Title
Evaluation of mechanisms of non-canonical G-protein signaling, membrane transporter dysregulation, and bacterial sialidase substrate preference using X-ray crystallography

Author
Rees, Steven Douglas

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Evaluation of mechanisms of non-canonical G-protein signaling, membrane transporter dysregulation, and bacterial sialidase substrate preference using X-ray crystallography

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Steven D. Rees

Committee in charge:

Professor Geoffrey A. Chang, Chair
Professor Kevin Corbett
Professor Amro Hamdoun
Professor Milton Saier
Professor Dong Wang

2018

This Dissertation of Steven D. Rees is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

DEDICATION

To my family and friends, who with their support and actions big and small made me who I am today. To my mom, for giving the (good) aforementioned contributions roots. To Sissy, I wish you could have seen this. To my brother, my model for strength of character and an endless source of information, ideals, and heckling. And to Anna, for braving this strange new world with me, and all its trials and tribulations.

EPIGRAPH

“Sometimes life seems uncomfortable and that’s actually when life starts.”

M. H. Rakib

“Someone once told me that time was a predator that stalked us all our lives. I rather believe that time is a companion who goes with us on the journey and reminds us to cherish every moment, because it will never come again. What we leave behind is not as important as how we’ve lived. After all, Number One, we’re only mortal.”

Captain Jean-Luc Picard, *Star Trek: The Next Generation*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ALS	Advanced Light Source
AMP	adenosine monophosphate
ANOSIM	analysis of similarities
APS	Advanced Photon Source
ATP	adenosine tri-phosphate
ATPase	adenosine triphosphatase
CLS	Canadian Light Source
CsA	cyclosporine A
DDD	dichlorodiphenyldichloroethane
DDE	dichlorodiphenyldichloroethylene
DDT	dichlorodiphenyltrichloroethane
DEHP	di(2-ethylhexyl)phthalate
DL	dioxin-like PCB congeners
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DOX	doxorubicin
EBI	European Bioinformatics Institute
EBN	edible bird's nest
EMBL	European Molecular Biology Laboratory
EPA	Environmental Protection Agency
GDI	guanine dissociation inhibitor
GDP	guanine diphosphate
GEF	guanine exchange factor
GEM	guanine-nucleotide exchange modulator
GOM	Gulf of Mexico
GPCR	G protein-coupled receptor
GST	glutathione-*S*-transferase
GTP	guanine triphosphate
HCH	hexachlorocyclohexane
IACUC	Institutional Animal Care and Use Committee
MALDI MS	matrix-assisted laser desorption/ionization mass spectrometry
MD	molecular dynamic
N	no
Neu5Ac	N-acetylneuraminic acid
Neu5Gc	N-glycolylneuraminic acid
NI	noninteraction
NMQ	*N*-methylquinidine
NMR	nuclear magnetic resonance
NSL	nucleotide-stabilizing linker
OCP	organochlorine pesticide

PBDE	polybrominated diphenyl ether
PCA	principal component analysis
PCB	polychlorinated biphenyl
PDB	Protein Data Bank
PFC	perfluorochemicals
PFOA	perfluorooctanoic acid
PFOS	perfluorooctane sulfonate
P-gp	P-glyoprotein
POP	persistent organic pollutant
PSM	porcine submaxillary mucin
RMSF	root mean square fluctuation
SD	standard deviation
SEC	size exclusion chromatography
sia	sialidase
SSRL	Stanford Synchrotron Radiation Lightsource
SwI	switch I
SwII	switch II
SwIII	switch III
TM	transmembrane
Y	yes

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis committee for advice, critiques, and alternate points of view throughout these past five years. In particular, thank you to my mentor and PI, Dr. Geoffrey Chang. Your tireless efforts to secure projects and funding for the lab, and willingness to pass on your expertise and life lessons, allowed me to pursue my own interests for my thesis work and view science, and my career, with a refreshing perspective.

Thanks as well to the Chang lab; it certainly takes a village. To Mark and Beverly, thank you for late-night lab support (and pineapple/bacon/ham/mystery pizzas). To Cristina, you one of the most organized and focused researchers I have ever met, and an incredibly helpful project partner. To Paul, for asking the critical questions of and bringing practical advice to many of the Chang lab projects. To Aaron, for answering research questions simple and complex, for helping to train me from recent undergraduate to crystallographer, and for adding endless heckling and excitement to the lab.

I would also like to thank the beamline scientists and support staff at the Advanced Light Source, Advanced Photon Source, the Canadian Light Source, and the Stanford Synchrotron Radiation Lightsource, for late-night/weekend troubleshooting and lending their expertise for over three dozen beamline trips over the past five years.

Thank you to the Pharmacology, Molecular Biophysics, and Interfaces Training Programs, and the Center for Microbiome Innovation, for both administrative and financial support. Thank you to the Biomedical Sciences Graduate Program, for absolutely everything you have done since I began (both known and unknown), and for offering such unique opportunities to pursue a scientific career. Last but not least, the BMS graduate students themselves; this is certainly a lively crew, with fantastic ideas on work-life balance.

Chapter 2, in full, is a reprint of the material as it appears in *Science Advances*. Nicklisch SC, Rees SD*****, McGrath AP*, Gökirmak T, Bonito LT, Vermeer LM, Cregger C, Loewen G, Sandin S, Chang G, Hamdoun A. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and co-crystal structure. April 2016. This dissertation author is a co-secondary author of this paper.

Chapter 3, in full, is a reprint of material written for *Science*. Martino C, Alisson-Silva F, Zaramela LS, Rees SD, Diaz SL, Chang G, Varki A, Zengler K. Red meat consumption promotes gut microbiota containing unconventional sialidases. This dissertation author is the co-secondary author of this paper.

Chapter 4, in full, is a reprint of material written for *Nature Structural and Molecular Biology.* Rees SD, Kalogriopoulos NA, Ngo T, Kopcho N, Ilatovskiy A, Sun N, Komives E, Chang G, Ghosh P, Kufareva I. Structural basis for GPCR-independent activation of heterotrimeric Gi proteins. This dissertation author is the co-primary author of this paper.

VITA

2012	Bachelor of Arts, Biology and Chemistry, St. Mary's College of Maryland, St. Mary's City, MD
2013-2014	Training Grant in Pharmacology
2014-2015	Training Grant in Molecular Biophysics
2018	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Rees SD*, Kalogriopoulos NA*, Ngo T, Kopcho N, Ilatovskiy A, Sun N, Komives E, Chang G, Ghosh P, Kufareva I. Structural basis for GPCR-independent activation of heterotrimeric Gi proteins. (In Preparation) **(*Co-Primary Author)**

Martino C, Alisson-Silva F, Zaramela LS, **Rees SD***, Diaz SL, Chang G, Varki A, Zengler K. Red meat consumption promotes gut microbiota containing unconventional sialiases. (In Preparation) **(*Co-Secondary Author)**

Nicklisch SC, **Rees SD***, McGrath AP*, Gökirmak T, Bonito LT, Vermeer LM, Cregger C, Loewen G, Sandin S, Chang G, Hamdoun A. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and co-crystal structure. *Sci Adv.* 2(4), e1600001. April 2016. **(*Co-Second Author)** (PMID: 27152359).

Szewczyk P, Tao H, McGrath AP, Villaluz M, **Rees SD**, Lee SC, Doshi R, Urbatsch IL, Zhang Q, Chang G. Snapshots of ligand entry, malleable binding and induced helical movement in P-glycoprotein. *Acta Crystallogr. D* 51(Pt 3), 732-41. March 2015. **(Third Author)** (PMID: 25760620).

Honors Thesis: "Ethanol Influence Modeling: Investigation and Experimentation*"* – St. Mary's College of Maryland (May 2012)

ABSTRACT OF THE DISSERTATION

Evaluation of mechanisms of non-canonical G-protein signaling, membrane transporter dysregulation, and bacterial sialidase substrate preference using X-ray crystallography

by

Steven D. Rees

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2018

Professor Geoffrey A. Chang, Chair

X-ray crystallography remains a popular technique for achieving atomic resolution of macromolecular interactions. Protein crystal structures offer a snapshot of how these complex molecules fold in three-dimensional space, often too difficult to simulate *de novo* with high accuracy, as well as where substrates, activators, inhibitors, regulators, and other regulatory partners bind them. While the physics of crystallography can be a steep learning curve, the methodology offers a powerful tool for analyzing protein crystals.

This dissertation is a summation of three major projects I undertook in Dr. Geoffrey Chang's lab at UCSD. These projects differ highly in their focus areas, moving from environmental pollution to sugar-derived microbiome changes, and still further to cancer-linked non-canonical G protein signaling. All three of these projects were linked not by topic, but by their problems: all required a means of seeing a protein-ligand interaction at high resolution, one that had not yet been previously achieved. Each required its own solutions to difficult crystallographic situations, including low resolution, poor ligand occupancy, abnormal

crystal packing, and subtle binding pocket changes, yet all benefitted from applying the crystallographic method to solve them.

This dissertation began by exploring the hypothesis that persistent organic pollutants (POPs) dysregulate efflux transporters at the barriers between main circulation and the body's major organ systems. We identified a high number of POPs in yellowfin tuna, one of the most consumed fish species in the US, many of which inhibited xenobiotic efflux by P-glycoprotein, a well-characterized efflux transporter. We utilized five asymmetrically placed Br atoms around PBDE-100, a common POP, to place this ligand in the density of a P-gp co-crystal structure, the first known structure of an efflux transporter with a POP. This structure confirmed occupancy of this ligand in an area known to be bound by inhibitors, which corroborated biochemical characterization of this interaction.

My work in crystallography continued with a collaboration with the Karsten Zengler lab at UCSD. In this work, metagenomics studies identified changes in the gut microbiome of mice in response to diets enriched with Neu5Ac or Neu5Gc, the two common mammalian sialidases with disparate presence in red meat. In particular, the first microbial sialidases with specificity for Neu5Gc were identified. I led the protein purification and structure determination efforts in this project. The structures for this project, including the first structure of any protein with a Neu5Gc-like ligand, indicated which residues likely play a role in Neu5Gc preference for these sialidases, paving the way for future work into engineering Neu5Gc-specific enzymes, which could play a role in reducing gut inflammation due to red meat-containing diets.

Another sought-after collaboration with the Pradipta Ghosh lab at UCSD explored non-canonical signaling of G protein systems by Guanine nucleotide exchange modulators (GEMs). Previous work has shown that G proteins can be activated in signaling cascades independent of canonical GPCR signaling, but little was known about how G proteins interacted with GEMs. We serendipitously discovered a GDI-like motif that aided in crystallization of a complex between the alpha G protein subunit Gi with the C-terminus of GIV/Girdin, a well-characterized GEM. This structure represents the first of a Gi protein with a natural GEM.

These projects, while rooted in highly diverse scientific fields, found commonality in their unique crystallographic challenges. A crystallographer, or protein chemist, could not have asked for a better means

of not only learning the intricacies of this complex field, but also how to apply them in new ways to solve old and new structural problems.

CHAPTER 1: Introduction

This dissertation details three unique projects on which I have worked during my graduate career. All three differ in their scientific focus, but share common ground in the structural method used to characterize interactions among the proteins and compounds involved: X-ray crystallography. In the first study (detailed in Chapter 2), anomalous scattering by particular atoms in a persistent organic pollutant (POP), PBDE-100, is used to identify where this compound interacts with, and inhibits, P-glycoprotein (P-gp), a critical xenobiotic regulator in the body. In the second (detailed in Chapter 3), the first crystals of a Neu5Gc-specific sialidase expressed by microbes in the gut are grown, and co-crystallization with sialic acid-like inhibitors is used to show how these proteins possess such preference. Finally, in the third study (detailed in Chapter 4), a serendipitous linker sequence introduced to the N-terminus of a well-characterized G protein subunit finally allows for its long-sought co-crystallization with GIV/Girdin, a natural GEM that engages G proteins in a non-canonical, recently characterized fashion. In all three cases, X-ray crystallography offered a unique, critical perspective to understanding the mechanisms explored in the study. Backgrounds on each of these studies, including their unique flavors of crystallographic problem solving, are below.

Persistent organic pollutants and P-glycoprotein

Persistent organic pollutants (POPs) are widespread environmental contaminants that bioaccumulate to toxic levels in animals and humans (1-5). Originally developed for a variety of uses, including surfactants, flame retardants, and agricultural insect repellents, POPs were largely banned in the 1970s and 1980s by the U.S. Environmental Protection Agency (EPA) and other countries with the hope of mitigating their spread (6-7). These compounds, including organochlorine pesticides, polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), and polychlorinated biphenyls (PCBs), are still present in the environment despite widespread restriction in the developed world, not only spreading from underdeveloped countries that still use them but also making their way to human populations from bioaccumulation in fatty tissues over continued consumption.

A variety of POPs spread over long distances in an aerosolized form, moving to colder areas at the earth's poles and slowly condensing in areas of lower temperature, spreading thousands of miles from their

original area of release and ubiquitously draining into bodies of water as atmospheric deposition, surface runoff, groundwater, and effluent (11-12). While some POPs have shown a decrease in U.S. populations, others like polybrominated diphenyl ethers (PBDEs) remain 10 to 100 times more concentrated in the U.S. compared to Europe, China, and Japan, most likely due to continued production for export despite restrictions on domestic use (9, 11). Resistant to typical degradation in the environment and in living organisms, POPs accumulate in fatty tissues of animals and humans due to their lipophilic nature (13, 14), and bioconcentrate with increasing trophic levels. While meat consumption is the biggest contributor to POP levels in the body, they are commonly found in marine organisms, dairy products, vegetable-based foods, and (consequently) pizza. The most common source of these compounds is believed to be dietary (9).

P-glycoprotein (P-gp), a 170-kDa transmembrane protein found in a variety of tissues, has been associated with regulation of toxic compounds in the body (21-23). A multi-drug efflux transporter of the ATP-Binding Cassette (ABC) Superfamily, P-gp plays a major role in xenobiotic regulation across barriers between the body and major organs, including the brain, liver, testes, kidneys, and gastrointestinal (GI) tract (24). Encoded by the *mdr1* gene in humans, the protein is arranged in two homologous halves connected by a phosphorylatable linker region, possessing a total of 12 transmembrane α-helices and two nucleotide-binding domains (NBDs) that cleave ATP to drive compound transport from the cell (23, 25).

P-gp has been shown to bind a variety of compounds in its ubiquitous, 6000-$Å^3$ binding pocket. While this also complicates design of drugs intended for major organs due to their extrusion by this protein, it reflects its natural role maintaining barrier integrity to toxic compounds. Studies of mouse P-gp ATPase activity (an indicator of compound transport) suggested inhibition of P-gp by PBDE-209, PCB-145, Dieldrin, and Endrin at micromolar concentrations (though no effect was seen by DEHP and BPA) (2). P-gp has also been well-documented as a binder of a variety of insecticides that inhibit P-gp's function, including methyl parathion, endosulfan, cypermethrin, diazinon, antihelminthics, ivermectin, and fenvalerate, while also being induced by DDT (4, 26). The transporter's role in environmental toxin regulation has even shown correlation with Parkinson's Disease incidence in an Italian study, suggesting a breakdown in P-gp regulation with age or toxin exposure can have major health effects (27). Another study in the 1990s found

interaction between P-gp and 38 pesticides, suggesting some effect on normal xenobiotic transport by these POPs (28, 29). Treating P-gp with verapamil or quinidine (known P-gp inhibitors) has shown to increase toxicity of insecticides like thiodicarb, and absorption of ivermectin increased (18-20).

POPs' involvement with P-gp has been suggested, but such results beg the question of how these compounds specifically interact. This would further elucidate the effects of exposure to various POPs and, consequently, the extent of barrier disruption between the body and its major organs, not only in humans but also other organisms affected by POP exposure.

For the study detailed in Chapter 2, our goal was to visualize how POPs interact with P-gp. X-ray crystallography offers the highest atomic resolution for proteins to date. However, P-gp itself is limited to typically 3.5-4.0Å, not an uncommon limitation for membrane transporters; this resolution, in a binding pocket so large, would make identifying POPs extremely difficult. However, one pollutant in particular, PBDE-100, presented us with a solution. Not only was this compound prevalent in fish samples we analyzed for the study, but it also possessed unique properties for crystallography. PBDE-100 possesses five Br atoms, placed asymmetrically around a diphenyl core. Br atoms have a degree of anomalous scattering that can be measured above a particular excitation wavelength, one we can reach at nearly all of the synchrotron sources in the world. We could use these data, combined with knowledge of the known geometry of the PBDE-100 ligand, to place PBDE-100 in the electron density of a P-gp-PBDE-100 cocrystal structure.

On the beamline, we conducted an excitation scan of each crystal to first validate if there even is Bromine present (when you're sending a couple dozen crystals and you have limited time, collecting on crystals without Br is pointless, and the hit rate was typically 50% or less...usually less). Briefly, you expose the crystal to X-rays at an energy higher than the absorption edge of that element, in this case the K edge of Br. At an energy greater than the ionization potential, the atoms ionize, or eject a lower-orbital electron (in this case, 1s). Higher-orbital electrons fall into the lower orbital to fill the "hole" left behind, and that released photon equal to the energy difference of the orbitals is detectable. This was used successfully to identify PBDE-100's five Br atoms, and place the ligand.

Neu5Gc and the microbiome

Sialic acids (Sias) are 9-carbon sugars typically found at the terminal position of N- and O-linked glycans attached to the cell surface, secreted glycoproteins, and cell surface glycosphingolipids (27-28). Typically found in vertebrates and higher-order invertebrates, Sias are introduced on proteins and lipids in the ER/Golgi biosynthetic pathway (29-30) and play major roles in cell-cell interactions. These include host communication and host-pathogen interactions, as some bacteria and protozoa have evolved the means of recognizing particular sialic acids on cell surfaces, a mechanism of molecular mimicry for invasion. Sias also play a role in mammalian embryonic development, the alternative complement pathway of the innate immune system, metastasis, and chronic and acute inflammation (31-32).

Neu5Ac and Neu5Gc are two major mammalian sialic acids. Following Neu5Ac synthesis, CMP-Neu5Ac-hydroxylase introduces a hydroxyl group onto the acetamido group on C5. Humans possess a 92-bp deletion in the CMAH gene located on chromosome 6, causing a frameshift mutation that renders CMP-Neu5Ac-hydroxylase inactive (33). This likely occurred early in the hominid lineage, after divergence with non-human primates several million years ago (34). This caused major changes in expression and function of a number of significant genes related to sialic acid biology. In particular, this included a loss of Neu5Gc expression on the cell surface, and an increased presence of Neu5Ac.

In the past 15 years, evidence has emerged of Neu5Gc display in human cells, despite no other mechanism for Neu5Gc production in humans (35-36, 38). Antibodies specific for Neu5Gc-terminating sugar moieties showed reactivity to human tissue in the GI tract, kidneys, immune cells, and embryonic stem cells (37-38). Cell culture and human volunteer studies suggested a potential dietary mechanism for Gc uptake and display, and cell culture studies revealed that a lysosomal sialidase would release free Neu5Gc from pinocytosed glycosidyl-Neu5Gc, and a lysosomal sialic acid transporter would release free Neu5Gc into the cytosol for glycan incorporation (37, 40). The gut microbiome also plays a major role in Neu5Gc release and subsequent uptake in humans. A variety of commensal and pathogenic bacteria secrete or display sialidases, enzymes that release free Sias from their glycosidically-bound form upon dietary uptake in the gut, which are then introduced into gut epithelial cells by a variety of PM- and lysosomal SA transporters (SiaPQM, NanT and SatABCD).

For the study detailed in Chapter 3, we worked with Dr. Karsten Zengler at UCSD to identify sialidases in the gut microbiome that could play a role in liberating Neu5Gc from cell surfaces, which would play a critical role in reducing antibody-driven inflammation as described above. Dr. Zengler's lab had tested Neu5Gc- and Neu5Ac-containing diets (or neither, as controls) in mice deficient in CMAH, rendering them with similar sensitivities as humans. Metagenomic analyses identified a set of sialidases enriched in the microbiome of Neu5Gc-fed mice compared to Neu5Ac-fed. All five exhibited Neu5Gc preference over Neu5Ac in biochemical assays, yet none showed much sequence difference from previously studied sialidases; in fact, the residues predicted from known sialidase structures to interact with sias were found to be 100% conserved in these sialidases. To better understand this interaction, we co-crystallized one of the sialidases with DANA, a potent inhibitor of sialidases with high similarity to Neu5Ac, save an eliminated C2 that retains the inhibitor in the binding pocket. Additionally, we co-crystallized the same sialidase with DANA-Gc, a Neu5Gc-like version of the inhibitor.

While DANA co-crystallization showed strong occupancy of the small molecule in the sialidase binding pocket, DANA-Gc's occupancy was low. This could have been due to lower purity of the compound, a crystallographic defect (unfortunately more common than one would hope), or simply lower affinity for the sialidase. However, the unit cell of this co-crystal was quite large, possessing four sialidase molecules in the asymmetric unit. By overlaying the asymmetric unit members, combined with multi-crystal averaging, we were able to see enough of DANA-Gc to model it successfully. This indicated an unexpected residue that was close enough to Neu5Gc's hydroxyl group (not found on Neu5Ac) that could increase substrate contact in the binding pocket, offering an explanation for this preference.

G Proteins and GEMs

Heterotrimeric G-protein signaling (subunits α, β, and γ) plays a role in many of the body's major signaling pathways. Canonical G-protein signaling is mediated by G-protein coupled receptors. The binding of an agonist-bound GPCR to resting state G protein complex (GDP-bound) induces an exchange for GTP at Gα's nucleotide binding site, which promotes dissociation of the G protein heterotrimer to a variety of cell signaling pathways, after which GTP is hydrolyzed by Ga to GDP to return the mechanism to resting state

(41-42). This all happens on the order of milliseconds. This canonical pathway has been well-studied, and led to the nearly 40% of drugs on the market that target GPCRs (43).

A major factor in this exchange is Gα, which interacts directly with GPCRs as well as a variety of nucleotide exchange regulators, including GDIs (44), GEFs (45), and RGS proteins (46). Gα possesses a Ras-like domain which aids in GTP hydrolysis, an all-helical domain which binds several of the regulators and downstream effectors like the Gβ subunit, and a GDP cleft in its interdomain sequences (42).

In the past 10 years, a non-canonical mechanism of G protein signaling has emerged. In 2005, a peptide called KB-752, from a peptide bacteriophage library from Karo*Bio USA, was identified to bind Gα (47). The sequence of this peptide was found later to be homologous to another protein not previously mapped to G protein signaling, called GIV/Girdin (48). This protein was found, like KB-752, to increase nucleotide exchange of Gα, like GPCRs, but in an independent mechanism that could also engage other signaling pathways, including TLRs and RTKs, which had not been linked to G protein signaling. This signaling pathway branches across multiple signaling pathways on the order of seconds, and a variety of cancers have been linked to faults in this system. This work has led to the discovery of other proteins with similar motifs, including Calnuc (49) and Daple (50). These proteins bind GDP-bound Gαi to enhance nucleotide exchange, and lower nucleotide exchange upon binding Gαs (51). This dual mechanism of action, not found in previous G protein signaling regulators, classified these proteins into the new class of GEMs.

While GIV and KB-752 share a conserved motif that interacts with Gα, structures of Gα with GIV or other natural GEMs had to date not been achieved. Even truncating GIV to this conserved motif proved unsuccessful, requiring modeling simulations or other techniques to attempt to understand this mechanism. The goal of the study detailed in Chapter 4 was to achieve a structure of Gα with a GEM. We succeeded in co-crystallizing Gαi with the conserved GEM motif of GIV, but through a serendipitous event that can occur in crystal packing; the N-terminus of Gα, which had been modified with a His tag and tag-orf linker sequence to ease purification, bound its symmetry neighbor in the crystal, acting as a GDI and (theoretically) slowing nucleotide exchange. While this behavior was not observed to impact biochemical studies, crystallization itself introduced this phenomenon, allowing us to achieve the co-crystal structure. We compared this to a

similarly crystallized KB-752 co-crystal, to account for any confounding variables introduced by this N-terminal modification.

REFERENCES

1. Miller, D. S. and X. Wang. 2010. The aryl hydrocarbon receptor (AhR) upregulates P-glycoprotein (P-gp) at the blood-brain barrier (BBB). *FASEB J.* 24: 604.7.
2. Nicklisch, S. C. T. and A. Hamdoun. 2013. Identification and characterization of the molecular interactions of persistent organic pollutants with the multidrug resistance transport protein ABCB1. *Submission, 10th Annual ABC Genetic Workshop*.
3. Wang, X., Hawkins, B. T., and D. S. Miller. 2011. Aryl hydrocarbon receptor-mediated up-regulation of ATP-driven xenobiotic efflux transporters at the blood-brain barrier. *FASEB J.* 25(2): 644-52.
4. Sreeramulu, K., Liu, R., and F. J. Sharom. 2007. Interaction of insecticides with mammalian P-glycoprotein and their effect on its transport function. *BBA-Biomembranes* 1768(7): 1750-57.
5. Abu-Qare, A., Elmasry, E., and M. Abou-Donia. 2003. A role for P-glycoprotein in environmental toxicology. *J. Tox and Environ. Health* 6(3): 279-88.
6. U.S. EPA. 2010. Blood Persistent Organic Pollutants Level. *Report on the Environment*.
7. Chandra, R. and S. Chaudhary. 2013. Persistent organic pollutants in environment and their health hazards. *Int J Bioassays* 02(09): 1232-8.
8. He, X., Nie, X., Wang, Z., Cheng, Z., Li, K., Li, G., Hung Wong, M.,Liang, X., and M. T. Tsui, 2011. Assessment of typical pollutants in waterborne by combining active biomonitoring and integrated biomarkers response. *Chemosphere* 84(10): 1422-31.
9. Schecter, A., Haffner, D., Colacino, J., Patel, K., Päpke, O., Opel, M., and L. Birnbaum, 2010. Polybrominated diphenyl ethers (PBDEs) and Hexabromocyclodecane (HBCD) in composite U.S. food samples. *Environ Health Perspect.* 118(3): 357-62.
10. Someya, M., Ohtake, M., Kunisue, T., Subramanian, A., Takahashi, S., Chakraborty, P., Ramachandran, R., and S. Tanabe, 2010. Persistent organic pollutants in breast milk of mothers residing around an open dumping site in Kolkata, India: Specific dioxin-like PCB levels and fish as a potential source. *Environ Int.* 36(1): 27-35.
11. F. Wania and D. Mackay. 1996. Tracking the distribution of persistent organic pollutants. *ACS Environ Sci and Tech* 30(9): 390A-396A.
12. Zhou, H., Wu, H., Liao, C., Diao, X., Zhen, J., Chen, L., and Q. Xue, 2010. Toxicology mechanism of the persistent organic pollutants (POPs) in fish through AhR pathway. *Tox. Mech. And Methods* 1-8.
13. Berntssen, M. H. G., Olsvik, P. A., Torstensen, B. E., Julshamn, K., Midtun, T., Goksoyr, A., Johansen J., Sigholt, T., Joerum, N., Jakobsen, J. V., Lundebye, A. K., and E. J. Lock, 2010. Reducing persistent organic pollutants while maintaining long chain omega-3 fatty acid in farmed Atlantic salmon using decontaminated fish oils for an entire production cycle. *Chemosphere* 81(2): 242-52.
14. Rudge, C. V. C., Sandanger, T., Rollin, H. B., Calderon, I. M. P., Volpato, G., Silva, J. L. P., Duarte, G., Neto, C. M., Sass, N., Nakamura, M. U., Odland, J. O., and M. V. C. Rudge. 2012. Levels of selected persistent organic pollutants in blood from delivering women in seven selected areas of Sao Paulo State, Brazil. *Environ Int.* 40: 162-9.
15. Breslin, W., Kirk, H., and Zimmer, M., 1989. Teratogenic evaluation of a polybromodiphenyloxide mixture in New Zealand white rabbits following oral exposure. *Fund. Appl. Toxicol.* 12: 151–157.
16. Eriksson, P., Viberg, H., Jakobsson, E., Orn, U., and A. Fredriksson. 2002. A brominated flame retardant, 2, 20, 4, 40, 5-pentabromodiphenyl ether: uptake, retention, and induction of neurobehavioral alterations in mice during a critical phase of neonatal brain development. *Toxicol. Sci.* 67: 98–103.

17. Turyk, M. E., Persky, V. W., Imm, P., Knobeloch, L., Chatterton, R., and H. A. Anderson. 2008. Hormone disruption by PBDEs in adult male sport fish consumers. *Environ Health Perspect.* 116(12): 1635-41.

18. Joenson, U. N., Bossi, R., Leffers, H., Jensen, A. A., Skakkebaek, N. E., and N. Jorgensen, 2009. Do perfluoroalkyl compounds impair human semen quality? *Environ Health Perspect.* 117: 923-7.

19. Kimura-Kuroda J., Nagata, I., and Y. Kuroda. 2007. Disrupting effects of hydroxypolychlorinated biphenyl (PCB) congeners on neuronal development of cerebellar Purkinje cells: a possible causal factor for developmental brain disorders? *Chemosphere* 67: S412–S420.

20. Ribas-Fito, N., Sala, M., Kogevinas, M., and J. Sunyer. 2001. Polychlorinated biphenyls (PCBs) and neurological development in children: a systematic review. *J Epidemiol Comm Health* 55: 537-46.

21. Juliano, R. L. and Ling, V. 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *BBA* 455: 152-62.

22. Roninson, I. B., Chin, J. E., Choi, K. G., Gros, P., Housman, D. E., Fojo, A., Shen, D. W., Gottesman, M. M., and I. Pastan, 1986. Isolation of human *mdr* DNA sequences amplified in multidrug-resistant KB carcinoma cells. *PNAS* 83: 4538-42.

23. Aller, S. G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhou, R., Harrell, P. M., Trinh, Y. T., Zhang, Q., Urbatsch, I. L., and G. Chang. 2009. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323(5922): 1718-22.

24. Leslie, E. M., Deeley, R. G., and S. P. Cole. 2004. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxico and Applied Pharm* 204: 216-37.

25. Ward, A. B., Szewczyk, P., Grimard, V., Lee, C. W., Martinez, L., Doshi, R., Caya, A., Villaluz, M., Pardon, E., Cregger, C., Swartz, D. J., Falson, P. G., Urbatsch, I. L., Govaerts, C., Steyaert, J., and G. Chang. 2013. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *PNAS* 110(33): 13386-91.

26. Shabbir, A., DiStasio, S., Zhao, J., Cardozo, C. P., Wolff, M. S., and A. J. Caplan. 2005. Differential effects of the organochlorine pesticide DDT and its metabolite p,p′-DDE on p-glycoprotein activity and expression. *Toxicol. Appl. Pharmacol.* 203: 91–98.

27. Varki A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature. 2007; 446:1023–1029.

28. Varki A, Schauer R. 2008. *Essentials of Glycobiology*. Cold Spring Harbor Press. pp. Ch. 14.

29. Fulcher CA, "MetaCyc Chimeric Pathway: superpathway of sialic acid and CMP-sialic acid biosynthesis", 2009.

30. Warren L, Felsenfeld H. 1962. The Biosynthesis of Sialic Acids. *The Journal of Biological Chemistry.* 237 (5): 1421.

31. Severi E, Hood DW, Thomas GH. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology* 153(9): 2817–2822.

32. Varki A, Gagneux P. 2012. Multifarious roles of sialic acids in immunity. *Ann. N.Y. Acad. Sci*. 1253(1): 16–36.

33. Ghaderi D, Taylor RE, Padler-Karavani V, Diaz S, Varki A. 2010. Implications of the presence of N-glycolyneuraminic acid in recombinant therapeutic glycoproteins. *Nature Biotechnology*. 28(8): 863–7.

34. Chou HH, Takematsu H, Diaz, S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc. Natl. Acad. Sci.* 95, 11751–11756.

35. Bardor, M., Nguyen, D. H., Diaz, S., and Varki, A. (2005) Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. J. Biol. Chem. 280, 4228–4237

36. Malykh YN, Schauer R, Shaw L. 2001. N-Glycolylneuraminic acid in human tumors. *Biochimie* 83, 623–634.

37. Hedlund M, Padler-Karavani V, Varki NM, Varki A. 2008. Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. *Proc. Natl. Acad. Sci.* 105, 18936–18941.

38. Diaz SL, Padler-Karavani V, Ghaderi D, Hurtado-Ziola N, Yu H, Chen X, Brinkman-Van der Linden, EC, Varki A, Varki NM. 2009. Sensitive and specific detection of the non-human sialic acid Nglycolylneuraminic acid in human tissues and biotherapeutic products. *PLoS ONE* 4, e4241.

39. Tangvoranuntakul P, Gagneux P, Diaz S, Bardor M, Varki N, Varki A, Muchmore E. 2003. Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc. Natl. Acad. Sci.* 100, 12045–12050.

40. Allison-Silva F, Kawanishi K, Varki A. 2016. Human risk of diseases associated with red meat intake: analysis of current theories and proposed rule for metabolic incorporation of a non-human sialic acid. *Mol Aspects Med.* 51: 16-30.

41. Siderovski DP, Willard FS. 2005. The GAPs, GEFs, and GDIs of heterotrimeric G-protein alpha subunits. *Int J Biol Sci* 1:51-66.

42. Ghosh P, Rangamani P, Kufareva I. 2017. The GAPs, GEFs, GDIs, and…now, GEMs: New kids on the heterotrimeric G protein signaling block. *Cell Cycle* 16(7): 607-612.

43. Hopkins AL, Groom CR. 2002. The druggable genome. *Nat Rev Drug Discov* 1:727-730.

44. Brandt DR. and Ross EM. 1985. GTPase activity of the stimulatory GTP-binding regulatory protein of adenylate cyclase, Gs. Accumulation and turnover of enzyme-nucleotide intermediates. *J Biol Chem.* 260(1):266–72.

45. Gilman AG. 1987. G proteins: transducers of receptor-generated signals. *Annu Rev Biochem.* 56: 615–49.

46. Siderovski DP., Hessel A., Chung S., Mak TW., and Tyers M. 1996. A new family of regulators of G-protein-coupled receptors? *Curr Biol.* 6(2):211–2.

47. Johnston CA, Willard FS, Jezyk MR, Fredericks Z, Bodor ET, Jones MB, Blaesius R, Watts VJ, Harden TK, Sondek J, Ramer JK, Siderovski, DP. 2005. Structure of Gai1 bound to a GDP-selective peptide provides insight into guanine nucleotide exchange. *Structure* 13(7): 1069-1080.

48. Garcia-Marcos M, Ghosh P, Farquhar MG. 2009. GIV is a nonreceptor GEF for G alpha I with a unique motif that regulates Akt signaling. *Proc Natl Acad Sci* 106(9): 3178-83.

49. Garcia-Marcos M, Kietrsunthorn PS, Wang H, Ghosh P, Farquhar MG. 2011. G Protein binding sites on Calnuc (nucleobindin 1) and NUCB2 (nucleobindin 2) define a new class of G(alpha)i-regulatory motifs. *J Biol Chem.* 286(32):28138-49.

50. Aznar N, Midde KK, Dunkel Y, Lopez-Sanchez I, Pavlova Y, Marivin A, Barbazán J, Murray F, Nitsche U, Janssen KP, Willert K, Goel A, Abal M, Garcia-Marcos M, Ghosh P. 2015. Daple is a novel non-receptor GEF required for trimeric G protein activation in Wnt signaling. *Elife* 4:e07091.

51. Gupta V, Bhandari D, Leyme A, Aznar N, Midde KK, Lo IC, Ear J, Niesman I, López-Sánchez I, Blanco-Canosa JB, von Zastrow M, Garcia-Marcos M, Farquhar MG, Ghosh P. 2016. GIV/Girdin activates Gαi and inhibits Gαs via the same motif. *Proc Natl Acad Sci* 113(39):E5721-30.

CHAPTER 2

Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure

Sascha C. T. Nicklisch[a], Steven D. Rees[b,c], Aaron P. McGrath[b,c], Tufan Gökirmak[a], Lindsay T. Bonito[a], Lydia M. Vermeer[d], Cristina Cregger[c], Greg Loewen[d], Stuart Sandin[a], Geoffrey Chang[c,e], Amro Hamdoun[a]

[a] Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0202, USA.
[b]These authors contributed equally to this work.
[c]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093-0657, USA.
[d]Sekisui XenoTech, LLC, 1101 West Cambridge Circle Drive, Kansas City, KS 66103, USA.
[e]Department of Pharmacology, School of Medicine, University of California San Diego, La Jolla, CA 92093-0657, USA.

Published in *Science Advances*

April 2016

PREFACE TO CHAPTER 2

Chapter 2, in full, is an article published in *Science Advances* entitled *"*Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and cocrystal structure". This chapter was the first project I worked on upon entering the lab and represents my first work in the crystallographic method. I am thankful to Dr. Sascha Nicklisch for leading the project on the side of protein and pollutant characterization, and to Dr. Aaron McGrath for his guidance on handling anomalous data, critical to this paper's structural insights. This work is, to our knowledge, the first to show how the cellular gatekeeper P-glycoprotein interacts with persistent organic pollutants in the environment, and that pollutants at their current levels in the environment can affect our ability to regulate xenobiotics.

ABSTRACT

The world's oceans are a global reservoir of persistent organic pollutants to which humans and other animals are exposed. Although it is well known that these pollutants are potentially hazardous to human and environmental health, their impacts remain incompletely understood. We examined how persistent organic pollutants interact with the drug efflux transporter P-glycoprotein (P-gp), an evolutionarily conserved defense protein that is essential for protection against environmental toxicants. We identified specific congeners of organochlorine pesticides, polychlorinated biphenyls, and polybrominated diphenyl ethers that inhibit mouse and human P-gp, and determined their environmental levels in yellowfin tuna from the Gulf of Mexico. In addition, we solved the cocrystal structure of P-gp bound to one of these inhibitory pollutants, PBDE (polybrominated diphenyl ether)–100, providing the first view of pollutant binding to a drug transporter. The results demonstrate the potential for specific binding and inhibition of mammalian P-gp by ubiquitous congeners of persistent organic pollutants present in fish and other foods, and argue for further consideration of transporter inhibition in the assessment of the risk of exposure to these chemicals.

IMPACT STATEMENT

Here we provide the first data to demonstrate how the cellular gatekeeper P-glycoprotein interacts with persistent organic pollutants in the environment, and that pollutants at their current levels in the environment can affect our ability to regulate xenobiotics.

INTRODUCTION

Persistent organic pollutants (POPs) are hazardous, man-made chemicals that endure in the environment and bioaccumulate in animals. Their environmental persistence ensues from properties such as halogenation and hydrophobicity that slow degradation and promote partitioning into organisms. At the same time, these properties also favor POP bioaccumulation by slowing their elimination. Indeed, although all animals have numerous metabolic enzymes, conjugation systems, and transporter proteins that normally act to eliminate xenobiotics, these systems appear ineffective at limiting POP bioaccumulation.

A critical step toward understanding the persistence and organismal impacts of POPs is defining their interactions with xenobiotic elimination systems. Drug transporters are plasma membrane proteins that both limit the entry of foreign chemicals into the body and speed their clearance, and are already well studied for their roles in drug disposition (1). Previous studies have suggested that environmental chemicals can also interact with drug transporters, such as P-glycoprotein (P-gp), but that they are poorly transported, and that these interactions ultimately lead to inhibition of transporter function (2–10). Of concern is that this inhibition reduces the efficacy of transport, thereby sensitizing animals to toxic chemicals that would otherwise be effluxed (11, 12).

Here, we took a multilevel approach to examine transporter-pollutant interactions, from levels in the environment down to the cocrystal structure of an environmental chemical bound to the transporter. We focused on P-gp, an adenosine triphosphate (ATP)–binding cassette (ABC) transporter (13–15), which plays a major role in the disposition of xenobiotics (1, 16) and which is one of the best-studied drug transporters to date. P-gp has a large binding pocket that interacts with a wide variety of structurally divergent hydrophobic molecules (16–19), and binding within this large pocket can have different impacts on the transporter, from stimulation to inhibition of function (20, 21). P-gp is conserved and is typically expressed at environmental barrier tissues, such as the small intestine or gills (22–24).

To identify POPs that interact with P-gp, we used robust biochemical and cellular assays of mouse and human P-gp and identified specific congeners that inhibit this transporter. Using x-ray crystallography, we validated the binding of one of these chemicals, polybrominated diphenyl ether (PBDE)–100, deep within the ligand pocket of the transporter, providing the first snapshots of P-gp bound to a pollutant. To

gain insight into the environmental relevance of P-gp inhibitors, we measured their levels in yellowfin tuna (*Thunnus albacares*) from the Gulf of Mexico (GOM) and used these data to examine the effects of a representative POP mixture on the transport function of the human P-gp.

RESULTS

Persistent pollutants inhibit P-gp

A gap in knowledge exists with regard to the identity and levels of environmental modulators of P-gp. To begin to address this question, we applied a systematic screen to test ubiquitous chemicals for effects on P-gp and to identify specific congeners most relevant to human and environmental health (Figure 2.1). In this screen, we used two independent assays to test 37 pollutants for inhibitory effects, and published data on the incidence of the specific congeners in humans to identify relevant compounds (Figure 2.1 and Table 2.1). We took advantage of mouse P-gp (Figure 2.S1), an established model of mammalian P-gp that is amenable to high-resolution x-ray crystallography (25, 26) and has a high amino acid sequence homology with human P-gp (18).

For chemosensitization assays, we expressed mouse P-gp in drug-sensitive yeast (*Saccharomyces cerevisiae*). In these strains, three endogenous transporters are deleted, rendering the cells sensitive to the cytotoxic effects of doxorubicin (DOX) (27). Expression of P-gp protects the cells against DOX (Figure 2.S2), and inhibition of recombinant P-gp by exogenous inhibitors restores DOX sensitivity, causing a dose-dependent decrease in their growth (Figure 2.2A). For ATPase assays, we used purified, detergent-solubilized protein and measured the liberation of inorganic phosphate in the presence or absence of chemicals (28, 29). To determine the potency of chemicals as inhibitors, P-gp was prestimulated with verapamil (30, 31) and then incubated with each pollutant individually.

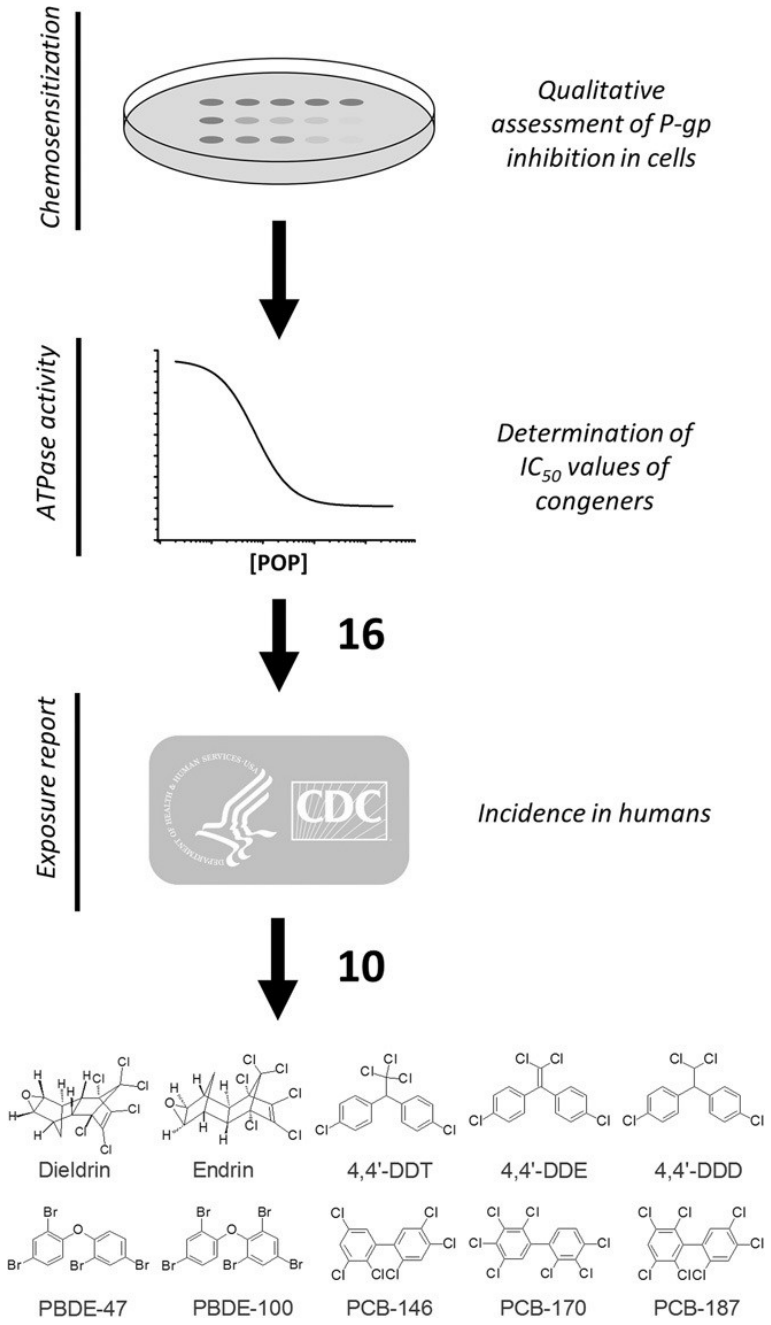


Figure 2.1. Identification of P-gp-inhibiting POPs. Thirty-seven pollutants were tested for interactions with mouse P-gp using two independent assays (see Table 2.1). Sixteen compounds were identified as inhibitors in both assays. We focused on 10 congeners reported in humans on the basis of the literature (86, 87) and the *Fourth National Report on Human Exposure to Environmental Chemicals* of the U.S. Centers for Disease Control and Prevention (88). DDD, dichlorodiphenyldichloroethane; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; PCB, polychlorinated biphenyl; PBDE, polybrominated diphenyl ether.

Table 2.1. Summary of the interaction kinetics of environmental chemicals with mouse P-gp. Compounds that were inhibitors in both assays and present in humans are in boldface. Compounds that were inhibitors in both assays but not known to be detected in humans are italicized. Y, yes; N, no; HCH, hexachlorocyclohexane; DDD, dichlorodiphenyldichloroethane; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; PFOA, perfluorooctanoic acid; PFOS, perfluorooctane sulfonate; DEHP, di(2-ethylhexyl)phthalate; NI, noninteraction; DL, dioxin-like PCB congeners.

Chemical	IC_{50} (μM)*	I_{Max} (%)†	I_{Yeast}‡
Organochlorine pesticides			
Aldrin	26.2 (±1.9)	0.93	Y
Dieldrin	21.8 (±4.2)	0.69	Y
Endrin	1.1 (±0.7)	0.51	Y
α-HCH	26.8 (±18.0)	0.09	N
β-HCH	NI	0.00	N
γ-HCH (lindane)	82.6 (±9.9)	0.58	N
Heptachlor	10.4 (±0.9)	0.98	N
Hexachlorobenzene	NI	0.00	N
Methoxychlor	21.7 (±2.4)	0.57	N
Mirex	3 (±0.2)	0.24	N
4,4′-DDD	72.5 (±5.7)	0.45	Y
4,4′-DDE	31.3 (±3.7)	0.63	Y
4,4′-DDT	25.6 (±4.8)	0.61	Y
Polybrominated diphenyl ethers			
PBDE-3	NI	0.00	N
PBDE-47	22.6 (±6.2)	0.52	Y
PBDE-49	35.6 (±5.4)	0.52	Y
PBDE-100	23.2 (±2.9)	0.68	Y
PBDE-209	6.5 (±0.4)	0.94	N
Perfluorochemicals			
PFOA	156.5 (±6.1)	1.20	N
PFOS	NI	0.00	N
Plastic-related compounds			
Bisphenol A	NI	0.00	N
DEHP	NI	0.00	N
Polychlorinated biphenyls			
PCB-118[DL]	15.9 (±1.0)	0.89	N
PCB-134§	12.5 (±0.8)	0.82	Y
PCB-142	6.1 (±0.7)	0.90	Y
PCB-145	4.4 (±0.4)	0.92	Y
PCB-146	12.8 (±1.9)	0.50	Y
PCB-147§	23.6 (±3.1)	0.82	Y
PCB-152	22 (±4.2)	0.84	N

Table 2.1 (continued). Summary of the interaction kinetics of environmental chemicals with mouse P-gp.

Compounds that were inhibitors in both assays and present in humans are in boldface. Compounds that were inhibitors in both assays but not known to be detected in humans are italicized. Y, yes; N, no; HCH, hexachlorocyclohexane; DDD, dichlorodiphenyldichloroethane; DDE, dichlorodiphenyldichloroethylene; DDT, dichlorodiphenyltrichloroethane; PFOA, perfluorooctanoic acid; PFOS, perfluorooctane sulfonate; DEHP, di(2-ethylhexyl)phthalate; NI, noninteraction; DL, dioxin-like PCB congeners.

Chemical	**IC_{50} (μM)***	**I_{Max} (%)†**	**I_{Yeast}‡**
PCB-153§	21.8 (±3.1)	0.46	N
PCB-154§	14.3 (±1.1)	0.54	N
PCB-161	43.2 (±8.3)	0.40	N
PCB-168§	25.8 (±3.7)	0.66	N
PCB-169DL	9.7 (±0.5)	0.93	N
PCB-170	9.2 (±0.8)	0.73	Y
PCB-186	6.9 (±0.5)	0.91	N
PCB-187	11.6 (±0.6)	0.56	Y

*Inhibition coefficients (IC_{50}) (±SD) of verapamil-stimulated ATPase activity.

†Percent maximum inhibition (I_{Max}).

‡Inhibition in the yeast cytotoxicity assay (I_{Yeast}).

§Co-elusion with other PCB congeners.

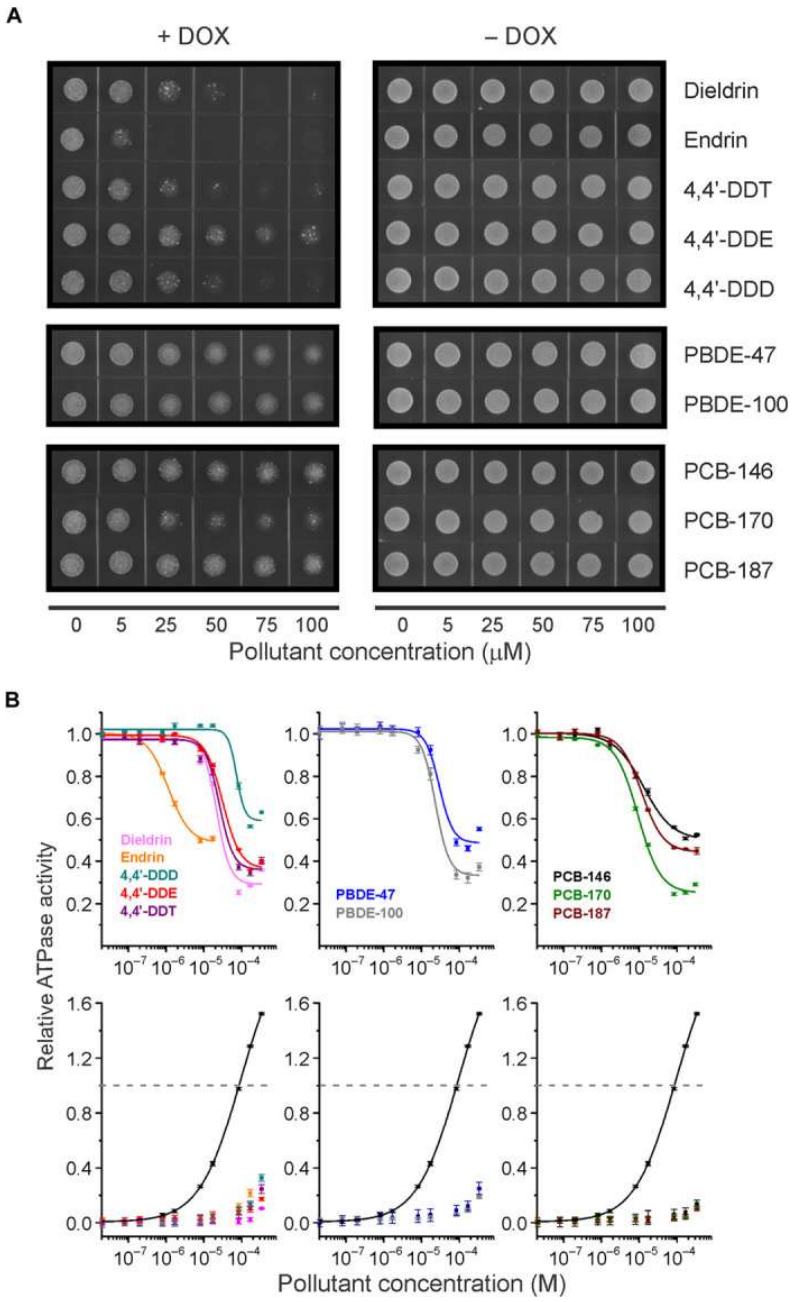


Figure 2.2. P-gp-inhibiting effects of POPs.
(**A**) Representative images showing chemosensitization of P-gp–expressing yeast by POPs. Inhibition of mouse P-gp heterologously expressed in drug (+DOX)–sensitive yeast. Inhibition is indicated by the reduction in yeast growth with increasing POP concentration. None of the POPs were toxic to yeast in the absence of DOX (−DOX). Yeast assays were replicated three times and representative micrographs are shown. (**B**) Upper panel: Inhibition of verapamil-stimulated P-gp adenosine triphosphatase (ATPase) activity by POPs. Graphs show P-gp ATPase inhibition kinetics with the 10 transporter-inhibiting POPs. ATPase assays were performed with purified, recombinant mouse P-gp protein. Lower panel: Lack of P-gp ATPase activation by POPs. ATPase activation of P-gp–inhibiting organochlorine pesticides (OCPs), PBDEs, and PCBs was determined with increasing concentrations of each compound and without verapamil prestimulation. The black curves show verapamil stimulation. Points were normalized to 100 μM verapamil stimulation (gray dashed line) and represent the average ATPase activity ± SD from three to six experiments. Where not visible, error bars are smaller than symbols; R^2 values were all >0.99.

From our assays, 16 chemicals that exhibited inhibitory action in both assays were identified. Of these, 10 have been reported in humans (Figure 2.1 and Table 2.1), including DDT, DDE, DDD, dieldrin, endrin, PCB-146, PCB-170, PCB-187, PBDE-47, and PBDE-100 (Figure 2.1). All 10 compounds chemosensitized P-gp–overexpressing yeast toward DOX and had IC_{50} (median inhibitory concentration) values of 1.1 to 72.5 μM in the ATPase assays (Figure 2.2 and Table 2.1). We also observed that congener-specific effects were similar in the two assays, suggesting common modes of interaction with the transporter. For instance, in the cellular assay, endrin had the most pronounced chemosensitizing effect, whereas its stereoisomer, dieldrin, had a markedly lower inhibitory effect (Figure 2.2A). Similarly, in ATPase assays, endrin was nearly 20 times more potent than dieldrin (Figure 2.2B and Table 2.1).

We also examined whether the 10 compounds might stimulate the basal (that is, unstimulated) activity of P-gp (32), which is sometimes seen when P-gp is incubated with its transported substrates, and we found that none of them significantly stimulated activity (Figure 2.2B). This, along with their ready bioaccumulation, suggests that POPs are not well transported; however, such a possibility is not completely ruled out because the ATPase assay does not directly measure transport. In addition, none of these 10 compounds were toxic to yeast in the absence of DOX (Figure 2.2A), indicating that the observed chemosensitizing effects were not related to nonspecific effects, such as membrane perturbation.

PBDE-100 binds to a conserved region deep within the drug-binding pocket of P-gp

To further investigate these interactions, we next sought to visualize the binding of a pollutant with P-gp. Until recently, little has been known about the mechanisms by which environmental chemicals might interact with P-gp, and an unresolved question has been whether their interactions involve specific binding to the ligand-binding domain or nonspecific interactions with the transporter. Although solving structures of large transmembrane proteins bound to inhibitors is a challenge because of the resolution range of ~3 to 4 Å, this limitation can be circumvented by the use of compounds that have an anomalous x-ray scatter signal from atoms such as selenium, mercury, or bromine. For this reason, we chose to cocrystallize and solve the structure of mouse P-gp bound to the penta-brominated compound PBDE-100. The resulting crystal diffracted to 3.5 Å resolution (Table 2.S1). The PBDE-100 binding site was located on one-half of the transporter's pseudo-symmetric dimer structure (Figure 2.3A and B). Fifteen residues from transmembrane

segments 5, 6, 7, 8, and 12 mediated hydrophobic interactions with the diphenyl core of PBDE-100 (Figure 2.3C and D), burying more than 90% of its solvent-accessible surface. Eleven of these residues (Y^{303}, Y^{306}, A^{307}, F^{310}, F^{331}, F^{724}, I^{727}, V^{731}, S^{752}, F^{755}, and S^{975}) were unique to this structure and not shared with those previously identified in structures of P-gp crystallized with a series of selenium-labeled cyclopeptide ligands (26). Alternative side-chain conformations were observed for all interacting residues upon PBDE-100 binding (Figure 2.S3).

To compare PBDE-100 binding sites of human and mouse P-gp, we aligned the interacting regions identified in our cocrystal structure (Figure 2.3E). In addition, we examined potential conservation or divergence of this region in human, mouse, zebrafish, and sea urchin P-gp (Figure 2.S4). These comparisons revealed a high degree of similarity in PBDE-100 binding residues, with 11 of the 15 residues being identical in vertebrates and with 13 being identical in humans and mice (Figure 2.3E). Nine of these residues were conserved in sea urchins (Figure 2.S4), which diverge from humans at the base of the deuterostome lineage (33). Collectively, these results suggest evolutionary conservation of the newly identified PBDE-100 binding sites.

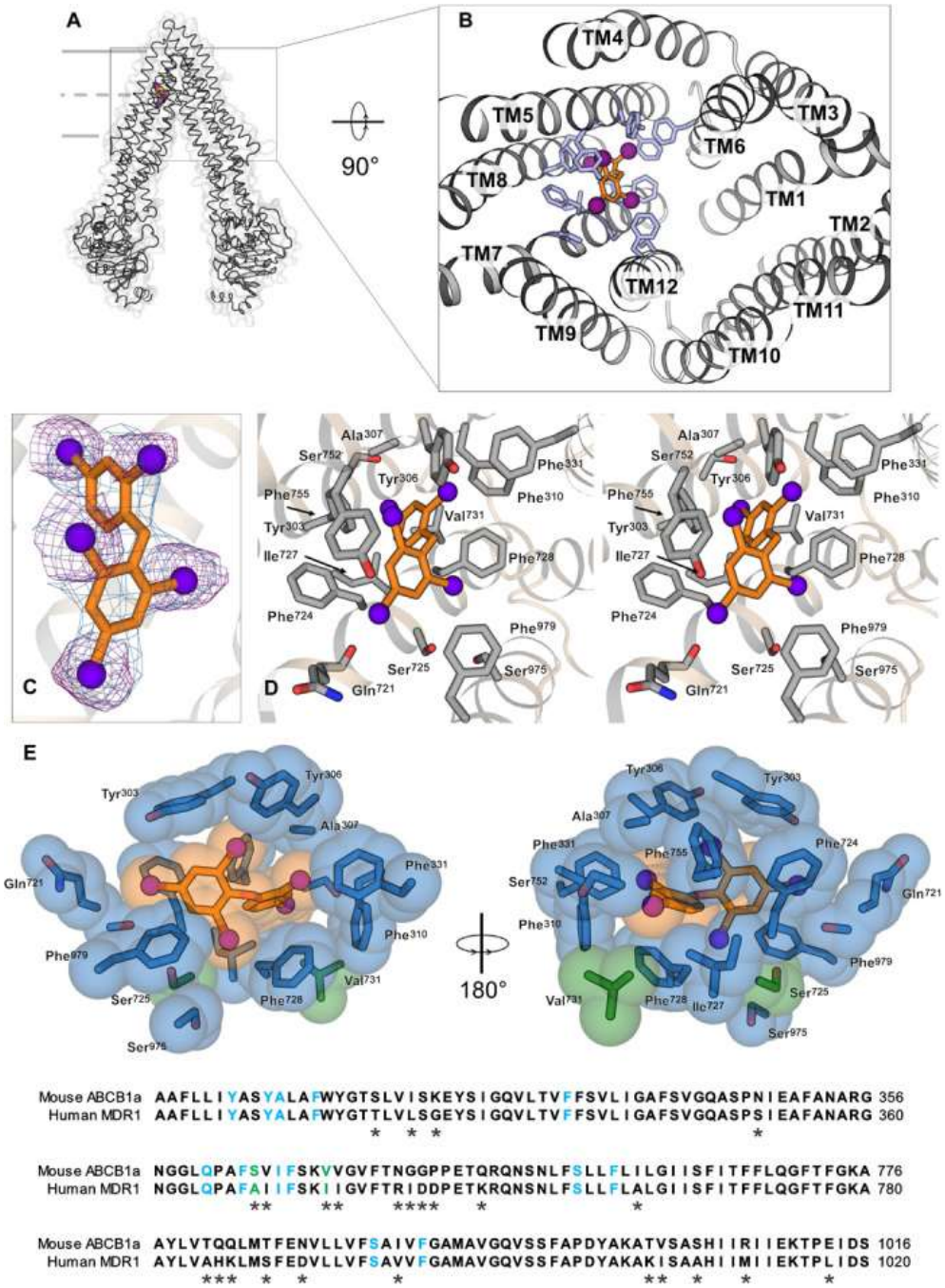


Figure 2.3. POP interactions at the substrate-binding site of mouse P-gp.
(**A**) Structure of mouse P-gp cocrystallized with PBDE-100. (**B**) Location of PBDE-100 at a distinct binding site in the internal cavity of P-gp, viewed from the intracellular side. TM, transmembrane. (**C**) $2mF_o - DF_c$ electron density (where m is the figure of merit and D is the Sigma-A weighting factor) for PBDE-100 (blue; contour level of 1.2σ) and anomalous difference density peaks (purple; contour level of 3.5σ). (**D**) Stereo view of the binding pocket, with key residues important for the interaction with the diphenyl backbone of PBDE-100 shown as sticks. (**E**) Conserved binding site for PBDE-100. Top: Side chains found to interact with PBDE-100 are shown in blue (conserved in human and mouse) or green (not conserved). These residues are Y^{303}, Y^{306}, A^{307}, F^{310}, F^{331}, Q^{721}, F^{724}, S^{725}, I^{727}, F^{728}, V^{731}, S^{752}, F^{755}, S^{975}, and F^{979}. Bottom: Amino acid sequence alignment of mouse and human P-gp highlighting the 15 interacting residues with PBDE-100 in TM5, TM6, TM7, TM8, and TM12.

Environmentally relevant inhibitor mixtures impair the transport function of human P-gp

The chemicals that we found to inhibit P-gp are ubiquitous, legacy pollutants that have been widely reported in humans and wildlife. Humans are likely to be exposed to these compounds through consumption of contaminated food, raising questions on the environmental levels and impacts of these inhibitor mixtures. Previous studies have demonstrated that marine environments are major global repositories of POPs (34) and that fish can have high levels of these compounds as compared to other foods (35, 36). Thus, to determine the environmental levels of the 10 inhibitors, we measured them in wild-caught yellowfin tuna. We focused on tuna because they are among the most widely consumed fish in the world (37). More than 1 million metric tons of yellowfin tuna are caught annually, accounting for 27% of the global tuna catch (38).

White dorsal muscle fillets of eight fish (Figure 2.4A and Table 2.S2) were collected and analyzed for levels of OCPs, PCBs, PBDEs, and perfluorochemicals (PFCs) (Figure 2.S5). Nine of the 10 inhibitory POPs were detected in these tuna, and the mean lipid-normalized concentration of the nine inhibitors was 1.01 μM, accounting for approximately 23% of total POPs (Figure 2.4B and C, and Table 2.S2). PCBs accounted for 79% of total POPs (Figure 2.S5 and Table 2.S2) and 44% of inhibitors (Figure 2.4C and Table 2.S2).

On the basis of these data, a pollutant mixture was formulated, reflecting relative levels of inhibitors found in tuna (Figure 2.5A). We used this mixture to examine effects on the transport function of human P-gp. In this assay, inside-out membrane vesicles of mammalian cells overexpressing human P-gp were generated, and effects of POPs were assayed by measurement of the accumulation of the P-gp substrate *N*-methylquinidine (NMQ) into the vesicles (Figure 2.S6A). In the presence of the POP mixture, NMQ transport was inhibited with an IC_{50} of 28.7 μM and an IC_{10} of 7.1 μM (Figure 2.5B). We also verified the inhibition of NMQ transport by the individual compounds (Figure 2.S6B). The effects of the POP mixture on mouse P-gp ATPase activity were similar to those seen on human P-gp, with an IC_{50} of 25 μM and an IC_{10} of 5.7 μM (Figure 2.5C). Collectively, these results indicate that POPs identified in our assays are likely to exhibit similar inhibitory effects on human P-gp.

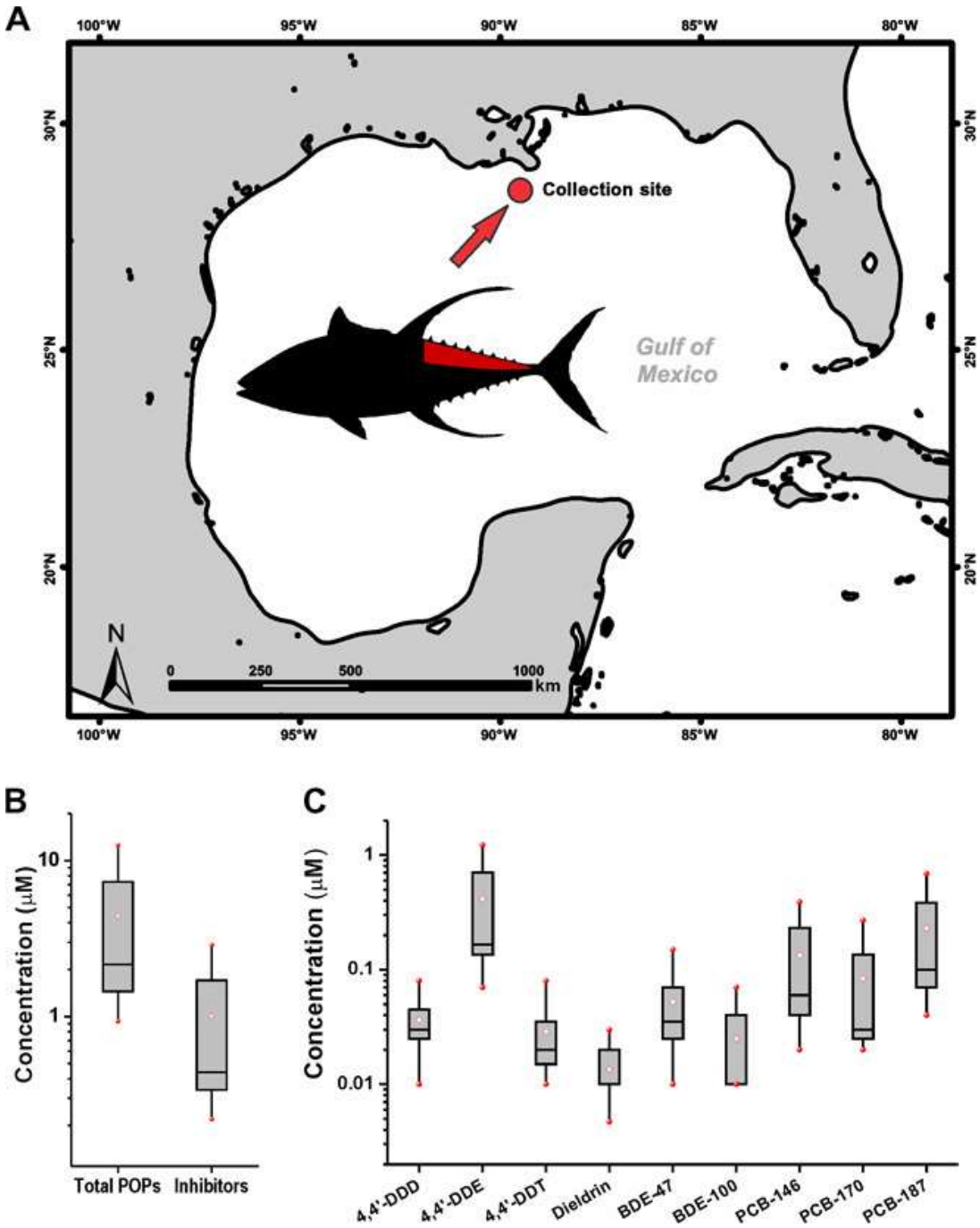


Figure 2.4. Levels of P-gp inhibitors in yellowfin tuna (*T. albacares*).
(**A**) Sampling site for the eight yellowfin tunas (*T. albacares*) caught in the GOM. The inset shows a yellowfin tuna with the sampled dorsal muscle tissue marked in red. (**B**) Lipid-normalized concentrations of the total POPs and the 10 P-gp inhibitors. The red-filled circles represent the minimum and maximum values. The white diamonds represent the mean value. The horizontal lines represent the 50th percentile, and the boxes represent the 25th and 75th percentiles. (**C**) Range of concentrations of nine inhibitory POPs measured in yellowfin tuna muscle from the GOM.

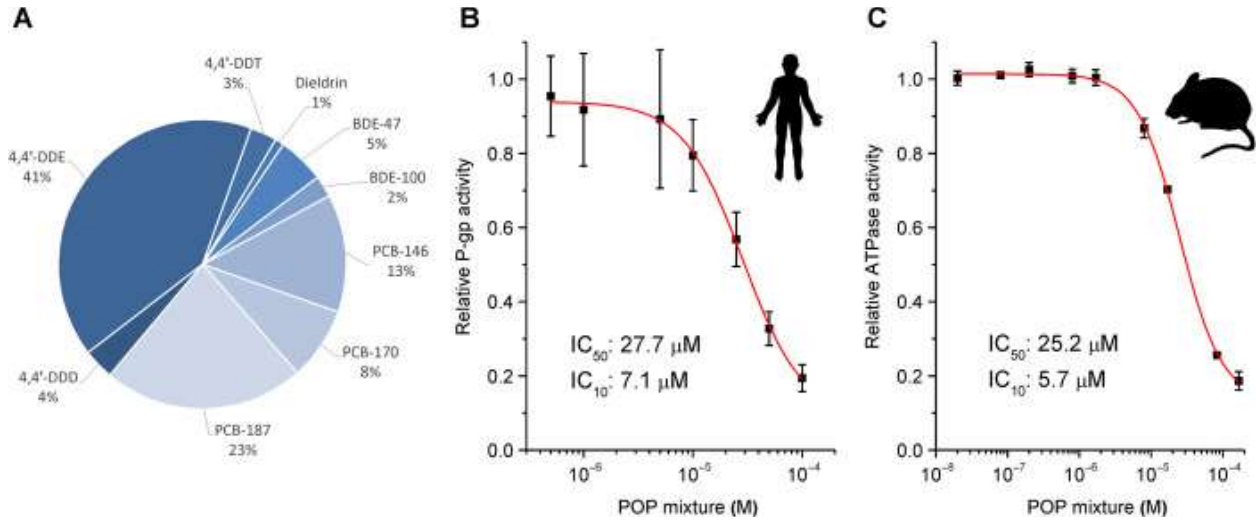


Figure 2.5. An environmentally relevant POP mixture inhibits the transport function of human and mouse P-gp.
(A) Relative ratio of the mean concentrations of P-gp inhibitors found in yellowfin tuna. (B) Inhibition of human P-gp by the POP mixture. Points represent the average percentage of NMQ uptake ± SD relative to the control from nine different experiments and with increasing concentration of the POP mixture. (C) Inhibition of verapamil-stimulated ATPase activity of mouse P-gp by the POP mixture. Shown is the respective dose-response curve as ATPase activity relative to 100 μM verapamil stimulation. The ATPase activity of the purified protein was measured in the presence of increasing concentrations of the POP mixture on the basis of the relative concentration of nine inhibitory POPs identified in this study. All data were fitted using a Hill function [$y = v_1 + (v_2 - v_1) * x^n/(k^n + x^n)$]. The R^2 value was >0.99.

DISCUSSION

Here, we show the direct binding of a pollutant to the ligand-binding pocket of P-gp and the potential for POPs to inhibit this transporter. POPs are ubiquitous contaminants and P-gp is a key protein for xenobiotic elimination in all animals. P-gp, along with cytochrome P450, is typically expressed on apical plasma membranes at sites of xenobiotic uptake (39, 40), such as the intestine, and plays a key role in first-pass elimination of dietary toxins. Thus, the inhibition of P-gp by POPs could represent a mechanism by which this critical cellular defense becomes compromised.

A notable finding of this study was that the levels of transporter inhibitors were sometimes high, raising concerns about transporter inhibition after consumption of highly contaminated foods and about exposure to inhibitors in vulnerable populations of humans and wildlife. High exposure could result from consumption of individual fish with elevated levels of inhibitors. For instance, in this study, although the mean tuna inhibitor concentration was 1.1 μM, two of our eight fish had 2.8 and 2.9 μM POP inhibitors and 11.9 and 12.6 μM total POPs (table S2). This is notable considering that we measured levels in lean cuts of tuna and that more fatty cuts or other types of fish (41) could have even higher levels.

Vulnerable populations could include human neonates, who can have limited intestinal xenobiotic metabolism activity in the first several months of life (42, 43) and, in certain places, can be exposed to high levels of POPs through breast milk. Indeed, in areas where DDT was used for malaria control, high concentrations (>28 μM) were reported in milk fat (44, 45). In wildlife, P-gp inhibitors may pose similar dangers to vulnerable life history stages and geographic populations. For instance, in the GOM, there are high levels of oil hydrocarbons (46, 47) and P-gp could be important for protection against their toxic effects (48–51). Inhibition of P-gp by POPs could be especially problematic for early life history stages because the developing fish heart is sensitive to polycyclic aromatic hydrocarbons (52, 53), and high levels of POPs could be maternally loaded through the yolk (54, 55).

Although our study sheds light on the potential effects of specific POPs on P-gp, the mechanisms of their inhibitory interactions remain incompletely understood. The compounds that we identified are hydrophobic halogenated compounds (Table 2.S3), with relatively high potential for passive movement from the environment into tissues (56). Although we did not directly determine the binding constants of these

POPs to P-gp, the cocrystal structure demonstrates that a representative compound from our study can form intimate interactions with the ligand-binding domain of the transporter. The high passive flux of POPs into cells and their specific binding properties could be factors that reduce the effectiveness of P-gp in eliminating POPs (57–59). In addition, the conformational changes of the protein that occur upon ligand binding (Figure 2.S3) might also play some as yet unknown role in the observed inhibition by POPs (60, 61).

Finally, it will be important for future studies to consider how metabolism could alter the inhibitory potential and/or transport of pollutants. For instance, at sites of excretion, metabolism pathways, including those involved in conjugation and oxidation, might mitigate some of the inhibitory effects of pollutants and alter their potential for transport. Indeed, in mice, PBDE-47 is conjugated to carrier proteins that improve its efflux by P-gp into urine (62). In addition, other ABC transporters, such as ABCC1, ABCC2, and ABCG2, show a broad tissue distribution and can serve as an additional line of defense by expelling modified and/or unmodified pollutants (16, 24, 40, 63, 64).

CONCLUSIONS

The idea that environmental pollutants could perturb P-gp was first postulated nearly 20 years ago (65); however, the precise mechanisms and environmental relevance have largely remained a mystery. Here, we show the inhibition of the transporter by ubiquitous pollutants found in fish, and likely in other foods. Transporter interactions are already among the criteria evaluated in assessing drug availability (1, 66), and our results argue for consideration of these interactions in risk assessment of environmental chemicals.

An important step taken by this study was solving the cocrystal structure of P-gp bound to a pollutant. The structure revealed a high degree of conservation in the PBDE-100 binding residues and provides a useful tool to help explain and potentially predict both similarities and differences in compound selectivity across species (67). These interactions could also help experimentally define binding sites that are responsible for the adaptive evolution of P-gp into anthropogenic pollutants (68) and could prove applicable to the design of chemicals (69–71) with better potential for elimination by transporters. Indeed, with other transporter structures almost within reach, transporter interactions could soon represent an exciting new avenue for green chemistry.

EXPERIMENTAL PROCEDURES

Chemicals

All OCPs, cyclosporine A (CsA), verapamil, DOX, Triton X-100, pesticides, and dimethyl sulfoxide (DMSO) were purchased from Sigma. All PCB and PBDE congeners were purchased from AccuStandard. With the exception of verapamil, which was dissolved in water, all stock solutions were prepared in DMSO and diluted to the final concentrations in reaction buffer (ATPase assay) or sterile water (yeast assay). The yeast nitrogen base without amino acids was purchased from Fluka, and the amino acid dropout supplement (−Leu) was obtained from Clontech. The final DMSO concentration in the ATPase and yeast cell assays did not exceed 2 and 0.5%, respectively.

Expression of mouse P-gp in DOX-sensitive yeast

Mouse P-gp (MmABCB1a) was cloned into Bam HI/Hind III restriction sites of the constitutive yeast expression vector p415GPD (Ampr LEU2 ARS/CEN) (72). The hypersensitized MATa S. cerevisiae strain lacking three ABC transporters (Δpdr5 Δsnq2 Δyor1) (27) was transformed with p415GPD plasmids containing MmABCB1a using the standard lithium acetate/polyethylene glycol method (73). The empty p415GPD plasmid was transformed as a control. Transformed cells were selected on synthetic complete dropout (SC-Leu) plates. A single colony from each transformation was picked and cultured overnight in SC-Leu media at 30°C for the spot assays. Overnight cultures were washed in sterile water and their concentrations were adjusted to an OD600 (optical density at 600 nm) of 0.1. To test the functional expression of mouse ABCB1 in yeast cells, fivefold serial dilutions of transformed cells were spotted on SC-Leu and SC-Leu + 10 μM DOX plates and incubated at 30°C for 36 and 48 hours, respectively. For the chemosensitization assay, 15 μl of the culture of p415GPD and p415GPD-MmABCB1a transformed cells was soaked in indicated concentrations of pollutants and spotted on SC-Leu and SC-Leu + 10 μM DOX plates. SC-Leu plates were incubated at 30°C for 40 hours and SC-Leu + DOX plates were incubated at 30°C for 48 to 72 hours.

Expression and purification of mouse P-gp protein

To generate purified mouse P-gp, we overexpressed both genes in Pichia pastoris and purified the protein using combined affinity tag and size exclusion chromatography (SEC). The expression and purification of mouse P-gp in P. pastoris were described previously (18, 25, 26). Briefly, P. pastoris

transformed with yeast codon–optimized mouse P-gp (mouse ABCB1a, GenBank JF834158) harboring a C-terminal 6×His-tag was grown as 10-liter cultures in a Bioflow 415 bioreactor (New Brunswick Scientific), induced using slow methanol induction, and harvested. Cells were lysed at 40,000 psi by a single pass through a constant cell disrupter (TS-Series; Constant Systems Inc.), cell debris was separated by centrifugation at 12,500g, and membranes were isolated by centrifugation at 38,400g. Membranes were then solubilized and P-gp was purified using a Ni–nitrilotriacetic acid Superflow resin (Qiagen) via fast protein liquid chromatography (AKTA, GE Life Sciences). The protein was concentrated (Centricon YM-100; Millipore), ultracentrifuged, and subjected to SEC (Superdex 200 16/60; GE Healthcare). The protein concentration was determined using the Micro BCA Protein Assay Kit (Pierce).

To evaluate the purity of the recombinant P-gp, 5 μg was separated by electrophoresis in a 7.5% SDS–polyacrylamide gel electrophoresis gel and subjected to immunoblotting using a 6×His epitope tag antibody (Figure 2.S1A) (Pierce). The proteins were transferred to a polyvinylidene difluoride membrane by wet electroblotting (tank transfer) for 30 min at 100 V in transfer buffer [25 mM tris-base, 150 mM glycine, and 20% (v/v) methanol (pH 7.4)]. After the transfer, the membranes were blocked in 5% skim milk powder/1× TBST [0.1 M tris-base, 150 mM NaCl, and 0.05% Tween 20 (pH 7.4)] overnight at 23°C. After three washes with 1× TBST, the membranes were incubated for 2 hours with a 6×His epitope tag monoclonal antibody (mouse) in 5% skim milk powder/1× TBST at a ratio of 1:2000 (Pierce). Goat anti-mouse immunoglobulin G–horseradish peroxidase was used as secondary antibody in 5% skim milk powder/1× TBST at a ratio of 1:5000 for 1 hour at 23°C (Pierce). The proteins were visualized using the SuperSignal West Pico Chemiluminescent Substrate Kit (Pierce).

The yeast codon–optimized mouse P-gp used in this study has three N-glycosylation sites converted into glutamine (N83Q, N87Q, and N90Q) and a C-terminal His6-tag fusion. The calculated molecular mass of the resulting protein is ~142 kD, which was confirmed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) (Figure 2.S1B). MALDI MS analysis was used with a time-of-flight mass detector on a Voyager Mass Spectrometer DE-STR (Applied Biosystems) and an effective ion path length of 2 m in the positive ion reflector mode. Sinapinic acid in 50% acetonitrile and 0.1% trifluoroacetic acid was used as matrix solution. Samples (1 mg/ml) were diluted 1:20 with the matrix solution and 1 μl was spotted onto the MALDI sample target plates and air-dried. Spectra were obtained in

the mass range between 5000 and 200,000 daltons with 256 laser shots per spectrum. Accelerating voltage was 25,000 V with a grid voltage of 93% and a guide wire percentage of 0.3%. Spectra were recorded in delayed extraction with a delay time of 700 ns. Internal calibration was performed using bovine serum albumin standard (Sigma-Aldrich) with a calculated molecular mass of 66.5 kD. All data were analyzed using Voyager Data Explorer 4.0.0.0 (Applied Biosystems) and plotted using Origin 7.0 (OriginLab).

ATPase activity of purified mouse P-gp

To test for molecular interactions of pollutants with the P-gp transporters, we optimized an ATPase assay based on the malachite green method (28, 74). This assay is well established (29, 75), requiring only 1 to 2 µg of total protein per reaction for measurements of the ATPase activity of efflux transporters. Figure S1C shows the respective dose-response curves of mouse P-gp with the model ATPase stimulator verapamil and the model inhibitor CsA. Activation of mouse P-gp ATPase using verapamil resulted in a half maximal stimulation concentration of 9.4 ± 0.7 µM, which is in agreement with that found previously (76, 77). Verapamil-stimulated (100 µM) ATPase activity was inhibited by CsA with an IC50 of 1.3 ± 0.1 µM. Briefly, we used 2 µg of purified, solubilized mouse P-gp and added the protein to the wells of a chilled 96-well plate containing 60 µl of ATP-free reaction buffer [10 mM MgSO4, 0.05% (w/v) DDM, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and Escherichia coli polar extract lipids (0.1 mg/ml) in 50 mM tris-Cl buffer (pH 7.5)] with serial dilutions of verapamil (control activator). To test inhibition, we used serial dilutions of CsA (control inhibitor) or pollutant compounds plus 100 µM verapamil. Then, 60 µl of ATP solution [5 mM Na-ATP, 10 mM MgSO4, 0.05% (w/v) DDM, 1 mM TCEP, and E. coli polar extract lipids (0.1 mg/ml) in 50 mM tris-Cl buffer (pH 7.5)] was added, mixed, and incubated for 3 min on ice. After incubation, the reaction mixtures in the 96-well polymerase chain reaction plate were transferred to a thermocycler and the reaction was started with the following cycling parameters: 3 s at 4°C, 5 min at 37°C, 15 s at 80°C (heat inactivation), and hold at 4°C. ATPase reactions (30 µl) were transferred to a 96-well enzyme-linked immunosorbent assay plate and the liberated inorganic phosphate was measured by adding 150 µl of an activated color development solution [17 mg of malachite green in 3.75 ml of Milli-Q H2O and 0.525 g of ammonium molybdate tetrahydrate in 12.5 ml of 4N HCl, activated with 0.02% (v/v) Triton X-100] in each sample well. The absorbance of each sample was immediately measured at 600 nm in a microplate reader (Spectramax M2) to minimize acid-catalyzed ATP hydrolysis in the color development solution (pH ~1 to 2). Control

samples containing buffer and DMSO (CsA, pollutants) or H2O (verapamil) without added P-gp protein were subtracted as background values. Inorganic phosphate standards (KH2PO4) from 0.125 to 2 nmol served as controls. The ATPase activity data are given as mean ± SEM from three to six different measurements, and representative experiments are shown. For the calculation of IC50 values, the data were fitted to a Hill function: $y = v1 + (v2 - v1) * xn/(kn + xn)$, where v1 and v2 are the initial and final reaction velocities, respectively, n is the Hill coefficient or the cooperativity of the dependence on x, and k is the concentration of the inhibitor (IC50) or the stimulator (EC50) that corresponds to 50% of the maximum (Michaelis constant). All calculations were performed using Origin software.

Reductive methylation and crystallization of mouse P-gp

We added freshly made borane dimethylamine (50 mM) and formaldehyde (100 mM) to purified mouse P-gp (~1 to 2 mg). The mixture was incubated (2 hours at 4°C), and the reaction was quenched with glycine, washed via concentration in SEC buffer [20 mM tris-HCl (pH 8.0), 20 mM NaCl, 0.01% LMNG (lauryl maltose neopentyl glycol), 0.01% sodium cholate, 0.2 mM TCEP, and 0.5 mM EDTA], and treated with 0.2 to 0.5 mM PBDE-100 dissolved in DMSO for an overnight incubation in the dark (12 to 16 hours at 4°C). The sample was then diluted 1:10 in SEC buffer, concentrated to ~12 to 15 mg/ml, and used for crystal trials. P-gp crystals were grown in 24-well Cryschem plates (Hampton Research) at a protein concentration of ~12 to 15 mg/ml using 4-μl sitting drops at a ratio of 1:1 protein/mother liquor [0.1 M Hepes, 50 mM lithium sulfate, 10 mM EDTA, and 24 to 29.5% (w/v) polyethylene glycol (PEG) 600 (pH 7.0 to 8.4)]. Crystals grown at 4°C typically appeared after 1 to 3 days and continued to grow to full size in approximately 2 weeks. Collected crystals were first cryoprotected by soaking in 0.1 M Hepes (at a pH identical to the crystal growth condition), 50 mM lithium sulfate, 10 mM EDTA, and 32% PEG 600. Collected crystals were typically ~650 μm × 400 μm × 300 μm.

X-ray data collection, structure determination, and refinement of the mouse P-gp/PBDE-100 cocrystal structure

X-ray diffraction data were collected at 100 K at the Canadian Light Source (08ID-1). X-ray fluorescence spectra were collected to verify the presence of bromine in P-gp crystals soaked with PBDE-100, and subsequent multiwavelength anomalous dispersion scans were conducted around the Br-K edge to maximize their anomalous signal contribution during data collection. All diffraction data were indexed and

integrated with MOSFLM (78), processed with AIMLESS (79), and truncated with TRUNCATE within the CCP4 suite of programs (80). Phases were determined using the recently improved model of P-gp (26) as a search model in molecular replacement in Phaser (81). The resultant model underwent rigid-body and restrained positional refinement, with hydrogens applied in their riding positions, using PHENIX.REFINE (82) against a maximum likelihood target function with grouped B factors and secondary structure restraints. Rounds of refinement were interspersed with manual inspection and correction against Sigma-A–weighted electron density maps in Coot (83), and improvements to model geometry and stereochemistry were monitored using MolProbity (84). The side chains of residues proximal to PBDE-100 were modeled during the final rounds of refinement to avoid biasing their placement. Ligand description dictionaries were calculated using PHENIX.ELBOW (82), and the crystallographic position of PBDE-100 was validated using the anomalous scattering from the bromine atoms (Fig. 3C). The refined structure was judged to have excellent geometry as determined by MolProbity (84). The resulting refinement statistics are listed in Table 2.S1. Figures displayed in this paper were prepared using PyMOL (http://www.pymol.org). Atomic coordinates and structure factors for the cocrystal structure presented here were deposited with the Protein Data Bank (accession code 4XWK). Sequence alignments were performed using ClustalW2 Version 2.1 [European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI)], and molecular interfaces were examined using PDBePISA (EMBL-EBI).

Tuna collection and pollutant data analysis

Yellowfin tuna (T. albacares) were collected from the GOM (28° 29′ N, 89° 27′ W). Tuna muscles were sent to AXYS (AXYS Analytical Services Ltd.) for analysis. Pollutants were measured according to the following EPA (U.S. Environmental Protection Agency) methods: for OCPs, 608, 625, 1625, 8081, and 8270; for PBDEs, 1614 and 625; and for PCBs, 1668 and 8270. For PFCs, methods developed by AXYS were used (AXYS MLA-041, AXYS MLA-042, AXYS MLA-043, and AXYS MLA-060). Analyses of pollutant data were performed according to previous studies (35, 85). Where pollutant values were below the detection limit, the values were treated as nondetectable (ND). Most blank measurements were at or below the detection limits and thus not subtracted from samples. In runs where blank values were above detection limits, only samples with values more than twice the blank value were reported (after subtraction of the blank value). The remaining values were treated as ND. The “ND” values were assigned

zero when calculating the molar concentrations of POPs in the lipid fraction of the muscle tissue and for further statistical analysis, so as not to overestimate the level of pollutant in samples.

All tuna pollutant data were converted into micromolar lipid-normalized concentrations. Because the lipid composition of the tuna tissue samples was unknown, we assumed a lipid mass density of 1 for volume conversion (that is, 1 g of lipid is equal to 1 ml). To meet assumptions of normality and homogeneity of variance for statistical analyses, all data were log10-transformed. Zero values were excluded from the analyses. To avoid overestimating pollutant levels, one of nine fish initially collected from the GOM was treated as an outlier because of its high levels (more than two times the interquartile range) that did not appear to represent the concentration signature for the group. All data were analyzed and plotted using Origin software.

Vesicle inhibition assays and data analysis

Inhibition of the efflux of probe substrates into membrane vesicles expressing P-gp was carried out according to the manufacturer's instructions, with some modifications (Sigma-Aldrich). Briefly, vesicle membrane suspensions were added to a 96-well plate stored on ice. Incubation media containing the single POP congener, the POP mixture, or verapamil as positive control were added to the plate and incubated for 15 min. Substrate solutions containing either Mg-ATP or Mg-AMP (adenosine monophosphate) and probe substrate (NMQ) were added to the plate for the designated time (3 min). The incubation was ended by the addition of the ice-cold wash mix. The sample solution was transferred to a filter plate and washed five times with wash mix. Plates were allowed to dry at room temperature for approximately 1 hour, after which a 50:50 methanol/water solution containing the internal standard was added to the filter plate wells and incubated for 15 min before it was transferred to a 96-well analytical plate. For each inhibitor, three assays were performed in duplicate. IC50 values were determined from the average percent inhibition values from each experiment ± SE of the measurement.

Analyst Instrument Control and Data Processing Software (AB SCIEX, version 1.6.1) was used to analyze the unlabeled probe substrate in P-gp vesicle assays (NMQ) for data collection and integration, which were then processed with Microsoft Excel 2007 (Microsoft). Calibration standards were used to calculate concentration on the basis of analyte/internal standard peak-area ratios with Analyst Instrument Control and Data Processing Software (AB SCIEX, version 1.6.1). A Shimadzu API 4000 mass

spectrometer in positive mode (4500 V) was used in tandem with a Waters Atlantis (dC18, 5 μm, 100 × 2.1 mm) column (at 40°C) and a Luna C8 guard column (4.0 × 2.0 mm) for separation in a mobile phase of 0.2% formic acid in water and 0.2% formic acid in methanol and an injection volume of 1 μl. The mobile phase flow rate was 0.6 ml/min and the mass transitions used to identify NMQ were m/z (mass/charge ratio) = 339.1 and 339.3. Deuterated NMQ (d3-NMQ) was used as an internal standard and identified with mass transitions of m/z = 342.1 and 342.3.

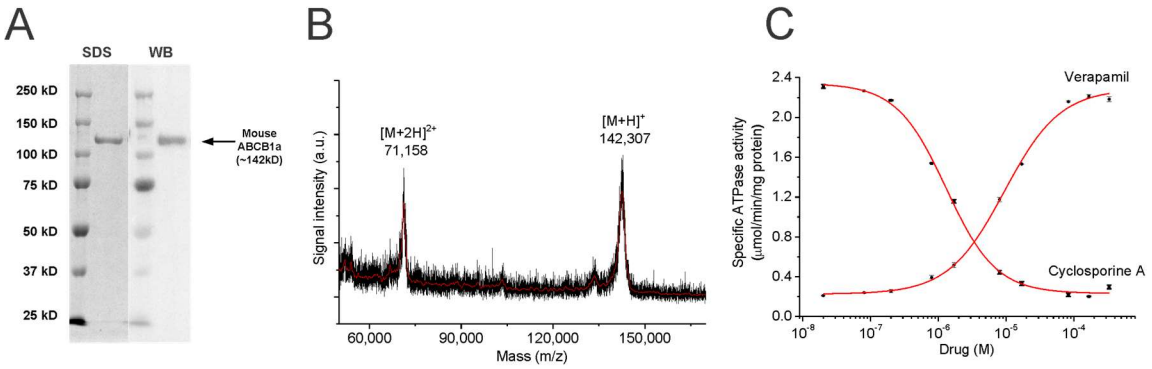


Figure 2.S1. ATPase activity of purified, recombinant mouse P-gp. (A) 7.5% SDS-PAGE with Coomassie blue staining and an immunoblot (WB) of mouse P-gp fused to a C-terminal His10-tag by anti-His antibody of purified P-gp protein. (B) MALDI TOF mass spectra of the purified mouse P-gp. Shown are two differently charged species of the protein (single and double charged ions). (C) ATPase activity of purified mouse P-gp using the malachite green method. ATPase activation and inhibition was determined with increasing concentrations of verapamil or cyclosporine A. Data points show specific ATPase activity from at least three independent experiments. Lines represent non-linear regression analysis of the data points with a Hill equation ($y=v1+(v2-v1)*x^n/(k^n+x^n)$). R2 values for the data fits were >0.99.

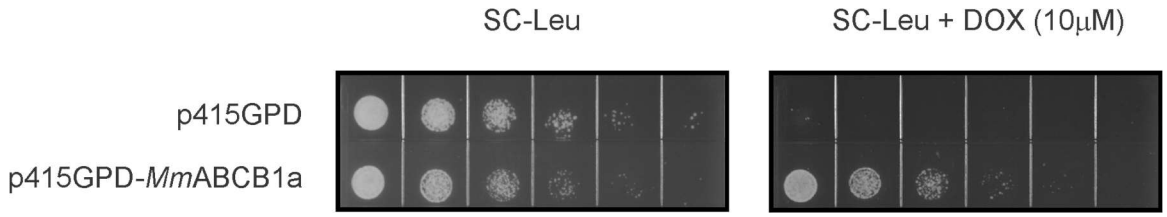


Figure 2.S2. Functional expression of mouse P-gp in yeast cells.
Hyper-sensitized MATα *S. cerevisiae* control strains and strains transformed with the p415GPD plasmid containing mouse ABCB1 were spotted in five-fold serial dilutions on SC-Leu (left panel) and SC-Leu + 10 μM doxorubicin (right panel) plates**.**

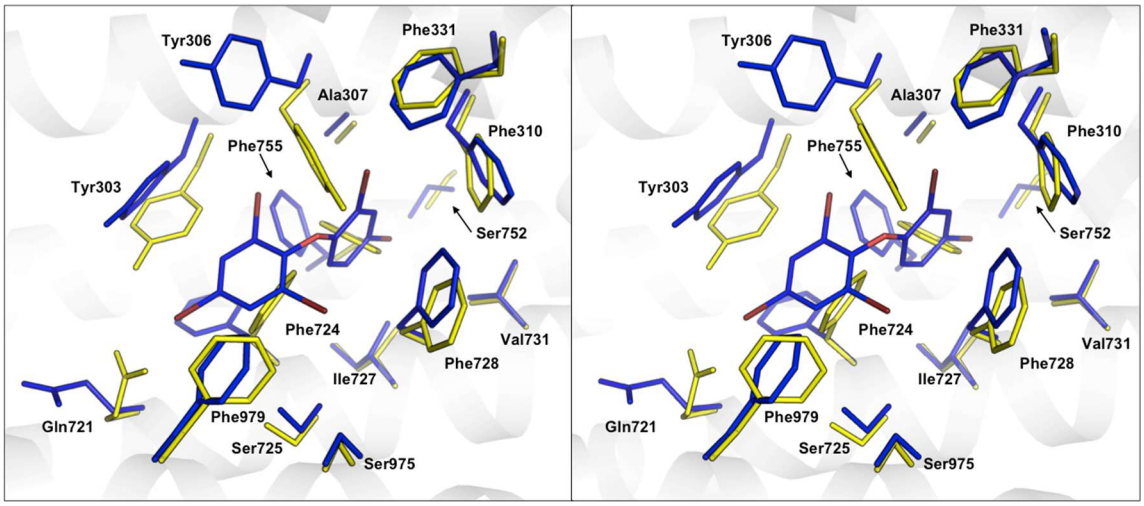


Figure 2.S3. Comparison of PBDE-100 interacting residues. Alternative side-chain conformations adopted by P-gp ligand-interacting residues upon PBDE-100 binding are shown in blue compared to ligand-free P-gp (PDB code 4Q9H) shown in yellow.

```
Human MDR1        YASYALAFWYGTTLVLSGEYSIGQVLTVFFSVLIGAFSVGQASPSIEAFANARGAAYEIF 366
Mouse ABCB1a      YASYALAFWYGTSLVISKEYSIGQVLTVFFSVLIGAFSVGQASPNIEAFANARGAAYEVF 362
Zebrafish ABCB4   YMSYALAFWYGSTLILGGEYTIGMLLTIFFAVLIGAFGLGQTSPNIQTFSSARGAAHKVF 371
Sea urchin ABCB1a FSSYAIAFWYGTVLYLDNEITPGDILTTFLAVLFGAFAIGQAGPNYSDFTTARAAASSIW 410
Conservation

Human MDR1        INGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLLFLALGIISFITFFLQGFTFGK 779
Mouse ABCB1a      INGGLQPAFSVIFSKVVGVFTNGGPPETQRQNSNLFSLLFLILGIISFITFFLQGFTFGK 775
Zebrafish ABCB4   INGGMQPAFAVIFSKIIAVFAE-PDQNLVRQRCDLYSLLFAGIGVLSFFTLFLQGFCFGK 777
Sea urchin ABCB1a INGAVQPGFAVVFSKILGAYS-ITDRAALFDEVTIYCVLFAALGLLSLLASIIQGVGFGK 828
Conservation

Human MDR1        FGAYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSFAPDYAKAKISAAHIIMIIEKTPLI 1018
Mouse ABCB1a      FGAYLVTQQLMTFENVLLVFSAIVFGAMAVGQVSSFAPDYAKATVSASHIIRIIEKTPEI 1014
Zebrafish ABCB4   FGSWLIEQKLMTFEGVFLVISAVVYGAMAVGEANSFTPNYAKAKMSASHVLMLINRAPAI 1016
Sea urchin ABCB1a LGAHLVGTGDLTFPDVFLSFSALMFGAFGLGRAAGSVPDFSKAKVATGELFYLVDRSPDI 1068
Conservation
```

Figure 2.S4. Amino acid sequence alignment of human, mouse, zebrafish, and sea urchin P-gp. PBDE-interacting residues are marked in red (conserved) and orange (non-conserved).

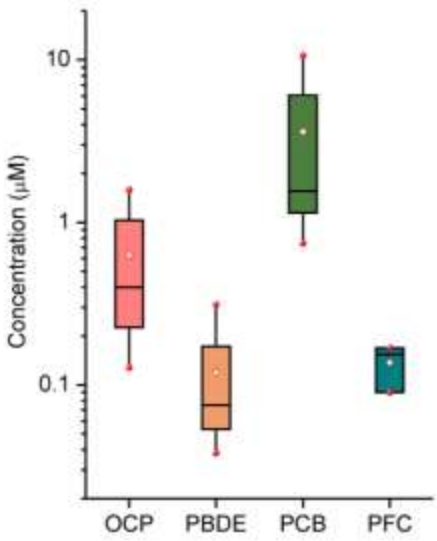


Figure 2.S5. Levels of POPs in wild-caught yellowfin tuna. Box and whisker graphs represent the lipid normalized range of concentrations in micromolar of OCPs, PBDEs, PCBs, and PFC for eight fish caught in the Gulf of Mexico. The red-filled circles represent the minimum and maximum values. The white diamond represents the mean value. The horizontal line represents the 50th percentile, and the box represents the 25th and 75th percentiles.

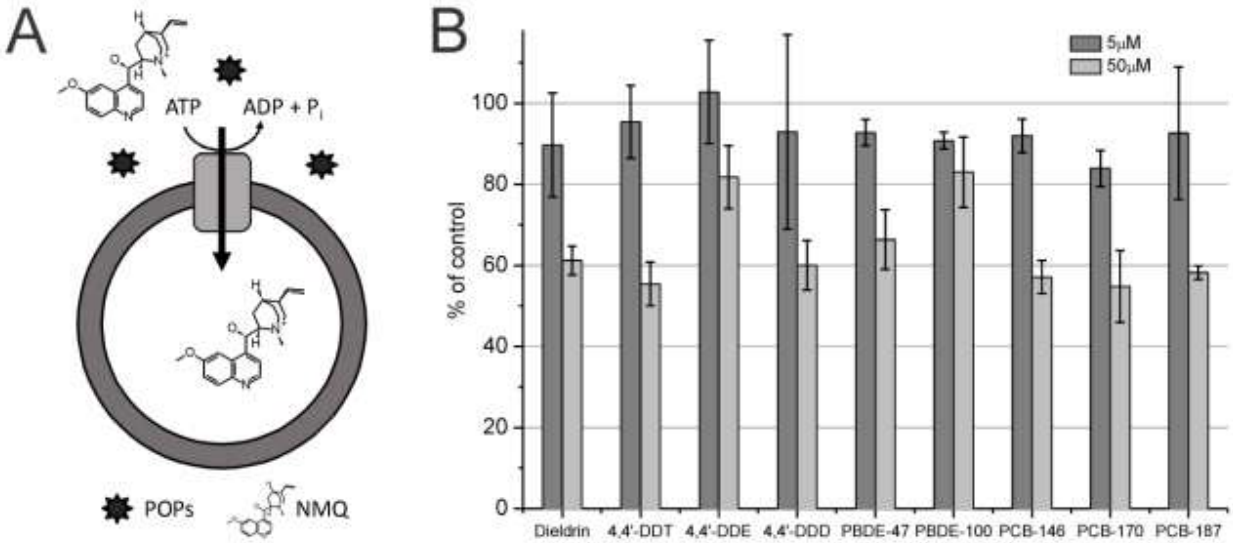


Figure 2.S6. Individual POP congeners from the mixture inhibiting human P-gp. (A) Schematic illustration of the vesicular inhibition assay using human P-gp and NMQ as a substrate. (B) Inhibition of human P-gp by the 9 different congeners from the POP mixture. Bar graphs represent the average percentage of NMQ uptake ±S.D. relative to the control from 9 different experiments and at two different concentrations for each compound.

Table 2.S1. Data collection and refinement statistics of the mouse P-gp/PBDE-100 cocrystal structure.

	BDE100-Bound
PDB ID	
Data collection	
Wavelength	0.91929
Beamline	CLS-08ID
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	89.0, 138.2, 185.1
α, β, γ (°)	90, 90, 90
Resolution range (Å)	92.6-3.5 (3.7-3.5)
Number of crystals	1
R_{merge} (%)	7.7 (74.3)*
R_{pim}	5.5 (51.8)
Observed reflections	156577
Unique reflections	29261
Mean ((I/sd(I))	10.2 (2.2)
$CC_{1/2}$	0.998 (0.754)
Completeness (%)	99.3(99.9)
Redundancy	5.4 (5.4)
Refinement	
Resolution range (Å)	92.6-3.5 (3.6-3.5)
Reflections in working set	29204
Reflections in test set	1379
R_{work}/ R_{free}	26.3/28.2
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.988
<*B*> (Å^2), P-gp	139.17
<*B*> (Å^2), BDE100	178.16
Ramachandran statistics	
Outliers %	0
Favoured %	95.8
Rotomer Outliers %	0.6
Cβ Deviations	0

*Highest resolution shell is shown in parenthesis.

Table 2.S2. Metadata on yellowfin tuna specimens used in this study. Listed are the individual standard lengths and % of lipid from the fish caught in the Gulf of Mexico. The lipid-normalized levels of OCPs, PBDEs, PCBs, and PFCs as well as the sum of total POPs and the 10 P-gp inhibitors for each fish are shown in micromolar concentration. The fraction of the concentration of the ten inhibitors relative to the total POPs is listed in percent. ND = not detected.

Fish [#]	Standard length [cm]	Lipid [%]	ΣOCP [μM]	ΣPBDE [μM]	ΣPCB [μM]	ΣPFC [μM]	Total POPs [μM]	Inhibitors [μM]	Fraction [%]
1	114.3	0.36	1.59	0.26	10.10	ND	11.94	2.83	23.7
2	108.0	0.39	1.59	0.31	10.65	ND	12.56	2.88	22.9
3	86.4	1.29	0.48	0.09	2.04	0.09	2.69	0.59	22.1
4	99.1	2.33	0.47	0.08	1.53	0.15	2.23	0.44	19.7
5	99.1	1.19	0.33	0.07	1.53	ND	1.94	0.44	22.9
6	91.4	0.68	0.27	0.06	1.59	0.17	2.09	0.44	20.9
7	101.6	0.47	0.13	0.05	0.76	ND	0.93	0.22	23.6
8	99.1	0.72	0.18	0.04	0.74	ND	0.96	0.24	25.0

Table 2.S3. Physical and chemical properties of the 10 POP inhibitors. Listed are the molecular weights (MW) in g/mol, the octanol/water partition coefficients (Log K_{ow}), and the water solubilities in mg/L for each of the inhibitory compounds according to the respective ATSDR toxicological profile database (89-94).

POP congener	MW	Log K_{ow}	Water solubility at 25°C
Dieldrin	380.91	6.2	0.110
Endrin	380.91	5.34-5.6[a]	0.2
4,4' DDT (p,p' DDT)	354.49	6.91	0.025
4,4' DDE (p,p' DDE)	318.03	6.51	0.12
4,4' DDD (p,p' DDD)	320.05	6.02	0.090
BDE-47	485.79	6.81	0.015
BDE-100	564.69	7.24	0.04
PCB-146	360.88	6.7[b]	n.a.
PCB-170	395.32	7.1[b]	n.a.
PCB-187	395.32	7.1[b]	0.00047[c]

[a]Calculated values according to (91). [b]Values obtained from the Certificate of Analysis, Accustandard Inc. [c]Values according to (89).

ACKNOWLEDGEMENTS

We thank G. Moy for technical assistance, R. Doshi and I. L. Urbatsch for helpful discussions on ATPase assays, and R. J. Ferl and J. D. Rine for the yeast strain and the p415GPD plasmid. We thank B. Zgliczynski for providing the GOM map. We thank D. Epel, V. D. Vacquier, and J. Stegeman for discussions of earlier versions of the manuscript. **Funding:** This work was supported by the program on Oceans and Human Health through NIH grant ES021985 and NSF grant 1314480 to A.H. and G.C., by a WAITT Foundation grant to A.H. and S.S., and by a University of California San Diego Academic Senate Grant to A.H. A.P.M. was supported by a National Health and Medical Research Council CJ Martin Postdoctoral Research Fellowship. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. The GOM map file was created using ArcGIS Desktop v10.3.1 and is licensed under the Educational Site License.

AUTHOR CONTRIBUTIONS

S.C.T.N. designed the research, conducted the assays, analyzed the data, and co-authored the manuscript. S.D.R. and A.P.M. conducted x-ray and protein purification work, and edited the manuscript. T.G. designed and performed the yeast assays, and edited the manuscript. L.T.B. collected tuna samples and analyzed the data. L.M.V. performed vesicle assays and analyzed the data. C.C. conducted protein purification work. G.L., S.S., and G.C. designed and supervised the research, and edited the manuscript. A.H. analyzed the data, designed the research, co-authored the manuscript, and supervised the project.

DISSERTATION ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in *Science Advances*. Nicklisch SC, Rees SD*, McGrath AP*, Gökirmak T, Bonito LT, Vermeer LM, Cregger C, Loewen G, Sandin S, Chang G, Hamdoun A. Global marine pollutants inhibit P-glycoprotein: Environmental levels, inhibitory effects, and co-crystal structure. April 2016. This dissertation author is a co-secondary author of this paper.

REFERENCES

1. International Transporter Consortium Giacomini K. M., Huang S.-M., Tweedie D. J., Benet L. Z., Brouwer K. L. R., Chu X., Dahlin A., Evers R., Fischer V., Hillgren K. M., Hoffmaster K. A., Ishikawa T., Keppler D., Kim R. B., Lee C. A., Niemi M., Polli J. W., Sugiyama Y., Swaan P. W., Ware J. A., Wright S. H., Yee S. W., Zamek-Gliszczynski M. J., Zhang L., Membrane transporters in drug development. Nat. Rev. Drug Discov. 9, 215–236 (2010).
2. Galgani F., Cornwall R., Toomey B. H., Epel D. D., Interaction of environmental xenobiotics with a multixenobiotic defense mechanism in the bay mussel *Mytilus galloprovincialis* from the coast of California. Environ. Toxicol. Chem. 15, 325–331 (1996).
3. Bain L. J., McLachlan J. B., LeBlanc G. A., Structure-activity relationships for xenobiotic transport substrates and inhibitory ligands of P-glycoprotein. Environ. Health Perspect. 105, 812–818 (1997).
4. Luckenbach T., Epel D., Nitromusk and polycyclic musk compounds as long-term inhibitors of cellular xenobiotic defense systems mediated by multidrug transporters. Environ. Health Perspect. 113, 17–24 (2005).
5. Shabbir A., DiStasio S., Zhao J., Cardozo C. P., Wolff M. S., Caplan A. J., Differential effects of the organochlorine pesticide DDT and its metabolite *p*,*p'*-DDE on p-glycoprotein activity and expression. Toxicol. Appl. Pharmacol. 203, 91–98 (2005).
6. Sreeramulu K., Liu R., Sharom F. J., Interaction of insecticides with mammalian P-glycoprotein and their effect on its transport function. Biochim. Biophys. Acta 1768, 1750–1757 (2007).
7. Stevenson C. N., MacManus-Spencer L. A., Luckenbach T., Luthy R. G., Epel D., New perspectives on perfluorochemical ecotoxicology: Inhibition and induction of an efflux transporter in the marine mussel, *Mytilus californianus*. Environ. Sci. Technol. 40, 5580–5585 (2006).
8. Anselmo H. M. R., van den Berg J. H. J., Rietjens I. M. C. M., Murk A. J., Inhibition of cellular efflux pumps involved in multi xenobiotic resistance (MXR) in echinoid larvae as a possible mode of action for increased ecotoxicological risk of mixtures. Ecotoxicology 21, 2276–2287 (2012).
9. Bircsak K. M., Richardson J. R., Aleksunes L. M., Inhibition of human MDR1 and BCRP transporter ATPase activity by organochlorine and pyrethroid insecticides. J. Biochem. Mol. Toxicol. 27, 157–164 (2013).
10. Dankers A. C. A., Roelofs M. J. E., Piersma A. H., Sweep F. C. G. J., Russel F. G. M., van den Berg M., van Duursen M. B. M., Masereeuw R., Endocrine disruptors differentially target ATP-binding cassette transporters in the blood-testis barrier and affect leydig cell testosterone secretion in vitro. Toxicol. Sci.136, 382–391 (2013).
11. Smital T., Luckenbach T., Sauerborn R., Hamdoun A. M., Vega R. L., Epel D., Emerging contaminants—Pesticides, PPCPs, microbial degradation products and natural substances as inhibitors of multixenobiotic defense in aquatic organisms. Mutat. Res. 552, 101–117 (2004).
12. Buss D. S., Callaghan A., Interaction of pesticides with p-glycoprotein and other ABC proteins: A survey of the possible importance to insecticide, herbicide and fungicide resistance. Pestic. Biochem. Phys.90, 141–153 (2008).
13. Juliano R. L., Ling V., A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim. Biophys. Acta 455, 152–162 (1976).
14. Ambudkar S. V., Dey S., Hrycyna C. A., Ramachandra M., Pastan I., Gottesman M. M., Biochemical, cellular, and pharmacological aspects of the multidrug transporter. Annu. Rev. Pharmacol. Toxicol. 39, 361–398 (1999).
15. Dean M., Annilo T., Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu. Rev. Genomics Hum. Genet. 6, 123–142 (2005).

16. Leslie E. M., Deeley R. G., Cole S. P. C., Multidrug resistance proteins: Role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. Toxicol. Appl. Pharmacol. 204, 216–237 (2005).

17. Gottesman M. M., Fojo T., Bates S. E., Multidrug resistance in cancer: Role of ATP–dependent transporters. Nat. Rev. Cancer 2, 48–58 (2002).

18. Aller S. G., Yu J., Ward A., Weng Y., Chittaboina S., Zhuo R., Harrell P. M., Trinh Y. T., Zhang Q., Urbatsch I. L., Chang G., Structures of P-glycoproteins reveals a molecular basis for poly-specific drug binding. Science 323, 1718–1722 (2009).

19. Borst P., Elferink R. O., Mammalian ABC transporters in health and disease. Annu. Rev. Biochem. 71, 537–592 (2002).

20. Martin C., Berridge G., Higgins C. F., Mistry P., Charlton P., Callaghan R., Communication between multiple drug binding sites on P-glycoprotein. Mol. Pharmacol. 58, 624–632 (2000).

21. Ferreira R. J., Ferreira M.-J. U., dos Santos D. J. V. A., Molecular docking characterizes substrate-binding sites and efflux modulation mechanisms within P-glycoprotein. J. Chem. Inf. Model. 53, 1747–1760 (2013).

22. A. Sturm, H. Segner, in *Biochemistry and Molecular Biology of Fishes*, T. P. Mommsen, T. W. Moon, Eds. (Elsevier B.V., Amsterdam, Netherlands, 2005), vol. 6, pp. 495–533.

23. Epel D., Luckenbach T., Stevenson C. N., MacManus-Spencer L. A., Hamdoun A., Smital T., Efflux transporters: Newly appreciated roles in protection against pollutants. Environ. Sci. Technol. 42, 3914–3920 (2008).

24. Döring B., Petzinger E., Phase 0 and phase III transport in various organs: Combined concept of phases in xenobiotic transport and metabolism. Drug Metab. Rev. 46, 261–282 (2014).

25. Ward A. B., Szewczyk P., Grimard V., Lee C.-W., Martinez L., Doshi R., Caya A., Villaluz M., Pardon E., Cregger C., Swartz D. J., Falson P. G., Urbatsch I. L., Govaerts C., Steyaert J., Chang G., Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. Proc. Natl. Acad. Sci. U.S.A. 110, 13386–13391 (2013).

26. Szewczyk P., Tao H., McGrath A. P., Villaluz M., Rees S. D., Lee S. C., Doshi R., Urbatsch I. L., Zhang Q., Chang G., Snapshots of ligand entry, malleable binding and induced helical movement in P-glycoprotein. Acta Crystallogr. D Biol. Crystallogr. 71, 732–741 (2015).

27. Jeong H., Herskowitz I., Kroetz D. L., Rine J., Function-altering SNPs in the human multidrug transporter gene *ABCB1* identified using a *Saccharomyces*-based assay. PLOS Genet. 3, e39 (2007).

28. Van Veldhoven P. P., Mannaerts G. P., Inorganic and organic phosphate measurements in the nanomolar range. Anal. Biochem. 161, 45–48 (1987).

29. Henkel R. D., VandeBerg J. L., Walsh R. A., A microassay for ATPase. Anal. Biochem. 169, 312–318 (1988).

30. Sarkadi B., Price E. M., Boucher R. C., Germann U. A., Scarborough G. A., Expression of the human multidrug resistance cDNA in insect cells generates a high activity drug-stimulated membrane ATPase. J. Biol. Chem. 267, 4854–4858 (1992).

31. Doige C. A., Yu X., Sharom F. J., ATPase activity of partially purified P-glycoprotein from multidrug-resistant Chinese hamster ovary cells. Biochim. Biophys. Acta 1109, 149–160 (1992).

32. Tampal N. M., Robertson L. W., Srinivasan C., Ludewig G., Polychlorinated biphenyls are not substrates for the multidrug resistance transporter-1. Toxicol. Appl. Pharmacol. 187, 168–177 (2003).

33. Lowe C. J., Clarke D. N., Medeiros D. M., Rokhsar D. S., Gerhart J., The deuterostome context of chordate origins. Nature 520, 456–465 (2015).

34. Lohmann R., Jurado E., Pilson M. E. Q., Dachs J., Oceanic deep water formation as a sink of persistent organic pollutants. Geophys. Res. Lett. 33, L12607 (2006).

35. Schecter A., Päpke O., Harris T. R., Tung K. C., Musumba A., Olson J., Birnbaum L., Polybrominated diphenyl ether (PBDE) levels in an expanded market basket survey of U.S. food and estimated PBDE dietary intake by age and sex. Environ. Health Perspect. 114, 1515–1520 (2006).

36. Schecter A., Colacino J., Haffner D., Patel K., Opel M., Päpke O., Birnbaum L., Perfluorinated compounds, polychlorinated biphenyls, and organochlorine pesticide contamination in composite food samples from Dallas, Texas, USA. Environ. Health Perspect. 118, 796–802 (2010).

37. Loke M., Geslani C., Takenaka B., Leung P., An overview of seafood consumption and supply sources: Hawai'i versus U.S. Econ. Issues 22, 1–9 (2012).

38. ISSF Tuna Stock Status Update, 2015: Status of the world fisheries for tuna (ISSF Technical Report 2015-03A, International Seafood Sustainability Foundation, Washington, 2015).

39. Lin J. H., Yamazaki M., Role of P-glycoprotein in pharmacokinetics: Clinical implications. Clin. Pharmacokinet. 42, 59–98 (2003).

40. Dietrich C. G., Geier A., Oude Elferink R. P. J., ABC of oral bioavailability: Transporters as gatekeepers in the gut. Gut 52, 1788–1795 (2003).

41. Hites R. A., Foran J. A., Carpenter D. O., Hamilton M. C., Knuth B. A., Schwager S. J., Global assessment of organic contaminants in farmed salmon. Science 303, 226–229 (2004).

42. Johnson T. N., Tanner M. S., Taylor C. J., Tucker G. T., Enterocytic CYP3A4 in a paediatric population: Developmental changes and the effect of coeliac disease and cystic fibrosis. Br. J. Clin. Pharmacol. 51, 451–460 (2001).

43. Johnson T. N., Thomson M., Intestinal metabolism and transport of drugs in children: The effects of age and disease. J. Pediatr. Gastroenterol. Nutr. 47, 3–10 (2008).

44. Smith D., Worldwide trends in DDT levels in human breast milk. Int. J. Epidemiol. 28, 179–188 (1999).

45. Bouwman H., Kylin H., Sereda B., Bornman R., High levels of DDT in breast milk: Intake, risk, lactation duration, and involvement of gender. Environ. Pollut. 170, 63–70 (2012).

46. Crone T. J., Tolstoy M., Magnitude of the 2010 Gulf of Mexico oil leak. Science 330, 634 (2010).

47. Reddy C. M., Arey J. S., Seewald J. S., Sylva S. P., Lemkau K. L., Nelson R. K., Carmichael C. A., McIntyre C. P., Fenwick J., Ventura G. T., Van Mooy B. A. S., Camilli R., Composition and fate of gas and oil released to the water column during the *Deepwater Horizon* oil spill. Proc. Natl. Acad. Sci. U.S.A. 109, 20229–20234 (2012).

48. Lampen A., Ebert B., Stumkat L., Jacob J., Seidel A., Induction of gene expression of xenobiotic metabolism enzymes and ABC-transport proteins by PAH and a reconstituted PAH mixture in human Caco-2 cells. Biochim. Biophys. Acta 1681, 38–46 (2004).

49. Vaché C., Camares O., De Graeve F., Dastugue B., Meiniel A., Vaury C., Pellier S., LEoz-Garziandia E., Bamdad M., *Drosophila melanogaster* P-glycoprotein: A membrane detoxification system toward polycyclic aromatic hydrocarbon pollutants. Environ. Toxicol. Chem. 25, 572–580 (2006).

50. Hamdoun A. M., Griffin F. J., Cherr G. N., Tolerance to biodegraded crude oil in marine invertebrate embryos and larvae is associated with expression of a multixenobiotic resistance transporter. Aquat. Toxicol. 61, 127–140 (2002).

51. Valton E., Amblard C., Wawrzyniak I., Penault-Llorca F., Bamdad M., P-gp expression in brown trout erythrocytes: Evidence of a detoxification mechanism in fish erythrocytes. Sci. Rep. 3, 3422 (2013).

52. Incardona J. P., Gardner L. D., Linbo T. L., Brown T. L., Esbaugh A. J., Mager E. M., Stieglitz J. D., French B. L., Labenia J. S., Laetz C. A., Tagal M., Sloan C. A., Elizur A., Benetti D. D., Grosell M., Block B. A., Scholz N. L., *Deepwater Horizon* crude oil impacts the developing hearts of large predatory pelagic fish. Proc. Natl. Acad. Sci. U.S.A. 111, E1510–E1518 (2014).

53. Brette F., Machado B., Cros C., Incardona J. P., Scholz N. L., Block B. A., Crude oil impairs cardiac excitation-contraction coupling in fish. Science 343, 772–776 (2014).

54. Mull C. G., Lyons K., Blasius M. E., Winkler C., O'Sullivan J. B., Lowe C. G., Evidence of maternal offloading of organic contaminants in white sharks (*Carcharodon carcharias*). PLOS One 8, e62886 (2013).

55. Lyons K., Adams D. H., Maternal offloading of organochlorine contaminants in the yolk-sac placental scalloped hammerhead shark (*Sphyrna lewini*). Ecotoxicology 24, 553–562 (2015).

56. Mackay D., Fraser A., Bioaccumulation of persistent organic chemicals: Mechanisms and models. Environ. Pollut. 110, 375–391 (2000).

57. Cole B. J., Hamdoun A., Epel D., Cost, effectiveness and environmental relevance of multidrug transporters in sea urchin embryos. J. Exp. Biol. 216, 3896–3905 (2013).

58. Eytan G. D., Regev R., Oren G., Assaraf Y. G., The role of passive transbilayer drug movement in multidrug resistance and its modulation. J. Biol. Chem. 271, 12897–12902 (1996).

59. Regev R., Katzir H., Yeheskely-Hayon D., Eytan G. D., Modulation of P-glycoprotein-mediated multidrug resistance by acceleration of passive drug permeation across the plasma membrane. FEBS J.274, 6204–6214 (2007).

60. Raub T. J., P-glycoprotein recognition of substrates and circumvention through rational drug design. Mol. Pharm. 3, 3–25 (2006).

61. Chufan E. E., Kapoor K., Ambudkar S. V., Drug–protein hydrogen bonds govern the inhibition of the ATP hydrolysis of the multidrug transporter P-glycoprotein. Biochem. Pharmacol. 101, 40–53 (2015).

62. Emond C., Sanders J. M., Wikoff D., Birnbaum L. S., Proposed mechanistic description of dose-dependent BDE-47 urinary elimination in mice using a physiologically based pharmacokinetic model. Toxicol. Appl. Pharmacol. 273, 335–344 (2013).

63. van Herwaarden A. E., Schinkel A. H., The function of breast cancer resistance protein in epithelial barriers, stem cells and milk secretion of drugs and xenotoxins. Trends Pharmacol. Sci. 27, 10–16 (2006).

64. Chhabra R. S., Intestinal absorption and metabolism of xenobiotics. Environ. Health Perspect. 33, 61–69 (1979).

65. Kurelec B., A new type of hazardous chemical: The chemosensitizers of multixenobiotic resistance. Environ. Health Perspect. 105 (Suppl. 4), 855–860 (1997).

66. Montanari F., Ecker G. F., Prediction of drug–ABC-transporter interaction—Recent advances and future challenges. Adv. Drug Deliv. Rev. 86, 17–26 (2015).

67. Gökirmak T., Campanale J. P., Shipp L. E., Moy G. W., Tao H., Hamdoun A., Localization and substrate selectivity of sea urchin multidrug (MDR) efflux transporters. J. Biol. Chem. 287, 43876–43883 (2012).

68. Wirgin I., Roy N. K., Loftus M., Chambers R. C., Franks D. G., Hahn M. E., Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River. Science 331, 1322–1325 (2011).

69. Anastas P., Eghbali N., Green chemistry: Principles and practice. Chem. Soc. Rev. 39, 301–312 (2010).

70. Zimmerman J. B., Anastas P. T., Toward substitution with no regrets. Science 347, 1198–1199 (2015).

71. *A Framework to Guide Selection of Chemical Alternatives* (National Research Council, The National Academies Press, Washington, DC, 2014), 334 pp.

72. Mumberg D., Müller R., Funk M., Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156, 119–122 (1995).

73. R. D. Gietz, in *Methods in Molecular Biology*, W. Xiao, Ed. (Humana Press, Clifton, NJ, 2014), vol. 1205, pp. 1–12.

74. Urbatsch I. L., Al-Shawi M. K., Senior A. E., Characterization of the ATPase activity of purified Chinese hamster P-glycoprotein. Biochemistry 33, 7069–7076 (1994).

75. Venter H., Velamakanni S., Balakrishnan L., van Veen H. W., On the energy-dependence of Hoechst 33342 transport by the ABC transporter LmrA. Biochem. Pharmacol. 75, 866–874 (2008).

76. Swartz D. J., Weber J., Urbatsch I. L., P-glycoprotein is fully active after multiple tryptophan substitutions. Biochim. Biophys. Acta 1828, 1159–1168 (2013).

77. Bai J., Swartz D. J., Protasevich I. I., Brouillette C. G., Harrell P. M., Hildebrandt E., Gasser B., Mattanovich D., Ward A., Chang G., Urbatsch I. L., A gene optimization strategy that enhances production of fully functional P-glycoprotein in *Pichia pastoris*. PLOS One 6, e22577 (2011).

78. Battye T. G. G., Kontogiannis L., Johnson O., Powell H. R., Leslie A. G. W., *iMOSFLM*: A new graphical interface for diffraction-image processing with *MOSFLM*. Acta Crystallogr. D Biol. Crystallogr.67, 271–281 (2011).

79. P. Evans, in *Acta Crystallographica Section D: Biological Crystallography* (International Union of Crystallography, Chester, UK, 2006), vol. 62, pp. 72–82.

80. Winn M. D., Ballard C. C., Cowtan K. D., Dodson E. J., Emsley P., Evans P. R., Keegan R. M., Krissinel E. B., Leslie A. G. W., McCoy A., McNicholas S. J., Murshudov G. N., Pannu N. S., Potterton E. A., Powell H. R., Read R. J., Vagin A., Wilson K. S., Overview of the *CCP4* suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242 (2011).

81. McCoy A. J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C., Read R. J., *Phaser*crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007).

82. Adams P. D., Afonine P. V., Bunkóczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L.-W., Kapral G. J., Grosse-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C., Zwart P. H., *PHENIX*: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).

83. Emsley P., Lohkamp B., Scott W. G., Cowtan K., Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 (2010).

84. Chen V. B., Arendall W. B. III, Headd J. J., Keedy D. A., Immormino R. M., Kapral G. J., Murray L. W., Richardson J. S., Richardson D. C., *MolProbity*: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).

85. Hites R. A., Foran J. A., Schwager S. J., Knuth B. A., Hamilton M. C., Carpenter D. O., Global assessment of polybrominated diphenyl ethers in farmed and wild salmon. Environ. Sci. Technol. 38, 4945–4949 (2004).

86. Johansen H. R., Alexander J., Rossland O. J., Planting S., Løvik M., Gaarder P. I., Gdynia W., Bjerve K. S., Becher G., PCDDs, PCDFs, and PCBs in human blood in relation to consumption of crabs from a contaminated Fjord area in Norway. Environ. Health Perspect. 104, 756–764 (1996).

87. Bedi J. S., Gill J. P. S., Kaur P., Sharma A., Aulakh R. S., Evaluation of pesticide residues in human blood samples from Punjab (India). Vet. World 8, 66–71 (2015).

88. *Fourth National Report on Human Exposure to Environmental Chemicals* (Centers for Disease Control and Prevention, Atlanta, GA, 2009), 530 pp.

89. Mackay D., Mascarenhas R., Shiu W. Y., Aqueous solubility of polychlorinated biphenyls. Chemosphere 9, 257–264 (1980).

90. ATSDR, Toxicological profile for aldrin/dieldrin. *Agency Toxic Subst. Dis. Regist.*, 1–354 (2002).

91. ATSDR, Toxicological profile for endrin. *Agency Toxic Subst. Dis. Regist.*, 1–227 (1996).

92. ATSDR, Toxicological profile for DDT, DDE, and DDD. *Agency Toxic Subst. Dis. Regist.*, 1–497 (2002).

93. ATSDR, Toxicological profile for polychlorinated biphenyls (PCBs). *Agency Toxic Subst. Dis. Regist.*, 1–948 (2000).

94. ATSDR, Toxicological profile for polybrominated diphenyl ethers (PBDEs). *Agency Toxic Subst. Dis. Regist.*, 1–532 (2015).

CHAPTER 3

Red meat consumption promotes gut microbiota containing unconventional sialidases

Livia S. Zaramela[1#], Cameron Martino[1#], Frederico Alisson-Silva[2,3#], Steven Rees[4], Sandra L. Diaz[2,3], Léa Chuzel[5], Mehul B. Ganatra[5], Christopher H. Taron[5], Cristal Zuñiga[1], Jianbo Huang[4], Dionicio Siegel[4], Geoffrey Chang[4], Ajit Varki[2,3] and Karsten Zengler[1,6*]

[1]Department of Pediatrics, University of California, San Diego; [2]Department of Medicine and Cellular & Molecular Medicine, University of California, San Diego; [3]Glycobiology Research and Training Center (GRTC); [4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego; [5]New England Biolabs; [6]Center for Microbiome Innovation.

#These authors contributed equally to this work.
*Corresponding author.

In preparation

September 2018

PREFACE TO CHAPTER 3

Chapter 3, in full, is an article in preparation entitled "Red meat consumption promotes gut microbiota containing unconventional sialidases". This chapter resulted from collaborative work among the Chang, Zengler, Siegel, and Varki labs at UCSD. I am thankful to Cameron and Lívia for their metagenomics expertise, Dio and Jianbo (David) for synthesizing the modified inhibitor, and the Center for Microbiome Innovation Fellowship program, for funding me to do this work. This work is, to our knowledge, the first crystal structure of a Neu5Gc-preferring sialidase, and the first structure of any protein with the Neu5Gc-like inhibitor Neu5Gc2en.

ABSTRACT

High dietary consumption of red meat is associated with increased risk of inflammation-associated diseases in humans. The inflammation induced by metabolic incorporation of an antigenic red meat-derived glycan *N*-glycolylneuraminic acid ('xenosialitis') is a recently proposed mechanism to explain this association. We asked in commensal gut microorganisms play a role in the release of Neu5Gc from red meat glycans and therefore modifying Neu5Gc incorporation into human tissues. We identified novel Neu5Gc-preferring sialidases in samples from a human-like *Cmah-/-* mouse model and further confirmed sialidase activities, substrate preferences and performed X-ray crystallography for protein characteristics. Furthermore, we identified red meat-dependent enrichment of these novel sialidases in humans. Rapid release of bound Neu5Gc by bacteria expressing such sialidases could potentially protect humans from xenosialitis and help to reduce risk of inflammatory diseases associated with red meat consumption.

IMPACT STATEMENT

Here we present evidence of a Neu5Gc-preferring sialidase, including the first crystal structure, to our knowledge, of any protein with the Neu5Gc-like ligand Neu5Gc2en.

INTRODUCTION

The lifestyle and dietary habits of Western civilization have been extensively associated with alteration of the genetic composition and metabolic activity human gut resident microorganisms contributing to the high incidence of obesity, diabetes, and cancer (1). Red meat (*i.e.*, beef, pork, lamb, veal, and mutton) is one of the most appreciated and frequently eaten types of food in Western diets. Unequivocally, red meat is a rich food source for infants and childbearing women because it is highly enriched in nutrients such as proteins, vitamin B6 and B12, phosphorus, zinc, and other micronutrients. However, long-term consumption of red meat is associated with significant increase in all-cause mortality, likely contributing to the current epidemic of cardiovascular diseases and type 2 diabetes, and to increased risk of certain kinds of adenocarcinomas (2,3). Carcinogenic compounds generated by high temperature cooking or by processing methods, high saturated fat content, heme and heme-derived iron are the red meat components often associated to disease risk (4). However, none of the above are specific for red meat and their presence alone does not explain the human-specific propensity for increases in these diseases (12). Previous studies from our group have demonstrated that red meat is enriched in *N*-glycolylneuraminic acid (Neu5Gc), an isoform of sialic acid (Sia) that should be completely absent in humans (4). Humans cannot endogenously synthesize Neu5Gc, due to an evolutionary loss of a functional CMP-Neu5Ac hydroxylase (CMAH) that promotes the conversion of *N*-acetylneuraminic acid (Neu5Ac) to Neu5Gc. However, the consumption of red meat can cause Neu5Gc incorporation in cell surface glycans of human epithelial and endothelial cells (5), especially in human carcinomas (13,14). As a consequence, a chronic inflammatory response is triggered when Neu5Gc-containing glycans encounter circulating anti-Neu5Gc antibodies (3). Notably, for reasons that are currently unknown, incorporation is primarily caused by glycosidically-bound Neu5Gc. In contrast, free Neu5Gc would be utilized by gut microbes or cleared rapidly by the kidneys if it enters the circulation (Fig. 1A).

Several commensal and pathogenic intestinal bacteria can release host-derived sialic acids (Sias) from mucosal mucins and glycolipids by expressing a wide range of sialidases (6,7). Bacteria that lack *de novo* biosynthetic pathways for Sias can use their sialidase(s) to release it from the host glycans (8,15). Once free, Sias can be taken up through membrane-associated transporters and utilized as carbon, nitrogen or energy sources, or used to sialylate bacterial cell surface glycans (7). In addition, changes in

the intestinal concentration of Sias, for example induced by inflammation, can alter the expression of bacterial genes involved in Sia catabolism promoting intestinal dysbiosis (17).

RESULTS

Based on these important roles for Sias in bacteria-host interaction in the gut, we investigated if a Neu5Gc-rich diet, mimicking red meat intake, could provoke changes in bacterial metabolism in the gut and alter the gut microbiome. We compared the microbiota composition in fecal samples of human-like *cmah-/-* mice (18) fed a Sias-free (soy), a Neu5Gc-rich porcine submaxillary mucin (PSM), or a Neu5Ac-rich edible bird's nest (EBN) diet. The PSM and EBN diets contain 2.5% (wt/wt) of glycosidically-bound Neu5Gc and Neu5Ac, respectively (Table 3.S1). Fecal samples from mice of each diet group were collected by scraping the colon. The microbiota composition of *cmah-/-* mice and WT mice on soy, PSM, and EBN diets were determined through 16S rRNA gene amplicon sequencing. Analysis of similarities (ANOSIM) on Bray-Curtis dissimilarity showed significant difference between the genotypes of *cmah-/-* and WT (ANOSIM R = 0.979, pvalue < 0.01), indicating that the lack of endogenous Neu5Gc by itself shapes the microbiome (Figure 3.1A), substantiating the need for a human-like mouse model. We found that changes in microbial community composition within *cmah-/-* mice were diet dependent (ANOSIM R = 0.831, pvalue < 0.01), with Clostridiales and Bacteroidales contributing significantly to variations amongst diets (Figure 3.1B, 3.S1). Independent of genotype, the microbiome of PSM-fed mice was significantly less diverse compared to the microbiome of those fed soy and EBN diets (pvalue < 0.05, Wilcoxon) (Figure 3.1B, 3.S1). Human-like *cmah-/-* mice revealed similar taxonomic profiles at the family level amongst the three diet groups (Figure 3.1C). However, at the genus level, *Helicobacter*, *Intestinimonas*, and *Saccharibacteria incertae sedis* were significantly enriched in the EBN group compared to PSM. Moreover, *Bacteroides*, *Barnesiella*, *Clostridium* group III, *Parabacteroides*, *Roseburia,* and *Turicibacter* were significantly enriched in the PSM group compared to EBN (pvalue < 0.05, Wilcoxon) (Figure 3.1D, 3.S3). Bacteroidales were previous associated with high intake of animal-based food due their diverse enzymatic repertoire for utilizing carbohydrates (19). Additionally, by simulating 773 metabolic models from human gut microbiome, we found that Bacteroidetes was one of the most efficient microorganisms at metabolizing sialic acids (Figure 3.S4, Table 3.S2).

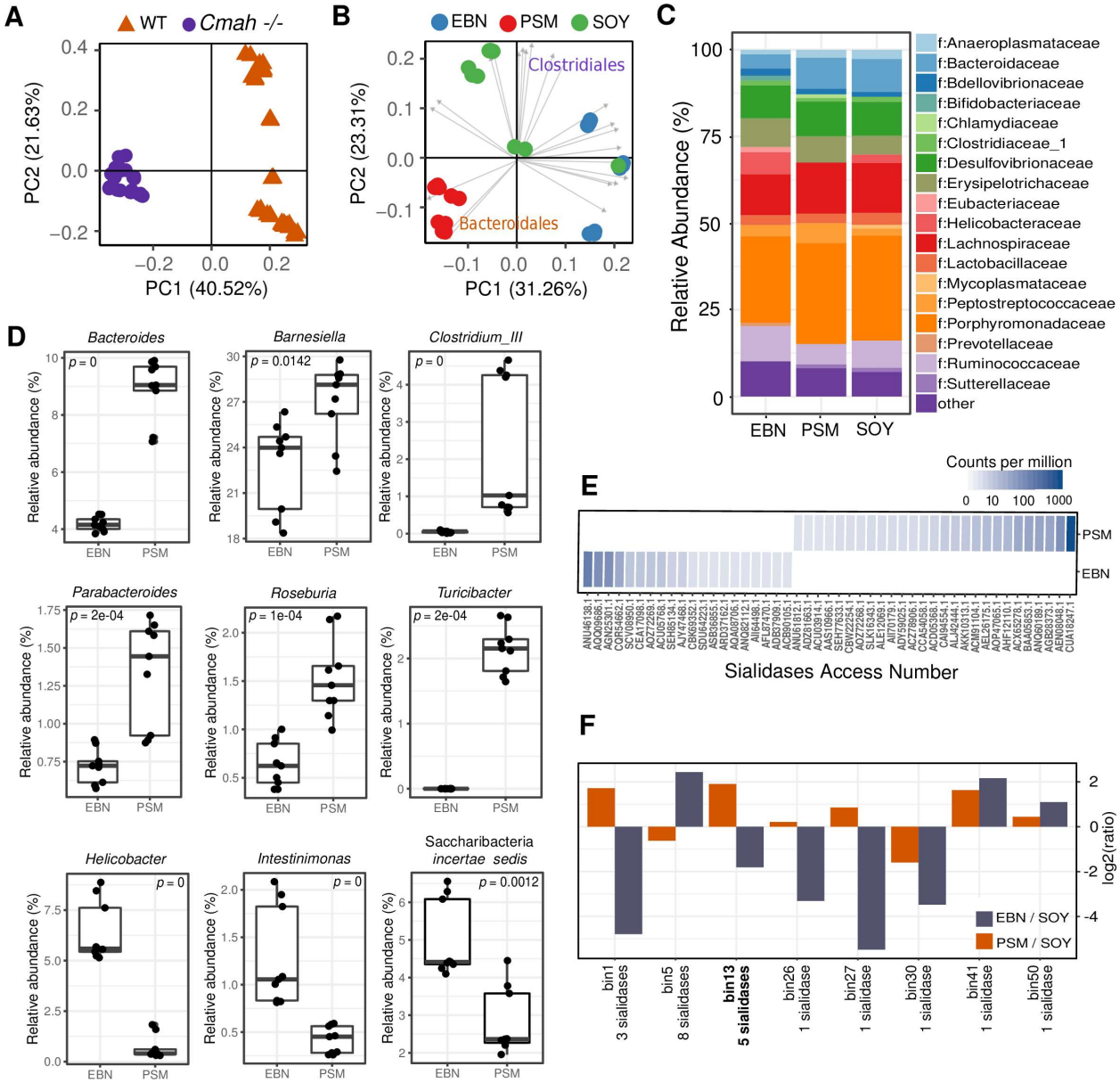


Figure 3.1. Gut microbial community composition on Sias-free (soy), Neu5Gc-rich (PSM), or a Neu5Ac-rich (EBN) diet. Composition of gut microbial community on Sia-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet. . **(A)** Beta-diversity analysis of WT vs *Cmah-/-* mouse. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R = 0.979, *pvalue* < 0.01). **(B)** Beta-diversity analysis of *Cmah-/-* mouse fed with Sia-free (soy), Neu5Gc-rich (PSM), or Neu5Ac-rich (EBN) diet. Pairwise Bray-Curtis dissimilarities were plotted against the first and second principal coordinates (ANOSIM R = 0.831, *pvalue* < 0.01). Significant vectors (R >= 0.7, *pvalue* =< 0.01) were obtained using *envfit* function from the R package vegan version 2.5-2 and the most representative taxa are indicated on the plot. (**C**) Taxonomic distribution at family level. (**D**) Differentially abundant bacterial genus on Neu5Gc-rich (PSM) or Neu5Ac-rich (EBN) diet. (**E**) Sialidases diet-dependent using CAZYme database. (**F**) Relative abundance of the bins with annotated sialidases. *Pvalues* were computed using the Wilcoxon rank sum test. The significance levels are indicated as follow: *pvalue* ≤ 0.05 (*), *pvalue* ≤ 0.01 (**), and *pvalue* ≤ 0.001 (***).

Shotgun metagenomic sequencing was performed to evaluate the functional capacity of the microbial community on soy, PSM, and EBN. To evaluate the enzymatic repertoire for utilizing carbohydrates, raw reads from soy, PSM, and EBN metagenomes were aligned to a carbohydrate active enzymes (CAZymes) database (11). A Principal Component Analysis (PCoA) of rarified read counts per CAZymes classes revealed a slightly significant clustering in gene function between diet types (ANOSIM R = 0.246, pvalue = 0.06), with a higher similarity between the microbial community on soy and EBN diets (Figure 3.S4). The mean of the relative abundance of all sialidases were similar between PSM and EBN; however, inspecting individual sialidase genes revealed several diet-dependent sialidases (Figure 3.1E).

To retrieve genes related to Sia metabolism, combined metagenomes were co-assembled and the resulting contigs were binned into 51 unique genomes (Table 3.S3). A total of 21 sialidase genes distributed in eight bins were identified (Table 3.S4, Figure 3.S6). While most bacterial genomes contain only one sialidase, we identified hundreds of examples of bacterial genomes with multiple sialidases in PATRIC (Figure 3.S6). Taking into account that all of our diets were mainly composed of soy, the relative abundance of the bins in PSM and EBN was normalized to soy. Amongst the bins with annotated sialidase genes, bin13, whose closest relative was *Bacteroides caccae* (Table 3.S3), was the highest abundant bin in PSM compared to EBN (Figure 3.1F). Bin13 contains five sialidases (sialidase23, sialidase24, sialidase26, sialidase60, and sialidase65). Sialidase26 exhibited high sequence conservation to sialidase CUA18247.1 from *Bacteroides fragilis*, identified as the most abundant sialidase in the PSM diet (Figure 3.1E, 3.S7).

The increased abundance of the bin 13 genome in a Neu5Gc-high diet (PSM) is a strong indicator of its role in Sia metabolism in the gut. To investigate preferential sialidase activity for Neu5Gc, a feature that has not been reported previously, all five sialidases from bin 13 were heterologously expressed *in vitro* and purified. Sialidase activity measurements were performed as previously reported using different enzyme concentrations (0.5 µg, 2.5 µg, 5 µg, and 10 µg). Four out of five sialidases showed preferential Neu5Gc activity in at least one of the pHs tested (Fig. 3.2A) indicating that bin13 is significantly enriched for exosialidases that prefer Neu5Gc over Neu5Ac as a substrate. Importantly, except for a sialyltransferase from the animal commensal bacteria *Pasteurella multocida*, no previously characterized exosialidases have been shown to prefer Neu5Gc over Neu5Ac. In fact, previous studies revealed that many bacterial species

such as *Arthrobacter ureafaciens*, *Clostridium perfringens*, *Streptococcus sp*., *Salmonella typhimurium, Streptococcus pneumonia, Ruminococcus gnavus* and *Vibrio cholera* express sialidases that hydrolyse Neu5Ac-terminated glycans more efficiently than Neu5Gc-terminated glycans, a substrate preference that is also shared by the murine sialidases Neu1, Neu2 and Neu4. These data points for the functionally distinct substrate preference of the bin13 sialidases are described herein.

The most compelling enzyme from the substrate specificity study was Sialidase 26, which showed significant preference for Neu5Gc from pH 6.5 to 8. Sialidase26 possessed protein sequence motifs characteristic of the GH33 family of bacterial sialidases, including a RIP/RLP motif (AAs 203-205), four Asp-boxes (AAs 240-247, 319-326, 379-387, and 486-493), and conserved nucleophilic tyrosine and acid/based glutamic acid residues at AAs 509 and 398, respectively (Figure. 3.S9). Furthermore, despite this novel substrate preference, sequence residues that are predicted to interact with terminal Sias in the catalytic site are highly conserved with structurally studied sialidases exhibiting no known Neu5Gc preference over Neu5Ac, including those from *B. caccae* (PDB 4Q6K), *B. thetaiotamicron* (PDB 4BBW), and *P. distasonis* (PDB 4FJ6). Other known sialidases with very low sequence identity (<30%) nevertheless shared 50-75% identity with residues predicted to interact with sialic acid, including NanA and NanB from *S. pneumoniae*, NanI from *C. perfringens*, an IT-sialidase from *R. gnavus*, and a trans-sialidase from *M. decora*, but suggest no clear sequence motif or mutation causing Neu5Gc preference.

To understand the structural underpinnings of Neu5Gc preference, we used X-ray crystallography to determine the structure of sialidase 26 both alone (PDB 6MRX, 2.0 Å resolution) and in complex with DANA-Ac (Neu5Ac2en) and DANA-Gc (Neu5Gc2en), known transition state mimics that closely resemble the structure of Sias, and are inhibitors of hydrolytic sialidases (PDB 6MRV - 1.8 Å resolution and PDB 6MYV - 2.2 Å resolution, respectively). Sialidase 26 possesses an overall structure and Sia engagement common to GH33 sialidases (Figure 3.2B), including Y509 nucleophilic engagement of C5 following E398 charge activation, D228 acid-base catalysis of the glycosidic bond at C5, and C2 stabilization by an Arg triad (R203, R414, R478). However, typical Sia-stabilizing interactions are lost, including Glu engagement of the glycerol moiety C7-C9 (T397 in sialidase26), C10 stability by an Arg residue (now a Leu), and inward movement of W507 into the binding pocket (though this is restored in the DANA-Gc co-crystal structure).

Co-crystallization of sialidase 26 with DANA-Gc indicated an overall fold and ligand placement similar to that of DANA-Ac (Figure 3.2C). The hydroxyl group at the end of the C5 acetamido group of DANA-Gc (a H atom in DANA) is pointed towards the binding pocket residues, forming H contacts with D271 and likely increasing its stability. This, in combination with the changes described above, likely explains sialidase26's preference for Neu5Gc.

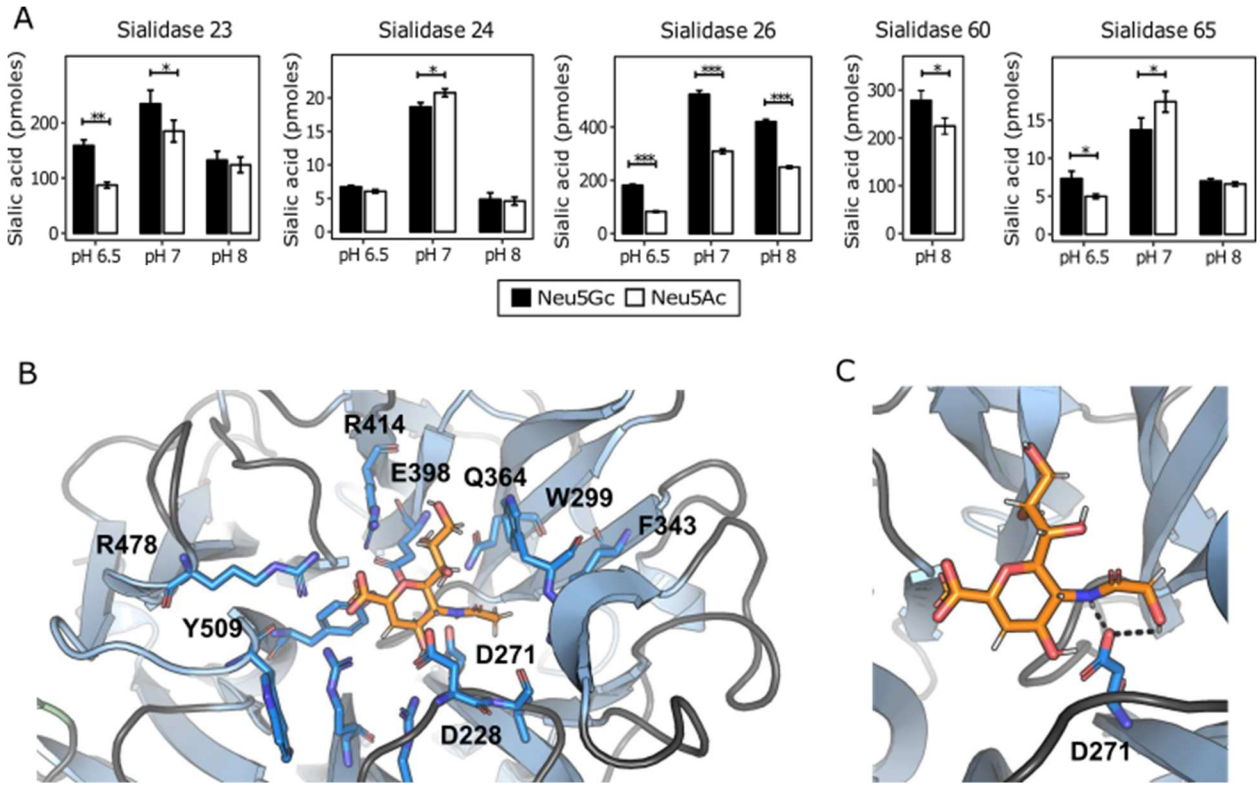


Figure 3.2. Sialidase Sia preference characterization. (**A**) Sialidase activity assay using 0.5 ug of each sialidase in three different pHs (6.5, 7.0, and 8.0). Sialidases were incubated for one hour with mouse serum containing either Neu5Ac or Neu5Gc-containing glycoproteins. Data are represented as triplicates and the statistical analysis were performed by Student's 2-tailed t test. (**B and C**) Crystal structures of sialidase26 co-crystallized with DANA-Ac **(B)** and DANA-Gc **(C)**. Interacting side chains are represented as sticks (blue), as well as the ligand (orange). The significance levels for (**A**) are indicated as follow: *pvalue* ≤ 0.05 (*), *pvalue* ≤ 0.01 (**), and *pvalue* ≤ 0.001 (***).

To extend our findings to human fecal samples, we reanalyzed 40 metagenomes from the Hadza. These hunter-gatherers live in a remote place in Tanzania and change their diet periodically according to food availability. In dry seasons, their diet is predominantly based on red meat, whereas in wet seasons the diet is based on honey and berries. Raw reads from Hadza metagenomes were mapped to bin13 and we observed that bin13 was significantly more abundant in microbiome samples taken during the dry season compared to wet season (pvalue = 0.015, Wilcoxon) (Figure 3.3A). To evaluate the sialidases genes present on Hadza microbiome, the combined metagenomes were co-assembled and the resulting genomes were annotated. In total, 52 sialidase genes were identified and these genes were distributed in 39 genome bins (Table 3.S6). The binHz19, whose closest relative was *Alistipes* sp., was equally abundant in microbiomes from the dry season compared to the wet season (Figure 3.3B). BinHz19 also contained the most similar sequence (sialidaseHz136) amongst all annotated sialidases to sialidase26 from our mouse study (Figure 3.S12), suggesting widespread distribution of this sialidase amongst mammals. The sialidaseHz136 was heterologously expressed *in vitro* and the activity assay revealed a 1.3 fold-change preferential activity to Neu5Gc compared to Neu5Ac (Figure 3.3C). We used X-ray crystallography to determine the structure of SialidaseHz136 (PDB 6MNJ, 2.2 Å resolution) (Figure 3.3E), and found similar predicted engagement of conserved residues with sialic acid substrate when compared to sialidase26, albeit a shift in the acetamido-interacting Asp residue. This result supports the notion that binHz19 sialidases are capable of metabolizing structurally diverse glycans, a concept that would be of dietary benefit to an individual in both wet and dry seasons. This result indicates that binHz19 are capable of metabolizing glycans from different sources, which is consistent with the similar relative abundance in both seasons.

Our findings indicate that sialidases exhibiting Neu5Gc preference are widespread in the mammalian intestine. To evaluate if Neu5Gc-specific sialidases are restricted to the intestine, we generated a fosmid library using DNA isolated from soil obtained from a commercial organic composting facility (Brick End Farms, Hamilton, MA). Sialidase activity of clones was differentially screened with fluorogenic substrates, i.e. 4MU-α-Neu5Ac and 4MU-α-Neu5Gc (Figure 3.S12 and 3.S13). Two clones designated C19 and C22 showed significant activity on 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac (Figure 3.3D). In addition, we tested the preferential activity of five commercial sialidases (Figure 3.3D). All

enzymes exhibited similar or minor activity to Neu5Gc compared to Neu5Ac (Figure 3.3D). Phylogenetic analysis of all seven Neu5Gc-preferential sialidases revealed differential clustering of these sialidases. Neu5Gc-preferential sialidases obtained from the gut microbiota of human and mouse differed from the ones found in soil, hinting at parallel evolution events (Figure 3.4).

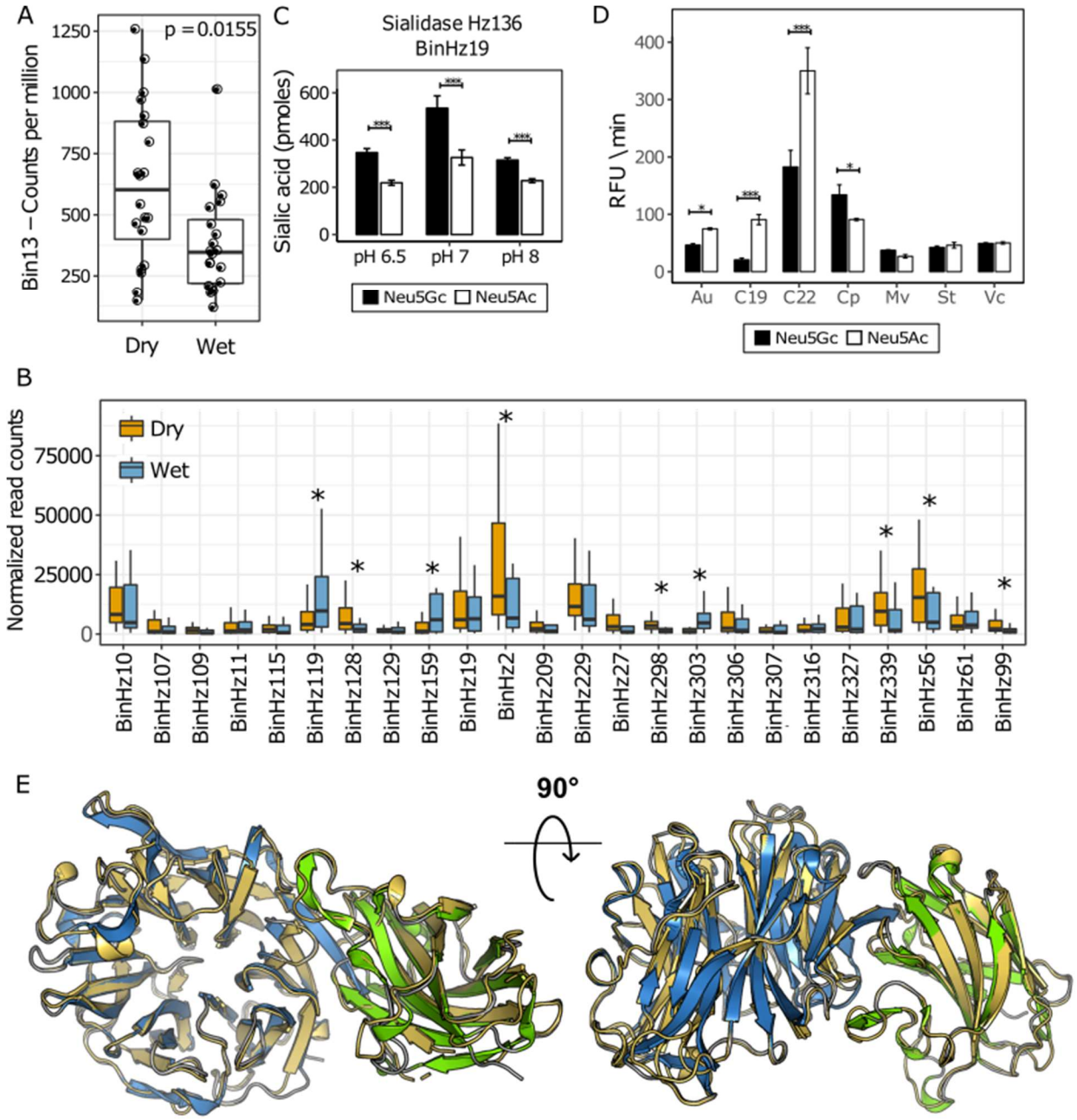


Figure 3.3. Analysis extended to other environments. (A) Relative abundance of bin13 and bins containing other sialidades genes **(B)** in samples from Hadza hunther-gatherers population. Bins were assembled from Hadza shotgun metagenome (28839072). Statictical significance were computed using the non-parametric Wilcoxon rank sum test. **(C)** Screen for the activity of sialidase Hz136 retrieved from Hadza shotgun metagenome in three different pHs (6.5, 7.0, and 8.0) and from composting sialidases **(D)** both using 0.5µg of each enzyme. **(E)** Crystal structure of sialidaseHz136 (yellow) aligned to the catalytic site (blue) and carbohydrate binding motif (green) of sialidase26. The experiments were performed in triplicates and the statistical significance determined by Student's 2-tailed t test. The significance levels are indicated as follow: *pvalue* ≤ 0.05 (*), *pvalue* ≤ 0.01 (**), and *pvalue* ≤ 0.001 (***).

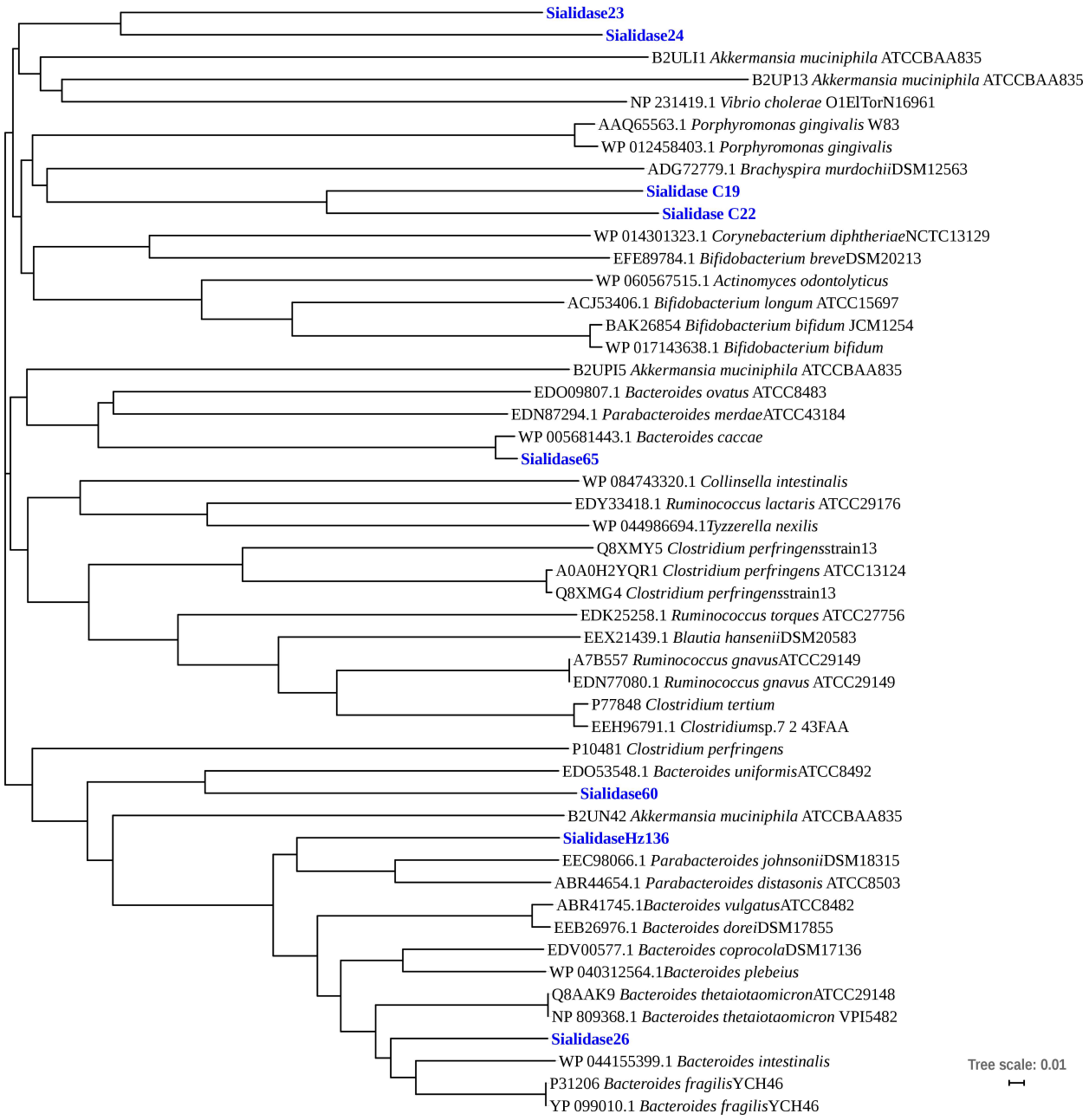


Figure 3.4. Phylogenetic relationship of sialidases encoded by gut bacteria. Branch lengths are in the same units as those of the evolutionary distances used to construct the tree. The organism and loci are listed where available. Sialidases identified in this study are highlighted in blue.

CONCLUSION

In summary, we discovered several new sialidases with preference for Neu5Gc that are enriched in the gut microbiota of mice and humans consuming a Neu5Gc-rich diet. Rapid cleavage of Neu5Gc prevents incorporation of this non-human sugar into tissues because unbound Neu5Gc is not to contributing to xenosialitis. We thus hypothesize that lack of bacteria with Neu5Gc-preferential sialidase could result in increased xenosialitis and be one contributing factor to colon inflammation and colorectal cancer. Our results lay the foundation to define a strategy for translation of pre- or probiotics to treat red meat-related colon inflammation.

EXPERIMENTAL PROCEDURES

In Vivo Sampling

Human-like *Cmah*$^{-/-}$ mice generated as previously described (25) were bred onto a congenic C57BL/6 background and maintained in the University of California, San Diego vivarium according to Institutional Animal Care and Use Committee (IACUC) guidelines, with 12-h diurnal lighting and access to food and water ad libitum. All mice used in the study were maintained in sialic acid free soy based diet (Dyets, Inc.; 110951) from weaning until 8 to 10 weeks of age to prevent previous exposure to sialic acid. To evaluate the effect of the dietary Neu5Gc in the gut microbiome, sex matched *Cmah-/-* mice with 10 weeks of age were caged in three groups of 5 mice each fed during the four weeks with the same soy based diet either enriched in Neu5Gc (PSM) (26), Neu5Ac, or only soy as control. The animals were euthanized in CO2 and the colon tissues were cut open with blunted scissors for fresh collection of fecal samples by scraping it straight from the tissue. Colonic fecal scraping from five mice of each diet type were used for 16S rRNA gene amplicon and metagenomic shotgun sequencing analysis.

16S rRNA sequencing and analysis.

MoBio PowerFecal DNA isolation kit (MoBio, Carlsbad, CA, USA) was used per manufacturer's instructions to extract genomic DNA in all samples collected. Purified DNA was amplified and processed according to the Illumina 16S protocol (17). Libraries were quality assessed using quantitative PCR (qPCR) and a Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and subsequently sequenced using two MiSeq 600 cycle kits (Illumina). All data was trimmed of adapter sequences and quality controlled using Trim Galore (18). 16S analysis was performed with search denoising and the RDP 16S rRNA database (19). All downstream analysis was performed using scikit bio.

Metagenome sequencing and analysis.

MoBio PowerFecal DNA isolation kit (MoBio, Carlsbad, CA, USA) was used per manufacturer's instructions to extract genomic DNA in all samples collected. Purified DNA was prepared for shotgun metagenomic sequencing using the Nextera XT library preparation method (Illumina, San Diego, CA, USA). Libraries were quality assessed using quantitative PCR (qPCR) and a Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and subsequently sequenced using two MiSeq 500 cycle kits (Illumina). All data was

trimmed of adapter sequences and quality controlled using Trim Galore. Samples were filtered of possible human and mouse contamination through Knead Data. Functional annotation of unassembled reads were analyzed using the python 2.7 version of MetaPhlAn2 (20). Reads were assembled using Spades (21) with the metagenomics flag and Binning was performed using MetaBat (22). Each bin was assessed for functional profiles separately using RAST (23). Taxonomic classification was assessed using phylophlan (24). Relative bin sample abundance was determined by aligning binned genes to the general assembly with Bowtie2 (25). Differentially abundant raw read functional genes between sampling groups were identified using random forests feature selection. Identified genes were then filtered for biological significance and correlated to functional profiles of assembled bins. Bins with relevant genes were compared for relative abundance between sample groups.

Protein Expression and Assay

Target sialidase sequences from metagenomic bin 45 were PCR amplified from genomic DNA isolated as described above, subcloned into a pET19b expression vector with a C-terminal 10xHis tag and N-terminal truncation to remove the signal peptide sequence (predicted by SignalP 4.1, CBS) (26), and transformed into BL21(DE3) *E. coli* (MilliporeSigma) using established heat-shock methods. Cells were grown to OD 0.6-0.8 in multiple 1 L cultures at 37℃, and induced overnight at 25℃ with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Harvested cells were resuspended in lysis buffer (50 mM HEPES pH 8.0, 50 mM NaCl, and 1 mM TCEP) with DNaseI and hen egg white lysozyme, lysed with a TS-Series cell disruptor (Constant Systems, Inc.) at 15 KPSI, and spun for 45 minutes at 186,000xg with a Ti45 ultracentrifugation rotor (Beckman Coulter, Inc.) to remove cell debris. Purification was performed as below and based on purification of a putative *Bacteroides* neuraminidase as provided by the Protein Structure Initiative (BACCAC_01090, Joint Center for Structural Genomics, to be published), with modifications to imidazole stringency based on the sialidase purified. Supernatant was loaded on a 5-mL HisTrap Ni affinity column on an Akta Explorer purification system (GE Healthcare Life Sciences) with 20-40 mM imidazole added, washed with Running Buffer (50 mM HEPES pH 8.0, 300 mM NaCl, 40-60 mM imidazole, 10% glycerol, and 1 mM TCEP), and eluted with Elution Buffer (20 mM HEPES pH 8.0, 300 mM imidazole, 10% glycerol, and 1 mM TCEP). Samples were concentrated using 10-30 kDa Amicon centrifugal filters

(MilliporeSigma) at 1500xg to 1 mL, and desalted over a 5-mL Desalting column using the Akta system into Desalting Buffer (20 mM HEPES pH 8.0, 200 mM NaCl). Resulting protein sample was diluted as needed for functional studies.

Assay for sialidase activity *ex vivo*

The five sialidases purified as described above were quantified using SDS-PAGE image analysis with BSA references (Bio-Rad) and absorbance using the Nanophotometer P330 (Implen), with extinction coefficients calculated using ExPASy Translate (https://web.expasy.org/translate/). Sialidase activity was performed in a series of dilution. 0.5µg, 2.5µg, 5µg and 10µg of each enzyme were incubated with equal amount of human-like Cmah$^{-/-}$ mouse serum and wild-type mouse serum for one hour at 37°C. Additional 10 ug of each enzyme was inactivated by heat for 5 minutes at 95°C. The samples were kept at -20°C until derivatization and analyzed by HPLC as described below. The DMB reagent was made with the following recipe (27): 14 mM DMB (Sigma D4787), 18 mM sodium hydrosulfite (Sigma 157953), 1.0 M 2-mercaptoethanol (Sigma M3148), and 40 mM trifluoroacetic acid (Sigma T6508), and it was incubated at 50 °C for 2.5 h. The DMB-derivatized samples were analyzed on a Dionex Ultra3000 HPLC System using a Phenomenex Gemini 5µ C18 250 × 4.6-mm HPLC column at room temperature, eluted in isocratic mode with 85% water, 7% methanol, 8% acetonitrile. The fluorescence was detected using emission at 448 nm and excitation at 373 nm.

Neu5Gc2en Synthesis

Neu5Gc2en (DANA-Gc) was synthesized as previously published, with some modifications. Briefly, Neu5Gc (Sigma-Aldrich) was treated with Dowex 5OW-X8 (H+) resin in MeOH for 20 h at 20℃ to methylate C1's hydroxyl group, then treated with acetic anhydride and pyridine for 42 h at 20℃ to generate the paracetylated methyl ester, purified by column chromatography (5O:l CHC13- MeOH). This sample was treated with TMSOTf under nitrogen at 0℃ with MeCN for 6 h to eliminate the acetylated hydroxyl group on C2, purified by chromatography (toluene/acetone, 3:1 → 2:1 gradient following pyridine and toluene addition to the sample). The sample was deprotected overnight with NaOH/MeOH pH 11, and neutralized with H+ resin.

Crystallization and Structure Determination of Sialidase26 and SialidaseHz136

Purified sialidase26 was concentrated to 8 g/L and set on sitting drop trays in 1:1 volume ratios with mother liquor (20% PEG 6000, 0.1 M Tris-HCl pH 8.0) at 16℃. For ligand co-crystals, concentrated sialidase26 was incubated for one hour at room temperature with 5 mM N-Acetyl-2,3-dehydro-2-deoxyneuraminic acid (Neu5Ac2en or DANA, Sigma-Aldrich) or DANA-Gc (synthesized). Crystals appeared after 3-4 days and grew to full size in 8-10 days. Crystals were soaked briefly in mother liquor supplemented with 10% glycerol and flash-frozen with liquid nitrogen. Purified sialidaseHz136 was crystallized similarly (20% PEG 3350, 0.1 M Bis-Tris-Propane pH 7.5, 0.2 M sodium citrate); additional crystals in 20% PEG 3350 and 0.2 M sodium sulfate were soaked for 14-21 days in 8 mM DANA-Gc for the co-crystal structure. X-ray diffraction data were collected at 100 K at the Lawrence Berkeley National Laboratory Advanced Light Source (8.2.1 and 8.2.2) at a single wavelength. Preliminary diffraction data were collected at the Stanford Synchrotron Radiation Lightsource and Advanced Photon Source. All diffraction data were indexed and integrated with XDS or MOSFLM, processed with AIMLESS, and truncated with CTRUNCATE within the CCP4 suite of programs. Phases were estimated via molecular replacement in PHENIX.PHASER, using a previously published model of an uncharacterized Bacteroides-derived sialidase with high sequence homology to sialidase26 (PDB 4q6k) as a search model. Models underwent rigid-body and restrained positional refinement using PHENIX.REFINE in the PHENIX software suite against a maximum likelihood target function, alternated with manual inspection against electron density maps in Coot. Geometry restraints for DANA were generated using PHENIX.eLBOW, with manual inspections in Coot and refined in the final rounds of refinement, which also included the application of hydrogens to their riding positions and simulated annealing. The resulting refinement statistics for each model are included in table S7. Figures displaying crystal packing were prepared using PyMOL (http://www.pymol.org), and atomic coordinates and structure factors were desposited with the Protein Data Bank (accession codes 6MNJ, 6MRV, 6MRX, and 6MYV).

Analysis of previously published shotgun metagenomic data

Previously published stool shotgun metagenomic data from Hadza hunter-gatherer individuals was obtained from Sequence Read Archive (SRA) repository under the project IDs PRJNA392012,

PRJNA392180. In total were analysed 40 shotgun metagenomic data, in which 20 samples were collected during wet season and 20 during dry season. All data was trimmed of adapter sequences and quality controlled using Trimmomatic v0.36. Samples were filtered of possible human contamination using bowtie2 v2-2.2. with the following parameters (-D 20 -R 3 -N 1 -L 20 --very-sensitive-local). Relative abundance was obtained by aligning the metagenomic trimmed reads to the assembled bins. Statistical differences in abundance between wet and dry seasons were calculated using the nonparametric two-sided Wilcoxon rank sum test with Holm correction for multiple hypotheses.

Enzyme Discovery - Metagenomic DNA library construction

Environmental DNA was isolated from soil obtained from a commercial organic composting facility in Hamilton, MA (Brick End Farms) by phenol:chloroform extraction and isopropanol precipitation. A fosmid library was produced using the CopyControl™ Fosmid Library Production Kits (Lucigen Corporation, Middleton, WI) as recommended. Briefly, DNA was end-repaired and size-selected using a 1% Low Melting Point agarose gel run overnight at 35 V. DNA fragments from 30-70 kb were isolated from the gel using 1 U of β-agarase I (New England Biolabs, Ipswich, MA) for each 100 µL of melted agarose. The end-repaired and size-selected DNA was ligated to the pCC1 FOS cloning vector. Resulting clones were packaged in phage particles. Escherichia coli EPI300 T1R cells were transfected with the packaging reaction and plated on LB agar medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g MgCl2•6H2O, chloramphenicol 12.5 µg/mL, 2 mL 2M NaOH and 20 g of agar per liter,) and incubated overnight at 37°C. A total of 5376 colonies were archived in fourteen 384-well plates in sterile 20% (v/v) glycerol.

Enzyme Discovery - Screening for sialidase activity

The compost metagenomic library was differentially screened with fluorogenic 2′-(4-methylumbelliferyl)-α-D-N-glycolylneuraminic acid (4MU-α-Neu5Gc) (Sussex Research, Ottawa, CA) and 2′-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid 4MU-α-Neu5Ac (Toronto Research Chemicals, Toronto, CA) substrates. In a primary screen, library clones were grown in 384-well plates containing 50 µL LB liquid cultures (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g MgCl2-6H2O, 2 mL of 2M NaOH per liter, containing chloramphenicol 12.5 µg/mL and 1X inducing solution (Lucigen Corporation))

overnight at 37°C. Fifty microliters of Y-per lysis buffer (Thermo Fischer Scientific, Waltham, MA) containing 40 μg/mL of 4MU-α-Neu5Gc was added to each well. The mixtures were incubated overnight at 37°C in a static incubator. Fluorescence at λex = 365 nm and λem = 445 nm was read with a SpectraMax Plus 384 Microplate Reader (Molecular Devices, Sunnyvale, CA) at 6 h, 24 h and 48 h. Positive clones were defined as those showing fluorescence greater than 3 standard deviations above the mean background, and each was re-archived in a fresh 384-well plate in sterile 20% (v/v) glycerol. Each of the positive clones was grown and comparatively re-screened in separate assays containing 4MU-α-Neu5Ac and 4MU-α-Neu5Gc substrates (reactions run in duplicate). Two clones designated C19 and C22 showed significant activity on 4MU-α-Neu5Gc but only minor activity on 4MU-α-Neu5Ac and were subjected to further study.

Phylogenetic analysis

Multiple sequence alignment was performed with MUSCLE and manually curated. The phylogeny of bacterial sialidase enzyme sequences was inferred by Maximum Likelihood using the JTTmatrix-based model. 500 replicates were bootstrapped to assess statistical confidence in inferred phylogenetic relationships. Distances reflect the number of amino acid substitutions per position. All gaps and incomplete positions were excluded from the final analysis that yielded 274 aligned positions.

Reconstructed metabolic models analysis

Microorganisms part of the gut microbiome containing the sialidase activity were identified by scanning the repository of gut microbiome metabolic models. Chosen genome-scale metabolic models were simulated using the Gurobi Optimizer Version 5.6.3 (Gurobi Optimization Inc., Houston, Texas) solver in MATLAB (The MathWorks Inc., Natick, MA) with the COBRA Toolbox. Growth rates were simulated using flux balance analysis (FBA). All metabolic models were constrained using a western diet (45% fat, 35% carbohydrate, 20% protein). Additionally, the contribution to growth of the metabolites associated with the sialidase activity (e.g. N-acetylneuraminate) was determined using shadow prices simulations.

DISSERTATION ACKNOWLEDGEMENTS

Chapter 3, in full, is a reprint of material in preparation for submission. Livia S. Zaramela, Cameron Martino, Frederico Alisson-Silva, Steven Rees, Sandra L. Diaz, Léa Chuzel, Mehul B. Ganatra, Christopher H. Taron, Cristal Zuñiga, Jianbo Huang, Dionicio Siegel, Geoffrey Chang, Ajit Varki and Karsten Zengler. Red meat consumption promotes gut microbiota containing unconventional sialidases. This dissertation author is a co-author of this paper.

REFERENCES

1. Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18,** 117–143 (1993).
2. Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69,** 57–68 (1987).
3. Severi, E., Hood, D. W. & Thomas, G. H. Sialic acid utilization by bacterial pathogens. *Microbiology* **153,** 2817–2822 (2007).
4. Almagro-Moreno, S. & Boyd, E. F. Insights into the evolution of sialic acid catabolism among bacteria. *BMC Evol. Biol.* **9,** 118 (2009).
5. Steenbergen, S. M., Lichtensteiger, C. A., Caughlan, R., Garfinkle, J., Fuller, T. E., and E. R. Vimr. Sialic Acid metabolism and systemic pasteurellosis. *Infect. Immun.* **73,** 1284–1294 (2005).
6. Annunziato, P. W., Wright, L. F., Vann, W. F. & Silver, R. P. Nucleotide sequence and genetic analysis of the neuD and neuB genes in region 2 of the polysialic acid gene cluster of Escherichia coli K1. *J. Bacteriol.* **177,** 312–319 (1995).
7. Hirmo, S., Kelm, S., Wadström, T. & Schauer, R. Lack of evidence for sialidase activity in Helicobacter pylori. *FEMS Immunol. Med. Microbiol.* **17,** 67–72 (1997).
8. Vimr, E. R., Lawrisuk, L., Galen, J. & Kaper, J. B. Cloning and expression of the Vibrio cholerae neuraminidase gene nanH in Escherichia coli. *J. Bacteriol.* **170,** 1495–1504 (1988).
9. Sakamoto, M. & Benno, Y. Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. *Int. J. Syst. Evol. Microbiol.* **56,** 1599–1605 (2006).
10. Park, K. H., Kim, M. G., Ahn, H. J., Lee, D. H., Kim, J. H., Kim, Y. W., and E. J. Woo. Structural and biochemical characterization of the broad substrate specificity of Bacteroides thetaiotaomicron commensal sialidase. *Biochim. Biophys. Acta* **1834,** 1510–1519 (2013).
11. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: a mini-review. *Biochem. Soc. Trans.* **44,** 166–175 (2016).
12. Daniel, C. R., Cross, A. J., Koebnick, C. & Sinha, R. Trends in meat consumption in the USA. *Public Health Nutr.* **14,** 575–583 (2011).
13. Aykan, N. F. Red Meat and Colorectal Cancer. *Oncol. Rev.* **9,** 288 (2015).
14. Singh, R. K., Chang, H. W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., Abrouk, M., Farahnik, B., Nakamura, M., Zhu, T. H., Bhutani, T., and W. Liao. Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15,** 73 (2017).
15. Bernstein, A. M., Song, M., Zhang, X., Pan, A., Wang, M., Fuchs, C. S., Le, N., Chang, A. T., Willett, W. C., Ogino, S., Giovannucci, E. L., and K. Wu. Processed and Unprocessed Red Meat and Risk of Colorectal Cancer: Analysis by Tumor Location and Modification by Time. *PLoS One* **10,** e0135959 (2015).
16. Santarelli, R. L., Pierre, F. & Corpet, D. E. Processed meat and colorectal cancer: a review of epidemiologic and experimental evidence. *Nutr. Cancer* **60,** 131–144 (2008).
17. Illumina 16S Metagenomics Sequencing Protocol - support.illumina.com.
18. Krueger, F. Trim Galore. *A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files* (2015).
19. Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., and J. M. Tiedje. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42,** D633–42 (2014).

20. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and N. Segata. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12,** 902–903 (2015).

21. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and P. A. Pevzner. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

22. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3,** e1165 (2015).

23. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olson, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and O. Zagnitko. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9,** 75 (2008).

24. Segata, N., Börnigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* **4,** 2304 (2013).

25. Langdon, W. B. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min.* **8,** 1 (2015).

26. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786 (2011).

27. Inoue, S. & Inoue, Y. Ultrasensitive Analysis of Sialic Acids and Oligo/ Polysialic Acids by Fluorometric High-Performance Liquid Chromatography. in *Methods in enzymology* **362,** 543–560 (Elsevier, 2003).

28. Hedlund, M., Tangvoranuntakul, P., Takematsu, H., Long, J. M., Housley, G. D., Kozutsumi, Y., Suzuki, A., Wynshaw-Boris, A., Ryan, A. F., Gallo, R. L., Varki, N., and A. Varki. (2007) N-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution. Mol Cell Biol 27(12):4340–4346.

29. Banda, K., Gregg, C. J., Chow, R., Varki, N. M., and A. Varki. (2012) Metabolism of vertebrate amino sugars with N-glycolyl groups: mechanisms underlying gastrointestinal incorporation of the non-human sialic acid xeno-autoantigen N-glycolylneuraminic acid. Aug 17;287(34):28852-64.

CHAPTER 4

Structural basis for GPCR-independent activation of heterotrimeric Gi proteins

Nicholas A. Kalogriopoulos[a,f], Steven D. Rees[b,f], Tony Ngo[b], Noah Kopcho[b, c], Andrey Ilatovskiy[b], Nina Sun[a], Elizabeth Komives[c], Geoffrey Chang[b], Pradipta Ghosh[a, d, e] and Irina Kufareva[b]

[a]Department of Medicine, University of California, San Diego, La Jolla, CA, USA.
[b]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA.
[c]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA.
[d]Moores Cancer Center, University of California, San Diego, La Jolla, California, USA.
[e]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.
[f]These authors contributed equally: Nicholas A. Kalogriopoulos, Steven D. Rees.

In preparation

September 2018

PREFACE TO CHAPTER 4

Chapter 4, in full, is an article in preparation entitled "Structural basis for GPCR-independent activation of heterotrimeric Gi proteins". This chapter resulted from collaborative work between the Chang and Ghosh labs at UCSD, and was spearheaded by myself and fellow BMS graduate student Nicholas Kalogriopoulos. I am thankful to Nick for his expertise in the world of GPCRs, and Dr. Irina Kufareva on advice related to the structural models. This work is, to our knowledge, the first co-crystal of a G protein subunit with a natural GEM, and offers much-needed insight in how relevant GEMs engage G protein systems in non-canonical ways.

ABSTRACT

Heterotrimeric G proteins are key molecular switches that control cell behavior. The structural basis for canonical heterotrimeric G protein activation by agonist-occupied G protein-coupled receptors (GPCRs) has been resolved at an atomic level. In contrast, the structural basis for GPCR-independent G protein activation via a novel family of Guanine-nucleotide Exchange Modulators (GEMs), remains unknown. Here, we present a 2.1 Å crystal structure of Gαi in complex with the GEM motif of GIV/Girdin. Nucleotide exchange assays, molecular dynamics simulations, and hydrogen-deuterium exchange experiments demonstrate that GEM binding causes conformational changes in Sw-II that are allosterically propagated through Sw-I and the β2-β3 strands to the hydrophobic core of the Gαi GTPase domain. Perturbation of the hydrophobic core appears to be a common mechanism by which GPCRs and GEMs activate G proteins. Atomic-level insights presented here will aid structure-based efforts to selectively target the non-canonical G protein activation pathway.

IMPACT STATEMENT

Here we present the first structure of a G protein subunit with a natural GEM, offering much needed insight into non-canonical means of G protein signaling.

INTRODUCTION

Heterotrimeric G proteins act as molecular switches that gate the flow of information from extracellular cues to intracellular effectors that control cell behavior (1, 2). Canonically, heterotrimeric G protein signaling is initiated exclusively at the plasma membrane where agonist-activated G-Protein Coupled Receptors (GPCRs) activate Gαβγ trimers by promoting the exchange of GDP for GTP on the Gα subunit (1). Heterotrimeric G proteins are expressed in virtually all cell and tissue types in the body and are involved in most physiologic, and many pathologic, processes; thus, the molecular mechanism and structural determinants of G protein activation and action has been a top priority in the field, yielding over 70 publicly available structures in various conformations and complex composition (Table 4.S1). The structural basis for GPCR-dependent G protein activation had challenged the field for decades, but in the past 8 years, breakthrough structural studies have demonstrated that GPCR-dependent G protein activation occurs via direct disruption the hydrophobic core of the Gα GTPase domain by removal of the C-terminal α5 helix and insertion of the GPCR's intracellular loop 2 (3-6).

In contrast, GPCR-independent activation of heterotrimeric G proteins can occur downstream of non-GPCRs via a novel family of Guanine-nucleotide Exchange Modulators (GEMs) (7-8). Members of the GEM family are cytosolic proteins that are defined by their unique ability to act as Guanine-nucleotide Exchange Factors (GEFs) to activate Gαi and to act as Guanine-nucleotide Dissociation Inhibitors (GDIs) to inhibit Gαs using the same evolutionarily conserved GEM motif (9-10), identified based on its homology to the synthetic peptide KB752 (a peptide that can bind and activate Gαi) (11). GEMs activate heterotrimeric G proteins downstream diverse classes of receptors, resulting in spatio-temporal signaling that is distinct from GPCR-driven G protein signaling. Furthermore, GEM-dependent G protein activation has been shown to play a pivotal role in pathophysiology (12), including cancer, organ fibrosis, and diabetes. A structural understanding of the mechanism of GEM-dependent G protein activation would be invaluable towards pharmacologic targeting of GEMs to treat these diseases. Despite vast progress in understanding GEM biology and demonstrable translational relevance of dysregulated GEM signaling in disease, structural mechanistic insights into GEM-dependent G protein activation is lacking. This study fills this void, providing high resolution insight into GPCR-independent, GEM-driven G protein activation.

RESULTS

Unique from GPCRs, GIV-GEM binds and stabilizes Switch II of Gαi

To understand structurally the mechanism of GEM-dependent heterotrimeric G protein activation, the 31-aa GEM motif of GIV/Girdin (aa 1671-KTGSPGSEVVTLQQFLEESNKLTSVQIKSSS-1701) was co-crystallized with GDP-bound rat Gαi3 (henceforth Gαi•GDP) (Table 4.S2). In the crystallization construct, the flexible 25-aa long N-terminal helix of Gαi was deleted as done previously (11), and replaced by a His-tag followed by a short linker (SSGLVPRGSHM; Figure 4.S1A). The structure was solved to 2.1 Å resolution Figure 4.1A-B, Figure 4.S1).

The structure demonstrated that GIV-GEM binds at the typical effector binding interface: the hydrophobic pocket between Sw-II and the α3-helix of Gαi (Figure 4.S1A-B). By forming a short anti-parallel β-sheet with Sw-II residues 204-208, the peptide stabilizes Sw-II in a previously unseen elevated conformation. Key polar contacts at the interface include hydrogen-bonding of GIV E1678 and E1688 to Gαi R208, around which the peptide folds in a loop-helix conformation, and a hydrogen bond from GIV Q1683 with Gαi Q204, a residue known for its role in GTP-hydrolysis (13). The interface also features hydrophobic packing of GIV's F1685 against W211, I212, F215, and W258 of Gαi. Residues L1682-N1690 of GIV form an α-helix that packs favorably across the α3-helix of Gαi (Figure 4.1B-C). Many, but not all, Gαi residues engaged by GIV-GEM are shared by Gβγ and GoLoco GDIs (Figure 4.1C-F, 4.2A), explaining the ability of GIV to dissociate Gαi from both (9, 14).

The basis for the previously described phosphoregulation of GEM activity of GIV (15-16) is evident from the structure and molecular modeling. A phosphate on the N-terminal S1674 of GIV-GEM is predicted to improve binding by creating an additional polar contact with Gαi R208 (Figure 4.2B-C). By contrast, a phosphate on the C-terminal S1689 of GIV-GEM would disrupt a key hydrogen bond that this residue forms with W258 of Gαi (Figure 4.2D). These findings explain the opposing roles of the two phosphoevents: the former enhances and the latter abrogates the ability of GIV to bind and activate Gαi (15-16).

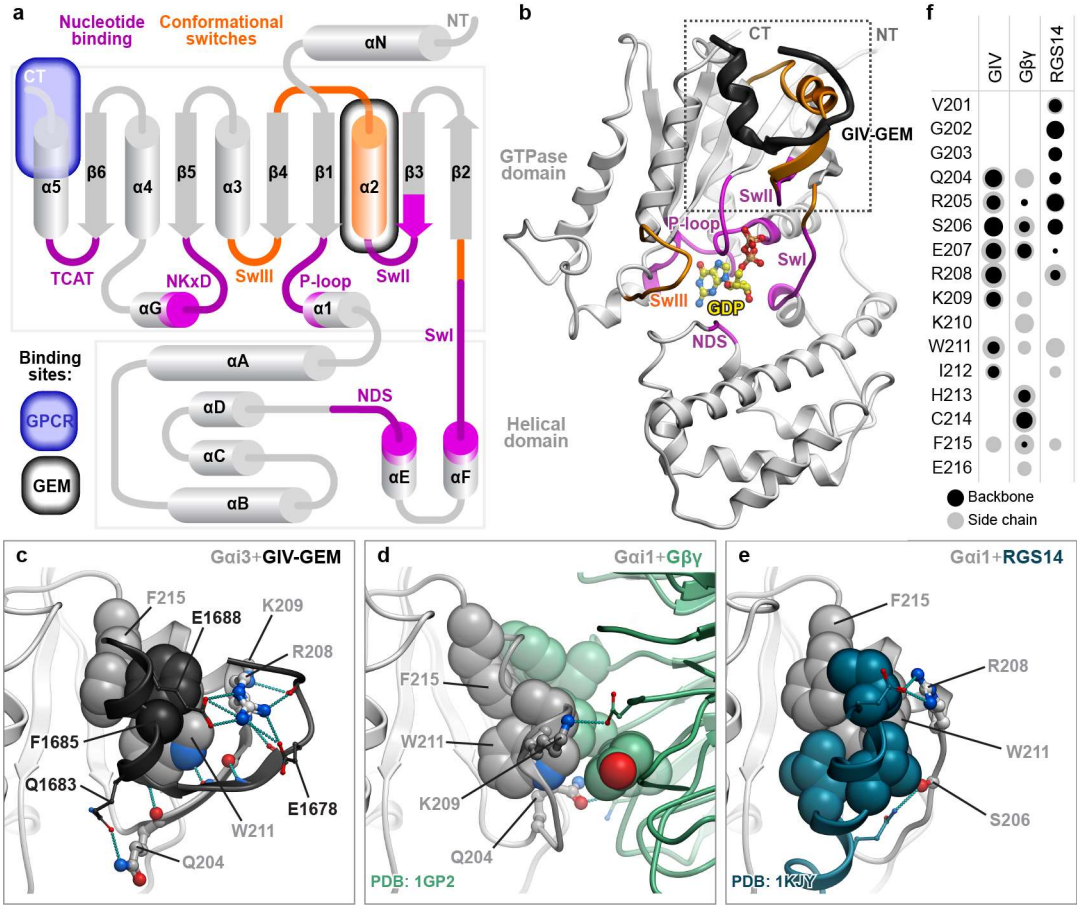


Figure 4.1: GIV-GEM binds Sw-II of Gαi. A) Topology of the Gαi protein with conformational switches and binding sites of key interactors marked. B) Crystal structure of Gαi with GIV-GEM peptide bound at Switch (Sw)-II. C) A close-up view of the interaction interface between Gαi and GIV-GEM. D-E) Close-up views of Gαi Sw-II bound to Gβγ (D, PDB 1GP2) or GoLoco-motif GDI RGS14 (E, PDB 1KJY). Key Sw-II residues shared by GIV and at least one of Gβγ or RGS14 are shown as spheres (aromatic/aliphatic) or sticks (polar). F) Bubble plot displaying the strength and the nature of contacts that Gαi Sw-II residues make with GIV-GEM, Gβγ or RGS14. The size of the dot is proportional to the strength of the contact (36); backbone and side-chains contacts are shown in black and grey, respectively.

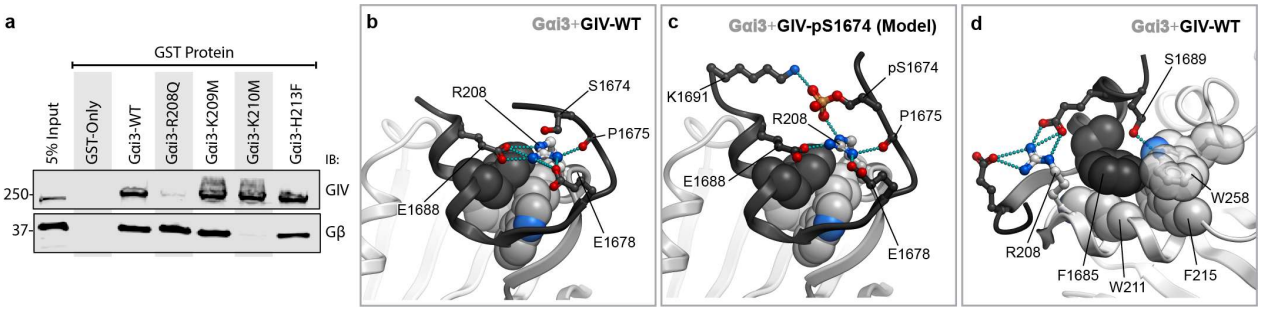
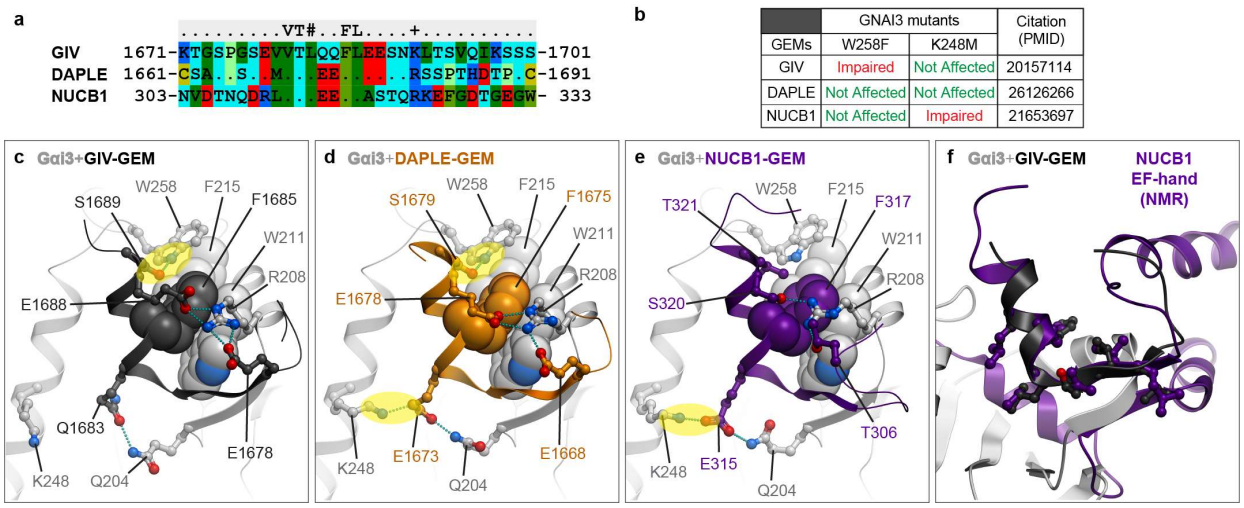


Figure 4.2: Structural basis for phosphoregulation of GIV binding and activity towards Gαi. A) Western blot of GST pull-down assay of WT and mutant GST-tagged Gαi with GIV and Gβγ from HeLa lysate. B) Structure of WT GIV-GEM, highlighting unphosphorylated S1674 and the various contacts of R208 of Gαi. C) Model of (pS1674)GIV-GEM highlighting the formation of an additional direct contact with R208. D) Structure of WT GIV-GEM, highlighting a polar contact that unphosphorylated S1689 makes with W258 of Gαi.

Homology modeling of other GEM family members, Daple and NUCB1, suggested a conserved mode of binding with a few subtle differences that corroborate prior mutagenesis findings (17-19) (Figure 4.3A-E). Interestingly, the GEM motif of NUCB1 maps onto one of the EF-hand motifs of this protein (20); modeling suggests not only full compatibility of the EF-hand topology with Gαi Sw-II binding, but also structural mimicry between such binding and the canonical EF-hand-mediated molecular fold (Figure 4.3F).

In our structure, the N-terminal linker of each Gαi molecule binds to its symmetry neighbor, positioning the linker Arg and surrounding residues across the nucleotide cleft in a manner similar to GoLoco GDIs (21) (Figure 2.S2A-C). Because this interaction was predicted to stabilize the bound GDP, we henceforth refer to the N-terminal linker as the nucleotide-stabilizing linker (NSL). Removal of the NSL (M. Garcia-Marcos, personal correspondence) or changing its position (this work) produced no crystals, suggesting that the linker trapped an otherwise transient and likely non-crystallizable GEF-induced conformation of Gαi•GDP. To determine whether the NSL has confounded structural observations at the GIV-GEM interface with Gαi, we solved the structure of the NSL-containing Gαi•GDP with KB752 (Figure 4.S2A, D-E) and compared it to a previously published complex *without* the NSL (11). No discernible differences were noted in the Gαi-KB752 interface (Figure 4.S2F), suggesting that the observed features at the Gαi-GIV-GEM interface are also representative of the native interactions.



	GNAI3 mutants		Citation (PMID)
GEMs	W258F	K248M	
GIV	Impaired	Not Affected	20157114
DAPLE	Not Affected	Not Affected	26126266
NUCB1	Not Affected	Impaired	21653697

Figure 4.3: Homology models of Gαi•GDP bound to the various members of the GEM family suggest a conserved mechanism of binding and action. A) Sequence alignment of the GEM motifs within human GIV, Daple, and NUCB1 (Calnuc) sequences. B) Table summarizing previous mutagenesis studies. C) Crystal structure of GIV-GEM bound to Gαi. D-E) Homology models of (d) Daple and (e) NUCB1 bound to Gαi created using the GIV-GEM-bound structure as template. Hydrogen bonds explaining the mutagenesis in b are highlighted. F) Overlay of our GIV-GEM-bound Gαi structure with the EF-hand motif of NUCB1, previously determined by NMR (PDB 1SNL).

GIV-GEM binding disfavors the high-GDP-affinity conformations of Gαi SwII and Q204

Upon binding, GIV-GEM accelerates the basal nucleotide exchange of monomeric Gαi (9). To understand the structural basis for this phenomenon, we compared the newly solved structure with all previously crystallized GDP-bound complexes of Gαi. The complexes were organized in order of decreasing GDP affinity, from GoLoco GDI-bound and Gβγ-bound (high GDP affinity), through GDP-only (basal affinity) to KB752- and GIV-bound (low GDP affinity). A clear trend emerged in the position of Sw-I and the molecular contacts of Q204 in Sw-II. In high-GDP-affinity states, Q204 appears to stabilize Sw-I in an outward position, away from the nucleotide-binding pocket (Figure 4.4A-B). By contrast, in the KB752-bound Gαi structure, Q204 is displaced away from Sw-I allowing the latter to "collapse" towards the bound nucleotide (Figure 4.4D). GIV-GEM produces a similar but more exacerbated effect: it stabilizes an elevated conformation of Sw-II, hydrogen-bonds to Gαi Q204 via Q1683 and pulls it ~11 Å away from Sw-I, leading to an even greater contraction of the GDP binding site that also involves a displacement of the β2-strand (Figure 4.4E-F). Despite this collapse, the N-terminal part of Sw-I is found in the outward position, likely due to the presence of the NSL (Figure 4.4E).

These observations prompted us to probe the role of Gαi Q204 in GIV-GEM-mediated GDP-release. A Q204A mutant was generated and tested in a kinetic assay where GDP released from Gαi is replaced by MANT-GTPγS, a non-hydrolyzable fluorescent GTP analogue (22-23). Because GDP release is the rate-limiting step of nucleotide exchange, increases in MANT-GTPγS incorporation rate by Gαi reflect the acceleration of GDP release (24). Compared to WT, Gαi-Q204A displayed a small but consistent increase in the basal nucleotide exchange rate (1.28-fold; Figure 4.4G-H); however, it was significantly more sensitive to activation by GIV-GEM (3.25-fold compared to 1.84-fold for WT Gαi; Figure 4.4G-H). These findings suggest that Q204 indeed negatively regulates nucleotide exchange, likely by stabilizing Sw-I in the high-GDP-affinity state. Interestingly, the direct contact between GIV Q1683 and Gαi Q204 appears unnecessary for accelerated nucleotide exchange because a GIV-Q1683A mutant fully retained its GEF function (Figure 4.S3).

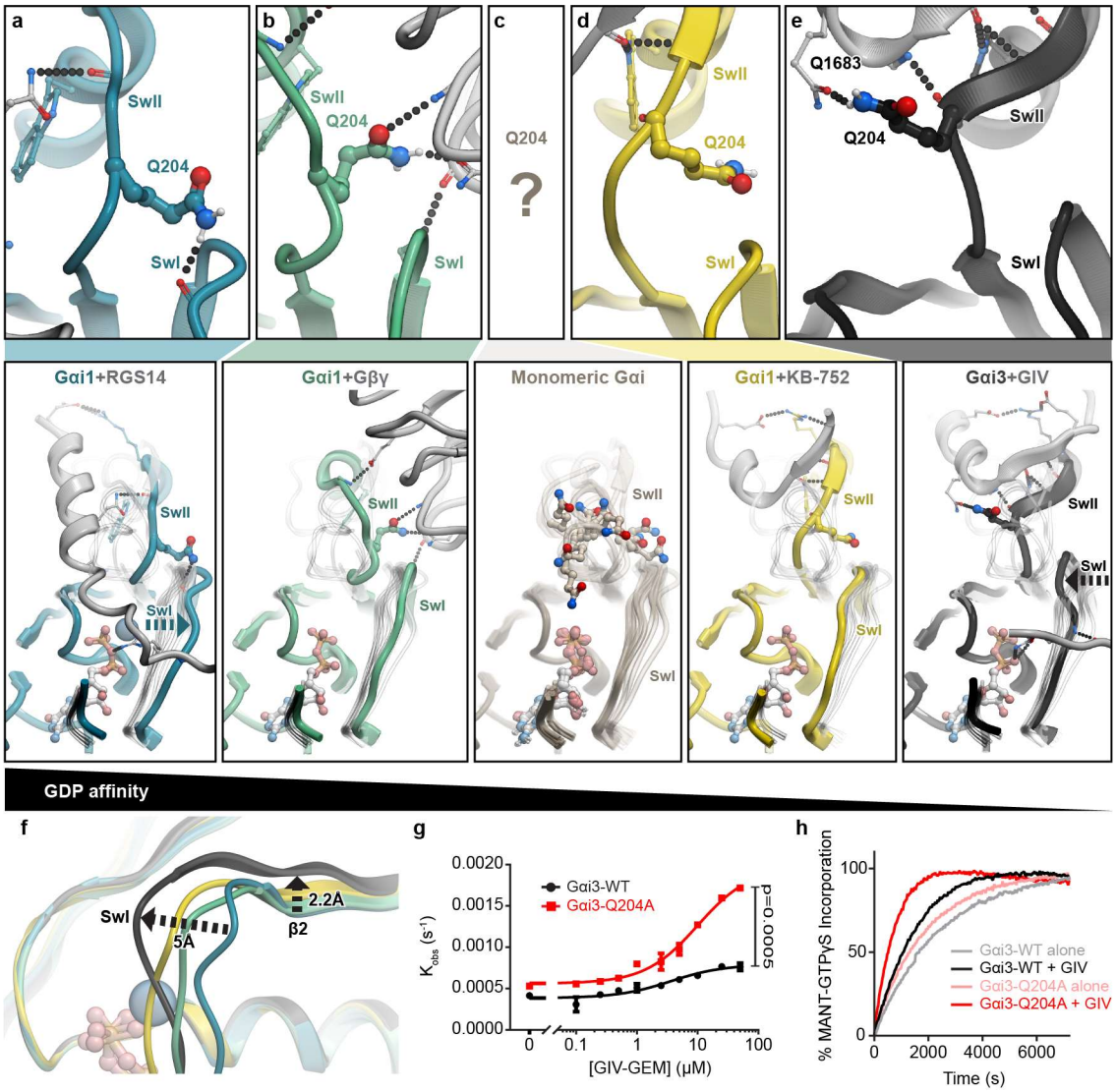


Figure 4.4: GIV binding to Sw-II of Gαi disrupts GDP-stabilizing interactions between Sw-II and Sw-I and induces a low-GDP-affinity conformation of Gαi. A-E) Comparison of Sw-I, Sw-II, and Q204 in various GDP-bound structures of Gαi, arranged from high (left) to low (right) GDP-affinity states. F) Overlay of structures shown in a-b and d-e, highlighting differences in Sw-I and the β2-strand. G) MANT-GTPγS incorporation into WT and Q204A Gαi proteins was assessed in the presence of varying concentrations of WT GIV-GEM peptide. Findings are displayed as a line graph showing observed rates (k_{obs}, s^{-1}) for nucleotide incorporation. Data shown is triplicates from a representative experiment; n = 3. H) Same data as in g presented as a line graph showing average nucleotide incorporation over time in the presence or absence of 50 μM WT GIV-GEM peptide.

Binding of GIV-GEM to Gαi overcomes the allosteric GDP-stabilizing role of hydrophobic residues in SwII

Besides Q204, GIV-GEM also directly engages the aromatic residues W211 and F215 in Sw-II of Gαi; these residues were previously proven critical for GIV-GEM binding (9). We hypothesized that the packing of these bulky hydrophobic residues against the β-barrel of the GTPase domain may stabilize GDP in the basal state (Figure 4.5A). If so, binding of GIV-GEM to Sw-II may neutralize such GDP-stabilizing effects to stimulate GDP release. Alanine substitutions, W211A or F215A, resulted in substantial increases in the basal nucleotide exchange rate of Gαi in MANT-GTPγS incorporation assays (2.48- and 1.84-fold increases, respectively; Figure 4.5B-C). Mutation of a hydrophobic residue on Sw-II that is not necessary for GIV-GEM binding, V218A, showed a small decrease in nucleotide exchange rate (Figure 4.5B-C). Consistent results were obtained in thermal stability assays where the two fast-exchanging Gαi mutants, W211A and F215A, displayed lower melting temperatures in both native and GDP-bound state compared to WT and V218A Gαi (Figure 4.S4). These results suggest that W211 and F215 on Sw-II contribute to stabilization of the bound GDP, an effect that is neutralized by GIV-GEM binding.

To understand the global allosteric changes in Gαi caused by the loss of bulky hydrophobics in Sw-II, we subjected WT and mutant Gαi to hydrogen-deuterium exchange mass spectrometry (HDXMS): a sensitive technique that uses deuterium labeling of protein backbone amides (25) to probe conformational dynamics and mutation-induced allostery (26-27) (Table 4.S3). The V218A mutant showed no measurable difference in deuterium uptake compared to WT Gαi (Figure 4.5D, 4.S5). By contrast, the fast-exchanging W211A mutant exhibited regions of higher deuterium uptake indicative of increased dynamics. The fragment spanning Sw-I and the β2-strand (aa 176-191) showed the highest increase in deuterium uptake in the W211A mutant compared to the WT protein (7.20% increase; Figure 4.5E, 4.S5). Other regions with increased deuterium uptake in the mutant include the C-terminal end of the α4-helix through the β6-strand (peptides 311-323 and 308-323, 5.23% increase) and the αD-αE (NDS) loop (peptide 140-154, 4.24% increase) (Figure 4.5E, 4.S5). Although it is impossible to state whether these changes are a trigger or a consequence of GDP release, the findings are consistent with the role of W211 on Sw-II as an allosteric stabilizer of Sw-I and the β2-strand of Gαi, and thus, of the overall high-GDP-affinity state of the protein.

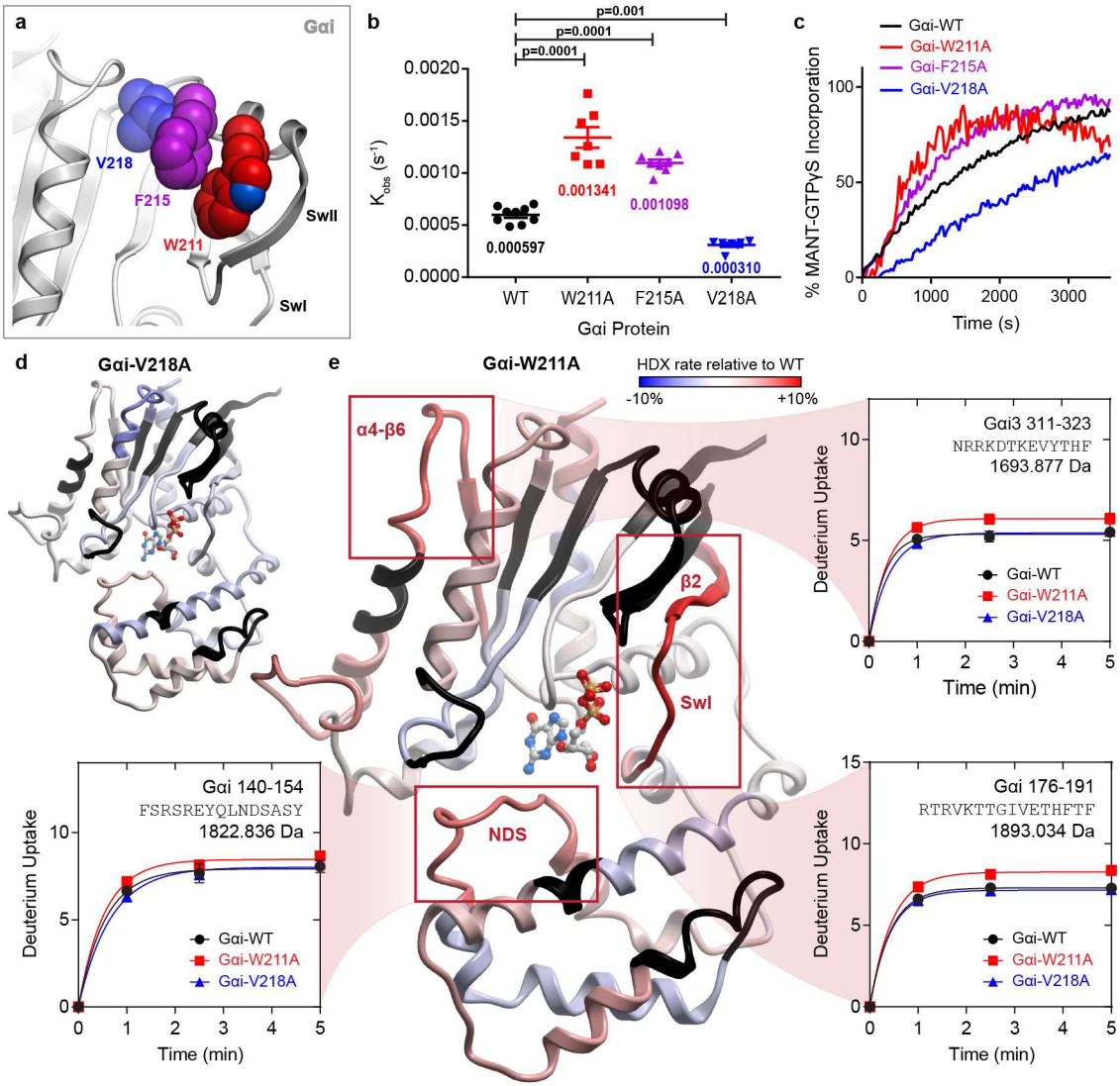


Figure 4.5: Bulky hydrophobic residues in Sw-II of Gαi that are engaged by GIV stabilize GDP and influence the dynamics of Sw-I and the β2-strand. A) Structure showing hydrophobic residues in Sw-II of Gαi that were subjected to mutagenesis. B-C) MANT-GTPγS incorporation into WT, W211A, F215A, and V218A Gαi. Findings are displayed as a dot plot (b) showing the observed nucleotide incorporation rates (k_{obs}, s^{-1}) and as line graphs (c) showing average nucleotide incorporation over time. Data shown is from three independent experiments; n = 9, 7, 8 and 7 for WT, W211A, F215A, and V218A, respectively. D-E) Differences in relative deuterium uptake between V218A and WT Gαi (d) and between W211A and WT Gαi (e) at 5 min, as determined by triplicate HDXMS assays. Blue and red coloring corresponds to -10% and +10% change, respectively, black indicates regions that were not mapped. Regions exhibiting increased uptake in the W211A mutant are highlighted and the corresponding deuterium uptake plots shown (standard deviation error bars are within the symbols).

Understanding the mechanism of nucleotide release acceleration with molecular dynamics

To gain further insights into the allosteric regulation of Gαi and the mechanism by which GIV-GEM accelerates GDP release, we carried out molecular dynamic (MD) simulations. Using the crystallized conformation of Gαi•GDP as a starting point, 350 ns of protein dynamics were simulated in triplicates for the GDP only and GIV-GEM-bound states (1050 ns total for each state), and 3×200 ns were simulated in the GIV-GEM+NSL bound state. A root mean square fluctuation (RMSF) analysis of the centers of mass of Gαi residues demonstrates that Sw-II is highly dynamic in the GDP-only simulation (Figure 4.6A), in agreement with its invariably disordered state in WT Gαi•GDP crystal structures (28-30) (Figure 4.4C). Binding of GIV-GEM to Gαi Sw-II increased its rigidity as expected, but it also unexpectedly stabilized Sw-III that has no direct contact with the peptide (Figure 4.6A-C). The most striking increase in dynamics was observed in the C-terminal region of Sw-I and the β2-β3 strands (Figure 4.6B-C), which normally pack against the α1 and α5 helices of Gαi to form the hydrophobic core (31) of the GTPase domain. When simulations were run in the presence of the NSL, the dynamics of Sw-II was unchanged with respect to the GIV-bound state, the high mobility of Sw-III was restored to the GDP-only level, and the GIV-induced increase in Sw-I and β2-β3 strand dynamics was partially negated, in agreement with the NSL's role in nucleotide stabilization (Figure 4.S6A). These data support the idea that binding of GIV-GEM to Sw-II allosterically perturbs Sw-I and the β2-strand; it also suggests that the perturbation is further propagated to the hydrophobic core of the GTPase domain of Gαi.

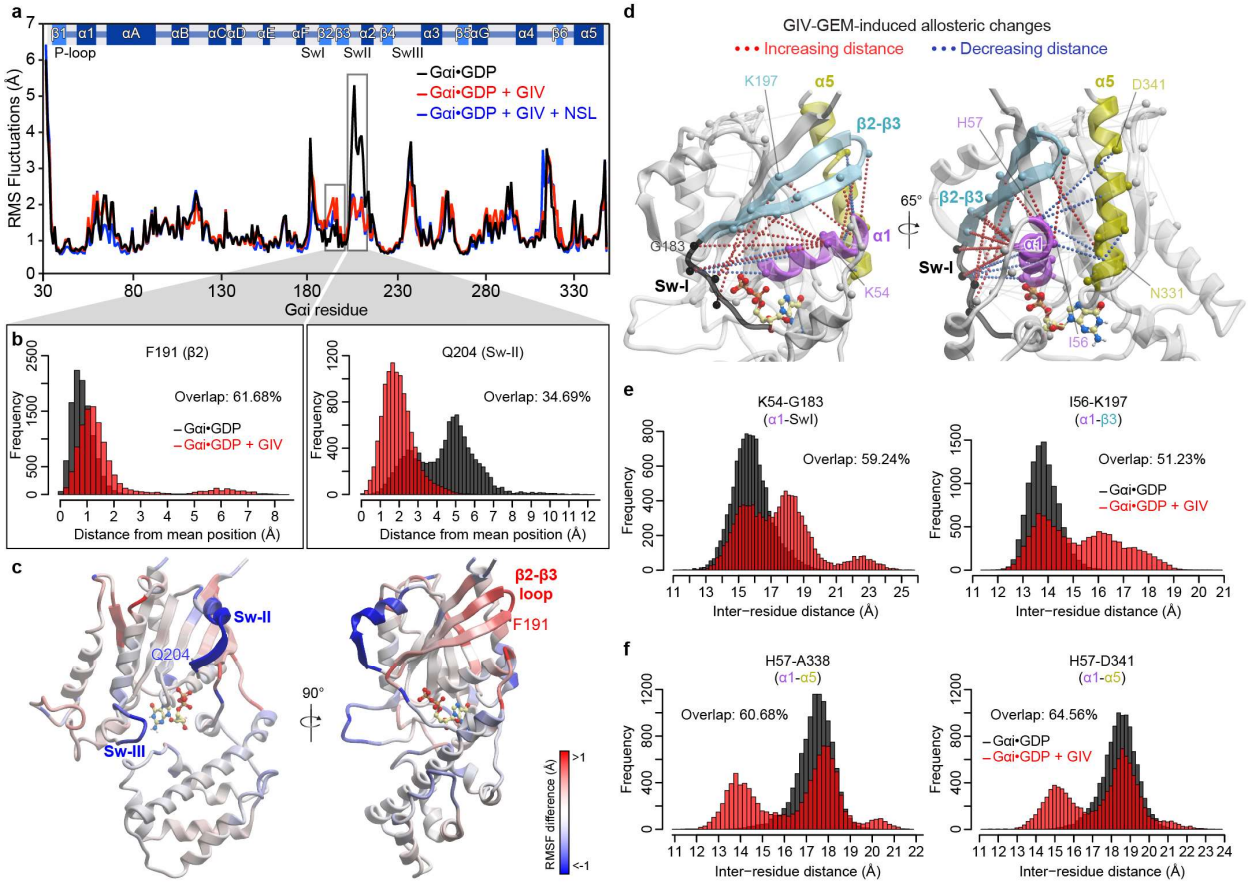


Figure 4.6: Binding of GIV-GEM overcomes the GDP-stabilizing role of Sw-II and releases conformational constraints on Sw-I, β2-β3 strands, and the hydrophobic core of the GTPase domain. A) Root Mean Square Fluctuations (RMSF, Å) of Gαi residues as determined by molecular dynamics simulations under the three specified conditions. B) Representative histograms of residue deviations from their respective mean positions across all simulations under given conditions. C) Residue RMSF differences between the GIV-GEM-bound Gαi•GDP and Gαi•GDP alone mapped onto the structure of Gαi. D) Intramolecular distances where the most significant changes between the two simulation conditions (as in c), as determined by PCA, are shown as dotted lines; significant distances beyond the hydrophobic core are colored silver. E-F) Distribution of inter-residue distances for the indicated residue pairs throughout the molecular dynamics simulations.

To pinpoint the dominant allosteric changes in Gαi induced by GIV-GEM, we projected the pairwise Gαi residue (center of mass) distances onto a lower-dimension space via principal component analysis (PCA, Figure 4.S6B, Table 4.S4). Sw-II was excluded from the PCA to selectively detect allosteric changes rather than direct consequences of GIV-GEM binding. In the first principal component, the largest contributions were from the distances within the hydrophobic core of the GTPase domain that changed consistently and substantially upon GIV-GEM-binding: those from the β2-β3 strands to helix α1 systematically increased, and those from α1 to α5 systematically decreased (Figure 4.6D-F, 4.S6C). In addition, GIV-GEM binding resulted in a distance increase between GDP and R178 (a residue known to stabilize GDP) and a concomitant decrease in distance between the GDP and the αF-helix (preceding Sw-I) (Figure 4.S6D-F), indicative of inward collapse of Sw-I as predicted (Figure 4.4). Many of the Gαi residues highlighted by this analysis were retrospectively found to play important roles in GDP binding (32).

GDP dissociation did not occur over the course of our 3×350 ns simulations with GIV-GEM, consistent with the reported high affinity of GDP to Gαi and its ability to stay bound through much longer simulations (total 42 μs simulation for Gαi-GDP only) unless the protein conformation is substantially perturbed (33).

DISCUSSION

Overall, this work provides the first atomic level structure of a naturally occurring GEM bound to Gαi. The structure provides mechanistic insights into key aspects of GEM biology, including the ability of GEMs to dissociate Gαi complexes with Gβγ (thereby initiating Gβγ signaling) and GoLoco-containing proteins (thereby antagonizing the GDI action of such proteins) (9, 14). The structure also explains the basis for phosphoregulation of GIV-GEM.

Furthermore, this study elucidates the mechanism by which GEMs accelerate GDP release from Gαi. MD simulations, HDXMS, and nucleotide exchange experiments reveal a previously unknown role of Gαi Sw-II in nucleotide affinity. Stabilization of the elevated Sw-II conformation by GIV-GEM releases conformational constraints on Sw-I and β2-β3 strands of Gαi, allowing for inward collapse of the former and higher mobility of the latter. This perturbation propagates to the hydrophobic core in the center of the GTPase domain that was previously shown to contribute to both basal and GPCR-accelerated nucleotide exchange in Gαi (31, 34). Furthermore, structures of GPCR-bound G proteins demonstrate that GPCRs directly disrupt this core by inserting into it a hydrophobic residue from the intracellular loop 2 (3-6). Thus, our findings suggest that despite binding at non-overlapping interfaces on Gαi, GEMs and GPCRs share a part of their allosteric mechanism for acceleration of GDP release by both disrupting the hydrophobic core of the GTPase domain of Gαi (Figure 4.7). These similarities escaped detection in earlier studies employing molecular modeling (9) and NMR (35).

Because nucleotide exchange is an inherently dynamic process, our serendipitously identified NSL has likely facilitated the crystallization of an otherwise unstable and transient complex, much like the intentionally introduced conformation-specific nanobodies in other GEF-bound structures of G proteins (3). While the insights from the structure alone may be limited by its static nature and possibly the presence of the NSL, the complementary computational, biophysical, and biochemical experiments provide a holistic understanding of the diverse mechanisms for allosteric regulation of Gαi.

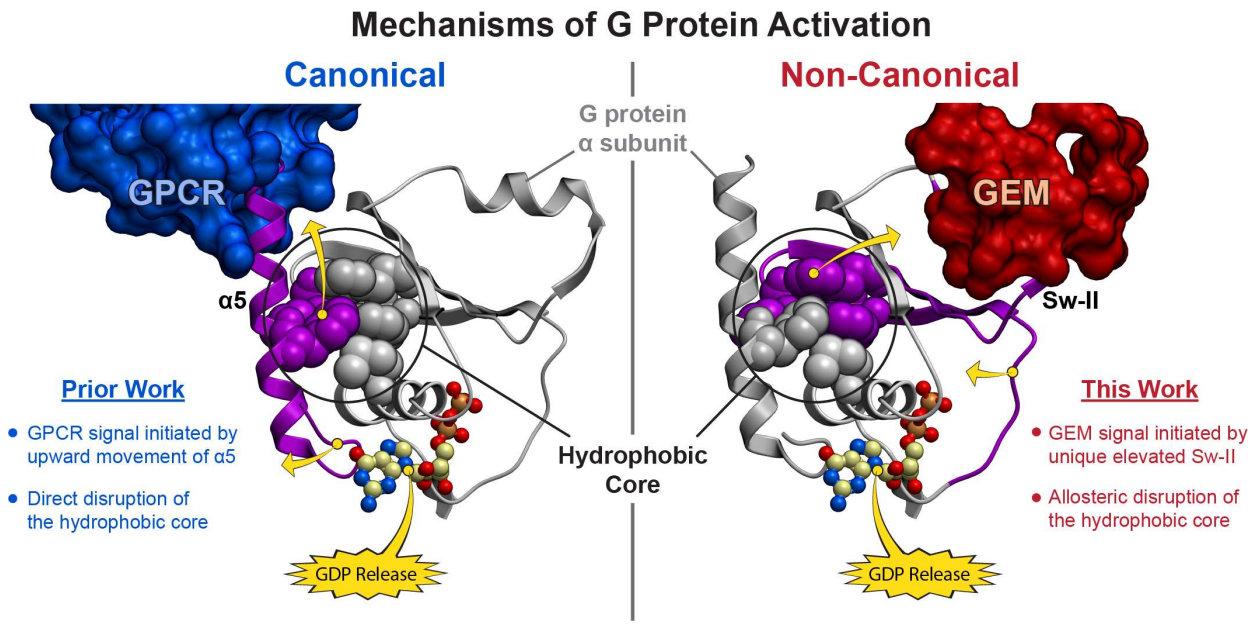


Figure 4.7: GEMs and GPCRs bind at non-overlapping interfaces on Gαi but both disrupt the hydrophobic core of the GTPase domain to stimulate GDP release. **Left**, Structure displaying GPCR interface and subsequent Gαi dynamics that ultimately result in GDP release. **Right**, Structure displaying GEM interface and subsequent Gαi dynamics that ultimately result in GDP release. Purple color highlights regions of Gαi that move during activation, while yellow arrows describe the direction those regions move. For clarity, only part of the GTPase domain of Gαi is shown.

EXPERIMENTAL PROCEDURES

Plasmid constructs and mutagenesis

All restriction endonucleases and *Escherichia coli* strain DH5α were purchased from New England Biolabs (Ipswich, MA). For crystallization, biochemical experiments, and HDXMS, rat Gαi3 (Uniprot P08753-1) was cloned into a pET28b vector using NdeI and XhoI restriction sites, resulting in an N-terminal 6xHis tag separated from the Gαi3 protein by the sequence SSGLVPRGSHM. The nucleotide sequences encoding for the tag and the linker was:

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCAT-

followed by the start ATG codon of WT Gαi3; this construct is referred to 6xHis-Gαi3. In the construct used for crystallization, the N-terminal 25 amino acids of Gαi3 were removed to facilitate crystallization as previously done (11); this construct will be referred to as 6xHis-ΔN25-Gαi3. For GST pull down assays, full length Gαi3 was cloned into a pGEX vector with an N-terminal GST-tag, resulting in a GST-Gαi3. All site-directed mutagenesis (Gai3 Q204A, R208Q, K209M, K210M, W211A, H213F, F215A, and V218A, were carried out using QuikChange II site-directed mutagenesis kit (Agilent Technologies; Santa Clara, CA; Ca#200524) as per the manufacturer's protocol. In the main text, all three constructs (6xHis-Gαi3, 6xHis-ΔN25-Gαi3, and GST-Gαi3) are referred to as Gαi; in the methods below, specific constructs used in each experiment are detailed.

Sequences

Cloned full length *Rattus norvegicus* Gαi3 sequence:

MGSSHHHHHHSSGLVPRGSHMGCTLSAEDKAAVERSKMIDRNLREDGEKAAKEVKLLLLGAGESGKSTI
VKQMKIIHEDGYSEDECKQYKVVVYSNTIQSIIAIIRAMGRLKIDFGEAARADDARQLFVLAGSAEEGVMTS
ELAGVIKRLWRDGGVQACFSRSREYQLNDSASYYLNDLDRISQTNYIPTQQDVLRTRVKTTGIVETHFTFK
ELYFKMFDVGGQRSERKKWIHCFEGVTAIIFCVALSDYDLVLAEDEEMNRMHESMKLFDSICNNKWFTDT
SIILFLNKKDLFEEKIKRSPLTICYPEYTGSNTYEEAAAYIQCQFEDLNRRKDTKEVYTHFTCATDTKNVQFV
FDAVTDVIIKNNLKECGLY

Cloned 25aa N-terminally truncated *Rattus norvegicus* Gαi3 sequence:

MGSSHHHHHHSSGLVPRGSHMDGEKAAKEVKLLLLGAGESGKSTIVKQMKIIHEDGYSEDECKQYKVVV
YSNTIQSIIAIIRAMGRLKIDFGEAARADDARQLFVLAGSAEEGVMTSELAGVIKRLWRDGGVQACFSRSR
EYQLNDSASYYLNDLDRISQTNYIPTQQDVLRTRVKTTGIVETHFTFKELYFKMFDVGGQRSERKKWIHCF
EGVTAIIFCVALSDYDLVLAEDEEMNRMHESMKLFDSICNNKWFTDTSIILFLNKKDLFEEKIKRSPLTICYP
EYTGSNTYEEAAAYIQCQFEDLNRRKDTKEVYTHFTCATDTKNVQFVFDAVTDVIIKNNLKECGLY

Peptides

Peptides were synthesized by three companies independently [LifeTein (Somerset, NJ), Chempeptide (Shanghai, China), and AbClonal (Woburn, MA)] and all displayed comparable effects in assays. Peptides were synthesized with L-amino acids at >95% purity and kept frozen at -80°C as 10 mM stocks in DMSO.

KB-752 peptide sequence: 1-NH2-SRVTWYDFLMEDTKSR-COOH-16

GIV-WT GEM-motif peptide sequence: 1671-NH2-KTGSPGSEVVTLQQFLEESNKLTSVQIKSSS-COOH-1701

GIV-Q1683A GEM-motif peptide sequence: 1671-NH2-KTGSPGSEVVTLAQFLEESNKLTSVQIKSSS –COOH-1701

Expression and purification of Gαi3

6xHis-tagged Gαi3 constructs (6xHis-Gαi3, 6xHis-ΔN25-Gαi3, or single-point mutants of thereof) were transformed into *E. coli* BL21 (DE3; Invitrogen) cells. Cells were grown in 1 L flasks at 37°C until OD reached 0.8-1.0, then induced overnight at 25°C with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Cells were harvested via centrifugation and lysed at 15,000 PSI by a single pass through a cell disruptor (TS-Series; Constant Systems, Inc) in Running Buffer (RB; 50 mM NaH_2PO_4 pH 7.4, 300 mM NaCl, and 0.5 mM EDTA) supplemented with 2x Protease Inhibitors (Roche Life Science) and 10 mM imidazole. Cell debris was removed by ultracentrifugation at 45,000×g for 40 min, and the supernatant was loaded on a Ni–NTA His60 Superflow resin (Qiagen) affinity column via fast protein liquid chromatography (AKTA, GE Life Sciences). The resin was washed with RB+60 mM imidazole, and eluted with RB+300 mM imidazole. The eluted protein was concentrated at 1500×g (Amicon Ultra-15 30 MWCO centrifugal filter; Millipore) and

subjected to size exclusion chromatography via Superdex 200 resin (GE Healthcare) equilibrated with storage buffer (20 mM Tris-HCl pH 7.4, 20 mM NaCl, 1 mM $MgCl_2$, and 5% glycerol). Fractions from major peak were pooled, resulting in usually ~1-5 mg/mL Gαi protein. Protein was then aliquoted, flash frozen, and stored at −80°C. Protein concentration and purity were checked throughout purification via SDS-PAGE and comparison to known amounts of Bovine Serum Albumin (BSA).

GST-alone and GST-tagged Gαi3 constructs (wild type and mutant proteins) were expressed and purified from *Escherichia coli* strain BL21 (DE3; Invitrogen) as described previously. Briefly, cells were grown in 1 L flasks at 37°C until OD reached 0.8-1.0, then induced overnight at 25°C with 1 mM IPTG. A bacterial pellet from 1 L of culture was resuspended in 10 ml of GST-lysis buffer (25 mM Tris-HCl, pH 7.5, 20 mM NaCl, 1 mM EDTA, 20% [vol/vol] glycerol, 1% [vol/vol] Triton X-100, 2× protease inhibitor cocktail [Complete EDTA-free; Roche Diagnostics]). Cell lysates were sonicated (4 × 20 s, 1 min between cycles) and then centrifuged at 12,000 × *g* at 4°C for 20 min. Solubilized proteins were affinity purified on glutathione-Sepharose 4B beads (GE Healthcare) by incubation for 4 hours at 4°C. Beads were washed 3 x with 50mM Tris pH 8 and then eluted with GST elution buffer (50 mM Tris pH 8, 10 mM reduced glutathione). Eluted proteins were dialyzed overnight at 4°C against phosphate-buffered saline (PBS), and stored at −80°C in aliquots.

Co-crystallization of Gαi3 with KB-752 and GIV-GEM

Purified 3 mg/mL 6xHis-ΔN25-Gαi3 (either freshly prepped or freeze-thawed once) was incubated overnight in storage buffer (20 mM Tris-HCl pH 7.4, 20 mM NaCl, 1 mM $MgCl_2$, and 5% glycerol) at a 3:1 (peptide:Gαi3) molar ratio at 4°C, then concentrated to ~15 mg/mL and set on 288-well Intelli-Plate trays (Art Robbins Instruments) in 1:1, 1.5:1, and 2:1 volume ratios with mother liquor (12-16% PEG 3350, 0.2 M NH_4Cl) at room temperature. Crystals appeared after 1-2 days and grew to full size in 5-7 days. Crystals were cryoprotected by soaking in mother liquor supplemented with 10% glycerol and flash-frozen with liquid nitrogen.

X-ray data collection and structure determination of Gαi3-peptide co-crystal structures

X-ray diffraction data were collected at 100 K at the Lawrence Berkeley National Laboratory Advanced Light Source (8.2.2) and Stanford Synchrotron Radiation Lightsource (9-2) at a single wavelength. All diffraction data were indexed and integrated with MOSFLM, processed with AIMLESS, and

truncated with CTRUNCATE within the CCP4 suite of programs (37-39). Phases were estimated via molecular replacement in Phaser (40), using a previously published model of human Gαi1 (PDB 1y3a, for Gαi3•GDP with KB-752) or human Gαi3 (PDB 4g5r, for Gαi3•GDP with GIV-GEM) as a search model. Models underwent rigid-body and restrained positional refinement using PHENIX.REFINE in the PHENIX software suite (41) against a maximum likelihood target function, alternated with manual inspection against electron density maps in Coot (42). Peptides were manually modeled in Coot and refined in the final rounds of refinement, which also included the application of hydrogens to their riding positions and simulated annealing. The resulting refinement statistics for each model are included in Table 4.S1. Figures displaying crystal packing were prepared using PyMOL (http://www.pymol.org), and atomic coordinates and structure factors were deposited in the Protein Data Bank (accession codes 6MHE and 6MHF for KB752 and GIV-GEM co-crystal structures, respectively,).

Cell culture

Cells were cultured according to American Type Culture Collection (ATCC) guidelines. Briefly, HeLa cells were grown in high glucose DMEM (Sigma; Ca#D5796) supplemented with 10% (vol/vol) FBS (HyClone; Ca#SH30071.03) and penicillin-streptomycin-glutamine (Gibco; Ca#10378-016). For cell lysates, HeLa cells were grown on 10-cm plates and harvested by scraping into 0.5 mL of lysis buffer [20 mM HEPES pH 7.4, 5 mM Mg-acetate, 125 mM K-acetate, 0.4% Triton X-100, 1 mM DTT, 1× Complete Protease Inhibitor Mixture (Roche; Ca#11873580001), and 1× Phosphatase Inhibitor Mixtures 2 and 3 (Sigma; Ca#P5726 and P0044, respectively)] on ice. Cell lysates were incubated for 10 min at 4°C and were centrifuged at 12,000×g for 10 min. Clarified cell lysates were subsequently used in GST pulldown assays.

In vitro GST pulldown assays

Purified GST-Gαi3 or GST-alone (5 µg) were immobilized on glutathione-Sepharose beads and incubated with binding buffer [50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 0.4% (vol/vol) Nonidet P-40, 10 mM $MgCl_2$, 5 mM EDTA, 30 µM GDP, 2 mM DTT, 1× Complete Protease Inhibitor Mixture (Roche; Ca#11873580001)] for 90 min at room temperature as described before (14, 43-45). Lysates (~250 µg protein) of HeLa cells were added to each tube, and binding reactions were carried out for 4 hr at 4ºC with constant tumbling in binding buffer [50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 0.4% (vol/vol) Nonidet P-40,

10 mM $MgCl_2$, 5 mM EDTA, 30 μM GDP, 2 mM DTT]. Beads were washed (4X) with 1 mL of wash buffer [4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4 (pH 7.4), 137 mM NaCl, 2.7 mM KCl, 0.1% (vol/vol) Tween-20, 10 mM $MgCl_2$, 5 mM EDTA, 30 μM GDP, 2 mM DTT] and boiled in Laemmli's sample buffer for 10min. For immunoblotting, rabbit anti-Gβ primary antibody (M-14; Ca#sc-261) and anti-GIV-CT primary antibody (T-13; Ca#sc-133371) were obtained from Santa Cruz Biotechnology (Dallas, TX). IRDye 680RD goat anti-rabbit secondary antibody (Ca#926-68071) and IRDye 800 goat anti-mouse secondary antibody (Ca#926-32210) were from Li-Cor Biosciences (Lincoln, NE). Protein samples were separated by SDS-PAGE and transferred to PVDF membranes (Millipore). Membranes were blocked with PBS supplemented with 5% nonfat milk before incubation with primary antibodies (1:500 dilutions overnight at 4ºC). Blots were washed 3 times in PBS-T [4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4 (pH 7.4), 137 mM NaCl, 2.7 mM KCl, 0.1% (vol/vol) Tween-20] and incubated with secondary antibodies (1:20,000 dilutions at room temperature for 1 hour). Blots were then washed 3 times in PBS-T and once with PBS before infrared imaging following the manufacturer's protocols using an Odyssey imaging system (Li-Cor Biosciences).

Molecular modeling

Models of Gαi•GDP with (pS1674)GIV-GEM, Gαi•GDP with Daple-GEM, and Gαi•GDP with NUCB1-GEM were constructed by homology with the structure of Gαi•GDP with GIV-GEM using ICM versions 3.8-6 to 3.8-7a (Molsoft LLC, San Diego, CA).

The GEM motif peptides from (pS1674)GIV (1671-KTG-pS1674-PGSEVVTLQQFLEESNK-1691) and Daple (1663-ASPSSEMVTLEEFLEESNR-1681) were built *ab initio*; the GEM motif peptide from NUCB1 (305-DTNQDRLVTLEEFLASTQRKEF-326) was extracted from the NMR structure of NUCB1 (PDB 1snl (20)). The backbone atoms of the peptides were confined to the crystallographic coordinates of the corresponding atoms of GIV-GEM (residues 1676-GSEVVTLQQFLEES-1689 only) via a set of harmonic distance restraints (tethers); the peptide flanks and side-chains were kept unrestrained. Full-atom conformational sampling of the peptides (backbone, side-chains, and positional variables) and the surrounding side-chains of Gαi was performed using 5×10^6 steps of biased probability Monte Carlo optimization (46) as implemented in ICM, with the repulsive part of the Van der Waals potential capped at 20 kJ/mol. The top scoring pose of each peptide was selected for analysis.

Gαi3-limited proteolysis assay

6xHis-Gαi3 or 6xHis-ΔN25-Gαi3 (0.25 mg ml^{-1}) was incubated for 150 min at 30°C in buffer (20 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA, 10 mM $MgCl_2$, and 1 mM DTT) supplemented with GDP (30 μM) or GTPγS (30 μM). After incubation trypsin was added to the tubes (final concentration 6.25 μg ml^{-1}) and samples were incubated for 10 min at 30°C. Samples were rapidly transferred to ice, reactions were stopped by the addition of Laemmli sample buffer, after which the samples were boiled for 10 min. Proteins were separated by SDS–PAGE and stained with Coomassie blue.

MANT-GTPγS incorporation assays

MANT-GTPγS incorporation assays in Figure 4.2 were done using full length 6xHis-ΔN25-Gαi3, whereas incorporation assays in Figure 4.3 were done using the 6xHis-Gαi3 construct. For Gαi3 incorporation assays in the presence of peptide, peptide was pre-bound to Gαi3 prior to running the assay. To equilibrate and pre-bind peptide to Gαi3, 111 nM His-Gαi3 WT or mutants were first incubated in reaction buffer (20 mM HEPES pH 8, 100 mM NaCl, 1 mM EDTA, 10 mM $MgCl_2$, and 1 mM DTT) in 30°C water bath for 30 min with or without varying concentrations of peptide in a final incubation volume of 250 μL. After equilibration, 72 μL protein-peptide complexes were transferred to a pre-warmed 384-well black flat-bottom plate (in triplicates). The reaction was initiated by injecting 8 μL of 250 nM MANT-GTPγS (Abcam, Cambridge, MA) in each well for a final reaction volume of 80 μL and final concentrations of 100 nM Gαi3, 25 nM MANT-GTPγS and the indicated concentrations of the peptide. MANT-GTPγS incorporation into Gαi3 was quantified, either by FRET (ex = 280; em = 440) or by direct MANT excitation (ex = 350; em = 440), using a microplate fluorescence reader (TECAN Spark 20M). Fluorescence was measured every 30 sec starting immediately after injection of MANT-GTPgS. Raw fluorescence was plotted over time and observed rates (k_{obs}) were determined by fitting a one-phase association curve to the data (GraphPad Prism 7).

Differential scanning fluorimetry (thermal shift assays)

6xHis-Gαi3 WT and mutants (5 μM) were taken in their native state (as purified) or loaded with GDP by incubating it for 150 min at 30 °C in buffer (20 mM HEPES, pH 8, 100 mM NaCl, 1 mM EDTA, 10 mM MgCl2, and 1 mM DTT) supplemented with GDP (1 mM). After loading, 45 μL of 5 μM His-Gαi3 was pipetted into PCR tubes (in triplicates) and 5 μL 200X SYPRO Orange solution freshly made in the same buffer from 5000X stock (Life Technologies S-6650) was added to the protein. A buffer + dye only (no

protein) control was also included. Thermal shift assays were run on an Applied Biosystems StepOnePlus Real-Time PCR machine. Mixed protein and dye samples were subjected to increasing temperatures from 25 to 95°C in half degree increments, holding each temperature for 30 sec and measuring SYPRO fluorescence (using filter 3 for TAMRA™ and NED™ dyes) at each temperature. Melting temperatures were defined as the temperature at which the maximum value for the derivative of signal fluorescence (dF/dt) is achieved (GraphPad Prism 7).

Hydrogen-deuterium exchange mass spectrometry (HDXMS)

HDXMS measurements were made using a Synapt G2Si system (Waters Corporation). Deuterium exchange reactions were carried out by a Leap HDX PAL autosampler (Leap Technologies, Carrboro, NC). Deuterated buffer was prepared by lyophilizing 10 mL of 20 mM Tris-HCl pH 7.4, 20 mM NaCl, 5 μM GDP and 5% glycerol and resuspending it in 10 mL 99.96% D_2O immediately before use. Each deuterium exchange time point (0 min, 1 min, 2.5 min, 5 min) was measured in triplicate. For each measurement, 5 μL of 100 μM 6xHis-Gαi3 protein [in storage buffer (20 mM Tris-HCl pH 7.4, 20 mM NaCl, 1 mM $MgCl_2$, and 5% glycerol) was mixed with 55 μL of D_2O buffer at 25°C. Deuterium exchange was quenched by combining 50 μL of the deuterated sample with 50 μL of 0.1% formic acid and 3M guanidinum-HCl for 1 min at 1°C. The quenched sample was then injected in a 50 μL sample loop and digested by an inline pepsin column (Pierce, Inc.) at 15°C. The resulting peptides were captured on a BEH C18 Vanguard precolumn, separated by analytical chromatography (Acquity UPLC BEH C18, 1.7 μm, 1.0 × 50 mm, Waters Corporation) using 7−85% acetonitrile in 0.1% formic acid over 7.5 min, and analyzed in a Waters Synapt G2Si quadrupole time-of-flight mass spectrometer following electrospray injection.

Data were collected in Mobility, ESI+ mode, mass acquisition range of 200−2000 (m/z), scan time 0.4 sec. Continuous lock mass correction was performed using infusion of leu-enkephalin (m/z = 556.277) every 30 sec (mass accuracy of 1 ppm for calibration standard). For peptide identification, data were instead collected in MS E (mobility ESI+) mode. Peptides masses were identified following triplicate analysis of 10 μM Gαi3, and were analyzed using PLGS 2.5 (Waters Corporation). Peptides masses were identified using a minimum number of 250 ion counts for low energy peptides and 50 ion counts for their fragment ions; with the additional constraint that peptide size was greater than 1500 Da. The following parameters were used to filter peptide sequence matches: minimum products per amino acid of 0.2, minimum score of 7,

maximum MH+ error of 5 ppm, and a retention time RSD of 5%, and the peptides had to be present in two of the three ID runs collected. After identification in PLGS, peptides were analyzed in DynamX 3.0 (Waters Corporation). Deuterium uptake for each peptide was calculated by comparing the centroids of the mass envelopes of the deuterated samples with the undeuterated controls. To account for back-exchange and systematic autosampler sample handling differences between the shorter 1 min and longer 2.5 min and 5 min deuteration times, the uptake and standard deviation values were divided by 0.79 and 0.75, respectively. Data were plotted as number of deuterons incorporated vs time. The Y-axis limit for each plot reflects the total number of amides within the peptide that can possible exchange. Each plot includes the peptide MH+ value, sequence, and sequential residue numbering.

Molecular dynamic (MD) simulations

MD simulations were performed with AMBER package (v. 16) in periodic boundary conditions. Three different complexes were created and refined in ICM, and then used to create Amber topology files: Gαi•GDP, Gαi•GDP with GIV-GEM, and Gαi•GDP with GIV-GEM and NSL (amino acid residues GLVPRGS from the linker of the crystallographic neighbor molecule). In all cases, the Gαi molecule contained residues 30-347 only. Residue protonation states were assigned by the "convertObject" utility in ICM (Molsoft LLC, San Diego, CA), followed by *pdb4amber* conversion. The proteins were represented using the ff14SB force-field. The parameters for GDP were taken from the AMBER parameter database (www.pharmacy.manchester.ac.uk/bryce/amber) (47).

Each system was solvated with explicitly represented water (TIP3P model, 12 Å margin). K^+ and Cl^- ions were added to maintain neutrality of the system and represent approximately 150 mM K^+. The system compositions were as follows. The Gαi•GDP system consisted of 47,571 atoms in the 97×83×73 Å box, including 14,121 water molecules, 39 K^+ ions and 30 Cl^- ions. The Gαi•GDP+GIV-GEM system consisted of 50,454 atoms in the 97×83×77 Å box, including 15,000 water molecules, 42 K^+ ions and 30 Cl^- ions. The Gαi•GDP+GIV-GEM+NSL system consisted of 50,463 atoms in the 97×83×77 Å box, including 14,969 water molecules, 42 K^+ ions and 31 Cl^- ions. The systems were minimized with restrained heavy atoms of the solute, followed by an unrestrained minimization.

The systems were slowly heated (1 K/ps) up to 310 K with Langevin thermostat (γ = 1 ps^{-1}). The density of the systems was equilibrated in a series of short (20 ps) simulations at isotropic pressure of 1

bar with MC barostat. All bond lengths to hydrogen atoms were constrained with SHAKE algorithm (ntc = 2, ntf = 2). The non-bonded interactions cutoff was 12 Å. The time step was 0.5 fs. Additionally, the systems were equilibrated in the NPT ensemble for 10 ns with time step of 2 fs.

The production runs were started in triplicates from the last frame of the equilibration stage. The simulations were carried out in the same NPT ensemble (310 K, 1 bar), non-bonded cutoff (12 Å) and time step (2 fs) for 100 ns. Each run started with a unique random seed that affects the thermostat.

The MD simulations were performed with GPU-accelerated PMEMD using SPFP precision model on Nvidia GeForce GTX 680 and GTX TITAN cards.

MD trajectory analysis

MD simulation analyses were performed in ICM v3.8-7a (Molsoft LLC, San Diego, CA), unless otherwise stated. Replicate simulations of a single condition were concatenated together for analysis.

For Root Mean Square Fluctuation (RMSF) analysis, MD frames from each condition were superimposed by the backbone (C, N, O and Cα atoms), using *cpptraj*, within the AMBER package (48). The coordinates of the center of mass of each Gαi residue *j* at frame *f*, $\boldsymbol{r}_f(j) = \langle x_f(j), y_f(j), z_f(j) \rangle$, were given by:

$$\boldsymbol{r}_f(j) = \frac{\sum_{a \in A_j} m(a) \times \boldsymbol{r}(a)}{\sum_{a \in A_j} m(a)}$$

where A_j is the set of all non-hydrogen atoms in residue *j*, $m(a)$ is the atomic weight of atom *a*, and $\boldsymbol{r}(a) = \langle x(a), y(a), z(a) \rangle$ is the vector of Cartesian coordinates of atom *a*. The mean coordinates of residue *j* over all *F* frames of the trajectory were calculated as $\bar{\boldsymbol{r}}(j) = \frac{1}{F}\sum_{f=1}^{F} \boldsymbol{r}_f(j)$, and the RMSF of residue *j* as:

$$RMSF_j = \sqrt{\frac{1}{F}\sum_{f=1}^{F} \left|\boldsymbol{r}_f(j) - \bar{\boldsymbol{r}}(j)\right|^2}$$

where || denotes the length of the vector in Cartesian coordinates. Differences in residue RMSF between the MD conditions were mapped onto the crystal structure and visualized.

To trace intramolecular motions in Gαi induced by GIV-GEM, Euclidean distances between centers of mass of amino-acid residue pairs, or between residues and GDP (50,721 pairs total), were calculated for each frame of the simulation. Residue pairs were filtered to retain only 1658 pairs that satisfied the following criteria: (1) they were at least two residues apart in the sequence; (2) they were separated by less than 12

Å in at least one MD frame; (3) their distances displayed less than 75% overlap in frequency distribution between the Gαi•GDP and Gαi•GDP+GIV-GEM simulations; (4) they did not involve Sw-II residues (residues 202 to 218). To calculate the overlap between the distance frequency distributions, the distance range was broken into *b* 0.2 Å intervals and the binned relative frequencies of the center of mass distance were computed for each pair of residues *i*, *j* over the course of the Gαi•GDP trajectory:

$$\langle p_1^{GDP}(i,j), p_2^{GDP}(i,j), \dots, p_b^{GDP}(i,j)\rangle, \qquad \sum_{k=1}^{b} p_k^{GDP}(i,j) = 1$$

and the same was done for the Gαi•GDP+GIV-GEM trajectory giving $p_k^{GIV}(i,j), k \in \{1, \dots, b\}$. The overlap was given by $\sum_{k=1}^{b} \min\left(p_k^{GDP}(i,j), p_k^{GIV}(i,j)\right)$. The 1658 non-trivial residue pairs were subjected to principal component (PC) analysis to identify those pairs whose changing distances contribute the most to the dominant modes of motion upon GIV-GEM binding, in an unbiased manner. Only the first PC was analyzed because it correctly discriminated the simulation conditions (Figure 4.S8B). Residue pairs assigned with the largest weights and associated with the first PC were mapped onto the crystal structure for visualization.

Statistical analysis

Each experiment presented in the figures is representative of at least three independent repeats (with at least two technical repeats for each condition within each repeat). Statistical significance between the differences of means was calculated using multiple comparisons in one-way nonparametric ANOVA. All statistics and graphical data presented were prepared using GraphPad Prism. Histograms of MD simulation data were generated in R.

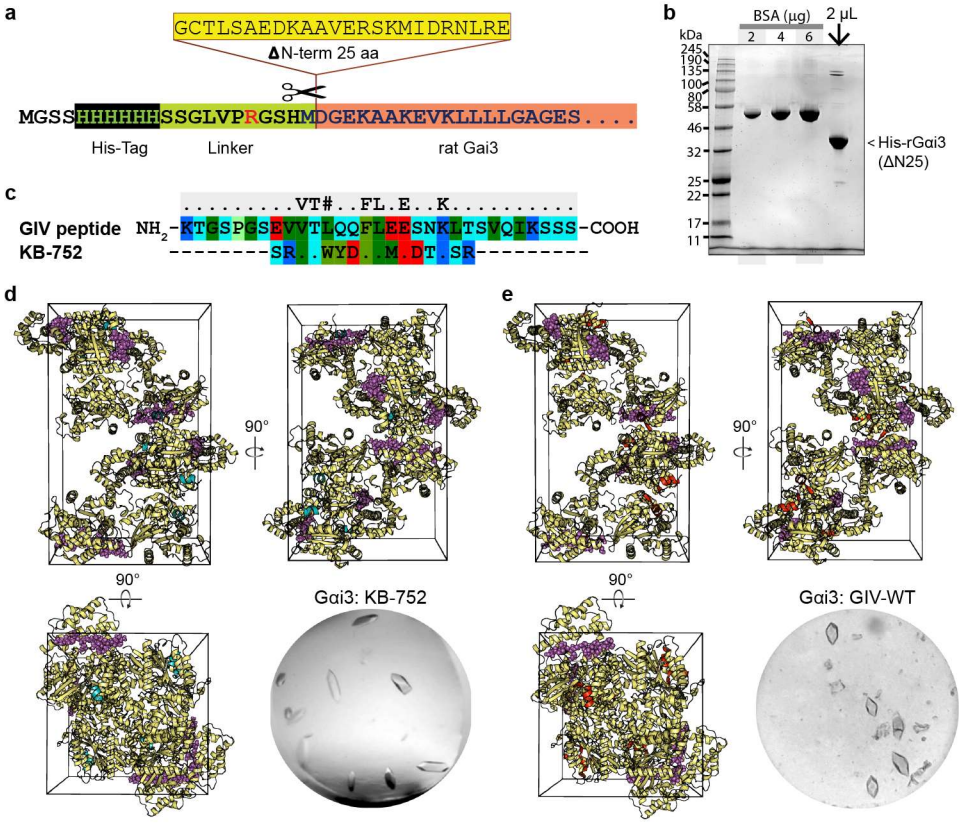


Figure 4.S1: Gαi construct design and crystallization. A) Design of the rat Gαi3 construct that was used to produce protein for crystallography in this study. The 25 aa N-terminal helix of Gαi3 was removed and replaced by an N-terminal His-tag separated from the β1-strand by an 11-residue linker (referred to as ΔN25-HisGαi3). B) SDS-PAGE analysis and Coomassie-Blue staining of the purified ΔN25-HisGαi3 side-by-side with bovine serum albumin (BSA) standards confirms the expected molecular weight, purity, and high yield for ΔN25-HisGαi3. C) Sequence alignment of GIV-GEM with the previously described synthetic peptide, KB752[11]. D) Unit cell and crystal image for our solved KB752-bound Gαi•GDP structure. Yellow, Gαi; purple, NSL; cyan, KB752. E) Unit cell and crystal image for our solved GIV-GEM-bound Gαi•GDP structure. Structure is colored as in d except GIV-GEM is in red.

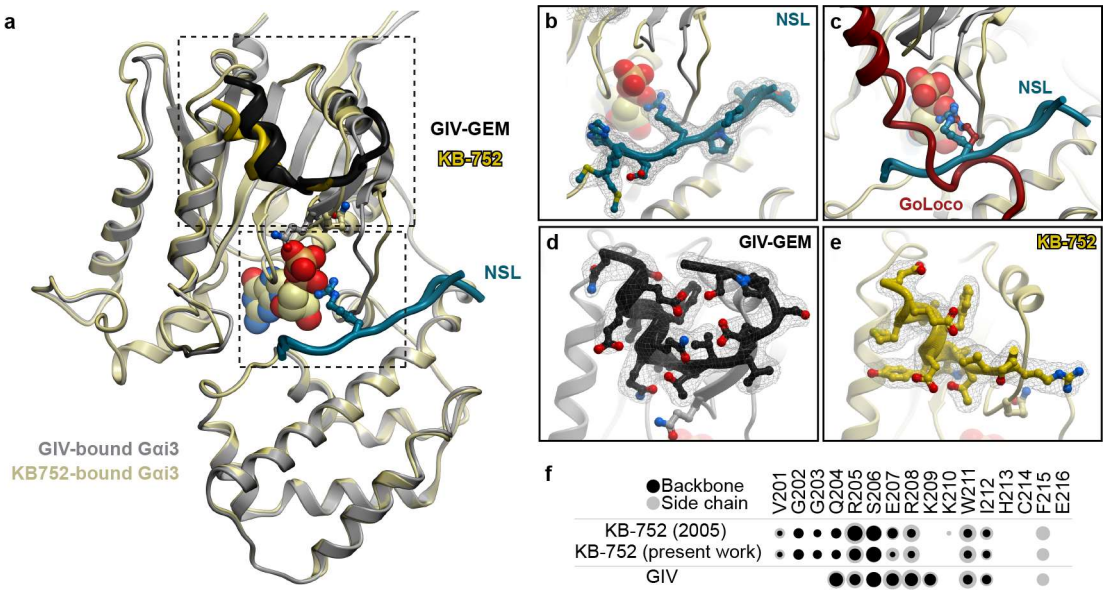


Figure 4.S2: Structures of KB752-bound and GIV-GEM-bound Gαi•GDP. A) Overlay of our solved KB752-bound and GIV-GEM-bound Gαi•GDP structures. Boxed regions are highlighted in b-e. B-C) 2Fo-Fc electron density maps around the GIV-GEM peptide (b) or KB752 (c) are contoured at 3σ. D) 2Fo-Fc electron density map around the NSL in the Gαi•GDP structure with GIV-GEM is contoured at 3σ. E) Overlay of the NSLs with the GoLoco motif of GDI RGS14 (PDB 1KJY). F) Bubble plot of the contacts between Sw-II residues of Gαi and the KB752 peptide, as seen in the previously published (2005) KB752-bound structure *without* the NSL and our solved structures *with* the NSL. The size of the dot is directly proportional to the strength of the contact (36).

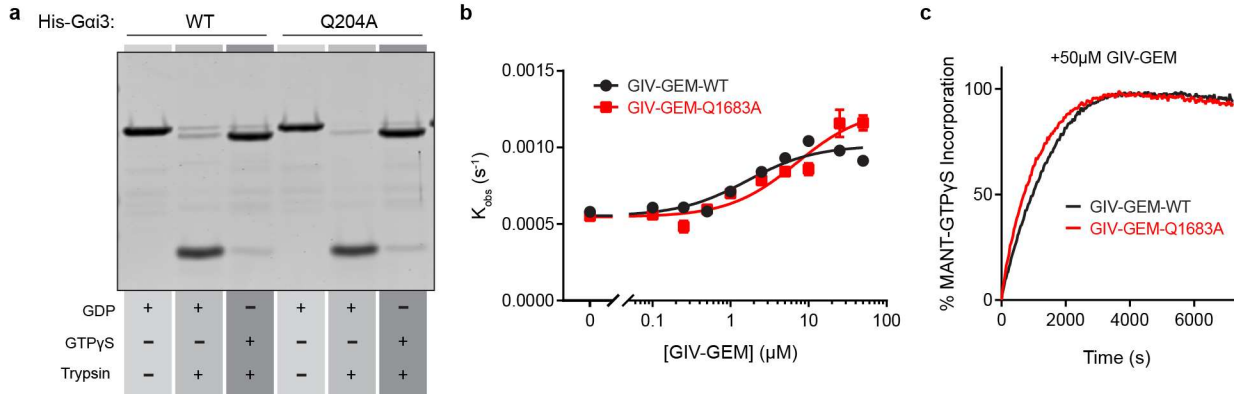


Figure 4.S3: The polar contact observed in the crystal structure between GIV Q1683 and Gαi Q204 is dispensable for activation of Gαi by GIV-GEM. A) Coomassie stain of a trypsin proteolysis assay performed on WT and Q204A mutant Gαi proteins loaded with GDP or GTPγS. B-C) MANT-GTPγS incorporation into Gαi was assessed in the presence of increasing concentrations of WT and Q1683A mutant GIV-GEM proteins. Findings are displayed as a line graph (b) showing observed rates (k_{obs}, s^{-1}) for nucleotide incorporation and as line graphs (c) showing average nucleotide incorporation over time. Data shown is triplicates from a representative experiment; n = 3.

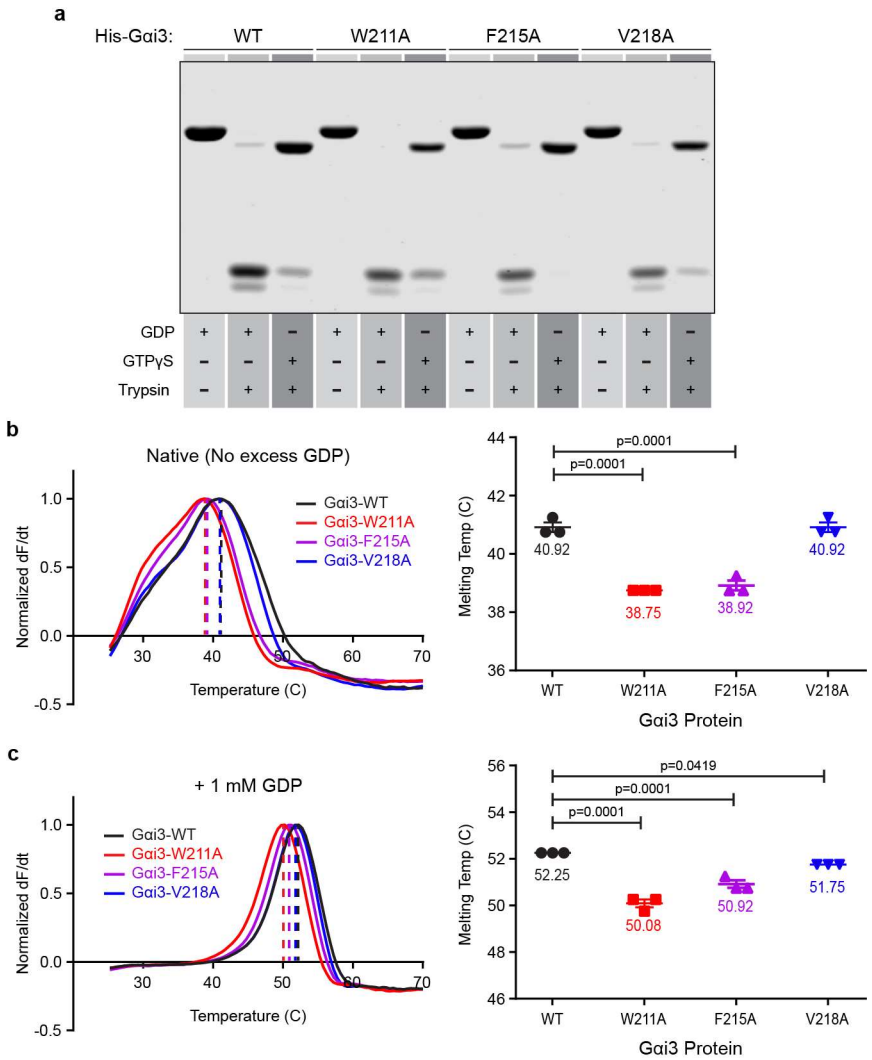


Figure 4.S4: Trypsin proteolysis and thermal shift assays support the GDP-stabilizing role of Sw-II residues W211 and F215. A) Coomassie stain of a trypsin proteolysis assay performed on WT and Sw-II mutant Gαi proteins loaded with GDP or GTPγS. B) WT and Sw-II mutant Gαi proteins were subjected to increasing temperatures in differential scanning fluorimetry (thermal shift) assay. Findings are displayed as a line graph (left) showing average normalized dF/dt curves of *Native* (no excess GDP added) or as dot plot (right) showing the melting temperatures of *Native* WT and Sw-II mutant Gαi proteins. Peaks of the curves (left) represent protein melting temperatures. Data shown are triplicates from a representative experiment; n = 3. C) Same as in b except 1 mM GDP was added.

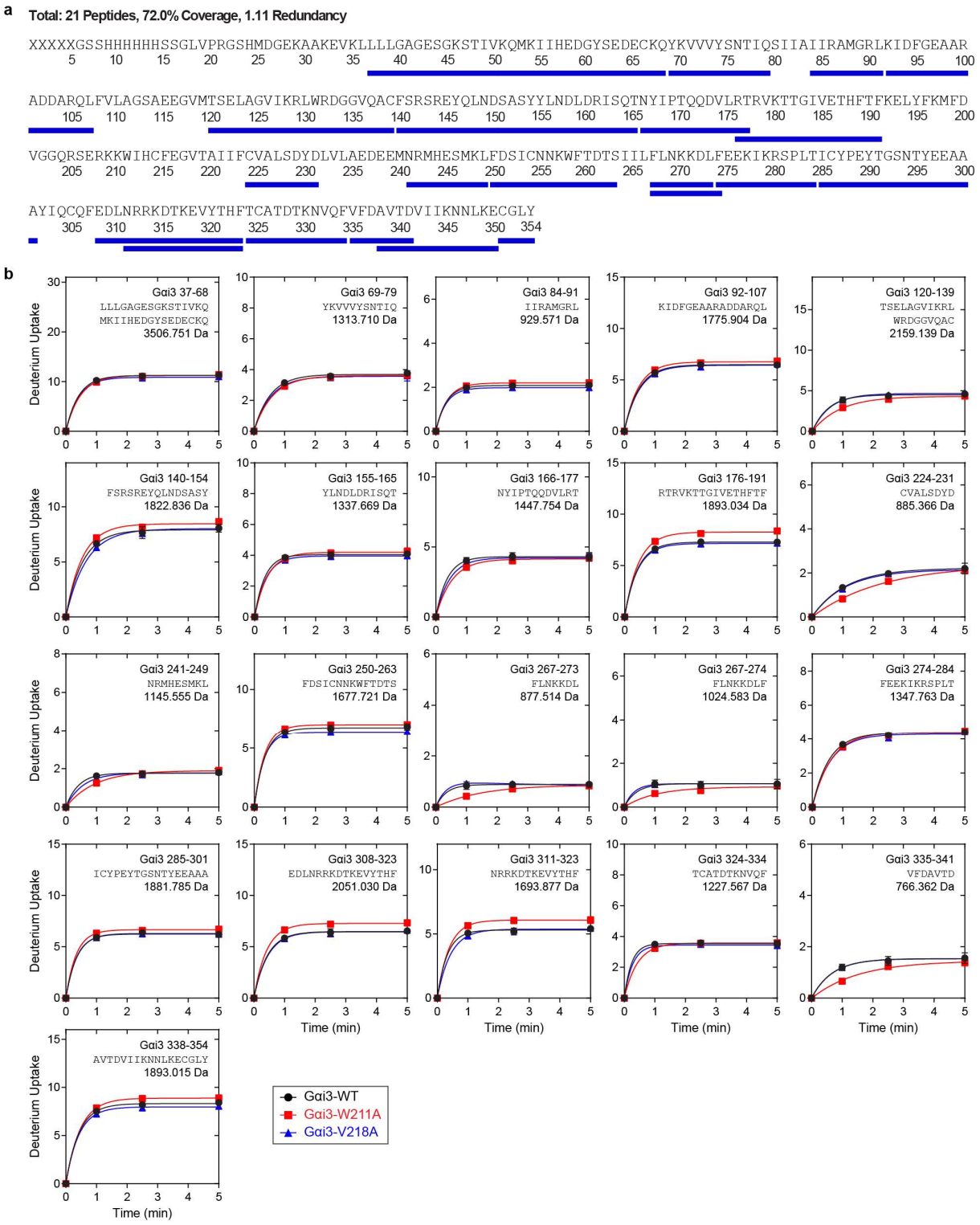


Figure 4.S5: Coverage map and deuterium uptake plots for HDX assays. A) Sequence coverage of Gαi following pepsin digestion, LC separation, and MS/MS detection. Blue bars represent distinct peptides. **B)** Plots showing deuterium uptake for distinct peptides at various time points. Experiments were performed in triplicate and standard deviation error bars are shown (although most are within the symbols).

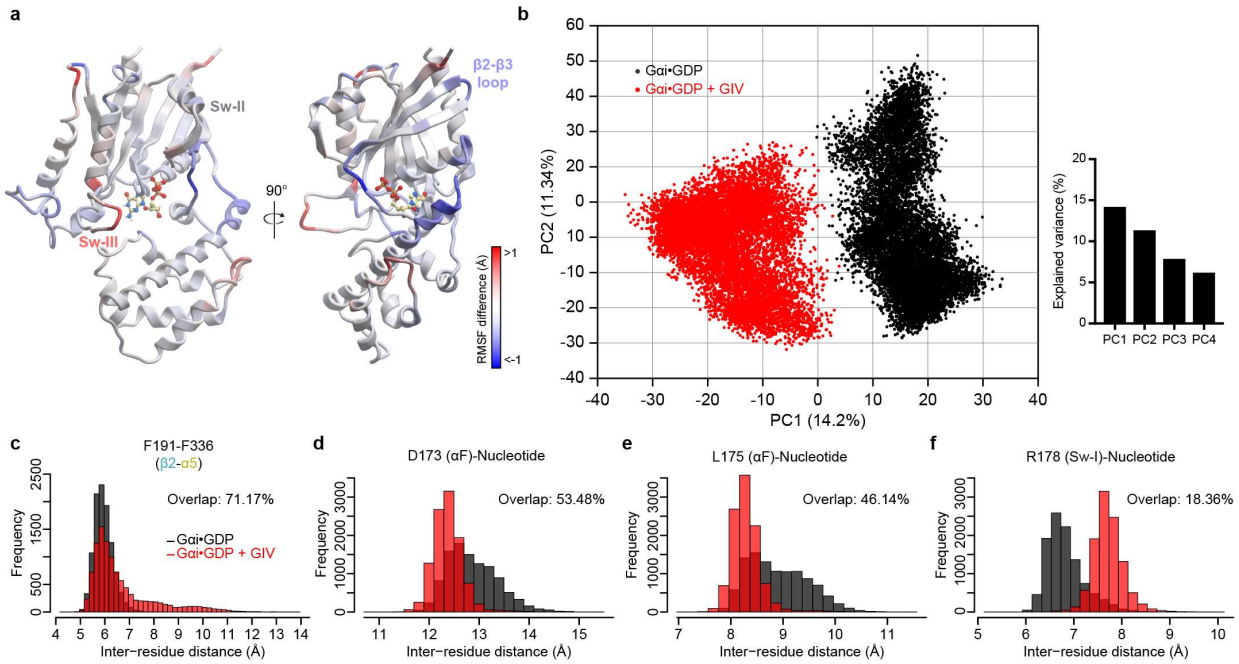


Figure 4.S6: MD simulations support the stabilizing role of the NSL and identify GIV-GEM-induced structural changes. A) Residue RMSF differences between the GIV-GEM-bound Gαi•GDP *with* the NSL and GIV-GEM-bound Gαi•GDP *without* the NSL are mapped onto the structure of Gαi. B) (Left) Plot displaying first two principal components as identified by the Principle Component Analysis (PCA) of inter-residue distances in the Gαi•GDP and GIV-GEM-bound Gαi•GDP simulations. Each dot represents an MD frame taken with an interval of 100 ps. (Right) Bar graph showing % variance explained by each of the first four principal components. C-F), Histograms showing inter-residue distances (c) or the distances between the indicated residues and GDP (d-f) throughout the duration of molecular dynamics simulations.

Table 4.S1: Counts of publicly available structures of G proteins in the PDB, by subfamily and complex composition.

	# of PDB structures with:						
Subfamily	GDP	GTP	effector	GAP	GDI	GPCR	non-GPCR
G(s)		1	16			4 (3 cryoEM) + 1* (X-ray)	
G(i/o/t/z)	15	17	3	8	15	3 + 1* (all cryoEM)	THIS WORK
G(q)			9	2	1		
G(12)	1	1	4				

* Structure obtained with a miniG (Nehmé et al., PLOS ONE, 2017).

Table 4.S2: X-ray crystallography data collection and refinement statistics.

	Gαi3:GIV-GEM (6MHF)	Gαi3:KB752 (6MHE)
Data Collection		
Space group	$P4_32_12$	$P4_32_12$
Cell dimensions		
a,b,c (Å)	83.78, 83.78, 141.37	83.24, 83.24, 133.83
α,β,γ (°)	90, 90, 90	90, 90, 90
Beamline	ALS-8.2.2	ALS-8.2.2
Wavelength	0.99997	0.99999
Resolution (Å)	72.08-2.00	66.91-2.20
Rmerge (%)	2.3 (95.7)[1]	5.0 (84.9)[1]
Rpim (%)	1.7 (79.6)[1]	3.3 (60.3)[1]
I/σ (I)	29.5 (1.4)[1]	15.1 (1.6)[1]
CC1/2	0.999 (0.446)[1]	0.994 (0.131)[1]
Completeness (%)	93.9 (99.7)[1]	94.0 (99.7)[1]
Redundancy	4.0 (4.4)[1]	3.7 (3.7)[1]
Observed reflections	147287	100669
Unique reflections	34165	24419
Refinement		
Resolution (Å)	72.08-2.00 (2.05-2.00)[1]	66.91-2.20 (2.27-2.20)[1]
Reflections (working set)	34765	24594
Reflections (test set)	1738	1229
Rwork/Rfree	0.20/0.24	0.23/0.26
R.m.s deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.515	0.504
Ramachandran statistics		
Outliers %	0.29	0.29
Favored %	98.84	98.82
Rotamer Outliers %	1.61	1.66
Cβ Deviations	0	0

[1]Values in parentheses are for the highest-resolution shells.

Table 4.S3: Raw hydrogen-deuterium exchange uptake data.
Excel File; see publication to which this dissertation chapter references.

Table 4.S4: Top 60 intramolecular distances, which contribute to GIV-GEM action, as determined by PCA.
Excel File; see publication to which this dissertation chapter references.

ACKNOWLEDGEMENTS

Authors thank Drs. Barry Grant and Xinqiu Yao for fruitful discussions and help with Bio3D software, and Dr. Suchismita Roy for help with HDXMS interpretation. This work was partially supported by NIH R01 grants AI118985 and R01 GM117424 to I.K, CA100768, CA160911 and DK099226 to P.G, and NSF grant IOS-1444435 to G.C. and S.D.R. N.A.K was supported by a NIH predoctoral fellowship (F31 CA206426), and T32 training grants T32CA067754 and T32DK007202. S.D.R. was supported by the Frontiers of Innovation Scholarship Program, the UCSD Stem Cell Innovative Projects Award, and T32 training grants T32GM007752 and T32 GM008326. T.N. was supported by a NHMRC C. J. Martin Early Career Fellowship 1145746. N.K. was supported by the Frontiers of Innovations Scholarship Program, the UCSD Stem Cell Innovative Projects Award, and T32 training grant T32GM008326. A.I was supported by NIH R01 GM071872. The Synapt G2Si was obtained from NIH grant S10 OD016234. This research was supported in part by the W. M. Keck Foundation through computing resources at the W. M. Keck Laboratory for Integrated Biology at UCSD. We thank Advanced Light Source (ALS), Advanced Photon Source (APS), Stanford Synchrotron Radiation Laboratory (SSRL), and Canadian Light Source (CLS) for screening crystals and subsequent data collection.

AUTHOR CONTRIBUTIONS

N.A.K., S.D.R., P.G. and I.K. participated in research design. N.A.K. and S.D.R. purified WT Gαi protein for crystallography and optimized conditions for crystal formation. S.D.R. harvested crystals, collected and processed diffraction data, and solved and refined the structures with guidance from G.C and I.K. N.A.K. purified Gαi mutants for biochemical experiments and HDXMS, and conducted biochemical experiments, including in vitro binding, trypsin proteolysis, nucleotide exchange, and thermal shift assays. N.K. conducted and analyzed results from HDX experiments with guidance from E.K. A.I. ran and T.N. analyzed molecular dynamics simulations with the guidance from I.K. I.K. supervised analysis of all structural data, including crystal structures, HDXMS, and MD simulations. N.S. assisted with in vitro binding experiments. N.A.K., S.D.R., P.G., and I.K. wrote the manuscript. P.G. and I.K. conceived and supervised the project.

DISSERTATION ACKNOWLEDGEMENTS

Chapter 4, in full, is a reprint of material written for *Nature Structural and Molecular Biology.* Rees SD, Kalogriopoulos NA, Ngo T, Kopcho N, Ilatovskiy A, Sun N, Komives E, Chang G, Ghosh P, Kufareva I. Structural basis for GPCR-independent activation of heterotrimeric Gi proteins. This dissertation author is the co-primary author of this paper.

REFERENCES

1 Gilman, A. G. G proteins: transducers of receptor-generated signals. *Annu Rev Biochem* **56**, 615-649, doi:10.1146/annurev.bi.56.070187.003151 (1987).

2 Morris, A. J. & Malbon, C. C. Physiological regulation of G protein-linked signaling. *Physiol Rev* **79**, 1373-1430, doi:10.1152/physrev.1999.79.4.1373 (1999). Crystal structure of the beta2 adrenergic receptor-Gs protein complex.

3 Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Char, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and B. K. Kobilka. *Nature* **477**, 549-555, doi:10.1038/nature10361 (2011).

4 Koehl, A.., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J. M., Latorraca, N. R., Hilger, D., Dawson, R., Matile, H., Schertler, G. F. X., Granier, S., Weis, W. I., Dror, R. O., Manglik, A., Skiniotis, G., and B. K. Kobilka. Structure of the micro-opioid receptor-Gi protein complex. *Nature*, doi:10.1038/s41586-018-0219-7 (2018).

5 Carpenter, B., Nehme, R., Warne, T., Leslie, A. G. & Tate, C. G. Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* **536**, 104-107, doi:10.1038/nature18966 (2016).

6 Draper-Joyce, C. J., Khoshouei, M., Thal, D. M., Liang, Y., Nguyen, A. T. N., Furness, S. G. B., Venugopal, H., Baltos, J., Plitzko, J. M., Danev, R., Baumeister, W., May, L. T., Wootten, D., Sexton, P. M., Glukhova, A., and A. Christopoulos. Structure of the adenosine-bound human adenosine A1 receptor-Gi complex. *Nature* **558**, 559-563, doi:10.1038/s41586-018-0236-6 (2018).

7 Aznar, N., Kalogriopoulos, N., Midde, K. K. & Ghosh, P. Heterotrimeric G protein signaling via GIV/Girdin: Breaking the rules of engagement, space, and time. *Bioessays* **38**, 379-393, doi:10.1002/bies.201500133 (2016).

8 Ghosh, P., Rangamani, P. & Kufareva, I. The GAPs, GEFs, GDIs and...now, GEMs: New kids on the heterotrimeric G protein signaling block. *Cell Cycle* **16**, 607-612, doi:10.1080/15384101.2017.1282584 (2017).

9 Garcia-Marcos, M., Ghosh, P. & Farquhar, M. G. GIV is a nonreceptor GEF for G alpha i with a unique motif that regulates Akt signaling. *Proc Natl Acad Sci U S A* **106**, 3178-3183, doi:10.1073/pnas.0900294106 (2009).

10 Gupta, V., Bhandari, D., Leyme, A., Aznar, N., Midde, K. K., Lo, I. C., Ear, J., Niesman I., Lopez-Sanchez, I., Blanco-Canosa, J. B., von Zastrow, M., Garcia-Marcos, M., Farquhar, M. G., and Ghosh, P. GIV/Girdin activates Galphai and inhibits Galphas via the same motif. *Proc Natl Acad Sci U S A* **113**, E5721-5730, doi:10.1073/pnas.1609502113 (2016).

11 Johnston, C. A., Willard, F. S., Jezyk, M. R., Fredericks, Z., Bodor, E. T., Jones, M. B., Blaesius, R., Watts, V. J., Harden, T. K., Sondek, J.,Ramer, J. K., and D. P. Siderovski. Structure of Galpha(i1) bound to a GDP-selective peptide provides insight into guanine nucleotide exchange. *Structure* **13**, 1069-1080, doi:10.1016/j.str.2005.04.007 (2005).

12 Ghosh, P. Heterotrimeric G proteins as emerging targets for network based therapy in cancer: End of a long futile campaign striking heads of a Hydra. *Aging (Albany NY)* **7**, 469-474, doi:10.18632/aging.100781 (2015).

13 Sprang, S. R. Invited review: Activation of G proteins by GTP and the mechanism of Galpha-catalyzed GTP hydrolysis. *Biopolymers* **105**, 449-462, doi:10.1002/bip.22836 (2016).

14 Garcia-Marcos, M., Ear, J., Farquhar, M. G. & Ghosh, P. A GDI (AGS3) and a GEF (GIV) regulate autophagy by balancing G protein activity and growth factor signals. *Mol Biol Cell* **22**, 673-686, doi:10.1091/mbc.E10-08-0738 (2011).

15 Bhandari, D., Lopez-Sanchez, I., To, A., Lo, I. C., Aznar, N., Leyme, A., Gupta, V., Niesman, I., Maddox, A. L., Garcia-Marcos, M., Farquhar, M. G., and P. Ghosh. Cyclin-dependent kinase 5

activates guanine nucleotide exchange factor GIV/Girdin to orchestrate migration-proliferation dichotomy. *Proc Natl Acad Sci U S A* **112**, E4874-4883, doi:10.1073/pnas.1514157112 (2015).

16 Lopez-Sanchez, I., Garcia-Marcos, M., Mittal, Y., Aznar, N., Farquhar, M. G., and P. Ghosh. Protein kinase C-theta (PKCtheta) phosphorylates and inhibits the guanine exchange factor, GIV/Girdin. *Proc Natl Acad Sci U S A* **110**, 5510-5515, doi:10.1073/pnas.1303392110 (2013).

17 Garcia-Marcos, M., Ghosh, P., Ear, J. & Farquhar, M. G. A structural determinant that renders G alpha(i) sensitive to activation by GIV/girdin is required to promote cell migration. *J Biol Chem* **285**, 12765-12777, doi:10.1074/jbc.M109.045161 (2010).

18 Aznar, N., Midde, K. K., Dunkel, Y., Lopez-Sanchez, I., Pavlova, Y., Marivin, A., Barbazan, J., Murray, F., Nitsche, U., Janssen, K., Willert, K., Goel, A., Abal, M., Garcia-Marcos, M., and P. Ghosh. Daple is a novel non-receptor GEF required for trimeric G protein activation in Wnt signaling. *Elife* **4**, e07091, doi:10.7554/eLife.07091 (2015).

19 Garcia-Marcos, M., Kietrsunthorn, P. S., Wang, H., Ghosh, P. & Farquhar, M. G. G Protein binding sites on Calnuc (nucleobindin 1) and NUCB2 (nucleobindin 2) define a new class of G(alpha)i-regulatory motifs. *J Biol Chem* **286**, 28138-28149, doi:10.1074/jbc.M110.204099 (2011).

20 de Alba, E. & Tjandra, N. Structural studies on the Ca2+-binding domain of human nucleobindin (calnuc). *Biochemistry* **43**, 10039-10049, doi:10.1021/bi049310a (2004).

21 Kimple, R. J., Kimple, M. E., Betts, L., Sondek, J. & Siderovski, D. P. Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* **416**, 878-881, doi:10.1038/416878a (2002).

22 Remmers, A. E. Detection and quantitation of heterotrimeric G proteins by fluorescence resonance energy transfer. *Anal Biochem* **257**, 89-94, doi:10.1006/abio.1997.2533 (1998).

23 Goricanec, D., Stehle, R., Egloff, P., Grigoriu, S., Pluckthun, A., Wagner, G., and F. Hagn. Conformational dynamics of a G-protein alpha subunit is tightly regulated by nucleotide binding. *Proc Natl Acad Sci U S A* **113**, E3629-3638, doi:10.1073/pnas.1604125113 (2016).

24 Ferguson, K. M., Higashijima, T., Smigel, M. D. & Gilman, A. G. The influence of bound GDP on the kinetics of guanine nucleotide binding to G proteins. *J Biol Chem* **261**, 7393-7399 (1986).

25 Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* **25**, 158-170, doi:10.1002/mas.20064 (2006).

26 Peacock, R. B., Davis, J. R., Markwick, P. R. L. & Komives, E. A. Dynamic Consequences of Mutation of Tryptophan 215 in Thrombin. *Biochemistry* **57**, 2694-2703, doi:10.1021/acs.biochem.8b00262 (2018).

27 Kant, R., Zeng, B., Thomas, C. J., Bothner, B. & Sprang, S. R. Ric-8A, a G protein chaperone with nucleotide exchange activity induces long-range secondary structure changes in Galpha. *Elife* **5**, doi:10.7554/eLife.19238 (2016).

28 Raw, A. S., Coleman, D. E., Gilman, A. G. & Sprang, S. R. Structural and biochemical characterization of the GTPgammaS-, GDP.Pi-, and GDP-bound forms of a GTPase-deficient Gly42 --> Val mutant of Gialpha1. *Biochemistry* **36**, 15660-15669, doi:10.1021/bi971912p (1997).

29 Coleman, D. E. & Sprang, S. R. Crystal structures of the G protein Gi alpha 1 complexed with GDP and Mg2+: a crystallographic titration experiment. *Biochemistry* **37**, 14376-14385, doi:10.1021/bi9810306 (1998).

30 Kapoor, N., Menon, S. T., Chauhan, R., Sachdev, P. & Sakmar, T. P. Structural evidence for a sequential release mechanism for activation of heterotrimeric G proteins. *J Mol Biol* **393**, 882-897, doi:10.1016/j.jmb.2009.08.043 (2009).

31 Kaya, A. I., Lokits, A. D., Gilbert, J. A., Iverson, T. M., Meiler, J., and H. E. Hamm. A Conserved Hydrophobic Core in Galphai1 Regulates G Protein Activation and Release from Activated Receptor. *J Biol Chem* **291**, 19674-19686, doi:10.1074/jbc.M116.745513 (2016).

32 Sun, D.,Flock, T., Deupi, X., Maeda, S., Matkovic, M., Mendieta, S., Mayer, D., Dawson, R., Schertler, G. F. X., Babu, M. M., and D. B. Veprintsev. Probing Galphai1 protein activation at single-amino acid resolution. *Nat Struct Mol Biol* **22**, 686-694, doi:10.1038/nsmb.3070 (2015).

33 Dror, R. O., Mildorf, T. J., Hilger, D., Manglik, A., Borhani, D. W., Arlow, D. H., Philippsen, A., Villanueva, N., Yang, Z., Lerch, M. T., Hubbell, W. L., Kobilka, B. K., Sunahara, R. K., and D. E. Shaw. SIGNAL TRANSDUCTION. Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* **348**, 1361-1365, doi:10.1126/science.aaa5264 (2015).

34 Kaya, A. I., Lokits, A. D., Gilbert, J. A., Iverson, T. M., Meiler, J., and H. E. Hamm. A conserved phenylalanine as a relay between the alpha5 helix and the GDP binding region of heterotrimeric Gi protein alpha subunit. *J Biol Chem* **289**, 24475-24487, doi:10.1074/jbc.M114.572875 (2014).

35 de Opakua, A. I., Parag-Sharma, K., DiGiacomo, V., Merino, N., Leyme, A., Marivin, A., Villate, M., Nguyen, L. T., de la Cruz-Morcillo, M. A., Blanco-Canosa, J. B., Ramachandran, S., Baillie, G. S., Cerione, R. A., Blanco, F. J., and M. Garcia-Marcos. Molecular mechanism of Galphai activation by non-GPCR proteins with a Galpha-Binding and Activating motif. *Nat Commun* **8**, 15163, doi:10.1038/ncomms15163 (2017).

36 Kufareva, I., Rueda, M., Katritch, V., Stevens, R. C., Abagyan, R., GPCR Dock 2010 participants. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* **19**, 1108-1126, doi:10.1016/j.str.2011.05.012 (2011).

37 Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**, 271-281, doi:10.1107/S0907444910048675 (2011).

38 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82, doi:10.1107/S0907444905036693 (2006).

39 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and K. S. Wilson. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

40 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and R. J. Read. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

41 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Offner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

42 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

43 Ghosh, P., Beas, A. O., Bornheimer, S. J., Garcia-Marcos, M., Forry, E. P., Johannson, C., Ear, J., Jung, B. H., Cabrera, B., Carethers, J. M., and M. G. Farquhar. A G{alpha}i-GIV molecular complex binds epidermal growth factor receptor and determines whether cells migrate or proliferate. *Mol Biol Cell* **21**, 2338-2354, doi:10.1091/mbc.E10-01-0028 (2010).

44 Ghosh, P., Garcia-Marcos, M., Bornheimer, S. J. & Farquhar, M. G. Activation of Galphai3 triggers cell migration via regulation of GIV. *J Cell Biol* **182**, 381-393, doi:10.1083/jcb.200712066 (2008).

45 Lin, C., Ear, J., Pavlova, Y., Mittal, Y., Kufareva, I., Ghassemian, M., Abagyan, R., Garcia-Marcos, M., and P. Ghosh. Tyrosine phosphorylation of the Galpha-interacting protein GIV promotes activation of phosphoinositide 3-kinase during cell migration. *Sci Signal* **4**, ra64, doi:10.1126/scisignal.2002049 (2011).

46 Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* **235**, 983-1002, doi:10.1006/jmbi.1994.1052 (1994).

47 Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **24**, 1016-1025, doi:10.1002/jcc.10262 (2003).

48 Roe, D. R. & Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084-3095, doi:10.1021/ct400341p (2013).