Genetic Incorporation of Unnatural Amino Acids for the Study of Protein-Protein Interactions

By

Rachel E. Pricer

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2016

Doctoral Committee:

Professor Anna K. Mapp, Chair
Assistant Professor Jolanta Grembecka
Assistant Professor Brent R. Martin
Assistant Professor Zaneta Nikolovska-Coleska

Acknowledgements

I have a lot of people to thank for making it through graduate school. It is probably the hardest thing I have done and there were many points where I thought I might not finish. Thinking of discussions I had at critical points in my time here, it was always with, my advisor, Anna Mapp. When I had trouble during rotations, switching smoothly between labs, and even just with research troubles along the way, a meeting with Anna has always helped fix things. I am extremely grateful I was able to join her lab and it has made all the difference for my career.

I also have to thank my second mentor, Brent Martin who has always jumped in with helpful ideas, research suggestions, and advice. I appreciate the time I have spent in your lab and with your group. It has brought a unique perspective to my research and reenergized me often when I'm feeling stuck. I am also grateful to rest of my committee and the PCB staff who have provided much needed help and guidance along the way whether it be with research or getting paid on time.

Over the years I have had wonderful friends and labmates who have helped shape me as a scientist and person. I have loved sharing the graduate experience with my PCB cohort. Everyone was friendly and welcoming from day 1 and were a great support group through the first few years. I owe much of my molecular biology knowledge to Dr. Keith Green who was an amazing mentor and friend. He really shaped the scientist I am today and I am forever grateful for my time with him. I also owe much my current biochemistry knowledge and training to Dr. Amanda Dugan. She took me on as a refugee student when she was trying to graduate and was already swamped. I had always looked up to her since she had been my recruitment host and had no idea I would end up working with her which was one of the best experiences of my PhD. I also am thankful for the remainder of the Mapp lab who were so welcoming and just all around a fun group. I also have to thank the new students that have brought life back to a lot of the projects especially during times I was struggling. Melanine and Omari have become amazing scientists and friends. I look forward to seeing the amazing things both of you

continue to do. Also I have to thank Jean who has been an endless place for support and friendship, I'm so glad we went on that Madison conference together and had those Spotted Cows.

Finally, my family has always been there for me. My mom especially has just listened to frustrated rants on the phone patiently to let me vent or flown out here to sit and help me work. You guys have been a part of all of my Ann Arbor housing traumas and have always helped pick up the pieces, of my house if need be. I love you and always look forward to the next time we see each other which never seems to be soon enough. Rob you have been the rock I have always needed even when I've been an overtired, psychotic, and stressed-out wreck. I don't know where I'd be without you, you are my soulmate. Finally, I dedicate this thesis to Nora, my ever patient scruffy little dog. For all nights I came home crabby and didn't want to play or go for a walk and you still excitedly jumped up to see me and for all the nights now when I'm lonely and can snuggle up in bed with you, you've brought joy to this whole difficult endeavor.

Table of Contents:

List of Figures:

List of Tables:

List of Appendices:

Abstract

Genetic Incorporation of Unnatural Amino Acids for the Study of Protein-Protein Interactions

Proteins comprise the majority of the cell and are vital to all cellular functioning. Protein-protein interactions (PPIs) are the communication networks behind cellular processes, often functioning in machine-like complexes with exchangeable subunits or parts to convey different messages. PPIs exhibit a wide range of structural features, surface areas, and affinities with some displaying dynamic interfaces allowing multiple binding partners to interact depending on cellular conditions. This makes some PPIs more difficult to study than others. Understanding these PPIs and exploring larger PPI networks has been a challenge without considering the cellular context in which they belong. Methods to study difficult PPIs in their native environments have thus been instrumental advancing the field.

The predominant theme of this work is the demonstration of the utility of genetically incorporated photolabile unnatural amino acids for the study of the difficult PPIs between transcriptional activator-coactivator complexes. Covalent chemical capture of protein binding partners in live cells is combined with mass spectrometry to discover novel PPIs and further expanded to include new ways to visualize direct PPIs on DNA. Caveats to the covalent capture method are also explored with an illustration of capture efficiencies of two common photolabile groups across various PPI binding affinities and surface areas.

The work presented here displays a thorough examination of the use and application of chemical capture for the study of PPIs in a cellular context. The methods established within this work add to the foundation for the study of difficult PPIs and demonstrates the ability to understand new networks of low affinity, dynamic interactions. The presentation of novel binding partners for the well-studied transcriptional activator, Gal4, expands traditional beliefs on transcriptional activator participation in binding dynamic complexes as well as highlights the potential of these PPIs for later therapeutic points of intervention. In addition, the groundwork for guidelines on using covalent chemical capture in various PPIs was established which, when

completed, will enable not only easier use but also hopefully lead to the ability to tailor selection of a photocrosslinker based on the specific PPIs under study.

Chapter 1: Introduction – Defining Protein-Protein Interactions (PPIs): Challenges and Strategies in the Field[1]

1A: Abstract

Protein-protein interactions underlie every biological function in the cell, and as a result, dysregulation of PPIs accompanies many human diseases. For example, approximately 30% of human PPIs have been linked to cancer.[3] Due to their fundamental importance and role in disease, PPIs are a logical target for drug modulation; however, they are often challenging to study due to limited structural data, transient binding, and numerous binding partners, thus presenting many hurdles to a comprehensive understanding of the PPI network.[6] One category that has been a considerable struggle to study are large, transient PPIs with this group frequently lacking small molecules even for tool purposes.[8] Transcriptionally-related proteins are often within this large and transient interface category, leaving one of the most biologically important processes without methods to study their PPIs.[11] Employing *in vivo* techniques to investigate these proteins has vastly expanded our ability to study their PPIs with the focus of this thesis being the utilization of genetically incorporated photocrosslinking amino acids to further the understanding of transcriptional activator PPIs.

1B: Defining PPIs

Proteins make up a large portion of a cell, comprising approximately 60% of its dry mass.[12] The interactions between these proteins are integral to cellular functioning and control many cellular processes. Signals transmitted through PPIs control everything from localization and activation of specific cellular constituents to the timing of events within the cell. As a result, dysregulation of PPIs accompanies many human diseases. One example is the disruption of c-Myc in many human cancers. C-Myc is a linchpin in many cellular processes ranging from proliferation to apoptosis, and disruption of its

[1] Portions of this chapter are taken from: Mapp, A. K.; Pricer, R.; Sturlis, S., Targeting transcription is no longer a quixotic quest. *Nat Chem Biol* **2015**, 11 (12), 891-4. My contributions to this work included Figure 1-1, Figure 1-2, Figure 1-4, Figure 1-5 and assisting in writing.

function leads to tumor regression, underscoring the importance normal functioning PPIs. Other examples include c-Myb, a transcription factor that is often required for the transformation of myeloid progenitor cells into a leukemic state;[13] the tumor suppressor p53, is misregulated in more than half of all human malignancies;[14] and the Ets family of transcription factors, which affect the status of growth receptors such as EGFR and Her2 and contribute to metastasis.[3] Given the fundamental importance of PPIs, they are a logical target of study for understanding disease, however, the broad landscape of PPIs has made it difficult to parse out clear interaction networks.

Several features of PPIs make their study a challenging endeavor. Proteins in many PPIs participate in a wide variety of interactions with multiple interacting partners that have varying cellular roles.[15] Interactions may occur only under particular cellular conditions or regions of a cell. In addition to compositional dynamics within PPI complexes, individual subunits may exhibit significant conformational plasticity as a means to interact specifically with a variety of binding partners. Although challenging, studying complex PPI networks and understanding their purposes within a cell provides a foundation for understanding what malfunctioning interactions cause disease. Regaining normal PPI functioning or blocking unwanted PPIs has been a long-term goal of drug development. Although many years of research have gone into understanding PPIs, developing ways to better study and understand interaction networks still needs to be done.

In order to examine the critical features of PPIs, several classification systems have been analyzed. Several defining features for classifying PPIs include biological role, protein structure, interface surface area, and interaction binding affinity. Classification of PPIs by biological role has proven useful, as similar function within a cell often means a similar mode of action and structure. In addition to functional classification, structural homology is a useful tool for proteins that have well-ordered structural features. A molecular basis for the interaction can be established, since pinpointing specific requirements and contact points between the proteins is possible. For proteins that have less-defined structural features, in many cases it is possible to determine a range of amino acid residues that comprise the minimal interaction sequence and thus allows an estimated surface area of interaction. An illustrative example of this is

the interaction between the yeast activator Gal4 and its repressor Gal80 for which limited structural data is *Ava*ilable. Gal80 is a readily crystalized protein but Gal4 is largely unstructured and only a few residues have been well-defined in crystallography attempts.[16, 17] In addition to useful properties gleaned from the amino acid content such as potential surface area and hydrophobicity of the interaction interface, more traditional biochemical methods can then be applied to deepen the understanding of the interface interactions. For instance, mutational studies have been extensively used to assess a specific residue's binding contributions to an interaction.[18] Although understanding of specific molecular interactions between two proteins may be limited, this information aids in moving forward in accessing the size and potential for small molecule targeting of the interface. Finally, PPIs may be classified by the binding affinity of an interaction. Binding affinity is a measure of how tightly two partners interact and often gives clues to the nature of an interaction. For example, in processes that may only occur for a short period of time, such as gene expression, the "turning on" of the molecular switch requires

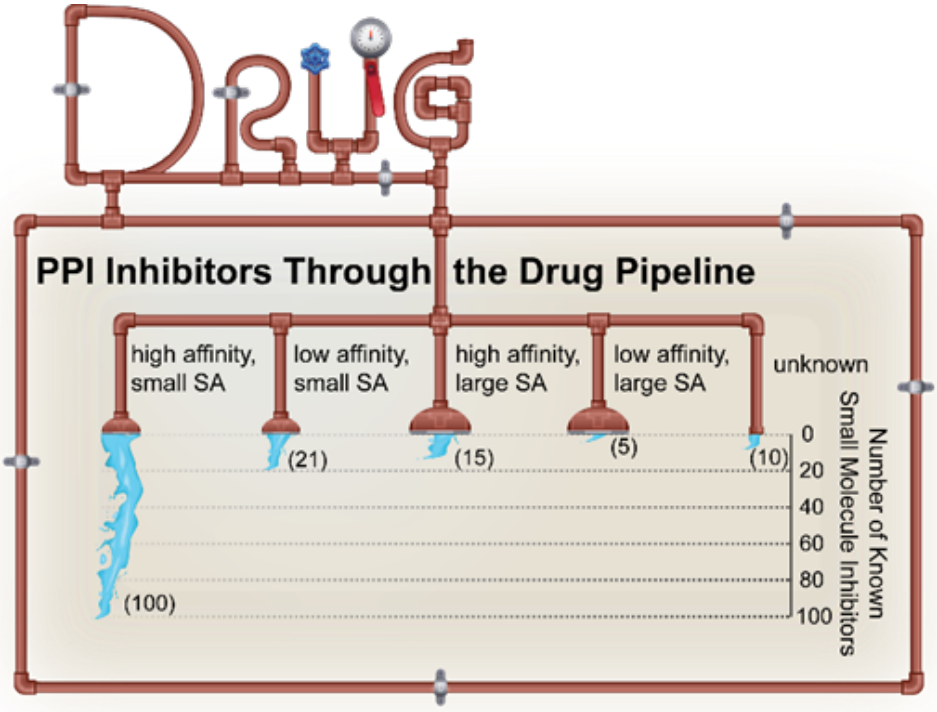


Figure 1-1 Inhibitors Targeting Different Areas of PPI Chemical Space.
Recent advances in protein-protein interaction inhibitor design and screening methodologies have led to the discovery of a number of new small-molecule inhibitors, although success in targeting high-affinity, small-surface-area interactions has far outpaced broader or weaker interactions. SA, surface area.

binding affinity to be weak and transient, otherwise, the molecular switch would be very difficult to "shut off."[19]

Using surface area and binding affinity to plot the number of drugs discovered for PPIs, we can see a trend. The largest body of successful work has focused on PPIs that have tighter interactions with a smaller surface area (Figure 1-1).[8, 10] These PPIs are usually the most enzyme-like with defined binding pockets more amenable to small molecules. One area that has particularly struggled are PPIs with large binding surfaces and transient affinities.[15] Recent advances in biochemical methods, however, have allowed for the easier study of larger, more transient PPIs. Although, the general advancement of laboratory instrument technology (cryo-electron microscopy and mass spectrometry for example) has allowed for the study of larger protein complexes, the *Ava*ilability and application of biochemical probes has dramatically increased the convenience of doing *in vivo* studies. This has allowed the field to advance toward conducting experiments involving PPIs within their native environments which decreases the chances of overlooking a weaker PPI and keeps larger interacting protein complexes assembled and in their functional forms.

An additional complicating factor in many PPI studies is the dynamic interfaces that exist for proteins that are involved in multiple PPIs. One example of this can be found by looking at the GACKIX (Gal11, Arc105, CBP/p300, kinase-inducible domain interacting) domain of the master coactivators and histone acetyltransferases CBP and p300. GACKIX is highly plastic, and its two binding surfaces can accommodate more than 15 distinct binding partners, in the context of binary or ternary complexes, that are involved in a

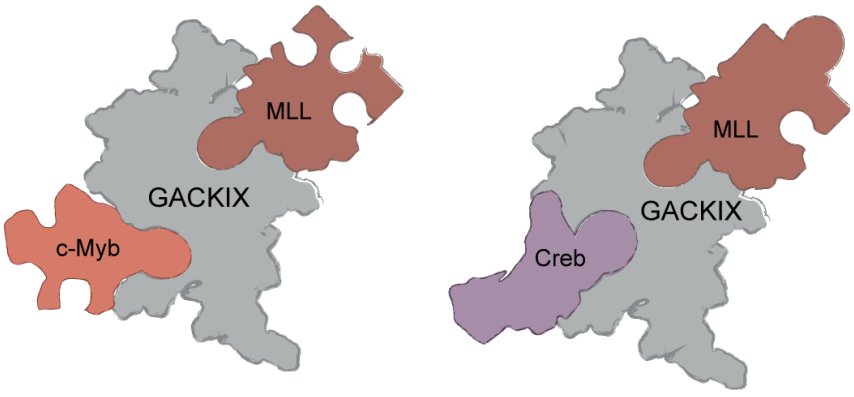


Figure 1-2 Dynamic Protein Complexes.
GACKIX has two binding sites to interact with more than 10 distinct transcriptional activators with each ternary complex having a signature conformation.[1]

variety of physiological processes and implicated in diseases from cancer to neuropathic pain (Figure 1-2).[1] Low energy barriers between individual conformations mean that a participant can use the same group of amino acids to recognize a variety of binding partners, with each complex requiring a distinct conformation. Much like the puzzle depiction of this process, synthetic regulation of complex assembly represents an opportunity to "freeze" the protein so only certain partners (or puzzle pieces) are able to bind furthering the selectivity of a molecule and give it context specificity.

1C: Transcriptional Activators and Their PPIs

Transcription-related proteins which are involved in the process of moving from information encoded within the DNA to messenger RNA (mRNA) for later translation into functional proteins combine aspects of the more difficult interactions to study mentioned previously such as large, low affinity, and dynamic PPI interfaces. Although necessary for their biological function, to assemble the necessary proteins and protein complexes to eventually recruit RNA Polymerase II and begin the transcription process,[20](Figure 1-3), the proteins in this process often contain dynamic binding interfaces allowing them to participate in many different interactions.[21] One study found that of 401 transcription factors computationally analyzed, approximately 49% of the sequences indicated intrinsically disordered regions.[6] In fact, a disordered to ordered transition is assumed to occur upon binding of proteins such that alterations in structure

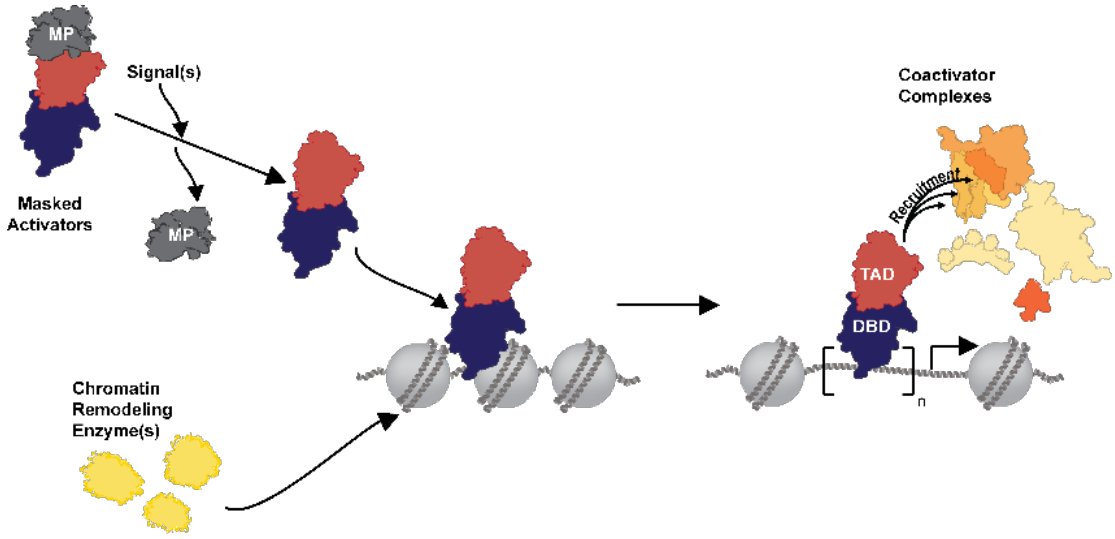


Figure 1-3 Transcription Initiation Process.
Transcriptional activators participate bind DNA and begin recruiting coactivator complexes for the pre-initiation complex (middle). This collection of complexes involves those such as Saga, Swi/Snf, Mediator, and RNA Pol II. Once RNA Pol II has been recruited, transcription can proceed.

may exist between different binding partners.[22, 23] These broad and weak interactions lead to distinct networks of PPIs being used to assemble the transcriptional machinery and usually begin with one protein, a transcriptional activator. To recruit proteins to a specific DNA sequence, a transcriptional activator will bind to DNA and begin recruiting protein complexes to modify and remodel the chromatin to make room for the recruitment of other proteins such as the general transcription factors and the large coactivator complex, Mediator[24, 25]. The recruitment of these factors leads to the binding of RNA Polymerase II and allows the initiation of transcription (Figure 1-3).

The proteins at the root of this process, transcriptional activators, wield great power within a cell by promoting or inhibiting the expression of genes. This powerful role of activators places a bull's eye on them for controlling and understanding transcriptional PPI regulation. However, transcriptional activators have several features that make them difficult to study. Traditional probe discovery or design methods are ill equipped to target these complexes because one or both binding partners are classified as intrinsically disordered proteins, and the complexes often form transiently in the cell, with affinities one to two orders of magnitude weaker than the category of small surface area and high affinity interactions previously discussed.[8, 22] Further, the interaction surface is often considerably larger, with interaction energies shared over greater numbers of amino acids, defying hotspot analysis. Using the prototypical transcriptional activator p53 as an example, we can see the wide range of affinity and surface areas a protein involved in many transcriptional PPIs can exhibit, (Figure 1-4). The complexes formed between p53 and its regulatory partners such as MDM2 are typically high affinity with a relatively small surface area, and their interaction energy largely resides in a small number of residues (Figure 1-4, upper left image). These are structurally well-organized interfaces and are highly amenable to structural characterization. As with many activators, p53 typically functions as a multimer, and the oligomerization interface is another high affinity, well-organized interface, although it takes place over a considerably larger binding surface (Figure 1-4, upper right image). The final group of PPIs that activators such as p53 utilize are those with coactivator proteins, typically involved as part of the assembly of the transcriptional machinery in the early stages of

transcription (Figure 1-4, lower image). These interactions are most poorly characterized, likely because the binding partners are conformationally dynamic, inhibiting high-resolution structural studies, and many are of only moderate affinity.

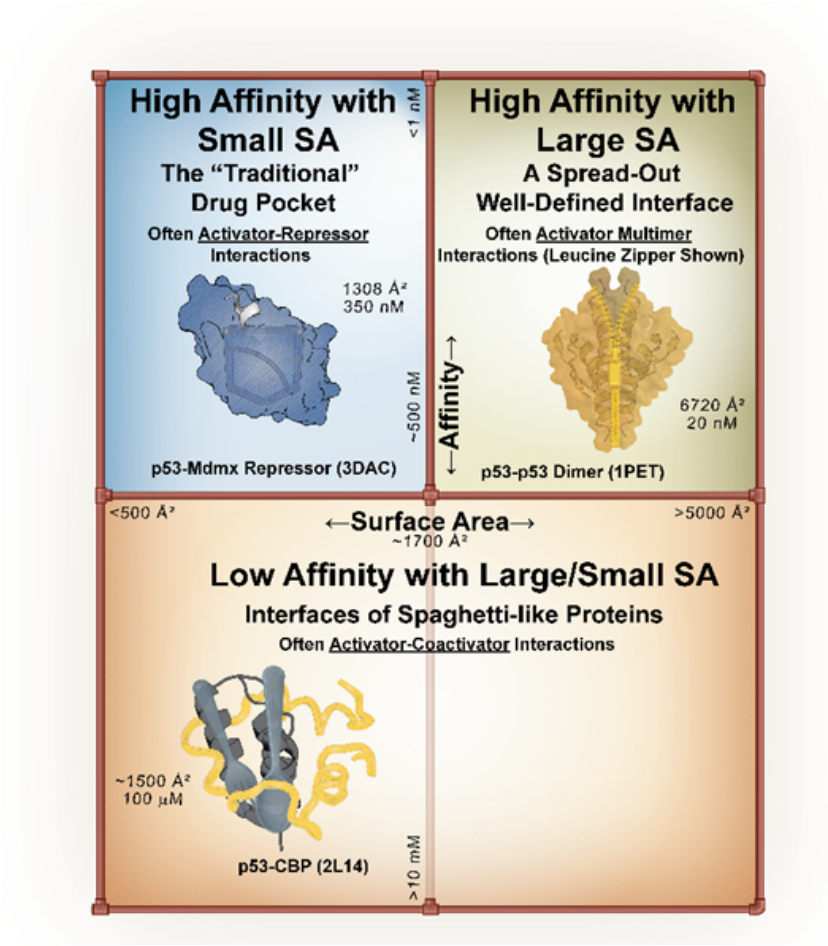


Figure 1-4 Chemical Space of Protein-Protein Interactions. Protein-protein interactions can be effectively classified by the strength of the complex (Y-axis) and the surface area over which the interaction occurs (X-axis). Using p53 as an example, transcriptional activator interaction networks span a broad range of strength and surface area. PBD ID codes for each structure are as follows: p53-Mdm2 repressor, PDB 1YCR; p53 dimer, PDB 1PET; p53 with CBP 2L14.

Although these features of transcriptional PPIs make them difficult to study, they are critical to the function of the activator and are responsible for their ability to control the level of gene expression. While a single interface is utilized for interacting with many proteins it also allows for exceptional specificity. The larger interfaces that take on a dynamic structure, many varying with each binding partner or regulatory signal allows for the very specific targeting of a given interaction, theoretically without disturbing other interactions in which a given protein may participate in.[22, 26] For example, screening techniques that directly address the conformational plasticity of coactivators have been effective for identifying small-molecule modulators that capture distinct conformers. The site-directed fragment screening strategy of tethering first developed at Sunesis is one such strategy. When applied to the conformationally dynamic GACKIX motif, (Figure 1-2) researchers identified chemical co-chaperones that

stabilize a range of GACKIX conformations and dictate the formation of particular GACKIX– transcription factor assemblies either positively or negatively.[27] In addition to conformational specificity, context specify is also a factor in transcriptional proteins. One example of this is the BAF-type chromatin remodeling complexes that play a key role in transcription initiation. Most share a core enzymatic subunit (the ATPase Brg), but there are exchangeable subunits that vary according to tissue (Figure 1-5).[2] Recently, evidence has emerged that transcription factors target both the enzymatic component and exchangeable modules in such complexes as part of the assembly of the transcriptional machinery. Thus, one could imagine blocking a PPI that would affect the localization of a complex at particular gene promoters but would leave the core enzyme unaltered, providing functional capacity in other contexts. The ability to prevent or initiate specific transcription events through controlling this process would be a huge gain for the medical field as the misregulation of transcription is critical to many human diseases (Figure 1-6).

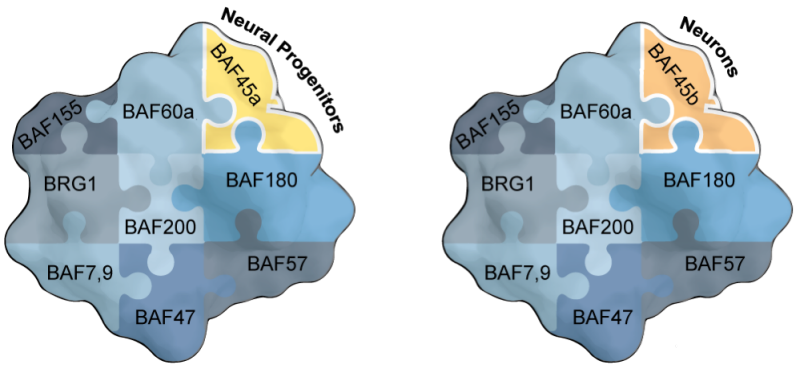


Figure 1-5 Dynamic Composition of Transcriptional Complexes Offers a Route to Specificity.
The BAF chromatin remodeling complexes contain the same enzymatic core (BRG1) but have exchangeable subunits such as BAF45a and BAF45b that define tissue specificity.[2]

To begin deciphering the PPIs that are part of the network of transcriptional activators, this body of work focuses on amphipathic activators which are modular in makeup. Amphipathic activators have a structured DNA-binding domain (DBD) tethered to an intrinsically disordered transcriptional activation domain (TAD) that usually only gains a concrete structure upon binding a transcriptional coactivator (Figure 1-3).[21] As their name suggests, amphipathic activators are characterized by the predominance of amphipathic amino acid residues within the TAD that are important for coactivator binding. A large body of classical transcriptional studies focused on amphipathic activators giving the foundation for subsequent methods to investigate transcriptional

PPIs. Most early studies in this area focused on yeast, an easy to manipulate model, since the transcriptional apparatus is conserved across species. As it is responsible for producing all mRNA for gene translation, RNA Pol II and its associated factors is one of the most universally homologous protein complexes across species, and important for understanding misregulation in human disease[28, 29]. Over the years, using yeast as a model, researchers have looked at many of the components needed by RNA Pol II for transcription as well as start to understand their assembly[20, 28]. The amphipathic yeast activator, Gal4, has been historically used to study transcription[30] and is particularly useful due to its activation of gene expression in response to nutrient *Ava*ilability in yeast.[31-35] This allows for easy induction of this activator and its target genes. The action of Gal4 has been well studied and an abundance of data exists pertaining to the activation

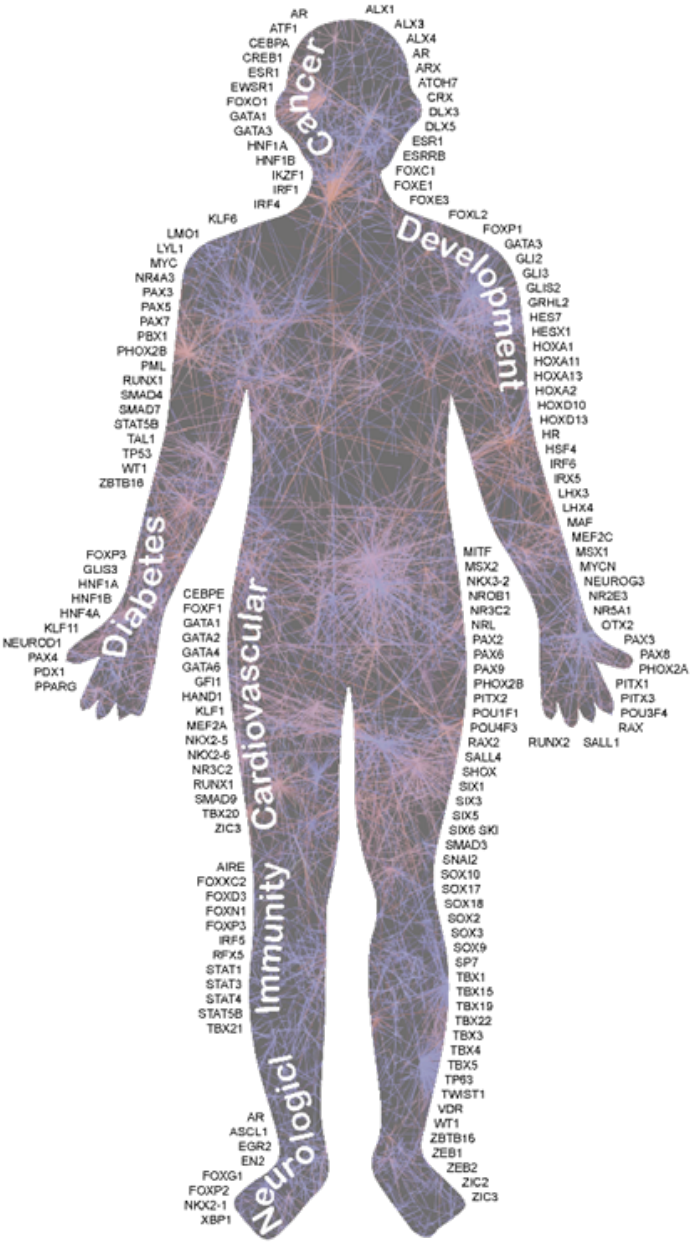


Figure 1-6 Transcriptional Activators and Consequences of Misregulation.
With many components needing to come together, there are many places for error during the transcription initiation process from lack of initiation to over expression to alteration of complex components. Outlining the body, activators involved in specific diseases are shown indicating how many activators are involved in disease not accounting for other factors such as coactivators, and chromatin remodeling factors.[7]

domain including mutational studies, binding sites in the genome, interactions with repressor protein Gal80, and pathways activated by this gene[17, 18, 31, 32, 34-48]. Using this wealth of information, the Gal4 transcriptional activator in yeast is an ideal system for pursuing methods to look at undiscovered binding partners and to gain insight into novel proteins/protein complexes activators may recruit to a promoter.

Contrary to the seemingly well-rounded model activator Gal4 presents, there are still many questions surrounding activator recruitment during the transcriptional activation process. One activator, discussed further in Chapter 2, VP16 is a viral activator from the Herpes simplex virus and has had years of swirling debate regarding the recruitment of TATA-binding protein (TBP) at the GAL1 promoter. The traditional view of activator recruitment is the binding of SAGA, a histone acetyl transferase that is believed to aid in mediating interactions with the general transcriptional machinery, and Mediator, which attracts the nucleosome remodeling complex SWI/SNF, followed by the rest of the general transcription factors and RNA Pol II. Based off this model, several groups have shown using ChIP and crosslinking methods *in vivo* that TBP may be recruited to the promoter through interactions with the SAGA complex.[49-52]. In contrast with the accepted model, using purified protein, different groups have shown the ability to co-precipitate TBP from lysate using the activator VP16 as well as view binding with immobilization assays and fluorescence polarization.[53-56]. As will be highlighted in Chapter 2, the methods that give us the ability to decipher what is recruited by transcriptional activators during transcription initiation allows us to edge our way closer to understanding the process and developing molecular tools to aid in its modulation.

1D: Strategies for a New Age of Transcriptional Activator PPI Studies

1DI: Current Methods Applied to PPIs

With the described difficulties of investigating PPIs, a menagerie of methods has been created. Some of the traditional biochemical methods used to study protein-protein interactions include co-immunoprecipitation and affinity purification followed by Western blotting. These methods involve the selection of a protein through the attachment of epitope or affinity tags such as a Flag tag and poly-histidine tag or biotin, respectively[57, 58]. The binding of an antibody or resin then allows the pull-down of the protein of interest for Western blot analysis (Figure 1-7). While this method provides a

very clear display of interacting proteins, it often misses those that have weaker binding affinities that cannot withstand washing during a pull-down. These methods are also not amenable to proteins that aggregate or easily become misfolded. Moreover, this technique is performed *in vitro* using cell lysates and thus interactions that are only observed in the native context are not seen.

The desire to keep proteins in their native environments while evaluating PPIs has led to the development of several *in vivo* techniques such as the yeast two-hybrid, Förster resonance energy transfer (FRET), and protein complementation methods. These methods involve expression of the binding partners of interest within yeast or mammalian cells, allowing a native environment to be maintained, while looking for the interaction of interest (Figure 1-7). The yeast two-hybrid involves the expression of the protein of interest attached to a DNA-binding domain, often of Gal4, while co-expressing the hypothesized binding partner fused to a transcriptional activation domain[59]. Thus, if the proteins are interacting partners, a complete transcriptional activator is present and transcribes a reporter gene of choice. In a similar fashion, protein complementation methods fuse half of a fluorescent protein such as GFP or a bioluminescent protein such as luciferase to the protein of interest with the other half fused to the hypothesized binding partner. Once the two binding partners interact, a complete protein of fluorescent or bioluminescent nature is made and can be detected[60, 61]. FRET applies the use of two fluorescent proteins or molecular probes such that, one fluorophore can be excited and will emit at the excitation level of the second fluorophore which will emit at a wavelength characteristic to that fluorophore[62, 63]. As with pulldown methods, these *in vivo* techniques are not often amenable to protein structures or interactions sensitive to changes, as they involve the attachment of large tags or proteins. It also requires the knowledge of both binding partners, so it limits a discovery motivated search coupled with the fact that using fluorescent readouts can often produce results that have a high background, making them difficult to apply to certain cellular conditions.

In addition to general PPI methods that have been developed, further methods have been established for DNA-bound proteins. Classical methods such as chromatin immunoprecipitation, ChIP, which employs the non-specific crosslinker formaldehyde to crosslink protein complexes to each other and DNA for immunoprecipitation and

verification at a given promoter (usually with an established activator) by PCR amplification of pulled down DNA pieces, have elucidated many transcriptional PPIs[64, 65]. This approach gives information about PPIs at the gene promoter under selected cellular conditions by "freezing" live cells to crosslink the PPIs in their native environment (Figure 1-7). New techniques along this line of thought such as ChIP-Seq, also employ the principles of pulled-down DNA to map proteins to specific genomic locations[66, 67]. One disadvantage of these methods is the lack of selectivity. The formaldehyde crosslinking enables the pull-down of all proteins in a complex and doesn't necessarily indicate a direct interaction if found. Expanding on this method, a new method to improve the ability to observe direct binding partners at the promoter has been developed and discussed in detail in Chapter 2.

The use of chemical crosslinkers such as amine-reactive N-hydroxysuccinimide esters, sulfhydryl-reactive maleimides, photocrosslinkers or formaldehyde (as mentioned above) have also been employed for the covalent attachment of binding partners within a cell[68-70]. These methods have proven successful in identifying many PPIs and essentially suspend the protein interactions in their current state (Figure 1-7). Although small and providing minimal structure perturbation, these crosslinkers do suffer from a lack of selectivity. Most crosslinkers employed are non-specific or attack a specific amino acid type such that entire complexes are crosslinked and it can be hard to parse out direct PPIs versus indirect PPIs through the complex.[26] As will be discussed in the next section, the recent application of crosslinkers to *in vivo* studies has greatly expanded their usefulness in looking at direct PPI binding partners in their native context.

A final and relatively new approach is the use of proteomics to identify novel PPIs. This employs a mass spectrometer to analyze digested peptide fragments that have resulted from a pull-down of the protein of interest. Quantitative measurements can also be made by using stable isotope labeling with amino acids in cell culture (SILAC) which applies a "heavy" or isotopically labeled amino acid to one sample to compare to a "light" or normal sample control. The abundance ratio between light and heavy can then be used to quantify protein amounts between samples (Figure 1-7)[71]. Similarly, crosslinked protein or reactive tags that are isotopically labeled or mass spec-cle*Ava*ble are also used for identification of PPIs[70]. Though most studies done to map activator-

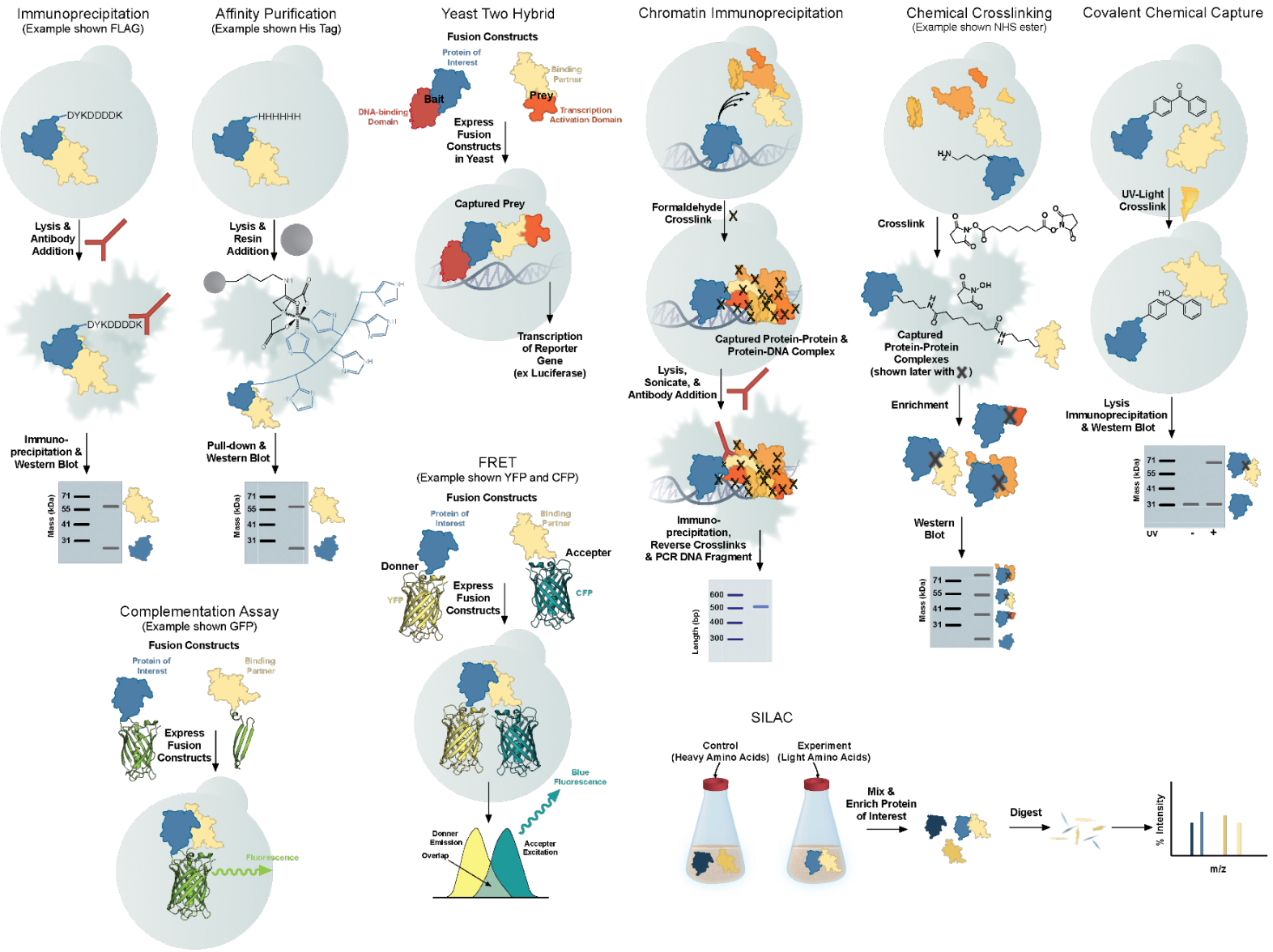


Figure 1-7 Methods Schematic. Displays methods discussed and illustrates what hypothetical results of the method may look like.

coactivator interfaces and activator interactions have been done using cell lysates or intact protein complexes, as pull-down efficiencies have increased with newer techniques and mass spectrometry technology has advanced, the ability to detect lower abundance proteins within the cell has improved as this has historically been a limiting factor to this strategy.

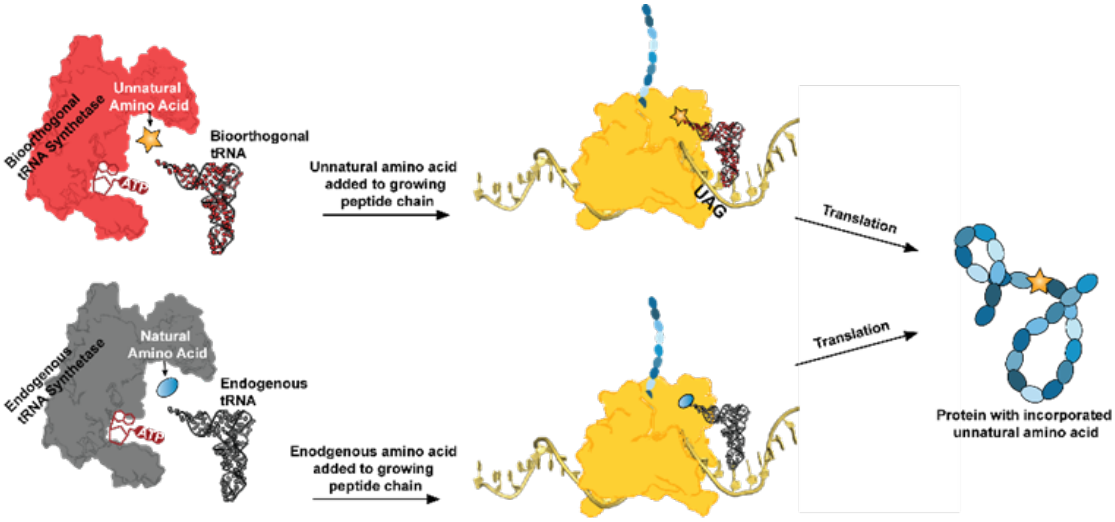


Figure 1-8 Genetic Incorporation of Unnatural Amino Acids by Nonsense Suppression. Inserting a biorthogonal tRNA and tRNA synthetase pair into the expression system of choice with the complementary UAA allows for the site-specific incorporation of the UAA at the amber stop codon TAG without any interference to the normal translational machinery or tRNA pairs.

1DII: Genetic Incorporation of Photolabile Unnatural Amino Acids

In addition to the methods described above, another *in vivo* method of PPI identification applying crosslinking chemistry in cells has been developed. This method inserts photocrosslinking amino acids in a site-specific fashion to allow for their use to covalently capture binding partners surrounding them. Genetic incorporation of unnatural amino acids (UAAs) was first used by the Peter Schultz laboratory[72, 73] and has since expanded to include over 100 different UAAs[73]. This method relies on the fact that there are 3 stop codons present within the genetic code which are conserved throughout all of life. The least used amber stop codon or TAG is nonsense suppressed by an orthogonal tRNA carrying the UAA of choice. This tRNA also has a companion tRNA synthetase which loads the UAA onto the tRNA for incorporation (Figure 1-8.)[74]. Although this method often suffers from production loss due to truncation products (100% suppression of the amber codon is not possible) it has the advantage of producing protein with minimal tags or changes to the sequence. Due to the large increase in

species capable of incorporating unnatural amino acids, often the protein can even be expressed in its native organism.

Once produced, the protein can function in the cell as it naturally would and bind to its partners in the native cellular context. This makes the method extremely useful as it enables *in vivo* characterization, is amenable to all protein types, and disruption of normal functioning is minimal. In addition to providing a method to capture binding events in the native cellular context, photocrosslinkers also have a fast reaction rate (Figure 1-9). This permits the capture of more transient, short-lasting PPIs (Figure 1-9).

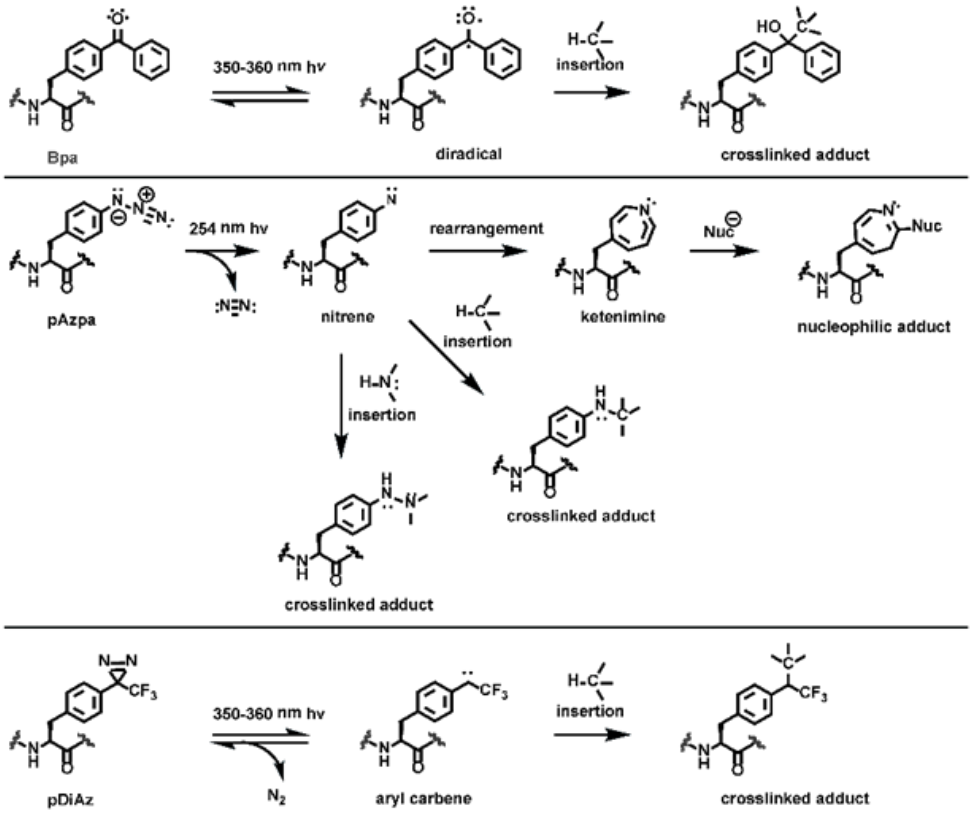


Figure 1-9 Activation of Popular Photocrosslinking Amino Acids.
Benzophenones (top) are a diradical reactive species with a 3.1Å reactivity radius and 120 μs reactive form until relaxation back to the inactivated ketone[4,5]. Azides (middle) lose N_2 and form a nitrene which is reactive for 0.1 ms before rearranging to the less reactive but more stable ketenimine that can only react with nucleophiles[4]58585858485. The aromatic diazirine (bottom) forms a highly reactive singlet carbene which has a lifetime of approximately 1 ns or a diazoisomer which ultimately converts to the carbene as well[9,10].

1E: Overview of Thesis

In the following chapters, it will be demonstrated how the use of genetically incorporated unnatural amino acids has expanded our ability to study difficult PPIs. In

particular, examination of transcriptional activator-coactivator complexes in their native environment has been successfully demonstrated through the use of photolabile UAAs to covalently capture protein binding partners in live cells. Chapters 2-3 contain a discussion of covalent capture methods combined with mass spectrometry to discover novel PPIs as well as show new ways to visualize direct PPIs on DNA. Chapter 4 further explores a caveat to covalent capture, the lack of *in vivo* characterization of crosslinking moieties. The window of capture for each moiety was determined across a various PPI binding affinities and surface areas. The methods developed in this thesis provide a sound demonstration of the power of using covalent chemical capture for the study of PPIs in a cellular context. This work is significant to the field of biochemistry, as it adds to the framework indicating that difficult PPIs can be studied and into the future, understood. Retaining the native environment of these PPIs not only allows much easier study but also can indicate their importance to cellular functioning as many of these PPIs are part of critical cell processes that would allow researchers to understand diseases and potentially find better drug targets. In addition, I have discovered novel binding partners of our yeast model activator, Gal4, which broadens the perspective on transcriptional activator participation in dynamic complexes implicated in the transcription initiation process.

References Chapter 1:

1. Thakur, J. K., Yadav, A., and Yadav, G. (2014) Molecular recognition by the KIX domain and its role in gene regulation, *Nucleic Acids Research 42*, 2112-2125.
2. Hargreaves, D. C., and Crabtree, G. R. (2011) ATP-dependent chromatin remodeling: genetics, genomics and mechanisms, *Cell Res 21*, 396-420.
3. Kar, A., and Gutierrez-Hartmann, A. (2013) Molecular mechanisms of ETS transcription factor-mediated tumorigenesis, *Critical Reviews in Biochemistry and Molecular Biology 48*, 522-543.
4. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008) Photocrosslinkers illuminate interactions in living cells, *Molecular BioSystems 4*, 473-480.
5. Dorman, G., and D., P. G. (1994) Benzophenone photophores in biochemistry, *Biochemistry 33*, 5661-5573.
6. Sammak, S., and Zinzalla, G. (2015) Targeting protein-protein interactions (PPIs) of transcription factors: Challenges of intrinsically disordered proteins (IDPs) and regions (IDRs), *Prog Biophys Mol Biol 119*, 41-46.

7. Lee, T. I., and Young, R. A. (2013) Transcriptional regulation and its misregulation in disease, *Cell 152*, 1237-1251.
8. Thompson, A. D., Dugan, A., Gestwicki, J. E., and Mapp, A. K. (2012) Fine-tuning multiprotein complexes using small molecules, *ACS Chem Biol 7*, 1311-1320.
9. Geurink, P. P., Prely, L. M., van der Marel, G. A., Bischoff, R., and Overkleeft, H. S. (2012) Photoaffinity Labeling in Activity-Based Protein Profiling, In *Activity-Based Protein Profiling* (Sieber, A. S., Ed.), pp 85-113, Springer Berlin Heidelberg, Berlin, Heidelberg.
10. Cesa, L. C., Mapp, A. K., and Gestwicki, J. E. (2015) Direct and Propagated Effects of Small Molecules on Protein-Protein Interaction Networks, *Front Bioeng Biotechnol 3*, 119.
11. Liu, J. P., N.B.; Oldfield, C.J.; Su, E.W.; Uversky, V.N.; Dunker, A.K. (2006) Intrinsic Disorder in Transcription Factors, *Biochem. 45*, 6873-6888.
12. Prescher, J. A., and Bertozzi, C. R. (2005) Chemistry in living systems, *Nat Chem Biol 1*, 13-21.
13. Pattabiraman, D. R., McGirr, C., Shakhbazov, K., Barbier, V., Krishnan, K., Mukhopadhyay, P., Hawthorne, P., Trezise, A., Ding, J., Grimmond, S. M., Papathanasiou, P., Alexander, W. S., Perkins, A. C., Levesque, J.-P., Winkler, I. G., and Gonda, T. J. (2014) Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes, *Blood 123*, 2682-2690.
14. Khoo, K. H., Verma, C. S., and Lane, D. P. (2014) Drugging the p53 pathway: understanding the route to clinical efficacy, *Nat Rev Drug Discov 13*, 217-236.
15. Laraia, L., McKenzie, G., Spring, D. R., Venkitaraman, A. R., and Huggins, D. J. (2015) Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions, *Chem Biol 22*, 689-703.
16. Pilauri, V., Bewley, M., Diep, C., and Hopper, J. (2005) Gal80 Dimerization and the Yeast GAL Gene Switch, *Genetics 169*, 1903-1914.
17. Thoden, J. B., Ryan, L. A., Reece, R. J., and Holden, H. M. (2008) The interaction between an acidic transcriptional activator and its inhibitor. The molecular basis of Gal4p recognition by Gal80p, *J Biol Chem 283*, 30266-30272.
18. Salmeron, J. M., Leuther, K. K., and Johnston, S. A. (1990) GAL4 mutations that separate the transcriptional activation and GAL80-interactive functions of the yeast GAL4 protein, *Genetics 125*, 21-27.
19. Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., and Orengo, C. (2010) Transient protein-protein interactions: structural, functional, and network properties, *Structure 18*, 1233-1243.
20. Lee, T. I. Y., R.A. (2000) Transcription of Eukaryotic Protein-Coding Genes, *Annu. Rev. Genet. 34*, 77-137.
21. Elio, C., Mikko, K., and Wing-Yiu, C. (2014) Long Molecular Dynamics Simulations of Intrinsically Disordered Proteins Reveal Preformed Structural Elements for Target Binding, In *Computational Approaches to Protein Dynamics*, pp 233-256, CRC Press.

22. Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C., and Asturias, F. J. (2008) Malleable machines take shape in eukaryotic transcriptional regulation, *Nat Chem Biol 4*, 728-737.
23. Scholes, N. S., and Weinzierl, R. O. (2016) Molecular Dynamics of "Fuzzy" Transcriptional Activator-Coactivator Interactions, *PLoS Comput Biol 12*, e1004935.
24. Thomas, M. C., and Chiang, C. M. (2006) The general transcription machinery and general cofactors, *Crit Rev Biochem Mol Biol 41*, 105-178.
25. Weake, V. M., and Workman, J. L. (2010) Inducible gene expression: diverse regulatory mechanisms, *Nat Rev Genet 11*, 426-437.
26. Mapp, A. K., Pricer, R., and Sturlis, S. (2015) Targeting transcription is no longer a quixotic quest, *Nat Chem Biol 11*, 891-894.
27. Wang, N., Majmudar, C. Y., Pomerantz, W. C., Gagnon, J. K., Sadowsky, J. D., Meagher, J. L., Johnson, T. K., Stuckey, J. A., Brooks, C. L., Wells, J. A., and Mapp, A. K. (2013) Ordering a Dynamic Protein Via a Small-Molecule Stabilizer, *Journal of the American Chemical Society 135*, 3363-3366.
28. Hahn, S. (2004) Structure and mechanism of the RNA Polymerase II transcription machinery, *Nat Struct Mol Biol. 11*, 394-403.
29. Huang, Y., and Maraia, R. J. (2001) Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human, *Nucleic Acids Research 29*, 2675-2690.
30. Ptashne, M. G., A. (2002) Yeast: A Single-celled Eukaryote, In *Genes & Signals*, Cold Spring Harbor Laboratory Press.
31. Stone, G. S., I. (1993) Gal4 is regulated by a glucose-responsive functional domain, *EMBO 12*, 1375-1385.
32. Klar, A., and Halvorson, H. (1974) Studies on the positive regulatory gene gal4 and regulation of galactose catabolic enzymes in Saccharomyces cerevisiae, *Mol. Gen. Genet. 135*, 203-212
33. Brent, R., and Ptashne, M. (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor, *Cell 43*, 729-736.
34. Melcher, K., and Johnston, S. A. (1995) GAL4 interacts with TATA-binding protein and coactivators, *Molecular Cellular Biology 15*, 2839-2848.
35. Johnston, S. A. Z., M.J.; Debouck, C.; Hopper, J.E. (1986) Functional domains of the yeast regulatory protein GAL4, *Proc. Natl. Acad. Sci. 83*, 6553-6557.
36. Ma, J., and Ptashne, M. (1987) The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell 50*, 137-142.
37. Flick, J. S., and Johnston, M. (1990) Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae, *Molecular Cell Biology 10*, 4757-4769.
38. Griggs, D. W., and Johnston, M. (1991) Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression, *Proceedings of the National Academy of Sciences 88*, 8597-8601.
39. Parthun, M. R. J., J.A. (1992) A Transcriptionally Active Form of GAL4 Is Phosphorylated and Associated with GAL80, *Mol. Cell. Biol.*, 4981-4987.
40. Lamphier, M. S., and Ptashne, M. (1992) Multiple mechanisms mediate glucose repression of the yeast GAL1 gene, *Proceedings of the National Academy of Sciences of the United States of America 89*, 5922-5926.

41. Lohr, D. V., P.; Zlantanova, J. (1995) Transcriptional regulation in yeast GAL gene family: a complex genetic network, *FASEB 9*, 777-787.
42. Zenke, F. T. E., R.; Vollenbroich, V.; Meyer, J.; Hollenberg, K.D. (1996) Activation of Gal4p by Galactose-Dependent Interaction of Galactokinase and Gal80p, *Science 272*, 1662-1665.
43. Platt, A., and Reece, R. J. (1998) The yeast galactose genetic switch is mediated by the formation of a Gal4p–Gal80p–Gal3p complex, *The EMBO Journal 17*, 4086-4091.
44. Melcher, K., and Xu, H. (2001) Gal80–Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression, *The EMBO Journal 20*, 841-851.
45. Bryant, G. O., and Ptashne, M. (2003) Independent Recruitment In Vivo by Gal4 of Two Complexes Required for Transcription, *Cell 11*, 1301-1309.
46. Traven, A., Jelicic, B., and Sopta, M. (2006) Yeast Gal4: a transcriptional paradigm revisited, *EMBO Rep 7*, 496-499.
47. Thoden, J. B., Sellick, C. A., Reece, R. J., and Holden, H. M. (2007) Understanding a Transcriptional Paradigm at the Molecular Level: The Structure of Yeast Gal80p, *Journal of Biological Chemistry 282*, 1534-1538.
48. Egriboza, O. G., S.; Taod, X.; Dottsb, K.; Schaefferb, C.; Pilaurie, V.; Hopperb, J.E. (2013) Self-Association of the Gal4 Inhibitor Protein Gal80 Is Impaired by Gal3: Evidence for a New Mechanism in the GAL Gene Switch, *Mol. Cell. Biol. 33*, 3667-3674.
49. Mohibullah, N., and Hahn, S. (2008) Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3, *Genes Dev 22*, 2994-3006.
50. Dudley, A. M. R., C.; Winston, F. (1999) The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step in vivo, *Genes Dev. 13*, 2940-2945.
51. Larschan, E., and Winston, F. (2001) The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4, *Genes & Development 15*, 1946-1956.
52. Bhaumik, S. R., and Green, M. R. (2001) SAGA is an essential in vivo target of the yeast acidic activator Gal4p, *Gene Development 15*, 1935-1945.
53. Klein, J., Nolden, M., Sanders, S. L., Kirchner, J., Weil, P. A., and Melcher, K. (2003) Use of a genetically introduced cross-linker to identify interaction sites of acidic activators within native transcription factor IID and SAGA, *J Biol Chem 278*, 6779-6786.
54. Stringer, K. F., Ingles, C. J., and Greenblatt, J. (1990) Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID, *Nature 345*, 783-786.
55. Nedialkov, Y. A., and Triezenberg, S. J. (2004) Quantitative assessment of in vitro interactions implicates TATA-binding protein as a target of the VP16C transcriptional activation region, *Arch Biochem Biophys 425*, 77-86.
56. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) Transcriptional Activation Domain of the Herpesvirus Protein VP16 Becomes

Conformationally Constrained upon Interaction with Basal Transcription Factors, *Journal of Biological Chemistry 271*, 4827-4837.
57. Schmitt, J., Hess, H., and Stunnenberg, H. G. (1993) Affinity purification of histidine-tagged proteins, *Molecular Biology Reports 18*, 223-230.
58. Sambrook, J. R., D.W. (2001) *Molecular Cloning: A Laboratory Manual*, Spring Harbor Laboratory Press.
59. Fields, S., and Song, O.-k. (1989) A novel genetic system to detect protein-protein interactions, *Nature 340*, 245-246.
60. Kim, S. B., Otani, Y., Umezawa, Y., and Tao, H. (2007) Bioluminescent Indicator for Determining Protein−Protein Interactions Using Intramolecular Complementation of Split Click Beetle Luciferase, *Analytical Chemistry 79*, 4820-4826.
61. Huang, Y.-m., and Bystroff, C. (2009) Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein, *Biochemistry 48*, 929-940.
62. Piston, D. W., and Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly, *Trends in Biochemical Sciences 32*, 407-414.
63. Roy, R., Hohng, S., and Ha, T. (2008) A practical guide to single-molecule FRET, *Nat Meth 5*, 507-516.
64. Kustatscher, G., Wills, K. L. H., Furlan, C., and Rappsilber, J. (2014) Chromatin enrichment for proteomics, *Nat. Protocols 9*, 2090-2099.
65. Das, P. M., Ramachandran, K., vanWert, J., and Singal, R. (2004) Chromatin immunoprecipitation assay, *Biotechniques 37*, 961-969.
66. Park, P. J. (2009) ChIP-seq: advantages and challenges of a maturing technology, *Nat Rev Genet 10*, 669-680.
67. Whitington, T., Frith, M. C., Johnson, J., and Bailey, T. L. (2011) Inferring transcription factor complexes from ChIP-seq data, *Nucleic Acids Research 39*, e98.
68. Metz, B., Kersten, G. F. A., Baart, G. J. E., de Jong, A., Meiring, H., ten Hove, J., van Steenbergen, M. J., Hennink, W. E., Crommelin, D. J. A., and Jiskoot, W. (2006) Identification of Formaldehyde-Induced Modifications in Proteins: Reactions with Insulin, *Bioconjugate Chemistry 17*, 815-822.
69. Kunkel, G. R. M., M.; Martinson, H.G. (1981) Contact-site cross-linking agents, *Mol. Cell. Biochem. 34*, 3-13.
70. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions, *Mass Spectrometry Reviews 25*, 663-682.
71. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review, *Analytical and Bioanalytical Chemistry 389*, 1017-1031.
72. Noren, C. J. A.-C., S.; Griffith, M.C.; Schultz, P.G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins, *Science 244*, 182-190.
73. Xie, J., and Schultz, P. G. (2005) An expanding genetic code, *Methods 36*, 227-238.
74. Wang, L., and Schultz, P. G. (2002) Expanding the genetic code, *Chemistry Communications*, 1-11.

Chapter 2: TRIC - Capturing the Direct Cellular Targets of Promoter-Bound Transcriptional Activators[2]

2A: Abstract

Transcriptional activators coordinate the dynamic assembly of multi-protein coactivator complexes required for gene expression to occur. Here we combine the power of *in vivo* covalent chemical capture with *p*-benzoyl-L-phenylalanine (Bpa), a genetically incorporated photocrosslinking amino acid, and chromatin immunoprecipitation (ChIP) to capture the direct protein interactions of the transcriptional activator VP16 with the general transcription factor TBP at the GAL1 promoter in live yeast.

2B: Background

Transcriptional activators are essential in directing the assembly of the RNA polymerase transcriptional machine at individual promoters within the genome. Though it is well understood that this occurs through the formation of protein-protein interactions (PPIs) between DNA-bound activators and individual components of the transcriptional machinery, resolving the network of activator PPIs that underpin this process has been technically challenging[1]. For example, *in vivo* co-localization and chromatin immunoprecipitation (ChIP) studies have identified the complexes that are recruited by activators to the promoter, but distinguishing the individual subunits within these complexes that serve as the activator binding partner(s) *in vivo* has not been possible. Toward this goal, *in vivo* covalent chemical capture with genetically incorporated photocrosslinking amino acids has emerged as an important tool in the identification of PPIs in cells[2-5]. Here we describe an expansion of this strategy that utilizes the rapid, reversible formaldehyde crosslinking of ChIP together with the site-specific, irreversible crosslinking of the non-natural amino acid p-benzoyl-L-phenylalanine (Bpa) to investigate the direct, *in vivo* PPIs of DNA-bound activators. We demonstrate that this

[2] Portions of this chapter are taken from: Dugan, A.; Pricer, R.; Katz, M.; Mapp, A.K., TRIC: Capturing the direct cellular targets of promoter-bound transcriptional activators. *Protein Sci*. **2016**, 25, *8*, 1371-1377. My contributions to this work include initial optimization experiments contributing to the TRIC method, data in Figure 2-1D, Figure 2-2, and assisting in writing.

tandem reversible and irreversible crosslinking (TRIC) approach is able to capture the interaction of the viral activator VP16 and the general transcription factor TBP at the GAL1 promoter in live yeast.

The identities and compositions of the multi-protein complexes that assemble at genes in order to initiate gene expression can vary significantly from promoter to promoter[6-9]. Due to a lack of methods appropriate for determining the individual contacts made by activators at a promoter, these PPIs remain poorly defined. As an example, the recruitment of TATA-binding protein (TBP) by the canonical transcriptional activator VP16 has been studied extensively, yet the mechanism by which TBP arrives at promoters is still contested, even in the case of well-studied promoter contexts such as GAL1.[10-15] A significant body of *in vitro* data exists to support a direct interaction between VP16 and TBP[16-19]. However, *in vivo* ChIP studies suggest that TBP recruitment to GAL1 is mediated via interactions with the SAGA complex, supporting a model in which the activator is not directly involved in TBP recruitment[20-23].

2C: Covalent Chemical Capture of TBP-VP16 Interaction

To distinguish between the two mechanisms of TBP recruitment, *in vivo* covalent chemical capture was first used to determine if VP16 directly contacts TBP in living yeast. In this approach, the photolabile unnatural amino acid Bpa is genetically incorporated into the transcriptional activation domain (TAD) of VP16 using an engineered nonsense suppression system[24, 25]. Upon irradiation of live yeast with UV light, Bpa is activated and forms an irreversible covalent bond with any protein partners that are in direct contact with VP16 at the site of incorporation. The covalently captured binding partners can then be isolated and analyzed to determine their identities. As previously described, we used a chimeric activator in which the amino terminal half of the extended VP16 activation domain is fused to the LexA DNA binding motif[2]. We chose to incorporate Bpa at position 444 in the VP16N (413-456) subdomain, as this residue has been shown to be involved in maintaining activator interactions and, consistent with this, our previous work showed a robust and repeatable multi-protein crosslinking profile at this position. Furthermore, work from our group has demonstrated that incorporation at this site does not impair activator function[2]. LexA+VP16 L444Bpa was expressed in yeast and the cells were irradiated with UV light to activate Bpa and

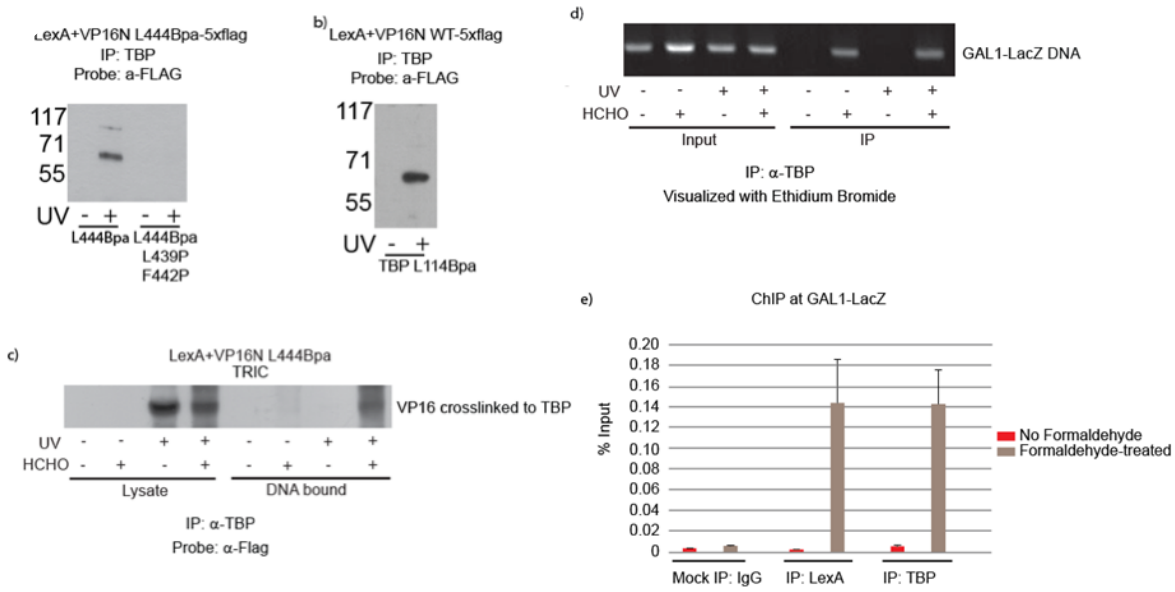


Figure 2-1 VP16 Directly Contacts TBP in Yeast.
The VP16-TBP interaction was captured *in vivo* using covalent chemical capture in yeast expressing (a) LexA+VP16N L444Bpa and (b) LexA+VP16N WT and Myc-TBP L114Bpa. Yeast were irradiated with UV light and the lysates immunoprecipitated with a TBP antibody. The covalent VP16-TBP products were visualized on Western blot with an α-Flag HRP antibody. (c,d) Tandem reversible and irreversible crosslinking captures the direct targets of DNA bound transcriptional activators. Identical cultures of yeast expressing LexA+VP16N L444Bpa were either crosslinked with UV, formaldehyde, or treated with a combination of both procedures (formaldehyde followed by UV crosslinking). The chromatin fractions of these cultures were washed to remove non-covalently bound protein and then the chromatin was sheared and solubilized using sonication. Soluble chromatin was then immunoprecipitated with an α-TBP antibody and the formaldehyde crosslinks were reversed. Western blots (c) were probed with an α-Flag antibody and immunoprecipitated DNA was amplified with GAL1-LacZ specific primers and visualized on an agarose gel stained with ethidium bromide (d). (e) Chromatin immunoprecipitation at the GAL1 promoter in yeast. Bars are the mean of at least three biological replicates. Error bars indicate the standard error of the mean.

covalently capture any proteins directly contacting VP16. Yeast were then lysed and the lysates immunoprecipitated with an antibody showed a robust and repeatable multi-protein crosslinking profile at this position. Furthermore, work from our group has demonstrated that incorporation at this site does not impair activator function[2].

LexA+VP16 L444Bpa was expressed in yeast and the cells were irradiated with UV light to activate Bpa and covalently capture any proteins directly contacting VP16. Yeast were then lysed and the lysates immunoprecipitated with an antibody against TBP. After washing away any non-covalently bound proteins, the immunoprecipitated TBP was run on SDS-PAGE and the resulting Western blot was probed with a Flag-HRP antibody to detect the presence of a covalently bound VP16-TBP species. Indeed, a crosslinked band is observed at approximately 60 kDa that corresponds to the additive molecular weight of the LexA+VP16 construct and TBP (Figure 2-1A). This data indicates that VP16 directly

contacts TBP in live yeast. Consistent with previous reports, introduction of a double point mutation in the VP16 activation domain (L439P and F442P) that abolishes activity also abrogates crosslinking to TBP, indicating that this interaction is specific (Figure 2-1A)[26, 27].

To further substantiate this interaction, covalent chemical capture in which TBP is the captor was carried out. Biochemical data suggest that residues along the concave face of TBP are involved in dynamic exchange with the other transcriptional proteins, including activators, which compete for binding at this site[28-30]. We therefore co-expressed LexA+VP16 WT alongside a Myc-TBP construct in which the leucine at position 114 was mutated to Bpa (L114Bpa) and carried out covalent chemical capture experiments. We found that incorporation of Bpa at this position in TBP resulted in crosslinking TBP-VP16 adduct (Figure 2-1B). Thus in support of previous biochemical data, our results indicate that this site on TBP is indeed a contact point for VP16 *in vivo*[30].

2D: TRIC-Tandem Reversible and Irreversible Crosslinking in Yeast

While demonstrating a direct interaction between VP16 and TBP, the above experiments do not report on any one context. In other words, the crosslinked species likely originate from a variety of different promoter-localized and non-DNA bound VP16 species. We next chose to design a straightforward method that would allow for the direct examination of the VP16-TBP interaction at the GAL1 promoter present in the yeast strain. The first step of this strategy borrows principles from classic chromatin immunoprecipitation (ChIP) protocols in which formaldehyde is first administered to cells to rapidly stabilize dynamic protein-DNA interactions. The fixed cells are then subjected to our covalent chemical capture workflow in order to activate Bpa and site-specifically crosslink the direct targets of our Bpa-containing activators within the immobilized transcription complexes. The advantage of this tandem crosslinking approach is rooted in the nature of the covalent bonds that are formed by each crosslinking moiety; more specifically, the formaldehyde crosslinks can be readily reversed in an appropriately nucleophilic buffer, whereas the Bpa crosslinks are irreversible under such treatment conditions. Thus, after tandem crosslinking treatment, we could lyse the yeast, isolate the insoluble chromatin-containing pellet, and wash the pellet extensively to remove any non-DNA bound complexes. Similar to ChIP, the DNA

could then be solubilized via sonication and immunoprecipitation carried out against the protein of interest, in this case TBP. The formaldehyde crosslinks on the isolated complexes are then reversed, leaving only the irreversible photocrosslinked activator-coactivator interactions intact and able to be visualized via Western blotting (Figure 2-2).

When we executed our tandem reversible and irreversible crosslinking (TRIC) approach, we found that only under conditions where both formaldehyde and UV

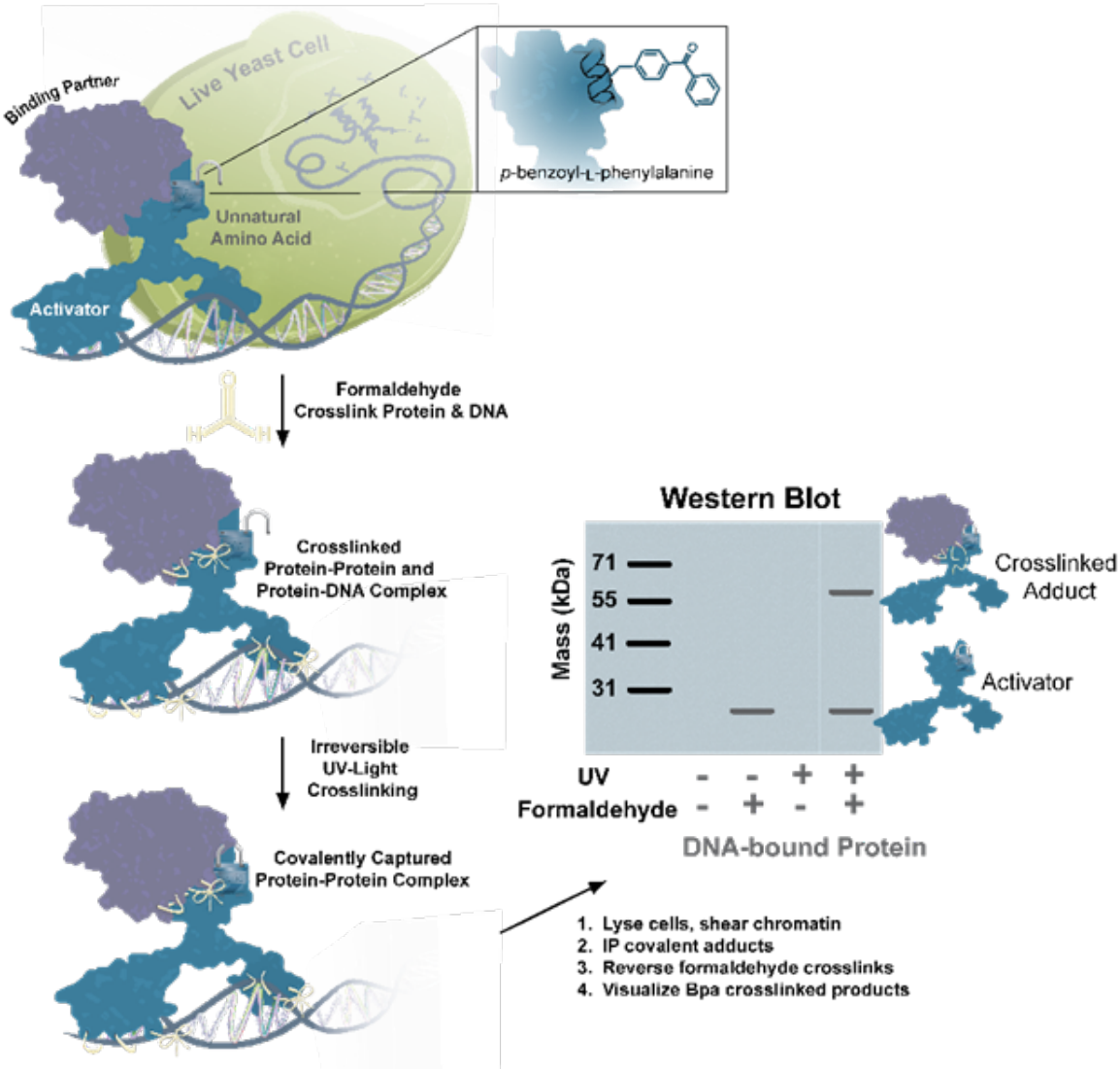


Figure 2-2 Tandem Reversible and Irreversible Crosslinking (TRIC) Allows for the Covalent Capture of Direct Targets of Transcriptional Activators at the Promoter.
Yeast cells are treated with formaldehyde to stabilize protein-DNA and protein-protein interactions at the promoter. Treated cells are then gently irradiated with UV light to activate a genetically encoded photocrosslinking amino acid in the transcriptional activation domain (TAD) of the activator coordinating complex assembly. The cells are lysed and the chromatin isolated and sheared via sonication. Following immunoprecipitation of the immobilized complexes, the formaldehyde crosslinks are reversed and the irreversibly covalently linked PPIs are identified by Western blotting.

treatment were used were we able to observe a VP16-TBP crosslink in the DNA fraction, whereas UV alone was sufficient to capture this interaction in yeast lysate (Figure 2-1C). When this protocol was repeated to examine the DNA that co-purified during immunoprecipitation, we observed that the GAL1-LacZ gene could only be visualized under conditions in which formaldehyde was present. This indicates that formaldehyde stabilization of promoter-bound proteins is required to observe the interactions of DNA-bound activators (Figure 2-1D). Finally, we executed a straightforward ChIP protocol to show that both LexA+VP16 and TBP co-localize to our reporter gene and confirm that we are indeed observing a direct protein-protein interaction at this specific location in the yeast genome (Figure 2-1E).

We note that although the protocol is a hybrid of two powerful approaches, optimization efforts in several areas were required in order for this approach to work. The first point of optimization was the duration of formaldehyde crosslinking. We were unable to observe a VP16-TBP interaction using previously reported one-minute rapid mixing protocols.[31] We therefore tested several formaldehyde treatment times and found that five-minute treatment with formaldehyde was the shortest interval that could be used to achieve consistent results. As has been noted elsewhere, formaldehyde crosslinking optimization will have to be examined on an individual protein basis in order to achieve sufficient crosslinking results.[32, 33] Another key point of optimization for this study involved the extent of yeast lysis prior to sonication of the insoluble chromatin-containing fraction. Yeast have a tough cell wall that is notoriously difficult to break open. We found that under conditions of incomplete lysis, any remaining cells that pelleted with the insoluble fraction would release their cellular contents upon sonication, resulting in contaminating bands in the UV treated lanes of the Western. We thus screened a number of commercially *Ava*ilable chemical lysis reagents and mechanical lysis conditions to determine a method that would provide us with the most complete yeast lysis. We evaluated the efficacy of lysis under each condition by measuring the total concentration of protein released in the lysate as well as by visually assessing the extent of lysis by monitoring cells under a microscope. We found that only extended mechanical disruption with glass beads consistently achieved >95% cell lysis and yielded maximal protein release in the cell lysate.

2E: Conclusion

In conclusion, we have described a new *in vivo* tandem crosslinking approach that is useful in capturing the PPIs of DNA bound activators, as demonstrated by the capture of VP16 and TBP at the GAL1 promoter. The advantage of this strategy over previously described dual crosslinking applications (such as those using formaldehyde and disuccinimidyl suberate, for example) is that the use of a site-specific photocrosslinking amino acid allows one to distinguish the direct targets of a given protein within a DNA-bound multiprotein complex. This allows a level of resolution that, to the best of our knowledge, has not been achieved prior to this work.

TBP is an essential transcription factor required for gene expression in yeast, yet, as is the case with other requisite transcriptional complexes, the mechanism by which it is recruited to promoters has remained elusive, even in the most well-studied promoter contexts. Using TRIC, we were able to stabilize transcription factor-DNA contacts and then covalently capture proteins that were in direct contact with the Bpa-containing activator VP16 in live yeast. This work suggests that VP16 directly contacts TBP in yeast to recruit the general transcription factor to the GAL1 promoter. While our results thus suggest a direct mechanism of recruitment to GAL1, earlier observations that SAGA is required for TBP localization to GAL1 likely point to a role for SAGA in stabilizing TBP residency at this promoter.[20-23] This work is an important first step toward resolving the PPI map of activators at a single promoter, work that will be useful in building a more complete picture of activator interactions as a whole. Future work will be focused on combining TRIC with DNA microarray technologies to facilitate the identification of the direct interactions of a single activator at individual promoters across the genome.

2F: Experimental

All covalent chemical capture and TRIC experiments were carried out in yeast strain LS41 [JPY9::pZZ41, Mat his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4 URA::pZZ41]. Bpa was purchased from Chem-Impex International (Wood Dale, IL). All plasmids used in this study were constructed using standard molecular biology techniques. Sanger sequencing verifying plasmid sequences was performed by the University of Michigan Core Facility (Ann Arbor, MI).

Construction of plasmids

pLexA+VP16N WT

A high copy plasmid expressing LexA(1-202)+VP16N (413-456)+5x FLAG tag under the control of the ADH1 promoter was created via ligation of the fusion gene into a pCLexA-5xFlag backbone containing BamHI and SalI sites. Primers 5'-catgaattcATGGCCCCCCCGACCGATGTC-3' and 5'-catGTCGACTTACTTGTCATCGTCGTCCTTGTAGTCTCCCGGCCCCGGGGAATCCC-3' were used to amplify VP16 (413-456) from a pCLexA-VP16-1xFlag template. The amplified PCR product was digested with SalI and BamHI and inserted into pCLexA digested with SalI and BamHI and treated with calf intestinal phosphate to create pLexA+VP16N WT-5X Flag.

pLexAVP16N 444TAG

To create each plasmid, site-directed mutagenesis was used to replace an existing amino acid codon with TAG codon within the VP16C or VP16N TAD. In general, PCR primers were designed to have ~15 bases of homology on either side of the TAG mutation. QuikChange protocol was used to incorporate the TAG mutants using manufacturer recommended conditions.

pGADT7 Myc-TBP

A high copy plasmid pGADT7 expressing an N-terminally Myc-tagged TBP was constructed by amplifying the DNA sequence encoding TBP from yeast genomic DNA using primers (5'- catCATATGATGGCCGATGAGGAACGTTTAAAGG-3') and (5'-atgCTCGAGTCACATTTTTCTAAATTCACTTAGC -3'). The purified PCR product was ligated into a Myc-pGADT7 vector digested with NdeI and XhoI using standard molecular biology techniques. The Myc-pGADT7 cloning vector was created by inserting a c-Myc epitope tag in pGADT7 (Clontech) using site-directed mutagenesis with the following primers: (5'-AGCTATGGAACAAAAGTTGATTTCTGAAGAAGATTTGGGATCCAATGCATATGATCT -3') and (5'-AGCTTGATCATATGCATTGGATCCCAAATCTTCTTCAGAAATCAACTTTTGTTCCAT- 3'). Leu114 in TBP was mutagenized to an amber stop codon using site-directed mutagenesis (Qiagen Quikchange Protocol).

***In vivo* covalent chemical capture**

In vivo covalent chemical capture experiments for pLexA+VP16N L444Bpa was carried out as previously described.[25] For crosslinking studies with Myc-TBP L114Bpa, the procedure was identical except that cells were grown in SC media lacking histidine, leucine, and tryptophan. For lysis, cells were resuspended in 600 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche) and lysed using glass beads by vortexing at 4°C. Subsequently, the lysate was pelleted and the supernatant incubated with 2 µg of LexA antibody (sc-1725, Santa Cruz Biotechnologies) for 2 hours at 4°C for immunoprecipitation. The protein bound to the antibody was isolated by incubation for 1 hour with 50 µL of prewashed protein G magnetic beads slurry (Dynal Corporation, Invitrogen, Carlsbad, CA) and the beads were washed 6 times with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -80˚C until elution. The crosslinked sample was eluted from the beads by heating at 95˚C for 10 minutes in NuPAGE 4x LDS Sample buffer (Invitrogen, Carlsbad, CA) containing 250 mM DTT and probed using Western Blot analysis using anti-FLAG (M2) antibody (Sigma, St. Louis, MO).

Tandem reversible and irreversible crosslinking

To perform TRIC, 100 mL cultures of yeast were grown in SC media containing 2% raffinose, 2% galactose, 1 mL of 100 mM Bpa dissolved in 1M NaOH, and 1 mL 1M HCl. The cultures were incubated overnight at 30°C with agitation and grown to mid-log phase (OD660 ~1.0). Cultures receiving UV treatment only were spun down by centrifuging at 3901 rcf at 4˚C for 5 minutes following which the cell pellets were washed with SC media lacking histidine and tryptophan. These cell pellets were resuspended in 2mL SC media lacking histidine and tryptophan + 2% raffinose, 2% galactose and transferred to small cell culture dishes and subjected to UV irradiation at 365 nm light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour.

For cultures receiving only formaldehyde treatment, 3 mL of 37% formaldehyde solution was added directly to the culture and allowed to remain in the incubator for 20 minutes. Repeat of this procedure with a 5-minute formaldehyde incubation yielded the same results as a 20-minute incubation. Cultures were then quenched with 15 mL of 2M

Glycine. Cells were then centrifuged and washed with 50 mL SC media lacking histidine and tryptophan. Samples intended to additionally receive UV crosslinking were resuspended in 2 mL SC media (His-,Trp-) containing 2% raffinose and 2% galactose and transferred to a small cell culture dish and subjected to UV irradiation at 365 nm UV light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour.

For lysis, cells were resuspended in 600 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche) and lysed using glass beads mechanical disruption at 4°C until >95% lysis was observed. We found in these studies that complete cellular lysis is necessary to eliminate background signal caused by cell lysis during sonication. Subsequent lysates were immunoprecipitated with 8 µL TBP antibody (Santa Cruz, sc-33736) and incubated for 2 hours at 4°C. The remaining pellet was then washed 4x with "Harsh" ChIP buffer (50 mM HEPES-KOH pH 7.5, 1M NaCl, 1 mM EDTA, 1% Triton X-100, 1% Na-Deoxycholate) followed by 2 washes with standard ChIP buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche)). Pellets were resuspended in 600 µL standard ChIP buffer containing protease inhibitor and sonicated at a setting of 10% for 2 minutes with 30 second pulses on/off (double-step microtip, Fisher Scientific Dismembrator Model 500). Samples were then centrifuged at 4°C for 20 minutes at 14,000 rpm (Eppendorf 5417C). Soluble chromatin (supernatant) was immunoprecipitated with TBP antibody (Santa Cruz, sc-33736) for 2 hours, 4°C. The protein bound to the antibody was isolated by incubation for 1 hour with 50 µL of prewashed protein G magnetic beads (Dynal Corporation, Invitrogen, Carlsbad, CA). After immunoprecipitation, the beads were washed 6 times with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -80˚C until elution. The crosslinked sample was eluted from the beads and formaldehyde crosslinks reversed by heating at 95˚C for 20 minutes in SUTEB buffer containing 250 mM DTT and probed using Western Blot analysis using anti-FLAG (M2) antibody (Sigma, St. Louis, MO). For studies examining the DNA immunoprecipitated during TRIC, the TRIC protocol was

followed with the exception of 50 µL of solubilized chromatin being saved prior to immunoprecipitation. Additionally, lysates were discarded in these experiments.

To examine the size of the sheared chromatin, 50 µL TE/SDS was added to the Input samples and incubated overnight at 65°C to reverse crosslinks. 2.5 µL proteinase K (20 mg/mL stock) was then added and incubated at 50°C for 3 hours to digest proteins, followed by a PCR cleanup. 0.5 µL RNAse A (1 mg/mL stock) was added and then incubated at 37°C for 30 minutes. Samples were visualized on 1% agarose gel stained with Ethidium bromide. Smears showed between 300-900 bps.

For PCR on TRIC samples, 90 µL TE/SDS was added to 50 µL input and incubated overnight at 65°C followed by PCR Cleanup and elution in 58 µL EB buffer (Qiagen). For immunoprecipitated samples, beads were washed 2x with lysis buffer, 1 time with 500 mM NaCl lysis buffer, 1 time with wash buffer, and two times with TE buffer (10 mM Tris-HCl, 1 mM EDTA, 0.01% SDS (5 g in 500 mL ex). 50 µL elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS) was then added to the beads and vortexed briefly before incubating at 65°C for 30 minutes, with vortexing every 5 minutes to resuspend the beads. Beads were centrifuged for 30 seconds at 3000 rpm and the eluent transferred to a new tube. 120 µL TE/SDS was added and the samples incubated overnight at 65°C. Samples were purified using a PCR cleanup kit (Qiagen) and eluted in 58 µL EB buffer. DNA was quantified and PCR reactions were set up with GAL1- LacZ specific primers (5' CCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGC 3' and 5'GGGCGATCGGTGCGGGCCTCTTCGC 3'). Products were visualized on a 1% agarose gel stained with ethidium bromide.

Chromatin Immunoprecipitation at GAL1-LacZ

For chromatin immunoprecipitation, cultures were grown and formaldehyde crosslinked as was done for TRIC experiments. After mechanical glass bead shearing, the lysate and insoluble pellet were resuspended by gentle pipetting. Samples were sonicated and centrifuged as described earlier. Soluble chromatin was separated from the pellet and 10 µL was saved as an input sample. The remaining soluble chromatin was split equally between three 1.75 mL tubes which were then immunoprecipitated with either 2 µg Snf1 antibody (sc-15621, Santa Cruz Biotechnologies), LexA antibody (sc-

1725, Santa Cruz Biotechnologies), or control IgG (SC- 2027, Santa Cruz Biotechnologies). Immunoprecipitations were allowed to proceed for 2 hours at 4˚C followed by incubation for 1 hour with ~40 µL of prewashed protein G magnetic Dynabeads slurry (Life Technologies). Beads were washed in the same manner as TRIC DNA samples. Immunoprecipitated complexes were eluted in 50 µL elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS) at 65˚C for 30 minutes. The eluate was transferred to a new tube and 120 µL of TE-SDS (10 mM Tris-HCl, 1 mM EDTA, 0.1% SDS) buffer was added. To the saved inputs, 90 µL TE-SDS was added. Formaldehyde crosslinks were reversed overnight in a 65°C water bath. Samples were purified using Qiagen PCR Clean Up protocol and eluted in 58 µL Buffer EB. qPCR on all samples and inputs was run using Promega GoTaq qPCR master mix (A6001, Promega) using primers specific for GAL1-LacZ (sequences above). All qPCR runs were carried out on an Applied Biosystems StepOnePlus instrument. At least three independent biological replicates were run for each condition, with each biological replicate run in triplicate for qPCR quantitation. PCR amplification with the designed primers yielded a single band around 450 bps. This band was gel purified and submitted for sequencing. Results returned a sequence for the GAL1-LacZ gene, as expected.

Chapter 2: References

1. Mapp, A. K., and Ansari, A. Z. (2007) A TAD further: exogenous control of gene activation, *ACS Chemical Biology 2*, 62-75.
2. Krishnamurthy, M., Dugan, A., Nwokoye, A., Fung, Y. H., Lancia, J. K., Majmudar, C. Y., and Mapp, A. K. (2011) Caught in the act: covalent cross-linking captures activator-coactivator interactions in vivo, *ACS Chemical Biology 6*, 1321-1326.
3. Mori, H., and Ito, K. (2006) Different modes of SecY–SecA interactions revealed by site-directed in vivo photo-cross-linking, *Proceedings of the National Academy of Sciences 103*, 16159-16164.
4. Berg, M., Michalowski, A., Palzer, S., Rupp, S., and Sohn, K. (2014) An In Vivo Photo-Cross-Linking Approach Reveals a Homodimerization Domain of Aha1 in S. cerevisiae, *PloS one 9*, e89436.
5. Ting, S. Y., Schilke, B. A., Hayashi, M., and Craig, E. A. (2014) Architecture of the TIM23 inner mitochondrial translocon and interactions with the matrix import motor, *J Biol Chem 289*, 28689-28696.

6. Raab, J. R., Resnick, S., and Magnuson, T. (2016) Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes, *PLoS Genet 11*, e1005748.
7. Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003) The evolution of transcriptional regulation in eukaryotes, *Mol Biol Evol 20*, 1377-1419.
8. Cheng, J. X., Floer, M., Ononaji, P., Bryant, G., and Ptashne, M. (2002) Responses of Four Yeast Genes to Changes in the Transcriptional Machinery Are Determined by Their Promoters, *Current Biology 12*, 1828-1832.
9. Poss, Z. C., Ebmeier, C. C., and Taatjes, D. J. (2013) The Mediator complex and transcription regulation, *Critical Reviews in Biochemistry and Molecular Biology 48*, 575-608.
10. Flick, J. S., and Johnston, M. (1990) Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae, *Molecular Cell Biology 10*, 4757-4769.
11. Griggs, D. W., and Johnston, M. (1991) Regulated expression of the GAL4 activator gene in yeast provides a sensitive genetic switch for glucose repression, *Proceedings of the National Academy of Sciences 88*, 8597-8601.
12. Lamphier, M. S., and Ptashne, M. (1992) Multiple mechanisms mediate glucose repression of the yeast GAL1 gene, *Proceedings of the National Academy of Sciences of the United States of America 89*, 5922-5926.
13. Lemieux, K., and Gaudreau, L. (2004) Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II, *The EMBO journal 23*, 4040-4050.
14. Wands, A. M., Wang, N., Lum, J. K., Hsieh, J., Fierke, C. A., and Mapp, A. K. (2010) Transient-state kinetic analysis of transcriptional activator-DNA complexes interacting with a key coactivator, *Journal of Biological Chemistry 286*, 16238-16245.
15. Scholes, N. S., and Weinzierl, R. O. (2016) Molecular Dynamics of "Fuzzy" Transcriptional Activator-Coactivator Interactions, *PLoS Comput Biol 12*, e1004935.
16. Klein, J., Nolden, M., Sanders, S. L., Kirchner, J., Weil, P. A., and Melcher, K. (2003) Use of a genetically introduced cross-linker to identify interaction sites of acidic activators within native transcription factor IID and SAGA, *J Biol Chem 278*, 6779-6786.
17. Stringer, K. F., Ingles, C. J., and Greenblatt, J. (1990) Direct and selective binding of an acidic transcriptional activation domain to the TATA-box factor TFIID, *Nature 345*, 783-786.
18. Nedialkov, Y. A., and Triezenberg, S. J. (2004) Quantitative assessment of in vitro interactions implicates TATA-binding protein as a target of the VP16C transcriptional activation region, *Arch Biochem Biophys 425*, 77-86.
19. Shen, F., Triezenberg, S. J., Hensley, P., Porter, D., and Knutson, J. R. (1996) Transcriptional Activation Domain of the Herpesvirus Protein VP16 Becomes Conformationally Constrained upon Interaction with Basal Transcription Factors, *Journal of Biological Chemistry 271*, 4827-4837.

20. Mohibullah, N., and Hahn, S. (2008) Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3, *Genes Dev 22*, 2994-3006.
21. Dudley, A. M. R., C.; Winston, F. (1999) The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step in vivo, *Genes Dev. 13*, 2940-2945.
22. Larschan, E., and Winston, F. (2001) The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4, *Genes & Development 15*, 1946-1956.
23. Bhaumik, S. R., and Green, M. R. (2001) SAGA is an essential in vivo target of the yeast acidic activator Gal4p, *Gene Development 15*, 1935-1945.
24. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science 301*, 964-967.
25. Majmudar, C. Y., Lee, L. W., Lancia, J. K., Nwokoye, A., Wang, Q., Wands, A. M., Wang, L., and Mapp, A. K. (2009) Impact of nonnatural amino acid mutagenesis on the in vivo function and binding modes of a transcriptional activator, *Journal of the American Chemical Society 131*, 14240-14242.
26. Ingles, C. J., Shales, M., Cress, W. D., Triezenberg, S. J., and Greenblatt, J. (1991) Reduced binding of TFIID to transcriptionally compromised mutants of VP16, *Nature 351*, 588-590.
27. Regier, J., Shen, F., and Triezenberg, S. J. (1993) Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator, *Proceedings of the National Academy of Science USA 90*, 883-887.
28. Nishikawa, J. I., Kokubo, T., Horikoshi, M., Roeder, R. G., and Nakatani, Y. (1997) Drosophila TAFII230 and the transcriptional activator VP16 bind competitively to the TATA box-binding domain of TATA box-binding protein, *Proceedings of the National Academy of Science USA 94*, 85-90.
29. Mal, T. K., Takahata, S., Ki, S., Zheng, L., Kokubo, T., and Ikura, M. (2007) Functional silencing of TATA-binding protein (TBP) by a covalent linkage of the N-terminal domain of TBP-associated factor 1, *J Biol Chem 282*, 22228-22238.
30. Kim, T. K., Hashimoto, S., Kelleher, R. J. r., Flanagan, P. M., Kornberg, R. D., Horikoshi, M., and Roeder, R. G. (1994) Effects of activation-defective TBP mutations on transcription initiation in yeast, *Nature 369*, 252-255.
31. Bryant, G. O., and Ptashne, M. (2003) Independent Recruitment In Vivo by Gal4 of Two Complexes Required for Transcription, *Cell 11*, 1301-1309.
32. Das, P. M., Ramachandran, K., vanWert, J., and Singal, R. (2004) Chromatin immunoprecipitation assay, *Biotechniques 37*, 961-969.
33. Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation, *Trends in Biochemical Sciences 25*, 99-104.

Chapter 3: Discovery of Enzymatic Targets of Transcriptional Activators via *In vivo* Covalent Chemical Capture[3]

3A: Abstract

The network of activator protein-protein interactions (PPIs) that underpin transcription initiation is poorly defined, particularly in the cellular context. Through the coupling of *in vivo* covalent chemical capture and shotgun LC-MS/MS (MuDPIT) analysis, we discover that the prototypical activators Gal4 and VP16 target the Snf1 (AMPK) kinase complex via direct interactions with both the core enzymatic subunit Snf1 and the exchangeable subunit Gal83. Further, the Gal4-Snf1 interaction is localized to the GAL1 promoter and visualized through a tandem reversible formaldehyde and irreversible covalent chemical capture approach (TRIC). Together, these data support a critical role for activator PPIs in both the recruitment and positioning of important enzymatic complexes at a gene promoter and represents the discovery of new cellular binding targets of transcriptional activators.

3B: Background

Executive orders within a cell are implemented by multi-component complexes whose functional state and subcellular location are often influenced by association with exchangeable subunits.[1-3] In the process of transcription initiation, for example, several lines of evidence indicate that the Snf1/AMPK kinase complex arrives at the promoters of galactose catabolism genes only when the exchangeable subunit Gal83 is present within the complex[4]. Although the molecular mechanism by which enzymatic complexes such as Snf1/AMPK localize to promoters has been the subject of debate, biochemical and genetic data suggests that the most likely suspects performing the recruitment function are one or more transcriptional activators via protein-protein interactions (PPIs).[5-9] Distinguishing which of the components of an enzymatic (or coactivator) complex is the binding target of a promoter-bound activator is, however, quite difficult using conventional strategies. This is particularly true in cellular settings, where the

[3] Portions of this chapter are from published work: Dugan, A.; Majmudar, C.Y.; Pricer, R.; Niessen, S.; Lancia, J.K.; Fung, H.; Cravatt, B.F.; & Mapp, A. K. Discovery of enzymatic targets of transcriptional activators via *in vivo* covalent chemical capture. *JACS* Submitted. My contributions to this work include data in Figure 3-3 and Figure 3-5 and assisting in writing and figure creation.

transient nature of the activator-PPI network and the structural plasticity of the individual binding partners proves especially challenging.[10-12] For this reason, the PPI network of transcriptional activators that underpins transcriptional activation is limited to a handful of transcriptional machinery proteins.[13]

The prototypical transcriptional activator Gal4 stimulates the co-localization of a long list of enzymatic complexes at the promoters of yeast galactose catabolism genes.[14] In the presence of glucose, Gal4 is tightly repressed by its masking protein Gal80; however, when galactose becomes the primary carbon source, repression of Gal4 is relieved, enabling recruitment of complexes involved in galactose sensing and catabolism.[15] To define the protein-protein interaction network that Gal4 engages to recruit these functions to promoters, we sought an approach that would mitigate false positives caused by the often promiscuous *in vitro* binding profile of transcriptional activators. Thus, we focused on the interrogation of activator PPIs in the native cellular environment.

Here we describe the combination of *in vivo* covalent chemical capture with the photocrosslinking amino acid *p*-benzoyl-L-phenylalanine (Bpa) and label-free shotgun liquid chromatography-mass spectrometry analysis using the multi-dimensional protein identification technology (MuDPIT) to identify novel binding targets of the yeast activator Gal4 in live *Saccharomyces cerevisiae* that accompany gene activation (Figure 3-1). Both Gal4 and the viral activator VP16 associate directly with two components within the central metabolic regulator Snf1/AMPK: the kinase Snf1 and an exchangeable sub-unit Gal83.

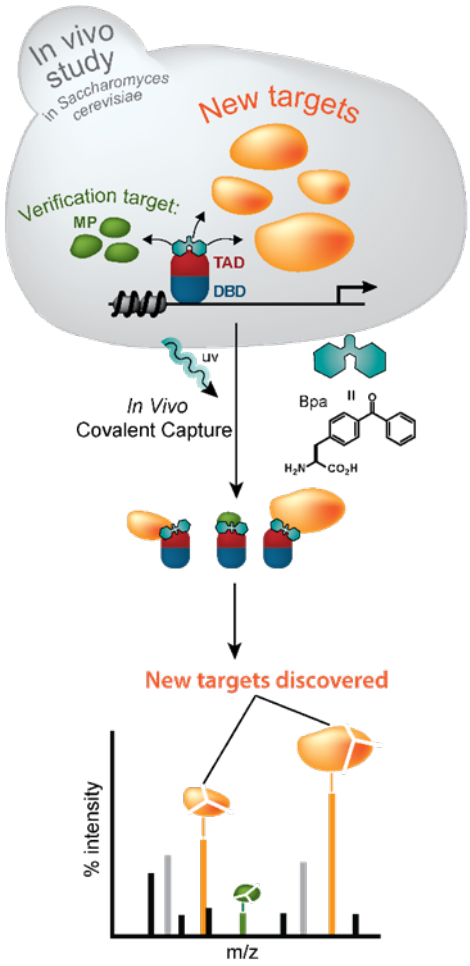


Figure 3-1 An In Vivo Covalent Chemical Capture and Mass-Spectrometric Based Approach for the Identification of the Cellular Targets of Transcriptional Activators.
Live yeast ex-pressing *p*-benzoyl-L-phenylalanine (Bpa)-containing activators (blue and red) are irradiated with UV light to covalently capture the spectrum of protein binding partners that activators directly contact in cells. Following cell lysis, affinity chromatography and immunoprecipitation are used to enrich for cross-linked complexes. The purified products are then subjected to mass spectrometric-based analysis such as Mud-PIT, thus allowing for the concurrent verification of previously identified activator targets (green) and the discovery of novel binding partners (orange).

Additionally, the combined use of reversible formaldehyde-based crosslinking and covalent chemical capture pinpoints the location of the Gal4-Snf1 PPI directly within the Gal1 promoter region. These data support a central role for amphipathic activators in the functional response of Snf1/AMPK to changes in cellular environment. As protein-protein interaction networks emerge as viable therapeutic tar-gets, covalent chemical capture *in vivo* will be a powerful tool for the discovery of molecular interaction points, particularly in dynamic assemblies.

3C: Combined Covalent Chemical Capture and MuDPIT Identifies Gal4-Gal80 Interaction In Vivo.

For *in vivo* covalent chemical capture, the strategy of unnatural amino acid incorporation using nonsense suppression was employed.[16-18] Given the complexity of the activator PPI interaction network, it was critical to incorporate Bpa at a position in Gal4 that displayed a robust and repeatable crosslinking profile. For this reason, Bpa incorporation at position Phe849 in Gal4 was selected for the combined covalent chemical capture MuDPIT experiments.[16] Initial optimization studies with this approach examined the sugar-dependent changes of the complex between Gal4 and Gal80 at a genomically integrated GAL1-LacZ reporter gene.[19] Importantly, earlier work by our group demonstrated that Bpa containing Gal4 mutants retain responsiveness to changes in nutrient *Ava*ilability (glucose, raffinose, and raffinose-galactose)[16]. Thus, the Bpa-containing Gal4 mutants in our yeast expression system behave in accordance with the findings of others who have studied activation and repression at the GAL1 promoter under various nutrient contexts.[15, 20] Gal4 Phe849Bpa is functionally active in galactose but inhibited in glucose, indicating a change in the Gal4-Gal80 interaction that is reflected in the *in vivo* covalent chemical capture-MuDPIT data in which Gal80 was significantly enriched under

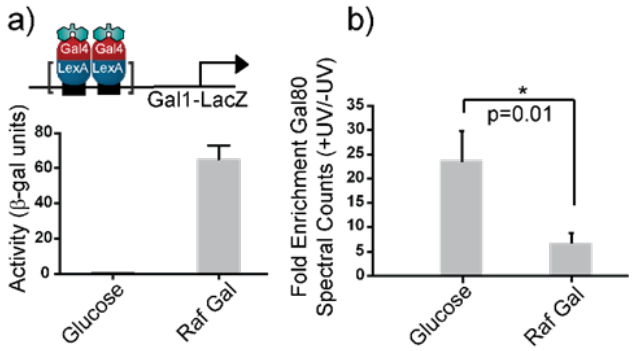


Figure 3-2 In Vivo Covalent Chemical Capture and MuDPIT Validate the Masking Protein Gal80 as a Direct Target of Gal4 in Live Yeast.
(a) Incorporation of the unnatural amino acid Bpa in at position 849 in the Gal4 transcriptional activation domain yields a functional activator that retains its ability to sense changes in carbon source availability, as determined via β-galactosidase assays on an integrated pGal1-LacZ reporter gene. In the presence of glucose, Gal4 is transcriptionally repressed but is activated in the presence of the inducing sugar, galactose, indicating that Gal4 remains responsive to repression by its masking protein Gal80. b) In line with the activity data, *in vivo* covalent chemical capture followed by MuDPIT analysis of purified Gal4 covalent adducts shows significant enrichment of the Gal4-Gal80 interaction under repressive glucose conditions (N=3) biological replicates.

glucose growth conditions (Figure 3-2). Together, these data indicate that the Gal4 mutant retains the responsiveness of its wild-type counterpart and additionally that Bpa crosslinking coupled with MuDPIT serves as a reliable platform upon which to examine activator PPIs *in vivo*.

3D: Identification of Novel Targets of Gal4 via In Vivo Covalent Chemical Capture and MuDPIT

Large scale (10L) transcriptionally active yeast cultures were subjected to the covalent chemical capture-MuDPIT workflow. The resulting datasets for non UV-treated (N=2 biological replicates) and UV-treated (N=3 biological replicates) samples were compiled and the spectral counts detected for each protein under each condition were averaged. Proteins demonstrating a maximal mean of less than two spectral counts were removed from further analysis. The remaining 638 proteins were subjected to stringent filter cutoffs according to the following criteria (Figure 3-3a): (i) Proteins must display a minimum average of five spectral counts (ii) Proteins must show a fold enrichment (+UV/-UV) greater than or equal to five and (iii) a t-test between UV-treated and non-UV control data sets for a protein must generate a p-value of less than or equal to 0.05[21]. A selection of protein targets meeting these criteria are displayed in (Figure 3-3a).

These crosslinking-MS studies revealed several previously unidentified targets of Gal4, all of which are part of larger complexes (Figure 3-3a). To validate a direct interaction between Gal4 and each of the proteins, Myc-tagged versions of each identified target were co-expressed alongside LexA+Gal4 F849Bpa and covalent chemical capture experiments were carried out. Following irradiation of live cells, yeast lysates were enriched with a LexA antibody and enriched proteins were analyzed via Western blot with a Myc-antibody to detect the presence of a covalently bound activator-target complex (Figure 3-3b). Two chaperones, Mas5 and Sti1, directly interact with Gal4; consistent with this result, deletion of these chaperone proteins in yeast has been previously demonstrated to significantly delay or abrogate induction of GAL1[22]. We additionally observe proteins involved in kinase complexes (Gal83)[23, 24], RNA binding complexes (Rrp5)[25], and those that are regulated by the TOR pathway (Nog1) directly interact, suggesting that previously identified crosstalk between the TOR and Snf1 nutrient sensing pathways extends to the level of transcription.[26-29]

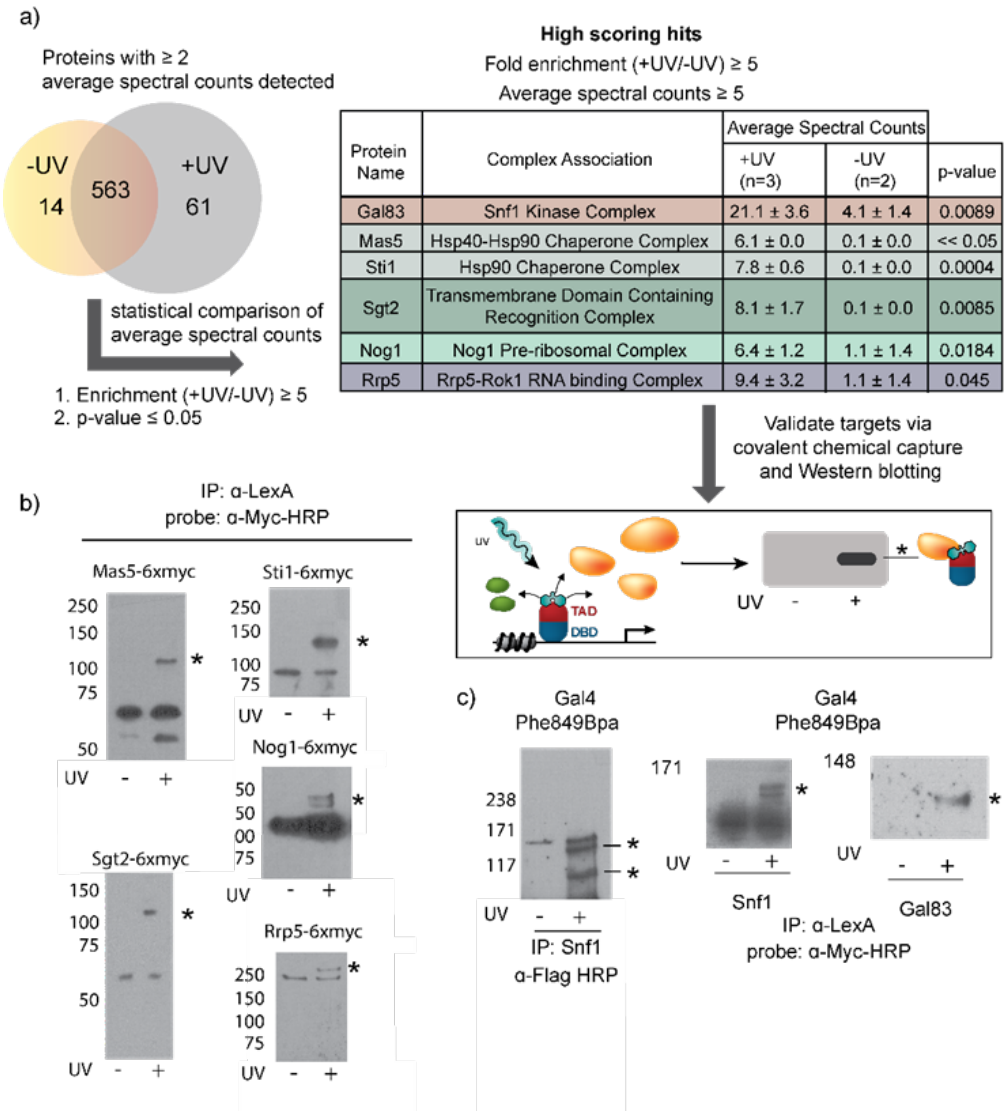


Figure 3-3 Covalent Chemical Capture and MuDPIT Reveals Several Novel Targets of the Transcriptional Activator Gal4 In Vivo.

a) Datasets from UV treated (n=3) and non-UV control (n=2) datasets were compiled and filtered to remove proteins with less than two average spectral counts detected. Of the proteins remaining, 14 were unique to the non-UV treated samples, 61 proteins were unique to the UV-treated condition, and 563 proteins were detected under both conditions. Stringent filtering criteria were applied to generate a list of proteins that were significantly enriched in the UV treated sample set. Proteins that were less than 5 fold enriched (+UV/-UV) and with p-values greater than 0.05 were removed from further analysis. A selection of previously unidentified direct targets of Gal4 Phe849Bpa are displayed as the number of spectral counts detected in UV treated and non-UV treated conditions. Novel targets detected via MuDPIT should be validated using covalent chemical capture Western blotting. See Supporting Information for additional details and complete MS results. b) Covalent chemical capture Western blot-ting validates the proteins identified from MuDPIT analysis are genuine targets of Gal4 within their respective complexes. Proteins identified from MuDPIT analysis were co-expressed with Gal4 Phe849Bpa in yeast and then subjected to the covalent chemical capture workflow. Yeast lysates were immunoprecipitated with a LexA antibody to enrich for covalent activator species and the subsequent Westerns were probed with a Myc-HRP antibody to detect the presence of covalently bound Myc-tagged proteins. c) Gal4 directly contacts the catalytic subunit Snf1 and the exchangeable subunit Gal83 in live yeast. *In vivo* covalent chemical capture was carried out in yeast expressing Gal4 Phe849Bpa. Immunoprecipitation of yeast lysates with a Snf1 anti-body and Western detection with a Flag-HRP antibody indicated that Gal4 contacts two proteins within the Snf1 complex. Myc-tagged versions of the Snf1 and Gal83 subunits were co-expressed alongside Gal4 Phe849Bpa in yeast. The resulting yeast lysates were immunoprecipitated with a LexA antibody as indicated and the immunoprecipitated complexes were analyzed by Western blot with a Myc-HRP antibody to detect covalently bound Gal4 complexes.

3E: Gal4 Directly Contacts the Snf1 Kinase and Gal83 Exchangeable Subunit of the Snf1/AMPK Kinase Complex

Of the targets identified in this study, one protein of particular note belongs to the Snf1 kinase complex. The Snf1 complex is a heterotrimeric complex containing a catalytic alpha subunit (Snf1), a regulatory gamma subunit (Snf4), and a third beta subunit that exchanges between Sip1, Sip2, and Gal83 to regulate subcellular localization of the complex. Importantly, both Gal4 and the Snf1 complex regulate galactose-inducible genes such as GAL1.[14, 30] The Snf1 complex plays a critical role in the response to environmental changes in carbon source.[4, 24] Several lines of evidence suggest that Snf1 phosphorylation of the Mig1 repressor contributes to the de-repression of carbon source-regulated genes.[24, 31, 32] The Gal83-containing isoform has been shown to be the complex that predominantly localizes to the nucleus, and, consistent with this model, our MuDPIT data shows a five-fold increase in spectral counts over background, thus supporting a nuclear Gal4-Gal83 interaction when cells are grown in galactose.[4] The Snf1 subunit of the complex is also present in the Mud-PIT data set but it appeared robustly under both conditions. A closer examination of the Snf1 sequence revealed the presence of a polyhistidine stretch on its amino terminus, and we thus postulated that it was likely enriched during the denaturing Ni-affinity column purification and additionally retained throughout the milder FLAG affinity purification, presumably via an interaction with Flag-tagged Gal4. In support of Gal4 targeting multiple proteins within the Snf1 complex, enrichment of yeast lysates with a Snf1 antibody revealed two proteins that Gal4 directly contacts within the complex under irradiating conditions (Figure 3-3c). We were able to confirm Snf1 and Gal83 as direct targets of Gal4 by co-expressing Gal4 Phe849Bpa with either Myc-Snf1 or Myc-Gal83 and following our established covalent chemical capture workflow (Figure 3-3c). Furthermore, a functionally inactive mutant, Gal4 867-869Ala, was unable to crosslink to Snf1, demonstrating that this interaction is specific and is not mediated by the presence of Bpa (Figure 3-4).

To further refine the Snf1 model, we performed chromatin immunoprecipitation (ChIP) to determine if Gal4 and Snf1 co-localize at the GAL1-LacZ gene used in our experiments. Consistent with earlier studies, both proteins showed significant enrichment, supporting a model where Gal4 and the Snf1 complex work collaboratively at GAL1 to drive reporter gene expression (Figure 3 5a).[6] Additional ChIP experiments with the Gal4 triple alanine mutant show a substantial loss in Snf1 localization to GAL1, supporting the importance of Gal4 in

coordinating Snf1 localization to the promoter (Figure 3-7). β-Galactosidase experiments with WT Gal4 in a Snf1 delete strain complemented with either WT Snf1 or a kinase dead T210A Snf1 mutant additionally support this model (Figure 3-7).

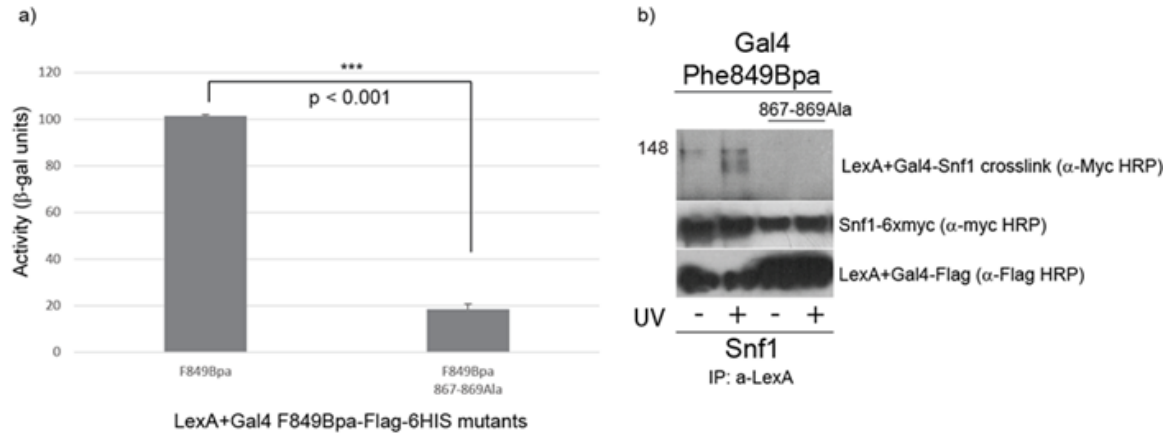


Figure 3-4 Activity and Expression of LexA+Gal4 F849Bpa Mutants.
(a) Bars represent the mean of at least three biological replicates. Error bars are the standard error of the mean (SEM). p-values were obtained by running an unpaired t-test between the two sample sets. (b) Expression and crosslinking of LexA+Gal4 F849Bpa mutants. *In vivo* covalent chemical capture using LexA+Gal4 F849Bpa or its cognate 867-869 triple alanine mutant was carried out. Yeast lysates were immunoprecipitated with a-LexA antibody and the blots were probed with a-Myc HRP antibody (top and middle panels) and then stripped and reprobed with a-Flag HRP antibody (bottom panel). The top panel shows crosslinking between Gal4 F849Bpa and Snf1. The middle panel shows free Snf1-6xMyc in yeast lysate input lanes. The bottom panel shows levels of Gal4 F849Bpa and its cognate mutant after immunoprecipitation with LexA antibody. Both mutants express in yeast, indicating that the loss of activity observed in (a) is not due to lack of mutant protein expression.

3F: The Gal4-Snf1 PPI is Sustained at the Gal1 Promoter

ChIP assays are suitable for determining the co-localization of proteins at the same promoter over a defined period of time, but they lack the resolution to determine if these proteins are in direct contact with one another when bound to DNA.[33] In the absence of methods to accomplish this goal and determine if the Gal4-Snf1 interaction is sustained at GAL1, we developed a new strategy which we have coined TRIC (tandem reversible and irreversible crosslinking) that combines the irreversible crosslink arising from UV-activated Bpa and the reversible protein-protein and protein-DNA crosslinks formed upon formaldehyde (HCHO) treatment. In this experiment, cells bearing Gal4 modified to have Bpa at position 849 were first treated with formaldehyde and then subjected to our standard UV-irradiation protocol. The treated cells were then lysed, the chromatin fraction isolated and sheared, and then immunoprecipitation with α-Snf1 antibody was carried out. Next, the formaldehyde crosslinks were reversed, thus leaving only the irreversible Bpa-crosslinks intact, and a Western blot of the Bpa crosslinked species with α-FLAG antibody conclusively demonstrated that Gal4 directly

interacts with the Snf1 complex on the DNA (Figure 3-5b). Taken together, these data support a model of recruitment for the Snf1 complex in which Gal4 directly contacts the catalytic Snf1 subunit and forms additional contacts with the exchangeable regulatory subunit Gal83.

3G: Snf1 and Gal83 are Shared Targets of Amphipathic Activators

Given the high homology between yeast Snf1 and mammalian AMPK, the interactions of Snf1 and VP16, an amphipathic viral activator that functions in the HSV-1 infection of mammalian cells, was examined.[34] The VP16 transcriptional activation domain is large (77 amino acids) and consists of two subdomains that can function independently. As has been done previously, we examined the binding behavior of each subdomain individually.[35] *In vivo* covalent chemical capture with VP16N (413-456) bearing Bpa at position 444 resulted in an observable crosslink with Snf1 (Figure 3-5c) but no interaction with Gal83 was identified for this subdomain under any conditions examined. In contrast, VP16C (446-490) with a Phe475Bpa mutation interacts with both of the subunits (Figure 3-5d). These data suggest that the Snf1 and Gal83 subunits of

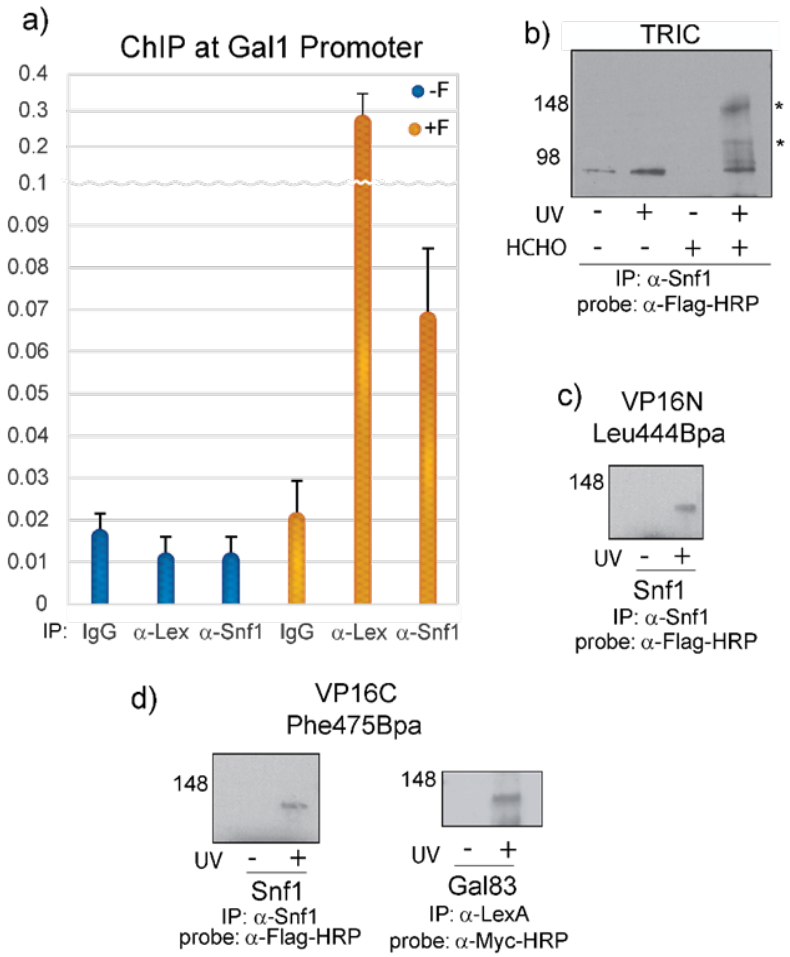


Figure 3-5 In Vivo Photocrosslinking with Bpa Captures the Direct Targets of Amphipathic Activators within the Snf1 Kinase Complex. (a) Chromatin immunoprecipitation with Snf1 and LexA antibodies verify that Gal4 and the Snf1 kinase complex co-localize to the Gal1-LacZ reporter gene used in these studies. (b) The Gal4-Snf1 interaction is maintained on DNA. Tandem reversible and irreversible crosslinking (TRIC) was performed to investigate the interaction of DNA bound Gal4 with Snf1 at the Gal1 promoter. (c,d) The Snf1 complex is a common target of amphipathic activators. The amino- and carboxy-terminal subdomains of the amphipathic activator VP16 were tested for crosslinking to the catalytic subunit of the Snf1 kinase complex as well as a Myc-Gal83 construct in yeast. Covalent complexes were analyzed by Western blot with a Flag or Myc antibody to detect covalently bound Snf1-VP16 and Gal83-VP16, respectively.

Snf1/AMPK are a common target for these two activators specifically and perhaps for amphipathic activators as a class.

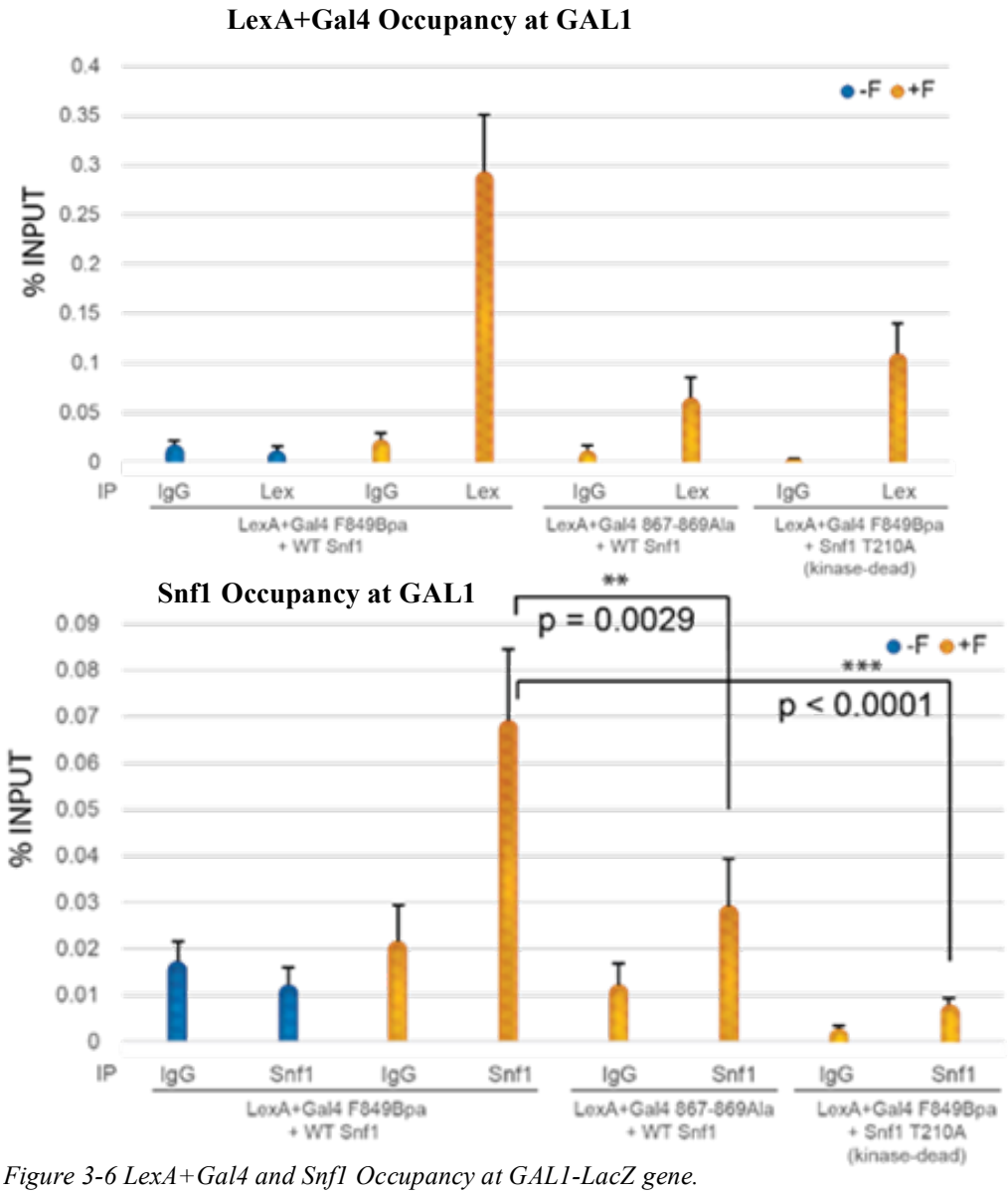


Figure 3-6 LexA+Gal4 and Snf1 Occupancy at GAL1-LacZ gene.
All experiments were carried out in a yeast strain lacking Snf1 (Δsnf1) and either WT Snf1-6xMyc or a kinase dead Snf1-6xMyc T210A mutant complemented back in. With the exception of the Gal4 867-869 triple alanine mutant, all ChIP studies were done with the Gal4 F849Bpa construct. Cultures were grown to mid-log phase in glucose and then induced with galactose prior to treatment with formaldehyde. Each bar on the graph represents the % input reported as the mean and standard error from at least three independent experiments. (Left) ChIP analysis of Gal4 occupancy at the GAL1-LacZ gene as determined by immunoprecipitation with an antibody for LexA. (right) ChIP analysis of Snf1 occupancy at the GAL1-LacZ gene as determined by immunoprecipitation with an antibody for Snf1 protein.

3H: Conclusion

Snf1/AMPK and Gal4 each play a central role in the ability of yeast to rapidly respond to changes in nutrient context.[14, 36] Upon the switch to galactose, the Gal83-containing isoform of

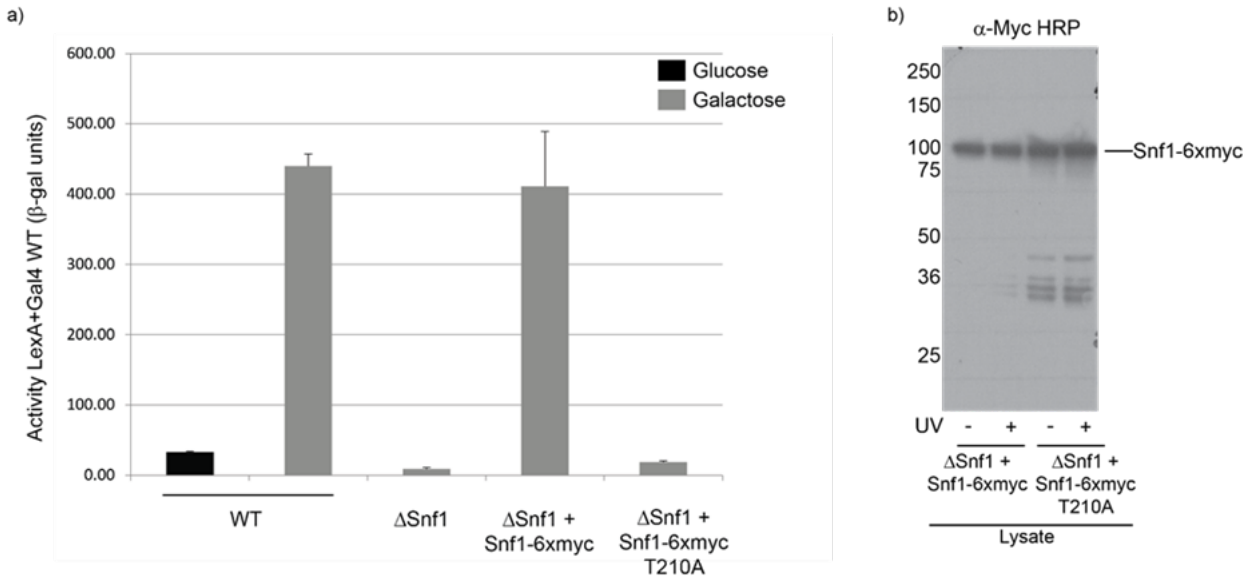


Figure 3-7 Activity and Expression of Snf1 Kinase Mutants.
(a) Beta-galactosidase assays were carried out for LexA+Gal4 WT in the indicated yeast strains. Cultures were grown in glucose to mid-log phase and then induced with galactose prior to measuring activation potential. Data labeled 'WT' were generated in a yeast strain (LS41) that expresses endogenous Snf1 kinase. All other data was generated in a yeast strain that has had Snf1 kinase deleted (ΔSnf1) and either WT Snf1-6xMyc or a kinase dead Snf1-6xMyc T210A mutant co-expressed alongside the activator. (b) Expression of WT Snf1-6xMyc and Snf1-6xMyc T210A mutant. An aliquot of whole cell lysate from cultures expressing either the WT or kinase-dead Snf1-6xMyc protein was separated on SDS-PAGE and analyzed via Western blot with a Myc-HRP antibody. The free Snf1 protein runs at approximately 100 kDa, with both constructs expressing in the ΔSnf1 yeast strain.

Snf1/AMPK transits to the nucleus where it phosphorylates targets at galactose-catabolism genes, contributing to de-repression of nutrient response genes.[4, 24] Galactose similarly signals a change in the Gal4-Gal80 complex, leading to rapid up-regulation of the GAL genes[14]. The covalent chemical capture data with Gal4 reveals two direct PPIs that connect these two critical nutrient response pathways. Our results show that the activator Gal4 directly contacts not only the kinase Snf1 within the Snf1/AMPK complex, but also the exchangeable subunit Gal83. It is conceivable that by contacting two components of a single complex, Gal4 is not only able to cooperatively recruit the Snf1/AMPK complex and its associated kinase activities, but also orient the complex in its appropriate configuration, ultimately providing a rapid transcriptional response to process the external stimulus. In retrospect, we observed the same phenomenon with the Swi/Snf chromatin modifying complex; in that case two regions of VP16 contacted the enzymatic subunit and a second scaffolding component.[35] An open question is what role these two distinct types of interactions play. Certainly, proper positioning at the promoter is critical and, in the case of the VP16-Swi/Snf interaction, EM studies support this role.[37] An additional

possibility is that the PPIs alter the activity and/or substrate profile of the enzymatic subunit through allosteric effects. Studies investigating this latter possibility are currently underway.

The use of *in vivo* covalent chemical capture followed by adduct analysis via shotgun LC-MS methods (e.g., MuDPIT) represents a significant step towards the development of a complete *in vivo* interaction map of transcriptional activators. While a powerful approach, it should be noted that issues of relative abundance, stability, and ease of digestion may con-tribute to uneven coverage of proteins across the proteome.[38] Indeed, many of the 'typical' targets of transcriptional activators are of lower abundance including, for example, Gal11/Med15.[39, 40] In the case of Gal11/Med15, even in an *in vitro* crosslinking experiment in which an abundance of crosslinked products (Gal11/Med15-activator) were isolated and purified prior to MS analysis it was necessary to use two different digestion strategies and two different ionization techniques in order to achieve sufficient sequence coverage.[41] Thus, continued advances in MS methods coupled with fine-tuning crosslinker reactivities will undoubtedly enhance the utility of tandem covalent chemical capture-MS approaches.[42-45] Additionally, this study captured binding partners of Gal4 that rely upon the sequence surrounding position 849 for interaction. As our previous work and that of others has shown, future experiments with Bpa placed at distinct positions will likely lead to additional binding partners.[15-18, 35]

In this study, this strategy resulted in the identification of two subunits of the Snf1 kinase complex as novel, direct targets of the amphipathic activators Gal4 and VP16 *in vivo*. The Snf1 complex is highly conserved among eukaryotes and the mammalian counterpart of the yeast Snf1 kinase complex, AMPK, plays a significant role in maintaining cellular homeostasis by functioning in some cases as a tumor suppressor and additionally as a regulator of energy response.[36, 46] Given its important role in the cell, AMPK is a critical target in the treatment of diseases that exhibit abnormal metabolic profiles including certain cancers and diabetes.[47, 48] The establishment of a direct link between the Snf1/AMPK complex and transcriptional activators suggests a novel intervention point for extrinsic regulation of Snf1/AMPK promoter activity. This will be the focus of future efforts.

3I: Experimental

Yeast strains and plasmid construction

Yeast strains used in this study were LS41 [JPY9::pZZ41, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4 URA::pZZ41] and ΔSnf1 LS41 [JPY9::pZZ41, Matα his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4 URA::pZZ41 SNF1::TRP1].

All plasmids were generated using standard restriction enzyme digestion and ligation reactions and the sequences of all the isolated plasmids were verified by sequencing at the University of Michigan Core Facility (Ann Arbor, MI). T4 DNA ligase and all restriction enzymes were all purchased from New England Biolabs. Mutagenesis of Snf1 and Gal4 was carried out using a two-step site directed mutagenesis protocol (Quikchange, Qiagen).

pLexAVP16N Leu444Bpa, pLexAVP16C Phe475Bpa, and pSNRtRNA-pBpaRS plasmids were constructed as previously described.[16, 35] Briefly, the LexA fusion genes were cloned into vector pCLexA, which has a 2μ origin of replication and an ADH1 promoter that drives activator expression. Importantly, in our previous work we demonstrate that maximal Bpa incorporation is about 20% compared to the wild-type protein expression. Thus, while the vector used in these studies is high copy, the expression of Bpa-containing activators is likely much lower than that achieved with canonical protein expression. pLexA+Gal4 Phe849Bpa-Flag-6HIS was created by modifying previously described pLexA+Gal4 Phe849Bpa to include a 6xHis tag 3' to the Gal4-Flag open reading frame.[16] Insertion of the 6xHis tag was achieved by site-directed mutagenesis with primers P11/P12 and Qiagen Quikchange two-step SDM protocol. For TRIC experiments, a 3x Flag tag was additionally inserted using site directed mutagenesis with primers P9 and P10 (Qiagen Quikchange). To create p6xMyc-Snf1, primers P1 and P2 were used to clone out the Snf1 gene from yeast genomic DNA. The purified PCR product was then digested with XhoI and NdeI and ligated into XhoI and NdeI digested pGADT7-6xMyc backbone using standard molecular cloning techniques. To create Gal83-6xMyc, primers P3 and P4 were used to clone out the Gal83 gene from yeast genomic DNA. The purified Gal83 PCR product was then digested with HindIII and NdeI and ligated into a HindIII/NdeI digested pGADT7-6xMyc backbone using standard molecular cloning techniques. Snf1-6xMyc T210A was made using the Qiagen Quikchange two-step mutagenesis protocol with primers P5 and P6. All sequences were verified by the University of Michigan DNA Sequencing Core.

ΔSnf1 LS41 was constructed using a PCR-mediated gene deletion strategy with a cassette containing the full TRP1 open reading frame and ~40 bp complementarity to the 5' and 3' flanking regions of genomic Snf1. The SNF1::Trp1 deletion cassette was transformed into LS41 using a lithium acetate transformation protocol and transformants were selected on plates lacking uracil and tryptophan.[49] Colonies were screened using polymerase chain reaction (PCR) with PCR products visualized on a 1% agarose gel stained with ethidium bromide.

Primers P13 and P14 were used to clone out the Trp1 gene off pFA6a-TRP1-PGAL1-HBTH (Addgene, plasmid # 26898). The PCR product was verified on 1% agarose gel stained with ethidium bromide and the product was gel purified (QIAquik gel extraction kit and protocol, Qiagen). The purified Snf1-Trp1 cassette was transformed into LS41 yeast using standard lithium acetate/PEG transformation methods and plated on plates containing 2% glucose and lacking uracil and tryptophan for selection.[49] Transformants were screened via colony PCR using primer combinations P15/P16, which are specific for the Snf1 locus, and colonies showing a PCR product of ~750 bps were screened further using primer combinations P15/P18, P16/P17, P15/P20, and P16/P19 to verify that Snf1 was definitively replaced by the TRP1 gene at the Snf1 locus. Further confirmation was obtained by submitting the P15/P16 gel purified PCR product for sequencing, which returned a TRP1 sequence.

Table 3-1 Primers Used in Study

Primer ID	Sequence (5'-3')	Description
P1	GTACATCTCGAGATGAGCAGTAACAACAACACAAAC	XhoI-Snf1 F
P2	GCTACGCATATGATTGCTTTGACTGTTAACGGCTAATTC	NdeI-Snf1 R
P3	GACTTAAAGCTTATGGCTGGCGACAACCCTGA	HindIII-Gal83 F
P4	GCTACGCATATGATTACTGATCTGAGACTTTTG	NdeI-Gal83 R
P5	CATTACCTCGAGATGTCAGCATCAAAAGAAGAAATTGCTGCCC	XhoI-Sgt2 F
P6	GTAATGGGATCCCTATTGCTTGTTCTCATTGTCTGGTGTTTCATCTG	BamHI-Sgt2 R
P7	CATTACCTCGAGATGTCACAAGACGCTGCTATTGCAGAGC	Xho1-Hxt6 F
P8	GTAATGGGATCCTTTGGTGCTGAACATTCTCTTGTACAATGGC	BamHI-Hxt6 R
P9	CATTACCTCGAGATGGTTAAAGAAACTAAGTTTTACGATATTCTAGGTG	XhoI-Mas5 F
P10	GTAATGGGATCCTCATTGAGATGCACATTGAACACCTTCG	BamHI-Mas5 R
P11	CATTACCTCGAGATGTCATTGACAGCCGATGAATACAAACAACAAGG	XhoI-Sti1 F
P12	GTAATGGGATCCTTAGCGGCCAGTCCGGATGATACCAG	BamHI-Sti1 R
P13	CATTACCTCGAGATGCAACTTTCATGGAAGGATATCCCTACTGTCGCTCC	XhoI-Nog1-F
P14	GTAATGGGATCCTCAACGGAAATCTGTCTTACCGACACCACGCTTACCAC	BamHI-Nog1-R
P15	CATTACGGATCCATGGTAGCTTCCACCAAAAGAAAGAGAGATGAAGATTTTCC	BamHI-Rrp5-F
P16	GTAATGCCCGGGTTCGTCTGCTTTTTGAGATTCATGGCTAGCGAC	XmaI-Rrp5 R
P17	GACTGATGGTAATTTCTTAAAGGCTTCTTGTGGTTCTCCCAATTATGCG	SDM Snf1 T210A F
P18	CGCATAATTGGGAGAACCACAAGAAGCCTTTAAGAAATTACCATCAGTC	SDM Snf1 T210A R
P19	CAATGGATGATGTATATAACGCTGCTGCTGATGATGAAGATACCCC	SDM Gal4 867-869 AAA F
P20	GGGGTATCTTCATCATCAGCAGCAGCGTTATATACATCATCCATTG	SDM Gal4 867-869 AAA R
P21	CAAACCCAAAAAAAGAGGATTACAAGGATCACGATGGTGATTACAAGGA TCACGATATTGACTATAAAGACGACGAC	SDM 3xFlag insert F
P22	GTCGTCGTCTTTATAGTCAATATCGTGATCCTTGTAATCACCATCGTGAT CCTTGTAATCCTCTTTTTTTGGGTTTG	SDM 3xFlag insert R
CTATAAAGACGACGACGACAAACATCATCATCATCATCATTAAGGATCCGTCGACCTGCAG		SDM 6HIS F
CTGCAGGTCGACGGATCCTTAATGATGATGATGATGATGTTTGTCGTCGTCGTCTTTATAG		SDM 6HIS R
GTAACAAGTTTTGCTACACTCCCTTAATAAAGTCAACATGTCTGTTATTAATTTCACAGGTAGT TCTGGTCC		5' Snf1-TRP1 cassette F
CATAAAAAAAAGGGAACTTCCATATCATTCTTTTACGTTCCACCATTATTTCTTAGCATTTTTGA CGAAATTTGC		3' Snf1-TRP1 cassette R
CTGCGCATTCGTGTCCAAACAGTCATTCAGG		gSnf1 +210 seq F
CATAGAGCGTGAAATTTGCTTTTCATCCGAAG		gSnf1 -193 seq R
GCCTACATCTTTACCTCAGATCCACAGAGC		Snf1 int. F seq
GCCTATGGCAGTACTCGACGGCAC		Snf1 int. R seq
CTGCATGGAGATGAGTCGTGGC		Trp1 int. F seq
GCCACGACTCATCTCCATGCAG		Trp1 int. R seq
CCTTCTCTTTGGAACTTTCAGTAATACGCTTAACTGC		GAL1-LacZ F
GGGCGATCGGTGCGGGCCTCTTCGC		GAL1-LacZ R

***In vivo* covalent chemical capture**

In vivo covalent chemical capture was carried out as previously described, but on larger scale (1-10 L).[16] For all covalent chemical capture experiments, Bpa purchased from Chem Impex International (cat.# 05110) was used. To perform *in vivo* cross-linking, an individual colony expressing pLexA+Gal4 F849TAG-Flag-6xHis was inoculated in 10 mL SC media containing 2% raffinose and lacking histidine, tryptophan, and uracil (H-W-U-) for selection. This culture was incubated overnight at 30 °C with agitation and used to inoculate 1L cultures of H-W-U- SC media containing either 2% glucose or 2% raffinose and 2% galactose. To each 1 L culture, 10 mL of 100 mM Bpa dissolved in 1M NaOH and 10 mL 1M HCl were added. Cultures were incubated overnight at 30°C with agitation and grown to mid-log phase OD660 (~0.8). Yeast were harvested by centrifugation at 6000 rcf at 4˚C for 15minutes (Beckman *Ava*nti J-201, JLA 8.1000). Cell pellets were washed with SC media lacking histidine and tryptophan and resuspended in 2mL H-W-U- SC media with either 2% raffinose and 2% galactose or 2% glucose. Yeast were transferred to cell culture dishes and subjected to UV irradiation at 365 nm light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour. Irradiated cells were then harvested by centrifugation and stored at -80˚C until lysis.

To obtain each MS data set, 10 L yeast cultures were grown and irradiated as described previously but on larger scale.[16] Briefly, yeast were grown in SC media containing 2% raffinose, 2% galactose with 1 mM Bpa but lacking histidine, tryptophan, and uracil (H-W-U-) for selection and grown at 200 rpm, 30 ˚C with constant air bubbling using a bioreactor (New Brunswick, Bioflo 300) to mid-log phase OD_{660} (~0.8). Yeast were harvested by centrifugation at 6000 rpm for 15 minutes and resuspended in SC media with 2% raffinose and galactose but lacking histidine, tryptophan, uracil (H-W-U-), Bpa and irradiated for 1 hour at 365 nm with cooling using a Euro Solar UV lamp. Control samples, where no irradiation was to be performed were directly taken to the next step. After irradiation, the yeast were centrifuged and washed 2X with 100 mM PBS and resuspended in ~ 40 mL MS Lysis Buffer (100 mM PBS pH 7.0, 150 mM NaCl, 10% glycerol, 0.5% NP-40 and 10 mM beta-mercaptoethanol) containing 2x protease inhibitor (complete mini EDTA free protease inhibitor tablets, Roche) and flash frozen with liquid Nitrogen and stored at -80˚C until lysis.

Cryolysis and Affinity Purification of Crosslinked Samples for MuDPIT Analysis

Cryolysis was performed using a Retsch planetary ball mill PM 200. All equipment was cooled in liquid nitrogen prior to use. Frozen yeast cells were crushed in a chilled mortar and pestle with liquid nitrogen added. Ground cells were added to the chilled lysis chambers and loaded into the ball mill. PM-200 was run three times for two minutes/cycle at 500 rpm with liquid nitrogen cooling performed between each cycle. Chambers were removed from machine after final spin and the frozen lysed yeast were recovered from the chambers and placed in a 50 mL centrifuge tube on dry ice.

Lysed yeast were allowed to thaw in a cold water bath and then centrifuged for 30 minutes at 4˚C, 9299 rcf (Beckman Coulter Allegra X-22R, fixed rotor). While cells were spinning, 300-500 mL Ni-agarose bead slurry was washed three times with 1 mL chilled PBS. Supernatant from lysed yeast was transferred to a clean 50 mL centrifuge tube and the Ni-agarose beads (resuspended in 1 mL chilled PBS) were added to the supernatant. Beads and supernatant were incubated for 1 hour on rotating carousel in 4˚C cold room. After incubation, beads were spun down (low rpm) and the supernatant was poured off. Ni-agarose beads were then resuspended in the residual supernatant and washed in 5 mL chilled Nickel Wash Buffer (100 mM PBS pH 7.0, 150 mM NaCl, 10% glycerol, 3M Guanidine-HCl, 30 mM imidazole, 0.2% Tween-20, 10 mM Beta-mercaptoethanol) in a 15 mL centrifuge tube. Washed beads were centrifuged at 2500 rpm for 2 minutes at 4˚C (Beckman Coulter Allegra X-22R, bucket rotor) and the supernatant decanted. This procedure was repeated 4 more times for a total of five-5 mL washes. The beads were then washed twice with 5 mL cold 100 mM PBS. PBS was decanted and the beads were resuspended in 500 µL 100 mM PBS and transferred to a chilled 1.75 microcentrifuge flip cap tube. Residual beads on the 15 mL tube were washed with about 500 µL 100 mM PBS and transferred to a microcentrifuge tube. Beads were spun briefly at low rpm (Eppendorf 5417C microfuge).

Beads were resuspended in 500 µL chilled Nickel Elution Buffer (100 mM PBS pH 7.0, 500 mM Imidazole, 0.5% Tween-20) and mixed gently for 1 minute by inversion and then re-centrifuged. The supernatant (eluant) was pipetted off and collected in a separate, clean 1.75 mL microcentrifuge tube. Another 500 µL chilled Nickel Elution Buffer was added to the beads and mixed by inversion for 10 minutes, 4˚C in a cold room. During the second elution, a 10K microcentrifuge tube concentrator (Vivascience 10K MWCO) was washed with 100 mM PBS.

The first elution was centrifuged to remove any lingering Ni-agarose and the eluent was pipetted out and added to the concentrator. Concentrator was spun at 15,000xg for 5 minutes at 4˚C. Elutions were repeated for a total of 3 elutions, each elution added to the concentrator after centrifugation to remove residual Ni-agarose. 500 µL of 100 mM chilled PBS was added to the protein and then run through the concentrator. This was repeated for a total of 3 PBS buffer exchanges to get a final volume of 1000-1500 µL of protein.

400 µL of Flag-agarose (Sigma, M2) was washed three times with 1 mL 100 mM PBS pH 7.0. Concentrated sample was added to the Flag-agarose and incubated for 2 hours at 4˚C, with rotation. Flag-agarose was then washed ten times with 1 mL/wash Flag Wash Buffer (100 mM PBS pH 7.0, 150 mM NaCl, 0.5% NP-40), spinning between washes (5000 rpm, 30 second, Eppendorf 5417C). Next, Flag-agarose was washed twice with 100 mM PBS to dilute out the NP-40. Flag Elution Buffer (100 mM PBS pH 7.0, 350 mM NaCl, 0.1% NP-40) containing 10 mg/mL Flag peptide was used for elution. 333 µL Flag elution buffer was added to washed Flag-agarose and incubated for 30 minutes, 4˚C, with rotation. Beads were briefly centrifuged and the elution transferred to a new 1.75 mL Eppendorf tube. Elution was quick frozen in liquid nitrogen and stored in -80˚C freezer. Second elution was performed in the same fashion and the final 333 µL was added to beads for a third elution that was allowed to run overnight. The next morning, the two frozen elutions were thawed and spun down to remove residual Flag-agarose. Elutions were transferred to new tubes and then centrifuged again to remove any chance of residual beads remaining. These first two elutions, cleared of all beads, were then concentrated in a 10K concentrator at 15,000 xg until down to 100 µL. The overnight elution was then added and centrifuged down to 100 µL. 500 µL of ammonium bicarbonate (50 mM) was then added to the concentrator for a total of 4 buffer exchange cycles with the protein. Concentrated sample was then transferred to a clean microcentrifuge tube. Sample was quick frozen in liquid nitrogen, packed in dry ice, and shipped for mass-spectrometry analysis.

β-Galactosidase assays

To evaluate the ability of each LexA+Gal4 F849Bpa-Flag-6HIS to activate transcription in the presence or absence of glucose, saturated cultures (SC media + 2% raffinose) of yeast expressing Gal4 Phe849Bpa were used to inoculate 5 mL SC media containing either 2% glucose or 2% raffinose & 2% galactose but lacking histidine and tryptophan for selection. The cells

were grown to an OD660 of 0.8-1 and harvested. The activity of each construct was monitored using β-galactosidase assays as previously described.[16]

For activity of LexA+Gal4 WT in ΔSnf1strains, a single colony of each construct was inoculated in 5 mL yeast dropout media containing 2% glucose and grown overnight at 30°C. These cultures were then used to inoculate additional 5 mL cultures of dropout media containing 2% glucose and 1 mM Bpa. Cultures were grown to mid-log phase, harvested via centrifugation, washed with dropout media, and, with the exception of the glucose control culture, were then resuspended in media containing 2% raffinose, 2% galactose, and 1 mM Bpa. Cultures were induced for 3 hours at 30°C before being harvested. The activity of each construct was monitored using β-galactosidase assays as previously described.[35]

MuDPIT analysis

All MuDPIT analyses were performed at the Center for Physiological Proteomics in La Jolla, California. For both the glucose-grown and galactose-grown cultures, three biological replicates were analyzed for UV-treated samples and two biological replicates were analyzed for non-UV control samples. Spectral counting was used for identifying hits. Crosslinked products were denatured by adding urea to a final concentration of 8 M, in 50 mM Tris (pH 8.0). The protein was then reduced with 10 mM Tris(2-carboxyethyl) phosphine HCl (TCEP, Sigma) for 30 minutes at room temperature. The proteins were subsequently alkylated with fresh 12.5 mM iodoacetamide (IAA; Sigma) and the concentration of urea was reduced to 2 M by adding 50 mM Tris (pH 8.0). Next, crosslinked proteins were digested overnight with 1 mM $CaCl_2$ and 2 μg Trypsin at 37˚C. Digested peptides were acidified to a final concentration of 5% formic acid and centrifuged at 17,000xg for 15 minutes. Half of the digested mixture was pressure loaded onto a biphasic strong cation exchange/reverse phase capillary column and separated by 2D liquid chromatography and tandem MS using an 11 step gradient on an LTQ mass spectrometer. Full MS spectra were acquired with a mass range of 400-1,800 followed by 7 MS/MS scans. All MS/MS spectra were collected using a normalized collision energy of 35% and an isolation window of 2 Da. One microscan was applied for all experiments in the LTQ. Spray voltage was set to 2.50 kV, and the flow rate through the column was 0.20 μL/minute. The data was analyzed against the entire yeast genome using the SEQUEST algorithm and the analysis was performed using the DTASelect software package. For both the glucose-grown and galactose-grown cultures, three biological replicates were analyzed for UV-treated samples and two

biological replicates were analyzed for non-UV control samples. Spectral counting was used for identifying hits. Hits from each replicate were compiled and the mean spectral counts for each protein were calculated. Proteins with less than two average spectral counts were removed from further analysis. UV-treated and non-UV samples within a single carbon source condition were compared to determine enrichment values (Average +UV SC/average –UV SC). Proteins were considered significant if the enrichment ratio (+UV/-UV) was greater than or equal to 5 and if a t-test comparing UV-treated and non-UV data for a protein generated a p-value of less than or equal to 0.05. Hits with unknown function and proteins already identified as targets of Gal4 were not displayed in.

To avoid zero values in the denominator, all spectral count values were adjusted to contain an additional 0.1. Spectral counts for each protein were averaged within the non-UV (N=2) and +UV (N=3) datasets. Proteins with a spectral count average less than 2 in both the –UV and +UV datasets were removed from further analysis. Enrichment ratios were calculated for proteins with at least two average spectral counts in either condition (Average +UV SC/Average non-UV SC.) and t-tests were carried out in Excel to generate p-values for the enrichment ratios. Proteins with an enrichment ratio of five or greater and a p-value of 0.05 or less were considered to be significant hits. Those with an average of at least five spectral counts were considered high-scoring hits and those with an average between 2 and 5 spectral counts were considered moderate-scoring hits. Proteins that were previously identified as Gal4 targets were excluded from the list (Gal80) as were proteins with no characterized function.

Preparation and Western blotting of Snf1 and Gal83 crosslinked samples

An individual colony expressing either pLexA+Gal4 Phe849Bpa-6xHis-Flag, pLexA+VP16 Leu444Bpa-Flag, or pLexA+VP16C Phe475Bpa-Flag was inoculated in 5 mL SC media containing 2% raffinose but lacking histidine, tryptophan, and uracil (H-W-U-) for selection. This culture was incubated overnight at 30°C with agitation and then used to inoculate 100 mL cultures of H-W-U- SC media containing 2% raffinose and 2% galactose. To each culture, 1 mL of 100 mM Bpa dissolved in 1M NaOH and 1 mL 1M HCl were added. Cultures were incubated overnight at 30°C with agitation and grown to mid-log phase OD660 (~0.8). Yeast were harvested by centrifugation at 3901 rcf at 4˚C for 5minutes (Beckman-Coulter Allegra X-22R). Cell pellets were washed with H-W-U- SC media and resuspended in 2mL H-W-U- SC media containing 2% raffinose and 2% galactose. Yeast were transferred to cell

culture dishes and subjected to UV irradiation at 365 nm light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour. Irradiated cells were then harvested by centrifugation and stored at -20˚C until lysis. For lysis, cell pellets were resuspended in 600 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche)) and lysed using glass beads by vortexing at 4°C. The lysate was pelleted and the supernatant incubated with 10 µL (2 ng) of Snf1 antibody (sc-15621, Santa Cruz Biotechnologies) for 2 hours at 4°C for immunoprecipitation. The protein bound to the antibody was isolated by incubation for 1 hour with 40 µL of prewashed protein G magnetic Dynabeads (Life Technologies) at 4˚C. After immunoprecipitation, the beads were washed 6 times with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -20˚C until elution. The crosslinked sample was eluted from the beads by heating at 95˚C for 10 minutes in Laemmli 2x Buffer (BioRad) containing 250 mM DTT and run on a 4-20% Tris-Glycine TGX gel (BioRad). The separated proteins were then transferred to PVDF membrane and probed using Western Blot analysis with a 1:1000 dilution of anti-FLAG (M2) antibody (Sigma) in 5% milk PBST.

Crosslinking to Gal83 was carried out in the same way, except that p6xMyc-Gal83 was co-expressed alongside each activator construct in yeast. Cells were additionally grown in media lacking histidine, tryptophan, leucine and uracil for selection and lysates were immunoprecipitated with a LexA antibody (sc-1725, Santa Cruz Biotechnologies). Resulting Western blots were probed with a 1:1000 dilution of Myc-HRP antibody (sc-40 HRP, Santa Cruz Biotechnologies) in 5% milk PBST.

Chromatin Immunoprecipitation (ChIP) at GAL1-LacZ

All ChIP experiments were conducted in ΔSnf1 LS41 strain with pGADT7-Snf1-6xMyc complemented back in. An individual colony of pLexA+Gal4 Phe849Bpa-Flag-6HIS or its cognate triple alanine mutant was grown in 5 mL SC media containing 2% glucose but lacking histidine, tryptophan, leucine, lysine and uracil for selection. The culture was incubated overnight at 30 °C with agitation. This culture was used to inoculate 100 mL cultures of SC media containing 2% glucose, 1 mL 100 mM Bpa dissolved in 1M NaOH and 1 mL 1M HCl. The cultures were incubated overnight at 30 °C with agitation to an OD660 of ~0.8. Cultures were harvested via centrifugation, washed with media, and added to induction media composed

of SC media lacking HWKLU and containing 2% raffinose, 2% galactose and 1 mM Bpa. Cells were induced for 3 hours prior to formaldehyde treatment. Non-formaldehyde treated cultures were harvested by centrifugation at 3901 rcf (Beckman Coulter Allegra X-22R, 4˚C) and washed with 25 mL media lacking his, trp, lys, leu, and ura. Cell pellets were transferred to a 1.75 mL tube, re-centrifuged, and the pellet frozen and stored at -80˚C until lysis. Formaldehyde treated samples were equipped with a stirbar on a stir plate. To each culture, 2.9 mL 37% formaldehyde solution was then added and allowed to stir for 5 minutes at room temperature, at which point the reaction was quenched with excess 2 M glycine for 5 minutes with stirring. The treated culture was harvested via centrifugation, washed with 50 mL media lacking his, trp, lys, leu and ura, transferred to 1.75 mL tube and the cell pellet frozen and stored at -80˚C until lysis.
Cell pellets were thawed on ice and lysed as previously described.[35] After mechanical shearing and whole cell lysate separation, the lysate and insoluble pellet were resuspended by pipetting and each sample was sonicated at a setting of 10% (double-step microtip, Fisher Scientific Dismembrator Model 500) for 2 minutes on ice with 30 second pulse on/off. Samples were then centrifuged at 4˚C for 20 minutes at max speed. Soluble chromatin was separated from the pellet and 10 µL removed as an input sample. The remaining soluble chromatin was split equally between three 1.75 mL tubes. Immunoprecipitation with either 2 µg Snf1 antibody (sc-15621, Santa Cruz Biotechnologies), LexA antibody (sc-1725, Santa Cruz Biotechnologies), or control IgG (sc-2027, Santa Cruz Biotechnologies) proceeded for 2 hours, 4˚C followed by incubation for 1 hour with ~40 µL of prewashed protein G magnetic Dynabeads slurry (Life Technologies). Beads were washed twice with 1X ChIP Lysis Buffer ((50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche)), twice with 1X ChIP Lysis buffer containing 0.5M NaCl, twice with 1x ChIP Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 1% Na-Deoxycholate and 1 mM EDTA), and twice with TE Buffer (10 mM Tris-HCl, 1 mM EDTA). Immunoprecipitated complexes were eluted in 50 µL elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS) at 65˚C for 30 minutes. The eluate was transferred to a new 1.75 mL tube to which 120 µL of TE-SDS (10 mM Tris-HCl, 1 mM EDTA, 0.1% SDS) buffer was added and 90 µL TE-SDS was added to the 10 µL of input. Formaldehyde crosslinks were reversed overnight in a 65 °C water bath. Samples were purified using Qiagen PCR Clean protocol and eluted in 58 µL Buffer EB. qPCR on all samples and inputs was run using Promega GoTaq qPCR master

mix (A6001, Promega) using primers P21 and P22 on an Applied Biosystems StepOnePlus instrument. At least three independent biological replicates were run for each condition, with each biological replicate run in triplicate for qPCR quantitation. PCR amplification with primer P21 and P22 yielded a single band around 450 bps. This band was gel purified and submitted for sequencing. Results returned a sequence for the GAL1-LacZ gene, as expected (below).

Sequence of GAL1-LacZ PCR amplified product:

5'TGCATAANCCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTC
AAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAA
GGAGAAAAAACTATAATGACTAAATCTCATTCAGAAGAAGTGATTGTACCTGAGTTC
AATTCTAGCGCAAAGGAATTACCAAGACCATTGGCCGAAAAGTGCGGAATTCCAAGC
TTGGCCAAGCCCGGATCCGGAGCTTGGCTGTTGCCCGTCTCACTGGTGAAAAGAAAA
ACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAA
TGCAGCTGGCACGACAGGTTTCCCGACTTAATCGCCTTGCAGCACATCCCCCTTTCGC
CAGCTGGCGTAATAGCGAAGAGGCCCGCACCG 3'

Input CT values for a single biological replicate were averaged and log2(1/60) was subtracted from this value to generate a normalized Input value. The individual CT values generated for a set of technical triplicates were subtracted from the respective normalized input values to generate a Δ CT value (negative number) for each measurement. The Δ CT value was then log transformed (100*2^(Δ CT)) to generate individual % input values. The % input values generated across at least three biological replicates were averaged and the standard error of the mean calculated for graphical display. p-values were calculated using an unpaired t-test.

Tandem Reversible and Irreversible Crosslinking (TRIC) at GAL1-LacZ

An individual colony of pLexA+Gal4 Phe849Bpa-6HIS-3xFLAG was grown in 5 mL SC media containing 2% raffinose but lacking histidine, tryptophan and uracil for selection. The culture was incubated overnight at 30°C with agitation. Following incubation, this culture was used to inoculate 100 mL cultures of SC media containing 2% raffinose and 2% galactose. For Bpa incorporation, 1 mL 100 mM Bpa dissolved in 1M NaOH and 1 mL 1M HCl were added to the above cultures. The cultures were incubated overnight at 30°C with agitation to an OD660 of ~0.8. When cultures reached the appropriate OD660, the cells were treated in the following manner: The cultures receiving only UV treatment were centrifuged at 3901 rcf, 4˚C (Beckman Coulter Allegra X-22R), washed with SC media lacking histidine, tryptophan and uracil, and recentrifuged. Cell pellets were resuspended in 2mL SC media lacking histidine, tryptophan and

uracil and containing 2% raffinose, 2% galactose. Cells were transferred to a small cell culture dish and subjected to UV irradiation at 365 nm light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour. The cultures receiving formaldehyde treatment were equipped with a stirbar and placed on a stirplate in a fume hood. 2.9 mL of 37% formaldehyde was added directly to the culture and allowed to stir for 5 minutes before being quenched with 15 mL of 2 M Glycine for 5 minutes. Cells were then centrifuged and cell pellets were washed with 50 mL H-W-U- media and recentrifuged. Samples intended to additionally receive UV crosslinking were then resuspended in 2 mL H-W-U- media containing 2% raffinose, 2% galactose and transferred to a small cell culture dish and subjected to UV irradiation at 365 nm UV light (Eurosolar 15 W UV lamp with cooling for 0.5hour.

For lysis, cells were resuspended in 600 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini EDTA Free Protease Inhibitor (Roche) and lysed using glass beads and vortexing at 4°C. We found in these studies that complete cellular lysis is necessary to eliminate background signal caused by cell lysis during sonication. Subsequent lysates were removed and the remaining pellet was then washed twice with "Harsh" ChIP buffer (50 mM HEPES-KOH pH 7.5, 1M NaCl, 1 mM EDTA, 1% Triton X-100, 1% Na-Deoxycholate) followed by 2 washes with regular ChIP buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini EDTA Free Protease Inhibitor (Roche)). Pellets were resuspended in 600 µL ChIP lysis buffer containing protease inhibitor and sonicated at a setting of 10% (double-step microtip, Fisher Scientific Dismembrator Model 500) for 2 minutes on ice with 30 second pulses on/off. Samples were then centrifuged at 4˚C for 20 minutes at max speed. Soluble chromatin was then removed from the pellet and immunoprecipitated with Snf1 antibody (sc-15621, Santa Cruz Biotechnologies) for 2 hours, 4˚C. The protein bound to the antibody was isolated by incubation for 1 hour with ~50 µL of prewashed protein G magnetic Dynabeads (Life Technologies). After immunoprecipitation, the beads were washed 6 times with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -20˚C until elution. The crosslinked sample was eluted from the beads and formaldehyde crosslinks reversed by heating at 95˚C for 20 minutes in NuPAGE 4x LDS Sample buffer (Life Technologies) containing 250 mM DTT. Samples were run on a 4-20% Tris-Glycine TGX gel (Bio-Rad) and transferred to a PVDF

membrane. Western Blot analysis was carried out using a 1:1000 dilution of anti-FLAG (M2) antibody (Sigma) in 5% PBST.

References 3:

1. Thompson, A. D., Dugan, A., Gestwicki, J. E., and Mapp, A. K. (2012) Fine-tuning multiprotein complexes using small molecules, *ACS Chem Biol 7*, 1311-1320.
2. Paoletti, A., Parmely, T., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R., Conaway, J., Florens, L., and Washburn, M. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors, *Proc Natl Acad Sci USA 103*, 18928-18933.
3. Sikorski, T. W., and Buratowski, S. (2009) The basal initiation machinery: beyond the general transcription factors, *Curr Opin Cell Biol 21*, 344-351.
4. Vincent, O., Townley, R., Kuchin, S., and Carlson, M. (2001) Subcellular localization of the Snf1 kinase is regulated by specific beta subunits and a novel glucose signaling mechanism, *Genes & Development 15*, 1104-1114.
5. Chen, X. F., Lehmann, L., Lin, J. J., Vashisht, A., Schmidt, R., Ferrari, R., Huang, C., McKee, R., Mosley, A., Plath, K., Kurdistani, S. K., Wohlschlegel, J., and Carey, M. (2012) Mediator and SAGA have distinct roles in Pol II preinitiation complex assembly and function, *Cell Rep 2*, 1061-1067.
6. Lo, W.-S., Gamache, E. R., Henry, K. W., Yang, D., Pillus, L., and Berger, S. L. (2005) Histone H3 phosphorylation can promote TBP recruitment through distinct promoter-specific mechanisms, *EMBO J 24*, 997-1008.
7. Ptashne, M., and Gann, A. (1997) Transcriptional activation by recruitment, *Nature 386*, 569-577.
8. Vincent, O., and Carlson, M. (1999) Gal83 mediates the interaction of the Snf1 kinse complex with the transcriptional activator Sip4, *EMBO J 18*, 6672-6681.
9. Sikorski, T. W., Joo, Y. J., Ficarro, S. B., Askenazi, M., Buratowski, S., and Marto, J. A. (2012) Proteomic analysis demonstrates activator- and chromatin-specific recruitment to promoters, *Journal of Biological Chemistry 287*, 35397-35408.
10. Tsai, K.-L., Tomomori-Sato, C., Sato, S., Conaway, Ronald C., Conaway, Joan W., and Asturias, Francisco J. Subunit Architecture and Functional Modular Rearrangements of the Transcriptional Mediator Complex, *Cell 157*, 1430-1444.
11. Wands, A. M., Wang, N., Lum, J. K., Hsieh, J., Fierke, C. A., and Mapp, A. K. (2010) Transient-state kinetic analysis of transcriptional activator-DNA complexes interacting with a key coactivator, *Journal of Biological Chemistry 286*, 16238-16245.
12. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008) Photocrosslinkers illuminate interactions in living cells, *Molecular BioSystems 4*, 473-480.
13. Mapp, A. K., and Ansari, A. Z. (2007) A TAD further: exogenous control of gene activation, *ACS Chemical Biology 2*, 62-75.
14. Traven, A., Jelicic, B., and Sopta, M. (2006) Yeast Gal4: a transcriptional paradigm revisited, *EMBO Rep 7*, 496-499.
15. Ansari, A. Z. R., R.J.; Ptashne, M. (1998) A transcriptional activating region with two contrasting modes of protein interaction, *Proc. Natl. Acad. Sci. 95*, 13543-13548.

16. Majmudar, C. Y., Lee, L. W., Lancia, J. K., Nwokoye, A., Wang, Q., Wands, A. M., Wang, L., and Mapp, A. K. (2009) Impact of nonnatural amino acid mutagenesis on the in vivo function and binding modes of a transcriptional activator, *Journal of the American Chemical Society 131*, 14240-14242.
17. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science 301*, 964-967.
18. Schmidt, M. J., and Summerer, D. (2014) Genetic Code Expansion as a Tool to Study Regulatory Processes of Transcription, *Frontiers in Chemistry 2*.
19. Zaman, Z., Ansari, A. Z., Koh, S. S., Young, R., and Ptashne, M. (2001) Interaction of a transcriptional repressor with the RNA polymerase II holoenzyme plays a crucial role in repression, *Proc Natl Acad Sci USA 98*, 2550-2554.
20. Johnston, M., Flick, J. S., and Pexton, T. (1994) Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae, *Mol Cell Biol 14*, 3834-3841.
21. Martin, B. R., and Cravatt, B. F. (2009) Large-scale profiling of protein palmitoylation in mammalian cells, *Nature Methods 6*, 135-138.
22. Floer, M., Bryant, G. O., and Ptashne, M. (2008) HSP90/70 chaperones are required for rapid nucleosome removal upon induction of the GAL genes of yeast, *Proc Natl Acad Sci USA 105*, 2975-2980.
23. Young, E. T., Zhang, C., Shokat, K. M., Parua, P. K., and Braun, K. A. (2012) The AMP-activated protein kinase Snf1 regulates transcription factor binding, RNA polymerase II activity, and mRNA stability of glucose-repressed genes in Saccharomyces cerevisiae, *J Biol Chem 287*, 29021-29034.
24. Papamichos-Chronakis, M., Gligoris, T., and Tzamarias, D. (2004) The Snf1 kinase controls glucose repression in yeast by modulating interactions between the Mig1 repressor and the Cyc8-Tup1 co-repressor, *EMBO Reports 5*, 368-372.
25. Hierlmeier, T., Merl, J., Sauert, M., Perez-Fernandez, J., Schultz, P., Bruckmann, A., Hamperi, S., Ohmayer, U., Rachel, R., Jacob, A., Hergert, K., Deutzmann, R., Griesenbeck, J., Hurt, E., Milkereit, P., Bassler, J., and Tschochner, H. (2013) Rrp5, Noc1p and Noc2p form a protein module which is part of early large ribosomal subunit precursors in S. cerevisiae, *Nucl Acids Res 41*, 1191-1210.
26. Hedbacker, K., and Carlson, M. (2006) Regulation of the Nucleocytoplasmic Distribution of Snf1-Gal83 Protein Kinase, *Euk Cell 5*, 1950-1956.
27. Bertram, P. G., Choi, J. H., Carvalho, J., Chan, T. F., Ai, W., and Zheng, X. F. (2002) Convergence of TOR-nitrogen and Snf1-glucose signaling pathways onto Gln3, *Mol Cell Biol 22*, 1246-1252.
28. Mitchell, S. F., Jain, S., She, M., and Parker, R. (2012) Global analysis of yeast mRNPs, *Nat Struct Mol Biol 20*, 127-133.
29. Shashkova, S., Welkenhuysen, N., and Hohmann, S. (2015) Molecular communication: crosstalk between the Snf1 and other signaling pathways, *FEMS Yeast Res 15*, fov026.
30. Flick, J. S., and Johnston, M. (1990) Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae, *Molecular Cell Biology 10*, 4757-4769.
31. Treitel, M. A., Kuchin, S., and Carlson, M. (1998) Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in Saccharomyces cerevisiae, *Mol Cell Biol 18*, 6273-6280.

32. Ostling, J., and Ronne, H. (1998) Negative control of the Mig1p repressor by Snf1p-dependent phosphorylation in the absence of glucose, *Eur J Biochem 252*, 162-168.
33. Hall, D. B., and Struhl, K. (2002) The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo, *Journal of Biological Chemistry 277*, 46043-46050.
34. Woods, A., Munday, M. R., Scott, J., Yang, X., Calson, M., and Carling, D. (1994) Yeast Snf1 is functionally related to mammalian AMP-activating protein kinase and regulates acetyl-CoA carboxylase in vivo, *J Biol Chem 269*, 19509-19515.
35. Krishnamurthy, M., Dugan, A., Nwokoye, A., Fung, Y. H., Lancia, J. K., Majmudar, C. Y., and Mapp, A. K. (2011) Caught in the act: covalent cross-linking captures activator-coactivator interactions in vivo, *ACS Chemical Biology 6*, 1321-1326.
36. Hardie, D. G. (2007) AMP-activated SNF1 protein kinases: conserved guardians of cellular energy, *Nat Rev Mol Cell Biol 8*, 774-785.
37. Dechassa, M. L., Zhang, B., Horowitz-Scherer, R., Persinger, J., Woodcock, C. L., Peterson, C. L., and Bartholomew, B. (2008) Architecture of the SWI/SNF-nucleosome complex, *Mol Cell Biol 28*, 6010-6021.
38. Falsone, S. F., Gesslbauer, B., Tirk, F., Piccinini, A. M., and Kungl, A. J. (2005) A proteomic snapshot of the human heat shock protein 90 interactome, *FEBS Lett 579*, 6350-6354.
39. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast, *Nature 425*, 737-741.
40. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014) Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells, *Nat Methods 11*, 319-324.
41. Majmudar, C. Y., Wang, B., Lum, J. K., Hakansson, K., and Mapp, A. K. (2009) A high-resolution interaction map of three transcriptional activation domains with a key coactivator from photo-cross-linking and multiplexed mass spectrometry, *Angew Chemie Int Ed 48*, 7021-7024.
42. Lancia, J. K., Nwokoye, A., Dugan, A., Joiner, C., Pricer, R., and Mapp, A. K. (2014) Sequence context and crosslinking mechanism affect the efficiency of in vivo capture of a protein-protein interaction, *Biopolymers 101*, 391-397.
43. Hulce, J. J., Cognetta, A. B., Niphakis, M. J., Tully, S. E., and Cravatt, B. F. (2013) Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells, *Nat Meth 10*, 259-264.
44. Mann, M. (2006) Functional and quantitative proteomics using SILAC, *Nat Rev Mol Cell Biol 7*, 952-958.
45. Preston, G. W., Radford, S. E., Ashcroft, A. E., and Wilson, A. J. (2014) Analysis of amyloid nanostructures using photo-cross-linking: In situ comparison of three widely used photo-cross-linkers, *ACS Chem Biol 9*, 761-768.
46. Faubert, B., Boily, G., Izreig, S., Griss, T., Samborska, B., Dong, Z., Dupuy, F., Chambers, C., Fuerth, B. J., Viollet, B., Mamer, O. A., Avizonis, D., DeBerardinis, R. J., Siegel, P. M., and Jones, R. G. (2013) AMPK is a negative regulator of the Warburg effect and suppresses tumor growth in vivo, *Cell Metab 17*, 113-124.
47. Liang, J., and Mills, G. B. (2013) AMPK: A contextual oncogene or tumor suppressor?, *Cancer Res 73*, 2929-2935.

48. Viollet, B., Lantier, L., Devin-Leclerc, J., Hebrard, S., Amouyal, C., Mounier, R., Foretz, M., and Andreelli, F. (2009) Targeting the AMPK pathway for the treatment of type 2 diabetes, *Front Biosci 14*, 3380-3400.
49. Gietz, R. D., and Schiestl, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method, *Nature Protocols 2*, 31-34.

Chapter 4: Impact of Genetically Incorporated Photocrosslinker on PPI Capture Efficiency4

4A: Abstract

The genetic incorporation of photocrosslinking amino acids has been applied to the study of difficult PPIs, but we recently found that this method did not capture a PPI for which there was abundant functional evidence. The factors that influence the success of covalent chemical capture and the innate reactivity of the two UAAs utilized, *p*-benzoyl-L-phenylalanine (Bpa) and *p*-azido-L-phenylalanine (Azpa), plays a profound role in the capture of protein binding partners. Based upon these data, we propose the first guidelines for the successful use of *in vivo* photocrosslinking to capture novel PPIs and to characterize the interfaces has been developed.

4B: Background

Photocrosslinking to covalently chemically capture proteins interacting with one another has increased in popularity. Photoreactive groups have been seen in everything from bifunctional ligand molecules to genetically incorporated unnatural amino acids. Although photocrosslinking is not a new technique, its application to *in vivo* experiments is very recent. One consequence of this relatively new application is a shortage of data indicating crosslinker efficiencies and reactivity preferences within a cellular environment. The enormous difference in the cellular environment and a controlled test tube reaction means that although a foundation is in place for understanding crosslinking moieties and their mechanisms, we still need to explore the reasons behind differences in the crosslinking efficiencies.

Earlier work from our lab stimulated interest in exploring some of the reactivity considerations of genetically incorporated p-benzoyl-l-phenylalanine (Bpa) along the interface of a classical PPI. Using the transcriptional activator Gal4 and its repressor protein Gal80 in *Saccharomyces cerevisiae*, Bpa was incorporated into 10 different positions within the transcriptional activation domain of Gal4 and photocrosslinking was carried out in living yeast

[4] Portions of this chapter are from published work: Lanacia, J.K.; Nwokoye, A.; Dugan, A.; Joiner, C.; Pricer, R.; Mapp, A.K., Sequence context and crosslinking mechanism affect the efficiency of *in vivo* capture of a protein-protein interaction. *Biopolymers* **2014**, 101 (4), 391-397. Figure 4-1 and Figure 4-2 are from this work and I contributed all data to these figures. Work from section 4D was done in conjunction with Melanine Cheung-See-Kit.

under conditions in which Gal80 was expected to bind Gal4. While complex formation via *in vivo* crosslinking was detected at most positions, several instances were noted where functional data and existing structural data for the Gal4–Gal80 complex supported a direct interaction, yet little to no crosslinking was observed.[1-6] For example, Phe856 has been observed in structural studies to be buried within the Gal4-Gal80 binding interface.[5, 6] When Phe856 was replaced with Bpa, the resulting Gal4 mutant was repressed in the presence of Gal80, suggesting that the mutation did not negatively impact the Gal80 binding interaction, yet no Gal4–Gal80 covalent complex could be observed after crosslinking. In the case of the Phe856Bpa Gal4 mutant, there was functional evidence excluding the lack of a binding interaction as an explanation. Thus, other parameters were examined that would produce a negative result, focusing on the crosslinking mechanism of Bpa in addition to a second photocrosslinking UAA, *p*-azido-L-phenylalanine (Azpa).

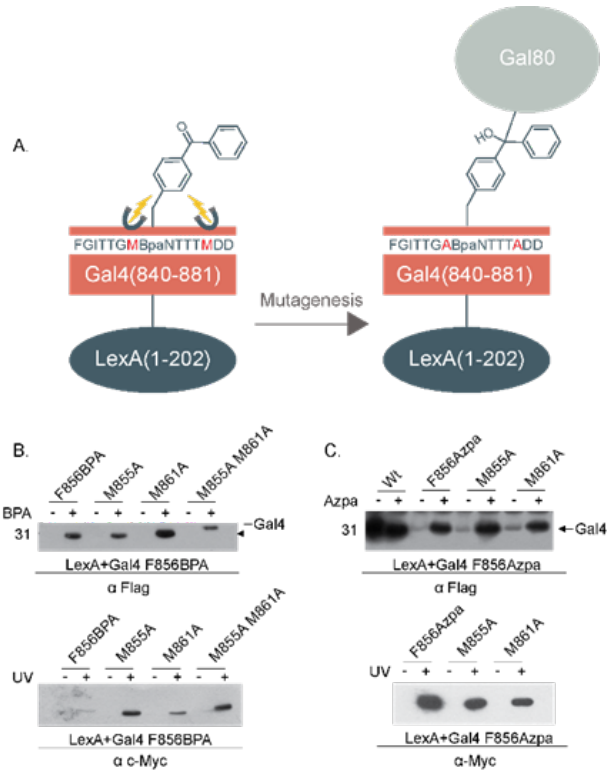


Figure 4-1 The Effect of Methionine on LexA+Gal4 F856Bpa Crosslinking to Gal80.
(a) Scheme showing experiment results. (b) Expression of LexA+Gal4 F856Bpa as detected by Western blot of cell lysates with α-Flag and crosslinking of LexA+Gal4 F856Bpa to Myc-tagged Gal80 as shown by Western blot of cell lysates with α-Myc antibody. (c) Expression of LexA+Gal4 F856Azpa as detected by Western blot of cell lysates with α-Flag antibody and crosslinking of LexA+Gal4 F856Azpa to Myc-Gal80 as shown by Western blot of cell lysates with α-Myc antibody.

In addition to the exploration of this very specific case of an apparent crosslinking false negative, a larger surface area PPI between Gal80-Gal80 dimers was examined. Looking at an interface differing in size and affinity gives an interesting comparison to the Gal4-Gal80 binding interface which is small and very high affinity. We wanted to see if there were particular interface types that are more amenable to Bpa or Azpa to enable easier application of this method to less well-studied PPIs.

4C: Covalent Capture of Gal4-Gal80 Varies Between Bpa and Azpa

To delve further into the unexpected lack of a crosslinked adduct at the Gal4 interface with Gal80, we first began examining the potential mechanistic causes. Bpa forms a diradical

upon UV irradiation at 350–365 nm and then undergoes a C-H insertion reaction with nearby backbones and amino acid side chains.[7] Although Bpa is capable of inserting into most C-H bonds, experimentally, Bpa reacts preferentially with methionine (Met) where it will react at distances beyond the 3.1 Å reactivity radius.[8] Specifically, the apparent preference of Bpa for methionine suggests that Bpa efficiency can be altered dramatically when placed in close proximity to methionine's thioether side chain.[8-10] Further examination of position 856 in the Gal4 TAD reveals two methionines in close proximity to the Bpa side chain; thus, we hypothesized that these methionines at positions 855 and 861 are internally "quenching" Bpa, thereby preventing it from crosslinking to Gal80. Consistent with this hypothesis, when Met855 and Met861 are mutated either individually or collectively to alanine, we see that the resulting mutants are functional and, importantly, that Bpa crosslinking to Gal80 is restored (Figure 4-1). These data are consistent with a model in which an intramolecular crosslink was competing with the intermolecular reaction in the LexA1Gal4 Phe856Bpa mutant and led to a false negative in our original experiments.

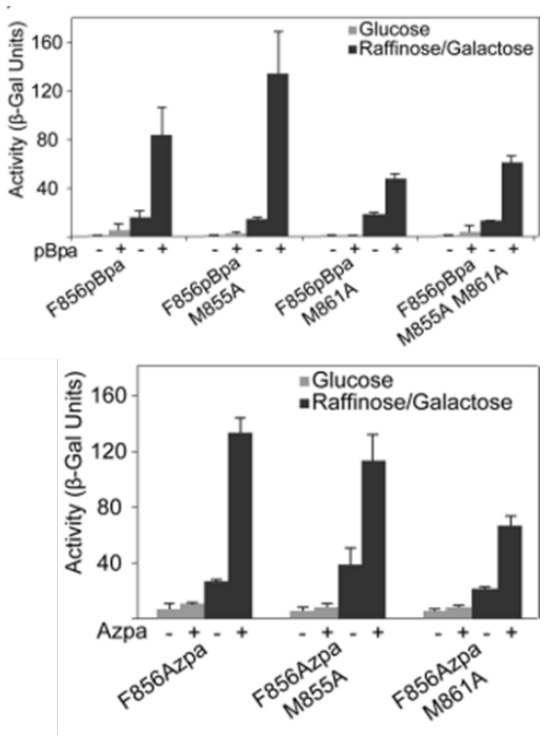


Figure 4-2 Activation Potential by Liquid β-Galactosidase Assays of Gal4 Mutants.
In the yeast strain tested, β-galactose expression was controlled by a Gal1 promoter containing two LexA binding sites for LexA+Gal4 binding. Each value is the average of at least three independent experiments with the indicated error (SDOM). As displayed above, all constructs used were able to activate transcription.

To test this hypothesis further, a second photoactivatable UAA, Azpa, was incorporated into Gal4 in *S. cerevisiae* using the nonsense suppression method. The amino acid preference for Azpa crosslinking is less clear because it has a more complex crosslinking mechanism compared to Bpa.[7] During excitation at 254 nm light it forms a nitrene and it is in this state that insertion into C-H or heteroatom-H bonds occurs. If, however, insertion does not take place during the 1024 second excitation lifetime (determined for simple nitrenes in a polystyrene matrix) it will rearrange to a more stable ketenimine which only reacts with nucleophiles.[7, 11] Utilization of the same expression conditions for Bpa led to the incorporation of Azpa at position 856 and the

resulting mutant was fully functional (Figure 4-2, Figure 4-1). A direct comparison of Azpa and Bpa crosslinking at position 856 reveals that Azpa crosslinks readily with Gal80, whereas Bpa does not (Figure 4-1). As expected, introduction of alanine at positions 855 and 861 yielded no changes in Azpa crosslinking, consistent with the reactivity profile of this amino acid. These data indicate that the difference in reactive mechanisms of crosslinkers plays a critical role in the outcome of crosslinking experiments. The results shown here illustrate that not only is optimization of UAA incorporation a key factor in successful application of the strategy, but careful consideration of the innate reactivity of the UAA utilized is also critical.

4D: Covalent Capture Efficiency of Varied PPI Interfaces Comparing Bpa and Azpa

Although these studies illustrated a difference in reactivity between Bpa and Azpa, a complete analysis of differences in binding partner capture ability between incorporated amino acids on different PPI interfaces has yet to be done. To understand these preferences such that the proper crosslinker can be used with the interactions under study, a series of PPIs that span a range of affinities and surface area must be compared. The Gal4-Gal80 interaction used to test the methionine preferences of Bpa has a native K_D is 5 nM with a surface area of approximately 1500 Å^2(Figure 4-3). To look at a larger surface area interface, the Gal80-Gal80 dimerization domain was chosen with a K_D of 1 nM and surface area of 5000 Å^2. In both systems, a series of mutations can be made to access a range of lowered binding affinities.

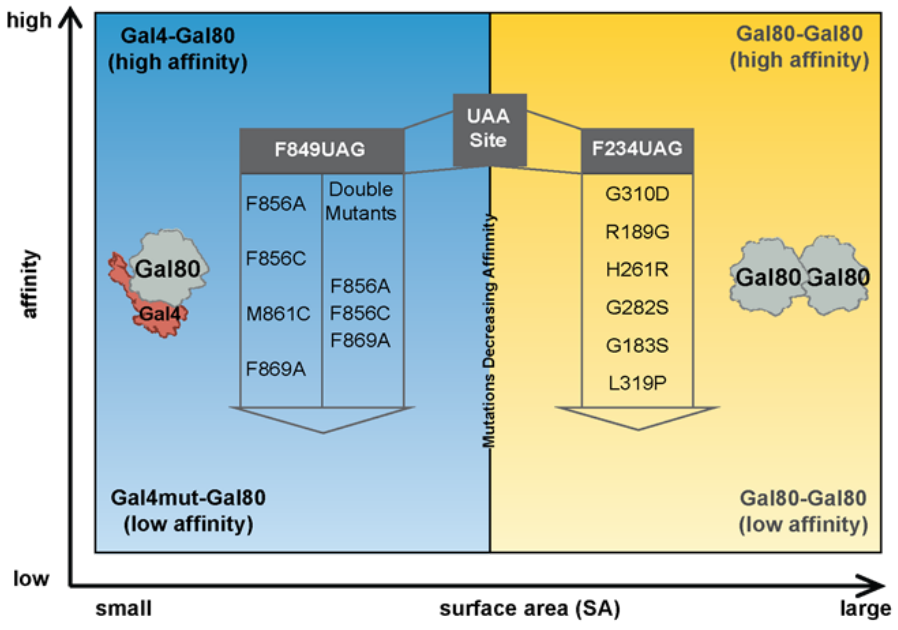


Figure 4-3 Interactions with Potential Mutants to Highlight a Range of PPI Interfaces.
The first model PPI is the high affinity/small surface area complex of Gal4 and its repressor protein Gal80. Mutations introduced into Gal4 will be used to progressively weaken this complex to move this into the lower light blue quadrant. The second model PPI, fitting into the strong and broad category of interactions, utilizes the Gal80-Gal80 dimer with Gal80 mutations listed to create weaker binding.

Mutations for the Gal4-Gal80 interface were selected based on

previous literature citing a dramatic decrease in Gal4 binding to Gal80. These mutants were then run as a panel on a preliminary Western blot to see which mutants had the most effect on binding. Contrary to literature, we were still able to capture Gal80 with the single mutations (Figure 4-4). Although surprising, when looking at the context of previous literature experiments it is easier to understand the potential differences. To express and test the binding affinities, proteins are often truncated so the native environment with full-length protein will have a much higher binding affinity and thus be more difficult to attenuate. A panel of double mutants will need to be made to attempt to attenuate binding of Gal80 such that a noticeable difference can be observed by Western Blot.

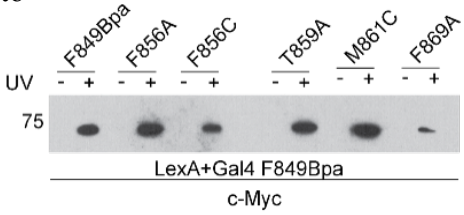


Figure 4-4 Crosslinking of LexA+Gal4 F849Bpa mutants with Gal80.
In vivo covalent chemical capture using LexA+Gal4 F849Bpa and the panel of Gal4 mutants was carried out. An α-LexA antibody was used for immunoprecipitation and the blot was probed with a c-Myc-HRP antibody. The only mutant that appears to cause weaker binding may be F869A but this is not a quantitative blot so no firm conclusion can be made. No loss of binding was observed as expected, however.

Another observation made during the execution of these experiments, is that the traditional Western blotting method used with chemiluminescence is not quantifiable (Figure 4-4). Extensive optimization went into choosing antibodies for multiplex fluorescent Western blotting. Although the yeast model is a simpler system, not many companies produce yeast antibodies making antibody selection a much more difficult process. The first antibody pairs found to successfully detect both Gal4 and Gal80 protein were anti-Flag and Myc primary antibodies with rabbit and mouse Alexafluor conjugated secondaries. Although Alexafluor gave quantifiable detection, a significant background was observed, which may interfere with visualization and calculations (Figure 4-5). The pair of antibodies that produced the best signal with minimal background was

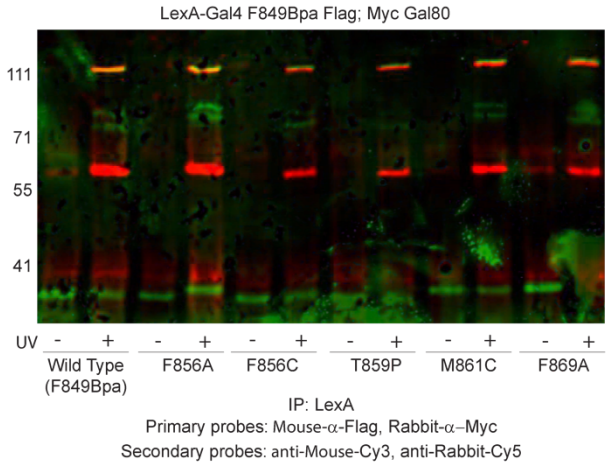


Figure 4-5 Duplex Visualization of Gal4 Mutant Attenuation of Gal4-Gal80 Interaction.
Single point mutants within the Gal4 TAD were examined for their effects on Myc-Gal80 binding compared to the wild-type LexA-Gal4 849Bpa-Flag construct. These mutants were able to attenuate binding to Gal4 endogenous binding partners, but had little effect on the Gal4-Gal80 binding event.

found to be secondaries conjugated to near-IR dyes with 680 and 800 nm excitations (Figure 4-5).

Using the optimized antibodies, we tested several Gal80 mutants proposed to decrease dimerization (Figure 4-6).[12, 13] We found that overall Apza had poorer capture of the Gal80 dimer then Bpa (Figure 4-7). This is surprising, as the main defect of Azpa is thought to be crosslinking of accessible solvent but the site of incorporation F234 is proposed to be fully buried in the dimer interface using Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute.[14]

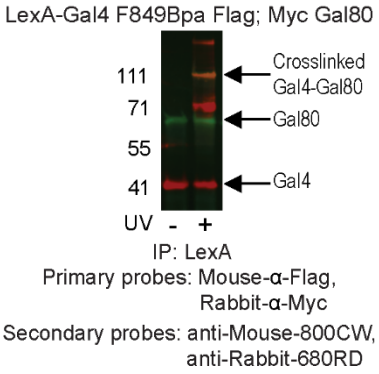


Figure 4-5 Visualization of Gal4-Gal80 with near-IR Antibodies.
The near-IR secondary antibodies at 680 and 800 nm showed the best specificity and signal.

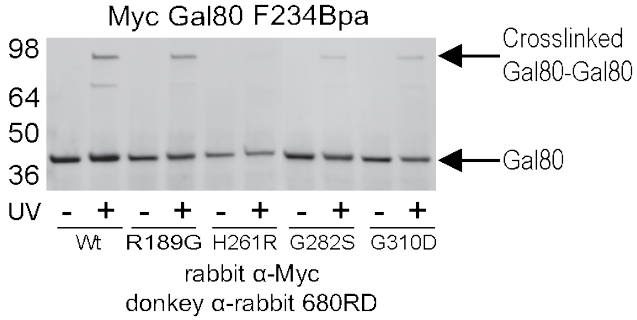


Figure 4-6 Visualization of Gal80-Gal80 Mutations Attenuating Binding.
Gal80 containing Bpa was visualized to see which mutations attenuated dimer binding the most.

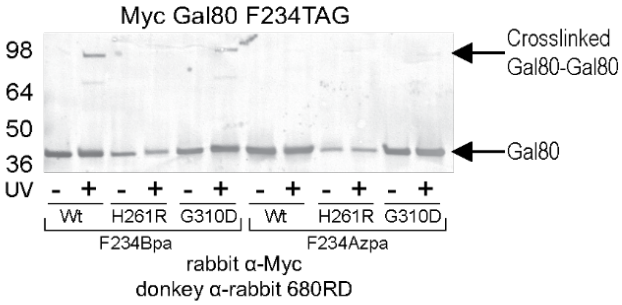


Figure 4-7 Gal80-Gal80 Crosslinking Comparing Bpa and Azpa.
Mutations selected to attenuate Gal80 binding were compared between Bpa and Azpa crosslinkers. We can see bands for Bpa but extremely faint bands for Azpa.

4E: Conclusion

Although the Gal4-Gal80 PPI that is the focus of this study is a well-characterized complex, the present and future applications of *in vivo* covalent capture are in the discovery of novel PPIs. Thus, it is important to understand the meaning behind a negative result. An absence of crosslinking adduct could arise from a variety of factors, including the lack of a binding interaction, low unnatural amino acid (UAA) incorporation yield and/or fidelity, poor positioning of the UAA, and the poor reactivity of the activated UAA with the amino acids in the binding partner. To avoid false negatives, it is critical to carry out crosslinking with more than a single UAA mutant, since, as illustrated here and in a previous study, a small change in position can have a dramatic effect on crosslinking.

When selecting an amino acid for a particular experiment, the longer lifetime and lower reactivity toward solvent makes Bpa an attractive choice, particularly for PPIs that occur through shallow, exposed binding sites. However, the marked preference of Bpa for methionine raises some concern that crosslinking results could be influenced by the presence of methionine in the UAA-containing protein as shown here with Gal4 or by a lack of methionines in potential binding partners.[15] Relevant binding partners could be missed in an unbiased study due to either of these factors.

The panel of Gal80-Gal80 mutants tested presents a contrary view of Azpa which theoretically should excel in a buried site such as F234. Looking closer at the Gal80 crystal structure, it is plausible that this mutation along with some of the others we selected for this study could drastically change the structure. If this were the case, Azpa may be more solvent accessible than predicted resulting in non-specific crosslinking with solvent. Further studies will need to be done to look at potential structural effects these mutations have on Gal80 as well as the creation and testing of double mutants within the Gal4 binding interface to enable a direct comparison between small and large PPIs over a spectrum of affinities. Additional interfaces will also need to be examined to see if any trends are maintained over a wider variety of PPIs. If trends for photocrosslinking amino acids use can be established, this will have widespread applications for laboratories interested in using this technology, as the selection of a crosslinking amino acid could thus be tailored to the PPI being studied.

4F:Experimental

All covalent chemical capture experiments were carried out in yeast strain LS41 [JPY9::pZZ41, Mat his3Δ200 leu2Δ1 trp1Δ63 ura3-52 lys2Δ385 gal4 URA::pZZ41]. Bpa and Azpa was purchased from Chem-Impex International (Wood Dale, IL). All plasmids used in this study were constructed using standard molecular biology techniques. Sanger sequencing verifying plasmid sequences was performed by the University of Michigan Core Facility (Ann Arbor, MI).

Plasmid Construction:

All plasmids used were previously reported. All mutations were made using site-directed mutagenesis (Qiagen Quikchange Protocol). Insertion of the 6xHis tag was achieved by site-directed mutagenesis and Qiagen Quikchange two-step SDM protocol.

***In vivo* Covalent Chemical Capture**

In vivo covalent chemical capture experiments was carried out as previously described.[1] To perform *in vivo* crosslinking, an individual colony expressing pLexA+Gal4 F849TAG-Flag and pGADT7-Myc-Gal80 was inoculated in 10 mL SC media containing 2% glucose and lacking histidine, tryptophan, leucine, and uracil (H-W-L-U-) for selection. This culture was incubated overnight at 30 °C with agitation and used to inoculate 100 mL cultures of H-W-L-U- SC media containing 2% glucose. To each 100 mL culture, 1 mL of 100 mM Bpa dissolved in 1M NaOH and 1 mL 1M HCl were added. Cultures were incubated overnight at 30 °C with agitation and grown to mid-log phase OD660 (~0.8). Yeast were harvested by centrifugation at 6000 rcf at 4˚C for 15 minutes (Beckman *Ava*nti J-201, JLA 8.1000). Cell pellets were washed with SC media and resuspended in 2mL H-W-L-U- SC media with 2% glucose. Yeast were transferred to cell culture dishes and subjected to UV irradiation at 365 nm light (Eurosolar 15 W UV lamp) with cooling for 0.5 hour. Irradiated cells were then harvested by centrifugation and stored at -80˚C until lysis.

For lysis, cells were resuspended in 600 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate and 2X Complete Mini, EDTA Free Protease Inhibitor (Roche) and lysed using glass beads by bead beating at 2800 rpms for 10 minutes at 4°C. Subsequently, the lysate was pelleted and the supernatant incubated with 2 µg of LexA antibody (sc-1725, Santa Cruz Biotechnologies) for 2 hours at 4°C for immunoprecipitation. The protein bound to the antibody was isolated by incubation for 1 hour with 40 µL of prewashed protein G magnetic beads slurry (Dynal Corporation, Invitrogen, Carlsbad, CA). The beads were washed 6 times with 1 mL Wash Buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.1% Na-Deoxycholate and 1 mM EDTA) and stored dry at -80 ˚C until elution.

Preparation and Western Blotting of Crosslinked Samples

The crosslinked sample was eluted from the beads by heating at 95˚C for 10 minutes in NuPAGE 4x LDS Sample buffer (Invitrogen, Carlsbad, CA) containing 250 mM DTT. Protein amounts were measure using the Pierce 660 Assay so equal concentrations of protein could be loaded per lane. Samples were run on a 4-8% Tris Acetate gel and transferred to PDVF fluorescent membrane. Blots were probed using a 1:500 dilution of mouse-anti-FLAG (M2) antibody (Sigma, St. Louis, MO) and rabbit-anti-Myc (Pierce) in 5% BSA PBST. The secondary

detection antibodies used were a 1:10,000 dilution of donkey anti-mouse 680LT IRDye and donkey anti-rabbit 800CW IRDye (Lycore). Blots were visualized using an Azure C600 Imaging System.

β-Galactosidase assays

To evaluate the ability of each LexA+Gal4 F849Bpa-Flag mutants to activate transcription in the presence or absence of glucose, saturated cultures (SC media + 2% raffinose) of yeast expressing Gal4 Phe849Bpa mutants were used to inoculate 5 mL SC media containing either 2% glucose or 2% raffinose & 2% galactose but lacking histidine and tryptophan for selection. The cells were grown to an OD660 of 0.8-1 and harvested. The activity of each construct was monitored using β-galactosidase assays as previously described.[1]

References Chapter 4:

1. Majmudar, C. Y., Lee, L. W., Lancia, J. K., Nwokoye, A., Wang, Q., Wands, A. M., Wang, L., and Mapp, A. K. (2009) Impact of nonnatural amino acid mutagenesis on the in vivo function and binding modes of a transcriptional activator, *Journal of the American Chemical Society 131*, 14240-14242.
2. Ansari, A. Z. R., R.J.; Ptashne, M. (1998) A transcriptional activating region with two contrasting modes of protein interaction, *Proc. Natl. Acad. Sci. 95*, 13543-13548.
3. Ma, J., and Ptashne, M. (1987) The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell 50*, 137-142.
4. Lohr, D. V., P.; Zlantanova, J. (1995) Transcriptional regulation in yeast GAL gene family: a complex genetic network, *FASEB 9*, 777-787.
5. Thoden, J. B., Ryan, L. A., Reece, R. J., and Holden, H. M. (2008) The interaction between an acidic transcriptional activator and its inhibitor. The molecular basis of Gal4p recognition by Gal80p, *J Biol Chem 283*, 30266-30272.
6. Kumar, P. R. Y., Y.; Sternglanz, R.; Johnston, S.A.; Joshua-Tor, L. (2008) NADP regulates the yeast GAL induction system, *Science 319*, 1090-1092.
7. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008) Photocrosslinkers illuminate interactions in living cells, *Molecular BioSystems 4*, 473-480.
8. Wittelsberger, A., Thomas, B. E., Mierke, D. F., and Rosenblatt, M. (2006) Methionine acts as a "magnet" in photoaffinity crosslinking experiments, *FEBS Letters 580*, 1872-1876.
9. Rihakova, L., Deraët, M., Auger-Messier, M., Pérodin, J., Boucard, A. A., Guillemette, G., Leduc, R., Lavigne, P., and Escher, E. (2002) METHIONINE PROXIMITY ASSAY, A NOVEL METHOD FOR EXPLORING PEPTIDE LIGAND–RECEPTOR INTERACTION, *Journal of Receptors and Signal Transduction 22*, 297-313.
10. Wittelsberger, A., Mierke, D. F., and Rosenblatt, M. (2008) Mapping Ligand–receptor Interfaces: Approaching the Resolution Limit of Benzophenone-based Photoaffinity Scanning, *Chemical Biology & Drug Design 71*, 380-383.

11. Knowles, J. R. (1972) Photogenerated reagents for biological receptor-site labeling, *Accounts of Chemical Research 5*, 155-160.
12. Melcher, K. (2005) Mutational Hypersensitivity of a Gene Regulatory Protein: Saccharomyces cerevisiae Gal80p, *Genetics 171*, 469-476.
13. Pilauri, V., Bewley, M., Diep, C., and Hopper, J. (2005) Gal80 Dimerization and the Yeast GAL Gene Switch, *Genetics 169*, 1903-1914.
14. Krissinel, E. H., K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.
15. Ofran, Y., and Rost, B. (2003) Analysing Six Types of Protein–Protein Interfaces, *Journal of Molecular Biology 325*, 377-387.

Chapter 5: Conclusion

5A: Introduction

Protein-protein interactions support all cellular functioning and execute everything from signaling pathways to ensuring the correct proteins localize to the proper locations within a cell. Because of their extremely critical nature, any dysregulation of PPIs can quickly drive a cell into a disease state. The vast span of PPIs and their myriad of roles in disease make them logical points for drug intervention. This has not been the reality for many PPIs, however. Compared to traditional enzymes or enzyme-like PPI that have been successful drug targets, many PPIs have much larger surface areas with no clear binding pockets or hot-spots of inhibition. This can often be compounded with the fact that many PPIs have dynamic structures allowing them to bind multiple binding partners sometimes with transient affinities. Due to these complications more commonly used biochemical techniques to study proteins may miss important PPIs altogether or provide an incomplete picture of their binding network. Therefore, methods that look at PPIs in their native environments with minimal perturbation to their structures/function are needed to form more complete pictures of PPI networks and their potential for therapeutic interaction.

5B: In Vivo Covalent Chemical Capture Expands PPI Studies

One such method employed throughout this thesis is the use of genetically incorporated photocrosslinking unnatural amino acids for the covalent chemical capture of PPIs. This technique site-specifically incorporates a UAA, causing minimal perturbation to the protein and allowing it to still function as intended within the cell. This method was used to specifically look at the PPIs of transcriptional activators, as their intrinsically disordered nature, multi-protein complex recruitment, and transient associations make them a difficult to investigate by traditional means.

In Chapters 2-3 covalent chemical capture was combined with other techniques to illustrate not only the power of using genetically incorporated UAAs for PPI studies but also the ability to fuse with other techniques to answer specific questions. In Chapter 2, the question being asked, how to we look at direct interactions of DNA-bound proteins, stemmed from a

long-time debate of general transcription factor TBP's recruitment to a the GAL1 promoter. We were able to resolve this question using TRIC, a method to reversibly formaldehyde crosslink followed by covalent chemical capture to isolate DNA-bound PPIs. This led us to show direct binding of DNA-bound VP16 to TBP indicating the activator also plays a role in TBP recruitment. Although TRIC was developed in yeast, its application could extend to other systems especially since the more difficult lysis of yeast provided substantial difficulties for method development initially. It's use with a mammalian system could be further optimized to include a dual isolation of DNA and protein. This would be an exciting use of this methodology as it would provide a way to visualize direct DNA-bound interactions as well as split to a ChIP-like experiment validating the promoter localization.

In Chapter 3 we questioned the ability to isolate novel binding partners without a bias search by Western blot. MuDPIT mass spectrometry with covalent chemical capture was utilized and yielded the novel Gal4 binding partners Snf1 and Gal83 of the Snf1/AMPK complex. Furthermore, these discovered binding partners demonstrate the participation of transcriptional activators in complex recruitment beyond what was traditionally believed and broaden the perspective of transcriptional PPIs. Although demonstrated with a yeast model, this study also points to the use of the Snf1/APMK complex as a potential place for therapeutic intervention and additional studies to elucidate the complete involvement of AMPK in transcription. The mammalian version of this complex exists in many forms with varying subunits, potentially providing even further specificity for a drug target as our developed UAA and mass spectrometry technique could be employed to look at which subunits are part of transcriptionally related interactions.[1, 2] This would allow the targeting of specific gene activation instead of interfering with the global action of this multipurpose complex.

In Chapter 4, we began to think more analytically about covalent chemical capture and explored the potential for specific uses for individual UAAs. Several guidelines emerged from these studies including the need for testing experiments with multiple UAAs as the benzophenone tested has a high methionine preference sometimes leading to false negatives and the azide used can have lowered reactivity in certain interfaces. A quick check, if possible for methionines near the incorporation site to prevent intramolecular crosslinking as well as on the binding partner interface would be largely beneficial to ensure successful Bpa crosslinking. The specific reasons for lowered azide reactivity need to be further studied as the instance shown in

chapter 4 is the first example we have encountered a complete lack of crosslinking. Although additional work needs to be done to establish more firm guidelines, throughout this thesis I have demonstrated the power of covalent chemical capture as a technique for PPI studies, especially those that are more challenging to study by traditional means. It begs for us to ask, what questions do we want to answer next?

5C: Future Directions

Applications of the genetic incorporation of UAAs can extend far beyond the discussed covalent chemical capture methods. New UAAs that are capable of being incorporated into cells are constantly being added with new and exciting applications with over 100 different UAAs that have been successfully incorporated. Incorporation has also expanded to easily include mammalian cells. The work discussed in this thesis was done using a simplified yeast model, but switching to using mammalian cell line models would greatly increase the applicability of this work directly to human disease PPIs. Yeast were chosen for their slightly simpler PPIs networks and relative ease of use but using a mammalian system will greatly increase the number and variations of PPI networks. Understanding mammalian PPI networks will take considerable effort and time to decipher.

Looking even further into the future, past understanding individual cellular PPI networks, I think a movement to looking at the biological system as a whole will be possible using UAA incorporation. The ability to expand UAA incorporation to whole model organism such as *Caenorhabditis elegans* (worms) and *Drosophila melanogaster* (fruit flies) has already been demonstrated.[3, 4] Ideal additions to this set of tools would be the ability to incorporate UAAs within mouse and zebra fish models as well. The capability to probe PPIs over larger biological systems such as organs or even throughout the body would be a large leap in the field's ability to relate to whole-organism biology.

Multi-addition systems are becoming more practical to use as well.[5-7] This has been accomplished through the use of the amber stop codon with an additional sense codon rarely used or the addition of a blank quadruplet codon. In the future, one could envision this applying to PPI studies through the incorporation of a photocrosslinking amino acid for covalent chemical

capture, followed by an exposed UAA incorporating a bioorthogonal moiety such as a click group to allow extra chemical steps to be done like pull-downs or fluorescent tagging (Figure 5-1).

UAAs have other interesting possibilities for the manipulation of post-translational modification signaling within a cell. For example, photo-control over protein phosphorylation has been established with tyrosine (for use in receptors)[8], and serine (often part of signaling cascades).[9] Applications of this type of signaling and cellular control could be interesting especially through the observation of PPI changes mediated by the turning on of specific pathways. Examining a pathway within the cell in this fashion may also provide information about overlooked proteins within a signaling cascade/pathway such as scaffolding proteins or other critical binding partners we do not normally study through traditional means.

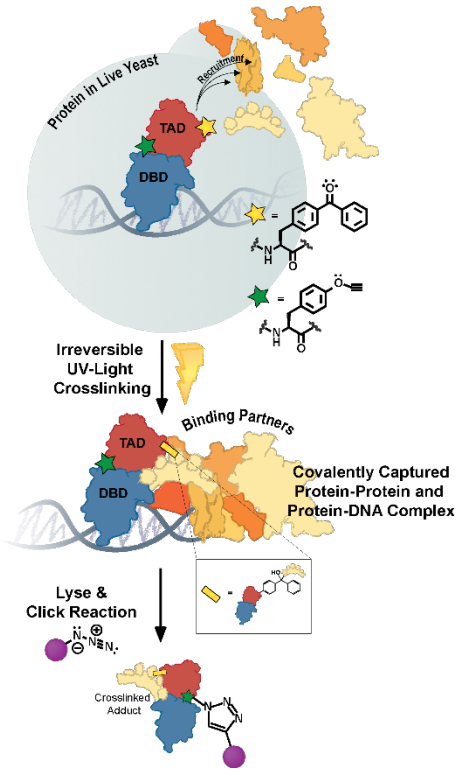


Figure 5-1 Schematic Displaying Hypothetical Dual UAA Incorporation for Covalent Chemical Capture and Enhanced Enrichment with the Click Reaction.

Looking forward, there are many exciting possibilities using the demonstrated power of covalent chemical capture and genetic incorporation of UAAs for the studying of PPIs in a cellular context. This will expand our knowledge and understanding of PPIs in their native environments, allowing the further elucidation of the dynamic interworking's of a cell. With the vast number of PPIs that have already been linked to disease states, a more complete understanding of PPI networks will aid researchers in finding better therapeutic targets for drugs in the future.

References Chapter 5:

1. Zhang, B. B., Zhou, G., and Li, C. (2009) AMPK: An Emerging Drug Target for Diabetes and the Metabolic Syndrome, *Cell Metabolism 9*, 407-416.
2. Hardie, D. G. (2007) AMP-activated SNF1 protein kinases: conserved guardians of cellular energy, *Nat Rev Mol Cell Biol 8*, 774-785.

3. Parrish, A. R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S. P., Dillin, A., and Wang, L. (2012) Expanding the Genetic Code of Caenorhabditis elegans Using Bacterial Aminoacyl-tRNA Synthetase/tRNA Pairs, *ACS Chemical Biology 7*, 1292-1302.
4. Bianco, A., Townsley, F. M., Greiss, S., Lang, K., and Chin, J. W. (2012) Expanding the genetic code of Drosophila melanogaster, *Nat Chem Biol 8*, 748-750.
5. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome, *Nature 464*, 441-444.
6. Niu, W., Schultz, P. G., and Guo, J. (2013) An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon, *ACS Chemical Biology 8*, 1640-1645.
7. Nikić, I., and Lemke, E. A. (2015) Genetic code expansion enabled site-specific dual-color protein labeling: superresolution microscopy and beyond, *Current Opinion in Chemical Biology 28*, 164-173.
8. Arbely, E., Torres-Kolbus, J., Deiters, A., and Chin, J. W. (2012) Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine, *Journal of the American Chemical Society 134*, 11912-11915.
9. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M., and Schultz, P. G. (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid, *Nat Chem Biol 3*, 769-772.

Appendix 1: Work Accomplished in the Garneau-Tsodikova Lab

1a: Biochemical and Structural Analysis of Aminoglycoside Acetyltransferase Eis from Anabaena variabilis[5]

1ai: Abstract

The Mycobacterium tuberculosis enhanced intracellular survival (Eis) protein is a clinically important aminoglycoside (AG) multi-acetylating enzyme. Eis homologues are found in a variety of mycobacterial and non-mycobacterial species. Variation of the residues lining the AG-binding pocket and positions of the loops bearing these residues in the Eis homologues dictates the substrate specificity and, thus, Eis homologues are Nature-made tools for elucidating principles of AG recognition by Eis. Here, we demonstrate that the Eis from *Anabaena variabilis* (Eis_*Ava*), the first non-mycobacterial Eis homologue reported, is a multi-acetylating AG-acetyltransferase. Eis_*Ava*, Eis from *Mycobacterium tuberculosis* (Eis_*Mtb*), and Eis from *Mycobacterium smegmatis* (Eis_*Msm*) have different structures of their AG-binding pockets. We perform comparative analysis of these differences and investigate how they dictate the substrate and cosubstrate recognition and acetylation of AGs by Eis.

1aii: Introduction

Upregulation of the enzyme Eis from *Mycobacterium tuberculosis* (Eis_*Mtb*) has been shown to be responsible for kanamycin A (KAN) resistance in a large fraction of KAN-resistant clinical isolates from tuberculosis patients around the globe.[3, 4] We recently reported the unique and unprecedented ability of Eis_*Mtb*[1] and of its homologue from *Mycobacterium smegmatis* (Eis_*Msm*)[5] to multi-acetylate a variety of aminoglycosides (AGs). Even though Eis_*Mtb* and Eis_*Msm* are structurally very similar,[7] we identified differences in the substrate recognition by these two Eis homologues and in their inhibition.[1] Di-acetylation of the rigid fused-ring AG apramycin (APR) was observed with Eis_*Msm*, but APR was found to not be a substrate of Eis_*Mtb*. Moreover, the potency of common inhibitors of these two Eis proteins varied between

[5] Portions of this appendix are from published work: Pricer, R.E.; Houghton, J.L.; Green, L.D.; Mayhoub, A.S.; Garneau-Tsodikova, S., Biochemical and structural analysis of aminoglycoside-acetyltransferase Eis from Anabaena variabilis. *Mol. Biosyst.*. **2012**, 8, 3305-3313. I contributed to all aspects of this work and was kindly instructed by Dr. Green and Dr. Houghton in learning the techniques presented within.

the two homologues. These data were consistent with structural differences between the AG-binding cavities of the two enzymes.

Due to the rising threat of mycobacterial infection from non-tuberculosis mycobacteria and the ever-growing emergence of resistance to AG antibiotics, it is important to extend the analysis of Eis homologues beyond Eis_*Mtb* and Eis_*Msm* to aid in the understanding of Eis substrate recognition principles across the bacterial kingdom.

As a focus of this study, we chose a putative GCN5-related-N-acetyltransferase (GNAT) from the cyanobacterium *Anabaena variabilis* ATCC 29413, (termed Eis_*Ava* herein) structurally similar to Eis_*Mtb*. A crystal structure of Eis_*Ava* was deposited into the Protein Data Bank (PDB) by the Joint Center for Structural Genomics (PDB code: 2OZG). In this work, to gain insight into the structural requirements dictating substrate and cosubstrate profile as well as the inhibition of Eis proteins, we performed structural and biochemical comparative studies of Eis_*Ava* with Eis_*Mtb* and Eis_*Msm*. Cyanobacteria do not cause any clinical infections that are treated with AGs. However, the *A. variabilis* ATCC 29413 strain was isolated from a sewage oxidation pond,[8] and the development of resistance to AGs through water contamination has been previously documented in various bacterial species.[9] Thus, exposure to AGs through sewage waste provides a potential explanation as to why a putative AG-acetylating enzyme, Eis_*Ava*, has arisen in this cyanobacterial strain.

1aiii: Results and Discussion

In Silico Analysis of Eis Homologues Across 29 Bacterial Species

To evaluate the evolutionary relationship among the 3 Eis homologues of interest (Eis_*Ava*, Eis_*Mtb*, and Eis_*Msm*), we first performed a phylogenetic tree analysis using 29 Eis homologue sequences from a variety of mycobacterial and non-mycobacterial species (Figure A1-1). In general, the mycobacterial and non-mycobacterial Eis homologues can be divided into two distinct clades, with Eis_*Ava* and Eis_*Mtb* being two of the most evolutionary distinct proteins in this set. Interestingly, the Eis homologue from *Mycobacterium abscessus* (Eis_*Mab*) is clustered with the Eis homologues from the non-mycobacterial clade instead of the expected mycobacterial group. The genetic proximity of Eis_*Ava* to Eis_*Mab* further highlights the importance of studying non-mycobacterial enzymes such as Eis_*Ava* to understand the mycobacterial AG-acetylation resistance mechanisms, due to the current increase in drug-

resistant *Mab* infections.[10] In the future, the information learned about Eis_*Ava* could additionally be applied to problematic bacterial species closely related to *Ava* that also contain Eis, such as *Enterococcus faecalis* (*Efa*).

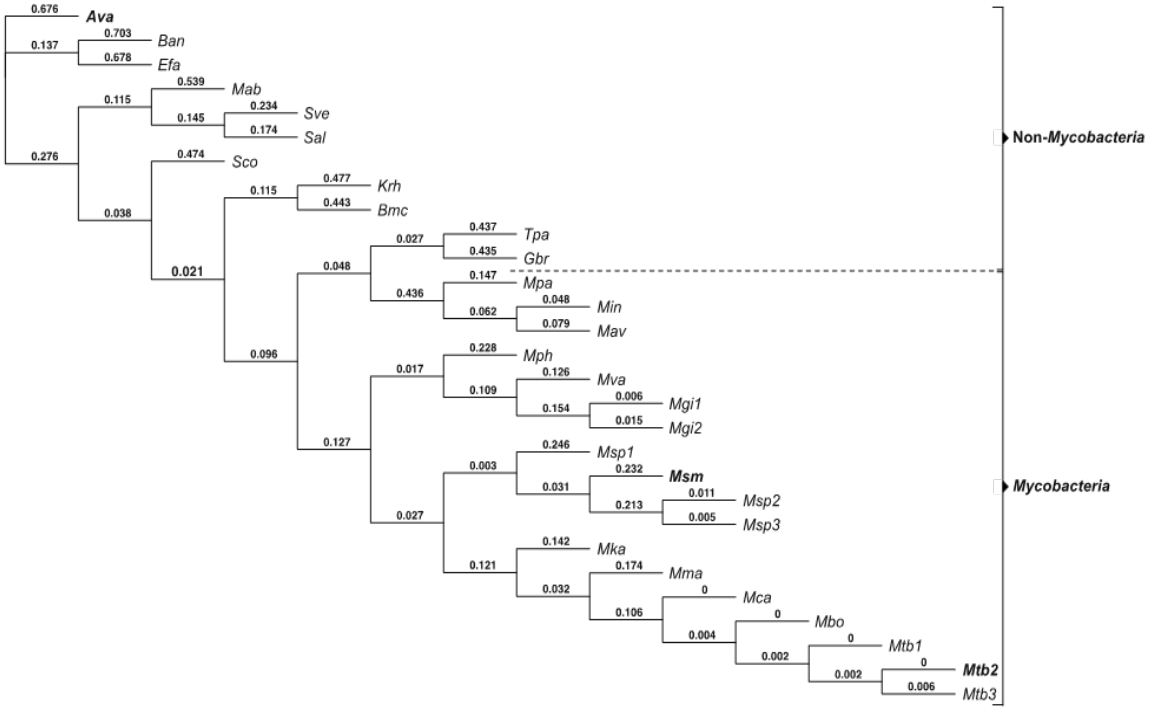


Figure A1-1 Phylogenetic Tree Analysis of 29 Eis Homologues.
The numbers are proportional to the evolutionary distance between the branches of the tree. The Eis proteins discussed in this manuscript are in bold. The abbreviated names of the bacterial species correspond to the following: *Anabaena variabilis* ATCC 29413 (*Ava*), *Bacillus anthracis* str. Sterne (*Ban*), *Enterococcus faecalis* V583 (*Efa*), *Mycobacterium abscessus* ATCC 19977 (*Mab*), *Streptomyces venezuelae* ATCC 10712 (*Sve*), *Streptomyces albus* J1074 (Sal), *Streptomyces coelicolor* A3(2) (*Sco*), *Kocuria rhizophila* DC2201 (*Krh*), *Brevibacterium mcbrellneri* ATCC 49030 (*Bmc*), *Tsukamurella paurometabola* ATCC 8368 (*Tpa*), *Gordonia bronchialis* DSM 43247 (*Gbr*), *Mycobacterium parascrofulaceum* ATCC BAA-614 (*Mpa*), *Mycobacterium intracellulare* ATCC 13950 (*Min*), *Mycobacterium avium* subsp. *avium* ATCC 25291 (*Mav*), *Mycobacterium phlei* RIVM601174 (*Mph*), *Mycobacterium vanbaalenii* PYR-1 (*Mva*), *Mycobacterium gilvum* Spyr1 (*Mgi1*), *Mycobacterium gilvum* PYR-GCK (*Mgi2*), *Mycobacterium* sp. JDM601 (*Msp1*), Mycobacterium smegmatis str. MC2 155 (*Msm*), *Mycobacterium* sp. JLS (*Msp2*), *Mycobacterium* sp. MCS (*Msp3*), *Mycobacterium kansasii* ATCC 12478 (Mka), *Mycobacterium marinum* M (*Mma*), *Mycobacterium canettii* CIPT 140010059 (*Mca*), *Mycobacterium bovis* BCG str. Pasteur 1173P2 (*Mbo*), *Mycobacterium tuberculosis* SUMu002 (*Mtb1*), *Mycobacterium tuberculosis* H37Rv (*Mtb2*), and *Mycobacterium tuberculosis* T17 (*Mtb3*).

Structural Comparison of Eis Homologues

Eis_*Ava*, Eis_*Mtb*, and Eis_*Msm* all share a tripartite monomer structure containing N-terminal and central GNAT regions and a similar homo-hexameric organization.[1, 7] By performing structure-based sequence alignment of these three Eis homologues, we observed that Eis_*Ava* greatly differs in amino acid composition when compared to Eis_*Mtb* (18% sequence identity) and Eis_*Msm* (19% sequence identity; Figure A1-2). In contrast, Eis_*Mtb* and Eis_*Msm* are more similar (53% identical). In all 3 Eis proteins, the key catalytic residue involved in catalysis (Tyr125 in Eis_*Ava*) (Figure A1-2, yellow circles/squares) is strictly

conserved, and those involved in CoA binding (Figure A1-2, orange circles/squares) are highly conserved. However, the amino acid residues predicted to bind AGs through site-directed mutagenesis studies of Eis_*Mtb*[1] and structure analysis of Eis_*Mtb* and Eis_*Msm*[5] are highly divergent between Eis_*Ava* and Eis_*Mtb*/Eis_*Msm* (Figure A1-2, red squares). Taken together,

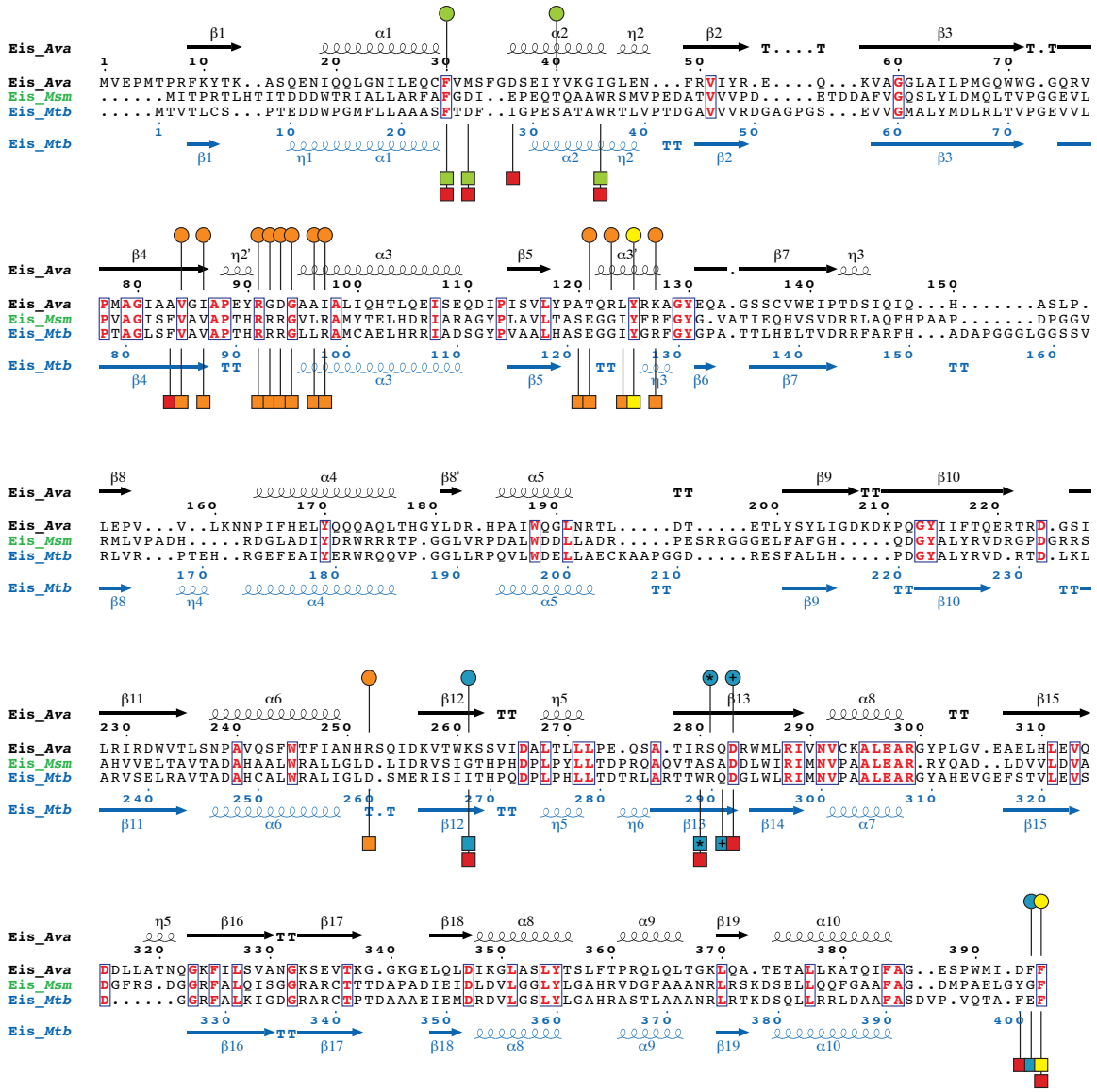


Figure A1-2 Structure-based Sequence Alignment of Eis_Ava, Eis_Msm, and Eis_Mtb.
Eis_*Ava* is from *A. variabilis* ATCC 29413, Eis_*Msm* is from *M. smegmatis* str. MC2 155, and Eis_*Mtb* is from *M. tuberculosis* H37Rv with alignment generated using Secondary-structure Matching (SSM).[6] Residues in bold red in blue boxes are conserved between the 3 Eis homologues. The circles above and the squares below the Eis_*Ava* and Eis_*Mtb* sequences, respectively, correspond to important residues in these sequences. Based on structural and mutagenesis studies of Eis_*Mtb*, the residues proposed to be involved in catalysis, in CoA binding, and in the formation of the AG-binding pocket are marked by yellow, orange, and red circles/squares, respectively.[1] The AG-binding pocket of Eis_*Mtb* is divided into two channels. Residues lining these two channels are marked by green and turquoise circles/squares. The * and the + symbols indicate residues that structurally aligned when superimposing the crystal structures of Eis_*Ava* and Eis_*Mtb*.

these observations demonstrate the evolutionary divergence between the Eis proteins from mycobacterial and non-mycobacterial species observed in Figure A1-1.

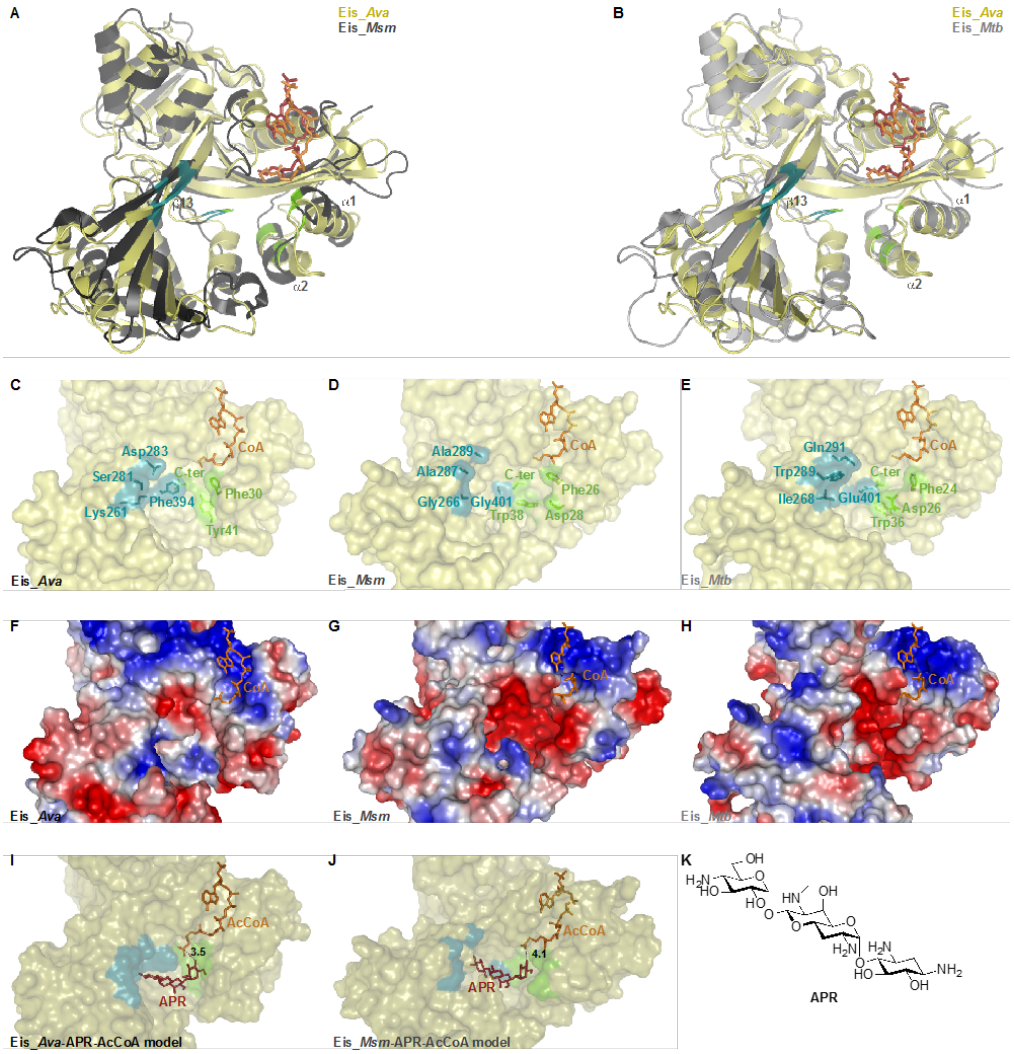


Figure A1-3. Structural Comparison of Eis Homologues from Ava, Msm, and Mtb with Molecular Models of Eis-APR-AcCoA Complexes.
A. Structural alignment of Eis_*Ava* (PDB: 2OZG) with Eis_*Msm* (PDB: 3SXN) (The alignment of one of the monomers of the hexameric structures is shown). B. Structural alignment of Eis_*Ava* (PDB: 2OZG) with Eis_*Mtb* (PDB: 3R1K) (The alignment of one of the monomers of the hexameric structures is shown). The active site of C. Eis_*Ava*, D. Eis_*Msm*, and E. Eis_*Mtb* with residues lining the two channels (green and blue) of the AG-binding pocket highlighted. Surface representation of the Eis monomer active sites of F. Eis_*Ava*, G. Eis_*Msm*, and H. Eis_*Mtb* colored according to their electrostatic potential, positive in blue, negative in red, and hydrophobic in white. I. A model of AcCoA and APR bound to Eis_*Ava*. J. A model of AcCoA and APR bound to Eis_*Msm*. K. Structure of APR.

Even though Eis_*Ava* greatly differs in amino acid composition from both Eis_*Mtb* and Eis_*Msm*, its structure is generally very similar to those of Eis_*Mtb* and Eis_*Msm* (Figure A1-3A and B). To explore the structural similarities and differences among the 3 studied Eis proteins and their relationship to function, we compared the structures of their substrate-binding cavities (Figure A1-3C-E) and observed striking differences. We previously reported that the AG-

binding site of Eis_*Mtb* is divided into two distinct narrow channels (highlighted in green and blue in Figure A1-3E), while the AG-binding pocket of Eis_*Msm* consists of one wide and open cavity (Figure A1-3D), as a result of the small amino acid side-chains of the residues lining one of the two Eis_*Msm* channels (blue channel of Eis_*Msm*: Ala289, Ala287, Gly266, and Gly401; corresponding residues in Eis_*Mtb*: Gln291, Trp289, Ile268, and Glu401). As a result of the broadening of its AG-binding site relative to that in Eis_*Mtb*, Eis_*Msm* can accommodate the structurally rigid APR AG while Eis_*Mtb* cannot.[5] Interestingly, the substrate-binding cavity of Eis_*Ava* is divided into two distinct channels, as in the case of Eis_*Mtb*, but it differs structurally. As in Eis_*Mtb*, the blue channel of Eis_*Ava* composed of Asp283, Ser281, Lys261, and Phe394 (corresponding residues in Eis_*Mtb*: Gln291, Trp289, Ile268, and Glu401) (Figure A1-2, blue circles/squares and Figure A1-3C and E) is narrow. However, the green channel of Eis_*Ava* is much larger than that of Eis_*Mtb* and could potentially accept APR as a substrate as does Eis_*Msm*. Different positions of the loop between the α1 and α2- helices and the β13-sheet in Eis_*Ava* and Eis_*Msm* (Figure A1-3B) are also in part responsible for the broadening of the substrate-binding cavity of these two Eis homologues relative to that in Eis_*Mtb* helices and the β13-sheet in Eis_*Ava* and Eis_*Msm* (Figure A1-3B) are also in part responsible for the broadening of the substrate-binding cavity of these two Eis homologues relative to that in Eis_*Mtb*.

Multi-acetylation, Substrate, and Cosubstrate Specificity Profiles of Eis_*Ava*

We recently reported the substrate and multi-acetylation profiles of Eis_*Mtb* and Eis_*Msm*, demonstrating the unique ability of the Eis enzymes to multi-acetylate many AGs.[1, 5] To investigate a potential effect of the broadening of the substrate-binding pocket of Eis_*Ava* on its substrate specificity profile, we monitored the acetylation by Eis_*Ava* of 11 AGs: amikacin (AMK), APR, KAN, neamine (NEA), neomycin B (NEO), netilmicin (NET), paromomycin (PAR),

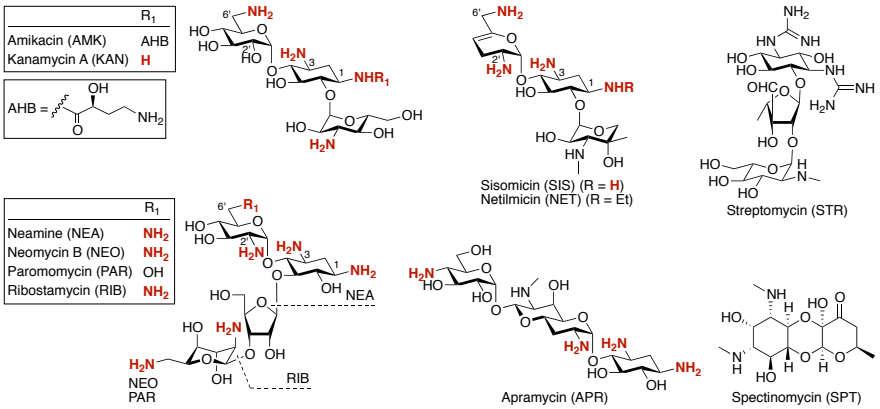


Figure A1-4 Structures of Aminoglycosides (AGs) Tested.
The amine functionalities that could potentially be modified by Eis proteins are highlighted in bold red.

ribostamycin (RIB), sisomicin (SIS), spectinomycin (SPT), and streptomycin (STR) (Table A1: 1and Figure A1-4, Figure A1-6), by UV-Vis and mass spectrometry assays. To ensure the validity of the comparisons made in this work, we cloned and overexpressed Eis_*Ava* in an identical fashion to Eis_*Mtb* and Eis_*Msm*, which we previously reported (Figure A1-5). As predicted, based on the structural homology with other Eis proteins, Eis_*Ava* is a functional AG-acetyltransferase. As we previously observed with Eis_*Mtb* and Eis_*Msm*, we found SPT and STR to not be substrates of Eis_*Ava*. By mass spectrometry, we observed that the number of acetylated positions on AMK, NEA, NEO, NET, and RIB by the 3 Eis homologues studied was the same for the 3 enzymes. With other AGs, such as KAN and SIS, we found the number of acetylations with Eis_*Ava* to be less than those observed with Eis_*Mtb* and Eis_*Msm*, which were equal for any given AG. Of particular note is the tri-acetylation of PAR by Eis_*Ava* and Eis_*Msm*, in contrast to only di-acetylation of PAR by Eis_*Mtb*. This result suggests that the broader substrate-binding cavities of Eis_*Ava* (Figure A1-3C and F) and Eis_*Msm* (Figure A1-3D and G) accommodate the large PAR molecule in more orientations than Eis_*Mtb* does. In addition, we found APR to be mono-acetylated by Eis_*Ava*, which corroborated the previously proposed role of the expanded substrate-binding cavity in binding this rigid extended AG. The presence of two channels (one narrow (blue) and one large (green) in Figure A1-3C) in the substrate-binding site of Eis_*Ava* versus a single AG-binding pocket in Eis_*Msm* that is larger than the green channel of Eis_*Ava* (Figure A1-3D) could explain the mono-acetylation and di-acetylation of APR observed with Eis_*Ava* and Eis_*Msm*, respectively.

Table A1-1 Comparison of Number of Acetylations of Various AGs by Eis Homologues from A. variabilis, M. tuberculosis, and M. smegmatis.

AG	Eis_*Ava*	Eis_*Mtb*[a]	Eis_*Msm*[b]
AMK	Tri	Tri	Tri
APR	Mono	X[c]	Di
KAN	Mono	Di	Di
NEA	Tri	Tri	Tri
NEO	Tri	Tri	Tri
NET	Di	Di	Di
PAR	Tri	Di	Tri
RIB	Tri	Tri	Tri
SIS	Di	Tri	Tri
SPT	X[c]	X[c]	X[c]
STR	X[c]	X[c]	X[c]

[a] These data were previously reported.[1] [b] These data were previously reported.[5] [c] X indicates no observed acetylation.

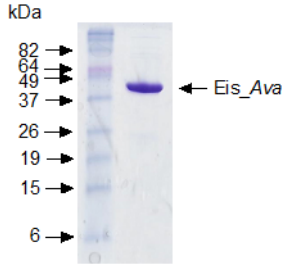


Figure A1-5 Coomassie Blue-stained SDS-PAGE of Purified Eis_Ava.

To explore how APR may be accommodated in the binding pocket of Eis_*Ava* for acetylation, we performed docking of APR to

Eis_*Ava* and Eis_*Msm* in complex with AcCoA (Figure A1-3I and J). The shorter loop between the α1 and α2-helices of Eis_*Ava* that is responsible for the expansion of the green channel of this enzyme when compared to its counterpart in Eis_*Msm* (Figure A1-3B, I and J) could account for the different conformations adopted by APR in the AG-binding site and the different number of acetylations of APR by these two Eis enzymes, mono- versus di-acetylation by Eis_*Ava* and Eis_*Msm*, respectively (Table A1-1). Taken together, the structural analysis of the substrate-binding cavities of Eis homologues and their multi-acetylation profiles revealed that Eis_*Ava* behaves more similarly to Eis_*Msm* than to Eis_*Mtb*.

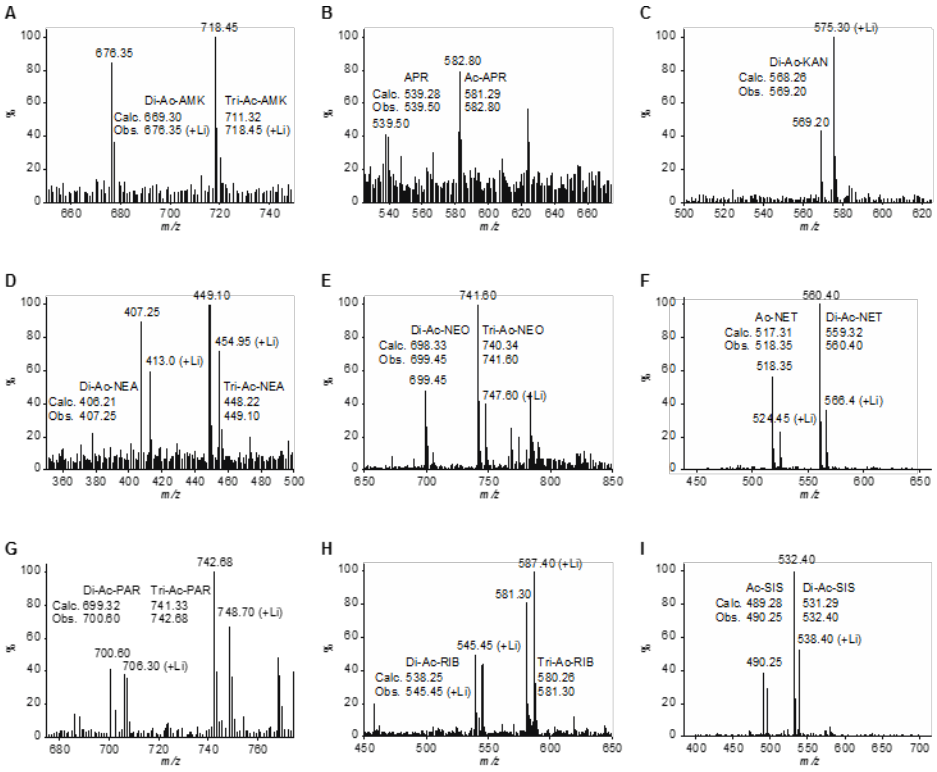


Figure A1-6 Mass Spectra of AGs Multi-acetylated by Eis_Ava.

In addition to studying the AG substrate promiscuity of Eis_*Ava*, we also investigated its cosubstrate profile. We recently showed that Eis_*Mtb* is limited in its cosubstrates, efficiently accepting only AcCoA and n-propionyl-CoA and moderately so crotonyl-CoA and malonyl-CoA for transfer onto a few AGs.[11] We determined that among the 6 acyl-CoA cosubstrates tested with Eis_*Ava* (AcCoA, butyryl-CoA, crotonyl-CoA, malonyl-CoA, myristoyl-CoA, and n-propionyl-CoA), only the natural AcCoA cosubstrate and its close structural homologue n-propionyl-CoA were accepted by the cyanobacterial Eis.

Steady-state kinetic parameters for Eis_*Ava*

To further understand the action of Eis_*Ava* and the subtleties differentiating the 3 investigated Eis homologues, we performed steady-state kinetic measurements of AG acetylation by Eis_*Ava* and compared the Michaelis-Menten parameters, K_m and k_{cat}, to those of Eis_*Mtb*

and Eis_*Msm* (Table A1-2 and Figure A1-7). In this study, we determined all kinetic parameters (for Eis_*Ava*, Eis_*Mtb*, and Eis_*Msm*) for 500 µM AcCoA, in a range of 0-2500 µM AG, and 0.25 µM Eis. In a previous report, we determined and compared the K_m and k_{cat} values for Eis_*Mtb* and Eis_*Msm* using 100 µM AcCoA, 0-2500 µM AG, and 0.25 µM Eis.[5] AcCoA was used at a higher concentration in this study to ensure that all enzyme was bound to AcCoA thereby allowing a direct interpretation of kinetics as a property of this holo-Eis. For APR, a substrate of Eis_*Ava* and Eis_*Msm*, but not of Eis_*Mtb*, we found that K_m and k_{cat} values are very similar (337 ± 98 µM and 0.019 ± 0.002 s-1 for Eis_*Ava* and of 150 ± 43 µM and 0.019 ± 0.002 s-1 for Eis_*Msm*). For KAN and NEO, the K_m and k_{cat} values for acetylation by Eis_*Ava* were also more similar to those of Eis_*Msm* than to those of Eis_*Mtb*. Overall, the kinetic parameters of Eis_*Ava* appear to be most similar to those of Eis_*Msm*, which can be explained by the larger substrate-binding cavities of these two enzymes compared to the divided AG-binding site of Eis_*Mtb*.

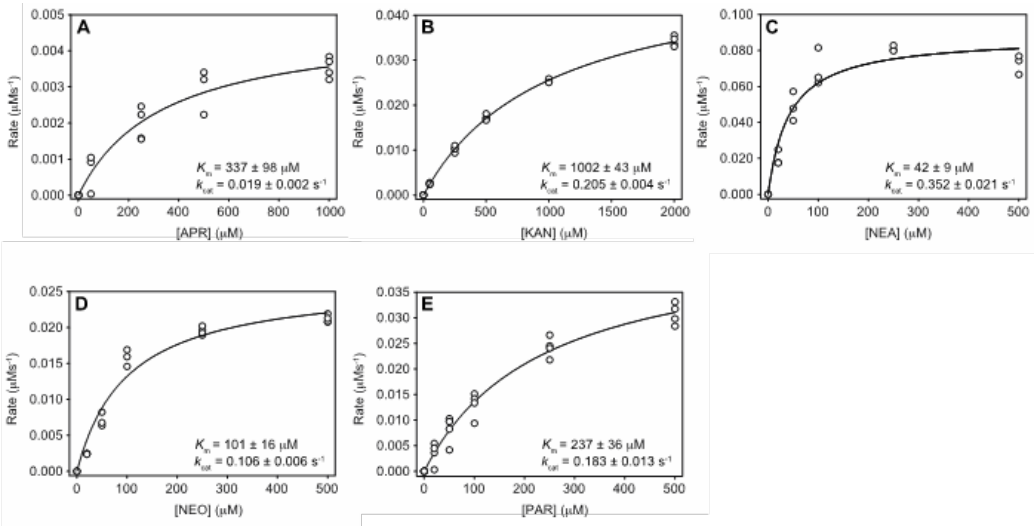


Figure A1-7 Michaelis-Menten Analysis of the Eis_Ava Catalyzed Acetylation. A. APR, B. KAN, C. NEA, D. NEO, and E. PAR.

Table A1-2 Steady-State Kinetic Parameters for AG Concentration-Dependent Acetylation by Eis Homologues from A. variabilis, M. smegmatis, and M. tuberculosis.

	Eis_*Ava*		
AG	**K_m (µM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
APR	337 ± 98	0.019 ± 0.002	57 ± 18
KAN	1002 ± 43	0.205 ± 0.004	205 ± 9
NEA	42 ± 9	0.352 ± 0.021	7413 ± 1938
NEO	101 ± 16	0.106 ± 0.006	1052 ± 184
PAR	237 ± 36	0.183 ± 0.013	773 ± 132
	Eis_*Msm*		
AG	**K_m (µM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
APR	150 ± 43	0.019 ± 0.002	127 ± 39
KAN	665 ± 42	0.360 ± 0.010	541 ± 37
NEA	1302 ± 312	0.687 ± 0.083	528 ± 142
NEO	110 ± 14	0.148 ± 0.006	1345 ± 180
PAR	738 ± 158	0.235 ± 0.033	318 ± 82
	Eis_*Mtb*		
AG	**K_m (µM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_m ($M^{-1}s^{-1}$)**
APR	---[a]	---	---
KAN	330 ± 40	0.526 ± 0.027	1594 ± 250
NEA	1315 ± 492	0.699 ± 0.136	532 ± 224
NEO	122 ±23	0.610 ± 0.029	5000 ± 972
PAR	110 ± 21	0.136 ± 0.012	1236 ± 260

[a] --- indicates that APR is not a substrate for Eis_*Mtb*.

Regio-Specificity of NEA Di-Acetylation by Eis_*Ava*.

We previously demonstrated that Eis_*Mtb*[1] and Eis_*Msm*[5] sequentially tri-acetylate NEA at the 2'-, 6'-, and 1-position. Seeking a deeper understanding of the differences and similarities between the mycobacterial and non-mycobacterial Eis proteins, we sought to compare the regio-specificity and order of multi-acetylation of NEA by Eis_*Ava* to those by Eis_*Mtb*/Eis_*Msm*. Interestingly, by thin-layer chromatography (TLC) and NMR spectroscopy (1H, gCOSY, zTOCSY, and gHSQC), we could only observe di-acetylation of NEA by Eis_*Ava* (Figure A1-8, Figure A1-9, Figure A1-10, and Table A1-3). Remarkably, we also established that the first position of the NEA scaffold acetylated was the 2'-position, as previously observed with Eis_*Mtb* and Eis_*Msm*, but that the second site modified was the 1-position instead of the expected 6'-position. It is important to note that, as established by mass spectrometry, a third position can be acetylated by Eis_*Ava* on the NEA scaffold. The change in order of the position acetylated by Eis_*Ava* could potentially explain why for this enzyme we did not observe by UV-Vis and mass spectrometry assays a third acetylation, which could proceed at the much slower rate. These data suggest that the order and number of acetylations may vary for each AG based on the Eis homologue utilized. Further studies aimed at establishing the number and positions of acetylation on multiple AGs by a variety of Eis homologues are currently underway in our laboratory.

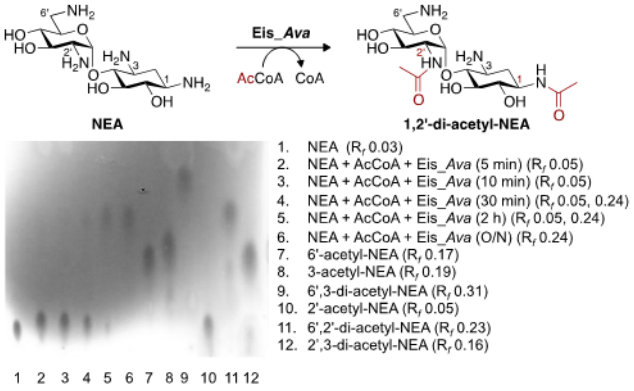


Figure A1-8 Formation of 1,2'-di-acetyl-NEA by Eis_Ava Monitored by a TLC Assay.
Lanes 1-6: a time course showing the mono- and di-acetyl-NEA products of the Eis_*Ava* reaction. Lanes 7-12: Controls for mono- and di-acetylation of NEA by AAC(2')-Ic, AAC(3)-IV, and AAC(6') used individually or sequentially.

Inhibition of Eis_*Ava*

We recently reported the discovery and characterization of Eis_*Mtb* inhibitors,[2] which also inhibited Eis_*Msm* (Figure A1:11).[5] As Eis homologues are found in a variety of bacterial strains (Figure A1-1), including emerging resistant species such as *Mab*, the importance of finding Eis inhibitors that efficiently work across various species is increasing with the number of resistant bacterial species that harbour Eis homologues continuously rising. To perform preliminary exploration of the potential of the identified Eis_*Mtb* inhibitors against Eis

homologues from various bacterial species, we tested 5 Eis_*Mtb* inhibitors against Eis_*Ava* and Eis_*Msm*. For chlorhexidine (1), the IC_{50} value of 20 ± 7 µM observed for Eis_*Ava* was was 107-fold higher than that for Eis_*Mtb* (188 ± 30 nM) and 11-fold higher than that against Eis_*Msm* (1.85 ± 0.39 µM). For compounds 2 and 3, the IC_{50} values were comparable for both mycobacterial strains. Interestingly, for compound 4, the IC_{50} value of 455 ± 92 nM observed for Eis_*Msm* was 4.4-fold lower than that for Eis_*Mtb* (2.01 ± 0.12 µM). The poor inhibition of Eis_*Ava* by all 5 inhibitors (IC_{50} values of 20 ± 7 µM for compound 1 and >200 µM for compounds 2-5) tested could be attributed to the electrostatics of its substrate-binding cavity, which is more hydrophobic (Figure A1-3F) than in Eis_*Mtb* (Figure A1-3H) and in Eis_*Msm* (Figure A1-3G). For example, the hydrophobic Met32 and Pro185 residues of the substrate-binding pocket in Eis_*Ava* replace the structurally homologous negatively-charged Asp28 and

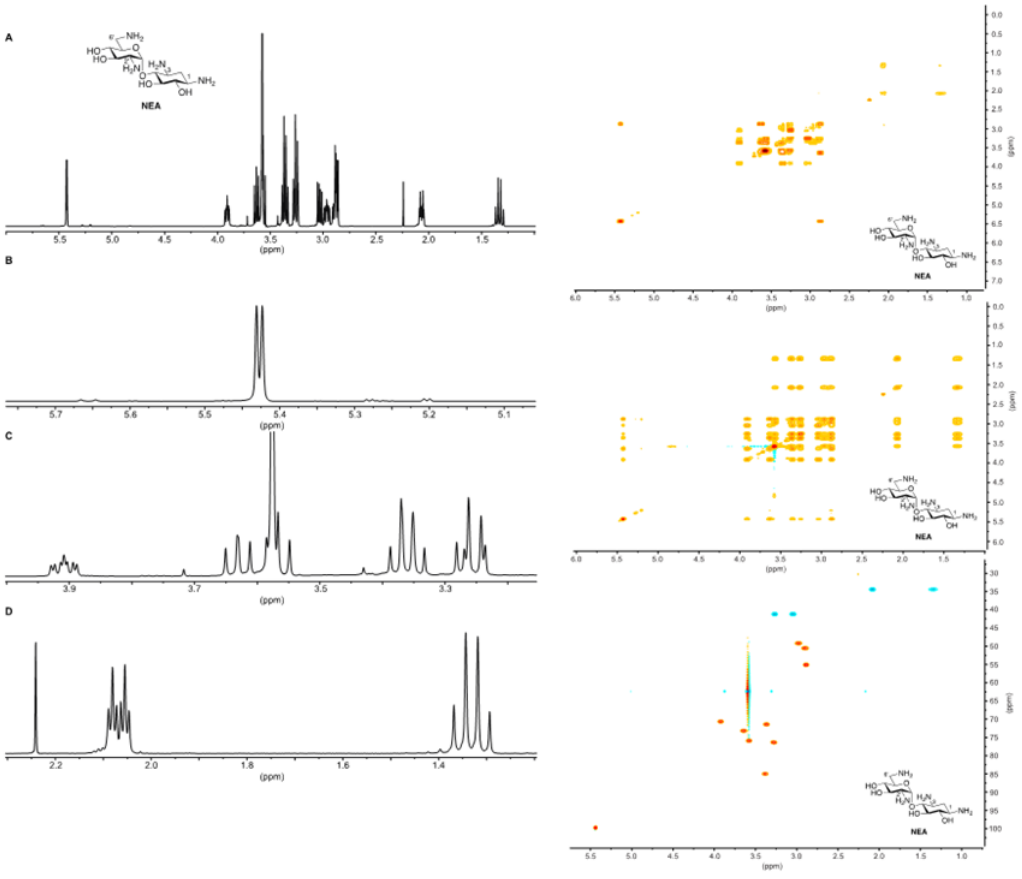


Figure A1-9 NMR of NEA.
Left: 1H NMR of NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz). The full spectrum is shown in panel A and the expansions in panels B-D. Right Top: gCOSY of NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz). Right Middle: zTOCSY of NEA in 1.2:8.8/D_2O:H_2O at pH 7.5

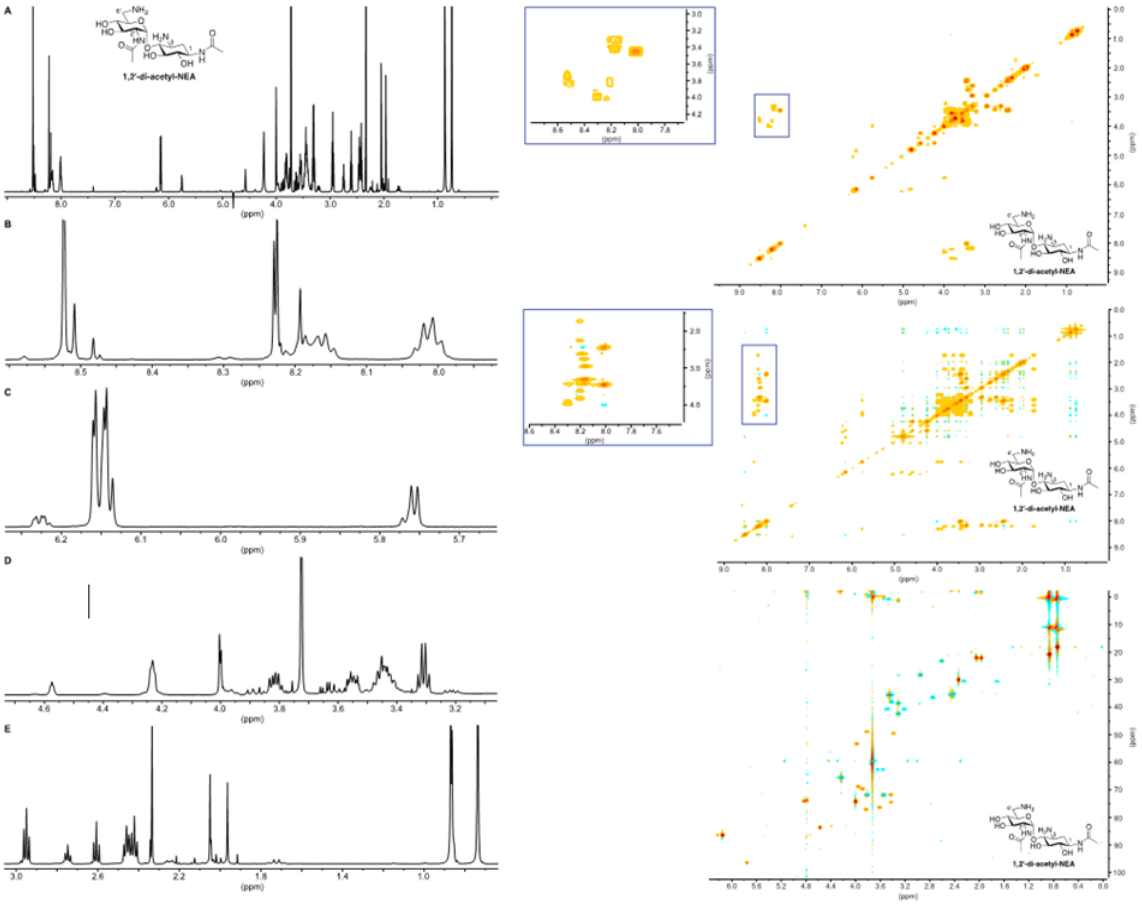


Figure A1-10 NMR of 1,2'-di-acetyl-NEA.
Left: 1H NMR of 1,2'-di-acetyl-NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz). The full spectrum is shown in panel A and the expansions in panels B-D. Right Top: gCOSY of 1,2'-di-acetyl-NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz). Right Middle: zTOCSY of 1,2'-di-acetyl-NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz). Right Bottom: gHSQC of 1,2'-di-acetyl-NEA in 1.2:8.8/D_2O:H_2O at pH 7.5 (500 MHz).

Table A1-3 Proton Chemical Shifts determined for NEA and 1,2'-di-acetyl-NEA.[a]

Ring	H position	NEA	1,2'-di-acetyl-NEA	Δppm
II	1	2.91-2.88[b] (m) [2.89][d]	3.70-3.84 (m) [3.77]	**0.88**
	2_{ax}	1.33 (ddd (app. q), $J_{2ax,2eq} = J_{2ax,1} = J_{2ax,3} = 12.0$ Hz)[c]	1.72 (ddd (app. q), $J_{2ax,2eq} = J_{2ax,1} = J_{2ax,3} = 13.0$ Hz)	0.39
	2_{eq}	2.08 (ddd (app. dt), $J_{2eq,2ax} = 12.0$ Hz, $J_{2eq,1} = J_{2eq,3} = 4.0$ Hz)	2.25 (ddd (app. dt), $J_{2eq,2ax} = 13.0$ Hz, $J_{2eq,1} = J_{2eq,3} = 4.1$ Hz)	0.17
	3	2.96[b] (ddd, $J_{3,2eq} = 4.0$ Hz, $J_{3,2ax} = 12.0$ Hz, $J_{3,4} = 9.5$ Hz)	3.33-3.41 (m) [3.38]	0.42
	4	3.35[c] (dd, (app. q) $J_{4',3'} = J_{4',5'} = 9.5$)	3.91-3.79 (m) [3.82]	0.45
	5	3.59-3.55 (m) [3.57]	3.66-3.58 (m) [3.62]	0.05
	6	3.29-3.23[c] (m) [3.26]	3.51-3.37 (m) [3.42]	0.16
I	1'	5.42 (d, $J_{1',2'} = 4.0$ Hz)	5.76 (d, $J_{1',2'} = 3.8$ Hz)	0.34
	2'	2.87 (dd, $J_{2',1'} = 4.0$ Hz, $J_{2',3'} = 10.5$ Hz)	4.01-3.93 (m) [3.98]	**1.11**
	3'	3.63 (dd (app. t), $J_{3',4'} = 9.5$ Hz, $J_{3',2'} = 10.5$ Hz)	3.92-3.85 (m) [3.89]	0.26
	4'	3.35 (dd, (app. q) $J_{4',3'} = J_{4',5'} = 9.5$)	3.51-3.37 (m) [3.44]	0.09
	5'	3.91 (ddd, $J_{5',4'} = 9.5$ Hz, $J_{5',6'a} = 7.5$ Hz, $J_{5',6'b} = 3.0$ Hz)	4.01-3.93 (m) [3.96]	0.05
	$6'_a$	3.04 (dd, $J_{6'a,6'b} = 14.0$ Hz, $J_{6'a,5'} = 7.5$ Hz)	3.22 (dd, $J_{6'a,6'b} = 13.0$ Hz, $J_{6'a,5'} = 7.5$ Hz)	0.18
	$6'_b$	3.29-3.23 (m) [3.25]	3.51-3.37 (m) [3.49]	0.24
Acetyl	NH-2'	×	8.30 (d, $J_{NH,2'} = 8.0$ Hz)	
	NH-1	×	8.24-8.14 (m) [8.20]	
	CH_3C=O on 2'	×	2.05 (s)	
	CH_3C=O on 1	×	1.97 (s)	

[a]The chemical shift were established based on ^{1}H, zTOCSY, gCOSY, and gHSQC NMR (500 MHz). [b]Could be analogous position of the 2-deoxystreptamine (DOS) ring. [c]Multiplicity and *J* are given in (). [d]The numbers in [] were determined from gCOSY and/or zTOCSY. [e]Could be analogous position of the DOS ring. ×Indicates that the acetyl moiety is not present in the molecule.

Asp195 in Eis_*Msm*. Therefore, to assess the utility of discovered Eis_*Mtb* inhibitors on new potential Eis homologues, a thorough analysis of the substrate-binding pocket residues is needed.

Reduced Activity of Eis Homologues Towards 6'-glycinyl-KAN.

The development of AGs capable of avoiding acetylation by Eis enzymes could present an alternative strategy to the discovery of Eis inhibitors to combat the Eis resistance problem. We recently reported a simple chemical approach for the production of biologically active 6'-N-acylated AG derivatives.[12] To explore the potential of these acylated AGs to evade Eis multi-acetylation, we tested 6'-glycinyl-KAN against Eis_*Ava*, Eis_*Msm*, and Eis_*Mtb* (Figure A1-12). We observed that in contrast to KAN that efficiently gets multi-acetylated by all Eis homologues tested (blue, red, and green circles in Figure A1-12), 6'-glycinyl-KAN is not a substrate for these enzymes (blue, red, and green triangles in Figure A1-12). These data demonstrate the potential of developing novel AG derivatives towards combating AG acetyltransferase enzymes.

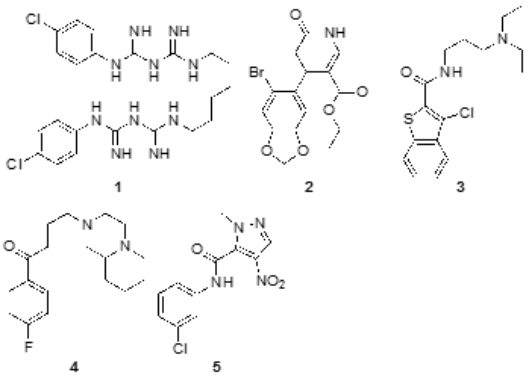


Figure A1-11 Structures of Inhibitors Tested Against Eis_Ava, Eis_Msm, and Eis_Mtb.

Table A1-4 Inhibition of Eis Homologues from A. variabilis, M. smegmatis, and M. tuberculosis by Compounds 1-5 for NEO Acetylation.

Compound	IC_{50} [μM] Eis_*Ava*	Eis_*Msm*	Eis_*Mtb*[a]
1	20 ± 7	1.85 ± 0.39	0.188 ± 0.030
2	>200	0.925 ± 0.201	1.09 ± 0.14
3	>200	1.69 ± 0.32	1.24 ± 0.16
4	>200	0.455 ± 0.092	2.01 ± 0.12
5	>200	>200	2.29 ± 0.52

[a] These data were previously reported.[2]

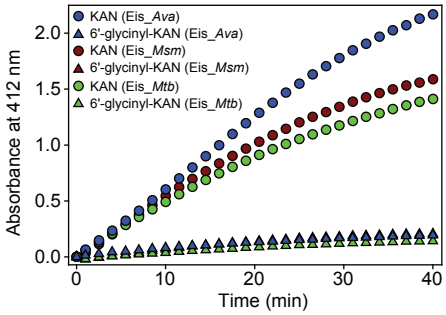


Figure A1-12 UV-Visible Spectrum Showing the Activity of Eis Homologues Against KAN and 6'-glycinyl-KAN.
All Eis homologues (Eis_*Ava* - blue, Eis_*Msm* - red, and Eis_*Mtb* -green) are active against KAN (circles) and almost inactive against 6'-glycinyl-KAN (triangles). As a result, the red triangles are completely hidden behind the blue triangles.

1aiv: Conclusion

In this work, we refined our understanding of the details of the substrate recognition for multi-acetylation of AGs by Eis homologues from mycobacterial and non-mycobacterial Eis homologues. We also explored the applicability of Eis_*Mtb* inhibitors to the inactivation of Eis homologues from other bacterial species. From these studies we concluded that (i) the substrate-binding cavities of Eis homologues vary greatly even though the overall structures of the enzymes are highly comparable, (ii) Eis_*Ava* behaves more similarly to Eis_*Msm* than to Eis_*Mtb*, (iii) the order of acetylation for a given AG can vary for different Eis homologues, and

(iv) inhibitors of Eis_*Mtb* could be applied with varying degree of efficiency against bacterial strains containing Eis homologues. Further studies are underway in our laboratory aimed at establishing the positions acetylated by Eis homologues on a number of AG scaffolds. Structure activity relationship studies towards the development of Eis inhibitors with high potency and broad application are also ongoing.

1av: Experimental

Bacterial Strains, Plasmids, Materials, and Instrumentation

Chemically competent *Escherichia coli* TOP10 and BL21 (DE3) strains were purchased from Invitrogen (Carlsbad, CA). *A. variabilis* ATCC 29413 genomic DNA was purchased from the American Type Culture Collection (ATCC). The pET28a plasmid used in this study was purchased from Novagen (Gibbstown, NJ). Primers used for PCR were purchased from Integrated DNA Technologies (Coralville, IA). All reagents and enzymes used for cloning, including restriction enzymes, Phusion DNA polymerase, and T4 DNA ligase, were purchased from New England Biolabs (Ipswich, MA). DNA sequencing was performed at the University of Michigan DNA Sequencing Core. Dithionitrobenzoic acid (DTNB), AcCoA, acyl-CoAs (butyryl-CoA, crotonyl-CoA, malonyl-CoA, myristoyl-CoA, and n-propionyl-CoA), and aminoglycosides (AGs) (AMK, APR, KAN, NEO, RIB, SIS, SPT, and STR; Figure A1-4) were purchased from Sigma-Aldrich (Milwaukee, WI) and used without further purification. The remaining AGs (NEA, NET, and PAR; Figure A1-4) were purchased from AK Scientific (Mountain View, CA). Eis_*Mtb* and Eis_*Msm* were purified as previously described using the pEis_*Mtb*-pET28a[1] and pEis_*Msm*-pET28a[5] constructs, respectively. All UV-Vis spectrophotometric assays were performed in 96-well plates (Fisher Scientific; Pittsburg, PA) using a multimode SpectraMax M5 plate reader. A Shimadzu LCMS-2019EV equipped with a SPD-20AV UV-Vis detector and a LC-20AD liquid chromatograph was used for all liquid chromatography mass spectrometry (LCMS) measurements. The PDB structures 3R1K (Eis_*Mtb*), 3SXN (Eis_*Msm*), and 2OZG (Eis_*Ava*) were analyzed using Coot[13] and PyMOL (The PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC).

Phylogenetic Tree Creation for Eis Homologues

The amino acid sequences of 29 Eis homologues were retrieved from NCBI through the Geneious Pro 4.8.5 program. The phylogenetic tree (Figure A1-1) was built by neighbor-joining using the Geneious Pro 4.8.5 program with the following parameters: Jukes-Cantor genetic

distance and the cost matrix Blosum45. The NCBI accession numbers for the Eis homologues were YP_325469 (*Ava*), YP_029001 (*Ban*), NP_814755 (*Efa*), YP_001705255 (*Mab*), CCA55694 (*Sve*), ZP_06590713 (*Sal*), NP_628362 (*Sco*), YP_001855041 (*Krh*), ZP_06804610 (*Bmc*), YP_003645809 (*Tpa*), YP_003272441 (*Gbr*), ZP_06852004 (*Mpa*), ZP_05228133 (*Min*), ZP_05216001 (*Mav*), ZP_09975446 (*Mph*), YP_956800 (*Mva*), YP_004079667 (*Mgi1*), YP_001132138 (*Mgi2*), YP_004522750 (*Msp1*), YP_887817 (*Msm*), YP_001071002 (*Msp2*), YP_639864 (*Msp3*), ZP_04749484 (*Mka*), YP_001852011 (*Mma*), YP_004745879 (*Mca*), YP_978521 (*Mbo*), ZP_07418795 (*Mtb*1), CAB03742 (*Mtb*2), and ZP_06450792 (*Mtb*3).

Preparation of pEis_*Ava*-pET28a Overexpression Construct

PCR for the amplification of the homologous eis gene from *A. variabilis* was performed using Phusion DNA polymerase according to instructions provided by NEB. *A. variabilis* ATCC 29413 (*Ava*) genomic DNA was used as the template for PCR with forward primer 5'-GTGCTTCATATGGTAGAACCAATGAC-3' and reverse primer 5'-CTATGCCTCGAGTTAAAAGAAATCAATC-3'. The amplified PCR product, eis_*Ava*, was inserted into a linearized pET28a vector between NdeI and XhoI restriction sites. The plasmid containing eis_*Ava* was transformed into chemically competent *E. coli* TOP10 cell and used to confirm the sequence of the Eis_*Ava* gene (gene locus *Ava*_4977) (University of Michigan DNA Sequencing Core).

Overproduction and Purification of Eis_*Ava* Protein

The Eis_*Ava* protein, containing a N-terminal His6-tag, was overexpressed and purified using the pEis_*Ava*-pET28a overexpression construct as previously reported for its homologue in *Mycobacterium tuberculosis* (Eis_*Mtb*).[1] The Eis_*Ava* protein was dialyzed in Tris-HCl buffer (50 mM, pH 7.5 adjusted at rt), flash frozen in the dialysis buffer containing 10% glycerol, and stored at -80 °C. The protein yield after purification was 1.5 mg per L of culture (Figure A1-5).

UV-Vis Spectrophotometric Assay Determination of AG Selectivity Profile of Eis_*Ava*

The Eis_*Ava* acetyltransferase activity was monitored by using the Ellman's method in which the CoA thiol, freed by the Eis_*Ava* catalyzed reaction, is reacted with DTNB to produce an increase in absorbance that can be monitored at 412 nm ($\varepsilon 412$ = 13,600 M-1cm-1).[14] Reactions (200 μL) containing AG (0.1 mM, 1 eq), acyl-CoA (0.5 mM, 5 eq), DTNB (2 mM), and Tris-HCl (50 mM, pH 7.5 adjusted at rt) were initiated by the addition of Eis_*Ava* (0.5 μM)

at 25 °C. The reactions were monitored by taking readings every 30 seconds for 1 hour in 96-well plate format.

Determination by Mass Spectrometry of the Number of AG Amine Functionalities Acetylated by Eis_*Ava*

To determine the number of acetylations performed by Eis_*Ava* on each AG substrate, we used LCMS. Reactions (30 µL) containing AG (0.67 mM, 1 eq), AcCoA (6.7 mM, 10 eq), Tris-HCl (50 mM, pH 8.0 adjusted at rt), and Eis_*Ava* (10 µM) were run overnight at rt. To prepare samples for the LCMS, ice-cold MeOH (30 µL) was added to the reaction mixture, followed by incubation for 20 minutes at -20 °C to precipitate the protein. Centrifugation (13,000 rpm, rt, 10 minutes) was then used to pellet the protein and the supernatant was used for LCMS injection after its dilution in H_2O (10 µL of supernatant into 20 µL of H_2O). The whole sample (30 µL) was injected onto the LCMS, used in positive mode with H_2O (0.1% formic acid). Mass spectra showing the number acetylations on AMK, APR, KAN, NEA, NEO, NET, PAR, RIB, and SIS are presented in Figure A1-6. These data are also summarized in Table A1-1.

Molecular Modeling of APR with Eis_*Ava* and Eis_*Msm*

The APR structure was built using the Sybyl-X software and minimized to 0.01 kcal/mol by the Powell method, using Gasteiger-Hückel charges and the Tripos force field. The coordinates of the Eis_*Ava* and Eis_*Msm* (PDB codes: 2OZG and 3SXN, respectively) were downloaded from the Protein Data Bank website. Except for CoA, the H_2O molecules and all other substructures were removed from the two proteins. In the structure of Eis_*Ava* (PDB code: 2OZG), the selenium atoms were replaced manually by sulfur. An acetyl group was added to CoA in both Eis_*Ava* and Eis_*Msm*. The surrounding protein residues were then kept frozen and the AcCoA was subjected to energy minimization using the Steepest descent method. Hydrogen atoms were added and the energy of the Eis_*Ava*-AcCoA and Eis_*Msm*-AcCoA complexes were minimized using the Amber force fields with Amber charges. The energy-optimized APR was docked into the AG-binding site in both Eis_*Ava*-AcCoA and Eis_*Msm*-AcCoA minimized complexes using GOLD.[15] The parameters were set as the default values for GOLD. The maximum distance between hydrogen bond donors and acceptors for hydrogen bonding was set to 3.5 Å. After docking, the first-ranked conformation of APR in both cases was merged into the corresponding APR-free Eis_*Ava*-AcCoA and Eis_*Msm*-AcCoA complexes. The new Eis_*Ava*-APR-AcCoA and Eis_*Msm*-APR-AcCoA complexes were subsequently subjected to energy

minimization using the Amber force fields with Amber charges. During the energy minimization, the structure of APR and residues within an 8-Å radius were allowed to move. The remaining residues were kept frozen in order to save calculation time. The energy minimization, in both cases, was performed using the Powell method with a 0.05 kcal/mol energy gradient convergence criterion and a distance dependent dielectric function. The Eis_*Ava*-APR-AcCoA and Eis_*Msm*-APR-AcCoA complexes are depicted in Figure A1-3I and J.

Characterization of Eis_*Ava* Steady-State Kinetic Parameters

For several AGs tested in this study (APR, KAN, NEA, NEO, and PAR) the kinetic parameters (K_m and k_{cat}) were determined for acetylation by Eis_*Ava* in reactions (200 µL) with a fixed AcCoA (0.5 mM) concentration. For NEA, NEO, and PAR the AG concentration range used was 0, 20, 50, 100, 250, and 500 µM. Due to their higher K_m values, APR and KAN were tested in a higher concentration range of 0, 50, 250, 500, 1000, and 2000 µM. The reaction mixture containing DTNB (2 mM), Tris-HCl (50 mM, pH 7.5 adjusted at rt), and Eis_*Ava* (0.25 µM) was initiated by the addition of AG. Reactions were monitored by taking readings every 20 seconds for 20 minutes at 25 °C. For determination of the K_m and k_{cat} parameters, a non-linear regression fit to the Michaelis-Menten dependence was carried out by using Sigma Plot 11.0 software (Systat Software Inc.; San Jose, CA) (Figure A1-7and Table A1-2). The same procedure was used to determine the kinetic parameters (K_m and k_{cat}) reported in Table A1-2 for Eis_*Mtb* and Eis_*Msm*.

Monitoring of NEA Acetylation by TLC

The eluent system utilized for all TLCs of NEA reactions was 92.5:7.5/MeOH:NH4OH. The TLCs showing various mono-, and di-acetylated NEA along with their respective Rf values are depicted in Figure A1-8. The mono-acetylated standards 2'-, 3-, and 6'-acetyl-NEA were obtained by using AAC(2')-Ic, AAC(3)-IV, and AAC(6')-Ie/APH(2")-Ia, respectively. These enzymes were overexpressed and purified as previously reported.[1, 16] The acetylation reactions (30 µL) by these enzymes were run at rt in MES buffer (50 mM, pH 6.0 adjusted at rt) (AAC(3)-IV and AAC(6')) or in potassium phosphate buffer (100 mM, pH 7.0 adjusted at rt) (AAC(2')-Ic) with NEA (5 mM, 1 eq), AcCoA (25 mM, 5 eq) and AAC enzyme (10 µM). After overnight incubation at rt, the protein was precipitated by addition of MeOH (30 µL; 20 minutes, -20 °C). The protein was pelleted by centrifugation (13,000 rpm, rt, 10 minutes). An aliquot (5 µL) of the

supernatant was diluted with MeOH (5 µL) before loading onto SiO2 TLC plates. The di-acetylation control reactions (30 µL) were carried out in a similar fashion with the addition of the second AAC enzyme (10 µM) and an additional portion of AcCoA (25 mM, 5 eq) after a 24 hour incubation period.

The reactions with Eis_*Ava* (30 µL) were run in Tris-HCl (50 mM, pH 7.5 adjusted at rt) with AcCoA (25 mM, 5 eq), NEA (5 mM, 1 eq), and Eis_*Ava* (10 µM). At different time points: 0 (before Eis_*Ava* addition), 5, 10, 30, 120 minutes, and overnight, aliquots (5 µL) were removed and processed as described above before loading onto SiO2 TLC plates.

Determination by NMR of NEA Amines Acetylated by Eis_*Ava*

To determine which two positions of NEA are acetylated by Eis_*Ava*, a reaction was designed for characterization in solution by NMR without the need for purification. A reaction (200 µL) containing NEA (10 mM, 1 eq), AcCoA (50 mM, 5 eq), and Eis (10 µM) in Tris-HCl (50 mM, pH 7.5 adjusted at rt) containing 12% D_2O was prepared and allowed to proceed to completion as monitored by TLC. After 24 h, 1H, gCOSY, zTOCSY, and gHSQC NMR experiments were performed on the reaction solution, applying PURGE solvent suppression to establish the positions of acetylation. LCMS was used to confirm the structure identification. Proton connectivity was assigned using zTOCSY, gCOSY, and gHSQC spectra. Representative spectra for 1,2'-di-acetyl-NEA are provided in Figure A1-10.

To establish unambiguously the two acetylated positions on the NEA scaffold, the NMR spectra of 1,2'-di-acetyl-NEA were compared to those of a standard of pure NEA [(10 mM in Tris-HCl (50 mM, pH 7.5 adjusted at rt) containing 12% D_2O], which was prepared identically to the reaction mixture, but omitted AcCoA and Eis_*Ava* (Table A1-3). Representative spectra for the standard reaction solution of NEA in the absence of AcCoA and Eis_*Ava* are provided in Figure A1-9.

Inhibition of Eis_*Ava*

IC_{50} values were measured by the UV-Vis assay described above. The reactions (200 µL total) were initiated in several steps. The inhibitors 1-5 (Figure A1-11) were dissolved in Tris-HCl (50 mM, pH 8.0 adjusted at rt with 10% DMSO) (100 µL) with a fivefold serial dilution with the highest concentration of inhibitor being 250 µM. A second mixture (50 µL) containing Eis_*Ava* or Eis_*Msm* (2 µM), NEO (400 µM), and Tris-HCl (50 mM, pH 8.0 adjusted at rt) was added to the inhibitor solutions and incubated for 10 minutes at rt. Reactions were initiated by

the addition of a third mixture (50 µL) containing AcCoA (8 mM), DTNB (8 mM), and Tris-HCl (50 mM, pH 8.0 adjusted at rt). The final concentrations of compound 1 ranged from 250 µM to 650 pM. The final concentrations of compounds 2-5 ranged from 200 µM to 102 pM. The assay was run in triplicate and the data was normalized to DMSO control reactions. A Hill plot analysis performed by using Kaleidagraph 4.1 software yielded the IC_{50} values.

UV-Vis Spectrophotometric Assay Determination of Eis_*Ava*, Eis_*Msm*, and Eis_*Mtb* Activity Against KAN and 6'-glycinyl-KAN

The activity of Eis homologues against KAN and 6'-glycinyl-KAN was tested as described for the determination of the AG selectivity profile of Eis_*Ava* described above.

References Appendix 1a:

1. Chen, W., Biswas, T., Porter, V. R., Tsodikov, O. V., and Garneau-Tsodikova, S. (2011) Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB, *Proceedings of the National Academy of Sciences of the United States of America 108*, 9804-9808.
2. Green, K. D., Chen, W., and Garneau-Tsodikova, S. (2012) Identification and characterization of inhibitors of the aminoglycoside resistance acetyltransferase Eis from Mycobacterium tuberculosis, *ChemMedChem 7*, 73-77.
3. Zaunbrecher, M. A., Sikes, R. D., Jr., Metchock, B., Shinnick, T. M., and Posey, J. E. (2009) Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences of the United States of America 106*, 20004-20009.
4. Campbell, P. J., Morlock, G. P., Sikes, R. D., Dalton, T. L., Metchock, B., Starks, A. M., Hooks, D. P., Cowan, L. S., Plikaytis, B. B., and Posey, J. E. (2011) Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis, *Antimicrob Agents Chemother 55*, 2032-2041.
5. Chen, W., Green, K. D., Tsodikov, O. V., and Garneau-Tsodikova, S. (2012) Aminoglycoside Multiacetylating Activity of the Enhanced Intracellular Survival Protein from Mycobacterium smegmatis and Its Inhibition, *Biochemistry*.
6. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.
7. Kim, K. H., An, D. R., Song, J., Yoon, J. Y., Kim, H. S., Yoon, H. J., Im, H. N., Kim, J., Kim do, J., Lee, S. J., Lee, H. M., Kim, H. J., Jo, E. K., Lee, J. Y., and Suh, S. W. (2012) Mycobacterium tuberculosis Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7, *Proceedings of the National Academy of Sciences of the United States of America 109*, 7729-7734.
8. Tischer, R. G. (1965) Pure Culture of Anabaena Flos-Aquae a-37, *Nature 205*, 419-420.

9. Rice, E. W., Messer, J. W., Johnson, C. H., and Reasoner, D. J. (1995) Occurrence of high-level aminoglycoside resistance in environmental isolates of enterococci, *Appl Environ Microbiol 61*, 374-376.
10. Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., and Gicquel, B. (2012) Mycobacterium abscessus: a new antibiotic nightmare, *J Antimicrob Chemother 67*, 810-818.
11. Chen, W., Green, K. D., and Garneau-Tsodikova, S. (2012) Cosubstrate tolerance of the aminoglycoside resistance enzyme Eis from Mycobacterium tuberculosis., *Antimicrob. Agents Chemother. 56*, 5831.
12. Shaul, P., Green, K. D., Rutenberg, R., Kramer, M., Berkov-Zrihen, Y., Breiner-Goldstein, E., Garneau-Tsodikova, S., and Fridman, M. (2011) Assessment of 6'- and 6"'-N-acylation of aminoglycosides as a strategy to overcome bacterial resistance, *Org Biomol Chem 9*, 4057-4063.
13. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.
14. Ellman, G. L. (1958) A colorimetric method for determining low concentrations of mercaptans, *Archives of biochemistry and biophysics 74*, 443-450.
15. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD, *Proteins 52*, 609-623.
16. Green, K. D., Chen, W., Houghton, J. L., Fridman, M., and Garneau-Tsodikova, S. (2010) Exploring the substrate promiscuity of drug-modifying enzymes for the chemoenzymatic generation of N-acylated aminoglycosides, *Chembiochem : a European journal of chemical biology 11*, 119-126.

1b: Unexpected N-acetylation of Capreomycin by Mycobacterial Eis Enzymes[6]

1bi: Abstract

The enhanced intracellular survival (Eis) protein from *Mycobacterium tuberculosis* (Eis_*Mtb*), a regio-versatile N-acetyltransferase active towards many aminoglycosides (AGs), confers resistance to kanamycin A in some cases of extensively drug-resistant tuberculosis (XDR-TB). We assessed the activity of Eis_*Mtb* and of its homologue from *Mycobacterium smegmatis* (Eis_*Msm*) against a panel of anti-tuberculosis (TB) drugs and lysine-containing compounds. Both enzymes acetylated capreomycin but not other non-AG non-lysine-containing drugs tested. Using *Msm*, we also demonstrated for the first time to our knowledge that acetylation of capreomycin results in deactivation of the drug. Eis is a unique acetyltransferase capable of inactivating the anti-TB drug capreomycin, AGs and other lysine-containing compounds.

1bii: Introduction

Causing 1.4 million deaths globally in 2010, tuberculosis (TB) is the second leading pathogen-related cause of death, surpassed only by HIV/AIDS. Although worldwide incidences of TB are beginning to modestly decline, occurrences of drug-resistant TB are increasing rapidly. Drug resistance to TB occurs when infected patients are treated with an improper regimen of antibiotics or when the patients fail to comply with the proper multi-drug regimen, which occurs most frequently in areas with sub-standard TB control programmes. Bacteria that are resistant to at least isoniazid and rifampicin, the two most powerful first-line treatments for TB, are designated as multidrug-resistant TB (MDR-TB). Strains of *Mycobacterium tuberculosis* (*Mtb*) are classified as extensively drug-resistant (XDR) if they are shown to be resistant to rifampicin and isoniazid, to one of the fluoroquinolones and to at least one of the second-line anti-TB drugs such as kanamycin A, amikacin or capreomycin. The cure rates for MDR- and XDR-TB are ~50% and ~30%, respectively.

In one-third of kanamycin A-resistant *Mtb* clinical isolates from a large and diverse set, the resistance to kanamycin A was shown to be a result of up-regulation of the enhanced intracellular survival (eis) gene due to mutations in its promoter.[1, 2] The Eis protein is a unique

[6] Portions of this appendix are from published work: Houghton, J.L.; Green, K.D.; Pricer, R.E.; Mayhoub, A.S.; Garneau-Tsodikova, S. Unexpected N-acetylation of capreomycin by mycobacterial Eis enzymes. *J. Antimicrob. Chemother*. **2013**, 68, 800-805 and have been edited to reflect my contributions which included purifying protein used in the study and setting up enzymatic reactions for the testing of various drugs.

aminoglycoside (AG) acetyltransferase (AAC)[2] that inactivates AGs by unusual regio-versatile acetylation of AGs at multiple amine positions.[3] Specifically, kanamycin A and amikacin, which are both multi-acetylated by Eis,[3-6] are two important second-line anti-TB drugs. The structural and functional characterization of Eis from *Mtb* (Eis_*Mtb*) has been the focus of several recent studies.[2, 3, 5, 7-12] Eis homologues can be found in mycobacteria as well as in several non-mycobacterial prokaryotes. We have shown that Eis homologues from *Mycobacterium smegmatis* (Eis_*Msm*) and *Anabaena variabilis* (Eis_*Ava*) also act as a regio-versatile AG acetyltransferase (AAC) and that they can be inhibited by Eis_*Mtb* inhibitors.[4, 6] Additionally, it was recently reported that Eis_*Msm* and Eis_*Mtb* are also capable of acetylating lysine residues of dual-protein phosphatase 16/mitogen-activated protein kinase phosphatase-7 (DUSP16/MPK-7).[5]

The broad specificity of Eis towards a large number of AGs, its acetylation function towards some proteins and previous reports of acetylation by other AACs of drugs other than AGs, including the second-line anti-TB drug ciprofloxacin,[13] prompted us to ask whether Eis is capable of conferring resistance to other non-AG anti-TB drugs. As Eis was shown to acetylate protein lysine residues,[5] in this study, we sought to explore the potential of Eis to acetylate capreomycin, a cyclic peptide antibiotic with a β-lysine side chain (commercially *Ava*ilable as a mixture of capreomycin IA and capreomycin IB; particularly effective against *Mtb* and used as a second-line anti-TB drug. We demonstrated that Eis_*Mtb* and Eis_*Msm* are both capable of acetylating capreomycin IA and capreomycin IB. The position of acetylation on capreomycin was confirmed as the ε-amine of the β-lysine side chain using Eis_*Mtb*.

1biii: Results and Discussion

Exploration of Non-AG Anti-TB Drugs as Potential Substrates of Eis[7]

In addition to capreomycin, we chose other first- and second-line non-AG anti-TB drugs as candidate Eis acetylation substrates, as follows. Previous studies have shown that

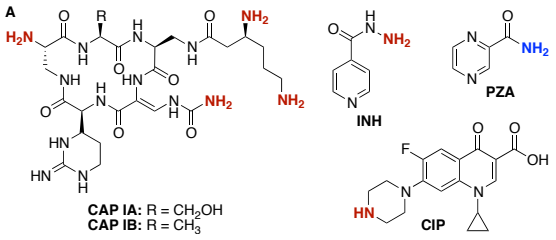


Figure A1-13 Structures of non-AG anti-TB drugs tested against Eis proteins.
CAP = capreomycin; INH = isoniazid; PZA = pyrazinamide; CIP = ciprofloxacin. The amine functionalities of CAP, CIP, and INH that could potentially be acetylated by Eis are highlighted in bold red. The amide functionality of PZA in blue is unlikely to be acetylated due to a lack of nucleophilicity

[7] My contributions to this work included testing CAP for acetylation with Eis_*Mtb* and Eis_*Msm* using the Ellman method described and aiding in making Figure A1: 14.

isoniazid (Figure A1-13) becomes acetylated by arylamine N-acetyltransferase 2 on the hydrazide functional group, demonstrating its potential for deactivation via acetylation.[14] Pyrazinamide (Figure A1-13) was chosen for its structural similarity to isoniazid, allowing comparison of isoniazid with a similar drug that is not likely to be acetylated. Ciprofloxacin was selected because its acetylation by another resistance-conferring AAC was previously reported.[13]

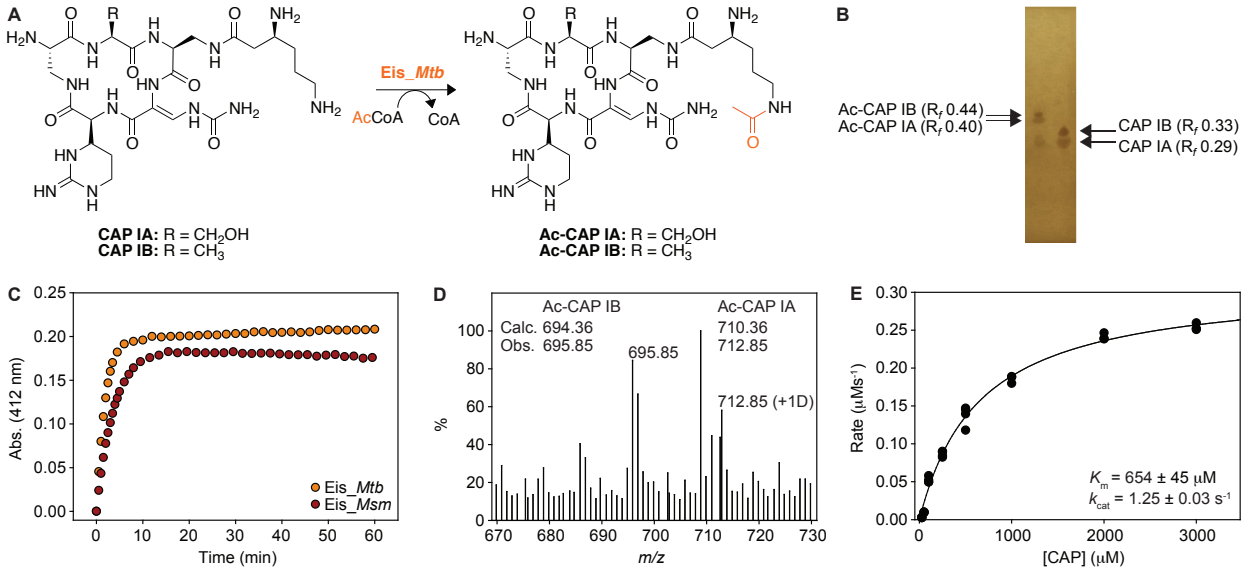


Figure A1-14 6′-N-Acetylation of Capreomycin (CAP) by Eis.
(a) Scheme showing the conversion of CAP IA and CAP IB into 6′-N-acetyl-CAP IA (Ac-CAP IA) and 6′-N-acetyl-CAP IB (Ac-CAP IB), respectively, by Eis_*Mtb*. (b) TLC showing the Ac-CAP IA and Ac-CAP IB products (left-hand lane) generated by Eis_*Mtb* using CAP (mixture of CAP IA and CAP IB) (right-hand lane) with 2.5 eq of AcCoA. (c) Spectrophotometric assay plot monitoring the conversion of CAP into Ac-CAP by Eis_*Mtb* (black circles) and Eis_*Msm* (grey circles). (d) Mass spectrum of Ac-CAP (a mixture of Ac-CAP IA and Ac-CAP IB). (e) Michaelis–Menten analysis of the Eis_*Mtb*-catalyzed N-acetylation of CAP.

We first tested acetylation of capreomycin, ciprofloxacin, isoniazid and pyrazinamide by Eis proteins from *Mtb* and *Msm* by using a UV-Vis spectrophotometric assay. We found that Eis_*Mtb* and Eis_*Msm* did not acetylate ciprofloxacin, isoniazid and pyrazinamide to any observable levels. For pyrazinamide, this result is not surprising, given the lack of nucleophilic amines likely to undergo acetylation. In contrast, we observed efficient acetylation of capreomycin by both Eis_*Mtb* and Eis_*Msm* (Figure A1-14A and C). Even though Eis can acetylate AGs at multiple positions, we confirmed, by TLC (Figure A1-14B) and LCMS (Figure A1-14D), that both capreomycin IA and capreomycin IB were converted to mono-acetylated products. Steady-state kinetic experiments (Figure A1-14E) yielded the following values of Michaelis–Menten parameters for capreomycin acetylation by Eis_*Mtb*: K_m=654±45 μM and k_{cat}=1.25±0.03 s−1. Both K_m and k_{cat} values are slightly greater than the respective values for AG acetylation by Eis_*Mtb* or Eis_*Msm*.[3-5] The catalytic efficiency (k_{cat}/K_m) of 1911±139 M−1

s−1 is equivalent to or better than that for many of the AGs previously tested,[3, 4] suggesting that a somewhat lower affinity of Eis for capreomycin than for AGs is over-compensated by the more efficient acetyl transfer onto capreomycin or a quicker product release.

1biv: Conclusion

We have assessed the potential of Eis proteins from *Mtb* and *Msm* to acetylate various first- and second-line non-AG anti-TB drugs, finding that capreomycin is the only non-AG anti-drug tested to date that can be modified by either protein. These observations expand upon our understanding of Eis proteins as acetyltransferases, specifically in terms of their ability to modify a broad range of small molecules as well as proteins; all known acetylations occur at a primary amine. It was also shown that acetylation of capreomycin inactivates the drug against *Msm* and supports the potential role of acetylation in resistance to capreomycin.

1bv: Experimental

The acetylation of the non-AG anti-TB drugs isoniazid, pyrazinamide, ciprofloxacin and capreomycin was explored using the previously described Ellman method,[4, 7] monitoring the reaction of coenzyme A (CoA) released by the Eis enzymes with Ellman's reagent (DTNB). Briefly, Eis (from *Mtb* or *Msm*, 0.5 μM) was added to a mixture containing Tris (50 mM, pH 8.0, adjusted at room temperature), AcCoA (500 μM), anti-TB drug or lysine-containing compound (100 μM) and DTNB (2 mM) to initiate reactions (200 μL total volume). The reaction progress at 25°C was monitored at 412 nm (ε412=14150 M−1 cm−1) by taking measurements every 30 seconds for 1 hour (Figure 1c and Figure 2a and b).

References Appendix 1b:

1. Campbell, P. J., Morlock, G. P., Sikes, R. D., Dalton, T. L., Metchock, B., Starks, A. M., Hooks, D. P., Cowan, L. S., Plikaytis, B. B., and Posey, J. E. (2011) Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis, *Antimicrob Agents Chemother 55*, 2032-2041.
2. Zaunbrecher, M. A., Sikes, R. D., Jr., Metchock, B., Shinnick, T. M., and Posey, J. E. (2009) Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences of the United States of America 106*, 20004-20009.
3. Chen, W., Biswas, T., Porter, V. R., Tsodikov, O. V., and Garneau-Tsodikova, S. (2011) Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB, *Proceedings of the National Academy of Sciences of the United States of America 108*, 9804-9808.

4. Chen, W., Green, K. D., Tsodikov, O. V., and Garneau-Tsodikova, S. (2012) Aminoglycoside Multiacetylating Activity of the Enhanced Intracellular Survival Protein from Mycobacterium smegmatis and Its Inhibition, *Biochemistry*.
5. Kim, K. H., An, D. R., Song, J., Yoon, J. Y., Kim, H. S., Yoon, H. J., Im, H. N., Kim, J., Kim do, J., Lee, S. J., Lee, H. M., Kim, H. J., Jo, E. K., Lee, J. Y., and Suh, S. W. (2012) Mycobacterium tuberculosis Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7, *Proceedings of the National Academy of Sciences of the United States of America 109*, 7729-7734.
6. Pricer, R. E., Houghton, J. L., Green, K. D., Mayhoub, A. S., and Garneau-Tsodikova, S. (2012) Biochemical and structural analysis of aminoglycoside acetyltransferase Eis from Anabaena variabilis, *Mol Biosyst 8*, 3305-3313.
7. Green, K. D., Chen, W., and Garneau-Tsodikova, S. (2012) Identification and characterization of inhibitors of the aminoglycoside resistance acetyltransferase Eis from Mycobacterium tuberculosis, *ChemMedChem 7*, 73-77.
8. Lella, R. K., and Sharma, C. (2007) Eis (Enhanced Intracellular Survival) Protein of Mycobacterium tuberculosis Disturbs the Cross Regulation of T-cells, *Journal of Biological Chemistry 282*, 18671-18675.
9. Roberts, E. A., Clark, A., McBeth, S., and Friedman, R. L. (2004) Molecular Characterization of the eis Promoter of Mycobacterium tuberculosis, *Journal of Bacteriology 186*, 5410-5417.
10. Samuel, L. P., Song, C. H., Wei, J., Roberts, E. A., Dahl, J. L., Barry, C. E., Jo, E. K., and Friedman, R. L. (2007) Expression, production and release of the Eis protein by Mycobacterium tuberculosis during infection of macrophages and its effect on cytokine secretion, *Microbiology 153*, 529.
11. Shin, D.-M., Jeon, B.-Y., Lee, H.-M., Jin, H. S., Yuk, J.-M., Song, C.-H., Lee, S.-H., Lee, Z.-W., Cho, S.-N., Kim, J.-M., Friedman, R. L., and Jo, E.-K. (2010) Mycobacterium tuberculosis Eis Regulates Autophagy, Inflammation, and Cell Death through Redox-dependent Signaling, *PLoS Pathog 6*, e1001230.
12. Wei, J., Dahl, J. L., Moulder, J. W., Roberts, E. A., O'Gaora, P., Young, D. B., and Friedman, R. L. (2000) Identification of a Mycobacterium tuberculosis Gene That Enhances Mycobacterial Survival in Macrophages, *Journal of Bacteriology 182*, 377-384.
13. Vetting, M. W., Park, C. H., Hegde, S. S., Jacoby, G. A., Hooper, D. C., and Blanchard, J. S. (2008) Mechanistic and Structural Analysis of Aminoglycoside N-Acetyltransferase AAC(6′)-Ib and Its Bifunctional, Fluoroquinolone-Active AAC(6′)-Ib-cr Variant, *Biochemistry 47*, 9825-9835.
14. Sim, E., Walters, K., and Boukouvala, S. (2008) Arylamine N-acetyltransferases: From Structure to Function, *Drug Metabolism Reviews 40*, 479-510.

:

1c: Comparative Study of Eis-like Enzymes from Pathogenic and Nonpathogenic Bacteria.[8]

1ci: Abstract

Antibiotic resistance is a growing problem worldwide. Of particular importance is the resistance of *Mycobacterium tuberculosis* (*Mtb*) to currently *Ava*ilable antibiotics used in the treatment of infected patients. Up-regulation of an aminoglycoside (AG) acetyltransferase, the enhanced intracellular survival (Eis) protein of *Mtb* (Eis_*Mtb*), is responsible for resistance to the second-line injectable drug kanamycin A in a number of *Mtb* clinical isolates. This acetyltransferase is known to modify AGs, not at a single position, as usual for this type of enzyme, but at multiple amine sites. We identified, using *in silico* techniques, 22 homologues from a wide variety of bacteria, that we then cloned, purified, and biochemically studied. From the selected Eis homologues, 7 showed the ability to modify AGs to various degrees and displayed both similarities and differences when compared to Eis_*Mtb*. In addition, an inhibitor proved to be active against all homologues tested. Our findings show that this family of acetyltransferase enzymes exists in both mycobacteria and non-mycobacteria and in both pathogenic and nonpathogenic species. The bacterial strains described herein should be monitored for rising resistance rates to AGs.

1cii: Introduction

Resistance to aminoglycosides (AGs, Figure A1-15) has been a growing problem with numerous species of bacteria, driving the development of novel antibiotics and research into the causes of resistance. *Enterococcus faecium*, *Staphylococcus a*ureus, *Klebsiella p*neumoniae, *Acinetobacter b*aumannii, *Pseudomonas ae*ruginosa, and *Enterobacter* species, all together known as the "ESKAPE"

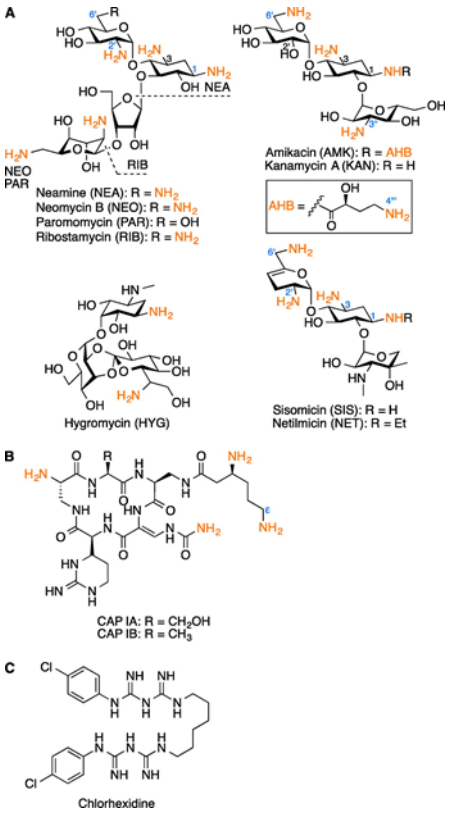


Figure A1-15 Chemical Structures of Drugs Used in this Study.
(A) all AGs found to be modified by Eis homologues, (B) the polypeptide capreomycin (sold as a mixture of CAP IA and CAP IB), and (C) an Eis_*Mtb* inhibitor, chlorhexidine.

[8] Portions of this appendix come from Green, K.D.; Pricer, R.E.; Stewart, M.N.; Garneau-Tsodikova, S. Comparative Study of Eis-like Enzymes from Pathogenic and Nonpathogenic Bacteria. *ACS Infect. Dis.* **2015**, *1* (6), 272-283. I contributed to the cloning, purification, and testing/analysis of various enzymes presented in this work.

pathogens, have been particularly difficult to treat.[8] The most common mode of resistance to AGs is their inactivation by AG-modifying enzymes (AMEs), such as the AG acetyltransferases (AACs), AG phosphotransferases (APHs), and AG nucleotidyltransferases (ANTs). AACs, comprising the majority of these AMEs, are known for their monoacetylating specificities toward the 2′-, 3-, or 6′-positions of a variety of AGs. More recently, a unique AAC, the enhanced intracellular survival (Eis) protein of *Mycobacterium tuberculosis* (*Mtb*), was found capable of multiacetylating AGs.[2] It was originally shown that it is the up-regulation of the eis gene that is responsible for resistance to kanamycin A (KAN) in a large number of clinical isolates,[11] and since then, numerous additional papers on mutation in the eis promoter have been published.[12-14]

We have shown that multiacetylation by Eis completely inactivates AGs.[15, 16] We have also performed studies monitoring the effect that different modifications have on a second modification and its catalyzing enzyme.[16] We found that Eis from *Mtb* (Eis_*Mtb*) is capable of multiacetylating AGs,[2] and we observed a similar multiacetylating behavior when studying Eis from *Mycobacterium smegmatis* (Eis_*Msm*1),[9] Eis from *Anabaena variabilis* (Eis_*Ava*),[1] and the Eis homologue from *Bacillus anthracis* (Eis_*Ban*).[5] AACs are known to be cosubstrate promiscuous,[15] although we have shown that Eis_*Mtb* has a much more limited cosubstrate tolerance.[4] In addition, we have studied the metabolically possible competition between acetyl-CoA (AcCoA) and n-propionyl-CoA (ProCoA) and the effects it has on multiacetylation of AGs.

Although the cellular role of Eis is still unclear, there have been several proposed functions and substrates for this enzyme. Eis_*Mtb* has been found in the sera of infected patients, suggesting it may play a role in the persistence of bacterial infection.[17] The eis protein product that is secreted from the organism has been proposed to modulate the host inflammatory response.[18] Additionally, T-cells treated with Eis_*Mtb* have an inhibited ERK1/2, JAK pathway, which subsequently produces TNF-α and IL-4, thus disrupting the cross-regulation of T-cells.[19] This effect of aiding the survival of the *Mtb* bacteria in macrophages is thought to be associated with Eis_*Mtb*'s extreme stability under harsh conditions.[20] Suh and co-workers have shown that this modulation of the host immune system may happen through the acetylation of DUSP16/MKP-7 by Eis_*Mtb*[3] and have performed docking studies that corroborate this hypothesis.[21]

In an effort to better understand Eis's role in bacteria and attempt to understand the evolution that occurred with this gene in *Mtb* to modify AGs along with its other substrates, we looked to other bacterial species containing Eis homologues. Through a rigorous *in silico* search, we found Eis homologues in a variety of mycobacteria and non-mycobacteria (Figure 2). Herein, we examine several properties of these enzymes and compare them to Eis_*Mtb*. We present *in silico* results for bacterial species containing Eis homologues. We cloned and purified 18 new Eis homologues. We tested these proteins against a small group of antibacterial compounds checking for any acetylation activity. For each AG-active enzyme, we determined the kinetic parameters for a representative of each structural class of AG, kanamycin (KAN) for 4,6-disubstituted deoxystreptamine (DOS) and PAR for 4,5-disubstituted DOS. For combinations producing a positive UV–vis test, we determined the degrees of acetylation and n-propionylation. We also explored the cosubstrate tolerance of AG-active enzymes and examined the competition of AcCoA and ProCoA binding in the GNAT domain. In an effort to identify broad or specific inhibitors of these enzymes, a compound known to inhibit Eis_*Mtb*[6] was tested against the AG-active homologues.

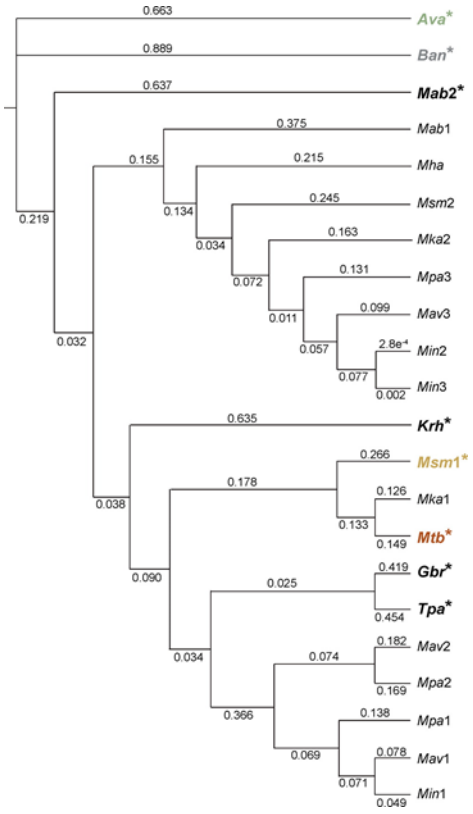


Figure A1-16 Phylogenetic Tree Analysis of the 22 Eis Homologues Studied Based on Sequence. The numbers are proportional to the distance of the evolutionary branches of the tree. The abbreviated names of the tree correspond to the following species and gene: *Anabaena variabilis* ATCC 29413 (*Ava*); *Bacillus anthracis* 34F2 Sterne strain (*Ban*); *Gordonia bronchialis* ATCC 25592 (*Gbr*); *Kocuria rhizophila* ATCC 9341 (*Krh*); *Mycobacterium abscessus* ATCC 19977 genes *MAB4124* (*Mab1*) and *MAB4532c* (*Mab2*); *Mycobacterium avium* subsp. *avium* ATCC 25291 genes *MaviaA2_07418* (*Mav1*), *MaviaA2_15865* (*Mav2*), and *MaviaA2_*00691 (*Mav3*); *Mycobacterium hassiacum* DSM 44199 (*Mha*); *Mycobacterium intracellulare* ATCC 13950 genes *MintA_24600* (*Min1*), *MintA_21739* (*Min2*), and *OCU_01560* (*Min3*); *Mycobacterium kansassii* ATCC 12478 genes *Mkan A1_16037* (*Mka1*) and *Mkan A1_05800* (*Mka2*); *Mycobacterium parascrofulaceum* ATCC BAA-614 genes *ZP_06852004* (*Mpa1*), *ZP_06848356* (*Mpa2*), and *ZP_06849426* (*Mpa3*); *Mycobacterium smegmatis* MC2 155 genes *MSMEG_3513* (*Msm1*) and *MSMEG_4540* (*Msm2*); *Mycobacterium tuberculosis* H37Rv (*Mtb*); *Tsukamurella paurometabola* ATCC 8368 (*Tpa*). Active enzymes are indicated with an asterisk. The percent identity of each homologue compared to that of Eis_*Mtb* is presented in the Table A1-5.

1ciii: Results and Discussion

In Silico Analysis of Potential Eis Homologues

Table A1-5 NCBI Nucleotide and Protein Sequence Numbers, Alternate Gene Names, and Identity to Eis_Mtb for Eis Homologues.

Protein	NCBI nucleotide sequence #	NCBI protein sequence #	Alternate gene identifier	% Identity to Eis_*Mtb*
Eis_*Ava* [a]	NC_007413.1	YP_325469		18
Eis_*Ban* [a]	NC_005945.1	YP_029001		18
Eis_*Gbr* [a]	NC_013441.1	YP_003272441		41
Eis_*Krh* [a]	NC_010617.1	YP_001855041	KRH_11880	34
Eis_*Mab*1	NC_010397.1	YP_001704851	MAB4124	33
Eis_*Mab*2	NC_010397.1	YP_001705255	MAB4532c	29
Eis_*Mav1* [b]	NZ_ACFI01000055.1	ZP_05216001*	MaviaA2_07418	35
Eis_*Mav2* [b]	NZ_ACFI01000149.1	ZP_05217647	MaviaA2_15865	33
Eis_*Mav3* [b]	NZ_ACFI01000006.1	ZP_05214704*	MaviaA2_00691	32
Eis_*Mha* [b]	AMRA01000079.1	EKF23162		33
Eis_*Min1* [b]	NZ_ABIN01000322.1	ZP_05228133*	MintA_24600	36
Eis_*Min2* [b]	NZ_ABIN01000280.1	ZP_05227568*	MintA_21739	32
Eis_*Min3* [b]	NC_016946.1	YP_005335697*	OCU_01560	32
Eis_*Mka1* [b]	NZ_ACBV01000045.1	ZP_04749484*	Mkan A1_16037	74
Eis_*Mka2* [b]	NZ_ACBV1000010.1	ZP_04747464*	Mkan A1_05800	32
Eis_*Mpa1*	NZ_GG770554.1	ZP_06852004*		34
Eis_*Mpa2*	NZ_GG770564.1	ZP_06848356*		34
Eis_*Mpa3*	NZ_GG770553.1	ZP_06849426*		34
Eis_*Msm*	NC_008596.1	YP_887817	MSMEG_3513	53
Eis_*Msm2*	NC_008596.1	YP_888812	MSMEG_4540	34
Eis_*Mtb* [a]	AF144099.1	AAF03768.1		100
Eis_*Tpa* [a]	NC_014158.1	YP_003645809		40

[a] Indicates Eis homologues found to be active in acetylating AGs in this work.
[b] Indicates Eis homologues found to be inactive in acetylating AGs in this work.
* These proteins (all of which we found to be inactive) were recently marked as obsolete in NCBI

A BLAST search against Eis_*Mtb* revealed a myriad of Eis homologues. Eis proteins were found in the genome of several mycobacteria including *Mycobacterium avium* subsp. *avium* ATCC 25291 (*Mav*), *Mycobacterium kansasii* ATCC 12478 (*Mka*), *Mycobacterium abscessus* ATCC 19977 (*Mab*), *Mycobacterium hassiacum* DSM 44199 (*Mha*), *Mycobacterium intracellulare* ATCC 13950 (*Min*), and *Mycobacterium parascrofulaceum* ATCC BAA-614 (*Mpa*). Because Eis was originally found in *Mtb*[22] and later in *Msm*,[9] it is not surprising Eis homologues were found in other mycobacteria. Curiously, some species were found to have two (*Mab* and *Mka*) or three (*Mav*, *Min*, and *Mpa*) Eis homologues, and upon further examination of the genomic sequence, a second Eis-like protein was found in *Msm*. Many of the listed mycobacteria are involved in atypical mycobacterial infections,[23] and *Mab* has been noted, particularly, as being "a new antibiotic nightmare";[24] however, this is likely due to mutations in the rRNA,[25] yet drug combinations are being investigated for this pathogenic species.[26] *Mav–Min* complex, *Mka*, and *M. fortuitum* extracts have been analyzed for AMEs, and an unspecified AAC was found to act on KAN, tobramycin (TOB), and neomycin B (NEO) in *Mka* and *M. fortuitum*.[27] Outside the mycobacterial genus, three other species were found to harbor Eis homologues in their respective

genomes: *Gordonia bronchialis* ATCC 25592 (*Gbr*), *Kocuria rhizophila* ATCC 9341 (*Krh*), and *Tsukamurella paurometabola* ATCC 8368 (*Tpa*). As the name suggests, *Gbr* is a bacterium that often affects the lungs; the sample used here was originally isolated from the sputum of a patient

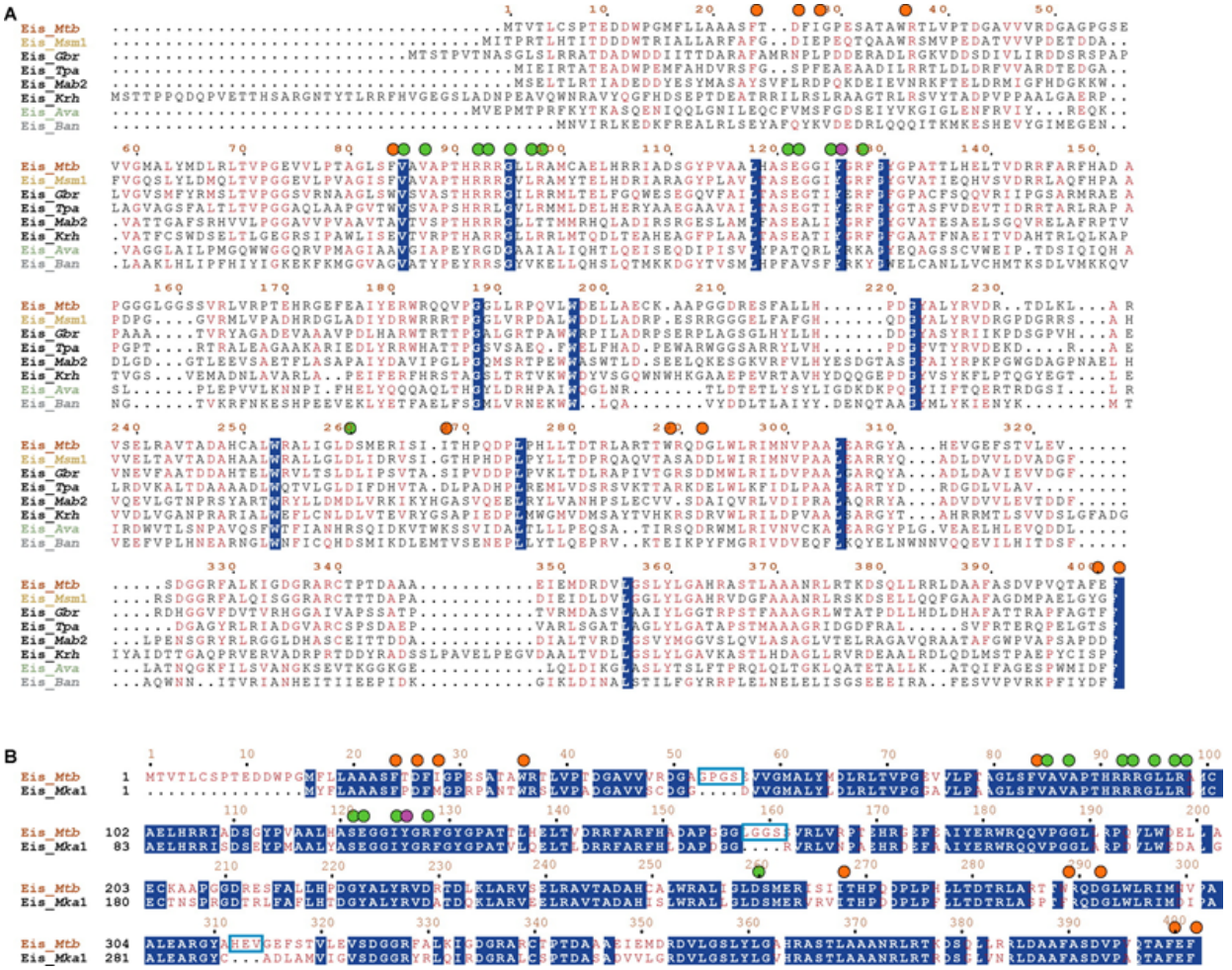


Figure A1-17 Alignment of All Eis Homologue Protein Sequences Found to be Active in Acetylating AGs. Alignment of Eis_*Mtb* and its closest homologue in amino acid sequence tested, Eis_*Mka1*. Turquoise boxes indicate areas of difference for the *Mka1* homologue, which are hypothesized to be important in protein–protein interactions of the Eis monomers. For both alignments, the catalytic tyrosine (Y) residue is indicated with a purple circle, AG-binding site residues are indicated with orange circles, and AcCoA-binding site residues are indicated with green circles. All of the residues marked with colored circles were originally identified by site-directed mutagenesis of the Eis_*Mtb* enzyme.[2] White letters with blue background indicate amino acids completely conserved throughout the sequences, whereas red residues indicate amino acids highly conserved throughout the sequences.

with cavitary pulmonary disease. *Tpa* was originally isolated from the microflora of a bedbug,[28] and *Krh* is a soil microorganism,[29] but is now appearing more often in the clinic.[30, 31] A phylogenetic tree (Figure A1-16) of these proteins was constructed, revealing that these enzymes break into two major groups and three outliers (Eis_*Ava*, Eis_*Ban*, and Eis_*Mab*2). One of the groups consists of strictly proteins from mycobacteria that align well (35% exact amino acid

match) with each other and include Eis_*Mab*1, Eis_*Mav*3, Eis_*Mha*, Eis_*Min*2, Eis_*Min*3, Eis_*Mka*2, Eis_*Msm*2, and Eis_*Mpa*3. The second group contains all of the remaining Eis homologues that align poorly (7% exact amino acid match). Taking a closer look at their amino acid sequences, we found most homologues to have the AcCoA binding residues and the catalytic residues in the appropriate position. However, the residues proposed to bind the AG are found to a lesser extent in the homologous enzymes.

AG Activity of Eis Enzymes

All Eis homologues were tested against a small library of antibiotic compounds including amikacin (AMK), chloramphenicol (CAM), capreomycin (CAP), ciprofloxacin, hygromycin (HYG), KAN, neamine (NEA), NEO, netilmicin (NET), norfloxacin, paromomycin (PAR), ribostamycin (RIB), sisomicin (SIS), spectinomycin (SPT), and streptomycin (STR) to establish the substrate profile of these enzymes (Table A1-6). CAM was included in our study because there are known CAM-modifying enzymes;[32] CAP was included because of the recent discovery that Eis_*Mtb* can modify this cyclic peptide;[33] and the fluoroquinolones were included as an AAC mutant was found to acetylate these drugs.[34] Enzymes purified from *Mav*, *Mka*, *Min*, and *Mpa*, along with one enzyme from *Mab* (Eis_*Mab*1) and the new *Msm* enzyme (Eis_*Msm*2), were all found to be AG inactive. These enzymes also all proved to be less soluble than the AG active homologues, with Eis_*Mka*2 being completely insoluble and remaining in the pelleted cell debris under the conditions used to purify Eis_*Mtb*. Eis_*Mha* had excellent expression and solubility, but was also AG inactive under conditions identical to those used in acetylating assays with Eis_*Mtb*. Eis_*Gbr*, Eis_*Krh*, Eis_*Mab*2, and Eis_*Tpa* all showed various degrees of activity with the AGs tested (AMK, HYG, KAN, NEA, NEO, NET, PAR, RIB, and SIS) with the exception of Eis_*Mab*2, which proved to be inactive when tested with AMK. Although the explanation of inactivity of Eis_*Mab*2 toward AMK would be highly speculative without a crystal structure, the 14 Eis homologues showing complete AG inactivity can be analyzed on the basis of their amino acid sequence. Eis_*Mab*1, Eis_*Mav*3, Eis_*Mha*, Eis_*Min*2, Eis_*Min*3, Eis_*Mka*2, Eis_*Msm*2, and Eis_*Mpa*3 all find themselves in one main branch of the phylogenetic tree (Figure A1-16). It is highly possible that they have some other unknown biological purpose that differs from that of Eis_*Mtb*. One residue that is found to differ from Eis_*Mtb* in all of these AG-inactive Eis homologues is F24 (Eis_*Mtb* numbering used throughout). In the majority of these homologues, an arginine residue is found in place of F24. This arginine would repel AGs

Table A1-6 Comparison of Number of Modifications for Reactions of Eis Homologues with AGs and AcCoA or ProCoA

AcCoA.

AG	Eis_*Ava*[b]	Eis_*Ban*[c]	Eis_*Gbr*	Eis_*Krb*	Eis_*Mab*	Eis_*Msm1*[b,d]	Eis_*Mtb*[b,c,d,e]	Eis_*Tpa*
AMK	Tri	Di	Di	Di	X	Tri	Tri	Tri
HYG	X	Di	Mono	Mono	Mono	Mono	Di	Di
KAN	Mono	Di	Mono	Di	Di	Di	Di	Di
NEA	Tri	Tri	Tri	Tri	Tri	Tri	Tri	Tri
NEO	Tri	Tri	Tri	Tetra	Tetra	Tri	Tri	Tetra
NET	Mono	Mono	Mono	Di	Di	Di	Di	Di
PAR	Tri	Tri	Tri	Tri	Tri	Tri	Di	Tri
RIB	Tri	Tri	Di	Di	Di	Tri	Tri	Di
SIS	Di	Di	Tri	Tri	Di	Tri	Tri	Di

ProCoA

AG	Eis_*Ava*[b]	Eis_*Ban*[c]	Eis_*Gbr*	Eis_*Krb*	Eis_*Mab*	Eis_*Msm1*[b,d]	Eis_*Mtb*[b,c,d,e]	Eis_*Tpa*
AMK	Mono	Di	Di	MOno	X	Di	Di	Di
HYG	X	Di	Mono	Mono	Mono	Mono	Mono	Mono
KAN	Mono	Mono	Mono	Mono	Di	Di	Di	Di
NEA	Di	Tri	Di	Di	Tri	Di	Di	Tri
NEO	Di	Tri	Tri	Di	Di	Tri	Tri	Di
NET	Mono	Mono	Mono	Mono	Di	Mono	Mono	Di
PAR	Di	Di	Di	Di	Tri	Di	Mono	Tri
RIB	Di	Tri	Di	Di	Di	Di	Di	Di
SIS	Di	Di	Di	Di	Di	Mono	Mono	Di

X indicates that the AG is not a substrate for the enzyme tested.
[b]Data from [1]
[c] Data from [5]
[d] Data from [9, 10]
[e] Data from [4]

under physiological conditions. The remaining inactive enzymes are in the second branch of the phylogenetic tree and include the AG-active homologues Eis_*Mtb*, Eis_*Msm*1, Eis_*Gbr*, Eis_*Krh*, and Eis_*Tpa*. Whereas the homologues from the first branch were missing an important AG binding residue, Eis_*Mav*1, Eis_*Mav*2, Eis_*Min*1, Eis_*Mpa*1, and Eis_*Mpa*2 are all missing one residue for AcCoA binding, R128, which is replaced by either glycine or isoleucine. This lack of charge at this position may not allow AcCoA to bind with affinity similar to that with Eis_*Mtb*. Most surprising is the AG inactivity of Eis_*Mka*1, which has 74% homology to Eis_*Mtb*. The only explanation we can propose from the sequence alignment is that Eis_*Mka*1 is 26 amino acids shorter and is missing amino acid sequences from G52 to S55, from L158 to S161, and from H312 to V314. Upon examination of the Eis_*Mtb* crystal structure (PDB ID 3R1K)[2] these residues form three sites for protein–protein interaction, possibly responsible for oligomerization (Figure A1-17B) as each missing sequence is in an area where protein–protein interactions are likely to be. This might indicate that Eis_*Mka* does not form hexamers. It has been shown that the Eis_*Mtb* hexameric form is

required for activity.[20] It is known that the Eis_*Mtb* hexamer is extremely resilient to a variety of conditions,[20] and based on what we can hypothesize from the Eis_*Mka*1 data, this area could be heavily involved in activity, despite its remote location in relationship to the enzyme active site. Crystal structures of all enzymes would allow for a better understanding of inactivity.

There are structures *Ava*ilable for three Eis_*Mtb* homologues, Eis_*Ava*, Eis_*Ban*, and Eis_*Msm*1, which have been previously published with a comparative structural analysis.[1, 2, 5, 9, 10, 35] Shown in Figure A1-18 are the four currently known crystal structures of Eis_*Mtb* and its three crystallized homologues. The cartoon overlay of the four crystal structures (Figure A1-18A) shows only minor changes in the backbone of the structure with only one or two

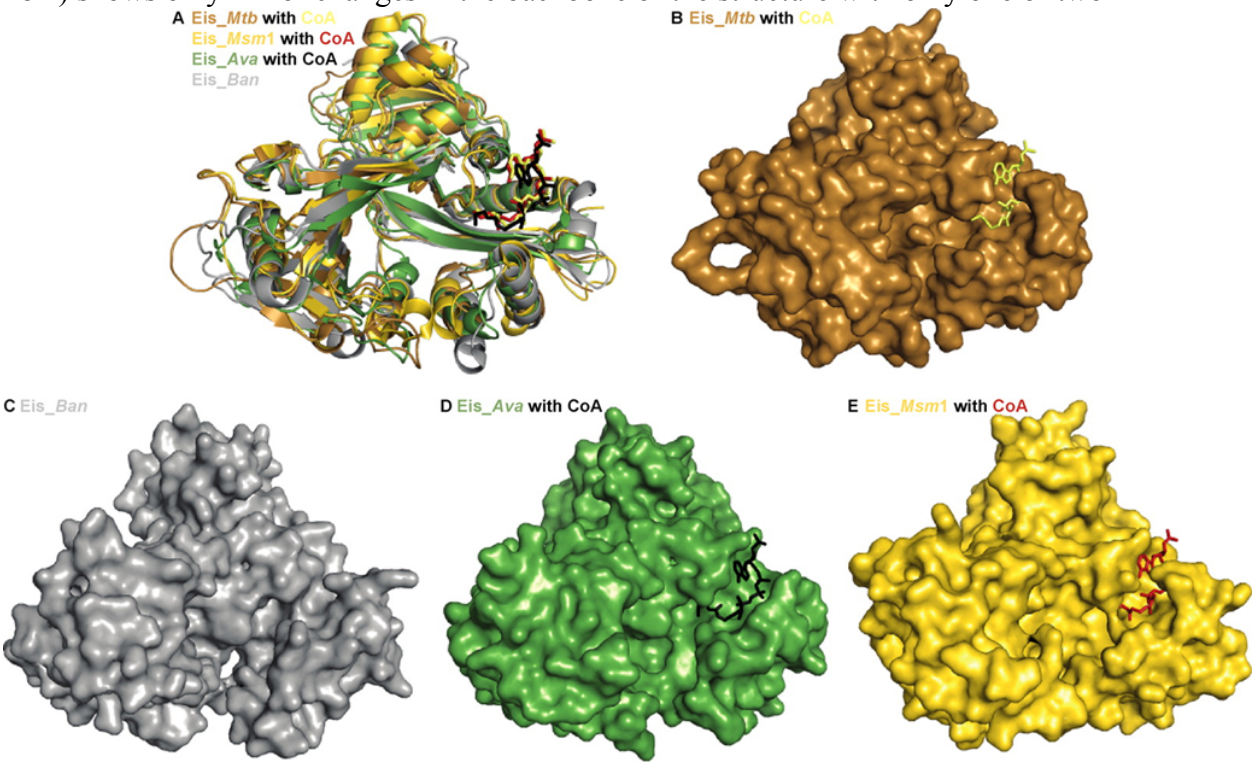


Figure A1-18 Comparison of Crystal Structures of Eis Homologues.
(Only one monomer of these hexameric structures is shown for clarity.) Panel A depicts the overlay of one monomer from each of the four structures and shows that there is little variation in the overall structure of this family of enzymes. Surface cartoons shown in panels B–E demonstrate subtle differences in the active sites of the Eis homologues from (B) *Mycobacterium tuberculosis* (shown in brown with CoA as yellow stick), (C) *Bacillus anthracis* (shown in gray in the absence of CoA), (D) *Anabaena variabilis* (shown in green with CoA as black stick), and (E) *Mycobacterium smegmatis* (shown in gold with CoA as red stick). Eis_*Mtb* (PDB 3R1K, 1.95 Å),[2] Eis_*Msm1* (PDB 3SXN, 2.03 Å),[3] Eis_*Ava* (PDB 2OZG, 2.00 Å),[1] and Eis_*Ban* (PDB 3N7Z, 2.75 Å).

misaligned α-helices and a few disordered loops. The differences are significantly more apparent when looking at the space-filled models (Figure A1-18B–E), where the AG-binding site (left of the AcCoA-binding site in Figure A1-18B,D,E) varies. The variance in the AG-binding site is likely why each enzyme reacts with the same AGs differently. Recent work has shown

that the AG TOB can bind to Eis_*Mtb* in two distinct orientations, placing two amino groups in the appropriate configuration for acetylation at the 6′- and 3″-positions.[10] This emphasizes the fact that the AG-binding site can dramatically affect not only the identity of AGs that can bind to the Eis enzymes but also the various orientations that any given AG can adopt in the AG-binding pocket.

Acetylation of the Polypeptide Antibiotic CAP

We recently showed that in addition to modifying AGs, Eis_*Mtb* is also capable of acetylating the nonribosomal peptide CAP.[33] In addition to the panel of AGs (Figure A1-15A), we also tested the seven homologues able to acetylate AGs for their ability to modify CAP. Just as we observed for the AGs, there were several differences within this family of enzymes. Eis_*Krh*, Eis_*Mab*, Eis_*Msm*, Eis_*Mtb*, and Eis_*Tpa* all had very similar initial rates in the UV–vis assay, when CAP was used as the substrate (Figure A1-20). This suggests that these enzymes all catalyze the modification of CAP similarly. Eis_*Ava* and Eis_*Gbr* both showed significantly reduced reactions rates when modifying CAP, suggesting that these two homologues bind CAP less effectively than the previous five or have a lower catalytic efficiency. Interestingly, Eis_*Ban* did not react with CAP at all. These data supplement the AG data showing that the differences observed in the crystal structures of Eis proteins (Figure A1-18) can significantly alter the ability of said enzymes to modify different amino groups.

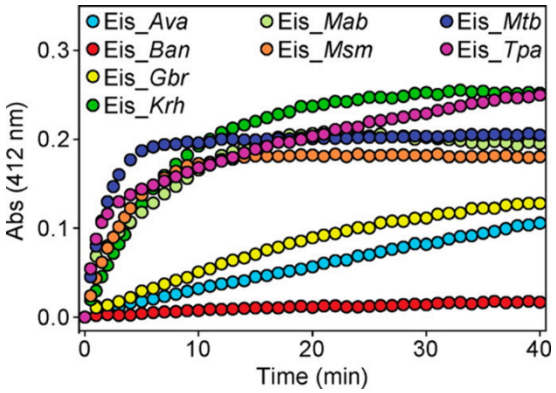


Figure A1-20 UV–Vis Curves Monitoring the Acetylation of CAP by Various Eis Homologues.

Kinetic Parameters of AG-Active Eis Homologues for KAN and PAR

Kinetic parameters were determined for Eis_*Gbr*, Eis_*Krh*, Eis_*Mab*2, and Eis_*Tpa* and compared to those previously published for Eis_*Ava*,[1] Eis_*Ban*,[5] Eis_*Msm*1,[9] and Eis_*Mtb*.[2] Kinetic values are presented in Table A1-7 along with a comparison of the catalytic turnover (k_{cat}) and catalytic efficiency (k_{cat}/K_m) to Eis_*Mtb*. KAN bound with the highest affinity to Eis_*Tpa*, binding 9-fold better than Eis_*Mtb*. Meanwhile, PAR displayed 2.5-fold tighter binding to Eis_*Ban* over Eis_*Mtb*. KAN reacted at a rate of around 0.7 reaction per second with Eis_*Gbr*, 1.28 times faster than Eis_*Mtb*, whereas PAR had the quickest turnover time with

Eis_*Tpa* at 5.92 reactions per second, ~44-fold faster than Eis_*Mtb*. The most efficient combinations were with AGs and Eis_*Tpa*, showing ~3- and ~27-fold increases in efficiency for KAN and PAR when compared to Eis_*Mtb*, respectively. Additionally, Eis_*Krh* showed poor kinetic parameters for KAN and heightened parameters for PAR. This is somewhat surprising considering that Eis_*Krh* is roughly 10-kDa larger than Eis_*Mtb*, with two extraneous peptide chains throughout the amino acid sequence (Figure A1-18A). Although an in-depth kinetic study was not performed for each enzyme, there are no data that indicate that these enzymes do not react similarly to Eis_*Mtb*, having a random sequential mechanism.[36] Whereas Eis_*Mtb* may have evolved to modify AGs due to selective pressure, it seems that the homologues from *Gbr* and *Tpa* have a greater affinity for AGs and these species should be watched for development of AG resistance.

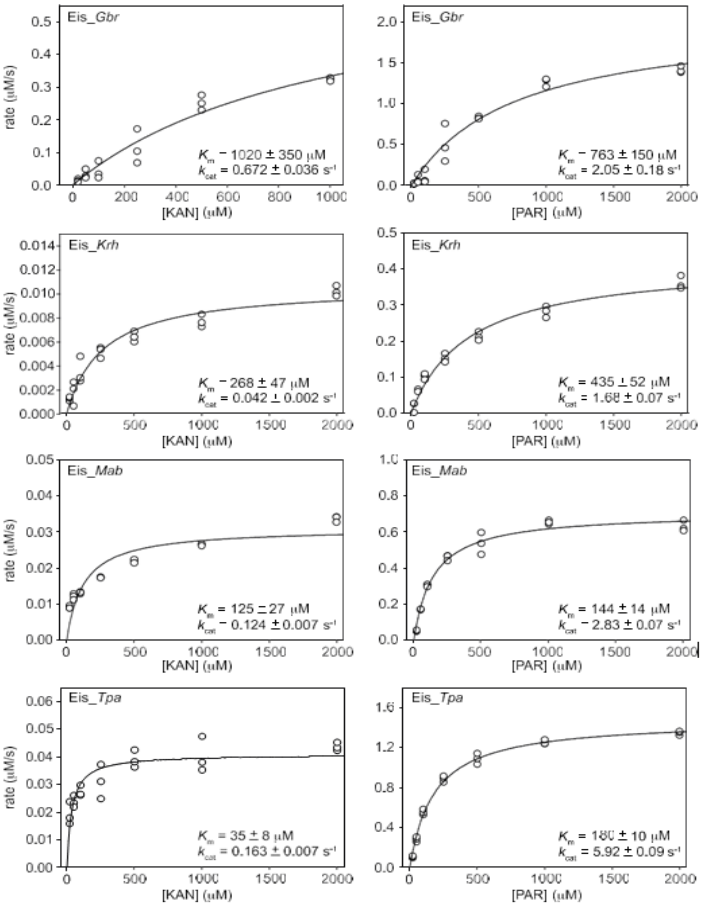


Figure A1-19 Michaelis-Menten Kinetic Plots for All Enzyme-AG Combinations Not Previously Published.
KAN and PAR were used as representatives of the two major structural classes of AGs.

Number of Acetylations

To investigate the extent to which the AG-active Eis homologues can acetylate AGs, we determined the number of sites acetylated by these enzymes on AMK, HYG, KAN, NEA, NEO, NET, PAR, RIB, and SIS (Table A1-6). All Eis homologues were found to acetylate NEA three times. HYG, KAN, and NET all showed mono- or diacetylation by the enzymes, whereas AMK, PAR, RIB, and SIS were observed to be either di- or triacetylated. Interestingly, NEO was acetylated between two and four times. This tetra modification was observed previously only for Eis_*Mtb* and Eis_*Msm*1 with TOB. Only Eis_*Krh*, Eis_*Mab*, and Eis_*Tpa* were found to tetra-acetylate NEO. This suggests that the binding pockets or at least the AG-binding portion of

Table A1-7 Kinetic Parameters Determined for Eis Homologues by Using AcCoA with Different AGs.

Eis homologue	AG	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat}*Xxx*/ k_{cat}*Mtb*	k_{cat}/K_m*Xxx*/ k_{cat}/K_m*Mtb*
Eis_*Ava*[a]	KAN	1002 ± 43	0.205 ± 0.004	205 ± 9	0.39	0.13
	PAR	237 ± 36	0.183 ± 0.013	773 ± 132	1.35	0.48
Eis_*Banb*	KAN	57 ± 9	0.135 ± 0.008	2368 ± 399	0.26	1.49
	PAR	44 ± 11	0.019 ± 0.001	432 ± 110	0.14	0.35
Eis_*Gbr*	KAN	1020 ± 350	0.672 ± 0.036	659 ± 229	1.28	0.41
	PAR	763 ± 150	2.05 ± 0.18	2687 ± 579	15.07	2.17
Eis_*Krh*	KAN	268 ± 47	0.042 ± 0.002	157 ± 28	0.08	0.1
	PAR	435 ± 52	1.68 ± 0.07	3862 ± 406	12.35	3.12
Eis_*Mab*2	KAN	125 ± 27	0.124 ± 0.007	992 ± 377	0.24	0.62
	PAR	144 ± 14	2.83 ± 0.07	19653 ± 1972	20.8	15.9
Eis_*Msm*1[a,c]	KAN	665 ± 42	0.360 ± 0.010	541 ± 37	0.68	0.34
	PAR	738 ± 158	0.235 ± 0.033	318 ± 82	1.73	0.26
Eis_*Mtb*[a,c]	KAN	330 ± 40	0.526 ± 0.027	1594 ± 250	1	1
	PAR	110 ± 21	0.136 ± 0.012	1236 ± 260	1	1
Eis_*Tpa*	KAN	35 ± 8	0.163 ± 0.007	4657 ± 1083	0.31	2.92
	PAR	180 ± 10	5.92 ± 0.09	32889 ± 1894	43.53	26.6

[a] Data from [1]
[b] Data from [5]
[c] Data from [7]

these enzymes are more flexible in their substrate binding. Eis_*Gbr* showed the lowest number of acetylations, in total having monoacetylated three of the AGs tested (HYG, KAN, and NET), suggesting a more limited enzyme–AG interaction. The only other mycobacterial enzyme, Eis_*Mab*, showed similar patterns of acetylation to those of Eis_*Mtb* and Eis_*Msm*1, but was more limited in its substrate profile, not accepting AMK and only diacetylating RIB.

Cosubstrate Tolerance of Eis Enzymes

The AG-active Eis homologues were tested with a variety of acyl-CoAs to establish their cosubstrate promiscuity (Table A1-8). As we have previously shown, Eis_*Mtb* had a very limited cosubstrate tolerance, transferring acyl groups from AcCoA, ProCoA, crotonyl-CoA, and malonyl-CoA only.[4] In addition to the four acyl-CoAs used by Eis_*Mtb*, n-butyryl-CoA, n-hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, lauroyl-CoA, myristoyl-CoA, palmitoyl-CoA, and i-valeryl-CoA were tested as potential cosubstrates. All enzymes could only transfer acyl groups from AcCoA and ProCoA, and Eis_*Krh* could utilize malonyl-CoA to a lesser extent. These data confirm that despite the poor sequence alignment, the CoA-binding pocket of the Eis homologues studied is very likely similar to that of Eis_*Mtb*.

Table A1-8 Comparison of the Cosubstrate Tolerance for Eis Homologues Using NEO as the Substrate.

acyl-CoA	Eis_*Ava*[b]	Eis_*Ban*	Eis_*Gbr*	Eis_*Krh*	Eis_*Mab*	Eis_*Msm1*	Eis_*Mtb*[c]	Eis_*Tpa*
Ac	√	√	√	√	√	√	√	√
Pro	√	√	√	√	√	√	√	√
n-butanoyl	×	×	×	×	×	×	×	×
n-hexanoyl	×	×	×	×	×	×	×	×
octanoyl	×	×	×	×	×	×	×	×
decanoyl	×	×	×	×	×	×	×	×
lauroyl	×	×	×	×	×	×	×	×
myristoyl	×	×	×	×	×	×	×	×
palmitoyl	×	×	×	×	×	×	×	×
crotonyl	×	×	×	×	×	×	√	×
malonyl	×	×	×	√	×	×	√	×
i-valeryl	×	×	×	×	×	×	×	×

× indicates that the acyl-CoA is not a cosubstrate of the Eis enzyme tested
√ indicated that the acyl-CoA is a cosubstrate of the Eis enzyme tested
[b] Data from [1]
[c] Data from [4, 7]

Degree of n-Propionylation

The number of n-propionylations each enzyme catalyzes with the different AGs was also analyzed to glean more insight into the size of the catalytic pocket. ProCoA, being slightly larger than AcCoA, is not likely to be used as efficiently or result in as many modifications. As expected, all n-propionylations were of equal number or less than the number of acetylations observed for the Eis homologue–AG combinations tested. n-Propionylation was equal to acetylation for Eis_*Mab*2 in all cases, for seven of nine combinations using Eis_*Ban* (all except KAN and PAR) and Eis_*Gbr* (all except NEA and PAR), six combinations for Eis_*Tpa* (all except AMK, KAN, and NEO), four cases for Eis_*Msm*1 (HYG, KAN, NEO, and NET), and two combinations for Eis_*Ava* (NET and SIS) and Eis_*Krh* (HYG and RIB). These results were equal to or greater than the two cases where n-propionylation and acetylation were equal with Eis_*Mtb* (KAN and NEO).[4]

For those combinations that displayed a number of n-propionylations lower than that of acetylations, additional experiments were performed to establish if Eis homologues could further acetylate these n-propionylated AGs. Sequential acylation reactions in which the enzymes were first allowed to catalyze the n-propionylation of the AG for 24 hours prior to addition of AcCoA were performed (Table A1-10). From these sequential experiments, one could expect to see that (1) no acetyl moiety can be transferred to an n-propionylated AG, indicating that n-propionylation prevents further acetylation; (2) the total number of modifications (acetylation

and n-propionylation) equals that of acetylations when only AcCoA is used, indicating that the AG-binding site is likely sufficiently large to accommodate the n-propionylated AGs in the same orientation as their corresponding acetylated versions for additional acetylations; or (3) the total number of modifications is less than the number of acetylations observed in the presence of AcCoA alone, likely pointing to the fact that the AG-binding pocket is not large enough to accommodate the extra bulkiness of the n-propionyl group on modified AGs in specific orientations in the cavity. In most cases, additional acetylation of n-propionylated AGs did occur. The one exception was that of Eis_*Ban*, for which no additional acetylations were observed after n-propionylation with the two AGs tested, KAN and PAR. Other combinations where the n-propionylation stopped any further modifications included Eis_*Ava* with AMK, NEA, and PAR; Eis_Krh with SIS; and finally, Eis_*Msm*1 with NET, RIB, and SIS. Eis_*Mtb* showed limitation with propionylated NET, RIB, and SIS. Although additional acetylations are possible, these results demonstrate that the inability of AGs to become modified with ProCoA as efficiently as AcCoA may lie in the kinetics of the n-propionyl transfer rather than the bulkiness of the extra methyl group of the cosubstrate. The remaining results are spread between scenarios 2 and 3. These data indicate that the loss of modification is due primarily to the binding of the modified AG to the enzyme (scenario 3) or that n-propionylation has no effect on acetylation (scenario 2). All observed combinations are presented in Table A1-10.

Cosubstrate Competition of Eis Enzymes

Due to the abundance of both AcCoA and ProCoA in the cellular environment, a competition experiment was performed to gain insight into the kinetic mechanism of the enzymes (Table A1-10). AGs were incubated with Eis homologues and a 1:1 mixture of AcCoA and ProCoA and the results analyzed. From these competition experiments, one could expect to observe (1) no n-propionylation only acetylation, (2) no acetylation only n-propionylation, (3) mixed n-propionylation and acetylation, or (4) lower levels of acetylation than incubation with AcCoA alone, due to a competition with the ProCoA for the CoA-binding site. There were no observed cases that follow the second hypothesized observation, meaning that the presence of AcCoA limits the amount of n-propionylation that can occur. The most common observation was scenario 3, where there is a majority of acetylation with some n-propionylation (usually mono-), followed closely by scenario 4, where less acetylation is observed, likely due to reaction inhibition due to an abundance of CoA-containing molecules. From all these data arise a few

interesting cases including NEO with Eis_*Ban*, by which under pure reaction, either AcCoA or ProCoA, only trimodification was observed. When AcCoA and ProCoA are combined, the maximum modification observed is still trimodification (mono-Ac-di-Pro-NEO); however, only diacetyl or dipropionyl-NEO is observed for NEO molecules modified by a single CoA species. Another interesting outcome was the reaction of Eis_*Ban* with PAR; in this reaction we saw monoacetyl-dipropionylated PAR. In the sequential reaction only dipropionylated PAR was observed. These data suggest that acetylation occurred either first or second and n-propionylation then occurred. Also of interest is the reaction of Eis_*Gbr* with NEO, after which a whole spectrum of modifications was observed. Everything from triacetyl-NEO to tripropionyl-NEO, including the two possible acetyl-propionyl mixed products, was observed as seen in Table A1-10.

Inhibition of Eis Homologues with Eis_*Mtb* Inhibitors

Knowing that the active sites of the Eis homologues are similar enough to acetylate a variety of AGs and CAP, but different enough to display diverse results, we wondered how these similarities and differences would be reflected in the inhibition of enzymatic activity. Chlorhexidine (Figure A1-15C), an inhibitor of Eis_*Mtb*,[6] was tested against all seven AG-modifying Eis homologues (Table A1-9). Although chlorhexidine is used as a topical antibiotic, we used the compound here simply to test if the inhibitor would broadly inhibit all Eis homologues or be specific to Eis_*Mtb* or Eis from mycobacteria. Chlorhexidine displayed a wide range of activities for the Eis homologues, ranging from low nanomolar (Eis_*Mtb*) to micromolar (Eis_*Ava* and Eis_*Ban*). This is an approximate 100-fold range in IC_{50} value, indicating that this inhibitor targets a broad range of Eis homologues. Generally the trend seems to follow that the less homologous the protein is to Eis_*Mtb*, the higher the IC_{50} value, implying a less potent inhibitor. Whereas chlorhexidine would be a poor candidate as an antitubercular drug, it does tell us that inhibitors of Eis_*Mtb* may inhibit other homologues, but an increase in the IC_{50} value could be expected.

Table A1-9 IC_{50} Values of the Known Eis_Mtb Inhibitor Chlorhexidine with Eis Homologues.

Eis species	Chlorhexidine (μM)	IC_{50}*Xxx*/ IC_{50}*Mtb*
Ava	20 ± 7[a]	111
Ban	14 ± 4[b]	77
Gbr	4.4 ± 0.6	23
Krh	17 ± 4	91
Mab	4.4 ± 0.9	23
*Msm*1	1.9 ± 0.4[a]	10
Mtb	0.188 ± 0.030[c]	1
Tpa	7.4 ± 1.6	40

[a] Data from [1]
[b] Data from [5]
[c] Data from [6]

1civ: Conclusion

In summary, the consolidation of mass spectral analysis, kinetic, inhibitory, and activity data tells us that the Eis homologues share activity, cosubstrate tolerance, and modification multiplicity. We have shown that deletion of 13 residues can render an Eis homologue inactive, and an additional 10 kDa can have a limited effect on activity. Although without structural data these hypotheses are speculative, we have learned that Eis homologues can be found in a diverse community of bacteria and that they behave similarly to Eis_*Mtb*. They also have very distinct sequences, kinetic profiles, and inhibition. Whereas the enzymes do not all come from pathogenic bacteria, the similarity of the enzymes encoded in the genomes of these nonpathogenic species may be a warning that although currently the species do not pose a threat, they may become more aggressive in the future.

Table A1-10 Sequential Reactions Using ProCoA Followed by AcCoA and Competition Assays with both AcCoA and ProCoA.

ProCoA → AcCoA[b]

AG	Eis_*Ava*	Eis_*Ban*	Eis_*Gbr*	Eis_*Krh*	Eis_*Mab*	Eis_*Msm*1	Eis_*Mtb*[c]	Eis_*Tpa*
AMK	mono-Ac mono-Pro	ND	ND	mono-Ac-mono-Pro	×	mono-Ac-mono-Pro di-Pro mono-Ac-di-Pro	mono-Ac-di-Pro	di-Pro mono-Ac-di-Pro
HYG	×	ND	ND	ND	ND	ND	mono-Pro	mono-Pro mono-Ac-mono-Pro
KAN	ND	mono-Pro	ND	di-Ac mono-Ac-mono-Pro	ND	ND	di-Pro	ND
NEA	di-Pro	ND	mono-Pro mono-Ac-mono-Pro di-Pro di-Ac-mono-Pro mono-Ac-di-Pro	mono-Ac-mono-Pro	ND	mono-Ac-Mono-Pro di-Pro di-Ac-mono-Pro mono-Ac-di-Pro	mono-Ac-di-Pro	ND
NEO	di-Ac mono-Ac-mono-Pro mono-Ac-di-Pro	ND	ND	mono-Ac-di-Pro	mono-Ac-di-Pro	ND	tri-Pro	di-Ac-Pro mono-Ac-di-Pro tri-Ac-Pro di-Ac-di-Pro
NET	ND	ND	ND	mono-Ac mono-Pro di-Ac mono-Ac-mono-Pro	ND	mono-Pro	mono-Pro	ND
PAR	mono-Ac-mono-Pro di-Pro	di-Pro	mono-Ac-di-Pro	mono-Ac-di-Pro	ND	mono-Ac-mono-Pro di-Pro di-Ac-mono-Pro mono-Ac-di-Pro	mono-Ac-mono-Pro	ND
RIB	mono-Pro di-Ac mono-Ac-mono-Pro di-Pro mono-Ac-di-Pro	ND	ND	ND	ND	di-Pro	di-Pro	ND
SIS	ND	ND	ND	mono-Pro mono-Ac-mono-Pro di-Pro	ND	mono-Pro	mono-Pro	ND

AcCoA + ProCoA[d]

AG	Eis_*Ava*	Eis_*Ban*	Eis_*Gbr*	Eis_*Krh*	Eis_*Mab*	Eis_*Msm*1	Eis_*Mtb*[c]	Eis_*Tpa*
AMK	mono-Ac di-Ac mono-Ac-mono-Pro	di-Ac	mono-Ac-mono-Pro	di-Ac mono-Ac-mono-Pro	×	di-Ac	di-Ac tri-Ac	di-Ac
HYG	×	mono-Ac di-Ac	mono-Ac	mono-Ac	mono-Ac	mono-Ac	mono-Ac mono-Ac-mono-Pro	mono-Ac di-Ac
KAN	mono-Ac	mono-Ac	mono-Ac mono-Pro	di-Ac	di-Ac	di-Ac	di-Ac	di-Ac
NEA	mono-Ac mono-Pro di-Ac tri-Ac	di-Ac tri-Ac di-Ac-mono-Pro mono-Ac-di-Pro	mono-Ac mono-Pro	di-Ac	tri-Ac di-Ac-mono-Pro	di-Ac	tri-Ac di-Ac-mono-Pro	di-Ac tri-Ac
NEO	mono-Ac di-Ac mono-Ac-mono-Pro	di-Ac mono-Ac-mono-Pro di-Pro mono-Ac-di-Pro	di-Ac di-Pro tri-Ac di-Ac-mono-Pro mono-Ac-di-Pro tri-Pro	di-Ac mono-Ac-mono-Pro	tri-Ac di-Ac-mono-Pro tetra-Ac tri-Ac-mono-Pro di-Ac-di-Pro	tri-Ac	tri-Ac tri-Pro	mono-Ac mono-Pro di-Ac mono-Ac-mono-Pro mono-Ac-mono-Pro tri-Ac
NET	mono-Ac mono-Pro mono-Ac-mono-Pro	mono-Ac mono-Pro	mono-Ac mono-Pro	mono-Ac	mono-Ac di-Ac mono-Ac-mono-Pro di-Pro	mono-Ac	mono-Ac di-Ac mono-Ac-mono-Pro	mono-Ac di-Ac mono-Ac-mono-Pro
PAR	di-Ac tri-Ac	di-Ac mono-Ac-di-Pro	di-Ac mono-Ac-mono-Pro	di-Ac	tri-Ac	di-Ac	di-Ac	di-Ac tri-Ac
RIB	di-Ac mono-Ac-mono-Pro	mono-Ac mono-Pro di-Ac	mono-Ac mono-Pro di-Ac	di-Ac	di-Ac	mono-Ac di-Ac	di-Ac	di-Ac
SIS	mono-Ac di-Ac mono-Ac-	mono-Ac di-Ac mono-Ac-mono-Pro	mono-Ac di-Ac	mono-Ac di-Ac mono-Ac-mono-Pro	di-Ac	mono-Pro	mono-Ac di-Ac mono-Ac-mono-Pro	di-Ac

× indicates that the reaction was not tested as the AG was found to not be a substrate of the enzyme
ND indicates that the reaction was not tested as the number of *n*-propionylations was equal to acetylations
[b] reactions with ProCoA were followed by incubation with AcCoA [c] Data from [4] [d] AcCoA and ProCoA were incubated with Eis and AG in a 1:1 ratio.

1cv: Experimental

Phylogenetic Tree Generation

The amino acid sequences of 22 Eis homologues were obtained from NCBI through the Geneious Pro 4.8.5 program to construct a phylogenetic tree (Figure A1-16). The parameters used were as follows: Jukes–Cantor genetic distance and cost matrix Blosum45. The Eis homologue accession numbers from NCBI were YP_325469 (*Ava*), YP_029001 (*Ban*), YP_003272441 (*Gbr*), YP_001855041 (*Krh*), YP_00170485 (*Mab*1), YP_001705255 (*Mab*2), ZP_05216001 (*Mav*1), ZP_05217647 (*Mav*2), ZP_05214704 (*Mav*3), EKF23162 (*Mha*), ZP_05228133 (*Min*1), ZP_05227568 (*Min*2), YP_005335697 (*Min*3), ZP_04749484 (*Mka*1), ZP_04747464 (*Mka*2), ZP_06852004 (*Mpa*1), ZP_06848356 (*Mpa*2), ZP_06849426 (*Mpa*3), YP_887817 (*Msm*1), YP_888812 (*Msm*2), AAF03768.1 (*Mtb*), and YP_003645809 (*Tpa*). The percent identity of these homolgues to Eis_*Mtb* is presented in Table A1-5.

Materials and Instrumentation

Chemically competent *Escherichia coli* (TOP10 and BL21 (DE3)) cells were purchased from Invitrogen (Carlsbad, CA, USA). Genomic DNA (*Krh* and *Tpa*) and bacterial strains (*Gbr*, *Mab*, *Mav*, *Min*, *Mka*, and *Mpa*) were purchased from American Type Culture Collection (ATCC; Manassas, VA, USA). Genomic DNA for *Mha*[37] was a gift from Dr. Nuno Empadinhas (University of Coimbra, Portugal). The pET28a plasmid used in this study was obtained from Novagen (Gibbstown, NJ, USA). All primers used were obtained from Integrated DNA Technologies (Coraville, IA, USA). All cloning enzymes including restriction enzymes, Phusion DNA polymerase, and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA, USA). Sequencing of generated plasmids was performed at the University of Michigan DNA sequencing core. AMK, CAM, CAP, CIP, KAN, NEO, NOR, RIB, SIS, SPT, STR, AcCoA, acyl-CoAs (ProCoA, n-butyryl-CoA, n-hexanoyl-CoA, octanoyl-CoA, decanoyl-CoA, lauroyl-CoA, myristoyl-CoA, palmitoyl-CoA, crotonyl-CoA, malonyl-CoA, and i-valeryl-CoA), and dithionitrobenzoic acid (DTNB) were purchased from Sigma-Aldrich (Milwaukee, WI, USA) and used without further purification. NEA, NET, and PAR were purchased from AK Scientific (Mountain View, CA, USA). All UV–vis assays and kinetic experiments were performed using a multimode SpectraMax M5 plate reader in 96-well plates from Fisher Scientific (Pittsburgh, PA, USA). Liquid chromatography–mass spectrometry (LC-MS) chromatograms were obtained on a Shimazdu LC-MS-2019EV equipped with an SPD-20AV

UV–vis detector and an LC-20AD liquid chromatograph. The PDB structures 2OZG (Eis_*Ava*),[1] 3N7Z (Eis_*Ban*),[15] 3SXN (Eis_*Msm*1),[35] and 3R1K (Eis_*Mtb*)[2] were visualized using PyMOL (The PyMOL Molecular Graphics System, version 1.4.1, Schrödinger, LLC).

DNA Isolation from Bacteria

DNA was isolated from the following bacteria: *M. abscessus* ATCC 19977, *M. kansasii* ATCC 12478, *M. parascrofulaceum* ATCC BAA-614, *M. intracellulare* ATCC 13950, *M. avium* ATCC 25291, and *G. bronchialis* ATCC 25592. Mycobacteria were grown in 7H9 medium with albumin–dextrose–catalase (ADC). One milliliter of turbid culture was pelleted (13000 rpm, 1 minutes, room temperature) and resuspended in TE buffer (400 μL). The bacteria were inactivated with heat (30 minutes, 80 °C). After cooling to room temperature, to avoid clumping of the cells, four to five glass beads were added and the cells were briefly vortexed. Lysozyme (30 μL, 50 mg/mL) was added, and the suspension was further incubated for 1 hour at 37 °C. At this point SDS (10 μL, 10%) and proteinase K (10 μL, 20 mg/mL) were added, and the suspension was gently vortexed, followed by a brief incubation period (15 minutes, 65 °C). NaCl (100 μL, 5 M) and CTAB/NaCl (100 μL, 10% CTAB in 0.7 M NaCl) were added sequentially, and the suspension was further incubated (10 minutes, 65 °C). The mixture was extracted with phenol/chloroform/isoamyl alcohol 25:24:1 (3 × 750 μL). The DNA was precipitated using isopropanol (0.6×volume) and incubated (20 minutes, −20 °C). The DNA was collected by centrifugation (12000 rpm, 10 minutes, room temperature), the isopropanol was removed, and the DNA pellet was washed with ice-cold ethanol (75%, 1 mL) and centrifuged (12000 rpm, 5 minutes, room temperature). The ethanol was removed and allowed to air-dry (10 minutes, room temperature). The DNA was redissolved in TE buffer (100 μL, 30 minutes, 37 °C). G. bronchialis was grown in BHI, and DNA was isolated as above.

Cloning, Expression, and Purification of Eis Proteins

Eis_*Mtb*,[2] Eis_*Msm*1,[9] Eis_*Ban*,[5] and Eis_*Ava*[1] were cloned and purified as previously described. The remaining Eis proteins were cloned using the isolated DNA and Phusion DNA polymerase. The isolated PCR fragments were inserted into a linearized pET28a vector between the NdeI (all genes) and XhoI (Eis_*Gbr*, Eis_*Krh*, Eis_*Mab*1, Eis_*Mab*2, Eis_*Mha*, and Eis_*Tpa*) or HindIII (Eis_*Mav*1, Eis_*Mav*2, Eis_*Mav*3, Eis_*Min*1, Eis_*Min*2–3, Eis_*Mka*, Eis_*Mpa*, and Eis_*Msm*2) restriction sites. The plasmids containing the eis genes were then transformed into *E. coli* TOP10 chemically competent cells. Sequencing of the isolated plasmid revealed perfect

alignment with the reported gene sequences. Plasmids were transformed into *E. coli* BL21 (DE3) chemically competent cells for protein expression. All new Eis enzymes were purified exactly as indicated for Eis_*Mtb*.[2] All proteins were dialyzed in Tris-HCl buffer (50 mM, pH 8.0 adjusted at room temperature) and stored at 4 °C. After purification, 2.2 mg (Eis_*Gbr*), 1.8 mg (Eis_*Krh*), 0.1 mg (Eis_*Mab*1), 0.4 mg (Eis_*Mab*2), 3.8 mg (Eis_*Mav*1), 0.8 mg (Eis_*Mav*2), 0.5 mg (Eis_*Mav*3), 7.0 mg (Eis_*Mha*), 2.6 mg (Eis_*Min*1), 1.2 mg (Eis_*Min*2/3), 1.0 mg (Eis_*Mka*1), 0 mg (Eis_*Mka*2, insoluble), 0.4 mg (Eis_*Mpa*1), 0.5 mg (Eis_*Mpa*2), 0.3 mg (Eis_*Mpa*3), 0.5 mg (Eis_*Msm*2), and 1.9 mg (Eis_*Tpa*) of protein were obtained per liter of culture.

Spectrophotometric Measurements of Acyltransferase Activity

The acyltransferase activity of Eis enzymes was tested against a variety of antibiotics using AcCoA, by monitoring the 2-nitro-5-thiobenzoate (NTB–) produced by the reaction of free CoA with DTNB. Reactions were monitored at 412 nm (ε412 nm = 14150 M–1 cm–1) for 1 h, taking measurements every 30 seconds at 25 °C. Reaction mixtures (200 μL) containing antibiotic (100 μM), AcCoA (500 μM), DTNB (2 mM), and Tris-HCl buffer (50 mM, pH 8.0 adjusted at room temperature) were initiated by the addition of Eis enzyme (0.5 μM). The only non-AG to be modified by Eis enzymes was CAP (Figure A1-15B); UV–vis curves for this antibiotic are presented in Figure A1-12. All enzymes displaying activity against AGs were also tested for cosubstrate promiscuity (Table A1-8). Here acyl-CoAs (ProCoA, n-butyryl-CoA, n-hexanoyl-CoA, octanoyl-CoA, decaonyl-CoA, lauroyl-CoA, myristoyl-CoA, palmitoyl-CoA, crotonyl-CoA, malonyl-CoA, and i-valeryl-CoA) (500 μM) were tested against NEO (100 μM), the most dynamic AG with all AG-active enzymes.

Determination of Number of Acetylations and n-Propionylations as well as Sequential and Competitive Modifications of AGs by Eis Proteins via LC-MS

The extent of acetylation for each AG that was found to be active by UV–vis assays (Figure A1-15A) was determined by using LC-MS. Reactions (30 μL) containing AG (AMK, HYG, KAN, NEA, NEO, NET, PAR, RIB, and SIS, 0.67 mM), AcCoA (3.35 mM), Eis enzyme (5 μM), and Tris-HCl buffer (50 mM, pH 8.0 adjusted at room temperature) were incubated overnight at room temperature. Reactions were quenched by the addition of ice-cold methanol (30 μL) and were chilled (−20 °C) for at least 20 minutes. To remove excess enzyme from the solution, the reactions were then centrifuged (13000 rpm, 10 minutes, room temperature) and

diluted 1:1 with H_2O before loading onto the LC-MS. Samples were run using H_2O (0.1% formic acid). A summary of the level of acetylation is presented in Table A1-6. The n-propionylation of AGs (AMK, HYG, KAN, NEA, NEO, NET, PAR, RIB, and SIS) was also monitored using the above method substituting ProCoA for AcCoA. A summary of n-propionylation is also presented in Table A1-6. Sequential acylations (n-propionylation followed by acetylation) were performed and monitored as above. After the first overnight incubation with ProCoA, AcCoA (3.35 mM) and additional Eis enzyme (5 μM) were added, bringing the total volume up to 40 μL, and incubated for an additional 24 h. A summary of the acylated AG species obtained during sequential experiments is presented in Table A1-10. Competition experiments were also performed as above in the simultaneous presence of both AcCoA (3.35 mM) and ProCoA (3.35 mM). A summary of the acylated AG species obtained during competition experiments is also presented in Table A1-10. Calculated and observed masses for all AGs against all Eis homologue proteins are presented at the end of this section, with each table summarizing findings for a specific Eis homologue. Mass spectra for all experiments for Eis_*Ava*, Eis_*Ban*, Eis_*Gbr*, Eis_*Krh*, Eis_*Mab*2, Eis_*Msm*1, and Eis_*Tpa* are presented in the following order: (i) acetylation, (ii) n-propionylation, (iii) sequential n-propionylation followed by acetylation, and (iv) competition between n-propionylation and acetylation.

Determination of Steady-State Kinetic Parameters of Eis Proteins for AGs

The kinetic parameters (k_{cat} and K_m) were determined for acetylation with KAN and PAR for all enzymes, in reactions (200 μL) with a fixed AcCoA (0.5 μM) concentration. For Eis_*Krh*, Eis_*Mab*2, and Eis_*Tpa*, concentrations of 0, 20, 50, 100, 250, 500, 1000, and 2000 μM were used with both AGs. With Eis_*Gbr* the ranges varied, KAN concentrations were 0, 20, 50, 250, 500, and 1000 μM, and PAR concentrations were 0, 20, 50, 100, 250, 500, 1000, and 2000 μM. The reaction cocktail containing DTNB (2 mM), Tris-HCl buffer (50 mM, pH 8.0 adjusted at room temperature), AcCoA (0.5 μM), and Eis enzyme (0.25 μM) was initiated by the addition of the AGs. The reactions were monitored at 412 nm by taking readings every 20 seconds for 20 minutes at 25 °C. The determination of kinetic parameters was done with a nonlinear regression using Sigma-Plot 11.0 software (Systat Software Inc., San Jose, CA, USA) (Table A1-7).

Inhibition of Eis Enzymes

IC_{50} values for chlorhexidine (Figure A1-15C) were determined as previously reported.[1, 6, 9] Briefly, reactions (200 μL) were initiated in a stepwise manner. Chlorhexidine (100 μL) was

dissolved in Tris-HCl buffer (50 mM, pH 8.0, 10% DMSO) and subjected to a 5-fold serial dilution. To the chlorhexidine solution was added a mixture (50 μL) of Tris-HCl buffer (50 mM, pH 8.0), Eis enzyme (2 μM), and NEO (400 μM) followed by incubation (10 minutes, room temperature). A third reaction mixture consisting of Tris-HCl buffer (50 μM, pH 8.0), AcCoA (2 mM), and DTNB (8 mM) was used to initiate the reactions. The reactions were monitored as described for the kinetic analysis. The determination of IC_{50} values was done with a Hill plot analysis using Kaleidagraph 4.1 software, and the data are summarized in Table A1-9.

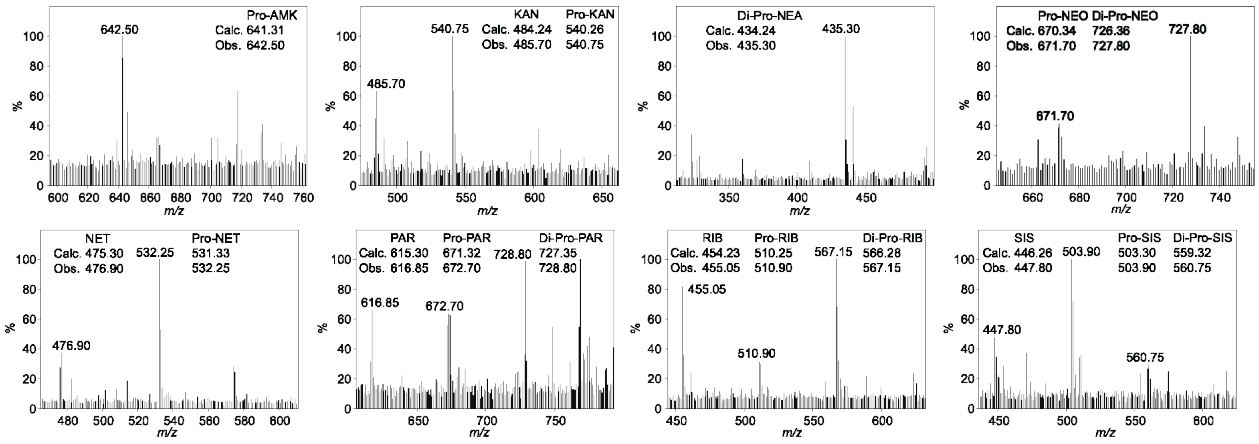
1cvi: Additional Data

Mass results are summarized in a table followed by the mass spectra for each enzyme.

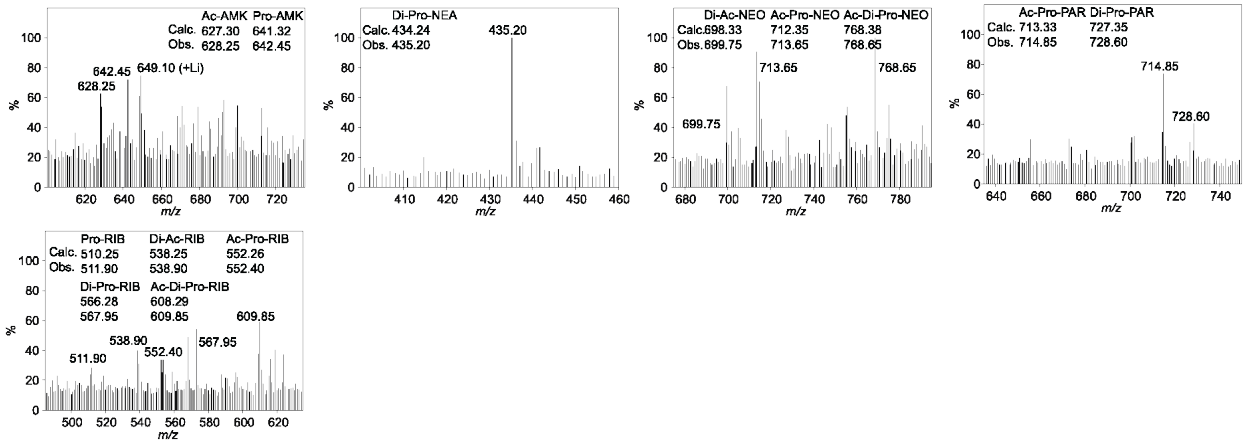
Mass analysis of AGs acylated by Eis *Ava*.

		n-Propionylation		Pro → Ac			Ac + Pro		
AG		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$
AMK	Mono	642.32	642.50	Ac	628.30	628.25	Ac	628.30	628.85
				Pro	642.32	642.45	Di-Ac	670.31	670.65
							Ac-Pro	684.33	683.85
KAN	Mono	541.27	540.75				Ac	527.26	527.10
NEA	Di	435.25	435.30	Di-Pro	435.25	435.20	Ac	365.20	365.95
							Pro	379.22	383.95 (+Li)
							Di-Ac	407.21	407.20
							Tri-Ac	449.22	450.40
NEO	Mono	671.35	671.70	Di-Ac	699.34	699.75	Ac	657.33	657.85
	Di	727.37	727.80	Ac-Pro	713.36	713.65	Di-Ac	699.34	699.70
				Ac-Di-Pro	769.38	768.65	Ac-Pro	713.36	713.45
NET	Mono	532.33	532.25				Ac	518.32	518.40
							Pro	532.33	532.30
							Ac-Pro	574.35	574.30
PAR	Mono	672.33	672.70	Ac-Pro	714.34	714.85	Di-Ac	700.32	701.45
	Di	728.36	728.80	Di-Pro	728.36	728.60	Tri-Ac	742.34	742.80
RIB	Mono	511.26	510.90	Pro	511.26	511.90	Di-Ac	539.26	539.30
	Di	567.29	567.15	Di-Ac	539.26	538.90	Ac-Pro	553.27	553.20
				Ac-Pro	553.27	552.40			
				Di-Pro	567.29	567.95			
				Ac-Di-Pro	609.30	609.85			
SIS	Mono	504.30	503.90				Ac	490.29	490.40
	Di	560.33	560.75				Di-Ac	532.30	532.35
							Ac-Pro	546.31	546.90

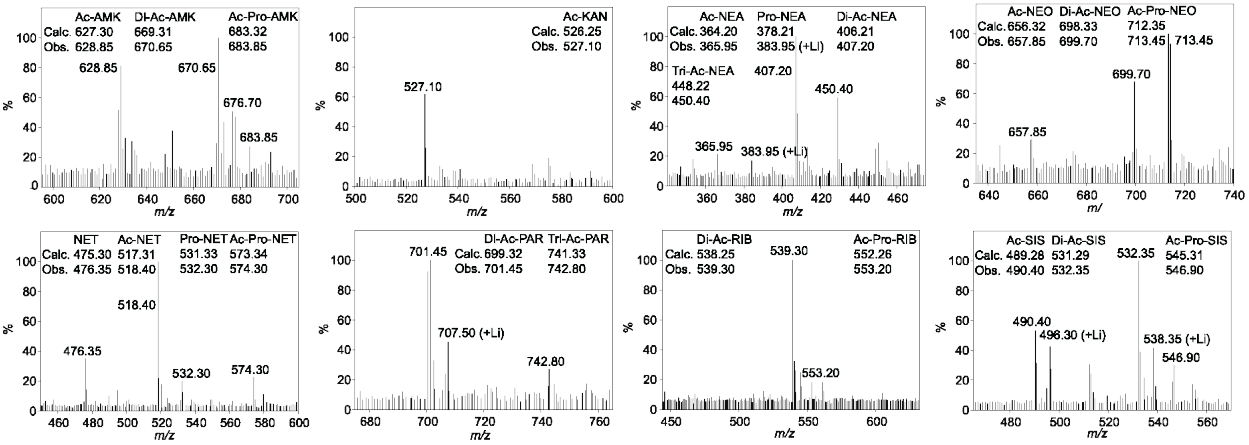
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Ava*.



Mass spectra of AGs that were first *n*-propionylated by Eis_*Ava*, then reacted with AcCoA in the presence of Eis_*Ava*.

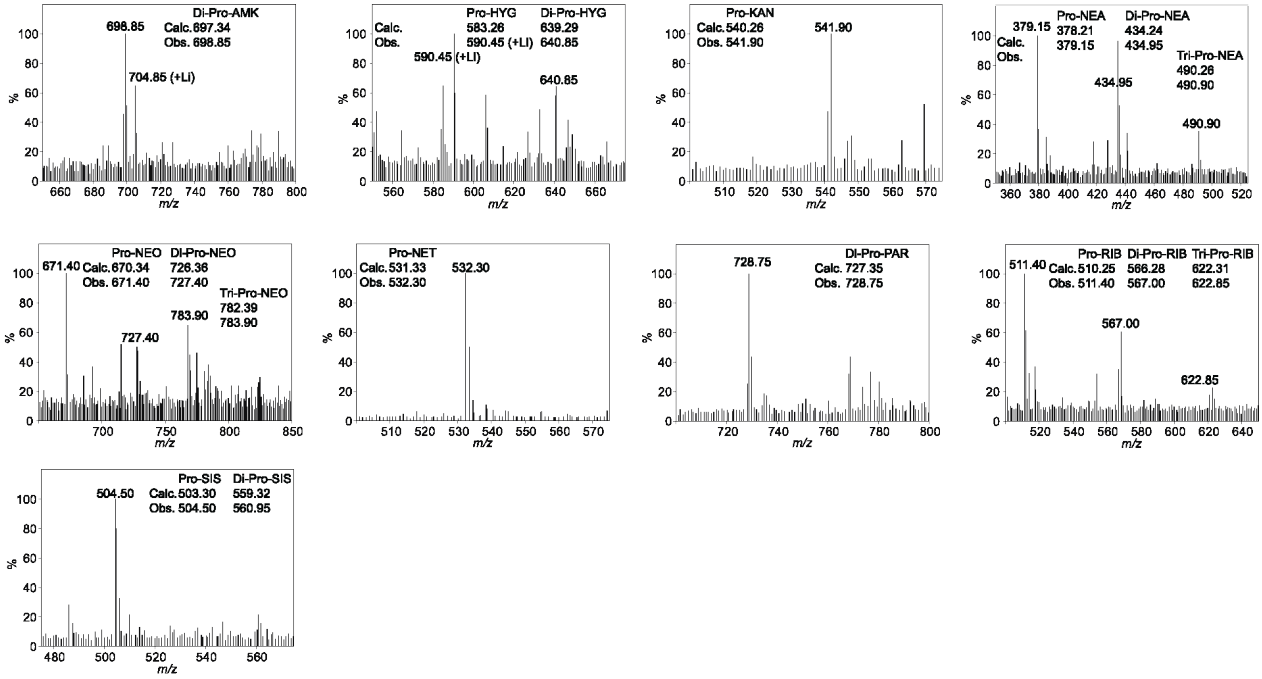


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Ava*.

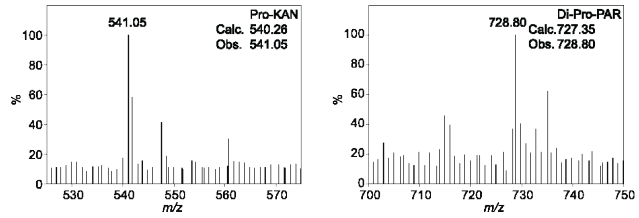
Mass analysis of AGs acylated by Eis *Ban*.

AG		*n*-Propionylation Calc $[M+H]^+$	*n*-Propionylation Obs $[M+H]^+$	Pro → Ac	Pro → Ac Calc $[M+H]^+$	Pro → Ac Obs $[M+H]^+$	Ac + Pro	Ac + Pro Calc $[M+H]^+$	Ac + Pro Obs $[M+H]^+$
AMK	Di	698.35	698.85				Di-Ac	670.31	670.40
HYG	Mono	584.27	590.45 (+Li)				Ac	570.25	576.40 (+Li)
	Di	640.29	640.85				Di-Ac	612.26	612.35
KAN	Mono	541.27	541.90	Pro	541.27	541.05	Ac	527.26	526.90
NEA	Mono	379.22	379.15				Di-Ac	407.21	406.95
	Di	435.25	434.95				Tri-Ac	449.22	448.90
	Tri	491.27	490.90				Di-Ac-Pro	463.24	462.90
							Ac-Di-Pro	477.26	483.80 (+Li)
NEO	Mono	671.35	671.40				Di-Ac	699.34	699.65
	Di	727.37	727.40				Ac-Pro	713.36	713.85
	Tri	783.40	783.90				Di-Pro	727.37	733.65 (+Li)
							Ac-Di-Pro	769.38	768.85
NET	Mono	532.33	532.30				Ac	518.32	518.15
							Pro	532.33	532.45
PAR	Di	728.36	728.75	Di-Pro	728.36	728.80	Di-Ac	700.32	700.60
							Ac-Di-Pro	770.37	770.70
RIB	Mono	511.26	511.40				Ac	497.25	496.60
	Di	567.29	567.00				Pro	511.26	511.90
	Tri	623.31	622.85				Di-Ac	539.26	539.15
SIS	Mono	504.30	504.50				Ac	490.29	490.20
	Di	560.33	560.95				Di-Ac	532.30	531.90
							Ac-Pro	546.31	546.25

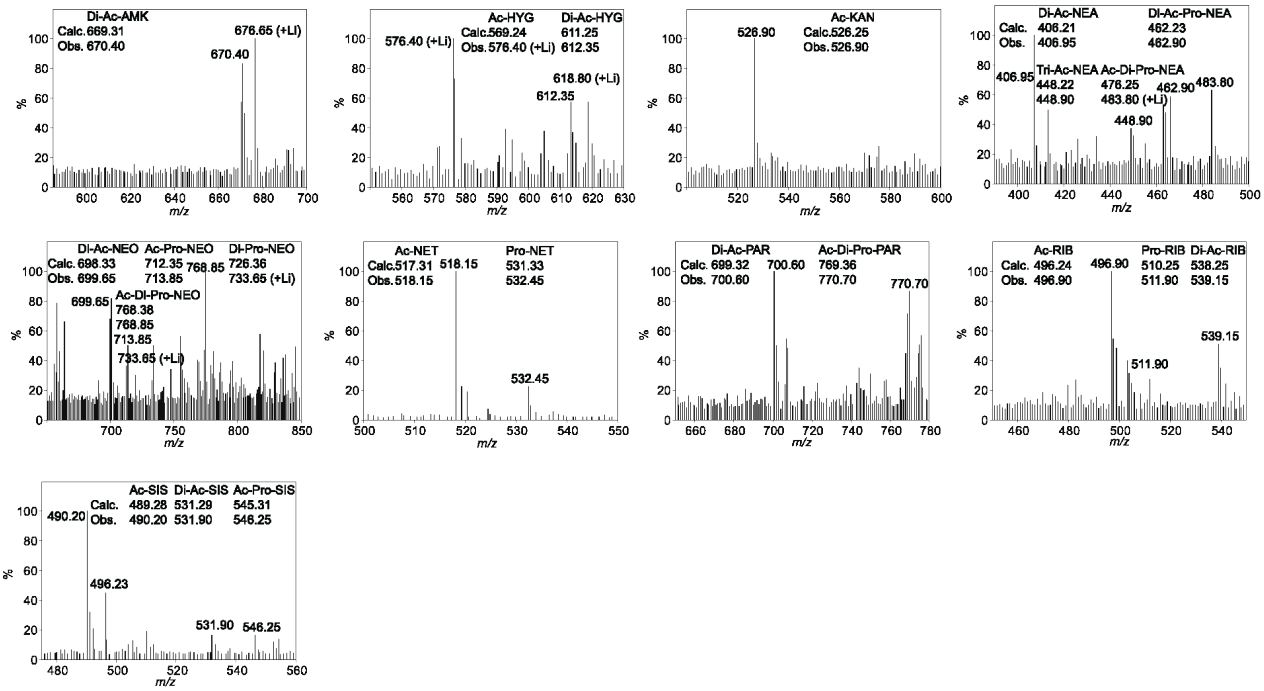
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Ban*.



Mass spectra of AGs that were first *n*-propionylated by Eis_*Ban*, then reacted with AcCoA in the presence of Eis_*Ban*.

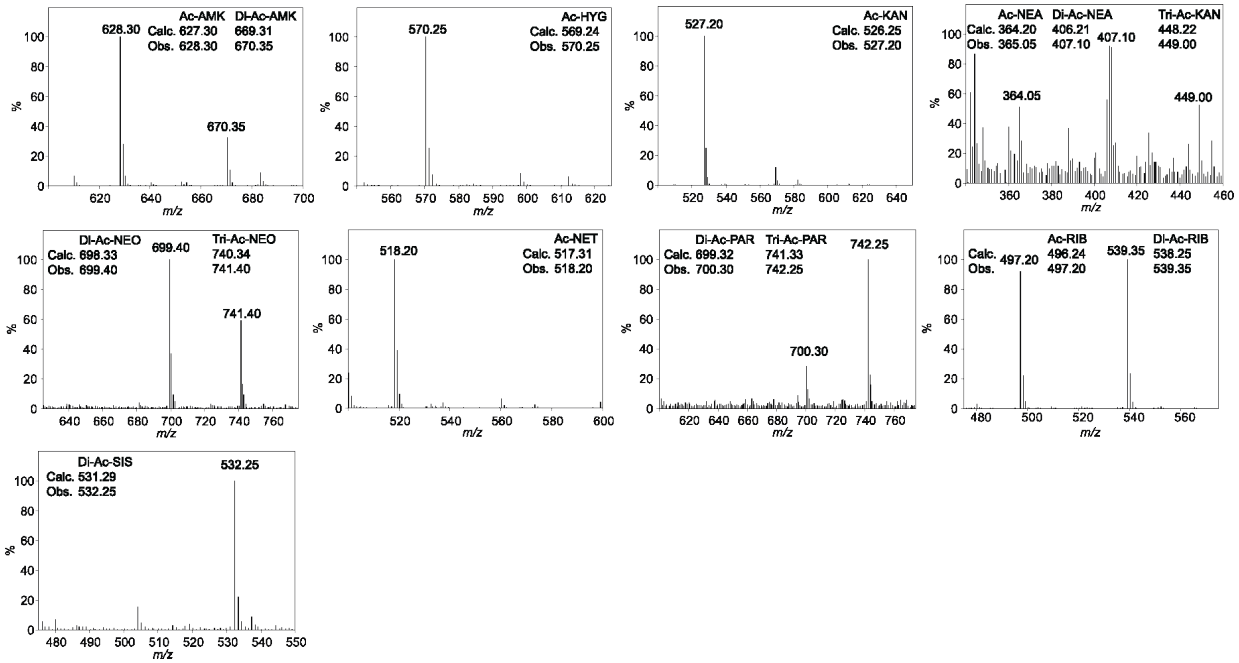


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Ban*.

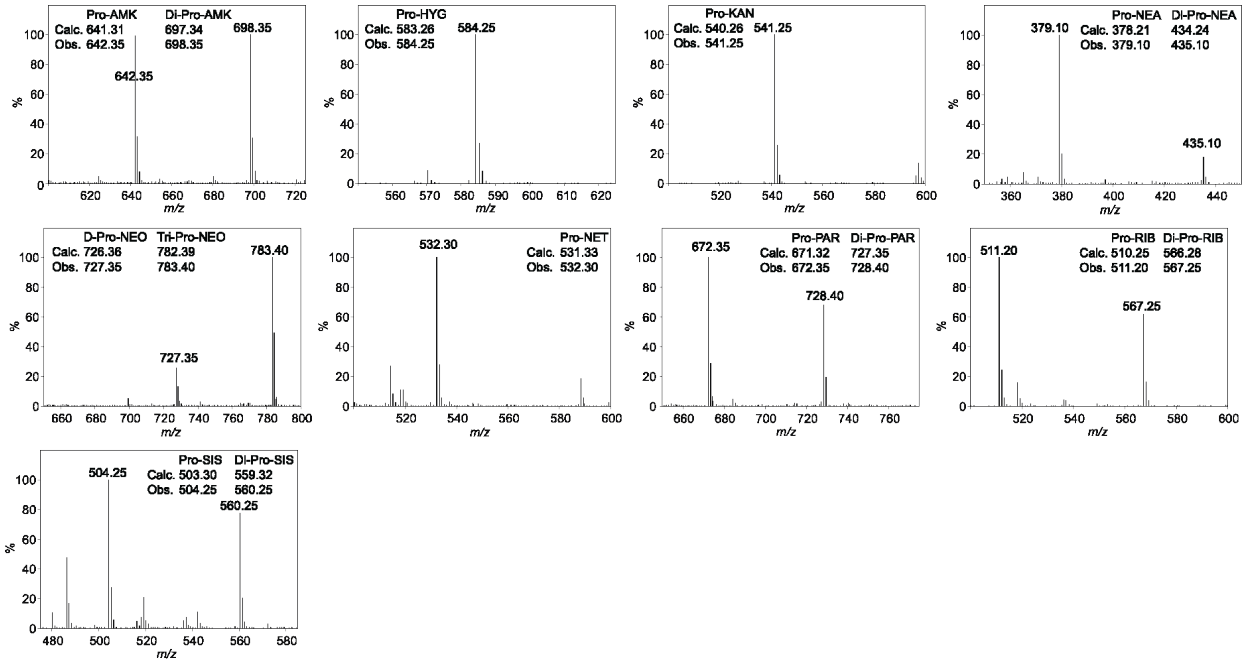
Mass analysis of AGs acylated by Eis_*Gbr*.

AG		Acetyl		*n*-Propionylation			Pro → Ac			Ac + Pro	
		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$
AMK	Mono	628.30	628.30	Mono	642.32	642.35				Ac	628.30	628.30
	Di	670.31	670.35	Di	698.35	698.35				Ac-Pro	684.33	684.35
HYG	Mono	570.25	570.25	Mono	584.27	584.25				Ac	570.25	570.25
KAN	Mono	527.26	527.20	Mono	541.27	541.25				Ac	527.26	527.25
										Pro	541.27	541.25
NEA	Mono	365.20	365.05	Mono	379.22	379.10	Pro	379.22	379.05	Ac	365.20	365.05
	Di	407.21	407.10	Di	435.25	435.10	Ac-Pro	421.23	421.05	Pro	379.22	379.05
	Tri	449.22	449.00				Di-Pro	435.25	435.15			
							Di-Ac-Pro	463.24	463.05			
							Ac-Di-Pro	477.26	477.25			
NEO	Di	699.34	699.40	Di	727.37	727.35				Di-Ac	699.34	699.55
	Tri	741.35	741.40	Tri	783.40	783.40				Di-Pro	727.37	727.30
										Tri-Ac	741.35	741.30
										Di-Ac-Pro	755.37	755.35
										Ac-Di-Pro	769.38	769.40
										Tri-Pro	783.40	783.35
NET	Mono	518.32	518.20	Mono	532.33	532.30				Ac	518.32	518.20
										Pro	532.33	537.10 (+Li)
PAR	Di	700.32	700.30	Mono	672.33	672.35	Ac-Di-Pro	770.37	770.45	Di-Ac	700.32	700.40
	Tri	742.34	742.25	Di	728.36	728.40				Ac-Pro	714.34	714.40
RIB	Mono	497.25	497.20	Mono	511.26	511.20				Ac	497.25	497.15
	Di	539.26	539.35	Di	567.29	567.25				Pro	511.26	518.15 (+Li)
										Di-Ac	539.26	539.25
SIS	Di	532.30	532.25	Mono	504.30	504.25				Ac	490.29	490.25
				Di	560.33	560.25				Di-Ac	532.30	532.25

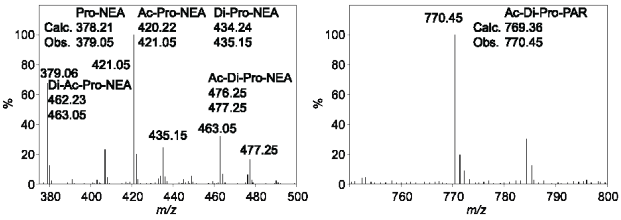
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



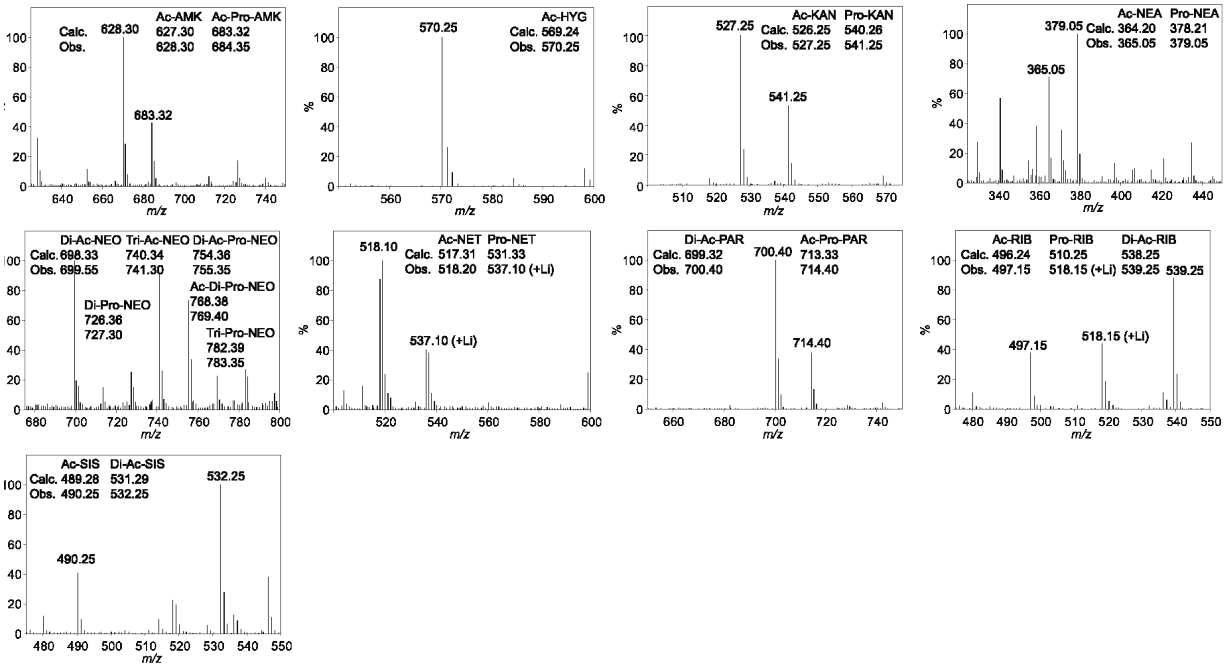
Mass spectra of AG acetylation catalyzed by Eis_*Gbr*.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Gbr*.



Mass spectra of AGs that were first *n*-propionylated by Eis *Gbr*, then reacted with AcCoA in the presence of Eis_*Gbr*.

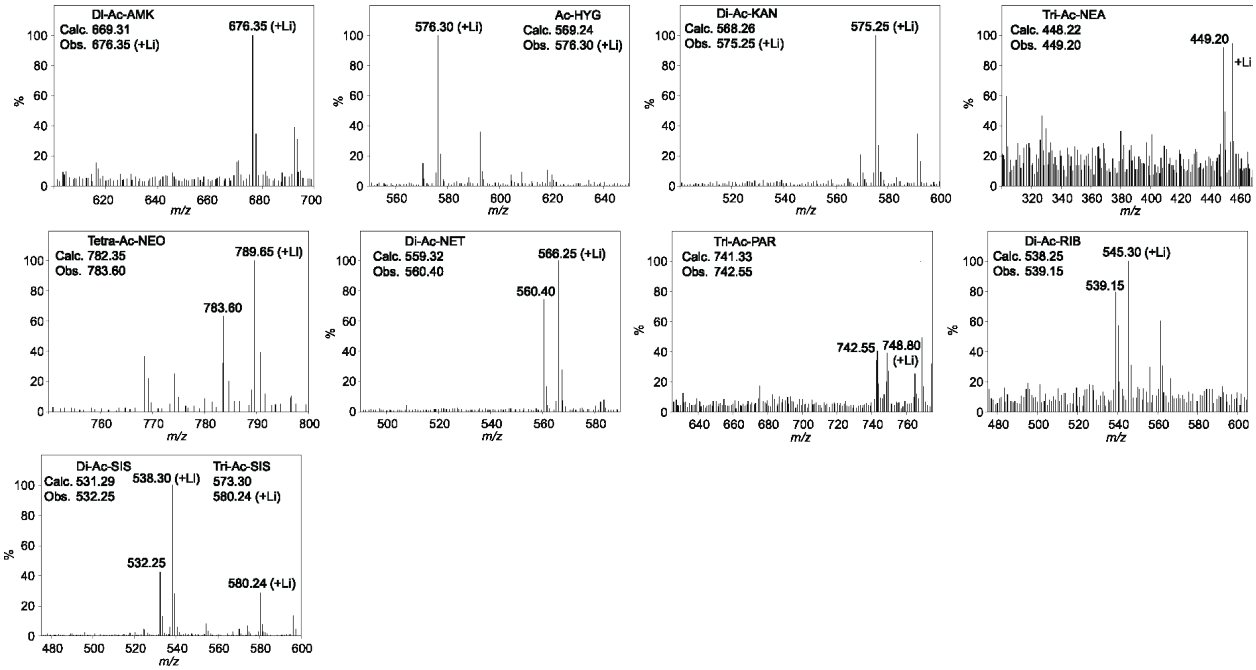


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Gbr*.

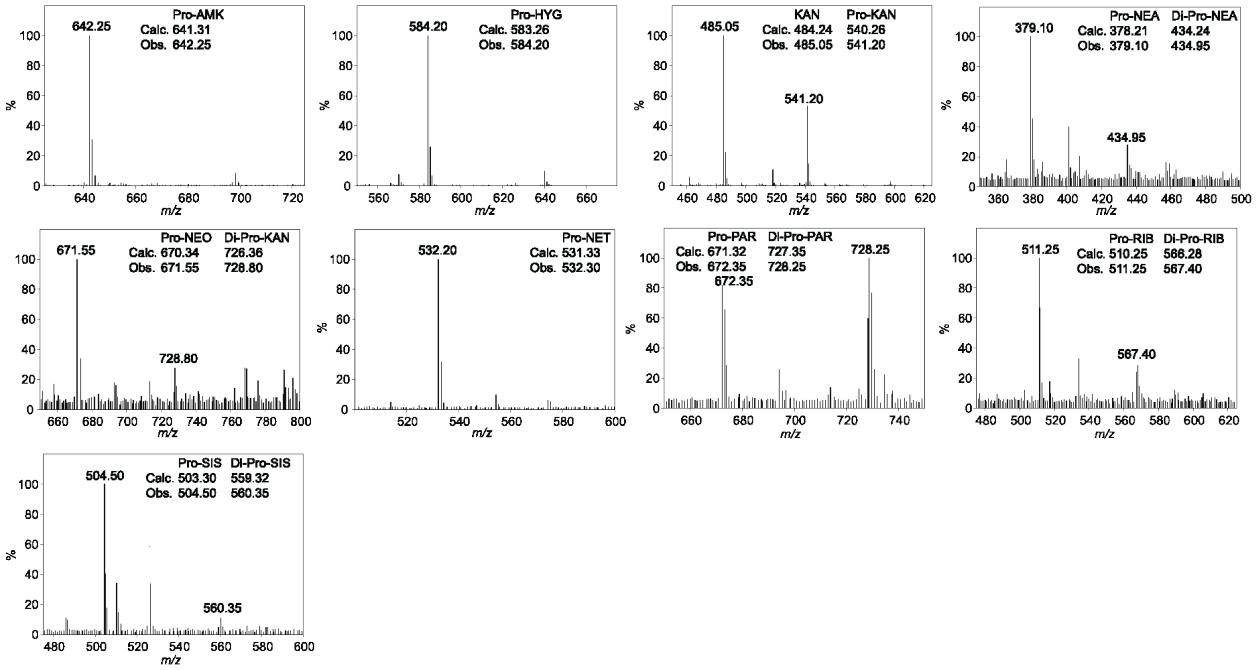
Mass analysis of AGs acylated by Eis *Krh*.

AG		Acetyl		*n*-Propionylation			Pro → Ac			Ac + Pro	
		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$
AMK	Di	670.31	676.35 (+Li)	Mono	642.32	642.25	Ac-Pro	684.33	684.90	Ac	628.30	628.30
										Di-Ac	670.31	670.25
										Ac-Pro	684.33	684.35
HYG	Mono	570.25	576.30 (+Li)	Mono	584.27	584.20				Ac	570.25	570.20
KAN	Di	569.27	575.25 (+Li)	Mono	541.27	541.20	Di-Ac	569.27	569.25	Di-Ac	569.27	568.90
							Ac-Pro	583.28	583.20			
NEA	Tri	449.22	449.20	Mono	379.22	379.10	Ac-Pro	421.23	421.05	Di-Ac	407.21	407.20
				Di	435.25	434.95						
NEO	Tetra	783.36	783.60	Mono	671.35	671.55	Ac-Di-Pro	769.38	769.35	Di-Ac	699.34	699.70
				Di	727.37	728.80				Ac-Pro	713.36	713.45
NET	Di	560.33	560.40	Mono	532.33	532.20	Ac	518.32	518.25	Ac	518.32	518.30
							Pro	532.33	532.30			
							Di-Ac	560.33	560.30			
							Ac-Pro	574.35	574.35			
PAR	Tri	742.34	742.55	Mono	672.33	672.35	Ac-Di-Pro	770.32	770.85	Di-Ac	700.32	700.40
				Di	728.36	728.25						
RIB	Di	539.26	539.15	Mono	511.26	511.25				Di-Ac	539.26	539.30
				Di	567.29	567.40						
SIS	Di	532.30	532.35	Mono	504.30	504.50	Pro	504.30	504.15	Ac	490.29	490.60
	Tri	574.31	580.24 (+Li)	Di	560.33	560.35	Ac-Pro	546.31	546.20	Di-Ac	532.30	532.35
							Di-Pro	560.33	560.15	Ac-Pro	546.31	546.90

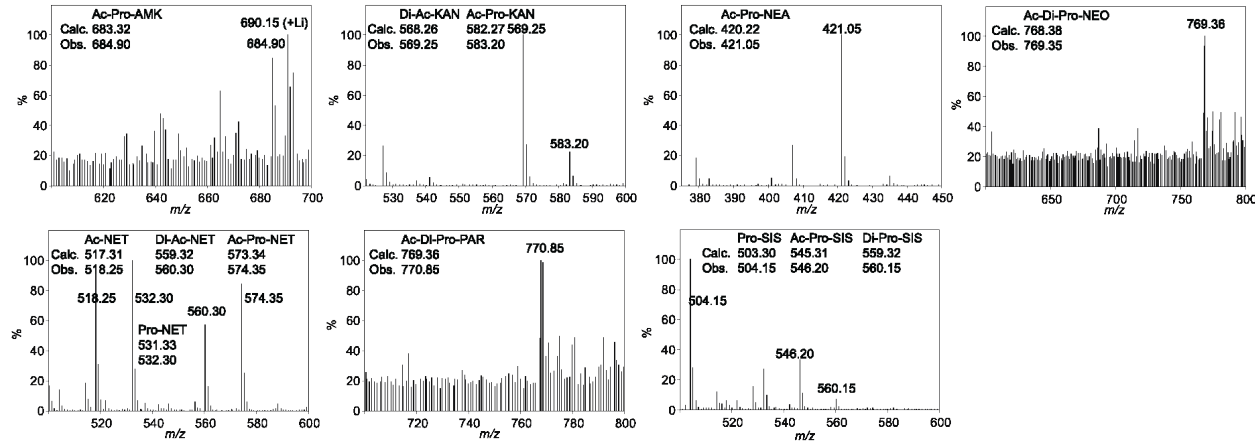
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



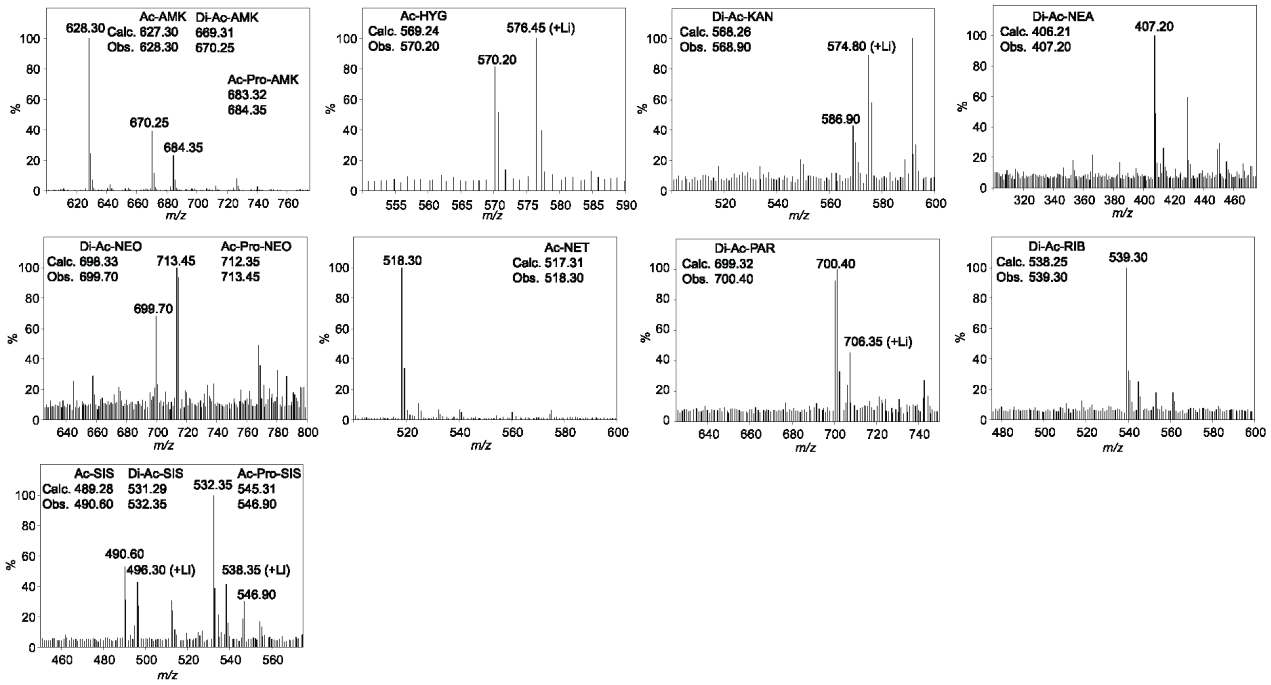
Mass spectra of AG acetylation catalyzed by Eis_*Krh*.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Krh*.



Mass spectra of AGs that were first *n*-propionylated by Eis_*Krh* then reacted with AcCoA in the presence of Eis_*Krh*.

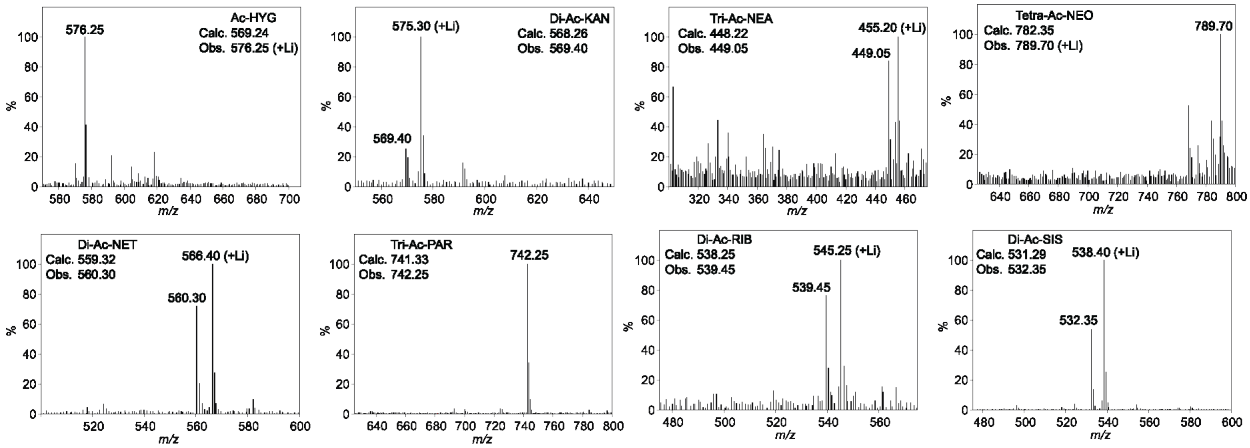


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Krh*.

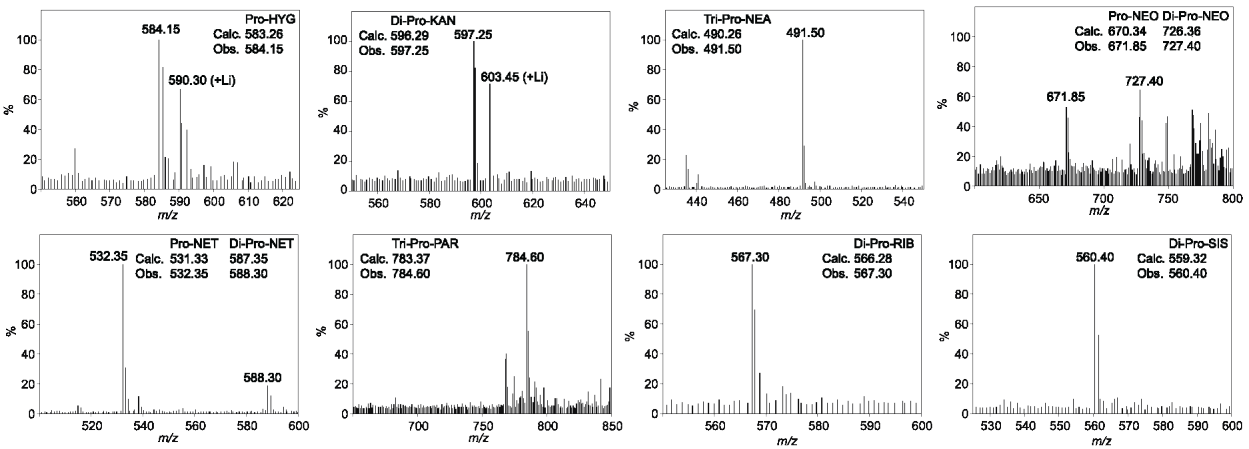
Mass analysis of AGs acylated by Eis_*Mab*.

AG	Acetyl		*n*-Propionylation			Pro → Ac			Ac + Pro	
	Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$
HYG Mono	570.25	576.25 (+Li)	Mono	584.27	584.15				Ac	570.25	570.15
KAN Di	569.27	569.40	Di	597.30	597.25				Di-Ac	569.27	569.25
NEA Tri	449.22	449.05	Tri	491.27	491.50				Tri-Ac	449.22	449.20
									Di-Ac-Pro	463.24	463.15
NEO Tetra	783.36	789.70 (+Li)	Mono	671.35	671.85	Ac-Di-Pro	769.38	769.85	Tri-Ac	741.35	741.35
			Di	727.37	727.40				Di-Ac-Pro	755.37	755.30
									Tetra-Ac	783.36	783.35
									Tri-Ac-Pro	797.38	797.25
									Di-Ac-Di-Pro	811.39	811.25
NET Di	560.33	560.30	Mono	532.33	532.35				Ac	518.32	518.65
			Di	588.36	588.30				Di-Ac	560.33	560.45
									Ac-Pro	574.35	574.00
									Di-Pro	588.36	588.50
PAR Tri	742.34	742.25	Tri	784.38	784.60				Tri-Ac	742.34	742.65
RIB Di	539.26	539.45	Di	567.29	567.30				Di-Ac	539.26	539.10
SIS Di	532.30	532.35	Di	560.33	560.40				Di-Ac	532.30	532.25

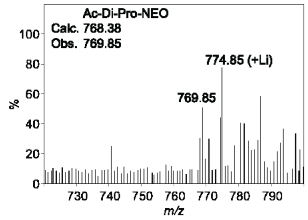
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



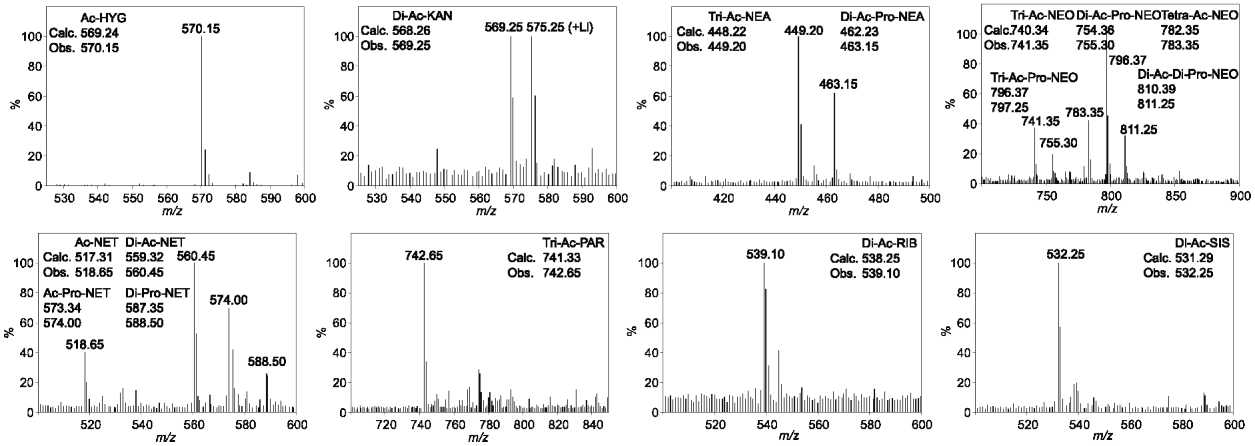
Mass spectra of AG acetylation catalyzed by Eis_*Mab*.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Mab*.



Mass spectrum of NEO that was first *n*-propionylated by Eis *Mab* then reacted with AcCoA in the presence of Eis_*Mab*

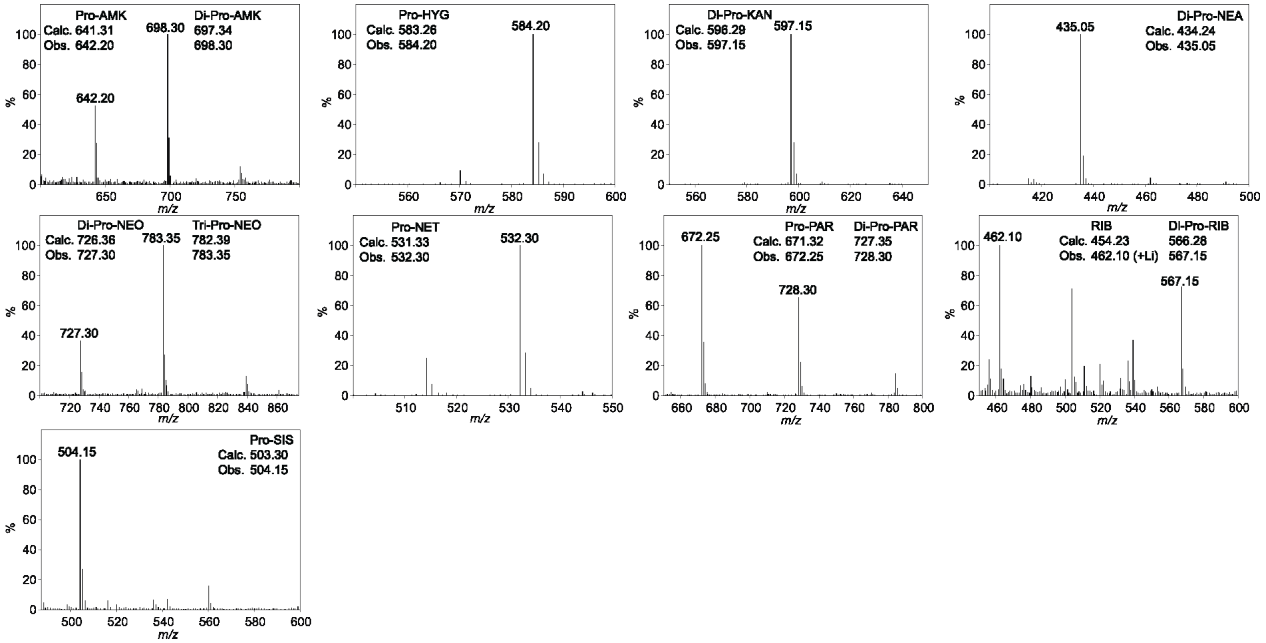


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Mab*.

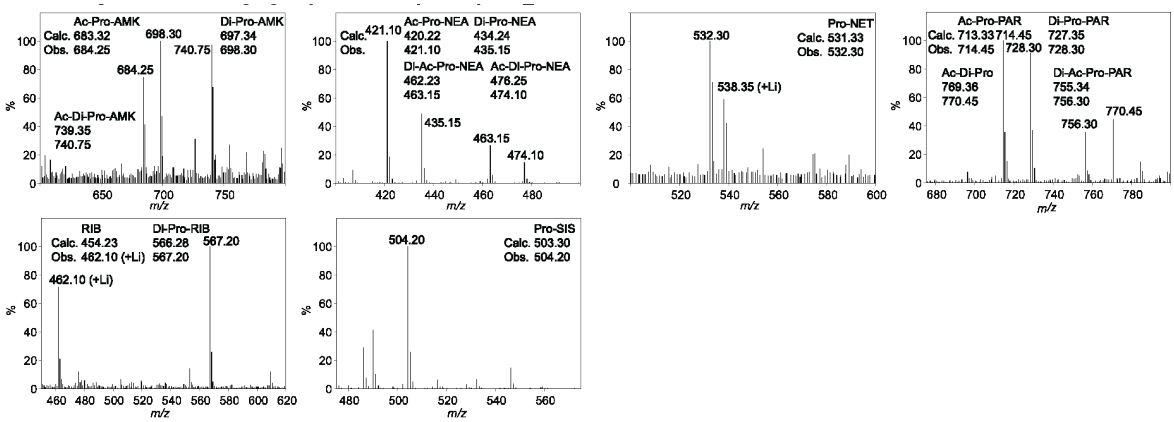
Mass analysis of AGs acylated by Eis *Msm*1.

AG		*n*-Propionylation Calc $[M+H]^+$	Obs $[M+H]^+$	Pro → Ac	Calc $[M+H]^+$	Obs $[M+H]^+$	Ac + Pro	Calc $[M+H]^+$	Obs $[M+H]^+$
AMK	Mono	642.32	642.20	Ac-Pro	684.33	684.25	Di-Ac	670.31	670.20
	Di	698.35	698.30	Di-Pro	698.35	698.30			
				Ac-Di-Pro	740.36	740.75			
HYG	Mono	584.27	584.20				Ac	570.25	570.15
KAN	Di	597.30	597.15				Di-Ac	569.27	569.10
NEA	Di	435.25	435.05	Ac-Pro	421.23	421.10	Di-Ac	407.21	407.00
				Di-Pro	435.25	435.15			
				Di-Ac-Pro	463.24	463.15			
				Ac-Di-Pro	477.26	474.10			
NEO	Di	727.37	727.30				Tri-Ac	741.35	741.25
	Tri	783.40	783.35						
NET	Mono	532.33	532.30	Pro	532.33	532.30	Ac	518.32	518.15
PAR	Mono	672.33	672.25	Ac-Pro	714.34	714.45	Di-Ac	700.32	700.20
	Di	728.36	728.30	Di-Pro	728.36	728.30			
				Di-Ac-Pro	756.35	756.30			
				Ac-Di-Pro	770.37	770.45			
RIB	Di	567.29	567.15	Di-Pro	567.29	567.20	Ac	497.25	503.95 (+Li)
							Di-Ac	539.26	539.15
SIS	Mono	504.30	504.15	Pro	504.30	504.20	Pro	504.30	503.95

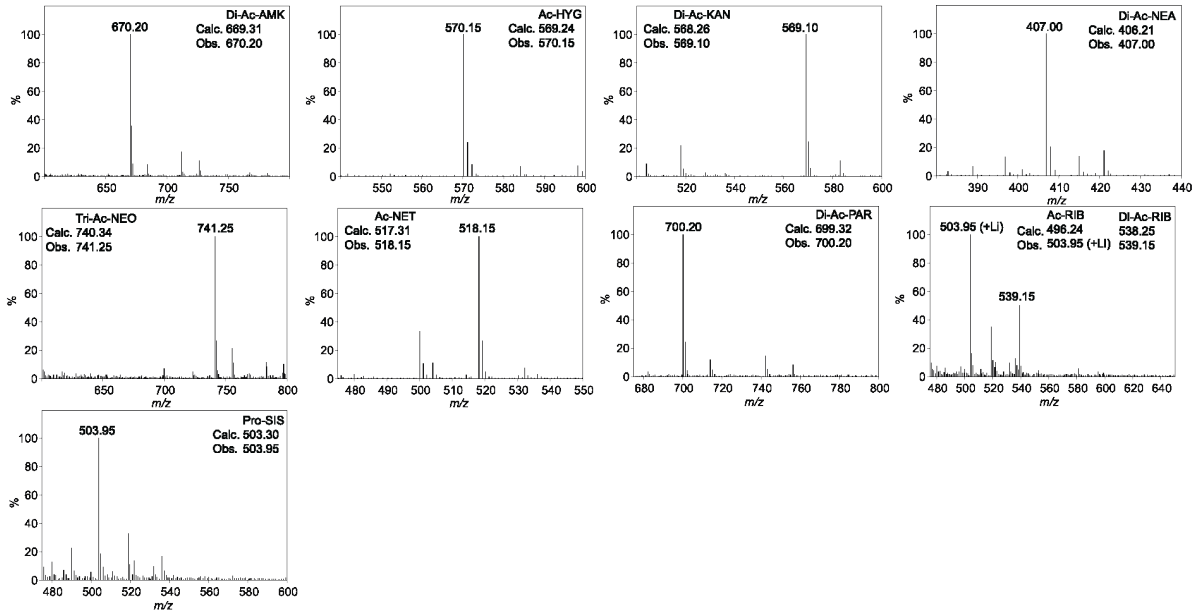
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Msm*1.



Mass spectra of AGs that were first *n*-propionylated by Eis *Msm*1, then reacted with AcCoA in the presence of Eis_*Msm*1.

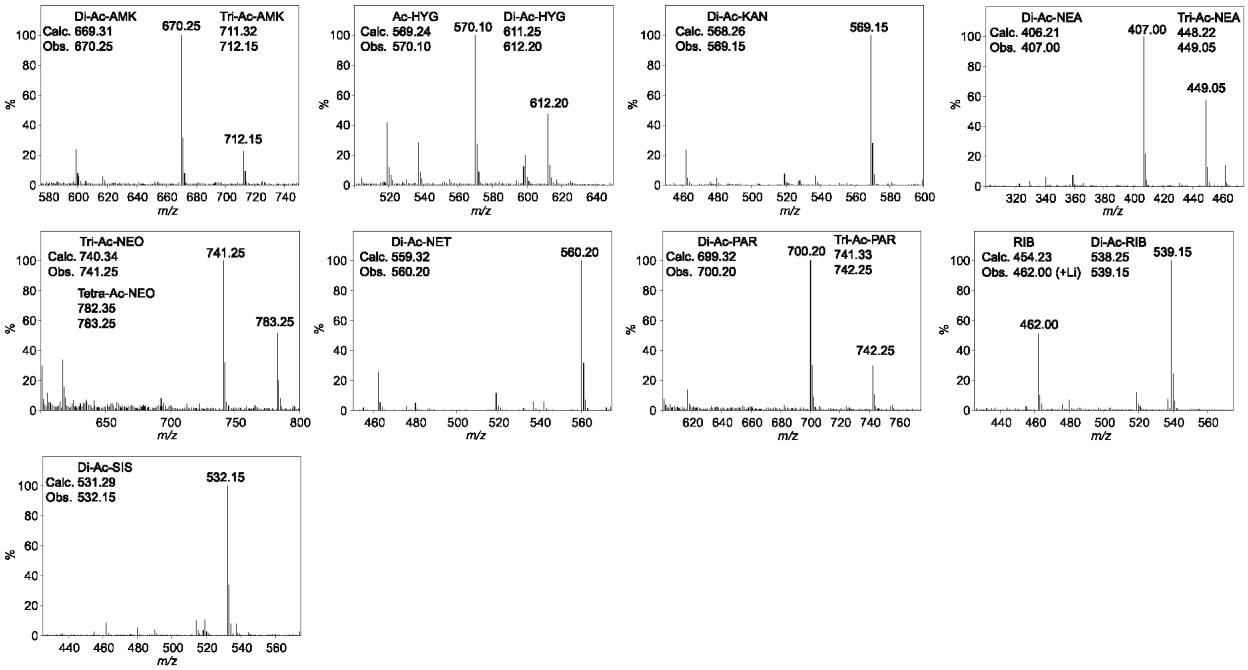


Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Msm*1.

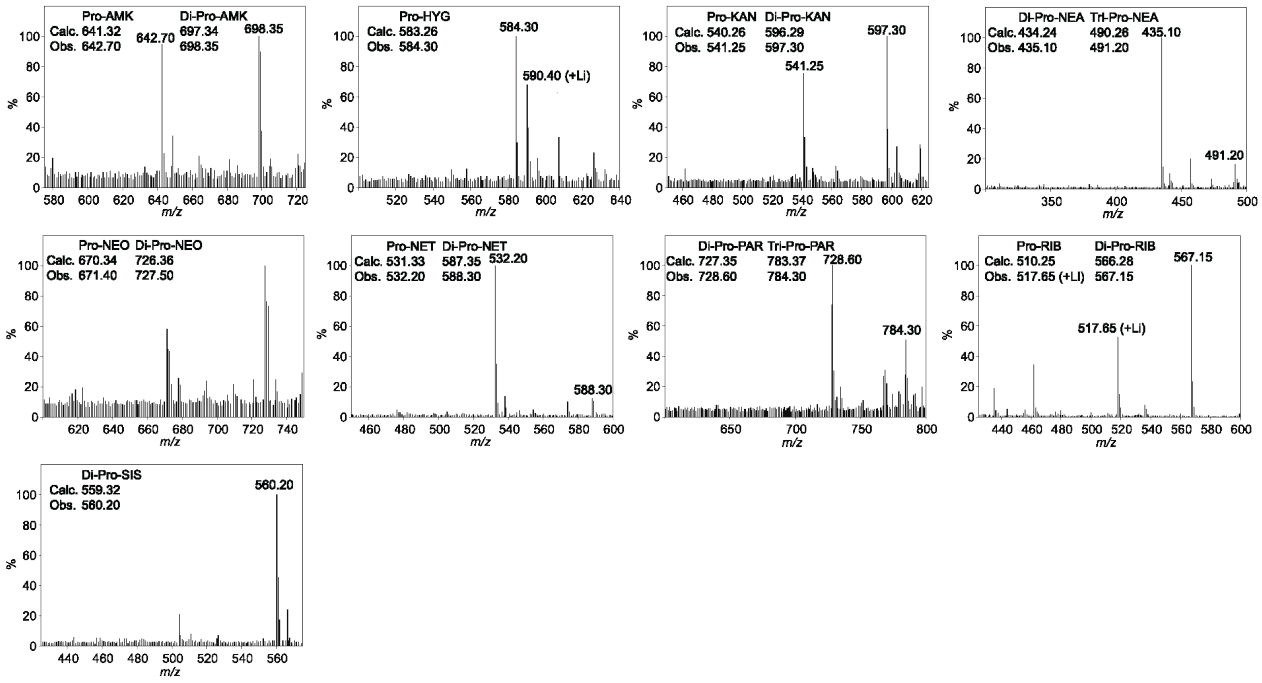
Mass analysis of AGs acylated by Eis_*Tpa*.

AG	Acetyl			*n*-Propionylation			Pro → Ac			Ac + Pro	
	Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$		Calc $[M+H]^+$	Obs $[M+H]^+$
AMK Di	670.31	670.25	Mono	642.32	642.70	Di-Pro	698.35	698.45	Di-Ac	670.31	670.85
Tri	712.33	712.15	Di	698.35	698.35	Ac-Di-Pro	740.36	740.65			
HYG Mono	570.25	570.10	Mono	584.27	584.30	Pro	584.27	584.30	Ac	570.25	570.00
Di	612.26	612.20				Ac-Pro	626.28	626.55	Di-Ac	612.26	612.85
KAN Di	569.27	569.15	Mono	541.27	541.25				Di-Ac	569.27	569.50
			Di	597.30	597.30						
NEA Di	407.21	407.00	Di	435.25	435.10				Di-Ac	407.21	407.00
Tri	449.22	449.05	Tri	491.27	491.20				Tri-Ac	449.22	449.10
NEO Tri	741.35	741.25	Mono	671.35	671.40	Di-Ac-Pro	755.37	755.35	Ac	657.33	657.65
Tetra	783.36	783.25	Di	727.37	727.50	Ac-Di-Pro	769.38	769.45	Pro	671.35	671.85
						Tri-Ac-Pro	797.38	797.40	Di-Ac	699.34	699.75
						Di-Ac-Di-Pro	811.39	811.45	Ac-Pro	713.36	713.85
									Tri-Ac	741.35	741.85
NET Di	560.33	560.20	Mono	532.33	532.20				Ac	518.32	518.00
			Di	588.36	588.30				Di-Ac	560.33	560.30
									Ac-Pro	574.35	574.40
PAR Di	700.32	700.20	Di	728.36	728.60				Di-Ac	700.32	700.50
Tri	742.34	742.25	Tri	784.38	784.30				Tri-Ac	742.34	742.85
RIB Di	539.26	539.15	Mono	511.26	517.65 (+Li)				Di-Ac	539.26	539.90
			Di	567.29	567.15						
SIS Di	532.30	532.15	Di	560.33	560.20				Di-Ac	532.30	532.25

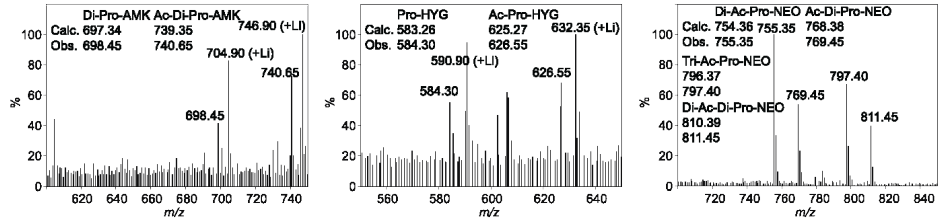
Mono indicates mono-acylation.
Di indicates di-acylation.
Tri indicates tri-acylation.



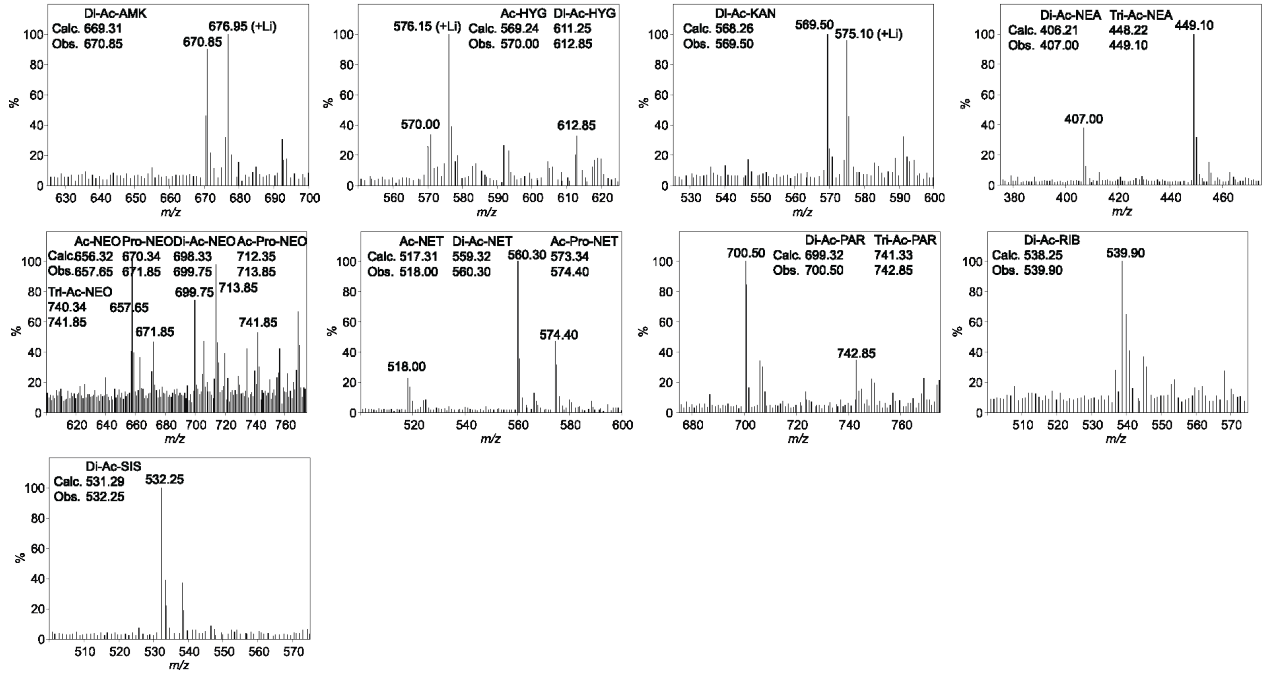
Mass spectra of AG acetylation catalyzed by Eis_*Tpa*.



Mass spectra of AG *n*-propionylation catalyzed by Eis_*Tpa*.



Mass spectra of AGs that were first *n*-propionylated by Eis_*Tpa*, then reacted with AcCoA in the presence of Eis_*Tpa*.



Mass spectra of AGs that were treated with a 1:1 mixture of AcCoA and ProCoA in the presence of Eis_*Tpa*.

References Appendix 1c:

1. Pricer, R. E., Houghton, J. L., Green, K. D., Mayhoub, A. S., and Garneau-Tsodikova, S. (2012) Biochemical and structural analysis of aminoglycoside acetyltransferase Eis from Anabaena variabilis, *Mol Biosyst 8*, 3305-3313.
2. Chen, W., Biswas, T., Porter, V. R., Tsodikov, O. V., and Garneau-Tsodikova, S. (2011) Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB, *Proceedings of the National Academy of Sciences of the United States of America 108*, 9804-9808.
3. Kim, K. H., An, D. R., Song, J., Yoon, J. Y., Kim, H. S., Yoon, H. J., Im, H. N., Kim, J., Kim do, J., Lee, S. J., Lee, H. M., Kim, H. J., Jo, E. K., Lee, J. Y., and Suh, S. W. (2012) Mycobacterium tuberculosis Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7, *Proceedings of the National Academy of Sciences of the United States of America 109*, 7729-7734.
4. Chen, W., Green, K. D., and Garneau-Tsodikova, S. (2012) Cosubstrate tolerance of the aminoglycoside resistance enzyme Eis from Mycobacterium tuberculosis., *Antimicrob. Agents Chemother. 56*, 5831.
5. Green, K. D., Biswas, T., Chang, C., Wu, R., Chen, C., Janes, B. K., Chalupska, D., Gornicki, P., Hanna, P. C., Tsodikov, O. V., and Joachimiak, A. (2015) Biochemical and structural analysis of an Eis family aminoglycoside acetyltransferase from Bacillus anthraci, *Biochemistry*.
6. Green, K. D., Chen, W., and Garneau-Tsodikova, S. (2012) Identification and characterization of inhibitors of the aminoglycoside resistance acetyltransferase Eis from Mycobacterium tuberculosis, *ChemMedChem 7*, 73-77.
7. Jennings, B. C., Labby, K. J., Green, K. D., and Garneau-Tsodikova, S. (2013) Redesign of substrate specificity and identification of the aminoglycoside binding residues of Eis from Mycobacterium tuberculosis, *Biochemistry 52*, 5125.
8. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America., *Clin. Infect. Dis. 48*, 1.
9. Chen, W., Green, K. D., Tsodikov, O. V., and Garneau-Tsodikova, S. (2012) Aminoglycoside Multiacetylating Activity of the Enhanced Intracellular Survival Protein from Mycobacterium smegmatis and Its Inhibition, *Biochemistry*.
10. Houghton, J. L., Biswas, T., Chen, W., Tsodikov, O. V., and Garneau-Tsodikova, S. (2013) Chemical and structural insights into the regioversatility of the aminoglycoside acetyltransferase Eis., *ChemBioChem 14*, 2127.
11. Zaunbrecher, M. A., Sikes, R. D., Jr., Metchock, B., Shinnick, T. M., and Posey, J. E. (2009) Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis, *Proceedings of the National Academy of Sciences of the United States of America 106*, 20004-20009.
12. Georghiou, S. B., Magana, M., Garfein, R. S., Catanzaro, D. G., Catanzaro, A., and Rodwell, T. C. (2012) Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review., *PloS one 7*, e33275.

13. Rodwell, T. C., Valafar, F., Douglas, J., Qian, L., Garfein, R. S., Chawla, A., Torres, J., Zadorozhny, V., Kim, M. S., Hoshide, M., Catanzaro, D., Jackson, L., Lin, G., Desmond, E., Rodrigues, C., Eisenach, K., Victor, T. C., Ismail, N., Crudu, V., Gler, M. T., and Catanzaro, A. (2014) Predicting extensively drug-resistant Mycobacterium tuberculosis phenotypes with genetic mutations., *J. Clin, Microbiol 52*, 781.
14. Gikalo, M. B., Nosova, E. Y., Krylova, L. Y., and Moroz, A. M. (2012) The role of eis mutations in the development of kanamycin resistance in Mycobacterium tuberculosis isolates from the Moscow region., *J. Antimicrob. Chemother. 67*, 2107.
15. Green, K. D., Chen, W., Houghton, J. L., Fridman, M., and Garneau-Tsodikova, S. (2010) Exploring the substrate promiscuity of drug-modifying enzymes for the chemoenzymatic generation of N-acylated aminoglycosides, *Chembiochem : a European journal of chemical biology 11*, 119-126.
16. Green, K. D., Chen, W., and Garneau-Tsodikova, S. (2011) Effects of altering aminoglycoside structures on bacterial resistance enzyme activities., *Antimicrob. Agents Chemother. 55*, 3207.
17. Dahl, J. L., Wei, J., Moulder, J. W., Laal, S., and Friedman, R. L. (2001) Subcellular localization of the intracellular survival-enhancing Eis protein of Mycobacterium tuberculosis, *Infect. Immun. 69*, 4295.
18. Samuel, L. P., Song, C. H., Wei, J., Roberts, E. A., Dahl, J. L., Barry, C. E., Jo, E. K., and Friedman, R. L. (2007) Expression, production and release of the Eis protein by Mycobacterium tuberculosis during infection of macrophages and its effect on cytokine secretion, *Microbiology 153*, 529.
19. Lella, R. K., and Sharma, C. (2007) Eis (Enhanced Intracellular Survival) Protein of Mycobacterium tuberculosis Disturbs the Cross Regulation of T-cells, *Journal of Biological Chemistry 282*, 18671-18675.
20. Ganaie, A. A., Lella, R. K., Solanki, R., and Sharma, C. (2011) Thermostable hexameric form of Eis (Rv2416c) protein of M. tuberculosis plays an important role for enhanced intracellular survival within macrophages, *PloS one 6*, e27590.
21. Yoon, H. J., Kim, K. H., Yang, J. K., Suh, S. W., Kim, H., and Jang, S. (2013) A docking study of enhanced intracellular survival protein from Mycobacterium tuberculosis with human DUSP16/MKP-7, *J. Synchrotron Radiat. 20*, 929.
22. Wei, J., Dahl, J. L., Moulder, J. W., Roberts, E. A., O'Gaora, P., Young, D. B., and Friedman, R. L. (2000) Identification of a Mycobacterium tuberculosis Gene That Enhances Mycobacterial Survival in Macrophages, *Journal of Bacteriology 182*, 377-384.
23. Esteban, J., and Ortiz-Perez, A. (2009) Current treatment of atypical mycobacteriosis, *Expert Opin. Pharmacother. 10*, 2787.
24. Nessar, R., Cambau, E., Reyrat, J. M., Murray, A., and Gicquel, B. (2012) Mycobacterium abscessus: a new antibiotic nightmare, *J Antimicrob Chemother 67*, 810-818.
25. Nessar, R., Reyrat, J. M., Murray, A., and Gicquel, B. (2011) Genetic analysis of new 16S rRNA mutations conferring aminoglycoside resistance in Mycobacterium abscessus, *J. Antimicrob. Chemother. 66*, 1719.
26. Cremades, R., Santos, A., Rodriguez, J. C., Garcia-Pachon, E., Ruiz, M., and Royo, G. (2009) Mycobacterium abscessus from respiratory isolates: activities of drug combinations, *J. Infect. Chemother. 15*, 46.

27. Ho, I. I. Y., Chan, C. Y., and Cheng, A. F. B. (2000) Aminoglycoside resistance in Mycobacterium kansasii, Mycobacterium avium-M. intracellulare, and Mycobacterium fortuitum: are aminoglycoside-modifying enzymes responsible?, *Antimicrob. Agents Chemother. 44*, 39.
28. Steinhaus, E. A. (1941) A study of the bacteria associated with thirty species of insects, *J. Bacteriol. 42*, 757.
29. Lecomte, J., St-Arnaud, M., and Hijri, M. (2011) Isolation and identification of soil bacteria growing at the expense of arbuscular mycorrhizal fungi, *FEMS Microbiol. Lett. 317*, 43.
30. Becker, K., Rutsch, F., Uekotter, A., Kipp, F., Konig, J., Marquardt, T., Peters, G., and von Eiff, C. (2008) Kocuria rhizophila adds to the emerging spectrum of micrococcal species involved in human infections, *J. Clin. Microbiol. 46*, 3537.
31. Moissenet, D., Becker, K., Merens, A., Ferroni, A., Dubern, B., and Vu-Thien, H. (2012) Persistent bloodstream infection with Kocuria rhizophila related to a damaged central catheter, *J. Clin. Microbiol. 50*, 1495.
32. Biswas, T., Houghton, J. L., Garneau-Tsodikova, S., and Tsodikov, O. V. (2012) The structural basis for substrate versatility of chloramphenicol acetyltransferase CATI, *Protein Sci. 21*, 520.
33. Houghton, J. L., Green, K. D., Pricer, R. E., Mayhoub, A. S., and Garneau-Tsodikova, S. (2013) Unexpected N-acetylation of capreomycin by mycobacterial Eis enzymes, *Journal of Antimicrobial Chemotherapy 68*, 800-805.
34. Vetting, M. W., Park, C. H., Hegde, S. S., Jacoby, G. A., Hooper, D. C., and Blanchard, J. S. (2008) Mechanistic and Structural Analysis of Aminoglycoside N-Acetyltransferase AAC(6′)-Ib and Its Bifunctional, Fluoroquinolone-Active AAC(6′)-Ib-cr Variant, *Biochemistry 47*, 9825-9835.
35. Kim, K. H., An, D. R., Yoon, H. J., Yang, J. K., and Suh, S. W. (2014) Structure of Mycobacterium smegmatis Eis in complex with paromomycin, *Acta Crystallogr., Sect. F: Struct. Biol. Commun. 70*, 1173.
36. Tsodikov, O. V., Green, K. D., and Garneau-Tsodikova, S. (2014) A random sequential mechanism of aminoglycoside acetylation by Mycobacterium tuberculosis Eis protein, *PloS one 9*, e92370.
37. Tiago, I., Maranha, A., Mendes, V., Alarico, S., Moynihan, P. J., Clarke, A. J., Macedo-Ribeiro, S., Pereira, P. J., and Empadinhas, N. (2012) Genome sequence of Mycobacterium hassiacum DSM 44199, a rare source of heat-stable mycobacterial protein, *J. Bacteriol. 194*, 7010.

Appendix 2: Use of Photocrosslinking Amino Acids for Quantitation of Activated Ras Isoforms in Live Cells[9]

2a: Abstract

Applying the knowledge gained from previous studies with difficult PPIs in yeast, covalent chemical capture was applied to a high-affinity, well-structured PPI. Photocrosslinking was used to improve upon current methods to isolate activated Ras within cells. Extensive optimization went into the production of protein containing Bpa within *E. coli* as well as the construct used. We have shown that using photocrosslinking, we can directly visualize activated Ras crosslinked adducts on a Western blot. In addition to this, we have worked on developing a liquid chromatography mass spectrometry (LC-MS) method to detect Ras isoforms for future work that will apply this method to quantify the amount of activated Ras isoforms.

2b: Introduction

The recent development of the methods to efficiently genetically incorporate a wider variety of unnatural amino acids into a wealth of cell types has finally begun to allow the expansion of the technique into other research applications. Uses have ranged from photocrosslinking moieties for covalent capture, fluorescently tagging proteins, or addition of a chemical handle for biorthogonal chemistry in cells. In the Mapp lab, this technique has been applied to the covalent capture of transient, lower affinity PPIs. We decided to extend this method to a high affinity, well-characterized PPI. Leveraging the ability of the Ras Binding Domain (RBD) to capture activated GTPases, we wanted to test the potential for improvement to current methods using photocrosslinkers. The specific PPI chosen for this study was the interaction between a kinase in the MAPK pathway,

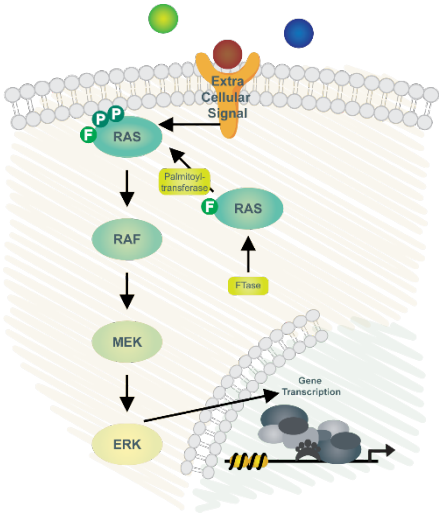


Figure A2-1 Schematic of the MAPK Pathway.

[9] The work presented here was done in partnership with Omari Baruti and mass spectrometry work done by Sarah Haynes of the Martin lab.

Raf, and Ras which only occurs once Ras is activated and GTP-bound (Figure A2-1).[2] Using the Ras-binding domain (RBD) of Raf has become a well-established probe for isolating activated Ras. The binding interaction is at the low nm level and traditionally has been utilized with a GST-Raf-RBD fusion protein.[2] By adding an unnatural amino acid to covalently capture this interaction, we can improve pulldown efficiency for mass spectrometry purposes as well as directly visualize the percentage of activated Ras via Western blot.

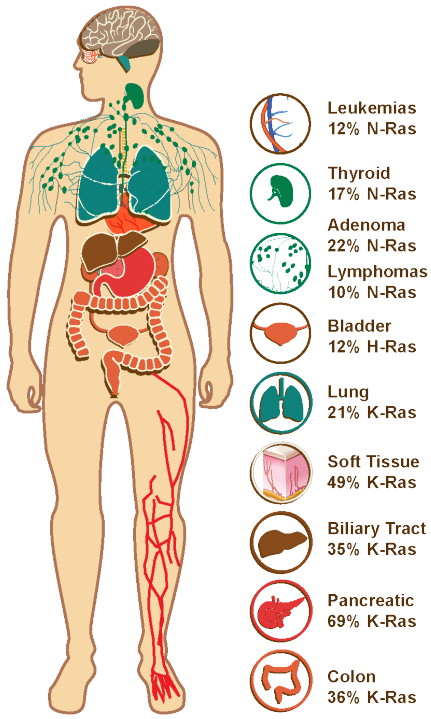


Figure A2-2 Percentages of Ras Mutation Involvement in Various Cancers.[1]

The biological importance of this PPI is its critical involvement in many cancers, as the activation of Ras leads to cell proliferation. One element of this interaction that has yet to be fully elucidated is the exact participation of the 4 major Ras isoforms, H-, N-, K4a-, and K4b. Many of these isoforms have been shown to be a major contributor in different types of cancer (Figure A2-2). K-Ras has been indicated to contain oncogenic mutations in over 90% of pancreatic caners, 45% of colorectal cancers, and 35% of lung cancers.[1] N-Ras mutations are most notable in melanomas while H-Ras mutations are present in 10% of bladder cancers.[1] Although these percentages are known, it has been difficult to study which isoforms of Ras become activated under specific conditions and in what ratios compared to other isoforms. This has proven difficult due to the similarity of the isoforms. H-, N-, and K4a-, and K4b-Ras, the main Ras isoforms, share 80% homology except for a C-terminal hypervariable region which contains almost all of their sequence variance.[3] In being able to more successfully pulldown activated Ras, we hope to be able to extend this method for the study of different cancers in the future.

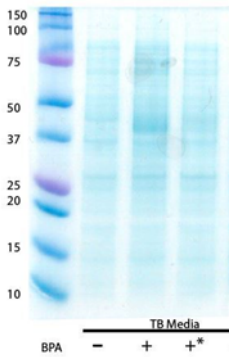


Figure A2-3 Expression Test of Raf-RBD N64Bpa.
Bpa was added immediately or *at induction point.

2c: Genetic Incorporation of Bpa in E. coli

Before beginning this project, bacterial expression of UAA containing proteins needed to be optimized and established within our laboratory. Previously, only yeast genetic incorporation of UAAs had

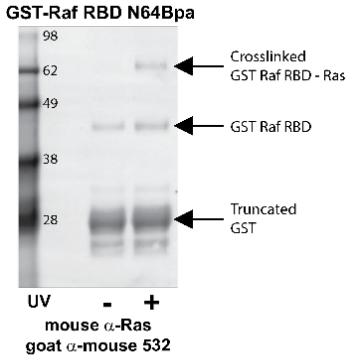


Figure A2-4 Crosslinking of GST-Raf RBD to Ras. Crosslinking of the original GST construct was tested to demonstrate that crosslinking at this position was possible for future studies.

been successfully done. To start using bacterial expression, we purchased the *Methanococcus jannaschii* tyrosyl-tRNA synthetase and tRNA pair on the pEvol backbone (Addgene #31190). However, pEvol has a chloramphenicol resistance marker such that using the regular pRARE expression aiding-vector was not possible. We tested expression without the codon supplementing pRARE but saw no expression, leading us to contact the Protein Core in the Center for Structural Biology (Life Science Institute, University of Michigan). They were able to provide us with a better expression strain suited for our use, BL21-AI cells expressing a Streptomycin resistance pRARE plasmid. This not only gave us a non-interfering codon supplementing plasmid but also tighter control over expression by placing the T7 RNA polymerase in the araBAD operon (arabinose induction is also required for protein expression).

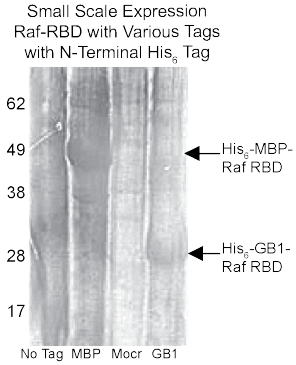


Figure A2-5 Small Scale Expression of Raf RBD Tag Testing. Coomassie stained 4-12% Bis Tris gel showing 5 mL expression test from BL21-AI *E. coli*.

Once our strain was established, we had to optimize expression conditions since varying methods for how to best incorporate UAAs were present in the literature. The recommended minimal media and auto-induction media were both tested, and yielded little expression. We settled upon the regular rich media Terrific Broth for expression and then moved on to testing the amounts of arabinose and IPTG to add for induction. This optimization resulted in the addition of 0.5 mM IPTG with 0.02% arabinose for induction. Literature protocols were also varied in the time of addition of UAA. In testing addition to media upon immediately inoculation compared with addition at the time of induction, we saw best expression with addition to the media upon inoculation (Figure A2-3).

2d: Incorporation of Bpa into the RBD of Raf

Expression of the c-Raf RBD (amino acids 1-202) has traditionally been done using a GST solublization tag. This was not applicable to genetic incorporation studies, as truncated protein is produced as a by-product in the event of unsuccessful non-sense suppression of the amber TAG stop codon (Figure A2-4). In the case of multimeric fusion tags such as GST, it can

become difficult to obtain accurate measurements of protein concentration due to the presence of truncated and full-length oligomeric species. For this reason, a variety of solublization tags (MBP, Mocr, and GB1) were tested for replacement of GST (Figure A2-5).

The final construct used, consisted of the GB1 followed by Raf RBD TAG mutants and a His6 tag. This fusion construct was seen to have one of the better expressions, solubility, and pulldown efficiency (Figure A2-6). This optimized construct allowed us to have a small solubilizing unit with our protein and a C-terminal His tag to permit efficient pulldowns using Ni-NTA agarose resin. With the optimized expression conditions, we were able to produce approximately 6 mg/L of both wild-type Raf RBD fusion protein and TAG Raf RBD with incorporated Bpa.

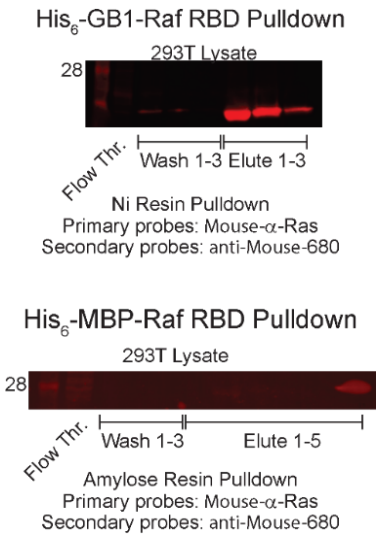


Figure A2-6 Pulldown Test Comparing a Ni-His GB1 Isolation to Amylose-MBP Isolation.
The elution off of the Ni resin for the GB1 pulldown has a much stronger band than the amylose-MBP pulldown using HEK293T lysate as a test for pulldown efficiency.

2e: Establishment of Ras Standards for Mass Spectrometry Analysis

To establish the proper standards and test our ability to differentiate between Ras isoforms on the LC-MS, the 4 major Ras isoforms were purified, digested, and analyzed by LC-MS. A variety of proteases were tested including Glu-C (C-terminal cle*Ava*ge at aspartic or glutamic acids), Asp-N (N-terminal cle*Ava*ge at aspartic acid and cysteine), Trypsin (C-terminal cle*Ava*ge at lysine and arginine), Chymotrypsin (C-terminal cle*Ava*ge at tryptophan, phenylalanine, and tyrosine), and α-Lytic (C-terminal cle*Ava*ge at alanine, threonine, serine, and valine). The best results were yielded with Asp-N and Trypsin, giving coverage of unique peptides for each isoform (Figure A2-4).

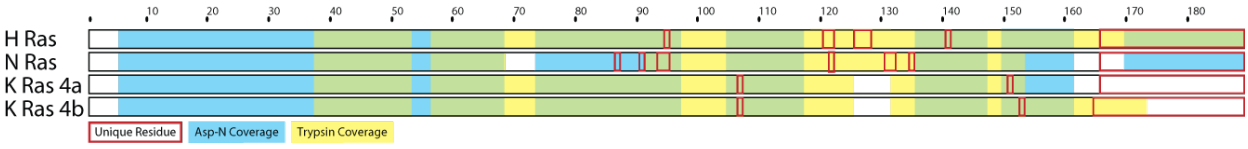


Figure A2-8 Mass Spectrometry Results for Ras Isoforms Digested with Asp-N and Trypsin.
Sequence coverage is displayed overtop a sequence scale that highlights unique residues for each isoform. Mass spectrometry methods for detecting and selecting peptides can be found in the experimental section.

2f: Crosslinking of GB1-Raf RBD V69Bpa to Ras

Once we established that we could successfully express the Raf RBD with the incorporated Bpa, we went on to make two TAG mutations, N64Bpa and V69Bpa. V69Bpa was

tested against a panel of various lysates overexpressing H-Ras, H-Ras G12V (constitutively active), and H-Ras S17N (inactive). The results were compared to a non-TAG mutant and demonstrate the ability to "quick detect" activated Ras by visualizing the crosslinked adduct without a pulldown (Figure A2-7). This technique could further be used to identify active isoforms by using the Ras isoform-specific antibodies for probing the Western blot.

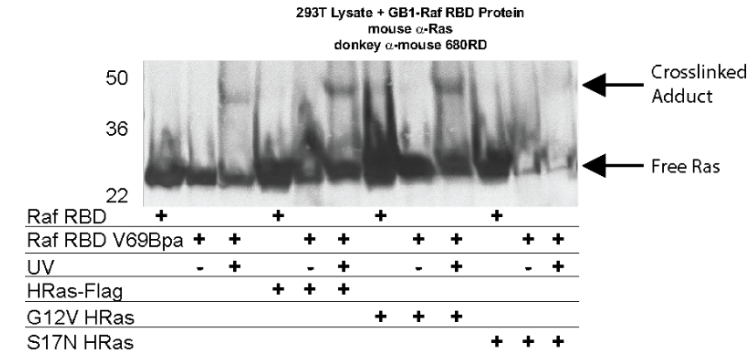


Figure A2-7 Detection of Activated Ras with Raf RBD V69Bpa. For a full comparison, each lysate was compared to wild type Raf RBD, uncrosslinked Raf RBD V69Bpa, and crosslinked Raf RBD V69Bpa. Lysates tested included HEK293T with no overexpression as a control and HEK293T overexpression of H-Ras, H-Ras G12V, and H-Ras S17N. This panel displays how much endogenous Ras is present and able to be crosslinked in lysate without any overexpression as well as showing constitutively active and inactive controls. We were able to see crosslinking in all lanes except for the inactive H-Ras S17N. Due to the similar molecular weights of the Raf RBD and Ras, protein standards of the wild type and V69Bpa were run on the last two lanes to ensure we were detecting proteins.

2g: Conclusion

Using a higher affinity PPI that also has a crystal structure, presented a unique opportunity to test the ability of photocrosslinking to improve upon already well-established methods. Probing with the RBD of Raf, we aimed to be able to isolate and crosslink Ras which only binds this domain while activated. Through the work on this project we were able to expand the genetic incorporation of photocrosslinkers to bacteria within the lab as well as develop a construct for optimal Bpa incorporation, protein expression, and purification in bacterial cells. This allowed us to express and purify the Raf RBD with Bpa incorporated at several sites along the Ras binding interface. Preliminary crosslinking studies demonstrate the ability not only to crosslink activated Ras but also to detect activated Ras via Western blot without any additional pulldown or enrichment.

To extend this method further, future experiments will focus on a LC-MS analysis of the crosslinked Raf RBD samples. Work done with purified Ras isoform standards has shown the ability to detect and search for unique peptides of each Ras isoform, allowing detection and eventually quantitation of individual isoforms by LC-MS. Although the focus here has been the Raf RBD, other GTPases including Rak, Rho, and Ral have RBDs that could be used to isolate the activated species. Understanding the amount of activated GTPase as well as individual

isoform activation will allow for a clearer picture of cellular signaling along these pathways to aid in the eventual goal of small molecule modulation.

2h: Experimental

Protein expression was carried out in BL21-AI cells expressing an additional pRareCDF vector (Protein Core in the Center for Structural Biology, Life Sciences Institute, University of Michigan). Bpa was purchased from Chem-Impex International (Wood Dale, IL). All plasmids used in this study were constructed using standard molecular biology techniques. All Ras *E. coli* expression plasmids were a kind gift from the Fierke lab (University of Michigan). Sanger sequencing verifying plasmid sequences was performed by the University of Michigan Core Facility (Ann Arbor, MI).

Construction of plasmids

pGB1+Raf-RBD-His6

A high copy plasmid expressing Raf-RBD (residues 2-149 of Raf-1) + His6 was created via ligation of the fusion gene into a pGB1(Protein Core in the Center for Structural Biology (Life Science Institute, University of Michigan) via LIC. Primers 5'-TACTTCCAATCCAATG CAGAGCACATTCAGGGTGC-3' and 5'-TTATCCACTTCCAATGTTAGTCGCGATGAATT CCC-3' were used to amplify Raf-RBD from a codon optimized gene fragment purchased from IDT. Once the Raf-RBD was ligated into pGB1, site directed mutagenesis was used to remove the His tag from the N-terminus with the primers 5'-CTTTAAGAAGGAGATATACATATGTC TTCTGGTGTAGATCTGCAGTAC-3' and 5'-GTACTGCAGATCTACACCAGAAGACATAT GTATATCTCCTTCTTAAAG-3'. A His tag was then added back onto the C-terminus with the primers 5'-CTGGGAATTCATCGCGACCACCATCATCATCATCATTAACATTGGAAGTG GGATAACGG-3' and 5'- CCGTTATCCCACTTCCAATGTTAATGATGATGATGATGGTG GTCGCGATGAATTCCCAG-3' to aid in the pulldown of only full-length protein.

pGB1+Raf-RBD-His6 TAG Mutants

To create each plasmid, site-directed mutagenesis was used to replace an existing amino acid codon with TAG codon within Raf-RBD. In general, PCR primers were designed to have ~15 bases of homology on either side of the TAG mutation and mutagenesis reactions were carried out using manufacturer recommended conditions (Qiagen).

Expression and purification of Bpa containing Raf-RBD

A 30 mL starter culture of LB with Ampicillin and Streptomycin and Chloramphenicol was inoculated and grown overnight at 37˚C. The starter culture was used to inoculate a 1 L Terrific Broth culture containing antibiotic and 1 mM Bpa. This was grown to an OD600 of 0.8 and chilled at 22˚C. Induction was carried out with 0.5 mM IPTG and 0.2% arabinose. After growth overnight, cells were collected by centrifugation for 20 minutes at 6,500 x g and 4˚C. The cell pellet was dissolved in 50 mM $NaPO_4$ pH 8, 300 mM NaCl, and 10% glycerol and with a Roche cOmplete, EDTA-free protease inhibitor tablet. Lysis was done by sonication for 4 minutes with 3 second pulses. To remove insoluble cell debris, lysate was centrifuged at 12,000 x g and 4˚C for 30 minutes. The supernatant was incubated with 1 mL of Ni-NTA agarose resin (Qiagen) for 2 hours at 4˚C. A column was used to wash the resin with 10 mL of lysis buffer with 5 mM imidazole. An imidazole gradient was used to elute protein beginning with 10 mL of 20 mM imidazole, 3 5mL fractions of 40 mM imidazole, and 3 5 mL fractions of 250 mM imidazole for elution. To remove imidazole, protein as dialyzed into $NaPO_4$ pH 8, 150 mM NaCl, and 5% glycerol overnight. Concentration was done using Amicon Ultra PL-10 (10,000 MWCO).

Covalent chemical capture and Pulldown of Crosslinked Adducts.

Mammalian cells overexpressing Ras or Ras mutants were resuspended in 100 µL of lysis buffer (25 mM Tris HCl, 150 mM NaCl, 5 mM $MgCl_2$, 1% NP-40, 5% glycerol). Cells were sonicated at 10% amplitude for 20 seconds. To the lysate, 100 µg of GB1-Raf-RBD (or GB1-Raf-RBD TAG mutants) was added and incubate for 20 minutes at 4˚C. +UV samples were then irradiated for 20 minutes on ice at 365 nm while -UV samples were maintained on ice on bench for this time. Half of the samples were removed to run on the gel without a pulldown of any kind.

Pulldown samples were incubated with 40 µL of Dynabeads magnetic Ni beads for 2 hours at 4˚C. Dynabeads were pre-washed and equilibrated in 3x1mL of lysis buffer. Samples were then washed with 3 x 100µL lysis buffer + 20 mM imidazole and eluted with 3 x 50µL of lysis buffer + 250 mM imidazole with a 5 minute incubation for each wash and elution.

Western Blots

Samples were denatured by heating at 95˚C for 10 minutes in Laemmli 2x Buffer (BioRad) containing 250 mM DTT and run on a 5-15% Bis-Tris gel (BioRad). The separated

proteins were then transferred to PVDF membrane and probed using Western Blot analysis with a 1:1000 dilution of rabbit-anti-His antibody (Abcam) and mouse-anti-Ras (Millipore) in Odessey Blocking Buffer (Lycore) followed by a 1:10,000 dilution of donkey-anti-mouse IRDye 680LT and donkey-anti-rabbit IRDye 800CW. Imaging was done on an Azure C600.

Preparation of Ras Isoforms

Using the pETM11-Ras plasmids generously given by the Fierke Lab, recombinant protein was grown in BL21-AI cells and Terrific Broth rich media. Cells were grown at 37˚C until an OD600 of 0.8 was reached. We induced with 1 mM IPTG and 0.1% arabinose with growth overnight at 18˚C. Cells were harvested via centrifugation for 20 minutes at 6,500 x g and 4˚C.

To lyse cells, the pellet was first dissolved in 10 mM Tris/HCl pH 7.6, 0.1 mM GDP, 20 mM sodium citrate, 50 mM KCl, and 5 mM $MgCl_2$ and 2 mM β-mercaptoethanol with a Roche cOmplete, EDTA-free protease inhibitor tablet. Lysis was done by passing the dissolved pellet through an Avestin EmulsiFlex-C3 high-pressure homogenizer at 10,000-15,000 psi. Homogenization was used to aid in the isolation of inclusion bodies by giving a more uniform lysis. Lysate was centrifuged at 12,000 x g and 4˚C for 30 minutes and supernatant was incubated with 1 mL of Ni-NTA agarose resin (Qiagen) for 3 hours at 4˚C. The pellet was extracted a second time with 10 mM Tris/HCl pH 7.6, 0.1 mM GDP, 20 mM sodium citrate, 50 mM KCl, and 5 mM $MgCl_2$ and 2 mM β-mercaptoethanol for 1 hour at 4˚C with shaking and then re-centrifuged as before. Using a column, the resin was washed with 10 mL of lysis buffer with 5 mM imidazole. An imidazole gradient was used to elute protein beginning with 10 mL of 20 mM imidazole, 3 x 5mL fractions of 40 mM imidazole, and 3 x 5 mL fractions of 250 mM imidazole for elution. To remove imidazole, protein as dialyzed into 50 mM tris-citrate pH 6.5, 50 mM NaCl, 5 mM $MgCl_2$, 0.01 mM GDP and 10 mM β-mercaptoethanol overnight. Concentration was done using Amicon Ultra PL-10 (10,000 MWCO).

Protease Digestion

All protease digestions were prepared according to Promega manufacturer instructions. Briefly, ~8 μg of protein was incubated at 37˚C for 1 hour with 7.5 M urea. This was followed by 20 minutes at 60˚C with 15 mM DTT and finally, 15 minutes in the dark at room temperature with 60 mM iodoacetamide. This final mixture was then diluted with 100 mM tris buffer pH 8 and 1 mM $CaCl_2$. Protease (2 μL ~ 2 μg) was added and digested overnight at 37˚C. Peptides

were desalted with an Oasis 96-well plate (Waters) on a vacuum manifold. 70% acetonitrile was used to elute peptides. After transfer to a vial, a Thermo S*Ava*nt Vacuum Concentrator was used to remove solvent at 45°C.

LC-MS Ras Isoform Identification

Protein digests were prepared in 3% acetonitrile with 0.1% formic acid and 10 fmol/μL alcohol dehydrogenase (ADH, Waters Corp.) in water. In triplicate, peptides were separated using a one-dimensional Waters nanoAcquity UPLC system fitted with a 5 μm C18 (180 μm x 20 mm) trap column and a 1.8 μm HSS T3 analytical column (75 μm x 250 mm). Tryptic peptides were loaded onto the trap column over 3 minutes, followed by analytical separation over a 110 minute gradient (3% acetonitrile to 40% acetonitrile over 92 minutes).
Peptides were analyzed in positive mode on a Synapt G2-S HDMS traveling wave ion mobility time-of-flight mass spectrometer (Waters Corporation). The nanoESI source was set to 70°C and a 3kV potential was applied. Data-independent acquisition was performed as described previously.[4]

Raw data were analyzed with ProteinLynx Global SERVER version 3.0.2 (PLGS, Waters Corporation), searching against a database of the human proteome (downloaded from UniProt on November 9, 2015). The following criteria were applied to perform the search, in addition to the appropriate enzyme specificity: (i) one missed cle*Ava*ge allowed, (ii) carbamidomethyl cysteine as a fixed modification and methionine oxidation as a variable modification, (iii) a minimum of two identified fragment ions per peptide and a minimum of five fragments per protein, and (iv) at least two identified peptides per protein. The false discovery rate (FDR) for peptide and protein identification was set at 1% using a reversed database. Any identified peptides with a calculated mass error greater than 10 ppm were not considered.

References Appendix 2:

1. Prior, I. A., Lewis, P. D., and Mattos, C. (2012) A Comprehensive Survey of Ras Mutations in Cancer, *Cancer Research 72*, 2457-2467.
2. Herrmann, C., Martin, G. A., and Wittinghofer, A. (1995) Quantitative Analysis of the Complex between p21 and the Ras-binding Domain of the Human Raf-1 Protein Kinase, *Journal of Biological Chemistry 270*, 2901-2905.

3. Castellano, E., and Santos, E. (2011) Functional Specificity of Ras Isoforms: So Similar but So Different, *Genes & Cancer 2*, 216-231.
4. Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H., and Tenzer, S. (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics, *Nat Meth 11*, 167-170.