Chemical and Cellular Defenses against Foreign Pathogens

Professor Thomas Walz Xiaonan Lin

Chemical and Cellular Defenses Against Foreign Pathogens

Abstract

Bacterial and viral infections affect billions of people each year. While bacterial infections are treated by the use of antibiotics, viral infections are eradicated by the immune system. This thesis comprises two parts. Part I presents the reconstitution of enzymes required for the synthesis of the minimal pharmacophore of moenomycin A (MmA), a molecule with antibacterial activity. Part II details single-particle electron microscopy studies of MDA5 alone and in complex with double-stranded RNA (dsRNA).

MmA is a natural product antibiotic from *Streptomyces ghanaensis* that possesses remarkable potency against clinically relevant Gram-positive bacteria. MmA exerts its antibacterial activity by binding directly to peptidoglycan glycosyltransferases, enzymes involved in the synthesis of the glycan strands of peptidoglycan. The genes responsible for MmA biosynthesis have been identified, and a complete biosynthetic pathway has been proposed. Part I of this thesis describes the reconstitution of enzymes required for the synthesis of two trisaccharide scaffolds of MmA that retain antibacterial activity. It also describes the synthesis of unnatural phosphoglycerate lipid acceptors and UDP-amino sugars that can be used to probe the substrate tolerances of key glycosyltransferases in MmA biosynthesis. This work lays the foundation for the

synthesis of unnatural MmA analogs that may possess better pharmacokinetic properties than the parent molecule.

MDA5 is a helicase that detects viral dsRNA in the cytoplasm of vertebrate cells. Upon dsRNA recognition, MDA5 initiates a series of signal transduction events that activate the innate immune response. Part II of this thesis presents the structures of apo MDA5 protein and the MDA5-dsRNA complex obtained by using single-particle electron microscopy. Two-dimensional averages of apo MDA5 revealed that the protein is very flexible and can adopt multiple conformations. This finding suggests that MDA5 does not adopt an autorepressed conformation in the absence of viral dsRNA. When MDA5 is incubated with dsRNA, the protein assembles onto the dsRNA to form a linear oligomer. Two-dimensional averages and a three-dimensional reconstruction reveal the complex to consist of seven to eight stacked discs per strand of 112 base pair dsRNA. This work lays the foundation for further structural studies aimed at elucidating the mechanism by which MDA5 is activated.

Table of Contents

List of Figures

List of Tables

Glossary of Abbreviations

1D	1-dimensional
2D	2-dimensional
3D	3-dimensional
5'-ppp	5'-triphosphorylated
ADP.AlF_4	ADP-aluminum fluoride
ADPCP	β,γ-methyleneadenosine 5′-triphosphate
BMH	Bismaleimide
Bp	Base pair
CARD	Caspase activation and recruitment domain
Ch	Chinovosamine
CTD	C-terminal domain
DEAD	Asp-Glu-Ala-Asp
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
ECD	Ectodomain
E. coli	Escherichia coli
EM	Electron microscopy
EMCV	Encephalomyocarditis virus
EMSA	Electrophoretic mobility shift assay
ESI-MA	Electrospray ionization mass spectrometry
Hel	Helicase domain of RIG-I

Hel2i	Helicase insertion domain of RIG-I
IFN	Interferon
IKK-ε	I-kappa B kinase-epsilon
IRES	Internal ribosome entry sites
IRF	Interferon regulatory factor
K_d	Dissociation constant
LCMS	Liquid chromatography mass spectrometry
LRR	Leucine rich repeat motif
MALS	Multiangle light scattering
MAVS	Mitochondria antiviral-signaling protein
MDA5	Melanoma differentiation-associated protein 5
MEFs	Mouse embryonic fibroblasts
MmA	Moenomycin A
MRSA	Methicillin-resistant S. aureus
MV	Mengo virus
MyD88	Myeloid differentiation primary response gene 88
NAP1	NAK-associated protein 1
OG	n-octyl-β-D-glucopyranoside
PAMPs	Pathogen-associated molecular patterns
PGT	Peptidoglycan glycosyltransferase
Poly (I:C)	Polyriboinosinic-polyribocytidylic acid
RecA	Recombinase A
RIG-I	Retinoic acid inducible gene-I

RIP I	Receptor interacting protein 1
RLR	RIG-I-like receptor
SAR	Structure-activity relationship
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ssRNA	Singlel-stranded RNA
S. aureus	Staphylococcus aureus
S. ghanaensis	Streptomyces ghanaensis
S. lividans	Streptomyces lividans
SNP	Single nucleotide polymorphism
T1D	Type 1 diabetes
TBk1	TANK-binding kinase-1
TG	Transglycosylase
TLR	Toll-like receptor
TP	Transpeptidase
TRAF3	Tumor necrosis factor receptor-associated 3
TRIF	TIR-domain-containing adaptor-inducing interferon
Trx	Thioredoxin
UDP-GlcNAc	Uridine diphosphate N-acetyl glucosamine
UDP-GalUA	Uridine diphosphate galacturonic acid

UDP-GlcUA	Uridine diphosphate glucuronic acid
VRE	Vancomycin-resistant enterococci
VRSA	Vancomycin-resistant MRSA
WT	Wild type

Acknowledgements

My thesis work would not have been possible without the help of many people.

First and foremost, I would like to thank my advisor, Professor Thomas Walz, for all the support and advice that he has given me during my time in his laboratory. I have received invaluable guidance from him both as a scientist and person.

I would like to thank Professors Suzanne Walker and Daniel Kahne for giving me the opportunity to work in their laboratories during the first three years of my graduate career. I would also like to thank my committee members. Professor Christopher Walsh has offered support and intellectual advice on all of the projects that are described in the text. Professor Tom Rapoport has offered his thoughts and ideas on numerous projects that I describe. Professor Alan Saghatelian has always been helpful with mass spectrometry trouble-shooting and has generously agreed to serve on my dissertation committee at the last minute.

The projects described in the text were done in collaboration with many people. I would like to thank Dr. Bohdan Ostash, Dr. Emma Doud, Dr. Deborah Perlstein, Dr. Shinichiro Fuse, Dr. Manuel Wolpert, and Dr. Hirokazu Tsukamoto for all the help and advice that they have given me on the moenomycin project. I would also like to thank Professor Sun Hur, Dr. Alys Peisley, and Dr. Bin Wu. The MDA5 project would not have been possible without them.

I am indebted to the members of the Walz lab for their support and camaraderie. I would like to thank Dr. Danijeal Dukovski for her assistance with anything and everything related to the lab, big or small. I would also like to thank Dr. Zongli Li who has patiently taught me how to prepare EM samples and use the microscopes. Many post-doctoral associates in the lab have offered me valuable guidance, in particular, Dr. Po-Lin Chiu, Dr. Andreas Schenk, and Dr. Debbie Kelly.

Chapter 1: Introduction to moenomycin and MDA5

Bacterial and viral infections affect billions of people every year. While bacterial infections are treated by the administration of antibiotics, most viral infections can be eradicated by the innate immune response (Dasgupta, 2012). This thesis consists of two parts, the first one describing efforts towards the development of new antibiotics and the second one describing research to advance our understanding of the innate immune system.

Part I of this thesis (Chapters 2-4) is concerned with the characterization of the biosynthetic pathway of moenomycin A (MmA) and the involved enzymes. MmA, a natural product produced by *Streptomyces ghanaensis*, exhibits potent antibacterial activity against a diverse range of gram-positive bacteria (Wallhausser et al, 1965; Slusarchyk et al, 1971). Interest in MmA has surged in the past 25 years as the rapid development of resistance to existing antibiotics prompted a renewed search for novel molecules with antibacterial potency (Bren, 2002; Halliday et al, 2006). MmA harbors great promise, because it prevents the synthesis of peptidoglycan, a structure that is essential for bacterial survival and growth (Bugg, 1999; Halliday et al, 2006). MmA also represents a novel structural class of antibacterials that poses no risk towards the development of cross-resistance from existing antibiotics (Halliday et al, 2006). Even though it exhibits impressive potency, MmA has not been developed for use in humans, because the molecule has poor pharmacokinetic properties (Wasielewski et al, 1965). Chapter 2 of this thesis provides a general introduction to MmA and summarizes major research findings. Chapter 3 presents genetic experiments

performed to characterize the biosynthetic pathway of MmA. Chapter 4 describes efforts toward the reconstitution of enzymes required for the synthesis of the minimal pharmacophore of MmA, and also lays out initial research performed to produce novel analogs of MmA. Taken together, the research described in the first part of this thesis is aimed at developing a rapid method for the chemenzymatic synthesis of MmA analogs with improved pharmacokinetic properties.

Part II of this thesis (Chapters 5-7) is concerned with the biochemical and structural characterization of MDA5, an enzyme that detects non-self, double-stranded RNA (dsRNA). dsRNA is the genetic material of many RNA viruses and elicits a powerful innate immune response in vertebrate organisms (Wang et al, 2011). Effective immune defense depends on the rapid recognition of viral dsRNA by pattern recognition receptors, of which there are two major classes. The toll-like receptors (TLRs) are integral membrane proteins that detect phagocytosed viral components in the extracelluar matrix (Schulz et al, 2005). The RIG-I-like receptors, RIG-I and MDA5, are helicases that detect foreign dsRNA in the cytoplasm of cells (Wilkins et al, 2010). Upon activation, RIG-I and MDA5 interact with a common signaling adaptor protein, mitochondria antiviral-signaling protein (MAVS), which ultimately leads to the expression of type I interferons and proinflammatory cytokines (Rehwinkel et al, 2010). While much effort focused on understanding the mechanism and structure of RIG-I, little work has addressed MDA5. Chapter 5 introduces the innate immune system with an emphasis on our current knowledge of RIG-I and MDA5, and Chapters 6 and 7

present biochemical and structural studies of the mechanism that underlies MDA5 activation.

References

Bren, L. (2002) Battle of the bugs: fighting antibiotic resistance. *FDA Consum* 36:28-34.

Bugg, T. Bacterial peptidoglycan biosynthesis and its inhibition. In Comprehensive Natural Products Chemistry, Pinto, B., Ed Elsevier: Oxford, 1999, 3, 241-294.

Dasgupta, A. (2012) Advances in antibiotic measurement. *Adv Clin Chem* 56:75-104.

Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P., and Meutermans, W. (2006) Targeting the forgotten transglycosylases. *Biochem Pharmacol* 71:957-967.

Rehwinkel, J., and Reis e Sousa, C. (2010) RIGorous detection: exposing virus through RNA sensing. *Science* 327:284-286.

Schulz, O., Diebold, S., Chen, M., Naslund, T., Nolte, M., Alexopoulou, L., Azuma, Y., Flavell, R., Lilijestrom, P., and Sousa R. (2005) Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature* 433:887-892.

Slusarchyk, W. (1971) Chemical and biological aspects of a new family of phosphorus-containing antibiotics. *Biotechnol Bioeng* 13:399-407.

Wallhausser, K., Nesemann, G., Prave, P., and Steigler, A. (1965) Moenomycin, a new antibiotic: Fermentation and Isolation. *Antimicrob Agents Chemother* 5:734-736.

Wasielewski, E., Muschaweck, R., and Schutze, E. Moenomycin, a new antibiotic. *Antimicrob Agents Chemother* 5:743-748.

Wang, Y., Swiecki, M., McCartney, S., and Colonna, M. dsRNA sensors and plasmacytoid dendritic cells in host defense and autoimmunity. *Immunol Rev* 243:74-90.

Wilkins, C., and Gale, M, Jr. (2010) Recognition of viruses by cytoplasmic sensors. *Curr Opin Immunol* 22:41-47.

Chapter 2: Introduction to moenomycin A and the peptidoglycan glycosyltransferases

2.1: The problem of antibiotic resistance

The discovery of antibiotics in the early 20th century revolutionized the medical sciences and prolonged human life. Before this discovery, patients with any form of bacteremia had very little chance of survival. In 1970, the Surgeon General of the United States declared that infectious diseases were no longer a major health threat and stipulated research efforts to be redirected to emerging frontiers of health issues (Breithaupt, 1999). Since then, however, resistance to available antibiotics has emerged, and the search for novel antibiotics has been reasserted as a pressing medical need. Today, five to ten percent of hospital patients develop bacterial infections, and 70 percent of the causative pathogens exhibit resistance to at least one of the antibiotics in current use (Bren, 2002).

Amongst all bacteremia-causing pathogens, *Staphylococcus aureus* has emerged as the leading cause of life-threatening infections in hospitals and the community (Gould, 2006; Taubes, 2008). *S. aureus* is highly virulent and can spread simply by contact (Boucher et al, 2008; Stryjewski et al, 2008). The bacteria also infect burn and surgical wounds and thus are endemic in hospitals. *S. aureus* has the propensity to develop resistance towards antibiotics, however, and methicillin-resistant *S. aureus* (MRSA) appeared in the United States in 1981. Since then, invasive MRSA infection has become widespread in hospitals and

accounts for more deaths in the United States than that from AIDS, tuberculosis and viral hepatitis combined (Klevens et al, 2007; Boucher, 2008).

For a period of time, MRSA infections could be treated with the glycopeptide vancomycin, commonly known as the last line of defense against MRSA. In 2002, however, a high level vancomycin-resistant MRSA (VRSA) strain was isolated from a dialysis patient in Michigan (Weigel et al, 2003). This VRSA strain harbored a plasmid that encoded resistance genes for vancomycin, trimethoprim, aminoglycosides, beta-lactams, and disinfectants (Ferber, 2003; Weigel et al, 2003). It is believed that this MRSA strain has acquired vancomycin resistance from vancomycin-resistant enterococci, which have become common in the 1980s. Since then, a handful of additional cases of VRSA have been documented (Bozdogan et al, 2003; Lowy, 2003; Pfeltz et al, 2004; Walsh et al, 2004). Although several of these cases involve common features (feet wounds), they are believed to be epidemiologically unrelated events. The emergence and spread of VRSA in the community is now inevitable, and the development of antibiotics to treat these infections is a high priority (Coates et al, 2002; Brown et al, 2005).

The introduction of linezolid, cubicin (daptomycin), and tigecycline into clinical use affords some short-term security against resistant pathogens. The oxazolidinone linezolid was approved in 2000 for the treatment of resistant bacterial infections, including those caused by MRSA and vancomycin-resistant enterococci (VRE) (Clemett et al, 2000; Brickner et al, 2008). However, resistance to linezolid has already developed, although the frequency of such

occurrences remains low (Tsiodras et al, 2001; Meka et al, 2004a; Meka et al, 2004b). Cubicin is a natural product antibiotic that was introduced in 2003 to treat skin and soft tissue infections caused by MRSA (Kirkpatrick et al, 2003). Like linezolid, resistance against cubicin developed quickly and a cubicin-resistant MRSA strain has already appeared in the clinic (Skiest et al, 2006). More recently, in 2005, the tetracycline derivative tigecycline was made available after fast-track approval by the FDA. However, since it is a structural analog of the tetracycline family of antibiotics, cross-resistance is expected to become an issue in the near-term future (Silver, 2003). The projected dearth of available treatments against resistant VRSA infections in the future makes the development of antibiotics that belong to novel structural classes or bind to new protein targets an urgent medical need.

2.2: Peptidoglycan biosynthesis: a validated target for antibiotic action

Peptidoglycan biosynthesis is a validated pathway for antibiotic action, because it is essential for bacterial growth and survival and is conserved even amongst organisms whose evolution diverged a billion years ago (Bugg, 1999). It is thus one of four metabolic pathways traditionally targeted by antibacterial drugs. The other three pathways are bacterial protein biosynthesis, DNA replication and repair, and folate coenzyme biosynthesis (Walsh, 2003).

Peptidoglycan is an essential, cross-linked polymer composed of carbohydrate and peptide units that prevent bacterial cells from lysing due to high internal osmotic pressures. It is biosynthesized in three stages. The first two stages

occur on the cytoplasmic side of the bacterial inner membrane and consist of the formation of Lipid II, a lipid-linked disaccharide pentapeptide molecule, from UDP-N-acetyl-glucosamine (UDP-GlcNAc) (Bugg et al, 1992; van Heijenoort, 2001). It is then flipped to the outside of the plasma membrane, where the disaccharide is polymerized into glycan chains by peptidoglycan glycosyltrasnferases (PGTs, also known as transglycosylases TG), and the pentapeptide is cross-linked by transpeptidases (TPs) (Koch, 2000; van Heijenoort, 2001). PGT and TP activities can either be mediated by two distinct proteins or sometimes by two separate domains of one bifunctional enzyme.

The formation of mature peptidoglycan from Lipid II on the extracellular side of the bacterial plasma membrane is a particularly good drug target. First, since the process is extracellular, drugs targeted at gram-positive pathogens will not have to cross any membrane barriers, allowing greater variability in the size and charge of the potential drug. This localization also allows the drug to evade efflux pumps that actively transport toxic substances from the cells' interior into the extracellular environment (Webber et al, 2003). Second, bacteria have multiple PGTs and TPs encoded by different genes (Zapun et al, 2012). The emergence of resistance against drugs that target these enzymes should thus be slow, since multiple spontaneous mutations must occur before all PGTs and TPs become unresponsive to the drug (Silver, 2007). Finally, since peptidoglycan is unique to bacterial cells, there is nearly no risk that a drug can elicit an autoimmune response by cross-reacting with a similar human enzyme.

The targeting of the final step of peptidoglycan biosynthesis by antibiotics is validated by the large number of antibacterials that block this step. Penicillin and other beta-lactams (cephalosporins, monobactams, and cabapenems) inhibit peptide cross-linking by binding to the transpeptidases (Essack, 2001; Fisher et al, 2005), vancomycin blocks mature peptidoglycan formation by sequestering Lipid II (Breukink et al, 2006), and finally, compounds of the moenomycin family bind to the PGTs, thus preventing the polymerization of glycan chains (Huber et al, 1968; Jackson et al, 1984).

2.3: Moenomycin A

Moenomycin A (MmA) is a phosphoglycolipid natural product produced by *Streptomyces ghanaensis*. It is one of more than ten structurally similar compounds isolated from the *Streptomyces* species in the 1960s (Wallhausser et al, 1965). MmA is composed of an isoprene-derived C25 lipid tail (moenocinol) attached through a phosphoglycerate linkage to a pentasaccharide that is covalently linked to a 1-amino-2,5-dioxocyclopentane moiety (Figure 2.1). Other phosphoglycolipid natural products have similar scaffold structures with minor differences in the lipid and pentasaccharide regions (Torikata et al, 1977).

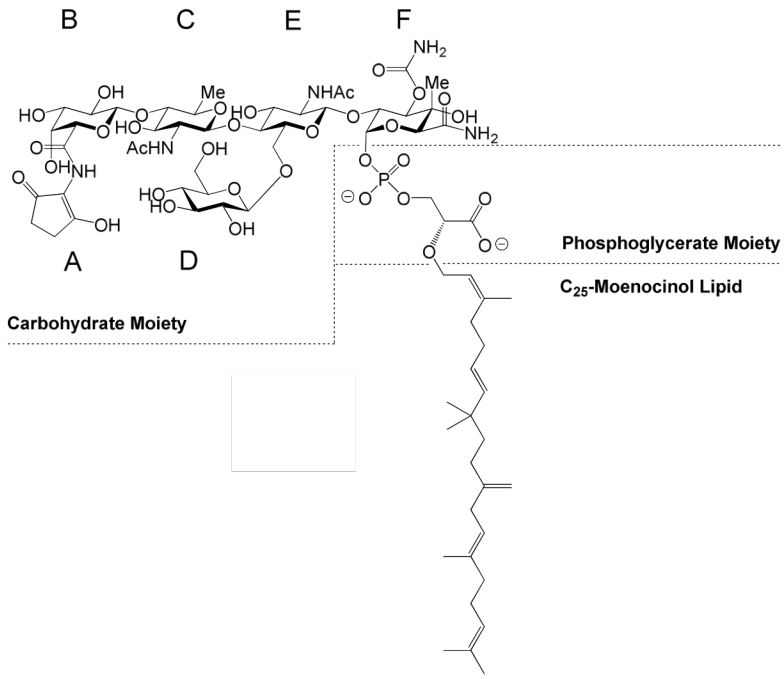

Figure 2.1: Chemical structure of moenomycin A.

2.4: Structure-activity relationship studies

MmA displays impressive activity against a broad spectrum of Gram-positive bacteria, including bacilli, staphylococci (and MRSA), streptococci, and diplococci. The minimum inhibitory concentration (MIC) values reported for MmA are 10-1000 lower than that for vancomycin (Slusarchyk et al, 1971; Halliday et al, 2006). Despite its potency, however, MmA has never been developed for human use because the molecule displays poor pharmacokinetic properties (Wasielewski et al, 1965; Halliday et al, 2006). It has, however, been

commercialized and used successfully for many years as a growth promoter in animal feed (Pfaller, 2006).

Interest in developing MmA as an antibacterial drug for human use has surged in the last 25 years. To improve the pharmacokinetic properties of the molecule, extensive structure-activity relationship (SAR) studies have been performed. Welzel and colleagues performed extensive synthesis and degradation studies of MmA to define the minimum active unit (El-Abadla et al, 1999; Welzel et al, 1984; Welzel et al, 1987). These studies concluded that the disaccharide unit consisting of the E and F glycan rings has full activity (when compared with MmA) as a PGT inhibitor. Unexpectedly, the disaccharide analog with an F-ring 4'-H replacing the 4'-methyl group exhibited full inhibitory activity as well (El-Abadla et al, 1999). This result indicated that the axial methyl group on the F-ring is not involved in eliciting binding energy. Welzel also showed that the carbamate at C3 of the F-ring, the acetyl group at C2 of the E-ring, and the phosphoglycerate moiety are required for inhibitory activity (1987). Since Welzel obtained disaccharide and trisaccharide products through chemical degradation, all analogs contained a fully saturated C25 lipid tail. This is due to the labile nature of the double bond, which Welzel hydrogenated in the first step of his degradation studies (1984). The role of the lipid structure in enzyme inhibition was not studied until 2006, when Kahne and coworkers designed a degradation scheme that enabled the detachment of the C25 moenocinol lipid and the reattachment of any synthetic lipid to the pentasaccharide core (Adachi et al, 2006). By doing so, they found that the enzyme inhibitory activity of neryl-MmA (MmA with a C10 neryl

lipid tail) is comparable to that of the parent molecule. However, the MIC of neryl-MmA against *S. aureus* and *E. facalis* are a few orders higher than for MmA (Adachi et al, 2006). This result indicates that 10 carbon units is the minimum lipid length required for full enzyme inhibition in the absence of membrane and detergent. The reason for why a longer lipid is required for biological activity may be that longer lipids enhance the ability of the molecule to partition into biological membranes (Marzian et al, 1994; Volke et al, 1997).

The extensive amount of SAR knowledge that has been gathered on MmA aids its development as an antibiotic for human use. Recent years have also seen the development of a strategy for the total synthesis of MmA, enabling further flexibility and variety in analog design (Taylor et al, 2006). The recent identification of the MmA gene cluster from *S. ghanaensis* opens yet another new avenue for accessing novel bioactive compounds that may be used in conjunction with chemical synthesis and degradation (Ostash et al, 2007). The next two chapters of this thesis discuss the characterization and initial reconstitution efforts of the MmA biosynthetic enzymes.

2.5: References

Adachi, M., Zhang, Y., Leimkuhler, C., Sun, B., LaTour, J., and Kahne, D. (2006) Degradation and reconstruction of moenomycin A and derivatives: dissecting the function of the isoprenoid chain. *J Am Chem Soc* 128:14012-14013.

Boucher, H., and Corey, G. (2008) Epidemiology of methicillin-resistant Staphylococcus aureus. *Clin Infect Dis* 46 Suppl 5:S344-S349.

Breithaupt, H. (1999) The new antibiotics. *Nat Biotechnol* 17:1165-1169.

Bren, L. (2002) Battle of the bugs: fighting antibiotic resistance. *FDA Consum* 36:28-34.

Breukink, E., and de Kruijiff, B. (2006) Lipid II as a target for antibiotics. *Nat Rev Drug Discov* 5:321-332.

Brown, E., and Wright, G. (2005) New targets and screening approaches in antimicrobial drug discovery. *Chem Rev* 105:759-774.

Bozdogan, B., Esel, D., Whitener, C., Browne, F., and Appelbaum, P. (2003) Antibacterial susceptibility of a vancomycin-resistant Staphylococcus aureus strain isolated at the Hershey Medical Center. *J Antimicrob Chemother* 52:864-868.

Brickner, S., Barbachyn, M., Hutchinson, D., and Manninen, P. (2008) Linezolid (ZYVOX), the first member of a completely new class of antibacterial agents for treatment of serious gram-positive infections. *J Med Chem* 51:1981-1990.

Bugg, T. (1999) Bacterial peptidoglycan biosynthesis and its inhibition. In Comprehensive Natural Products Chemistry, Pinto, B., Ed Elsevier: Oxford, 3, 241-294.

Bugg, T., and Walsh, C. (1992) Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: Enzymology, antibiotics, and antibiotic resistance. *Nat Prod Rep* 9:199-215.

Coates, A., Hu, Y., Bax, R., and Page, C. (2002) The future challenges facing the development of new antimicrobial drugs. *Nat Rev Drug Discov*, 1:895-910.

Clemett, D., and Markham, A. (2000) Linezolid. *Drugs* 50:815-828.

El-Abadla, N., Lampilas, M., Hennig, L., Findeisen, M., Welzel, P., Muller, D., Markus, A., and van Heijenoort, J. (1999) Moenomycin A: the role of the methyl group in the moenuronamide unit and a general discussion of structure-activity relationships. *Tetrahedron* 55:699-722.

Essack, S. (2001) The development of beta-lactam antibiotics in response to the evolution of beta-lactamases. *Pharm Res* 18:1391-1399.

Ferber, D. (2003) Triple-threat microbe gained powers from another bug. *Science* 302:1488.

Fisher, J., Meroueh, S., and Mobashery, S. (2005) Bacterial resistance to beta-lactam antibiotics: compelling opportunism, compelling opportunity. *Chem Rev* 105:395-424.

Gould, I. (2006) Costs of hospital-acquired methicillin-resistant Staphylococcus aureus (MRSA) and its control. *Int J Antimicrob Agents* 28:379-384.

Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P., and Meutermans, W. (2006) Targeting the forgotten transglycosylases. *Biochem Pharmacol* 71:957-967.

Huber, G., and Nesemann, G. (1968) Moenomycin, an inhibitor of cell wall synthesis. *Biochem Biophys Res Commun* 30:7-13.

Jackson, G., and Strominger, J. (1984) Synthesis of peptidoglycan by high molecular-weight penicillin-binding proteins of Bacillus-Subtilis and Bacillus-Stearothermophilus. *J Biol Chem* 259:1483-1490.

Kirkpatrick, P., Raja, A., LaBonte, J., and Lebbos, J. (2003) Daptomycin. *Nat Rev Drug Discov* 2:943-944.

Klevens, R., Morrison, M., Nadle, J., Petit, S., Gershman, K., Ray, S., Harrison, L., Lynfield, R., Dumyati, G., Townes, J., Craig, A., Zell, E., Fosheim, G., McDougal, L., Carey, R., and Fridkin, S. (2007) Invasive methicillin-resistant Staphylococcus aureus infections in the United States. *Jama* 298:1763-1771.

Koch, A. (2000) Penicillin binding proteins, beta-lactams, and lactamases: offensives, attacks, and defensive countermeasures. *Crit Rev Microbiol* 26:205-220.

Lowy, F. (2003) Antibicrobial resistance: the example of Staphylococcus aureus. *J Clin Invest* 111:1265-1273.

Marzian, S., Happel, M., Wagner, U., Muller, D., Welzel, P., Fehlhaber, H., Stark, A., Schutz, H., Markus, A., Limbert, M., van Heijenoort, Y., and van Heijenoort, J. (1994) Moenomycin A – reactions at the lipid part-new structure-activity relations. *Tetrahedron* 50:5299-5308.

Meka, V., and Gold, H. (2004a) Antimicrobial resistance to linezolid. *Clin Infect Dis* 39:1010-1015.

Meka, V., Pillai, S., Sakoulas, G., Wennersten, C., Venkataraman, L., DeGirolami, P., Eliopoulos, G., Moellering, R., and Gold, H. (2004b) Linezolid resistance in sequential Staphylococcus aureus isolates associated with a T2500A mutation in the 23S rRNA gene and loss of a single copy of rRNA. *J Infect Dis* 190:311-317.

Ostash B., Saghatelian, A., and Walker, S. (2007) A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chem Biol* 14:257-267.

Pfaller, M. (2006) Flavophospholipol use in animals: positive implications for antimicrobial resistance based on its microbiologic properties. *Diagn Microbiol Infect Dis* 56:115-121.

Pfeltz, R., and Wilkinson, B. (2004) The escalating challenge of vancomycin resistance in Staphylococcus aureus. *Curr Drug Targets Infect Disord* 4:273-294.

Silver, L. (2003) Novel inhibitors of bacterial cell wall synthesis. *Curr Opin Microbiol* 6:431-438.

Silver, L. (2007) Multi-targeting by monotherapeutic antibacterials. *Nat Rev Drug Discov* 6:41-55.

Skiest, D. (2006) Treatment failure resulting from resistance of Staphylococcus aureus to daptomycin. *J Clin Microbiol* 44:655-656.

Slusarchyk, W. (1971) Chemical and biological aspects of a new family of phosphorus-containing antibiotics. *Biotechnol Bioeng* 13:399-407.

Stryjewski, M., and Chambers, H. (2008) Skin and soft-tissue infections caused by community-acquired methicillin-resistant staphylococcus aureus. *Clin Infect Dis* 46 Suppl 5:S368-377.

Taubes, G. (2008) The bacteria fight back. *Science* 321:356-361.

Taylor, J., Li, X., Oberthur, M., Zhu, W., and Kahne, D. (2006) The total synthesis of moenomycin A. *J Am Chem Soc* 128:15084-15085.

Torikata, A., Yoshikawa, H., Katayama, T., and Arai, M. (1977) Pholipomycin a new member of phosphoglycolipid antibiotics. *J Antibiot* 30:1060-1063.

Tsiodras, S., Gold, H., Sakoulas, G., Eliopoulos, G., Wennersten, C., Venkataraman, L., Moellering, R., and Ferraro, M. (2001) Linezolid resistance in a clinical isolate of Staphylococcus aureus. *Lancet* 358:207-208.

Van Heijenoort, J. (2011) Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* 11:25R-36R.

Volke, F., Waschipky, R., Pampel, A., Donnerstag, A., Lantzsch, G., Pfeiffer, H., Richter, W., Klose, G., and Welzel, P. (1997) Characterisation of antibiotic moenomycin A interaction with phospholipid model membranes. *Chem Physics Lipids* 85:115-123.

Wallhausser, K., Nesemann, G., Prave, P., and Steigler, A. (1965) Moenomycin, a new antibiotic: Fermentation and Isolation. *Antimicrob Agents Chemother* 5:734-736.

Walsh, C. (2003) Where will new antibiotics come from. *Nat Rev Microbiol* 1:65-70.

Walsh, F., and Amyes, S. (2004) Microbiology and drug resistance mechanisms of fully resistant pathogens. *Curr Opin Microbiol* 7:439-444.

Wasielewski, E., Muschaweck, R., and Schutze, E. (1965) Moenomycin, a new antibiotic. *Antimicrob Agents Chemother* 1965:743-748.

Webber, M., and Piddock, L. (2003) The importance of efflux pumps in bacterial antibiotic resistance. *J Antimicrob Chemother* 51:9-11.

Weigel, L., Clewell, D., Gill, S., Clark, N., McDougal, L., Flannagan, S., Kolonay, J., Shetty, J., Killgore, G., and Tenover, F. (2003) Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. *Science* 302:1569-1571.

Welzel, P., Kunisch, F., Kruggel, F., Stein, H., Ponty, A., and Duddeck, H. (1984) Stepwise degradation of moenomycin A. *Carbohydr Res* 126:C1-C5.

Welzel, P, Kunisch, F., Kruggel, F., Stein, H., Scherkenbeck, J., Hiltmann, A., Duddeck, H., Muller, D., Maggio, J., Fehlhaber, H., Seibert, G., Heijenoort, Y., and Heijenoort J. (1987) Moenomycin A: minimum structural requirements for biological activity. *Tetrahedron* 43:585-598.

Zapun, A., Noiclerc-Sayoye, M., Helassa, N. and Vernet, T. (2012) Peptidoglycan assembly machines: the biochemical evidence. *Microb Drug Resist* 18:1-5.

Chapter 3: Complete characterization of the seventeen step moenomycin biosynthetic pathway

This chapter is a reprint of the following publication: Ostash, B., Doud, E., Lin, C., Ostash, I., Perlstein, D., Fuse, S., Wolpert, M., Kahne, D., and Walker, S. (2009) Complete characterization of the seventeen step moenomycin biosynthetic pathway. *Biochem* 48:8830-8841. I contributed to this work by reconstituting MoeGT1 and MoeE5 and testing their biochemical activities. The supplemental information that accompanies this publication is not reprinted here due to its large size.

The moenomycins are phosphoglycolipid antibiotics produced by *Streptomyces ghanaensis* and related organisms. The phosphoglycolipids are the only known active site inhibitors of the peptidoglycan glycosyltransferases, an important family of enzymes involved in the biosynthesis of the bacterial cell wall. Although these natural products have exceptionally potent antibiotic activity, pharmacokinetic limitations have precluded their clinical use. We previously identified the moenomycin biosynthetic gene cluster in order to facilitate biosynthetic approaches to new derivatives. Here, we report a comprehensive set of genetic and enzymatic experiments that establish functions for the 17 moenomycin biosynthetic genes involved in the synthesis of moenomycin and variants. These studies reveal the order of assembly of the full molecular scaffold and define a subset of seven genes involved in the synthesis of bioactive analogues. This work will enable both *in vitro* and fermentation-based reconstitution of phosphoglycolipid scaffolds so that chemoenzymatic approaches to novel analogues can be explored.

3.1: Introduction

Peptidoglycan glycosyltransferases (PGTs) are a family of enzymes that assemble the glycan chains of the bacterial cell wall from disaccharide precursors (Goldman et al, 2000; Ostash et al, 2005). They are potentially important antibacterial targets because their functions are essential, their structures are highly conserved, they have no eukaryotic counterparts, and peptidoglycan biosynthesis is a well-validated pathway for antibiotics (Ostash et al, 2005; Halliday et al, 2006; Wright, 2007). Although there are not yet any clinically used antibiotics that target these enzymes, a family of natural products called the phosphoglycolipids inhibit the PGTs and have potent antibacterial activity against Gram-positive pathogens, including methicillin-resistant *Staphylococcus aureus* and both vancomycin-sensitive and vancomycin-resistant pathogens (Goldman et al, 2000; Ostash et al, 2005; Welzel, 2005).

Moenomycin A (**1**) (MmA, Figure 3.1a) is the prototypical member of the phosphoglycolipid family of antibiotics. It has been widely used as an animal growth promoter but was never developed for human use because it has poor physicochemical properties, including a very long half-life that is presumed to be related to the long lipid chain on the molecule (Adachi et al, 2006). It may be possible to develop phosphoglycolipid analogues that have better properties than MmA by altering particular structural features. Recently reported structures of co-complexes of PGTs with bound moenomycin and neryl-moenomycin provide information on the mode of binding and on critical protein-ligand contacts, which should facilitate efforts to design analogues (Lovering et al, 2007; Yuan et al,

2007; Lovering et al, 2008; Yuan et al, 2008; Heaslet et al, 2009; Sung et al, 2009). The challenge now is to develop approaches to make a wide range of analogues efficiently so that the potential of the phosphoglycolipid antibiotics can be explored (Goldman et al, 2000; Ostash et al, 2005; Welzel, 2005).

Efforts to synthesize MmA analogues go back many years, and the total synthesis of MmA was reported in 2006 (Goldman et al, 2000; Ostash et al, 2005; Welzel, 2005; Adachi et al, 2006; Taylor et al, 2006). Thus, fully synthetic approaches to make analogues are available. However, the total chemical synthesis of MmA analogues remains an inefficient process because of the structural complexity of the molecule (Adachi et al, 2006; Taylor et al, 2006). Biosynthetic or chemoenzymatic approaches provide an alternate, potentially more efficient means, to produce analogues. With this possibility in mind, we recently sequenced the producing organism, *Streptomyces ghanaensis*, in order to identify the genes involved in MmA biosynthesis (Ostash et al, 2007). We proposed functions for many of the biosynthetic genes based on sequence analysis, but experimental validation of these functions was not completed. Furthermore, specific functions were not assigned for the five putative glycosyltransferases that make the pentasaccharide, the direction of pentasaccharide assembly was not established, the order of the tailoring modifications was unknown, and the genes involved in the biosynthesis of biologically active scaffolds were not identified. Here, we report a comprehensive set of genetic and enzymatic experiments that illuminate the moenomycin biosynthetic pathway. This work lays the foundation for the development of

chemoenzymatic approaches to the synthesis of new PGT inhibitors.

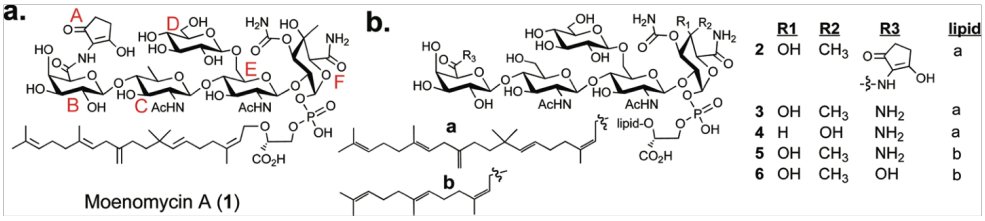

Figure 3.1: Key structures and DNA constructs. (a) Structure of moenomycin A (**1**, MmA). (b) Phosphoglycolipid derivatives of moenomycin: **2** (pholipomycin) produced by strain moeno38-1^{+}pOOB64b^{+}; **3** produced by cosmid moeno38-1; **4** produced by moeno38-1^{+}Δ*moeK5*; **5** and **6** produced by moeno38-1^{+}Δ*moeN5*.

3.2: Materials and methods

Antibiotics. Pure MmA was provided by M. Adachi (Dept. of Chemistry and Chemical Biology, Harvard University). The following concentrations of commercially available antibiotics were used for strain selection (μg/mL): ampicillin (100), carbenicillin (100), chloramphenicol (35), kanamycin (50), apramycin (50), hygromycin (100), spectinomycin (200), streptomycin (100), thiostrepton (50), and nalidixic acid (50) (Sigma-Aldrich). UDP-GlcUA was purchased from Sigma-Aldrich. Isoprenoid pyrophosphates were purchased from Isoprenoids, Lc.

Strains and vector DNAs. *Streptomyces ghanaensis* ATCC14672 and *Bacillus cereus* ATCC19637 were obtained from ATCC. *S. lividans* TK24 was kindly provided by M. Kobayashi (University of Tsukuba, Japan). The methylation-

deficient conjugative strain *E. coli* ET12567 (pUB307) (Flett et al, 1997) was obtained from Professor C. P. Smith (Manchester University, U.K.). *E. coli* BW25113 (pIJ790) was obtained from the John Innes Center (Norwich, U.K.). The *S. ghanaensis* strain MO12, which contains a disrupted *moeGT3* gene, and the *S. lividans* strains expressing various subsets of *moe* genes were constructed in this work. The vectors pKC1139, pSET152, pMKI9, and pOOB40 were obtained as described previously (Ostash et al, 2007). Plasmid pWQ67 was obtained from Professor C. Whitfield (University of Guelph, Canada) for the expression of epimerase GlaKP (Frirdich et al, 2005). The integrative vector pSOK804 (Sekurova et al, 2004) was a gift from S. Zotchev (Norwegian University of Science and Technology, Trondheim, Norway). The expression vector pAF1 (ori^{pIJ101} bla tsr, PermE*, 6His tag) was provided by A. Bechthold (Freiburg University, Germany). Plasmids pKD4 and pCP20 (Datsenko et al, 2000) were obtained from J. Beckwith (Harvard Medical School, USA). The spectinomycin resistance cassette pHP45 (BlondeletRouault et al, 1997) was obtained from J.-L. Pernodet (Universite Paris-Sud, France). Plasmid pTB146 was obtained from T.G. Bernhardt (Harvard Medical School).

Cloning of moeO5, moeE5, and moeGT1. The *moeO5* gene was initially cloned into the pAF1 vector for expression as a N-terminal His6 fusion in *S. lividans* (pED2). An *E. coli* codon optimized *moeO5* gene (*synO5*) was designed and synthesized by Mr.Gene. The gene was delivered in an ampicillin resistant plasmid named pMA_moeO5. *SynO5* was PCR amplified from pMA_moeO5

using primers synO5SUMOf (TTGGTAGAAGAGCAGTTAACGCTAGTCCTCAACTG) and synO5SUMOr (CCCGGGTCAACGGCCAGAACC). These primers introduced the respective restriction enzyme sites SapI and XmaI (underlined). PCR amplification was performed with KOD Hot Start polymerase (Novagen). The PCR product was gel purified, digested with SapI and XmaI, and ligated using T4 DNA ligase (NEB) into the linearized expression plasmid pTB146 to give the N-terminal SUMO/His6-tagged pTBsynO5.

MoeE5 was PCR amplified from cosmid moeno38-1 using primers MoeE5BamHI (AAAAAGGATCCGGTGTCGAGCGATACACACGG) and MoeE5XhoI (AAAAACTCGAGCTACAGCCGCGGCACGGAC), introducing BamHI and XhoI restriction sites, respectively (underlined). The PCR product was gel purified, digested with BamHI and XhoI, and ligated into linearized vector pET48b using T4 DNA ligase (NEB) to give N-terminal thioredoxin/His6-tagged pET48b-MoeE5.

MoeGT1 was PCR amplified from cosmid moeno38-1 using primers MoeGT1Nde1 (AAACATATGGCTGCCCCCGACCGAC) and MoeGT1NotI (ATAAAGCGGCCGCTCGGGCGTC), introducing NdeI and NotI restriction sites, respectively (underlined). The PCR product was gel purified, digested with NdeI and NotI, and ligated using T4 DNA ligase (NEB) into linearized vector pET24b to give N-terminal His6-tagged pET24b-MoeGT1.

Expression and purification of MoeO5, MoeE5, and MoeGT1. The pTBsynO5 expression construct was transformed into BL21(DE3) competent cells (Invitrogen) for protein overproduction. pET48b-MoeE5 and pET24b-MoeGT1 constructs were transformed into Rosetta2(DE3)pLysS (Novagen) competent cells for protein overproduction. Transformants harboring the desired constructs were grown at 37°C in LB containing 100 μg/mL carbenicillin (for pTBsynO5) or 50 μg/mL kanamycin and 35 μg/mL cloramphenicol (for pET48b-MoeE5 and pET24b-MoeGT1) to an OD_{600} of 0.6. The temperature was reduced to 16°C, and the cells were induced by the addition of isopropyl-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. After an additional 16 h at 16°C, the cells were harvested by centrifugation (20 min at 5000g) and frozen at -80°C.

pTBsynO5 cells were resuspended in buffer A (50 mM Tris-HCl at pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, and 3% CHAPS) and incubated at room temperature with rlysozyme, benzonase, and protease inhibitor complexes (Novagen) for 1 h. Cells were then lysed by sonication, and the cell debris was removed by centrifugation (30 min at 14,000g). The supernatant was diluted with 1.5 volumes buffer B (50 mM Tris-HCl at pH 7.5, 400 mM NaCl, and 5 mM $MgCl_2$) and rocked with TALON resin (Clontech) for 1 h at 4°C. The resin was collected and washed with buffer B supplemented with 5 mM imidazole by rocking for 30 min at 4°C, packed into a column, and eluted with 20-200 mM imidazole in buffer B. Fractions containing the target protein (identified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)) were pooled, desalted, concentrated, flash frozen in liquid nitrogen, and stored at -

80°C. The protein concentration was determined using the Dc Protein Assay (Biorad) using BSA as the standard. The yield of purified MoeO5 was approximately 4 mg/L.

pET48b-MoeE5 and pET24b-MoeGT1 cell pellets were resuspended in buffer C (50 mM Tris-HCl at pH 7.5, 400 mM NaCl, 5 mM MgCl2, and 1% CHAPS) and D (25 mM Tris-HCl at pH 8, 200 mM NaCl, 5% glycerol, 0.5% CHAPS, and 20 mM imidazole), respectively, supplemented with rlysozyme, benzonase, and protease inhibitor complex (Novagen). After 1 h at room temperature, the cells were lysed by sonication, and the cell debris was removed by centrifugation (30 min at 14,000g). The supernatant was incubated with Ni-NTA resin (Qiagen) for 30 min at 4°C. The recovered resin was washed with buffer C or D and packed into a column, and protein was eluted using a stepwise gradient of 5-500 mM imidazole. Fractions containing the target protein (identified by SDS-PAGE) were pooled, desalted, and concentrated. MoeE5 was flash frozen in liquid nitrogen in 20 mM Tris-HCl at pH 7.5, 150 mM NaCl, 10% glycerol and stored at -80°C. MoeGT1 was stored at 4°C in 20 mM Tris-HCl at pH 8, 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 10% glycerol. The protein concentration was determined using the Dc Protein Assay (Biorad) using BSA as the standard. The purified yield for MoeE5 was approximately 15 mg/L. The purified yield for MoeGT1 was approximately 4 mg/L.

Characterization of MoeO5. All LC/MS experiments were performed using a Phenomenex Gemini 5μ C18 100 Å column (50 mm x 4.6 mm) at a flow rate of 1

mL/min (solvent A: 95:5 water/methanol; solvent B, 60:35:5 isopropanol/methanol/water both with 0.1% ammonium hydroxide as a solvent modifier).

MoeO5 activity assays were performed in 50 mM Tris-HCl at pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$ with 2 mM 3-D-phosphoglyceric acid, and 1 μM MoeO5. Pyrophosphates tested were geranyl, farnesyl, geranylgeranyl, and moenocinyl pyrophosphate. Reactions were initiated with the addition of 40 μM of isoprenoid pyrophosphate. The reactions were incubated for 1 h at 37°C and quenched with an equal volume of methanol. The quenched reactions were centrifuged (5 min at 10,000g) to remove precipitated protein. Reactions were analyzed on an Agilent 1200 LC/MS (conditions described above) with a linear gradient over 10 min. For the farnesyl pyrophosphate reaction, no detectable FPP starting material was observed following a 1 h incubation. Reactions were performed with MoeO5 purified from *S. lividans* and with Sumo-MoeO5 purified from *E. coli*, and the results were identical.

For NMR analysis, compound **7** was scaled up and purified by C18 column (Extract-Clean SPE C18 column (Alltech), eluted in methanol) followed by preparative TLC (solvent system: 6:4:0.85 chloroform/methanol/water, run and dried three times). ^1H resonances for compound **7** were assigned from one-dimensional (1D) and two-dimensional (2D) COSY spectra based on reported data. Multiplicities are reported by using the following abbreviations: s = singlet, d = doublet, t = triplet, m = multiplet, and br = broad. δ_H (500 MHz, D_2O) 1.618 (s, 3H); 1.625 (s, 3H); 1.693 (s, 3H); 1.774 (s, 3H); 2.007-2.035 (m, 2H); 2.094-

2.157 (m, 6H); 3.858 (br s, 1H); 3.956-4.038 (m, 3H); 4.126 (br s, 1H); 5.181 (br s, 2H); 5.415 (s, 1H). The stereochemistry of the allylic bond was confirmed to be cis based on 1D NOESYs.

Characterization of MoeE5. Reconstitution of MoeE5 activity was performed in reactions containing 20 mM Tris-HCl at pH 7.5 and 1 mM UDP-GlcUA. The reactions were initiated by the addition of 1 μM MoeE5, incubated at 37°C for 2 h, and quenched by boiling for 5 min followed by centrifugation (5 min at 10,000g) to remove precipitated protein. Reactions were analyzed by anion-exchange HPLC (Phenomenex Phenosphere SAX 5 μm, 100 Å column, 250 mm 4.6 mm) on an Agilent 1100 HPLC over a 30 min salt gradient flow rate of 1 mL/min (solvent A: 5 mM $NH_4H_2PO_4$ at pH 4.5; solvent B: 750 mM $NH_4H_2PO_4$ at pH 3.7) (Frirdich et al, 2005).

Characterization of MoeGT1. Reconstitution of MoeGT1 was carried out in reactions with 50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM UDP-sugar (UDP-GlcUA, UDP-GalUA, or an epimeric mixture of UDP-GlcUA and UDP- GalUA), and 40 μM **7**. Reactions were initiated by the addition of 800 nM MoeGT1, incubated at 37°C, quenched at specific time points by the addition of 1 volume methanol (see Figure 3.5 for time points), and centrifuged (5 min at 10,000g) to remove precipitated protein. Reactions were analyzed on an Agilent 6200 LCMS (conditions described above for MoeO5) with a linear gradient over 10 min. Product and substrate peaks were integrated, normalized, and used to

calculate percent conversion in μmols.

Generation of recombinant S. lividans and S. ghanaensis strains. Standard molecular biology procedures were used throughout the work (Kieser et al, 2000; Sambrook et al, 2001). For all λ-RED-assisted deletions of *moe* genes within cosmid moeno38-1 (except for *moeGT3*), the entire open reading frame was replaced with a kanamycin resistance cassette (pKD4). The mutated cosmid was introduced into strain DH5α (pCP20) to evict *kanR* as described (Datsenko et al, 2000). Deletions within the cosmids were confirmed by PCR. To create the Δ*moeGT5*Δ*moeGT3* strain, λ-RED recombination was used to replace *moeGT3* with the disrupted allele *moeGT3::aadA* in the Δ*moeGT5* derivative of moeno38-1. All constructs were transferred into *S. lividans* via intergeneric conjugation. Integration of moeno38-1 and its derivatives into the *S. lividans* genome was confirmed as described (Ostash et al, 2007). Gene *moeGT3* was insertionally inactivated in the *S. ghanaensis* genome according to an established procedure (Ostash et al, 2007). Plasmid borne copies of the genes were used to complement the gene deletions.

Moenomycin production and analysis. Small-scale fermentation and purification of moenomycins was performed as described previously (Ostash et al, 2007). To obtain pure moenomycin intermediates (>90% as judged by TLC) from recombinant *S. lividans* strains, the following procedure was used. TSB medium (30 mL) in a 250 mL flask containing 70 glass beads (5 mm) was inoculated with

100 μL (approximately 10^4-10^5 cfu) of stock culture (kept in 10.3% sucrose at -20°C). The flask was incubated on an orbital shaker (240 rpm) for 2 days at 37°C and then used as a preculture to start the fermentation. Slightly modified R5 medium (Kieser et al, 2000) (sucrose: 6% instead of 10.3%; 1 mg/L $CoCl_2$ added after autoclaving) was used as a fermentation medium. Eight 4 L flasks (500 mL of medium per flask) containing beads were grown for 6 days at 37°C. Using the floor shaker in a warm room fixed at 37°C provided maximized aeration and optimized antibiotic production compared to smaller incubated shakers at 30°C. The mycelium was collected by centrifugation and extracted exhaustively with methanol-water (9:1) at 37°C (when necessary, the pH of the extraction mixture was adjusted to 7 - 7.5 with Tris-HCl). The extract was concentrated, reconstituted in water, and extracted with dichloromethane. The aqueous phase was loaded on a XAD-16 column (30x400 mm), washed with water (300 mL), and eluted with methanol (500 mL). Methanol fractions containing the desired compound were combined, concentrated and purified on a Sep-Pak C18 SPE cartridge (Waters) as described (Eichhorn et al, 2005). Further silica gel flash chromatography or preparative TLC of the extract was performed according to Adachi et al. (2006) and yielded pure compound (0.025-0.1 mg/L, depending on the strain). Antibiotic disk diffusion assays, LC-MS, MS/MS, and determination of accurate mass spectra of moenomycins were carried out as described previously (Ostash et al, 2007). MIC values (μg/mL) were obtained using a standard microdilution assay. The MIC is defined as the lowest antibiotic concentration that resulted in no visible growth after incubation at 37°C for 22 h.

^{1}H NMR spectrum of compound **20** was recorded on a Varian Inova 500 (500 MHz) instrument in D_2O (4.80 ppm).

3.3: Results

Description of the approach. The MmA biosynthetic genes are located in two clusters on the *S. ghanaensis* chromosome: a three gene operon involved in A ring assembly (*moe* cluster 2) and a larger cluster containing the genes involved in the assembly of the phosphoglycolipid pentasaccharide scaffold (*moe* cluster 1) (Adachi et al, 2006; Taylor et al, 2006; Ostash et al, 2007). We have now reconstituted the biosynthesis of MmA in *S. lividans* using a three-component heterologous expression system. The three components include the previously described integrative cosmid moeno38-1, which contains all but two of the genes of *moe* cluster 1, the replicative plasmid pOOB49f, which supplies the two missing *moe* cluster 1 genes, and the integrative plasmid pOOB64b, which contains the three genes in *moe* cluster 2 (Figure 3.2) (Ostash et al, 2007). To probe the functions of the putative biosynthetic genes, we have constructed a series of derivatives carrying λ-RED-induced single or double gene deletions in cosmid 38-1 or other components of the heterologous expression system (Table 3.1) and have prepared recombinant *S. lividans* strains expressing the subsets of genes. One gene deletion (*moeGT3*) was prepared in the producing organism, *S. ghanaensis*. We then analyzed the phosphoglycolipids produced by the resulting strains. Because production levels in the heterologous host are low, we made use of exact mass data and previously established fragmentation patterns of the

moenomycins to assign structures. MS^2 fragmentation patterns of phosphoglycolipids from fermentation have been extremely well characterized by both Eichhorn et al. (2005) and Zehl et al. (2006), allowing us to confirm the structures of our compounds. One example is shown in Figure 3.3. Proton NMR spectra were used to confirm assignments in cases, in which sufficient material was obtained (compounds **7** and **20**). To address polar effects, deletion strains were complemented with plasmid-borne copies of the genes. When genetic approaches could not reveal function (with MoeO5, MoeGT1, and MoeE5), we used biochemical reconstitution of enzymatic activity *in vitro* to establish substrates and products.

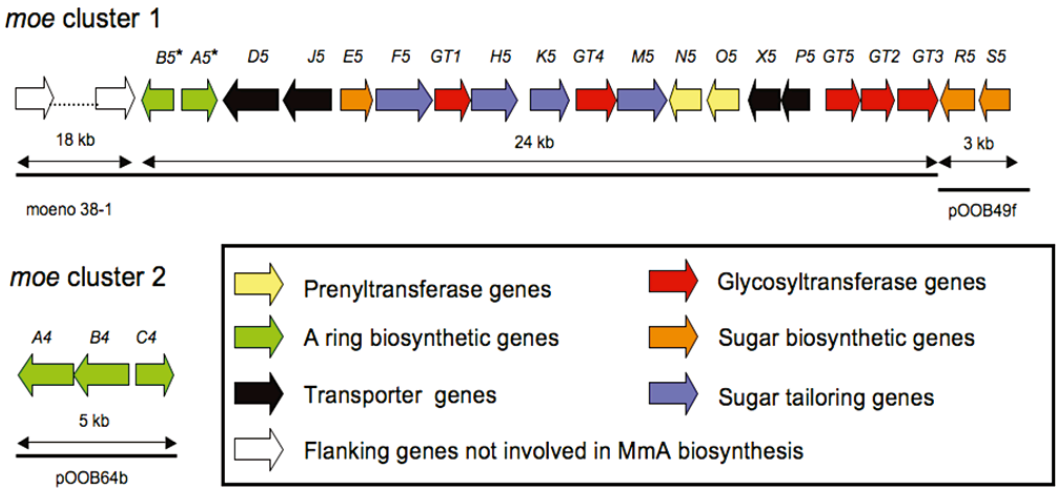

Figure 3.2: *Moe* clusters 1 & 2. Three parental DNA constructs were used to generate all other recombinant cosmids and plasmids described in this study. Cosmid moeno38-1 carries all of *moe* cluster 1 except *moeR5moeS5*. Plasmid pOOB49f carries *moeR5moeS5* under the control of the *ermE* promoter. The three genes in *moe* cluster 2 are carried on plasmid pOOB64b.

Table 3.1: Data for phosphoglycolipids from recombinant strains and *in vitro* reconstitution

Source	Compound	Retention time (min)	m/z [M - H]- Calculated	Observed
38-1+Δ*moeA5*Δ*B5*	**3**	9.9[b]	1500.6273	1500.6272
38-1+Δ*moeF5*	**8**	3.9[b]	565.2050	565.2055
38-1+Δ*moeGT4*	**9**[a]	4.7[b]	564.2210	564.2221
38-1+Δmoe*GT5*Δ*GT3*	**11**	4.8[b]	781.3160	781.3147
38-1+Δmoe*GT5*	**20**	10.4[b]	1122.4998	1122.5005
38-1+Δmoe*GT2*	**22**	10[b]	1325.5792	1325.5786
38-1+*moeR5*+Δ*moeH5*	**23**	9.4[b]	1485.6164	1485.6199
38-1+Δ*moeH5*	**24**	9.3[b]	1501.6113	1501.6113
38-1+pBOO64b+Δ*moeH5*	**24**	9.3[b]	1501.6113	1501.6139
38-1+*moeR5*+Δ*moeA5*Δ*B5*	**25**	10[b]	1484.6324	1484.6329
38-1+pBOO64b+Δ*moeA5*Δ*B5*	**2**	9.3[b]	1596.6490	1596.6439
38-1+Δ*moeK5*	**4**	9.6[b]	1486.6116	1486.6118
38-1+Δ*moeN5*	**5**	4.1[b]	1365.4861	1365.4867
	6	4.2[b]	1364.5021	1364.5023
S. ghanaensis MO12	**26**[d]	9.2[b]	1418.6007	1418.6016
MoeO5 in vitro	**7**	6.5[c]	389.1729	389.1737
MoeE5 in vitro	UDP-GalUA	5.6	579.2793	579.279
	UDP-GlcUA	16.0		
MoeGT1 in vitro	**8**	6.6[c]	565.2050	565.2066

[a]The C4 equatorial epimer of **9** cannot be excluded, but it is not consistent with the reported desmethyl MmA analogue obtained via fermentation nor with evidence that UDP-GalUA is the F ring sugar donor. [b]Column: 250 x 4.6mm Agilent C_{18} column; conditions are as previously reported (Ostash et al, 2007). [c]Column: 100 x 4.6 mm Gemini C_{18} column; HPLC conditions are described in Materials and Methods.

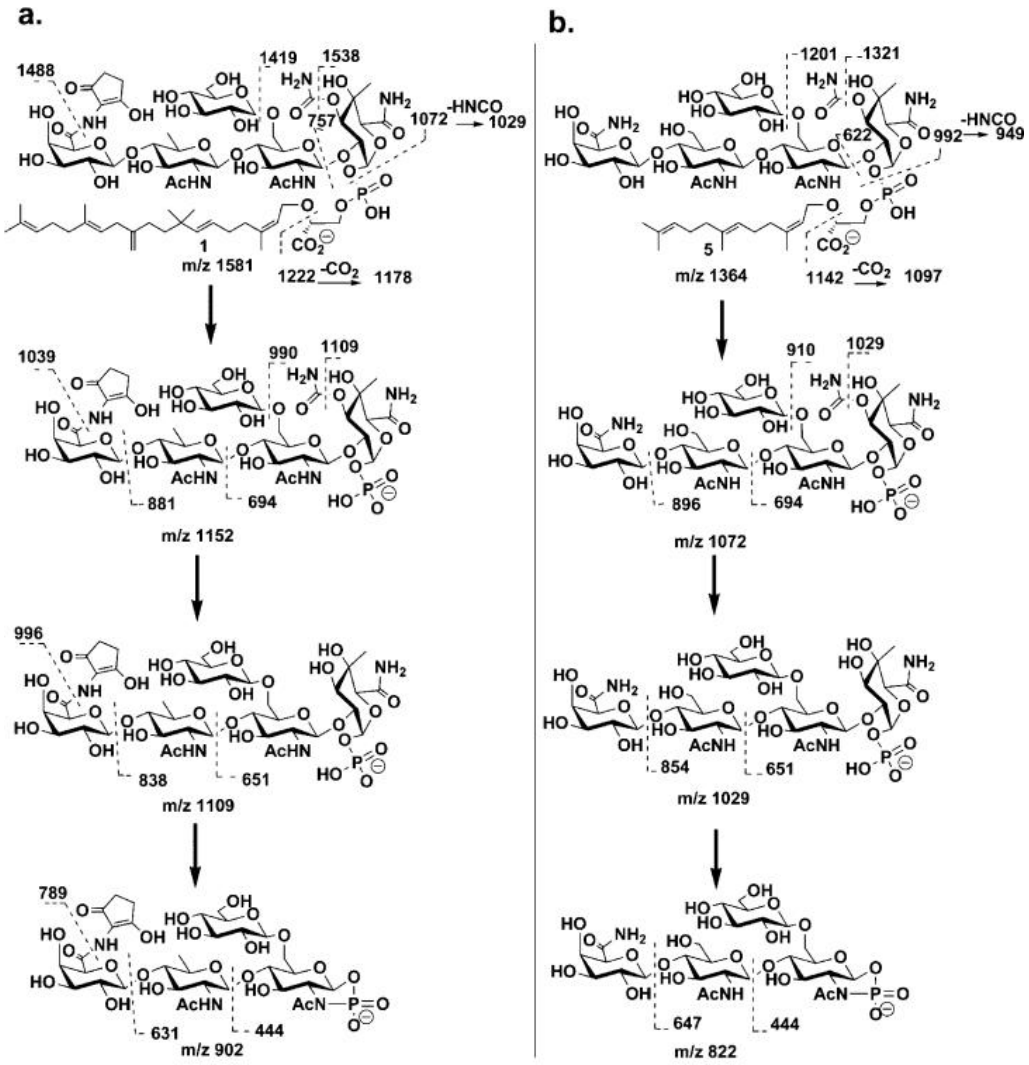

Figure 3.3: Fragmentation patters of moenomycin A. Observed fragmentation patterns of (a) moenomycin A (m/z observed, 1580.6430; calculated, 1580.6535) and (b) compound 5 (m/z observed, 1364.5023; calcd, 1364.5021), which lacks the C5Nunit (A ring), contains farnesyl (C15) rather than moenocinyl (C25) lipid, and has a GlcNAc rather than a chinovosamine C ring for a mass difference of 217. The fragment ions observed for moenomycin A are assigned on the basis of published data by Eichorn et al. (2005). The corresponding fragment ions are observed for compound **5**.

Functions of the prenyltransferases MoeO5 and MoeN5. Moenomycin contains an irregular isoprenoid chain attached via a cis-allylic ether linkage to phosphoglycerate. On the basis of *in silico* analysis of *moe* cluster 1, we previously proposed that *moeO5* and *moeN5*, two putative prenyl transferases, are involved in the assembly of the moenocinyl chain (Ostash et al, 2007). MoeO5 is a predicted TIM-barrel protein that shows homology to the prenyl transferases that transfer geranylgeranyl to glycerol phosphate in the first step of archael membrane lipid biosynthesis (Zhang et al, 1993; Payandeh et al, 2006). On the basis of this resemblance, we suggested that MoeO5 forms an ether linkage between D-3-phosphoglyceric acid and an activated moenocinyl chain (Ostash et al, 2007). To test this hypothesis, we cloned and expressed MoeO5 in both *S. lividans* and *E. coli*. After purification, MoeO5 was incubated at 37°C with 2 mM D-3-phosphoglyceric acid and 40 μM of moenocinyl pyrophosphate. No product was observed even after an overnight incubation. MoeO5 was then incubated with one of several commercially available prenyl pyrophosphates, including geranyl, farnesyl, and geranylgeranyl pyrophosphate. MoeO5 completely converted farnesyl pyrophosphate to compound **7** within 1 h. The product was purified and shown by NMR to contain a cis-allylic ether bond like the natural product (Figure 3.4, Table 3.1). It is worth noting that double bond isomerization, although common in prenyl cyclase reactions, has not previously been observed for a prenyl transferase that catalyzes an intermolecular coupling (Dewick, 2002; Liang et al, 2002; Domingo et al, 2009). Furthermore, it does not occur in the reactions catalyzed by the archaeal homologues of MoeO5 (Zhang et al, 1993; Payandeh et

al, 2006). This and other differences in the substrates and products of the archaeal enzymes compared with MoeO5 suggest that TIM-barrel prenyl transferases, a largely unexplored family of enzymes, may display considerable variability in the types of reactions they catalyze (Wierenga, 2001; Nagano et al, 2002; Anantharaman et al, 2003; Caetano-Anolles et al, 2009).

Since MoeO5 generates a C15 lipid-phosphoglycerate while MmA contains a C25 lipid, we surmised that the other prenyltransferase, MoeN5, must couple the C10 lipid from a pyrophosphate donor to a z-farnesylated intermediate in the biosynthetic pathway. Efforts to reconstitute the activity of MoeN5 using starter unit **7** and geranyl pyrophosphate as substrates were unsuccessful. However, we found that strain 38-1+Δ*moeN5* accumulates compounds **5** and **6**, which contain farnesyl instead of moenocinyl moieties (Figure 3.1b and Table 3.1). Isolation of **5** and **6** from the 38-1+Δ*moeN5* strain confirms that MoeN5 catalyzes C10 prenyl transfer, presumably using geranyl pyrophosphate from primary metabolism as a substrate. A mechanism for the unusual rearrangement that produces the moenocinyl chain was proposed by Arigoni, Welzel, and co-workers (Neundorf et al, 2003), and this mechanism is consistent with the finding that the C25 chain is constructed via C10 prenyl transfer to a C15-containing intermediate. In cells, MoeN5 apparently only acts on glycosylated intermediates (see below), a finding that may explain our inability to reconstitute its activity using **7** as a substrate.

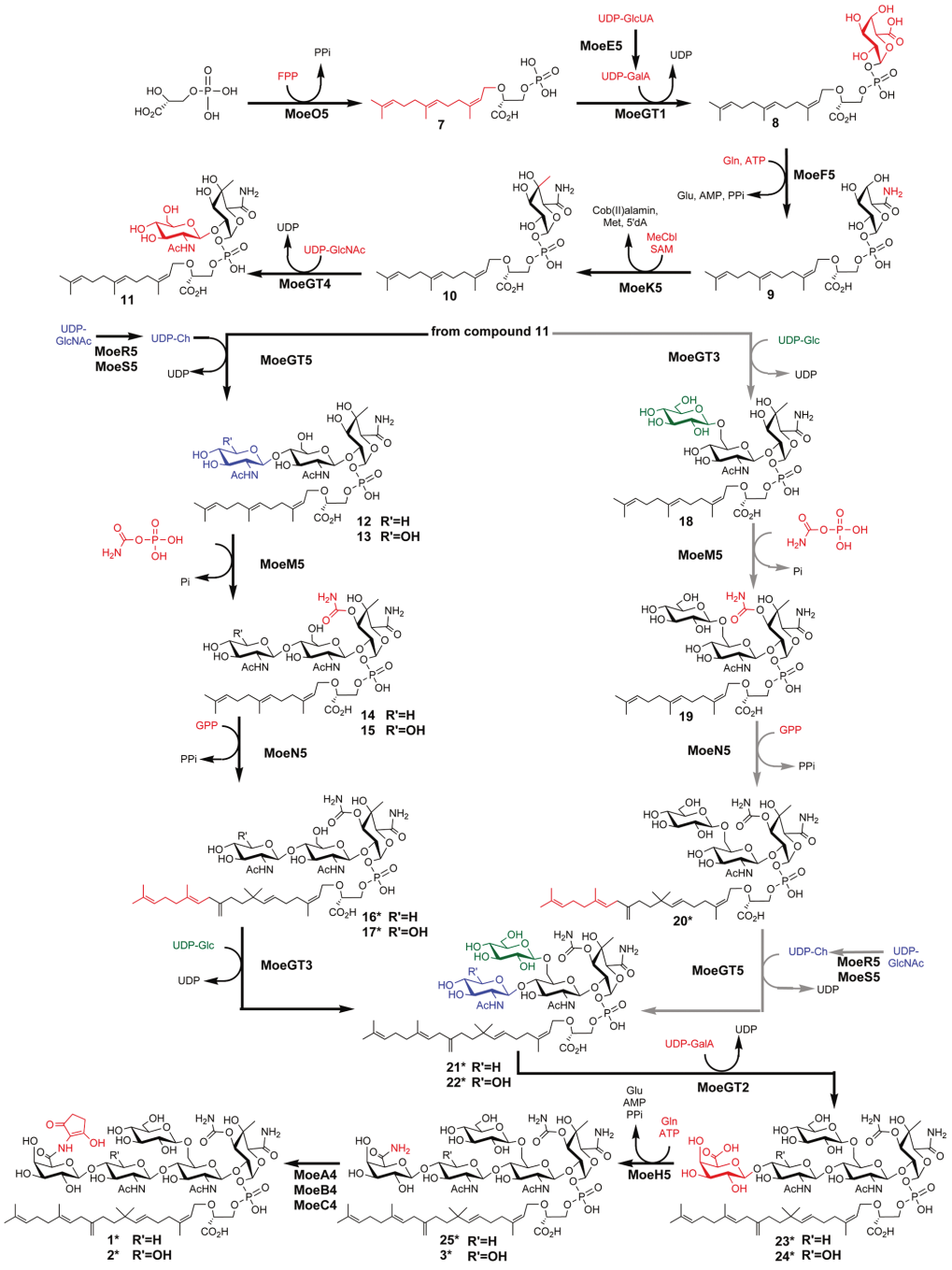

Figure 3.4: Proposed pathway for moenomycin biosynthesis. Solid arrows indicate the order of moenomycin assembly. Structural components shown in red, green, and blue indicate the positions of new moieties in the biosynthetic intermediates. Asterisks (*) indicate biologically active compounds. Structures shown are consistent with exact mass and MS^2 fragmentation data. Stereochemical assignments are consistent with the extensive body of literature on moenomycins isolated from fermentation broths (Donnerstag et al, 1996; Hennig et al, 1998). We cannot exclude the possibility that the tailoring enzymes are able to act at multiple points during assembly.

Functions of the glycosyltransferases. To probe the functions of each of the five putative glycosyltransferase genes in *moe* cluster 1, we disrupted each of them either singly or in combination (Figure 3.2). We were able to isolate glycosylated MmA intermediates from strains deficient in MoeGT2, MoeGT3, MoeGT4, and MoeGT5, but not from a strain deficient in MoeGT1. Taken together, these results suggested that MoeGT1 transfers the first sugar to the lipid phosphoglycerate. Based on the structure of the F ring, we speculated that the natural substrate for MoeGT1 is either UDP-glucuronic acid (UDP-GlcUA) or UDP-galacturonic acid (UDP-GalUA). To verify the function of MoeGT1 and assess its donor sugar substrate specificity, we cloned and overexpressed the enzyme as a His-tag fusion in *E. coli*. Following purification, MoeGT1 was incubated with 40 μM of farnesyl phosphoglycerate **7** and 1 mM of either UDP-GlcUA or a mixture of UDP-GlcUA and UDP-GalUA, prepared by incubating UDP-GlcUA with a well-characterized UDP-GalUA C4-epimerase from *Klebsiella pneumoniae*, Gla_{KP} (Frirdich et al, 2005). After 2 h, a product having an exact mass corresponding to compound **8** was detected in both reactions; however, the epimeric mixture of sugars reacted to >80% completion while pure UDP-GlcUA only went to 10% completion (Table 3.1). Although Michaelis-Menten analysis of MoeGT1 is currently hampered by the low activity of the reconstituted enzyme, a comparison of the two sugar donors following separation of the epimeric mixture showed a clear preference for UDP-GalUA over UDP-GlcUA (0.34 μmol/min for UDP-GalUA and 0.08 μmol/min for UDP-GlcUA, Figure 3.5). Therefore, UDP-GalUA is assigned as the natural sugar for MoeGT1, consistent with the presence, demonstrated below,

of a UDP-GalUA C4 epimerase in *moe* cluster 1.

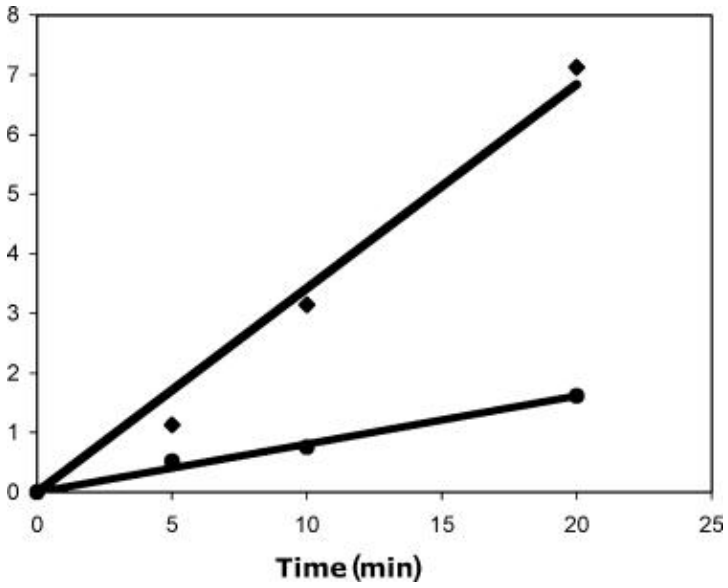

Figure 3.5: Kinetics of MoeGT1. Comparison of the MoeGT1 reaction carried out at 37 °C in 50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 40 μM compound **7**, 800 nM MoeGT1, and 1 mM of either UDP-GlcUA (•) or UDP-GalUA (♦).

The functions of the other glycosyltransferases were established by analyzing MmA intermediates produced by various gene disruption strains. The recombinant 38-1+Δ*moeGT4* strain was found to accumulate monosaccharide **9** (Figure 3.4), suggesting that it attaches the E ring sugar. Consistent with this assignment, all strains containing disruptions of *moeGT5*, *moeGT2*, and *moeGT3*, either individually or in combination, were found to produce moenomycin precursors containing the E ring (Figure 3.4 and Table 3.1). Furthermore, complementation of the deletion strain with a plasmid-borne copy of moeGT4 produced the desmethylated pentasaccharide **4**. On the basis of these results, we conclude that MoeGT4 acts after MoeGT1 to produce the C15 disaccharide **11**.

Upon complementation of the 38-1+Δ*moeGT4* strain, desmethyl compound **4** was produced rather than **3** because the 3' end of the putative methyltransferase *moeK5* gene overlaps with *moeGT4*. Although the *moeGT4* gene deletion affected the expression of *moeK5*, the functional assignments of these genes are not affected (see below; see also Figures 3.2 and 3.4).

The order of biosynthesis with respect to the remaining three glycosyltransferases, MoeGT2, MoeGT3, and MoeGT5, indicates a branching pathway. Strains 38-1+Δ*moeGT5* and 38-1+Δ*moeGT2* were found to produce trisaccharide **20** (Figure 3.4) and tetrasaccharide **22** (Figure 3.4), respectively. However, the double mutant strain 38-1+Δ*moeGT5*Δ*moeGT3* accumulated C15 disaccharide **11** (Figure 3.4 and Table 3.1). Moenomycin analogues lacking the branching glucose (the D ring) are naturally produced in *S. ghanaensis* (Welzel, 2005). We found that they also accumulate when *moeGT3* is disrupted in the producing organism (*S. ghanaensis* MO12, Table 3.1). Therefore, we propose that MmA biosynthesis can follow two branches from **11**, depicted in Figure 3.4, which merge at the stage of the tetrasaccharide **21**/**22**. In one branch, MoeGT5 attaches the C ring, which can be either chinovosamine (Ch) or N-acetylglucosamine (GlcNAc), depending on the presence of *moeR5moeS5* (see below) before MoeGT3 attaches the D ring. In the other branch, MoeGT3 attaches the D ring glucose before MoeGT5 attaches the C ring. MoeN5 then extends the lipid chain on trisaccharides **14**/**15** and **19** to produce **16**/**17** and **20**. The absence of mono- and disaccharide phosphoglycolipids having the full lipid chain suggests that the moenocinyl chain is assembled in two distinct phases, although the reason

for this unusual order of assembly is unclear. Following chain elongation, F ring tailoring (see below), and attachment of the fourth sugar (by either MoeGT5 or MoeGT3 depending on which acted previously), MoeGT2 attaches the B ring sugar to produce the pentasaccharide precursors **23**/**24** (Figure 3.4).

Functions of the sugar biosynthesis enzymes. While the gene clusters for most carbohydrate-containing secondary metabolites contain numerous genes involved in the synthesis of nucleotide-sugar building blocks, there are only three genes in the *moe* cluster that appear to be involved in producing NDP-sugar donors. Apparently, the nucleotide-sugar building blocks for moenomycin are siphoned directly from primary metabolism and used largely without further modification. The three sugar biosynthetic genes identified in the *moe* cluster are *moeE5*, which encodes a putative nucleotide sugar epimerase, and *moeR5*-*moeS5*, which encode a 4,6-dehydratase/keto-reductase pair. To evaluate the function of these three putative NDP-sugar biosynthetic genes, we constructed the appropriate recombinant strains.

The 38-1+ strain, which is lacking *moeR5moeS5*, accumulates MmA derivative **3** containing GlcNAc in place of chinovosamine (C ring, Figure 3.1), implying that the missing genes control the conversion of UDP-GlcNAc to UDP-chinovosamine (UDP-Ch, Figure 3.4). Consistent with this notion, we found that supplying the missing genes on a plasmid produces a strain, 38-1+*moeR5*+-*moeS5*+, which accumulates a compound (**25**) (Figure 3.4 and Table 3.1) containing a methyl at C6 of the C ring. Further studies showed that a 38-

1+*moeS5*+ strain yielded only the hydroxy-methyl variant **3** (Ostash et al, 2007); however, a 38-1+*moeR5*+ strain produced **25** in addition to **3**. It is possible that a homologue of *moeS5* in the *S. lividans* genome complements the loss of *moeS5* function in the 38-1+*moeR5*+ strain. Alternatively, MoeR5 may contain both ketoreductase and dehydratase activities. In any event, these results show that moeR5 is required for the biosynthesis of UDP-chinovosamine, the C ring donor for MmA biosynthesis, from UDP-GlcNAc.

To determine the function of the putative nucleotide sugar epimerase *moeE5*, we constructed strain 38-1+Δ*moeE5*, and analyzed its cell extracts. We were unable to detect any phosphoglycolipid compounds in these extracts. Since the apparent glycosyl donor substrate for MoeGT1 is UDP-GalUA (see above), we hypothesized that MoeE5 is responsible for epimerizing UDP-GlcUA obtained from primary metabolism to UDP-GalUA. To test this hypothesis, we cloned and expressed MoeE5 as a His-tagged fusion in *E. coli*. MoeE5 was incubated with UDP-GlcUA, and the nucleotide-sugar products were then analyzed by anion-exchange HPLC. Pure UDP-GlcUA and the epimeric mixture produced by Gla_{KP}, a well-characterized UDP-GalUA C4-epimerase from *K. pneumoniae*, were used as authentic standards (Frirdich et al, 2005). MoeE5 was found to convert UDP-GlcUA to a 60:40 mixture of compounds having the same exact mass and identical retention times as those of the authentic standards (Figure 3.6). Therefore, we conclude that MoeE5 epimerizes UDP-GlcUA to UDP-GalUA to provide the first sugar in MmA biosynthesis.

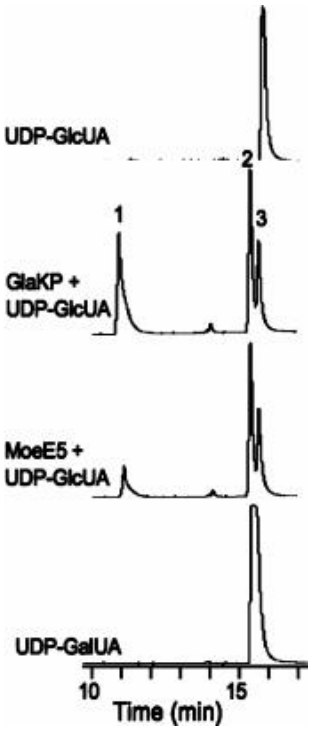

Figure 3.6: HPLC of UDP-GlcUA and UDP-GalUA. Anion-exchange HPLC (see Materials and Methods for conditions) of commercial UDP-GlcUA (peak 3), Gla_{KP} incubated for 2 h at 37 °C with 1 mM UDP-GlcUA in 20 mM Tris-HCl at pH 7.5, MoeE5 incubated for 2 h at 37 °C with 1 mM UDP-GlcUA in 20mM Tris-HCl at pH 7.5, and UDP-GalUA (peak 2) purified from the MoeE5 reaction. NAD+/NADH (peak 1) is strongly associated with both epimerases, and, although it is likely a necessary coenzyme, it was not necessary to add it exogenously.

Functions of the sugar tailoring enzymes. *Moe* cluster 1 contains several genes proposed to encode tailoring enzymes that modify sugars on the growing phosphoglycolipid scaffold. These tailoring enzymes include two genes (*moeA5* and *moeB5*) that are truncated homologues of two of the three genes in *moe* cluster 2 (*moeA4* and *moeB4*, Figure 3.2). The three genes in *moe* cluster 2 encode a putative aminolevulinate synthase (*moeA4*) as well as two other enzymes proposed to be involved in the assembly and attachment of the A ring. Deletion of *moeA4* from *moe* cluster 2 in the producing organism abolishes A ring formation (Sung et al, 2009). Therefore, the putative aminolevulinate synthase in *moe*

cluster 1, *moeA5*, cannot complement the loss of *moeA4* and is likely nonfunctional. Expression of *moe* cluster 2 in the 38-1+Δ*moeA5*Δ*moeB5* deletion strain produces analogues containing the A ring. These observations indicate that *moe* cluster 2 is required for A ring assembly. Neither *moeA5* nor *moeB5* of *moe* cluster 1 is required for MmA biosynthesis.

The remaining putative tailoring enzymes in *moe* cluster 1 include *moeF5*, *moeH5*, *moeK5*, and *moeM5*. *MoeF5* and *moeH5* both encode amidotransferases proposed to form primary amides from carboxylic acids. However, MmA contains only one unsubstituted amide (on the F ring), which led us to propose previously that MoeF5 and MoeH5 may work together to catalyze amidotransfer, although we also considered the possibility that only one of the genes was functional. To probe whether both genes are required for MmA biosynthesis, we disrupted each one individually. Strain 38-1+Δ*moeF5* accumulated monosaccharide **8**, which has a carboxyl moiety at C6 instead of a carboxamide group (Figure 3.4 and Table 3.1). No methylated monosaccharide precursors or larger phosphoglycolipids could be detected. These results imply that MoeF5 is necessary for F ring carboxamidation and imply that the absence of the carboxamide moiety abolishes unit F methylation and all subsequent transformations.

Strain 38-1+Δ*moeH5* accumulated phosphoglycolipid **24** (Figure 3.4 and Table 3.1), indicating that MoeH5 catalyzes carboxamidation of the C6 position of the B ring. This finding may explain the observation that C6 B ring primary amides and acids are observed in addition to MmA in fermentations of moenomycins (see Figure 3.1) (Eichhorn et al, 2005; Zehl et al, 2006). Several

other secondary metabolites contain C5N units identical to the A ring of MmA, and it has been proposed that these subunits are attached in a process involving attack of an amine on an activated acid to form the corresponding amide (Cho et al, 1993; Hu et al, 2004; McAlpine et al, 2005; Ostash et al, 2007). Since an activated acid is a putative precursor in the biosynthesis of MmA, we wondered whether B ring amides are intermediates in the MmA pathway or belong to a different branch of moenomycin metabolism. To determine whether the MoeH5 reaction catalyzes an essential biosynthetic step required for attachment of the unit A chromophore, we coexpressed the genes for unit A biosynthesis (pOOB64b) in strain 38-1+, which produces **3** in the absence of the A ring genes, and also in strain 38-1+Δ*moeH5*, which produces **24** in the absence of the A ring genes (Figure 3.4). We observed that pholipomycin **2** (Figure 3.1b), which contains the A ring, accumulates in the 38-1+pOOB64b+ strain but not in the 38-1+Δ*moeH5*pOOB64b+ strain (Table 3.1) (Welzel, 2005). The inability of the strain deficient in *MoeH5* to make a product containing the A ring implies that carboxylic acid **24** is not the direct precursor to **2** (Schuricht et al, 2000; Petricek et al, 2006; Ostash et al, 2007). We propose that MoeH5 catalyzes the formation of carboxamides **3**/**25** as necessary precursors to MmA and pholipomycin, respectively (compounds **1** and **2**, Figure 3.4). The details of the chemistry involved are not understood, but evidently do not involve attack of an amine on an activated acid. These results may have implications for the biosynthesis of other secondary metabolites containing C5N subunits.

Gene *moeK5* encodes a protein homologous to a recently discovered

family of putative radical-SAM, methyl-cobalamin-dependent methyltransferases involved in the biosynthesis of fosfomycin, pactamycin, and a handful of other secondary metabolites (Kuzuyama et al, 1992; Hidaka et al, 1995; Kuzuyama et al, 1995; Nunez et al, 2003; Unwin et al, 2004; Kudo et al, 2007; Woodyer et al, 2007). We have proposed that MoeK5 controls the methylation of the first sugar (unit F) (Ostash et al, 2007). Indeed, strain 38-1+Δ*moeK5* accumulates the desmethylated MmA derivative **4** (Table 3.1 and Figure 3.1b). Since previous studies have shown that desmethyl MmA derivatives have the axial stereochemistry at C4 of the F ring (El-Abadla et al, 1999; Eichhorn et al, 2005), as shown in compound **4** (and consistent with our results showing that UDP-GalUA is the F ring precursor), MoeK5-catalyzed methylation must proceed with inversion at the C4 hydroxyl. Because we identified methylated C15-monosaccharide-phosphoglycerate compounds in extracts of several glycosyltransferase deletion strains, we have concluded that methylation of the F ring sugar occurs at the monosaccharide stage. It must occur following MoeF5-catalyzed carboxamidation of the F ring since only desmethylated compound **8** could be detected in the 38-1+Δ*moeF5* strain (Figure 3.4 and Table 3.1).

The remaining tailoring enzyme, MoeM5, was established in our previous study to be a carbamoyltransferase since extracts from a *S. ghanaensis* strain in which the *moeM5* gene was deleted produced only an inactive phosphoglycolipid pentasaccharide lacking the F ring carbamate (see Figure 3.1a) (Ostash et al, 2007). However, the previous study provided no information on when MoeM5 acts in MmA biosynthesis. We analyzed extracts from the set of *moeM5*+

glycosyltransferase deletion strains produced in this study (see Table 3.1) to try and determine when the carbamate is installed. We detected only descarbamoylated products in extracts accumulating disaccharides; however, only carbamoylated products were detected in extracts accumulating trisaccharides and larger scaffolds. Therefore, we have concluded that carbamoylation occurs primarily at the trisaccharide stage during fermentation (Figure 3.4). This modification, however, is not required to complete the synthesis of the pentasaccharide scaffold since full-length analogues lacking the carbamate are produced in the Δ*moeM5* strain.

On the basis of the preceding studies, we have concluded that *moe* cluster 1 contains 14 demonstrably functional biosynthetic genes, while *moe* cluster 2 contains 3. As described below, fewer than half of these 17 biosynthetic genes are required to produce bioactive MmA analogues, an observation that should facilitate efforts to produce novel analogues by pathway engineering.

Antibacterial activity of moenomycins produced in recombinant strains. In order to define the subsets of genes required to make bioactive MmA analogues, we assessed the activity of phosphoglycolipids isolated from selected recombinant *S. lividans* strains on a *Bacillus cereus* reporter strain using a disk diffusion assay. Relative compound potencies were assessed by comparing the antibiotic concentrations required to achieve clear zones of inhibition. The results (Figure 3.7 and Table 3.2) confirm previous structure-activity relationships reported by Welzel and others but also provide additional insights (Welzel et al, 1987; Welzel,

2005). As expected, monosaccharide compounds **8** and **9** were found to be inactive, while moenocinyl-linked tetra- and pentasaccharides **22** and **3** were found to have activity similar to that of MmA itself. These results confirm the modest role of the A ring in antibiotic activity. The C ring sugar was previously proposed to be an essential part of the minimal pharmacophore, but we have found that DEF trisaccharide **20** also has activity (Welzel, 2005). In fact, the MIC of **20** against a methicillin sensitive *Staphylococcus aureus* strain was found to be comparable to that of the CEF trisaccharide (8 μg/mL for 20 versus 3 μg/mL for a CEF trisaccharide analogue) (Welzel et al, 1987; Ritzeler et al, 1997; Welzel, 2005; Adachi et al, 2006; Yuan et al, 2008). The observation that both trisaccharides have biological activity is consistent with structural work showing that the majority of directional contacts between moenomycin and its enzyme target involve only the EF-phosphoglycerate portion of the molecule, suggesting that this is the most critical portion of moenomycin for specificity (Yuan et al, 2008; Heaslet et al, 2009; Sung et al, 2009). Since both trisaccharides are active, both could serve as starting points for the elaboration of phosphoglycolipid analogues.

Some of the compounds produced in our studies also provide information about the minimum length of the lipid required for biological activity. It was established previously that MmA analogues containing neryl (C10) chains are biologically inactive even though they have midnanomolar enzyme inhibitory potencies (Adachi et al, 2006). This and other work have indicated an important role for the lipid chain in biological activity (Lovering et al, 2007; Lovering et al,

2008; Yuan et al, 2008; Heaslet et al, 2009). It is presumed that the prenyl chain must be long enough to interact with the cell membrane when the pentasaccharide scaffold is bound in the PGT active site so that it can compete with the lipid-anchored substrates, which contain a C55 prenyl chain. We have found that farnesylated compounds **5** and **6**, produced by deleting *moeN5*, show clear zones of inhibition. Therefore, three isoprene units in the prenyl chain are necessary and sufficient to confer biological activity, albeit with a loss in potency. The C25 chain is believed to be responsible for many of the undesirable properties of MmA, including its excessively long half-life and propensity to bind serum. Biologically active analogues containing shorter lipid chains as well as other modifications to improve potency may have more favorable properties than MmA itself, and the studies reported here indicate that it should be possible to reduce the size of the lipid considerably.

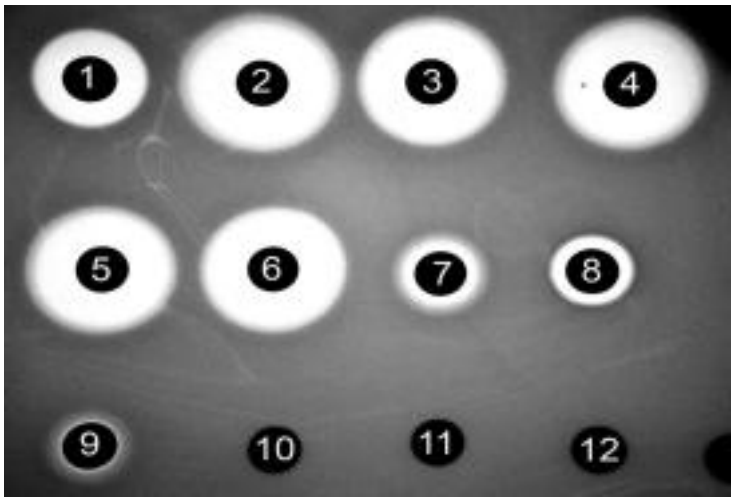

Figure 3.7: Disk diffusion assay of MmA intermediates against reporter strain B. cereus ATCC19637. Disks 1 and 2, MmA (**1**), 10 and 100 nmol, respectively; disk 3, compound **22**, 100 nmol; disk 4, compound **25**, 100 nmol; disk 5, compound **23**, 100 nmol; disk 6, compound **4**, 100 nmol; disk 7, mixture of compounds **5** and **6**, 200 nmol; disk 8, compound **20**, 100 nmol; disk 9, mixture of compounds **4** and **5**, 50 nmol; disk 10, extract from 5 g of *S. lividans* TK24 mycelial cake; disk 11 compound **9**, 200 nmol; disk 12, compound **8**, 200 nmol.

Table 3.2: Relative antibiotic activity of MmA intermediates based on disc diffusion

Compound	Antibiotic activity[a]
1	+++
2	+++
3	+++
4	+++
5 and 6	+
7	-
8	-
9	-
11	-
20	++
22	+++
23	+++
24	+++
25	+++

[a]+++ indicates at least 90% as active as MmA, ++ indicates 50-89% activity, + indicates 15-49% activity, and - indicates no visible bioactivity.

3.4: Discussion

Summary of the biosynthetic pathway and implications for analogue production. The phosphoglycolipids, which target the active site of peptidoglycan glycosyltransferases, remain the only known starting points for the production of potentially useful antibiotics that target these enzymes. In this article, we have described a set of studies that have allowed us to propose a complete biosynthetic pathway for the synthesis of MmA, which should in turn enable the chemoenzymatic production of analogues. The pathway order is reconstructed from deletions of 17 *moe* genes and analysis/identification of accumulating

intermediates, bolstered by initial enzymatic assays of MoeO5, MoeGT1, and MoeE5. Biosynthesis begins with the coupling of farnesyl pyrophosphate and phosphoglycerate to form starter unit **7**. Sugars are then sequentially attached by a series of glycosyltransferases, starting with the MoeGT1-catalyzed transfer of galacturonic acid (F ring precursor, Figure 3.4) to **7**. Before additional processing can occur, the galacturonic acid must undergo an obligatory tailoring step by MoeF5, which converts the C6 carboxylic acid to a carboxamide. Methylation at the C4 position, catalyzed by MoeK5, then takes place but is not required for further biosynthetic steps to proceed. MoeGT4 transfers the second sugar, N-acetylglucosamine (unit E, GlcNAc), to form the EF disaccharide. MoeGT5 and MoeGT3 then catalyze the addition of the C and D sugar units, respectively (Figure 3.4), but the order of addition is flexible. At the trisaccharide stage, MoeN5 catalyzes extension of the lipid chain to form the mature C25 isoprenoid chain, and MoeM5 catalyzes carbamoylation of the F ring. Neither modification is required for the other one to occur nor for subsequent glycosylations. After tetrasaccharide **21**/**22** is produced (by either MoeGT3 or MoeGT5), MoeGT2 transfers another galacturonic acid subunit to the scaffold (B ring, Figure 3.4). The B ring C6 acid is then converted to the corresponding amide by MoeH5; this transformation is required for subsequent attachment of the C5N chromophore by the gene products encoded in *moe* cluster 2.

The results described reveal considerable promiscuity in the moenomycin biosynthetic pathway, which may explain the observation that moenomycins are typically isolated as a complex mixture of related compounds (Schuricht et al,

2000; Zehl et al, 2006). We have found that except for unit F carboxamidation, every one of the sugar tailoring reactions can be bypassed without preventing the assembly of the phosphoglycolipid pentasacharide scaffold. Furthermore, both MoeGT4 and MoeGT5 can accept either UDP-GlcNAc or UDP-chinovosamine as donor substrates, explaining the production of pholipomycin (**2**) and moenomycin C3 as well as moenomycin A (**1**) in both producing organisms and the heterologous expression host (Welzel, 2005). The complexity of moenomycin mixtures obtained from fermentation has been a major hindrance to efforts to purify intermediates for various uses, including studies directed toward selective chemical derivatization to make novel phosphogycolipid analogues. It should now be possible to simplify the spectrum of moenomycin metabolites produced via fermentation or to optimize the production of new phosphoglycolipid scaffolds for chemoenzymatic synthesis by judicious deletion or overexpression of selected genes in either the producing organism or a suitable heterologous host (Schuricht et al, 2000).

In addition to establishing gene function and the order of assembly of the pentasaccharide, we have identified the subsets of genes (7 out of the 17 total) required to form biologically active scaffolds. These genes include moeE5, moeO5, moeGT1, moeF5, moeGT4, moeM5, and either moeGT3 or moeGT5. It should be possible to produce phosphoglycolipid scaffolds for further chemical or enzymatic elaboration with 7 or fewer enzymes, depending on the exogenous substrates supplied, making *in vitro* reconstitution imminently feasible. The ability to reconstitute *in vitro* the production of small phosphoglycolipid scaffolds

will make it possible to explore the substrate tolerances of each enzyme in detail. This information should, in turn, be useful in guiding efforts to engineer heterologous hosts to produce novel phosphoglycolipid antibiotics or precursors for further chemical elaboration.

3.5: Conclusions

We have established functions for the 17 moenomycin biosynthetic genes, determined the order of assembly of the molecule, shown which transformations are required for full assembly of the pentasacharide scaffold, and identified the subsets of 7 genes involved in the production of bioactive analogues. This work will facilitate efforts to make analogues to explore the potential of phosphoglycolipids as useful antibiotics.

3.6: References

Adachi, M., Zhang, Y., Leimkuhler, C., Sun, B., LaTour, J. V., and Kahne, D. E. (2006) Degradation and reconstruction of moenomycin A and derivatives: dissecting the function of the isoprenoid chain. *J Am Chem Soc* 128:14012–14013.

Anantharaman, V., Aravind, L., and Koonin, E. V. (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* 7:12–20.

BlondeletRouault, M. H., Weiser, J., Lebrihi, A., Branny, P., and Pernodet, J. L. (1997) Antibiotic resistance gene cassettes derived from the Omega interposon for use in *E-coli* and *Streptomyces*. *Gene* 190:315–317.

Caetano-Anolles, G., Yafremava, L. S., Gee, H., Caetano-Anolles, D., Kim, H. S., and Mittenthal, J. E. (2009) The origin and evolution of modern metabolism. *Int J Biochem Cell Biol* 4:285–297.

Cho, H., Beale, J. M., Graff, C., Mocek, U., Nakagawa, A., Omura, S., and Floss, H. G. (1993) Studies on the biosynthesis of the antibiotic reductiomycin in *Streptomyces xanthochromogenus*. *J Am Chem Soc* 115:12296–12304.

Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 97:6640–6645.

Dewick, P. M. (2002) The biosynthesis of C-5-C-25 terpenoid compounds. *Nat Prod Rep* 19:181–222.

Domingo, V., Arteaga, J. F., del Moral, J. F. Q., and Barrero, A. F. (2009) Unusually cyclized triterpenes: occurrence, biosynthesis and chemical synthesis. *Nat Prod Rep* 26:115–134.

Donnerstag, A., Hennig, L., Findeisen, M., Welzel, P., and Haessner, R. (1996) H-1 NMR as a tool for the structure elucidation of moenomycin antibiotics. *Magn Reson Chem* 34:1031–1035.

Eichhorn, P., and Aga, D. S. (2005) Characterization of moenomycin antibiotics from medicated chicken feed by ion-trap mass spectrometry with electrospray ionization. *Rapid Commun Mass Spectrom* 19:2179–2186.

El-Abadla, N., Lampilas, M., Hennig, L., Findeisen, M., Welzel, P., Muller, D., Markus, A., and van Heijenoort, J. (1999) Moenomycin A: The role of the methyl group in the moenuronamide unit and a general discussion of structure-activity relationships. *Tetrahedron* 55:699–722.

Flett, F., Mersinias, V., and Smith, C. P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol Lett* 155:223–229.

Frirdich, E., and Whitfield, C. (2005) Characterization of Gla(KP), a UDP-galacturonic acid C4-epimerase from Klebsiella pneumoniae with extended substrate specificity. *J Bacteriol* 187:4104–4115.

Goldman, R., and Gange, D. (2000) Inhibition of transglycosylation involved in bacterial peptidoglycan synthesis. *Curr Med Chem* 7:801–820.

Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P., and Meutermans, W. (2006) Targeting the forgotten transglycosylases. *Biochem Pharmacol* 71:957–967.

Heaslet, H., Shaw, B., Mistry, A., and Miller, A. A. (2009) Characterization of the active site of S. aureus monofunctional glycosyl- transferase (Mtg) by site-directed mutation and structural analysis of the protein complexed with moenomycin. *J. Struct Biol* 167:129– 135.

Hennig, L., Findeisen, M., Welzel, P., and Haessner, R. (1998) H-1 NMR spectroscopic studies of the moenomycins. Magn Reson Chem 36:615–620.

Hidaka, T., Hidaka, M., Kuzuyama, T., and Seto, H. (1995) Sequence of a P-methyltransferase-encoding gene isolated from a bialaphos-producing *Streptomyces hygroscopius*. *Gene* 158:149–150.

Hu, Y. D., and Floss, H. G. (2004) Further studies on the biosynthesis of the manumycin-type antibiotic, asukamycin, and the chemical synthesis of protoasukamycin. *J Am Chem Soc* 126:3837– 3844.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., and Hopwood, D. A. (2000) Practical Streptomyces Genetics, The John Innes Foundation, Norwich, U.K.

Kudo, F., Kasama, Y., Hirayama, T., and Eguchi, T. (2007) Cloning of the pactamycin biosynthetic gene cluster and characterization of a crucial glycosyltransferase prior to a unique cyclopentane ring formation. *J Antibiot* 60:492–503.

Kuzuyama, T., Hidaka, T., Kamigiri, K., Imai, S., and Seto, H. (1992) Studies on the biosynthesis of fosfomycin: The biosynthetic origin of the methyl-group of fosfomycin. *J Antibiot* 45:1812–1814.

Kuzuyama, T., Seki, T., Dairi, T., Hidaka, T., and Seto, H. (1995) Nucleotide-sequence of fortimicin Kl1 methyltransferase gene isolated from *Micromonospora olivasterospora*, and comparison of its deduced amino-acid-sequence with those of methyltransferases involved in the biosynthesis of bialaphos and fosfomycin. *J Antibiot* 48:1191–1193.

Liang, P. H., Ko, T. P., and Wang, A. H. J. (2002) Structure, mechanism and function of prenyltransferases. *Eur J Biochem* 269:3339–3354.

Lovering, A. L., de Castro, L. H., Lim, D., and Strynadka, N. C. (2007) Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science* 315:1402–1405.

Lovering, A. L., De Castro, L., and Strynadka, N. C. J. (2008) Identification of dynamic structural motifs involved in peptidoglycan glycosyltransfer. *J Mol Biol* 383:167–177.

McAlpine, J. B., Bachmann, B. O., Piraee, M., Tremblay, S., Alarco, A. M., Zazopoulos, E., and Farnet, C. M. (2005) Microbial genomics as a guide to drug discovery and structural elucidation: ECO- 02301, a novel antifungal agent, as an example. *J Nat Prod* 68:493–496.

Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* 321:741–765.

Neundorf, I., Kohler, C., Hennig, L., Findeisen, M., Arigoni, D., and Welzel, P. (2003) Evidence for the combined participation of a C-10 and a C-15 precursor in the biosynthesis of moenocinol, the lipid part of the moenomycin antibiotics. *Chem Bio Chem* 4:1201– 1205.

Nunez, L. E., Mendez, C., Brana, A. F., Blanco, G., and Salas, J. A. (2003) The biosynthetic gene cluster for the beta-lactam carbapenem thienamycin in *Streptomyces cattleya*. *Chem Biol* 10:301–311.

Ostash, B., and Walker, S. (2005) Bacterial transglycosylase inhibitors. *Curr Opin Chem Biol* 9:459–466.

Ostash, B., Saghatelian, A., and Walker, S. (2007) A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chem Biol* 14:257–267.

Payandeh, J., Fujihashi, M., Gillon, W., and Pai, E. F. (2006) The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. *J Biol Chem* 281:6070–6078.

Petricek, M., Petrickova, K., Havlicek, L., and Felsberg, J. (2006) Occurrence of two 5-aminolevulinate biosynthetic pathways in *Streptomyces nodosus* subsp asukaensis is linked with the production of asukamycin. *J Bacteriol* 188:5113–5123.

Ritzeler, O., Hennig, L., Findeisen, M., Welzel, P., and Muller, D. (1997) Search for new moenomycin structure-activity relationships. Synthesis of a trisaccharide precursor of a moenomycin analogue. *Tetrahedron* 53:5357–5357.

Sambrook, J., and Russel, D. W. (2001) Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

Schuricht, U., Endler, K., Hennig, L., Findeisen, M., and Welzel, P. (2000) Studies on the biosynthesis of the antibiotic moenomycin A. *J Prakt Chem* 342:761–772.

Sekurova, O. N., Brautaset, T., Sletta, H., Borgos, S. E. F., Jakobsen, O. M., Ellingsen, T. E., Strom, A. R., Valla, S., and Zotchev, S. B. (2004) *In vivo* analysis of the regulatory genes in the nystatin biosynthetic gene cluster of Streptomyces noursei ATCC 11455 reveals their differential control over antibiotic biosynthesis. *J Bacteriol* 186:1345–1354.

Sung, M. T., Lai, Y. T., Huang, C. Y., Chou, L. Y., Shih, H. W., Cheng, W. C., Wong, C. H., and Ma, C. (2009) Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from *Escherichia coli*. *Proc Natl Acad Sci USA* 106:8824–8829.

Taylor, J. G., Li, X., Oberthur, M., Zhu, W., and Kahne, D. E. (2006) The total synthesis of moenomycin A. *J Am Chem Soc* 128:15084– 15085.

Unwin, J., Standage, S., Alexander, D., Hosted, T., Horan, A. C., and Wellington, E. M. H. (2004) Gene cluster in *Micromonospora echinospora* ATCC15835 for the biosynthesis of the gentamicin C complex. *J Antibiot* 57:436–445.

Welzel, P., Kunisch, F., Kruggel, F., Stein, H., Scherkenbeck, J., Hiltmann, A., Duddeck, H., Muller, D., Maggio, J. E., Fehlhaber, H. W., Seibert, G., Vanheijenoort, Y., and Vanheijenoort, J. (1987) Moenomycin-A: Minimum structural requirements for biological activity. *Tetrahedron* 43:585–598.

Welzel, P. (2005) Syntheses around the transglycosylation step in peptidoglycan biosynthesis. *Chem Rev* 105:4610–4660.

Wierenga, R. K. (2001) The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett* 492:193–198.

Woodyer, R. D., Li, G. Y., Zhao, H. M., and van der Donk, W. A. (2007) New

insight into the mechanism of methyl transfer during the biosynthesis of fosfomycin. *Chem Commun* 4:359–361.

Wright, G. (2007) A new target for antibiotic development. *Science* 315:1373–1374.

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D., Sliz, P., and Walker, S. (2007) Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proc Natl Acad Sci U.S.A.* 104:5348–5353.

Yuan, Y., Fuse, S., Ostash, B., Sliz, P., Kahne, D., and Walker, S. (2008) Structural analysis of the contacts anchoring moenomycin to peptidoglycan glycosyltransferases and implications for antibiotic design. *ACS Chem Biol* 3:429–436.

Zhang, D. L., and Poulter, C. D. (1993) Biosynthesis of archaebacterial ether lipids: Formation of ether linkages by prenyltransferases. *J Am Chem Soc* 115:1270–1277.

Zehl, M., Pittenauer, E., Rizzi, A., and Allmaier, G. (2006) Characterization of moenomycin antibiotic complex by multistage MALDI-IT/RTOF-MS and ESI-IT-MS. *J Am Soc Mass Spectrom* 17:1081–1090

Chapter 4: Studies towards the reconstitution of the minimal pharmacophores of moenomycin A.

4.1: Introduction

The moenomycin (MmA, **1**, Figure 4.1) gene cluster has been identified in the producing organism *Streptomyces ghanaensis* (Ostash et al, 2007), and an order of assembly for MmA biosynthesis has been proposed using genetic analysis (Ostash et al, 2009). Evaluation of an MmA analog containing a C15-farnesyl lipid suggested that truncation of the C25-moenocinol lipid to a C15-farnesyl lipid maintains antibacterial activity. These studies also showed that C25-moenocinol trisaccharide scaffolds consisting of either the CEF or DEF rings are biologically active based on minimum inhibitory concentration (MIC) measurements. We also isolated a desmethyl-C25-moenocinol-pentasaccharide analog from a Δ*moeK5* deletion strain to confirm that the 4'-methyl on glycan ring F is not required for biological activity (Ostash et al, 2009). The minimal MmA pharmacophore thus appears to be composed of the desmethyl-C15-farnesyl-CEF-trisaccharide **2a/b** and/or the desmethyl-C15-farnesyl-DEF-trisaccharide **3a/b** (Figure 4.1). Therefore, we set out to reconstitute the MmA enzymes that are required for the synthesis of these two putative minimal MmA pharmacophores (Figure 4.2). Successful reconstitution would allow verification of the proposed functions of the MmA enzymes that have not yet been verified, and also lay the groundwork for the exploration of substrate tolerance and the potential for the chemoenzymatic synthesis of MmA analogs.

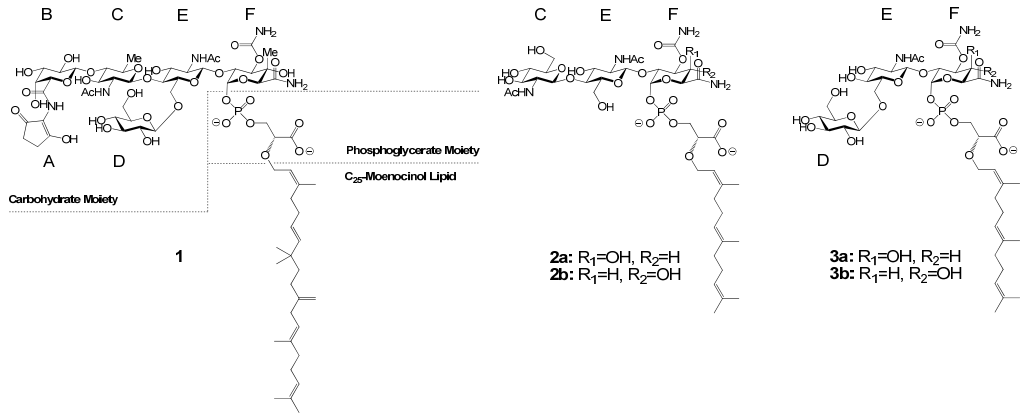

Figure 4.1: MmA and trisaccharide pharmacophores. Chemical structures of moenomycin A (**1**), desmethyl-C15-farnesyl-CEF-trisaccharide (**2**), and desmethyl-C15-farnesyl-DEF-trisaccharide (**3**).

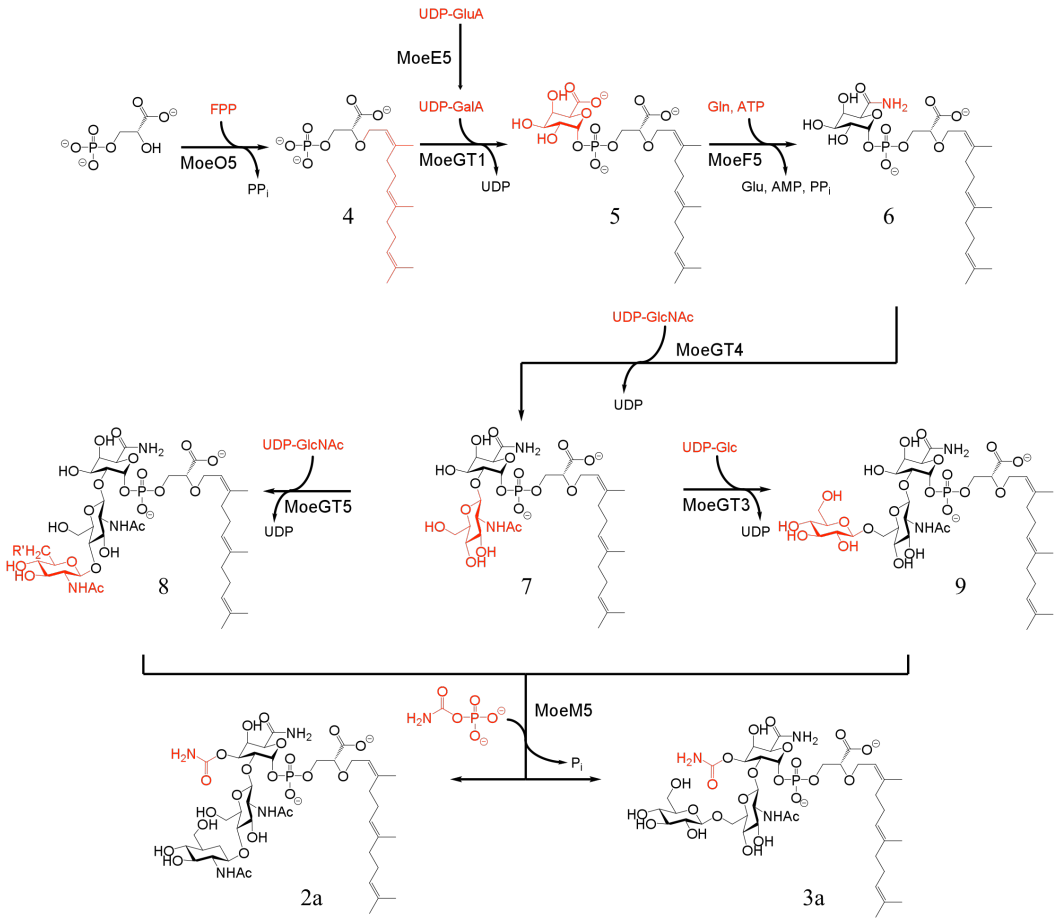

Figure 4.2: Enzymatic reconstitution scheme. Enzymatic synthesis of the putative MmA pharmacophores, desmethyl-C15-farnesyl-C, E, F-trisaccharide MmA (**2a**) and desmethyl-C15-farnesyl-D, E, F-trisaccharide MmA (**3a**).

Reconstitution would allow the development of a chemoenzymatic method to assemble a MmA scaffold, which would then enable rapid exploration of the chemical space around the MmA pharmacophore. From a medicinal chemistry perspective, reconstitution is a cost-effective methodology for the development of a lead compound that has both the desired antibacterial and pharmacokinetic properties for further evaluation in animal models. Our strategy involved the chemical synthesis of unnatural substrates that probe the promiscuity of the

biosynthetic enzymes involved in assembling the MmA pharmacophore. Structure-activity relationship (SAR) data suggest that modifications to the lipid moiety as well as to select regions of the glycan rings C, E and F may result in analogs with enhanced biological activity (Halliday et al, 2006). We therefore focused our investigation on enabling modifications to the lipid and carbohydrate regions of MmA. We pursued two approaches: (1) the synthesis of unnatural phosphoglycerate lipid acceptors that would enable us to evaluate the range of lipid analogs that may be enzymatically attached to the carbohydrate portion, and (2) the synthesis of unnatural carbohydrate donors featuring functionalizable chemical handles that would enable us to evaluate the potential of assembling derivatizable MmA scaffolds.

4.2: Results

4.2.1: MoeF5

Gene disruption studies indicated that MoeF5 is a putative amidotransferase necessary for the conversion of the F ring carboxylic acid **5** to the corresponding carboxyamide **6** (Ostash et al, 2009). These studies revealed that Δ*moeF5 S. lividans* strains accumulate monosaccharide **5**, suggesting that the absence of the carboxamide moiety abolishes downstream tailoring and glycosylation events. To verify the function of MoeF5, I reconstituted the activity of the protein. The protein was overexpressed in Rosetta2(DE3)pLysS *E. coli* cells, a strain optimized for the expression of high GC content proteins, and purified using Ni-

NTA affinity chromatography (Figure 4.3). The identity of MoeF5 was confirmed by Western blotting using an antibody that recognizes the His6 tag. MoeF5 consistently copurified with a slightly larger partner protein (asterix in Figure 4.3). A second protein band just below MoeF5 on the SDS-PAGE gel could either be a degradation product or another protein that copurifies with MoeF5 (arrowhead in Figure 4.3). Since MoeF5 has a theoretical isoelectric point of 9.4, cation exchange chromatography was tested to increase the purity of the sample, but the use of both strong (sulfate derivatives) and weak (carboxylic acid derivaties) cation exchange resins, was unsuccessful. Since the protein requires detergent (0.5% CHAPS) to remain soluble, it is possible that the contaminating proteins are chaperones or proteins that interact with the hydrophobic patches on MoeF5. Each liter of *E. coli* cells yielded approximately 0.2 mg MoeF5 protein.

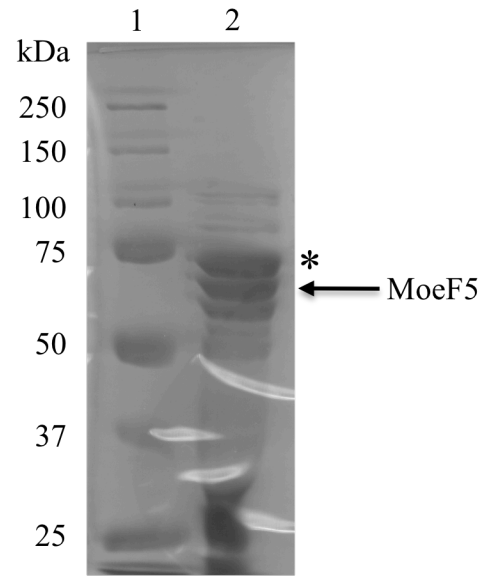

Figure 4.3: SDS-PAGE gel of purified MoeF5. Lane 1: Molecular weight standards; lane 2: Ni-NTA affinity purified MoeF5.

The function of MoeF5 was verified by incubating the protein with 40 μM C15-farnesyl-monosaccharide **5** and 10 mM glutamine in the presence of ATP. Liquid chromatography mass spectrometry (LCMS) of the product mixture revealed the appearance of product C15-farnesyl-F-amido-monosaccharide **6** after 2 hours (Figure 4.4). The protein is also able to use ammonia (in the form of NH_4OAc) as the nitrogen source. MoeF5 is indiscriminate towards Mg^{2+} and Mn^{2+} ions, but requires one or the other for activity, presumably to stabilize ATP in the catalytic cleft of the enzyme. MoeF5 also requires 10% DMSO, 10% glycerol, and 4 mM DTT in the buffer for efficient activity, as omission of any of these additives significantly reduced product turnover, which could then only be detected if a greater amount of starter material was used.

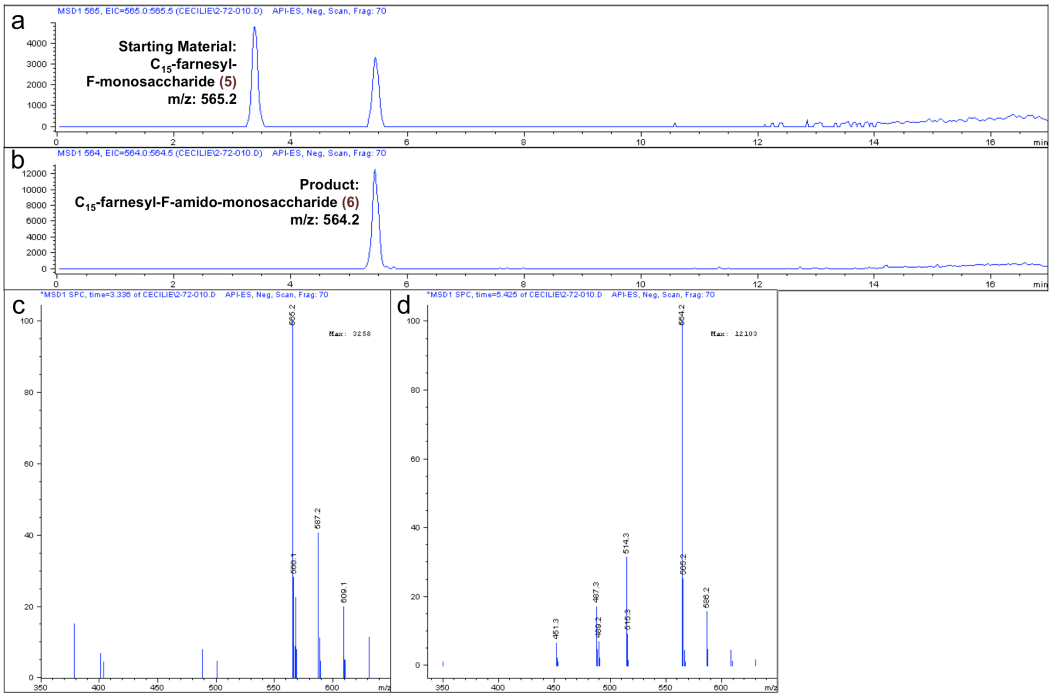

Figure 4.4: MoeF5 reaction: MoeF5 was incubated with 40 µM C15-farnesyl monosaccharide **5** and 10 mM glutamine in the presence of ATP. (a) The LCMS spectrum of the reaction after 2 h incubation reveals the presence of unreacted starting material **5**. (b) The LCMS of the reaction after 2 h incubation reveals the presence of product C15-farnesyl-F-amido-monosaccharide **6**. (c) Electron spray ionization mass spectrometry (ESI-MS) analysis of the starting material **5** peak: 565.2056 $[M-H]^-$ calculated, 565.2 observed. (d) ESI-MS of product **6** peak: 564.2215 $[M-H]^-$ calculated, 564.2 observed.

4.2.2: MoeGT4

Gene disruption studies indicated that MoeGT4 is a putative glycosyltransferase necessary for the coupling of the glycan ring E to the C15-farnesyl-F-amido-monosacchride **6** (Ostash et al, 2009). To verify the function of MoeGT4, I reconstituted the activity of the protein. My initial constructs, N-terminally His-tagged MoeGT4 (His_6-MoeGT4) and C-terminally His-tagged MoeGT4

(MoeGT4-His_6), expressed well but were insoluble, even in the presence of various ionic, non-ionic and zwitterionic detergents. A screen of ionic strength, pH, and the addition of divalent metal ions was also performed but unsuccessful. Hydropathy analysis of MoeGT4 in Expasy suggested that the protein contained a significant hydrophobic patch near its N terminus, which presumably anchors the protein to the periplasmic membrane in order to efficiently handle the lipophilic substrate and product molecules. I therefore constructed MoeGT4 with an N-terminal thioredoxin tag (Trx-MoeGT4). This construct expressed well and was soluble in the presence of CHAPS (0.5%). Interestingly, Trx-MoeGT4 was extremely sensitive to salt (NaCl) and precipitated even in the presence of 50 mM NaCl. Ni-NTA affinity purified Trx-MoeGT4 is shown in Figure 4.5. Each liter of *E. coli* cells yielded approximately 2 mg MoeGT4.

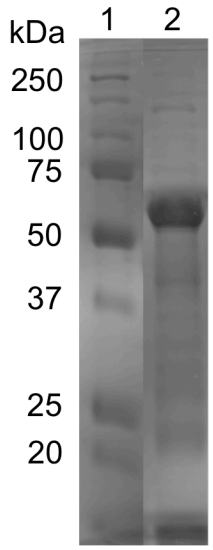

Figure 4.5: SDS-PAGE gel of purified MoeGT4. Lane 1: molecular weight standards; lane 2: Ni-NTA affinity purified Trx-MoeGT4.

The function of MoeGT4 was verified by incubating the enzyme with dried C15-farnesyl-F-amido-monosaccharide **6** and UDP-GlcNAc, and analyzing the reaction by LCMS (Figure 4.6). MoeGT4 was found to be active only in the presence of Mn^{2+}, and addition of glycerol to the reaction buffer increased enzymatic catalytic activity. The Mn^{2+} cations are likely required to coordinate the diphosphate moiety of UDP-GlcNAc during catalysis.

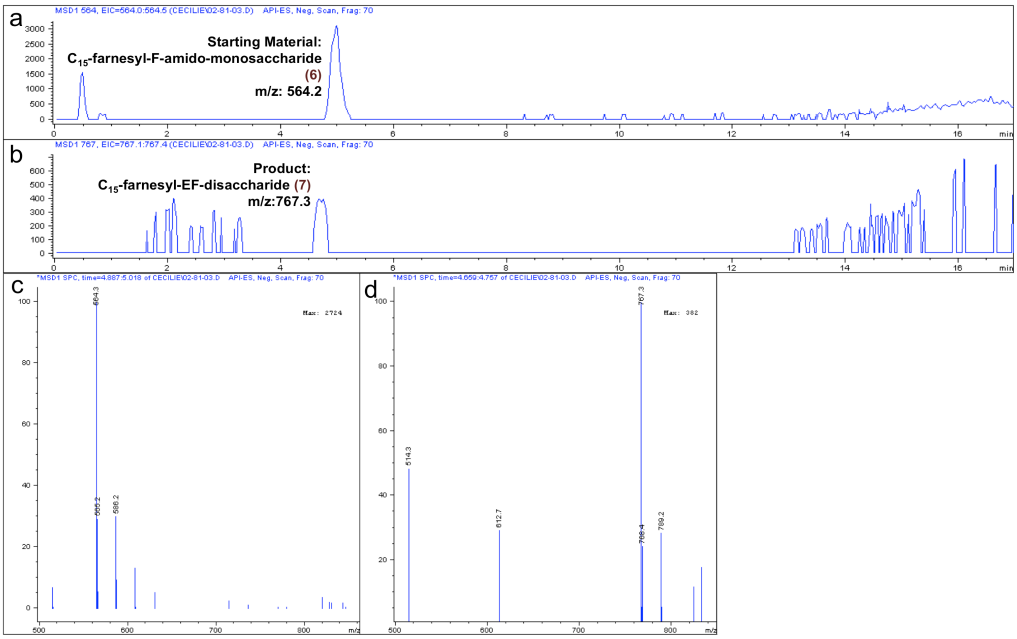

Figure 4.6: MoeGT4 reaction. MoeGT4 was incubated with dried C15-farnesyl F-amido monosaccharide **6** and UDP-GlcNAc in the presence of $MnCl_2$. (a) The LCMS spectrum of the reaction after 10 h incubation reveals the presence of unreacted starting material **6**. (b) The LCMS spectrum of the reaction after 10 h incubation reveals product C15-farnesyl-EF-disaccharide **7** formation. (c) ESI-MS analysis of the starting material **6** peak: 564.2215 $[M-H]^-$ calculated, 564.2 observed. (d). ESI-MS analysis of the product **7** peak: 767.3009 $[M-H]^-$ calculated, 767.3 observed.

4.2.3: MoeGT5

Gene disruption studies indicated that MoeGT5 is a putative glycosyltransferase necessary for the coupling of glycan ring C to the C15-farnesyl-EF monosacchride **7** (Ostash et al, 2009). To verify the glycosyltransferase activity of MoeGT5, I reconstituted the activity of the protein. MoeGT5 was difficult to express and purify. I first attempted to express MoeGT5 as either an N-terminally His-tagged protein (His_6-MoeGT5) or a C-terminally His-tagged protein (MoeGT5-His_6) in Rosetta2(DE3)pLysS *E. coli* cells, a strain that enhances the expression of high GC content genes. Unlike MoeGT1, MoeE5, MoeF5, and MoeGT4, which all exhibited higher soluble expression levels at 16°C as compared to 30°C or 37°C, MoeGT5 did not express at 16°C, even after an overnight incubation. MoeGT5 expressed at 37°C but exhibited unusual expression patterns. Expression levels peaked after 1 hour of induction with IPTG and steadily decreased to zero within 4 hours, presumably because the protein was degraded due to instability or misfolding. Expression of MoeGT5 in ArcticExpress *E. coli* cells, a strain that overexpresses chaperone proteins to enhance the folding and solubility of recombinant proteins, was, however, unsuccessful.

Since the goal was to confirm its function rather than to produce large amounts of pure MoeGT5, I decided to obtain protein by using large quantities of cell mass harvested 1 hour after induction. After screening a panel of ionic, nonionic and zwitterionic detergents for their ability to extract MoeGT5 from the bacterial membranes, I found n-octyl-β-D-glucopyranoside (OG) to be the most

potent, enabling the extraction of ~30% MoeGT5. MoeGT5 copurified with many proteins on Ni-NTA affinity chromatography, and final yields of MoeGT5 were less than 0.2 mg per liter of culture.

The function of MoeGT5 was probed by incubating the enzyme with dried C15-farnesyl-EF-disaccharide **7** and UDP-GlcNAc, and the reaction was analyzed by LCMS (Figure 4.7). Since only a small amount of starting material **7** could be obtained enzymatically, the amounts used for each MoeGT5 reaction were small (nM range). The limited starting material accounts for the appearance of multiple mass peaks at increasing elution times and the low ion count associated with the product **8** peak. Although the possibility remains that the peak associated with product represents noise rather than true product formation, the experiments were performed multiple times and the product **8** peak appeared reproducibly, suggesting that the enzyme was able to turnover substrate to product. For verification, the experiments need to be reproduced with more starting material **7**.

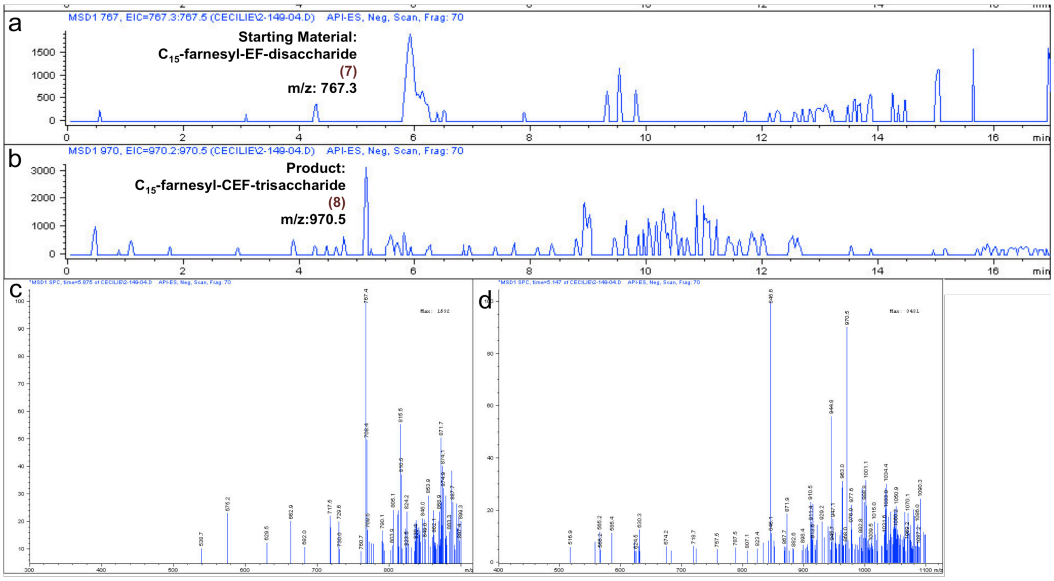

Figure 4.7: MoeGT5 reaction. MoeGT5 was incubated with dried C15-farnesyl EF-amido disaccharide **7** and UDP-GlcNAc. (a) The LCMS spectrum of the reaction after 10 h incubation reveals the presence of unreacted starting material **7**. (b) The LCMS spectrum of the reaction after 10 h incubation reveals product C15-farnesyl-CEF-trisaccharide **8** formation. (c) ESI-MS analysis of the starting material **7** peak: 767.3009 $[M-H]^-$ calculated, 767.3 observed. (d) ESI-MS analysis of the product **8** peak: $[M-H]^-$ calculated, 970.4, 970.5 observed.

4.2.4: MoeGT3

Gene disruption studies indicated that MoeGT3 is a putative glycosyltransferase necessary for the coupling of the D ring to the C15-farnesyl-EF monosacchride **7** (Ostash et al, 2009). To verify the function of MoeGT3, I tried to reconstitute the activity of the protein. Initial expression and solubilization attempts of N-terminally His_6-tagged MoeGT3 (His_6-MoeGT3), C-terminally His_6-tagged MoeGT3 (MoeGT3-His_6), thiroedoxin-MoeGT3, and nus-MoeGT3 failed because all constructs formed insoluble inclusion bodies. Inclusion bodies formed with all combinations of expression conditions (temperature, late *versus* early induction,

amount of IPTG used), *E. coli* strains and detergents used. Since *Streptomyces* proteins are known to express poorly in *E. coli*, I also tried to express MoeGT3 in *S. lividans*. MoeGT3 expressed well in *S. lividans*, but was only soluble when clarification was performed at below 10,000g. A panel of ionic, nonionic, and zwitterionic detergents was assayed for their ability to increase the solubility of the protein. Although CHAPS and Triton-X100 were able to solubilize a small amount of MoeGT3, the protein yield was too low to test its function.

4.2.5: MoeM5

Gene disruption studies indicate that MoeM5 is a putative carbamoyltransferase that attaches a carbamate functionality on the F glycan ring after the assembly of either the C15-farnesyl-CEF trisaccharide or the C15-farnesyl-DEF trisaccharide (Ostash et al, 2009). IC_{50} measurements have indicated that the carbamoyl group increases the binding affinity of meonomycin to the peptidoglycan glycosyltransferases by 100 fold. I attempted to overexpress C-terminally His_6-tagged MoeM5 (MoeM5-His_6), thioredoxin-MoeM5, and GST-MoeM5 in *E. coli*. All constructs formed insoluble inclusion bodies and the protein could not be refolded with detergent. In order to obtain soluble protein, I then attempted to express N-terminally His6-tagged MoeM5 and C-terminally His6-tagged MoeM5 in *S. lividans*. Both constructs expressed well, remained soluble, and could be purified to near homogeneity in the presence of CHAPS or Triton X-100 (Figure 4.8). Final yields approached 5 mg per liter of *S. lividans* culture.

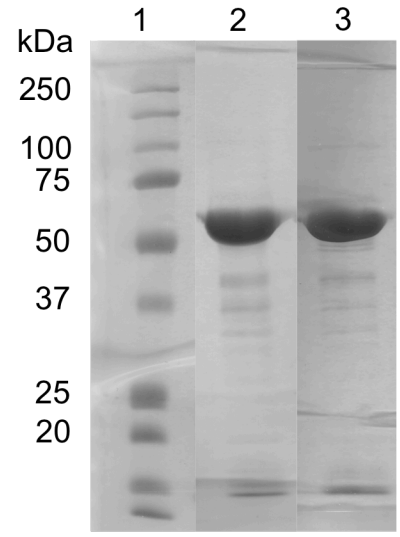

Figure 4.8: SDS PAGE gel of purified MoeM5. Lane 1: Molecular weight standards; lane 2: Ni-NTA affinity purified His6-MoeM5; lane 3: Ni-NTA affinity purified MoeM5-His6.

4.2.6: MoeGT4 substrate promiscuity

Novel MmA analogs that contain amino sugars are excellent scaffolds that enable rapid chemical diversification. I synthesized UDP-3-amino-3-deoxy-GlcNAc via an enzymatic route established by Raetz and coworkers (2004) (Figure 4.9). A preliminary activity assay using MoeGT4 reaction conditions suggested that the enzyme was unable to recognize this amino sugar as a substrate. It may be that enzymatic turnover is low and that optimization of reaction conditions is needed enhance the turnover rate of the enzyme.

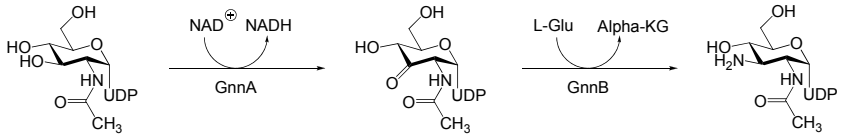

Figure 4.9: Enzymatic synthesis of UDP 3-amino-3-deoxy-GlcNAc.

4.2.7: Synthesis of unnatural phosphoglycerate lipid acceptors

The C25-moenocinol lipid of MmA contributes to the molecule's poor pharmacokinetic properties. In collaboration with Dr. Hiro Tsukamoto, I synthesized 10 unnatural phosphoglycerate lipid acceptors that vary in length and structure (Figure 4.10). Since MoeGT1 displays substrate promiscuity towards the UDP-sugar donor, it is likely that the enzyme may also accept variations in the lipid acceptor, especially since the lipid is distant from the site of catalysis. This would enable the rapid synthesis of MmA analogs with diverse lipids that may exhibit better pharmacokinetic properties compared to the parent molecule.

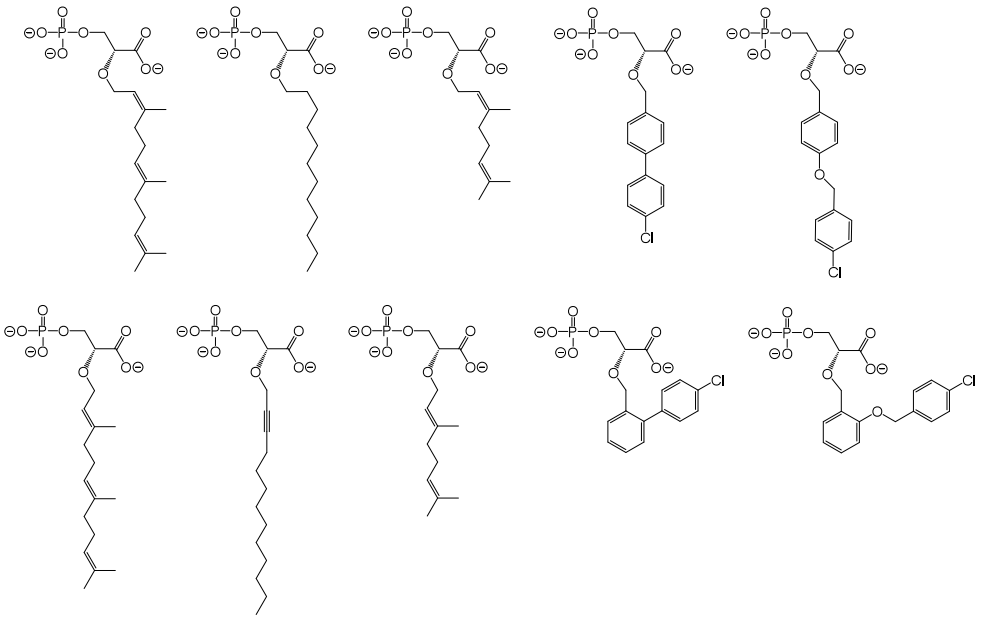

Figure 4.10: A panel of chemically synthesized unnatural lipid acceptors to be tested as substrates for MoeGT1.

4.3: Discussion

This chapter describes efforts aimed at reconstituting the minimal moenomycin pharmacophores required for *in vivo* antibacterial activity. Initial expression and purification conditions have been worked out for MoeF5, MoeGT4, MoeGT5, and MoeM5. The purified proteins allowed the enzymatic activity of MoeF5, MoeGT4, and MoeGT5 to be verified. The activity of MoeM5 could not be ascertained, because of substrate material constraints. The expression and purification of MoeGT3 was attempted in *E. coli* and *S. lividans*, but in both cases it was not possible to isolate soluble protein. This problem was not anticipated, because the cellular conditions of *S. lividans* are similar to those of the producing organism, *S. ghanaensis*. Furthermore, since MmA could be isolated when the monomycin producing gene cluster was conjugated into *S. lividans*, MoeGT3

must be expressed in an active form in *S. lividans*. It is thus likely that expression parameters can be found that will allow soluble protein to be obtained.

The reconstitution of glycosyltransferases MoeGT1, MoeGT4 and MoeGT5 enabled studies of their substrate promiscuity. Structural-activity relationship data suggest that modifications to the lipid moiety as well as to select regions of the glycan rings C, E and F may result in analogs with enhanced biological activity. Should the enzymes exhibit relaxed substrate specificity, a combination of two approaches can be used to generate a diverse array of moenomycin analogs: 1) Unnatural phosphoglycerate lipid acceptors can be synthesized to evaluate the range of lipid analogs that may be enzymatically attached by using MoeGT1, and 2) Unnatural carbohydrate donors can be synthesized with functionalizable chemical handles that enable chemical derivatization of the moenomycin scaffold.

Unnatural lipids. The C25-moenocinyl lipid of moenomycin is believed to be responsible for the molecule's undesirable pharmacokinetic properties (poor absorption, non-specific binding to cellular components) and inability to exhibit efficacy against gram-negative bacteria (Pfaller, 2006). The design of moenomycin analogs that overcome these limitations would be possible if MoeGT1 exhibits substrate promiscuity towards the acceptor lipid. The choice of unnatural lipid substrate should be motivated by the demands to 1) maintain or increase non-specific binding to the PGT active site as judged by IC_{50} measurements and 2) maintain antibacterial activity while improving

pharmacokinetics as judged by measuring the differential in MICs in the presence and absence of serum.

The semisynthesis of a C10-neryl moenomycin analog has shown that truncation of the C25-moenocinyl lipid to a C10-neryl lipid destroys antibacterial activity while exhibiting a negligible effect on peptidoglycan glycosyltransferase inhibition (Adachi et al, 2006). Recent studies suggest that a C15-farnesyl moenomycin analog isolated from genetically engineered *S. lividans* retains a similar level of antibacterial activity as moenomycin. It is likely that a certain amount of lipophilicity is required to target moenomycin to the membrane bound peptidoglycan glycosyltransferases. However, extensive lipophilicity, as on the parent molecule, contributes to non-specific binding to cellular components that results in a substantial reduction in the effective concentration of moenomycin. The currently available SAR data tentatively suggest that the C15-farnesyl lipid may represent the optimal isoprenyl lipid length required for antibacterial activity, and thus the panel of unnatural lipid substrates designed for MoeGT1 was based on this premise. It should be noted that MoeGT1 has predicted structural homology to MurG, a glycosyltransferase in peptidoglycan biosynthesis that exhibits remarkable acceptor lipid promiscuity. This similarity, together with the observation that the lipid is removed from the immediate site of catalysis, suggests that MoeGT1 has a high chance of exhibiting promiscuity.

Unnatural sugars. Recent SAR data indicate that a trisaccharide configuration consisting of either the CEF or DEF glycan rings forms the smallest carbohydrate fragment that retains antibacterial activity (Ostash et al, 2009).

Although structural studies of a C10-neryl moenomycin analog bound to the *A. aeolicus* peptidoglycan glycosyltransferase suggest that the functionalities on the glycan rings E and F are critical in forming specific polar contacts with the enzyme, studies by Sofia and coworkers have shown that slight changes and additional derivatization with a range of chemical moieties may effect an increase in antibacterial activity (Sofia et al, 1999; Yuan et al, 2007). Detailed analysis of the crystal structure also suggests that there is space around the 3'-hydroxyl group of both the glycan rings C and E for addition of larger substituents to increase non-specific binding to the enzyme (Yuan et al, 2007). This finding provides compelling motivation to explore the possibility of introducing a chemical handle at the 3' positions of the glycan rings C, E, and F to enable derivatization.

The generation of a panel of C15-farnesyl-C/DEF-trisaccharide analogs containing 3-amino-3-deoxy modifications would enable rapid diversification of the moenomycin scaffold. Amines are attractive functionalities in drug design because they are charged at physiological conditions and nucleophilic and thus amenable to chemical derivatization. I synthesized UDP-3-amino-3-deoxy-GlcNAc via an enzymatic route established by Sweet and coworkers (2004). Although an initial activity assay suggested that MoeGT4 is unable to utilize this UDP-amino sugar, this outcome could be due to unfavorable reaction conditions. More experiments are needed before definitive conclusions can be drawn. If conditions that induce MoeGT4 and MoeGT5 to accept unnatural UDP-amino sugars can be found, a diverse array of moenomycin derivatives can be made and tested for their antibacterial activities. Such derivatization could perhaps lead to

moenomycin analogs that exhibit high antibacterial acitivty and acceptable pharmacokinetic properties.

4.4: Materials and methods

Expression of Moe proteins. Plasmids encoding MoeGT1, MoeF5, MoeGT5, MoeE5, and MoeGT4 were transformed into Rosetta2(DE3)pLysS competent *E. coli* cells. Starting from a 10 mL overnight culture, MoeGT1, MoeF5, MoeE5, and MoeGT4 were grown at 37°C in Luria-Bertani (LB) medium to an OD_{600} of 0.6. The temperature was reduced to 16°C and the cells were induced by the addition of 1 mM IPTG and allowed to grow overnight before harvesting by centrifugation. The overexpression plasmid for MoeGT5 was transformed into Rosetta2(DE3)pLysS competent *E. coli* cells and grown at 37°C in LB medium to an OD_{600} of 0.6. At that point, the cells were induced with 1 mM IPTG and grown for an additional 1 h before the culture was harvested by centrifugation. The overexpression plasmid for MoeM5, pJVD53-MoeM5, was conjugated into *Streptomyces lividans*. Starting from a 1 mL starter culture, MoeM5 was grown at 30°C in YEME medium for 72 h before induction with e-caprolactam. The cells were then allowed to grow for an additional 24 h before harvesting by centrifugation.

Purification of Moe proteins. Cell pellets containing MoeE5, MoeGT1, and MoeF5 were lysed by freeze-thaw in Buffer A (20 mM Tris-HCl, pH 8, 200 mM NaCl, 5 mM imidazole, 5% glycerol, 0.5% CHAPS) supplemented with rlysozyme, benzonase, and protease inhibitor cocktail. Cellular debris and

membrane components were pelleted by centrifugation at 38,700 x g for 45 minutes. The clarified cell lysate was loaded onto a column containing Ni-NTA resin equilibrated with Buffer A. The column was washed with 8 column volumes of Buffer W (20 mM Tris-HCl, pH 8, 200 mM NaCl, 5 mM imidazole) followed by 6 column volumes of Buffer W containing 30 mM imidazole. The proteins were eluted with a step gradient of imidazole in Buffer W (75-500 mM imidazole). Purified MoeE5 and MoeGT1 were dialyzed twice against 4 L of 20 mM Tris-HCl, pH 7. Crudely purified MoeF5 was dialyzed twice against 3 L of 20 mM Tris-HCl, pH 7, 4 mM dithiothreitol (DTT), and 10% glycerol.

Cells from expression cultures of MoeGT4 were lysed by freeze-thaw in Buffer B (20 mM Tris-HCl, pH 8, 10% glycerol, 0.5% CHAPS) supplemented with rlysozyme, benzonase, and protease inhibitor cocktail. After centrifugation at 16,000 x g for 40 min to remove cellular debris and membrane components, the supernatant was rocked with Ni-NTA resin for 6 h. The resin/lysate was loaded onto a column and washed with 18 column volumes of Buffer V (20 mM Tris-HCl, pH 8, 5 mM imidazole) followed by 12 column volumes of Buffer V containing 30 mM imidazole. The purified protein was eluted with 500 mM imidazole in Buffer V, and dialyzed against 3 L of 20 mM Tris-HCl, pH 7, 5 mM $MnCl_2$, and 10% glycerol.

Cell pellets containing MoeGT5 were lysed by freeze-thaw in Buffer C (20 mM Tris-HCl, pH 8, 200 mM NaCl, 5 mM imidazole, 1% n-octyl-β-D-glucopyranoside) supplemented with rlysozyme, benzonase, and protease inhibitor cocktail. After clarification of the cell lysate at 38,700 x g for 30 min,

the supernatant was loaded onto a column containing Ni-NTA resin. The protein was washed and eluted in a manner identical to the MoeGT1/MoeE5/MoeF5 purification.

Cell pellets containing MoeM5 were lysed in Buffer D (50 mM Tris-HCl, pH 7, 10% glycerol, 0.5% CHAPS or 0.5% Triton X-100) supplemented with rlysozyme and protease inhibitor cocktail. The protein was then clarified and purified in a manner identical to the MoeGT1/MoeE5/MoeF5 purification.

C_{15}-farnesyl-F-monosaccharide ***5*** *biosynthesis.* To synthesize C_{15}-farnesyl-F-monosaccharide **5**, purified MoeGT1 was incubated with 0.1 mg C_{15}-farnesyl-phosphoglycerate **4** and 2 mM of a 3:2 mixture of UDP-galacturonic acid (UDP-GalUA) and UDP-glucuronic acid (UDP-GlcUA) at 30°C with mild shaking for 10 h. The reaction was stopped by passage through a Microcon filtration unit (10K molecular weight cut-off). Analytical HPLC was carried out on a Gemini C18 column (50 x 4.6 mm) using solvent A (95 water: 5 methanol + 0.1% ammonium hydroxide) and solvent B (60 isopropanol: 35 methanol: 5 water + 0.1% ammonium hydroxide). A gradient of 0-75% solvent B over 17 min was used at a flow rate of 1 ml/min. Product peaks were detected using absorbance at 210 nm, collected, and characterized with high-resolution mass spectrometry: 565.2056 $[M-H]^-$ calculated, 565.2066 observed.

The C_{15}-farnesyl-F-amido-monosaccharide **6** was synthesized by adding Buffer F (20 mM Tris-HCl, 4 mM DTT, 10% glycerol, 10% DMSO, 10 mM $MgCl_2$, 10 mM ATP) to a tube containing dried C_{15}-farnesyl-F-monosaccharide **5** and glutamine, pH 9 (10 mM final concentration) or ammonium acetate, pH 7

(100 mM final concentration). The reaction was initiated by addition of crudely purified MoeF5, and allowed to incubate at 30°C with mild shaking for 10 h. The reaction was stopped by immersion into an 85°C water bath for 5 min followed by centrifugation at 14,000 x g for 10 min. Analytical HPLC was carried out as described above. Product peaks were detected using absorbance at 210 nm, collected, and characterized with ESI-MS: 564.2215 $[M-H]^-$ calculated, 564.2 observed.

To synthesize C_{15}-farnesyl-EF-disaccharide **7**, purified MoeGT4 was incubated with dried C_{15}-farnesyl-F-amido-monosaccharide **6** and 5 mM UDP-GlcNAc in Buffer 4 (20 mM Tris-HCl, pH 7, 10% glycerol, 5 mM $MnCl_2$) at 30°C with mild shaking for 10 h. The reaction was quenched and analyzed as described for the MoeGT1 reaction. The formation of product was detected with ESI-MS: 767.3009 $[M-H]^-$ calculated, 767.3 observed.

To synthesize C_{15}-farnesyl-CEF-trisaccharide **8**, MoeGT5 was incubated with dried C_{15}-farnesyl-E, F-disaccharide **7** and 5 mM UDP-GlcNAc in MoeGT4 reaction buffer at 30°C with mild shaking for 10 h. The reaction was quenched and analyzed as described for the MoeGT1 reaction. The formation of product was detected with ESI-MS and HRMS.

4.5: References

Adachi, M., Zhang, Y., Leimkuhler, C., Sun, B., LaTour, J., and Kahne, D. (2006) Degradation and reconstruction of moenomycin A and derivatives: dissecting the function of the isoprenoid chain. *J Am Chem Soc* 128:14012-14013.

Ostash, B., Doud, E., Lin, C., Ostash, I., Perlstein, D., Fuse, S., Wolpert, M., Kahne, D., and Walker, S. (2009) Complete characterization of the seventeen step moenomycin biosynthetic pathway. *Biochem* 48:8830-8841.

Ostash, B., Saghatelian, A., and Walker, S. (2007) A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chem Biol* 14:257-267.

Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P., & Meutermans, W. (2006) Targeting the forgotten transglycosylases. *Biochem Pharmacol* 71:957-967.

Pfaller, M. (2006) Flavophospholipol use in animals: positive implications for antimicrobial resistance based on its microbiologic properties. *Diagn Microbio Infect Dis* 56:115-121.

Sofia, M., Allanson, N., Hatzenbuhler, N., Jain, R., Kakarla, R., Kogan, N., Liang, R., Liu, D., Silva, D., Wang, H., Gange, D., Anderson, J., Chen, A., Chi, F., Dulina, R., Huang, B., Kamau, M., Wang, C., Baizman, E., Brandstrom, A., Bristol, N., Goldman, R., Han, K., Longley, C., Midha, S., and Axelrod, H. (1999) Disvoery of novel disaccharide antibacterial agents using a combinatorial library approach. *J Med Chem* 42:3193-3198.

Sweet, C., Ribeiro, A., and Raetz, C. (2004) Oxidation and transmination of the 3'-position of UDP-N-acetylglucosamine by enzymes from *Acidithiobacillus ferrooxidans*. *J Biol Chem* 279:25400-25410.

Yuan, Y., Barrett, D., Zhang, Y., Kahne, D. & Walker, S. (2007) Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *Proc Natl Acad Sci* 104:5348-5353.

Chapter 5: Introduction to RIG-I-like helicases

5.1: The innate immune system

Double-stranded RNA (dsRNA), the genetic material of many RNA viruses, triggers a powerful innate immune response in vertebrate organisms (Wang et al, 2011). The innate immune response is rapid and coupled to adaptive immunity, which provides vertebrate organisms with long-term protection based on immunological memory (Rodriguez et al, 2012). The vertebrate immune system is able to distinguish between viral and host RNA. Characteristic viral components that are absent in the host organism include long dsRNA, 5'-triphosphorylated (ppp) RNA, and unmethylated CpG DNA (Zhong et al, 2010). Host mRNAs are typically capped and are bound by poly-A binding proteins (Afonina et al, 1998), and ribosomal RNAs are integrated into ribonucleoprotein complexes (Yusupov et al, 2001), thus providing the host organism protection from eliciting an autoimmune response.

Effective immune defense against viral infection depends on the rapid and robust detection of pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors. The innate immune system has developed two major pathways to respond to the detection of viral nucleic acids, each using a specific class of pattern recognition receptors. One pathway for the detection of viral RNA is mediated by Toll-like receptor 3 (TLR3), an integral membrane protein located in the endosome. TLR3 detects phagocytosed viral components that have been released into the extracellular space by infected cells that undergo lysis (Schulz et

al, 2005). TLR3 may also detect viruses that are internalized through receptor-mediated endocytosis. When TLR3 is activated, it signals through the protein TIR-domain-containing adaptor-inducing interferon β (TRIF) to activate two kinases, TANK-binding kinase-1 (TBk1) and I-kappa B Kinase-epsilon (IKK-ε). These kinases phosphorylate interferon (IFN) transcription factors, which induce the expression of type-I IFNs (Hoebe et al, 2003; Yamamoto et al, 2003).

The second pathway for the detection of viral RNA is mediated by the RIG-I-like receptors (RLRs), retinoic acid inducible gene-I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5). RIG-I and MDA5 are conserved viral RNA receptors in vertebrates, sharing similar domain architectures and approximately 30% sequence identity (Wilkins et al, 2010). RLRs are expressed ubiquitously in most cell types and sense viral RNA in the cytoplasm (Takeuchi et al, 2008; Yoneyama et al, 2007; Meylan et al, 2006). Upon viral RNA recognition, RIG-I and MDA5 interact with a common signaling adaptor, mitochondria antiviral-signaling protein (MAVS, also known as IPS1) and activate the NF-kB and IRF3/7 signaling pathways, resulting in the expression of type I IFNs and proinflammatory cytokines (Rehwinkel et al, 2010).

5.2: Toll-like receptors (TLRs)

TLRs are pattern recognition receptors that recognize microbial structures as well as host-derived molecules that are released from injured or apoptotic cells (Kariko et al, 2004). TLRs are located either on the cell surface (TLR1, TLR2, TLR4, and TLR6) or in endosomal membranes (TLR3, TLR7, TLR8, and TLR9). Cell

surface TLRs form homo- and heterodimers that recognize microbial membrane lipids, peptidoglycans and lipopolysaccharides, which is distinct from endosomal TLRs, which recognize microbial nucleic acids (Kaisho et al, 2006). Of the four endosomal TLRs, only TLR3 recognizes dsRNA and thus is the endosomal counterpart to the cytoplasmic RLRs. TLR7, TLR8, and TLR9 recognize ssRNA, G-rich oligonucleotides, and DNA, respectively.

TLR3 is expressed in the placenta, pancreas, lung, liver, heart, lymph nodes, and brain (Muzio et al, 2000; Rock et al, 1998; Zarember et al, 2002). TLR3 is also present in immune cells, including T lymphocytes, natural killer cells, macrophages, and mast cells (Hornung et al, 2002; Orinska et al, 2005; Tabiasco et al, 2006). Interestingly, human fibroblasts and epithelial cells express TLR3 on the cell surface (Matsumoto et al, 2002; Matsumoto et al, 2003).

TLR3 shows significant homology to TLRs that sense nucleic acids (Iwasaki et al, 2004). It is a 125 kDa type I transmembrane receptor with an extracellular ectodomain (ECD), a transmembrane region, and an intracellular Toll/IL-1 receptor (TIR) domain. The ECD is capped by leucine-rich repeat (LRR) motifs and dimerize upon ligand binding (Wang et al, 2010a). X-ray crystallography of two TLR3 ECDs bound to dsRNA revealed that each ECD binds dsRNA at two sites located at opposite ends of the ECD (Liu et al, 2008). This study also showed that dimeric ECD is stabilized by intermolecular interactions at the C terminus of the protein.

TLR3 recognizes viral dsRNA molecules that are at least 40 base pairs in length as well as polyriboinosinic-polyribocytidylic acid (poly(I:C)), a synthetic

mimic of dsRNA (Wang et al, 2010a). Poly(I:C) internalization into endosomes is mediated by co-receptor CD14, which then "hands over" the ligand to TLR3 through CD14-TLR3 interactions (Lee et al, 2006). TLR3 shows preference for cell-associated dsRNA over soluble dsRNA. Cell-associated dsRNA also produces a stronger type I IFN response in tumor-specific T cells (McBridde et al, 2006). Furthermore, RNA released from necrotic cells has been found to induce the activation of TLR3-mediated immunity, including the activation of dendritic cells (Kariko et al, 2004).

While most TLR signaling occurs through the adaptor protein myeloid differentiation primary response gene 88 (MyD88), dsRNA-mediated TLR3 signaling induces the recruitment of adaptor protein TRIF (Jiang et al, 2004). TLR3 interacts with TRIF through TIR-TIR domain interactions and causes TRIF to homodimerize through its TIR domain and C-terminal region. Homodimerization induces a conformational change in TRIF that exposes protein surfaces that are needed to bind downstream proteins (Tatematsu et al, 2010). After activation, TRIF recruits tumor necrosis factor receptor-associated 3 (TRAF3), NAK-associated protein 1 (NAP1), and two kinases, TBk1 and IKK-e. These interactions result in the phosphorlyation and dimeraization of transcription factors interferon regulatory factor 3 (IRF3) and IRF7, which subsequently translocate into the nucleus to induce the expression of type I IFNs (Honda et al, 2006). In a similar fashion, TRIF also interacts with receptor interacting protein 1 (RIP1) to activate transcription factor NF-kB, which causes the expression of pro-infammatory genes (Meylan et al, 2004).

5.3: RIG-I-like receptors

The RIG-I-like family of proteins is comprised of RIG-I and MDA5, helicases that detect viral RNA in the cytoplasm of vertebrate cells. RIG-I and MDA5 are members of the DEAD (Asp-Glu-Ala-Asp) box RNA helicase family (Linder & Jankowsky, 2011), and contain a central helicase domain that hydrolyzes ATP (Fairman-Williams et al, 2010). The N terminus of RIG-I and MDA5 contain two tandem caspase activation and recruitment domains (CARDs) that activate downstream signaling proteins via binding to the mitochondrial membrane adaptor protein, MAVS, through CARD-CARD interactions (Berke & Modis, 2012). RIG-I and MDA5 also contain a C-terminal domain (CTD), which is believed to regulate substrate binding and protein activation (Peisley et al, 2012). The most notable sequence differences between MDA5 and RIG-I are a significantly longer linker between the CARDs and the helicase domain (105 residues in human MDA5 compared with 55 residues in human RIG-I) and the slightly more acidic nature of MDA5's Hel2i domain (the last 150 amino acids of the helicase domain) (Peisley et al, 2012).

MDA5 and RIG-I belong to the DEAD box RNA helicase family. This family of proteins forms the largest helicase family, with 37 known members in humans, and family members are characterized by the presence of the Asp-Glu-Ala-Asp motif (Fairman-Williams et al, 2010). DEAD box helicases have essential functions in RNA metabolism that include transcription, translation initiation, pre-mRNA splicing, mature mRNA transport, and the stabilization of protein complexes on RNA (Linder et al, 2011). All DEAD box proteins consist

of a highly conserved helicase core with two nearly identical domains that resemble the bacterial recombinase A (RecA) protein (Singleton et al, 2007; Caruthers et al 2002). Structural and biochemical studies have shown the ATP- and RNA-binding sites to be localized across both RecA-like domains on many DEAD box helicases. These studies showed that the two domains form an open cleft in the absence of ligand that closes upon ligand binding. The functional groups that are required for ATP hydrolysis form a complex network of interactions in the closed conformation, and productive ATP hydrolysis can only occur in this closed cleft conformation (Hilbert et al, 2009).

Despite the highly conserved structure of the catalytic core, DEAD box helicases perform a diverse range of biological functions (Linder et al, 2011). This diversity is mediated in part by auxiliary domains that flank the core and have different functional roles (Linder et al, 2011). In addition, DEAD box helicases use the energy harnessed from ATP hydrolysis to perform a variety of biochemical activities. For example, the RNA clamp protein eIF4AIII is essential for the formation of a ribonucleoprotein complex at exon-exon junctions during pre-mRNA splicing while it is in the ATP bound state. ATP hydrolysis occurs only to induce dissociation of the protein from RNA (Liu et al, 2008a). This use of ATP hydrolysis differs from RIG-I or Ded1, which use ATP hydrolysis to power translocation and local strand denaturation, respectively (Myong et al, 2009; Yang et al, 2006). A main goal of our work was to elucidate the biochemical function of ATP hydrolysis in MDA5 function.

5.3.1: Ligand recognition by RIG-I-like receptors

In the last ten years, much effort has focused on identifying the PAMPs that trigger RIG-I and MDA5 signaling. The findings indicate that MDA5 and RIG-I detect different but overlapping sets of viruses, suggesting that they play different roles in the recognition of viral molecular features (Schmidt et al, 2012). RIG-I is essential in recognizing negative-sense ssRNA viruses of the paramyxoviridae, rhabdoviridae, orthomyxoviridae, and filoviridae families and positive-sense ssRNA viruses of the flaviviridae family (Kato et al, 2006). RIG-I-deficient mouse embryonic fibroblasts (MEFs) show abolished production of IFN-β and inflammatory cytokine IL-6 upon infection with these viruses (Kato et al, 2005). In addition, infection of RIG-I depleted MEFs with negative-sense ssRNA viruses abrogates the nuclear translocation of transcription factor IRF3 (Loo et al, 2008). MDA5, on the other hand, was shown through similar gene targeting experiments to specifically sense picornaviruses, a family of positive-sense ssRNA viruses that includes the rhinovirus, echovirus, encephalomyocaditis virus, Theiler's virus, and poliovirus (Kato et al, 2006). Interestingly, RIG-I and MDA5 show a concerted response to dsRNA reoviruses and the positive-sense ssRNA West Nile and Dengue viruses (Loo et al, 2008; Fredericksen et al, 2006).

It was originally thought that the essential component for recognition by RIG-I is ssRNA bearing a 5'-triphosphate group (Hornung et al, 2006; Pichlmair et al, 2006). The 5'-triphosphate moiety is unique to viral nucleic acids since vertebrate RNA is either capped in the nucleus or post-translationally modified to remove the triphosphate. However, more recent experiments using chemically

synthesized RNA indicate that only dsRNA with a 5'-triphosphate is recognized by RIG-I (Schmidt et al, 2009). The authors suggested that previous results were due to unintended hairpin structures and double-stranded byproducts resulting from *in vitro* transcription. To identify the exact RNA characteristics that are recognized by RIG-I, a panel of RNA molecules was synthesized and tested for their ability to induce type-I IFN activation in mouse monocytes (Schlee et al, 2009). The results established that RIG-I activation requires short stretches (at least 19 nucleotides) of RNA in double-strand conformation with a 5'-triphosphate. Chemically synthesized 5'-triphosphate ssRNA or even 5'-triphosphate dsRNA that contains a missing end base pair did not induce type-I IFN expression. These results are consistent with sequence analysis data of negative-sense ssRNA viruses indicating that their genomes contain blunt short double-stranded 5' triphosphate RNA in the panhandle regions (Hofacker et al, 2004).

Recent studies have shown that RIG-I preferentially recognizes uridine/adenosine rich regions of RNA. Saito et al (2008) showed that poly-(U/A) dsRNA forms a stable complex with RIG-I and produces a greater IFN response when transfected into human Huh7 cells compared with other dsRNA sequences. This result was subsequently confirmed by Uzri et al (2009). RIG-I also exhibits cell type-specific induction of antiviral responses. While it is essential for the induction of type I IFNs in fibroblasts and conventional dendritic cells, RIG-I-deficient plasmacytoid dendritic cells produced similar levels of type I IFNs as wild-type cells (Kato et al, 2005).

Ligand recognition requirements for MDA5 have not yet been extensively studied and are not well understood. Picornavirus RNA is specifically recognized by MDA5 and not by RIG-I. The genome of picornavirus is different from other families of RNA viruses because it has a protein, VPg, attached to the 5'-end of the viral RNA (Pathak et al, 2007). This difference has led to the hypothesis that RIG-I is unable to sense picornaviral RNA due to the blockage of the 5'-triphosphate group by VPg (Wilkins et al, 2010); however, this has yet to be experimentally verified. MDA5 is the primary sensor of poly-(I:C). Gitlin et al (2006) showed that while both RIG-I and MDA5 can bind to poly-(I:C) *in vitro*, *mda5*-deficient mice exhibit a more severely impaired production of IFNs. Kato et al (2008) subsequently showed that poly-(I:C) can be converted from an MDA5 to a RIG-I ligand by shortening its length by RNase III treatment. This result led to the current hypothesis that MDA5 preferentially senses long (greater than 1 kb) dsRNA, whereas RIG-I recognizes shorter fragments.

Recent experiments with recombinantly expressed human MDA5 support the notion that MDA5 is able to sense dsRNA length. Electrophoretic mobility shift assays (EMSA) revealed that MDA5 binds to 100-bp dsRNA with high affinity and is indiscriminate towards dsRNA termini structure. dsRNA molecules with 5'-monophosphate, 5'-triphosphate, 5'-OH, and 5'- or 3'- overhangs of 1, 2, or 12 nucleotides all bound to MDA5 with similar affinities. However, the results differed when the same experiments were performed with short molecules of dsRNA. When 12-bp dsRNA was used, MDA5 was still able to bind with high affinity to dsRNA molecules with 5'-monophosphate, 5'-triphosphate, 5'-OH, but

could no longer bind to dsRNA molecules with 5'- or 3'- overhangs (Peisley et al, 2012). This result is consistent with recent experiments showing that an 8-bp dsRNA with either 5' or 3' overhangs do not bind the CTD of MDA5. The authors concluded that the blunt termini of dsRNA with or without phosphate groups are the structural motif recognized by the MDA5 CTD (Li et al, 2009).

5.3.2: Functional analysis of RIG-I-like receptors

ATPase activity has been shown to be critical to RIG-I function as the catalytically inactive K270A mutant shows abrogated antiviral signaling (Yoneyama et al, 2004). However, why ATPase activity is needed for RIG-I function is not understood. Two major hypotheses have been presented in the literature. One hypothesis posits that ATP hydrolysis enables conformational switching between active and inactive forms of RIG-I (see further discussion in Section 5.3.3), while the other hypothesis assumes that ATP hydrolysis is used to unwind dsRNA. The latter hypothesis argues that RIG-I is similar to the RNA/DNA helicases that unwind double-stranded nucleic acids by using ATP hydrolysis to power translocation along a product single strand. However, these RNA/DNA helicases show ATP hydrolysis activity upon binding to single-stranded nucleic acids, whereas the helicase domain of RIG-I exhibits ATPase activity only upon binding to dsRNA (Yoneyama et al, 2005; Takahasi et al, 2008; Cui et al, 2008).

Myong et al (2009) used protein-induced fluorescence enhancement, a technique similar to the more widely used fluorescence resonance energy transfer,

to examine whether RIG-I exhibits translocation activity. By end-labeling 25- and 40-nucleotide dsRNA with a fluorophore, RIG-I dsRNA binding events were monitored by increases in fluorescence intensity and dissociation events by decreases in fluorescence intensity. These studies revealed that in the presence of ATP, RIG-I translocates along the axis of dsRNA but does not unwind it. Interestingly, a RIG-I truncation mutant that lacks both CARDs (RIGh) showed a 15-fold accelerated movement along the dsRNA. This finding is consistent with the observation that wild-type RIG-I stimulates ATP hydrolysis to a greater extent than the RIGh mutant (Cui et al, 2008), and supports the argument that the CARDs maintain RIG-I in an autosuppressed state (see Section 5.3.3).

The mechanism of MDA5 regulation and activation is largely unknown. Another goal of our work was to investigate the oligomeric state and nature of conformational changes in MDA5 associated with viral RNA binding and the respective roles of ATP binding and hydrolysis in MDA5 function.

5.3.3: Structural analysis of RIG-I-like receptors

In the last few years, many groups have independently provided crystallographic and biophysical studies of RIG-I in its free state and bound to dsRNA. The first crystal structure of the regulatory CTD of RIG-I bound to a chemically synthesized self-complementary 12-mer 5'-ppp dsRNA revealed a binding stoichiometry of two CTDs per duplex, with one CTD at each end of the dsRNA. The 5'-ppp ends are each anchored in a basic pocket of the CTD with non-bridging phosphate oxygens in close enough proximity to hydrogen bond to six

lysine side chains of the protein (Wang et al, 2010). These interactions are believed to account for the 10-fold higher binding affinity of RIG-I for blunt-end 5'-ppp dsRNA compared to 5'-OH dsRNA (Lu et al, 2010). The crystal structure also revealed pi stacking of the 5' nitrogenous base pair to the aromatic ring of a phenylalanine side chain. This interaction is believed to provide an explanation for the preference of RIG-I for blunt-end duplexes over ssRNA (Wang et al, 2010).

More recently, crystal structures were obtained for full-length RIG-I, deltaCTD RIG-I, and the helicase domain alone (Hel), as well as a ternary complex of RIG-I with an ATP analog and a 19-mer dsRNA (Kowalinski et al, 2011). These structures, along with prior biochemical observations, have led to the currently accepted mechanism for RIG-I activation. In the absence of bound ligand, the helicase domain exists in an open conformation with the CARDs bound to the helicase insertion domain (Hel2i, composed of the final 150 amino acids of the helicase domain) and the CTD flexibly linked to the helicase. This conformation has since been termed the "autorepressed state", because the tandem CARDs are inhibited from signaling by binding to a region of the protein itself. The fact that binding of the CARDs to Hel2i is strong enough to prevent signaling is supported by biochemical evidence. When a conserved interface residue on Hel2i (Phe540 in duck RIG-I) is substituted with alanine, RIG-I exhibits a constitutively active phenotype (Kowalinski et al, 2011).

In the open conformation, the helicase domain is sterically hindered from binding dsRNA, whereas the flexible CTD is available for sensing and capturing

dsRNA. It is believed that the CTD samples different nucleic acids and only presents those for which it has a high affinity to the helicase domain for binding. This notion is consistent with two seemingly contradictory observations, namely that the CTD has a significantly higher affinity and longer off-rate for 5'-ppp dsRNA compared to 5'-OH dsRNA (Lu et al, 2010) and that the helicase domain shows little preference for 5'-ppp dsRNA over 5'-OH dsRNA (Kowalinski et al, 2011). The final binding of dsRNA to the helicase domain induces a conformational change that leads to the closed state with concomitant release of the tandem CARDs. Since the CARDs are attached to the rest of the protein through a 55-amino acid linker, they are free to signal to downstream MAVS.

Compared with RIG-I, there have been fewer structural studies of MDA5 and little is known about the structural basis of MDA5 activation. In 2009, a 1.45 Å resolution structure of the MDA5 CTD was determined by X-ray crystallography and its binding surface for dsRNA was mapped by NMR spectroscopy (Li et al, 2009). Although the amino acid sequence of the MDA5 CTD is only 24% identical to that of RIG-I, the structure revealed similar folds that contains two four-stranded antiparallel β-sheets connected by a β-hairpin. Like in RIG-I, four conserved cysteine residues arranged in tetrahedral geometry coordinate a central zinc ion (Li et al, 2009). The zinc ion is believed to be essential for maintaining the fold of the CTD as mutation of the cysteine residues result in abrogation of RNA recognition (Cui et al, 2008).

The solution structure of the MDA5 CTD was determined by NMR spectroscopy by two independent research groups (Li et al, 2009; Takahasi et al,

2009). Both groups were able to map the dsRNA binding surface by titration experiments because the ligand-free protein showed good chemical shift dispersion. The results revealed that, like RIG-I, MDA5 uses a highly conserved positively charged groove to bind dsRNA. Interestingly, superposition of the two CTDs showed that the patch of lysine residues responsible for binding to the 5'-ppp moiety of dsRNA in RIG-I is replaced by neutral amino acids in MDA5. This difference likely explains the inability of MDA5 to discriminate between 5'-ppp and 5'-OH dsRNA ends. Takahasi et al (2009) also noted the absence of a critical phenylalanine residue in MDA5 that is conserved in RIG-I. It is thought that this residue anchors the position of dsRNA in RIG-I, and its absence in MDA5 is not yet understood.

5.3.4: Type 1 diabetes and MDA5

Type 1 diabetes (T1D) is a disorder characterized by the immune-mediated destruction of insulin-producing pancreatic cells (Nakayama, 2011). T1D is believed to develop in mid-life as a consequence of a complex interaction between genetic contributions and environmental factors (Forlenza, 2011). There are approximately 15 chromosomal loci polymorphisms associated with T1D (Nejentsev et al, 2007). In 2009, a series of genome-wide association studies were performed on over 3000 families that include one or more offspring with T1D and their parents. The results revealed that individuals carrying valine at position 923 of MDA5 (CTD) have only ~50% risk of developing T1D compared to those carrying the wild-type amino acid isoleucine (Nejentsev et al, 2009). The minor

alleles of this single nucleotide polymophism (SNP) has an occurrence frequency of less than 3% (Nejentsev et al, 2009), suggesting that SNPs that reduce MDA5 function decrease the risk for T1D while wild-type MDA5 function is associated with T1D (Liu et al, 2009). These results were subsequently confirmed by Barrett et al (2009).

The finding that the I923V SNP confers a protective effect against T1D spurred interest in understanding the function and structure of the mutant protein. Functional complementation using MDA5-deficient MEFs revealed significantly reduced (~40%) ability to activate downstream proteins. This finding, along with the observation that diabetic patients have a 3.5-fold increased frequency for having enteroviral RNA in their blood, suggests that MDA5 aids the onset of T1D through viral-induced autoimmunity (Chistiakov et al, 2010). It was thought that the I923V mutant confers partial resistance to T1D by reducing the affinity of MDA5 for viral dsRNA. In contradiction, however, EMSA using 32P-labeled poly(I:C) indicated that the I923V mutant bound to poly(I:C) as strongly as wild-type MDA5 (Shigemoto et al, 2009). This result was not expected as the MDA5 CTD is associated with dsRNA binding (Shigemoto et al, 2009).

5.4: References

Afonina, E., Stauber, R, and Pavlakis, G. (1998) The human Poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm. *J Biol Chem* 273:13015-13021.

Barrett, J., Clayton, D., Concannon, P., Akolkar, B., Cooper, J., Erlich, H. Julier, C., Morahan, G., Nerup, J., Nierras, C., Piagnol, V., Pociot, F., Schuilenburg, H., Smyth, D., Chistiakov, D., Voronova, N., Savost'Anov, K., and Turakulov, R. (2010) Loss-of-function mutations E6 27X and I923V of IFIH1 are associated with lower poly(I:C)-induced interferon-B production in peripheral blood mononuclear cells of type 1 diabetes patients. *Hum Immunol* 71:1128-1134.

Berke, I., and Modis, Y. (2012) MDA5 cooperatively forms dimmers and ATP-sensitive filaments upon binding double-stranded RNA. *EMBO J* 31:1714-1726.
Caruthers, J., and McKay, D. (2002) Helicase structure and mechanism. *Curr Opin Struct Biol* 12:123-133.

Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K., Krug, A., and Hopfner, K. (2008) The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Mol Cell* 29:169-179.

Fairman-Williams, M., Guenther, U., and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 3:419-429.

Forlenza, G., and Rewers, M. (2011) The epidemic of type 1 diabetes: what is it telling us? *Curr Opin Endocrinol Diabetes Obes* 18:248-251.

Fredericksen, B., and Gale, M. (2006) West Nile virus evades activation of interferon regulatory factor 3 through RIG-I-dependent and independent pathways without antagonizing host defense signaling. *J Virol* 80:2913-2923.

Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R., Diamond, M., and Colonna, M. (2006) Essential role of Mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc Natl Acad Sci* 103:8459-8464.

Hilbert, M., Karow, A., and Klostermier, D. (2009) The mechanism of ATP-dependent RNA unwinding by DEAD-box proteins. *Biol Chem* 390:1237-1250.

Hoebe, K., Du, X., Georgel, P., Janssen, E., Tabeta, K., Kim, S., Goode, J., Lin, P., Mann, N., Mudd, S., Crozat, K., Sovath, S., Han, J., and Beutler, B. (2003) Identification of Lps2 as a key transducer of MyD88-independent TIR signaling. *Nature* 424:743-748.

Hofacker, I., Stadler, P., and Stocsits, R. (2004) Conserved RNA secondary structures in viral genomes: a survey. *Bioinformatics* 20:1495-1499.

Honda, K., Takaoka, A., and Taniguchi, T. (2006) Type I interferon gene induction by the interferon regulatory factor family of transcription factors. *Immunity* 25:349-360.

Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K., Schlee, M., Endres, S., and Hartmann G. (2006) 5'-triphosphate RNA is the ligand for RIG-I. *Science* 314:994-997.

Hornung, V., Rothenfusser, S., and Britsch, S. (2002) Quantitative expression of toll-like receptor 1-10 mRNA in cellular subsets of human peripheral blood mononuclear cells and sensitivity to CpG oligodeoxynucleotides. *J Immunol* 168:4531-4537.

Iwasaki, A., and Medzhitov, R. (2004) Toll-like receptor control of the adaptive immune responses. *Nat Immunol* 5:987-995.

Jiang, Z., Mak, T., Sen, G., and Li, X. (2004) Toll-like receptor 3-mediated activation of NF-kappaB and IRF3 diverges at Toll-IL-1 receptor domain-containing adapter inducing IFN-beta. *Proc Natl Acad Sci USA* 101:3533-3588.

Kaisho, T., and Akira, S. (2006). Toll-like receptor function and signaling. *J Allergy Clin Imunol* 117:979-987.

Kariko, K., Ni, H., Capodici, J., Lamphier, M., and Weissman, D. (2004) mRNA is an endogenous ligand for Toll-like receptor 3. *J Biol Chem* 279:12542-12550.

Kato, H., Sato, S., Yoneyama, M., Yamamoto, M., Uematsu, S., Matsui, K., Tsujimura, T., Takeda, K., Fujita, T., Takeuchi, O., and Akira, S. (2005) Cell type-specific involvement of RIG-I in antiviral response. *Immunity* 23:19-28.

Kato H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C., Sousa, C., Matsuura, Y., Fujita, T., and Akira, S. (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441:101-105.

Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T., Fujita, T., and Aira, S. (2008) Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med* 205:1601-1610.

Kowalinski, E., Lunardi, T., McCarthy, A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D., and Cusack, S. (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147:423-435.

Lee, H., Dunzendorfer, S., Soldau, K., and Tobias, P. (2006) Double-stranded RNA-mediated TLR3 activation is enhanced by CD14. *Immunity* 24:153-163.

Linder, P., and Jankowsky, E. (2011) From unwinding to clamping – the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12:505-516.

Liu, F., Putnam, A, and Jankowsky, E. (2008a) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci USA* 105:20209-20214.

Liu, L., Botos, I., Wang, Y., Leonard, J., Shiloach, J., Segal, D., and Davies, D. (2008) Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320:379-381.

Liu, S., Wang, H., Jin, Y., Podolsky, R., Reddy, M., Pedersen, J., Bode, B., Reed, J., Steed, D., Anderson, S., Yang, P., Muir, A., Steed, L., Hopkins, D., Huang, Y., Purohit, S., Wang, C., Steck, A., Montemari, A., Eisenbarth, G., Rewers, M., and She, J. (2009) IFIH1 polymorphisms are significantly associated with type 1 diabetes and IFIH1 gene expression in peripheral blood mononuclear cells. *Hum Mol Gen* 18:358-365.

Loo, Y., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira Sh., Gill, M., Garcia-Sastre, A., Katze M., and Gale, M. (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol* 82:335-345.

Lu, C., Xu, H., Ranjith-Kumar, C., Brooks, M., Hou, T., Hu, F., Herr, A., Strong, R., Kao, C., and Li, P. (2010) The structural basis of 5'-triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. *Structure* 18:1032-1043.

Matsumoto, M., Kikkawa, S., Kohase, M., Miyake, K., and Seya, T. (2002) Establishment of a monoclonal antibody against human Toll-like receptor 3 that blocks double-stranded RNA-mediated signaling. *Biochem Biophys Res Comm* 239: 1364-1369.

Matsumoto, M., Funami, K., and Tanabe, M. (2003) Subcellular localization of Toll-like receptor 3 in human dendritic cells. *J Immunol* 171:3154-3162.

McBride, S., Hoebe, K., Georgel, P, and Janssen, E. (2006) Cell-associated double-stranded RNA enhances antitumor activity through the production of type I IFN. *J Immunol* 177:6122-6128.

Meylan, E., Burns, K., Hofmann, K., Blancheteau, V., Martinon, F., Kelliher, M., and Tschopp, J. (2004) RIG1 is an essential mediator of Toll-like receptor 3-induced NF-kappa B activation. *Nat Immunol* 5:503-507.

Meylan, E., Juerg, T., and Michael, K. (2006) Intracellular pattern recognition receptors in the host response. *Nature* 442:39-44.

Muzio, M., Bosisio, D., and Polentarutti, N. (2000) Differential expression and regulation of toll-like receptors (TLR) in human leukocytes: Selective expression of TLR3 in dendritic cells. *J Immunol* 164:5998-6004.

Myong, S., Cui, S., Cornish, P., Kirchhofer, A., Gack, M., Jung, J., Hopfner, K., and Ha, T. (2009) Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. *Science* 323:1070-1074.

Nakayama, M. (2011) Insulin as a key autoantigen in the development of type 1 diabetes. *Diabetes Metab Res Rev* 27:773-777.

Nejentsev, S., Howson, J., Walker, N., Szeszko, J., Field, S., Stevens, H., Reynolds, P., Hardy, M., King, E., Masters, J., Hulme, J., Maier, L., Smyth, D., Bailey, R., Cooper, J., Ribas, G., Campbell, R., Clayton, D., and Todd, J. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 450:887-U19.

Nejentsev, S., Walker, N., Riches, D., Egholm, M., and Todd, J. (2009a) Rare varients of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324:387-389.

Orinska, Z., Bulanova, E., and Budagian, V. (2005) TLR3-induced activation of mast cells modulates CD8+ T-cell recruitment. *Blood* 106:978-987.

Pathak, H., Arnold, J., Wiegand, P., Hargittai, M., and Cameron, C. (2007) Picornavirus genome replication – Assembly and organization of the VPg uridylylation ribonucleoprotein (initiation) complex. *J Biol Chem* 282:16202-16213.

Peisley, A., Wu, B., and Hur, S. (2012). Personal Communication.

Pichlmair, A., Schulz, O., Tan, C., Naslund, T., Liljestrom, P., Weber, F., and Reis e Sousa, C. (2006) RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* 314:997-1001.

Rehwinkel, J., and Reis e Sousa, C. (2010) RIGorous detection: exposing virus through RNA sensing. *Science* 327:284-286.

Rock, F., Hardiman, G., Timans, J., Kastelein, R., and Bazan, J. (1998) A family of human receptors structurally related to Drosophila toll. Proc Natl Acad Sci 95:588-593.

Rodriguez, R., Lopez-Vazquez, A., and Lopez-Larrea, C. (2012) Immune systems evolution. *Adv Exp Med Biol* 739:237-251.

Saito, T., Owen, D., Jiang, F., Marcotrigiano, J., and Gale, M. (2008) Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* 454:523-527.

Schmidt, A., Rothenfusser, S., and Hopfner, K. (2012) Sensing of viral nucleic acids by RIG-I: from translocation to translation. *Eur J Cell Biol* 91:78-85.

Schlee, M., Roth, A., Hornung, V., Hagmann, C., Wimmenauer, V., Barche, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G., Juranek, S., Kato, H., Kawai, T., Poeck, H., Fitzgerald, K., Takeuchi, O., Akira, S., Tuschl, T., Latz, E., Ludwig, J., and Hartmann, G. (2009) Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity* 31:25-34.

Schulz, O., Diebold, S., Chen, M., Naslund, T., Nolte, M., Alexopoulou, L., Azuma, Y., Flavell, R., Lilijestrom, P., and Sousa R. (2005) Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature* 433:887-892.

Shigemoto, T., Kageyama, M., Hirai, R., Zheng, J., Yoneyama, M., and Fujita, T. (2009) Identification of loss of function mutations in human genes encoding RIG-I and MDA5. *J Biol Chem* 284:13348-13354.

Singleton, M., Dillingham, M., and Wigley, D. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76:23-50.

Tabiasco, J., Devevre, E., and Rufer, N., (2006) Human effector CD8+ T lymphocytes express TLR3 as a functional coreceptor. *J Immunol 177*:8708-8713.

Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale, M., Inagaki, F., and Fujita, T. (2008) Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell* 29:428-440.

Takeuchi, O., and Shizuo, A. (2008) MDA5/RIG-I and virus recognition. *Curr Opin Immunol* 20:17-22.

Tatematsu, M., Ischii, A., Oshiumi, H., Horiuchi, M., Inagaki, F., Seya, T., and Matsumoto, M. (2010) A molecular mechanism for Toll-IL-1 receptor domain-containing adaptor molecule-1-mediated IRF-3 activation. *J Biol Chem* 285:20128-20136.

Uzri, D., and Gehrke, L. (2009) Nucleotide sequences and modifications that determine RIG-I/RNA binding and singaling activities. *J Virol* 83:4174-4184.

Wang, Y., Liu, L., Davies, D., and Segal, D. (2010a) Dimerization of Toll-like receptor 3 (TLR3) is required for ligand binding. *J Biol Chem* 285:36836-36841.

Wang, Y., Ludwig, J., Schuberth, C., Goldneck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., Hartmann, G., and Patel, D. (2010b) Structural and functional insight into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat Struct Mol Biol* 17:781-787.

Wang, Y., Swiecki, M., McCartney, S., and Colonna, M. (2011) dsRNA sensors and plasmacytoid dendritic cells in host defense and autoimmunity. *Immunol Rev* 243:74-90.

Wilkins, C., and Gale, M, Jr. (2010) Recognition of viruses by cytoplasmic sensors. *Curr Opin Immunol* 22:41-47.

Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., and Akira, S. (2003) Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science* 301:640-643.

Yang, Q., and Jankowsky, E. (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* 13:981-986.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., and Akira, S. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* 5:730-737.

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y., Gale, M., Akira S., Yonehara, S., Kato, A., and Fujita, T. (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol* 175:2851-2858.

Yoneyama, M., and Takashi, F. (2007) Function of RIG-I-like receptors in antiviral innate immunity. *J Biol Chem* 282:15315-15318.

Yusupov, M., Yusupova, G., Baucom, A., Lieberman, K., Earnest, T., Cate, J., and Noller, H. (2001) Crystal structure of the ribosome at 5.5 angstrom resolution. *Science* 292:83-896.

Zarember, K., and Godowski, P. (2002) Tissue expression of human toll-like receptors and differential regulation of toll-like receptor mRNAs in leukocytes in response to microbes, their products, and cytokines. *J Immunol* 168:554-561.

Zhong, B., Wang, Y., and Shu, H. (2010) Regulation of virus-triggered type I interferon signaling by cellular and viral proteins. *Immunity* 5:12-31.

Chapter 6: Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition

This chapter is a reprint of the following publication: Peisley, A., Lin, C., Wu, B., Orme-Johnson, M., Liu, M., Walz, T., and Hur, S. (2011) Cooperative assembly and dynamic disassembly of MDA5 filaments for viral dsRNA recognition *Proc Natl Acad Sci USA* 108:21010-21015. I contributed to this work by performing all the single particle electron microscopy experiments. The supplemental information that accompanies this publication is not reprinted here due to its large size.

MDA5, an RIG-I-like helicase, is a conserved cytoplasmic viral RNA sensor, which recognizes dsRNA from a wide-range of viruses in a length-dependent manner. It has been proposed that MDA5 forms higher-order structures upon viral dsRNA recognition or during antiviral signaling, however the organization and nature of this proposed oligomeric state is unknown. We report here that MDA5 cooperatively assembles into a filamentous oligomer composed of a repeating segmental arrangement of MDA5 dimers along the length of dsRNA. Binding of MDA5 to dsRNA stimulates its ATP hydrolysis activity with little coordination between neighboring molecules within a filament. Individual ATP hydrolysis in turn renders an intrinsic kinetic instability to the MDA5 filament, triggering dissociation of MDA5 from dsRNA at a rate inversely proportional to the filament length. These results suggest a previously unrecognized role of ATP hydrolysis in control of filament assembly and disassembly processes, thereby autoregulating the interaction of MDA5 with dsRNA, and provides a potential basis for dsRNA length-dependent antiviral signaling.

6.1: Introduction

Effective immune defense against viral infection depends on accurate and robust detection of viral RNAs by pattern recognition receptors in the innate immune system. RIG-I and MDA5 are conserved viral RNA receptors in vertebrates, sharing similar domain architectures and approximately 30% sequence identity. They are functional in the cytoplasm and responsible for detection of viral RNAs present during viral infection in a broad range of tissues (Wilkins et al, 2010). Upon viral RNA recognition, RIG-I and MDA5 interact with a common signaling adaptor, MAVS and activate NF-kB and IRF3/7 signaling pathways, resulting in the expression of type I interferon and proinflammatory cytokines (Rehwinkel et al, 2010).

Previous studies suggested that RIG-I and MDA5 recognize largely distinct groups of viruses through their divergent RNA specificity (Loo et al, 2008; Kato et al, 2006). RIG-I detects a variety of positive and negative strand viruses through recognition of a 5′-triphosphate group and blunt ends of genomic RNAs (Loo et al, 2008; Baum et al, 2010; Schlee et al, 2009). By contrast, MDA5 detects several positive strand and dsRNA viruses such as picornaviruses and reoviruses through recognition of dsRNA replication intermediates and genomic dsRNAs (Loo et al, 2008; Kato et al, 2006; McCartney et al, 2008). This viral RNA recognition by MDA5 is independent of 5′-triphosphate or blunt end, but instead depends on dsRNA length in a range of approximately 1–7 kb (Kato et al, 2008).

The precise mechanism for length-dependent dsRNA recognition by MDA5 is poorly understood. Previous studies suggest that ATP hydrolysis could play an important role in both interferon signaling and RNA specificity (Wilkins et al, 2010; Rehwinkel et al, 2010). The current model of MDA5 (or RIG-I) mediated signal activation is that binding of viral RNA to MDA5 triggers ATP hydrolysis in the conserved helicase domain, which then alters the receptor conformations or accessibility of the signaling domain (a tandem caspase activation recruitment domain) to allow its interaction with MAVS (Wilkins et al, 2010; Zeng et al, 2010). The proposed role of ATP hydrolysis as a conformational switch for signaling is supported by the requirement of ATP in the reconstituted signaling system of RIG-I (Hou et al, 2011; Zeng et al, 2010) and the observation that mutations in the active site for ATP hydrolysis either constitutively activate or inactivate interferon signaling by RIG-I and MDA5 (Bamming et al, 2009; Yoneyama et al, 2004). In addition, induction of ATP hydrolysis by dsRNA has been shown to correlate with stimulation of interferon signaling (Schlee et al, 2009; Cui et al, 2008).

It has been proposed that MDA5 (and RIG-I) forms a higher-order structure during viral RNA recognition for efficient antiviral signaling (Andrejeva et al, 2004; Saito et al, 2007; Binder et al, 2011). Receptor oligomerization or clustering, as demonstrated in other receptors, such as inflammasomes, Toll-like receptors, T- and B-cell receptors, is often important for ligand sensitivity and downstream signaling efficiency (Davis et al, 2011; Sigalov et al, 2010). To understand the assembly architecture of the postulated higher order structure and

a potential role for ATP hydrolysis in structural reorganization, we used a combination of EM, biochemical, and kinetic analyses. We demonstrate here that MDA5 assembles into a filamentous oligomer along the length of dsRNA and dynamically regulates its disassembly process through ATP hydrolysis in a length-dependent manner. This study provides a mechanism of how MDA5 both recognizes dsRNA and autoregulates its antiviral signaling potential in a dsRNA length-dependent manner.

6.2: Results

6.2.1: MDA5 binding to dsRNA is cooperative

Antiviral signaling through MDA5 is stimulated by dsRNA but not by ssRNA or dsDNA (Kato et al, 2008; Ablasser et al, 2009). To examine whether the differential binding affinities of MDA5 are responsible for dsRNA specificity, we purified recombinant human MDA5 from *Escherichia coli* and examined its interaction with 112 bp dsRNA and dsDNA, and 112 nt ssRNA using an EMSA. These model nucleic acids were derived from MDA5 cDNA, and no significant secondary structure was predicted for the ssRNA (mFold).

EMSA revealed that MDA5 is capable of high-affinity binding to dsRNA, dsDNA, and ssRNA with dissociation constants (K_d) of 22, 33, and 57 nM, respectively (Figure 6.1). Despite the comparable binding affinities between the nucleic acid types, only the interaction between MDA5 and dsRNA is highly cooperative. By EMSA, MDA5 addition to dsRNA resulted in a bimodal

distribution of mostly either free (RNA-only) or saturated (fully complexed) products (Figure 6.1A). By contrast, MDA5 complexed with ssRNA or dsDNA exhibits a broad distribution of intermediate sizes, indicative of less cooperative binding. In a quantitative analysis of the EMSA results (Figure 6.1B), the dsRNA binding curve is sharply sigmoidal with a Hill coefficient of 4.0, whereas the binding curves for ssRNA and dsDNA are more hyperbolic with Hill coefficients of 1.5 and 2.7, respectively. EMSA analysis of another 112 bp dsRNA with different sequence reveals a similar bimodal distribution of RNA upon MDA5 binding and sigmoidal binding curve, suggesting that highly cooperative binding of MDA5 is dependent on dsRNA structure but not sequence.

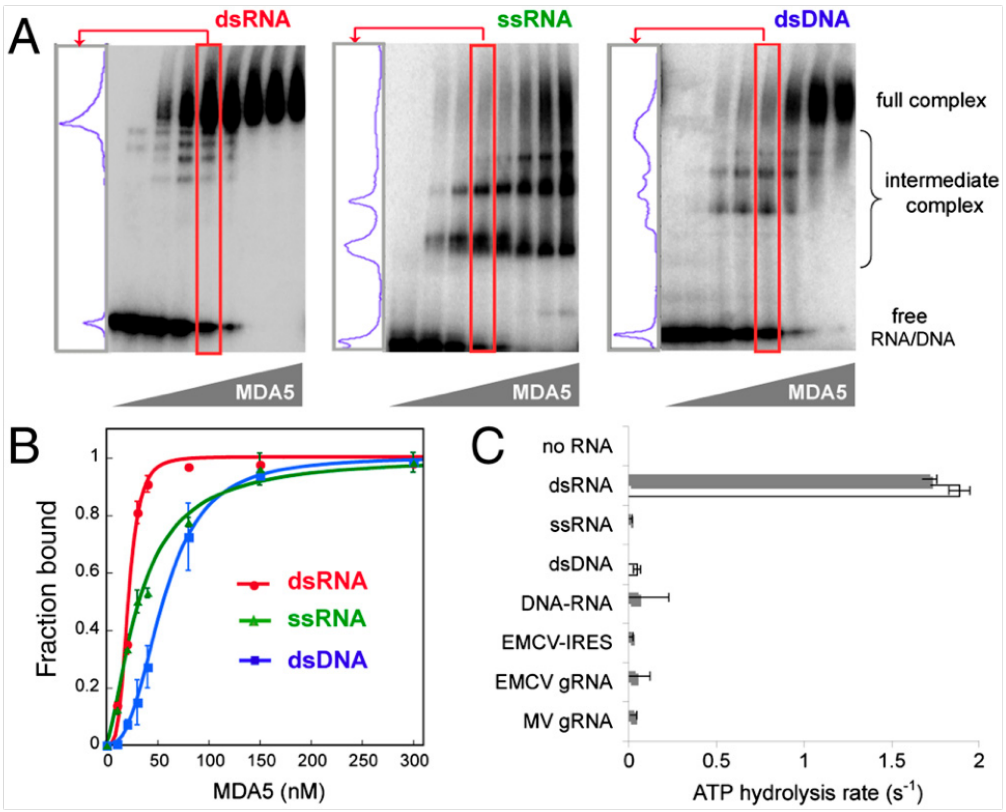

Figure 6.1: Only binding of MDA5 to dsRNA is cooperative and stimulates ATP hydrolysis. (A) EMSA with 112 bp/nt dsRNA, ssRNA, and dsDNA. A lane in which approximately 50% of MDA5 is bound to the respective nucleic acid (red box) was chosen for each sample, and the distribution of the complexes is displayed in the left-hand panel. (B) The EMSA results in A were fit to the Hill equation to obtain dissociation constant (Kd) and Hill coefficient (Nh). Estimated Kd and Nh are 22 nM and 4.0 for dsRNA, 33 nM and 1.5 for ssRNA, and 57 nM and 2.7 for dsDNA, respectively. Plotted values are mean ± SD (n = 2). (C) ATP hydrolysis rates (mean ± SD, n = 4) of MDA5 (0.3 μM) free of nucleic acids, MDA5 bound to 112 bp/nt dsRNA, ssRNA, dsDNA, DNA– RNA hybrid (▪ sequence 1, □ sequence 2, see Table 6.1), IRES (1 kb), and genomic RNA (gRNA) (~7 kb) of EMCV and MV. Equivalent mass concentration (4.8 μg/mL) was used for all nucleic acids.

6.2.2: Only dsRNA stimulates ATP hydrolysis by MDA5

We next asked whether the ATP hydrolysis activity of MDA5 is dependent on the type of bound nucleic acid. We measured the ATP hydrolysis rate of MDA5 using two sets of 112 bp/nt dsRNAs, ssRNAs, or dsDNAs derived from two different

sequences at an equivalent nucleotide concentration (4.8 μg/mL, i.e., 71.4 nM 112 bp dsRNA). The ATP hydrolysis activity of MDA5 was strongly stimulated by dsRNAs, but not by ssRNAs or dsDNAs (Figure 6.1C), consistent with previous observations (Kang et al, 2002). A 112 bp RNA-DNA hybrid and 7 kb genomic ssRNAs derived from two picornaviruses, encephalomyocarditis virus (EMCV) and Mengo virus (MV), which contain highly structured internal ribosome entry sites (IRES) were also unable to stimulate ATP hydrolysis (Figure 6.1C). This result suggests that dsRNA is the only type of nucleic acid that supports MDA5 ATP hydrolysis.

6.2.3: MDA5 forms filamentous oligomers along the length of dsRNA

To better understand the nature of the interaction of MDA5 with dsRNA in comparison with ssRNA and dsDNA, we visualized complexes of MDA5 with nucleic acids using negative-stain EM. We included the cross-linker reagent, bismaleimide (BMH), and ATP analogue, ADP-aluminum fluoride (ADP•AlF4) in the sample preparation to improve complex homogeneity and stability. The migration of the BMH-cross-linked MDA5：dsRNA complex was identical to that observed for the native complex by EMSA.

The electron micrographs of the MDA5：dsRNA：ADP•AlF4 complex revealed that MDA5 forms a filamentous oligomer upon binding to 112 bp dsRNA (Figure 6.2A). Rod-like oligomers were rarely observed with ssRNA and dsDNA, (<3% of dsRNA), and were more heterogeneous in size. Each nucleoprotein filament of MDA5 with 112 bp dsRNA is 32-37-nm long. RNA is

not directly visible in the negatively stained samples, but the agreement between the observed filament length and predicted length of 112 bp dsRNA (~33 nm, ~2.8 Å helical rise per base pair) supports the idea that the filament is arranged along the length of dsRNA. To test whether the filament extends with RNA length, we examined the complex formed with 512 bp dsRNA. Longer filaments of 150-170 nm were observed, consistent with the expected length of a 512 bp dsRNA molecule (Figure 6.2A). Thus, dsRNA promotes MDA5 filament assembly along the length of dsRNA.

To obtain a more detailed understanding of the filament assembly, we calculated class averages of the 112 bp complex. Each filament consists of seven or eight segments (Figure 6.2B). This result is in agreement with the number of intermediate complex bands in EMSA (Figure 6.2C), suggesting that each segment corresponds to the minimal binding unit. The tight packing of these segments along the filament is consistent with cooperative binding observed from EMSA. Each unit occupies approximately 1.3 helical turns of the RNA (15 bp), but no clear indications of helical symmetry could be detected in the averages or their Fourier transforms. Occasionally segments in the filaments slightly deviate from the central axis, causing a modest level of curvature in the filament; this curvature may result from heterogeneity in the ADP•AlF4 occupancy or in the dimer-dimer interface. A more detailed analysis of the structure was not feasible because of the low resolution of the averages.

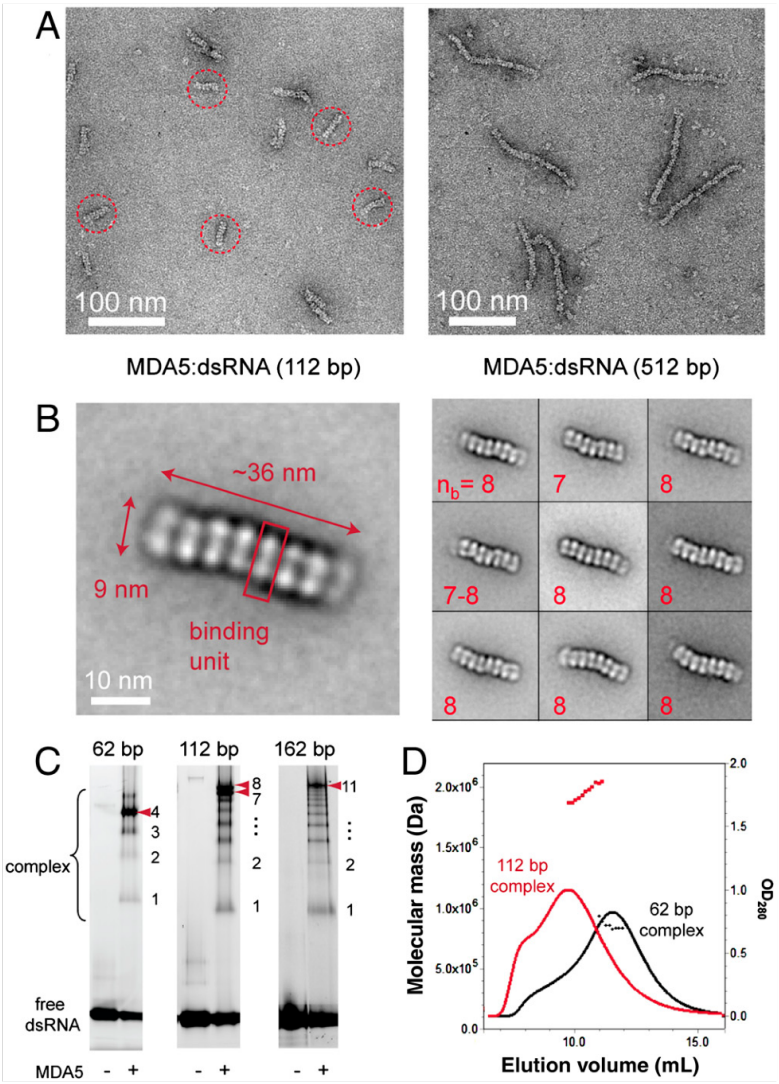

Figure 6.2: MDA5 forms filamentous oligomers along the length of dsRNA. (A) Representative electron micrographs of negatively stained MDA5 in complex with 112 and 512 bp dsRNA in the presence of ADP•AlF_4 . Representative particles used for averaging are circled in red. (B) Representative class averages of MDA5 in complex with 112 bp dsRNA. The red box in the magnified view corresponds to a minimum binding unit of MDA5 in EMSA (C). The number of binding units per complex (n_b) is displayed at the bottom left for each average. (C) EMSA of 62, 112, and 162 bp dsRNAs with MDA5 (0.3 μM) visualized with SybrGold stain to enhance the intermediate complexes that are otherwise difficult to detect. Enumeration of intermediate complexes reveals that the full complex contains 4, 7–8, and 11 binding units of MDA5 for 62, 112, and 162 bp dsRNAs, respectively. (D) Multiangle light scattering (−OD280 and ▪ M_r estimation) suggests that the molecular mass of the full complexes formed on 62 and 112 bp dsRNAs are 854 (±43) kDa and 1,878 (±13) kDa, respectively.

To determine the stoichiometry of the MDA5:dsRNA complex, we measured the molecular mass of the BMH-cross-linked filament using multiangle light scattering (MALS) (Figure 6.2D). We found that MDA5 forms complexes of 854 kDa with 62 bp and 1,878 kDa with 112 bp dsRNA, which corresponds to 6.7 and 15.5 MDA5 monomers per RNA molecule, respectively. From EMSA, we observed approximately 4 and 7-8 complex bands for 62 and 112 bp dsRNAs, respectively (Figure 6.2C). This result suggests that each segment in EM averages and each binding unit in EMSA represents an MDA5 dimer. Interestingly, our MALS analysis and a previous report (Cui et al, 2008) suggest that MDA5 is monomeric in solution, indicating that MDA5 dimerizes upon dsRNA binding. Although a complex of a single MDA5 monomer bound to dsRNA is often undetectable by EMSA with dsRNA longer than 62 bp, it is apparent with 15 bp dsRNA. This finding suggests that dimerization is cooperative only in the context of filament.

6.2.4: ATP hydrolysis occurs by monomers of MDA5 throughout the filament

The unique property of dsRNA in promoting both filament assembly and ATP hydrolysis by MDA5 raised the question of whether filament formation is required for ATP hydrolysis, and whether ATP hydrolysis of an MDA5 molecule depends on its position within a filament. To address these questions, we examined ATP hydrolysis activity of MDA5 with 15, 112, and 512 bp dsRNAs. To account for the low binding affinity of MDA5 to 15 bp (K_d of 949 nM) and to measure the intrinsic ATP hydrolysis activity of MDA5 (instead of binding or

dissociation kinetics which often dominate the steady-state reaction rate at low enzyme or substrate concentration), we measured ATP hydrolysis rates using high concentration of MDA5 (4 μM) or dsRNA (40 μg/mL, i.e., 10.7 μM MDA5 binding site) that would saturate 0.6 μg/mL dsRNA or 0.3 μM MDA5, respectively. Use of the same total base-pair (i.e., mass) concentration also ensures the same amount of MDA5 binding sites for all reactions. Note that 1 μM would occupy 3.75 μg/mL of dsRNA regardless of the RNA length (i.e., MDA5 dimer per 15 bp). We found that 15 bp dsRNA robustly stimulates MDA5 ATP hydrolysis (Figure 6.3A) and its rate is comparable to that of 112 and 512 bp. These observations suggest that filament formation is not required for ATP hydrolysis and that an individual dimer can hydrolyze ATP with little positional preference within a filament.

To examine the interdependence of the ATP hydrolysis activity of the two MDA5 monomers in a dimeric binding unit or MDA5 molecules within a filament, we measured the ATP hydrolysis activity of heterodimers or heterofilaments formed on 15 bp or 512 bp dsRNA by a mixture of WT and catalytically inactive mutants. We used alanine mutants of K335 and D443, the two conserved residues important for ATP binding and hydrolysis (Linder & Jankowsky, 2011), because these mutants do not hydrolyze ATP but retain high affinities to dsRNA. We confirmed that WT and K335A or D443A form heterodimers and heterofilaments using EMSA. We premixed WT and the mutants at various ratios with a fixed total MDA5 concentration, and formed a complex with a 15 or 512 bp dsRNA with the same two conditions used in Figure

6.3A (Figure 6.3B). In all cases, the ATP hydrolysis rate per WT (derived from the relative rate in Figure 6.3B divided by the WT concentration) remained constant. These results suggest that the intrinsic ATP hydrolysis activity of a WT subunit is unaffected by the catalytic activity of its dimeric partner or neighboring molecules within a filament.

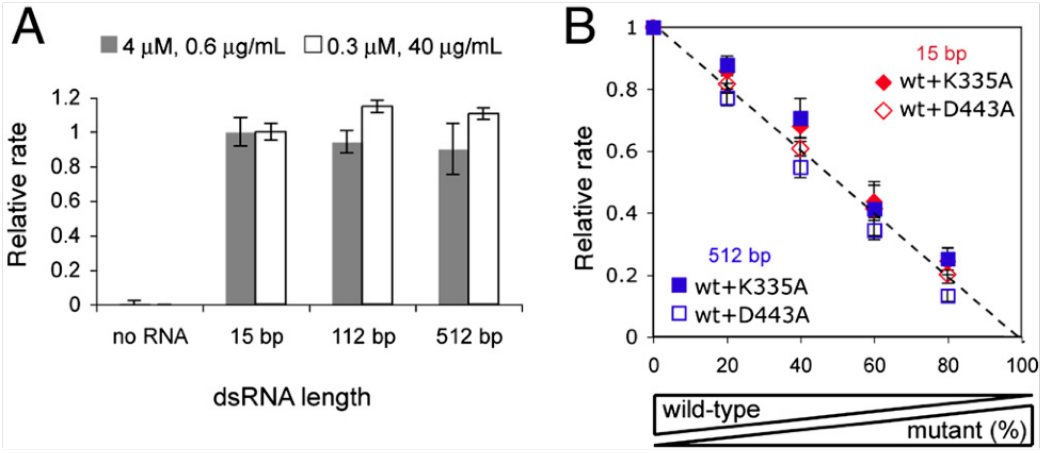

Figure 6.3: Intrinsic ATP hydrolysis activity of MDA5 is independent of filament length or catalytic activity of neighboring molecules. (A) Relative rates of ATP hydrolysis (mean ± SD, n = 3) of MDA5 (4 or 0.3 μM) with and without 15, 112, and 512 bp dsRNAs (0.6 or 40 μg⁄mL, equivalent to 0.16 or 10.7 μM MDA5 binding site), with respect to 15 bp rates. (B) Relative rates of ATP hydrolysis of mixtures of WT MDA5 and K335A or D443A at indicated ratios with total MDA5 concentration fixed at 0.3 μM. The protein mixture was incubated with 40 μg⁄mL (10.7 μM MDA5 binding site) of 15 or 512 bp dsRNA. Rates were normalized against that with 100% WT.

6.2.5: ATP hydrolysis by MDA5 promotes dissociation from dsRNA

Previous studies with several helicases suggested that ATP hydrolysis can trigger their dissociation from RNA (Linder & Jankowsky, 2011; Cordin et al, 2006). To test the effect of ATP hydrolysis on the interaction between MDA5 and dsRNA, we performed a time-dependent EMSA, in which the time evolution of the level of MDA5:dsRNA complex was monitored in the presence of a protein trap, heparin, which binds MDA5 but does not stimulate ATP hydrolysis. MDA5 was preincubated with 512 bp RNA, and the reaction was initiated by adding a mixture of ATP and a 166-fold excess of heparin relative to RNA. The ability of heparin to trap MDA5 at this ratio was confirmed from EMSA in which a premixture of MDA5 with heparin prevents MDA5 binding to dsRNA. In the absence of ATP (Figure 6.4A), MDA5:dsRNA complex is stable with less than 30% of conversion of total dsRNA to free dsRNA within 5 min. Similar stability was observed for both catalytic mutants, K335A and D443A, in the presence of ATP. This stability is in contrast to WT MDA5 with ATP where approximately 50% of dsRNA was free of MDA5 within 15 s (Figure 6.4A). To examine whether the rapid dissociation of WT is due to hydrolysis or binding of ATP, we replaced ATP by a nonhydrolyzable ATP analogue, β,γ-methyleneadenosine 5′-triphosphate (ADPCP) in the dissociation reaction. ADPCP binds to MDA5 as evidenced by its inhibition of ATP-hydrolysis activity but does not interfere with electrophoresis which is the case for ADP•AlF4. In the presence of ADPCP, approximately 50% of total dsRNA was converted to free RNA in 5 min, as

opposed to 15 s for ATP (Figure 6.4A). Thus, the ATP-dependent rapid dissociation WT MDA5 is primarily due to ATP hydrolysis.

Consistent with the above result, a single-round ATP hydrolysis assay also revealed similar dissociation kinetics during ATP hydrolysis. We monitored ATP hydrolysis by MDA5 in complex with 512 bp dsRNA in the presence and absence of heparin. The initial rates for both reactions were comparable. However, the rate with heparin declined to a halt within 1 min, whereas the rate without heparin remained constant over 3 min (Figure 6.4B). The overlapping time interval for ATP hydrolysis and dissociation kinetics suggests that the rapid decrease in the ATP hydrolysis rate in the heparin reaction is due to MDA5 dissociation from dsRNA.

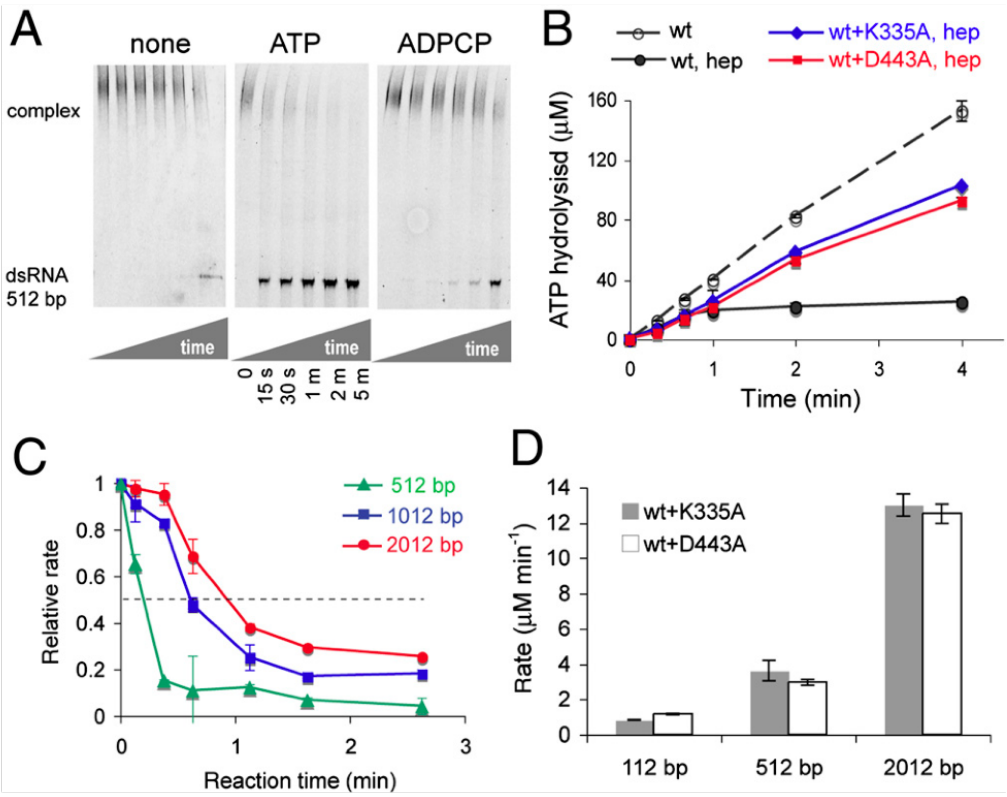

Figure 6.4: Disassembly of MDA5 filament is both ATP and RNA length-dependent. (A) Time-dependent EMSA monitoring of MDA5:RNA complex dissociation. The reaction was initiated by adding a mixture of 2 mM ATP (or ADPCP or no ligand) and 200 μg⁄mL heparin to MDA5 (1.35 μM) in complex with 512 bp dsRNA (1.2 μg⁄mL, 0.32 μM MDA5 binding site), quenched on ice at indicated time points and analyzed on native gel. (B) Single-round ATP hydrolysis kinetics (mean ± SD, n = 3) of WT MDA5 (0.1 μM) with and without K335A or D443A (0.1 μM) bound to 512 bp dsRNA (1.2 μg⁄mL) with and without heparin (hep, 200 μg⁄mL). (C) Time evolution of the ATP hydrolysis rate derived from the single-round kinetics (Fig. 6.10 F) of WT MDA5 (0.2 μM) bound to 512, 1,012, and 2,012 bp dsRNAs (1.2 μg⁄mL) using the finite difference method. The rates were normalized against the initial rate during the first 15 s. (D) Single-round ATP hydrolysis rate of WT MDA5 (0.1 μM) with K335A or D443A (0.1 μM) bound to 512, 1,012, and 2,012 bp dsRNAs (1.2 μg⁄mL) in the presence of heparin (200 μg⁄mL) during 4–10 min. No ATP hydrolysis was observed without the mutants during this time period.

6.2.6: Catalytic mutants inhibit ATP-dependent dissociation of wild-type MDA5 from dsRNA

To test interdependence of dissociation among individual MDA5 molecules within a filament, we examined the effect of K335A and D443A on ATP-dependent dissociation of WT MDA5. We first mixed WT MDA5 with the mutants, and performed the single round kinetic assays using 512 bp dsRNA. Unlike the reaction of WT alone with heparin, which stops within 1 min, the reactions with K335A and D443A progressed over 4 min at a rate 70% of that of WT alone without heparin. This result suggests that 70% of WT MDA5 remained bound to dsRNA over 4 min in the presence of K335A or D443A. This interpretation is supported by a pull-down assay with biotinylated RNA, which showed that WT copurifies with dsRNA after 3 min incubation with ADPCP but not with ATP. Addition of K335A/D443A to the ATP reaction restores the level of WT copurified with dsRNA. Thus, the catalytic mutants stabilize the WT complex and dissociation of individual MDA5 molecules is interdependent.

6.2.7: ATP-driven MDA5 dissociation is filament length-dependent

We next asked whether ATP-dependent MDA5 dissociation depends on the length of the filament. If every ATP hydrolysis event were equally capable of dissociating an individual MDA5 molecule from both filament interiors and ends, the dissociation rate would be independent of the initial filament length. This prediction would also be true if a filament disassembles cooperatively, either as an entire filament or in smaller fragments, when a fraction of MDA5 monomers

hydrolyzing ATP at a time reaches a certain threshold that is independent of filament length (i.e., linearly increasing stability). Alternatively, if filament disassembly occurs from an end as individual monomers or dimers, or disassembles cooperatively with a higher threshold for longer filaments (i.e., nonlinearly increasing stability), MDA5 would dissociate more slowly from longer dsRNA. We compared the dissociation rate of WT MDA5 using the single-round kinetic assay with 512, 1,012, and 2,012 bp dsRNAs. The time derivative of the single-round kinetics (Figure 6.10F) showed a more rapid reduction of the ATP hydrolysis rate for shorter dsRNAs with half-lives of 0.2, 0.6, and 1.3 min for 512, 1,012, and 2,012 bp dsRNAs, respectively (Figure 6.4C). The more rapid decline in the ATP hydrolysis rate with shorter dsRNAs suggests a more rapid dissociation of MDA5 from shorter dsRNAs, which is inconsistent with models in which MDA5 dissociates from the filament interiors or cooperatively with its threshold independent of filament length.

To further validate the length-dependent dissociation mechanisms, we utilized the K335A/D443A mutants to stall the dissociation reaction, and compared the level of WT that remained bound to dsRNAs of different lengths using the single-round kinetic assay. WT was premixed with an equal amount of K335A/ D443A, and complexes were formed with dsRNAs of 112, 512, and 2,012 bp at the equivalent total base-pair concentration. To ensure equivalent binding of WT to dsRNA at the beginning of the reaction, we used 0.2 μM MDA5, which is 10-fold higher than the value of Kd for 112 bp (Figure 6.1B). We measured ATP hydrolysis progression between 4 and 10 min during which no

ATP hydrolysis was observed for the WT alone reaction. In the presence of K335A or D443A, however, ATP hydrolysis continued during this time frame and the rate correlated with dsRNA length; the fold increase in the rate from 112 to 512 bp and 2,012 bp is 4.3 and 15.4 for K335A and 2.5 and 10.7 for D443A, in comparison to a 4.6- and 18-fold increase in length (Figure 6.4D). In principle, if disassembly occurs only from a filament end, the rate ratios would be equal to or greater than the length ratios. The closely matching of rate and length ratios suggests that end disassembly could be the primary mechanism with some level of interior dissociation, or that the filament disassembles cooperatively with nonlinearly increasing stability with filament length.

6.2.8: A partial loss-of-function mutant, I923V, exhibits impaired filament assembly and reduced kinetic stability

A partial loss-of-function SNP variant I923V (Figure 6.5A), which has been associated with a reduced risk for type 1 diabetes, was reported to have reduced signaling activity but retain similar affinity to a dsRNA mimic, poly I:C, as WT MDA5 (Nejentsev et al, 2009; Chistiakov et al, 2010). We asked whether the reduced signaling by I923V is due to impaired filament assembly and stability. We found that I923V binds to 112 bp dsRNA with similar affinity but with reduced cooperativity (K_d of 18 nM vs. 22 nM, Hill coefficient of 1.9 vs. 4.0) (Figure 6.5B).

Consistent with low cooperative binding, EM revealed that I923V is less efficient in filament formation with 112 bp dsRNA (Figure 6.5C). The class

averages revealed that the overall RNA binding modes of I923V and WT are similar, but the I923V complex contains 4–7 dimers rather than the 7–8 dimers seen in the WT complex. The histogram analysis of filament length distribution also indicates that 45% of the WT complexes are approximately 36 nm long, whereas only 15% of the I923V complexes are 36 nm and 60% are 27–31.5 nm (Figure 6.5D). Consistent with this finding, the molecular mass of the I923V:112 bp dsRNA complex estimated from MALS is 209 kDa (~ 1 dimer) less than that of the WT complex.

We next examined the kinetic stability of I923V using the single-round ATP hydrolysis assay. Both WT and I923V were complexed with 2,012 bp dsRNA, and the progression of the ATP hydrolysis reaction in the presence of heparin was monitored (Figure 6.5E). Although both reactions initiated with similar efficiencies, the I923V reaction reached a plateau faster than the WT reaction, resulting in a lower level of ATP hydrolyzed during the single-binding event. Rate analysis suggests that I923V dissociates from dsRNA approximately four times more rapidly than WT (Figure 6.5F). This finding suggests that I923V has a reduced kinetic stability compared with WT, and supports our hypothesis that extended filament formation is important for maintaining kinetic stability of the MDA5:dsRNA complex.

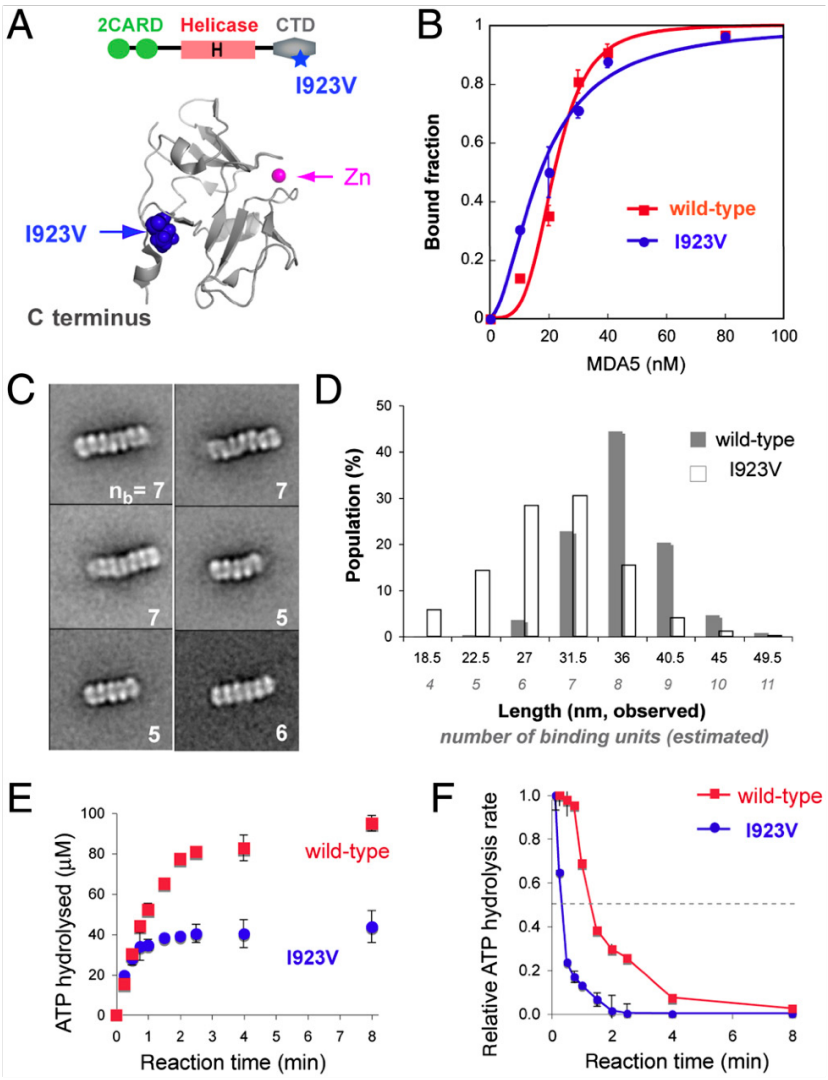

Figure 6.5: SNP I923V forms shorter filaments and exhibits decreased kinetic stability. (A) Cartoon representation of MDA5 CTD [Protein Data Bank (PDB) ID code 2RQB]. Residue I923, distal to the Zn coordination site but proximal to the C terminus, is partially solvent-exposed. (B) EMSA binding curves (mean ± SD, n = 2) of WT and I923V MDA5 with 112 bp dsRNA. (C) Representative class averages of I923V in complex with 112 bp dsRNA. The number of binding units (n_b) present for each complex is displayed at the bottom right. See also Fig. 6.11D. (D) Histogram of the length distribution of complexes formed by WT and I923V with 112 bp dsRNA. Each length represents the median value of each interval. The number of binding units per complex was estimated from the complex length. (E) Single-round ATP hydrolysis kinetics (mean ± SD, n = 3) of WT and I923V (0.2 µM) bound to 2,012 bp dsRNAs (1.2 µg⁄mL) in the presence of heparin (200 µg⁄mL). (F) Time evolution of the ATP hydrolysis rate derived from E using the finite difference method. Rates were normalized against the initial rate during the first 15 s.

6.3: Discussion

The molecular mechanism for recognition of long (>1–2 kb) viral dsRNA by MDA5 remains poorly understood, and our limited understanding has been largely inferred from knowledge of a paralogous receptor, RIG-I, despite its divergent RNA and viral specificity. We report here the previously uncharacterized architecture of a filamentous assembly of MDA5 on dsRNA, and its dynamic ATP-dependent disassembly to provide a mechanism of how MDA5 both recognizes dsRNA and auto-regulates its antiviral signaling activity in a length-dependent manner.

We demonstrate here that binding affinity alone is insufficient to confer dsRNA specificity to MDA5 with high-affinity binding observed for multiple types of nucleic acids. Despite low stringency nucleic acid discrimination on the basis of affinity, MDA5 is able to distinguish between dsRNA and other types of nucleic acids through cooperative filamentous assembly along the length of dsRNA. Analogous filament structures have been observed with RecA, which forms filamentous oligomers along DNA and promotes strand exchange during homologous recombination (Cox, 2007). MDA5 is structurally related to RecA in that its central helicase domain consists of two RecA-like subdomains. However, unlike the RecA filament whose formation requires ATP binding (Chen et al, 2008), MDA5 filament formation does not require ATP or its analogues. We speculate that the C-terminal domain (CTD) of MDA5, which is absent in RecA, plays an important role in the cooperative assembly process of the filament. A partial loss-of-function mutation, I923V, within the CTD lowers both the

cooperativity and the extent of filament assembly without affecting the dsRNA affinity of MDA5 (Figure 6.5). In addition, replacement of the MDA5 CTD by the RIG-I CTD abolishes the cooperativity while retaining high-affinity interaction with dsRNA. Filament formation was recently proposed for RIG-I based on atomic force microscopy (Binder et al, 2011); however, we did not observe either filament formation of RIG-I by EM or cooperative binding by EMSA.

We demonstrated that the MDA5:dsRNA filament is a highly dynamic structure, the stability of which is tightly regulated by ATP hydrolysis. Using quantitative single- and multiround kinetic analysis and catalytic mutant interference studies, we provide evidence supporting that ATP hydrolysis occurs throughout the filament with little coordination between neighboring MDA5 molecules (Figure 6.3). ATP hydrolysis-driven MDA5 dissociation, however, occurs in a coordinated manner as evidenced by a markedly increased stability of a filament upon incorporation of catalytically inactive mutants (Figure 6.4B). This coordinated dissociation occurs at a rate inversely proportional to the filament length (Figure 6.4C), suggesting that not every ATP hydrolysis event leads to MDA5 dissociation. We propose two mechanisms to explain this length-dependent dissociation. First, MDA5 may dissociate from dsRNA as monomers or dimer pairs from filament extremities. This end dissociation could be because MDA5 molecules in the filament ends are less stable than those in the interiors due to the lack of stabilizing contacts with neighboring molecules. Based on this end-disassembly model, the positive effect of the catalytic mutants on the filament stability could be explained by the mutants capping the filament ends and

preventing progression of filament disassembly. Second, MDA5 may cooperatively dissociate from dsRNA as an entire filament or smaller fragments whose stability increases nonlinearly with filament length. In this model, the catalytic mutants may increase filament stability by increasing the threshold required for filament disassembly by ATP hydrolysis.

The dual and seemingly opposing roles of ATP hydrolysis as both a consequence of filament formation (or dsRNA binding) and as a trigger for filament disassembly is reminiscent of the roles of nucleotide hydrolysis in other filamentous proteins, such as RecA, actin, and tubulin. These proteins utilize either repetitive or single-round of nucleotide hydrolysis as a counterbalance between filament growth and turnover (Cox, 2007; Kueh & Michison, 2009). Competition between the filament assembly and disassembly processes was also evident for MDA5 in EMSA during the ATP hydrolysis steady state, which revealed reduced apparent affinity in comparison without ATP. Although comprehensive understanding of the MDA5 filament dynamics would require more detailed knowledge on both assembly and disassembly processes, the observed length-dependent dissociation provides a basis for how MDA5 could self-regulate its interaction with dsRNA and thus antiviral signaling potential according to the underlying dsRNA length. This notion is in agreement with the known requirement of long (>1–2 kb) dsRNA to efficiently elicit interferon signaling activity of MDA5 in the cell (Kato et al, 2008). Thus, our observations suggest a previously unrecognized role of ATP hydrolysis in autoregulation of MDA5 filament growth and turnover.

The highly cooperative MDA5 filament formation raises intriguing possibilities for other antiviral functions of MDA5. As many receptors in the immune system cluster for effective signal transmission (Davis et al, 2011; Sigalov, 2010), clustering of MDA5 on dsRNA may facilitate recruitment of MAVS or subsequent oligomerization of MAVS, which itself was recently shown to form prion-like fibrillar structures during antiviral signaling (Hou et al, 2011). It has been proposed that recruitment of MAVS by MDA5 depends on a conformational change of MDA5 induced by ATP hydrolysis (Wilkins & Gale, 2010; Hou et al, 2011). It would be interesting to test whether ATP hydrolysis by individual MDA5 molecules within a filament recruits MAVS and promote their oligomerization through induced proximity. In addition, MDA5 filament formation may also perform a unique antiviral effector function by limiting access of viral proteins to RNAs, thereby interfering with viral replication or packaging. Interestingly, a recent study identified MDA5, along with RIG-I, as one of the most broadly acting antiviral effectors (Schoggins et al, 2011). Our present studies provide a basis for future investigations on more detailed mechanisms and functions of MDA5 in antiviral defense.

6.4: Materials and methods

Electron microscopy. MDA5 was incubated with a 20：1 molar ratio of MDA5 to 112 bp dsRNA (ssRNA or dsDNA) in 20 mM Hepes, pH 7.0, 150 mM NaCl, 5 mM ADP•AlF_4, and cross-linked using 0.4 mM BMH at room temperature (RT)

for 30 min. Negatively stained specimens were prepared as described in Ohi et al (2004). Images were collected with a Tecnai T12 electron microscope (FEI) and processed using the SPIDER software package.

Multiangle light scattering. The absolute molecular mass of the complexes formed by WT MDA5 or I923V with dsRNA was determined using a Superose 6 10/300 column (GE) attached to a miniDAWN TREOS detector (Wyatt Technology) in 20 mM Hepes, pH 7.0, 75 mM NaCl. MDA5：dsRNA complexes were prepared as for EM.

Electrophoretic mobility shift assay. EMSA was performed using 5′-^{32}P-, 3′-fluorescein- or Cy3-labeled RNAs. Unless mentioned otherwise, 0.2 nM of ^{32}P-labeled RNA was mixed with the indicated amount of MDA5 in buffer A (20 mM Hepes, pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, and 2 mM DTT) with 2 mM ADPCP. The mixture was incubated at RT for 10 min and were analyzed on BisTris native gels (Life). For time-dependent EMSA, reactions were incubated at 37°C for indicated time period.

ATP hydrolysis assay. The ATP hydrolysis rate (initial velocity) was determined using Green Reagent (Enzo) (Martin et al, 1985). Unless mentioned otherwise, MDA5 was preincubated with RNA in buffer A at 37°C, and the reaction was initiated by addition of 2 mM ATP. For single-round kinetic assays, a mixture of 200 μg⁄mL heparin and 2 mM ATP was added to the reaction. For initial velocity

measurement, aliquots were withdrawn at five time points between 15 s and 30 min, and were quenched with 50 mM EDTA on ice. The Green Reagent was added to the quenched reaction at a ratio of 9：1, and OD_{650} was measured using a Synergy2 plate reader (BioTek).

6.5: References

Ablasser, A., Bauernfeind, R., Hartmann, G., Latz, E., Fitzgerald, K. & Hornung, V. (2009) RIG-I-dependent sensing of poly (dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat Immunol* 10:1065–1072.

Andrejev, J., Childs, K., Young, D., Carlos, T., Stock, N., Goodbourn, S., and Randall, R. (2004) The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN−1 promoter. *Proc Natl Acad Sci USA* 101:17264–17269.

Bamming, D., and Horvath, C. (2009) Regulation of signal transduction by enzymatically inactive antiviral RNA helicase proteins MDA5, RIG-I, and LGP2. *J Biol Chem* 284:9700–9712.

Baum, A., Sachidanandam, R., and Garcia-Sastre, A. (2010) Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc Natl Acad Sci USA* 107:16303–16308

Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A., and Bartenschlager, R. (2011) Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid inducible gene-I. *J Biol Chem* 286:27278–27287.

Chen, Z., Yang, H., and Pavletich, N. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453:489–494.

Chistiakov, D., Voronova, N., Savost'Anov, K., and Turakulov, R. (2010) Loss-of-function mutations E627X and I923V of IFIH1 are associated with lower poly(I:C)-induced interferon−1 production in peripheral blood mononuclear cells of type 1 diabetes patients. *Hum Immunol* 71:1128–1134.

Cordin, O., Banroques, J., Tanner, N., and Linder P. (2006) The DEAD-box protein family of RNA helicases. *Gene* 367:17–37.

Cox, M. (2007) Motoring along with the bacterial RecA protein. *Nat Rev Mol Cell Biol* 8:127–138.

Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K., Krug, A., and Hopfner, K. (2008) The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Mol Cell* 29:169-179.

Davis, B., Wen, H., and Ting J. (2011) The inflammasome NLRs in immunity, inflammation, and associated diseases. *Annu Rev Immunol* 29:707–735.

Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q., and Chen Z. (2011) MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* 146:1–14.

Kang, D., Gopalkrishnan, R., Wu, Q., Jankowsky, E., Pyle, A., and Fisher P. (2002) mda-5: Aninterferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci USA* 99:637–642.

Kato H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C., Sousa, C., Matsuura, Y., Fujita, T. & Akira, S. (2006) Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* 441:101-105.

Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T., Fujita, T. & Akira, S. (2008) Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med* 205:1601–1610.

Kueh, H., and Michison, T. (2009) Structural plasticity in actin and tubulin polymer dynamics. *Science* 325:960–963.

Linder, P., and Jankowsky, E. (2011) From unwinding to clamping – the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12:505–516.

Loo, Y., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira Sh., Gill, M., Garcia-Sastre, A., Katze M., and Gale, M. (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol* 82:335-345.

Martin, B., Pallen, C., Wang, J., and Graves, D. (1985) Use of fluorinated tyrosine phosphates to probe the substrate specificity of the low molecular weight phosphatase activity of calcineurin. *J Biol Chem* 260:14932–14937.

McCartney, S., Thackray, L., Gitlin, L., Gilfillan, S., Virgin, H. & Colonna, M. (2008) MDA-5 recognition of a murine norovirus. *PLoS Pathogens* 4**:**e1000108.

Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324:387–389.

Ohi, M., Li, Y., Cheng, Y., and Walz T (2004) Negative staining and image classification—powerful tools in modern electron microscopy. *Biol Proced Online* 6:23–34.

Rehwinkel, J., and Reis e Sousa, C. (2010) RIGorous detection: Exposing virus through RNA sensing. *Science* 327:284–286.

Saito, T., Hirai, R., Loo, Y., Owen, D., Johnson, C., Sinha, S., Akira, S., Fujita, T., and Gale, M. (2007) Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. *Proc Natl Acad Sci USA* 104:582–587.

Sigalov AB (2010) The SCHOOL of nature. *Self Nonself* 1:4–39.

Schlee, M., Roth, A., Hornung, V., Hagmann, C., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G., Juranek, S., Kato, H., Kawai, T., Poeck, H., Fitzgerald, K., Takeuchi, O., Akira, S., Tuschi, T., Latz, E., Ludwig, J., and Hartmann, G. (2009) Recognition of 5′ triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. *Immunity* 31:25–34.

Schoggins, J., Wilson, S., Panis, M., Murphy, M., Jones, C., Bieniasz, P., and Rice, C. (2011) A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472:481–485.

Wilkins, C., and Gale Jr, M. (2010) Recognition of viruses by cytoplasmic sensors. *Curr Opin Immunol* 22:41–47.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004) The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* 5:730–737.

Zeng, W., Sun, L., Jiang, X., Chen, X., Hou, F., Adhikari, A., Xu, M., and Chen Z. (2010) Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* 141:315–330.

Chapter 7: Other structural studies of MDA5

7.1: Introduction

Several aspects of MDA5 activation remain unclear, including the nature of the associated conformational changes and the precise roles of ATP and dsRNA binding. Since MDA5 and RIG-I have similar domain architectures and high sequence similarity, we hypothesized that MDA5 may be regulated in the same manner as RIG-I. In the absence of dsRNA, RIG-I maintains an autorepressed conformation, with the CARDs unavailable for signaling. In this state, the helicase domain adopts an open conformation, in which the three individual subdomains, Hel1, Hel2, and Hel2i, are arranged in a semicircle with a central crevice (Kowalinski et al, 2011). The tandem CARDs, linked to the helicase by a flexible 55-residue linker, interact with the Hel2i insertion domain through multiple salt bridges and hydrophobic interactions (Kowalinski et al, 2011). Cooperative dsRNA and ATP binding induces a large conformational change in the RIG-I helicase domain, with all three helicase subdomains rearranging into a closed conformation to create the mature binding site. This rearrangement disrupts the interaction between the tandem CARDs and the Hel2i subdomain, freeing the CARDs for signaling (Kowalinski et al, 2011).

To gain insight into the conformational changes associated with MDA5 activation, it is essential to obtain structural information of MDA5 alone and in complex with ATP. Although the crystal structure of the MDA5 CTD has been determined and the dsRNA-binding site has been mapped, neither the overall

MDA5 structure nor the structure of any other domain is known. To obtain structural information that would allow us to develop a model for MDA5 activation, we used single-particle EM to visualize the structures of full-length MDA5, full-length MDA5 in complex with ADP.ALF4, and deltaCARD MDA5 in complex with ADPCP.

2D averages and MALS data revealed that MDA5 assembled onto 112bp dsRNA forms a filamentous complex that consists of eight stacked discs, with each disc being composed of an MDA5 dimer (Peisley et al, 2011). Since it is not clear how MDA5, which is monomeric in solution, assembles onto dsRNA as dimers, we also calculated a 3D reconstruction of the MDA-dsRNA complex in the hope that the density map of the filament would provide insight into intra- and inter-disc interactions that govern filament formation.

In addition to working towards a structural model of MDA5 activation, we also aimed at gaining an understanding of how MDA5 senses the length of dsRNA. Kato et al. (2008) reported that treatment of long dsRNA (greater than 1 kb) with RNase III converts it from an MDA5 ligand to a RIG-I ligand. This result suggests that MDA5 has a mechanism to measure the length of the dsRNA before accepting or rejecting it as an appropriate ligand. Furthermore, electrophoretic mobility shift assays (EMSA) revealed that MDA5 binds to 100-bp dsRNA with high affinity and is indiscriminate of termini structure (Peisley, 2012). dsRNA molecules with 5'-monophosphate, 5'-triphosphate, 5'-OH, and 5'- or 3'- overhangs of 1, 2, or 12 nucleotides all bound to MDA5 with similar affinities. However, when 12-bp dsRNA was used, MDA5 still bound with high

affinity to dsRNA molecules with 5'-monophosphate, 5'-triphosphate, and 5'-OH, but did no longer bind dsRNA molecules with 5'- or 3'- overhangs (Peisley, 2012). These results suggest that the mode of MDA5 binding varies based on dsRNA length. We sought to understand the basis for this phenomenon by using single-particle EM to follow the disassembly of MDA5-dsRNA filaments.

7.2: Results

7.2.1: Structural characterization of monomeric MDA5

We first used negative-stain EM to visualize recombinantly expressed human MDA5 in a low-salt buffer (75 mM NaCl). Multiangle light scattering (MALS) revealed that the majority of MDA5, a 1025 amino acid, 110 kDa protein, was monomeric in solution, but it also detected a small population of dimeric MDA5. In contrast, EM images of negatively stained MDA5 revealed particles of variable size and shape (Figure 7.1), and this structural heterogeneity was also seen in class averages (Figure 7.2). The heterogeneous size and shape of the particles suggested that MDA5 is prone to aggregation in low-salt buffer, and that different buffer conditions had to be explored to stabilize the protein.

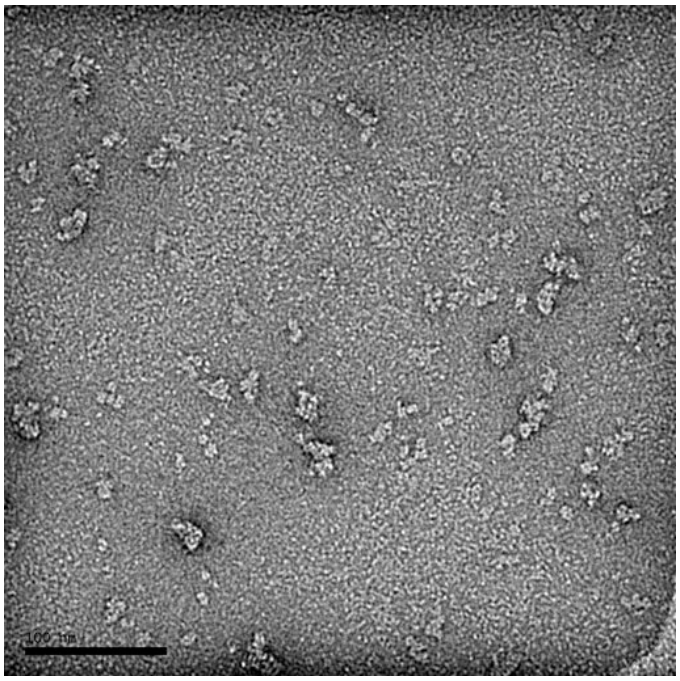

Figure 7.1: Representative raw image of negatively stained MDA5 in low-salt buffer (75 mM NaCl). Image was taken at a magnification of 52,000x, and the scale bar is 100 nm.

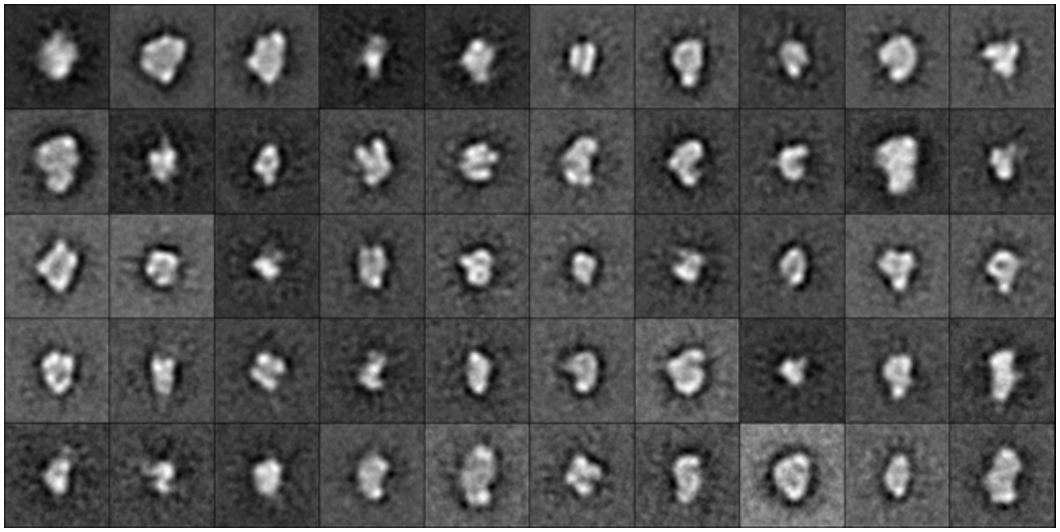

Figure 7.2: Averages obtained by classifying 5004 particle images of negatively stained MDA5 in low-salt buffer into 50 classes. The side length of the individual panels is 29 nm.

MDA5 in high-salt buffer: We tested whether a higher salt concentration in the sample buffer could reduce protein aggregation. Images of an MDA5 sample in 300 mM NaCl showed indeed less aggregation in high-salt buffer (Figure 7.3), and class averages revealed particles with clearly resolved protein domains (Figure 7.4). All further studies on MDA5 were therefore performed in high-salt buffer. Since MALS indicated that MDA5 exists predominantly in monomeric form, the class averages likely represent MDA5 monomers. The averages show that the protein adopts an elongated and flexible conformation. Although some classes show long molecules, presumably representing the full-length protein (Figure 7.5, a), most classes show smaller molecules and resolve two to four domains (Figure 7.5, b-d). In these classes, the remaining domains were likely averaged out due to the high flexibility of MDA5. Since the tandem CARD domain is connected to the central helicase domain through a 105 amino acid linker, this is the most likely domain that was averaged out in many classes.

The high degree of flexibility seen in soluble MDA5 suggests that MDA5 signaling may not be regulated in a manner similar to RIG-I signaling. Unbound RIG-I exists in an "autorepressed" conformation, in which the tandem CARDs are bound to the Heli subdomain of the helicase domain. If MDA5 were regulated in the same way, the class averages would be expected to show proteins in a more defined conformation. The structural heterogeneity seen instead in averages of unbound MDA5 thus suggest that both the tandem CARDs and the CTD are not constrained.

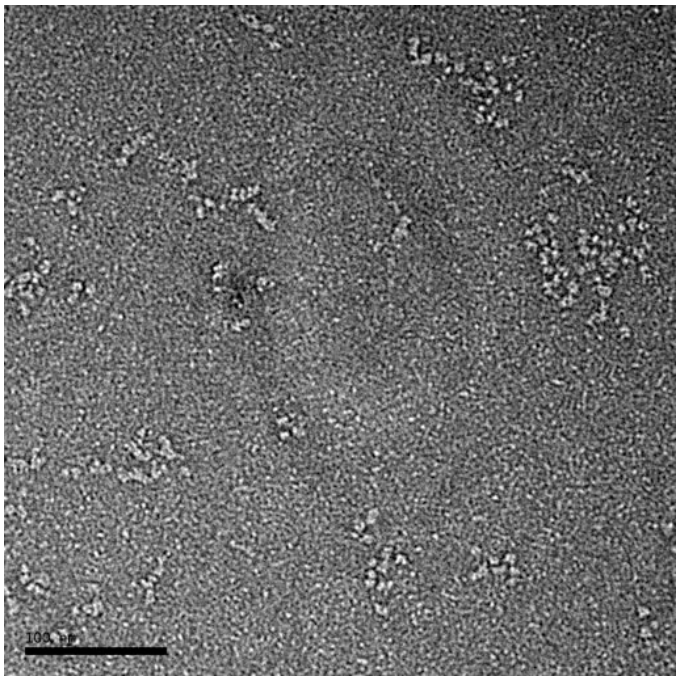

Figure 7.3: Representative raw image of MDA5 in high-salt buffer (300 mM NaCl). Image was taken at a magnification of 52,000x, and the scale bar is 100 nm.

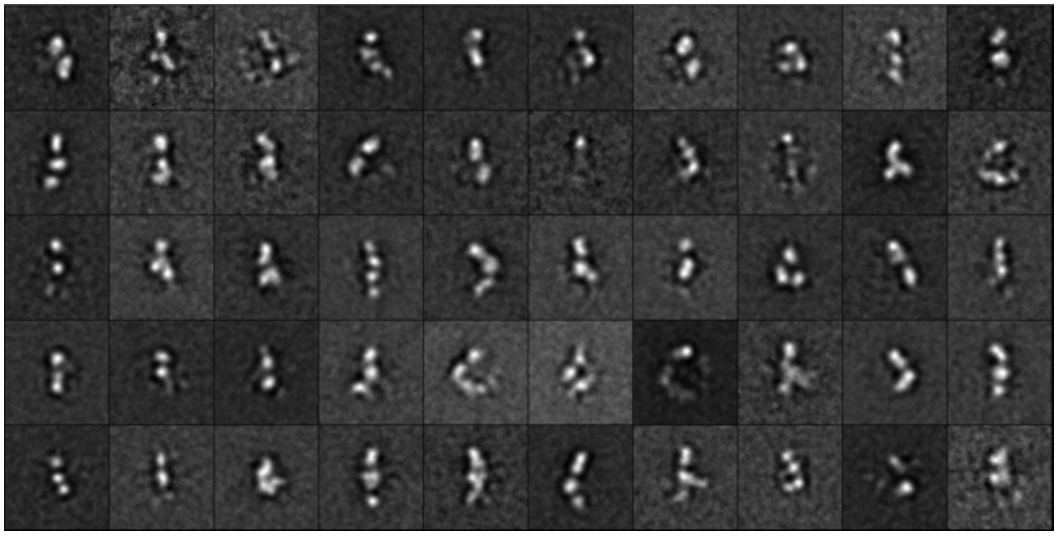

Figure 7.4: Averages obtained by classifying 4915 particle images of negatively stained MDA5 in high-salt buffer into 50 classes. The side length of the individual panels is 29 nm.

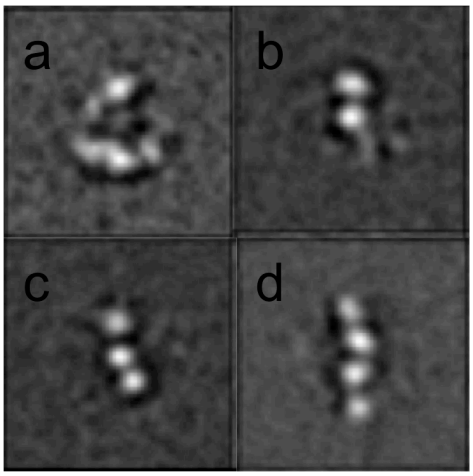

Figure 7.5: Selected averages of MDA5 in high-salt buffer. a) The largest particles likely represent full-length MDA5. b) MDA5 protein showing 2 domains. c) MDA5 protein showing 3 domains. d) MDA5 protein showing 4 domains. The side length of the individual panels is 29 nm.

MDA5 with ATP analogs: We next tested whether non-hydrolyzable ATP analogs could reduce the heterogeneity of the MDA5 population by locking the helicase domain in a defined conformation. We tested both the substrate analogue ADPCP and the transition state analogue ADP.ALF_4. Since ADP.ALF_4 mimics the tetrahedral geometry during the transition state of ATP hydrolysis, this ATP analog in particular may stabilize MDA5 in a dsRNA-bound conformation. A comparison of the structures of MDA5 in the free and ATP analog-bound forms thus had the potential to reveal the conformational changes that underlie activation of MDA5 upon ligand binding.

Representative raw images of negatively stained ADP.ALF_4- and ADPCP-bound MDA5 are shown in Figure 7.6. Although it was difficult to find well-stained areas for both samples (phosphates interfere with negative staining), the images suggested that the conformational variability of MDA5 with bound ATP analogs was similar to that of free MDA5. Averages of MDA5 with bound

ADP.ALF$_4$ indeed showed that the protein still adopts a flexible and elongated conformation (Figure 7.7). Thus, while ADP.AlF$_4$ presumably locked the helicase domain into a defined conformation, this did not translate to a reduction of the conformational variability displayed by the full-length protein.

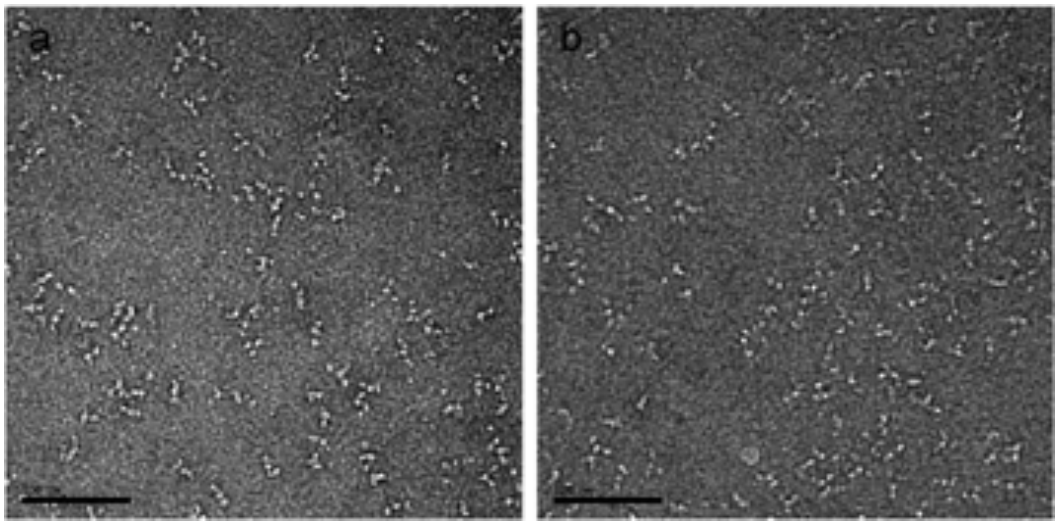

Figure 7.6: MDA5 in the presence of ATP analogs. a) Raw image of MDA5 in high-salt buffer with 5 mM ADP.AlF$_4$. b) Raw image of MDA5 in high-salt buffer with 5 mM ADPCP. Images were taken at a magnification of 52,000x, and the scale bars are 100 nm.

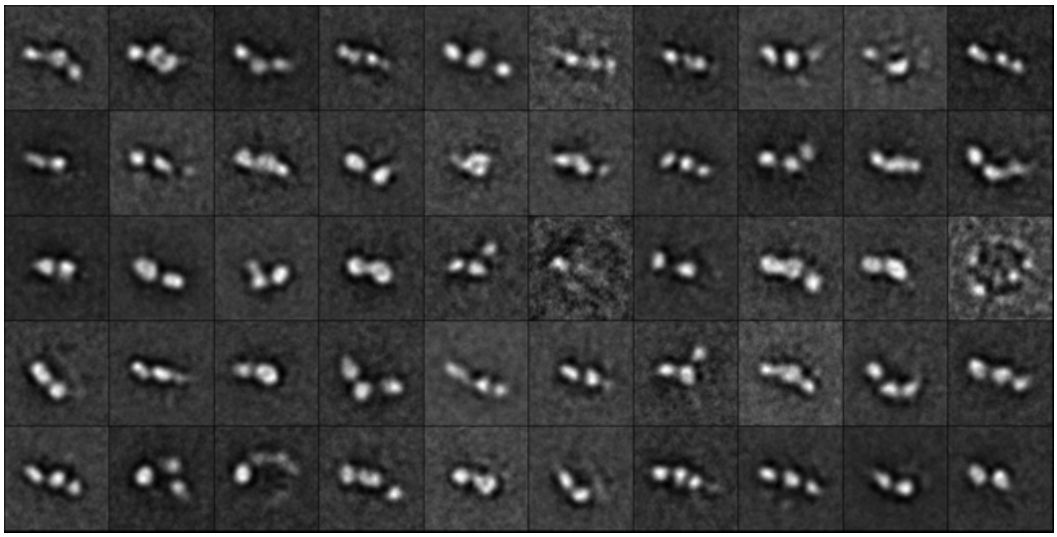

Figure 7.7: Averages obtained by classifying 5010 particle images of negatively stained MDA5 in the presence of ADP.AlF$_4$ into 50 classes. The side length of the individual panels is 29 nm.

DeltaCARD MDA5: To identify the position of the tandem CARDs in MDA5, we imaged an MDA5 construct lacking the CARD domain in the presence of ADP.AlF$_4$ or ADPCP (Figure 7.8). The images suggest that the conformation variability of this deltaCARD construct is similar in the presence of the two ATP analogs and to that of the full-length protein. This impression was confirmed by averages of deltaCARD MDA5 in the presence of ADPCP (Figure 7.9). The averages show particles with a similar structural heterogeneity seen for the full-length protein. This result suggests that the CARDs are not visible in the averages of full-length MDA5, most likely due to a combination of their small size and flexible linkage to the helicase domain. The conformational variability of MDA5 seen in averages of full-length and deltaCARD MDA5 thus likely arises from the flexible linker between the helicase domain and the CTD of MDA5.

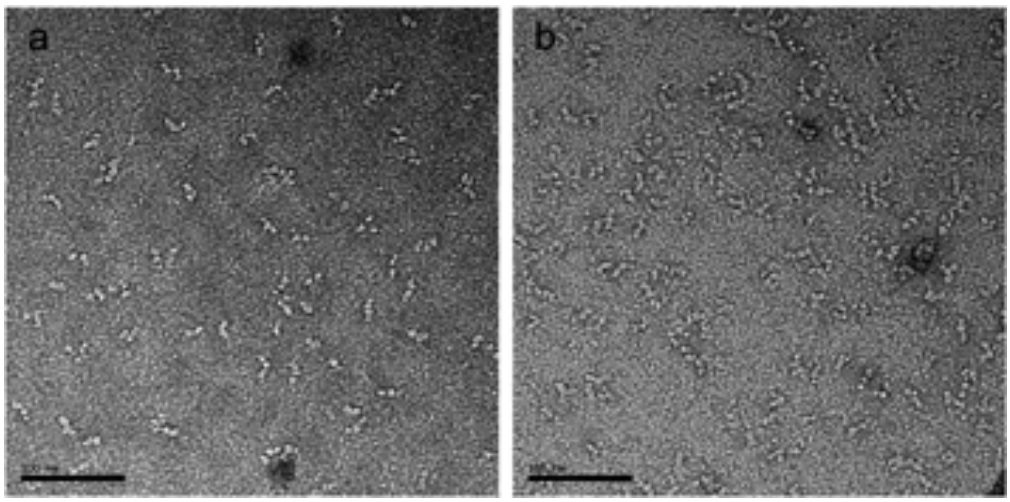

Figure 7.8: DeltaCARD MDA5 representative images. a) Raw image of deltaCARD MDA5 in high-salt buffer with 5 mM ADP.AlF_4. b) Raw image of deltaCARD MDA5 in high-salt buffer with 5 mM ADPCP. Images were taken at a magnification of 52,000x, and the scale bars are 100 nm.

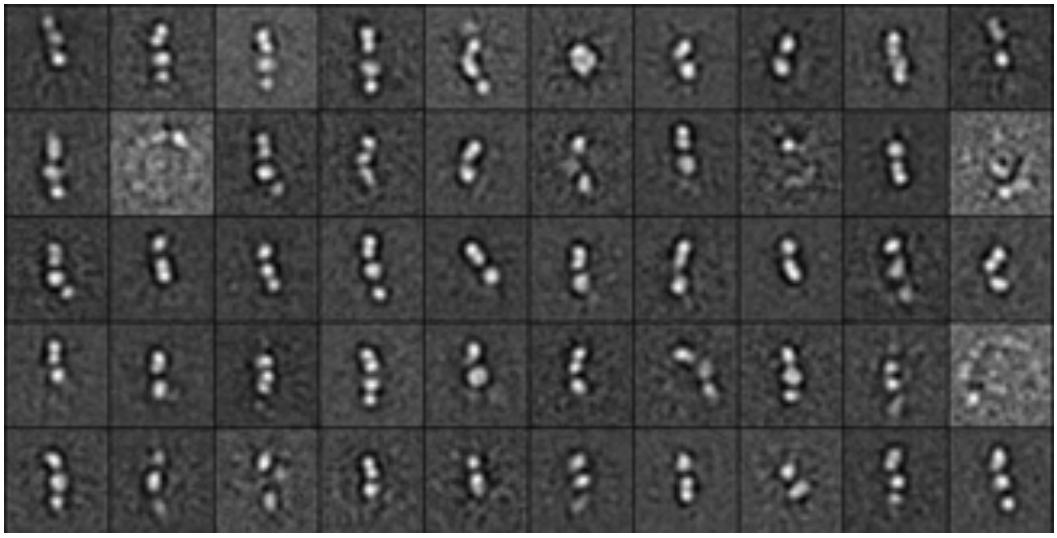

Figure 7.9: Averages obtained by classifying 4479 particle images of negatively stained deltaCARD MDA5 in the presence of ADPCP into 50 classes. The side length of the individual panels is 29 nm.

7.2.2: Three-dimensional reconstruction of the MDA5–dsRNA complex

We demonstrated that MDA5 forms filamentous oligomers along the length of dsRNA (Chapter 6). 2D averages of MDA5 in complex with 112-bp dsRNA show a filament comprised of seven to eight stacked discs that does not appear to have helical symmetry. MALS revealed that each of the stacks in the nucleoprotein complex is an MDA5 dimer (Peisley et al, 2011). In an effort to understand the intra- and inter-dimer protein-protein interactions that stabilize the complex, we calculated a 3D reconstruction of the complex formed by MDA5 on 112-bp dsRNA. 10,470 particle pairs were selected from 88 50°/0° image pairs of MDA5–dsRNA complexes prepared by cryo-negative staining (Cheng et al, 2006). The particles selected from the images of the untilted specimen were classified into 20 classes. While all class averages revealed 7 to 8 stacked discs, the complexes differed in the extent of bending and the presence of kinks (Figure 7.10). The class with the most particles and best resolved fine structure were chosen for 3D reconstruction by the random conical tilt approach (Radermacher, 1988). The best density map in terms of resolved structural features filtered to 30 Å resolution is shown in Figure 7.11. The map resolves 8 stacked discs but provided no additional insights into the architecture of the complex compared with the 2D averages.

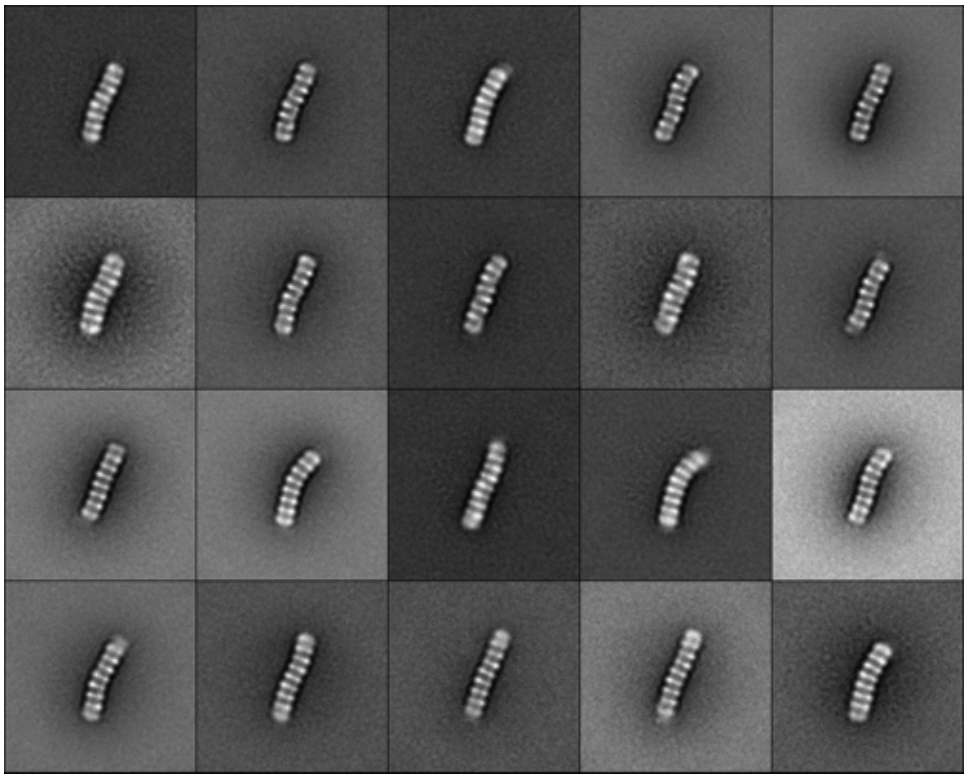

Figure 7.10: Averages obtained by classifying 10,470 particles of cryo-negatively stained MDA5-dsRNA complexes selected from untilted images. The particles were grouped into 20 classes. The side length of the individual panels is 54 nm.

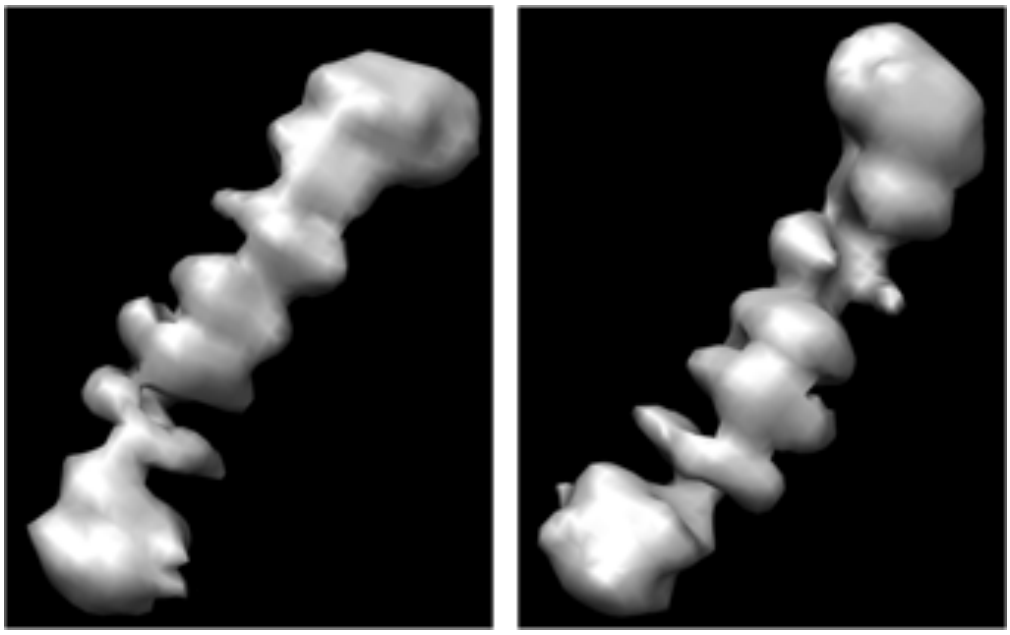

Figure 7.11: MDA5-dsRNA 3D density map. Different views of the 3D density map of MDA5 bound to 112-bp dsRNA at a resolution of 30 Å.

7.2.3: Dissociation of MDA5 from 112-bp dsRNA

We have shown previously that ATP hydrolysis promotes the dissociation of MDA5 from dsRNA. EMSA experiments revealed that 50% of complexes formed by MDA5 with 512-bp dsRNA dissociate fully within 15 s of incubation with ATP (Peisley et al, 2011). This finding is consistent with studies of several other helicases that require ATP hydrolysis for dissociation (Cordin et al, 2006; Linder et al, 2011). We have also shown that the rate of ATP hydrolysis is dependent on dsRNA length, where shorter dsRNAs experience a more rapid decline (Peisley et al, 2012). This result is consistent with two possible models for dissociation: 1) disassembly of MDA5 occurs mainly from the ends of the dsRNA (as opposed to indiscriminate dissociation from both the interior and the ends of the MDA5 filament) or 2) MDA5 disassembly occurs cooperatively with a higher threshold for longer filaments.

We sought to further increase our understanding of MDA5 dissociation from dsRNA by using EM to visualize the disassembly as a function of time. MDA5 was preincubated with 1-kbp or 2-kbp dsRNA, and dissociation was initiated by adding a mixture of ATP and a 166-fold excess of heparin relative to RNA. The ability of heparin at this ratio to trap MDA5 released from the dsRNA was confirmed previously (Peisley et al, 2011). Aliquots were taken at various time points (0 to 100 s) and imaged by negative-stain EM (Figures 7.12 and 7.13). The dissociation experiments corroborated our previous biochemical results that indicated that the dissociation rate is inversely proportional to the filament length. Complete dissociation of MDA5 from 1-kbp dsRNA occurred within 60 s, whereas dissociation from 2-kbp dsRNA required 80 s. Images taken at intermediate time points also revealed that complex disassembly occurs not only at filament ends but also at multiple points along the nucleoprotein complex. We currently do not understand the requirements for the disassembly of MDA5 from the interior of the filament, but possible models are outlined in the discussion section.

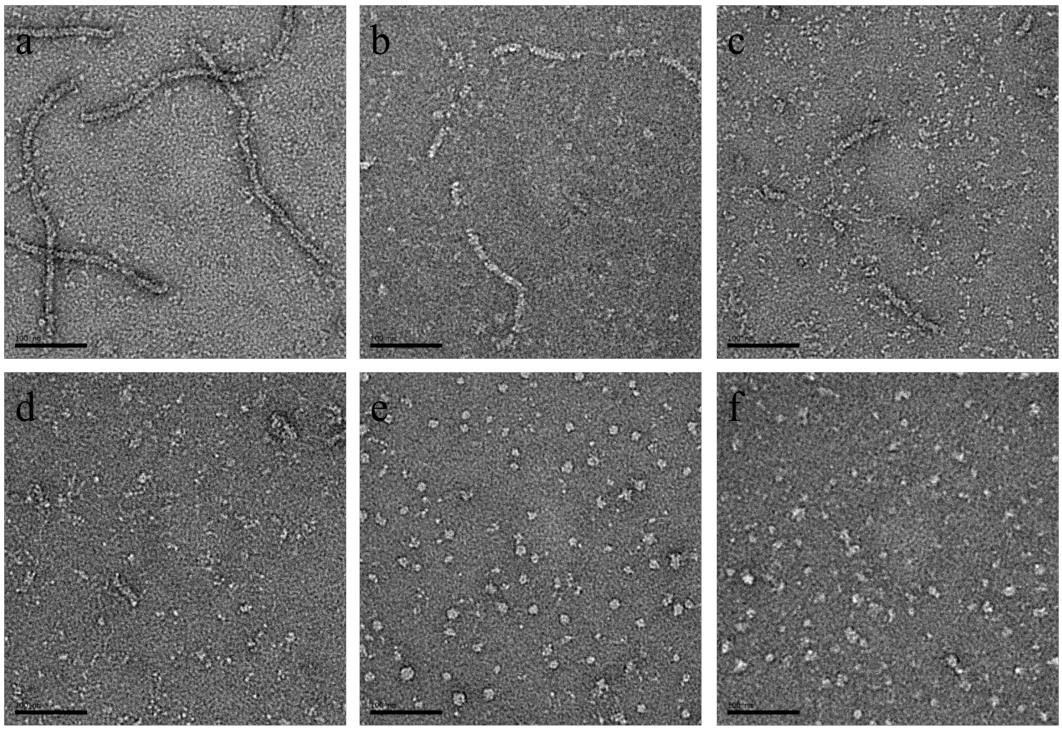

Figure 7.12: Dissociation of MDA5 from 1-kbp dsRNA. MDA5 was pre-incubated with 1-kbp dsRNA, and dissociation was initiated by adding a mixture of ATP and a 166-fold excess of heparin relative to the dsRNA. Aliquots were taken at defined time points and imaged by negative-stain EM. a) 0 s, b) 20 s, c) 40 s, d) 60 s, e) 80 s, and f) 100s. Images were taken at a magnification of 52,000x, and scale bars are 100 nm.

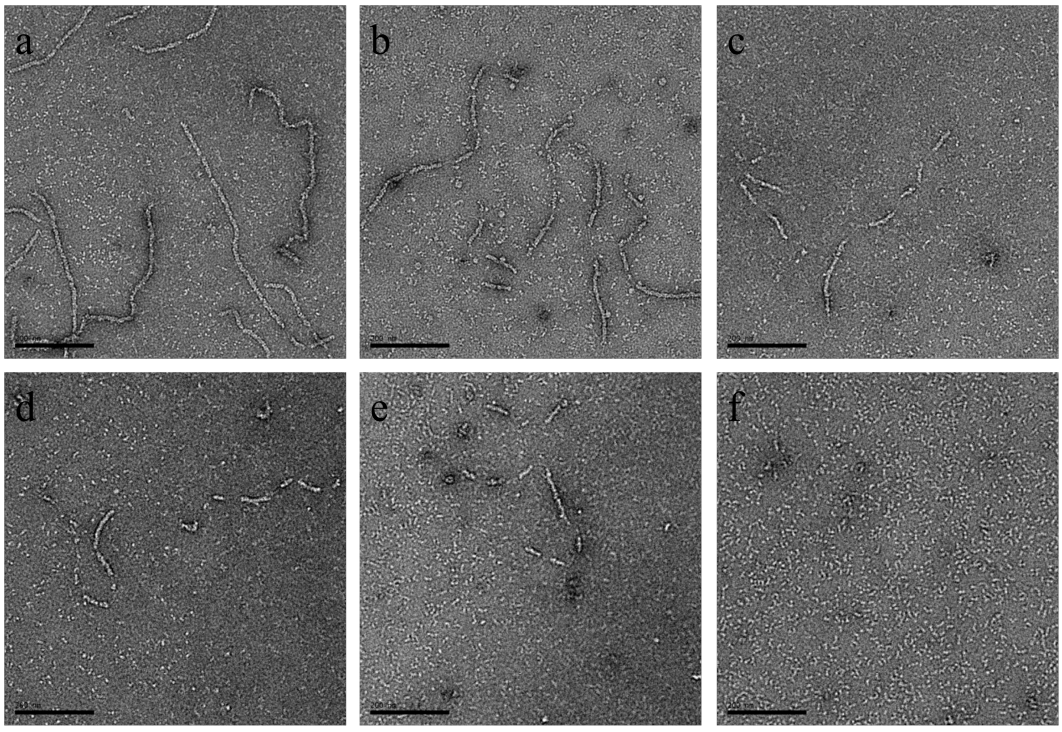

Figure 7.13: Dissociation of MDA5 from 2-kbp dsRNA. MDA5 was pre-incubated with 2-kbp dsRNA, and dissociation was initiated by adding a mixture of ATP and a 166-fold excess of heparin relative to the dsRNA. Aliquots were taken at defined time points and imaged by negative-stain EM. a) 0 s, b) 20 s, c) 40 s, d) 60 s, e) 80 s, and f) 100s. Images were taken at a magnification of 28,500x, and scale bars are 200 nm.

7.2.4: Averages of MDA5 complexes with ssRNA and dsDNA

Recent studies revealed that the ATPase activity of MDA5 depends on the presence of dsRNA and cannot be induced by single-stranded (ss) RNA or dsDNA (Kang et al, 2002; Peisley et al, 2011). To see whether the specificity of MDA5 for the different nucleic acid species is reflected in the structure of the formed complexes, we used negative-stain EM to visualize the structures of the complexes formed by MDA5 with ssRNA and dsDNA (Figure 7.14). The images suggest that MDA5 can form complexes with both ssRNA and dsDNA, but the complexes are smaller and structurally distinct from the MDA5-dsRNA complex.

Averages of the MDA5-ssRNA (Figure 7.15) and MDA5-dsDNA complexes (Figure 7.16) showed that the complexes varied greatly in size and had no indication of the ordered arrangement seen in MDA5-dsRNA complexes. We hypothesize that the imaged particles represent abortive complexes or complexes that are falling apart. The observed structural differences may explain the inability of MDA5 bound to ssRNA and dsDNA to hydrolyze ATP.

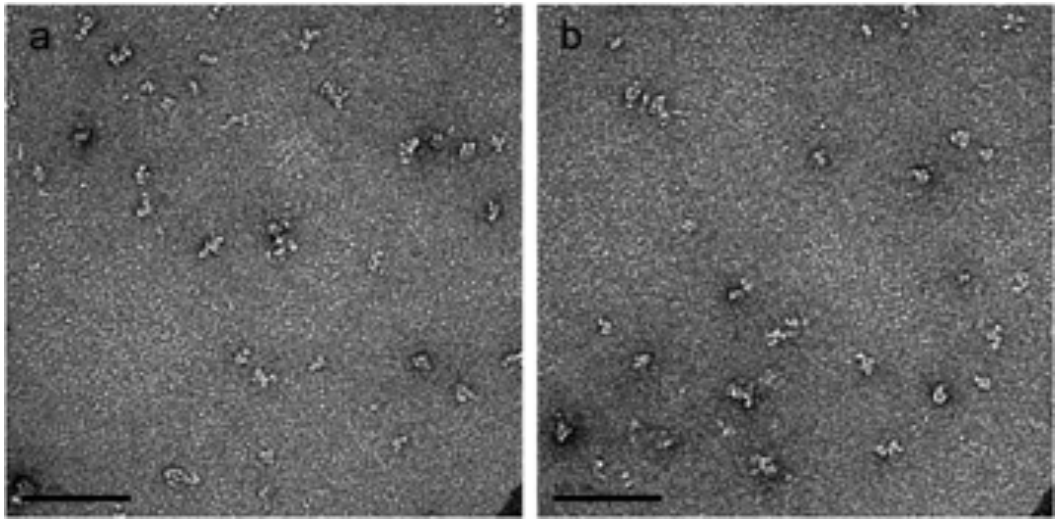

Figure 7.14: MDA5-ssRNA and MDA5-dsDNA complexes. a) Raw image of MDA5 bound to ssRNA in high-salt buffer with 5 mM ADP.AlF_4. (b) Raw image of MDA5 bound to dsDNA in high-salt buffer with 5 mM ADP.AlF_4. Images were taken at a magnification of 52,000x, and the scale bars are 100 nm.

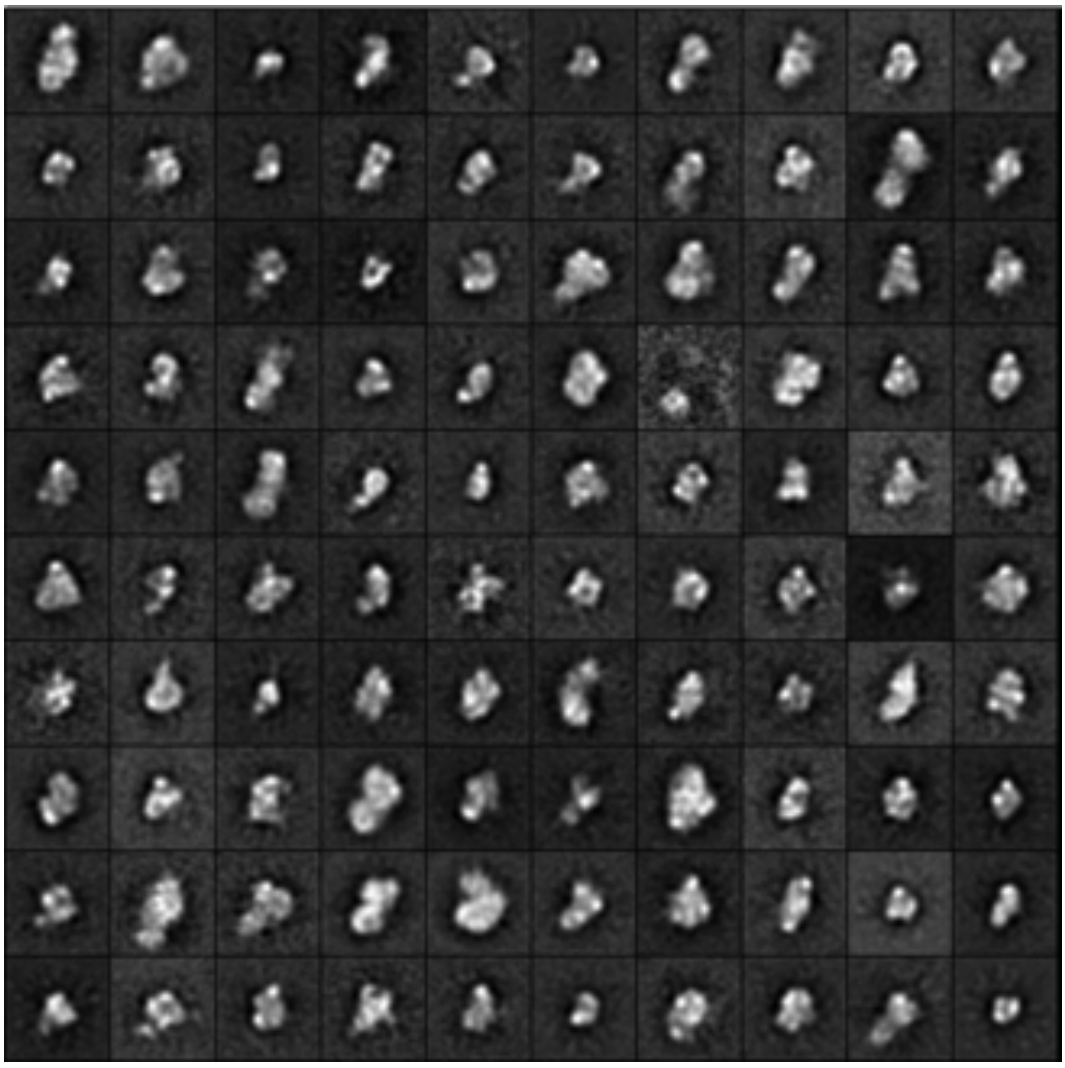

Figure 7.15: MDA5-ssRNA averages. Averages were obtained by classifying 10,902 particles of negatively stained MDA5-ssRNA complexes into 50 classes. Particles were in high-salt buffer containing 5 mM ADP-AlF_4. The side length of the individual panels is 54 nm.

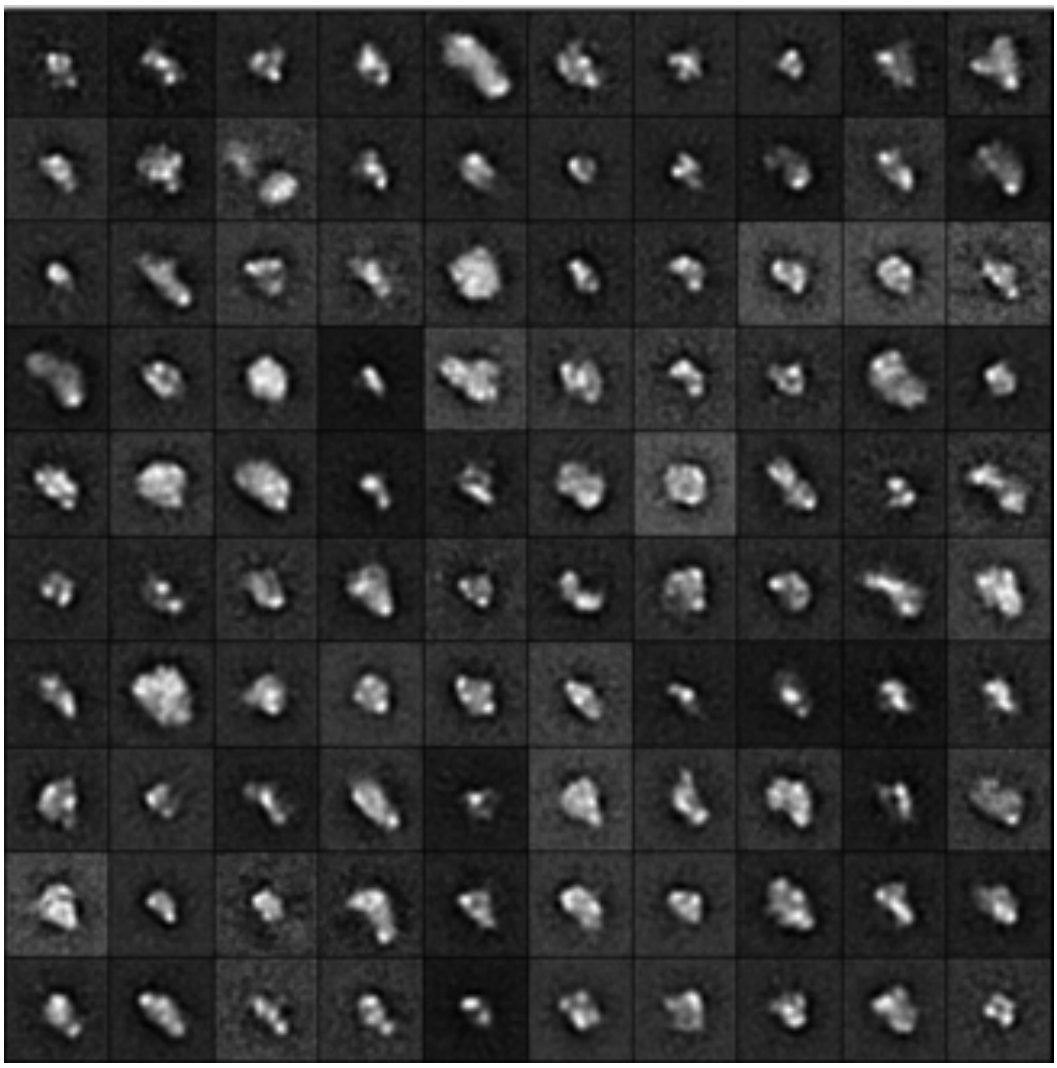

Figure 7.16: MDA5-dsDNA averages. Averages were obtained by classifying 10,187 particle of negatively stained MDA5-dsDNA complexes into 50 classes. Particles were in high-salt buffer containing 5 mM ADP-AlF_4. The side length of the individual panels is 54 nm.

7.3: Discussion

Structural information on MDA5 has thus far been limited to an atomic model of the CTD and the mapping of the dsRNA-binding site on the CTD by NMR spectroscopy. To gain insight into the conformational changes that accompany ligand binding and activation, we sought to use single-particle EM to capture MDA5 in distinct conformations representing different functional states of the protein. In particular, the apo-MDA5 structure and the ADP.ALF_4-bound MDA5 structure were assumed to represent the inactive and active states of the protein, respectively. A comparison of these two structures would thus provide insight into the ligand-dependent activation of the protein. We also imaged a deltaCARD version of MDA5 to identify the position of the tandem CARDs. Successful identification of the CARD domains would enable a comparison of the positions of the CARDs in the inactive and active forms of MDA5, potentially providing a structure-based explanation for the regulation of MDA5. However, the EM averages of all the MDA samples were very similar and showed the same structural heterogeneity. The structural similarity of full-length and deltaCARD MDA5 suggest that the CARDs are too small and/or too flexible to create a density in the averages, and the structural similarity of MDA with and without bound ADP.ALF_4 suggests that this ATP analog does not lock the domains of MDA5 in a defined conformation. Furthermore, the lack of density for the CARDs indicates that they do not interact with the helicase domain, suggesting that the CARDs do not regulate MDA5 in a manner similar to RIG-I (through

autorepression). Further studies are thus required to elucidate how the CARDs regulate the signaling by MDA5.

Binding of ADP.AlF_4 did not lock MDA5 in an active, ligand-bound conformation. This may be due to APD.ALF_4 binding to MDA5 in the ALF_3 form, thus stabilizing the ground state rather than the transition state of the protein. Alternatively, the rigidity conferred on the helicase domain by ADP.ALF_4 binding may simply not translate to the CARDs and CTD.

We also used EM to further study the mechanism by which MDA5 dissociates from dsRNA. We have previously shown that dissociation occurs at a rate inversely proportional to the dsRNA length, suggesting that disassembly only occurs through dissociation from the filament ends (Peisley et al, 2011). EM observation of filament disassembly confirmed the length dependence of filament dissociation. While MDA5 completely dissociated from 1-kbp dsRNA in 60 seconds, it required 80 seconds for complete disassembly from 2-kbp dsRNA. However, unlike proposed previously, MDA5 did not only dissociate from the filament ends but also from multiple positions along the dsRNA. It is currently unclear what factors initiate MDA5 disassembly from dsRNA and define the positions of where disassembly occurs on the dsRNA strands. Analogous filament disassembly has been seen for the eukaryotic RecA analogue, RAD51. Like in MDA5, the central helicase domain of RAD51 consists of two RecA-like subdomains. RAD51 forms nucleoprotein filaments on ssRNA during homologous recombination, and, like MDA5, requires ATP hydrolysis for dissociation. The monitoring of RAD51 filament disassembly using magnetic

tweezers revealed multiple points of disassembly (Heijden et al, 2007). The authors suggested that disassembly is initiated at positions, in which the RAD51 oligomer is interrupted and contains monomers that only interact with a single neighboring subunit. Disassembly initiation would thus occur at sites where RAD51 filament extensions from the 5'- and 3'-ends of the nucleic acid meet and where there is a small stretch of unoccupied nucleic acid that is not large enough to accommodate another RAD51 monomer. Currently, we have no experimental evidence to suggest that MDA5 filaments disassemble in a similar fashion. We are in the process of performing kinetic association experiments, in which we aim to see multiple nucleation sites per strand of dsRNA, which would lend support to a RAD51-like mechanism of filament disassembly. Ultimately, however, a single particle approach will be required to fully understand the dynamic interaction of MDA5 with dsRNA.

7.4: Electron microscopy experimental procedures

Negatively stained samples: Negatively stained specimens were prepared by adsorbing 3 mL of sample to a carbon-coated EM grid, washing with two drops of deionized water prior to staining it with two drops of 0.75% uranyl formate. Images were taken on a Tecnai T12 electron microscope (FEI) equipped with a LaB6 electron source and operated at 120 kV. Images were recorded on imaging plates at a magnification of 67,000x and a defocus of –1.5 mm. Imaging plates

were read out with a Ditabis scanner (Pforzheim, Germany) using a step size of 15 μm, a gain setting of 20,000 and a laser power setting of 30% (Li et al, 2009).

Cryo-negatively stained samples: Cryo-negatively stained samples were prepared as described in Ohi et al (2004). In brief, Quantifoil grids from SPI (400 mesh copper holey carbon R2/1) were used. 10% glycerol was added to the sample as a cryo-protectant prior to deposition on the grid. Following sample adsorption and staining, a second carbon layer was applied to the specimen by using the following technique. A small square of carbon was allowed to float off carbon-coated mica onto the surface of a third stain drop. After incubation for 20 seconds, the grid was used to pick up the floating carbon to complete the formation of a carbon sandwich. Excess stain was removed by blotting against filter paper. After drying, the grid was frozen in liquid nitrogen.

2D average image processing: Particles were interactively selected using WEB, the display program associated with the SPIDER software package (Frank et al, 1996). Particles were windowed into individual 64 x 64 pixel images, rotationally and translationally aligned, and subjected to 10 cycles of multi-reference alignment. Each round of multireference alignment was followed by K-means classification specifying a pre-determined number of output classes. The references used for the first multireference alignment were randomly chosen from the windowed raw images.

3D reconstruction: 88 pairs of micrographs at 0° and 50° were recorded. 10,470 pairs of particles were interactively selected and windowed from the collected micrographs. The particles from the 0° images were then subjected to ten cycles of multireference alignment and K-means classification to yield 20 classes. A single class was used for 3D reconstruction using the random conical tilt method (Radermacher et al, 1987). An initial back-projection map was calculated using only the particles from the images of the tilted specimen. The best 10% of the particles from the images of the untilted specimen, as judged by cross-correlation with the class average, were added to the particle image stack for refinement. Eight rounds of angular refinement were performed in SPIDER to produce the final volume. The resolution of the final density map was estimated using Fourier shell correlation (FSC) and the FSC = 0.5 criterion.

7.5: References

Cheng, Y., Wolf, E., Larvie, M., Zak, O., Aisen, P., Grigorieff, N., Harrison, S., and Walz, T. (2006) Single particle reconstructions of the transferring-transferrin receptor complex obtained with different specimen preparation techniques. *J Mol Biol* 355:1048-1065.

Cordin, O., Banroques, J., Tanner, N., and Linder P. (2006) The DEAD-box protein family of RNA helicases. *Gene* 367:17–37.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadi, M., and Leith, A. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116:190-199.

Gipson, B., Zeng, X., and Stahlberg, H. (2007) 2dx_merge: Data management and merging for 2D crystal images. *J Struct Biol* 160:375-384.

Gipson, B., Zeng, X., Zhang Z., and Stahlberg, H. (2007) 2dx – User-friendly image processing for 2D crystals. *J Struct Biol* 157:64-72.

Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R., Diamond, M., and Colonna M. (2006) Essential role of mda-5 in type IIFN responses to polyriboinosinic: polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc Natl Acad Sci* 103:8459-8464.

Heijden, T., Seidel, R., Modesti, M., Kanaar, R., Wyman, C., and Dekker, C. (2007) Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules. *Nuc Acids Res* 35:5646-5657.

Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q., and Chen, Z. (2011) MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* 146:448-461.

Kang, D., Gopalkrishnan, R., Wu, Q., Jankowsky, E., Pyle, A., and Fisher P. (2002) Mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci* 99:637-642.

Kowalinski, E., Lunardi, T., McCarthy, A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D., and Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147:423-435.

Li, Z., Hite, R., Cheng, Y., and Walz, T. (2009) Evaluation of imaging plates as recording medium for images of negatively stained single particles and electron diffraction patterns of two-dimensional crystals. *J Electron Microsc* 59:53-63.

Linder, P., and Jankowsky, E. (2011) From unwinding to clamping – the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12:505–516.

Liu, Z., Shen, J., Carbrey, J., Mukhopadhyay, R., Agre, P., and Rosen, B. (2002) Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proc Natl Acad Sci USA* 99:6053-6058.

Leung, J., Pang, A., Yuen, W., Kwong, Y., and Tse, E. (2007) Relationship of expression of aquaglyceroporin 9 with arsenic uptake and sensitivity in leukemia cells. *Blood* 109:740-746.

McWhirter, S., Tenoever, B., and Maniatis, T. (2005) Connecting mitochondria and innate immunity. *Cell* 122:645-647.

Ohi, M., Li, Y., Cheng, Y. & Walz, T. (2004) Negative staining and image classification – powerful tools in modern electron microscopy. *Bio Proc Online* 6:23-24.

Peisley, A., Lin, C., Wu, B., Orme-Johnson, M., Liu, M., Walz, T., and Hur, S. (2011) Cooperative assembly and dynamic disassembly of Mda5 filaments for viral dsRNA recognition. *Proc Natl Acad Sci USA* 108:21010-21015.

Peisley, A. (2012) Personal communication

Pichlmair, A., Schulz, O., Tan, C., Naslund, T., Liljestrom, P., Weber, F., and Sousa, C. (2006) RIG-I mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* 314:997-1001.

Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc* 146:113-136.

Radermacher, M. (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J Electron Microsc Tech* 9:359-394.

Tsukaguchi, H., Weremowicz, S., Morton, C. & Hediger, M. (1999) Functional and molecular characterization of the human neutral solute channel aquaporin-9. *Am J Physiol* 277:F685-F696.

Viadiu, H., Gonen, T., and Walz, T. (2007) Projection map of aquaporin-9 at 7 A resolution. *J Mol Biol* 367:80-88.

Werten, P., Hasler, L., Koenderink, J., Klaassen, C., de Grip, W., Engel, A., and Deen, P. (2001) Large-scale purification of functional recombinant human aquaporin-2. *FEBS Lett* 504:200-205.

Yang, B., Zhao, D., and Verkman, A. (2006) Evidence against functionally significant aquaporin expression in mitochondria. *J Biol Chem* 281:16202-16206.

Chapter Eight: Conclusion

Bacterial and viral infections affect large numbers of people each year. While bacterial infections can be treated by the administration of antibiotics, the innate immune response is the first line of defense against viral infections. Part I of this thesis described my work aimed at reconstituting the enzymes that are required for the biosynthesis of the minimal pharmacophore of moenomycin A (MmA), a natural product with potent antibacterial activity. Part II of this thesis described single-particle EM studies of MDA5, a receptor that recognizes microbial double-stranded RNA (dsRNA).

MmA is a phosphoglycolipid natural product that is produced by *Streptomyces ghanaensis* (Ostash et al, 2007). Despite its excellent antibacterial potency, MmA has not been developed as an antibiotic because it displays poor pharmacokinetic properties (Wasielewski et al, 1965; Halliday et al, 2006). In order to enable the chemoenzymatic synthesis of unnatural MmA analogs that may possess better pharmacokinetic properties, I first attempted to reconstitute the enzymes needed to synthesize C15-farnesyl-CEF-trisaccharide and C15-farnesyl-DEF-trisaccharide. I was able to overexpress and purify MoeGT1, MoeE5, MoeF5, MoeGT4, MoeGT5, and MoeM5, the enzymes required for the chemoenzymatic synthesis of the C15-farnesyl-CEF-trisaccharide. All enzymes except for MoeM5 were overexpressed and isolated from *E. coli* in small quantities. MoeM5 was overexpressed in *S. lividans* because the protein formed insoluble inclusion bodies in *E. coli*. I also attempted to express MoeGT3, the glycosyltransferase required for the attachment of the D glycan ring, but was

unsuccessful because of problems in the production of soluble protein in both *E. coli* and *S. lividans*. I was able to reconstitute the activities of all purified proteins except for MoeM5 and MoeGT3. The activity of MoeM5 was not tested because I was unable not obtain sufficient amounts of disaccharide substrates through the chemoenzymatic route. The reason for the low yield of MmA intermediates was that all enzymes expressed poorly, some yielding only 0.2 mg protein per liter of culture. The enzymes also exhibited a low turnover rate, such that a k_{cat} could not be measured for any of the enzymes. In order to produce large quantities of MmA intermediates, both expression and reaction conditions would have to be optimized.

I also initiated preliminary work to define the promiscuity of MoeGT1, MoeGT4 and MoeGT5, glycosyltransferases that attach the F and E glycan rings, respectively. MoeGT1 couples C15-farnesyl-phosphoglycerate with UDP-galacturonic acid (UDP-GalUA). If MoeGT1 exhibits relaxed substrate specificity, one could synthesize MmA analogs with unnatural lipids. This possibility is of interest because the C25-moenocinol lipid tail of MmA is thought to contribute to the poor pharmacokinetic properties of MmA, and replacing the C25-moenocinol lipid with lipids with other chemical structures may improve the pharmacokinetic properties of MmA. Towards this goal, I synthesized a panel of 10 phosphoglycerate lipid acceptors, so that it is now possible to test whether MoeGT1 accepts these lipids. I also chemoenzymatically synthesized UDP-3-amino-N-acetyl-glucosamine (UDP-3-amino-3-doxy-GlcNAc) and assayed MoeGT4 for whether it could accept this sugar donor. Amino sugars are useful

chemical functionalities since they enable further chemical derivatization. Although the preliminary result was negative, it may be possible to find reaction conditions that would enable the enzyme to turnover. Taken together, the work on MmA detailed in Part I of my thesis represent significant progress towards the chemoenzymatic synthesis of the C15-farnesyl-CEF-trisaccharide and the first steps towards the exploration of enzyme promiscuity.

Part II of this thesis concerns structural studies of MDA5, an innate immune receptor that recognizes dsRNA (Linder & Jankowsky, 2011). Upon binding to dsRNA, MDA5 interacts with the mitochondrial membrane protein, MAVS, which activates the innate immune response through a cascade of signaling events (Berke & Modis, 2012). I used single-particle EM to study the structures of both MDA5 alone and in complex with different nucleic acids.

2D averages of MDA5 alone revealed that the protein adopts an elongated conformation and is very flexible. I attempted to stabilize the structure of the protein by using ATP analogs and by using an MDA5 construct that lacked the CARD domains. Both approaches did not improve the conformational homogeneity of the protein. The extreme heterogeneity displayed by MDA5 suggests that MDA5 is not autoregulated in a similar manner as RIG-I. Structural studies of RIG-I have revealed that in the absence of dsRNA ligand, the CARD domain of RIG-I is bound to a portion of the helicase domain, thus preventing downstream signaling (Kowalinski et al, 2011). If MDA5 exhibits a similar type of autorepression in the absence of ligand, we would expect the protein to adopt a more homogenous conformation. Instead, our averages suggest that both the

CARD and C-terminal domains of apo-MDA5 are flexibly linked to the helicase domain and are exposed in solution. This finding raises the question of how MDA5 prevents signaling to IPS1 in the absence of viral infection, which would be an interesting topic for future studies.

Biochemical experiments revealed that MDA5 is able to bind to dsRNA, ssRNA, and dsDNA with high affinity and similar dissociation constants. I used single-particle EM to examine the complexes formed between MDA5 and the different nucleic acids. When MDA5 was incubated with 112-bp dsRNA in the presence of ADP.AlF_4, 16 MDA subunits assembled to form a linear oligomer consisting of eight stacked discs. I then calculated a 3D reconstruction for the MDA5-dsRNA complex using single-particle EM of cryo-negatively stained samples. The 3D map resolved the eight discs, but the flexibility of the linear oligomer limited the resolution of the 3D reconstruction and thus did not provide additional insights into the organization of the MDA5 subunits in the dsRNA-induced oligomer.

Filaments consisting of stacked discs were rarely observed with MDA5 bound to ssRNA and dsDNA. Instead, these samples contained mostly amorphous complexes of varying sizes and shapes. We concluded that these complexes represent abortive complexes that do not extend to cover the entire length of the nucleic acid. These structural observations are in agreement with biochemical results showing that only the productive MDA5-dsRNA complexes hydrolyze ATP. We currently do not understand how filament formation and ATP hydrolysis activate MDA5 downstream signaling. One possibility is that

productive filament formation results in a conformational change in MDA5 that exposes the CARD domains. Alternatively, ATP hydrolysis could be used to effect a similar conformational change that exposes the CARD domains. In order to understand the mechanism of MDA5 activation, it would be interesting to obtain atomic structures of the MDA5-dsRNA complex, the MDA5-dsRNA-ATP analogue ternary complex, and potentially complexes with the downstream protein MAVS bound. Such structures would greatly aid in delineating the roles of filament formation and ATP hydrolysis in downstream signaling by MDA5.

I also used single-particle EM to visualize the complexes that are formed by I923V MDA5 on dsRNA. The I923V mutation in the C-terminal domain of MDA5 has been associated with a reduced risk for developing Type I diabetes and a reduced ability for the complex to signal to downstream proteins (Nejentsev et al, 2009). The EM averages showed that the mutation disrupts complex formation. Whereas the wild-type protein contained seven to eight stacked discs per strand of 112-bp dsRNA, the mutant complex often contained only four to six stacked discs. These experiments suggest that disc stacking is required for proper interaction with IPS1. The mechanism of this interaction is not understood and is an intriguing question to explore.

References

Berke, I., and Modis, Y. (2012) MDA5 cooperatively forms dimmers and ATP-sensitive filaments upon binding double-stranded RNA. *EMBO J* 31:1714-1726.

Halliday, J., McKeveney, D., Muldoon, C., Rajaratnam, P., and Meutermans, W. (2006) Targeting the forgotten transglycosylases. *Biochem Pharmacol* 71:957-967.

Kowalinski, E., Lunardi, T., McCarthy, A., Louber, J., Brunel, J., Grigorov, B., Gerlier, D., and Cusack, S. (2011) Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell* 147:423-435.

Linder, P., and Jankowsky, E. (2011) From unwinding to clamping – the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12:505-516.

Nejentsev, S., Howson, J., Walker, N., Szeszko, J., Field, S., Stevens, H., Reynolds, P., Hardy, M., King, E., Masters, J., Hulme, J., Maier, L., Smyth, D., Bailey, R., Cooper, J., Ribas, G., Campbell, R., Clayton, D., and Todd, J. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* 450:887-U19.

Ostash B., Saghatelian, A., and Walker, S. (2007) A streamlined metabolic pathway for the biosynthesis of moenomycin A. *Chem Biol* 14:257-267.

Wasielewski, E., Muschaweck, R., and Schutze, E. (1965) Moenomycin, a new antibiotic. *Antimicrob Agents Chemother* 1965:743-748.