Breaking Ub with *Leishmania mexicana*:

A ubiquitin activating enzyme as a novel therapeutic target for leishmaniasis

Daniel James Harris (2021). Submitted in fulfilment of the requirements for the degree of PhD.

Abstract

Leishmaniasis is a neglected tropical disease, which inflicts a variety of gruesome pathologies on humans. The number of individuals afflicted with leishmaniasis is thought to vary between 0.7 and 1.2 million annually, of whom it is estimated that 20 to 40 thousand die. This problem is exemplary of inequality in healthcare – current leishmaniasis treatments are inadequate due to toxicity, cost, and ineffectiveness, so there is an urgent need for improved chemotherapies.

Ubiquitination is a biochemical pathway that has received attention in cancer research. It is the process of adding the ubiquitin protein as a post-translational modification to substrate proteins, using an enzymatic cascade comprised of enzymes termed E1s, E2s, and E3s. Ubiquitination can lead to degradation of substrate proteins, or otherwise modulate their function. As the name suggests, this modification can be found across eukaryotic cell biology. As such, interfering with ubiquitination may interfere with essential biological processes, which means ubiquitination may present a new therapeutic target for leishmaniasis.

Before ubiquitination inhibitors can be designed, components of the ubiquitination system must be identified. To this end, a bioinformatic screening campaign employed BLASTs and hidden Markov models, using characterised orthologs from model organisms as bait, to screen publicly-available *Leishmania mexicana* genome sequence databases, searching for genes encoding putative E1s, E2s, and E3s. To confirm some of these identifications on a protein level, activity-based probes, protein pulldowns, and mass spectrometry were used. Using an activity-based probe that emulates the structure of adenylated ubiquitin, E1s were identified, and their relative abundance quantified. A chemical crosslinker extended the reach of this probe, allowing the identification of an E2 (LmxM.33.0900). It is noted that *L. mexicana* has two E1s – unusual for a single celled organism. Of these E1s, LmxM.34.3060 was considerably more abundant than LmxM.23.0550 in both major life cycle stages of the *in vitro Leishmania* cultures.

It is important to describe the wider context of these enzymes – what is their interactome, what are their substrates? To study this, CRISPR was used to fuse a proximity-based labelling system, BioID, on genes of interest – LmxM.34.3060 and LmxM.33.0900. The E2 (LmxM.33.0900) was shown to interact with the E1 (LmxM.34.3060), validating the results from the activity-based probe and crosslinker experiments. Due to sequence homology with characterised orthologs, the E2 was hypothesised to function in the endoplasmic reticulum degradation pathway. Immunoprecipitations of a ubiquitin motif, diglycine, were conducted with a view to gathering information on the substrates of ubiquitin. Anti-diglycine peptides included some of those identified by BioID. Experiments examining ubiquitin's role in the DNA damage response were also initiated, as were improvements to the proximity-based labelling system, however these were not followed to completion due to a lack of time and resources.

To examine the possibility of finding novel drug targets in the ubiquitination cascade, recombinant proteins were expressed. LmxM.34.3060 was expressed in a functional form, while a putative SUMO E2 (LmxM.02.0390) was functional after refolding. Expressed LmxM.33.0900 was not functional and could not be refolded into a functional form. Drug assays were conducted on LmxM.34.3060, which found an inhibitor of the human ortholog, TAK-243, to be 20-fold less effective against the *Leishmania* enzyme. Additional assays found an inhibitor that was 50-fold more effective at inhibiting the *Leishmania* enzyme as opposed to its human equivalent - 5'O-sulfamoyl adenosine. Furthermore, a new mechanism of action, inhibiting the E1, for was identified for drugs previously characterised to inhibit protein synthesis. LmxM.34.3060 underwent biophysical characterisation, with structural information obtained using SAXS and protein crystallography. A crystal structure was solved to 3.1 Å, with the in-solution SAXS structure complementary to this. TAK-243 was modelled into the LmxM.34.3060 structure and clashes were predicted, concurring with TAK-243's reduced efficacy against the *Leishmania* enzyme in the drug assays.

This project aimed to characterise the potential of an understudied biochemical system to provide novel therapeutic targets for a neglected tropical pathogen. To achieve this aim it presents the identifications of two E1s, an interactome, a

structure, and a potent, selective inhibitor of a *Leishmania* ubiquitin activating enzyme.

Table of Contents

List of Tables

List of Figures

Acknowledgements

In a just world this section would form the majority of the thesis. For the reason that any degree of success presented here is due to the countless virtues of those who surround me, of which wisdom, patience, prudence, kindness, and humility have been exceptionally obvious. This has been a great privilege for which I will be eternally grateful.

Although I promised myself not to single any individuals out by name, for fear of doing anyone the grave disservice of allowing them to go unacknowledged, I am unable to keep this promise. These lists are in no particular order, and merely stating their name does not begin to do justice to the people involved. For professional mentorship and guidance, without which there would be no thesis, I must thank: Richard Burchmore, Mads Gabrielsen, Patrizia Di Crescenzio, Abdulbaset Kabli, David Ward, Jeziel Damasceno, Catarina De Almeida Marques, Ross Madden, Vincent Geoghegan, Helena De La Torre Olvera, Cesar Ramirez, Thimo Kurz, Ruaidhri Carmody, Suzanne McGill, Aruna Prakash, Danny Huang, Richard McCulloch, Jeremy Mottram, Boris Rodenko, Robert Liskamp, Ola Pasternak, and David Zechel. This thesis depended not only on your brilliance but on your kindness. I must also acknowledge those who made work such a pleasure over the years: Lauren Carruthers, Felix Warren, Will Beyer, Chris Klinger, Emily Allan, Snezhana Akpunarlieva, Marija Krasilnikova, Tahani Alsiari, and Jamal Asseri. You made every day a joy. Finally, the following is a list of individuals on whom I could rely on under all difficulties: Victoria Bolton, Hazel Hamilton, Mariana Pintado Zurita, Marie Lepelletier. I will be forever indebted to you.

More than deserving of their own paragraph are my parents, Diane and Gerald Harris. Without their perpetual support and encouragement my nascent career would have ended before it began. My gratitude cannot be adequately translated into language, so I am forced to leave it unwritten, but know that it is endless.

This work has depended on the help of so many individuals that I am sure I have neglected to thank at least some of them. Although this error will be unintentional, I am all too aware of my defects not to think it probable. I hope

these faults will be viewed with indulgence, for I am as apologetic as I am grateful to those I have missed out.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name: Daniel James Harris

Abbreviations

E1	Ubiquitin activating enzyme
E2	Ubiquitin conjugating enzyme
E3	Ubiquitin ligase
HECT	Homologous to E6AP C-terminus
HPLC	High-performance liguid chromatography
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
LmxM.02.0390	A putative SUMO conjugating enzyme
LmxM.33.0900	A putative ubiquitin conjugating enzyme
LmxM.34.3060	A putative ubiquitin activating enzyme
MS	Mass spectrometry
PBS	Phosphate buffered saline
PSM	Peptide spectral matches
RBR	RING-between-RING
RING	Really interesting new gene
SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
THP1	A human monocyte cell line
Ub	Ubiquitin
Ubl	Ubiquitin-like

Chapter 1 Introduction

1.1 Quipu

It is well known that the Inca empire featured vast stone cities, gold, and treasure. But, at its height in the 15th century, it also consisted of 10 – 12 million people governed by a federal system with power centralised in the capital city, Cusco. Food and commodities were centrally controlled, with subjects issued with necessities from state storehouses in exchange for labour (Cossins, 2018). These storehouses had inventories, just as the subjects had tax records. Yet, the Inca had no written script with which to conduct this extensive administration. What they did have were intricately knotted cords, called "quipu". Quipu encode information in the specific structures of knots that are tied at distinct positions – this is to say that detailed information is recorded in the way a cord is connected to other cords via knots (Harari, 2014). So why open a biochemistry thesis with anthropology? Because of the remarkable parallels between quipu, a pre-Columbian record system, with ubiquitination, the act of modifying proteins with ubiquitin to modulate their function.

1.2 The ubiquitin code

Like quipu, ubiquitination can be thought of as transmitting information, information encoded in the linkages between proteins. Ubiquitination is the term for when ubiquitin (Ub) is added to a substrate protein, typically through forming an isopeptide bond between ubiquitin's C-terminal glycine and the ε-amino group of a substrate's lysine residue (Pickart, 2001). Although, non-canonical ubiquitination deserves a mention, where ubiquitin can be added to the amino group of the protein's N-terminal, as well as cysteine, serine, or threonine residues (forming peptide, thioester, and hydroxyester bonds respectively) (McClellan, Laugesen and Ellgaard, 2019). But what is ubiquitin? Ubiquitin owes its name to its ubiquitous expression. It is a protein composed of 76 compact and tightly hydrogen bonded amino acids, which form a ß-sheet with five anti-parallel ß-strands and a single helical segment. Due to the way in which the ß-sheet appears to "grasp" the helical segment, this fold has become known as the ß-grasp (Vijay-Kumar, Bugg and Cook, 1987); and since first being documented in ubiquitin it has been found in a diverse array of proteins, becoming a defining feature of the ubiquitin-like (Ubl) protein family (Burroughs *et al.*, 2007). Notably, ubiquitin has seven lysine residues spread across all faces

of the molecule, pointing in different directions (Figure 1-1). As ubiquitin is a protein with lysine residues, and ubiquitin can be added to proteins at lysine residues, ubiquitin can be added to itself, forming polyubiquitin chains. The plethora of lysines on both substrate and ubiquitin moieties are integral to understanding the ubiquitin code – each linkage results in a distinct chain conformation, and the more distinct conformations the more diverse the information that can be encoded. For instance, a polyubiquitin chain may be compact, in which adjacent ubiquitin moieties interact with one another; or open, in which the only interface between moieties is the linkage site. Examples of compact chains include Lys48-linked diubiquitins, where interactions between hydrophobic patches centred on Ile44 cause the moieties to pack tightly together. And so in Lys48-linked tetraubiquitin, where two of these diubiquitins are closely arranged (Eddins *et al.*, 2007). Yet, in another compact chain, moieties linked through Lys11 have their hydrophobic Ile44 patch facing out into the solute (Bremm and Komander, 2011). In contrast, open conformations, like those made by Met1- and Lys63-linked chains, have high conformational freedom (Sims and Cohen, 2009). It is this structural diversity that forms the foundation of the ubiquitin code.

Figure has been removed due to copyright restrictions.

Figure 1-1 Ubiquitin and its topologies a) Structure of ubiquitin, showing its seven lysine residues and methionine 1, all free amino groups available for building polyubiquitin chains. **b)** The variety of topologies that ubiquitin can have in a cell. Both figures were taken from “The Ubiquitin Code” by Komander and Rape (Komander and Rape, 2012).

1.3 Activating ubiquitin

Of course, any functioning code demands specific assembly. In the same way a thesis introduction can only function if letters are arranged to form meaningful words, or a quipu can only function if the right knots are tied in the proper positions, ubiquitination can only function if the same enzymes catalyse the formation of the same products each time they act on a substrate. These enzymes can be divided into three cooperating families: firstly, there is the ubiquitin activating enzyme (E1), then the ubiquitin conjugating enzyme (E2), and finally the ubiquitin ligase (E3) (Figure 1-2). The goal of an E1 is to activate the C-terminus of ubiquitin, in preparation for its ultimate attack on a substrate's amino group. E1s achieve this through adenylation, which occurs when the hydrophobic patches between ubiquitin and the E1 interact, causing ubiquitin's carboxy-terminal flexible tail to reach into the E1's ATP binding pocket and attack the α-phosphate of the ATP. Remarkably, at this point the E1 undergoes massive domain rearrangements around the bound ubiquitin. By moving a catalytic cysteine approximately 130° from its prior position, over a distance of 30 Å, it enables this catalytic cysteine to attack the high-energy adenylate, producing a covalent thioester linkage between the cysteine sulphydryl and the C-terminus of ubiquitin (Schulman and Wade Harper, 2009). During this thioesterification reaction, the domain containing the cysteine active site changes from a helical conformation to an extended one, allowing it to reach into the adenylation pocket. So the movements of the cysteine domain can be categorised into two conformations: open (when adenylating ubiquitin), and closed (when the adenylated ubiquitin reacts with the cysteine active site) (Lorenz *et al.*, 2013). And so conformational flexibility permits biochemical flexibility - having various locales for multiple active sites allows the catalysis of three reactions: adenylation, thioesterification, and transthiolation (the latter being the transfer of ubiquitin from an E1 to an E2, more on which below).

Figure 1-2 Overview of the ubiquitination system. a) The functions of the enzymes involved in ubiquitination, show an E1 adenylating ubiquitin, and an E2/RING E3 complex conjugating ubiquitin to a substrate. Figure taken from Wikipedia article "Ubiquitin" (Rogerdodd, no date) **b)** The demographics of the ubiquitination enzymes. Figure taken from Cao and Mao (Cao and Mao, 2011).

In humans two E1s can activate ubiquitin – UBA1 and UBA6. Both enzymes are pseudosymmetrical monomers, both approximately 118 kDa. For more than two decades it was thought that UBA1 was solely responsible for charging E2s, as is the case in lower eukaryotes, but then UBA6 was discovered to also charge ubiquitin (Jin *et al.*, 2007). This was surprising as, despite both enzymes being monomers of nearly identical mass, UBA1 and UBA6 share only 42.41% identity. Furthermore, UBA6 activates the Ubl molecule FAT10 as well as ubiquitin. Nonetheless, UBA6 activates ubiquitin as efficiently as UBA1 and transfers it to an E2, UbcH5b, with equal efficiency (Pelzer *et al.*, 2007). So why have two enzymes doing the same job? Given that UBA6 is essential for early embryogenesis in mice (in a capacity unrelated to FAT10), it is not merely redundant (Chiu, Sun and Chen, 2007). In fact, UBA6 has its own E2, Use1, which UBA1 cannot charge (Jin *et al.*, 2007). The significance of this is that there is a level of control even at this earliest stage of ubiquitination – an E1 may not come into direct contact with ubiquitin's substrates, but nonetheless influences which substrates get ubiquitinated.

This may be the case in humans, but which other organisms possess two ubiquitin E1s? UBA6 can be found in vertebrates (and, curiously, the sea urchin)

but is not detected in insects, worms, fungi or plants, which have a single E1 to charge ubiquitin (Groettrup *et al.*, 2008). This regression through evolutionary history begs the question of the origin of E1s. Given that complete toolkits for conjugating ubiquitin and ubiquitin-like proteins exist in all main groups of eukaryotes, it would appear that ubiquitination predates the last eukaryotic common ancestor, with components even found in several archaeal groups (Grau-Bové, Sebé-Pedrós and Ruiz-Trillo, 2015). As such, ubiquitination must have its origin in prokaryotes. The bacterial proteins molybdopterin-converting factor subunit 1 (MoaD) and thiamine biosynthesis protein S (ThiS) both harbour the ß-grasp fold (Lake *et al.*, 2001; Wang *et al.*, 2001). These proteins carry sulphur to be assimilated into molybdopterin and thiazole, respectively; where molybdopterin is a small molecule cofactor and thiazole is a precursor in thiamine production. MoaD and ThiS are activated by C-terminal adenylation, which is carried out by the enzymes molybdopterin biosynthetic enzyme B (MoeB) and ThiF, respectively. Furthermore, MoeB and ThiF possess sequence homology with the E1 domain responsible for binding and adenylating ubiquitin, the common building block of all E1s. To reiterate, MoaD and ThiS possess a fold typical of ubiquitin and are only activated after adenylation of their C-terminals. The enzymes MoeB and ThiF are responsible for this, and embody the minimal component of E1s that can recognise and adenylate ubiquitin (Lake *et al.*, 2001; Wang *et al.*, 2001). Although MoeB and ThiF may be the ancestral E1s, there are no obvious prokaryotic antecedents of E2s. Perhaps this is why a uniting feature of all the various ubiquitin and ubiquitin-like pathways is activation by an E1-like enzyme (Hochstrasser, 2000). The far-reaching historical pedigree of the E1s, coupled with their lasting sequence similarities and conserved mechanisms, provides a sense of their importance to innumerable organisms, and certainly to all eukaryotic cells.

Looking toward the next stage of the ubiquitination cascade, before an E2 can accept an activated ubiquitin, a second ubiquitin must be bound to the E1 adenylation site. Therefore, the E1 is asymmetrically loaded with two ubiquitin molecules: one covalently bonded to a catalytic cysteine and ready to be handed off to an E2; and a second non-covalently associated at the adenylation active site. This second ubiquitin molecule does seem to accelerate the transfer of the ubiquitin at the cysteine active site, although what mechanism is at work is

unknown (Schulman and Wade Harper, 2009). When it comes to recognising the E2, the E1 does so in a tripartite manner, involving three domains. Of interest are the mobile domains: the ubiquitin fold domain of the E1 (a C-terminal domain, named due its use of a ß-grasp fold), and the ubiquitin molecule at the E1's cysteine active site. Upon transfer of ubiquitin from the E1 to the E2, a third of the tripartite interactions are lost. This facilitates movement of the ubiquitin fold domain, which swings towards the E1 and away from the E2 (Huang *et al.*, 2007). Having lost two thirds of its interacting surfaces, the E2~ubiquitin is released from the E1. The mobility of these two domains also adds directionality to the cascade, decreasing the chance of backwards movement of ubiquitin from an E2 to an E1.

1.4 Conjugating and ligating ubiquitin

E2s, ubiquitin conjugating enzymes, are often viewed as mere middle men in the ubiquitination cascade, simply carriers of ubiquitin. Yet, despite being only twice the size of their substrate, containing just a solitary active site, E2s generally catalyse two types of reactions: transthiolation (transfer from a thioester to a thiol group) and aminolysis (transfer from a thioester to an amino group) (Stewart *et al.*, 2016). The defining motif of these enzymes is their histidine–proline–asparagine tripeptide that generally appears eight amino acids, on the C-terminal side, away from an active site cysteine. This motif assists in the folding of the active site (Jones *et al.*, 2001). This motif is part of a domain common to all E2s – the approximately 150 amino acid UBC (UBiquitin Conjugating) domain, which is the catalytic core domain, the minimum unit of an E2 necessary for enzymatic activity (Valimberti *et al.*, 2015). The UBC domain has an α/ß-fold, usually with four α-helices and a four stranded ß-sheet (Stewart *et al.*, 2016) (Figure 1-3). The majority of human E2s comprise of only the UBC domain, yet many have short extensions on their N- and/or C-terminals to modulate their functions. Typically, these extensions are intrinsically disordered, but some have secondary structures that contact the UBC domain (Schelpe *et al.*, 2016). A few E2s have an additional structured domain that is connected to the UBC domain, and another significant minority are part of large multidomain proteins (Stewart *et al.*, 2016). All this is to say that the common UBC fold of an E2 has been adjusted for use in particular systems. Yet a fuller understanding of the approximately 40 human E2s can be better obtained when

the intrinsic reactivity of an E2 is contrasted with its reactivity when part of an E2-E3 complex (van Wijk and Timmers, 2010).

Figure has been removed due to copyright restrictions.

Figure 1-3 The UBC fold of UBE2D2 Conserved structural features are indicated by red α-helices and yellow β-sheets. The main site for interactions with E1s and E3s is indicated by the components labelled H1, L1, and L2. The catalytic cysteine is shown by magenta spheres. This figure was taken from van Wijk and Timmers (van Wijk and Timmers, 2010).

To state the obvious, E2s can exist as E2~ubiquitin conjugates. Yet, the implications of this are perhaps less conspicuous – if E2~ubiquitin conjugates are already wired to react, why don't they? What keeps their reactivity low, and how do E3s stimulate E2s? Even in the absence of an E3, most E2s show some innate ability to transfer ubiquitin to free lysine (Stewart *et al.*, 2016). Yet, unexpected data has still been obtained when assessing this intrinsic reactivity of E2~ubiquitin conjugates. One example is Ube2W, an E2 that functions as part of the Fanconi anaemia complex, which responds to DNA damage. Ube2W possesses no inherent ability to react with free lysine, instead attaching ubiquitin to the N-terminal α-amino group of proteins, forming fusion proteins (Vittal *et al.*, 2014). This is in direct contrast to Ube2W's reactivity in the presence of its cognate E3s (BRCA1/BARD1 and FANCL, the latter also a part of the Fanconi anaemia complex), which induce it to ubiquitinate the free amino group of lysine (Zhang *et al.*, 2011). The cause of this is Ube2W's unusual UBC domain, which has a disordered C-terminal extension that hovers near the active site. This extension can accommodate a variety of substrate N-terminal backbones, as Ube2W's disorder renders it non-specific (curiously, it cannot catalyse polyubiquitination, as the N-terminal of ubiquitin is too highly structured to be a good substrate for Ube2W) (Vittal *et al.*, 2014). A similar example is found with Ube2J2, an E2 that degrades misfolded proteins, functioning as part of the endoplasmic reticulum associated degradation pathway. Ube 2J2 can ubiquitinate lysine residues, but when exposed to an E3 of viral origin (the RING E3 murine K3 from the mouse γ-herpesvirus), Ube2J2 will

preferentially ubiquitinate serine and threonine residues on the major histocompatibility complex (Wang *et al.*, 2007). Supporting evidence is that the yeast homolog of Ube2J2, Ubc6, is essential to degrade a lysine-less substrate (Boban, Ljungdahl and Foisner, 2015). Yet, despite these interactions, any intrinsic ability of Ube2J2~ubiquitin conjugates to react with serine or threonine, and so catalyse hydroxyl reactions, has yet to be demonstrated (Stewart *et al.*, 2016). These cases, albeit from unusual E2s, underline the changeable natures of ubiquitin conjugating enzymes.

Working our way down the cascade we come to the ubiquitin ligases, E3s. These play a decisive role in selecting substrates. Given the innumerable substrates of ubiquitin, it is no surprise that E3s can number in their hundreds, with the human genome encoding more than 600 (Li *et al.*, 2008). Thankfully, these multitudes can be broadly grouped into three families, families that are dependent on the domain used to bind E2~ubiquitin. RING E3s (Really Interesting New Gene) act as a scaffold to catalyse the direct transfer of ubiquitin from an E2 to a substrate, lacking a catalytic cysteine themselves. This distinguishes them from the HECT (Homologous to E6AP C-Terminus) and RBR (RING-Between-RING) families, whose catalytic cysteine receives ubiquitin from an E2, before passing it on to the substrate (Buetow and Huang, 2016). However, E3s are far from equally distributed between these families, with the human genome encoding approximately 600 RING, 28 HECT, and 14 RBR E3s (Buetow and Huang, 2016). A natural question arises: when E2s and E3s are working together, which determines the substrate? That responsibility is left to the last party holding the activated ubiquitin – an E2 when RING E3s are involved, and an E3 in the circumstances when the E3 has a catalytic cysteine (Stewart *et al.*, 2016).

One function of E3s is to dramatically increase the intrinsic reactivity of many E2s. For example, E2s of the Ube2D family will react slowly with lysine if no E3 is present, but rapidly react when presented with the RING domain of an E3 (Wenzel *et al.*, 2011). Complexes of E2s and E3s form using the E2's UBC binding domain, (specifically its residues in loops four and seven, as well as those in helix one), which interact with either the E3's HECT or RING domain with a moderate to weak affinity (Zheng *et al.*, 2000). This same region of the E2 is also responsible for E1 interactions, which, given its necessity in multiple

protein interactions, may be why most E2s share similar electrostatic properties in this area (Sheng *et al.*, 2012). Furthermore, that an E2 uses one locale to interact with multiple partners means that binding an E1 or an E3 is a mutually exclusive event (Eletr *et al.*, 2005). This also means that an E1/E2/E3 complex is unlikely to form, so an E2 will need to be recycled before it can be used to add additional ubiquitin to a substrate.

1.5 Reading the ubiquitin code

So that is how the code is written, but to what end? Why bother with a multi-layered system comprising of hundreds of enzymes, what information could possibly justify such an elaborate system? How can information even be derived from the arrangement of small molecules? And are there other systems based on similar principles?

To answer the last question first, ubiquitin is far from alone. As mentioned above, the ß-grasp fold provides a similar core structure for the nebulous constellation of ubiquitin-like proteins, which features 17 members from 9 families. This provides evidence of common ancestry from pre-existing parts that had already experienced extensive diversification in prokaryotes (Hochstrasser, 2009). One such cousin of ubiquitin is SUMO (Small Ubiquitin-related MOdifier), which is also involved in a myriad of fundamental cellular processes, such as DNA replication, transcription, cell cycle regulation, DNA damage repair, chromatin organization, ribosome biogenesis, pre-mRNA splicing, nuclear trafficking, signal transduction and protein degradation (Pichler *et al.*, 2017). Ironic given its title, but mammalian SUMO proteins are typically larger than ubiquitin, extending to approximately 100 amino acids. Mammalian cells express up to five SUMO paralogs, although *Saccharomyces cerevisiae* has only one SUMO protein, which is essential (Pichler *et al.*, 2017). Typically, SUMO is a monomer conjugate, in which it is often attached at a SUMO consensus motif. Additionally, there is a growing body of evidence that poly-SUMO chains are employed in at least some aspects of SUMO function, although, unlike ubiquitin, there is no evidence that different linkages result in different outcomes for the substrate (Ulrich, 2008). A notable role for poly-SUMO chains is their ability to signal for ubiquitination of a substrate, thereby promoting its degradation (Ulrich, 2008). As for attaching SUMO, in humans the sole SUMO E1 is a heterodimer, comprised of subunits

termed Sae1 and Sae2, where only Sae2 is catalytically active. Furthermore, there is only one SUMOylation E2 present in humans, Ubc9.

To return to ubiquitin, even though the same principles underlie its cousins - information is encoded its structural diversity. This diversity can be read by effector proteins with ubiquitin binding domains, which translate the modifications into specific outcomes. There are various facets of the chain structure that can be recognised, for example the distance between ubiquitin molecules. Rap80 (a protein involved in recruiting proteins for DNA repair) has multiple ubiquitin binding domains that interact with ubiquitin's Ile44 patch, although these ubiquitin binding domains are separated by a helix. This helix acts somewhat like a ruler, spacing the domains in such a way that only the open conformation of Lys63-linked chains (as opposed to the closed conformation of Lys48-linked chains) have the correct distance between Ile44 patches to be recognised (Sims and Cohen, 2009). Alternatively, the flexibility of the chain can be used - the TAK1 binding protein 1 (used for signalling nuclear factor-kB) employs Npl4-like zinc fingers to distinguish between Lys63- and Met1- linked chains. This is achieved as the Npl4-like zinc fingers bind to two Ile44 hydrophobic patches on subsequent moieties, which contorts the ubiquitin chain in such a way as to displace Met1 from the distal unit's C-terminus. Given this, it is impossible for Met1-linked chains to bend in the same way, and so, despite their structural similarity to Lys63-linked chains, Met1-linked chains are not recognised by TAK1 binding protein 1 (Sato *et al.*, 2009). Interestingly, ubiquitin is highly conserved from yeast to human, featuring only three conservative changes, rendering it one of the most highly conserved of all eukaryotic proteins (Sharp and Li, 1987). This suggests ubiquitin's structure is subject to high evolutionary pressure, implying that ubiquitin binding domains recognise most of ubiquitin's surfaces.

As to the function of ubiquitin, it is an all-purpose molecule, touching on all aspects of eukaryotic biology. Myriad mechanisms are dependent on ubiquitin: cell cycle advancement, apoptosis, cell division, cellular differentiation, organelle biogenesis, quality control in protein synthesis, protein transport, stress responses, histone regulation, and DNA repair are just a few examples (Weissman, 2001). This multitasking is why equally many human pathologies are

linked to ubiquitination: muscle dystrophies, metabolic syndromes, inflammatory disorders, infections, autoimmunity, neurodegenerative diseases, and cancer (Popovic, Vucic and Dikic, 2014). When considering how many aspects of life involve ubiquitination, it comes as no surprise that, in 2004, the highest honour in science, the Nobel Prize, was awarded to Aaron Ciechanover, Avram Hershko, and Irwin Rose for their discovery of ubiquitin-mediated protein degradation (Editorial, 2004). Before this research protein degradation was an unfashionable topic, clearing the cell of rubbish was thought to be no more exciting than clearing household waste. Yet these investigators demonstrated the level of specificity and control operating in protein degradation. Furthermore, now with the benefit of additional decades of research, it is has become apparent that ubiquitin has truly lived up to its name - ubiquitous in both expression and function.

1.6 *Leishmania* and Leishmaniasis

The protozoan parasites of the *Leishmania* genus have had a colossal impact on humanity across aeons, and even today they continue to annually disfigure 0.7 – 1.2 million people, whilst killing between 20 to 40 thousand (Alvar *et al.*, 2012). *Leishmania* are the second largest parasitic killer in the world because they are the causative agents of various forms of leishmaniasis. *Leishmania* parasites, and the disease they cause, are spread by their vector: infected female sand flies of the phlebotomine subfamily. These flies are widely spread across the globe, afflicting vast swathes of people from Brazil to Ethiopia, from India to Iraq, and from Spain to Kazakhstan, to give a just a few examples of the 98 countries that reported endemic leishmaniasis transmission (Alvar *et al.*, 2012). Yet, although the disease can be found almost worldwide, the parasite discriminates against the poorest and most vulnerable in society disproportionately – becoming both a cause and a consequence of poverty (Okwor and Uzonna, 2016).

The terrible burden inflicted by *Leishmania* has a long and storied history, predating humanity and even the continents as we know them today. The earliest *Leishmania*-like fossil was found inside of an extinct species of sand fly that had been trapped in 100 million-year-old amber from the Cretaceous period (Poinar and Poinar, 2004). This ancient origin lead to one hypothesis that a monoxenous insect flagellate disseminated into mammals, whereby the

diversification of mammals resulted in the evolution of new genera of parasites (Espinosa et al., 2016). However, further variety in the *Leishmania* subgenera is due to the breakup of the supercontinent Gondwana, which placed some hosts and parasites in what is now South America and others into Africa, where distinct evolutionary paths were taken. Not to mention that further complexity is added as prehistoric hosts took migratory routes along land bridges that have not existed in our recorded history (Momen and Cupolillo, 2000). To skip over millions of years, this fascinating tale continues into the comparatively recent history of the ancient Egyptians, as four approximately 4000 year-old mummies were found to contain fragments of *Leishmania* mitochondrial DNA (Zink *et al.*, 2006). Subsequently, in the Middle Ages of the New World, Incas were using pottery figurines to depict two different forms of leishmaniasis, and the Conquistadores were warning one another of the "deep and fetid sores" insect bites could cause (Novo et al., 2016). Around the turn of the 20th century a Russian military surgeon named Piotr Fokich Borovsky was the first to acknowledge that the lesions typical of leishmaniasis contained protozoans, yet, as his discovery was published in an obscure Russian journal, his work went sadly unaccredited (Hoare, 1938). Those who received the acclaim a few years later were the Glaswegian Professor Sir William Boog Leishman, who was serving in the British army in colonial India, and an Irish medical officer in the Indian Medical Service named Charles Donovan. Donovan and Leishman made their breakthroughs independently mere weeks apart and had their discovery named in their honour – thus a new genus was born with *Leishmania donovani* (Leishman, 1903).

What made *Leishmania* so notable throughout human history is the variety of gruesome pathologies it incurs. There are over 20 *Leishmania* species that have had ample time to diversify, thereby providing a nuanced yet expansive range of symptoms. Nonetheless, the disease can be broadly classified into three main forms: cutaneous, mucocutaneous, and visceral. Cutaneous leishmaniasis is, in general, characterised by an erythema that appears at the site of a sand fly bite, after a variable prepatent period. This erythema in turn advances into a papule, then a nodule, before ulcerating to become the unsightly lesion that typifies cutaneous leishmaniasis (Reithinger *et al.*, 2007). Although rarely life threatening, this disease is nonetheless debilitating to those who suffer it

because, as with any disease that resembles leprosy, there is considerable stigma attached – examples include women being abandoned by their husbands and children being barred from school (Okwor and Uzonna, 2016). Additionally, a lesion on a hand or foot can be disabling for the sufferer, preventing them from working or walking. Mucocutaneous leishmaniasis is a serious complication of cutaneous leishmaniasis, in which the parasites spread to mucous tissues through the lymphatic or vascular systems. This can cause ulceration of the nasal mucosa, lips, cheeks, soft palate, pharynx or larynx. Not only is this difficult to treat and likely to develop a secondary bacterial infection, but it can also interfere with the patient's eating and respiration (Reithinger *et al.*, 2007). However, the vast majority of deaths are due to visceral leishmaniasis, which is when cells of the reticuloendothelial system are parasitized. This can affect a number of internal organs, although typically the liver and spleen are particularly susceptible (Allison, 1993). Unfortunately, this is not all the sufferer has to worry about - a study in Nepal found that average treatment cost for a bout of visceral leishmaniasis was more than the annual household per capita income, with some families forced to take out high-interest loans or sell their livestock (Rijal et al., 2006); and a similar study in India noted that, for households with a visceral leishmaniasis patient, the estimated annual household expenditure on visceral leishmaniasis was higher than the average household assets (Sarnoff *et al.*, 2010). Despite these high costs, the afflicted have little choice but to seek help as the disease is almost invariably fatal if left untreated over a two-year period (Allison, 1993).

1.7 Treating leishmaniasis

Treatment for leishmaniasis is far from ideal. Without a vaccine, treatment is based on chemotherapy. Yet this chemotherapy is based on very few drugs, and these few drugs have serious shortcomings. Toxicity, difficult administration, prolonged treatment, high production costs, and severe adverse reactions characterise current leishmaniasis treatments. Perhaps these flaws could be justified if the treatments were effective at completely eliminating parasites, but that is not the case for the most commonly used drugs. A problem which emerging drug resistance will only exacerbate (Sasidharan and Saudagar, 2021). Despite being over six decades old, with an obscure mechanism, pentavalent antimonials are the still the most frequently used anti-leishmanial drug. It has

been noted that pentavalent antimonials are prodrugs, as the pentavalent antimonite is reduced to trivalent antimonial, which is a more potent anti-leishmanial. This is hypothesised to occur in the acidic environment of a macrophage phagolysosome (the human host cell of the parasite) (Aruleba *et al.*, 2020). Yet, serious adverse effects can include muscle pain, joint pain, abdominal pain, headaches, lethargy, nausea, anorexia, vomiting, and even fatal cardiac arrhythmia (De Menezes *et al.*, 2015; Sasidharan and Saudagar, 2021). This is due to accumulation in tissues, that can induce pancreatitis, pancytopenia, nephrotoxicity, cardiotoxicity. In turn, this accumulation is caused by the prolonged course of drug administration – typically large daily parenteral doses of 20 mg of pentavalent antimonials/kg/day for 20–30 days (Frézard, Demicheli and Ribeiro, 2009). The severe side effects and long treatment courses therefore demand hospitalisation of patients, who may, understandably, abandon treatment, which furthers drug resistance (Sundar, 2001).

As drug resistance to pentavalent antimonals increases, amphotericin B becomes the go-to drug for visceral leishmaniasis (Mann *et al.*, 2021). Amphotericin B was discovered by Donovick, Gold, Pagano, and Stout in 1955 (Donovick *et al.*, 1955). A product of the bacteria *Streptomyces nodosus*, isolated from a soil sample in Venezuela's Orinoco River region, it was initially licensed as an antifungal (Cavassin *et al.*, 2021). Its mechanism revolves arounds its amphipathic nature, elongated structure, and ability to bind sterols. One such sterol is ergosterol, a major component of *Leishmania*'s cell membrane. When the hydrophobic portion of amphotericin B binds to ergosterol, within the cellular membrane, its polar head faces away from the membrane. As such, this elongated molecule inserts itself within the membrane, forming pores that lead to the egress of electrolytes and other small metabolites from the intracellular medium, ultimately causing cell death (Cavassin *et al.*, 2021). However, another sterol is cholesterol, which substitutes ergosterol in human cell membranes. Fortunately, amphotericin B has a lower affinity for cholesterol, but still enough of an affinity to damage the host kidneys, heart, and blood. (Cavassin *et al.*, 2021). As a consequence, inducing rigor, chills, and fever (De Menezes *et al.*, 2015). Initially, due to poor absorption through the gastrointestinal tract, it was administered intravenously, with a 95% cure rate obtained using daily doses of 1 mg/kg for 20 days (Aruleba

et al., 2020). Yet, the aforementioned side effects, exacerbated by the long treatment regimen, undermine its efficacy. As such, liposomal formulations of amphotericin were developed to reduce its adverse effects and better its pharmokinetic properties. These drugs are a success in that they are less toxic, but they are expensive and demand cooling due to instability at warmer temperatures, complicating their use in the developing, tropical nations in most need (De Menezes *et al.*, 2015).

Another treatment option is paromomycin. An aminoglycosidic antibiotic, with both antileishmanial and antibacterial capacities, it can cure both visceral and cutaneous leishmaniasis (although it is more effective with the latter). This drug can be administered intramuscularly, intravenously, and even topically (Matos *et al.*, 2020). It has an efficacy similar to amphotericin B, yet lacks the more severe side effects of that drug (Wiwanitkit, 2012). For visceral lesihmaniasis, dosing of 15 mg/kg for 21 days gives a 95% cure rate, with associated side effects of nephrotoxicity and ototoxicity (Aruleba *et al.*, 2020). Given the formation of skin lesions and paromomycin's effectiveness against both new and old world leishmaniasis, topical formulations are the most commonly used (Aruleba *et al.*, 2020). The mechanism of paromomycin is not fully elucidated. It has been suggested that mitochondria are targeted, although it is also known to inhibit translocation and recycling of ribosomes, so inhibiting protein synthesis is thought to be a primary mechanism (Singh, Kumar and Singh, 2012). Its inexpensive nature is a huge boon; however, resistant cell lines are easy to induce experimentally, and a Nepalese strain of *L. donovani* already shows high natural paromomycin resistance (Aruleba *et al.*, 2020). Furthermore, paromomycin resistant strains are every bit as transmissible as the paromomycin susceptible parental lines (Hendrickx *et al.*, 2020).

A recent addition to the armamentarium is miltefosine. An alkylphosphocholine, originally developed to treat cancer, it is considered a major breakthrough in leishmaniasis treatment – the first oral drug (Singh, Kumar and Singh, 2012). Coupled with its tolerable side effects (primarily mild vomiting and diarrhoea, though nephro- and hepatotoxicity have been reported), it is a recommended treatment for both cutaneous and visceral leishmaniasis (Sundar and Olliaro, 2007). As is often the case, the precise mechanism of action of is unclear.

Miltefosine seems to disrupt mitochondrial function, lipid metabolism, and cause apoptosis. Perhaps this is due to the drug's ability to disrupt calcium homeostasis, inducing intracellular calcium accumulation. Additionally, miltefosine may impact the host immune response, through increased IFN-γ, TNF-α, IL-12, and nitric oxide production (Aruleba *et al.*, 2020). Interplay with the host's immune system could explain a weakness of miltefosine – its variable cure rate. For example, in Colombia miltefosine achieved a cure rate of 91% for cutaneous leishmaniasis caused by *Leishmania panamensis*, but a cure rate of 53% was found in Guatemala against cutaneous disease caused by *Leishmania mexicana* and *Leishmania braziliensis*. To compare like-for-like, miltefosine had a 75% cure rate against *L. braziliensis* in Brazil (Aruleba *et al.*, 2020). Additionally, if host immunity contributes to miltefosine's mechanism of action, it could explain the high risk of relapse associated with this treatment. Alternatively, this could be explained by miltefosine resistant strains, of which several are already known (Sasidharan and Saudagar, 2021).

1.8 Leishmania biology

Leishmania have a fascinating digenetic life cycle, a single cell with multiple morphologies and mixed reproductive methods, depending on its host organism (Rougeron, De Meeûs and Bañuls, 2017). The two major morphological stages of the parasite are the promastigote and amastigote; the former present in the insect vector and the latter found in in the mammalian host (Mann *et al.*, 2021). Promastigotes are characterised by their slender form (a cell body measuring 15 to 20 μm in length and 1.5 to 3.5 μm in width) and flagella (15 to 28 μm in length) (Sasidharan and Saudagar, 2021). The flagella provides them with motility and allows them to attach to the gut of the sandfly. After migrating through the sandfly gut, the promastigotes reach the proboscis, ready for transmission to a mammalian host. With the bite of the sandfly the parasite is injected into the skin, where it is phagocytosed by macrophages and other phagocytic cells. Within this new environment the parasite adopts its amastigote form: ovoids measuring 2 to 4 μm (Torres-Guerrero *et al.*, 2017). Within this usually hostile environment, the amastigote proliferates, eventually causing its host cell to rupture. This begins a chain of further infection, leading to the pathologies discussed earlier. This elaborate life cycle, which demands

coordination of protein levels as the parasite develops in response to changing environments, is shown in Figure 1-4.

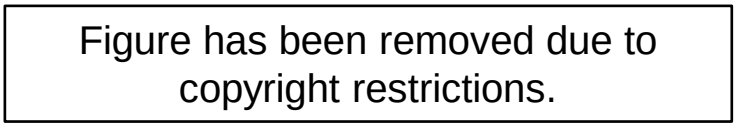


Figure 1-4 The life cycle of *Leishmania.* Image taken from CDC website (CDC, 2020).

There have been multiple roles noted for ubiquitin in *Leishmania*. Firstly, a typical eukaryotic proteasome, coupled with the toxicity of proteasome inhibitors, implies the parasite employs a conserved pathway for protein turnover (Paugam *et al.*, 2003). Although the precise role of the proteasome in *Leishmania*'s differentiation requires further exploration, it must be noted the parasite remodels itself over the course of its life cycle, not least diminishing its flagella, so it seems plausible that proteasome mediated degradation may be involved (Besteiro *et al.*, 2007). On the note of protein degradation, the ubiquitin-like protein ATG5 is required for autophagy. When ATG5 was knocked out, *Leishmania* cells were characterised by dysfunctional mitochondria, poorer differentiation, morphological abnormalities, and much reduced virulence (Williams *et al.*, 2012). This demonstrates the need for protein degradation and ubiquitin-like proteins in the parasite.

An additional ubiquitin-like protein is SUMO. However, for an examination of its roles we must turn to *Trypanosoma* parasites, which belong to the same kinetoplastid class as their *Leishmania* cousins. Liao *et al.* used RNAi interference to knock down SUMO in *T. brucei*, significantly inhibiting the growth of SUMO-deficient cells. Furthermore, with chromosome segregation failing, their cell cycle was arrested at the G_2/M phase, indicating an essential role for SUMO in *T. brucei*'s cell cycle (Liao *et al.*, 2010). Klein, Droll, and Clayton expand the SUMO repertoire by identifying a *T.brucei* SUMO specific protease and a SUMO E2 (Klein, Droll and Clayton, 2013). Extending this topic to *Typansoma cruzi*, Bayona *et al.* identify its SUMOylation machinery and substrates, indicating conserved functions for SUMO even in this early divergent eukaryote.

Another thought-provoking aspect of *Leishmania* biology is its constitutively expressed genome but differentially expressed proteome – if mRNA levels are constant, but protein levels vary, then there must be a post-transcriptional method controlling protein abundance (Leifso *et al.*, 2007). It is possible that one such method may be the post-translational degradation of proteins expressed unnecessarily.

Alternatively, once transcribed, unnecessary RNA may be degraded before translation. One such protein that could achieve this is LdCSBP, which binds to an octameric sequence in the UTR of mRNAs, and exhibits riboendonuclease activity. *In vitro*, LdCSBP is also mediated through ubiquitination, in which monoubiquitination downregulates its activity (Bhandari *et al.*, 2011). When LdCSBP's zinc finger domain is ubiquitinated intramolecular forces interact with the ubiquitin binding CUE domain, regulating the enzyme's endonuclease activity (Guha *et al.*, 2014). Furthermore, the anti-leishmanial drug paromomycin also inhibits the endonuclease activity of LdCSBP, suggesting that interfering with the parasite's RNA metabolism may provide novel therapeutics (Guha *et al.*, 2014).

Furthermore, the ubiquitin-like protein Ufm1, together with its corresponding E1 and E2, has been identified in *Leishmania donovani* (Gannavaram *et al.*, 2011). One protein modified by Ufm1 is the α-subunit of the mitochondrial trifunctional protein, and when Ufm1 was knocked out there was a decrease in ß-oxidation of fatty acids, and, consequently, a decrease in the amount of acetyl-CoA

produced. This resulted in amastigote growth defects both *in vivo* and *ex vivo*, and reduced survival of the parasite (Gannavaram *et al.*, 2012). Additionally, knockouts of the protease responsible for Ufm1 processing, Ufsp, also result in reduced survival. This is interesting as amphotericin B also seems to inhibit Ufsp activity, further demonstrating the feasibility of targeting leishmanial ubiquitin-like pathways with drugs (Gannavaram *et al.*, 2014).

Recently, Burge *et al.* constructed a null mutant library of ubiquitination enzymes identified through bioinformatic domain searches (Burge *et al.*, 2020). They describe how knockouts could not be generated for one of the two ubiquitin E1s and four of the thirteen ubiquitin E2s, indicating their potential essentiality. For the knockout of the other E1, a loss of fitness was displayed in *in vitro* models, and a 20-fold loss of fitness was apparent in an animal host model, in which knockouts were unable to sustain an infection. They also show how two E2s form a complex essential for the promastigote to amastigote transformation, and how this E2 complex has different substrates than its composite parts. All this implies a complex, essential system integral for parasite survival.

Other authors have reached the conclusion the *Leishmania* ubiquitination system could provide promising drug targets (Duncan *et al.*, 2011; Sinha and Sarkar, 2013; Muñoz *et al.*, 2015; Bibo-Verdugo *et al.*, 2017; Boer and Bijlmakers, 2019; Bijlmakers, 2021). In particular, Bijlmakers and Boer focus on the promise of ubiquitin activating enzymes as novel drug targets – demonstrating that *Trypanosoma brucei*, a parasite closely related to *Leishmania*, has an E1 that is refractory to the inhibitor of the human E1 (Boer and Bijlmakers, 2019). As such, implying a difference in the drug binding pockets of the human and trypanosomal E1 - a boon for drug design which could, in theory, design a selective compound. However, they were unable to obtain a structure of the enzyme, relying on homology models to make predictions about the drug binding site, undercutting the use of their work for structure-based drug design. If these enzymes could be better characterised it could aid in expanding the flawed chemotherapeutic arsenal currently in use, ultimately helping people with leishmaniasis.

1.9 Summary and aims

To summarise, leishmaniasis is a pernicious disease, preying on the world's most vulnerable communities. Many of its casualties can not or will not complete treatment, as the current standard of care is often unaffordable, ineffective, or toxic (De Menezes *et al.*, 2015). As such, raising the standard of care, perhaps with novel drugs targeting newly characterised pathways, should be a priority.

One such pathway is ubiquitination, the post-translational modification of proteins, involved in almost every aspect of eukaryotic cell biology. In *Leishmania,* the enzymes involved in charging and attaching ubiquitin to substrates are still largely a mystery, at most identified by searching genomes for annotated domains and knocking out the identified genes (Burge *et al.*, 2020). Although this provides valuable indications of essentiality for multiple enzymes in the ubiquitination pathway, it leaves room to describe the enzymes at a protein level – when are they expressed, how abundant are they, with what do they interact, what are their structures, and, most importantly, are there feasible drug targets?

This works aims to address some of those questions. To identify enzymes, BLAST searches and hidden Markov models were employed. To confirm the existence of E1s at a protein level, and measure their abundance at major life cycle stages, activity-based probes and mass spectrometry were used. To identify the interactome of these E1s, chemical cross linkers were used with these activity-based probes, and CRISPR was used to embed a proximity-based labelling system to an E1 and E2. To examine the substrates of these enzymes immunoprecipitations against a ubiquitin motif were performed. To test the feasibility of drug targets amongst the discovered enzymes proteins were recombinantly expressed and drug assays developed. To assist in structure-based drug design a structure of a key enzyme was solved and a compound modelled into the drug binding pocket.

Chapter 2 Materials and methods

2.1 Materials

Unless specified otherwise, all reagent and chemicals were obtained from Sigma-Aldrich.

2.2 Methods

2.2.1 Bioinformatics

FASTA sequences of genes used as bait in searches were obtained from UniProt (https://www.uniprot.org/). BLASTs were conducted using the TriTrypDB BLAST tool (https://tritrypdb.org/tritrypdb/app, release 54). Multiple sequence alignments were performed with T-Coffee from the tools available from the EMBL-EBI website (https://www.ebi.ac.uk/Tools/msa/tcoffee/). Alignments were visualised using Jalview (https://www.jalview.org/, release 2.11.1.3). Alignments were submitted to jackhmmer (https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer, version 2.41.1) to provide hidden Markov models. Data was graphed in GraphPad Prism (https://www.graphpad.com/scientific-software/prism/).

2.2.2 Parasite Culture

The species of *Leishmania* used in this study was largely *L. mexicana*, other than in cyclophosphamide related experiments (detailed below). For experiments that did not employ CRISPR, the strain was M379, an isolate originally obtained from a vesper rat in Belize in 1962 (Lainson and Strangways-Dixon, 1964). For experiments that did use CRISPR a modified strain of M379, LmexCas9T7, was kindly donated by the Gluenz lab (Beneke *et al.*, 2017). For experiments analysing the role of ubiquitin in responding to cyclophosphamide induced DNA damage, a *Leishmania major* cell line, modified to feature a dimerizable Cre recombinase system was used (Santos *et al.*, 2017). This cell line was kindly donated by the McCulloch lab at the University of Glasgow. For storage, cryopreservation of mid-log phase promastigotes used 1:1 dilution of parasite culture with 70% heat-inactivated foetal bovine serum (Life Technologies) and 30% glycerol. Promastigotes were recovered and cultured in haemoflagellate minimal essential medium (GE Healthcare) with 10% foetal bovine serum, at 25°C, with routine passages undertaken every two to three days. Axenic

amastigotes were grown in Schneider's Drosophila Media (Gibco) adjusted to pH 5.5 with 20% foetal bovine serum and 3 ml of 2.5 mg/ml haemin in 50 mM NaOH, at 32°C with 5% CO_2, with routine passages conducted weekly. LmexCas9T7 were also cultured with 32 µg/ml hygromycin B phosphotransferase. Parasite growth was monitored using a haemocytometer, diluting cultures in phosphate buffered saline (PBS) and 25% glycerol. Both promastigote and axenic amastigote cultures underwent a maximum of 15 passages before a new cell line was retrieved.

2.2.3 THP1 culture

THP1 cells, a monocyte cell line isolated from a patient with acute monocytic leukaemia (Tsuchiya *et al.*, 1980), were cultured in RPMI-1640 media with 10% heat-inactivated foetal bovine serum (Life Technologies) at 37°C with 5% CO_2. Routine passages were conducted every 3 – 4 days, maintaining cell at a density of 1 – 3x10^5 cells/ml.

2.2.4 Cell Lysis

After pelleting and supernatant removal, cell pellets were washed twice in PBS and collected by centrifugation at 1000 xg, 10 min, 4°C. Cells were lysed through water bath sonication (10 mins, 30 secs on/30 secs off, maximum setting). Lysis buffers included:

- Laemmli buffer (2% SDS, 10% Glycerol, 62.5 mM Tris-Cl pH 6.8, 200 mM DTT)

- Radioimmunoprecipitation assay (RIPA) buffer (1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 7.5, 125 mM NaCl, 0.1 mM EDTA)

- CHAPS buffer (0.5% w/v CHAPS, 100 mM NaCl, 50 mM HEPES pH 7.5)

The buffer used depended on downstream experiments. Protease inhibitor cocktail was added to any lysis buffer and included: 2% proteoloc protease inhibitor (Expedeon), 100 µM PMSF, 1 µM E64, 5 µg/ml Pepstatin A, 400 µM 1,10-Phenanthroline, 1× Roche inhibitor cocktail, 1× PhosSTOP solution. Protein content of lysates were measured using the Bio-Rad Protein Assay, a method based on the Bradford assay.

2.2.5 Activity-based probe

7.5×10^7 mid log phase promastigotes per condition were centrifuged at 1000 × g for 10 mins at 4°C and the supernatant discarded. Cells were washed twice in PBS, and resuspended in CHAPS based lysis buffer. When the negative control was iodoacetamide treated then 15 mM iodoacetamide was added to the lysis buffer and incubated with cells on ice for 20 mins in the dark. 34.5 µg/ml E1 specific probe was then added to the lysis buffer. Cells were lysed as detailed in 2.2.4 Cell Lysis. Afterwards, lysates were incubated at 25°C for 15 mins. If a crosslinker was used 1 – 5 mM were added after this 15 minute incubation, and lysates were incubated at room temperature for a further 30 mins, before being quenched with 50 mM Tris pH 7.5. 25 µl neutravidin agarose beads (ThermoFisher) were added to each ~100 µl condition. Beads were incubated with samples for 1 hour at 4°C with rotation. Beads were collected by centrifugation at 2500 × g for 2 mins, then washed with 750 µl of radioimmunoprecipitation assay buffer. This was repeated 4 times. Samples, including agarose beads, were resuspended in Laemmli buffer and incubated at 90°C for 10 mins. Aliquots were then run on gel and silver stained (see 2.2.6 SDS-PAGE), and further aliquots submitted for mass spectrometric analysis (see 2.2.8 Mass spectrometry sample preparation).

2.2.6 SDS-PAGE

For polyacrylamide gel electrophoresis, protein samples were mixed with Laemmli buffer (if not previously used as a lysis buffer). Samples were heated to 90°C for 5 minutes. Typically, ~ 10 µg of protein were run on a 4-20% gradient polyacrylamide gel. Gels were supplied by either BioRad or Expedeon and run according to the manufacturer's instructions. After running gels were imaged with either silver stain, Coomassie blue, or fluorescence scanning. Silver staining was done with the Pierce™ Silver Stain Kit (Thermo Fisher) according to the manufacturer's instructions. Coomassie blue stain Expedeon's Instant Blue. Fluorescence scanning was conducted using a Typhoon laser scanner (Cytiva).

2.2.7 Western blots

Anti-myc and anti-His tag antibodies were gifted by Dr Helena De La Torre Olvera, and anti-ubiquitin were gifted by Dr Jeziel Damasceno, both of the

University of Glasgow. For western blots of promastigote whole cell lysate, the lysates of 1×10^7 cells were run per well. These were resolved on 4-20% acrylamide gel from Expedeon, then transferred to a PVDF membrane. Membranes were blocked with 5% skimmed milk powder in PBS-T (0.05% Tween 20) before incubation with primary antibodies. Secondary antibodies were either anti-mouse or anti-rabbit, conjugated to horseradish peroxidase (Invitrogen). Imaging relied upon the Pierce™ ECL Plus Substrate (Thermo Fisher) and chemiluminescence imager.

2.2.8 Mass spectrometry sample preparation

Samples were processed for LC-MS/MS analysis using Expedeon's filter aided sample preparation method. Protein samples in Laemmli buffer with DTT were heated at 60°C for 5 mins. Samples were then loaded into a 30 kDa molecular weight cut off filter (Microcon YM-30) and centrifuged at 14,000 × g for 15 mins. Washes and buffer exchanges use 200 µl 8 M urea in 0.1 M Tris-Hcl pH 8.5 and centrifuged as above. Proteins were alkylated with 0.05 M iodoacetamide in urea buffer and incubated for 20 mins in the dark. Iodoacetamide is removed through further centrifugation and two washes in urea buffer. Then three washes with 100 µl of 50 mM ammonium bicarbonate with centrifugation at 14,000 × g for 10 mins. Digestion uses trypsin at an enzyme to protein ratio of 1:100, with incubation for 24 hours at 37°C. After digestion, filter units are placed in new collection tubes, where peptides are collected by resuspension in 40 µl of 50 mM ammonium bicarbonate and centrifuged at 14,000 × g for 10 mins. After repeating this step once more, peptides still bound to the filter were retrieved using 50 µl of 10% acetonitrile and centrifugation at 14,000 × g for 10 mins. Using a vacuum centrifuge, peptides were then dried down in the wells of 96 well plate ready for LC-MS/MS analysis.

2.2.9 TMT Labelling

Relative quantitation was achieved with the TMT Mass Tagging kit from Thermo Scientific, as per the manufacturer's instructions.

2.2.10 Peptide/protein identification by nLC-ESI-MS/MS

After filter aided sample preparation, tryptic peptides were analysed with nanoflow reversed-phase HPLC coupled to electrospray tandem mass spectrometry. 0.05% formic acid was used to solubilise peptides, before fractionization with a nanoflow uHPLC system (Thermo RSLCnano), then online analysis by electrospray ionisation mass spectrometry on an Orbitrap Elite (Thermo Scientific).

Before analysis on the Orbitrap Elite, peptides were desalted and concentrated for 4 mins on a trap column, then transferred to the analytical column under starting solvent conditions (4% solvent B). The water:acetonitrile gradient was: 4 – 40% v/v solvent B from 12 to 102 mins, 40 – 100% v/v solvent B from 102.1 to 116 mins, then held at 100% solvent B from 116 to 121 mins, before re-equilibration at starting conditions (4% solvent B) for a total time of 125 mins. Solvent A: water and 0.1% formic acid. Solvent B: 80% acetonitrile and 0.08% formic acid. The analytical column had a fixed flow rate of 0.3 µl/min. The Orbitrap Elite has a precursor scan at 60,000 resolving power (for a mass range of m/z 400 - 2000), with collision-induced dissociation fragmentation and detection of the top 20 precursors in the linear ion trap. Singly charged ions were excluded from selection, whereas selected precursors were added to a dynamic exclusion list for 180 secs.

Analysis of TMT labelled samples differed as described below. If not otherwise stated conditions were the same as for unlabelled samples. Starting solvent conditions were 5% solvent B. The water:acetonitrile gradient used was: 5 – 45% v/v solvent B from 4 to 154 mins, 45% - 100% v/v solvent B from 154 to 154.1 mins, then held at 100% v/v solvent B from 154.1 to 160 mins, before re-equilibration to starting conditions (5% solvent B) for a total time of 165 mins. The m/z mass range of the precursor scan was 380 – 1800. The top 3 precursor ions were selected from the linear ion trap. These were also subjected to higher-energy C-trap dissociation before detection in the Orbitrap.

2.2.11 Protein identification with Mascot

Mass spectra were searched using the Mascot search engine, with protein identities obtained using the LmexicanaMHOMGT2001U1103 database from TriTrypDB.org. The search criteria were as follows: Enzyme, Trypsin; Maximum missed cleavages, 1; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide mass tolerance, 0.4Da; Fragment mass tolerance, 0.4 Da. Significance threshold set to 0.05. For studies of the ubiquitinated peptides, the variable modification of GG (K) was also added.

Identifications over the 0.05 significance threshold and a protein score ≥30 were included in the protein list. The protein score factors the combined ion scores of each peptide mass spectra matched to a protein. In turn, the ion scores are based on the probability that an observed match between the experimental data and the database sequence is a random occurrence, using a 95% confidence threshold. Peptide matches were not used for protein identification if their *p*-value was greater than the significance threshold used for a false discovery rate of 1%.

2.2.12 Proteomics figures

Data was also analysed using Proteome Discoverer™ (Thermo Fisher Scientific) using Mascot, Sequest HT, and Amanda search engines, with the same search criteria as described in 2.2.11 Protein identification with Mascot. Figures were designed using Proteome Discoverer.

2.2.13 Amplification of CRISPR reagents

The pPLOT plasmids required for tagging genes of interest with standard, miniTurbo and Turbo BirA* were kindly provided by Dr Jeziel Damasceno of the University of Glasgow. *E. coli* containing these plasmids were cultured from glycerol stocks using 30 ml LB media with 100 µg/ml ampicillin, incubated at 37˚C overnight with shaking. Then cultures were pelleted by centrifugation at 3,300 × g for 5 mins, before the NucleoSpin Plasmid kit (Macherey-Nagel) was used according to the manufacturer's instructions to purify pPLOT plasmids from *E. coli*. Plasmid DNA was quantified using a NanoDrop spectrophotometer

(ND1000, Thermo Scientific) to measure the absorbance of the elution at 260 nm. Plasmids were stored at -80˚C.

Primers to amplify the sgRNA and template DNA were designed using http://leishgedit.net/. The primers used to tag each terminus of each gene of interest are shown in Table 2-1. To amplify the repair cassettes, using the primers in Table 2-1, the following standard 80 µl PCR master mix was used: 30 ng of pPLOT plasmid, 1 µl 10 mM dNTPs, 16 µl 10 µM forward primer, 16 µl 10 µM reverse primer, 2.4 µl DMSO, 6 µl 25 mM $MgCl_2$, 8 µl 10× HiFi buffer (supplemented with 15 mM $MgCl_2$, to a final $MgCl_2$ concentration of 3.375 mM), 4 µl High-Fidelity DNA polymerase, and water to a total volume of 80 µl. The thermocycler used was AlphaCycler (AC196, PCRMax), with the program set for 5 mins at 94˚C, then 40 cycles of the following: 30 secs at 94˚C, 30 secs at 65˚C, 2 mins 15 secs at 72˚C. A final elongation step of 72˚C for 7.5 mins was used to finish the cycle.

To amplify the sgRNA, using the primers in Table 2-1, the following standard 50 µl PCR master mix was used: 1 µl 10 mM dNTPs, 10 µl 10 µM G00 oligonucleotide, 10 µl 10 µM of the either the N- or C-terminal targeting sgRNA primers, 5 µl 10× HiFi buffer (supplemented with 15 mM $MgCl_2$, to a final $MgCl_2$ concentration of 3.375 mM), 2.5 µl High-Fidelity DNA polymerase, and water to a total volume of 50 µl. The thermocycler used was AlphaCycler (AC196, PCRMax), with the program set for 30 secs at 98˚C, then 35 cycles of the following: 30 secs at 98˚C, 30 secs at 60˚C, 15 secs at 72˚C. A final elongation step of 72˚C for 7.5 mins was used to finish the cycle.

Target gene	Terminal targeted	Primer	Sequence
LmxM.34.3060 (putative E1)	N-	Fwd	CAGACACGAAAACATAGCTTCAAACTACCGGTATAATGCAGACCTGCTGC
		Rev	GTCGATCAACACTTCCTTCTCAGCCTCCATACTACCCGATCCTGATCCAG
		sgRNA	GAAATTAATACGACTCACTATAGGTTGTGCTTTGCGCTGTGTTTGTTTTAGAGCTAGAAATAGC
	C-	Fwd	ACGGCCACTGTGTGCCTTGAAATTCAGCAGGGTTCTGGTAGTGGTTCCGG
		Rev	CCGGGTGGGGTGGGCATGGCCGGGGGGCGTCCAATTTGAGAGACCTGTGC
		sgRNA	GAAATTAATACGACTCACTATAGGGCAGCCAGCAGCTCCGGATTGTTTTAGAGCTAGAAATAGC
LmxM.33.0900 (putative E2)	N-	Fwd	ACCAAAGAGTTCCAGTCCAGACCTTCCCCAGTATAATGCAGACCTGCTGC
		Rev	TTTCAGGCATTGCGCGGTAGGGGCCGTCATACTACC CGATCCTGATCCAG
		sgRNA	GAAATTAATACGACTCACTATAGGACAGGTGGGAAGGCGAGTTAGTTTTAGAGCTAGAAATAGC
	C-	Fwd	ATAGGGCTGGGGCTCCTGCGGCAGCTGATGGGTTCTGGTAGTGGTTCCGG

		Rev	CTTCGTAGTTTATCAGAGGGAGGTTCGCCGCCAATTTGAGAGACCTGTGC
		sgRNA	GAAATTAATACGACTCACTATAGGAGTGAACACAGATCCCTAGGGTTTTAGAGCTAGAAATAGC
G00			AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

Table 2-1 Primers used to tag the genes of interest with BirA*. Fwd: Forward primer; Rev: Reverse primer. G00 is an oligonucleotide used as template DNA for the amplification of all sgRNAs used in this study.

2.2.14 Agarose gel electrophoresis

PCR products from amplification of repair cassettes were run on 1% (w/v) UltraPure Agarose (BioRad) gels formed using 1× UltraPure TAE buffer (BioRad) with 1× SYBR Safe DNA Gel Stain (ThermoFisher). 2 µl of PCR products were mixed with 1× gel loading dye (ThermoFisher), running alongside 6 µl 1 kb DNA ladder (New England Biolabs). Gels ran for 30 – 60 mins in 1× TAE buffer at 95 v using the Mini-Sub Cell GT Electrophoresis System (Bio Rad). Gels were visualised using a gel documentation system. For sgRNA PCR products, the same method was adopted. However, agarose gels used 2% (w/v) UltraPure Agarose (BioRad). After confirmation of PCR success, the remainder of the PCR products were heat sterilised by incubation at 94°C for 5 mins.

2.2.15 Transfections

1×10^7 of mid log phase LmexCas9T7 were used per transfection, collected through centrifugation at 800 × g for 5 mins. Pelleted cells were washed in 1 ml 1× transfection buffer (3× transfection buffer: 200mM Na_2HPO_4, 70mM NaH_2PO_4, 15mM KCl, 150mM HEPES pH 7.3, 1.5 mM $CaCl_2$) before resuspension in 200 µl transfection buffer. 20 µl of sgRNA PCR products and 40 µl of repair cassette PCR products were added to the cell resuspension. 250 µl of this mixture was then transferred to a Gene Pulser electroporation cuvette, and the transfection conducted using one pulse of the X-001 program in the Amaxa Nucleofector II (Lonza). Transfected cells were immediately transferred to 5 ml of warmed haemoflagellate minimal essential medium with 10% foetal bovine serum. Transfectants were incubated at 27°C for 16 hours, before selective antibiotics were added (for standard BirA* tagged transfectants: 20 µg/ml puromycin; for miniTurbo BirA* tagged transfectants: 5 µg/ml blasticidin). 10 to 14 days later drug resistant populations emerged. Clonal cell lines were generated of all the miniTurbo transfectants by serial dilution in 96 well plates. Diluting for an average of 0.3 cells per well obtained putatively clonal colonies, which were

moved into the 1 ml wells of 24 well plates after 2 weeks, then placed into 10 ml flasks, before stabilates were formed.

2.2.16 Diagnostic PCR

To confirm the insertion of the BirA* gene at the correct locus, genomic DNA was extracted from 1×10^8 cells after pelleting by centrifugation at 600 × g for 5 mins. Transfectants then had their DNA extracted using the DNeasy Blood and Tissue Kit (Qiagen), as per the manufacturer's instructions. DNA elutions were quantified using a NanoDrop spectrophotometer. 100 ng of extracted DNA was then added to a master mix containing the following: 10 µl of 10 µM of a forward primer, 10 µl of 10 µM of a reverse primer, 1 µl of 10mM dNTPs, 5 µl High-Fidelity DNA Polymerase, 5 µl 10× HiFi buffer supplemented with 15 mM $MgCl_2$, and water to a total volume of 50 µl. The primers used for diagnostic PCRs are shown in Table 2-2. PCR reactions used an AlphaCycler (PCRMax) with the following program: 5 mins at 94°C, then 35 cycles of 30 secs at 94°C, 5 secs at 60°C, and 2.5 mins at 72°C. Then a final elongation step of 7 mins at 72°C. 5 µl of PCR products were mixed with 1× Gel Loading Dye and run on a 1% agarose gel (as detailed in 2.2.14 Agarose gel electrophoresis).

Anneals to	Position	Sequence
BirA*	5'	GAGCCAACCTGTACCTGTCC
	3'	ACGCTCTCCTCCACTCTTCT
LmxM.34.3060 (putative E1)	5'	GGTGATGCGTGTCCTTGATG
	Mid-gene	GTAGAGGTGCTCGGTGATTG
	3'	GACTTCGCTCTCCTTATCCC
LmxM.33.0900 (putative E2)	5'	CGCCTCTACTCATCGTCTC
	Mid-gene	CAATGTAAAGCGGCTCCAC
	3'	CTCTCTTTGTGGGGTGACG

Table 2-2 Primers used to confirm the presence of the BirA* tag. 5': primer anneals near the 5' end of the gene; 3': primer anneals near the 3' end of the gene; mid-gene: primer anneals at a centre-point between the 5' and 3' primers.

2.2.17 BioID

100 ml of BirA* tagged promastigote culture at a density of 2×10^6 cells/ml had 150 µM biotin added. Cultures were left for 18 to 24 hours. Parasites were then washed twice in 10 ml PBS (centrifuged at 1000 × g for 10 mins). 4×10^8 parasites were crosslinked using 1 mM DSP in 10 ml PBS for 10 mins at 26°C. The reaction

was quenched with 200 µl 1 M Tris pH 7.5 incubated for 5 mins at room temperature. 4×10^8 parasites were lysed in 500 µl of RIPA buffer, as described in 2.2.4 Cell lysis. 1 µl of BaseMuncher (Expedeon), an endonuclease, was added to lysates and left on ice for 1 hour. Lysates were spun at 10,000 × g for 10 mins at 4°C, and the supernatants kept and quantitated with a BCA assay (ThermoFisher). Streptavidin agarose beads were incubated with the lysates overnight at 4°C with rotation. Agarose beads were then washed with the following wash buffers, each incubated with the beads for 5 mins: RIPA buffer, 4 M urea, 6 M urea, 0.1 M Na_2CO_3, 1 M KCl, 50 mM TEAB. Beads were then boiled at 90°C for 10 mins in Laemmli buffer, before being trypsin digested as detailed in 2.2.8 Mass spectrometry sample preparation.

2.2.18 Conjugating anti-diglycine antibodies to magnetic beads

Reagents were from the Pierce Crosslink Magnetic IP/Co-IP kit, with slight modifications to the manufacturer's instructions as detailed here. 13 ml of 1x modified coupling buffer was prepared through mixing 0.65 ml 20× coupling buffer, 0.65 ml IP lysis/wash buffer, and 11.7 ml water. 75 µl of bead solution was added to two microfuge tubes, then beads were collected using a magnetic stand and their storage solution discarded. 1.4 ml of modified coupling buffer was added to each tube and incubated for 1 min at room temperature with rotation. This step was repeated twice. 300 µl of 1× coupling buffer was formed with 489.4 µl of water, 30.8 µl of 20× coupling buffer, 30.8 µl IP lysis/wash buffer, 64 µl (32 µg) of anti-diglycine antibodies. 300 µl of this coupling buffer was added to each microfuge tube, and incubated with rotation for 20 mins at room temperature. Beads were then collected with a magnetic stand, the supernatant removed. 300 µl of modified coupling buffer was added and beads vortexed. Beads were then collected with a magnetic stand, the supernatant removed. 900 µl of modified coupling buffer was added and beads vortexed, before being collected on a magnetic stand and supernatant disposed. This step was repeated two more times. 1 ml of water was added to the beads, beads were vortexed and supernatant disposed. This step was repeated once more. If crosslinking was conducted, it was as follows. 217 µl of DMSO was added to a single tube of DSS (tubes included in the kit) to make a 25 mM solution. This solution is diluted 1:100 in DMSO (0.25 mM DSS). A crosslinking solution is then

formed with 24 µl 0.25 mM DSS, 15 µl 20× coupling buffer, 261 µl water. Antibody-bead conjugates are then incubated with the crosslinking solution for 30 mins at room temperature with rotation. Beads are then collected with a magnetic stand and supernatant discarded. 200 µl of elution buffer was incubated with the beads for 3 mins at room temperature with rotation, before beads were collected with a magnetic stand and supernatant discarded. This step was repeated once more. 900 µl of modified coupling buffer and 100 µl of neutralization buffer were mixed with beads, and removed through use of the magnetic stand. 600 µl of IP lysis/wash buffer was mixed with the beads and removed using the magnetic stand. Beads were stored in water at 4˚C until use.

2.2.19 Anti-diglycine immunoprecipitations

For promastigotes, 7.5×10^7 cells were treated with 50 µM MG132, or an equivalent volume of DMSO, and left overnight. Cells were then spun down at 1000 x g, 10 min, 4˚C, and washed twice in 2.5 ml PBS. Cells were then resuspended in 100 µl CHAPS lysis buffer and lysed as detailed in 2.2.4 Cell lysis. Lysates were centrifuged at the maximum speed of a benchtop centrifuge for 15 mins at 4˚C. The supernatant was kept and incubated for 1 hour with 10 mM DTT, then incubated for 20 mins with 15 mM iodoacetamide in the dark, both at room temperature. Protein content was extracted by acetone precipitation: 4 sample volumes of acetone at -20˚C added to protein solution, incubated at -20˚C for at least 1 hour, centrifuged 10 mins at 13,000 x g, pellet washed twice with 4 sample volumes of -20˚C 80% acetone, centrifuged as before and supernatant removed, pellets left to air dry. Pellet resuspended in 500 µl 6 M urea and 50 mM HEPES pH 7.5 and vortexed. Bradford assay was run to determine protein concentration. Urea solution was diluted to 1 M, trypsin added at a 1:50 enzyme:protein ratio, then incubated overnight at 37˚C. Trypsinisation was stopped with the addition of 2% (v/v) formic acid (pH 2). Solid phase extraction was conducted with standard density C18 cartridges (Empore). Before loading the tryptic peptide solution, the solid phase extraction cartridges were prepared by loading 1 ml of methanol and centrifuged at 1425 rpm for 1 min, then 0.5 ml of 0.1% trifluoroacetic acid and 70% acetonitrile and centrifuging as before, then loading 0.5 ml of 0.1% trifluoroacetic acid and centrifuging as before. Afterwards, the tryptic peptides were loaded into the cartridge, and centrifuged for 800 rpm for 10 mins. The cartridge was

transferred to a new collection tube, 0.5 ml of 70% acetonitrile was added, and centrifuged at 800 rpm for 10 mins. The eluate contains the desalted peptides. Peptides were then lyophilised in a vacuum centrifuge with heating, before being resuspended in 500 µl of the IP/wash buffer of Pierce Crosslink Magnetic IP/Co-IP Kit (ThermoFisher) and centrifuged at max speed for 5 mins at 4˚C in a benchtop centrifuge to remove insoluble material. Immunoprecipitations used 15 µg of anti-diglycine antibodies (Merck) conjugated to magnetic beads per immunoprecipitation. Incubation of peptides with beads occurred for 1 hour at room temperature with rotation and vortexing every 15 mins. A magnetic stand was used to collect beads, the supernatant removed and beads washed twice in 500 µl of IP/wash buffer. Beads were then washed with 500 µl ddH_2O. Peptides were then eluted using 100 µl of the elution buffer and incubating for 5 mins at room temperature with rotation, the low pH of the elution buffer was then neutralized with the neutralization buffer. The elution and neutralization was then repeated once more, and peptides quantified using the Pierce Quantitative Colorimetric Peptide Assay (ThermoFisher). Peptides were then submitted to mass spectrometry as stated in 2.2.10 Peptide/protein identification by nLC-ESI-MS/MS.

When anti-digylcine immunoprecipitations were performed with amastigotes the following alterations to the method were made:

- MG132 incubation was for 6 hours as opposed to overnight.
- Cells were lysed in radioimmunoprecipitation assay buffer as opposed to CHAPS.
- $8x10^8$ cells were used per condition.
- Anti-diglycine antibodies were chemically crosslinked to beads.
- Immunoprecipitations were performed in series, using 64 µg of antibodies conjugated to beads per immunoprecipitation.
- Peptides were TMT labelled (see 2.2.9 TMT Labelling).

2.2.20 Transforming *E. coli*

Gene synthesis and subcloning into pET15B (the plasmid used for expression of the proteins of interest) was performed by DC Biosciences. Subcloning was performed through restriction enzyme digests. Competent Rosetta *E. coli*, BL21 DE3 strain (Merck) were thawed on wet ice, and 1 – 10 ng of plasmid was added, then cells and plasmids were incubated on ice for 5 mins. Solutions were heated for exactly 30 secs in a 42°C water bath, then incubated on ice for 2 mins. 250 µl room temperature super optimal broth with catabolite repression media was added. Cultures were left at 37°C with shaking at 250 rpm for 60 mins. Cells were resuspended by finger flicking, and 50 µl placed on agar plates with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. Cells were equally distributed around the plate and left for 15 mins at room temperature. Plates were then inverted and placed in a 37°C incubator and left overnight.

2.2.21 Protein expression of LmxM.34.3060, LmxM.33.0900, and LmxM.02.0390 for functional assays

Cultures were grown in LB at 37°C until the optical density at 600 nm was 0.6. Then 1 mM IPTG was added and cultures were incubated overnight at 18°C with shaking. Cells were pelleted by centrifugation at 5000 × g for 15 mins at 4°C, then lysed as indicated in 2.2.4 Cell lysis. Lysates were clarified through centrifugation at 20,000 × g for 30 mins at 4°C. The insoluble fraction was the pellet formed after centrifugation, which was resuspended in Laemmli buffer and DTT for examination on SDS-PAGE. Proteins of interest were purified using a 5 ml Histrap column (General Electric), then size exclusion chromatography, both on an AKTA start. Refolding used a 5 ml Histrap column, equilibrated with 10 column volumes of resolubilisation buffer (50 mM Sodium Phosphate pH 8, 100 mM NaCl, 8 M Urea, 20 mM Imidazole, 10 mM 2-Mercaptoethanol). Sample was loaded onto the column at a 1 ml/min flow rate. Washing buffer was 50 mM Na-Phosphate pH 8, 100 mM NaCl, 3 M Urea, 10% Glycerol, 20 mM Imidazole, 10 mM 2-Mercaptoethanol. Elution was done step-wise, 1st with a gradient of 10% elution buffer for 6 column volumes, followed by 100 % elution buffer for 8 column volumes (elution buffer: 50 mM Na-Phosphate pH 8, 100 mM NaCl, 10% Glycerol, 500 mM Imidazole, 10 mM 2-Mercaptoethanol).

2.2.22 Functional assays

5 pMol of E1 was mixed with 33 pMol of E2, 100 µM of ATP, 200 µM of $MgCl_2$, 50 mM HEPES, 116 µM ubiquitin, and the relevant drug concentration, with the volume made to 20 µl with water. The reaction was started by adding E1. The reaction was incubated at 30°C for 30 mins and halted with the addition of 4× Laemmli buffer. Reactions were then run on gel (see 2.2.6 SDS-PAGE) and Coomassie stained using InstantBlue (Expedeon).

2.2.23 Protein expression of LmxM.34.3060 for crystallography

Cultures were grown in LB at 37°C until the optical density at 600 nm was 0.6. Then 1 mM IPTG was added and cultures were incubated overnight at 18°C with shaking. Cells were collected through centrifugation at 4000 rpm, 15 mins, at 4°C and lysed using a microfluidizer. Lysates of cells expressing GST-ubiquitin were mixed with lysates of cells expressing LmxM.34.3060 with 2.5 mM ATP and 2.5 mM $MgCl_2$. These were incubated for 2 hours at 4°C with stirring. Lysate mixture was then placed on a column packed with glutathione beads and incubated for 1 hour with stirring. These were washed with at least 12 column volumes of wash buffer (25 mM Tris 7.6, 200 mM NaCl, 10 mM Imidazole, 5 mM 2-Mercaptoethanol), until Bradford reagent did not change colour when exposed to flow-through. Elution of LmxM.34.3060 from GST-ubiquitin used the wash buffer with 25 mM DTT. Elution of GST-ubiquitin used the GST elution buffer (25 mM Tris 7.6, 150 mM NaCl, 5 mM 2-Mercaptoethanol, 10 mM glutathione). LmxM.34.3060 elution was incubated with HRV3C restriction enzyme overnight at 4°C in snakeskin dialysis tubing with 10 kDa molecular weight cut off (ThermoFisher), dialysis buffer: 25 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT. The dialysis products were then subject to strong anion exchange using a SOURCE Q column (buffer A: 25 mM Tris pH 7.6, 1 mM DTT; buffer B: 1 M NaCl, 25 mM Tris pH 7.6, 1 mM DTT). Fractions containing protein were then subject to size exclusion chromatography (buffer: 25 mM Tris pH 7.6, 150 mM NaCl, 1 mM DTT). Fractions containing protein were concentrated using spin columns, and flash frozen before storage at -80°C.

2.2.24 Protein crystals

The reaction mix that was incubated in crystallisation screening trays consisted of 10 mg/ml LmxM.34.3060, 2.5 mM ATP, 2.5 mM $MgCl_2$, and 50 mM HEPES. The 3.1 Å structure presented here was found in well E6 of a Morpheus screening tray (Ligands: 30 mM Diethylene Glycol, 30 mM Triethylene glycol, 30 mM Tetraethylene glycol, 30 mM Pentaethylene glycol. 100 mM sodium HEPES and MOPS pH 7.5. 20% (v/v) Ethylene glycol and 10% (w/v) PEG 8000. The 96 well tray was seeded with crystal fragments from a prior experiment (using the same conditions) at a 1:100 dilution using a Mosquito Crystal liquid handling robot. The tray was incubated at 6°C until crystals formed. Crystals were shot at the Diamond Light Source, Oxford, and the structure solved with the generous assistance of Mads Gabrielson of the Beatson Institute for Cancer Research. The structure was determined by molecular replacement, using the structure of 5TR4 (PDB file), and the program MORDA.

2.2.25 SAXS

Purified LmxM.34.3060 (see 2.2.23 Protein expression of LmxM.34.3060 for crystallography) in 25 mM Tris pH 7.6, 150 mM NaCl, 1 mM DTT was sent to the Diamond Light Source and data collected by Mads Gabrielson. Data was analysed with generous assistance from Mads Gabrielson using the ATSAS software package. Basic data analysis was conducted using ScÅtter IV, *ab initio* modelling conducted using DAMMIN, hypothetical scattering of the crystal structure computed by CRYSOL.

Chapter 3 Bioinformatic and proteomic identification of ubiquitination enzymes

3.1 Introduction

Before *Leishmania mexicana*'s ubiquitination machinery can be targeted in any drug discovery efforts, the machinery must be identified. This chapter will focus on bioinformatic and proteomic identification of those enzymes. For bioinformatics, the methods employed include BLAST searches and iterative searches using Hidden Markov Models. Bait sequences for bioinformatics were a representative selection of well-characterised ubiquitination enzymes, predominantly from humans. A putative SUMOylation E1 and E2 was also identified, as well as components of NEDDylation, urmylation, and ufmylation E1s. For proteomics, activity-based probes were used, targeting either the ubiquitin activating enzymes, all catalytically active ubiquitination enzymes (E1s, E2s, and HECT E3s), or SUMOylating enzymes. The success of these probes was judged through SDS-PAGE, affinity purification of tagged proteins, and mass spectrometry.

3.1.1 Bioinformatics

An early, transformative contribution to drug discovery from the then-nascent field of bioinformatics was the recognition of sequence homology between the simian sarcoma virus onc gene *v-sis* and a platelet-derived growth factor by simple string matching (Doolittle *et al.*, 1983). This platelet-derived growth factor became both a drug target and an influential lesson (demonstrating that any factor inducing aberrant gene expression patterns of growth factors can contribute to cancer) (Xia, 2017). This concept of searching for similarity has gone on to become a cornerstone technique for computational biologists, and has even found widespread use amongst general biologists, in a tool known as BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990). However, the similarity between proteins is more taxing to determine, as opposed to similarity between genes, as amino acid substitutions are affected by their chemical properties and frequency of occurrence (Pertsemlidis and Fondon, 2001). Using high-confidence alignments of many homologous proteins, then calculating the frequencies of substitutions, matrices that state the likelihood of various substitutions have been formed (States, Gish and Altschul, 1991). Typically, a BLAST operates in three steps (Pertsemlidis and Fondon, 2001):

1. Remove the statistically significant yet biologically uninteresting low complexity regions (such as CA repeats), then generate a list of all the short sequences that comprise the query. For protein queries, a short sequence is three amino acids. A scoring matrix is then used to determine high scoring matches (where the three amino acid short sequences are re-matched to the original query sequence) above a given threshold, termed the neighbourhood word score threshold. Under standard parameters for this threshold, every added residue in a query sequence results in approximately 50 more short sequences.

2. A target database is searched for exact matches to the list of short sequences. If a match occurs, it is used to seed potential alignments between the query and the database's target sequence.

3. If the score increases, the alignment is extended in both directions using the list of short sequences. Then the score is compared to a cut-off score, which is empirically determined through comparison of random sequences, and if the alignment score is significantly great it is reported. The statistical significance of a hit is determined by the probability that random sequences could produce the same score.

Despite the justified popularity of BLAST searches, a more sensitive method revolves around Hidden Markov Models (HMMs) (Johnson, Eddy and Portugaly, 2010). HMMs describe the evolution of observable events that depend on imperceptible factors. They achieve this through linking probability distributions of those observed events to the underlying unknowns (Yoon, 2009). What does this mean in a biological context? A protein's amino acid sequence counts as an observable event. This observable event has statistical properties from which domains, an unknown, can be inferred. So does a gene encode a protein of interest? It may if the encoded protein has the same statistical properties as characterised proteins (Yoon, 2009). So one method to identify proteins of interest in a novel organism would be to align characterised proteins from diverse model organisms, forming a general profile of a protein of interest, then submitting this profile to HMM searches. In fact, high-confidence HMM results could then be integrated into the original alignment, making a more general

query sequence, which is then resubmitted to find more divergent proteins in more species (Johnson, Eddy and Portugaly, 2010).

Proteins identified through bioinformatic screens could be checked using domain search engines, such as SMART (Letunic, Khedkar and Bork, 2021), which allows visualisation of the domain architecture of a protein, confirming the existence of the expected domains. Additionally, identified proteins could be further analysed by searching for their orthologs, using a tool such as OrthoMCL (Li, Stoeckert and Roos, 2003). Still, these tools demand a sequenced genome for an organism of interest. Thankfully, a functional genomics database for kinetoplastid parasites exists - TriTrypDB (Aslett *et al.*, 2009). Having such a database permits the searching, annotation, and analysis of genomes.

3.1.2 Proteomics

Of course, a bioinformatic screening campaign tells you what genes exist, but it does not tell you when their proteins are present, in what abundance, with which post-translational modifications, to name only a few factors worthy of investigation. To interrogate biology at the protein level, an appropriate tool would be the aptly named proteomics. The liberally used "-omics" suffix refers to studying a system in its totality, in this case its proteins, with a view to examining their abundances, interactions, locations, and/or modifications. How can any instrument or experiment measure something so nebulous and multifaceted as a protein network? Key to multiple approaches is a mass spectrometer (MS). Key to the optimal functioning of a mass spectrometer is sample preparation and fractionation (Matallana-Surget, Leroy and Wattiez, 2010). For bottom-up proteomics (the most widely used proteomics), proper sample preparation goes some way to solving the enduring issue of undetected minor or underrepresented proteins in the sample. "Proper" sample preparation can include separation of the proteome by protein mass and/or isoelectric point, as well as depletion and enrichment methodologies (such as use of an activity-based probe) (Matallana-Surget, Leroy and Wattiez, 2010). The end goal of sample preparation is to reduce the concealment of the spectral features of those low abundance proteins, which can be dominated by high abundance proteins. To aid in protein identification, proteins are cleaved, often by trypsin, into peptides (a surprising point on the most fundamental of biological

nomenclature, it is curiously ill-defined as to when a polypeptide qualifies as a protein). Peptides do not possess the complex structural aspects of proteins, are more chemically amenable, and their smaller size makes them more agreeable to MS analysis. Trypsin earned its place as the most commonly used cleavage agent as it forms small and predictable peptides, due to specifically cutting at the carboxyl side of lysine and arginine residues (except when they are followed by proline) (Karpievitch *et al.*, 2010).

Given that disparate peptides can have similar or identical molecular weights, and so produce a single intense peak in the initial MS spectrum, it is important to stagger their arrival to the MS. Frequently this is achieved using high-performance liquid chromatography (HPLC). This system is comprised of a column packed with hydrophobic beads (the stationary phase), a solvent (the mobile phase), a pump to move to mobile phase, and a detector that measures retention time (Karpievitch *et al.*, 2010). The sample is diluted in the mobile phase and pushed through the stationary phase, where peptides bind to the beads at a strength proportional to the peptides' hydrophobicity. This is to say that hydrophilic peptides will elute sooner than hydrophobic peptides, and so enter the initial MS run separately, increasing peak resolution (Karpievitch *et al.*, 2010).

Now the peptides are separated based on their hydrophobicity, how are they introduced into the mass spectrometer? Electrospray ionisation forces the sample through a spray needle or capillary tube where a high voltage (e.g. 3 – 6 kV) is applied and the sample travels through a warm, inert gas (perhaps nitrogen) (Awad, Khamis and El-Aneed, 2015). This induces solvent evaporation as the repulsive forces of the ions overcome the surface tension of the droplet, causing the ions to desorb into the gas phase (El-Aneed, Cohen and Banoub, 2009).

With ions in the correct phase, it is now appropriate to measure their charge to mass ratio. One such mass analyser is a quadrupole analyser, first described by the Nobel Prize winning physicist Wolfgang Paul in the 1950's (Paul and Steinwedel, 1953). Wolfgang's analyser contained four parallel electrical rods, with two supplied with direct current, and two linked to an alternating radio frequency. An ion enters the quadrupole analyser and is attracted to the rod

possessing its opposite charge. However, the polarity of the rods is then changed, which changes the trajectory of the ion. This happens repeatedly, causing the ions to attempt a complex path through the analyser. Only ions within a narrow range of mass to charge ratios successfully make it to the detector at the other end of the analyser, ions outside of this range collide with the rods and are lost. By changing the magnitude of the currents and radio frequencies a variety of ions can be transmitted to the detector (El-Aneed, Cohen and Banoub, 2009).

However, no matter how sensitive a detector or how long a gradient used on a chromatography system, if a protein mixture is too complex low abundance proteins of interest may go unidentified. Besides, if only a subset of proteins are of interest, why submit the entire proteome for a time consuming MS run and arduous data analysis? This need to isolate proteins of interest from complex biosystems, for labelling and identification, has lead to the invention of activity-based probes. An activity-based probe is a small molecule probe that typically includes a reactive group (or warhead) that targets specific amino acids on target proteins, a reporter group that enables detection or purification of labelled targets, and a linker that minimises undesirable interactions between the warhead and reporter group (Benns *et al.*, 2021). A relevant example of an activity-based probe is ubiquitin with a dehydroalanine warhead – if the C-terminal glycine of ubiquitin is replaced by dehydroalanine it is still recognised by E1s, E2s, and E3s but, upon activation by an E1, the electrophilic nature of the dehydroalanine primes it to covalently react with the cysteine active site of a ubiquitination enzyme (Mulder *et al.*, 2016). The probe may react with the cysteine active site of the E1 (forming a thioether-linked adduct), or it may progress along the cascade (forming the native thioester bonds). If the probe progresses along the cascade, there is a chance it will react with an E2 or E3 (notably, only HECT or RBR E3s, as those are the only E3 families that possess a catalytic cysteine). The reporter group used depends on subsequent experiments, so if purification is desired then a good reporter group would be biotin, enabling biotin-streptavidin pulldowns. This enables a network of ubiquitination machinery to be separated from the cellular milleu, integral for its study.

Science has progressed far thanks to reductionism, so why not focus on any one individual protein? Because any one individual protein does not work in isolation, it works as part of an information processing network – for example, in *E. coli*, glucose concentration can be considered a signal, a signal that affects transcription factors, transcription factors that affect protein expression, proteins that affect which sugar is metabolised (Davies, 2019). Perhaps the need to consider a system as a whole is best exemplified by a broken transistor radio (Lazebnik, 2002). If a radio is broken, does one fix it by removing parts one by one and studying what sound, if any, is produced? Would it be fixed by examining the insides of transistors and capacitors to the atomic level? These projects seem expensive, time consuming, and unlikely to repair it. Alternatively, any individual moderately trained in the principles of electronic circuitry, knowing how current flows from one component to another, converting electromagnetic waves into sound waves, would be infinitely better equipped to solve the problem. This is the case for omics, knowing how a system functions as a whole provides insights into its emergent properties. If reductionism is the only tool in the biologist's toolbox then we have reduced too much.

3.2 Aims and Hypotheses

Below is a study that identifies ubiquitination E1s, E2s, and HECT E3s through a bioinformatic screening campaign, as well as the E1 and E2 involved in SUMOylation. To obtain information at the protein level, these putative identifications are confirmed through activity-based probes and mass spectrometry.

- It is hypothesised that *Leishmania mexicana* ubiquitin and ubiquitin-like machinery will be sufficiently related to well-characterised orthologs that bioinformatic techniques will identify the parasite orthologs.

- It is hypothesised that expression at the protein level will be detectable through activity-based probes and mass-spectrometry.

3.3 Results

3.3.1 Bioinformatic screening

Using the full length of E1 encoding genes of well characterised organisms as bait, BLAST searches against the *L. mexicana* genome yielded the genes seen in Figure 3-1. Furthermore, after aligning the bait sequences, a profile hidden Markov model was searched against the entire UniProt database. The *L. mexicana* genes returned are represented with ticks in Figure 3-1.

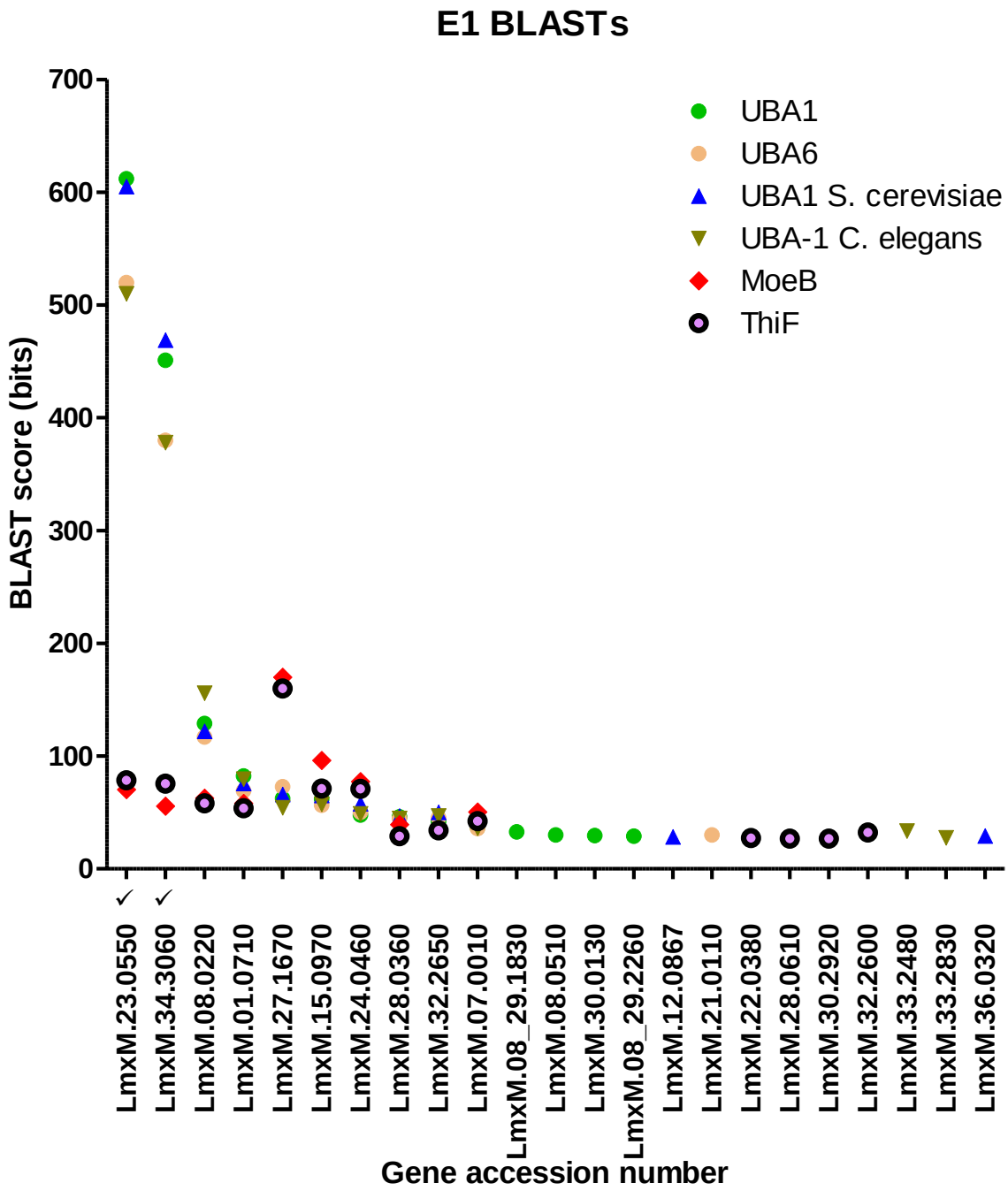


Figure 3-1 Screening the *L. mexicana* genome for putative ubiquitin activating enzyme encoding genes. Using the two human ubiquitin activating enzymes (UBA1 and UBA6), the ancestral prokaryotic enzymes (MoeB and ThiF, sequences from *E. coli*), as well as the *S. cerevisiae* and *C. elegans* orthologs as bait for BLASTs, and to generate a hidden Markov model profile, *L. mexicana* orthologs were identified. Ticks by the gene accession number denote results from hidden Markov models, while the BLAST scores are graphed.

The same method was used to screen for E2 encoding genes (Figure 3-2). In this case, bait sequences were from the human genome, and represented the three different classes of E2s (however, two enzymes were used to represent class I).

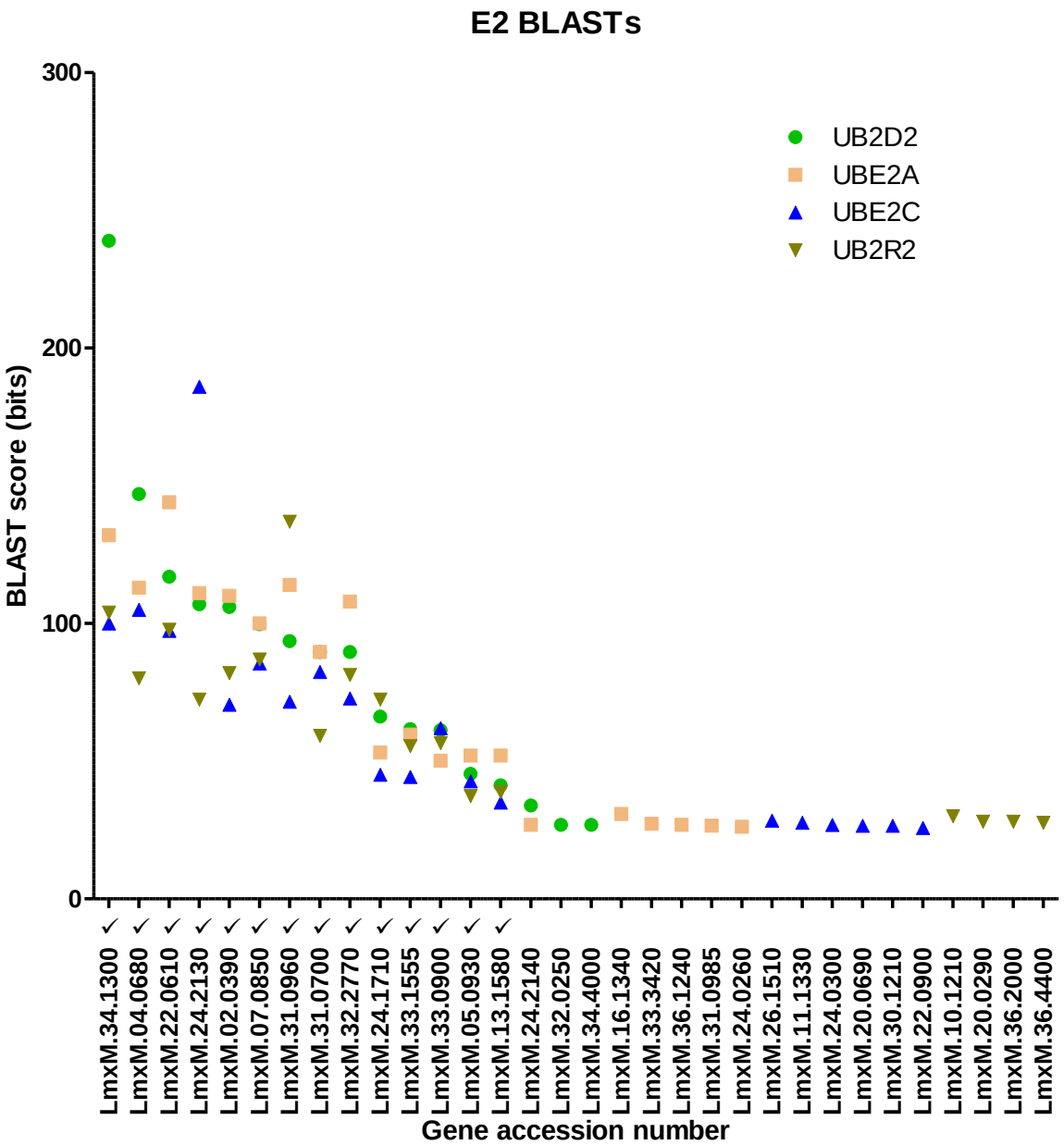


Figure 3-2 Screening the *L. mexicana* genome for ubiquitin conjugating enzyme encoding genes. Using the human conjugating enzymes from class I (UB2D2 and UBE2A), class II (UBE2C), and class III (UB2R2) as bait for BLASTs and to generate a hidden Markov model profile, *L. mexicana* orthologs were identified. Ticks by the gene accession number denote results from hidden Markov models, while the BLAST scores are graphed.

To identify HECT E3s the results portrayed in Figure 3-3 were obtained. Additionally, to identify the other catalytically active class of E3s, RBR, Figure 3-4 was formed.

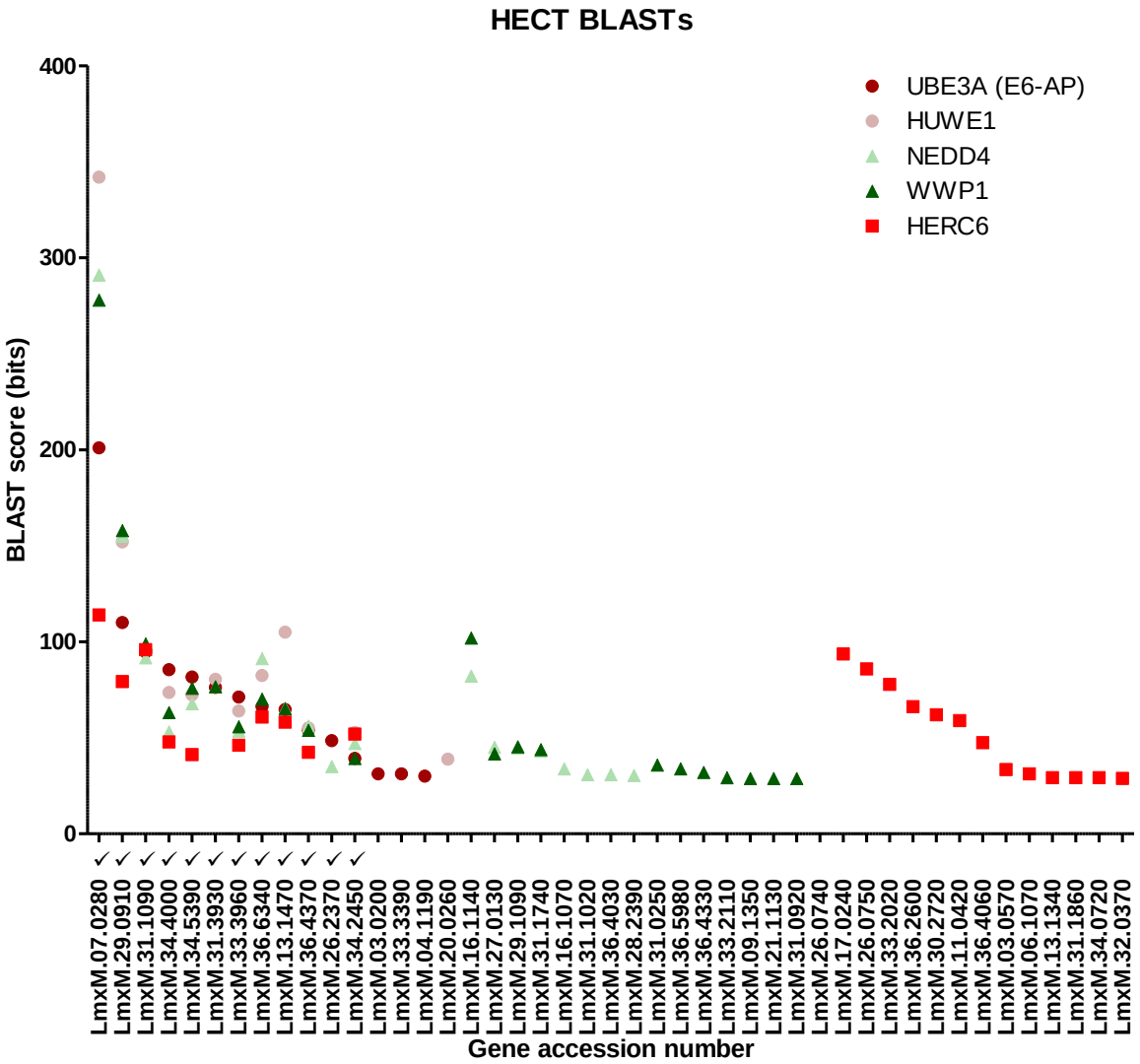


Figure 3-3 Screening the *L. mexicana* genome for ubiquitin ligase genes containing a HECT domain. To identify *L. mexicana* orthologs, human HECT E3s from three different subfamilies were employed: SI(ngle)-HECT (UBE3A and HUWE1), C2-WW-HECT (NEDD4 and WWP1), and those containing RCC1-like domains (HERC6). These were used as bait for BLASTs and to generate a hidden Markov model profile. Ticks by the gene accession number denote a hit from the hidden Markov models search, while the BLAST scores are graphed.

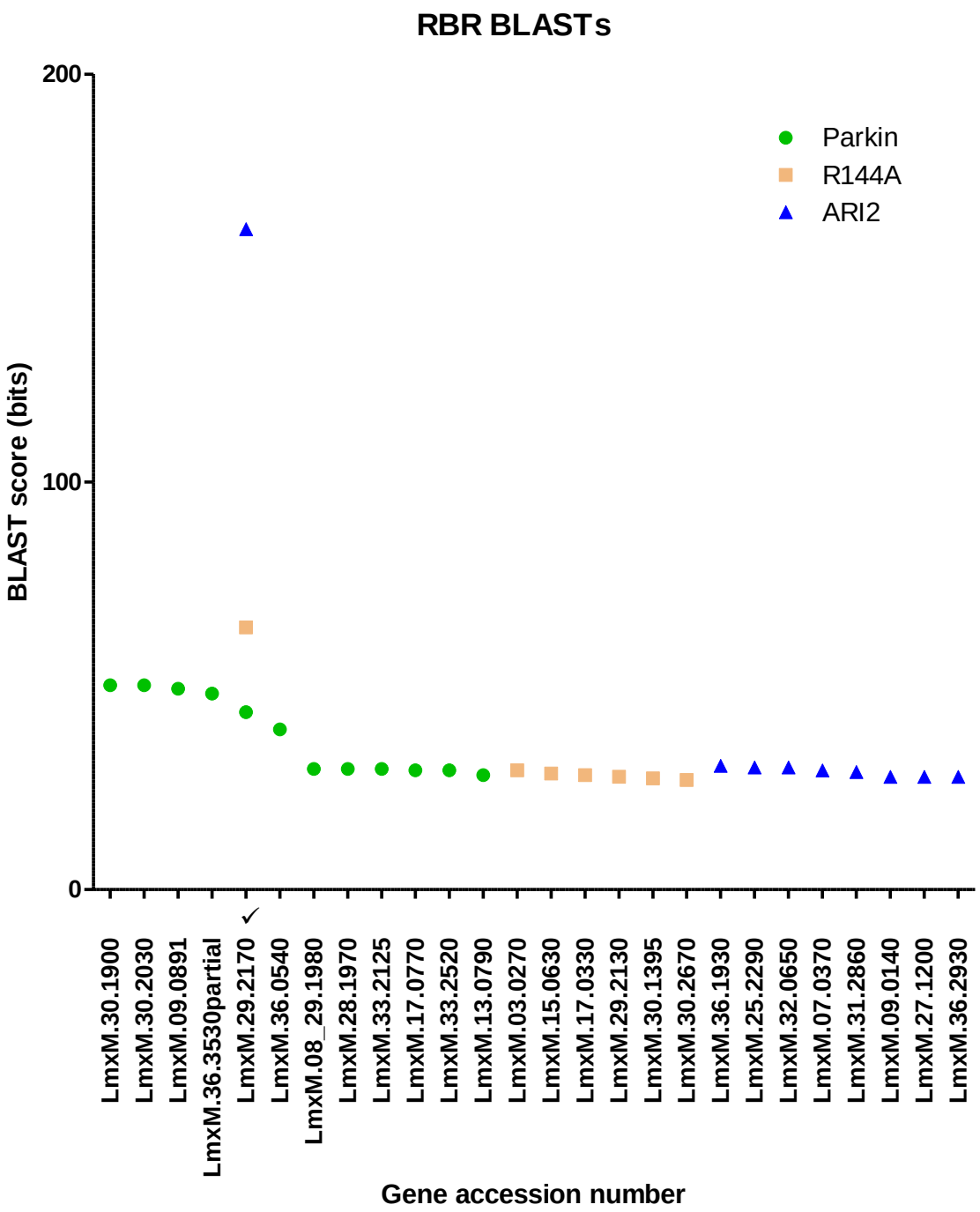


Figure 3-4 Screening the *L. mexicana* genome for ubiquitin ligase genes containing a RBR domain. To identify *L. mexicana* orthologs, three human RBR E3s with diverse domain architectures were used – PARKIN, R144A, and ARI2. These were used as bait for BLASTs and to generate a hidden Markov model profile. A tick by the gene accession number denotes a hit from the hidden Markov models search, while the BLAST scores are graphed.

Table 3-1 surmises the results of the bioinformatic screens. Of note are the ubiquitin-like modifications: SUMOylation, NEDDylation, urmylation, and ufmylation. These were first identified as low scoring hits from BLASTs on ubiquitination machinery, and further BLASTs using the human ubiquitin-like orthologs yielded these IDs. Furthermore, ortholog searches were conducted on

hits returned from the BLAST searches, providing further evidence as to whether a gene encoded enzymes involved in ubiquitin or ubiquitin-like pathways. To examine whether these hits contained relevant domains, domain searches were also performed.

	Ubiquitination	SUMOylation	NEDDylation	Urmylation	Ufmylation
E1	LmxM.23.0550 LmxM.34.3060	LmxM.08.0220 (catalytic) LmxM.28.0360 (regulatory)	LmxM.01.0710 (catalytic) LmxM.32.2650 (regulatory)	LmxM.27.1670	LmxM.15.0970
E2	LmxM.04.0680 LmxM.05.0930 LmxM.07.0850 LmxM.13.1580 LmxM.22.0610 LmxM.24.1710 LmxM.24.2130 LmxM.31.0700 LmxM.31.0960 LmxM.32.2770 LmxM.33.0900 LmxM.33.1555 LmxM.34.1300	LmxM.02.0390			
E3 (HECT)	LmxM.07.0280 LmxM.13.1470 LmxM.26.2370 LmxM.29.0910 LmxM.31.1090 LmxM.31.3930 LmxM.33.3960 LmxM.34.2450 LmxM.34.4000 LmxM.34.5390 LmxM.36.4370 LmxM.36.6340				
E3 (RBR)	LmxM.29.2170				

Table 3-1 A summary of the bioinformatic screening of the *Leishmania mexicana* genome, to identify catalytically active ubiquitination machinery. The terms "catalytic" and "regulatory" refer to subunits in the E1 dimer – "catalytic" refers to the subunit with the active site, while "regulatory" is the subunit without an active site.

Table 3-2 shows this summary combined with data from publicly accessible databases. The relative transcript expression levels were obtained from bulk RNA sequencing (Fiebig, Kelly and Gluenz, 2015). Genes were transformed into the orthologous *Trypanosoma brucei* protein using the TriTryp database, and the localisation of the ortholog searched for using the TrypTag database (Dean, Sunter and Wheeler, 2017). It is worth restating that this work has focused on

E3s with cysteine active sites, as those are potentially easier drug targets than E3s that do not possess an active site (RING E3s). This is in contrast to the Burge *et al.* paper, which included RING E3s (Burge *et al.*, 2020).

	Accession code	Relative transcript expression in different life cycle stages			Ortholog localisation
		Intracellular amas.	Axenic amast.	Promastigote	
E1	LmxM.23.0550	No significant differences			Cytoplasm, nucleoplasm, flagellar cytoplasm
	LmxM.34.3060	No significant differences			Nucleoplasm, peripheral nucleolus, posterior cell tip, cytoplasm, flagella cytoplasm
E2	LmxM.02.0390	No significant differences			Nucleoplasm, cytoplasm
	LmxM.04.0680	Lowest	Lowest	Highest	Cytoplasm, flagellar cytoplasm, nuclear lumen
	LmxM.05.0930	Highest	Intermediate	Lowest	Cytoplasm, flagellar cytoplasm, nuclear lumen
	LmxM.07.0850	No significant differences			Endoplasmic reticulum
	LmxM.13.1580	No significant differences			N/A - tagging failed
	LmxM.22.0610	Highest	Lowest	Lowest	Cytoplasm, flagellar cytoplasm, nuclear lumen, nucleoplasm
	LmxM.24.1710	No significant differences			N/A - ortholog unidentified
	LmxM.24.2130	No significant differences			Cytoplasm
	LmxM.31.0700	Highest	Highest	Lowest	Cytoplasm, endocytic, nucleoplasm, cytoplasm
	LmxM.31.0960	Highest	Intermediate	Lowest	Endocytic, cytoplasm
	LmxM.32.2770	No significant differences			Cytoplasm, flagellar cytoplasm, nuclear lumen
	LmxM.33.0900	Highest	Intermediate	Lowest	Endoplasmic reticulum, nuclear envelope, mitochondrion
	LmxM.33.1555	Highest	Intermediate	Lowest	Cytoplasm, flagellar cytoplasm, nuclear lumen
	LmxM.34.1300	Lowest	Intermediate	Highest	Cytoplasm
E3 (HECT)	LmxM.07.0280	No significant differences			Cytoplasm, nucleoplasm
	LmxM.13.1470	Lowest	Highest	Highest	Endocytic, cytoplasm
	LmxM.26.2370	No significant differences			Cytoplasm, flagellar cytoplasm
	LmxM.29.0910	Highest	Highest	Lowest	Cytoplasm, nucleoplasm
	LmxM.31.1090	Highest	Intermediate	Lowest	Cytoplasm, flagellar cytoplasm, nuclear lumen
	LmxM.31.3930	Lowest	Intermediate	Highest	Cytoplasm
	LmxM.33.3960	Highest	Lowest	Lowest	Nucleolus
	LmxM.34.2450	Highest	Intermediate	Lowest	Endocytic, cytoplasm, flagellar pocket, flagellum
	LmxM.34.4000	Highest	Highest	Lowest	N/A - ortholog unidentified
	LmxM.34.5390	Highest	Highest	Lowest	Cytoplasm
	LmxM.36.4370	Highest	Intermediate	Lowest	N/A - ortholog unidentified
	LmxM.36.6340	No significant differences			Cytoplasm
E3 (RBR)	LmxM.29.2170	Highest	Highest	Lowest	Paraflagellar rod, cytoplasm

Table 3-2 Bioinformatically identified ubiquitination E1/E2/E3s with additional information from database searches. The ubiquitination machinery putatively identified through bioinformatic methods, with details added from a transcriptomic study (Fiebig, Kelly and Gluenz, 2015) and the localisation of the *T. brucei* ortholog from the TrypTag database (Dean, Sunter and Wheeler, 2017). "Amast." refers to amastigotes.

3.3.2 Proteomic screening using an activity-based probe

3.3.2.1 An E1 probe with *Leishmania* lysate

Although bioinformatic screening yielded genes that putatively encoded ubiquitination machinery, whether these genes are expressed at the protein level, and when, is a very different matter. As such, proteomic screening was conducted using an activity-based probe. The results of using a probe specific for E1s is demonstrated in Figure 3-5. In more detail, Figure 3-5 shows a biotin-streptavidin pulldown when using a biotin labelled probe that interacts with E1s,

compared to a streptavidin pulldown on a promastigote lysate without a probe (to show endogenously biotinylated protein), and a pulldown on iodoacetamide treated lysate (to inactivate E1s through alkylation of their cysteine active sites, controlling for other protein-probe interactions). The banding pattern between the experimental and control conditions is clearly different, with several additional bands above the 100 kDa marker only present in experimental conditions. The two putative ubiquitin activating enzymes that were identified through bioinformatic screening were identified using mass spectrometry. Interestingly, using the MASCOT program's emPAI value (Ishihama *et al.*, 2005), LmxM.34.3060 was estimated to be 3-fold more abundant than LmxM.23.0550 in this analysis. However, both of these proteins were present in the iodoacetamide treated negative control, albeit at a lower abundance – for LmxM.34.3060, of the 65 unique peptide groups identified, 60 were found only in the experimental condition; for LmxM.23.0550, of the 32 unique peptides groups identified, 22 of these were only found in the experimental condition.

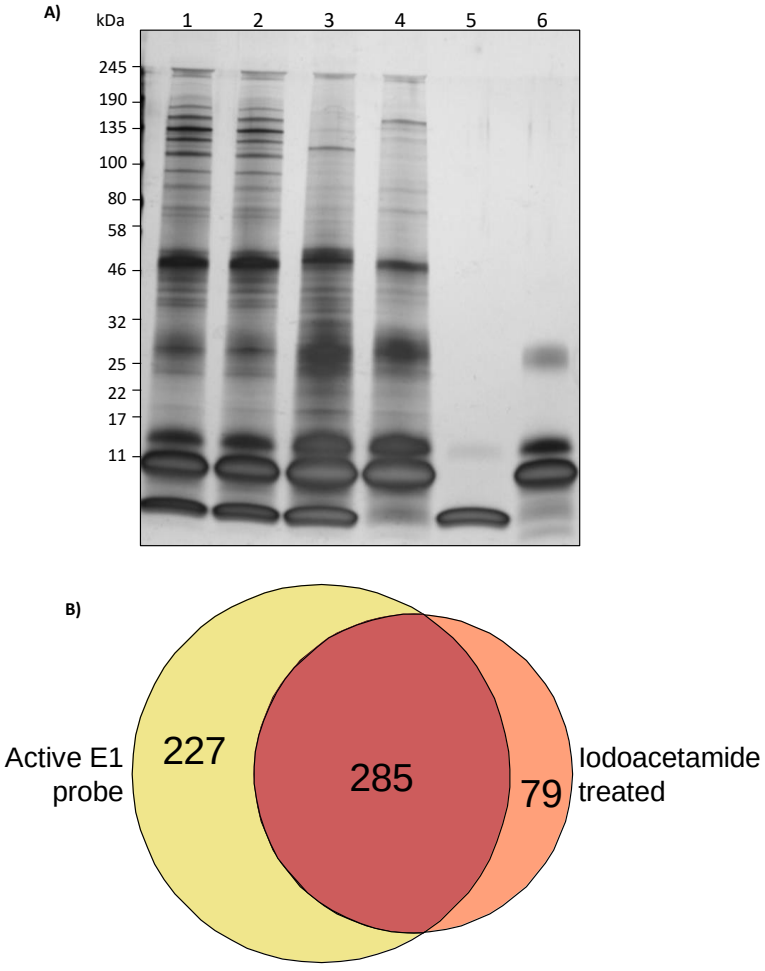


Figure 3-5 An activity-based probe, specific for ubiquitin activating enzymes, reacting with *L. mexicana* promastigote lysate. A) 34.5 µg/ml (1) and 69 µg/ml (2) of E1 probe (product code

UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. These reactions were compared against iodoacetamide treated lysate with 69 µg/ml of probe (3), cell lysate without probe or iodoacetamide (4), probe without lysate (5), and the supernatant present after boiling streptavidin beads in Laemmli buffer for 10 minutes (6). All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** Samples from lane 2 (69 µg/ml of E1 probe) and lane 3 (iodoacetamide treated lysate) were analysed by mass spectrometry, with the results shown in a Venn diagram

Given that of the 512 total proteins only two were of interest (interest defined as present in the bioinformatic screening campaign), further attempts were made to reduce the background, in case this background masked a signal from other proteins of interest. Additionally, optimising this probe, which reacts with a subset of ubiquitination enzymes, may aid in the use of other activity-based probes that interact with more ubiquitination enzymes. To reduce the background, RIPA buffer was used to wash the streptavidin beads after pulldown, as opposed to PBS in prior iterations. Figure 3-6 shows how this impacted the results. The number of wash steps alongside the volume of streptavidin beads were optimised in Figure 3-7.

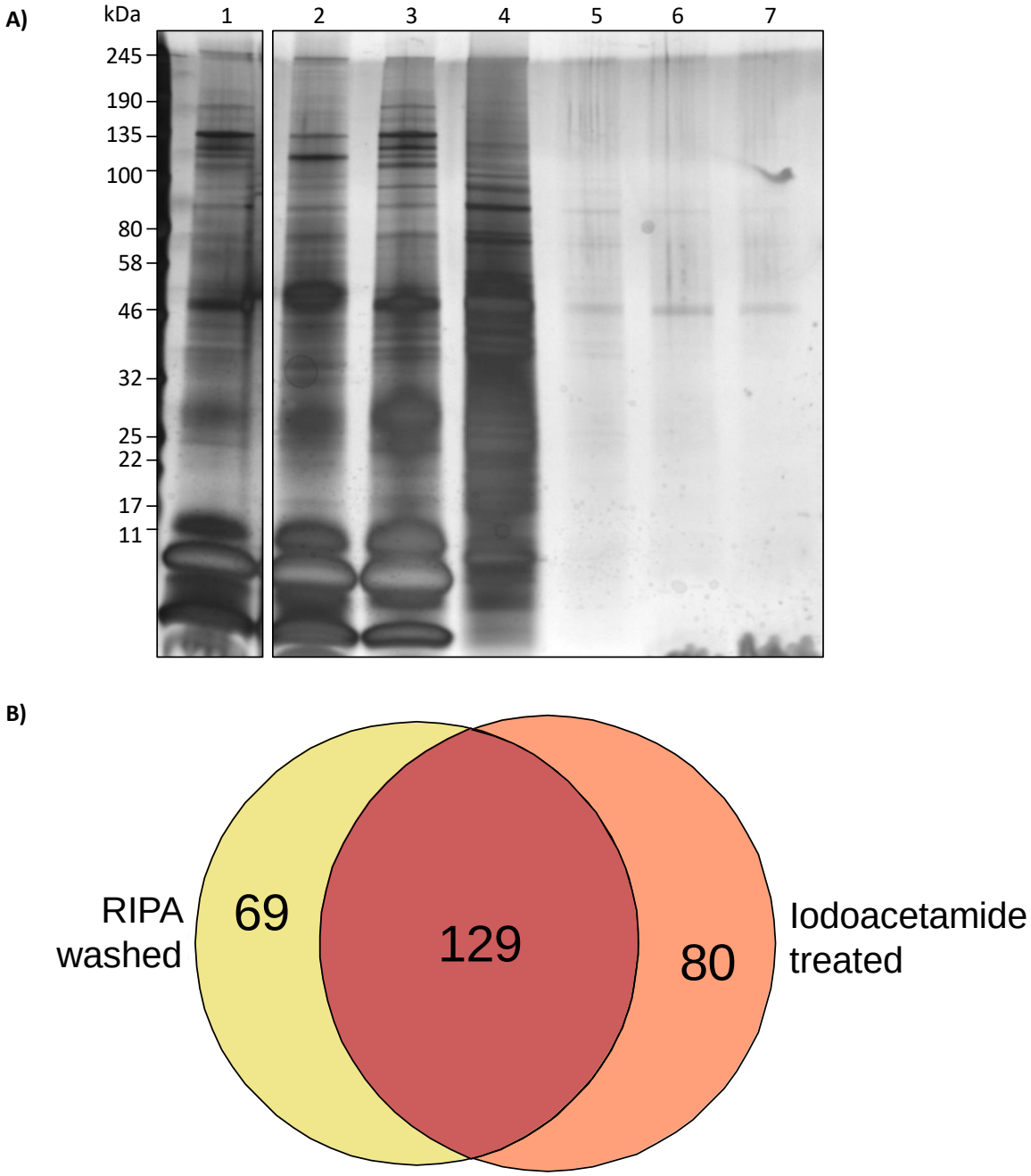


Figure 3-6 Optimising wash buffers, to remove more contaminants present in a reaction of an E1 specific activity-based probe with *L.mexicana* promastigote lysate. A) 34.5 µg/ml of E1 probe (product code UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. Lane 1 features the eluate from streptavidin beads, after four washes in RIPA buffer. This is compared to an iodoacetamide treated lysate (2), and a probe-lysate reaction washed with four lots of PBS (3). Respectively, lanes 4, 5, 6, and 7, show protein removed in the 1st, 2nd, 3rd, and 4th washes in RIPA buffer. All samples were run on the same 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** Samples from (1) and (2) were sent for mass spectrometry, with the resulting protein IDs displayed as a Venn diagram.

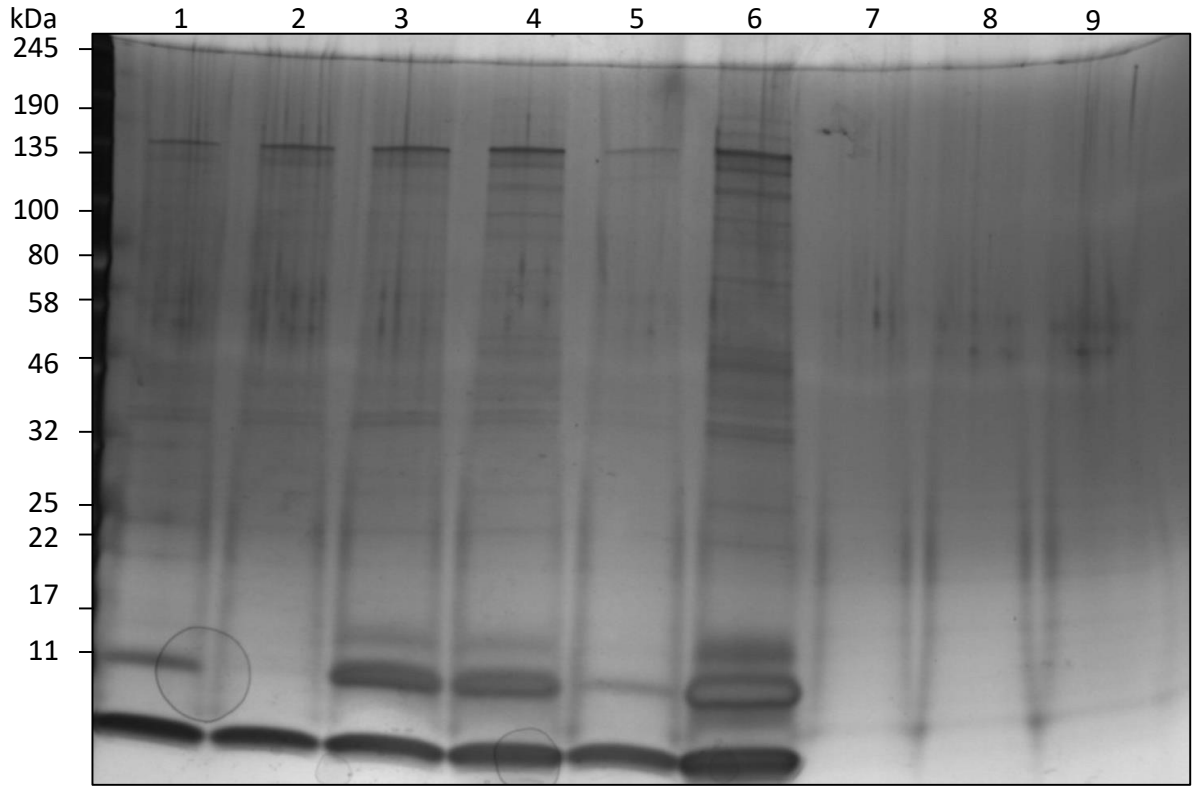


Figure 3-7 Optimising streptavidin bead volume and the number of washes for an activity-based probe, specific for ubiquitin activating enzymes, reacting with *L. mexicana* promastigote lysate. 34.5 µg/ml of E1 probe (product code UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. These reactions differed in the volume of streptavidin beads employed in the pulldown: 15 µl (1), 20 µl (2), 25 µl (3), or 30 µl (4). Furthermore, these reactions were washed more thoroughly than in prior experiments, undergoing 10 washes in RIPA buffer. These were compared against probe incubated with iodoacetamide treated lysate, eluted with 30 µl of streptavidin beads, after 10 RIPA washes (5). Additionally, samples were contrasted against a sample representative of a previous protocol, using 2 × PBS washes and 3 × RIPA washes (6). Lanes 7, 8, and 9 show the protein removed in the final three washes (washes 8, 9, and 10 respectively). All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

To further reduce non-specific binding two further method developments were trialled: adding Tween 20 during the streptavidin bead and biotin probe incubation, and eluting with a biotin based elution buffer (as opposed to boiling the beads in Laemmli buffer) (Figure 3-8). Eluting using biotin returned the fewest protein IDs, 25 in total, whereas boiling in Laemmli returned approximately 90 identifications (in both the Tween 20 and iodoacetamide control). In all samples both LmxM.34.3060 and LmxM.23.0550 were present, though the Tween 20 sample had the most peptides from these putative E1s identified. The Tween 20 sample also had an estimated 12-fold difference in abundance of LmxM.34.3060 to LmxM.23.0550, whereas the proteins were estimated to be of almost equal abundance when eluted using biotin.

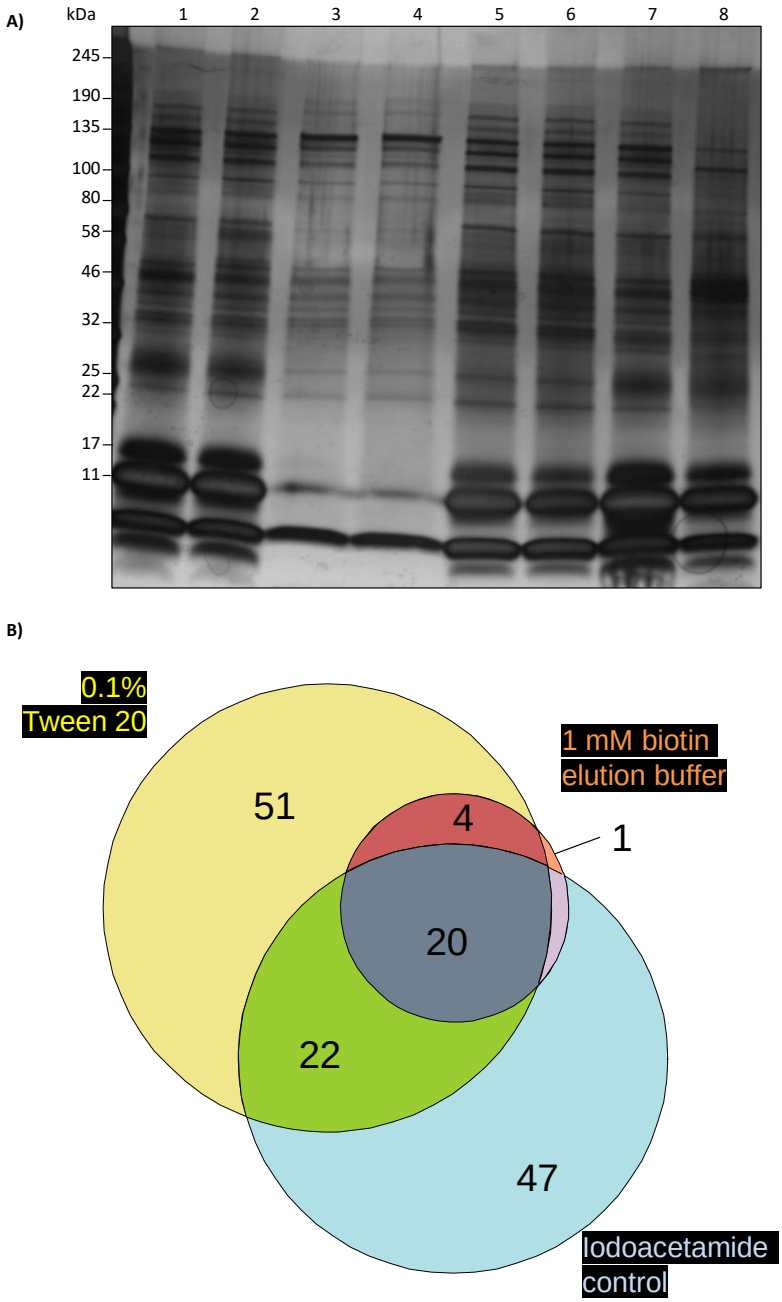


Figure 3-8 Optimising binding and elution methods for an activity-based probe, specific for ubiquitin activating enzymes, reacting with *L. mexicana* promastigote lysate. A) 34.5 µg/ml of E1 probe (product code UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. In lane 1 0.05% Tween 20 was added alongside the streptavidin beads, lane 2 had 0.1% added at the same point. Lane 3 shows an elution buffer using 0.5 mM biotin. Lane 4 shows an elution buffer using 1 mM of biotin. After incubation with the biotin elution buffers, the same streptavidin beads were boiled in Laemmli buffer for 10 mins. Supernatant from boiled beads used with 0.5 mM biotin elution buffer are shown in lane 5, and those beads used with 1 mM of biotin are shown in lane 6. Lane 7 shows a positive control, using prior elution methods. Lane 8 is a negative control, with iodoacetamide treated lysate. All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** Samples from lane 2 (0.1% Tween 20 during pulldown), lane 4 (1 mM biotin elution buffer), and lane 8 (iodoacetamide treated lysate) underwent mass spectrometry, with protein identifications shown in a Venn diagram.

Using axenic amastigotes in lieu of promastigotes yields similarly distinct bands above 100 kDa, not present in control conditions, as seen in Figure 3-9.

Furthermore, a new negative control was trialled - a chemically inactive probe, that was unable to form a covalent bond with E1s, replacing the iodoacetamide control. Mass spectrometric analysis revealed two of the predicted ubiquitin activating enzymes: LmxM.23.0550 and LmxM.34.3060. Of note are the estimated abundances within the sample, providing emPAI scores of 0.04 and 0.63 respectively. These scores suggest LmxM.34.3060 is approximately 16-fold more abundant in the sample than LmxM.23.0550. These enzymes did not appear in the negative control.

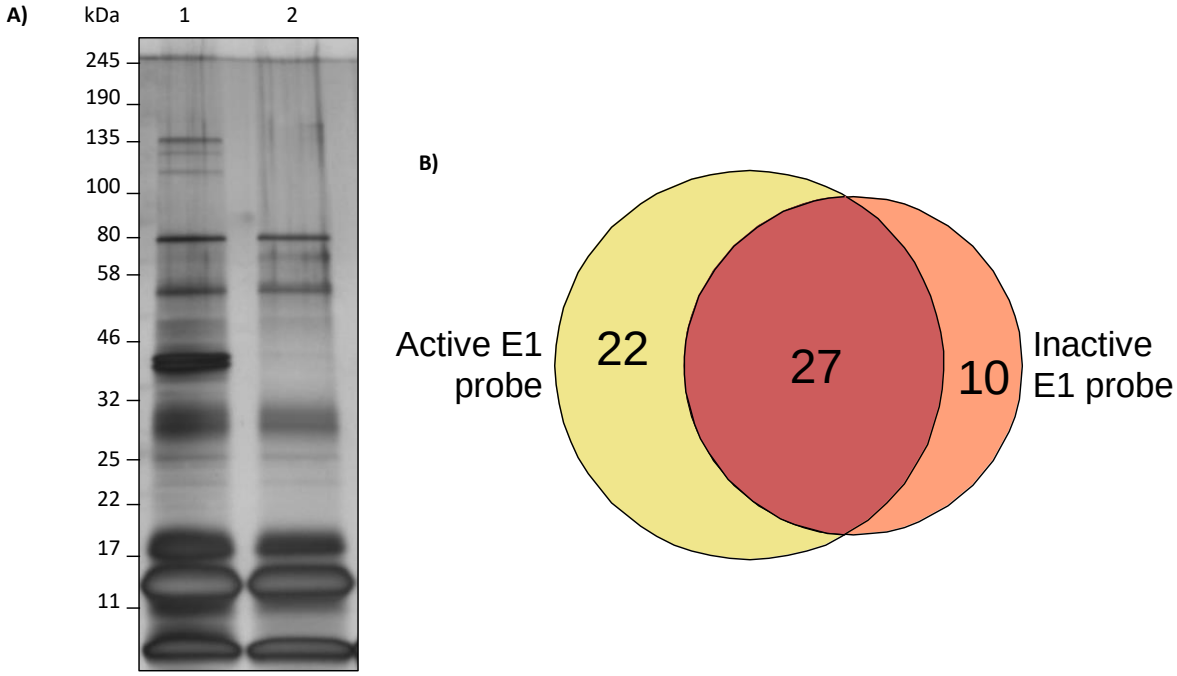


Figure 3-9 An activity-based probe, specific for ubiquitin activating enzymes, reacting with *L. mexicana* amastigote lysate. A) 34.5 µg/ml of E1 probe (product code UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* axenic amastigotes (1). This reaction was compared against cell lysate reacting with an inactivated probe (2). All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** Venn diagram showing the distribution of mass spectrometric protein identifications when using chemically active and chemically inactive probes.

3.3.2.2 A cascade probe with *Leishmania* cell lysate

In order to identify E2s and HECT/RBR E3s, an activity-based probe was employed that should be transferred through the enzymatic cascade referred to here as a cascade probe. Initially, fluorescent probes were employed to optimise the reaction. Figure 3-10 shows varying concentrations of a rhodamine labelled probe, and Figure 3-11 varies the concentration of a Cy5 labelled probe. Notable bands are the three bold bands under 25 kDa in all conditions of all experiments, including the negative control (apyrase was used to remove ATP from the lysate,

thereby precluding the probe from entering the enzymatic cascade). Furthermore, these bands are present when lysate was not used in the reaction mixture (lane 5).

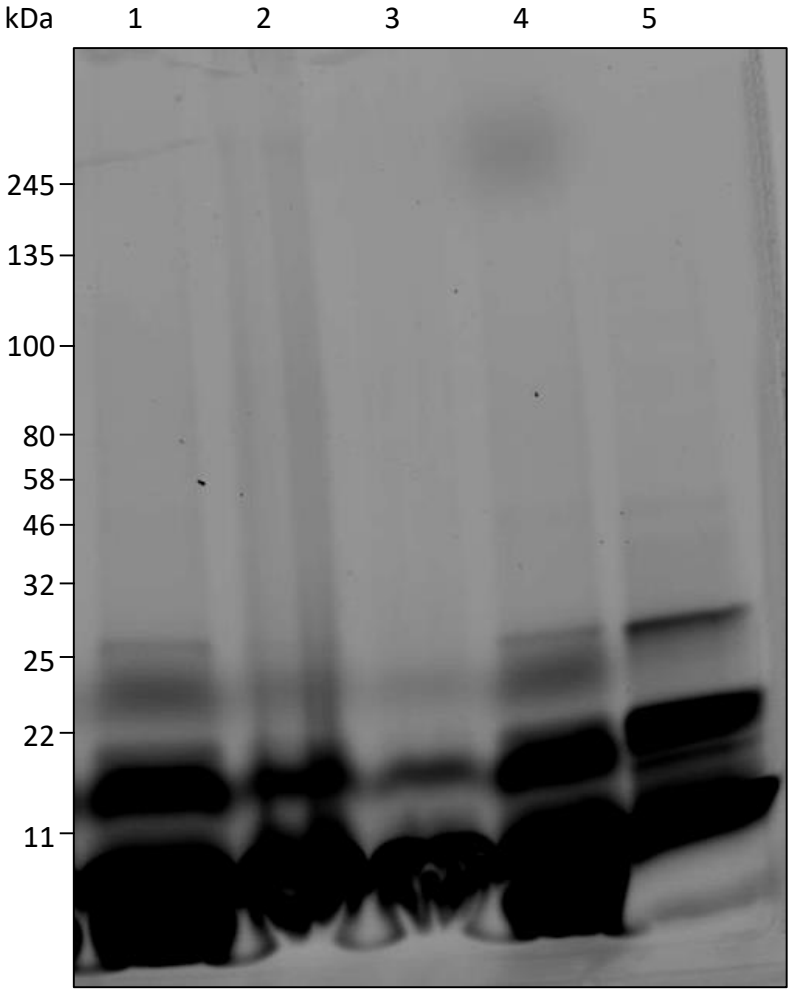


Figure 3-10 Rhodamine fluorescence scan of various concentrations of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate. Rhodamine labelled cascade probe (product code UbiQ-131, supplied by UbiQ) was incubated with various concentrations of *L. mexicana* promastigote lysate. Lane one shows a reaction that featured probe at 20 µg/ml, lane two a reaction with probe at 10 µg/ml, and lane three used 5 µg/ml probe. Lane four also used 20 µg/ml probe, but the lysate was incubated with apyrase. Lane five shows a sample of probe run without lysate. All samples were run on the same 4-20% SDS-PAGE gel and scanned for rhodamine fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

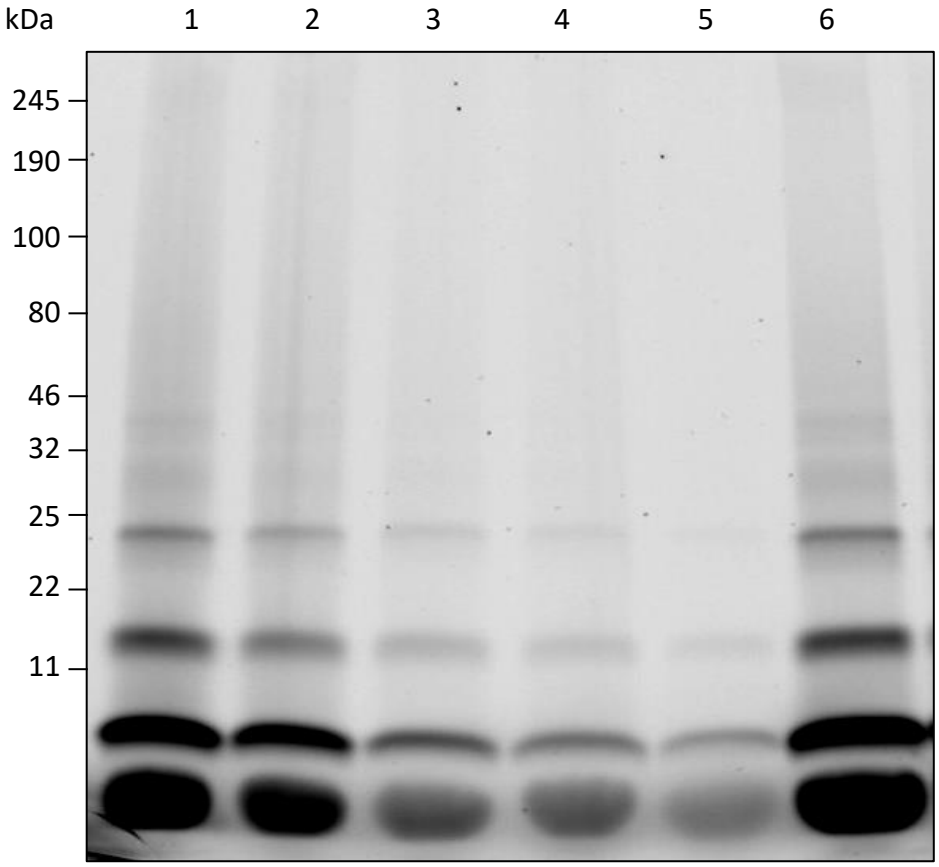


Figure 3-11 Cy5 fluorescence scan of various concentrations of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate. Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with various concentrations of *L. mexicana* promastigote lysate. Lane one shows a reaction that featured probe at 20 µg/ml, lane two a reaction with probe at 10 µg/ml, lane three used 5 µg/ml probe, lane four 2.5 µg/ml probe, lane five 1 µg/ml probe. Lane six used 20 µg/ml of probe, after the lysate was incubated with apyrase. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

An experiment to vary the concentration of promastigote lysate is shown in Figure 3-12. Again, the most distinct bands are present at and below 25 kDa. At low concentrations of lysate there is some faint banding at 46 kDa and above, however these bands are present when the probe is without lysate.

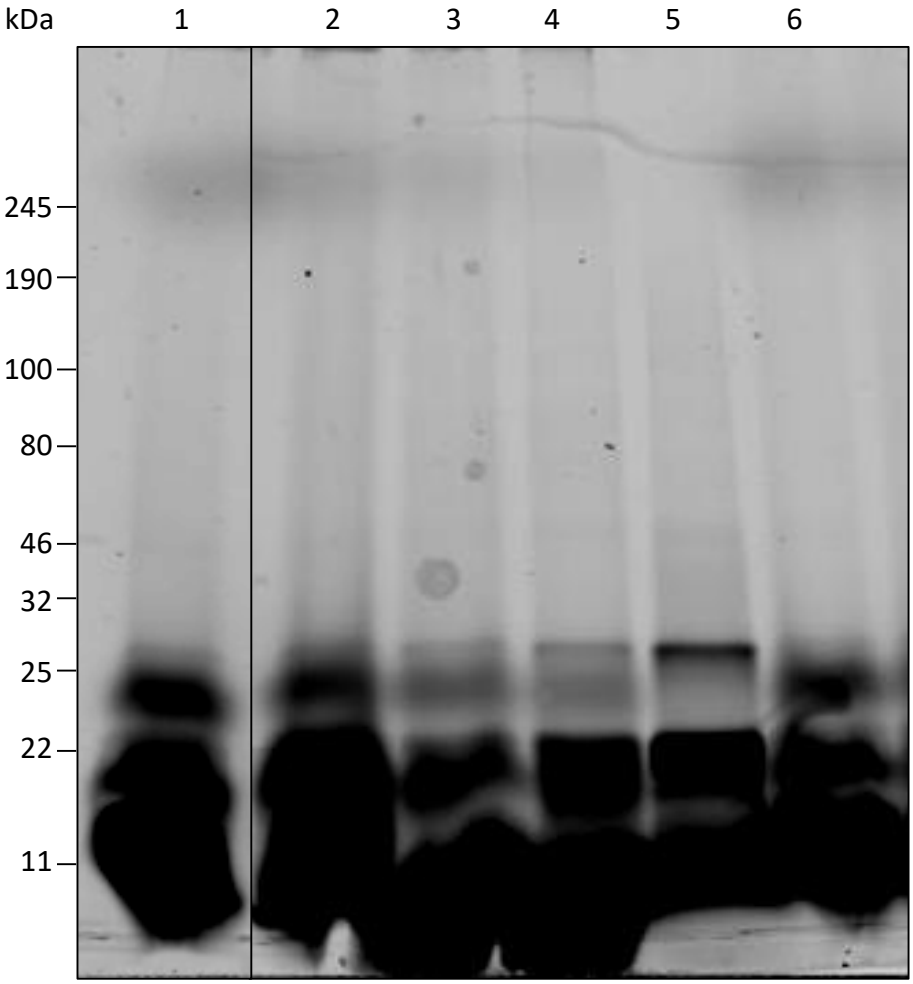


Figure 3-12 Rhodamine fluorescence scan of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate, varying the concentration of cell lysate. 20 µg/ml of rhodamine labelled cascade probe (product code UbiQ-131, supplied by UbiQ) was incubated with various concentrations of *L. mexicana* promastigote lysate. Lane one shows a reaction that featured cell lysate at 2 mg/ml, lane two a reaction with lysate at 1.5 mg/ml, lane three 1 mg/ml lysate, and lane four 0.5 mg/ml lysate. Lane five shows a sample of probe run without lysate. Lane six was treated with 2 U of apyrase before sonication, and used 2 mg/ml lysate. All samples were run on the same 4-20% SDS-PAGE gel and scanned for rhodamine fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

A revision to the protocol was made around cell lysis – adding the activity-based probe pre-lysis, then lysing using a water bath sonicator (as opposed to lysing with a probe sonicator, then adding the probe). Figure 3-13 shows the impact of these method improvements, on a Cy5 labelled probe. Although faint, bands not visible in previous experiments can be seen between 32 and 46 kDa, which are absent from the negative control.

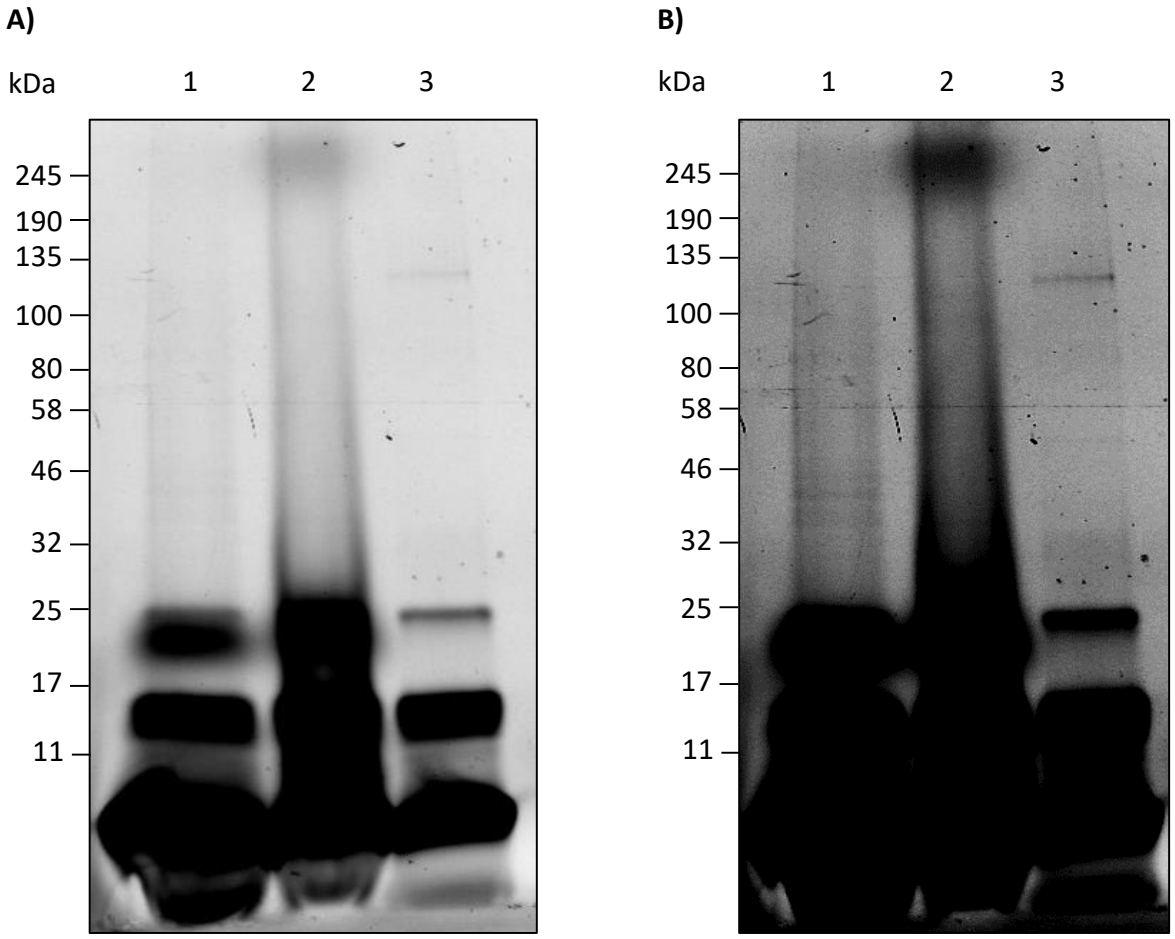


Figure 3-13 Cy5 fluorescence scan of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate. A) 20 µg/ml of Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows the aforementioned reaction, lane two shows the same reaction after treatment with 2 U of apyrase. Lane three shows the probe reacting with human UBA1 (product code 61-0001-050, supplied by Ubiquigent). **B)** The same image as in A), but with increased contrast. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

Given an identifiable signal, present only in experimental conditions, optimisation could begin. Figure 3-14 shows varying concentrations of a Cy5 labelled probe – while existing bands get darker the more probe is used, no new bands appear. Figure 3-15 shows the effect of modulating the concentration of essential ligands – ATP and $MgCl_2$ – in the reactions. The clearest banding is found without the addition of ATP or $MgCl_2$. Figure 3-16 shows the effect of varying the incubation time of the reaction, there does not seem to be any visible differences in the banding patterns. Additionally, a methodological improvement applied in these experiments was to run proteins <25 kDa off the polyacrylamide gel, permitting greater visibility of bands at higher molecular weights. Although a banding pattern becomes gradually more distinct over the course of these experiments, there is no differentiation between the negative control and experimental conditions.

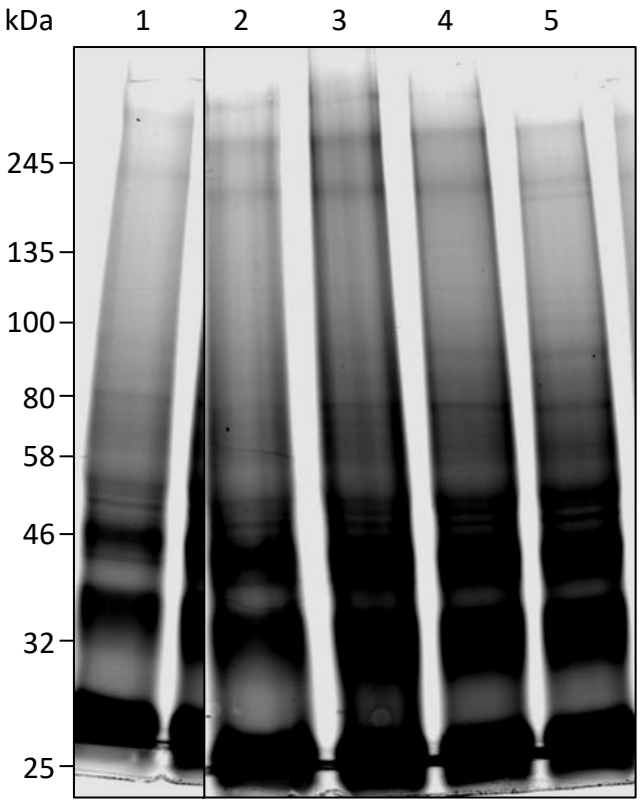


Figure 3-14 Cy5 fluorescence scan of various concentrations of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate. Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows a reaction that used 37.5 µg/ml of probe, lane two employed 54 µg/ml of probe, lane three 69 µg/ml, lane four 83 µg/ml, and lane five 83 µg/ml of probe with treatment with 2 U of apyrase. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

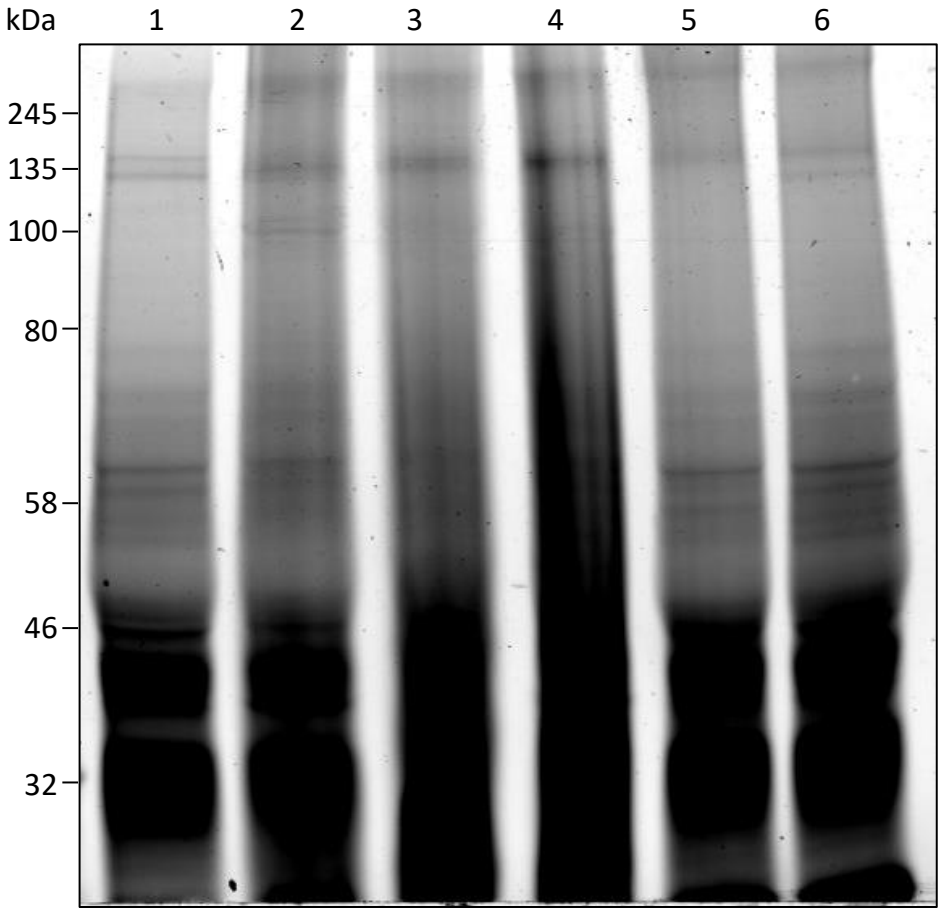


Figure 3-15 Cy5 fluorescence scan of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate, using various concentrations of ATP and $MgCl_2$. 34.5 µg/ml of Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows a reaction that had no added ATP or $MgCl_2$, lane two pictures a reaction with 7.5 mM ATP and 15 mM $MgCl_2$ added, lane three 15 mM ATP and 30 mM $MgCl_2$, and lane four used 22.5 mM ATP and 45 mM $MgCl_2$. Lane five used 15 mM ATP and 30 mM $MgCl_2$ as well, though addition was before cell lysis (contrary to other reactions). Lane six was treated with 2 U of apyrase before sonication, and did not have any additional ATP. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

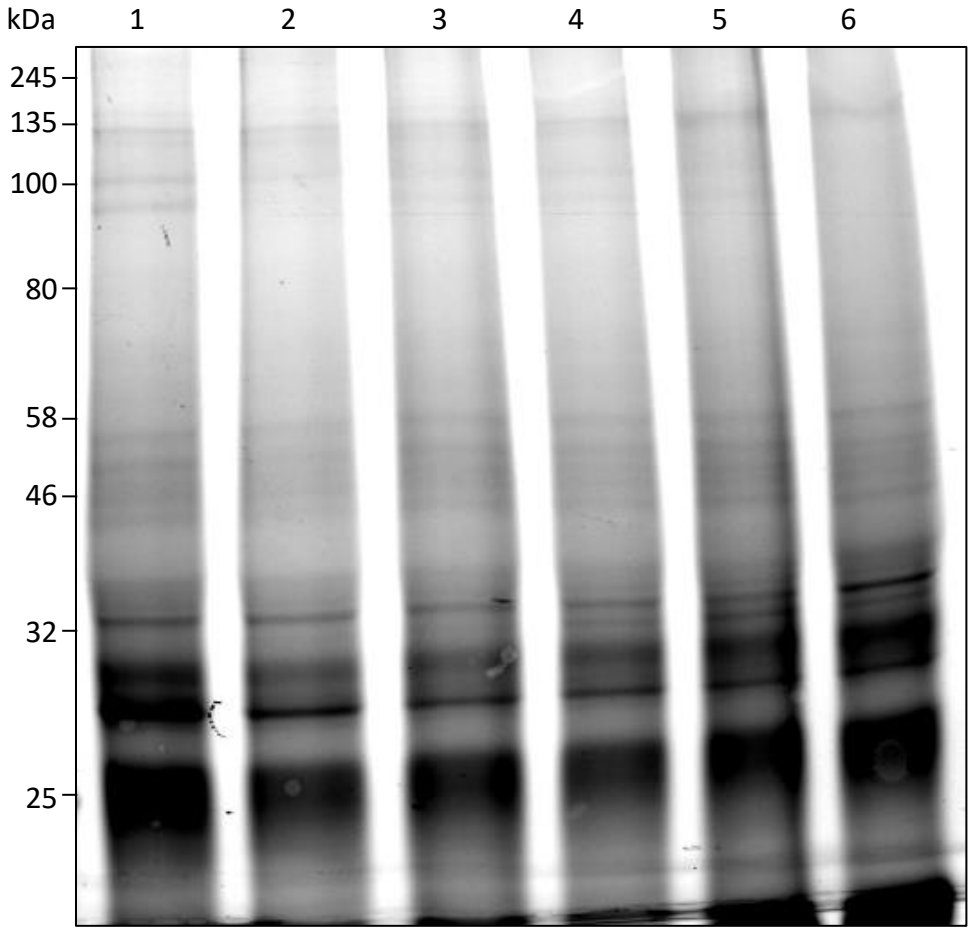


Figure 3-16 Cy5 fluorescence scan of an activity-based ubiquitin probe that reacted with *L. mexicana* promastigote lysate for various time periods. 34.5 µg/ml of Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows a reaction that ran for 15 minutes, lane two reacted for 30 minutes, lane three for 45 minutes, lane four for 60 minutes, and lane five for 90 minutes. Lane six also ran for 90 minutes, but was first treated with 2 U of apyrase. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

To compare the fluorescence between rhodamine and Cy5 labelled probes, the same lysate, reacting with each probe under the same method, was run on the same gel. This can be seen in Figure 3-17. Note the distinct banding patterns, entirely dependent on whether a rhodamine or Cy5 labelled probe is employed.

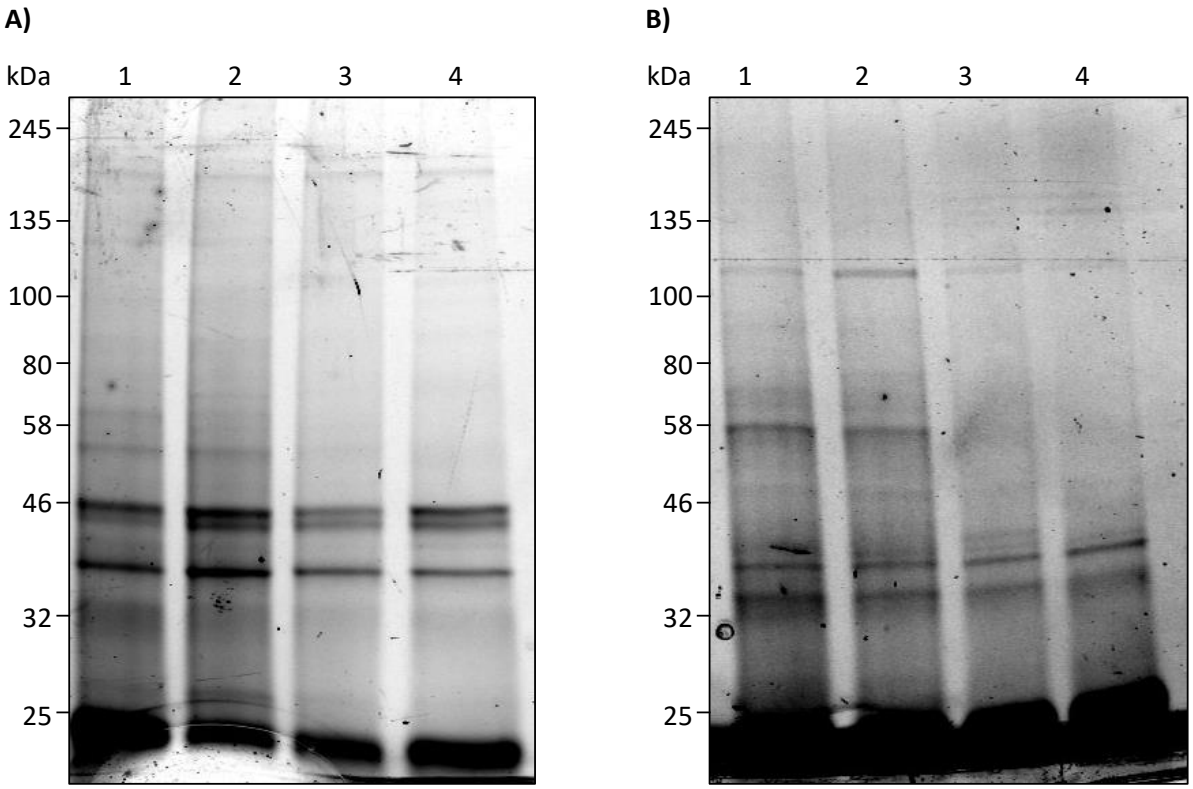


Figure 3-17 Cy5 and rhodamine fluorescence scans of an activity-based ubiquitin probe reacting with *L. mexicana* promastigote lysate, varying the concentration of both probes. A) Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ) was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows a reaction with 1 mg/ml of probe, lane two used 3 mg/ml probe, lane four used 3 mg/ml probe after treatment with apyrase. **B)** Rhodamine labelled cascade probe (product code UbiQ-131, supplied by UbiQ was incubated with lysate from 3e7 *L. mexicana* promastigotes. Lane one shows a reaction with 1 mg/ml of probe, lane two used 3 mg/ml probe, lane four used 3 mg/ml probe after treatment with apyrase. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 and rhodamine fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

Given that the banding pattern was dependent on the fluorescent label employed, one probe's optimisation may not be generalisable to other probes, so further optimisation was conducted using the biotin labelled probe (the same probe that would be employed for biotin-streptavidin pulldowns and subsequent mass spectrometry). Further parameters to be optimised were the cell number needed to give a signal after a biotin-streptavidin pulldown, and the elution method to retrieve the biotin labelled E1/E2/E3s from the streptavidin beads (Figure 3-18). Despite distinct banding patterns on a silver stained gel between the negative control (lane 6) and its most similar experimental condition (lane 4), none of the anticipated target proteins were identified through mass spectrometry.

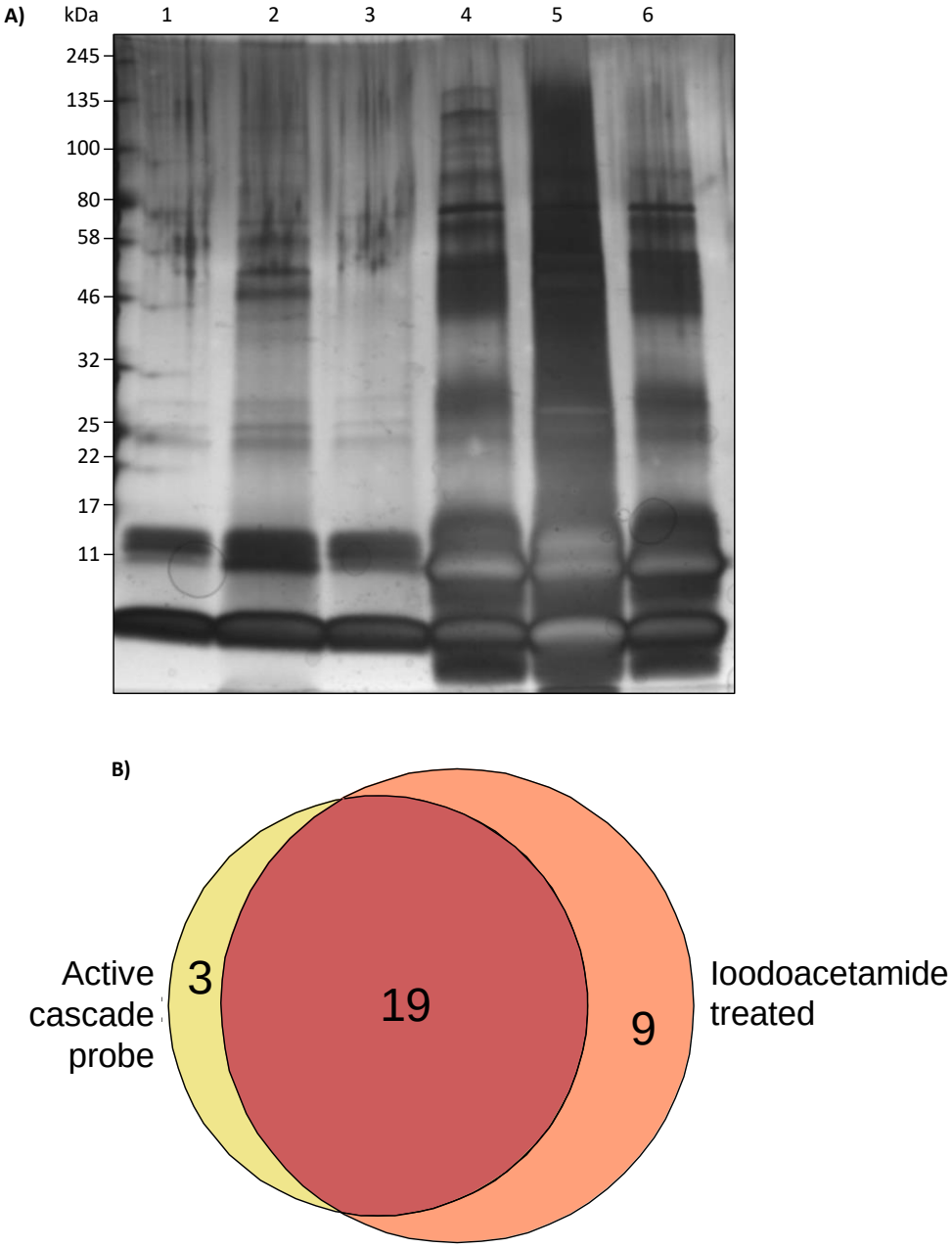


Figure 3-18 Optimising cell number and elution for an activity-based probe, designed to interact with ubiquitin conjugation machinery, reacting with *L. mexicana* promastigote lysate. A) 34.5 µg/ml of cascade probe (product code UbiQ-102, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. In lane one 7.5e8 cells were used, and elution conducted with a biotin-based elution buffer. In lane two 7.5e7 cells were used, and elution conducted with a biotin-based elution buffer. In lane three 7.5e8 cells were treated with iodoacetamide before sonication, and eluted with a biotin based elution buffer. Lane four used 7.5e8 cells for the reaction mix, and elution was conducted by boiling streptavidin beads in Laemmli buffer. Lane five used 7.5e7 cells for the reaction mix, with elution conducted by boiling streptavidin beads in Laemmli buffer. In lane six 7.5e8 cells were treated with iodoacetamide before sonication, and eluted by boiling streptavidin beads in Laemmli buffer. All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** The conditions in lane four (7.5e8 cells, eluted by boiling in Laemmli buffer) and lane six (7.5e8 cells treated with iodoacetamide), underwent mass spectrometry, with the number of protein identifications in both samples shown in a Venn diagram.

The volume of streptavidin beads used for pulldowns was also optimised, as was the concentration of ATP and $MgCl_2$ (Figure 3-19). It was thought that a deficit of

ATP could be preventing the cascade probe from interacting with the ubiquitination machinery, and so extra ATP and $MgCl_2$ were added to the reactions (7.5mM ATP and 10 mM $MgCl_2$), or perhaps a bead proteome was masking any signal from any proteins of interest (Trinkle-Mulcahy *et al.*, 2008). Although there was a correlation between the microlitres of beads used for pulldowns and the number of protein identifications obtained, neither the addition of extra ATP nor the use of fewer beads resulted in any identifications of proteins of interest.

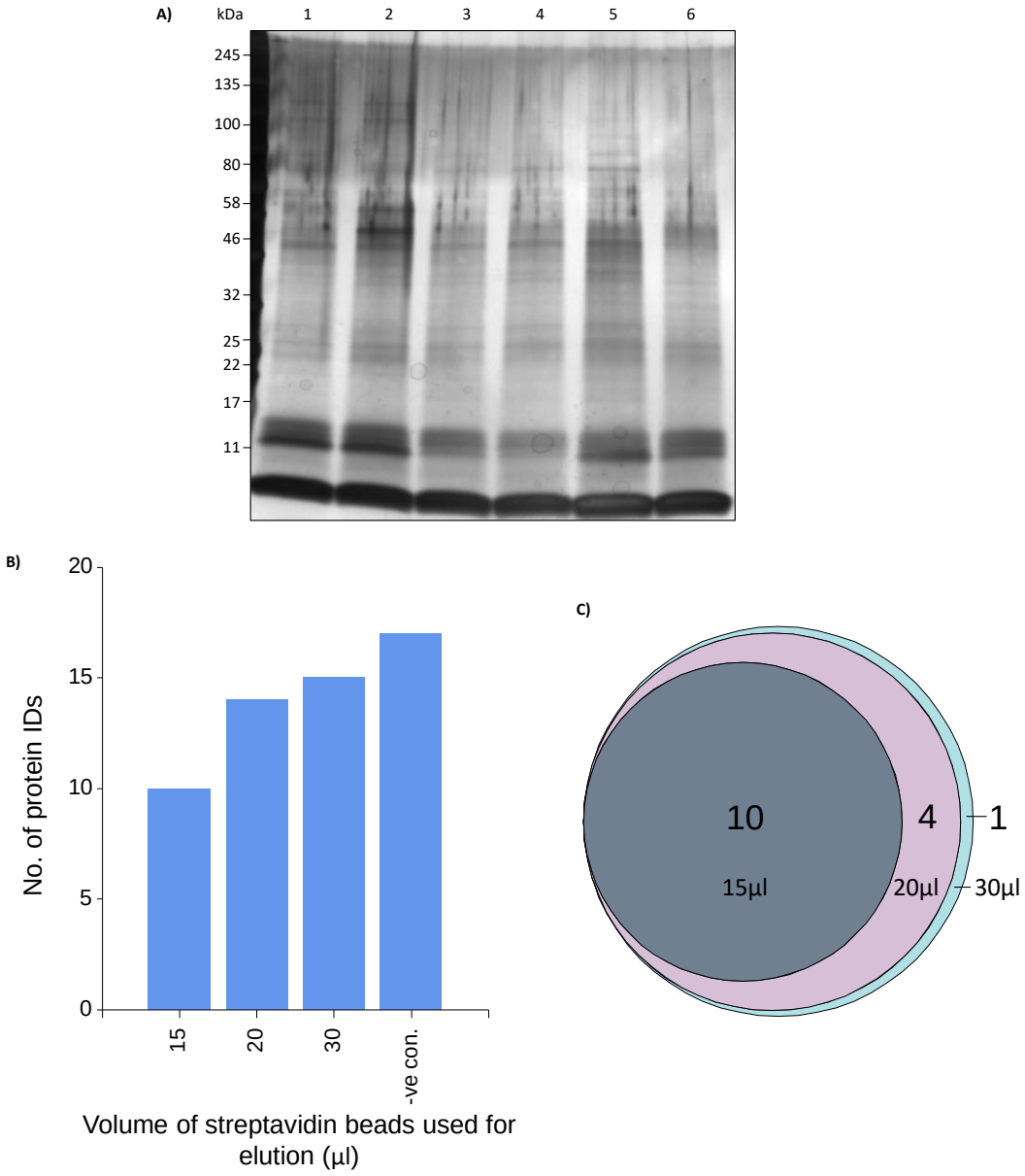


Figure 3-19 Optimising bead volume and ATP concentration for an activity-based probe, designed to interact with ubiquitin conjugation machinery, reacting with *L. mexicana* promastigote lysate. A) 34.5 µg/ml of cascade probe (product code UbiQ-102, supplied by UbiQ) was incubated with lysate from 7.5e7 *L. mexicana* promastigotes. In lane one 15 µl of streptavidin beads were used to elute, lane two 20 µl, lane three 25µl, lane four 30µl. All four reactions also had 7.5mM ATP and 10 mM $MgCl_2$ added. Lane five used 25 µl of beads, without any additional ATP or $MgCl_2$. Lane six features a reaction that was treated with iodoacetamide before sonication. All

samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** Number of protein identifications achieved through mass spectrometry with various volumes of streptavidin beads, displayed as a bar chart. **C)** A Venn diagram showing the similarities between the samples.

3.3.2.3 A cascade probe with human cell lysate

In an attempt to reproduce published data using the cascade probe (Mulder *et al.*, 2016), THP1 cells, a human monocyte line (Tsuchiya *et al.*, 1980), were used with a fluorescently labelled probe. This is shown in Figure 3-20.

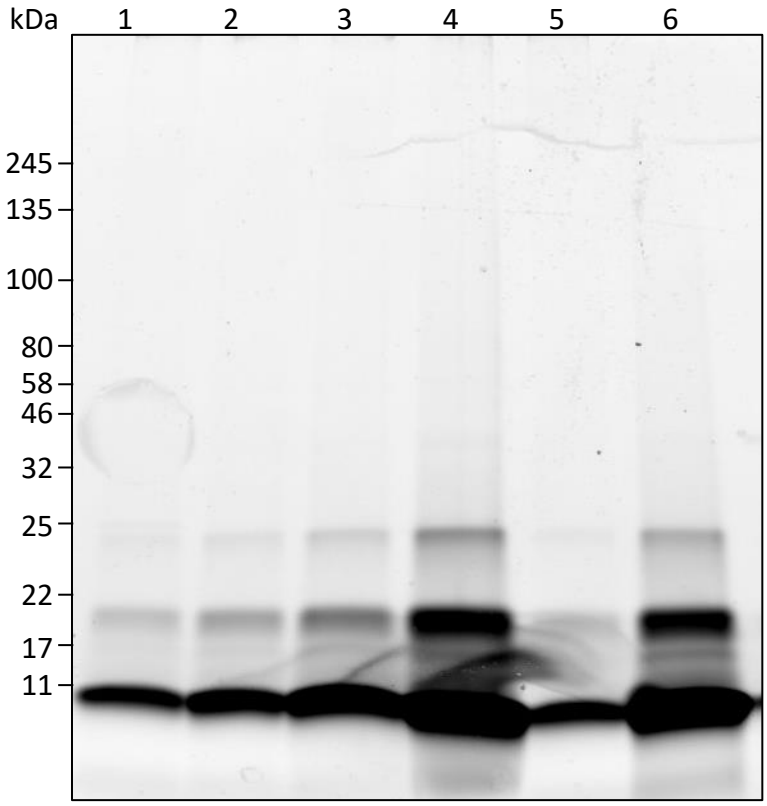


Figure 3-20 Cy5 fluorescence scan of various concentrations of an activity-based ubiquitin probe reacting with THP1 lysate. Various concentrations of THP1 lysate were incubated with 20 µg/ml of Cy5 labelled cascade probe (product code UbiQ-104, supplied by UbiQ). Lane one shows a reaction using lysate at 20 mg/ml, lane two had lysate at 10 mg/ml, lane three at 5 mg/ml, and lane four at 1 mg/ml. Lane five also used lysate at 20 mg/ml, after treatment with 2 U of apyrase. Lane six is a sample of the probe, without lysate. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

THP1 cells were used again, with an activity-based cascade probe that had undergone further purification. The aim of further purification was to remove probe oligomers. These results are demonstrated in Figure 3-21. A banding pattern is present <25 kDa, which is independent of the presence of lysate.

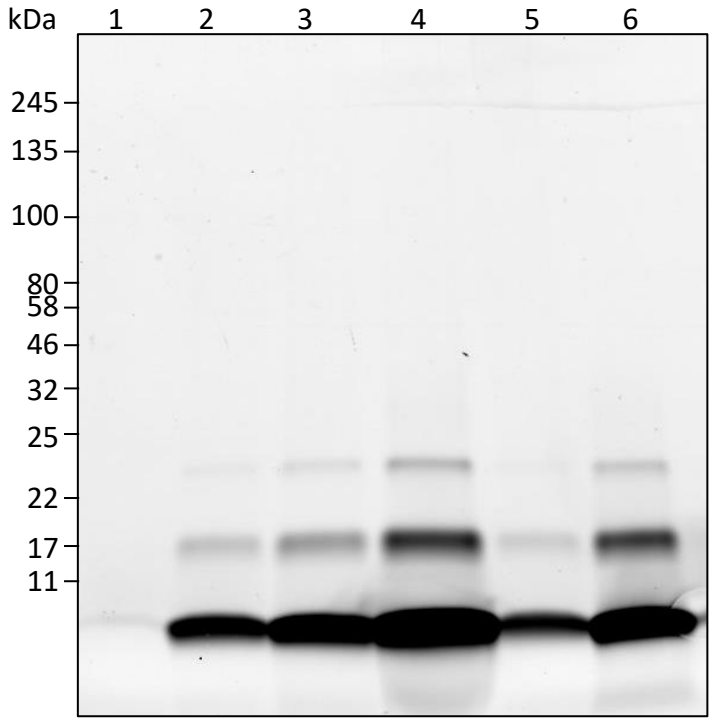


Figure 3-21 Cy5 fluorescence scan of various concentrations of a purified, activity-based ubiquitin probe reacting with THP1 lysate. Various concentrations of THP1 lysate were incubated with 20 µg/ml of (product code UbiQ-104, supplied by UbiQ) probe, purified to remove oligomers, to confirm whether published data could be replicated. Lane one shows a reaction using lysate at 20 mg/ml, lane two had lysate at 10 mg/ml, lane three at 5 mg/ml, and lane four at 1 mg/ml. Lane five also used lysate at 20 mg/ml, after treatment with 2 U of apyrase. Lane six is a sample of the probe, without lysate. All samples were run on the same 4-20% SDS-PAGE gel and scanned for Cy5 fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

3.3.2.4 A cascade probe based on the sequence of *Leishmania* ubiquitin

Of the 76 amino acid protein, two amino acids differ between *L. mexicana* ubiquitin and human ubiquitin – one conserved mutation, and one non-conserved. A cascade probe that was based on the sequence of *L. mexicana* ubiquitin (prior probes using the sequence of human ubiquitin) was provided by UbiQ. Furthermore, this probe contained two tags: TAMRA and biotin, permitting detection via fluorescent scanning and affinity purification. Various concentrations of this probe, with some samples shown before and after pulldowns, are shown in figure Figure 3-22, with fluorescent scanning in Figure 3-22A, and affinity purification in Figure 3-22B. To determine the optimal volume of streptavidin beads figure Figure 3-23 was formed. In both figures, it is important to note that the fluorescent signal, apparent before the biotin-streptavidin pulldown, is no longer detectable after the pulldown. Of the 56 proteins detected through mass spectrometry, none were deemed proteins of interest.

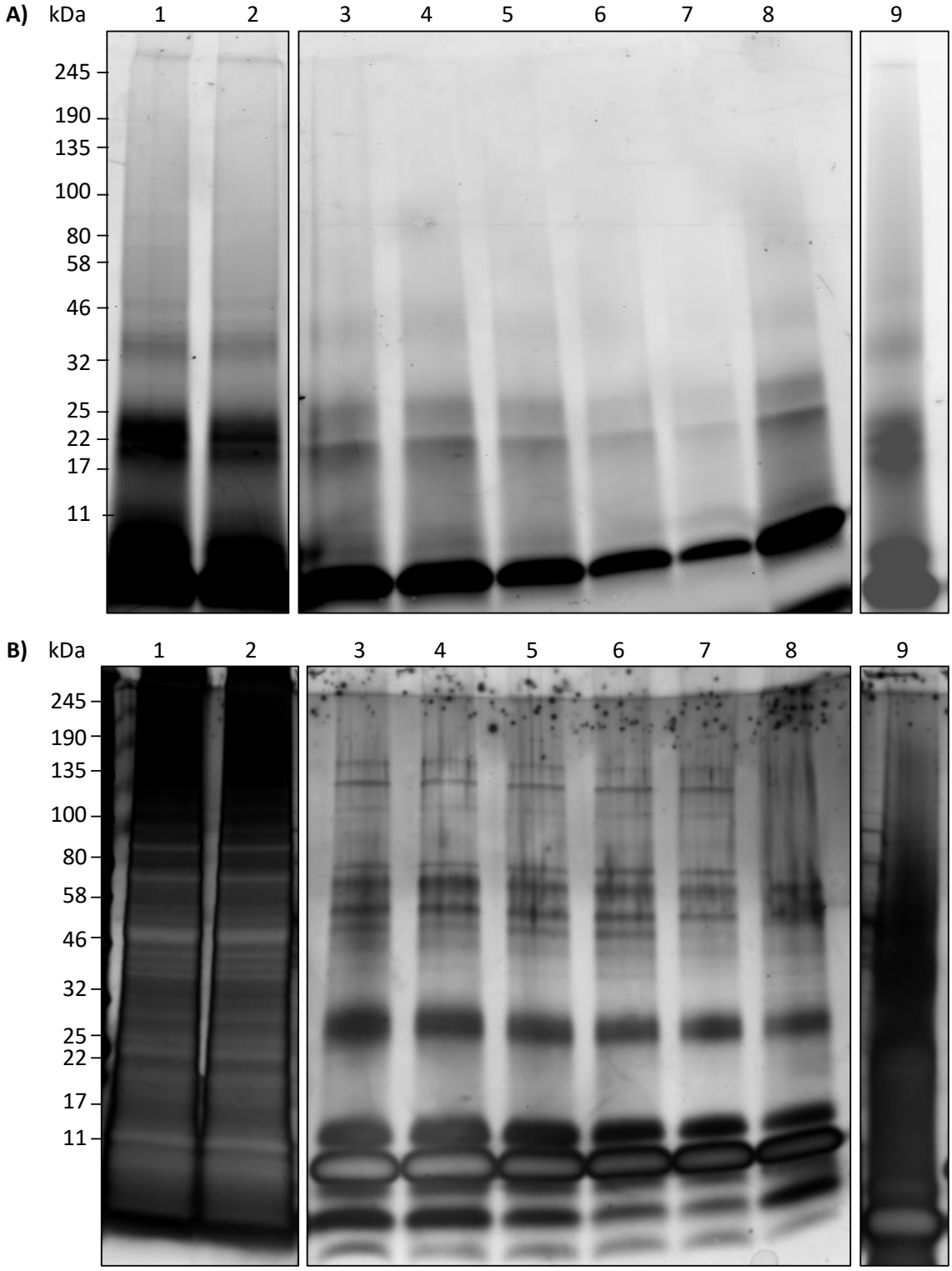


Figure 3-22 Various concentrations of an activity-based cascade probe, based on the sequence of *L. mexicana* ubiquitin with dual tags, reacting with *L. mexicana* lysate.
A TAMRA and biotin tagged cascade probe (product code UbiQ-246, supplied by UbiQ) was incubated with lysate from 7.5e7 *L. mexicana* promastigotes. **A)** TAMRA fluorescence scan **B)** The silver stain of the same gel. Lane 1 shows lysate is incubated with 40 µg/ml probe. Lane 2 shows lysate with 30 µg/ml of probe. Lanes 3,4,5,6,7, and 8 all show samples after a biotin-streptavidin pulldown. Lane 3 used probe at 40 µg/ml, lane 4 at 30 µg/ml, lane 5 at 20 µg/ml, lane 6 at 10 µg/ml, and lane 7 at 5 µg/ml. Lane 8 uses 40 µg/ml of probe to pulldown an iodoacetamide treated lysate. Lane 9 is solely the activity-based probe. All samples were run on the same 4-20% SDS-PAGE gel and scanned for TAMRA fluorescence before being silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

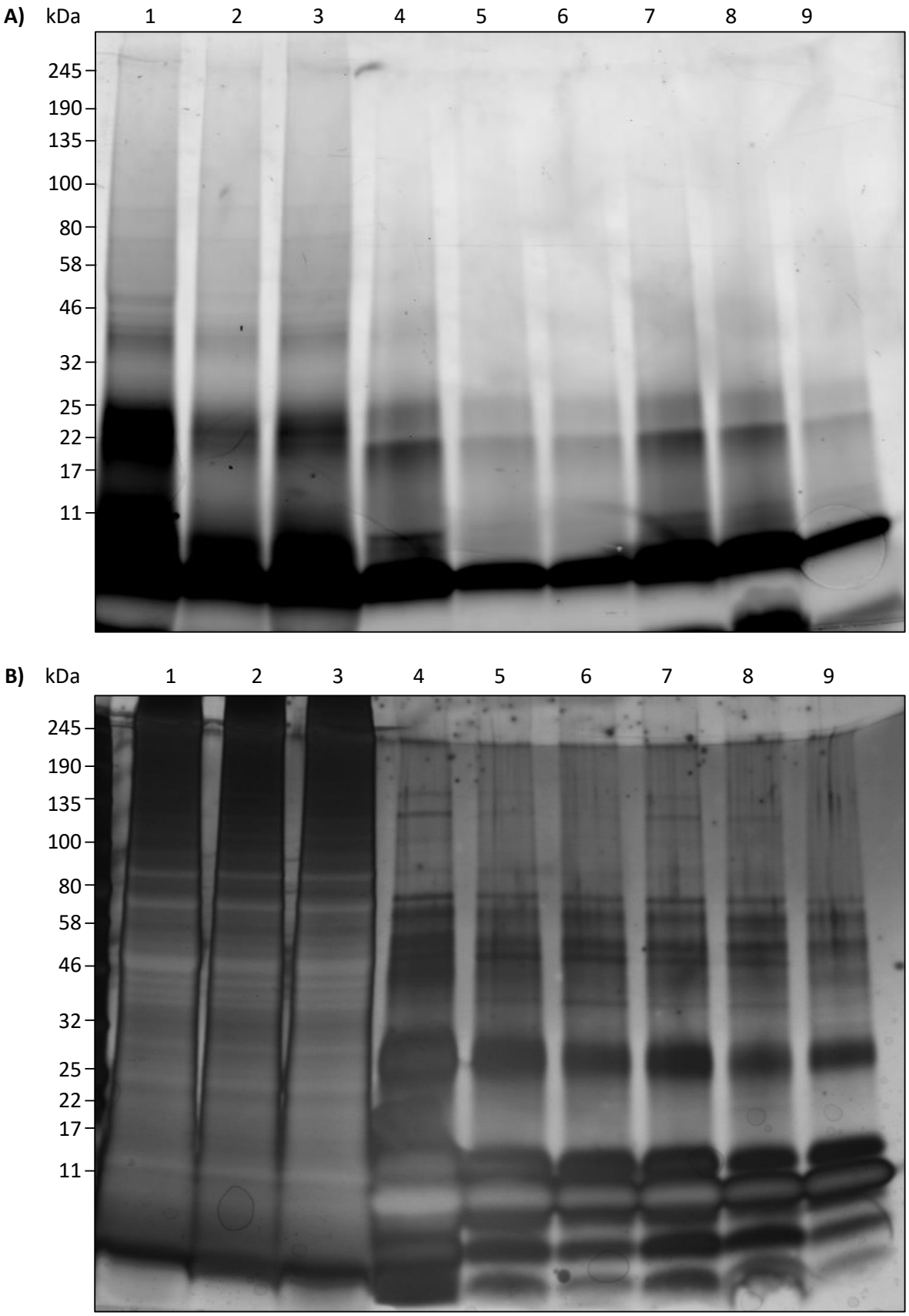


Figure 3-23 An activity-based cascade probe, based on the sequence of *L. mexicana* ubiquitin with dual tags, reacting with *L. mexicana* lysate with various elution conditions. A TAMRA and biotin tagged cascade probe (product code UbiQ-246, supplied by UbiQ) was incubated with lysate from 7.5e7 *L. mexicana* promastigotes. **A)** TAMRA fluorescence scan **B)** The silver stain of the same gel. Lane 1 shows lysate and probe mixture, before pulldown. Lane 2 shows left over supernatant, after pulldown with 75 µl of streptavidin beads. Lane 3 shows the leftover supernatant after 15 µl of beads are used for a pulldown. Lane 4 is the elution using 75 µl

beads. Lane 5 the elution from 60 µl beads Lane 6 the elution from 45 µl of beads. Lane 7 the elution from 30 µl of beads. Lane 8 the elution from 15 µl of beads. Lane 9 used 75 µl of beads to elute from an iodoacetamide treated lysate. All samples were run on the same 4-20% SDS-PAGE gel and scanned for TAMRA fluorescence before being silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

3.3.2.5 E1 probe with crosslinker

To identify E2s, whilst using the E1 specific probe, chemical crosslinkers were employed. Various concentrations of a non-cleavable crosslinker, DSS, are shown in Figure 3-24. These samples were sent for mass spectrometric analysis. Proteins identified in both the bioinformatic screening and mass spectrometry, but absent in negative control, are listed in Table 3-3. Of particular interest is a putative E2, LmxM.33.0900, which is present in all three experimental samples. Also interesting, is the appearance of a putative HECT E3, LmxM.07.0280, in all three experimental conditions. Table 3-3 does not feature the putative E1 LmxM.34.3060 – this was abundant in all three experimental samples, but a peptide of LmxM.34.3060 was identified in the negative control condition, and so LmxM.34.3060 was excluded from Table 3-3.

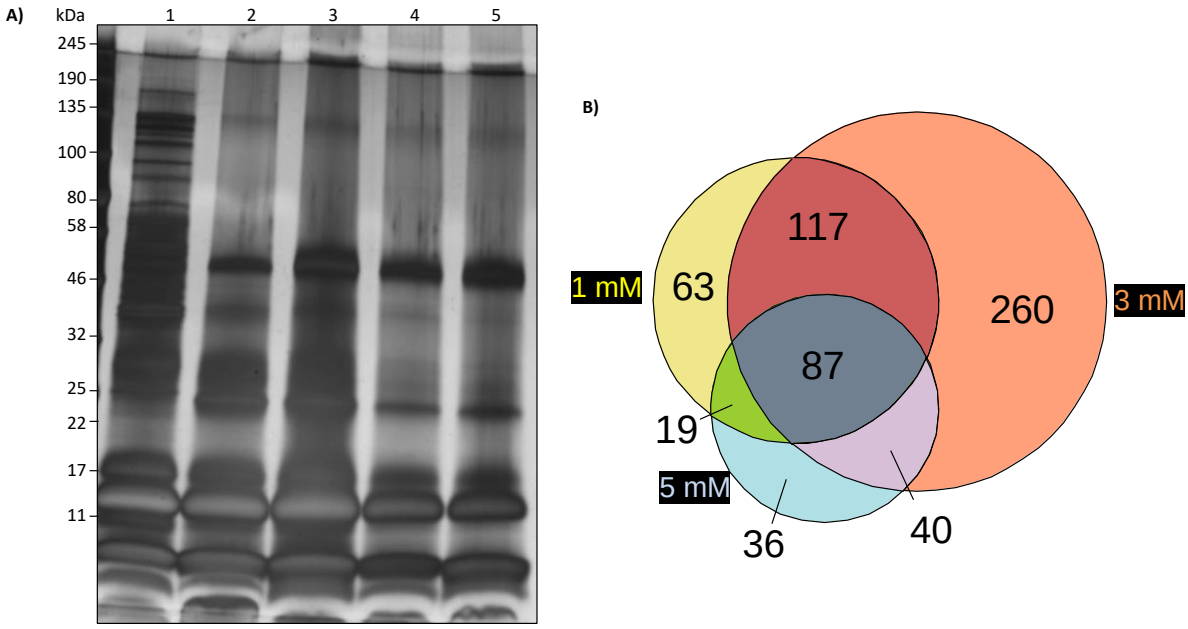


Figure 3-24 Optimising a non-cleavable crosslinker, DSS, with an activity-based probe specific for ubiquitin activating enzymes, in *L. mexicana* promastigote lysate. A) 34.5 µg/ml of E1 probe (product code UbiQ-221, supplied by UbiQ) was incubated with lysate from *L. mexicana* promastigotes. Lane 1 contains a reaction without DSS. This is compared to lane two with 1 mM, lane three with 3mM, and lane four with 5 mM of DSS. Lane 5 has an unreactive probe with lysate. All samples were run on the same 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** A Venn diagram showing the distribution of mass spectrometric protein identifications, in samples using 1, 3, and 5 mM of DSS, with samples also identified in the unreactive probe condition removed.

	Gene accession no.	Present in samples (mM of DSS)
E1	LmxM.23.0550	1, 3, 5
E2	LmxM.07.0850	3
	LmxM.33.0900	1, 3, 5
E3 (HECT)	LmxM.07.0280	1, 3, 5

Table 3-3 Summary of mass spectrometric results obtained when using DSS, and an E1 specific activity-based probe to identify components of *L. mexicana*'s ubiquitination machinery. The listed proteins appeared in the bioinformatic screening, but did not appear in a negative control (a chemically inactive activity-based probe incubated with 5mM DSS).

A cleavable crosslinker, DTSSP, is shown in figure Figure 3-25. A key distinction between the non-cleavable and cleavable crosslinkers are the number of bands, with DTSSP Figure 3-25 having more. However, the presence of distinct bands around 245 kDa are only seen with DSS in figure Figure 3-24. The number of proteins that were identified in the experimental samples, but not present in the negative control, are shown in figure Figure 3-25B. Other than the putative E1s already identified using this probe, crosslinking with DTSSP provided no further proteins of interest. This experiment was repeated, washing with PBS as opposed to RIPA buffer (in case RIPA buffer caused premature cleavage of DTSSP), and seeding the reaction with 350 pmol of recombinant LmxM.34.3060 (see 5.3.1). This did not improve on the initial DTSSP results, also yielding no additional proteins of interest.

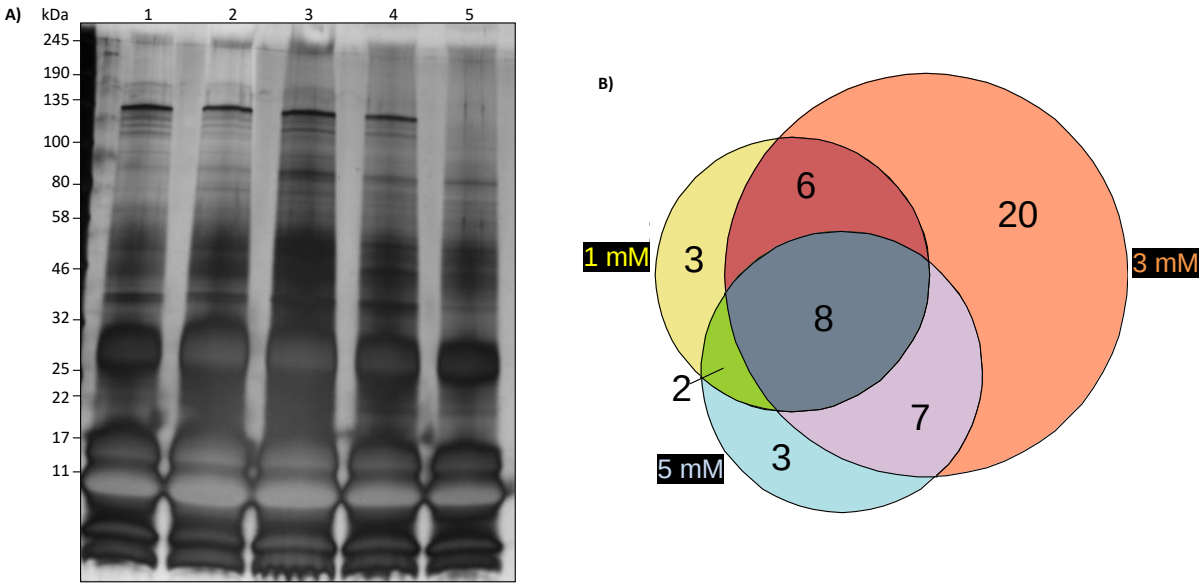


Figure 3-25 Optimising a cleavable crosslinker, DTSSP, with an activity-based probe specific for ubiquitin activating enzymes, in *L. mexicana* promastigote lysate. A) 34.5 µg/ml of E1 probe product code UbiQ-221, supplied by UbiQ) was incubated with lysate from 1.5e8 *L. mexicana* promastigotes. Lane 1 contains a reaction without DTSSP. This is compared to lane two with 1 mM, lane three with 3mM, and lane four with 5 mM of DTSSP. Lane 5 has an unreactive probe with lysate. All samples were run on a 4-20% SDS-PAGE gel and silver stained. Molecular weight markers had product code #P7712, and were supplied by NEB. **B)** A Venn diagram showing the distribution of mass spectrometric protein identifications, in samples using 1, 3, and 5 mM of DTSSP, with proteins also identified in the iodoacetamide condition removed.

3.3.2.6 A SUMO cascade probe with *Leishmania* lysate

Given the existence of an activity-based probe based on the sequence of SUMO2, SUMOylation was also investigated using similar principles. Figure 3-26 shows a various concentrations of a fluorescently labelled probe, based on the sequence of human SUMO2. Faint bands are denoted by asterisks.

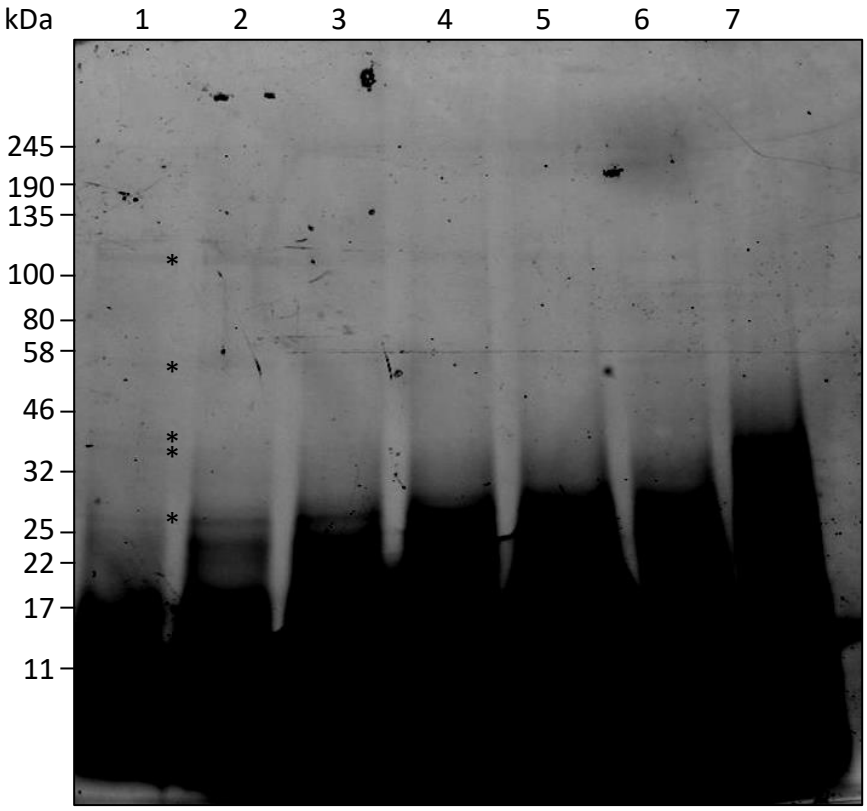


Figure 3-26 Rhodamine fluorescence scan of various concentrations of an activity-based SUMO1 probe reacting with *Leishmania mexicana* lysate. Various concentrations of SUMO1 probe (product code UbiQ-130, supplied by Ubiq) were incubated with lysate from 7.5e7 *L. mexicana* promastigotes, in order to identify SUMOylation machinery in the parasite. Lane one shows 5 µg/ml of probe interacting with lysate, lane two shows 10 µg/ml, lane three 20 µg/ml, lane four 30 µg/ml, and lane five 40 µg/ml. Lane six also shows 40 µg/ml of probe, however the lysate was alkylated with iodoacetamide. Lane seven displays the fluorescence from 40 µg/ml of probe without lysate present. Asterisks are used to draw attention to otherwise subtle bands in lane 1. All samples were run on the same 4-20% SDS-PAGE gel and scanned for rhodamine fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

A following experiment varied the cell number, with a proportional increase in the amount of probe used. The results of this are shown in Figure 3-27. The banding pattern is more distinct as more cells are used, more than were necessary when studying ubiquitination.

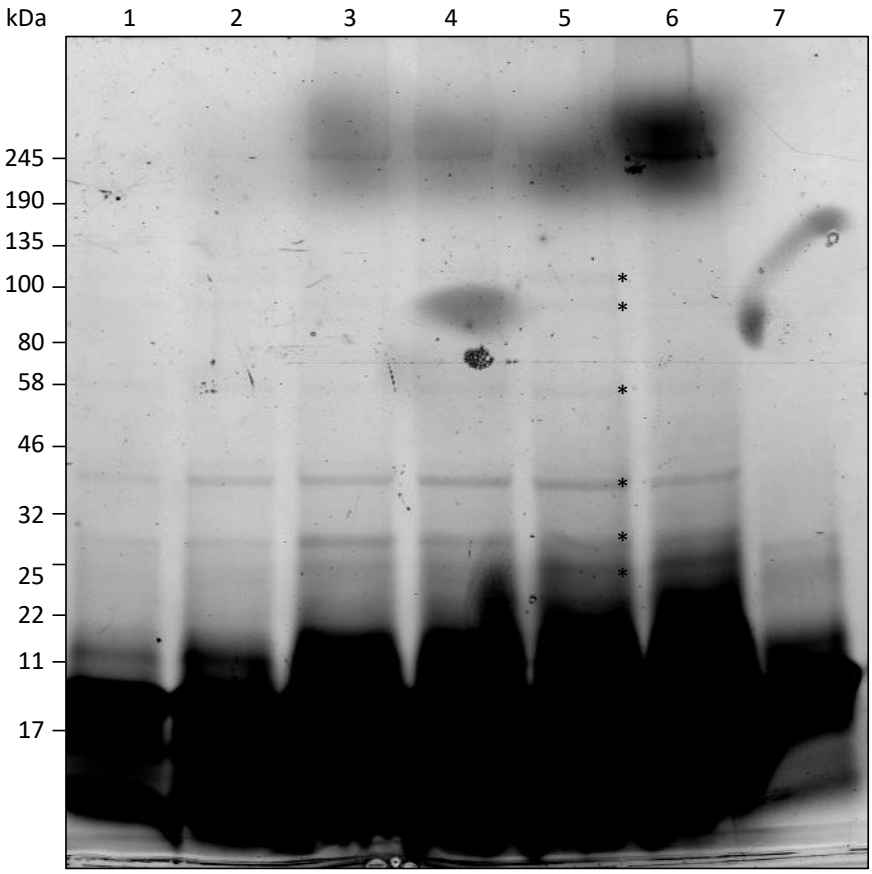


Figure 3-27 Rhodamine fluorescence scan of various concentrations of an activity-based SUMO1 probe reacting with *Leishmania mexicana* lysate. Various concentrations of (product code UbiQ-130, supplied by Ubiq) probe and *L. mexicana* promastigote lysate were incubated together, in order to identify SUMOylation machinery in the parasite. Lane1: 7.5e7 cells with 0.5 µg of probe (1 × reaction). Lane 2: 1.875e8 cells with 1.25 µg probe (2.5 × reaction). Lane 3: 3.75e8 cells with 2.5 µg probe (5 × reaction). Lane 4: 5.625e8 with 3.75 µg probe (7.5 × reaction). Lane 5: 7.5e8 cells with 5 µg probe (10 × reaction). Lane 6: Iodoacetamide treated lysate in a 10 × reaction. Lane 7: the probe without lysate. Asterisks are used to draw attention to otherwise subtle bands in lane 5. All samples were run on the same 4-20% SDS-PAGE gel and scanned for rhodamine fluorescence. Molecular weight markers had product code #P7712, and were supplied by NEB.

A biotin streptavidin pulldown was conducted to enrich the SUMO machinery, with this fraction undergoing mass spectrometry. However, a biotin labelled probe was only available with the SUMO2 sequence, as opposed for the SUMO1-based probe that was used for fluorescent labelling. The results are shown as a Venn diagram in Figure 3-28. None of the proteins identified were proteins of

interest, as indicated in the bioinformatic screening.

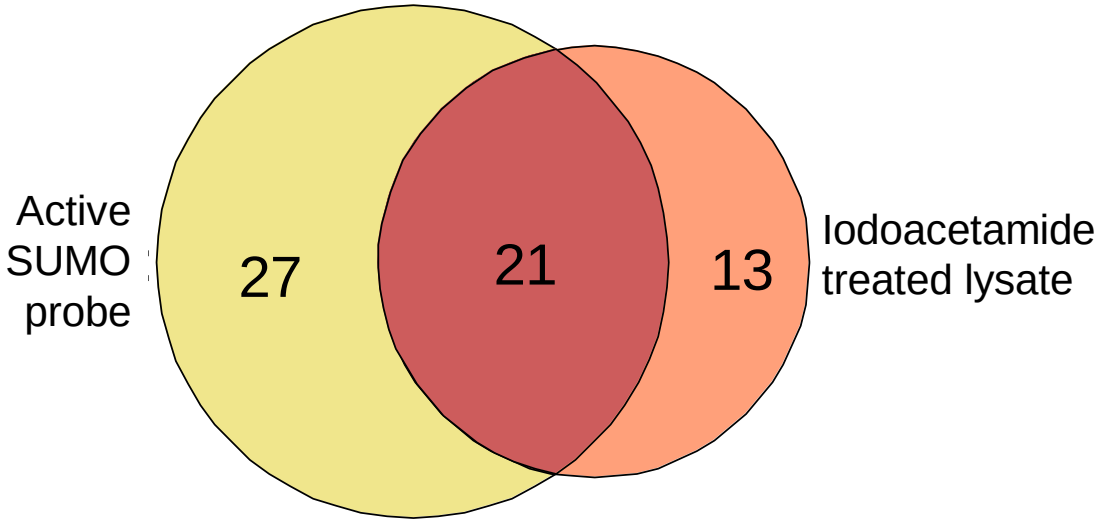


Figure 3-28 Venn diagram of the protein identifications obtained using a biotin tagged activity-based probe, targeting *Leishmania*'s SUMOylation machinery.

3.4 Discussion

3.4.1 Ubiquitin activating enzymes

Bioinformatic analyses suggest that a ubiquitination system likely exists in *L. mexicana*, with some components of these pathways being labelled using activity-based probes that allowed selection by affinity chromatography and subsequent identification through proteomic approaches. This study predominantly details ubiquitin activating enzymes, which are identified as LmxM.23.0550 and LmxM.34.3060. Although these genes have been identified before, the bioinformatic methods employed here do not rely on domain searches, and so do not rely on prior annotation of the *Leishmania* genome (Burge *et al.*, 2020). This is also the first time these E1s have been characterised at a protein level, with their relative abundance noted. The idea that there are two E1s is itself interesting – single celled organisms typically suffice with one E1, as is the case with *Schizosaccharomyces pombe* (Olsen and Lima, 2013) and *Saccharomyces cerevisiae* (McGrath, Jentsch and Varshavsky, 1991). In fact, some multicellular organisms only have a single E1 enzyme, such as *C. elegans* (Jones *et al.*, 2002) and *Drosophila* (Watts, Hoopfer and Luo, 2003). The use of two E1s is usually limited to vertebrates, from humans to zebrafish (although sea urchins provide an interesting exception) (Jin *et al.*, 2007). So, the question naturally arises: why does the protozoan organism *Leishmania* need two E1s? The related organism *Trypanosoma brucei* also has two E1 genes, which are each syntenic orthologs of the putative *Leishmania* E1 genes (Boer and Bijlmakers, 2019). The percentage identity matrix in Table 3-4 compares the two putative *Leishmania* E1s with two human orthologs. It is interesting that the putative *Leishmania* E1s share 29.46% identity, whereas the two human E1s share 42.41%. It is difficult to determine simple orthologous relationships between these two species as both *Leishmania* enzymes share greater similarity to UBA1 than UBA6. Nonetheless, in the human ubiquitination system UBA1 is widely considered the work horse of two ubiquitin activating enzymes. Its ortholog is essential in yeast (McGrath, Jentsch and Varshavsky, 1991), and RNAi with UBA1 stops embryogenesis in *C. elegans* before killing adult worms, the severest phenotype of all the E1s tested (Jones *et al.*, 2002). Furthermore, at non-permissive temperatures, a temperature sensitive UBA1 mutant mammalian cell line displays a dramatic decrease in ubiquitination, arresting the cell at the G2-M

phase transition (Ciechanover, Finley and Varshavsky, 1984). UBA1 knockdown also eliminates charging of the cell cycle E2s CDC34 and CDC34B, (Jin *et al.*, 2007). Yet UBA6 is not entirely redundant, as it is required for charging the E2 Use1, a job that UBA1 is unable to perform (Jin *et al.*, 2007). So, if a percentage identity difference of 57.59 % results in distinct specificities in the human system, it is plausible that a difference of 70.54% could mean differentiation in the *Leishmania* system. Does *L. mexicana* have an ortholog of Use1? Using Use1 as bait, a BLAST returns LmxM.34.1300 as the top scoring hit, a gene which was also returned in the initial bioinformatic screening campaign, with an expectation value of 9e-18. However, it is far from the only promising result, with the next two top hits having e values of 5e-16 and 7e-16, so drawing simple parallels between species would be unwise. Another role for UBA6 is charging the Ubl FAT10 (Chiu, Sun and Chen, 2007) – could one of the putative *Leishmania* E1s similarly charge a FAT10 ortholog? If there is a FAT10 ortholog in *L. mexicana* it is not immediately obvious, as a BLAST search with the human FAT10 sequence yields the same *L. mexicana* genes as a BLAST search using the human ubiquitin sequence, but the FAT10 hits have substantially lower scores. This is supported by the work of Karpiyevich *et al.*, who state that, as of 2020, FAT10 has yet to be found in any parasitic protozoan (Karpiyevich and Artavanis-Tsakonas, 2020). Additionally, FAT10 appears to function in response to cancer stimuli, so it is unsurprising that it has yet to be found in a unicellular organism (Ponder and Bogyo, 2007). An alternative explanation for the existence of two E1s would be gene duplication. This is most likely the case in plants, whose two or three different E1s are 90% identical to each other, and transfer ubiquitin with similar efficiency to several E2s (Hatfield *et al.*, 1997). It is worth noting that the *Leishmania* genome is plastic, presenting a plausible mechanism for gene duplication (Sterkers *et al.*, 2012). However, the two *Leishmania* E1s are quite dissimilar in terms of amino acid identity, and there is at least one *Leishmania* E2 that is preferentially charged by LmxM.23.0550 (Burge *et al.*, 2020). One scenario would be that a gene duplication event occurred in an ancestral organism of *Trypanosoma* and *Leishmania*, an event entirely unrelated to the dual E1s in vertebrates. Once this duplication event had taken place specificity for each E1 could begin to evolve. Support for this theory is provided by the phylogenetic tree shown in Figure 3-29. Here LmxM.23.0550 is shown to be as distantly related to LmxM.34.3060 as it is to orthologous human proteins,

implying that LmxM.23.0550 has an evolutionary origin in the deep past to have allowed such divergence, an origin that is unconnected to the development of a second E1 in vertebrates.

	LmxM.23.0550	LmxM.34.3060	UBA1 (*H.s.*)	UBA6 (*H.s.*)
LmxM.23.0550	-	29.46	38.49	35.20
LmxM.34.3060	29.46	-	33.03	29.71
UBA1 (H.s.)	38.49	33.03	-	42.41
UBA6 (H.s.)	35.20	29.71	42.41	-

Table 3-4 A percentage identity matrix for published ubiquitin E1 structures, and the *L. mexicana* E1 characterised in this study. H.s. stands for Homo sapiens.

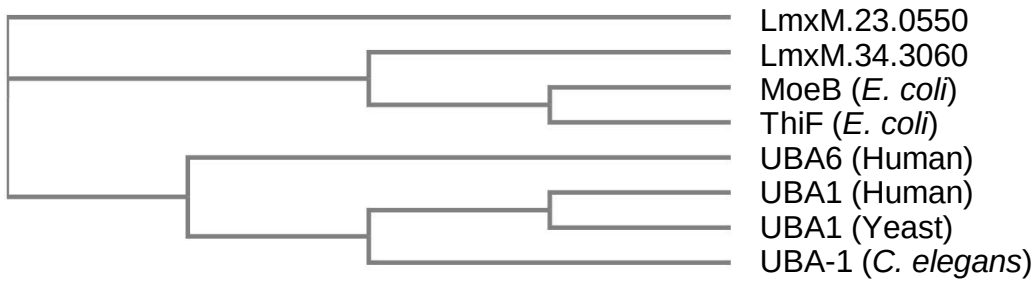


Figure 3-29 Phylogenetic tree showing the relationship between E1s. The inferred evolutionary relationship of the two *Leishmania* ubiquitin activating enzymes (LmxM.23.0550 and LmxM.34.3060) with well-characterised orthologs.

If abundance can be used as a crude proxy for importance (a more abundant E1 may be expected to charge more E2s, resulting in more ubiquitination than its less abundant counterpart), then LmxM.34.3060 could be the more promising drug target of the two putative *Leishmania* E1s. It is interesting that the *T. brucei* orthologs of LmxM.23.0550 and LmxM.34.3060, TbUBA1a and TbUBA1b respectively, both displayed severe growth defects in a large-scale RNAi study. Of the ~10,000 genes studied TbUBA1b was in the top 1% of genes with the most impact on *T. brucei* fitness, and TbUBA1a in the top 12% (Alsford *et al.*, 2011). In *L. mexicana*, null mutants of LmxM.34.3060 had a loss of fitness phenotype when promastigotes were transformed into axenic amastigotes, and when promastigotes underwent macrophage infection. On mouse infection LmxM.34.3060 null mutants were undetectable in mouse footpads three weeks post-infection (the earliest time point collected after infection). That being said, these experiments were only conducted because null mutants could be generated for LmxM.34.3060, whereas they could not be generated for LmxM.23.0550 (Burge *et al.*, 2020). Nonetheless, the *Leishmania major* orthologs of both LmxM.34.3060 and LmxM.23.0550 are listed in the Tropical

Disease Research drug target database (Urán Landaburu *et al.*, 2020). Still, as LmxM.34.3060 shares fewer amino acids with common identities to the human orthologs, and so could have a dissimilar structure, it may be less challenging to design a *Leishmania* selective LmxM.34.3060 inhibitor than a selective LmxM.23.0550 inhibitor. Other authors have also reached this conclusion (Boer and Bijlmakers, 2019).

In all mass spectrometry data where the N-terminal peptide of either LmxM.34.3060 or LmxM.23.0550 is present it is consistently acetylated. The act of joining an acetyl group (CH_3CO) to the amino group (NH_3^+) N-terminal end of a polypeptide neutralises the otherwise positive charge of the amino group, changing the electrostatic properties of the protein (Varland, Osberg and Arnesen, 2015). This alteration of a protein's electrostatic properties can affect a range of attributes, such as stability, folding, protein-protein interactions, and subcellular targeting (Ree, Varland and Arnesen, 2018). Perhaps it is because of this range of roles that 80 - 90% of the soluble human proteome, and 50 – 70% of the yeast proteome, are acetylated at their N-termini (Varland, Osberg and Arnesen, 2015). To the best of the author's knowledge, the abundance of N-terminal acetylation in the *L. mexicana* proteome is unknown. However, during the differentiation of *Leishmania donovani*, one study listed 26 acetylated proteins, of which three were part of the ubiquitination-proteasome system (Rosenzweig *et al.*, 2008). Elsewhere, 144 N-terminally acetylated proteins were identified in the *Leishmania infantum* proteome, among which was the *L. infantum* ortholog to LmxM.23.0550 (the ortholog to LmxM.34.3060 was not listed) (Sanchiz *et al.*, 2020). The authors state that, when N-terminal acetyltransferase B is putatively thought to be responsible for acetylation, then 17.5% of those proteins have glutamic acid as their second amino acid (after methionine). Glutamic acid is the second amino acid in LmxM.34.3060. When N-terminal acetyltransferase C is putatively thought to be responsible then the amino acid following methionine is leucine 5% of the time. This is the case with LmxM.23.0550. In this dataset, the consistency of the modification, appearing, without fail, whenever the N-terminal peptide of either putative E1 was identified, gives reason to believe N-terminal acetylation may be integral to the functioning of a *L. mexicana* ubiquitin activating enzyme. Whatever its role may

be, this is evidence of crosstalk between post-translational modifications – acetylation impacting ubiquitination.

3.4.2 Method development

It is clear that only a fraction of the activity-based probe experiments yielded the expected results – namely, those tailored to interact with an E1 (results summarised in **Error! Reference source not found.**). Why could this be? An identical protocol, applied to chemical probes that operate on similar principles, generates wildly different results. It is notable that when using a fluorescent probe based on the sequence of human ubiquitin there is prominent banding below 25 kDa irrespective of the presence of lysate. These bands appear at regular intervals, approximately 9 kDa apart, which is the weight of the activity-based probe. Perhaps these bands represent the activity-based probe as a monomer, dimer, and trimer. These bands are also present irrespective of whether the probe underwent further rounds of purification before use. As such, the chosen remedy was to run the anomalous bands off the polyacrylamide gel, so that this noise did not overwhelm any other signal. However, any other signal achieved by fluorescent labelling was largely dependent on the fluorescent label employed – Cy5 labelled probes generating an entirely distinct banding pattern to rhodamine labelled probes. Needless to say, this did not instil confidence in the data, and use of fluorescently labelled probes was soon abandoned. Use of the biotin labelled cascade probe did not fare much better, failing to identify a single protein that was suggested by the bioinformatic screening. While the bioinformatic screen might have overlooked very divergent components of the *Leishmania* ubiquitination pathway, both BLAST and HMM searches yielded the same putative genes, of which E1s were validated by an activity-based probe. Broadly, with some exceptions, mass spectrometry results that provided the expected identifications (i.e. those from the E1 probe) obtained more protein identifications from the experimental condition than the negative control, which was not the case when cascade probe was used. Despite the fact that *Trypanosoma cruzi* ubiquitin is immunologically distinct from its human counterpart, only differing by three amino acids, it seems unlikely that the two amino acid difference between human and *Leishmania* ubiquitin is to blame for the failure of the probe (Télles *et al.*, 1999). This is because an E1 specific probe, also based on the sequence of human ubiquitin worked fine, and a probe

based on the sequence of *L. mexicana* ubiquitin did not. Although this dual labelled *Leishmania* probe may help explain the lack of mass spectrometric identifications – its fluorescent signal disappears after the biotin-streptavidin pulldown, indicating that any enzymes labelled in the lysate did not interact with the streptavidin beads, or were lost during the wash steps. Given the abundance of ubiquitin in the mass spectrometry results, a likely explanation would be that enzyme-probe bond was broken during the wash step, however this should not have happened if the covalent bond stated in the literature was present (Mulder *et al.*, 2016). Of course, this assumes that the pre-pulldown fluorescent signal in the lysate is valid, which cannot be taken for granted, and if the pre-pulldown fluorescent signal is misleading it may be that the activity-based probe never bound to the enzymes in the first place. The multiple issues with the activity-based probe make it difficult to determine the root cause of the failure. However, a key difference between the E1 specific probe, which worked, and the cascade probe, which did not, is the presence of an electrophilic AMP analog in the E1 specific probe. This enables the E1 specific probe to examine the first half of the E1 chemical reactions, the adenylation reaction, and binds the activity-based probe to the E1 adenylation site (Lu *et al.*, 2010). In contrast, the cascade probe uses dehydroalanine to bind to the catalytic cysteine of ubiquitination enzymes, theoretically causing a thioether linkage between the probe and any ubiquitination enzyme with a cysteine active site (Mulder *et al.*, 2016). It is unknown whether something was amiss with the synthesis of the cascade probe, or the theory of using dehydroalanine to form a covalent bond is incorrect, but only one mechanism of activity-based probes obtained results in this study – using a ubiquitin-adenylate mimic to probe the E1 adenylation site.

Probe number	Cascade or E1	Protein	Label	Additional notes
UbiQ-102	Cascade	Human ubiquitin	Biotin	Did not obtain any proteins of interest
UbiQ-104	Cascade	Human ubiquitin	Cy5	Fluorescent labelling distinct from rhodamine counterpart
UbiQ-130	Cascade	Human SUMO1	Rhodamine	Fluorescent labelling achieved
UbiQ-131	Cascade	Human ubiquitin	Rhodamine	Fluorescent labelling distinct from Cy5 counterpart
UbiQ-159	Cascade	Human SUMO2	Biotin	Did not obtain any proteins of interest
UbiQ-221	E1	Human ubiquitin	Biotin	Obtained the E1s predicted from bioinformatic screening
UbiQ-246	Cascade	*L. mexicana* ubiquitin	TAMRA and biotin	Did not obtain any proteins of interest, fluorescent signal lost after pull down

Table 3-5 The activity-based probes used in this study. Summary of the activity-based probes used to examine Leishmania mexicana's ubiquitination system. Cascade refers to the probe's potential to interact with the entire cascade of ubiquitination enzymes – E1s, E2s, and HECT/RBR E3s.

As for method optimisation, the method improvements employed in subsequent iterations of the E1 probe methods employed were successful in driving down the number of identifications of proteins unrelated to ubiquitination. However, a lower signal to noise ratio did not lead to increased identifications, implying that the signal to noise was not a problem. Subsequent iterations of the crosslinking experiment used a cleavable crosslinker, DTSSP, as it was thought the more predictable peptides (relative to a non-cleavable crosslinker) would lead to increased identifications of ubiquitination components. This was not the case, with MS results looking similar to results without any crosslinker use. The implication of this is that the crosslinker was cleaved prematurely, perhaps during the washing stages, or never crosslinked the lysate at all. It is unknown what could have caused premature cleavage, as cleavage should be dependent on a reducing agaent and PBS was used in wash steps. Alternatively, why DTSSP would fail to crosslink the lysate is unknown – its mechanism, reacting with lysine residues, is identical to the successful DSS crosslinker, and DTSSP was solubilised according to the manufacturer's instructions.

3.5 Summary

This study has identified two putative *Leishmania mexicana* ubiquitin activating enzymes: LmxM.23.0550 and LmxM.34.3060. These enzymes were first identified in a bioinformatic screening, then validated using an activity-based probe. The activity-based probe indicates they are expressed during both the promastigote and axenic amastigote life cycle stages. The mass spectrometry data obtained suggests that LmxM.34.3060 is considerably more highly expressed than LmxM.23.0550.

This study has identified an E2: LmxM.33.0900. First identified in a bioinformatic screening, and repeatedly identified through use of an activity-based probe and a chemical crosslinker.

Future work should focus on identifying the other E2s and E3s that likely exist in *L. mexicana*, whether they belong to a ubiquitin or a ubiquitin-like pathway.

Chapter 4 Biochemical characterisation of the ubiquitination system

4.1 Introduction

The previous chapter identified components of the ubiquitination pathway using bioinformatics and activity-based probes. This chapter will further characterise the ubiquitination system at a biochemical level, identifying interacting proteins and substrates of ubiquitin. To identify a wider interactome, fusion proteins will be formed between proteins of interest and a promiscuous biotin ligase. These fusion proteins will be formed using CRISPR. To identify ubiquitinated proteins, immunoprecipitations will be conducted, precipitating tryptic peptides that contain a ubiquitination motif. The data will be used to validate the results of the activity-based probes with chemical crosslinkers, and screen E1-E2-substrate cascades.

4.1.1 CRISPR

Has there been any revolution in biology that has swept the field more swiftly than CRISPR? If there has been, the author is unaware of it. It was only in 2012 that Jinek *et al*. published the concept of repurposing a prokaryotic adaptive immune system into a genome editing tool (Jinek *et al.*, 2012). At this point, the papers mentioning CRISPR totalled 127. As of August 2019, seven years later, they amount to over 14,000 (Cohen and Desai, 2019). This could be why the 2020 Nobel Prize in Chemistry deservedly went to Jennifer Doudna and Emmanuelle Charpentier, both of whom pioneered this innovative technology (Ledford and Callaway, 2020). The first discovery of the Clustered Regularly InterSpaced Palindromic Repeats from which CRISPR get its name were thanks to Francisco Mojica, who noted curious structures of multiple copies of roughly palindromic repeated sequences of 30 base pairs, separated by spacers of roughly 36 base pairs that did not resemble any known family of microbial repeats (Mojica, Juez and Rodriguez-Valera, 1993). Although similar structures (with no sequence similarity) had been mentioned earlier in the literature (Ishino *et al.*, 1987), it was Mojica who first noted their significance. With these mysterious structures described, researchers set about cataloguing the key features of the loci – importantly, the presence of specific CRISPR-associated (Cas) genes in the immediate vicinity of these structures (Jansen *et al.*, 2002). Nonetheless, labelling the system does not explain it. Perhaps the first experimental evidence for the purpose of CRISPR came from food science. Philippe Hovarth was

interested in overcoming the frequent bacteriophage infections that plagued the industrial cultures of bacteria used in the dairy industry, bacteria such as *Streptococcus thermophilus* that provide us with yoghurt and cheese. So Hovarth and others employed a bacteriophage sensitive strain of *S. thermophilus* and two bacteriophages, using genetic selections to isolate phage-resistant bacteria. The resistant strains had acquired phage-derived sequences at their CRISPR loci. The experimenters had seen acquired immunity in action (Barrangou *et al.*, 2007).

Acknowledging that CRISPR is essentially a programmable restriction enzyme, and employing it *in vitro* for genetic modifications on demand were Doudna and Charpentier (Jinek *et al.*, 2012). Using recombinant Cas9 protein from *Streptococcus pyogenes* and custom designed guide RNA strands, they showed that Cas9 could cut purified DNA *in vitro*. Furthermore, they showed that the various RNA components used *in vivo* by *S. pyogenes* could be fused into one single guide RNA (sgRNA). This sgRNA comprises of the spacer sequence (which, *in vivo*, is the viral DNA to be cleaved) and the repeated sequence (which forms a hairpin structure, necessary to interact with the Cas9 protein). So the CRISPR-Cas9 system could be induced to inflict a double stranded break in the DNA sequence of choice. This break can then be repaired by either nonhomologous end joining, an error prone mechanism, or by homology directed repair, which has higher fidelity (Jiang and Doudna, 2017). Provided with enough donor DNA to use as a template, the incidence of homology directed repair can be increased. Should this template DNA, with ends homologous to either side of the double strand break, possess an insert sequence (such as an antibiotic resistance gene and a tag) then gene editing can occur (Sansbury, Hewes and Kmiec, 2019).

Fortunately, with the ongoing CRISPR revolution comes increasing applications of the technology. The system has now been established in pathogenic organisms such as *Plasmodium falciparum* (Ghorbal *et al.*, 2014) and *Cryptosporidium parvum* (Vinayak *et al.*, 2015), as well as traditionally neglected organisms such as *Toxoplasma gondii* (Sidik *et al.*, 2016) and *Leishmania mexicana* (Beneke *et al.*, 2017). In fact, the system has become so user friendly that modular plasmid systems exist, where each plasmid serves as a template for PCR-amplification of a drug-selectable repair cassette, with selectivity determined by the primers designed with automated online software (Beneke *et al.*, 2017; Beneke and

Gluenz, 2019). This once-obscure microbial system has gone from a niche object of curiosity in basic research (revealed through the bioinformatic screening of bacterial genomes), to industrial applications (to improve yoghurt production), to the focus of special issues of scientific journals, headlines in the *New York Times*, and the subject of international ethics summits (Lander, 2016). Perhaps this story is instructive – that scientific breakthroughs are ensemble acts, that less popular topics are worthy of study, and that medical breakthroughs can come from unpredictable origins.

4.1.2 BioID

As stated previously (see 3.1.2), proteins do not work in isolation, but in information processing networks. To build on this, information processing requires organisation, as a system can only be perturbed if there is a system to start with. Collisions in a random maelstrom of movement have no value, merely exchanging one indiscriminate disorder for another indiscriminate disorder. Proteins, and any inherent information they may contain, may be organised in two ways – spatially and temporally. As such, this has given rise to spatial and temporal proteomics, of which spatial proteomics is of primary concern to this project.

Two proteins expressed at the same time can only interact if they are in the same space. Yet a cell is subdivided into numerous organelles, complete with elaborate transport mechanisms, all to control the movement of proteins (Scott and Pawson, 2009). Hence the spatial organisation of the proteome is integral to understanding the cell. To address this need, several techniques have been developed. There is the yeast two-hybrid system (Brückner *et al.*, 2009), affinity purification (Xie *et al.*, 2009), crosslinking mass spectrometry (Holding, 2015), and luminescence-based interactome mapping (Blasche and Koegl, 2013) to name only a few. However, some of these techniques struggle to identify weak, transient interactions, or detect protein-protein interactions outside of the a protein's natural cellular context (as is the case when lysis is required before screening for an interactome) (Snider *et al.*, 2015). A method that is able to detect weak protein-protein interactions *in vivo* is proximity-dependent biotin identification, or BioID (Roux *et al.*, 2012). This uses *Escherichia coli*'s biotin ligase BirA to biotinylate proteins. In the presence of ATP and biotin BirA will

generate the reactive species biotinoyl-5'-AMP, hold it at its reactive site, then facilitate its transfer to the amine groups of exposed lysines on target proteins (Roux *et al.*, 2012). To decrease the affinity of BirA for biotinoyl-5'-AMP the mutant protein BirA* was formed, where the wild type's arginine 118 was mutated to glycine. This mutation permits the biotinylation of any protein within a 10 nm radius of BirA* (Kim *et al.*, 2014). When this BirA* is fused to a protein of interest it will biotinylate the interactome of the protein of interest, or at least those proteins that have an exposed lysine and come within 10 nm of the protein of interest-BirA* fusion protein. The use of biotin for labelling presents another benefit to the method – the use of stringent lysis conditions and harsh wash buffers. However, it is not without its drawbacks. Namely, a long labelling time of 24 hours. This means the technique may not capture those interactions that have fast kinetics and cannot determine the chronology of any interactions. Furthermore, the 32 kDa BirA* may modify the structure or activity of its fusion partner. To address the reaction rate shortcoming Branon *et al.* used directed evolution to produce two new biotin ligases: the 35 kDa TurboID, and the 28 kDa miniTurbo (Branon *et al.*, 2018). Both enzymes reduce the tagging time to a few minutes, permitting snapshots of an interactome to be captured. Given these advantages, BioID stands out as a first rate approach for analysing interaction networks. In the context of a living cell, it allows experimenters to see what and when proteins interact, it allows them to organise the proteome.

4.1.3 The DNA damage response and anti-diglycine immunoprecipitations

Having acknowledged the need to consider the cell as an information processing network (see 3.1.2), having acknowledged the need to consider the spatial and temporal organisation of the proteome (above), the need to measure the ebb and flow of a signalling cascade must now be acknowledged. To respond to cues about their internal and external environments proteins must interact with one another, and this can involve covalently modifying one another, leading to activation or deactivation, perhaps a change in localisation or interactors. Mass spectrometry is uniquely well placed to analyse these fluctuations in protein modifications. Given a stimulus, is a protein modified? One way to determine whether a protein has been modified with ubiquitin is to look for a mass shift in tryptic peptides, specifically a 114.04 Da mass increase on lysines (Xu, Paige and

Jaffrey, 2010). This 114.04 Da mass increase corresponds to the weight of two glycines, the glycines that comprise the penultimate amino acids of ubiquitin. These are conjoined to the lysine of a substrate protein, remnants of ubiquitin left behind following a tryptic digest (the amino acid that precedes the two glycine is arginine). Should a trypsinised peptide weigh 114.04 Da more than predicted, there is a good chance that ubiquitin was present. However, it should not be forgotten that this diglycine motif is isobaric with dicarbamidomethylation (Kim, Zhong and Pandey, 2016). Furthermore, ubiquitin is not the only Ubl to generate to this motif – NEDD8 and ISG15 leave diglycine motifs upon trypsinisation (Kim *et al.*, 2011a).

An important signalling cascade, communicating that something is fundamentally wrong with the cell, is the DNA damage response. Ubiquitin is liberally used within this context. Firstly, within 15 seconds of damage, a massive accumulation of ubiquitin surrounds the site of DNA damage (Feng and Chen, 2012). This could be due to monoubiquitination of histone subunits H2A, H2B, and H2AX, which destabilises the nucleosome (Li *et al.*, 1993). However, a specific pathway that is highly dependent on ubiquitin is the repair of interstrand crosslinks, which relies on FANCD2 and a collection of partner proteins. A deficiency in these proteins can lead to Fanconi anaemia, which is typified by sensitivity to DNA crosslinking agents, bone marrow failure, increased risk of cancer, and developmental abnormalities (Kottemann and Smogorzewska, 2013). Exposure to a DNA crosslinking agent (such as cisplatin or cyclophosphamide) results in FANCD2 becoming monoubiquitinated on lysine 561 and its association with damaged DNA. Monoubiquitinated FANCD2 co-localises with other proteins that promote repair, perhaps by homologous recombination, or translesion DNA synthesis, or through nuclease activity (Longerich *et al.*, 2014). Multiple proteins with ubiquitin binding motifs interact with FANCD2's monoubiquitin and are thereby recruited to damaged DNA.

4.2 Aims and hypotheses

Below is a study that characterises the interactome for components of the ubiquitination system, namely the putative E1 LmxM.34.3060, and the putative E2 LmxM.33.0900. This information was obtained by using CRISPR to fuse a BirA*

to the proteins of interest, then mass spectrometry to analyse the pull downs. Additionally, attempts were made to immunoprecipitate ubiquitinated peptides.

- It was hypothesised that the putative E1 and E2 will have an interactome that features other components and substrates of the ubiquitin-proteasome system.

- It was hypothesised that *Leishmania* will utilise ubiquitin to respond to interstrand crosslinking DNA damage.

4.3 Results

4.3.1 Standard BioID

4.3.1.1 CRISPR

Usage of CRISPR requires DNA sequences to guide the Cas9 protein, and DNA sequences to guide DNA damage repair (complete with the sequences for insertion). Using online tools (http://leishgedit.net/), primers were designed for a modular plasmid based CRISPR system. sgRNA was designed to direct the Cas9 protein to the 5'- and 3'-ends of the gene encoding LmxM.34.3060, a gene encoding a putative ubiquitin activating enzyme. A schematic is shown in Figure 4-1, demonstrating these intended cleavage sites. Also shown are the PCR products that will guide the Cas9 protein to the 5'-end of the gene.

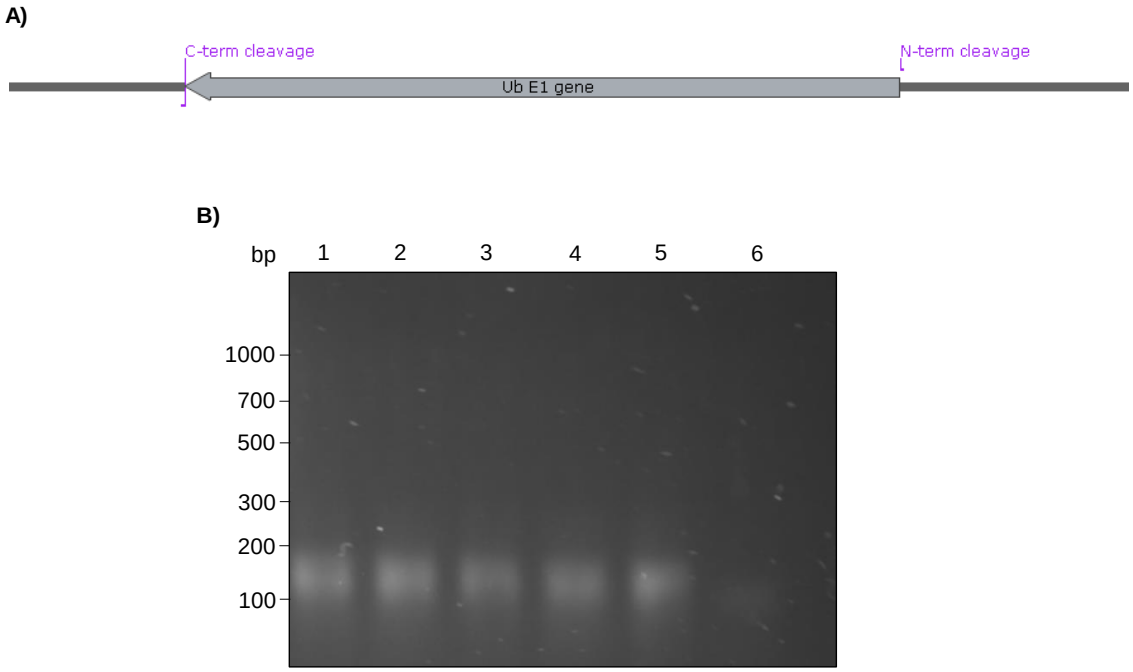


Figure 4-1 Cleavage sites and guide DNA generation used for CRISPR modification of LmxM.34.3060. A) Intended cleavage sites for the Cas9 protein. "Ub E1 gene" refers to ubiquitin activating enzyme gene, "N-term cleavage" and "C-term cleavage"refer to whether modifications will be at the N- or C-terminal of the LmxM.34.3060 protein. **B)** PCR amplification of DNA that will guide Cas9 to the 5'-end of LmxM.34.3060. Lanes 1 – 5 show these PCR reactions, lane 6 shows a negative control that used water instead of template DNA. Molecular weight markers had product code N3231, and were supplied by NEB.

To guide the double strand break repair process, template DNA was amplified from a plasmid that encoded a promiscuous biotinylating enzyme, BirA* (Figure 4-2). The predicted binding sites of primers used to create this construct are shown – note the two binding sites for the "(1) Ub E1 N term Forward" primer, and the two PCR products generated (fragments approximately 600 bp and 3000 bp). Due to an undesired fragment of 600 bp, a gradient PCR was conducted, varying the annealing temperature from 65°C to 72°C. However, this did not affect the ratio of the fragments.

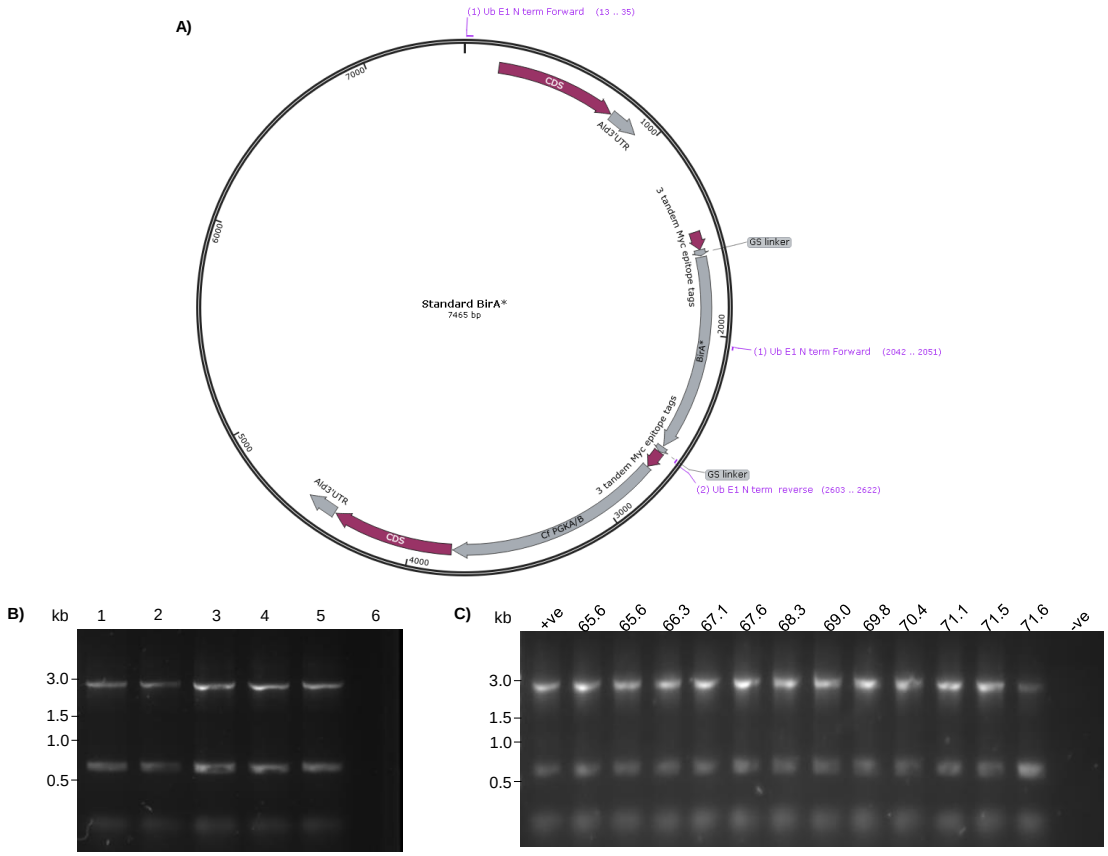


Figure 4-2 DNA amplified to fuse a biotinylating enzyme to the N-terminal of LmxM.34.3060. **A)** Schematic of predicted primer binding sites on a plasmid containing the BirA* gene. **B)** PCR products from this amplification run on an agarose gel. Lanes 1 – 5 show the PCR products obtained, and lane 6 is a negative control. **C)** A gradient PCR run on an agarose gel. Lane labels refer to the annealing temperature used, with "+ve" and "-ve" referring to positive and negative controls respectively. Molecular weight markers had product code N0552 and were supplied by NEB.

To fuse BirA* to the C-terminal of LmxM.34.3060 repair templates were amplified (Figure 4-3). Although only one annealing site per primer is predicted, two PCR products were identified by gel electrophoresis – one at the expected 3 kb, and another at 1.4 kb.

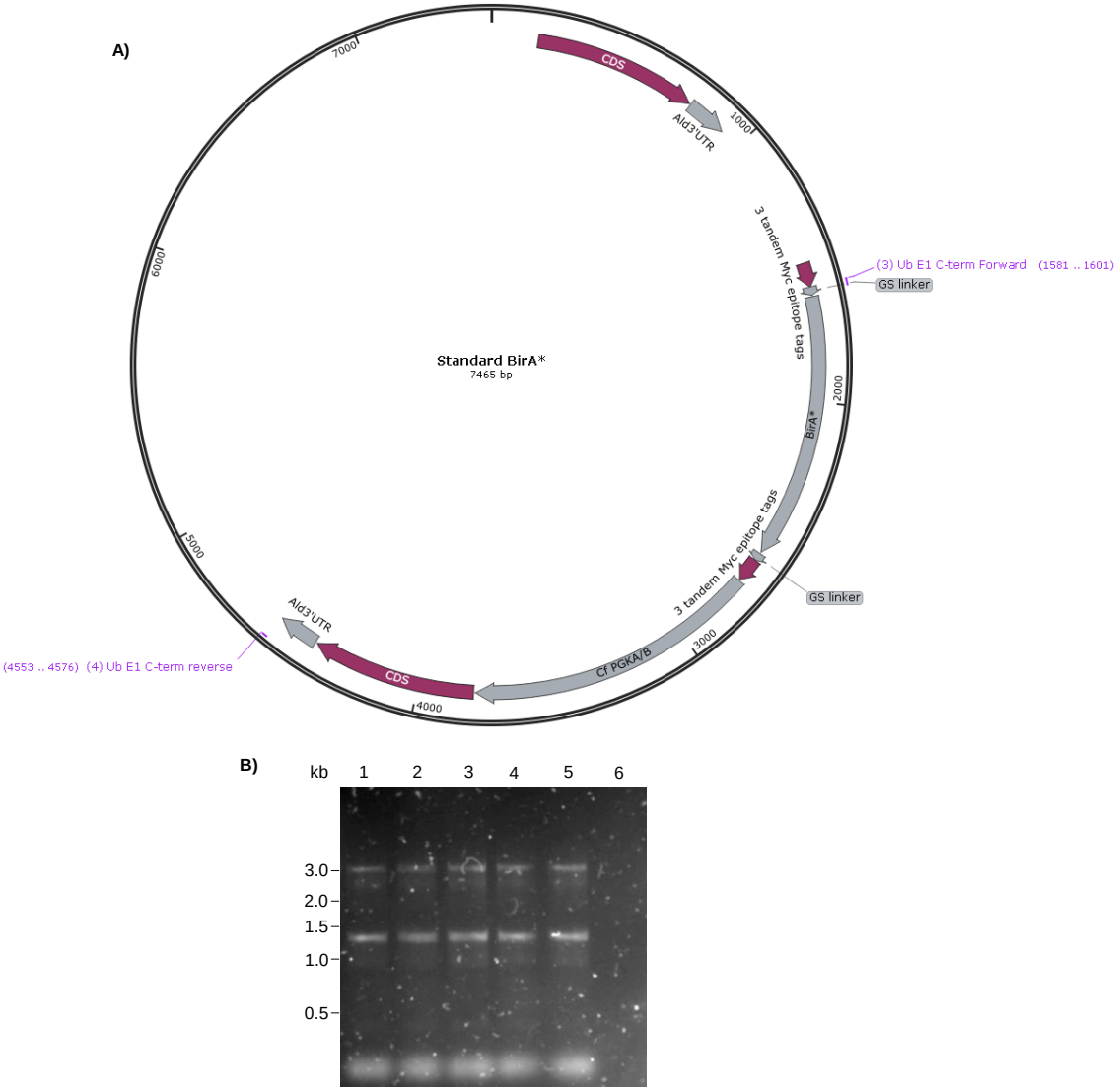


Figure 4-3 DNA amplified to fuse a biotinylating enzyme to the C-terminal of LmxM.34.3060. A) Schematic of predicted primer binding sites on a plasmid containing the BirA* gene. **B)** PCR products from this amplification run on an agarose gel. Lanes 1 – 5 show the PCR products obtained, and lane 6 is a negative control, a PCR reaction without template DNA. Molecular weight markers had product code N0552 and were supplied by NEB.

Using the PCR reactions containing 3 kb and 0.6 kb for N-terminal tagging, and 3 kb and 1.4 kb for C-terminal tagging, transfections commenced. To determine the success of the transfections western blots were conducted (Figure 4-4). For both N- and C-terminally tagged cell lines a myc tagged protein of the expected mass is present (approximately 168 kDa), however there is also a myc tagged protein of approximately 92 kDa, which is smaller than the wild type LmxM.34.3060. A second western blot conducted on subsequent progeny of the same cell line, showed increased abundance of the 92 kDa protein and decreased abundance of the 168 kDa protein.

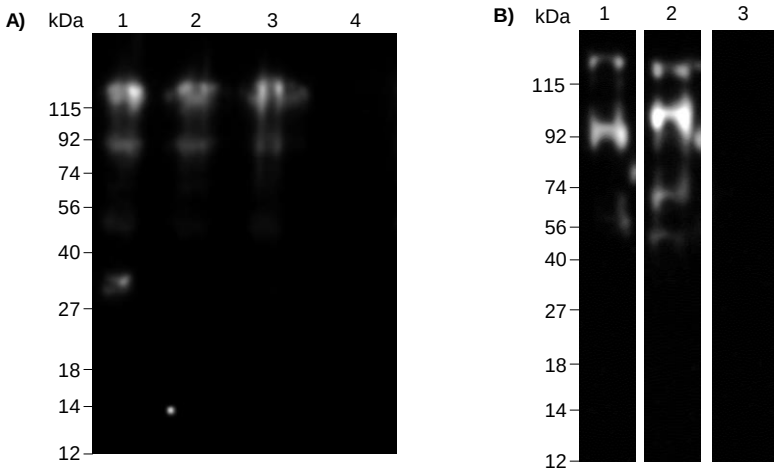


Figure 4-4 Western blots detecting the presence of modified LmxM.34.3060. A) Myc tagged BirA* fused to the N-terminal of LmxM.34.3060. Lanes 1 – 3 feature lysates from various transfected cultures, lane 4 is a lysate from the Cas 9 parental cell line. **B)** Myc tagged BirA* on N- and C-terminal of LmxM.34.3060. Lane 1 is lysate from N-terminal tagged LmxM.34.3060, lane 2 is C-terminal tagged LmxM.34.3060, lane 3 is lysate from the Cas 9 parental cell line. The lysates probed in B) were obtained a few months after A). All samples were run on 4-20% SDS-PAGE gels, transferred to a PVDF membrane, and the protein of interest detected using α-myc antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Additionally, PCRs were performed to validate the insertion of the BirA* gene into the correct location of the *Leishmania* genome. These are shown in Figure 4-5. Although amplifications of the target sequence (the BirA* gene) could be made from the plasmid, amplification using genomic DNA was never successful. On one occasion, amplification of a biopterin transporter, a positive control unrelated to the CRISPR modifications, was successful when amplification of the BirA* gene failed Figure 4-5(Figure 4-5C). However, the success of this positive control was not able to replicated in subsequent experiments (Figure 4-5D).

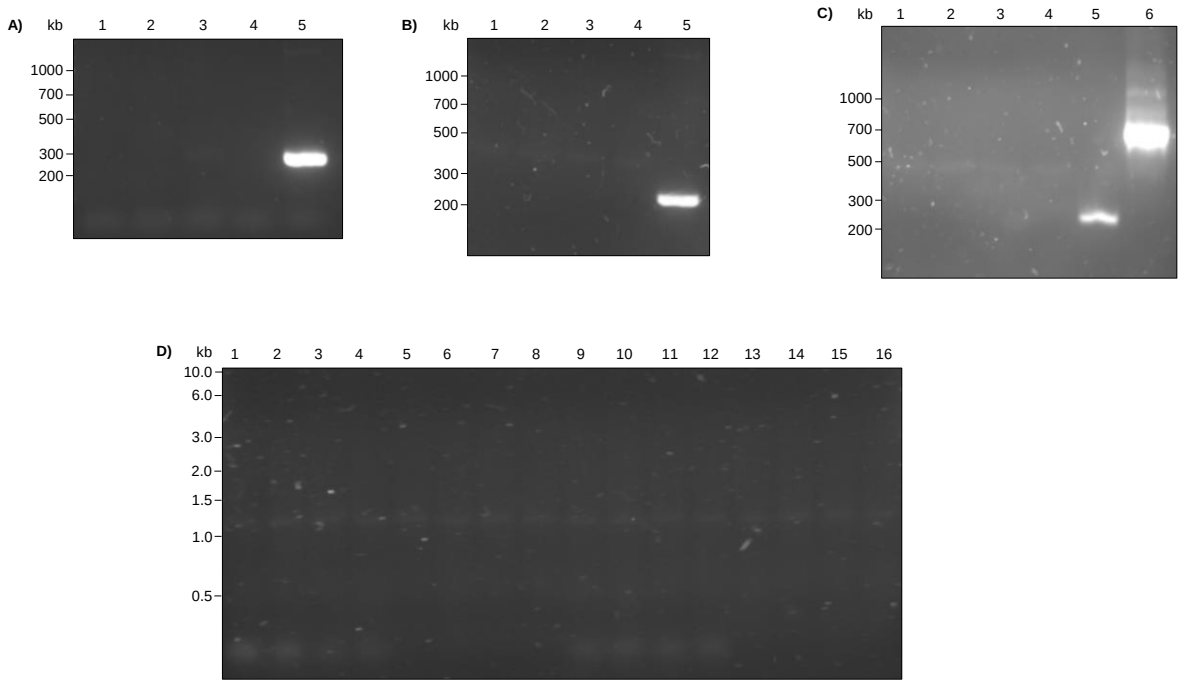


Figure 4-5 Leishmania genomic DNA amplified determine the insertion of BirA* to the N-terminal of LmxM.34.3060. A) Using "confirmation" primers that anneal to the BirA* gene. Lanes 1 – 3 feature gDNA from various transfected cell lines. Lane 4 used gDNA from the Cas 9 parental cell line, and lane 5 used the "Standard BirA*" plasmid as a template. **B)** Using the same reaction conditions as A), but used more gDNA as a template, and more PCR products loaded onto the gel. **C)** Using the same reaction conditions as A) and B), however lane 6 uses primers that bind to a gene encoding a biopterin transporter. **D)** A combination of primers used to amplify various regions of gDNA. Lanes 1 – 4 use the same primers that were employed for the amplification of repair template, lanes 5 – 8 used both "confirmation" primers and "repair template" primers as the forward and reverse primers respectively, lanes 9 – 12 used "repair template" and "confirmation" primers as the forward and reverse primers respectively, lanes 13 – 16 used primers that anneal to a biopterin transporter gene. For A), B), and C) molecular weight markers had product code N0467, for D) molecular weight markers had product code N0552. Both were supplied by NEB.

4.3.1.2 BioID

Despite the failure of the PCRs, a BioID experiment was conducted. This was due to the presence of a myc-tagged protein of the expected size in western blots. Figure 4-6 shows the number of protein IDs obtained when the BioID experiment was conducted using BirA* as an N-terminal tag or as a C-terminal tag. Of note is that the putative E1 itself was pulled down in both conditions. Four other proteins were also pulled down with both N- and C-terminally tagged LmxM.34.3060: a putative heat shock protein 70 related protein (LmxM.18.1370), calmodulin (LmxM.09.0910), a Valosin containing protein homolog (LmxM.36.1370), and the gamma subunit of protein transport protein sec61 (LmxM.25.1015). These are listed in Table 4-1. There were 54 proteins identified using an N-terminal tag, and 26 proteins identified with a C-terminal tag. These are listed in Table 4-2 and Table 4-3 respectively. Of note is

LmxM.34.1300, a putative E2 that was revealed in the bioinformatic screening, and was identified only when using the N-terminally tagged putative E1.

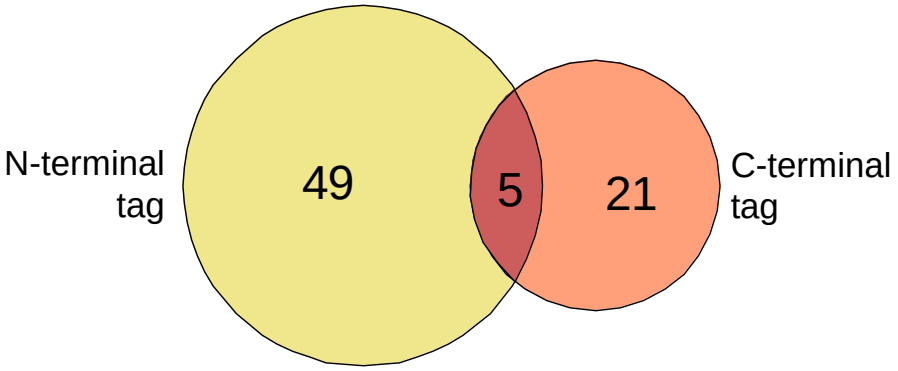


Figure 4-6 A Venn diagram of protein identifications from mass spectrometry conducted on cell lines with LmxM.34.3060-BirA* fusion proteins. N-terminal tag and C-terminal tag refer to whether the BirA* was fused to the N- or C-terminal of LmxM.34.3060. Proteins also identified in a negative control, the Cas 9 parental cell line, have been removed.

LmxM.34.3060 - both termini	
Accession	Description
LmxM.09.0910	calmodulin, putative
LmxM.18.1370	heat shock protein, putative
LmxM.25.1015	protein transport protein Sec61 gamma subunit, putative
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.36.1370	valosin-containing protein, putative

Table 4-1 Proteins identified by both N- and C-terminally BirA* tagged LmxM.34.3060.

LmxM.34.3060 - N-terminal	
Accession	Description
LmxM.03.0710	hypothetical protein, conserved
LmxM.03.0840	hypothetical protein, conserved
LmxM.05.0830	methylthioadenosine phosphorylase, putative
LmxM.06.0880	acyl-coenzyme a dehydrogenase, putative
LmxM.07.0350	hypothetical protein, conserved
LmxM.07.0570	hypothetical protein, conserved
LmxM.07.0870	splicing factor ptsr1-like protein
LmxM.08.0430	hypothetical protein, conserved
LmxM.08_29.0250	oxidase-like protein
LmxM.08_29.0880	ADP-ribosylation factor-like protein 3A, putative (arl3a)
LmxM.08_29.1580	hypothetical protein, conserved
LmxM.08_29.2630	hypothetical protein, conserved
LmxM.09.0910	calmodulin, putative

LmxM.11.1130partial	unspecified product
LmxM.14.1210	hypothetical protein, conserved
LmxM.17.0870	hypothetical protein, conserved
LmxM.18.0610	ATP-dependent zinc metallopeptidase, putative, metallo-peptidase, Clan MA(E), Family M41
LmxM.18.1340	hypothetical protein, conserved
LmxM.18.1370	heat shock protein, putative
LmxM.18.1400	60S ribosomal protein L34, putative
LmxM.19.0650	hypothetical protein, conserved
LmxM.19.1120	proteasome regulatory non-ATP-ase subunit 9, putative,19S proteasome regulatory subunit 9, putative (RPN9)
LmxM.21.0810	methionyl-tRNA synthetase, putative
LmxM.21.0845	hypoxanthine-guanine phosphoribosyltransferase
LmxM.22.0720	hypothetical protein, conserved
LmxM.23.0640	hypothetical protein, conserved
LmxM.24.0140	ankyrin/TPR repeat protein
LmxM.25.1015	protein transport protein Sec61 gamma subunit, putative
LmxM.27.1300	hypothetical protein, conserved
LmxM.27.1630	hypothetical protein, conserved
LmxM.28.0960	40S ribosomal protein S14
LmxM.28.1030	ribosomal protein S20, putative
LmxM.28.2470	MRP protein-like protein
LmxM.28.2510	acyl-CoA dehydrogenase, putative
LmxM.29.3340	60S ribosomal protein L9, putative
LmxM.30.0410	calpain-like cysteine peptidase, putative,cysteine peptidase, Clan CA, family C2, putative
LmxM.30.0670	hypothetical protein, conserved
LmxM.30.1670	hypothetical protein, conserved
LmxM.31.0360	hypothetical protein, conserved
LmxM.31.0870	phenylalanyl-tRNA synthetase alpha chain, putative
LmxM.31.2180	hypothetical protein, conserved
LmxM.31.2690	ribosomal protein L27, putative
LmxM.32.2270	hypothetical protein, conserved
LmxM.32.2800	hypothetical protein, conserved
LmxM.33.2560	hypothetical protein, conserved
LmxM.34.0760	CBS domain protein, conserved
LmxM.34.1300	ubiquitin-conjugating enzyme E2, putative
LmxM.34.1880	60S ribosomal protein L5, putative
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.34.3530	hypothetical protein, unknown function
LmxM.34.3790	60S ribosomal protein L23, putative
LmxM.34.4430	mitochondrial phosphate transporter, putative
LmxM.36.0250	EIF3-interacting protein-like protein

LmxM.36.1370	valosin-containing protein, putative

Table 4-2 Proteins identified by N-terminally tagged BirA*-LmxM.34.3060.

C-terminal	
Accession	Description
LmxM.02.0550	hypothetical protein, unknown function
LmxM.03.0030	D-3-phosphoglycerate dehydrogenase-like protein
LmxM.08.0940	hypothetical protein, conserved
LmxM.08_29.2140	C-8 sterol isomerase-like protein
LmxM.09.0910	calmodulin, putative
LmxM.10.0405	GP63, leishmanolysin
LmxM.11.0960	40S ribosomal protein S5
LmxM.13.1210	nucleobase transporter
LmxM.14.1120	kinesin K39, putative
LmxM.18.1370	heat shock protein, putative
LmxM.24.0820	inositol polyphosphate phosphatase, putative
LmxM.25.1015	protein transport protein Sec61 gamma subunit, putative
LmxM.25.1210	hypothetical protein, conserved
LmxM.26.1020	dynein heavy chain, putative
LmxM.26.1240	heat shock protein 70-related protein
LmxM.27.0760	ras-related protein RAB1A, putative (Rab1A)
LmxM.27.1870	trypanothione synthetase, putative
LmxM.28.2740	activated protein kinase c receptor (LACK)
LmxM.30.1220	vacuolar-type proton translocating pyrophosphatase 1, putative
LmxM.30.2790	ADP-ribosylation factor, putative
LmxM.34.1410	threonyl-tRNA synthetase, putative
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.34.3230	cystathione gamma lyase, putative
LmxM.36.1370	valosin-containing protein, putative
LmxM.36.2360	tyrosine aminotransferase, putative
LmxM.36.3180	receptor-type adenylate cyclase a-like protein

Table 4-3 Proteins identified by C-terminally tagged LmxM.34.3060-BirA*.

An identical strategy was followed to obtain N- and C-terminal tags on the E2 identified from the E1-probe and crosslinking strategy – LmxM.33.0900 (Figure 4-7). Although a similar phenomenon occurred when amplifying the repair cassettes to tag the N-terminal of the E2 as occurred when tagging the N-terminal of the E1 – namely, two fragments of 3 and 0.6 kb were predicted and observed, of which only the 3 kb was desired. When amplifying the repair

cassette to tag the C-terminal of LmxM.33.0900 only one fragment of the expected length was observed.

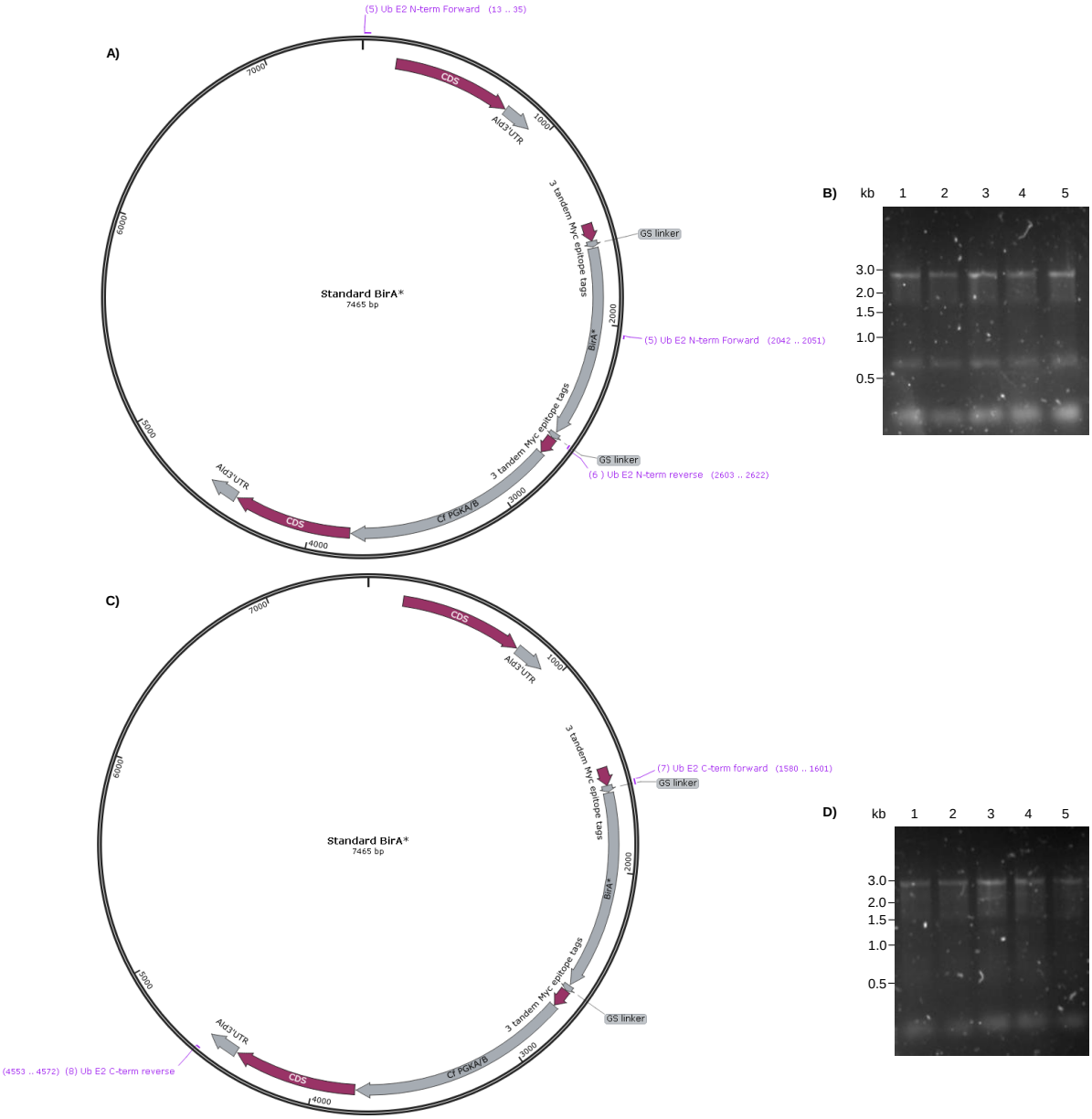


Figure 4-7 DNA constructs formed to tag LmxM.33.0900 with BirA*.
A) Schematic of predicted primer binding sites on a plasmid containing the BirA* gene, with the amplification products used to tag the N-terminal of LmxM.33.0900. **B)** PCR products from the amplification described in A) run on an agarose gel. Lanes 1 – 5 show the PCR products obtained. **C)** Schematic of predicted primer binding sites on a plasmid containing the BirA* gene, with the amplification products used to tag the C-terminal of LmxM.33.0900. **D)** PCR products from the amplification described in C) run on an agarose gel. Lanes 1 – 5 show the PCR products obtained. Molecular weight markers had product code N0552 and were supplied by NEB.

The success of transfections was determined by western blotting, blotting for a myc-tagged protein of the expected mass (73 kDa) (Figure 4-8). The N-terminally tagged LmxM.33.0900 is greater than 74 kDa, whereas C-terminally tagged is just smaller than 74 kDa. Nonetheless, transfectant cell lines possessed a myc-tagged protein of approximately the right size.

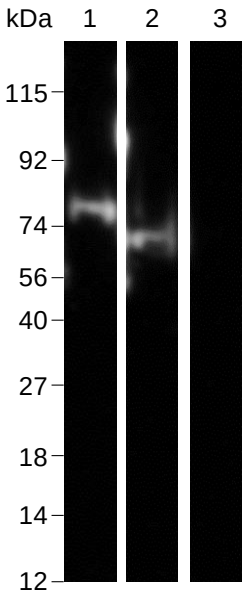


Figure 4-8 Western blot detecting the presence of modified LmxM.33.0900. Lane 1 shows a lysate from a transfectant cell line that had myc tagged BirA* fused to the N-terminal of LmxM.33.0900. Lane 2 shows the myc tagged BirA* on the C-terminal of LmxM.33.0900. Lane 4 is a lysate from the Cas 9 parental cell line. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-myc antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Using these cell lines, streptavidin pull downs and mass spectrometry were conducted after spiking in biotin (Figure 4-9). Proteins also identified from a Cas 9 parental cell line were excluded. Both the N- and C-terminally tagged LmxM.33.0900 featured LmxM.34.3060, the putative E1, in their pull downs. Other proteins identified in both the N- and C-terminally tagged pull downs include GP63 (LmxM.10.0405), a heat shock protein 70 related protein (LmxM.26.1240), cystathione gamma lyase (LmxM.34.3230), poly(a) binding protein (LmxM.34.4130), dynein heavy chain (LmxM.25.0980), 3-ketoacyl-CoA thiolase like protein (LmxM.23.0690), the gamma subunit of protein transport protein sec61 (LmxM.25.1015), and a hypothetical protein of unknown function (LmxM.02.0550). All proteins identified by both transfectants are shown in Table 4-4, with proteins also identified by anti-diglycine immunoprecipitations (see 4.3.3 The ubiquitinome) highlighted in red. Proteins identified by each of the N- and C-terminally tagged LmxM.33.0900 fusion proteins are shown in Table 4-5 and Table 4-6 respectively, with proteins identified by anti-diglycine immunoprecipitations (see 4.3.3 The ubiquitinome) again highlighted in red.

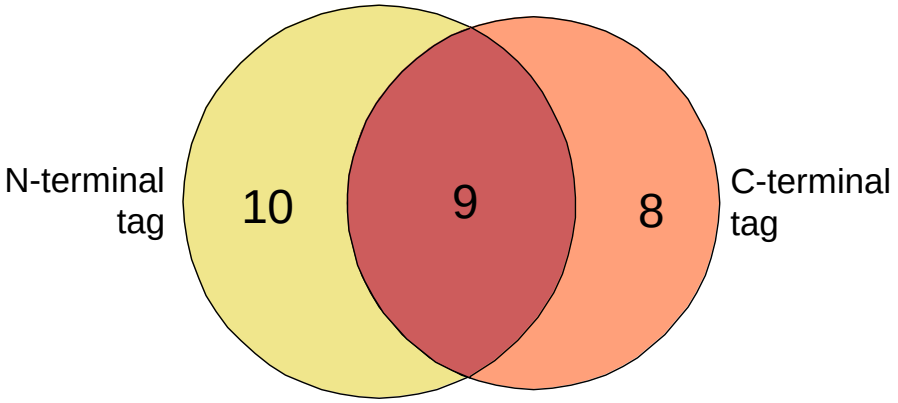


Figure 4-9 A Venn diagram of protein identifications from mass spectrometry conducted on cell lines with LmxM.33.0900-BirA* fusion proteins. N-terminal tag and C-terminal tag refer to whether the BirA* was fused to the N- or C-terminal of LmxM.33.0900. Proteins also identified in a negative control, the Cas 9 parental cell line, have been removed.

Both	
Accession	Description
LmxM.02.0550	hypothetical protein, unknown function
LmxM.10.0405	GP63, leishmanolysin
LmxM.23.0690	3-ketoacyl-CoA thiolase-like protein
LmxM.25.0980	dynein heavy chain, putative
LmxM.25.1015	protein transport protein Sec61 gamma subunit, putative
LmxM.26.1240	heat shock protein 70-related protein
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.34.3230	cystathione gamma lyase, putative
LmxM.34.4130	polyadenylate-binding protein 2

Table 4-4 Proteins identified by both N- and C-terminally BirA* tagged LmxM.33.0900. Proteins highlighted in red also appear in anti-diglycine immunoprecipitations.

N-terminal	
Accession	Description
LmxM.02.0550	hypothetical protein, unknown function
LmxM.10.0405	GP63, leishmanolysin
LmxM.10.0870	histone H3
LmxM.13.1210	nucleobase transporter
LmxM.13.1650	dynein heavy chain, putative
LmxM.16.0730	ubiquitin hydrolase, putative,cysteine peptidase, Clan CA, family C19, putative
LmxM.23.0690	3-ketoacyl-CoA thiolase-like protein
LmxM.25.0980	dynein heavy chain, putative
LmxM.25.1015	rotein transport protein Sec61 gamma subunit, putative
LmxM.26.1240	heat shock protein 70-related protein
LmxM.27.1870	trypanothione synthetase, putative
LmxM.28.0930	hypothetical protein, conserved

LmxM.30.2790	ADP-ribosylation factor, putative
LmxM.34.1410	threonyl-tRNA synthetase, putative
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.34.3230	cystathione gamma lyase, putative
LmxM.34.3700	Gim5A protein, putative,glycosomal membrane protein
LmxM.34.4130	polyadenylate-binding protein 2
LmxM.36.2360	tyrosine aminotransferase, putative

Table 4-5 Proteins identified by N-terminally tagged BirA*-LmxM.33.0900. Proteins highlighted in red also appear in anti-diglycine immunoprecipitations.

C-terminal	
Accession	Description
LmxM.02.0550	hypothetical protein, unknown function
LmxM.10.0405	GP63, leishmanolysin
LmxM.12.0590	hypothetical protein, conserved
LmxM.23.0690	3-ketoacyl-CoA thiolase-like protein
LmxM.24.1630	succinate dehydrogenase flavoprotein, putative
LmxM.25.0420	hypothetical protein, conserved
LmxM.25.0980	dynein heavy chain, putative
LmxM.25.1015	protein transport protein Sec61 gamma subunit, putative
LmxM.26.1240	heat shock protein 70-related protein
LmxM.27.0760	ras-related protein RAB1A, putative (Rab1A)
LmxM.27.0930	isovaleryl-coA dehydrogenase, putative
LmxM.28.0960	40S ribosomal protein S14
LmxM.30.1220	vacuolar-type proton translocating pyrophosphatase 1, putative
LmxM.34.0790	actin-like protein, putative
LmxM.34.3060	ubiquitin-activating enzyme E1, putative
LmxM.34.3230	cystathione gamma lyase, putative
LmxM.34.4130	polyadenylate-binding protein 2

Table 4-6 Proteins identified by N-terminally tagged LmxM.33.0900-BirA*. Proteins highlighted in red also appear in anti-diglycine immunoprecipitations.

4.3.2 MiniTurbo BioID

4.3.2.1 CRISPR

To build on these results, a smaller, faster-labelling BirA* was used – miniTurbo. Although the same sgRNA could be used as used for standard BirA* (as the same genes were being targeted), new repair templates had to be formed that encoded the miniTurbo BirA*. Figure 4-10 shows the repair templates amplified

to tag the N-terminal of the putative E1 LmxM.34.3060 in a gradient PCR. It is evident that two PCR products are expected, of 2204 bp and 2119 bp, however only the former is desired. Although the similar sizes makes the fragments difficult to resolve on gel, they should be amplified at different annealing temperatures, given differences in the number of primer bases that can anneal in each location. To amplify the desired 2204 bp fragment, the "Ub E1 N term reverse" primer is predicted to have an annealing temperature of 57°C; whereas to anneal at the truncated location the same primer is predicted to have an annealing temperature of 42°C. As such a gradient PCR was conducted, with temperatures ranging from 65.6°C to 71°C (with the higher temperatures possibly decreasing the abundance of the truncated fragment). The product size may increase marginally at higher temperatures, however any small difference could easily be an experimental artifact. For transfections, the reaction in lane 4 was used.

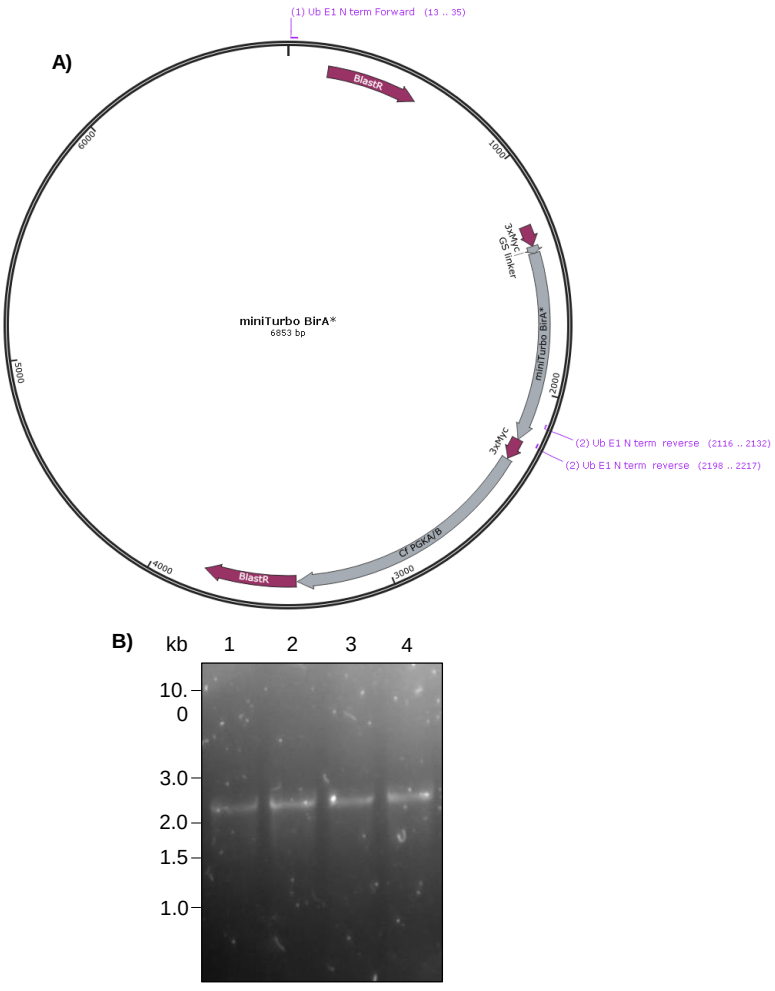


Figure 4-10 DNA amplified to fuse a biotinylating enzyme to the N-terminal of LmxM.34.3060.
A) Schematic of predicted primer binding sites on a plasmid containing the miniTurbo BirA* gene.
B) PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products

obtained at various annealing temperatures - lane 1: 65.6°C, lane 2: 67.1°C, lane 3: 69.8°C, lane 4: 71.6°C. Molecular weight markers had product code N0552 and were supplied by NEB.

An identical PCR gradient was used to amplify the repair templates for C-terminal tagging of LmxM.34.3060. This is shown is Figure 4-11. Although only one PCR product is predicted (2590 bp), Figure 4-11B shows products at approximately 2500 bp and 1300 bp, with the ratio shifting in favour of the smaller fragment at higher temperatures (a gradient of 65.6˚C to 71˚C). As such, Figure 4-11C was formed, using PCR gradient with lower annealing temperatures (50.6˚C to 63.6˚C).

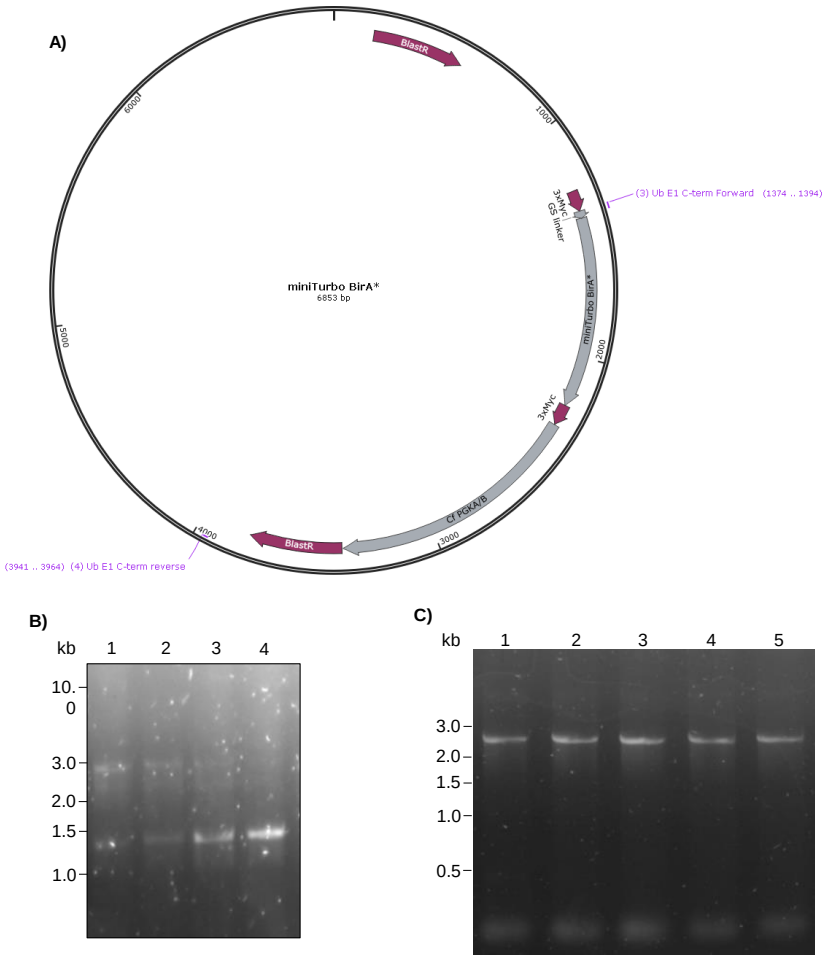


Figure 4-11 DNA amplified to fuse a biotinylating enzyme to the C-terminal of LmxM.34.3060. **A)** Schematic of predicted primer binding sites on a plasmid containing the miniTurbo BirA* gene. **B)** PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products obtained at various annealing temperatures - lane 1: 65.6°C, lane 2: 67.1°C, lane 3: 69.8°C, lane 4: 71.6°C. **C)** PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products obtained at various annealing temperatures - lane 1: 50.6°C, lane 2: 53.8°C, lane 3: 56.8°C, lane 4: 60.0°C, lane 5: 63.3°C. Molecular weight markers had product code N0552 and were supplied by NEB.

The miniTurbo BirA* was also used to tag the putative E2 LmxM.33.0900. When tagging the N-terminal of the putative E2, the problems encountered (two similarly sized PCR products) and solutions developed (a gradient PCR) were

identical as when tagging the N-terminal of the putative E1. This can be seen in Figure 4-12. As before, although it is difficult to confirm the success of the method, the reaction in lane 4 was used for transfections, due to it using the highest annealing temperature.

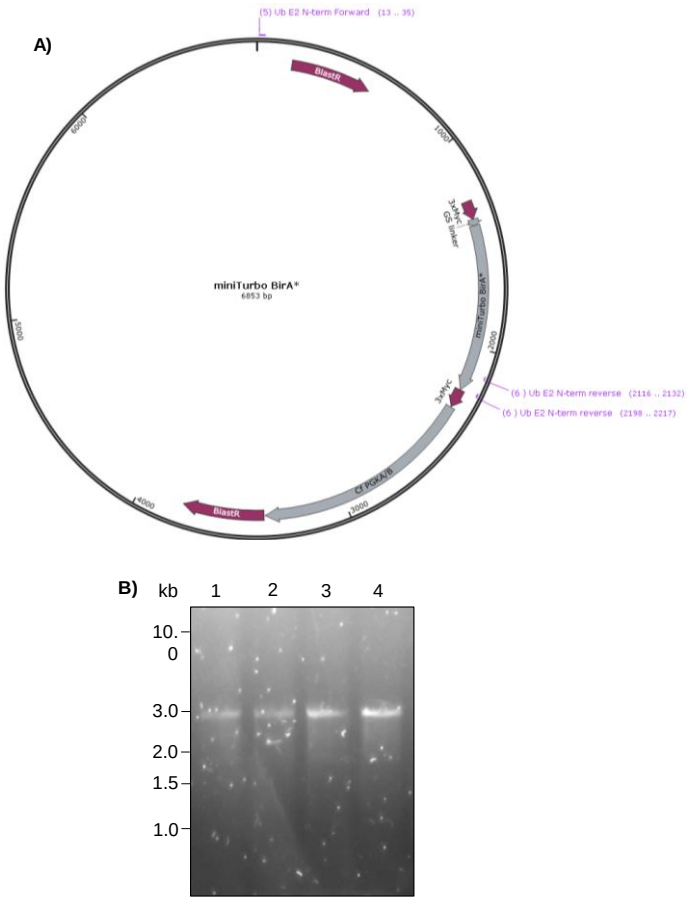


Figure 4-12 DNA amplified to fuse a biotinylating enzyme to the N-terminal of LmxM.34.3060. **A)** Schematic of predicted primer binding sites on a plasmid containing the miniTurbo BirA* gene. **B)** PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products obtained at various annealing temperatures - lane 1: 65.6°C, lane 2: 67.1°C, lane 3: 69.8°C, lane 4: 71.6°C. Molecular weight markers had product code N0552 and were supplied by NEB.

The miniTurbo BirA* was used to tag the C-terminal of LmxM.33.0900 as well. The amplification of the repair templates necessary for this are shown in Figure 4-13. As found previously, the PCR amplification performed poorly and generated multiple fragments at higher annealing temperatures (65.6˚C – 71˚C), but amplified fragments of the expected size at lower temperatures (50.6˚C – 63.6˚C).

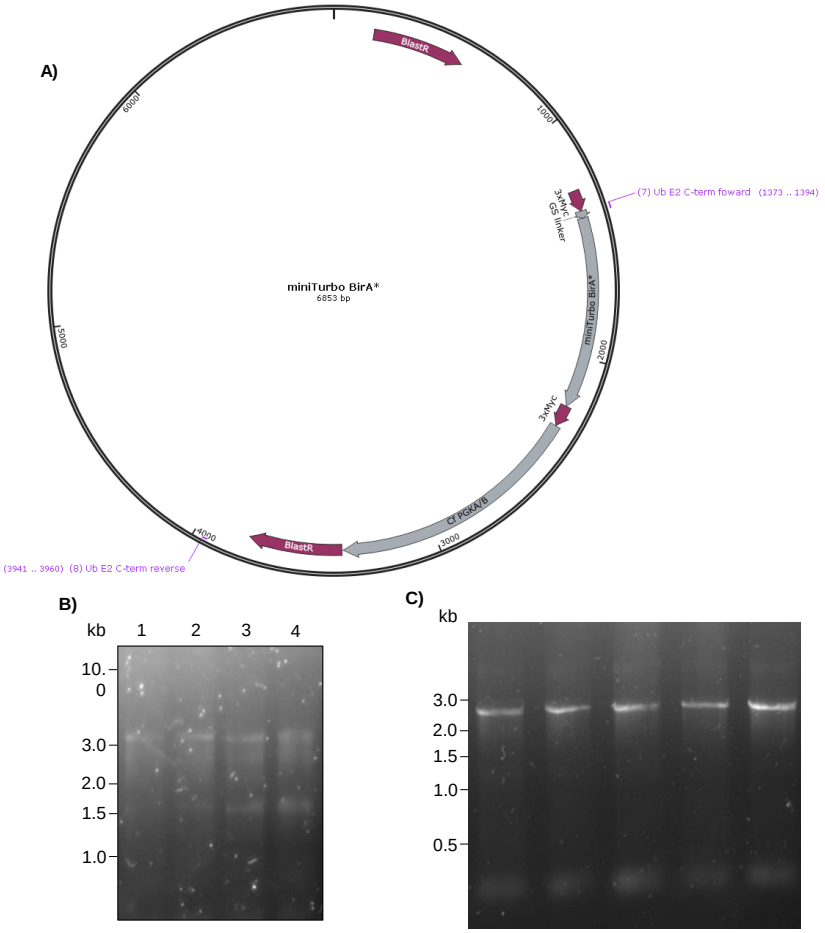


Figure 4-13 DNA amplified to fuse a biotinylating enzyme to the C-terminal of LmxM.33.0900. **A**) Schematic of predicted primer binding sites on a plasmid containing the miniTurbo BirA* gene. **B**) PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products obtained at various annealing temperatures - lane 1: 65.6°C, lane 2: 67.1°C, lane 3: 69.8°C, lane 4: 71.6°C. **C)** PCR products from a gradient PCR run on an agarose gel. Lanes 1 – 4 show the PCR products obtained at various annealing temperatures - lane 1: 50.6°C, lane 2: 53.8°C, lane 3: 56.8°C, lane 4: 60.0°C, lane 5: 63.3°C. Molecular weight markers had product code N0552 and were supplied by NEB.

After transfections western blots were performed to confirm the existence of tagged proteins in cell cultures (Figure 4-14). The putative E1 with a miniTurbo tag is predicted to have a mass of 160 kDa, while the putative E2~miniTurbo conjugate is predicted to be 65.3 kDa. Tagging of each terminal of the putative E1 seems to have resulted in a variety of myc-tagged proteins, with a protein of approximately 45 kDa particularly abundant. However, the range of proteins does not seem to include to a mass of 160 kDa. The western blot against the modified putative E2 shows a myc tagged protein around 74 kDa for both N- and C-terminally tagged protein.

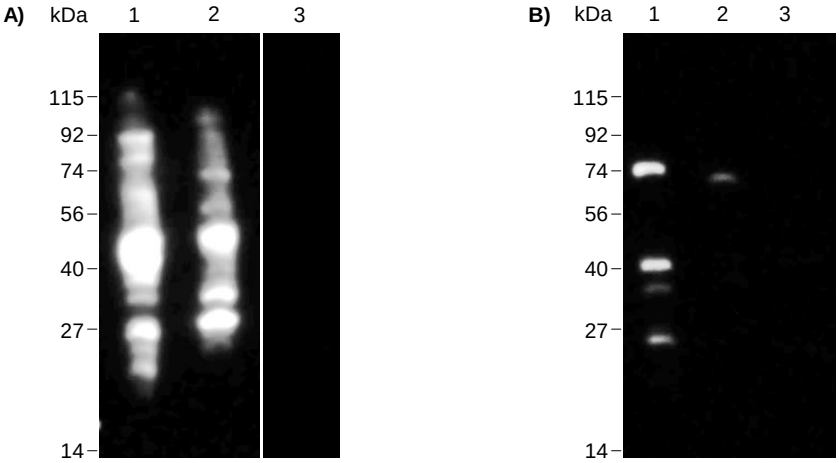
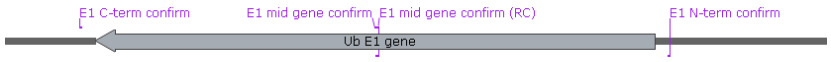


Figure 4-14 Western blots detecting the presence of LmxM.34.3060 and LmxM.33.0900 tagged with miniTurbo BirA*. A) Lysates from transfections involving LmxM.34.3060. Lane 1 shows a lysate from N-terminal tagging, lane 2 has a lysate from C-terminal tagging, and lane 3 shows a lysate of the Cas 9 parental cell line. **B)** Lysates from transfections involving LmxM.33.0900. Lane 1 shows a lysate from N-terminal tagging, lane 2 has a lysate from C-terminal tagging, and lane 3 shows a lysate of the Cas 9 parental cell line. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-myc antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

A method improvement employed when using miniTurbo BirA* was the use of serial dilutions after the transfections and western blots were conducted. The aim being to isolate single cells, and thereby generate clonal cell cultures. Additionally, the PCR confirmation system was redesigned, with primers predicted to bind at the midpoint of a gene of interest, and then at the intergenic regions at either the 5' or 3' prime end of the gene (Figure 4-15). This would allow a PCR product to be produced irrespective of the presence of the miniTurbo insert – a smaller product if the transfection failed to have the desired effect, and a larger one if the miniTurbo insert was found at the correct locus.

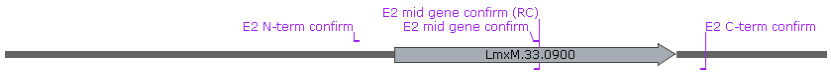
A)



B)

E1	N-terminal (bp)	C-terminal (bp)
Wild type	1824	1849
With miniTurbo	4029	4440

C)



D)

E2	N-terminal (bp)	C-terminal (bp)
Wild type	576	538
With miniTurbo	2781	3126

Figure 4-15 DNA amplified to confirm the success of the miniTurbo transfections. A) Schematic of wild type LmxM.34.3060 and surrounding intergenic region, showing where the confirmation primers are predicted to bind. RC stands for Reverse Complement. **B)** Table of the sizes of the predicted PCR products when tagging LmxM.34.3060, depending on where the tag was inserted (N- or C-terminal) and whether transfections were successful (with miniTurbo or wild type). **C)** Schematic of wild type LmxM.33.0900 and surrounding intergenic region, showing where the confirmation primers are predicted to bind. RC stands for Reverse Complement. **D)** Table of the sizes of the predicted PCR products when tagging LmxM.33.0900, depending on where the tag was inserted (N- or C-terminal) and whether transfections were successful (with miniTurbo or wild type).

After clones were generated, DNA extraction and PCRs commenced. Figure 4-16 shows PCRs conducted on genomic DNA from clonal cell lines, where cells were transfected with the constructs to tag the C-terminal of the putative E2. Should this transfection have been successful, there would be fragment of approximately 3 kbp, whereas wild type genomic DNA would produce a fragment of 500 bp. Of the 23 cell lines shown in Figure 4-16, none possess a 3 kbp fragment. Curiously, most cell lines in Figure 4-16B possess a fragment of approximately 1.5 kbp. It is also worth noting that production of any fragment, even a 500 bp fragment, is not uniform across all PCR reactions. Furthermore, PCR reactions conducted on the putative E1 with a potential C-terminal miniTurbo tag were inconclusive. Due to external circumstances, and the time consuming nature of screening, the miniTurbo project was ended here.

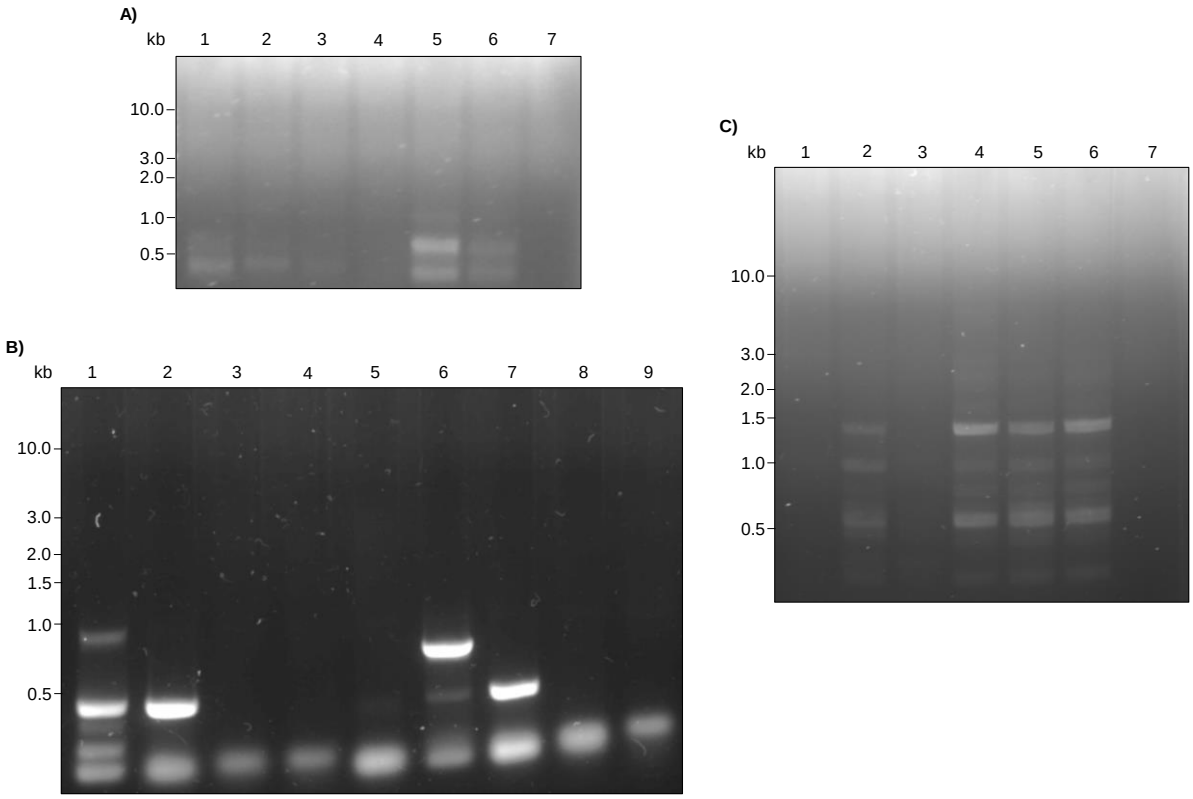


Figure 4-16 Leishmania genomic DNA amplified determine the insertion of miniTurbo BirA* to the C-terminal of LmxM.33.0900. A), B), and **C)** All lanes show PCR products amplified using the genomic DNA of various transfectant clonal cell lines. Molecular weight markers had product code N0552. Both were supplied by NEB.

4.3.3 The ubiquitinome

4.3.3.1 DNA damage and ubiquitination

To examine the substrates of ubiquitin, particularly those involved in the DNA damage response, cyclophosphamide, a DNA alkylating agent, was added to cell lines featuring a dimerizable Cre recombinase system (used to induce conditional knockouts). Anti-ubiquitin western blots were conducted to see whether addition of cyclophosphamide resulted in increased ubiquitination. Figure 4-17 shows western blots conducted by adding 5 mM cyclophosphamide, with samples collected at various time points. Curiously, in Figure 4-17A there seems to be more ubiquitin before addition of the DNA damaging reagent than in subsequent time points, although the second most is found three hours after addition. In Figure 4-17B, no differences in ubiquitin abundance can be seen in the 15-minute time points for the first three hours.

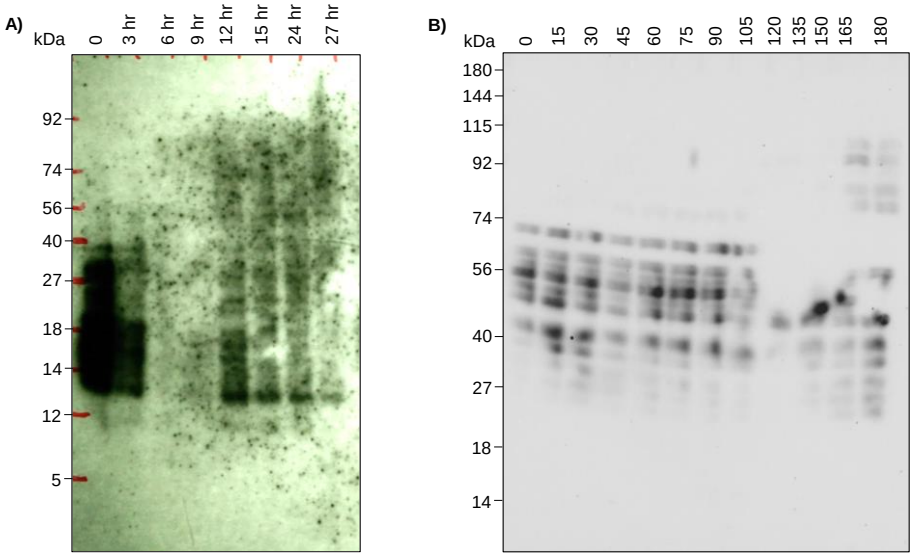


Figure 4-17 Anti-ubiquitin western blots conducted on Leishmania lysates after the addition of 5 mM cyclophosphamide. A) Cell lysates were collected every three hours for 15 hours, then at 24 and 27 hours. Lane labels refer to the hours post cyclophosphamide addition. The blot was imaged using X-ray film. **B)** Cell lysates were collected every 15 minutes for three hours. Lane labels refer to the minutes post cyclophosphamide addition. The blot was imaged using a gel documentation imaging system. All samples were run on 4-20% SDS-PAGE gels, transferred to a PVDF membrane, and the proteins of interest detected using α-ubiquitin antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Subsequent experiments varied the concentration of cyclophosphamide. Figure 4-18 shows these experiments. The concentrations used here are 2.5, 5, and 10 mM. The only image that could possibly show any difference in ubiquitination is using cyclophosphamide at 10 mM (Figure 4-18C), where there is potentially more ubiquitin below 27 kDa at 195 and 210 minutes.

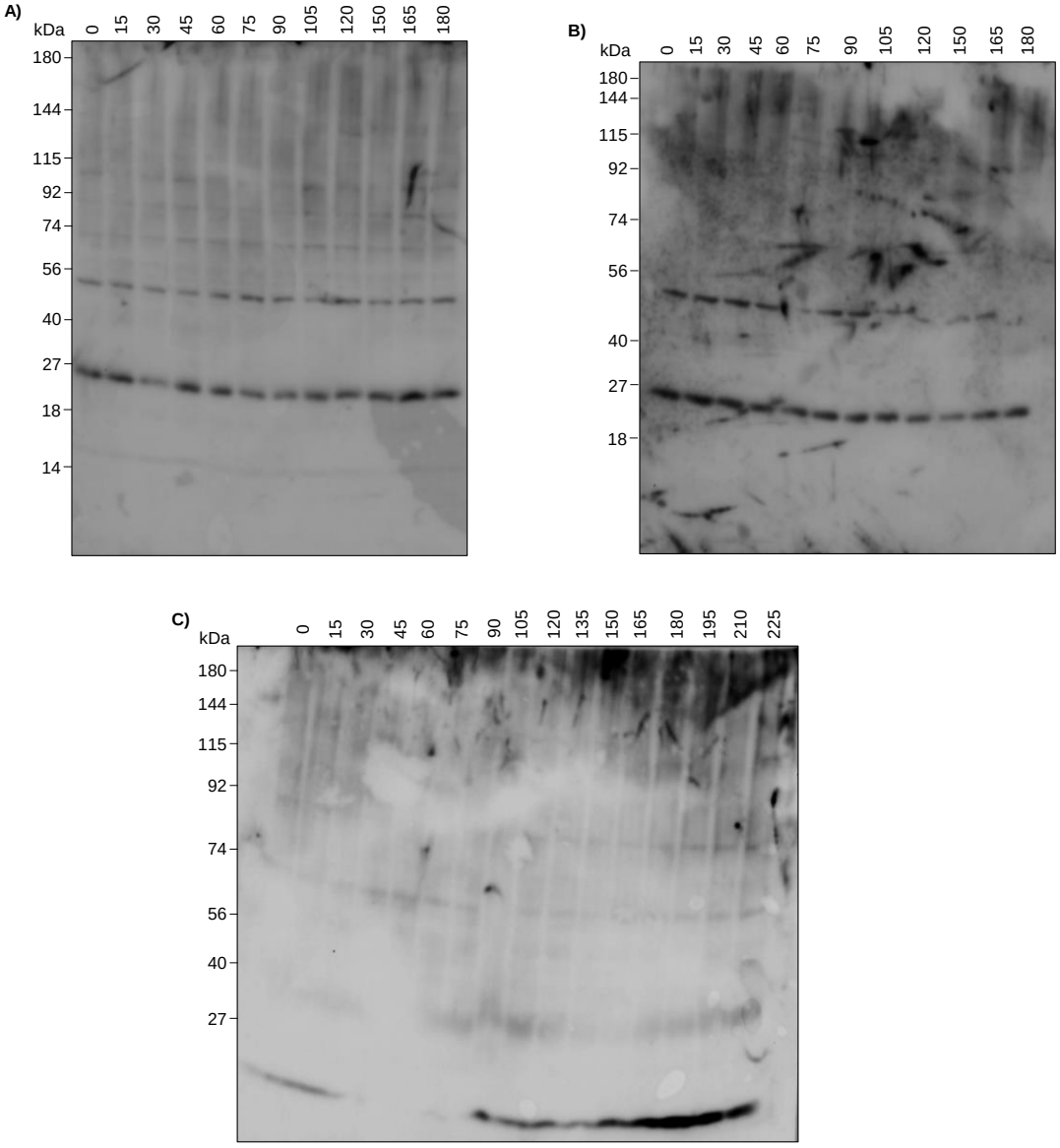


Figure 4-18 Anti-ubiquitin western blots conducted on Leishmania lysates after the addition of various concentrations of cyclophosphamide. A) Cells were exposed to 2.5 mM cyclophosphamide and collected every 15 minutes for three hours. **B)** Cells were exposed to 5 mM cyclophosphamide and collected every 15 minutes for three hours. **C)** Cells were exposed to 10 mM cyclophosphamide and collected every 15 minutes for 3.75 hours. All lane labels refer to minutes post cyclophosphamide addition. All samples were run on 4-20% SDS-PAGE gels, transferred to a PVDF membrane, and the proteins of interest detected using α-ubiquitin antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Building on this, higher concentrations of cyclophosphamide were trialled, 15 and 20 mM, with time points extending until four hours and 55 minutes. This is seen in Figure 4-19. There is possibly a subtle increase in banding in Figure 4-19A at approximately 12 kDa at 105 and 120 minutes; and in Figure 4-19B at 105, 120, and 135 minutes between 40 kDa and 5 kDa. However, any potential increases in ubiquitin are ambiguous at best. As such, the effect of cyclophosphamide on ubiquitin's role in the DNA damage response was not explored further.

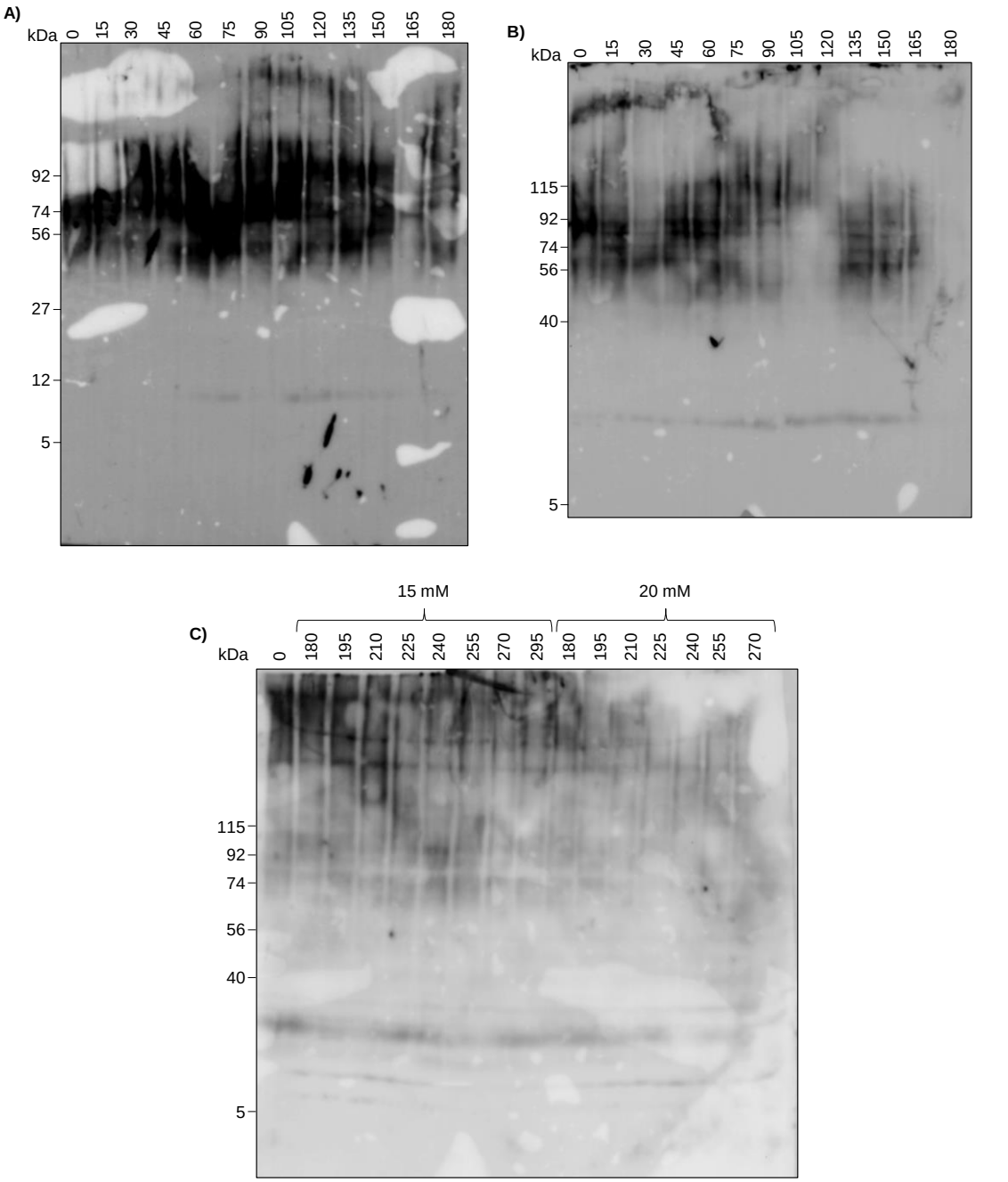


Figure 4-19 Anti-ubiquitin western blots conducted on Leishmania lysates after the addition of 15 and 20 mM of cyclophosphamide. A) Cells were exposed to 15 mM cyclophosphamide and collected every 15 minutes for three hours. **B)** Cells were exposed to 20 mM cyclophosphamide and collected every 15 minutes for three hours. **C)** Cells were exposed to 15 or 20 mM cyclophosphamide and collected every 15 minutes after 3 hours of cyclophosphamide incubation. All lane labels refer to minutes post cyclophosphamide addition. All samples were run on 4-20% SDS-PAGE gels, transferred to a PVDF membrane, and the proteins of interest detected using α-ubiquitin antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

4.3.3.2 Anti-diglycine immunoprecipitations

To catalogue the ubiquitinome, anti-diglycine immunoprecipitations were planned. To perform the immunoprecipitations anti-diglycine antibodies (supplied by Millipore, cat. no.: MABS27) were conjugated to magnetic beads coated with Protein A/G (supplied by ThermoFisher Scientific, cat. no.: 88805).

Figure 4-20 shows the results of this procedure – although there were anti-diglycine antibodies in the initial antibody solution before incubation with beads, there are no detectable antibodies in the solution after incubation with beads, and no antibodies detected in subsequent washes.

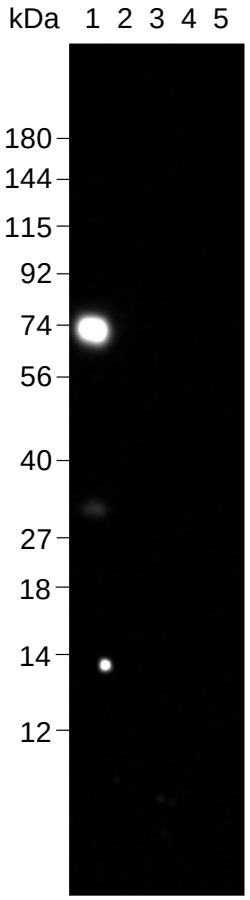


Figure 4-20 Conjugation of anti-diglycine antibodies to Protein A/G beads. Lane 1 is the initial antibody solution before incubation with Protein A/G beads, lane 2 is the antibody solution after incubation with beads. Lanes 3, 4, and 5 are three subsequent washes of the beads after incubation with antibodies. All samples were run on 4-20% SDS-PAGE gels, transferred to a PVDF membrane, and the antibodies detected using α-mouse antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

The anti-diglycine antibodies conjugated to magnetic beads were used to immunoprecipitate ubiquitinated peptides. These peptides came from *Leishmania* cultures that had overnight treatment with the proteasome inhibitor MG132 (to prevent the degradation of ubiquitinated proteins), or treatment with the drug vehicle DMSO. There were eight high confidence (false discovery rate of 1%) diGLY modified peptide spectral matches (PSMs) in the MG132 condition and 5 in the DMSO condition. If the confidence limit is relaxed to include PSMs made at medium confidence then there are 1222 and 1346 PSMs in the MG132 and

DMSO conditions respectively. This data is summarised in Table 4-7. The distribution of unique peptides identified with at least medium confidence (false discovery rate of 5%) can be seen in the Venn diagram in Figure 4-21. This makes it apparent that the MG132 and DMSO treated samples contained distinct populations of peptides with little overlap. It is also worth noting that none of the diGLY modified PSMs resulted in any protein identifications. Which is to say, none of the identifications made at the protein level used diglycine modified peptides.

	MG132	DMSO
High confidence digly modified PSMs	8	5
Medium confidence digly modified PSMs	1222	1346
Total PSMs identified to at least medium confidence	3003	3272
Maximum % of PSMs that may be digly modified	41.0	41.3

Table 4-7 Summary of the first anti-diglycine immunoprecipitation.

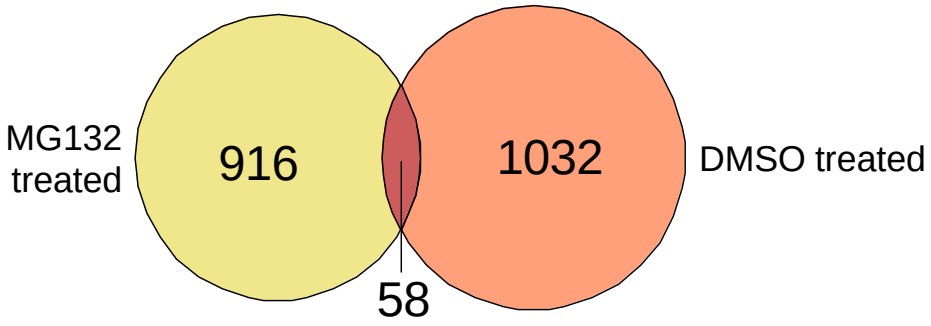


Figure 4-21 Distribution of unique diglycine modified peptides identified to at least medium confidence in MG132 and DMSO treated cells.

This dataset was integrated with the hits from the BioID screen of the E2. There were nine hits that were present in both the N- and C-terminal tagged versions of LmxM.33.0900. Of these, three were also present in the PSMs of the diGLY dataset (33%): a putative dynein heavy chain (LmxM.25.0980), a putative cystathione gamma lyase (LmxM.34.3230), and the E1 of interest to this thesis (LmxM.34.3060). Approximately 16% of the proteins identified by the N-terminal tagged E2 (3 out of 19), and 29% of the proteins identified by the C-terminal tagged E2 (5 out of 17), were present in the PSMs of this dataset. These proteins are highlighted in the BioID tables for LmxM.33.0900, where they are highlighted in red (Table 4-4, Table 4-5, Table 4-6).

A number of alterations were made for a second iteration of this experiment:

- The anti-diglycine antibodies were crosslinked to the magnetic beads (the initial experiment did not have the antibodies crosslinked to the beads). As such, immunoprecipitations could be conducted in series as co-elution of antibodies was not a concern. Therefore, 64 µg of antibody could be used per immunoprecipitation, as opposed to 15 µg when immunoprecipitations were conducted in parallel.

- Amastigotes were used instead of promastigotes.

- Parasites were lysed in RIPA buffer as opposed to CHAPS-based buffer.

- 1 mg of protein was used, whereas the initial study used 200 µg.

- Parasites were incubated with MG132 for six hours as opposed to overnight.

- TMT labelling was employed to quantify any changes induced by MG132.

The effect of these changes is displayed in Table 4-8. Furthermore, the labelling of those diglycine modified peptides was inconsistent, so quantitative information could not be derived. Given the inconsistency of the labelling, it was also difficult to determine which sample a obtained any particular peptide. This project did not proceed further.

High confidence digly modified PSMs	5
Medium confidence digly modified PSMs	2
Total PSMs identified to at least medium confidence	489
Maximum % of PSMs that may be digly modified	1.4

Table 4-8 Summary of the second anti-diglycine immunoprecipitation.

4.4 Discussion:

4.4.1 BioID

Here it is shown that a putative ubiquitin activating enzyme (LmxM.34.3060) and a ubiquitin conjugating enzyme (LmxM.33.0900) feature in the same interactome, validating results from crosslinking studies (see 3.3.2.5). When BLASTed against the human proteome, LmxM.33.0900 shares 42.2% identity with the human ubiquitin-conjugating enzyme E2 J2. A reciprocal BLAST, searching the amino acid sequence of E2 J2 against the *Leishmania* proteome yields LmxM.33.0900 as the top hit. In humans E2 J2 is a component of the endoplasmic reticulum associated degradation pathway, a quality control system that sends misfolded or unassembled proteins to the proteasome for destruction (Lenk *et al.*, 2002). This is also the case for its yeast ortholog UBC6 (Sommer and Jentsch, 1993). Transforming LmxM.33.0900 into its *Trypanosoma brucei* ortholog, Tb927.4.3190, then searching the *Trypanosome* online protein localisation database (http://tryptag.org/) for this ortholog's localisation shows it is associated with the endoplasmic reticulum and nuclear envelope (Dean, Sunter and Wheeler, 2017). E2 J2 is anchored into the endoplasmic reticulum through a transmembrane domain at its carboxy terminal (Claessen *et al.*, 2010). Aligning LmxM.33.0900 against characterised orthologs from humans, mice, fruit flies and yeast shows a conserved hydrophobic C-terminal sequence (Figure 4-22). This implies a transmembrane domain, a transmembrane domain that may anchor LmxM.33.0900 to the endoplasmic reticulum. If LmxM.33.0900 has a similar function as its orthologs, its substrates would likely be varied, given that any protein may be subject to misfolding. It is also interesting that other authors have noted that the minimal endoplasmic reticulum associated degradation networks of protozoan parasites have little redundancy in their components, making them attractive drug targets. As such, a compound that has low nanomolar activity against the *Plasmodium falciparum* (the causative agent of malaria) endoplasmic reticulum associated degradation pathway has been validated as an anti-parasitic (Harbut *et al.*, 2012).

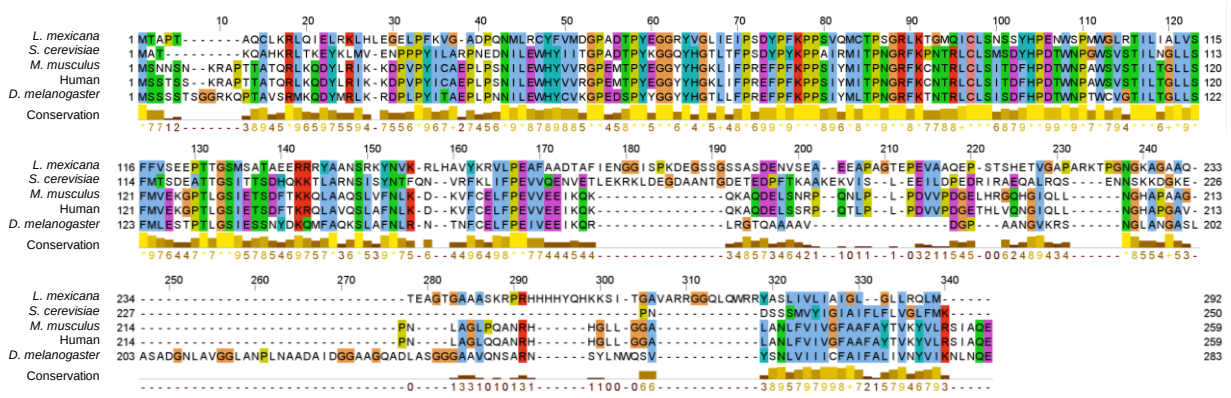


Figure 4-22 Alignment of LmxM.33.0900 to orthologs from *Saccharomyces cerevisiae*, *Mus musculus* (mouse), humans, and *Drosophila melanogaster* (fruit fly). Colour coding according to the default Clustal X Colour Scheme, which colours groups of amino acids based on shared chemical properties – notably, blue indicates hydrophobicity.

The lack of specificity in this hypothetical interactome makes it difficult to compare with the experimentally derived interactome obtained from BioID screens. However, searches of the protein-protein interaction database STRING (https://string-db.org/) did not predict the same substrates as identified by BioID. If further testing of the hypothesis that LmxM.33.0900 features in the endoplasmic reticulum associated degradation pathway is desired, perhaps the endoplasmic reticulum could be stressed, possibly by hydrogen peroxide. This could increase the number substrates that LmxM.33.0900 would ubiquitinate (Lam *et al.*, 2014). Having stated how important the carboxy terminal is, it is also worth considering what effects conjugating a BirA* may have on the putative E2. Furthermore, if the carboxy terminal is a transmembrane sequence then a BirA* conjugated to it would access an entirely separate space to a BirA* conjugated to the N-terminal, which extends into the cytoplasm. If this is true, it is perhaps surprising that approximately 50% of the proteins obtained when tagging one terminal were identified when the opposite terminal was tagged (approximately 18 proteins were identified with any one tag, of which 9 are found with both tags). An additional consideration would be that an E2 and a BirA* may compete for substrate lysines, the former to add ubiquitin and the latter adding a biotin to the lysine's ε-amino group. This could reduce the efficiency of any E2 or E3 BioID screen.

The BioID screen of the putative E1, LmxM.34.3060, yielded fewer proteins shared between the N- and C-terminal tags. Perhaps this is unsurprising, as an E1 would be predicted to have fewer substrates than an E2 – namely ubiquitin and E2s. As such, it is something of a concern that neither E2s nor ubiquitin

consistently appeared in the results (ubiquitin not appearing at all, as was the case with the putative E2). Although the E1 itself did consistently appear in its own datasets, the BirA* likely biotinylating its fusion protein partner, and so providing confidence that the BioID system was working. Another identification of note is LmxM.34.1300, a putative E2 identified in the bioinformatic screen (see 3.3.1), was found in the BioID dataset when the N-terminal of the E1 was tagged. This putative E2 was also identified by the protein-protein interaction database STRING as a likely interactor. However, this is the first time LmxM.34.1300 has been seen in data sets from the wet lab, not being present in the activity-based probe studies, or even when the putative E1 is tagged at its C-terminal. A protein that did appear with tags at both termini was LmxM.36.1370, a putative p97 ortholog, p97 is also known as valosin containing protein or Cdc48. This is fascinating as p97 is also part of the ubiquitination pathway, in which one of its functions is to transport ubiquitinated substrates to the proteasome for degradation. Additionally, it has a characterised role in the endoplasmic reticulum degradation pathway, extracting misfolded proteins from the endoplasmic reticulum membrane (Sun and Qiu, 2020). This is curious as why would a protein that functions at the very beginning of the ubiquitin-proteasome pathway, an E1, associate with a protein at its end, p97? To the best of the author's knowledge, there is no precedent for a complex that features both E1 and p97 orthologs. Perhaps then E1 and p97 do not function as part of a larger complex, but the E1 is a substrate of p97. If the E1 were to be misfolded, as could happen when a 41 kDa BirA* is fused to it, then p97 may be responsible for degradation of the malformed protein. An alternative hypothesis is that p97 is nothing more than noise. After all, p97 is one of the most abundant proteins in human cells, and is constitutively expressed in all eukaryotic cells (Ye, 2006). Should p97 be as abundant in *Leishmania*, the BirA* tagged E1 may simply bump into it, without any biochemical or biophysical reactions taking place. The control, an anti-biotin pulldown, merely controls for endogenous biotinylation, so does not control for the Brownian motion of a biological molecule. To control for the riotous maelstrom of collisions in a biological system a BirA* by itself or as a component of a fusion protein unrelated to ubiquitination, would need to be expressed in the same locale as the target protein. However, perhaps an answer can be inferred from the E2 data – this does not feature p97, if it did then p97 could be dismissed as noise (although its absence from the E2 dataset does not

necessarily mean p97 can be interpreted as signal, given potentially distinct localisations of the E1 and E2). What both the E1 and the E2 have in common is LmxM.25.1015, the putative gamma subunit of protein transport protein Sec61. Assuming that the E1 and E2 have different substrates and do not form an E1-E2-substrate complex (which would be novel), the presence of LmxM.25.1015 in the interactomes of both E1 and E2 enzymes may indicate it is noise, merely an abundant protein.

An easy improvement to this method would be to deplete the media of biotin before the labelling reaction, which could lower the background biotinylation of proteins unrelated to ubiquitination. Further work to be done on this system would revolve around applying the improved transfection methods from miniTurbo to use with the standard BirA*. It is assumed that generating a clonal culture, in this case using serial dilution, would increase the signal from a BioID experiment. An alternative method to generate clonal cultures, in which both alleles of a gene of interest are tagged, would be to use two selectable markers – perhaps puromycin and blasticidin resistance genes – thereby forcing a cell to insert two sequences of interest.

4.4.2 Anti-diglycine immunoprecipitations and the DNA damage response

At best, the take home message from this project is that some peptides are diglycine modified. A very general statement, that will not surprise anyone. Perhaps the most surprising thing about this project is cyclophosphamide's lack of effect on ubiquitination in *Leishmania*. In its role as a DNA alkylating agent, cyclophosphamide was thought to induce interstrand crosslinks in DNA (Kondo *et al.*, 2010). This is a particularly pernicious form of DNA damage, demanding a variety of proteins, functioning in a variety of pathways, to repair the damage (Hashimoto, Anai and Hanada, 2016). In humans, one repair complex that involves ubiquitination is the Fanconi anaemia complex, which are eight proteins that constitute a multi-subunit ubiquitin E3. This complex monoubiquitinates another Fanconi anaemia protein FANCD2, which colocalises in nuclear foci with other proteins involved in DNA damage repair (Jacquemont and Taniguchi, 2007). Other proteins linked with DNA damage are the E2 RAD6 and the E3 RAD18, which monoubiquitinate PCNA (Hoege *et al.*, 2002). Yet, in this study, no

substantial increase in ubiquitination was seen upon addition of a DNA damaging agent. How could this be? Are *Leishmania* remarkably resistant to cyclophosphamide? Or do they not respond to interstrand crosslinks with ubiquitination? Future studies would do well to try other DNA damaging agents, but still with the aim of inducing interstrand crosslinks. Components of a stripped down Fanconi anaemia complex have been found in *C. elegans*, *D. melanogaster*, as well as budding and fission yeast (McHugh, Ward and Chovanec, 2012). If *Leishmania* does not possess any orthologs it would have diverged significantly from other eukaryotes. This in itself would be notable.

The anti-diglycine immunoprecipitations yielded approximately 1300 PSMs with a false discovery rate of 5%, in both drug and vehicle treated conditions. However, turning PSMs into protein identifications proved to be challenging, as mass spectrometric software typically requires multiple peptides from any given protein in order to provide a confident protein identification. The experimental method demanded selecting for modified peptides as opposed to proteins, and, due to their stoichiometry, there is unlikely to be more than one ubiquitinated peptide per protein. Yet this method has been published elsewhere (Xu, Paige and Jaffrey, 2010; Rose *et al.*, 2016; Udeshi *et al.*, 2020). Although the key distinction between those studies and this one remain unknown, the vast majority of published methods use a kit from Cell Signalling Technology (product number 5562), whereas a home-made kit was used here. Although wet lab reagents do not assist with dry lab data analysis, they do provide data to analyse. Perhaps a store brought kit results in more peptides, some of which are unique and fragment well, and so more confident identifications can be made in the dry lab. A further concern is that the maximum percentage of diglycine spectral matches dropped precipitously from 41% to 1.4%, despite what were considered logical improvements between experiments. It seems unlikely that using more protein, or using amastigotes as opposed to promastigotes, caused this decreased efficiency. A potential cause is the act of crosslinking the antibodies to beads – anecdotally, in other cases, the author has been informed that crosslinking antibodies to these agarose beads caused a sharp drop in efficiency. It is also worth noting that the kit from Cell Signalling Technology does not have the anti-diglycine antibodies crosslinked to the agarose beads. This is likely because, while a chemical crosslinker may covalently bind an

antibody to Protein A/G, it will also cause intra-antibody crosslinking, which could affect the structure of the antibody. Additionally, although the combination of TMT labelling with diglycine peptides has been published (Rose *et al.*, 2016), it is not in frequent use, whereas the majority of the literature uses label free methods. This could be due to the relative recency of the publication of isobaric tags being used with diglycine peptides, and old habits of label free quantitation dying hard. Or it could be due to the risk of the two amino termini on diglycine modified peptides (one on the peptide backbone, one on the diglycine residues) both being labelled and confounding data analysis. Whatever the reason, the addition of TMT labels was one factor among many that preceded poorer quality data. Given the failure of stimuli to increase ubiquitination (both cyclophosphamide and MG132), and the failure to immunoprecipitate diglycine modified peptides to a high confidence, this project was abandoned.

4.5 Summary

This study has presented the interactomes of two putative components of *Leishmania*'s ubiquitination pathway. It was hypothesised that such interactomes would feature other components of the ubiquitin-proteasome system. This was shown to be true, in which an E2 was shown to associate with an E1, and the E1 shown to associate with another E2. However, this study would benefit from clear evidence that the BirA* gene was inserted at the correct locus, something that PCRs conducted on transfectant DNA were unable to demonstrate here.

This study has presented the number of ubiquitinated peptides obtained from immunoprecipitations, and the impact of cyclophosphamide on ubiquitination. It was hypothesised that cyclophosphamide would perturb ubiquitination. This was not proven, as western blots did not show any notable changes in ubiquitination. It was hypothesised that the substrates of ubiquitin could be identified from immunoprecipitations. This was true to an extent, as >1000 peptides were identified to a 5% false discovery rate. However, these peptide spectral matches were not turned into protein identifications, and attempts to quantify them failed.

Future work should focus on validating the BioID screen and troubleshooting the anti-diglycine immunoprecipitations.

Chapter 5 Biophysical characterisation of LmxM.34.3060

5.1 Introduction

The kinetoplastid ubiquitination system comprises multiple druggable targets (Gupta *et al.*, 2018; Boer and Bijlmakers, 2019; Bijlmakers, 2021). Previous studies have identified some of these targets, not least the ubiquitin activating enzymes. This chapter will confirm the results of bioinformatic and proteomic screening (Chapter 3), as well as provide further insight into drug development, through enzyme expression, and the development of *in vitro* functional assays. In these assays an inhibitor of the human ortholog of LmxM.34.3060 was tested (TAK-243), as well as structural analogs nucleocidin and 5′-Sulfamoyladenosine. Furthermore, biophysical characterisation of LmxM.34.3060 was conducted to aid rational drug development. This characterisation relied on small angle X-ray scattering, protein crystallography, and drug modelling.

5.1.1 Ubiquitination inhibitors

Activation of ubiquitin is a multistep process, demanding a multidomain enzyme. Broadly, the role of a ubiquitin-activating enzyme, an E1, can be divided into four distinct processes (Schulman and Wade Harper, 2009):

1. Adenylation of the first ubiquitin

2. Thioester formation of the first ubiquitin to the E1 cysteine active site

3. Adenylation of the second ubiquitin

4. Transfer of the first ubiquitin to the E2 by transthiolation

The complex biochemistry at play provides ample opportunity for inhibition, a point already demonstrated by the numerous inhibitors of multifarious mechanisms. Figure 5-1 lists some these inhibitors alongside their sites of action. To start with an obvious, if not easy, target: the catalytic cysteine at the cysteine active site. The thiol group, that which defines cysteine and renders it crucial in biology, offers a target for an electrophillic inhibitor. Such an inhibitor could covalently modify the active site, inflicting irreversible and potent inhibition on its target (Visscher, Arkin and Dansen, 2016).Yet, this strength is also their greatest weakness – cysteine reactive sites are abundant throughout

nature, so an off-target effect of an irreversible inhibitor could have dire consequences for an organism. Nevertheless, the Food and Drug Administration have approved the kinase inhibitors rociletinib and osimertinib, which selectively target cysteine residues to a high degree (Visscher, Arkin and Dansen, 2016). More pertinant are the inhibitors PYR-41 (Yang *et al.*, 2007) and PYZD-4409 (Xu *et al.*, 2010), both of which inhibit ubiquitin E1s. Both compounds preferentially kill malignant cell lines, although their exact mechanism of action is obscure (Huang and Dixit, 2016). However, there are concerns about their specificity – for example, despite a decrease in UBA1 activity, there is an increase in ubiquitinated proteins with mass >250 kDa, explained by PYR-41 interacting with deubiquitnating enzymes (Barghout and Schimmer, 2021). As such, the anti-cancer effect of PYR-41 may be due to inhibition of off-target proteins, alongside UBA1.

A far more difficult, yet conceptually interesting, target would be the ubiquitin fold domain, the E2's docking site integral for the ubiqutin transfer. The same problem applies here as when targeting most protein-protein interactions: a vast surface area deplete of the pockets necessary for small molecule inhibitors (Jin, Wang and Fang, 2014).Even so, a peptide that inhibits the NEDD8 E1-E2 interaction, composed of the 26 amino acids in the E2's N-terminal extension, has generated structural insights into the selectivity of the NEDD8 pathway (Huang *et al.*, 2004).

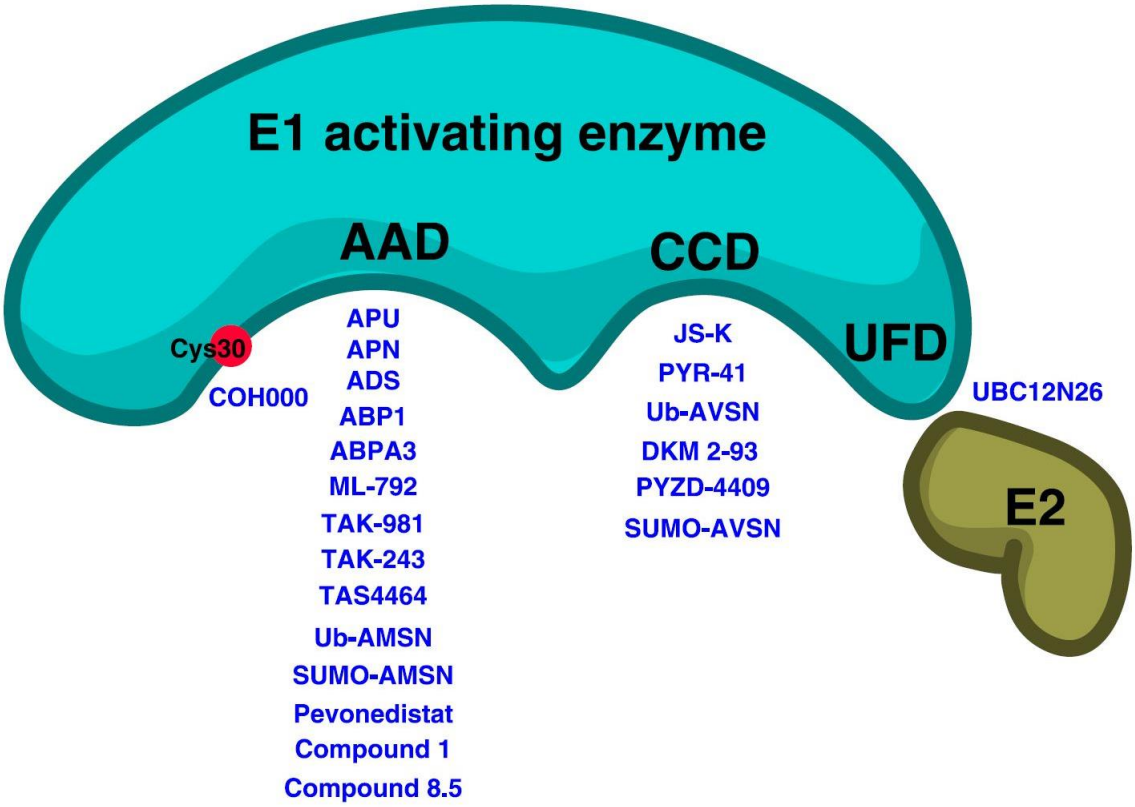


Figure 5-1 Inhibitors of ubiquitin and ubiquitin-like activating enzymes, and their sites of action. AAD: Active Adenylation Domain, CCD: Catalytic Cysteine Domain, UFD: Ubiquitin Fold Domain. Figure taken from Barghout and Schimmer (Barghout and Schimmer, 2021).

Given the drawbacks when focusing an inhibitor on the catalytic cysteine or ubiquitin fold domain, it is no wonder that greater attention has been directed towards the adenylation domain. The precedent for small molecule ATP mimetics was set by targeting tyrosine kinases in cancer therapy, when the anti-tumour drug Glivec was rationally designed to interact with an ATP-binding pocket (Capdeville *et al.*, 2002). As such, it does not require a stretch of the imagination to consider an ATP mimetic that might interact with an E1's ATP binding pocket. Furthermore, the unique mechanism, whereby AMP is conjugated to ubiquitin at the ATP binding site, represents an opportunity to develop inhibitors with little impact on other ATP dependent enzymes. Yet, despite all the promise, we are left with the challenge of selecting for one E1 among many – whether that is targeting a ubiquitination E1 instead of its NEDD8 equivalent, or targeting a *Leishmania* E1 instead of its human ortholog.

Nonetheless, selectivity for specific E1 enzymes has been demonstrated. Firstly, Ubl~AMP analogs, such as those activity based E1 probes discussed in Chapter 3, inhibit specific E1s at the adenylation site (specificity determined by the proteinaceous Ubl component) (Lu *et al.*, 2010). The trouble is the size of the

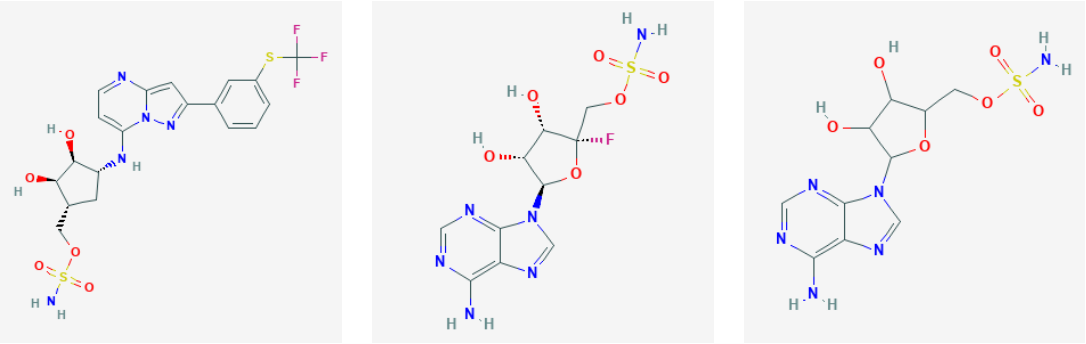
molecule – it is difficult to pass an entire protein through a cell's membrane, and as such it has little use as a drug. Nonetheless, of note is this molecule's chemical warhead: a sulfamide group. The closely related sulfamate group (the difference between the sulfamide and sulfamate groups being an amino group in the former replaced with an oxygen atom in the latter), when conjugated to a nucleobase analog, permits an E1 to synthesise its own Ubl~AMP analog inhibitor *in situ*. This principle, in the form of a compound termed MLN4924, was used to inhibit the NEDD8 E1, which in turn inhibited the growth of human tumour xenografts on mice (Soucy *et al.*, 2009). Currently, MLN4924 has progressed to phase III clinical trials against leukemias (https://clinicaltrials.gov/, 2021). The recently characterised and related compound TAK-243 inhibits the human ubiquitin activating enzyme UBA1 (Hyer *et al.*, 2018), and has succeeded in phase I trials against advanced solid tumours (Barghout and Schimmer, 2021).

Why does this sulfamate group conjoined to a nucleobase analog, often adenosine, work? That is due to its mimcry of AMP. Imitating the phosphate group in the native Ubl~AMP intermediate, the sulfamate's nitrogen forms a covalent bond with ubiquitin's C-terminal glycine (Hyer *et al.*, 2018). How does this mimicry inhibit an E1? In a mechanism so novel it deserved its own name – substrate assisted inhibition (Brownell *et al.*, 2010; Chen *et al.*, 2011). When the Ubl is bonded to the E1's cysteine active site, the now empty nucleotide binding pocket of the E1 binds the sulfamate based inhibitor. Once in place, the sulfamate group acts as a nucleophile, attacking the thioester bond joining Ubl to E1. Thus, a Ubl-inhibitor adduct is formed. This adduct then moves to the adenylation site, forming a quasi-stable complex that precludes the further uptake of either ATP or Ubl by the E1 (Brownell *et al.*, 2010).

Yet, this well characterised mechanism does not explain selectivity – all E1s have corresponding Ubls and adenylation pockets, so all could be inhibited by the same sulfamate inhibitor. This can be avoided thanks to extensions applied to the nucleobase analog, which, through multiple non-covalent forces, interact more or less favourably with the variable ATP binding pockets of E1's (Misra *et al.*, 2017). As such, connecting a sulfamate warhead to a nucleobase analog represents a general mechanism for potent yet specific inhibition of E1s. This provides a need for detailed examination of the ATP binding pocket of any E1 for

which an adenosine sulfamate inhibitor is to be designed. A need which demands both biochemical and biophysical methods.

The compounds 5'O-sulfamoyl adenosine and nucleocidin share TAK-243's adenosine sulfamate group, but lack TAK-243's extension after the nucleobase. (Figure 5-2). Nucleocidin is an antibiotic produced by *Streptomyces calvus*, a bacteria isolated from soil in Dinepur, India (Thomas *et al.*, 1956). It was shown to be a potent anti-trypanosmal compound (trypanosomes being a sister organism of *Leishmania*), but was reported to be too toxic to the host for clinical use. It is thought to have a mechanism that inhibits protein synthesis, though as Florini, Bird, and Bell note in their publication on the topic: "Nevertheless, it remains difficult to understand how inhibition of protein synthesis can result in death in 2 hours" (Florini, Bird and Bell, 1966). Other authors have noted the potential of the sulfamate group to inhibit E1s, as Brownall *et al.* quote Bloch and Coutsogeorgopoulos when they write: "We suspect that this pan-specific E1 inhibition may explain why one such analog, the natural product nucleocidin … gained notoriety as 'One of the most toxic nucleosides encountered in nature.'" (Bloch and Coutsogeorgopoulos, 1971; Brownell *et al.*, 2010). Lux, Standke, and Tan hypothesise the same in their review article of sulfonyladenosine inhibitors (Lux, Standke and Tan, 2019). 5'O-sulfamoyl adenosine is a readily synthesised analog of nucleocidin, lacking only nucleodin's fluorine atom, also thought to inhibit protein synthesis (Shuman, Robins and Robins, 1969; Bloch and Coutsogeorgopoulos, 1971). Experimental work, originally conducted in 1971 or earlier, has focused on the ability of 5'O-sulfamoyl adenosine and nucleocidin to inhibit protein synthesis. References to the compounds in regards to E1 inhibition have been untested hypotheses.



TAK-243 Nucleocidin 5'O-sulfamoyl adenosine

Figure 5-2 The compounds TAK-243, nucleocidin, and 5'O-sulfamoyl adenosine. All compounds share a sulfamate group (H_2NSO_3), but nucleocidin and 5'O-sulfamoyl adenosine both lack an extension off of their nucleobase group (present in TAK-243). Compound structures were taken from the PubChem database (*[(2R,3S,4R,5R)-5-(6-Amino-9H-purin-9-YL)-3,4-dihydroxytetrahydro-2-furanyl]methyl sulfamate | C10H14N6O6S - PubChem*, no date; *Tak-243 | C19H20F3N5O5S2 - PubChem*, no date; *Nucleocidin | C10H13FN6O6S - PubChem*, no date).

5.1.2 Protein Expression And Purification

Nothing embodies the revolutionary power of recombinant protein expression better than insulin. When insulin production switched from mammalian pancreatic extracts to recombinant bacterial and yeast systems the gulf between supply and demand closed (Ladisch and Kohlmann, 1992). This exemplifies a broader transformation in biochemistry, enabled by researchers expressing and purifying proteins in industrial quantities. On a theoretical level, the steps necessary for successful expression of recombinant protein are simple: take your gene of interest, clone it into an expression vector, transform the construct into a relevant host, induce protein expression, purify the protein. However, this simplification belies the necessary optimisation and frequent troubleshooting steps.

Nonetheless, the outline of a protein expression project is defined by the host organism, as all molecular tools, reagents, and equipment must be compatable with the machinery used to synthesise the protein of interest. Due to a plethora of advantages, a good starting point for protein expression is often found in *Escherichia coli*. Examples of these advantages include a doubling time of 20 minutes (given an optimised environment) (Sezonov, Joseleau-Petit and D'Ari, 2007), achievable high cell densities in laboratory cultures, readily available inexpensive media, and simple transformation with exogeneous DNA (Rosano and Ceccarelli, 2014).

To use *E. coli*, an expression vector must be chosen from a vast catalogue. A commonly used vector series are termed pET. Their common usage can be justified through their application of the T7 transcription and translation regulatory system, which permits strong, tightly controlled expression of the protein of interest (Studier *et al.*, 1990; Dubendorf and Studier, 1991). Under the T7 system, expression of the recombinant protein is typically induced through addition of isopropyl ß-D-1-thiogalactopyranoside (IPTG) (Rosano and Ceccarelli, 2014). After expression, the challenge becomes separating the protein of interest from the medley of proteins in an *E. coli* lysate. To aid in this, it is invaluable to have an affinity tag on the expressed protein. One such tag is the polyhistidine, or His, tag. As the name would suggest, consecutive histidine residues are attached to a recombinant protein's terminus, permitting its immobilisation on a fixed matrix of transition metal ions, in a process known as Immobilised Affinity Metal Chromatography (Terpe, 2003). Unfortunately, it is likely further rounds of purification will be necessary. Additional orthogonal variables can be selected for with techniques such as size exclusion chromatography, in which proteins are fractionated according to size, due to a relationship between protein size and protein flow rate through porous beads (Burgess, 2018). Ion exchange chromatography enables fractionation by isoelectric point, where, at a given pH, a protein is electrostatically bound to a column possesing an opposite charge. Through varying the ionic strength of the elution buffer, the protein is eluted (Ullah, 2012). Thanks to this variety of techniques polypeptides are produced in the vast quantities suitable for industrial and commercial activities, and have driven academic research for over 40 years (Itakura *et al.*, 1977).

5.1.3 Small Angle X-ray Scattering

In the appropiately titled 1929 article "A New X-ray Effect", Raman and Krishnamurti note note how smaller particles of graphite powder scatter X-rays more intensely than larger particles, with this intense scattering visible at small angles from the primary X-ray beam (Raman and Krishnamurti, 1929). Although this was the first experimental observation of X-rays scattering at small angles, the theoretical framework waited for André Guinier to write his doctoral thesis on the subject. A thesis, successfully defended in 1939, that gave the world Guinier's approximation (more on which below), a powerful tool that remains in use over eight decades later (Ravy, 2015). It was Guinier's insight that first

permitted interpretation of the diffuse X-ray scattering signals found close to the primary beam, and extended their uses into the biological realm. Among Guinier's realisations were that biological macromolecules in solution would be ideal systems to be studied by the then new method, due to both their typical size ranges, and the potential for highly pure samples (Weiss, 2017). Thus, a new biophysical method was born: Small Angle X-ray Scattering (SAXS).

SAXS is based on the elastic scattering of X-rays by macromolecules: when a monochromatic wave hits an object, its constituent electrons become sources of secondary waves. The scattering of interest in SAXS occurs due to the discrete electron density contrast between the solvent and macromolecule, which is to say the difference in the number of electrons per unit volume between the sample and the solvent (Putnam, Lowe and Meiler, 2013). As such, the parameter that demands measurement in a SAXS experiment is momentum transfer (q), shown in equation 1. This shows how momentum transfer relates to the wavelength of the incident X-rays (λ) and half their scattering angle (θ). Figure 5-3 demonstrates how this measurement is made.

$$q = \frac{4\pi \ \sin\theta}{\lambda} \qquad \textbf{Equation 1}$$

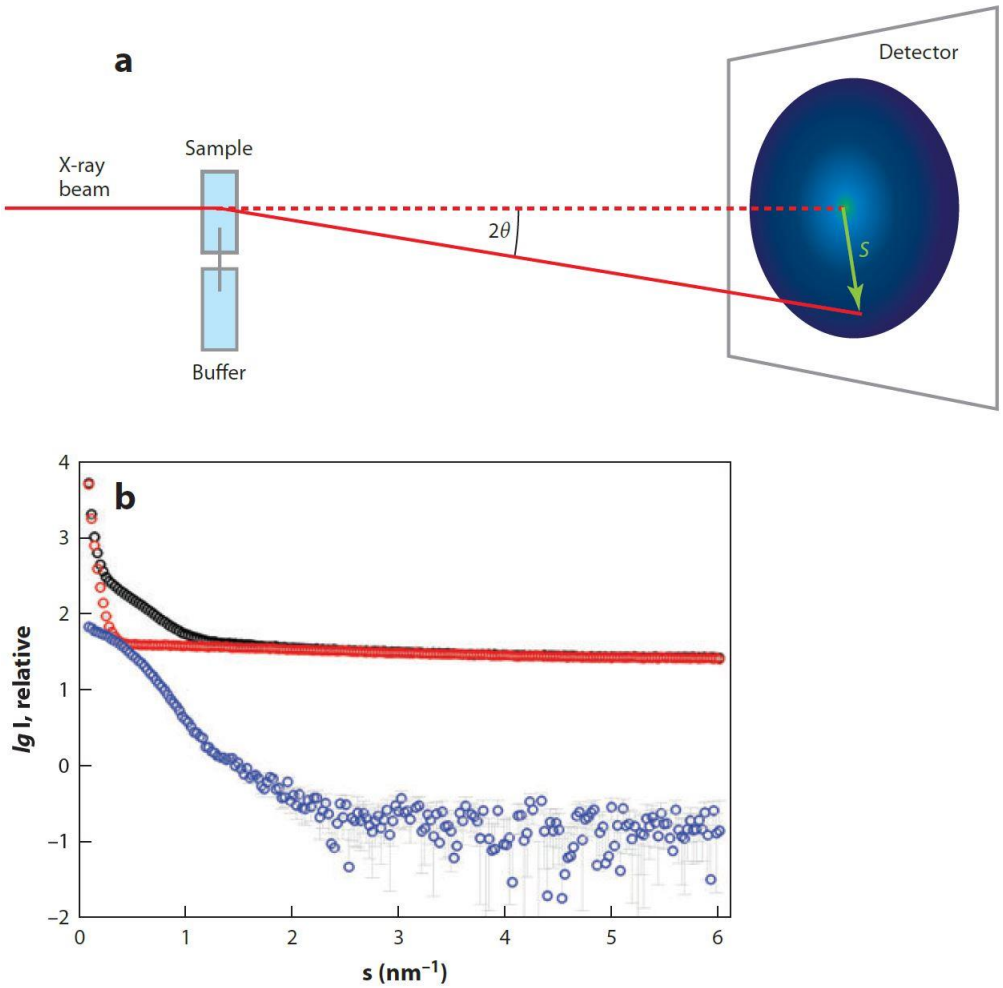


Figure 5-3 A typical SAXS experiment. a) A monochromatic beam of X-rays hits a sample, and scattered photons are collected by a 2D detector. The data is radially averaged to obtain the scattering curve. **b)** The scattering curve of a protein sample (blue data points) is obtained after subtracting the solvent scattering (red data points) from the sample scattering (black data points). This figure is taken from Blanchet and Svergun, and refers to momentum transfer as s (as opposed to q) (Blanchet and Svergun, 2013).

From here, the scattering intensity, I (q), can be calculated. It is proportional to the scattering from a single macromolecule averaged over all orientations, and is found through the Debye equation (equation 2). This shows how the sum of scattering intensity from *N* atoms is multiplied by the scattering factors (*f*), and the distance between the scattering atoms (r_{ij}). The free movement and random orientations of biological molecules necessitate averaging scattering data over all orientations, meaning detailed orientational information is lost. Yet information on interatomic distances is maintained, which is enough to calulate serveral structural parameters, and even a low resolution shape of the macromolecule of interest.

$$I(q) = \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}} \qquad \textbf{Equation 2}$$

Still, the most taxing difficulty with SAXS is not data collection, but data analysis. Typically, analysis starts with Guinier's approximation (equation 3). This allows estimation of the Radius of Gyration (R_g) and the forward scattering, called *I*(0). R_g is the average root-mean- square distanced to the centre of density in the molecule weighted by the scattering length density, which is to say the particle's distribution of mass around its center, or, even more crudely, its overall size. *I* (0) refers to the scattering that would take place if q = 0 could be measured (given the beamstop, scattering at q = 0 cannot be directly observed), it is proportional to the molecular weight and concentration of the protein (Putnam, Lowe and Meiler, 2013; Kikhney and Svergun, 2015).

$$\ln I(q) = \ln I(0) - \frac{{R_g}^2}{3} \times q^2 \qquad \textbf{Equation 3}$$

However, Guinier's linear relationship only holds when $q \times R_g < 1.3$ (for a spherical protein at least). As such, R_g and *I*(0) are estimated from the low q portion of the scattering data (after excluding any noise from the beamstop), and are found, respectively, through the slope and intercept of the resulting linear plot (Kikhney and Svergun, 2015). The Guinier plot can also be used to infer sample quality – a non-linear Guinier region could be indicative of poor sample quality (or a very elongated protein). Poor sample quality refers to non-specific interactions between particles, such as aggregation (increased scattering at low angles) and inter-particle repulsion (decreased scattering at low angles) (Blanchet and Svergun, 2013). Figure 5-4a shows Guinier plots for samples with and without linear Guinier regions.

Despite the many uses of a Guinier plot, the R_g and *I*(0) values obtained are approximations. More accurate assessments can often be obtained through the pair-distance distribution, p(r) (equation 4), as all the scattering data may be used (as opposed to the Guinier approximation, which holds only for low values of q).

$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin(qr)}{qr}dq \qquad \textbf{Equation 4}$$

The p(r) function provides a histogram of distances between all possible pairs of atoms in a particle. This is a representation of the SAXS data in real space, achieved through an indirect Fourier transformation (Kikhney and Svergun, 2015). A spherical particle will yield a bell shaped distribution, which smoothly finishes at the maximum distance between atoms, the D_{max}. Again for a spherical particle, its peak should be at approximately half the D_{max} (Blanchet and Svergun, 2013). Figure 5-4c provides example p(r) distributions for various particle morphologies.

Yet more information on morphology can be deduced from a Kratky plot, which enables distinction of a compact, globular protein from a disordered, unfolded one. For example, the scattering profile for a compact protein decays at $1/q^4$, whereas the profile for disordered protein decays at 1/q. Furthermore, a compact protein often yields a Kratky plot that peaks at low angles of q, as opposed to disordered proteins that do not have a defined peak, preferring to increase continously at high q (Blanchet and Svergun, 2013). So, a globular protein will display a bell-shaped Kratky plot with a well defined maximum (Figure 5-4b). An additional note, if one wishes to compare Kratky plots of different proteins: it is benficial to multiply q by R_g prior to plotting. This normalises the data, and results in a dimensionless Kratky plot (Kikhney and Svergun, 2015).

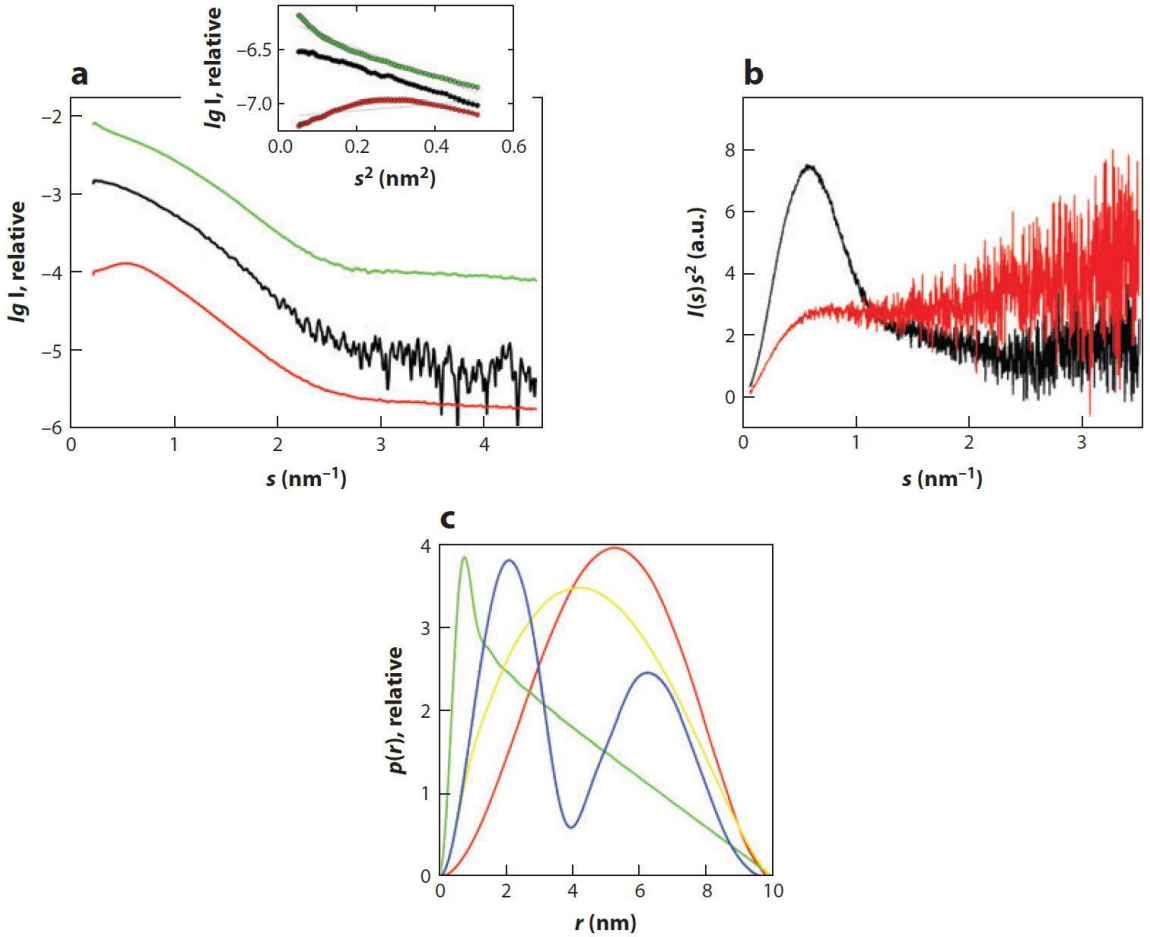


Figure 5-4 Examples of SAXS data from various proteins. a) SAXS scatter curve, with Guinier plot (inset). An ideal protein solution is in black, a protein solution with non-specific aggregation in green, and a protein solution subject to inter-particle replusion in red. **b)** Kratky plot for a folded protein (black) and an unfolded protein (red). **c)** Example p(r) distributions for a spherical (red), dumbbell-shaped (blue), cylindrical (green), and disk-shaped (yellow) proteins. This figure is taken from Blanchet and Svergun (Blanchet and Svergun, 2013), and refers to momentum transfer as s (as opposed to q).

Nevertheless, no number of biophysical equations can tell us why we should use SAXS. Why bother to apply a method that can only ever provide a low resolution structure? The value of SAXS lies in the samples it examines – proteins in solution. Proteins do not merely wallow around in their chosen solvent. It is a fundamental part of the biochemistry of life, life and liquid interwoven with such complexity and subtlety that the solvent is part of the machinery (Cockell, 2018). Contrast this with protein crystallography, which can provide atomic level detail in a highly artificial environment (there are rarely protein crystals *in vivo*, and rarer yet if those crystals are non-pathological), and SAXS is a powerful addition to the biophysicists toolkit.

5.1.4 Protein Crystallography

In 1912 Friedrich, Knipping, and Laue made the seminal discovery of the pronounced diffraction patterns exhibited when X-rays passed through crystals of copper sulphate pentahydrate and zinc sulphide; a discovery that pleased both physicists (providing evidence of X-rays as electromagnetic waves), and crystallographers (as evidence that atoms arrange in a space-lattice configuration in crystals) (Eckert, 2012). However, it was graduate student Lawrence Bragg who provided a mathematical framework to explain diffraction patterns of the like observed by Laue (Thomas, 2012). Yet when Bragg employed his eponymous law to solve the arrangement of atoms in a crystal of sodium chloride, using its X-ray diffraction pattern, the results were contentious. His proposed structure, a three-dimensional checkerboard arrangement of sodium and chloride ions, was attacked as “more than repugnant to common sense” by the top chemists of the day (Armstrong, 1927). Yet, as we know now, Bragg was correct. From his theory of X-ray crystallography, Lawrence Bragg went on to become the youngest science Nobel laureate, at the mere age of 25 (Ramakrishnan, 2018). This is emblematic of the effectiveness of protein crystallography, which makes the invisible visble.

To see protein structure at the atomic level is to have insight into enzymatic mechanisms, protein interactions, and a foundation for rational drug design. Yet, despite its manifold uses, there is no comprehensive theory regarding protein crystallisation. As such, crystal growth is largely empirical in nature, demanding a user’s intuition. In the absence of insight (or even with), a good approach is largely trial-and-error based. *A priori* knowledge of crystallisation conditions would be hard-won, as even for a small protein, such as myoglobin, there are roughly a thousand atoms, resulting in thousands of bonds, with thousands of degrees of freedom (McPherson and Gavira, 2014). Furthermore, the non-biological parameters that affect crystallisation are diverse, including, but not limited to, incubation time and temperature, precipitant type and concentration, mother-liquor volume and macromolecule concentration. Due to the vast range of parameters that require searching, extensive series of parrallel trials are often performed, with successive rounds of trials building on any promising results (McPherson and Gavira, 2014).

So a general approach to obtaining protein crystals demands as much protein as patience from the investigator. But what is happening in a crystal screening tray? In one way or another, the end goal is to create supersaturation, significantly reducing the solubility of the protein in solution. There are diverse approaches to accomplish this, but a popular way is vapour diffusion – demanding a small droplet (1 – 10 µl) of protein solution and a larger liquid reservoir (0.5 – 1 ml), which often contains a precipitant, salt, buffer, and ligands. The protein solution is mixed 1:1 with the precipitant solution, then equilibrated against the larger liquid reservoir. Given time, hours or days, water is lost from the protein solution and gained by the liquid reservoir. Correspondingly, the concentration of reagents in the 1:1 protein solution:reservoir mix increases (McPherson, 2004). With luck, this should aid crystal growth, a process that occurs with two distinct yet inseperable stages: nucleation and growth. Intriguingly, nucleation remains one of the least understood processes in nature (Vekilov, 2016). Still, classical nucleation theory holds that mesoscopic, disordered protein clusters form in a dense liquid, after which crystal nuclei form within the clusters (Vekilov, 2016).

This brings us to crystal growth, which requires lower supersaturation levels than nucleation, and features as many mechanisms as there are paramerters to vary (a lot). Although the precipitant is often key to crystallisation, for brevity, only two popular, broad categories will be considered: organic solvents, and long chain polymers. Organic solvents decrease the dielectric constant of the media, this perturbs the electric fields that mediate the interactions of proteins in solution. In order to satisfy their electrostatic demands, the proteins associate with one another (McPherson and Gavira, 2014). On the other hand, long chain polymers function through volume exclusion – by taking up more room than they otherwise deserve, long chain polymers deny proteins space in the solvent, causing protein aggregation (McPherson and Gavira, 2014). Protein solubility is also dramatically impacted by pH, as this modifies the chance of salt bridges and hydrogen bonds, thereby influencing protein-protein contacts (Russo Krauss *et al.*, 2013).

Although the physics and chemistry at play in protein crystallisation is truly fascinating, all the more so for its mystery, the topic is sadly too vast to be

covered in detail. Nonetheless, a final note, what makes a good protein crystal? A highly ordered crystal. If the atoms of its constituent proteins are in the same defined positions over time and throughout the crystal, it can diffract to high resolution. But disorder, either in the form of atoms moving over time or the contents of one unit cell differing from another, lowers the intensity of diffraction spots, and so the resolution of the diffraction image (Acharya and Lloyd, 2005).

Still, with luck and perseverance, protein crystals may form. As touched upon previously, the next challenge faced by the investigator is converting their protein crystal into diffraction data. This is typically achieved by shooting the protein crystal with high intensity X-rays from a syncotron. How many times does a crystal need to be shot? A complete data set contains all reflections of the reciprocal space asymmetric unit. At least two terms in the preceeding sentence need explanation: “reciprocal space”, and “asymmetric unit”. Reciprocal space is a mathematical construct, based on direct space (the space of everyday life), that aids in understanding diffraction phenomena. An asymmetric unit is the smallest portion of the crystal that, through the action of symmetry operations, can produce the unit cell of the crystal (the unit cell is the smallest portion of the crystal that, through symmetry operations, can reproduce the entire crystal) (Dauter, 2017). This brings us to Bragg’s law, displayed in all its elegant simplicity in equation 5.

$$n\lambda = 2d \sin\theta \qquad \textbf{Equation 5}$$

This states that when an X-ray of wavelength λ is incident to a crystal surface, its angle of incidence (θ) will reflect with an equal angle of scattering (θ). Furthermore, when the path difference (d) is an integer (n) constructive interference will occur (Bragg, 1934). Bragg’s law is well illustrated by the Ewald construction (Figure 5-5). Constructive interference is, of course, what results in the spot on the detector, a datapoint closer to solving the structure of a protein. On which note, the strategy for data collection depends on several conditions (Smyth and Martin, 2000):

1. Crystallographic symmetry: The symmetry present in the crystal system and space group. A high symmetry crystal system, such as a cubic one,

could have data collected through as little as 35° of rotation. Compare to a low symmetry system, such as monoclinic, that could require data collection through 180°.

2. Non-crystallographic symmetry: The amount of symmetry present in the asymmetric unit. A particle such as a virus would have a high level of non-crystallographic symmetry (thanks to identical subunits), whereas a protein monomer may possess no non-crystallographic symmetry. The monomer would thus require a more complete data set than the virus.

3. The possibility of molecular replacement. If the structure of a similar protein has already been solved, it can be used as a starting point for new data and "fill-in" any gaps in the data.

4. The upper resolution limit of the crystal.

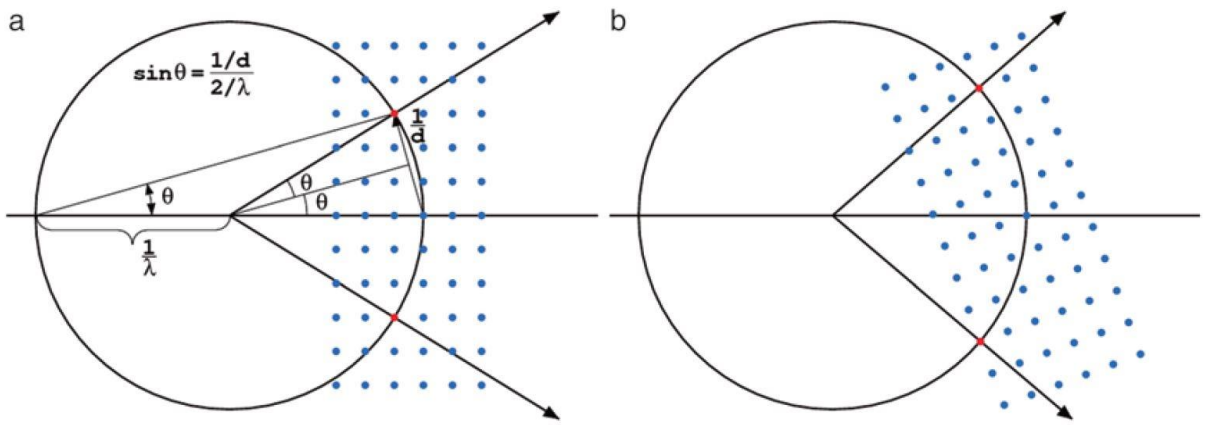


Figure 5-5 Bragg's law illustrated by the Ewald construction. The central section of an Ewald sphere of radius 1/λ, representing the X-rays, diffracting off a crystal (not drawn) in the centre of the sphere. The reflections in the reciprocal space are represented by the lattice of dots. **a)** If a point in reciprocal lattice lies at the surface of the Ewald sphere (red dot) the trigonometric conditions of Bragg's law are met. **b)** To obtain more reflections the crystal is rotated, bringing new points onto the surface of the Ewald sphere, satisfying Bragg's law. This figure is taken from Dauter (Dauter, 2017).

Of course, it is also perfectly feasible to simply collect data blindly. Starting from an arbitrary crystal orientation, using full beam intensity, rotating through through 180°, with an image every 0.1°, although this would likely be suboptimal. Still, armed with a diffraction pattern, the investigator is equipped to solve the structure of life's machinery.

5.2 Aims and Hypotheses

Below is a study that expresses functional *Leishmania* ubiquitin activating enzyme, LmxM.34.3060, and a functional SUMOylation E2, LmxM.02.0390. It goes on to chart LmxM.34.3060's response to various adenosine sulfamate based inhibitors. To further understand LmxM.34.3060's capacity as a novel drug target, protein crystallography was conducted, with drug modelling performed on the resultant LmxM.34.3060 structure. To validate the crystallographic model, small angle X-ray scattering was performed.

- It is hypothesised that LmxM.34.3060 will respond differently to drugs that are tailored to the human ortholog.

- It is hypothesised any difference will be due to differences in the structure of the ATP binding pocket.

5.3 Results

5.3.1 Expression of LmxM.34.3060 for drug assays

Given the data gathered from an activity-based probe, the most abundantly expressed ubiquitin activating enzyme *in vitro* cultures, LmxM.34.3060, was expressed recombinantly. LmxM.34.3060 is also the *Leishmania* ubiquitin activating enzyme that is most dissimilar to human/host orthologs. A small-scale test was run using 10 ml of *E. coli*, to compare the abundance of LmxM.34.3060 in soluble (so potentially functional) and insoluble (presumably misfolded) conditions, in addition to examining leaky expression (defined as expression without IPTG stimulation). A Western blot probed with anti-His antibody can be seen in Figure 5-6, suggesting that expression is entirely dependent on IPTG induction, and a His tagged protein of the expected size is found in both soluble and insoluble fractions of lysates. Of note is the approximately 56 kDa band present wherever the full-length protein (~ 127 kDa) is expressed.

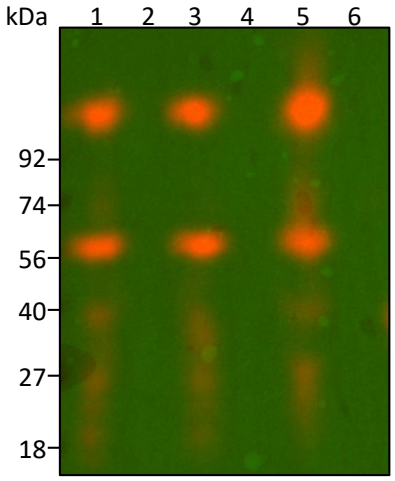
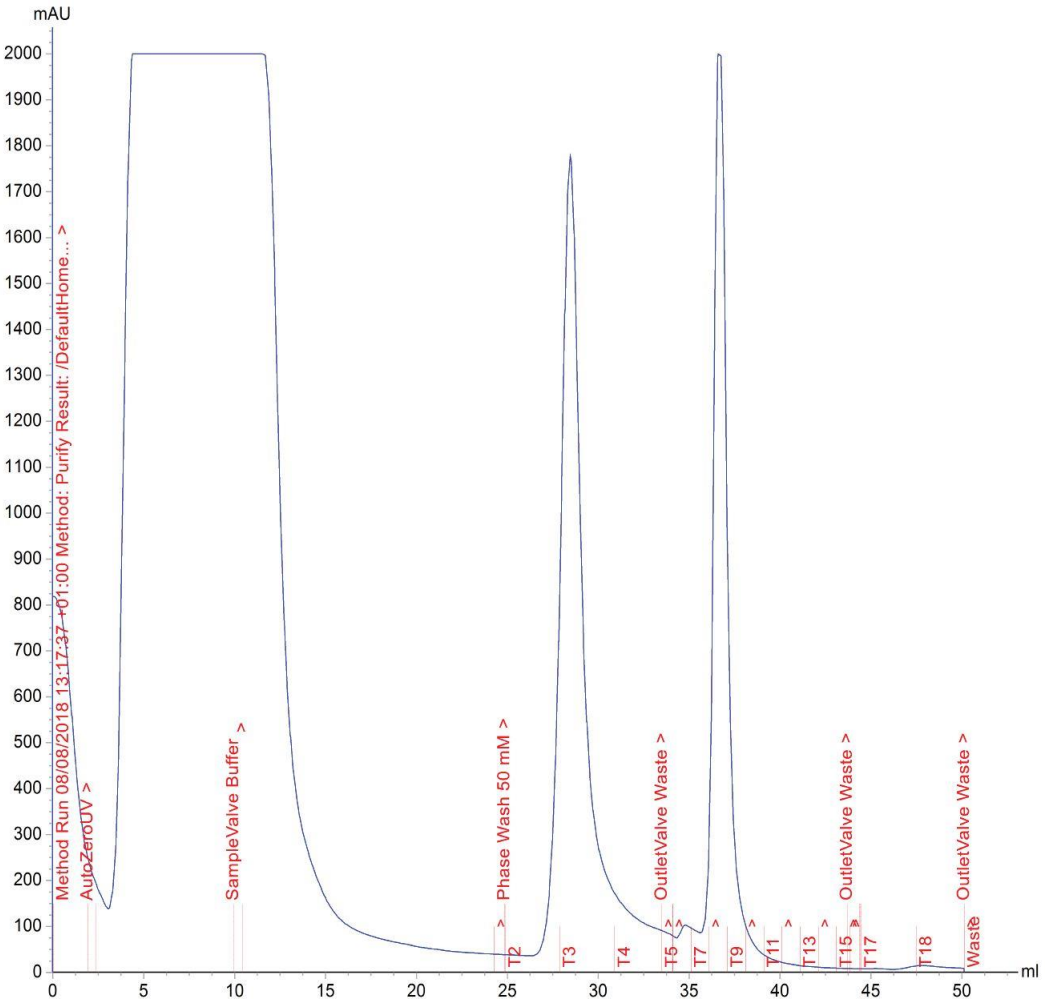


Figure 5-6 Western blot against E. coli lysates, after induction of *Leishmania* E1 expression. After culturing 10 ml of recombinant *E. coli*, expression of LmxM.34.3060 (a putative *L. mexicana* ubiquitin activating enzyme) either induced using IPTG, or without IPTG to check for leaky expression. Lanes 1 and 2, show *E.coli* lysates from induced and non-induced cultures respectively. Lanes 3 and 4 are the soluble fractions of lysates from induced and non-induced cultures. Lanes 5 and 6 are the insoluble fractions from induced and non-induced cultures. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Given the partial success of the small scale expression - demonstrating that at least some protein is expressed in a soluble, so potentially functional, form – the 10 ml experiment was scaled up to 400 ml. Purification of the His tagged protein using a nickel column can be seen in Figure 5-7. A protein of the expected mass

in fraction 5 is abundant enough to warp the neighbouring gel lanes. Again, a protein of ~ 56 kDa is also found wherever this protein is highly expressed. Additionally, there are plenty of proteins of other molecular weights to be found within fraction 5, so further purification was necessary.

a)



b)

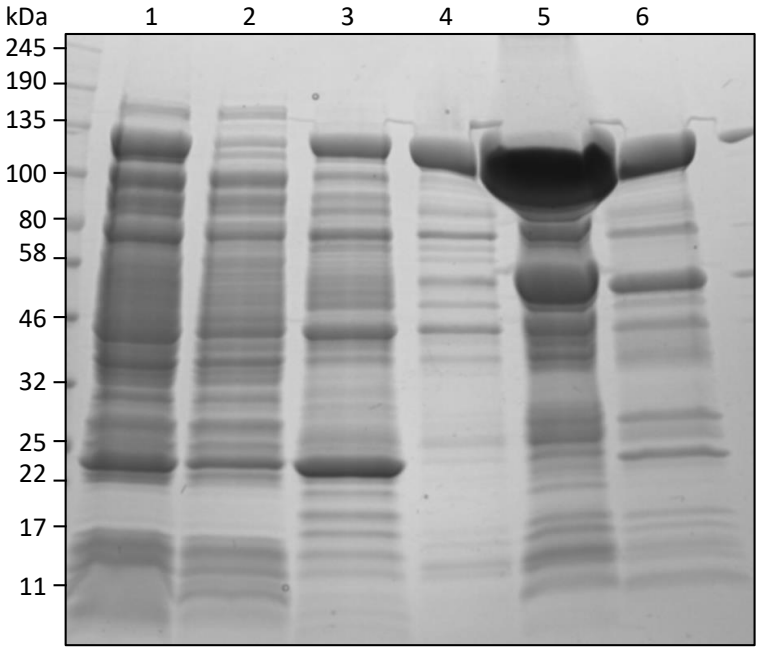
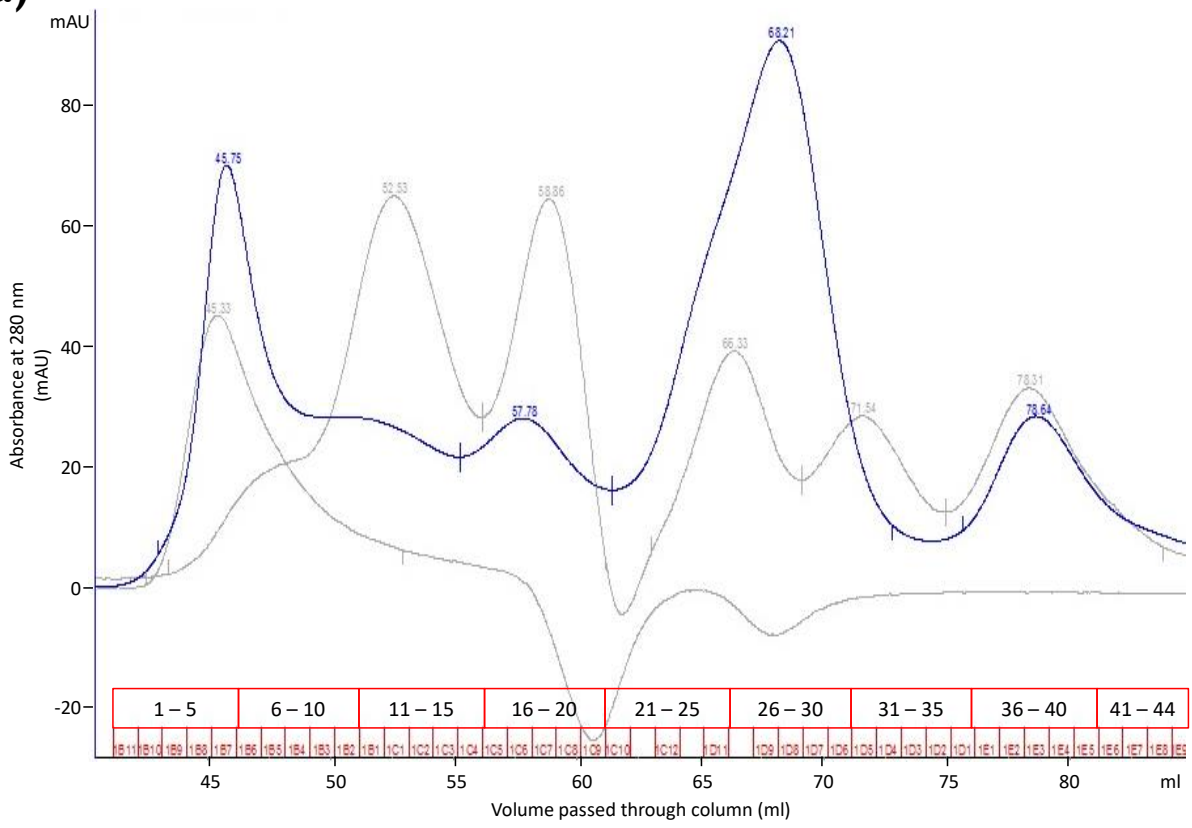


Figure 5-7 Purification by immobilized metal affinity chromatography of a recombinant ubiquitin activating enzyme from *Leishmania mexicana*. a) Lysate from *E. coli* expressing *L. mexicana* ubiquitin E1, was passed over a nickel column, and fractions eluted using various concentrations of an imidazole-based buffer were collected at various times. The blue line shows how the UV absorbance, related to protein concentration, varies in each fraction. Fraction numbers are in red, starting with the letter "T", along the x axis. **b)** These fractions were run on gel, with the lane numbers referring to the fraction. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers were supplied by NEB, product code #P7712.

Size exclusion chromatography (SEC) was employed for further purification, using fraction 5. A chromatograph and its corresponding gel in Figure 5-8. Although not found in isolation, a protein of the expected size can be seen across fractions C11 to D5 (inclusive). Fractions C11 to D5 were pooled, in the expectation that the residual amounts of unwanted protein would not affect the biochemical assays required to demonstrate functionality.

a)



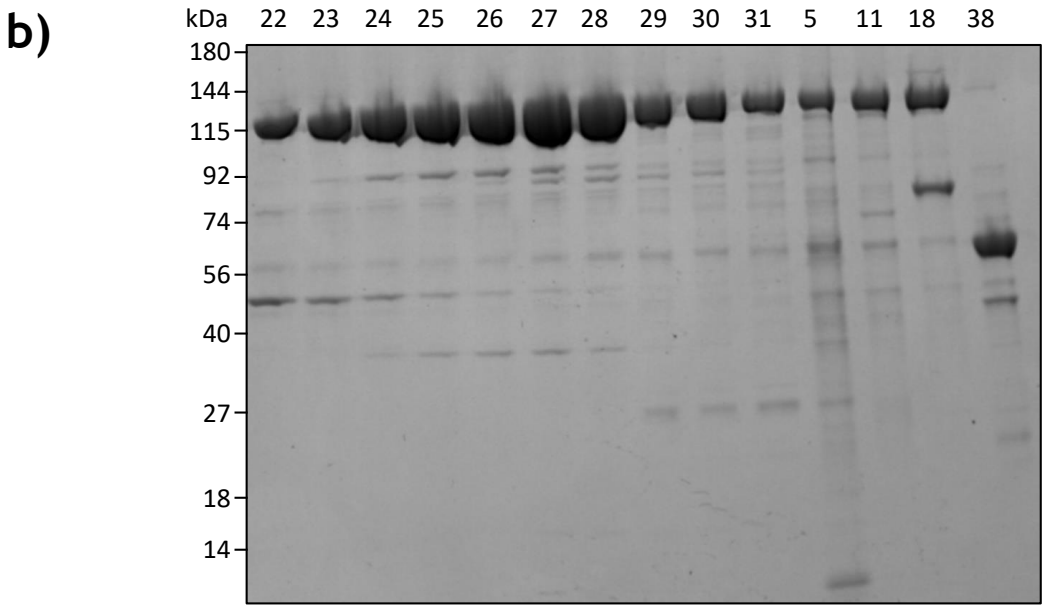


Figure 5-8 Purification by size exclusion chromatography of a recombinant ubiquitin activating enzyme from *Leishmania mexicana*. a) After undergoing His tag purification, a fraction of *E. coli* lysate was passed through gel filtration, with fractions collected over time. The blue line shows how the UV absorbance, related to protein concentration, varies in each fraction. Fraction numbers are in the red boxes on the x axis. The grey line relates to templates of known molecular weights. **b)** Fractions from size exclusion chromatography were run on gel and Coomassie stained. The lane labels relate to the fraction numbers that can be seen in the chromatograph. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

To discern whether the synthesised LmxM.34.3060 was operational, and could indeed activate ubiquitin, a functional assay was conducted (Figure 5-9). Without a reducing agent, a mass shift could be seen when LmxM.34.3060 was incubated with human ubiquitin and the necessary cofactors. However, when a reducing agent was included in the loading buffer, the mass shift typical of a LmxM.34.3060~ thioester bond was absent. Furthermore, addition of the human E2 UbcH5a results in another, reducing agent dependent, mass shift – that of UbcH5a with and without Ub. However, it is worth attention that, despite an abundance of cofactors and ubiquitin, not all expressed LmxM.34.3060 reacted with ubiquitin, as evidenced by the double banding in lanes 1 and 3.

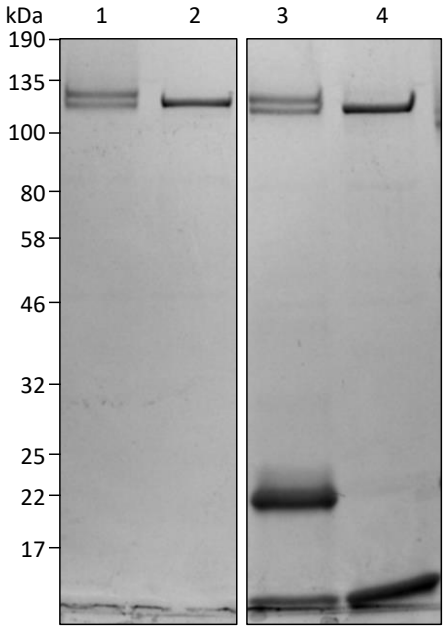


Figure 5-9 : Coomassie stain of a ubiquitin charging and transfer assay, employing expressed *Leishmania* E1. *L. mexicana* ubiquitination E1, expressed using *E. coli*, underwent a functional assay, to determine whether it could charge human ubiquitin and transfer it to the human ubiquitin E2 UbcH5a. Lane 1 indicates a reaction in which the E1 is incubated with human ubiquitin, ATP, $MgCl_2$, and HEPES. Lane 2 demonstrates the same reaction, but with the presence of 100 mM DTT. Lane 3 also shows the interaction of human ubiquitin ATP, $MgCl_2$, and the *Leishmania*l E1, but additionally has a human E2. Lane 4 is identical to lane 3, but features 100 mM of DTT. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

5.3.2 Inhibitors of LmxM.34.3060

To investigate LmxM.34.3060's potential as a novel therapeutic target TAK-243, an inhibitor active against the human E1, UBA1, was tested. Figure 5-10 indicates the concentration of TAK-243 necessary to inhibit *Leishmania* and human ubiquitin activating enzymes. It is apparent that a higher concentration of TAK-243 is needed to inhibit the *Leishmania* enzyme, as opposed to its human counterpart. This difference may be more noticeable if all of the LmxM.34.3060 employed was functional, however this is not the case demonstrated by the double banding even with 0 µM of inhibitor. Even so, with 5 µM of TAK-243 the human E1 displays no visible UBA1~ conjugates while *Leishmania*'s LmxM.34.3060 is still able to react with ubiquitin at this inhibitor concentration.

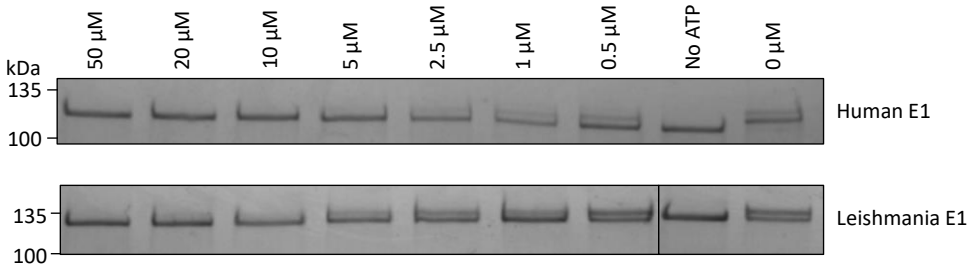


Figure 5-10 Coomassie stain of human and *Leishmania* E1 ubiquitin charging, when incubated with various concentrations of a human E1 inhibitor (TAK-243) Micromolar values along the lanes indicate the concentration of TAK-243 employed in that reaction. The upper lanes demonstrate the effect of TAK-243 on human E1 UBA1, and the lower set show the interactions of the same inhibitor on the *Leishmania* E1. Samples were run on two 4-20% SDS-PAGE gels (one gel featured reactions involving human E1, the other *Leishmania* E1), and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

Given that not all the expressed LmxM.34.3060 was functional, a commercial human E2, UbcH5a, was added to the inhibition assay. The greater homogeneity of UbcH5a, combined with a greater relative weight increase when bound to ubiquitin, permits easier interpretation of results. In Figure 5-11, it can be seen that 2.5 µM of TAK-243 added to the Human E1 results in approximately a 1:1 ratio of UbcH5a with and without ubiquitin. However, an approximately 1:1 ratio of activated and inactivated UbcH5a is obtained after 50 µM of TAK-243 was added to the *Leishmania* E1.

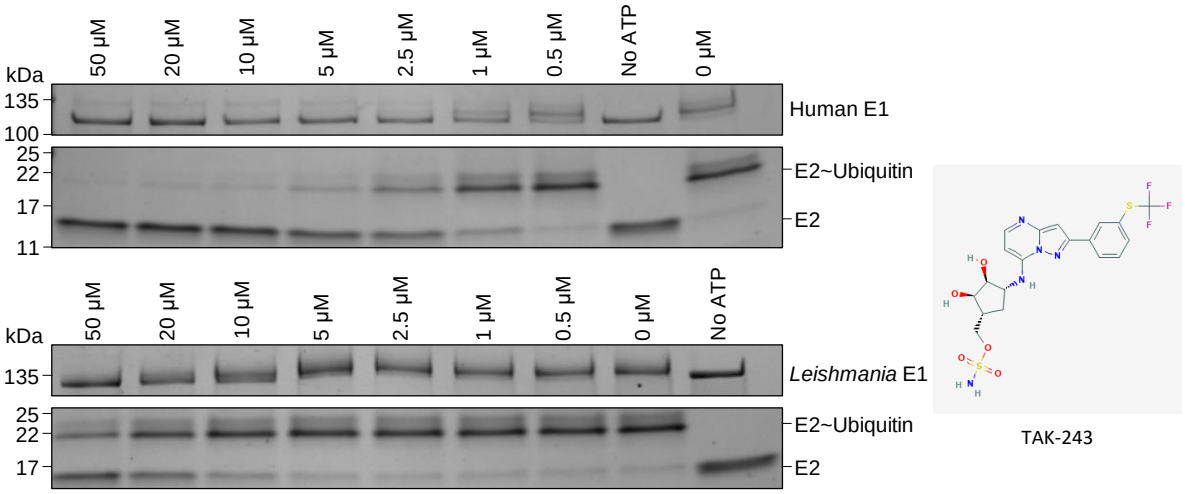


Figure 5-11 Coomassie stain of human and *Leishmania* E1 ubiquitin charging, and transfer to the human E2 UbcH5a, when incubated with various concentrations of a human E1 inhibitor (TAK-243). Micromolar values along the lanes indicate the concentration of TAK-243 employed in that reaction. The upper lanes demonstrate the effect of TAK-243 on human E1 UBA1, and the lower set show the interactions of the same inhibitor on the *Leishmania* E1. Samples were run on two 4-20% SDS-PAGE gels (one gel featured reactions involving human E1, the other *Leishmania* E1), and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB. TAK-243 image taken from PubChem entry on TAK-243 (*Tak-243 | C19H20F3N5O5S2 - PubChem*, no date).

Analysing the density of the E2 without ubiquitin band led to the densitometric profiles shown in Figure 5-12. It can be seen that the abundance of uncharged E2 increases quicker when the human E1 is inhibited by TAK-243 as opposed to LmxM.34.3060.

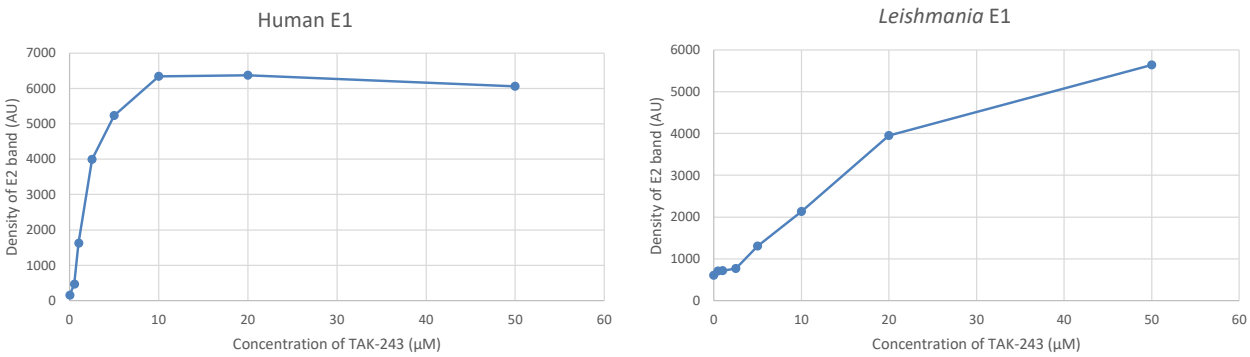


Figure 5-12 Densitometric analysis of E2 bands with TAK-243. Densitometry performed on the uncharged E2 band, when the human or *Leishmania* E1 is inhibited with TAK-243. Image analysis performed with ImageJ.

Using this improved method, additional drug assays were conducted. Figure 5-13 shows that semi-pure nucleocidin (kindly provided by Ms Ola Pasternak and Prof. David Zechel, Queen's University, Canada), a previously characterised protein synthesis inhibitor that has an adenosine sulfamate group, also inhibits the *Leishmania* E1. A 1:1 ratio of UbcH5a with and without ubiquitin was achieved with approximately 50 μM nucleocidin, whereas the human E1 did not achieve equivalent inhibition with 200 μM, the maximum concentration used.

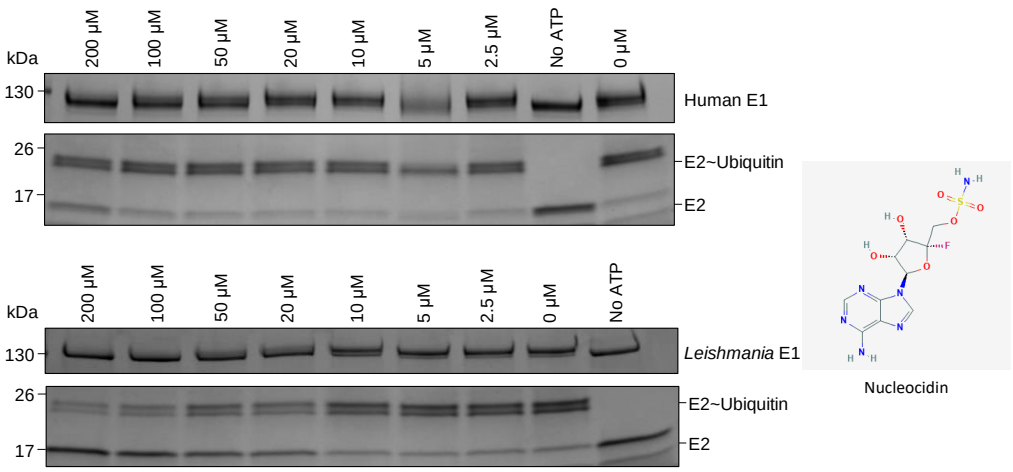


Figure 5-13 Coomassie stain of human and *Leishmania* E1 ubiquitin charging, and transfer to the human E2 UbcH5a, when incubated with various concentrations of nucleocidin. Micromolar values along the lanes indicate the concentration of nucleocidin employed in that reaction. The upper lanes demonstrate the effect of nucleocidin on human E1 UBA1, and the lower set show the interactions of the same inhibitor on the *Leishmania* E1. Samples were run on two 4-20% SDS-PAGE gels (one gel featured reactions involving human E1, the other *Leishmania* E1), and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied

by NEB. Nucleocidin image taken from PubChem (*Nucleocidin | C10H13FN6O6S - PubChem*, no date).

Again, analysing the density of the uncharged E2 bands formed a densitometric profile (Figure 5-14). The abundance of uncharged E2 increases quicker when LmxM.34.3060 is exposed to nucleocidin, relative to the human E1. However, both parasite and mammal E1s are inhibited by nucleocidin.

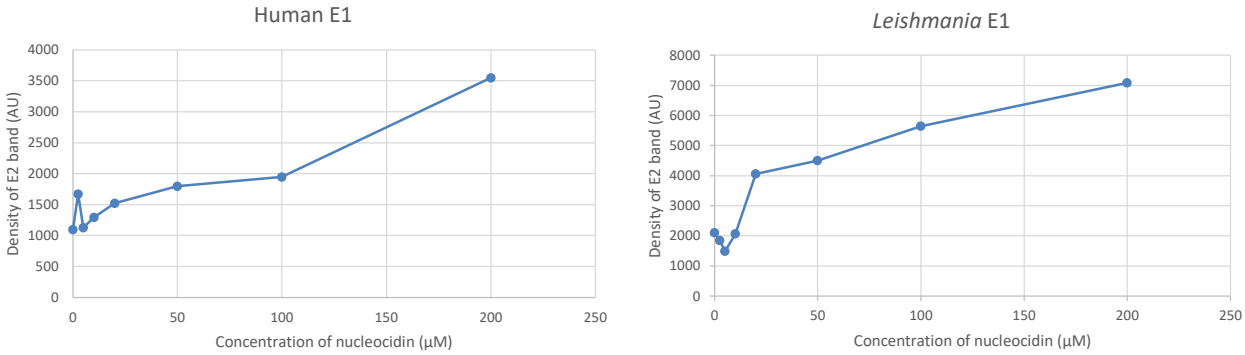


Figure 5-14 Densitometric analysis of E2 bands with nucleocidin. Densitometry performed on the uncharged E2 band, when the human or *Leishmania* E1 is inhibited with nucleocidin. Image analysis performed with ImageJ.

5'O-sulfamoyl adenosine, another compound with a sulfamate active group, but lacking nucleocidin's fluorine, was examined using the same method (Figure 5-15). 1 μM of 5'O-sulfamoyl adenosine achieved approximately 50% inhibition of the *Leishmania* E1, while the human E1 was inhibited to a similar extent using 50 μM.

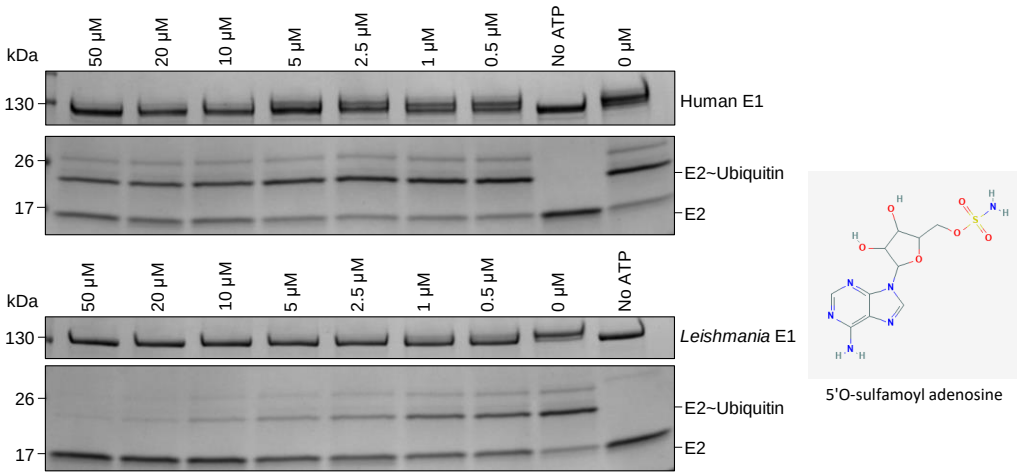


Figure 5-15 Coomassie stain of human and *Leishmania* E1 ubiquitin charging, and transfer to the human E2 UbcH5a, when incubated with various concentrations of 5'O-sulfamoyl adenosine. Micromolar values along the lanes indicate the concentration of 5'O-sulfamoyl adenosine employed in that reaction. The upper lanes demonstrate the effect of 5'O-sulfamoyl adenosine on human E1 UBA1, and the lower set show the interactions of the same inhibitor on the *Leishmania* E1. Samples were run on two 4-20% SDS-PAGE gels (one gel featured reactions involving human E1, the other *Leishmania* E1), and Coomassie stained. Molecular weight markers

had product code #P7712, and were supplied by NEB. 5'O-sulfamoyl adenosine image taken from PubChem entry on 5'O-sulfamoyl adenosine (*[(2R,3S,4R,5R)-5-(6-Amino-9H-purin-9-YL)-3,4-dihydroxytetrahydro-2-furanyl]methyl sulfamate | C10H14N6O6S - PubChem*, no date).

Plotting the density of the uncharged E2 bands forms the densitometry profile in Figure 5-16. The abundance of uncharged E2 increases quicker when LmxM.34.3060 is inhibited by 5'O-sulfamoyl adenosine, compared to 5'O-sulfamoyl adenosine's inhibition of the human E1. Although, both parasite and mammal E1s are inhibited by 5'O-sulfamoyl adenosine.

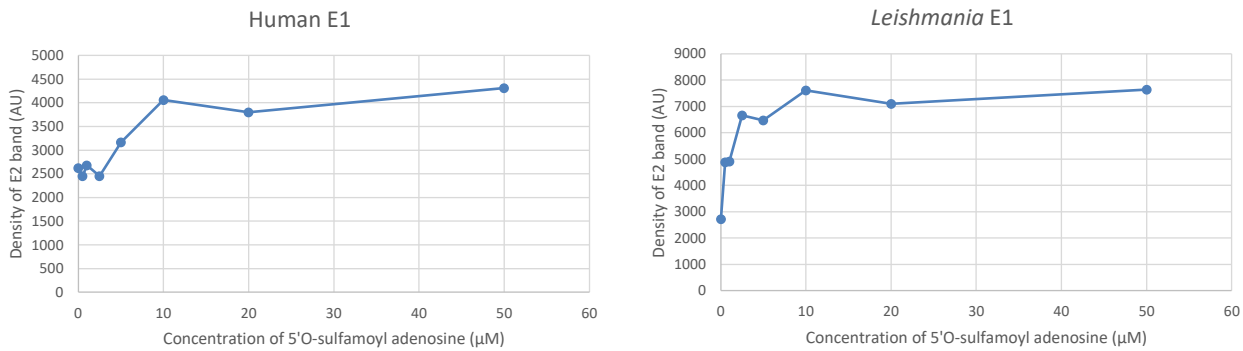


Figure 5-16 Densitometric analysis of E2 bands with 5'O-sulfamoyl adenosine. Densitometry performed on the uncharged E2 band, when the human or *Leishmania* E1 is inhibited with 5'O-sulfamoyl adenosine. Image analysis performed with ImageJ.

5.3.3 Expression of LmxM.34.3060 for protein crystallography

A crystal structure of LmxM.34.3060 would aid in drug development so, given the relatively good expression of functional LmxM.34.3060, it seemed prudent to attempt protein crystallography. To achieve the quantity of LmxM.34.3060 required for protein crystallography, the 400 ml *E. coli* cultures were scaled up to 12 L. Furthermore, following an initial lack of success in generating crystals using the aforementioned protein purification workflow, further purification steps were taken to increase homogeneity of the recombinant LmxM.34.3060. Chief among these additional steps was reacting LmxM.34.3060 with GST-tagged ubiquitin, then purifying with glutathione beads in a gravity flow column. To remove LmxM.34.3060 from the bead bound GST-Ub~LmxM.34.3060 complex, DTT was added, thereby removing LmxM.34.3060 whilst leaving GST-Ub attached to the solid support. This had the benefit of selecting only LmxM.34.3060 that was functional, as opposed to the prior method that purified any soluble LmxM.34.3060. Additionally, the His tag was cleaved from recombinant LmxM.34.3060, using the protease HRV3C, to present an enzyme as close as

possible to its *in vivo* equivalent. Key steps from this improved protocol are represented in Figure 5-17. There was an absence of a ~ 127 kDa band in the flow through from the glutathione beads (lane 1), and a similar absence of such a band remaining on the beads after glutathione addition (lane 4). This likely represents the absence of LmxM.34.3060 in the glutathione bead flow through, and shows LmxM.34.3060 was removed from the beads with DTT addition. Post His tag cleavage (lane 3), there were few visible proteins, but a band at approximately 127 kDa is clearly visible.

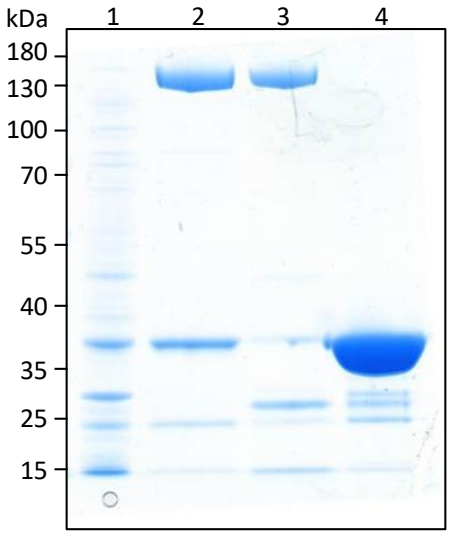


Figure 5-17 Coomassie stain from GST-Ub purification and HRV3C cleavage of *Leishmania* E1. An improved protocol to select for functional LmxM.34.3060, as opposed to soluble LmxM.34.3060, was employed and samples from various stages of that protocol presented here. Lane 1 is flow-through, lysate that did not interact with glutathione beads, diluted 1:100. Lane 2 is LmxM.34.3060 that still possess a His tag, as opposed to the sample in lane 3 that is LmxM.34.3060 without a His tag. Lane 4 is the elution from the GST beads using a glutathione based elution buffer. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #26616, and were supplied by ThermoFisher.

Another additional purification step added was ion exchange chromatography, further increasing the purity of LmxM.34.3060 used in crystal screening trays by selecting for an additional orthogonal variable, the isoelectric point (according to ProtParam (Gasteiger *et al.*, 2005), expected to be around 5.25 for LmxM.34.3060). A chromatograph of this process, with fractions of interest run on the gel, can be seen in Figure 5-18, which show that there is increased removal of contaminants. From this point on fractions were pooled to create two samples: fractions C25, C26, and C27 formed a “peak” sample, while fractions C22, C23, and C24 comprised the “off-peak” sample. Given the shoulder at C24, which is potentially representative of aberrant protein due to abnormal elution properties, this seemed a sensible precaution.

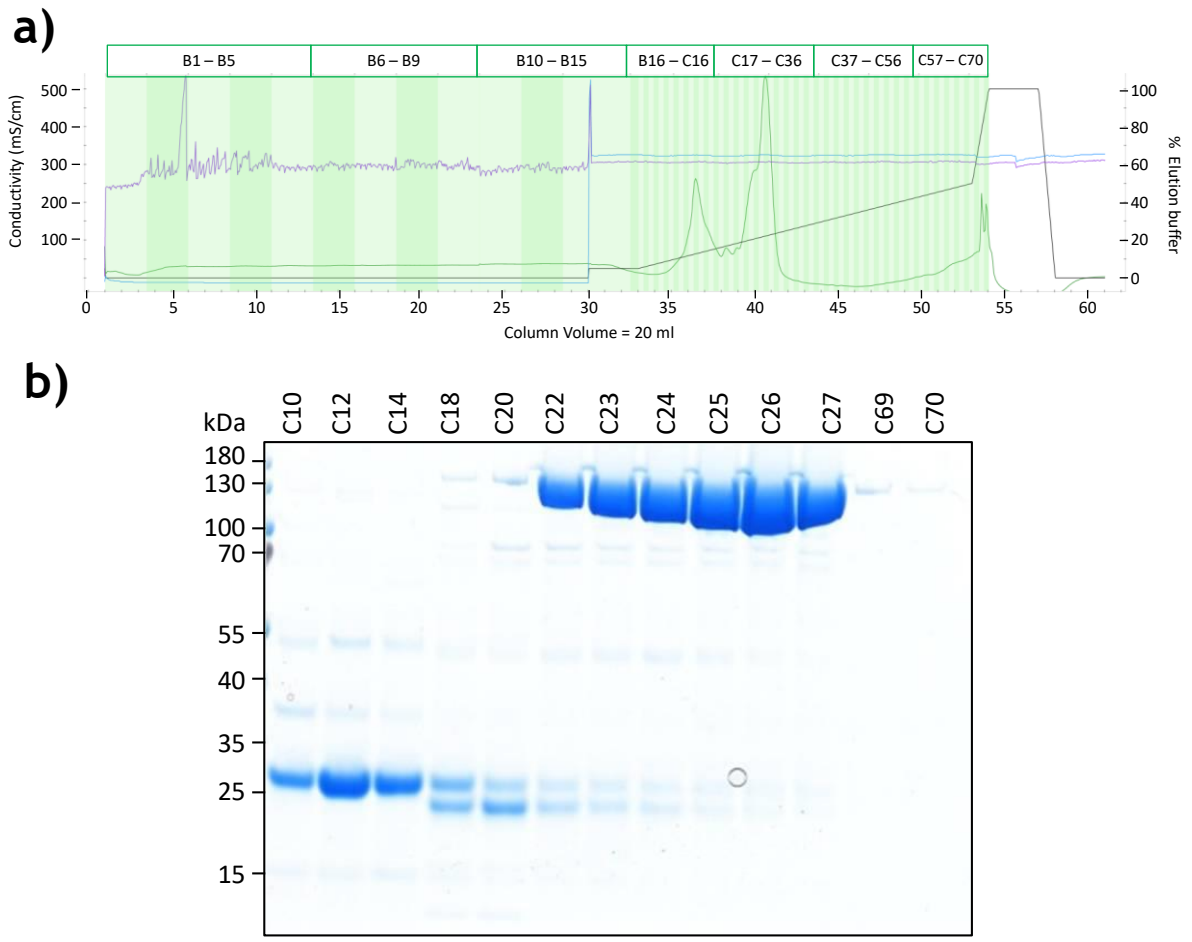


Figure 5-18 Purification by ion exchange chromatography of a recombinant ubiquitin activating enzyme from *Leishmania mexicana*. a) A chromatograph of UV absorbance during ion exchange chromatography to purify LmxM.34.3060. The green line shows how UV absorbance, and so protein concentration, at the column outlet varies; whilst the black line shows the increasing concentration of the sodium chloride-based elution buffer. The purple and blue lines show the changes in the column and system pressures respectively. Catchments of fraction identities can be seen at the top of the chromatograph. **b)** Coomassie stain of fractions obtained from ion exchange chromatography. Fractions featuring abundant protein were run on SDS-PAGE, to confirm presence of LmxM.34.3060. Lane labels refer to the fraction identities, naming that correlates with the fractions shown in the chromatograph. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #26616, and were supplied by ThermoFisher.

Both peak and off-peak samples required further processing, which necessitated size exclusion chromatography. The chromatograph of this process for the peak sample, and its accompanying gel, can be seen in Figure 5-19. Curiously, despite this in theory being the purest sample, there is a noticeable abundance of unwanted protein just below 70 kDa. The off-peak sample also underwent its own size exclusion chromatography, the chromatograph and gel portrayed in Figure 5-20. Although the maximum absorbance found in any fraction from the off-peak sample is lower than the maximum absorbance found in peak sample, this protein appears to be much purer under Coomassie staining. An overview of the entire LmxM.34.3060 purification process can be seen in Figure 5-21.

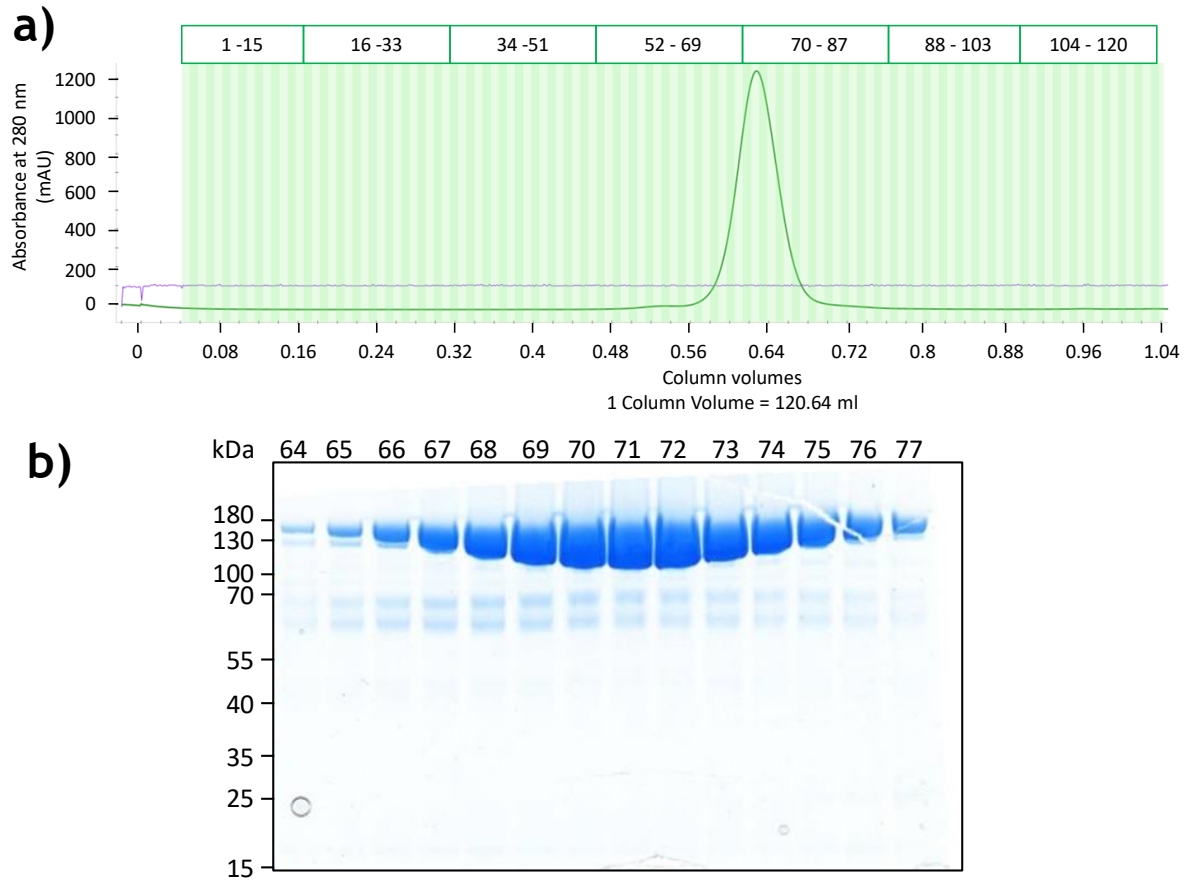
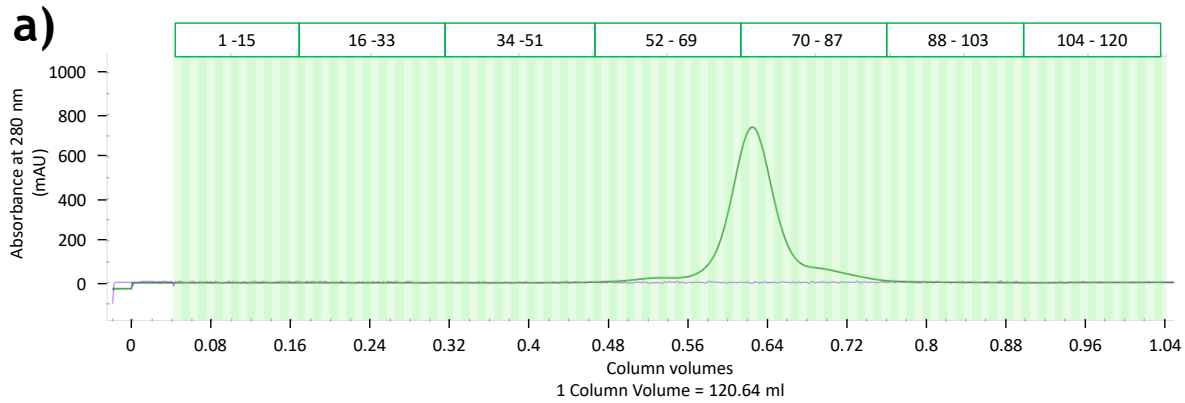


Figure 5-19 Size exclusion chromatography on a ubiquitin activating enzyme from *Leishmania mexicana*, performed on the peak sample obtained after ion exchange. a) To further purify LmxM.34.3060 for crystallographic purposes, the peak sample obtained after ion exchange was passed through gel filtration. The green line represents UV absorbance, and how it varies between fractions. The fractions numbers can be found in the green boxes along the top of the graph. **b)** Coomassie stain of fractions from size exclusion chromatography. Fractions featuring abundant protein were run on SDS-PAGE, to confirm presence of LmxM.34.3060 along with any contaminants. Lane labels refer to the fraction identities. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #26616, and were supplied by ThermoFisher.



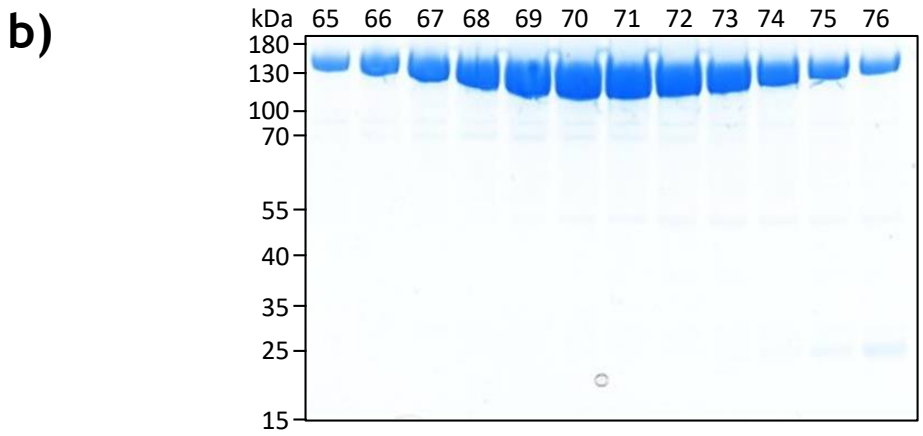


Figure 5-20 Size exclusion chromatography on a recombinant *Leishmania mexicana* ubiquitin activating enzyme, performed on the off-peak sample obtained after ion exchange. **a)** To further purify LmxM.34.3060 for crystallographic purposes, and minimise loss of protein, the off-peak sample obtained after ion exchange was passed through gel filtration. The green line represents UV absorbance, and how it varies between fractions. The fractions numbers can be found in the green boxes along the top of the graph. **b)** Coomassie stain of fractions from size exclusion chromatography. Fractions featuring abundant protein were run on SDS-PAGE, to confirm presence of LmxM.34.3060 along with any contaminants. Lane labels refer to the fraction identities. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #26616, and were supplied by ThermoFisher.

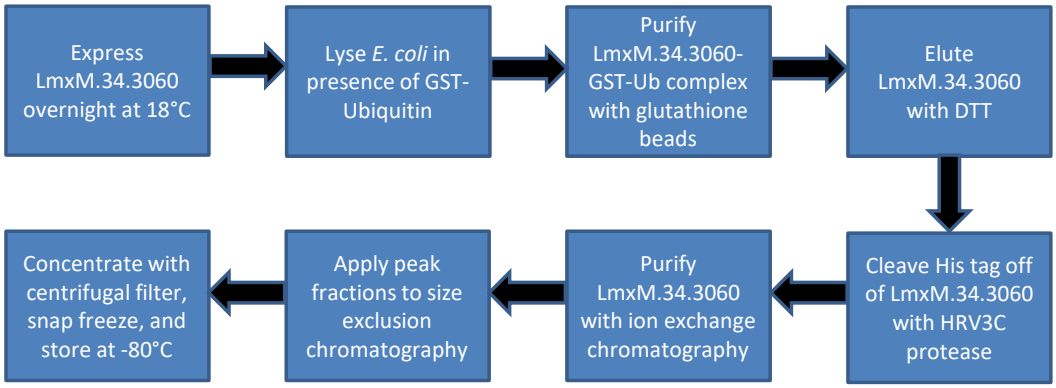


Figure 5-21 Summary of the workflow used to produce to a homogenous, functional *Leishmania* ubiquitin activating enzyme suitable for protein crystallography.

5.3.4 Small Angle X-ray Scattering of LmxM.34.3060

With the generous assistance of Dr Mads Gabrielson (University of Glasgow), SAXS analysis enabled determination of the solution structure of LmxM.34.3060. Figure 5-22 demonstrates an X-ray scattering curve of LmxM.34.3060. The scattering curve is as expected for a globular protein, however there is an "uptick" at very small values of the scattering vector (q).

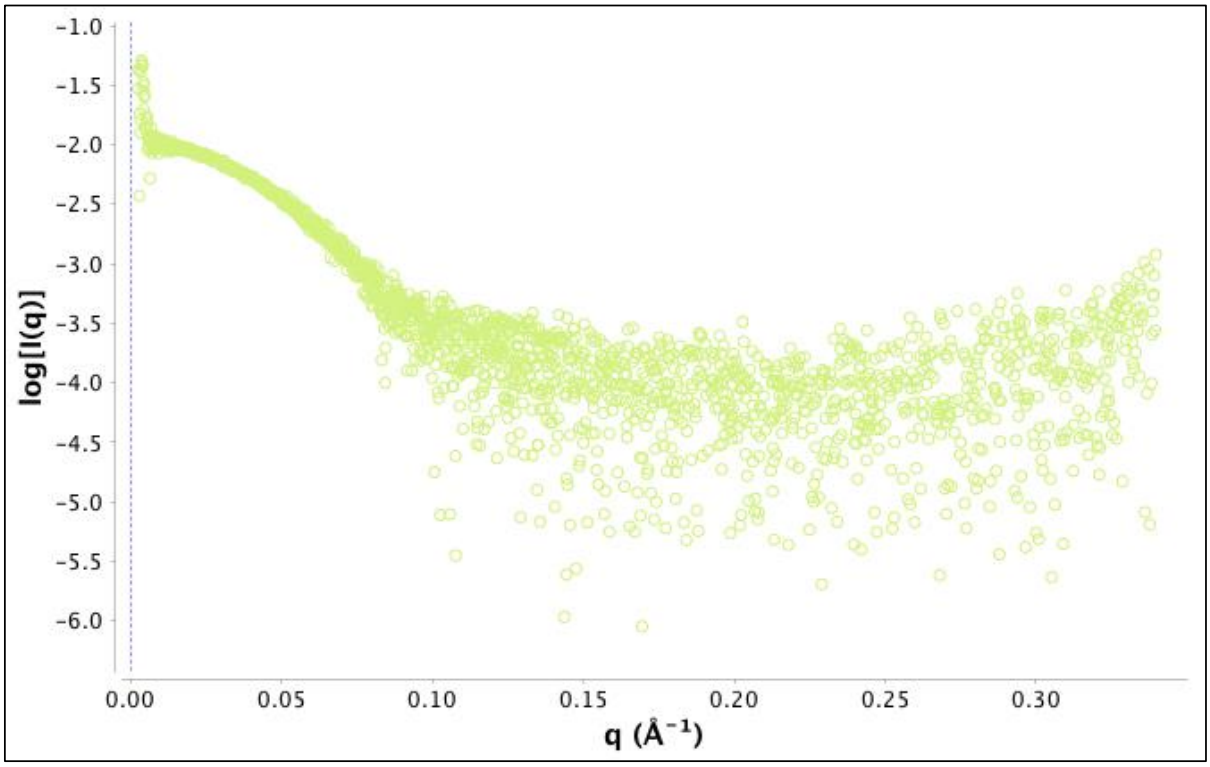


Figure 5-22 Intensity plot of X-ray scattering for an *L. mexicana* E1 Intensity (I) as a function of the amplitude of the scattering vector (q), charting how incident X-rays interact with electrons in LmxM.34.3060.

To evaluate the impact of issues that may have affected the scattering data, such as sample aggregation, radiation damage, buffer mismatch, and interparticle interaction, a Guinier plot was formed Figure 5-23. Using lower and upper q × Rg limits of 0.64 and 1.29 respectively, linear scaling is clearly evident. Furthermore, the residuals are flat and randomly distributed around 0.

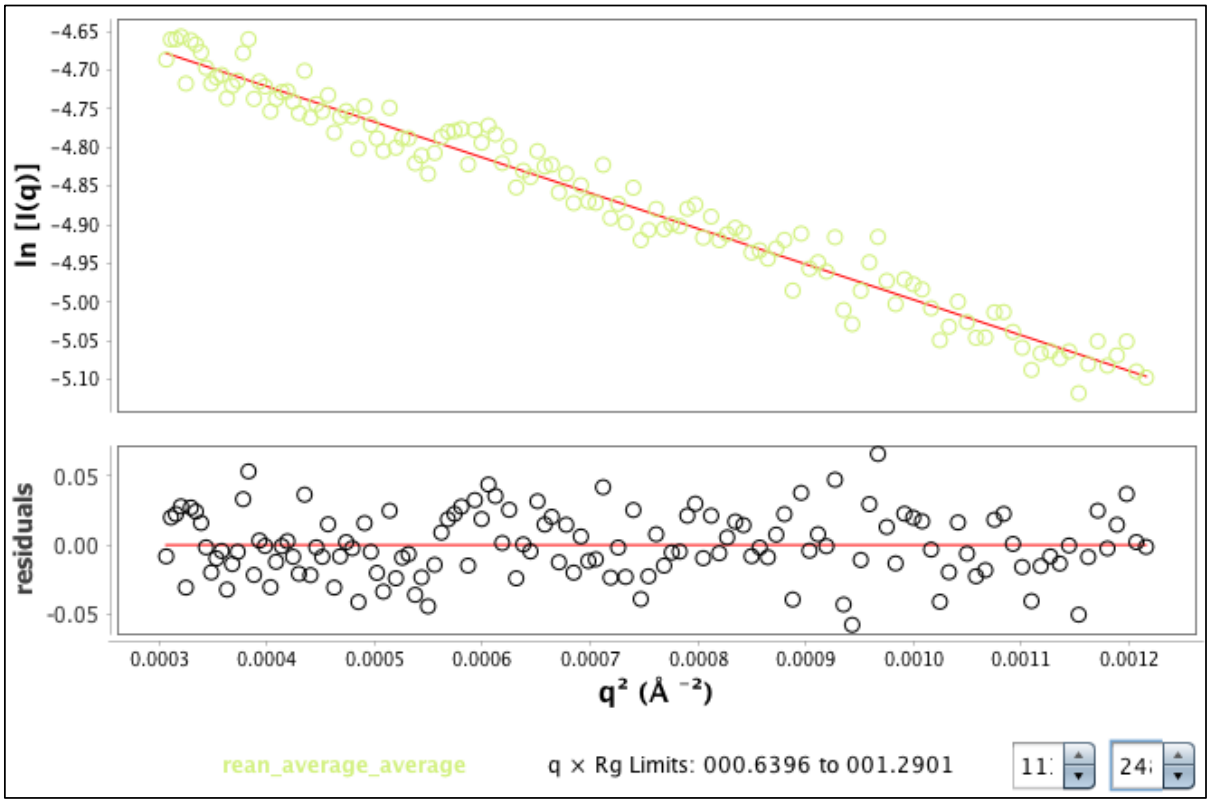


Figure 5-23 Guinier fitting of the X-ray scattering data for recombinant *L. mexicana* E1. The natural logarithm of the scattering intensity, I(q), versus the square of the amplitude of the scattering vector, q^2.

Using the Rg and I(0) values gathered from the Guinier plot, a dimensionless Kratky plot can be calculated (Figure 5-24), the shape of which provides an excellent tool for assessing the folding of samples. Note the Gaussian distribution of the plot.

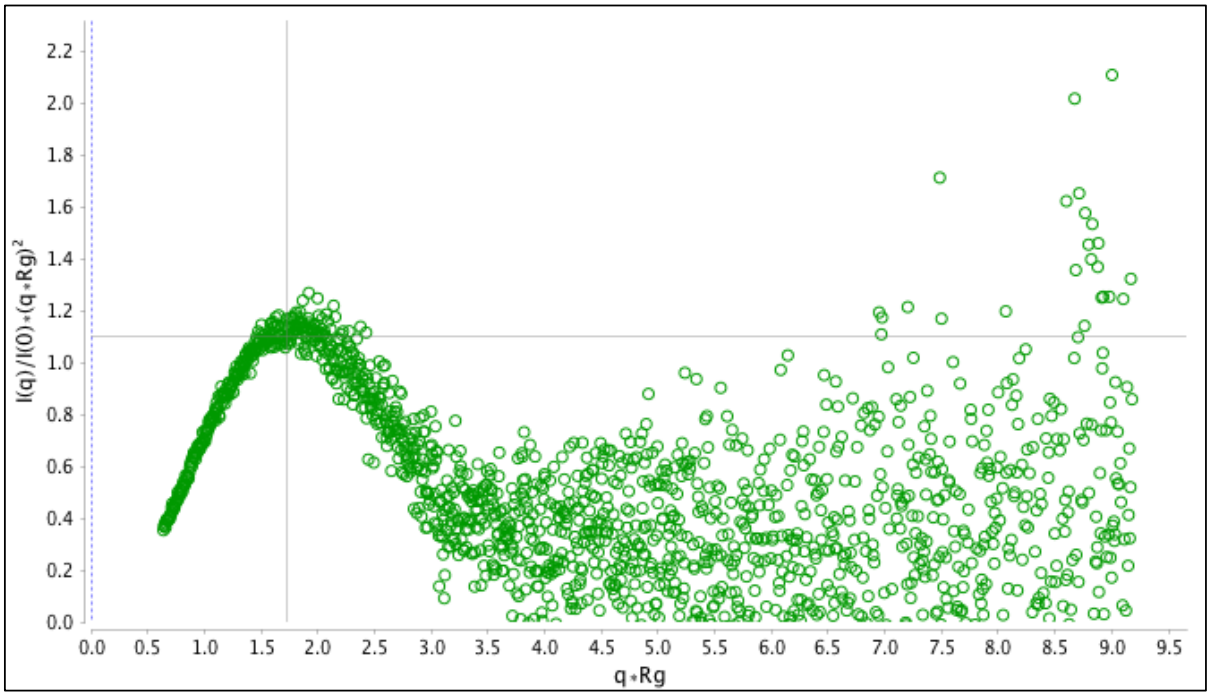


Figure 5-24 Dimensionless Kratky plot of scattering data for *L. mexicana* E1. Normalised scattering profile by mass and concentration of the LmxM.34.3060, using Rg and I(0) values from the Guinier fitting.

To further assess the folding of samples, by determining their flexibility, the decay of the scattering intensity can be used. Graphs that have different power-law relationships are shown in Figure 5-25. Of the four graphs, it is the Porod-Debye plot that plateaus first, stating that the scattering intensity decay is q^{-4}. The Kratky-Debye plot has a negative linear slope.

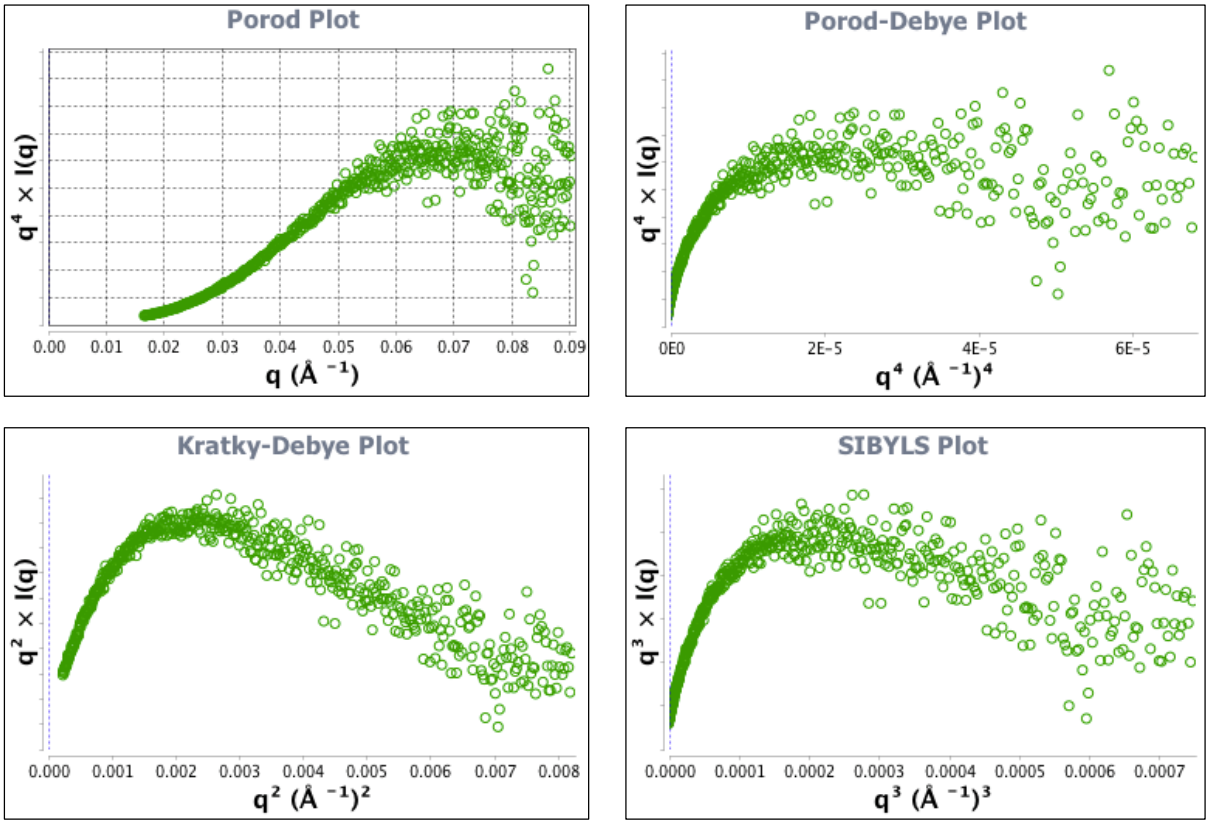


Figure 5-25 Flexibility of *L. mexicana* E1, determined by decay of q in various power-law distributions. SAXS data from LmxM.34.3060, transformed as $q^4 \times I(q)$ vs. q in the Porod plot; $q^4 \times I(q)$ vs. q^4 in the Porod-Debye plot; $q^2 \times I(q)$ vs. q^2 in the Kratky-Debye plot; $q^3 \times I(q)$ vs q^3 in the SIBYLS plot. The Porod-Debye plot plateaus first.

To confirm that buffer subtraction was successful, an integrated area plot of q × I(q) was formed (Figure 5-26). This flattens to a plateau at low values of q.

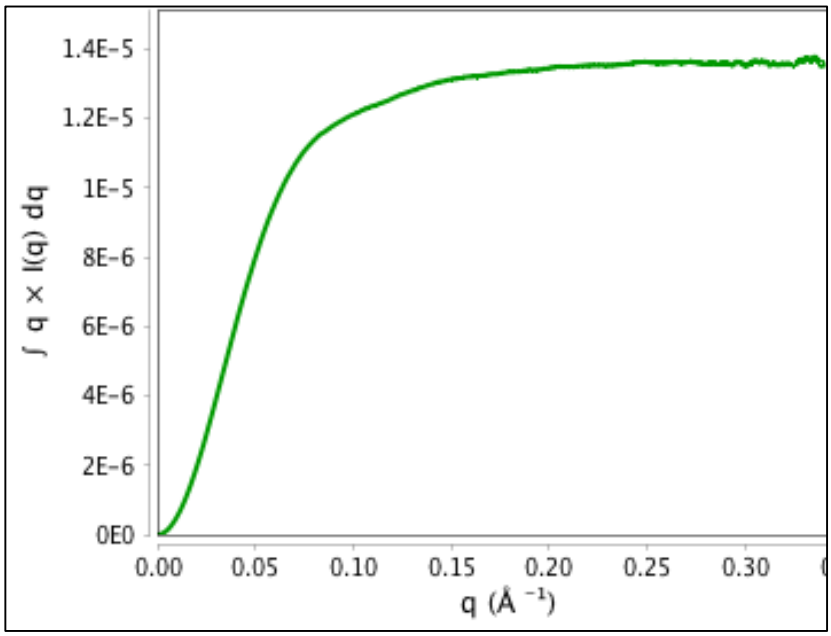


Figure 5-26 Integrated intensity plot of X-ray scattering data for *L. mexicana* E1. Total scattered intensity of SAXS data from LmxM.34.3060, used to determine the success of buffer subtraction.

A pair-distance distribution function, or P(r)-distribution, is plotted in Figure 5-27. Note the single peak, and the slightly longer tail at larger values of r. The distribution smoothly finishes at 108 Å, the d_{max}.

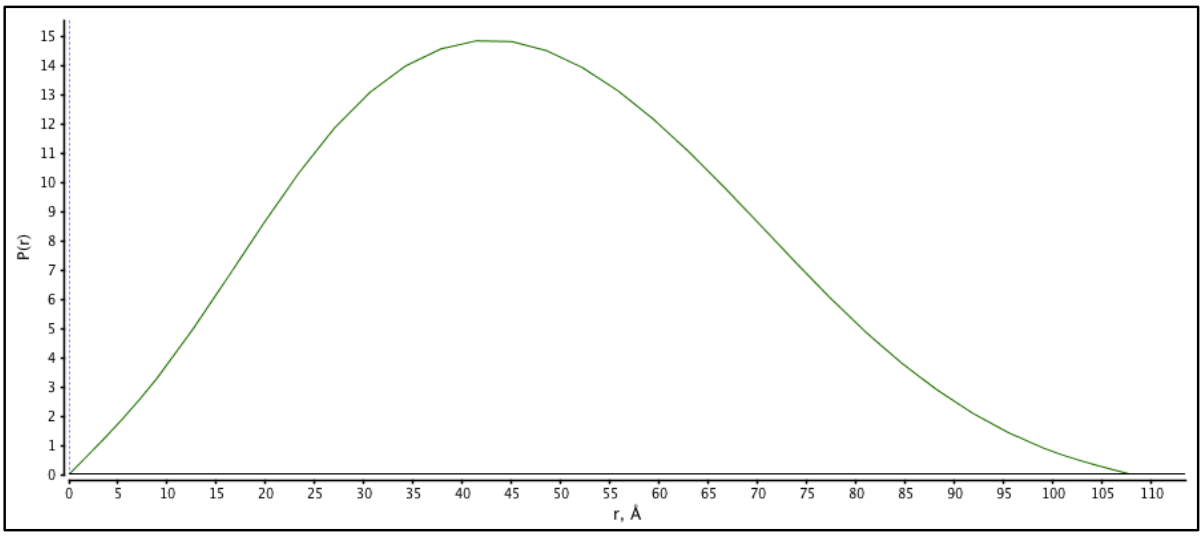


Figure 5-27 Integrated intensity plot of X-ray scattering data for *L. mexicana* E1. Total scattered intensity of SAXS data from LmxM.34.3060, used to determine the success of buffer subtraction.

Key biophysical parameters are presented in Table 5-1, providing a summary of the data obtained through SAXS.

		LmxM.34.3060
SAXS data collection		
I(0)		
	Real	1.2E-2
	Reciprocal	1.1E-2
R_g		
	Real	36.32
	Reciprocal	36.90
V_c		785.9
Volume (Å^3)		6.1E5
r		46.9
d_{max}		108
R_c		15.94
P_x		1.9

Table 5-1 Data collection from small angle X-ray scattering conducted on LmxM.34.3060.

This data was used to generate a model of LmxM.34.3060 in solution, which is pictured in Figure 5-28.

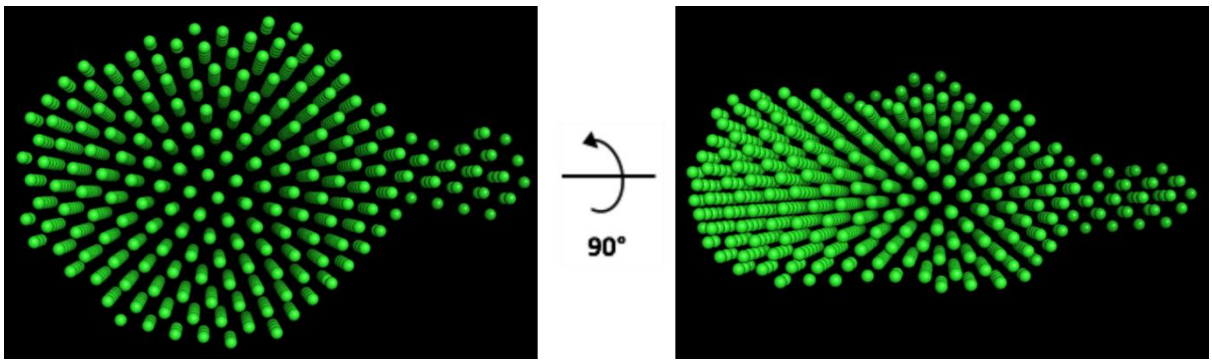


Figure 5-28 Model of an in-solution structure of LmxM.34.3060, constructed from SAXS data.

5.3.5 Crystal structure of LmxM.34.3060

For structure determination by X-ray crystallography, a wide variety of crystal screening trays were formed, using LmxM.34.3060 at multiple concentrations in multiple volumes, incubated at both 6°C and 19°C. However, these did not form any protein crystals. Alternatively, success occurred when LmxM.34.3060 was incubated with ATP, $MgCl_2$, and ubiquitin (further details in 2.2.24). The highest resolution dataset was 3.1 Å, using crystals formed in well E6 of the Morpheus screening tray (0.12 M ethylene glycols; 0.1 M Sodium HEPES and MOPS (acid) pH 7.5; 20% v/v Ethylene glycol and 10% w/v PEG 8000). The data collection and refinement statistics are summarised in Table 5-2. Using molecular replacement, the three-dimensional structure of LmxM.34.3060 was solved. This is seen in Figure 5-29. Remote data collection at the synchrotron, as well as structure refinement, was kindly conducted by Dr Mads Gabrielson (University of Glasgow).

	LmxM.34.3060
Protein crystal data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	22.53, 260.04, 138.46,
α, ß, γ (°)	90, 90, 90
Resolution (Å)	112.77 – 3.12 (3.17 – 3.12)[1]
Total number of reflections	489254
Unique reflections	72565
R_{merge} (%)	77.7 (695.3)
R_{pim} (%)	32.2 (280.4)
Completeness (%)	100.0 (100.0)
Multiplicity	6.7 (7.1)
I/σI	1.8 (0.1)
CC(1/2)	0.954 (0.214)
Wilson B ($Å^2$)	43.13

Current Refinement Statistics	
R_{work} (%)	38.4
R_{free} (%)	42.14
No. atoms	
Protein	12424
Water	-
R.m.s.d. bond (Å)	0.004
R.m.s.d. angle (°)	0.906
B-factors	
Main chain	80.17
Side chain	75.76
ATP	76.94
Ramachandran outliers (%)	12.25
Molprobity score	3.02

Table 5-2 Data collection and refinement statistics for the highest resolution protein crystal of LmxM.34.3060, in complex with ATP.

Data was collected at 0.91260 Å. [1]Values in parentheses are for the highest resolution shell.

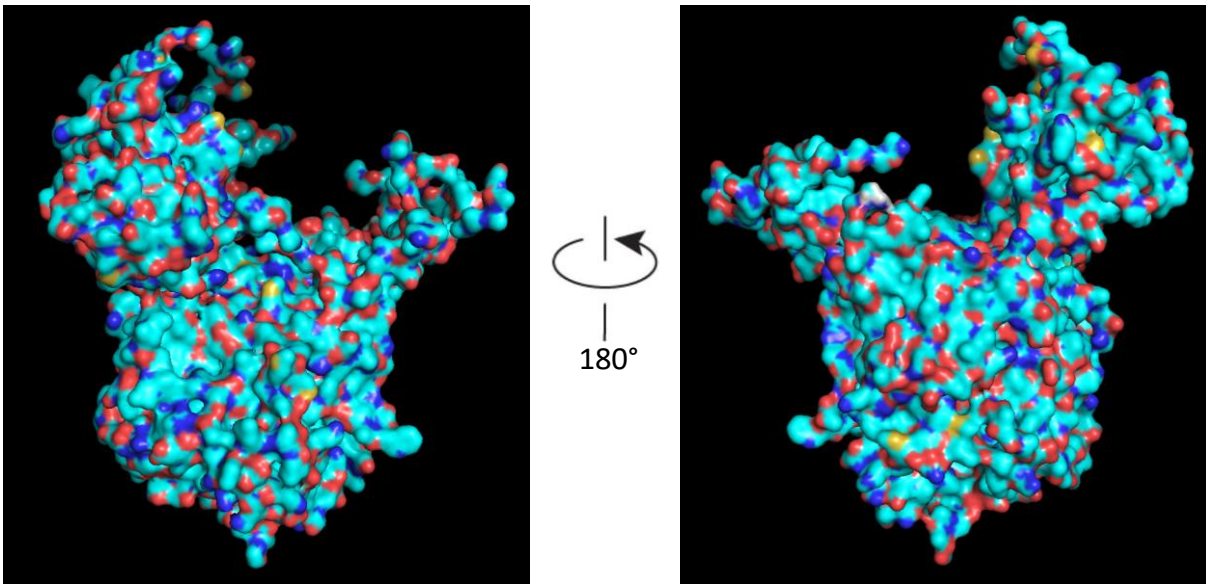


Figure 5-29 Surface of a *Leishmania* ubiquitin activating enzyme at 3.1 Å. The surface of LmxM.34.3060, obtained through protein crystallography.

Overlaying the structure obtained by SAXS with the crystallographic structure results in Figure 5-30. Although each structure overlaps considerably, white arrows point to locations that are exclusive to either the crystallographic or in-solution structures.

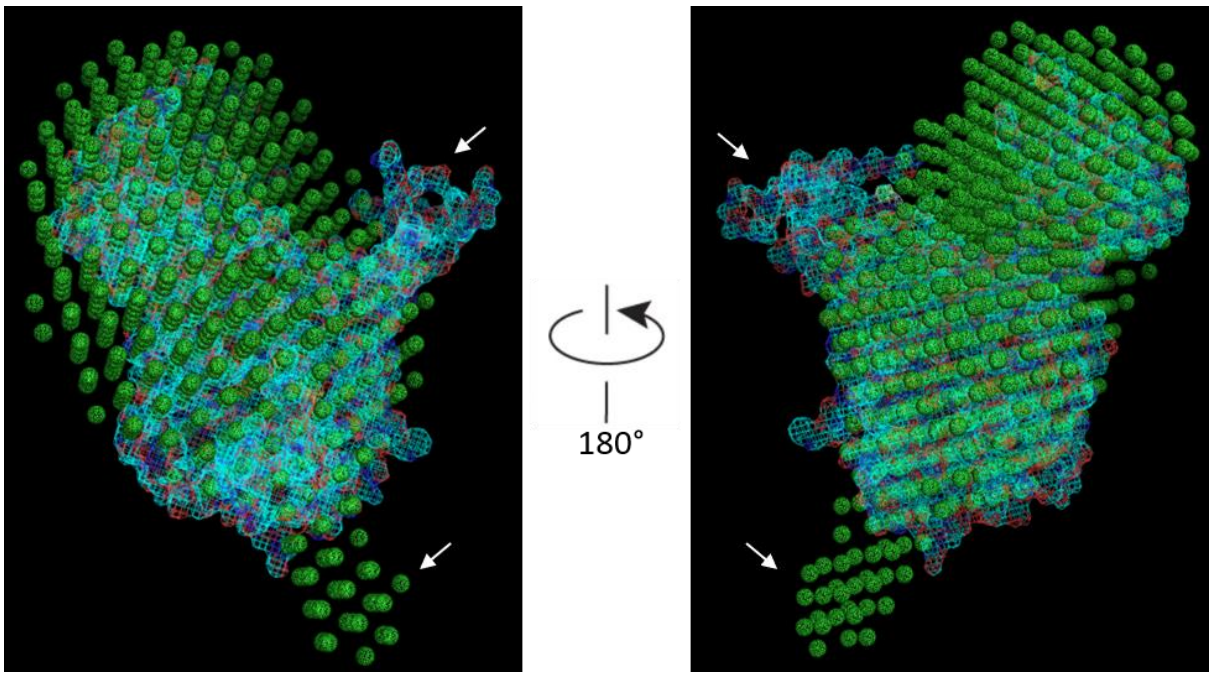


Figure 5-30 An overlay of crystallographic and SAXS structures of a *Leishmania* ubiquitin activating enzyme. Blue mesh represents the surface of LmxM.34.3060 at 3.1 Å, obtained through protein crystallography. Green spheres show the structure determined through small angle X-ray scattering. White arrows point to areas that appear in only the crystallographic or solution structures.

Using the CRYSOL software (Franke *et al.*, 2017), a SAXS scattering curve can be estimated from a protein's crystallographic structure. Figure 5-31 plots LmxM.34.3060's SAXS data against this theoretical scattering curve, determined using its crystal structure.

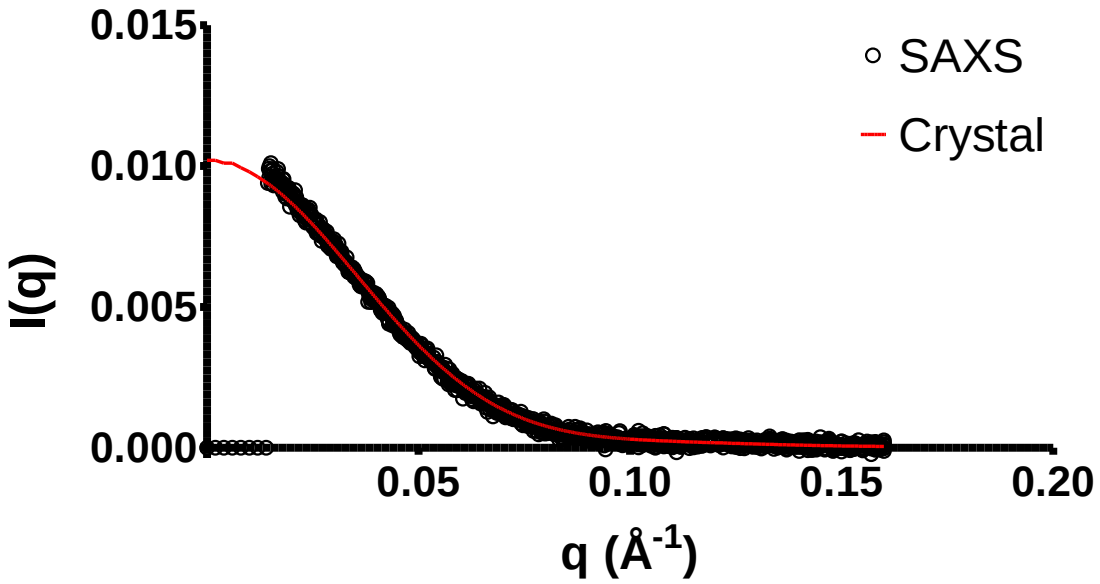


Figure 5-31 SAXS data from LmxM.34.3060 (black circles), a *Leishmania mexicana* ubiquitin activating enzyme, plotted against a theoretical scattering curve, determined using its crystallographic structure (red line). The theoretical curve was calculated using the CRYSOL software.

5.3.6 The drug binding pocket of LmxM.34.3060

Of particular interest to this project is the ATP binding site, where adenosyl sulfamate inhibitors are predicted to bind (Misra *et al.*, 2017). Figure 5-32 focuses on this site, while portraying interactions between the ligand and enzyme.

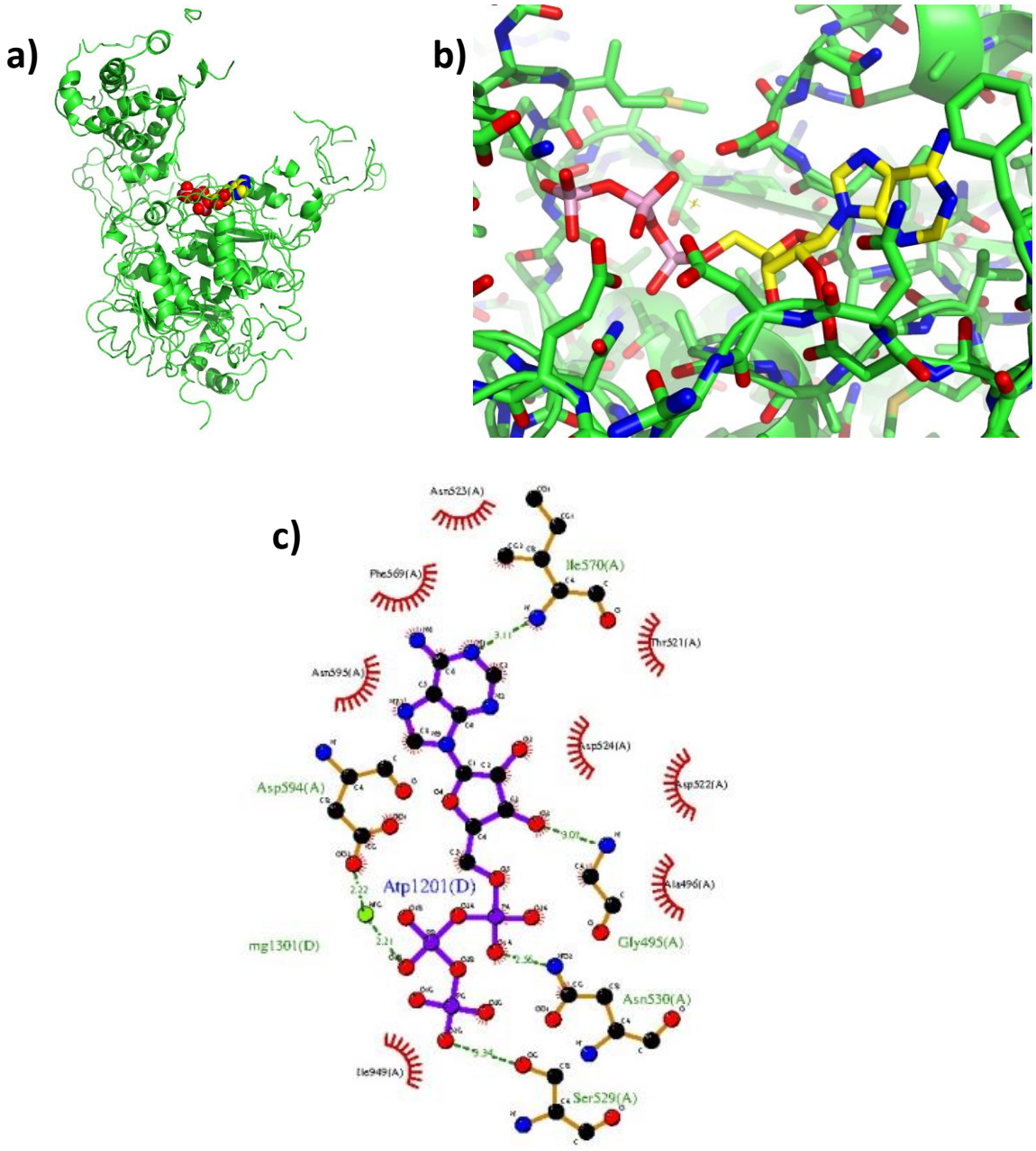


Figure 5-32 The ATP binding site of a *Leishmania* ubiquitin activating enzyme.
a) The structure of LmxM.34.3060, shown in green in cartoon form, with ATP, multicoloured, bound to the relevant pocket. **b)** A close up of ATP sitting in the ATP binding site. **c)** Interactions between ATP and the surrounding amino acids at the binding site. Red quarter circles represent hydrophobic interactions, dotted green lines denote hydrogen bonds, with the distance between atoms in Å.

Using this structure with location of ATP as a template, the inhibitor for the orthologous human E1, TAK-243, could be modelled in place of ATP. This concept is seen in Figure 5-33, where two methods of modelling TAK-243 are employed – fitting TAK-243 to the position of ATP in LmxM.34.3060, as well as copying the positioning of TAK-243 from an orthologous structure in the Protein Databank (5TR4). Hydrogen bonds 2.55 Å or smaller are highlighted, as are covalent bonds between drug and enzyme.

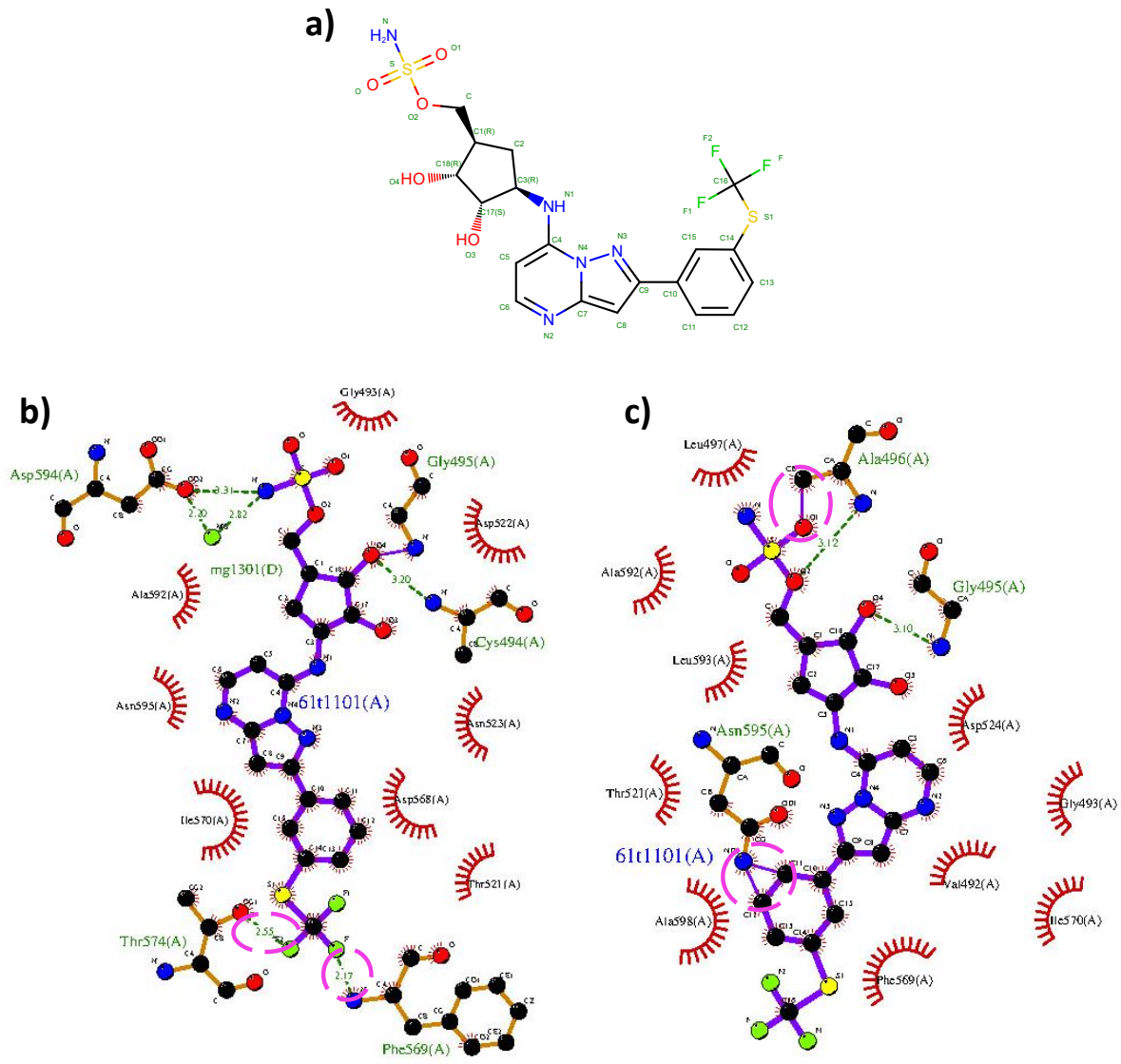


Figure 5-33 TAK-243 modelled into a *Leishmania* ubiquitin activating enzyme. a) The structure of TAK-243, as obtained from entry 61T in the Protein Databank. **b)** TAK-243 fitted to the position of ATP in LmxM.34.3060 crystallographic structure. **c)** The position of TAK-243 copied from the orthologous structure 5TR4 in the Protein Databank. Red quarter circles represent hydrophobic interactions, dotted green lines denote hydrogen bonds, with the distance between atoms in Å. Purple circles indicate possible clashes in the models.

5.3.7 Expression of LmxM.33.0900

A putative *L. mexicana* ubiquitin conjugating enzyme, identified through mass spectrometric and bioinformatic methods, was expressed with a polyhistidine tag in BL21 derivatives of *E. coli*. Figure 5-34 shows a small-scale test of E2 expression, which found a substantial portion of protein in the insoluble fraction. As such, the concentration of IPTG, as well as the growth media, were varied in an attempt to alter the solubility of expressed LmxM.33.0900 (Figure 5-35). It is notable that the majority of the E2 is present in an insoluble fraction of the bacterial lysate (defined as the pellet formed post-centrifugation of the lysates, as opposed to the soluble fraction present in the supernatant). That which is found in the soluble fraction exhibits an unexpected double band.

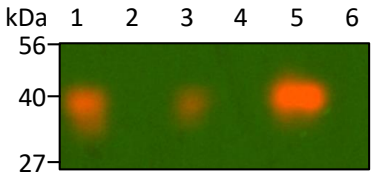


Figure 5-34 Western blot against E. coli lysates, after induction of *Leishmania* E2 expression. After culturing 10 ml of recombinant *E. coli*, production of LmxM.33.0900 (a putative *L. mexicana* ubiquitin conjugating enzyme) were either induced using IPTG, or without IPTG to check for leaky expression. Lanes 1 and 2, show *E.coli* lysates from induced and non-induced cultures respectively. Lanes 3 and 4 are the soluble fractions of lysates from induced and non-induced cultures. Lanes 5 and 6 are the insoluble fractions from induced and non-induced cultures. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

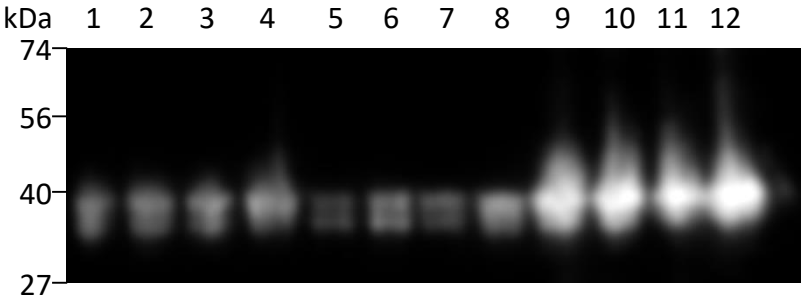
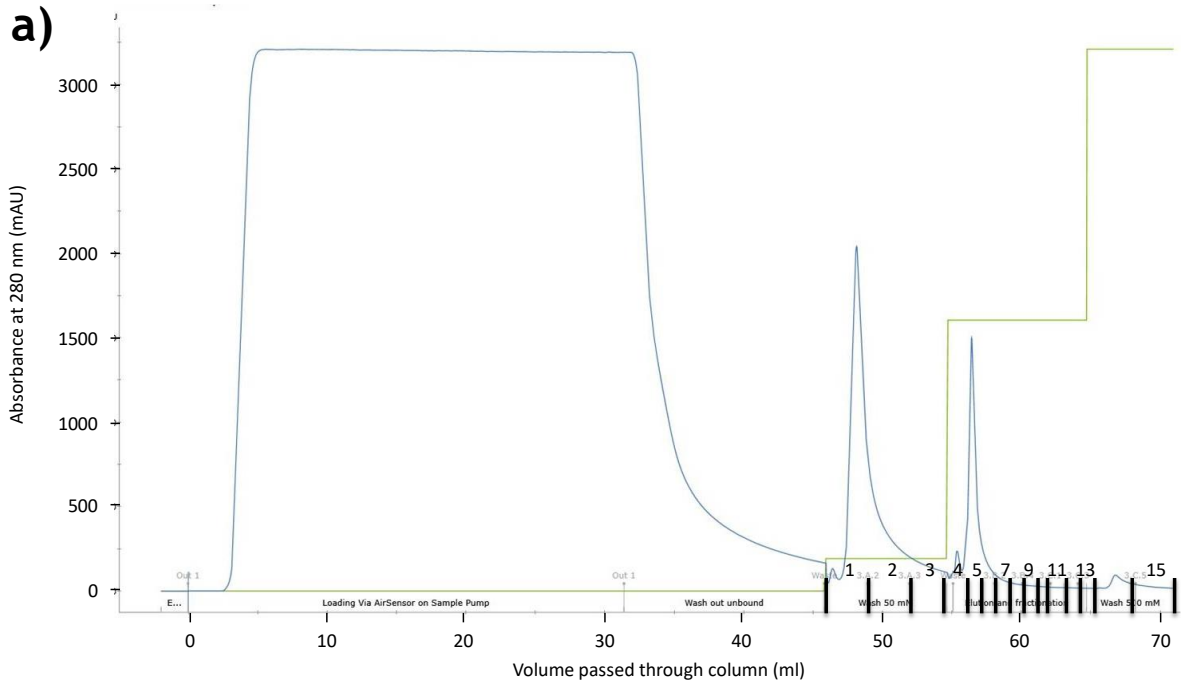


Figure 5-35 Western blot against E. coli lysates, after induction of *Leishmania* E2 expression. After culturing 10 ml of recombinant *E. coli*, production of LmxM.33.0900 (a putative *L. mexicana* ubiquitin conjugating enzyme) was induced using IPTG or auto-induction media (AIM). Lanes 1, 2, 3, and 4 show *E.coli* lysates, from bacteria induced using 1 mM, 0.5 mM, 0.2 mM, of IPTG respectively, or (in the case of lane 4) AIM. Lanes 5, 6, 7, and 8 show an identical order of IPTG and AIM conditions, but are soluble fractions of lysates, separated using centrifugation. Lanes 9, 10, 11, and 12 again show the four conditions in the same order, but present the insoluble fraction of the lysates. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Albeit not the majority, but a still significant proportion of target protein was found to be soluble, which made a nickel column purification of the His tagged protein worthwhile. This process is illustrated by Figure 5-36a, where the total protein content present in the column outlet is charted using UV absorbance, to determine which fractions contain the most protein. Fractions found to contain a higher protein content, and particularly those identified during the elution stages of the program, were analysed by gel electrophoresis to identify any bands corresponding to the predicted molecular weight of LmxM.33.0900 (approximately 40 kDa). This gel can be seen in Figure 5-36b. A relatively pure ~ 40 kDa protein is visible in fraction 14, the penultimate fraction to elute from the column, and so required a high concentration of imidazole to displace the protein from the column.



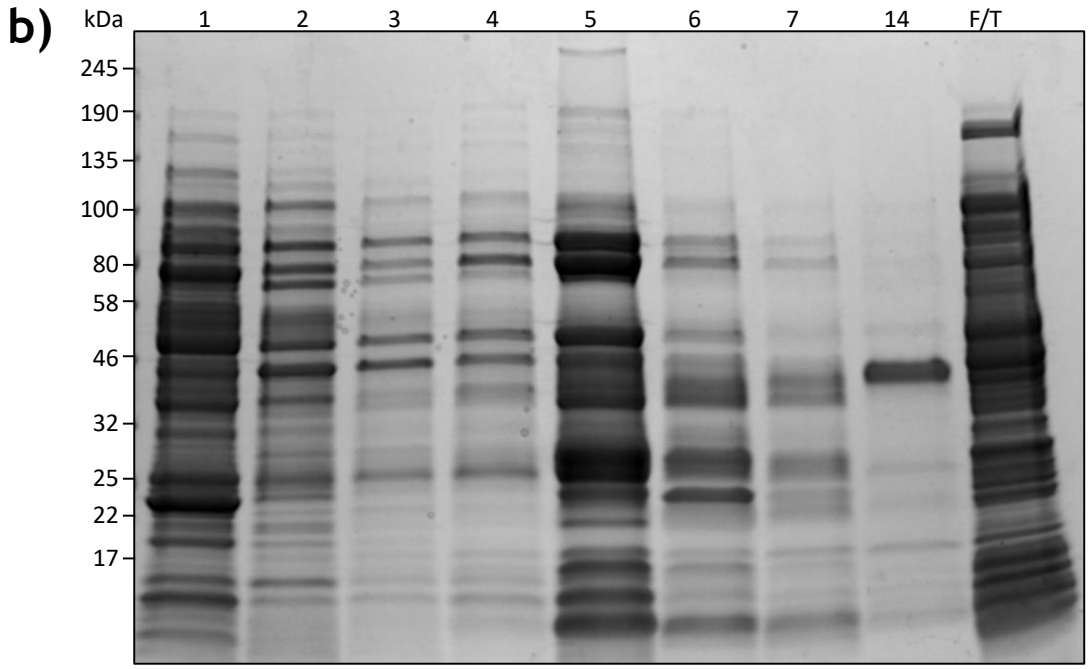


Figure 5-36 His tag purification of LmxM.33.0900, a putative of *Leishmania* E2. a) 800 ml of recombinant *E. coli* cultures were lysed, and the soluble fraction applied to an AKTA Start. Shown are the UV absorbance (blue line), an indicator of total protein abundance, and the concentration of the imidazole-based elution buffer (green line), over the volume passed through the column. The fractions numbers are indicated by the numbers 1 to 15, divided by the thick black lines on the x axis. **b)** Samples from various fractions obtained from AKTA His-tag purification were run on gel and Coomassie stained. Lane labels correspond to the fractions. Fractions 1, 2, and 3 are from washes. 4 is waste. Fractions 5, 6, 7, and 14 are from various elution peaks. F/T is flow through, lysate that did not bind to the column. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

To ascertain whether this ~ 40 kDa protein was indeed the protein of interest, the same fractions were assessed by Western blotting, using anti-His antibodies. It can be seen in Figure 5-37 that a His-tagged protein of the expected mass is present in the elution fractions, most notably so in fraction 14. It is interesting that the bands found are neither singular nor well-formed, with slight variations in their apparent molecular masses between fractions.

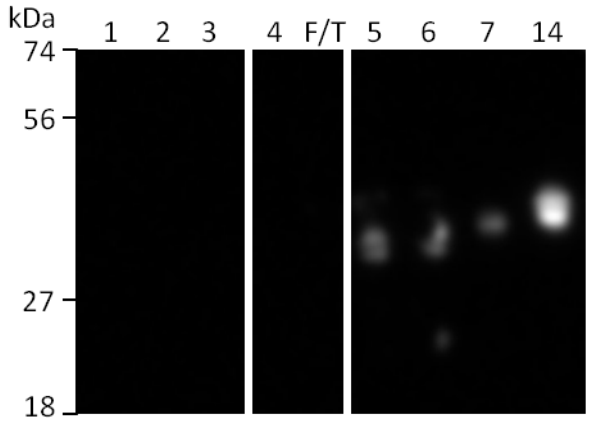


Figure 5-37 Western blot against various fractions from His tag purification of *Leishmania* E2. Samples from various fractions obtained from AKTA His-tag purification of a putative *Leishmania* ubiquitin conjugating enzyme. Lane labels correspond to the fractions present in Figure 5-36. Fractions 1, 2, and 3 are from washes. 4 is waste. F/T is flow through, lysate that did not bind to the column. Fractions 5, 6, 7, and 14 are from various elution peaks. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

Despite few obvious contaminants in fraction 14, size exclusion chromatography was conducted, both to remove the trace contaminants, and exchange the imidazole elution buffer. This process is shown in Figure 5-38a, which indicates the UV absorption of each fraction. As assessed by UV absorption, fractions containing a surfeit of protein were run on gel, to confirm the presence of a protein of the expected mass. Figure 5-38b shows a Coomassie stained gel of the fractions obtained from SEC, where the only detectable band is at the expected molecular weight of ~ 40 kDa. In addition, fractions 5, 6, and 7 from the His tag purification were pooled to undergo SEC as well, demonstrated in Figure 5-39a. However, after SEC, the pool of fractions 5, 6, and 7, do not contain detectable LmxM.33.0900, but a variety of proteins of various masses (Figure 5-39). Given this, work with fractions 5, 6, and 7 was discontinued, and work with fraction 14 continued.

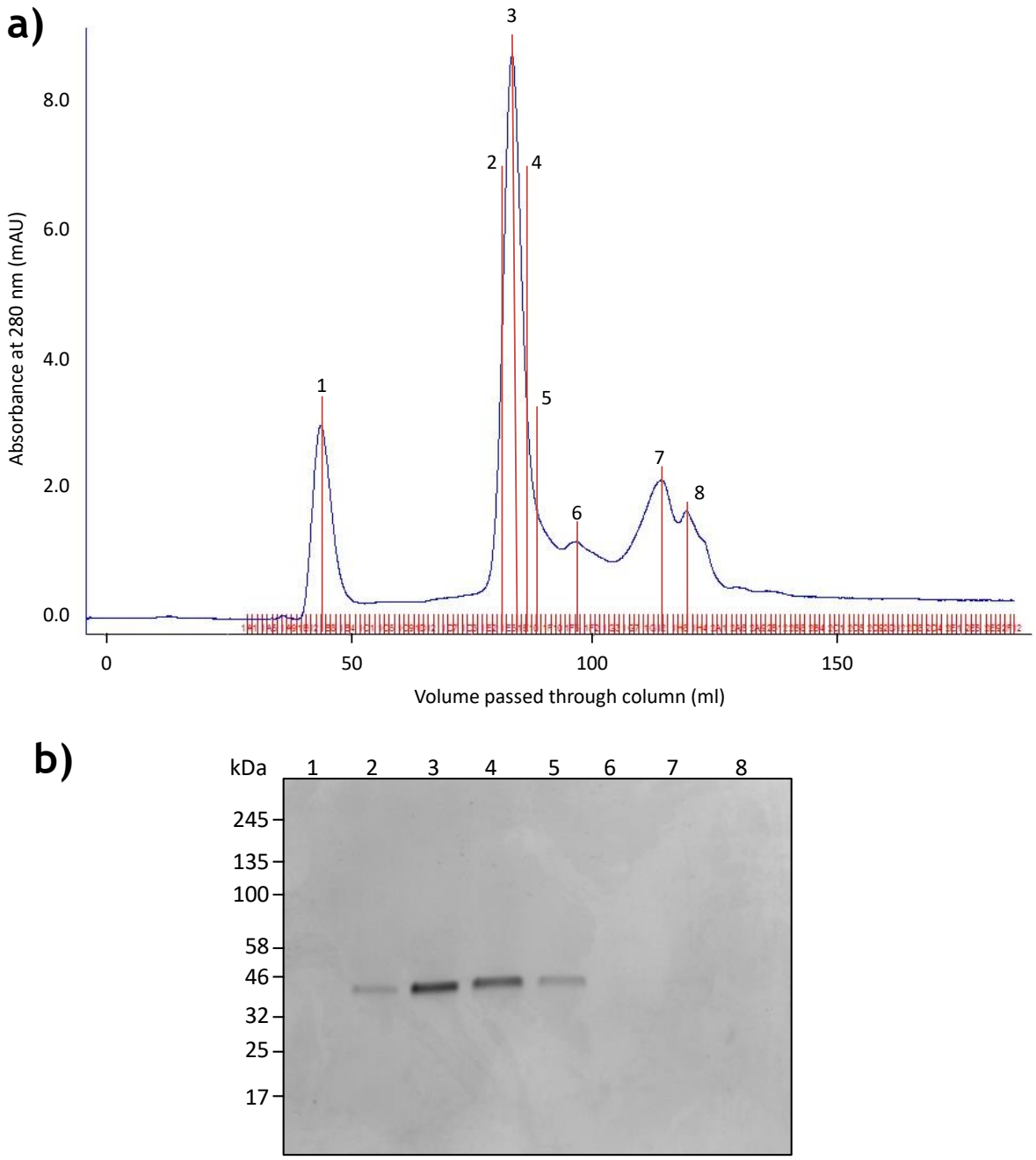
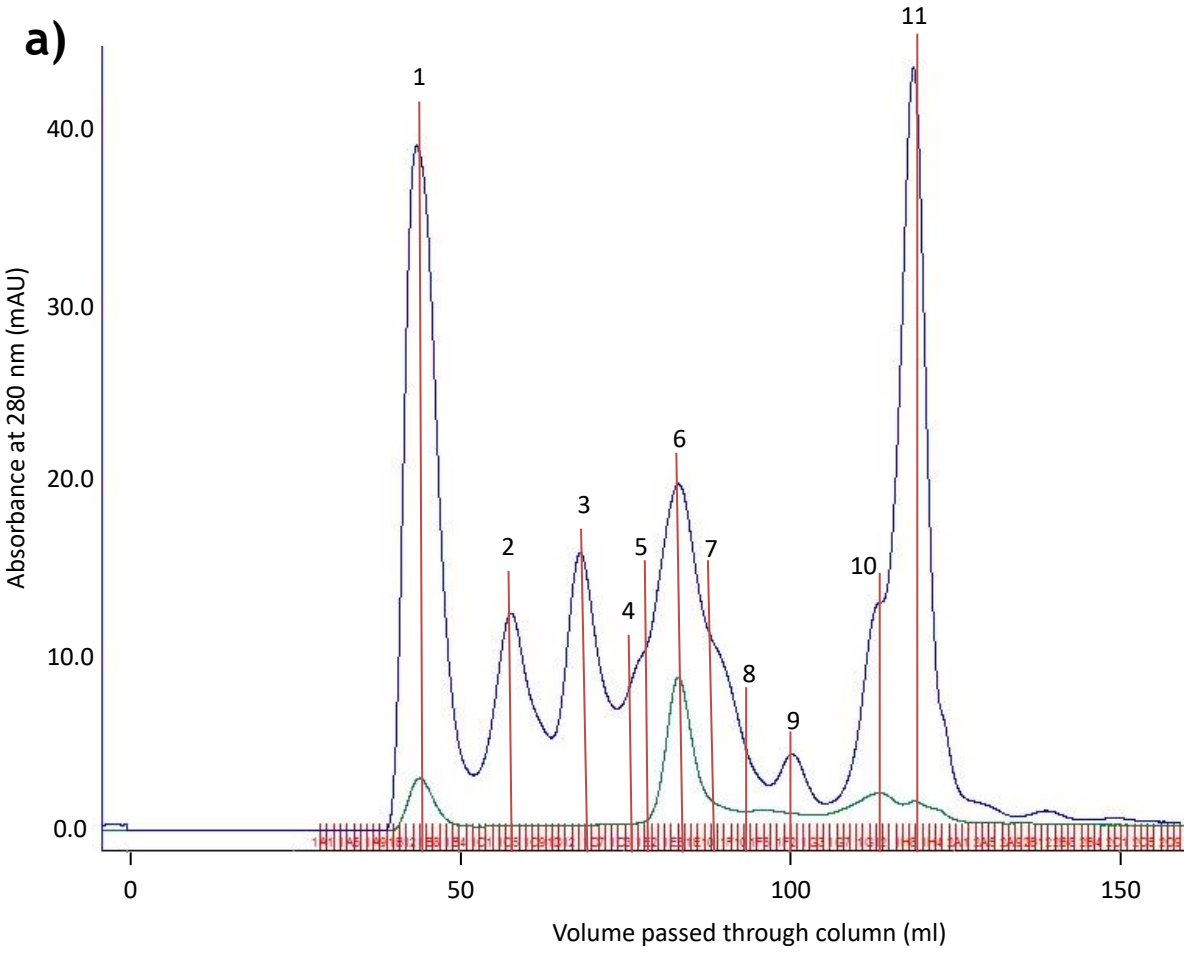


Figure 5-38 Size exclusion chromatography, conducted on an expressed ubiquitin conjugating enzyme from *L. mexicana*. a) Chromatograph of Fraction 14 (referring to prior His tag purification fraction) of LmxM.33.0900 underwent gel filtration. The blue curve indicates absorbance fluctuations at the column's outlet. The red lines and associated numbers indicate the position and identity of fractions that were selected for gel electrophoresis. **b)** Coomassie stain of fractions obtained from gel filtration The lane numbers correspond to the numbered lines in a), so represent the fraction identities. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB



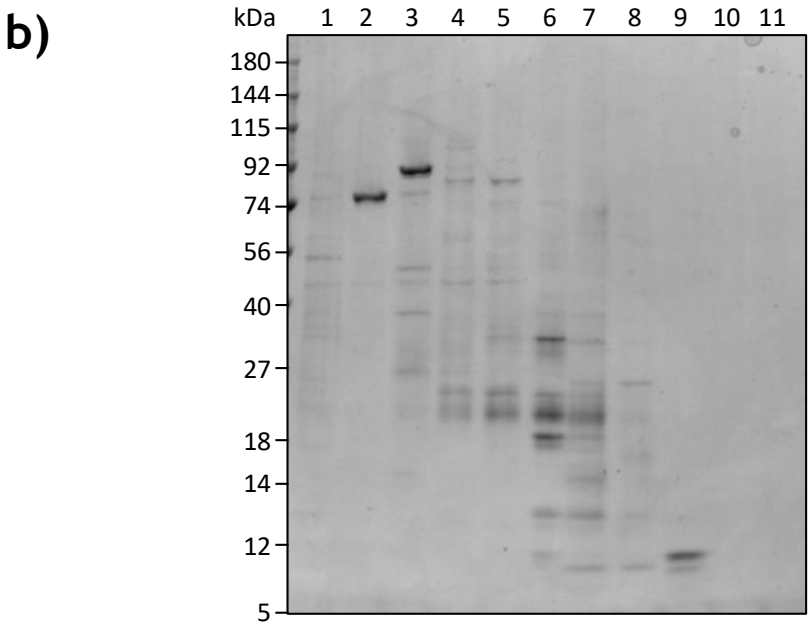


Figure 5-39 Size exclusion chromatography, conducted on pooled fractions of an expressed ubiquitin conjugating enzyme from *L. mexicana*. a) LmxM.33.0900 fractions 5, 6, and 7 (referring to prior His tag purification fraction) were pooled and underwent gel filtration. The blue curve shows how absorbance varies as buffer flows through the column; the green curve indicates the same, but for the earlier sample featured in Figure 5-38 (used to compare the samples). The red lines and associated numbers indicate the position and identity of fractions that were selected for gel electrophoresis. **b)** Coomassie stain of fractions obtained from gel filtration of pooled fractions of a ubiquitin conjugating enzyme of *Leishmania mexicana*. The lane numbers correspond to the numbered lines in a), so represent the fraction identities. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

The purified E2 was tested for activity in a functional assay, to see whether it could accept adenylated ubiquitin from the expressed *L. mexicana* LmxM.34.3060 or yeast E1. The results detailed in Figure 5-40 show an identical result in the experimental and negative control conditions (without and with a reducing sample buffer, respectively), indicating that LmxM.33.0900, expressed under these conditions, was not functionally active.

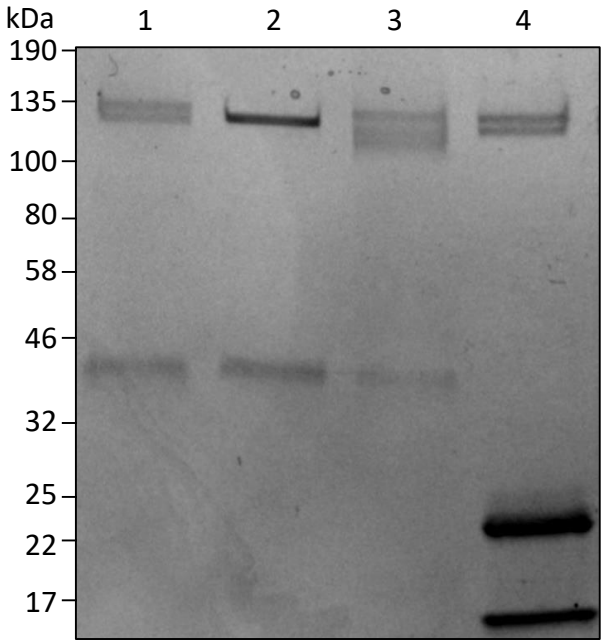


Figure 5-40 Coomassie stain of an activity assay involving synthesised *Leishmania* ubiquitination enzymes. An E1 and an E2 from *Leishmania mexicana* were expressed in *E. coli* and purified, with the enzyme's interactions with human ubiquitin shown here. Lane 1 features the *Leishmania* E1 and E2 in a non-reducing sample buffer, and lane 2 shows an identical reaction, but in a reducing sample buffer (100 mM of DTT). Lane 3 shows the *Leishmania* E2 coupled with an E1 from yeast. Lane 4 functions as positive control, using *Leishmania* E1 and a human E2 (UbcH5a), which are known to be capable of ubiquitin transfer. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

5.3.8 Refolding of LmxM.33.0900

Given the results obtained with LmxM.33.0900 thus far, an attempt was made to refold the enzyme into a functional conformation. LmxM.33.0900 expression was induced under identical conditions; however cells were lysed in a re-solubilisation buffer, containing 8 M urea. After centrifugation, the soluble fraction was again loaded onto a nickel column. Refolding was achieved through washing the column in 3 M urea, before eluting in a buffer without urea. After passing through the nickel column, fractions with a notable abundance of protein were subjected to SDS-PAGE, to confirm their purity. Figure 5-41 shows

these results, demonstrating an abundance of protein of various masses, including a distinct band around the expected size for LmxM.33.0900.

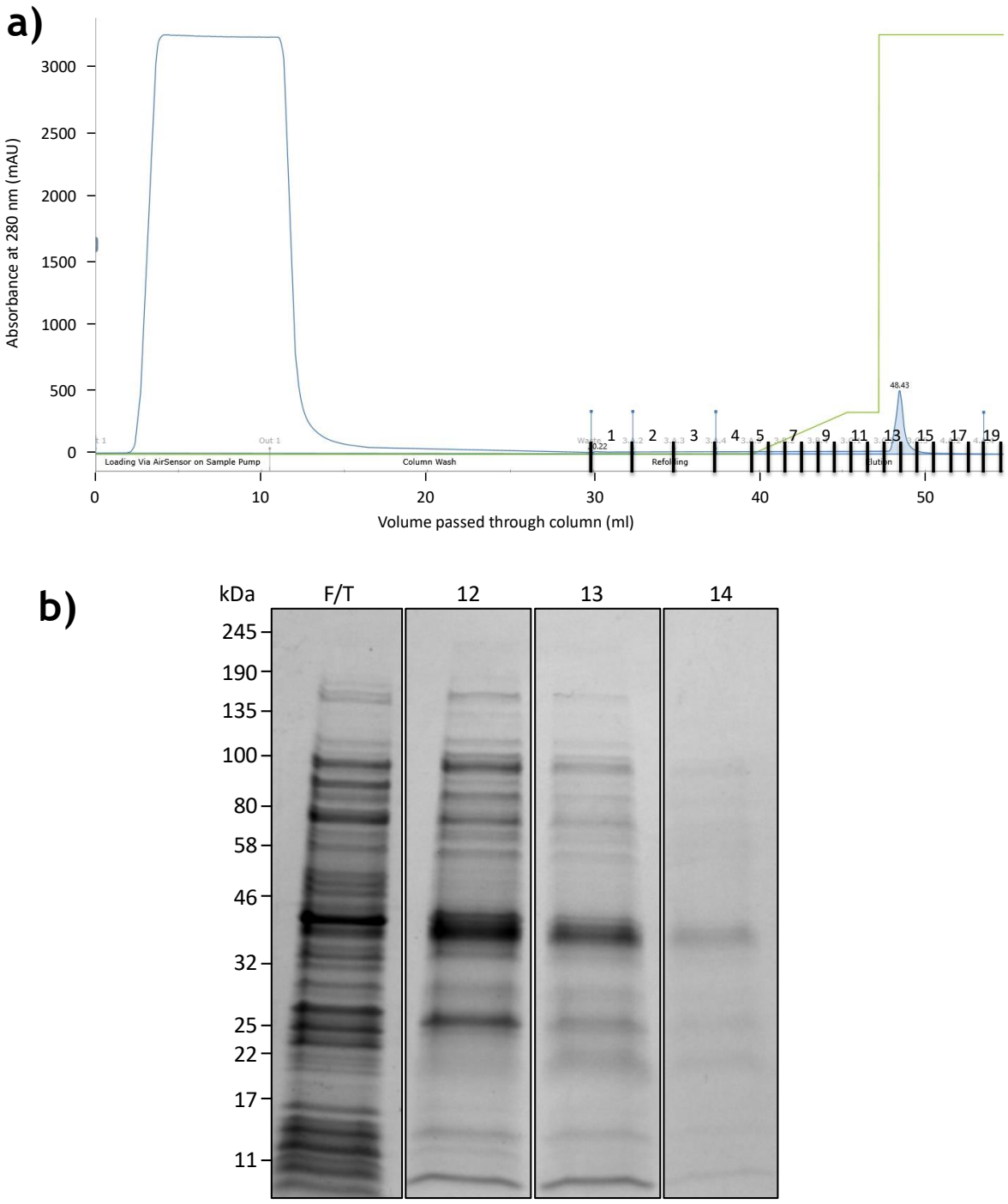
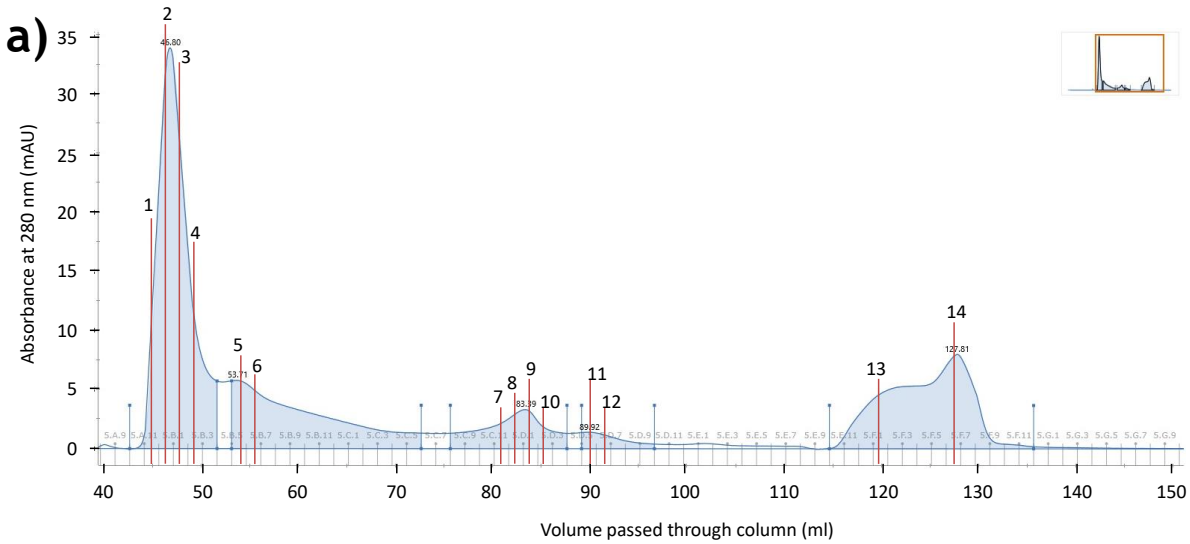


Figure 5-41 His tag purification and refolding of *Leishmania* E2, LmxM.33.0900 a) 800 ml of recombinant *E. coli* cultures were lysed in 8 M urea. The soluble fraction was applied to an AKTA Start and the column washed with 3 M urea refolding buffer, before elution without urea, using an imidazole buffer. Shown are the UV absorbance (blue line), an indicator of total protein abundance, and the concentration of the imidazole-based elution buffer (green line), over the volume passed through the column. The vertical black lines and number show the identity of each fraction. **b)** Coomassie stain of various fractions from the refolding and His tag purification. Lane labels correspond to fractions present in a). F/T refers to flow through. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB.

Given the presence of unexpected molecular weights, fractions 12, 13, and 14 were pooled, and run through size exclusion chromatography (Figure 5-42a). This was particularly revealing, due to the profusion of protein in the void volume (fractions 1 - 6, inclusive). Considering the void volume generally contains proteins too large to enter the gel filtration column, if a ~ 40 kDa protein, the expected size of recombinant LmxM.33.0900, is present it could indicate a misfolded protein. As before, fractions of interest were run on gel electrophoresis to confirm or deny the presence of a 40 kDa protein. Fractions from the void volume did include an abundant ~ 40 kDa protein (Figure 5-42b). When subjected to a functional assay this protein did not interact with LmxM.34.3060, which was known to be functional (data not shown).



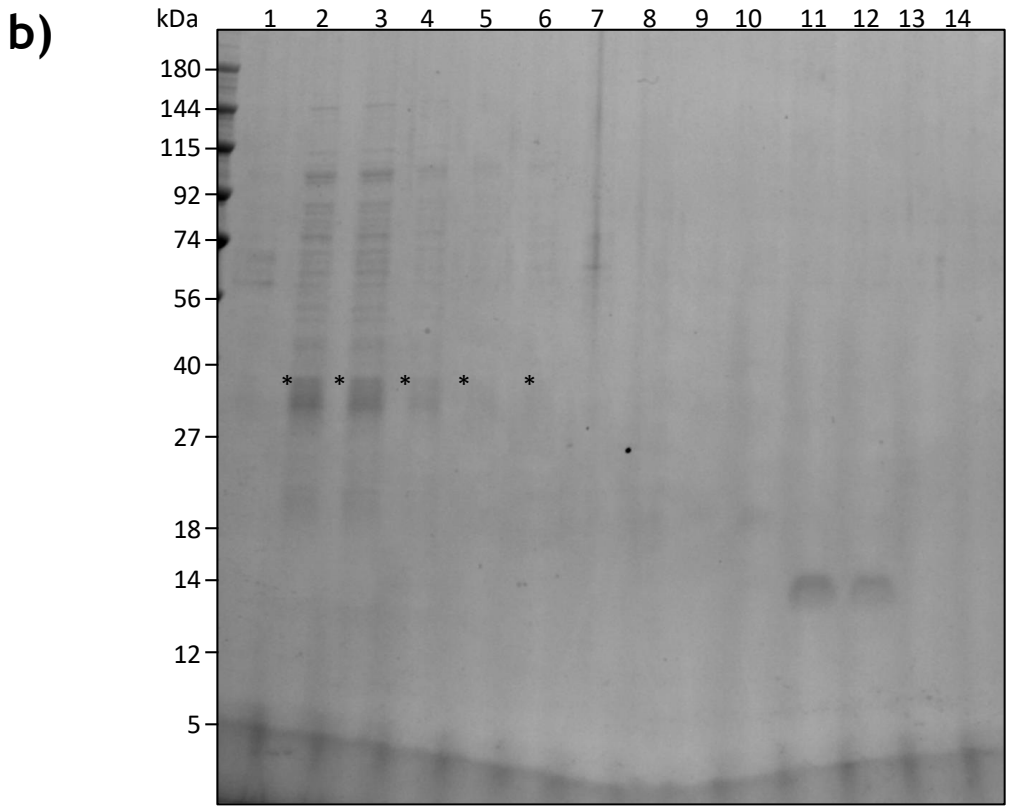


Figure 5-42 Size exclusion chromatography on refolded *Leishmania* E2. a) Refolded, His-tag purified *Leishmania* ubiquitin conjugating enzyme underwent size exclusion chromatography. The blue line indicates how UV absorption, an indicator of total protein content, varied over the volume of solution passing through the column outlet. The vertical red lines and corresponding numbers indicate the identity of a sample. **b)** Coomassie stain of fractions from size exclusion chromatography. Lane labels indicate which fraction is present. Asterisks are used to denote bands of the approximate size of the protein of interest. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB

5.3.9 Expression of LmxM.02.0390

Through bioinformatic strategies, a putative SUMOylation E2 was identified – LmxM.02.0390. However, upon expression in *E. coli*, the protein proved undetectable in the soluble fraction of lysates. Yet LmxM.02.0390 was detectable in the insoluble fraction of lysates (Figure 5-43). Figure 5-43 also shows the insolubility of LmxM.02.0390 is constant across various concentrations of IPTG, or the use of auto-induction media. While acknowledging the low chance of success, due to the low investment of resources required the experiment was repeated with four other transformants of *E. coli*, in the hope that LmxM.02.0390 may be marginally better expressed in a separate transformant line (Figure 5-44). This was not the case, and it became apparent

that LmxM.02.0390 would be unlikely to express in a functional conformation in *E. coli*.

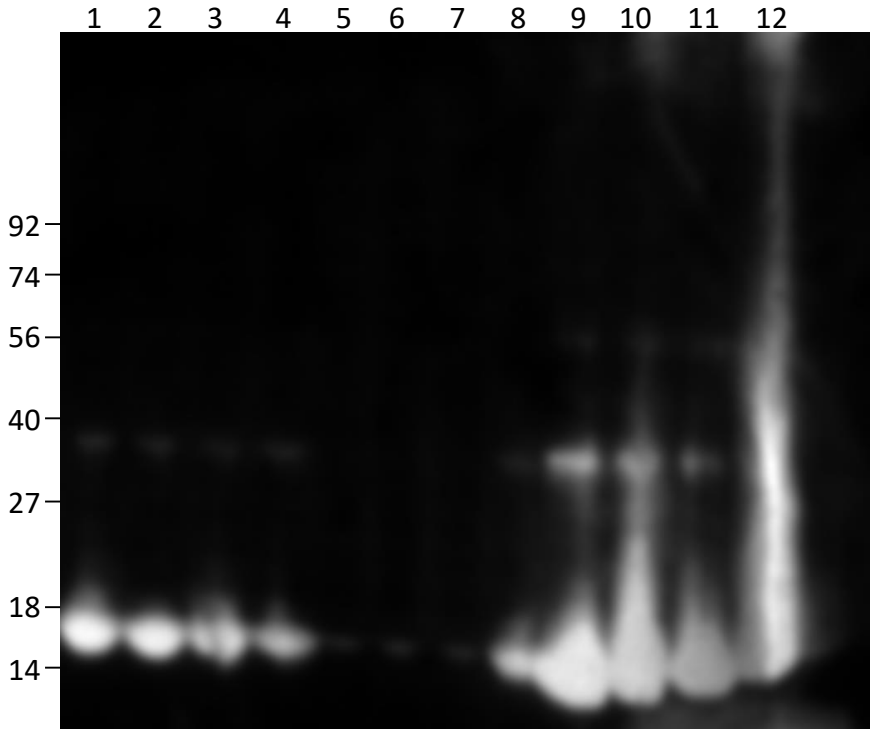


Figure 5-43 Western blot against E. coli lysates, after induction of *Leishmania* SUMO E2 expression. After culturing 10 ml of recombinant *E. coli*, production of LmxM.02.0390 (a putative *L. mexicana* SUMO conjugating enzyme) was induced using IPTG or auto-induction media (AIM). Lanes 1, 2, 3, and 4 show *E.coli* lysates, from bacteria that were induced using 1 mM, 0.5 mM, 0.2 mM, of IPTG respectively, or (in the case of lane 4) AIM. Lanes 5, 6, 7, and 8 show an identical order of IPTG and AIM conditions, but are soluble fractions of lysates, separated using centrifugation. Lanes 9, 10, 11, and 12 again show the four conditions in the same order, but present in the insoluble fraction of the lysates. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

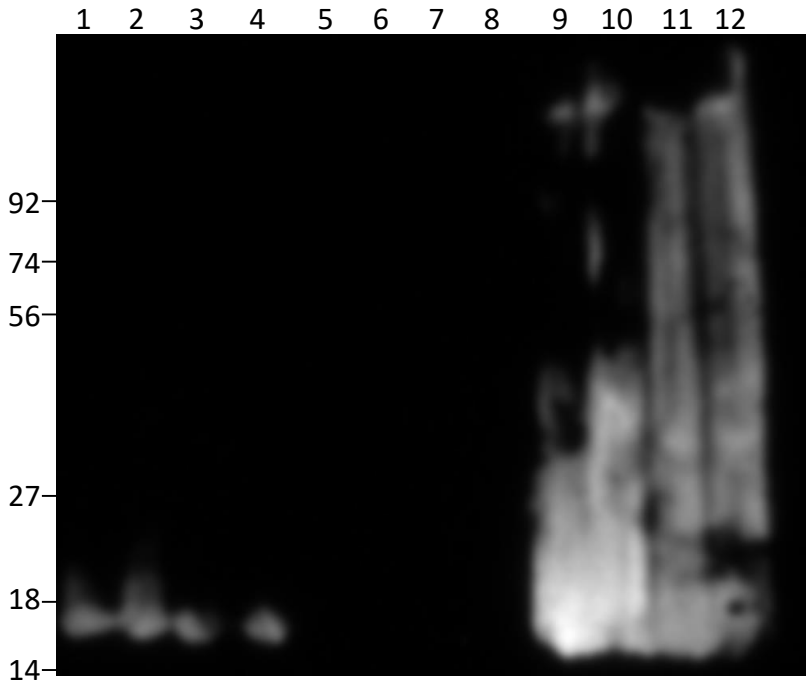


Figure 5-44 Multiple bacterial lysates after autoinduction of *Leishmania* SUMO E2, displayed on an anti-His Western blot. Ten millilitres of BL21 *E. coli*, grown in autoinduction media were incubated overnight at 18°C. Cultures were then lysed via sonication, and lysates from four different cultures are seen in lanes 1, 2, 3, and 4. After centrifugation at 20,000 ×g, soluble and insoluble fractions from the same four lysates were formed. Lanes 5, 6, 7, and 8 demonstrate the protein of interest found in the soluble fractions; and lanes 9, 10, 11, and 12 display the SUMO E2 found in the insoluble fractions. All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

5.3.10 Refolding of LmxM.02.0390

Given the near total insolubility of LmxM.02.0390 expressed in *E. coli*, employing the same refolding protocol as stated above seemed prudent. Although the protocol failed to produce a functional LmxM.33.0900, the smaller size of LmxM.02.0390 (~ 17 kDa) increased the likelihood of success. As before, LmxM.02.0390 was expressed in *E. coli* BL21 derivatives, which were lysed through sonication in a resolubilisation buffer containing 8 M urea. The lysate was then applied to a nickel column, washed with 3 M urea to refold, and eluted in buffer without imidazole, a process that can be seen in Figure 5-45a. The asymmetry of the absorbance peak across fractions 13 and 14 is noteworthy - particularly the protuberance on 14, which interrupts the otherwise smooth descent of the absorbance curve. Again, fractions deemed interesting by virtue of their protein content were run on gel (Figure 5-45b). Initial impressions were of fairly pure fractions, which largely contain protein of the expected mass.

Notably, fraction 14 has an aberration, which has caused an irregularity in the running of the sample.

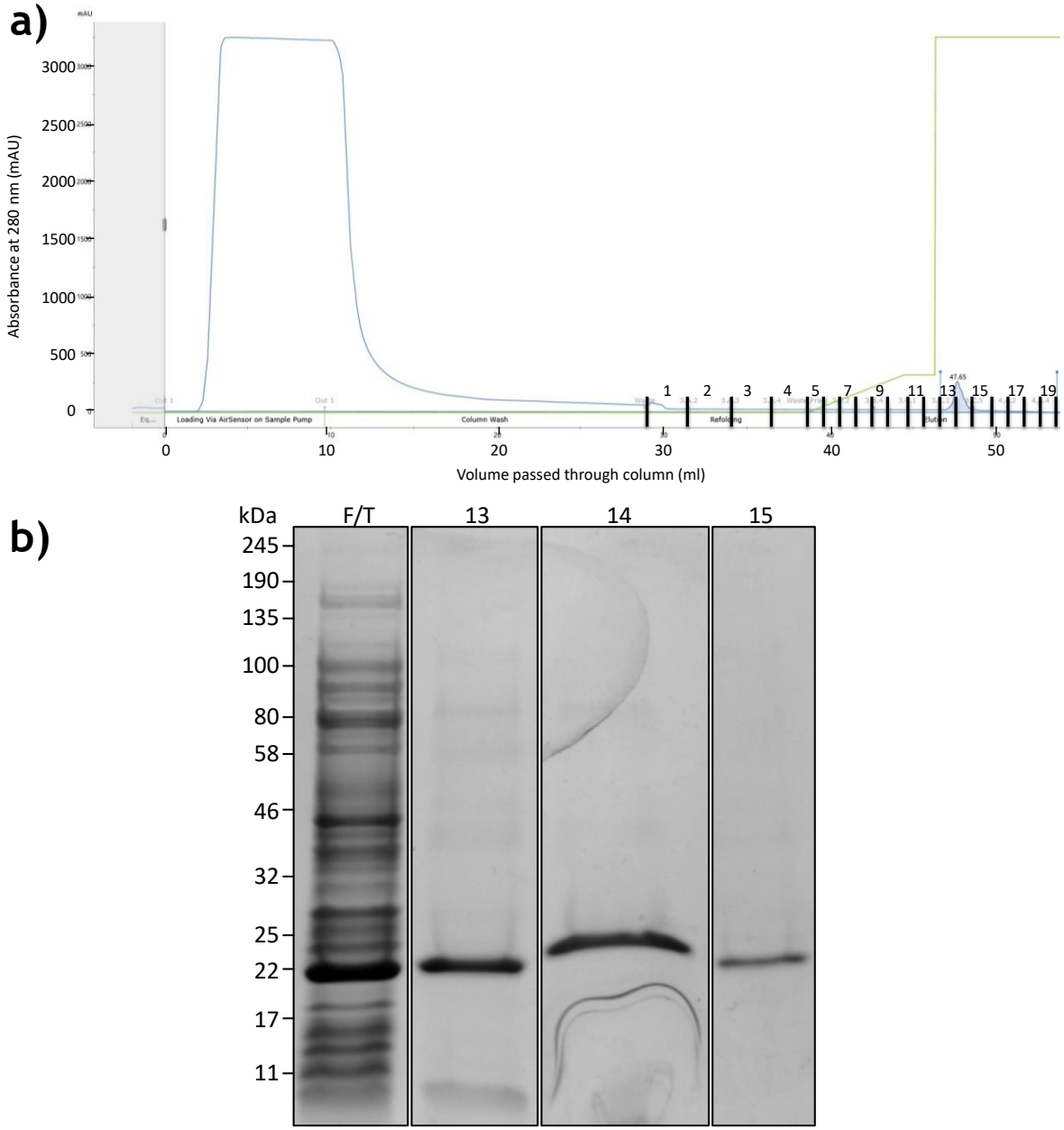


Figure 5-45 Chromatograph of His tag purification and refolding of *Leishmania* SUMOylation E2, after expression in E. coli. a) 800 ml of recombinant *E. coli* cultures were lysed in 8 M urea. The soluble fraction was applied to an AKTA Start, and the column washed with 3 M urea refolding buffer, before elution without urea, using an imidazole buffer. Shown are the UV absorbance (blue line), an indicator of total protein abundance, and the concentration of the imidazole-based elution buffer (green line) over the total volume passed. The vertical black lines and corresponding numbers indicate the fraction identities. **b)** Coomassie stain of fractions from the His tag purification. Lane labels correspond to fractions present in a). F/T refers to flow through, lysate that did not bind to the column. All other lanes feature fractions from the elution. All samples were run on the same 4-20% SDS-PAGE gel and Coomassie stained. Molecular weight markers had product code #P7712, and were supplied by NEB

The outcome of refolding LmxM.02.0390 was again determined by functional assay, using an anti-His Western blot to detect the His tagged LmxM.02.0390, and a His tagged known substrate of SUMO in the orthologous human pathway, termed RANGAP (Figure 5-46). For fraction 13, multiple masses of LmxM.02.0390 are present without a reducing agent, but all mass shifts are eliminated upon addition of DTT. However, LmxM.02.0390's ability to SUMOylate RANGAP is ambiguous. The evidence is a single faint band, which is resistant to reducing agents (as the RANGAP-SUMO covalent bond should be). Fraction 14 did not produce the multiple bands indicative of functionality (data not pictured), however it is unknown whether this is due to incorrect folding, or an unexpectedly low protein concentration in the stock solution.

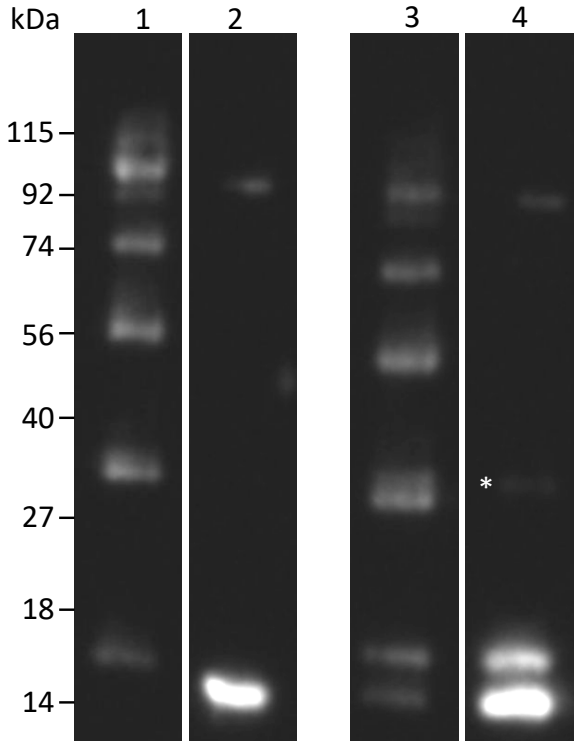


Figure 5-46 Anti-His Western blot, showing in vitro SUMOylation, against refolded His tagged *Leishmania* SUMO E2, and His tagged human RANGAP. Refolded *Leishmania* SUMO E2 was incubated with human SAE1, human SUMO1, ATP and $MgCl_2$. Lane 1 shows the reaction of these components in a non-reducing sample buffer, compared to lane 2 that included 111.1 mM DTT. Lane 3 also contained refolded *Leishmania* SUMO E2, human SAE1, human SUMO1, ATP, $MgCl_2$, as well as human RANGAP in a non-reducing sample buffer. Lane 4 features an identical reaction to 3, but in a reducing buffer, employing 111.1 mM DTT; an asterisk is used to denote a faint band All samples were run on the same 4-20% SDS-PAGE gel, transferred to a PVDF membrane, and the protein of interest detected using α-His antibodies. Molecular weight markers had product code NXA6050, and were supplied by Expedeon.

5.4 Discussion

5.4.1 An E1 identification validated

Here it is shown that a putative ubiquitin activating enzyme, LmxM.34.3060, identified through bioinformatics and proteomics, interacts with ubiquitin in a manner dependent on the presence of ATP and absence of a reducing agent – the behaviour expected of any Ubl E1 (Hjerpe *et al.*, 2012). Searching the UniProt database using the PFAM domain pf10585 (termed "UBA_e1_thiolCys", referring to the cysteine active site of the ubiquitin-activating enzyme), and filtering for only those results that have evidence at protein level, yields 45 results from 14 organisms. It is worth noting that most of these hits likely do not function to activate ubiquitin. If functional, they may activate a Ubl substrate (hence the ratio of results to organisms). Furthermore "evidence at protein level" does not demand expression of a protein alongside a functional assay. For that reason, the 14 organisms should be considered a generous upper limit of the number of organisms with a well-characterised ubiquitin activating enzyme. Still, these organisms are predominantly model organisms, with the possible exception of *Triticum aestivum* (wheat), and *Oryctolagus cuniculus* (rabbit). None are protozoan, and none are parasites. For this reason, a *L. mexicana* E1 adds diversity to the catalogue of characterised ubiquitin E1 orthologs.

It was also demonstrated this *Leishmania* E1 could activate human ubiquitin, then transfer it to the human E2 UbcH5a. Relative to *L. mexicana* ubiquitin, human ubiquitin has 2 differences in the 76 amino acid protein (97% identity shared), and one of those two divergent amino acid residues is a conservative substitution. As such, it is perhaps no surprise that an *L. mexicana* E1 is not selective between human and *Leishmania* ubiquitin. On a similar note, a BLAST on the *L. mexicana* genome, with UbcH5a as the query sequence, yields a top scoring hit with an E value of 2e-80 (LmxM.34.1300). This hit has 72% of amino acids identical to UbcH5a, and 88% of amino acids have similar chemical properties. It would be interesting to see how dissimilar a non-*Leishmania* E2 would have to be from any *Leishmania* orthologs before an inter-species E1-E2 assay stopped functioning.

5.4.2 Drug assays

A much lower concentration of TAK-243 is needed to inhibit the human E1 UBA1 than is needed to inhibit its *Leishmania* ortholog. From this it can be inferred that the ATP binding pockets of both enzymes differ. This is notable as it provides support for future drug discovery efforts – had both E1s been inhibited equally well by TAK-243 it would mean both ATP binding sites were similar, and the possibility of specificity for any adenosine sulfamate based drug remote.

Both nucleocidin and 5'O-sulfamoyl adenosine inhibit the *Leishmania* E1 more than the human ortholog. 5'O-sulfamoyl adenosine in particular was approximately 50-fold more potent against the parasite than human enzyme, inhibiting ~50% of the parasite's E1 activity at 1 μM. Although nucleocidin was less potent, it should be noted that the nucleocidin used came from a small-amount (0.193 mg) of semi-pure stock, which adds uncertainty to the measurements – a purer stock, with less associated error in weighing, may have better potency. Regardless, this is the first time these compounds have been shown to inhibit an E1, as opposed to protein synthesis. This additional mechanism of action could help resolve previously unexplained phenomena in the literature, such as the rapid action of nucleocidin, which authors acknowledge is difficult to explain by inhibition of protein synthesis alone (Florini, Bird and Bell, 1966). Also, the differences in inhibition between the human and parasite E1s adds further weight to the possibility of designing an inhibitor specific for the parasite E1. Furthermore, this matches the literature that states these inhibitors were more toxic to the parasites than mammalian hosts – as the inhibitors were more effective against parasite enzymes than human enzymes (yet they still inhibited the human E1) (Thomas *et al.*, 1956; Florini, Bird and Bell, 1966; Shuman, Robins and Robins, 1969). Perhaps these molecules, particularly 5'O-sulfamoyl adenosine, could be treated as initial lead compounds, with future work focusing on tailoring the nucleobase extension to the drug binding pocket of LmxM.34.3060, adding additional selectivity and rendering these compounds less toxic. Of course, this may be premature, as toxicity to *Leishmania* has not been demonstrated by this assay. To do that, and bolster the case for further research, a cell viability assay, such as an alamar blue assay, would need to be conducted.

5.4.3 Crystal structure

A crystal structure of LmxM.34.3060 is presented here. This puts it amongst 15 other structures of full length ubiquitin activating enzymes (found through searching the InterPro database for structures under entry IPR018075, the code for "Ubiquitin-activating enzyme E1"). Of these structures eight are from *Saccharomyces cerevisiae*, six from *Schizosaccharomyces pombe*, and one from humans. Given the paucity of species represented in the database, the *L. mexicana* E1 adds some much-needed diversity. This is evidenced in the percentage identity matrix in Table 5-3. The *L. mexicana* E1 has, at most, 32.57% shared identity with any of the existing structures, whereas the existing structures share at least 53.37% of their identity with one another. The resolution of the existing structures range from 2.032 Å to 3.153 Å, so, at 3.1 Å, the *L. mexicana* structure is comparable. It is also worth noting that this is not the first time that protein crystallography has been attempted on a kinetoplastid ubiquitination E1, but this is the first time it has been successful (Boer and Bijlmakers, 2019).

	L. mexicana	*S. cerevisiae*	*S. pombe*	Human
L. mexicana	-	32.47	32.20	32.57
S. cerevisiae	32.47	-	59.31	53.37
S. pombe	32.20	59.31	-	54.17
Human	32.57	53.37	54.17	-

Table 5-3 A percentage identity matrix for published ubiquitin E1 structures, and the *L. mexicana* E1 characterised in this study.

Lessons learned from the crystallisation process include the importance of purification – His tag purification, ion exchange chromatography, and size exclusion chromatography did not result in a crystallisable protein. Perhaps this is because homogeneity was lacking. When an extra, function-based, purification step was introduced, reacting LmxM.34.3060 with GST tagged ubiquitin then purifying with glutathione-agarose, crystals were more forthcoming. Assuming that not all LmxM.34.3060 expressed in *E. coli* is functional, the Ub-GST purification removes non-functional protein, in a way no other purification method could manage (other methods selected for proteins with a similar tag, size or isoelectric point). Despite the otherwise overwhelming need for uniformity in the protein mixture, an unknown protein, likely an E1 fragment,

around 70 kDa was present after all rounds of purification. This did not seem to affect crystallisation. Another key ingredient for obtaining protein crystals of LmxM.34.3060 was the addition of ATP·Mg. ATP·Mg was also necessary for the formation of all other ubiquitin E1 protein crystals reported in the surveyed literature (Olsen *et al.*, 2010; Schäfer, Kuhn and Schindelin, 2014; Lv *et al.*, 2018). If, in the absence of ATP·Mg and Ub, the open and closed conformations of LmxM.34.3060 exist in equilibrium, as is the case in the *S. pombe* E1 (Lv *et al.*, 2017), the addition of those ligands should shift the equilibrium in favour of the open conformation. Yet, despite its presence in the protein mixture, ubiquitin was not present in any protein crystals. Perhaps the thioester Ub~E1 bond was too unstable to exist within the crystallisation conditions examined, so perhaps a crosslinker would be necessary to maintain the Ub~E1 structure. Still, ATP·Mg are visible in the crystal structure. This may have aided crystallisation by ensuring a homogenous open conformation throughout the crystallising E1s. If the three important factors of real estate are location, location, location, then the equivalent three factors of crystallography must be purity, purity, and homogeneity (McPherson and Gavira, 2014).

The reason the author and others (Bijlmakers, 2021) have displayed an interest in protein crystallography on E1s is because a structure permits drug modelling. Drug modelling on *Leishmania* LmxM.34.3060 has enormous potential as the enzyme is necessary for proliferation of the parasite in mice (Burge *et al.*, 2020). Within LmxM.34.3060, of particular interest is the adenylation site, which would be key to the success of any adenosine sulfamate based inhibitor. When TAK-243 was fitted to the structure of the bound ATP, two possible clashes were found; when the structure of TAK-243 was copied from an orthologous structure in the Protein Databank, two (different) clashes were nonetheless present. The clashes, defined as a covalent bond between drug and enzyme or a hydrogen bond of ≤ 2.55 Å, predominantly occur within the C-terminal of TAK-243. This is thought to be the specificity determining component of adenosine sulfamate based anti-E1 drugs (Misra *et al.*, 2017), and so would be where clashes would be predicted. This also correlates well with the biochemical assay, which found higher concentrations of TAK-243 necessary for inhibition of LmxM.34.3060 – a phenomenon explained by the clashes predicted in the drug model.

5.4.4 SAXS data

Using SAXS, an in-solution structure was also obtained. The scatter plot (Figure 5-22) shows an uptick at very small values of q – this could represent protein aggregation or noise from beamstop scattering. However, the linearity of the Guinier region indicates good sample quality, and the integrated intensity plot shows buffer subtraction worked as it plateaus at low values of q. Given this, and the coherence of later analyses, the uptick at very small values of q is likely due to beamstop scattering. The Kratky plot is typical of a globular protein, with its bell-shaped curve that peaks at low values of q. There is also no indication from the Kratky plot that LmxM.34.3060 is flexible, so the various molecules likely exist in the same conformation in the SAXS sample. This may be surprising, as it has been suggested that the open and closed conformations of the *S. pombe* E1 exist in equilibrium prior to the binding of ATP·Mg and ubiquitin (Lv *et al.*, 2017). Nonetheless, the Kratky plot concurs with the flexibility analysis – in which the Porod-Debye plot, which indicates a compact particle, plateaus before the Kratky-Debye plot, used to determine flexibility. Completing a coherent picture of a compact, globular particle is the p(r) distribution. Its slightly asymmetrical bell-shaped peak is indicative of a slightly elongated sphere. A further mark of consistency is the small difference between real and reciprocal parameters – whether calculated through the Guinier approximation or p(r) distribution there is only a difference of 1E-3 photons per second for I(0), and a difference of 0.58 Å for R_g. To the author's knowledge, this is the first time SAXS has been conducted on a ubiquitin E1, however a globular, near-spherical protein would fit with existing crystallographic E1 structures (Lv *et al.*, 2018).

Of particular interest are the non-overlapping domains between the SAXS and crystallographic structures. BLAST searches of the amino acids present in the crystal structure's non-overlapping domain aligns with the adenylation domains of characterised E1s. This is noteworthy as the adenylation domain has been shown to rotate by 106° in the *S. pombe* ubiquitin E1 (Lv *et al.*, 2017), and 130° in the human SUMO E1 (Olsen *et al.*, 2010). Therefore, it is possible that the *Leishmania* E1 has been captured in both open and closed conformations. Given that crystals were found only with the addition of ATP·Mg, whilst SAXS was conducted on a sample without those ligands, it is not implausible. As the addition of ATP·Mg should shift the open:closed equilibrium toward the open

conformation, any structure disparate from that (i.e. the SAXS structure) would represent the closed conformation. Alternatively, the spatial demands imposed by the crystallisation process may have forced the flexible adenylation domain into an artificial position, one that is not representative of either open or closed conformations. Finally, the agreement between the crystal and SAXS structures is quantified using the CRYSOL software. A theoretical SAXS dataset was calculated, based on predictions as to how the LmxM.34.3060 crystal structure would scatter X-rays in a SAXS experiment. It can be seen the theoretical scattering curve overlaps considerably with the scattering curve obtained from the actual SAXS experiment, providing further validity to the model. Given this concurrence between the crystal and SAXS models, it can be said that the crystal structure is broadly representative of an E1 in solution.

5.4.5 SUMO E2

An expressed SUMOylation E2, a *L. mexicana* LmxM.02.0390, is presented here. It is able to accept activated human SUMO1 from the human SUMO E1. BLASTing the human SUMO1 amino acid sequence against the *L. mexicana* genome yields a top hit with an E value of 9e-23 (gene ID: LmxM.08.0470). This hit shares 51% identity with the human SUMO1, which increases to 74% when permitting conservative mutations. BLASTing the human SUMO2 isoform yields the same top hit, which maintains a similar relationship to the human protein (±4% for both identity and similarity). Under standard search parameters, this hit is not returned when the human ubiquitin sequence is used as bait in a BLAST. As such, it is likely that *L. mexicana* has a separate SUMOylation system, with possibly just the one SUMO isoform. A solitary SUMO isoform would fit the published literature on *Trypanosoma brucei*, and stand in contrast to the four human isoforms (Liao *et al.*, 2010). It is interesting that, unlike ubiquitin, which is highly conserved between organisms, SUMO proteins seem to have diverged more. Perhaps this attests to ubiquitin's essentiality, and SUMO's less necessary functions (though still essential in *Trypanosoma brucei* for cell cycle regulation (Liao *et al.*, 2010), and proteomics has revealed diverse roles for SUMO in *Trypanosoma cruzi* (Bayona *et al.*, 2011)).

Of course, the *Leishmania* E2 worked alongside a human E1 in this assay, and a similar bioinformatic comparison between characterised human and putative *L.*

mexicana E1s can be conducted. The human catalytically active subunit UBA2 displays 40% identity and 54% similarity to its putative *Leishmania* ortholog. Curiously, the catalytically inactive human subunit SAE1 displays more similarity to a *Leishmania* ubiquitin E1 (LmxM.23.0550) than to its predicted ortholog (LmxM.28.0360). Further evidence that the discipline of bioinformatics requires more than BLAST searches. Still, the predicted ortholog possesses 43% identity and 59% similarity. Why this fixation on percentage identities? Because it follows that the *L. mexicana* E1 and E2 of SUMOylation have diverged alongside their SUMO protein. But, more importantly, because this study has built two inter-species Ubl cascades, and it is interesting to note how much variability the components can tolerate before their orthologs are no longer recognised. It is also stimulating to speculate about the selective pressure placed on a system by considering its capacity to diverge.

Nonetheless, this has been a digression from *L. mexicana* SUMO E2 that is demonstrated here. Why did refolding the SUMO E2 work when refolding the ubiquitin E2 failed? It is likely to do with the shorter length of the former – as shorter proteins have fewer amino acids, it seems plausible that this results in fewer biochemical interactions between those amino acids. Therefore, when denaturing conditions are gradually alleviated, the odds of a few amino acids adopting the proper conformation seem greater than the odds of many amino acids spontaneously adopting their proper structure (given a lack of chaperone proteins, at least).

A surprise finding was the apparent assembly of reducing agent sensitive polySUMO chains in the functional assay. An interesting and plausible explanation for this is the *in vitro* assembly of polySUMO chains by LmxM.02.0390, chains which continue to interact with the LmxM.02.0390 active site, ready to be transferred *en bloc* to a substrate. That is not to say that this biochemistry exists *in vivo*, as it could well be an artifact caused by the presence of a SUMO conjugating enzyme and SUMO1, without a substrate. Still, the formation of polySUMO chains forming on a SUMO E2 during an *in vitro* assay has been noted by other authors (Yang *et al.*, 2006). Other authors have noted the existence of polySUMO chains *in vivo*, and presented a structure of human SUMO E2 non-covalently interacting with SUMO to promote the formation of

polySUMO chains (Knipscheer *et al.*, 2007). Could the *Leishmania* LmxM.02.0390 possess the same domain, used to the same effect? Figure 5-47 presents an alignment of the *L. mexicana* and human LmxM.02.0390s. If anything, the lack of conservation of the human E2's non-covalent binding site is all the more conspicuous given how well conserved the rest of the enzyme is. It is perhaps more profitable to compare the *Leishmania* SUMOylation system to a closer relative. Fortunately, the SUMOylation enzymes in *T. brucei* have been characterised (Ye *et al.*, 2015). Adding further weight to this characterisation of a *Leishmania* SUMO E2 is the fact that it is a syntenic ortholog of the *T. brucei* SUMO E2.

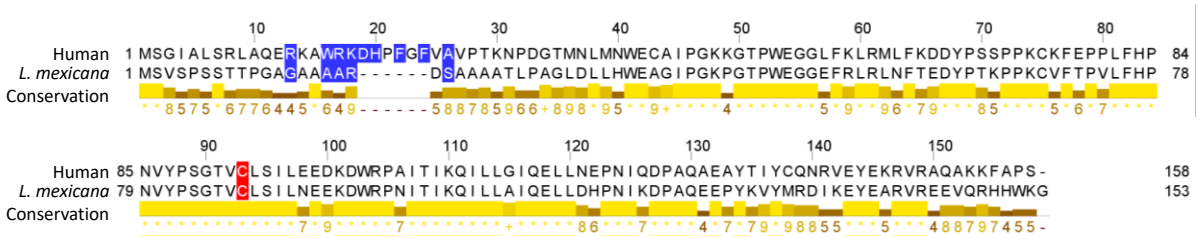


Figure 5-47 Sequence alignment of human and *L. mexicana* SUMO E2s. Highlighted are the non-covalent LmxM.02.0390-SUMO interaction site (blue), and the cysteine active site (red), which were characterised in the human enzyme.

What is in more doubt is LmxM.02.0390's ability to SUMOylate RANGAP. There is evidence in a single faint band the expected weight of SUMOylated RANGAP, which is resistant to reducing agents (as the RANGAP-SUMO covalent bond should be). Yet, given the weakness of this band (Figure 5-46), it is not conclusive evidence of LmxM.02.0390's ability to SUMOylate a mammalian substrate. However, this could be explained as an inefficient reaction. The greater divergence of the SUMOylation machinery compared to the ubiquitination enzymes has already been noted, perhaps this differentiation results in reduced reaction kinetics when disparate, multi-species components are expected to work in tandem.

5.4.6 Structural predictions

Recently, advances in the field of artificial intelligence, namely from Google's DeepMind subsidiary, lead to the creation of AlphaFold – a software capable of highly accurate protein structure predictions (Jumper *et al.*, 2021). Theoretically, all the information about a protein's final three-dimensional structure should be obtainable *a priori* from its amino acid sequence.

Practically, understanding the thermodynamics of the innumerable interatomic forces that determine this structure is immensely challenging, to say the least. As such, the protein folding problem has remained intractable since the 1950's. However, AlphaFold evades that question through standard machine learning – identifying correlations between the sequence and structure of the approximately 170,000 proteins in the protein databank, then applying these correlations to sequences with unknown structures (Ball, 2020). Yet the strength of this technique – avoiding the quagmire of complex thermodynamics in favour of noting correlations – is also its weakness: machine learning is only competent within the bounds of its training set. If the training set used by AlphaFold to make predictions is focused on human proteins, featuring few early branching organisms, then predictions for early branching organisms will be less accurate. As such, efforts have been made to incorporate data on early branching organisms, such as *Leishmania* (Wheeler, 2021). Using this tailored version of AlphaFold, predictions have been made for some of the proteins of interest listed here.

A particularly interesting candidate for structural prediction would be LmxM.34.3060 – given the empirically determined structure published here, a similar prediction from AlphaFold would validate the program. Unfortunately, such a prediction cannot be provided. Obviously, running AlphaFold is computationally intensive. Access to suitable processors is provided by Google, free of charge, with certain terms and conditions. One term is a maximum usage time of 12 hours. Perhaps due to their demanding nature and fair usage of resources, larger proteins are broken into discrete parts. This in itself is problematic, as a large protein may not function as the sum of disconnected blocks, but as a holistic entity in which domains, although distant on a bioinformatic string of amino acids, interact with one another due to their proximity in a three-dimensional structure. Regardless, LmxM.34.3060, comprised of 1154 amino acids at full length, is broken into three distinct overlapping chunks of 399, 552, and 604 amino acids. The smallest segment required approximately 10.5 hours to be modelled. The second largest segment failed to be modelled within 12 hours. This precludes obtaining a full length model of LmxM.34.3060, and shows that as advanced as an algorithm may be, it is only helpful if you have the resources to run it.

A far smaller protein is LmxM.02.0390, the SUMO E2, which is 153 amino acids. This required around 3.5 hours of processing time. The predicted structure is shown in **Error! Reference source not found.**. The confidence of this AlphaFold structure corresponds to the sequence alignment between human and parasite enzymes, shown in Figure 5-47. A poor alignment is shown between residues 12 to 19 (inclusive) of the *Leishmania* enzyme with its human ortholog, corresponding to a deletion of the sequence the human enzyme employs to non-covalently interact with SUMO (Knipscheer *et al.*, 2007). A similar stretch (residues 10 to 24, inclusive) has poor confidence in its structural prediction as well, although it is predicted to form a flexible loop. Due to their disorder, the structure of flexible loops is difficult to predict. However, whether it has a low confidence because it is a flexible loop, or whether it is predicted to be a loop because it is in a low confidence region (due to poorly matching homology models) is debatable. It is perhaps noteworthy that, when searching the Protein Data Bank, none of the orthologous structures possess such a loop. If the *Leishmania* enzyme also interacts with SUMO via this site, it would be unusual if this protein-protein binding site had such disorder. Nonetheless, the ß-grasp fold has its structure predicted with high confidence, although this is unsurprising for a well conserved, highly ordered domain with well-characterised functions. It is the less well characterised domains, such as the backside binding site of an E2, that remain the most interesting. Yet here it is also the most the divergent area, and so it is here that structural prediction falters.

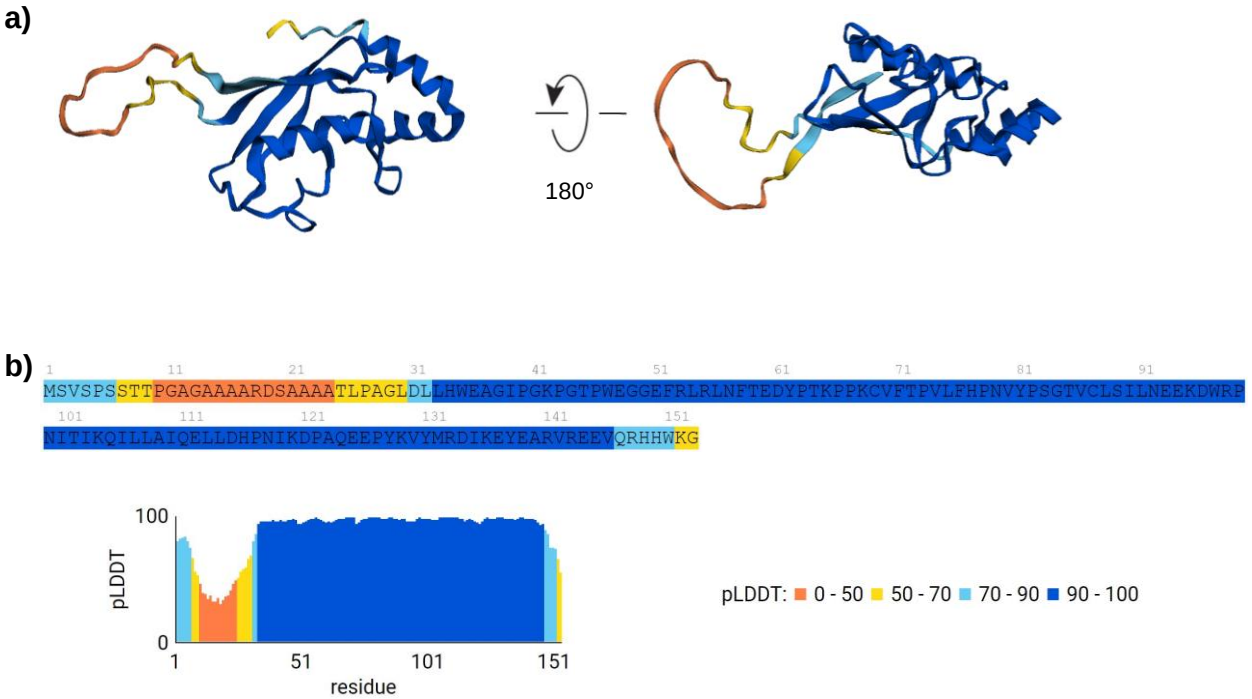


Figure 5-48 AlphaFold structural prediction for LmxM.02.0390. a) Structural prediction by AlphaFold for a *Leishmania* SUMO E2. **b)** Depictions of a per residue confidence metric, pLDDT. Higher scores indicate greater confidence.

The ubiquitination E2, LmxM.33.0900, at 292 amino acids, demanded approximately 6.5 hours of processing time (although it should be noted that the initial run of AlphaFold failed to produce data after apparent "completion", so the total time spent on this structure was ~13 hours). This structure is shown in Figure 5-49. Immediately obvious is the flexible loop that encircles the main body of the protein. This disordered segment starts at residue 172 and continues until residue 267. This segment, approximately a third of the protein, lacks any secondary structure at all, producing a structure that is preposterous even at first glance. Of course, it should be noted that the confidence of this structure is rated as low, so AlphaFold is aware of the structure's absurdity. This lends credence to the theory that regions of low confidence are predicted to be flexible loops (as opposed to flexible loops being hard to predict). So why is this region difficult to predict? Well the sequence alignment (shown in Figure 4-22) shows this region differs substantially from those of model organisms. Compounding this problem is that the only published structure of an ortholog, the human enzyme E2 J2, only features the core domain – missing the last 74 amino acids of the C-terminal. As stated earlier (4.4.1), orthologs of LmxM.33.0900 possess a transmembrane domain at their C-terminal (Claessen *et al.*, 2010). Given the abundance of hydrophobic residues, transmembrane

domains are notoriously difficult to crystalise, which is perhaps why a full length E2 J2 has yet to be solved through protein crystallography. Having said that, there is a high confidence structure at the C-terminal – the terminal 20 amino acids, of which the majority are hydrophobic, are confidently predicted to form a helix. This is also a conspicuously well-conserved component when viewing the sequence alignment (Figure 4-22), and was hypothesised to form a transmembrane domain. The core UBC domain, conserved among orthologs, is predicted with high confidence. Again, this structure demonstrates that predictions are hard to make, especially about structures you don't already know.

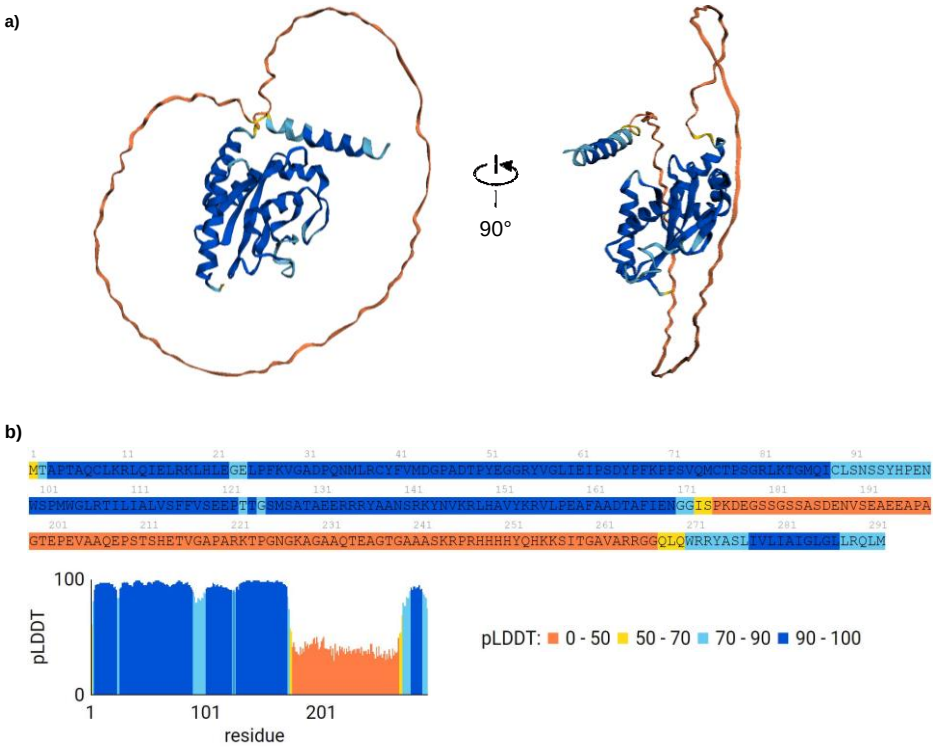


Figure 5-49 AlphaFold structural prediction for LmxM.33.0900.a) Structural prediction by AlphaFold for a Leishmania ubiquitination E2. **b)** Depictions of a per residue confidence metric, pLDDT. Higher scores indicate greater confidence.

5.5 Summary

This study has presented a functional *Leishmania mexicana* ubiquitin activating enzyme. It was hypothesised that this E1 would respond differently from its human counterpart when faced with potent inhibitors of its ortholog. This was shown to be true, with the *Leishmania* enzyme considerably less affected than the human enzyme. Inhibitors that affected the *Leishmania* enzyme more than the human ortholog were also identified.

It was hypothesised that differences in response to inhibition would be due to differences in the ATP binding site. To visualise the ATP binding site, a 3.1 Å structure was obtained through protein crystallography. To validate the crystallographic structure, SAXS was conducted. The data from both methodologies concurred. Using the validated crystallographic structure, drug modelling was conducted. This predicted clashes between the *Leishmania* enzyme and human E1-specific drug, TAK-243.

Furthermore, a SUMOylation enzyme has been identified. A *Leishmania* SUMO conjugating enzyme, able to form polySUMO chains *in vitro*. This provides a glimpse of ubiquitin-like pathways in *Leishmania mexicana*.

To build on this, future work could use radiolabelled ATP to quantify the effect of various inhibitors, as well use cell viability assays such as alamar blue to determine the toxicity of inhibitors. However, the priority should be to utilise the crystal structure presented here, to aid in the rational design of drugs to inhibit LmxM.34.3060.

Chapter 6 Concluding Remarks

The experiments detailed here are focused on the components of the *Leishmania* ubiquitination system and characterise the potential of a key enzyme in this system as a novel therapeutic target. Genes encoding putative enzymes of the ubiquitination system were identified through BLAST searches and hidden Markov models. At the protein level, two ubiquitin activating enzymes were detected using activity-based probes. The reach of these activity-based probes was extended using a chemical crosslinker, identifying a ubiquitin conjugating enzyme. This interaction between the ubiquitin activating and ubiquitin conjugating enzyme was validated *in vivo* through interactome screening using BioID. BioID also provided a list of putative substrates of the ubiquitin conjugating enzyme, some of which were validated by anti-diglycine immunoprecipitations, which selected for the fingerprints of ubiquitinated peptides.

Having described components and substrates of the ubiquitination system, the most abundant ubiquitin activating enzyme was recombinantly expressed in a functional form (as was a SUMO conjugating enzyme). The structure of this ubiquitin activating enzyme was solved through protein crystallography and small angle X-ray scattering. This structure was used to explain the activities of small molecule inhibitors of the ubiquitin activating enzyme, inhibitors identified through a biochemical assay developed in this study. This biochemical assay allowed comparisons to be made between the inhibitors of the *Leishmania* enzyme and its human ortholog, and selective inhibitors of the parasite enzyme were identified.

6.1 A ubiquitination system exists in *Leishmania mexicana*

Leishmania can thrive in the disparate environments of a human macrophage and a sandfly midgut. To achieve this they must remodel themselves, from their molecular biology to their morphology (Coelho *et al.*, 2012; Sundar and Singh, 2018). It has been inferred that there is a ubiquitin-proteasome system that mediates protein degradation (Pérez-Pertejo *et al.*, 2011), and ubiquitin-like pathways have been identified (Gannavaram *et al.*, 2012). Recently, Burge *et al.*

constructed and screened a knock out library of genes thought to be involved in the ubiquitination pathway, identifying the same genes of interest reported herein, using an orthologous method (Burge *et al.*, 2020). Notably, cell lines missing an E1 could not establish infections in mice, if knock out cell lines could be generated at all.

This document uses BLASTs and hidden Markov models to identify genes of interest, then confirms that predicted E1s interact with an activity-based probe specific for E1s. It shows an E1-probe complex interacts with a predicted E2. This validates the bioinformatic work, demonstrates that both putative E1s are expressed at both major stages of the parasite's life cycle, and suggests which E1 is the most abundantly expressed. As such, it provides previously unknown information at the protein level of these potential drug targets.

Further work could use an activity-based probe that interacts with all components of the ubiquitination cascade, not just E1s. This would allow information on stage dependent expression, and the relative abundances of every catalytically active ubiquitination enzyme to be obtained. If more detailed information on the relative abundances of each E1 is necessary then a quantitative proteomic approach, such as TMT labelling, could be used to provide greater confidence than the emPAI score used here. Using the mass spectrometry data from the E1-E2 crosslinker experiment, the peptides crosslinked between each enzyme should identify their interaction sites, so it could be worth reanalysing this existing data with structural biology in mind. On a separate note, ubiquitin-like pathways exist. This document touches upon them, acknowledging their appearances on the fringes of the bioinformatic screening campaign, and putatively assigning pathways to the identified enzymes. These deserve further attention - particularly NEDD8. Given NEDD8's crosstalk with ubiquitination, its role in DNA repair (Brown *et al.*, 2015), and the existence of a mammalian NEDD8 E1 inhibitor in clinical trials for an array of cancers (Shah *et al.*, 2016), further experiments could characterise another potential drug target.

6.2 Biochemical characterisation of the ubiquitination system

Ubiquitination is an ordered system, responding to specific stimuli at a particular place and a precise time. As such it is important to appreciate the wider context of these enzymes - when they function, with which partners, and to what end? Geoghegan *et al.* unravel protein kinase signalling in *Leishmania*'s kinetochore using BioID (Geoghegan *et al.*, 2021). Kim *et al.* explain ubiquitin's role on a substrate through monitoring the ubiquitinome in the presence of a proteasome inhibitor (Kim *et al.*, 2011b). Tiengwe, Muratore and Bangs (2016) identify components of the endoplasmic reticulum associated degradation pathway in trypanosomes, and note conditions where this pathway is essential (Tiengwe, Muratore and Bangs, 2016).

Work in this thesis screens for the interactomes of the enzymes identified in Chapter 3. This is achieved by using CRISPR to fuse a promiscuous biotin ligase to either the N- or C-termini of an E1 and E2. This shows an *in vivo* interaction between an E1 and its cognate E2. Furthermore, this identifies putative substrates of the E2, and hypothesises that the E2 may belong to an endoplasmic degradation pathway. Ubiquitinated substrates were also identified through immunoprecipitation of peptides with a ubiquitin motif. As such, the wider context of ubiquitination is provided, which follows ubiquitin from an E1, to an E2, to a substrate.

Further work should prioritise repeating and validating the BioID screen. Repetition of the BioID screen could be conducted using the miniTurbo BirA*, its quicker reaction kinetics potentially generating more results (which is to say, a lower background coupled with more transient interactors). Alternatively, other proximity labelling systems, such as SPPLAT (Rees *et al.*, 2015) and APEX (Rhee *et al.*, 2013) exist, which provide orthologous methods to confirm the results. A screening method could even be combined with the immunoprecipitations shown here, first selecting for biotin at the protein level, then the ubiquitin motif at the peptide level, identifying the substrates of a specific E2. Screening methods aside, validating particular interactions between screening hits could be achieved using BRET (Xu, Piston and Johnson, 1999) or PLA (Söderberg *et al.*, 2006). However, a far more ambitious project, utilising the

immunoprecipitations detailed here, would be to identify the ubiquitination that occurs when a parasite transitions through its major life cycle stages. Entire organelles, such as the flagella, can be drastically remodelled, so it is likely that ubiquitination plays an important role as the parasite manoeuvres between insects and mammals.

6.3 Biophysical characterisation of a ubiquitin activating enzyme

E1s are already validated drug targets, with inhibitors existing for the NEDD8 and ubiquitin E1s (Yang *et al.*, 2007; Soucy *et al.*, 2009). These inhibitors are in either early or late stage clinical trials for a variety of cancers (Barghout and Schimmer, 2021). The structure-activity relationship of these inhibitors has been well-characterised, with the drug binding site identified as the ATP binding pocket of an E1 (Misra *et al.*, 2017). Given their characterised mechanisms and clinical success, other authors have been interested in the E1s of kinetoplastid parasites for the purpose of drug design (Bijlmakers, 2021). However, they have failed to generate a crystal structure, relying on homology based modelling to compare human and parasite E1s (Boer and Bijlmakers, 2019). The difficulty in obtaining crystal structures of full length E1s may be why the only published structures are from humans and two species of yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*).

This thesis reports the recombinant expression of a putative ubiquitin activating enzyme and a SUMO conjugating enzyme, both in functional forms, and develops an activity assay for each. Using the activity assay, drugs are trialled against ubiquitin activating enzyme, and a drug that is a potent, specific inhibitor of the *Leishmania* enzyme is identified. Furthermore, concurring structures of this enzyme are found through protein crystallography and SAXS. The crystal structure is used to explain the activity of one of these drugs.

Future work could consider optimising the crystallisation conditions, to achieve a higher resolution than the 3.1 Å presented here. If it is of interest, the structure of the E1 in complex with ubiquitin and an E2 could also be sought. If the structures of complexes are desired, a technique that melds mass spectrometric and structural approaches is HDX-MS, which would advise on the conformational

changes an E1 goes through while charging and transferring ubiquitin. On which note, only bacterial expression systems were trialled for E2 expression, and so there is plenty of room for alternative expression systems if a functional *Leishmania* ubiquitin conjugating enzyme is desired. However, more pressing than any of these suggestions would be structure-based drug design. Indeed, this application, and its potential impact in clinical settings, was the primary drive for research into the structure of this E1.

6.4 A novel drug target

The aim of this project was to characterise an understudied system in a neglected tropical pathogen, with the ultimate goal of contributing a novel drug target to the literature. Having identified two ubiquitin activating enzymes at the protein level, having described their interactome, solved the structure of an enzyme, and analysed its inhibitors, it is hoped that this study has had some success towards that lofty goal. Of course, further research is needed before any tangible benefits reach the millions of leishmaniasis patients worldwide. Yet, it is the sincerest hope of the author that this work may be cited by someone, who is cited by another someone, who develops any form of aid for a terrible disease that kills tens of thousands of the world's poorest people annually. If this work can feature as a footnote to research that alleviates human suffering then the author will be content.

List of References

[(2R,3S,4R,5R)-5-(6-Amino-9H-purin-9-YL)-3,4-dihydroxytetrahydro-2-furanyl]methyl sulfamate | C10H14N6O6S - PubChem (no date) *PubChem*. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/97576#section=2D-Structure (Accessed: 25 September 2021).

Acharya, K. R. and Lloyd, M. D. (2005) 'The advantages and limitations of protein crystal structures', *Trends in Pharmacological Sciences*, 26(1), pp. 10–14. doi: 10.1016/j.tips.2004.10.011.

Allison, M. (1993) 'Leishmaniasis', in Kiple, K. (ed.) *The Cambridge World History of Human Disease*. Cambridge: Cambridge University Press, pp. 832–834. doi: https://doi.org/10.1017/CHOL9780521332866.141.

Alsford, S. *et al.* (2011) 'High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome', *Genome Research*, 21(6), pp. 915–924. doi: 10.1101/gr.115089.110.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Alvar, J. *et al.* (2012) 'Leishmaniasis worldwide and global estimates of its incidence', *PLoS ONE*, 7(5). doi: 10.1371/journal.pone.0035671.

Armstrong, H. E. (1927) 'Poor common salt! [9]', *Nature*. Nature Publishing Group, p. 478. doi: 10.1038/120478b0.

Aruleba, R. T. *et al.* (2020) 'Can We Harness Immune Responses to Improve Drug Treatment in Leishmaniasis?', *Microorganisms*, 8(7), p. 1069. doi: 10.3390/MICROORGANISMS8071069.

Aslett, M. *et al.* (2009) 'TriTrypDB: A functional genomic resource for the Trypanosomatidae', *Nucleic Acids Research*, 38(SUPPL.1), pp. D457–D462. doi: 10.1093/nar/gkp851.

Awad, H., Khamis, M. M. and El-Aneed, A. (2015) 'Mass spectrometry, review of

the basics: Ionization', *Applied Spectroscopy Reviews*, 50(2), pp. 158–175. doi: 10.1080/05704928.2014.954046.

Ball, P. (2020) *Behind the screens of AlphaFold, Chemistry World*. Available at: https://www.chemistryworld.com/opinion/behind-the-screens-of-alphafold/4012867.article (Accessed: 8 March 2022).

Barghout, S. H. and Schimmer, A. D. (2021) 'E1 enzymes as therapeutic targets in cancers', *Pharmacological Reviews*, 73(1), pp. 1–56. doi: 10.1124/pharmrev.120.000053.

Barrangou, R. *et al.* (2007) 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes', *Science*, 315(5819), pp. 1709–1712. doi: 10.1126/SCIENCE.1138140.

Bayona, J. C. *et al.* (2011) 'SUMOylation pathway in Trypanosoma cruzi: Functional characterization and proteomic analysis of target proteins', *Molecular and Cellular Proteomics*, 10(12), p. M110.007369. doi: 10.1074/mcp.M110.007369.

Beneke, T. *et al.* (2017) 'A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids', *Royal Society Open Science*. doi: 10.1098/RSOS.170095.

Beneke, T. and Gluenz, E. (2019) 'LeishGEdit: A Method for Rapid Gene Knockout and Tagging Using CRISPR-Cas9', *Methods in Molecular Biology*, 1971, pp. 189–210. doi: 10.1007/978-1-4939-9210-2_9.

Benns, H. J. *et al.* (2021) 'Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery', *Current Opinion in Chemical Biology*, 60, pp. 20–29. doi: 10.1016/J.CBPA.2020.06.011.

Besteiro, S. *et al.* (2007) 'Protein turnover and differentiation in Leishmania', *International Journal for Parasitology*, 37(10), pp. 1063–1075. doi: 10.1016/j.ijpara.2007.03.008.

Bhandari, D. *et al.* (2011) 'Ubiquitination of mRNA cycling sequence binding

protein from Leishmania donovani (LdCSBP) modulates the RNA endonuclease activity of its Smr domain', *FEBS Letters*, 585(5), pp. 809–813. doi: 10.1016/j.febslet.2011.02.007.

Bibo-Verdugo, B. *et al.* (2017) 'Targeting proteasomes in infectious organisms to combat disease', *The FEBS Journal*, 284(10), pp. 1503–1517. doi: 10.1111/FEBS.14029.

Bijlmakers, M.-J. (2021) 'Ubiquitination and the Proteasome as Drug Targets in Trypanosomatid Diseases', *Frontiers in Chemistry*, 8, p. 1307. doi: 10.3389/fchem.2020.630888.

Blanchet, C. E. and Svergun, D. I. (2013) 'Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution', *Annual Review of Physical Chemistry*, 64(November 2012), pp. 37–54. doi: 10.1146/annurev-physchem-040412-110132.

Blasche, S. and Koegl, M. (2013) 'Analysis of protein-protein interactions using LUMIER assays', *Methods in molecular biology*, 1064, pp. 17–27. doi: 10.1007/978-1-62703-601-6_2.

Bloch, A. and Coutsogeorgopoulos, C. (1971) 'Inhibition of protein synthesis by 5'-sulfamoyladenosine', *Biochemistry*, 10(24), pp. 4394–4398. doi: 10.1021/BI00800A007.

Boban, M., Ljungdahl, P. O. and Foisner, R. (2015) 'Atypical Ubiquitylation in Yeast Targets Lysine-less Asi2 for Proteasomal Degradation', *The Journal of Biological Chemistry*, 290(4), p. 2489. doi: 10.1074/JBC.M114.600593.

Boer, D. R. and Bijlmakers, M. J. (2019) 'Differential Inhibition of Human and Trypanosome Ubiquitin E1S by TAK-243 Offers Possibilities for Parasite Selective Inhibitors', *Scientific Reports*, 9(1), pp. 1–14. doi: 10.1038/s41598-019-52618-3.

Bragg, W. L. (1934) *The Crystalline State: Volume I*. New York: The Macmillan Company.

Branon, T. C. *et al.* (2018) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–887. doi: 10.1038/nbt.4201.

Bremm, A. and Komander, D. (2011) 'Emerging roles for Lys11-linked polyubiquitin in cellular regulation', *Trends in Biochemical Sciences*, 36(7), pp. 355–363. doi: 10.1016/j.tibs.2011.04.004.

Brown, J. S. *et al.* (2015) 'Neddylation Promotes Ubiquitylation and Release of Ku from DNA-Damage Sites', *Cell Reports*, 11(5), pp. 704–714. doi: 10.1016/J.CELREP.2015.03.058.

Brownell, J. E. *et al.* (2010) 'Substrate-Assisted Inhibition of Ubiquitin-like Protein-Activating Enzymes: The NEDD8 E1 Inhibitor MLN4924 Forms a NEDD8-AMP Mimetic In Situ', *Molecular Cell*, 37(1), pp. 102–111. doi: 10.1016/j.molcel.2009.12.024.

Brückner, A. *et al.* (2009) 'Yeast Two-Hybrid, a Powerful Tool for Systems Biology', *International Journal of Molecular Sciences*, 10(6), p. 2763. doi: 10.3390/IJMS10062763.

Buetow, L. and Huang, D. T. (2016) 'Structural insights into the catalysis and regulation of E3 ubiquitin ligases', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 626–642. doi: 10.1038/nrm.2016.91.

Burge, R. J. *et al.* (2020) 'Leishmania differentiation requires ubiquitin conjugation mediated by a UBC2-UEV1 E2 complex', *PLoS Pathogens*, 16(10), p. e1008784. doi: 10.1371/journal.ppat.1008784.

Burgess, R. R. (2018) 'A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting', *Protein Expression and Purification*. Academic Press Inc., pp. 81–85. doi: 10.1016/j.pep.2018.05.007.

Burroughs, A. M. *et al.* (2007) 'Small but versatile: the extraordinary functional and structural diversity of the ß-grasp fold', *Biology Direct*, 2(1), p. 18. doi: 10.1186/1745-6150-2-18.

Cao, B. and Mao, X. (2011) ‘The ubiquitin-proteasomal system is critical for multiple myeloma: implications in drug discovery’, *American Journal of Blood Research*, 1(1), pp. 46–56. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/pmid/22432065/?tool=EBI (Accessed: 20 September 2021).

Capdeville, R. *et al.* (2002) ‘Glivec (ST1571, imatinib), a rationally developed, targeted anticancer drug’, *Nature Reviews Drug Discovery*. Nature Publishing Group, pp. 493–502. doi: 10.1038/nrd839.

Cavassin, F. B. *et al.* (2021) ‘Sixty years of Amphotericin B: An Overview of the Main Antifungal Agent Used to Treat Invasive Fungal Infections’, *Infectious Diseases and Therapy*, 10(1), pp. 115–147. doi: 10.1007/S40121-020-00382-7.

CDC (2020) *CDC - Leishmaniasis - Biology*.

Chen, J. J. *et al.* (2011) ‘Mechanistic studies of substrate-assisted inhibition of ubiquitin-activating enzyme by adenosine sulfamate analogues’, *Journal of Biological Chemistry*, 286(47), pp. 40867–40877. doi: 10.1074/jbc.M111.279984.

Chiu, Y. H., Sun, Q. and Chen, Z. J. (2007) ‘E1-L2 Activates Both Ubiquitin and FAT10’, *Molecular Cell*, 27(6), pp. 1014–1023. doi: 10.1016/j.molcel.2007.08.020.

Ciechanover, A., Finley, D. and Varshavsky, A. (1984) ‘Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85’, *Cell*, 37(1), pp. 57–66. doi: 10.1016/0092-8674(84)90300-3.

Claessen, J. H. L. *et al.* (2010) ‘The Transmembrane Segment of a Tail-anchored Protein Determines Its Degradative Fate through Dislocation from the Endoplasmic Reticulum’, *Journal of Biological Chemistry*, 285(27), pp. 20732–20739. doi: 10.1074/JBC.M110.120766.

Cockell, C. (2018) *The Equations of Life*. Atlantic Books.

Coelho, V. T. S. *et al.* (2012) ‘Identification of Proteins in Promastigote and

Amastigote-like Leishmania Using an Immunoproteomic Approach', *PLoS Neglected Tropical Diseases*, 6(1). doi: 10.1371/JOURNAL.PNTD.0001430.

Cohen, J. and Desai, N. (2019) 'With its CRISPR revolution, China becomes a world leader in genome editing', *Science*. doi: 10.1126/SCIENCE.AAY9689.

Cossins, D. (2018) *We thought the Incas couldn't write. These knots change everything*, *New Scientist*. Available at: https://www.newscientist.com/article/mg23931972-600-we-thought-the-incas-couldnt-write-these-knots-change-everything/ (Accessed: 3 September 2021).

Dauter, Z. (2017) 'Collection of X-ray diffraction data from macromolecular crystals', in *Methods in Molecular Biology*. Humana Press Inc., pp. 165–184. doi: 10.1007/978-1-4939-7000-1_7.

Davies, P. (2019) *The Demon in the Machine*. Penguin Random House UK.

Dean, S., Sunter, J. D. and Wheeler, R. J. (2017) 'TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource', *Trends in Parasitology*, 33(2), pp. 80–82. doi: 10.1016/J.PT.2016.10.009.

Donovick, R. *et al.* (1955) 'Amphotericins A and B, antifungal antibiotics produced by a streptomycete. I. In vitro studies.', *Antibiotics Annual*, 3, pp. 579–586. Available at: https://europepmc.org/article/med/13355330 (Accessed: 21 September 2021).

Doolittle, R. F. *et al.* (1983) 'Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor', *Science*, 221(4607), pp. 275–277. doi: 10.1126/science.6304883.

Dubendorf, J. W. and Studier, F. W. (1991) 'Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor', *Journal of Molecular Biology*, 219(1), pp. 45–59. doi: 10.1016/0022-2836(91)90856-2.

Duncan, R. *et al.* (2011) 'Identification and Characterization of Genes Involved in

Leishmania Pathogenesis: The Potential for Drug Target Selection', *Molecular Biology International*, 2011, pp. 1–10. doi: 10.4061/2011/428486.

Eckert, M. (2012) 'Max von Laue and the discovery of X-ray diffraction in 1912', *Ann. Phys. (Berlin)*, 524(5), pp. 83–85. doi: 10.1002/andp.201200724.

Eddins, M. J. *et al.* (2007) 'Crystal Structure and Solution NMR Studies of Lys48-linked Tetraubiquitin at Neutral pH', *Journal of Molecular Biology*, 367(1), pp. 204–211. doi: 10.1016/j.jmb.2006.12.065.

Editorial (2004) 'A prize for protein degradation', *Nature Cell Biology*, 6(11), pp. 1011–1011. doi: 10.1038/ncb1104-1011.

El-Aneed, A., Cohen, A. and Banoub, J. (2009) 'Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers', *Applied Spectroscopy Reviews*, 44(3), pp. 210–230. doi: 10.1080/05704920902717872.

Eletr, Z. M. *et al.* (2005) 'E2 conjugating enzymes must disengage from their E1 enzymes before E3-dependent ubiquitin and ubiquitin-like transfer', *Nature Structural & Molecular Biology*, 12(10), pp. 933–934. doi: 10.1038/nsmb984.

Espinosa, O. A. *et al.* (2016) 'An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as Leishmania and Endotrypanum', *Parasitology*, pp. 1–13. doi: 10.1017/S0031182016002092.

Feng, L. and Chen, J. (2012) 'The E3 ligase RNF8 regulates KU80 removal and NHEJ repair', *Nature Structural & Molecular Biology*, 19(2), pp. 201–206. doi: 10.1038/nsmb.2211.

Fiebig, M., Kelly, S. and Gluenz, E. (2015) 'Comparative Life Cycle Transcriptomics Revises Leishmania mexicana Genome Annotation and Links a Chromosome Duplication with Parasitism of Vertebrates', *PLOS Pathogens*, 11(10), p. e1005186. doi: 10.1371/JOURNAL.PPAT.1005186.

Florini, J. R., Bird, H. H. and Bell, P. H. (1966) 'Inhibition of Protein Synthesis in Vitro and in Vivo by Nucleocidin, an Antitrypanosomal Antibiotic', *Journal of*

Biological Chemistry, 241(5), pp. 1091–1098. doi: 10.1016/S0021-9258(18)96807-9.

Franke, D. *et al.* (2017) 'ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions', *Journal of Applied Crystallography*, 50(4), pp. 1212–1225. doi: 10.1107/S1600576717007786.

Frézard, F., Demicheli, C. and Ribeiro, R. R. (2009) 'Pentavalent Antimonials: New Perspectives for Old Drugs', *Molecules*, 14(7), pp. 2317–2336. doi: 10.3390/MOLECULES14072317.

Gannavaram, S. *et al.* (2011) 'Mitochondrial Associated Ubiquitin Fold Modifier-1 Mediated Protein Conjugation in Leishmania donovani', *PLOS ONE*, 6(1), p. e16156. doi: 10.1371/JOURNAL.PONE.0016156.

Gannavaram, S. *et al.* (2012) 'Deletion of mitochondrial associated ubiquitin fold modifier protein Ufm1 in Leishmania donovani results in loss of ß-oxidation of fatty acids and blocks cell division in the amastigote stage', *Molecular Microbiology*, 86(1), pp. 187–198. doi: 10.1111/j.1365-2958.2012.08183.x.

Gannavaram, S. *et al.* (2014) 'Deletion of Ubiquitin Fold Modifier Protein Ufm1 Processing Peptidase Ufsp in L. donovani Abolishes Ufm1 Processing and Alters Pathogenesis', *PLoS Neglected Tropical Diseases*, 8(2). doi: 10.1371/journal.pntd.0002707.

Gasteiger, E. *et al.* (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in *The Proteomics Protocols Handbook*. Humana Press, pp. 571–607. doi: 10.1385/1-59259-890-0:571.

Geoghegan, V. *et al.* (2021) 'Protein kinase signalling at the Leishmania kinetochore captured by XL-BioID', *bioRxiv*, p. 2021.07.08.451598. doi: 10.1101/2021.07.08.451598.

Ghorbal, M. *et al.* (2014) 'Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system', *Nature Biotechnology*, 32(8), pp. 819–821. doi: 10.1038/nbt.2925.

Grau-Bové, X., Sebé-Pedrós, A. and Ruiz-Trillo, I. (2015) 'The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin', *Molecular Biology and Evolution*, 32(3), pp. 726–739. doi: 10.1093/molbev/msu334.

Groettrup, M. *et al.* (2008) 'Activating the ubiquitin family: UBA6 challenges the field', *Trends in Biochemical Sciences*, 33(5), pp. 230–237. doi: 10.1016/j.tibs.2008.01.005.

Guha, K. *et al.* (2014) 'Ubiquitination-mediated interaction among domains is responsible for inhibition of RNA endonuclease activity of mRNA cycling sequence binding protein from L. donovani (LdCSBP)', *Parasitology Research*, 113(8), pp. 2941–2949. doi: 10.1007/s00436-014-3956-z.

Gupta, I. *et al.* (2018) 'Ubiquitin Proteasome pathway proteins as potential drug targets in parasite Trypanosoma cruzi', *Scientific Reports*, 8(1), pp. 1–12. doi: 10.1038/s41598-018-26532-z.

Harari, Y. N. (2014) *Sapiens: A Brief History of Humankind*.

Harbut, M. B. *et al.* (2012) 'Targeting the ERAD pathway via inhibition of signal peptide peptidase for antiparasitic therapeutic design', *Proceedings of the National Academy of Sciences*, 109(52), pp. 21486–21491. doi: 10.1073/PNAS.1216016110.

Hashimoto, S., Anai, H. and Hanada, K. (2016) 'Mechanisms of interstrand DNA crosslink repair and human disorders', *Genes and Environment*, 38(1), pp. 1–8. doi: 10.1186/S41021-016-0037-9.

Hatfield, P. M. *et al.* (1997) 'The ubiquitin-activating enzyme (E1) gene family in Arabidopsis thaliana', *Plant Journal*, 11(2), pp. 213–226. doi: 10.1046/j.1365-313X.1997.11020213.x.

Hendrickx, S. *et al.* (2020) 'Transmission potential of paromomycin-resistant Leishmania infantum and Leishmania donovani', *Journal of Antimicrobial Chemotherapy*, 75(4), pp. 951–957. doi: 10.1093/JAC/DKZ517.

Hjerpe, R. *et al.* (2012) 'Changes in the ratio of free NEDD8 to ubiquitin triggers NEDDylation by ubiquitin enzymes', *Biochemical Journal*, 441(3), pp. 927–939. doi: 10.1042/BJ20111671.

Hoare, C. A. (1938) 'Early discoveries regarding the parasite of oriental sore', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 32(1), pp. IN12-92. doi: 10.1016/S0035-9203(38)90097-5.

Hochstrasser, M. (2000) 'Evolution and function of ubiquitin-like protein-conjugation systems', *Nature Cell Biology*. doi: 10.1038/35019643.

Hochstrasser, M. (2009) 'Origin and function of ubiquitin-like proteins', *Nature*, 458(7237), pp. 422–429. doi: 10.1038/nature07958.

Hoege, C. *et al.* (2002) 'RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO', *Nature*, 419(6903), pp. 135–141. doi: 10.1038/nature00991.

Holding, A. (2015) 'XL-MS: Protein cross-linking coupled with mass spectrometry', *Methods (San Diego, Calif.)*, 89, pp. 54–63. doi: 10.1016/J.YMETH.2015.06.010.

https://clinicaltrials.gov/ (2021) *Pevonedistat Plus Azacitidine Versus Single-Agent Azacitidine as First-Line Treatment for Participants With Higher-Risk Myelodysplastic Syndromes (HR MDS), Chronic Myelomonocytic Leukemia (CMML), or Low-Blast Acute Myelogenous Leukemia (AML) - Full Text V*. Available at: https://clinicaltrials.gov/ct2/show/NCT03268954?intr=MLN4924&phase=2&draw=2&rank=2 (Accessed: 26 March 2021).

Huang, D. T. *et al.* (2004) 'A unique E1-E2 interaction required for optimal conjugation of the ubiquitin-like protein NEDD8', *Nature Structural and Molecular Biology*, 11(10), pp. 927–935. doi: 10.1038/nsmb826.

Huang, D. T. *et al.* (2007) 'Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity', *Nature*, 445(7126). doi: 10.1038/nature05490.

Huang, X. and Dixit, V. M. (2016) ‘Drugging the undruggables: Exploring the ubiquitin system for drug development’, *Cell Research*. Nature Publishing Group, pp. 484–498. doi: 10.1038/cr.2016.31.

Hyer, M. L. *et al.* (2018) ‘A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment’, *Nature Medicine*, 24(2), pp. 186–193. doi: 10.1038/nm.4474.

Ishihama, Y. *et al.* (2005) ‘Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein’, *Molecular and Cellular Proteomics*, 4(9), pp. 1265–1272. doi: 10.1074/mcp.M500061-MCP200.

Ishino, Y. *et al.* (1987) ‘Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product’, *Journal of Bacteriology*, 169(12), pp. 5429–5433. doi: 10.1128/JB.169.12.5429-5433.1987.

Itakura, K. *et al.* (1977) ‘Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin’, *Science*, 198(4321), pp. 1056–1063. doi: 10.1126/science.412251.

Jacquemont, C. and Taniguchi, T. (2007) ‘The Fanconi anemia pathway and ubiquitin’, *BMC Biochemistry 2007 8:1*, 8(1), pp. 1–10. doi: 10.1186/1471-2091-8-S1-S10.

Jansen, R. *et al.* (2002) ‘Identification of genes that are associated with DNA repeats in prokaryotes’, *Molecular microbiology*, 43(6), pp. 1565–1575. doi: 10.1046/J.1365-2958.2002.02839.X.

Jiang, F. and Doudna, J. A. (2017) ‘CRISPR–Cas9 Structures and Mechanisms’, *The Annual Review of Biophysics*, 46, pp. 505–529. doi: 10.1146/ANNUREV-BIOPHYS-062215-010822.

Jin, J. *et al.* (2007) ‘Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging’, *Nature*, 447(7148), pp. 1135–1138. doi:

10.1038/nature05902.

Jin, L., Wang, W. and Fang, G. (2014) ‘Targeting protein-protein interaction by small molecules’, *Annual Review of Pharmacology and Toxicology*. Annual Reviews Inc., pp. 435–456. doi: 10.1146/annurev-pharmtox-011613-140028.

Jinek, M. *et al.* (2012) ‘A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity’, *Science*, 337(6096), pp. 816–821. doi: 10.1126/SCIENCE.1225829.

Johnson, L. S., Eddy, S. R. and Portugaly, E. (2010) ‘Hidden Markov model speed heuristic and iterative HMM search procedure’, *BMC Bioinformatics*, 11(1), pp. 1–8. doi: 10.1186/1471-2105-11-431.

Jones, D. *et al.* (2001) ‘Functional and phylogenetic analysis of the ubiquitylation system in Caenorhabditis elegans: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes, and ubiquitin-like proteins’, *Genome Biology*, 3(1), p. research0002.1. doi: 10.1186/gb-2001-3-1-research0002.

Jones, D. *et al.* (2002) ‘Functional and phylogenetic analysis of the ubiquitylation system in Caenorhabditis elegans: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes, and ubiquitin-like proteins.’, *Genome biology*, 3(1), p. RESEARCH0002. doi: 10.1186/gb-2001-3-1-research0002.

Jumper, J. *et al.* (2021) ‘Highly accurate protein structure prediction with AlphaFold’, *Nature*, 596(7873), pp. 583–589. doi: 10.1038/S41586-021-03819-2.

Karpievitch, Y. V. *et al.* (2010) ‘Liquid chromatography mass spectrometry-based proteomics: Biological and technological aspects’, *Annals of Applied Statistics*, 4(4), pp. 1797–1823. doi: 10.1214/10-AOAS341.

Karpiyevich, M. and Artavanis-Tsakonas, K. (2020) ‘Ubiquitin-Like Modifiers: Emerging Regulators of Protozoan Parasites’, *Biomolecules*, 10(10), pp. 1–20. doi: 10.3390/BIOM10101403.

Kikhney, A. G. and Svergun, D. I. (2015) ‘A practical guide to small angle X-ray

scattering (SAXS) of flexible and intrinsically disordered proteins', *FEBS Letters*, 589(19), pp. 2570–2577. doi: 10.1016/j.febslet.2015.08.027.

Kim, D. I. *et al.* (2014) 'Probing nuclear pore complex architecture with proximity-dependent biotinylation', *Proceedings of the National Academy of Sciences*, 111(24), pp. E2453–E2461. doi: 10.1073/PNAS.1406459111.

Kim, M. S., Zhong, J. and Pandey, A. (2016) 'Common errors in mass spectrometry-based analysis of post-translational modifications', *Proteomics*, 16(5), pp. 700–714. doi: 10.1002/PMIC.201500355.

Kim, W. *et al.* (2011a) 'Systematic and quantitative assessment of the ubiquitin-modified proteome', *Molecular Cell*, 44(2), pp. 325–340. doi: 10.1016/j.molcel.2011.08.025.

Kim, W. *et al.* (2011b) 'Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome', *Molecular Cell*, 44(2), pp. 325–340. doi: 10.1016/J.MOLCEL.2011.08.025.

Klein, C. A., Droll, D. and Clayton, C. (2013) 'SUMOylation in Trypanosoma brucei', *PeerJ*, 1(1). doi: 10.7717/PEERJ.180.

Knipscheer, P. *et al.* (2007) 'Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation', *The EMBO Journal*, 26(11), pp. 2797–2807. doi: 10.1038/sj.emboj.7601711.

Komander, D. and Rape, M. (2012) 'The Ubiquitin Code', *Annual Review of Biochemistry*, 81(1), pp. 203–229. doi: 10.1146/annurev-biochem-060310-170328.

Kondo, N. *et al.* (2010) 'DNA damage induced by alkylating agents and repair pathways', *Journal of Nucleic Acids*, 2010. doi: 10.4061/2010/543531.

Kottemann, M. C. and Smogorzewska, A. (2013) 'Fanconi anemia and the repair of Watson and Crick crosslinks', *Nature*, 493(7432), p. 356. doi: 10.1038/NATURE11863.

Ladisch, M. R. and Kohlmann, K. L. (1992) ‘Recombinant human insulin’, *Biotechnology Progress*, 8(6), pp. 469–478. doi: 10.1021/bp00018a001.

Lainson, R. and Strangways-Dixon, J. (1964) ‘The epidemiology of dermal leishmaniasis in British Honduras: Part II. Reservoir-hosts of Leishmania mexicana among the forest rodents’, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 58(2), pp. 136–153. doi: 10.1016/0035-9203(64)90003-3.

Lake, M. W. *et al.* (2001) ‘Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex.’, *Nature*, 414(6861), pp. 325–9. doi: 10.1038/35104586.

Lam, S. Y. *et al.* (2014) ‘The human ubiquitin conjugating enzyme UBE2J2 (Ubc6) is a substrate for proteasomal degradation’, *Biochemical and Biophysical Research Communications*, 451(3), pp. 361–366. doi: 10.1016/J.BBRC.2014.07.099.

Lander, E. S. (2016) ‘The Heroes of CRISPR’, *Cell*, 164(1–2), pp. 18–28. doi: 10.1016/J.CELL.2015.12.041.

Lazebnik, Y. (2002) ‘Can a biologist fix a radio? - Or, what I learned while studying apoptosis’, *Cancer Cell*. Cell Press, pp. 179–182. doi: 10.1016/S1535-6108(02)00133-2.

Ledford, H. and Callaway, E. (2020) ‘Pioneers of revolutionary CRISPR gene editing win chemistry Nobel’, *Nature*, 586(7829), pp. 346–347. doi: 10.1038/D41586-020-02765-9.

Leifso, K. *et al.* (2007) ‘Genomic and proteomic expression analysis of Leishmania promastigote and amastigote life stages: The Leishmania genome is constitutively expressed’, *Molecular and Biochemical Parasitology*, 152(1), pp. 35–46. doi: 10.1016/j.molbiopara.2006.11.009.

Leishman, W. B. (1903) ‘On the possibility of the occurrence of trypanosomiasis in India.’, *The British Medical Journal*, 1, pp. 1252–4. doi: 10.1136/bmj.1.2213.1252.

Lenk, U. *et al.* (2002) 'A role for mammalian Ubc6 homologues in ER-associated protein degradation', *Journal of Cell Science*, 115(14), pp. 3007–3014. doi: 10.1242/JCS.115.14.3007.

Letunic, I., Khedkar, S. and Bork, P. (2021) 'SMART: Recent updates, new developments and status in 2020', *Nucleic Acids Research*, 49(D1), pp. D458–D460. doi: 10.1093/nar/gkaa937.

Li, L., Stoeckert, C. J. and Roos, D. S. (2003) 'OrthoMCL: Identification of ortholog groups for eukaryotic genomes', *Genome Research*, 13(9), pp. 2178–2189. doi: 10.1101/gr.1224503.

Li, W. *et al.* (1993) 'Effects of histone acetylation, ubiquitination and variants on nucleosome stability', *Biochemical Journal*, 296(3), pp. 737–744. doi: 10.1042/BJ2960737.

Li, W. *et al.* (2008) 'Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling', *PloS one*, 3(1). doi: 10.1371/JOURNAL.PONE.0001487.

Liao, S. *et al.* (2010) 'The small ubiquitin-like modifier (SUMO) is essential in cell cycle regulation in Trypanosoma brucei', *Experimental Cell Research*, 316(5), pp. 704–715. doi: 10.1016/j.yexcr.2009.12.017.

Longerich, S. *et al.* (2014) 'Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA', *Nucleic Acids Research*, 42(9), p. 5657. doi: 10.1093/NAR/GKU198.

Lorenz, S. *et al.* (2013) 'Macromolecular juggling by ubiquitylation enzymes', *BMC Biology*, 11. doi: 10.1186/1741-7007-11-65.

Lu, X. *et al.* (2010) 'Designed semisynthetic protein inhibitors of Ub/Ubl E1 activating enzymes', *Journal of the American Chemical Society*, 132(6), pp. 1748–1749. doi: 10.1021/ja9088549.

Lux, M. C., Standke, L. C. and Tan, D. S. (2019) 'Targeting adenylate-forming enzymes with designed sulfonyladenosine inhibitors', *The Journal of Antibiotics*, 72(6), pp. 325–349. doi: 10.1038/s41429-019-0171-2.

Lv, Z. *et al.* (2017) 'Domain alternation and active site remodeling are conserved structural features of ubiquitin E1', *Journal of Biological Chemistry*, 292(29), pp. 12089–12099. doi: 10.1074/jbc.M117.787622.

Lv, Z. *et al.* (2018) 'Crystal structure of a human ubiquitin E1– ubiquitin complex reveals conserved functional elements essential for activity', *Journal of Biological Chemistry*, 293(47), pp. 18337–18352. doi: 10.1074/jbc.RA118.003975.

Mann, S. *et al.* (2021) 'A Review of Leishmaniasis: Current Knowledge and Future Directions', *Current Tropical Medicine Reports*, 8(2), pp. 121–132. doi: 10.1007/S40475-021-00232-7.

Matallana-Surget, S., Leroy, B. and Wattiez, R. (2010) 'Shotgun proteomics: Concept, key points and data mining', *Expert Review of Proteomics*, 7(1), pp. 5–7. doi: 10.1586/epr.09.101.

Matos, A. P. S. *et al.* (2020) 'A review of current treatments strategies based on paromomycin for leishmaniasis', *Journal of Drug Delivery Science and Technology*, 57, p. 101664. doi: 10.1016/J.JDDST.2020.101664.

McClellan, A. J., Laugesen, S. H. and Ellgaard, L. (2019) 'Cellular functions and molecular mechanisms of non-lysine ubiquitination', *Open Biology*, 9(9), p. 190147. doi: 10.1098/RSOB.190147.

McGrath, J. P., Jentsch, S. and Varshavsky, A. (1991) 'UBA1: An essential yeast gene encoding ubiquitin-activating enzyme', *EMBO Journal*, 10(1), pp. 227–236. doi: 10.1002/j.1460-2075.1991.tb07940.x.

McHugh, P. J., Ward, T. A. and Chovanec, M. (2012) 'A prototypical Fanconi anemia pathway in lower eukaryotes?', *Cell Cycle*, 11(20), p. 3739. doi: 10.4161/CC.21727.

McPherson, A. (2004) ‘Introduction to protein crystallization’, *Methods*, 34(3), pp. 254–265. doi: 10.1016/j.ymeth.2004.03.019.

McPherson, A. and Gavira, J. A. (2014) ‘Introduction to protein crystallization’, *Acta Crystallographica Section F:Structural Biology Communications*, 70(1), pp. 2–20. doi: 10.1107/S2053230X13033141.

De Menezes, J. P. B. *et al.* (2015) ‘Advances in development of new treatment for leishmaniasis’, *BioMed Research International*, 2015. doi: 10.1155/2015/815023.

Misra, M. *et al.* (2017) ‘Dissecting the Specificity of Adenosyl Sulfamate Inhibitors Targeting the Ubiquitin-Activating Enzyme’, *Structure*, 25(7), pp. 1120-1129.e3. doi: 10.1016/j.str.2017.05.001.

Mojica, F. J. M., Juez, G. and Rodriguez-Valera, F. (1993) ‘Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites’, *Molecular Microbiology*, 9(3), pp. 613–621. doi: 10.1111/J.1365-2958.1993.TB01721.X.

Momen, H. and Cupolillo, E. (2000) ‘Speculations on the origin and evolution of the genus Leishmania’, *Memorias do Instituto Oswaldo Cruz*, 95(4), pp. 583–588. doi: 10.1590/S0074-02762000000400023.

Mulder, M. P. C. *et al.* (2016) ‘A cascading activity-based probe sequentially targets E1–E2–E3 ubiquitin enzymes’, *Nature Chemical Biology*, 12(7), pp. 523–530. doi: 10.1038/nchembio.2084.

Muñoz, C. *et al.* (2015) ‘Role of the ubiquitin-proteasome systems in the biology and virulence of protozoan parasites’, *BioMed Research International*, 2015. doi: 10.1155/2015/141526.

Novo, S. P. C. *et al.* (2016) ‘The process of Leishmania infection - Disease and new perspectives of paleoparasitology’, *Revista do Instituto de Medicina Tropical de Sao Paulo*, 58(45). doi: 10.1590/S1678-9946201658045.

Nucleocidin | C10H13FN6O6S - PubChem (no date) *PubChem*. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/72299 (Accessed: 25 September 2021).

Okwor, I. and Uzonna, J. (2016) 'Social and economic burden of human leishmaniasis', *American Journal of Tropical Medicine and Hygiene*, 94(3), pp. 489–493. doi: 10.4269/ajtmh.15-0408.

Olsen, S. K. *et al.* (2010) 'Active site remodelling accompanies thioester bond formation in the SUMO E1', *Nature*, 463(7283), pp. 906–912. doi: 10.1038/nature08765.

Olsen, S. K. and Lima, C. D. (2013) 'Structure of a Ubiquitin E1-E2 Complex: Insights to E1-E2 Thioester Transfer', *Molecular Cell*, 49(5), pp. 884–896. doi: 10.1016/j.molcel.2013.01.013.

Paugam, A. *et al.* (2003) 'Characterization and role of protozoan parasite proteasomes', *Trends in Parasitology*, 19(2), pp. 55–59. doi: 10.1016/S1471-4922(02)00064-8.

Paul, W. and Steinwedel, H. (1953) 'Ein neues Massenspektrometer ohne Magnetfeld', *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences*, 8(7), pp. 448–450. doi: 10.1515/zna-1953-0710.

Pelzer, C. *et al.* (2007) 'UBE1L2, a novel E1 enzyme specific for ubiquitin', *Journal of Biological Chemistry*, 282(32). doi: 10.1074/jbc.C700111200.

Pérez-Pertejo, Y. *et al.* (2011) 'Leishmania donovani: proteasome-mediated down-regulation of methionine adenosyltransferase.', *Parasitology*, 138(9), pp. 1082–92. doi: 10.1017/S0031182011000862.

Pertsemlidis, A. and Fondon, J. W. (2001) 'Having a BLAST with bioinformatics (and avoiding BLAST phemy)', *Genome Biology*. BioMed Central, pp. 1–10. doi: 10.1186/gb-2001-2-10-reviews2002.

Pichler, A. *et al.* (2017) 'SUMO conjugation - A mechanistic view', *Biomolecular*

Concepts, 8(1), pp. 13–36. doi: 10.1515/bmc-2016-0030.

Pickart, C. M. (2001) 'Mechanisms Underlying Ubiquitination', *Annual Review of Biochemistry*, 70(1), pp. 503–533. doi: 10.1146/annurev.biochem.70.1.503.

Poinar, G. and Poinar, R. (2004) 'Paleoleishmania proterus n. gen., n. sp., (Trypanosomatidae: Kinetoplastida) from Cretaceous Burmese Amber', *Protist*, 155(September), pp. 305–310. doi: S1434-4610(05)70187-5 [pii]\r10.1078/1434461041844259.

Ponder, E. L. and Bogyo, M. (2007) 'Ubiquitin-like modifiers and their deconjugating enzymes in medically important parasitic protozoa', *Eukaryotic Cell*, 6(11), pp. 1943–1952. doi: 10.1128/EC.00282-07.

Popovic, D., Vucic, D. and Dikic, I. (2014) 'Ubiquitination in disease pathogenesis and treatment.', *Nature medicine*, 20(11), pp. 1242–53. doi: 10.1038/nm.3739.

Putnam, D. K., Lowe, E. W. and Meiler, J. (2013) 'Reconstruction of SAXS profiles from protein structures', *Computational and Structural Biotechnology Journal*, 8(11), p. e201308006. doi: 10.5936/csbj.201308006.

Ramakrishnan, V. (2018) *Gene Machine*. First edit. Basic Books.

Raman, C. V. and Krishnamurti, P. (1929) 'A new X-ray effect', *Nature*, 124(3115), pp. 53–54. doi: 10.1038/124053b0.

Ravy, S. (2015) 'André Guinier (1911-2000): A physicist among crystallographers', *Physica Scripta*, 90(3), p. 38001. doi: 10.1088/0031-8949/90/3/038001.

Ree, R., Varland, S. and Arnesen, T. (2018) 'Spotlight on protein N-terminal acetylation', *Experimental and Molecular Medicine*. Nature Publishing Group, p. 90. doi: 10.1038/s12276-018-0116-z.

Rees, J. S. *et al.* (2015) 'Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters', *Current Protocols in Protein Science*, 80(1), pp. 19.27.1-19.27.18. doi: 10.1002/0471140864.PS1927S80.

Reithinger, R. *et al.* (2007) 'Cutaneous leishmaniasis', *The Lancet Infectious Diseases*, 7(9), pp. 581–596. doi: 10.1016/S1473-3099(07)70209-8.

Rhee, H.-W. *et al.* (2013) 'Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging', *Science*, 339(6125), pp. 1328–1331. doi: 10.1126/SCIENCE.1230593.

Rijal, S. *et al.* (2006) 'The economic burden of visceral leishmaniasis for households in Nepal', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(9), pp. 838–841. doi: 10.1016/j.trstmh.2005.09.017.

Rogerdodd (no date) *Ubiquitin*, *Wikipedia*. Available at: https://commons.wikimedia.org/w/index.php?curid=7677277.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia coli: Advances and challenges', *Frontiers in Microbiology*. Frontiers Research Foundation, p. 172. doi: 10.3389/fmicb.2014.00172.

Rose, C. M. *et al.* (2016) 'Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes', *Cell Systems*, 3(4), pp. 395–403. doi: 10.1016/j.cels.2016.08.009.

Rosenzweig, D. *et al.* (2008) 'Post-translational modification of cellular proteins during Leishmania donovani differentiation', *Proteomics*, 8(9), pp. 1843–1850. doi: 10.1002/pmic.200701043.

Rougeron, V., De Meeûs, T. and Bañuls, A. L. (2017) 'Reproduction in Leishmania: A focus on genetic exchange', *Infection, Genetics and Evolution*, 50, pp. 128–132. doi: 10.1016/J.MEEGID.2016.10.013.

Roux, K. J. *et al.* (2012) 'A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells', *Journal of Cell Biology*, 196(6), pp. 801–810. doi: 10.1083/JCB.201112098.

Russo Krauss, I. *et al.* (2013) 'An Overview of Biological Macromolecule Crystallization', *International Journal of Molecular Sciences*, 14(6), pp. 11643–

11691. doi: 10.3390/ijms140611643.

Sanchiz, Á. *et al.* (2020) 'The experimental proteome of Leishmania infantum promastigote and its usefulness for improving gene annotations', *Genes*, 11(9), pp. 1–20. doi: 10.3390/genes11091036.

Sansbury, B. M., Hewes, A. M. and Kmiec, E. B. (2019) 'Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair', *Communications Biology*, 2(458). doi: 10.1038/s42003-019-0705-y.

Santos, R. E. R. S. *et al.* (2017) 'A DiCre recombinase-based system for inducible expression in Leishmania major', *Molecular and Biochemical Parasitology*, 216, pp. 45–48. doi: 10.1016/J.MOLBIOPARA.2017.06.006.

Sarnoff, R. *et al.* (2010) 'The economic impact of visceral leishmaniasis on rural households in one endemic district of Bihar, India', *Tropical Medicine and International Health*, 15(SUPPL. 2), pp. 42–49. doi: 10.1111/j.1365-3156.2010.02516.x.

Sasidharan, S. and Saudagar, P. (2021) 'Leishmaniasis: where are we and where are we heading?', *Parasitology Research*, 120(5), pp. 1541–1554. doi: 10.1007/S00436-021-07139-2.

Sato, Y. *et al.* (2009) 'Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by NZF domains of TAB2 and TAB3.', *The EMBO journal*, 28(24), pp. 3903–9. doi: 10.1038/emboj.2009.345.

Schäfer, A., Kuhn, M. and Schindelin, H. (2014) 'Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules', *Acta Crystallographica Section D: Biological Crystallography*, 70(5), pp. 1311–1320. doi: 10.1107/S1399004714002910.

Schelpe, J. *et al.* (2016) 'Structure of UBE2Z Enzyme Provides Functional Insight into Specificity in the FAT10 Protein Conjugation Machinery', *The Journal of Biological Chemistry*, 291(2), p. 630. doi: 10.1074/JBC.M115.671545.

Schulman, B. A. and Wade Harper, J. (2009) 'Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways', *Nature Reviews Molecular Cell Biology*, 10(5), pp. 319–331. doi: 10.1038/nrm2673.

Scott, J. D. and Pawson, T. (2009) 'Cell Signaling in Space and Time: Where Proteins Come Together and When They're Apart', *Science*, 326(5957), p. 1220. doi: 10.1126/SCIENCE.1175668.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. (2007) 'Escherichia coli physiology in Luria-Bertani broth', *Journal of Bacteriology*, 189(23), pp. 8746–8749. doi: 10.1128/JB.01368-07.

Shah, J. J. *et al.* (2016) 'Phase I Study of the Novel Investigational NEDD8-Activating Enzyme Inhibitor Pevonedistat (MLN4924) in Patients with Relapsed/Refractory Multiple Myeloma or Lymphoma', *Clinical Cancer Research*, 22(1), pp. 34–43. doi: 10.1158/1078-0432.CCR-15-1237.

Sharp, P. M. and Li, W. H. (1987) 'Molecular evolution of ubiquitin genes', *Trends in Ecology & Evolution*, 2(11), pp. 328–332. doi: 10.1016/0169-5347(87)90108-X.

Sheng, Y. *et al.* (2012) 'A Human Ubiquitin Conjugating Enzyme (E2)-HECT E3 Ligase Structure-function Screen', *Molecular & Cellular Proteomics*, 11(8), pp. 329–341. doi: 10.1074/MCP.O111.013706.

Shuman, D. A., Robins, R. K. and Robins, M. J. (1969) 'The synthesis of adenine 5'-O-sulfamoyl nucleosides related to nucleocidin', *Journal of the American Chemical Society*, 91(12), pp. 3391–3392. doi: 10.1021/JA01040A062.

Sidik, S. M. *et al.* (2016) 'A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes', *Cell*, 166(6), pp. 1423-1435.e12. doi: 10.1016/J.CELL.2016.08.019.

Sims, J. J. and Cohen, R. E. (2009) 'Linkage-Specific Avidity Defines the Lysine 63-Linked Polyubiquitin-Binding Preference of Rap80', *Molecular Cell*, 33(6), pp. 775–783. doi: 10.1016/j.molcel.2009.02.011.

Singh, N., Kumar, M. and Singh, R. K. (2012) 'Leishmaniasis: Current status of available drugs and new potential drug targets', *Asian Pacific Journal of Tropical Medicine*, 5(6), pp. 485–497. doi: 10.1016/S1995-7645(12)60084-4.

Sinha, A. and Sarkar, S. (2013) 'Ubiquitin-Proteasome System- a target to control pathogenic protozoa', in *Microbial pathogens and strategies for combating them: science, technology and education*. Available at: https://www.researchgate.net/publication/279847103 (Accessed: 23 September 2021).

Smyth, M. S. and Martin, J. H. J. (2000) 'x Ray crystallography', *Molecular Pathology*, 53(1), pp. 8–14. doi: 10.1136/mp.53.1.8.

Snider, J. *et al.* (2015) 'Fundamentals of protein interaction network mapping', *Molecular Systems Biology*, 11(12), p. 848. doi: 10.15252/MSB.20156351.

Söderberg, O. *et al.* (2006) 'Direct observation of individual endogenous protein complexes in situ by proximity ligation', *Nature Methods*, 3(12), pp. 995–1000. doi: 10.1038/nmeth947.

Sommer, T. and Jentsch, S. (1993) 'A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum', *Nature*, 365(6442), pp. 176–179. doi: 10.1038/365176A0.

Soucy, T. A. *et al.* (2009) 'An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer', *Nature*, 458(7239), pp. 732–736. doi: 10.1038/nature07884.

States, D. J., Gish, W. and Altschul, S. F. (1991) 'Improved sensitivity of nucleic acid database searches using application-specific scoring matrices', *Methods*, 3(1), pp. 66–70. doi: 10.1016/S1046-2023(05)80165-3.

Sterkers, Y. *et al.* (2012) 'Novel insights into genome plasticity in Eukaryotes: Mosaic aneuploidy in Leishmania', *Molecular Microbiology*. John Wiley & Sons, Ltd, pp. 15–23. doi: 10.1111/j.1365-2958.2012.08185.x.

Stewart, M. D. *et al.* (2016) 'E2 enzymes: more than just middle men', *Nature Publishing Group*, 26(4), pp. 423–440. doi: 10.1038/cr.2016.35.

Studier, F. W. *et al.* (1990) 'Use of T7 RNA polymerase to direct expression of cloned genes', *Methods in enzymology*, 185, pp. 60–89. doi: 10.1016/0076-6879(90)85008-C.

Sun, X. and Qiu, H. (2020) 'Valosin-Containing Protein, a Calcium-Associated ATPase Protein, in Endoplasmic Reticulum and Mitochondrial Function and Its Implications for Diseases', *International Journal of Molecular Sciences*, 21(11). doi: 10.3390/IJMS21113842.

Sundar, S. (2001) 'Drug resistance in Indian visceral leishmaniasis', *Tropical Medicine & International Health*, 6(11), pp. 849–854. doi: 10.1046/J.1365-3156.2001.00778.X.

Sundar, S. and Olliaro, P. L. (2007) 'Miltefosine in the treatment of leishmaniasis: Clinical evidence for informed clinical risk management', *Therapeutics and Clinical Risk Management*, 3(5), p. 733. Available at: /pmc/articles/PMC2376078/ (Accessed: 1 March 2022).

Sundar, S. and Singh, B. (2018) 'Understanding Leishmania parasites through proteomics and implications for the clinic', *Expert review of proteomics*, 15(5), p. 371. doi: 10.1080/14789450.2018.1468754.

Tak-243 | C19H20F3N5O5S2 - PubChem (no date) *PubChem*. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/71715374#section=2D-Structure (Accessed: 25 September 2021).

Télles, S. *et al.* (1999) 'Trypanosoma cruzi and human ubiquitin are immunologically distinct proteins despite only three amino acid difference in their primary sequence', *FEMS Immunology and Medical Microbiology*, 24(2), pp. 123–130. doi: 10.1016/S0928-8244(99)00011-5.

Terpe, K. (2003) 'Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems', *Applied Microbiology and*

Biotechnology, 60(5), pp. 523–533. doi: 10.1007/s00253-002-1158-6.

Thomas, J. M. (2012) 'The birth of X-ray crystallography', *Nature*. Nature Publishing Group, pp. 186–187. doi: 10.1038/491186a.

Thomas, S. O. *et al.* (1956) 'Nucleocidin, a new antibiotic with activity against Trypanosomes', *Antibiotics annual*, pp. 716–21.

Tiengwe, C., Muratore, K. A. and Bangs, J. D. (2016) 'Surface proteins, ERAD and antigenic variation in Trypanosoma brucei', *Cellular Microbiology*, 18(11), pp. 1673–1688. doi: 10.1111/CMI.12605.

Torres-Guerrero, E. *et al.* (2017) 'Leishmaniasis: a review', *F1000Research*, 6. doi: 10.12688/F1000RESEARCH.11120.1.

Trinkle-Mulcahy, L. *et al.* (2008) 'Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes', *Journal of Cell Biology*, 183(2), pp. 223–239. doi: 10.1083/jcb.200805092.

Tsuchiya, S. *et al.* (1980) 'Establishment and characterization of a human acute monocytic leukemia cell line (THP-1)', *International Journal of Cancer*, 26(2), pp. 171–176. doi: 10.1002/ijc.2910260208.

Udeshi, N. D. *et al.* (2020) 'Rapid and deep-scale ubiquitylation profiling for biology and translational research', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-019-14175-1.

Ullah, H. (2012) 'The Role of Ion Exchange Chromatography in Purification and Characterization of Molecules', in *Ion Exchange Technologies*. InTech. doi: 10.5772/52537.

Ulrich, H. D. (2008) 'The Fast-Growing Business of SUMO Chains', *Molecular Cell*, 32(3), pp. 301–305. doi: 10.1016/j.molcel.2008.10.010.

Urán Landaburu, L. *et al.* (2020) 'TDR Targets 6: Driving drug discovery for human pathogens through intensive chemogenomic data integration', *Nucleic Acids Research*, 48(D1), pp. D992–D1005. doi: 10.1093/nar/gkz999.

Valimberti, I. *et al.* (2015) 'E2 superfamily of ubiquitin-conjugating enzymes: constitutively active or activated through phosphorylation in the catalytic cleft', *Scientific Reports*, 5(April), p. 14849. doi: 10.1038/srep14849.

Varland, S., Osberg, C. and Arnesen, T. (2015) 'N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects', *Proteomics*. Wiley-VCH Verlag, pp. 2385–2401. doi: 10.1002/pmic.201400619.

Vekilov, P. G. (2016) 'Nucleation of protein crystals', *Progress in Crystal Growth and Characterization of Materials*. Elsevier Ltd, pp. 136–154. doi: 10.1016/j.pcrysgrow.2016.04.007.

Vijay-Kumar, S., Bugg, C. E. and Cook, W. J. (1987) 'Structure of ubiquitin refined at 1.8 Å resolution', *Journal of Molecular Biology*, 194(3), pp. 531–544. doi: 10.1016/0022-2836(87)90679-6.

Vinayak, S. *et al.* (2015) 'Genetic modification of the diarrhoeal pathogen Cryptosporidium parvum', *Nature*, 523(7561), pp. 477–480. doi: 10.1038/nature14651.

Visscher, M., Arkin, M. R. and Dansen, T. B. (2016) 'Covalent targeting of acquired cysteines in cancer', *Current Opinion in Chemical Biology*. Elsevier Ltd, pp. 61–67. doi: 10.1016/j.cbpa.2015.11.004.

Vittal, V. *et al.* (2014) 'Intrinsic disorder drives N-terminal ubiquitination by Ube2w', *Nature Chemical Biology*, 11(1), pp. 83–89. doi: 10.1038/nchembio.1700.

Wang, C. *et al.* (2001) 'Solution structure of ThiS and implications for the evolutionary roots of ubiquitin', *Nature Structural Biology*, 8(1). doi: 10.1038/83041.

Wang, X. *et al.* (2007) 'Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3', *Journal of Cell Biology*, 177(4), pp. 613–624. doi: 10.1083/JCB.200611063.

Watts, R. J., Hoopfer, E. D. and Luo, L. (2003) 'Axon pruning during Drosophila metamorphosis: Evidence for local degeneration and requirement of the ubiquitin-proteasome system', *Neuron*, 38(6), pp. 871–885. doi: 10.1016/S0896-6273(03)00295-2.

Weiss, T. M. (2017) 'Small angle scattering: Historical perspective and future outlook', in *Advances in Experimental Medicine and Biology*. Springer New York LLC, pp. 1–10. doi: 10.1007/978-981-10-6038-0_1.

Weissman, A. M. (2001) 'Themes and Variations on Ubiquitylation', *Nature Reviews Molecular Cell Biology*, 2(3), pp. 169–178. doi: 10.1038/35056563.

Wenzel, D. M. *et al.* (2011) 'UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids', *Nature*, 474(7349), pp. 105–108. doi: 10.1038/nature09966.

Wheeler, R. J. (2021) 'A resource for improved predictions of Trypanosoma and Leishmania protein three-dimensional structure', *PLOS ONE*, 16(11), p. e0259871. doi: 10.1371/JOURNAL.PONE.0259871.

van Wijk, S. J. L. and Timmers, H. T. M. (2010) 'The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins', *The FASEB Journal*, 24(4), pp. 981–993. doi: 10.1096/fj.09-136259.

Williams, R. A. M. *et al.* (2012) 'ATG5 Is Essential for ATG8-Dependent Autophagy and Mitochondrial Homeostasis in Leishmania major', *PLOS Pathogens*, 8(5), p. e1002695. doi: 10.1371/JOURNAL.PPAT.1002695.

Wiwanitkit, V. (2012) 'Interest in paromomycin for the treatment of visceral leishmaniasis (kala-azar)', *Therapeutics and Clinical Risk Management*, 8, p. 323. doi: 10.2147/TCRM.S30139.

Xia, X. (2017) 'Bioinformatics and Drug Discovery', *Current Topics in Medicinal Chemistry*, 17(15), pp. 1709–1726. doi: 10.2174/1568026617666161116143440.

Xie, S. *et al.* (2009) 'Emerging Affinity-Based Techniques in Proteomics', *Expert*

review of proteomics, 6(5), p. 573. doi: 10.1586/EPR.09.74.

Xu, G., Paige, J. S. and Jaffrey, S. R. (2010) 'Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling', *Nature Biotechnology*, 28(8), pp. 868–873. doi: 10.1038/nbt.1654.

Xu, G. W. *et al.* (2010) 'The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma', *Blood*, 115(11), pp. 2251–2259. doi: 10.1182/blood-2009-07-231191.

Xu, Y., Piston, D. W. and Johnson, C. H. (1999) 'A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins', *Proceedings of the National Academy of Sciences*, 96(1), pp. 151–156. doi: 10.1073/PNAS.96.1.151.

Yang, M. *et al.* (2006) 'Assembly of a polymeric chain of SUMO1 on human topoisomerase I in vitro', *Journal of Biological Chemistry*, 281(12), pp. 8264–8274. doi: 10.1074/jbc.M510364200.

Yang, Y. *et al.* (2007) 'Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics', *Cancer Research*, 67(19), pp. 9472–9481. doi: 10.1158/0008-5472.CAN-07-0568.

Ye, K. *et al.* (2015) 'Identification of enzymes involved in SUMOylation in Trypanosoma brucei', *Scientific Reports*, 5(1), p. 10097. doi: 10.1038/srep10097.

Ye, Y. (2006) 'Diverse functions with a common regulator: Ubiquitin takes command of an AAA ATPase', *Journal of Structural Biology*, 156(1), pp. 29–40. doi: 10.1016/J.JSB.2006.01.005.

Yoon, B.-J. (2009) 'Hidden Markov Models and their Applications in Biological Sequence Analysis', *Current Genomics*, 10(6), pp. 402–415. doi: 10.2174/138920209789177575.

Zhang, Y. *et al.* (2011) 'UBE2W Interacts with FANCL and Regulates the

Monoubiquitination of Fanconi Anemia Protein FANCD2', *Mol. Cells*, 31, pp. 113–122. doi: 10.1007/s10059-011-0015-9.

Zheng, N. *et al.* (2000) 'Structure of a c-Cbl–UbcH7 Complex: RING Domain Function in Ubiquitin-Protein Ligases', *Cell*, 102(4), pp. 533–539. doi: 10.1016/S0092-8674(00)00057-X.

Zink, A. R. *et al.* (2006) 'Leishmaniasis in ancient Egypt and Upper nubia.', *Emerging infectious diseases*, 12(10), pp. 1616–7. doi: 10.3201/eid1210.060169.