

STRUCTURAL STUDIES OF INTEGRAL MEMBRANE PROTEINS INVOLVED IN  
GPCR SIGNALING AND STEROL HOMEOSTASIS

APPROVED BY SUPERVISORY COMMITTEE

---

Dan Rosenbaum, Ph.D

---

Luke Rice, Ph.D

---

Kevin Gardner, Ph.D

---

Youxing Jiang, Ph.D

## DEDICATION

I would like to dedicate this dissertation to all of my incredible friends, family, and mentors that supported me in this journey.

I've been so lucky to have friends like Shirin and Scarlett, my old ski buddies and still close friends after 20+ years; Todd, Mike, Tom, (and Corey) who make me laugh more than anyone else and have kept the party going since college even though we all live apart; Mary, Eddie, Ryan, and Carina who are my Big D crew; Andre, Mika, Mameko, Nick, and Scott who supported my decision to leave the great northeast for Dallas; Beth and Allyson for being great listeners and not judging me when I keep ordering my Thai food too spicy; Matt and the rest of the MIT crew for all their advice to not grow "bitter" during this process (and teaching me about ssNMR along the way); Jamila for keeping the goofs alive across the ocean; and of course Jake for being my best grad school buddy and making NMR vector diagrams out of cocktail straws with me.

I would also like to thank my family. They have always been there for me, willing and devoted to celebrate every success of mine, either professional or personal, and regardless of size. I appreciate their love and support and our shared love of baseball.

I've been so fortunate to have a good deal of experience before starting my PhD, and I truly thank my former mentors for steering me towards graduate school, and biophysics in particular. My first technician job was with Soumya Ray at HMS, and I still remember feeling the spark that made me change my decision to strive for a PhD instead of an MD. I also owe a great deal of gratitude to Bob Griffin at MIT for his incredible support both past and present. I would be remiss to not thank my former and current lab members, and particularly Karen for being my right hand (wo)man for the last few years.

And lastly, I would like to thank the members of my committee for their feedback and encouragement during this process. I especially am grateful for my mentor, Dan, for being my sounding board for the past seven years and being willing to jump on the bench when I need an extra pair of hands or just have poor time management skills. I appreciate his support particularly in this last final stretch, and I am looking forward to see how these projects continue on after I pass the baton.

STRUCTURAL STUDIES OF INTEGRAL MEMBRANE PROTEINS INVOLVED IN  
GPCR SIGNALING AND STEROL HOMEOSTASIS

by

LINDSAY DIANE CLARK

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2018

Copyright

by

Lindsay Diane Clark, 2018

All Rights Reserved

STRUCTURAL STUDIES OF INTEGRAL MEMBRANE PROTEINS INVOLVED IN  
GPCR SIGNALING AND STEROL HOMEOSTASIS

Lindsay Diane Clark, Ph.D

The University of Texas Southwestern Medical Center at Dallas, 2018

Supervising Professor: Dan Rosenbaum, Ph.D

Membrane proteins are crucial molecules for cellular survival, and can take on multiple and diverse roles within the native membrane. In this dissertation, I will detail my efforts to understand and study two different types of membrane proteins. First, I will discuss my research developing and applying a strategy to use NMR spectroscopy to study specific receptors within the large family of G protein-coupled receptors. This strategy enabled the first methyl-TROSY experiments on a wild-type human GPCR, and have significant value for future drug discovery efforts on this important class of membrane proteins. Second, I will discuss my endeavors to understand the important role of the protein Scap, which can both sense and respond to differences in cholesterol levels within the ER membrane. Scap is a central player in the SREBP pathway, which is targeted by multiple classes of pharmaceuticals, including statins. Through efforts described in the second half of this dissertation, I have been able to demonstrate the first biochemical

characterization of the full-length mammalian Scap/Insig complex, which has led to the first structural characterization of this important machinery. The long-term goal of both of these projects is aimed at having a more complete understanding of how these important membrane proteins respond to ligands and other environmental changes within their native cell membrane. This information will further our ability to diagnose and treat diseases ranging from insomnia and chronic pain to atherosclerosis and hypercholesterolemia.

## TABLE OF CONTENTS

|  |     |
|--|-----|
| PRIOR PUBLICATIONS.....  | xiv |
| LIST OF FIGURES AND TABLES.....  | xvi |
| LIST OF ABBREVIATIONS.....   | xxv |
| CHAPTER 1: INTRODUCTION .....  | 2   |
| BACKGROUND ON MEMBRANE PROTEINS.....   | 2   |
| G PROTEIN-COUPLED RECEPTOR SIGNALING.....  | 3   |
| CHOLESTEROL HOMEOSTASIS MACHINERY.....   | 5   |
| CHAPTER 2: METHYL LABELING AND TROSY NMR SPECTROSCOPY OF<br>PROTEINS EXPRESSED IN THE EUKARYOTE PICHIA PASTORIS..... | 9   |
| ABSTRACT.....  | 9   |
| RESULTS AND DISCUSSION.....  | 9   |
| MATERIALS AND METHODS.....   | 15  |
| MBP EXPRESSION AND PURIFICATION.....   | 15  |
| ACTIN EXPRESSION AND PURIFICATION.....   | 17  |
| MASS SPECTROMETRY.....   | 20  |
| NMR SPECTROSCOPY.....  | 21  |
| REFERENCES.....  | 24  |
| FIGURES.....   | 28  |
| CHAPTER 3: LIGAND MODULATION OF SIDECHAIN DYNAMICS IN A WILD-<br>TYPE HUMAN GPCR.....                                | 37  |
| ABSTRACT.....  | 37  |

|  |     |
|--|-----|
| INTRODUCTION.....  | 38  |
| RESULTS .....  | 44  |
| DISCUSSION.....  | 54  |
| MATERIALS AND METHODS .....  | 60  |
| CONSTRUCT DESIGN.....  | 60  |
| EXPRESSION OF PROTONATED CULTURES .....  | 61  |
| EXPRESSION OF DEUTERATED CULTURES.....   | 61  |
| PURIFICATION .....   | 63  |
| G PROTEIN EXPRESSION AND PURIFICATION .....  | 65  |
| A <sub>2A</sub> R-G <sub>s</sub> COMPLEX FORMATION .....   | 66  |
| NMR SPECTROSCOPY AND ANALYSIS.....   | 68  |
| COMPUTATIONAL METHODS.....   | 71  |
| MEMBRANE BINDING .....   | 74  |
| RECONSTITUTION AND GTPGAMMAS BINDING .....   | 75  |
| REFERENCES.....  | 77  |
| FIGURES.....   | 87  |
| CHAPTER 4: ISOTOPIC LABELING OF EUKARYOTIC MEMBRANE PROTEINS<br>FOR NMR STUDIES OF INTERACTIONS AND DYNAMICS ..... | 104 |
| ABSTRACT.....  | 104 |
| INTRODUCTION.....  | 105 |
| EXPRESSION AND ISOTOPIC LABELING IN PICHIA .....   | 109 |
| REQUIRED EQUIPMENT AND REAGENTS .....  | 110 |

|   |     |
|---|-----|
| PROCEDURE.....  | 111 |
| PURIFICATION/NMR SAMPLE PREPARATION.....  | 119 |
| REQUIRED EQUIPMENT AND REAGENTS .....   | 119 |
| PROCEDURE.....  | 121 |
| RUNNING <sup>13</sup> C/ <sup>1</sup> H HMQC AND METHYL RELAXATION EXPERIMENTS .....                | 126 |
| REQUIRED EQUIPMENT AND SOFTWARE.....  | 127 |
| PROCEDURE.....  | 127 |
| CONCLUDING REMARKS.....   | 135 |
| REFERENCES.....   | 138 |
| FIGURES .....   | 147 |
| CHAPTER 5: ON THE USE OF PICHIA PASTORIS FOR ISOTOPIC LABELING OF HUMAN GPCRS FOR NMR STUDIES ..... | 155 |
| ABSTRACT.....   | 155 |
| INTRODUCTION.....   | 156 |
| MATERIALS AND METHODS .....   | 158 |
| CONSTRUCT DESIGN.....   | 158 |
| EXPRESSION OF DEUTERATED CULTURES.....  | 159 |
| RECEPTOR PURIFICATION.....  | 160 |
| NMR EXPERIMENTS.....  | 161 |
| MEMBRANE PROTEIN LABELING IN INSECT CELLS.....  | 162 |
| MEMBRANE PROTEIN LABELING IN MAMMALIAN CELLS.....   | 163 |

|   |     |
|---|-----|
| MEMBRANE PROTEIN LABELING IN YEAST.....   | 164 |
| APPLICATION OF METHYL LABELING IN PICHIA TO GPCRS .....                             | 166 |
| CONCLUSIONS.....  | 170 |
| REFERENCES.....   | 172 |
| FIGURES .....   | 180 |
| CHAPTER 6: STRUCTURAL AND FUNCTIONAL STUDIES OF INDIVIDUAL<br>DOMAINS OF SCAP ..... | 183 |
| INTRODUCTION.....   | 183 |
| BACKGROUND ON SREBP PATHWAY .....   | 183 |
| CHARACTERIZATION OF SOLUBLE DOMAINS OF SCAP .....                                   | 187 |
| LOOP1-LOOP7 CONSTRUCT .....   | 188 |
| PURIFICATION AND CONSTRUCT OPTIMIZATION .....                                       | 190 |
| CRYSTALLIZATION .....   | 196 |
| CHOLESTEROL BINDING ASSAYS.....   | 196 |
| NMR SPECTROSCOPY .....  | 197 |
| C-TERMINAL WD DOMAIN.....   | 201 |
| CHARACTERIZATION OF TRANSMEMBRANE DOMAIN OF SCAP .....                              | 203 |
| TM1-8 CONSTRUCT.....  | 203 |
| TM1-6MINI CONSTRUCT .....   | 205 |
| REFERENCES.....   | 207 |
| FIGURES .....   | 210 |

|   |     |
|---|-----|
| CHAPTER 7: CHARACTERIZATION OF FULL LENGTH SCAP, INSIG AND THE SCAP/INSIG COMPLEX ..... | 222 |
| INTRODUCTION.....   | 222 |
| SCAP ORTHOLOGUE SCREEN .....  | 223 |
| THERMOPHILIC SCAP .....   | 223 |
| SCAP FROM HIGHER EUKARYOTES .....   | 227 |
| CHARACTERIZATION OF FULL-LENGTH SCAP.....   | 228 |
| THERMOSTABILITY SCREEN .....  | 229 |
| VIRUS PRODUCTION AND GNT11- EXPRESSION SYSTEM .....                                     | 241 |
| ADDITION OF N-TERMINAL FUSION PROTEIN .....   | 245 |
| PURIFICATION PROTOCOL AND OPTIMIZATION .....  | 247 |
| CHARACTERIZATION FULL LENGTH INSIG.....   | 251 |
| PURIFICATION OF FULL LENGTH SCAP/INSIG COMPLEX.....                                     | 254 |
| PROTEIN BIOCHEMISTRY.....   | 254 |
| ADDITIONAL POINT MUTATIONS .....  | 257 |
| STEROL BINDING ASSAYS .....   | 259 |
| RECONSTITUTION INTO LIPID NANODISCS .....   | 260 |
| BACKGROUND .....  | 261 |
| PURIFICATION OF MEMBRANE SCAFFOLD PROTEINS.....   | 264 |
| RECONSTITUTION OF SCAP/INSIG COMPLEX .....  | 267 |
| GENERATION OF CONFORMATIONALLY SPECIFIC NANOBODIES .....                                | 269 |
| MACS/FACS SCREENING.....  | 270 |

|  |     |
|--|-----|
| EXPRESSION AND CHARACTERIZATION OF INDIVIDUAL CLONES ...   | 271 |
| REFERENCES.....  | 274 |
| FIGURES.....   | 281 |
| CHAPTER 8: CRYO-EM OF FULL LENGTH SCAP/INSIG COMPLEX ..... | 309 |
| INTRODUCTION.....  | 309 |
| GRID PREPARATION .....                                     | 309 |
| DATA COLLECTION.....                                       | 310 |
| CURRENT MODEL.....   | 311 |
| REFERENCES.....  | 316 |
| FIGURES .....  | 317 |
| CHAPTER 9: CONCLUSIONS AND FUTURE DIRECTIONS .....         | 321 |
| NMR STUDIES OF GPCRS.....                                  | 321 |
| CONCLUSIONS .....  | 321 |
| FUTURE DIRECTIONS .....                                    | 324 |
| STUDIES OF CHOLESTEROL HOMEOSTASIS MACHINERY .....         | 326 |
| CONCLUSIONS .....  | 326 |
| FUTURE DIRECTIONS .....                                    | 327 |



## PRIOR PUBLICATIONS

Clark, L., Dikiy, I., Rosenbaum D. M., Gardner, K. H. (2018). On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies. *Journal of Biomolecular NMR*, 71(4), 203–211. <http://doi.org/10.1007/s10858-018-0204-3>

Dikiy, I.\* ,Clark, L. D.\* , Gardner, K. H., Rosenbaum, D. M. Isotopic labeling of integral membrane proteins for study by solution NMR relaxation methods. Accepted to *Methods in Enzymology*. <https://doi.org/10.1016/bs.mie.2018.08.030>

Clark, L. D.\* , Dikiy, I.\* , Chapman, K., Rodstrom, K. E., Aramini, J., LeVine, M. V., Khelashvili, G., Rasussen, S. G. F., Gardner, K. H., Rosenbaum, D. M. Ligand modulation of sidechain dynamics in a wild-type human GPCR. *eLife* (2017) vol. 6

Shao, Z., Yin, J., Chapman, K., Grzemska, M., Clark, L., Wang, J., Rosenbaum, D. M. High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* (2016) vol. 540 (7634) pp. 602-606

Zhang, Y., Lee, K. M., Kinch, L. N., Clark, L., Grishin, N. V., Rosenbaum, D. M., Brown, M. S., Goldstein, J. L., Radhakrishnan, A. Direct Demonstration that Loop1 of Scap Binds to Loop7, a Crucial Event in Cholesterol Homeostasis. *Journal of Biological Chemistry* (2016) vol. 291 (24) pp. 12888-12896

Yin, J., Babaoglu, K., Brautigam, C. A., Clark, L., Shao, Z., Scheuermann, T. H., Harrell, C. M., Gotter, A. L., Roecker, A. J., Winrow, C. J., Renger, J. J., Coleman, P. J., Rosenbaum, D. M. Structure and ligand-binding mechanism of the human OX1 and OX2 orexin receptors. *Nature Structural and Molecular Biology* (2016) vol. 23 (4) pp. 293-301

Clark, L., Zahm, J. A., Ali, R., Kukula, M., Bian, L., Gardner, K. H., Rosen, M. K., Rosenbaum, D. M. Methyl labeling and TROSY NMR spectroscopy of protein expressed in the eukaryote *Pichia pastoris*. *Journal of Biomolecular NMR* (2015) vol. 62 (3) pp. 239-245

Eddy, M. T., Su, Y., Silvers, R., Andreas, L., Clark, L., Wagner, G., Pintacuda, G., Emsley, L., Griffin, R. G. Lipid-bilayer-bound conformation of an integral membrane  $\beta$ -barrel protein by multidimensional MAS NMR. *Journal of Biomolecular NMR* (2015) vol. 61 (3) pp. 299-310

Eddy, M. T., Andreas, L., Hermida, O., Su, Y., Clark, L., Noskov, S., Rostovtseva, T. K., Wagner, G., Griffin, R. G. Magic angle spinning NMR characterization of VDAC gating in 2D lipid crystalline bilayers. *Biochemistry* (2015) vol. 54 (4) pp. 994-1005

Eddy, M. T., Ong, T-C., Clark, L., Teijido, O., van der Wel, P. C. A., Garces, R., Wagner, G., Rostovtseva, T. K., Griffin, R. G. Lipid dynamics and protein-lipid interactions in 2D crystals formed with the  $\beta$ -barrel integral membrane protein VDAC1. *Journal of the American Chemical Society* (2012) vol.134 (14) pp. 6375-6387

Logan, T., Clark, L., Ray, S. S. Engineered disulfide bonds restore chaperone-like function of DJ-1 mutants linked to familial Parkinson's Disease. *Biochemistry* (2010) vol. 49 (27) pp. 5624-5633

## LIST OF FIGURES AND TABLES

### CHAPTER TWO

Figure 1 – Incorporation of  $^{13}\text{C}$ -methyl groups at the  $\delta 1$  position of isoleucine residues in proteins expressed in *P. pastoris* (pg. 28)

Figure 2 – Labeling of  $\delta 1$ -methyl groups of MBP expressed in *Pichia pastoris* (pg. 29)

Figure 3 – Expanded view of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of isoleucine  $\delta 1$ -methyl labeled maltose binding protein (pg. 30)

Figure 4 – NMR spectra of *Drosophila* actin labeled and overexpressed in *Pichia pastoris* (pg. 31)

Supplemental Figure S1 – Comparison of MBP Ile  $\delta 1$  chemical shifts (pg. 32)

Supplemental Figure S2 – Intensity comparison of ten well-resolved cross peaks in methyl HMQC spectra (pg. 33)

Supplemental Figure S3 – Mass spectra of deuterated  $^{13}\text{C}$ -methyl-Ile-labeled MBP (pg. 34)

Supplemental Figure S4 – Comparison of MBP expressed in *E. coli* and *P. pastoris* (pg. 35)

Supplemental Figure S5 – Recombinant actin for NMR (pg. 36)

### CHAPTER THREE

Figure 1 –  $A_{2A}R$  expression and  $G_s$  activation (pg. 87)

Figure 1 supp. 1 –  $A_{2A}R$  processing and purification (pg. 88)

Figure 1 supp. 2 – Ligand binding assays on A<sub>2A</sub>R yeast membranes (pg. 89)

Figure 2 – NMR spectra of A<sub>2A</sub>R in different liganded states (pg. 90)

Figure 2 supp. 1 – A<sub>2A</sub>R and G<sub>s</sub> complex formation (pg. 91)

Figure 2 supp. 2 – Assignments of select Ile- $\delta$ 1 methyl resonances in A<sub>2A</sub>R (pg. 92)

Figure 2 supp. 3 – Lifetime of A<sub>2A</sub>R NMR sample (pg. 93)

Figure 3 – Ligand- and cation-dependent chemical shift changes in A<sub>2A</sub>R (pg. 94)

Figure 3 supp. 1 – Overlays of ligand- and cation-dependent chemical shift changes in A<sub>2A</sub>R (pg. 95)

Figure 4 – Solvent PRE analysis of A<sub>2A</sub>R (pg. 96)

Figure 5 – Modified 3Q relaxation for MBP (pg. 97)

Figure 5 supp. 1 – Adaptation of 3Q methyl relaxation experiment to A<sub>2A</sub>R samples (pg. 98)

Figure 6 – Changes in dynamics of methyl groups of A<sub>2A</sub>R in the presence of different ligands (pg. 100)

Figure 6 supp. 1 – Changes in dynamics of methyl groups of A<sub>2A</sub>R in the presence of different ligands from MD trajectories (pg. 101)

Figure 7 – Structural contexts of isoleucine peak assignments (pg. 102)

## CHAPTER FOUR

Table 1 – Typical spectral parameters for  $^{13}\text{C}/^1\text{H}$  HMQC spectra of U- $^2\text{H}$  ( $^{13}\text{C},^1\text{H}$  Ile  $\delta 1$   $\text{CH}_3$ ) labeled GPCR samples (pg. 147)

Figure 1 – Optimization of MF $\alpha$  signal sequence (pg. 148)

Figure 2 – Adaptation of *Pichia pastoris* to deuterated media (pg. 149)

Figure 3 – Receptors reconstituted with CHS are more stable but have crowded spectra (pg. 150)

Figure 4 – Assignment of selected A $_2\text{A}$ R Ile- $\delta 1$  methyl peaks by site-directed mutagenesis (pg. 152)

Figure 5 – Schematic of fast motions relaxation experiment (pg. 153)

## CHAPTER FIVE

Figure 1 – Distribution and  $^1\text{H}/^{13}\text{C}$  HMQC spectra of isoleucine  $\delta 1$  methyl groups within representative GPCRs (pg. 180)

Figure 2 – Sample gel filtration chromatograms of GPCR purifications (pg. 181)

Table 1 – Summary of strengths and weaknesses of eukaryotic systems for the expression and isotopic labeling of integral membrane proteins for NMR studies (pg. 182)

## CHAPTER SIX

Figure 1 – Purification of L1L7 (pg. 210)

Figure 2 – Limited proteolysis of optimized L1L7 (pg. 211)

Figure 3 – Circular dichroism of optimized L1L7 (pg. 212)

Figure 4 – Crystallization of optimized L1L7 crystals (pg. 213)

Figure 5 – L1L7 monoclonal antibody details (pg. 214)

Figure 6 – L1L7 crystallization conditions (pg. 215)

Figure 7 – Saturation binding of optimized L1L7 (pg. 216)

Figure 8 – Competition binding of optimized L1L7 (pg. 217)

Figure 9 – HSQC spectra of <sup>15</sup>N-Leu labeled L1L7 with and without cholesterol (pg. 218)

Figure 10 – Overlay of L1L7 NMR spectra (pg. 219)

Figure 11 – SDS-PAGE gels of purified Scap16mini constructs (pg. 220)

Figure 12 – SDS-PAGE gel of M2-purified Scap16miniGFP co-expressed with untagged Insig2 (pg. 221)

## CHAPTER SEVEN

Figure 1 – Negative stain and cryo-EM micrographs of FC-13 solubilized TCT (pg. 281)

Figure 2 – Negative stain EM of amphipol-reconstituted TCT (pg. 282)

Figure 3 – Cryo-EM of amphipol-reconstituted TCT (pg. 283)

Figure 4 – Orthologue screening of ScapFL with species-matched Insig-2 (pg. 284)

Figure 5 – Mapping of mutations screened for in FSEC thermostability assay (pg. 285)

Figure 6 – Example overlay traces of individual FSEC runs (pg. 286)

Figure 7 – Summary of FSEC results (pg. 287)

Figure 8 – Summary of additional substitutions at position 428 (pg. 288)

Figure 9 – Full  $T_m$  curve of ScapTM16mini constructs (pg. 289)

Figure 10 – *In vivo* SREBP trafficking assay of single and double point mutations in CHO cells (pg 290)

Figure 11 – Comparison of ScapTM 1-6 with the SSD of NPC1 (pg. 291)

Figure 12 – Effect of additional glutamate mutations on Insig binding (pg. 292)

Figure 13 – SREBP trafficking assays of WT and glutamate mutant forms of Scap (pg. 293)

Figure 14 – Optimization of N-terminal fusion protein onto ScapFL (pg. 294)

Figure 15 – Purification of ScapFL 2M + Insig2 with and without N-terminal fusion protein (pg. 295)

Figure 16 – Comparison of ScapFL WT vs 2M (pg. 296)

Figure 17 – SEC traces of Insig (pg. 297)

Figure 18 – SEC-MALS of Insig-2 (pg. 298)

Figure 19 – SEC traces of QC6-ScapFL 2M + Insig2 (pg. 299)

Figure 20 – Ion exchange of ScapFL 2M + Insig2 in LMNG (pg. 300)

Figure 21 – SEC-MALS data of ScapFL 2M alone or ScapFL 2M + Insig2 (pg. 301)

Figure 22 – SEC traces of point mutations in QC6 ScapFL 2M + Insig2 (pg. 302)

Figure 23 – Saturation binding and competition assays of ScapFL 2M with or without Insig2 (pg. 303)

Figure 24 – Additional binding assays with ScapFL 2M and ScapFL 2M + Insig2 (pg. 304)

Figure 25 – Reconstitution of ScapFL 2M + Insig2 complexes into nanodiscs (pg. 305)

Figure 26 – Sequence alignment of identified nanobody clones (pg. 306)

Figure 27 – Nanobody screening overview (pg. 307)

Figure 28 – Sypro Ruby stains of all 19 nanobody clones (pg. 308)

## CHAPTER EIGHT

Figure 1 – Example micrograph and 2D classes of ScapFL 2M + Hamster Insig 2 (pg. 317)

Figure 2 – Representative 3D classes arising from 2D classes shown in Figure 1 (pg. 318)

Figure 3 – Final best model of Hamster ScapFL 2M + Hamster Insig2 in saposin nanodiscs (pg. 319)

Figure 4 – Preliminary cryo-EM data from Hamster ScapFL 2M + Hamster Insig2 containing an N-terminal PGS fusion protein prior to TM1 (pg. 320)







## LIST OF ABBREVIATIONS

A<sub>2A</sub>R – adenosine A<sub>2A</sub> receptor

AOX – alcohol oxidase

ABC transporter – ATP-binding cassette transporter

ACN – acetonitrile

ATP – adenosine triphosphate

β<sub>2</sub>AR – beta-2 adrenergic receptor

BMG – buffered minimal glycerol media

BMGH – buffered minimal glycerol histidine media

BMM – buffered minimal methanol media

BRIL – *E. coli* apocytochrome *b<sub>562</sub>RIL*

CHO – Chinese hamster ovary

CHS – cholesterol hemisuccinate

CIP – alkaline phosphatase, calf intestinal

CDR – complementarity determining region

CHARMM – Chemistry at Harvard Macromolecular Mechanics

DDM – n-Dodecyl-β-D-Maltopyranoside

DEER – double electron electron resonance

DMF – dimethylformamide

DMNG – decyl maltose neopentyl glycol

DMSO – dimethyl sulfoxide

EDTA – Ethylenediaminetetraacetic acid

EM – electron microscopy

ER – endoplasmic reticulum

FA – formic acid

FACS – fluorescence activated cell sorting

FC13 – Fos-choline 13

FL – full-length

FPLC – fast protein liquid chromatography

FT – Fourier transform

FTIR – Fourier transform infrared spectroscopy

GDP – guanosine diphosphate

GFP – green fluorescent protein

GPCR – G protein-coupled receptor

GTP – guanosine triphosphate

HEK – human embryonic kidney

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HMQC – Heteronuclear Multiple Quantum Coherence

HPLC – high pressure liquid chromatography

HSQC – Heteronuclear Single Quantum Coherence

Ile – isoleucine

IMAC – immobilized metal-ion affinity chromatography

IPA – isopropanol

IPTG – Isopropyl  $\beta$ -D-1-thiogalactopyranoside

LB – Luria broth

LCP – lipidic cubic phase

LDL – low density lipoprotein

Leu – leucine

L1L7 – loop 1 – loop 7 construct

LMNG – lauryl maltose neopentyl glycol

MACS – magnetic activated cell sorting

MBP – maltose binding protein

MD – molecular dynamics

MDH – minimal dextrose histidine media

MF $\alpha$  – mating factor alpha

MP – membrane protein

MSP – membrane scaffold protein

MWCO – molecular weight cut-off

NMR – nuclear magnetic resonance

NOESY – Nuclear Overhauser Effect Spectroscopy

OD – optical density

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PGS – *Pyrococcus abyssi* glycogen synthase

POPC – 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine

POPE – 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine

PRE – paramagnetic relaxation enhancement

RMSD – root mean square deviation

RT – room temperature

SAXS – small angle X-ray scattering

Scap – SREBP-cleavage activating protein

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

SMA – styrene maleic acid

SMALP – styrene maleic acid lipoprotein particle

SRE – sterol regulatory element

SREBP – sterol regulatory element binding protein

SQ – single quantum

SSD – sterol sensing domain

TB – terrific broth

TCEP – tris(2-carboxyethyl)phosphine

TEV protease – tobacco etch virus protease

TFA – trifluoroacetic acid

TK – thymidine kinase

T<sub>m</sub> – melting temperature

T<sub>M</sub> – transmembrane

TROSY – Transverse Relaxation Optimized Spectroscopy

YPD – yeast extract peptone dextrose media

YPDS – yeast extract peptone dextrose sorbitol media

3Q – triple quantum



# CHAPTER ONE

## Introduction

### Background on membrane proteins

Cell membranes serve multiple functions crucial for proper cellular health and integrity. The plasma membrane forms a boundary between the cytoplasm and the extracellular environment, is semi-permeable, and facilitates active and passive transport of metabolites and signaling molecules in and out of the cell. Membranes also comprise the boundaries of intracellular organelles in eukaryotic cells, including the nucleus, the mitochondria, and the endoplasmic reticulum (ER).

Each of these membranes contain a specific mixture of lipids as well as a significant protein component made up of peripheral and integral proteins. Peripheral membrane proteins contain hydrophobic sequences or post-translational modifications that drive association with the periphery of a lipid bilayer for proper stability and function. Integral membrane proteins (hereby referred to as MPs) have entire segments of amino acids embedded through the cell membrane, typically  $\alpha$ -helical segments called transmembrane helices (TMs). Many MPs have multiple TMs, providing a large surface area to interact with the acyl chains of lipid molecules comprising the membrane bilayer. While this property is crucial for MP function, it makes them challenging to work with *in vitro*. Each specific sample requires

membrane mimetics and additional accommodations for biochemical and biophysical characterizations.

The work described in this dissertation will comprise two of my main focuses during my PhD research, both centering around membrane protein structure and function. The first project details my work enabling isotopic labeling and dynamics studies of G Protein-Coupled Receptors (GPCRs), while the second project describes structural efforts focused on the mammalian cholesterol homeostasis machinery. Both of these projects required significant optimization to study membrane proteins outside of their native membrane environment with a variety of biophysical techniques including Nuclear Magnetic Resonance Spectroscopy (NMR), X-ray crystallography, and cryo-Electron Microscopy (cryo-EM).

### **G Protein-Coupled Receptor Signaling**

G Protein-Coupled Receptors are the largest family of membrane proteins and play a crucial role in cellular signaling pathways. GPCRs sit at the plasma membrane and mediate the cell's response to a diverse array of ligands (e.g. light, hormones, etc). Binding of ligands on the extracellular face of the receptor leads to conformational changes on the cytoplasmic side that interface with cytosolic protein binding partners as heterotrimeric G proteins, G-protein coupled receptor kinases, and arrestins. Through activation by its bound receptor, these cytosolic proteins can then induce various downstream cellular signaling pathways in response to the

receptor-bound ligand. Since GPCRs play such a huge part of cell signaling, they constitute about a third of all current clinical drug targets, and having a full understanding of the mechanism of GPCR activation is key to the drug design process.

An interesting characteristic of GPCRs is that most of them exhibit a small level of basal activity, where they are able to signal even without a ligand bound. However, when ligands are bound, the signaling response within the cell varies based on the type of ligand bound. Antagonists reduce overall signaling and have two main types. An inverse agonist will reduce basal signaling levels, while a neutral antagonist will maintain basal signaling levels. Agonists also can have different efficacies. A full agonist will achieve that maximum signaling response when bound to the receptor, while a partial agonist will max out its signaling response at an attenuated level.

In the last fifteen years, our understanding of GPCRs has exploded due to breakthroughs in structural biology, primarily through X-ray crystallography (Rosenbaum *et al.*, 2007; Rasmussen *et al.*, 2011) and more recently, cryo-EM (Koehl *et al.*, 2018). Crystal structures of receptors showed the conformational changes associated with active and inactive states, as well as the molecular details of the orthosteric ligand binding pockets and allosteric binding sites that have guided structure-based drug design (SBDD). However, there is a large body of biochemical evidence that GPCRs are dynamic molecules (Manglik and Kobilka, 2014), and

crystal structures only offer a static glimpse into the lowest energy states at either end of an activation pathway. Due to their dynamic nature, receptors are able to adopt multiple states during their activation that are not represented in crystal structures.

It currently is still not fully understood how different types of ligands can exert different levels of signaling through a single receptor. A common hypothesis for this is that the binding of various ligands stabilizes distinct receptor conformations and affects the internal dynamics of the receptor, which then affects its signaling abilities. We sought to test this hypothesis and work towards understanding how the receptor dynamics change when bound to different ligands or effector proteins. Chapter 2 encompasses my efforts to develop a method allowing us to answer these questions through solution NMR, and chapters 3 through 5 focus on the application of this method to wild-type human GPCRs.

### **Cholesterol homeostasis machinery**

Each cellular membrane requires a specific lipid composition for proper function. In order to maintain the proper balance of lipids within cellular membranes, cells have a feedback mechanism in place to ensure appropriate lipid composition. One of the critical components in membranes that maintain their fluidity and integrity is cholesterol. Cholesterol can either be synthesized *de novo* by a series of enzymes or absorbed from the diet, and the levels must be maintained within a suitable level

for cellular health (Brown and Goldstein, 1986; Brown and Goldstein, 1997; Brown *et al.*, 2018).

In mammals, the feedback mechanism for maintenance of cholesterol levels is termed the SREBP pathway, centering around the eponymous Sterol Regulatory Element Binding Proteins. SREBPs are integral membrane proteins with a domain containing a transcription factor that targets the SRE promoter in the nucleus. There are about 30 genes under control of the SRE promoter and are all involved in lipid and cholesterol homeostasis (Horton *et al.*, 2003). The cell modulates activity of these genes by controlling the activation of the SREBPs through interactions with the protein Scap (SREBP cleavage-activating protein). Scap is constitutively bound to SREBPs in the ER, and is the key player in the SREBP pathway. Scap is able to bind and sense cholesterol levels in the ER membrane and facilitates SREBP activation via transportation to Golgi where proteases cleave and release the transcriptionally active domain to travel to the nucleus. Scap's transportation to the Golgi is inhibited by the presence of sufficient levels of cholesterol in the ER membrane, as well as interactions with an additional integral membrane protein, Insig, that acts as an ER "anchor" (Yang *et al.*, 2002). Together, Scap, Insig, and SREBP form a core unit that responds to levels of cholesterol in the cell and modulate gene expression that influences sterol synthesis and uptake (Brown *et al.*, 2018).

The SREBP pathway has been the main focus of the Brown and Goldstein lab at UTSW over multiple decades, and their efforts have generated a large amount of biochemical data towards understanding how these proteins work together to maintain cholesterol homeostasis. My contribution to this effort has been focused on obtaining structural information on this machinery to fully understand the mechanism of cholesterol homeostasis at a molecular level. Structures of Scap along with Insig and SREBP combined with functional data would help us understand how this machinery senses sterols in the membrane. Specifically, how and where does cholesterol bind to Scap? And then how does that binding ultimately influence Scap's conformation to control SREBP activation?

My dissertation research has worked towards answering these questions by overcoming multiple challenges with protein biochemistry and obtaining the first preliminary structural information on the Scap/Insig complex. While as of this writing we do not have a publishable structure of this complex, chapters 6-8 will address the efforts to get to our current state and detail future efforts moving forward.

## References

Brown, M. S. and Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* (1986) vol. 232 (4746) pp. 34-47

Brown, M. S. and Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* (1997) vol. 89 (3) pp. 331-340

Brown, M. S., Radhakrishnan, A., & Goldstein, J. L. (2017). Retrospective on Cholesterol Homeostasis: The Central Role of Scap. *Annual Review of Biochemistry*, 87(1), annurev-biochem-062917-011852-25.  
<http://doi.org/10.1146/annurev-biochem-062917-011852>

Horton, J. D., Shah, N. A., Warrington, J. A., Anderson, N. N., Park, S. W., Brown, M. S., & Goldstein, J. L. (2003). Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. *Proceedings of the National Academy of Sciences*, 100(21), 12027–12032.  
<http://doi.org/10.1073/pnas.1534923100>

Koehl, A., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J. M., et al. (2018). Structure of the  $\mu$ -opioid receptor–Gi protein complex. *Nature*, 558(7711), 547–552.  
<http://doi.org/10.1038/s41586-018-0219-7>

Manglik, A., & Kobilka, B. (2014). The role of protein dynamics in GPCR function: insights from the  $\beta$ 2AR and rhodopsin. *Current Opinion in Cell Biology*, 27, 136–143.  
<http://doi.org/10.1016/j.ceb.2014.01.008>

Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. 2011b. Crystal structure of the  $\beta_2$ -adrenergic receptor-Gs protein complex. *Nature* 477:549–555. DOI: <https://doi.org/10.1038/nature10361>

Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. 2007. GPCR engineering yields high-resolution structural insights into  $\beta_2$ -adrenergic receptor function. *Science* 318:1266–1273. DOI: <https://doi.org/10.1126/science.1150609>

## CHAPTER TWO

### Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*

#### Abstract

$^{13}\text{C}$  Methyl TROSY NMR spectroscopy has emerged as a powerful method for studying the dynamics of large systems such as macromolecular assemblies and membrane proteins. Specific  $^{13}\text{C}$  labeling of aliphatic methyl groups and perdeuteration are limited primarily to proteins expressed in *E. coli*, preventing studies of many eukaryotic proteins of physiological and biomedical significance. We demonstrate the feasibility of efficient  $^{13}\text{C}$  isoleucine  $\delta$ 1-methyl labeling in an established eukaryotic expression host, *Pichia pastoris*, and show that this method can be used to label a eukaryotic protein, actin, which cannot be expressed in bacteria. This approach will enable NMR dynamics studies of previously intractable targets.

#### Results and Discussion

Well-resolved two-dimensional (2D) NMR spectra are essential for measuring the dynamics of backbone and sidechain moieties within proteins. Such motions are fundamental to the function and regulation of large protein complexes and membrane proteins. NMR spectra of large macromolecules suffer from poor dispersion and line broadening due to rapid transverse relaxation of nuclear

magnetization and spectral crowding. To overcome this problem, proteins can be specifically labeled with  $^{13}\text{C}$  in the methyl groups of isoleucine, leucine, and valine residues using  $^{13}\text{C}$   $\alpha$ -ketoacid precursors in *E. coli* (Gardner and Kay 1997; Goto *et al.* 1999). When paired with selective protonation in an otherwise deuterated background (Rosen *et al.* 1996), this approach takes advantage of the favorable relaxation properties of  $^{13}\text{C}$ -methyl groups with the application of Transverse Relaxation Optimized Spectroscopy (TROSY) (Pervushin *et al.*, 1997; Ollerenshaw *et al.*, 2003). However these methods have remained unavailable for many eukaryotic proteins due to poor expression and folding in *E. coli* resulting from lack of required chaperones, lack of proper post-translational modifications, or improper membrane composition.

Several different eukaryotic hosts, including fungi (Miyazawa-Onami *et al.* 2013), insect cells (Nygaard *et al.* 2013; Kofuku *et al.* 2014), and mammalian cells (Werner *et al.* 2008), have been used to overexpress proteins for NMR. While these systems have succeeded in producing amino acid-specific and uniformly  $^{15}\text{N}$  or  $^{13}\text{C}$  labeled material (Chen *et al.* 2005; Fan *et al.* 2011; Gossert *et al.* 2011; Strauss *et al.* 2005; Hansen *et al.* 1992), the high expense and difficulties in perdeuteration have limited their widespread use for larger eukaryotic proteins. The methylotrophic yeast *Pichia pastoris* is a well established expression host (Cereghino and Cregg 2000) for proteins that cannot be made in *E. coli*- eukaryotic membrane proteins such as ATP transporters (Lee *et al.* 2002), ion pumps (Strugatsky *et al.* 2003) and G-protein

coupled receptors (Shimamura *et al.* 2011; Hino *et al.* 2012) have been successfully overexpressed in and purified from this organism. Genetic manipulation, transformation, and growth of *P. pastoris* are more rapid than for higher eukaryotes such as insect cells and mammalian cells.

Overexpression using the tightly regulated AOX1 promoter often yields milligram quantities of recombinant protein per liter of suspension culture (Cereghino and Cregg 2000). *P. pastoris* is also favorable for NMR studies given its ability to grow on defined minimal media, uptake isotope-containing precursors, and efficiently incorporate deuterium at non-exchangeable sites (Morgan *et al.* 2000). Despite conservation of branched-chain amino acid biosynthesis pathways from *E. coli* (Figure 1), site-specific methyl labeling using  $\alpha$ -ketoacid precursors has not been reported in *P. pastoris*.

We initially explored the use of  $^{13}\text{C}$ -methyl  $\alpha$ -ketobutyrate in *P. pastoris* cultures to label maltose binding protein (MBP) with  $^{13}\text{C}$  at the  $\delta^1$ -methyl groups of Isoleucine (Ile) residues. MBP has well-characterized  $^1\text{H}$ - $^{13}\text{C}$  2D NMR spectra (Gardner *et al.* 1998) and is highly expressed in *P. pastoris* (Li *et al.* 2010). We collected  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear single quantum coherence (HSQC) spectra on MBP that was labeled by addition of  $^{13}\text{C}$ -methyl  $\alpha$ -ketobutyrate to the culture media (Figure 2a).

Resonances for all 22 Ile  $\delta^1$ -methyl groups of MBP (Gardner *et al.* 1998) are observed in our spectrum (Fig. 2a, Fig. S1), while little signal is present in other

regions (indicating lack of “bleed-through” of the isotope into other amino acids - see Fig. 3). Based on tryptic peptide mass spectra (Figure 2b), we estimate the efficiency of incorporation for the  $\alpha$ -ketobutyrate-derived  $^{13}\text{C}$  methyl probe to be  $51\pm 7\%$  in a protonated background (see Supporting Information). The power of TROSY to obtain spectra of high-molecular weight species can only be exploited in the context of partial or full deuteration (Gardner *et al.* 1997; Wider and Wüthrich 1999; Ruschak and Kay 2010), which eliminates dipolar relaxation effects of surrounding protons on a given  $^{13}\text{C}$ -methyl spin system. To assess simultaneous  $^{13}\text{C}$  methyl labeling and perdeuteration in our system, we made samples of MBP in both *P. pastoris* and *E. coli*. We quantified the level of Ile $\delta$ 1 labeling in *P. pastoris*-derived deuterated MBP by comparing intensities to a concentration-matched *E. coli* sample (with assumed full incorporation at Ile $\delta$ 1 sites and deuteration), yielding a labeling efficiency of  $45\pm 6\%$  (Figure S2). The total deuteration level of *P. pastoris*-expressed MBP was estimated at 90% through ESI-LC-MS analysis (Figure S3; a comparison of labeling efficiency and yields of recombinant MBP from *P. pastoris* vs. *E. coli* is shown in Figure S4). Addition of  $\alpha$ -ketoisovalerate led to very modest labeling of leucine  $\delta$ - and valine  $\gamma$ -methyl groups ( $< 5\%$ ), suggesting that these sites could be labeled with significant future optimization.

The impetus for using *P. pastoris* for  $^{13}\text{C}$  methyl labeling is to access proteins that are not amenable to expression and purification from *E. coli* – for example, the eukaryotic cytoskeletal protein actin. Actin’s capacity to shuttle between monomeric

and polymeric states arises from its conformational dynamics between distinct globular and filamentous forms (Oda *et al.* 2010; Fujii *et al.* 2010; Nair *et al.* 2008). NMR dynamics measurements would represent a significant new tool to study the biophysics of actin polymerization and interactions with regulatory molecules (Schmid *et al.* 2004; Kudryashov and Reisler 2013). While the structure of actin monomer has been determined by X-ray crystallography (Otterbein *et al.* 2001, Rould *et al.* 2006) and actin filaments have been characterized by electron microscopy (Fujii *et al.* 2010, Ecken *et al.* 2015), expression of isotopically labeled actin for NMR has not been reported. Actin cannot be expressed at high levels in *E. coli* because of the lack of eukaryotic chaperone systems that are necessary for folding.

Biophysical characterization of actin is intrinsically difficult because actin polymerizes at concentrations above 100 nM. We therefore attempted to express a non-polymerizable *Drosophila* 5C actin (94% identity to human actin) mutant in *P. pastoris* with mutations that impair the fast growing “barbed-end” of the filament (Zahm *et al.* 2013). However, the mutant proved toxic, presumably because it interferes with the polymerization of endogenous actin. To solve this problem, we generated a C-terminal fusion to human thymosin  $\beta$ 4, an actin binding protein that blocks the intact, slow-growing “pointed-end” and thus ameliorates toxicity (Noguchi *et al.* 2007). This strategy resulted in high expression levels (10 mg/L of culture) and purification to homogeneity (Supporting Information and Fig. S5).

A representative HMQC spectrum of  $^{13}\text{C}$ - $\delta^1$ -methyl -Ile actin is shown in Figure 4a. Notably, for a protein with 27 Ile residues, we observe 30 peaks in the  $^1\text{H}$ - $^{13}\text{C}$  spectrum, likely reflecting slow-exchange processes at some sites. Taking advantage of the ability to highly deuterate proteins in *P. pastoris*, we repeated expression of *Drosophila* 5C actin in cultures where cells were adapted to  $\text{D}_2\text{O}$ -containing media prior to induction, resulting in 2.5 mg/L of  $^{13}\text{C}$ -methyl perdeuterated actin. Spectral lines in the  $^1\text{H}$ - $^{13}\text{C}$  TROSY HMQC of the deuterated sample were much narrower than lines in the HMQC spectrum of non-deuterated actin (Figures 4a and 4b).

Future use of TROSY NMR methods to study the dynamics of high-MW mammalian protein complexes and membrane proteins will depend on the tractability of isotope incorporation. We have demonstrated efficient incorporation of  $^{13}\text{C}$  at the Ile  $\delta^1$ -methyl groups of proteins expressed in *P. pastoris*, a robust eukaryotic expression host. In conjunction with perdeuteration, we acquired high-quality  $^1\text{H}$ - $^{13}\text{C}$  methyl TROSY spectra on *Drosophila* actin, which were unobtainable before. This development, along with similar approaches using other yeast systems (Miyazawa-Onami *et al.* 2013), will allow 2D NMR spectroscopy to be applied to many previously intractable proteins.

## Materials and methods

### *MBP expression and purification*

The gene encoding *E. coli* maltose binding protein with a C-terminal hexahistidine tag (41.5 kDa) was inserted into the pPICZ B vector (Invitrogen) under control of the AOX1 promoter and transformed into the KM71H strain. Expression was screened with increasing concentrations of Zeocin selection marker (100- 1000 mg/mL), and a high expressing clone was chosen for future experiments. Cells were cultured at 28°C in BMG media (1% glycerol, 100 mM potassium phosphate pH 6.0, 1.34% YNB, 4 x 10<sup>-5</sup>% biotin) with shaking at 250rpm. Once cells reached saturation (~36hrs), cells were spun down and resuspended in an equal volume of BMM media without methanol and supplemented with DMSO (100mM potassium phosphate pH 6.0, 1.34% YNB, 4 x 10<sup>-5</sup>% biotin, 2.5% DMSO). Cells were incubated for an additional 4-5 hrs to allow for metabolism of residual glycerol. One hour prior to induction with methanol, a sterile- filtered solution of  $\alpha$ -ketobutyric acid (methyl-<sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; Cambridge Isotope) was added to a final concentration of 100 mg/L. Protein expression was induced with 0.5% methanol and was maintained by additions of 0.5% methanol every 12 hours. After 36 hours, cells were harvested by centrifugation at 5000g and frozen at -80°C. Pichia were resuspended in lysis buffer containing 200 mM sodium chloride, 20 mM Tris pH 7.4, and protease inhibitors (160  $\mu$ g/mL benzamidine, 2.5  $\mu$ g/mL leupeptin, 1 mM PMSF,

1  $\mu\text{M}$  E-64) and lysed by five passes through a microfluidizer (Microfluidics M-110P) at 25,000 psi. Insoluble material was removed by centrifugation at 10,000 g, and MBP was purified through Ni-NTA and amylose affinity chromatography, followed by size exclusion chromatography using a Superdex 200 10/300 column (GE Healthcare). MBP was concentrated to 500  $\mu\text{L}$  (measured at 225  $\mu\text{M}$  by UV Abs<sub>280</sub>) in Amicon 30kDa MWCO concentrators. An equimolar amount of  $\beta$ -cyclodextrin (Acros Organics) was added to the MBP sample, and the mixture was dialyzed overnight against 20 mM HEPES pH 7.2 at 4°C.

To generate the deuterated MBP sample, cells were adapted to deuterated media as follows. 100  $\mu\text{L}$  of an initial starter culture in protonated BMGH was used to inoculate 50 mL of BMGH media containing 90% D<sub>2</sub>O (Cambridge Isotope Laboratories) and grown to an OD of 7. 100  $\mu\text{L}$  of the 90% D<sub>2</sub>O culture was used to inoculate 50 mL BMGH containing 100% D<sub>2</sub>O, and again grown to an OD<sub>600</sub> of 7. 100  $\mu\text{L}$  of the fully deuterated culture was used to inoculate 50 mL BMGH containing 100% D<sub>2</sub>O and 1% deuterated d<sub>8</sub>- glycerol (CIL) in place of protonated glycerol. This culture was grown to an OD<sub>600</sub> of 7 and used in its entirety to inoculate 250 mL of BMGH containing 100% D<sub>2</sub>O and 1% d<sub>8</sub>-glycerol and grown to an OD<sub>600</sub> of 20. This culture was spun down at 3000 g for 10 min and resuspended in 250 mL methanol-free BMMH containing 100% D<sub>2</sub>O. The culture continued shaking for 12 hours to ensure full metabolism of glycerol. One hour prior to induction, a sterile-filtered solution of  $\alpha$ -ketobutyric acid (methyl-<sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; CIL) in D<sub>2</sub>O was added

to a final concentration of 200 mg/L. Induction was achieved by addition of d<sub>4</sub>-methanol (CIL) to a final concentration of 0.5%. Expression proceeded for a total of 36 hours at 28°C, with a supplementation of 0.5% 100% d<sub>4</sub>-methanol every 12 hrs. Cells were harvested by centrifugation and purified in a manner identical to that used for the protonated sample, except that the final gel filtration buffer was made using 100% D<sub>2</sub>O instead of water and the final dialysis buffer prior to NMR was made with 100% D<sub>2</sub>O.

Deuterated MBP generated in *E. coli* was grown as described previously (Gardner *et al.* 1998). MBP was purified from 1L of *E. coli* in an identical manner to deuterated MBP expressed in *P. pastoris*.

#### *Actin expression and purification*

The gene encoding the polymerization-incompetent *Drosophila* 5C actin mutant [D287A, V288A, D289A] was cloned into pPICZ B (Invitrogen) using EcoRI and NotI restriction sites. The construct used to generate the protonated actin sample contained a C-terminally fused human  $\beta$ -thymosin-4 followed by a hexahistidine tag. The construct was transformed via electroporation into the GS115 strain of *Pichia pastoris*. Expression was screened with increasing concentrations of Zeocin selection marker (0.5-2 mg/mL) and a high expressing clone was chosen for future experiments. A single colony was selected from a BMGH (BMG + 0.004% histidine) agar plate streaked with the high-expressing clone, and grown at 30°C to

an OD<sub>600</sub> of 5-7 in 35 mL BMGH media in a 125 mL baffled flask. The entire starter culture was used to inoculate 2L of BMGH media, and cells were grown until OD<sub>600</sub> ~17. Cells were collected by centrifugation at 3000 g for 15 minutes and resuspended in 4L of BMMH without methanol (BMM + 0.004% histidine).

Immediately after resuspension, a sterile-filtered solution of  $\alpha$ -ketobutyric acid (methyl- <sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; Cambridge Isotope Laboratories) was added to a final concentration of 100 mg/L. Induction was initiated one hour later by addition of methanol to a final concentration of 0.5% and expression continued for 24 hours.

Cells were harvested by centrifugation at 3000 g for 10 minutes and resuspended in 200 mL 50 mM Tris pH 8.0, 250 mM potassium chloride, 0.1 mM calcium chloride, 0.2 mM ATP, and 0.5 mM  $\beta$ -mercaptoethanol, supplemented with protease inhibitors (1  $\mu$ g/mL leupeptin, 500 ng/mL pepstatin, 1 mM benzamidine, 1  $\mu$ g/mL antipain, and 1 mM PMSF). Cells were lysed by four passes through a microfluidizer (Microfluidics M-110P) at 25,000 psi. Insoluble material was removed by centrifugation at 50,000 g, and the supernatant was subjected to Ni-NTA affinity chromatography. The resulting material was further purified on a 4 mL Source 15Q anion exchange column. Fractions containing actin were pooled and concentrated to 1 mL and exchanged via size exclusion chromatography (Figure S5) on a Superdex 200 10/300 column (GE Healthcare) into a buffer containing 20 mM potassium phosphate pH 7.0, 50 mM potassium chloride, 0.2 mM ATP, 1 mM TCEP, 0.1 mM calcium chloride, and 1 mM sodium azide.

To generate the deuterated actin sample, cells were grown in BMGH and were adapted to increasing concentration of D<sub>2</sub>O in the media as follows. 100 μL of the initial starter culture was used to inoculate 35 mL of BMGH media containing 90% D<sub>2</sub>O (Cambridge Isotope Laboratories) and grown to an OD of 5-7. 100 μL of the 90% D<sub>2</sub>O culture was used to inoculate 35 mL BMGH containing 100% D<sub>2</sub>O, and again grown to an OD<sub>600</sub> of 5-7. 100 μL of the fully deuterated culture was used to inoculate 35 mL BMGH containing 100% D<sub>2</sub>O and 1% deuterated d<sub>8</sub>-glycerol (Sigma) in place of protonated glycerol. This culture was grown to an OD<sub>600</sub> of 5-7 and used in its entirety to inoculate 1L of BMGH containing 100% D<sub>2</sub>O and 1% d<sub>8</sub>-glycerol and grown to an OD<sub>600</sub> of 17. This culture was spun down at 3000 g for 10 min and resuspended in 30 mL methanol-free BMMH containing 100% D<sub>2</sub>O. The resuspended culture was distributed equally among three 4L flasks, each containing 1L of methanol-free BMMH. The cultures continued shaking for 12 hours to ensure full metabolism of glycerol. One hour prior to induction, a sterile- filtered solution of α-ketobutyric acid (methyl-<sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; Cambridge Isotope) in D<sub>2</sub>O was added to a final concentration of 200 mg/L. Induction was achieved by addition of d<sub>4</sub>-methanol (Sigma) to a final concentration of 0.5%. Expression proceeded for a total of 48 hours at 30°C, with a supplementation of 5 mL 100% d<sub>4</sub>-methanol at 24 hrs. Cells were harvested by centrifugation and purified in a manner identical to that used for the protonated sample, except that the final gel filtration buffer was made using 100% D<sub>2</sub>O instead of water.

### *Mass spectrometry*

Purified labeled and unlabeled MBP samples were reduced in 100 mM DTT and alkylated with iodoacetamide. Samples were trypsinized at 37 °C overnight. Digestion products were analyzed on a Shimadzu IT-TOF LC-MS system equipped with NESP-100 NANO-ESI ion source. LC was run on a Chromolith CapROD RP-18e column with a H<sub>2</sub>O/Acetonitrile gradient containing 0.1% Formic Acid. Incorporation efficiency of <sup>13</sup>C at the δ1 position of Isoleucine residues was quantified in the following manner: Five peptides were identified that had high signal/noise in the tryptic LC-MS data for both labeled and unlabeled MBP, and which had a unique Isoleucine present in the sequence (z = +2 or +3). For each matched peptide, an isotopic distribution was observed, corresponding to increasing numbers of natural- abundance <sup>13</sup>C atoms incorporated into the peptide – M+1 has one <sup>13</sup>C atom, M+2 has two <sup>13</sup>C atoms, etc. The specific labeling of an Ile residue with <sup>13</sup>C shifts this distribution to the right by 1/z. We normalized the intensities of the peaks in each of these distributions by dividing by the intensity of the M+0 peak (i.e. no <sup>13</sup>C atoms) – see Fig. 2b. The % incorporation of <sup>13</sup>C by addition of the label was then determined by: [(Normalized intensity M+1 labeled) – (Normalized Intensity M+1 unlabeled)] ÷ (Normalized Intensity M+1 labeled). Similar values could be calculated for the M+2 peaks. Reported efficiency is the average value for the M+1 and M+2 peaks for the five peptides.

Online LC-MS on intact MBP samples was performed as follows. Purified deuterated  $^{13}\text{C}$ -methyl-Ile MBP samples from *P. pastoris* and *E. coli* were dialyzed into  $\text{H}_2\text{O}$  buffer containing 10 mM HEPES pH 7.0, 10 mM DTT, and 2M urea. Samples were directly subjected to online LC-MS analysis. The mobile phase compositions were as follows: solvent A – 0.025% trifluoroacetic acid (TFA), 0.3% formic acid (FA), and 20% acetonitrile (ACN) in water; solvent B – 0.025% TFA, 0.3% FA, and 20% isopropanol (IPA) in ACN. After online desalting at 0% B for 5 min at a flow rate of 200  $\mu\text{L}/\text{min}$ , proteins were separated with a 30 min gradient (from 20 to 40% B). LC was performed on an Agilent 1100 Series HPLC, with a 0.5x75 mm, Poroshell 300SB-C8 column with 300 Å pore size, 5  $\mu\text{m}$  diameter particles (Agilent, Santa Clara, CA), heated to 70 °C. Column eluate was directed into the mass spectrometer by a TriVersa NanoMate nano-electrospray robot (Advion Biosystems, Ithica, NY). The LTQ Orbitrap XL (Thermo Fisher, Waltham, MA) was operated at a resolving power of 15,000 at  $m/z$  400, positive ion mode, with 700-2,000 mass-to-charge ( $m/z$ ) range. Intact mass data was collected with 300 ms maximum ion injection time, 3 microscans per data point, and  $2 \times 10^5$  automatic gain control (AGC). Average masses reported are by deconvolution of low res charge states.

### *NMR Spectroscopy*

All NMR data on MBP and actin were acquired on a Varian Inova 800 MHz NMR spectrometer equipped with a cryogenically-cooled  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  probe operating

at 25 °C. NMR experiments on MBP were carried out on a sample containing a 1:1 ratio of labeled MBP and  $\beta$ -cyclodextrin (225  $\mu$ M each) in 20 mM HEPES pH 7.2, 10% D<sub>2</sub>O. <sup>1</sup>H-<sup>13</sup>C HSQC spectra were collected with spectral widths of 11990 Hz and 4026 Hz and acquisition times of 85 ms and 24 ms in the <sup>1</sup>H and <sup>13</sup>C dimensions, respectively. An inter-scan delay of 1 s was employed between successive transients. Total acquisition time was approximately 2.2 hr.

Actin NMR data were acquired on samples at 180  $\mu$ M (protonated, Ile $\delta$ 1-methyl <sup>1</sup>H/<sup>13</sup>C labeled) and 150  $\mu$ M (deuterated, Ile  $\delta$ 1-methyl <sup>1</sup>H/<sup>13</sup>C labeled) concentrations in 90%H<sub>2</sub>O/10%D<sub>2</sub>O and 100% D<sub>2</sub>O, respectively. <sup>1</sup>H-<sup>13</sup>C HMQC (Bax *et al.* 1983; Ollerenshaw *et al.* 2003) spectra were acquired with 9615.4 Hz and 4026 Hz in <sup>1</sup>H and <sup>13</sup>C dimensions, respectively. The acquisition times in directly detected <sup>1</sup>H dimension and indirectly detected <sup>13</sup>C dimension were 106 ms and 16 ms, respectively. An inter-scan delay of 1 s was employed between successive transients during data acquisition.

Data were processed using NMRPipe (Delaglio *et al.* 1995). Data sets were zero-filled prior to Fourier transformation. The directly and indirectly detected time domain data were processed by applying a 90° phase-shifted squared sine bell or a Gaussian filter.

In order to compare the incorporation of <sup>1</sup>H/<sup>13</sup>C at Ile  $\delta$ 1-methyl positions in a deuterated background, MBP was overexpressed in *E.coli* and *P. pastoris*, purified and complexed with a 1:1 molar ratio of  $\beta$ - cyclodextrin (Acros Organics), and

dialyzed overnight against 20mM HEPES pH 7.2 prepared in 100% D<sub>2</sub>O. NMR samples of each protein were prepared at identical concentrations (240 μM). The 2D <sup>1</sup>H-<sup>13</sup>C HMQC spectra on both samples were acquired at 298 K using identical acquisition parameters. The <sup>1</sup>H-<sup>13</sup>C HMQC spectra were acquired with sweep widths of 9615.4 Hz and 4026 Hz in <sup>1</sup>H and <sup>13</sup>C dimensions, respectively. Acquisition times in the directly detected <sup>1</sup>H dimension and indirectly detected <sup>13</sup>C dimension were 106 ms and 16 ms, respectively. An inter-scan delay of 1 s was employed between successive transients during data acquisition.

Both datasets were processed identically using NMRPipe (Delaglio *et al.* 1995). Ten well-resolved cross peaks in methyl HMQC spectra were identified for intensity analysis. The intensities of individual cross-peaks were measured in terms of peak height (data not shown), peak volume and signal-to-noise ratio (S/N). The intensity analysis was performed using NMRPipe as well as using the Analysis module in CCPNMR (Vranken *et al.* 2005). The incorporation of <sup>1</sup>H/<sup>13</sup>C at Ile δ1-methyl positions in *P. pastoris* MBP was measured relative to *E.coli* MBP.

## References

Bax AD, Griffey RH, Hawkins BL (1983) Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J Magn Reson* 55(2):301-315

Cereghino JL, Cregg JM (2000) Heterologous expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24(1):45-66

Chen C-Y, Cheng C-H, Chen Y-C, Lee J-C, Chou S-H, Huang W, Chuang W-J (2005) Preparation of amino-acid-type selective isotope labeling of protein expressed in *Pichia pastoris*. *Proteins* 62(1):279-287

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax AD (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293

Ecken J, Müller M, Lehman W, Manstein DJ, Penczek PA, Raunser, S (2015) Structure of the F-actin-tropomyosin complex. *Nature* 519(7541):114-117

Fan Y, Shi L, Ladizhansky V, Brown LS (2011) Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment. *J Biomol NMR* 49(2):151-161  
Fujii T, Iwane AH, Yanagida T, Namba K (2010) Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* 467(7316):724-728

Gardner KH, Kay LE (1997) Production and incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $1\text{H}-\delta^1$  methyl) isoleucine into proteins for multidimensional NMR studies. *J Am Chem Soc* 119(32):7599-7600

Gardner KH, Rosen MK, Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* 36(6):1389-401

Gardner KH, Zhang X, Gehring K, Kay LE (1998) Solution NMR Studies of a 42 KDa *Escherichia coli* Maltose Binding Protein/ $\beta$ -Cyclodextrin Complex: Chemical Shift Assignments and Analysis. *J Am Chem Soc* 120(45):11738-11748

Gardner KH, Kay LE (1998) The Use of  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  Multidimensional NMR to Study the Structure and Dynamics of Proteins. *Ann Rev Biophys Biomol Struct* 120(45):11738-11748

- Gossert AD, Hinniger A, Gutmann S, Jahnke W, Strauss A, Fernández C (2011) A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *J Biomol NMR* 51(4): 449-456
- Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta^1$ ) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ -,  $^2\text{H}$ -labeled proteins. *J Biomol NMR* 13(4):369-74
- Hansen AP, Petros AM, Mazar AP, Pederson TM, Rueter A, Fesik, SW (1992) A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells. *Biochemistry* 31(51):12713-8
- Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, et al (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482(7384):237-240
- Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Mizumura T, Suzuki S, Shimada I (2014) Functional Dynamics of Deuterated  $\beta_2$ -Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angew Chem Int Ed* 53(49):13376-13379
- Kudryashov DS, Reisler E (2013) ATP and ADP actin states. *Biopolymers* 99(4):245-256
- Lee J-Y, Urbatsch IL, Senior AE, Wilkens S (2002) Projection Structure of P-glycoprotein by Electron Microscopy: Evidence for a Closed Conformation of the Nucleotide Binding Domains. *J Biol Chem* 277(42):40125-40131
- Li Z, Leung W, Yon A, Nguyen J, Perez VC, Vu J, Giang W, Luong LT, Phan T, Salazar KA, et al (2010) Secretion and proteolysis of heterologous proteins fused to the *Escherichia coli* maltose binding protein in *Pichia pastoris*. *Protein Expression Purif* 72(1):113-124
- Miyazawa-Onami M, Takeuchi K, Takano T, Sugiki T, Shimada I, Takahashi H (2013) Perdeuteration and methyl-selective ( $^1\text{H}$ ), ( $^{13}\text{C}$ )-labeling by using a *Kluyveromyces lactis* expression system. *J Biomol NMR* 57(3):297-304
- Morgan WD, Kragt A, Feeney J (2000) Expression of deuterium-isotope-labelled protein in the yeast *pichia pastoris* for NMR studies. *J Biomol NMR* 17(4):337-47
- Nair UB, Joel PB, Wan Q, Lowey S, Rould MA, Trybus KM (2008) Crystal Structures of Monomeric Actin Bound to Cytochalasin D. *J Mol Biol* 384(4):848-864

Noguchi TQP, Kanzaki N, Ueno H, Hirose K, Uyeda TQP. (2007) A Novel System for Expressing Toxic Actin Mutants in Dictyostelium and Purification and Characterization of a Dominant Lethal Yeast Actin Mutant. *J Biol Chem* 282(38):27721-27727

Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A, Pan AC, Liu CW, Fung JJ, Bokoch MP, et al (2013) The Dynamic Process of  $\beta$ 2-Adrenergic Receptor Activation. *Cell* 152(3):532-542

Oda T, Iwasa M, Aihara T, Maéda Y, Narita A (2009) The nature of the globular- to fibrous-actin transition. *Nature* 457(7228):441-445

Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* 41(10):843-852

Otterbein LR, Graceffa P, Dominguez R. (2001) The crystal structure of uncomplexed actin in the ADP state. *Science* 293(5530):708-11

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in  $^{13}\text{C}$ -labeled proteins. *Proc Natl Acad Sci USA* 94(23):12366-71

Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* 263(5):627-636

Rould MA, Wan Q, Joel PB, Lowey S, Trybus KM (2006) Crystal Structures of Expressed Non-polymerizable Monomeric Actin in the ADP and ATP States. *J Biol Chem* 281(42):31909-31919

Ruschak AM, Kay LE (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR* 46(1):75-87

Schmid MF, Sherman MB, Matsudaira P, Chiu W. (2004) Structure of the acrosomal bundle. *Nature* 431(7004):104-7

Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, et al. (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475(7354):65-70

Strauss A, Bitsch F, Fendrich G, Graff P, Knecht R, Meyhack B, Jahnke, W (2005) Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *J Biomol NMR* 31(4):343-349

Strugatsky D, Gottschalk K-E, Goldshleger R, Bibi E, Karlish SJD (2003) Expression of Na<sup>+</sup>,K<sup>+</sup>-ATPase in *Pichia pastoris*: Analysis of Wild Type and D369N Mutant Proteins by Fe<sup>2+</sup>-Catalyzed Oxidative Cleavage and Molecular Modeling. *J Biol Chem* 278(46):46064-46073

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* 59(4):687-96

Werner K, Richter C, Klein-Seetharaman J, Schwalbe H (2008) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* 40(1):49-53

Wider G, Wüthrich K (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* 9(5):594-601

Zahm JA, Padrick SB, Chen Z, Pak CW, Yunus AA, Henry L, Tomchick DR, Chen Z, Rosen MK (2013) The Bacterial Effector VopL Organizes Actin into Filament-like Structures. *Cell* 155(2):423-434

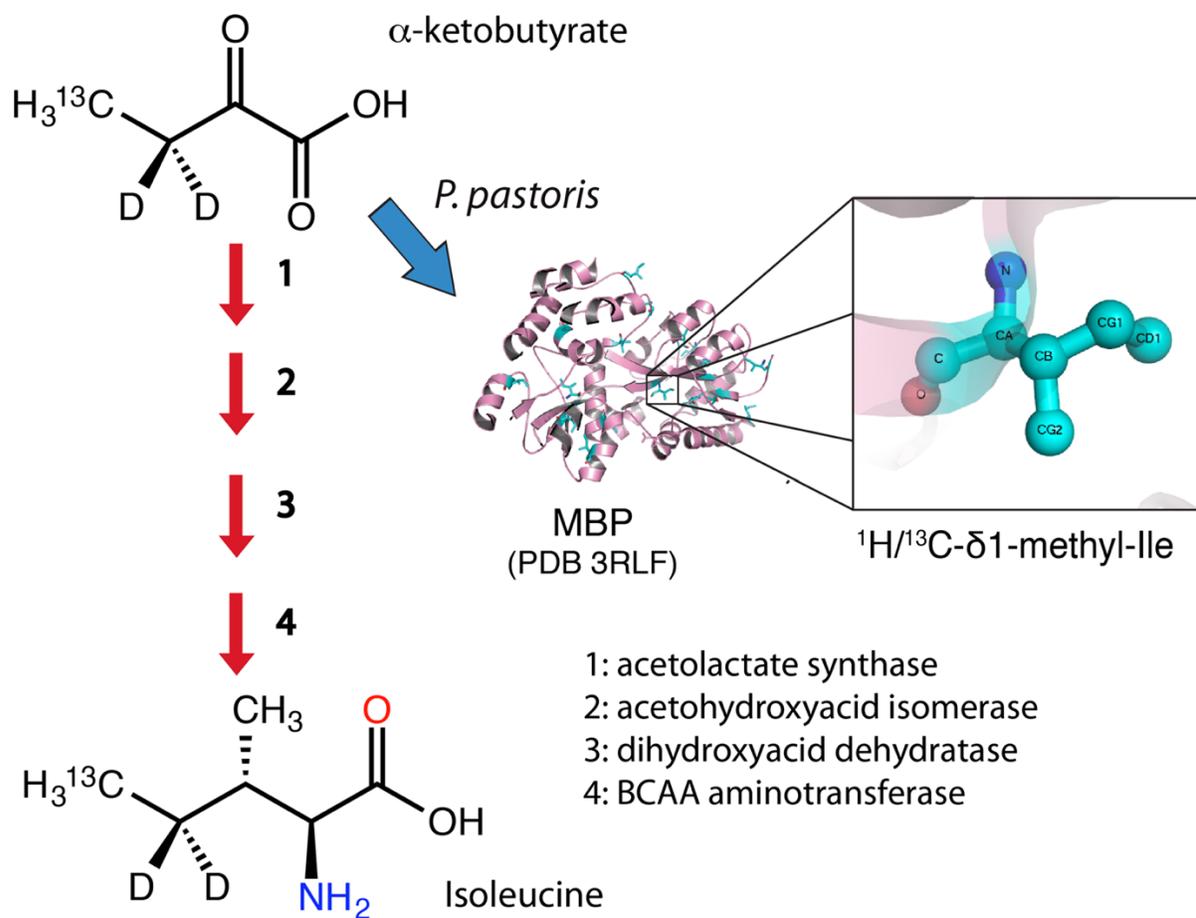


Fig. 1 Incorporation of  $^{13}\text{C}$ -methyl groups at the  $\delta$ 1 position of isoleucine residues in proteins expressed in *P. pastoris* (PDB: 3RLF)

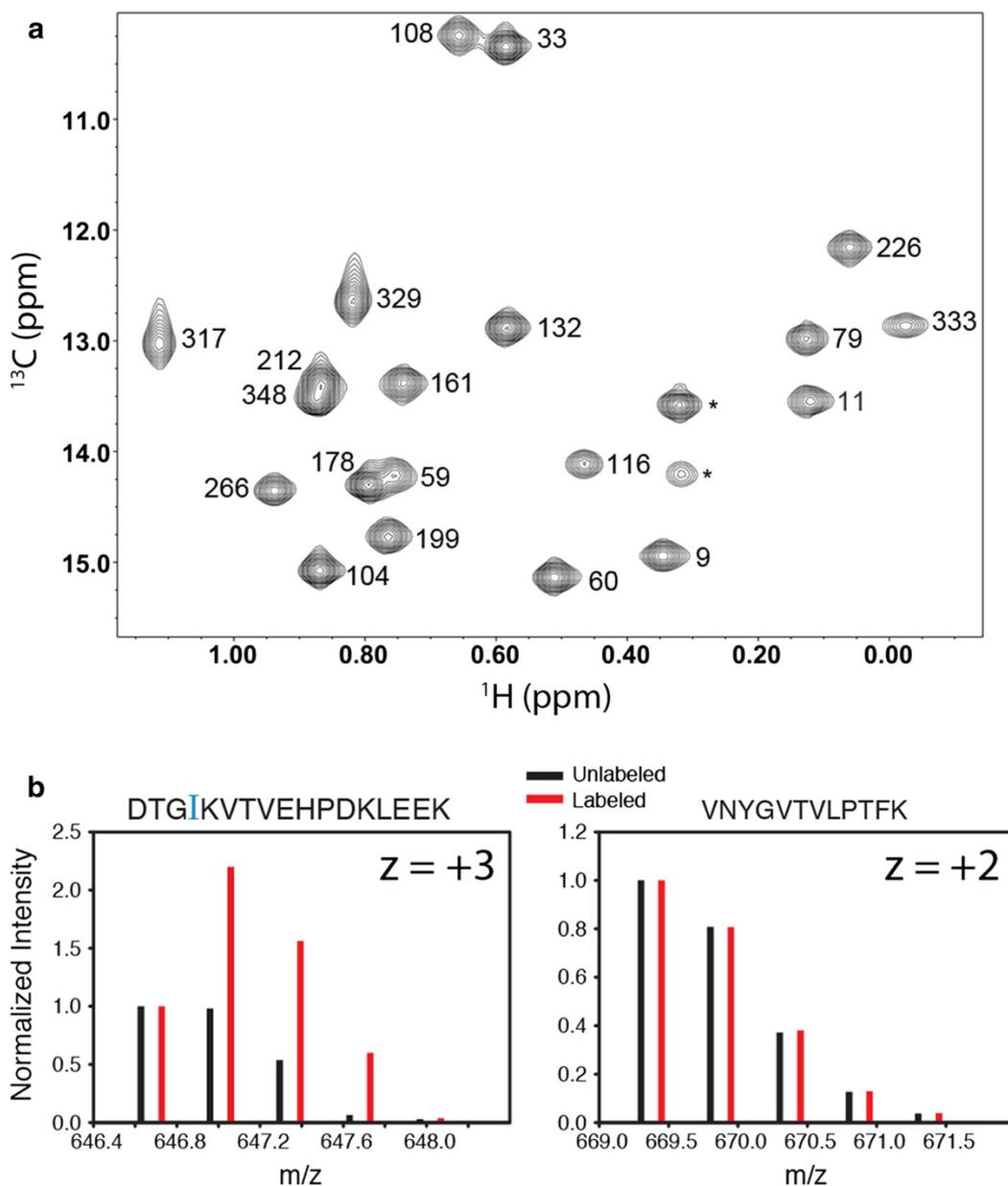
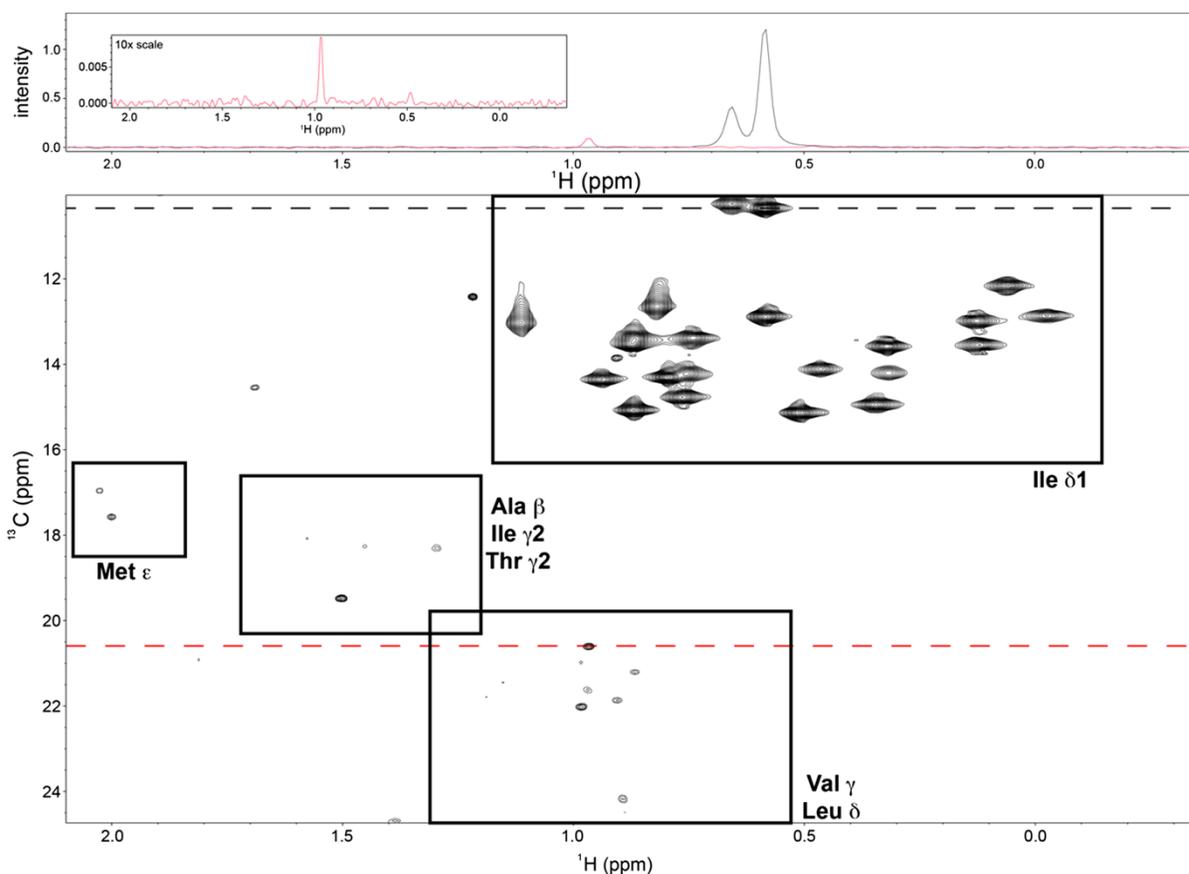
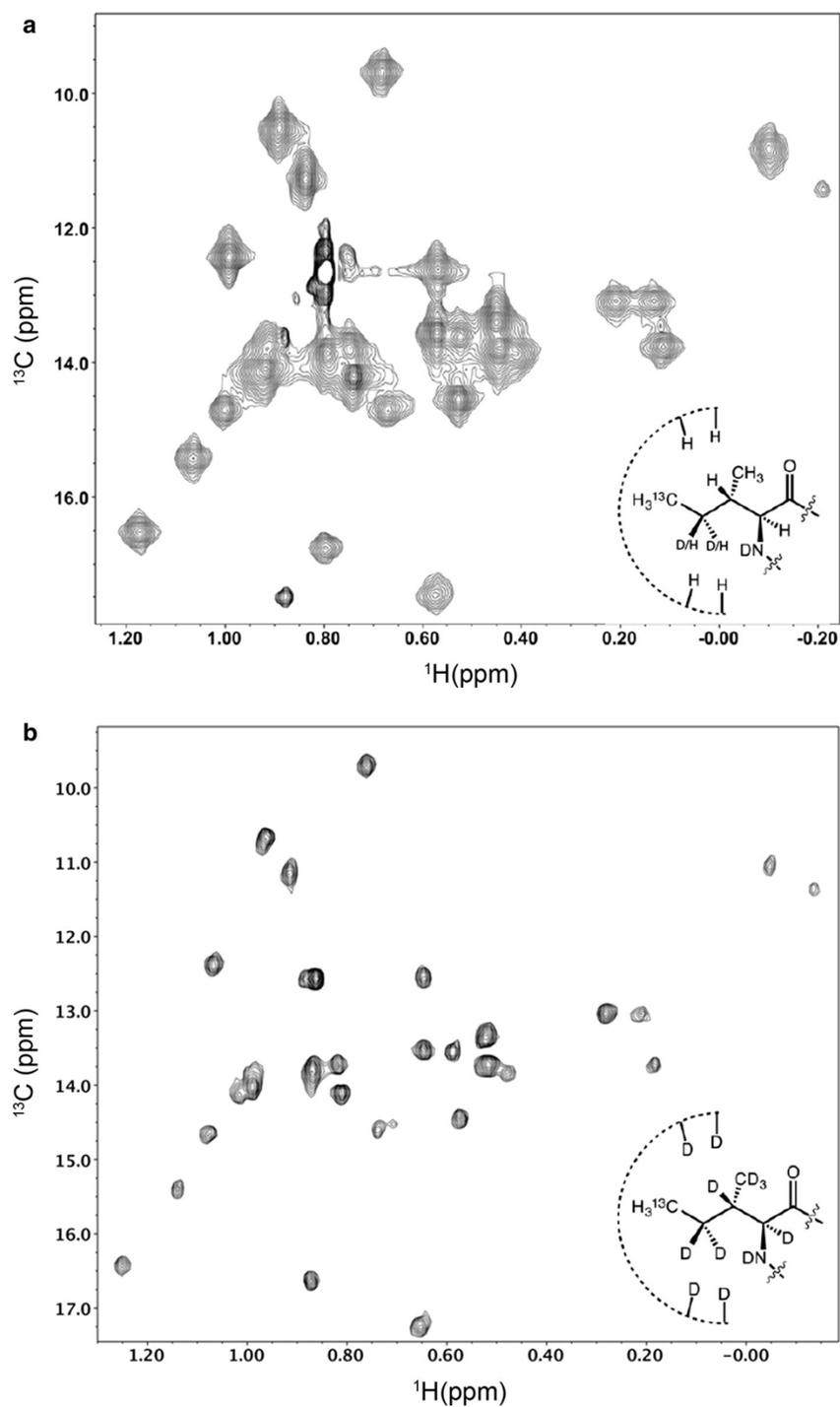


Fig. 2 Labeling of  $\delta$ 1-methyl groups of MBP expressed in *Pichia pastoris*. (a)  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of 225  $\mu\text{M}$  MBP labeled with  $\alpha$ -ketobutyrate. Spectrum was recorded at 25°C on a Varian 800 MHz spectrometer. Peaks corresponding to Ile  $\delta$ 1 methyl groups are labeled in reference to assigned spectra (Gardner *et al.* 1998). Two unassigned peaks likely arising from differences in constructs are denoted with an asterisk. (b) Mass spectra of tryptic peptides containing Ile (left) and not containing Ile (right).



**Fig. 3** Expanded view of the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of isoleucine  $\delta$ 1-methyl labeled maltose binding protein shown in Fig. 2a. Top panel shows horizontal slices of the 2D dataset (bottom panel), taken at approximately  $^{13}\text{C} = 10.3$  ppm (black; Ile  $\delta$ 1) and 20.6 ppm (red; Val/Leu) to show representative signal-to-noise in the spectrum for the labeled Ile  $\delta$ 1 methyl groups versus the unlabeled (natural abundance  $^{13}\text{C}$ ) methyl groups of other amino acids. Inset of the top panel shows the  $^{13}\text{C} = 20.6$  ppm trace at 10x vertical scale of the surrounding panel to provide a clearer sense of signal-to-noise. Signal-to-noise measurements for all 22 Ile  $\delta$ 1 methyl peaks resulted in an average S/N ratio of 280.



**Fig. 4** NMR spectra of *Drosophila* actin labeled and overexpressed in *Pichia pastoris*. (a)  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum of  $^{13}\text{C}$ -Ile  $\delta$ 1-methyl-labeled actin ( $180\ \mu\text{M}$ ). (b) TROSY-HMQC spectrum of perdeuterated,  $^{13}\text{C}$ -Ile  $\delta$ 1-methyl-labeled actin ( $150\ \mu\text{M}$ ). Spectra were recorded at  $25^\circ\text{C}$  on a Varian 800 MHz spectrometer.

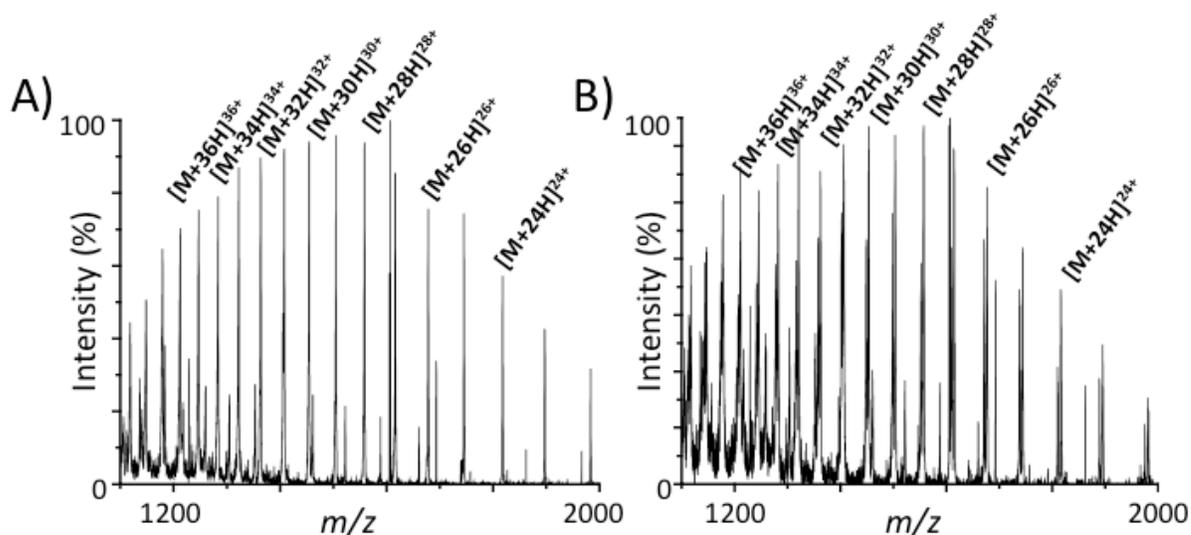
| <i>P. pastoris</i><br>(MBP res #) | 1H    | 13C   | <i>E. coli</i><br>(MBP res #) | 1H     | 13C     | $\delta$ 1H | $\delta$ 13C |
|-----------------------------------|-------|-------|-------------------------------|--------|---------|-------------|--------------|
| 9                                 | 0.34  | 14.96 | 9                             | 0.325  | 14.504  | 0.02        | 0.46         |
| 11                                | 0.12  | 13.56 | 11                            | 0.108  | 13.1151 | 0.01        | 0.44         |
| 33                                | 0.58  | 10.36 | 33                            | 0.576  | 10.246  | 0.00        | 0.11         |
| 59                                | 0.75  | 14.24 | 59                            | 0.739  | 13.813  | 0.01        | 0.43         |
| 60                                | 0.51  | 15.15 | 60                            | 0.408  | 14.627  | 0.10        | 0.52         |
| 79                                | 0.13  | 13    | 79                            | 0.171  | 12.326  | -0.04       | 0.67         |
| 104                               | 0.87  | 15.09 | 104                           | 0.856  | 14.654  | 0.01        | 0.44         |
| 108                               | 0.66  | 10.26 | 108                           | 0.65   | 10.048  | 0.01        | 0.21         |
| 116                               | 0.46  | 14.13 | 116                           | 0.462  | 13.724  | 0.00        | 0.41         |
| 132                               | 0.58  | 12.9  | 132                           | 0.564  | 12.401  | 0.02        | 0.50         |
| 161                               | 0.74  | 13.4  | 161                           | 0.761  | 13.139  | -0.02       | 0.26         |
| 178                               | 0.79  | 14.32 | 178                           | 0.857  | 13.918  | -0.07       | 0.40         |
| 199                               | 0.76  | 14.79 | 199                           | 0.749  | 14.231  | 0.01        | 0.56         |
| 212                               | 0.87  | 13.44 | 212                           | 0.872  | 13.1    | 0.00        | 0.34         |
| 226                               | 0.06  | 12.17 | 226                           | 0.072  | 11.812  | -0.01       | 0.36         |
| 266                               | 0.94  | 14.37 | 266                           | 0.925  | 13.82   | 0.01        | 0.55         |
| 317                               | 1.11  | 13.04 | 317                           | 1.105  | 12.54   | 0.01        | 0.50         |
| 329                               | 0.82  | 12.66 | 329                           | 0.821  | 12.343  | 0.00        | 0.32         |
| 333                               | -0.03 | 12.88 | 333                           | -0.149 | 12.276  | 0.12        | 0.60         |
| 348                               | 0.87  | 13.54 | 348                           | 0.872  | 13.073  | 0.00        | 0.47         |
| ---                               | ---   | ---   | 368                           | 0.684  | 15.275  | ---         | ---          |
| *                                 | 0.32  | 14.22 | ---                           | ---    | ---     | ---         | ---          |
| *                                 | 0.32  | 13.6  | ---                           | ---    | ---     | ---         | ---          |

|                    |      |      |
|--------------------|------|------|
| Average $\delta$ : | 0.01 | 0.43 |
|--------------------|------|------|

**Supplemental Figure S1:** Comparison of MBP Ile  $\delta$ 1 chemical shifts. Chemical shifts from MBP expressed in *P. pastoris* and *E. coli* (Gardner *et al.* 1998) are listed in blue and yellow, respectively. Although the constructs of MBP used in each study vary slightly, residue numbering is kept in reference to the *E. coli* spectra for simplicity. The construct used in our study does not include Ile368 but contains two additional peaks (notated with an asterisk) that did not readily correlate with previously published spectra and likely arises from small differences in the constructs. The  $^1\text{H}$  chemical shifts between samples remain nearly identical, however the *E. coli*-expressed sample spectra display a nearly uniform shift upfield of 0.43ppm in the  $^{13}\text{C}$  dimension. This shift is likely a combination of small differences in referencing and the effects of deuteration (Gardner and Kay 1998), as the *E. coli*-expressed sample is highly deuterated and the *P. pastoris*-expressed sample is protonated.

| Peak number | Volume ( <i>E. coli</i> ) | Volume ( <i>P. pastoris</i> ) | Percentage | S/N ( <i>E. coli</i> ) | S/N ( <i>P. pastoris</i> ) | Percentage |
|-------------|---------------------------|-------------------------------|------------|------------------------|----------------------------|------------|
| 1           | 7.93E+07                  | 3.40E+07                      | 42.94      | 1.69E+04               | 7.45E+03                   | 44.15      |
| 2           | 5.96E+07                  | 2.36E+07                      | 39.64      | 1.27E+04               | 5.18E+03                   | 40.77      |
| 3           | 5.07E+07                  | 2.21E+07                      | 43.50      | 1.08E+04               | 4.83E+03                   | 44.73      |
| 9           | 1.73E+07                  | 8.99E+06                      | 52.06      | 3.68E+03               | 1.97E+03                   | 53.53      |
| 12          | 1.23E+08                  | 4.85E+07                      | 39.54      | 2.61E+04               | 1.06E+04                   | 40.66      |
| 15          | 4.56E+07                  | 2.06E+07                      | 45.13      | 9.72E+03               | 4.51E+03                   | 46.41      |
| 17          | 1.25E+07                  | 7.04E+06                      | 56.17      | 2.67E+03               | 1.54E+03                   | 57.76      |
| 18          | 5.18E+07                  | 2.21E+07                      | 42.63      | 1.10E+04               | 4.83E+03                   | 43.84      |
| 19          | 7.13E+07                  | 2.79E+07                      | 39.13      | 1.52E+04               | 6.11E+03                   | 40.24      |
| 20          | 4.74E+07                  | 1.96E+07                      | 41.26      | 1.01E+04               | 4.29E+03                   | 42.43      |
|             |                           | Average                       | 44.34      |                        | Average                    | 45.60      |

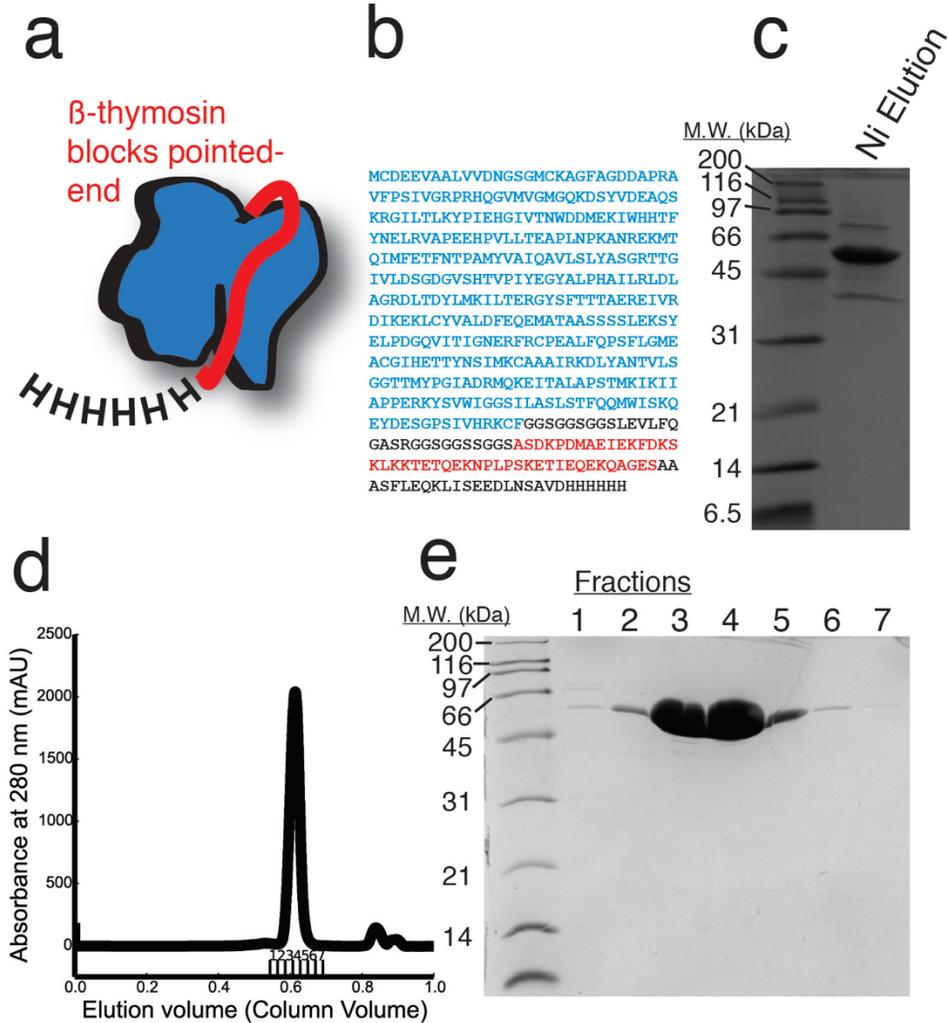
**Supplemental Figure S2:** Intensity comparison of ten well-resolved cross peaks in methyl HMQC spectra. On the basis of measured intensity in terms of peak volume and S/N, the percentage incorporations of  $^1\text{H}/^{13}\text{C}$  at the Ile  $\delta$ 1-methyl positions are  $44\% \pm 6\%$  and  $46\% \pm 6\%$ , respectively. Listed peak numbers are arbitrary.



**Supplemental Figure S3:** Mass Spectra of deuterated  $^{13}\text{C}$ -methyl-Ile-labeled MBP. (a) ESI mass spectrum of the principal species in the LC-MS on labeled MBP purified from *E. coli*. (b) Mass spectrum from analogous sample from *P. pastoris*. Protonated charge states are annotated above each peak in the mass to charge spectrum ( $m/z$ ). Calculated mass from the *E. coli* spectrum is 43,606 Da, corresponding to 91% deuterium labeling at non-exchangeable sites. Calculated mass from the *P. pastoris* spectrum is 43,566 Da, corresponding to 90% deuterium labeling at non-exchangeable sites.

|                    | Protonated growth |   | Deuterated growth |   |                       |
|--------------------|-------------------|---|-------------------|---|-----------------------|
|                    | Yield (mg/L)      | Ile( $\delta$ 1) incorporation rate (%) | Yield (mg/L)      | Ile( $\delta$ 1) incorporation rate (%) | Deuteration level (%) |
| <i>E. coli</i>     | Not tested        | Not tested                              | 40                | 100                                     | 91                    |
| <i>P. pastoris</i> | 60                | 51 $\pm$ 7                              | 10                | 45 $\pm$ 6                              | 90                    |
| <i>K. lactis</i>   | 10                | Not tested                              | 3.6               | 67 $\pm$ 6                              | >90                   |

**Supplemental Figure S4:** Comparison of MBP expressed in *E. coli* and *P. pastoris* from this manuscript and wild-type *K. lactis* from a recent study (Miyazawa-Onami *et al.* 2013). 100% Ile $\delta$ 1 incorporation in *E. coli* was assumed for comparison.



**Supplemental Figure S5:** Recombinant actin for NMR. (a) Cartoon and (b) sequence of actin-thymosin fusion expressed in *P. pastoris*. Actin is in blue, thymosin is in red, with linker or tagged regions in black. (c) SDS-PAGE gel showing actin-thymosin (51.5kD) as the main component in the elution off of Ni-NTA. (d) Size exclusion profile of purified actin-thymosin injected on a Superdex 200 10/300 column. Fractions collected are notated. (e) SDS-PAGE gel of fractions from (d).

## CHAPTER THREE

### Ligand modulation of sidechain dynamics in a wild-type human GPCR

#### Abstract

G protein-coupled receptors (GPCRs) constitute the most abundant superfamily of signaling receptors in the human proteome, regulating nearly every aspect of human physiology. Despite a wealth of crystallographic and spectroscopic studies on regulation of GPCR conformational states by different ligands, there is no experimental evidence for how sidechain dynamics control the allosteric transitions between these states. We generated samples of a model wild-type GPCR ( $A_{2A}$  adenosine receptor -  $A_{2A}R$ ) that are deuterated apart from  $^1H/^{13}C$  NMR probes at isoleucine  $\delta 1$  methyl groups, and used  $^1H/^{13}C$  methyl TROSY NMR (Tugarinov *et al.*, 2003) to probe the structure and dynamics of these residues while the receptor is saturated with a full agonist (NECA) or an inverse agonist (ZM241385), in the presence or absence of  $Na^+$  ions. Our data indicate that the loss of  $Na^+$  is required to allow significant structural rearrangements in NECA-bound  $A_{2A}R$ , and that the NECA- and ZM241385-bound receptors can be differentiated (even in the presence of  $Na^+$ ) by their patterns of sidechain dynamics, with the inverse agonist suppressing fast motions at the G protein binding site. Our approach to GPCR NMR spectroscopy creates a framework for exploring how different regions of a receptor

respond to different ligands or signaling proteins through modulation of fast ps-ns sidechain dynamics.

## Introduction

Our understanding of the molecular underpinnings of GPCR function has been greatly advanced over the past two decades through a combination of X-ray crystal structures, computational simulations, and spectroscopic studies of protein dynamics. Crystals of bovine rhodopsin provided the first high-resolution picture of a GPCR's architecture (Tugarinov *et al.*, 2003; Palczewski *et al.*, 2000; Li *et al.*, 2004), and structures of photoactivation intermediates (Nakamichi and Okada, 2006; Salom *et al.*, 2006) and retinal-free opsin (Park *et al.*, 2008; Scheerer *et al.*, 2008) further documented the structural transitions involved in rhodopsin activation. For GPCRs activated by diffusible ligands, crystal structures of the human  $\beta_2$  adrenergic receptor ( $\beta_2$ AR) with inverse agonists (Rosenbaum *et al.*, 2007; Cherezov *et al.*, 2007), agonists (Rosenbaum *et al.*, 2011; Rasmussen *et al.*, 2011a; Warne *et al.*, 2011), and bound G protein (Rasmussen *et al.*, 2011b) provided a molecular basis for understanding how diffusible agonist binding can promote structural changes in a receptor to enhance signaling. Subsequent crystal structures have revealed the ligand binding pockets for GPCRs of diverse function, responding to biogenic amines, purines, lipids, peptides, and proteins. While the sequences of these GPCRs and their orthosteric ligand binding pockets are highly diverse, the overall

structures are remarkably similar, with  $C\alpha$  RMSD values between unrelated receptors in the 2-3 Å range. How these diverse ligands activate signaling by a small number of G proteins and arrestins through a common structural scaffold remains a central problem for the GPCR field. Contributing to this problem is the fact that most of the existing X-ray GPCR structures are of antagonist-bound receptors locked in inactive conformations. For the few GPCRs that have been captured in both inactive and active conformations, common structural changes include inward movements of transmembrane (TM) helices at the orthosteric pocket, rearrangements of common core residues through the helical bundle, and rigid-body outward movements of TM6 at the cytoplasmic surface to expose G protein binding epitopes (Katritch *et al.*, 2013).

Beyond the static pictures of GPCR conformations seen in these crystal structures, experimental evidence for complex dynamic behavior has emerged from spectroscopic studies on purified receptors. Unlike rhodopsin, which exhibits an efficient and ordered photon-induced transition from the dark state conformation to metarhodopsin II (Choe *et al.*, 2011a; Choe *et al.*, 2011b), ligand-activated GPCRs generally show considerable basal activity that is reduced by inverse agonists.  $^{19}\text{F}$  NMR and EPR experiments on the  $\beta_2\text{AR}$  (Liu *et al.*, 2012a; Manglik *et al.*, 2015) and  $\text{A}_{2\text{A}}\text{R}$  (Ye *et al.*, 2016) provided evidence for at least 4 different conformational states populated by these receptors, which are differentially stabilized by ligands of different efficacy.  $^1\text{H}/^{13}\text{C}$  2D NMR studies on  $^{13}\text{C}_\epsilon$ -methionine-labeled  $\beta_2\text{AR}$

(Nygaard *et al.*, 2013; Kofuku *et al.*, 2014) showed that a high free-energy barrier exists between inactive and active conformations (with exchange on the millisecond timescale), and agonist binding alone is only capable of weakly populating the active conformation. These studies, in addition to similar measurements on  $\mu$ -opioid receptor chemically modified to introduce NMR-active  $^{13}\text{C}$ -methyl labels onto lysine residues (Sounier *et al.*, 2015), suggest that ligand-activated GPCRs are weakly coupled allosteric systems in which multiple energetic inputs (i.e. agonist and G protein binding) are required to predominantly populate the active conformation.

Within the slow millisecond-timescale global interconversion between functional GPCR conformations (Villardaga *et al.*, 2003), local structural rearrangements of sidechains in the receptor (referred to here as ‘microswitches’) have been observed in comparisons of inactive and active GPCR crystal structures (Katritch *et al.*, 2013), NMR data (Liu *et al.*, 2012a; Manglik *et al.*, 2015; Ye *et al.*, 2016; Bokoch *et al.*, 2010) and MD simulations (Vanni *et al.*, 2009; Dror *et al.*, 2011a). The repacking of sidechains from TM3, TM5, and TM6 beneath the ligand binding pocket, known as the ‘conserved core triad,’ was first observed in structures of  $\beta_2\text{AR}$  (Rosenbaum *et al.*, 2011; Rasmussen *et al.*, 2011a), and later seen in structures of the M2 muscarinic acetylcholine receptor (Kruse *et al.*, 2013) and the  $\mu$ -opioid receptor (Huang *et al.*, 2015). Further down the TM bundle, an activated microswitch near the site of G protein binding has been found to represent a common feature of GPCRs that have been crystallized in both active and inactive

conformations (Venkatakrisnan *et al.*, 2016). A working hypothesis for GPCR activation is that a set of loosely coupled microswitches connecting the orthosteric pocket and G protein binding site are activated by agonists in a non-concerted fashion, and stabilizing subsets of these rearrangements can lead to alternate overall conformations that have different signaling properties.

One of the most powerful ways to probe sidechain dynamics and their contributions to biological processes is provided by NMR-based studies of isotopically-labeled methyl groups (Ruschak and Kay, 2010). To apply these techniques to study the sidechains in these microswitches, one approach would entail isoleucine/leucine/valine (ILV) labeling (Goto *et al.*, 1999; Gardner *et al.*, 1998) and perdeuteration of a wild-type GPCR, to generate proteins with  $^1\text{H}/^{13}\text{C}$ -labeled methyl groups within an otherwise  $^2\text{H}/^{12}\text{C}$ -labeled background. Such samples are ideally suited for acquiring  $^1\text{H}/^{13}\text{C}$  methyl TROSY-based (Tugarinov *et al.*, 2003a; Ollerenshaw *et al.*, 2003) relaxation NMR data, allowing the quantitative determination of methyl order parameters and relative motion of sidechains on the ps-ns timescale (Tugarinov *et al.*, 2005; Ishima and Torchia, 2000). Advances in labeling methodology and pulse sequences have enabled such measurements on large macromolecular systems, such as the 80 kDa enzyme malate synthase (Tugarinov and Kay, 2003b; Sandala *et al.*, 2007), the 670 kDa archaeal proteasome core particle (Sprangers and Kay, 2007; Religa *et al.*, 2010), and the 1 MDa GroEL/GroES complex (Fiaux *et al.*, 2002; Horst *et al.*, 2005). NMR spectroscopic

interrogation of methyl groups has also been used to dissect the energetic contribution of sidechain entropy to important biological phenomena such as protein-protein interactions (Marlow *et al.*, 2010) and transcription factor binding to double-stranded DNA (Tzeng and Kalodimos, 2012). These previous applications suggest that characterization of sidechain dynamics using methyl NMR approaches should be feasible for a GPCR purified in a detergent micelle, with an aggregate molecular weight in the 100-150 kDa range.

A major hurdle in carrying out such experiments is that perdeuteration and ILV labeling conventionally rely on expression in *E. coli* (Goto *et al.*, 1999), while wild-type GPCRs typically require a eukaryotic host to ensure proper folding, glycosylation, trafficking, and stability in the plasma membrane. GPCR NMR studies have primarily used supplementation with labeled amino acids ( Nygaard *et al.*, 2013; Kofuku *et al.*, 2014; Okude *et al.*, 2015; Isogai *et al.*, 2016; Kofuku *et al.*, 2012) or covalent modification (Liu *et al.*, 2012a; Sounier *et al.*, 2015; Bokoch *et al.*, 2010; Eddy *et al.*, 2016) to incorporate isotopic probes into protein from Sf9 cells, and perdeuteration with incorporation of labeled ILV methyl probes is largely intractable for this established eukaryotic system. Another limitation is that purified wild-type GPCRs are typically prone to aggregation and are not highly thermostable, limiting acquisition times and sample concentration. While high quality 2D NMR spectra of GPCRs have been obtained for receptors with thermostabilizing mutations

(Isogai *et al.*, 2016), such mutations can affect the receptor's dynamics and functional properties, including basal activity and maximal G protein stimulation.

We have addressed many of these practical issues with our generation (Clark *et al.*, 2015) of highly-deuterated  $^1\text{H}/^{13}\text{C}$  Ile  $\delta 1$ -methyl labeled proteins in *Pichia pastoris*, a methylotrophic yeast (Cereghino and Cregg, 2000) that is adaptable to  $\text{D}_2\text{O}$ -based minimal media, will utilize the Ile precursor  $\alpha$ -ketobutyrate, and enables expression of GPCRs at levels sufficient for crystallography (Shimamura *et al.*, 2011; Hino *et al.*, 2012, Yurugi-Kobayashi *et al.*, 2009). Here we apply this labeling method to the wild-type human  $\text{A}_{2\text{A}}\text{R}$ , an important regulator of vascular function with a well-developed pharmacology including the natural hormone adenosine and high-affinity synthetic agonists and antagonists (Prete *et al.*, 2015, de Lera Ruiz *et al.*, 2014).  $\text{A}_{2\text{A}}\text{R}$  has been crystallized in inactive (Jaakola *et al.*, 2008; Liu *et al.*, 2012b), intermediate (Lebon *et al.*, 2011; Xu *et al.*, 2011) and active (Carpenter *et al.*, 2016) conformations, and has been studied by  $^{19}\text{F}$  NMR to characterize its conformational equilibria (Ye *et al.*, 2016). We expressed and purified milligram quantities of labeled wild-type human  $\text{A}_{2\text{A}}\text{R}$  and were able to resolve 20 out of 29 expected peaks in the Ile  $\delta 1$  region in  $^1\text{H}/^{13}\text{C}$  TROSY HMQC experiments. Through structure-guided mutagenesis, we assigned four of these signals to specific functionally important residues, including I92<sup>3.40</sup> of the conserved core triad (Huang *et al.*, 2015) and I292 at the cytoplasmic surface. We collected spectra in the presence of the high-affinity full agonist NECA and the inverse agonist ZM241385, and also measured the effects

of monovalent cations ( $\text{Na}^+$  versus  $\text{K}^+$ ) on these spectra. We further carried out a modified triple-quantum (3Q) relaxation experiment (Sun *et al.*, 2011) to quantify the relative fast ps-ns motions of these sidechains while the receptor is bound to ligands of opposing efficacy. These data represent an important first step towards understanding how agonists can activate the fast motions of specific sidechains to facilitate conformational changes of a GPCR.

## Results

The first challenge in collecting methyl NMR spectra on a wild-type GPCR was achieving high-level expression and milligram scale purification from *Pichia pastoris*. We initially cloned several GPCRs (including  $\text{A}_{2\text{A}}\text{R}$ ) into a modified methanol-inducible expression vector with the wild-type  $\text{MF}\alpha$  signal sequence at the N-terminus to direct incorporation into the membrane. Several receptors, such as the wild-type M2 muscarinic receptor and  $\text{A}_{2\text{A}}\text{R}$ , showed reliable and reproducible expression, however much of the expressed proteins were present in the immature unprocessed form (without  $\text{MF}\alpha$  cleavage) indicating that they had not been properly localized to the plasma membrane (Figure 1—figure supplement 1). To improve the efficiency of production of mature, folded receptors, we made several modifications to the  $\text{MF}\alpha$  sequence, based on biochemical and genetic precedents (Rakestraw *et al.*, 2009; Lin-Cereghino *et al.*, 2013) for improved processing efficiency (Figure 1A, Materials and methods). To validate that *Pichia*-expressed  $\text{A}_{2\text{A}}\text{R}$  is functional, we

carried out radioligand binding assays measuring [<sup>3</sup>H]-ZM241385 saturation and NECA competition on membranes after cell wall disruption and zymolyase treatment. The  $K_d$  for [<sup>3</sup>H]-ZM241385 was 0.42 nM in NaCl-containing buffer, while the  $K_i$  for NECA was 210 nM in low ionic strength and 420 nM in KCl-containing buffer, in agreement with previous binding studies (Xu *et al.*, 2011; Carpenter *et al.*, 2016; Bertheleme *et al.*, 2013) on heterologously expressed A<sub>2A</sub>R (Figure 1—figure supplement 2).

We developed a purification protocol for A<sub>2A</sub>R from *Pichia*-derived membranes, in which we can solubilize and purify ~0.5 mg of monodisperse biochemically pure wild-type A<sub>2A</sub>R in dodecylmaltoside (DDM) detergent from 1 L of culture (Figure 1—figure supplement 1). Coupling this protocol with our previously-described labeling method (Clark *et al.*, 2015), we were able to generate samples of wild-type A<sub>2A</sub>R that are highly deuterated both within the protein (<sup>2</sup>H at non-exchangeable sites except Ile δ1 methyl groups) and the surrounding buffer (99% D<sub>2</sub>O and deuterated DDM). We prioritized the use of a wild-type A<sub>2A</sub>R construct to report on the dynamics of a physiologically and functionally relevant GPCR sample. To confirm that the high level of deuteration did not alter the receptor's functional properties, we reconstituted the perdeuterated purified A<sub>2A</sub>R together with purified G<sub>s</sub> heterotrimer in phospholipid vesicles and carried out activation assays by [<sup>35</sup>S]GTP<sub>γ</sub>S binding. Saturating the receptor with the full agonist NECA led to an 8-fold stimulation of [<sup>35</sup>S]GTP<sub>γ</sub>S binding, and the inverse agonist ZM241385 produced

a 50% reduction in basal activity, similar to values measured for protonated A<sub>2A</sub>R (Figure 1B). These data show that the perdeuterated A<sub>2A</sub>R is capable of activating G<sub>s</sub> at wild-type levels in vitro, and can serve as a valid model for GPCR dynamics.

Our first NMR characterization of A<sub>2A</sub>R used <sup>1</sup>H/<sup>13</sup>C TROSY HMQC spectra of the protein in the presence of different ligands or G<sub>s</sub>. We initially expressed and solubilized the receptor with the low-affinity antagonist theophylline, and exchanged other ligands onto the receptor during washing in the Co<sup>2+</sup>-affinity and gel filtration steps (Materials and methods). In this way we were able to purify A<sub>2A</sub>R samples with the inverse agonist ZM241385, the full agonist NECA, or no ligand. The resulting 2D HMQC spectra, requiring ~100 mM GPCR and a minimum of 2 hr acquisition at 30° C on a cryo-probe-equipped 800 MHz spectrometer, are shown in Figure 2. Overall, in a liganded sample, we could resolve 20 peaks in the Ile δ1 region of the spectrum, out of 29 Ile residues present in our construct (28 from the receptor and one from the C-terminal Protein C tag). The chemical shifts of the agonist-bound and inverse agonist-bound A<sub>2A</sub>R are similar (Figure 2A, B), showing subtle changes that are comparable in magnitude to those observed previously for <sup>13</sup>C-methyl-methionine probes in β<sub>2</sub>AR (Kofuku *et al.*, 2014; Kofuku *et al.*, 2012). The spectrum of the unliganded (apo) A<sub>2A</sub>R shows a loss of most of the well-dispersed peaks present in the ligand-bound spectra (Figure 2C), which could result from protein instability or conformational exchange on the μs-ms timescale. Lastly, we formed and isolated a high-affinity complex between purified NECA-bound A<sub>2A</sub>R and purified G<sub>s</sub>

heterotrimer in the absence of GDP or GTP (Figure 2—figure supplement 1, Materials and methods). Like the apo spectrum, the G protein complex spectrum had fewer well-dispersed Ile peaks than receptor/ligand complexes (Figure 2D), consistent with the increase in molecular weight of the complex. We also note the appearance of a few distinct peaks (most prominently at 0.85 ppm  $^1\text{H}$ , 11.8 ppm  $^{13}\text{C}$ ) in the G protein complex, consistent with the conformation of the complexed receptor being distinguishable from the apo receptor (Rasmussen *et al.*, 2011b; Carpenter *et al.*, 2016).

Assignment of  $^1\text{H}/^{13}\text{C}$  peaks in the Ile  $\delta 1$  region by NMR was hampered by the modest dispersion of our spectra and the limited stability of A<sub>2</sub>AR during data acquisition. Therefore, we took a site-specific mutagenesis approach, wherein we collected HMQC spectra of samples containing point mutations at a small subset of functionally important Ile sites (Figure 2 and Figure 2—figure supplement 2). This process was complicated by the sensitivity of the receptor to mutations at several of these Iles: changes to Val, Leu, or Met often drastically reduced purification yields and/or led to significantly altered spectra; for an example, see the I106V spectra in Figure 2—figure supplement 2. Three of the assignments (I92<sup>3.40</sup>, I238<sup>6.40</sup>, and I292) could be made primarily through observation of ZM241385-bound mutant spectra. By comparing WT and I92V spectra in the presence of ZM241385, we see a clear loss of signal at  $\sim 0.68\text{ppm}/13.6\text{ppm}$  ( $^1\text{H}/^{13}\text{C}$ ) that we can assign to the  $\delta 1$  methyl of Ile92. The NECA-bound I92V spectrum is missing a subset of peaks compared to

the ZM241385-bound spectrum, possibly due to compromised affinity of NECA for the I92V mutant (Figure 1—figure supplement 2D). Ile238 and Ile292 are well-dispersed peaks, allowing for unambiguous assignment through loss of peaks at 0.72 ppm/10.2 ppm and 1.12 ppm/12.7 ppm, respectively ( $^1\text{H}/^{13}\text{C}$ ). The mutant spectra of I238<sup>6,40</sup> and I292 have reciprocal effects on each other, where I238V causes a full peak-width shift in the I292 signal (upfield in  $^1\text{H}$ ), and I292L causes a  $\sim$  0.5 peak-width shift in the I238 signal (upfield in  $^{13}\text{C}$ ). The reciprocity between these mutant spectra is consistent with the proximity of these residues in the A<sub>2A</sub>R structure. Mutation of I274<sup>7,39</sup> in the ligand binding pocket (within van der Waals contact of the ligands in the respective crystal structures) to Val or Met led to a disappearance of some peaks, reminiscent of the apo spectrum. However, comparison of ZM241385-bound I274M and I274V spectra narrowed the choice of the assignment of I274<sup>7,39</sup> to one of two peaks, with the greater solvent accessibility as revealed by solvent PRE effect (see below and Figure 4) of one peak serving as the deciding factor of assignment at 0.55 ppm/14.1 ppm ( $^1\text{H}/^{13}\text{C}$ ). As well as disappearance of many peaks, the spectrum of I274M A<sub>2A</sub>R in the presence of ZM241385 displays several apparent peak doublings, suggesting that the mutant protein is undergoing slow conformational exchange, potentially related to the slow interconversion between different inactive states of the receptor (Ye *et al.*, 2016). In addition, we assigned a peak to the Ile residue of the C-terminal Protein C tag in our construct by comparing spectra between differently tagged constructs (not shown).

The remaining unassigned peaks are referred to as a-o in the rest of this work (Figure 3A). The chemical shift changes between ZM241385-bound and NECA-bound  $A_{2A}R$  are small overall (Figure 2A, B). In the  $A_{2A}R$ - $G_s$  complex spectrum (Figure 2D), the peak for I92<sup>3,40</sup> appears to shift and simultaneously drop in intensity, strongly suggesting that I92<sup>3,40</sup> is involved in conformational changes on several timescales while the receptor is bound to  $G_s$ .

$A_{2A}R$  contains a well-characterized  $Na^+$  binding site within the receptor core (Liu *et al.*, 2012b), and  $Na^+$  has been shown to act as a negative allosteric modulator of  $A_{2A}R$  activation (Gao and Ijzerman, 2000). Due to the presence of 150 mM NaCl in the experiments shown in Figure 2, one possible explanation for the relatively small chemical shift differences between NECA- and ZM241385-bound spectra (Figure 3A, B) is that high  $[Na^+]$  is suppressing structural changes in the agonist-bound receptor. To test this possibility, we acquired HMQC spectra for  $A_{2A}R$  samples with 150 mM KCl substituting for NaCl. The spectrum of ZM241385-bound receptor in KCl is very similar to that in the presence of NaCl (Figure 3C and Figure 3—figure supplement 1B), with the peak assigned to I92<sup>3,40</sup> shifting slightly upfield in the  $^{13}C$  dimension. The spectrum of NECA-bound  $A_{2A}R$ , on the other hand, is dramatically different in KCl and NaCl (Figure 3D, Figure 3—figure supplement 1D), with multiple peaks shifting and decreasing in intensity. However, the KCl samples were significantly less stable than those in NaCl, showing markedly decreased spectral quality after 1 (NECA-bound) or 2 (ZM241385-bound) hours of acquisition.

Given the weak signal-to-noise and poor stability of the KCl samples, we were only able to perform NMR relaxation experiments (below) on A<sub>2A</sub>R in the presence of NaCl.

In addition to ligand-induced chemical shift changes, we used our system to characterize the solvent accessibility of residues within the ZM241385-bound A<sub>2A</sub>R (with NaCl) using solvent paramagnetic relaxation enhancements (PREs). To do so, we used the soluble paramagnetic Gd<sup>3+</sup>-DTPA probe (Petros *et al.*, 1990), which enhances the relaxation rates of protons in a distance dependent manner within roughly 15 Å of the protein/solvent interface (Madl *et al.*, 2011). By acquiring <sup>1</sup>H/<sup>13</sup>C TROSY HMQC spectra in the absence and presence of Gd<sup>3+</sup>-DTPA, we could compare peak intensities between the two to establish solvent PRE levels. Examining our data, we see that the average peak intensity in the Gd<sup>3+</sup>-DTPA containing spectrum is about 0.4 of that in the control spectrum, suggesting that the Ile δ1 methyl groups of A<sub>2A</sub>R in the DDM micelle are quite close to solvent on average (Figure 4). Of the four most strongly affected peaks, three peaks correspond to assigned residues (I238<sup>6.40</sup>, I274<sup>7.39</sup>, and I292). The accessibility of I274<sup>7.39</sup> is expected due this amino acid's presence in the solvent- exposed ligand-binding pocket. Broadening of I292 at the junction between TM7 and Helix 8 is also somewhat expected due to its position at the cytoplasmic surface. The broadening observed for I238<sup>6.40</sup> is more surprising, given that it is further embedded within the membrane and buried within the protein surface of A<sub>2A</sub>R. Crystal structures of

multiple GPCRs, including A<sub>2A</sub>R, β<sub>2</sub>AR, μOR, and rhodopsin have all revealed ordered solvent networks within the protein core in contact with the position equivalent to I238<sup>6,40</sup>. In ZM241385-bound A<sub>2A</sub>R, this residue is also in close proximity to the Na<sup>+</sup> site (Liu *et al.*, 2012b), one packing layer toward the cytoplasmic surface. The solvent PRE effect seen for this residue could reflect breathing of the structure to expose this region to the Gd<sup>3+</sup>-DTPA complex.

To characterize the motions of the Ile sidechains within A<sub>2A</sub>R, we sought to carry out a triple quantum (3Q) relaxation experiment developed by Kay, Tugarinov and colleagues (Sun *et al.*, 2011) to quantify sidechain dynamics in large macromolecules. In this experiment, two related series of 2D <sup>1</sup>H/<sup>13</sup>C HMQC-based spectra were acquired, both of which start by generating transverse <sup>1</sup>H magnetization composed of single quantum (SQ) coherences. During a subsequent variable delay period, intra-methyl <sup>1</sup>H-<sup>1</sup>H dipolar cross-correlated relaxation mechanisms lead to a portion of these SQ coherences evolving into 3Q coherences; the rate of this conversion ( $\eta$ ) is proportional to the  $S^2_{axis}$  methyl order parameter. Practically, the degree of SQ->3Q evolution can be quantitated by examining paired ‘forbidden’ and ‘allowed’ spectra, which report on the fraction of SQ coherences which do and do not evolve into 3Q coherences, respectively. As previously reported (Sun *et al.*, 2011), this allows the simple calculation of the  $\eta$  rate for each specific methyl group from the ratio of peak intensities in the forbidden and allowed spectra.

Our initial attempts to apply this method to purified A<sub>2A</sub>R were hampered by

the fact that the wild-type receptor is insufficiently stable at 30° C to permit the measurement of the requisite paired spectra at many relaxation delays and with a high enough signal-to-noise ratio. We therefore explored the possibility of reducing the number of forbidden spectra required to accurately measure  $\eta$  values. Using similarly  $^2\text{H}$ ,  $^{12}\text{C}$  ( $^1\text{H}$ ,  $^{13}\text{C}$   $\delta$ 1 methyl) labeled maltose binding protein (MBP) as a test case (Materials and methods), we collected 3Q datasets with different numbers of forbidden spectra paired with a constant number of allowed spectra (Figure 5, Figure 5—figure supplement 1). We found that pairing five allowed spectra with a single forbidden spectrum, we were able to faithfully recapitulate the  $\eta$  values measured using five pairs of forbidden and allowed experiments (Figure 5B, C). This  $\eta$  value is proportional to the  $S^2_{\text{axis}}$  order parameter, which quantifies the amplitude of motion of the methyl group on the ps-ns timescale that is faster than global molecular tumbling.

With this modified analysis of 3Q relaxation data and our Ile  $\delta$ 1  $^1\text{H}/^{13}\text{C}$ -methyl-labeled and deuterated GPCR samples, we measured  $\eta$  values for Ile sidechains in  $A_{2A}R$  bound to either ZM241385 or NECA in NaCl. The resulting values for the 20 Ile peaks and relative changes between the samples with different ligands are shown in Figure 6. The error in these measurements is large due to the limited signal-to-noise of our  $A_{2A}R$  spectra, and we did not attempt to convert these  $\eta$  measurements to  $S^2_{\text{axis}}$  values. However, in comparing the two datasets some features of the Ile dynamics can be discerned. The average  $\eta$  value for the

ZM241385-bound sample is higher than for NECA, indicative of greater overall rigidity of sidechains with inverse agonist. At the individual residue level, we observe a diversity of ligand-dependent changes in  $\eta$ , with some Ile residues becoming more rigid with agonist while most become more flexible. As an internal control, the peak we assigned to the C-terminal Protein C tag shows the lowest  $\eta$  value measured (i.e. greatest flexibility) with little ligand-dependence. Among the other assigned peaks, I292 shows the largest difference between ligands, becoming more flexible in the NECA-bound sample. I92<sup>3.40</sup> and I274<sup>7.39</sup> display more modest increases in flexibility with agonist, while the dynamics of I238<sup>6.40</sup> are largely unchanged by ligand (Figure 6B). Interestingly, these differences in fast timescale dynamics occur in the presence of NaCl, in which chemical shifts for the agonist- and inverse agonist-bound states were nearly identical. To our knowledge, these data represent the first reported effort to experimentally quantify site-specific sidechain dynamics in a GPCR or comparable human integral membrane protein.

While we were able to measure sidechain dynamics in this challenging A<sub>2A</sub>R sample, the associated errors are quite large. As such, we turned to molecular dynamics (MD) simulations to provide independent validation of our dynamics measurements. We extracted  $S^2_{\text{axis}}$  order parameters for Ile  $\delta^1$  methyl groups from ~80 ns trajectories (post-equilibrium) of ZM241385- and NECA-bound A<sub>2A</sub>R in DDM micelles, using overlapping 30-ns windows to estimate the standard deviation in the order parameters (Figure 6—figure supplement 1). The  $S^2_{\text{axis}}$  order parameters for

some methyl groups (for example, I60 and I80 in ZM241385; I106 and I302 in NECA) show significant variations across different windows, suggesting that slower timescale motions involving these residues contribute to the sidechain dynamics measured on the fast timescale. Interestingly, the differences between ZM241385- and NECA-bound dynamics at three of the four assigned sites (I238<sup>6.40</sup>, I274<sup>7.39</sup>, and I292) are qualitatively similar to what we see by NMR. The fourth site, I92<sup>3.40</sup>, shows the opposite trend, with higher  $S^2_{axis}$ , i.e. more rigidity, in the NECA-bound state. The simulations were set up with bulk Na<sup>+</sup> ions, but none specifically occupying the binding site near I92<sup>3.40</sup>, which could account for this discrepancy.

## Discussion

Sidechain dynamics represent an important functional component of protein behavior. In cases of protein-ligand (Marlow *et al.*, 2010; Frederick *et al.*, 2007) and protein-DNA (Tzeng and Kalodimos, 2012) interactions that have been intensively studied, changes in entropy arising from modified sidechain dynamics in complexes were found to substantially contribute toward the overall free energy of binding. For GPCRs, changes in sidechain dynamics may play an energetic role in binding to ligands and G proteins, however we have focused on allosteric mechanisms connecting these two functionally important binding sites. To experimentally assess the roles of sidechain dynamics in a wild-type GPCR, we set out to create a labeled sample that would be amenable to relaxation NMR methods. The spectra we

obtained for the labeled Ile  $\delta 1$  methyl groups of perdeuterated  $A_{2A}R$  are comparable or superior to previously published NMR spectra on other wild-type GPCRs, however the limited dispersion and signal-to-noise contributed to significant error in the relaxation values derived from the data (Fig. 6A).

The  $Na^+$  site first seen in the high-resolution structure of  $A_{2A}R$  bound to ZM241285 is conserved throughout most Class A GPCRs, and  $Na^+$  exerts a negative allosteric effect on  $A_{2A}R$  activation by bridging residues on TM3 and TM7 and stabilizing the inactive conformation (Liu *et al.*, 2012b). This effect can be observed pharmacologically as a NaCl-dependent decrease in agonist affinity (Carpenter *et al.*, 2016; Gao and Ijzerman, 2000). For  $A_{2A}R$  expressed in *P. pastoris*, we observe a 9-fold decrease in NECA affinity when membranes are incubated in 150 mM NaCl buffer versus 150 mM KCl buffer (Figure 1—figure supplement 2), similar to previously reported values (Carpenter *et al.*, 2016). Several of the Ile residues that we assigned (i.e. I92<sup>3.40</sup> and I238<sup>6.40</sup>) are in close proximity to the  $Na^+$  binding site, which should make for sensitive probes of local structure in this part of the receptor core. Our 2D NMR spectra indicate a strong dependency on combined lack of  $Na^+$  and binding of agonist to stabilize significant structural changes in the receptor core (Figure 3). This observation is also consistent with the active conformation seen in the crystal structure of the  $A_{2A}R$ /mini- $G_s$  complex (Carpenter *et al.*, 2016), which results in collapse of the  $Na^+$  site due to inward movements of TM3 and TM7. In contrast to our data, the crystal structure of a NECA-bound

thermostabilized A<sub>2A</sub>R mutant construct (Lebon *et al.*, 2011) showed significant structural changes relative to the inactive conformation even with a high [Na<sup>+</sup>] beyond its EC<sub>50</sub> (~50 mM) for negative allosteric modulation (Gao and Ijzerman, 2000). Unfortunately the weak signal and stability of our samples in KCl precluded measurement of 3Q relaxation dynamics under these conditions.

For the Na<sup>+</sup>-bound A<sub>2A</sub>R samples, we can correlate our observations of ligand-dependent changes in the relaxation rate of specific sites to previous structural and biophysical studies of GPCRs. I274<sup>7.39</sup> at the ligand binding site makes direct contact with the adenine or adenine analogue rings of both NECA and ZM241385 in their respective crystal structures (Jaakola *et al.*, 2008; Liu *et al.*, 2012b; Lebon *et al.*, 2011; Carpenter *et al.*, 2016) (Figure 7A). In our relaxation datasets, I274<sup>7.39</sup> was more rigid in terms of its fast sidechain motions with inverse agonist bound, and more flexible with NECA occupying the binding pocket. NMR experiments previously showed different ligand-dependent conformations of the extracellular surface region of β<sub>2</sub>AR bound to agonists compared to inverse agonists (Bokoch *et al.*, 2010). In addition, MD simulations on β<sub>2</sub>AR predicted greater mobility of agonists relative to inverse agonists in the orthosteric pocket of the inactive-state structure (i.e. not bound to G<sub>s</sub>) (Dror *et al.*, 2011b). Our data suggests that the fast dynamics of residues at the orthosteric pockets of Class A GPCRs may be correlated with the functional efficacy of the bound ligand. The residue I92<sup>3.40</sup> is one of the three residues of the conserved core triad in A<sub>2A</sub>R, along with P189<sup>5.50</sup> and

F242<sup>6.44</sup>, which interact at a layer beneath the orthosteric pocket further toward the cytoplasmic surface (Figure 7B). Consistent with the rearrangements of this micro-switch region between the inactive and active conformations of multiple GPCRs, we observe a lower  $\eta$  value for I92<sup>3.40</sup> in the NECA dataset versus ZM241385, indicating greater sidechain flexibility when A<sub>2A</sub>R is agonist-bound. In several GPCRs, including  $\beta_2$ AR and  $\mu$ OR, the conserved core triad is unchanged at a static structural level with an agonist bound but without a G protein or nanobody to further stabilize the active conformation. In contrast, crystal structures of A<sub>2A</sub>R bound to agonists alone (Figure 7B) showed an intermediate active-like conformation in the region surrounding I92<sup>3.40</sup> (Lebon *et al.*, 2011; Xu *et al.*, 2011). Our data indicate that low [Na<sup>+</sup>] is required for NECA alone to stabilize structural rearrangements surrounding I92 (Figure 3, Figure 3—figure supplement 1). However agonist alone (even with high [Na<sup>+</sup>]) is enough to at least promote the fast motions of the sidechains in this microswitch, which may reduce the activation energy for the observed packing rearrangement to the active conformation.

The residue I238<sup>6.40</sup> is situated further toward the G protein binding site at a critical region for GPCR activation, where it packs against TM7 and undergoes significant outward movement in the transition to the active conformation (Figure 7C). I238<sup>6.40</sup> is also one helical turn on TM6 above L235<sup>6.37</sup>, which participates in a conserved microswitch between inactive and active conformations for multiple GPCRs (Venkatakrisnan *et al.*, 2016). I292 is present at the linker between TM7

and Helix 8, where it packs against Y288<sup>7.53</sup> of the highly conserved NPXXY motif (Katritch *et al.*, 2013) (Figure 7D). NMR studies of <sup>13</sup>C-dimethyllysine-labeled  $\mu$ OR showed that peak broadening of a lysine probe in Helix 8 (near the position equivalent to I292) was more sensitive to agonist than probes at the cytoplasmic ends of TM5 and TM6 (Sounier *et al.*, 2015). Further, in the structure of A<sub>2A</sub>R/mini-G<sub>s</sub> (Carpenter *et al.*, 2016), the engineered G<sub>s</sub> protein makes direct contact with I292 at the TM7-Helix 8 junction. In our relaxation dataset, the  $\eta$  value of I238<sup>6.40</sup> is low (i.e. more flexible) and essentially independent of the ligand, while I292 undergoes a large change from more rigid to more flexible from ZM241385 to NECA. As mentioned above, the environment surrounding I238<sup>6.40</sup> is loosely packed in A<sub>2A</sub>R structures, with an ordered solvent network that may allow for relative freedom of motion for this sidechain. In contrast, despite the solvent exposure of I292 (Figure 4), inverse agonist binding suppresses the fast dynamics of this residue relative to agonist binding (Figure 6). Outward movement of TM6 is one of the hallmarks of GPCR activation seen in crystal structures, and the cytoplasmic ends of TM6 and TM7 are separated from each other in the structures of  $\beta_2$ AR and A<sub>2A</sub>AR bound to G<sub>s</sub> (Rasmussen *et al.*, 2011b; Carpenter *et al.*, 2016). Our data suggest that the responsiveness of sidechain dynamics to ligands in these two regions may be largely uncoupled, and that the inverse agonist activity of ZM241385 could partly arise from its allosteric suppression of dynamics at the cytoplasmic end of TM7. Changes in  $\eta$  for other peaks in our dataset may provide further insights into the

regulation of different regions of the receptor by ligands, depending on assignment of the other Ile residues in our spectra.

Beyond our studies of the A<sub>2A</sub>R, we can now apply the methods for labeling and NMR spectroscopy described here to other GPCRs and eukaryotic membrane proteins. Since the microswitches discussed above were identified by comparison of different GPCR structures (Huang *et al.*, 2015; Venkatakrisnan *et al.*, 2016), it will be instructive to apply our methods to other receptors to see if the same patterns of ligand-dependent changes apply across the GPCR superfamily or change depending on the cognate ligand or preferred G protein signaling partner. Further, sidechain dynamics may be among the biophysical properties of GPCRs that are altered when receptors are bound to allosteric modulators or biased agonists, classes of ligands that are of increasing value and importance in GPCR pharmacology and drug development. In addition to GPCRs, many other disease-relevant human integral membrane proteins (such as ABC transporters and ion channels) are currently impossible to study by NMR methods, largely due to the challenges of expression, labeling, and perdeuteration in *E. coli*. Our approach has the potential to make these systems tractable for similar NMR measurements of sidechain dynamics.

## Materials and Methods

### *Construct design*

The cDNA for wild-type human ADORA2A adenosine A<sub>2A</sub> receptor was cloned into the pPICZ vector for expression in *Pichia pastoris* with a modified MF $\alpha$  secretion signal combining previous precedents (Rakestraw *et al.*, 2009; Lin-Cereghino *et al.*, 2013). Modifications to the MF $\alpha$  greatly increased the amount of fully-processed receptor present at the plasma membrane. Briefly, the mutations V22A, G40D, L42S, V50A, V52A, and F55L were introduced, and residues 57-70 were removed from the signal sequence. The receptor expression construct was terminated at residue 316 and the N-glycosylation site at Asn154 was mutated to glutamine (Jaakola *et al.*, 2008). The gene for the receptor was followed by a 8x His tag and a Protein C tag at the C-terminus. All point mutations were created using standard QuikChange protocols. Plasmids were linearized by incubation with *PmeI* (NEB) and inserted via electroporation into freshly prepared competent KM71H cells (Invitrogen). Clones were screened for integration efficiency with increasing amount of Zeocin (Invitrogen) in selection media and further selected through expression screens and western blots. The best expressing clones were stored as glycerol stocks at -80° C.

### *Expression of protonated cultures*

For large scale growth in natural abundance media, a small amount of freshly streaked cells was inoculated into a 10 mL culture of BMG media (1% glycerol, 100 mM potassium phosphate pH 6.0, 1.34% YNB (yeast nitrogen base), 0.004% histidine,  $4 \times 10^{-5}$ % biotin) and shaken at 28° C overnight at 250 rpm. The pilot culture was used to inoculate multiple liters of BMG and shaken until saturation is reached ( $OD_{600}$  ~20-30). The total culture was spun down in sterile bottles at 4000 rpm for 30 min and resuspended in equal volume of BMM media (100 mM potassium phosphate pH 6.0, 1.34% YNB, 0.004% histidine,  $4 \times 10^{-5}$ % biotin) without methanol. Cultures continued shaking for ~8 hr at 28° C to allow for metabolism of residual glycerol. Protein expression was induced with the addition of 0.5% v/v methanol and the temperature was reduced to 20° C. An additional 0.5% methanol was added every 12 hr to maintain expression. Cells were harvested after 36-48 hr and pelleted by centrifugation. Pellets were stored at -80° C.

### *Expression of deuterated cultures*

For expression of cells in deuterated cultures, the cells were first adapted to deuterated media as follows. A small amount of freshly streaked cells was inoculated into a 50 mL culture of BMGY containing 90% D<sub>2</sub>O/10% H<sub>2</sub>O (Cambridge Isotope Laboratories, Inc.). The culture was shaken at 28° C until an  $OD_{600}$  of 8-10 was reached (typically ~24 hr). Once that OD was reached, 200  $\mu$ L of the 90%/10%

culture was inoculated into 50 mL of BMGY media made with 100% D<sub>2</sub>O and protonated glycerol. The culture was shaken at 28° C until an OD<sub>600</sub> of 8-10 was reached (typically ~48 hr). 200 µL of the 100% culture was inoculated into 50 mL of BMGY media made with 100% D<sub>2</sub>O including d<sub>8</sub>-glycerol as the carbon source (Cambridge Isotope Laboratories, Inc.). This culture was incubated until reaching an OD<sub>600</sub> of ~10 and the entire culture was inoculated into large scale cultures of BMGY media again made with 100% D<sub>2</sub>O including d<sub>8</sub>-glycerol. The large scale cultures were shaken until saturation (OD<sub>600</sub> of ~20-30) and then spun down in sterile bottles at 4000 rpm for 30 min. The cells were resuspended in BMMY made in 100% D<sub>2</sub>O without methanol and continued shaking for 12-16 hr to metabolize residual d<sub>8</sub>-glycerol. One hour prior to induction, 200 mg/L of labeled α-ketobutyric acid (methyl-<sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; Cambridge Isotope Laboratories, Inc.) was added to the culture. Ten minutes prior to induction, dry theophylline was added to the culture to a final concentration of 4 mM. Protein expression was induced with the addition of 0.5% d<sub>4</sub>-methanol (Cambridge Isotope Laboratories, Inc.) and the temperature was reduced to 20°C. Expression was maintained by further additions of 0.5% d<sub>4</sub>-methanol every 12 hr, and cells were harvested by centrifugation after 36-48 hr and stored at -80°C.

### *Purification*

Cell pellets were thawed and resuspended in lysis buffer (PBS containing 10% glycerol, 4 mM theophylline, 2 mM EDTA, and protease inhibitors (160  $\mu$ g/mL benzamidine, 2.5  $\mu$ g/mL leupeptin, 1 mM PMSF, 1  $\mu$ M E-64)). Cells were passed through a high-pressure microfluidizer (Microfluidics M-110P) three times at 24,000 psi with a cooling period between passes. LongLife Zymolyase (G-Bio Sciences) was added to the total lysate at a concentration of 15 U/mL and stirred at 37° C for 1 hr. Total membranes were isolated by centrifugation at 140,000 rcf for 30 min and then washed by douncing in an equal volume of lysis buffer followed by an additional centrifugation step. Membranes were then resuspended in hypotonic buffer (10 mM HEPES pH 7.5, 2 mM EDTA, 4 mM theophylline, protease inhibitors) by douncing and stirred at 4° C for 30 min, followed by centrifugation again to isolate membranes. Membranes were dounced in buffer containing 500 mM NaCl, 50 mM HEPES pH 7.5, 20% glycerol, 1% DDM (Anatrace), 4 mM theophylline, and protease inhibitors and stirred for 2 hours at 4° C. Insoluble material was spun out by centrifugation at 140,000 rcf for 30 min. The resultant supernatant was incubated with TALON resin (Clontech) pre-equilibrated in 250 mM NaCl, 50 mM HEPES 7.5, 0.05% protonated DDM, 5% glycerol, 4 mM theophylline, and 30 mM imidazole. Additional imidazole was added to the supernatant to the final concentration of 30 mM to minimize background binding. Batch binding continued overnight at 4° C.

Following batch binding, the resin was washed with a series of buffers to exchange the protonated DDM into deuterated DDM and exchange on the high affinity ligands ZM241385 or NECA. Buffers were made in D<sub>2</sub>O and all contain 250 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 0.05% DDM (protonated or deuterated; Anatrace), 20 mM imidazole, and 10 μM ZM241385 or 20 μM NECA (Tocris). Buffers contain different ratios of protonated:deuterated detergents, and were added sequentially: (A) 4:0; (B) 3:1; (C) 2:2; (D) 1:3; (E) 0:4. Protein was eluted from TALON with buffer E + 250 mM imidazole. Eluted A<sub>2A</sub>R was concentrated in 100 kDa MWCO Amicon concentrators (Millipore) and injected on a Superdex200 column (GE Healthcare) equilibrated in 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% deuterated DDM, and 10 μM ZM241385 or 20 μM NECA made in D<sub>2</sub>O.

Previous studies have shown the importance of using cholesterol hemisuccinate (CHS) as a component of the micelle to preserve A<sub>2A</sub>R function (Weiss and Grisshammer, 2002) and increase the thermostability of the receptor (Liu *et al.*, 2012b). However, when CHS was included in our purification, we observed significant artifacts in the Ile δ1 region of our NMR spectra that hampered data collection and analysis. Ye *et al.* utilized an XAC ligand affinity column as an additional chromatography step during purification (Ye *et al.*, 2016) that we did not include in this study. Our purification is similar to what has been published in structural studies of A<sub>2A</sub>R, following a general purification scheme of IMAC followed by size exclusion chromatography prior to crystallization in detergent (Hino *et al.*,

2012; Carpenter *et al.*, 2016; Sun *et al.*, 2017) or lipidic cubic phase (Jaakola *et al.*, 2008; Liu *et al.*, 2012b; Xu *et al.*, 2011). These structures contain A<sub>2A</sub>R bound to agonists, antagonists, or in complex with an engineered 'mini-G<sub>s</sub>' protein. Given our biochemical evidence (Figure 1—figure supplement 1, Figure 2—figure supplement 1), we are confident that our purification protocol outlined above produces a high level of folded and functional A<sub>2A</sub>R, which is able to efficiently bind to G<sub>s</sub> in solution.

#### *G protein expression and purification*

G<sub>s</sub> heterotrimer was purified from *Trichoplusia ni* cells grown in ESF921 media (Expression Systems). Cells were lysed in hypotonic buffer containing 10 mM Tris pH 7.4, 100 μM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 10 μM GDP, and protease inhibitors. After centrifugation, membranes were dounced and solubilized in buffer containing 100 mM NaCl, 20 mM HEPES pH 7.5, 1% sodium cholate, 0.05% DDM, 5 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 10 μM GDP, 5 mM imidazole, protease inhibitors, and a 1:150,000 volume dilution of CIP (NEB).

The soluble material was incubated with pre-equilibrated Ni<sup>2+</sup>-NTA resin for 2 hours at 4°C. Detergent exchange into deuterated DDM was performed on-column by mixing volumes of buffer E1 (100 mM NaCl, 20 mM HEPES pH 7.5, 1% sodium cholate, 0.05% dDDM, 5mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 10 μM GDP, 20 mM imidazole, protease inhibitors) and buffer E2 (50 mM NaCl, 20 mM HEPES pH 7.5, 0.1% dDDM, 1 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol, 10 μM GDP, 20 mM imidazole,

protease inhibitors) in the following E1:E2 ratio: 1:1; 1:3; 1:9; 1:19 at a flow rate of 1 mL/min.  $G_s$  was eluted with buffer E2 supplemented with 200 mM imidazole.

Following elution,  $MnCl_2$  was added to a final concentration of 1 mM, and the pooled eluate (~10 mL volume) was incubated with 10  $\mu$ L lambda phosphatase, 1  $\mu$ L CIP, and 1  $\mu$ L Antarctic phosphatase for 30' at 4° C. Sample was diluted with buffer E2 to reduce imidazole concentration and applied to a pre-equilibrated 2 mL Q-sepharose column. The resin was washed with 6 CV of 100 mM NaCl, 20 mM HEPES pH 7.5, 0.04% dDDM, 1 mM  $MgCl_2$ , 100  $\mu$ M TCEP, 10  $\mu$ M GDP and eluted with wash buffer supplemented with an additional 250 mM NaCl. Eluted sample was concentrated in 10 kDa MWCO concentrator to 1 mL and diluted 1:1 with buffer containing 20 mM HEPES pH 7.5, 1.5 mM EDTA, 1.65 mM  $MgCl_2$ , 100  $\mu$ M TCEP, 1  $\mu$ M GDP to reduce final DDM and NaCl concentrations. Glycerol was added to 20%, and aliquots were flash frozen in  $LN_2$  and stored at -80° C until needed. Active, functional  $G_s$  was conservatively estimated to constitute at least 40% of the total purified sample.

#### *A<sub>2A</sub>R-G<sub>s</sub> complex formation*

$A_{2A}R$ - $G_s$  complex formation was carried out as follows. SEC-purified  $A_{2A}R$  was mixed with  $G_s$  at a 1:8 w/w ratio in the presence of 1 mM EDTA, 3 mM  $MgCl_2$ , 10  $\mu$ M NECA, 1 mM  $MnCl_2$ , and a 1:100 dilution of lambda phosphatase (NEB). The complex was incubated at room temperature for 1 hour, then apyrase (NEB) was added a 1:2000 volume dilution followed by an additional hour incubation at room

temperature.  $\text{CaCl}_2$  was added to a final concentration of 2 mM and loaded at a flow rate of 10 mL/hr on Protein C antibody resin (Sigma) pre-equilibrated in 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% DDM, 2 mM  $\text{CaCl}_2$ , 10  $\mu\text{M}$  NECA. ProC resin was washed with 1 CV buffer at 10 mL/hr, followed by 5 CV buffer at 30 mL/hr. Complex was eluted in 2 CV with 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% DDM, 10  $\mu\text{M}$  NECA, 5 mM EDTA, 0.2 mg/mL ProC peptide (EDQVDPRLIDGK). Freshly made TCEP was added to a final concentration of 100  $\mu\text{M}$  and was spin-concentrated prior to injection on a Superdex200 column equilibrated in 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% DDM, and 10  $\mu\text{M}$  NECA.

$\text{A}_{2\text{A}}\text{R-G}_s$  complex formation for NMR proceeded as above with a few modifications.  $\text{G}_s$  heterotrimer purification was carried out as above with the final exchange steps containing deuterated DDM instead of protonated DDM. Deuterated, labeled, SEC-purified  $\text{A}_{2\text{A}}\text{R}$  was mixed with  $\text{G}_s$  at a 1:8 w/w ratio in the presence of 3 mM  $\text{MgCl}_2$  in a final volume of 1 mL. The complex was incubated at room temperature for five minutes, and then a 1/2000 volume of apyrase (NEB) was added. The complex was incubated for an additional hour at room temperature, and then was injected on a Superdex200 column incubated in 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% deuterated DDM, 20  $\mu\text{M}$  NECA. Peak fractions were concentrated to  $\sim 100$   $\mu\text{L}$  for NMR experiments.

### *NMR spectroscopy and analysis*

NMR spectra were collected at 30 °C on 50-100  $\mu\text{M}$   $A_{2A}R$  (complexed with ZM241285 or NECA) and 100  $\mu\text{M}$  MBP (complexed with  $\beta$ -cyclodextrin) using a Bruker AVANCE III HD 800 MHz spectrometer with a cryogenically-cooled TCI probe. All NMR samples were approximately 100  $\mu\text{l}$  in a 3 mm Shigemi tube. NMR data were processed using NMRpipe (Delaglio *et al.*, 1995) and analyzed using NMRViewJ (Johnson, 2004). Parameters derived from relaxation data were obtained with a non-linear least-squares fitting programmed in Python implemented in the Python SciPy library (Jones *et al.*, 2001; More, 1977). Errors in these parameters were generated using a nonparametric Monte Carlo bootstrap approach (Efron, 1981) with 1000 simulated datasets, each of which consisted of synthetic data-points with generated values centered on the measured peak intensities (and errors determined from the noise of the experimental spectra). Each simulated dataset was individually fit; parameters from these fits were averaged to generate the reported values (as average  $\pm$  standard deviation). Unassigned isoleucine methyl peaks in  $A_{2A}R$  are referred to by a peak ID (Figure 3A), sorted by increasing difference between inverse agonist- and agonist-bound relaxation rates (Figure 6B).

$^1\text{H}/^{13}\text{C}$ -HMQC spectra of the methyl region were collected with a  $^{13}\text{C}$  spectral width of 22 ppm (8 ppm) centered at 14 ppm (9.5 ppm) for  $A_{2A}R$  (MBP), with 32 complex pairs collected. Due to short sample lifetimes, some  $A_{2A}R$  HMQC spectra were collected as several sequential experiments of 64 scans; if the spectra

remained consistent, they were summed after processing. Solvent PRE experiments were carried out by collecting an HMQC spectrum on an A<sub>2</sub>A<sub>R</sub> sample, then adding Gd<sup>3+</sup> complexed with diethylenetriaminepentaacetic acid (DTPA; Sigma) to a final concentration of 1 mM and collecting a matched HMQC spectrum. The PRE effect was measured as the ratio of peak intensities in the paramagnetic sample to those in the diamagnetic sample.

Triple quantum (3Q) relaxation experiments were conducted using a pulse sequence kindly provided by Prof. Lewis Kay (University of Toronto) (Sun *et al.*, 2011). The carrier was centered on the methyl region (0.8 ppm) in the <sup>1</sup>H dimension, while <sup>13</sup>C dimension center and spectral width were as above for HMQC spectra. The 3Q relaxation experiments were run as pseudo-4D experiments with the relaxation delays and alternating forbidden/allowed experiments as the 3<sup>rd</sup> and 4<sup>th</sup> dimensions. An NMRpipe (Delaglio *et al.*, 1995) script was used to divide the data and process each spectrum. For MBP, spectra were collected with relaxation delays of 0.8, 2, 4, 8, and 16 ms for both forbidden and allowed experiments. Peaks (corresponding to I317 and I333) with low intensities (defined as less than ten-fold greater than noise in the 16 ms forbidden spectrum) were not used for relaxation analysis. Peak intensities were measured using the NMRViewJ (Johnson, 2004) Rate Analysis module by fitting each peak to an ellipse and calculating the volume. Ratios of peak intensities were fit to the following equation as a function of relaxation delay (Sun *et al.*, 2011):

$$\left| \frac{I_{forb}}{I_{all}} \right| = \frac{3N_{all}}{4N_{forb}} \frac{\eta \tanh(\sqrt{\eta^2 + \delta^2} T)}{\sqrt{\eta^2 + \delta^2} - \delta \tanh(\sqrt{\eta^2 + \delta^2} T)} \quad (1)$$

where  $N$  is the number of scans for each experiment,  $T$  is the relaxation delay,  $\delta$  ( $< 0$ ) is a term for the coupling between rapidly and slowly decaying single-quantum coherences, and  $\eta$  is a relaxation rate, defined as the difference between slow and fast relaxation rates of single-quantum transitions for methyl protons:

$$\eta = \frac{R_{2,H}^F - R_{2,H}^S}{2} \propto \tau_c S_{axis}^2 \quad (2)$$

proportional to the methyl axis order parameter  $S_{axis}^2$  and correlation time  $\tau_c$ . All relaxation rates were constrained to be  $> 0$  while  $\delta$  was constrained to be  $< 0$ .

Due to short lifetimes of A<sub>2</sub>AR samples (~12-14 hr at 30°C), the relaxation experiment was modified and run with 5 allowed experiments (0.8, 2, 4, 8, and 14 ms) and fewer forbidden experiments. To extract values of  $\eta$ , the forbidden:allowed peak intensity ratios at those few relaxation delays were fit to equation 1 as above. Simultaneously, the allowed experiment peak intensities were fit to:

$$I_{all} = A \frac{3}{2} [\exp(-R_{2,H}^S T) + \exp(-R_{2,H}^F T)] \quad (3)$$

where  $A$  is a scaling constant,  $T$  is the relaxation delay, and  $R^S$  and  $R^F$  are as above (Fig. 5 supp. 1).

To determine which relaxation delays for the limited forbidden experiments gave values of  $\eta$  that best agreed with those determined from a full dataset, the MBP relaxation dataset was analyzed using all 5 allowed experiments and different

combinations of 1 or 2 forbidden experiments. The differences in  $\eta$  values for each peak, as well as the sum of squared differences across all peaks, were compared for each analysis and a single relaxation delay of 8 ms was selected for  $A_{2A}R$  relaxation experiments (Fig. 5, Fig. 5, supp. 1). For processing, two separate scripts were used to extract the 5 allowed experiments and sum up the 5 forbidden experiments.

Allowed peak intensity and forbidden:allowed peak intensity ratios were calculated and fit to equations 2 and 1, respectively, as above, to derive values for  $\eta$  at each peak.

#### *Computational Methods: Molecular Constructs*

To construct systems for atomistic molecular dynamics (MD) simulations, we used the X-ray structures of the human  $A_{2A}$  GPCR in complex with agonist NECA or inverse agonist ZM241385 (PDB accession codes 2YDV (Lebon *et al.*, 2011) and 4E1Y (Liu *et al.*, 2012b), respectively). The thermostabilizing agents present in these X-ray models were removed and thermostabilizing mutations in the  $A_{2A}R$  introduced during the crystallization experiments were reverted back to the corresponding wild-type residues using VMD mutator plugin (Humphrey *et al.*, 1996). Several short residue segments missing from the X-ray structures were added with Modeller (Sali and Blundell, 1993) to complete the full-length atomistic models of the human  $A_{2A}R$ . The  $Na^+$  ion bound to the protein in the 4E1Y structure was removed so that the

effects of the agonist and the inverse agonist on the protein dynamics could be directly compared.

Using the protocol we have described earlier (Khelashvili *et al.*, 2013), the NECA- and ZM241385-bound A<sub>2A</sub>R structures were inserted into a detergent micelle containing 246 DDM molecules. The resulting proteomicelles were placed in  $\sim 157^3$  Å<sup>3</sup> size cubic water box containing 150 mM NaCl salt. The final molecular systems contained  $\sim 330,000$  atoms and resulted in detergent concentration of  $\sim 0.1$  M, well above established critical micelle concentration (CMC) for DDM (170 mM) (Kaufmann *et al.*, 2006).

#### *Computational Methods: Force fields and MD simulations*

The all-atom MD simulations were performed with the NAMD 2.7 package (Phillips *et al.*, 2005) using the all-atom CHARMM27 force-field with CMAP corrections for proteins (Brooks *et al.*, 2009), and a CHARMM-compatible force-field parameter set for detergents (Abel *et al.*, 2011). CHARMM- suitable parameters for NECA and ZM241385 compounds were generated with the program MATCH (Yesselman *et al.*, 2012). Molecular constructs were initially equilibrated using a two-phase protocol implemented by us earlier for MD simulations of leucine transporter (LeuT)/micelle complexes (Khelashvili *et al.*, 2013; LeVine *et al.*, 2016): (i) short energy minimization was carried out during which protein, water, and ion atoms were fixed and the coordinates of only DDM molecules were allowed to evolve freely; and

(ii) 1.5 ns long MD simulations were conducted with the protein backbone harmonically constrained. The constraints were released gradually, in 0.5 ns steps, with decreasing force constants of 1, 0.5, and 0.01 kcal/ (mol•Å<sup>2</sup>).

After the equilibration phase, both the NECA-bound and ZM241385-bound A<sub>2A</sub>R systems were subjected to ~80 ns unbiased MD simulations. Integration steps were 1 fs for the equilibration stage and 2 fs thereafter. The simulations implemented PME for electrostatics interactions (Essmann *et al.*, 1995) and were carried out in NPT ensemble under isotropic pressure coupling conditions, and at 310 K temperature. The Nose-Hoover Langevin piston algorithm (Phillips *et al.*, 2005) was used to control the target p=1 atm pressure with the Langevin-PistonPeriod set to 100 fs and LangevinPistonDecay to 50 fs. The first ~25 ns of each trajectory was discarded based on the convergence of the C $\alpha$  RMSD.

Each MD trajectory was split into six 32-ns windows overlapping by 28 ns. The S<sup>2</sup><sub>axis</sub> order parameters for Ile  $\delta$ 1 methyls were calculated from C $\gamma$ 1-C $\delta$ 1 bond vectors extracted from four of these windows using the trjtool software (Xue *et al.*, 2014; Bremi *et al.*, 1997). The average  $\pm$ standard deviation of these order parameters over the four windows is reported for each Ile  $\delta$ 1 methyl.

PDB files used in MD simulations and referenced throughout this manuscript (i.e. in Figure 7) can be accessed at <http://www.rcsb.org>.

### *Membrane binding*

Ligand binding experiments on membranes containing A<sub>2A</sub> receptor were carried out based on previously published protocols. *Pichia* cells expressing each construct were used to generate membranes as follows. Cells were resuspended in lysis buffer (PBS containing 10% glycerol, 2 mM EDTA, and protease inhibitors) and incubated for 2 hr at 37° C with Zymolyase 20T (AMS Bio) at a final concentration of 50 U/mL. Crude membranes were isolated through centrifugation at 40,000 rcf for 30 min and dounced in storage buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 10% glycerol, protease inhibitors). Large cell debris was removed by a low speed spin at 1000 rcf for 10 min, and remaining membranes were subjected to a high speed spin at 140,000 rcf for 30 min. Pellets were dounced in a minimal volume of storage buffer, flash frozen in LN<sub>2</sub>, and stored at -80° C until needed. Saturation binding was carried out by incubating 1.5-2 µg of membranes of membranes with different concentrations of [<sup>3</sup>H]-ZM241385 (50 Ci/mmol, American Radiolabeled Chemicals, Inc.) between 0.019 and 10 nM in assay buffer (50 mM Tris pH 7.4, 10 mM MgCl<sub>2</sub>, 1 mM EDTA, ±150 mM NaCl or KCl) containing 0.1% protease-free BSA in a final volume of 250 µL per tube. Reactions were incubated at room temperature for 1 hr. Non-specific binding was determined using reactions containing 10 mM theophylline. Reactions were separated on a vacuum manifold using GF/C filters (pre-soaked in assay + 0.5% PEI) to retain membranes and discard unbound ligand. After washing 4 times with cold assay buffer, bound radioactivity was quantified using a scintillation

counter. For competition binding experiments, aliquots of membranes were incubated with 0.5 nM [ $^3\text{H}$ ]-ZM241385, and varying concentrations of cold NECA, from 0.5 nM to 300  $\mu\text{M}$ , were included in the binding reactions. All binding experiments were carried out as three independent experiments, each performed in duplicate. Data analysis and fitting was performed with GraphPad Prism (GraphPad Software Inc.).

#### *Reconstitution and GTP $\gamma$ S binding*

9  $\mu\text{g}$  of  $G_s$  trimer was added to a tube and preincubated on ice for 15 min with 75  $\mu\text{L}$  of 1650  $\mu\text{M}$  SAPE (1-stearoyl-2-arachidonoyl-*sn*-glycero-3-phosphoethanolamine; Avanti Lipids), 980  $\mu\text{M}$  porcine brain phosphatidylserine (Avanti Lipids), 180  $\mu\text{M}$  cholesteryl hemisuccinate (Steraloids) in 20 mM HEPES pH 8.0, 100 mM NaCl, 0.4 % deoxycholate, 0.04% sodium cholate. 35 pmol of  $A_{2A}$  receptor and HMEN buffer (20 mM HEPES pH 8.0, 100 mM NaCl or KCl, 3 mM  $\text{MgCl}_2$ , 1 mM EDTA) was added to the tube for a total volume of 150  $\mu\text{L}$ . The sample was incubated an additional 5 min and applied to an Ultrogel AcA34 column (Sigma) equilibrated in HMEN buffer. BSA was added to fractions at a concentration of 0.1 mg/mL and vesicles were flash-frozen for storage at  $-80^\circ\text{C}$ . Receptor recovery was monitored by total ZM241385 binding and  $\sim 1$  mL pool just after void volume was used in the GTP $\gamma$ S binding assays. The recovery of  $A_{2A}$  receptor in this pool was  $\sim 13\%$ .

5  $\mu\text{L}$  of vesicles was assayed in a total of 50  $\mu\text{L}$  with 20 mM Hepes pH 8.0, 100 mM NaCl or KCl, 2 mM  $\text{MgCl}_2$ , 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, 100 nM  $\text{GTP}\gamma\text{S}$  with  $^{35}\text{S}$   $\text{GTP}\gamma\text{S}$  as a tracer in either no ligand, 10  $\mu\text{M}$  NECA agonist, or 100 nM ZM241385 antagonist. Samples were incubated at 30° C for 10 minutes and stopped with 50  $\mu\text{L}$  of quench buffer (20 mM Tris pH 8, 100 mM NaCl, 10 mM  $\text{MgCl}_2$ , 0.1 mM DTT, 1 mM GTP, and 0.1% Lubrol and incubated on ice for 10 minutes. Samples were filtered over BA85 nitrocellulose and the filters counted on a liquid scintillation counter after washing 4 times with 20 mM Tris pH 8.0, 100 mM NaCl, 10 mM  $\text{MgCl}_2$ .

## References

- Abel S, Dupradeau FY, Raman EP, MacKerell AD, Marchi M. 2011. Molecular simulations of dodecyl- $\beta$ -maltoside micelles in water: influence of the headgroup conformation and force field parameters. *The Journal of Physical Chemistry B* 115:487–499. DOI: <https://doi.org/10.1021/jp109545v>
- Bertheleme N, Singh S, Dowell SJ, Hubbard J, Byrne B. 2013. Loss of constitutive activity is correlated with increased thermostability of the human adenosine A2A receptor. *British Journal of Pharmacology* 169:988–998. DOI: <https://doi.org/10.1111/bph.12165>
- Bokoch MP, Zou Y, Rasmussen SG, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi HJ, Thian FS, Kobilka TS, Puglisi JD, Weis WI, Pardo L, Prosser RS, Mueller L, Kobilka BK. 2010. Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463:108–112. DOI: <https://doi.org/10.1038/nature08650>
- Bremi T, Bruschweiler R, Ernst RR. 1997. A protocol for the interpretation of side-chain dynamics based on NMR relaxation: application to phenylalanines in antamanide. *Journal of the American Chemical Society* 119:4272–4284. DOI: <https://doi.org/10.1021/ja9636505>
- Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, et al. 2009. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry* 30:1545–1614. DOI: <https://doi.org/10.1002/jcc.21287>
- Carpenter B, Nehmé R, Warne T, Leslie AG, Tate CG. 2016. Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* 536:104–107. DOI: <https://doi.org/10.1038/nature18966>
- Cereghino JL, Cregg JM. 2000. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews* 24:45–66. DOI: <https://doi.org/10.1111/j.1574-6976.2000.tb00532.x>
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC. 2007. High-resolution crystal structure of an engineered human  $\beta_2$ -adrenergic G protein-coupled receptor. *Science* 318:1258–1265. DOI: <https://doi.org/10.1126/science.1150577>

- Choe HW, Kim YJ, Park JH, Morizumi T, Pai EF, Krauss N, Hofmann KP, Scheerer P, Ernst OP. 2011b. Crystal structure of metarhodopsin II. *Nature* 471:651–655. DOI: <https://doi.org/10.1038/nature09789>
- Choe HW, Park JH, Kim YJ, Ernst OP. 2011a. Transmembrane signaling by GPCRs: insight from rhodopsin and opsin structures. *Neuropharmacology* 60:52–57. DOI: <https://doi.org/10.1016/j.neuropharm.2010.07.018>
- Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, Gardner KH, Rosen MK, Rosenbaum DM. 2015. Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *Journal of Biomolecular NMR* 62:239–245. DOI: <https://doi.org/10.1007/s10858-015-9939-2>
- de Lera Ruiz M, Lim YH, Zheng J. 2014. Adenosine A2A receptor as a drug discovery target. *Journal of Medicinal Chemistry* 57:3623–3650. DOI: <https://doi.org/10.1021/jm4011669>
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6:277–293. DOI: <https://doi.org/10.1007/BF00197809>
- Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, Borhani DW, Shaw DE. 2011a. Activation mechanism of the  $\beta_2$ -adrenergic receptor. *PNAS* 108:18684–18689. DOI: <https://doi.org/10.1073/pnas.1110499108>
- Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H, Shaw DE. 2011b. Pathway and mechanism of drug binding to G-protein-coupled receptors. *PNAS* 108:13118–13123. DOI: <https://doi.org/10.1073/pnas.1104614108>
- Eddy MT, Didenko T, Stevens RC, Wuë thrich K. 2016.  $\beta_2$ -adrenergic receptor conformational response to fusion protein in the third intracellular loop. *Structure* 24:2190–2197. DOI: <https://doi.org/10.1016/j.str.2016.09.015>
- Efron B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* 68:589–599. DOI: <https://doi.org/10.1093/biomet/68.3.589>
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103:8577–8593. DOI: <https://doi.org/10.1063/1.470117>

Fiaux J, Bertelsen EB, Horwich AL, Wu' thrich K. 2002. NMR analysis of a 900K GroEL GroES complex. *Nature* 418:207–211. DOI: <https://doi.org/10.1038/nature00860>

Frederick KK, Marlow MS, Valentine KG, Wand AJ. 2007. Conformational entropy in molecular recognition by proteins. *Nature* 448:325–329. DOI: <https://doi.org/10.1038/nature05959>

Gao ZG, Ijzerman AP. 2000. Allosteric modulation of A(2A) adenosine receptors by amiloride analogues and sodium ions. *Biochemical Pharmacology* 60:669–676. DOI: [https://doi.org/10.1016/S0006-2952\(00\)00360-9](https://doi.org/10.1016/S0006-2952(00)00360-9)

Gardner KH, Zhang X, Gehring K, Kay LE. 1998. Solution nmr studies of a 42 kDa *Escherichia Coli* maltose binding protein/b-cyclodextrin complex: chemical shift assignments and analysis. *Journal of the American Chemical Society* 120:11738–11748. DOI: <https://doi.org/10.1021/ja982019w>

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. 1999. A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta$  1) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ -,  $^2\text{H}$ -labeled proteins. *Journal of Biomolecular NMR* 13:369–374.

Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, Cameron AD, Kobayashi T, Hamakubo T, Iwata S, Murata T. 2012. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482:237–240. DOI: <https://doi.org/10.1038/nature10750>

Horst R, Bertelsen EB, Fiaux J, Wider G, Horwich AL, Wu' thrich K. 2005. Direct NMR observation of a substrate protein bound to the chaperonin GroEL. *PNAS* 102:12748–12753. DOI: <https://doi.org/10.1073/pnas.0505642102>

Huang W, Manglik A, Venkatakrishnan AJ, Laeremans T, Feinberg EN, Sanborn AL, Kato HE, Livingston KE, Thorsen TS, Kling RC, Granier S, Gmeiner P, Husbands SM, Traynor JR, Weis WI, Steyaert J, Dror RO, Kobilka BK. 2015. Structural insights into  $\mu$ -opioid receptor activation. *Nature* 524:315–321. DOI: <https://doi.org/10.1038/nature14886>

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14:33– 38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)

Ishima R, Torchia DA. 2000. Protein dynamics from NMR. *Nature Structural Biology* 7:740–743. DOI: <https://doi.org/10.1038/78963>

Isogai S, Deupi X, Opitz C, Heydenreich FM, Tsai CJ, Brueckner F, Schertler GF, Veprintsev DB, Grzesiek S. 2016. Backbone NMR reveals allosteric signal transduction networks in the  $\beta_1$ -adrenergic receptor. *Nature* 530:237–241. DOI: <https://doi.org/10.1038/nature16577>

Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EY, Lane JR, Ijzerman AP, Stevens RC. 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322:1211–1217. DOI: <https://doi.org/10.1126/science.1164772>

Johnson BA. 2004. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods in Molecular Biology* 278:313–352. DOI: <https://doi.org/10.1385/1-59259-809-9:313>

Jones E, Oliphant T, Peterson P. 2001. SciPy: Open Source Scientific Tools for Python. <http://www.scipy.org/> Katritch V, Cherezov V, Stevens RC. 2013. Structure-function of the G protein-coupled receptor superfamily. *Annual Review of Pharmacology and Toxicology* 53:531–556. DOI: <https://doi.org/10.1146/annurev-pharmtox-032112-135923>

Kaufmann TC, Engel A, Remigy HW. 2006. A novel method for detergent concentration determination. *Biophysical Journal* 90:310–317. DOI: <https://doi.org/10.1529/biophysj.105.070193>

Khelashvili G, LeVine MV, Shi L, Quick M, Javitch JA, Weinstein H. 2013. The membrane protein LeuT in micellar systems: aggregation dynamics and detergent binding to the S2 site. *Journal of the American Chemical Society* 135:14266–14275. DOI: <https://doi.org/10.1021/ja405984v>

Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Maeda M, Tsujishita H, Shimada I. 2012. Efficacy of the  $\beta_2$ -adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nature Communications* 3:1045. DOI: <https://doi.org/10.1038/ncomms2046>

Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Mizumura T, Suzuki S, Shimada I. 2014. Functional dynamics of deuterated  $\beta_2$ -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angewandte Chemie International Edition* 53:13376–13379. DOI: <https://doi.org/10.1002/anie.201406603>

Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AG, Tate CG. 2011. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* 474:521–525. DOI: <https://doi.org/10.1038/nature10136>

- LeVine MV, Khelashvili G, Shi L, Quick M, Javitch JA, Weinstein H. 2016. Role of Annular Lipids in the Functional Properties of Leucine Transporter LeuT Proteomicelles. *Biochemistry* 55:850–859. DOI: <https://doi.org/10.1021/acs.biochem.5b01268>
- Li J, Edwards PC, Burghammer M, Villa C, Schertler GF. 2004. Structure of bovine rhodopsin in a trigonal crystal form. *Journal of Molecular Biology* 343:1409–1438. DOI: <https://doi.org/10.1016/j.jmb.2004.08.090>
- Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chang JW, Chavan A, Tsai JW, Lin-Cereghino J. 2013. The effect of  $\alpha$ -mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene* 519:311–317. DOI: <https://doi.org/10.1016/j.gene.2013.01.062>
- Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K. 2012a. Biased signaling pathways in  $\beta_2$ -adrenergic receptor characterized by  $^{19}\text{F}$ -NMR. *Science* 335:1106–1110. DOI: <https://doi.org/10.1126/science.1215802>
- Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, IJzerman AP, Cherezov V, Stevens RC. 2012b. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* 337:232–236. DOI: <https://doi.org/10.1126/science.1219218>
- Madl T, Guttler T, Gorlich D, Sattler M. 2011. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angewandte Chemie International Edition* 50:3993–3997. DOI: <https://doi.org/10.1002/anie.201007168>
- Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK. 2015. Structural Insights into the Dynamic Process of  $\beta_2$ -Adrenergic Receptor Signaling. *Cell* 161:1101–1111. DOI: <https://doi.org/10.1016/j.cell.2015.04.043>
- Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. 2010. The role of conformational entropy in molecular recognition by calmodulin. *Nature Chemical Biology* 6:352–358. DOI: <https://doi.org/10.1038/nchembio.347>
- Mittermaier A, Kay LE, Forman-Kay JD. 1999. Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *Journal of Biomolecular NMR* 13:181–185. DOI: <https://doi.org/10.1023/A:1008387715167>

More JJ. 1977. The Levenberg-Marquardt Algorithm: Implementation and Theory, in Numerical Analysis. In: Lecture Notes in Mathematics 630. Watson GA: Springer Verlag. p. 105–116.

Nakamichi H, Okada T. 2006. Local peptide movement in the photoreaction intermediate of rhodopsin. PNAS 103:12729–12734. DOI: <https://doi.org/10.1073/pnas.0601765103>

Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A, Pan AC, Liu CW, Fung JJ, Bokoch MP, Thian FS, Kobilka TS, Shaw DE, Mueller L, Prosser RS, Kobilka BK. 2013. The dynamic process of  $\beta_2$ -adrenergic receptor activation. Cell 152:532–542. DOI: <https://doi.org/10.1016/j.cell.2013.01.008>

Okude J, Ueda T, Kofuku Y, Sato M, Nobuyama N, Kondo K, Shiraishi Y, Mizumura T, Onishi K, Natsume M, Maeda M, Tsujishita H, Kuranaga T, Inoue M, Shimada I. 2015. Identification of a conformational equilibrium that determines the efficacy and functional selectivity of the  $\mu$ -opioid receptor. Angewandte Chemie International Edition 54:15771–15776. DOI: <https://doi.org/10.1002/anie.201508794>

Ollerenshaw JE, Tugarinov V, Kay LE. 2003. Methyl TROSY: explanation and experimental verification. Magnetic Resonance in Chemistry 41:843–852. DOI: <https://doi.org/10.1002/mrc.1256>

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. Science 289: 739–745. DOI: <https://doi.org/10.1126/science.289.5480.739>

Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP. 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. Nature 454:183–187. DOI: <https://doi.org/10.1038/nature07063>

Petros AM, Mueller L, Kopple KD. 1990. NMR identification of protein surfaces using paramagnetic probes. Biochemistry 29:10041–10048. DOI: <https://doi.org/10.1021/bi00495a005>

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. 2005. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 26:1781–1802. DOI: <https://doi.org/10.1002/jcc.20289>

Preti D, Baraldi PG, Moorman AR, Borea PA, Varani K. 2015. History and perspectives of A2A adenosine receptor antagonists as potential therapeutic agents. *Medicinal Research Reviews* 35:790–848. DOI: <https://doi.org/10.1002/med.21344>

Rakestraw JA, Sazinsky SL, Piatasi A, Antipov E, Wittrup KD. 2009. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* 103:1192–1201. DOI: <https://doi.org/10.1002/bit.22338>

Rasmussen SG, Choi HJ, Fung JJ, Pardon E, Casarosa P, Chae PS, Devree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK. 2011a. Structure of a nanobody-stabilized active state of the  $\beta_2$ -adrenoceptor. *Nature* 469:175–180. DOI: <https://doi.org/10.1038/nature09648>

Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. 2011b. Crystal structure of the  $\beta_2$ -adrenergic receptor-Gs protein complex. *Nature* 477:549–555. DOI: <https://doi.org/10.1038/nature10361>

Religa TL, Sprangers R, Kay LE. 2010. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328:98–102. DOI: <https://doi.org/10.1126/science.1184991>

Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Yao XJ, Weis WI, Stevens RC, Kobilka BK. 2007. GPCR engineering yields high-resolution structural insights into  $\beta_2$ -adrenergic receptor function. *Science* 318:1266–1273. DOI: <https://doi.org/10.1126/science.1150609>

Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, Devree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK. 2011. Structure and function of an irreversible agonist-  $\beta_2$  adrenoceptor complex. *Nature* 469:236–240. DOI: <https://doi.org/10.1038/nature09665>

Ruschak AM, Kay LE. 2010. Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of Biomolecular NMR* 46:75–87. DOI: <https://doi.org/10.1007/s10858-009-9376-1>

Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234:779–815. DOI: <https://doi.org/10.1006/jmbi.1993.1626>

Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K. 2006. Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *PNAS* 103:16123–16128. DOI: <https://doi.org/10.1073/pnas.0608022103>

Sandala GM, Smith DM, Marsh EN, Radom L. 2007. Toward an improved understanding of the glutamate mutase system. *Journal of the American Chemical Society* 129:1623–1633. DOI: <https://doi.org/10.1021/ja066432c>

Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP. 2008. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455:497–502. DOI: <https://doi.org/10.1038/nature07330>

Sharff AJ, Rodseth LE, Quijcho FA. 1993. Refined 1.8-Å structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein. *Biochemistry* 32:10553–10559. DOI: <https://doi.org/10.1021/bi00091a004>

Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S. 2011. Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65–70. DOI: <https://doi.org/10.1038/nature10236>

Sounier R, Mas C, Steyaert J, Laeremans T, Manglik A, Huang W, Kobilka BK, Demene H, Granier S. 2015. Propagation of conformational changes during  $\mu$ -opioid receptor activation. *Nature* 524:375–378. DOI: <https://doi.org/10.1038/nature14680>

Sprangers R, Kay LE. 2007. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618–622. DOI: <https://doi.org/10.1038/nature05512>

Sun B, Bachhawat P, Chu ML, Wood M, Ceska T, Sands ZA, Mercier J, Lebon F, Kobilka TS, Kobilka BK. 2017. Crystal structure of the adenosine A2A receptor bound to an antagonist reveals a potential allosteric pocket. *PNAS* 114:2066–2071. DOI: <https://doi.org/10.1073/pnas.1621423114>

Sun H, Kay LE, Tugarinov V. 2011. An optimized relaxation-based coherence transfer NMR experiment for the measurement of side-chain order in methyl-

protonated, highly deuterated proteins. *The Journal of Physical Chemistry B* 115:14878–14884. DOI: <https://doi.org/10.1021/jp209049k>

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. 2003a. Cross-correlated relaxation enhanced  $^1\text{H}$ - $^{13}\text{C}$  NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society* 125:10420–10428. DOI: <https://doi.org/10.1021/ja030153x>

Tugarinov V, Kay LE. 2003b. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *Journal of the American Chemical Society* 125:13868–13878. DOI: <https://doi.org/10.1021/ja030345s>

Tugarinov V, Ollerenshaw JE, Kay LE. 2005. Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: an application to an 82-kDa enzyme. *Journal of the American Chemical Society* 127:8214–8225. DOI: <https://doi.org/10.1021/ja0508830>

Tzeng SR, Kalodimos CG. 2012. Protein activity regulation by conformational entropy. *Nature* 488:236–240. DOI: <https://doi.org/10.1038/nature11271>

Vanni S, Neri M, Tavernelli I, Rothlisberger U. 2009. Observation of "ionic lock" formation in molecular dynamics simulations of wild-type beta 1 and beta 2 adrenergic receptors. *Biochemistry* 48:4789–4797. DOI: <https://doi.org/10.1021/bi900299f>

Venkatakrishnan AJ, Deupi X, Lebon G, Heydenreich FM, Flock T, Miljus T, Balaji S, Bouvier M, Veprintsev DB, Tate CG, Schertler GF, Babu MM. 2016. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* 536:484–487. DOI: <https://doi.org/10.1038/nature19107>

Villardaga JP, Bunemann M, Krasel C, Castro M, Lohse MJ. 2003. Measurement of the millisecond activation switch of G protein-coupled receptors in living cells. *Nature Biotechnology* 21:807–812. DOI: <https://doi.org/10.1038/nbt838>

Warne T, Moukhametzianov R, Baker JG, Nehmé R, Edwards PC, Leslie AG, Schertler GF, Tate CG. 2011. The structural basis for agonist and partial agonist action on a  $\beta_1$ -adrenergic receptor. *Nature* 469:241–244. DOI: <https://doi.org/10.1038/nature09746>

Weiss HM, Grisshammer R. 2002. Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *European*

Journal of Biochemistry 269:82–92. DOI: [https://doi.org/ 10.1046/j.0014-2956.2002.02618.x](https://doi.org/10.1046/j.0014-2956.2002.02618.x)

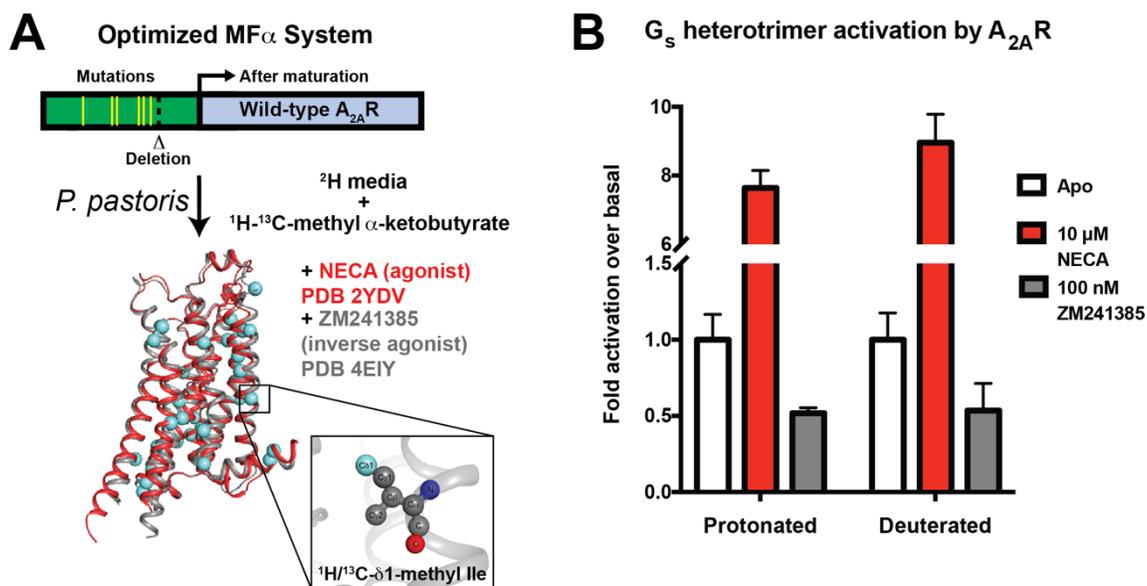
Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC. 2011. Structure of an agonist-bound human A2A adenosine receptor. *Science* 332:322–327. DOI: <https://doi.org/10.1126/science.1202793>

Xue Y, Yuwen T, Zhu F, Skrynnikov NR. 2014. Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos. *Biochemistry* 53:6473–6495. DOI: <https://doi.org/10.1021/bi500904f>

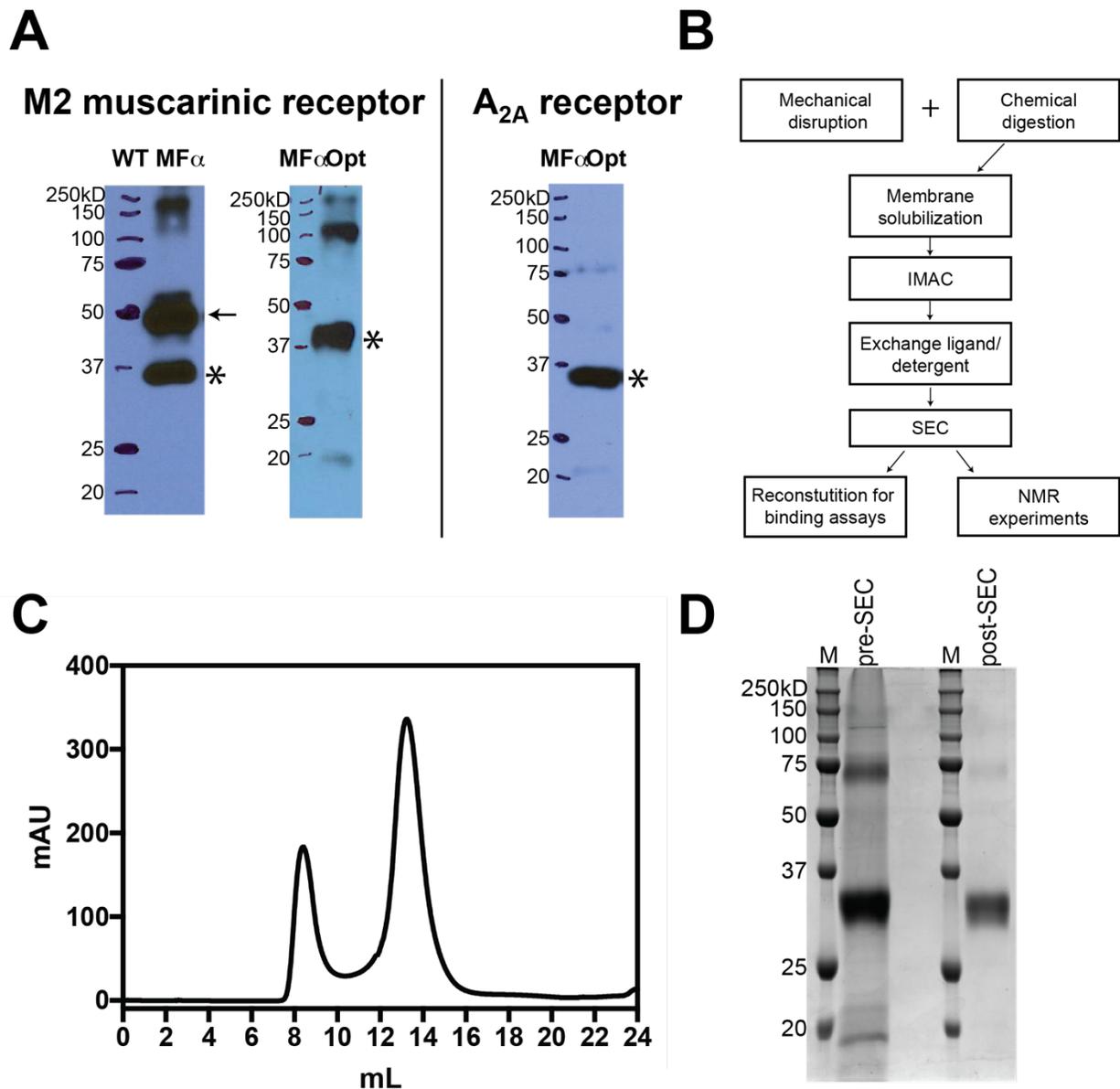
Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS. 2016. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* 533:265–268. DOI: <https://doi.org/10.1038/nature17668>

Yesselman JD, Price DJ, Knight JL, Brooks CL. 2012. MATCH: an atom-typing toolset for molecular mechanics force fields. *Journal of Computational Chemistry* 33:189–202. DOI: <https://doi.org/10.1002/jcc.21963>

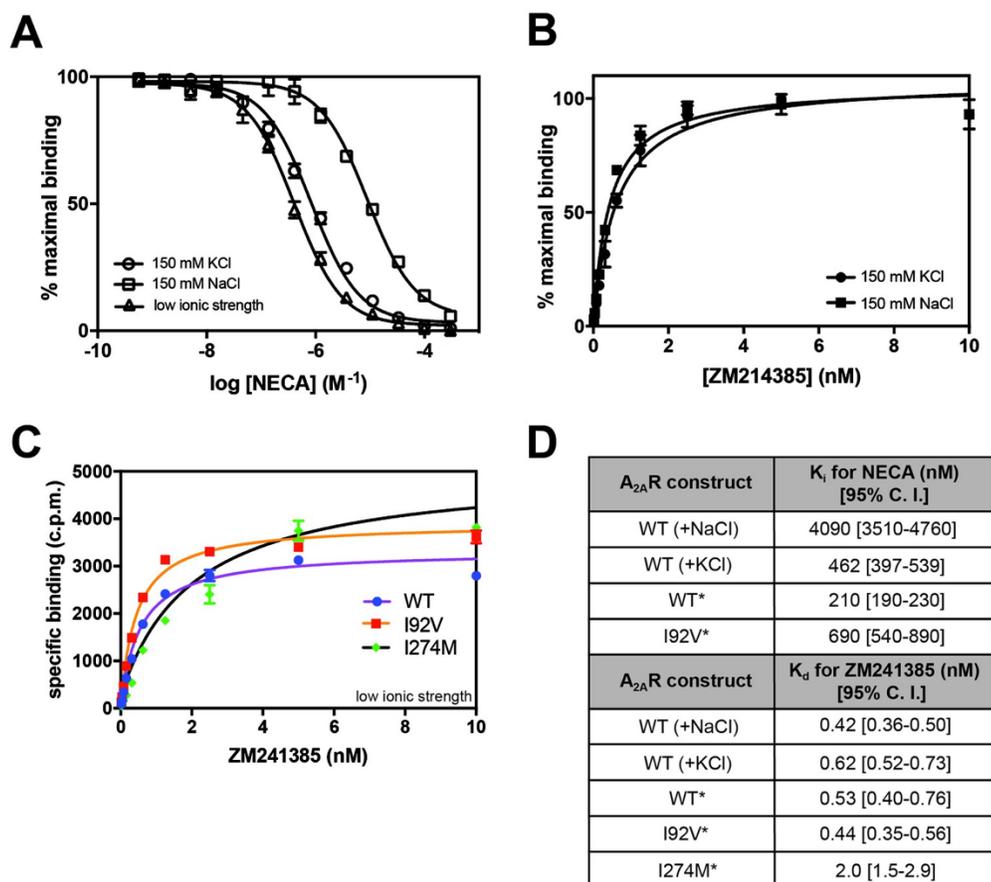
Yurugi-Kobayashi T, Asada H, Shiroishi M, Shimamura T, Funamoto S, Katsuta N, Ito K, Sugawara T, Tokuda N, Tsujimoto H, Murata T, Nomura N, Haga K, Haga T, Iwata S, Kobayashi T. 2009. Comparison of functional non-glycosylated GPCRs expression in *Pichia pastoris*. *Biochemical and Biophysical Research Communications* 380: 271–276. DOI: <https://doi.org/10.1016/j.bbrc.2009.01.053>



**Figure 1.** A<sub>2A</sub>R expression and G<sub>s</sub> activation. (A) Wild-type A<sub>2A</sub>R was expressed in *P. pastoris* as a fusion with an optimized version of the  $\alpha$ -mating factor (MF $\alpha$ ) signal sequence that contains mutations and a deletion that increase receptor trafficking to the plasma membrane. For NMR experiments, *Pichia* cultures were grown in deuterated media supplemented with (<sup>1</sup>H/<sup>13</sup>C-methyl)  $\alpha$ -ketobutyrate to facilitate <sup>1</sup>H/<sup>13</sup>C labeling of Ile  $\delta$ 1 methyl groups in a deuterated background as previously described (Clark *et al.*, 2015). Crystal structures of A<sub>2A</sub>R are shown complexed with agonist NECA (red, 2YDV (Lebon *et al.*, 2011)) and inverse agonist ZM241385 (gray, 4EIY (Liu *et al.*, 2012b)), with isoleucine residues displayed as spheres. Labeled Ile  $\delta$ 1 carbon atoms are shown in cyan (see inset). (B) [<sup>35</sup>S]GTP $\gamma$ S binding to purified, protonated or perdeuterated A<sub>2A</sub>R reconstituted with purified G<sub>s</sub> heterotrimer. Agonist NECA stimulates [<sup>35</sup>S]GTP $\gamma$ S binding, while inverse agonist ZM241385 inhibits basal levels of binding.

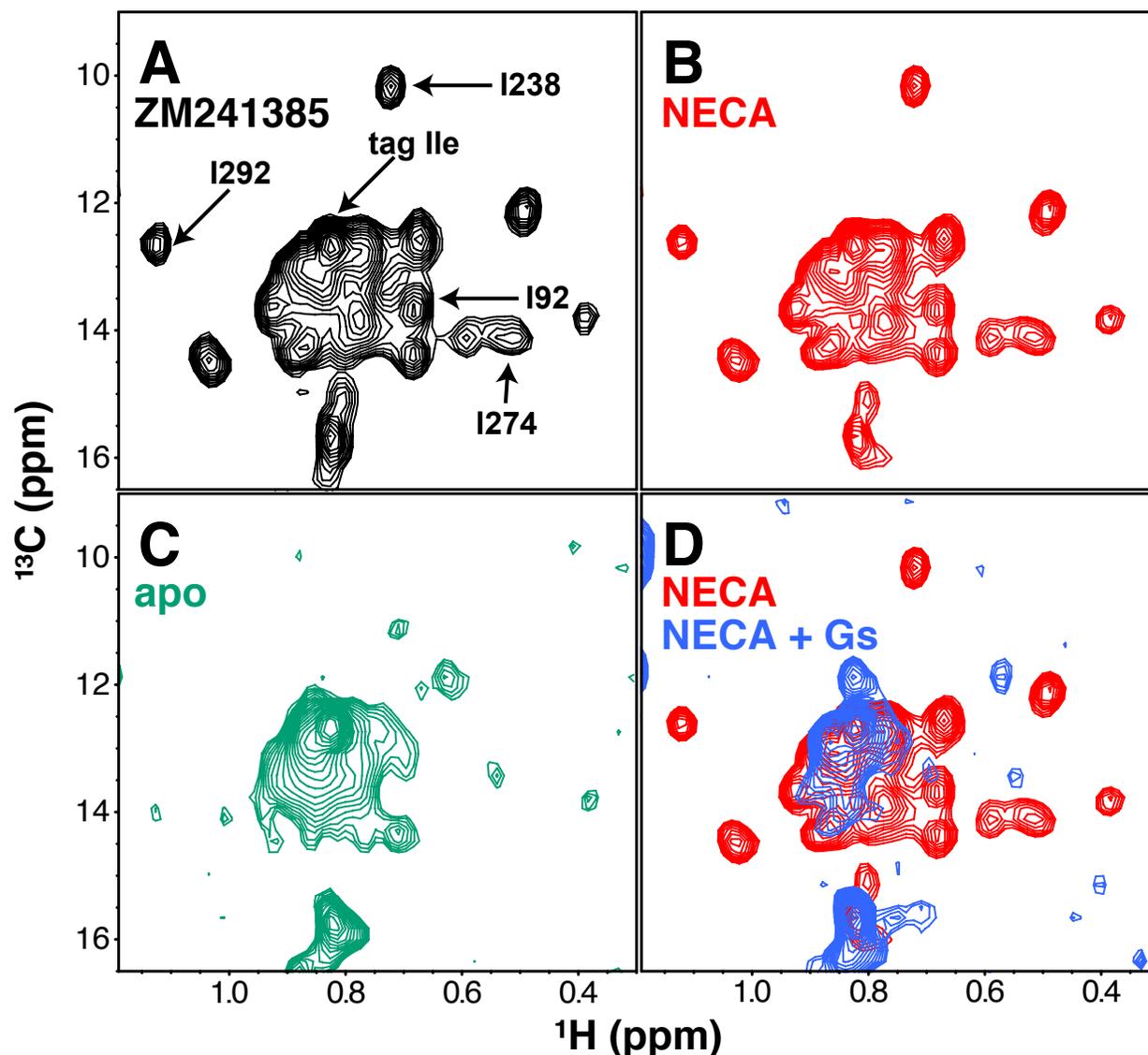


**Figure 1 supp. 1: A<sub>2A</sub>R processing and purification.** A) Anti-His<sub>6</sub> Western blots of wild-type GPCRs expressed in *Pichia* as a fusion with the wild-type (WT) and optimized (MF $\alpha$ Opt)  $\alpha$ -mating factor secretion signal. Receptor that has not been fully processed and trafficked to the plasma membrane is marked with an arrow. Fully processed, mature receptor is marked with an asterisk. B) Purification scheme of A<sub>2A</sub>R from *Pichia* cultures. C) Gel filtration trace of IMAC-purified A<sub>2A</sub>R injected on a Superdex200 10/300 column. The void volume is 8.3 mL and monomer A<sub>2A</sub>R elutes at approximately 13.7 mL. D) SDS-PAGE of IMAC-purified A<sub>2A</sub>R before and after size exclusion chromatography shown in C).

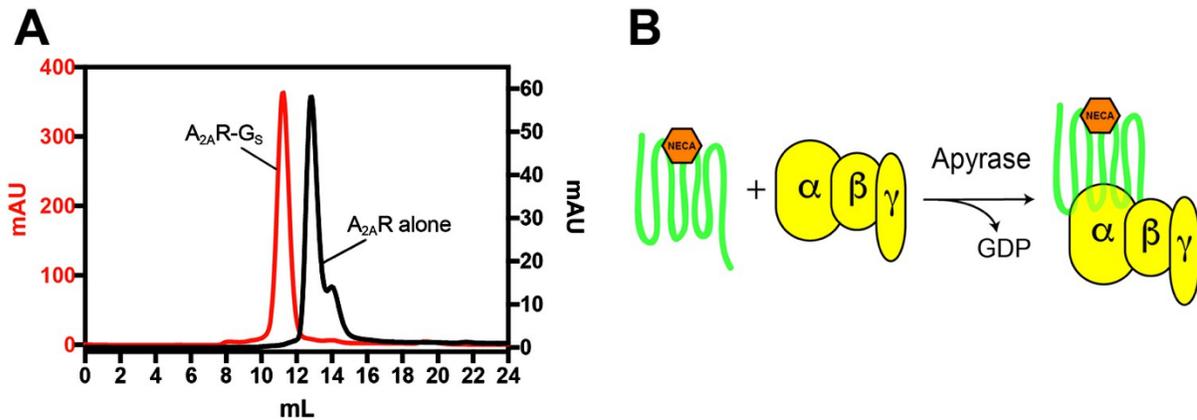


**Figure 1 supp. 2: Ligand binding assays on  $A_{2A}R$  yeast membranes. A)**

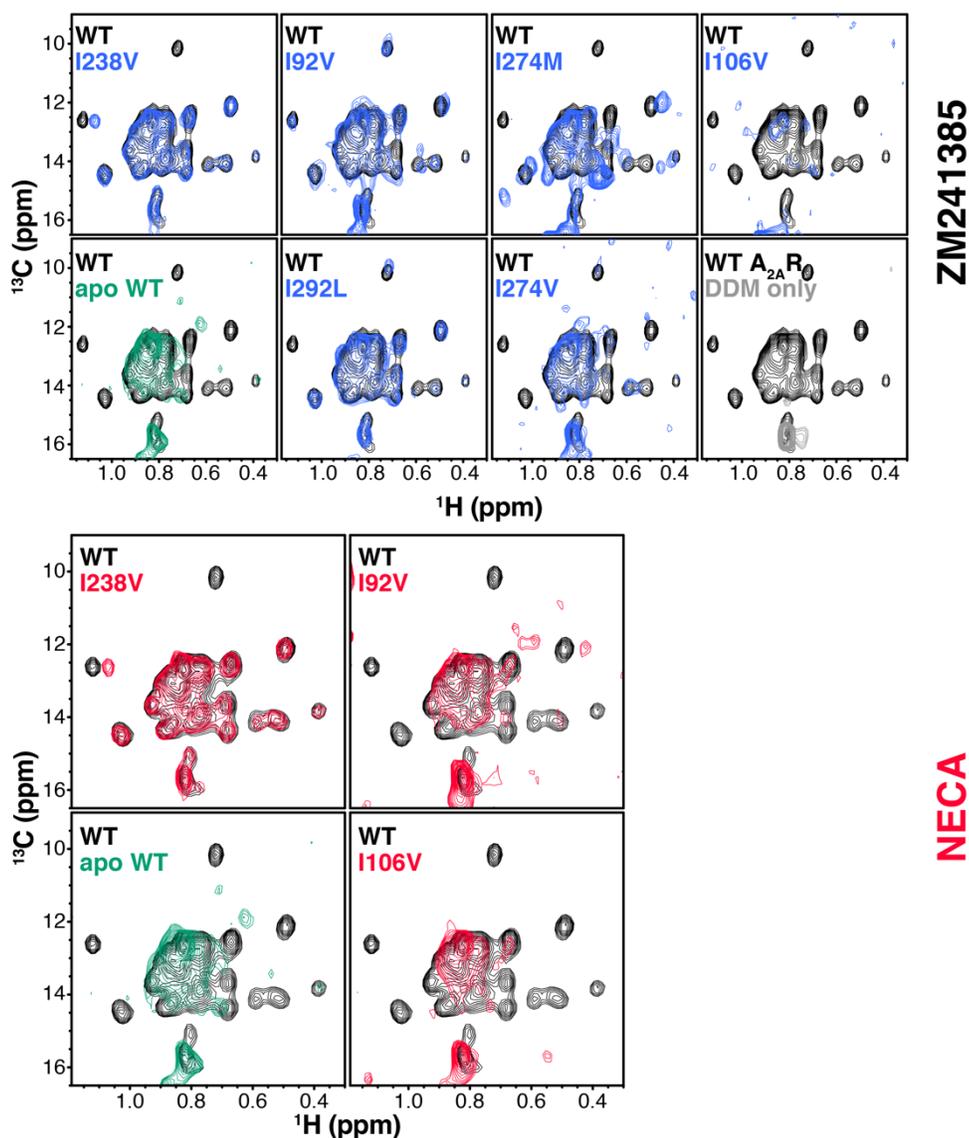
Competition binding curve of WT  $A_{2A}R$  membranes incubated with [ $^3H$ ]-ZM214385 and increasing amounts of unlabeled NECA in low ionic strength assay buffer (50 mM Tris pH 7.4, 10 mM  $MgCl_2$ , 1 mM EDTA) alone or supplemented with 150 mM NaCl or KCl. B) Saturation curve of WT  $A_{2A}R$  membranes incubated with increasing amounts of [ $^3H$ ]-ZM214385 in assay buffer containing 150 mM NaCl or KCl. Error bars in A) and B) represent standard deviation for three separate experiments, each performed in duplicate. C) [ $^3H$ ]-ZM214385 saturation binding of two mutants that showed spectral differences. Binding experiments were carried out in assay low ionic strength buffer without additional NaCl or KCl added, and corresponding values are indicated with asterisks in panel D. D) Table of  $K_i$  and  $K_d$  values of WT and mutant constructs tested in A-C with 95% confidence interval indicated. Additional buffer components in each assay are notated in parentheses, with an asterisk indicating low ionic strength buffer with no additional salt added.



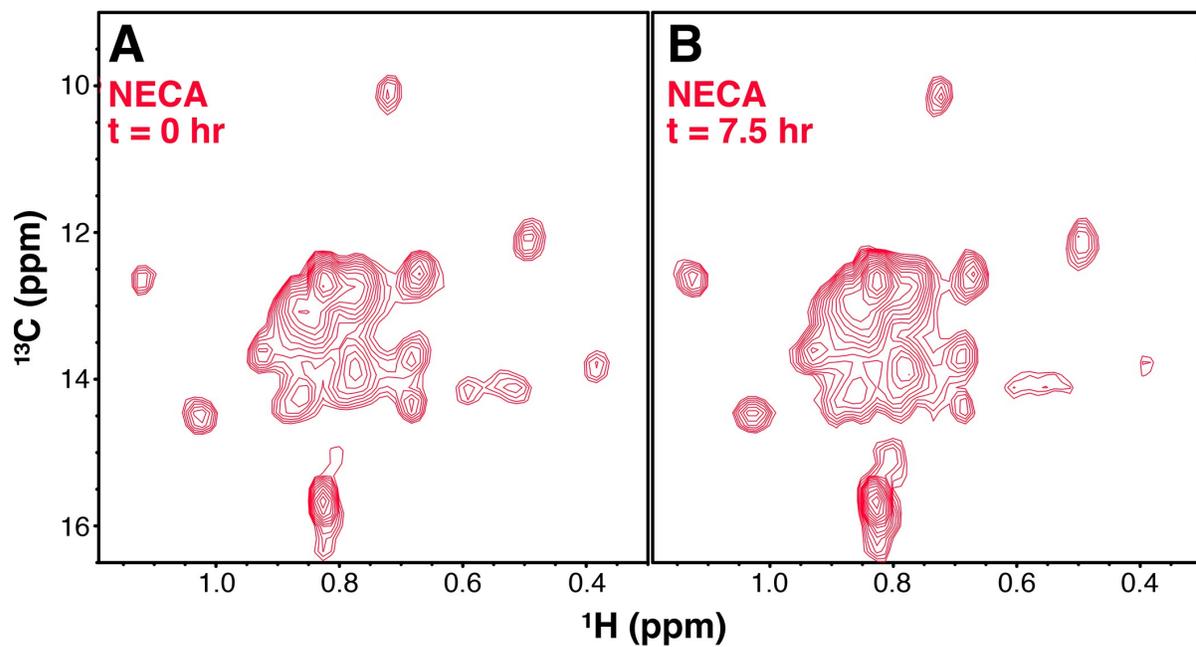
**Figure 2: NMR spectra of  $\text{A}_2\text{A}$ R in different liganded states with NaCl.** A)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_2\text{A}$ R in DDM micelles and 150 mM NaCl with inverse agonist ZM241385 (black). Resonances assigned in this work (I92<sup>3,40</sup>, I238<sup>6,40</sup>, I274<sup>7,39</sup>, I292, and the isoleucine residue in the protein C tag) are indicated with arrows. B)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_2\text{A}$ R in DDM micelles and 150 mM NaCl with agonist NECA (red). C)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_2\text{A}$ R in DDM micelles and 150 mM NaCl without ligand (green). D)  $^1\text{H}/^{13}\text{C}$  HMQC spectra of Ile  $\delta 1$ -labeled WT  $\text{A}_2\text{A}$ R in DDM micelles and 150 mM NaCl with agonist NECA (red) and with agonist NECA and  $\text{G}_s$  heterotrimer (blue).



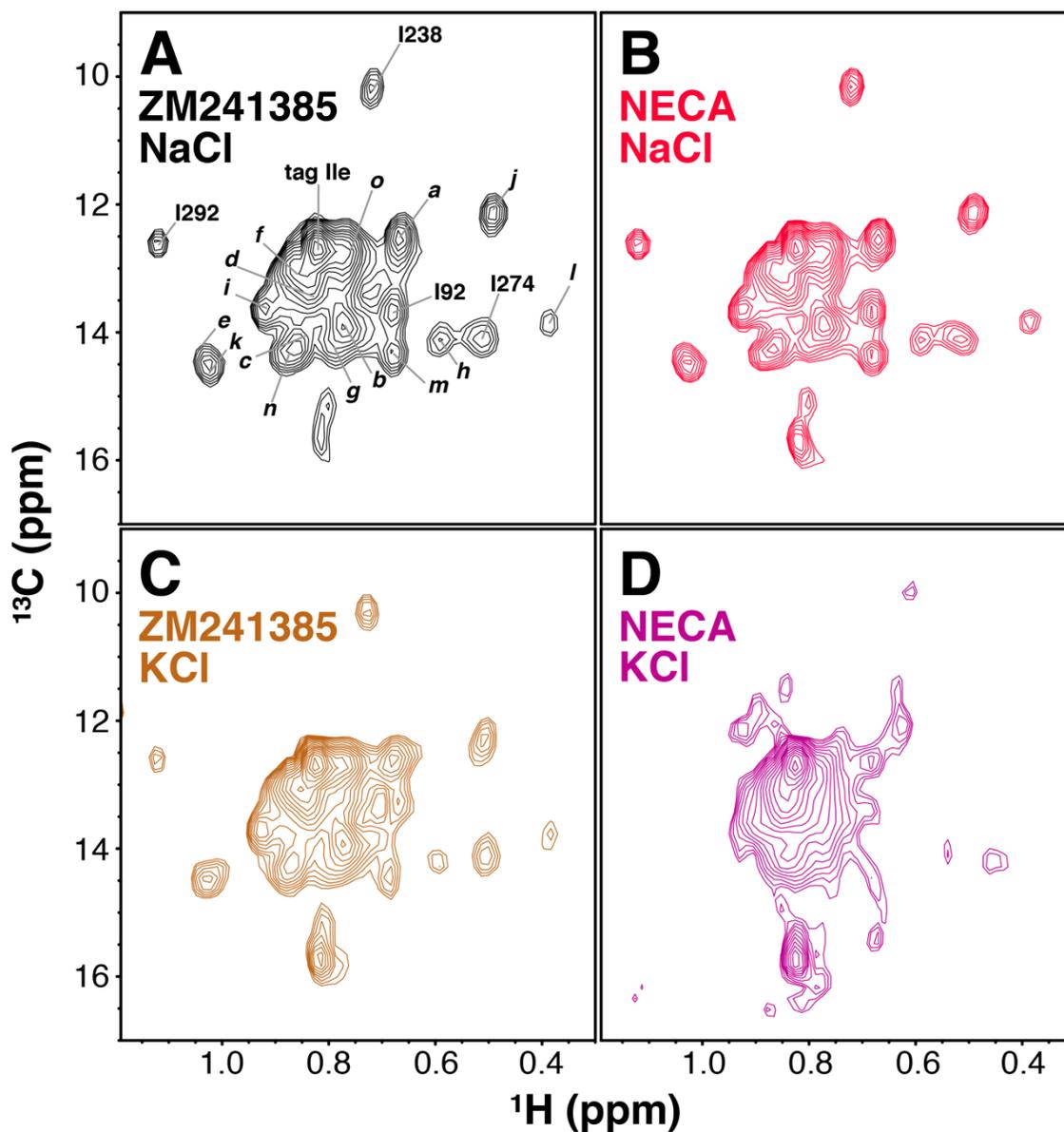
**Figure 2 supp. 1:  $A_{2A}R$  and  $G_s$  complex formation.** A) Gel filtration trace of purified  $A_{2A}R$  alone (black) and  $A_{2A}R$  complexed with  $G_s$  heterotrimer (red) injected on a Superdex200 10/300 column. Total protein inputs were 360  $\mu\text{g}$  and 833  $\mu\text{g}$ , respectively. No free  $A_{2A}R$  is detectable in the complex (red), demonstrating that the NECA-bound receptor is functional and the  $A_{2A}R-G_s$  complex is stable. B) Simplified schematic of  $A_{2A}R-G_s$  complex formation. After incubation, apyrase is added which hydrolyzes nucleotide released during complex formation. Free nucleotide can cause dissociation of the  $A_{2A}R-G_s$  complex.



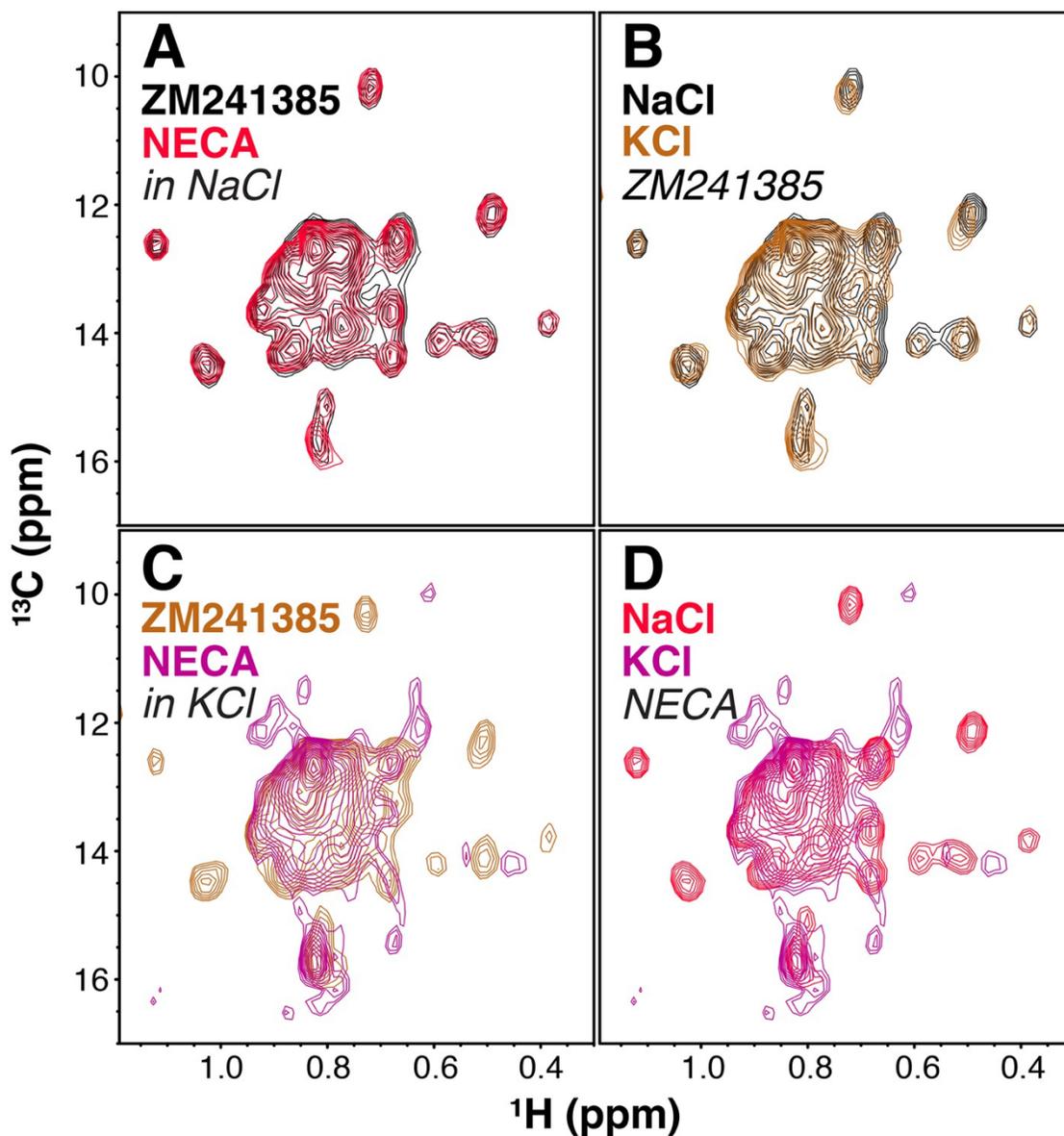
**Figure 2 supp. 2: Assignments of select Ile- $\delta$ 1 methyl resonances in  $A_{2A}R$ .** *top* -  $^1H/^{13}C$  HMQC spectra of Ile  $\delta$ 1-labeled mutants (blue) of  $A_{2A}R$  in DDM micelles with inverse agonist ZM241385 collected to make the assignments shown in Figure 2A, compared to the WT spectrum (black). The bottom-left panel shows the spectrum of apo WT (green)  $A_{2A}R$  in DDM micelles compared to the spectrum in the presence of ZM241385 (black). The lower right panel shows the spectrum of DDM alone (gray) to demonstrate the non-protein origin of the peak at  $\sim 0.82$  ppm  $^1H$  and 15-16 ppm  $^{13}C$ . *bottom* -  $^1H/^{13}C$  HMQC spectra of Ile  $\delta$ 1-labeled mutants (red) of  $A_{2A}R$  in DDM micelles with agonist NECA collected, compared to the WT spectrum (black). The bottom-left panel shows the spectrum of apo WT (green)  $A_{2A}R$  in DDM micelles compared to the spectrum in the presence of NECA (black).



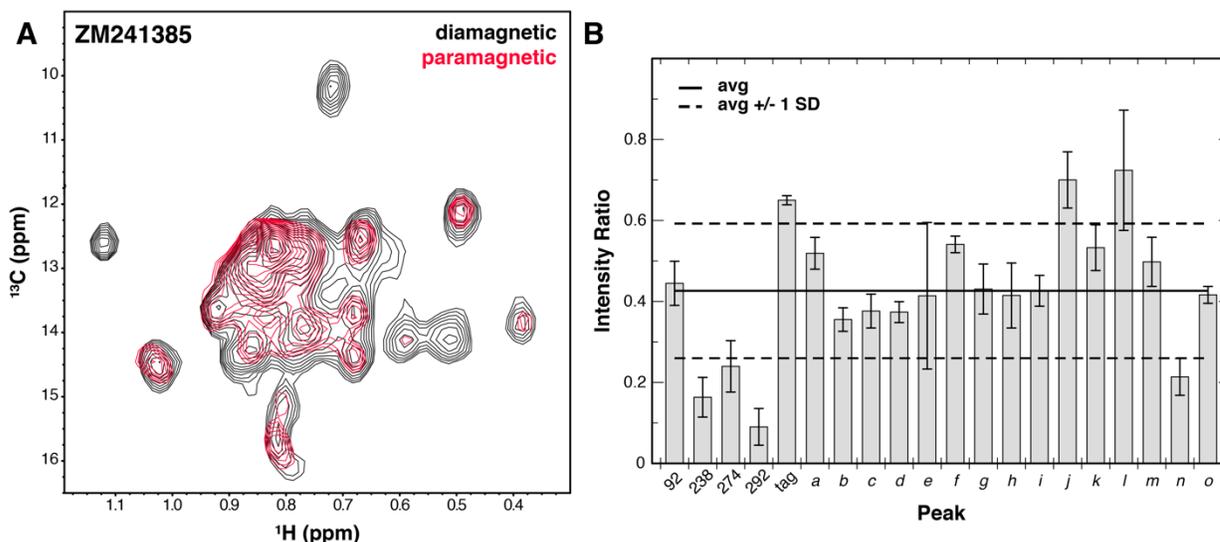
**Figure 2 supp. 3: Lifetime of A<sub>2A</sub>R NMR sample.** A)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles with agonist NECA collected immediately after final SEC column. B)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of the same Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles with agonist NECA collected after 7.5 hr of NMR experiment acquisition, showing loss of peak intensity but no major chemical shift changes.



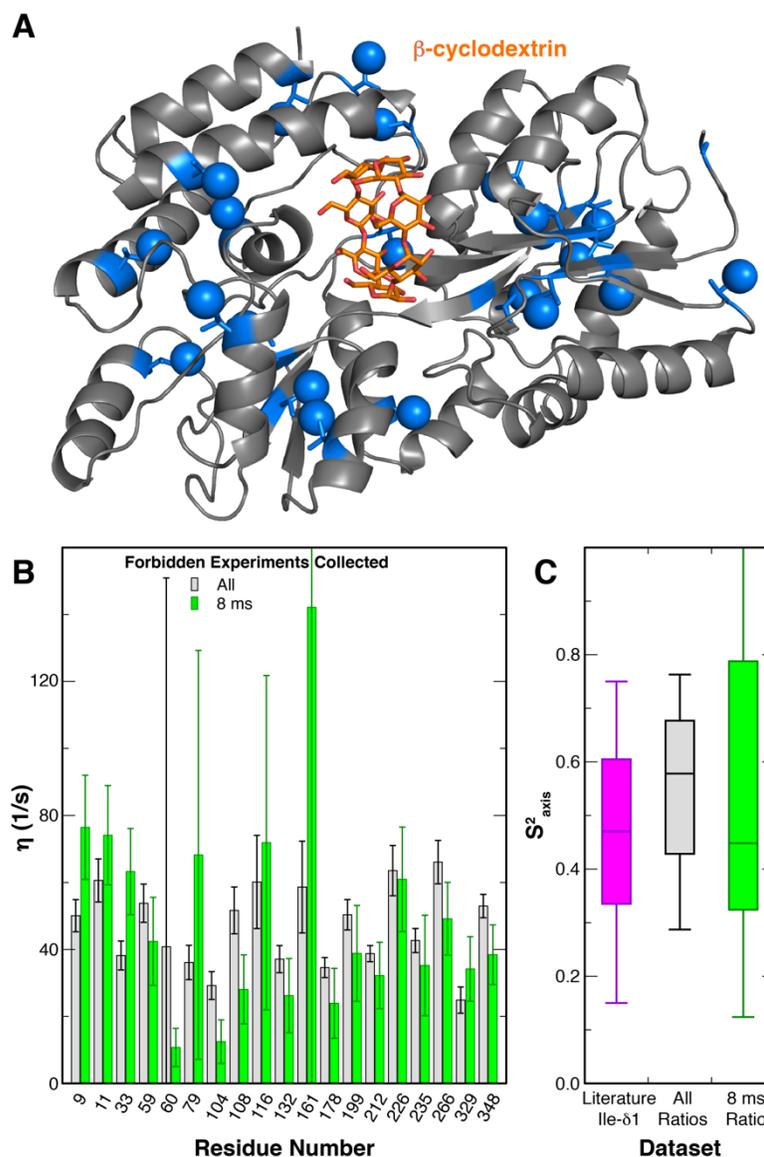
**Figure 3: Ligand- and cation-dependent chemical shift changes in  $\text{A}_{2\text{A}}\text{R}$ .** A)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles and 150 mM NaCl with inverse agonist ZM241385 (black). Residue numbers of assigned peaks and peak IDs of unassigned peaks are indicated in black. B)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles and 150 mM NaCl with agonist NECA (red). C)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles and 150 mM KCl with inverse agonist ZM241385 (orange). D)  $^1\text{H}/^{13}\text{C}$  HMQC spectrum of Ile  $\delta 1$ -labeled WT  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles and 150 mM KCl with agonist NECA (purple).



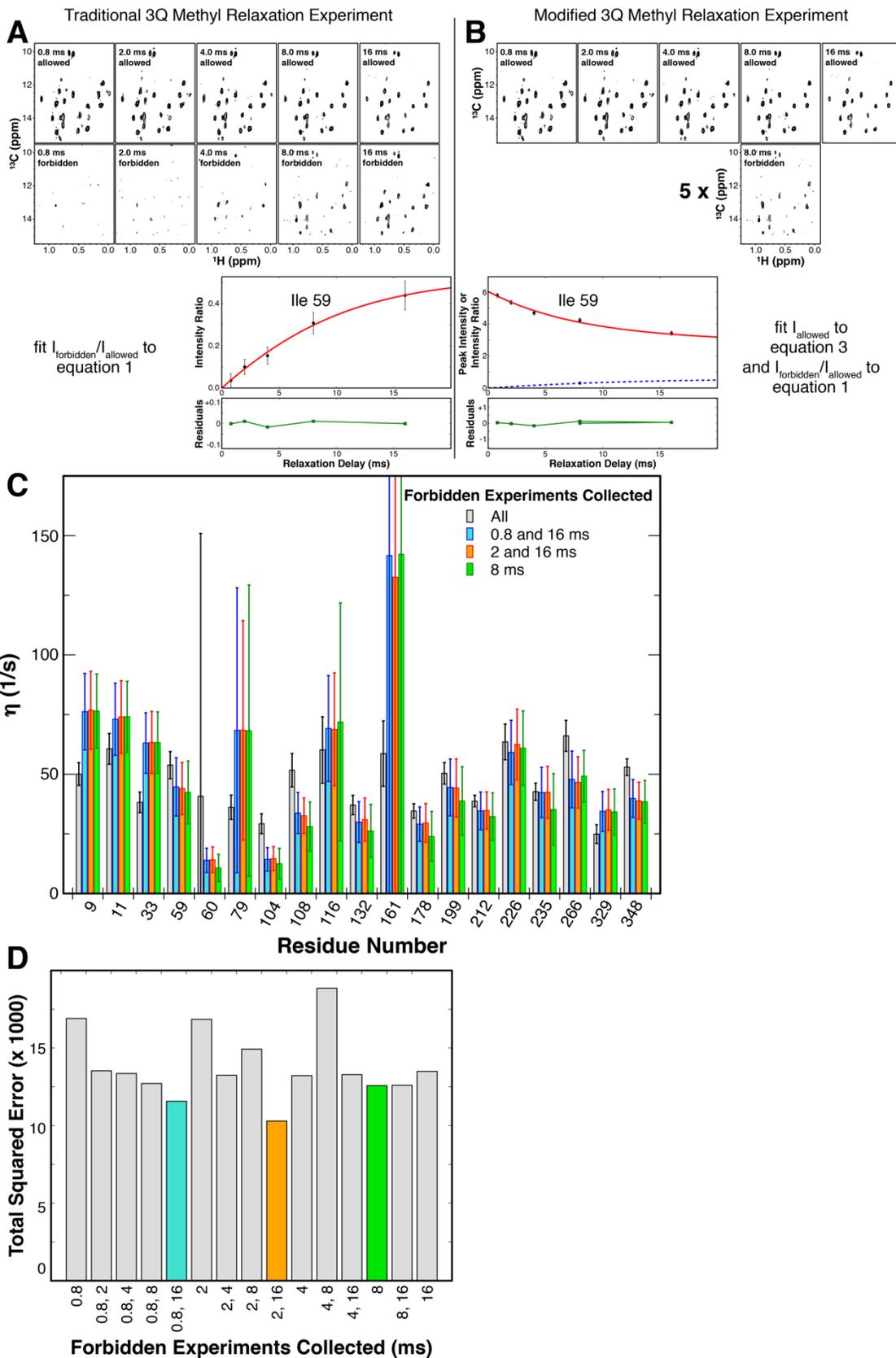
**Figure 3 supp. 1: Overlays of ligand- and cation-dependent chemical shift changes in A<sub>2A</sub>R.** Overlays of pairs of spectra in Figure 3. A)  $^1\text{H}/^{13}\text{C}$  HMQC spectra of Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles and 150 mM NaCl with inverse agonist ZM241385 (black) and agonist NECA (red). B)  $^1\text{H}/^{13}\text{C}$  HMQC spectra of Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles and 150 mM NaCl (black) or 150 mM KCl (orange) with inverse agonist ZM241385 (red). C)  $^1\text{H}/^{13}\text{C}$  HMQC spectra of Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles and 150 mM KCl with inverse agonist ZM241385 (burnt orange) and agonist NECA (purple). D)  $^1\text{H}/^{13}\text{C}$  HMQC spectra of Ile  $\delta$ 1-labeled WT A<sub>2A</sub>R in DDM micelles and 150 mM NaCl (red) or 150 mM KCl (purple) with agonist NECA.



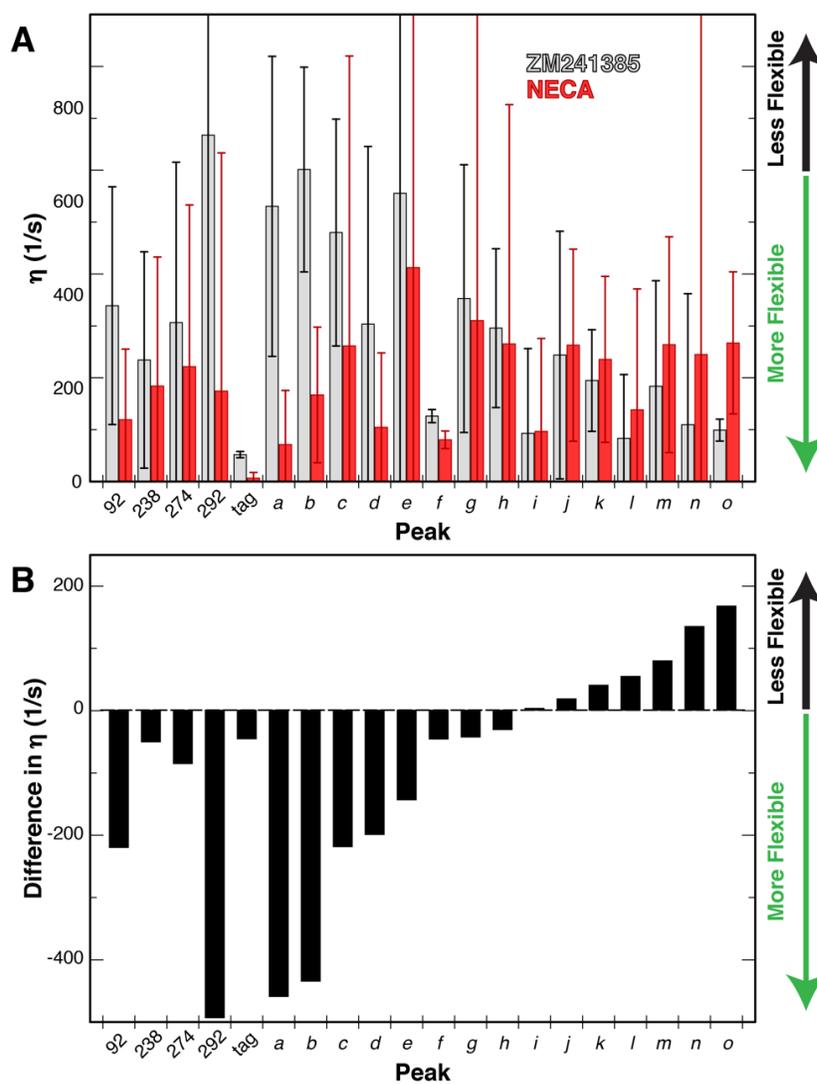
**Figure 4: Solvent PRE analysis of A<sub>2A</sub>R.** A) Solvent PRE <sup>1</sup>H/<sup>13</sup>C HMQC spectra of Ile δ<sub>1</sub>-labeled WT A<sub>2A</sub>R in DDM micelles with inverse agonist ZM241385 in the presence (red) and absence (black) of paramagnetic Gd<sup>3+</sup>-DTPA. Disappearance of peaks suggests solvent exposure. B) Plot of intensity ratios between paramagnetic and diamagnetic samples for 20 peaks in the Ile δ<sub>1</sub> region of the HMQC for A<sub>2A</sub>R complexed with inverse agonist ZM241385. The average ( $\pm 1$  standard deviation) intensity ratio is shown as solid (dashed) black lines. Error bars show errors propagated from the noise of the NMR spectra. Assigned peaks are on the left of the plot, unassigned on the right. Peak IDs correspond to those in Figure 3A.



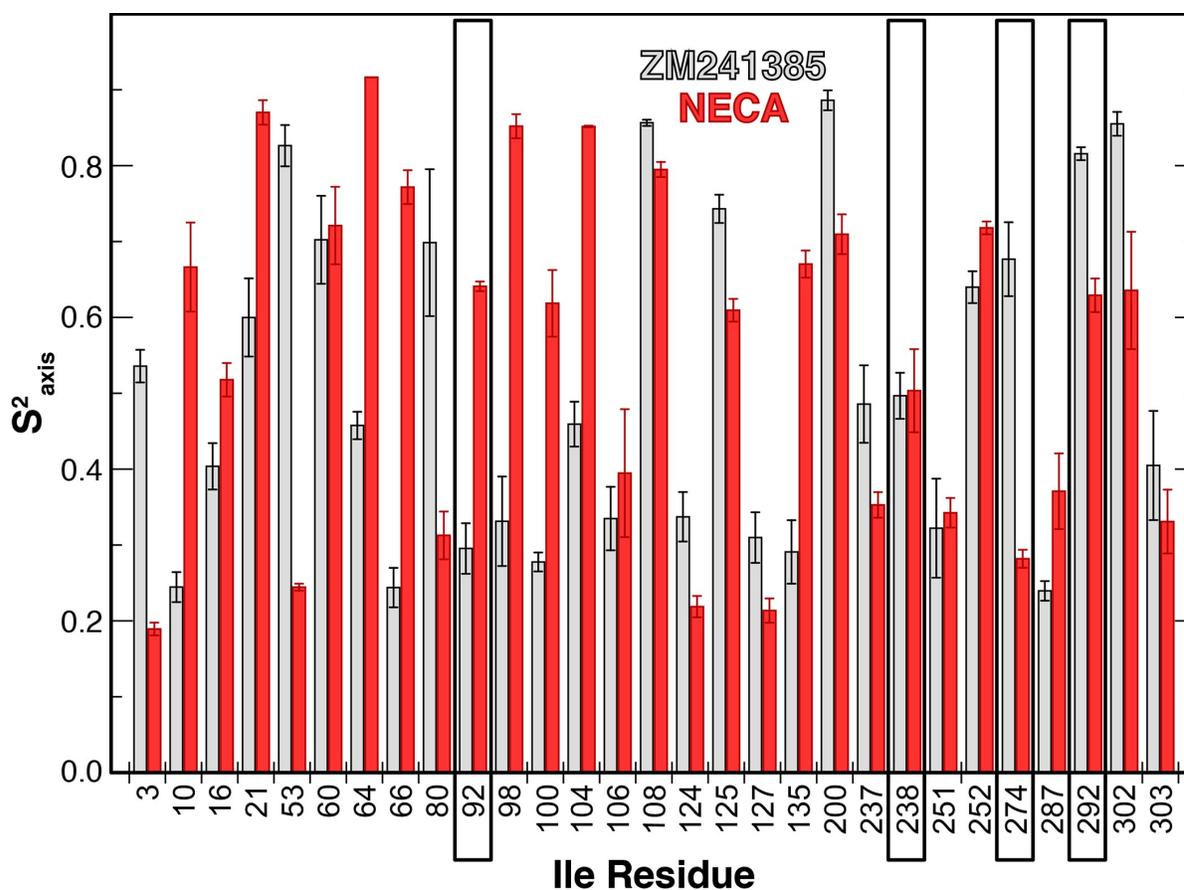
**Figure 5: Modified 3Q relaxation for MBP.** A) Crystal structure of MBP complexed with  $\beta$ -cyclodextrin (PDB 1DMB (Sharff *et al.*, 1993)). The  $\beta$ -cyclodextrin ligand is shown as orange sticks, while protein isoleucine residues are shown as blue sticks, with  $\delta 1$  carbon atoms as spheres. B) Plot of  $\eta$  values for each peak in MBP obtained from forbidden:allowed ratios at all five time points (gray) and the single 8 ms time point (green). Values calculated using the ratio at the single 8 ms time point show good agreement with those calculated using ratios at all five time points. C) Box-and-whisker plot showing Ile  $\delta 1$   $S^2_{axis}$  values from a set of globular proteins (Mittermaier *et al.*, 1999) (purple) and the  $S^2_{axis}$  values calculated from the  $\eta$  values using ratios at all time points (gray) and the single 8 ms time point (green).



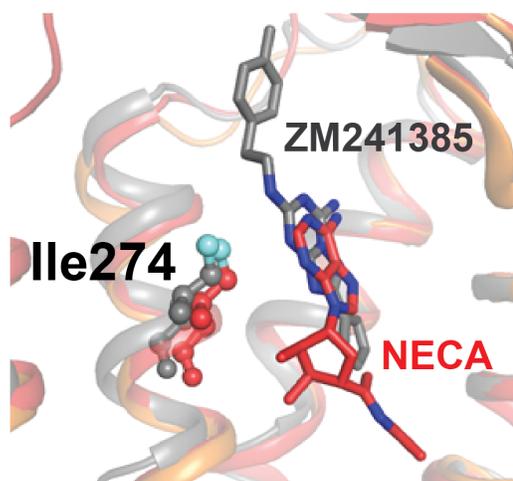
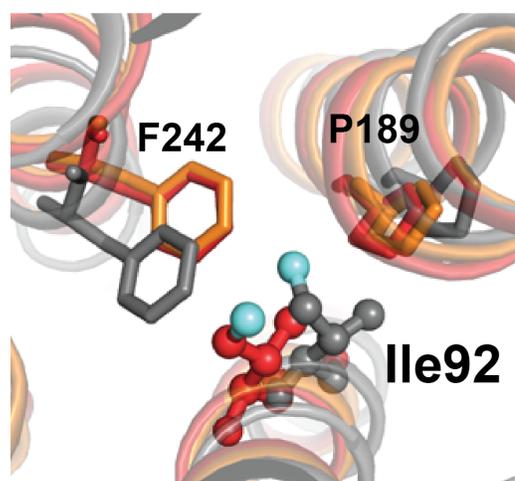
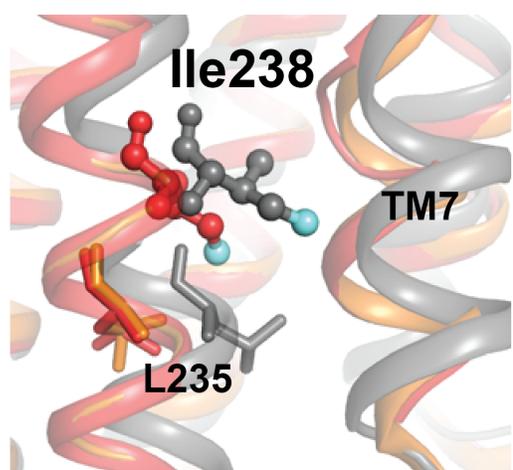
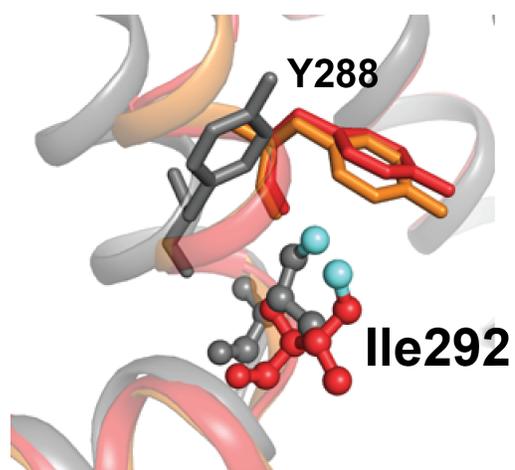
**Figure 5 supp. 1: Adaptation of 3Q methyl relaxation experiment to A<sub>2A</sub>R samples.** A) Traditional 3Q methyl relaxation experiments (Sun *et al.*, 2011) involve collecting matched pairs of allowed (SQ) and forbidden (MQ) spectra at several relaxation delay times, calculating the ratio of peak intensities, and fitting the buildup to a function of  $\eta$ , the difference between fast and slow R<sub>2</sub> relaxation rates. B) Due to short lifetime of A<sub>2A</sub>R samples (~12-14 hr at 30°C), the experiment was modified to reduce number of forbidden experiments, which have very low signal-to-noise and accordingly require long collection times. To do so, we evaluated alternatives requiring the collection of fewer forbidden spectra (down to the limiting case of a single spectrum), allowing for more scans for that time point. The decay in allowed spectra intensities was fit to a function of the fast and slow R<sub>2</sub> rates, and the one calculated forbidden:allowed ratio to the same function of  $\eta$ . C) A relaxation experiment with allowed and forbidden spectra at 5 relaxation delays was collected on MBP and  $\eta$  values were fit from forbidden:allowed ratios at all time points. The plot shows  $\eta$  values for each peak in MBP fit from forbidden:allowed ratios at all time points and three selected combinations of limited timepoints.  $\eta$  values calculated using the ratio at the single 8 ms time point show good agreement with the comparable values collected using other strategies, but can be collected in less time than the ones requiring two timepoints. D) Plot of the sum of squared differences between  $\eta$  values calculated using ratios at all timepoints and each combination of limited timepoints. The two combinations of 2 time points giving the lowest error, and the one single time point giving the lowest error are shown in blue, orange, and green respectively.



**Figure 6: Changes in dynamics of methyl groups of A<sub>2A</sub>R in the presence of different ligands.** A) Plot of the <sup>1</sup>H-<sup>1</sup>H dipolar cross-correlation rate  $\eta$ , which is proportional to the  $S^2_{axis}$  order parameter, for 20 peaks in the Ile  $\delta 1$  region for A<sub>2A</sub>R in DDM micelles with agonist NECA (red) and inverse agonist ZM241385 (grey). A higher  $\eta$  value corresponds to more rigid methyl groups. Error bars show errors of the fit parameters, propagated from the noise of the NMR spectra. Assigned peaks are on the left of the plot, unassigned on the right. Peak IDs correspond to those in Figure 3A. B) Plot of the ligand-induced differences ( $\Delta\eta = \eta[\text{NECA}] - \eta[\text{ZM241385}]$ ) for 20 peaks in the Ile  $\delta 1$  region for A<sub>2A</sub>R in DDM micelles. Negative  $\Delta\eta$  values, which are observed for the majority of peaks, indicate increased motions in the presence of agonist. Assigned peaks are on the left of the plot, unassigned on the right. Peak IDs correspond to those in Figure 3A.



**Figure 6 supp. 1: Changes in dynamics of methyl groups of A<sub>2A</sub>R in the presence of different ligands from MD trajectories.** Plot of the  $S^2_{axis}$  order parameter for all 29 Ile  $\delta$ 1 methyl groups extracted from two  $\sim$ 100-ns MD simulations of A<sub>2A</sub>R in DDM micelles with agonist NECA (red) and inverse agonist ZM241385 (grey). A higher  $S^2_{axis}$  value corresponds to more rigid methyl groups. Order parameters were extracted from overlapping 40-ns windows after simulations converged; bars show the average of these values  $\pm$  SD. Residues corresponding to assigned peaks are boxed.

**A****B****C****D**

**Figure 7: Structural contexts of isoleucine peak assignments.** A<sub>2A</sub>R ribbon diagrams are shown for structures solved in complex with inverse agonist ZM241385 (gray; PDB 4EIY (Liu et al., 2012b), agonist NECA (red; PDB 2YDV (Lebon et al., 2011), and agonist UK432097 (orange; 3QAK (Xu et al., 2011). Assigned isoleucine residues are shown as spheres for the ZM241385- and NECA-bound structures using the same coloring scheme as in Figure 1A. Additional highlighted residues are shown as sticks in B-D. **(A)** Isoleucine 274 is part of the orthosteric binding pocket and makes direct contact with the adenine analogue and adenine rings in ligands ZM241385 and NECA, respectively (shown as sticks). **(B)** Isoleucine 92 sits below the binding pocket and interacts with nearby residues Pro189 and Phe242. Ile92, Pro189, and Phe242 all make structural rearrangements upon receptor activation. **(C)** On TM6, assigned peak Ile238 is one helical turn above Leu235. Leu235 is part of a conserved microswitch region that undergoes outward motion during the transition to the active state. **(D)** Isoleucine 292 is at the cytoplasmic end of TM7 where it interacts with Tyr288, also part of the conserved NPXXY motif.

## CHAPTER FOUR

### Isotopic labeling of eukaryotic membrane proteins for NMR studies of interactions and dynamics

#### Abstract

Membrane proteins, and especially G-protein coupled receptors (GPCRs), are increasingly important targets of structural biology studies due to their involvement in many biomedically-critical pathways in humans. These proteins are often highly dynamic and thus benefit from studies by NMR spectroscopy in parallel with complementary crystallographic and cryo-EM analyses. However, such studies are often complicated by a range of practical concerns, including challenges in preparing suitably isotopically-labeled membrane protein samples, large sizes of protein/detergent or protein/lipid complexes, and limitations on sample concentrations and stabilities. Here we describe our approach to addressing these challenges via the use of simple eukaryotic expression systems and modified NMR experiments, using the human adenosine A<sub>2A</sub> receptor as an example. Protocols are provided for the preparation of U-<sup>2</sup>H (<sup>13</sup>C, <sup>1</sup>H-Ile δ1) labeled membrane proteins from overexpression in the methylotrophic yeast *Pichia pastoris*, as well as techniques for studying the fast nsec-psec sidechain dynamics of the methyl groups of such samples. We believe that, with the proper optimization, these protocols should be generalizable to other GPCRs and human membrane proteins.

## Introduction

Integral membrane proteins make up a large proportion of the genomes of many organisms – approximately 25% of the human genome – and perform a diverse range of functions, including key steps in the communication of a cell with its environment. Because of their biological and therapeutic importance (Almén, Nordström, Fredriksson, & Schiöth, 2009), membrane proteins are the focus of fundamental and applied biophysical research to characterize three-dimensional structures, dynamics, and interactions in native-like environments.

Solution state NMR spectroscopy has played a critical role in membrane protein biophysical studies, as the site-specific dynamic and interaction information provided by such approaches nicely complements structural data obtained from X-ray diffraction, cryo-EM and computational analyses (Cuniasse, Tavares, Orlova, & Zinn-Justin, 2017; Opella & Marassi, 2017). Fundamental to such studies are several 2D “fingerprint spectra,” most often  $^{15}\text{N}/^1\text{H}$  HSQC (heteronuclear single quantum coherence) spectra (for backbone amide plus Trp, Asn and Gln sidechains) or methyl  $^{13}\text{C}/^1\text{H}$  HMQC (heteronuclear multiple quantum coherence) spectra for sidechain methyl groups (Pellecchia, Bertini, Cowburn, Dalvit, Giralt, Jahnke *et al.*, 2008). These methyl-directed experiments are especially advantageous for large, slow-tumbling membrane protein/lipid complexes; experiments directed to other sidechain and mainchain sites have been successfully applied as well. NMR experiments can provide information about protein dynamics over many time scales,

from fast (psec-nsec) sidechain motions to slow conformational changes ( $\mu$ sec-msec) (Kasinath, Sharp, & Wand, 2013; Liang & Tamm, 2016; Palmer, 2012; Wand, Moorman, & Harpole, 2013). Many of these dynamics experiments, often using sidechain methyl groups as probes, have been adapted and developed for large biomolecular systems and can be used for membrane proteins (Rosenzweig & Kay, 2014; Sun, Kay, & Tugarinov, 2011; Tugarinov, Hwang, Ollerenshaw, & Kay, 2003).

A particular advantage of solution state NMR is that proteins are studied in a native-like solution state where they can interconvert among multiple conformations. However, membrane proteins must be solubilized in a suitable membrane mimetic that maintains native structure and dynamics. Different options include detergent micelles, amphipols, bicelles, nanodiscs, SMALPs, and lipid vesicles, each having their own benefits and drawbacks (Liang & Tamm, 2016, 2018; Zhou & Cross, 2013). It is often necessary to test different solubilization strategies for a given protein sample for stability, signal intensity and resolution, and native structure/activity. Two important considerations for all membrane mimetics are 1) a uniform and small particle size and 2) a high extent of deuteration.

High-level deuteration, both within the membrane mimetic and protein itself, is critical to reduce the number of  $^1\text{H}$  signals present in spectra (including those from lipids, which can be intense) and to improve the relaxation characteristics of the remaining NMR-active spins in the sample. While deuteration is possible for the membrane-mimetic through the purchase / synthesis of deuterated compounds,

replacing  $^1\text{H}$  with  $^2\text{H}$  in proteins requires biosynthetic incorporation. For backbone experiments in eukaryotic expression systems, one can label uniformly with  $^{15}\text{N}$  to observe all amides (Eddy, Lee, Gao, White, Didenko, Horst *et al.*, 2018; Opitz, Isogai, & Grzesiek, 2015) or through addition of specifically-labeled amino acids (Isogai, Deupi, Opitz, Heydenreich, Tsai, Brueckner *et al.*, 2016). For methyl groups, one can provide either appropriately-labeled amino acids or amino acid precursors (particularly alpha-keto acids) to growth media to access various labeling patterns in the sidechains of several amino acids (Kofuku, Ueda, Okude, Shiraishi, Kondo, Mizumura *et al.*, 2014; Kofuku, Yokomizo, Imai, Shiraishi, Natsume, Itoh *et al.*, 2018). We find isoleucine  $\delta 1$  methyl groups particularly useful given 1) the abundance of Ile residues in integral membrane proteins including GPCRs (Ulmschneider & Sansom, 2001), 2) the far upfield  $^{13}\text{C}$  shift of isoleucine  $\delta 1$  methyl groups [average  $13.5 \pm 3.6$  ppm  $^{13}\text{C}$  according to BioMagResBank (Ulrich, Akutsu, Doreleijers, Harano, Ioannidis, Lin *et al.*, 2008)], putting them in a particularly uncrowded region of 2D  $^{13}\text{C}/^1\text{H}$  spectra, 3) the lack of need to stereospecifically assign these methyl groups, unlike Val and Leu, and 4) the presence of multiple, freely-rotatable bonds between the methyl group and protein backbone, providing substantial independence of dynamics at these sites (Kasinath *et al.*, 2013).

While many of the aforementioned labeling strategies have been well developed for *E. coli*, many integral membrane proteins can only be expressed at high levels in eukaryotic hosts. Among these, the methylotrophic yeast *Pichia*

*pastoris* is a convenient host for heterologous expression and isotopic labeling of eukaryotic membrane proteins (Clark, Dikiy, Rosenbaum, & Gardner, 2018). Advantages of *Pichia* include rapidity of genetic manipulation, high yields of recombinant protein, existence of post-translational modification and chaperone machinery necessary for eukaryotic membrane proteins, and ability to grow on defined minimal media allowing for perdeuteration (J. L. Cereghino & Cregg, 2000; Morgan, Kragt, & Feeney, 2000). Uniform isotopic labeling in *Pichia* has been well established (Morgan *et al.*, 2000; Pickford & O'Leary, 2004). We have extended this work by demonstrating the  $^{13}\text{C}$ ,  $^1\text{H}$  labeling of isoleucine  $\delta^1$ -methyl groups in a perdeuterated background by adding labeled  $\alpha$ -ketobutyrate (~50% labeling, ~90% deuteration) to highly deuterated growth media (Clark, Dikiy, Chapman, Rodstrom, Aramini, LeVine *et al.*, 2017; Clark, Zahm, Ali, Kukula, Bian, Patrie *et al.*, 2015). In contrast, simultaneous labeling of leucine  $\delta$ - and valine  $\gamma$ -methyl groups with  $\alpha$ -ketoisovalerate is inefficient but can be achieved by adding labeled valine directly to the growth media or modifying culture conditions (Clark *et al.*, 2015; Suzuki, Sakakura, Mori, Fujii, Akashi, & Takahashi, 2018; Zhang, Yu, Xu, Jouhten, Swapna, Glaser *et al.*, 2017). *Pichia* can readily take up additional amino acids from media, with a general correlation between uptake efficiency and the energetic cost to synthesize that amino acid type *de novo* (Heyland, Fu, Blank, & Schmid, 2011). However, after uptake into cells, labeled amino acids can be fed into metabolic pathways (Solà, Maaheimo, Ylonen, Ferrer, & Szyperski, 2004), diluting signal of

desired amino acids and complicating data analysis by isotopic scrambling.

Alternatively, auxotrophic strains can be developed for labeling a specific amino acid; however, care must be taken to confirm that off-target effects in other metabolic pathways do not arise (Whittaker, 2007).

Here we provide detailed protocols needed to generate such U-<sup>2</sup>H (<sup>13</sup>C, <sup>1</sup>H-Ile δ1 methyl) labeled integral membrane proteins by overexpression in *Pichia*, using the human adenosine A<sub>2A</sub> receptor [A<sub>2A</sub>R] as a model system. We further detail how such samples can be used in solution NMR studies, from acquiring simple <sup>13</sup>C/<sup>1</sup>H HMQC spectra, through chemical shift assignments by site-directed mutagenesis, to analyses of <sup>1</sup>H-<sup>1</sup>H cross-relaxation measurements of fast sidechain dynamics.

### **Expression and isotopic labeling in *Pichia***

The overall strategy for expression of membrane proteins in *Pichia* is outlined below. After cloning the gene of interest into a *Pichia* expression vector [see Note 1 below and several literature reviews where such reagents are comprehensively discussed al.(Byrne, 2015; Cos, Ramon, Montesinos, & Valero, 2006; Fan, Emami, Munro, Ladizhansky, & Brown, 2015)], the total process to obtaining initial expression cultures takes about a week. While *Pichia* can be grown in large-scale fermenters (G. P. Cereghino, Cereghino, Ilgen, & Cregg, 2002), this protocol describes growth in shake-flasks, as this is the most straightforward and cost-efficient way to initiate *Pichia* expression at levels sufficient for structural biology

work. Although the process outlined here is a good general starting point for membrane protein expression, it is likely that each sample will require optimization to achieve high levels of expression, and pertinent notes are indicated after each step.

### **Equipment**

1. Plate incubator (28° C)
2. DNA agarose gel apparatus
3. Electroporator (e.g. Bio-Rad Micropulser) and cuvettes
4. Temperature-controlled shake flask incubator, ideally dedicated for *Pichia* expression

### **Required buffers and media (all solutions 0.2 $\mu$ m sterile-filtered, store at 4° C)**

1. 1 M sorbitol
2. 20% w/v dextrose (use at 10x stock)
3. 10x YNB (per 1L: 34 g yeast nitrogen base without amino acids or nitrogen, 100 g ammonium sulfate)
4. 1 M Potassium phosphate pH 6.0 (use as 10x stock)
5. 2000x biotin (0.08% in 0.05 M NaOH)
6. 200x histidine (0.8% in water)
7. MD agar (7.5% agar, 1x YNB, 1x dextrose, 1x biotin, 1x histidine if required ("MDH" plates); add all stock solutions after autoclaving agar and water)

8. 10% w/v glycerol (use as 10 x stock)
9. YPD (1% yeast extract, 2% peptone, 1 x dextrose; dextrose solution added after autoclaving other components)
10. YPDS (YPD + 2% agar + 1 M sorbitol; sorbitol added after autoclaving)
11. BMG (1x glycerol, 1x potassium phosphate pH 6.0, 1x YNB, 1x biotin, and 1x histidine if required)
12. BMM without methanol (1x potassium phosphate pH 6.0, 1x YNB, 1x biotin, 1x histidine if required)

## Procedure

1. Clone the gene of interest into the desired vector for expression in *Pichia*. The cDNA for wild-type human ADORA2A adenosine A<sub>2A</sub> receptor with C-terminal His<sub>6</sub> and Protein C affinity tags was cloned into the pPICZ vector containing a modified MF $\alpha$  secretion signal (see **Note 1** below). Molecular biology steps were carried out using standard techniques and utilizing Zeocin selection.
  - a. **Note 1:** Although some membrane proteins have been successfully purified when expressing intracellularly (Byrne, 2015; J. Y. Lee, Kinch, Borek, Wang, Wang, Urbatsch *et al.*, 2016), we have observed that the best expression of A<sub>2A</sub>R occurred when the gene was preceded by the MF $\alpha$  signal sequence from *Saccharomyces cerevisiae*. Furthermore, the highest amount of fully-processed (e.g. plasma membrane

localized) receptor occurred when the signal sequence was modified for more efficient processing. These modifications were based on previous literature precedents (Lin-Cereghino, Stark, Kim, Chang, Shaheen, Poerwanto et al., 2013; Rakestraw, Sazinsky, Piatasi, Antipov, & Wittrup, 2009), comprising the V22A, G40D, L42S, V50A, V52A, and F55L mutations plus the deletion of residues 57-70 (**Fig. 1**).

- b. **Note 2:** GPCRs and other membrane proteins commonly contain post-translational modifications (PTMs) and/or large flexible regions that are problematic for NMR or other structural analyses (Yurugi-Kobayashi, Asada, Shiroishi, Shimamura, Funamoto, Katsuta *et al.*, 2009). PTM target residues can be removed through point mutations and flexible regions can be truncated during this stage. While *Pichia* does not hyperglycosylate overexpressed proteins to the extent of other yeast such as *S. cerevisiae* (Pickford & O'Leary, 2004), a simple way to minimize this possibility is to remove targeted asparagine residues. To avoid such issues with A<sub>2A</sub>R, we mutated an N-glycosylation site at Asn154 to glutamine and truncated the construct at residue 316 as previously described (Jaakola, Griffith, Hanson, Cherezov, Chien, Lane *et al.*, 2008).

2. Prepare a fresh batch of competent KM71H cells. Briefly, cells should be grown from a freshly streaked MD plate in YPD to an OD<sub>600</sub> of 1, then spun

down under sterile conditions at 4000 rcf for 10 minutes. Cells were washed and repelleted in ice-cold, sterile Milli-Q water twice, followed by a wash in ice-cold 1 M sorbitol and resuspension in a minimal volume of 1 M sorbitol to create a thick, but pipettable, cell paste (e.g. ~500  $\mu$ L for 500 mL pellet).

Competent cells may be kept on ice at this point until electroporation.

- a. **Note 1:** It is important that the competent cells are grown and prepared fresh for efficient transformation, preferably no more than 24 hours prior to electroporation. Sterility must be maintained as much as possible.
  - b. **Note 2:** The popularity of *Pichia* as an expression host has expanded the options for expression and culturing conditions. *Pichia* is commercially available in various cell strains from several vendors including Invitrogen, including the most commonly used strains for MP expression (KM71/H, GS115, SMD1163/8, and wild-type X-33). KM71H is the strain that gave us the highest expression levels of A<sub>2A</sub>R, although other membrane proteins have been successfully overexpressed and purified from alternative strains for structural studies (Byrne, 2015).
3. Purify 10-20  $\mu$ g of plasmid prior to linearization. Linearize with *PmeI* or appropriate restriction enzyme and use a small amount of plasmid to verify efficient digestion with a DNA agarose gel. Re-purify linearized plasmid with

- PCR Clean-up Kit (Qiagen), with the final elution step carried out in a minimal volume of sterile Milli-Q water.
- a. **Note:** It is important to use a large amount of linearized plasmid (5  $\mu\text{g}$  is ideal) for each electroporation to maximize efficiency of incorporation into the *Pichia* genome. If screening multiple *Pichia* strains, it is helpful to increase the scale of plasmid linearization and divide the resultant DNA between strains.
4. Add linearized plasmid to  $\sim 100 \mu\text{L}$  of competent *Pichia*, and transfer to an electroporation cuvette with a 0.2 cm gap. Include an extra cuvette containing cells and an equal volume of sterile water instead of plasmid as a negative control. Process each cuvette with 1 pulse of 2.00 kV (“Pic” program on Bio-Rad MicroPulser system), and immediately transfer cuvette back to ice. Add 1 mL of ice-cold 1 M sorbitol and transfer the contents of the cuvette to a 14 mL round-bottom culture tube.
    - a. **Note:** Arcing during electroporation is rare but can happen if there is salt in the sample. This risk can be minimized by ensuring that the paste of competent cells has been thoroughly rinsed of media and eluting linearized plasmid with water instead of elution buffer.
  5. Incubate cells at 28 °C for 30-60 minutes without shaking, and then add 1 mL YPD to each tube and shake at 200 rpm for 4 hours at 28 °C.

6. Plate cells on YPDS plates containing increasing amounts of Zeocin (100  $\mu\text{g}/\text{mL}$  – 1000+  $\mu\text{g}/\text{mL}$ ) to check for efficiency of gene integration. Incubate plates at 28 °C for 3-4 days. “Pinprick” colonies should appear after 48 hours and individual colonies should be obvious (~1 mm diameter) after 72 hours.
7. Individual colonies should be expanded and screened for expression levels, generally through western blots, as simply using the copy number as a proxy for expression levels is not advisable (Aw & Polizzi, 2013). Typically, 10-12 clones from each sample are selected mostly (but not solely) from the plates containing the higher levels of Zeocin and easily expanded on a single MD (or MDH) plate. Individually pick colonies with a sterile pipet tip and gently spread each colony on a labeled area of the plate. After 24 hours of incubation at 28 °C, each colony will grow into a small “patch” that can be easily looped for screening cultures and subsequent preservation (see **Step 10** below).
8. Grow a small culture (~10 mL) of each clone by inoculating a loopful of yeast into sterile BMG media and incubating at 200 rpm/28 °C. After growing to saturation ( $\text{OD}_{600}$  ~20-30), spin down cells at 4000 rcf for 5 minutes. While maintaining sterility, remove media and replace with an equal volume of BMM media without methanol. After continued shaking at 28 °C for 4-6 hours to allow for metabolism of residual glycerol, induce expression with the addition of 0.5% methanol. Additional aliquots of 0.5% methanol should be added

every 12 hours to maintain expression, and cells can be harvested and subjected to western blots after 48 hours.

- a. **Note 1:** Although each sample will likely require individual optimization to achieve the best expression conditions, the strategy outlined above can be viewed as a general starting point. Common additives to increase expression levels are DMSO (André, Cherouati, Prual, Steffan, Zeder-Lutz, Magnin *et al.*, 2006), histidine (even if not required by cell strain as it is an energetically “expensive” amino acid to synthesize; (Heyland *et al.*, 2011)), and individual ligands, especially in the case of GPCRs (André *et al.*, 2006; Clark *et al.*, 2017; Hino, Arakawa, Iwanari, Yurugi-Kobayashi, Ikeda-Suno, Nakada-Nakura *et al.*, 2012; Shimamura, Shiroishi, Weyand, Tsujimoto, Winter, Katritch *et al.*, 2011). It is also common to use unbuffered media (*J. Y. Lee et al.*, 2016), BMG/BMM at different pH levels (known to influence expression and activity levels of various GPCRs (André *et al.*, 2006; Yurugi-Kobayashi *et al.*, 2009)), or alter the temperature and length of expression (André *et al.*, 2006; Yurugi-Kobayashi *et al.*, 2009). In all cases, effects on protein expression can be assayed for with western blots.
- b. **Note 2:** If methanol is not suitable for expression induction due to toxicity or other sample constraints, proteins can be expressed

constitutively through use of the commercially available pGAPZ vectors (Invitrogen).

9. Once the best expression clone is identified, it can be stored as a glycerol stock at  $-80\text{ }^{\circ}\text{C}$  and streaked onto fresh MD (or MDH) plates as needed. To prepare a culture for a glycerol stock, inoculate a loopful of the colony “patch” in YPD and shake for 24 hours at  $28\text{ }^{\circ}\text{C}$ . Store culture in cryo vials with a final glycerol concentration of 25%.
10. To grow deuterated and isotopically labeled cultures for NMR experiments, use a loopful of freshly streaked cells from an MD (or MDH) plate and inoculate 50 mL of BMG media made with 90%  $\text{D}_2\text{O}$  / 10%  $\text{H}_2\text{O}$ . Shake the culture at  $28\text{ }^{\circ}\text{C}$  until it reaches an  $\text{OD}_{600}$  of  $\sim 10$  (approximately 36-48 hours). Use 200  $\mu\text{L}$  of this culture to inoculate 50 mL of BMG media made with 100%  $\text{D}_2\text{O}$ . Shake the culture at  $28\text{ }^{\circ}\text{C}$  until it reaches an  $\text{OD}_{600}$  of  $\sim 10$  (approximately 36-48 hours). Use 200  $\mu\text{L}$  of this culture to inoculate 50 mL of BMG media made with 100%  $\text{D}_2\text{O}$ , but replacing the 1% glycerol with 1%  $\text{d}_8$ -glycerol. Shake the culture until reaching an  $\text{OD}_{600}$  of  $\sim 10$ , and then use the culture in its entirety to inoculate large-scale (1-2 L) cultures of identical media. Shake the large-scale cultures at  $28\text{ }^{\circ}\text{C}$  until reaching saturation ( $\text{OD}_{600}$  of  $\sim 20$ -30), and then pellet cells in sterile 1 L bottles at 4000 rcf for 30 minutes. Remove the media and gently resuspend cell pellets in BMM media without methanol. Continue to shake flasks at  $28\text{ }^{\circ}\text{C}$  for  $\sim 12$  more hours to

metabolize any remaining d<sub>8</sub>-glycerol. One hour prior to induction, add 200 mg/L of labeled  $\alpha$ -ketobutyric acid (methyl-<sup>13</sup>C, 99%; 3,3-D<sub>2</sub>, 98%; Cambridge Isotope Laboratories). For A<sub>2A</sub>R cultures, add 4 mM dry theophylline (low-affinity antagonist) ten minutes prior to induction, and induce with 0.5% d<sub>4</sub>-methanol. Reduce the temperature to 20 °C after induction. As with the small test cultures, add additional aliquots of 0.5% d<sub>4</sub>-methanol roughly every 12 hours to maintain expression levels. Harvest cells after 36-48 hours and store at -80 °C. See **Fig. 2** for a schematic of the deuterated media adaptation procedure.

- a. **Note 1:** Sterility should be maintained as much as possible during adaptation due to the lack of antibiotics in the media. Typically, this requires a dedicated shaker for *Pichia* cultures and a clean flame area for inoculations and media transfers.
- b. **Note 2:** Instead of d<sub>8</sub>-glycerol, deuterated glucose can be used as an alternative (and somewhat less expensive) carbon source in *Pichia* cultures (Emami, Fan, Munro, Ladizhansky, & Brown, 2013; Fan *et al.*, 2015; Fan, Shi, Ladizhansky, & Brown, 2011). However, glucose is a very strong repressor of the AOX promoter, thus any residual glucose in the cells or media can result in toxic conditions during induction (Wood & Komives, 1999).

- c. **Note 3:** Deuterated media can be saved and recycled to offset some of the costs of deuteration. Recycling can be done in-house by flash distillation (Moore, 1979) or shipped to Cambridge Isotope Laboratories for their D<sub>2</sub>O recovery program (<https://www.isotope.com/sales-technical-support/D2O.cfm>).

### **Purification/NMR sample preparation**

During optimization of culture and expression conditions (prior to purification), it is important to validate expression of A<sub>2A</sub>R or other target proteins via western blots using specific antibodies to the protein or affinity tag(s). In our experience, a 10-20  $\mu$ L volume of cells from the final harvested culture is typically sufficient to obtain a robustly-visible signal with an anti-His<sub>6</sub> antibody (Pierce) for A<sub>2A</sub>R or other His-tagged membrane proteins. Once an optimized level of expression is demonstrated, the receptor can be purified from *Pichia* pellets following the procedure described below. Ideally, the purification should be “straight-through,” without any substantial delays from cell disruption to final sample; if required, the affinity-purified receptor can be flash-frozen in liquid N<sub>2</sub> prior to SEC with minimal (10-15%) losses of final yield.

### Equipment

Stir plate

Glass Dounce homogenizer with loose and tight plungers

Refrigerated ultracentrifuge (100,000+ rcf)

Orbital shaker

FPLC with Superdex200 or comparable size exclusion column

### Buffers

1. Lysis buffer: PBS containing 10% glycerol, 4 mM theophylline (low-affinity antagonist for A<sub>2A</sub>R), 2 mM EDTA, and protease inhibitors (160 µg/mL benzamidine, 2.5 µg/mL leupeptin, 1 mM PMSF, 1 µM E-64)
2. Hypotonic buffer: 10 mM HEPES pH 7.5, 2 mM EDTA, 4 mM theophylline, protease inhibitors as above.
3. Solubilization buffer: 500 mM NaCl, 50 mM HEPES pH 7.5, 20% glycerol, 1% w/v DDM (Anatrace), 4 mM theophylline, and protease inhibitors as above.
4. TALON equilibration buffer: 250 mM NaCl, 50 mM HEPES pH 7.5, 0.05% w/v protonated DDM, 5% glycerol, 4 mM theophylline, and 30 mM imidazole.
5. Buffers A-E: 250 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 20 mM imidazole, ligand of interest (e.g., 10 µM of the A<sub>2A</sub>R antagonist ZM241385 or 20 µM of the agonist NECA), made in D<sub>2</sub>O, and containing 0.05% w/v DDM in

the following protonated:deuterated ratios: (A) 4:0; (B) 3:1; (C) 2:2; (D) 1:3; (E) 0:4.

6. SEC buffer: 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% w/v deuterated DDM, ligand of interest (e.g., 10  $\mu$ M ZM241385 or 20  $\mu$ M NECA), 99% D<sub>2</sub>O.  
Store at 4 °C after 0.2  $\mu$ m sterile filtering.

## Procedure

### *Day 1*

1. Thaw cell pellets and resuspend in 200-300 mL cold lysis buffer. All volumes in this protocol are given for a culture of 1 L, and should be scaled appropriately for different volumes.
2. Lyse cells with a high-pressure microfluidizer (Microfluidics M-110P or similar), using three passes at 24,000 psi, allowing the lysate to cool between passes. Pack ice or ice packs around microfluidizer coils to minimize sample heating.
3. To digest yeast cell wall glycans, add zymolyase (LongLife Zymolyase from G-Bio Sciences or similar) to 15 U/mL and incubate at 37 °C for 1 hour with stirring.
4. Separate out total membranes by centrifugation at 140,000 rcf for 30 minutes. Carefully decant the supernatant and keep membrane pellets on ice.

5. To wash membranes, resuspend the pellets in 200-300 mL of cold lysis buffer using a Dounce homogenizer (loose, 10 passes), then centrifuge at 140,000 rcf for 30 minutes again. Carefully decant the supernatant and keep membrane pellets on ice.
6. To lyse any intact spheroplasts and further wash the membranes, resuspend the pellets in 200-300 mL of cold hypotonic buffer using a Dounce homogenizer (loose, 10 passes). Incubate at 4 °C for 30 min with stirring, then centrifuge at 140,000 rcf for 30 minutes. Carefully decant the supernatant and keep membrane pellets on ice.
7. To solubilize membranes, resuspend the pellets in cold solubilization buffer using a Dounce homogenizer (loose, 10 passes; tight, 30 passes) and incubate at 4 °C for 2 hours with stirring. Centrifuge at 140,000 rcf for 30 minutes to remove insoluble material.
  - a. **Note:** The addition of cholesteryl hemisuccinate (CHS) to the micelles as a cholesterol analogue has been demonstrated to increase the stability and activity of A<sub>2A</sub>R and other GPCRs (Liu, Chun, Thompson, Chubukov, Xu, Katritch *et al.*, 2012; Weiss & Grisshammer, 2002). We initially solubilized and purified A<sub>2A</sub>R in DDM buffer with CHS included at a concentration of 0.2% and 0.01% w/v, respectively (**Fig. 3A**). While the inclusion of CHS improved NMR sample stability (**Fig. 3B**), it gave rise to an intense methyl peak at around 25 ppm <sup>13</sup>C/0.85 ppm <sup>1</sup>H

that interfered with analyses of NMR spectra (particularly relaxation measurements detailed below, which were extremely difficult to analyze until CHS was removed from the preparation). We note that a small peak at the same chemical shift remained in our CHS-free samples (**Fig. 3A**), suggesting that the receptor copurified with a cholesterol-like molecule.

8. During the final centrifugation step, pre-equilibrate 10-15 mL of fresh immobilized cobalt resin (such as TALON resin from Clontech) in TALON equilibration buffer. After centrifugation, decant supernatant into 250 mL tubes and add 30 mM imidazole to reduce background nonspecific binding. Incubate the supernatant with the equilibrated resin and bind overnight at 4 °C with gentle agitation.

## *Day 2*

1. Spin down loose TALON resin at 2000 rcf for 10 minutes, remove the majority of the unbound material, and transfer to a gravity flow column. Through a series of wash steps, exchange the protonated DDM for deuterated DDM and simultaneously exchange theophylline (low-affinity antagonist) for a high-affinity ligand of interest (or no ligand). This is achieved by washing resin sequentially with 4 column volumes (CV) each of buffers A-E. Wash steps

- should be carried out at 4 °C and at a flow rate of 1-1.5 mL/min to ensure efficient exchange.
2. Elute tagged protein with Buffer E + 250 mM imidazole. A<sub>2A</sub>R should elute within 2 CV. If necessary, purified A<sub>2A</sub>R can be snap frozen in liquid N<sub>2</sub> in a buffer containing glycerol at a final concentration of 15% (e.g. addition of 10% glycerol to eluate).
    - a. **Note:** Another option for solution state NMR is reconstituting the proteins into nanodiscs, in which the protein is surrounded by lipids held together by an apolipoprotein band (Bayburt & Sligar, 2002, 2010; Denisov, Grinkova, Lazarides, & Sligar, 2004; Puthenveetil, Nguyen, & Vinogradova, 2017). We carried out experiments in which A<sub>2A</sub>R was reconstituted into deuterated DMPC nanodiscs containing 10% cholesterol. While these samples exhibited a substantially-improved stability under NMR conditions, this advantage was negated by the presence of many intense lipid-derived peaks in the methyl region which significantly complicated spectral analysis. As a result, we decided to use deuterated DDM micelles as our membrane mimetic.
  3. Concentrate eluted protein using 100 kDa MWCO centrifugal concentrators to 1 mL for injection on a Superdex200 or comparable column (e.g., S200 GL Increase 10/300). As with many other membrane proteins, it is important to

- frequently resuspend the sample during concentration and spin at a slow speed (<3500 rcf) to reduce aggregation.
4. Pre-equilibrate Superdex200 column with SEC buffer, inject concentrated protein, and collect elution peak. The peak should elute at a volume commensurate with monomeric receptor with bound micelle.
    - a. **Note:** Sodium has been shown to be an allosteric modulator of A<sub>2A</sub>R and other Class A GPCRs (Gao & Ijzerman, 2000; Liu *et al.*, 2012). We obtained high-quality spectra of A<sub>2A</sub>R in the presence of 150 mM NaCl, although the minimal chemical shift differences observed between agonist- and antagonist-loaded samples suggest that cation choice has a substantial effect in this system. In contrast, purification in 150 mM KCl led to more heterogeneous spectra and less stable samples that were resistant to relaxation experiments and analysis, although there was a clear difference between agonist- and antagonist-bound spectra under these conditions (Clark *et al.*, 2017).
  5. Concentrate eluted protein using 100 kDa MWCO centrifugal concentrators to ~100  $\mu$ L for NMR samples. Frequently resuspend sample and spin at a slow speed (<3500 rcf) to reduce aggregation. Using 3 mm Shigemi tubes (100  $\mu$ L volume) allows us to concentrate samples to a concentration of ~100  $\mu$ M, leading to adequate signal-to-noise ratios for relaxation data.

**Final Note:** After establishing a purification protocol, it is important to verify that the protein sample is functional. In the case of GPCRs, the standard readout is coupling to G-protein. This is achieved by liposome reconstitution and G-protein activation assays (Cerione & Ross, 1991) and/or G-protein complex formation in detergent followed by subsequent gel filtration chromatography steps (Rasmussen, DeVree, Zou, Kruse, Chung, Kobilka *et al.*, 2011). For A<sub>2A</sub>R, we verified functionality with both methods (Clark *et al.*, 2017) and found that the purification yielded a high level of functional receptor that can efficiently couple to the G<sub>s</sub> heterotrimer.

### **Running <sup>13</sup>C/<sup>1</sup>H HMQC and methyl relaxation experiments**

Once the protein sample is generated, the protein can be analyzed by NMR. A <sup>13</sup>C/<sup>1</sup>H HMQC experiment reports on overall folding and stability of the protein sample, as well as the effects of any perturbations, such as addition of binding partners. To be able to interpret the NMR spectra, some or all of the methyl peaks must be assigned to the residues that they arise from. Finally, residue-specific dynamics at various time scales can be measured by specialized NMR experiments. Here, we focus on our application of a methyl relaxation experiment measuring the fast (psec-nsec) sidechain motions (Sun *et al.*, 2011; Tugarinov, Sprangers, & Kay, 2007) and adaption to short-lived membrane protein samples.

## Equipment and Software

High field NMR spectrometer with cryoprobe (e.g. Bruker AVANCE III HD 800 MHz with TCI probe)

3 mm solvent-matched Shigemi NMR tube

NMRpipe software (Delaglio, Grzesiek, Vuister, Zhu, Pfeifer, & Bax, 1995)

Python installation with NumPy and SciPy modules

## **Procedure**

1. Load the sample into an NMR tube (e.g. 3 mm solvent-matched Shigemi tube). Insert tube into spinner and spinner into high field NMR spectrometer with cryogenically-cooled probe.
2. Set temperature to 30 °C and allow to equilibrate for ~5 minutes.
3. Lock and shim the sample, tune the probe head for  $^1\text{H}$  and  $^{13}\text{C}$ , and determine the  $^1\text{H}$  90° pulse width as per usual NMR protocols.
4. Acquire a 1D  $^1\text{H}$  spectrum to look for detectable isoleucine methyl signals (approx. -0.5 to 1.2 ppm) that are not swamped by other signals (buffer, detergent, lipids).
5. Acquire a  $^{13}\text{C}/^1\text{H}$  HMQC (Ollerenshaw, Tugarinov, & Kay, 2003) spectrum centered on the isoleucine methyl region to assess sample quality, stability, and degree of secondary and tertiary structure.

- a. **Note 1:** The optimal spectral parameters may vary a bit from sample to sample. The  $^{13}\text{C}$  and  $^1\text{H}$  carrier frequencies, spectral widths, and number of points collected for our  $A_{2A}R$  samples are shown in **Table 1**. For a new sample, it is recommended to collect a spectrum with increased spectral width in the  $^{13}\text{C}$  dimension (e.g., 40 ppm centered at 20 ppm) to determine where all isoleucine  $\delta^1$  methyl peaks appear and where other signals arising from the sample (e.g., lipid, detergent, buffer, ligand) may fall. This initial spectrum can then be used to design a combination of spectral width and center frequency that avoids unintended aliasing in contaminating peaks over the peaks of interest.
- b. **Note 2:** Wild-type eukaryotic membrane proteins, and especially GPCRs, have limited stability in the micelle-solubilized form required for obtaining well-resolved spectra in solution state NMR studies. As shown in **Fig. 3B**, the intensities of several  $A_{2A}R$  signals dropped significantly at  $\sim 12$  hours. With this in mind, it is highly recommended to collect consecutive somewhat short (1-2 hr) HMQC experiments, process them independently, and look whether peaks are decreasing or changing radically with time. Those experiments that result in similar spectra can then be summed and processed together to increase the

signal-to-noise ratio. NMRpipe (Delaglio *et al.*, 1995) offers simple commands (“addNMR”) for summing NMR data.

6. To assign isoleucine  $\delta$ 1 methyl peaks, generate mutants in which a single isoleucine has been mutated to a different residue type, then express, purify, and collect a  $^{13}\text{C}/^1\text{H}$  HMQC spectrum on each as described above. Ideally, one peak will disappear and that peak can be assigned to the mutated residue.
  - a. **Note:** This method relies on the mutation perturbing as little of the overall protein structure as possible. Thus, it is important to choose residues to mutate (and thus assign) that are biologically interesting but not indispensable for the proper folding of the protein. The identity of the mutant also plays a large role; we have found that in general isoleucine to valine mutations work well; however, in certain cases it may be necessary to test multiple mutants, such as leucine or methionine (**Fig. 4**). This is often the most time- and cost-limiting step of a study. It is recommended to limit assignments to the most interesting or important residues.
7. On the same, or ideally on another freshly-prepared sample, acquire relaxation measurements. Here we focus on the relaxation violated coherence-transfer experiment and its measurements of methyl order parameters ( $S^2_{\text{axis}}$ ) which report on fast (psec-nsec) motions of the side-

chains (Sun *et al.*, 2011; Tugarinov *et al.*, 2007). The experiment and analysis are outlined in **Fig. 5A**. Other experiments for measuring slower ( $\mu\text{sec}$ - $\text{msec}$ ) dynamics exist, but will not be detailed here (Korzhev, Kloiber, Kanelis, Tugarinov, & Kay, 2004).

- a. **Note 1:** Relaxation experiments rely on collecting multiple spectra from experiments which differ solely by the duration of a relaxation delay and fitting spectral peak intensities as a function of this delay to the appropriate equation to extract relaxation parameters which in turn provide insights into dynamics. It is important to collect sufficient data points to generate a good fit (number of delays) and measure a long enough delay for the (usually exponential) function to decay (longest delay). These values vary based on the size of the particle and rotational tumbling time. For the class of experiments described here, a recommended set of delays for  $\sim 100$  kDa particle is 0.8, 2, 3, 4, 6, 8, 10, 14 msec (Sun *et al.*, 2011). However, since a pair of forbidden and allowed experiments needs to be collected for each delay in this class of relaxation experiment, this large number of time delays can lead to long experimental times, which may be problematic for samples with relatively short lifetimes (see below).
- b. **Note 2:** In our experience, the combination of sample concentration and stability of the membrane protein sample significantly limited our

ability to collect a full set of sixteen spectra as described in **Note 1**. As detailed above (**Note 2** on **Step 5** and **Fig. 3B**), our samples had a useful lifetime of 12-14 hours in the spectrometer at 30 °C, limiting the number of relaxation delays that could be tested. To overcome this, we modified the experimental setup to collect allowed experiments (the higher signal-to-noise experiment) at 5 relaxation delays and a forbidden experiment with 5 times the number of scans at 1 relaxation delay. The choice of forbidden relaxation delay is important; an intermediate value (for us, 8 msec) that has good peak intensity for both forbidden and allowed spectra is highly preferable. The pulse sequence was modified accordingly and the experiment was processed using a custom-written NMRpipe (Delaglio *et al.*, 1995) macro to appropriately split the experiments. Both pulse sequence and macro are available upon request. This change requires modifying the analysis and fitting protocol (see **Note 2** on **Step 10** below).

8. This relaxation experiment is implemented as a pseudo-4D experiment with alternating forbidden/allowed experiments and relaxation delays as the third and fourth dimensions, respectively. To process relaxation experiments, either extract the 2D spectrum at each relaxation delay (allowed and forbidden) using the Bruker TopSpin **rser2d** command and process each dataset as an individual 2D spectrum, or use a NMRpipe script that iterates

over the relaxation delays to simultaneously split and process (a sample can be provided on request). NMRpipe (Delaglio *et al.*, 1995) and NMRFX Processor (Norris, Fetler, Marchant, & Johnson, 2016) are both good options to process extracted 2D datasets.

9. To analyze the relaxation data, first extract peak intensities for each peak to be analyzed for each relaxation delay (allowed and forbidden) using an analysis program such as NMRViewJ (Johnson, 2004). To estimate peak intensity errors in NMRViewJ, either select an empty region of the spectrum and measure the standard deviation of signal across it (in the **Analysis** tab in NMRView J), or use the command `nv_dataset noise data_name.nv` in the NMRViewJ console. This noise estimate may be different for each plane or spectrum and will be used for error analysis of the fit.
10. To extract relaxation parameters, take the ratio of forbidden to allowed peak intensities to the following Equation 1 (Sun *et al.*, 2011), where  $N$  is the number of scans for each experiment,  $T$  is the relaxation delay,  $\delta$  ( $< 0$ ) is a term for the coupling between rapidly and slowly decaying single-quantum coherences, and  $\eta$  is a relaxation rate:

$$\left| \frac{I_{forb}}{I_{all}} \right| = \frac{3N_{all}}{4N_{forb}} \frac{\eta \tanh(\sqrt{\eta^2 + \delta^2} T)}{\sqrt{\eta^2 + \delta^2} - \delta \tanh(\sqrt{\eta^2 + \delta^2} T)} \quad (1)$$

The two fit parameters are  $\eta$  and  $\delta$ , which are constrained to be  $> 0$  and  $< 0$ , respectively, within the fit.

- a. **Note 1:** There are multiple approaches to estimate errors in the fit parameters. One is to propagate noise estimates into the intensity ratios, then use those errors to weight the data points during the least-squares minimization and determine the errors of the fit parameters from the covariance matrix. The `optimize.curve_fit` function in the SciPy Python module provides reasonably simple functionality for this process (Oliphant, 2007). Another approach is a **nonparametric Monte Carlo bootstrap** (Efron, 1981), in which for each set of datapoints to fit, many (100-1000) artificial datasets are simulated centered on the measured peak intensities (and within error estimates from the noise) and then fit individually. The mean and standard deviation of the fit parameters provide the  $\eta$  and  $\delta$  values and errors. A comparison of these two methods applied to data from A<sub>2</sub>AR Is shown in **Fig. 5B**; the error estimates in the bootstrap approach are sometimes larger and sometimes smaller than those from the covariance matrix.
- b. **Note 2:** For the modified relaxation experiment with only one forbidden relaxation delay (together with a complement of several allowed relaxation spectra) described above in **Note 2** on **Step 7**, the processing step (**Step 8**) should be modified to correctly split allowed experiments from the one forbidden experiment. A sample NMRpipe

script can be provided upon request. Peak intensity quantification and error estimates (**Step 9**) are unaffected. The analysis (**Step 10**) involves simultaneous fitting of the single forbidden to allowed intensity ratio to Equation 1 and all allowed peak intensities to the following Equation 2 (Sun *et al.*, 2011), where  $A$  is a constant,  $T$  is the relaxation delay, and  $R_{2,H}^S$  and  $R_{2,H}^F$  are the slow and fast relaxation rates of methyl proton single-quantum transitions:

$$I_{all} = A \frac{3}{2} [\exp(-R_{2,H}^S T) + \exp(-R_{2,H}^F T)] \quad (2)$$

Unfortunately, this approach introduces more fit parameters:  $A$ ,  $R_{2,H}^S$  and  $R_{2,H}^F$  (all of which were constrained to be  $> 0$ ). Using data collected on *E. coli* maltose binding protein as a test case, we achieved good correspondence between  $\eta$  values calculated from 5 pairs of allowed and forbidden experiments and from 5 allowed and 1 forbidden experiment (Clark *et al.*, 2017). In the case of A<sub>2A</sub>R, the isoleucine residue in the flexible C-terminal Protein C tag served as a serendipitous control with markedly reduced relaxation rate than the other residues.

11. The value of  $\eta$  can be used to extract the  $S^2_{axis}$  order parameter of the methyl group using the following Equation 3 (Sun *et al.*, 2011), where  $\tau_c$  is the molecular rotational correlation time,  $\mu_0$  is the vacuum permittivity constant,  $\gamma_H$  is the proton gyromagnetic ratio,  $r_{HH}$  is the distance between pairs of methyl

protons (estimated as 1.813 Å), the function  $P_2(x) = 1/2(3x^2 - 1)$ , and  $\theta_{axis,HH}$  is the angle between the methyl axis and a line between two methyl  $^1\text{H}$  nuclei ( $90^\circ$ ) (Tugarinov & Kay, 2006):

$$\eta = \frac{9}{10} \left( \frac{\mu_0}{4\pi} \right)^2 \left[ P_2(\cos \theta_{axis,HH}) \right]^2 \frac{S_{axis}^2 \gamma_H^4 \hbar^2 \tau_c}{r_{HH}^6} \quad (3)$$

The rotational correlation time  $\tau_c$  can be measured using a  $^{15}\text{N}/^1\text{H}$ -TRACT experiment (D. Lee, Hilty, Wider, & Wuthrich, 2006) or  $^{15}\text{N}$   $T_1/T_2$  relaxation analysis (Farrow, Muhandiram, Singer, Pascal, Kay, Gish *et al.*, 1994). However, like all global estimates of tumbling time, these methods may underestimate the rotational correlation time due to contributions from fast internal dynamics.

**Final Note:** Pulse sequence, macros, and scripts all available on request.

### Concluding remarks

As detailed here, we have developed a protocol using  $\alpha$ -ketobutyrate (the keto acid precursor for isoleucine) supplementation of the growth media for *Pichia pastoris*, to prepare Ile  $\delta^1$  methyl-labeled GPCRs (Clark *et al.*, 2017; Clark *et al.*, 2018; Clark *et al.*, 2015). This protocol is analogous to methods previously developed for expression and labeling in *E. coli* (Gardner & Kay, 1997); however, it allows access to integral membrane proteins such as GPCRs that cannot be efficiently expressed in a prokaryotic host. The samples we have generated are

suitable for a range of solution NMR measurements, including relaxation-based determination of ligand-dependent changes in receptor dynamics. These data complement other NMR-based measurements of receptor structure, dynamics, and interactions using probes at other  $^{13}\text{C}$  methyl groups (*Kofuku et al., 2014; Kofuku et al., 2018; Solt, Bostock, Shrestha, Kumar, Warne, Tate et al., 2017*), exogenously added  $^{19}\text{F}$  probes (*Ye, Neale, Slijoka, Lyda, Pichugin, Tsuchimura et al., 2018; Ye, Van Eps, Zimmer, Ernst, & Prosser, 2016*), or  $^{15}\text{N}/^1\text{H}$  pairs in receptors stabilized by point mutations or insertion into covalently-circularized nanodiscs (*Isogai et al., 2016; Nasr, Baptista, Strauss, Sun, Grigoriu, Huser et al., 2017*).

Taken together, these studies show the rich diversity of GPCR conformations and dynamics, taking advantage of the power of solution NMR to get site-specific information from receptors in equilibrium and complementing structural studies from X-ray diffraction and cryo-EM. In addition, we believe that, properly modified, these approaches can be applied to study the dynamics and interactions of other eukaryotic membrane proteins as well. We anticipate that further technological developments, including those in sample preparation [e.g., additional labeling strategies (*Robson, Takeuchi, Boeszoermenyi, Coote, Dubey, Hyberts et al., 2018*), ways to improve sample stability or concentration via improved membrane-mimetics such as smaller, more “NMR-friendly” nanodiscs (*Chien, Helfinger, Bostock, Solt, Tan, & Nietlispach, 2017; Hagn, Etzkorn, Raschle, & Wagner, 2013*), NMR data acquisition (e.g., ultrahigh field NMR systems (*Quinn, Wang, & Polenova, 2018*))],

improved  $^{13}\text{C}/^{15}\text{N}$  direct detection methods (Takeuchi, Heffron, Sun, Frueh, & Wagner, 2010), and analysis (e.g., improved methods for chemical shift assignment through PRE or NOESY experiments (Lescanne, Skinner, Blok, Timmer, Cerofolini, Fragai *et al.*, 2017)), will only further improve such prospects in the future.

## References

Almén, M. S., Nordström, K. J., Fredriksson, R., & Schiöth, H. B. (2009). Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*, 7, 50. doi:10.1186/1741-7007-7-50

André, N., Cherouati, N., Prual, C., Steffan, T., Zeder-Lutz, G., Magnin, T., . . . Reinhart, C. (2006). Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Science*, 15(5), 1115-1126. doi:10.1110/ps.062098206

Aw, R., & Polizzi, K. M. (2013). Can too many copies spoil the broth? *Microbial Cell Factories*, 12, 128. doi:10.1186/1475-2859-12-128

Bayburt, T. H., & Sligar, S. G. (2002). Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10), 6725-6730. doi:10.1073/pnas.062565599

Bayburt, T. H., & Sligar, S. G. (2010). Membrane protein assembly into Nanodiscs. *FEBS Letters*, 584(9), 1721-1727. doi:10.1016/j.febslet.2009.10.024

Byrne, B. (2015). *Pichia pastoris* as an expression host for membrane protein structural biology. *Current Opinion in Structural Biology*, 32, 9-17. doi:10.1016/j.sbi.2015.01.005

Cereghino, G. P., Cereghino, J. L., Ilgen, C., & Cregg, J. M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. *Current Opinion in Biotechnology*, 13(4), 329-332.

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*, 24(1), 45-66.

Cerione, R. A., & Ross, E. M. (1991). Reconstitution of receptors and G proteins in phospholipid vesicles. *Methods in Enzymology*, 195, 329-342.

Chien, C. H., Helfinger, L. R., Bostock, M. J., Solt, A., Tan, Y. L., & Nietlispach, D. (2017). An adaptable phospholipid membrane mimetic system for solution NMR studies of membrane proteins. *Journal of the American Chemical Society*, 139(42), 14829-14832. doi:10.1021/jacs.7b06730

- Clark, L., Dikiy, I., Chapman, K., Rodstrom, K. E., Aramini, J., LeVine, M. V., . . . Rosenbaum, D. M. (2017). Ligand modulation of sidechain dynamics in a wild-type human GPCR. *Elife*, 6. doi:10.7554/eLife.28505
- Clark, L. (2018). On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies. *Journal of Biomolecular NMR*, 71(4), 203–211. <http://doi.org/10.1007/s10858-018-0204-3>
- Clark, L., Zahm, J. A., Ali, R., Kukula, M., Bian, L., Patrie, S. M., . . . Rosenbaum, D. M. (2015). Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *Journal of Biomolecular NMR*, 62(3), 239-245. doi:10.1007/s10858-015-9939-2
- Cos, O., Ramon, R., Montesinos, J. L., & Valero, F. (2006). Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microbial Cell Factories*, 5, 17. doi:10.1186/1475-2859-5-17
- Cuniasse, P., Tavares, P., Orlova, E. V., & Zinn-Justin, S. (2017). Structures of biomolecular complexes by combination of NMR and cryoEM methods. *Current Opinion in Structural Biology*, 43, 104-113. doi:10.1016/j.sbi.2016.12.008
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3), 277-293. doi:10.1007/BF00197809
- Denisov, I. G., Grinkova, Y. V., Lazarides, A. A., & Sligar, S. G. (2004). Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *Journal of the American Chemical Society*, 126(11), 3477-3487. doi:10.1021/ja0393574
- Eddy, M. T., Lee, M. Y., Gao, Z. G., White, K. L., Didenko, T., Horst, R., . . . Wüthrich, K. (2018). Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. *Cell*, 172(1-2), 68-80 e12. doi:10.1016/j.cell.2017.12.004
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68(3), 589-599. doi:10.1093/biomet/68.3.589
- Emami, S., Fan, Y., Munro, R., Ladizhansky, V., & Brown, L. S. (2013). Yeast-expressed human membrane protein aquaporin-1 yields excellent resolution of solid-

state MAS NMR spectra. *Journal of Biomolecular NMR*, 55(2), 147-155.  
doi:10.1007/s10858-013-9710-5

Fan, Y., Emami, S., Munro, R., Ladizhansky, V., & Brown, L. S. (2015). Isotope labeling of eukaryotic membrane proteins in yeast for solid-state NMR. *Methods in Enzymology*, 565, 193-212. doi:10.1016/bs.mie.2015.05.010

Fan, Y., Shi, L., Ladizhansky, V., & Brown, L. S. (2011). Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment. *Journal of Biomolecular NMR*, 49(2), 151-161. doi:10.1007/s10858-011-9473-9

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., . . . Kay, L. E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src Homology 2 domain studied by <sup>15</sup>N NMR relaxation. *Biochemistry*, 33(19), 5984-6003. doi:10.1021/bi00185a040

Gao, Z. G., & Ijzerman, A. P. (2000). Allosteric modulation of A(2A) adenosine receptors by amiloride analogues and sodium ions. *Biochemical Pharmacology*, 60(5), 669-676.

Gardner, K. H., & Kay, L. E. (1997). Production and incorporation of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H (<sup>1</sup>H- $\delta$ 1 methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society*, 119(32), 7599-7600. doi:10.1021/ja9706514

Hagn, F., Etzkorn, M., Raschle, T., & Wagner, G. (2013). Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *Journal of the American Chemical Society*, 135(5), 1919-1925. doi:10.1021/ja310901f

Heyland, J., Fu, J., Blank, L. M., & Schmid, A. (2011). Carbon metabolism limits recombinant protein production in *Pichia pastoris*. *Biotechnology and Bioengineering*, 108(8), 1942-1953. doi:10.1002/bit.23114

Hino, T., Arakawa, T., Iwanari, H., Yurugi-Kobayashi, T., Ikeda-Suno, C., Nakada-Nakura, Y., . . . Murata, T. (2012). G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature*, 482(7384), 237-240. doi:10.1038/nature10750

Isogai, S., Deupi, X., Opitz, C., Heydenreich, F. M., Tsai, C. J., Brueckner, F., . . . Grzesiek, S. (2016). Backbone NMR reveals allosteric signal transduction networks in the beta1-adrenergic receptor. *Nature*, 530(7589), 237-241. doi:10.1038/nature16577

Jaakola, V. P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y., Lane, J. R., . . . Stevens, R. C. (2008). The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322(5905), 1211-1217. doi:10.1126/science.1164772

Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods in Molecular Biology*, 278, 313-352. doi:10.1385/1-59259-809-9:313

Kasinath, V., Sharp, K. A., & Wand, A. J. (2013). Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *Journal of the American Chemical Society*, 135(40), 15092-15100. doi:10.1021/ja405200u

Kofuku, Y., Ueda, T., Okude, J., Shiraishi, Y., Kondo, K., Mizumura, T., et al. (2014). Functional Dynamics of Deuterated  $\beta$  2-Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angewandte Chemie International Edition*, 53(49), 13376–13379. <http://doi.org/10.1002/anie.201406603>

Kofuku, Y., Yokomizo, T., Imai, S., Shiraishi, Y., Natsume, M., Itoh, H., . . . Shimada, I. (2018). Deuteration and selective labeling of alanine methyl groups of beta2-adrenergic receptor expressed in a baculovirus-insect cell expression system. *Journal of Biomolecular NMR*. doi:10.1007/s10858-018-0174-5

Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V., & Kay, L. E. (2004). Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *Journal of the American Chemical Society*, 126(12), 3964-3973. doi:10.1021/ja039587i

Lee, D., Hilty, C., Wider, G., & Wuthrich, K. (2006). Effective rotational correlation times of proteins from NMR relaxation interference. *Journal of Magnetic Resonance*, 178(1), 72-76. doi:10.1016/j.jmr.2005.08.014

Lee, J. Y., Kinch, L. N., Borek, D. M., Wang, J., Wang, J., Urbatsch, I. L., . . . Rosenbaum, D. M. (2016). Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature*, 533(7604), 561-564. doi:10.1038/nature17666

Lescanne, M., Skinner, S. P., Blok, A., Timmer, M., Cerofolini, L., Fragai, M., . . . Ubbink, M. (2017). Methyl group assignment using pseudocontact shifts with PARAssign. *Journal of Biomolecular NMR*, 69(4), 183-195. doi:10.1007/s10858-017-0136-3

- Liang, B., & Tamm, L. K. (2016). NMR as a tool to investigate the structure, dynamics and function of membrane proteins. *Nature Structural and Molecular Biology*, 23(6), 468-474. doi:10.1038/nsmb.3226
- Liang, B., & Tamm, L. K. (2018). Solution NMR of SNAREs, complexin and alpha-synuclein in association with membrane-mimetics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 105, 41-53. doi:10.1016/j.pnmrs.2018.02.001
- Lin-Cereghino, G. P., Stark, C. M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., . . . Lin-Cereghino, J. (2013). The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*. *Gene*, 519(2), 311-317. doi:10.1016/j.gene.2013.01.062
- Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., . . . Stevens, R. C. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*, 337(6091), 232-236. doi:10.1126/science.1219218
- Moore, P. B. (1979). The preparation of deuterated ribosomal materials for neutron scattering. *Methods in Enzymology*, 59, 639-655.
- Morgan, W. D., Kragt, A., & Feeney, J. (2000). Expression of deuterium-isotope-labelled protein in the yeast *pichia pastoris* for NMR studies. *Journal of Biomolecular NMR*, 17(4), 337-347. doi:10.1023/A:1008313530207
- Nasr, M. L., Baptista, D., Strauss, M., Sun, Z. J., Grigoriu, S., Huser, S., . . . Wagner, G. (2017). Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nature Methods*, 14(1), 49-52. doi:10.1038/nmeth.4079
- Norris, M., Fetler, B., Marchant, J., & Johnson, B. A. (2016). NMRFx Processor: a cross-platform NMR data processing program. *Journal of Biomolecular NMR*, 65(3-4), 205-216. doi:10.1007/s10858-016-0049-6
- Oliphant, T. E. (2007). Python for Scientific Computing. *Computing in Science & Engineering*, 9(3), 10-20. doi:10.1109/mcse.2007.58
- Ollerenshaw, J. E., Tugarinov, V., & Kay, L. E. (2003). Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry*, 41(10), 843-852. doi:10.1002/mrc.1256
- Opella, S. J., & Marassi, F. M. (2017). Applications of NMR to membrane proteins. *Archives of Biochemistry and Biophysics*, 628, 92-101. doi:10.1016/j.abb.2017.05.011

Opitz, C., Isogai, S., & Grzesiek, S. (2015). An economic approach to efficient isotope labeling in insect cells using homemade <sup>15</sup>N-, <sup>13</sup>C- and <sup>2</sup>H-labeled yeast extracts. *Journal of Biomolecular NMR*, 62(3), 373-385. doi:10.1007/s10858-015-9954-3

Palmer, A. G. (2012). NMR Spectroscopy: NMR Relaxation Methods. In H. J. Dyson (Ed.), *Biophysical Techniques for Structural Characterization of Macromolecules* (pp. 216-244). Oxford: Academic Press.

Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., . . . Siegal, G. (2008). Perspectives on NMR in drug discovery: a technique comes of age. *Nature Reviews Drug Discovery*, 7(9), 738-745. doi:10.1038/nrd2606

Pickford, A. R., & O'Leary, J. M. (2004). Isotopic labeling of recombinant proteins from the methylotrophic yeast *Pichia pastoris*. *Methods in Molecular Biology*, 278, 17-33. doi:10.1385/1-59259-809-9:017

Puthenveetil, R., Nguyen, K., & Vinogradova, O. (2017). Nanodiscs and solution NMR: preparation, application and challenges. *Nanotechnology Reviews*, 6(1), 111-126. doi:10.1515/ntrev-2016-0076

Quinn, C. M., Wang, M., & Polenova, T. (2018). NMR of macromolecular assemblies and machines at 1 GHz and beyond: New transformative opportunities for molecular structural biology. *Methods in Molecular Biology*, 1688, 1-35. doi:10.1007/978-1-4939-7386-6\_1

Rakestraw, J. A., Sazinsky, S. L., Piatasi, A., Antipov, E., & Wittrup, K. D. (2009). Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 103(6), 1192-1201. doi:10.1002/bit.22338

Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., . . . Kobilka, B. K. (2011). Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature*, 477(7366), 549-555. doi:10.1038/nature10361

Robson, S. A., Takeuchi, K., Boeszoermyeni, A., Coote, P. W., Dubey, A., Hyberts, S., . . . Arthanari, H. (2018). Mixed pyruvate labeling enables backbone resonance assignment of large proteins using a single experiment. *Nature Communications*, 9(1), 356. doi:10.1038/s41467-017-02767-8

- Rosenzweig, R., & Kay, L. E. (2014). Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annual Review of Biochemistry*, 83, 291-315. doi:10.1146/annurev-biochem-060713-035829
- Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., . . . Iwata, S. (2011). Structure of the human histamine H1 receptor complex with doxepin. *Nature*, 475(7354), 65-70. doi:10.1038/nature10236
- Solà, A., Maaheimo, H., Ylonen, K., Ferrer, P., & Szyperski, T. (2004). Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *European Journal of Biochemistry*, 271(12), 2462-2470. doi:10.1111/j.1432-1033.2004.04176.x
- Solt, A. S., Bostock, M. J., Shrestha, B., Kumar, P., Warne, T., Tate, C. G., & Nietlispach, D. (2017). Insight into partial agonism by observing multiple equilibria for ligand-bound and Gs-mimetic nanobody-bound beta1-adrenergic receptor. *Nature Communications*, 8(1), 1795. doi:10.1038/s41467-017-02008-y
- Sun, H., Kay, L. E., & Tugarinov, V. (2011). An optimized relaxation-based coherence transfer NMR experiment for the measurement of side-chain order in methyl-protonated, highly deuterated proteins. *The Journal of Physical Chemistry B*, 115(49), 14878-14884. doi:10.1021/jp209049k
- Suzuki, R., Sakakura, M., Mori, M., Fujii, M., Akashi, S., & Takahashi, H. (2018). Methyl-selective isotope labeling using alpha-ketoisovalerate for the yeast *Pichia pastoris* recombinant protein expression system. *Journal of Biomolecular NMR*. doi:10.1007/s10858-018-0192-3
- Takeuchi, K., Heffron, G., Sun, Z. Y., Frueh, D. P., & Wagner, G. (2010). Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *Journal of Biomolecular NMR*, 47(4), 271-282. doi:10.1007/s10858-010-9430-z
- Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E., & Kay, L. E. (2003). Cross-correlated relaxation enhanced <sup>1</sup>H-<sup>13</sup>C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society*, 125(34), 10420-10428. doi:10.1021/ja030153x
- Tugarinov, V., & Kay, L. E. (2006). A <sup>2</sup>H NMR relaxation experiment for the measurement of the time scale of methyl side-chain dynamics in large proteins. *Journal of the American Chemical Society*, 128(38), 12484-12489. doi:10.1021/ja063071s

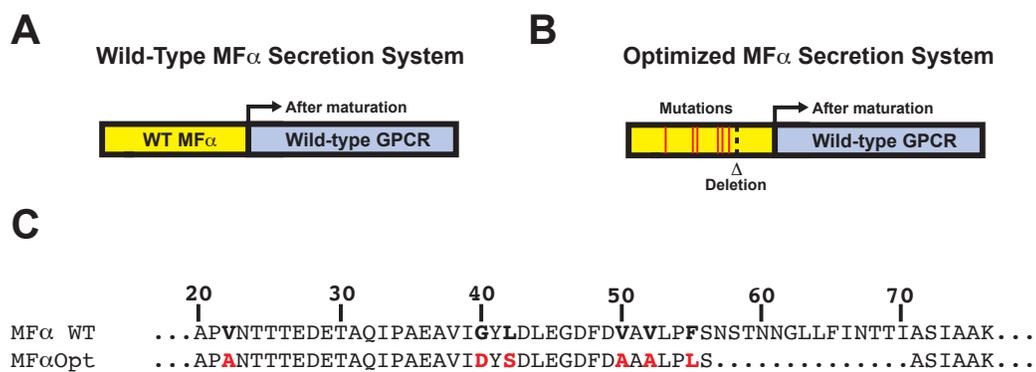
- Tugarinov, V., Sprangers, R., & Kay, L. E. (2007). Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. *Journal of the American Chemical Society*, 129(6), 1743-1750. doi:10.1021/ja067827z
- Ulmschneider, M. B., & Sansom, M. S. (2001). Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta*, 1512(1), 1-14. doi:10.1016/S0005-2736(01)00299-1
- Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., . . . Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research*, 36(Database issue), D402-408. doi:10.1093/nar/gkm957
- Wand, A. J., Moorman, V. R., & Harpole, K. W. (2013). A surprising role for conformational entropy in protein function. *Topics in Current Chemistry*, 337, 69-94. doi:10.1007/128\_2012\_418
- Weiss, H. M., & Grisshammer, R. (2002). Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*. *European Journal of Biochemistry*, 269(1), 82-92.
- Whittaker, J. W. (2007). Selective isotopic labeling of recombinant proteins using amino acid auxotroph strains. *Methods in Molecular Biology*, 389, 175-188. doi:10.1007/978-1-59745-456-8\_13
- Wood, M. J., & Komives, E. A. (1999). Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. *Journal of Biomolecular NMR*, 13(2), 149-159.
- Ye, L., Neale, C., Sljoka, A., Lyda, B., Pichugin, D., Tsuchimura, N., . . . Prosser, R. S. (2018). Mechanistic insights into allosteric regulation of the A2A adenosine G protein-coupled receptor by physiological cations. *Nature Communications*, 9(1), 1372. doi:10.1038/s41467-018-03314-9
- Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P., & Prosser, R. S. (2016). Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature*, 533(7602), 265-268. doi:10.1038/nature17668
- Yurugi-Kobayashi, T., Asada, H., Shiroishi, M., Shimamura, T., Funamoto, S., Katsuta, N., . . . Kobayashi, T. (2009). Comparison of functional non-glycosylated GPCRs expression in *Pichia pastoris*. *Biochemical and Biophysical Research Communications*, 380(2), 271-276. doi:10.1016/j.bbrc.2009.01.053

Zhang, M., Yu, X. W., Xu, Y., Jouhten, P., Swapna, G. V. T., Glaser, R. W., . . . Szyperski, T. (2017). (13) C metabolic flux profiling of *Pichia pastoris* grown in aerobic batch cultures on glucose revealed high relative anabolic use of TCA cycle and limited incorporation of provided precursors of branched-chain amino acids. *The FEBS Journal*, 284(18), 3100-3113. doi:10.1111/febs.14180

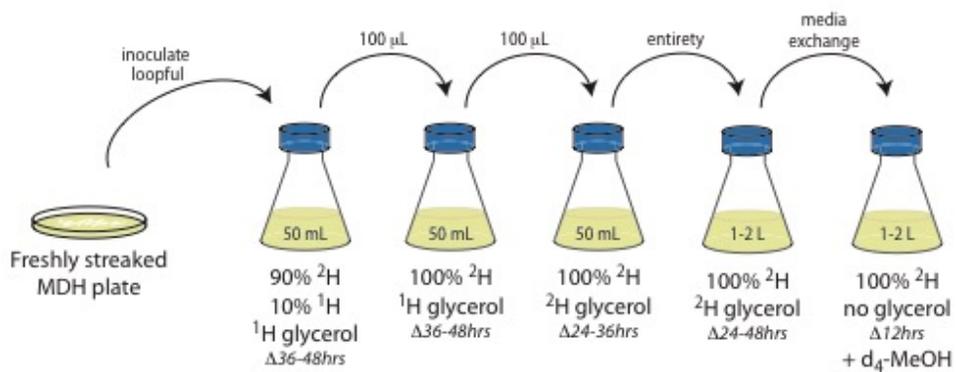
Zhou, H. X., & Cross, T. A. (2013). Influences of membrane mimetic environments on membrane protein structures. *Annual Review of Biophysics*, 42, 361-392. doi:10.1146/annurev-biophys-083012-130326

**Table 1: Typical spectral parameters for  $^{13}\text{C}/^1\text{H}$  HMQC spectra of U- $^{2}\text{H}$  ( $^{13}\text{C}$ ,  $^1\text{H}$  Ile  $\delta 1$   $\text{CH}_3$ ) labeled GPCR samples.**

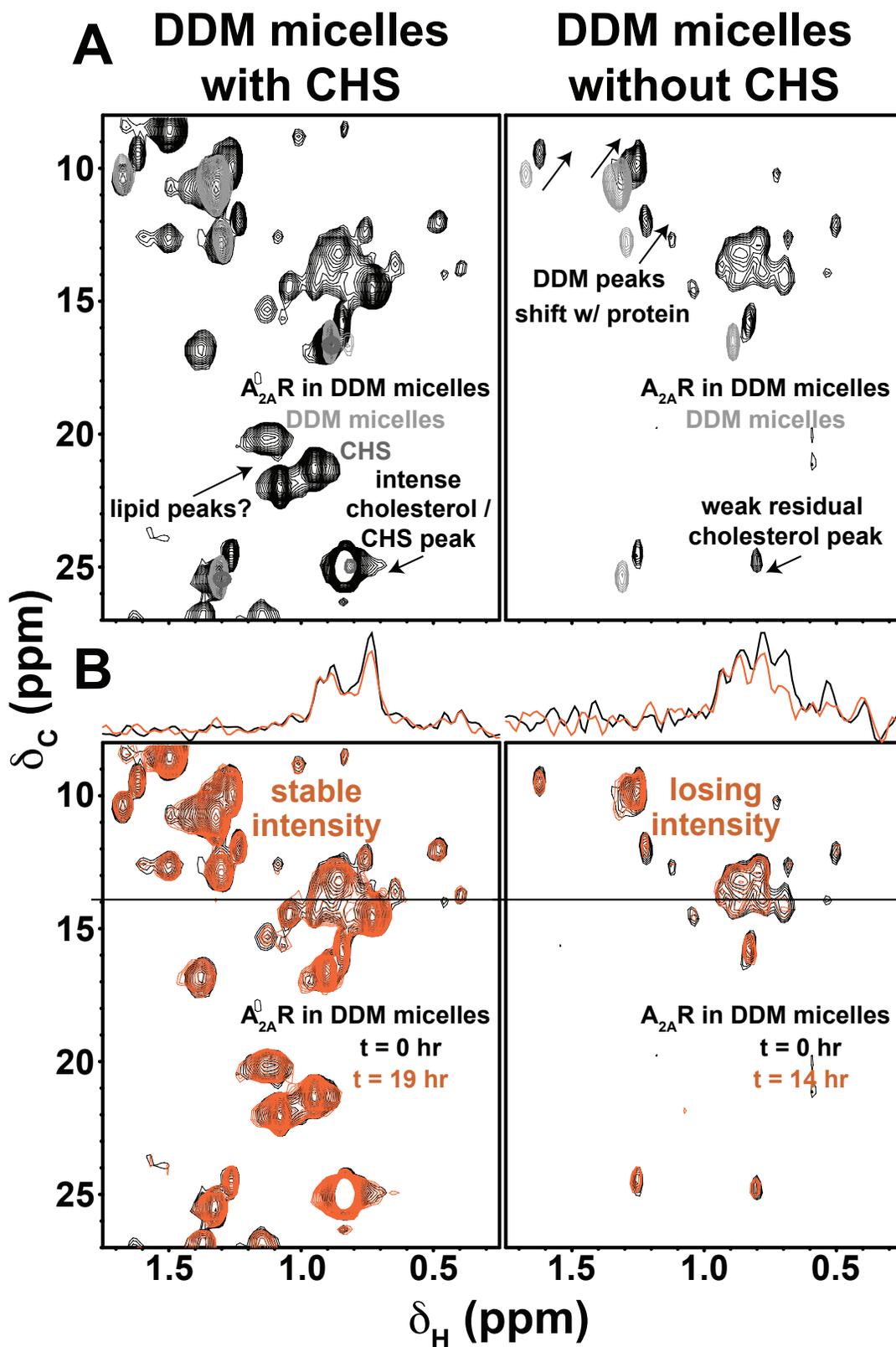
| Nucleus         | Center Frequency (ppm) | Spectral Width (ppm) | Number of Points | Acquisition Time (msec) |
|-----------------|------------------------|----------------------|------------------|-------------------------|
| $^1\text{H}$    | 4.7                    | 13.35                | 1024             | 7                       |
| $^{13}\text{C}$ | 14                     | 22                   | 64               | 48                      |



**Figure 1: Optimization of MF $\alpha$  signal sequence.** A) and B) General schematics of WT and optimized MF $\alpha$  signal sequences prior to start of WT GPCR gene of interest. As shown in alignments in C), six mutations (V22A, G40D, L42S, V50A, V52A, and F55L) plus a deletion of residues 57-70 generate an optimized version of the MF $\alpha$  signal sequence (“MF $\alpha$ Opt”) that was used for efficient A<sub>2A</sub>R expression at the plasma membrane.

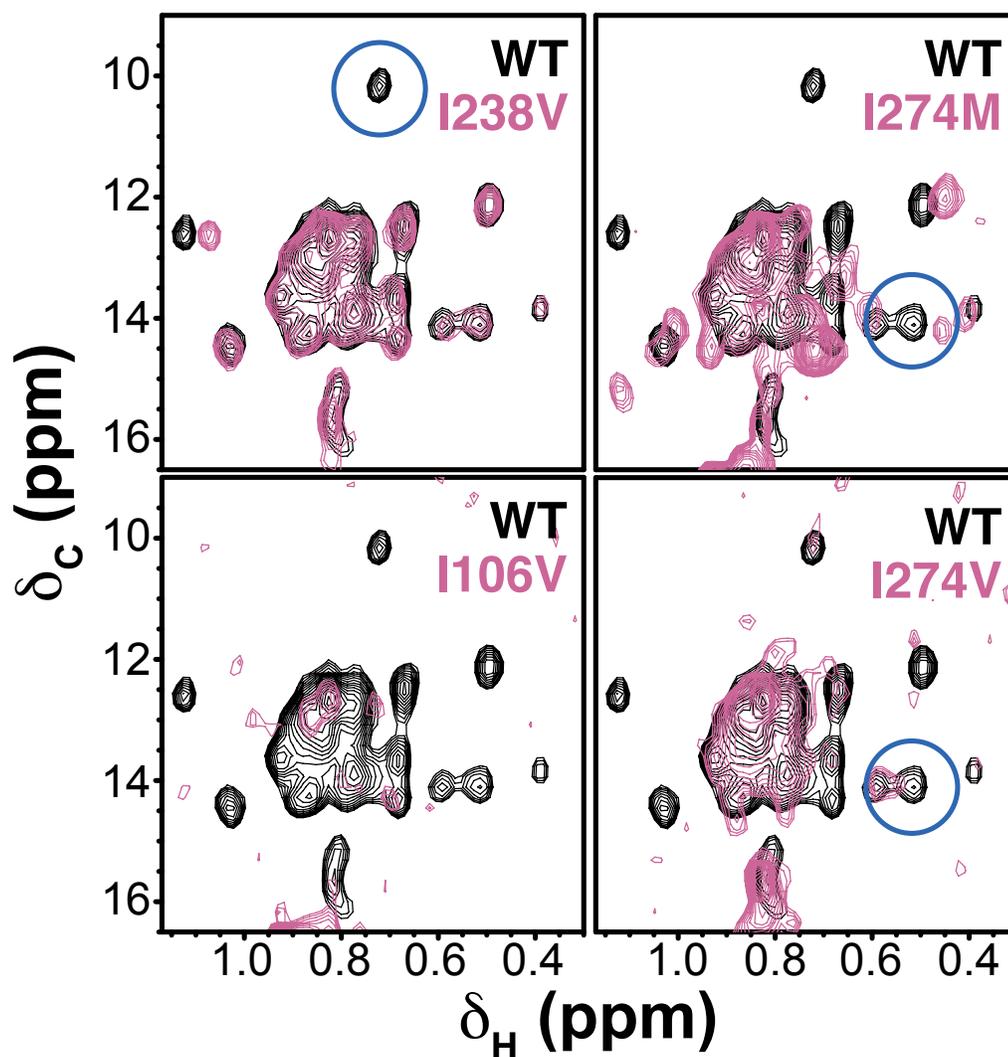


**Figure 2: Adaptation of *Pichia pastoris* to deuterated media.** Summary of schemes used to adapt *Pichia* cultures to increasingly deuterated media to ensure efficient deuteration levels and protein expression. See section on **Expression and isotopic labeling in *Pichia*** for more details.

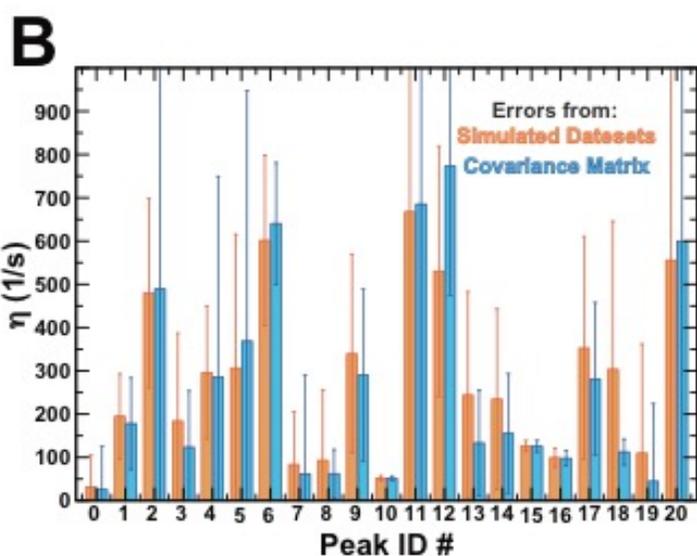
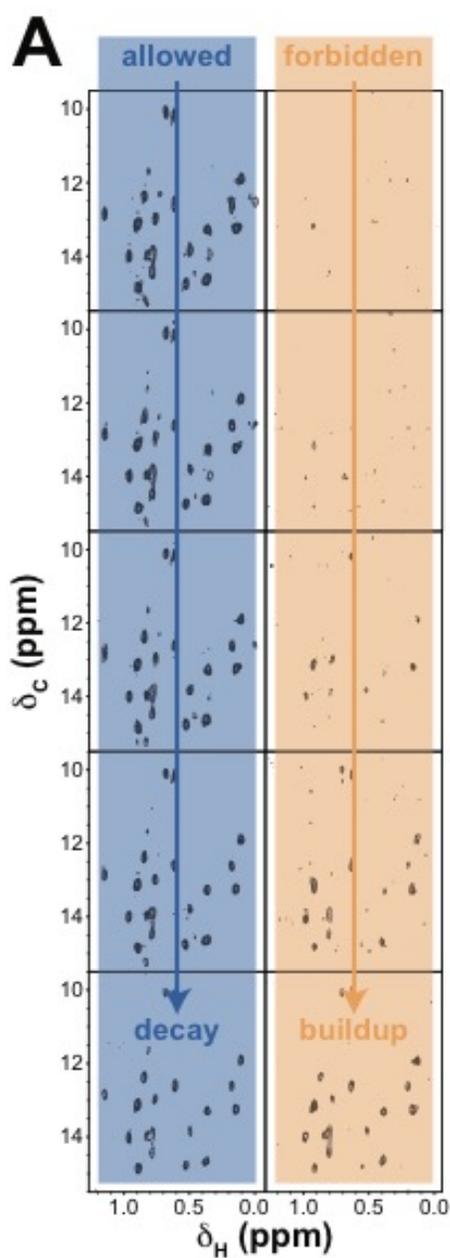


**Figure 3: Receptors reconstituted with CHS are more stable but have crowded**

**spectra.** A)  $^{13}\text{C}/^1\text{H}$  HMQC spectra of  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles with CHS (left) and without added CHS (right). Reconstitution with CHS resulted in spectra with intense extraneous signals, likely arising from lipids, as well as a very intense cholesteryl methyl peak (dark gray). In DDM micelles without added CHS, the dispersion of protein-derived peaks is comparable, but there are fewer extraneous peaks in the region. A weak signal at the CHS/cholesterol methyl chemical shift suggests that some sterol was copurified with the protein. B)  $^{13}\text{C}/^1\text{H}$  HMQC spectra of  $\text{A}_{2\text{A}}\text{R}$  in DDM micelles with CHS (left) and without added CHS (right) collected after different lengths of time at 30 °C. Several protein signals in the DDM micelles without CHS drop significantly after 14 hours (dark orange).



**Figure 4: Assignment of selected A<sub>2A</sub>R Ile- $\delta$ 1 methyl peaks by site-directed mutagenesis.** <sup>13</sup>C/<sup>1</sup>H HMQC spectra of wild-type A<sub>2A</sub>R (black) are shown overlaid with four mutants (purple). In some cases, such as I238, only one peak disappeared and was assigned (blue circle). In others, such as I274, manifold spectral changes were present, and multiple mutants were needed, along with solvent PRE data (Clark et al., 2017) to make the assignment (blue circle). In other cases, such as I106, the mutation completely abrogated protein folding or stability.

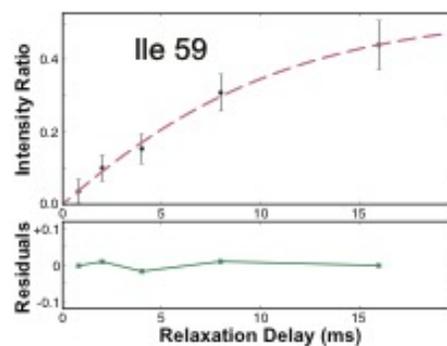


- Extract intensities

- Take ratios

- Fit  $I_{\text{forbidden}}/I_{\text{allowed}}$  to Equation 1

$$\left| \frac{I_{\text{forb}}}{I_{\text{all}}} \right| = \frac{3N_{\text{all}}}{4N_{\text{forb}}} \frac{\eta \tanh(\sqrt{\eta^2 + \delta^2} \tau)}{\sqrt{\eta^2 + \delta^2} - \delta \tanh(\sqrt{\eta^2 + \delta^2} \tau)}$$



**Figure 5: Schematic of fast motions relaxation experiment.** A) Sample  $^{13}\text{C}/^1\text{H}$  planes at the indicated relaxation delays for *E. coli* MBP are shown to illustrate the decay of allowed signal intensity and buildup of forbidden signal intensity. To derive relaxation rate  $\eta$ , peak intensities are extracted and the forbidden:allowed ratio is fit to Equation 1 as a function of relaxation delay. B) Plot of extracted  $\eta$  relaxation rates and their errors using either a Monte Carlo bootstrapping approach with 1000 selected datasets (red/yellow) or covariance matrix of the fit (blue/cyan) for A<sub>2</sub>R. Ile- $\delta$ 1 methyl peaks are listed by an arbitrary peak ID number.

## CHAPTER FIVE

### On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies

#### Abstract

NMR studies of human integral membrane proteins provide unique opportunities to probe structure and dynamics at specific locations and on multiple timescales, often with significant implications for disease mechanism and drug development. Since membrane proteins such as G protein-coupled receptors (GPCRs) are highly dynamic and regulated by ligands or other perturbations, NMR methods are potentially well suited to answer basic functional questions (such as addressing the biophysical basis of ligand efficacy) as well as guiding applications (such as novel ligand design). However, such studies on eukaryotic membrane proteins have often been limited by the inability to incorporate optimal isotopic labels for NMR methods developed for large protein/lipid complexes, including methyl TROSY. We review the different expression systems for production of isotopically labeled membrane proteins and highlight the use of the yeast *Pichia pastoris* to achieve perdeuteration and  $^{13}\text{C}$  methyl probe incorporation within isoleucine sidechains. We further illustrate the use of this method for labeling of several biomedically significant GPCRs.

## Introduction

Membrane proteins constitute a third of the human proteome and play key roles in diverse aspects of physiology and disease (Almén *et al.* 2009). Because they require a lipid environment or membrane mimetic to maintain native-like structure and function, integral membrane proteins are particularly challenging targets for structural and biophysical studies. Structures solved by crystallography and cryo-EM have provided molecular details of how these proteins bind to natural and artificial ligands and respond to changes in their environment. Despite these advances, most membrane proteins have remained difficult to study by NMR spectroscopy due to their high molecular weight and difficulties of producing samples labeled with specific  $^2\text{H}$  and  $^{13}\text{C}$  patterns that facilitate such studies. While some membrane proteins can be natively expressed or refolded from prokaryotic systems where such isotopic labeling is well-established (Casiraghi *et al.* 2016; Hiller *et al.* 2008; Park *et al.* 2012), most eukaryotic membrane proteins require a eukaryotic expression host for proper translation, folding, and stability.

One such expression host is the methylotrophic yeast *Pichia pastoris* ('*Pichia*'), which has been widely used to express a wide variety of proteins, including human membrane proteins (Byrne 2015). *Pichia* contains the machinery required for important post-translational modifications such as disulfide formation and glycosylation, as well as appropriate chaperones to facilitate protein folding (Cereghino and Cregg 2000). Further, *Pichia*-based overexpression is efficient

enough to produce a range of human integral membrane proteins at levels sufficient for structural studies by X-ray crystallography and cryo-EM. Some examples are H<sub>1</sub> histamine receptor (Shimamura *et al.* 2011), A<sub>2A</sub> adenosine receptor (Hino *et al.* 2012), ABC G5G8 sterol transporter (Lee *et al.* 2016), GIRK2 potassium channel (Whorton and MacKinnon 2011), Orai store-operated calcium channel (Hou *et al.* 2012), P-glycoprotein (Jin *et al.* 2012), TPC potassium channel (Guo *et al.* 2016), and TAP transporter (Oldham *et al.* 2016).

*Pichia* has also proven valuable for production of isotopically labeled samples for NMR spectroscopy. The first successful demonstration of uniform <sup>13</sup>C/<sup>15</sup>N labeling in *Pichia* was carried out for the tick anticoagulant peptide (Laroche *et al.* 1994), and subsequently for a number of aqueous-soluble, low-MW targets. Membrane proteins are often inherently larger on their own and even further when considering the micelle or membrane mimetic required for purification in a functional state, with accompanying slow rotational tumbling and broad NMR signals. Coupled with the poor chemical shift dispersion and substantial peak overlap associated with NMR spectra of many multiple transmembrane (TM) helix proteins (Kim *et al.* 2009), these considerations require sophisticated isotopic labeling patterns involving deuteration and site-specific isotope incorporation (e.g. methyl labeling) to facilitate NMR studies of these proteins.

In this brief communication, we summarize recent applications of isotopic labeling of membrane proteins in several types of eukaryotic expression hosts, including insect cells, mammalian cells, and yeast. We particularly focus on the yeast *Pichia pastoris*, and detail the general strengths and weaknesses of its use for membrane protein expression for NMR studies. Additionally, we include further background on our recently-published methyl labeling approach in *Pichia* (Clark *et al.* 2015; Clark *et al.* 2017) and include new examples of spectra collected on labeled wild-type human GPCRs produced from this system.

## Materials and Methods

### *Construct design*

cDNAs for human CB1 (residues 74-430) and human OX<sub>2</sub>R (residues 1-386) were cloned into the MF $\alpha$ Opt vector described previously (Clark *et al.* 2017) containing a C-terminal 10x-His tag for affinity purification. Plasmids were linearized with *PmeI* (NEB) and 5-10  $\mu$ g of linearized DNA was inserted via electroporation (2.00 kV pulse) into freshly-prepared competent KM71H (Mut<sup>s</sup> phenotype) *Pichia pastoris* cells (Invitrogen). Immediately following electroporation, 1 mL of ice-cold 1M sorbitol was added to the cuvette, and cells were transferred to a sterile culture tube and incubated at 28 °C for 60 minutes. Following incubation, 1 mL of YPD (1% yeast extract, 2% peptone, 2% dextrose) was added to each tube and cultures were

shaken at 200 rpm for 4 hours at 28 °C to recover. Cells were plated on YPDS (1% yeast extract, 2% peptone, 2% dextrose, 2% agar, 1M sorbitol) agar plates containing increasing amounts of Zeocin (Invitrogen; typically 100–1000  $\mu\text{g}/\text{mL}$ ) and individual clones were screened for expression through western blots using an anti-His<sub>6</sub> antibody (Pierce). Of note, clones from the KM71H cell line gave consistently higher expression levels for all three receptors compared to GS115 and SMD1163 cell lines (data not shown). The best expressing clones were stored as 20% glycerol stocks at -80 °C.

#### *Expression of deuterated cultures*

Protein expression and culture adaptation was carried out as described for the A<sub>2A</sub> receptor (Clark *et al.* 2017) with a few modifications. Briefly, a loopful of freshly streaked cells were inoculated into a sterile 50 mL volume of BMG (1% glycerol, 100 mM potassium phosphate pH 6.0, 1.34% YNB (yeast nitrogen base), 0.004% histidine, 4 x 10<sup>-5</sup>% biotin) containing 90% D<sub>2</sub>O/10% H<sub>2</sub>O (Cambridge Isotope Laboratories). The 50 mL culture was shaken at 28° C until an OD<sub>600</sub> of ~10 was reached. At this point, 200  $\mu\text{L}$  of the 90%/10% culture was inoculated into 50 mL of BMG made with 100% D<sub>2</sub>O. The 50 mL culture was shaken at 250 rpm at 28° C until an OD<sub>600</sub> of ~10 was reached, and then 200  $\mu\text{L}$  of that culture was used to inoculate 50 mL of BMG made with 100% D<sub>2</sub>O and d<sub>8</sub>-glycerol (Cambridge Isotope Laboratories). The culture was shaken until reaching an OD<sub>600</sub> of ~10, and then was

used in its entirety to inoculate large scale cultures (typically 1-2 L) of identical media. The large-scale cultures were shaken until saturation ( $OD_{600}$  of 20-30), and then spun down in sterile bottles and the media was removed. The cells were resuspended in BMM (100 mM potassium phosphate pH 6.0, 1.34% YNB, 0.004% histidine,  $4 \times 10^{-5}$ % biotin) made in 100%  $D_2O$  without methanol. The cells continued to shake at saturation density ( $OD_{600}$  of 20-30) for  $\sim 12$  hours to metabolize residual  $d_8$ -glycerol. One hour before induction, 200 mg/L of labeled  $\alpha$ -ketobutyric acid (methyl- $^{13}C$ , 99%; 3,3- $D_2$ , 98%; Cambridge Isotope Laboratories) was added to the culture. Immediately prior to induction, 500 nM taranabant (MedChemExpress) or 5  $\mu M$  suvorexant (Adooq Biosciences) were added to CB1 or  $OX_2R$  cultures, respectively. Expression was induced by adding 0.5%  $d_4$ -methanol (Cambridge Isotope Laboratories) and the temperature was reduced to 20° C. Additional aliquots of 0.5%  $d_4$ -methanol were added every 12 hours, and cells were harvested after 36 hours and stored at -80° C. Following the slow deuteration adaptation procedure as described should help reduce cellular stress from adaptation and increase the final overall deuteration level by minimizing contamination from  $^1H$ -containing media and metabolites into later cultures.

### *Receptor purification*

Deuterated CB1 and  $OX_2R$  cultures were purified in an identical manner to  $A_{2A}$  (Clark et al. 2017), only substituting the ligand present throughout the purification

with 1  $\mu\text{M}$  taranabant or 5  $\mu\text{M}$  suvorexant for CB1 or OX<sub>2</sub>R, respectively. IMAC-purified samples were injected on a Superdex200 column (GE Healthcare) equilibrated in D<sub>2</sub>O containing 150 mM NaCl, 20 mM HEPES pH 7.5, 0.05% deuterated DDM (Anatrace), and 1  $\mu\text{M}$  taranabant or 5  $\mu\text{M}$  suvorexant (Fig. 2). Fractions containing receptor were collected and concentrated immediately prior to NMR experiments. Final yields ranged from 0.3-0.5 mg of purified receptor per liter of deuterated culture.

#### *NMR experiments*

Receptor samples were concentrated to a volume of  $\sim 100$   $\mu\text{L}$  and loaded into 3 mm Shigemi tubes. Final concentrations ranged from  $\sim 30$   $\mu\text{M}$  (CB1) to  $\sim 100$   $\mu\text{M}$  (OX<sub>2</sub>R and A<sub>2A</sub>). NMR spectra were collected at 30°C using a Bruker AVANCE III HD 800 MHz spectrometer with a cryogenically-cooled TCI probe. <sup>1</sup>H/<sup>13</sup>C-HMQC spectra of the methyl region centered at 14 ppm in the <sup>13</sup>C dimension were collected with a <sup>13</sup>C spectral width of 30 ppm for CB1 and OX<sub>2</sub>R and 22 ppm for A<sub>2A</sub>. In all cases, 32 complex points were collected in the <sup>13</sup>C dimension. HMQC spectra for CB1 and OX<sub>2</sub>R were collected as sequential experiments with 64 scans per t<sub>1</sub> point ( $\sim 1.25$  hr); peak intensities and locations in these spectra remained consistent for the first four experiments, so these were summed and processed with NMRpipe (Delaglio et al. 1995) and analyzed with NMRViewJ (Johnson 2004). For both proteins, peak intensities greatly decreased after four 2D spectra were acquired (ca.

5 hr), leaving only few peaks above the noise level. For A<sub>2A</sub>, while the spectra remained consistent and free of extensive signal loss for a longer time (~12 hr), a single 2.5 hr HMQC experiment sufficed to generate enough signal-to-noise for a high quality spectrum due to higher starting concentration. All spectra were processed without linear prediction in the indirect dimension to optimize signal-to-noise ratio.

### **Membrane protein labeling in insect cells**

*Spodoptera frugiperda* ('insect cells' or Sf9) remains the most common expression system used for GPCRs and other membrane proteins (Massotte 2003), with a number of routes developed to facilitate isotopic labeling for NMR studies (Franke *et al.* 2018). Simplest among these is supplementation of growth media with <sup>13</sup>C-labeled amino acids (e.g. methionine), enabling 80-90% labeling efficiencies, as demonstrated for NMR studies of the  $\beta_2$ -adrenergic receptor (Kofuku *et al.* 2012; Kofuku *et al.* 2018; Nygaard *et al.* 2013). Similarly, insect cells will uptake deuterated amino acids from (protonated) growth media and incorporate them into recombinant proteins albeit with backprotonation at certain sites (Kofuku *et al.* 2014; Kofuku *et al.* 2018). Such samples have been used to probe the conformational equilibrium of the  $\beta_2$ -adrenergic receptor (Kofuku *et al.* 2014), the  $\mu$ -opioid receptor (Okude *et al.* 2015), and the P2X<sub>4</sub> purinergic receptor (Minato *et al.* 2016) in various membrane mimetics. As an alternative to <sup>13</sup>C labeling, several groups have successfully carried

out  $^{15}\text{N}$  labeling in insect cells (*Brüggert et al. 2003; Gossert et al. 2011; Strauss et al. 2003; Strauss et al. 2005*). Isogai *et al.* produced the first backbone NMR spectra of a GPCR with the addition of  $^{15}\text{N}$ -valine (Isogai *et al. 2016*). Finally, uniformly labeled samples can be purified from insect cells grown in media formulated with “homemade” yeast extract from labeled *Pichia pastoris* cultures (Opitz *et al. 2015*). It is likely that further developments in media formulation will permit other isotopic labeling schemes and broaden the utility of insect cells for NMR sample preparation.

### **Membrane protein labeling in mammalian cells**

Some eukaryotic membrane proteins require a mammalian expression host for efficient overexpression. Like insect cells, mammalian cells are unable to efficiently grow in minimal media or media containing high concentrations of  $\text{D}_2\text{O}$  (*Murphy et al. 1977*), complicating attempts at isotopic labeling. Further, low yields of purified protein can make NMR samples prohibitively expensive. Nevertheless, specifically-labeled amino acids can be taken up by mammalian cells and incorporated into recombinant proteins, exemplified by NMR analyses of rhodopsin from a stable HEK293 cell line (*Ahuja et al. 2009; Dutta et al. 2012; Goncalves et al. 2010; Werner et al. 2007*). Depending on the formulation of the amino acid lysate added to the culture media, isotopic labeling efficiency can reach up to 90% as determined by FTIR spectroscopy (*Egorova-Zachernyuk et al. 2010*). While cost concerns have limited the use of this approach to NMR studies of rhodopsin at this

time, advances in labeling efficiency and NMR sensitivity may enable broader application in the future.

### **Membrane protein labeling in yeast**

Yeast are the most commonly used eukaryotic systems for expression of NMR samples. Although a variety of yeast can be employed for protein expression and NMR isotope labeling (Kamiya *et al.* 2011; Miyazawa-Onami *et al.* 2013; Sugiki *et al.* 2012), the most commonly used host is the methylotrophic yeast *Pichia pastoris*. *Pichia* can be grown to high cell densities in shake flasks or fermentation, and robust levels of expression are achieved through target gene insertion under control of the AOX promoters and subsequent induction with methanol (Cereghino and Cregg 2000). Proteins can be expressed in a variety of *Pichia* cell strains with wild-type AOX1 and AOX2 proteins (Mut<sup>+</sup> phenotype), or strains with a mutant AOX1, facilitating slow growth upon induction with methanol (Mut<sup>s</sup> phenotype) (Byrne 2015); due to strain-to-strain variations in protein expression, it is common to empirically test several options in parallel to identify those with optimal yield. Contrary to plasmid expression in bacteria, the gene-of-interest is integrated into the *Pichia* genome to create a stable cell line, often with multiple copy integration. After integration, individual clones are typically screened for copy number through increasing antibiotic concentration on selection media. This said, it is not recommended to rely on copy number (e.g. through PCR quantitation) as a proxy for

protein expression level (Aw and Polizzi 2013), and subsequent screening should be performed by small-scale expression tests in either protonated or deuterated media. For more detailed reviews of available *Pichia* vectors and expression lines, see several recent reports and review articles (Byrne 2015; Cos *et al.* 2006; Fan *et al.* 2015).

*Pichia* is advantageous for the expression of isotopically-labeled samples, due to its capacity to grow in either protonated or deuterated minimal media (Massou *et al.* 1999; Morgan *et al.* 2000; Sugiki *et al.* 2012). Accordingly, uniform  $^{15}\text{N}$  labeling is straightforwardly accomplished with the addition of 0.5% to 1% w/w of  $^{15}\text{N}$  ammonium sulfate to yeast nitrogen base lacking amino acids and ammonium sulfate (Rodriguez and Krishna 2001); correspondingly, different  $^{13}\text{C}$ -labeled carbon sources can be used to achieve various labeling patterns as well (Wood and Komives 1999). Examples of these approaches as applied to membrane proteins include solid state NMR studies of U- $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled aquaporin-1 (Emami *et al.* 2013) and rhodopsin (Fan *et al.* 2011), both of which were reconstituted into proteoliposomes. More complex  $^{13}\text{C}$  labeling patterns are achievable through the use of specifically-labeled carbon sources, such as  $^{13}\text{C}$  sparse labeling from media containing alternatively-labeled glycerol and  $^{13}\text{C}$  methanol to reduce peak overlap and improve spectral resolution (Liu *et al.* 2016), or by the combination of amino-acid auxotrophic *Pichia* strains and labeled amino acids, leading to >99% labeling efficiencies (Whittaker and Whittaker 2005).

We summarize the general pros and cons of using insect, mammalian, and yeast expression systems in Table 1.

### **Application of methyl labeling in *Pichia* to GPCRs**

With these advantages, *Pichia* has proven a viable expression host for the production of a range of eukaryotic membrane proteins, including ion channels and GPCRs, for X-ray crystallography efforts (Guo *et al.* 2016; Hino *et al.* 2012; Shimamura *et al.* 2011; Yurugi-Kobayashi *et al.* 2009). *Pichia*-generated samples of the adenosine A<sub>2A</sub> receptor have also enabled solution NMR studies, using C<sup>19</sup>F<sub>3</sub> labels chemically added *in vitro* (Ye *et al.* 2016) or uniformly-labeled <sup>15</sup>N samples monitored at well-dispersed glycine backbone and tryptophan indole sites (Eddy *et al.* 2018).

To extend the applicability of these studies, we sought to take advantage of the favorable NMR properties and locations of methyl groups within GPCRs. While such a route requires optimizing expression conditions to produce highly deuterated, methyl-labeled samples with improved NMR relaxation characteristics, it allows the use of protein samples containing solely the native wildtype sequence. In contrast, other strategies used for solution NMR studies of GPCRs have involved either post-purification chemical modifications (Ye *et al.* 2016) or the introduction of thermostabilizing point mutations to lengthen sample lifetimes and increase signal-

to-noise by permitting NMR data collection at elevated temperatures (Isogai *et al.* 2016; Solt *et al.* 2017).

Methyl groups are highly sensitive probes for NMR, enabling NMR studies of very high-MW complexes such as the proteasome (Religa *et al.* 2010; Sprangers and Kay 2007). Methyl-TROSY experiments with perdeuteration capitalize on the slow relaxation properties of methyl groups to increase both resolution and signal-to-noise (Ollerenshaw *et al.* 2003). Methyl probes are attractive for probing different aspects of receptor activation, but applying methyl-TROSY experiments require an expression system that can facilitate specific  $^{13}\text{C}$ -methyl labeling of samples in an otherwise deuterated background. Labeling the terminal methyl groups of the branched-chain amino acids isoleucine, leucine, and valine can be routinely performed in bacteria, through addition of  $^{13}\text{C}$   $\alpha$ -ketoacids into minimal protonated or deuterated media (Gardner and Kay 1997; Goto *et al.* 1999). However, many eukaryotic membrane proteins cannot be sufficiently overexpressed in bacteria, and although there has been recent success with partial deuteration in Sf9 systems (Kofuku *et al.* 2014; Kofuku *et al.* 2018; Minato *et al.* 2016; Okude *et al.* 2015), the metabolic pathways required for efficient usage of  $\alpha$ -keto acids are not conserved in insect cells.

With this motivation, we recently developed approaches in *Pichia* to produce proteins specifically labeled with  $^1\text{H},^{13}\text{C}$   $\delta^1$  methyl groups of isoleucine residues in a perdeuterated background for both challenging soluble proteins (e.g. actin) (Clark *et*

*al.* 2015) and membrane proteins (Clark *et al.* 2017). This is achieved by addition of methyl-<sup>13</sup>C labeled  $\alpha$ -ketobutyrate to *Pichia* cultures adapted to deuterated media, leading to ~50% labeling of Ile  $\delta$ 1 methyl groups in a ~90% deuterated background (Clark *et al.* 2015). Adaptation of *Pichia* to deuterated media takes about a week, incrementally replacing protons for deuterons in the media and in glycerol, the sole carbon source during biomass generation. Although there are cases where glucose has been successfully used with lower cost (Emami *et al.* 2013; Fan *et al.* 2015; Fan *et al.* 2011), glycerol is generally the preferred carbon source for *Pichia* cultures because the AOX1 promoter is severely repressed by residual glucose and can cause cell death upon methanol induction (Wood and Komives 1999).

Using this labeling strategy, we collected 2D NMR spectra on *Pichia*-expressed deuterated, methyl-labeled A<sub>2A</sub> receptor and were able to assess how the ps-ns dynamics of different isoleucine side chains changed as a result of the ligand bound at the orthosteric pocket (Clark *et al.* 2017). Following this precedent, we collected preliminary 2D <sup>1</sup>H/<sup>13</sup>C HMQC spectra on two other wild-type GPCRs, the CB1 cannabinoid receptor and the OX<sub>2</sub>R orexin receptor bound to their respective antagonists taranabant or suvorexant (Fig. 1). CB1 and OX<sub>2</sub>R receptors have been structurally characterized by X-ray crystallography (Shao *et al.* 2016; Yin *et al.* 2015) and are drug targets for pain and epilepsy (CB1) or sleep disorders (OX<sub>2</sub>R). While the number of dispersed peaks in OX<sub>2</sub>R and CB1 are fewer than for A<sub>2A</sub>, we are able to distinguish specific peaks that could be monitored as probes of dynamics within

the TM bundles. In each of our three samples, the number of visible methyl resonances is fewer than the number of Ile residues present in each receptor. Several possible explanations for this effect are possible. First, a substantial number of Ile residues in these proteins are pointed towards the lipid (Fig. 1), likely with few differences in local environments (and chemical shifts) among these sites, leading to highly overlapped peaks in the center of the Ile  $\delta 1$  methyl region. Another contributor is likely to be chemical exchange among multiple different GPCR conformations and accompanying peak broadening. Such conformational interconversion has been documented at multiple timescales using NMR and MD approaches (Clark *et al.* 2017; Ye *et al.* 2016) and can be mitigated by changing bound ligands (including bound Na<sup>+</sup>), optimizing sample conditions (e.g. salt, buffer, detergent, or temperature) or reconstituting receptors into membrane mimetics such as nanodiscs. While membrane proteins reconstituted into nanodiscs would have longer rotational correlation times and slower tumbling than those in protein-detergent micelles, the protein stability and sample lifetime may be increased, allowing longer data acquisitions (and correspondingly better sensitivity in a range of NMR spectra). Additionally, new SaposinA-based nanodiscs have smaller particle sizes, leading to high quality spectra of GPCRs (Chien *et al.* 2017).

Isoleucines are prevalent in the TM helices of integral membrane proteins, with an average of 30 residues in Class A GPCRs (Pándy-Szekeres *et al.* 2018), providing abundant native NMR probes throughout receptors and other membrane

proteins expressed in this manner. Notably, cell-based labeling approaches do not require any post-purification chemical modification step, as required with several  $^{13}\text{C}$ -formaldehyde or  $^{19}\text{F}$  labeling strategies for membrane proteins (Bokoch *et al.* 2010; Eddy *et al.* 2016; Liu *et al.* 2012; Manglik *et al.* 2015; Sounier *et al.* 2015; Staus *et al.* 2016; Ye *et al.* 2016). Further spectral probes are likely to be accessible in *Pichia*-expressed samples through a recently-developed protocol for  $^1\text{H}$ ,  $^{13}\text{C}$ -methyl labeled valine and leucine using low pH to facilitate uptake of  $\alpha$ -ketoisovalerate as a precursor to these amino acids (Suzuki *et al.* 2018).

## Conclusions

Membrane proteins have remained challenging targets for NMR spectroscopy due to their large size and typical requirement for a eukaryotic host for heterologous expression. *Pichia pastoris* is a promising expression host for biophysical studies of membrane proteins and has been fruitful for structural and dynamic characterizations of human GPCRs. In particular, methyl labeling of isoleucine residues in a deuterated background permit sensitive methyl-TROSY experiments (Ollerenshaw *et al.* 2003) that can help deepen our understanding of GPCR dynamics at the atomic level, including probing the transition between inactive and active conformations. We and other groups have shown that the  $\text{A}_{2\text{A}}$  receptor is a tractable model system to explore the basis of ligand modulation of GPCRs (Clark *et al.* 2017; Eddy *et al.* 2018; Ye *et al.* 2016), and we demonstrate here that other wild-

type GPCRs can be expressed and labeled in the same manner for NMR (Fig. 1).

Further developments are underway to expand our labeling capabilities in *Pichia* and help reduce costs of isotopic labeling in *Pichia* and in other eukaryotic expression systems.

## References

- Ahuja S, Hornak V, Yan ECY, Syrett N, Goncalves JA, Hirshfeld A, Ziliox M, Sakmar TP, Sheves M, Reeves PJ, Smith SO, Eilers M (2009) Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation. *Nat Struct Mol Biol* 16:168-175
- Almén MS, Nordström KJ, Fredriksson R, Schiöth HB (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* 7:50
- Aw R, Polizzi KM (2013) Can too many copies spoil the broth? *Microbial Cell Factories* 12:128
- Bokoch MP, Zou Y, Rasmussen SGF, Liu CW, Nygaard R, Rosenbaum DM, Fung JJ, Choi H-J, Thian FS, Kobilka TS, Puglisi JD, Weis WI, Pardo L, Prosser RS, Mueller L, Kobilka BK (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463:108-112
- Brüggert M, Rehm T, Shanker S, Georgescu J, Holak TA (2003) A novel medium for expression of proteins selectively labeled with <sup>15</sup>N-amino acids in *Spodoptera frugiperda* (Sf9) insect cells. *J Biomol NMR* 25:335-348
- Byrne B (2015) *Pichia pastoris* as an expression host for membrane protein structural biology. *Curr Opin Struct Biol* 32:9-17
- Casiraghi M, Damian M, Lescop E, Point E, Moncoq K, Morellet N, Levy D, Marie J, Guittet E, Baneres JL, Catoire LJ (2016) Functional modulation of a G Protein-Coupled Receptor conformational landscape in a lipid bilayer. *J Am Chem Soc* 138:11170-11175
- Cereghino JL, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24:45-66
- Chien CH, Helfinger LR, Bostock MJ, Solt A, Tan YL, Nietlispach D (2017) An adaptable phospholipid membrane mimetic system for solution NMR studies of membrane proteins. *J Am Chem Soc* 139:14829-14832
- Clark L, Zahm JA, Ali R, Kukula M, Bian L, Patrie SM, Gardner KH, Rosen MK, Rosenbaum DM (2015) Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*. *J Biomol NMR* 62:239-245

Clark LD, Dikiy I, Chapman K, Rödström KE, Aramini J, LeVine MV, Khelashvili G, Rasmussen SG, Gardner KH, Rosenbaum DM (2017) Ligand modulation of sidechain dynamics in a wild-type human GPCR. *Elife* 6:e28505

Cos O, Ramon R, Montesinos JL, Valero F (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microb Cell Fact* 5:17

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-293

Dutta A, Saxena K, Schwalbe H, Klein-Seetharaman J (2012) Isotope labeling in mammalian cells. *Methods Mol Biol* 831:55-69

Eddy Matthew T, Didenko T, Stevens Raymond C, Wüthrich K (2016)  $\beta$ 2-Adrenergic receptor conformational response to fusion protein in the third intracellular loop. *Structure* 24:2190-2197

Eddy MT, Lee MY, Gao ZG, White KL, Didenko T, Horst R, Audet M, Stanczak P, McClary KM, Han GW, Jacobson KA, Stevens RC, Wüthrich K (2018) Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. *Cell* 172:68-80 e12

Egorova-Zachernyuk TA, Bosman GJCGM, DeGrip WJ (2010) Uniform stable-isotope labeling in mammalian cells: formulation of a cost-effective culture medium. *Appl Microbiol Biotechnol* 89:397-406

Emami S, Fan Y, Munro R, Ladizhansky V, Brown LS (2013) Yeast-expressed human membrane protein aquaporin-1 yields excellent resolution of solid-state MAS NMR spectra. *J Biomol NMR* 55:147-155

Fan Y, Emami S, Munro R, Ladizhansky V, Brown LS (2015) Isotope labeling of eukaryotic membrane proteins in yeast for solid-state NMR. *Methods Enzymol* 565:193-212

Fan Y, Shi L, Ladizhansky V, Brown LS (2011) Uniform isotope labeling of a eukaryotic seven-transmembrane helical protein in yeast enables high-resolution solid-state NMR studies in the lipid environment. *J Biomol NMR* 49:151-161

Franke B, Opitz C, Isogai S, Grahl A, Delgado L, Gossert AD, Grzesiek S (2018) Production of isotope-labeled proteins in insect cells for NMR. *J Biomol NMR* 71:173-184

Gardner KH, Kay LE (1997) Production and incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $^1\text{H}$ -  $\delta^1$  methyl) isoleucine into proteins for multidimensional NMR studies. *J Am Chem Soc* 119:7599-7600

Goncalves JA, South K, Ahuja S, Zaitseva E, Opefi CA, Eilers M, Vogel R, Reeves PJ, Smith SO (2010) Highly conserved tyrosine stabilizes the active state of rhodopsin. *Proc Natl Acad Sci U S A* 107:19861-19866

Gossert AD, Hinniger A, Gutmann S, Jahnke W, Strauss A, Fernández C (2011) A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility. *J Biomol NMR* 51:449-456

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta^1$ ) methyl-protonated  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  labeled proteins. *J Biomol NMR* 13:369-374

Guo J, Zeng W, Chen Q, Lee C, Chen L, Yang Y, Cang C, Ren D, Jiang Y (2016) Structure of the voltage-gated two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature* 531:196-201

Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G (2008) Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* 321:1206-1210

Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, Cameron AD, Kobayashi T, Hamakubo T, Iwata S, Murata T (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482:237-240

Hou X, Pedi L, Diver MM, Long SB (2012) Crystal structure of the calcium release-activated calcium channel Orai. *Science* 338:1308-1313

Isogai S, Deupi X, Opitz C, Heydenreich FM, Tsai C-J, Brueckner F, Schertler GFX, Vepintsev DB, Grzesiek S (2016) Backbone NMR reveals allosteric signal transduction networks in the  $\beta$ 1-adrenergic receptor. *Nature* 530:237-241

Jin MS, Oldham ML, Zhang Q, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from *Caenorhabditis elegans*. *Nature* 490:566-569

Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313-352

Kamiya Y, Yamamoto S, Chiba Y, Jigami Y, Kato K (2011) Overexpression of a homogeneous oligosaccharide with  $^{13}\text{C}$  labeling by genetically engineered yeast strain. *J Biomol NMR* 50:397-401

Kim HJ, Howell SC, Van Horn WD, Jeon YH, Sanders CR (2009) Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. *Prog Nucl Magn Reson Spectrosc* 55:335-360

Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Maeda M, Tsujishita H, Shimada I (2012) Efficacy of the  $\beta$ 2-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nature Communications* 3:1045-1049

Kofuku Y, Ueda T, Okude J, Shiraishi Y, Kondo K, Mizumura T, Suzuki S, Shimada I (2014) Functional dynamics of deuterated  $\beta$ 2-adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angewandte Chemie International Edition* 53:13376-13379

Kofuku Y, Yokomizo T, Imai S, Shiraishi Y, Natsume M, Itoh H, Inoue M, Nakata K, Igarashi S, Yamaguchi H, Mizukoshi T, Suzuki EI, Ueda T, Shimada I (2018) Deuteration and selective labeling of alanine methyl groups of beta2-adrenergic receptor expressed in a baculovirus-insect cell expression system. *J Biomol NMR* 71:185-192

Laroche Y, Storme V, De Meutter J, Messens J, Lauwereys M (1994) High-level secretion and very efficient isotopic labeling of tick anticoagulant peptide (TAP) expressed in the methylotrophic yeast, *Pichia pastoris*. *Biotechnology (N Y)* 12:1119-1124

Lee JY, Kinch LN, Borek DM, Wang J, Wang J, Urbatsch IL, Xie XS, Grishin NV, Cohen JC, Otwinowski Z, Hobbs HH, Rosenbaum DM (2016) Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* 533:561-564

Liu J, Liu C, Fan Y, Munro RA, Ladizhansky V, Brown LS, Wang S (2016) Sparse  $^{13}\text{C}$  labelling for solid-state NMR studies of *P. pastoris* expressed eukaryotic seven-transmembrane proteins. *J Biomol NMR* 65:7-13

Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012) Biased signaling pathways in  $\beta$ 2-adrenergic receptor characterized by  $^{19}\text{F}$ -NMR. *Science* 335:1106-1110

- Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK (2015) Structural insights into the dynamic process of  $\beta$ 2-adrenergic receptor signaling. *Cell* 161:1101-1111
- Massotte D (2003) G protein-coupled receptor overexpression with the baculovirus-insect cell system: a tool for structural and functional studies. *Biochim Biophys Acta* 1610:77-89
- Massou S, Puech V, Talmont F, Demange P, Lindley ND, Tropis M, Milon A (1999) Heterologous expression of a deuterated membrane-integrated receptor and partial deuteration in methylotrophic yeasts. *J Biomol NMR* 14:231-239
- Minato Y, Suzuki S, Hara T, Kofuku Y, Kasuya G, Fujiwara Y, Igarashi S, Suzuki E-i, Nureki O, Hattori M, Ueda T, Shimada I (2016) Conductance of P2X 4purinergic receptor is determined by conformational equilibrium in the transmembrane region. *Proceedings of the National Academy of Sciences of the United States of America* 113:4741-4746
- Miyazawa-Onami M, Takeuchi K, Takano T, Sugiki T, Shimada I, Takahashi H (2013) Perdeuteration and methyl-selective  $^1\text{H}$ ,  $^{13}\text{C}$ -labeling by using a *Kluyveromyces lactis* expression system. *J Biomol NMR* 57:297-304
- Morgan WD, Kragt A, Feeney J (2000) Expression of deuterium-isotope-labelled protein in the yeast *pichia pastoris* for NMR studies. *J Biomol NMR* 17:337-347
- Murphy J, Desai C, Giaretti W, Kendall F, Nicolini C (1977) Experimental results on mammalian cells growing in vitro in deuterated medium for neutron-scattering studies. *J Cell Sci* 25:87-94
- Nygaard R, Zou Y, Dror RO, Mildorf TJ, Arlow DH, Manglik A, Pan AC, Liu CW, Fung JJ, Bokoch MP, Thian FS, Kobilka TS, Shaw DE, Mueller L, Prosser RS, Kobilka BK (2013) The dynamic process of  $\beta$ 2-adrenergic receptor activation. *Cell* 152:532-542
- Okude J, Ueda T, Kofuku Y, Sato M, Nobuyama N, Kondo K, Shiraishi Y, Mizumura T, Onishi K, Natsume M, Maeda M, Tsujishita H, Kuranaga T, Inoue M, Shimada I (2015) Identification of a conformational equilibrium that determines the efficacy and functional selectivity of the  $\mu$ -opioid receptor. *Angewandte Chemie* 127:15997-16002
- Oldham ML, Hite RK, Steffen AM, Damko E, Li Z, Walz T, Chen J (2016) A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. *Nature* 529:537-540

Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* 41:843-852

Opitz C, Isogai S, Grzesiek S (2015) An economic approach to efficient isotope labeling in insect cells using homemade <sup>15</sup>N-, <sup>13</sup>C- and <sup>2</sup>H-labeled yeast extracts. *J Biomol NMR* 62:373-385

Pándy-Szekeres G, Munk C, Tsonkov TM, Mordalski S, Harpsøe K, Hauser AS, Bojarski AJ, Gloriam DE (2018) GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res* 46:D440-D446

Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM, Opella SJ (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491:779-783

Religa TL, Sprangers R, Kay LE (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328:98-102

Rodriguez E, Krishna NR (2001) An economical method for (<sup>15</sup>N)/(<sup>13</sup>C) isotopic labeling of proteins expressed in *Pichia pastoris*. *J Biochem* 130:19-22

Shao Z, Yin J, Chapman K, Grzemska M, Clark L, Wang J, Rosenbaum DM (2016) High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature* 540:602-606

Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475:65-70

Solt AS, Bostock MJ, Shrestha B, Kumar P, Warne T, Tate CG, Nietlispach D (2017) Insight into partial agonism by observing multiple equilibria for ligand-bound and Gs-mimetic nanobody-bound  $\beta$ 1-adrenergic receptor. *Nature Communications* 8:1795

Sounier R, Mas C, Steyaert J, Laeremans T, Manglik A, Huang W, Kobilka BK, Déméné H, Granier S (2015) Propagation of conformational changes during  $\mu$ -opioid receptor activation. *Nature* 524:375-378

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618-622

Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, Wingler LM, Ahn S, Chatterjee A, Masoudi A, Kruse AC, Pardon E, Steyaert J, Weis WI, Prosser RS, Kobilka BK, Costa T, Lefkowitz RJ (2016) Allosteric nanobodies reveal the dynamic

range and diverse mechanisms of G-protein-coupled receptor activation. *Nature* 535:448-452

Strauss A, Bitsch F, Cutting B, Fendrich G, Graff P, Liebetanz J, Zurini M, Jahnke W (2003) Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *J Biomol NMR* 26:367-372

Strauss A, Bitsch F, Fendrich G, Graff P, Knecht R, Meyhack B, Jahnke W (2005) Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells. *J Biomol NMR* 31:343-349

Sugiki T, Ichikawa O, Miyazawa-Onami M, Shimada I, Takahashi H (2012) Isotopic labeling of heterologous proteins in the yeast *Pichia pastoris* and *Kluyveromyces lactis*. *Methods Mol Biol* 831:19-36

Suzuki R, Sakakura M, Mori M, Fujii M, Akashi S, Takahashi H (2018) Methyl-selective isotope labeling using alpha-ketoisovalerate for the yeast *Pichia pastoris* recombinant protein expression system. *J Biomol NMR* in press

Werner K, Richter C, Klein-Seetharaman J, Schwalbe H (2007) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* 40:49-53

Whittaker MM, Whittaker JW (2005) Construction and characterization of *Pichia pastoris* strains for labeling aromatic amino acids in recombinant proteins. *Protein Expression Purif* 41:266-274

Whorton MR, MacKinnon R (2011) Crystal structure of the mammalian GIRK2 K<sup>+</sup> channel and gating regulation by G proteins, PIP<sub>2</sub>, and sodium. *Cell* 147:199-208

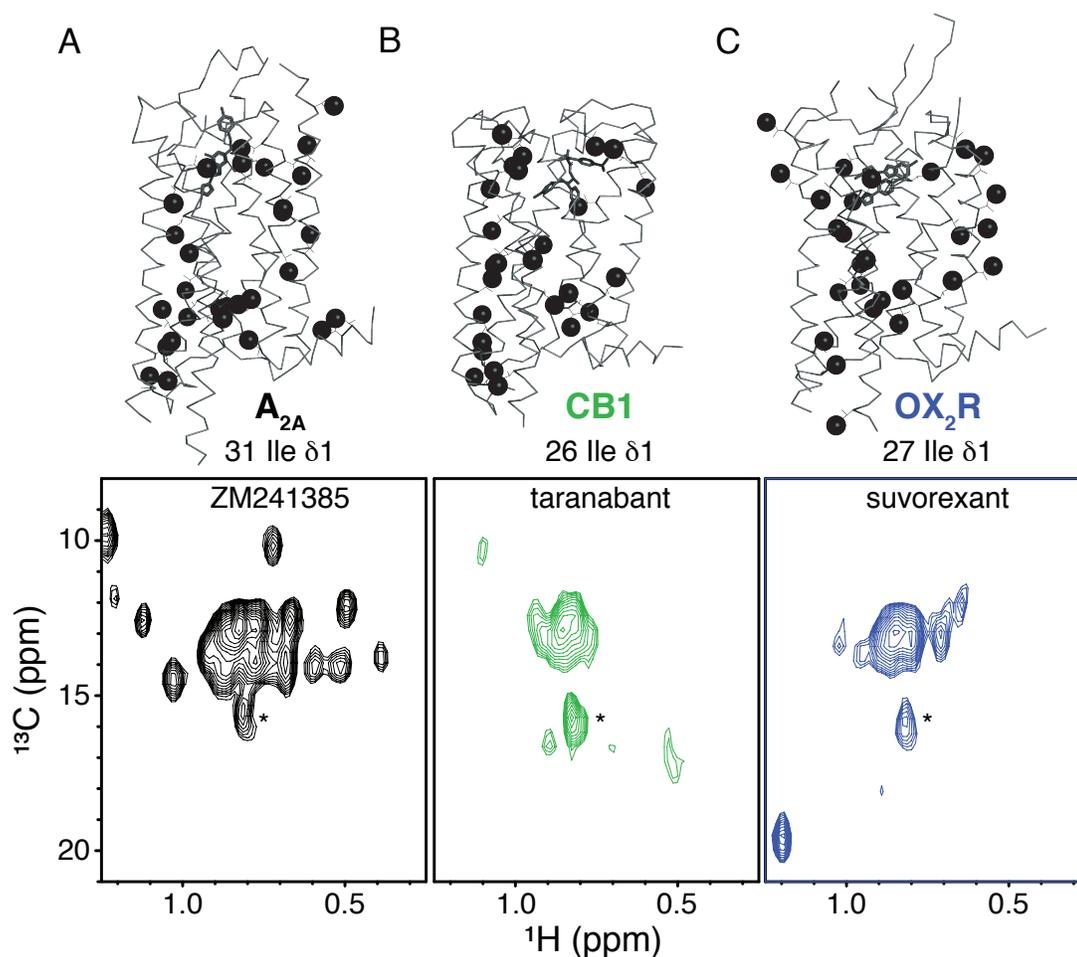
Wood MJ, Komives EA (1999) Production of large quantities of isotopically labeled protein in *Pichia pastoris* by fermentation. *J Biomol NMR* 13:149-159

Ye L, Van Eps N, Zimmer M, Ernst OP, Prosser RS (2016) Activation of the A<sub>2A</sub> adenosine G-protein-coupled receptor by conformational selection. *Nature* 533:265-268

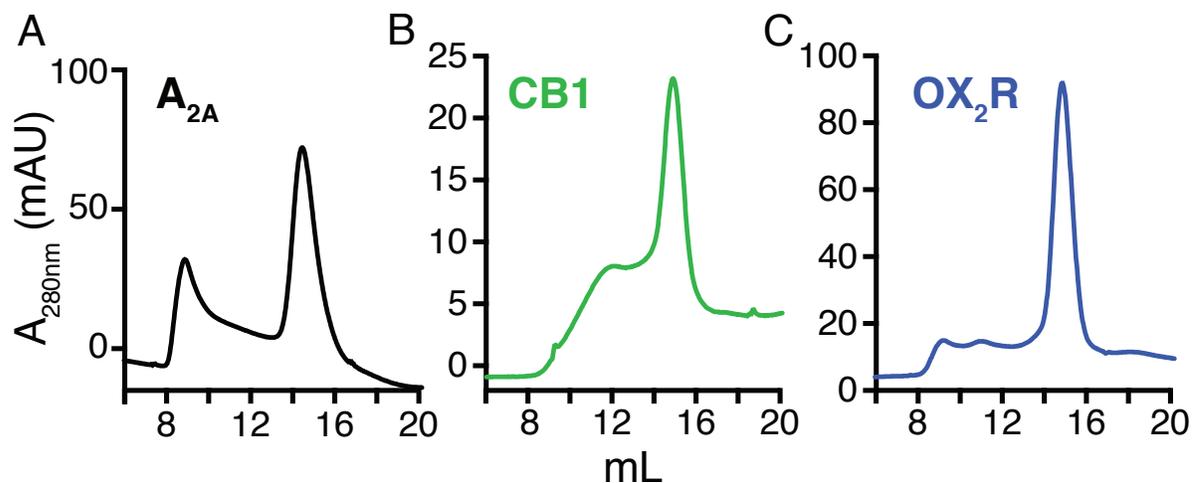
Yin J, Mobarec JC, Kolb P, Rosenbaum DM (2015) Crystal structure of the human OX<sub>2</sub> orexin receptor bound to the insomnia drug suvorexant. *Nature* 519:247-250

Yurugi-Kobayashi T, Asada H, Shiroishi M, Shimamura T, Funamoto S, Katsuta N, Ito K, Sugawara T, Tokuda N, Tsujimoto H, Murata T, Nomura N, Haga K, Haga T,

Iwata S, Kobayashi T (2009) Comparison of functional non-glycosylated GPCRs expression in *Pichia pastoris*. *Biochem Biophys Res Commun* 380:271-276



**Figure 1: Distribution and <sup>1</sup>H/<sup>13</sup>C HMQC spectra of isoleucine δ1 methyl groups within representative GPCRs.** Top – Wire model figures of A<sub>2A</sub> (A) (PDB: 4EIY), CB1 (B) (PDB: 5U09), and OX<sub>2</sub>R (C) (PDB: 4S0V) and structures with ligands shown as sticks, isoleucine sidechains as lines, and isoleucine δ1 carbon atoms as spheres. The number of isoleucine residues is shown as well. Bottom – <sup>1</sup>H/<sup>13</sup>C HMQC spectra of deuterated, Ile δ1-labeled A<sub>2A</sub> (A), CB1 (B), and OX<sub>2</sub>R (C) in DDM micelles with inverse agonists ZM241385, taranabant, and suvorexant, respectively. The spectrum of A<sub>2A</sub> was collected on a sample independently prepared from those previously published (Clark *et al.* 2017), leading to slight differences. In each spectrum, an asterisk shows a peak arising from DDM detergent.



**Figure 2: Sample gel filtration chromatograms of GPCR purifications.** Gel filtration traces of affinity-purified  $A_{2A}$  (A), CB1 (B), and  $OX_2R$  (C) samples injected over a Superdex200 column prior to NMR spectroscopy. In all three samples, fractions comprising the peak eluting at 15 mL were concentrated and used for NMR experiments.

| Expression system   | Pro   | Con  |
|---------------------|---|--|
| <b>Insect cells</b> | <ul style="list-style-type: none"> <li>•Most commonly used system for GPCRs and other MPs</li> <li>•Easy to grow/manipulate</li> </ul>  | <ul style="list-style-type: none"> <li>•ILV methyl labeling not currently possible</li> <li>•High cost to deuterate</li> <li>•Complete deuteration not currently possible</li> </ul> |
| <b>Mammalian</b>    | <ul style="list-style-type: none"> <li>•Required expression system for many MPs</li> <li>•Ideal host for human MPs that are sensitive to lipid environment or PTMs</li> </ul>   | <ul style="list-style-type: none"> <li>•Unable to culture in deuterated media</li> <li>•Lack of strong precedent for difficult systems</li> </ul>                                    |
| <b>Yeast</b>        | <ul style="list-style-type: none"> <li>•Can express many different types of MPs, including GPCRs</li> <li>•Easy to grow/manipulate</li> <li>•Cultured in minimal media, with or without deuteration</li> <li>•Can readily label Ile <math>\delta</math>1 methyl groups</li> </ul> | <ul style="list-style-type: none"> <li>•Complete labeling capabilities not established yet</li> <li>•Not suitable for all MPs</li> </ul>   |

**Table 1: Summary of strengths and weaknesses of eukaryotic systems for the expression and isotopic labeling of integral membrane proteins for NMR studies.**

## **CHAPTER SIX**

### **Structural and functional studies of individual domains of Scap**

#### **INTRODUCTION**

Cells require each of their membranes to contain a specific concentration range and composition of lipids for their function (Brown and Goldstein, 1997), and this is modulated in mammals by the SREBP pathway. The core proteins in this pathway are the SREBP transcription factors, the ER anchor protein Insig, and the cholesterol sensor, Scap (Brown *et al.*, 2017). This chapter provides background on the SREBP pathway and its importance to cellular (and human) health. It will also cover my biochemical characterization of individual domains of Scap during my PhD research, which began as a way to get an initial foothold into understanding the mechanism of how this important protein works. This work is still ongoing, and can be helpful in parallel studies with the full-length protein.

#### **BACKGROUND ON SREBP PATHWAY AND SCAP**

Cholesterol homeostasis is maintained by the SREBP pathway that couples cellular cholesterol levels to transcriptional activation of genes under control of the sterol regulatory element (SRE) that regulate sterol uptake and biosynthesis (Brown and Goldstein, 1986; Brown and Goldstein, 1997). SREBPs are membrane-bound

transcription factors that activate SRE genes, and are bound to Scap in the ER membrane shortly after their synthesis (Sakai *et al.*, 1997). Scap is the key regulator in the SREBP pathway, acting as both the cholesterol sensor (Radhakrishnan *et al.*, 2004) and mediator of intracellular SREBP trafficking through sterol-induced conformational changes (Brown, *et al.*, 2002). When cholesterol levels are low, the ER-luminal domains of Scap communicate this information to a cytoplasmic loop (loop 6) that contains a hexapeptide binding site (MELADL) for COPII vesicle proteins (Sun *et al.*, 2007). The Scap-SREBP complex is then incorporated into vesicles for transport from the ER to the Golgi. In the Golgi membrane, SREBP is cleaved sequentially by two site-specific proteases, S1P and S2P, to release its transcriptionally-active bHLH domain that can travel to the nucleus and recognize and activate SRE-controlled genes, resulting in an increase of cellular sterol levels (Sakai *et al.*, 1996). However, when cholesterol levels are above the required concentration threshold (>5% of total ER lipids), Scap adopts an alternative conformation that hides the MELADL sequence in loop 6 (Radhakrishnan *et al.*, 2008). This change in conformation prevents COPII vesicle proteins from binding to Scap, leading to ER retention and prevention of SREBP transport and ultimately downregulating sterol biosynthesis and uptake (Brown *et al.*, 2002).

Scap is central to this feedback mechanism as it can both sense cholesterol levels and respond through conformational changes that affect SREBP processing. While there is very little structural information known about Scap, the topology and

domain organization have been delineated through biochemical and other assays (Nohturfft *et al.*, 1998). Scap consists of an N-terminal domain containing eight transmembrane helices joined by four short loops and three large loops. Of the three large loops, one of them is cytoplasmic (loop 6) and two are luminal (loops 1 and 7). Of the luminal loops, loop 1 is the larger of the two, consisting of 245 amino acids and harboring the cholesterol-binding site (Motamed *et al.*, 2011). Loop 7 has 173 amino acids and has previously been implicated in regulation of SREBP trafficking (Zhang *et al.*, 2013). The third large loop, loop 6, faces the cytoplasm and contains the COPII binding motif (MELADL) required for ER-to-Golgi transport (Sun *et al.*, 2005). Evidence has shown that cholesterol binding to the luminal loops causes a conformational change in loop 6 that precludes incorporation into vesicles bound for the Golgi apparatus (Brown *et al.*, 2002). Within the eight transmembrane helices, helices 2 through 6 comprise a conserved domain called the sterol sensing domain (SSD) (Kuwabara and Labouesse, 2002). SSDs are seen in other proteins involved in cholesterol regulation, such as HMG CoA reductase as mentioned above, as well as the protein Niemann-Pick disease, type C1 (abbreviated NPC1), responsible for transporting LDL-derived cholesterol out of the lysosome (Li *et al.*, 2016). The transmembrane region of Scap is followed by a cytoplasmic C-terminal WD repeat domain that mediates interaction with SREBP in the ER membrane (Nohturfft *et al.*, 1998). The WD domain is predicted to form a beta propeller structure, similar to other proteins containing WD domains.

The feedback mechanism used by Scap and the SREBP pathway is exploited in treatments with statin drugs, one of the most commonly prescribed medications in the US (Taylor *et al.*, 2013). Statins work by inhibiting cholesterol biosynthesis via direct interactions with the enzyme HMG-CoA reductase (Endo, 1992; Istvan and Deisenhofer, 2001), the rate limiting step in the cholesterol synthesis pathway. By inhibiting cholesterol synthesis, the cell increases the amount of LDL receptors at the cell surface. LDL receptors then can internalize excess LDL from the bloodstream (Brown and Goldstein, 1986), reducing the risk of plaque buildup and atherosclerosis.

Statins work by activating the SREBP pathway, but there is therapeutic potential inhibiting this pathway as well. If the gene encoding Scap is knocked out, shown in a mouse model for obesity, the level of liver triglycerides is drastically reduced (Moon *et al.*, 2012). This effect suggests that developing an inhibitor of Scap could be used to treat fatty liver disease, which can lead to cirrhosis and is currently untreatable. Since Scap is the direct cholesterol sensor in our cells, a detailed molecular understanding of Scap is required in order to develop selective and potent pharmaceuticals that can specifically target Scap and modulate the SREBP pathway.

## CHARACTERIZATION OF SOLUBLE DOMAINS OF SCAP

When working with complicated molecules containing multiple domains, a common tactic is to take a “divide-and-conquer” approach. This type of approach entails “dividing” up your target into its individual domains that are more biochemically tractable, and typically allows for faster characterization of each domain compared to in the full-length context. Ultimately, the long-term goal is to circle back and combine all of the data together to understand how each domain fits together, as well as provide initial data on your system of interest. Scap is a perfect candidate for this approach due to its large size, multiple soluble domains, and inherent flexibility. I built upon previous work from the Brown and Goldstein lab to characterize the soluble domains of Scap (Loop,1 Loop7, C-terminal WD domain) as well as the transmembrane region containing up to all 8 TM helices. While all the information gained from these studies didn’t propel our studies of the full-length system as initially hoped, it provided very important clues into how Scap is working in the ER, and poses interesting new questions that could be answered by future experiments.

## Loop1-Loop7 construct

Cholesterol binds to Scap in the ER lumen, inducing a conformational change that is communicated to the cytosolic loop 6 and leads to its preclusion from COPII vesicle binding (Brown *et al.*, 2002). Previous studies have shown that Scap can bind to [<sup>3</sup>H]-cholesterol with a  $K_d$  of 100nM, yet other sterols that vary slightly in chemical structure have very little to no affinity, suggesting a highly specific binding pocket (Radhakrishnan *et al.*, 2004). The Brown and Goldstein lab have identified this pocket as residing in the luminal loop 1 (Motamed *et al.*, 2011). Through extensive mutagenesis, subsequent work also identified the other large luminal loop, loop 7, as having a regulatory role in cholesterol binding (Zhang *et al.*, 2013). Taking this into account, we engineered a fusion protein containing loop 1 and loop 7 connected through a short linker region (abbreviated L1L7), with a total molecular weight of around 50kD. What is striking about this construct was that it was soluble and could be secreted into the media of Sf9 cells for a simple, one-step purification. This is in contrast to expressing loop 1 by itself, which required detergent for purification, suggesting that this loop is intimately associated with the membrane. Considering that loop 1 contains the cholesterol binding site (Motamed *et al.*, 2011), it would not be surprising if that loop was able to interact with the lipid bilayer. Loop 7 remains soluble if expressed by itself in either Sf9 or mammalian cell lines (Zhang *et al.*, 2016). The requirement of loop7 to keep loop 1 soluble led us to our hypothesis that these two loops are interacting with one another. Binding between these loops

has been reproduced through 1) simple co-expression of the individual loops in mammalian cells and assaying the media for co-secretion (Zhang *et al.*, 2016), and 2) in a L1L7 construct with a protease cleavage site in the linker to determine that the complex stays together (Zhang *et al.* 2016).

Loop 1 contains the cholesterol binding site in Scap (Motamed *et al.*, 2011), and we previously demonstrated that L1L7 constructs maintain the saturable and specific cholesterol binding properties of full-length Scap (Zhang *et al.*, 2016). This was carried out with saturation and competition binding assays using L1L7 and various [<sup>3</sup>H]-labeled sterols. Scintillation counts from experiments done with WT L1L7 resulted in a  $K_d$  of 100nM for cholesterol, consistent with measurements taken with full-length Scap (Radhakrishnan *et al.*, 2004). L1L7 was also able to bind cholesterol specifically, with no appreciable binding to epicholesterol or 25-hydroxysterol, which are extremely structurally similar (Zhang *et al.*, 2016).

Given that this L1L7 construct is soluble, high-expressing, and contains the sterol binding site, obtaining structural information would be incredibly valuable. The size of L1L7 is around 45kD, making it quite small for cryo-EM. However, due to its robust biochemical behavior, it is amenable to other structural methods such as X-ray crystallography and NMR spectroscopy. L1L7 readily crystallized in preliminary screens, but these initial crystals did not diffract. In this section, I will discuss my efforts to improve the construct for crystallization, as well as enable methods to incorporate isotopic labels and collect preliminary 2D NMR spectra.

### *Purification and construct optimization*

For our initial L1L7 construct, we engineered and cloned a synthetic gene encoding loops 1 and 7 from *Cricetulus griseus* (hamster) Scap into a single continuous polypeptide joined by a 15-aa Gly/Ser linker. This construct was inserted into a baculovirus vector for Sf9 infection with an N-terminal hexahistidine tag preceded by a mellitin signal sequence that will direct L1L7 for secretion into the media. After 48 hours of infection, the Sf9 media is filtered and loaded overnight at 4°C onto Ni-NTA resin. After elution, L1L7 is injected over gel filtration (See Fig. 1) to remove any aggregates and migrates as a monomer at the appropriate size (~50kD). The peak is collected and treated with TEV protease to remove the hexahistidine tag followed by PNGase F to remove N-linked sugars (See Fig. 1). The purification yield is approximately 1mg of pure protein per liter of culture, which is sufficient for crystallization experiments.

Previous *in vivo* alanine scanning mutagenesis studies of Scap have shown that regions of loop 7 can tolerate mutation and still retain native SRE activation levels, suggesting that not all regions of loops 1 and 7 are participating in sterol recognition. In contrast, mutations to a central “core” of amino acids in loop 7 result in a loss of function (Zhang *et al.*, 2013). Interestingly, while loop 7 is predicted to mainly exist unstructured, this “core” that is sensitive to mutation is the only region of

loop 7 that is predicted to adopt defined secondary structure. These data together suggest that this structured “core” is necessary for native function of Scap. By removing any unnecessary flexible regions of loop 7, the likelihood of forming an ordered crystal lattice will be increased. In order to find a minimal region of loop 7 that can be used in the L1L7 construct, I systematically screened variants of L1L7 that included truncations of loop 7, removing amino acids from the N- and C-terminal end of the loop. Each of the truncations of loop 7 include the functionally important and structured “core” of amino acids. I expressed each of these L1L7 variants in Sf9 cells and first used secretion efficiency as a proxy for proper folding, followed by large scale purifications.

I found that I could delete up to residues 574-670 (DPAF...LEGR) in loop 7 while still maintaining interaction with Loop1, and the resulting optimized construct gave similar yields and biochemical behavior. Limited proteolysis showed that the resulting optimized L1L7 was generally resistant to proteases (Fig. 2), and thermal and chemical melts with CD spectroscopy demonstrated a non-classical globular fold and a  $T_m$  of 55°C (Fig. 3). This is removing ~75 amino acids of flexible peptide compared to the original construct, improving our chances of obtaining diffraction-quality crystals. The crystals from the minimized construct adopted an overall more uniform appearance, with larger size and more three-dimensional qualities (Fig. 4). Through iterative rounds of crystallization trials followed by cryo-preservation and exposure to high-energy synchrotron beams, the optimized crystals remained

resistant to diffraction. Attempts to co-crystallize L1L7 with cholesterol, as demonstrated with the N-terminal domain of NPC1 (Kwon *et al.*, 2009), were not fruitful. This setback prompted additional alterations to our strategy, described below.

To improve our chances for crystallization, we raised monoclonal antibodies against our optimized L1L7 construct. These efforts were outsourced to a facility at OHSU and returned 17 monoclonal antibodies that made it through multiple screening steps (ELISAs, etc) and in theory their binding should be conformationally specific. However, to our surprise, all 17 of the initial antibodies seemed to recognize L1L7 in Western blots. We were worried that SDS could have been stabilizing some secondary structure within the protein as sometimes seen with  $\alpha$ -helical proteins, and so I repeated the experiment with dot blots, where the protein was directly applied to a nitrocellulose membrane. To mimic the ELISA conditions and further ensure that the protein was completely denatured, it was incubated at 85°C for 30 minutes in 8M urea with 1%  $\beta$ -mercaptoethanol prior to application on the membrane. After this step, there were five antibodies that performed better than negative controls and were our best candidates moving forward, although there was still some recognition of denatured protein (albeit lower than the other candidates) indicating that L1L7 has linear epitopes that the antibodies were raised against. See Figure 5 for antibody details and isotypes. Antibodies were purified from hybridoma supernatants using ProteinG beads and low pH elution, followed by dialysis into

phosphate buffer. Fab fragments were generated from purified antibodies by papain cleavage and ion exchange. Of the five Fab fragments, Fab2 (from Ab2) was the best candidate as determined by SEC. Crystallization setups yielded many hits, but to confirm that our crystals contained L1L7 and not just a Fab fragment (which is easily crystallizable), I labeled L1L7 with tetramethylrhodamine (TMR), which attaches a rhodamine fluorophore to free cysteines. The efficiency of labeling was ~30% in each prep, but yielded a bright pink protein that would be easily visible in any crystals. I confirmed with fluorescence microscopy that subsequent crystals contained L1L7 due to positive rhodamine signal. Our best looking, pink crystals gave sufficient diffraction to solve the structure, but unfortunately each lattice contained only the Fab fragment and not L1L7. While we were not able to locate L1L7 in the crystal lattice, it is likely that the TMR-labeled protein was nonspecifically binding to the Fabs, and likely coated the crystals. After multiple iterations of this without improvement, we stopped attempts at co-crystallization and looked for other strategies.

One of those strategies entailed making individual point mutations within L1L7 to potentially change crystal contacts or the overall behavior of the protein. Loop1 contains three cysteines – two are in a disulfide bond, and one is free and solvent exposed, as demonstrated by reaction with PEG maleimides (data not shown). In order to prevent any aberrant disulfides being formed, I mutated the third cysteine in loop1 within the L1L7 construct (referred to as L1L7 CCA). This CCA mutation

slightly improved the yield and biochemical behavior of the construct, but did not improve the crystal morphology beyond that of wild-type L1L7. In addition, alanine-scanning mutagenesis was previously performed on loop 1 (Motamed *et al.*, 2011) and identified two single point mutations that affected the SREBP trafficking function of Scap. One of these mutations was Y234A, which caused a defect in SREBP trafficking. It should be noted in a followup publication carrying out alanine-scanning in the loop 7, that an additional mutation, Y640S was described that carried the same phenotype (Zhang *et al.*, 2013). Both of these tyrosine mutations (individually or together), prevent secretion of L1L7 in Sf9 or mammalian cells and therefore were not useful for this project. However, an additional mutation in loop 1, V98A, was identified in Motamed *et al.* but was never followed up on in the L1L7 context. Given its phenotype with SREBP trafficking being similar to the tyrosine mutations, I expected the same negative result in the L1L7 background – no secretion. I was surprised to see that the V98A did not fully repress secretion, although it should be noted that the yield was lower than with the wild-type L1L7. L1L7 V98A was able to crystallize readily, as seen with WT, but in a separate set of conditions and with different overall morphology. Unfortunately, while the secretion behavior of this mutant was interesting, these crystals did not yield useable diffraction.

As an additional strategy, I covalently attached various fusion proteins to L1L7. Fusion proteins have proven very useful as folding and crystallization chaperones, providing a well-folded globular domain to stabilize flexible regions and

form crystal contacts during lattice formation. A group of fusion proteins are regularly used to assist with crystallization, stemming from the successes of GPCR crystallization (Rosenbaum *et al.*, 2007). I used three different fusion proteins in the optimized L1L7 construct: T4 lysozyme (T4L), *Pyrococcus abyssi* glycogen synthase (PGS aka G6), and *E. coli* apocytochrome *b*<sub>562</sub>RIL (BRIL). These three fusion proteins have facilitated crystallization of various proteins through improved stabilization and reduced flexibility (Rosenbaum *et al.*, 2007; Yin *et al.*, 2015; Liu *et al.*, 2012). All three fusion proteins were placed at the C-terminus of L1L7 and all three versions were able to efficiently secrete as well as the construct lacking the fusion protein. After purification, by analysis on SDS-PAGE gels, it appears that the fusion protein increased the amount of proteolysis and instability. Fusion proteins were also placed in between the loops with flexible linkers attached, but these constructs also behaved poorly and were not tractable.

Although some of the modifications described above improved the overall yield and behavior of the protein, as well as increasing the three-dimensionality of the crystals, we are still currently unable to obtain any usable diffraction from these crystals. The optimized L1L7 construct is still possible to study by NMR (see below), and while is still biochemically valuable for binding experiments and screening tools (see nanobody chapter), it currently does not seem like a viable target for X-ray crystallography.

### *Crystallization*

The original L1L7 construct, the optimized (aka “v7”) L1L7 construct (WT and CCA versions), and the L1L7 V98A (original and optimized versions) all crystallized in various conditions, with some overlapping components in the mother liquor. Figure 6 contains a table that summarizes the conditions that grew the crystals with the best overall morphology. In order to improve these crystals from these conditions, I varied crystallization temperatures, tray format (hanging drop vs. sitting drop), protein concentration, and component concentrations. In addition, I used Hampton Research additive and detergent screens from the base conditions, and created crystal seeds to perturb crystal nucleation rates. Prior to crystal tray setups, I added extra biochemical steps often employed to improve crystal formation, including deglycosylation of N-linked sugars, reductive methylation of lysine residues and the N-terminal free amine, and various reducing conditions. As mentioned above, despite my best efforts over multiple years, I was unable to obtain any usable diffraction from any of the optimized crystals.

### *Cholesterol binding assays*

We tested the optimized L1L7 for its ability to saturably and specifically bind cholesterol. This was carried out in the same manner as published in Zhang *et al.*,

2016, but will be briefly described below. Each binding reaction was set up in individual Eppendorf tubes in a 100 $\mu$ L volume containing a background buffer of 150mM NaCl, 50mM Tris pH 7.4, 0.004% NP40, and 0.002% FC13. Each reaction contained ~10pmol of L1L7 and increasing amounts of “hot” ( $[^3\text{H}]$ -labeled) or “cold” or (unlabeled) sterol, depending on the type of experiment. Each reaction mixture was incubated at room temperature for 4 hours prior to application to 300 $\mu$ L of  $\text{Ni}^{2+}$ -NTA agarose slurry, pre-equilibrated with the background buffer. Since L1L7 contains a His-tag for purification purposes, it and its bound sterol will bind to the resin. After controlled wash steps, L1L7 and bound sterol are eluted with imidazole-containing buffer and the levels of tritium are measured in a scintillation counter and back-calculated to determine binding curves. In our assays, optimized L1L7 saturated at ~150nM cholesterol, and showed specific binding in competition assays with epicholesterol and 25-hydroxysterol (Zhang *et al.*, 2016). The sterol binding properties of L1L7 are comparable with Scap containing the transmembrane domain (Radhakrishnan *et al.*, 2004). See figures 7 and 8 for binding curves.

### *NMR Spectroscopy*

L1L7 is a soluble protein, and is readily secreted into Sf9 media. It expresses highly and is biochemically well-behaved, but as described above did not yield any diffractable crystals. Given this, I turned to a different method, NMR spectroscopy, to understand if L1L7 could be studied in solution and provide clues on how the loops

interact and where cholesterol could be binding. The drawback to NMR is that it often requires biosynthetic incorporation of NMR active probes, limiting the expression systems that can be readily used. Typically, NMR samples are overexpressed and purified from *E. coli*, where the isotopic labeling is simple and routine. Our attempts to express L1L7 in *E. coli* were not fruitful, and only resulted in misfolded protein in inclusion bodies. I next turned to the yeast *Pichia pastoris*, given my experience with it in my other PhD project. Using the same MF $\alpha$  secretion signal as described in chapter 3, I was able to purify L1L7 from the media, but the yield was quite low and therefore likely not a viable option financially.

As an alternative, I capitalized on the recent availability of different types of “dropout” media for insect cell expression. It is possible to purchase “ILV” dropout media that are missing the branched chain amino acids, and then supplementing the media with isotopically labeled versions of those amino acids for biosynthetic incorporation into your target protein. Following this method, I discovered that labeling isoleucines or valines was not very efficient likely due to low uptake of those amino acids, but  $^{15}\text{N}$ -labeled leucine was readily incorporated into cells and labeled L1L7 was secreted back into the media. After routine purification, the sample was concentrated to  $\sim 150\mu\text{M}$  for 2D HSQC experiments carried out at  $25^\circ\text{C}$  at 800MHz. Compared to 43 leucines present in the optimized L1L7 construct, we saw  $\sim 50$  distinct  $^{15}\text{N}$ - $^1\text{H}$  peaks, although roughly half of them lie within  $\sim 7.8$ - $8.4$ ppm in the proton dimension, suggesting that those amides are arising from unfolded

polypeptide. This is consistent with secondary structure prediction, which predicts 19 out of 43 leucines are present in an unstructured region. There is little to no overlap with previous HSQC spectra taken on FC13-solubilized loop 1 alone. I incubated L1L7 with an excess (20x) of cholesterol, supplemented in ethanol with background levels of NP40 and FC13 of 0.004% and 0.001%, respectively. After incubation at room temperature for an hour, an HSQC spectrum was again taken. Fig. 9 shows the overlay of apo and cholesterol-incubated L1L7. We see a loss (or near-loss) of many peaks downfield of 8.5ppm in the proton dimension, and weaker intensity of some peaks around and upfield of 7ppm. There is still strong intensity for amide peaks in the “unstructured” region, suggesting that cholesterol is not rigidifying this region of the protein into a canonical secondary structure.

As an alternative to solely labeling leucines, a recent study described the ability to uniformly label all nitrogens and carbons within a protein sample (Opitz *et al.*, 2015). This is in contrast to commercially available media, which can run almost \$7,000 per liter. With uniform labeling of nitrogens and carbons in a sample, it is possible to use multi-dimensional NMR methods to solve the 3D structure of protein samples. This is admittedly a lofty goal for a sample like L1L7 that has no structural homology to anything previously solved, and at a somewhat large size for NMR (~50kD). Nevertheless, I initially explored the feasibility of this method with uniform nitrogen labeling.

Uniform labeling in Sf9 cells requires a more drastic “dropout” media that is depleted of all amino acids, not only branched chain amino acids as needed with  $^{15}\text{N}$ -leucine labeling. In order to maintain cell growth and protein expression, the media requires supplementation of a uniform  $^{15}\text{N}$ -labeled yeast extract (8g/L) as well as  $^{15}\text{N}$ -labeled glutamine (1g/L).  $^{15}\text{N}$ -labeled yeastolate is not commercially available, but can be prepared from the yeast *Pichia pastoris*. Since *Pichia* are able to grow on minimal media with a defined nitrogen source, it is simple to exchange the  $^{14}\text{N}$ -ammonium chloride with  $^{15}\text{N}$ -ammonium chloride.  $^{15}\text{N}$ -ammonium chloride is relatively inexpensive due to its routine use in *E. coli* cultures. Wild-type *Pichia* (cell strain X-33) is grown in minimal media in shake flasks to a high OD, with periodic supplementation with a “feeding” solution containing glucose and additional nutrients. Once the glucose is depleted (easily detected on a patient urine strip), the cells are harvested and lysed through multiple steps to break down the yeast to minimal components (zymolyase treatment for 5hrs at 37°C, followed by papain treatment for 5 days at 50°C). The resulting material can then be lyophilized, weighed, and sterile filtered for media supplementation.

Through this strategy, I was able to collect 2D spectra on a uniform  $^{15}\text{N}$ -labeled sample of L1L7, demonstrating that the labeling efficiency with this method is feasible for labeling in Sf9 cells. The 2D spectrum of U- $^{15}\text{N}$  labeled optimized L1L7 shows peaks mostly collapsed within the “unstructured” region, with significant peak overlap (see Fig. 10 for an overlay with  $^{15}\text{N}$ -Leu labeled protein). While there are

peaks dispersed outside of this region (including three well-resolved peaks likely attributable to tryptophan indole NH groups), the amount of peak overlap and overall poor dispersion will make this sample extremely difficult to use for multidimensional NMR experiments. However, this sample could be used to monitor chemical shift changes in specific amino acids, such as the three well-dispersed peaks in the tryptophan region of the spectrum, which is quite downfield from the rest of the peaks. There are 7 total tryptophans in the optimized L1L7 construct so it is likely that signal-to-noise is not high enough to observe them all, or there are intermediate exchange processes that are averaging out our signal with this experiment.

### **C-terminal WD domain**

The C-terminus of Scap contains a WD domain, predicted to form a  $\beta$ -propeller structure based on sequence and homology models. This domain has been shown to interact with the C-terminal domain of SREBP1 and SREBP2, and this complex is formed immediately after protein synthesis in the ER. The *S. pombe* Scap WD domain crystal structure was solved in 2015 (Gong *et al.*, 2015), but there is limited homology to the mammalian Scap WD domain. In addition, the SREBP pathway in yeast is a hypoxia sensing pathway instead of a sterol sensing pathway (Espenshade and Hughes, 2007). Due to these discrepancies, the mammalian Scap WD domain is a valuable structural target. It is quite large (~500 amino acids), and comprises almost half the molecular weight of the full-length Scap. However, since it

is predicted to be a globular domain, it can easily be separated from the transmembrane region of Scap and purified as a separate entity.

Previous unpublished work from the Radhakrishnan lab at UTSW has used the WD domain of Scap, with a construct comprising amino acids 766-1276 (the C-terminus of Scap). This construct expresses well in Sf9 cells and can be purified through a one-step affinity purification. However, the gel filtration traces showed a significant amount of aggregation, both eluting in the void and as larger soluble aggregates. After careful review of the secondary structure prediction for this domain, I changed the N-terminal boundary to contain more amino acids as it appeared that starting at 766 would be in the middle of a  $\beta$ -strand and could cause the construct to be more aggregation prone. I added back amino acids to begin at 759 and 751, and found that both of these constructs improved the biochemical behavior of the WD domain by gel filtration, with a start at 751 being the better of the two.

I initially set up crystallization trials with the 751-end construct, and found that it was able to readily crystallize in a variety of conditions, often in small crystal “showers” that are not large enough for structure determination. The WD domain contains a high amount of cysteine residues, and I found that “capping” the cysteines with iodoacetamide prior to crystallization setups generated larger, more three-dimensional crystals. At this point, this project was handed over to a new postdoc (Dan Kober), who is continuing the characterization of this domain with the

long-term goal of solving the structure of the WD domain bound to a soluble form of the C-terminal domains from SREBP1/2.

## **CHARACTERIZATION OF TRANSMEMBRANE DOMAIN OF SCAP**

The N-terminal domain of Scap contains the 8 transmembrane helices, including the three large loops described earlier. These three loops contain the cholesterol binding site, as well as the COPII binding motif that is required for SREBP trafficking. It has been previously shown that Scap does not require the C-terminal WD domain for cholesterol binding and specificity, leading to a long tenure of study on the transmembrane domain of Scap (Scap<sup>TM</sup>). This domain is about 750 amino acids long, and has typically been expressed and purified from Sf9 cells, which generally has required FC13 detergent. In this section, I will briefly describe my work to characterize this domain on its own, prior to my work with the full-length construct (see next chapter).

### **TM1-8 construct**

Scap<sup>TM</sup> has been the subject of experiments for many years in the Brown and Goldstein lab, and is readily purifiable from Sf9 cells with the detergent FC13. FC13 and other members of the Fos-Choline family are very efficient detergents for membrane protein extraction and purification, but have hampered function in

multiple cases of membrane proteins and/or caused aberrant protein folding or stability (Chipot *et al.*, 2018). Fos-choline detergents are zwitterionic detergents containing a choline and a phosphate group in the headgroup, and typically have large micelles with low critical micelle concentrations (CMCs). The first binding study showing cholesterol binding to Scap<sup>TM</sup> relied on the detergent FC13 for solubilization, indicating that FC13 is not disrupting the ability for Scap<sup>TM</sup> to bind sterols. Based on this, many initial purifications during my dissertation research relied on FC13 as the primary detergent for purification. However, FC13 is an extremely disadvantageous detergent for structural studies due to the zwitterionic headgroup, particularly the phosphate which can cause artifacts in cryo-EM studies.

Considering this fact, my experience with this construct focused on a version cloned from a thermophilic ortholog (see next chapter for more details), as well as the hamster version with a C-terminal protein replacing the WD domain. I replaced the WD domain with maltose binding protein (MBP) with varied lengths of C-terminal linker, with the hope that MBP would be a more stable domain than the WD domain, and could also be used as a chromatography tag in conjunction with amylose resin. However, MBP appeared to be destabilizing to the rest of the protein, possibly due to unforeseen steric clashes with other regions of Scap. I also attached green fluorescent protein (GFP) to the C-terminus, and found that GFP was better tolerated than MBP, but the biochemical behavior was not ideal. The one striking difference came when I included a thermostabilizing mutation (D428V; see next

chapter), that permitted extraction for the first time in a detergent other than FC13. Scap<sup>TM</sup>-GFP was able to be solubilized to a significant degree in DDM and DMNG, but displayed signs of aggregation and heterogeneity during gel filtration and was not further pursued due to other more promising leads with the full-length protein.

### **TM1-6mini construct**

Out of Scap's 8 transmembrane helices, TM2-6 form the sterol sensing domain, present in other integral membrane proteins involved in lipid homeostasis. These helices are predicted to adopt a similar fold as other SSDs, such as in NPC1 or Patched. Expression of Scap from the N-terminus until the middle of loop 6 has been demonstrated in mammalian cells at sufficient enough levels for pulldown experiments (Zhang *et al.*, 2013), but overexpression in Sf9 cells or GnT1- HEK cells is poor. This was quite surprising given the fact that helices 2-6 should form a conserved "unit", and we hypothesized that having loop 1 without loop 7 was the culprit. To test this, we removed loop 1, leaving enough native amino acids to comfortably connect TMs 1 and 2, and tested for expression in mammalian cells with transient transfections. We included a GFP tag on the C-terminus to use for subsequent fluorescence-detected size exclusion chromatography (FSEC) experiments. Compared to the previous construct, we were now getting robust expression in CHO and HEK cell lines, which permitted a thermostability screen by

FSEC which is detailed in the next chapter. We denoted this construct as ScapTM16mini, with the construct containing loop 1 as ScapTM16.

ScapTM16mini expresses well and can be purified to homogeneity in GnT1-cells in DDM (Fig. 11). This holds true for constructs lacking or containing a C-terminal GFP, which in the latter case provides an easy way to track this process and can be removed through proteases if desired. This construct can be concentrated to levels sufficient for lipidic cubic phase (LCP) crystallization (~20mg/mL), but never formed crystals in both monoolein or 7.7 MAG lipids containing cholesterol. In addition, through overexpression in GnT1-cells in the presence of an untagged hamster Insig2 construct, I was able to get interaction of these two proteins to a level that can be monitored by gel, indicating for the first time *in vitro* that this region of Scap is all that is needed to bind Insig2 (Fig. 12), and does not require the other two TMs or any of the large loops.

## REFERENCES

- Brown, A. J., Sun, L., Feramisco, J. D., Brown, M. S., Goldstein, J. L. Cholesterol addition to ER membranes alters conformation of SCAP, the SREBP escort protein that regulates cholesterol metabolism. *Molecular Cell* (2002) vol. 10 (2) pp. 237-245
- Brown, M. S. and Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* (1986) vol. 232 (4746) pp. 34-47
- Brown, M. S. and Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* (1997) vol. 89 (3) pp. 331-340
- Brown, M. S., Radhakrishnan, A., & Goldstein, J. L. (2017). Retrospective on Cholesterol Homeostasis: The Central Role of Scap. *Annual Review of Biochemistry*, 87(1), annurev-biochem-062917-011852-25. <http://doi.org/10.1146/annurev-biochem-062917-011852>
- Chipot, C., Dehez, F., Schnell, J. R., Zitzmann, N., Pebay-Peyroula, E., Catoire, L. J., et al. (2018). Perturbations of Native Membrane Protein Structure in Alkyl Phosphocholine Detergents: A Critical Assessment of NMR and Biophysical Studies. *Chemical Reviews*, 118(7), 3559–3607. <http://doi.org/10.1021/acs.chemrev.7b00570>
- Endo, A. (1992). The discovery and development of HMG-CoA reductase inhibitors. *Journal of Lipid Research*, 33(11), 1569–1582.
- Espenshade, P. J., & Hughes, A. L. (2007). Regulation of Sterol Synthesis in Eukaryotes. *Annual Review of Genetics*, 41(1), 401–427. <http://doi.org/10.1146/annurev.genet.41.110306.130315>
- Gong, X., Li, J., Shao, W., Wu, J., Qian, H., Ren, R., et al. (2015). Structure of the WD40 domain of SCAP from fission yeast reveals the molecular basis for SREBP recognition. *Cell Research*, 25(4), 401–411. <http://doi.org/10.1038/cr.2015.32>
- Istvan, E. S. and Deisenhofer, J. Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* (2001) vol. 292 (5519) pp. 1160-1164
- Kuwabara, P. E., & Labouesse, M. (2002). The sterol-sensing domain: multiple families, a unique role? *Trends in Genetics : TIG*, 18(4), 193–201.

Kwon, H. J., Abi-Mosleh, L., Wang, M. L., Deisenhofer, J., Goldstein, J. L., Brown, M. S., Infante, R. E. Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. *Cell* (2009) vol. 137 (7) pp. 1213-1224

Li, X., Wang, J., Coutavas, E., Shi, H., Hao, Q., & Blobel, G. (2016). Structure of human Niemann–Pick C1 protein. *Proceedings of the National Academy of Sciences*, 113(29), 8212–8217. <http://doi.org/10.1073/pnas.1607795113>

Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., et al. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*, 337(6091), 232–236. <http://doi.org/10.1126/science.1219218>

Moon, Y.-A., Liang, G., Xie, X., Frank-Kamenetsky, M., Fitzgerald, K., Koteliansky, V., et al. (2012). The Scap/SREBP Pathway Is Essential for Developing Diabetic Fatty Liver and Carbohydrate-Induced Hypertriglyceridemia in Animals. *Cell Metabolism*, 15(2), 240–246. <http://doi.org/10.1016/j.cmet.2011.12.017>

Motamed, M., Zhang, Y., Wang, M. L., Seemann, J., Kwon, H. J., Goldstein, J. L., Brown, M. S. Identification of Luminal Loop 1 of Scap Protein as the Sterol Sensor That Maintains Cholesterol Homeostasis. *Journal of Biological Chemistry* (2011) vol. 286 (20) pp. 18002-18012

Nohturfft, A., Brown, M. S., & Goldstein, J. L. (1998). Topology of SREBP Cleavage-activating Protein, a Polytopic Membrane Protein with a Sterol-sensing Domain. *Journal of Biological Chemistry*, 273(27), 17243–17250. <http://doi.org/10.1074/jbc.273.27.17243>

Opitz, C., Isogai, S., & Grzesiek, S. (2015). An economic approach to efficient isotope labeling in insect cells using homemade <sup>15</sup>N-, <sup>13</sup>C- and <sup>2</sup>H-labeled yeast extracts. *Journal of Biomolecular NMR*, 62(3), 373–385. <http://doi.org/10.1007/s10858-015-9954-3>

Radhakrishnan, A., Sun, L-P., Kwon, H. J., Brown, M. S., Goldstein, J. L. Direct binding of cholesterol to the purified membrane region of SCAP: mechanism for a sterol-sensing domain. *Molecular Cell* (2004) vol. 15 (2) pp. 259-68

Radhakrishnan, A., Goldstein, J. L., McDonald, J. G., Brown, M. S. Switch-like Control of SREBP-2 Transport Triggered by Small Changes in ER Cholesterol: A Delicate Balance. *Cell Metabolism* (2008) vol. 8 (6) pp. 512-521

Sakai, J., Duncan, E. A., Rawson, R. B., Hua, X., Brown, M. S., Goldstein, J. L. Sterol-regulated release of SREBP-2 from cell membranes requires two sequential cleavages, one within a transmembrane segment. *Cell* (1996) vol. 85 (7) pp. 1037-1046

Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C., Kobilka, B. K. GPCR engineering yields high-resolution structural insights into  $\beta$ 2-adrenergic receptor function. *Science* (2007) vol. 318 (5854) pp. 1266-73

Sakai, J., Nohturfft, A., Cheng, D., Ho, Y. K., Brown, M. S., Goldstein, J. L. Identification of complexes between the COOH-terminal domains of sterol regulatory element-binding proteins (SREBPs) and SREBP cleavage-activating protein. *Journal of Biological Chemistry* (1997) vol. 272 (32) pp. 20213-20221

Sun, L.-P., Li, L., Goldstein, J. L., Brown, M. S. Insig required for sterol-mediated inhibition of Scap/SREBP binding to COPII proteins in vitro. *Journal of Biological Chemistry* (2005) vol. 280 (28) pp. 26483-26490

Sun, L.-P., Seemann, J., Goldstein, J. L., Brown, M. S. Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: Insig renders sorting signal in Scap inaccessible to COPII proteins. *Proceedings of the National Academy of Sciences* (2007) vol. 104 (16) pp. 6519-6526

Taylor, F. C., Huffman, M., Ebrahim, S. Statin therapy for primary prevention of cardiovascular disease. *JAMA* (2013) vol. 310 (22) pp. 2451-2452

Yin, J., Mobarec, J. C., Kolb, P., & Rosenbaum, D. M. (2015). Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature*, 519(7542), 247–250. <http://doi.org/10.1038/nature14035>

Zhang, Y., Motamed, M., Seemann, J., Brown, M. S., Goldstein, J. L. Point mutation in luminal loop 7 of Scap protein blocks interaction with loop 1 and abolishes movement to Golgi. *Journal of Biological Chemistry* (2013) vol. 288 (20) pp. 14059-14067

Zhang, Y., Lee, K. M., Kinch, L. N., Clark, L., Grishin, N. V., Rosenbaum, D. M., et al. (2016). Direct Demonstration That Loop1 of Scap Binds to Loop7: A CRUCIAL EVENT IN CHOLESTEROL HOMEOSTASIS. *Journal of Biological Chemistry*, 291(24), 12888–12896. <http://doi.org/10.1074/jbc.M116.729798>

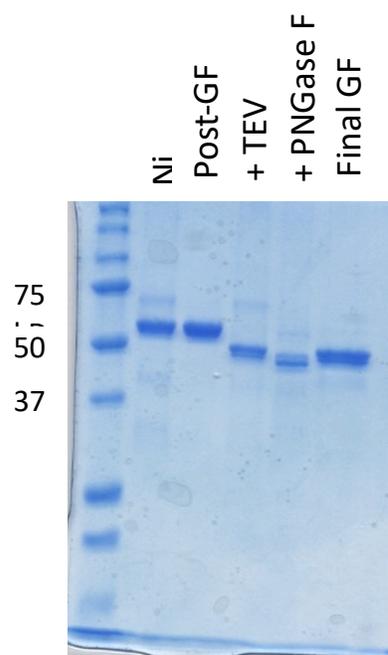
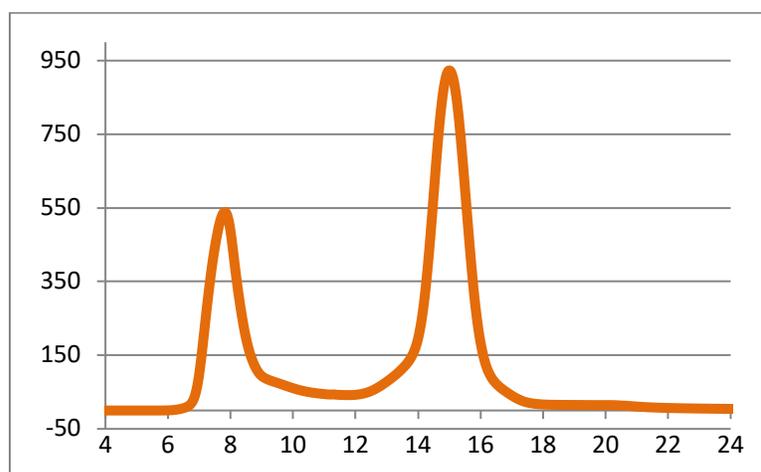


Figure 1. Top: SEC trace of optimized L1L7 (aka v7) on a Superdex 200 10/300 column equilibrated in 150mM NaCl, 20mM HEPES 7.5. Bottom: SDS-PAGE gel of purification steps.

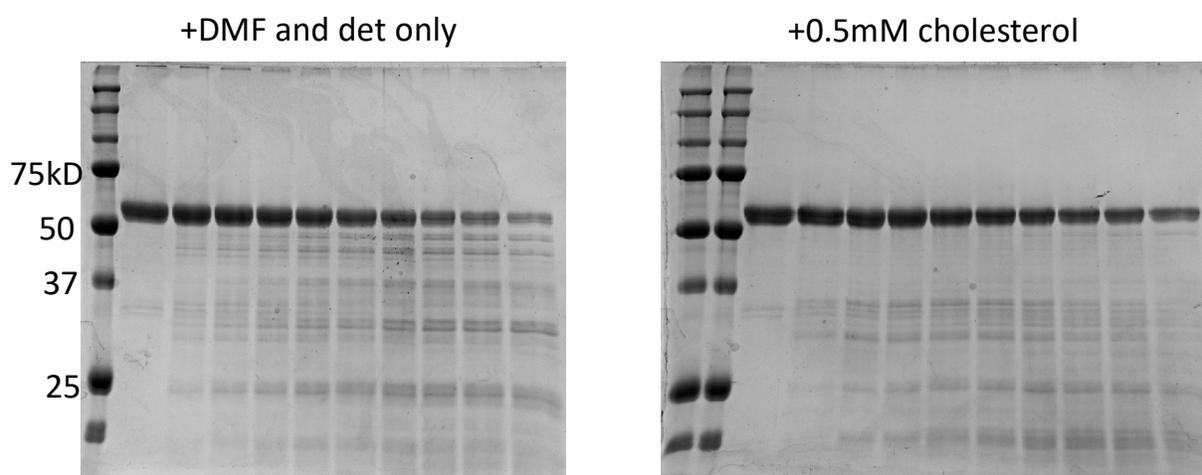


Figure 2. Limited proteolysis of optimized L1L7. Trypsin was incubated at RT with purified L1L7 at a 1:200 w/w ratio of L1L7:trypsin. Time points were taken at 0', 2' 5', 10', 20', 30', 60', 90', 120', 180'. Cholesterol (DMF-solubilized) was incubated with L1L7 in the gel on the right, with vehicle only on the left.

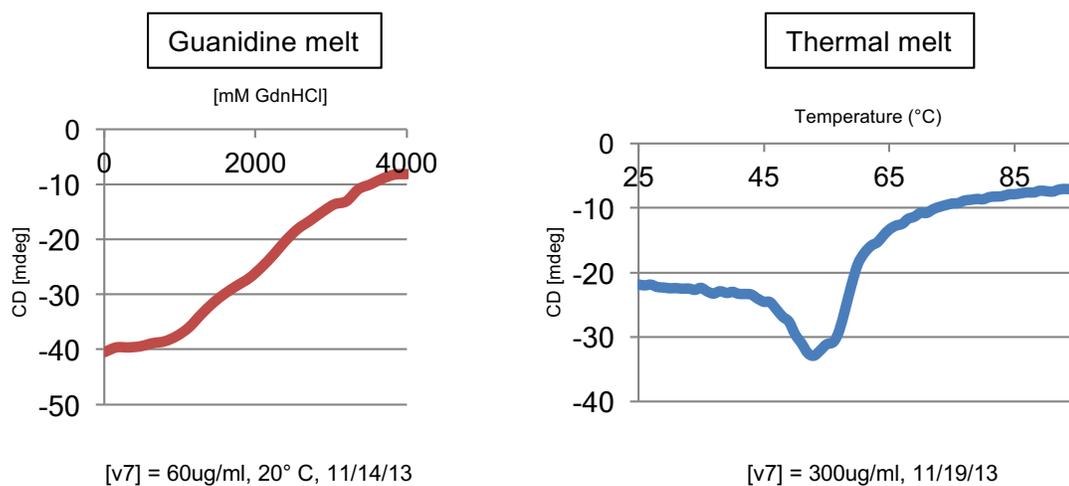


Figure 3. Circular dichroism of optimized L1L7 (aka v7), monitored at a wavelength of 222nm. Chemical melting via GdnHCl is shown on the left, and the thermal melt spectrum is shown on the right. Pertinent experimental details are included below each spectra.

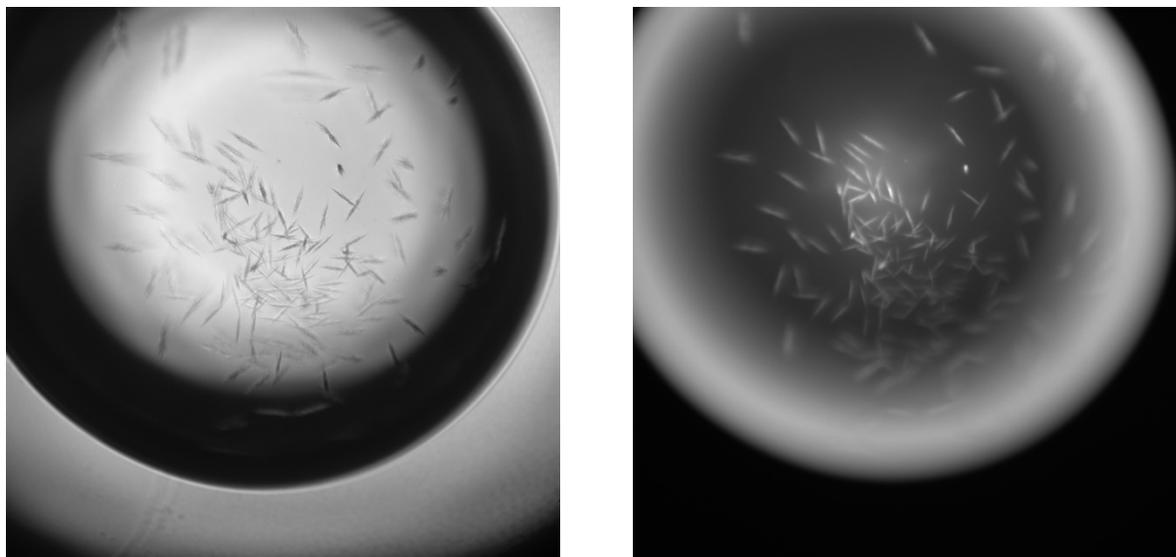


Figure 4. Optimized L1L7 crystals are shown here in large format hanging drop format. Crystals are easily visualized by bright field (left) and UV (right) microscopy.

| shorthand | Antibody | Heavy chain | Light chain |
|-----------|----------|-------------|-------------|
| Ab1       | 10C6.1   | IgG2b       | Kappa       |
| Ab2       | 10G10.1  | IgG2a       | Kappa       |
| Ab3       | 14F6.1   | IgG2a       | Kappa       |
| Ab4       | 16E8.1   | IgG1        | Kappa       |
| Ab5       | 17E12.2  | IgG1        | Kappa       |

Figure 5. L1L7 monoclonal antibody details.

| Construct                   | Best condition  |
|-----------------------------|---|
| Optimized L1L7              | ~20% PEG 3350, 0.1M HEPES pH 7.0, 0.2M Ammonium sulfate or formate<br><br>20% PEG 1000, 0.1M Tris pH 7.0<br><br>0.2M MgCl <sub>2</sub> , 0.1M Tris pH 7.0, 10% PEG 8000 |
| Both optimized and original | 20mM Tris pH 7.0, 0.1M NaCl, 7.7% PEG 4000<br><br>~24% PEG 4000, 0.15M AmSO <sub>4</sub> , 15% glycerol<br><br>1M LiCl, variety of buffers from pH 6.5-9, 10% PEG 6000  |
| Original L1L7               | 0.1M Li <sub>2</sub> SO <sub>4</sub> , 0.1M NaCl, 0.1M Tris pH 8.5, 30% v/v PEG 400   |
| Optimized L1L7 + V98A       | 0.1M tri-Na citrate 5.6, 12% PEG 4000, ±0.1M NaCl, ±0.1M Li <sub>2</sub> SO <sub>4</sub>  |

Figure 6. L1L7 crystallization conditions that produced the best looking crystals described in this chapter. Unfortunately, none of these crystals diffracted.

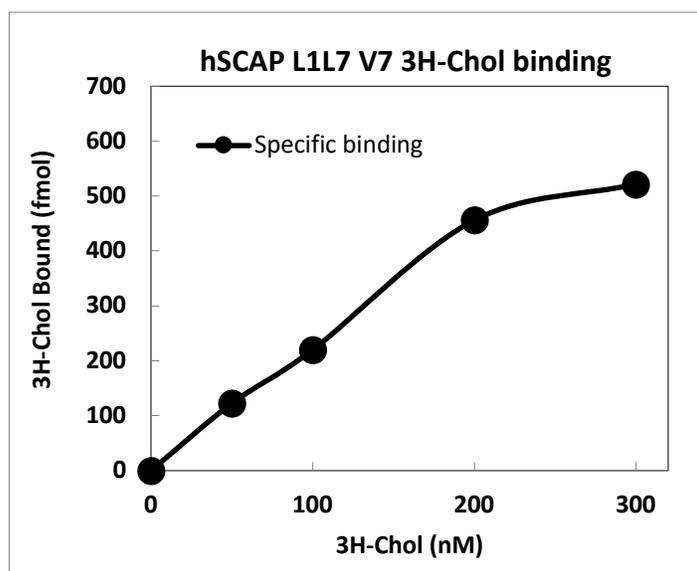
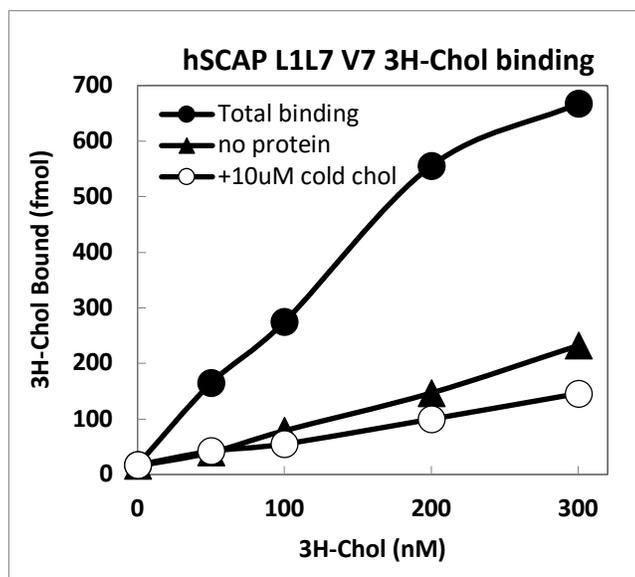


Figure 7. Saturation binding of optimized L1L7 (aka v7). Top: Samples were incubated with increasing amounts of tritiated cholesterol with or without excess cold cholesterol. Bottom: background binding from the “no protein” sample is subtracted to give specific binding. No line fit is imposed.

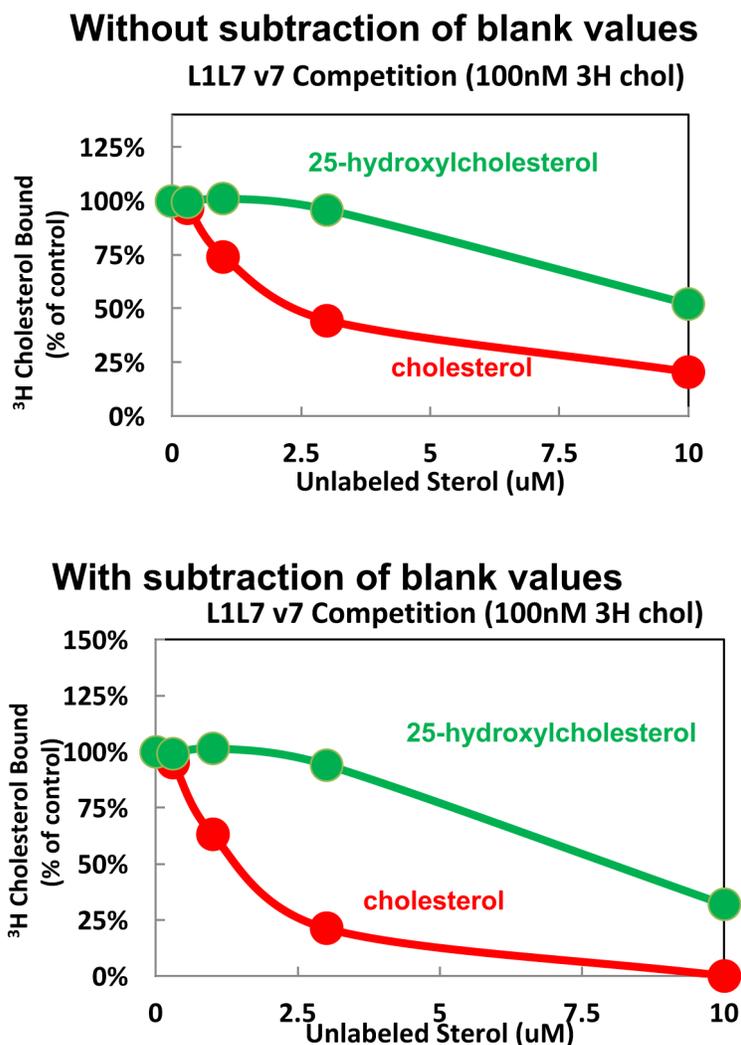


Figure 8. Competition binding of optimized L1L7 (aka v7). Top: Samples were incubated with 100nM tritiated cholesterol with increasing amounts of cold cholesterol or 25-hydroxycholesterol. Bottom: background binding is subtracted to give specific binding. No line fit is imposed.

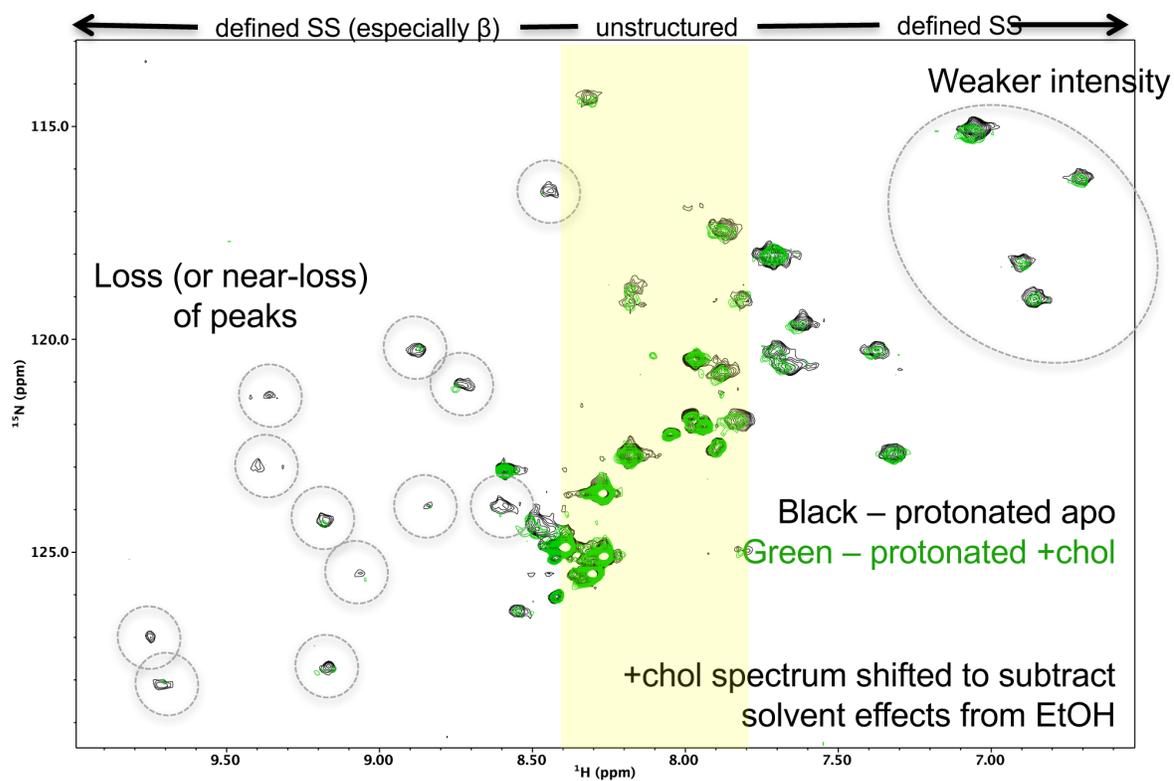


Figure 9. HSQC spectra of  $^{15}\text{N}$ -Leu labeled L1L7 v7 with and without cholesterol incubation. The unstructured region in the proton dimension is approximated in yellow. Slight solvent effects from ethanol were corrected for in the proton dimension.

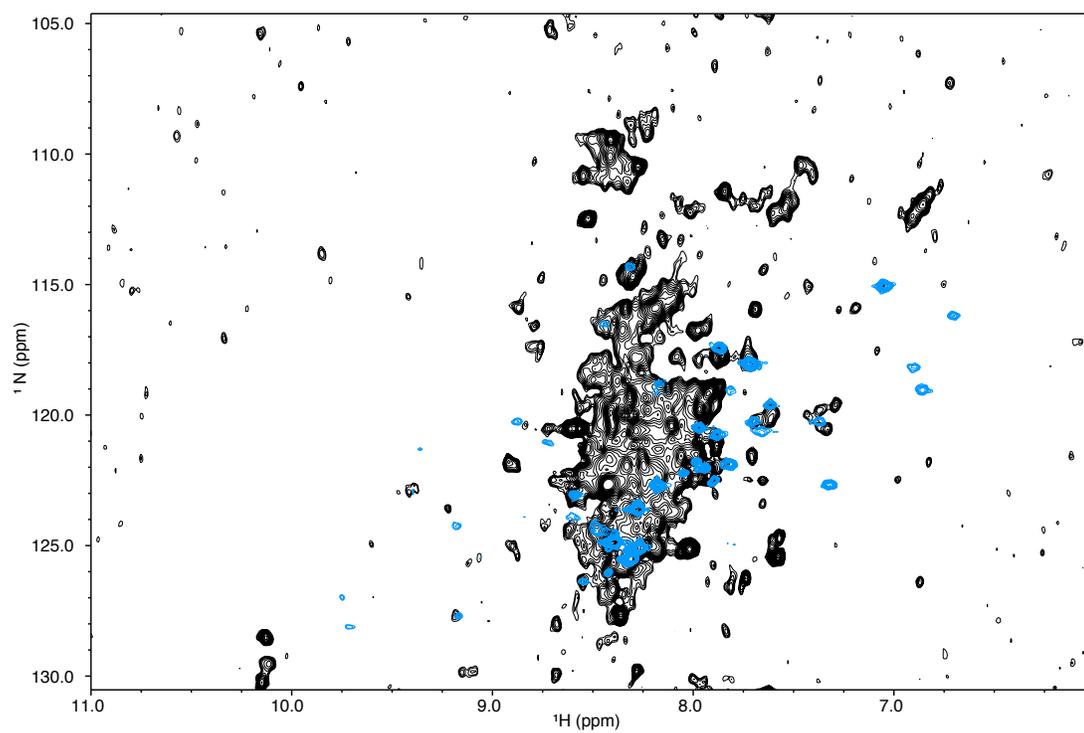


Figure 10. Overlay of L1L7 v7 NMR spectra. Uniform  $^{15}\text{N}$  labeled protein is shown in black, and  $^{15}\text{N}$ -Leu labeled protein is overlaid in blue.

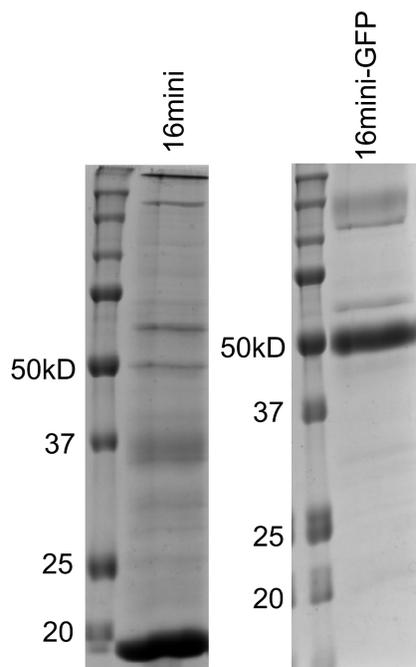


Figure 11. SDS-PAGE gels of purified 16mini constructs, purified in DDM. The strongest bands correlate to the expected MW for each construct.

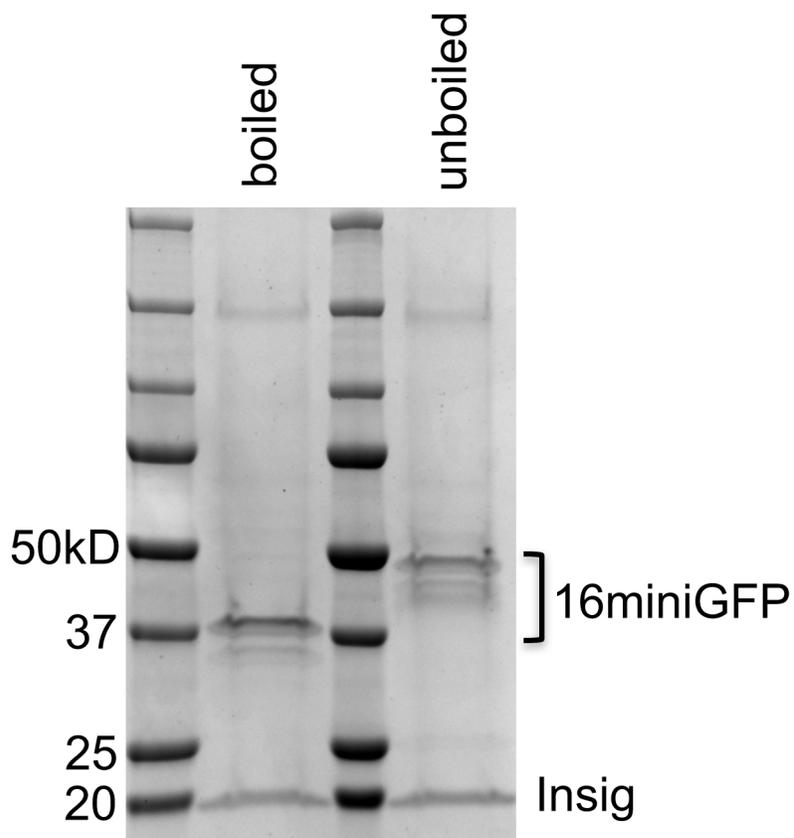


Figure 12. SDS-PAGE gel of M2-purified 16miniGFP co-expressed with untagged Insig2. Following M2 pulldown, Insig can be visualized as a band on the gel running at 20kD. This is the first time that this minimal construct of Scap has been used to pull down Insig.

## CHAPTER SEVEN

### Characterization of Full-length Scap, Insig, and the Scap/Insig Complex

#### INTRODUCTION

The previous chapter discussed experiments carried out focusing on individual domains of Scap, following a strong precedent in structural biology using the “divide-and-conquer” approach. As concluded above, we were not able to obtain large amounts of stable protein amenable to structure determination using this strategy. As an alternative approach, guided by promising initial experiments using the BacMam expression system, we moved towards focusing on the full-length construct of Scap. Full-length Scap was previously never purified, and the yield, purity, and homogeneity displayed great potential to achieve our long-term goals on this project. Subsequently, I was able to improve upon this further by including single point mutations in Scap, stemming from previous *in vivo* studies from the Brown and Goldstein lab as well as *de novo* mutations discovered from an FSEC screen. Building upon this breakthrough in protein biochemistry, I was able to co-purify Scap with its ER binding partner, Insig, and demonstrate the first purification of this important complex. This chapter will touch on some of the earlier efforts that provided clues and pointed us in this direction, as well as cover the experiments and practical considerations taken to bring us to our current state.

## SCAP ORTHOLOGUE SCREEN

Scap was initially discovered and cloned from a line of Chinese Hamster Ovary (CHO) cells (Hua *et al.*, 1996; Brown *et al.*, 2018). Due to this, functional and biochemical experiments have traditionally been carried out using the Hamster form of Scap, which bears a high level of homology (>90% identity) to other mammalian forms of Scap including mouse and human. Many membrane protein structures have emerged using orthologues from related species that can provide physiologically relevant but often more biochemically-tractable versions of the desired target protein. This section will discuss my efforts to follow this precedent, first focusing on a more distant relative and following up with closer vertebrate orthologs.

### **Thermophilic Scap**

Scap is highly conserved across eukaryotes. By comparing genome sequences from mammals down to thermophilic eukaryotes, Scap maintains a ~25-30% homology in its TM region. While the sequence divergence is significant, obtaining a structure of Scap from a thermophilic organism would substantially improve our understanding of how this protein exists and functions. Thermophilic fungi are often used due to the fact they are 1) eukaryotic organisms and can still utilize chaperones and post-translational modifications as required, and 2) are naturally able to grow at higher temperatures, imparting a native level of

thermostability within the amino acid sequence that is likely to improve biochemical tractability. Multiple membrane structures have been solved from thermophilic orthologs (Rheinberger, *et al.*, 2018; Noll *et al.*, 2017), providing some precedence for this strategy.

Through sequence searching, we identified two thermophilic fungi with 25-30% homology to the TM region of Scap, *Chaetomium thermophilum* and *Myceliophthora thermophile*. Both of these fungi had their genomes published in 2011 (Amlacher *et al.*, 2011, Berka *et al.*, 2011), and are cultured at 45°C in either potato dextrose media or rabbit food media, both prepared in my personal apartment's kitchen and then sterilized in the lab prior to culturing. Once enough biomass was generated, we prepared cDNA libraries from each organism and were able to clone out the gene for thermophilic Scap<sup>TM</sup> (abbreviated CScap<sup>TM</sup> or MScap<sup>TM</sup>). These constructs were subcloned into vectors for expression in yeast (*Pichia pastoris*) and insect cell systems, and while they were both able to express in both systems, Sf9 cells gave a higher yield and were focused on for future experiments described below.

Thermophilic Scap (CScap<sup>TM</sup> and MScap<sup>TM</sup>) purified from Sf9 cells were the focus of purifications for approximately two years of my PhD. All thermophilic Scap constructs relied on a His tag for purification by Ni-affinity chromatography, and were able to be purified from Sf9 cells. Purification was, as expected, very efficient in the detergent FC13, but extraction was possible to a lower degree in other detergents

such as DDM, NP40, or a mixture of DDM and CHAPS. We attempted to exchange the primary detergent from FC13 to a more gentle, nonionic detergent during chromatography steps, but this was generally not tolerated. In an attempt to increase the stability of Scap as well as provide rigid globular domains, we inserted the fusion protein T4L into loop 7, at the N-terminus of the protein, or both. The rationale for inserting into loop 7 is that the thermophilic version of loop 7 was much less conserved than the rest of the protein, and therefore could be replaced, similar to intracellular loop 3 in GPCRs (Rosenbaum *et al.*, 2007). CScap<sup>TM</sup> with T4L at the N-terminus as well as in loop 7 was the highest expressing and best-behaved construct, and was abbreviated in lab at “TCT”. Like other constructs, TCT was efficiently extracted in FC13, but it required multiple rounds of injection on SEC to isolate a pure, homogenous species. Proteolysis plagued this purification and a C-terminal fragment (likely cutting at or near loop 7) was routinely co-purified due to the C-terminal affinity tag. In addition, there were varying (but always significant) levels of soluble and insoluble oligomer formation which is not ideal for structure determination.

Nevertheless, we moved forward with our best sample of TCT, purified in FC13. Given the problematic nature of FC13, we tried multiple avenues to reduce the probability of detergent artifacts. We were able to exchange FC13 into amphipols, which are popular synthetic polymers used in place of detergents, typically for cryo-EM purposes. Alternatively, we can inject TCT onto gel filtration in

purely aqueous buffer (NaCl and HEPES only), and retain almost full recovery of homogenous protein. This would, in theory, reduce the background detergent in the sample. Through collaboration with Qiu-Xing Jiang (then a faculty member at UTSW), we examined our samples with cryo-EM on our original JEOL microscope (prior to the Krios and Talos installations). While contrast was very low, we were able to determine that the FC13 sample looked amenable to structure determination, and we were able to secure time on the Krios at Janelia Farm. From that dataset, I was able to process the images with help from two postdocs in the Jiang lab as well as from Qiu-Xing. The end result of our analysis was that the signal was being dominated by the presence of FC13 micelles, and we were not able to resolve any structural features of TCT. Followup analyses of detergent only grids (at a very low level to mimic any background detergent), showed very similar looking objects as the original images, leading us to conclude that our sample was not going to lead to a structure (see Fig. 1). In collaboration with Yiorgo Skiniotis and his postdoc Alpay Seven (then at University of Michigan, now at Stanford), we collected initial negative stain images of amphipol-reconstituted TCT which provided promising looking class averages (see Fig. 2). However, once we took the sample to cryo-EM, we were not able to resolve particles well enough to pursue a dataset due to low contrast and insufficient instrumentation at that time (see Fig. 3 for an example image).

Given the presence of two copies of a known crystallization chaperone, we worked towards optimizing TCT for crystallography as an alternative to cryo-EM. We

utilized both classical vapor diffusion and lipid mediated crystallography setups for TCT. FC13 is known to be a nonideal detergent for crystallography, specifically lipidic cubic phase (LCP) and bicelle crystallization, and was likely the deterrent for crystal formation. Vapor diffusion yielded mainly detergent crystals, again due to the interference from FC13. Results from all these experiments above had the end result of us focusing our future efforts on a form of Scap from a higher eukaryote.

### **Scap from higher eukaroytes**

While thermophilic Scap was a dead end, we wanted to further explore the feasibility of using a higher eukaryote, since experiments with Hamster Scap had reached a bit of a stall at this point in time. Membrane protein structures have been solved by other eukaryotic species besides human (or hamster), and have provided a significant amount of information on each system. MP structures can be solved from chicken (Hite *et al.*, 2015), cockroach (Shen *et al.*, 2017), electric eel (Yan *et al.*, 2017), shark (Shinoda *et al.*, 2009), rat (Martin *et al.*, 2017), mouse (Aller *et al.*, 2009), and zebrafish (Yuan *et al.*, 2012), among many others. Based on sequence conservation to the original Hamster form, we decided to pursue forms of Scap from human (92%), zebrafish (68%), python (72%), Chinese alligator (75%), *Xenopus* (69%), and chicken (76%), with the percent identity across the full-length protein shown in parentheses.

Each of these genes were cloned into the pEZT vector containing a single mutation stemming from a thermostability screen (see below), and followed by a C-terminal 3x-FLAG tag. The species-matched versions of Insig2 were also cloned for subsequent complex formation. I generated virus for each of these clones (Scap and Insig), and purified the Scap/Insig complex for each species from GnT1<sup>-</sup> cells. At the time of this writing, the chicken ortholog has not been cloned or tested, but the others have. The summary of those results is in Fig. 4. The Human ScapFL 2M + Human Insig 2 is the best candidate of the orthologues tested, but displayed a higher void peak and more aggregation compared to the Hamster sample. The other tested orthologues (Xenopus, Zebrafish, Python, Gator) all contained only the D428V mutation (“1M”), but exhibited worse behavior compared to Hamster and will likely not improve upon further optimization. Based on these experiments, we have chosen to focus on the matched Hamster constructs.

### **CHARACTERIZATION OF FULL-LENGTH SCAP**

Full-length (FL) Hamster Scap has been a long sought-after target in the Molecular Genetics department at UTSW. In addition to experiments on Scap<sup>TM</sup> in Sf9 cells described earlier, multiple human and mammalian stable cell lines were generated with the goal of providing a more native-like membrane to increase Scap expression and biochemical behavior. The throughput from these stable lines was

very low – a 9-10mL pellet of cells would be the result of growing and scraping about 250 dishes worth of cells. Efforts to make stable lines in suspension were not as successful. Overall expression of FL Scap from these cells was still quite low, but the striking difference was that Scap expressed in this manner was now soluble in nonionic detergents such as DDM or DMNG. Around this time, we were preparing to pilot expression in the Bacmam system, which was the major breakthrough in this project. This section describes the major discoveries and obstacles overcome to generate the first purification of full-length Hamster Scap.

### **Thermostability screen**

In order to maintain maximum physiological relevance, it is important to study proteins that are as native as possible to wild-type. Of course, this needs to be balanced with the feasibility of such studies, as many wild-type eukaryotic membrane proteins are low expressing and unstable. A common approach to this conundrum is to search for individual point mutations in the amino acid sequence that confer an increased level of stability. Any positive (or negative) effect can be assayed for with incubation at increasing temperatures, and monitoring the resulting chromatography trace on an HPLC that can detect fluorescence at a tunable wavelength. This technique, fluorescence-detected size exclusion chromatography (FSEC), has recently gained popularity because it can be used to rapidly screen

different constructs, detergents, or mutations in a high throughput fashion (Goehring *et al.*, 2014).

As described previously, TMs 2-6 in Scap comprise the sterol sensing domain (SSD). We wanted to use a minimal construct of Scap that included this domain due to its importance, and first generated a construct that contained the native sequence of Scap up until loop 6, where it was truncated. The expression and biochemical behavior of this construct is poor, possibly due to the presence of loop 1 not having a putative binding partner. This prompted the generation of the construct denoted ScapTM16mini, which is truncated after the MELADL sequence in loop 6, and has loop 1 replaced by a small linker. This construct also contained a GFP fused to the C-terminus of the protein, for ease of detection and also for use in FSEC.

We sought to use the ScapTM16mini as a background construct to conduct a thermostability screen to identify potential mutations that would stabilize the SSD, and should, in theory, translate to stabilization of the FL construct. We focused on mutating residues comprising the TM regions of the protein, as this has been a successful approach with GPCRs (Strege *et al.*, 2017; Shao *et al.*, 2016). We identified 14 polar (serine or threonine) residues within the TMs of the SSD, and also included four additional “functional” residues. Each of these four residues has been identified previously by *in vivo* experiments, and mutation of any of them affects the function of Scap and its ability to transport SREBP in a regulated fashion. All of these 18 residues were initially mutated to alanine, and are all mapped on a model

of the SSD in Fig. 5. Each of these mutations was also cloned into the Scap<sup>TM</sup>16 construct, but displayed very low levels of expression in HEK293 cells, and generated a very messy FSEC trace.

Each mutant was cloned into the pEZT vector and purified DNA was used for transient transfection into 10cm dishes of HEK293 cells that were seeded with  $1.25 \times 10^6$  cells one day prior to transfection. The protocol for each transient transfection is as follows, using Lipofectamine 3000 as the transfection reagent and Optimem as the culture media. The final volume of transfection mix per dish is 500  $\mu\text{L}$ . First, for each transfection (2 x 10cm plates), 31.5  $\mu\text{g}$  of DNA is incubated with 63  $\mu\text{L}$  of P3000 reagent (2  $\mu\text{L}$  per  $\mu\text{g}$  of DNA), in Optimem to a final volume of 562.5  $\mu\text{L}$ . Then a master mix of Optimem media and Lipofectamine 3000 is made and incubated for 5 minutes (3  $\mu\text{L}$  of Lipofectamine is needed for every 1  $\mu\text{g}$  of DNA in a final volume of 562.5  $\mu\text{L}$  per condition, allowing ~10% extra for pipetting). Add 562.5  $\mu\text{L}$  LF3K master mix to each tube (1.125 mL final volume), and incubate for 15 minutes. During this incubation step, change the media on each of the plates for 10 mL of fresh media). After incubation, add 500  $\mu\text{L}$  of transfection mixture per plate to two plates per construct. Each dish has 14  $\mu\text{g}$  DNA, 28  $\mu\text{L}$  P3000, and 42  $\mu\text{L}$  of Lipofectamine 3000. Incubate for ~44 hours at 37°C with 5% CO<sub>2</sub> and look for GFP-positive cells with a fluorescence microscope. To harvest cells, scrape with a sterile cell scraper and transfer to a conical tube. Wash plate once with PBS, and add to the cells. Spin cells for 10 minutes at 2000 rpm, carefully decant the media, and then

wash with PBS and re-pellet. Both plates of cells are combined (i.e. one tube per transfection), and pellets can be stored at  $-80^{\circ}\text{C}$  until needed.

The cell pellet for each transfection is resuspended in 1.5 mL of solubilization buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 2% w/v DDM, 0.5 mM TCEP) and rotated at  $4^{\circ}\text{C}$  for 45 minutes to solubilize membranes. To separate out insoluble material, each mixture is spun at 47,000 rpm for 40 minutes. The supernatant contains solubilized membranes and then split into 10 identical PCR tubes for heating. The temperature range used for this thermostability screen is from  $25^{\circ}$  to  $80^{\circ}\text{C}$ , using the gradient program on the PCR block. After heating for 10 minutes, each tube is spun again at 47,000 rpm for 20 minutes to pellet any precipitated material. The supernatant is then ready for injection on the FSEC system. When testing different mutants, it's important to toggle the steps so that the resultant material is treated the same and isn't sitting around for extended periods of time. Typically, when the last spin is started, the next cell pellet can begin solubilization. In addition, injecting the highest temperature sample from each set first will provide the most reproducible results. The injection volume was 20  $\mu\text{L}$  per sample, using a Shodex KW-803 column equilibrated and run in 150mM NaCl, 50mM HEPES 7.5, 0.1% w/v DDM, 0.5mM TCEP. As the temperature increases, the amount of soluble material decreases. Quantification of the amount of material leftover compared to the lowest temperature sample ( $25^{\circ}\text{C}$ ) gives us the  $T_m$  of each sample. ScapTM16 did not give a clear peak, preventing quantification of a melting temperature. From

this point forward, all experiments are referring to mutations carried out in the ScapTM16mini construct.

This protocol allows rapid screening of mutations and reproducible measurement of  $T_m$  values. The beauty of FSEC is that this screening can be performed without purifying the protein – crude solubilized membranes are sufficient. Since the detector can be set to only observe GFP fluorescence ( $\lambda_{exc}$ , 483 nm;  $\lambda_{em}$ , 510 nm), solubilized ScapTM16mini is the only species detected, and runs as a distinct peak with an elution time of about 8.5 minutes (see Fig. 6). By normalizing the height of each peak as a function of temperature, the  $T_m$  can be calculated by finding the temperature where half of the signal is remaining. The  $T_m$  of WT ScapTM16mini is  $\sim 38^\circ\text{C}$ , and each melting curve was highly reproducible. The summary of each mutant and its corresponding  $T_m$  is listed and a scatterplot of each  $T_m$  with respect to WT is shown in Figure 7.

Most of the mutations tested had no thermostabilizing effect, but two mutations immediately jumped out as increasing the  $T_m$ : S427A, a “polar” mutation, and D428A, a “functional” mutation. This result was especially striking because these two mutations are adjacent to each other in sequence, located in the middle of TM6. In our model, these mutations are facing the core of the protein, and it is logical that replacing a polar or acidic side chain with a small hydrophobic side chain would increase the overall stability of the protein. See Fig. 5 for a mapping of these residues on our working model of the SSD.

The most stabilizing mutation of the two best candidates is D428A, which was originally identified by the Brown and Goldstein lab in 2005 (Feramisco *et al.*, 2005). The presence of an acidic residue is a rare feature to see in the middle of a transmembrane helix where most amino acids are hydrophobic. However, this residue is widely conserved across species, even down to bread mold, and is also present in some (but not all of the related SSD-containing proteins. The 2005 study characterized this mutation in Scap using a CHO cell line that is deficient in Scap (Feramisco *et al.*, 2005). The D428A mutation causes Scap to be tightly bound to Insig in cells, when the cells are depleted of cholesterol and Scap would normally be transporting SREBP out of the ER. This effect has been shown with cell-based assays monitoring SREBP trafficking (see Feramisco *et al.*, 2005), monitoring the difference in processed (and thereby trafficked) versus retained forms of SREBP. The readout for this assay is with a western blot using an antibody against SREBP. Through cell fractionation, the nuclear (N), processed form of SREBP is a smaller species compared to the pellet (P) form still in its membrane-bound state. Cells are depleted of sterols through the addition of  $\beta$ -cyclodextrin, and then can have various sterols added back to the cells prior to analysis. Oxysterols are the sterols added back to cells in this case, and promote Insig binding to Scap. Expression of wild-type Scap produces a normal trafficking pattern, where SREBP is trafficked to the nucleus when there is no oxysterol added. When cells have oxysterol added back, we see a significant decrease in the nuclear form of SREBP as a result of retention

of Scap in the ER through interactions with Insig. This normal regulation pattern is lost when we mutate D428 to an alanine – now all of SREBP, and Scap by extension, is retained in the ER. To probe this further, Feramisco *et al.* mutated the aspartate residue to a glutamate to determine if this loss of regulation is due to the negative charge or another factor like side chain size. D428E produced a wild-type like behavior and normal sterol regulation, suggesting that this effect primarily relied on the charge. Furthermore, if D428 was mutated to a neutral or positively charged residue, we have again lost the normal trafficking ability of Scap, ultimately suggesting that Scap requires an acidic residue at the 428 position to be able to dissociate from Insig and restore sterol regulated transport of SREBP.

Other lines of evidence support the importance of D428 in normal sterol regulation. Through pulldown assays using immobilized Insig as bait, an increased level of Scap is pulled down with the D428A construct compared to WT. This effect is also seen in large-scale purifications of the complex, discussed later in this thesis. In addition, D428A causes a conformational change in Scap that can be monitored by a trypsin cleavage assay targeting loop 6 that contains the COPII-binding motif, MELADL (Brown *et al.*, 2002). There are two trypsin cleavage sites within loop 6, and the accessibility of these cleavage sites can be modulated by the addition of sterol to Scap-containing membranes. Compared to WT, Scap with the D428A mutation significantly increased the amount of Scap in an Insig-binding, ER-retained conformation, mimicking a cholesterol-bound form (Feramisco *et al.*, 2005).

Residue 428, as well as the serine at position 427, are in TM6 (see Fig. 5).

TM6 is directly connected to loop 6, so these mutations within this helix could be removing some of the inherent dynamics within Scap, explaining the conformational bias seen in the trypsin cleavage assay. Given the conservation of an acidic residue at position 428, a rare feature in transmembrane helices, we hypothesize that the presence of this residue is related to the mechanism of Scap. It is possible that Scap has a level of built-in destabilization by having this aspartate in TM6, acting as a sort of “fulcrum” within the protein and allowing Scap to sample multiple conformations within loop 6 to be able to bind to COPII when necessary. Removal of that charge by mutation to alanine is abrogating these inherent dynamics, and in theory should be able to form a more homogenous conformation. In addition, the presence of these multiple conformations is a likely explanation for some of the difficulties we and other labs have had expressing and purifying WT Scap over the years.

To further probe mutation at this site with the thermostability assay, I mutated the aspartate to other types of residues. The underlying idea behind this is that we could possibly get an even stronger stabilizing effect with a larger hydrophobic residue than simply alanine. As shown in Figure 8, mutation to asparagine didn't garner that much of an effect, acting as a good negative control for the importance of the negative charge. However, mutation to valine and leucine were more stabilizing than alanine. While leucine had the largest stabilizing effect, the total expression of D428L in HEK293 cells was significantly lower than D428V or D428A, suggesting

that the size of the pocket we can “build” into is not very large. Based on this additional mutational analysis, we focused on D428V as our best candidate mutation. When combined in one construct, both D428V and S427A had additive stabilizing effects. Figure 9 shows the full  $T_m$  curve for WT Scap<sup>TM16mini</sup>, each of the individual mutations, as well as the double mutation (abbreviated as “2M”).

I incorporated the 2M mutations into full-length Scap for purification (see below) as well as in TK-driven plasmids for *in vivo* functional assays. Since S427A is a novel mutation untested in SREBP trafficking assays, we wanted to understand if it had an effect on Scap’s function. Figure 10 contains the western blots from these experiments, which vary slightly from the assay previously described. In the case of WT Scap, when there is no Insig expression, SREBP is trafficked by Scap with or without sterol added. However, when Insig is expressed, we now have sterol regulation where without cholesterol we see trafficking of SREBP compared to ER retention when cholesterol is added. With our D428V mutation in Scap, this regulation is quite different – everything is retained in the ER with or without cholesterol, and with or without Insig expression. D428V prevents trafficking of SREBP to the nucleus, which is also the case with the previously published D428A mutation. S427A had an intermediate effect on SREBP trafficking. We see no sterol regulation of trafficking without Insig added, similar to with WT Scap. In contrast, if Insig is being expressed, now Scap and SREBP are retained in the ER, similar to with D428V Scap. When both mutants are present (2M), everything is again retained

in the ER, demonstrating that the 2M are strongly biasing Scap to be in a conformation that is retained in the ER, bound to Insig.

In addition to D428, there are two other acidic amino acids in the TM helices of the SSD: E359 and E399. Both glutamates are also conserved in mammalian Scap and in the SSD of the protein NPC1 (albeit one of the glutamates is an aspartate in NPC1). Mutation of either of these glutamates to alanine in Scap<sup>TM16mini</sup> conferred additional thermostability when tested in the 2M background. This increase in thermostability was also additive with the previous 2M construct, so we initially included them to create “3M” (2M + E359A) and “4M” (2M + E359A + E399A). Since these glutamate residues are conserved in NPC1, we can further analyze how they fit into the TM region. NPC1 has two bundles of TM helices, one composed of the SSD, formed in a pseudosymmetrical intramolecular dimer. As shown in Fig. 11, the residues analogous to the Scap glutamates are facing away from the core of the SSD and towards the other TM bundle. If we extrapolate this observation onto our model of Scap, this would suggest that these two glutamate residues are also facing out, possibly towards a potential binding partner.

2M, 3M, and 4M mutations were incorporated into the full-length Scap protein and purified with or without co-expression of Insig (see next section). We initially focused on the 4M construct due to the highest level of thermostability, but we were having problems with aggregation and a very low efficiency pulldown of Insig. We

hypothesized that this instability and poor pulldown could be due to the loss of a native protein-protein interface with Insig. I tested this in a small scale experiment comparing the ability of 1M (D428V only), 2M, 3M and 4M Scap FL to pull down Insig after M2 affinity chromatography. Eluate fractions after M2 are shown in Fig. 12. Indeed, it appears that while 1M and 2M are able to pull down a stoichiometric amount of Insig during purification, the 3M and 4M versions are significantly hampered in this ability. A saturating amount of 27-hydroxysterol, which should further drive the Scap/Insig interaction, is maintained throughout the purification. This suggests that our hypothesis is correct, and these two glutamate mutations are possibly at the interface with Insig, and we decided to focus on the 2M construct for future expressions and purifications.

Since these E359A and E399A mutants have been previously uncharacterized, we followed up with cell-based assays to understand what effect they have on SREBP trafficking. The results are shown in Fig. 13, as western blots with increasing amounts of cholesterol added back to sterol-depleted membranes. With WT Scap, we see that the amount of nuclear SREBP detected is “shut off” after 10 $\mu$ M cholesterol. With either of the two glutamates mutated to alanine, it takes a higher amount of cholesterol to prevent SREBP trafficking, suggesting that the ability of Scap to remain anchored to Insig in the ER membrane is affected. The double mutant was not able to be expressed, and was therefore not tested with this assay.

In addition to using FSEC to test for the mutants described above, I also used it to screen for an increase in stability with three pairs of designed disulfide bonds. Using our model and the published NPC1 structure as a guide, I incorporated the cysteine pairs M329C/D428C, V396C/D428C, and G357C/D428C into the ScapTM16mini construct and tested their thermostability (abbreviated CC1, CC2, and CC3). All three pairs increased the thermostability compared to WT, but it was unclear if that was due to simple replacement of D428 with a cysteine or if an actual disulfide was being formed in the protein. This could be tested with observing the effects of D428C alone. However, the 2M combination seemed to be the most promising, especially considering the culmination of functional assays performed with this mutant. I also used FSEC to test for a change in  $T_m$  in ScapTM16mini purified from sterol-depleted (i.e.  $\beta$ -cyclodextrin treated) cells, or from cells treated with saturating amounts of cholesterol. Both of these conditions did not change the  $T_m$  of WT ScapTM16mini. And lastly, given the possibility that the D428 could be coordinating a calcium ion as seen in other proteins, I performed the thermal melt and FSEC runs in the presence of 1mM calcium. It also did not increase the  $T_m$  of WT ScapTM16mini. All of these conditions are included in the summary table in Figure 7.

## **Virus production and GnT1<sup>-</sup> expression system**

Previous studies on Scap have generally focused on expression and purification from the Sf9 expression system using pFastBac plasmids. As noted above, these studies were hampered by low yields and poor biochemical behavior. As an alternative, we were inspired to move towards a mammalian expression system, motivated by recent successes in MP expression and purification on our campus (Morales-Perez *et al.*, 2016a, Morales-Perez *et al.*, 2016b). Specifically, we utilized the pEZT-BM expression vector developed by Ryan Hibbs' lab, which is based on the pEG vector utilized by Eric Gouaux's lab (Goehring *et al.*, 2014). The general idea behind the BacMam system is that viruses can be propagated and titered in Sf9 cells for infection of HEK cells. The pEZT vector was originally derived from a pFastBac vector, but contains a CMV-driven promoter instead of a polyhedrin promoter, and contains an additional GFP in the reverse orientation under control of the insect cell p10 promoter. The addition of GFP under the insect cell promoter provides a rapid way to titer virus without complicated and time-consuming plaque assays. Producing high titer virus is not as critical in insect cells expression as the baculovirus can replicate during the course of the infection, further amplifying the level of total protein expression.

Bacmam virus propagation takes roughly two weeks from DNA to large scale P2 virus. The initial DNA is transformed into the DH10Bac *E. coli* strain that contains additional plasmids containing virus components. These cells are transformed with a

traditional heat shock protocol, but with a longer recovery period (4-5hrs). The resulting cells are then plated and “Bacmid” blue/white agar selection plates containing the antibiotics kanamycin, gentamicin, and tetracycline to maintain all plasmids, in addition to 100 $\mu$ g/mL Blu-gal and 40 $\mu$ g/mL IPTG. Plates are incubated in the dark at 37°C for ~48hrs to fully grow out and digest plate sugars, and colonies with appropriately recombined bacmid are white, while blue colonies should not be picked as they have not recombined properly. Typically two (or more) white bacmid colonies are grown overnight in LB + all three selection antibiotics and are prepped in a similar fashion as a typical mini-prep with a final step of isopropanol precipitation and final wash with 70% ethanol. Bacmid pellets are dried in a sterile hood and resuspended in a minimal volume of EB buffer (Qiagen). Around 8 $\mu$ L of each bacmid is incubated with 8 $\mu$ L of Cellfectin II in 200 $\mu$ L of Grace’s media for 30 minutes, and then diluted to a 1mL volume to infect  $\sim 1e^6$  Sf9 cells, pre-attached to a cell culture plate. This initial transfection is incubated for 4-5 hours at 27°C, and then the 1mL is replaced with 2-3mL of fresh Sf9 media and incubated for 4-5 days at 27°C. For pEZT viruses, it was crucial to use the Invitrogen Sf900 cells instead of the Expression Systems ESF921 cells. Although the reasons are unknown, the titers were consistently and reproducibly higher in Sf900 cells across multiple constructs. After 3 days, the cells should begin to become GFP-positive, with a full infection around 5 days. At this point, the 2-3mL of media should be carefully and sterilely removed, and transferred to a separate tube. Spin the tube for 1500rpm for 10min,

and then carefully remove the supernatant and transfer to a fresh conical tube. This tube now contains the P0 virus. Viruses should all be stored at 4°C, away from light. The amount of P0 added to start a P1 is typically based on how GFP positive the cells were from the initial transfection, but an appropriate range is ~1-2mL for every 50mL of cells (at a density of  $1e^6$  cells/mL). The P1 should shake at 27°C for about 2 days, where the cells again should start to become GFP positive, generally becoming 100% positive after 3 days. The P1 is then spun down again at 1500rpm for 10min, followed by sterile filtering in the hood. Similar to the P1 step, the amount of P1 added to create a P2 should reflect how efficient the P1 appeared, but a typical range is 3-7mL per 1L of cells (at  $1e^6$  cells / mL). The P2 cells should be monitored for GFP fluorescence after 2 days, and should be harvested at the timepoint of maximum GFP positive cells, or after 3 days. The P2 should be spun down and sterile filtered in the hood, and tested to evaluate each viruses' titer.

Virus titering should be performed for every unique P2, even if produced from the same P1 as it is common to get variability between batches. Titering is performed in 96-well black plates, with 8 different virus dilutions and in triplicate. Cells used for the titer should be at a density of  $0.75e^6$  cells/mL, and typically 10mL is needed per plate which can titer four viruses. 100 $\mu$ L of cells should be plated in each well and allowed to attach. Serial dilutions of each virus should be made, covering a  $10^{-1}$  to  $10^{-8}$  range of dilutions (e.g. 360 $\mu$ L of media + 40 $\mu$ L of virus, mix, then take 40 $\mu$ L of that virus to dilute into the next tube). After making serial dilutions,

carefully remove media in each well and replace with 100 $\mu$ L of media containing the viruses, from row A (as  $10^{-1}$ ) to H (as  $10^{-8}$ ), and in triplicate (columns 1-3 contain the first virus, 4-6 contain the second, etc). The titer plate should be incubated at 27°C for 48hrs and then read by carefully counting the number of GFP-positive cells at each dilution. Typically this is done on the row that contains less than 10 positive cells for accuracy. To calculate the titer, take the average number of GFP positive cells multiplied by the dilution factor (e.g.  $10^{-6}$ ), multiplied by 10 to account for the initial dilution, and the result is the titer (pfu/mL) of that virus. To calculate the amount of virus needed to add to cells to achieve a specific MOI, multiple the desired MOI (e.g. 1) by the number of cells infecting (e.g.  $3 \times 10^9$  (for a L of cells at  $3e^6$ /mL), and then divide that value by the titer to get the volume of virus to add. For expression in GnT1- cells, it is important to not add more than 10% of the final volume in virus in order to maintain sufficient cell health. It is also common (and helpful) to include a small amount of sodium butyrate to each culture (1-10mM is common; we consistently found 2.5mM to be the best in our hands). Sodium butyrate is a histone deacetylase inhibitor, and addition to cell culture typically increases the amount of total protein expression.

The mammalian cells we used for all of our Bacmam expression were a variant of HEK293S cells mentioned earlier, referred to as GnT1- cells. This cell strain is deficient for the enzyme N-acetylglucosaminyl transferase I, which has the downstream effect of reducing the level of protein glycosylation (Reeves *et al.*,

2002). Limiting protein glycosylation is helpful for structural biology because it increases the likelihood of producing a more uniform sample after purification, and improving the efficiency of the enzymatic deglycosylation step via PNGaseF or other enzymes. GnT1<sup>-</sup> cells are grown and maintained in FreeStyle293 media, supplemented by 2-5% FBS or NCS. Cells typically divide every ~24 hours when shaken at 37°C, and after transduction by Bacmam virus, the temperature is typically reduced to 30°C. Along with MOI and sodium butyrate concentration, length of expression is typically a variable that should be tested for each sample, but I found reproducible levels of expression with an MOI of 1-2, 2.5mM sodium butyrate, and expression for 48-60 hours. After this length of time, cells are spun down for an hour at 4000rpm, and then resuspended in PBS and transferred to conical tubes for subsequent centrifugation. Pellets are stored at -80°C until needed for purification.

### **Addition of N-terminal fusion protein**

As mentioned previously, fusion proteins have emerged as a useful technique in structural biology to reduce flexibility in membrane proteins and act as crystallization chaperones. One of these fusion proteins, abbreviated as PGS (see page XX), is able to form a continuous  $\alpha$ -helix from the C-terminus of PGS into the N-terminus of the target protein. This property facilitated the crystallization and structure determination of the orexin receptors (type 1 and 2) (Yin *et al.*, 2015; Yin *et al.*, 2016), and the cannabinoid-1 receptor (Shao *et al.*, 2016). Initial cryo-EM efforts

(see below) on full-length Scap showed a high level of flexibility within soluble domains as well as the membrane region. This level of flexibility complicated the particle alignment as the signal-to-noise was averaged out and only about 100kD of total molecular weight was resolved in the best, but low-resolution 3D map.

A common way to overcome flexibility within proteins in cryo-EM is to provide a rigid body within the sample that can act as a fiducial marker and improve overall particle alignment. This approach has been taken in negative stain and cryo-EM studies of GPCR-G protein complexes (Shukla *et al.*, 2014; Garcia-Nafria *et al.*, 2018). This can also be accomplished by the addition of conformationally specific antibody fragments or nanobodies (see later section).

We hypothesized that we could stabilize TM1 of Scap by attaching PGS to the N-terminus. Assuming we could find the correct attachment point to lock it in the correct register, we could then capitalize on the ability of PGS to form a continuous  $\alpha$ -helix with its binding partner. This could also provide some rigidity to the complex to improve particle alignment during later image processing steps. To test for the correct linking residue, we cloned attached PGS at various points just prior to the predicted start of TM1 (see Fig. 14, panel A). A Scap-KO CHO cell line was used for transient transfections for each of the PGS-fused variants, and cells were subjected to western blots to assay for expression levels with a Scap-specific antibody (7G5). It was apparent that expression was dropping off if we attached the fusion protein too far into the predicted TM1 (see Fig. 14, panel B), and subsequent experiments

demonstrated that the glycosylation pattern of Scap was altered at a certain point (data not shown), suggesting that we were drastically changing the protein folding or damaging Scap's stability in the ER membrane. Based on these data points, we moved forward with a PGS-fusion construct that attaches at the amino acid sequence HGLL (resulting construct also referred to as QC6).

The QC6 construct expressed at a higher level than constructs without the fusion protein added, and purified with less of a left shoulder by size exclusion chromatography (see Fig. 15). The addition of the fusion protein also provided an additional mass in our 2D and 3D classifications during cryo-EM processing, which can assist with model building and general alignment improvements. See the next chapter on cryo-EM for more details.

### **Purification protocol and optimization**

Full-length Scap can be purified from GnT1<sup>-</sup> cells in a variety of detergents, including DDM, GDN, and the maltose neopentyl glycol (MNG) class of detergents. The detergent that was the most promising is the 12-carbon length version of MNG, called lauryl maltose neopentyl glycol (LMNG). LMNG has a low CMC (0.001%), which makes it nonideal for a purification where the detergent should be later removed (e.g. detergent exchange, reconstitution), but we found that it gave the most reproducible protein behavior and additional considerations were taken during subsequent lipid reconstitution (see later section).

The general protocol for purifying Scap from GnT1<sup>-</sup> is as follows. Cells are thawed from the -80°C in a beaker of water and then subjected to a hypotonic lysis. The lysis buffer (10mM HEPES pH 7.5, 2mM EDTA, 0.5mM TCEP, protease inhibitors) should be kept on ice until use. When the cells are thawed, dounce loose ~10 times in the buffer to break up any cells clumps and let stir at 4°C for 30 minutes. Spin out membranes in 250mL bottles at 12,000rpm for 30 minutes. The cell membranes at this point are quite soft, so the supernatant should be decanted off carefully. The basic solubilization buffer contains 0.5M NaCl, 50mM HEPES pH 7.5, 1% w/v detergent (LMNG), 0.2% w/v cholesterol hemisuccinate (CHS), 0.2% w/v sodium cholate, 20% v/v glycerol, 0.5mM TCEP, and protease inhibitors. This buffer typically requires >6 hours to go into solution, and should be made ahead of time and then chilled prior to use. Dounce membranes with buffer on loose (~10 strokes) and tight (~30 strokes) to fully break up membranes. Add ~40-50µL of nuclease to the membranes to help reduce viscosity. The solubilized membranes are quite viscous and may take a few rounds of douncing to allow for stirring. Once at that point, membranes need 1-1.5 hours of stirring at 4°C for efficient protein extraction. Insoluble material is then removed by ultracentrifugation (typically 35,000rpm for 30 minutes in a Ti45 rotor), and clarified supernatant is then transferred to large 250mL conical tubes. M2 resin (pre-equilibrated in M2 buffer containing 150mM NaCl, 50mM HEPES pH 7.5, 0.05% detergent (LMNG), 0.01% CHS, 0.01% sodium cholate, 10% glycerol, 0.5mM TCEP) is then added to tubes

and allowed to batch bind at 4°C for 4-6 hours. After binding is complete, the tubes are spun at low speed (1500rpm for 15 minutes), and the resin is transferred to a glass gravity column. The resin is washed with approximately 10 column volumes of M2 buffer, with occasional “spot checks” of flow-through to ensure all background, unbound protein is washed away. Scap is eluted from the resin with M2 buffer containing 0.4mg/mL FLAG peptide and generally comes off completely in ~2 column volumes in a sharp peak, especially if eluting at a low flow rate (<1 mL/min). At this point, the M2-purified protein can be flash frozen in LN<sub>2</sub> and stored at -80°C for future experiments. If the purification is continuing, add ~20-30µL of the enzyme Peptide N-Glycosidase F (PNGaseF; from New England Biolabs) is added and the protein is incubated on ice, at 4°C overnight. The next morning, the protein is concentrated to ~1mL (or 2 x 1mL if the yield is very high), and injected on a Superose 6 10/300 GL equilibrated in 150mM NaCl, 20mM HEPES pH 7.5, 0.01% LMNG, 0.002% CHS, 0.002% sodium cholate, 0.5mM TCEP. A typical trace of Scap alone is shown in Fig. 16, panel A, with a void peak eluting around 9mL and a main peak around 15.5mL, with a prominent left shoulder peak. The protein is quite pure at this stage, and thus any heterogeneity seen during SEC is a result of aggregation or oligomerization. For comparison, an overlay of WT vs 2M Scap alone is shown in Fig. 16, panel B.

Through the history of this project, many additives and other solubilization strategies were attempted before pursuing the main protocol described above. One

of the more promising additives was the soy L- $\alpha$ -phosphatidylcholine (soy PC, from Avanti Lipids). Soy PC is a natural lipid mixture from soybeans, and contains predominantly the 18:2 fatty acid composition, although others are present. The impetus for using soy PC as an additive stemmed from the success of recent cryo-EM and X-ray crystallography structures such as NOMPC (Jin *et al.*, 2017), TRPV1 (Gao *et al.*, 2016), and TPC1 (Kintzer and Stroud, 2016). Through parallel purifications with and without soy PC, we determined that it was not necessary and didn't impart an obvious positive benefit.

It is possible that any detergent (including LMNG) could be stripping away native lipids from Scap that are important for its stability. A method to circumvent this step is by using styrene maleic acid polymers (SMA) to solubilize membranes (Lee *et al.*, 2016; Knowles *et al.*, 2009). These polymers are able to form "native nanodiscs" by solubilizing the entirety of the cell's lipid membranes into small regions that enveloped by SMA polymers. This requires no detergent, and can be kept in aqueous buffer through subsequent chromatography and purification steps. I was able to have limited success with this method, but through optimization found that the best buffer for SMA solubilization included 250mM NaCl, 50mM HEPES pH 8.0, 10% glycerol, 2.5% w/v SMA, 100mM arginine, and protease inhibitors. I found that the NaCl concentration, the pH, and the SMA concentration was optimal at these amounts, and through a small homemade additive screen, found that arginine significantly increased the amount of protein solubilized via western blot. Ultimately,

the resulting yield of SMA-solubilized Scap was significantly lower than with detergent, and cryo-EM analyses determined that the size of our particles was not as homogenous as described (Lee *et al.*, 2016). Therefore, we determined that we should continue with the LMNG purification described above.

### CHARACTERIZATION OF FULL-LENGTH INSIG

Cholesterol and oxysterols reduce SREBP processing by promoting the interaction of Scap with the protein Insig. Insig is an ER anchored protein and exists in two isoforms, Insig-1 and Insig-2, that can both bind to Scap and retain the Scap/SREBP complex in the ER (Yabe *et al.*, 2002; Adams *et al.*, 2003). Although this won't be discussed in this dissertation, Insig can also bind HMG CoA reductase, causing it to become ubiquitinated and degraded (Goldstein *et al.*, 2006).

Overexpression of either isoform of Insig promotes the binding and retention of Scap, and induces conformational changes in Scap that mimic the D428A mutated form discussed earlier (Adams *et al.*, 2003). An ultimate goal of this project is to understand how Insig and sterols influence the conformation of Scap, and therefore obtaining structural information on Insig could provide clues into how it achieves this in the ER membrane.

Insig is a small protein, 277 and 225 amino acids for the two human isoforms, respectively, and is predicted to contain size transmembrane helices. Insig-1 is

slightly larger due to a longer N-terminus, but the TM regions are highly identical between the two isoforms. Traditionally, the human forms of Insig were used in cell-based assays in the Brown and Goldstein department, likely due to this high homology. Human Insig-2 can be purified from Sf9 cells with an N-terminal His-tag, but can only be solubilized in the detergent FC13 (Radhakrishnan *et al.*, 2007). Despite this, it could be purified to homogeneity as a monomer, and was shown to be able to bind oxysterol (25-hydroxysterol) saturably and specifically (Radhakrishnan *et al.*, 2007). This sample was never carried forward for structural studies, but eight years later, a crystal structure of the mycobacterial form of Insig was published (Ren *et al.*, 2015). The mycobacterial Insig shares fairly low identity to mammalian Insigs (20-30%), but likely adopt the same fold as the mammalian form. The structure shows Insig as a bundle of six TM helices, arranged in a trimer. The trimer formation was shown to potentially be functionally relevant, as a cross-linked form that forces it to be a trimer displays similar levels of homogeneity (Ren *et al.*, 2015). Currently, there is no evidence that mammalian Insig forms a trimer in cells or in solution.

Initial efforts to purify Insig for structural studies was focused on the Human form of Insig-2. The construct contained an N-terminal single FLAG tag, and was expressed to high levels in GnT1<sup>-</sup> cells. M2-Human Insig2 can be readily purified as a biochemically pure species through batch binding to M2 resin, in the detergents DDM and LMNG. However, the SEC profile was consistently very broad, signaling

that the protein was heterogeneous in size (see Fig. 17). Following up on this finding, I performed multi-angle light scattering (SEC-MALS) with Chad Brautigam at the Macromolecular Biophysics Resource here at UTSW. The results are shown in Figure 18. While the mycobacterial form of Insig formed a consistent trimer, the human form of Insig-2 displayed the properties of a “classical polydisperse polymer”, and had evidence of masses from 20kD (consistent with a monomer) up to ~500kD (20+-mer). Following results studying oligomerization of Insig-2 in CHO cells (Lee and Boyd, 2010), I introduced the G39E mutation into Insig-2 construct to determine if this would bias Insig to form mainly a monomeric species and therefore be less oligomerization-prone, as suggested by their results in pull down experiments from CHO cells. I expressed and purified this mutant form of Insig-2, and the results are shown in 7-17. The gel filtration trace of G39E Insig-2 still showed a high degree of heterogeneity, and the yields mimic the notoriously low-expressing Insig 1.

The main takeaway from this experiment is that any free Insig will be problematic because of its affinity for itself and capability of forming heterogenous assemblies. Due to this (and for the obvious scientific importance), we moved towards forming a Scap/Insig complex, which required optimization and is described in the next section. To avoid any future reviewer questions, we opted to work with Scap and Insig from matched species. In contrast to prior experiments, we ultimately decided upon using the hamster forms of both proteins for structural studies of the complex.

## PURIFICATION OF FULL-LENGTH SCAP/INSIG COMPLEX

Scap is able to sample multiple conformations as evidenced through difference in trypsinization patterns carried out in membranes (Brown *et al.*, 2002). The point mutation D428A biases Scap to adopt a “cholesterol-bound” conformation, and is retained in the ER, bound to Insig (Feramisco *et al.*, 2005). Through the thermostability screen that led us to include the point mutations S427A and D428V to create the “2M” constructs, we have created a specific construct of Scap that should be in the same “cholesterol-bound” form that is primed to bind Insig. Based on the improved biochemistry of the 2M construct compared to WT, it appears that we the instability of Scap could be tied to its flexibility, and thus our 2M constructs have increased our likelihood of obtaining a structure. Furthermore, we have created a construct of Scap that is primed to bind Insig, and in this section I will describe my efforts to produce the first co-purification and reconstitution of these two full-length proteins.

### **Protein biochemistry**

Viruses encoding full-length Hamster Scap 2M (containing a C-terminal 3x-FLAG tag) and full-length Hamster Insig-2 (untagged) were used to transfect GnTI-cells ( $2-3 \times 10^6$  cells/mL) in a ratio of 1:4. Sodium butyrate was added to the media to

2.5mM along with a final concentration of 300nM 25-hydroxycholesterol (added from a 1mg/mL stock in ethanol). After transfection, the temperature is dropped to 30°C and cells are harvested after 60 hours. The purification can be carried out in DDM or LMNG as described above for full-length Scap alone, with 0.5mM TCEP and 600nM 27-hydroxycholesterol present through the entire purification. TCEP proved to be a crucial component of this purification. Experiments without TCEP or with a cysteine-capping step utilizing the alkylating properties of iodoacetamide, commonly used in GPCR purifications, yielded protein with much worse behavior (see Fig. 19).

After the overnight PNGase purification, purified complex can be subjected to ion exchange chromatography (MonoQ) prior to gel filtration. In LMNG and DDM, protein is able to efficiently bind to MonoQ in a background of 150mM NaCl, 50mM HEPES pH 7.5, 0.05% detergent, 0.01% CHS, 0.01% sodium cholate, 0.5mM TCEP. Over a shallow gradient to 40% B (400mM NaCl), a very sharp and distinct peak elutes at ~250-300mM NaCl that contains the complex (see Fig. 20). The initial hope with this ion exchange step was that it would remove some of the underlying heterogeneity prior to SEC and provide a cleaner profile with less of a left “shoulder”. While this step did lessen the size of the void peak and some of the shoulder, it did not completely ablate it. Also, for unknown reasons, the loss of protein at this step did not compensate for any benefit received, and was removed from subsequent purifications.

SEC-purified protein elutes at around 15.5mL, consistent with a 1:1 Scap:Insig complex. To confirm this, we again turned to SEC-MALS that showed a very tight size distribution corresponding to a 1:1 complex. See Fig. 21 for Scap alone and Scap/Insig SEC-MALS traces. The mass range for Scap alone is 156-161kD, and Scap/Insig gave a mass range of 167-185kD. This is the first time this stoichiometry has been demonstrated with purified components, and is a striking result given the ratio of Scap to Insig viruses added to cells. We see no evidence of Scap being able to bind and pull down more than one Insig, even with saturating amounts of Insig being co-expressed.

While the level of purity was consistently clean after the M2 resin (see Fig. 15), some of the SEC traces displayed more heterogeneity than others. This effect seems to be correlated to the scale of the purification, as this requires more concentration of the purified protein prior to injection on gel filtration. In addition, the resolution becomes limited on gel filtration columns when loaded with a high amount of protein. This heterogeneity could be easily cleaned up by a subsequent injection on the same column, and this is particularly useful when studying the complex in detergent versus in a lipid reconstituted system (see below).

The “QC6” version of hamster ScapFL 2M (containing an N-terminal PGS fusion protein) is also able to pull down untagged hamster (and human) Insig-2 from GnT1<sup>-</sup> cells following the same infection and purification protocol as the non-fusion protein construct. As mentioned previously, QC6 hamster ScapFL 2M (or “QC6” for

short) expresses better and is more homogeneous than ScapFL 2M alone, and this effect carries through to purifying it with Insig-2 present. The impetus for including the N-terminal fusion protein was to provide a rigid handle that can be helpful for particle alignment during cryo-EM processing.

#### *Additional point mutations*

To further improve upon the 2M construct, I included other published point mutations that should also help bias Scap to be in the cholesterol-bound mutation and in theory should be compatible with S427A and especially D428V. However, this was not the case. Y234A is a mutation in loop 1 discovered through an alanine scan of the entire loop (Motamed *et al.*, 2011). Y640S is a mutation in loop 7, discovered through an alanine scan of loop 7 (Zhang *et al.*, 2013). Each of these mutations, either by themselves or combined, cause loop 6 to adopt the cholesterol-bound conformation, even when the cells have been depleted of cholesterol. They also cause Scap to be retained in the ER and SREBP trafficking to the Golgi is prevented (Motamed *et al.*, 2011; Zhang *et al.*, 2013).

Our hypothesis is that either of these mutations would help further stabilize Scap by reducing flexibility in loops 1 and 7, with loop 1 (Y234A) providing the highest benefit due to it occurring in the larger of the two loops. I incorporated Y234A into the 2M constructs, with and without the N-terminal fusion proteins. Those results are shown in Fig. 22. In summary, the Y234A mutation significantly reduced

the expression and purification yields of the Scap/Insig complex, with a larger negative effect evident when mutated the QC6 background. Surprisingly, a similar result was found with two additional mutations in loop 1: V98A and C264A (“CCA”). Both of these mutations were described previously in the L1L7 section. V98A arose from the same loop 1 alanine scanning study that identified Y234A (Motamed *et al.*, 2011), and produced the same ER-retained Scap phenotype. The CCA mutation is simply mutating the third cysteine in loop 1 to an alanine. C264 is free and could potentially be forming aberrant disulfide bonds with other free cysteines in our sample. I generated constructs of 2M with and without the QC6 fusion protein with each of these mutations incorporated. Similar to the Y234A mutation, both of these mutations reduced the expression and worsened the biochemical behavior, with a more pronounced effect in the QC6 construct. It is still unclear as to why these mutations impair the stability of 2M, as we hypothesized they would further stabilize the “cholesterol-bound” conformation of Scap, or at the very least be compatible. It is possible that loop 1 is much more sensitive to mutation than expected, and the combination of a mutation in that loop and an N-terminal fusion protein are destabilizing, as evidenced by the strong negative effect when added in QC6 constructs. It is our hope that a structure of the Scap/Insig complex will help us understand the reason(s) why this is the case.

## Sterol binding assays

Through functional assays shown in Fig. 10, our hamster ScapFL 2M construct was locked in an ER-retained form, and our purification shows that it is capable of efficiently pulling down Insig in a 1:1 ratio. In order to fully characterize this construct, we needed to confirm that it was capable of binding cholesterol with specificity. Scap<sup>TM</sup> 1-8 containing the mutation D428A was shown to be able to bind cholesterol as well as wild-type (Feramisco *et al.*, 2005), we expected our full-length 2M construct to also retain this ability. Due to prior difficulties with protein expression and purification, full-length Scap (or the Scap/Insig complex) have never been able to have their binding characteristics studied. The experiments described here and shown in Fig. 23 were carried out by Arun Radhakrishnan, using protein purified no more than 24 hours beforehand.

Each reaction was set up with 2-5  $\mu$ g of Scap or Scap/Insig in each tube, similar to the binding experiments described earlier for L1L7. Each sample was incubated with increasing amounts of tritiated sterols, kept in concentrated ethanol stocks. The binding buffer contained 150 mM NaCl, 20 mM HEPES pH 7.5, and 0.05% DDM. After incubation with sterol at RT, the sample was loaded over  $\sim$ 300  $\mu$ L of equilibrated Ni resin, washed, eluted, and subjected to scintillation counts. Competition experiments were completed with 200 nM <sup>3</sup>H-cholesterol and a range of unlabeled sterols. Experiments were completed in triplicate, and are shown in Fig. 23. In summary, our 2M construct (alone or in complex with human Insig-2), displays

saturating and specific binding behavior, similar to prior experiments carried out with Scap<sup>TM</sup> 1-8 (Radhkarishnan *et al.*, 2007).

In addition, Arun performed experiments on full-length Scap 2M ( $\pm$  human Insig-2) where 2-5  $\mu$ g of sample was incubated overnight with 5  $\mu$ M of various unlabeled (i.e. not tritiated) sterols. The next day, saturation experiments were performed with increasing amounts of  $^3$ H-cholesterol. The results from these experiments are shown in Fig. 24. When pre-incubated with 25-hydroxycholesterol or vehicle, Scap 2M alone is able to still bind cholesterol. However, if the Scap 2M/Insig complex is pre-incubated with 25-hydroxycholesterol, cholesterol is no longer able to bind, suggesting some sort of competition for binding between the two sterols. This result has never been shown before, and our hope is that a structure of Scap and/or the Scap/Insig complex would help us understand if possibly there is a shared binding site or if there are two distinct binding sites with some sort of allosteric mechanism connecting the two. Future experiments are planned that do the converse experiment – incubation overnight with cholesterol, following by saturation binding with  $^3$ H-25-hydroxycholesterol.

### **Reconstitution into lipid nanodiscs**

A major challenge in biophysical studies of membrane proteins is finding an appropriate membrane mimetic that keeps the sample stable while maintaining semblance to the original membrane composition. The most common membrane

mimetic is simply the detergent used throughout the purification, and has been useful in structural determinations of membrane proteins in pure (Parker and Newstead, 2016; Li *et al.*, 2017) or mixed micelles (Schmidt *et al.*, 2016; Liu *et al.*, 2017; Koehl *et al.*, 2018). However, there have been many cases of membrane protein structures containing perturbations from the detergent used (Zhou and Cross, 2013) and the general trend is to move towards a lipid bilayer system that better recapitulates the properties of the native membrane environment. In this section, I will introduce different types of lipid membrane mimetic systems and discuss the optimization and reconstitution of the Scap/Insig complex into both MSP-based and Saposin-based nanodiscs.

### *Background*

The most commonly used lipid-based membrane mimetic is the nanodisc, sometimes also referred to as rHDL particles (Denisov and Sligar, 2016). Nanodiscs were developed by the Sligar lab, and rely on the membrane scaffold protein MSP derived from the apolipoprotein family (Bayburt *et al.*, 2002; Denisov *et al.*, 2004). The size of the nanodisc formed depends on the length of the MSP being used, with reported sizes ranging from 6nm (Hagn *et al.*, 2013) to 50nm (Nasr *et al.*, 2017). Nanodiscs have been utilized for a variety of biophysical methods, ranging from single molecule studies (Ritchie *et al.*, 2011), double electron-electron resonance (DEER) studies (Kang *et al.*, 2015), NMR (Hagn *et al.*, 2013; Kofuku *et al.*, 2014),

SAXS (Bayburt *et al.*, 2006), and of course cryo-EM (Gao *et al.*, 2016; Shen *et al.*, 2016; Taylor *et al.*, 2017). MSP-based nanodiscs are compatible with virtually any type of lipid, including additives such as cholesterol. There are two MSP molecules that wrap around each lipid acyl chain in the bilayer, and based on the size of the MSP chosen, a finite amount of lipids can fit within the disc. Of course, some of those lipids will be displaced by the target protein, and during optimization the ratios between the sample, lipids, and the amount of MSP added are typically varied to determine the optimal quantities for efficient reconstitution. In addition to being one of the pioneers in using nanodiscs for NMR of MPs, Gerhard Wagner's group has further engineered this system to create circularized nanodiscs through a sortase-based ligation step (Nasr *et al.*, 2017). It is likely that further improvements or alterations to MSP-based nanodisc technology will continue to be developed and sustain its popularity as a membrane mimetic.

An alternative lipid-based membrane mimetic system has been recently described using the scaffold protein saposin-A (Frauenfeld *et al.*, 2016; Chien *et al.*, 2017). Saposin proteins are lipid binding proteins present in the lysosome (Olmeda *et al.*, 2013) and multiple copies of this small protein are able to form a "custom" sized nanodisc around the membrane protein and a belt of lipids. Like with MSP-based systems, saposin nanodiscs are suitable for use with major lipid classes and the ratio of sample:lipid:saposin must be optimized for each sample. Saposin has been recently used in the structure of the fungal mitochondrial calcium uniporter

(Nguyen *et al.*, 2018), and is a promising option for membrane proteins that require a membrane mimetic with an atypical geometry or size not tolerated by the standard MSP-based systems.

As mentioned previously, “native” nanodiscs have also been an emerging option for membrane protein stabilization, capitalizing on the properties of styrene maleic acid (SMA) polymers (Knowles *et al.*, 2009; Lee *et al.*, 2016). SMA lipid particles (SMALPs) differ from nanodiscs in that the target membrane proteins are never exposed to detergent and remain in contact with their native lipids from cells to reconstitution. This is an ideal characteristic for many membrane proteins, as they have evolved to exist in a membrane bilayer which contains many different types of lipids typically not added back in classic nanodisc reconstitutions. In addition, membrane proteins often require lipids or other components of that bilayer for their proper function or as a ligand (as in the case of Scap and Insig). Utilizing detergent to solubilize samples from the membrane can often strip away these important lipids, causing downstream problems in purification or subsequent analyses. Although they have become more popular, SMALPs have not been as widely adapted for membrane protein structural biology. Homogeneity seems to be problematic, despite published reports of narrow diameter size distributions based on the type of polymer used (Esmaili and Overduin, 2018). As of now, SMALPs have been used for only a handful of structural efforts in cryo-EM (Parmar *et al.*, 2018; additional personal communication), but have been demonstrated as compatible with reconstitution into

lipidic cubic phase (Broecker *et al.*, 2017). The premise of “native” nanodiscs is intriguing, and with further improvements could become a more popular option in membrane protein structural biology.

#### *Purification of nanodisc scaffold proteins*

Each nanodisc scaffold protein can be overexpressed and readily purified from *E. coli* cells. MSP1D1 and MSP1E3D1 (pET28a) are expressed in BL21(DE3) cells with kanamycin resistance, while saposin (pNIC28) is expressed in Rosetta gami-2 (DE3) cells with kanamycin, tetracycline, and chloramphenicol. All three proteins can have starter cultures in Luria broth (LB) media, but should be grown in terrific broth (TB) for large scale cultures, typically 2 L. After inoculation of the large scale cultures with the starter cultures, cells are grown at 37°C until the OD<sub>600</sub> reaches 0.6-0.8. Induction varies slightly for each sample. For MSPs, expression is induced with 1 mM IPTG and continued to grow at 37°C for 4 more hours. For saposin, expression is induced with 0.4 mM IPTG, temperature is reduced to 20°C and allowed to continue shaking overnight. Cells are harvested by centrifugation, and pellets are stored at -80°C until needed.

Purification for all three scaffold proteins is similar, with the only divergence occurring during the lysis. The buffer volumes listed here are the typical requirements for a 2 L pellet purification. After pellets are thawed, they are dounced in 30-40 mL lysis buffer (50 mM HEPES pH 7.5, 300 nM NaCl, 1% Triton; make 250

mL), with 10  $\mu$ L of Pierce Universal Nuclease and 1 mg/mL lysozyme added. Lysis should continue for 30 minutes by spinning at 4°C. Insoluble material is removed by centrifugation at 10,000 rcf and the supernatant is saved. For saposin samples, an additional heating step is performed at this point. The saposin-containing supernatant is heated to 85°C for 10 minutes. This is best carried out by heating a beaker of water in the microwave to 85°C, and placing the supernatant in a falcon tube in the beaker on a stir plate to circulate the heat. Material will be visibly precipitating from the supernatant at this point, but saposin will remain in solution. Precipitated material is easily removed by centrifuging at  $\sim$ 40,000 rcf for 30 minutes.

Clarified supernatant is incubated with 8-10 mL of Ni resin, pre-washed with lysis buffer. Batch binding should occur for 2 hours at 4°C (RT is acceptable for saposin). Following batch binding, the slurry is applied to a gravity column to collect resin. Wash steps are performed at 4°C with a flow rate of  $\sim$ 2 mL/min. Resin is then washed with 150 mL lysis buffer, followed by 150 mL of buffer B (50 mM HEPES pH 7.5, 300 mM NaCl, 20 mM imidazole, 50 mM sodium cholate), and 150 mL of buffer C (50 mM HEPES pH 7.5, 300 mM NaCl, 50 mM imidazole). Protein is eluted at a slower flow rate ( $\sim$ 1 mL/min), with buffer C supplemented with 400 mM imidazole. The eluted protein should come off within one column volume.

After elution, both MSPs should be dialyzed in 10kDa MWCO against 3L of 20 mM Tris pH 7.4, 100 mM NaCl, and 0.5 mM EDTA. After a few hours of dialysis, the samples are exchanged into a fresh 3L of dialysis buffer. Protein should not be

injected over gel filtration as it can form messy oligomers and should be only checked by SDS-PAGE for biochemical purity. Typical yields are ~60 mg per liter of culture. The molecular weight of MSP1D1 is 24662 Da, with an extinction coefficient of 21430. The molecular weight of MSPE3D1 is 32600 Da, with an extinction coefficient of 29400. Following dialysis, protein is concentrated to ~100  $\mu$ M, flash frozen in 1 mL aliquots in LN<sub>2</sub>, and stored at -80°C until needed.

After elution of saposin protein, use the elution buffer to estimate the protein concentration and yield prior to TEV cleavage. The uncleaved/cleaved molecular weight of saposin is 11607/9141 Da, with extinction coefficients of 10345/8855. A typical yield is about 20 mg of saposin per liter of culture. An addition of 1:10 w/w of TEV protease is then added to the eluate. Using 7kD MWCO dialysis tubing, the sample is dialyzed overnight at 4°C in against 2 L of 150 mM NaCl, 20 mM HEPES pH 7.5. The next morning dialyzed protein is concentrated in 3kD MWCO concentrators to 3-4 mL and injected in 1 mL amounts over a Superose200 10/300 column, using 150 mM NaCl, 20 mM HEPES pH 7.5 as running buffer. A peak should elute around 18 mL, spread over about 2 mL. Fractions comprising this peak from each run can be combined and quantified. Aliquots of ~500  $\mu$ L can be flash frozen in LN<sub>2</sub> and stored at -80°C until needed.

*Reconstitution of Scap/Insig complex*

Following gel filtration in LMNG, the Scap/Insig complex can be either frozen directly onto cryo-EM grids for analysis or reconstituted into a lipid bilayer membrane mimetic. For the latter, the protein must be carefully quantified using  $A_{280}$ . Typically, I aimed to keep the protein input for every reconstitution of at least 1 mg to produce enough reconstituted material to analyze by cryo-EM with minimal (or no) concentration. The two MSP scaffolds used for reconstitution of the complex were MSP1D1, which forms ~10 nm nanodiscs, and the larger MSPE3D1, which forms ~12 nm nanodiscs. Both of these scaffold proteins were purified previously using the protocol described above. Prior to reconstitution (at least a day beforehand), lipid mixtures are dissolved in chloroform and dried with argon gas to form a film in a small glass vial, placed in a vacuum dessicator overnight (or longer), and stored at -20°C. While we initially used native lipids purified by Arun Radhakrishnan's lab from rabbit liver, we ultimately switched to a synthetic lipid mixture of 3:1:0.3 POPC:POPE:cholesterol. This lipid mixture, while not perfect, is a good representation of the ER lipid composition with saturating amounts (10%) of cholesterol. The lipid mixture is brought up in a solution containing 150 mM NaCl, 20 mM HEPES pH 7.5, 1% LMNG, 0.5 mM TCEP, and 600  $\mu$ M 27-hydroxycholesterol to a final lipid concentration of 10 mM. The vial is sonicated in an ice bath for an hour to form a homogenous, but cloudy, solution.

After preparing all the components, the required amounts of each component are calculated. The optimal molar ratios (complex : scaffold protein : lipid) for each scaffold protein are as follows: MSP1D1 – 1:3:60; MSP1E3D1 – 1:3:130; saposin – 1:3:60. The final volume for each reconstitution should be no more than 1 mL final volume, although taking into account losses from each transfer step, volumes up to 1.2 mL are acceptable. Depending on the concentration of the Scap/Insig complex after gel filtration, it might need to be concentrated prior to mixing. The components are added sequentially, followed by slow detergent removal with adsorbent BioBeads. Scap/Insig sample is first incubated with detergent-solubilized lipids on ice for 10 minutes. Then the scaffold protein is added and the sample is slowly rotated at 4°C for 1 hour to fully equilibrate. Following this hour, 5-6 additions of pre-washed BioBeads are added (~60 mg of BioBeads/addition), over the course of about 18 hours. BioBeads must be washed prior to addition twice with methanol, five times with Milli-Q water, and two times with a detergent free buffer (typically the final nanodisc gel filtration buffer). After 18 hours, the sample is then injected over a Superose6 10/300 column in 150 mM NaCl, 20 mM HEPES pH 7.5, 0.5 mM TCEP, and 1.2  $\mu$ M 27-hydroxycholesterol. The nanodisc reconstituted protein elutes as a single species around 15.5 mL, with a slight right shoulder that contains excess empty nanodiscs. An overlay of gel filtration traces and SDS-PAGE gels are shown in Fig. 25. Typical concentrations of the reconstituted complex after gel filtration range from 0.15-0.5 mg/mL.

As mentioned in the Scap purification section, I attempted to use SMA polymers to purify Scap in “native” nanodiscs. Through optimization, I found buffer conditions that improved solubilization, but the cryo-EM grids showed particles that were very homogenous in size and in general too large to be a Scap particle. Based on these results, we did not pursue SMALPs further.

### **GENERATION OF CONFORMATIONALLY-SPECIFIC NANOBODIES**

Conformationally-specific antibodies are enormously helpful tools in structural biology, as they can act as both rigid bodies to assist with particle alignment and as stabilizing binding partners to help lock a flexible protein into a single conformation. Antibodies are typically generated through an animal host inoculated with the target protein, followed by tedious and laborious screening procedures to hopefully be left with a handful of candidates. During my tenure working on Scap, we went through three separate attempts to generate antibodies. After years of purifying protein and performing a large amount of follow-up experiments, all three rounds yielded antibodies that were not conformationally specific (i.e. western blot-positive), and ultimately were not useful for structural studies.

As a faster and less frustrating route to obtain antibodies, nanobodies have gained popularity recently. Nanobodies differ from conventional antibodies because they only contain a single variable heavy chain that contains three complementarity-

determining loops (CDRs). McMahon *et al.* describe a yeast surface display platform that can be used in conjunction with fluorescence-activated cell sorting (FACS) to rapidly screen through up to  $5 \times 10^8$  unique sequences (McMahon *et al.*, 2018). We turned to this platform as a potential fourth attempt to generate conformationally specific nanobodies against the Scap/Insig complex. To achieve this, I worked with Dan Kober to initiate these efforts and isolate nanobody candidates that I tested with detergent-purified complex. While these efforts are currently ongoing at the time of this writing, I will touch on the general workflow and include preliminary data.

### **MACS/FACS screening**

The sample used for nanobody generation is Hamster Scap FL 2M + Hamster Insig-2, purified in LMNG as described previously. The initial requirements for sequential MACS steps required 5-10 mL of 1  $\mu$ M protein (172  $\mu$ g/mL). In order to satisfy these requirements, I carried out multiple large-scale purifications of protein and froze the M2 eluate after overnight PNGaseF treatment. I injected a small amount of eluate on the Superose column to verify purity and homogeneity, and get an approximate idea of the yield.

While Insig is contained mostly within the membrane, Scap has multiple soluble regions that can be used as epitopes. The luminal loops, loop 1 and loop 7, have been the culprit of multiple failures in antibody generation due to their flexibility and hydrophobic regions that can give rise to linear epitopes and false positives.

Additionally, the C-terminus of Scap contains the large WD domain that is likely flexible within our complex. Both of these regions of the protein would not be ideal to raise nanobodies against for these reasons. Ideally, we would like to have a nanobody bind in a small groove within loops, or possibly even interface with the TM region as seen with the  $\beta$ 2 adrenergic receptor-Gs complex (Rasmussen *et al.*, 2011). To increase our chances, Dan Kober and I decided to use purified L1L7 and WD domain as counterscreening tools. This required 0.5-1 mg amounts of purified protein that we directly labeled with the flurophore dye Alexa 647 (made at a 10 mM stock in DMSO). L1L7 was labeled at a 3-fold excess, and the WD domain was labeled at a substoichiometric amount given the number of free cysteines in that domain. Combined with depletion of the majority of M2-reactive clones, our counterscreen was able to remove many of our false positives and leave us with a small but workable population of clones. Individual clones were picked and screened against purified protein, and after sequencing, 19 unique clones (from 21 colonies) were chosen for biochemical characterization and are currently being tested.

### **Expression and characterization of individual clones**

Given the hydrophobicity of membrane proteins, it is common to pull out generally “greasy” sequences in the CDRs of nanobody candidates. Of our 19 clones, it does not seem immediately apparent that that is our result (e.g. no poly-tryptophan sequences). The 19 sequences (from 21 colonies) are shown in Fig. 26.

These sequences were subcloned into the pMES vector used for Nb35 and Nb80 in the  $\beta$ 2 adrenergic receptor studies (Rasmussen *et al.*, 2010; Rasmussen *et al.*, 2011). Plasmids were transformed into the WK6 *E. coli* strain, and cultured in terrific broth (TB) containing 0.1% glucose, 2 mM MgCl<sub>2</sub>, and ampicillin antibiotics. When the cell density reached an OD<sub>600</sub> of 0.6-0.8, 1 mM IPTG and additional ampicillin were added to induce expression, and the temperature was decreased from 37°C to 25°C overnight. The next morning, cells were harvested by centrifugation.

Resulting *E. coli* cell pellets are lysed by osmotic shock by gentle douncing in cold lysis buffer containing 125 mM sucrose, 50 mM Tris pH 8.0, 12.5 mM EDTA, and protease inhibitors, followed by stirring for 30 minutes at 4°C. Cell debris is separated by centrifugation, and the supernatant is saved as it contains the periplasmic proteins. It is common for this to have a very light color. The solution is then loaded onto ~3-5 mL of fresh Ni resin, equilibrated in 150 mM NaCl, 20 mM HEPES pH 7.5, and 20 mM imidazole. Following loading, the resin is washed by 10 column volumes of equilibration buffer, and is eluted with the same buffer spiked with 300 mM imidazole. Given the strong prevalence of beta sheet structure in the nanobodies, it is not expected to see a high amount of protein staining, but should be visible on an SDS-PAGE gel. Following elution, the protein should be dialyzed against 100 mM NaCl, 20 mM HEPES pH 7.5, and 5-10% glycerol. Protein yields

have ranged from 1-40 mg/L, and seems to be dependent on protein sequence as well as other unknown factors.

Final yields have been quite low for our nanobodies, but I have purified enough of them to initially test for binding with Scap/Insig. To do this, I incubated ~200 µg of purified Scap/Insig with ~90 µg of purified nanobody for 1 hr, rotating at 4°C. Following incubation, I injected each sample on a Superose 6 10/300 column equilibrated in the same LMNG gel filtration buffer used for protein purification and MACS/FACS screening. Due to the vast differences in molecular weight (172kD for Scap/Insig; ~14kD for nanobodies), I do not expect to be able to see a shift in retention volume for the main Scap/Insig peak. Instead, I can monitor the disappearance of the nanobody peak, or alternatively run Scap/Insig peak fractions on an SDS-PAGE gel, followed by Sypro Ruby staining. I opted for the latter, and from the first set of 19 nanobodies tested, I have evidence of 4 nanobodies showing strong co-elution with Scap/Insig, which would only occur at the retention volume if the nanobodies were binding to the complex (versus a non-specific oligomerization effect). The remaining nanobodies either show weak (but non-zero) binding, or do not bind at all. Figures 27 and 28 show a summary of this analysis. I'm currently working towards following up these candidates on a larger scale to verify this result, as well as testing the rest of our nanobody clones. Our hope is that one of these clones can help reduce flexibility in our sample by binding and stabilizing a specific

region of our complex. In addition, an ideal nanobody can potentially act as a fiducial marker to improve upon our current cryo-EM studies, discussed in the next chapter.

## REFERENCES

- Adams, C. M., Goldstein, J. L., & Brown, M. S. (2003). Cholesterol-induced conformational change in SCAP enhanced by Insig proteins and mimicked by cationic amphiphiles. *Proceedings of the National Academy of Sciences*, *100*(19), 10647–10652. <http://doi.org/10.1073/pnas.1534833100>
- Aller, S. G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., et al. (2009). Structure of P-glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. <http://doi.org/10.2210/pdb3g60/pdb>
- Amlacher, S., Sarges, P., Flemming, D., van Noort, V., Kunze, R., Devos, D. P., et al. (2011). Insight into Structure and Assembly of the Nuclear Pore Complex by Utilizing the Genome of a Eukaryotic Thermophile. *Cell*, *146*(2), 277–289. <http://doi.org/10.1016/j.cell.2011.06.039>
- Bayburt, T. H., Grinkova, Y. V., & Sligar, S. G. (2002). Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with Membrane Scaffold Proteins. *Nano Letters*, *2*(8), 853–856. <http://doi.org/10.1021/nl025623k>
- Bayburt, T. H., Grinkova, Y. V., & Sligar, S. G. (2006). Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs. *Archives of Biochemistry and Biophysics*, *450*(2), 215–222. <http://doi.org/10.1016/j.abb.2006.03.013>
- Berka, R. M., Grigoriev, I. V., Otilar, R., Salamov, A., Grimwood, J., Reid, I., et al. (2011). Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nature Biotechnology*, *29*(10), 922–927. <http://doi.org/10.1038/nbt.1976>
- Broecker, J., Eger, B. T., & Ernst, O. P. (2017). Crystallography of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. *Structure/Folding and Design*, *25*(2), 384–392. <http://doi.org/10.1016/j.str.2016.12.004>

Brown, A. J., Sun, L., Feramisco, J. D., Brown, M. S., & Goldstein, J. L. (2002). Cholesterol addition to ER membranes alters conformation of SCAP, the SREBP escort protein that regulates cholesterol metabolism. *Molecular Cell*, *10*(2), 237–245.

Brown, M. S., Radhakrishnan, A., & Goldstein, J. L. (2017). Retrospective on Cholesterol Homeostasis: The Central Role of Scap. *Annual Review of Biochemistry*, *87*(1), annurev-biochem-062917-011852-25.  
<http://doi.org/10.1146/annurev-biochem-062917-011852>

Chien, C.-T. H., Helfinger, L. R., Bostock, M. J., Solt, A., Tan, Y. L., & Nietlispach, D. (2017). An Adaptable Phospholipid Membrane Mimetic System for Solution NMR Studies of Membrane Proteins. *Journal of the American Chemical Society*, *139*(42), 14829–14832. <http://doi.org/10.1021/jacs.7b06730>

Denisov, I. G., Grinkova, Y. V., Lazarides, A. A., & Sligar, S. G. (2004). Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *Journal of the American Chemical Society*, *126*(11), 3477–3487.  
<http://doi.org/10.1021/ja0393574>

Denisov, I. G., & Sligar, S. G. (2016). Nanodiscs for structural and functional studies of membrane proteins. *Nature Structural & Molecular Biology*, *23*(6), 481–486.  
<http://doi.org/10.1038/nsmb.3195>

Feramisco, J. D., Radhakrishnan, A., Ikeda, Y., Reitz, J., Brown, M. S., & Goldstein, J. L. (2005). Intramembrane aspartic acid in SCAP protein governs cholesterol-induced conformational change. *Proceedings of the National Academy of Sciences*, *102*(9), 3242–3247. <http://doi.org/10.1073/pnas.0500206102>

Frauenfeld, J., Löving, R., Armache, J.-P., Sonnen, A. F.-P., Guettou, F., Moberg, P., et al. (2016). A saposin-lipoprotein nanoparticle system for membrane proteins. *Nature Methods*, *13*(4), 345–351. <http://doi.org/10.1038/nmeth.3801>

Gao, Y., Cao, E., Julius, D., & Cheng, Y. (2016). TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature*, *534*(7607), 347–351.  
<http://doi.org/10.1038/nature17964>

García-Nafria, J., Lee, Y., Bai, X., Carpenter, B., & Tate, C. G. (2018). Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. *eLife*, *7*. <http://doi.org/10.7554/eLife.35946.001>

Goehring, A., Lee, C.-H., Wang, K. H., Michel, J. C., Claxton, D. P., Bacongus, I., et al. (2014). Screening and large-scale expression of membrane proteins in

mammalian cells for structural studies. *Nature Protocols*, 9(11), 2574–2585.  
<http://doi.org/10.1038/nprot.2014.173>

Hagn, F., Etkorn, M., Raschle, T., & Wagner, G. (2013). Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *Journal of the American Chemical Society*, 135(5), 1919–1925.  
<http://doi.org/10.1021/ja310901f>

Hasan MT, Chang TY. (1994). Somatic cell genetic analysis of two classes of CHO cell mutants expressing opposite phenotypes in sterol-dependent regulation of cholesterol metabolism. *Somat. Cell Mol. Genet.* 20:481–91

Hite, R. K., Yuan, P., Li, Z., Hsuing, Y., Walz, T., & MacKinnon, R. (2015). Cryo-electron microscopy structure of the Slo2.2 Na<sup>+</sup>-activated K<sup>+</sup> channel. *Nature*, 527(7577), 198–203. <http://doi.org/10.1038/nature14958>

Hua, X., Nohturfft, A., Goldstein, J. L., & Brown, M. S. (1996). Sterol resistance in CHO cells traced to point mutation in SREBP cleavage-activating protein. *Cell*, 87(3), 415–426.

Jin, P., Bulkley, D., Guo, Y., Zhang, W., Guo, Z., Huynh, W., et al. (2017). Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature Publishing Group*, 547(7661), 118–122. <http://doi.org/10.1038/nature22981>

Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., et al. (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*, 523(7562), 561–567. <http://doi.org/10.1038/nature14656>

Kintzer, A. F., & Stroud, R. M. (2016). Structure, inhibition and regulation of two-pore channel TPC1 from *Arabidopsis thaliana*. *Nature*, 531(7593), 258–264.  
<http://doi.org/10.1038/nature17194>

Knowles, T. J., Finka, R., Smith, C., Lin, Y. P., Dafforn, T., & Overduin, M. (2009). Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. *Journal of the American Chemical Society*, 131(22), 7484–7485. <http://doi.org/10.1021/ja810046q>

Kofuku, Y., Ueda, T., Okude, J., Shiraishi, Y., Kondo, K., Mizumura, T., et al. (2014). Functional Dynamics of Deuterated  $\beta$  2-Adrenergic Receptor in Lipid Bilayers Revealed by NMR Spectroscopy. *Angewandte Chemie International Edition*, 53(49), 13376–13379. <http://doi.org/10.1002/anie.201406603>

Lee, P. C. W., & DeBose-Boyd, R. A. (2010). Intramembrane glycine mediates multimerization of Insig-2, a requirement for sterol regulation in Chinese hamster ovary cells. *Journal of Lipid Research*, *51*(1), 192–201.

<http://doi.org/10.1194/jlr.M900336-JLR200>

Lee, S. C., Knowles, T. J., Postis, V. L. G., Jamshad, M., Parslow, R. A., Lin, Y. P., et al. (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols*, *11*(7), 1149–1162.

<http://doi.org/10.1038/nprot.2016.070>

Li, N., Wu, J.-X., Ding, D., Cheng, J., Gao, N., & Chen, L. (2017). Structure of a Pancreatic ATP-Sensitive Potassium Channel. *Cell*, *168*(1-2), 101–110.e10.

<http://doi.org/10.1016/j.cell.2016.12.028>

Liu, F., Zhang, Z., Csanády, L., Gadsby, D. C., & Chen, J. (2017). Molecular Structure of the Human CFTR Ion Channel. *Cell*, *169*(1), 85–91.e8.

<http://doi.org/10.1016/j.cell.2017.02.024>

Martin, G. M., Yoshioka, C., Rex, E. A., Fay, J. F., Elife, Q. X., 2017. (n.d.). Cryo-EM structure of the ATP-sensitive potassium channel illuminates mechanisms of assembly and gating. *Elife* <http://doi.org/10.7554/eLife.24149.001>

McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., et al. (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nature Structural & Molecular Biology*, 1–14.

<http://doi.org/10.1038/s41594-018-0028-6>

Morales-Pérez, C. L., Noviello, C. M., & Hibbs, R. E. (2016a). Manipulation of Subunit Stoichiometry in Heteromeric Membrane Proteins. *Structure/Folding and Design*, *24*(5), 797–805. <http://doi.org/10.1016/j.str.2016.03.004>

Morales-Pérez, C. L., Noviello, C. M., & Hibbs, R. E. (2016b). X-ray structure of the human  $\alpha 4\beta 2$  nicotinic receptor. *Nature*, *538*(7625), 416–416.

<http://doi.org/10.1038/nature19785>

Motamed, M., Zhang, Y., Wang, M. L., Seemann, J., Kwon, H. J., Goldstein, J. L., & Brown, M. S. (2011). Identification of Luminal Loop 1 of Scap Protein as the Sterol Sensor That Maintains Cholesterol Homeostasis. *Journal of Biological Chemistry*, *286*(20), 18002–18012. <http://doi.org/10.1074/jbc.M111.238311>

Nasr, M. L., Baptista, D., Strauss, M., Sun, Z.-Y. J., Grigoriu, S., Huser, S., et al. (2017). Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nature Methods*, *14*(1), 49–52. <http://doi.org/10.1038/nmeth.4079>

Nguyen, N. X., Armache, J.-P., Lee, C., Yang, Y., Zeng, W., Mootha, V. K., et al. (2018). Cryo-EM structure of a fungal mitochondrial calcium uniporter. *Nature*, 1–19. <http://doi.org/10.1038/s41586-018-0333-6>

Nöll, A., Thomas, C., Herbring, V., Zollmann, T., Barth, K., Mehdipour, A. R., et al. (2017). Crystal structure and mechanistic basis of a functional homolog of the antigen transporter TAP. *Proceedings of the National Academy of Sciences*, 114(4), E438–E447. <http://doi.org/10.1073/pnas.1620009114>

Olmeda, B., García-Álvarez, B., & Pérez-Gil, J. (2012). Structure–function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins. *European Biophysics Journal*, 42(2-3), 209–222. <http://doi.org/10.1007/s00249-012-0858-9>

Parker, J. L., & Newstead, S. (2017). Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature Publishing Group*, 1–19. <http://doi.org/10.1038/nature24464>

Parmar, M., Rawson, S., Scarff, C. A., Goldman, A., Dafforn, T. R., Muench, S. P., & Postis, V. L. G. (2018). Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *BBA - Biomembranes*, 1860(2), 378–383. <http://doi.org/10.1016/j.bbamem.2017.10.005>

Radhakrishnan, A., Ikeda, Y., Kwon, H. J., Brown, M. S., & Goldstein, J. L. (2007). Sterol-regulated transport of SREBPs from endoplasmic reticulum to Golgi: oxysterols block transport by binding to Insig. *Proceedings of the National Academy of Sciences*, 104(16), 6511–6518. <http://doi.org/10.1073/pnas.0700899104>

Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., et al. (2011). Crystal structure of the  $\beta$ 2 adrenergic receptor-Gs protein complex. *Nature*, 477(7366), 549–555. <http://doi.org/10.1038/nature10361>

Rasmussen, S. G. F., Choi, H.-J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., et al. (2010). Structure of a nanobody-stabilized active state of the  $\beta$ 2 adrenoceptor. *Nature*, 469(7329), 175–180. <http://doi.org/10.1038/nature09648>

Reeves, P. J., Callewaert, N., Contreras, R., & Khorana, H. G. (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proceedings of the National Academy of Sciences*, 99(21), 13419–13424. <http://doi.org/10.1073/pnas.212519299>

- Rheinberger, J., Gao, X., Schmidpeter, P. A., & Nimigean, C. M. (2018). Ligand discrimination and gating in cyclic nucleotide-gated ion channels from apo and partial agonist-bound cryo-EM structures. *eLife*, 7, 213. <http://doi.org/10.7554/eLife.39775>
- Ritchie, T. K., Kwon, H., & Atkins, W. M. (2011). Conformational analysis of human ATP-binding cassette transporter ABCB1 in lipid nanodiscs and inhibition by the antibodies MRK16 and UIC2. *Journal of Biological Chemistry*, 286(45), 39489–39496. <http://doi.org/10.1074/jbc.M111.284554>
- Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., et al. (2007). GPCR Engineering Yields High-Resolution Structural Insights into  $\beta$ -2-Adrenergic Receptor Function. *Science*, 318(5854), 1266–1273. <http://doi.org/10.1126/science.1150609>
- Schmidt, H. R., Zheng, S., Gurpinar, E., Koehl, A., Manglik, A., & Kruse, A. C. (2016). Crystal structure of the human  $\sigma$ 1 receptor. *Nature*, 532(7600), 527–530. <http://doi.org/10.1038/nature17391>
- Shao, Z., Yin, J., Chapman, K., Grzemska, M., Clark, L., Wang, J., & Rosenbaum, D. M. (2016). High-resolution crystal structure of the human CB1 cannabinoid receptor. *Nature*, 540(7634), 602–606. <http://doi.org/10.1038/nature20613>
- Shen, P. S., Yang, X., DeCaen, P. G., Liu, X., Bulkley, D., Clapham, D. E., & Cao, E. (2016). The Structure of the Polycystic Kidney Disease Channel PKD2 in Lipid Nanodiscs. *Cell*, 167(3), 763–767.e11. <http://doi.org/10.1016/j.cell.2016.09.048>
- Shen, H., Zhou, Q., Pan, X., Li, Z., Wu, J., & Yan, N. (2017). Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science*, 355(6328), eaal4326–14. <http://doi.org/10.1126/science.aal4326>
- Shinoda, T., Ogawa, H., Cornelius, F., & Toyoshima, C. (2009). Crystal structure of the sodium-potassium pump at 2.4 Å resolution. *Nature*, 459(7245), 446–450. <http://doi.org/10.1038/nature07939>
- Shukla, A. K., Westfield, G. H., Xiao, K., Reis, R. I., Huang, L.-Y., Tripathi-Shukla, P., et al. (2014). Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature*, 512(7513), 218–222. <http://doi.org/10.1038/nature13430>
- Strege, A., Carpenter, B., Edwards, P. C., & Tate, C. G. (2017). Strategy for the Thermostabilization of an Agonist-Bound GPCR Coupled to a G Protein. *G Protein*

*Coupled Receptors* (1st ed., Vol. 594, pp. 243–264). Elsevier Inc.

<http://doi.org/10.1016/bs.mie.2017.05.014>

Taylor, N. M. I., Manolaridis, I., Jackson, S. M., Kowal, J., Stahlberg, H., & Locher, K. P. (2017). Structure of the human multidrug transporter ABCG2. *Nature*, 546(7659), 504–509. <http://doi.org/10.1038/nature22345>

Yabe, D., Brown, M. S., & Goldstein, J. L. (2002). Insig-2, a second endoplasmic reticulum protein that binds SCAP and blocks export of sterol regulatory element-binding proteins. *Proceedings of the National Academy of Sciences*, 99(20), 12753–12758. <http://doi.org/10.1073/pnas.162488899>

Yan, Z., Zhou, Q., Wang, L., Wu, J., Zhao, Y., Huang, G., et al. (2017). Structure of the Nav1.4- $\beta$ 1 Complex from Electric Eel. *Cell*, 170(3), 470–475.e11. <http://doi.org/10.1016/j.cell.2017.06.039>

Yin, J., Mobarec, J. C., Kolb, P., & Rosenbaum, D. M. (2015). Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant. *Nature*, 519(7542), 247–250. <http://doi.org/10.1038/nature14035>

Yin, J., Babaoglu, K., Brautigam, C. A., Clark, L., Shao, Z., Scheuermann, T. H., et al. (2016). Structure and ligand-binding mechanism of the human OX1 and OX2 orexin receptors. *Nature Structural & Molecular Biology*, 23(4), 293–299. <http://doi.org/10.1038/nsmb.3183>

Yuan, P., Leonetti, M. D., Hsiung, Y., & MacKinnon, R. (2011). Open structure of the Ca<sup>2+</sup> gating ring in the high-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel. *Nature*, 481(7379), 94–97. <http://doi.org/10.1038/nature10670>

Zhang, Y., Motamed, M., Seemann, J., Brown, M. S., & Goldstein, J. L. (2013). Point mutation in luminal loop 7 of Scap protein blocks interaction with loop 1 and abolishes movement to Golgi. *Journal of Biological Chemistry*, 288(20), 14059–14067. <http://doi.org/10.1074/jbc.M113.469528>

Zhang, Y., Lee, K. M., Kinch, L. N., Clark, L., Grishin, N. V., Rosenbaum, D. M., et al. (2016). Direct Demonstration That Loop1 of Scap Binds to Loop7: A CRUCIAL EVENT IN CHOLESTEROL HOMEOSTASIS. *Journal of Biological Chemistry*, 291(24), 12888–12896. <http://doi.org/10.1074/jbc.M116.729798>

Zhou, H.-X., & Cross, T. A. (2013). Influences of Membrane Mimetic Environments on Membrane Protein Structures. *Annual Review of Biophysics*, 42(1), 361–392. <http://doi.org/10.1146/annurev-biophys-083012-130326>

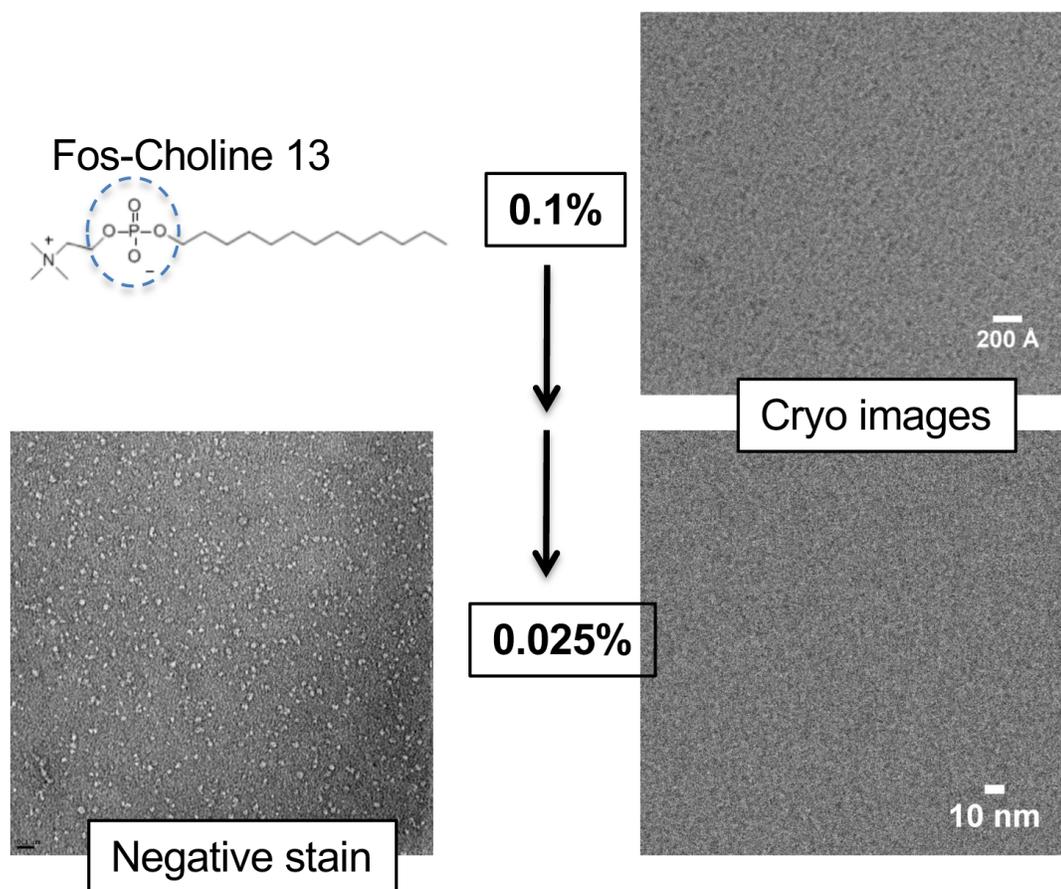


Figure 1. Artifacts arising from FC-13 detergent in negative stain and cryo-EM of TCT. Objects are visible in cryo-EM images at 0.1% detergent, but appear to be reduced at 0.025%. However, negative stain images at 0.025% still contain a high number of micelles.

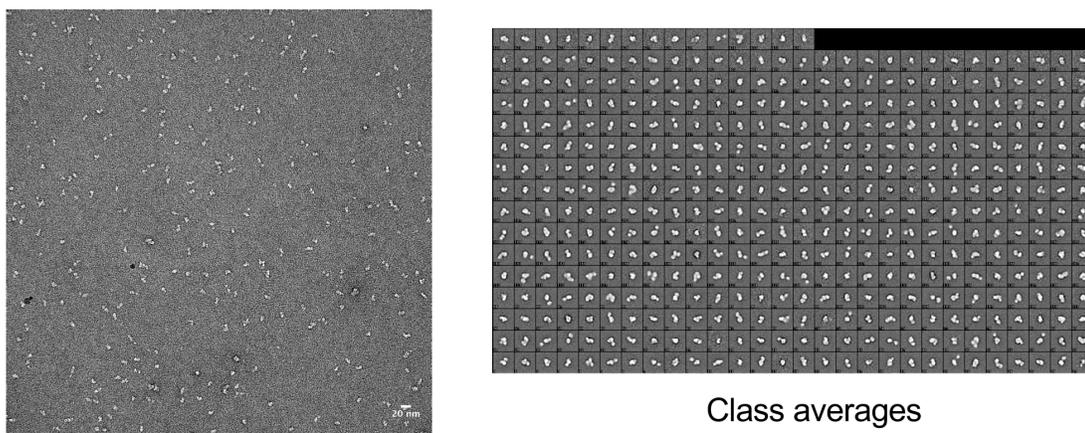


Figure 2. Negative stain EM of amphipol-reconstituted TCT. Example micrograph is shown on the left, with 2D class averages shown on the right.

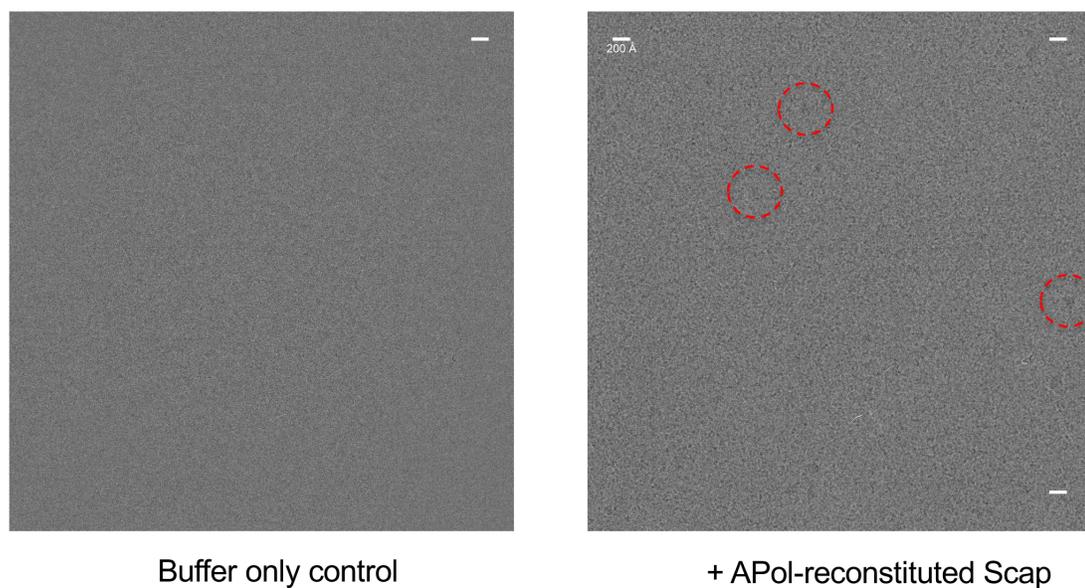


Figure 3. Cryo-EM of amphipol-reconstituted TCT. The left micrograph shows a buffer only control to confirm we weren't going to face the same artifact issue as seen with FC-13 samples. The right micrograph contains a small number of particles, circled in red. The density and contrast were too low for structure determination.

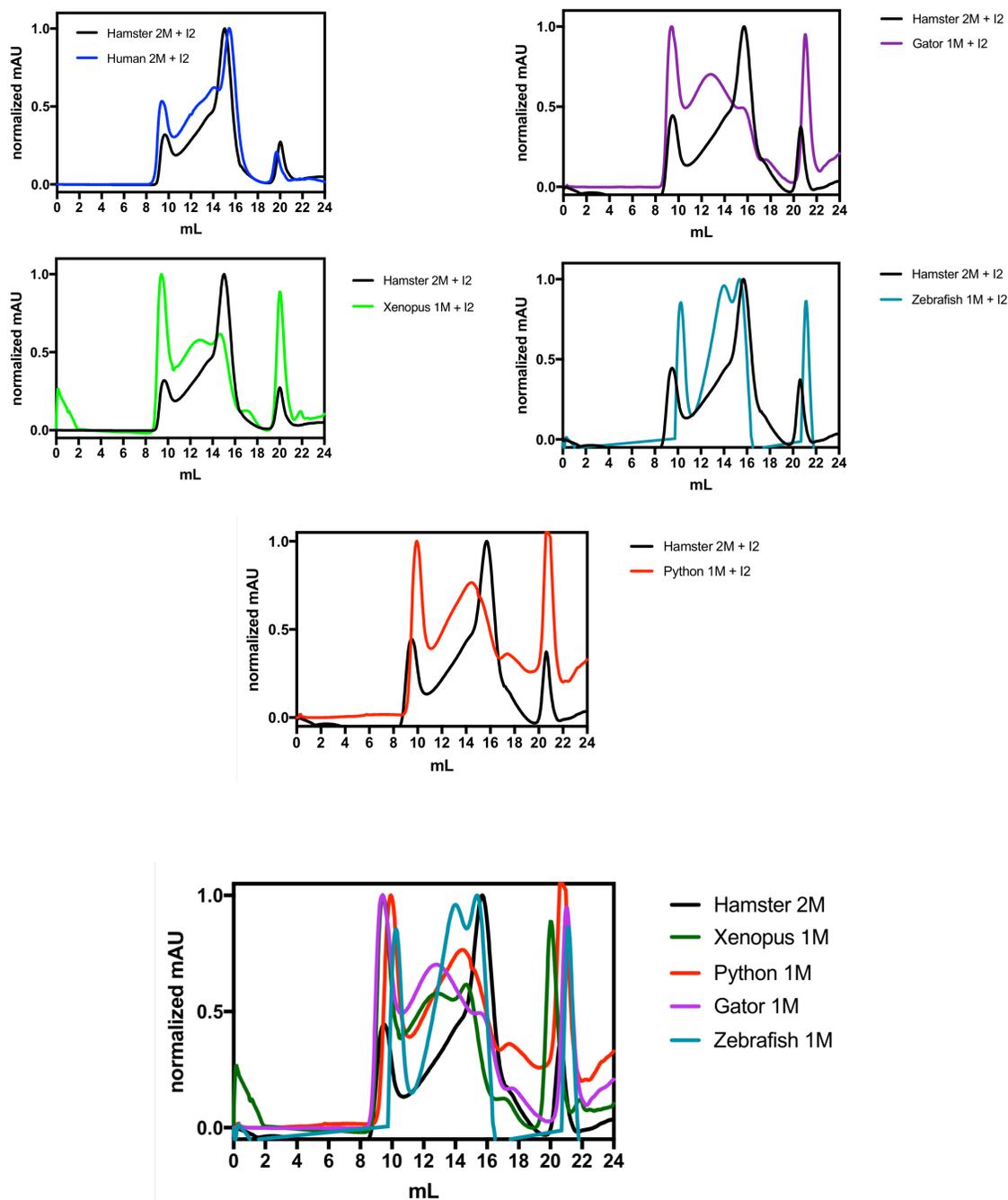


Figure 4. Orthologue screening of ScapFL with species-matched Insig-2. Normalized SEC traces in DDM buffer are overlaid with Hamster 2M + Insig-2 samples. Overlay of the more sequence-divergent orthologues are shown in the bottom panel.

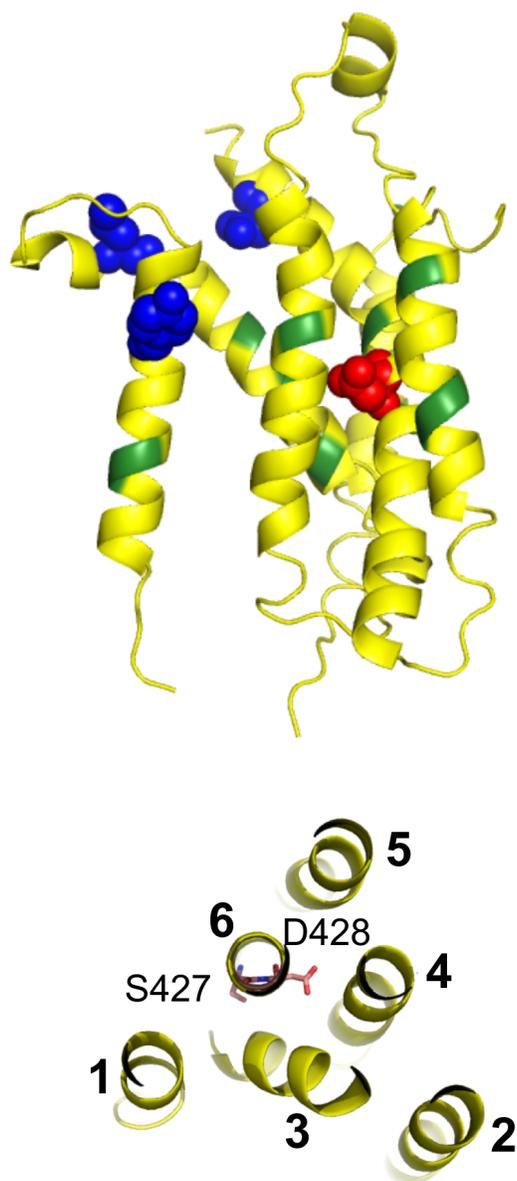


Figure 5. Mutations screened for in FSEC thermostability assay are mapped onto the model of Scap TM 1-6, based on the structural homology of this region to the protein NPC1. Top panel: functionally identified mutations to residues are highlighted in blue (and red), and polar residues mutated are shown in green. D428 is shown in red. Bottom panel: top down view of model in the top panel, with helices labeled 1-6. Mutations comprising the 2M construct (S427 and D428) are labeled and shown as sticks.

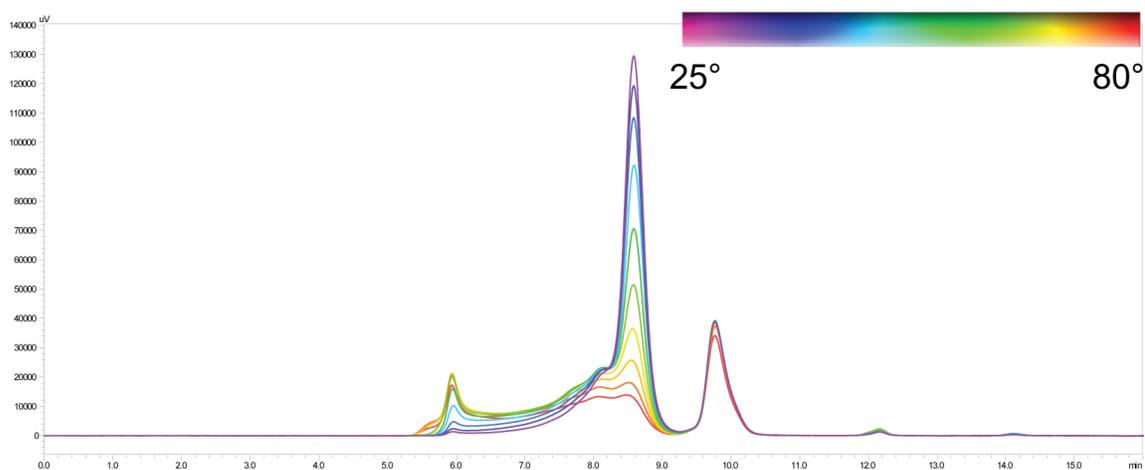


Figure 6. Overlay traces of individual FSEC runs. Samples are heated between 25°C and 80°C and injected on a Shodex KW-805 column in DDM buffer. ScapTM16mini elutes around 8.5 minutes, and this peak is reduced with increasing temperature.

| mutation/additive         | T <sub>m</sub> effect? | notes  |
|---------------------------|------------------------|--|
| Y298C                     | no change              |  |
| L315F                     | no change              |  |
| D428A                     | ~20C increase          | reproducible, significant boost in yield over WT         |
| D443N                     | no change              |  |
| S24A                      | no change              |  |
| T33A                      | no change              |  |
| T290A                     | no change              |  |
| T291A                     | no change              |  |
| T322A                     | ~1-2C increase         |  |
| S325A                     | no change              |  |
| S326A                     | no change              | boost in yield over WT                                   |
| S330A                     | no change              | boost in yield over WT                                   |
| S391A                     | no change              |  |
| T398A                     | no change              |  |
| S427A                     | ~10C increase          |  |
| D428V                     | ~22C increase          | ~5x increase in yield over WT, 2x over D428A at 50C      |
| D428N                     | ~6-7C increase         |  |
| D428L                     | ~27C increase          | less expression overall, lower yield, reproducible       |
| Q432L                     | ~3C increase           |  |
| E359A                     | ~11-12C increase       |  |
| G357A/E359A               | unclear                | very little yield (two separate experiments)             |
| S427A/D428A               | ~31C increase          | not a huge boost in yield, reproducible                  |
| S427A/D428V               | ~33C increase          | not a huge boost in yield                                |
| E359A/S427A/D428V         | ~36C increase          | below avg yield  |
| E359A/D428V               | ~33C increase          | not a huge boost in yield                                |
| E359A/S427A               | ~20C increase          | not a huge boost in yield                                |
| E399A/S427A/D428V         | ~35C increase          | higher yield than 359/427/428                            |
| E359A/E399A/S427A/D428V   |                        |  |
| M329C/D428C               | ~16C increase          | very low yield, reproducible                             |
| V396C/D428C               | ~18C increase          | reproducible, significant boost in yield, slight R shift |
| G357C/D428C               | ~21C increase          | significant boost in yield                               |
| betaCD treated            | no change              |  |
| saturating chol           | no change              |  |
| WT + 1mM Ca <sup>2+</sup> | no change              |  |

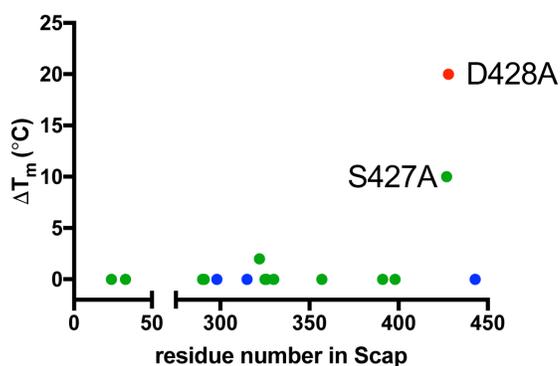


Figure 7. Summary of FSEC results. Top: table of mutations/conditions tested with FSEC, along with any increase in thermostability and pertinent notes. Below: Scatterplot of individual mutations (colored the same as in Fig. 5), with any increase in T<sub>m</sub> compared to WT noted on the Y axis.

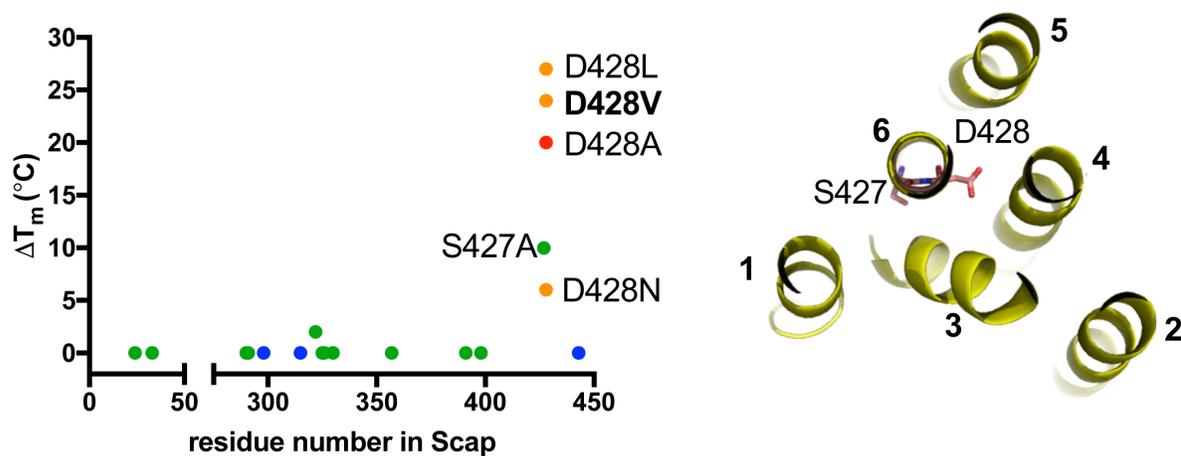


Figure 8. Summary of additional substitutions at position 428. Left: scatterplot as shown in Fig. 7, with additional data points for mutations to neutral charged and larger hydrophobic residues. Right: top view of ScapTM16mini model (as shown in Fig. 5) to show helical core and spatial limits of our side chains.

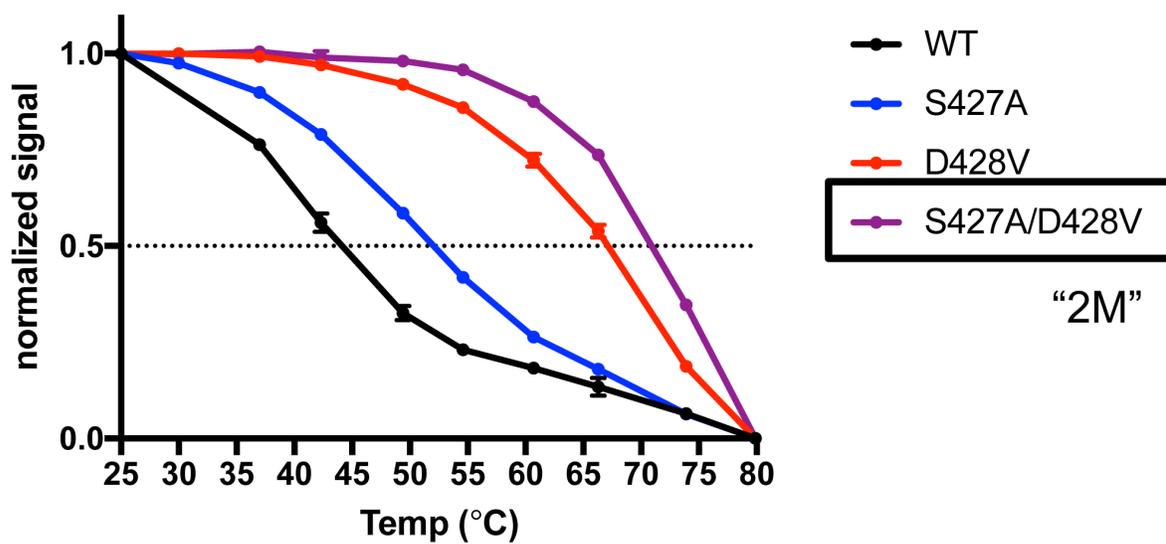


Figure 9. Full  $T_m$  curve of Scap<sup>TM16mini</sup> WT constructs shown in black, with individual point mutations and the double mutation (2M). The D428V mutation provides the highest amount of stabilization, but the effects of both mutations are additive, as shown in purple.

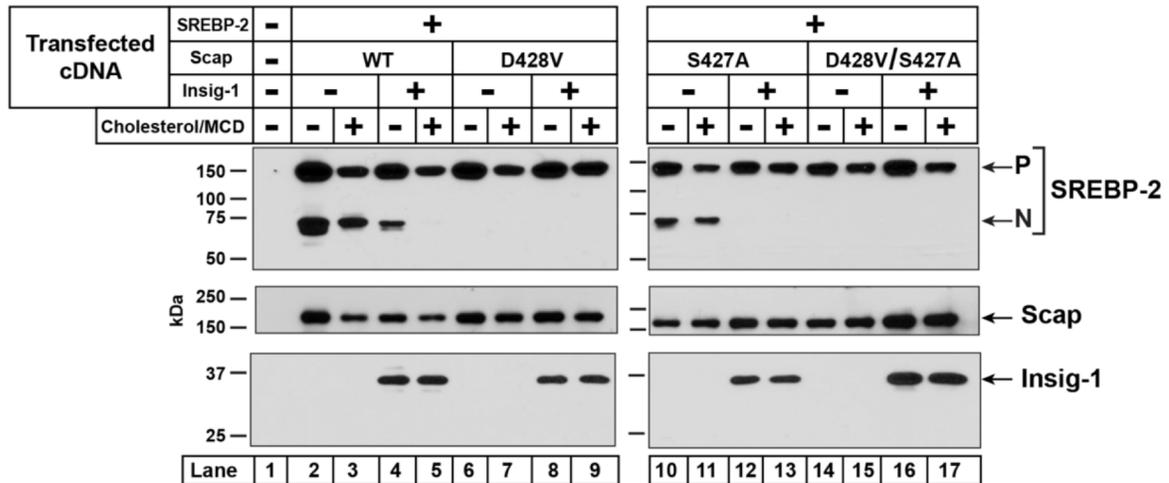


Figure 10. *In vivo* SREBP trafficking assay of single and double point mutations in CHO cells. Precursor (P) and nuclear (N) forms of SREBP-2 are monitored as a function of Insig expression, cholesterol levels, and Scap mutation. WT Scap displays normal trafficking patterns that rely on Insig expression or depletion of cholesterol for correct retention or trafficking, respectively. S427A showed an intermediate effect on SREBP trafficking, while D428V or S427A/D428V retained all SREBP in the ER even when Insig is not expressed or sterol levels are depleted.

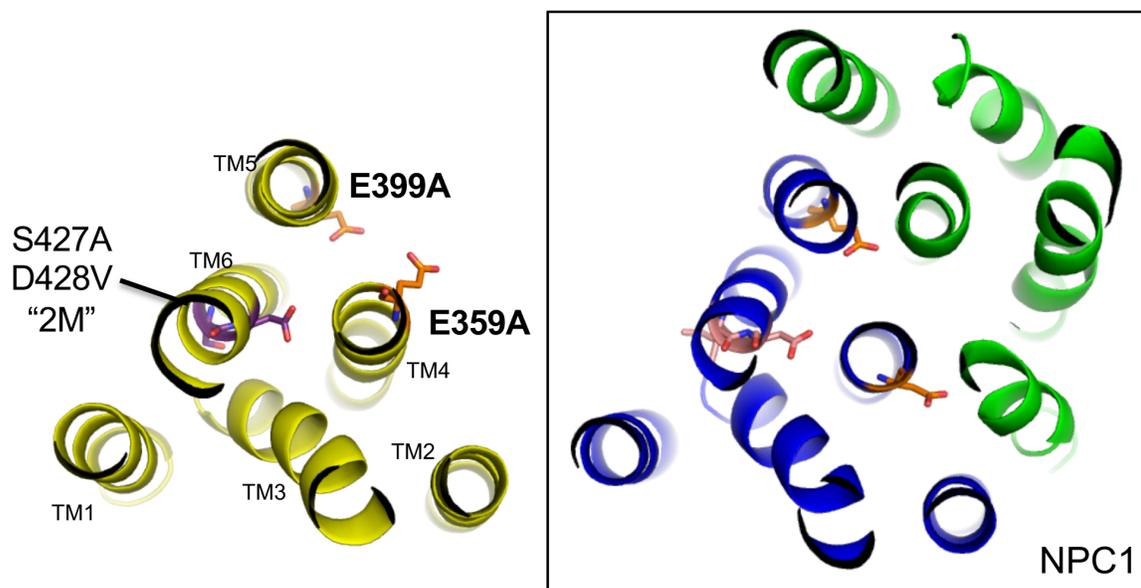


Figure 11. Comparison of Scap TM 1-6 with the SSD of NPC1. TM helices 1-6 of Scap are predicted to adopt a very similar fold to the SSD of NPC1, shown in blue on the right. NPC1 forms a pseudosymmetrical intramolecular dimer with additional helices shown in green. The residues focused on in this section are highlighted in Scap on the left and the homologous locations are also shown as sticks within NPC1.

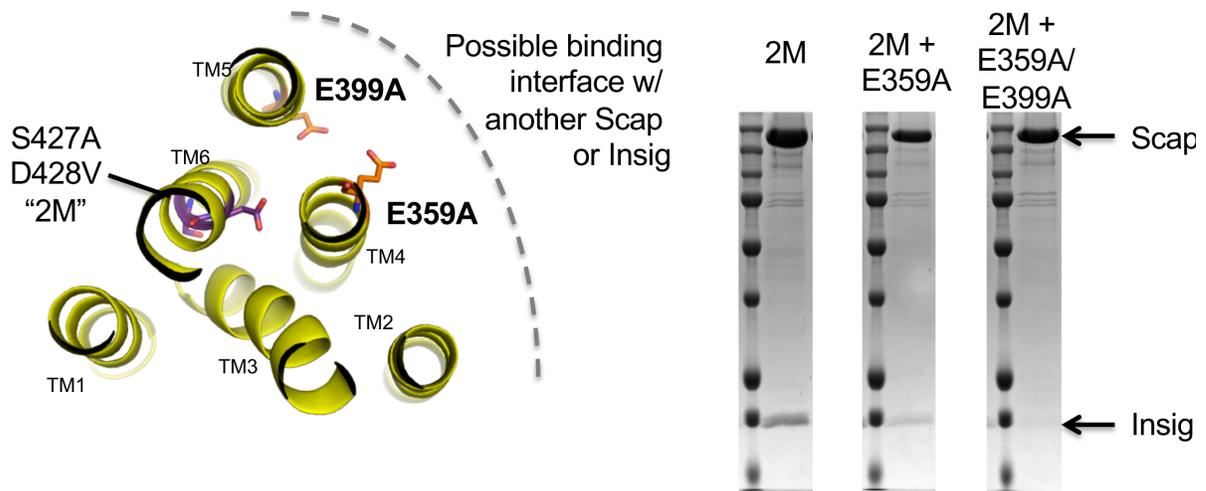


Figure 12. Model of Scap TM 1-6 shown on the left, as in Figure 11. The possible binding interface with another Scap or Insig is noted, with glutamate residues predicted to be facing out. Right: Mutant forms of ScapFL were co-expressed with untagged Insig-2 and affinity purified with M2 resin. Non-normalized eluate fractions are shown here on an SDS-PAGE gel. 2M alone is able to efficiently pull down Insig, where 2M + E359A or 2M + E359A/E399A display a reduced affinity for Insig, suggesting that this is in fact a binding interface with Insig.

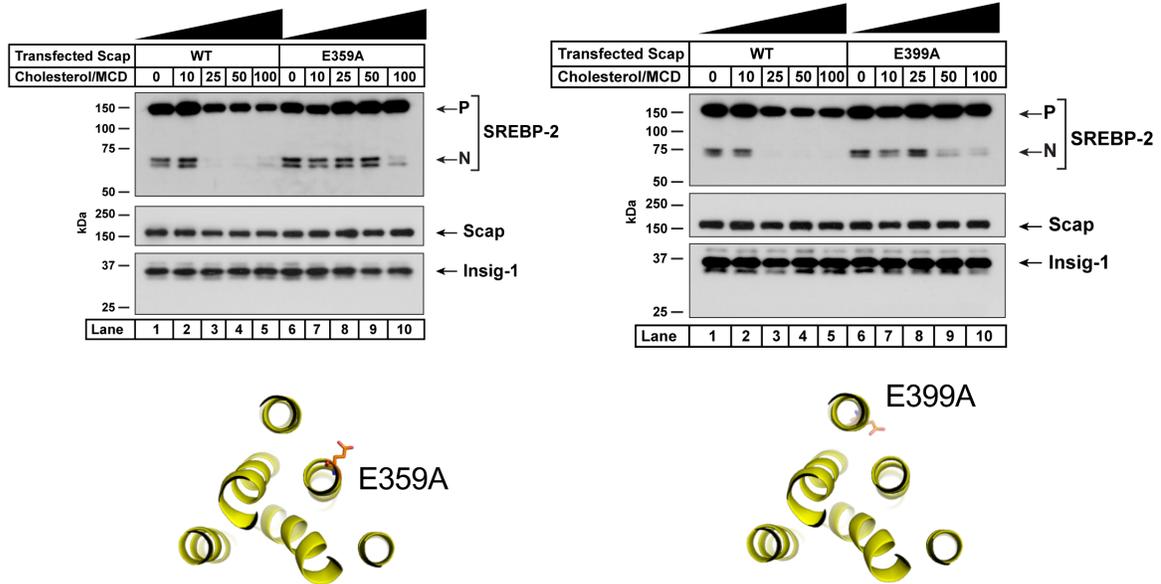


Figure 13. SREBP trafficking assays of WT and glutamate mutant forms of Scap. Increasing amounts of cholesterol are added back to sterol-depleted cells transfected with plasmids encoding Insig, SREBP-2, and either WT or mutant Scap. Each of the glutamate mutants increased the amount of cholesterol required for retention of SREBP in the ER, suggesting a decreased affinity for the retention protein Insig. The double mutant (E359A/E399A) was not expressed efficiently and thus was not tested in this assay.

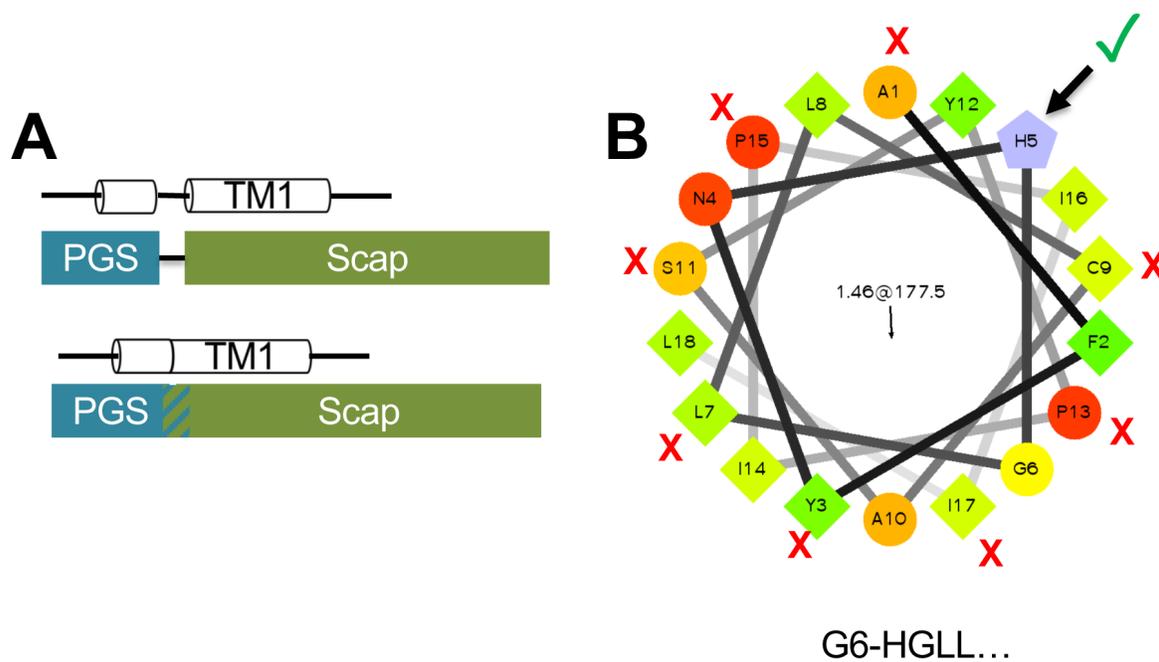


Figure 14. Optimization of N-terminal fusion protein. Panel A shows the overall schematic of finding the correct register of the C-terminus of PGS (aka G6) to create a continuous helix with TM1 of Scap. B) Helical wheel projection showing attachment points that were detrimental to expression shown as red Xs and the best attachment point shown as a green check mark. This places G6 (aka PGS) to attach at the sequence HGLL... in Scap.

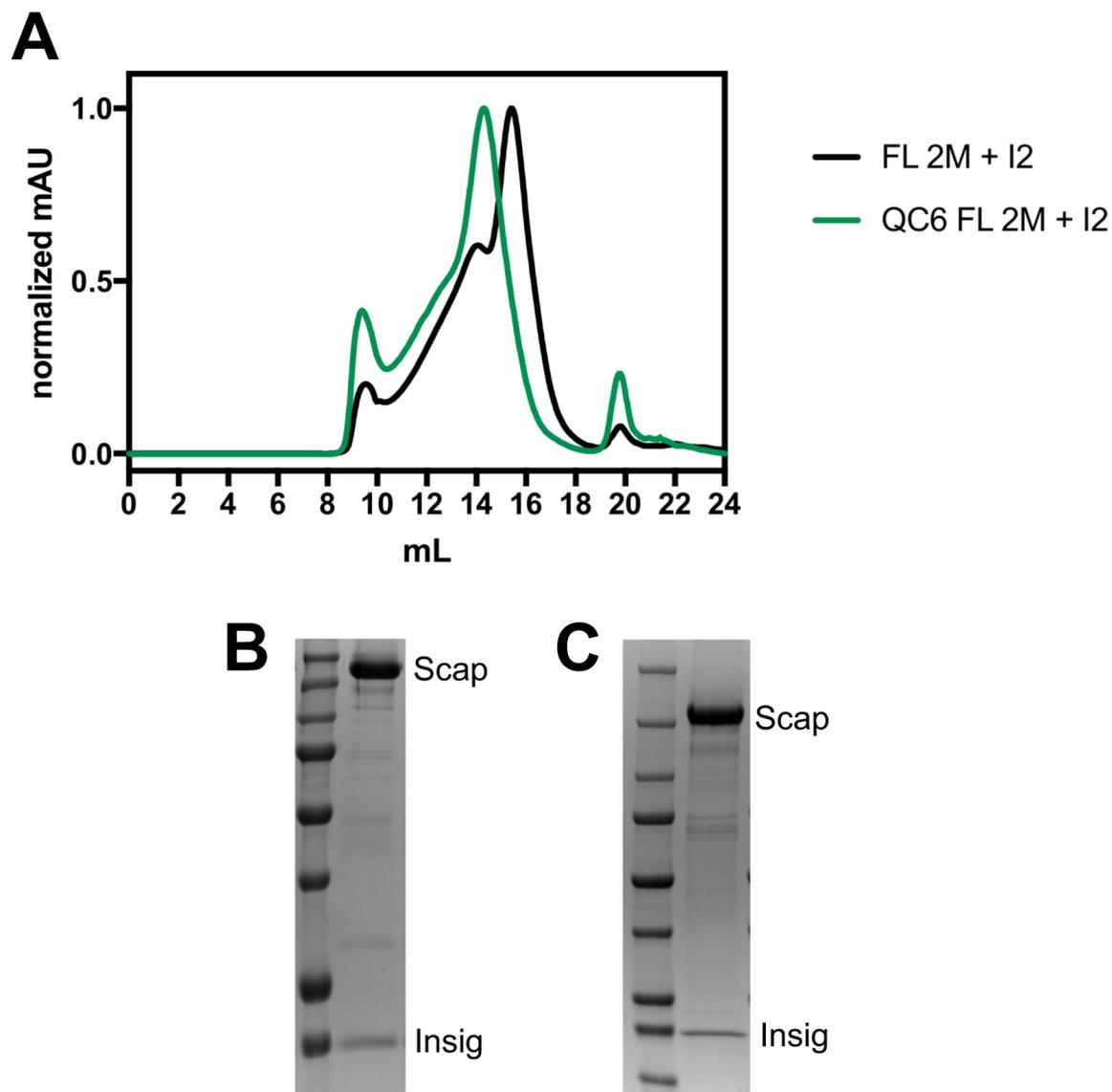


Figure 15. Purification of ScapFL 2M + Insig2 with and without N-terminal fusion protein. A) Normalized SEC traces of the Scap/Insig complex. Scap with the N-terminal fusion (aka QC6) elutes earlier on a Superose 6 10/300 column, but with less of a defined shoulder. ScapFL 2M + Insig2 (B) and QC6-ScapFL 2M + Insig2 (C) complexes shown by SDS-PAGE.

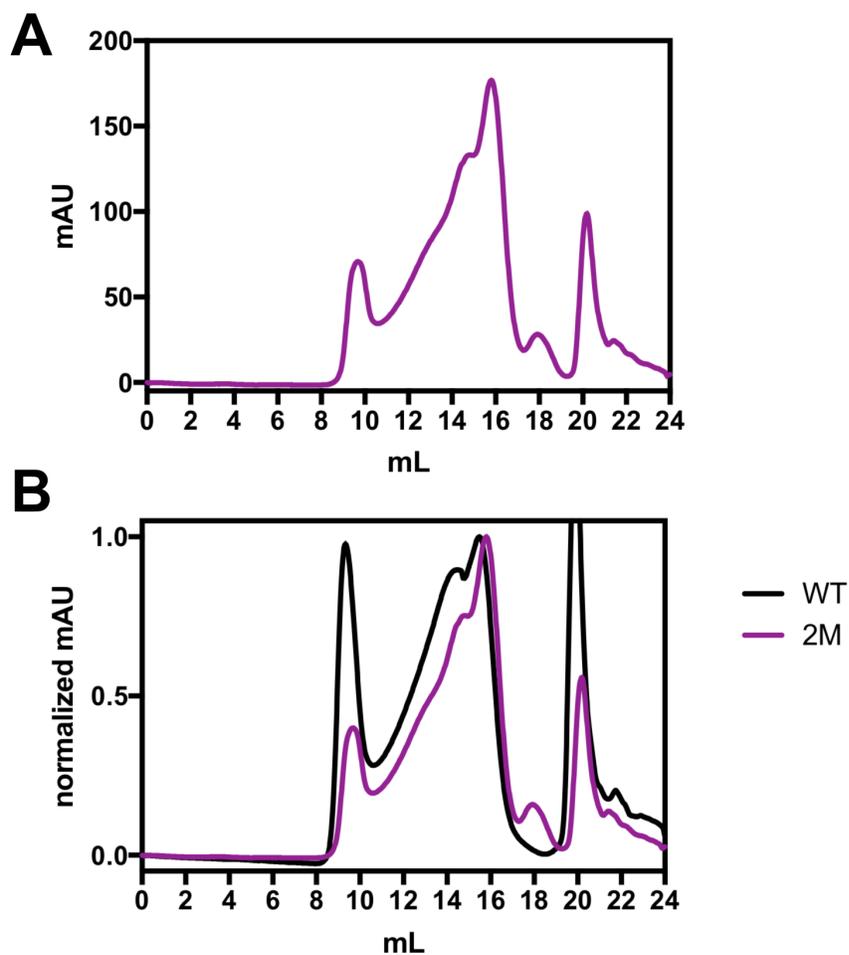


Figure 16. Comparison of ScapFL WT vs 2M. A) SEC trace of ScapFL 2M alone, displaying a significant shoulder but a defined monomeric peak. B) Normalized overlay of 2M shown in panel A with ScapFL WT. WT Scap is much more heterogeneous than 2M, exemplified by the high void:peak ratio and more defined left shoulder.

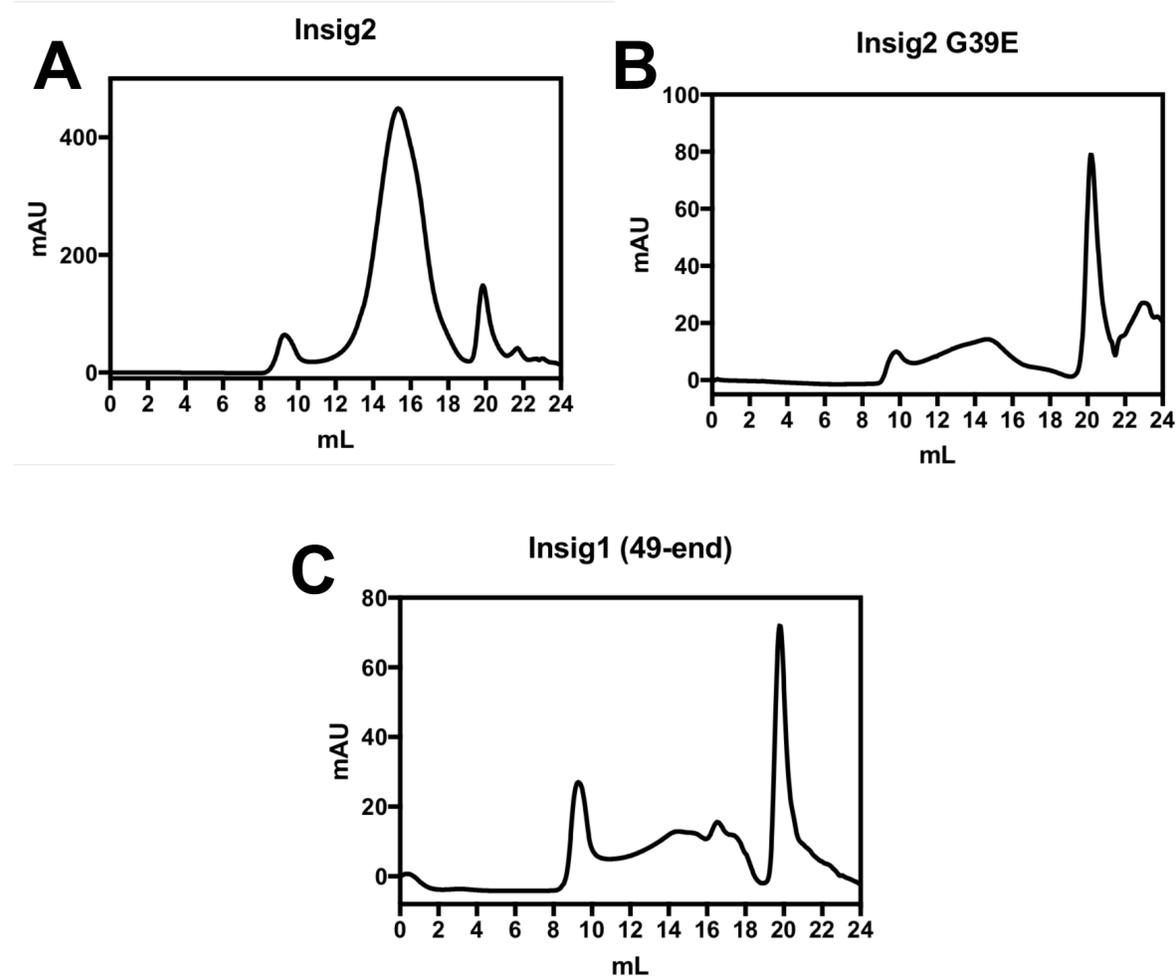


Figure 17. SEC traces of Insig. A) Insig-2 is biochemically pure but displays a high degree of heterogeneity due to its broad peak. B) G39E was previously described as a mutant of Insig-2 that reduced its oligomerization capability, but incorporation of this mutation drastically reduced expression level and still retained a broad distribution by SEC. C) Insig1 (lacking the N-terminal tail) also expressed at a low level, and did not behave as a single species by gel filtration.

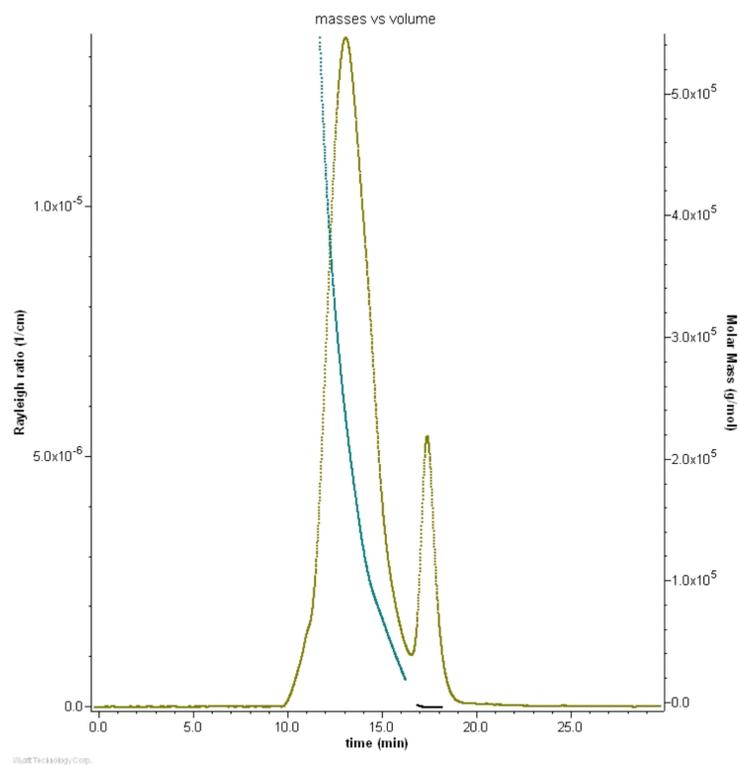


Figure 18. MALS of Insig-2 shown in panel A of Figure 17. The mass distribution (blue line) ranged from a monomer to at least a 20-mer, and displayed characteristics of a highly heterogeneous sample.

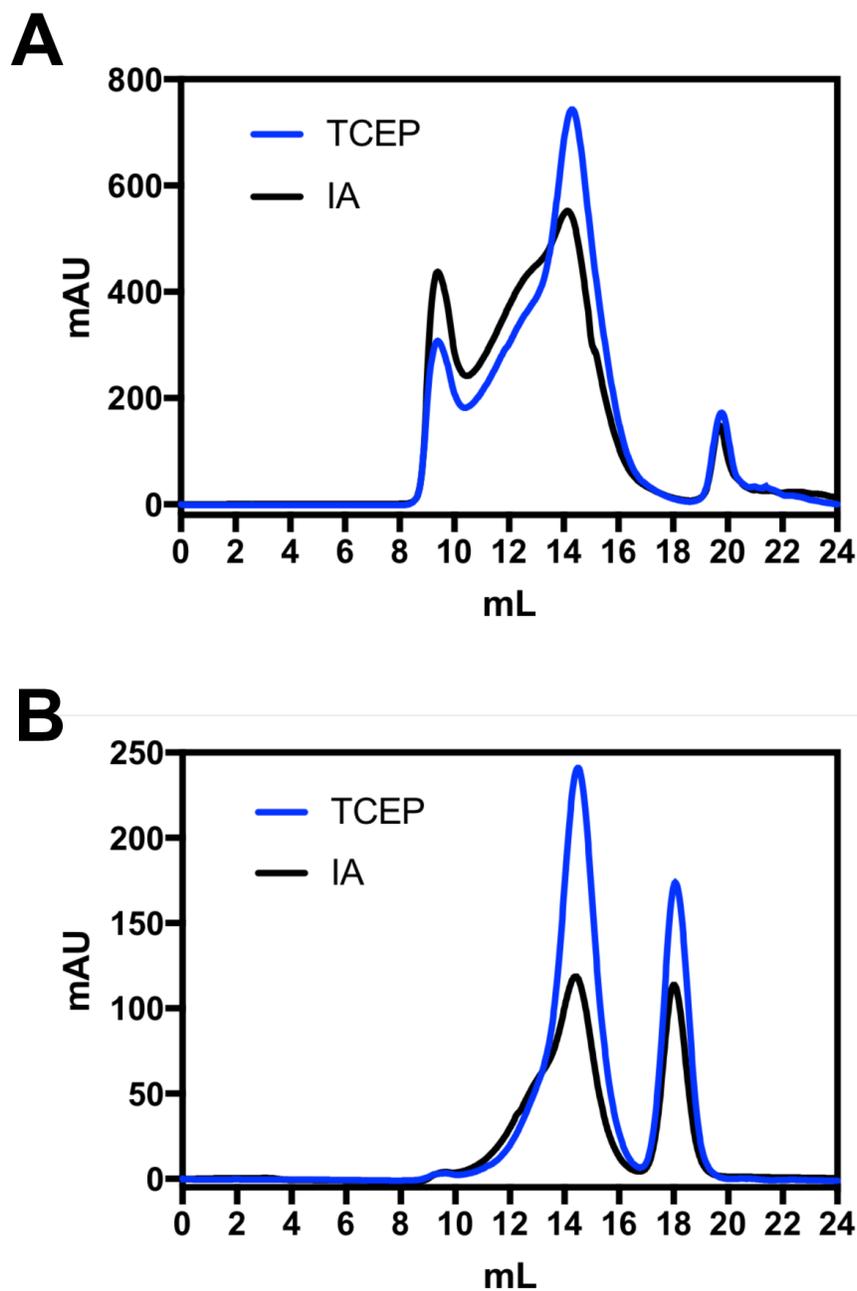


Figure 19. SEC traces of QC6-ScapFL 2M + Insig2. A) Overlay of LMNG-purified protein prepared with TCEP throughout the purification in blue, with protein capped with iodoacetamide shown in black. B) Overlay of protein from A) reconstituted into saposin nanodiscs. TCEP improves protein quality and reconstitution efficiency over iodoacetamide-treated samples.

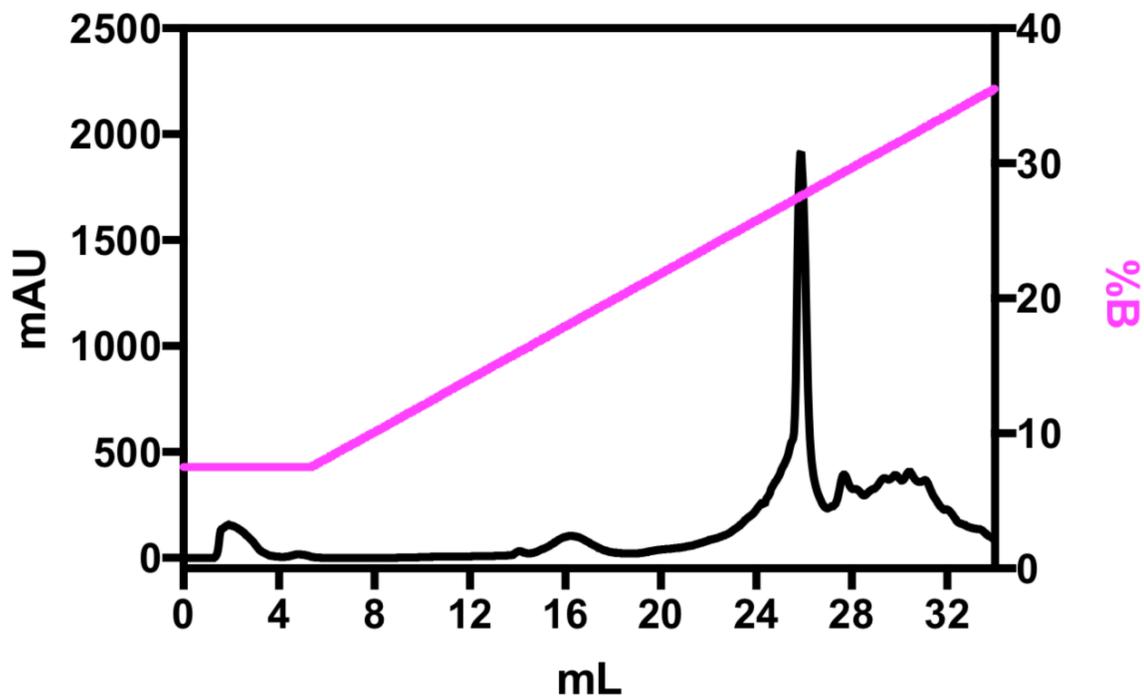


Figure 20. Ion exchange of ScapFL 2M + Insig 2 in LMNG. The complex elutes in a very sharp peak around 26-28% B (260-280mM NaCl).

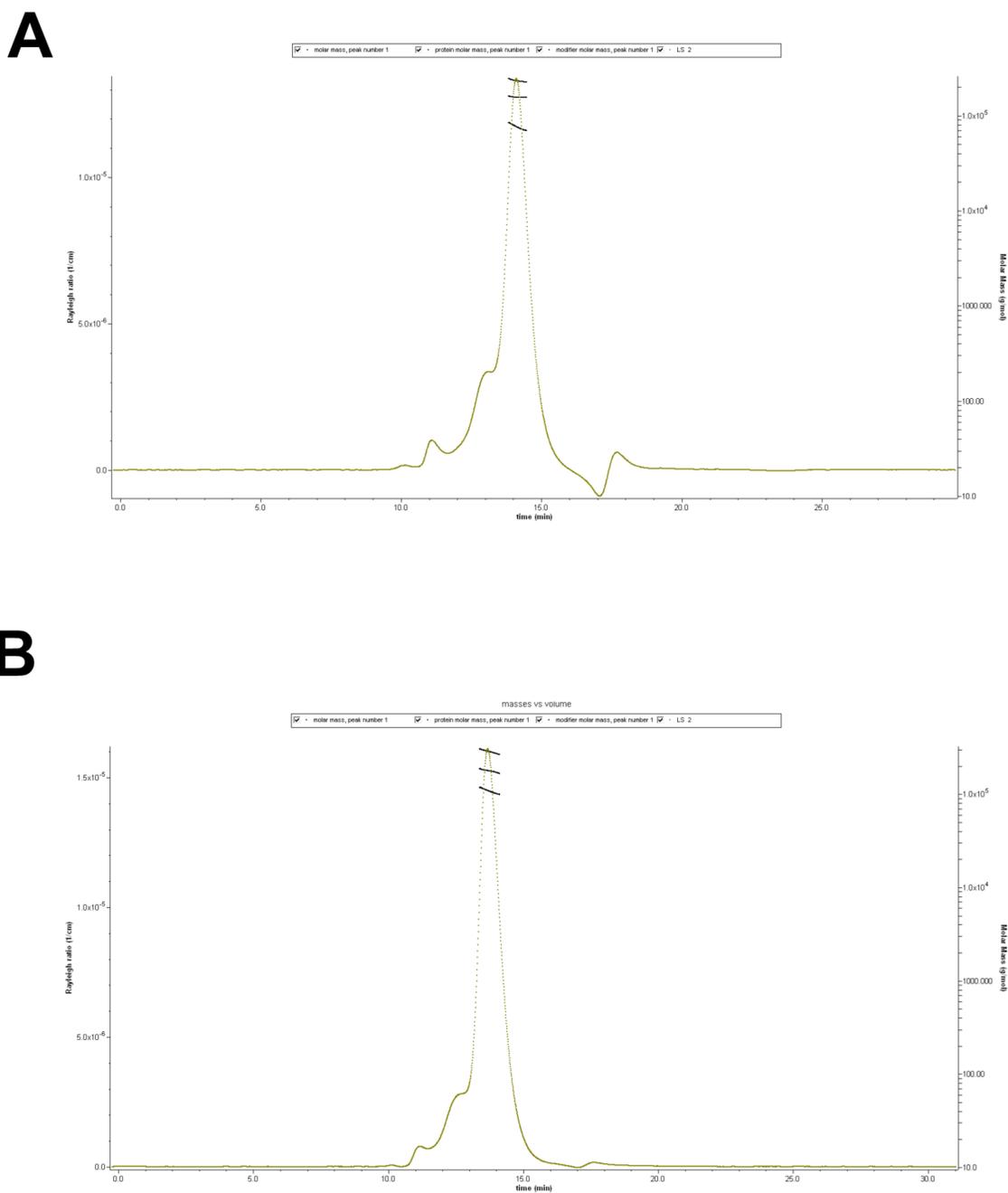


Figure 21. SEC-MALS data of ScapFL 2M alone (A) or ScapFL 2M + Insig 2 (B), both purified in LMNG. The mass ranges for A) unambiguously defines Scap as a monomer, and this mass is shifted by the mass for one bound Insig in B), confirming a 1:1 stoichiometry.

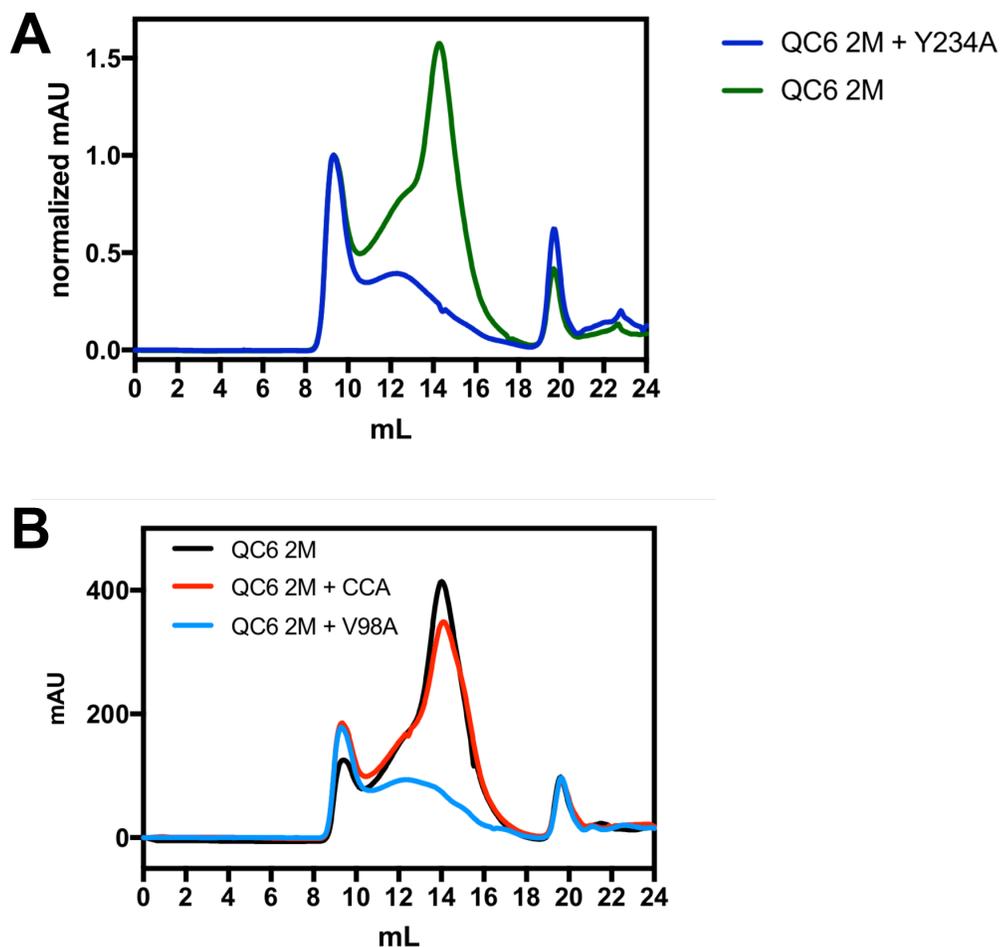


Figure 22. SEC traces of point mutations in QC6 ScapFL 2M + Insig2. A) Void-normalized traces of Scap/Insig complexes with or without the additional mutation of Y234A in loop 1. Y234A drastically decreased sample stability and yield. B) Unnormalized SEC traces of Scap/Insig complexes with or without C264A (“CCA”) or V98A mutations in loop 1. CCA showed a slight broadening of the peak, while V98A appeared to have a similar destabilized effect as Y234A.

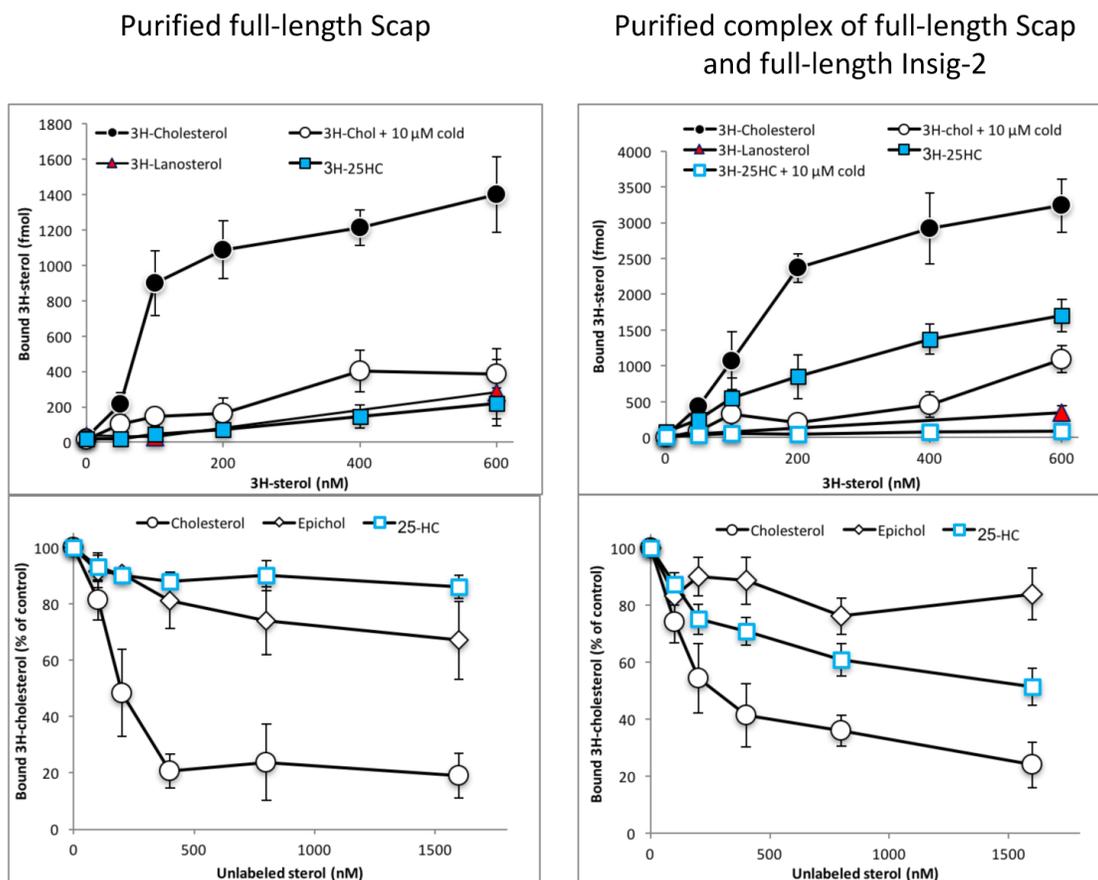


Figure 23. Saturation binding and competition assays of ScapFL 2M with or without Insig-2. Increasing amounts of tritiated sterols are incubated with samples in the top panels, with only appreciable binding demonstrated with cholesterol. Bottom panels contain data from competition binding experiments, where samples were pre-incubated with 200nM tritiated cholesterol, followed by increasing amounts of unlabeled sterols. Again, only cholesterol showed any appreciable competition behavior with Scap alone, but 25-hydroxysterol did demonstrate an intermediate ability to compete off cholesterol in the complex.

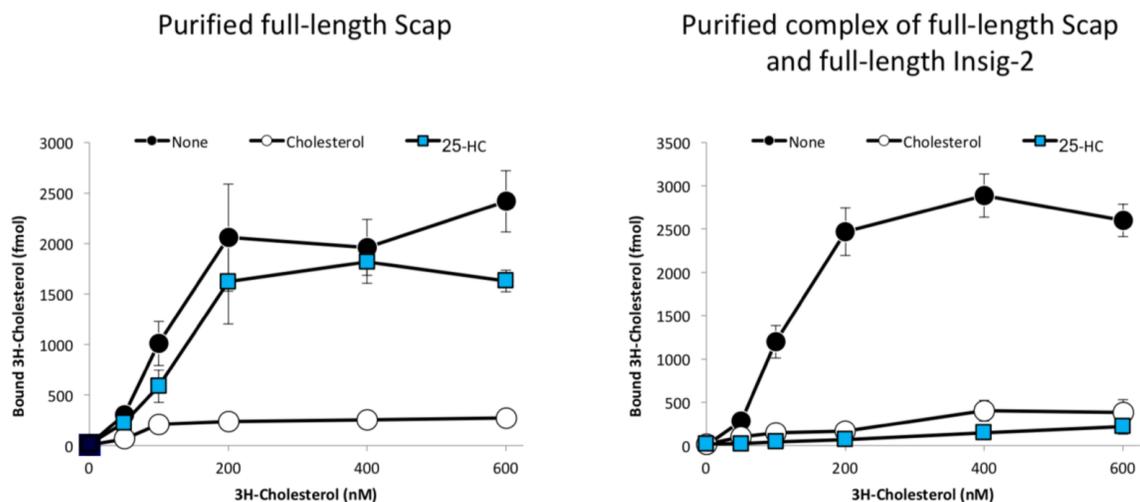


Figure 24. Additional binding assays with ScapFL 2M and ScapFL 2M + Insig-2. Samples were incubated overnight on ice with 5 $\mu$ M unlabeled sterols (from EtOH stocks), followed by a saturation binding assay using tritiated cholesterol. Scap alone is able to efficiently bind tritiated cholesterol if pre-incubated with 25-hydroxycholesterol or the EtOH vehicle. In contrast, the complex is only able to efficiently bind tritiated cholesterol if incubated with vehicle alone, demonstrating that pre-incubation with 25-hydroxycholesterol precludes efficient cholesterol binding in the complex.

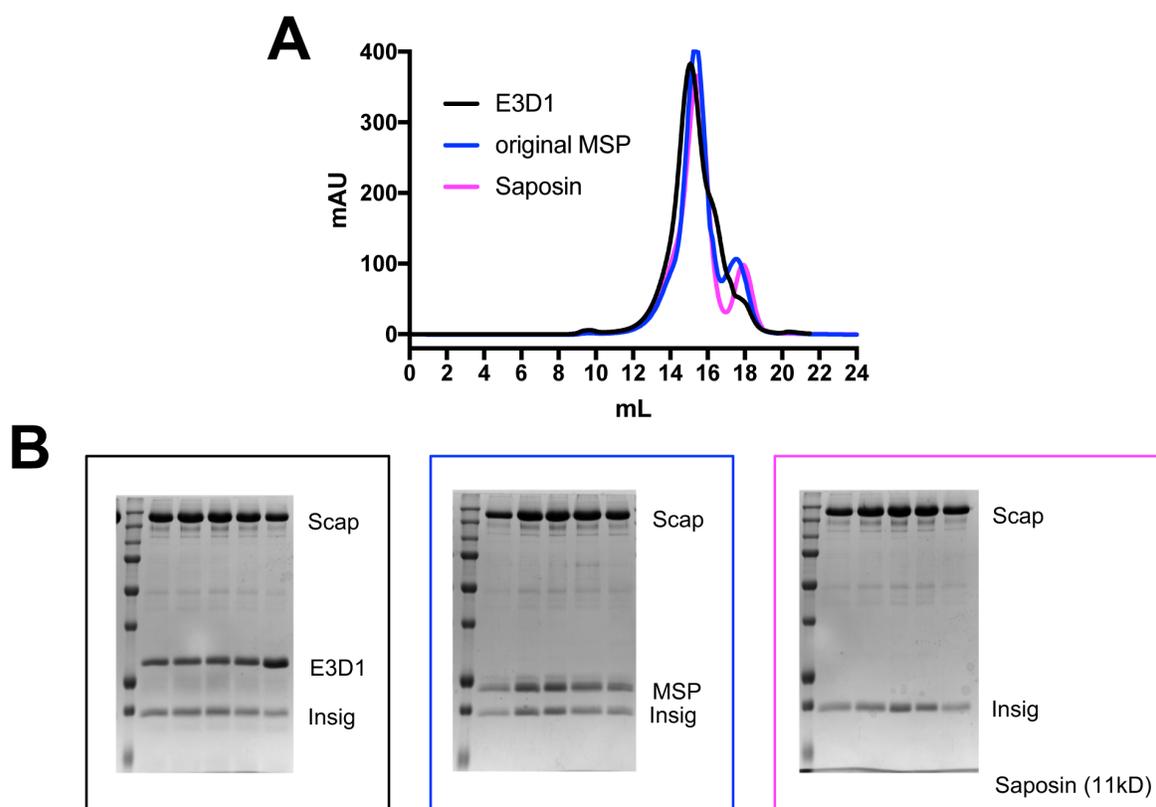


Figure 25. Reconstitution of ScapFL 2M + Insig-2 complexes into nanodiscs. A) Overlay of reconstituted samples injected in aqueous buffer over a Superose 6 10/300 column. The small peak to the right of the main peak contains empty/free nanodiscs. B) SDS-PAGE of fractions of the corresponding peaks in A). Saposin is 11kD and is not visible on the right-most gel.

```

1      10      20      30      40      50      60
Clone19 QVQLQESGGGLVQAGGSLRISCAASGNTSDLDTCMGWYRQAPKEREFAVGIISDGSSTHYVA
Clone21 QVQLQESGGGLVQAGGSLRISCAASGNTSDSSNMGWYRQAPKEREFAVGIIDGSSHYVA
Clone3   QVQLQESGGGLVQAGGSLRISCAASGTFYYPFGMGWYRQAPKEREFAVGIISDGTINVA
Clone5   QVQLQESGGGLVQAGGSLRISCAASGTFYYPFGMGWYRQAPKEREFAVGIISDGTINVA
Clone9   QVQLQESGGGLVQAGGSLRISCAASGTFSEFDSMGWYRQAPKEREFAVGIITGATINVA
Clone11  QVQLQESGGGLVQAGGSLRISCAASGTFSEFDSMGWYRQAPKEREFAVGIITGATINVA
Clone4   QVQLQESGGGLVQAGGSLRISCAASGTFSEIYLMGWYRQAPKEREFAVGIITAGSTINVA
Clone6   QVQLQESGGGLVQAGGSLRISCAASGTFSEYAYGMGWYRQAPKEREFAVGIIDAGATINVA
Clone8   QVQLQESGGGLVQAGGSLRISCAASGTFSEYAYGMGWYRQAPKEREFAVGIIDAGATINVA
Clone20  QVQLQESGGGLVQAGGSLRISCAASGTFSEPYIMGWYRQAPKEREFAVGIINVGATINVA
Clone2   QVQLQESGGGLVQAGGSLRISCAASGTFDIARMGWYRQAPKEREFAVGIIDDGSTINVA
Clone17  QVQLQESGGGLVQAGGSLRISCAASGTFNFGQSMGWYRQAPKEREFAVGIIDTGSNINVA
Clone13  QVQLQESGGGLVQAGGSLRISCAASGTFSPVTVMGWYRQAPKEREFAVGIISAGSTINVA
Clone16  QVQLQESGGGLVQAGGSLRISCAASGTFSSGYDMGWYRQAPKEREFAVGIIDHGSNINVA
Clone10  QVQLQESGGGLVQAGGSLRISCAASGTFSSGYDMGWYRQAPKEREFAVGIIDHGSNINVA
Clone18  QVQLQESGGGLVQAGGSLRISCAASGTFSRVEDMGWYRQAPKEREFAVGIITARGASINVA
Clone12  QVQLQESGGGLVQAGGSLRISCAASGTFDRFVVMGWYRQAPKEREFAVGIISPGSSINVA
Clone15  QVQLQESGGGLVQAGGSLRISCAASGTFSEAVMGWYRQAPKEREFAVGIITGANSINVA
Clone7   QVQLQESGGGLVQAGGSLRISCAASGTFSELDGMGWYRQAPKEREFAVGIITATGSSINVA
Clone1   QVQLQESGGGLVQAGGSLRISCAASGTFSEYTDGMGWYRQAPKEREFAVGIITNGSSINVA
Clone14  QVQLQESGGGLVQAGGSLRISCAASGNTFLRYQMGWYRQAPKEREFAVGIITVGGSSINVA

70      80      90      100     110
Clone19 DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVLLGGRWR.....THWYWGQGT
Clone21 DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAAAARHRYP.....RLTYWGQGT
Clone3   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVIGVPPDSDYP...LWYLYYWGQGT
Clone5   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVIGVPPDSDYP...LWYLYYWGQGT
Clone9   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVTFGVDP.....TLFYWGQGT
Clone11  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVTFGVDP.....TLFYWGQGT
Clone4   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVKVDHSD.....SYWYWGQGT
Clone6   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVYNDEHR.....YLYGWGQGT
Clone8   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVSKDE.....YPRLSYWGQGT
Clone20  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVSVTV.....YPRLAYWGQGT
Clone2   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVEKAW.....YPRLYYWGQGT
Clone17  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVSRDK.....YPRLVYWGQGT
Clone13  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVQRRTY.....YPRFQYWGQGT
Clone16  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVRRRTY.....YPRFVYWGQGT
Clone10  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVAGKWDYLFYDPAYYDWTYWGQGT
Clone18  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVRRFRD.....RPYFYWGQGT
Clone12  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVYRDR.....TGGFFYWGQGT
Clone15  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVHTWD.....TGGFFYWGQGT
Clone7   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVSTDQVP.....HWYWGQGT
Clone1   DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVATNYPQDLWG...YYLRYWGQGT
Clone14  DSVKGRFTISRDNKNTVYLOMNSLKPEDTAVYYCAVRHYIVYPKD...DRHYWYGGGT

120
Clone19 QVTVSSHHHHHH
Clone21 QVTVSSHHHHHH
Clone3   QVTVSSHHHHHH
Clone5   QVTVSSHHHHHH
Clone9   QVTVSSHHHHHH
Clone11  QVTVSSHHHHHH
Clone4   QVTVSSHHHHHH
Clone6   QVTVSSHHHHHH
Clone8   QVTVSSHHHHHH
Clone20  QVTVSSHHHHHH
Clone2   QVTVSSHHHHHH
Clone17  QVTVSSHHHHHH
Clone13  QVTVSSHHHHHH
Clone16  QVTVSSHHHHHH
Clone10  QVTVSSHHHHHH
Clone18  QVTVSSHHHHHH
Clone12  QVTVSSHHHHHH
Clone15  QVTVSSHHHHHH
Clone7   QVTVSSHHHHHH
Clone1   QVTVSSHHHHHH
Clone14  QVTVSSHHHHHH

```

Figure 26. Sequence alignment of identified nanobody clones. 21 individual colonies tested positive after iterative rounds of selection, and subsequent subcloning and sequencing revealed 19 unique sequences (clones 3 and 11 are duplicates of 5 and 9, respectively). Consensus sequences are notated in red with variable regions in the CDR loops are not colored.

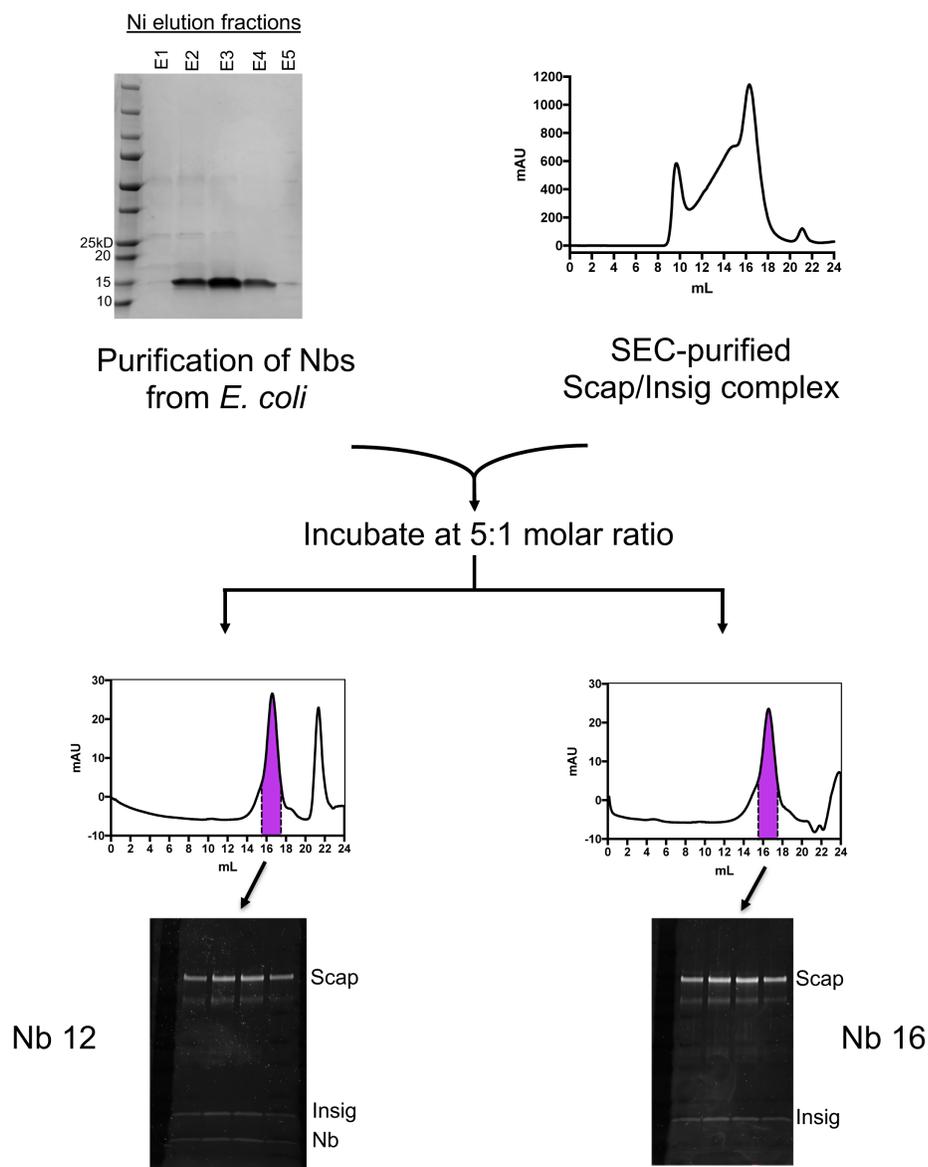


Figure 27. Nanobody screening overview. After iterative rounds of FACS selection, individual nanobody clones were expressed in *E. coli* and purified with IMAC. Purified nanobodies are incubated with purified complex, and then reinjected over SEC. Peak fractions are examined by SDS-PAGE followed by Sypro Ruby. Nb 12 shows co-elution of all three components, where Nb 16 does not show any Nb binding.



Figure 28. Sypro Ruby stains of all 19 clones screened in the manner described in Figure 27. Promising nanobodies are denoted with a green asterisk, with weak binders shown with an orange asterisk.

## **CHAPTER EIGHT**

### **Cryo-EM of Full-Length Scap/Insig Complex**

#### **INTRODUCTION**

The previous chapter outlined years of work to optimize the first full-length purification of Scap, as well as the first purification of the Scap/Insig complex. The main motivation behind this optimization is to obtain a sample that is amenable to structural determination. Success with X-ray crystallography is unlikely due to the large, flexible domains outside of the membrane region, and no crystallization hits were achieved in either vapor diffusion or lipid-mediated crystallization trials. Cryo-EM has become the new go-to structure determination method for membrane proteins, and has shown great promise for challenging targets. While this is currently an active work in progress, this chapter will detail our best efforts to date to obtain a structure of the Scap/Insig complex by cryo-EM.

#### **GRID PREPARATION**

After the final gel filtration step, concentration should be minimized (or avoided altogether) as this increases the amount of protein aggregates, even after a high-speed spin prior to preparing grids. Following an ideal reconstitution into saposin or MSP-based nanodiscs, the sample should be around 0.5 mg/mL

immediately off of gel filtration. Samples frozen in detergent typically are in the range of 0.8-1.2 mg/mL after gel filtration. In either case, the peak should be collected in small (~300  $\mu$ L) fractions and then double-checked with an  $A_{280}$  measurement to determine the fraction with the highest protein concentration. After this, the sample is immediately frozen onto grids with typically less than five minutes between elution and plunge freezing. This coordinated effort is usually a two-person job, with myself typically handling the gel filtration and Dan Rosenbaum or Xiaochen Bai preparing the ethane container and getting equipment prepared and ready. All of our samples were frozen on Quantifoil R1.2/1.3 300-mesh gold holey carbon grids, glow discharged immediately prior to freezing for 60-90 seconds at 30 mA. 3  $\mu$ L of sample was applied to grids and were plunge-frozen into liquid ethane with a Mark IV Vitrobot. While conditions varied slightly for each sample, grids were blotted for 4-5 seconds with approximately a +20 blotting force. Grids were frozen in multiples and stored in liquid nitrogen for future use. Representative grids from each batch were clipped into autogrids and screened on the Talos Arctica microscope in low dose mode for rapid assessment of grid quality.

## **DATA COLLECTION**

Our current best sample is Hamster ScapFL 2M complexed with Hamster Insig-2, reconstituted into saposin-based nanodiscs containing

POPC:POPE:cholesterol at a ratio of 3:1:1. Micrographs were acquired on our Titan Krios microscope, operating at 300 kV and equipped with a K2 Summit direct electron detector and a GIF-Quantum energy filter with a slit width of 20 eV. Screening images showed that our best particles came from grid squares with thicker (but still collectable) areas of ice, and those areas were selected from the grid for data collection. Particle contrast was greatly improved by using the Volta phase plate, which introduces a phase shift as the electron beam passes through the carbon film. The best contrast is achieved when the phase shift is 90°, but is still effective at shifts from 60°-110° (Danev *et al.*, 2016), which was a range we utilized for phase plate data collection. In addition, we also used a defocus value of -0.5 μm to assist with contrast transfer function correction during data processing steps as described in Danev *et al.* 2017. EPU software was used for automatic data collection over 2.5 days, with three micrographs collected per hole. Micrographs were collected at 165,000x magnification (1 pixel = 0.84 Å), using spot size 9 and a 0.6 μm illumination area. Each micrograph was dose-fractionated into 30 frames over a 15 second exposure with a dose rate of 5-6 e-/Å<sup>2</sup>.

## CURRENT MODEL

After data collection, the dose-fractionated movies were fed into the MotionCor2 program to subtract our any beam-induced motion (Zheng *et al.*, 2017),

followed by the program GCTF to estimate CTF parameters of each micrograph (Zhang, 2016). Micrographs (see Figure 1 for an example micrograph) were then imported into RELION and processed following the general pipeline We used RELION (Scheres, 2016) to process corrected micrographs. Initial 2D classes were generated from a small subset of particles (~1,000), and used as picking templates for automated particle picking on ~3,000 micrographs. Autopicked particles were subjected to iterative rounds of 2D classification, discarding classes containing bad particles after each round. 2D classes showed density in the TM region that correspond to TM helices, a very promising sign for membrane proteins that can often be lacking clear density in this region. Outside of the TM region, two features are clearly visible. A strong globular density is visible on one side of the TM which we can likely attribute to the WD domain. On the same opposite face of the TM, there is a fuzzier density that we hypothesize represents signal from the luminal loops. Both of these densities appear to be arising from the same region of TM density, suggesting that Scap is contained within one half of the TM region while Insig, which has very little density outside of the TM region, is contained within the other half of the TM region. Representative 2D classes are shown in Figure 1.

Approximately 800,000 particles were then used for 3D classification, which generated eight classes (Figure 2). Upon further inspection of these classes, only one out of the eight contained adequate density for the TM helices. We expect to have 14 TM helices (eight from Scap and six from Insig), but given the noisiness of

the map, we can only distinguish about 12 helices, although some of the density is large enough to potentially contain two helices. Density for the presumed WD domain is quite heterogeneous between each 3D class. Since the WD domain is our most prominent feature, its significant flexibility of the WD domain compared to the rest of our molecule is limiting our ability to correctly align particles and is likely can be blamed for our poor resolution. Further attempts at 3D classification did not improve our model.

Nevertheless, the single 3D class still contained ~12% of our particles and was subjected to 3D refinement which generated a final model rising from ~55,000 particles. The final model is shown in Figure 3. Our presumed WD domain is sitting above one side of the TM region, but with no connecting density visible. The linker between TM8 and the WD domain contains a palmitoylation site, but is still likely quite flexible and it is not surprising that that is not seen in our relatively low-resolution model. Contrary to the MCU structure that is solved with saposin nanodiscs (Nguyen *et al.*, 2018), our saposin belt appears more disordered, with a potential “bald spot” where no peptide is visible covering the lipid bilayer (Fig. 3, panel C). While it is possible that this “bald spot” is simply a result of our poor alignment, it has prompted future efforts to revisit MSP-based nanodiscs in case the geometry of our system is not compatible with saposin (see next chapter). While we don't see any resolvable density for the luminal loops, loop 1 and loop 7, it appears that there is some density extending from the TM region and possibly additional

density packed against the lipid bilayer. This is consistent with previous studies of loop 1 (see Chapter 6) that demonstrate its interaction with the membrane, and its amino acid sequence contains many hydrophobic regions that could be directly interacting with lipids.

Based on the lack of useful structural data, we currently do not have very reliable docking models. Our best options are 1) a computational model of the SSD, stemming from crystal structures of known SSD-containing proteins, 2) the mycobacteria structure of Insig (Ren *et al.*, 2015), and 3) the crystal structure of the *S. pombe* WD domain (Gong, *et al.*, 2015). There are multiple caveats with using these as docking models. While Scap contains an SSD, our 2M construct contains mutations that could potentially change the arrangement or angle of this region compared to WT structures of other SSDs. The mycobacteria Insig shares a low homology with hamster Insig, and displays conflicting oligomeric behavior compared to our sample. And lastly, the yeast WD domain also shares a low level of homology with the hamster WD domain, and is predicted to form a different number of beta repeats within the domain. However, we attempted to dock these models into our density, and those results are shown in Fig. 3. While certainly not close to an accurate fit for any of the individual pieces, it seems like the approximate sizes for these proteins is consistent, and a higher resolution model will likely guide future efforts and clarify differences between our structure and these docking models.

The molecular weight of Scap and Insig is ~172kD, plus the added mass of the saposin and lipid belt, yet we only can resolve about 100kD of density making it a very small particle for even the most cutting-edge cryo-EM software and analyses. Even with this small density, our current model is calculated at a resolution of ~6 Å using the gold-standard Fourier Shell Correlation (FSC) = 0.143 criterion. This is still an active work in progress, and our ultimate goal is to improve this resolution to a point where we can at least confidently trace and assign the helices. Current efforts to overcome our poor particle alignment have been detailed in the previous chapter, and a preliminary model of the complex containing the N-terminal PGS fusion protein is shown in Figure 4. While this resolution is low, efforts are currently ongoing to improve this resolution, and further optimization strategies will be briefly reviewed in the following conclusions chapter.

## REFERENCES

- Danev, R., & Baumeister, W. (2016). Cryo-EM single particle analysis with the Volta phase plate. *eLife*, 5. <http://doi.org/10.7554/eLife.13046>
- Danev, R., Tegunov, D., & Baumeister, W. (2017). Using the Volta phase plate with defocus for cryo-EM single particle analysis. *eLife*, 6. <http://doi.org/10.7554/eLife.23006>
- Gong, X., Li, J., Shao, W., Wu, J., Qian, H., Ren, R., et al. (2015). Structure of the WD40 domain of SCAP from fission yeast reveals the molecular basis for SREBP recognition. *Nature Publishing Group*, 25(4), 401–411. <http://doi.org/10.1038/cr.2015.32>
- Ren, R., Zhou, X., He, Y., Ke, M., Wu, J., Liu, X., et al. (2015). PROTEIN STRUCTURE. Crystal structure of a mycobacterial Insig homolog provides insight into how these sensors monitor sterol levels. *Science*, 349(6244), 187–191. <http://doi.org/10.1126/science.aab1091>
- Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3), 519–530. <http://doi.org/10.1016/j.jsb.2012.09.006>
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology*, 193(1), 1–12. <http://doi.org/10.1016/j.jsb.2015.11.003>
- Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4), 331–332. <http://doi.org/10.1038/nmeth.4193>

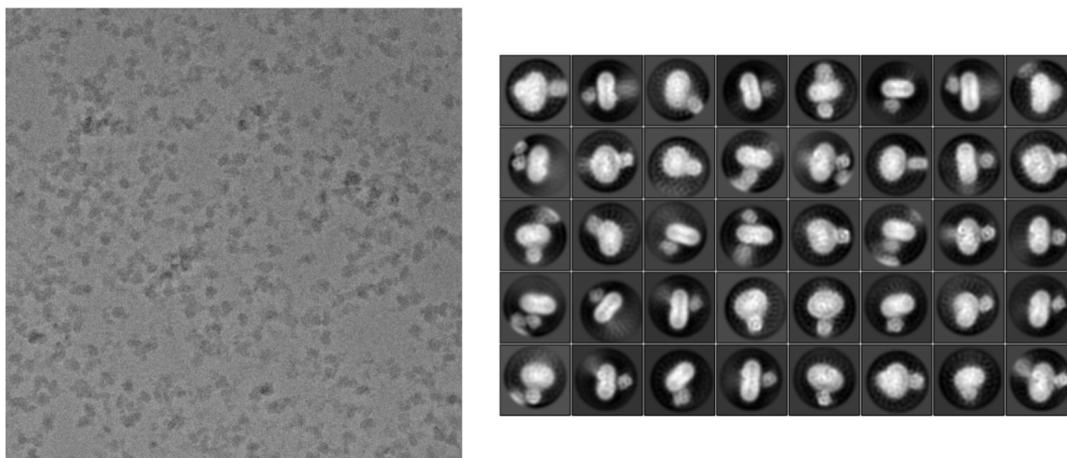


Figure 1. Left: example micrograph of ScapFL 2M + Hamster Insig2, taken on the Titan Krios microscope using a Volta phase plate. Right: Representative 2D classes from phase plate-collected dataset, showing density for TM helices and strong density for globular WD domain.

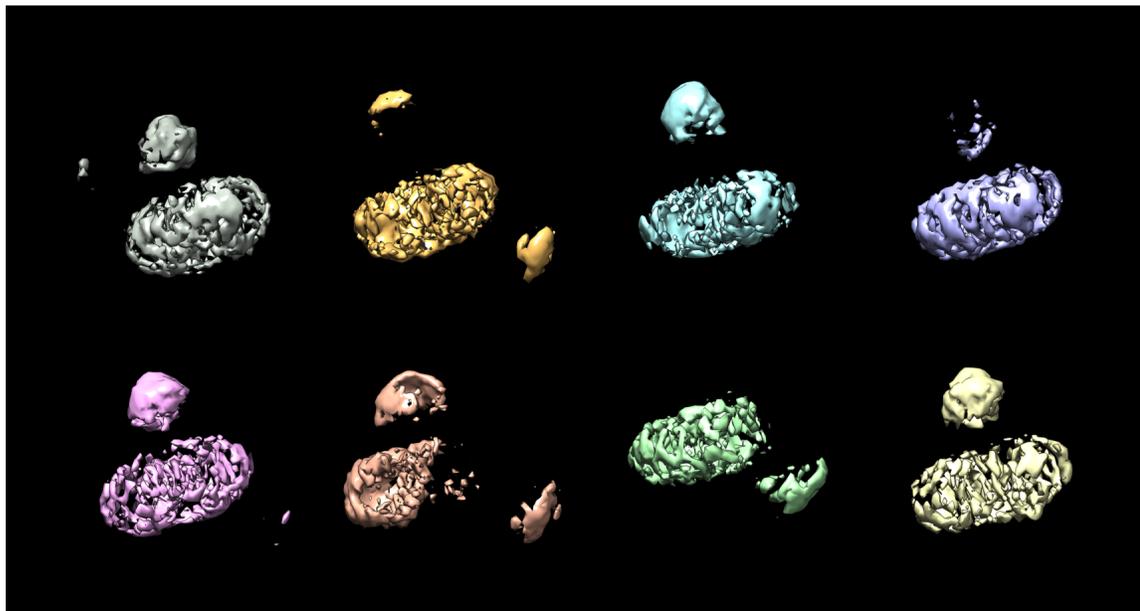


Figure 2. Representative 3D classes arising from 2D classes shown in Figure 1. The only class that showed strong TM density is the class on the bottom right, containing 12% of the total particles.

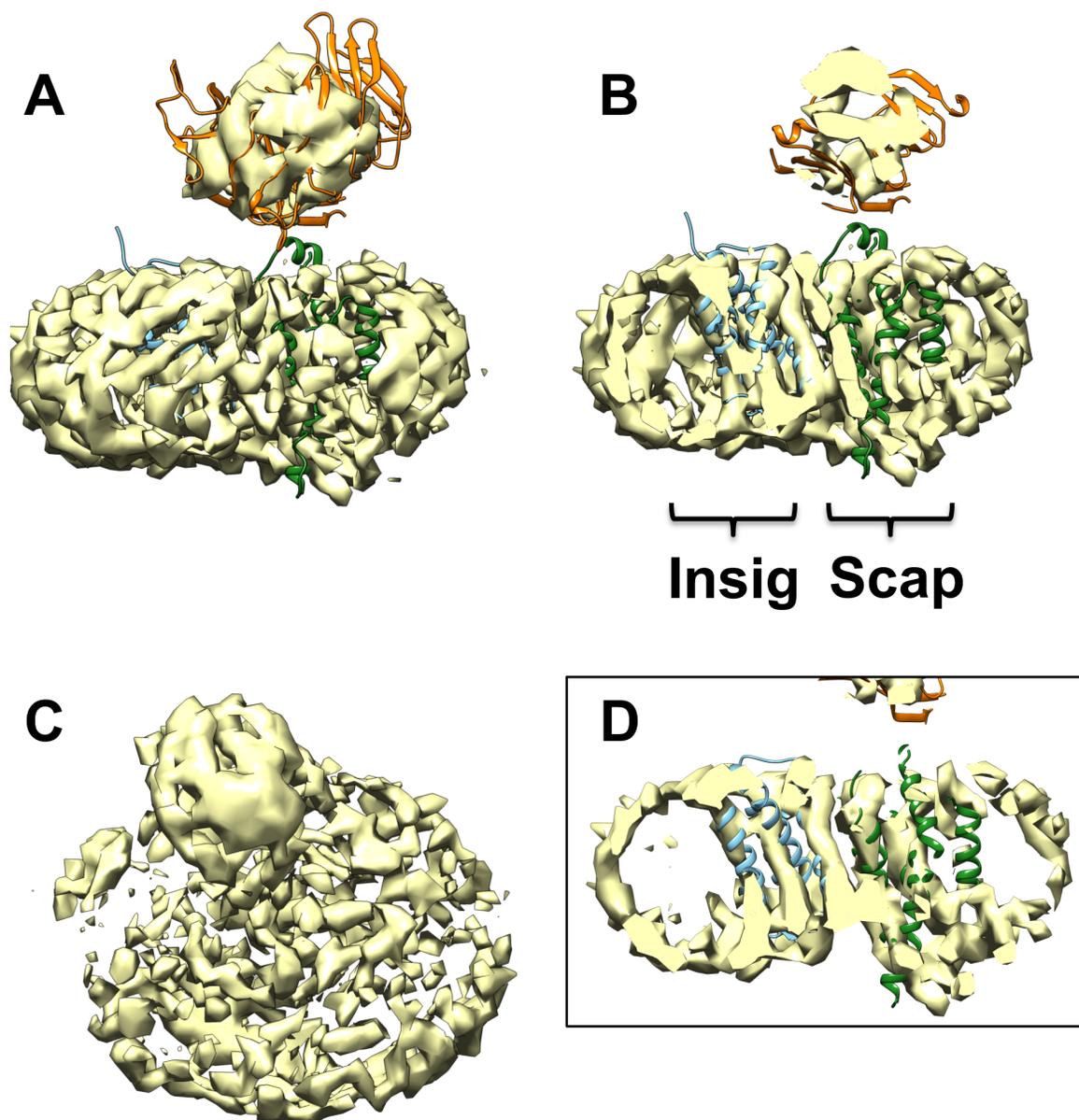


Figure 3. Final best model of Hamster ScapFL 2M + Hamster I2 in saposin nanodiscs. A) Overall view of model. B) Cutaway of A to show docked models of Insig in light blue, and Scap in green. The WD domain is shown in orange. C) Top view of A to show the saposin “bald spot” that could be limited resolution. D) Lower contour cutaway image of B to show TM densities.

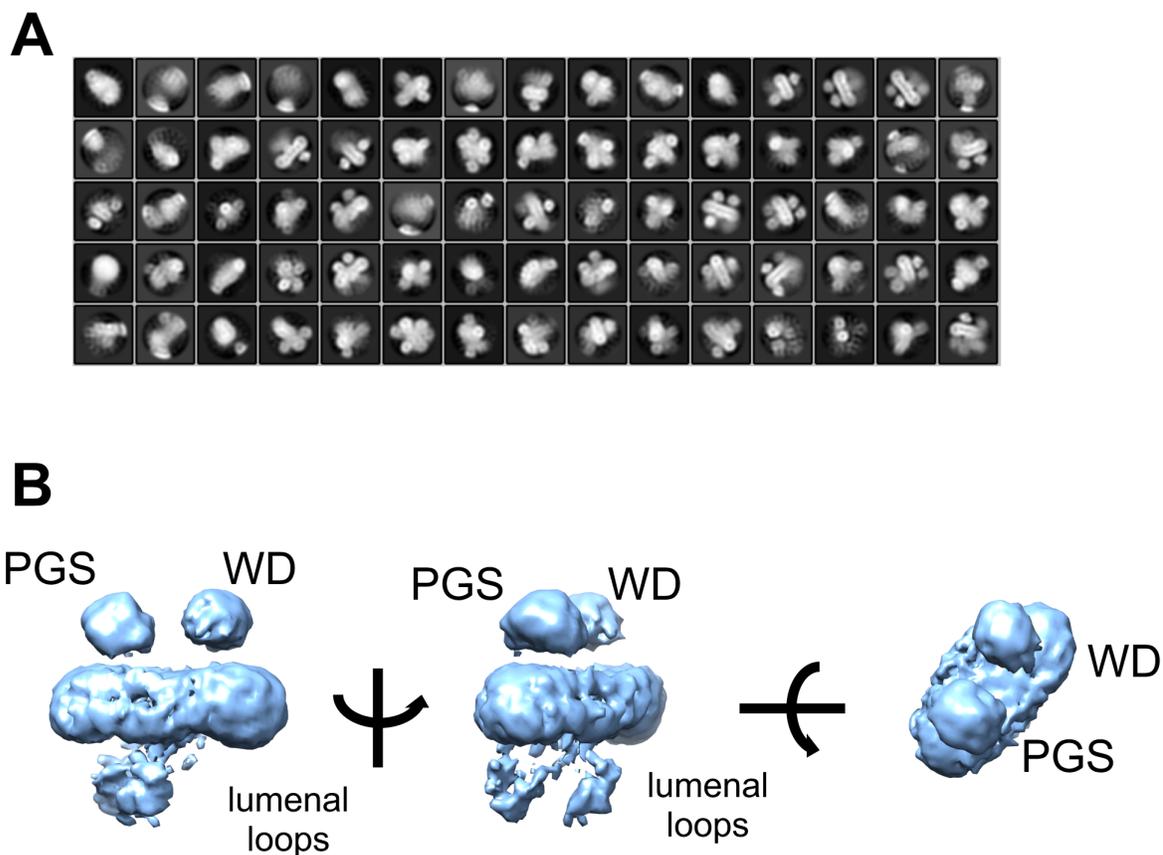


Figure 4. Preliminary cryo-EM data from Hamster ScapFL 2M + Hamster Inisig-2 containing an N-terminal PGS fusion protein prior to TM1. A) Representative 2D classes after autopicking that show structural heterogeneity in the TM region, but an additional density attributed to the PGS domain. B) Low resolution 3D classification of 2D classes from A). Due to improved particle classification and alignment, our 3D classes now contain some weak density attributable to the luminal loops, as well as globular density corresponding to the PGS domain and WD domains on the other side of the TM region.

## **CHAPTER NINE**

### **Conclusions and Future Directions**

#### **NMR STUDIES OF GPCRS**

##### **Conclusions**

G Protein-Coupled Receptors are the largest family of membrane proteins and are the target of approximately a third of all pharmaceuticals in development. Due to this, having a comprehensive understanding of how these molecules respond to their ligands is crucial for the drug development process. Our current understanding of how GPCRs work is largely guided by landmark crystal structures that illustrated how ligands bind to receptors and the conformational rearrangements required to bind and signal through G proteins. However, these crystal structures represent the endpoints of an activation pathway and an increasing amount of research has demonstrated that these receptors have a diverse energy landscape and can adopt multiple conformations that are not easily captured during crystallization. In addition, while there have been studies measuring large conformational changes occurring during receptor activation, there have been no studies addressing the changes in dynamics occurring at the side chains within these receptors.

NMR spectroscopy is a useful technique that can be used to study transient states in molecules as well as site-specific information about dynamics occurring at

multiple timescales. In this thesis, I have outlined the work I undertook to enable NMR studies of GPCRs, and specifically, the first measurements of fast (psec-nsec) sidechain dynamics in a receptor. These studies relied on engineering a method to express, and purify receptors that are  $^{13}\text{C}$ -methyl labeled at specific sites within the protein, with every other atom remaining NMR silent. Described in chapter 2, I determined that I could use the methylotrophic yeast *Pichia pastoris* as my expression system, and by systematically adapting the yeast to increasingly deuterated media followed by supplementation of  $^{13}\text{C}$ -labeled  $\alpha$ -ketobutyric acid, I could achieve ~55%  $^{13}\text{C}$ -Ile $\delta$ 1 methyl and >90%  $^2\text{H}$  labeling efficiencies. This method was first established using maltose binding protein (MBP) as a test case, and subsequently validated with the cytoskeletal protein actin.

With the labeling capabilities in hand, I then turned to using this method to GPCRs (described in chapters 3 through 5). Other groups that have used NMR to study GPCRs have relied on the introduction of thermostabilizing mutations to enable high temperature data collection and improve overall signal-to-noise. However, the information learned from these studies is quite limited, as these mutations have been shown to significantly perturb native function of the receptors, namely basal levels of activation. As a result, any dynamics measurements would require subsequent validation to claim physiological relevance. We wanted to avoid any such questions and study a native, wild-type system, and through screening of ~10 WT GPCRs, we chose the human adenosine A2A receptor as our model

system. Through collaboration with Kevin Gardner and his postdoc Igor Dikiy, we were able to collect the first 2D spectra on this receptor in various liganded states and assign four functionally important isoleucine residues to monitor as probes in dynamics studies. We collected two dynamics datasets of A2A in the presence of an agonist or antagonist, and noticed some interesting differences. While sample stability limited us to a specific condition (sodium-containing buffer, detergent micelles), we could still tease apart differences in side chain dynamics as a readout of the ligand bound.

In general, the receptor was more flexible in the presence of agonist compared to antagonist, and this is in agreement with previously published studies monitoring fluorescence or hydrogen-deuterium exchange. One interesting difference was at isoleucine 292, which sits at the end of TM7 and is part of the interface with the Gs heterotrimer. When A2A was bound to ZM241385, an inverse agonist, the dynamics were highly attenuated, more so than any other monitored isoleucine. This finding could begin to shine some light on connecting side chain dynamics to ligand efficacy, as the inverse agonist reduces basal activity levels of receptors, and this could be a result of reducing dynamics at that potential interface. While further studies are needed to follow up on this observation, this represents the first step towards understanding how dynamics could be playing a part in receptor activation.

## Future Directions

The strategy outlined in this thesis combined with NMR techniques geared towards large samples (e.g. methyl-TROSY) have the potential to permit NMR studies of many large membrane proteins that have long been considered “off-limits” to NMR due the lack of a suitable expression system or isotopic labeling strategy. I demonstrated the applicability of this method to additional GPCRs in chapter 5 – the cannabinoid receptor type 1 and the orexin receptor type 2. Both of these receptors have immense pharmacological importance and currently have drug candidates in development. With the expression and purification already demonstrated, the obvious next steps in this project would be to carry out the same type of analysis on these two receptors and determine if the same patterns of dynamics exist. Furthermore, additional studies on any three of these receptors could be carried out with different types of ligands (e.g. partial agonists, neutral antagonists) that weren’t included in the A2A study and further expand our understanding of how ligand efficacy is tied to dynamics. And while more optimization will likely be required to achieve this, obtaining a parallel relaxation dataset of G-protein bound receptors would be informative. The full G protein heterotrimer is ~90kD, which is not ideal for NMR due to the added bulk weight and reduced tumbling. A promising alternative to the full heterotrimer is to use the recently developed mini-G proteins, which can be expressed in deuterated *E. coli* cultures and complexed with purified receptors.

All of our studies on A2A, CB1, and OX2R have relied on the use of deuterated DDM micelles as a membrane mimetic. A2A is able to be reconstituted into nanodiscs containing deuterated lipids, but the reconstitution efficiency and resultant signal-to-noise was prohibitively low. With the increased popularity of nanodiscs and SMA-based particles, this efficiency could likely be improved upon. It has been shown that GPCRs can be strongly influenced by their lipid environment and it would be incredibly informative to compare data taken on a detergent-solubilized receptor versus a receptor in a lipid bilayer (either reconstituted or native).

And lastly, our method could easily be applied to any membrane protein that can be expressed in *Pichia*. During my PhD, I have also been involved in the structural characterization of the sterol transporter G5G8, an ABC transporter that was first structurally characterized by our lab. G5G8 is readily expressed and purified from *Pichia* cultures, and although it is not included in this thesis, I have shown that it is possible to reconstitute into MSP-based nanodiscs for cryo-EM analyses leading to resolutions of  $\sim 5$  Å or better. Given this success and general stability, it could likely be utilized for NMR experiments. The large size of G5G8 ( $\sim 150$  kD + the membrane mimetic) will require a methyl-TROSY based approach but it is certainly a feasible and promising candidate for future NMR experiments.

## STUDIES OF CHOLESTEROL HOMEOSTASIS MACHINERY

### Conclusions

Cholesterol levels are tightly regulated to ensure membrane integrity and proper cell function. The SREBP pathway is tasked with this responsibility in mammals, relying on the protein Scap to both sense cholesterol levels and respond accordingly by escorting SREBP to the Golgi for proteolytic processing of its transcription factor domains. Due to this importance, Scap has been a long sought-after structural target for both understanding the mechanism of this process as well as for pharmacological targeting. Scap is a large, polytopic membrane protein that exists in the ER membrane and contains large, flexible domains extending into the cytoplasm and ER lumen. Scap has been shown to adopt multiple conformations based on the sterol content of the ER membrane, and it is likely that that inherent flexibility has contributed to how challenging it has been to purify over decades of effort.

During my PhD, I have devoted the majority of my efforts into overcoming this challenge and developing the first purification of full-length Scap, as well as the first purification of the complex of Scap and Insig. This required multiple approaches and technological developments in construct design, orthologue screening, functional assays, testing of various expression systems, thermostability screening, and purification optimization, detailed in chapter 7. The ultimate goal of this project is to

obtain the first structural information on mammalian Scap, and our initial efforts were focused on a divide-and-conquer approach, which is commonly taken with large, modular structural targets. However, as described in chapter 6, this approach has been largely unsuccessful. These years of effort were not totally futile though, as they generated valuable information about individual domains of Scap that have been useful data points for subsequent studies with the full-length protein.

While this is largely still a work in progress, our most promising construct is the Hamster ScapFL 2M complexed with Hamster Insig-2. As discussed in chapter 8, we have collected our best dataset on this construct in a saposin-based reconstitution using a lipid mixture of POPC:POPE:cholesterol at a 3:1:1 ratio, yielding an approximate resolution of 6 Å. Our current efforts are being focused towards improving this resolution to a level where we can (at the very least) trace helices with confidence. The addition of the N-terminal fusion protein PGS may be of use to assist with improving particle alignment, and this work is still ongoing.

## **Future Directions**

Moving forward, we are actively designing ways to improve the resolution of our cryo-EM maps. A large amount of recent energy has been spent towards developing nanobodies against the full-length Scap/Insig complex. This is in collaboration with the postdoc Dan Kober, who has adapted and developed the yeast display library published from Andrew Kruse's lab. We have pulled out initial

binders from the yeast library and I have characterized their behavior with purified complex. While this in no means a guarantee that we will find a perfect nanobody candidate that could stabilize the complex and reduce flexibility, it is a promising and high-throughput approach that could yield a successful binding partner.

In parallel, we are currently collecting data on the PGS N-terminal fusion construct, which has shown preliminary promise of improving the alignment of our particles. In addition, since PGS is covalently attached to TM1, there is a possibility that it could be reducing the flexibility of loop 1, resulting in better density for the luminal region in our 2D classes. This construct can be reconstituted into nanodiscs containing lipids with and without cholesterol present, which could potentially inform on a cholesterol binding site assuming a high enough resolution. These efforts are ongoing, and I am optimistic that we will be successful.



