UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Targetin deubiquitinating enzymes in kinetoplastids with small molecules

Las enzimas desubiquitinadoras como dianas de moléculas pequeñas en kinetoplástidos

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Nitin Jindal

Director

Fernando Antonio Ramón Olayo

Madrid, 2018

UNIVERSIDAD COMPLUTENSE MADRID

Targeting deubiquitinating enzymes in Kinetoplastids with small molecules

Las enzimas desubiquitinadoras como dianas de moléculas pequeñas en Kinetoplástidos

by

Nitin Jindal

Thesis director

Fernando Antonio Ramón Olayo

Submitted to

Biochemistry, Molecular Biology and Biomedicine PhD program

Faculty of Biological Sciences

Complutense University of Madrid

Dedicated to the patients who need new treatments to lead a better life

Acknowledgements

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

5-TAMRA	5-carboxytetramethylrhodamine
ABPP	Activity-based protein profiling
AML	Acute myeloid leukaemia
APF-1	ATP-dependent proteolysis factor 1
BGG	Bovine gamma globulin
BSA	Bovine serum albumin
CDC	Centers for disease control and prevention
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CL	Confidence limit
CRL	Cullin-ring ligase
dBSA	Heat-denatured bovine serum albumin
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
DUB	Deubiquitinase
EDTA	Ethylenediaminetetraacetic acid
ELT	Encoded library technology
E_{opt}	Optimum enzyme concentration
FP	Fluorescence polarization
FPLC	Fast protein liquid chromatography
GSK	GlaxoSmithKline
HAT	Human African trypanosomiasis
hCTSB	human cathepsin B
HECT	Homologous to E6AP C-terminus
HTS	High-throughput screening
IMAC	Immobilized metal ion affinity chromatography
IV	Intravenous
L-cys	L-cysteine
MCL	Mucocutaneous leishmaniasis
MERS-CoV	Middle East respiratory syndrome coronavirus
Min.	Minutes
MJD	Machado-Joseph domain
MOA	Mechanism of action

MOI	Multiplicity of infection
NB	Neuroblastoma
NECT	Nifurtimox–eflornithine combination therapy
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
OTU	Ovarian tumour
PCNA	Proliferating cell nuclear antigen
PD	Pharmacodynamics
PFI	Property forecast index
PI	Post-infection
PK	Pharmacokinetics
PTM	Post-translational modification
RBR	RING between RING
RING	Really interesting new gene
RT	Room temperature
S/B	Signal-to-background ratio
SAR	Structure-activity relationship
SARS-CoV	Severe acute respiratory syndrome coronavirus
SD	Standard deviation
SEC	Size exclusion chromatography
SOD	Superoxide dismutase
STING	Stimulator of interferon genes
TGF-β	Transforming growth factor beta
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
t_{reac}	Reaction time
Ub	Ubiquitin
Ub-FP	5 -TAMRA-Lys(Ub)-Gly-OH
UBIP	Ubiquitous immunopoietic polypeptide
UCH	Ubiquitin c-terminal hydrolases
USP	Ubiquitin-specific proteases
VL	Visceral leishmaniasis
WHO	World Health Organisation

Definitions

Assay window: The difference in signal of positive and negative controls.

Negative control or control 1: A negative control in a screening assay is the control that shows response as expected from an inactive compound in the assay.

Pan-DUB inhibitor: An inhibitor that specifically inhibits the activity of DUBs but not other enzymes.

Positive control or control 2: A positive control in a screening assay is the control that shows a response as expected from an active compound in the assay. For example, if the screening is for an inhibitor, the positive control will be the wells that show fully inhibited reaction which can be obtained by removing the enzyme or by treatment with a known inhibitor.

***Tbr*DUB1** is identified in Uniprot database with id Q386W6 and in TriTrypDB with id number Tb927.11.1930.

Summary

Millions of people suffer from and more than 1 billion people live at the risk of contracting human African trypanosomiasis (HAT), Chagas disease and leishmaniasis. These diseases are caused by *Trypanosoma brucei*, *T. cruzi* and *Leishmania* spp. respectively - members of the class Kinetoplastida. The current treatments against these diseases are old and have complex administration profiles, low efficacy and/or serious side effects. The causative parasites are also developing resistance against the available treatments. Therefore, there is an urgent need to discover new treatments for these diseases. For this it is important to focus on new targets that have proven to be tractable. The ubiquitin machinery provides an exciting opportunity as the target for development of new therapies. Deubiquitinases are the enzymes responsible for removing ubiquitin from target proteins and regulate a whole lot of cellular processes. They are being pursued as targets in several therapeutic areas. It was shown that some deubiquitinases are essential in *T. brucei*. The three parasites have similar genomic sequence and biology (El-Sayed *et al.*, 2005). Thus compounds targeting a conserved target could be active against all three parasites (Khare *et al.*, 2016).

We focussed our efforts on the discovery of new and selective small molecules capable of killing the parasite by ablating the activity of an essential deubiquitinase in *T. brucei*. For this we expressed and purified the enzyme and screened the GSK library of ~1.7 million compounds. After confirming the hits and determining their potencies, we followed a systematic triage strategy to finally obtain 64 hits with the IC_{50} in biochemical assay between 3 to 50 μM. These hits were tested for their potencies against whole cell *T. brucei*, *T. cruzi* and *L. donovani* and for their cytotoxic effect on HepG2 cells. Finally, we were able to identify 14 selective hits that showed activity against whole cell *T. brucei* and were either not cytotoxic or the IC_{50} was at least 2.5 times higher against *T. brucei* compared to HepG2 cells. Also, we were able to identify 2 hits that showed IC_{50} of 8.7 μM and 6.6 μM against *T. brucei*, 3.98 μM and 1.99 μM against *L. donovani*, 15 μM and 10.4 μM against *T. cruzi* respectively. These compounds provide very promising starting points for medicinal chemistry efforts.

Resumen

Millones de personas sufren y más de 1000 millones están en riesgo de contraer la tripanosomiasis Africana, el mal de Chagas o la leishmaniasis. Estas enfermedades son causadas por los kinetoplástidos *Trypanosoma brucei*, *T. cruzi* y *Leishmania* spp., respectivamente. Los tratamientos actualmente disponibles para estas enfermedades son viejos y poseen perfiles complejos de administración, baja eficacia y/o serios efectos adversos. Además, los parásitos causantes están desarrollando resistencias a los tratamientos existentes. Por estas razones hay una necesidad urgente de descubrir nuevos tratamientos para estas dolencias. Es por tanto importante enfocarse en nuevas dianas que demuestren ser tratables. La maquinaria de ubiquitina ofrece una oportunidad excitante como diana para el desarrollo de nuevas terapias. Las desubiquitinasas son enzimas responsables de retirar moléculas de ubiquitina de proteínas diana y regulan numerosos procesos celulares. Estas enzimas ya están siendo exploradas como dianas en otras áreas terapéuticas. Recientemente se ha demostrado que algunas desubiquitinasas son esenciales en *T. brucei*. Los 3 parásitos tienen genomas y biologías similares (El-Sayed *et al.*, 2005). Así, cabe esperar que las moléculas que modulen una diana conservada sean activas frente a los 3 patógenos (Khare *et al.*, 2016).

Hemos enfocado nuestros esfuerzos en el descubrimiento de compuestos de bajo peso molecular que sean novedosos y capaces de matar selectivamente al parásito mediante la anulación de la actividad de una desubiquitinasa esencial de *T. brucei*. Para ello hemos expresado y purificado la enzima y hemos llevado a cabo una campaña de bioprospección molecular a gran escala usando la colección de 1.7 millones de compuestos de GSK. Tras confirmar los activos iniciales y determinar sus potencias inhibitorias, seguimos una estrategia de perfilado biológico sistemática hasta obtener la selección final de 64 compuestos con potencias en el ensayo bioquímico comprendidas entre 3 y 50 µM. Procedimos entonces a probar las actividades tripanocidas de estos compuestos enfrentándolos a *T. brucei*, *T. cruzi* y *L. donovani* y también estudiamos sus efectos citotóxicos sobre células HepG2. Finalmente, fuimos capaces de identificar 14 moléculas selectivas que además mostraban actividad en ensayos con célula entera de *T. brucei* y mostraban una citotoxicidad inapreciable o de una magnitud al menos 2.5 veces inferior que frente a *T. brucei* en comparación con células HepG2. Cabe destacar que hemos sido capaces de identificar 2 compuestos pan-kinetocidas que mostraron valores de IC_{50} de 8.7 µM y 6.6 µM frente a *T. brucei*,

3.98 µM y 1.99 µM frente a *L. donovani*, 15 µM y 10.4 µM frente a *T. cruzi*, respectivamente. Estas estructuras químicas suponen puntos de partida muy prometedores para esfuerzos de química médica que conduzcan a su desarrollo.

Chapter 1 Introduction

1.1 The parasite burden

This work is an attempt to identify novel starting points for developing new treatments against human African trypanosomiasis (HAT), Chagas disease and leishmaniasis. These diseases are caused by *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* spp. respectively, all three of which are dixenous *i.e.* have two hosts in their life cycle. These organisms are classified in the family Trypanosomatidae, of the class Kinetoplastida, phylum Euglenozoa, super group Excavata of the domain Eukaryota. The Kinetoplastida have a defining characteristic which is the kinetoplast, an unusually structured DNA granule inside the mitochondria. All the members of Trypanosomatidae are obligate parasites.

HAT or sleeping sickness is caused in humans by two subspecies of *T. brucei-T. b. gambiense* and *T. b. rhodesiense*. Most cases of sleeping sickness are caused by *T. b. gambiense*. The disease is largely prevalent in the sub-Saharan Africa. Every year ~10,000 new cases of HAT are reported, around 50-60,000 people suffer from the disease and 70 million people are at the risk of contracting it (WHO, CDC). Both subspecies of *T. brucei* are transmitted by the tsetse fly (genus *Glossina*). An infected tsetse fly injects metacyclic trypanomastigotes in humans when it takes a blood meal. Once in humans they transform into bloodstream form. In stage I of the disease the parasite stays in the peripheral circulation- for example in blood and lymph and multiplies by binary fission. Typical symptoms during stage I infection include swollen lymph nodes, fever, headache and muscle and joint ache. After 1-2 years in the case of *T. b. gambiense* and after only a few weeks in the case of *T. b. rhodesiense*, the parasite enters the central nervous system and causes severe neurological disorders. This is the stage II of the disease. If left untreated, infection from either subspecies will prove fatal.

Chagas disease or American trypanosomiasis is caused by *Trypanosoma cruzi*. The disease is widespread in Central and South America. The number of people suffering from Chagas disease is approximated to be around 7 to 8 million, with

50,000 new cases each year. The number of people living at the risk of contracting Chagas disease is around 120 million ([WHO](), [CDC]()). *T. cruzi* is transmitted by the triatomine bug (genus *Triatoma*). When an infected triatomine bug takes a blood meal it passes *T. cruzi* metacyclic trypanomastigotes in the faeces. These trypanomastigotes enter the host through the wound or through mucosal membranes. Once in the host, the trypanomastigotes differentiate into intracellular amastigotes. The amastigotes will multiply by binary fission and transform into trypanomastigotes that are released in the bloodstream and can infect more cells and the cycle will continue. Chagas disease has 2 phases: acute and chronic. Acute phase is mild or asymptomatic. The symptoms could include fever, headache, swollen lymph nodes and swelling at the site of infection. Post acute phase an infected person can stay asymptomatic for years and might not even progress to chronic phase. In chronic phase the patients suffer from cardiac, digestive or neurological disorders. If left untreated, the infection can cause death or heart failure due to degradation of heart muscles.

In humans leishmaniasis can be caused by more than 20 species of *Leishmania*. Leishmaniasis is spread in around 90 tropical and sub-tropical countries. An estimated 12 million people are infected with *Leishmania* and more than 1 million new cases are reported every year ([Alvar *et al.*, 2012](), [WHO](), [CDC]()). More than 1 billion people live at a risk of being infected by *Leishmania spp*. The parasites are transmitted by the sandflies (genus *Phlebotomus*). An infected sandfly injects promastigotes in humans when it takes a blood meal. These promastigotes either actively invade or are phagocytosed into macrophages. Inside the macrophage the promastigotes differentiate into amastigotes. The amastigotes multiply inside the macrophages by binary fission and can go on to infect more cells. Depending on the species and environmental factor, the disease can present three main clinical manifestations: visceral leishmaniasis (VL) or kala-azar, cutaneous leishmaniasis and mucocutaneous leishmaniasis (MCL). VL affects internal organs like spleen and liver and is fatal if left untreated. The symptoms include fever, swollen spleen and liver, weight loss, anaemia, leucopoenia and thrombocytopenia. Cutaneous

leishmaniasis is the most common form of leishmaniasis. Cutaneous leishmaniasis patients develop skin lesions or sores. The lesions heal over time but can leave unpleasant scars and cause serious disabilities. MCL causes destruction of mucosal membranes of the nose, mouth or throat.

These three diseases put together have a very significant prevalence and mortality. They pose a serious socio-economic burden which has been estimated to be tens of billions of dollars every year. Despite this, as will be discussed in detail in the next section, a look at the currently available treatments points towards an urgent need to discover new treatments.

1.2 Available treatments: the unmet medical need

Kinetoplastid diseases have been neglected for decades in terms of research and development into drugs. Consequently, there are only a few treatments available that are old and have several disadvantages like high cost, toxicity, long administration profiles, a difficult route of administration and/or low efficacy in endemic areas.

For treatment of HAT, only five drugs are available. These drugs and the stages targeted are listed in **Table 1.1**. Melarsoprol is the only available treatment for stage II infections of *T. b. rhodesiense* and is also effective against stage II of *T. b. gambiense*. Suramin and pentamidine are used to treat stage I of *T. b. gambiense* infection. Eflornithine is effective against *T. b. gambiense*. However all of these drugs either have a complex dosage regimen and/or have very serious side effects. NECT (nifurtimox–eflornithine combination therapy) is relatively safer, has shorter administration profile as compared to eflornithine and has been shown not to be inferior to standard eflornithine therapy (Priotto *et al.*, 2009). More importantly, the parasite has started showing drug resistance in several foci (Barrett *et al.*, 2011, Baker *et al.*, 2013).

For the acute stage and congenital infection of Chagas disease benznidazole and nifurtimox are effective treatments (Bahia-Oliveira *et al.*, 2000). However they show poor efficacy against the chronic stages (Cancado, 2002). The side effects of benznidazole treatment include bone marrow depression, anorexia, peripheral neuropathy and allergic dermatitis. Nifurtimox treatment can cause anorexia, polyneuropathy, psychic alterations and nausea amongst other effects (Castro *et al.*, 2006, Wilkinson and Kelly, 2009). The dosage regimen for benznidazole and nifurtimox can last up to 60 and 90 days respectively. Due to the long administration profile the treatment is often not completed and increases the chances of emergence of resistance. Resistance against both these drugs has been reported (Wilkinson *et al.*, 2008, Campos *et al.*, 2014).

Table 1.1: WHO list of essential medicines for human African trypanosomiasis.

Drug	Year	Parasite targeted	Drawbacks/Side effects
Suramin	1920	Stage I *T. b. rhodesiense*	IV infusion - 5 days. Nephrotoxicity, peripheral neurotoxicity and bone marrow toxicity, which are usually mild and reversible (Malvy and Chappuis, 2011).
Pentamidine	1940	Stage I *T. b. gambiense*	IV infusion or intramuscular injection – 10 days. Pain at the injection site (rarely complicated by aseptic or septic abscess), hypoglycemia and hypotension (Malvy and Chappuis, 2011).
Melarsoprol	1949	Stage II *T. b. gambiense, T. b. rhodesiense*	Arsenic-based derivative. Lengthy IV dosing schedule (for 3 days every 7 days- total 9 days). A poorly tolerated drug: encephalopathic syndrome in 5–10% patients with high case-fatality rate, peripheral neuropathy, hepatic toxicity, skin rash, acute phlebitis and vein sclerosis.
Eflornithine	1990	Stage II *T. b. gambiense*	Safer than melarsoprol. But 56 intravenous infusions, >30 min each over 14 days. Bone marrow suppression and seizures. (Balasegaram *et al.*, 2009, Chappuis *et al.*, 2005, Malvy and Chappuis, 2011)
Nifurtimox–eflornithine combination	2009	*T. b. gambiense*	It simplifies the use of eflornithine by reducing the duration of treatment and the number of IV perfusions, but it was not studied for *T. b. rhodesiense*.

For the treatment of different clinical manifestations of leishmaniasis pentavalent antimonials, amphotericin B, paromycin and miltefosine are used. Most of these

treatments are old, have long dosage regimens, some have very serious side effects and the parasite has developed resistance against some (den Boer *et al.,* 2009, Nagle *et al.,* 2014).

Clearly, there is an urgent need for new drugs that are more effective, have higher potency and therefore require a lower dosage or shorter administration profiles, and have less side effects. Furthermore, it is very important that the drugs are inexpensive because these diseases are closely associated with poverty and are primarily found in areas with underdeveloped public health systems. For the development of new drugs with novel mechanisms of actions, it is important to dig deeper into the parasite cellular biology and identify new targets.

1.3 Ubiquitin machinery as drug target

Ubiquitin is a 76 residues long polypeptide with a molecular weight of around 8.5 kDa. The structure of ubiquitin is shown in **Figure 1.1**. Ubiquitin was first reported as a ubiquitous immunopoietic polypeptide (UBIP) with lymphocyte-differentiating properties (Goldstein *et al.,* 1975). Later, Ciechanover *et al.* identified a heat-stable ATP-dependent proteolysis factor (APF-1) in reticulocytes (Ciehanover *et al.,* 1978) which formed high molecular weight covalent conjugates with reticulocyte proteins in an ATP-dependent manner (Ciechanover *et al.,* 1980). APF-1 was later identified to be ubiquitin (Wilkinson *et al.,* 1980). Hershko *et al.* (Hershko *et al.,* 1980) first demonstrated two distinct enzymatic activities of the ubiquitin system. First they showed that the APF-1 conjugation to target proteins was enzyme-catalysed. They called this enzyme APF-1-protein amide synthetase and showed that the conjugation is processive *i.e.* several APF-1 moieties can be conjugated to a single target. The second enzymatic activity demonstrated was that of an amidase by showing that the APF-1 was regenerated from APF-1-protein conjugates. Aaron Ciechanover, Avram Hershko and Irwin Rose were awarded the Nobel Prize in chemistry in 2004 for the discovery of ubiquitin-mediated protein degradation. The field continues to grow immensely. For example, a combined

broad literature search using Embase (Elsevier), for the terms 'ubiquitin' or 'ubiquitination' in January 2017 yielded around 67000 records with 5000 to 6000 papers being published each year.
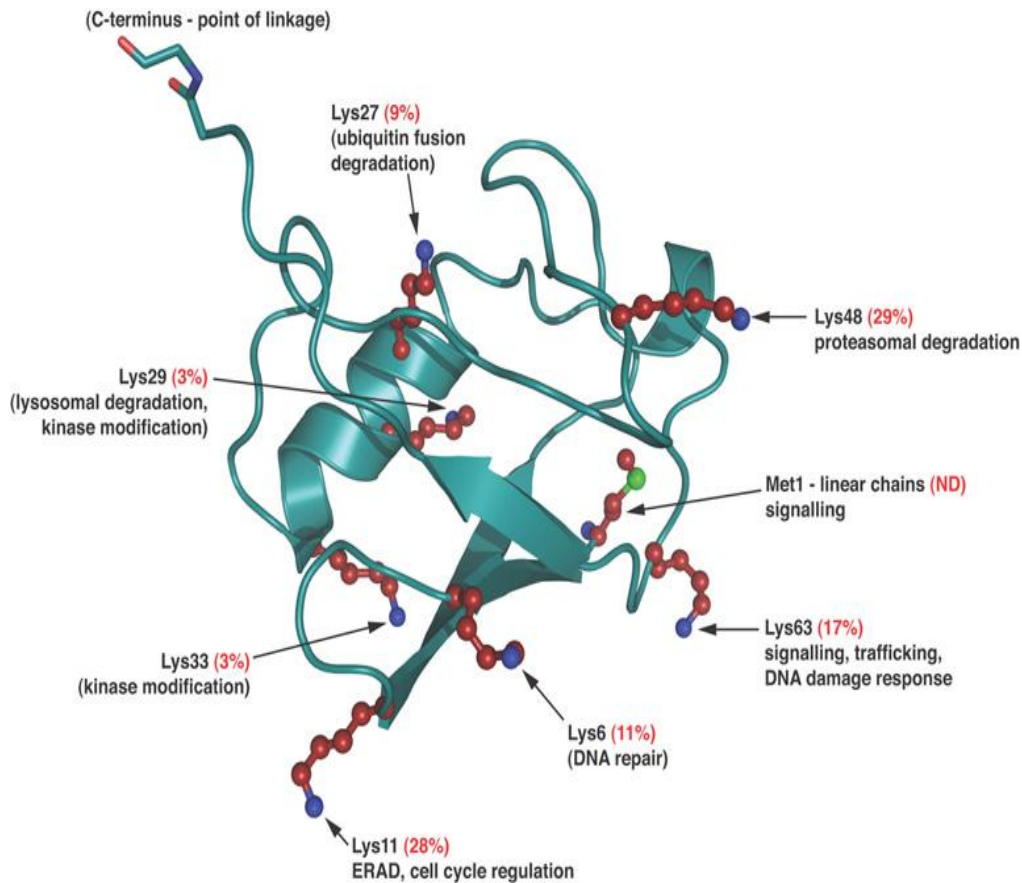
Figure 1.1 Structure of ubiquitin
The representation of the structure shown above is taken from (Komander, 2009) with the structure PDB code 1UBQ (Vijay-Kumar *et al.*, 1987). Ubiquitin has seven lysines and one methionine that act as substrates for another ubiquitin to attach and hence allow chain formation.

Ubiquitin covalently attaches to the ε-amino group of a lysine of the target protein via its C-terminal glycine forming an isopeptide linkage. As we now know, ubiquitination is a three-step process (**Figure 1.2**) involving activation, conjugation and ligation. Ubiquitin first covalently binds to the active site cysteine of an E1 enzyme or ubiquitin activating enzyme in an ATP-dependent manner. This ubiquitin is then transferred to an E2 enzyme referred to as ubiquitin conjugating enzyme.

The E2 then transfers the ubiquitin to an E3 ligase, which in turn transfers it to a lysine of the target protein, where it attaches via an isopeptide linkage. The human genome encodes for two E1 enzymes, 37 E2 enzymes and >600 E3 enzymes (Deshaies and Joazeiro, 2009, Markson *et al.*, 2009, Michelle *et al.*, 2009, Groettrup *et al.*, 2008). Like most post-translational modifications (PTMs) ubiquitination is a reversible process. Enzymes that cleave ubiquitin from the target are called deubiquitinases (DUBs). DUBs were identified by Hershko *et al.* (Hershko *et al.*, 1980) as amidases. These are discussed in greater detail in section 1.4.
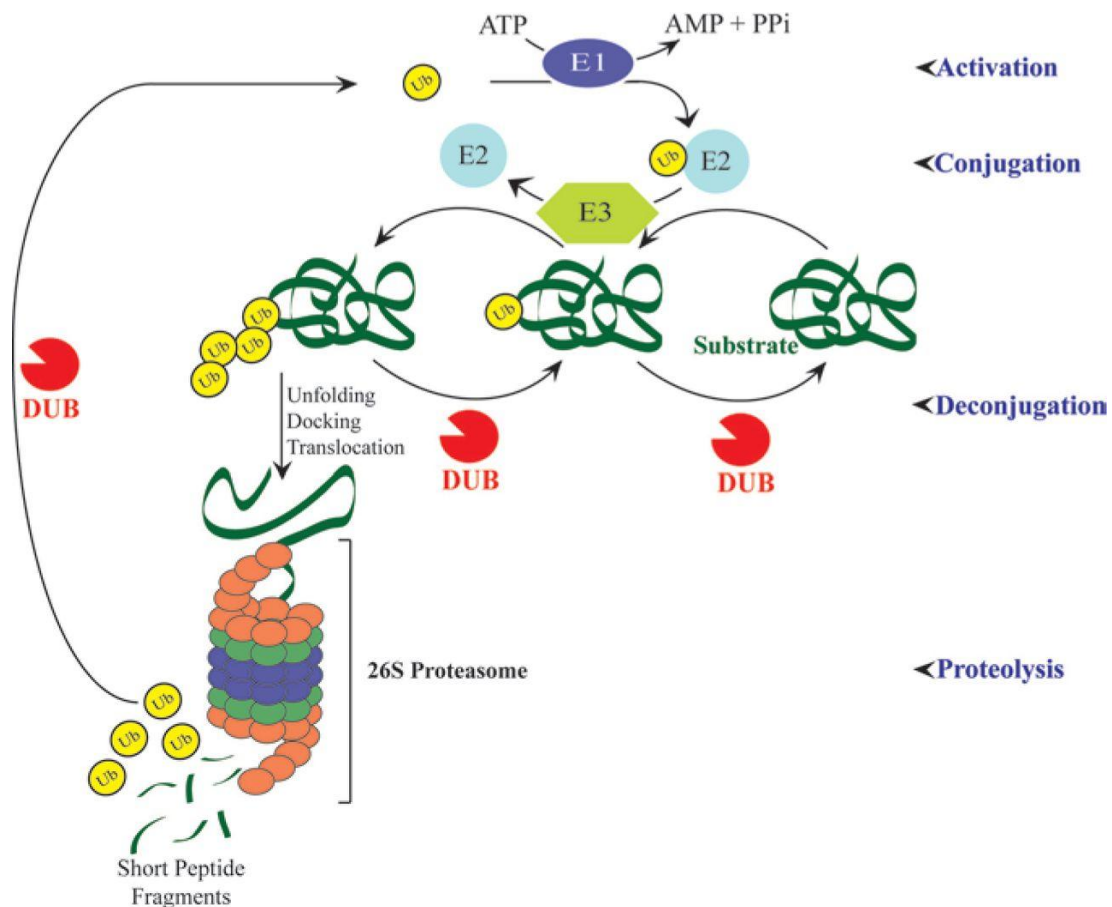
Figure 1.2 The ubiquitination pathway
Ubiquitin activation by E1s is an ATP-dependent process. This activated ubiquitin is transferred to an E2, and further an E3 couples it to the target protein. To remove the ubiquitin, DUBs come into action (Nicholson *et al.*, 2014).

The ubiquitin code is very complex ([Yau and Rape, 2016](#), [Komander, 2009](#)). Ubiquitin itself has 7 lysine residues *viz.* K6, K11, K27, K29, K33, K48 and K64 and also a terminal methionine (M1) residue which all can act as substrates for another ubiquitin to attach (**Figure 1.1**). This results in the formation of ubiquitin chains on a target protein. A protein can either be monoubiquitinated *i.e.* a single Ub on a single lysine, or multimonoubiquitinated *i.e.* a single Ub on multiple lysine residues or polyubiquitinated *i.e.* a chain of ubiquitin gets attached to a lysine residue. The polyubiquitin chains on a target can in turn either be homotypic or heterotypic. In homotypic chains all ubiquitins in the chain are attached by the same lysine or methionine. In a heterotypic chain the ubiquitins are attached by different lysines. The nature of ubiquitination on the target proteins can determine the fate of the target in different ways by signalling it for various roles or to different cellular localizations (reviewed in ([Komander, 2009](#), [Yau and Rape, 2016](#))). For example, K48 and K11-linked homotypic polyubiquitin chains target a protein primarily to the proteasome for degradation ([Chau *et al.*, 1989](#)). However, a minimum tetraubiquitin chain is required to efficiently target a protein for proteasomal degradation ([Thrower *et al.*, 2000](#)). Polyubiquitin chains of other linkage types regulate proteins by signalling for non-proteasomal fate. For example, K63-linked chains target receptors to the lysosome ([Lauwers *et al.*, 2009](#)) and play an important role in a variety of cellular processes like DNA repair ([Hoege *et al.*, 2002](#)), NF-κB activation ([Deng *et al.*, 2000](#)) and mitophagy ([Ordureau *et al.*, 2015](#), [Cunningham *et al.*, 2015](#)). K6 and K11-linked chains also regulate mitophagy ([Cunningham *et al.*, 2015](#)). K27-linked linkages regulate DNA damage response ([Gatti *et al.*, 2015](#)). Thus, the ubiquitin machinery has been shown to regulate a wide variety of cellular functions ([Ikeda and Dikic, 2008](#), [Haglund *et al.*, 2003](#), [Mukhopadhyay and Riezman, 2007](#), [Hicke, 2001](#)). The ubiquitin machinery is composed of more than one thousand proteins and many of them have been validated as drug targets in several therapeutic areas (reviewed in ([Tcherpakov, 2013](#), [Nalepa *et al.*, 2006](#), [Cohen and Tcherpakov, 2010](#))). These are discussed in detail in the following sections.

1.3.1 Proteasome as drug target

The global research and development market for ubiquitin proteasome is expected to grow with a five-year compound annual growth rate of 14.2% from estimated USD 2.9 billion in 2013 to USD 5.5 billion by 2018 (Tcherpakov, 2013). The success of **Bortezomib (PS-341)** (Adams, 2002) and the recent approval of **Carfilzomib** (Herndon *et al.*, 2013), both proteasome inhibitors for the treatment of multiple myeloma, have unequivocally established the proteasome as an attractive drug target (Nalepa *et al.*, 2006, Cohen and Tcherpakov, 2010). Additionally, there are several proteasome inhibitors in different stages of clinical trials (Cohen and Tcherpakov, 2010, Tcherpakov, 2013, Kisselev *et al.*, 2012). These include:

- **MLN9708** and **CEP18770** (Delanzomib) against multiple myeloma, non Hodgkin lymphoma, and solid tumours
- **Onyx0912** (Oprozomib) against haematological malignancies and solid tumours
- **Onyx0914** against autoimmune diseases
- **NPI-0052** (Marizomib/Salinosporamide A) against multiple myeloma

Furthermore, proteasome inhibitors have shown promising results in animal models against autoimmune and inflammatory diseases like lupus nephritis, lupus, myasthenia gravis, polyarthritis, rheumatoid arthritis, irritant sensitivity, psoriasis, asthma, colitis etc. (Tcherpakov, 2013, Neubert *et al.*, 2008, Ichikawa *et al.*, 2012, Gomez *et al.*, 2011, Fissolo *et al.*, 2008, Palombella *et al.*, 1998, Muchamuel *et al.*, 2009, Elofsson *et al.*, 1999, Elliott *et al.*, 2003, Elliott *et al.*, 1999, Schmidt *et al.*, 2010, Basler *et al.*, 2010, Kisselev *et al.*, 2012). Recently, Novartis has shown that **triazolo pyrimidine derivatives** inhibit proteasome activity in protozoan parasites and filed a patent on using them for the treatment of diseases like leishmaniasis (Biggart *et al.*, 2015). It is therefore reasonable to consider the application of proteasome inhibitors in anti-parasite drug discovery, as was patented in 1998 (Daniel *et al.*, 1998).

1.3.2 E1 enzymes as drug targets

E1 enzymes are also known as ubiquitin activating enzymes. They catalyse the first step of the ubiquitination pathway in an ATP-dependent manner. Pyrazolidine compounds **PR41** (Yang *et al.*, 2007) and **PYZD-4409** (Xu *et al.*, 2010) have been shown to inhibit the activity of E1 enzymes. PR41 prevents the loss of p53 and can inhibit cytokine induced activation of NF-κB (Yang *et al.*, 2007). It also blocks the activation of dendritic cells via IκBa/NF-κB and MKP1/ERK/STAT1 pathways thus showing potential as a therapeutic agent in autoimmune diseases (Chen *et al.*, 2014). PYZD-4409 preferentially induces cell death in malignant cell lines and prevents growth of primary AML cells (Xu *et al.*, 2010). **Benzothiazole** and **thiazole [5,5-b] pyridine** compounds were also identified as inhibitors of E1 enzyme (Look *et al.*, 2005).

1.3.3 E2 enzymes as drug targets

E2 enzymes transfer the ubiquitin from the E1 enzymes to E3 enzymes. They regulate the length of the ubiquitin chains, processivity of chain formation and chain topology (reviewed in (Ye and Rape, 2009)). Different E2 enzymes could work together, with one being used for chain initiation and another for chain elongation with a specific topology (Windheim *et al.*, 2008). Given the diverse regulatory functions of E2 enzymes, efforts are being made to find inhibitors of E2 enzymes.

hCDC34 is an E2 enzyme that works with cullin-ring (E3) ligases (CRLs) to ubiquitinate several targets. Ceccarelli *et. al.* (Ceccarelli *et al.*, 2011) were working towards blocking the SCF^{skp2} (an E3 ligase) dependent ubiquitination of p27^{Kip1} as a way of therapeutic intervention into different cancer types (Nalepa *et al.*, 2006, Frescas and Pagano, 2008). They identified **CC0651**, an allosteric inhibitor of hCDC34 that acts by stabilizing the E2-ubiquitin interaction and inhibits the proliferation of cancer cell lines (Ceccarelli *et al.*, 2011, Huang *et al.*, 2014).

Rad6B is another important E2 enzyme. It stabilizes β-catenin by ubiquitination (Shekhar *et al.,* 2008). It also ubiquitinates proliferating cell nuclear antigen (PCNA) and hence plays an important role in translesion synthesis DNA repair (Hoege *et al.,* 2002, Ulrich, 2005). Sanders *et. al.* (Sanders *et al.,* 2013) found **triazine analogs** can inhibit the activity of endogenous RaD6B and can inhibit proliferation, colony formation and migration of a breast cancer cell line.

Ubc13 or **UBE2N** is another E2 enzyme that has emerged as a very important therapeutic target due to its role in both oncology and immune disorders that have a combined market of USD 100 billion. Ubc13 catalyses K63-linked polyubiquitin chain formation (Hofmann and Pickart, 1999). It regulates p53 localization and activity (Laine *et al.,* 2006, Topisirovic *et al.,* 2009) and immune response (Fukushima *et al.,* 2007, Yamamoto *et al.,* 2006). Ubc13 catalyses the chain formation in the presence of either of the 2 cofactors: Uev1A or Mms2. Association with either cofactor regulates the chain length differently (Andersen *et al.,* 2005). Also, **Ubc13-Uev1A** is involved in NF- κB activation, while **Ubc13-Mms2** is involved in DNA repair (Andersen *et al.,* 2005). Pulvino *et. al.* (Pulvino *et al.,* 2012) identified **NSC697923**, a small molecule inhibitor of Ubc13-Uev1A which inhibits the constitutive NF-κB activation in diffuse large B-cell lymphoma (DLBCL) cells. Later, Cheng *et. al.* (Cheng *et al.,* 2014) identified Ubc13 as a target in neuroblastoma (NB) cells due to its role in p53 sequestration. They found that NSC697923 induces apoptosis in NB cell lines by promoting nuclear localisation of p53 (Cheng *et al.,* 2014). Several other inhibitors of the Ubc13-Uev1A complex are being developed (Tsukamoto *et al.,* 2008, Scheper *et al.,* 2010, Ushiyama *et al.,* 2012).

1.3.4 E3 enzymes as drug targets

E3 ligases catalyse the last step of the ubiquitination cascade. E3 enzymes bind to the E2-Ub complex and the substrate and this ubiquitin is then transferred from the

E2 to the substrate. As mentioned earlier, more than 600 E3 ligases are expressed in humans. These ligases are classified in three classes: the really interesting new gene (RING) family, the RING between RING (RBR) family and the homologous to E6AP C-terminus (HECT) family (Deshaies and Joazeiro, 2009, Berndsen and Wolberger, 2014). RING family ligases bind to both the E2 and the target simultaneously and catalyse direct transfer of ubiquitin from the E2 to the target (Deshaies and Joazeiro, 2009). The RBR and HECT family ligases catalyse a two step transfer of ubiquitin: first, the ubiquitin is transferred from the E2 to the active site cysteine of the E3 ligase, followed by its transfer to the target (Spratt *et al.*, 2014, Huibregtse *et al.*, 1995). E3 ligases offer greater specificity for drug discovery compared to E1 and E2 enzymes because they are relatively more specific for the target they ubiquitinate (Li *et al.*, 2008). E3 ligases have been implicated in several therapeutic areas and hence are important targets (Cohen and Tcherpakov, 2010, Tcherpakov, 2013, Skaar *et al.*, 2014, Metzger *et al.*, 2012, Morrow *et al.*, 2015, Lipkowitz and Weissman, 2011).

An important example of an E3 ligase as a drug target is **HDM2** or MDM2. In healthy cells, p53 and HDM2 form an auto-regulatory feedback to regulate one another's levels (Patel and Player, 2008, Brooks and Gu, 2011). Activation of p53 by stress signals leads to upregulation of HDM2, which in turn blocks its transcriptional activity (Oliner *et al.*, 1993) and marks it for degradation (Haupt *et al.*, 1997, Honda *et al.*, 1997, Kubbutat *et al.*, 1997). Consequently, inhibition of HDM2 can lead to higher levels of p53 and exert anti-tumour effect. Several small molecules and peptides have thus been developed as inhibitors of HDM2 activity or HDM2-p53 protein-protein interaction (reviewed in (Patel and Player, 2008, Cohen and Tcherpakov, 2010, Tcherpakov, 2013)). Most notable of these are **JNJ26854165** from Johnson & Johnson and **Nutlin** from Roche that are in clinical trials or in early development for different cancer types like solid tumours, multiple myeloma, neuroblastoma (Cohen and Tcherpakov, 2010, Tcherpakov, 2013, Patel and Player, 2008, Lakoma *et al.*, 2015).

Inhibitors of Apoptosis proteins (**IAPs**) are another important class of E3 ligases from therapeutic intervention perspective. IAPs have been shown to be key regulators of several signal transduction pathways like NF-κB, JNK1, TGF-β, Myc, PI3K/Akt and the MAPK pathway (reviewed in (Flygare and Fairbrother, 2010)). Several small molecules targeting IAPs are in different phases of clinical trials against several cancer types. These include Ascenta Therapeutics/Debiopharma's **AT-406/Debio 1143**, Genentech Inc./Curis' **GDC-0917/CUDC-427**, Novartis' **LCL161**, Tetralogic Pharmaceuticals' **TL-32711** and Aegera Therapeutics/Human Genome Sciences' **AEG40826/HGS1029** (Cohen and Tcherpakov, 2010, Tcherpakov, 2013, Flygare and Fairbrother, 2010, Hird *et al.*, 2015). **GDC-0152**, another small molecule inhibitor of IAPs, showed toxicity (Hird *et al.*, 2015). IAPs are also being pursued against autoimmune and inflammatory diseases, with Aegera pharmaceuticals have obtained a proof of concept data in rodent models of rheumatoid arthritis (Aegera's website and (Tcherpakov, 2013)). Several other E3 ligases like SMURF1, SCF^{Skp1}, Cereblon, anaphase-promoting complex (APC) and Parkin are also being pursued as therapeutic targets (Tcherpakov, 2013).

Different components of the ubiquitination machinery are thus validated targets in multiple therapeutic areas owing to their roles in regulation of various critical cellular pathways. With the examples of discovered and marketed small molecules, it is clear that all components of the ubiquitination machinery are tractable therapeutic targets in multiple therapeutic areas. Cross-talk between ubiquitination and other PTMs like phosphorylation (Yau and Rape, 2016) adds another layer of complexity to the ubiquitination process and offers additional ways of targeting ubiquitin-based regulation of cellular pathways.

1.4 Deubiquitinases as drug targets

As mentioned above, ubiquitination is a reversible post-translational modification, with deubiquitinases (DUBs) mediating the removal of ubiquitin from the target proteins. DUBs or ubiquitin-specific processing proteases are one of the largest

classes of proteases. The human genome encodes around 95 DUBs which are classified into five families based on the structure of their catalytic domain (Nijman *et al.*, 2005). Four families are cysteine proteases: the ubiquitin-specific proteases (USP/UBP), the ovarian tumour (OTU) domain-containing proteases, the ubiquitin c-terminal hydrolases (UCH) and the Machado-Joseph domain (MJD) proteases. The fifth family is formed by metalloproteases, the JAMM (JAB1/MPN/MOV34) motif proteases.
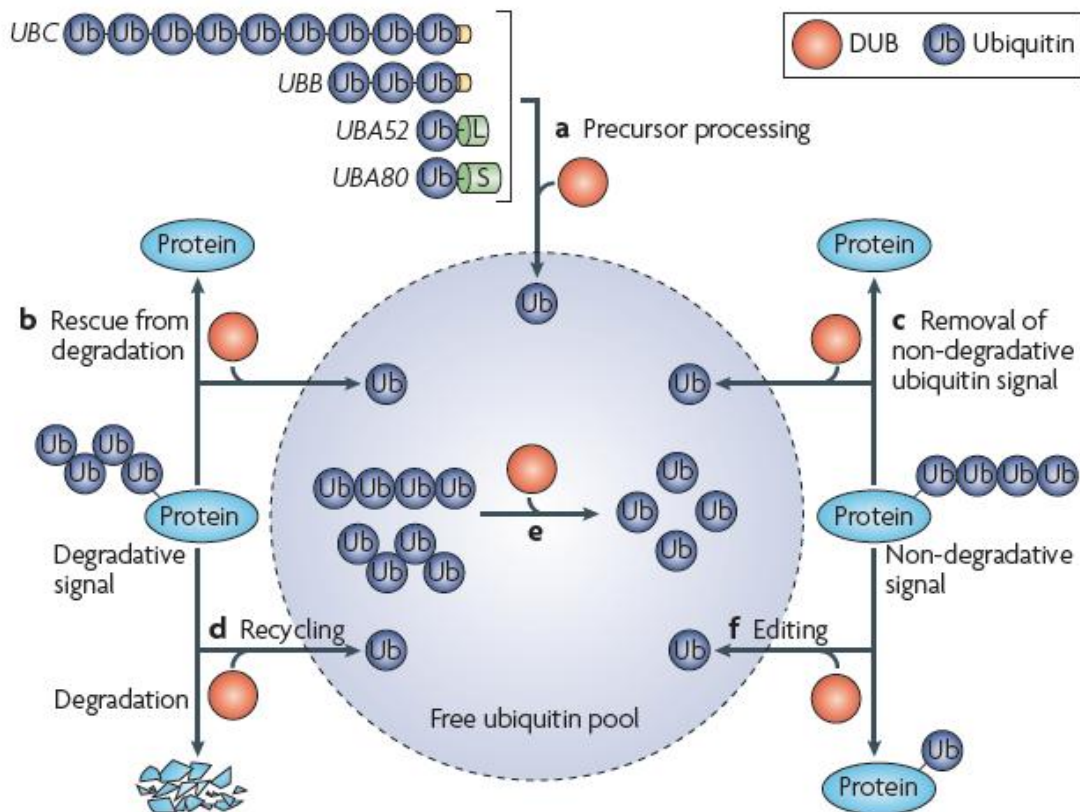
Figure 1.3 Roles of DUBs.
DUBs play several roles in the maintenance of ubiquitin homeostasis in the cell (Komander *et al.*, 2009).
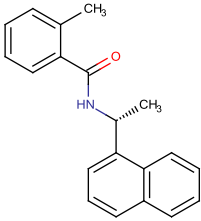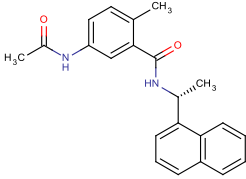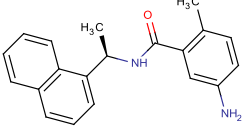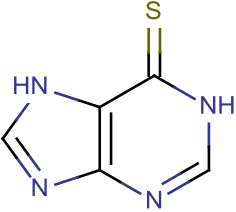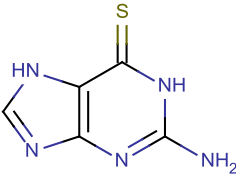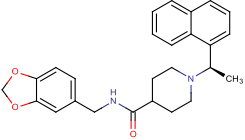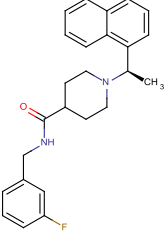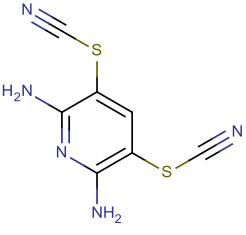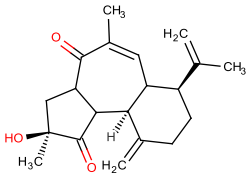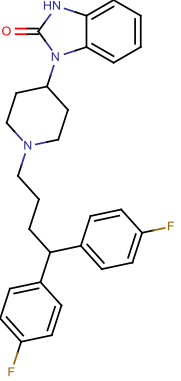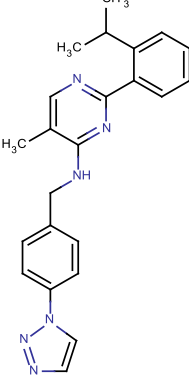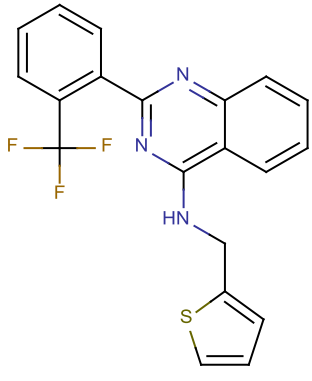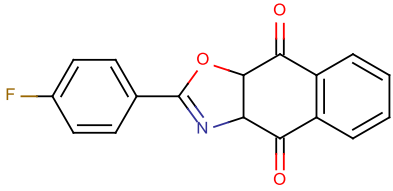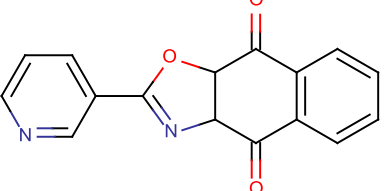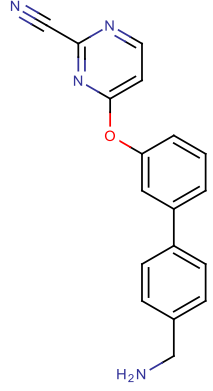
As shown in **Figure 1.3** DUBs play several roles in the ubiquitination pathway (reviewed in (Komander *et al.*, 2009)) and are thereby crucial regulators of diverse cellular processes like maintenance of monoubiquitin levels, substrate degradation at the proteasome, apoptosis and cell death, histone deubiquitination and hence

chromatin remodelling, cell cycle regulation, DNA damage repair response, activation of kinases, endocytosis, spermatogenesis and multiple signalling pathways (reviewed in (Komander *et al.*, 2009, Reyes-Turcu *et al.*, 2009, Wei *et al.*, 2015)). Nearly 40 deubiquitinases are implicated in the regulation of cancer cells and are reviewed extensively in (Wei *et al.*, 2015). Therefore, DUBs are being pursued as targets in several therapeutic areas like cancer, neurodegenerative disorders, haematological disorders, viral respiratory infections and bacterial gastroenteritis (reviewed in (Nicholson *et al.*, 2014, Tcherpakov, 2013, Farshi *et al.*, 2015, D'Arcy *et al.*, 2015, Ndubaku and Tsui, 2015, Colland, 2010, Daviet and Colland, 2008, Wei *et al.*, 2015)). Some examples of deubiquitinases as targets and their identified inhibitors are discussed below. The structures of known DUB inhibitors are presented in **Table 1.2**.
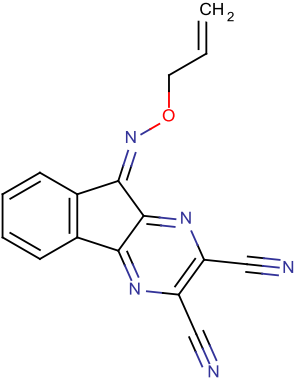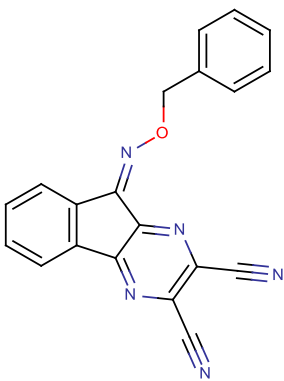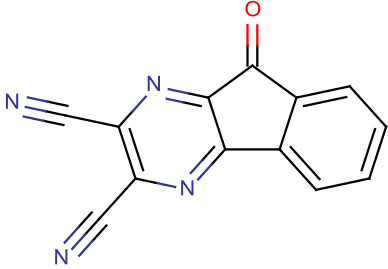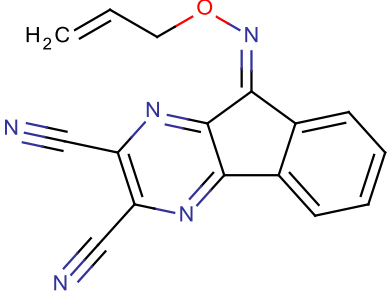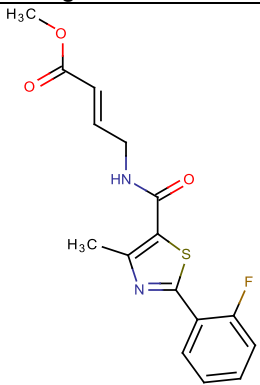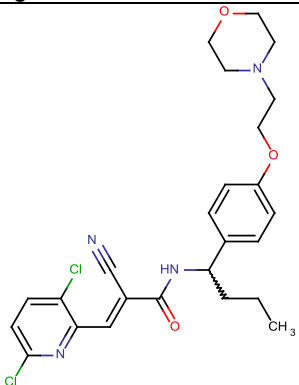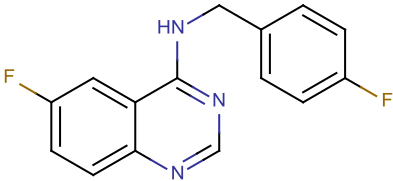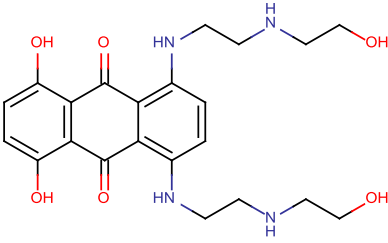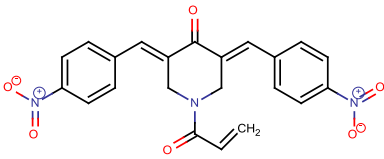
Table 1.2: Known inhibitors of human deubiquitinases and PLpro-a viral cysteine protease with deubiquitinase activity.
Each compound is identified by a unique serial number. Also, mentioned are the commercial names, if available, and the DUBs they target.
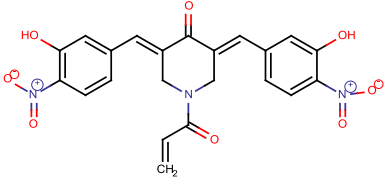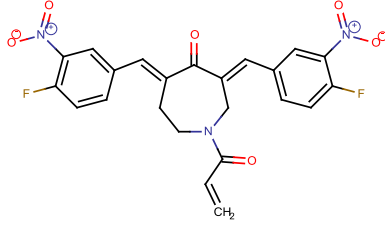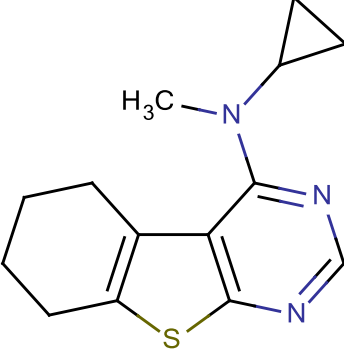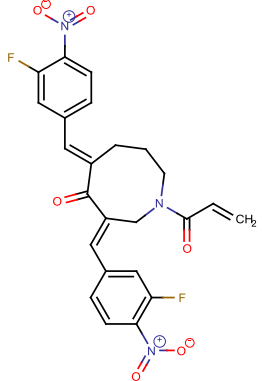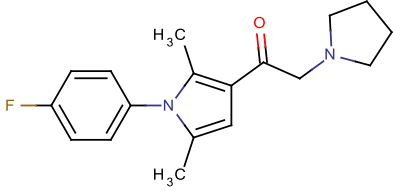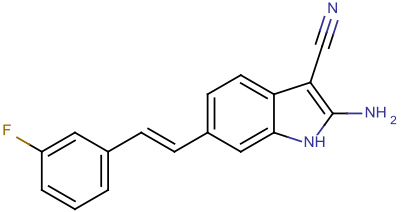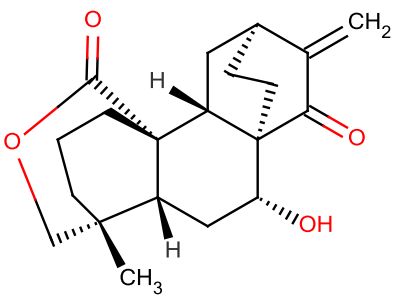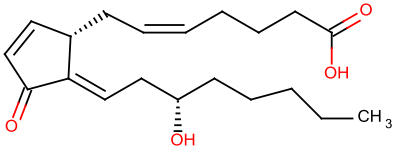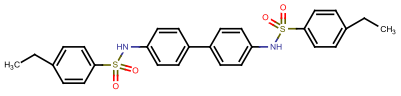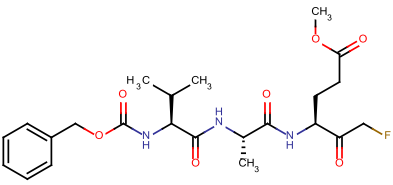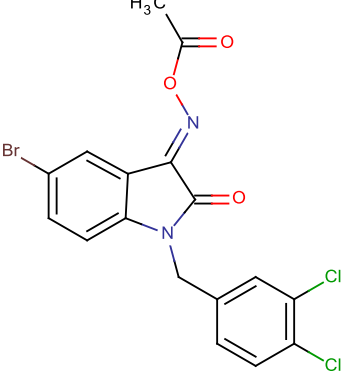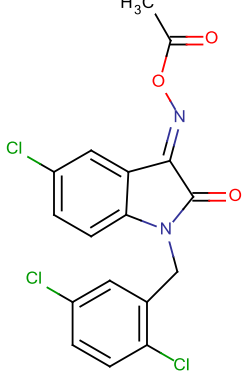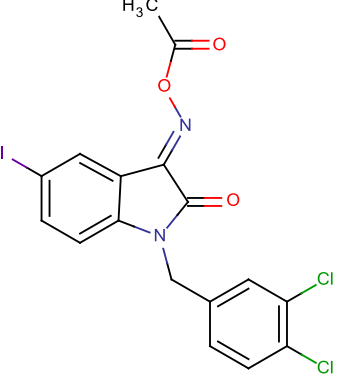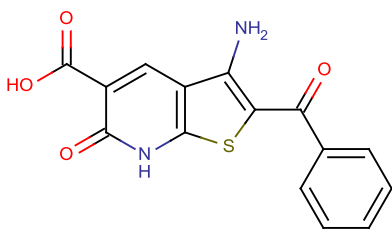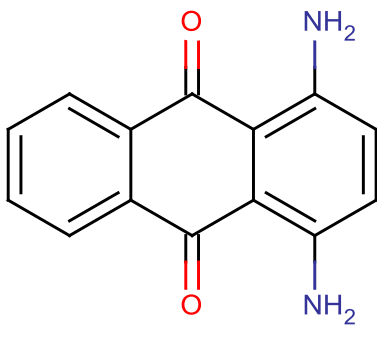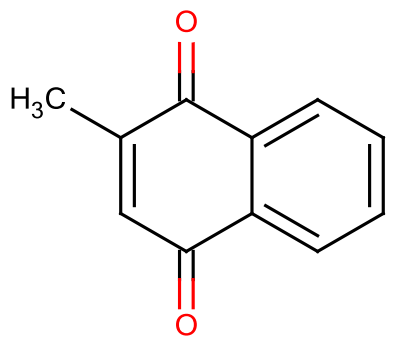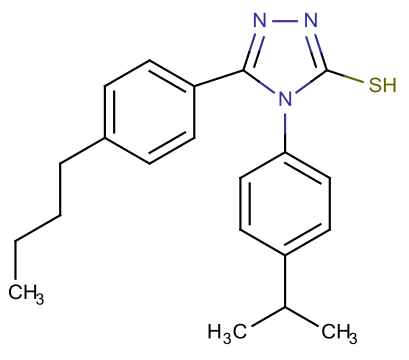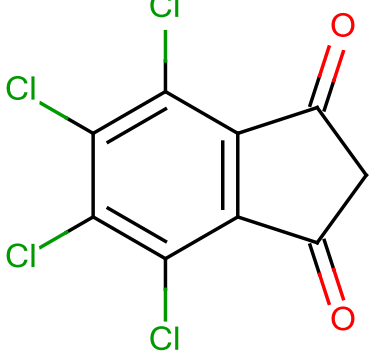
1 Target DUB: PLpro	**2** Target DUB: PLpro	**3** GRL0617 Target DUB: PLpro
4 6-mercaptopurine Target DUB: PLpro	**5** 6-thioguanine Target DUB: PLpro	**6** Target DUB: PLpro

7 Target DUB: PLpro	**8** PR-619 Target DUB: Non selective	**9** Curcusone D Target DUB: Non selective
10 Pimozide Target DUB: USP1/UAF1	**11** ML323 Target DUB: USP1/UAF1	**12** Target DUB: USP1/UAF1
13 C527 Target DUB: USP1/UAF1	**14** SJB3-019A Target DUB: USP1/UAF1	**15** 2-cyanopyrimidine Target DUB: USP1 and UCHL-3

16

Target DUB: USP1/UAF1 DUB complex

17
HBX 41,108
Target DUB: USP7

18

Target DUB: USP7

19
P5091
Target DUB: USP7 and USP47

20
P22077
Target DUB: USP7 and USP47

21

Target DUB: USP7 and USP47

22
Spongiacidin C
Target DUB: USP7

23
HBX19818
Target DUB: USP7

24

Target DUB: USP7

25 Target DUB: USP8	**26** Target DUB: USP8	**27** Target DUB: USP8 and USP7
28 Target DUB: USP8 and USP7	**29** Target DUB: USP8 binder	**30** EOAI3402143 Target DUB: USP9X
31 Spautin-1 Target DUB: USP10 and USP13	**32** Mitoxantrone Target DUB: USP11	**33** b-AP15 Target DUB: UCH-L5 and USP 14

34
b-AP15 analogue
Target DUB: UCH-L5 and USP
14

35
VLX1579, b-AP15 analogue
Target DUB: UCH-L5 and USP 14

36
IU2-6
Target DUB: USP14

37
Compound 1570 in patent US
20140228354 A1
Target DUB: UCH-L5 and USP
14

38
IU1

Target DUB: USP14

39
GSK2643943A

Target DUB: USP20

40

Target DUB: USP30

41
Δ12-PGJ2
Target DUB: UCHL1

42

Target DUB: UCHL1

43
Z-VAE(OMe)-FMK
Target DUB: UCHL1

44

Target DUB: UCHL1

45

Target DUB: UCHL1

46

Target DUB: UCHL1

47

Target DUB: UCHL1

48

Target DUB: UCHL3

49

Target DUB: UCHL3

50

Target DUB: UCHL3

51

Target DUB: UCHL3

41

52 Target DUB: UCHL3	**53** Target DUB: UCHL3	**54** Target DUB: UCHL3
55 Target DUB: RPN11	**56** PX-478 Target DUB: Not known	**57** WP1130 Target DUB: USP14, UCHL5, USP9x, USP5
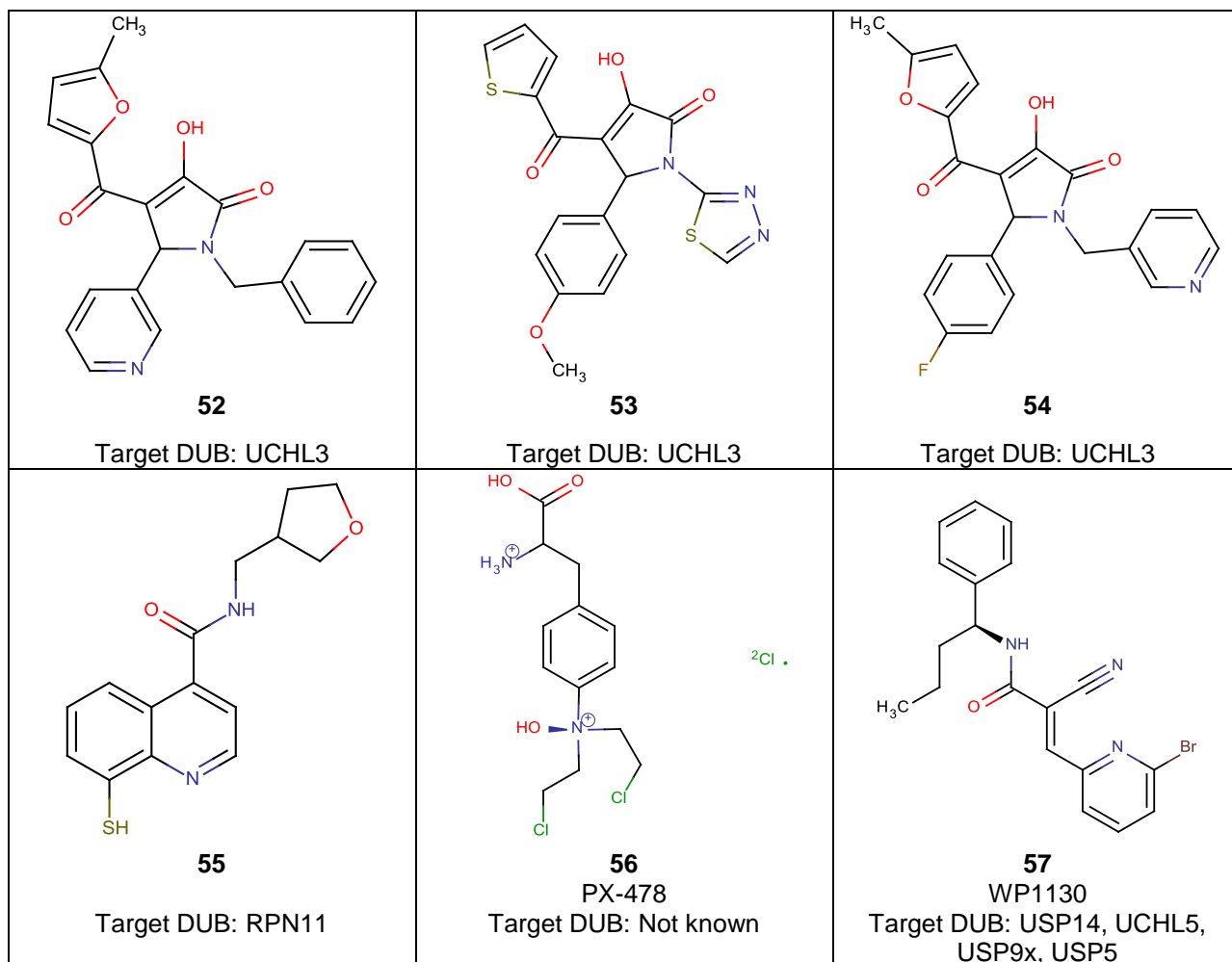
USP7

Deubiquitinase activity of USP7 stabilises HDM2 which in turn marks p53 for degradation (Cummins *et al.,* 2004, Li *et al.,* 2004). Inhibition of USP7 can lead thus to p53 stabilisation and tumour suppression. USP7 has also been shown to downregulate the activity of a transcription factor FOXO4 and thus regulates FOXO mediated stress response (van der Horst *et al.,* 2006). Several attempts have been made for the identification of inhibitors of USP7. The first identified inhibitor of USP7 was **HBX 41,108 (17)**, developed by Hybergenics (Colland *et al.,* 2009). It is a cyano-indinopyrazine derivative and is an uncompetitive reversible inhibitor of USP7. It was shown to stabilize p53 and induce p53 mediated apoptosis in cancer cell lines (Colland *et al.,* 2009). **P5091 (19)** and its analogues are another series of inhibitors of USP7 that were developed by Progenra (Chauhan *et al.,* 2012). P5091

is a trisubstituted thiophene with dichlorophenylthio, nitro and acetyl substitutions. It was shown to induce apoptosis in multiple myeloma cells and also overcome resistance against known therapies (Chauhan *et al.*, 2012). **HBX 19,818 (23)** and **HBX 28,258** are two specific inhibitors of USP7 developed by Hybergenics. They contain tetrahydroacridine core (Reverdy *et al.*, 2012).

USP14 and UCHL5

USP14 and UCHL5/Uch37 are two deubiquitinases that associate reversibly to the 19S regulatory particle of the proteasome (Yao *et al.*, 2006, Lee *et al.*, 2010, Borodovsky *et al.*, 2001). **b-AP15 (13)** is a small molecule inhibitor of the proteasome activity that acts via the inhibition of USP14 and UCHL5 (D'Arcy *et al.*, 2011). b-AP15 induces p53 independent apoptosis (Erdal *et al.*, 2005), causes tumour cell apoptosis and inhibits tumour growth *in vivo* (D'Arcy *et al.*, 2011) and shows inhibitory effect on multiple myeloma cell lines and patient cells (Tian *et al.*, 2014). **Auranofin** is a gold containing compound that has been used for treatment of rheumatoid arthritis since several decades. Recently auranofin was shown to inhibit proteasome activity via the inhibition of USP14 and UCHL5. This suppresses tumours *in vivo* and causes cytotoxicity in AML patient cells (Liu *et al.*, 2014). **WP1130 (57)** or degrasyn is a small molecule identified from a screen for inhibitors of cytokine induced activation of JAK/STAT pathway (Bartholomeusz *et al.*, 2007a). It is a second-generation tyrphostin derivative (Bartholomeusz *et al.*, 2007b). It has been shown to have apoptotic, anti-tumour and/or anti proliferative effects against chronic myelogenous leukaemia (Bartholomeusz *et al.*, 2007b), melanoma (Bartholomeusz *et al.*, 2007a) and mantle cell lymphoma (Pham *et al.*, 2010). It has been shown to be a partially selective deubiquitinase inhibitor targeting four deubiquitinases namely USP14, UCHL5, USP9X and USP5 (Kapuria *et al.*, 2010).

USP9X

USP9X/Fam, a deubiquitinase, has been implicated in the regulation of several cellular processes. For example, it plays a role in the positive regulation of the

TGF-β signalling (Dupont *et al.*, 2009), stabilizes β-catenin (Taya *et al.*, 1999), stabilizes Mcl1 (myeloid cell leukaemia 1) leading to increased survival of tumour cells (Schwickart *et al.*, 2010) and promotes polarity and self-renewal in stem cell derived neural progenitors (Jolly *et al.*, 2009). USP9X knockdown can have therapeutic effects in case of B-cell acute lymphoblastic leukaemia (Zhou *et al.*, 2015) and hepatocellular carcinoma (Hu *et al.*, 2015) while its activation can be beneficial for the treatment of pancreatic ductal adenocarcinoma (Perez-Mancera *et al.*, 2012). As mentioned above **WP1130 (57)** has been shown to inhibit USP9X activity (Kapuria *et al.*, 2010). Another identified inhibitor of USP9X activity is **EOAI3402143 (30)** which was shown to induce tumour cell apoptosis and suppressed growth of tumours in mice grafted with multiple myeloma tumours (Peterson *et al.*, 2015).

PLpro

PLpro is a cysteine protease in the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV). PLpro has a deubiquitinating activity against the host proteins. This deubiquitinating activity of PLpro inhibits the STING (<u>st</u>imulator of <u>i</u>nterferon <u>g</u>enes) signalling in host cells leading to the disruption of the innate antiviral immune response (Sun *et al.*, 2012). Thus, PLpro is considered an antiviral target. **6-mercaptopurine (4)** and **6-thioguanine (5)** were identified as inhibitors of both SARS-CoV PLpro and MERS-CoV PLpro (Chou *et al.*, 2008, Cheng *et al.*, 2015). **GRL0617 (3)** is another identified inhibitor of SARS-CoV PLpro (Ratia *et al.*, 2008). It is a competitive inhibitor with sub-micromolar potency (IC_{50} 0.6 µM). Several inhibitors with higher potencies have been developed by modifying GRL0617 (Kemp, 2016).

USP30

USP30 has been shown to be a therapeutic target against Parkinson's disease. In neurodegenerative diseases like Parkinson's disease, mitochondrial defects and dysfunction play a key role. Parkin, an E3 ligase, marks damaged mitochondria for

mitophagy by ubiquitinating outer mitochondrial membrane proteins (Narendra and Youle, 2011). USP30 localises to mitochondria and opposes Parkin mediated mitophagy by deubiquitinating outer membrane proteins (Bingol *et al.*, 2014, Cunningham *et al.*, 2015). A diterpenoid derivative, **15-oxospiramilactone (40)**, was shown to induce mitochondrial fusion and restore mitochondrial network (Yue *et al.*, 2014). It was identified to inhibit USP30 (Yue *et al.*, 2014).

Even though it has proven challenging to develop small molecule inhibitors with specificity against DUBs, there are several successful examples to conclude that drug discovery efforts against DUBs have potential and it is worth pursuing them as targets.

1.5 *Tbr*DUB1 is essential for parasite survival

While DUBs present an attractive and novel class of tractable targets in a wide range of therapeutic areas, they have remained largely unexplored in parasitic diseases. In a recent work, Rodenko *et. al.* identified at least 21 DUBs (*unpublished data*) in *T. brucei* by activity-based protein profiling (ABPP) (de Jong *et al.*, 2012). This was followed by RNAi knockdown screen to identify if any DUB is essential for the survival of the parasite. It was seen that when the RNAi was induced against *Tbr*DUB1, one of the 21 *Tbr*DUBs, it led to parasite clearance both *in vitro* and *in vivo* within 48h, suggesting that this DUB is essential for parasite survival *(**Figure 1.4** with permission from Rodenko, B. et. al. unpublished data)*.
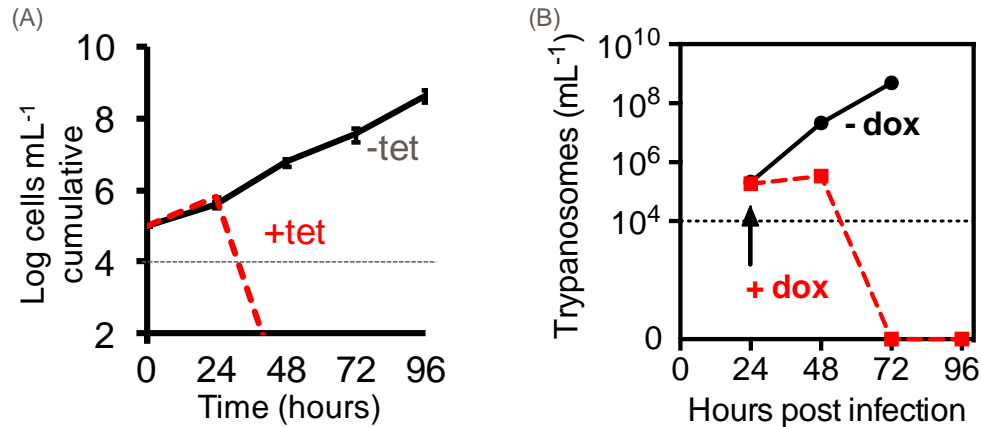
Figure 1.4 *Tbr*DUB1 is essential for parasite survival.
(a) *in vitro* RNAi knockdown shows that *Tbr*DUB1 knockdown is trypanocidal
(b) *in vivo* RNAi knockdown of *Tbr*DUB1 in parasites infecting mice (n=2 per group) shows that the parasitemia is again cleared within 48h. (--------- parasite detection limit: 10^4 cells/mL)
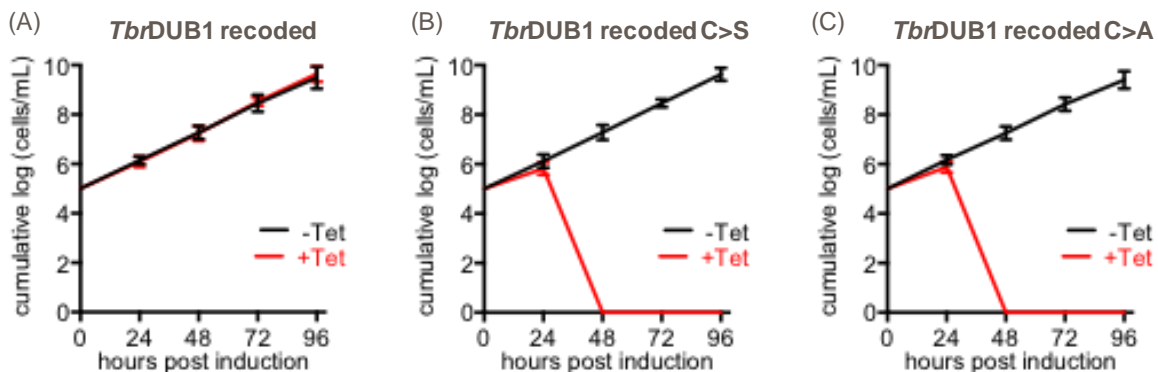
Figure 1.5 *Tbr*DUB1 activity is essential for parasite survival.
(a) A recoded *Tbr*DUB1 leads to phenotype rescue while (b) and (c) catalytically null mutants of *Tbr*DUB1 lead to parasite clearance within 48 hours.

This was followed by the generation of mutant parasites that either express catalytically active or inactive *Tbr*DUB1. In catalytically null mutants the active site cysteine of *Tbr*DUB1 was mutated to either serine or alanine. Parasites expressing catalytically inactive *Tbr*DUB1 die within 48h while parasites expressing wild-type *Tbr*DUB1 are unaffected (**Figure 1.5**). These experiments clearly establish that not just the expression, but also the activity of *Tbr*DUB1 is essential for the survival of the parasite. In addition to the strong genetic validation of the *Tbr*DUB1 as a

potential target, Rodenko *et.al.* also screened a small library of small-molecule pan-DUB inhibitors and found several molecules with trypanocidal activity. This observation chemically validates DUBs as a class as drug targets in *T. brucei*.

In summary, these results indicate that DUBs are chemically validated target class in *T. brucei* and that *Tbr*DUB1 is essential for parasite survival. These evidence are thus sufficient to launch a drug discovery effort targeting *Tbr*DUB1.

1.6 The process of drug discovery

Drug discovery is a complex interdisciplinary field involving numerous disciplines with biology and chemistry at the core working together to identify novel treatments. The cost associated with bringing a new treatment to the market is estimated at 2.6 billion US dollars (DiMasi *et al.,* 2016, DiMasi *et al.,* 2003, Munos, 2009). Additionally, the cost per new molecular entity has been increasing exponentially since 1950 at an annual rate of 13.4% (Munos, 2009). To get a perspective, nearly 1200 drugs have been approved by the FDA between 1950 and 2008, an average of 21 drugs per year (Munos, 2009). On an average time it takes more than 13 years from early-stage studies to bringing the drug to the market (Paul *et al.,* 2010). The process of drug discovery typically starts with the identification and validation of a target which could be molecular or cellular. This is followed by the identification of a small molecule or a biopharmaceutical that has the desired activity on the target, *i.e.* either inhibits or activates it depending on the therapeutic need. The small molecules at this stage are called hits. The next step involves medicinal chemistry efforts to optimise the potency, selectivity, pharmacokinetics and pharmacodynamics of these molecules at which stage they are declared as 'candidate'. These candidates are then tested in preclinical studies *in vitro* and/or in animals for their safety and optimum dosage for the first-time-in-humans. If successful in preclinical studies, the compound goes to first-time-in-human trials (phase I) to determine the safe dosage range and to assess the drug safety in a small group of healthy individuals. In phase II trials, the efficacy of the

drug is tested and the safety is further evaluated, involving up to hundreds of patients. Phase III trials are conducted in a much larger group of patients which could be from 300 to 3000 and can last up to 4 years. The efficacy, especially compared to already marketed treatments, and safety and side effects are tested in this phase. If successful, the drug is approved by regulatory bodies for treatment of the condition it was tested for. Post the launch, during the phase IV of the trials, further information is gathered about the safety, side effects, long term risks and benefits and efficacy.

The work in this thesis is focused on the early phases of drug discovery up to hit selection for *Tbr*DUB1. To identify small molecule inhibitor against the *Tbr*DUB1 several approaches can be taken and each has its advantages and disadvantages in terms of cost, speed, flexibility and the quality of the candidates delivered. The different approaches include full-diversity high-throughput screening (HTS), encoded library technology (ELT) screening, focused set screening, fragment based screening, virtual screening and knowledge-based or rational drug design. Since the structure of *Tbr*DUB1 is yet to be resolved, virtual screen, fragment based screening and rational drug design are not the approaches that we could take. An ELT screen identifies compounds that bind to the target with or without any effect on the activity of the target. A biochemical high-throughput screen will however focus on identifying inhibitors of enzyme activity. Starting with a diverse chemical library increases the chances of identifying tractable hits. A focused set is a set of compounds that could hit a target class. A HTS library contains all the compounds from focused sets, hence if feasible, HTS will have higher chances of success, because a larger diversity of compounds is tested. We focused our efforts on HTS at this stage and ELT screen could be carried out at a later stage, if required. It is financially not viable to apply all screening approaches to each target and hence often a choice has to be made based on feasibility and availability of tools. HTS has become a popular approach to early drug discovery over the past two decades and several marketed drugs originate from HTS campaigns (Macarron *et al.*, 2011, Swinney and Anthony, 2011). HTS involves screening a

library of small molecules against a target- using either biochemical or cell-based assays. The outcome of a HTS not only depends on the nature of the target and the quality of the assay, but is also heavily influenced by the nature of the screening library. The estimated number of molecules that could be theoretically synthesized ranges between 10^{20} and 10^{60}. In comparison, the number of seconds that have passed since the Big Bang is around 10^{17} (Valler and Green, 2000). As it is impossible to either synthesize or test the vast numbers of molecules, pharmaceutical companies build and expand their libraries with a focus on compound diversity in order to sample a broad chemical space. Several pharmaceutical companies have libraries up to one million compounds and only very few have larger collections. GlaxoSmithKline (GSK) has a HTS library of around 1.7 million compounds. This library has been improved over decades to be chemically diverse and is enriched in 'drug-like' molecules with desirable physicochemical properties. The importance of physicochemical properties is discussed in section 5.4.

1.7 The aim of the thesis

The questions that are explored in this thesis are:

Can we find small molecule inhibitors of TbrDUB1 that are specific and cidal against T. brucei? And are the hits obtained as an outcome of these efforts cidal against T. cruzi and/or L. donovani?

Chapter 3 of this thesis describes the enzyme purification and assay development. Chapter 4 describes the primary HTS of GSK's small molecule library (~1.7 million compounds) against *Tbr*DUB1 in a biochemical assay. Chapter 5 describes the process of hit characterisation and prioritisation through multiple secondary biochemical assays. The prioritised hits are tested in whole-cell assays against *T. brucei* and the closely related parasites *T. cruzi* and *L. donovani* to identify compounds with cross-species activity.

Chapter 2 Materials and Methods

2.1 Buffers and culture media

His trap buffer A
40 mM Tris pH 7.6
300 mM NaCl
10 mM imidazole
5 mM βME (add fresh)
pH 7.6
Filter through 0.2 micron filter

His trap buffer B
40 mM Tris pH 7.6
300 mM NaCl
500 mM imidazole
5 mM βME (add fresh)
pH 7.6
Filter through 0.2 micron filter

SEC buffer
20 mM Tris pH7.6
50 mM NaCl
5 mM βME (add fresh)
Autoclave

Dulbeco's PBS
Potassium Chloride 2.7 mM
Potassium Phosphate monobasic (KH_2PO_4) 1.47 mM
Sodium Chloride 137.9 mM
Sodium Phosphate dibasic (Na_2HPO_4-7H2O) 8 mM

4X Laemmli Sample Buffer
250 mM Tris-HCl, pH 6.8
4% Lithium dodecyl sulfate (LDS)
40% (w/v) glycerol
0.018% bromophenol blue
10% v/v (1.42 M) β mercaptoethanol

SDS Tank buffer
25 mM Tris
192 mM glycine
0.1% SDS
pH 8.6

DUB assay buffer
50 mM Tris pH 7.6
50 mM NaCl
1 mg/mL (1.626 mM) CHAPS

0.132 mg/mL (2 µM) dBSA
2 mM DTT
Filter through 0.2 micron filter. Degas for at least 10 minutes.

***Sans* DTT assay buffer**
50 mM Tris pH 7.6
50 mM NaCl
1 mg/mL (1.626 mM) CHAPS
0.132 mg/mL (2 µM) dBSA
Filter through 0.2 micron filter. Degas for at least 10 minutes.

SOD/Catalase assay buffer
50 mM Tris pH 7.6
50 mM NaCl
1 mg/mL (1.626 mM) CHAPS
0.132 mg/mL (2 µM) dBSA
10 µg/mL SOD
500 µg/mL catalase
2 mM DTT
Filter through 0.2 micron filter. Degas for at least 10 minutes.

L-cys assay buffer
50 mM Tris pH 7.6
50 mM NaCl
1 mg/mL (1.626 mM) CHAPS
0.132 mg/mL (2 µM) dBSA
2 mM L-cysteine
Filter through 0.2 micron filter. Degas for at least 10 minutes.

Resazurin assay buffer
50mM Hepes
50mM KCl
pH 7.5

hCTSB assay buffer
100 mM Sodium Acetate
150 mM NaCl
5 mM EDTA
1 mM CHAPS
40 µM L-cysteine
pH 5.5

T. brucei culture medium
1mM Hypoxanthine
0.16 mM Thymidine
1.5 mM L- Cysteine
0.05 mM Bathocuproinedisulfonic acid disodium salt

1 mM Sodium pyruvate solution
0.0014% 2-Mercaptoethanol
Make up volume with Iscove's Modified Dulbecco's Medium
pH 7.5
Filter sterilize through 0.2 micron filter

T. brucei complete culture medium
90% v/v _T. brucei_ culture medium
10% v/v foetal bovine serum (tetracycline free- Gibco 26140-079).

HepG2 cell culture medium
Eagles MEM with L-Glutamine and Earle's Salts
10 % FBS
1% of 100X non essential amino acid solution

2.2 Materials

2.2.1 Chemical reagents

4x Laemmli Sample Buffer Biorad 161-0747; 7-Amino-4-(trifluoromethyl)coumarin Sigma 248924; Bathocuproinedisulfonic acid disodium salt Sigma B1125-500mg; BCA Protein Assay Reagent Pierce 23225; Bradford's reagent Sigma B6916; BSA Sigma A3803; Catalase Sigma C100; Cellfectin II Life technologies 10362-100; CellTitre-Glo luminescent cell viability assay reagent Promega G7571; CHAPS Sigma C-3023; Digitoxin Sigma D-5878; DMSO Sigma 34943; DTT Sigma D-0632; E-64 Sigma E3132; Eagles MEM with L-Glutamine and Earle's Salts Gibco 31095-052; EDTA Sigma E7889; Gentamicin solution (50 mg/mL) Life technologies 15750-037; Hepes Sigma H-3375; HyClone Sfx-Insect cell culture medium Fisher scientific SH30278LS; Hydrogen peroxide Sigma H1009; Hygromycin B, _Streptomyces_ sp. Calbiochem 400050; Hypoxanthine Sigma H9636-5G; Imidazole Sigma I0125; Instant blue Expedeon ISB1L; Iodoacetamide Sigma I1149; IPTG (Isopropyl β-D-1-thiogalactopyranoside) Sigma I6758; Iscove's Modified Dulbecco's medium Gibco 21980-065; Kanamycin sulfate Sigma K1377; KCl Merck 104936; L-Cysteine Sigma C7352; Magic Mark XP western standards Life technologies LC5603; Mini-PROTEAN® TGX Stain-Free™ Precast Gels Biorad 4568093 or

4568096; NaCl Sigma S5886; Non essential amino acid solution Gibco 11140-068; Novex Tris glycine gels Life technologies EC60285BOX; dPBS Gibco 14190-136; Phleomycin from *Streptomyces verticillus* Sigma P9564; Phorbol 12-myristate 13-acetate (PMA, Sigma); Precision plus protein dual colour standards Biorad 161-0374; QIAprep spin miniprep kit Qiagen 27104; Resazurin Mol probes R12204; Resazurin Sigma R7017; Resorufin Sigma R3257; SeeBlue Plus2 molecular weight markers Life technologies LC5925; Super Optimal broth with Catabolite repression (SOC) medium Life technologies 15544-034; Sodium acetate Ambion AM9740; Sodium pyruvate solution Sigma S8636-100ml; Superoxide dismutase Sigma S5389; SuperSignal West Pico Chemiluminescent Substrate Pierce 34077; Tetracycline hydrochloride Sigma T7660; Thymidine Sigma T1895-1G; Tris/Glycine/SDS 10X electrophoresis buffer Biorad 1610732 or Sigma T7777; Trizma pH 7.6 Sigma T2444; Tween 20 Sigma P5927; Ubiquitin bound to 5-TAMRA or Ub-FP substrate UbiQ bio Ubi-012; XGal Promega V394A; Z-Phe-Arg-AFC Bachem; β mercaptoethanol Sigma M3148.

2.2.2 Biological and Molecular reagents

Anti-myc tag antibody, clone 4A6, alexa fluor® 555 conjugate Merck Millipore 16-225; Anti-polyhistidine-peroxidase monoclonal antibody produced in mouse- clone HIS-1 Sigma A7058; *E. coli* positive control whole cell lysate Abcam ab5395; Fetal Bovine Serum, qualified, US origin (Tet free) Gibco 26140-079; HepG2 cells ATCC HB-8065; MAX Efficiency® DH10Bac™ Competent Cells Invitrogen 10361-012; *Spodoptera frugiperda* pupal ovarian tissue (Sf9) cell line from European Collection of Authenticated Cell Cultures (ECACC) 89070101.

2.2.3 Equipment and materials

1536-well, black, medium-binding plates (Greiner 782076); 1536-well CELLSTAR®, HiBase, tissue culture treated, flat bottom, white microplates (Greiner 782073) with sterile lids (Greiner 656191); 384 well low volume black plates (Greiner 784076); 6

well plate Fisher Scientific TKT-520-030T; Agilent direct drive robot; AKTA pure 25 from GE coupled to a fraction collector F9-C and controlled by Unicorn 6.3 ; Baffled Polycarbonate Erlenmeyer Flask with Vent Cap Corning 431405, 431407, 431401, 431403 and 431253; Beckman Allegra 25R centrifuge; CASY cell counter (Roche diagnostics); Cedex AS20 cell counter (Innovatis); Chemidoc MP (Biorad); disposable hemocytometer (Incyto DHC-N01-5); Echo liquid handling system (Labcyte Inc.); EnVision™ 2104 multilabel reader with laser (Perkin Elmer); Galaxy R CO_2 incubator (R S biotech); Heraeus Cytomat incubator; HiLoad 16/600 Superdex 200 pg (GE healthcare 28989335); His-trap HP 5 mL columns (GE healthcare 17-5248-02); Life technologies iBlot system with gel transfer stacks with PVDR membrane (Life technologies IB401002); Thermo Electron Multidrop Combi nL low volume reagent dispenser; Thermo Electron Multidrop® Combi using a small tube metal tip dispensing cassette (Cat # 24073295); Wallac ViewLux 1430 ultraHTS microplate imager (Perkin Elmer).

2.3 Methods

2.3.1 Protein generation

2.3.1.1 Expression in insect cells

MAX Efficiency® DH10Bac™ competent cells were used to generate the bacmids. 50 µL of DH10Bac cells were thawed on ice for 5 to 10 minutes and 0.75 µg of pFastbac construct DNA was added to these DH10Bac cells and left on ice for 15 minutes. A 45 second heat shock at 42 °C was then given to the cells which were then left on ice for 5 minutes. 900 µL of S.O.C. medium was added and cells were incubated for 4 h at 37 °C. Different dilutions, *viz.*, 1/1000, 1/100, 1/10 and the rest of the cells, were plated on LB agar plates containing 50 µg/mL kanamycin, 10 µg/mL tetracycline, 7 µg/mL gentamicin, 100 µg/mL X-gal and 40 µg/mL IPTG. The plates were incubated at 37 °C for 24 hours and some blue and white colonies

were selected and restreaked on fresh plates to confirm the colour. The bacmid was obtained from these colonies using the Qiagen mini prep kit following the manufacturer's instructions. These bacmid were used to transfect insect cells.

Insect cells (Sf9) were maintained in logarithmic growth phase by maintaining the cell counts between 3 x 10^5 and 7 x 10^6 cells/mL at 27 °C, 100-120 rpm in maximum 1/3 full Erlenmeyer flasks. The cells were maintained in HyClone Sfx-Insect cell culture medium. To transfect the insect cells, 0.45 x 10^6 Sf9 cells in 2 mL of Hyclone medium were seeded per well in a sterile 6-well tissue culture plate (Fisher Scientific TKT-520-030T). Two additional control wells were always kept: cellfectin control (cells + CellFectin II + medium) and cells only control (cells + medium). The plate was incubated 27 °C for 1 h to allow cells to attach. During this incubation period, 6 μL CellFectin II reagent was added to 200μL Hyclone medium. 5 μL bacmid DNA was added to this mixture and incubated for at least 15 min. at room temperature. 800 μL Hyclone medium was added to this mixture. Once the Sf9 cells were properly attached to the surface of the 6 well plates, the medium was removed from the top and the 1mL DNA/CellFectin II mixture was added to the cells. The plates were incubated for 5 h at 27 °C. Then the transfection mixture was removed and cells were washed with 2 mL medium before adding the final 2 mL medium. The plates were placed in a humid environment in a box and incubated for five days at 27 °C. After five days, the supernatant was collected. This was the P0 stock which was used to generate P1.

To make P1 baculovirus, 2% v/v P0 virus was mixed with typically 50 to 75 mL SF9 cells at 1x10^6cells/mL. The cell count and viability was checked every day using Cedex AS20. On the day five the P1s were harvested by centrifuging the culture at 3500 rpm for 20 minutes to pellet the Sf9 cells. The supernatant is the P1 virus stock which was stored at 4 °C protected from light.

1/1000 v/v dilution of P1 virus was used to infect SF9 cells at a cell count of 2x10^6 cells/mL. The cells were monitored for increase in diameter indicating successful

infections. The cells were harvested 48 hours post infection and frozen at -80 °C till used for lysis for protein purification.

2.3.1.2 Cell lysis

The insect cell pellet was resuspended in his trap buffer A at approximately 180 mg cells/mL. Cell lysis was carried out by sonication on ice at 40% amplitude with a pulse of 30 seconds on and 59 seconds off with total pulse on time of 10 minutes. Following this, the lysate was ultracentrifuged for 90 mins at 4 °C at 300,000 g on a Beckman coulter ultracentrifuge XL90 using Type 70 Ti rotor with Beckman Coulter centrifuge tubes number 355618. The supernatant was collected and used to purify *Tbr*DUB1.

2.3.1.3 Affinity purification

The supernatant obtained after lysis of insect cell pellet was subjected to affinity purification using a GE His Trap HP 5 mL column coupled to a GE AKTA pure. The column was first prepared for the purification run according to manufacturer's instructions. This included 5 column volume (CV, 1 CV is 5 mL) wash with 20% ethanol, 5 CV wash with water, 5 CV wash with his trap buffer B without reducing agent and 10 CV wash with his trap buffer A without reducing agent. Following this, before loading the sample, the column was equilibrated with 5 CV of his trap buffer A. UV 280 was monitored throughout the run. The lysate was then applied to the column at a flow rate of 5 mL/min. The column was washed with 19.8 mM imidazole (obtained by 98 : 2 buffer A : buffer B mix) for at least 25 CV. The UV280 readout was monitored and the washes were maintained till the readout became stable indicating that all non specific proteins have washed out. For elution the direction of the flow was reversed to obtain a concentrated protein stock when the protein elutes. A shallow linear gradient followed by a step gradient was used to elute the bound proteins. The gradient used is presented in **Table 2.1**. 1 mL fractions were collected in 96 well plates.

Table 2.1: Elution gradient for affinity purification.
A combination of linear and step gradients was used for cleaning the column and eluting the bout *Tbr*DUB1. (CV-column volume).

% buffer B	Gradient Type	Length of step (CV)	Effective imidazole conc. (mM)
2-4%	Linear	5	19.8-29.6
10 %	Step	10	59
30%	Step	5	157
60%	Step	5	304
100%	Step	5	500

2.3.1.4 Size exclusion chromatography

A GE HiLoad 16/600 Superdex 200 pg column coupled to a GE AKTA pure was used for size exclusion chromatography. The column was first equilibrated with 2 column volumes of the SEC buffer before starting the run. The sample was loaded in to a 10 mL superloop by manual loading. Following the loading of the sample on the column, an isocratic 2 column volume elution at a flow rate of 0.5 mL/min (15 cm/h) was used to obtain high resolution. 1 mL fractions were collected in 96 well plates. UV280 was monitored throughout the run. The purified *Tbr*DUB1 was quantitated using absorbance at 280 nm on a nano-drop using the Bear- Lambert law:

$$A = \varepsilon.l.c$$

Where A is the Absorbance at a given wavelength, ε is the absorption coefficient at a given wavelength (ε=688180 for *Tbr*DUB1 at 280 nm), l is the path length and c is the concentration of the sample (MW= 66 kDa).

2.3.1.5 SDS-PAGE and western blotting

SDS-PAGE was used to resolve proteins in different fractions. For SDS-PAGE 10 or 15 well 4-20% tris-glycine gels were used. For gels to be stained with Instant blue, SeeBlue Plus2 molecular weight markers or Precision plus protein dual colour standards were typically used. For gels used for western blotting, Magic

Mark XP western standards were used. Samples were prepared by diluting appropriately in 4x Laemmli Sample Buffer and then heating them at 95 °C for 10 minutes. Following this, the samples were resolved on the gel. The proteins were first allowed to stack at a constant voltage of 100V till a uniform front was obtained. Then the proteins were resolved at a constant voltage of typically 200V.

For western blotting the iBlot western blotting system from Life technologies was used. We used the ready to use kit following manufacturer's instructions. A 7 minute transfer protocol was used. The membrane was then blocked with 5% non-fat milk for at least 2 hours. This was followed by 3 washes of 5 minutes each with 0.05% tween-20 in PBS (PBST). The membrane was then probed with 1:2000 dilution of anti-his antibody in PBS. The membrane was washed again thrice for 5 minutes each with PBST. SuperSignal west pico chemiluminescent substrate was used to develop the blot and the blot image was obtained using chemiluminescence mode (no illumination, no excitation filter) on Biorad Chemidoc MP.

2.3.2 Assay development

2.3.2.1 Buffer optimisation

To optimize the buffer for maximum assay signal window different additives were tested. Different additives and the highest concentration tested were: EDTA 100 mM, heat-denatured BSA (dBSA 2.2 mg/mL), Pluronic Acid 2.5%, CHAPS 6 mM, Tween-20 60 μM, Tween-80 12 μM, Triton X-100 400 μM and DTT 1 M. Eleven concentrations of each of these additives at ½ dilution were tested. Three different assay buffers were used because a detergent and a reducing agent are essential for enzyme activity. Buffer 1 was 50 mM Tris pH 7.6, 1 mg/mL CHAPS, 2 mM DTT and was used to test EDTA and dBSA. Buffer 2 was 50 mM Tris pH 7.6, 2 mM DTT and was used to test different detergents. Buffer 3 was 50 mM Tris pH 7.6, 1 mg/mL CHAPS which was used to test DTT concentration. The test was

performed in a 384 well, low volume, black plates (Greiner 784076). The plate map is shown in **Figure 2.1**.

Figure 2.1: Plate map for buffer optimisation.
Different additives were tested for their effect on the *Tbr*DUB1 activity, measured by observing the assay window obtained. Each additive was tested at eleven ½ dilutions, in duplicates. For each conc. of additive that was tested, a no enzyme control was also tested.

The final assay volume was 15 µL and the enzyme and substrate concentrations were 1.875 nM and 100 nM respectively. 5 µL each of 3X additive in respective buffer, 3X enzyme and 3X substrate were dispensed in this order to start the reaction. Each concentration was tested in duplicates. No enzyme control, in duplicates, was included for each concentration of each additive. The plates were then read on EnVision™ 2104 multilabel reader with laser. The excitation filter used was 531 ± 25 nm (Perkin Elmer BODIPY TMR FP 531 bar code 105). The emission filters for the s and p channels were 579 ± 25 nm: Perkin Elmer TAMRA FP S-pol 579 bar code 245 and Perkin Elmer TAMRA FP P-pol 579 bar code 246 respectively. The mirror module used was D555fp/D595 (Perkin Elmer BODIPY TMR FP Dual Enh bar code 682). Other settings used for the optics were measurement height 8.91 mm, number of flashes 10, PMT gain 325, second PMT gain 225, G-factor 0.8 and excitation light 60%.

2.3.2.2 Product standard curve

The product standard curve was performed in 1536-well, black, medium-binding plate (Greiner 782076). Stock solution of product (5-TAMRA) at 10 mM and substrate (Ub-FP) at 100 μM were prepared in DMSO. Further dilutions were prepared in the DUB assay buffer and 2.5 μL each of product and substrate dilutions were mixed to obtain different product to substrate ratios. The final substrate and product concentrations ranged between 0 to 100 nM. The plate was read on two different EnVision™ 2104 multilabel readers with laser. The excitation and emission filters and the mirror module used were the same as described above. Other settings used for the optics for the first instrument were measurement height 8.91 mm, number of flashes 10, PMT gain 305, second PMT gain 225, G-factor 0.8 and excitation light 60%. These settings different for the second EnVision were measurement height 11 mm and G-factor 1.

2.3.2.3 Determination of kinetic parameters

The experiment was performed in a 384-well, black, medium-binding plate (Greiner 784076) at a final assay volume of 15 μL. Different concentrations of substrate *viz.* 1, 5, 10, 50, 100, 500 and 1000 nM were tested. 2X working stock for the top concentration was prepared in DUB assay buffer and the 2X stocks for the rest of the concentrations were prepared by serial dilutions. 5 nM (2X) working stock of the enzyme was prepared in DUB assay buffer. 7.5 μL of 2X enzyme and substrate working solutions each were dispensed in quadruples. No enzyme blanks for each concentration of substrate tested were included in duplicates. The plate was centrifuged for 30 seconds to remove any air bubbles and then was read on Envision every 5 minutes for 5 hours. The Envision optics settings were as defined in the previous section except the measurement height which was set to 11 mm.

The polarisation values were obtained in mP and converted to processed substrate at time t, using the following equation,

$$S_t = S_0 - [S_0 * \frac{P_t - P_{min}}{P_{max} - P_{min}}]$$

Where

S_t: substrate processed till time t,

S_0:   initial substrate concentration,

P_t: polarization value at time t (in mP),

P_{min}:   polarization value of 100% processed substrate (50 for free TAMRA),

P_{max}:   polarization value of 100 % unprocessed substrate (~200).

The initial rates obtained by a linear fit from 15 to 60 minutes of the reaction, were plotted against the substrate concentration used and fitted to enzyme kinetics Michaelis-Menten equation using Erithacus GraFit 7.0.2.

2.3.2.4 Selection of enzyme concentration and reaction time

The experiment was performed in a 1536-well, black, medium-binding plate (Greiner 782076) at a final assay volume of 6 µL. 200 nM of 2X working stock of substrate was prepared in the DUB assay buffer from 100 µM stock in DMSO. Different concentrations of enzyme *viz.* 0, 2.5, 5, 10, 15 and 20 nM were tested. 2X working stock for the top concentration was prepared in DUB assay buffer from 4.697 µM stock and the 2X stocks for the rest of the concentrations were prepared by serial dilutions. The experiment was performed in the presence of 1% DMSO which will be the concentration of DMSO in final HTS. 3 µL of 2X enzyme and substrate working solutions each were dispensed in 12 replicates. The final substrate concentration was 100 nM, the recommended concentration by the manufacturer for this class of enzymes. The plate was centrifuged for 1 minute to remove any air bubbles and then was read on Envision every 10 minutes for approximately 8 hours. The Envision optics were configured as defined in the previous section.

2.3.2.5 Selection of final reaction volume, controls and effect of centrifugation

The experiment was performed in a 1536-well, black, medium-binding plate (Greiner 782076) and the different final assay volumes that were tested were 4, 5 and 6 µL. The experiment was performed in the presence of 1% DMSO which is the concentration of DMSO in HTS reactions. The different positive controls tested were a) minus enzyme b) excess iodoacetamide (10 mM) and c) 50 µM H1, which is a dechloro analogue of HBX 41,108 (Colland *et al.*, 2009, de Jong *et al.*, 2012). Final assay concentrations for enzyme and substrate were 5 and 100 nM respectively. To test these conditions an experiment was designed as shown in the plate map in **Figure 2.2**.

Figure 2.2: Plate map for comparing different positive controls.
Each 1536 plate was prepared as shown above for different volumes and the assay performance was compared by comparing three different positive controls and negative (+E) control (384 wells per condition).

The plates were either centrifuged or not for 1 min at 179 g (1000 rpm) using Beckman Allegra 25R centrifuge with TS-5.1-500 rotor to remove air bubbles and then were read on Envision every 18.3 minutes for 6 hours. The settings used for the optics on Envision were: measurement height 8.91 mm, number of flashes 10, PMT gain 305, second PMT gain 225, G-factor 0.8 and excitation light 60%.

2.3.3 HTS

2.3.3.1 Screening at single-concentration and dose response

The primary screening was performed at a single concentration of compound at 10 µM. Dose response studies were performed with 100 µM compound as the typical highest tested concentration followed by ten 1/3 dilutions and hence 1.7 nM was typically the lowest tested concentration for each compound. 50 nL of compounds dissolved in DMSO at 100x the final assay concentration were dispensed in 1536-well plates using Echo liquid handling system. The Echo audits each well of the mother plates and calculates the level of hydration of the DMSO and the depth of the fluid in the well; this survey is used to ensure accurate and precise nanoliter transfer. The Echo also reports empty wells when the volume in the well is not enough or when the compound precipitates into the well and the solution is not homogeneous. There is no contact between the ejection mechanism and the sample, eliminating tips and hazardous waste and cross-contamination issues.

The screening was performed in 1536-well, black, medium-binding plates (Greiner 782076). In all the plates tested, columns 11 and 12, *i.e.* 64 wells were used as negative control or control 1 and columns 35 and 36 *i.e.* 64 wells were used as positive control or control 2. The assay we used is a fluorescence polarization assay. The substrate used was Ubiquitin bound to 5-TAMRA (Ub-FP substrate) (Geurink *et al.*, 2012, Tirat *et al.*, 2005). The assay buffer used to prepare enzyme and substrate dilutions was 50 mM Tris pH 7.6, 50 mM NaCl, 1 mg/mL (1.626 mM) CHAPS, 0.132 mg/mL (2 µM) dBSA, 2 mM DTT, filtered and degassed. All stock solutions of buffer components were prepared in 50 mM Tris pH 7.6. DTT should always be prepared fresh. 50 nL of 1 M iodoacetamide in DMSO was dispensed in positive control wells using Multidrop Combi nL low volume reagent dispenser. Enzyme and substrate dispensations were carried out using Multidrop® Combi using a small tube metal tip dispensing cassette. 2 µL of 2.5x (12.5 nM) enzyme

was dispensed in all wells. The plates were then incubated at room temperature for 30 minutes. Following this 3 µL of 1.67x (166.67 nM) substrate was dispensed in full plate. Hence the final assay concentrations of enzyme and substrate were 5 nM and 100 nM respectively. The reaction was then incubated for 200 mins at room temperature (25 ± 2 °C). The plates were stacked or covered during this incubation to avoid evaporation. The plates were then read on EnVision™ 2104 multilabel reader with laser. The excitation filter used was 531 ± 25 nm (Perkin Elmer BODIPY TMR FP 531 bar code 105). The emission filters for the s and p channels were 579 ± 25 nm: Perkin Elmer TAMRA FP S-pol 579 bar code 245 and Perkin Elmer TAMRA FP P-pol 579 bar code 246 respectively. The mirror module used was D555fp/D595 (Perkin Elmer BODIPY TMR FP Dual Enh bar code 682). Other settings used for the optics were measurement height 8.91 mm, number of flashes 10, PMT gain 305, second PMT gain 225, G-factor 0.8 and excitation light 60%.

2.3.3.2 Analysis of data by IDBS activity base and Spotfire-single concentration and dose response

The data analysis was carried out using IDBS activity base XE. All plates were bar coded and the compound information, which includes concentration and compound ID, were uploaded to activity base when the plates were prepared. Upon running the HTS the data were acquired on activity base XE and the compound ID and concentration data was matched automatically to the data collected depending on the plate barcode and well number, hence maintaining highest standards of data integrity with minimum manual intervention.

The s and p channel data obtained from envision was uploaded to the IDBS activity base XE. All plates from a single batch (60 plates typically) were processed together on activity base XE. The expected outcome of the data analysis is percent inhibition. The collected data from the parallel and the perpendicular channels were used to calculate anisotropy (r) using the equation:

$$Anisotropy\ (r) = \frac{(s - p * G)}{[s + (2 * p * G)]}$$

Where s is the parallel fluorescence intensity, ex 531 nm, em 579 nm

p is the perpendicular fluorescence intensity, ex 531 nm, em 579 nm

G is the G-factor or the grating factor.

FP assays are subject to various types of interference, one of the primary being compound auto-fluorescence/quench which can interfere with fluorescence polarization-based readout. The total fluorescence intensity (I) of the emission beam is used to determine if a compound is auto-fluorescing or auto-quenching, and is expressed as:

$$Total\ Intensity\ (I) = s + (2 * p * G)$$

The anisotropy data hence obtained were used to calculate the normalised response using the equation:

$$Normalised\ response = 100 - \left(100 \times \frac{data - control\ 2}{control\ 1 - control\ 2}\right)$$

Where "control1" corresponds to fluorescence polarization measurements in the absence of an inhibitor (higher mobility, higher depolarization, lower anisotropy values) and

"control2" corresponds to fluorescence polarization measurements in presence of 10 mM Iodoacetamide, which is an inhibitor of DUB activity (lower mobility, lower depolarization, higher anisotropy values).

The quality control we used for the performance of the assay is Z' (Zhang *et al.,* 1999). Z' accounts for the standard deviation of both the positive and negative controls and the assay window. It is calculated as:

$$Z' = 1 - \left(\frac{3 \times (\sigma_{c+} + \sigma_{c-})}{|\mu_{c+} - \mu_{c-}|} \right)$$

Where σ_{c+} and σ_{c-} are the standard deviations for the positive and negative control populations respectively and μ_{c+} and μ_{c-} are the robust means of positive and negative control populations respectively.

Since Z' is calculated using the standard deviations and means of control populations, it is not affected by the test compounds and hence is a good measure of the assay performance when screening a diverse chemical library (Zhang *et al.*, 1999). We typically use Z' ≥ 0.4 as the quality control cut-off for the performance of an assay. Activity base XE calculates the Z', signal to background ratio, robust means for the negative and positive control populations and the response as mentioned in the equation above. Robust statistics are used to diminish the impact of outliers, which in our case are the active compounds. The hit cut-off is calculated as

$$Hit\ cutoff = Mean + (3 * SD)$$

Compounds showing response greater than or equal to the hit cut-off were marked as hits.

For dose response studies, as mentioned above, 100 µM was the highest concentration tested followed by ten 1:3 dilutions. As for single shot studies, for analysing the dose response data, the s and p channel data were uploaded to the IDBS activity base XE. The FP, normalisation and Z' calculations were the same as described above. The expected outcome of this analysis is pIC_{50}.

The IC_{50} values are calculated by fitting the normalised data, calculated as described above, to a four parameter logistic fit equation:

$$y = \frac{A - D}{(1 + (x/C)^B)} + D$$

The equation can be rewritten as:

$$y = \frac{Range}{(1 + (x/C)^B)} + Background$$

In both the equations above, x is the concentration of the test compound, y is the inhibition or normalised response (%), A is the maximum signal, D is the minimum signal (background), C is the IC_{50} and B is the Hill slope. IC_{50} is the concentration of the compound at which it shows half (50%) of the maximum response observed.

Expressing IC_{50} values in molar units, the pIC_{50} values are calculated as:

$$pIC_{50} = \log\left(\frac{1}{IC_{50}}\right)$$

For validating the fit to a four parameter logistic equation the following parameters were defined. The threshold above which points were excluded automatically from the data analysis was set to 300. The software was allowed to remove a maximum of 2 points per curve. At least 8 data points were required to fit a curve. A compound was marked inactive if the maximum response was less than 30 %. If all observed points were below 50 % (but at least some points above 30 %) response, the compound was marked as poorly active. Additionally, the compound was labelled poorly active if a) the curve fit failed due to XC_{50} confidence limit (CL) b) the R square was > 0.8 defining a concentration dependent effect, c) the fitted $pXC_{50} < 3^{rd}$ highest concentration tested. A compound was labelled highly active if the minimum observed response was greater than 50%.

For limiting the curve fitting algorithm the minimum response was constrained to be between -20 % and 20 % and the maximum response was constrained to be between 80 % and 120 %. The hill slope was constrained between 0.5 and 5. Lastly, the XC_{50} 95 % confidence limit was set to 10. This meant that curve fit will fail QC testing if the ratio of the XC_{50} upper 95% confidence limit divided by X_{C50} lower 95% confidence limit is greater than 10. Finally, despite these settings there are always cases of points that fail to fit. These curves were manually checked and some outlier points were excluded to obtain curve fits.

2.3.3.3 Pattern recognition and correction

For the single concentration data, the patterns in the plates were identified and removed (Coma *et al.*, 2009). The smoothing algorithm for pattern correction uses a smoothed running median in 2D to calculate a pattern that it is then subtracted from the response (Coma *et al.*, 2009). Additionally, the data were corrected for region dependent hit rates using a robust algorithm to mark hits (ANALYTICAL METHODS COMMITTEE, 1989) using all the samples in a well position (and alternatively the surrounding wells) for each well; in this way, the cut-off adapts to the inhomogeneities of the average response across the well positions. By using this "transverse-cut-off" approach the pattern disappeared. The calculations were performed using a C program written to this effect.

2.3.4 *Sans* DTT assay

The assay and data analysis is exactly the same as described in section 2.3.3 except that DTT was not included in the assay buffer. Hence the assay buffer used was 50 mM Tris pH 7.6, 50 mM NaCl, 1 mg/mL (1.626 mM) CHAPS, 0.132 mg/mL (2 µM) dBSA, filtered and degassed.

2.3.5 SOD/Catalase assay

The assay and data analysis is exactly the same as described in section 2.3.3 except that superoxide dismutase and catalase were added to the assay buffer. Hence the assay buffer used was 50 mM Tris pH 7.6, 50 mM NaCl, 1 mg/mL (1.626 mM) CHAPS, 0.132 mg/mL (2 µM) dBSA, 10 µg/mL SOD, 500 µg/mL catalase, 2 mM DTT, filtered and degassed.

2.3.6 L-Cys assay

The assay and data analysis is exactly the same as described in section 2.3.3 except that DTT was replaced by L-cysteine in the assay buffer. Hence the assay buffer used was 50 mM Tris pH 7.6, 50 mM NaCl, 1 mg/mL (1.626 mM) CHAPS, 0.132 mg/mL (2 µM) dBSA, 2 mM L-Cys, filtered and degassed.

2.3.7 Resazurin assay

The plate preparation was the same as described in section. The assay buffer was 50mM Hepes, 50mM KCl, pH 7.5. 1 mM stocks of resazurin and resorufin and 1 M stock of DTT were prepared in assay buffer. 5 µL of 5 µM resazurin and 2 mM DTT in assay buffer was dispensed in full plate except positive control wells (columns 35 and 36). For positive control 5 µL of 5 µM resorufin and 2 mM DTT in assay buffer was dispensed. The plate was then incubated at room temperature for 1 hour protected from light. The plate was then read on EnVision™ 2104 multilabel reader with laser. The excitation and emission filters used were 535 ± 40 nm (Perkin Elmer Resorufine/Amplex Red FP 535 bar code 124) and 595 ± 60 nm (Perkin Elmer Cy3 595 bar code 229) respectively. The mirror module used was D555 (Perkin Elmer Bodipy TMR bar code 405). Other settings used for the optics were measurement height 10 mm, number of flashes 1, PMT gain 1 and excitation light 1%. High concentration mode was enabled to avoid saturation of optics.

The fluorescence intensity data hence obtained were used to calculate normalised response using IDBS activity base XE as described above and the curve fitting and pIC_{50} calculations were done as described above.

2.3.8 hCTSB assay

The assay was carried out in 384 well low volume black plates (Greiner 784076). 100 nL of compounds dissolved in DMSO at 100x the final assay concentration

were dispensed in 384 well plates using Echo liquid handling system. The assay buffer was 100 mM sodium acetate, 150 mM NaCl, 5 mM EDTA, 1 mM CHAPS, 40 µM L-cys, pH 5.5. A stock solution of Cathepsin B (14 µM, obtained from GSK in-house repository of reagents, expressed and purified in-house) was diluted in assay buffer to give a conc. of 150 pM. The substrate used was Z-Phe-Arg-AFC. A substrate stock of 20mM (M.W. 710) was prepared in neat DMSO. This was then diluted to 50 µM in assay buffer. In 384 well format column 6 (16 wells) was used for negative control while column 18 (16 wells) was used for positive control. 5 µL of 150 pM CTSB was dispensed to all wells except to column 18. 5 µL of assay buffer was dispensed to positive control wells. The plate was then incubated for 30 minutes at room temperature. Then 5 µL of 50 µM substrate was dispensed to all wells. The reaction was incubated at room temperature for 2 hours after which it was stopped with 2 µL of 100 µM E64 in 25% DMSO. The plates were then read on EnVision™ 2104 multilabel reader with laser. The excitation and emission filters used were 380 ± 10 nm (Perkin Elmer Fura2, BFP 380 bar code 112) and 495 ± 10 nm (Perkin Elmer TRF Emission 495 bar code 276) respectively. The mirror module used was D425 (Perkin Elmer Beta-Lactamase bar code 418). Other settings used for the optics were measurement height 9.4 mm, number of flashes 10, PMT gain 271 and excitation light 10%.

The fluorescence intensity data hence obtained were used to calculate normalised response using IDBS activity base XE as described above and the curve fitting and pIC_{50} calculations were done as described above.

2.3.9 *T. brucei* cell culture

2T1 *Trypanosoma brucei brucei* cells derived from strain s427 were modified to integrate the *Tbr*DUB1 gene under the control of a tetracycline operator in the ribosomal RNA locus on chromosome 2a (Jones *et al.*, 2014, Alsford and Horn, 2008) (received from Dr. Boris Rodenko). This allows tetracycline induced overexpression of the myc$_6$*Tbr*DUB1 gene product. The selection markers used to

maintain this cell line are Hygromycin (2.5 µg/mL) and Phleomycin (0.5 µg/mL). The cells were maintained in a continuous culture in T-flasks at 37 °C, 100% humidity, 5% CO_2 in a Heraeus Cytomat incubator. The cell density was maintained below the stationary phase ($2x10^6$ cells/mL) by diluting them in fresh pre-warmed (to 37 °C) *T. brucei* complete culture medium. The cells can be maintained in exponential growth phase by maintaining the cell count between $2x10^5$ and $2x10^6$ cells/mL with typically 3 passages per week. Cells were counted using either CASY cell counter or disposable hemocytometer.

2.3.10 Overexpression of *Tbr*DUB1 in *T. brucei* and cell lysis

Overexpression of *Tbr*DUB1 was induced by adding 1 µg/mL tetracycline from a 1000x stock of tetracycline in 70% ethanol. Samples were taken just before screening, and the after another 24 hours which marks the end of screening. $2X10^7$ cells were pelleted by centrifugation at 800 g, 10 minutes at 4 °C. The cells were given a quick wash in PBS. Following this the cells were lysed in 1X SDS loading buffer to obtain lysate equivalent to $2x10^5$ cells/µL. $4x10^6$ cell equivalent lysate and *E. coli* positive control whole cell lysate (Abcam ab5395, serves as positive control for anti-myc antibody) were loaded on a Biorad mini protean TGX stain free precast gel and resolved using a constant voltage of typically 200 V. Following this, the gel was imaged in stain-free mode on Biorad ChemidocMP. This gel was then used for western transfer using the dry transfer technique with Invitrogen iBlot kit and instrument. The transfer time used was 7 minutes. The blot was then incubated with 5% fat free milk in PBS at room temperature for at least 1 h, followed by 3 washes of 5 minutes each with 0.05% PBST. The blot was then incubated with 1:200 dilution in PBS of anti-myc tag antibody, clone 4A6, alexa fluor® 555 conjugate for 1 hour at room temperature. This was followed by 4 washes of 5 minutes each with 0.05% PBST. The western blot was then imaged on Biorad ChemidocMP using green epi illumination and 605 ± 50 emission filter. Also, a multichannel image of the western blot was taken combining a stain free image of

the blot and alexa 555 signal. This allows us to compare the total protein in a lane, across the lanes and acts a loading control.

2.3.11 *T. brucei* whole cell screening

We tested the effect of all *Tbr*DUB1 hits on the parasite using an ATP based viability assay described in ([Sykes and Avery, 2009](#)) and used as an orthogonal assay in ([Pena *et al.*, 2015](#)). The screening was performed in a 1536-well, CELLSTAR®, HiBase, tissue culture treated, flat bottom, white microplates (Greiner 782073) covered with sterile lids (Greiner 656191). The highest compound concentration tested was 41.67 µM, followed by ten 1/3 dilutions. 25 nL of 240x compound stock in DMSO was dispensed in the plates using Echo liquid handling system. As usual, columns 11 and 12 of each plate were used as negative controls and columns 35 and 36 as positive controls. The negative control we used was cells growing in the absence of any compound. For positive control we used complete culture medium without any cells. To set up the screening, the cells were diluted to 1000 cells/mL and maintained in presence of selection markers for 72 hours. In case of induction of overexpression 1 µg/mL tetracycline was also added. The cells were further maintained in log-phase of growth (seed at 1×10^5 cells/mL) for at least 24 hours in the absence of selection markers before the screening although tetracycline was maintained in the case of induction. At this point a sample of cells (pre-screening) was taken to check for *Tbr*DUB1 overexpression as described in the previous section. For screening, all reagents dispensations were carried out using Multidrop® Combi using a small tube metal tip dispensing cassette. 6 µL of fresh complete culture medium were dispensed to positive control wells. 6 µL of 50,000 cells/mL (300 cells/well) were dispensed in the plate except to positive control wells. The plates were covered with sterile lids and incubated for 24 hours at 37 °C, 100% humidity, 5% CO_2 in a Galaxy R CO_2 incubator. The parasite load in the wells was determined by measuring the ATP content in each well using 4 µL of CellTitre-Glo luminescent cell viability assay reagent. The plate was then allowed to incubate at room temperature for 60 minutes to allow the

signal to stabilize. The plates were then read on Wallac ViewLux 1430 ultraHTS microplate imager. Settings used for the optics were: exposure time 15 s, readout speed 10 µs, readout gain 50X, image binning 3X and emission filter 613 ± 55.

The chemiluminescence data hence obtained were used to calculate normalised response using IDBS activity base XE as described above and the curve fitting and pIC_{50} calculations were done as described above, except that hill slope was constrained between 0.5 and 10 and the threshold to exclude a data point was set very high so as to avoid loss of data points. These were manually checked later.

2.3.12 *In silico* clustering of compounds

For clustering similar compounds, a complete link agglomerative hierarchical clustering was performed (Leach and Gillet, 2007). Agglomerative hierarchical clustering method is the one that starts by considering each compound a different cluster and then merges most similar clusters to form bigger clusters. The algorithm continues merging similar clusters together till no pairs of clusters are more similar than the defined cut-off. The complete link method of clustering is the one in which the similarity between clusters is defined as the minimum similarity between pairs of compounds in two different clusters. The similarity between compounds was defined by the tanimoto similarity between the fingerprints of compounds. The cut-off for tanimoto similarity was defined to be 0.55 to perform this clustering.

2.3.13 HepG2 cytotoxicity assay

The assay was carried out in 384 well low volume black plates (Greiner 781091). 250 nL of compounds dissolved in DMSO at 100x the final assay concentration were dispensed in 384 well plates using Echo liquid handling system. Cell culture medium was Eagles MEM with L-Glutamine and Earle's Salts, 10% FBS, 1% of 100X non essential amino acid solution. 100 µM digitoxin in DMSO was used to

inhibit growth in positive control wells. Actively growing HepG2 cells were removed from a T-175 flask using culture medium (5mL), and made as disperse as possible by repeated pipetting. Cell suspension was added to 500 mL culture medium + 1% Penicillin + Streptomycin. 25 µL were dispensed into all wells of Greiner black 384-well plates using a Multidrop. Seeding density should be checked to ensure that new monolayers are not more than ~50% confluent at the time of seeding (typically 3000 cells per well), before completing preparation of plates. Cells are left at 37 °C, 5% CO2 in a humidified incubator in the presence of compound for 48 hours. ATP content in each well was measured using 25 µL of CellTitre-Glo luminescent cell viability assay reagent. The plates were centrifuged and then allowed to incubate at room temperature for 10 minutes to allow the signal to stabilize. The plates were then read on Wallac ViewLux 1430 ultraHTS microplate imager. Settings used for the optics were: exposure time 5 s, readout speed 10 µs, readout gain 50X, image binning 2X and emission filter 613±55.

The chemiluminescence data hence obtained were used to calculate normalised response using IDBS activity base XE as described above and the curve fitting and pIC_{50} calculations were done as described above, except that a) the threshold to exclude a data point was set very high so as to avoid loss of data points b) the threshold to mark poorly active compounds was set to 60% c) the threshold for marking inactives was set to 60% and d) the hill slope was constrained between 0.5 and 6.

2.3.14 Intra-macrophage *L. donovani* assay

The assay is described in detail in Peña *et al.* (Pena *et al.*, 2015) and adapted from Rycker *et al.* (De Rycker *et al.*, 2013). THP-1 cells differentiated to macrophage-like cells using phorbol 12-myristate 13-acetate (PMA) were used. The cells were maintained at 37 °C, 5% CO_2 for 24h and cell differentiation was confirmed using an optical microscope. After washing, the cells were infected using eGFP LdBOB axenic amastigotes at a multiplicity of infection of 10 (*i.e.* 6 x 10^6 parasites/mL).

After overnight incubation, the remaining parasites were removed by 3x washes with sterile D-PBS. The infected cells were harvested using 0.05% (w/v) trypsin and 0.48 mM EDTA for 5 min. An aliquot was fixed and counted in a CASY cell counter. Cells were diluted to a final concentration of 1.6×10^5 cells/mL in assay medium consisting of RPMI (InvitrogenTM), 2% FBS (Gibco), 25 mM sodium bicarbonate (InvitrogenTM) and 30 nM of PMA. Compound plates were prepared by dispensing 250 nL of compound in DMSO and for a dose response eleven point one in three dilutions with highest concentration of 50 µM were tested. The cells were plated onto assay plates containing compounds (3000 cells/well, 50 µL) using a Multidrop Combi in all columns except the positive control. For positive control, Amphotericin B was added to an aliquot of infected cells to a final concentration of 2 mM and dispensed to the control column using a Multidrop Combi. Plates were incubated for 96 h at 37 °C, 5% CO2 and then fixed with 4% (v/v) formaldehyde-PBS for 30 min at room temperature. After fixation, the wells were washed twice with 100 mL PBS using an EL406 multi well plate washer (BioTek), stained with 10 mg/mL DAPI in PBS with 0.1% (v/v) Triton X-100 for 30 min at room temperature, and lastly washed twice with 50 mL PBS. Finally, 50 mL of PBS was added to each well and the plates were sealed. The plates were read on a high-content microscope (Opera QEHS) using a 20x objective, 3 fields per well. Two exposure images were taken for each well using 405 nm and 488 nm laser excitation. Automated image analysis was performed with a script developed on AcapellaH High Content Imaging and Analysis Software (PerkinElmer). THP-1 cell count, average number of amastigotes per macrophage and percent of infected cells were calculated for each well.

2.3.15 *T. cruzi* intracellular imaging assay

The assay is described in detail in Peña *et al.* (Pena *et al.*, 2015). It has been adapted from Alonso-Padilla *et al.* (Alonso-Padilla *et al.*, 2015) . Briefly, H9c2 cells were seeded in T-225 flasks in DMEM-10% FBS for 4 h to allow attachment. Cells were then washed once with PBS before infection. *T. cruzi* trypomastigotes were

collected at days 5 to 8 post infection from LLC-MK2 parasite infected cultures by centrifuging the culture at 2500 rpm for 10 minutes at room temperature followed by allowing the trypomastigotes to swim out for 4 h at 37 °C. These were then collected and counted in a CASY Cell Counter. Trypomastigotes, in supplemented DMEM, were added to H9c2 cultures in a multiplicity of infection of 1 and incubated for 18 h. Cells were washed once with PBS. The infected cells were harvested by trypsin treatment. Cells were counted in a CASY Cell Counter using a 150 mm and diluted to 5 x 10^4 cells per mL in supplemented assay DMEM. 50 μL of this cell preparation was dispensed per well into 384-well poly-lysine coated assay plates using a Multidrop Combi. The plates were incubated at 37 °C, 5% CO2 for 72 h. Cultures were then fixed and stained by addition of 50 μL of a solution containing 8% formaldehyde and 4 μM DRAQ5 DNA dye (BioStatus, UK) per well. Plates were incubated for 1h at room temperature and imaged using a Perkin-Elmer Opera microscope with 20x air objective (NA 0.4) using a 635 nm laser excitation and a 690/50 emission detection filter for DRAQ5 detection. Five images were collected per well for reliable statistical analysis. Automated image analysis was performed with a script developed on AcapellaH High Content Imaging and Analysis Software (PerkinElmer). Three outputs were reported for each well: (1) number of host cells nuclei to determine drug-related cytotoxicity; (2) number of amastigotes per cell as infection level measurement; and (3) percentage of infected cells per well as a second infection marker.

Results

Chapter 3 Purification of *Tbr*DUB1 and assay development

Abstract

*Tbr*DUB1 is a validated target in *T. brucei*. This chapter summarises the purification and characterisation of recombinantly expressed *Tbr*DUB1 and the development and validation of a reproducible and robust assay for screening the GSK compound collection to identify inhibitors of *Tbr*DUB1. Screening a small set of 9838 compounds that sample HTS library gave a hit rate of 0.26% with a confirmation rate of 69%.

3.1 Protein expression and purification

3.1.1 Expression of *Tbr*DUB1 in insect cells (Sf9)

After several unsuccessful attempts to express soluble codon-optimized *Tbr*DUB1 in *E. coli*, we utilized insect cells (Sf9) to express recombinant *Tbr*DUB1, as the feasibility of this system to express soluble recombinant *Tbr*DUB1 was previously demonstrated (*Dr. Boris Rodenko, oral communication*). The full-length *Tbr*DUB1 gene was cloned in the pFastBac-Htb vector, with N-terminal His-tag (the construct was received from Dr. Boris Rodenko). This construct was used to generate the bacmid by transforming competent DH10Bac *E. coli* cells. The detailed method is described in section 2.3.1.1. Following this, the bacmid was used to transfect insect cells (Sf9) maintained in a logarithmic growth phase, to obtain the first generation of baculovirus- the P0, in replicates. These P0 baculovirus were used to infect a larger batch of Sf9 cells to amplify the amount of the virus (P1 generation). These P1 stocks were then tested in multiplicity of infection (MOI) studies to identify the best conditions for the expression of the protein. For this, Sf9 cells were infected with two different P1 baculovirus stocks (named Bac1 and NKIBac) each at a dilution of 1/500 and 1/1000, and the levels of protein expression were analysed at 24, 48 and 72 h post-infection (PI) by western blotting. Highest protein expression level was observed using 1/1000 dilution of NKIBac 48h PI (**Figure 3.1**).

3.1.2 Purification of *Tbr*DUB1

Using the optimised expression conditions, *Tbr*DUB1 was produced in large scale cultures (5 L) of Sf9 cells and purified via immobilized metal ion affinity chromatography (IMAC) followed by size exclusion chromatography (SEC).
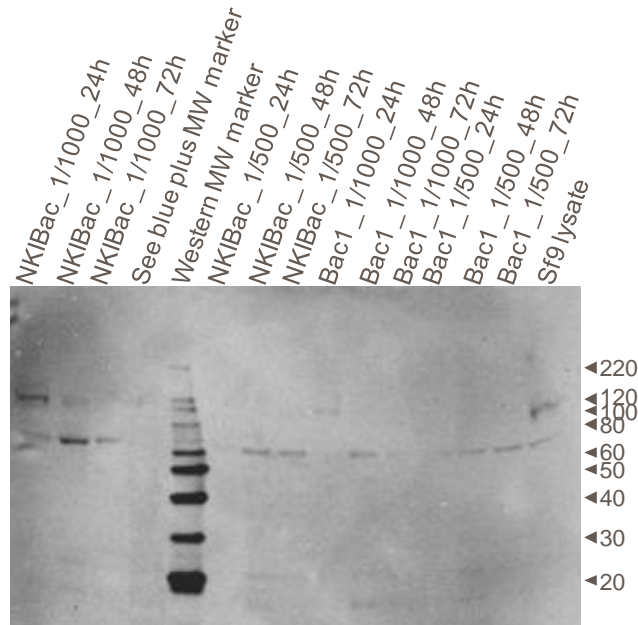
Figure 3.1: Determination of optimum conditions for expression of *Tbr*DUB1.
Different concentrations of two P1 stocks were checked for expression of protein at indicated time points. Shown here is an immunoblot probed with anti-his antibodies. Lanes are named as Bacmid name_dilution_time post-infection. NKIBac at 1/1000 dilution at 48 h PI showed highest expression levels of *Tbr*DUB1 (lane 2). The molecular weight is expressed in kDa.

Briefly, insect cell pellet was resuspended in lysis buffer and cell lysis was carried out by sonication. Following removal of cell debris by ultracentrifugation, the supernatant was subjected to affinity purification using a Ni^{2+} sepharose IMAC column coupled to a FPLC system (GE AKTA pure). A shallow linear gradient followed by a step gradient of imidazole was used to elute the bound proteins (**Figure 3.2(A)**), and 1 mL fractions were collected in 96 well plates. The collected gradient fractions were then analysed by SDS-PAGE and western blotting using anti-his antibodies to identify *Tbr*DUB1-containing fractions and assess their purity. As seen in **Figure 3.2(B)**, several fractions showed a band at the expected molecular weight of 66 kDa. and were of higher relative purity compared to crude cell lysate (lane 1 in **Figure 3.2(B)**). In order to increase the purity of *Tbr*DUB1 the fractions were pooled and further resolved using SEC.
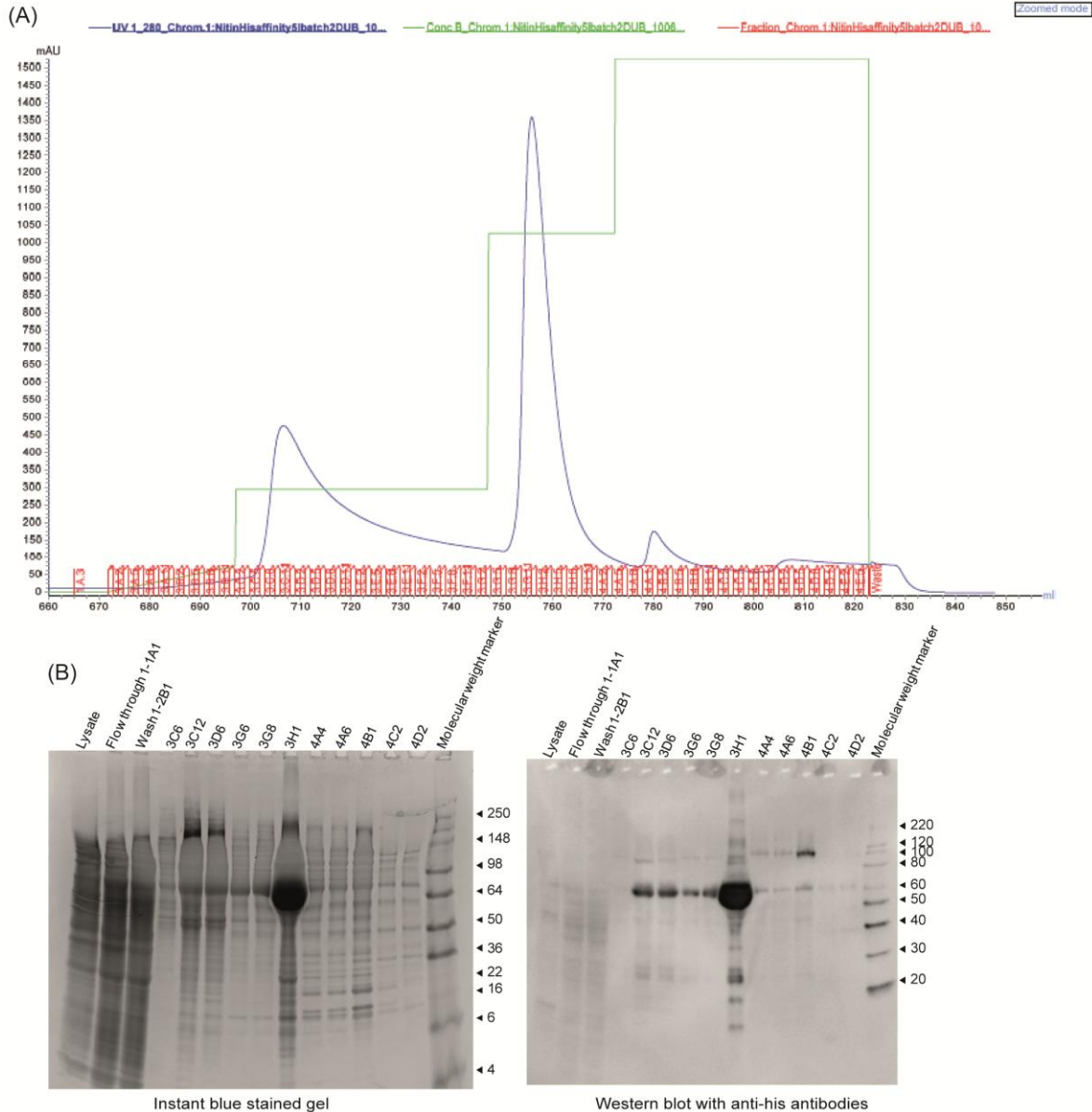
Figure 3.2 Affinity purification of *Tbr*DUB1.
(A) The chromatogram for the affinity purification. Shown is the UV280 chromatogram of the eluted proteins (blue) and the gradient of the elution buffer containing imidazole (green). 1 mL fractions in four 96 well plates were collected. Each fraction is named as plate number followed by well number. For example, first well of the first plate is 1A1. The fractions are shown in red. (B) SDS-PAGE gel and western blot of selected fractions, as indicated above respective lanes, obtained after affinity purification. Fractions showing an intense band at around 66 kDa were pooled and subjected to size exclusion chromatography. Molecular weight is expressed in kDa.

The pooled fractions were concentrated using a regenerated cellulose membrane with 50 kDa. cut-off (Millipore product code: UFC905024), by centrifugation. For SEC, a column with cross-linked agarose and dextran (GE HiLoad 16/600 Superdex 200 pg column) coupled to a FPLC system (GE AKTA pure) was used with a flow rate of 0.5 mL/min (15 cm/h). An isocratic 2 column volume elution was used. Shown in **Figure 3.3(A)** is the UV280 chromatogram for the SEC run of two different batches. The chromatograms show that the run is reproducible. A western blot analysis (**Figure 3.3(B)**) showed the *Tbr*DUB1 band at the expected molecular weight with higher purity than those obtained after affinity purification. Finally, SEC fractions containing *Tbr*DUB1 were combined (**Figure 3.3(C)**) and yielded a total of 10.85 mg *Tbr*DUB1 at a concentration of 310 µg/mL (4.697 µM) and a purity of 75%, from five litre cell culture *i.e.* ~2 mg *Tbr*DUB1 per litre of cell culture.

In summary, protein expression in Sf9 cells and the described two-step purification yielded a sufficient quantity of high-purity *Tbr*DUB1 for the development of a screening assay and HTS purposes.

3.2 *Tbr*DUB1 sequence homology

The sequence of *Tbr*DUB1 in presented in Figure 3.4. *Tbr*DUB1 is identified in Uniprot database with id Q386W6 and in TriTrypDB (a Kinetoplastid genomics resource) with id number Tb927.11.1930. A blast was performed to identify the closest orthologues in *T. cruzi*, *Leishmania* spp. and humans. The *Tbr*DUB1 shows 48%, 38.6% and 22.1% identity with the closest orthologues in *T. cruzi*, *L. infantum* and humans respectively. *T. brucei*, *T. cruzi* and *Leishmania* have similar genomic sequence and biology (El-Sayed *et al.*, 2005). It has been shown that compounds targeting a conserved target could be active against all three parasites (Khare *et al.*, 2016). Since the target DUB shows higher similarity across these species compared to the closest human orthologue, it is likely that the inhibitors identified against one species would be efficacious against the others as well.
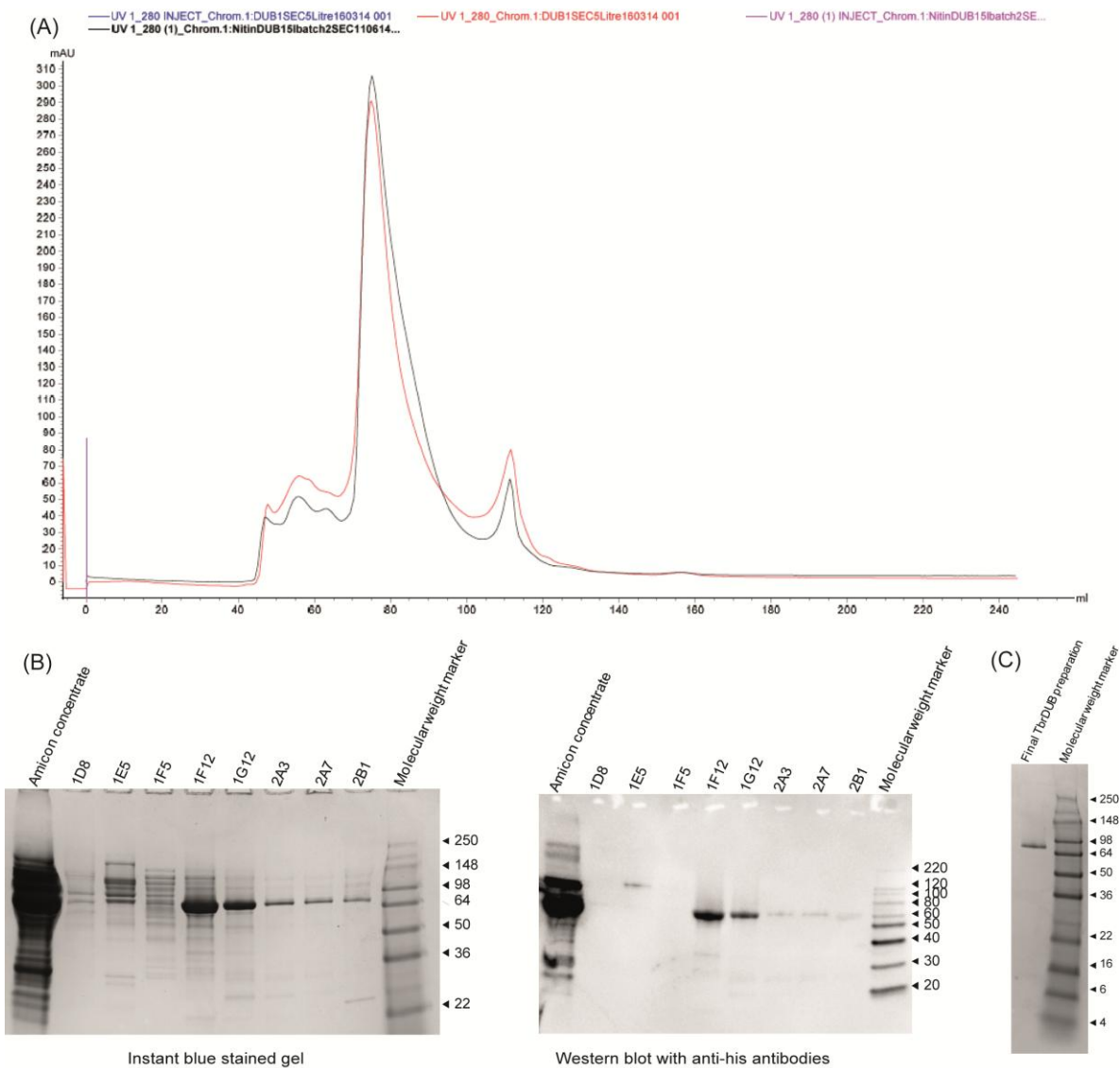
Figure 3.3: Size exclusion chromatography to purify *Tbr*DUB1.
(A) UV280 chromatogram of eluted proteins during size exclusion chromatography. Shown above are the chromatograms from 2 different runs, which show high reproducibility. The injection and elution phase for the first run are shown in blue and red respectively while for the second run are shown in purple and black respectively. (B) SDS-PAGE and western blot of fractions representative of different peaks. A strong signal with anti-his antibodies at the expected molecular weight indicates purification of *Tbr*DUB1. The percent purity improved after gel filtration. Fractions showing *Tbr*DUB1 were pooled and the final preparation of *Tbr*DUB1 shows high purity as seen in a coomassie stained gel (C). Molecular weight is expressed in kDa.
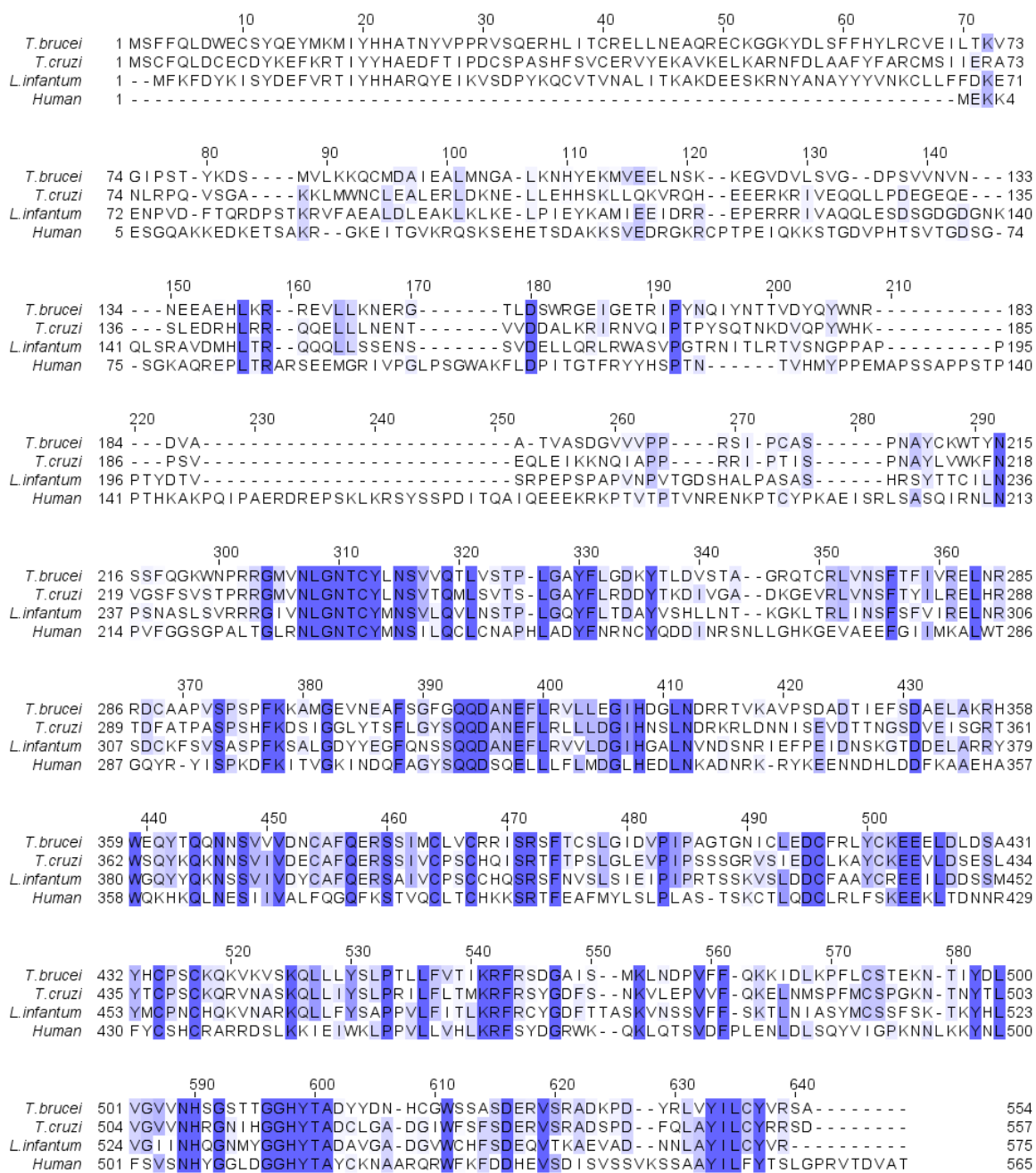
Figure 3.4: Multiple sequence alignment of *Tbr*DUB1 orthologues of *T. cruzi*, *L. infantum* and *Homo sapiens*.
*Tbr*DUB1 shows 48%, 38.6% and 22.1% identity with the closest orthologues in *T. cruzi*, *L. infantum* and humans, respectively

3.3 Enzyme characterisation and assay development for HTS

Assay development is an iterative process that can be very challenging as numerous variables are involved. It typically includes optimisation of the assay buffer, the plate reader optics, the enzyme and substrate concentration, the time of reaction, enzyme kinetics parameters, reaction volume, controls etc. The biochemical assay we optimised for the HTS is a fluorescence polarization (FP) assay. In FP assays the excitation light is plane polarized. Polarization of the emitted light depends on the fluorophore lifetime (τ) and the rotational relaxation time (μ). μ is the time an excited molecule takes to rotate through an angle of 68.5° and is defined by the equation

$$\mu = \frac{3\eta V}{RT}$$

Where η is viscosity of the solvent, V is the molar volume, R is the universal gas constant and T is the temperature. FP (expressed in mP) is proportional to the rotation relaxation time which in turn is proportional to the size/volume of the fluorescent molecule. A large molecule rotates slow and therefore emits light in the same plane as the excitation light, yielding a high FP value. In contrast, a small molecule rotates fast and emits depolarized light (in a different plane from the excitation light), leading to a low FP value (**Figure 3.5(B)**). The FP readout is independent of the concentration of the fluorophore.

The DUB substrate Ub-FP is a 5-carboxytetramethylrhodamine (5-TAMRA) modified Lys-Gly sequence that is linked to a ubiquitin via a native isopeptide bond with the lysine side-chain (5 -TAMRA-Lys(Ub)-Gly-OH) (Geurink *et al.,* 2012, Tirat *et al.,* 2005). An active DUB cleaves the isopeptide linkage to release 5-TAMRA which is significantly smaller than the substrate. At the time of setting up the assay, this was the only available DUB substrate with an isopeptide linkage. Since naturally DUBs cleave isopeptide linkages, this substrate has a higher

physiological relevance than others where the DUBs have to cleave a peptide linkage. The next sections describe the optimisations of assay conditions that were performed in order to obtain a miniaturised, robust and reproducible high-throughput DUB activity assay.
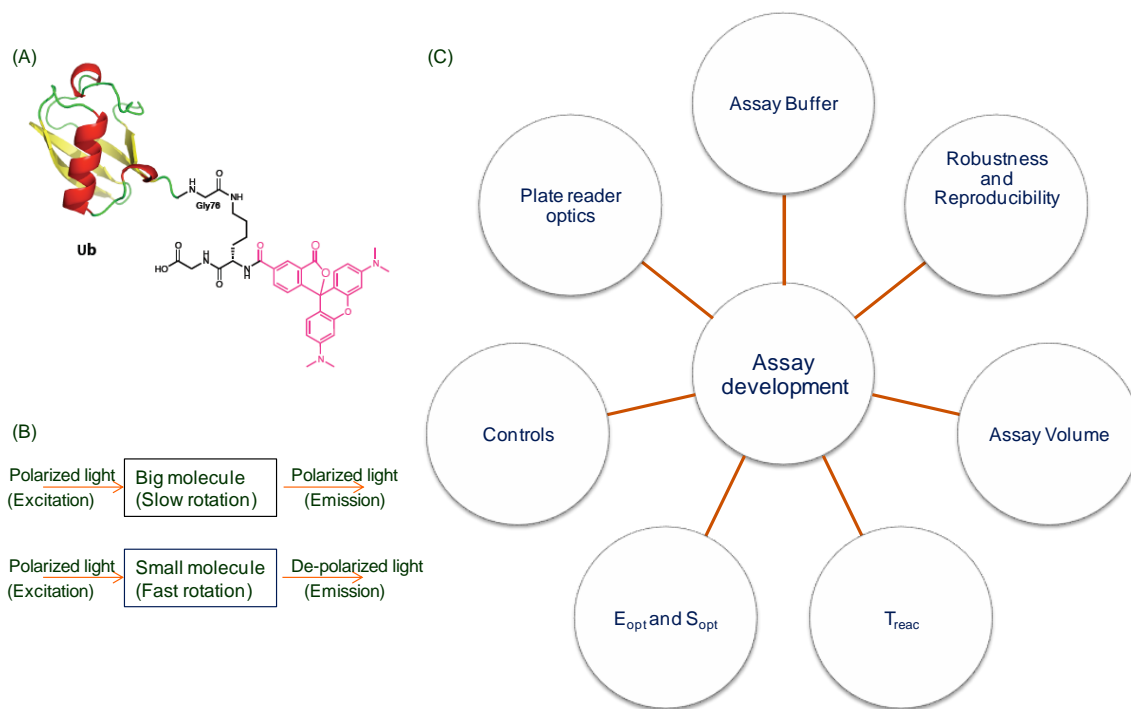
Figure 3.5 Assay principle and different assay parameters that need to be optimised.
(A) Structure of Ub-FP substrate. The substrate is 5-TAMRA modified Lys-Gly sequence that is linked to a ubiquitin via a native isopeptide bond with the lysine side-chain. An active DUB will cleave the isopeptide bond and release free 5-TAMRA. Since 5-TAMRA is smaller in size compared to the substrate, it depolarizes the light more (B) and this difference can be measured to determine the activity. (C) Overview of parameters that were optimised during the development of the DUB activity assay

3.3.1 Buffer optimisation

For any assay it is critical to optimise the reaction buffer to ensure optimal enzyme activity. For a high throughput DUB assay, the assay buffer should contain a buffering agent, a reducing agent, an electrolyte, a detergent and a carrier protein. The buffering agent is required to maintain the pH of the reaction mix, the reducing agent to maintain the DUB in a reduced state and the electrolyte, detergent and the

carrier protein to minimize variability caused by interactions with the tubing and plastic material that is used for the screening.

For an optimal composition of the *Tbr*DUB1 activity assay buffer, we first determined the components that are essential for enzyme activity, which were found to be the detergent CHAPS with DTT as reducing agent (data not shown). We then tested multiple buffer additives at different concentrations (**Figure 2.1**) in order to obtain a large enough assay window for a robust assay. Assay window is the difference between the signal of the positive and negative controls.

To evaluate the effect of additives, enzyme activity was normalised to the activity observed in the assay buffer recommended by the commercial suppliers of the substrate (50 mM Tris pH 7.6, 100 mM NaCl, 1 mg/mL (1.6 mM) CHAPS, 0.5 mg/mL (3.33 µM) bovine gamma globulin (BGG), 2 mM DTT). This gave us the percent remaining activity. The results are presented in **Figure 3.6**. EDTA had no effect on enzyme activity. 2 µM heat-denatured BSA (dBSA) performed as well as 3.33 µM BGG. Thus BGG was replaced with dBSA because BSA is cheaper than bovine gamma globulin and more importantly dBSA overcomes the compounds masking effects of BSA (Pena and Dominguez, 2010). None of the detergents that were tested gave a significantly better performance than CHAPS. The optimum concentration of DTT is 2-4 mM and we decided to use 2 mM in the final assay buffer. It is important to include an electrolyte in the assay buffer to reduce the interaction with plastic material of the assay plate. Since reducing the concentration of NaCl did not affect enzyme activity, its concentration was halved to 50 mM.

The final DUB assay buffer is thus composed of 50 mM Tris pH 7.6, 50 mM NaCl, 1 mg/mL (1.6 mM) CHAPS, 0.132 mg/mL (2 µM) heat denatured BSA and 2 mM DTT. To prevent interferences with the assay caused by particles or air micro-bubbles, the buffer is filtered and degassed prior to use.
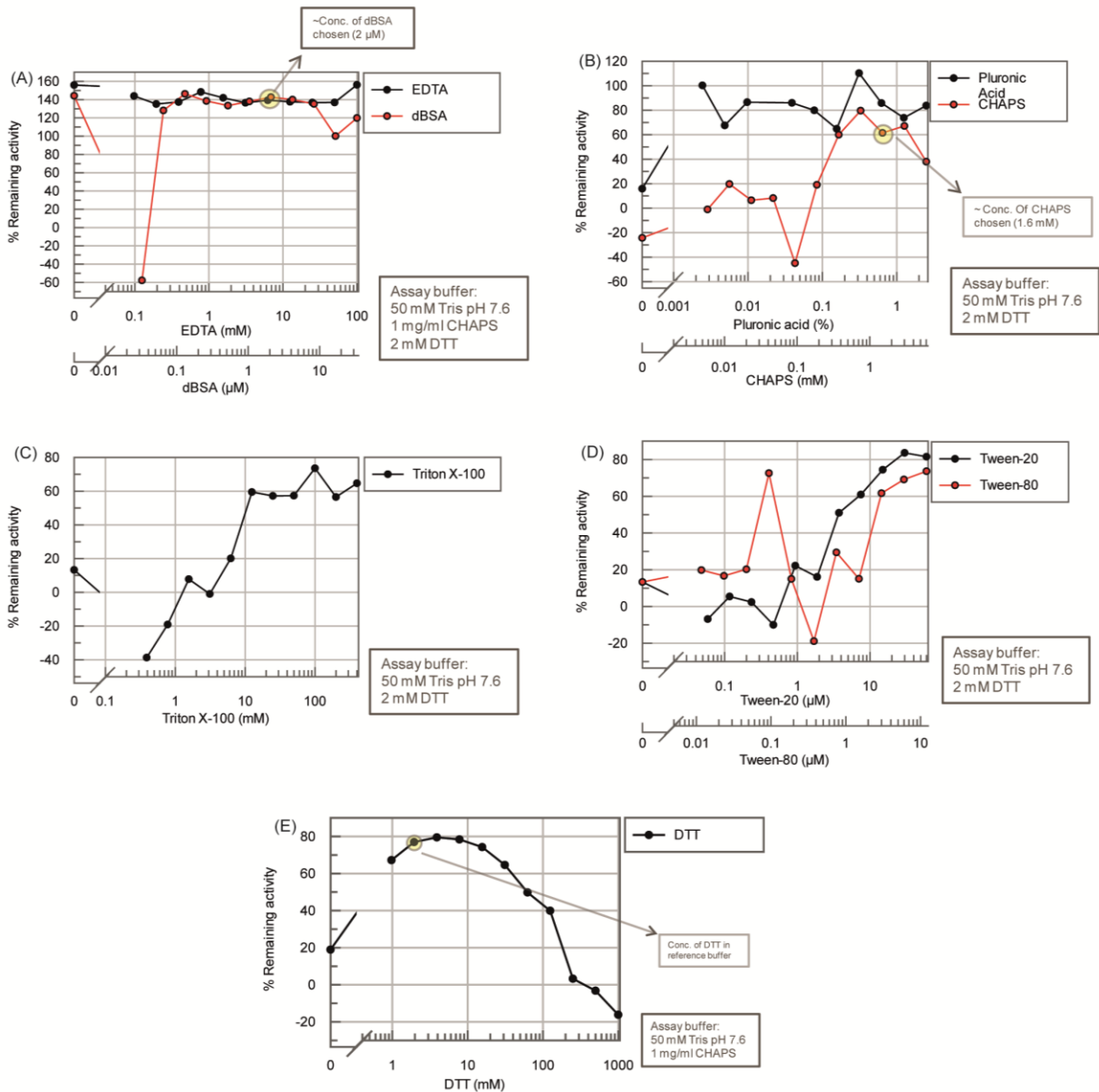
Figure 3.6: Buffer additive screen.
Different components of the assay buffer were tested to determine the conditions that yield a good assay window. (A) EDTA has no effect on the enzyme activity and dBSA at 2 µM performs as well as BGG. (B), (C) and (D) Pluronic acid, CHAPS, TritonX-100, Tween-20 and Tween-80 all give 80 to 100 % remaining activity at certain concentrations. (E) The optimum concentration of DTT is in the range of 2-4 mM.

3.3.2 Product standard curve

A very important factor for the successful performance of an assay in the HTS is the plate reader. Several high throughput capable plate readers are available like

the EnVision™ 2104 MultiLabel Reader (henceforth called EnVision), ViewLux, Spectramax, Analyst GT etc. Different readers are better suited for different assay formats. We used the EnVision plate reader as it had shown robust performance in FP-based assays in the past. As already mentioned, the DUB assay is a FP assay that is based on the fact that cleaved and uncleaved fluorophore-labelled DUB substrate emit light differently when excited by plane polarized light.

FP values (in mP) are determined from measurements of fluorescence signal in the planes parallel and perpendicular to the excitation plane (known as s and p channel signals), using the equation

$$FP = 1000 \times \left[(s - (G \times p))/(s + (G \times p)) \right]$$

where s is the Parallel fluorescence intensity, ex 531 nm, em 579 nm
p is the perpendicular fluorescence intensity, ex 531 nm, em 579 nm
G is the G-factor or the grating factor

To consider a plate reader ready for assay readout, it is important to optimise the settings of the optics so as to obtain a linear product standard curve between the maximum and minimum expected signal in the assay. For a FP assay, the EnVision optics need to be optimised for the G-Factor, measurement height, excitation light %, the number of flashes per well and detector (PMT) gain. Since there are two channels, there are two respective detectors (PMTs). An instrument can have different sensitivities for the two channels (s and p). The G-factor is used to correct for the bias in polarization values caused by differences in instruments and assay conditions. For the EnVision plate readers, the G-factor was determined following the manufacturer's instructions. To calculate the G-factor, the target mP values for 5-TAMRA and the substrate (Ub-FP) used were 50 and 200 mP, respectively. Optics settings were optimised for two EnVision plate readers, with EnVision 1 as the main reader for the HTS and EnVision 2 as the backup reader. Iterative readouts were obtained while varying the settings of the optics to obtain

readouts as close as possible to the target mP for the substrate and product. Optimal optics settings for EnVision 1 were a measurement height of 8.91 mm, 10 light flashes, a PMT gain of 305, a second PMT gain of 225, a G-factor of 0.8 and 60% excitation light. The same settings were used for EnVision 2 except for the measurement height which is 11 mm and a G-factor of 1.

With the optics settings in place, we then tested if the two readers for the linearity of the product standard curve. As shown in **Figure 3.7**, the two readers gave a similar product standard curve and identical slopes using linear regression, confirming that the readers can be considered optically identical.
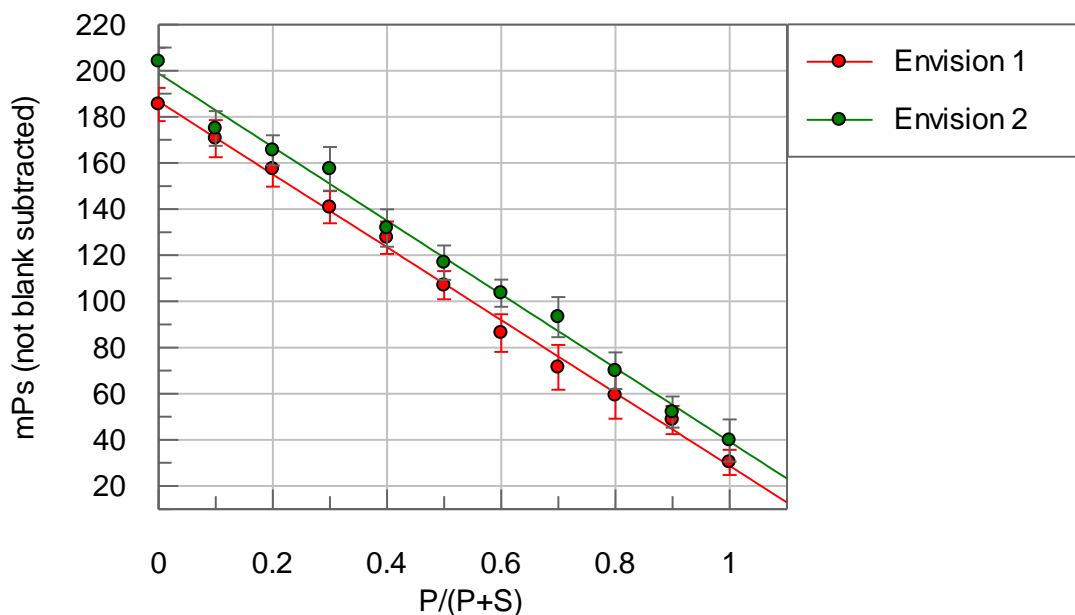
Parameter	Value	Std. Error	Parameter	Value	Std. Error
a (intercept)	186.6627	1.8782	a (intercept)	198.8250	2.6483
b (gradient)	-157.9964	3.1747	b (gradient)	-159.6673	4.4765

Figure 3.7: Product standard curve.
Two EnVision readers were prepared to read the plates for the HTS. EnVision 1 was used as the reader during HTS and Envision 2 served as a backup reader. Both the readers have identical optics settings and gave a similar standard curve slopes. (P: product, S: substrate)

3.3.3 Determination of kinetic parameters

To characterize an enzyme for an activity assay it is very important to determine the Michaelis-Menten kinetics of the enzyme. To determine the Michaelis-Menten kinetics for *Tbr*DUB1, we tested varying concentrations of the substrate (1, 5, 10, 50, 100, 500 and 1000 nM) at an enzyme concentration of 2.5 nM. The mP values obtained were converted to processed substrate at time t as described in section 2.3.2.4. Since the first 15 minutes of reaction showed large variations, the initial rates were obtained by a performing linear fit of remaining substrate concentration from 15 to 60 minutes. The initial rates thus obtained were plotted against the tested substrate concentrations and fitted to Michaelis-Menten enzyme kinetics using Erithacus GraFit 7.0.2 software. The calculated values of V_{max} and the K_M from this fit are 6.2 ± 0.4 nM.min^{-1} and 1373 ± 156 nM, respectively (**Figure 3.8**). Considering that we just begin to see a plateau of the rate of reaction (V) at the highest tested substrate concentration, it is clear that a more precise estimation of these constants requires more data points at higher substrate concentrations than those tested. Fluorescence polarization assays are very attractive for HTS for many reasons but show some limitations too. In our assay, the use of very high concentrations of substrate in kinetic studies is hampered by the fact that the resulting polarization value of reaction mixtures is dominated by the anisotropy of the substrate. As shown in **Figure 3.7**, a high substrate to product ratio in the reaction mix gives a read out of around 200 mP and thus gives a false impression that the reaction did not progress. Despite this limitation, the accuracy of the resulting kinetic parameters was deemed sufficient for the purpose of adjusting the sensitivity of the assay to different types of inhibitors. The final substrate concentration chosen for the assay was 100 nM which is around 0.1 x the observed K_M. At this concentration of substrate our screen will have higher sensitivity towards competitive and noncompetitive inhibitors. The lower sensitivity towards uncompetitive inhibitors was considered a worthy sacrifice in context of high cost of the substrate and the limitations of the readout technology.
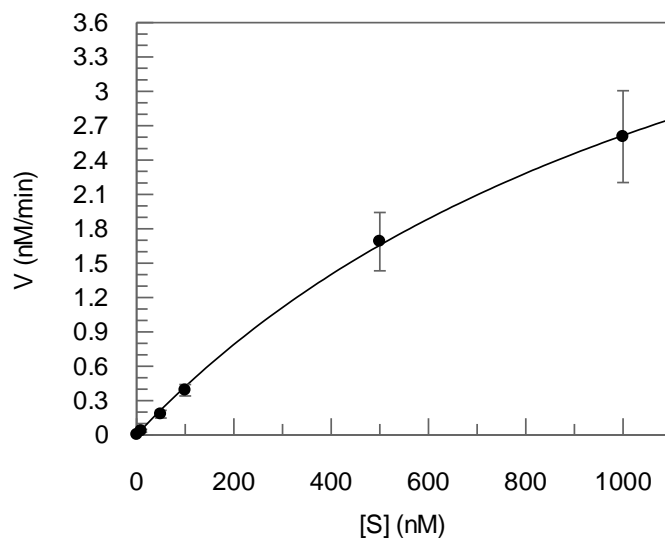
Figure 3.8: Michaelis-Menten kinetics of *Tbr*DUB1.
2.5 nM *Tbr*DUB1 was used to process different substrate concentrations at room temperature and the respective rate of reactions were obtained by calculating substrate processed at time t from the mP values. The rates of these reactions were plotted against the respective substrate concentrations. The calculated approximate values of V_{max} and the K_M are 6.2 nM.min^{-1} and 1373 nM, respectively.

3.3.4 Selection of enzyme concentration and reaction time

For developing a robust assay, it is essential to determine the optimum enzyme concentration (E_{opt}) and the reaction time (t_{reac}) while staying in the steady state of enzyme kinetics. These two factors are typically inversely related *i.e.* an increase in the enzyme concentration allows for a reduction of the reaction time. To find the optimum *Tbr*DUB1 concentration and time of reaction, that ensure reaction linearity and a good assay window, we tested different enzyme concentrations and monitored the enzyme activity over time up to 500 min post reaction start. A substrate concentration of 100 nM was used for these experiments, as high substrate concentration (1 µM) interfered with the assay and resulted in an observed small assay window (not shown). As shown in **Figure 3.9(A, B)**, different enzyme concentrations yielded different initial rates of reaction with assay windows between 80 and 100 mP over time. Theoretically, the maximum observable window is 150 mPs sine 5-TAMRA and Ub-FP substrate yield 50 and 200 mP, respectively. Additionally, it was observed that the plot of rate of reaction (V) *vs.* enzyme

concentration is not linear at concentrations higher than 10 nM. Therefore, to ensure steady state kinetics, a final assay concentration of 5 nM enzyme was chosen with a reaction time of 2 h, yielding an assay window of 80 mP. Depending on Z' values (**Figure 3.9(C)**) the time of reaction can be increased to 200 minutes as the reaction is still in the linear phase at this time point.
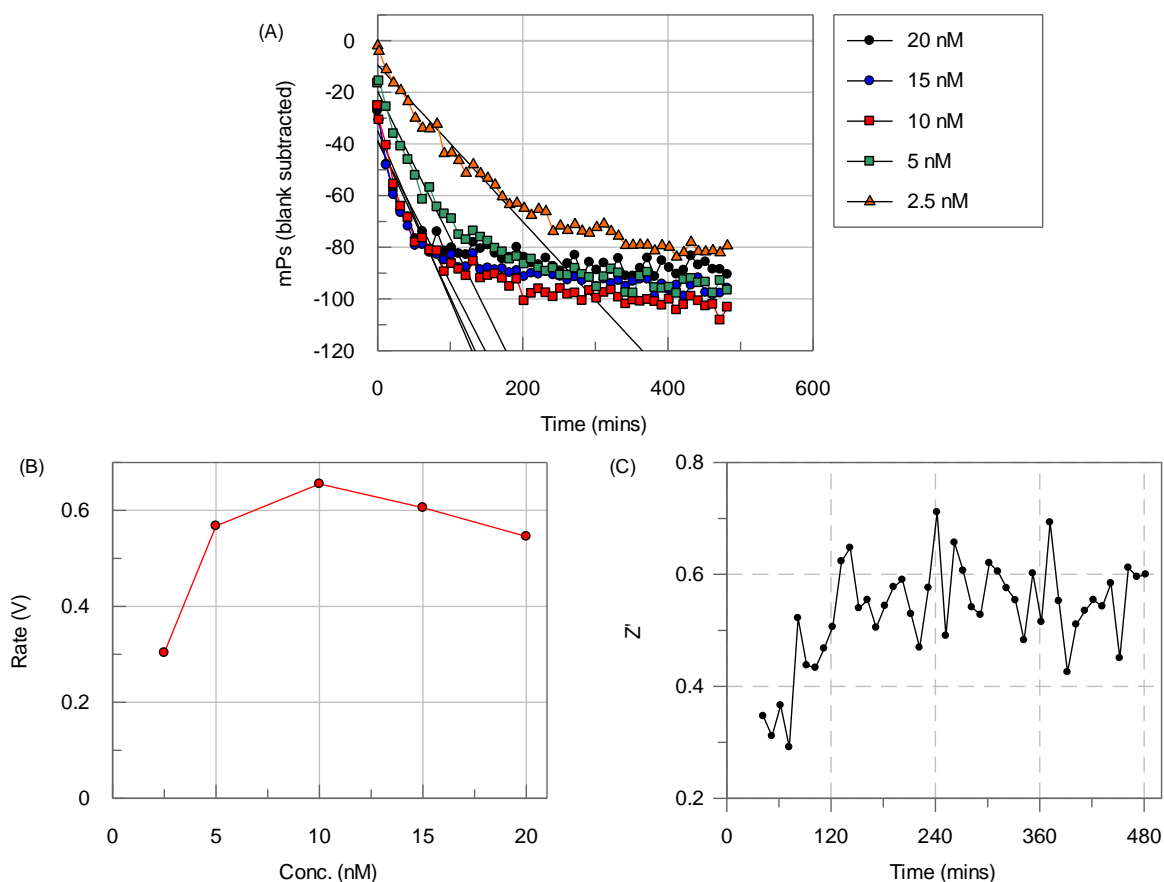
Figure 3.9: Selection of E$_{opt}$ and t$_{reac}$.
(A), (B) and (C) - Different enzyme concentrations were checked for linearity of initial rates and assay window for up to 500 minutes. Also, the Z' was calculated at each time point (D).

3.3.5 Selection of final reaction volume, controls and the effect of centrifugation

The last steps of assay development were to determine robust controls, the final assay volume and to test if centrifugation has a positive effect on the assay

performance through removal of air bubbles. The negative control or control 1 is the reaction in the absence of compounds and hence does not require optimisation. However, for the positive control, multiple options can be considered to achieve full inhibition of the reaction. We compared three different positive controls, a) omitting the enzyme, b) the promiscuous enzyme inhibitor iodoacetamide, a cysteine-alkylating agent and c) a pan-DUB inhibitor H1. H1 is a dechloro analogue of HBX 41,108- a USP7 inhibitor (compound **17** in **Table 1.2**, (Colland *et al.*, 2009, de Jong *et al.*, 2012)). Reactions for each positive control were set up in a final reaction volume of 4, 5 and 6 µL in 1536-well plates in duplicate, with one plate being centrifuged while the other was not. Reactions were allowed to proceed for 6 hours and the signals were monitored at intervals of 18.3 min in order to determine the Z' values over time.

As can be seen in **Figure 3.10**(A), Z' values in all plates increased over time, owing to an increase of assay window, and reached a maximum at 200 min, a time point at which the reaction is still in the linear phase (see **Figure 3.9**, section 3.3.4) This prompted us to change the end point of the assay from 120 min (section 3.3.4) to 200 min. Amongst the positive controls tested, iodoacetamide gave the best Z' values at 200 min, which is explained by the larger assay window and lower coefficient of variation compared to the other controls (**Figure 3.10** B-D). Hence, we decided to use 10 mM iodoacetamide as the positive control. Regarding the reaction volume, a general trend of higher Z' values in 6 µL reactions was observed (in the different controls) compared to reactions in smaller volumes. For iodoacetamide, the reaction volume of 5 µL yields a Z' > 0.52, which is increased by ~0.4 in 6-µL reactions (**Figure 3.10** B). Considering that a total of approximately two million wells are tested in the HTS, a 1 µL increase in reaction volume significantly increases the total amount of enzyme and substrate needed. To balance reagent costs and assay robustness, we decided to use a final reaction volume of 5 µL. Similarly, even though centrifugation slightly improves the assay performance, the gain in robustness does not justify the impact on the logistics of

the HTS, considering that approximately 1200 plates are tested. We therefore did not include the centrifugation step in the final assay.
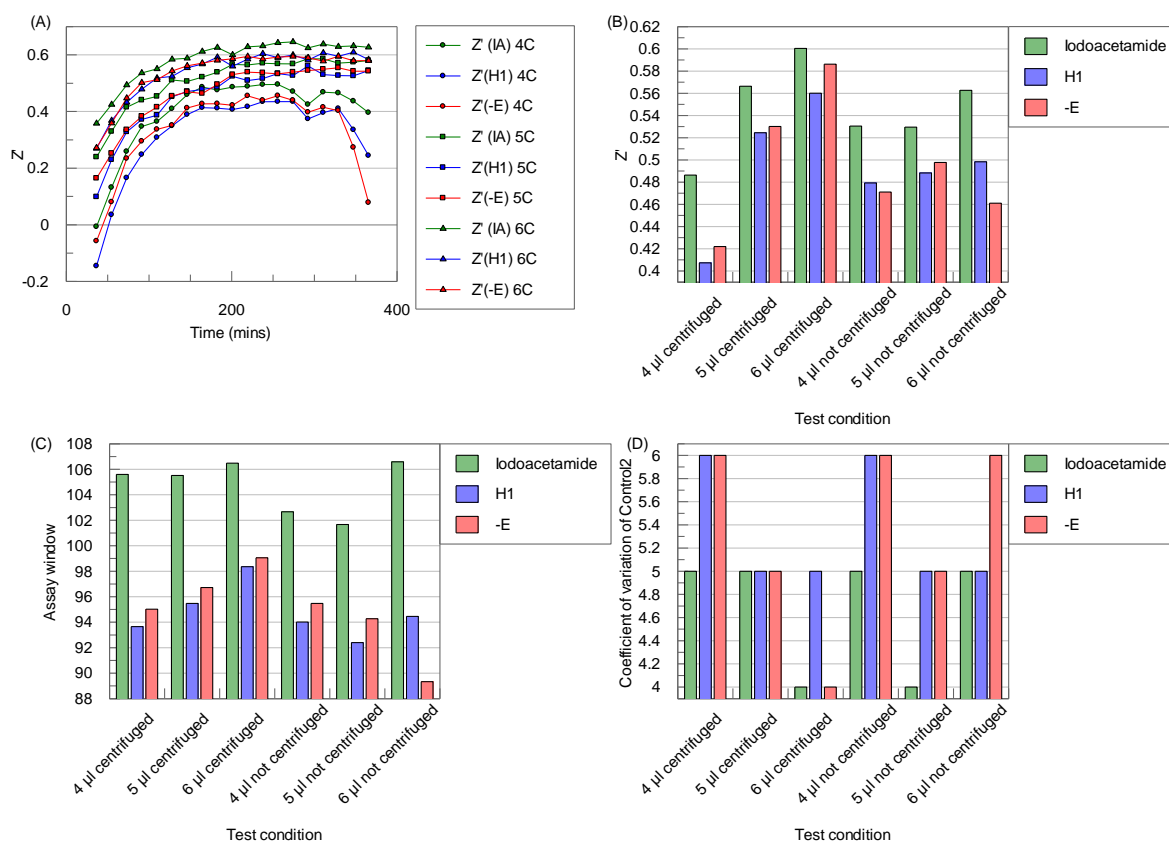
Figure 3.10: Selection of reaction volume and positive control and testing the effect of centrifugation.
(A) The Z' trend over time. The Z' (B), assay window (C) and coefficient of variation (D) for each control at t = 200 mins. Iodoacetamide shows a larger assay window and lower coefficient of variation, leading to higher Z' values, compared to the other two controls. IA: iodoacetamide; H1: specific DUB inhibitor; -E: reaction without enzyme.

In conclusion, we successfully set up a robust FP-based DUB activity assay in a 1536-well plate format, using a final assay volume of 5 µL with 5 nM *Tbr*DUB1 and 100 nM substrate, and a reaction time of 200 min. After the test compounds and iodoacetamide have been dispensed (assay ready plate) the assay protocol *per se* has just two additions and makes it easy to achieve a high throughput. Typical Z' range for this assay is 0.55-0.6. Iodoacetamide serves as positive control. Lastly,

for the HTS, a 30-min preincubation of enzyme with compounds was added to the assay protocol to allow for detection of slow-binding compounds.

3.3.6 Assay automation

To run the HTS in a semi-automated manner, we utilized an in-house assembled platform consisting of an Agilent direct drive robot arm, several Thermo Electron MultiDrop® Combi dispensers and plate racks for holding stacked plates. The platform was controlled by Agilent VWorks automation control software. The Agilent direct drive robot arm works by defining the coordinates of each location within a 3D grid. The robotic arm was used to move each plate from a stack of assay ready plates to the first MultiDrop Combi dispenser which adds 2 µL of enzyme solution to the plate. Following this, the robotic arm moves the plate to a rack for a 30-min incubation of compounds and enzyme. Meanwhile, enzyme dispensation continues in more plates, each separated by 3 min, which is the time it takes to read each plate on the EnVision. This ensured that each plate has the exact same incubation time of enzyme and compounds and when read at the end of the reaction. After 30 min of preincubation, the plate is moved to the second Multidrop Combi where 3 µL of substrate solution is dispensed and the plate is stacked in a rack where it is incubated for 200 min, protected from light. At the end, a restacking step is required to put the first plate at the bottom in order to be read first by the EnVision plate reader (**Figure 3.11(A, B)**).

Each plate tested contains two columns (64 wells) for each control, columns 11, 12 for the negative control and columns 35 and 36 for the positive control iodoacetamide (**Figure 3.11(C)**). With the assay development and automation phases complete, the next step is to test the robustness of the assay in the presence of the compounds.

(A)

MultiDrop nano   Agilent Direct drive robot   EnVision

Incubate 200 min
•Restack plates
•Move to Envision

Prepare   Stack   Incubate
reagents   plates   30 min

Dispense   Dispense   Dispense   Read
50 nL IA   3 µL   2 µL   plates
                  Enzyme   Substrate

(B)

Add 2 µl Enzyme (5 nM final
assay conc.)

Incubate 30 min (RT)

Add 3 µl substrate (100 nM
final assay conc.)

Incubate 200 min (RT)

Read on Envision (measure
emission light intensity in
both parallel and
perpendicular channels)

(C)

Test (320 wells) | Negative control (64 wells) | Test (704 wells) | Positive control (64 wells) | Test (384 wells)

Figure 3.11: The process workflow and the final assay design.
(A) The series of working steps on a typical screening day. Plates pre-dispensed with compounds were tested in batches of 60 (IA: Iodoacetamide). (B) The assay protocol. (C) The plate design for the HTS. In a 1536-well plate, columns 11 and 12 (a total of 64 wells) serve for negative control (control 1) while columns 35 and 36 contain iodoacetamide as positive control (control 2).

3.4 Assay validation

3.4.1 Robustness set

To consider an assay ready for HTS, it is crucial to check how robustly the assay performs when exposed to the organic load of compounds in DMSO. For this, the normal practice at GSK is to test the assay against a set of 1408 compounds (1 x 1536-well plate) in duplicate. This set is a subset of the validation set (next section) and the full HTS collection, and consists of compounds with a broad range of chemical properties. While relatively small, this set allows for identification of major

problems in the assay which then can be revisited and optimised. The main requirement for passing the robustness test is a Z' > 0.4 in each replicate. Additional information that is obtained from testing the robustness set is the mean SD of the controls, hit rate, response cut-off for marking hits, correlation coefficient, outliers etc. We coupled the robustness test to the test of reagent stability, as in a HTS scenario, approx. 60 plates are dispensed over a period of 3 h at room temperature. For this, we interspersed the robustness set plates with several plates containing the DUB inhibitor H1 in concentrations ranging from 0-100 µM in 1:3 serial dilutions (dose response plates). Dispensation of these plates was spread over a period of ~3 h to simulate the timing during the HTS. In addition to the stability of the reagents, the dose response plates of H1 further allow us to assess the sensitivity and the reproducibility of the assay.

As shown in **Figure 3.12(A)**, the H1 dose-response tests resulted in reproducible IC_{50} values with an average of 8.8 x 10^{-7} ± 4.8 x 10^{-8} M (pIC_{50} = 6.1). Robust statistics are often used in HTS scenario, wherein outliers are not included when calculating the Z' and coefficient of variation. Despite a relatively low signal-to-background ratio in all tested plates (1.97 ± 0.10), a good average robust Z' value of 0.6 ± 0.03 was obtained, demonstrating assay robustness (**Figure 3.12(B)**). Importantly, the Z' value did not decrease over time, indicating that the reagents are stable at RT for at least 3 hours, and are therefore suitable for the HTS.
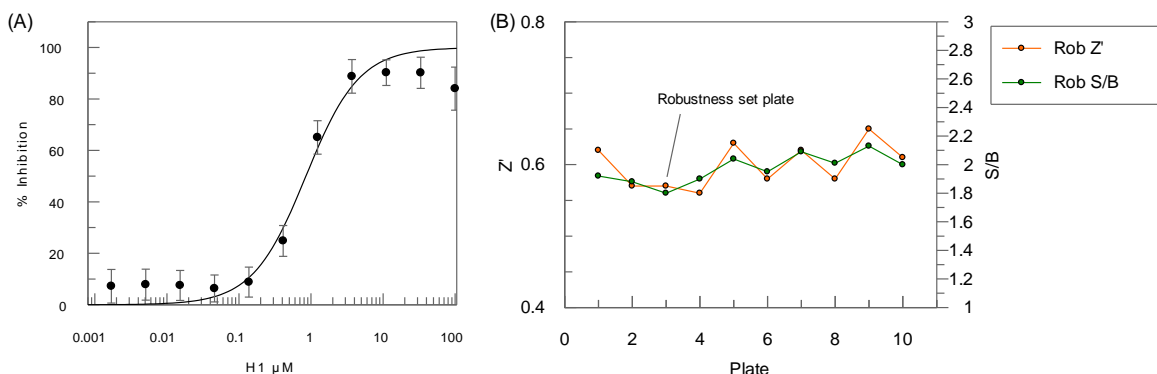
Figure 3.12: Assay performance during robustness test.
(A) H1 dose response curves (n = 24, 8 plates with 3 replicates per plate) were reproducible and gave an average IC_{50} value of 8.8 x10^{-7} ± 4.8 x10^{-8} M. (B) The Z' value and signal-to-background ratio (S/B) are stable over time and indicate a robust assay.

As seen in **Figure 3.13**, only two of the tested compounds showed robust activity against *Tbr*DUB1. Compounds are considered as robustly active when the response is greater than three times the SD of the mean response of the negative control. The small number of actives indicates that either the assay is not sensitive to detecting the inhibitors against *Tbr*DUB1 or that the assay is not prone to high hit rate. Given the sensitivity to H1, as seen in the H1 dose response in Figure 3.12(A), the latter might be the case. However, with Z' > 0.4 the assay passed the robustness test and is therefore ready to be tested more extensively on a larger compound set (validation set) with respect to reproducibility, hit rate, false positive and false negative rates.
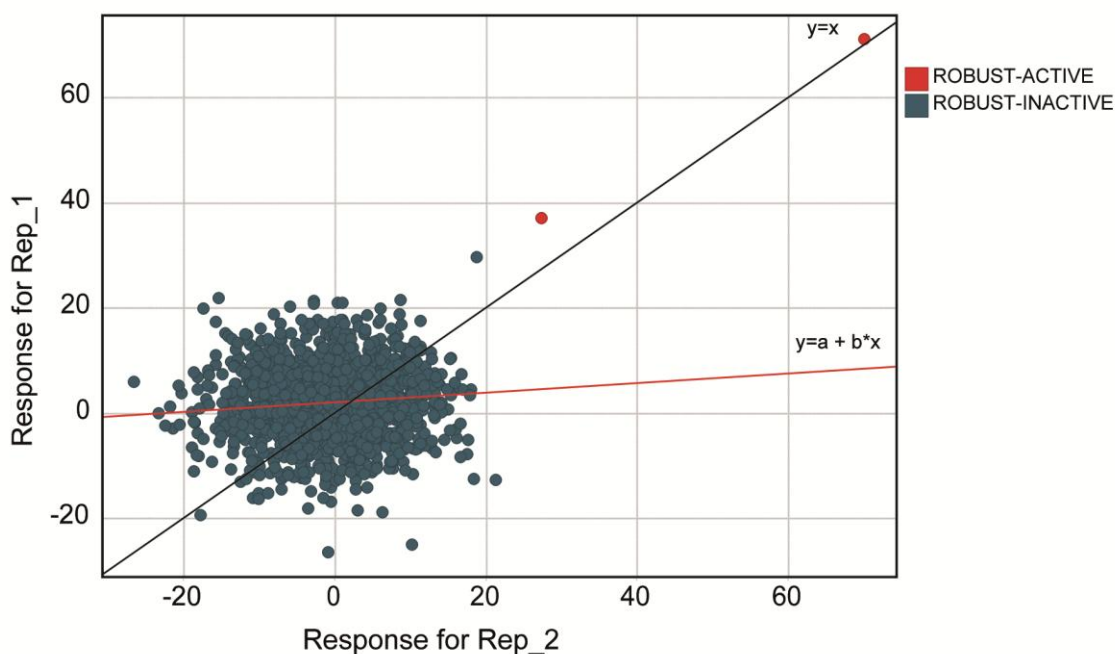
Figure 3.13: Replicate response correlation for the robustness set.
Robustness set was tested in duplicate. The assay performance was robust and gave only 2 robust hits.

3.4.2 Validation set

The validation set consists of 9838 diverse compounds (7 x 1536-well plates) that are representative of the full screening collection, and serves to validate the assay

and test its statistical quality before embarking on the HTS. For this, the set is tested in triplicate on at least two different days to account for potential day-to-day assay variability. The stability and separation of controls (Z'), reproducibility of hits and number of outliers are checked. Replicates with less than 80 % of plates having acceptable Z' values (> 0.4) or replicates with many extreme errors are causes of concern and indicate the requirement for further assay optimisation.

For the *Tbr*DUB1 assay, the validation replicates were run on two days, one replicate on day 1 and two replicates on day 2. The Z' values were similar in all plates of the replicates and above the quality control limit of 0.4, (**Figure 3.14**) indicating a good separation of the controls.
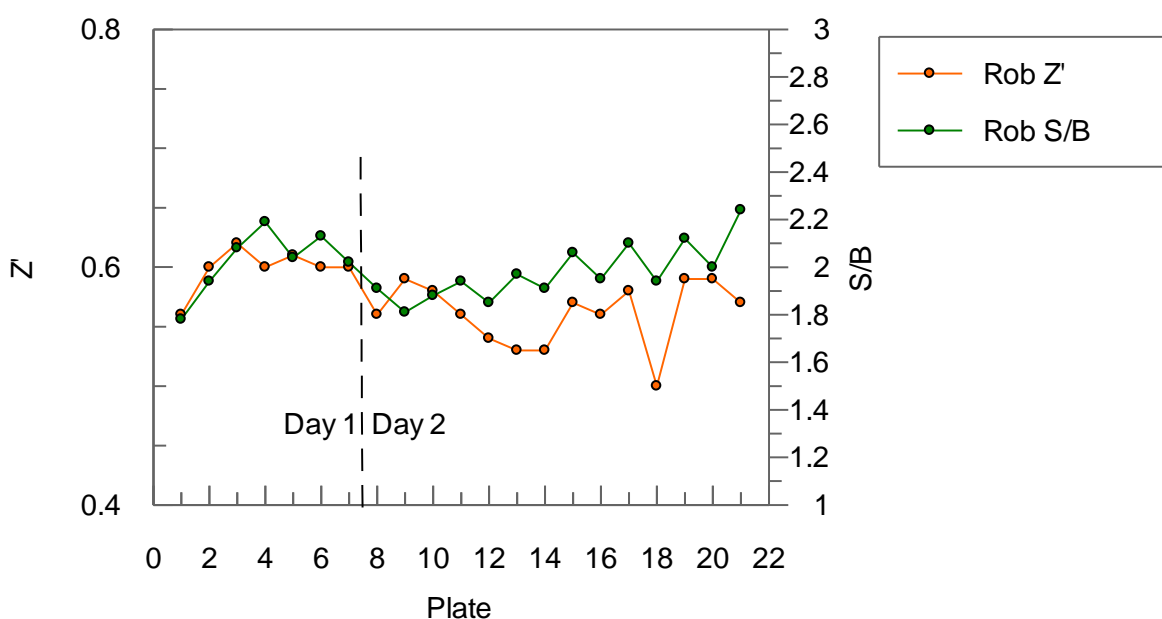
Figure 3.14: Assay performance during validation set screening.
The robust Z' and robust signal-to-background ratio are similar on the 2 days. The Z' never falls below the quality control limit of 0.4 which indicates that the assay is robust.

Furthermore, the replicates showed a strong correlation of response (**Figure 3.15**). The statistical parameters from the analysis of the replicates are presented in **Table 3.1**. As mentioned before, the cut-off for marking robustly active compounds is determined as 3x SD of the mean response of the negative control and was

24.5 % in the validation set. We obtained a hit rate of 0.26 % with a high predicted (for the HTS) confirmation rate of around 69 %. This implies that for a HTS of approximately 1.7 million compounds we should expect approximately 4000 hits. The false positive and the false negative rates of 0.05 % and 3.51 % are not very high. The extreme false positive rate is near zero, confirming the assay's reproducibility. The extreme false negative rate (EFNR) was high (5.26 %) in one replicate. However, the other two replicates did not contain any extreme false negatives, leading to an acceptable average EFNR of 1.75 %.

In conclusion, the *Tbr*DUB1 assay passed the assay validation and can be considered ready for the HTS campaign which is described in the following chapter.
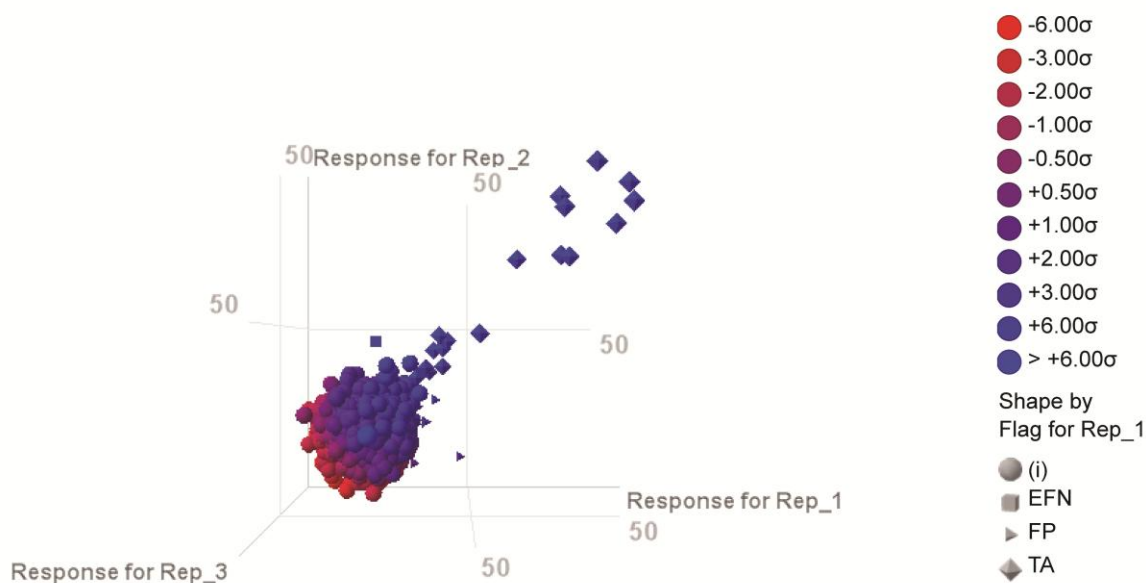
Figure 3.15: Replicate response correlation of the three copies of the validation set.
The response of each replicate is represented on a different axis, showing good correlation. EFN: Extreme false negative, FP: False positive, TA: True active, (i): No flag.

Table 3.1: Validation set statistical parameters.

Descriptor Name	Descriptor	Rep_1	Rep_2	Rep_3	Averages	Val. Indexes
Number of data points	ni	9838	9838	9838		
Number of considered triplets	nx3					9838
Estimated Robust mean	r mean	2.3397	1.391	1.7628		
Estimated Robust Standard Deviation (SD)	r sd	7.2412	7.7211	7.7514	7.57	
Cut Off 3 x SD	cut-off	24.06	24.55	25.02	24.54	
Robust Average Range (%)	robAR					12.29
Inter Class Coefficient (ICC)	ICC					0.15
Number of False Positives	FP	8	2	6		
Number of False Negatives	FN	1	0	1		
Number of Extreme False Positives	EFP	0	0	1		
Number of Extreme False Negatives	EFN	1	0	0		
Number of Hits	nHits	27	27	23		
Number of Triplicated Hits	nx3Hits	27	27	23		
Number of Statistical Actives	TA					19
Number of Statistical Inactives	TI					9819
Statistical Active Rate (%)	TAR (%)					0.19
Predicted Confirmation Rate (%)	PCR (%)	66.67	70.37	69.57	68.87	
Hit Rate (%)	HR (%)	0.27	0.27	0.23	0.26	
False Positive Rate (%)	FPR (%)	0.08	0.02	0.06	0.05	
False Negative Rate (%)	FNR (%)	5.26	0	5.26	3.51	
Extreme False Positive Rate (%)	EFPR (%)	0	0	0.01	0	
Extreme False Negative Rate (%)	EFNR (%)	5.26	0	0	1.75	

Chapter 4 Hit identification

Abstract

In this chapter, we present the results from the HTS of the GSK collection of approximately 1.7 million compounds against *Tbr*DUB1 using a FP assay. The primary screen was performed at a 10 µM compound concentration. 9405 hits were identified and were re-tested in duplicate at the same concentration, which yielded a total of 2020 confirmed, robustly active compounds. Their potencies and physicochemical properties are presented.

4.1 High throughput screening

As highlighted in the introduction, one of the proven approaches in drug discovery is to screen a chemical library against a target of interest. The chances of identifying tractable hits increase when using libraries of structurally diverse compounds that sample the chemical space, which in its entirety is too large to manage.

The GSK screening library is a dynamic, high-quality collection of around 1.7 million compounds that has been enriched in structurally diverse drug-like molecules over many years. More than 90 % of compounds have a purity of > 90 % at the time of incorporation, with the remaining ~10 % of compounds being > 80 % pure. To identify inhibitors of *Tbr*DUB1, we screened the full GSK library using a FP-based DUB activity assay that we developed and validated (Chapter 3). The primary screen was performed at a final compound concentration of 10 µM, according to the assay protocol described in section 2.3.3.1. Briefly, compound stocks in DMSO were dispensed in 1536-well plates using an automated Echo liquid handling system. Following dispensation of *Tbr*DUB1 (5 nM final concentration), the enzyme and compounds were incubated for 30 min before starting the reaction by adding the 5-TAMRA-labelled ubiquitin substrate (100 nM final concentration). After a reaction time of 200 min, fluorescent signal was measured using an EnVision plate reader, and data analysis was performed with ActivityBase XE software (see section 2.3.3.2 for details). In each plate, columns 11 and 12 were used for the negative control, i.e. the uninhibited reaction, while columns 35 and 36 contained the positive control iodoacetamide which inhibits the enzyme through cysteine alkylation.

A total of 1,893,888 wells in 1233 plates were tested. We screened a total of 1,668,637 unique compounds. The HTS was carried out in a semi-automated manner (section 3.3.6). The average robust Z' value for all valid plates was 0.57 with a SD of 0.03 (**Figure 4.1(B)**). Only seven plates did not pass the quality

control criterion of a Z' > 0.4, resulting in a very low failure rate of 0.57 %. However, one plate with a Z' of 0.39 was included in the data analysis, as it had only five hits that were easily re-tested in the confirmation step.
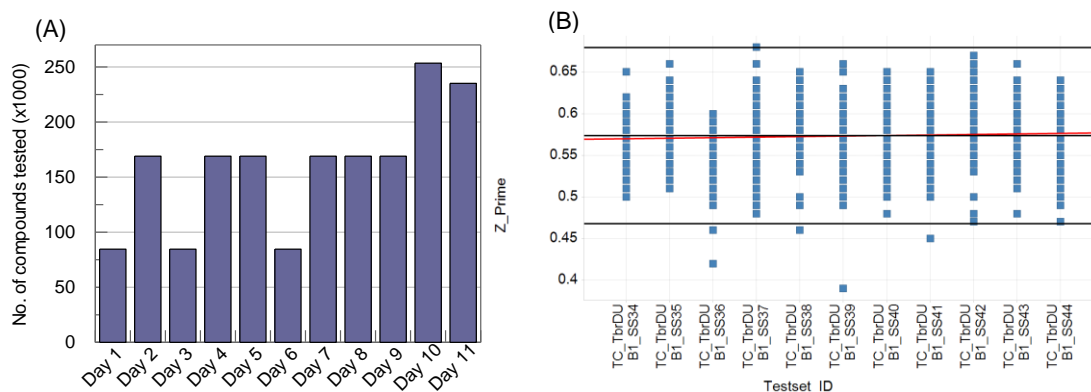
Figure 4.1: HTS chronogram and Z'.
(A) The total number of compounds tested each day of the screening. (B) Z' trend of plates tested in the HTS. Each test set represents plates from each day of screening.

As seen in **Figure 4.2**, the control populations show a very good separation with each of them showing a bell-shaped normal distribution of responses. As discussed earlier the cut-off for marking the hits is calculated as:

$$Hit\ cutoff = Mean\ + \ (3 * SD)$$

The reasoning is that a response > 3SD is less likely to be a chance fluctuation from the population of inactive compounds. The hit cut-off was calculated to be 24.23% response, resulting in 5331 robust actives *i.e.* a hit rate of 0.3 %.

109
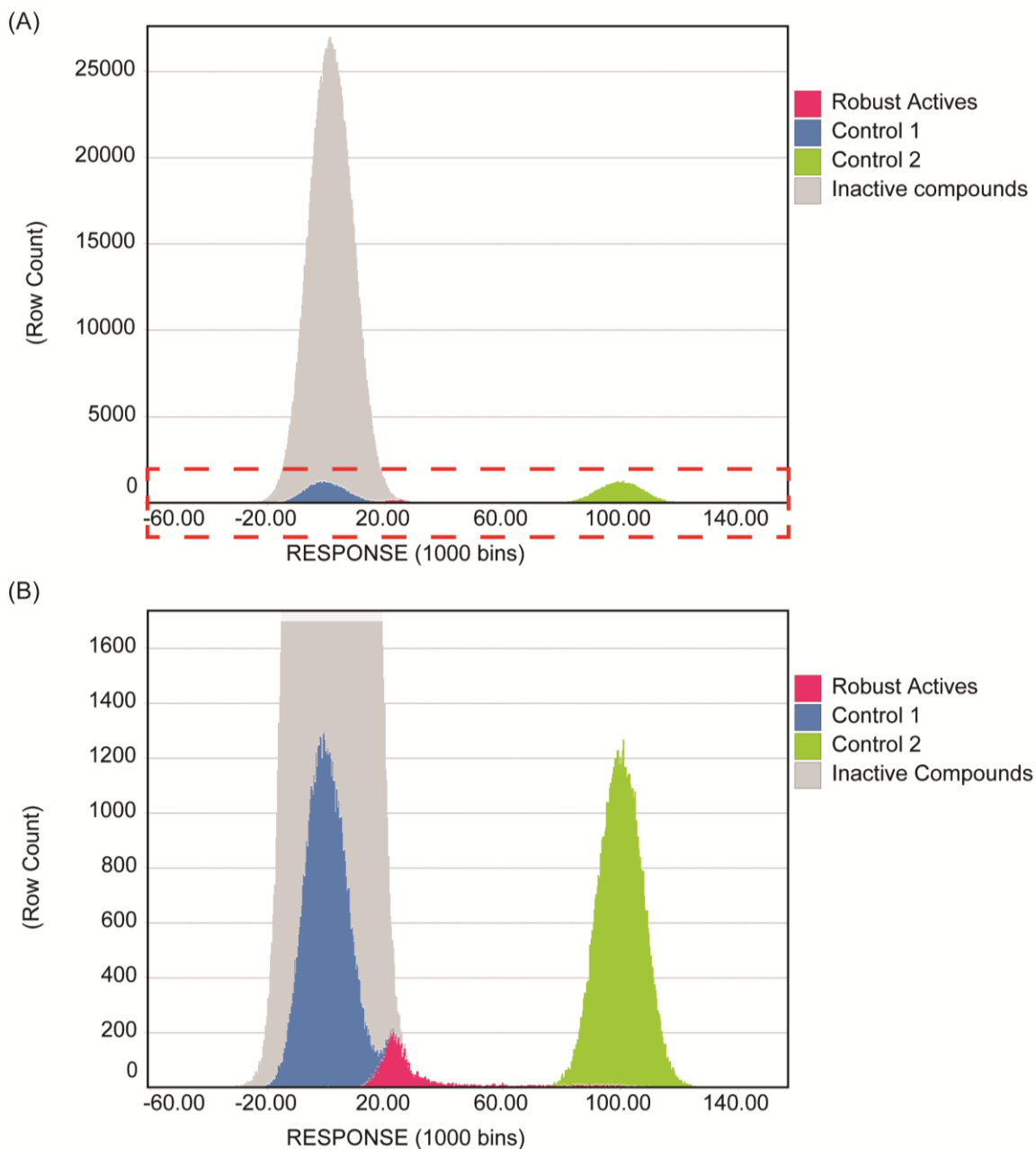
Figure 4.2: Response in the primary HTS.
(A) and (B) The x axes represent the normalised response and the y axes represent the number of rows, equal to the number of compounds showing that response. Marked in pink are the robust actives (9405). The control populations are shown in blue and green and the inactive population in grey. (A) The response of the entire collection in the HTS. A zoom of the area marked in dotted red rectangle is shown in (B) to show the distribution of robust actives.

As frequently observed in HTS, the plates showed a pattern (**Figure 4.3**), which is likely attributed to the dispensation technique used (Coma *et al.,* 2009). The pattern was identified and corrected for as described in Section 2.3.3.3 and as shown in **Figure 4.3**. The pattern correction reduced the variance of response and thereby lowered the 3SD cut-off from 24.43 % to 19.15 %, yielding 8315 hits. Transverse cut-off correction (hit rate by position) was also applied (section 2.3.3.3) and led to 5268 hits at a 3.5SD cut-off. There is a significant overlap in the lists of compounds selected after each of these corrections. Additionally, for the most potent compounds (> 50 % response), we looked for analogues, *i.e.* tanimoto similarity > 0.8, in the 2-3 SD region. A total of 397 analogues were found and were included in the final selection. Combined, a total of 9405 hits were selected, giving a hit rate of 0.56 %. The hits were next tested in duplicate at the same compound concentration of 10 μM to check for reproducibility of inhibition and to remove false positives.
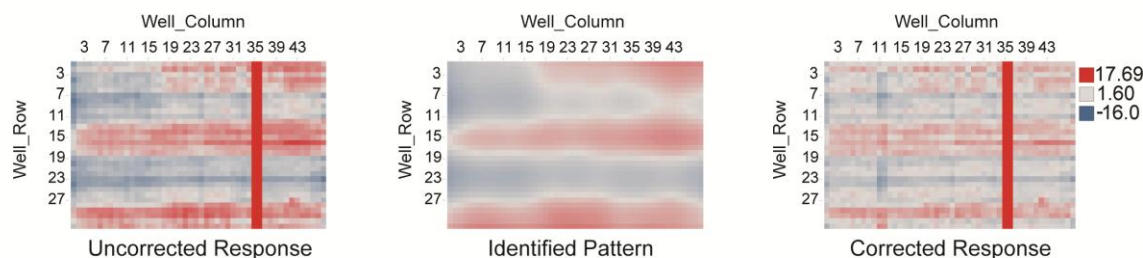
Figure 4.3: Identification and correction of patterns in the plates from the HTS.
The left panel shows an average response at each position of all tested plates. The identified response pattern is shown in the centre panel, while the right panel shows the responses after pattern correction. For the purpose of visualisation of the pattern, a response of 17.69 % is marked in red, 1.6 % in grey and a response of -16 % is marked in blue.
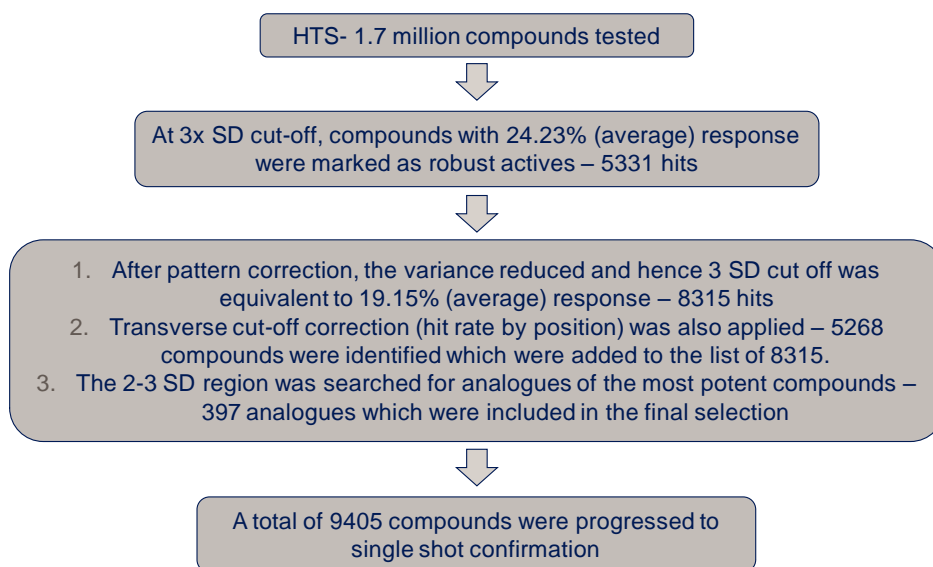
Figure 4.4: Hit identification.

4.2 Hit confirmation at single concentrations

We selected 9405 compounds to be tested again in duplicate under the same conditions as the primary screen, *i.e.* a compound concentration of 10 µM, to confirm compound activity against *Tbr*DUB1 and identify false positives. Of these, 221 compounds could not be tested due to limited compound availability at the time of the confirmation screen, leaving 9184 compounds. However, these 221 compounds were (later) incorporated in downstream dose response studies. None of the tested plates failed the quality control and the average Z' value of each replicate set was 0.60 and 0.59 with a SD of 0.03 and 0.04, respectively. In contrast to the primary screen, no strong plate patterns were recognized in the confirmation screen. Hence, the response data did not require correction. The 3SD cut-off in each replicate was similar to that in the primary screen (19.15 %), 19.76 % in one replicate and 20.15 % in the other. Robust activity was confirmed for a total of 2020 compounds (22 % of the primary hits), with 1325 compounds being active in both replicates. As seen in **Figure 4.5(A)**, the correlation of responses between the replicates was excellent. The correlation between the primary HTS response and the replicates is shown in **Figure 4.5(B)**. For the analogs of potent compounds rescued from the 2-3 SD region of response in primary HTS, the

112

confirmation rate was surprisingly high (18.2 %), amounting to 71 compounds. Therefore, a good number of compounds were rescued through this strategy and could help to better understand the structure-activity relationship (SAR) around the most potent compounds.
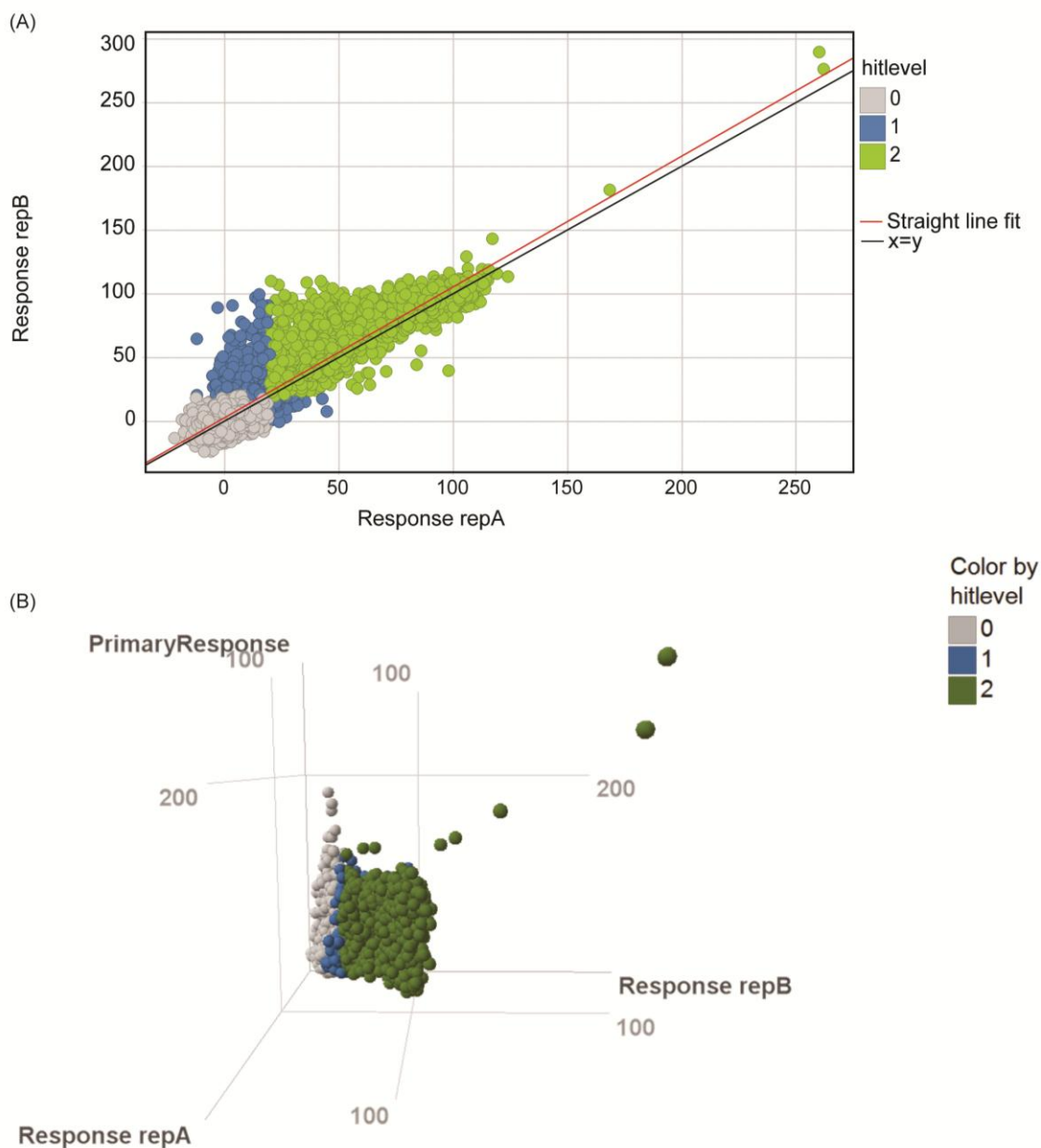
Figure 4.5: Hit confirmation studies.
A) Correlation of responses between the replicates. B) Correlation of responses between the primary screen and confirmation screen replicates. A hit level of zero (grey) represents primary hits that were not confirmed in either replicate. A hit level of one (blue) describes primary hits that were confirmed in one of the two replicates, while hits in green (hit level = 2) were confirmed in both replicates. A total of 2020 primary hits (22 %) were confirmed.

4.3 Analysis of physicochemical properties of confirmed hits

The importance of the physicochemical properties of compounds cannot be overstated. Several studies analysing approved oral small-molecule drugs or molecules in development, showed that the likelihood of a compound to progress from early development to clinical trials increases when certain physicochemical properties are controlled (Lipinski *et al.*, 1997, Lipinski, 2000, Lipinski *et al.*, 2001, Leeson and Springthorpe, 2007, Gleeson, 2008, Hughes *et al.*, 2008, Peters *et al.*, 2009, Gleeson *et al.*, 2011, Luker *et al.*, 2011, Waring *et al.*, 2015). The links of physicochemical properties to absorption, distribution, metabolism and excretion (ADME) (Lipinski *et al.*, 1997, Lipinski, 2000, Lipinski *et al.*, 2001, Leeson and Springthorpe, 2007, Gleeson, 2008), promiscuity (Leeson and Springthorpe, 2007, Peters *et al.*, 2009), *in vivo* toxicology (Hughes *et al.*, 2008, Luker *et al.*, 2011, Waring *et al.*, 2015) and clinical safety (Waring *et al.*, 2015) are well established. In 1997, Lipinski *et al.* (Lipinski *et al.*, 1997, Lipinski *et al.*, 2001) established a rule of five by analysing a set of 2245 small molecules in phase II studies or beyond. The rule of five states that the likelihood of good permeation and absorption of a compound is higher when its

a) molecular weight is < 500 Da,
b) the calculated octanol-water partition coefficient (clogP) is < 5,
c) the number of hydrogen bond acceptors is < 10 and
d) the number of hydrogen bond donors is < 5.

It is worth noting that this rule only applies to compounds that passively diffuse while compounds that are actively transported across membranes are excluded from this rule. In addition, the property forecast index (PFI), defined as $clogD_{pH7.4}$ + number of aromatic rings, has been shown to be an effective indicator of compound solubility (Hill and Young, 2010, Young *et al.*, 2011). Using a combined data set of 812 oral, small-molecule drug candidates nominated for development between 2000 and 2010 by Astra Zeneca, Eli Lilly, GlaxoSmithKline and Pfizer, it was observed that compounds failing in phase I due to clinical safety issues, have

a higher clogP than compounds that progressed to phase II ([Waring *et al.*, 2015](#)). The same study also showed that zwitterionic compounds fare better compared to acidic, basic or neutral compounds in preclinical and clinical toxicology studies, but perform worse in pharmacokinetics studies in phase I ([Waring *et al.*, 2015](#)). Using the above information as guideline, we analysed the physicochemical properties of the *Tbr*DUB1 hits to identify compounds with the highest chances of progressing to the next phases. Hits with undesirable physicochemical properties were not discarded but were 'red-flagged'.

The distribution of several physicochemical properties of the confirmed 2020 hits is shown in **Figure 4.6**. The majority of *Tbr*DUB1 hits have desirable drug-like physicochemical properties in conformation with the rule of five and have number of aromatic rings < 3 and PFI < 8. Looking at these distributions it is clear that the *Tbr*DUB1 assay is not biased for any physicochemical property for the selection of hits. This also indicates that if some of our hits pass all the secondary filters described in the next chapter, there will be high chances of the selected hit being a good starting point for lead discovery and lead optimization studies.
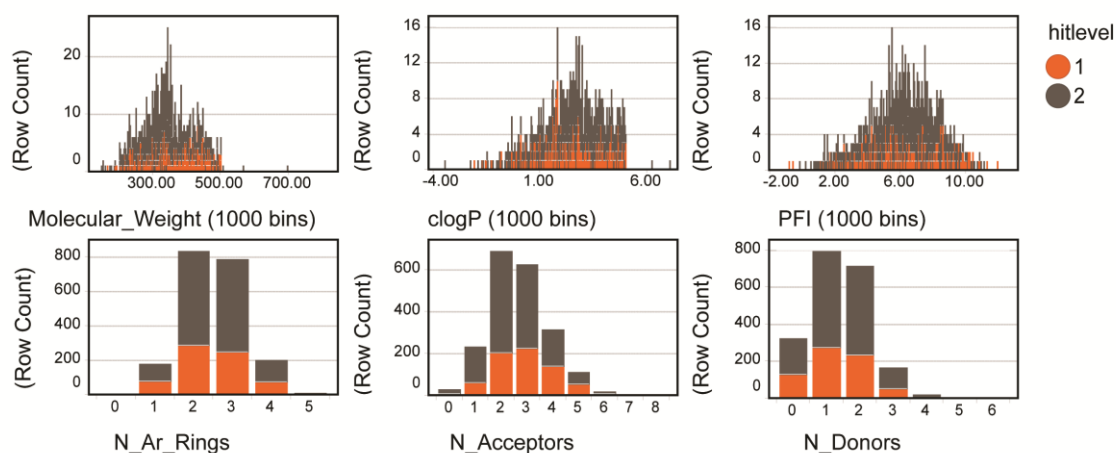
Figure 4.6: Physicochemical properties of confirmed hits.
Most hits meet the criteria of drug-like physicochemical properties, *i.e.* molecular weight < 500 Da, clogP < 5, PFI < 8, no. of aromatic rings < 3, no. of hydrogen bond acceptors < 10, number of hydrogen bond donors < 5. In orange are the compounds that were active in one of the confirmation screen replicates (hit level = 1) and in brown are those that were active in both (hit level = 2).

4.4 Determination of potency of confirmed hits

In order to determine the potency (IC_{50}) of the confirmed hits, each compound was tested in duplicate at 11 concentrations in 1:3 serial dilutions, with a maximum concentration of 100 μM. In addition to the 2020 confirmed hits, 221 primary hits that could not be tested in the confirmation screen, were included to list of compounds for dose-response studies. However owing to compound availability issues, finally 2142 compounds were tested in dose response studies.

The IC_{50} values were calculated by fitting the normalised data to a four parameter logistic fit equation defined as

$$y = \frac{A - D}{(1 + (x/C)^B)} + D$$

Where, x is the concentration of the test compound, y is the inhibition or activation (%), A is the maximum signal, D is the minimum signal (background), C is the IC_{50} and B is the Hill slope. IC_{50} is the concentration of the compound at which it shows half (50%) of the maximum response observed. Expressing IC_{50} values in molar units, the pIC_{50} values are calculated as:

$$pIC_{50} = \log\left(\frac{1}{IC_{50}}\right)$$

As seen in **Figure 4.7**, the compounds' pIC_{50} values show a very good reproducibility. As can be seen there is one compound with very good potency (7 < pIC50 ≤ 8) and 29 and 36 compounds respectively in the 2 replicates with pIC_{50} between 6 and 7. In total 1412 and 1423 compounds respectively in the two replicates showed a dose response behaviour.
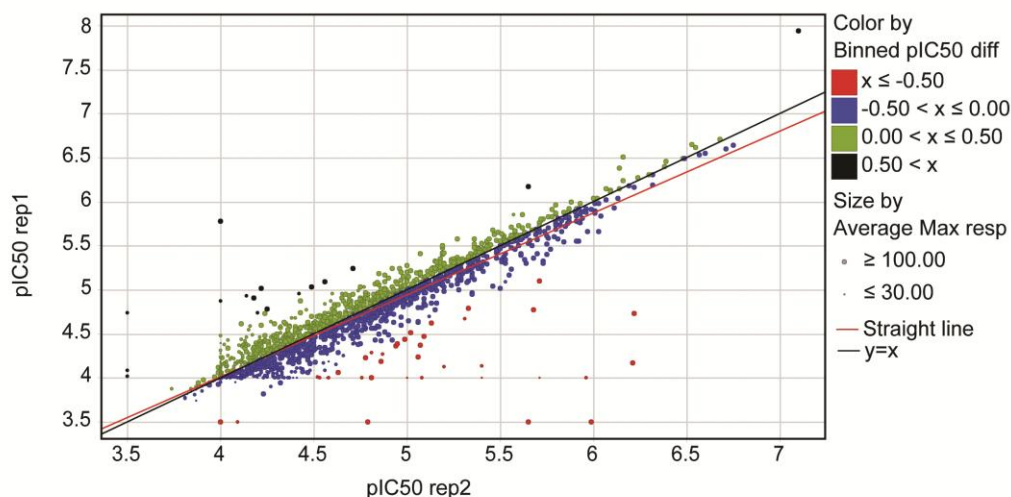
Figure 4.7: Correlation of pIC50 values between the two replicates.
The pIC_{50} values of compounds were calculated from 11-point dose-response curves with a maximum compound concentration of 100 µM. The compounds are coloured by difference in pIC_{50} values in the 2 replicates- blue and green points show a difference of less than 0.5 in the pIC_{50} values, while red and black show a difference greater than 0.5.

Table 4.1: The Z' and distribution of potencies of *Tbr*DUB1 hits in the two replicates.

	Replicate 1	Replicate 2
Robust mean Z'	0.53	0.52
Robust SD Z'	0.04	0.03
$pIC_{50} \leq 4$	731	720
$4 < pIC_{50} \leq 5$	994	976
$5 < pIC_{50} \leq 6$	388	410
$6 < pIC_{50} \leq 7$	29	36
$7 < pIC_{50} \leq 8$	1	1
Total	2143	2143

Shown in **Figure 4.8** are the dose response curves for all the compounds for each replicate. As discussed in section 2.3.3.2 based on the maximum response observed and the quality of fit the compounds were categorised as: inactive compounds, poorly active compounds, compounds that failed XC50 CL ratio, compounds with inappropriate fit, compounds with negative slope and lastly compounds that passed all fit parameters.
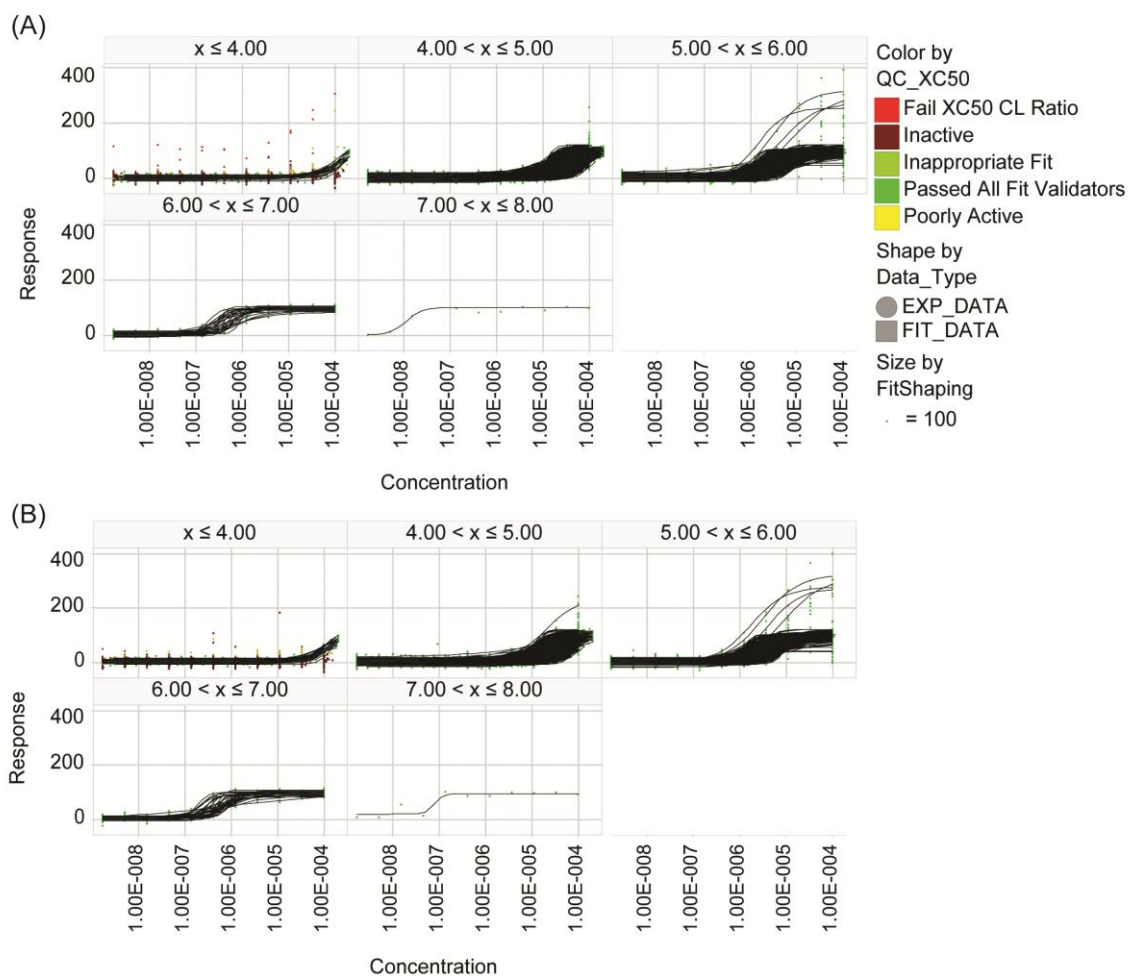
Figure 4.8 Dose response curves of the tested compounds.
(A) and (B) show the dose response curves for the 2 replicates, respectively. The experimental and fitted data are plotted. Different panels show different ranges of the calculated pIC_{50} values. The compounds were categorised based on the maximum response observed and the quality of data fitting. These categories are coloured as indicated.

In summary, we ran a successful HTS campaign and identified 2020 robustly active compounds, several of which showed high inhibition. We also determined the potencies of these confirmed hits that will be used for final selection of potent and selective hits. We also analysed these confirmed hits for their physicochemical properties and found them to be drug like which implies that we have good chances of identifying tractable hits. The next step is to characterise these hits to deprioritise compounds with undesirable mechanisms of action and low selectivity.

Chapter 5 Hit characterisation

Abstract

In this chapter, we present results from different biochemical assays that aimed at filtering out compounds with undesirable redox-based mechanism of action (MOA) and those with non-selective activity against cysteine proteases. We further tested the *Tbr*DUB1 hits against the parasite *T. brucei* to evaluate their whole-cell activity/effect on parasite growth and against *T. brucei* overexpressing *Tbr*DUB1 to confirm on-target activity. Selected compounds were also tested against the related parasites *Leishmania donovani* and *T. cruzi* for cross-species activity, and in cytotoxicity studies against the mammalian cell line HepG2. A final selection of prioritised hits and their physicochemical properties is presented.

5.1 Secondary biochemical assays

Compound interference is an unavoidable nuisance in most, if not all, HTS. Successful characterisation of hits from an HTS involves identifying what is undesirable and also what is of value. There are several modes of compound interference-both physical and chemical. Physical modes of interference could include fluorescence from the compound itself (autofluorescence) and compound aggregation. Autofluorescence is typically seen at the lower end of the visible spectrum. Since the fluorophore (5-TAMRA) in the Ub-FP substrate shows emission at a high wavelength (579 nm), we observed very few compounds with autofluorescence, which were removed. The strict quality control measures for the compound library maintenance and dispensation endeavour to remove compounds forming aggregates.

Chemical interference happens if a compound interacts with the buffer components and causes non-specific inhibition of the enzyme due to reactive products of such interactions. Deubiquitinases are cysteine proteases. Cysteine-containing enzymes are susceptible to DTT mediated oxidation of the cysteine by single electron donors (Lor *et al.*, 2007) and oxidative inactivation by hydrogen peroxide that is generated by redox-active compounds in the presence of a strong reducing agent like DTT (Johnston *et al.*, 2008, Mirkovic *et al.*, 2011, Johnston, 2011, Soares *et al.*, 2010). Redox compounds are undesirable non-tractable hits because of known toxicity (Rana *et al.*, 2013). In order to identify and remove such compounds from the *Tbr*DUB1 hits, we tested them in eleven-point dose response (100 µM maximum conc., 1:3 serial dilution) in four complementing biochemical assays. Three of them were performed using HTS conditions, each with one of the following modifications:

1) The assay buffer lacks DTT.
2) The assay buffer additionally contains superoxide dismutase (SOD) and catalase to convert potentially formed superoxide radicals and hydrogen peroxide to water and molecular oxygen.

3) DTT in the assay buffer is replaced by the weaker reducing agent L-cysteine.

Compounds that remain active against *Tbr*DUB1 in these assays are considered valid hits, while inactive ones are flagged as presumably redox-active compounds. In the fourth assay ([Lor *et al.*, 2007](#)), the compounds are incubated with resazurin in the presence of DTT. Redox-active compounds catalyse the DTT mediated reduction of resazurin to the highly fluorescent resorufin and can be easily identified through measurement of fluorescence intensity. The compounds that show activity in this assay will be deemed undesirable and removed from the final selection of compounds.

When working with targets from a pathogen, another important filter is activity against a human orthologue of the target, if known, or against a relevant class of human enzymes. *Tbr*DUB1 belongs to the large family of cysteine proteases, which are named after the conserved cysteine residue in their active site. Compounds that are active against *Tbr*DUB1 might be non-selective for *Tbr*DUB1 and also inhibit other cysteine proteases in the parasite and, more importantly, in the human host where they might lead to toxicity through various off-target effects. To exclude such compounds, we also tested the *Tbr*DUB1 hits in 11-point dose-response (100 µM maximum conc., 1:3 serial dilutions) against human Cathepsin B, a lysosomal cysteine protease.

The results of these five assays are summarized in **Table 5.1** and **Figure 5.1**. As expected, a large number of compounds showed high potencies in the resazurin to resorufin redox assay, highlighting the importance of this assay as a filter. Several compounds showed bell-shaped dose response curves in this assay. Thus, instead of pIC_{50}, the maximum response > 30% in this assay is used as a cut-off for marking redox active compounds. Most compounds that were active against *Tbr*DUB1 in the primary assay were inactive when DTT was replaced with the weaker reducing agent L-cysteine. 173 of the 2143 tested compounds showed a $pIC_{50} > 4$ in this assay, suggesting that most hits from the primary screen inhibit the

enzyme via an undesirable mechanism of action. A total of 1970 compounds were either inactive of failed to give a response > 30% at any concentration tested in the L-cys assay. Similar results were obtained from the assay with superoxide dismutase and catalase, confirming a significant number of non-specific compounds. As expected, the assay performance was poor in the absence of DTT due to a small assay window, which is attributed to the lower enzyme activity in a non-reducing environment. The results from this assay are not considered when finally selecting compounds, however this assay could again provide activity information for some compounds which will be useful for the chemists to make an informed decision when optimising the hits, at a later stage. Lastly, 82% (1785) of the tested compounds were inactive against human cathepsin B, indicating that most of the *Tbr*DUB1 hits might have specific activity against DUBs.

Together, the results demonstrate that a large number of hits show undesirable MOAs and could have liabilities due to their activity against a host cysteine protease. This highlights the importance of including secondary assays in the characterisation of identified hits.

Table 5.1 The Z' and distribution of potencies of *Tbr*DUB1 hits in different secondary biochemical assays.

Assay	L-Cys	Sod/Cat	- DTT	Redox	hCTSB
Robust Mean Z'	0.4	0.48	-0.21	0.62	0.75
Robust SD Z'	0.04	0.02	0.12	0.04	0.06
$pIC_{50} \leq 4$	1970	2037	1274	899	1785
$4 < pIC_{50} \leq 5$	146	76	665	290	332
$5 < pIC_{50} \leq 6$	27	30	194	606	55
$6 < pIC_{50} \leq 7$	0	0	8	302	2
$7 < pIC_{50} \leq 8$	0	0	2	42	0
$8 < pIC_{50}$	0	0	0	3	0
Total	2143	2143	2143	2142	2174
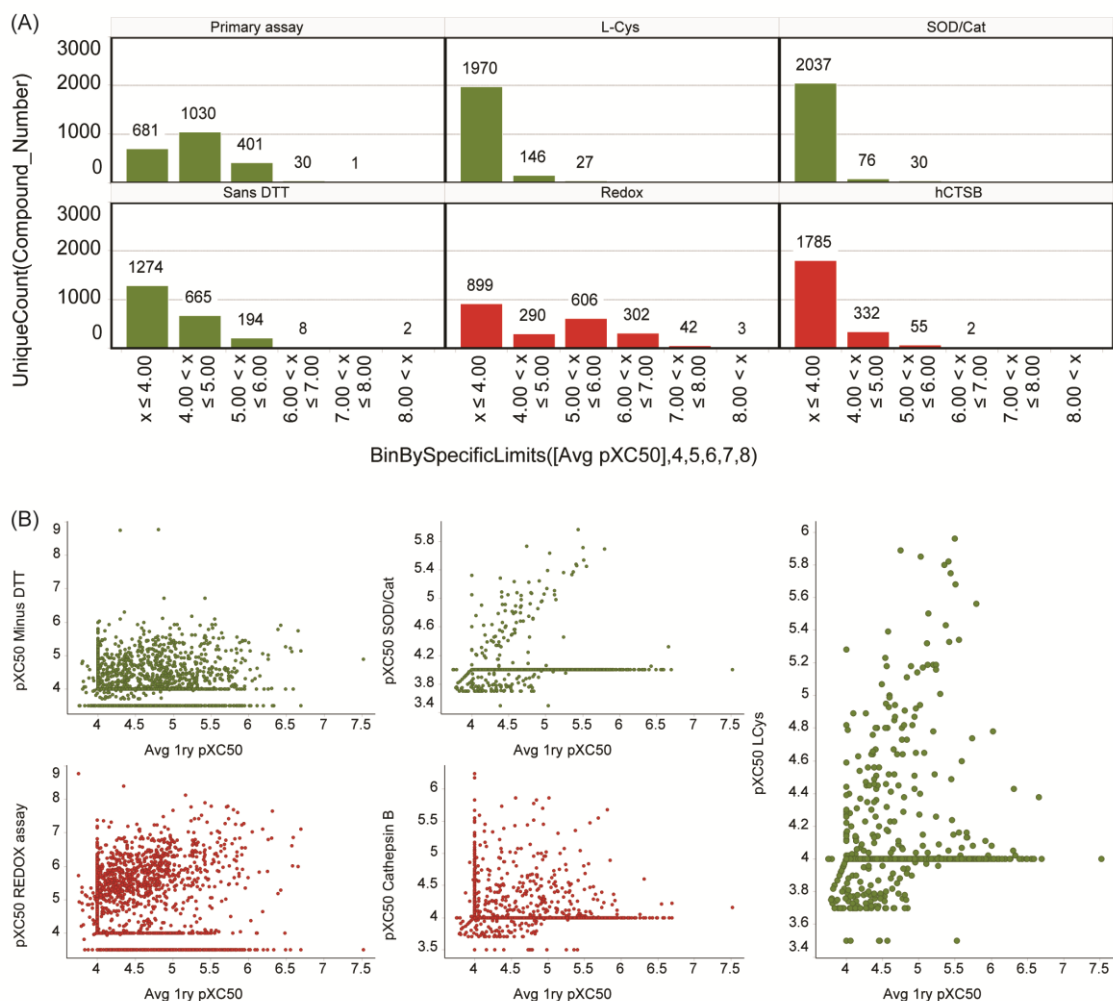
Figure 5.1: pIC$_{50}$ values in secondary biochemical assays.
(A) Numbers of compounds showing pIC$_{50}$ values in the ranges binned by a pIC$_{50}$ difference of 1 in different secondary biochemical assays. (B) Comparison of observed pIC$_{50}$ values from the different biochemical secondary assays to those from the primary (1ry) biochemical assay. Marked in green are the hit validation assays and in red the hit invalidation assays.

We filtered the hits by overlaying the results of the primary and secondary assays (**Figure 5.2**). pIC$_{50}$ values from the L-cysteine assay were used as indicator of compound activity with no undesirable mechanism of action. Compounds that were inactive against human cathepsin B (pIC$_{50} \leq 4$) and those that showed a maximum response of less than 30 % in the resazurin redox assay were selected. The selectivity window (in pIC$_{50}$) was defined as:

$$Selectivity\ window\ (SW) = pIC_{50}\ Lcys\ assay - pIC_{50}\ hCTSB$$

with the following tiers:

$SW \geq 1$: Most selective

$0.5 \leq SW < 1$: Selective

$0 < SW < 0.5$: Least selective

The analysis led to 20 most selective compounds, 23 selective compounds and 21 least selective compounds. The highest observed selectivity window was a pIC_{50} difference of 1.89.

To summarise, starting from a library of ~1.7 million compounds, we identified 64 *Tbr*DUB1 hits which do not show redox MOA and do not show activity against a host cysteine protease.
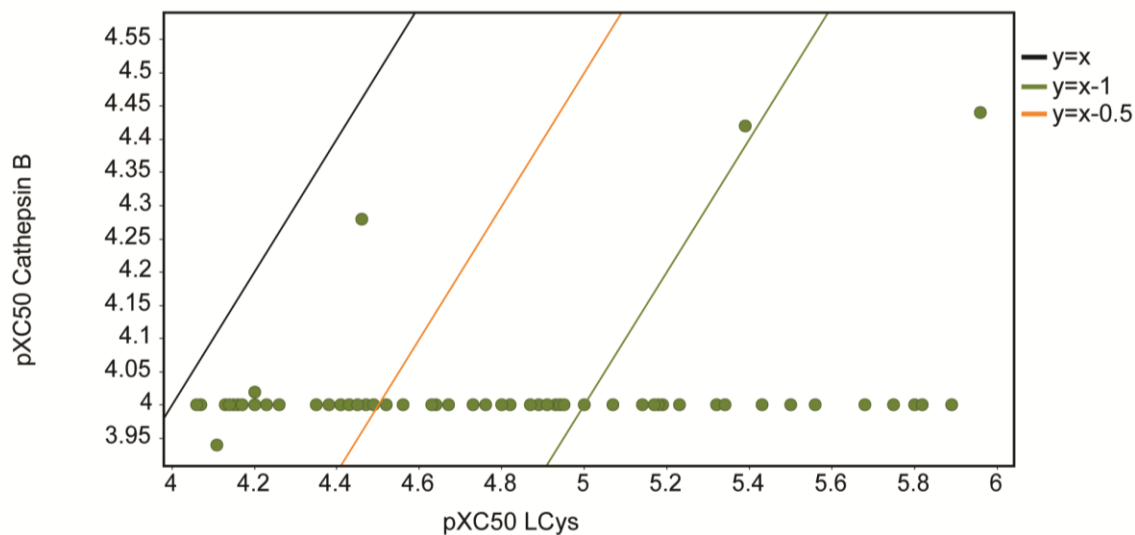
Figure 5.2: Hit selection.
pIC_{50} values in the L-Cys assay and the selectivity windows of the selected 64 *Tbr*DUB1 hits.

5.2 Whole-cell assays

5.2.1 Compound activity against wild-type versus DUB-overexpressing *T. brucei*

We checked the cidal activity of *Tbr*DUB1 hits against *T. brucei* and compared it to the activity against *T. brucei* cells overexpressing the *Tbr*DUB1. The rationale (**Figure 5.3**) for this assay is that if the compound acts on-target and is cidal, overexpression of the target DUB should make the parasite less sensitive to the compound. This is measured as the shift in pIC_{50} values between wild-type and DUB overexpressing parasites. However, in case of redundancy amongst parasite DUBs or if the levels of overexpression are low, the shift in pIC_{50} is expected to be subtle.
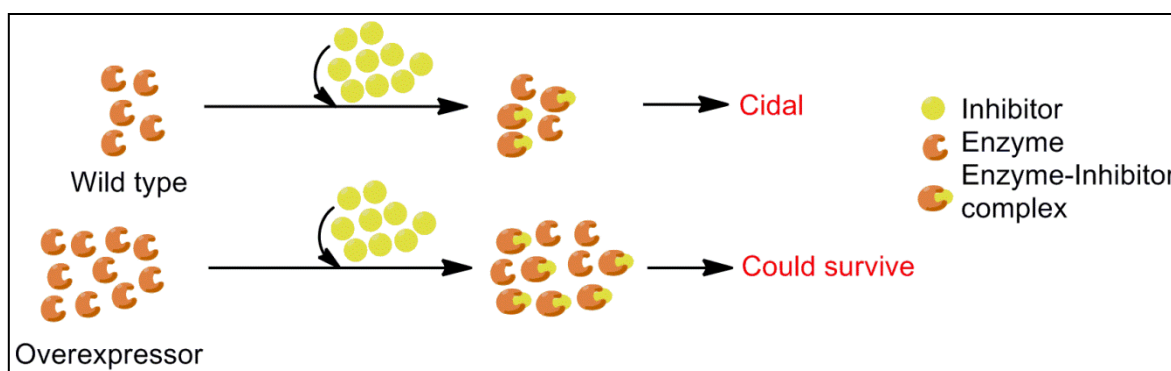
Figure 5.3: The principle of *T. brucei* whole cell DUB-overexpressor *vs.* wild-type assay.

5.2.1.1 Demonstration of myc$_6$ *Tbr*DUB1 overexpression

Mutant 2T1 *Trypanosoma brucei brucei* cell line derived from strain s427 were kindly provided by Dr. Boris Rodenko. The mutant contains the *Tbr*DUB1 gene under the control of a tetracycline operator in the ribosomal RNA locus on chromosome 2a (Jones *et al.*, 2014, Alsford and Horn, 2008). This allows tetracycline induced overexpression of the myc$_6$*Tbr*DUB1 gene product. As described in the section 2.3.10, overexpression of myc$_6$*Tbr*DUB1 was confirmed in

the parasite cells used for screening using SDS-PAGE and western blotting. The uninduced (tet-) cells do not show any expression of myc$_6$TbrDUB1, while induced cells (tet+) show a band at the expected molecular weight when probed with anti-myc tag antibodies indicating the expression of myc$_6$TbrDUB1 (**Figure 5.4**). Furthermore the myc$_6$TbrDUB1 band intensity is comparable when compared for equivalent cell lysate pre-screening and after another 24 hours- the duration of the screening. Using activity based protein profiling (ABPP) it has been shown that this mutant line shows 2-3 fold overexpression of TbrDUB1 (data not shown, oral communication from Dr. Boris Rodenko).
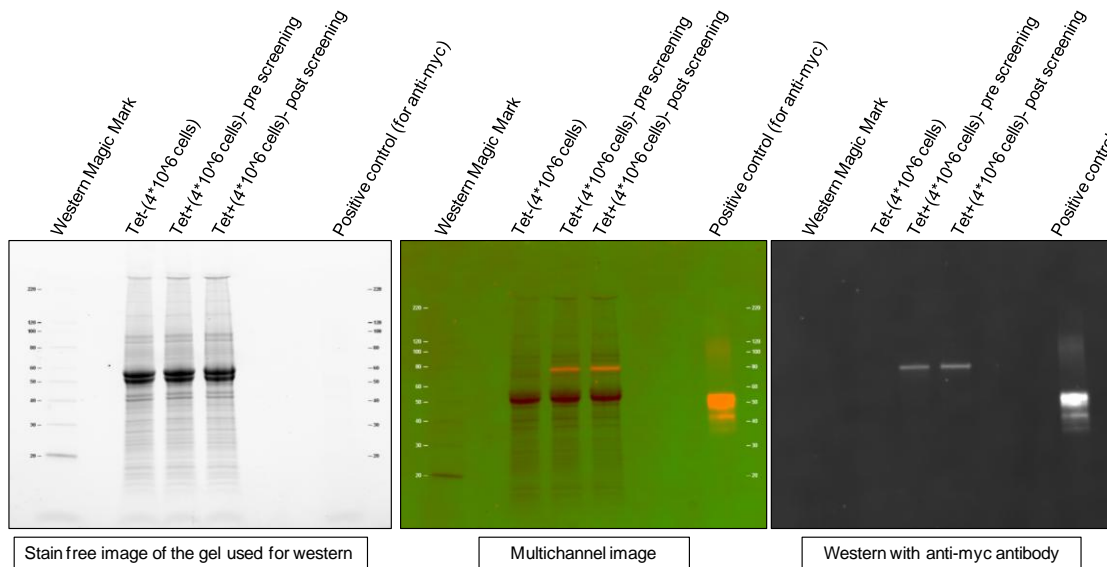
Figure 5.4 Overexpression of myc$_6$TbrDUB1 in the parasites.
The band intensities in stain free image and also in the stain free channel in the second panel serve as loading controls. The positive control for the anti-myc antibody is a commercial *E. coli* lysate containing a protein expressing the myc$_6$-tag. Pre and post screening timepoints show similar myc$_6$TbrDUB1 (measured using Biorad Image Lab 5.1 software package).

5.2.1.2 Screening

We tested the effect of all ~2000 TbrDUB1 hits on the parasite using an ATP based viability assay described in (Sykes and Avery, 2009) and used as an

orthogonal assay in (Pena *et al.*, 2015). In brief, the screening was performed in a 1536-well plate at a final volume of 6 µL in the presence of 25 nL of compound in DMSO. Since the assay has a tolerance limit of 0.4% towards DMSO (determined previously) and since the highest available compound stock concentrations are 10 mM, the highest compound concentration tested was 41.67 µM, followed by ten 1/3 dilutions. In the case of overexpression of myc$_6$ *Tbr*DUB1, it was induced with 1 µg/mL tetracycline (from a 1000x stock in 70% EtOH) at least 96 hours before screening. Cells were maintained in log-phase of growth without selection antibiotics for at least 24 h before exposing them to the compounds. 300 cells in 6 µL growth medium were dispensed per well and after 24 h of incubation with compound, viable cells were quantified by measuring the ATP content in each well using Promega CellTitre-Glo luminescent cell viability assay reagent. The luminescence was measured using ViewLux microplate readers. The assay performed robustly with a Z' of ~0.8 and was quite reproducible as seen by a good correlation of the pIC$_{50}$ values. Owing to the absence of known specific *Tbr*DUB1 inhibitors, we tested the performance of the assay using 27 trypanocidal compounds (commercial and GSK compounds previously identified in a phenotypic screen) and the pan-DUB-specific inhibitor H1, which was confirmed to be active against *Tbr*DUB1 (section 3.4.1). However, H1 will affect parasite viability by non-selectively inhibiting several DUBs and thus we would not expect a pIC$_{50}$ shift with H1. As can be seen in **Figure 5.5**, the correlation of tool compound performance in wild-type and DUB-overexpressing parasites is excellent. These trypanocidal compounds are expected to affect parasite viability in a DUB-independent fashion and thus not show a shift in the pIC$_{50}$ values. Several of the trypanocidal compounds showed a slightly higher potency in DUB-overexpressing parasites, suggesting that overexpression of *Tbr*DUB1 might impact parasite fitness. We thus believe that a small increase in the pIC$_{50}$ value against wild type parasites could mean that the compound is on-target. H1 showed an average pIC$_{50}$ of 4.9 and 5.2 in the wild-type and DUB-overexpressing parasites, respectively. As expected, it does not show a pIC$_{50}$ shift when comparing the two conditions.
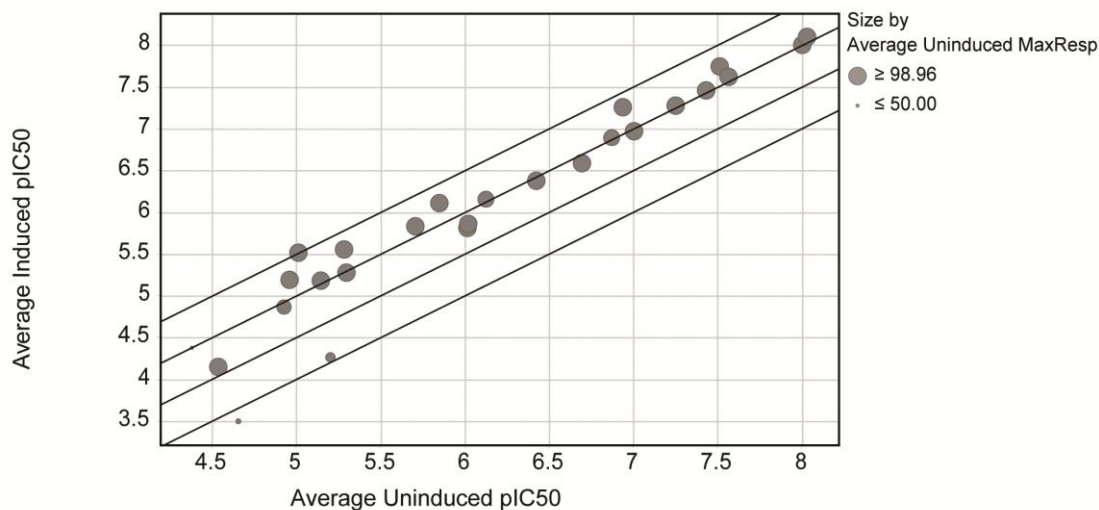
Figure 5.5: Correlation of pIC$_{50}$ values of trypanocidal compounds and H1 in the whole-cell assay.
Wild-type (uninduced) and DUB-overexpressing (induced) parasites were treated with trypanocidal compounds to evaluate the performance of the assay. The sizes of the points vary depending on the maximum response in the wild-type parasites. The lines are drawn at y = x+0.5, x, x-0.5 and x-1 to see a pIC$_{50}$ shift, if any.
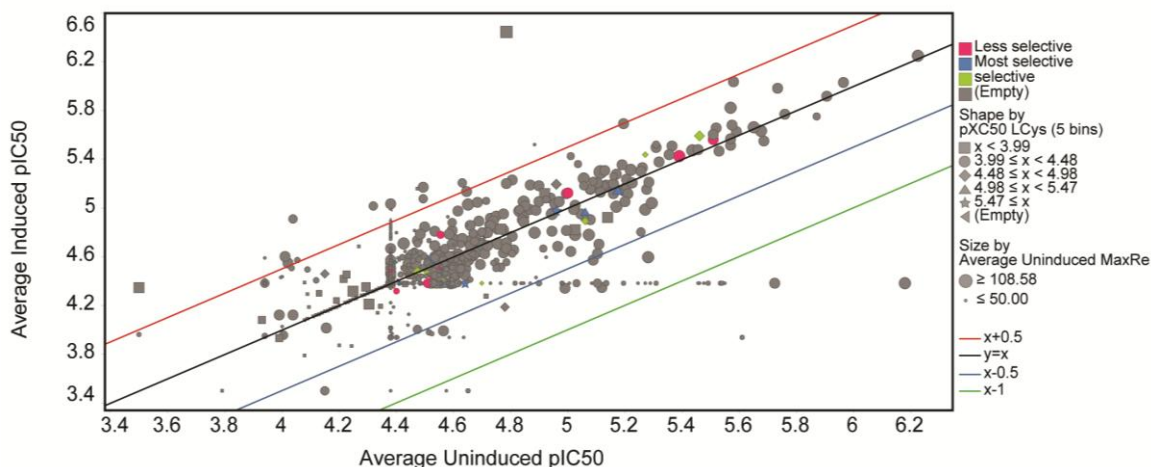
Figure 5.6: Correlation of pIC$_{50}$ values of *Tbr*DUB1 hits in the whole-cell assay.
All compounds were tested in the whole cell assay comparing the pIC$_{50}$ values in wild type and parasites overexpressing the DUB. The data shown above are average of pIC$_{50}$ values observed in 2 replicates. The colouring is according to the definitions in section 5.1 for most selective, selective and least or less selective compounds. The shape is defined by the pIC$_{50}$ observed in the L-cys assay. The size is defined by the maximum response observed against the wild-type parasites.

128
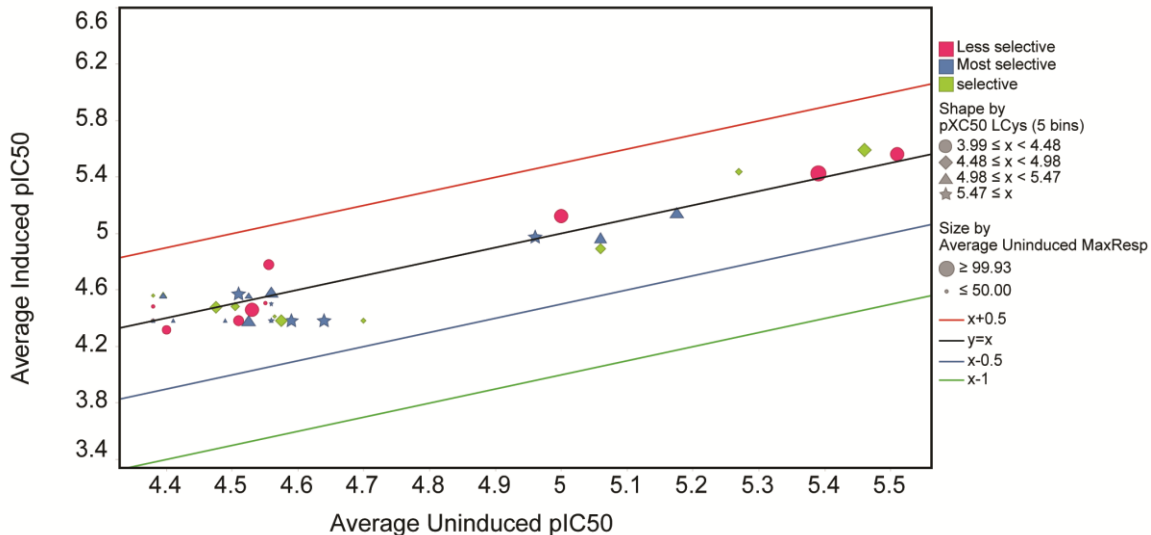
Figure 5.7: Response of the 64 selected hits in the whole cell assay.
The colouring is according to the definitions in section 5.1 for most selective, selective and less selective compounds. The shape is defined by the pIC_{50} observed in the L-cys assay. The size is defined by the maximum response observed in the wild-type parasites.

Shown in **Figure 5.6** are the potencies of all *Tbr*DUB1 hits against the wild-type and DUB-overexpressing parasite. pIC_{50} values of up to 6.2 were observed. In **Figure 5.7** we show the 64 compounds that passed all the biochemical filters. The most potent compounds in the whole cell assay fail in some of the biochemical assays and hence did not get selected. The selected 64 compounds show a low, but decent at hit stage, potency.

We failed to observe big shifts in pIC_{50} values from wild type to overexpressor. This could be due to a) low-level overexpression of *Tbr*DUB1 (2-3 fold), insufficient to confer drug tolerance and cause a change in the pIC_{50} value, b) off-target activity or c) fitness loss, *i.e.* increased drug sensitivity, of *Tbr*DUB1 overexpressing parasites which would lead to a reverse shift of pIC_{50} values, yielding a small net shift. Considering that we observed the trypanocidal compounds to show higher potencies against *Tbr*DUB1-overexpressing parasites, the latter scenario seems plausible. Nevertheless, we have identified *Tbr*DUB1 hits which show a cidal effect against *T. brucei*.

5.2.2 Cytotoxicity profiling using HepG2 cells

To obtain information about possible acute *in vitro* cytotoxicity of the *Tbr*DUB1 hits, we tested them on the human liver-derived HepG2 cell line in 11-point dose responses (100 µM max compound concentration; section 2.3.13). Compounds' cytotoxicity is measured by intracellular ATP concentrations after 48 h incubation with compounds, using a luciferin/luciferase reaction. Mitochondrial function and cell redox state are both markers for cell health, and impairment of either result in a decrease in intracellular [ATP] (Crouch *et al.*, 1993). Whether *in vitro* cytotoxicity translates to *in vivo* toxicity depends on a number of factors including therapeutic dose, free plasma concentration, metabolism, excretion etc. However, the compounds that are active in this assay can be flagged as undesirable, unless proven otherwise in *in vivo* studies. . As is seen in **Figure 5.8**, no compound has a $pIC_{50} > 6$ in the cytotoxicity assay and hence all hits qualify the absolute threshold for marking positives in this assay. Looking at relative cytotoxicity response, i.e. pIC_{50} values in the parasite versus HepG2 cells, several compounds, including some of the compounds in the 'most-selective' tier show up to a ~ 3-times lower IC_{50} (~ 0.5 log units higher pIC_{50}) against the parasite. These compounds could be the promising hits to follow up with chemistry efforts to increase potency and improve selectivity.
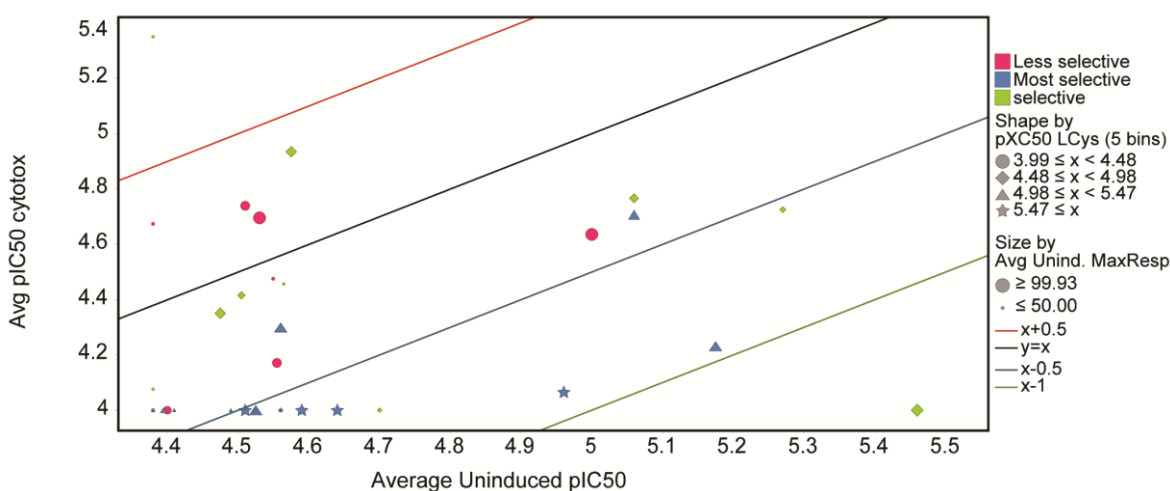
Figure 5.8: pIC_{50} values of the 64 selected compounds in the HepG2 cytotoxicity assay.

5.2.3 *L. donovani* and *T. cruzi* whole-cell assays

The orthologues of *Tbr*DUB1 in the related Kinetoplastid parasites *L. donovani* and *T. cruzi* are highly similar in amino acid sequence, sharing 48 % and 38.6 % identity, respectively, with *Tbr*DUB1 (**Figure 3.4**) It is therefore reasonable to assume that *Tbr*DUB1 hits also inhibit the respective orthologues in *L. donovani* and *T. cruzi*. As described in the materials and methods both the whole cell assays have three readouts each: percentage of infected cells, ratio of parasites per host cell and percent response against the host cell. Thus this assay gives us response against the parasite and host cells which indicates cytotoxicity. These responses are normalised to the appropriate controls and used to calculate the pIC_{50}. Of the 64 prioritized compounds, 61 were available for testing them in dose response against *L. donovani* and *T. cruzi*. As shown in **Figure 5.9**, we see some potent compounds against both the parasite, but most of them also showed activity against host cells. Two compounds were selected that showed significantly higher pIC_{50} against the parasites compared to the host cell. Both compounds belong to the same cluster (cluster 24; **Figure 5.12**) and hence are structural analogues (**Figure 5.23**). Interestingly, they do not cluster with known DUB inhibitor (discussed later, **Figure 5.12**), suggesting that they are novel chemical entities. Furthermore, these compounds have passed a very thorough biochemical profiling and show no cytotoxicity against HepG2 cells, making them promising compounds for progression.
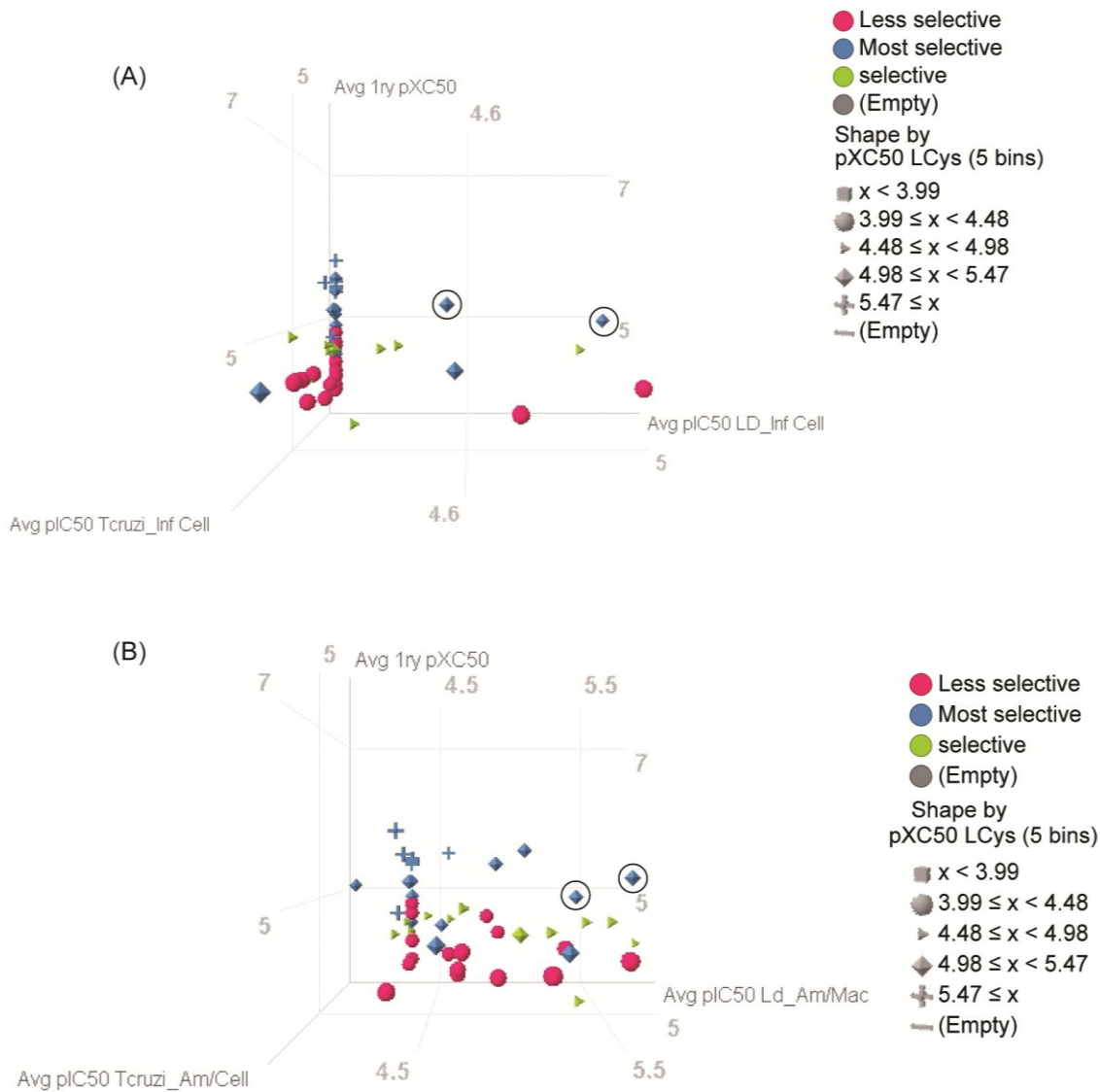
Figure 5.9 Response in *L. donovani* and *T. cruzi* whole cell assays.
The *Tbr*DUB1 hits were tested in whole cell assays and shown here are pIC$_{50}$ values in these assays compared to pIC$_{50}$ in the primary biochemical assay. Since there were 2 outputs against the parasite in the whole cell assay, presented are (A) pIC$_{50}$ values calculated using % infected cells as the output and (B) pIC$_{50}$ values calculated using amastigotes per host cell. The two most promising compounds are circled.

5.3 *In-silico* clustering of compounds

In order to identify structural analogs amongst *Tbr*DUB1 hits, we performed a complete-link hierarchical clustering of the 64 prioritized compounds, based on the tanimoto distance ([Leach and Gillet, 2007](#)). Studying the differences in the

132

potencies of structural analogs in different assays provides very valuable SAR information for the chemists to try to optimise the potency and selectivity of the compounds. For clustering, a digital fingerprint is generated for each compound based on its chemical structure and then a tanimoto score is obtained based on similarity of these digital fingerprints. Tanimoto score can range between 0 and 1, with 1 indicating highest similarity. For the clustering we defined a cut-off of tanimoto score > 0.55 for compounds to fall in the same cluster. As shown in **Figure 5.10**, we identified eleven clusters and 26 singletons. The largest family contains nine compounds, followed by a family of six and four compounds, respectively, and several families with two or three compounds each. It is interesting to note that within some families there are compounds with higher and lower selectivity and compounds from same cluster show different potencies in the *T. brucei* whole cell assay (**Figure 5.11**). This offers hints towards the SAR based on the functional groups these compounds contain.
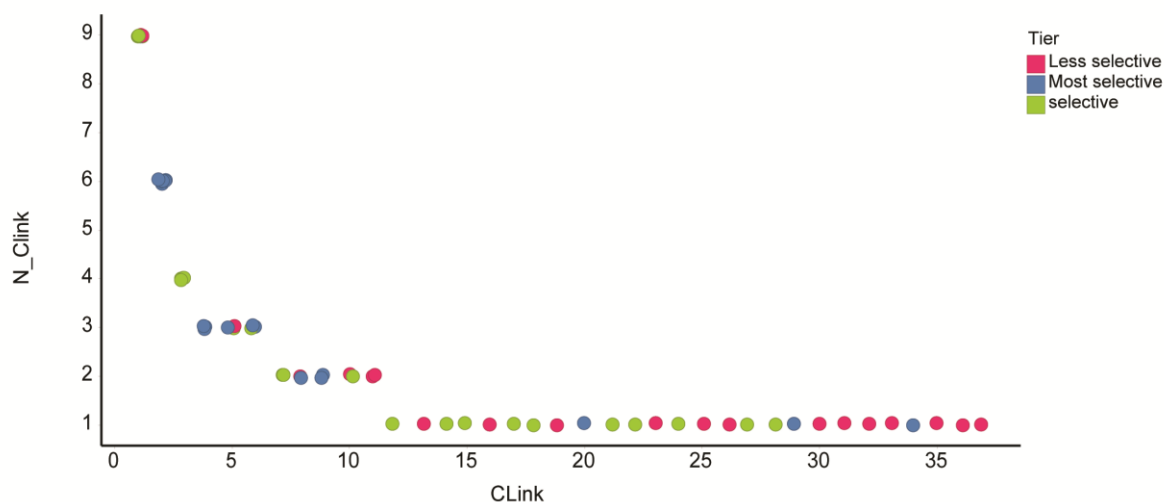
Figure 5.10: Complete-link hierarchical clustering of the 64 selected hits. X-axis (Clink) is the cluster in which the compounds fall and the y-axis (N_Clink) is the number of compounds per cluster. The colours are defined by the selectivity window cut-offs described above.
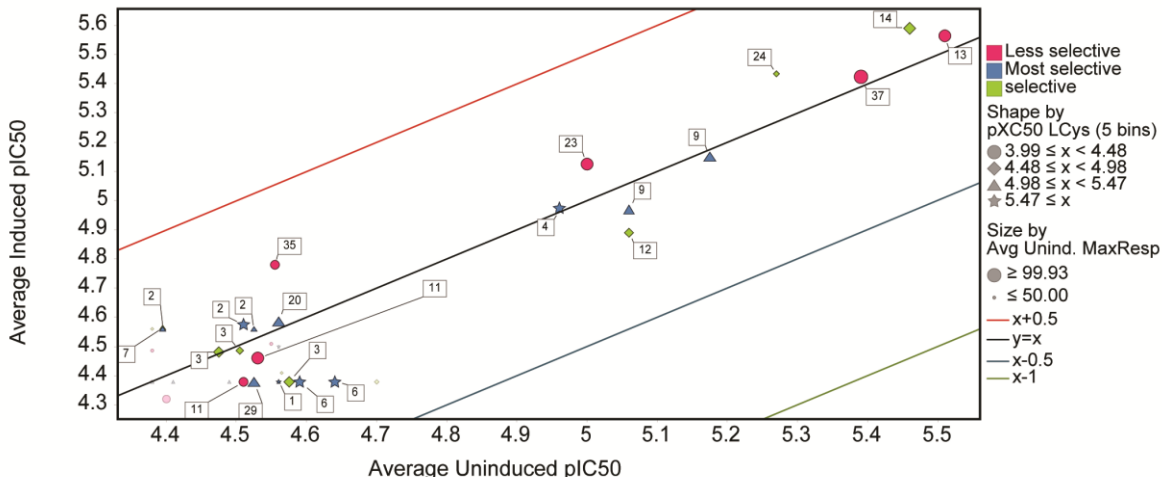
Figure 5.11: Compounds from same cluster show different but similar potencies.

The next question we tried to answer was whether the discovered *Tbr*DUB1 hits are new chemotypes from the GSK library or are they similar to known DUB inhibitors. For this, a clustering attempt was made with known inhibitors of DUBs and *Tbr*DUB1 hits. The list of known DUB inhibitors that were used in this clustering is presented in **Table 1.2**.

Only two of our hits clustered each with one of the known USP7 inhibitors: spongiacidin C (**22**) and HBX19818 (**23**). All the other *Tbr*DUB1 hits clustered into different clusters or as singletons. The results are shown in **Figure 5.12**. Cluster numbers 15 and 16 represent the 2 hits that cluster with spongiacidin C and HBX19818 respectively. It can be concluded that most of the *Tbr*DUB1 hits that we have discovered new families of compounds as *Trypanosoma brucei* deubiquitinase inhibitors. Whether these compounds are selective against *Tbr*DUB1 compared to human DUBs still needs to be tested.
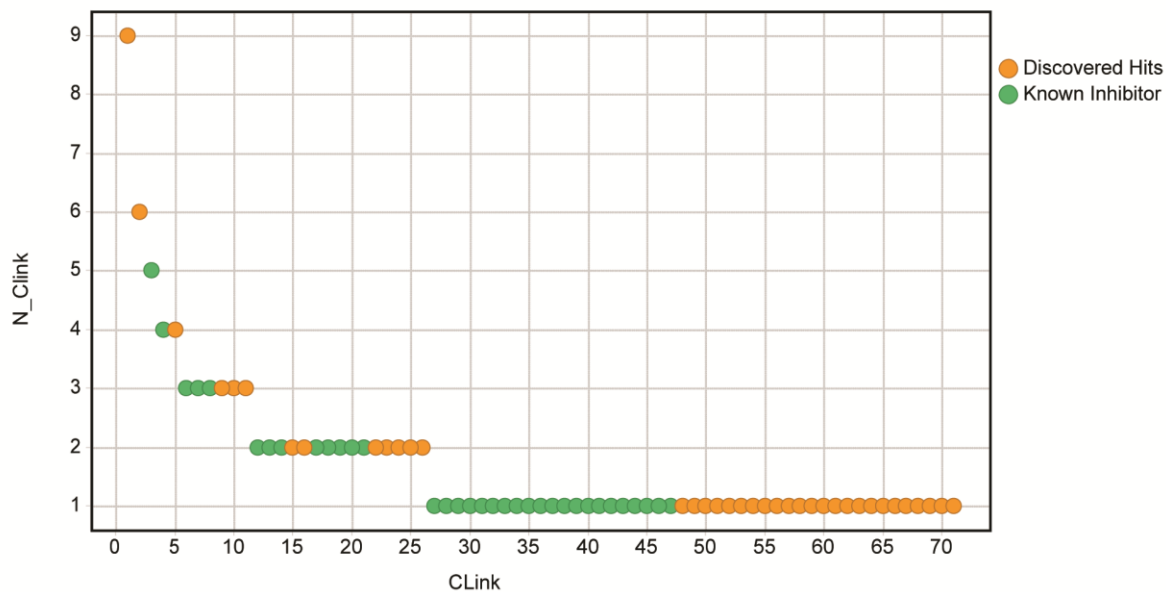
Figure 5.12: Clustering of 64 selected hits with known inhibitors of DUBs.
Only 2 of our hits cluster with known inhibitors: spongiacidin C and HBX19818 shown in clusters 15 and 16 respectively. Both of these are USP7 inhibitors.

5.4 Physicochemical properties of the selected compounds

The physicochemical properties of these compounds were checked to look for potential red flags due to undesirable properties. The physiochemical properties have been described and discussed in detail in the previous chapter. The different physicochemical properties of the 24 *Tbr*DUB1 hits out of 64 selected after biochemical assays (**Figure 5.2**) can be seen in **Figure 5.13** - **Figure 5.18**. These 24 hits are chosen based on their pIC_{50} values and maximum response against wild-type parasites. The physicochemical properties presented are clogP, molecular weight, number of aromatic rings, number of hydrogen bond acceptors, number of hydrogen bond donors and property forecast index (PFI). Overall, the *Tbr*DUB1 hits have desirable physicochemical properties and, if progressed, offer a chance of chemistry efforts for further optimisations. Additionally, from the distribution of the compounds we can conclude that our assay is not biased for any particular physicochemical property.
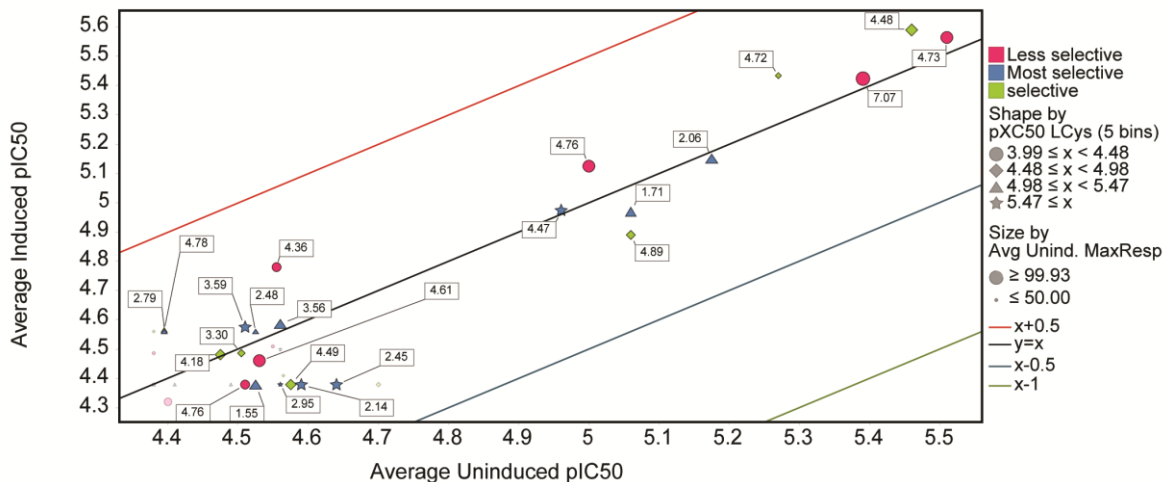
Figure 5.13: clogP of selected compounds
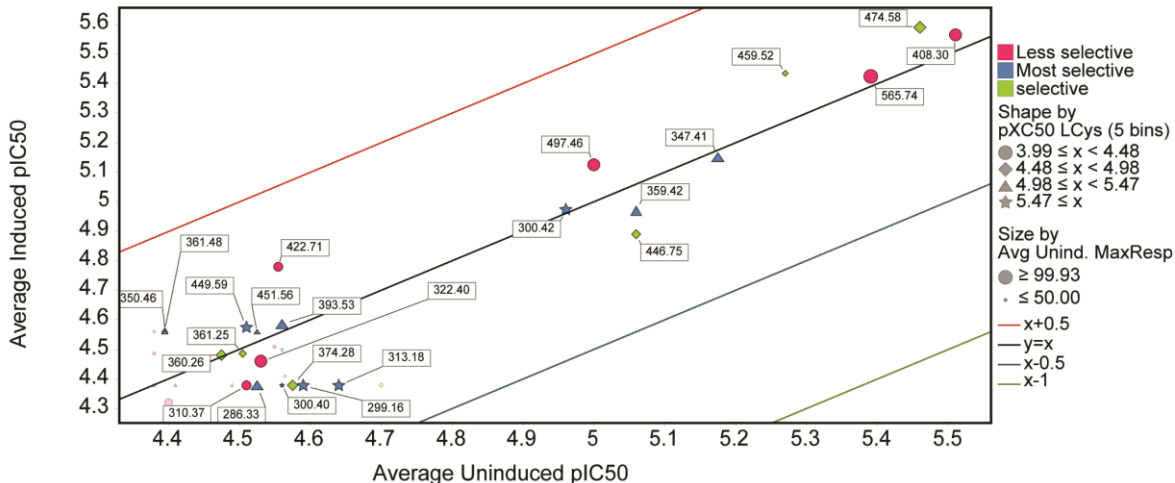
Figure 5.14: Molecular weight of selected compounds
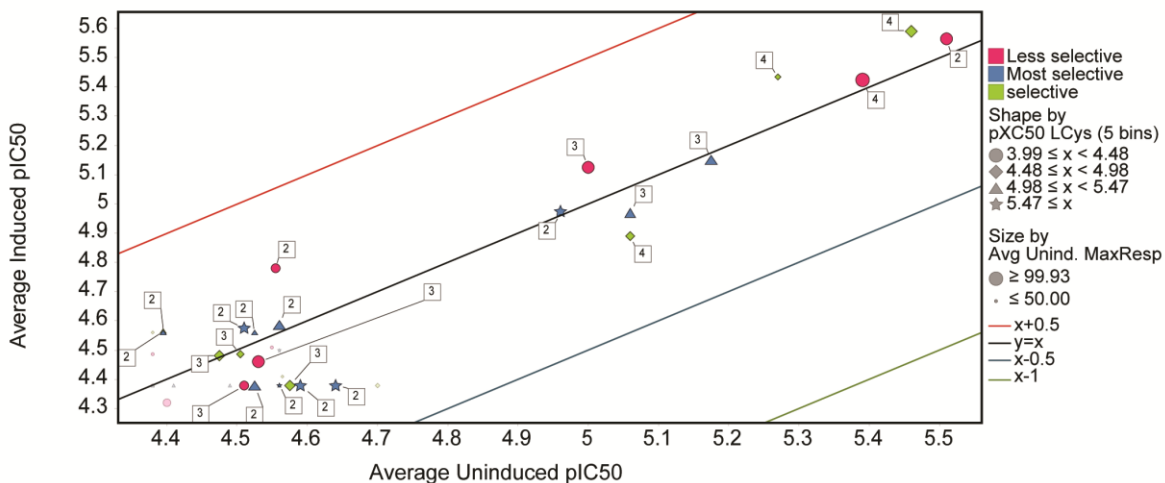
Figure 5.15: Number of aromatic rings in selected compounds
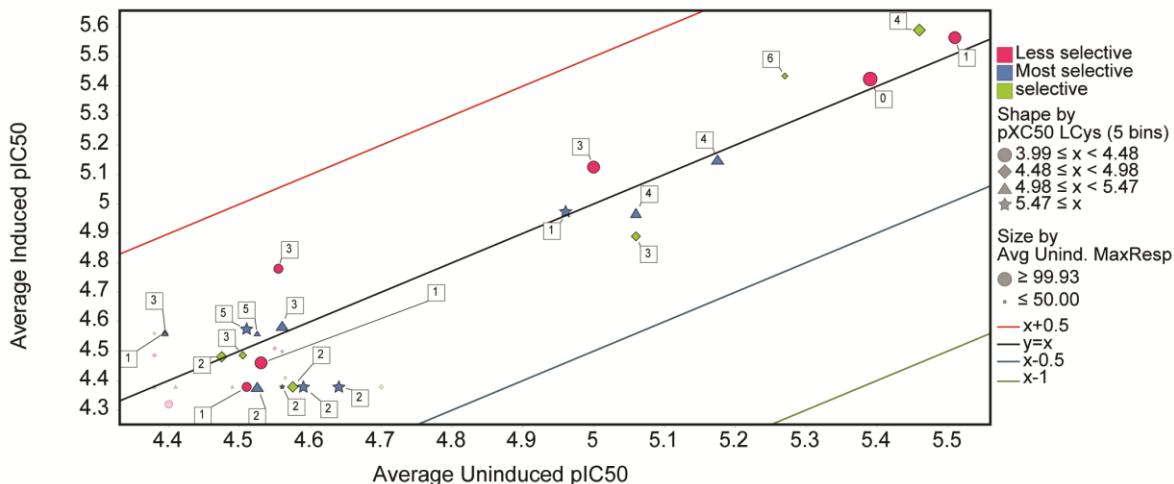
Figure **5.16: Number of hydrogen bond acceptors in selected compounds**
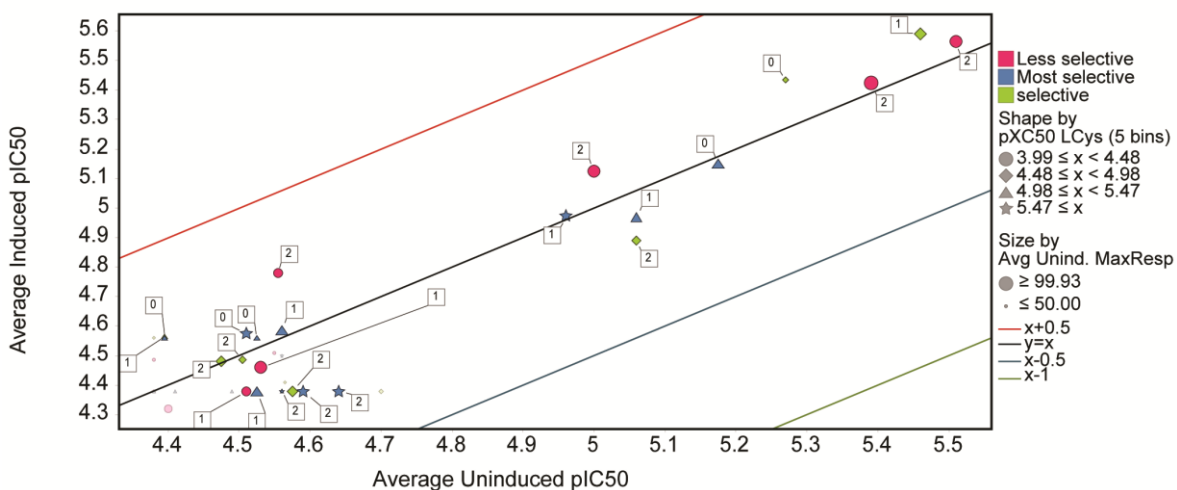
Figure 5.17: Number of hydrogen bond donors in selected compounds
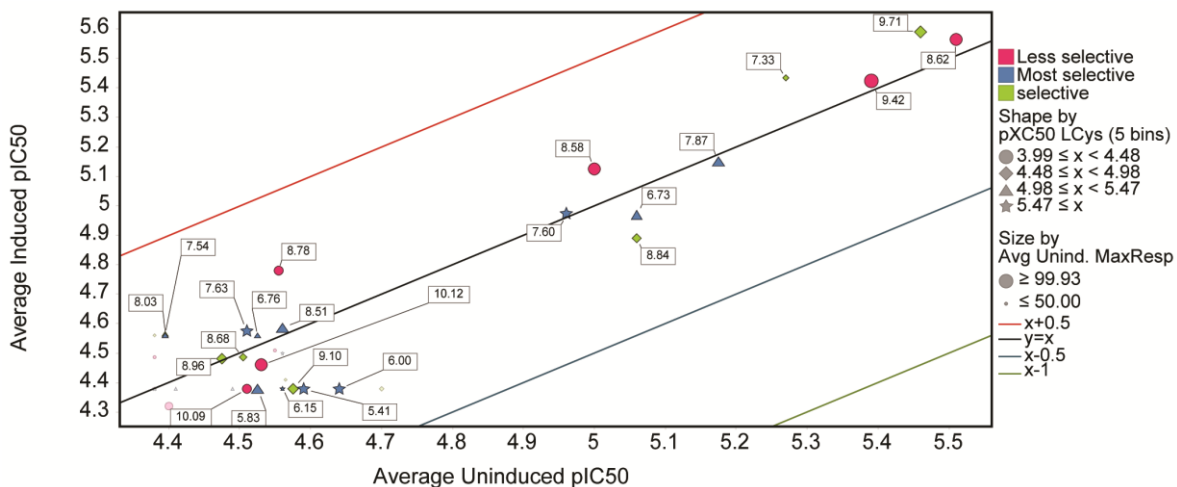
Figure 5.18: PFI (property forecast index) of the selected compounds

We also retrieved information about the promiscuity of *Tbr*DUB1 hits from historical data from the HTS carried out at GSK in the past. Inhibition Frequency Index (IFI) is defined as the % of assays in which a compound showed > 50 % inhibition in HTS screening run at 10 µM.

$$IFI = \frac{B}{A}\%$$

Where

A= Number of % inhibition instances recorded in unique non kinase Assays

B= Number of times the Median value of % inhibition > 50 %

IFI values below 5% are considered "clean", between 5 and 10% "suspicious", and > 10% are "potential nuisance". It is important to note that the statistical significance of IFI values should be judged by the number of tests recorded per compound. As a rule of thumb, at least 50 tests should be available to obtain an accurate measure of IFI, though extreme values with lower counts can be meaningful as well. The IFI values for *Tbr*DUB1 compounds are shown in **Figure 5.19**. The number of assays used to calculate the IFI values are shown in **Figure 5.20**. *Tbr*DUB1 hits have small IFI values which increases the confidence that they could be on target.
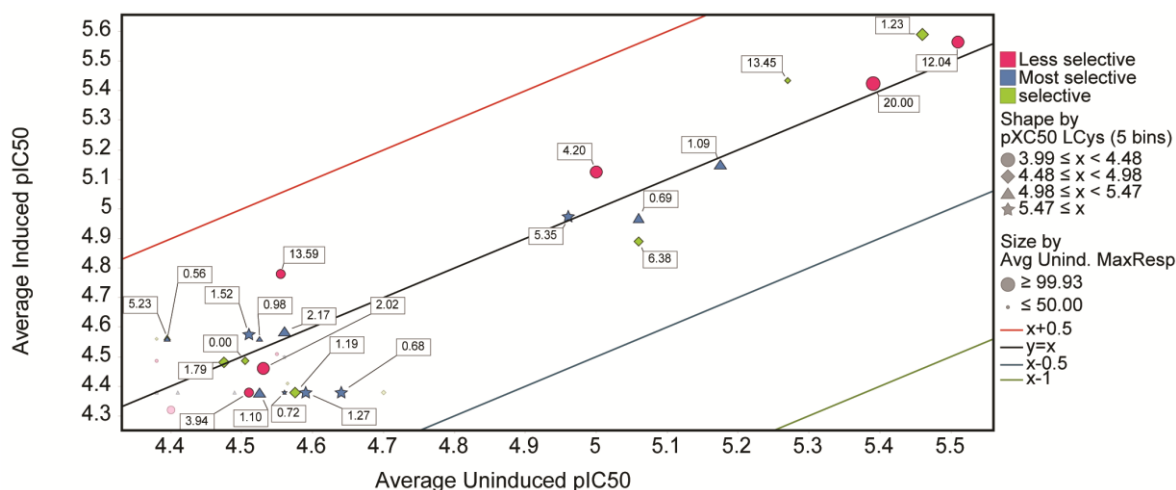
Figure 5.19: IFI (inhibition frequency index) of the selected compounds.
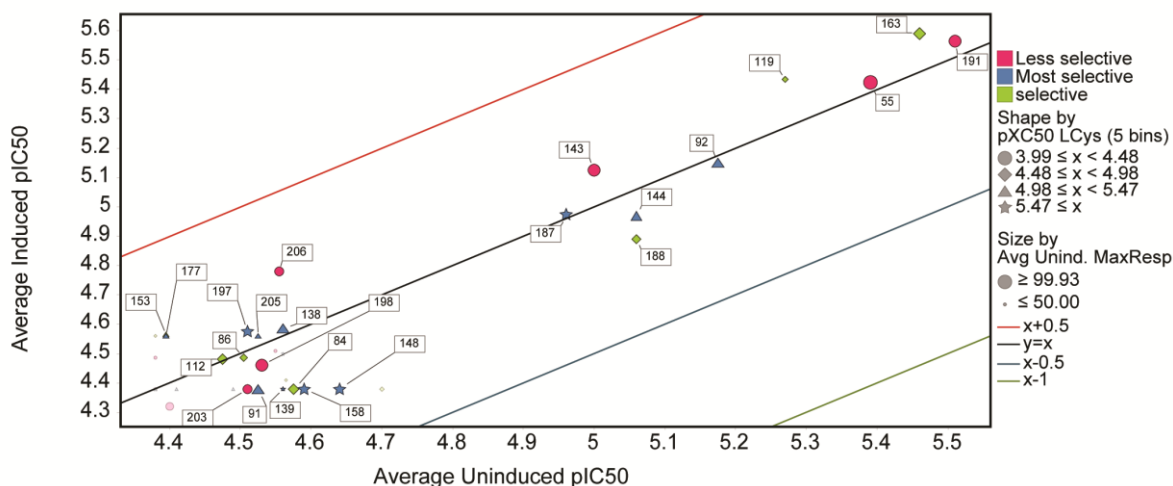
Figure 5.20: Assay count used to calculate the IFI for the selected compounds.

Lastly, we compared the distribution of physicochemical properties of 64 *Tbr*DUB1 hits to those of 771 marketed drugs. As can be seen in **Figure 5.21**, the distribution of physicochemical properties of *Tbr*DUB1 hits overlaps with the distribution of marketed drugs which gives confidence that the selected 64 *Tbr*DUB1 hits are indeed drug-like.
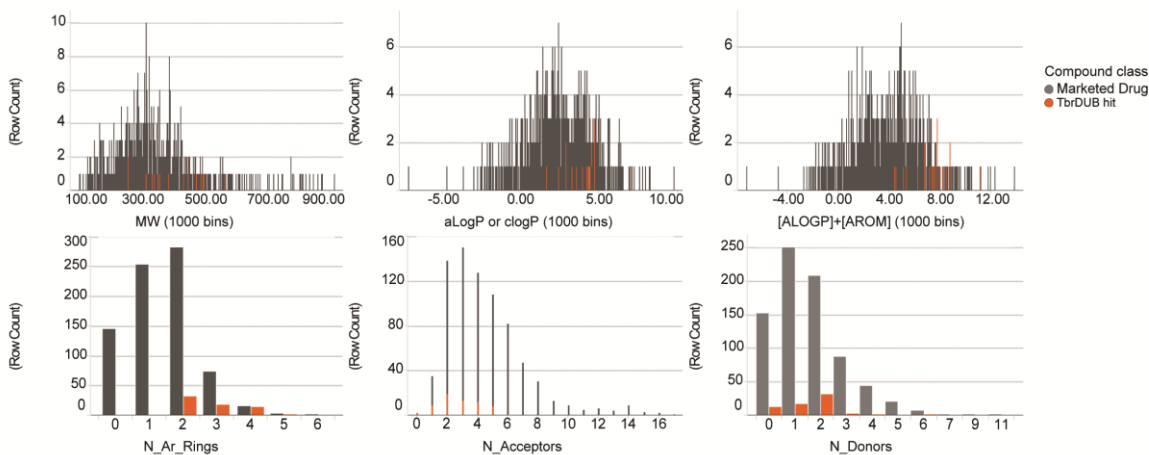
Figure 5.21: Comparison of physicochemical properties of *Tbr*DUB1 hits with those of marketed drugs.
Marketed drugs (grey) and *Tbr*DUB1 selected hits (orange) show a similar distribution of physicochemical properties. *Tbr*DUB1 hits are thus good starting points for medicinal chemistry efforts.

5.5 Selection of high-priority hits

Keeping in mind all the biochemical assays, the *T. brucei* whole assay and the cytotoxicity assay, a selection of the high priority hits was made. These hits stand the highest chance of success if followed up with efforts to increase potency and selectivity and obtain a good dose, metabolism and pharmacokinetics (DMPK) profile.

The criteria used for selection of high priority hits from the 64 that have already been selected (**Figure 5.2**) are:

- $0.4 \leq$ (pIC50 Uninduced-pIC50 cytotox) ≤ 1.46
- Tier: Most selective and selective

A pIC_{50} difference of 0.4 represents approximately 2.5 times difference in the IC_{50} value. Using these criteria we were able to select 14 compounds that have a range of pIC_{50} values against the wild-type *T. brucei* from 4.41 to 5.56. It is interesting to note that three compounds each from cluster numbers 2, 4 and 6 (as shown in **Figure 5.10**) are in the final selection which means these clusters contain valuable SAR information for follow-up studies. Presented in **Figure 5.22** are the structures of the compounds that we consider high priority.
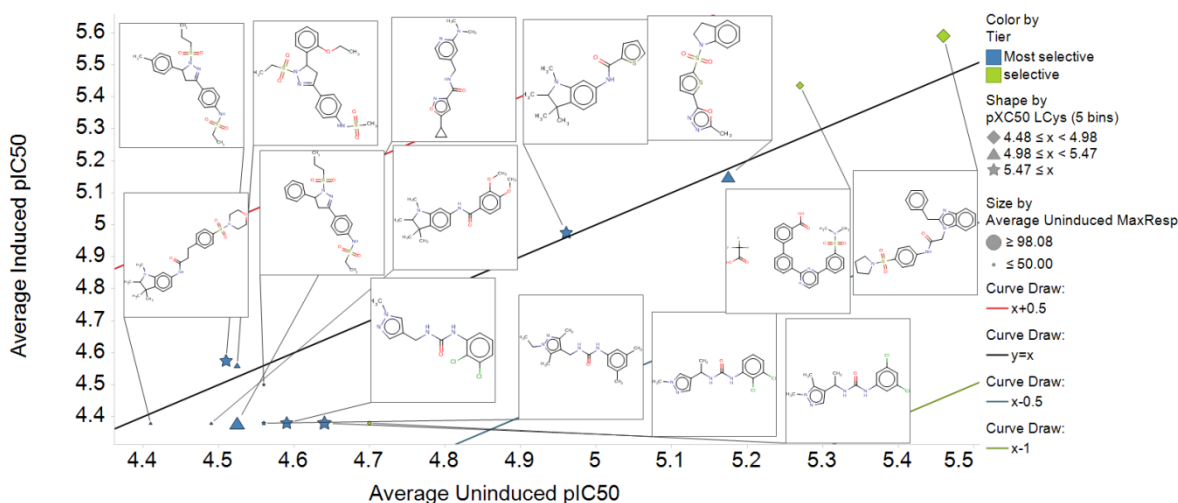
Figure 5.22: Structures of the high priority compounds.

Lastly, we also selected two most promising *Tbr*DUB1 hits that elicit a response against the three parasites tested *viz. T. brucei*, *T. cruzi* and *L. donovani* and show little or no response against the host cells. They have passed a very thorough biochemical profiling and don't show cytotoxicity against HepG2 cells. The pIC_{50} values of these compounds are shown in **Table 5.2** and the structures of these compounds are shown in **Figure 5.23**. These compounds have IC_{50} values of 3.98 and 1.99 µM, respectively against *L. donovani* and 15 and 10.4 µM, respectively against *T. cruzi*. Both of these compounds are structural analogs. They cluster together in cluster number 24 (**Figure 5.12**) and do not cluster with any of the known DUB inhibitors. They are thiophene phenyl derivatives.

In summary, after a set of five secondary biochemical assays and *T. brucei* whole cell assays, we identified fourteen high-priority compounds. Furthermore, after also including the results against whole cell *L. donovani* and *T. cruzi*, we identified two most promising hits that show activity against all three parasites.

58   59
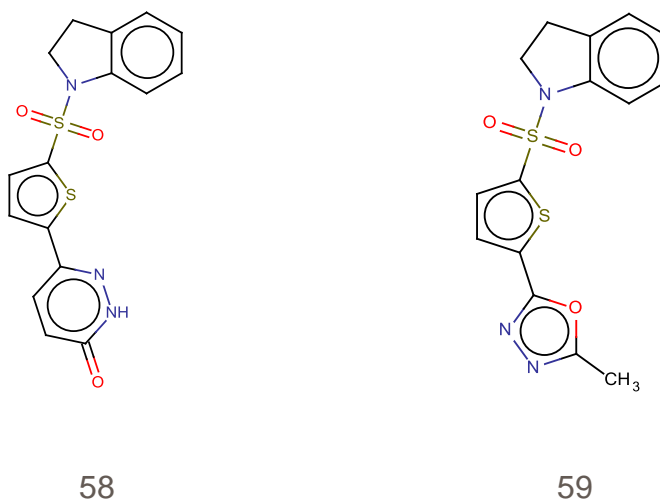
Figure 5.23: Structures of the highest priority *Tbr*DUB1 hits.
These compounds show cidal activity against all the three parasites tested and no cytotoxicity.

Table 5.2 Potencies of the two most promising compounds.

pIC$_{50}$ values of the two most promising compounds in different assays- primary biochemical assay, wild type (WT) and DUB-overexpressing *T. brucei*, *T. cruzi*, *L. donovani* and HepG2 cytotoxicity assay. For *T. cruzi* and *L. donovani* 3 pIC$_{50}$ values are reported calculated respectively from the 3 outcomes as response: % infected cells, amastigotes per host cell and response against host cell.

Compound Number	Biochemical assay	*T. brucei* WT	*T. brucei* DUB overexpressor	*T. cruzi* (%inf / Am per cell / host cell)	*L. donovani* (%inf / Am per cell / host cell)	Cytotox assay
58	5.14	5.06	4.97	4.8/4.4/4.3	5.4/5/4.3	4.7
59	5.19	5.18	5.15	4.98/4.74/4.3	5.7/4.6/4.3	4.23

Supplementary data

For the selected high-priority and most promising compounds (section 5.5), presented here are the structures, pIC$_{50}$ values and physicochemical properties.

Legend for the following tables: tier M: most selective; tier S: selective; CLink: cluster number; N_Ar_Rings: no. of aromatic rings; N_Chiral_Atoms/Bonds: no. of chiral atoms/bonds; clogP: calculated logP; cmr: calculated molar refractivity; N_Acceptors and N_Donors: no. of hydrogen bond acceptors and donors; N_Heavy_Atoms: no. of heavy atoms; N_Neg_Charge and N_Pos_Charge: no. of negative and positive charge; N_Rot_Bonds: no. of rotatable bonds; CNS_Penetration: Score of chances of penetrating the CNS; PFI: property forecast index; Assay_count: no of assays used to calculate IFI which is the inhibition frequency index.

Compound number	Structure
58	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.975	5.14	5.4
pIC$_{50}$ SOD/Cat	**pIC$_{50}$ hCathepsin B**	**pIC$_{50}$ REDOX assay**
5.06	4	4
Selectivity diff.	**Average Induced pIC$_{50}$**	**Average Uninduced pIC$_{50}$**
1.14	4.97	5.06
Avg. pIC$_{50}$ cytotox	**pIC$_{50}$ diff. uninduced and cytotox**	**Tier**
4.705	0.355	M
CLink	**N_Ar_Rings**	**N_Chiral_Atoms/Bonds**
9	3	0
clogP	**cmr**	**Flexibility_Index**
1.714	9.302	11
N_Acceptors	**N_Donors**	**N_Heavy_Atoms**
4	1	24
Molecular_Weight	**N_Neg_Charge**	**N_Pos_Charge**
359.42	0	0
N_Rot_Bonds	**Total_Polar_Surface_Area**	**CNS_Penetration**
3	83.13	
PFI	**Assay_count**	**IFI_50**
6.73	144	0.69
Chemotype		
thiophene phenyl derivative		

Compound number	Structure
59	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.24	5.19	5.62
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.36	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.19	5.15	5.175
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4.23	0.945	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
9	3	0
clogP	cmr	Flexibility_Index
2.064	8.767	11
N_Acceptors	N_Donors	N_Heavy_Atoms
4	0	23
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
347.41	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
3	76.3	
PFI	Assay_count	IFI_50
7.87	92	1.09
Chemotype		
thiophene phenyl derivative		

Compound number	Structure
60	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.5	5.96	3.5
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.71	4.44	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.52	4.975	4.96
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4.065	0.895	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
4	2	1
clogP	cmr	Flexibility_Index
4.465	8.851	8
N_Acceptors	N_Donors	N_Heavy_Atoms
1	1	21
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
300.42	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
2	32.34	High
PFI	Assay_count	IFI_50
7.6	187	5.35
Chemotype		
thiophene amide phenyl derivative		

Compound number	Structure
61	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.435	5.75	5.46
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.96	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.75	4.38	4.59
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.59	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
6	2	0
clogP	cmr	Flexibility_Index
2.138	7.565	15
N_Acceptors	N_Donors	N_Heavy_Atoms
2	2	19
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
299.16	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
3	58.95	High
PFI	Assay_count	IFI_50
5.41	158	1.27
Chemotypes		
urea phenyl derivative		

147
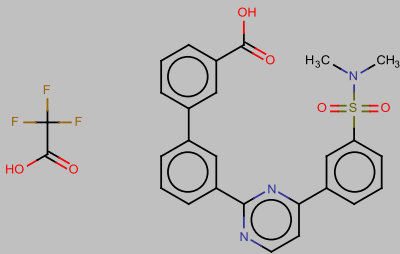
Compound number	Structure
62	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.67	4.87	5.2
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
4.98	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
0.87	4.38	4.7
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.7	S
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
6	2	1
clogP	cmr	Flexibility_Index
3.346	8.493	13
N_Acceptors	N_Donors	N_Heavy_Atoms
2	2	21
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
327.21	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
3	58.95	
PFI	Assay_count	IFI_50
6.74	164	1.83
Chemotypes		
urea phenyl derivative		

Compound number	Structure	
63		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.795	5.56	3.5
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.69	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.56	4.5	4.56
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.56	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
2	2	1
clogP	cmr	Flexibility_Index
3.089	11.711	19
N_Acceptors	N_Donors	N_Heavy_Atoms
5	0	29
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
435.56	1	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
6	95.91	
PFI	Assay_count	IFI_50
7.32	154	1.95
Chemotypes		
sulfonamide phenyl derivative		

Compound number	Structure	
64		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.585	4.93	5.17
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
4.57	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
0.93	5.59	5.46
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	1.46	S
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
14	4	0
clogP	cmr	Flexibility_Index
4.477	13.332	15
N_Acceptors	N_Donors	N_Heavy_Atoms
4	1	34
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
474.58	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
6	84.3	
PFI	Assay_count	IFI_50
9.71	163	1.23
Chemotypes		
benzimidazole amide phenyl derivative		

Compound number	Structure	
65		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.41	5.82	5.37
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.48	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.82	4.38	4.64
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.64	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
6	2	1
clogP	cmr	Flexibility_Index
2.447	8.029	14
N_Acceptors	N_Donors	N_Heavy_Atoms
2	2	20
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
313.18	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
3	58.95	
PFI	Assay_count	IFI_50
6	148	0.68
Chemotypes		
urea phenyl derivative		
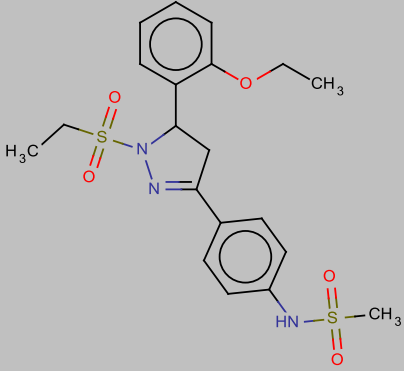
Compound number	Structure	
66		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.51	5.68	5.92
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.53	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.68	4.38	4.56
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.56	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
1	2	0
clogP	cmr	Flexibility_Index
2.949	8.902	17
N_Acceptors	N_Donors	N_Heavy_Atoms
2	2	22
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
300.4	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
4	58.95	
PFI	Assay_count	IFI_50
6.15	139	0.72
	Chemotypes	
	urea phenyl derivative	

Compound number	Structure	
67		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.365	4.76	4.89
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
4.39	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
0.76	5.435	5.27
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4.725385	0.544615	S
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
24	4	0
clogP	cmr	Flexibility_Index
4.724	12.621	11
N_Acceptors	N_Donors	N_Heavy_Atoms
6	0	33
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
459.52	1	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
4	100.46	Low
PFI	Assay_count	IFI_50
7.33	119	13.45
Chemotypes		
pyrimidyl carboxylate sulfonamide derivative		

Compound number	Structure
68	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.115	5.32	5.63
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.11	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.32	4.38	4.525
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.525	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
29	2	0
clogP	cmr	Flexibility_Index
1.545	7.874	17
N_Acceptors	N_Donors	N_Heavy_Atoms
2	1	21
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
286.33	0	1
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
4	71.26	
PFI	Assay_count	IFI_50
5.83	91	1.1
Chemotypes		
pyridyl isoxazole amide derivative		

154

Compound number	Structure	
69		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.56	5.18	5.3
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
4.1	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.18	4.38	4.41
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.41	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
4	2	1
clogP	cmr	Flexibility_Index
4.376	13.042	16
N_Acceptors	N_Donors	N_Heavy_Atoms
4	1	33
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
471.61	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
6	78.95	
PFI	Assay_count	IFI_50
7.58	217	0.46
Chemotypes		
morpholino amide phenyl derivative		

Compound number	Structure	
70		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.375	5.43	5.35
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.37	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.43	4.38	4.49
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.49	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
4	2	1
clogP	cmr	Flexibility_Index
4.506	10.276	14
N_Acceptors	N_Donors	N_Heavy_Atoms
1	1	26
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
354.44	0	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
4	50.8	
PFI	Assay_count	IFI_50
7.21	203	3.94
Chemotypes		
amide ether phenyl derivative		

Compound number	Structure	
71		

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
5.55	5.34	6.09
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.45	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.34	4.56	4.525
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.525	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
2	2	1
clogP	cmr	Flexibility_Index
2.479	11.864	18
N_Acceptors	N_Donors	N_Heavy_Atoms
5	0	30
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
451.56	1	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
6	105.14	Low
PFI	Assay_count	IFI_50
6.76	205	0.98
Chemotypes		
sulfonamide ether phenyl derivative		

Compound number	Structure
72	

Avg 1ry pIC$_{50}$	pIC$_{50}$ LCys	pIC$_{50}$ Sans DTT
4.745	5.89	3.5
pIC$_{50}$ SOD/Cat	pIC$_{50}$ hCathepsin B	pIC$_{50}$ REDOX assay
5.73	4	4
Selectivity diff.	Average Induced pIC$_{50}$	Average Uninduced pIC$_{50}$
1.89	4.575	4.51
Avg. pIC$_{50}$ cytotox	pIC$_{50}$ diff. uninduced and cytotox	Tier
4	0.51	M
CLink	N_Ar_Rings	N_Chiral_Atoms/Bonds
2	2	1
clogP	cmr	Flexibility_Index
3.588	12.174	18
N_Acceptors	N_Donors	N_Heavy_Atoms
5	0	30
Molecular_Weight	N_Neg_Charge	N_Pos_Charge
449.59	1	0
N_Rot_Bonds	Total_Polar_Surface_Area	CNS_Penetration
6	95.91	Low
PFI	Assay_count	IFI_50
7.63	197	1.52
Chemotypes		
sulfonamide phenyl derivative		

Chapter 6 Discussion

Human diseases caused by Kinetoplastids have been neglected for several decades and hence the available treatments against HAT, Chagas disease and Leishmaniasis are obsolete. Serious efforts are needed for the discovery of novel treatments against these diseases. For the discovery of novel small treatments, both target based approach and phenotypic screening are very well validated approaches, each with its own advantages (Swinney and Anthony, 2011, Eder *et al.*, 2014). Both of these approaches are being pursued at GSK for drug discovery efforts against Kinetoplastids. Phenotypic screening of the GSK compound collection led to identification of compounds that act against these three parasites (Pena *et al.*, 2015). These compounds are available as an open resource for labs all over the world. Efforts are on for target deconvolution for these compounds. As a parallel effort, target based approaches are being used for the identification of novel and tractable hits. Target based approaches have an advantage over the phenotypic screening for intracellular targets and for intracellular parasites. Biochemical assays can provide diverse chemical starting points against a target. This can be followed with better informed follow-up chemistry efforts for lead generation. In this thesis we follow a target based approach focussed on an essential deubiquitinase in *T. brucei*.

Ubiquitin was first discovered in 1975, much after most of the available treatments against these diseases were discovered. Ubiquitin biology has been mostly unveiled in the last two decades. Even though the field has shown a huge potential for discovery of novel therapeutic target, still not many compounds are available to inhibit the activity of various components of the ubiquitin pathway. The work in this thesis is an attempt to bring the ubiquitin drug discovery and Kinetoplastid drug discovery together. Deubiquitinases or ubiquitin specific proteases catalyse the removal of ubiquitin from the target proteins. Proteases, in general, have been highly researched and pursued as successful therapeutic targets. It is estimated that 5-10% of all pharmaceutical targets being pursued are proteases (Drag and Salvesen, 2010). With at least 30 marketed drugs targeting proteases against various diseases like hypertension, myocardial infarction, AIDS, respiratory

diseases and cancer, to count a few, proteases have emerged as the most prominent target family (Drag and Salvesen, 2010, Imming *et al.*, 2006, Turk, 2006). DUBs form one of the largest classes of proteases. DUBs are genetically and chemically validated targets in several therapeutic areas and even though it is difficult to achieve specificity, they have proven to be tractable targets. The work presented in this thesis is based on the initial discovery of essential deubiquitinases in *T. brucei* by Dr. Boris Rodenko *et. al*. Their work established deubiquitinases as genetically validated targets in *T. brucei* by RNAi screening. They also showed that inhibition of DUBs by pan-DUB inhibitors in *T. brucei* is cidal.

We started with the aim of identifying trypanocidal inhibitors of *Tbr*DUB1 and checking cross-species activity of these small molecules. For this we chose to perform a high throughput screening using a deubiquitinase biochemical assay and ~1.7 million compounds from the GSK collection. The chances of discovering tractable hits from a HTS campaign increase with the size and quality of the library screened. HTS is thus a good example of chance discovery as opposed to knowledge based drug design. However, HTS has proven to be a successful approach for identification of small molecule inhibitors, with several marketed drugs having their origins from target based screening efforts (Macarron *et al.*, 2011, Swinney and Anthony, 2011).

The first phase of the work involved reagent generation and assay development. We expressed and purified active recombinant *Tbr*DUB1 by expressing it in insect cells (Sf9). Insect cells were chosen as an expression system because *Tbr*DUB1 was insoluble in *E. coli* and it is likely that some post-translational modifications in a eukaryotic expression system improve its solubility. We were then able to set up a robust assay for the high throughput screening as shown in **Figure 3.12** to **Figure 3.15** and **Table 3.1**. The assay uses a substrate that has an isopeptide linkage, the biologically relevant substrate for the DUBs. When the performance of the assay was tested with 3 replicates of the validation set, a hit rate of 0.26% was

observed with a predicted confirmation rate of 69%. This meant that the HTS would give us around 3000 confirmed hits. A hit rate of 0.26% could be considered low because in some screens the hit rate is typically around 1%. However, this expected hit rate is very similar to a hit rate of 0.2% seen in USP7 screening using a fluorometric assay (Colland *et al.*, 2009).
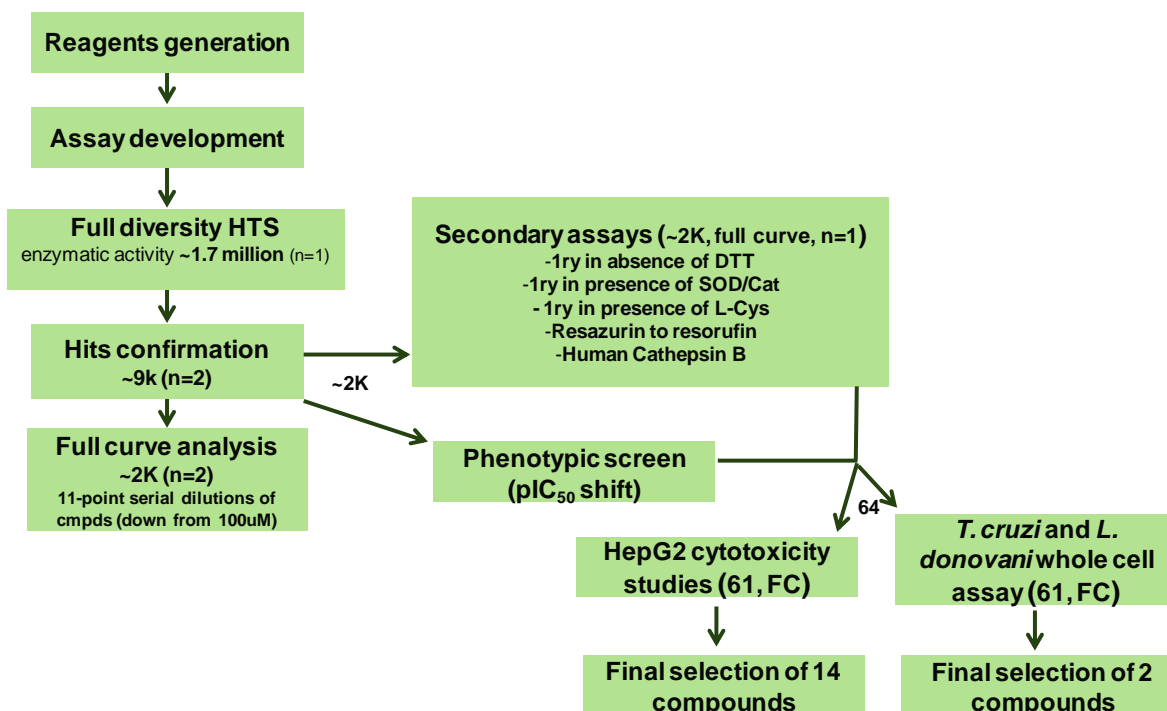
Figure 6.1: The project plan.
The plan of the work carried out and presented in this thesis. Starting from around 1.7 million compounds we have been able to prioritise 15 compounds (14 from one selection and 2 from another, 1 compound overlaps in these two).

The HTS itself was set up to run in a semi-automated manner. The test compound dispensation in 1536-well plates was performed by acoustic transfer using Echo liquid handling system. The plate handling was done by Agilent direct drive robot and liquid dispensation using Thermo Multidrop® Combi, both controlled by Agilent VWorks automation software. For the screening, the maximum achieved throughput was of around 250,000 compounds on one day. The assay performed robustly for the HTS, with only 0.56% plates failing. The HTS yielded 5331 hits,

162

which calculates to a hit rate of 0.3%. The observed cut-off at 3X SD was 24.23% inhibition at the tested concentration of 10 µM. To select compounds for confirmation studies, the patterns were corrected and weak analogs of potent inhibitors were also included and a total of 9405 hits were progressed to hit confirmation studies as shown in **Figure 4.4**. Of these, 2020 hits were confirmed as robust actives. To determine the potency and efficacy of these compounds, dose response studies were carried out and potency was determined by calculating the pIC_{50} values. These hits showed pIC_{50} values ranging from 4 to approximately 8.

It is very important to remove compounds that act via undesirable molecular mechanisms of action and also compounds that could have potential liabilities because of their activity on host proteins. As discussed earlier, HTS is serendipity based drug discovery. Therefore, it is desirable to maximise the chances of hit identification in the primary HTS. The undesirable hits can be easily filtered using secondary assays; however in this process valuable SAR information is generated that could be very useful for the chemists when working on optimising the physicochemical properties of the hits. Using L-cysteine as the reducing agent in the primary assay would have significantly reduced the hit rates. Thus, we decided to use it as a secondary assay along with other assays to filter out or flag the undesirable compounds. In one of the assays we replaced the reducing agent from DTT to L-cys. In another assay we removed the reducing agent from the buffer. In a third assay we added superoxide dismutase and catalase to the assay buffer to remove superoxides and peroxide. In the fourth assay we profiled compounds for their redox potential by testing their ability to reduce resazurin to resorufin in the presence of DTT. And lastly, we tested the compounds for their activity against human cathepsin B. 64 compounds were selected at this stage after profiling in these five secondary assays. Based on their selectivity against human cathepsin B, we classified the 64 *Tbr*DUB1 hits as- 20 most selective compounds, 23 selective compounds and 21 least selective compounds. A complete link hierarchical clustering of the 64 identified *Tbr*DUB1 hits and 57 known DUB inhibitors,

presented in **Table 1.2**, was performed. Out of the 64 *Tbr*DUB1 hits only 2 clustered with known inhibitors of USP7- namely spongiacidin C (**22**) and HBX19818 (**23**). Neither of these 2 hits made it to our list of finally selected hits. HBX19818 has been shown to irreversibly inhibit USP7 by covalently binding Cys223 with an IC_{50} of 28.1 µM ([Reverdy *et al.,* 2012](#)). The remaining 62 *Tbr*DUB1 clustered separately, in clusters or as singletons, indicating that these are novel DUB inhibitors.

All the confirmed hits from the primary assay were also tested for their potency against *T. brucei* and we measured the change in this potency when the *Tbr*DUB1 was overexpressed. This is a powerful technique frequently referred to as "target-biased phenotypic assay". A mutant *T. brucei* cell line was used that allowed inducible *Tbr*DUB1 overexpression. The 64 hits selected after biochemical profiling showed a pIC_{50} range of between 5.5 and 4.3 *i.e.* an IC_{50} range of 3.16 to 50 µM. However, we could not detect highly significant shifts in the pIC_{50} values of *Tbr*DUB1 hits in wild type *vs.* overexpressor comparison. This could largely be attributed to low expression levels of the *Tbr*DUB1 (2-3 fold overexpression). Additionally, as presented in **Figure 5.5**, the known trypanocidal compounds show a trend of higher potencies against the overexpressor, indicating that DUB overexpression could have a negative impact on the parasite's "fitness". Hence even a small decrease in pIC_{50} of *Tbr*DUB1 inhibitors against the overexpressor could be significant. This leads us to ask the question that if a compound, having survived a thorough biochemical profiling, is not showing a shift in pIC_{50} when the target is overexpressed, then should it be considered on-target or not? This can be answered, in part, by using biophysical techniques such as affinity selection mass spectrometry and surface plasma resonance to demonstrate the binding of the compounds to the target protein. Activity based protein profiling can also be done with these hits to demonstrate on-target activity. Also, even though the pIC_{50} value is the same, maybe the compounds could show a different rate of kill of the parasite when the target is overexpressed.

The compounds selected after biochemical profiling were passed to cytotoxicity studies in HepG2 cells. We were able to identify 14 small molecules that were either not cytotoxic or had at least 2.5 times higher IC_{50} values against the *T. brucei* compared to HepG2 cells.

Lastly, we also tested the selected 64 hits against *T. cruzi* and *Leishmania donovani* in whole cell assays to test for cross-species activity of the *Tbr*DUB1 inhibitors. We were able to identify 2 compounds that had potencies in the range of 2 - 15 µM against *T. cruzi and L. donovani*, while being inactive against the host cells. These compounds are thiophene phenyl derivatives. Katsuno *et al.* ([Katsuno](#) *[et al., 2015](#)*) describe the hit selection criteria for visceral leishmaniasis and Chagas disease that has been agreed by various international agencies involved in drug development against neglected diseases. Both compound **58** and **59** conform to the hit selection criteria for visceral leishmaniasis (IC_{50} 3.98 and 1.99 µM respectively). Their potencies against *T. cruzi* are 15 and 10.4 µM respectively. For compound **59** this is very close to the described cut-off – 10.4 µM *vs.* desired 10 µM. Since these compounds act on all three parasites, they could hold a great value in follow-up efforts. Even though it might prove challenging, they present a good chance of developing one pill that treats three diseases.

Even though the remaining *Tbr*DUB1 hits have been put in lower priority groups, these hits still hold a lot of value. Efforts on improving their membrane permeability, for example, could lead to a better response in the whole cell assays. The selected high priority hits can go through a hit to lead campaign wherein medicinal chemists attempt to optimise the potency, permeability, solubility and selectivity of these hits to obtain compounds that have the chances of showing optimum pharmacokinetics (PK) and pharmacodynamics (PD). These hits and the resulting lead compounds could be tested in *in vivo* models to study the PK and PD. The mechanism of inhibition or action of these compounds also is still to be determined. A co-crystal structure of the *Tbr*DUB1 with its inhibitors could help identify the binding sites. Also, it will be very valuable to screen the *Tbr*DUB1 hits against a panel of human

165

DUBs. Such a screen will help us characterise the inhibitors for their selectivity. One possible outcome of this screen could also be activity against a validated target human DUB with implications in other therapeutic areas like oncology. Furthermore, since our target are diseases caused by Kinetoplastids, for which the existing treatments are old and have serious side effects, the therapeutic index of the compounds should be carefully taken into account before de-prioritising them for the fear of potential liabilities in case of any activity against human deubiquitinases. A very interesting line of follow up research could be to identify the effect of these compounds on the ubiquitome of the parasites. This could help us identify downstream targets of the *Tbr*DUB1 and give hints towards the biological role and the targets of the *Tbr*DUB1 which still need to de determined. Efforts are on for the genetic validation of *Leishmania* and *T. cruzi* DUB as a target, even though our work provides hints towards that via chemical validation given that the compounds are on-target in these parasites. In conclusion, the work presented in this thesis paves a very promising way forward for drug discovery efforts against the Kinetoplastids.

Conclusions

As a result of the work presented in this thesis we can conclude that -

- We identified 64 *Tbr*DUB1 small molecule inhibitors after biochemical screening and triaging. Only two *Tbr*DUB1 hits cluster with known inhibitors of human USP7 while the remaining 62 cluster separately. Thus, we have identified novel inhibitors against deubiquitinases that could have applications against Kinetoplastids and could also hold potential against other therapeutic areas like oncology where DUBs play the roles of key regulators of cellular processes.

- Out of the 64 *Tbr*DUB1 hits we identified fourteen *Tbr*DUB1 that do not act via redox mechanisms, show specificity when compared to activity a human cysteine protease and are cidal against *T. brucei.* Additionally, our work further chemically validates *Tbr*DUB1 as a target in *T. brucei.*

- Out of the 64 *Tbr*DUB1 hits we identified two small molecule *Tbr*DUB1 inhibitors that are cidal against the three Kinetoplastid parasites – *T. brucei*, *T. cruzi* and *L. donovani.* These small molecules provide starting points for efforts to develop treatments against HAT, Chagas disease and leishmaniasis. This also indicates that the orthologues of *Tbr*DUB1 in *T. cruzi* and *L. donovani* could be potential targets for therapeutic intervention.

- All identified *Tbr*DUB1 hits have drug-like physicochemical properties and are not promiscuous enzyme inhibitors.

Conclusiones

De los resultados obtenidos en este trabajo de tesis se derivan las siguientes conclusiones:

- Hemos identificado 64 moléculas pequeña que inhiben la *Tbr*DUB1. Solo dos de los compuestos seleccionados para *Tbr*DUB1 son análogos de inhibidores conocidos de una desubiquitinasa humana (USP7) mientras que los restantes 62 son estructuras novedosas. De esta forma, hemos identificado inhibidores novedosos de desubiquitinasas que podrían tener aplicación contra parásitos kinetoplástidos y podrían presentar también potencial en otras áreas terapéuticas como la oncología, donde las desubiquitinasas juegan importantes papeles como reguladores clave de procesos celulares.

- De estos 64 moléculas, hemos identificado 14 moléculas que inhiben la *Tbr*DUB1, no lo hacen mediante un mecanismo redox, muestran especificidad respecto a una cisteín-proteasa humana y son capaces de matar a *T. brucei*. Adicionalmente, nuestro trabajo avanza en la validación química de la *Tbr*DUB1 como diana farmacológica en *T. brucei*.

- De estos 64 moléculas, hemos identificado 2 inhibidores de la *Tbr*DUB1 de molécula pequeña que matan a los tres parásitos kinetoplástidos de interés – *T. brucei*, *T. cruzi y L. donovani.* Estas moléculas pequeñas representan puntos de partida para esfuerzos de desarrollo de nuevos tratamientos para la enfermedad del sueño, el mal de Chagas y la leishmaniasis. Este hecho sugiere además la presencia de ortólogos de *Tbr*DUB1 en *T. cruzi* y *L. donovani* como dianas potenciales para la intervención terapéutica.

- Todas las moléculas seleccionadas para *Tbr*DUB1 poseen propiedades físico-químicas atractivas desde el punto de vista farmacológico y no son inhibidores enzimáticos promiscuos.

References

ADAMS, J. 2002. Development of the proteasome inhibitor PS-341. *Oncologist.,* 7**,** 9-16.

ALONSO-PADILLA, J., COTILLO, I., PRESA, J. L., CANTIZANI, J., PENA, I., BARDERA, A. I., MARTIN, J. J. & RODRIGUEZ, A. 2015. Automated high-content assay for compounds selectively toxic to Trypanosoma cruzi in a myoblastic cell line. *PLoS Negl Trop Dis,* 9**,** e0003493.

ALSFORD, S. & HORN, D. 2008. Single-locus targeting constructs for reliable regulated RNAi and transgene expression in Trypanosoma brucei. *Mol Biochem Parasitol,* 161**,** 76-9.

ALVAR, J., VELEZ, I. D., BERN, C., HERRERO, M., DESJEUX, P., CANO, J., JANNIN, J., DEN BOER, M. & TEAM, W. H. O. L. C. 2012. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One,* 7**,** e35671.

ANALYTICAL METHODS COMMITTEE, O. T. R. S. O. C. U. 1989. Robust statistics–how not to reject outliers. Part 1. Basic concepts. *The Analyst,* 114**,** 1693-1697.

ANDERSEN, P. L., ZHOU, H., PASTUSHOK, L., MORAES, T., MCKENNA, S., ZIOLA, B., ELLISON, M. J., DIXIT, V. M. & XIAO, W. 2005. Distinct regulation of Ubc13 functions by the two ubiquitin-conjugating enzyme variants Mms2 and Uev1A. *J Cell Biol,* 170**,** 745-55.

BAHIA-OLIVEIRA, L. M., GOMES, J. A., CANCADO, J. R., FERRARI, T. C., LEMOS, E. M., LUZ, Z. M., MOREIRA, M. C., GAZZINELLI, G. & CORREA-OLIVEIRA, R. 2000. Immunological and clinical evaluation of chagasic patients subjected to chemotherapy during the acute phase of Trypanosoma cruzi infection 14-30 years ago. *J Infect Dis,* 182**,** 634-8.

BAKER, N., DE KONING, H. P., MASER, P. & HORN, D. 2013. Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. *Trends Parasitol,* 29**,** 110-8.

BALASEGARAM, M., YOUNG, H., CHAPPUIS, F., PRIOTTO, G., RAGUENAUD, M. E. & CHECCHI, F. 2009. Effectiveness of melarsoprol and eflornithine as first-line regimens for gambiense sleeping sickness in nine Medecins Sans Frontieres programmes. *Trans.R.Soc.Trop.Med.Hyg.,* 103**,** 280-290.

BARRETT, M. P., VINCENT, I. M., BURCHMORE, R. J., KAZIBWE, A. J. & MATOVU, E. 2011. Drug resistance in human African trypanosomiasis. *Future Microbiol,* 6**,** 1037-47.

BARTHOLOMEUSZ, G., TALPAZ, M., BORNMANN, W., KONG, L. Y. & DONATO, N. J. 2007a. Degrasyn activates proteasomal-dependent degradation of c-Myc. *Cancer Res,* 67**,** 3912-8.

BARTHOLOMEUSZ, G. A., TALPAZ, M., KAPURIA, V., KONG, L. Y., WANG, S., ESTROV, Z., PRIEBE, W., WU, J. & DONATO, N. J. 2007b. Activation of a novel Bcr/Abl destruction pathway by WP1130 induces apoptosis of chronic myelogenous leukemia cells. *Blood,* 109**,** 3470-8.

BASLER, M., DAJEE, M., MOLL, C., GROETTRUP, M. & KIRK, C. J. 2010. Prevention of experimental colitis by a selective inhibitor of the immunoproteasome. *J Immunol,* 185**,** 634-41.

BERNDSEN, C. E. & WOLBERGER, C. 2014. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol,* 21**,** 301-7.

BIGGART, A., LIANG, F., MATHISON, C. J. N., MOLTENI, V., NAGLE, A. S., SUPEK, F. & YEH, V. 2015. *[1,2,4]triazolo[1,5-a]pyrimidine derivatives as protozoan proteasome inhibitors for the treatment of parasitic diseases such as leishmaniasis.* PCT/US2014/071077.

BINGOL, B., TEA, J. S., PHU, L., REICHELT, M., BAKALARSKI, C. E., SONG, Q., FOREMAN, O., KIRKPATRICK, D. S. & SHENG, M. 2014. The mitochondrial deubiquitinase USP30 opposes parkin-mediated mitophagy. *Nature,* 510**,** 370-5.

BORODOVSKY, A., KESSLER, B. M., CASAGRANDE, R., OVERKLEEFT, H. S., WILKINSON, K. D. & PLOEGH, H. L. 2001. A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *EMBO J,* 20**,** 5187-96.

BROOKS, C. L. & GU, W. 2011. p53 regulation by ubiquitin. *FEBS Lett,* 585**,** 2803-9.

CAMPOS, M. C., LEON, L. L., TAYLOR, M. C. & KELLY, J. M. 2014. Benznidazole-resistance in Trypanosoma cruzi: evidence that distinct mechanisms can act in concert. *Mol Biochem Parasitol,* 193**,** 17-9.

CANCADO, J. R. 2002. Long term evaluation of etiological treatment of chagas disease with benznidazole. *Rev Inst Med Trop Sao Paulo,* 44**,** 29-37.

CASTRO, J. A., DE MECCA, M. M. & BARTEL, L. C. 2006. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). *Hum Exp Toxicol,* 25**,** 471-9.

CDC. Available: https://www.cdc.gov/parasites/ [Accessed].

CECCARELLI, D. F., TANG, X., PELLETIER, B., ORLICKY, S., XIE, W., PLANTEVIN, V., NECULAI, D., CHOU, Y. C., OGUNJIMI, A., AL-HAKIM, A., VARELAS, X., KOSZELA, J., WASNEY, G. A., VEDADI, M., DHE-PAGANON, S., COX, S., XU, S., LOPEZ-GIRONA, A., MERCURIO, F., WRANA, J., DUROCHER, D., MELOCHE, S., WEBB, D. R., TYERS, M. & SICHERI, F. 2011. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell,* 145**,** 1075-87.

CHAPPUIS, F., UDAYRAJ, N., STIETENROTH, K., MEUSSEN, A. & BOVIER, P. A. 2005. Eflornithine is safer than melarsoprol for the treatment of second-stage Trypanosoma brucei gambiense human African trypanosomiasis. *Clin.Infect.Dis.,* 41**,** 748-751.

CHAU, V., TOBIAS, J. W., BACHMAIR, A., MARRIOTT, D., ECKER, D. J., GONDA, D. K. & VARSHAVSKY, A. 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science,* 243**,** 1576-83.

CHAUHAN, D., TIAN, Z., NICHOLSON, B., KUMAR, K. G., ZHOU, B., CARRASCO, R., MCDERMOTT, J. L., LEACH, C. A., FULCINNITI, M., KODRASOV, M. P., WEINSTOCK, J., KINGSBURY, W. D., HIDESHIMA, T., SHAH, P. K., MINVIELLE, S., ALTUN, M., KESSLER, B. M., ORLOWSKI, R., RICHARDSON, P., MUNSHI, N. & ANDERSON, K. C. 2012. A small molecule inhibitor of ubiquitin-specific protease-7 induces apoptosis in multiple myeloma cells and overcomes bortezomib resistance. *Cancer Cell,* 22**,** 345-58.

CHEN, C., MENG, Y., WANG, L., WANG, H. X., TIAN, C., PANG, G. D., LI, H. H. & DU, J. 2014. Ubiquitin-activating enzyme E1 inhibitor PYR41 attenuates angiotensin II-induced activation of dendritic cells via the IkappaBa/NF-kappaB and MKP1/ERK/STAT1 pathways. *Immunology,* 142**,** 307-19.

CHENG, J., FAN, Y. H., XU, X., ZHANG, H., DOU, J., TANG, Y., ZHONG, X., ROJAS, Y., YU, Y., ZHAO, Y., VASUDEVAN, S. A., ZHANG, H., NUCHTERN, J. G., KIM, E. S., CHEN, X., LU, F. & YANG, J. 2014. A small-molecule inhibitor of UBE2N induces neuroblastoma cell death via activation of p53 and JNK pathways. *Cell Death Dis,* 5**,** e1079.

CHENG, K. W., CHENG, S. C., CHEN, W. Y., LIN, M. H., CHUANG, S. J., CHENG, I. H., SUN, C. Y. & CHOU, C. Y. 2015. Thiopurine analogs and mycophenolic acid synergistically inhibit the papain-like protease of Middle East respiratory syndrome coronavirus. *Antiviral Res,* 115**,** 9-16.

CHOU, C. Y., CHIEN, C. H., HAN, Y. S., PREBANDA, M. T., HSIEH, H. P., TURK, B., CHANG, G. G. & CHEN, X. 2008. Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus. *Biochem Pharmacol,* 75**,** 1601-9.

CIECHANOVER, A., HELLER, H., ELIAS, S., HAAS, A. L. & HERSHKO, A. 1980. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci U S A,* 77**,** 1365-8.

CIEHANOVER, A., HOD, Y. & HERSHKO, A. 1978. A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochem Biophys Res Commun,* 81**,** 1100-5.

COHEN, P. & TCHERPAKOV, M. 2010. Will the ubiquitin system furnish as many drug targets as protein kinases? *Cell,* 143**,** 686-693.

COLLAND, F. 2010. The therapeutic potential of deubiquitinating enzyme inhibitors. *Biochem Soc Trans,* 38**,** 137-43.

COLLAND, F., FORMSTECHER, E., JACQ, X., REVERDY, C., PLANQUETTE, C., CONRATH, S., TROUPLIN, V., BIANCHI, J., AUSHEV, V. N., CAMONIS, J., CALABRESE, A., BORG-CAPRA, C., SIPPL, W., COLLURA, V., BOISSY, G., RAIN, J. C., GUEDAT, P., DELANSORNE, R. & DAVIET, L. 2009. Small-molecule inhibitor of USP7/HAUSP ubiquitin protease stabilizes and activates p53 in cells. *Mol.Cancer Ther.,* 8**,** 2286-2295.

COMA, I., HERRANZ, J. & MARTIN, J. 2009. Statistics and Decision Making in High-Throughput Screening. *In:* JANZEN, W. P. & BERNASCONI, P. (eds.) *High Throughput Screening.* Humana Press.

CROUCH, S. P., KOZLOWSKI, R., SLATER, K. J. & FLETCHER, J. 1993. The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J Immunol Methods,* 160**,** 81-8.

CUMMINS, J. M., RAGO, C., KOHLI, M., KINZLER, K. W., LENGAUER, C. & VOGELSTEIN, B. 2004. Tumour suppression: disruption of HAUSP gene stabilizes p53. *Nature,* 428**,** 1 p following 486.

CUNNINGHAM, C. N., BAUGHMAN, J. M., PHU, L., TEA, J. S., YU, C., COONS, M., KIRKPATRICK, D. S., BINGOL, B. & CORN, J. E. 2015. USP30 and parkin homeostatically regulate atypical ubiquitin chains on mitochondria. *Nat Cell Biol,* 17**,** 160-9.

D'ARCY, P., BRNJIC, S., OLOFSSON, M. H., FRYKNAS, M., LINDSTEN, K., DE CESARE, M., PEREGO, P., SADEGHI, B., HASSAN, M., LARSSON, R. & LINDER, S. 2011. Inhibition of proteasome deubiquitinating activity as a new cancer therapy. *Nat Med,* 17**,** 1636-40.

D'ARCY, P., WANG, X. & LINDER, S. 2015. Deubiquitinase inhibition as a cancer therapeutic strategy. *Pharmacol Ther,* 147**,** 32-54.

DANIEL, E., JORGE, G., VICTOR, N. & PHOTINI, S. 1998. *Method for treating parasitic diseases with proteasome inhibitors.* PCT/US1997/017136.

DAVIET, L. & COLLAND, F. 2008. Targeting ubiquitin specific proteases for drug discovery. *Biochimie,* 90**,** 270-83.

DE JONG, A., MERKX, R., BERLIN, I., RODENKO, B., WIJDEVEN, R. H., EL, A. D., YALCIN, Z., ROBSON, C. N., NEEFJES, J. J. & OVAA, H. 2012. Ubiquitin-based probes prepared by total synthesis to profile the activity of deubiquitinating enzymes. *Chembiochem.,* 13**,** 2251-2258.

DE RYCKER, M., HALLYBURTON, I., THOMAS, J., CAMPBELL, L., WYLLIE, S., JOSHI, D., CAMERON, S., GILBERT, I. H., WYATT, P. G., FREARSON, J. A., FAIRLAMB, A. H. & GRAY, D. W. 2013. Comparison of a high-throughput high-content intracellular Leishmania donovani assay with an axenic amastigote assay. *Antimicrob Agents Chemother,* 57**,** 2913-22.

DEN BOER, M. L., ALVAR, J., DAVIDSON, R. N., RITMEIJER, K. & BALASEGARAM, M. 2009. Developments in the treatment of visceral leishmaniasis. *Expert Opin Emerg Drugs,* 14**,** 395-410.

DENG, L., WANG, C., SPENCER, E., YANG, L., BRAUN, A., YOU, J., SLAUGHTER, C., PICKART, C. & CHEN, Z. J. 2000. Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell,* 103**,** 351-61.

DESHAIES, R. J. & JOAZEIRO, C. A. 2009. RING domain E3 ubiquitin ligases. *Annu.Rev.Biochem.,* 78**,** 399-434.

DIMASI, J. A., GRABOWSKI, H. G. & HANSEN, R. W. 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ,* 47**,** 20-33.

DIMASI, J. A., HANSEN, R. W. & GRABOWSKI, H. G. 2003. The price of innovation: new estimates of drug development costs. *J Health Econ,* 22**,** 151-85.

DRAG, M. & SALVESEN, G. S. 2010. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov,* 9**,** 690-701.

DUPONT, S., MAMIDI, A., CORDENONSI, M., MONTAGNER, M., ZACCHIGNA, L., ADORNO, M., MARTELLO, G., STINCHFIELD, M. J., SOLIGO, S., MORSUT, L., INUI, M., MORO, S., MODENA, N., ARGENTON, F., NEWFELD, S. J. & PICCOLO, S. 2009. FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell,* 136**,** 123-35.

EDER, J., SEDRANI, R. & WIESMANN, C. 2014. The discovery of first-in-class drugs: origins and evolution. *Nat Rev Drug Discov,* 13**,** 577-87.

EL-SAYED, N. M., MYLER, P. J., BLANDIN, G., BERRIMAN, M., CRABTREE, J., AGGARWAL, G., CALER, E., RENAULD, H., WORTHEY, E. A., HERTZ-FOWLER, C., GHEDIN, E., PEACOCK, C., BARTHOLOMEU, D. C., HAAS, B. J., TRAN, A. N., WORTMAN, J. R., ALSMARK, U. C., ANGIUOLI, S., ANUPAMA, A., BADGER, J., BRINGAUD, F., CADAG, E., CARLTON, J. M., CERQUEIRA, G. C., CREASY, T., DELCHER, A. L., DJIKENG, A., EMBLEY, T. M., HAUSER, C., IVENS, A. C., KUMMERFELD, S. K., PEREIRA-LEAL, J. B., NILSSON, D., PETERSON, J., SALZBERG, S. L., SHALLOM, J., SILVA, J. C., SUNDARAM, J., WESTENBERGER, S., WHITE, O., MELVILLE, S. E., DONELSON, J. E., ANDERSSON, B., STUART, K. D. & HALL, N. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science,* 309**,** 404-9.

ELLIOTT, P. J., PIEN, C. S., MCCORMACK, T. A., CHAPMAN, I. D. & ADAMS, J. 1999. Proteasome inhibition: A novel mechanism to combat asthma. *J Allergy Clin Immunol,* 104**,** 294-300.

ELLIOTT, P. J., ZOLLNER, T. M. & BOEHNCKE, W. H. 2003. Proteasome inhibition: a new anti-inflammatory strategy. *J Mol Med (Berl),* 81**,** 235-45.

ELOFSSON, M., SPLITTGERBER, U., MYUNG, J., MOHAN, R. & CREWS, C. M. 1999. Towards subunit-specific proteasome inhibitors: synthesis and evaluation of peptide alpha',beta'-epoxyketones. *Chem Biol,* 6**,** 811-22.

ELSEVIER. *Embase* [Online]. Available: https://www.elsevier.com/solutions/embase-biomedical-research [Accessed].

ERDAL, H., BERNDTSSON, M., CASTRO, J., BRUNK, U., SHOSHAN, M. C. & LINDER, S. 2005. Induction of lysosomal membrane permeabilization by compounds that activate p53-independent apoptosis. *Proc Natl Acad Sci U S A,* 102**,** 192-7.

FARSHI, P., DESHMUKH, R. R., NWANKWO, J. O., ARKWRIGHT, R. T., CVEK, B., LIU, J. & DOU, Q. P. 2015. Deubiquitinases (DUBs) and DUB inhibitors: a patent review. *Expert Opin Ther Pat,* 25**,** 1191-208.

FISSOLO, N., KRAUS, M., REICH, M., AYTURAN, M., OVERKLEEFT, H., DRIESSEN, C. & WEISSERT, R. 2008. Dual inhibition of proteasomal and lysosomal proteolysis ameliorates autoimmune central nervous system inflammation. *Eur J Immunol,* 38**,** 2401-11.

FLYGARE, J. A. & FAIRBROTHER, W. J. 2010. Small-molecule pan-IAP antagonists: a patent review. *Expert Opin Ther Pat,* 20**,** 251-67.

FRESCAS, D. & PAGANO, M. 2008. Deregulated proteolysis by the F-box proteins SKP2 and beta-TrCP: tipping the scales of cancer. *Nat Rev Cancer,* 8**,** 438-49.

FUKUSHIMA, T., MATSUZAWA, S., KRESS, C. L., BRUEY, J. M., KRAJEWSKA, M., LEFEBVRE, S., ZAPATA, J. M., RONAI, Z. & REED, J. C. 2007. Ubiquitin-conjugating enzyme Ubc13 is a critical component of TNF receptor-associated factor (TRAF)-mediated inflammatory responses. *Proc Natl Acad Sci U S A,* 104**,** 6371-6.

GATTI, M., PINATO, S., MAIOLICA, A., ROCCHIO, F., PRATO, M. G., AEBERSOLD, R. & PENENGO, L. 2015. RNF168 promotes noncanonical K27 ubiquitination to signal DNA damage. *Cell Rep,* 10**,** 226-38.

GEURINK, P. P., EL, O. F., JONKER, A., HAMEED, D. S. & OVAA, H. 2012. A general chemical ligation approach towards isopeptide-linked ubiquitin and ubiquitin-like assay reagents. *Chembiochem.,* 13**,** 293-297.

GLEESON, M. P. 2008. Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem,* 51**,** 817-34.

GLEESON, M. P., HERSEY, A., MONTANARI, D. & OVERINGTON, J. 2011. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov,* 10**,** 197-208.

GOLDSTEIN, G., SCHEID, M., HAMMERLING, U., SCHLESINGER, D. H., NIALL, H. D. & BOYSE, E. A. 1975. Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc Natl Acad Sci U S A,* 72**,** 11-5.

GOMEZ, A. M., VROLIX, K., MARTINEZ-MARTINEZ, P., MOLENAAR, P. C., PHERNAMBUCQ, M., VAN DER ESCH, E., DUIMEL, H., VERHEYEN, F., VOLL, R. E., MANZ, R. A., DE BAETS, M. H. & LOSEN, M. 2011. Proteasome inhibition with bortezomib depletes plasma cells and autoantibodies in experimental autoimmune myasthenia gravis. *J Immunol,* 186**,** 2503-13.

GROETTRUP, M., PELZER, C., SCHMIDTKE, G. & HOFMANN, K. 2008. Activating the ubiquitin family: UBA6 challenges the field. *Trends Biochem.Sci.,* 33**,** 230-237.

HAGLUND, K., SIGISMUND, S., POLO, S., SZYMKIEWICZ, I., DI FIORE, P. P. & DIKIC, I. 2003. Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat.Cell Biol.,* 5**,** 461-466.

HAUPT, Y., MAYA, R., KAZAZ, A. & OREN, M. 1997. Mdm2 promotes the rapid degradation of p53. *Nature,* 387**,** 296-9.

HERNDON, T. M., DEISSEROTH, A., KAMINSKAS, E., KANE, R. C., KOTI, K. M., ROTHMANN, M. D., HABTEMARIAM, B., BULLOCK, J., BRAY, J. D., HAWES, J., PALMBY, T. R., JEE, J., ADAMS, W., MAHAYNI, H., BROWN, J., DORANTES, A., SRIDHARA, R., FARRELL, A. T. & PAZDUR, R. 2013. U.s. Food and Drug Administration approval: carfilzomib for the treatment of multiple myeloma. *Clin.Cancer Res.,* 19**,** 4559-4563.

HERSHKO, A., CIECHANOVER, A., HELLER, H., HAAS, A. L. & ROSE, I. A. 1980. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc Natl Acad Sci U S A,* 77**,** 1783-6.

HICKE, L. 2001. Protein regulation by monoubiquitin. *Nat.Rev.Mol.Cell Biol.,* 2**,** 195-201.

HILL, A. P. & YOUNG, R. J. 2010. Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discov Today,* 15**,** 648-55.

HIRD, A. W., AQUILA, B. M., HENNESSY, E. J., VASBINDER, M. M. & YANG, B. 2015. Small molecule inhibitor of apoptosis proteins antagonists: a patent review. *Expert Opin Ther Pat,* 25**,** 755-74.

HOEGE, C., PFANDER, B., MOLDOVAN, G. L., PYROWOLAKIS, G. & JENTSCH, S. 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature,* 419**,** 135-41.

HOFMANN, R. M. & PICKART, C. M. 1999. Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell,* 96**,** 645-53.

HONDA, R., TANAKA, H. & YASUDA, H. 1997. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett,* 420**,** 25-7.

HU, H., TANG, C., JIANG, Q., LUO, W., LIU, J., WEI, X., LIU, R. & WU, Z. 2015. Reduced ubiquitin-specific protease 9X expression induced by RNA interference inhibits the bioactivity of hepatocellular carcinoma cells. *Oncol Lett,* 10**,** 268-272.

HUANG, H., CECCARELLI, D. F., ORLICKY, S., ST-CYR, D. J., ZIEMBA, A., GARG, P., PLAMONDON, S., AUER, M., SIDHU, S., MARINIER, A., KLEIGER, G., TYERS, M. & SICHERI, F. 2014. E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat Chem Biol,* 10**,** 156-63.

HUGHES, J. D., BLAGG, J., PRICE, D. A., BAILEY, S., DECRESCENZO, G. A., DEVRAJ, R. V., ELLSWORTH, E., FOBIAN, Y. M., GIBBS, M. E., GILLES, R. W., GREENE, N., HUANG, E., KRIEGER-BURKE, T., LOESEL, J., WAGER, T., WHITELEY, L. &

ZHANG, Y. 2008. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg Med Chem Lett,* 18**,** 4872-5.

HUIBREGTSE, J. M., SCHEFFNER, M., BEAUDENON, S. & HOWLEY, P. M. 1995. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci U S A,* 92**,** 2563-7.

ICHIKAWA, H. T., CONLEY, T., MUCHAMUEL, T., JIANG, J., LEE, S., OWEN, T., BARNARD, J., NEVAREZ, S., GOLDMAN, B. I., KIRK, C. J., LOONEY, R. J. & ANOLIK, J. H. 2012. Beneficial effect of novel proteasome inhibitors in murine lupus via dual inhibition of type I interferon and autoantibody-secreting cells. *Arthritis Rheum,* 64**,** 493-503.

IKEDA, F. & DIKIC, I. 2008. Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. *EMBO Rep.,* 9**,** 536-542.

IMMING, P., SINNING, C. & MEYER, A. 2006. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov,* 5**,** 821-34.

JOHNSTON, P. A. 2011. Redox cycling compounds generate H2O2 in HTS buffers containing strong reducing reagents--real hits or promiscuous artifacts? *Curr Opin Chem Biol,* 15**,** 174-82.

JOHNSTON, P. A., SOARES, K. M., SHINDE, S. N., FOSTER, C. A., SHUN, T. Y., TAKYI, H. K., WIPF, P. & LAZO, J. S. 2008. Development of a 384-well colorimetric assay to quantify hydrogen peroxide generated by the redox cycling of compounds in the presence of reducing agents. *Assay Drug Dev Technol,* 6**,** 505-18.

JOLLY, L. A., TAYLOR, V. & WOOD, S. A. 2009. USP9X enhances the polarity and self-renewal of embryonic stem cell-derived neural progenitors. *Mol Biol Cell,* 20**,** 2015-29.

JONES, N. G., THOMAS, E. B., BROWN, E., DICKENS, N. J., HAMMARTON, T. C. & MOTTRAM, J. C. 2014. Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. *PLoS Pathog,* 10**,** e1003886.

KAPURIA, V., PETERSON, L. F., FANG, D., BORNMANN, W. G., TALPAZ, M. & DONATO, N. J. 2010. Deubiquitinase inhibition by small-molecule WP1130 triggers aggresome formation and tumor cell apoptosis. *Cancer Res,* 70**,** 9265-76.

KATSUNO, K., BURROWS, J. N., DUNCAN, K., HOOFT VAN HUIJSDUIJNEN, R., KANEKO, T., KITA, K., MOWBRAY, C. E., SCHMATZ, D., WARNER, P. & SLINGSBY, B. T. 2015. Hit and lead criteria in drug discovery for infectious diseases of the developing world. *Nat Rev Drug Discov,* 14**,** 751-8.

KEMP, M. 2016. Recent Advances in the Discovery of Deubiquitinating Enzyme Inhibitors. *Prog Med Chem,* 55**,** 149-92.

KHARE, S., NAGLE, A. S., BIGGART, A., LAI, Y. H., LIANG, F., DAVIS, L. C., BARNES, S. W., MATHISON, C. J., MYBURGH, E., GAO, M. Y., GILLESPIE, J. R., LIU, X., TAN, J. L., STINSON, M., RIVERA, I. C., BALLARD, J., YEH, V., GROESSL, T.,

FEDERE, G., KOH, H. X., VENABLE, J. D., BURSULAYA, B., SHAPIRO, M., MISHRA, P. K., SPRAGGON, G., BROCK, A., MOTTRAM, J. C., BUCKNER, F. S., RAO, S. P., WEN, B. G., WALKER, J. R., TUNTLAND, T., MOLTENI, V., GLYNNE, R. J. & SUPEK, F. 2016. Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness. *Nature,* 537**,** 229-233.

KISSELEV, A. F., VAN DER LINDEN, W. A. & OVERKLEEFT, H. S. 2012. Proteasome inhibitors: an expanding army attacking a unique target. *Chem Biol,* 19**,** 99-115.

KOMANDER, D. 2009. The emerging complexity of protein ubiquitination. *Biochem.Soc.Trans.,* 37**,** 937-953.

KOMANDER, D., CLAGUE, M. J. & URBE, S. 2009. Breaking the chains: structure and function of the deubiquitinases. *Nat.Rev.Mol.Cell Biol.,* 10**,** 550-563.

KUBBUTAT, M. H., JONES, S. N. & VOUSDEN, K. H. 1997. Regulation of p53 stability by Mdm2. *Nature,* 387**,** 299-303.

LAINE, A., TOPISIROVIC, I., ZHAI, D., REED, J. C., BORDEN, K. L. & RONAI, Z. 2006. Regulation of p53 localization and activity by Ubc13. *Mol Cell Biol,* 26**,** 8901-13.

LAKOMA, A., BARBIERI, E., AGARWAL, S., JACKSON, J., CHEN, Z., KIM, Y., MCVAY, M., SHOHET, J. M. & KIM, E. S. 2015. The MDM2 small-molecule inhibitor RG7388 leads to potent tumor inhibition in p53 wild-type neuroblastoma. *Cell Death Discov,* 1.

LAUWERS, E., JACOB, C. & ANDRE, B. 2009. K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. *J Cell Biol,* 185**,** 493-502.

LEACH, A. R. & GILLET, V. J. 2007. Selecting diverse set of compounds. *An Introduction To Chemoinformatics.* Springer Netherlands.

LEE, B. H., LEE, M. J., PARK, S., OH, D. C., ELSASSER, S., CHEN, P. C., GARTNER, C., DIMOVA, N., HANNA, J., GYGI, S. P., WILSON, S. M., KING, R. W. & FINLEY, D. 2010. Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature,* 467**,** 179-84.

LEESON, P. D. & SPRINGTHORPE, B. 2007. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov,* 6**,** 881-90.

LI, M., BROOKS, C. L., KON, N. & GU, W. 2004. A dynamic role of HAUSP in the p53-Mdm2 pathway. *Mol Cell,* 13**,** 879-86.

LI, W., BENGTSON, M. H., ULBRICH, A., MATSUDA, A., REDDY, V. A., ORTH, A., CHANDA, S. K., BATALOV, S. & JOAZEIRO, C. A. 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**,** e1487.

LIPINSKI, C. A. 2000. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods,* 44**,** 235-49.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* 23**,** 3-25.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev,* 46**,** 3-26.

LIPKOWITZ, S. & WEISSMAN, A. M. 2011. RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. *Nat Rev Cancer,* 11**,** 629-43.

LIU, N., LI, X., HUANG, H., ZHAO, C., LIAO, S., YANG, C., LIU, S., SONG, W., LU, X., LAN, X., CHEN, X., YI, S., XU, L., JIANG, L., ZHAO, C., DONG, X., ZHOU, P., LI, S., WANG, S., SHI, X., DOU, P. Q., WANG, X. & LIU, J. 2014. Clinically used antirheumatic agent auranofin is a proteasomal deubiquitinase inhibitor and inhibits tumor growth. *Oncotarget,* 5**,** 5453-71.

LOOK, G. C., LOWE, R., PARLATI, F., PAYAN, D. G., RAMESH, U. V. & SINGH, R. 2005. *Benzothiazole and thiazole[5,5-b] pyridine compositions and their use as ubiquitin ligase inhibitors.* PCT/US2004/034397.

LOR, L. A., SCHNECK, J., MCNULTY, D. E., DIAZ, E., BRANDT, M., THRALL, S. H. & SCHWARTZ, B. 2007. A simple assay for detection of small-molecule redox activity. *J Biomol Screen,* 12**,** 881-90.

LUKER, T., ALCARAZ, L., CHOHAN, K. K., BLOMBERG, N., BROWN, D. S., BUTLIN, R. J., ELEBRING, T., GRIFFIN, A. M., GUILE, S., ST-GALLAY, S., SWAHN, B. M., SWALLOW, S., WARING, M. J., WENLOCK, M. C. & LEESON, P. D. 2011. Strategies to improve in vivo toxicology outcomes for basic candidate drug molecules. *Bioorg Med Chem Lett,* 21**,** 5673-9.

MACARRON, R., BANKS, M. N., BOJANIC, D., BURNS, D. J., CIROVIC, D. A., GARYANTES, T., GREEN, D. V., HERTZBERG, R. P., JANZEN, W. P., PASLAY, J. W., SCHOPFER, U. & SITTAMPALAM, G. S. 2011. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov,* 10**,** 188-95.

MALVY, D. & CHAPPUIS, F. 2011. Sleeping sickness. *Clin.Microbiol.Infect.,* 17**,** 986-995.

MARKSON, G., KIEL, C., HYDE, R., BROWN, S., CHARALABOUS, P., BREMM, A., SEMPLE, J., WOODSMITH, J., DULEY, S., SALEHI-ASHTIANI, K., VIDAL, M., KOMANDER, D., SERRANO, L., LEHNER, P. & SANDERSON, C. M. 2009. Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. *Genome Res.,* 19**,** 1905-1911.

METZGER, M. B., HRISTOVA, V. A. & WEISSMAN, A. M. 2012. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci,* 125**,** 531-7.

MICHELLE, C., VOURC'H, P., MIGNON, L. & ANDRES, C. R. 2009. What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor? *J.Mol.Evol.,* 68**,** 616-628.

MIRKOVIC, B., SOSIC, I., GOBEC, S. & KOS, J. 2011. Redox-based inactivation of cysteine cathepsins by compounds containing the 4-aminophenol moiety. *PLoS One,* 6**,** e27197.

MORROW, J. K., LIN, H. K., SUN, S. C. & ZHANG, S. 2015. Targeting ubiquitination for cancer therapies. *Future Med Chem,* 7**,** 2333-50.

MUCHAMUEL, T., BASLER, M., AUJAY, M. A., SUZUKI, E., KALIM, K. W., LAUER, C., SYLVAIN, C., RING, E. R., SHIELDS, J., JIANG, J., SHWONEK, P., PARLATI, F., DEMO, S. D., BENNETT, M. K., KIRK, C. J. & GROETTRUP, M. 2009. A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat Med,* 15**,** 781-7.

MUKHOPADHYAY, D. & RIEZMAN, H. 2007. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science,* 315**,** 201-205.

MUNOS, B. 2009. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov,* 8**,** 959-68.

NAGLE, A. S., KHARE, S., KUMAR, A. B., SUPEK, F., BUCHYNSKYY, A., MATHISON, C. J., CHENNAMANENI, N. K., PENDEM, N., BUCKNER, F. S., GELB, M. H. & MOLTENI, V. 2014. Recent developments in drug discovery for leishmaniasis and human African trypanosomiasis. *Chem Rev,* 114**,** 11305-47.

NALEPA, G., ROLFE, M. & HARPER, J. W. 2006. Drug discovery in the ubiquitin-proteasome system. *Nat.Rev.Drug Discov.,* 5**,** 596-613.

NARENDRA, D. P. & YOULE, R. J. 2011. Targeting mitochondrial dysfunction: role for PINK1 and Parkin in mitochondrial quality control. *Antioxid Redox Signal,* 14**,** 1929-38.

NDUBAKU, C. & TSUI, V. 2015. Inhibiting the deubiquitinating enzymes (DUBs). *J Med Chem,* 58**,** 1581-95.

NEUBERT, K., MEISTER, S., MOSER, K., WEISEL, F., MASEDA, D., AMANN, K., WIETHE, C., WINKLER, T. H., KALDEN, J. R., MANZ, R. A. & VOLL, R. E. 2008. The proteasome inhibitor bortezomib depletes plasma cells and protects mice with lupus-like disease from nephritis. *Nat Med,* 14**,** 748-55.

NICHOLSON, B., KUMAR, S., AGARWAL, S., EDDINS, M. J., MARBLESTONE, J. G., WU, J., KODRASOV, M. P., LAROCQUE, J. P., STERNER, D. E. & MATTERN, M. R. 2014. Discovery of Therapeutic Deubiquitylase Effector Molecules: Current Perspectives. *J.Biomol.Screen.,* 19**,** 989-999.

NIJMAN, S. M., LUNA-VARGAS, M. P., VELDS, A., BRUMMELKAMP, T. R., DIRAC, A. M., SIXMA, T. K. & BERNARDS, R. 2005. A genomic and functional inventory of deubiquitinating enzymes. *Cell,* 123**,** 773-86.

OLINER, J. D., PIETENPOL, J. A., THIAGALINGAM, S., GYURIS, J., KINZLER, K. W. & VOGELSTEIN, B. 1993. Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53. *Nature,* 362**,** 857-60.

ORDUREAU, A., HEO, J. M., DUDA, D. M., PAULO, J. A., OLSZEWSKI, J. L., YANISHEVSKI, D., RINEHART, J., SCHULMAN, B. A. & HARPER, J. W. 2015. Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. *Proc Natl Acad Sci U S A,* 112**,** 6637-42.

PALOMBELLA, V. J., CONNER, E. M., FUSELER, J. W., DESTREE, A., DAVIS, J. M., LAROUX, F. S., WOLF, R. E., HUANG, J., BRAND, S., ELLIOTT, P. J., LAZARUS, D., MCCORMACK, T., PARENT, L., STEIN, R., ADAMS, J. & GRISHAM, M. B. 1998. Role of the proteasome and NF-kappaB in streptococcal cell wall-induced polyarthritis. *Proc Natl Acad Sci U S A,* 95**,** 15671-6.

PATEL, S. & PLAYER, M. R. 2008. Small-molecule inhibitors of the p53-HDM2 interaction for the treatment of cancer. *Expert Opin Investig Drugs,* 17**,** 1865-82.

PAUL, S. M., MYTELKA, D. S., DUNWIDDIE, C. T., PERSINGER, C. C., MUNOS, B. H., LINDBORG, S. R. & SCHACHT, A. L. 2010. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov,* 9**,** 203-14.

PENA, I. & DOMINGUEZ, J. M. 2010. Thermally denatured BSA, a surrogate additive to replace BSA in buffers for high-throughput screening. *J Biomol Screen,* 15**,** 1281-6.

PENA, I., PILAR MANZANO, M., CANTIZANI, J., KESSLER, A., ALONSO-PADILLA, J., BARDERA, A. I., ALVAREZ, E., COLMENAREJO, G., COTILLO, I., ROQUERO, I., DE DIOS-ANTON, F., BARROSO, V., RODRIGUEZ, A., GRAY, D. W., NAVARRO, M., KUMAR, V., SHERSTNEV, A., DREWRY, D. H., BROWN, J. R., FIANDOR, J. M. & JULIO MARTIN, J. 2015. New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource. *Sci Rep,* 5**,** 8771.

PEREZ-MANCERA, P. A., RUST, A. G., VAN DER WEYDEN, L., KRISTIANSEN, G., LI, A., SARVER, A. L., SILVERSTEIN, K. A., GRUTZMANN, R., AUST, D., RUMMELE, P., KNOSEL, T., HERD, C., STEMPLE, D. L., KETTLEBOROUGH, R., BROSNAN, J. A., LI, A., MORGAN, R., KNIGHT, S., YU, J., STEGEMAN, S., COLLIER, L. S., TEN HOEVE, J. J., DE RIDDER, J., KLEIN, A. P., GOGGINS, M., HRUBAN, R. H., CHANG, D. K., BIANKIN, A. V., GRIMMOND, S. M., AUSTRALIAN PANCREATIC CANCER GENOME, I., WESSELS, L. F., WOOD, S. A., IACOBUZIO-DONAHUE, C. A., PILARSKY, C., LARGAESPADA, D. A., ADAMS, D. J. & TUVESON, D. A. 2012. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature,* 486**,** 266-70.

PETERS, J. U., SCHNIDER, P., MATTEI, P. & KANSY, M. 2009. Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds. *ChemMedChem,* 4**,** 680-6.

PETERSON, L. F., SUN, H., LIU, Y., POTU, H., KANDARPA, M., ERMANN, M., COURTNEY, S. M., YOUNG, M., SHOWALTER, H. D., SUN, D., JAKUBOWIAK, A., MALEK, S. N., TALPAZ, M. & DONATO, N. J. 2015. Targeting deubiquitinase activity with a novel small-molecule inhibitor as therapy for B-cell malignancies. *Blood,* 125**,** 3588-97.

PHAM, L. V., TAMAYO, A. T., LI, C., BORNMANN, W., PRIEBE, W. & FORD, R. J. 2010. Degrasyn potentiates the antitumor effects of bortezomib in mantle cell lymphoma cells in vitro and in vivo: therapeutic implications. *Mol Cancer Ther,* 9**,** 2026-36.

PRIOTTO, G., KASPARIAN, S., MUTOMBO, W., NGOUAMA, D., GHORASHIAN, S., ARNOLD, U., GHABRI, S., BAUDIN, E., BUARD, V., KAZADI-KYANZA, S., ILUNGA, M., MUTANGALA, W., POHLIG, G., SCHMID, C., KARUNAKARA, U., TORREELE, E. & KANDE, V. 2009. Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. *Lancet,* 374**,** 56-64.

PULVINO, M., LIANG, Y., OLEKSYN, D., DERAN, M., VAN PELT, E., SHAPIRO, J., SANZ, I., CHEN, L. & ZHAO, J. 2012. Inhibition of proliferation and survival of diffuse large B-cell lymphoma cells by a small-molecule inhibitor of the ubiquitin-conjugating enzyme Ubc13-Uev1A. *Blood,* 120**,** 1668-77.

RANA, P., NAVEN, R., NARAYANAN, A., WILL, Y. & JONES, L. H. 2013. Chemical motifs that redox cycle and their associated toxicity. *MedChemComm,* 4**,** 1175-1180.

RATIA, K., PEGAN, S., TAKAYAMA, J., SLEEMAN, K., COUGHLIN, M., BALIJI, S., CHAUDHURI, R., FU, W., PRABHAKAR, B. S., JOHNSON, M. E., BAKER, S. C., GHOSH, A. K. & MESECAR, A. D. 2008. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc Natl Acad Sci U S A,* 105**,** 16119-24.

REVERDY, C., CONRATH, S., LOPEZ, R., PLANQUETTE, C., ATMANENE, C., COLLURA, V., HARPON, J., BATTAGLIA, V., VIVAT, V., SIPPL, W. & COLLAND, F. 2012. Discovery of specific inhibitors of human USP7/HAUSP deubiquitinating enzyme. *Chem Biol,* 19**,** 467-77.

REYES-TURCU, F. E., VENTII, K. H. & WILKINSON, K. D. 2009. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu.Rev.Biochem.,* 78**,** 363-397.

SANDERS, M. A., BRAHEMI, G., NANGIA-MAKKER, P., BALAN, V., MORELLI, M., KOTHAYER, H., WESTWELL, A. D. & SHEKHAR, M. P. 2013. Novel inhibitors of Rad6 ubiquitin conjugating enzyme: design, synthesis, identification, and functional characterization. *Mol Cancer Ther,* 12**,** 373-83.

SCHEPER, J., GUERRA-REBOLLO, M., SANCLIMENS, G., MOURE, A., MASIP, I., GONZALEZ-RUIZ, D., RUBIO, N., CROSAS, B., MECA-CORTES, O., LOUKILI, N., PLANS, V., MORREALE, A., BLANCO, J., ORTIZ, A. R., MESSEGUER, A. & THOMSON, T. M. 2010. Protein-protein interaction antagonists as novel inhibitors of non-canonical polyubiquitylation. *PLoS One,* 5**,** e11403.

SCHMIDT, N., GONZALEZ, E., VISEKRUNA, A., KUHL, A. A., LODDENKEMPER, C., MOLLENKOPF, H., KAUFMANN, S. H., STEINHOFF, U. & JOERIS, T. 2010. Targeting the proteasome: partial inhibition of the proteasome by bortezomib or deletion of the immunosubunit LMP7 attenuates experimental colitis. *Gut,* 59**,** 896-906.

SCHWICKART, M., HUANG, X., LILL, J. R., LIU, J., FERRANDO, R., FRENCH, D. M., MAECKER, H., O'ROURKE, K., BAZAN, F., EASTHAM-ANDERSON, J., YUE, P., DORNAN, D., HUANG, D. C. & DIXIT, V. M. 2010. Deubiquitinase USP9X stabilizes MCL1 and promotes tumour cell survival. *Nature,* 463**,** 103-7.

SHEKHAR, M. P., GERARD, B., PAULEY, R. J., WILLIAMS, B. O. & TAIT, L. 2008. Rad6B is a positive regulator of beta-catenin stabilization. *Cancer Res,* 68**,** 1741-50.

SKAAR, J. R., PAGAN, J. K. & PAGANO, M. 2014. SCF ubiquitin ligase-targeted therapies. *Nat Rev Drug Discov,* 13**,** 889-903.

SOARES, K. M., BLACKMON, N., SHUN, T. Y., SHINDE, S. N., TAKYI, H. K., WIPF, P., LAZO, J. S. & JOHNSTON, P. A. 2010. Profiling the NIH Small Molecule Repository for compounds that generate H2O2 by redox cycling in reducing environments. *Assay Drug Dev Technol,* 8**,** 152-74.

SPRATT, D. E., WALDEN, H. & SHAW, G. S. 2014. RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochem J,* 458**,** 421-37.

SUN, L., XING, Y., CHEN, X., ZHENG, Y., YANG, Y., NICHOLS, D. B., CLEMENTZ, M. A., BANACH, B. S., LI, K., BAKER, S. C. & CHEN, Z. 2012. Coronavirus papain-like proteases negatively regulate antiviral innate immune response through disruption of STING-mediated signaling. *PLoS One,* 7**,** e30802.

SWINNEY, D. C. & ANTHONY, J. 2011. How were new medicines discovered? *Nat Rev Drug Discov,* 10**,** 507-19.

SYKES, M. L. & AVERY, V. M. 2009. A luciferase based viability assay for ATP detection in 384-well format for high throughput whole cell screening of Trypanosoma brucei brucei bloodstream form strain 427. *Parasit Vectors,* 2**,** 54.

TAYA, S., YAMAMOTO, T., KANAI-AZUMA, M., WOOD, S. A. & KAIBUCHI, K. 1999. The deubiquitinating enzyme Fam interacts with and stabilizes beta-catenin. *Genes Cells,* 4**,** 757-67.

TCHERPAKOV, M. 2013. *Ubiquitin Proteasome: Global Markets for Research Tools, Diagnostics and Drug Discovery*, BCC Research.

THROWER, J. S., HOFFMAN, L., RECHSTEINER, M. & PICKART, C. M. 2000. Recognition of the polyubiquitin proteolytic signal. *EMBO J,* 19**,** 94-102.

TIAN, Z., D'ARCY, P., WANG, X., RAY, A., TAI, Y. T., HU, Y., CARRASCO, R. D., RICHARDSON, P., LINDER, S., CHAUHAN, D. & ANDERSON, K. C. 2014. A novel small molecule inhibitor of deubiquitylating enzyme USP14 and UCHL5 induces apoptosis in multiple myeloma and overcomes bortezomib resistance. *Blood,* 123**,** 706-16.

TIRAT, A., SCHILB, A., RIOU, V., LEDER, L., GERHARTZ, B., ZIMMERMANN, J., WORPENBERG, S., EIDHOFF, U., FREULER, F., STETTLER, T., MAYR, L., OTTL, J., LEUENBERGER, B. & FILIPUZZI, I. 2005. Synthesis and characterization of fluorescent ubiquitin derivatives as highly sensitive substrates

for the deubiquitinating enzymes UCH-L3 and USP-2. *Anal.Biochem.,* 343**,** 244-255.

TOPISIROVIC, I., GUTIERREZ, G. J., CHEN, M., APPELLA, E., BORDEN, K. L. & RONAI, Z. A. 2009. Control of p53 multimerization by Ubc13 is JNK-regulated. *Proc Natl Acad Sci U S A,* 106**,** 12676-81.

TSUKAMOTO, S., TAKEUCHI, T., ROTINSULU, H., MANGINDAAN, R. E., VAN SOEST, R. W., UKAI, K., KOBAYASHI, H., NAMIKOSHI, M., OHTA, T. & YOKOSAWA, H. 2008. Leucettamol A: a new inhibitor of Ubc13-Uev1A interaction isolated from a marine sponge, Leucetta aff. microrhaphis. *Bioorg Med Chem Lett,* 18**,** 6319-20.

TURK, B. 2006. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov,* 5**,** 785-99.

ULRICH, H. D. 2005. The RAD6 pathway: control of DNA damage bypass and mutagenesis by ubiquitin and SUMO. *Chembiochem,* 6**,** 1735-43.

USHIYAMA, S., UMAOKA, H., KATO, H., SUWA, Y., MORIOKA, H., ROTINSULU, H., LOSUNG, F., MANGINDAAN, R. E., DE VOOGD, N. J., YOKOSAWA, H. & TSUKAMOTO, S. 2012. Manadosterols A and B, sulfonated sterol dimers inhibiting the Ubc13-Uev1A interaction, isolated from the marine sponge Lissodendryx fibrosa. *J Nat Prod,* 75**,** 1495-9.

VALLER, M. J. & GREEN, D. 2000. Diversity screening versus focussed screening in drug discovery. *Drug Discov Today,* 5**,** 286-293.

VAN DER HORST, A., DE VRIES-SMITS, A. M., BRENKMAN, A. B., VAN TRIEST, M. H., VAN DEN BROEK, N., COLLAND, F., MAURICE, M. M. & BURGERING, B. M. 2006. FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. *Nat Cell Biol,* 8**,** 1064-73.

VIJAY-KUMAR, S., BUGG, C. E. & COOK, W. J. 1987. Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol,* 194**,** 531-44.

WARING, M. J., ARROWSMITH, J., LEACH, A. R., LEESON, P. D., MANDRELL, S., OWEN, R. M., PAIRAUDEAU, G., PENNIE, W. D., PICKETT, S. D., WANG, J., WALLACE, O. & WEIR, A. 2015. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov,* 14**,** 475-86.

WEI, R., LIU, X., YU, W., YANG, T., CAI, W., LIU, J., HUANG, X., XU, G. T., ZHAO, S., YANG, J. & LIU, S. 2015. Deubiquitinases in cancer. *Oncotarget,* 6**,** 12872-89.

WHO. Available: http://www.who.int/entity/en/ [Accessed].

WILKINSON, K. D., URBAN, M. K. & HAAS, A. L. 1980. Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J Biol Chem,* 255**,** 7529-32.

WILKINSON, S. R. & KELLY, J. M. 2009. Trypanocidal drugs: mechanisms, resistance and new targets. *Expert Rev Mol Med,* 11**,** e31.

WILKINSON, S. R., TAYLOR, M. C., HORN, D., KELLY, J. M. & CHEESEMAN, I. 2008. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. *Proc Natl Acad Sci U S A,* 105**,** 5022-7.

WINDHEIM, M., PEGGIE, M. & COHEN, P. 2008. Two different classes of E2 ubiquitin-conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. *Biochem J,* 409**,** 723-9.

XU, G. W., ALI, M., WOOD, T. E., WONG, D., MACLEAN, N., WANG, X., GRONDA, M., SKRTIC, M., LI, X., HURREN, R., MAO, X., VENKATESAN, M., BEHESHTI ZAVAREH, R., KETELA, T., REED, J. C., ROSE, D., MOFFAT, J., BATEY, R. A., DHE-PAGANON, S. & SCHIMMER, A. D. 2010. The ubiquitin-activating enzyme E1 as a therapeutic target for the treatment of leukemia and multiple myeloma. *Blood,* 115**,** 2251-9.

YAMAMOTO, M., OKAMOTO, T., TAKEDA, K., SATO, S., SANJO, H., UEMATSU, S., SAITOH, T., YAMAMOTO, N., SAKURAI, H., ISHII, K. J., YAMAOKA, S., KAWAI, T., MATSUURA, Y., TAKEUCHI, O. & AKIRA, S. 2006. Key function for the Ubc13 E2 ubiquitin-conjugating enzyme in immune receptor signaling. *Nat Immunol,* 7**,** 962-70.

YANG, Y., KITAGAKI, J., DAI, R. M., TSAI, Y. C., LORICK, K. L., LUDWIG, R. L., PIERRE, S. A., JENSEN, J. P., DAVYDOV, I. V., OBEROI, P., LI, C. C., KENTEN, J. H., BEUTLER, J. A., VOUSDEN, K. H. & WEISSMAN, A. M. 2007. Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics. *Cancer Res,* 67**,** 9472-81.

YAO, T., SONG, L., XU, W., DEMARTINO, G. N., FLORENS, L., SWANSON, S. K., WASHBURN, M. P., CONAWAY, R. C., CONAWAY, J. W. & COHEN, R. E. 2006. Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1. *Nat Cell Biol,* 8**,** 994-1002.

YAU, R. & RAPE, M. 2016. The increasing complexity of the ubiquitin code. *Nat Cell Biol,* 18**,** 579-86.

YE, Y. & RAPE, M. 2009. Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol,* 10**,** 755-64.

YOUNG, R. J., GREEN, D. V., LUSCOMBE, C. N. & HILL, A. P. 2011. Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discov Today,* 16**,** 822-30.

YUE, W., CHEN, Z., LIU, H., YAN, C., CHEN, M., FENG, D., YAN, C., WU, H., DU, L., WANG, Y., LIU, J., HUANG, X., XIA, L., LIU, L., WANG, X., JIN, H., WANG, J., SONG, Z., HAO, X. & CHEN, Q. 2014. A small natural molecule promotes mitochondrial fusion through inhibition of the deubiquitinase USP30. *Cell Res,* 24**,** 482-96.

ZHANG, J. H., CHUNG, T. D. & OLDENBURG, K. R. 1999. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen,* 4**,** 67-73.

ZHOU, M., WANG, T., LAI, H., ZHAO, X., YU, Q., ZHOU, J. & YANG, Y. 2015. Targeting of the deubiquitinase USP9X attenuates B-cell acute lymphoblastic leukemia cell survival and overcomes glucocorticoid resistance. *Biochem Biophys Res Commun,* 459**,** 333-9.