from the target protein yielding the highly pure protein suitable for crystallization.

Incorporation of these large solubility tags has been shown to provide substantial benefits especially in bacterial expression systems. It has been estimated that nearly 50% of all overexpressed prokaryotic proteins have solubility issues using only a hexahistidine tag (His-tag) expression system [160,169,170]. Recombinant proteins of eukaryotic origin with a His-tag suffer even higher solubility issues [160,171-173]. Approaches that are often explored to resolve such issues include altering expression conditions such as temperature and induction strategy, and exploring alternative bacterial expression strains or eukaryotic expression systems [160,169]. However, the behavior of individual proteins can vary substantially, so that it is highly advantageous that a more universal strategy is implemented. Thus, the solubility tag approach has become widely adopted due to its noted success in protein structure elucidation.

Carrier Mediated Crystallography

As previously explained, large protein fusion tags are commonly used in structural biology for solubility enhancement and promoting proper folding of the target protein. It is common practice to remove these tags prior to crystal screening. This is because: (1) tagged proteins are less likely to form well-ordered diffracting crystals due to conformational heterogeneity resulting from the linker region; and (2) addition of a large fusion tag lends the possibility that the native structure of the target protein is changed or physiologically relevant interactions are altered. However, because the tags are often responsible for enhancing solubility and structural integrity, removal of them from the target protein can result in unwanted complications [160]. Common problems from tag removal include precipitation of the target protein and insufficient cleavage, both of which can result in reduced protein yield or poor quality of proteins. The alternative to such issues is to leave the protein tag attached for crystallization trials [160]. Although previously

thought to be undesirable, this practice, known as carrier mediated crystallography, is now being used to facilitate crystallization of proteins that have proven difficult to crystallize.

The concept of carrier mediated crystallography is that by leaving on the fusion tag, the tag not only promotes solubility of the target protein but also facilitates nucleation and crystal lattice formation by its mediated crystal contacts. This sequentially promotes the incorporation of the target protein into the crystal lattice which may not have been possible without the tag. An additional benefit of this technique is that the phase problem in X-ray crystallography can be easily solved by the molecular replacement method, since most commonly used fusion tags have previously solved structures. The structures of fusion tags also allow easy implementation of surface entropy reduction to further increase the chances of crystallization success.

Maltose Binding Protein

The first protein structure reported using a fusion tag approach was reported by Center in 1998 in which two fragments from the ectodomain of human T cell leukemia virus type 1 (HTLV-1) were crystallized with Maltose Binding Protein (MBP) as the tag [174]. Since then, it was described by Waugh in 2015 that over 100 crystal structures using MBP as fusion tag have been solved [175]. MBP is a 42.5kD *E. coli* protein responsible for the uptake of maltodextrin and promoting its catabolism. MBP exists as a monomer in solution and is divided into two distinct globular domains connected by three short polypeptide segments. The two globular domains are separated by a deep pocket that is responsible for the binding of its substrate maltose or other maltodextrins [176,177]. MBP can undergo a significant conformational change upon binding to its substrates (Fig. 14B). The substrate bound form displays a closed substrate binding pocket and is related to the substrate unbound form by a rigid motion of the two domains around the linking polypeptide hinge (Fig. 14B). The two MBP conformations can give rise to different crystal contacts. Thus, crystallization

of MBP fusion proteins is often screened with or without the addition of maltose in order to promote monodispersion of MBP's bound and unbound conformations. This was shown to be critical in at least one structure, (PDB code: 3WAI), in which only the ligand free form could be crystallized [175,178]. In other structures, different crystal forms were observed between the ligand bound and unbound forms [175,179].

The MBP conformational state is not the only factor that affects MBP mediated crystallization. The linker region between MBP and the target protein is also an important variable. Eleven MBP fusion structures as noted by Waugh exhibit relatively long linkers designed for proteolytic cleavage [175]. This suggests that crystallization with the MBP tag in these cases was a fall back approach when crystallization attempts of the cleaved protein failed. There is also a prevalent consensus of short linkers including N, NS, HM, GS, GSS, AMD, GSSGSS, and NSSS [175]. The linker NSSS is one of the common linkers deposited in the Protein Data Bank and is characteristic of the expression vector pMAL-c2 [175,180]. This linker is created by PCR amplification of the target protein sequence with introduction of an in-frame SacI restriction site at the N-terminus of the protein. However, the most common linker for MBP fusion constructs is that of NAAA which is the result of a three point mutations at the MBP C-terminus. These point mutations can be traced back to the crystal structure of HLTV-1 gp21 ectodomain fused to MBP [175,181]. It was anticipated that gp21 would exist as a homotrimer and thus three charged residues near the C-terminus of MBP were changed to alanine to avoid electrostatic repulsion. These three alanine residues code for a NotI restriction site which allows for the in-frame ligation of the target protein into the vector [175]. The asparagine that proceeds the three alanine residues is a cloning artifact introduced from the pMAL-c2 vector. Although there is no definitive answer to whether this linker

more likely facilitate MBP mediated crystallization, the predominance of solved structures with this linker sequence suggests a good starting point in pursuing MBP carrier crystallization.

To further the chances of obtaining a diffracting quality crystal, surface entropy reduction can be implemented together with the MBP carrier technique. In 2010, Moon describes a tandem fixed arm MBP/surface entropy reduction mutation system, in which both the linker region and MBP loop regions are optimized. Custom short linkers were designed to promote a "fixed arm" where the conformational flexibility of the linker would be minimized but not interfere with the structure of MBP or the protein itself [180]. Surface entropy reduction mutations were then introduced on the solvent exposed loop regions of MBP. Using this strategy, Moon was able to solve three previously unobtainable protein structures including 2-*O*-sulfotransferase (2OST) from *Gallus gallus*, receptor for activated C-kinase 1 (RACK1) from *Arabidopsis thaliana*, and Derp7 from *Dermatophagoides pteronyssinus*. Structural analysis showed that some of these surface entropy reduction mutations promoted crystal formation likely by reducing electrostatic repulsion in crystal contacts [180].

One major concern with the MBP mediated crystallization is potential interference of MBP with the native conformation of the target protein. Waugh in 2016 addressed this issue by investigation of 24 proteins with and without MBP as a fusion partner. Following analysis, it was found that the average r.m.s.d (root-mean-square-deviation) for the fused and non-fused structures averaged approximately 1 Å suggesting that MBP fusion caused little to no structural change [175]. Altogether, this suggests that the MBP mediated crystallization is a viable technique that can enhance the probability of crystallization success without influencing the native structure of the protein target.

Crystal Structures with Carriers other than MBP

Although the majority of crystal structures deposited in the PDB using the carrier approach utilize MBP as the fusion partner, a number of other structures exist using alternative protein tags. As described by Smyth in 2003, crystallization of GST fusion proteins have been reported for the DNA-binding domain of *Drosophila* DNA replication-related-element-binding factor and for mouse estrogen receptor hormone binding domain [160,182,183]. Carter and Schmidt in 2011 and 2012 revealed the crystal structures of the dynein motor domain of *Saccharomyces cerevisiae* with GST fusion [184,185]. In addition to GST, yeast SUMO was also reported as a fusion partner in crystallization. The SUMO tag is widely implemented in protein purification because of its high solubility; however, it is typically cleaved off prior to crystallization[186-188]. Regardless, at least nine unique SUMO fusion protein structures have been deposited in the PDB including C-terminal domain of Ebola virus VP30, thymidylate synthase, alpha-keto acid dehydrogenase phosphatase, and peptidyl-prolyl cis-trans isomerase. However, despite a limited amount of research on either GST or SUMO fusion, both tags may serve as a viable alternative for carrier mediated crystallography.

Antibodies as a Carrier Mediated Approach

An alternative carrier approach to the fusion protein is the use of antibody fragments in the facilitation of crystal formation. In this technique, crystal contacts are mediated between antibodies specifically bound to the protein of interest. One of the first uses of this technique was for the crystallization of the HIV capsid protein p24 where traditional crystallization methods were unsuccessful [189,190]. It was only after screening with antibody fragments that crystals suitable for X-ray diffraction were obtained. Antibodies can be divided into two regions that include the F_{ab} region and F_c regions. The F_{ab} region contains the sites that can bind to antigens, whereas the F_c region allows for the generation of an immune response [191]. The F_{ab} region is composed of one

variable domain and one constant domain from each the light chain and heavy chain of the antibody [192]. The variable domains are collectively known as the F_v region and is the most important region for binding antigens, constituting specificity and antigen discrimination.

When using antibodies as a carrier in crystallography, the most important part is selection and preparation of homogenous F_{ab} fragments specific for the protein target. The use of entire antibodies can hinder crystal lattice formation due to the flexibility between the F_{ab} and F_c [193,194]. Thus, the use of F_{ab} fragments is the most common approach in this technique. One approach that can be used to isolate the F_{ab} fragments is to subject the antibodies to papain digestion which cleaves the flexible domain between the F_{ab} and F_c regions [189]. The resulting F_{ab} fragments are then purified by ion exchange chromatography to remove the F_c region. Although cleaved F_{ab} fragments are relatively easy to obtain, care must be taken in the purification procedure to ensure homogeneity of the fragments. The more definitive approach to generating identical F_{ab} or even F_v regions is the use of recombinant methods that can ensure homogenous fragments for crystallization [193].

Once purified, the fragments are then mixed with the respective protein target and standard crystallization procedures are implemented. As shown by Hunte and Michel in 2002, the crystal contacts in the KcsA K^+ channel crystal are entirely mediated by the F_{ab} fragments, virtually suspending the target protein within the crystal lattice (Fig. 14C) [193,195]. The same was noted for the crystallization of cytochrome *c* oxidase (COX) where all crystal contacts are mediated by the bound F_v fragments (Fig. 14C) [193,196]. These observations suggest that the antibody fragments can provide substantial benefits in the mediation of protein crystallization. The high affinity of the antibody for its target eliminates the need for a linker as with the fusion method and reduces the conformational flexibility of the protein molecule. Additionally, as observed in the crystal

structures of KscA and COX, crystal lattice formation can be achieved purely by the antibody fragments eliminating the need for the manipulation of crystal contacts from the target molecule. However, even with recombinant methodologies, this approach can be very costly and labor intensive making this strategy often one of last resorts in protein crystallography.

Nanotechnology in Protein Crystallography

Regardless of the method used to facilitate crystallization of a protein, crystal formation is limited by the laws of chemistry and the insurmountable variables involved in macromolecular interactions. When simplified, the three key components to crystallization success is nucleation, conformational stability, and ordered protein-protein contacts. With every strategy, there are strengths and limitations. However, technological advancement continues to open alternate pathways to overcome this barrier. Nanotechnology and the use of nanoparticles has been extensively explored in recent years due to its wide range of practical application in physics, optics, electronics, and even medicine. Nanoparticles can be defined as an ordered cluster of atoms, typically inorganic materials, that have at least one dimension between 1 and 10 nanometers. They tend to be highly reactive and have been used for conjugation to a variety of molecules with applications in protein crystallography.

An example of such technology was described by Ko in 2016 in which nanoparticles served as an inducing reagent in protein nucleation and crystal growth [197]. Ko demonstrated that formation of lysozyme nucleation cores can be accelerated by decoration of gold nanoparticles with -COOH and Ni^{2+} ions. The interactions between lysozyme and the immobilized -COOH and Ni^{2+}ions can readily conjugate lysozyme to the nanoparticles creating the nucleation core [197]. Manipulation of nanoparticle size and shape increased the number of successful crystallization conditions by 24%. Chen in 2017 described another approach for nucleation induction using nanodiamond (ND)

carbon based particles. Like the gold particles described by Ko, NDs were modified with various oxygen containing groups including -COOH, -COH, and -C=O for protein conjugation [198]. Chen reported that the nanoparticles were able to increase crystallization efficiency of several proteins including lysozyme, ribonuclease A, proteinase K, and catalase [198]. NDs were also able to effectively induce crystallization of lysozyme at concentrations as low as 5 mg/mL. This finding indicates that the nanoparticles can facilitate the crystallization of proteins at low concentration or at ultra-low supersaturation.

In addition to the nanoparticles, several other technological innovations are being explored as potential facilitators of crystallization. Crystallization mediated by porous materials such as silicon has been examined. Microfluidic devices have also been developed to increase protein crystallization efficiency [199,200]. Crystallization of membrane proteins using lipidic bicelles have been reported with variable success [201-203]. Although no single technique has been successful for the crystallization of all proteins, each method can be implemented as a potential option for successful protein crystallization.

PDZ Domains in Nucleation and Crystal Facilitation

With efforts to expand the crystallization toolbox, our laboratory has recently begun to develop additional approaches to facilitate protein nucleation and crystal formation. One of our approaches was designed to simulate biological protein scaffolding in which protein-protein interaction known to mediate protein complex formation was exploited as a carrier for crystallization. Because the essence of nucleation is ordered protein contacts which results in the formation of a crystal lattice, the use of scaffolding proteins as a carrier may increase the chance of well-ordered interactions. The following will describe the potential of using the scaffolding properties of PDZ domains in facilitation of crystal lattice formation.

PDZ domains are composed of 80-90 amino acids that play critical roles in protein scaffolding and complex assembly at the cellular membrane. The acronym PDZ is derived from the first three proteins originally found to contain this domain. These three proteins are known as Postsynaptic density protein 95kD (PSD95), Drosophila disc large tumor suppressor (Dig1), and Zonula occludens-1 protein (ZO-1) [22,45]. PDZ domains bind to their targets by recognizing specific short C-terminal amino acid motifs of their targets. This structurally conserved interaction promotes scaffolding of protein complexes and is important for the assembly of signaling complexes, protein trafficking, and the recycling of cell receptors [23]. PDZ domains can be classified as either class I or class II based on their substrate specificity. Class I PDZ domains recognize the C-terminal peptide consensus sequence (S/T)X(V/I/L) (X denoting any amino acid) and class II recognizes (F/Y)X(F/V/A) [52,80]. All PDZ domains share an evolutionary conserved fold consisting of six β-strands (β1-β6) and two α-helical segments (αA and αB). A similar peptide recognition mode is shared among all PDZ domains with the target peptide inserted between the strand β2 and helix αB [1,2,43,45,52,80,116]. Many PDZ proteins increase their scaffolding capability through dimerization and promote formation of large macromolecular complexes [43,54-56].

Numerous PDZ-substrate complexes have been crystallized. One unique and effective strategy in these crystallizations is the use of a chimeric protein construct, i.e. the peptide ligand attached to the C-terminus of the PDZ molecule (Figure 15A) [1,2,43,116]. In the crystal, the chimeric protein displays a polymeric arrangement with the C-terminal ligand sequence bound to a neighboring PDZ, leading to the formation of a linear filament throughout the crystal (Fig. 15B). This repeated “pocket and tail” interaction appears to facilitate directional nucleation contributing to crystal contact formation between adjacent PDZ molecules. This strategy has proven essential

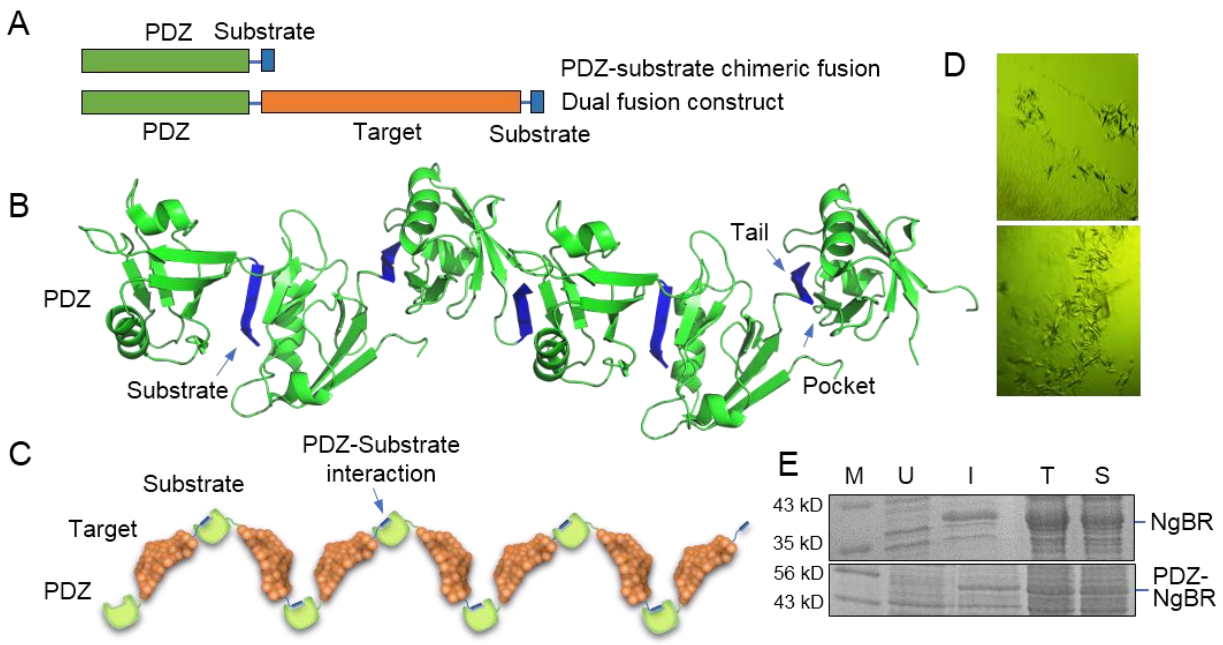

Figure 15. PDZ scaffold mediated protein crystallization. (A) A chimeric PDZ-substrate fusion construct (top) and a dual protein fusion construct with a target protein sandwiched by a PDZ domain and PDZ substrate peptide (bottom). (B) Crystal contacts mediated by the repeated "pocket and tail" interactions in the crystal of NHERF1 PDZ1-CXCR2 fusion protein (PDB code 4JL7). (C) Theoretical representation of crystal contact formation in PDZ scaffold mediated protein crystallization. (D) Crystals obtained for NHERF1/PDZ1-NgBR-CXCR2 (top) and NHERF1/PDZ1-SMYD5-CXCR2 (bottom). (E) Expression and solubility assessment of NgBR with and without N-terminal PDZ fusion. Lane M, molecular weight marker; U, uninduced cell culture; I, induced cell culture; T, total cell lysate; S, supernatant of cell lysate.

for the crystallization of NHERF1 PDZ1 which was not able to be crystallized without the peptide substrate fusion [2].

Inspired by this effective chimeric protein approach, we designed a dual fusion construct with the target protein sandwiched by a PDZ domain and its respective peptide ligand (Fig. 15A). The goal of building this construct is to explore the potential of specific PDZ-ligand interaction in promoting the crystallization of other proteins. Since the crystallization of PDZ domains is facilitated by the attachment of the C-terminal peptide substrate, our concept is that adding a PDZ molecule to the N-terminus of the target protein and a specific PDZ substrate peptide to the protein C-terminus may facilitate nucleation (Fig. 15C). The N-terminally fused PDZ domain would

recognize the C-terminally fused peptide substrate from another fusion protein, creating a chain of interactions that would facilitate crystal lattice formation. As with antibody mediated crystallization, the target protein could be virtually suspended within the crystal lattice reducing the necessity for the formation of crystal contacts from the target protein itself. Although in its infancy, our laboratory has employed this strategy for two proteins, NgBR (Nogo-B receptor) and SMYD5 (SET and MYND domain containing protein 5), for which standard crystallization methods have been unsuccessful. Interestingly, small crystals were obtained for both proteins in very similar conditions suggesting that the crystallization was mediated by PDZ-ligand interaction (Fig. 15D). In addition, PDZ domains are very soluble and stable in solution. Using them as a fusion partner could also increase the solubility of the target protein. This was verified in one instance when comparing the expression profile of NgBR with and without the N-terminally attached NHERF1 PDZ1 domain (Fig. 15E). Thus, our dual fusion protein approach not only provides a molecular scaffold for crystallization but also is able to increase protein solubility. Furthermore, optimization of this strategy may prove as an additional approach for crystallization of problematic proteins.

Concluding Remarks

X-ray crystallography continues to be the leading method for the elucidation of protein structure and rational drug design. However, the unpredictability of protein crystallization can significantly suppress the rate at which such discoveries are made. Although no crystallization method has guaranteed success, numerous strategies have been employed in order to increase the probability of which it can occur. Adjustment of crystallization components, modification of the protein construct, addition of carrier molecules, or even synthetic materials can be used alone or in combination to increase the odds at which the target protein can be crystallized. Additionally,

utilization of the natural PDZ scaffolding ability may be implicated as an additional strategy for the induction of nucleation as well as facilitating the formation of crystal contacts. Together, all the strategies reviewed here are viable approaches which may help evade the bottleneck of crystallography and advance the analysis of protein structures.

REFERENCES

1 Holcomb, J. *et al.* Structural insights into PDZ-mediated interaction of NHERF2 and LPA(2), a cellular event implicated in CFTR channel regulation. *Biochemical and biophysical research communications* **446**, 399-403, doi:10.1016/j.bbrc.2014.02.128 (2014).

2 Lu, G. *et al.* Structural insights into neutrophilic migration revealed by the crystal structure of the chemokine receptor CXCR2 in complex with the first PDZ domain of NHERF1. *PloS one* **8**, e76219, doi:10.1371/journal.pone.0076219 (2013).

3 Karthikeyan, S., Leung, T. & Ladias, J. A. Structural basis of the Na+/H+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. *The Journal of biological chemistry* **276**, 19683-19686, doi:10.1074/jbc.C100154200 (2001).

4 Karthikeyan, S., Leung, T. & Ladias, J. A. Structural determinants of the Na+/H+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors. *The Journal of biological chemistry* **277**, 18973-18978, doi:10.1074/jbc.M201507200 (2002).

5 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948, doi:10.1093/bioinformatics/btm404 (2007).

6 Garbett, D. & Bretscher, A. The surprising dynamics of scaffolding proteins. *Molecular biology of the cell* **25**, 2315-2319, doi:10.1091/mbc.E14-04-0878 (2014).

7 Romero, G., von Zastrow, M. & Friedman, P. A. Role of PDZ proteins in regulating trafficking, signaling, and function of GPCRs: means, motif, and opportunity. *Advances*

in pharmacology (San Diego, Calif.) **62**, 279-314, doi:10.1016/b978-0-12-385952-5.00003-8 (2011).

8 Lauffer, B. E. *et al.* SNX27 mediates PDZ-directed sorting from endosomes to the plasma membrane. *The Journal of cell biology* **190**, 565-574, doi:10.1083/jcb.201004060 (2010).

9 Thompson, B. J. Cell polarity: models and mechanisms from yeast, worms and flies. *Development (Cambridge, England)* **140**, 13-21, doi:10.1242/dev.083634 (2013).

10 Gumbiner, B. M. Regulation of cadherin adhesive activity. *The Journal of cell biology* **148**, 399-404 (2000).

11 Walther, C. & Ferguson, S. S. Minireview: Role of intracellular scaffolding proteins in the regulation of endocrine G protein-coupled receptor signaling. *Molecular endocrinology (Baltimore, Md.)* **29**, 814-830, doi:10.1210/me.2015-1091 (2015).

12 Moore, C. A., Milano, S. K. & Benovic, J. L. Regulation of receptor trafficking by GRKs and arrestins. *Annual review of physiology* **69**, 451-482, doi:10.1146/annurev.physiol.69.022405.154712 (2007).

13 Lefkowitz, R. J. & Shenoy, S. K. Transduction of receptor signals by beta-arrestins. *Science (New York, N.Y.)* **308**, 512-517, doi:10.1126/science.1109237 (2005).

14 Gurevich, V. V. & Gurevich, E. V. Extensive shape shifting underlies functional versatility of arrestins. *Current opinion in cell biology* **27**, 1-9, doi:10.1016/j.ceb.2013.10.007 (2014).

15 Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561-567, doi:10.1038/nature14656 (2015).

16 Smith, J. S. & Rajagopal, S. The beta-Arrestins: Multifunctional Regulators of G Protein-coupled Receptors. *The Journal of biological chemistry* **291**, 8969-8977, doi:10.1074/jbc.R115.713313 (2016).

17 Han, M., Gurevich, V. V., Vishnivetskiy, S. A., Sigler, P. B. & Schubert, C. Crystal structure of beta-arrestin at 1.9 A: possible mechanism of receptor binding and membrane Translocation. *Structure* **9**, 869-880 (2001).

18 Sutton, R. B. *et al.* Crystal structure of cone arrestin at 2.3A: evolution of receptor specificity. *Journal of molecular biology* **354**, 1069-1080, doi:10.1016/j.jmb.2005.10.023 (2005).

19 Shih, M., Lin, F., Scott, J. D., Wang, H. Y. & Malbon, C. C. Dynamic complexes of beta2-adrenergic receptors with protein kinases and phosphatases and the role of gravin. *The Journal of biological chemistry* **274**, 1588-1595 (1999).

20 Ali, M. S., Sayeski, P. P. & Bernstein, K. E. Jak2 acts as both a STAT1 kinase and as a molecular bridge linking STAT1 to the angiotensin II AT1 receptor. *The Journal of biological chemistry* **275**, 15586-15593, doi:10.1074/jbc.M908931199 (2000).

21 Cohen, B. D., Nechamen, C. A. & Dias, J. A. Human follitropin receptor (FSHR) interacts with the adapter protein 14-3-3tau. *Molecular and cellular endocrinology* **220**, 1-7, doi:10.1016/j.mce.2004.04.012 (2004).

22 Kennedy, M. B. Origin of PDZ (DHR, GLGF) domains. *Trends in biochemical sciences* **20**, 350 (1995).

23 Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *Journal of cell science* **114**, 3219-3231 (2001).

24 Cowburn, D. Peptide recognition by PTB and PDZ domains. *Current opinion in structural biology* **7**, 835-838 (1997).

25 Lee, H. W., Ko, J. & Kim, E. Analysis of PDZ domain interactions using yeast two-hybrid and coimmunoprecipitation assays. *Methods in molecular biology (Clifton, N.J.)* **332** (2006).

26 Niethammer, M. *et al.* CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90. *Neuron* **20**, 693-707 (1998).

27 Kim, E., Niethammer, M., Rothschild, A., Jan, Y. N. & Sheng, M. Clustering of Shaker-type K+ channels by interaction with a family of membrane-associated guanylate kinases. *Nature* **378**, 85-88, doi:10.1038/378085a0 (1995).

28 Dong, H. *et al.* GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors. *Nature* **386**, doi:10.1038/386279a0 (1997).

29 Perkinton, M. S., Sihra, T. S. & Williams, R. J. Ca(2+)-permeable AMPA receptors induce phosphorylation of cAMP response element-binding protein through a phosphatidylinositol 3-kinase-dependent stimulation of the mitogen-activated protein kinase signaling cascade in neurons. *J Neurosci* **19**, 5861-5874 (1999).

30 Shiraishi-Yamaguchi, Y. & Furuichi, T. The Homer family proteins. *Genome biology* **8**, 206, doi:10.1186/gb-2007-8-2-206 (2007).

31 Ranganathan, R. & Ross, E. M. PDZ domain proteins: scaffolds for signaling complexes. *Curr Biol* **7**, R770-773 (1997).

32 Brakeman, P. R. *et al.* Homer: a protein that selectively binds metabotropic glutamate receptors. *Nature* **386**, 284-288, doi:10.1038/386284a0 (1997).

33 Seidler, U. *et al.* The role of the NHERF family of PDZ scaffolding proteins in the regulation of salt and water transport. *Ann N Y Acad Sci* **1165**, doi:10.1111/j.1749-6632.2009.04046.x (2009).

34 Voltz, J. W., Weinman, E. J. & Shenolikar, S. Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation. *Oncogene* **20**, 6309-6314, doi:10.1038/sj.onc.1204774 (2001).

35 Weinman, E. J., Steplock, D. & Shenolikar, S. CAMP-mediated inhibition of the renal brush border membrane Na+-H+ exchanger requires a dissociable phosphoprotein cofactor. *J Clin Invest* **92**, 1781-1786, doi:10.1172/jci116767 (1993).

36 Lamprecht, G., Weinman, E. J. & Yun, C. H. The role of NHERF and E3KARP in the cAMP-mediated inhibition of NHE3. *The Journal of biological chemistry* **273**, 29972-29978 (1998).

37 Naren, A. P. *et al.* A macromolecular complex of beta 2 adrenergic receptor, CFTR, and ezrin/radixin/moesin-binding phosphoprotein 50 is regulated by PKA. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 342-346, doi:10.1073/pnas.0135434100 (2003).

38 Cinar, A. Differential effects of PDZ-adapter protein NHERF1, E3KARP and PDZK1 knockout on the regulation of NHE3 transport activity in native murine colonic epithelium [abstract]. *Gastroenterology* **130**, A99 (2006).

39 Broere, N. *et al.* Cystic fibrosis transmembrane conductance regulator activation is reduced in the small intestine of Na+/H+ exchanger 3 regulatory factor 1 (NHERF-1)-but Not NHERF-2-deficient mice. *The Journal of biological chemistry* **282**, 37575-37584, doi:10.1074/jbc.M704878200 (2007).

40 Favia, M. *et al.* NHE3 inhibits PKA-dependent functional expression of CFTR by NHERF2 PDZ interactions. *Biochemical and biophysical research communications* **347**, 452-459, doi:10.1016/j.bbrc.2006.06.112 (2006).

41 Li, C. *et al.* Lysophosphatidic acid inhibits cholera toxin-induced secretory diarrhea through CFTR-dependent protein interactions. *The Journal of experimental medicine* **202**, 975-986, doi:10.1084/jem.20050421 (2005).

42 Wu, Y. *et al.* A chemokine receptor CXCR2 macromolecular complex regulates neutrophil functions in inflammatory diseases. *The Journal of biological chemistry* **287**, 5744-5755, doi:10.1074/jbc.M111.315762 (2012).

43 Holcomb, J. *et al.* Crystal structure of the NHERF1 PDZ2 domain in complex with the chemokine receptor CXCR2 reveals probable modes of PDZ2 dimerization. *Biochemical and biophysical research communications* **448**, 169-174, doi:10.1016/j.bbrc.2014.04.085 (2014).

44 Li, C. & Naren, A. P. CFTR Chloride Channel in the Apical Compartments: Spatiotemporal Coupling to its Interacting Partners. *Integrative biology : quantitative biosciences from nano to macro* **2**, 161-177, doi:10.1039/b924455g (2010).

45 Holcomb, J. *et al.* PDZ Structure and Implication in Selective Drug Design against Cystic Fibrosis. *Current drug targets* **16**, 945-950 (2015).

46 Chapman, R. W. *et al.* CXCR2 antagonists for the treatment of pulmonary disease. *Pharmacology & Therapeutics* **121**, 55-68, doi:http://dx.doi.org/10.1016/j.pharmthera.2008.10.005 (2009).

47 Tsai, H.-H. *et al.* The Chemokine Receptor CXCR2 Controls Positioning of Oligodendrocyte Precursors in Developing Spinal Cord by Arresting Their Migration. *Cell* **110**, 373-383, doi:10.1016/S0092-8674(02)00838-3.

48 Gonsiorek, W. *et al.* Pharmacological Characterization of Sch527123, a Potent Allosteric CXCR1/CXCR2 Antagonist. *Journal of Pharmacology and Experimental Therapeutics* **322**, 477 (2007).

49 Wang, S. *et al.* CXCR2 Macromolecular Complex in Pancreatic Cancer: A Potential Therapeutic Target in Tumor Growth. *Translational Oncology* **6**, 216-225 (2013).

50 Magalhaes, A. C., Dunn, H. & Ferguson, S. S. G. Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins. *British Journal of Pharmacology* **165**, 1717-1736, doi:10.1111/j.1476-5381.2011.01552.x (2012).

51 Neel, N. F. *et al.* VASP is a CXCR2-interacting protein that regulates CXCR2-mediated polarization and chemotaxis. *Journal of cell science* **122**, 1882 (2009).

52 Sheng, M. & Sala, C. PDZ domains and the organization of supramolecular complexes. *Annual review of neuroscience* **24**, 1-29, doi:10.1146/annurev.neuro.24.1.1 (2001).

53 Ardura, J. A. & Friedman, P. A. Regulation of G Protein-Coupled Receptor Function by Na⁺/H⁺ Exchange Regulatory Factors. *Pharmacological Reviews* **63**, 882 (2011).

54 Fouassier, L., Yun, C. C., Fitz, J. G. & Doctor, R. B. Evidence for ezrin-radixin-moesin-binding phosphoprotein 50 (EBP50) self-association through PDZ-PDZ interactions. *The Journal of biological chemistry* **275**, 25039-25045, doi:10.1074/jbc.C000092200 (2000).

55 Umeda, K. *et al.* ZO-1 and ZO-2 Independently Determine Where Claudins Are Polymerized in Tight-Junction Strand Formation. *Cell* **126**, 741-754, doi:http://dx.doi.org/10.1016/j.cell.2006.06.043 (2006).

56 Chang, Bryan H. *et al.* A Systematic Family-wide Investigation Reveals that ~30% of Mammalian PDZ Domains Engage in PDZ-PDZ Interactions. *Chemistry & Biology* **18**, 1143-1152, doi:http://dx.doi.org/10.1016/j.chembiol.2011.06.013 (2011).

57 Sun, C. & Mierke, D. F. Characterization of interactions of Na+/H+ exchanger regulatory factor-1 with the parathyroid hormone receptor and phospholipase C. *The journal of peptide research : official journal of the American Peptide Society* **65**, 411-417, doi:10.1111/j.1399-3011.2005.00240.x (2005).

58 Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125-132, doi:doi:10.1107/S0907444909047337 (2010).

59 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674, doi:doi:10.1107/S0021889807021206 (2007).

60 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132, doi:doi:10.1107/S0907444904019158 (2004).

61 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213-221, doi:doi:10.1107/S0907444909052925 (2010).

62 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12-21, doi:doi:10.1107/S0907444909042073 (2010).

63 Runyon, S. T. *et al.* Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. *Protein science : a publication of the Protein Society* **16**, doi:10.1110/ps.073049407 (2007).

64 Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B. & Li, M. Peptide binding consensus of the NHE-RF-PDZ1 domain matches the C-terminal sequence of cystic fibrosis transmembrane conductance regulator (CFTR). *FEBS Lett* **427**, 103-108 (1998).

65 Karim, Z. *et al.* NHERF1 Mutations and Responsiveness of Renal Parathyroid Hormone. *New England Journal of Medicine* **359**, 1128-1135, doi:10.1056/NEJMoa0802836 (2008).

66 Wang, B. *et al.* Ezrin-anchored protein kinase A coordinates phosphorylation-dependent disassembly of a NHERF1 ternary complex to regulate hormone-sensitive phosphate transport. *The Journal of biological chemistry* **287**, 24148-24163, doi:10.1074/jbc.M112.369405 (2012).

67 Chen, J., Pan, L., Wei, Z., Zhao, Y. & Zhang, M. Domain-swapped dimerization of ZO-1 PDZ2 generates specific and regulatory connexin43-binding sites. *EMBO J* **27**, doi:10.1038/emboj.2008.138 (2008).

68 Im, Y. J. *et al.* Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. *The Journal of biological chemistry* **278**, doi:10.1074/jbc.M212263200 (2003).

69 Im, Y. J. *et al.* Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. *The Journal of biological chemistry* **278**, doi:10.1074/jbc.M306919200 (2003).

70 Fanning, A. S., Lye, M. F., Anderson, J. M. & Arnon, L. in *The Journal of biological chemistry* (2007).

71 Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science (New York, N.Y.)* **245**, 1066-1073 (1989).

72 Anderson, M. P. *et al.* Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science (New York, N.Y.)* **253**, 202-205 (1991).

73 Ballard, S. T. & Spadafora, D. Fluid secretion by submucosal glands of the tracheobronchial airways. *Respiratory physiology & neurobiology* **159**, 271-277, doi:10.1016/j.resp.2007.06.017 (2007).

74 Riordan, J. R. CFTR function and prospects for therapy. *Annual review of biochemistry* **77**, 701-726, doi:10.1146/annurev.biochem.75.103004.142532 (2008).

75 Li, C. & Naren, A. P. Macromolecular complexes of cystic fibrosis transmembrane conductance regulator and its interacting partners. *Pharmacol Ther* **108**, 208-223, doi:10.1016/j.pharmthera.2005.04.004 (2005).

76 Penmatsa, H. *et al.* Compartmentalized cyclic adenosine 3',5'-monophosphate at the plasma membrane clusters PDE3A and cystic fibrosis transmembrane conductance regulator into microdomains. *Molecular biology of the cell* **21**, 1097-1110, doi:10.1091/mbc.E09-08-0655 (2010).

77 Lin, F. T. & Lai, Y. J. Regulation of the LPA2 receptor signaling through the carboxyl-terminal tail-mediated protein-protein interactions. *Biochim Biophys Acta* **1781**, 558-562, doi:10.1016/j.bbalip.2008.04.013 (2008).

78 Shenolikar, S. & Weinman, E. J. NHERF: targeting and trafficking membrane proteins. *Am J Physiol Renal Physiol* **280**, F389-395 (2001).

79 Zhang, W. *et al.* Functional regulation of cystic fibrosis transmembrane conductance regulator-containing macromolecular complexes: a small-molecule inhibitor approach. *Biochem J* **435**, 451-462, doi:10.1042/bj20101725 (2011).

80 Lee, H. J. & Zheng, J. J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell communication and signaling : CCS* **8**, 8, doi:10.1186/1478-811x-8-8 (2010).

81 Tonikian, R. *et al.* A specificity map for the PDZ domain family. *PLoS Biol* **6**, doi:10.1371/journal.pbio.0060239 (2008).

82 Zhang, W., Fujii, N. & Naren, A. P. Recent advances and new perspectives in targeting CFTR for therapy of cystic fibrosis and enterotoxin-induced secretory diarrheas. *Future medicinal chemistry* **4**, 329-345, doi:10.4155/fmc.12.1 (2012).

83 Thevenet, L. *et al.* NHERF2/SIP-1 interacts with mouse SRY via a different mechanism than human SRY. *The Journal of biological chemistry* **280**, 38625-38630, doi:10.1074/jbc.M504127200 (2005).

84 Oh, Y. S. *et al.* NHERF2 specifically interacts with LPA2 receptor and defines the specificity and efficiency of receptor-mediated phospholipase C-beta3 activation. *Mol Cell Biol* **24**, 5069-5079, doi:10.1128/mcb.24.11.5069-5079.2004 (2004).

85 He, P. *et al.* Serum- and glucocorticoid-induced kinase 3 in recycling endosomes mediates acute activation of Na+/H+ exchanger NHE3 by glucocorticoids. *Molecular biology of the cell* **22**, 3812-3825, doi:10.1091/mbc.E11-04-0328 (2011).

86 Padanyi, R. *et al.* Apical scaffolding protein NHERF2 modulates the localization of alternatively spliced plasma membrane Ca2+ pump 2B variants in polarized epithelial

cells. *The Journal of biological chemistry* **285**, 31704-31712, doi:10.1074/jbc.M110.164137 (2010).

87 Jiang, Y. *et al.* New conformational state of NHERF1-CXCR2 signaling complex captured by crystal lattice trapping. *PloS one* **8**, e81904, doi:10.1371/journal.pone.0081904 (2013).

88 Gadsby, D. C., Vergani, P. & Csanady, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477-483 (2006).

89 Harris, A. Cystic fibrosis gene. *British medical bulletin* **48**, 738-753 (1992).

90 Flume, P. A. *et al.* Cystic fibrosis pulmonary guidelines: pulmonary complications: hemoptysis and pneumothorax. *American journal of respiratory and critical care medicine* **182**, 298-306, doi:10.1164/rccm.201002-0157CI

10.1164/rccm.201002-0157OC (2010).

91 Saiman, L. Microbiology of early CF lung disease. *Paediatric respiratory reviews* **5 Suppl A**, S367-369 (2004).

92 Rowe, S. M. & Verkman, A. S. Cystic fibrosis transmembrane regulator correctors and potentiators. *Cold Spring Harbor perspectives in medicine* **3**, doi:10.1101/cshperspect.a009761 (2013).

93 Radpour, R., Gourabi, H., Dizaj, A. V., Holzgreve, W. & Zhong, X. Y. Genetic investigations of CFTR mutations in congenital absence of vas deferens, uterus, and vagina as a cause of infertility. *Journal of andrology* **29**, 506-513, doi:10.2164/jandrol.108.005074 (2008).

94 Childers, M., Eckel, G., Himmel, A. & Caldwell, J. A new model of cystic fibrosis pathology: lack of transport of glutathione and its thiocyanate conjugates. *Medical hypotheses* **68**, 101-112, doi:10.1016/j.mehy.2006.06.020 (2007).

95 Van Goor, F. *et al.* Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18843-18848, doi:10.1073/pnas.1105787108 (2011).

96 Thomas, P. J., Ko, Y. H. & Pedersen, P. L. Altered protein folding may be the molecular basis of most cases of cystic fibrosis. *FEBS Lett* **312**, 7-9 (1992).

97 Du, K., Sharma, M. & Lukacs, G. L. The DeltaF508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. *Nature structural & molecular biology* **12**, 17-25, doi:10.1038/nsmb882 (2005).

98 Rowe, S. M., Miller, S. & Sorscher, E. J. Cystic fibrosis. *The New England journal of medicine* **352**, 1992-2001, doi:10.1056/NEJMra043184 (2005).

99 Bompadre, S. G., Sohma, Y., Li, M. & Hwang, T. C. G551D and G1349D, two CF-associated mutations in the signature sequences of CFTR, exhibit distinct gating defects. *The Journal of general physiology* **129**, 285-298, doi:10.1085/jgp.200609667 (2007).

100 Rosenstein, B. J. & Cutting, G. R. The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel. *The Journal of pediatrics* **132**, 589-595 (1998).

101 Davis, P. B., Drumm, M. & Konstan, M. W. Cystic fibrosis. *American journal of respiratory and critical care medicine* **154**, 1229-1256, doi:10.1164/ajrccm.154.5.8912731 (1996).

102 MacKenzie, T. *et al.* Longevity of patients with cystic fibrosis in 2000 to 2010 and beyond: survival analysis of the Cystic Fibrosis Foundation patient registry. *Annals of internal medicine* **161**, 233-241, doi:10.7326/m13-0636 (2014).

103 Stern, M. *et al.* The effect of mucolytic agents on gene transfer across a CF sputum barrier in vitro. *Gene therapy* **5**, 91-98, doi:10.1038/sj.gt.3300556 (1998).

104 Nafee, N. *et al.* Antibiotic-free nanotherapeutics: ultra-small, mucus-penetrating solid lipid nanoparticles enhance the pulmonary delivery and anti-virulence efficacy of novel quorum sensing inhibitors. *Journal of controlled release : official journal of the Controlled Release Society* **192**, 131-140, doi:10.1016/j.jconrel.2014.06.055 (2014).

105 Couetil, J. P. *et al.* Combined lung and liver transplantation in patients with cystic fibrosis. A 4 1/2-year experience. *The Journal of thoracic and cardiovascular surgery* **110**, 1415-1422; discussion 1422-1413 (1995).

106 Wong, J. Y. *et al.* Successful lung transplant in a child with cystic fibrosis and persistent Blastobotrys rhaffinosifermentans infection. *Pediatric transplantation* **18**, E169-173, doi:10.1111/petr.12294 (2014).

107 Bell, S. C., De Boeck, K. & Amaral, M. D. New pharmacological approaches for cystic fibrosis: promises, progress, pitfalls. *Pharmacol Ther* **145**, 19-34, doi:10.1016/j.pharmthera.2014.06.005 (2015).

108 Guggino, W. B. & Stanton, B. A. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nature reviews. Molecular cell biology* **7**, 426-436, doi:10.1038/nrm1949 (2006).

109 Li, C. *et al.* Spatiotemporal coupling of cAMP transporter to CFTR chloride channel function in the gut epithelia. *Cell* **131**, 940-951, doi:10.1016/j.cell.2007.09.037 (2007).

110 Shenolikar, S., Voltz, J. W., Cunningham, R. & Weinman, E. J. Regulation of ion transport by the NHERF family of PDZ proteins. *Physiology (Bethesda, Md.)* **19**, 362-369, doi:10.1152/physiol.00020.2004 (2004).

111 Weinman, E. J., Minkoff, C. & Shenolikar, S. Signal complex regulation of renal transport proteins: NHERF and regulation of NHE3 by PKA. *Am J Physiol Renal Physiol* **279**, F393-399 (2000).

112 Cushing, P. R., Fellows, A., Villone, D., Boisguerin, P. & Madden, D. R. The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling. *Biochemistry* **47**, doi:10.1021/bi8003928 (2008).

113 E, S. *et al.* Lysophosphatidic acid 2 receptor-mediated supramolecular complex formation regulates its antiapoptotic effect. *The Journal of biological chemistry* **284**, 14558-14571, doi:10.1074/jbc.M900185200 (2009).

114 Donowitz, M. *et al.* NHERF family and NHE3 regulation. *The Journal of physiology* **567**, 3-11, doi:10.1113/jphysiol.2005.090399 (2005).

115 Addison, C. L. *et al.* The CXC chemokine receptor 2, CXCR2, is the putative receptor for ELR+ CXC chemokine-induced angiogenic activity. *Journal of immunology (Baltimore, Md. : 1950)* **165**, 5269-5277 (2000).

116 Jiang, Y. *et al.* Crystallographic analysis of NHERF1-PLCbeta3 interaction provides structural basis for CXCR2 signaling in pancreatic cancer. *Biochemical and biophysical research communications* **446**, 638-643, doi:10.1016/j.bbrc.2014.03.028 (2014).

117 Ponting, C. P. Evidence for PDZ domains in bacteria, yeast, and plants. *Protein science : a publication of the Protein Society* **6**, doi:10.1002/pro.5560060225 (1997).

118 Miao, R. Q. *et al.* Identification of a receptor necessary for Nogo-B stimulated chemotaxis and morphogenesis of endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10997-11002, doi:10.1073/pnas.0602427103 (2006).

119 Park, E. J. *et al.* Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. *Cell Metab* **20**, 448-457, doi:10.1016/j.cmet.2014.06.016 (2014).

120 Zhao, B. *et al.* The Nogo-B receptor promotes Ras plasma membrane localization and activation. *Oncogene* **36**, 3406-3416, doi:10.1038/onc.2016.484 (2017).

121 Omerovic, J., Laude, A. J. & Prior, I. A. Ras proteins: paradigms for compartmentalised and isoform-specific signalling. *Cellular and molecular life sciences : CMLS* **64**, 2575-2589, doi:10.1007/s00018-007-7133-8 (2007).

122 Omerovic, J. & Prior, I. A. Compartmentalized signalling: Ras proteins and signalling nanoclusters. *The FEBS journal* **276**, 1817-1825, doi:10.1111/j.1742-4658.2009.06928.x (2009).

123 Prior, I. A. & Hancock, J. F. Ras trafficking, localization and compartmentalized signalling. *Seminars in cell & developmental biology* **23**, 145-153, doi:10.1016/j.semcdb.2011.09.002 (2012).

124 Chang, S. Y., Ko, T. P., Chen, A. P., Wang, A. H. & Liang, P. H. Substrate binding mode and reaction mechanism of undecaprenyl pyrophosphate synthase deduced from crystallographic studies. *Protein science : a publication of the Protein Society* **13**, 971-978, doi:10.1110/ps.03519904 (2004).

125 Chen, A. P. *et al.* Substrate and product specificities of cis-type undecaprenyl pyrophosphate synthase. *Biochem J* **386**, 169-176, doi:10.1042/bj20040785 (2005).

126 Guo, R. T. *et al.* Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. *The Journal of biological chemistry* **280**, 20762-20774, doi:10.1074/jbc.M502121200 (2005).

127 Li, M. & Song, J. Nogo-B receptor possesses an intrinsically unstructured ectodomain and a partially folded cytoplasmic domain. *Biochemical and biophysical research communications* **360**, 128-134, doi:http://dx.doi.org/10.1016/j.bbrc.2007.06.031 (2007).

128 Harrison, K. D. *et al.* Nogo-B Receptor stabilizes Niemann-Pick Type C2 protein and regulates intracellular cholesterol trafficking. *Cell metabolism* **10**, 208-218, doi:10.1016/j.cmet.2009.07.003 (2009).

129 Harrison, K. D. *et al.* Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. *The EMBO Journal* **30**, 2490-2500, doi:10.1038/emboj.2011.147 (2011).

130 Rambo, R. P. & Tainer, J. A. Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law. *Biopolymers* **95**, 559-571, doi:10.1002/bip.21638 (2011).

131 Vitrac, H., MacLean, D. M., Jayaraman, V., Bogdanov, M. & Dowhan, W. Dynamic membrane protein topological switching upon changes in phospholipid environment. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13874-13879, doi:10.1073/pnas.1512994112 (2015).

132 Bogdanov, M. & Dowhan, W. Lipid-dependent Generation of Dual Topology for a Membrane Protein. *The Journal of biological chemistry* **287**, 37939-37948, doi:10.1074/jbc.M112.404103 (2012).

133 Rapp, M., Granseth, E., Seppala, S. & von Heijne, G. Identification and evolution of dual-topology membrane proteins. *Nature structural & molecular biology* **13**, 112-116, doi:http://www.nature.com/nsmb/journal/v13/n2/suppinfo/nsmb1057_S1.html (2006).

134 Schuldiner, S. Undecided membrane proteins insert in random topologies Up, down and sideways: it doesn't really matter. *Trends in biochemical sciences* **37**, 215-219, doi:10.1016/j.tibs.2012.02.006 (2012).

135 Bogdanov, M., Xie, J., Heacock, P. & Dowhan, W. To flip or not to flip: lipid–protein charge interactions are a determinant of final membrane protein topology. *The Journal of cell biology* **182**, 925-935, doi:10.1083/jcb.200803097 (2008).

136 Hünefeld, F. L. *Der Chemismus in der thierischen Organisation : physiologisch-chemische Untersuchungen der materiellen Veränderungen oder des Bildungslebens im thierischen Organismus, insbesondere des Blutbildungsprocesses, der Natur der Blutkörperchen und ihrer Kernchen : ein Beitrag zur Physiologie und Heilmittellehre*. (1840).

137 Giege, R. A historical perspective on protein crystallization from 1840 to the present day. *The FEBS journal* **280**, 6456-6497, doi:10.1111/febs.12580 (2013).

138 McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications* **70**, 2-20, doi:10.1107/s2053230x13033141 (2014).

139 Funke, O. Über das Milzvenenblut. *Z Rat Medicin* **1**, 172-218 (1851).

140 Ritthausen, H., (1872).

141 Osborne, T. B. Crystallized vegetable proteids. *Amer Chem J* **14**, 662-689 (1892).

142 Kepler, J. *Ioannis Kepleris Strena, seu, De niue sexangula [microform]*. (Apud Godefridum Tampach, 1611).

143 Bragg, W. H. & Bragg, W. L. The Reflection of X-rays by Crystals. *Proceedings of the Royal Society of London. Series A* **88**, 428-438, doi:10.1098/rspa.1913.0040 (1913).

144 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).

145 Blake, C. C. *et al.* Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206**, 757-761 (1965).

146 Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature methods* **5**, 147-153, doi:10.1038/nmeth.f.203 (2008).

147 Sleutel, M. & Van Driessche, A. E. Role of clusters in nonclassical nucleation and growth of protein crystals. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E546-553, doi:10.1073/pnas.1309320111 (2014).

148 Gavira, J. A. Current trends in protein crystallization. *Archives of biochemistry and biophysics* **602**, 3-11, doi:10.1016/j.abb.2015.12.010 (2016).

149 Rayment, I. Small-Scale Batch Crystallization of Proteins Revisited. *Structure* **10**, 147-151, doi:http://dx.doi.org/10.1016/S0969-2126(02)00711-6 (2002).

150 Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. An automated system for micro-batch protein crystallization and screening. *Journal of Applied Crystallography* **23**, 297-302, doi:doi:10.1107/S0021889890003260 (1990).

151 Otalora, F., Gavira, J. A., Ng, J. D. & Garcia-Ruiz, J. M. Counterdiffusion methods applied to protein crystallization. *Progress in biophysics and molecular biology* **101**, 26-37, doi:10.1016/j.pbiomolbio.2009.12.004 (2009).

152 Gavira, J. A., Toh, D., Lopez-Jaramillo, J., Garcia-Ruiz, J. M. & Ng, J. D. Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling. *Acta crystallographica. Section D, Biological crystallography* **58**, 1147-1154 (2002).

153 Trakhanov, S. & Quiocho, F. A. Influence of divalent cations in protein crystallization. *Protein science : a publication of the Protein Society* **4**, 1914-1919, doi:10.1002/pro.5560040925 (1995).

154 Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. Screening and optimization strategies for macromolecular crystal growth. *Acta crystallographica. Section D, Biological crystallography* **50**, 414-423, doi:10.1107/s0907444994002660 (1994).

155 Tomcova, I., Branca, R. M., Bodo, G., Bagyinka, C. & Smatanova, I. K. Cross-crystallization method used for the crystallization and preliminary diffraction analysis of a novel di-haem cytochrome c4. *Acta crystallographica. Section F, Structural biology and crystallization communications* **62**, 820-824, doi:10.1107/s1744309106027710 (2006).

156 Sirinupong, N. *et al.* Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. *The Journal of biological chemistry* **285**, 40635-40644, doi:10.1074/jbc.M110.168187 (2010).

157 Hoeppner, A., Schmitt, L. & Smits, S. H. J. in *Advanced Topics on Crystal Growth* (ed Sukarno Olavo Ferreira) Ch. 01 (InTech, 2013).

158 Vedadi, M. *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15835-15840, doi:10.1073/pnas.0605224103 (2006).

159 Marsden, R. L., McGuffin, L. J. & Jones, D. T. Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein science : a publication of the Protein Society* **11**, 2814-2824, doi:10.1110/ps.0209902 (2002).

160 Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein science : a publication of the Protein Society* **12**, 1313-1322, doi:10.1110/ps.0243403 (2003).

161 Miyazaki, S., Kuroda, Y. & Yokoyama, S. Characterization and prediction of linker sequences of multi-domain proteins by a neural network. *Journal of structural and functional genomics* **2**, 37-51 (2002).

162 Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta crystallographica. Section D, Biological crystallography* **57**, 679-688 (2001).

163 Lawson, D. M. *et al.* Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541-544, doi:10.1038/349541a0 (1991).

164 McElroy, H. E., Sisson, G. W., Schoettlin, W. E., Aust, R. M. & Villafranca, J. E. Studies on engineering crystallizability by mutation of surface residues of human thymidylate

synthase. *Journal of Crystal Growth* **122**, 265-272, doi:http://dx.doi.org/10.1016/0022-0248(92)90255-H (1992).

165 Zhang, X. J., Wozniak, J. A. & Matthews, B. W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *Journal of molecular biology* **250**, 527-552, doi:10.1006/jmbi.1995.0396 (1995).

166 Zhang, F. *et al.* Crystal structure of the obese protein leptin-E100. *Nature* **387**, 206-209, doi:10.1038/387206a0 (1997).

167 Walter, T. S. *et al.* Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure* **14**, 1617-1622, doi:http://dx.doi.org/10.1016/j.str.2006.09.005 (2006).

168 Rypniewski, W. R., Holden, H. M. & Rayment, I. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **32**, 9851-9858 (1993).

169 Stevens, R. C. Design of high-throughput methods of protein production for structural biology. *Structure* **8**, R177-185 (2000).

170 Edwards, A. M. *et al.* Protein production: feeding the crystallographers and NMR spectroscopists. *Nature structural biology* **7 Suppl**, 970-972, doi:10.1038/80751 (2000).

171 Braun, P. *et al.* Proteome-scale purification of human proteins from bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2654-2659, doi:10.1073/pnas.042684199 (2002).

172 Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H. & Hard, T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in

Escherichia coli. *Protein science : a publication of the Protein Society* **11**, 313-321, doi:10.1110/ps.22102 (2002).

173 Shih, Y. P. *et al.* High-throughput screening of soluble recombinant proteins. *Protein science : a publication of the Protein Society* **11**, 1714-1719, doi:10.1110/ps.0205202 (2002).

174 Center, R. J. *et al.* Crystallization of a trimeric human T cell leukemia virus type 1 gp21 ectodomain fragment as a chimera with maltose-binding protein. *Protein science : a publication of the Protein Society* **7**, 1612-1619, doi:10.1002/pro.5560070715 (1998).

175 Waugh, D. S. Crystal structures of MBP fusion proteins. *Protein science : a publication of the Protein Society* **25**, 559-571, doi:10.1002/pro.2863 (2016).

176 Spurlino, J. C., Lu, G. Y. & Quiocho, F. A. The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *The Journal of biological chemistry* **266**, 5202-5219 (1991).

177 Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* **31**, 10657-10663 (1992).

178 Matsumoto, S., Shimada, A. & Kohda, D. Crystal structure of the C-terminal globular domain of the third paralog of the Archaeoglobus fulgidus oligosaccharyltransferases. *BMC structural biology* **13**, 11, doi:10.1186/1472-6807-13-11 (2013).

179 Cherry, A. L. *et al.* Structural basis of SUFU-GLI interaction in human Hedgehog signalling regulation. *Acta crystallographica. Section D, Biological crystallography* **69**, 2563-2579, doi:10.1107/s0907444913028473 (2013).

180 Moon, A. F., Mueller, G. A., Zhong, X. & Pedersen, L. C. A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. *Protein science : a publication of the Protein Society* **19**, 901-913, doi:10.1002/pro.368 (2010).

181 Kobe, B., Center, R. J., Kemp, B. E. & Poumbourios, P. Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 4319-4324 (1999).

182 Kuge, M. *et al.* Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. *Protein science : a publication of the Protein Society* **6**, 1783-1786, doi:10.1002/pro.5560060822 (1997).

183 Lally, J. M. *et al.* Crystallization of an intact GST-estrogen receptor hormone binding domain fusion protein. *Acta crystallographica. Section D, Biological crystallography* **54**, 423-426 (1998).

184 Carter, A. P., Cho, C., Jin, L. & Vale, R. D. Crystal structure of the dynein motor domain. *Science (New York, N.Y.)* **331**, 1159-1165, doi:10.1126/science.1202393 (2011).

185 Schmidt, H., Gleave, E. S. & Carter, A. P. Insights into dynein motor domain function from a 3.3-A crystal structure. *Nature structural & molecular biology* **19**, 492-497, s491, doi:10.1038/nsmb.2272 (2012).

186 Jiang, Y., Sirinupong, N., Brunzelle, J. & Yang, Z. Crystal Structures of Histone and p53 Methyltransferase SmyD2 Reveal a Conformational Flexibility of the Autoinhibitory C-Terminal Domain. *PloS one* **6**, e21640, doi:10.1371/journal.pone.0021640 (2011).

187 Jiang, Y. *et al.* Structural insights into estrogen receptor alpha methylation by histone methyltransferase SMYD2, a cellular event implicated in estrogen signaling regulation. *Journal of molecular biology* **426**, 3413-3425, doi:10.1016/j.jmb.2014.02.019 (2014).

188 Sirinupong, N., Brunzelle, J., Doko, E. & Yang, Z. Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. *Journal of molecular biology* **406**, 149-159, doi:10.1016/j.jmb.2010.12.014 (2011).

189 Kovari, L. C., Momany, C. & Rossmann, M. G. The use of antibody fragments for crystallization and structure determinations. *Structure* **3**, 1291-1293 (1995).

190 Prongay, A. J. *et al.* Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 9980-9984 (1990).

191 Heyman, B. Complement and Fc-receptors in regulation of the antibody response. *Immunology letters* **54**, 195-199 (1996).

192 Putnam, F. W., Liu, Y. S. & Low, T. L. Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain. *The Journal of biological chemistry* **254**, 2865-2874 (1979).

193 Hunte, C. & Michel, H. Crystallisation of membrane proteins mediated by antibody fragments. *Current opinion in structural biology* **12**, 503-508 (2002).

194 Lesk, A. M. & Chothia, C. Elbow motion in the immunoglobulins involves a molecular ball-and-socket joint. *Nature* **335**, 188-190, doi:10.1038/335188a0 (1988).

195 Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution. *Nature* **414**, 43-48, doi:10.1038/35102009 (2001).

196 Harrenga, A. & Michel, H. The cytochrome c oxidase from Paracoccus denitrificans does not change the metal center ligation upon reduction. *The Journal of biological chemistry* **274**, 33296-33299 (1999).

197 Ko, S., Kim, H. Y., Choi, I. & Choe, J. Gold Nanoparticles as Nucleation-Inducing Reagents for Protein Crystallization. *Crystal Growth & Design* **17**, 497-503, doi:10.1021/acs.cgd.6b01346 (2017).

198 Chen, Y.-W., Lee, C.-H., Wang, Y.-L., Li, T.-L. & Chang, H.-C. Nanodiamonds as Nucleating Agents for Protein Crystallization. *Langmuir*, doi:10.1021/acs.langmuir.7b00578 (2017).

199 Nanev, C. N., Saridakis, E. & Chayen, N. E. Protein crystal nucleation in pores. *Scientific reports* **7**, 35821, doi:10.1038/srep35821 (2017).

200 Abdallah, B. G., Roy-Chowdhury, S., Fromme, R., Fromme, P. & Ros, A. Protein Crystallization in an Actuated Microfluidic Nanowell Device. *Cryst Growth Des* **16**, 2074-2082, doi:10.1021/acs.cgd.5b01748 (2016).

201 Ujwal, R. & Bowie, J. U. Crystallizing membrane proteins using lipidic bicelles. *Methods (San Diego, Calif.)* **55**, 337-341, doi:10.1016/j.ymeth.2011.09.020 (2011).

202 Agah, S. & Faham, S. Crystallization of membrane proteins in bicelles. *Methods in molecular biology (Clifton, N.J.)* **914**, 3-16, doi:10.1007/978-1-62703-023-6_1 (2012).

203 Ujwal, R. & Abramson, J. High-throughput crystallization of membrane proteins using the lipidic bicelle method. *Journal of visualized experiments : JoVE,* e3383, doi:10.3791/3383 (2012).

ABSTRACT

STRUCTURAL BASIS OF MEMBRANE PROTEIN SCAFFOLDING AND SIGNALING IN HUMAN DISEASE

by

JOSHUA HOLCOMB

August 2017

Advisor: Dr. Zhe Yang

Major: Biochemistry and Molecular Biology

Degree: Doctor of Philosophy

Protein structural elucidation by means of X-ray crystallography is a powerful approach for both insight into a proteins biophysical properties and function. To date X-ray crystallography remains the gold standard in high resolution structural determination and serves as the basis for rational drug design for the purpose of combating a number of human diseases. Such an approach also allows for the exploitation of how various proteins interact with their substrates providing a molecular basis for their physiological function. In this dissertation, using X-ray crystallographic analysis along with other biophysical characterization methods, we seek to understand the mechanistic foundation for which scaffolding proteins such as NHERF1 and NHERF2 interact with their substrates in which leads to a wide variety of critical cellular events. Additionally, we elucidate the structure and function of a recently identified membrane protein, NgBR, with hopes to understand the molecular basis in which it influences the activation of the infamous oncoprotein Ras.

Protein scaffolding is a term that denotes the coming together of two or more proteins in which results in the formation of macromolecular complexes. Such events are known to regulate

several cellular processes including membrane protein recycling, protein stabilization, cell to cell adhesion, and regulation of signaling pathways. There are several families of scaffolding proteins with one of the most known to be the PDZ domain containing family. Two proteins within this family, NHERF1 and NHERF2, contain two PDZ domains in which are responsible for a variety of signaling events. The second PDZ domain of NHERF1 has been shown to complex with chemokine receptor CXCR2 and mediate the formation of the CXCR2-NHERF1-PLCβ2 complex. This complex formation has been shown to be influential in chemokine induced neutrophil migration and infiltration in regions of inflammation leading to the exacerbation of a variety of inflammatory diseases. Additionally, the first PDZ domain of NHERF2 has been shown to mediate the scaffolding of the CFTR-NHERF2-LAP2 complex which has been shown to inhibit CFTR activity by the LPA2 mediated inhibition of adenylate cyclase. Furthermore, disruption of this complex has been shown to be enough to augment CFTR activity *in vivo*. And thus, the determining the molecular basis by means both NHERF1 and NHERF2 scaffolding with its respective targets could lead to therapeutic targeting in a variety of human diseases. In this dissertation, we solved the co-crystal complexes between both NHERF1 PDZ2 with the C-terminal peptide of CXCR2 in addition to solving the co- crystal complex between NHERF2 PDZ1 with the C-terminal LPA2 peptide. Both structures reveal an expected mode of type 1 PDZ substrate recognition with various ligand specific interactions. However, both structures also reveal unexpected findings which may unveil additional therapeutic strategies other than targeting the PDZ substrate binding site.

The Nogo-B receptor is a recently identified type one transmembrane receptor with little known function besides its role in binding Nogo-B and promoting angiogenesis by chemotaxis *in vivo*. However, its most recently identified role is that of binding and facilitating the oncoprotein

Ras to the plasma membrane for activation. No X-ray crystallographic structure currently exists for NgBR but previous studies have confirmed to be composed of three primary domains including a cytosolic and extracellular domain with one transmembrane region. However, successive data have suggested that NgBR may exist in more than one topological orientation. Because previous attempts to crystallize NgBR have failed due to solubility issues and since it has been observed to possess various topological orientations, we chose to exploit NgBR's structural properties using a complimentary approach to X-ray crystallography known as small angle x-ray scattering (SAXS). By incorporation of residues in our expression construct in which share sequence identity to the *cis*-IPTase UPPs, we are able to achieve good solubility and purification of our NgBR construct. Additionally, SAXS analysis reveals that our NgBR construct exists as a globular macromolecule in solution which may pave the way for future high resolution structural studies.

Altogether, X-ray crystallography is a powerful technique for the determination of protein structure. The only bottleneck of this approach is the obtainment of a protein crystal. The final chapter in this thesis overviews well known and leading techniques in the evasion of this rate limiting step in the pursuit of solving molecular structure. And although there is still no one method for guaranteeing the formation of a protein crystal, X-ray crystallography continues to unveil new structural information and potential therapeutic targets which can be implicated in the advancement of molecular medicine.

AUTOBIOGRAPHICAL STATEMENT

Education

2008-2013 B.S. Biochemistry, Central Michigan University, Mount Pleasant, MI
2013-pres. Ph.D. Biochemistry and Molecular Biology; Wayne State University, Detroit,MI

Positions and Honors

2012- Member, American Chemical Society
2014 2nd Annual UAB Workshop on Metabolomics NIH travel fellowship
2016- Member, American Society for Biochemistry and Molecular Biology
2016-2017 Detroit Cardiovascular Training Program NIH T32 fellowship
2017 ASBMB Travel Award
2017 Experimental Biology 2017 poster presentation

Peer Reviewed Publications (8 of 15)

1. **Holcomb, J**. Spellmon, N., Trescott, L., Sun, F., Li, C., and Yang, Z. PDZ Structure and Implication in Selective Drug Design Against Cystic Fibrosis. *Current Drug Targets*, 16(9):945-50 (2015).
2. **Holcomb, J**., Jiang, Y., Lu, G., Trescott, L., Brunzelle, J., Sirinupong, N., Li, C., Naren, A. and Yang, Z. Structural Insights into PDZ-mediated Interaction of NHERF2 and LPA2, a Cellular Event Implicated in CFTR Channel Regulation. *Biochem Biophys Res Commun*, S0006-291X(14)00407-0. doi: 10.1016/j.bbrc. 2014.02.128 (2014).
3. **Holcomb, J**., Jiang, Y., Trescott, L., Lu, G., Brunzelle, J., Sirinupong, N., Li, C. and Yang, Z. Crystal Structure of the NHERF1 PDZ2 Domain in Complex with the Chemokine Receptor CXCR2 Reveals Probable Modes of PDZ2 Dimerization. *Biochem Biophys Res Commun*, 448(2):169-74. doi: 10.1016/j.bbrc.2014.04.085 (2014).
4. Kuiper BD, Slater K, Spellmon N, **Holcomb J**, Medapureddy P, Muzzarelli KM, Yang Z, Ovadia R, Amblard F, Kovari IA, Schinazi RF, Kovari LC. Increased activity of unlinked Zika virus NS2B/NS3 protease compared to linked Zika virus protease. *Biochem Biophys Res Commun*. 2017 Mar 22. pii: S0006-291X(17)30567-3. doi: 10.1016/j.bbrc.(**2017**).03.108
5. Spellmon, N., **Holcomb, J**., Niu, A., Choudhary, V., Sun, X., Zhang, Y., Wan, J., Doughan, M., Hayden, S., Hachem, F., Brunzelle, F., Li, C. and Yang, Z. Structural basis of PDZ-mediated chemokine receptor CXCR2 scaffolding by guanine nucleotide exchange factor PDZ-RhoGEF. *Biochem Biophys Res Commun*, doi: 10.1016/j.bbrc.2017.02.010, (2017).
6. Spellmon, N., Sun, X., Xue, W., **Holcomb, J**., Chakravarthy, S., Shang, W., Edwards, B., Sirinupong, N., Li, C. and Yang, Z**.** New open conformation of SMYD3 implicates conformational selection and allostery. *AIMS Biophysics*, 4(1):1-18, doi: 10.3934/biophy.2017.1.1, (2017).
7. Jiang, Y., **Holcomb, J**., Spellmon, N. and Yang, Z. Purification of Histone Lysine Methyltransferase SMYD2 and Co-Crystallization with a Target Peptide from Estrogen Receptor α. ***Methods Mol Biol.*** 1366:207-17, (**2016**)
8. Sun, X., Spellmon, N., **Holcomb, J.**, Xue, W., Li, C. and Yang, Z. Epigenetic landscape in embryonic stem cell. Stem Cells in Toxicology and Medicine, *Wiley* (2015).