The University Of Sheffield.

DNA replication inhibitor as a potential antimicrobial agent targeting neglected tropical pathogens

By:

Arturo Landeros de la Isla

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Medicine, Dentistry and Health

Department of Infection, Immunity and Cardiovascular Disease

November 2022

Abstract

Tropical diseases are a significant economic burden for developing nations, including Leishmaniasis, African sleeping sickness, and cryptococcosis caused by *Leishmania* spp., *Trypanosoma brucei*, and *Cryptococcus neoformans*, respectively. These diseases can be fatal without treatment. Current medicines are restricted by toxicity and resistance. Therefore, new treatments targeting these organisms are essential.

Flap endonucleases (FENs) form a class of enzymes presented in all living beings and are involved in maintenance of the genome, assuming a significant part in DNA replication and DNA repair. Thus, it is likely to be vital for parasite and fungal survival, and subsequently potential drugs can be designed to target FENs. This project was developed using experimental and *in silico* techniques to specifically target microbial FENs by understanding the molecular mechanism of the enzyme and its biological structure.

In silico studies were carried out using small molecule structures retrieved from publicly available databases and structures of the microbial FENs experimentally determined or predicted through AI algorithms. The *in silico* screening studies using Autodock and Autodock Vina identified potential inhibitors according to a binding energy score and further analysis was carried out. Additionally, over 1500 molecules were tested *in vitro*. Selective inhibitors were identified with IC_{50} as low as 23 μM.

Crystallisation screens for recombinant FEN proteins were executed with the aim of elucidating experimental structures of microbial FENs. Protein crystals of *Trypanosoma*, *Leishmania* and *Cryptococcus* FEN were produced under different screening conditions. One of the conditions formed a *Leishmania* FEN crystal with a diffracting resolution of 1.4 Å, in which the data was used to model the protein structure.

This work has potential impact in the future development of antimicrobial drugs and further studies can be performed (fragment structure optimization, *in vivo* assays and co-crystallisation of microbial FENs with inhibitors) to enrich structure-based drug design campaigns.

Acknowledgements

The work carried out was funded by the National Council of Science and Technology (CONACYT – CVU: 844464), to whom I really appreciate the support.

I would like to express my immensely gratitude to Prof. Jon Sayers for his continuous support and guidance throughout the project. I would also like to thank all the people from the Sayers group for the useful advice in regard to this work.

Additionally, I would like to express my sincere appreciation to all my friends that I have met during this PhD journey, from people that I met in one of many social events and night outs. Also, I would like to thank the Latin American Society from the University of Sheffield for organizing events and making me feel home when I was feeling homesick.

Also, I am very thankful to my closest circle of friends, or as I like to call it, the “Mexican PhD in Medicine friends”: Héctor, Brenda, Sebastiano, Ingrid, Dylan, Manuel, Claudia, and Ernesto, I really appreciate the friendship we established since the first day I met you and all the fun we had over a couple of pints, singing and dancing. I am certain that our friendship will last forever even if we follow different paths at the end of our PhDs.

Last but not least, I would like to thank my family back in Mexico, my parents and my sisters Angélica, Laura and Ana María for the immense support and the encouraging words to keep me concentrated in this work and make it possible. I do really appreciate all the gestures that my family had on me because this allowed me to continue and finally conclude this work.

Content page

List of Figures

List of Tables

Abbreviations

°C - Degrees Celsius

$A_{600\ nm}$ – Absorbance at 600 nm

AAT - Animal African trypanosomiasis

ADMET – Adsorption, distribution, metabolism, excretion, and toxicity

AI – Artificial Intelligence

AIDS - Acquired immunodeficiency syndrome

APE1 – Apurinic/apyrimidinic endonuclease

APS – Ammonium persulfate

ATP – Adenosine triphosphate

BLM – Bloom syndrome protein

BSA – Bovin serum albumin

cDNA – Complementary deoxyribonucleic acid

CL - Cutaneous leishmaniasis

CFU – Colony-forming unit

*Cn*FEN – *Cryptococcus neoformans* flap endonuclease

*Cn*GPD – *Cryptococcus neoformans* glyceraldehyde 3-phosphate dehydrogenase

CNS – Central Nervous System

CoA – Coenzyme A

CPB2 – Cysteine protease type 2

Da - Dalton

DALY - Disability adjusted life years lost

DHFR – Dihydrofolate reductase

DMSO – Dimethylsulfoxide

DNA - Deoxyribonucleic acid

dsDNA – Double-stranded DNA

dTMP – Deoxythymidine monophosphate

DTT – Ditriothreitol

dUMP – Deoxyuridine monophosphate

EDTA - Ethylenediaminetetraacetic acid

FAD – Flavin adenine dinucleotide

FBDD – Fragment based drug design

FBLD – Fragment based lead discovery

FDA - Food and Drug Administration

FEN - Flap endonuclease

FL – Full-length

FRET – Förster Resonance Energy Transfer

GBD - Global of Burden Disease Study

GPD – Glyceraldehyde 3-phosphate dehydrogenase

GUI – Graphical User Interface

HAT - Human African trypanosomiasis

His-tag – Polyhistidine tag

HIV - Human immunodeficiency virus

HRV – Human rhinovirus

*Hs*FEN - human flap endonuclease

Hsp90 – Heat shock protein 90

HTS – High-throughput screening

IFN – Interferon

IHME – Institution for Health Metrics and Evaluation

IL – interleukin

IPTG - Isopropyl β-D-1-thiogalactopyranoside

ISDD – *In silico* drug design

ITC – Isothermal titration calorimetry

Kbp – Kilobase pair

kDa - kilodalton

LB – Lysogeny broth

LBDD – Ligand based drug design

*Li*FEN - *Leishmania infantum* flap endonuclease

LP-BER – Long patch base excision repair

MAP – Methyonilaminopeptidase

MDA – Mass Drug Administration

mRNA – Messenger ribonucleic acid

NADH – Nicotinamide adenine dinucleotide

NADPH – Nicotinamide adenine dinucleotide phosphate

NCBI – National Center for Biotechnology Information

NCE – New chemical entity

NDH-2 – Type 2 NADH dehydrogenase

NME – Novel molecule entity

NMR – Nuclear Magnetic Resonance

NMT – N-Myristoyltransferase

nt – Nucleotide(s)

NTD - Neglected Tropical Disease

PAHO - Pan American Health Organization

PBS – Phosphate-buffered saline

PCNA – Proliferating cell nuclear antigen

PCR – Polymerase chain reaction

PDB – Protein Data Bank

PDBQT – Protein Data Bank, Partial Charge (Q), & Atom Type (T)

PEG – Polyethylene glycol

PEI – Polyethylenimine

PFK – Phosphofructokinase

pI – Isoelectric point

PMSF – Phenylmethanesulfonyl fluoride

Pol – Polymerase

PTR1 – Pteridine reductase

PyK – Pyruvate Kinase

QSAR – Quantitative structure-activity relationship

R&D – Research & development

RCSB – Research Collaboratory for Structural Bioinformatics

RFC – Replication factor C

RFU – Relative fluorescent units

RNA - Ribonucleic acid

RPA - Replication protein A

RT-PCR – Reverse transcription polymerase chain reaction

SBDD – Structure based drug design

SDF – Structure data file

SEC – Size exclusion chromatography

SMARTS – SMILES arbitrary target specification

SMILES – Simplified molecular-input line-entry system

SOD – Sodium deoxycholate

SOE – Splicing by Overlap Extension

SPR – Surface plasmon resonance

ssDNA – Single-stranded DNA

SVBS – Structure-based virtual screening

*Tb*FEN - *Trypanosoma brucei* flap endonuclease

TEMED – tetramethylethylenediamine

TGF – Transforming growth factor

TNF – Tumour necrosis factor

TRF2 – Telomeric-repeat binding factor 2

TS – Thymidylate synthase

UPGMA - Unweighted Pair Group Method with Arithmetic Mean

UV – Ultraviolet

vHTS – Virtual high-throughput screening

VL - Visceral leishmaniasis

w/v – Weight/volume

WHO - World Health Organization

WRN – Werner Syndrome Protein

WT – Wild type

YLD - Years lived with disability

YLL - Years of life lost

YPD – Yeast peptone dextrose

CHAPTER ONE: INTRODUCTION

1.1 Neglected Tropical Diseases

Neglected Tropical Diseases (NTDs) consist of a group of twenty infectious diseases which are caused by pathogens such as parasites, bacteria and viruses. They affect principally populations in tropical and subtropical region with low and middle-income such as various countries in Africa, Asia and America as shown in Figure 1.1 (Hollingsworth, 2018; Engels & Zhou, 2020).

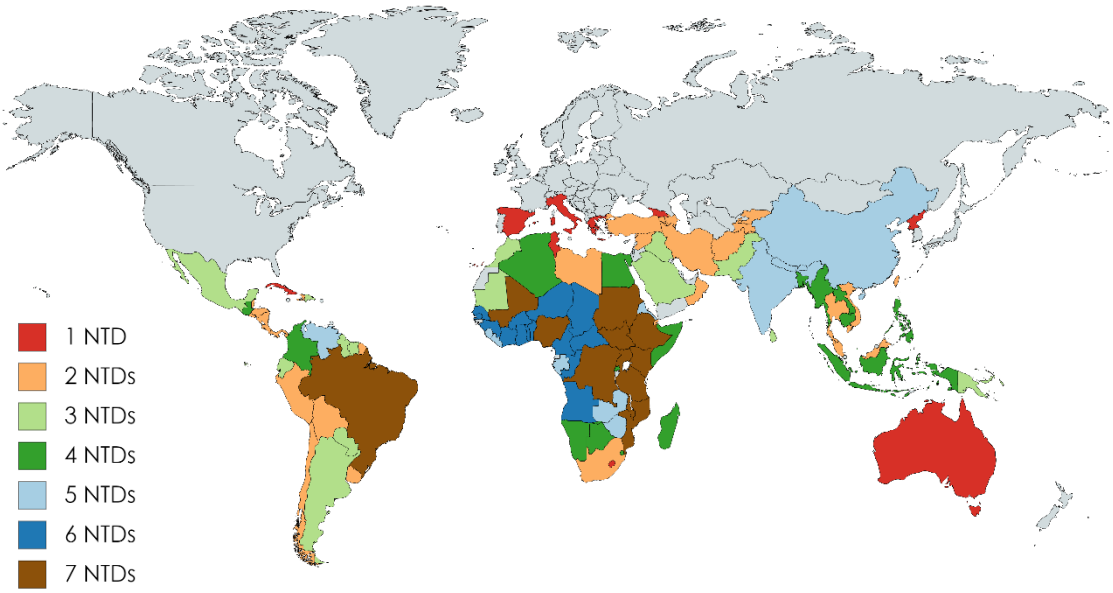


Figure 1.1 Presence of 10 NTDs around the world

This map displays countries endemic for ten NTDs according to the 2009-2010 data and international borders. (blinding trachoma, Chagas disease, guinea worm, sleeping sickness, lymphatic filariasis, leprosy, river blindness, soil-transmitted helminths, schistosomiasis, visceral leishmaniasis). Image reproduced from Uniting to Combat NTDs (2012) "Neglected tropical diseases burden map (2009/2010)" (https://unitingtocombatntds.org/wp-content/uploads/2017/11/burden_of_neglected_tropical_diseases.pdf).

These infections, which can be fatal in a significant proportion of cases, have affected over a billion individuals (Table 1.1). The majority of the population infected by NTDs live in precarious state, with hazardous water, poor sanitation and housing conditions, as well as a deficient wellbeing administration (Molyneux, 2013; Hollingsworth, 2018).

Table 1.1 Core group of the twenty NTDs according to WHO*

Infection type	Disease/Pathogen name	Number of cases/deaths reported
Helminth (worm) infection	Soil-transmitted helminth infection	Estimation of 1.5 billion people infected in 2016. About 6300 deaths reported in 2016.
	Schistosomiasis (snail fever)	Nearly 236 million people required treatment in 2019. Approximately 24,000 deaths registered in 2016.
	Lymphatic filariasis (elephantiasis)	Nearly 51.4 million people infected in 2018.
	Onchocerciasis (river blindness)	About 21 million estimated cases in 2017. 218 million people at risk in 2019.
	Food-borne trematode infections	200,000 new cases annually. Approximately 7000 deaths annually.
	Taeniasis and Cysticercosis	About 5.5 million people infected worldwide. 28 000 registered deaths in 2015.
	Human echinococcosis (hydatid cyst)	At least 1 million people infected in 2011. About 19,000 deaths registered in 2016.
	Dracunculiasis (guinea worm infection)	54 cases reported in 2019.
Protozoan infections	Leishmaniasis	Nearly 300,000 cases reported in 2019. 491 deaths reported among patients in seven countries in 2018 (Bangladesh, Brazil, Ethiopia, Nepal, Somalia South Sudan and Sudan).
	Chagas disease	6 – 7 million infected in 2019. About 10,000 deaths registered in 2017.
	Human African trypanosomiasis (sleeping sickness)	Fewer than 1000 reported cases of *T. b. gambiense* and 116 cases of *T. b. rhodesiense* in 2019. 57 million people at risk.
Bacterial infections	Trachoma	About 2.5 million people requiring surgery for trachomatous trichiasis in 2019.
	Buruli ulcer	2260 cases in 2019.
	Leprosy	202,226 infected people registered in 2019.
	Yaws and endemic treponematoses	1177 confirmed cases of yaws with 80 247 suspected cases during 2018.
Viral infections	Dengue and Chikungunya	Estimation of 104 million cases of dengue in 2017. Nearly 40,000 reported deaths in 2016. 8 million chikungunya infections reported during 2004-2017
	Rabies	Nearly 29 million people received urgent treatment in 2015. About 59,000 deaths registered in 2015.
Fungal infections	Mycetoma, chromoblastomycosis and other deep mycoses	At least 10 000 cases of chromoblastomycosis, paracoccidioidomycosis and sporotrichosis recorded globally since 1940s. The exact burden is unknown and probably higher.
Ectoparasitic infections	Scabies and other ectoparasites	About 200 million people affected by scabies in 2016.
Venomous threat	Snakebite envenoming	2.7 million people bitten with envenoming annually. Between 8000 and 140,000 registered deaths per year.

**(Hotez, 2013; WHO, 2020).*

The Institute for Health Metrics and Evaluation (IHME) monitors the prevalence, morbidity, mortality, injury and risk factors for hundreds of disease. In their 2010 study, Global Burden of Disease Study (GBD), they found that some NTDs such as rabies, African Sleeping sickness, and visceral leishmaniasis are lethal sickness as indicated by the high number of deaths from these diseases within the infected population (IHME, 2013; Hotez *et al.*, 2016). While the rest of the NTDs are exceedingly disabling and all the NTDs can be categorized as chronic in nature (Figure 1.2) contributing to massive economic burdens for those affected and their families.

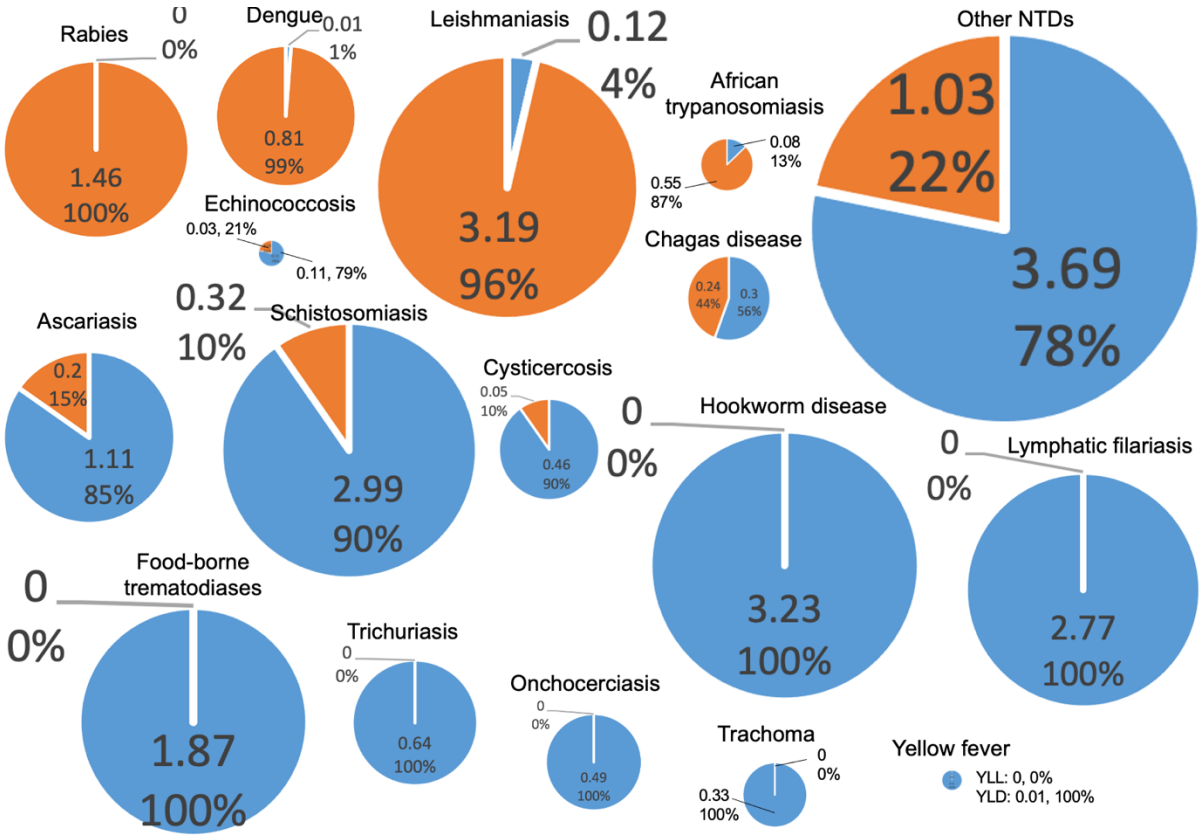


Figure 1.2 Fractions of years lived with disability (YLD), coloured in blue, and years of life lost (YLL), coloured in orange, in millions due to premature death

The figure shows components of disability-adjusted life years lost (DALYs) for each of the NTDs. Diagram based on the data from Hotez et al.*, (2014) obtained from the Global Burden of Disease Study 2010 (IHME, 2013).*

The relevant symptoms appearing during the development of the NTD can become progressively worse if the affliction has not been identified, and the individual has not been treated appropriately. This often results in an irreversible harm and severe pain, inability to function, adding to the misery of the patients concerned. In light of the symptoms, individuals with NTDs have displayed, for example, visual impairment, deformations, neurological issues, and cancers, often leading to frequent social stigmatization and social exclusion (Molyneux, 2013). For example, leprosy, caused by the bacterium *Mycobacterium leprae*, was handled by enforced exclusion of infected individual in leper colonies.

During the Middle Ages, leprosy was considered a contagious and incurable disease, widespread in European countries such as Denmark, southern Germany, England, Ireland, France, Italy, Czech Republic and Hungary (Petruševski, 2013; Spigelman & Rubini, 2016). An untreatable disease during medieval times, there was a misconception of lepers being cursed and living in sin, treating them inhumanly and finally excluding them from society in quarantined colonies (Santacroce *et al.,* 2021). It is considered that, by the year 1300, 19,000 leper colonies, often called lazar houses or leprosariums, existed in Europe and Mediterranean countries; with nearly 230 leper houses operating in the British Islands from the 10th century until the end of 19th century (Tulchinsky & Varavikova, 2014; Newman, 1895). Thankfully, attitudes to leprosy have largely changed and drug treatments such as rifampicin, dapsone and clofazimine have reduced this particular problem to a considerable extent (Lazo-Porras *et al.*, 2020). However, social exclusion is still a major problem with many NTDs throughout the world.

So why have NTDs proven to be so understudied? The answer is fairly obvious on the grounds that the victims of this group of diseases historically originate from poor communities in countries lacking advanced healthcare or even basic public health facilities. This meant that treatments for these diseases were not developed at the same rate and intensity as has been the case for infections such as HIV. Though originated in Africa (Sharp & Hahn, 2011) it became widespread around the world from the early 1980s, receiving significant attention from the pharmaceutical companies. The first antiviral drug (zidovudine, azidothymidine or AZT) being approved by FDA for the treatment of AIDS was in 1987 (Brook, 1987). Similarly, common western diseases such as cancer, cardiovascular diseases, diabetes, inflammatory disorders and

digestive diseases have proved financially attractive areas for drug development (Troulle*r et al.*, 2002; Pedrique *et al.*, 2013; Barrenho *et al.*, 2019). For example, drugs such as etanercept, adalimumab and infliximab, which are tumor necrosis factor (TNF) blockers and are used as treatments for rheumatoid arthritis, psoriasis and ankylosing spondylitis, have a cost ranging from $15,000 to $30,000 (Schabert *et al.*, 2013). Clearly, such high costs could never be afforded by patients in the majority of countries affected by NTDs (LMICs, low or middle income countries). This combined with the undoubtedly high costs of drug development for which the business model has required development costs are recovered from future sales. Given the very high costs of drug development, estimated between 1.2 and 1.3 billion dollars per drug, it is not surprising that drug companies have not been attracted to this area of research and development (DiMasi & Grabowski, 2007).

Clearly, in many of the regions affected by NTDs, the economic infrastructure needed to carry out drug discovery and development is lacking (Hotez; 2013). However, in recent years, possibly due to a combination of global population migrations, increased tourism from the richer nations into areas with endemic NTDs, and global warming, these infections have become the focus of significant international research efforts and some notable successes. Donation programs from 13 pharmaceutical companies, in collaboration with the governments of the United States, United Kingdom and the United Arab Emirates; the Bill and Melinda Gates foundation; the World Bank; and other global health organizations, have been set up aim of eliminating or controlling 10 NTDs by 2020 according to the London Declaration on Neglected Tropical Diseases, signed in 2012. Examples of those donation programs are: Merck's donation of ivermectin to treat onchocerciasis in 1987; and the Novartis Malaria Initiative, established in 2000, which has delivered more than 600 million artemether-lumefantrine fixed-dose combination treatments for malaria at reduced cost (Cohen *et al.*, 2016).

In summary, it is important to understand how NTDs had been ignored by different social strata, starting at a local level because NTDs can cause fear due to the developed symptoms. Likewise, NTDs had been overlooked at a national level because these diseases are usually spread in rural areas, and national health system did not prioritize them earlier enough (Hotez, 2013).

1.2 Kinetoplastid Infections

From the 20 NTDs that has been described, the project was focused mainly in the diseases from a specific infection type, the protozoan (or also known as kinetoplastids). The organisms of this core group are widespread around the world, with high infection rates and becoming an extensive threat to the population with poor health services. In addition, the current treatments for the diseases possess serious side effects and novel treatments have not been developed in the last 50 years. Therefore, looking for new treatments for these diseases has been a medical research priority.

The organisms of the kinetoplastids order are eukaryotes, which implies the prescence of certain characteristic subcellular structures, for example, the existence of a nucleus and organelles inside the cell, mitochondria, peroxisomes, rough and smooth endoplasmic reticulum, and a single morphologically identifiable Golgi apparatus (Clayton *et al.*, 1995). The kinetoplastids have particular features that can only be found within this group of parasites. For instance, these organisms have just a single mitochondrion in the cell, which is extremely large and elongated compared to a typical mammalian or yeast mitochondrion. For example, a typical yeast mitochondrial compartment might measure between 339 and 360 nm in diameter, whilst the kinetoplast one, depending of the organism, can measure 2 μm in diameter, in the case of *Perkinsela* sp., which is 4 times larger than *Trypanosoma brucei* mitochondrion. (Lukeš *et al*., 2018). However, the most obvious subcellular feature of these parasites is that the entire mitochondrial DNA is consolidated in a structure called the kinetoplast (Benne *et al.*, 1986; Lukeš *et al*., 2018).

Parasites of the Kinetoplastida class are responsible of infections affecting people, animals, and plants. In human infections, *Trypanosoma brucei* causes African sleeping sickness, *Trypanosoma cruzi* is the parasite responsible for Chagas disease and the *Leishmania* species causes the sickness called leishmaniasis (Hotez, 2013); every one of these parasites are transmitted by an insect vector and kill around 70,000 individuals yearly (Lozano *et al.*, 2012; Hotez *et al.*, 2014).

One of the greatest challenges to fight the NTDs is ready access to basic drugs, just as the advancement of new medicines and to improve the ones which already exist because the medications used to treat kinetoplastid infections were created in the mid 1900s (Johnston *et al.*, 2015)

1.2.1 Leishmania

Leishmania species are parasitic protozoans which belongs to the same family as the organisms *Trypanosoma brucei* and *Trypanosoma cruzi*, the Trypanosomatidae (Boucinha *et al.*, 2020). The former parasite can be transmitted by phlebotomine sand flies and infects animals and humans, inoculating flagellated forms of the parasite (known as promastigotes) into the skin. There are two main clinical manifestations of human infection: visceral leishmaniasis (VL) and cutaneous leishmaniasis (CL). There are different species of *Leishmania* parasites that cause either VL or CL (Table 1.2). The last one is not fatal, yet the disease can incite social stigmatization because of the disturbing appearance of the symptoms of this sickness. It can cause skin lesions, which develops into an ulcer over a period of weeks to months, leaving permanent scars. In the case of mucocutaneous leishmaniasis, the disease can cause ulcerations in the cartilage of the nose and subsequently deformation of the face (Van der Auwera and Dujardin, 2015; Hotez, 2013).

Table 1.2 Epidemiology characteristics of *Leishmania* species*

Species	Reservoir	Transmission[1]	Clinical Syndrome
L. donovani	Humans, rodents, canines	Anthroponotic, zoonotic	VL
L. infantum	Humans, canines	Anthroponotic, zoonotic	VL, CL
L. tropica	Humans	Anthroponotic	CL
L. major	Rodents, gerbils	Zoonotic	CL
L. aethiopica	Hyraxes	Zoonotic	CL

**Table adapted from Sundar (2015). [1]Anthroponotic: Disease that can be transmitted from a human host to an animal or human host; Zoonotic: Disease that can be transmitted from an animal host to a human or animal host.*

Around 12 million people are infected with *Leishmania* parasites around the world, and every year this sickness strikes between of 1.5 and 2 million individuals causing the death of a greater amount of 50,000 (Sundar, 2015; Hotez *et al.*, 2014). Leishmaniasis is a transcendently rural illness and is endemic in 75 countries for VL and 87 countries for CL. Nevertheless, in 2017, 95% of the new reported VL cases were concentrated

in only seven nations: Brazil, Ethiopia, India, Kenya, Somalia, South Sudan and Sudan. Simultaneously, 80% of the CL cases were registered in Afghanistan, Algeria, Brazil, Colombia, Iran, Iraq, Pakistan, Syria and Yemen during 2016 (WHO, 2020). Leishmaniasis is one of the most significant protozoan diseases that extensively affects people based on healthy life years lost from death or disability (Figure 1.2).

Leishmaniasis is transmitted by the bite of sandflies, which inoculate flagellated forms of the parasite (known as promastigotes) into the skin. Then, the *Leishmania* parasites attract the macrophages with the purpose of being ingested by them and once inside, these parasites can multiply as amastigote forms. Finally, the macrophages rupture, releasing new amastigotes that invade new macrophages. In the case of CL, the infected macrophages and amastigotes remain in the skin, while in VL, the macrophages travel to the liver, spleen, and bone marrow, causing a significant systemic disease. The life cycle of *Leishmania* spp. is shown in Figure 1.3 (Hotez, 2013; Lidani *et al.,* 2017).

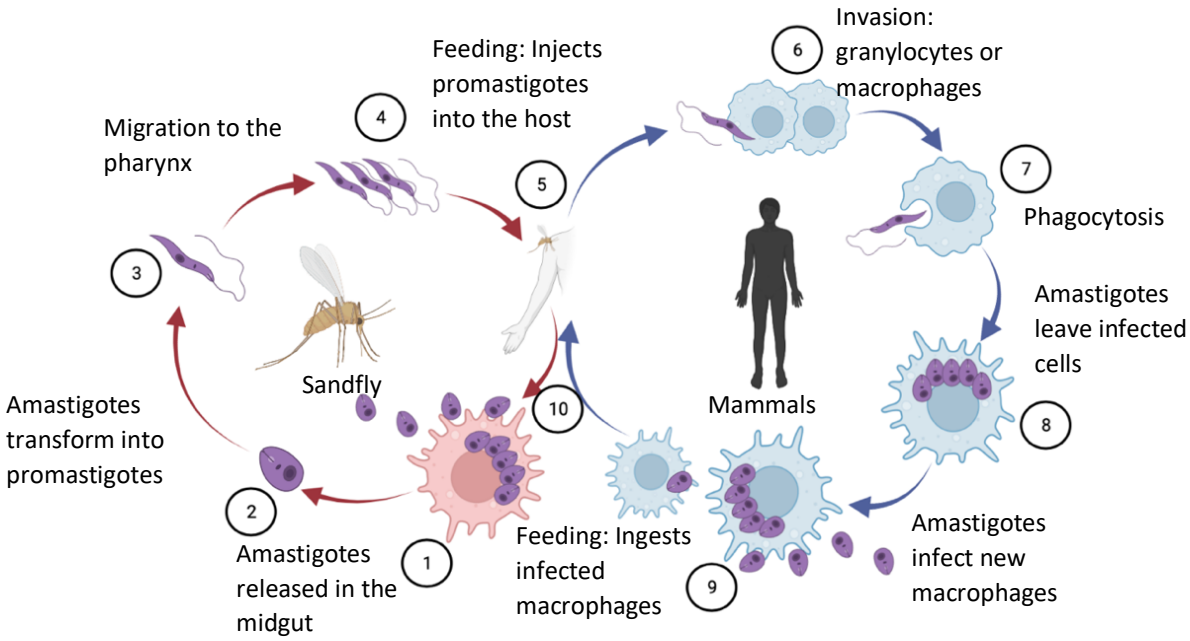


Figure 1.3 Life cycle of *Leishmania* spp.

Leishmaniasis is transmitted by the bites of infected sandflies during their blood meals (Lidani et al.*, 2017). The infection in humans start when the insect injects the metacyclic promastigote forms (5), which are engulfed by macrophages and neutrophils at the bite site (6). Inside the cells, promastigotes transform into amastigotes (7 & 8), which will then reproduce and progress to infect new phagocytic cells (9; Lidani* et al.*, 2017). Interactions between parasite, host and other factors will determine whether the infection progress to cutaneous or visceral leishmaniasis (Lidani* et al.*,*

2017). In the case of CL, the infected macrophages and amastigotes remain in the skin, while in VL, the macrophages travel to the liver, spleen, and bone marrow, causing a significant systemic disease. This figure was created using BioRender based on the life cycle depicted in Lidani et al.*, 2017.*

1.2.2 Trypanosomes

Trypanosomes are unicellular protozoan living beings within which 19 species and subspecies have been described in the literature (Kasozi *et al.*, 2021). Various types of trypanosomes have the facility to infect animals and humans (Bruckner & Labarca, 2015). However, there are two unique sorts of trypanosomes that affects humans and have become a concern to public health. Their life cycles and mechanisms of disease are distinctive one from the other (WHO, 2020; Table 1.3).

Table 1.3 Characteristics of the different parasites of trypanosomiasis*

Specie	Vector	Reservoir	Illness	Epidemiology[1]
T. brucei rhodesiense	Tsetse fly (*Glossina*)	Animals	Acute, <9 months	Anthropozoonosis
T. brucei gambiense	Tsetse fly (*Glossina*)	Humans	Chronic, months to years	Anthroponosis
T. cruzi	Reduviid insect (*Panstrongylus, Rhodnius, Triatoma*)	Animals	Acute, chronic	Anthropozoonosis

** Table adapted from Bruckner & Labarca (2015). [1]Anthropozoonosis, transmission involving a human-animal-human cycle; anthroponosis, transmission involving a human-human cycle.*

1.2.2.1 *Trypanosoma cruzi*

Chagas disease, or American trypanosomiasis, is caused by the protozoan parasite *Trypanosoma cruzi* (Bruckner & Labarca, 2015). Chagas disease is spread in the American continent, mostly in the endemic areas of twenty one Latin American countries: Argentina, Belize, Bolivia, Chile, Colombia, Costa Rica, Ecuador, El Salvador, French Guiana, Guatemala, Guyana, Honduras, Mexico, Nicaragua,

Panama, Paraguay, Peru, Suriname, Uruguay and Venezuela. From 1990 to 2010, the estimated number of infected people has fallen from approximately 20 million to around 10 million. The risk of transmission has been decreased by introducing vector-control measures such as spraying insecticide to remove triatomine insects and house cleanliness and improvements, e.g. crack-free walls and bednets. Another strategy implemented to prevent the risk of transmission is a better control over blood transfusion safety in Latin America (WHO, 2020).

T. cruzi is transmitted in trypomastigote forms previously ingested by a reduviid insect (*Panstrongylus, Rhodnius, Triatoma*) when it fed on an infected organism. The trypomastigotes inside the reduviid insect transform into epimastigotes which multiply in the hindgut. The infectious stage starts when metacyclic trypomastigotes develop from the epimastigotes and passed to the faeces of the reduviid insect. Humans contract the disease when the reduviid insect defecates while taking a blood meal and metacyclic trypomastigotes in the faeces are rubbed or scratched into the bite wound or onto mucosal surfaces (Bruckner & Labarca, 2015; Stevens, 1999). In contrast to the direct form of transmission (inoculative infection) from *Leishmania* spp. and *Trypanosoma brucei*, discussed later on, *Trypanosoma cruzi* relies on an indirect route of transmission (contaminative infection). The contaminative infection, also called "stercorian" infection, is considered as the ancestor of salivarian or inoculative infection, with low probability of transmission of 5.8×10^{-4} per contact with infected insects (Nouvellet *et al.*, 2013). Surprisingly, even with a low probability of transmission from vector to human, Chagas disease is still considered as a major public health concern.

1.2.2.2 *Trypanosoma brucei*

Trypanosoma brucei is a protozoan parasite which causes Animal African trypanosomiasis (AAT) and Human African trypanosomiasis (HAT), also known as sleeping sickness. HAT has currently affected around 37,000 people according to the estimations made by Hotez *et al.* (2014) from the Global Burden of Disease Study 2010 (IHME, 2013). Also, HAT is the second most lethal parasitic disease in the world affecting human health after rabies (Figure 1.2) i. e., rabies kills 100% of untreated victims while HAT has a mortality of 98% in untreated patients (Hotez *et. al.*, 2014).

There are two distinct subspecies of *Trypanosoma brucei*: *T. brucei gambiense*, which is concentrated in West and Middle Africa; and *T. brucei rhodesiense*, concentrated in East Africa (Figure 1.4). The two subspecies are deadly if the sickness stays untreated and Uganda is the only nation in which both species causing HAT are endemic (WHO, 2020). According to a study made in 2020 by Franco *et al.* (2022) announced a decrease of 97% of *T. brucei gambiense* cases and 85% of *T. brucei rhodesiense* since 2000, with the amount of 565 and 98 cases, respectively. The vast majority of *T. brucei gambiense* cases are gathered in the Democratic Republic of Congo (70%), while over 91% of the *T. brucei rhodesiense* cases were reported in Malawi.

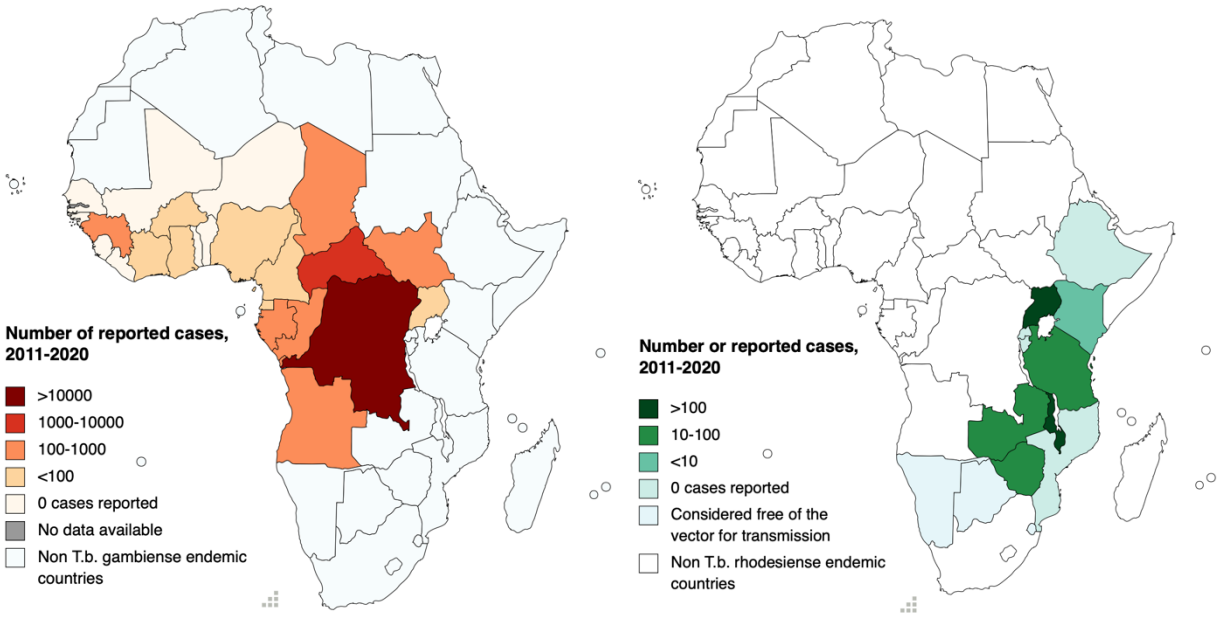


Figure 1.4 Distribution of *T. b. gambiense* (left) and *T. b.rhodesiense* (right) in the period of 2011-2020

Image depicted with the data provided by Franco et al*., 2022.*

The life cycle of *T. brucei* begins when a human or cattle become infected with the parasite in its infective form living in the salivary glands of the tsetse fly (*Glossina*), entering to the mammalian tissue when the fly bites. (Figure 1.5). After inflammation of the region surrounding the bite, the parasite enters the lymph nodes and after that to the circulatory system, where it starts to duplicate in the trypomastigote form (Hotez, 2013). Likewise, the parasite can be transmitted from a person or animal with parasitemia (presence of trypanosomes in the blood) to the tsetse fly, where it replicates in the insect gut and then relocates to the salivary glands.

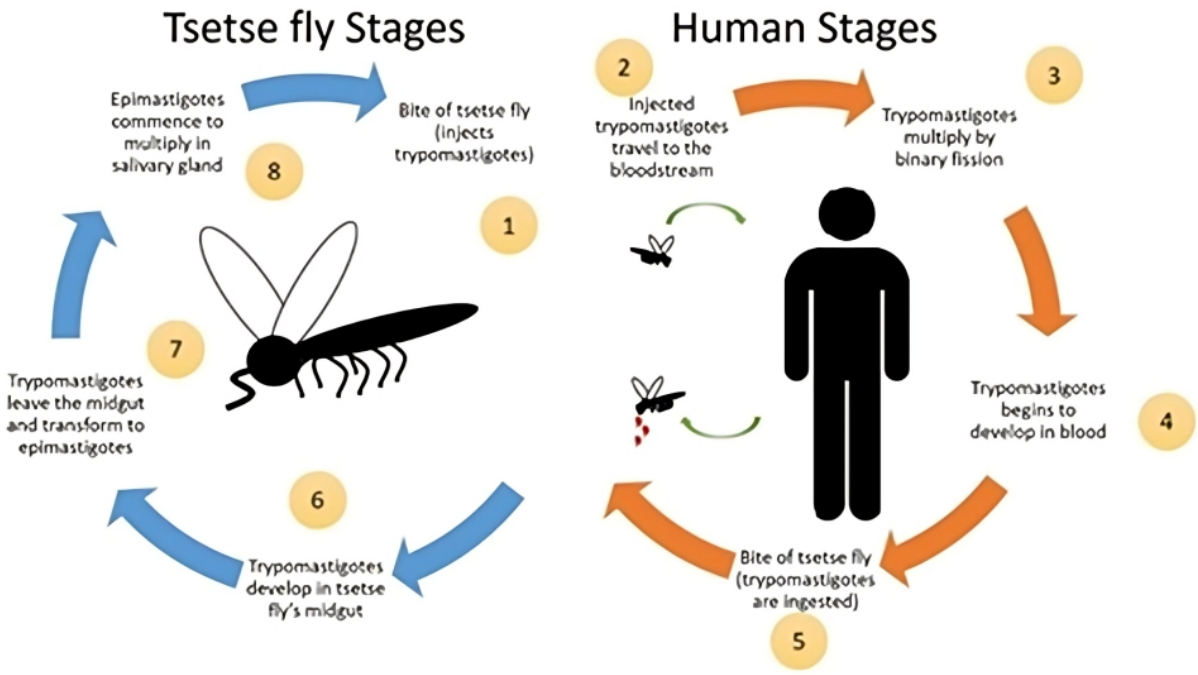


Figure 1.5 Life cycle of Human African Trypanosomiasis

When an infected tsetse fly bites (1), taking a blood meal, it injects metacyclic trypomastigotes into the skin tissue (2). The parasites enter the lymphatic system and pass into the bloodstream. Inside the host, they transform into bloodstream trypomastigotes and carried to other sites of the body while continuing the reproduction by binary fission (3 & 4). Also, a tsetse fly becomes infected with bloodstream trypomastigotes when taking a blood meal of an infected organism. In the fly's midgut, the parasites transform into procyclic trypomastigotes and multiply by binary fission (6). After that, the trypomastigotes leave the midgut and transform into epimastigotes (7). The epimastigotes reach the fly's salivary glands and continue multiplication by binary fission (8). (Image modified from Public Health Image Library, Centers for Disease Control and Prevention (CDC), 2020; https://www.cdc.gov/parasites/sleepingsickness/biology.html).

1.3 Cryptococcal infections

Despite not being formally recognised by the World Health Organisation (WHO) as a neglected tropical disease, it is listed in the Global Funding of Innovation for Neglected Diseases (G-FINDER) report since 2014 (FY report of 2013) and the Drugs for Neglected Diseases initiative (DNDi) since 2019 (Moran *et al.*, 2014; Drugs for Neglected Diseases initiative, 2020), affecting disproportionately people from low and middle-income countries.

Cryptococcus species are yeast-like fungal pathogens responsible for the disease known as cryptococcosis or cryptococcal meningitis. They belong to the phylum *Basidiomycota* and are differentiated from other pathogenic yeast by a polysaccharide capsule, formation of melanin, and urease and phospholipase B activity, which are considered virulence factors (Kwon-Chung *et al.*, 2014). The polysaccharide capsule of *Cryptococcus* species is formed mainly of two repeating constituents: glucuronoxylomannan (GXM) and glucuronoxylomannogalactan (GXMGal), which are key in the modulation of the immune response and resistance to phagocytosis (Almeida *et al.*, 2015). GXM and GXMGal have different immunomodulatory properties. GXM regulates the activity of monocytes/macrophages, neutrophils and dendritic cells, whereas GXMGal induces apoptosis of B-cell and T-cell and regulates the activity of macrophages (Vecchiarelli *et al.*, 2013). The formation of melanin has protective properties against environmental stress and protects the pathogen from phagocytosis, immune cells, oxidants, microbicidal peptides and antifungal agents (Kwon-Chung *et al.*, 2014). The production of the degradation enzymes urease and phospholipase B have different virulence functions such as intracellular survival of the pathogen, hydrolysis of host cell membranes, immunomodulation and promoting fungal dissemination from the lung to the brain (Kwon-Chung *et al.*, 2014).

The two species of Cryptococcus that are associated with human infections are *Cryptococcus neoformans* and *Cryptococcus gatti*. The organism is distributed throughout the world and is typically found in soil or in bird droppings (Sorrell & Ellis, 1997; Lazera *et al.*, 2000). *C. neoformans* infections are rare in people who have a healthy condition. Most cases of *C. neoformans* infections occur in patients who have been identified with an underlying immunocompromised condition, particularly those with HIV infection (Perfect 2010; Sorrell *et al.* 2011). Nevertheless, the proportional cases of *C. gatti* infections is significantly higher in immunocompetent individuals (Chen et al. 2000, 2008; Sorrell *et al.* 2011). The reason why this phenomenon occurs is unclear, but this might be explained because normal patients could have some unknown immunological defects that are not detected by routine tests (Kwon-Chung *et al.*, 2014) or *C. gatti* may exploit some aspect of the normal immune response to gain a foothold and evade host defences (Chen *et al.*, 2014).

Infection in *C. neoformans* usually initiates with inhalation of spores from the environment. Normally, the immune response is effective to contain and eliminate most inhaled cryptococci. However, in an immunocompromised host, cryptococcal cells can escape and proliferate in the lung before they disseminate to the blood and finally to the brain by crossing the blood – brain barrier (Eisenman *et al.*, 2007; Figure 1.6).

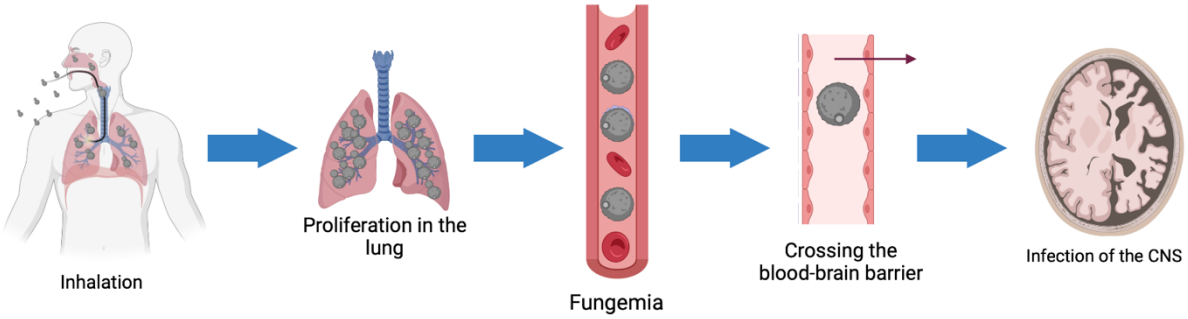


Figure 1.6 Infection route of cryptococcal meningitis

The host inhales airborne cryptococcal cells, which proliferate in the lung before hematogenously dissemination to the brain. (Image adapted from Kwon-Chung et al., *2014)*

Symptoms of *C. neoformans* infection in the lungs are often similar to many other illnesses, and can include: cough, shortness of breath, chest pain and fever. Cryptococcal meningitis is the advanced stage of infection when the spores have already spread from the lungs to the brain and the symptoms include: headache, fever, neck pain, nausea and vomiting, sensitivity to light and confusion or changes in behaviour (Chang *et al.*, 2006; Clark *et al.,* 1990; Bratton *et al.,* 2012). It is at this stage that cryptococcal meningitis can be lethal for the patient, especially if untreated or therapeutic intervention is delayed. Unfortunately, as the symptoms can be confused with infection with other pathogens, late diagnosis is not uncommon (Zhang *et al.*, 2019; Vechi *et al.*, 2019).

Since HIV infection is the most common predisposing factor for cryptococcal meningitis worldwide, most of cryptococcosis cases occur in regions where AIDS is prevalent (Park *et al.*, 2009; Park *et al.,* 2011). The region with the most cases registered therefore is eastern and southern Africa, where more than 20 million people live with AIDS and 16 million have access to treatment (UNAIDS, 2022) According to data obtained in 2014, the global incidence of cryptococcal meningitis is estimated at nearly

220,000 cases annually, with more than 180,000 deaths per year (Rajasinham *et al.*, 2017).

1.4 Current and novel treatments for kinetoplastid and cryptococcal infections

Since the development of the medications from the mid 1900s up to this days, the majority of the treatment alternatives for kinetoplastids are toxic and ill-adapted (Hotez *et al.*, 2016; Barrett *et al.*, 2007; Table 1.4). Because of the fears referenced previously, the interest in drug discovery and improvement has been increasing, searching for better and non-lethal choices (Table 1.5). Plus, numerous prospective leishmaniasis immunizations are in progress for both visceral and cutaneous forms and also for Chagas disease (Jennings & Urquhart, 1983; Jacobs *et al.*, 2011; Chatelain, 2015; Loiseau *et al.*, 2011; Ben Salah *et al.*, 2013).

Table 1.4 Chemical structure and known mechanism of action of anti-kinetoplastid and anticryptococcal treatments*

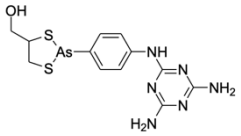
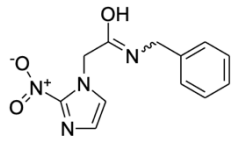
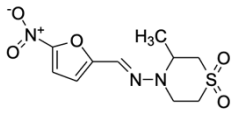
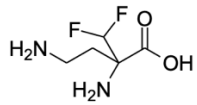
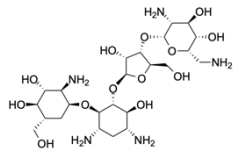
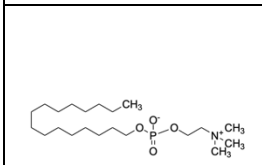
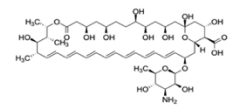
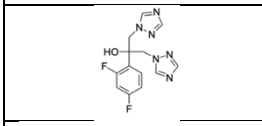
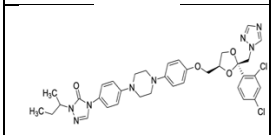
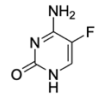
Current Treatment	Organism	Chemical structure	Mechanism of action
Melarsoprol: Trivalent arsenical compound.	*T. brucei rhodesiense*		Interacts with thiol groups of numerous proteins produced by trypanosomes. It has a high affinity to the active site of pyruvate kinase, interfering with the energy generation duplication of the parasite (Leder & Weller, 2015).
Benznidazole: 2-nitroimidazole derivative			Inhibits protein and RNA synthesis. Promotes phagocytosis process through cytokine discharge and synthesize of electrophilic metabolites (Murta *et al.*, 1999).
Nifurtimox: Synthetic nitrofuran.	*T. brucei gambiense*		Generates reactive oxygen species, i.e., superoxide, hydrogen peroxide, and hydroxyl radicals. These radicals damage the membrane and inactivate enzymes (Docampo & Moreno, 1984; Gutteridge, 1985).
Eflornithine			Inhibits ornithine decarboxylase required for the development of polyamines, significant for cell multiplication and differentiation in parasites (Kingsnorth, 1986).
Paromomycin	*Leishmania*		Inhibits the protein synthesis by binding to 16S rRNA, repressing the protein synthesis. This compound is a broad-spectrum aminoglycoside antimicrobial (Davidson *et al.*, 2009).
Miltefosine			Interferes with cell signaling pathways and acts on key enzymes involved in the metabolism of ether lipids present on the surfaces of the parasite (Jha *et al.*, 1999). Induce apoptotic cell death, being active against both extracellular promastigote and intracellular amastigote (Prasad *et al.*, 2004; Sindermann *et al.*, 2004).
Amphotericin B deoxycholate	*Cryptococcus neoformans*		Binds to ergosterol in the membrane cell of the fungus, perturbing its function and causing leakage of cellular content (Baginski *et al.*, 2005).
Liposomal amphotericin B			
Fluconazole			Inhibition of one of the key enzymes in the ergosterol biosynthesis, cytochrome P450-dependent lanosterol C14-alpha-demethylase (Sheng *et al.*, 2009).
Itraconazole			
Flucytosine			Competitive inhibition of purine and pyrimidine uptake through intracellular conversion to 5-fluorouracil. Hence, inhibits synthesis of DNA and RNA (Vermes *et al.*, 2000).

Table 1.5 Current and novel treatments developed for kinetoplastid and cryptococcal infections in humans*

	Human African trypanosomiasis (HAT)	Leishmaniasis	Cryptococcal meningitis
Current treatment	Pentamidine Suramin Melarsoprol Nifurtimox-eflornithine combination therapy (introduced in 2009)	Liposomal amphotericin B Paromomycin Miltefosine	Amphorecin B Flucytosine Fluconazole
Novel treatment	Fexinidazole (Phase IIb/III trials; Jennings & Urquhart, 1983; Torreele *et al.,* 2010) Oxaborole SCYX-7158 (Phase I; Jacobs *et al.,* 2011)	Sitamaquine (Loiseau *et al.,* 2011; Perez-Victoria *et al.,* 2011) Paromomycin and gentamicin topical cream (WR 279,396; Ben Salah *et al.,* 2013)	Itraconazole (Perfect *et al.*, 2010). Voriconazole (Boucher *et al.*, 2004; Theuretzbacher *et al.*, 2006) Tamoxifen (Hai *et al.*, 2019; Butts *et al.*, 2014) Miltefosine (Ravu *et al.*, 2013; Spadari *et al.*, 2018)

**(Hotez* et al.*, 2016; Bruckner & Labarca, 2015; WHO, 2018).*

However, under certain circumstances, mass drug administration (MDA) using the current treatments is neither possible or efficient to control or eradicate the disease. Those circumstances are related to high rates of drug failure or re-infection after treatment (Bethony *et al.*, 2011). At the same time, the cost of the treatments, mostly unaffordable for people from low-income communities, and their toxicity are further challenges that public health organizations have to deal with to deliver appropriate treatment to their patients. Because of this, there is an urgent need to develop new control tools, including vaccines, with the purpose to integrate them in the MDA or childhood vaccination programs (Bethony *et al.*, 2011; Hotez, 2020).

The development and improvement of vaccines targeting *Leishmania* spp. has been achieved through advancements made in uncovering *Leishmania* antigens that instigate a strong immune reaction against the parasite (Duarte *et al.*, 2016; Volpedo *et al.*, 2021). The most promising example is the development of the ChAd63-KH vaccine, a third-generation adenovirus-vectored vaccine that contains two *Leishmania* recombinant protein antigens: kinetoplastid membrane protein-11 (KMP-11) and hydrophilic acylated surface protein B (HASPB). KMP-11 is a membrane protein

expressed in the promastigote and amastigote stages of the parasite and is rich in CD8+ T cells epitopes. HASPB is a secreted protein that lacks a conventional signal peptide and is expressed in the infective stages of the parasite. This vaccine has proven to induce CD8+ and CD4+ T cell responses and antibodies with 7 patients showing more than 90% clinical improvement and 5 patients with partial improvement out of the 23 patients participating in a completed phase II clinical trial (NCT02894008; MacLean *et al.*, 2016; Younis *et al.*, 2021)

In addition, Sabin Vaccine Institute Product Development Partnership and Texas Children's Hospital Center for Vaccine Development in collaboration with an international consortium of academic and industrial partners in Mexico, Germany, Japan, and the USA have been working in the development of a vaccine targeting Chagas disease (Dumonteil *et al.,* 2012). The mechanism of action of this vaccine is through the stimulation of the CD8+ T cells and liberation of IFN-γ activated by the Th1 response. The vaccine contains two *T. cruzi* recombinant protein antigens, a *T. cruzi* 24 kDa trypomastigote excretory-secretory protein known as Tc24 and the *T. cruzi* trypomastigote surface transialidase known as TSA-1 (Dumonteil *et al.,* 2012). Tc24 is a 24 kDa protein of low-capacity, high-affinity calcium sensor in the flagellum while TSA-1 is a 85 kDa protein that plays an important role in the scavenging of sialic acid by the parasite (Biter *et al.*, 2018; De la Cruz *et al.*, 2019). The efficiency of this vaccine has been tested and confirmed the immune response in trials using mice, dogs and rhesus macaques as model organisms (Dumonteil *et al.*, 2004; Quijano-Hernandez *et al.*, 2008; Dumonteil *et al.*, 2020). However, clinical trials in humans have not been conducted yet due to the insufficient resources for the project and the lack of interest from the investors (Dumonteil & Herrera, 2021).

Furthermore, two passive vaccines targeting cryptococcosis have been developed from antibodies; for example Mycograb, a recombinant human antibody directed against fungal heat shock protein 90 (Hsp90, a chaperone required by fungal pathogens to maintain cellular homeostasis (Matthews *et al.*, 2003; Cordeiro *et al.*, 2016)). The efficacy and safety of this vaccine are under evaluation in ongoing phase II clinical trials (NCT00324025 and NCT00847678). And the vaccine 18B7, a monoclonal antibody targeting the capsular polysaccharide of *C. neoformans* (Casadevall *et al.*, 1998). This vaccine was evaluated in a phase I clinical trial being well tolerated and without evidence of toxicity. However, it is important to mention the

18B7 is a murine immunoglobulin G1 antibody, which would likely invoke a human anti-mouse antibody (HAMA). In the study, 10% of the patients developed HAMAs, which will occur more often in patients upon repeat injection. The decision to use a murine monoclonal antibody was based on the unknown efficacy and possible changes in the specificity that might occur during humanization of the mouse-human chimeric antibodies (Larsen *et al.*, 2005).

1.5 Resistance development in kinetoplastids and cryptococci

Antimicrobial resistance is a widely recognised and widespread problem not just in bacterial and viral infections but has also been observed in parasites and fungi.

As in the case of antibacterial resistance, eukaryotic pathogens such as *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania* spp., and *Cryptococcus neoformans*, have evolved mechanisms to escape the desired action of some drug therapies. There are complex biological interactions among parasites and their hosts which can affect the drug treatment. The response of the medication is not necessary reflected in the resistance of susceptibility of the parasite to the treatment. (Secor *et al.*, 2015).

Mechanisms for this kind of resistance are diverse relying upon the organism, acquiring some mechanism to escape the action of the drug. This might be by reduction in drug accumulation (efflux pump or receptor changes), or an alteration of the target protein structure or development of a metabolic pathway that reduces active drug concentration (Table 1.6; Secor *et al.*, 2015).

Table 1.6 Summary of the mechanisms of resistance of the kinetoplastid organisms to selected antiparasitic drugs

Disease	Drug(s)	Mechanism(s) of resistance
Leishmaniasis	Pentavalent antimonials	Decreased active intracellular medication fixation through diminished take-up (Decuypere *et al.*, 2012; do Monte-Neto *et al.*, 2011), increased efflux (Rai *et al.*, 2013; Mukherjee *et al.*, 2013), or diminished change to active trivalent form (Mukherjee *et al.*, 2007), glycans that increase host IL-10 generation (Mukhopadhyay *et al.*, 2011).
	Miltefosine	Resistance by augmented drug efflux or infectivity of the strain (Sánchez-Cañete *et al.*, 2009; Pérez-Victoria *et al.*, 2011).
	Amphotericin B	Expanded drug efflux, altered thiol metabolism (Purkait *et al.*, 2012).
African trypanosomiasis	Pentamidine (1st stage)	Mutation/loss of P2 adenosine as well as aquaglyceroporin 2 transporters that uptake drug (Stewart *et al.*, 2010; Baker *et al.*, 2012).
	Melarsoprol (2nd stage)	
	Eflornithine	Suitable just in *T. b. gambiense* because of *T. b. rhodesiense* resistance (Delespaux & de Koning, 2007). Deletion of the amino acid transporter gene *TbAAT6* confers resistance to this drug (Vincent *et al.*, 2010; Mathieu *et al.*, 2014).
	Eflornithine/ nifurtimox	
Chagas disease	Benznidazole	Mutation of nitroreductase type I NADH-dependent localized in the mitochondria (Wilkinson *et al.*, 2008).
	Nifurtimox	
Cryptococcosis	Fluconazole	The resistance mechanism to this drug can be classified as heritable or transient. Heritable fluconazole resistance occur due to a missense mutation in the ERG11 gene (Rodero *et al.*, 2003; Sionov *et al.*, 2012). Transient resistance happens in a subpopulation of cells capable to proliferate in the presence of high concentration of fluconazole (Xu *et al.*, 2001; Sionov *et al.*, 2009). The resistance of a specific population of fungal cells is mediated by aneuploidy of chromosomes 1, 4, 6 and 10 (Gerstein *et al.*, 2015; Sionov *et al.*, 2010).

1.6 Structure-based Drug Design (SBDD)

The process of a typical drug discovery from target identification to the FDA approval of the molecule takes up on average 10 years with the cost of 2.87 billion dollars according to an estimate in 2013 (DiMasi *et* al., 2016). The development of novel drugs with potential interaction on specific targets has become a matter of importance during the process of drug discovery. One conventional method of drug discovery is performed by physical high-throughput screening (HTS) of compound libraries, which can be time-consuming and expensive (Batool *et al.*, 2019; Cheng *et al.* 2012). For example, the drug discovery review performed by GlaxoSmithKline of industrial efforts to develop novel antibacterial drugs between the years 1995 and 2001 provides some useful insight. During this campaign, 70 HTS screening assays (of which 67 were target based and 3 on whole cells) using a range of proprietary libraries of up to 500,000 molecules were reviewed. They delivered only 5 leads for further development. The success rate of this campaign was four to five fold lower than other therapeutic areas at this time and highly expensive considering that each HTS assay costed around $1 million (Payne *et al.*, 2007).

Computational alternatives are relatively cheap, rapid and can sample much larger numbers and variety of compounds when compared to physical screening. They can be used to reduce large numbers of potential molecules down to a size that can be handled in the laboratory. Public databases such as Zinc20 (Irwin *et al.*, 2020) claim to provide over 200 million purchasable compounds in a format suitable for in silico screening (https://zinc20.docking.org). Such screens, if the docking programs can reliably predict inhibition are potentially a great help to counteract the limitations of the conventional method (Batool *et al.,* 2019). Using this approach, AstraZeneca has selected a subset of approximately 250,000 compounds from their entire screening collection (https://openinnovation.astrazeneca.com/preclinical-research/target-identification.html). Such a set has been screened to identify potential inhibitors of human FEN1 (McWhirter *et al.*, 2013).

Additionally, it is important to mention the advantage of computational alternatives compared to physical HTS in terms of results. The typical hit rate of a random screen is usually less than 1%, requiring compound libraries of hundreds of thousands to generate a minimum number of hits to continue onward with the next steps of drug

development (Dreiman *et al.*, 2021). However, using *in silico* approaches before continuing with physical HTS assays can give hit rates over 20% according to a case study searching for phosphodiesterase-5 (PDE5) inhibitors. Even higher hit rates (nearly 40%) were achieved in a case study seeking Melanin-Concentrating Hormone Receptor-1 (MCHR1) inhibitors (Szilágyi *et al.*, 2021).

One of the methods established for drug development is known as structure-based drug design (SBDD). SBDD is a more specific, efficient, and rapid process for lead discovery compared to random screening approaches because it deals first with the identification of promising target proteins through biological target validation. SBDD uses structural information on the target protein (X-ray crystallography, NMR or CryoEM) to aid inhibitor development, usually at an enzyme active site or receptor binding interface. It can also make use of knowledge of natural ligands (Swinney & Anthony, 2011).

1.6.1 Methodology of SBDD

The SBDD consists of an iterative approach that proceed through multiple cycles leading to a potential drug candidate for the clinical trials, accelerating the drug discovery process (Figure 1.7). One type of SBDD is known as virtual high throughput screening (vHTS). The first stage of vHTS is the cloning, purification and structure determination of the biological target protein. The structure determination step can be performed by X-ray crystallography, NMR or homology modelling. In the second stage a model of the desired binding site is constructed *in silico*. This describes the shape, charge distribution, hydrophobicity and possibly flexibility of the target site. Next, the computer is used to produce all possible conformations of each small molecule included in the library, as well as their charges and polarities. Then *in silico* docking of the small molecules is carried out and they are ranked based on their steric and electrostatic interaction with the target site. Usually, a selection of the highest ranking compounds is made, taking into account chemical diversity of the theoretical hits identified. Compounds then are tested in some sort of physical assay (*in vivo* or *in vitro*). Typically, the most active compounds are identified and a process called hit expansion is carried out. Here, the structures of the inhibitors are used to search chemical databases such as Zinc20 to provide a focused library of new potential hits,

which are once more subjected to physical screening. (Reddy *et al.,* 2007; Lavecchia & Di Giovanni, 2013; Lionta *et al.*, 2014). Some of the softwares tools used for structure-based vHTS are Autodock, Autodock Vina, GOLD, Glide and Rosetta Ligand, among others (Bajad *et al.*, 2021).

The previous two stages can be repeated several times which includes screening procedures, chemistry combination, and calculations of properties essential for drug development: such as absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics (Batool *et al.*, 2019). The final result of this process is an optimized compound which shows a significant improvement in binding and specificity for the target (Anderson, 2003).

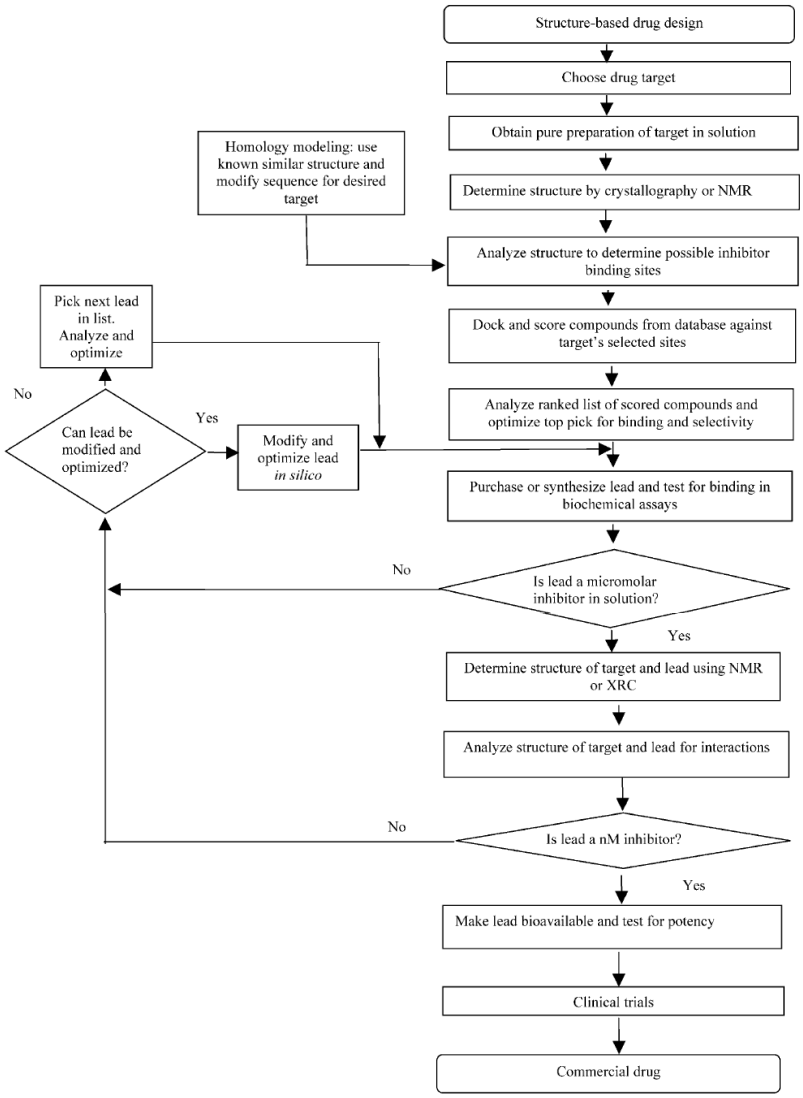


Figure 1.7 Iterative method of the Structure-based drug design

Image depicted from the study of Anderson (2003). License number 5367670031353.

Over the past twenty years, there has been an effort to apply computational analysis combined to the chemical and biological area with the purpose of speeding up drug discovery, design, development and optimization (Kapetanovic, 2008; Bergström & Lindmark, 2019). Computational methods play important roles in the identification of initial hits, designing improved leads from biochemically validated initial hits and in analysis of quantitative structure activity relationship (QSAR; Kwon *et al.*, 2019). The rational drug design method can be categorized in three types: ligand-based drug design (LBDD), structure-based virtual screening, and fragment-based drug design (FBDD).

Ligand-based drug design is founded on the extraction of essential chemical features from molecules known to interact with the target protein. X-ray crystallography and NMR structures of target ligand complexes are a common starting point for such projects. The medicinal chemistry approach here is to use the known interactions between ligand and protein to inform synthesis of new chemical entities (NCEs) which build upon features identified from relevant complexes (Aparoy *et al.*, 2012). A case of LBDD is the development of vasopressin 1a receptor antagonist, in which the optimization of the molecule was performed through an iterative method named scaffold hopping, shown in Figure 1.8 (Hu *et al.*, 2017; Ertl, 2014).

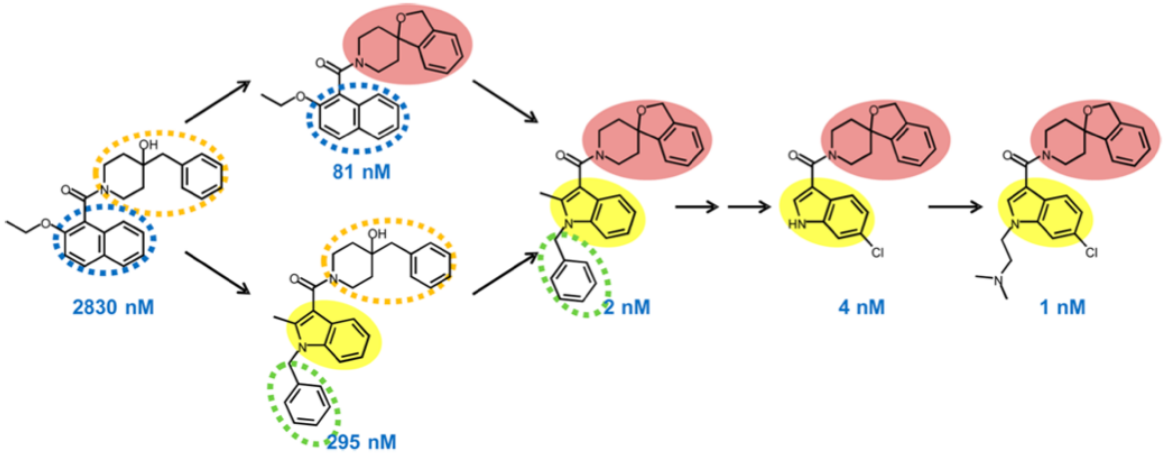


Figure 1.8 Vasopressin 1a receptor antagonists

Starting from a HTS hit with micromolar potency against human vasopressin 1a receptor, a series of highly potent antagonist were generated. For each antagonist, its Ki value is given. Dashed circles indicate substructures that were replace during optimization, and newly introduced fragments are highlighted in pink and yellow. Reprinted with permission from Hu et al.*, 2017. Copyright 2017 American Chemical Society.*

Several chemical characteristics such as molecular weight and lipophilicity are predictors of important properties of the designed drug (Lipinski, 2016). Also, the chemical similarity principle is applied in this method, which suggests that if two molecules share a similar structure, it will be likely that they will share similar biological properties (Yan *et al.,* 2016). The ligand-base drug design process follows these steps: 1. The target protein is used as a query for the chemical search, for example, search the Protein Data Bank for an enzyme X to identify similar or ideally identical protein structures which contain ligands. 2. The chemical structure of the ligand can then be used to search for similar molecules in databases such as Zinc20. 3. If available, these related molecules may be screened for activity or binding against the target protein. 4.The original ligand or any improved molecules emerging from the previous step may then be used to develop QSAR which can inform the design of molecules with improved activities (Lo *et al.,* 2016).

Fragment-based drug discovery (FBDD) starts with the physical screening of libraries of low-molecular weight compounds (so-called fragments) against the target molecule. (Erlanson *et al.,* 2016; Murray *et al.,* 2012). This physical screening may be carried out by a range of techniques but usually involves X-ray crystallography or NMR structure determination to identify molecules (hits) that bind directly to the protein of interest. For example, crystals of a protein may be soaked in mixtures of “fragments” typically at concentrations in the millimolar range. X-ray diffraction before and after the protein is exposed allows identification of any fragments that bind. It also provides information on where they bind. Fragment cocktails (mixtures) of compounds can be screened rapidly, e.g. 10 at a time, and if binding is detected, follow-up experiments are done with individual fragments to gain precise information on the binding sites and mode of fragment binding. Quantitative information on binding affinity using e.g. surface plasmon resonance, isothermal titration calorimetry or similar techniques are then used to quantify the interaction between fragment and target (Jarmoskaite *et al.*, 2020).

Usually, these first hits have a very weak affinity for the target protein. However, if multiple hits with different nearby binding sites are identified, fragment growing and fragment linking approaches can be adopted. Fragment growing is the process of adding functional groups or substituents to the core structure of the compound to increase the interaction with the binding site residues. Fragment linking is based on linking through covalent bonds two or more fragments with suitable linkers (Kumar *et*

al., 2012). In either case, this is accomplished by synthesis of new molecules. These new molecules typically have much lower dissociation constants (K_d) than the original fragments (Figure 1.9).

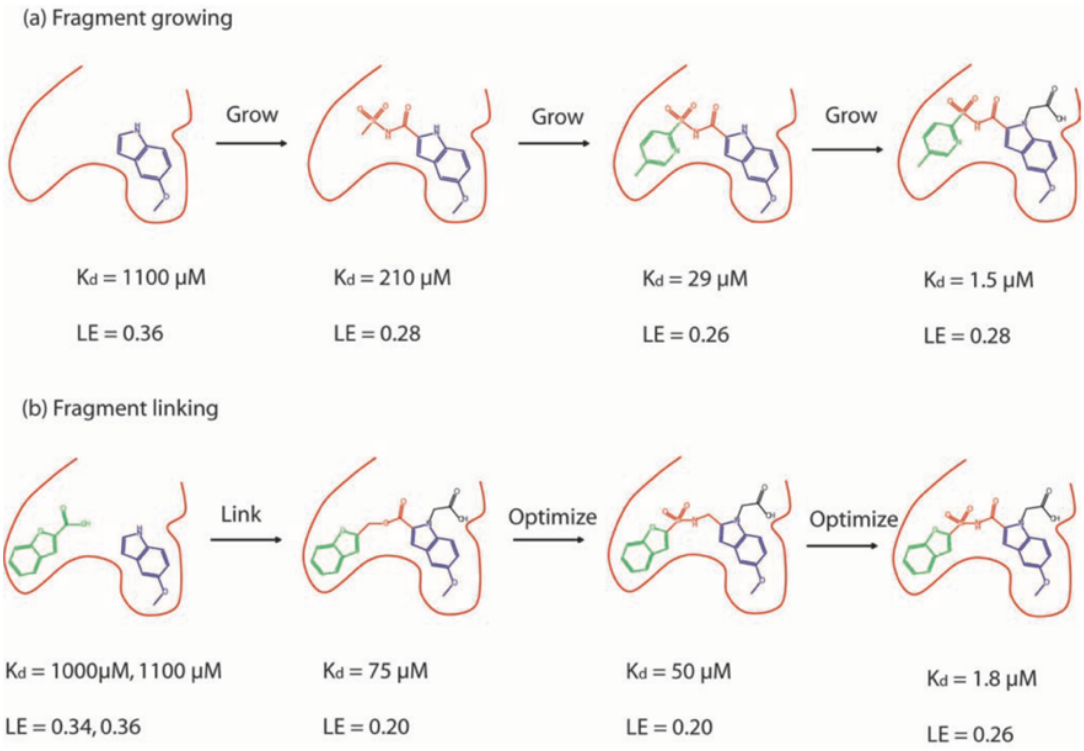


Figure 1.9 Schematic representation of fragment optimization through FBDD

(A) *Fragment growing: Initial fragment with low affinity is optimized by addition of functional groups to obtain a compound with higher affinity.* ***(B)*** *Fragment linking: Two or more fragments bound independently close to each other and are covalently linked to obtain a compound with higher affinity. Kd represents the constant of dissociation while LE represents the per atom ligand efficiency, defined by the free energy of binding divided by the number of heavy (non-hydrogen) atoms. Reprinted with permission from Kumar* et al*., 2012. Copyright 1994 Bentham Science Publishers Ltd.*

1.6.2 SBDD for drug development in NTDs

The identification of new macromolecular targets and small-molecule modulators has become a mission of high priority for the development of novel and effective antimicrobial agents. Therefore, some of the recent discoveries based on SBDD approaches for trypanosomiasis, leishmaniasis and cryptococcosis are presented in Table 1.7.

Table 1.7 Recent discoveries of potential drug targets for the development of novel therapies in Leishmaniasis, Chagas disease, African sleeping sickness and Cryptococcal meningitis

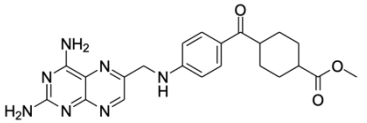
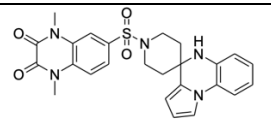
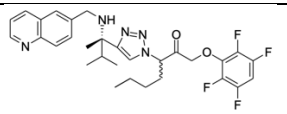
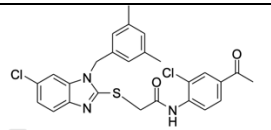
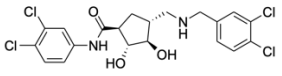
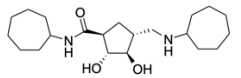
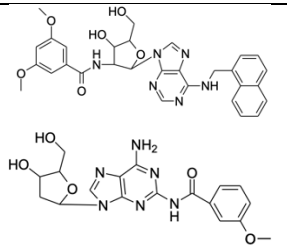
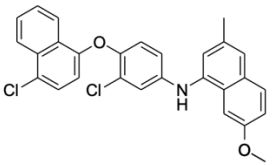
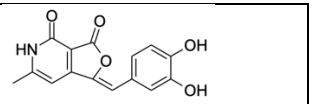
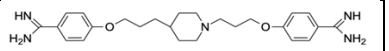
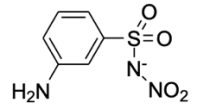
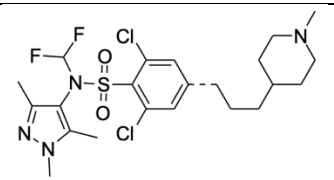
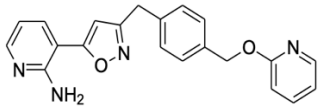
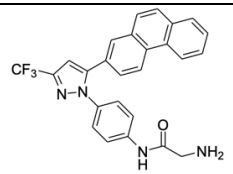
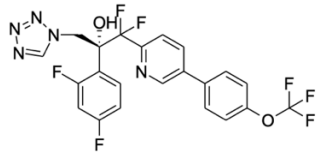
Enzyme	Function	Organisms	Molecule Designed
Dihydrofolate Reductase (DHFR)[1]	Both enzymes catalyse sequential reactions in the biosynthesis of dTMP. DHFR catalyzes the dihydrofolate by NADPH to regenerate the tetrahydrofolate. TS catalyzes the conversion of dUMP and 5,10-methylenetetrahydrofolate to dTMP and dihydrofolate (Ivanetich & Santi, 1990).	*L. major*	
Thymidylate Synthase (TS)[1]			
Pteridine Reductase (PTR1)[2]	Reduces folate and biopterin. In the presence of enzymatic inhibitors of DHFR and TS, PTR1 is overexpressed in trypanosomatids (Nare *et al.,* 1997; Bello *et al.,* 1994).	*L. panamensis*	
Cruzain (GP57/51)[3]	Both enzymes form part of the cysteine protease family, implicated in several biological processes such as virulence factor (Del Nery *et al.,* 1997; Mottram *et al.,* 1996; Casgrain *et al.,* 2016; Berasain *et al.,* 2003), evasion of immune system (Cameron *et al.,* 2004; Aparicio *et al.,* 2004), parasite differentiation and metabolism (Mahmoudzadeh-Niknam & McKerrow, 2004; Bonaldo *et al.*, 1991) and invasion of host cells (De Souza Leao *et al.,* 1995; Barral *et al,* 1993; Doyle *et al.,* 2011, Scharfstein *et al.*, 2000).	*T. cruzi*	
Cysteine protease type 2 (CPB2)[4]		*L. infantum*	
Phosphofructokinase (PFK)[5,6]	Enzymes implicated in the glycolytic pathway of trypanosomatids, which take place in specialized organelles called glycosomes (Opperdoes & Borst, 1977). During the bloodstream form of trypanosomatids, the tricarboxylic acid cycle is not functional, so the ATP production is highly dependent on glycolysis pathway (Bakker *et al.,* 2000). Without the glycosomes, the glycolysis process can be toxic to the cells due to dysfunctional regulation of the intermediates and dysfunctional production of ATP, by increasing the production of ATP or triggering an autocatalytic pathway due to the low concentration of ATP, unable to start the glycolytic route (Heinrich *et al.,* 1997; Bakker *et al.,* 2000).	*T. brucei* & *L. mexicana*	
Pyruvate Kinase (PyK)[5,6]			
Glyceraldehyde-3-phosphate dehydrogenase[7,8]		*T. brucei, T. cruzi* & *L. mexicana*	
Type 2 NADH dehydrogenase (NDH-2)[9]	Present in various microorganisms and plants, and a possible variation in humans (Marreiros *et al.,* 2016). This enzyme is directly responsible for the reduction of oxygen participating in the electron transfer process from soluble NADH via flavin adenine dinucleotide (FAD) to membrane-bound ubiquinone, essential for the respiration cycle (Blaza *et al.,* 2017).	*L. infantum*	

Table 1.7 Recent discoveries of potential drug targets for the development of novel therapies in Leishmaniasis, Chagas disease, African sleeping sickness and Cryptococcal meningitis

Topoisomerase 1[10,11,12,13]	Removes DNA supercoils formed during transcription and replication for the strand breakage during recombination and unknot and decatenate of chromosomes (Das *et al.,* 2008; Champoux, 2001).	*L. donovani*	
		C. neoformans	
Carbonic Anhydrase[14]	Metalloenzyme which catalyze the hydration of CO_2 to bicarbonate and proton, an important reaction in the metabolic pathways of pathogens such as gluconeogenesis, ureagenesis and biosynthesis of fatty acids, amino acids and nucleotides (Pan *et al.,* 2013; Syrjänen *et al.*, 2013; Capasso & Supuran, 2015).	*T. cruzi, L. amazonensis* & *L. infantum*	
N-Myristoyltransferase[15]	N-Myristoyltransferase (NMT) catalyses the transfer of myristate, a fatty acid formed by 14 carbons with saturated bonds, from CoA to proteins through the N-terminal glycine. This reaction facilitates the association of the proteins with membranes locations, mediating the interaction between proteins and stabilizing of protein structure (Brannigan *et al.*, 2010; Roberts *et al*. 2014).	*T. brucei*	
Inositol acylase (Gwt1 enzyme)[16,17,18]	Essential for the acetylation of the inositol ring of phosphatidylinositol during the biosynthesis of glycosylphosphatidylinositol (GPI)-anchored cell wall mannoproteins, required for the adhesion of pathogenic fungi to human epithelium (Umemura *et al.*, 2003; Tsukahara *et al.,* 2003).	*C. neoformans*	
Acetyl-CoA synthetase[19]	Enzyme required to generate acetyl-CoA from acetate and coenzyme A in an ATP-dependant reaction (Starai & Escalante-Semerena, 2004). Represents an important carbon source for production of acetyl-CoA (Strijbis & Distel, 2010). The activity of this enzyme is required for virulence in *C. neoformans* (Hu *et al.*, 2008).	*C. neoformans*	
Lanosterol 14-alpha demethylase (CYP51)[20,21,22,23]	CYP51 (Erg11) belongs to the group of cytochrome P450 monooxygenase. Catalyzes C14-demethylation of lanosterol and it is Involved in the synthesis of ergosterol (Aoyama & Yoshida, 1978). Ergosterol is considered an essential component of fungal cell membranes, determinating conformation, permeability and activity of membrane proteins (Jordá & Puig, 2020).	*C. neoformans*	

[1]Cavazzuti *et al.*, 2008; [2]Ochoa *et al.*, 2016; [3]Brak *et al.*, 2008; [4]De Luca *et al.*, 2018; [5]Nowicki *et al.*, 2008; [6]Claustre *et al.*, 2002; [7]Callens & Hannaert, 1995; [8]Bressi *et al.*, 2001; [9]Stevanović *et al.*, 2018; [10]Mamidala *et al.*, 2016; [11]Shibata *et al.*, 2012; [12]Mitsuyama *et al.*, 2008; [13]Nishikawa *et al.*, 2017; [14]Bonardi *et al.*, 2019; [15]Brand *et al.*, 2014; [16]Zhao *et al.*, 2019; [17]Pfaller *et al.*, 2019; [18]Shaw *et al.*, 2018; [19]Koselny *et al.*, 2016; [20]Lockhart *et al.*, 2016; [21]Nielsen *et al.*, 2017; [22]Wiederhold *et al.*, 2018; [23]Garvey *et al.*, 2018.

Although, target-based drug design against leishmaniasis, Chagas disease, African trypanosomiasis and cryptococcal meningitis is extensively reported (Ferreira & Andricopulo, 2018; Uzcanga *et al.,* 2017; Calderano *et al.,* 2011; Araujo-Zuma *et al.,* 2014; Spadari *et al.*, 2020), research on flap endonucleases as potential drug targets in this context has not been reported to our knowledge. Of the enzymes reviewed on these papers, none of them are directly involved in DNA replication.

1.7 Flap endonucleases

Flap endonucleases (FENs) are a class of enzymes identified in all organisms involved in the replication and maintenance of the genome (Figure 1.10). In the eukaryotic organisms FENs are characterised by the presence of three conserved domains; the N-terminal domain, an intermediate domain and the C-terminal domain (Harrington & Lieber, 1994; Lieber, 1997). FEN enzymes belongs to the family of metallonucleases as they require the presence of divalent metal cations to process the substrate and possess the ability to act as a 5'-3' exonuclease and as a structure-specific endonuclease (Allen *et al.,* 2009; Harrington & Lieber, 1994).

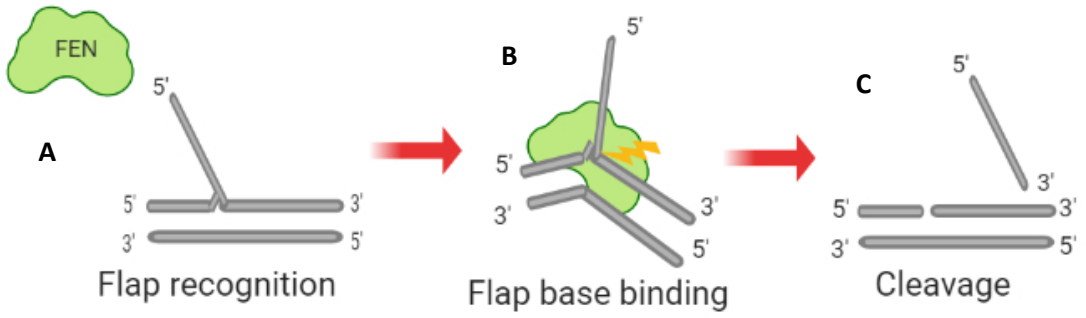


Figure 1.10 Simple model representation of flap endonuclease activity

(A) *Recognizes the displaced flap.* ***(B)*** *FEN binds to the base of the flap and displaced it by bending the substrate into a 100° angle to arrange the one-nucleotide 3' overhang.* ***(C)*** *Generation of a nick after the FEN cleaves and displaced the flap in the active site. Image created and modified according to Balakrishnan & Bambara, 2013.*

The term “flap endonuclease” was introduced by Harrington and Lieber (1994) in a paper describing cutting of branched substrates or flaps by a mouse enzyme which they called Flap Endonuclease (FEN-1). However, the phenomenon of flap endonuclease activity have been reported before this. For example, in studies carried out by Lundquist and Olivera (1982) it was demonstrated that the enzyme synthesizes new DNA at the 3' end of a nick and generates a flap with a 5' end in a process called nick translation. Clearly, this describes FEN-like activity. Later work on *Thermus aquaticus* DNA PolI (Lyamichev *et al.*, 1993), which contains a FEN domain and is highly homologous to the *E. coli* enzyme, provided similar results. They also suggested that the flap endonuclease enzyme operates by a threading mechanism, whereby the 5' end of the displaced strand passes through a domain of DNA PolI that displays 5' nuclease activity now better known as flap endonuclease activity. Earlier studies performed by Setlow and Kornberg (1972), found that the *E. coli* DNA PolI contains two distinctive enzymatic activities. In the C-terminal of the enzyme, it was demonstrated a polymerase and 3'-5' exonuclease activity, though the N-terminal domain was initially shown to possess a 5'-3' exonuclease activity. During proteolysis, DNA PolI is cleaved into two fragments. The large, or Klenow fragment, retains polymerase and 3'-5' exonuclease activity while the small fragment contains only 5'-3' exonuclease activity (Klenow & Henningsen, 1970; Klenow & Overgaard-Hansen, 1970). The 5'-3' exonuclease domain of the PolI enzyme shown the mechanism that was associated with the synthesis and cleavage of short flap chains, classifying the enzyme as a metallonuclease of the FEN family (Harrington & Lieber 1994).

1.7.1 Principal functions of Flap endonuclease

FENs are recognized a central components for DNA metabolism. They play a central role in DNA replication during the processing of Okazaki fragments. These are segments of DNA on the lagging strand that have been synthesized discontinuously as the replication fork moves in the opposite way to the polymerase, presenting both endo and exonuclease activity. Also, FENs have a role in the DNA repair process through the long-patch base excision repair, telomere maintenance and in the stalled replication forks rescue (Balakrishnan & Bambara, 2013).

1.7.1.1 DNA Replication

A precise replication and repair of DNA is essential for cell viability. The replication of DNA begins on a specific site of the genome acting as the origin and from this point the replication proceeds in two directions. The strands of the DNA helix have an antiparallel structure, which means replication can only occur from the direction of 5' to 3' (Alberts *et al.,* 1989). One strand is copied in the direction of the fork opening, which will not cause a problem because replication can be done continuously from a single RNA/DNA primer (5'-3'). To initiate the replication process, a primer is synthesized on this strand by the primase (a hetero-tetramer complex consisting of RNA polymerase and DNA polymerase α (Pol α)). After synthesis of the primer, Pol α is displaced by the combined action of replication factor C (RFC), proliferating cell nuclear antigen (PCNA) and polymerase ε (Pol ε) to synthesize what will be the leading strand in the same direction as the fork is opening (Lovett, 2007; Burgers & Kunkel 2017; Burgers, 2009; Pursell *et al.,* 2007; Nick McElhinny *et al.*, 2008).

In the case of the other strand (the lagging strand), continuous synthesizes is not possible. This is because the lagging strand is copied in the 5'-3' direction but the fork is moving in the opposite direction relative to this process. For this case, replication is again initiated by the primase. discontinuous segments called Okazaki fragments are produced from the primer by the displacement of Pol α through the action of RFC/PCNA with polymerase δ (Pol δ). Eventually, the newly extended lagging-strand copy made by Pol δ collides with the 5' end of an earlier primer/Okazaki fragment. A flap structure is generated by Pol δ. Then, FEN1 is recruited by PCNA to cut the displaced flap RNA/DNA primer yielding a nick. Finally, PCNA recruits DNA ligase I to seal the nick between the DNA fragments, forming the lagging strand (Waga & Stillman, 1994; Bambara *et al.*, 1997; Burgers, 2009; Figure 1.11).

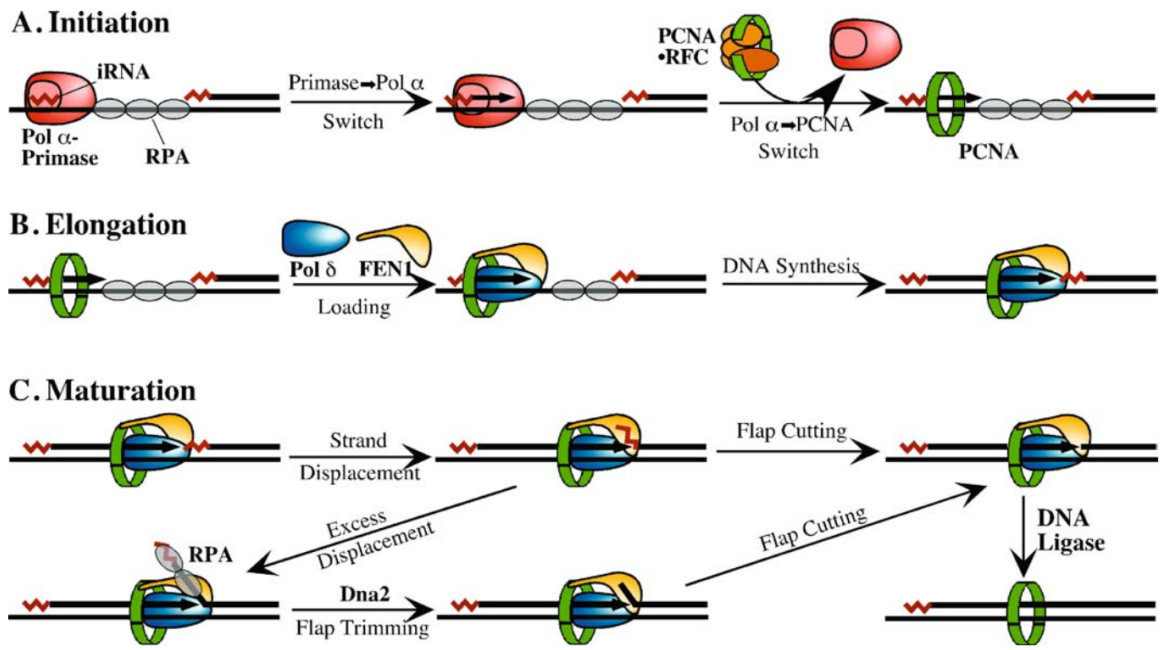


Figure 1.11. Role of FEN during the DNA replication stage of the lagging strand

(A) *Switch of the hetero-tetramer primase to PCNA switch promotes loading of Pol ε on the leading strand not shown), and Pol δ on the lagging strand.* ***(B)*** *Recruitment of Pol δ and FEN1 by PCNA for the elongation of the DNA strand made by Pol δ.* ***(C)*** *Activity of FEN1 to cut the displaced RNA/DNA primer and recruitment of DNA ligase I by PCNA to seal the spaces between the strands. RPA binds to long flaps only to prevent cleavage by FEN1 and stimulating cleavage by Dna2. Figure reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group: Critical Reviews in Biochemistry and Molecular Biology (Garg & Burgers, 2008; Reference code; ibmg/03106829).*

1.7.1.2 Stalled Replication Forks Rescue

The replication fork is controlled by the DNA helicase, which unwinds the double-stranded duplex during replication. During DNA replication, there are cases in which the replication forks can “stall” due to different conditions: alkylation on the DNA template, inhibition of the production of nucleotides or inhibition of polymerases but the helicase continues to unwind the DNA duplex (Katou *et al.,* 2003; Tercero & Diffley, 2001; Pacek *et al.,* 2006; Walter & Newport, 2000). When this happens, activation of checkpoint kinases is made by eukaryotic cells in response to the stalling replication fork (Branzei & Foiani, 2005; Li & Zou, 2005; McGowan & Russell, 2004). These checkpoint kinases are essential in such situations because they prevent an

irreversible collapse of the stalled forks that can be lethal for the eukaryotic cell (Lopes *et al.*, 2001; Tercero & Diffley, 2001).

According to the studies of Zheng *et al.* (2005) and Sharma *et al.* (2004), FEN1 is implicated in the rescue of stalled replication forks. FEN1 forms a complex with WRN (Werner Syndrome Protein), a helicase from the RecQ subfamily, in response to DNA damage that stall replication forks. FEN1 cleaves the single-stranded DNA through its activity of gap endonuclease, which is modulated by the presence of phosphorylated RPA, allowing to convert stalled replication forks into recombinant substrates (Zheng *et al.,* 2005).

1.7.1.3 DNA Repair

DNA lesions can block genome replication and transcription, and if they are not repaired or are repaired incorrectly, they can lead to mutations that can be lethal to the cell (Jackson & Bartek, 2009). Some of these DNA lesions can be caused due to physiological processes such as DNA mismatch, hydrolytic reactions, non-enzymatic methylations, interaction with reactive-oxygen compounds, among others. DNA lesions also can be produced because of external factors, for example, ultraviolet light, radioactive compounds and tobacco (Jackson & Bartek, 2009). To respond to the damage inflicted on the DNA, cells have developed a large number of DNA repair pathways including mismatch repair, base excision repair, nucleotide excision repair, and homologous recombination.

One of the major mechanisms involved in repair of damaged bases is the base excision repair pathway (Rothwell & Hickson, 1997). Lesions that need to be repaired by this mechanism are bases with a relatively small modification or apurinic/apyrimidinic (AP) sites (location in DNA which has neither a purine nor a pyrimidine base). According to a model presented by Lindahl (1993), an altered base is removed by DNA glycosylase, which can recognise and remove the damaged base by cleaving the N-glycosidic bond (Robertson *et al.,* 2009), leaving an AP site. The next step in this mechanism is the cleavage of the AP endonuclease (APE1) to the 5' of the AP site to produce a single nucleotide gap in the DNA. Then, the gap is filled by a DNA polymerase and finally,

the repair is completed by the action of a DNA ligase (Robertson *et al.,* 2009; Figure 1.12).

Studies have indicated that AP site repair in eukaryotes may proceed by two alternative pathways: a DNA polymerase β (Pol β)-dependent pathway (short-patch repair that removes only one nucleotide) and a proliferating cell nuclear antigen (PCNA)-dependent pathway (long-patch repair that removes from 2 to 13 nt; Matsumoto *et al.,* 1994, Frosina *et al.,* 1996; Huggins *et al.*, 2002). The Pol β -dependent pathway requires three proteins: APE1, Pol β and DNA ligase III. On the other hand, the PCNA-dependent pathway requires of more proteins during DNA repair: APE1, PCNA, Pol δ, Pol ε, FEN1 and DNA ligase I (Kim *et al.,* 1998: Robertson *et al.,* 2009).

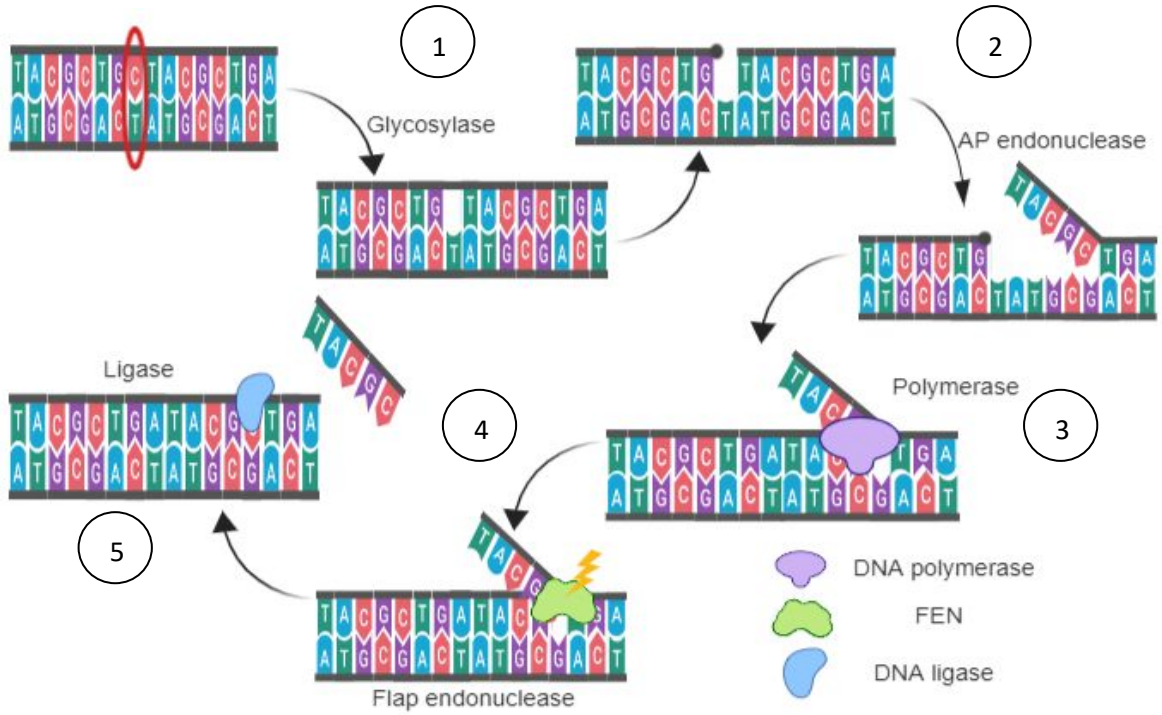


Figure 1.12 Flap endonuclease activity in long patch base excision repair

(1) An altered base is removed by DNA glycosylase, leaving an AP site. (2) Cleavage of the AP endonuclease (APE1) to the 5′ of the AP site to produce a single nucleotide gap in the DNA. (3) Gap is filled by a DNA polymerase. (4) Cut and displacement of the flap by FEN1. (5) Repair completed by the action of a DNA ligase (Robertson et al.*, 2009).*

If the cellular DNA repair pathways fail, it is likely that the organism will develop serious defects induced by lethal events happening on a cellular level. According to a study conducted by Shibata & Nakamura (2002), expression of a dominant nuclease-defective hFEN1 in a human cell line increases their sensitivity to DNA damage, causing a prolonged delay of S phase progression and impairment in colony-forming activity of cells. Also, FEN1 disfunction during DNA replication can cause DNA triplet repeat expansion promoted by alternative intraflap base pairing to form structures not accessible to FEN1, which leads to diseases like myotonic dystrophy, Huntington's disease, several ataxias, and fragile X syndrome (Gordenin *et al.,* 1997; Spiro *et al.*, 1999; Henricksen *et al.,* 2002).

1.7.1.4 Telomere Maintenance

The ends of each chromosome have a protective protein-DNA complex cap, known as telomere (Muftuoglu *et al.,* 2006). Telomeres protect the chromosome ends from being recognised as DNA damage and allow complete replication of the chromosome (Verdun & Karlseder, 2006). Consequences of telomere dysfunction include genomic instability that can contribute to neoplastic transformation and progression (Muftuoglu *et al.,* 2006).

Several studies have discovered that FEN1 plays an important role in the maintenance of telomere's integrity. During the phases S and G2 of the cell cycle, FEN1 interacts with Pol β and the telomeric-repeat binding factor 2 (TRF2) to stimulate the DNA synthesis on telomeric and nontelomeric primer/template substrates, repairing the damage on the telomeres caused by oxidative damage (Verdun & Karlseder, 2006; Muftuoglu *et al.,* 2006). Also, studies made by Saharia and collaborators (2008) revealed that the activity of FEN1 and its interaction with WRN and TRF2 are essential for it to function at telomeres. Mutations affecting the activity of FEN1 and the helicases WRN and BLM (Bloom syndrome protein) binding sites promoted telomere instability, suggesting that the interaction of FEN1 with the helicases WRN and BLM are essential for telomere maintenance due to binding of both WRN and BLM to the extreme C-terminal 18 amino acid tail of FEN1, which is next to the PCNA binding site of FEN1 (Sharma *et al.,* 2005).

1.7.2 Structure

Crystal structures of the FEN enzymes can differ depending on the organism, yet all of them share a similar core architecture (Finger *et al.*, 2012). For instance, the structure of the bacteriophage T5 5'-3' exonuclease, which was elaborated by Ceska and co-workers (1996), is comprised of a central β-sheet with 6 strands with additional α-helical features. The active site of this enzyme displays a triangular structure at the back of the nuclease and contains three α-helices that define a “helical arch” above a metal-binding domain. The divalent cations required for activation of the enzyme are bound by aspartic and glutamic acid residues (Asp26, Glu128, Asp130, Asp153, Asp155, Asp201 an Asp204; Figure 1.13). From the three α-helices involved in DNA binding, one of the helices (Helix 4) has positively charged residues (Lys83, Arg86, Lys89, and Arg93), which are directed towards the interior of the arch, a conserved tyrosine (Tyr82) located at the base of the arch, and the hydrophilic residues facing the exterior of the arch (Asp87 and Glu88). Helix 5 (Phe104 and Phe105) and Helix 1 (Phe32) are the two other helices which are involved in DNA binding (Figure 1.14).

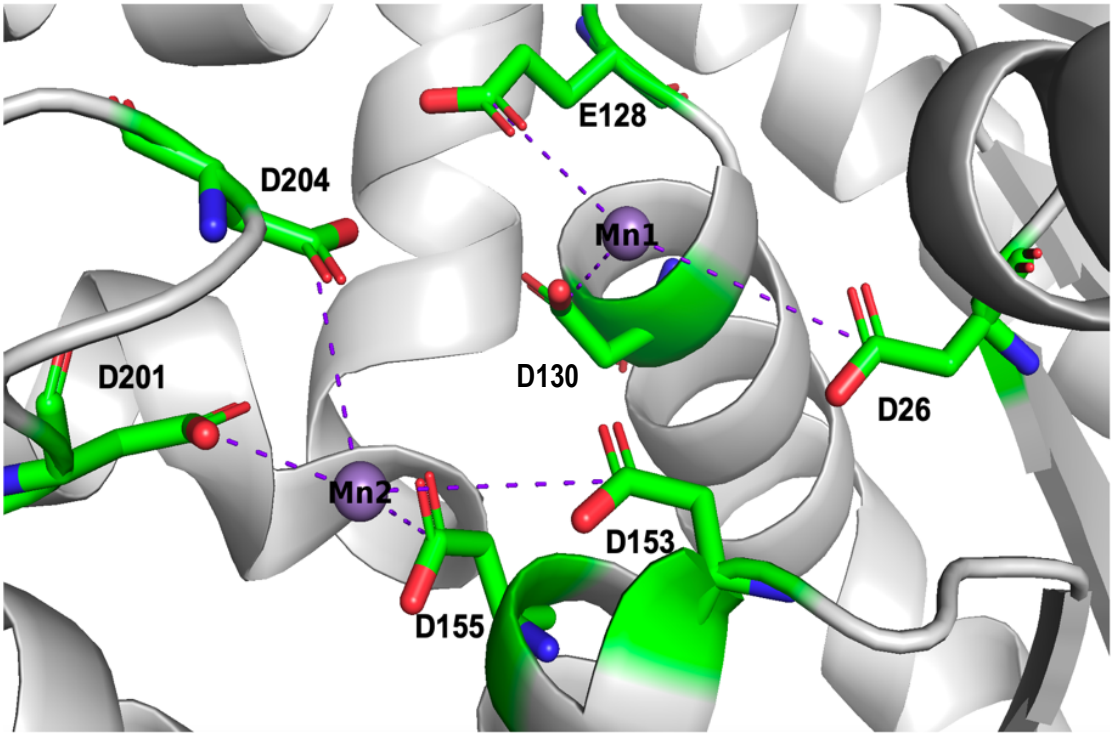


Figure 1.13 Active site of T5 5' exonuclease

The active site includes seven acidic amino acids, shown as sticks, that are involved in the coordination of two divalent metal ions of Mn^{2+} (purple dots). Image was rendered using structure 1UT5.pdb in PyMOL.

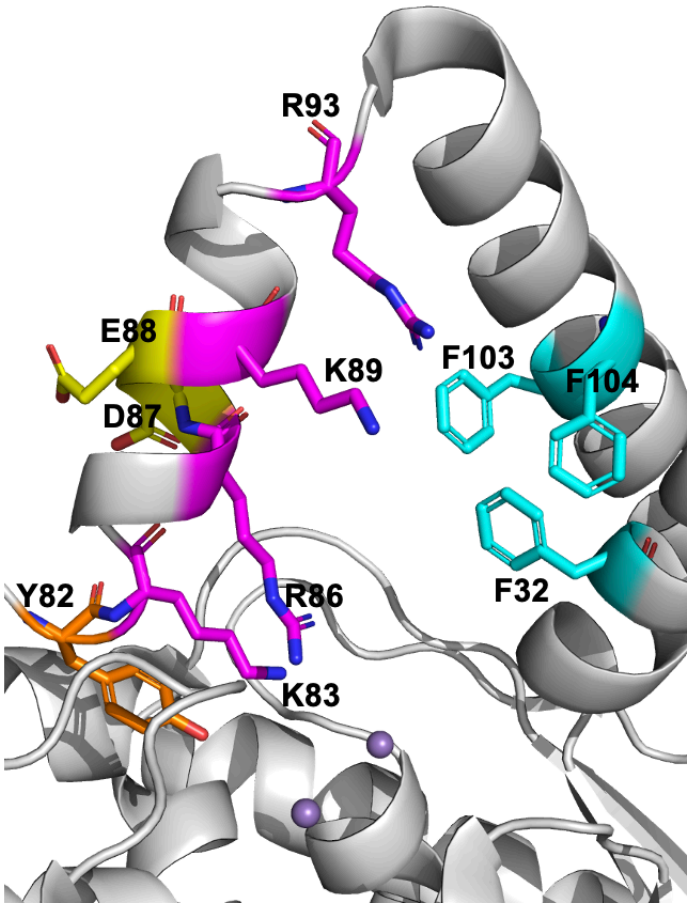


Figure 1.14: Residues involved in DNA binding and coordination

Shown as sticks: From Helix 4 at the base of the arch Tyr82 (orange), positively charged residues Lys83, Arg86, Lys89 and Arg93 (magenta) and hydrophobic residues Asp87 and Glu88 (yellow). From Helix 5 Phe104 and Phe105 and from Helix 1 Phe32 (cyan). Image was rendered using structure 1UT5.pdb in PyMOL.

The structure of the human FEN1 (HsFEN) enzyme, elucidated by Tsutakawa *et al.* (2011), was characterized by seven components: the helix-two turn-helix (H2TH; α-helices 10 and 11), a β pin (β-6 and 7 loop), an acid block (residues 56-59), a hydrophobic wedge (α helix 2 and α2 & 3 loop), a helical gateway (α-helix 4), an active site and cap (α-helices 4 and 5). For the active site, the presence of seven amino acids is required for metal ion coordination, which are Asp34, Asp86, Glu158, Glu160, Asp179, Asp181 and Asp233 (Figure 1.14). These can be part into two particular sites: metal ion binding site 1 and 2. Site 1 contains Asp86 and Glu158, while site 2 is comprises of Asp179, Asp181, and Glu160. The remaining amino acids, Asp34, Asp233, and Glu158, interact with both metal ions via water-mediated interactions (Figure 1.15; Tsutakawa *et al.*, 2011).

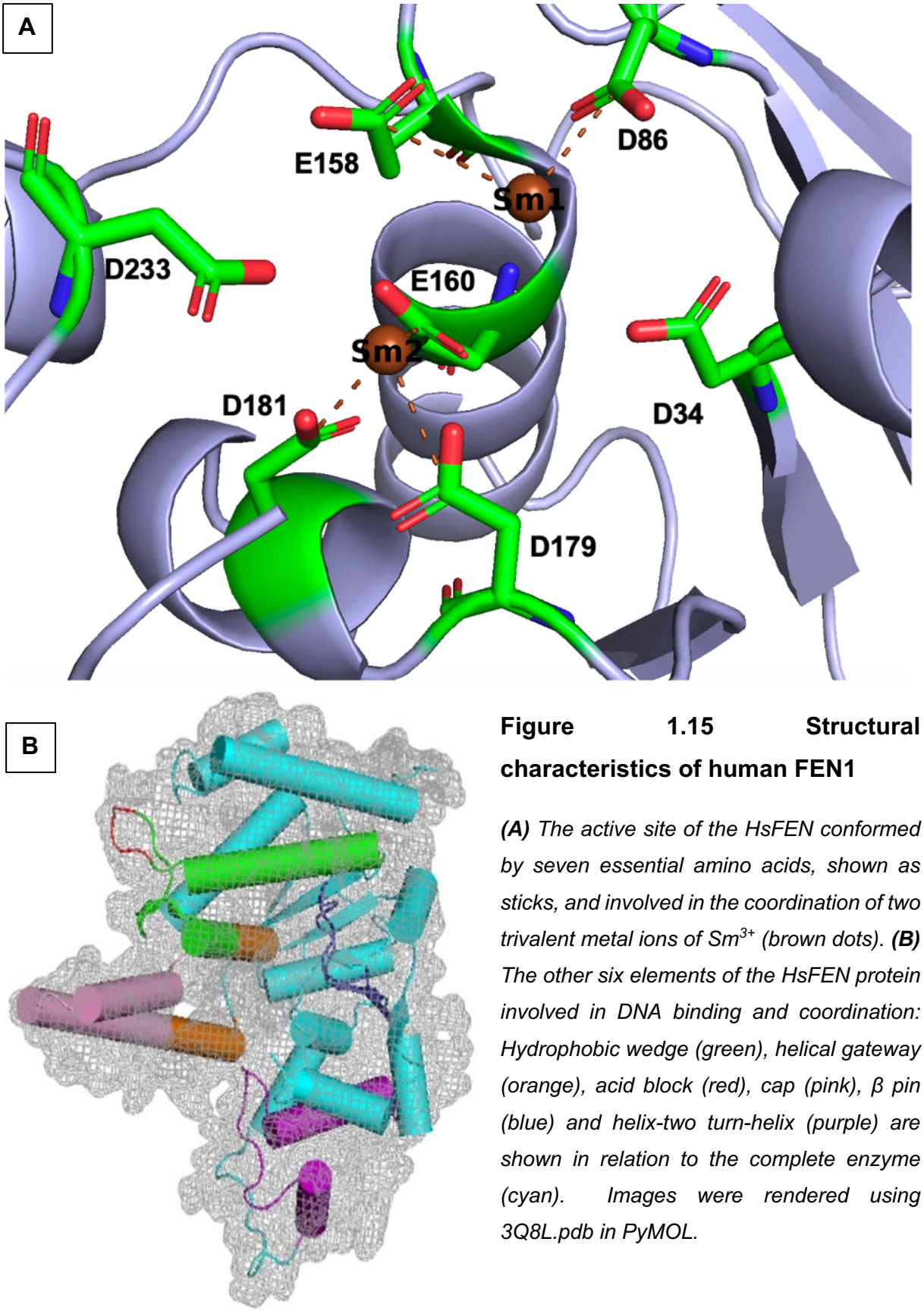


Figure 1.15 Structural characteristics of human FEN1

(A) *The active site of the HsFEN conformed by seven essential amino acids, shown as sticks, and involved in the coordination of two trivalent metal ions of* Sm^{3+} *(brown dots).* ***(B)*** *The other six elements of the HsFEN protein involved in DNA binding and coordination: Hydrophobic wedge (green), helical gateway (orange), acid block (red), cap (pink), β pin (blue) and helix-two turn-helix (purple) are shown in relation to the complete enzyme (cyan). Images were rendered using 3Q8L.pdb in PyMOL.*

The other six components of the HsFEN are associated with DNA recognition and interaction of the enzyme with the best possible substrate. The H2TH binds to the double-stranded DNA from the 5'-flap, though the helical cap functions as a filter, offering preference to process 5'-ends substrate (Tsutakawa *et al.*, 2011). The hydrophobic wedge is in charge of breaking the DNA path and create the 3' flap binding site and the acid block represses 3' folds that are longer than one nucleotide. The helical gateway allows the passage of a single-stranded DNA inside the enzyme, permitting the interaction with the active site, just as shown in Figure 1.16 (Tsutakawa *et al.*, 2011).

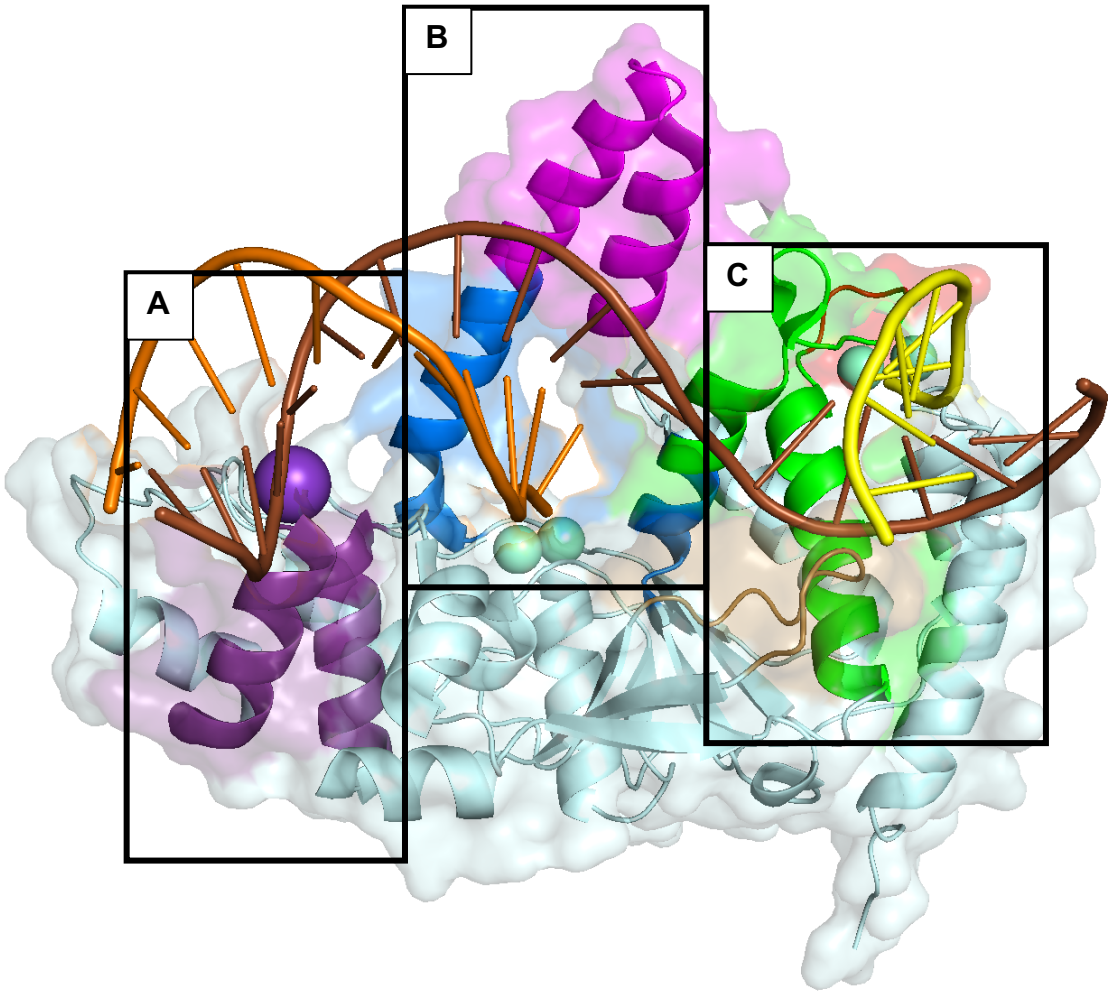
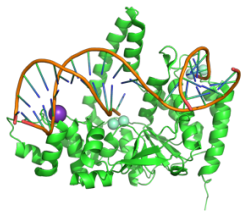
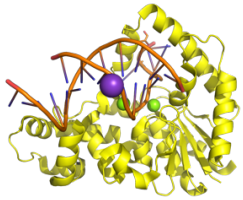
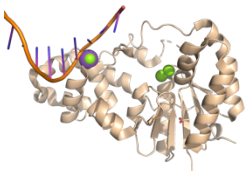
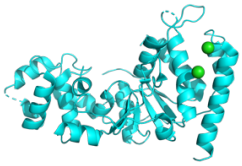
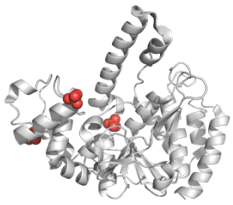
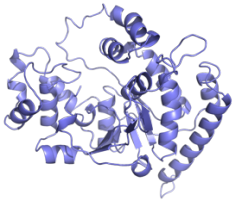


Figure 1.16 The six components of the human FEN1 associated with DNA recognition and processing (PDB 3Q8K)

(A) *Binding of the double-stranded DNA by H2TH (purple).* ***(B)*** *Passage of the single-stranded DNA and processing of the 5'-end substrate (orange) through the active site by the helical gateway (blue) and helical cap (magenta), respectively.* ***(C)*** *Blockage of the template DNA strand (brown) by the hydrophobic wedge (green), forming a 3' flap binding site. Inhibition of long 3'-flaps (yellow) by the acid block (red). Image was rendered using 3Q8K.pdb in PyMOL.*

As was mentioned previously, crystal structures of FEN enzymes can differ depending of the organism. The amino acid sequences of FEN from different organisms can vary in terms of length and in the percentage of identity. However, proteins share the same structural components and most of the conserved residues in the active site. Comparison in percentage of identity between the human and other organisms with known structures, according to BLASTp (Altschul *et al.*, 1990; Altschul *et al.*, 1997) is shown in Table 1.8. An in-depth comparison of sequence/structure of human and T5 FENs, a multiple sequence alignment (MSA) between the human and parasitic FENs and a predicted second structure domain MSA between the human and parasitic FENs are shown Figures 1.17, 1.18 and 1.19, respectively.

Table 1.8 FEN structures from different organisms and their percentage of identity against human FEN

Identity	Structure	Percentage of Identity	Identity	Structure	Percentage of Identity
Human FEN[1]		100%	T5FEN (5HNK)[5]		27%
E. coli ExoIX (3ZDB)[2]		35%	*M. kandleri* FEN (4WA8)[6]		42%
Desulfurococcus amylolyticus FEN (3ORY)[3]		36%	*M. jannaschii* FEN (1A76)[7]		35%
P. furiosus FEN (1B43)[4]		40%			

[1] Tsutakawa *et al.*, 2011. [2] Anstey-Gilbert *et al.*, 2013. [3] Mase *et al.*, 2011. [4] Shah *et al.*, 2015. [5] AlMalki *et al.*, 2016. [6] Hosfield *et al.*, 1998. [7] Hwang *et al.*, 1998.

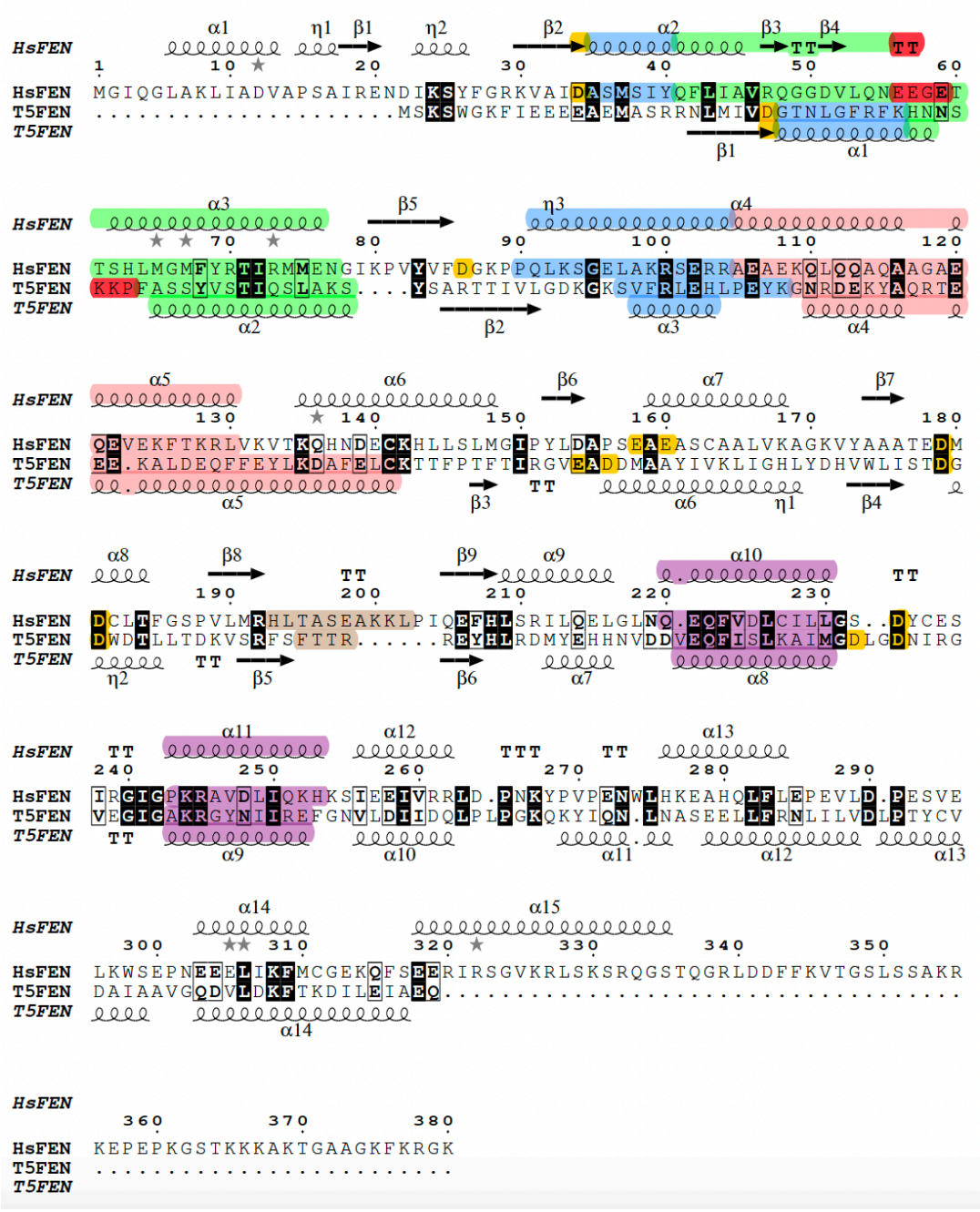


Figure 1.17 Sequence alignment of human and T5 FEN

Identical and similar residues are boxed in black and white, respectively. Secondary structures are represented as α (alpha-helix), β (beta-strand), η (3_{10}-helix), TT (strict beta-turns), TTT (strict alpha-turns). Alternate conformations of the residues are highlighted as gray stars. Components of the FENs are highlighted in different colours according to Figure 1.16: helical gateway (blue), hydrophobic wedge (green), acid block Hs*FEN (red; in* T5*FEN it is a basic block), helical cap (pink), β pin (brown) and helix-two turn-helix (purple). Conserved residues of the active site are highlighted in yellow. Sequence alignment of both proteins was performed using the online tool MultAlin (Corpet, 1988). Depiction of sequence alignment and secondary structure was carried out with the online tool ESPript 3.0 (Robert & Gouet, 2014).*

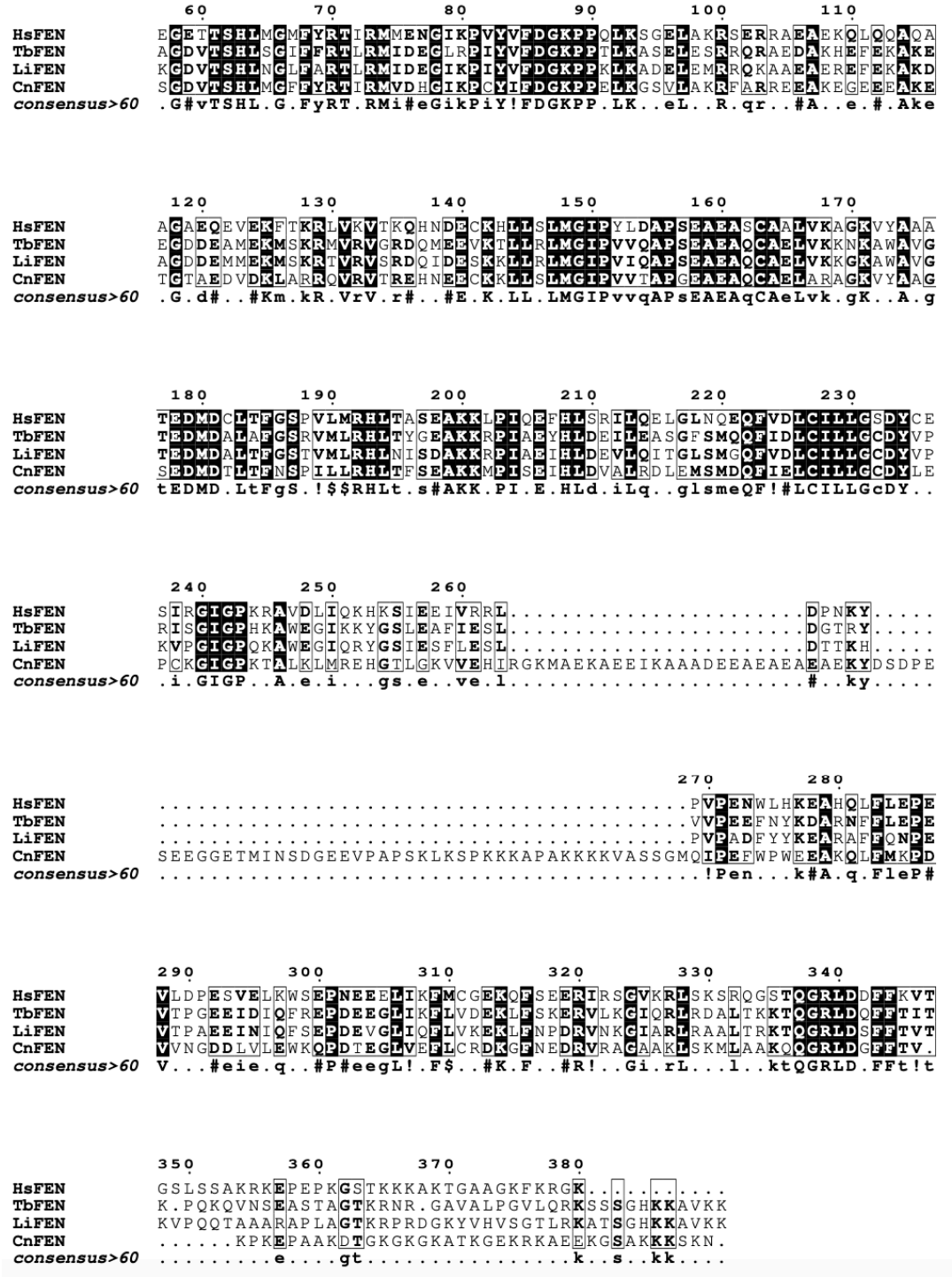


Figure 1.18 Multiple sequence alignment of human and parasitic FENs

Hs*FEN,* Tb*FEN,* Li*FEN and* Cn*FEN corresponds to human,* Trypanosoma brucei*,* Leishmania infantum *and* Cryptococcus neoformans *FENs, respectively. Identical are boxed in black while similar residues are boxed in white according to a 60% low level consensus. Consensus symbols: ! is anyone of Ile or Val; $ is anyone of Leu or Met; % is anyone of Phe or Tyr; # is anyone of Asn, Asp, Gln or Glu. Multiple sequence alignment of the proteins was performed using the online tool MultAlin (Corpet, 1988). Depiction of sequence alignment was carried out with the online tool ESPript 3.0 (Robert & Gouet, 2014).*

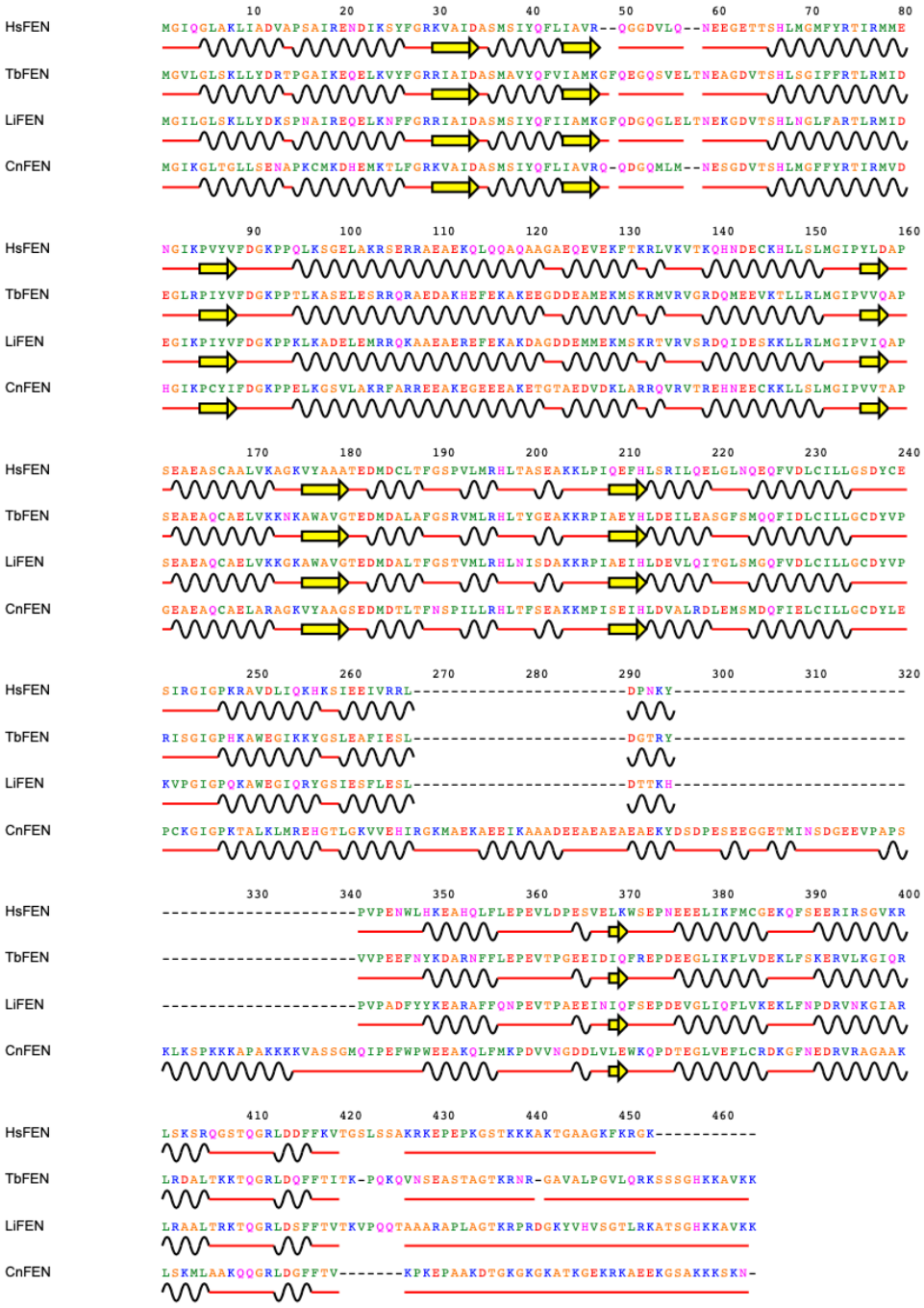


Figure 1.19 Secondary structure domain MSA between human and parasitic FENs

Alpha helices are represented as black waves and beta sheets as yellow arrows. Hydrophobic (Cys, Val, Ile, Leu, Pro, Phe, Tyr, Met, and Trp), small (Gly, Ala, Ser, and Thr), polar (Asn, Gln, His), negatively charged (Asp and Glu) and positively charged (Lys and Arg) residues are coloured in green, orange, pink, red and blue, respectively. Multiple sequence alignment of the proteins was performed using MultAlin (Corpet, 1988). Prediction of secondary structure domains was carried out using the online tool Ali2D (Zimmermann et al., *2018; Gabler* et al., *2020; Jones, 1999; Nugent & Jones, 2009). Depiction of MSA with the predicted domains was performed with 2dss (Lotun* et al., 2019*).*

1.7.3 Biochemical Activity

FENs cleave a large range of substrates in vitro with a 5' to 3' polarity with endo- and exonuclease activity. These properties are essential for the processing of DNA/RNA intermediates which form in cells during replication, as discussed above. (Finger *et al.* 2012).

The exonuclease activity consists in removing nucleotides from the end of a polynucleotide chain, cleaving nucleotides from the 5' end and able to release nucleotides from that point. FEN1 exonuclease activity is capable of processing single-stranded DNA and double-stranded DNA with branched structures (Harrington & Lieber, 1994). Interestingly, FEN1 exonuclease activity was not seen at blunt ends, 5' overhangs or RNA/DNA duplexes and the final product of the exonucleolytic reaction was, in most of the cases, dinucleotides (Harrington & Lieber, 1994). Even though FEN1 does not possess exonuclease activity on RNA/DNA duplexes, the enzyme binds to the double-stranded DNA flap junctions and removes both 5'-single stranded RNA and DNA flaps (Harrington & Lieber, 1994; Tsutakawa *et al.*, 2011).

Endonucleases are capable of cleaving phosphodiester bonds within a polynucleotide chain or circular structure (does not require a 3' end). FENs are characterized by their specificity for DNA flap structures (Harrington & Lieber, 1994) and require a free 5' end. The sequence and the length of the flap is not important as the structure of the substrate as these enzymes cleave specifically at the 5' flap strand bifurcation (Harrington & Lieber, 1994).

The crystal structure of the human FEN flap complex reveals an unpaired 3' flap that is surrounded by 10 residues, where the 3' hydroxyl group forms a bond with the backbone carbonyl of Lys314 and Thr61 hydroxyl (Tsutakawa *et al.*, 2011). For the 5' flap strand, this is unpaired from the template through a helical arch, comprised of two alpha-helices: α2 and α4. The size of the helical arch is about 13-15 Å, which allows only single-stranded DNA to pass through (Tsutakawa *et al.*, 2011; Figure 1.20).

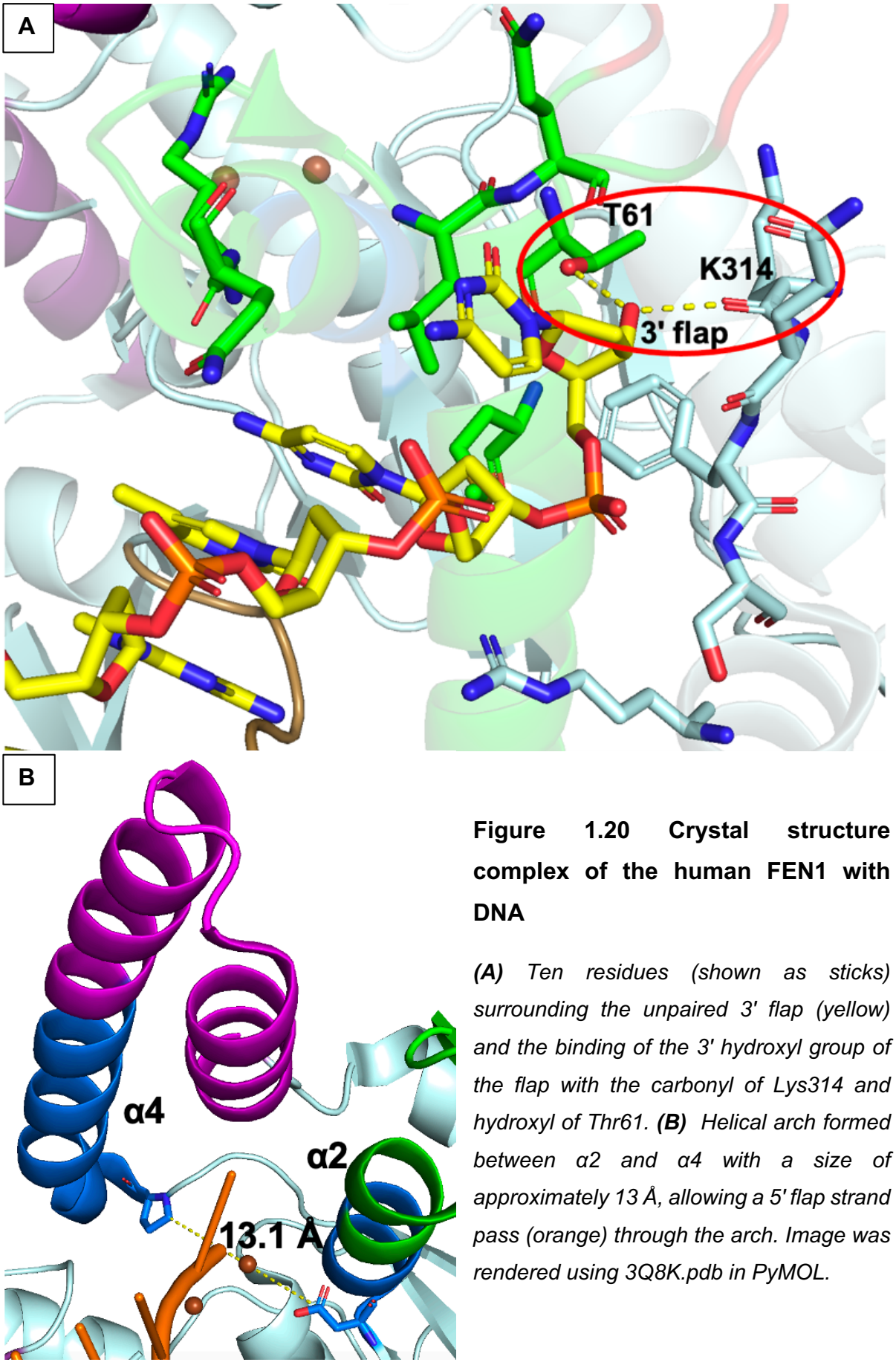


Figure 1.20 Crystal structure complex of the human FEN1 with DNA

(A) Ten residues (shown as sticks) surrounding the unpaired 3' flap (yellow) and the binding of the 3' hydroxyl group of the flap with the carbonyl of Lys314 and hydroxyl of Thr61. (B) Helical arch formed between α2 and α4 with a size of approximately 13 Å, allowing a 5' flap strand pass (orange) through the arch. Image was rendered using 3Q8K.pdb in PyMOL.

The best substrate for FEN1 of bacteria, yeast, archaebacteria, and human is a double-stranded DNA (dsDNA) substrate with a specific single-nucleotide 3' flap as well as 5' single stranded region. FEN1 acts by separating the hydrogen bonds of one or a couple of nucleotides to generate a small flap and after that it cuts (endonucleolytic cleavage) the phosphodiester bond at a point with the double-strand to release the flap (Balakrishnan & Bambara, 2013; Figure 1.21).

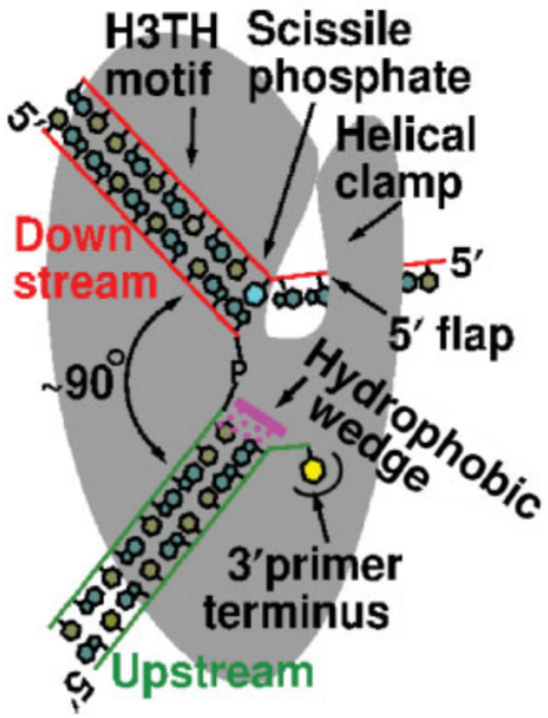


Figure 1.21 Interaction of FEN1 and DNA substrate

The image shows the cutting point of the phosphodiester bond and cleavage of the 5' single stranded flap. Image depicted from the study of Shen et al. *(2005). License number 5398870034849.*

The activity of FEN1 can only be functional in presence of metal ions, for example, magnesium (Mg^{2+}) and manganese (Mn^{2+}), but this activity can be inhibited by cations such as zinc (Zn^{2+}) and calcium (Ca^{2+}; Harrington & Lieber, 1994). The preference of specific metal ions, especially magnesium, may be due to the high abundance of this element inside the cells, as well as certain chemical properties such as a small ionic radius and lack of redox reaction (Cowan, 1998). *In vitro*, human FEN1 enzyme-mediated DNA hydrolysis requires divalent metal ions typically in the range from 1 mM to 10 mM Mg^{2+}, and an optimal temperature of 37°C, salt (NaCl or KCl) concentration of 50mM or lower, and a pH ranging from 6 to 10, with an optimal pH of 8 (Harrington & Lieber, 1994; Hosfield *et al.*, 1998). The importance of metal ions for FEN1 activity is due to phosphodiester hydrolysis (Figure 1.22). The enzyme generates a hydroxide ion from a water molecule bridging the two metal ions (Syson *et al.*, 2008).

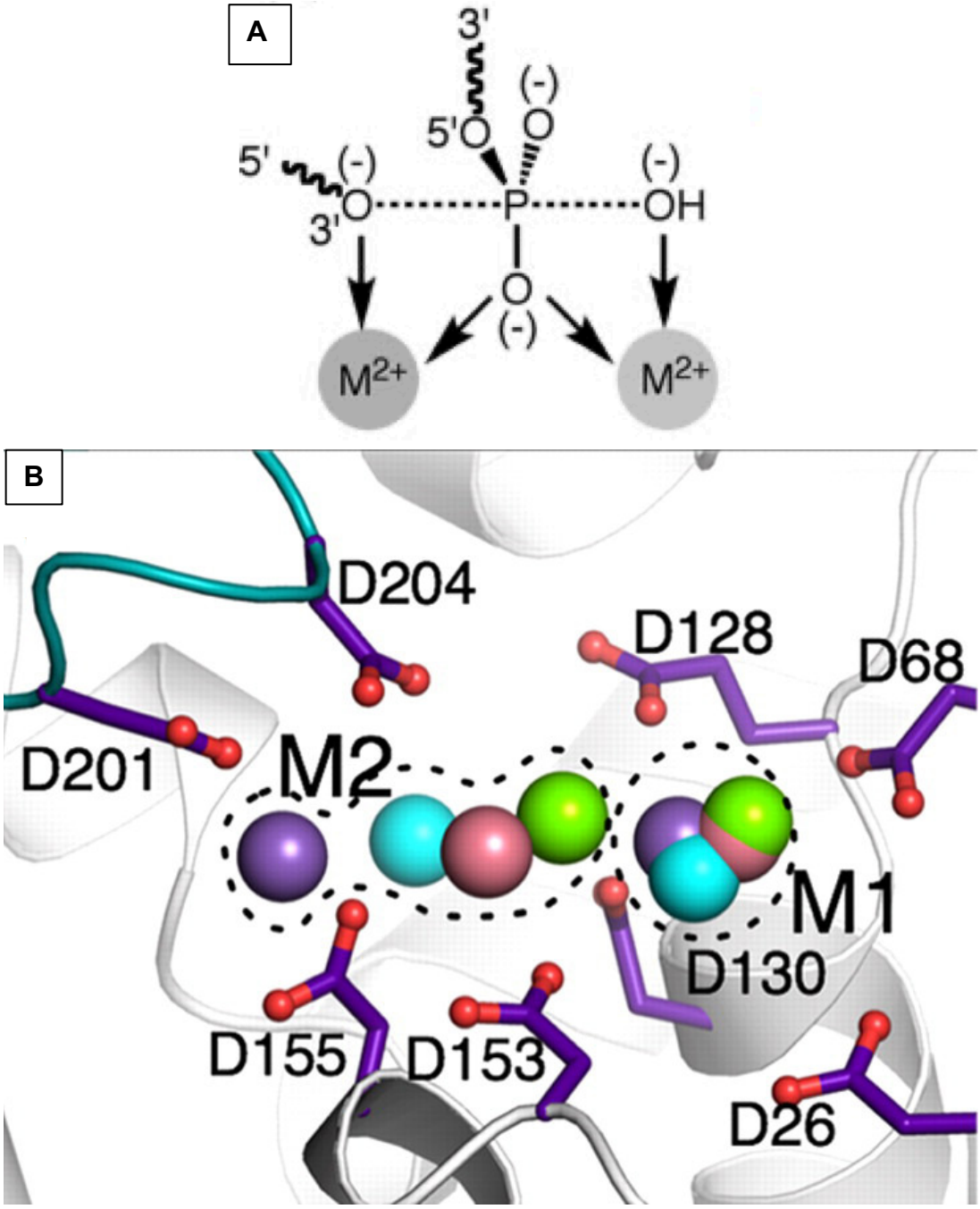


Figure 1.22 Two-metal ion mechanism in FEN structures

(A) *One metal acts as nucleophilic hydroxide ion and binds to a non-bridging oxygen of the phosphodiester. The same non-bridging oxygen is bound by a second metal ion that is coordinated to the leaving group oxygen.* ***(B)*** *Active site of the T5FEN (1UT5) shown as purple sticks with its two metal ions (M1 & M2 shown as purple spheres). Metal ions of T4FEN (1TFR, cyan),* M. jannashi *FEN (1A77, pink), human FEN (1UL1, green) are overlayed in the space of T5FEN active site. Images reproduced under the creactive commons license from Journal of Biological Chemistry (Syson* et al.*, 2008).*

1.7.4 DNA-Protein Binding: FEN Mechanism

It has been proposed that FEN mediated cleavage of flap substrates is initiated by binding of protein to double-stranded DNA, preferentially with one nucleotide 3'-flap and any length of 5'-flap (Tsutakawa *et al.*, 2017). When the enzyme diffuses to the fork conformational change happens, allowing the 5' single-stranded flap end entering through an opening involving the helical arch structure of the FEN (Shaw *et al.*, 2017). After the threading of the ssDNA into the helical arch, the enzyme can enclose the substrate, which is situated in the ideal contact with the active site, promoting hydrolysis of the phosphodiester bond and cleavage of the ssDNA 5' flap. Once the threading has been accomplished, the helical arch returns to the original conformation. This model, named as disorder thread-order model, explains why only the large 5' flap strands can bind to the flap endonuclease (Figure 1.23; Patel *et al.*, 2012; AlMalki *et al.*, 2016).

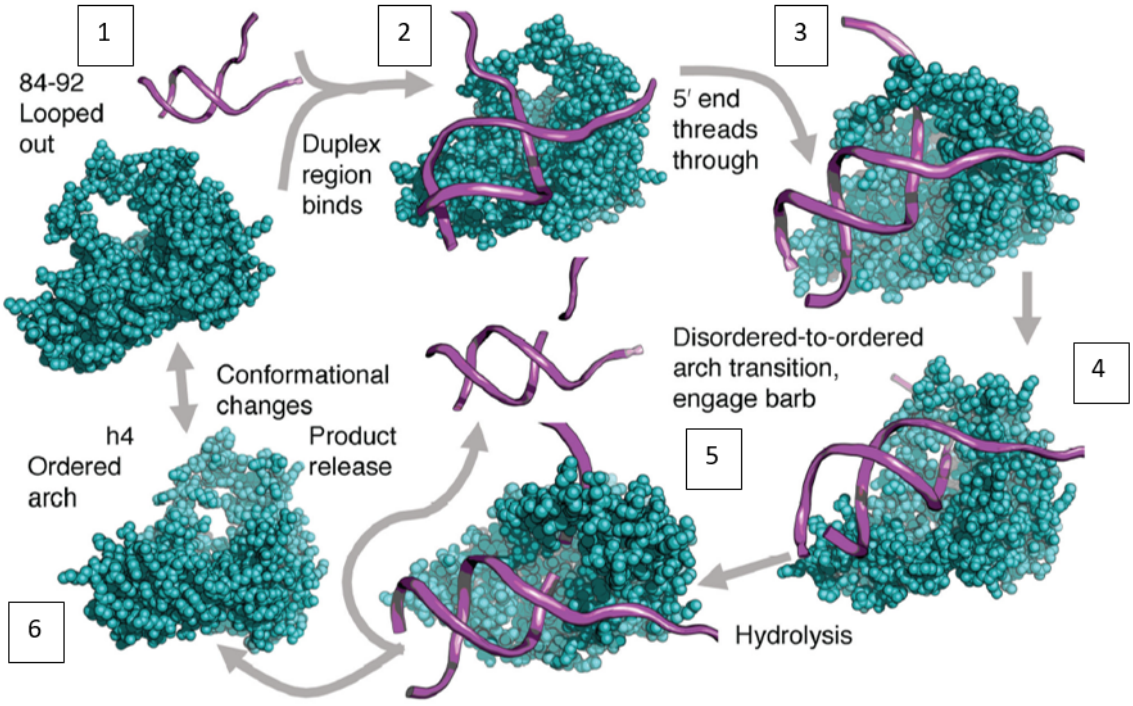


Figure 1.23 Mechanism of binding and threading of DNA in T5FEN

(1) *Recognition of the substrate by FEN.* ***(2)*** *FEN enzyme binds to the double-stranded DNA.* ***(3)*** *The 5' flap strand threads through the helical arch.* ***(4)*** *Distortion of the helical arch to process the 5' flap strand.* ***(5)*** *FEN endonuclease activity over the 5' flap strand, cutting the 5' flap end.* ***(6)*** *The helical arch of FEN returns to the original conformation. Image modified from the study of AlMalki* et al., *2016. License number 5367670242286.*

Nevertheless, a model proposed by Tsutakawa and co-workers (2017) suggests that the mechanism of action in eukaryotic FEN1 is not exactly the same compared to, for example, the bacteriophage T5 flap endonuclease previously described by AlMalki and collaborators (2016). What makes the eukaryotic FEN1 different from the other enzymes is a process called “phosphate steering” during threading of the ssDNA 5'-flap, orientating the flap during the threading and moving the target phosphodiester bond to the catalytic metal ions (Figure 1.24).

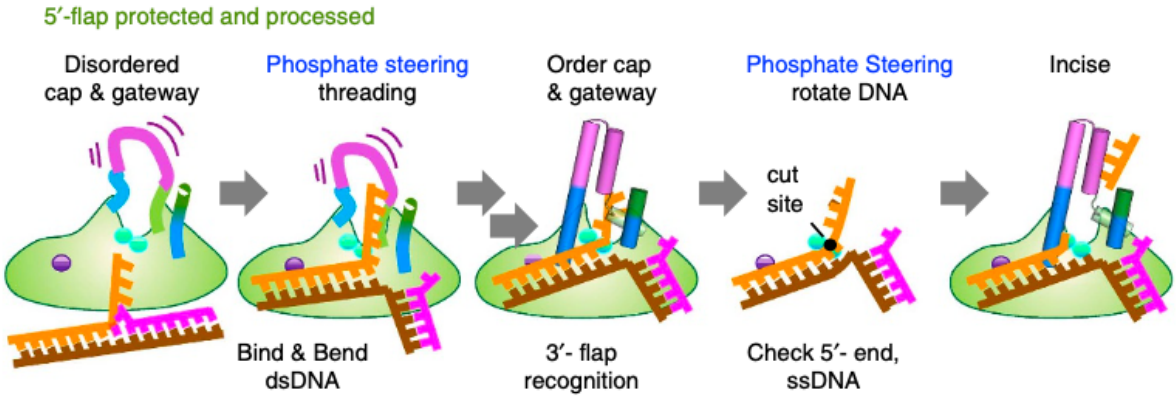


Figure 1.24. Mechanism of binding and threading of DNA in human FEN

Schematic model of the FEN1 mechanism highlighting the phosphate steering and 5'-flap inverted threading steps, guiding the DNA flap into the catalytic metals. Image reproduced under the creactive commons license from Nature Communications (Tsutakawa et al*., 2017).*

1.7.5 Human FEN1 as a potential target for drug development

As has been previously mentioned, flap endonucleases play an important role in DNA replication and repair processes. In studies conducted by Kucherlapati *et al.*, (2002) and Larsen *et al.*, (2003) demonstrated that *FEN1* knockout mice show early embryonic lethality, demonstrating that FEN1 is essential for the development of the organism.

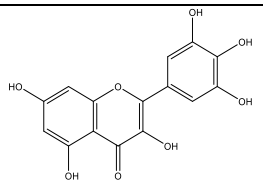
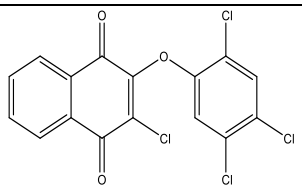
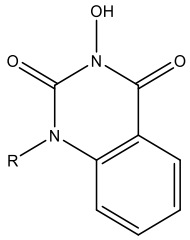
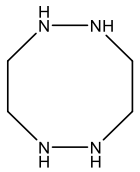
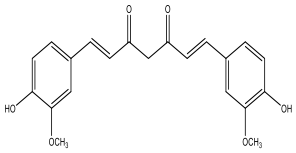
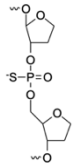
Furthermore, down-regulation of *FEN1* has a negative impact in the organism leading to increased apoptotic cell death after ionizing radiation and hypersensitivity to DNA alkylating agents and hydrogen peroxide (Larsen *et al.*, 2003; Matsuzaki *et al.*, 2002; Nikolova *et al.*, 2009). However, overexpression of the *FEN1* gene has been implicated

in different types of cancer, such as lung, gastric prostate, pancreatic, brain, and breast cancer (Zheng *et al.*, 2011; Illuzzi & Wilson 2012), development and/or progression of tumours (Lam *et al.*, 2006), and cancer cell survival in response to anticancer treatment (Freedland *et al.*, 2003).

Given that FEN1 is essential for the replicative and repairing process of DNA, it has been suggested that the protein can be further studied as a biomarker related to prognosis and disease progression, as well as a potential therapeutic target (Exell *et al.,* 2016). The interest of this protein in the therapeutical field arises from known synthetic lethal interactions with mutated genes in cancers (van Pel *et al.,* 2013; Illuzi & Wilson, 2012). For example: according to Shibata and Nakamura (2002) FEN1 has been identified as a potential chemosensitizing target due to its role in LP-BER from induced alkylation damaged caused by methyl methanesulfonate (MMS). Moreover, its knockdown or inhibition increases the sensitivity to temozolomide (TMZ) in glioblastoma (Nikolova *et al.*, 2009) and colorectal cancer cell lines (Panda *et al.,* 2009; van Pel *et al.,* 2013; McManus *et al.,* 2009) and the sensitivity to cisplatin in lung cancer cell lines (He *et al.*, 2017).

From the studies previously mentioned, an interest in developing FEN1 inhibitors has arisen in the pharmaceutical field. These potential inhibitors can be produced as single-form treatments for cancers that depend on its activity, as well as as a combined therapy with chemotherapeutic agents that induce DNA damage (McWhirter, 2013). Moreover, the significance of FEN activity in all organisms for DNA replication and repair, at the same time as the differences between FENs and other superfamily members, has become an interesting target to further develop medications apart from anticancer treatments, such as antibacterial and antiviral therapies (Patel *et al.*, 2012; Table 1.9).

Table 1.9 Development of some inhibitors targeting HsFEN

Drug inhibitor	Chemical structure	Mechanism of action
Myricetin[1]		The mode of action remains unknown. Notwithstanding, as indicated by the investigation of structure-activity relationship (SAR) the combination of 3', 4', 5' −OH is essential for the striking hindrance.
JFD00950[2]		Possible interaction with FEN1 through the tryptophan residues, which makes plausible the binding of JFD00950 with FEN1, making an important change in the secondary structure of the enzyme.
N-hydroxyurea compounds (SC13, C20)[3,4,5,6,7]		Prevents the entrance of the scissile phosphate diester of substrate DNA to the catalytic two Mg^{2+} metal ions. The action mechanism is not completely comprehended. It is estimated that there is an arrangement of a prereactive complex that requires the end of the DNA duplex so it can bind to metal ions, inhibiting the action of FEN1.
NSC-281680[8]		Blocks the long patch base excision repair (LP-BER). The NSC-281680 interacts with Asp181 of the FEN1 metal binding site (M2), causing some structural changes and the loss of Mg^{2+} cleavage, leading to the loss of the FEN1 activity.
Curcumin[9]		Unknown for the moment but according to latest studies, curcumin significantly diminish FEN1 promoter activity and protein expression interceded by Nrf2. Also, Nrf2 assumes a significant role in the development and progression of tumours.
Phosphorothioate -modified FEN1 substrate[10]		Used as a substrate competitor, the phosphorothioate bond replaces the phosphodiester bond at the FEN1 cleavage site of the DNA substrate, slowing down nuclease catalysis and tumor growth.

[1]Ma et al*., 2019; [2]Deshmukh* et al*., 2017; [3]Ward* et al*., 2017; [4]Exell* et al*., 2016; [5]Tumey* et al*., 2005; [6]He* et al*., 2016; [7]He* et al*., 2017; [8]Panda* et al*., 2009; [9]Chen* et al*., 2014; [10]Ba* et al*., 2019. Structures were rendered using Chemdraw 16.*

The considerable interest in human FEN1 (hFEN1) as a drug target has subsequently surged in the development of high-throughput screening procedures in order to identify potential inhibitors (McWhirter *et al.,* 2013; Dorjsuren *et al.,* 2011). And from the inhibitors developed to target hFEN1, the N-hydroxyurea based molecules are the most studied ones, with a known molecular mechanism between these compounds and hFEN1, as well as a demonstrated cellular activity and target engagement in live cells (Exell *et al.*, 2016).

Previously, two studies from Tumey and collaborators have shown that 2,4-diketobutyric acids and N-hydroxyurea series can inhibit FEN1 *in vitro* (Tumey *et al.*, 2004, Tumey *et al.*, 2005). Later, Exell and collaborators (2016) studied further the N-hydroxyurea series as potential inhibitors. In this study, it became possible to elucidate the crystal structure of a N-hydroxyurea molecule bound to the active site of FEN1. The inhibition of the protein is achieved through two different interactions: the coordination of Mg^{2+} ions with the N-hydroxyurea moiety, disrupting the conformational change within the substrate duplex to allow enzymatic cleavage; and the thiophene ring, which places in the hydrophobic pocket of FEN1 containing the conserved residues Y40, D181 and R100 essential for the positioning of the substrate (Algasaier *et al.*, 2016; Figure 1.25). The binding of N-hydroxyurea compounds to the active site of the protein occurs in a competitive and non-competitive reaction with DNA (Dovrat *et al.*, 2014; Stodola & Burgers, 2016). In addition, further members of the N-hydroxyurea series have shown to inhibit FEN1 *in vitro* (McWhirter *et al.*, 2013).

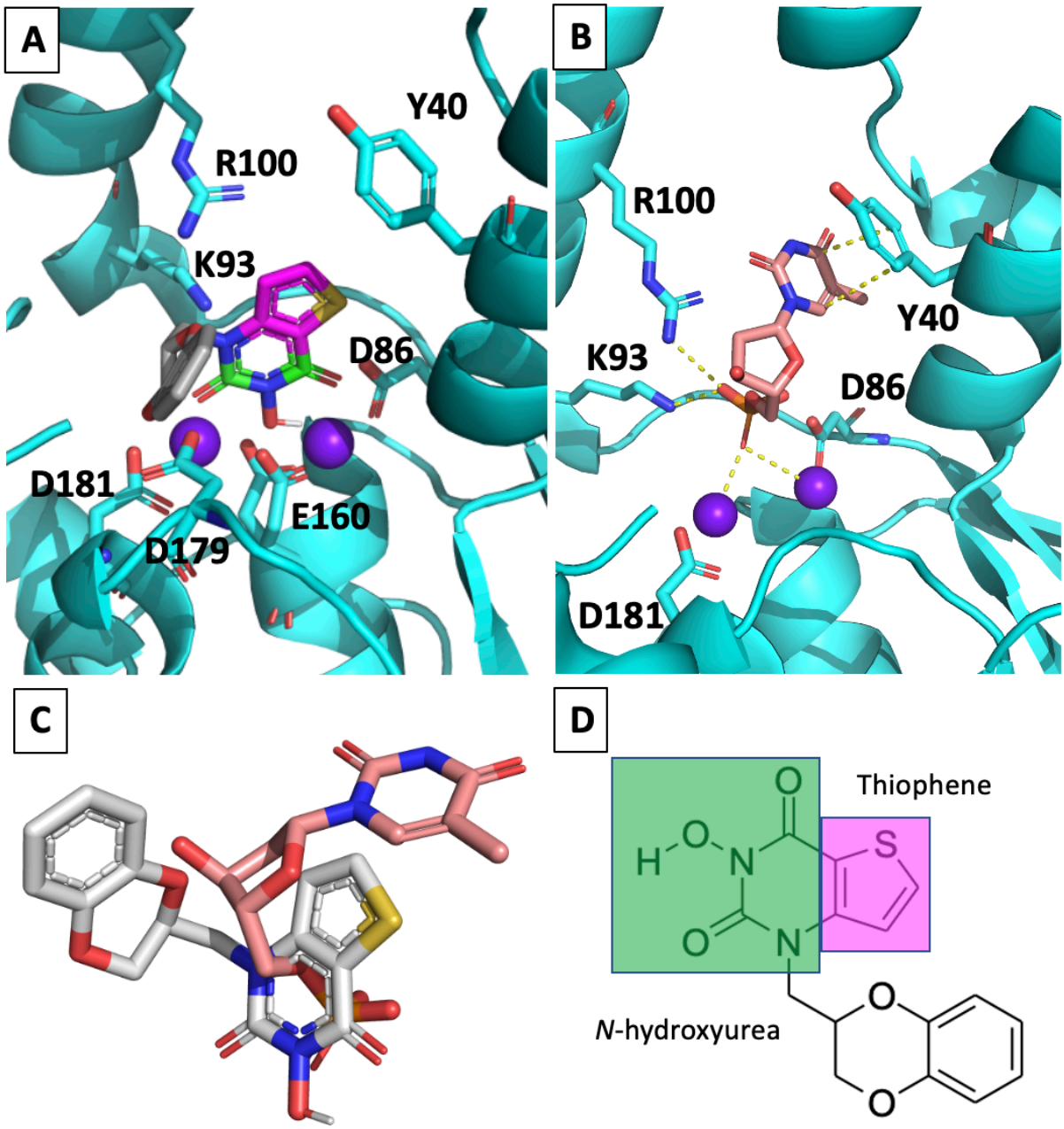


Figure 1.25 Binding of N-hydroxyurea to the active site of FEN1.

(A) *Chemical inhibitor of FEN1 (5FV7.pdb) binding to the active site and blocking the DNA cleavage through the helical gateway of FEN1 (3Q8K.pdb). Thiophene ring is depicted in magenta while N-hydroxyurea moiety is coloured in green.* ***(B)*** *FEN1-product structure (3Q8K.pdb) showing the phosphate monoester of the unpaired nucleotide (pink) in contact with the metal ions (purple) and the base of the helical gateway. For panels A and B, some residues of the active site and helical gateway are shown as sticks.* ***(C)*** *Spacial superposition of the N-hydroxyurea compound (gray) with the unpaired nucleotide generated from the DNA cleavage (pink).* ***(D)*** *Chemical structure of the N-hydroxyurea serie with the moieties of the molecule coloured according to panel A.*

Furthermore, *in vivo* assays have demonstrated that disruption or inhibition of FEN1 in organisms provokes DNA damage. Studies performed with *Saccharomyces cerevisiae* cells RAD27 deficient, a FEN1 homologue, presented temperature-sensitive lethality, sensitivity to DNA damaging agent methyl methanesulfonate (MMS) and damage of the post-replication repair (PRR) and DNA double-strand break (DSB) repair pathways followed by accumulation of unprocessed Okazaki fragments (Reagan *et al.,* 1995; Symington, 1998; Becker *et al.,* 2015). Additionally, Exell and collaborators (2016) have demonstrated that N-hydroxyurea molecules bind FEN1 protein *ex vivo* in SW620 colon cancer cells, starting a DNA damage response in a dose-response form. Inhibition of FEN1 leads to the interruption of the DSB repair pathway, which is caused by the lack of forming a complex with MRE11A, a member of the MRN complex (a protein complex consisting of Mre11, Rad50 and Nbs1), required for the initiation of DSB repair pathway and activation of cell-cycle checkpoints (van Pel *et al.,* 2013; Bekker-Jensen *et al.,* 2006; Hopfner, 2014; Grenon *et al.,* 2001; Porter-Goff *et al.,* 2009). Due to this event, alternative pathways are activated such as ATM checkpoint signalling pathway, phosphorylation of histone H2AX and ubiquitination of FANCD2, suggesting the initiation of the Fanconi anemia (FA) pathway (Exell *et al.* 2016). The pathway previously mentioned is required for the stabilisation of stalled replication forks and DNA damage associated with replication occurs in the presence of the N-hydroxyurea molecules.

Further investigation of the activity of N-hydroxyurea series inhibitors was carried out by Ward *et al.,* (2017) through high-throughput and targeted forms in human cancer cells. In the study it was identified synthetic genetic interactions between the inhibitors tested and disruption of the genes involved in DNA damage repair, suggesting FEN1 as a potential target for drug development in cancers. However, the study has highlighted the limitations in pharmacokinetics and efficacy. At the same time, promising results have only been demonstrated in animal models (Xin *et al.,* 2020; Bian *et al.,* 2022; Xu *et al.,* 2020) which could not be reproducible in humans and be considered as future clinical treatments. Despite the effectiveness of these inhibitors demonstrated *in vitro*, *in vivo* and in xenograft models, there is still the need of further studies to get to the point of conducting clinical trials and be considered as novel drugs (Yang *et al.*, 2022).

1.8 Aims and Objectives

1.8.1 Hypothesis

According to the previous studies pointing out that flap endonuclease is essential for DNA replication and repair in all organisms (Kucherlapati *et al.*, 2002; Larsen *et al.,* 2003) and specifically in kinetoplastids (Ponce *et al.*, 2017), it is plausible option to develop novel treatments against neglected tropical diseases by targeting this enzyme. The interest to increase the treatments options for NTDs arised from the WHO target to control, eliminate and eradicate them by 2030 (WHO, 2020).

To develop those novel treatments it is necessary to further understand the molecular mechanism of the flap endonuclease. This can be achieved by structure determination of the protein with mutations of specific metal binding residues in the active site, producing iso-structural proteins, capable of DNA binding but not hydrolysis. After obtaining high-resolution structure of *Tb*FEN, *Li*FEN and *Cn*FEN*, in silico* strategies can be used to improve the hit rate of inhibitor discovery (*in silico* drug design (ISDD)). Furthermore, compounds that selectively inhibit pathogen FENs rather than human FEN can be identified from the ISDD results and from *in vitro* assays. Specific inhibitors of flap endonuclease detected from the previous methods should kill disease-causing microorganisms *in vivo*.

1.8.2 Project aims

I planned to tackle the project and to test the research hypothesis through six project aims.

1. Identify the organisms to target in this project.

NTDs have been classified in twenty diseases according to WHO, which are caused by pathogens such as parasites, bacteria and viruses. From the group of 20 diseases, I decided to target two, the diseases caused by the organisms *Trypanosoma brucei* and *Leishmania.* The reason for this is because the two organisms belong to the same clagroup, the kinetoplastid, which makes them phylogenetically close. In addition, these two organisms where selected according their impact on public health. The cases of African trypanosomiasis have declined to their lowest level in 80 years and the WHO has targeted this disease to be eliminated by 2030 (WHO, 2020). However, Leishmaniasis has been reported to be endemic in 92 countries, registering more than a million cases annually. Due to the large number of cases worldwide, WHO has developed a programme to control this disease (WHO, 2020). Moreover, a third organism to target was selected in this project, the fungus *Cryptococcus neoformans*. Even though cryptococcosis is not considered as a NTD according to WHO, the disease has become a concern among international health organizations due to be an opportunistic infection that occurs primarily among people with advanced HIV disease and is an important cause of morbidity and mortality in this group (Moran *et al.*, 2014; Drugs for Neglected Diseases initiative, 2020).

2. Develop a plan to test the research hypothesis.

The plan developed to test the research hypothesis follows a similar pipeline utilized in the structure-based drug design campaign (Figure 1.7; Anderson, 2003) consisting of: (1) Identifying a drug target. (2) Obtain a pure sample of the target in solution (expression and purification of parasitic FENs). (3) Determine the structure by X-ray crystallography. (4) Evaluate compounds against the target selected site (*in silico* through determined crystal structures and *in vitro* assays).

3. Express and purify wild-type parasitic FENs, free of contaminants for further assays.

A T7 promoter vector containing the codon optimised sequence of *Tb*FEN, *Li*FEN and *Cn*FEN for over-expression in lactose-inducible *E. coli* cell system. After expression, the protein of interest is purified through chromatography techniques.

4. Crystallise and determine the structural conformation of parasitic FENs.

To determine the protein structure by X-ray diffraction, the purified protein is prepared for crystallization experiments. The diffraction data collected will help during characterisation of the protein and future drug design. Protein crystallization with and without DNA substrate will be useful in the understanding of molecular mechanisms utilized by flap endonucleases.

5. Develop selective inhibitors of pathogen FEN enzymes.

Two fragment libraries containing a total of more than 1500 molecules which reveal structural heterogeneity and are suitable for further compound design are physically screened against the parasitic FENs and human FEN through the Förster Resonance Energy Transfer (FRET) assay. The data obtained from the assays will determine the interaction of the enzyme with the compounds and will identify inhibitors.

6. Build and evaluate *in silico* screening protocols to facilitate physical compound selection.

Two fragment libraries containing a total of more than 1500 molecules were screened through virtual high-throughput screening against the parasitic FENs using different algorithms from a computational tool. The screening will be performed according to the structural information of the protein obtained experimentally or through online prediction tools and the chemical structure of the compounds. The data results obtained from the screenings predicting the binding energy between the protein and the molecules will facilitate the selection of compounds with the highest binding energy for physical screening.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cloning

Three *Trypanosoma brucei* flap endonuclease (*Tb*FEN) constructs were produced by gene synthesis (see Table 2.1). These constructs were designed to express the protein encoded by *T. brucei brucei* TREU927 FEN gene (NCBI Gene ID: 3656016). A *Leishmania infantum* flap endonuclease (*Li*FEN) construct was produced to express the protein of *L. infantum* JPCM5 FEN gene (NCBI Gene ID: 5070003). Five *Cryptococcus neoformans* flap endonuclease proteins (*Cn*FEN) were produced for biochemical and crystallisation assays. These constructs were modelled on the *C. neoformans* var. *neoformans* JEC21 FEN gene (NCBI Gene ID: 3257380). All the constructs were codon optimised for improved expression in *Escherichia coli* (Table 2.1).

Table 2.1 Constructs produced and/or used for this thesis

Name	Production
*Tb*FEN-pET	Synthesised by Eurofins MWG Operon in pUC19 vector.
*Tb*FEN D183K ΔC-pET	Synthesised by Eurofins MWG Operon in pUC19 vector.
*Tb*FEN D183K-pET	Produced from *Tb*FEN-pET and TbFEN D183K ΔC-pET.
*Li*FEN-pET	Synthesized by Eurofins MWG Operon in pEX-K4 vector.
*Cn*FEN-pET	Synthesized by Bio Basic Inc in pUC57 vector.
*Cn*FEN-pYM547c	Produced from *Cn*FEN-pET and inserted in pYM547c vector.
*Cn*FEN ΔC414-pYM547c	
*Cn*FEN ΔC405-pYM547c	

2.1.1 Bacterial strains and plasmids

Two *E. coli* strains were used in this project (Table 2.2): XL1 Blue for storage of the plasmid at -80°C (Bullock *et al*., 1987), and BL21(DE3), for over-expression of the protein of interest (Studier & Mofatt, 1986). The plasmids pET21a(+) and pYM547c (Studier *et al.,* 1990; Rosenberg *et al.*, 1987), which both of them are controlled by the T7 promoter system, were used for the expression of the *E. coli* recombinant protein. The plasmid pYM547c, a gift from the former post-doctoral researcher Dr John Darby, is a modified vector from the pET28a+, inserting a C-terminal double histidine tag (6x

His-tag) and a human rhinovirus (HRV) 3C protease site for cleavage of the histidine tag. The modification was previously performed through plasmid mutagenesis (Wu *et al.*, 2018).

Table 2.2 *Escherichia coli* strains used in this work and their genotypes

Strain	Genotype
XL1 Blue	(*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lac*I^q*ZΔM15* Tn*10*(Tet^r)])
BL21 (DE3)	(B F^- *ompT gal dcm lon hsd*S_B(r_B–m_B–) λ(DE3 [*lacI lacUV5-T7 gene 1 ind1 sam7 nin5*])[*malB*$^+$]$_{K-12}$(λS))

2.1.2 Chemically competent *E. coli* cells

From a glycerol stock of *E.* coli, either XL1 Blue or BL21(DE3), was streaked onto the LB agar plate (1 % (w/v) tryptone (Merck Millipore; 16922), 0.5% (w/v) yeast extract (Merck Millipore; 103753), 0.5% (w/v) NaCl and 1.5% (w/v) bacteriological agar (Oxoid; LP0011)) with the appropriate selection marker and incubated overnight at 37°C. From this plate, a single colony of *E. coli* was picked and then placed in 5 mL of LB media with the respective selection marker and incubated at 37°C, overnight. The overnight culture was diluted 1:50 in 100 mL of LB media containing the selection marker (LB media: 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl). The *E. coli* cells were grown at 37°C until the $A_{600\ nm}$ of 0.3-0.4 was reached. The culture was centrifuged at 645 x *g* for 20 minutes at 4°C. The cell pellet was re-suspended in 40 mL of ice-cold 100 mM $CaCl_2$ and incubated on ice at 4°C, overnight. The cell suspension was centrifuged at 215 x *g* for 20 minutes at 4°C. The pellet was re-suspended in 6 mL of ice-cold 85 mM of $CaCl_2$ with 15% glycerol (v/v). The cells were stored in 320 µL aliquots at -80°C.

2.1.3 Preparation of DNA plasmid

Plasmid DNA was isolated and purified using the commercial kit Monarch Midi and Mini prep (New England BioLabs, NEB). From a glycerol stock of *E.* coli, either XL1 Blue or BL21(DE3), was streaked onto the lysogeny broth (LB) agar plate with the appropriate selection marker and incubated overnight at 37°C. A single colony of *E. coli* was isolated and then placed in 10 mL of LB media with the respective selection marker and incubated at 37°C, overnight. Cells were harvested by centrifugation at

3400 x *g* for 10 minutes at room temperature, purified according to the manufacturer's instructions and stored in -20°C.

2.1.4 Transformation of chemically competent *E. coli* cells

Digestion of the DNA plasmids were made with the appropriate restriction enzymes from NEB according to each gene and as indicated to the manufacturer's protocol. One µg of DNA was incubated with 10 units of enzyme in the suitable buffer and incubated for one hour at 37°C. DNA fragments were separated by 1% agarose electrophoresis (see Section 2.4.1 for details), and the desired band was cut and isolated using the Monarch DNA Gel Extraction Kit (NEB). The DNA fragments were ligated into pET21a(+) and pYM547c using T4 DNA Ligase (NEB) and NEBuilder HiFi DNA Assembly Master Mix (NEB), respectively.

Ligation into pET21a(+) was carried out by adding 50 ng of digested pET21a(+) on a Eppendorf tube with 37.5 ng of digested insert, 2 µL of T4 DNA Ligase buffer (10x), 1 µL of T4 DNA Ligase and nuclease free water up to 20 µL. The reaction was gently mix by pipetting and incubated in ice overnight. The next day, the ligated vector was ready for transformation. Ligation into pYM547c was performed by setting up the reaction on ice in an Eppendorf tube containing 75 ng of pYM547c vector, 75 ng of insert (PCR product), 10 µL of NEBuilder HiFi DNA Assembly Master Mix and nuclease free water up to 20 µL. Then, the reaction was incubated in the thermocycler (Techne Techgene FTGENE2D) at 50ºC for 15 minutes and proceeded to transform chemically competent cells

Chemically competent *E. coli* cells were thawed on ice, and 5 µL of DNA ligation reaction was incubated with 50 µL of cells for one hour on ice. The cells were heat-shocked at 42°C for one minute, followed by 30 minutes of incubation at room temperature while adding 50 µL of LB media. After the 30 minutes of incubation, cells were diluted 1:10 and 50 µL of cells were plated onto LB agar plates containing the suitable antibiotic.

The plasmid was isolated from the colonies as described in Section 2.1.3 from the isolation of a single colony. Glycerol stocks of the bacterial cells carrying representative plasmids were also produced by adding 1:3 parts of glycerol to bacterial culture to a final volume of 0.6 mL and stored at -80°C. Plasmid DNA was analysed by digestion

with restriction enzymes and DNA sequencing (Core Genomics Facility, The University of Sheffield; GENEWIZ, Azenta Life Sciences).

2.2 Site Directed Mutagenesis

Site directed mutagenesis was performed through polymerase chain reaction (PCR) to create truncated forms and full length *Cn*FEN with a cleavable His-tag. Mutagenesis reactions were performed using the Q5 High Fidelity polymerase (NEB). Reactions were set up according to the manufacturer's protocols using the thermocycler Techne Techgene FTGENE2D. Typically 20 ng of DNA template was used in a 50 µL reaction, with 0.4-0.5 µM of each oligonucleotide primer and 1 unit of polymerase. The annealing temperature for the PCR was dependent on the primers previously designed (Table 2.3) using the NEBase Tm calculator tool (https://tmcalculator.neb.com/). PCR products were analysed on a 1% agarose gel and purified using the Monarch PCR & DNA Cleanup kit (NEB). Only one round of PCR was necessary for the production of the *Cn*FEN full length and truncated constructs with cleavable His-tag and sub-cloned into pYMC547c (Section 2.1.4).

2.3 Protein overexpression and purification

2.3.1 Protein overexpression

The expression of the proteins of interest *Tb*FEN, *Li*FEN and *Cn*FEN were performed from the pET21a(+) vector (and pYM547c vector for *Cn*FEN cleavable His-tag constructs) in the *E. coli* BL21(DE3) strain using either Studier's auto-inducting method (Studier, 2005), adding isopropyl β-D-1-thiogalactopyranoside (IPTG) in culture cells grown in 2YT media (1.6 % (w/v) vegetable tryptone, 1% (w/v) yeast extract and 0.5% (w/v) NaCl) with an $A_{600\ nm}$ of 1 or using an optimized media formula developed by Dr Domen Zafred, a post-doctoral researcher in the Sayers laboratory.

Prior the small-scale protein inductions, a single colony was isolated and cultured in 3 mL of LB (1 % (w/v) vegetable tryptone, 0.5% (w/v) yeast extract and 0.5% (w/v) NaCl) or 2YT media with 100 µg/mL of carbenicillin, 0.3% (w/v) glucose and 0.25% (w/v) aspartate for 37°C overnight.

For the Studier method, 250 µL of the 3 mL LB culture was incubated in 10 mL on a media based on ZYM-5052 media, substituting NZ-amine with vegetable tryptone, (1% (w/v) vegetable tryptone, 0.5% yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol (v/v), 0.05% glucose (w/v), 0.2% α-lactose (w/v), 0.25% aspartate, 0.2x trace metals, achieved by adding 1:5000 (v/v) of 1000 x trace metals stock (1000 x trace metal concentration: 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$-$4H_2O$, 10 mM $ZnSO_4$-$7H_2O$, 2 mM $CoCl_2$-$6H_2O$, 2 mM $CuCl_2$-$2H_2O$, 2 mM $NiCl_2$-$6H_2O$, 2 mM Na_2MoO_4-$2H_2O$, 2 mM Na_2SeO_3-$5H_2O$, 2 mM H_3BO_3) containing 100 µg/mL of carbenicillin at 37°C overnight.

For the IPTG technique, 250 µL of the 3 mL LB culture was incubated in 10 mL of LB or 2YT media containing the suitable antibiotic (100 µg/mL of carbenicillin or 50 µg/mL of kanamycin), 0.3% (w/v) glucose and 0.25% (w/v) aspartate and incubated at 37°C until $A_{600\ nm}$ reach approximately 1. Induction was initiated by addition of 1 µL of 1 M IPTG stock solution to the culture followed by incubation at 37°C overnight.

For the optimized media formula, 250 µL of the 3 mL LB culture was incubated in 10 mL of modified 2YT media (2% (w/v) vegetable tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.3% K_2HPO_4, 0.18% KH_2PO_4, 2 mM $MgSO_4$ and 0.2 mM $CaCl_2$) with 100 µg/mL of carbenicillin, 0.05% (w/v) glucose, 0.6% (w/v) glycerol and 0.25% (w/v) aspartate and incubated at 37ºC. After 2 hours, the same amount of glucose and glycerol were added. After 1.5 hours, 0.2% α-lactose (w/v) was added and the cells were incubated at 16ºC overnight. The next day in the morning, an extra 0.2% α-lactose (w/v) was added and 4 hours after the cells were harvested.

A large-scale protein induction was performed using the three methods from the small-scale induction (Vegetable tryptone was also used in all large-scale protein inductions).

For the Studier method, *E. coli* BL21(DE3) cells were cultured in 3 mL of LB or 2YT media with 100 µg/mL of carbenicillin, 0.3% (w/v) glucose and 0.25% (w/v) aspartate for 37°C, 12 hours. Then 1 mL of the previous cultured was transferred to 100 mL of LB or 2YT media with the proper antibiotic, 0.3% (w/v) glucose and 0.25% aspartate and was incubated for 37°C, 12 hours. The cell culture was transferred into 4.5 L of auto-induction media (Formedium) with 100 µg/mL of carbenicillin and incubated in a bioreactor with a speed stirring of 400 rpm, temperature of 30°C (16ºC for *Cn*FEN constructs), overnight.

For the IPTG induction, *E. coli* BL21(DE3) cells were cultured in 3 mL of MDG media (25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.2 x trace metals, 0.5% glucose and 0.25% aspartate) with the appropriate antibiotic (100 µg/mL of carbenicillin or 50 µg/mL of kanamycin) for 37°C, 8 hours. Then 1 mL of the previous cultured was transferred to 100 mL of MDG media with the proper antibiotic and incubated at 30°C, overnight. From the overnight culture, 15 mL were transferred into each of the 6 plastic baffled flasks of 2.5 L containing 0.75 L of 2YT media with 0.3% (w/v) glucose, 0.25% (w/v) aspartate and the appropiate antibiotic (100 µg/mL of carbenicillin or 50 µg/mL of kanamycin) and incubated at 37ºC until $A_{600\ nm}$ reach approximately 1. Induction was initiated by addition of 75 µL of 1 M IPTG stock solution in each flask to the culture followed by incubation at 30°C (16ºC for *Cn*FEN constructs) overnight.

For the optimized media formula, *E. coli* BL21(DE3) cells were cultured in 3 mL of MDG media for 37°C, 8 hours. Then 1 mL of the previous cultured was transferred 100 mL of MDG media with the proper antibiotic and for 30°C, overnight. The cell culture was transferred into 6 plastic baffled flasks containing 0.75 L of modified 2YT media with 100 µg/mL of carbenicillin, 0.05% (w/v) glucose and 0.6% (w/v) glycerol and incubated at 37ºC. After 2 hours, the same amount of glucose and glycerol were added. After 1.5 hours, 0.2% α-lactose (w/v) was added and the cells were incubated at 16ºC overnight. The next day in the morning, an extra 0.2% α-lactose (w/v) was added and 4 hours after the cells were harvested.

From the 3 protein expression methods, cells were harvested by centrifugation at 4000 x *g* for 20 minutes at 10ºC. Cell pellets were subsequently stored at -80ºC.

2.3.2 Cell lysis

Frozen cell paste was resuspended in lysis buffer (20 mM Tris pH 8.0, 5 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl, 5% (v/v) glycerol) at a volume of 5 mL per gram of cells. A 100 µL sample was taken for SDS-PAGE. Lysozyme (Sigma) was added to a final concentration of 400 µg/mL and incubated at room temperature for one hour. Phenylmethanesulfonyl fluoride (PMSF), freshly prepared in ethanol at a concentration of 25 mg/mL and kept as stock solution, was added to a final concentration of 25 µg/mL and sodium deoxycholate at a concentration of 25 mg/mL in water was added to a final concentration of 500 µg/mL. Afterwards, the

suspension was stirred at 4ºC for thirty minutes, or until the suspension had turned viscous.

The cell suspension was sonicated at 10 amplitude microns power for thirty-second pulses in three cycles using a Jencons Ultrasonic Processor (GE-50) sonicator (the sonication sample was kept on ice all throughout). A 100 µL sample was taken for SDS-PAGE analysis. The sonicated cells were centrifugated at 75600 x *g* at 4°C for 20 minutes to remove the cellular debris. Samples from the supernatant and the pellet fractions were analysed using SDS-PAGE (from this point if the protein was expressed using the pYM547c system, the recovered supernatant will go directly to the purification using chromatography columns).

Ammonium sulphate to a concentration of 0.5 M was stirred into the supernatant. The supernatant was kept on ice while a 5% polyethylenimine (PEI) (w/v) precipitation was performed. A titration of 1 mL sample of the supernatant was carried out to determine the appropriate volume of 5% PEI to use for the prescipitation, adding 10 µL of 5% PEI, vortexing and centrifuging at 15000 x *g* at room temperature for one minute. This process was repeated until no further pellet was visible. The precipitation protocol was scaled up to the whole sample, which was incubated in a roller for 30 minutes at room temperature, then centrifuged at 1677 x *g* for 20 minutes at room temperature.

The supernatant was then removed and placed on ice, whilst the pellet was retained for SDS-PAGE. Ammonium sulphate was added to the supernatant up to a concentration of 3.5 M (taking into consideration the 0.5 M already dissolved) and the supernatant was stirred for 40 minutes at room temperature. The supernatant was centrifuged at 75600 x *g* at room temperature for 20 minutes to precipitate the proteins. The supernatant was decanted to a new tube and kept in the fridge for further analysis. The pellet was solubilised in 20-40 mL of the non-salt buffer that is planned to use for purification. The solubilised protein was dialysed using a dialysis membrane tubing previously washed with 200 mL of 1 mM EDTA and 4 g sodium hydrogen carbonate ($NaHCO_3$). The wash step was repeated twice more with 200mLof 1 mM EDTA only. The dialysis of the solubilised protein was performed against 20-fold excess of the non-salt buffer overnight at 4°C with gentle stirring.

2.3.3 Ion exchange and Affinity Chromatography

Purification was performed using a combination of cation exchange, anion exchange, immobilized metal affinity and DNA binding affinity chromatography. First, the sample was centrifuged at 75600 x *g* at 4ºC to separate the supernatant form the precipitate formed during the overnight. Afterwards, the recovered supernatant of the protein sample and buffers were filtered using a 0.2-micron filter to remove particles that might damage the column.

All the purifications were carried out at room temperature, keeping the sample to load in ice to prevent its denaturation. Phosphate buffers (25 mM K_2HPO_4/KH_2PO_4, 1 mM EDTA, 1 mM DTT, 5% glycerol and 0 mM NaCl for non-salt buffer and 1000 mM NaCl for salt buffer) were used for elution of heparin and sulphopropyl (SP) chromatography columns (5 mL columns from Cytiva and the 20 mL heparin column from GEHealthCare) at the required pH. Tris buffers (20 mM Tris, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0 mM NaCl for non-salt buffer and 1000 mM NaCl for salt buffer) were used for elution of quaternary ammonium (Q) chromatography 5 mL column (Cytiva) at the required pH. The buffers in heparin, SP and Q columns were prepared at the required pH according to the pI of the proteins. Tris buffers (40 mM Tris pH 8.0, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol and 20 mM imidazole as the low-imidazole buffer; 40 mM Tris pH 8.0, 1.5 M NaCl, 0.5 mM TCEP, 5% glycerol and 20 mM imidazole as the high-salt DNA elution buffer; 40 mM Tris pH 8.0, 500 mM NaCl, 0.5 mM TCEP, 5% glycerol and 500 mM imidazole as the high-imidazole buffer) were used for elution of Ni Sepharose (HisTrap) 5 mL column (Cytiva).

The protein was loaded onto the columns at a flow rate of 0.5 column volumes per minute for 5 mL columns and 0.125 column volumes per minute for 20 mL columns. Fractionation was performed using ÄKTA systems (ÄKTAprime PLUS or ÄKTA pure). The system was washed before attaching the column and eluted with low salt buffer for about 5x and 2.5x column volumes for the 5 mL and 20 mL columns, respectively.

For the cation and anion exchange and the DNA binding affinity chromatography, the protein was eluted from the column using a 0 – 1 M salt gradient over a 10x column volume, collecting 1 column volume fractions for the ÄKTAprime PLUS and 2 mL fractions for the ÄKTA pure at a flow rate of 1 column volumes per minute for the 5 mL

columns and 0.5 volume fractions for the ÄKTAprime PLUS and 2 mL fractions for the ÄKTA pure at a flow rate of 0.1 column volumes per minute for the 20 mL column.

For the immobilized metal affinity chromatography, the DNA was first eluted from the column using a 1.5 M high salt buffer over a 5x column volumes at a flow rate of 1 column volumes per minute. After that, the protein was eluted from the column using a 20 – 500 mM imidazole gradient over a 10x column volume, collecting 2 mL fractions at a flow rate of 1 column volumes per minute. Cleavage of the C-terminal His-tag from the *Cn*FEN constructs was performed by addition of HRV 3C protease to a final concentration of 0.3 mg of protease per mg of purified protein (HRV 3C protease stocks of 3 mg/mL) and the cleavage reaction was carried out overnight at 4°C or at room temperature for two hours. A second immobilized metal affinity chromatography was performed to purify the protein from the cleaved tag using the buffers with low and high imidazole concentration without gradient.

2.3.4 Size-exclusion Chromatography

For further purification of the protein, the Superdex 200 10/300 GL size exclusion chromatography was used. The sample from the last purification step was concentrated to a maximum volume of 1 mL using a centrifugal filtration device (Sartorius Vivaspin) at 4°C at 15000 x *g* with a molecular weight cut off of 10 kDa. The column was previously equilibrated overnight by loading 4x column volume with buffer containing 20 mM Tris, 150 mM NaCl, 1 mM DTT, 1 M EDTA at a rate of 0.5 mL/min. The concentrated sample was injected into the loop. Samples were eluted in 1 mL fractions using the ÄKTA Pure system.

2.3.5 Protein storage

Purified protein was concentrated to 1 mL or less using the 10 kDa cut off centrifugal filtration device (Sartorius Vivaspin) at 4°C at 15000 x g. The concentrated protein was filtered and aliquoted approximately 100 µL of concentrated protein using filters Whatman FP 13/0.2 RC-S (regenerated cellulose membrane – sterile) in 500 µL sterile Eppendorf tubes. Then, the tubes were flash-frozen in liquid nitrogen and stored at -80°C.

2.4 Electrophoresis Methods

2.4.1 Agarose Gel Electrophoresis

DNA was separated according to its size using agarose gel electrophoresis. Powdered agarose was dissolved in TAE Buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA pH 8.0) in 1% (w/v) by boiling. Then, 0.15 µL/mL of Midori Green Advance DNA Stain was added and set at room temperature in its mold with a sample-well comb. The gel was placed in the electrophoresis tank containing TAE buffer and electrophoresis was carried out at 6 V/cm. The DNA bands were visualise by using the Bio-Rad Gel Doc EZ Imager and/or the UV trans-illuminator.

2.4.2 SDS-PAGE Electrophoresis

SDS-PAGE gels were used to visualise the proteins within samples, separating them based on their size and to confirm the presence and purity of the protein of interest according to the molecular weight. The gels were prepared with SDS-PAGE Resolving Gel (100 mM Tris (pH 8.3), 100 mM Bicine, 10% (w/v) acrylamide (37.5:1 acrylamide:bisacrylamide), 0.1% SDS, 0.5 mg/mL ammonium persulfate (APS) and 0.15% (v/v) tetramethylethylenediamine (TEMED)) and SDS-PAGE Stacking Gel (125 mM Tris (pH 6.9), 10% (w/v) acrylamide (37.5:1 acrylamide:bisacrylamide), 0.1% SDS, 0.5 mg/mL APS and 0.003% (v/v) TEMED). The samples were prepared by using 1:1 SDS-PAGE loading dye (1% sodium dodecyl sulphate (SDS), 10% glycerol, 62.5 mM Tris (pH 6.9), 0.025% (w/v) bromophenol blue and 1 mM DTT) and incubated at 90°C for 5 minutes. The samples and protein ladder (Bio-Rad Precision Plus 1610373) were loaded in the wells of the stacking gel and run with SDS-PAGE Running Buffer (50 mM Tris (pH 8.3), 50 mM Bicine and 0.1% SDS.) at 10 V/cm until the dye front has reached the bottom of the gel. The SDS-PAGE gel was stained using a combination of Brilliant Blue G and Coomasie Blue R250 stain (40% (v/v) methanol, 10% (v/v) acetic acid, 2 mg/mL Coomassie Blue R250, 2 mg/mL Brilliant Blue G), followed by a de-staining solution (40% (v/v) methanol, 10% (v/v) acetic acid). The gels were visualized using the Bio-Rad Gel Doc EZ Imager.

2.4.3 DNA Substrate Gel Electrophoresis (Zymogram)

This technique allows visualisation of proteins proteins with nuclease activity using a zymogram. The gel was prepared in the same way as the standard SDS-PAGE gel

(Section 2.4.2), but 800 µg of Type XIV Herring sperm DNA was added to the resolving gel. The samples were prepared and run in the same way as the SDS-PAGE technique. The gel was washed three times with 50 mL TBG buffer (100 mM Tris (pH 8.0), 100 mM Bicine, 10% (v/v) glycerol) for 15 minutes each time. The gel was incubated overnight at room temperature in the Reaction Buffer (100 mM Tris (pH 8.0), 100 mM Bicine, 10% (v/v) glycerol, 50 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, 100 mM KCl). Afterwards, the gel was washed with 50 mL TBG buffer and, later on, incubated with 40 mL of TBG buffer and 6 µL of Midori Green Advance DNA Stain for 30 minutes. The DNA substrate gel was visualised with the Bio-Rad Gel Doc EZ Imager.

2.4.4 Western Blot

Western blot membranes were used to detect a specific protein from a protein separation on a SDS-PAGE gel using a horizontal transfer system. First, a SDS-PAGE was performed as usual (Section 2.4.2). Then, the gel was washed for 15 minutes in Transfer Buffer (0.58% (w/v) Tris-base, 0.2% (w/v) glycine, 20% (v/v) methanol). Nitrocellulose membrane (Amersham Protran 0.45 NC) and blotting paper were soaked in Transfer Buffer for 5 minutes. The SDS-PAGE gel and the nitrocellulose membrane were setup in the horizontal transfer system, squeezing excess buffer out of the blotting paper (from top to bottom of the system: blotting paper, SDS-PAGE gel, nitrocelullose membrane, blotting paper), and transfer was done at 10 V for 1 hour. The membrane was blocked for 1 hour at room temperature using PBS-T (1x PBS, 0.05% (v/v) TWEEN-20) plus 5% (w/v) milk powder. The membrane was washed with PBS-T for 30 minutes and probed for 1 hour with the primary antibody made up in PBS-T plus 1% (w/v) milk powder. The membrane was washed again with PBS-T for 30 minutes and probed for 1 hour with the secondary antibody (horseradish peroxidase (HRP) linked) made up in PBS-T plus 1% (w/v) milk powder. The membrane was washed one last time with PBS-T for 30 minutes and developed bands were revealed by washing the membrane with Pierce 1-step Ultra TMB Blotting solution (ThermoFisher) for 1 minute or until bands appear. The membrane was washed with water to stop the reaction and allowed to dry.

2.5 Antibody production

Antibodies against *Cn*FEN protein were raised in rabbits by a contract research organisation, BioServUK (https://bioservuk.com). One rabbit was injected subcutaneously with 300 µg of purified protein in complete Freund's adjuvant and 5 immunisations were given in total over a 63-day schedule. Test sera was gathered on day 35, with final sera collection on day 63.

2.6 Spectroscopy Methods

2.6.1 UV Nuclease Assay

This technique is able to measure the activity of nucleases digesting large DNA strands and thereby releasing low molecular weight products which do not precipitate with acid. Reactions containing 25 mM HEPES-NaOH (pH 7.5) or 25 mM potassium glycinate (pH 9.3), 10 mM $MgCl_2$ or 0.75 mM $MnCl_2$, 100 mM KCl, 1 mM DTT and 667 µg/mL of herring sperm type XIV DNA were mixed well and incubated at 37°C for ten minutes. The recombinant FEN proteins were added (12.5 µg of *Tb*FEN per assay and 10 µg of *Li*FEN and *Cn*FEN per assays) and incubated at 37°C. Aliquots of 100 µL were taken over a time course of 3, 6, 12, 24 and 60 minutes) and added into tubes with equal volumes of 6% $HClO_4$ previously chilled in ice. These were incubated on ice for ten minutes and centrifuged at 14000 x *g* for 5 minutes. From the supernatant, a volume of 150 µL was taken from the reactions, De-ionised water was added to each sample up to 1 mL and the absorbance was measured at 260 nm in a 1 cm path length quartz cuvette. Increase in the absorbance against time was plotted and fitting by linear regression was used to calculate enzyme activity.

2.6.2 Förster Resonance Energy Transfer Cleavage Assay

2.6.2.1 Optimizing FRET Assay for *Li*FEN, *Tb*FEN and *Cn*FEN

Purified protein sample of *Li*FEN, *Tb*FEN and *Cn*FEN were diluted in FRET reaction buffer (25 mM HEPES, 0.5 mM EDTA, 100 mM NaCl, 10 mM $MgCl_2$ or 5 mM $MnCl_2$, 100 mM KCl, 2 mM DTT, 0.1 mg/mL BSA, pH 7) to four concentrations, 40, 4, 0.4 and 0.04 µg/mL. A 96-well plate was set up containing 8 negative controls (74 µL reaction buffer, 1 µL DMSO), and 3 repeats of each protein sample at the four different concentrations (25 µL of FEN, 49 µL reaction buffer, 1 µL DMSO) giving a final

concentration of 10, 1, 0.1 and 0.01 µg/mL. After 10-minute incubation at room temperature, 25 µL of 0.8uM A3InvD substrate was added to all wells. Fluorescence was read at 495nm absorbance/520 nm emission at every 30 seconds during 10 minutes using VarioSkan Flash instrument (Thermo Scientific, N06354). Data were analysed using Microsoft Excel.

Z' value for each assay was calculated with the equation below.

$$Z' = 1 - \frac{3(\sigma_{+RFU} + \sigma_{-RFU})}{|\mu_{+RFU} - \mu_{-RFU}|}$$

2.6.2.2 Screening 1166 compound library BioNet with FRET assay

A 96-well plate was set up containing 8 negative controls (74 µL reaction buffer, 1 µL DMSO), 8 positive controls (25 µL of protein at the concentration determined in Section 2.6.2.1, 49 µL reaction buffer, 1 µL DMSO), and 80 compounds at 1 mM (25 µL of FEN, 49 µL reaction buffer, 1 µL BioNet 100 mM compound). The assay was executed in a similar way as Section 2.6.2.1. The library is distributed in 15 plates and the assay was executed for each plate and each protein.

The RFU measurements were used to calculate the % inhibition, using the following equation.

$$\%\ inhibition = 100 - \frac{RFU - \mu_{-RFU}}{\mu_{+RFU} - \mu_{-RFU}} * 100$$

2.6.2.3 Selective screening of the 400 compound library HitExpansion with FRET assay

According from the results obtained from the molecular docking, a selective screening of the HitExpansion library was performerd setting up a 96-well plate containing 8 negative controls (74 µL reaction buffer, 1 µL DMSO), 8 positive controls (25 µL of protein at the concentration determined in Section 2.6.2.1, 49 µL reaction buffer, 1 µL DMSO), and 80 compounds at 1 mM (25 µL of FEN, 49 µL reaction buffer, 1 µL HitExpansion 100 mM compound). The assay was executed in a similar way as Section 2.6.2.1. The subset of compounds selected is distributed in 2 plates and the assay was executed for each plate and each protein.

2.6.2.4 Half-maximal inhibitory concentration (IC_{50}) of the potential hits obtained from BioNet and HitExpansion libraries

Potential hits selected from the screening of the 2 libraries were further analysed through IC_{50} assays. A 96-well plate was set up containing 8 negative controls (74 µL reaction buffer, 1 µL DMSO), 8 positive controls (25 µL of protein at the concentration determined in Section 2.6.2.1, 49 µL reaction buffer, 1 µL DMSO) and a 1:3 serial dilution throughout the column of the compound at 1mM (8 dilutions in total) (25 µL of FEN, 49 µL reaction buffer, 1 µL BioNet or HitExpansion compound with a concentration from 100 mM to 15 µM). Each compound was tested in triplicate, allowing to analyse up to 3 compounds per plate. Therefore, the BioNet hits were distributed in 8 plates while the HitExpansion hits were distributed in 5 plates and the assay was performed for each plate and each protein. The IC_{50} were calculated by non-linear regression in GraphPad Prism 9.4.1.

2.6.3 Mass Spectrometry Analysis

Purified protein from chromatography techniques (protocol Sections 2.3.3 and 2.3.4) was taken for mass spectrometry analysis. The protein concentration of the sample was measured with NanoDrop spectrophotometer (NanoDrop Technologies, ND-1000) by taking the absorbance read at 280 nm. The protein sample was analysed by the Mass Spectrometry Center of the Department of Chemistry, University of Sheffield.

2.7 Crystallisation Trials

2.7.1 Initial Screening crystallisation trial

Protein samples for crystallisation assays were prepared in low salt Tris buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT). Six commercial crystallisation screens were used: JCSG, PACT, Morpheus, Proplex (Molecular Dimensions), MPD (NeXtal Biotechnologie), and Natrix (Hampton Research). Initial crystallisation screens were set up at 50 µL of buffer in the reservoir of the 96 well crystallisation plates.

The protein sample at a concentration of approximately 12 mg/mL was seeded to each condition using Mosquito Crystal robot (TTP Labtech) with and without different DNA substrates (1:1 protein:DNA ratio (0.25 mM) using 5ov4 or 3ov6 and JT DNA substrates; Table 2.4). Protein or protein-DNA and buffer were pipetted on the wells

using 150 nL of the sample on a ratio of 1:1 protein-buffer. After seeding, the plates were covered with polyolefin film and stored at 17°C. A microscope was used to observe for crystal formations on the plates.

2.7.2 Optimizing crystallisation conditions

Conditions that have formed microcrystals were optimised by changing the buffer pH, salt or precipitant concentrations in small increments. The optimisation trials were performed in a 96-well plate considering the protocol from Section 2.7.1. For the optimisation trials, the mother liquor from the initial screening in which the conditions obtained crystals were used at a ratio of 30 nL of 150 nL of protein at the appropriate concentration with 120 nL of buffer (reservoir). After seeding, the plates were covered with polyolefin film and stored at 17°C. A microscope was used to observe for crystal formations on the plates.

2.8 Structure determination and refinement

X-ray diffraction data was collected from the crystals sent to the Diamond Light Source. Structures were determined by molecular replacement using the CCP4 suite programme (Winn *et al.*, 2011) with the previously determined 2.20 Å structure of the human FEN in complex with Sm^{3+} (PDB entry 3Q8K) as a model (Tsutakawa *et al.*, 2011). The molecules were found in the asymmetric unit and subsequent rounds of fitting and refinement were completed using COOT (Emsley *et al.,* 2010) and REFMAC5 (Murshudov *et al.,* 2011). Structures were validated within COOT and by using MOLPROBITY (Chen *et al.,* 2010). Identification of metal ions in all structures was performed on the basis of an analysis of coordination geometry, ion environment, bond lengths and refined B-factors and CheckMyMetal web server (Zheng *et al.,* 2014).

2.9 Virtual screening of microbial FENs and low-molecular weight fragment libraries

Structure-based virtual screening applying molecular docking simulations was performed using the PyRx interface (version 0.9.8 for Mac OS; Dallakyan & Olson, 2015) which streamLines access to Vina (Trott & Olson, 2010) and Autodock (Morris *et al.*, 2009).

2.9.1 Preparation of low-molecular weight molecules for docking

The structures of the BioNet library compounds were obtained from the manufacturer Key Organics Ltd (https://www.keyorganics.net/bionet-products/fragment-libraries/) in a SDF file. The SDF file was converted into PDBQT files using the graphical user interface (GUI) programme iBabel version 5.0 (from Open Babel version 3.1.1; O'Boyle *et al.,* 2011), adding hydrogens at the appropriate pH of 7.4, centring the coordinates, generating 3D molecules and removing salts. The HitExpansion library compounds are a set of 400 compounds from different manufacturers: Enamine (https://enamine.net), Chembridge (https://chembridge.com/screening-compounds/fragments/), Vitas (https://vitasmLab.biz) and Sigma (https://www.sigmaaldrich.com/GB/en/product/sial/dyna001), which are compounds that have a higher binding activity to flap endonucleases than the previous two libraries. The HitExpansion library was provided in a SMILES code by the postdoctoral researcher Dr. Srdjan Vitovski. The SMILES code for the compounds was converted into a single SDF file using the software Osiris DataWarrior (version 5.5.0; Sander *et al.*, 2015). Furthermore, the SDF file was converted into PDBQT files using the GUI programme iBabel version 5.0 (from Open Babel version 3.1.1), adding hydrogens at the appropriate pH of 7.4, centring the coordinates, generating 3D molecules and removing salts.

2.9.2 Preparation of FEN proteins as macromolecules

The FEN proteins used for the *in silico* studies are the TbFEN (previously crystallized by former postdoctoral researcher Jason Wilson), LiFEN (the partial crystallized structure obtained throughout this project) and CnFEN (predicted structure built using the tool Phyre2; Kelley *et al.*, 2015). The metal ions that interacts with the active site of the flap endonuclease were extracted from the crystallized human FEN (PDB ID: 3Q8K; Tsutakawa *et al.*, 2011). Samarium ions (Sm^{3+}) from the crystallized structure were changed to Mg^{2+} and added to the parasitic flap endonucleases (extraction and addition of the metal ions were done using PyMol and the change of samarium to magnesium ions as well as adding the charge were carried out using a text editor). The *in silico* studies were performed without the metal ions, with both metal ions or with individual metal ions (Magnesium 1 and 2).

2.9.3 Molecular docking process

Molecular docking was perform using the programmes AutoDock Vina version 1.2.0 (Trott & Olson, 2010) and AutoDock 4 (Morris *et al.*, 2009), both of them inside PyRx. The studies utilized a specific grid center and grid size for each protein, with the following dimensions in Å (The grid size for Vina was 2.5 times smaller with the purpose to resemble the same grid size used for AutoDock, which are the dimensions mentioned below):

- TbFEN: centre (x, y, z) = (-24.575, -6.378, -7.416), dimensions (x, y, z) = (57, 68, 86).
- LiFEN: centre (x, y, z) = (-37.135, -1.836, 27.491), dimensions (x, y, z) = (68, 60, 81).
- CnFEN: centre (x, y, z) = (-7.918, -0.772, 99.372), dimensions (x, y, z) = (81, 57, 67).

The AutoDock Vina programme was run according to the default parameters (exhaustiveness of 8 and number of modes of 9). The AutoDock 4 programme was run using the Lamarckian Generic Algorithm according to the default parameters.

2.9.4 Result Analyses

After the molecular docking was complete, the results were exported as a SDF file and visualized into Osiris DataWarrior (version 5.5.0), which were used to determine the compounds with the highest binding energy (those with the lowest Gibbs free energy), considering a cut off of -7 kcal/mol. In addition, the results were utilized to compare the binding energy predicted with the programmes AutoDock Vina and AutoDock 4 and the inhibition percentage obtained from the FRET assays (Section 2.6.2.2 for the BioNet library and Section 2.6.2.3 for the HitExpansion library). Finally, the selected hits from both programmes and both libraries were imported into PyMol to be visually inspected.

Table 2.3 Oligonucleotides used in all sub-cloning, site directed mutagenesis and sequencing protocols

Name	Sequence (5' – 3')	Technique
C-tag_for	CTTTAAGAAGGAGATATACCATGGGCATCAAAGGTCTGACC	Site directed mutagenesis
C-tag_453_rev	AGCACTTCTAAGCCCGAGCTGTTCTTGCTTTTCTTTTTCGCGCTAC	Site directed mutagenesis
C-tag_414_rev	AGCACTTCTAAGCCCGAGCTAAAGAAGCCGTCCAGACGAC	Site directed mutagenesis
C-tag_405_rev	AGCACTTCTAAGCCCGAGCTCTTCGCCGCCAGCATCTT	Site directed mutagenesis
T7_for	TAATACGACTCACTATAGGG	Sequencing
T7_term	GCTAGTTATTGCTCAGCGG	Sequencing

Table 2.4 Oligonucleotides used in analytical techniques

Name	Sequence (5' – 3')	Modification	Technique
Assay 3	TTTTCGCTGTCTCGCTGAGT	5' CY3, 3' FLU	FRET
Assay2Inv	ACTCAGCGAGACAGCGCCGGAACACACGCTGCGTGTGTTCCGGT	NA	FRET
5ov4	AAAAGCGTACGC	NA	Crystallization
3ov6	GATCTATATGCCATCGG	NA	Crystallization
JT1	ACCGTCC	NA	Crystallization
JT2+1	ATTGAGGCAGAGT	NA	Crystallization
JT3	ACTCTGCCTCAAGACGGT	NA	Crystallization

CHAPTER 3: EXPRESSION AND PURIFICATION OF WILD-TYPE MICROBIAL FLAP ENDONUCLEASES

3.1 Introduction

This chapter covers the over-expression and purification of flap endonucleases from relevant organisms (*T. brucei*, *L. infantum* and *C. neoformans*). This work was executed with the objective of obtaining pure and biochemically functional proteins for the crystallization assays and compound library screening in order to progress identification of inhibitors.

Enzymatically active proteins are needed for the FRET-based inhibitor screening assay and can also be used to attempt structure determination to facilitate this early-stage rational drug design project. Protein crystals can also be used for fragment-based screening assuming suitable conditions for soaking of small molecules can be obtained (Biggins *et al.*, 2000; Hartshorn *et al.*, 2005).

Furthermore, a catalytically inert *T. brucei* FEN, which was previously produced in the Sayers laboratory, was expressed and purified for crystallography. The aim is to obtain a crystal structure of a FEN protein with its DNA substrate bound to it without cleaving, in the presence of the divalent metal ions. This approach replaces a conserved aspartic acid in the active site with a lysine residue. The idea being to target a conserved amino acid that binds one of the active site magnesium ions. AlMalki and co-workers used this approach to replace one of two such residues in bacteriophage T5 D15 exonuclease (a flap endonuclease homologue). Replacement of either Asp153 (5HNK.pdb; Figure 3.1E) or Asp155 (5HP4.pdb; Figure 3.1F) with lysine resulted in proteins that were isostructural with wild type enzyme (1EXN.pdb; Figure 3.1A-D; Ceska *et al.,* 1996). Furthermore, the proteins are able to bind DNA but not cut it as the epsilon-amino group of the introduced lysine occupies a position that would otherwise host a magnesium ion. This allowed crystallization and successful structure determination of the resulting ternary complexes consisting of “inactive-mutant enzyme”, substrate and metal ions (AlMalki *et al*. 2016).

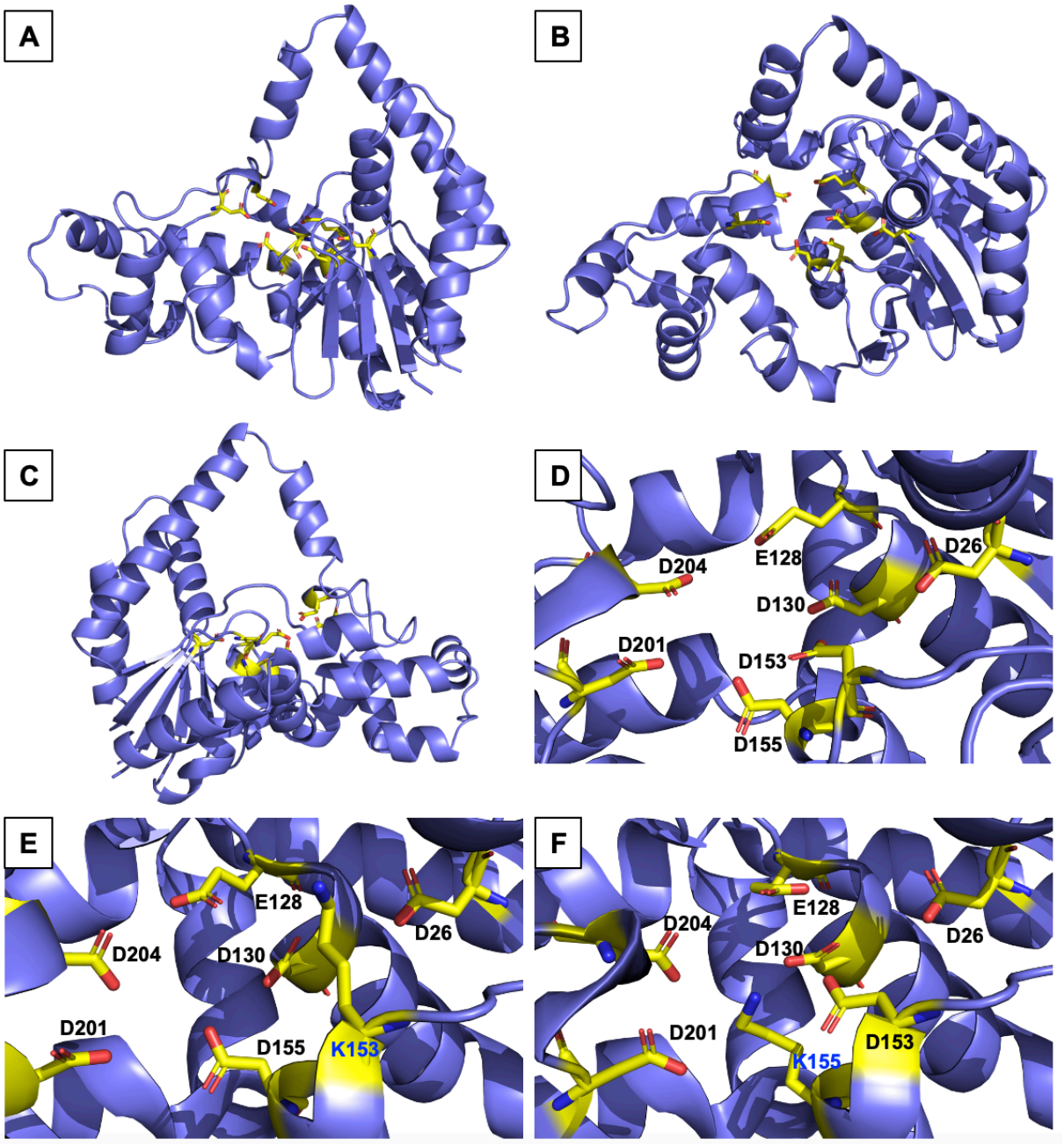


Figure 3.1 Crystal structures of T5 D15 exonuclease WT and isostructural proteins

(A – C) *Three orthogonal views of T5 D15 exonuclease WT (1EXN.pdb).* ***(D)*** *Shows the active site of T5 D15 exonuclease WT.* ***(E)*** *&* ***(F)*** *Depict the active site of the protein with the introduced mutations Asp153 and Asp155 to lysine, respectively (5HNK.pdb & 5HP4.pdb). The protein is coloured in lavender and residues involved in the active site are shown as yellow sticks. Residue mutation in Asp153 and Asp155 with lysine is highlighted in blue text in panels E (1EXN conformers A and B superimpose on 5HNK chains C and D with an RMSD in the range of 0.5 to 0.83 Å) and F (1EXN chains superimpose on 5HP4 with an RMSD in the range of 0.55 to 0.58 Å).*

At the same time, this approach was also conducted in the human FEN according to the enzymatic assays of Shen *et al.,* 1996 and the structural studies of Tsutakawa *et al.,* 2011. According to both studies, targeting one of the conserved amino acids from the active site led to DNA binding to the protein but not cleavage. Replacement of Asp181 (3Q8M.pdb; Figure 3.2D) with alanine resulted in a protein with an identical structural conformation as the wild type enzyme (3Q8L.pdb; Figure 3.2A-C; Tsutakawa *et al.,* 2011).

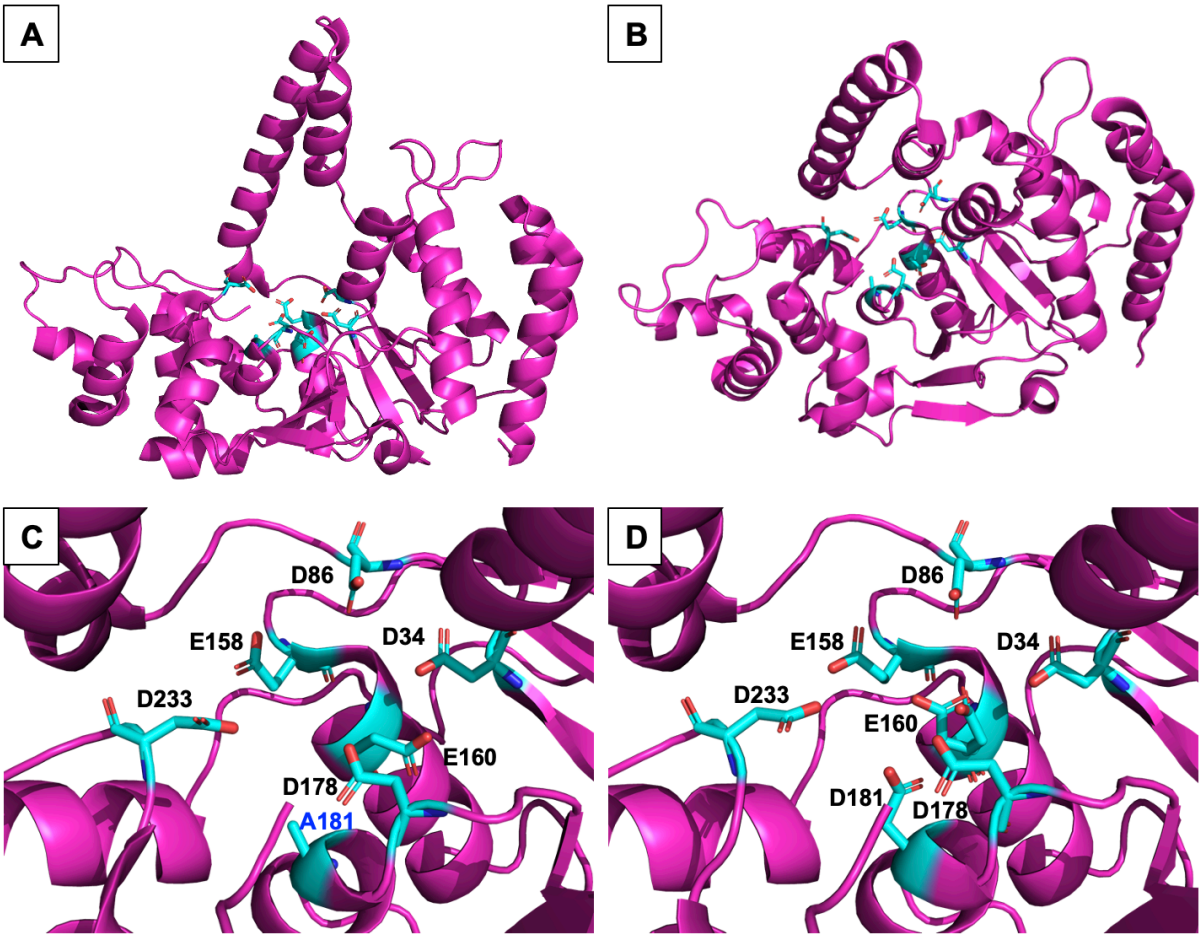


Figure 3.2 Crystal structures of human FEN WT and isostructural protein

(A) *&* ***(B)*** *Two orthogonal views of human FEN with the introduced mutation Asp181 to lysine (3Q8M.pdb).* ***(C)*** *Shows the active site of 3Q8M.pdb structure.* ***(D)*** *Depicts the active site of human FEN WT (3Q8L.pdb). The protein is coloured in magenta and residues involved in the active site are shown as cyan sticks. Residue mutation in Asp181 with alanine is highlighted in blue text in panel C (Human FEN WT superimposes on conformers A and B of human FEN D181A with an RMSD of 0.54 Å).*

Taking into consideration the analytical approach of the conserved residues in T5 D15 exonuclease and human FEN, Oates (2016) conducted a similar study with the *T. brucei* FEN. During her research, seven DNA constructs of *Tb*FEN with a single amino acid mutation on the active site were produced, but only the D183K variant was analysed in detail. A crystal structure of this variant was obtained in the Sayers' laboratory (data provided by Dr. Jason Wilson; Figure 3.3A-C). However, a structure of the wild-type *Tb*FEN has not been elucidated at the moment, so a predictive structure of this protein was built using the online tool Phyre2 (Kelley *et al.,* 2015) to compare the conformation of the residues within the active site (Figure 3.3D).

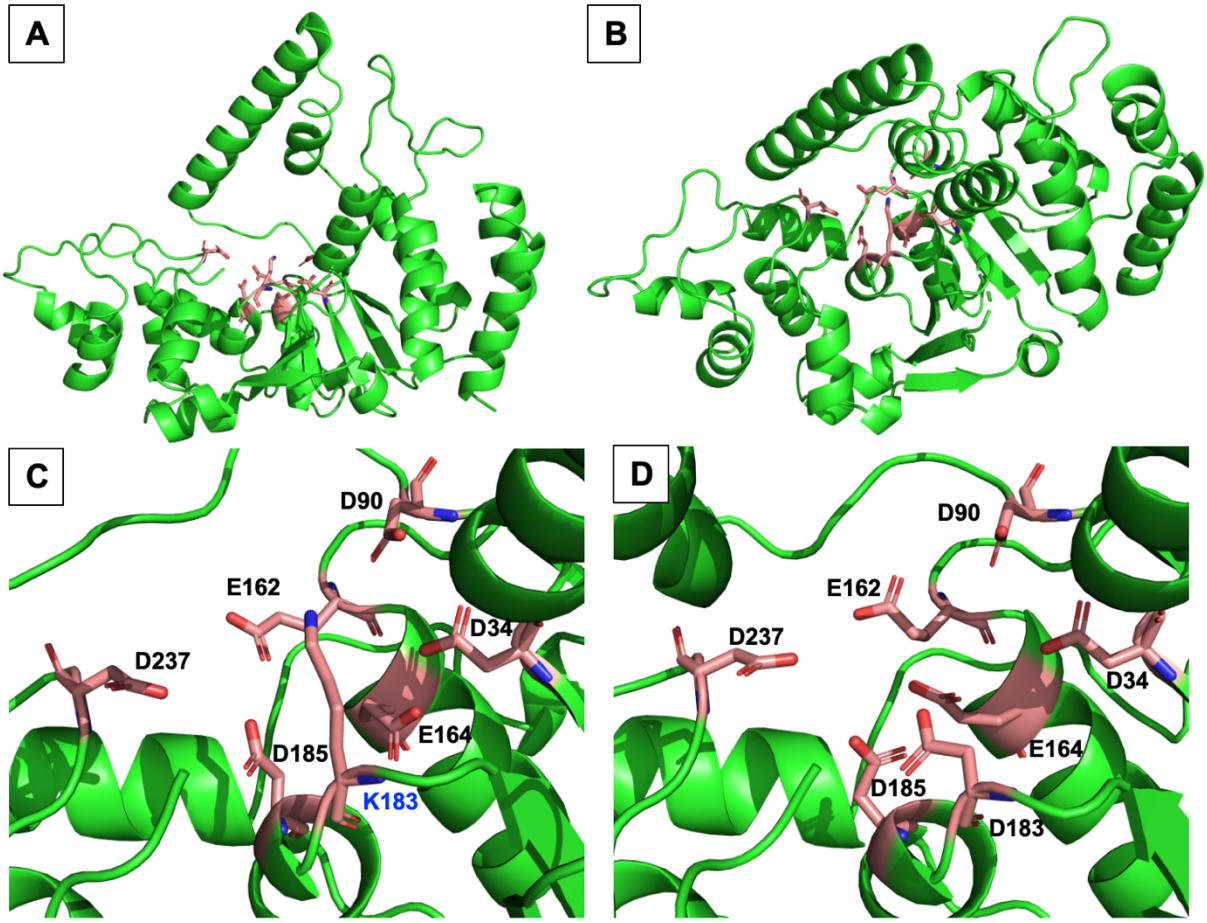


Figure 3.3 Crystal structure and structural prediction of *T. brucei* FEN

(A) *&* ***(B)*** *Two orthogonal views of* Tb*FEN with the introduced mutation Asp183 to lysine.* ***(C)*** *Shows the active site of the* Tb*FEN D183K variant.* ***(D)*** *Depicts the active site of the predicted* Tb*FEN FEN WT. The protein is coloured in green and residues involved in the active site are shown as pink sticks. Residue mutation in Asp183 with lysine is highlighted in blue text in panel C (*Tb*FEN D183K superimposes on the predicted* Tb*FEN WT with an RMSD of 0.894 Å).*

Expression of proteins of interest can be performed in a great variety of systems from prokaryotic and eukaryotic hosts, such as mammalian cells, bacteria, yeast, insect cells and transgenic plants (Tripathi & Shrivastava, 2019). From the systems already mentioned, bacteria (specifically *E. coli*) is one of the most popular systems for production of recombinant proteins in academia and pharmaceutical industry (Lozano Terol *et al.,* 2021). The reason why *E. coli* is widely used for protein expression is due to the various advantages this system possesses over other systems. Those advantages are, for example, the cell biology and genetic of this organism is well understood, easy to handle, simple and inexpensive fermentation process and large production of recombinant protein (Gupta & Shukla, 2015).

However, there is a possibility that the gene does not express efficiently when the expression host is switched over (Gupta & Shukla, 2015), which can cause problems when using *E. coli* as a host for the production of recombinant proteins. The reason behind this relies on the difference in the codon usage between the *E. coli* and eukaryotic genomes (Correddu *et al.,* 2020). Eukaryotic genes can contain codons that occur less frequently than bacterial genes, representing a challenge in the production of recombinant proteins. Some of the problems a researcher can encounter when expressing a protein of interest in *E. coli* are low production activity, loss of enzymatic activity, protein misfolding, aggregation, formation of inclusion bodies (water-insoluble protein aggregates formed in the bacterial cytoplasm), inefficient protein translocation, and metabolic burden (Kane, 1995; Kapust *et al.,* 2002; Malygin *et al.,* 2003; Sarilla *et al.,* 2010; Mairhofer *et al.,* 2013, Baig *et al.,* 2014; Marschall *et al.,* 2017).

During translation, the ribosome may experience translational stalling due to a lack of tRNAs that recognise uncommon codons. (Farabaugh, 2000). A list of rare codons was described by Kane (1995) from a compilation of several studies of heterologous protein expression in *E. coli*. The rare codons previously reported and enlisted in this study are: AGG, AGA and CGA coding for Arg; CUA coding for Leu; AUA coding for Ile; CCC coding for Pro; and GGA and GGG coding for Gly.

Nevertheless, in most of the cases there are strategies to overcome these problems. Due to the advances in synthetic biology and the declining cost of gene synthesis, it is now possible to create, edit and modify synthetic genes for the production of

recombinant proteins (Endy, 2005; Heinemann & Panke, 2006; Leonard *et al.*, 2008; McDaniel & Weiss, 2005). With this approach, is possible to edit the gene of interest to be codon optimised for the *E. coli* translation machinery, increasing the protein expression several fold (Elena *et al.,* 2014). Apart from codon optimising the gene of interest, there are regulatory elements that can be added when synthesizing the gene that can significantly improve the protein expression, including promoters, ribosomal binding sites (RBSs), transcriptional terminators, etc. (Shiue & Prather, 2012; Boyle & Silver, 2012; Meng *et al.*, 2013).

In the case of over-expression of nucleases, this can be problematic as they have the potential to interfere with host cell metabolism and degrade genomic DNA. Thus, it is essential to use a tightly regulated, inducible expression system. Suitable systems that have been reported in the literature include bacteriophage lambda leftward promoter (P_L) in conjunction with a temperature sensitive repressor, cl857 (Remaut *et al.*, 1981) and an antisense repression driven by convergent *lac* promotor (P_{lac}) (Alpers & Tomkins, 1965). Cells are grown at low temperature (under 30ºC) in the presence of IPTG (~0.1 μM) to induce antisense RNA production. Any leakage from the lambda promotor is thus counteracted by the complementary RNA produced from the opposing P_{lac}. Such systems have been used for flap endonucleases (Sayers & Eckstein, 1990) and restriction enzyme EcoRI (O'Connor & Timmis, 1987). Alternative systems such as the bacteriophage T7 promotor in conjunction with a lactose-inducible T7 RNA polymerase have also been successfully used for expression of flap endonucleases. For example, Tsutakawa and collaborators produced the human homologue FEN1 in recombinant *E. coli* using this system (Tsutakawa, 2011).

The T7 promoter system was used for this work using the pET21a(+) and pYM547c plasmids in BL21(DE3) (Studier & Mofatt, 1986; Wu *et al.*, 2018). This bacterial strain contains a chromosomal copy of the T7 RNA polymerase (T7RNAP) controlled by *lac* UV5 promoter (Studier *et al.*, 1990). The pET21a(+) has a beta-lactamase gene while the pYM547c has a aminoglycoside phosphotransferase gene (selection markers) conferring resistance to the antibiotic ampicillin (AmpR) and kanamycin (KanR), respectively. A lactose operon repressor (*lacI*), a T7 promoter sequence upstream of the multiple cloning site (MCS) into which the gene encoding the protein of interest is inserted, and a T7 terminator sequence downstream of the MCS.

In the absence of an inducer of the *lac* promoter, no T7 RNA polymerase is made. This phenomenon happens because the *E. coli* RNA polymerase does not recognize the T7 promotor on the plasmid. The presence of a T7 promoter does not allow the gene expression in the cell in the absence of T7 RNA polymerase. However, in BL21(DE3) addition of IPTG or lactose leads to the production of T7 RNAP (Studier & Moffat, 1986). An alternative to the IPTG induction for over-expression of recombinant protein is the Studier method. This method, described by Studier (2005), uses a defined media with carbon substrate mixtures of glucose, glycerol, and lactose. From this mixture, glucose is the preferred carbon substrate for the cells. Cells grow to high cell density but when the glucose is used up, the cells switch to lactose, which is an inducer of the *lac* promotor system. Glycerol improves biomass yields. Cell cultures with high densities grown in auto-induction media tend to produce more protein per volume of culture than IPTG induction, proving that this methodology is a convenient and inexpensive form of protein production in any scale (Studier, 2005).

Even when the gene has been codon optimised and regulatory elements have been added, it is possible that one of the scenarios during the production of heterologous proteins in a *E. coli* system is a poor or low expression as well as the formation of inclusion bodies (Gupta & Shukla, 2015). To address the issues previously mentioned, a variety of protein tags have been developed with the purpose of reducing aggregation and formation of inclusion bodies and producing protein in soluble form (Yang *et al.,* 2018). During cloning design, these tags are fused to the N- or C-terminal ends of the protein of interest (Eschenfeldt *et al.,* 2010). Numerous protein tags (His6, GST, MBP, NusA, Trx, and SUMO) have the potential to be employed as fusion tags in *E. coli* for the expression of soluble, active proteins. The His-tag (containing six or more consecutive histidine residues) is a popular tag for effective affinity purification of fused proteins. However, the problem of producing protein with fusion tags relies on the cleavage and removal of these tags at different stages of the purification process. This can be solved by engineering several protease cleavage sites for optimal protein production and cleavage (Gupta & Shukla, 2015; Mahmoudi Gomari *et al.,* 2020).

The over-expression and purification of microbial flap endonucleases was performed considering all the previous points to achieve a large quantity of pure protein for further enzymatic and structural analyses. Production of wild-type *Tb*FEN, *Li*FEN and *Cn*FEN;

a *Tb*FEN D183K truncated form and *Cn*FEN truncated forms were performed in this project (Figure 3.4; Appendix Table 9.2). Production of truncated forms of flap endonucleases was performed due to the conformational domains of the proteins, aiming for soluble protein with a high probability of crystallisation (Cooper *et al.,* 2017).

In terms of solubility and crystallisation potential, it is usually assumed that well-ordered or compact domains will provide proteins which are easier to produce and handle than full-length for protein expression and structural analysis (Gopal & Kumar, 2013). More ordered proteins, for example, are more likely to crystallise compared to flexible or highly disordered proteins (Dale *et al.,* 2003; Sagemark *et al.,* 2010). Flexibility presented either between domains in multi-domain proteins, or from a region in a single-domain protein (e.g., unstructured N- or C- termini or internal loops), creates a potential entropic impediment during the crystallisation process (Derewenda, 2010).

Therefore, in this project it was deemed desirable to produce truncated forms of microbial FENs to improve chances of success. For *Tb*FEN D183K, the truncation was executed before the PCNA binding region, corresponding to a protein of 341 amino acids. For the *Cn*FEN, the truncations were carried out before and after the PCNA binding region, corresponding to a protein of 405 and 414 amino acids respectively. However, production of truncated forms and a full-length *Cn*FEN were performed with a cleavable double His-tag. After expression of these constructs, the fusion proteins were purified using a Ni-NTA affinity purification method, and the His-tag was removed in a subsequent stage using the human rhinovirus (HRV) 3C protease. This protease cleaves a specific amino acid sequence (Leu-Glu-Val-Leu-Phe-Gln ↓ Gly-Pro). For these three constructs, once the purification was completed, they exhibited 9 extra amino acids on the C-terminal of the proteins, part of the already cleaved double His-tag (Ser-Ser-Gly-Leu-Glu-Val-Leu-Phe-Gln; Figure 3.4; Appendix Table 9.2).

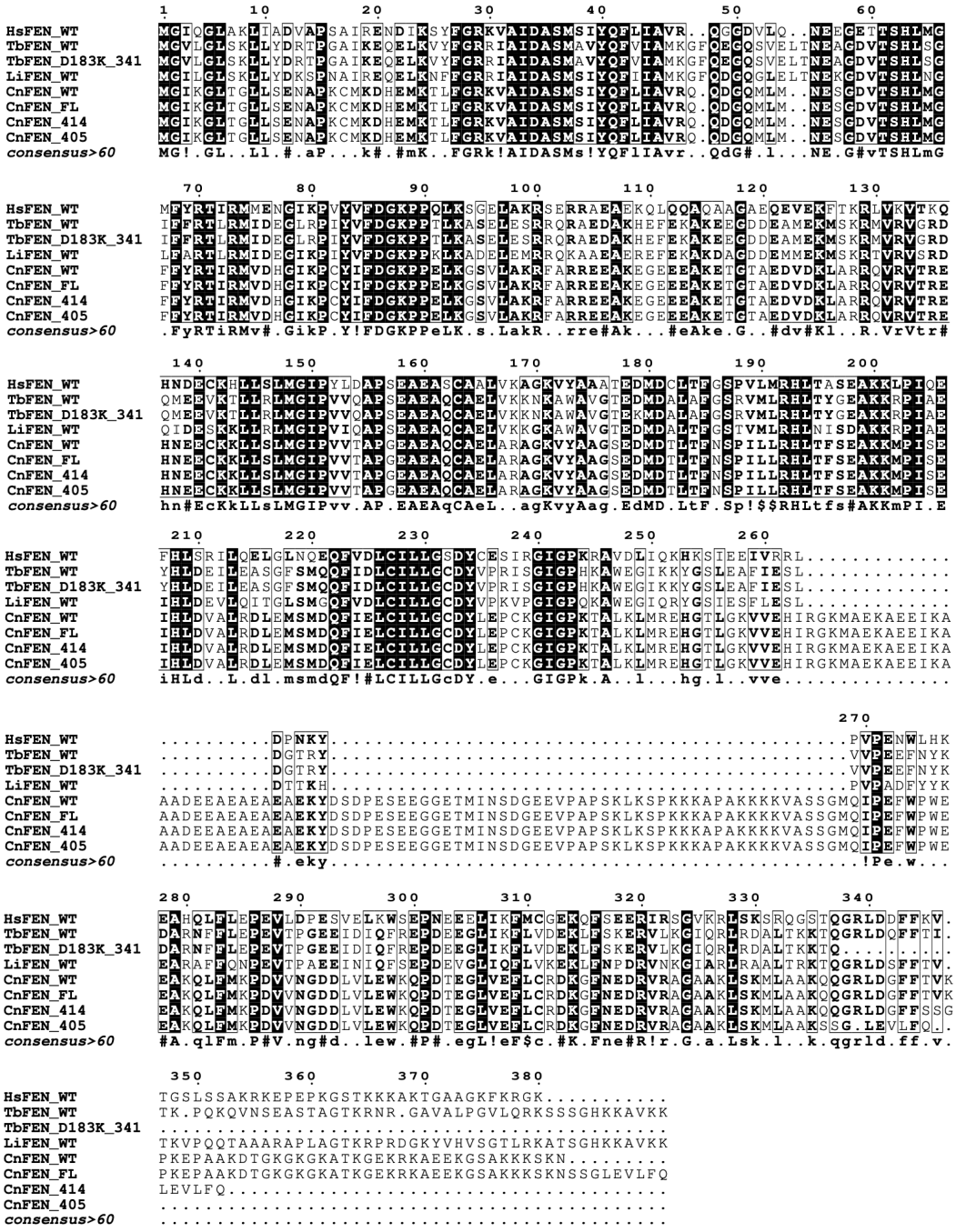


Figure 3.4 Multiple sequence alignment of human and parasitic FEN constructs expressed and purified in this project.

Hs*FEN,* Tb*FEN,* Li*FEN and* Cn*FEN corresponds to human,* T. brucei, L. infantum *and* C. neoformans *FENs, respectively. Identical are boxed in black while similar residues are boxed in white according to a 60% low level consensus. Consensus symbols: ! is anyone of Ile or Val; $ is anyone of Leu or Met; % is anyone of Phe or Tyr; # is anyone of Asn, Asp, Gln, Glu, Asx or Glx. Multiple sequence alignment of the proteins was performed using the online tool MultAlin (Corpet, 1988). Depiction of sequence alignment was carried out with the online tool ESPript 3.0 (Robert & Gouet, 2014).*

3.2 Over-expression and purification of *Tb*FEN proteins

3.2.1 Construct production of *Tb*FEN-WT in pET21a(+)

A codon-optimised gene fragment that encodes the *Trypanosoma* FEN WT, inserted in the pUC19 vector, was obtained from a commercial supplier (Eurofins). The plasmid pUC19 with the *Tb*FEN fragment and the empty pET21a(+) were digested using the restriction enzymes NdeI and HindIII (according to Section 2.1.4). After digestion, the samples were run on a 1% agarose gel to confirm the separation of the *Tb*FEN fragment from the pUC19 plasmid and to verify the size of the *Tb*FEN fragment and the plasmids, pUC19 and pET21a(+) (Figure 3.5).

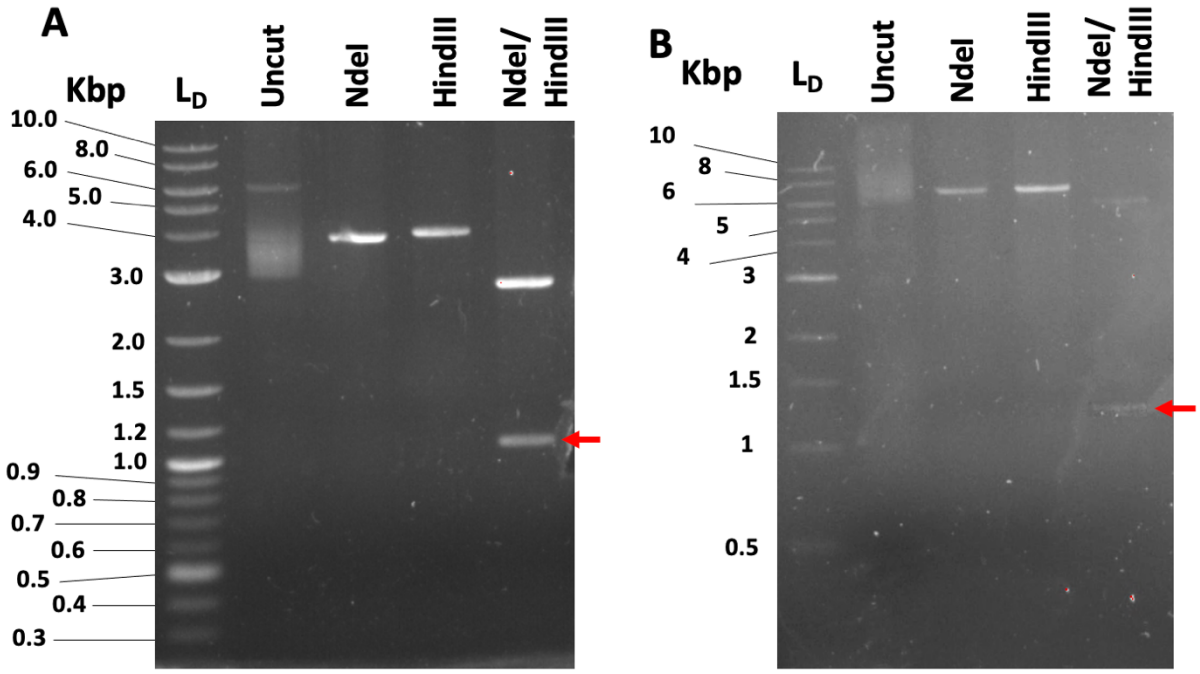


Figure 3.5 Agarose gel electrophoresis of the sub-cloning process to insert the *Tb*FEN fragment into the pET21a(+) vector

***(A)**: Digestion of pUC19-*Tb*FEN construct with indicated restriction enzymes. L_D corresponds to 2-log DNA ladder (NEB). **(B)**: Digestion with restriction enzymes after sub-cloning in pET21a(+)-*Tb*FEN construct. L_D corresponds to1 kbp DNA ladder (NEB). Red arrows point to the DNA fragment coding for *Tb*FEN (1182 bp). Images taken using the Bio-Rad Gel Doc EZ Imager.*

3.2.2 Construct production of *Tb*FEN-D183K full length and *Tb*FEN-D183K truncated in pET21a(+)

The production of constructs *Tb*FEN-D183K full length (FL) and *Tb*FEN-D183K truncated (ΔC) in pET21a(+) vector was previously performed in Sayers laboratory. A codon optimised *Tb*FEN-D183K ΔC inserted in pUC19 vector, was obtained from a commercial supplier (Eurofins). *Tb*FEN-D183K (FL) was produced from *Tb*FEN WT and *Tb*FEN-D183K ΔC through site-directed mutagenesis. The plasmid pET21a(+) coding for the genes *Tb*FEN-D183K FL and *Tb*FEN-D183K ΔC were digested using the restriction enzymes NdeI and BamHI (according to Section 2.1.4). After digestion, the samples were run on a 1% agarose gel to confirm the separation and the size of the *Tb*FEN fragments (Figure 3.6).

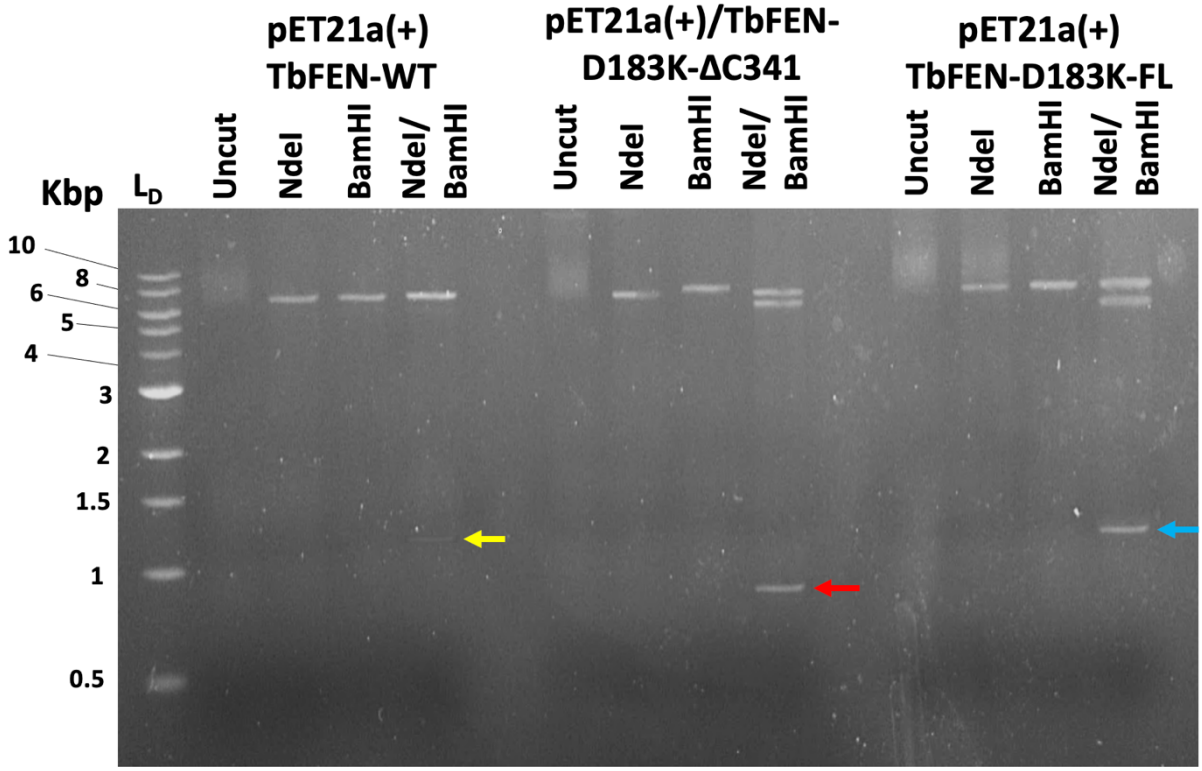


Figure 3.6 Agarose gel electrophoresis of *Tb*FEN-D183K FL and *Tb*FEN-D183K ΔC inserted in pET21a(+) vector

Annotations: L_D - 1kbp DNA ladder (NEB). Yellow arrow points to the DNA fragment coding for Tb*FEN WT (1182 bp); Red arrow points to the DNA fragment coding for* Tb*FEN D183K truncated (1026 bp); Blue arrow points to the DNA fragment coding for* Tb*FEN D183K full length (1182 bp). Image taken using the Bio-Rad Gel Doc EZ Imager.*

3.2.3 Over-expression of *Tb*FEN-WT

Over-expression of the protein was first analysed in a small-scale culture comparing two methods for the induction and expression of the recombinant protein: IPTG and Studier method (protocol in Section 2.3.1). SDS-PAGE analysis of this comparison is shown in Figure 3.7 A, which demonstrates that the over-expression of the protein can be executed by either of these two methods.

A large-scale cell culture for the over-expression of the protein was made only with IPTG (protocol section 2.3.1). A large-scale over-expression was carried out with IPTG as it gave similar results compared to the auto-induction method as this proved to be more convienient in my hands. Full induction with up to 40 μM IPTG start only after the depletion of glucose (Faust *et al.,* 2015). The protein over-expression was monitored by taking samples at the time 0, 0.5, 4, 8, 14 and 20 hours. An SDS-PAGE gel was run with the sample collected over the time course previously mentioned (Figure 3.7 B).

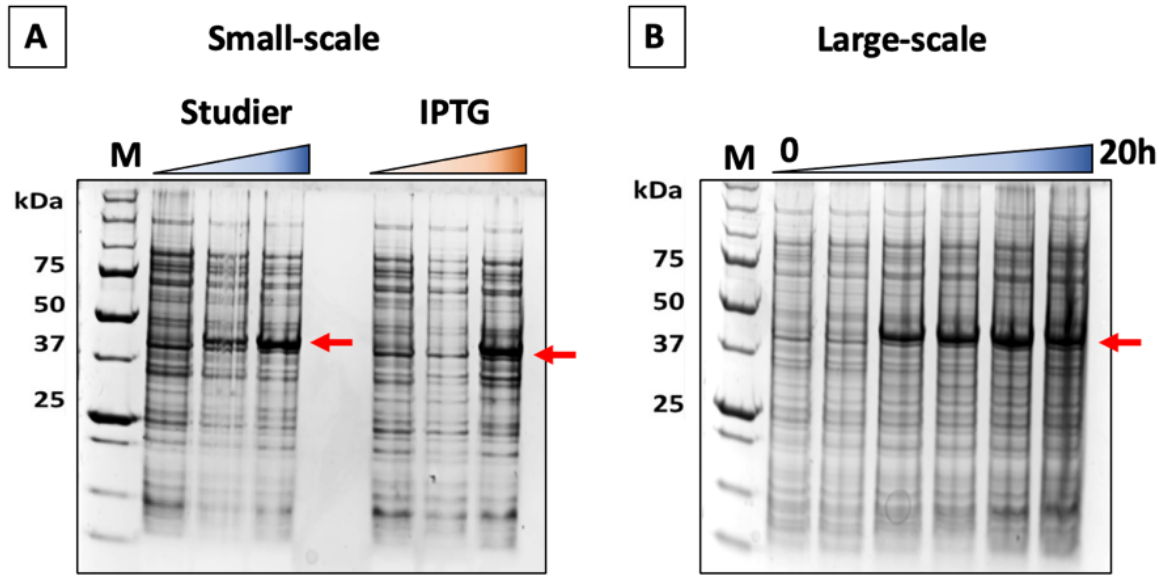


Figure 3.7 SDS-PAGE analysis of the over-expression of *Tb*FEN WT

***(A)**: Comparison between the two methods of protein induction and over-expression. **(B)**: Large scale over-expression of* Tb*FEN. Protein marker Bio-Rad Precision Plus 1610373 (M). Protein over-expression time in ascending order. Red arrows point expression of* Tb*FEN (44 kDa).*

3.2.4 Purification of *Tb*FEN-WT protein

*Tb*FEN WT protein was obtained from a re-suspended and lysed BL21(DE3) cell pellet expressing the protein from pET21a(+) vector (Section 2.3.2). The whole-cell lysis process, from sonication, PEI precipitation, and dialysis, is shown in Figure 3.8.

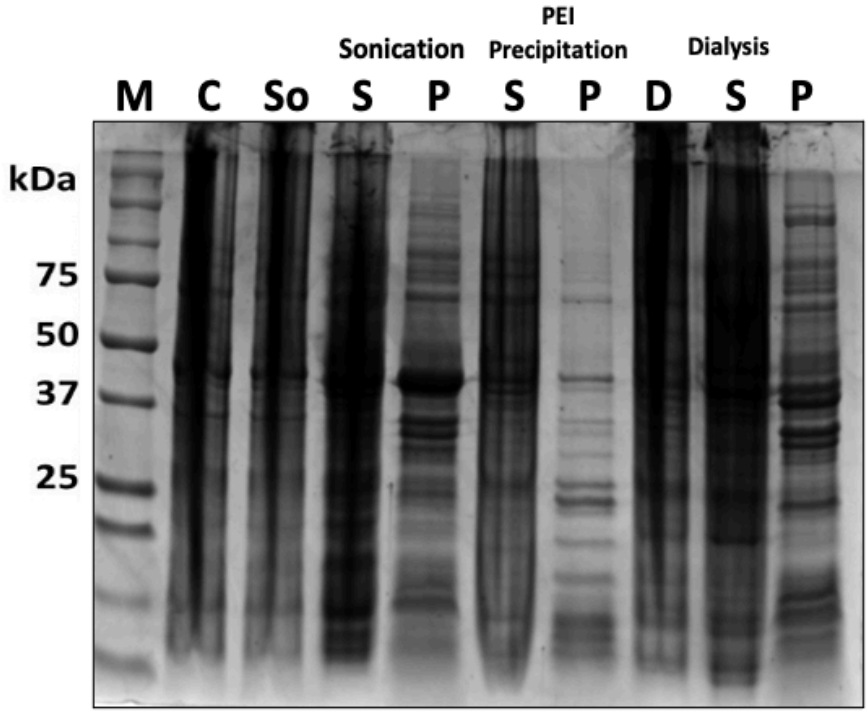


Figure 3.8 Cell lysis for the purification of *Tb*FEN WT

Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Cell pellet (C). Cell lysis sample after sonication (So). Dialysed protein sample (D). Supernatant (S). Pellet (P).

Purification of the *Tb*FEN protein was done with the supernatant obtained from the last step of the cell lysis process, after dialyzing the sample. The purification started with the SP column followed by the heparin column using phosphate buffe r at a pH of 6.7 (Figure 3.9). At this pH, the protein of interest was retained in both columns because proteins negatively charged binds to the SP column according to the estimated pI of the *Tb*FEN which is 7.79 and heparin column is designed to purify DNA-binding proteins. Estimation of the *Tb*FEN pI was conducted through the sequence analysis made in ExPASy ProtParam (Gasteiger *et al.*, 2005).

The fractions with the protein band around 44 kDa, which corresponds to the *Tb*FEN, were used for the next step of purification, which consisted of passing the dialysed protein into the Q column at pH 7 with Tris buffer (Figure 3.9). The last step of purification was done with size exclusion chromatography (SEC). The matrix of the gel filtration column allows smaller molecules to migrate into pores in the resin and their elution is delayed out while the larger molecules elute earlier due to their size stopping them from entering the pores in the resin beads. The SDS-PAGE analysis of the fractions took during SEC purification is presented in Figure 3.9.

To determine the nuclease activity of the purified protein, a zymogram was carried out to confirm the protein is active. The zymogram is a standard SDS-PAGE gel which also contains high molecular weight DNA (Type XIV, Sigma). After electrophoresis, the proteins are renatured and reaction buffer is added (Section 2.3.4). Nuclease activity is revealed by MIDORI Green Advance, a safe alternative to the traditional ethidium bromide for DNA staining. Darkened areas in the stained DNA-substrate gel demonstrate nuclease activity, and a duplicate SDS-PAGE gel containing the same samples was prepared as a reference. The protein ladder Bio-Rad Precision Plus 1610373 was used as negative control and *Pseudomonas aeruginosa* FEN was used as positive control on the zymogram assay (Figure 3.10 A and B).

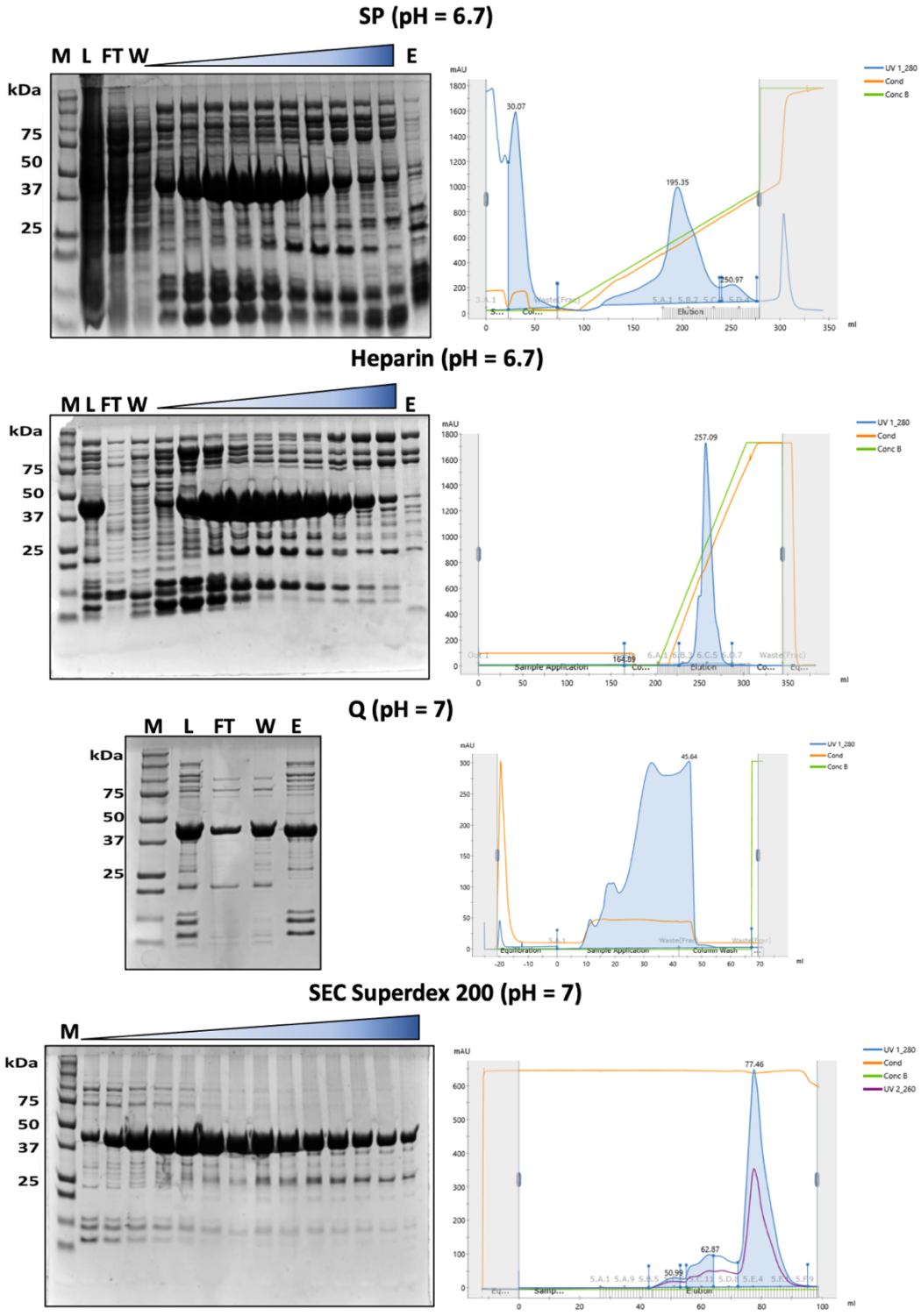


Figure 3.9 Purification of wild type *Tb*FEN

Panels on the right side represent SDS-PAGE gels of the purification of Tb*FEN. Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Load (L). Flow-through (FT). Wash (W). Elution (E). Salt concentration gradient (SP and heparin) and elution time (SEC) in ascending order (white → blue). Purification was carried out using the chromatography system ÄKTA pure™.*

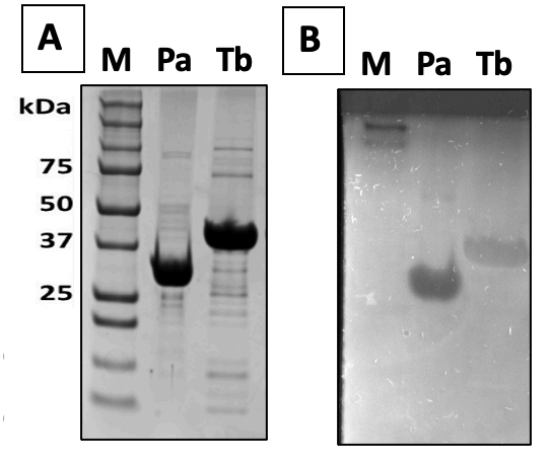


Figure 3.10 Determination of purity and nuclease activity of *Tb*FEN

(A) *&* ***(B)****: Purified* Tb*FEN concentrated and analysed by SDS-PAGE and zymogram, respectively. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Pseudomonas aeruginosa FEN (Pa). Purified* Tb*FEN (Tb).*

3.2.5 Purification of *Tb*FEN-D183K ΔC-341 protein

Cell lysis of BL21(DE3) expressing *Tb*FEN-D183K ΔC-341 protein was carried out on the same basis as the *Tb*FEN WT, considering the molecular weight of this protein of 39 kDa and pI of 5.38 according to ExPASy ProtParam (Section 2.3.2). The whole-cell lysis process, from sonication, PEI precipitation, and dialysis, is shown in Figure 3.11. Purification process was also performed following the same protocol as *Tb*FEN WT. Despite the fact that the pI of this protein is nearly 2.5 points below the pI of the WT *Tb*FEN, the purification protocol was similar to the wild-type with the exception of performing a salt gradient in the Q column. Purification was accomplished through heparin column, followed by Q column (with salt gradient) and finally size-exclusion chromatography (Figure 3.12). A final SDS-PAGE was performed to verify the purity of the protein and nuclease activity was determined through a zymogram (Figure 3.13).

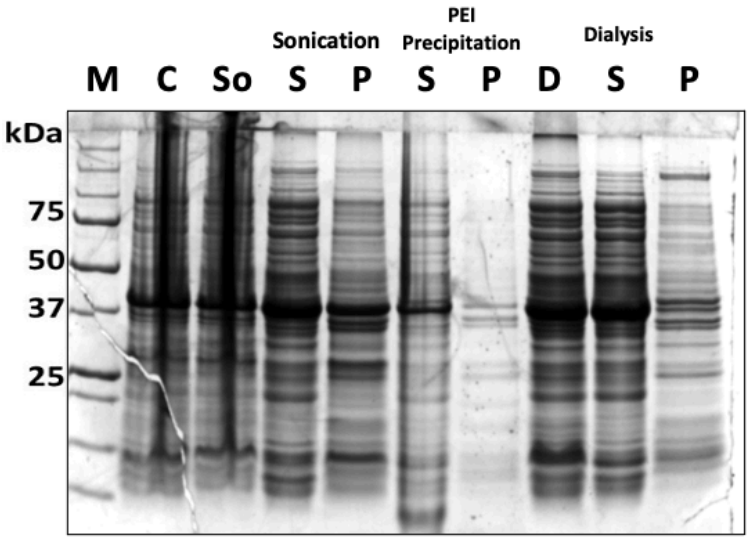


Figure 3.11 Cell lysis for the purification of *Tb*FEN D183K ΔC-341

Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Cell pellet (C). Cell lysis sample after sonication (So). Dialysed protein sample (D). Supernatant (S). Pellet (P).

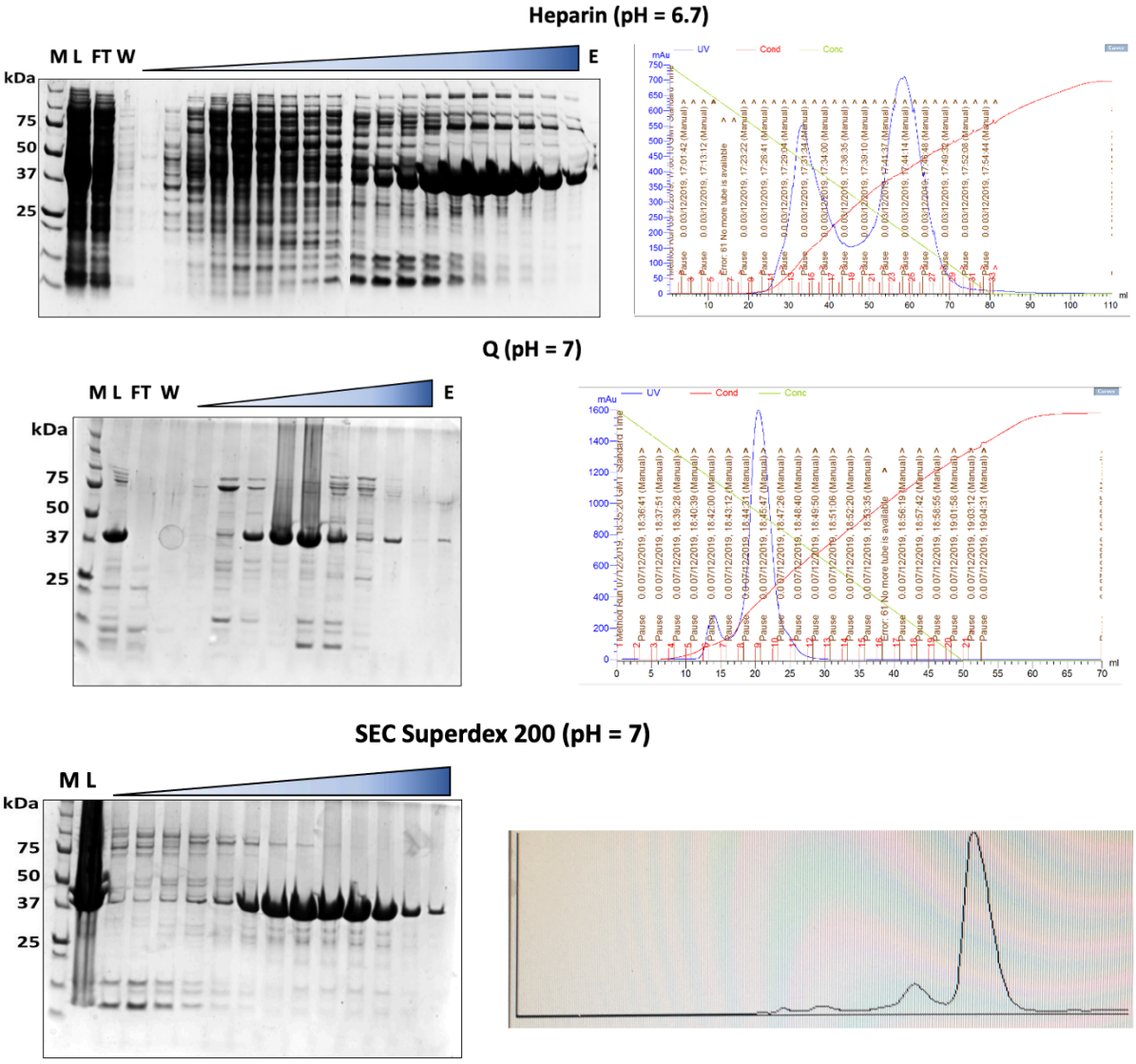


Figure 3.12 Purification of *Tb*FEN D183K ΔC-341

Panels on the right side represent SDS-PAGE gels of the purification of Tb*FEN D183K ΔC-341. Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Load (L). Flow-through (FT). Wash (W). Elution (E). Salt concentration gradient (heparin and Q) and elution time (SEC) in ascending order (white → blue). Purification was carried out using the chromatography system ÄKTAprime plus™.*

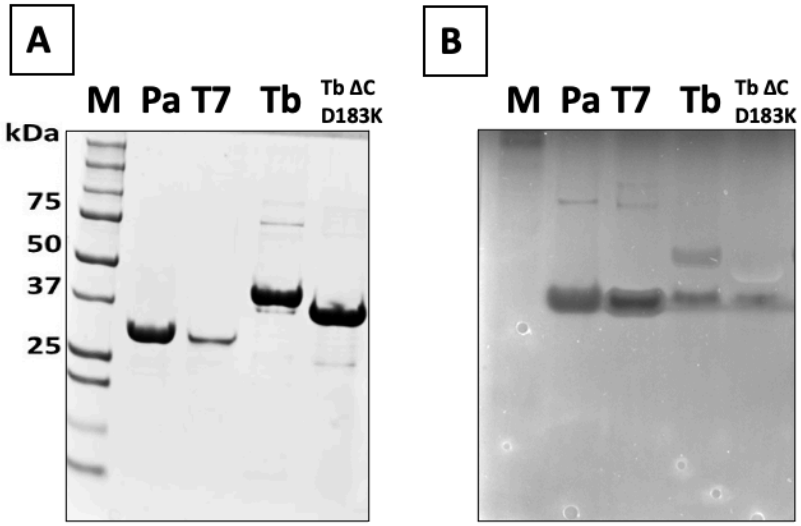


Figure 3.13 Determination of purity and nuclease activity of *Tb*FEN D183K ΔC-341

(A)* & *(B): *Purified* Tb*FEN D183K ΔC-341 concentrated and analysed by SDS-PAGE and zymogram, respectively. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Pseudomonas aeruginosa FEN (Pa). Bacteriophage T7 FEN (T7). Purified* Tb*FEN-WT (Tb). Purified* Tb*FEN D183K ΔC-341 (Tb ΔC D183K).*

In Figure 3.13, the presence of darkened bands in the lanes Tb and Tb ΔC D183K at the same molecular weight as T7 FEN is due to the activity leakage of this protein throughout the gel.

3.3 Over-expression and purification of *Li*FEN protein

3.3.1 Construct production of *Li*FEN-WT in pET21a(+)

A codon-optimised gene fragment that encodes the *Leishmania* FEN WT, inserted in the pEX-K4 plasmid, was obtained from a commercial supplier (Eurofins). The protocol to insert the *Li*FEN-WT fragment into the pET21a(+) plasmid was the same as the production of *Tb*FEN-WT in pET21a(+) (section 2.1.4; Figure 3.14).

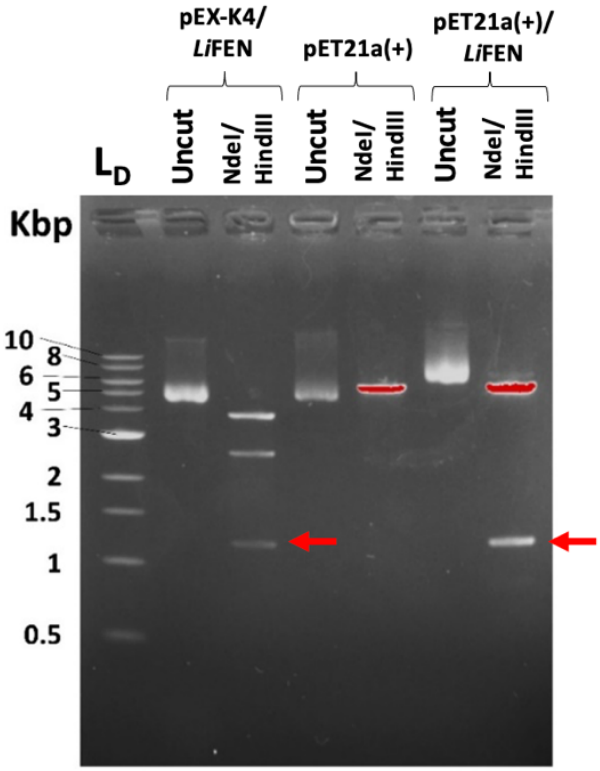


Figure 3.14 Agarose gel electrophoresis of the sub-cloning of *Li*FEN gene into the pET21a(+) expression vector

Annotations: L_D - 1 Kbp DNA ladder (NEB). Red arrows point to the DNA fragment coding for Li*FEN (1188 bp). Red bands correspond to overexposure of the gel in the Bio-Rad Gel Doc EZ Imager.*

3.3.2 Over-expression of *Li*FEN-WT protein

Over-expression of the *Li*FEN protein was performed in a small-scale culture through IPTG induction and Studier method, shown in Figure 3.15 A (protocol in section 2.3.1). A large-scale cell culture for the over-expression of the protein was made only with the Studier method (protocol section 2.3.1). For this occasion, the Studier method was selected for the large-scale cell culture to compare the expression with the large-scale *Tb*FEN expression. According to Studier (2005), this medium was calibrated to balance the carbon source of the cells by using glucose as a repressor and start the induction with natural lactose once glucose is depleted. In addition, this method possesses advantages over the IPTG culture. For example, (i) no need to monitor cell culture for optimal cell density induction, (ii) does not present technical issues for small scale, (iii) does not present toxicity limitations, (iv) inexpensive compared to IPTG (Briand *et al.,* 2015). The cell culture was left in incubation overnight and the protein over-expression was monitored by taking samples at the time 0, 3, 6 hours and overnight. A small sample was taken from the large-scale and proceeded to lyse the cells to test the solubility of the protein. First the cell pellet was resuspended in lysis buffer. Afterwards,

the cell suspension was sonicated, centrifuged and recovered the supernatant and pellet for SDS-PAGE analysis (Figure 3.15 B).

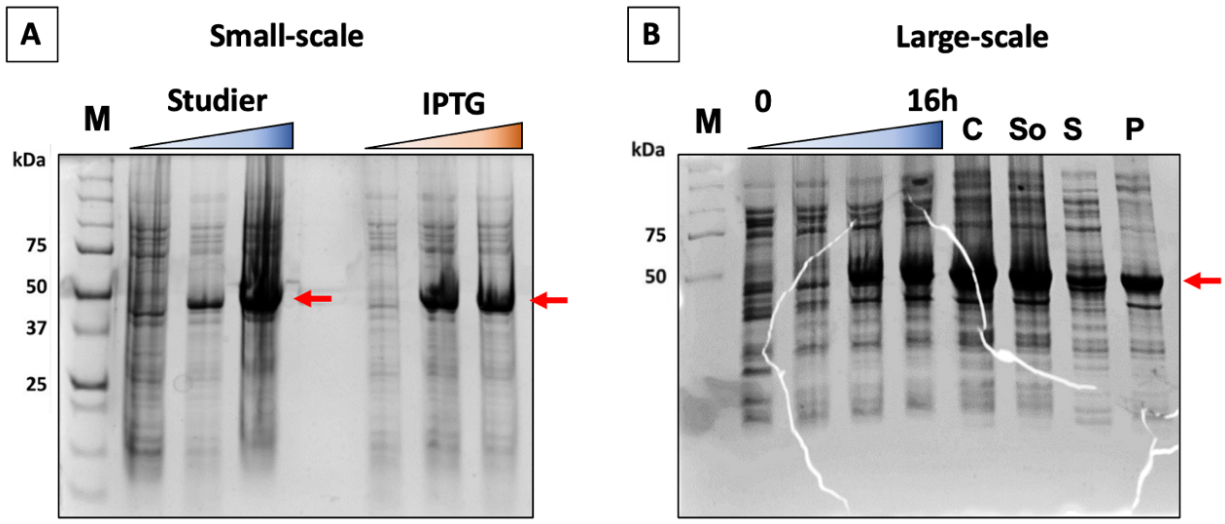


Figure 3.15 SDS-PAGE analysis of the over-expression of the recombinant WT *Li*FEN

***(A)**: Comparison between the two methods of protein induction and over-expression. **(B)**: Large scale over-expression of* Li*FEN. Protein marker Bio-Rad Precision Plus 1610373 (M). Protein over-expression time in ascending order. Cell suspension (C). Cell suspension after sonication (So). Supernatant (S). Pellet (P). Red arrows point expression of* Li*FEN (44 kDa).*

3.3.3 Purification of *Li*FEN-WT protein

First, BL21(DE3) cells expressing *Li*FEN protein were lysed according to the protocol on section 2.3.2, (Figure 3.16) and the supernatant was recovered. The isoelectric point of *Li*FEN is 9.12 and its molecular weight is 44 kDa according to ExPASy ProtParam. Considering that the pI of *Li*FEN is higher than the *Tb*FEN, the purification protocol for this protein was designed by firstly recovering the protein using a heparin column at pH 7.6 with phosphate buffer, followed by removal of the impurities with the Q column at pH 8 with Tris buffer. Using a buffer with the pH below the pI would make the protein pass through the column while many impurities would be removed. Finally, further purification and polishing was carried out with a size-exclusion chromatography column at pH 8 (Figure 3.17). SDS-PAGE and zymogram analysis were performed to determine the purity and nuclease activity of the protein (Figure 3.18 A & B).

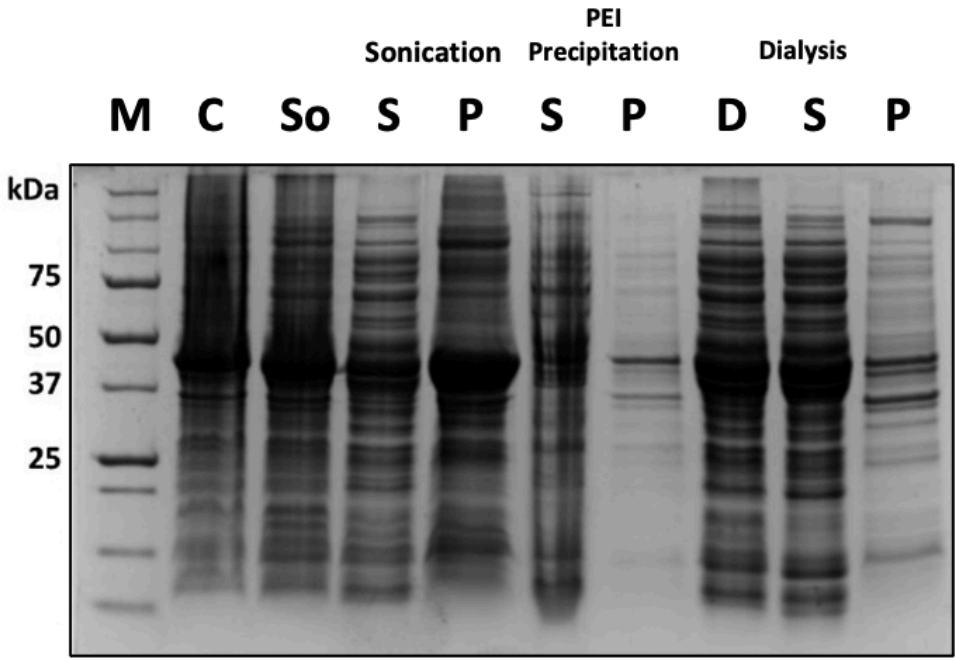


Figure 3.16 Cell lysis for the purification of *Li*FEN

Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Cell pellet (C). Cell lysis sample after sonication (So). Dialysed protein sample (D). Supernatant (S). Pellet (P).

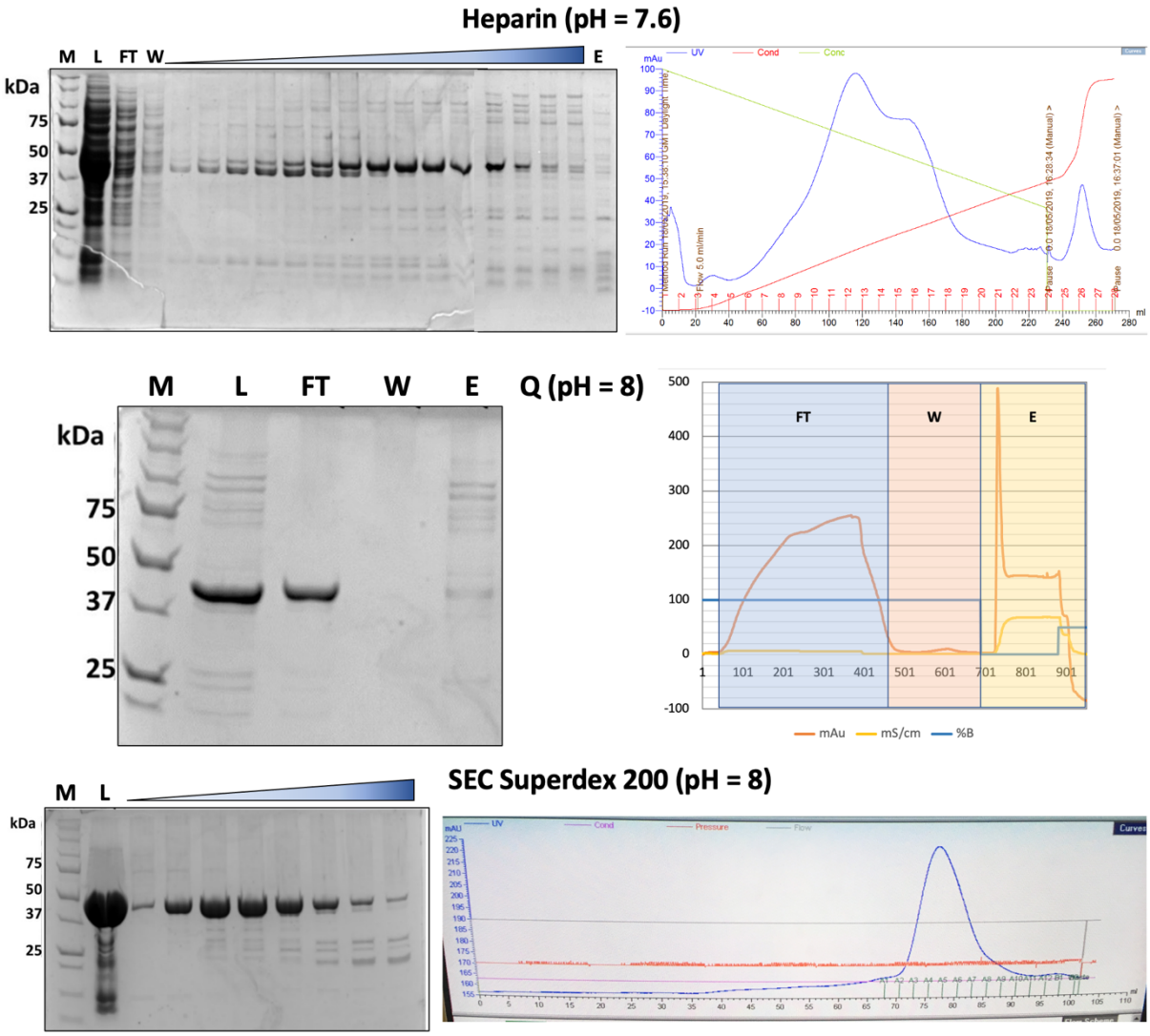


Figure 3.17 Purification of *Li*FEN-WT

Panels on the right side represent SDS-PAGE gels of the purification of LiFEN. *Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Load (L). Flow-through (FT). Wash (W). Elution (E). Salt concentration gradient (heparin) and elution time (SEC) in ascending order (white → blue). Purification was carried out using the chromatography system ÄKTAprime plus™.*

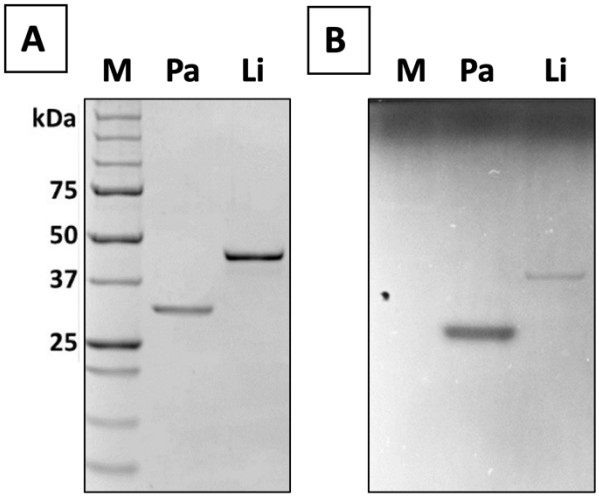


Figure 3.18 Determination of purity and nuclease activity of *Li*FEN

(A) *&* ***(B)****: Purified* Li*FEN concentrated and analysed by SDS-PAGE and zymogram, respectively. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Pseudomonas aeruginosa FEN (Pa). Purified* Li*FEN (Li).*

3.4 Over-expression and purification of *Cn*FEN protein

3.4.1 Construct production of *Cn*FEN-WT in pET21a(+)

A codon-optimised gene fragment that encodes the *Cryptococcus* FEN WT, inserted in the pUC57 plasmid, was obtained from a commercial supplier (Bio Basic Inc). Same protocol for insertion of *Cn*FEN-WT fragment into the pET21a(+) plasmid was followed (section 2.1.4). Production of *Cn*FEN-WT in pET21a(+) is shown in Figure 3.19.

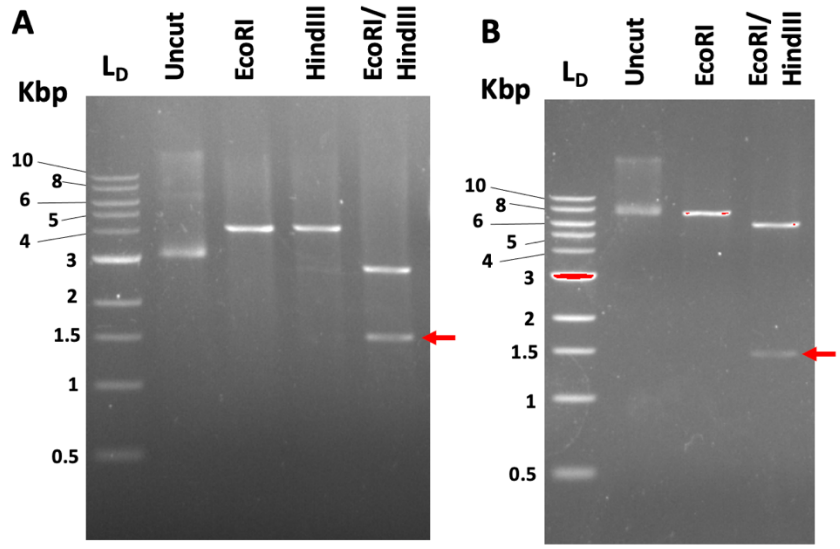


Figure 3.19 Agarose gel electrophoresis of the sub-cloning of *Cn*FEN fragment into the pET21a(+) vector

(A)*: Digestion of pUC57-*Cn*FEN construct with indicated restriction enzymes.* ***(B)****: Digestion with restriction enzymes after sub-cloning in pET21a(+)-*Cn*FEN construct.. Annotations:* L_D *-1 kbp DNA ladder (NEB). Red arrows point to the DNA fragment coding for* Cn*FEN (1452 bp). Images taken using the Bio-Rad Gel Doc EZ Imager.*

3.4.2 Over-expression of *Cn*FEN-WT protein

Over-expression of the *Cn*FEN protein was performed in a small-scale culture through IPTG induction, Studier method, and optimised auto-induction media (protocol in section 2.3.1). A large-scale cell culture for the over-expression of the protein was made only with the optimised auto-induction media (protocol section 2.3.1). This method was preferred over the Studier and IPTG methods because most of the protein expressed using these two methods remained in inclusion bodies (data not shown). The cell culture was left in incubation overnight and the protein over-expression was monitored by taking samples at the time 0, 2, 3.5, 21 and 25 hours (Figure 3.20).

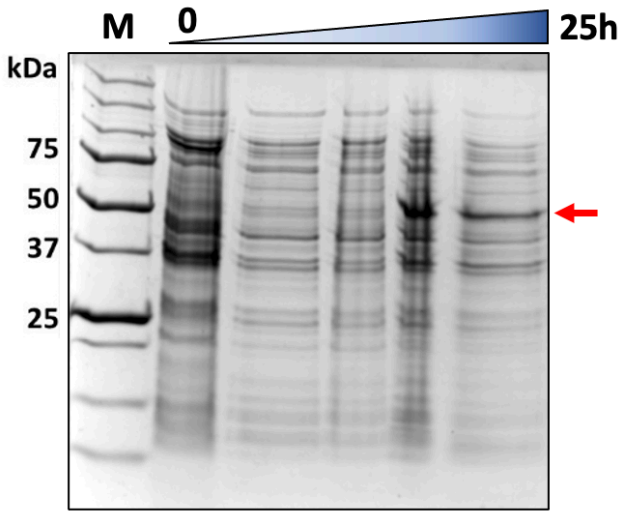


Figure 3.20 SDS-PAGE analysis of the *Cn*FEN over-expression

Large scale over-expression of Cn*FEN. Protein marker Bio-Rad Precision Plus 1610373 (M). Protein over-expression time in ascending order. Red arrow points expression of* Cn*FEN (50 kDa).*

3.4.3 Purification of *Cn*FEN-WT protein

BL21(DE3) cells expressing *Cn*FEN protein were lysed according to the protocol on section 2.3.2, (Figure 3.21 A) and the supernatant was recovered. The isoelectric point of *Cn*FEN is 6.56 and molecular weight of 50 kDa according to ExPASy ProtParam. For this protein, the purification began with ammonium sulphate precipitation, which is a step prior to the chromatography columns. This procedure will facilitate the purification of the protein by precipitating the contaminants or the protein of interest (Figure 3.11 A). The best approach for the next purification steps was to perform a sequential ammonium precipitation (at 1.25 M to precipitate some of the contaminants, which were removed by centrifugation, and at 2.5 M to precipitate *Cn*FEN with a reduced number of contaminants). Because the pI of this protein is lower than *Tb*FEN

and *Li*FEN, the purification protocol for this protein was adapted to take this into account. First, a tandem SP-heparin columns were used with a pH higher (pH of 8) than the pI of *Cn*FEN. With this approach some of the contaminants stuck to the SP column while the *Cn*FEN passed through, but was retained by the heparin column which in this case is acting as an affinity matrix for DNA binding proteins (Farooqui, 1980). The purification continued with an anion exchange step using a salt gradient and at the same pH. Finally, the last step of purification was performed using a size-exclusion chromatography column at the same pH (Figure 3.22 and 3.23). SDS-PAGE and zymogram analysis were performed to determine the purity and nuclease activity of the protein (Figure 3.24).

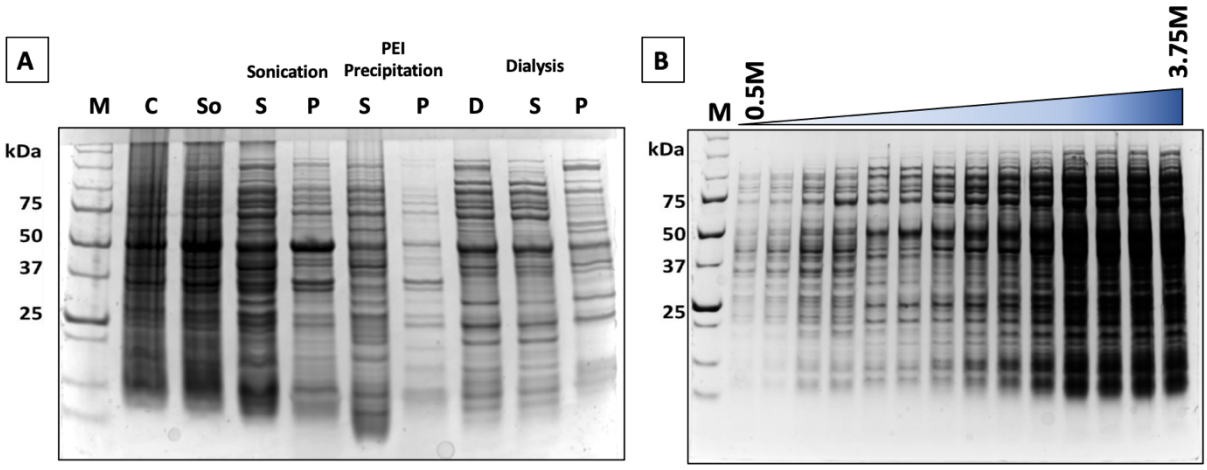


Figure 3.21 SDS-PAGE of the *Cn*FEN cell lysis and ammonium sulphate precipitation gradient

(A) Cell lysis for the purification of Cn*FEN-WT. Cell pellet (C). Cell lysis sample after sonication (So). Dialysed protein sample (D). Supernatant (S). Pellet (P). (B) Ammonium precipitation of* Cn*FEN-WT. Fractioning precipitation in ammonium sulphate concentration from 0.5 M to 3.75 M, with an increasing magnitude of 0.25 M and shown in ascending order (white → blue). Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M).*

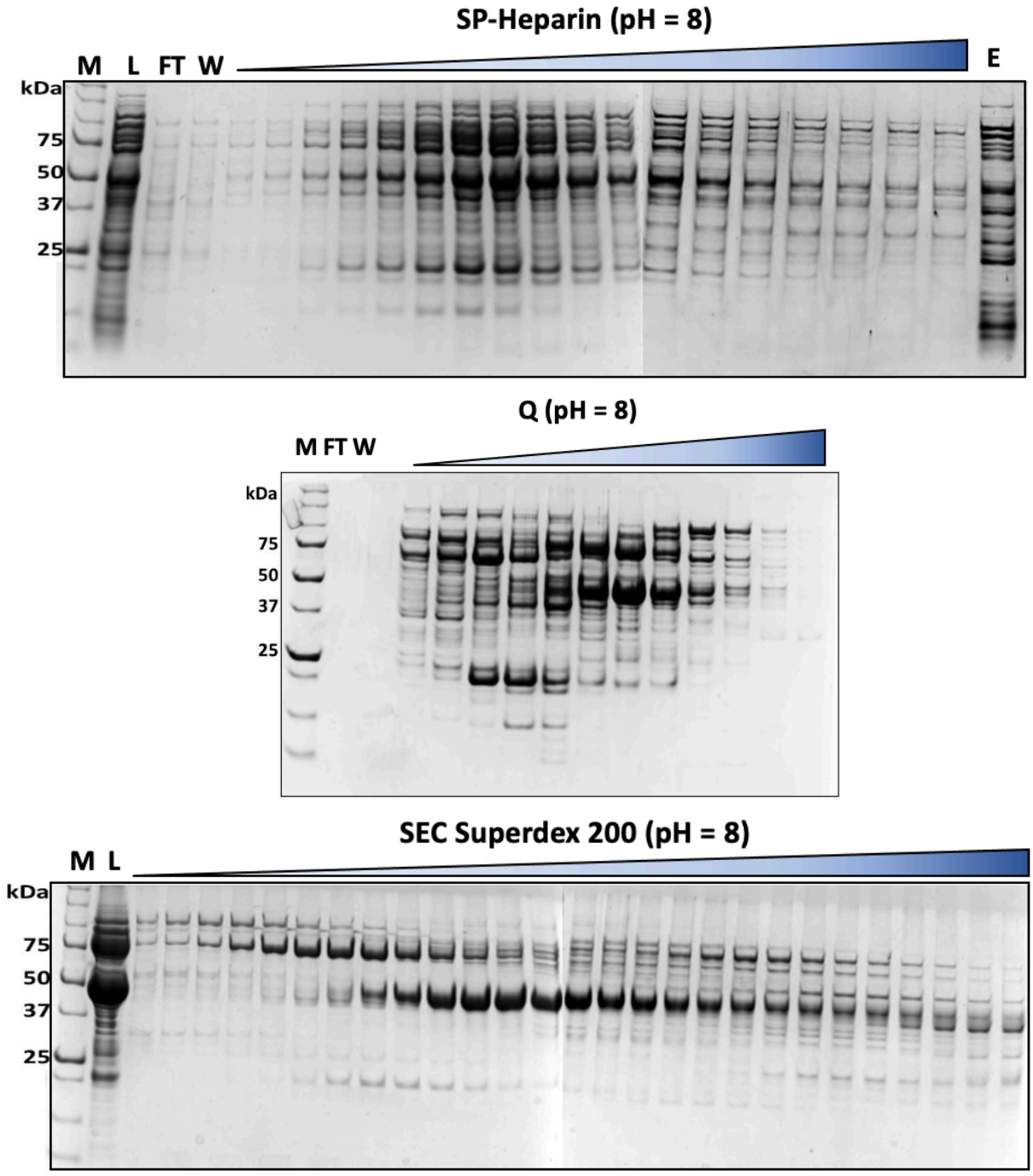


Figure 3.22 Purification of *Cn*FEN-WT

SDS-PAGE gels of the purification of Cn*FEN.. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Load (L). Flow-through (FT). Wash (W). Elution (E). Salt concentration gradient (SP-heparin & Q) and elution time (SEC) in ascending order (white → blue). Purification was carried out using the chromatography system ÄKTA pure™.*

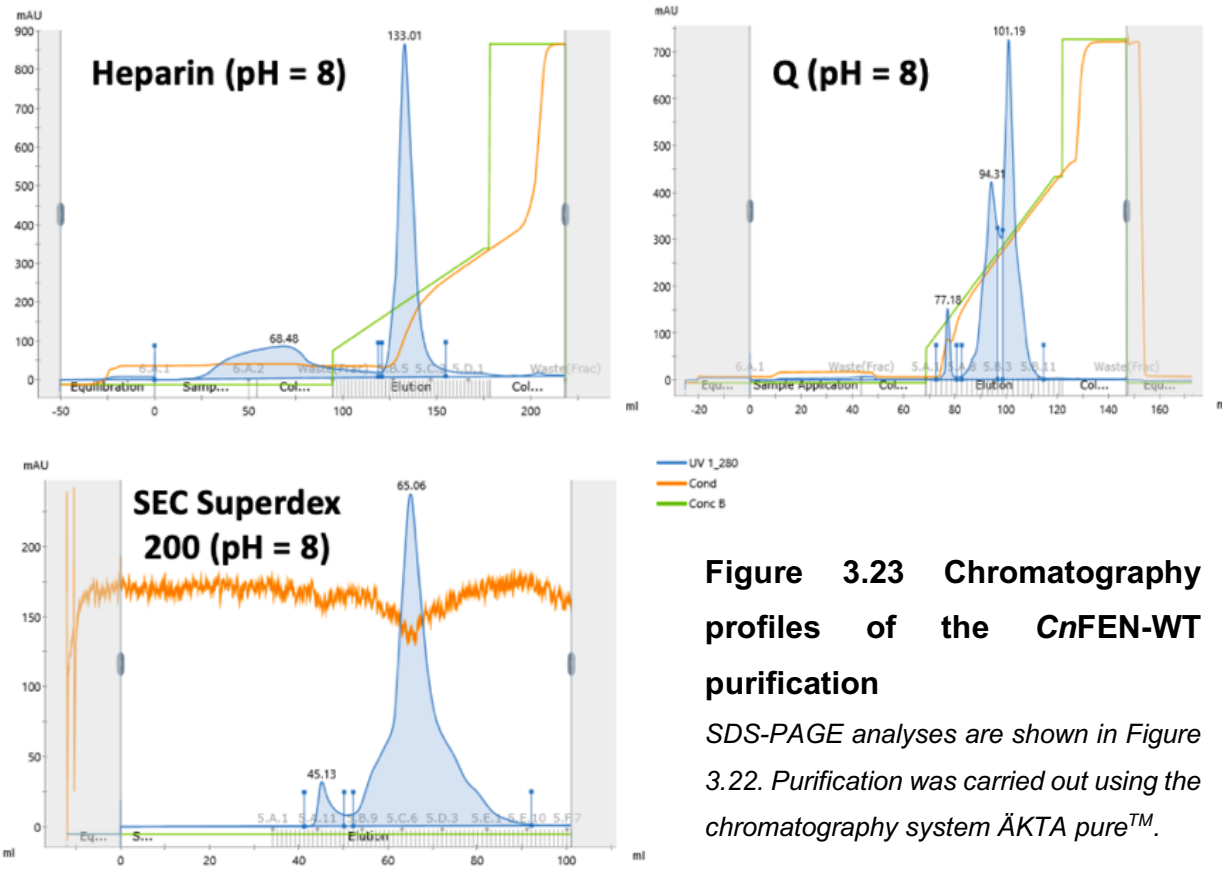


Figure 3.23 Chromatography profiles of the *Cn*FEN-WT purification

SDS-PAGE analyses are shown in Figure 3.22. Purification was carried out using the chromatography system ÄKTA pure™.

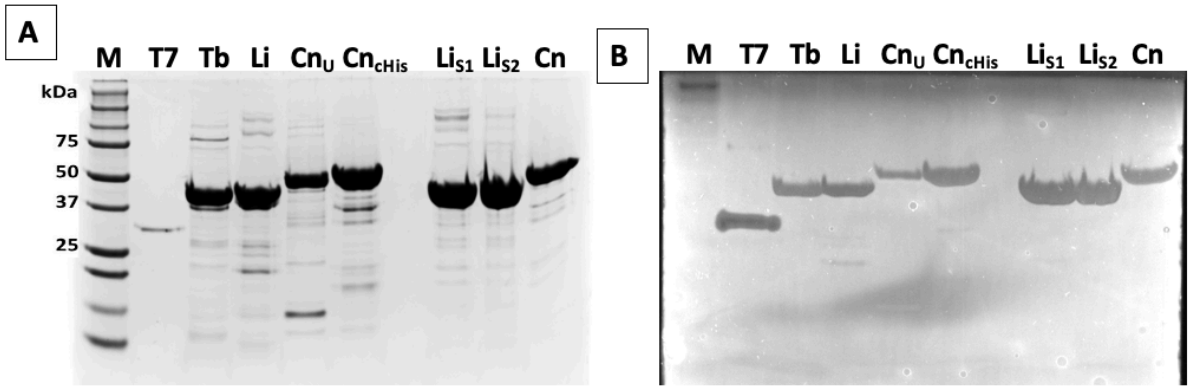


Figure 3.24 Determination of purity and nuclease activity of *Cn*FEN

(A) *&* ***(B)****: Purified* Cn*FEN concentrated and analysed by SDS-PAGE and zymogram, respectively. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Bacteriophage T7 FEN (T7). Purified* Tb*FEN-WT (Tb). Purified* Li*FEN-WT (Li). Unsuccessful purified* Cn*FEN-WT (Cn$_U$). Purified* Cn*FEN-FL with cleavable His-tag (Cn$_{His}$). Purified* Li*FEN-WT previously purified in Sayers laboratory (Li$_{S1}$ & Li$_{S2}$). Purified* Cn*FEN-WT (Cn).*

3.5 Over-expression and purification of *Cn*FEN full-length and truncated forms with cleavable His-tag

Protein purified from the above protocol (Section 3.4.3) is useful for biochemical and inhibition assays, but not enough material was obtained for large-scale crystallization trials. Due to the difficulties to optimize the purification of *Cn*FEN, the long protocol and obtaining a low yield of protein recovered from the induced cells (~10 mg per 15 g of cell paste), an alternative approach was developed. To overcome the problems, a different plasmid with a cleavable C-terminal double-histidine tag was used with the aim of purifying more protein using a shorter protocol. Truncations of the *Cn*FEN were performed before and after PCNA binding recognition site of the C-terminal, obtaining truncated forms of 405 and 414 amino acids, respectively, compared to the full-length protein of 453 amino acids.

3.5.1 Construct production of *Cn*FEN full length and truncated forms in pYM547c

*Cn*FEN gene was amplified from the pET21a(+)-*Cn*FEN using designed primers (Table 2.6) according to the protocol on section 2.2 (Figure 3.25 A). Later on, the amplified genes were inserted in the plasmid pYM547c and transformation of chemically competent *E. coli* cells was performed according to section 2.1.4. Three colonies of each construct (two in the case of ΔC 405) were grown in small cultures for further isolation of plasmid DNA and analyses of the constructs through restriction digestion (Figure 3.25 B).

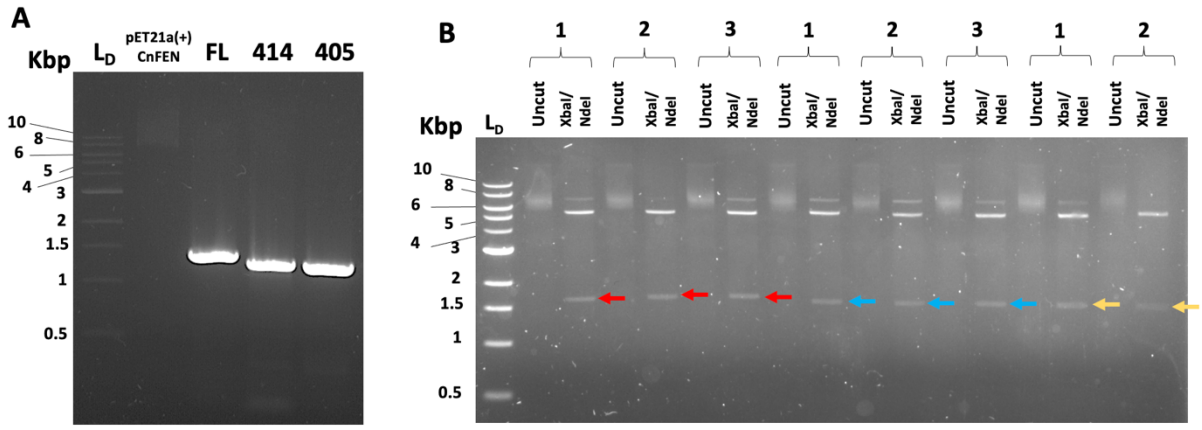


Figure 3.25 Agarose gel electrophoresis of *Cn*FEN full length and truncated constructs into the pYM547c vector

(A) *Uncut pET21a(+)-*Cn*FEN and PCR products* Cn*FEN-FL (FL),* Cn*FEN-ΔC 414 (414), and* Cn*FEN-ΔC 405 (405).* ***(B)*** *Cloned colonies of* Cn*FEN-FL and truncated construct in pYM547c. Annotations:* L_D *- 1 kbp DNA ladder (NEB). Red arrows point to the DNA fragment* Cn*FEN FL (1590 bp); Blue arrows point to the DNA fragment* Cn*FEN-ΔC 414 (1470 bp); Yellow arrows point to the DNA fragment* Cn*FEN-ΔC 405 (1443 bp). Images taken using the Bio-Rad Gel Doc EZ Imager.*

3.5.2 Over-expression of *Cn*FEN cleavable His-tag constructs

Over-expression of the *Cn*FEN protein was performed in a small-scale culture through IPTG induction, Studier method, and optimised auto-induction media (protocol in section 2.3.1). A large-scale cell culture for the over-expression of the protein was made only with the IPTG induction method (protocol section 2.3.1). The IPTG induction was preferred over the other two methods because the protein expression was absent in the Studier method and optimised auto-induction media (data not shown). The cell culture was incubated overnight, and protein over-expression was monitored by taking samples at the time 0, 1.25, 4.25, 23.5 and 27 hours (Figure 3.26).

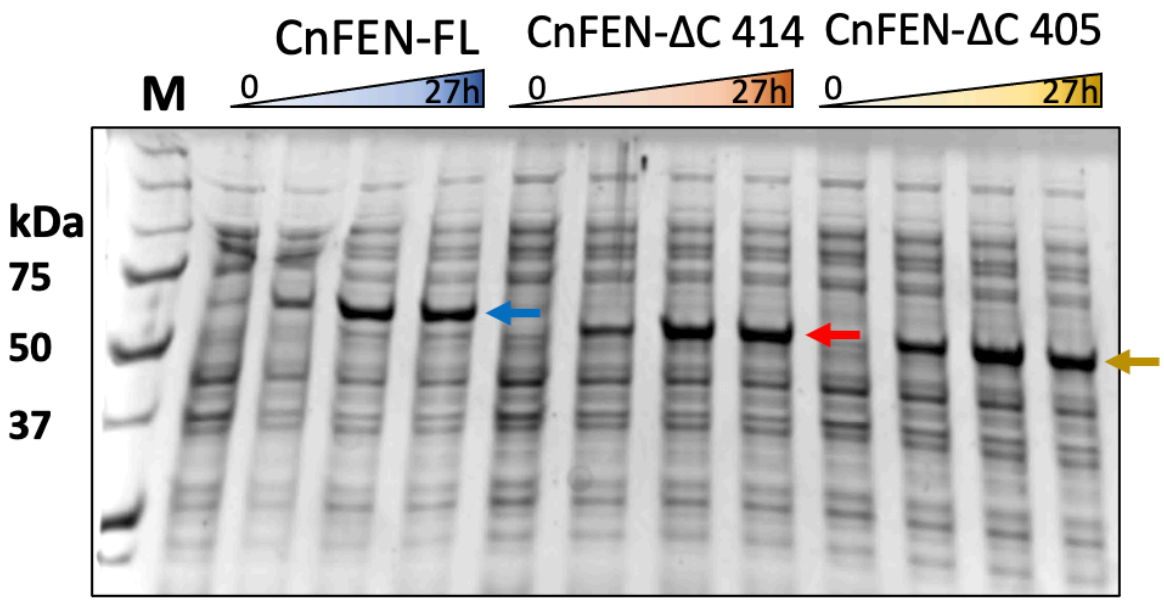


Figure 3.26 SDS-PAGE of the *Cn*FEN His-tag constructs over-expression

Analysis of protein over-expression using 0.1 mM IPTG in the large-scale cell culture. Protein over-expression time in ascending order. Blue arrow points expression of Cn*FEN-FL (55 kDa). Red arrow points expression of* Cn*FEN-ΔC 414 (51 kDa). Yellow arrow points expression of* Cn*FEN-ΔC 405 (50 kDa). Protein marker Bio-Rad Precision Plus 1610373 (M).*

3.5.3 Purification of *Cn*FEN cleavable His-tag proteins

BL21(DE3) cells expressing cleavable His-tag forms of *Cn*FEN FL, ΔC 414 and ΔC 405 proteins were lysed according to the first two paragraphs from the protocol on section 2.3.2. The supernatant from the sonicated cell suspension was recovered and the purification was performed according to section 2.3.3. The purification started through immobilized metal affinity chromatography using a nickel-chelate column at a pH of 8, eluting first the DNA of the supernatant using a high salt buffer followed by an imidazole gradient to elute the protein. Purification continued using a heparin column at pH 8. The histidine tag of the proteins was cleaved using human rhinovirus (HRV) 3C protease and separated of the proteins through a nickel-chelate column. The last purification step was a size-exclusion chromatography column (Figures 3.27 to 3.29).

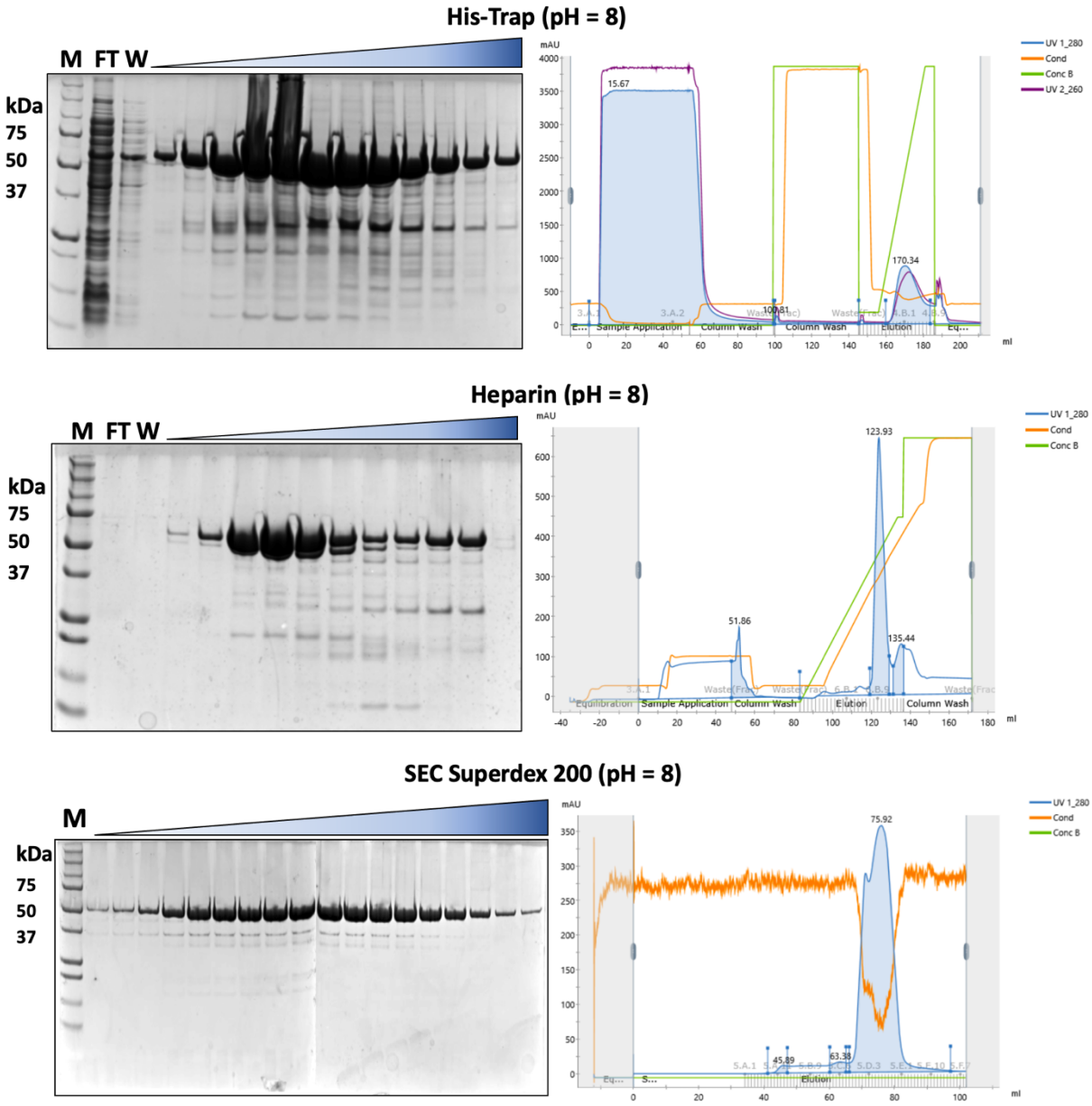


Figure 3.27 Purification of *Cn*FEN-FL

Panels on the right side represent SDS-PAGE gels of the purification of Cn*FEN-FL. Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Flow-through (FT). Wash (W). Imidazole concentration gradient (His-trap), salt concentration gradient (heparin) and elution time (SEC) in ascending order (white → blue). Purification was carried out using the chromatography system ÄKTA pure™.*

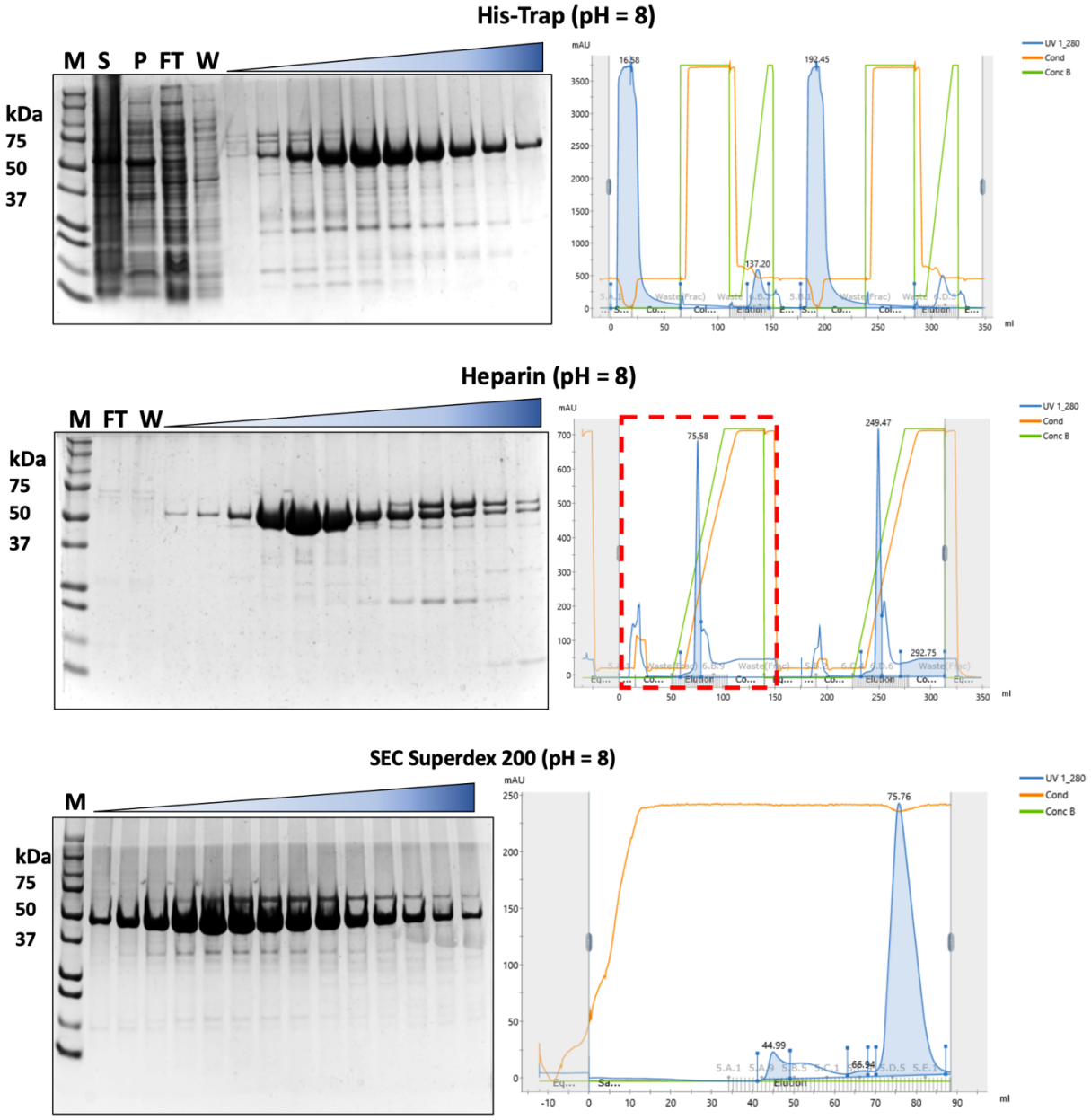


Figure 3.28 Purification of *Cn*FEN-ΔC 414

Panels on the right side represent SDS-PAGE gels of the purification of Cn*FEN-ΔC 414. Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Supernatant (S) and Pellet (P) after sonication. Flow-through (FT). Wash (W). Imidazole concentration gradient (His-trap), salt concentration gradient (heparin) and elution time (SEC) in ascending order (white → blue). Red dotted square correspond to the chromatography profile of* Cn*FEN-ΔC 414. Purification was carried out using the chromatography system ÄKTA pure™.*

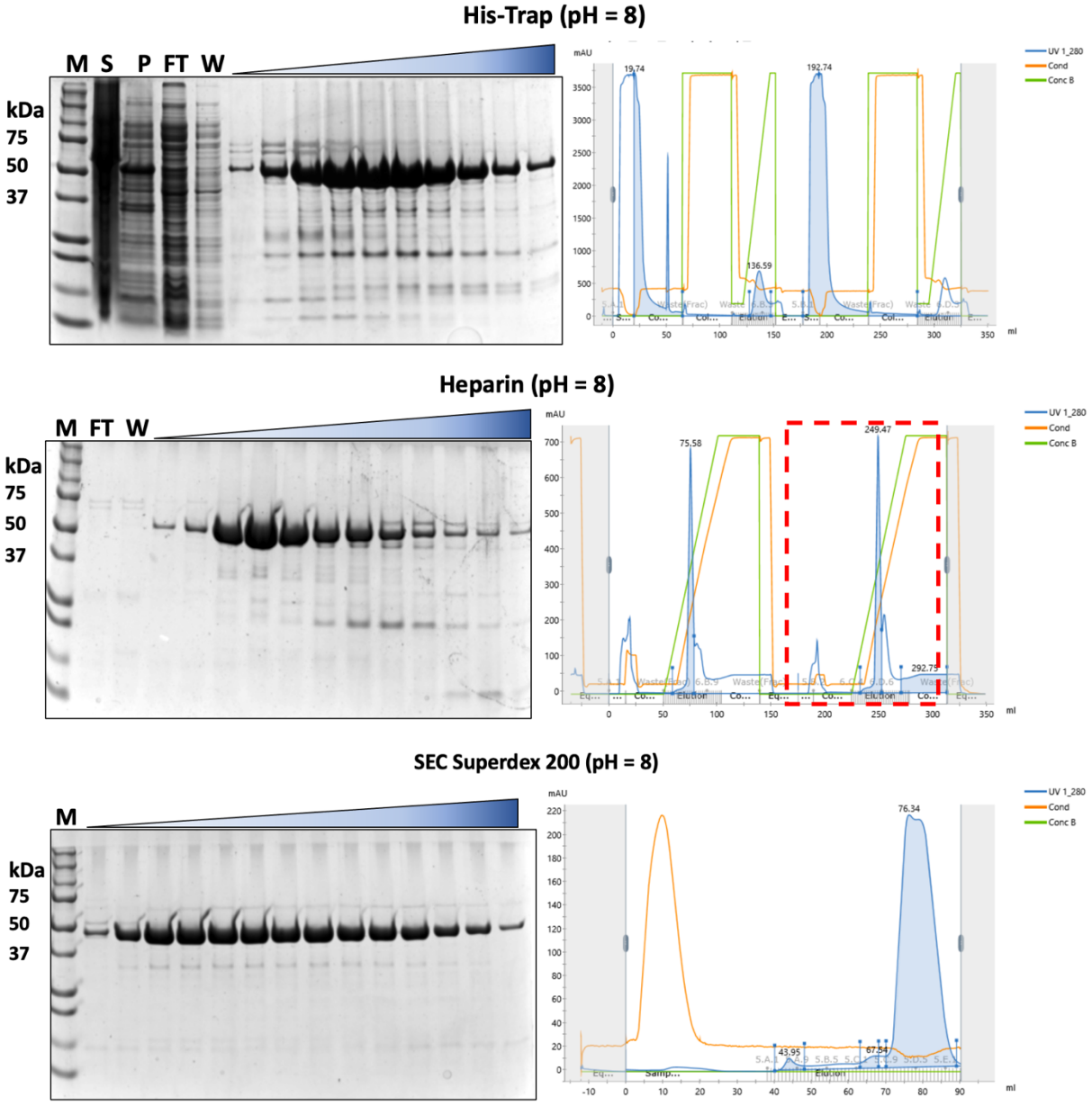


Figure 3.29 Purification of *Cn*FEN-ΔC 405

Panels on the right side represent SDS-PAGE gels of the purification of Cn*FEN-ΔC 405. Panels on the left side correspond to chromatography profiles from each column. Annotations: Protein marker Bio-Rad Precision Plus 1610373 (M). Supernatant (S) and Pellet (P) after sonication. Flow-through (FT). Wash (W). Imidazole concentration gradient (His-trap), salt concentration gradient (heparin) and elution time (SEC) in ascending order (white → blue). Red dotted square correspond to the chromatography profile of* Cn*FEN-ΔC 414. Purification was carried out using the chromatography system ÄKTA pure™.*

3.6 Production of anti-*Cn*FEN

Production of antibodies against *Cn*FEN protein was carried out in this project to purify the protein from a lysate of *Cryptococcus neoformans* cells through immunoprecipitation and further characterization and analysis of the protein, comparing it with the amino acid sequence deposited in the NCBI database. Due to time constriction, lysis of *C. neoformans* cells and immunoprecipitation were not performed. Antibodies against *Cn*FEN protein were raised in rabbits according to section 2.5. The sera obtained from the rabbits was tested in a Western Blot to confirm the specificity of the antibody to bind *Cn*FEN (section 2.4.4). BL21(DE3) with pET21a(+) and pYM547c empty vectors, as well as uninduced and induced cells were used to perform this test (Figure 3.30).

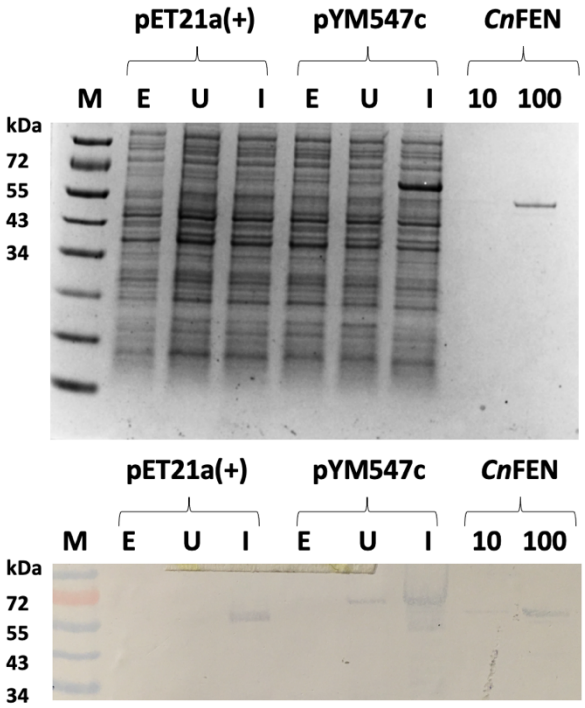


Figure 3.30 Western-Blot for detection of *Cn*FEN using anti-*Cn*FEN

(A) *&* ***(B)*** *SDS-PAGE and Western-Blot for detection of recombinant* Cn*FEN in BL21(DE3) cells. Annotations: Protein marker NEB Color Prestained Protein Standard, Broad Range (10-250 kDa) P7719S (M). Empty plasmid (E). Uninduced* Cn*FEN protein (U). Induced* Cn*FEN protein (I).. Purified* Cn*FEN FL with cleaved His-tag was used as control, loading 10 & 100 ng of protein in each well in the gel (10 & 100).*

3.7 Mass Spectrometry Analysis of microbial FENs

To assess that the purification of the protein was performed successfully and that protein purified corresponds to the predicted molecular weight from ExPASy web server, samples of purified protein were analyzed through mass spectrometry. The proteins were analyzed using the Bruker Reflex III mass spectrometer by the Facility of Mass Spectrometry (Department of Chemistry, University of Sheffield), for the determination of the molecular weight. Mass spectrometry data is summarized in Appendix (Tables 9.2 to 9.4 & Figures 9.3 to 9.4).

CHAPTER 4: VIRTUAL HIGH-THROUGHPUT SCREENING OF SMALL-MOLECULE LIBRARIES TARGETING MICROBIAL FLAP ENDONUCLEASES

4.1 Introduction

Drug discovery is a multidisciplinary, complex, expensive, and time-consuming process (Zoete *et al.*, 2009). It is commonly acknowledged that despite investing far more in R&D now than it did 20 years ago, the pharmaceutical industry is still facing challenges to develop new FDA-approved "novel molecule entities" (NMEs). According to Mullard (2022), the rolling 5-year average of drug approvals is 51 per year (38 of them were NMEs), while a decade ago it was 24 drugs per year (19 NMEs). However, the year with most registered FDA-approved NMEs was in 1996 with a number 47 NMEs. At the same time, it is important to mention that there are an estimated ~10,000 molecular targets that could be modulated by pharmacological drugs. Nevertheless, only 4% of these macromolecules are being studied in drug development programmes (Jin *et al.*, 2014).

Considering the previous points discussed, structure-based drug design (SBDD) is an important component in the pharmaceutical industry (Kinch & Hoyer, 2015; Batool *et al.*, 2019; Maia *et al.*, 2020). The integration of structural determination approaches, such as X-ray crystallography, with *in silico* techniques, such as molecular dynamics, homology modelling, and molecular docking, has aided progress in SBDD during the last two decades (Kalyaanamoorthy & Chen, 2011). Furthermore, it is projected that *in silico* technologies will become an essential tool to accelerate lead discovery and optimization through computational techniques and improvements in computational power, artificial intelligence (AI), cloud computing, neural network, (Zoete *et al.*, 2009; von Stosch *et al.*, 2021; Schanduagrat *et al.*, 2020).

The synergistic use of structural determination and *in silico* techniques, particularly during the preclinical phase of the R&D process, has enabled 3D structures of many biological macromolecules to be determined as well as the characterization of binding site features such as steric and electrostatic properties. This useful information has been crucial in understanding ligand-receptor molecular recognition events (Honarparvar *et al.*, 2014). For example, Stauch and collaborators (2010) used a

combination of NMR spectroscopy and molecular docking to characterize intermolecular interactions of the molecule argyrin A as an anticancer agent with the human S20 proteosome. Currently, these techniques have helped to develop 66 drugs that have gone through clinical trials, some of which are now FDA authorised (Brown & Boström, 2018). An example of the drugs developed through SBDD and approved by the FDA is the NME APG-115, later known as alrizomadlin and used for the treatment of stage IIB-IV melanoma and authorised in 2021 (Aguilar *et al.*, 2017). Another example is the molecule GSK1265744, an antiretroviral drug from the integrase strand transfer inhibitor family (INSTI) later named as cabotegravir and used for the treatment of HIV, authorised in 2021 (Johns *et al.*, 2013; Yoshinaga *et al.*, 2015).

In silico drug design should not be viewed as a straight-forward technique capable of directly suggesting a small number of compounds with high affinity and selectivity for the targeted macromolecule, with favourable pharmacokinetic and pharmacodynamic properties. Instead, it consists of the systematic application of computational tools such as molecular docking, virtual high-throughput screening (vHTS) and fragment-based ligand design (FBLD) (Zoete *et al.*, 2009).

Molecular docking is a method used to predict the most likely 3D conformations of small-molecule ligands within a target binding site, offering quantitative estimates of the energy involved in an intermolecular recognition event (Meng *et al.*, 2011). Furthermore, the docked molecules can be ranked according to their quantitative binding energy estimation, which is helpful in the selection of compounds for experimental tests. There are two distinctive steps in molecular docking which are as follows: exploration of the ligand conformational space within the binding cavity of the macromolecule and estimation of the binding energy for each expected conformation (Kitchen *et al.*, 2004; Ferreira *et al.*, 2015).

The use of quick and cost-effective computational approaches to identify potentially active compounds from virtual databases is known as virtual high-throughput screening (Lionta *et al.*, 2014). Exploring virtual chemical libraries is a common strategy in drug development, and some novel therapies may be traced back to virtual screening efforts (Blaney, 2012; Eder *et al.*, 2014). Examples of these achievements are the development of a glucocorticoid receptor antagonist by Corcept Therapeutics

for the treatment of patients with probable non-alcoholic steatohepatitis and in obese adults with schizophrenia treated with antipsychotic medications (Clinical Trial number: NCT03823703; Koorneef *et al.*, 2018), and a fibrinolysis inhibitor develop by AstraZeneca for the treatment of heavy bleeding and other blood disorders (Cheng *et al.*, 2014). Virtual HTS provides a rapid and cost-effective way to rank docked molecules by utilising scoring algorithms, choosing interesting compounds to try experimentally.

The following general protocol is often used in vHTS strategies: (1) molecular target selection and preparation, (2) druggability analysis of pocket finding (3) compound database selection, (4) molecular docking, and (5) outcome analysis (Figure 4.1).

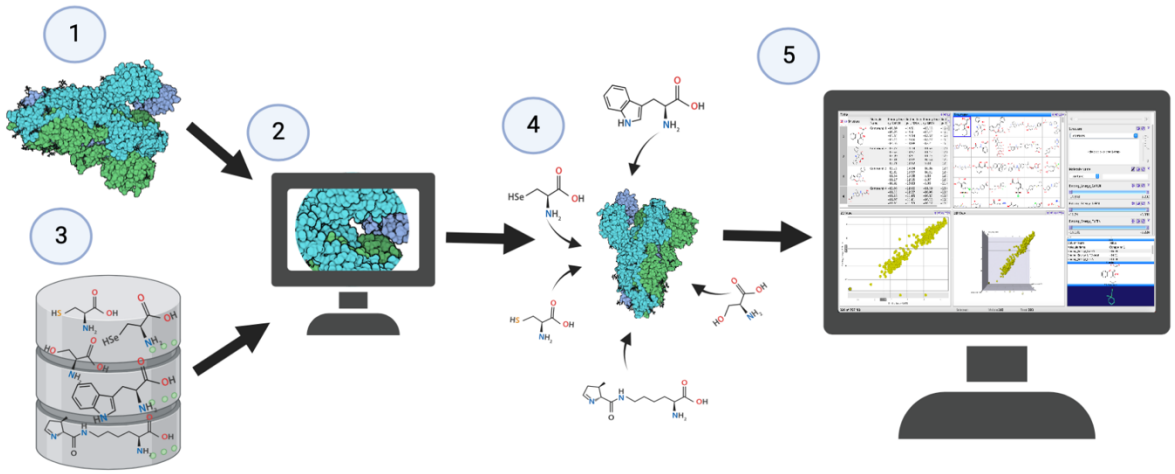


Figure 4.1 Workflow of the general protocol for virtual high-throughput screening

The first step (1) of vHTS is the selection of the molecular target and preparation of the structure according to the input required by the programme. (2) Identification of all hot spots of the target and druggability analysis of potential sites. (3) Selection of the required compound libraries and preparation of the molecules according to the programme input. (4) Molecular docking through different algorithms analysing a very large number of reasonable poses of the molecule interacting with the target. (5) Outcome analysis of the results obtained from the in-silico *studies, which is typically binding energy and 3D coordinates of each small molecule. This can be visualized in data analysis packages such as Datawarrior.*

When designing a structural-based virtual screening (SBVS) process, as much background information as possible should be considered. For example, it is desirable that the mechanism of the target is understood so a rational approach to inhibition site selection can be used and augmented by data on the molecular flexibility or alternative conformations that the target may adopt. Clearly, knowledge of any ligands, co-factors, metal ions, structurally important water molecules and of course substrate binding sites can all be useful in developing new inhibitors (Scior *et al.*, 2012). After selecting a molecular target structure, it goes through some preparation steps for molecular docking. Typically, these steps involve removal of non-polar hydrogen atoms, elimination of water molecules, assignment of partial charges, addition of polar hydrogens and specification of protonation states for amino acid residues (Jain & Nicholls, 2008).

Pocket finding and druggability analysis are the following stages in the protocol. In these stages it is important to identify an important region of the protein target. This might be a region that normally interacts with a substrate, cofactor, or other small ligand. This approach assumes that identification of the small molecule which might bind tightly to that region of the protein would then block its function (Nisius *et al.*, 2012). These interactions often occur in small areas (pockets) on the surface of the proteins. Clefts and holes correlate with biological relevant binding sites and often provide attractive targets for drug development (Wells & McClendon, 2007). However, not every binding site is ideal to develop a drug. Before looking for a molecule that can fit into a binding site, one should consider its druggability (Barakat, 2014). Due to the difficulties to predict the druggability of a binding site, the use of computational methods during this stage has become essential (Henrich *et al.*, 2010). These computational methods are divided in two steps: the identification of all possible binding sites within and on the surface of the target structure, and the ranking of these sites in terms of their druggability (Feng & Barakat, 2018). Nowadays, most of the computational tools to identify hot spots can be found on web servers, capable of defining a reasonably small 3D space and focusing on “important” regions (active, binding, or allosteric site). Examples of computational tools to perform the first step are Q-SiteFinder (Laurie & Jackson, 2005), GHECOM (Kawabata, 2010), POCASA (Yu *et al.*, 2010), COFACTOR (Roy et al., 2012), FINDSITE (Brylinski & Skolnick, 2009),

among others. Furthermore, computational tools have been developed to assess the druggability of a hot spot such as PockDrug (Hussein *et al.*, 2015), FTMap (Kozakov *et al.*, 2015) and Sitemap (Halgren, 2007; Halgren, 2009).

The selection of compound libraries to be used in molecular docking experiments is the next essential stage. Most frequently, freely accessible virtual databases with a large chemical variety are employed such as ZINC20, ChEMBL, DrugBank, PubChem, among others (Moura Barbosa & Del Rio, 2012; Singh *et al.*, 2021). These libraries are normally interactive interfaces that allow the use of chemical filters to search and pick specific subsets representing a certain chemical space. Virtual molecule libraries can be found as line notations (e.g., SMILES, SMARTS, or InChI files). When the files are downloaded, they are displayed in 2D structures which can immediately be transformed into 3D structures with suitable ionization states, partial charges, and stereochemistry (Ewing *et al.*, 2001; Moura Barbosa & Del Rio, 2012).

The next stage consists of docking the selected molecules into the target binding site. For this stage, several computer programmes with different algorithms have been developed over the past 40 years, starting with the revolutionary DOCK program developed by Kuntz and colleagues in 1982 (Kuntz *et al.*, 1982), followed by other programmes such as AutoDock (Morris *et al.*, 1998), FlexX (Rarey *et al.*, 1996; Rarey *et al.*, 1997), FRED (McGann *et al.*, 2003), Glide (Friesner *et al.*, 2004; Halgren *et al.*, 2004), GOLD (Jones *et al.*, 1997; Verdonk *et al.*, 2003), ICM (Totrov *et al.*, 1997), among many others.

The last stage consists of post-docking analysis. This is performed once the programme explores all the possible conformations of the ligand according to the interaction with the target site and compounds are classified as potential hits according to the scoring algorithm. The analysis involves viewing the anticipated ligand-receptor complexes and examining crucial aspects such as binding conformations and intermolecular interactions. In many cases, this is essential because some highly scoring docked conformations may include serious clashes between inhibitor and receptor atoms or distortions of the small molecule, which would make them unlikely to bind in reality. After this “sense-check”, selection of a chemically diverse set of top-scoring compounds should be prioritised for further physical screening (Meng *et al*.,

2011; Ferreira *et al.*, 2015; Lionta *et al.*, 2014). The prediction of free energy binding from target structures is still considered a big challenge in computational chemistry. The main problem lies in prioritising natural binding modes over irrelevant poses and, as a result, ranking the potential of pre-existing binding modes is fundamental to detect the most potent ligands from a large chemical space. This leads the researchers to observe a significant difference between the reported accuracy level in previous predictive studies and real performance in prospective virtual screening experiments, resulting in an overestimation of virtual screening accuracy in all current investigations (Tran-Nguyen *et al.*, 2021). This is why another element that may be evaluated by viewing the docking results is whether the solutions provided by the programme for known ligands correspond to crystallographic conformations, which can determine if the docking simulations can accurately recreate experimental data (dosSantos *et al.*, 2018).

In this project, the virtual high-throughput screening of compound libraries targeting *Tb*FEN, *Li*FEN and *Cn*FEN was performed using the GUI software PyRx (Dallakyan & Olson, 2015). The vHTS was carried out using the protein structures described on Section 2.10.2 as target molecules and small-molecule libraries from Section 2.10.1 as ligands. The programmes used through the PyRx interface were AutoDock 4 (Morris *et al.*, 2009) and AutoDock Vina (Trott & Olson, 2010). The two docking programmes were developed in parallel to address two different needs. AutoDock was the first docking programme developed at the Scripps Research Institute, and it is still a widely used, public-domain platform for experimental docking. AutoDock Vina was developed later to meet the demand for a docking algorithm that does not require expert knowledge from the users. It is highly optimized for docking experiments applying well-tested default methods. For most systems, AutoDock Vina is quick and effective, whereas AutoDock is offered for systems that require further methodological upgrades (Forli *et al.*, 2016). Here, it is important to mention that AutoDock 4 and AutoDock Vina have differences in their algorithm and, therefore, their scoring function. AutoDock 4 scoring function is semiempirical and involves parameters such as Coulomb potential term, Lennard-Jones 12-6 potential term, de-solvation associated with volume, and conformational entropy associated to the number of rotational bonds. On the other hand, AutoDock Vina scoring function is merely empirical and comprises Gaussian

steric interactions, repulsion, hydrogen bonds, and hydrophobic and torsion terms (Nguyen et al., 2020). For these reasons, AutoDock Vina is faster than AutoDock 4 but does not handle atom charges in a similar way, providing different results. This can lead to some targets getting a better prediction with the results from AutoDock Vina while others get a better result with AutoDock 4.

4.2 Results and experimental approach

4.2.1 Selection and preparation of molecular targets

As it has been mentioned throughout the introduction, flap endonucleases are essential for DNA replication and repair. The human homologue is the most studied member of the flap endonuclease family with the availability of several crystal structures and good understanding of the molecular mechanisms of DNA binding and hydrolysis. Furthermore, drug screening and drug discovery studies have been performed targeting this enzyme with the purpose of developing anticancer treatments. Examples of these studies involved known FEN1 inhibitors such as aurintricarboxylic acid (ATA) and 3-hydroxy-5-methyl-1-phenylthieno [2, 3-d] pyrimidin-e-2, 4(1H, 3H) dione (PTPD) (Bina-Stein & Tritton, 1976; Tumey *et al.*, 2005), or the high-throughput screening conducted by AstraZeneca, testing 850,000 compounds, and identifying over 6000 potential inhibitors (McWhirter *et al.*, 2013).

However, there are no published studies reported that describe inhibitors of FEN enzymes in other organisms. Nevertheless, flap endonuclease has been proposed as a potential drug target in *Trypanosoma cruzi* (Ponce *et al.*, 2017). In addition, Uzcanga and collaborators (2017) have proposed as drug targets the enzymes involved in the DNA replication and cell repair in *Leishmania* organisms, but they do not mention specifically FEN1. Furthermore, a study conducted by Tripathi and co-workers (2012) reported killing *Cryptococcus neoformans* in mice that were treated with hydroxyurea. This paper does not explicitly mention FEN1. Related larger N-hydroxyurea derivatives have been reported to be inhibitors of human FEN1, with extremely low IC_{50} values in the low nanomolar range (Exell *et al.*, 2016). However, the concentrations of

hydroxyurea, a rather small molecule with a molecular weight of 77 Da, used in the study from Tripathi were relatively high (25 – 200 mM). While the hydroxyurea might be targeting the FEN enzyme, there is no evidence to suggest that the mode of action of this compound involves FEN1. Nevertheless, another study has suggested a DNA replication enzyme in *C. neoformans* as a potential drug target (Boggs, 2017).

Taking into consideration the essential nature of FEN enzymes and the differences between the protein sequences from various organisms we decided to target these enzymes. This chapter describes the use of molecular docking algorithms AutoDock 4 and AutoDock Vina for virtual screening of a rule-of-three library and correlation of the results with physical screening of the same library carried out in Chapter 6. I aimed to compare different docking models and algorithms to identify those that have the most predictive power. An efficient *in silico* approach with a sufficiently high predictive efficiency would be useful in future large-scale virtual screening campaigns as it should maximise the percentage of hits obtained.

4.2.2 Pocket finding and druggability analysis of microbial FENs

Next, the structures of my target proteins were examined for the presence of potential pockets into which small molecules might bind. This was performed using the online tool POCASA (Yu *et al.*, 2010). The *Trypanosoma* FEN structure was previously determined in the Sayers' laboratory (data provided by Dr. Jason Wilson). The remaining structures were not available from experimentally determined models (the experimental model of *Li*FEN described in Chapter 5 was obtained after the *in silico* studies). Thus, their structures were predicted using Phyre2 server and AlphaFold 2 server & database (Kelley *et al.*, 2015; Jumper *et al.*, 2021; Varadi *et al.,* 2022). Results of pocket finding of the microbial FENs are shown in Figure 4.2.

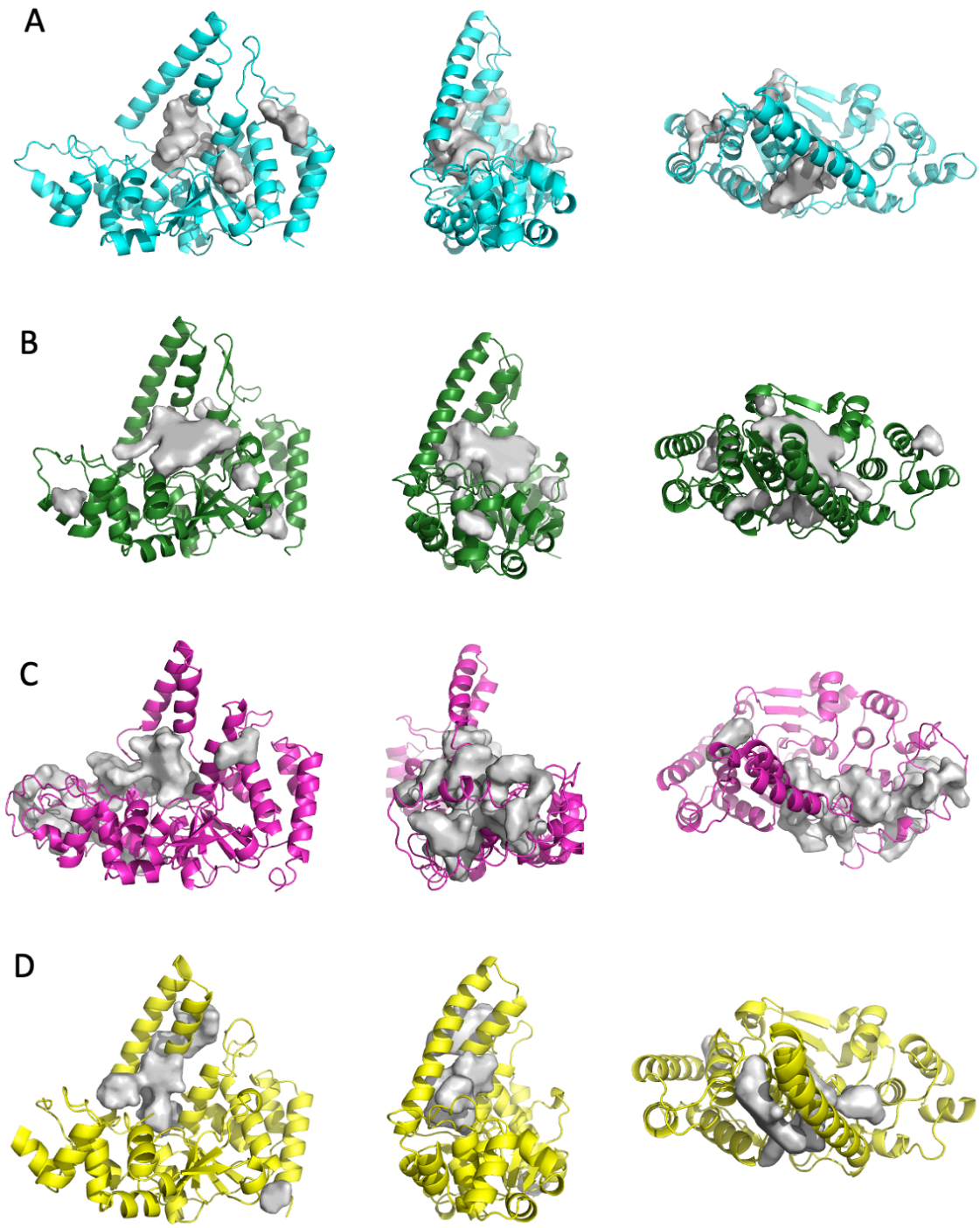


Figure 4.2 Predicted binding pockets in microbial FENs

POCASA was used to identify potential pockets. Panels show cartoon representations of the proteins and their predicted binding pockets (shown as grey solids surfaces). ***(A)*** *&* ***(B)*** *correspond to three orthogonal views of* Tb*FEN experimental structure and* Li*FEN AlphaFold 2 predicted structure, respectively. Phyre2 and AlphaFold 2 gave quite different predictions for some regions of CnFEN protein, therefore, both structures were analysed and shown in* ***(C)*** *&* ***(D)****.*

The results of the POCASA analysis suggest a large central pocket is present which overlaps with the DNA threading helical arch and active sites in each protein. The size of the pockets varies from protein to protein to some degree as it can be seen in the Figure 4.2. Other small pockets are also predicted in the proteins as shown in the previous diagrams.

4.2.3 Selection and preparation of the BioNet compound library

This section of the work was focused on a rule-of-three compound library from Key Organics, specifically their 2nd Generation BioNet Premium Fragment library (https://www.keyorganics.net/bionet-products/fragment-libraries/). This library contains 1166 small molecules with molecular weights from 95 to 290 Da, up to 3 hydrogen donors, up to 5 hydrogen acceptors and 3 or fewer rotatable bonds (approximately 60% of the molecules have one or no rotatable bonds). These also have good solubility in aqueous buffers, have non-hydrogen atom counts of 16 or less and polar surface area of 60 $Å^2$ or below. Promiscuous and reactive substructures are not full to be present in this library. Such libraries are used as part of fragment-based drug design approaches as well as starting points for standard hit expansion type studies. Preparation of the molecules was performed according to Section 2.10.1 from Chapter 2.

4.2.4 Molecular docking results and analysis

Molecular docking of the microbial FENs using the BioNet library as ligands was performed according to Section 2.10.3. Briefly, an sdf file containing the 1166 structures in the library was downloaded from the supplier and converted into individual pdbqt files using the program OpenBabel via GUI computer tool iBabel.

Each protein's pdbqt file was automatically obtained using the PyRx graphical interface which allows simplified access to the algorithms AutoDock 4 and AutoDock Vina, which uses the appropriate structure coordinates in pdb format. It is important to adequately prepare the input pdb files, such that only the receptor molecule is present. Where

experimental structures were used as input, all waters and other ligands were removed. For some models, one or both metal ions usually present in the active site were added back to the protein according to Section 2.10.2 in Chapter 2. The two programs were used to dock various models (no metals, two metals, or one metal in either site) as described in Section 2.10.3.

In silico validation of the POCASA pocket finding results

Before carrying out a focused docking, a preliminary "blind" docking was carried out on an entire FEN protein. This used the BioNet compounds. Results from the "blind" docking is shown in Figure 4.3. It reveals hot spots for binding which roughly correlate with the pockets predicted by POCASA. The most populated binding site is the one containing the active site metals and DNA binding arch.

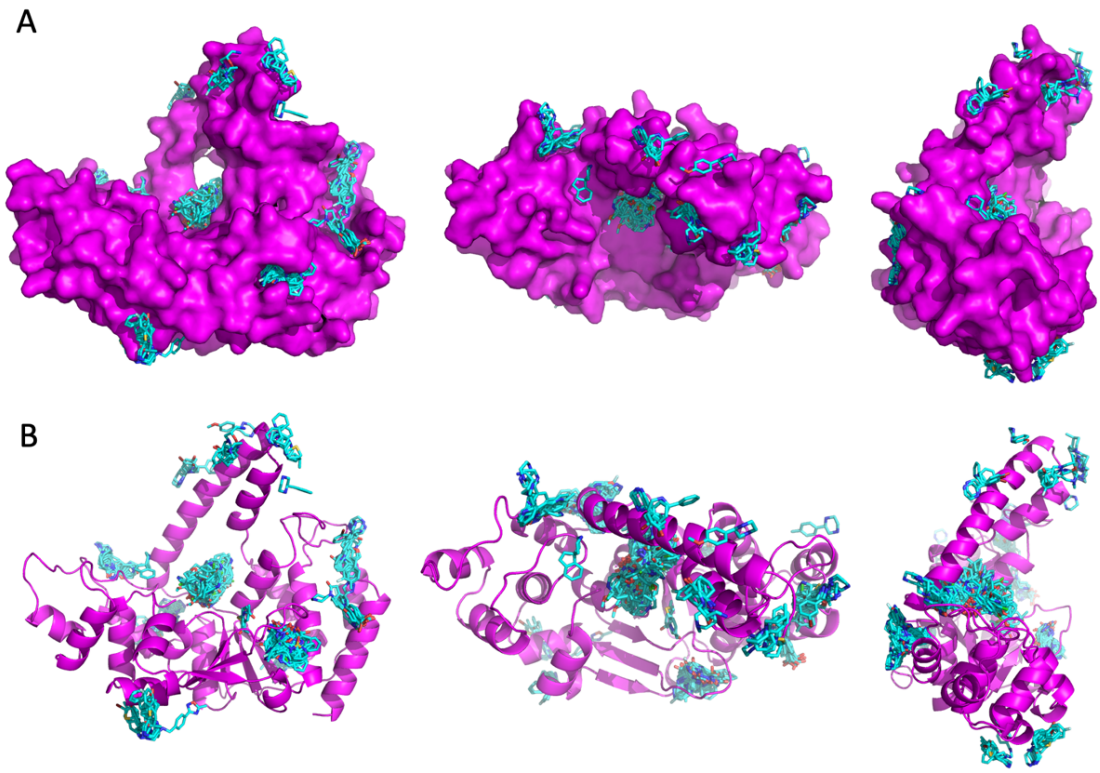


Figure 4.3 "Blind" docking results for *T. brucei* FEN model.

Docking was carried out using the BioNet library with AutoDock 4 and two metals in the active site. The top 20% of predictive binding modes are shown in cyan sticks. ***(A)*** *shows three orthogonal views of the protein surface with bound ligands.* ***(B)*** *represent the cartoon version.*

Validations of the predicted binding pockets in *Li*FEN and *Cn*FEN was assumed from the "blind" docking in the *Tb*FEN model. These validations were carried out considering the similar structure between the three microbial proteins and assuming the predicted binding pockets in *Li*FEN and *Cn*FEN correspond to similar ones predicted by *Tb*FEN.

Druggability analysis of the pockets

Once the pockets were predicted through POCASA and a "blind" docking with a grid covering the entire protein validated some of the binding pockets, druggability analysis of those pockets were carried out through the web server DoGSiteScorer (Volkamer *et al.*, 2012). The program gave a score of 0.69 for the central pocket. According to the web server DoGSiteScorer, a druggability score between 0 and 1 is returned. When the score is closer to 1, the more druggable the pocket is predicted to be. The druggability predictions are based on size and shape of the pocket, functional group and element descriptors, amino acid composition of the residues surrounding the pocket and a trained machine learning technique developed by Schmidtke and Barril (2010) setting a model trained and tested on druggable and non-druggable cavities. The accuracy of this computational tool to predict druggable pockets has been estimated to be of 88% (Volkamer *et al.*, 2012). The protein and the central pocket are visualized in Figure 4.4.

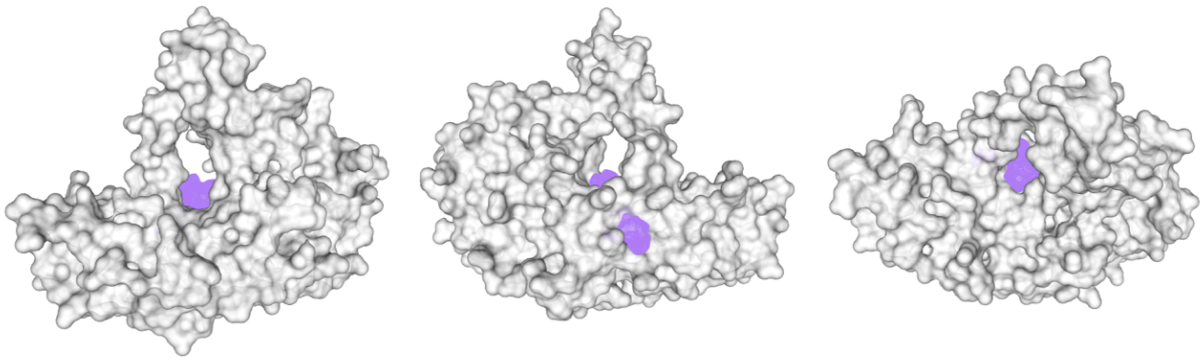


Figure 4.4 Druggability analysis of the central pocket in T. brucei FEN

Three approximately orthogonal views of the Tb*FEN protein presented as a grey surface. The DoGSiteScorer identified two main binding pockets with the central pocket shown as purple mesh.*

Given the results of the “blind” docking, pocket finder and druggability analysis, I concentrated the rest of the virtual screenings carried out in this project focusing on the central pocket, which may interfere with the DNA threading process and, therefore, inhibit the activity of the flap endonuclease in the microbial organisms.

In silico analysis of the BioNet library and microbial FEN docking results

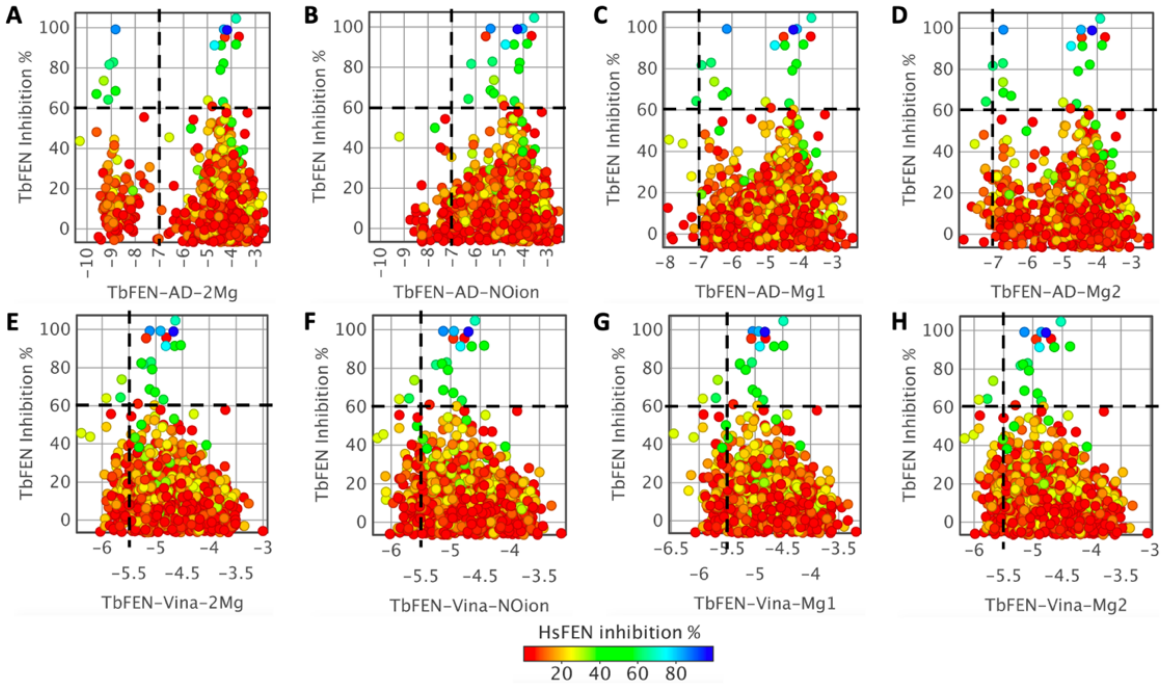
The figures below present an overview showing the range of binding scores obtained for each protein using AutoDock 4 and AutoDock Vina. The percentage of inhibition for each FEN protein determined in Chapter 6 is also plotted to allow assessment of each model/algorithms’ accuracy. Inspection of the results obtained from each model/algorithm combination was performed, and identification of potential inhibitors was carried out by establishing arbitrary binding energy and inhibition percentage cut-offs.

In reported studies, identification of virtual hits from different molecular targets and different compound libraries, arbitrary binding energies were set up between -5.0 to -9.0 kcal/mol (Srivastava *et al.*, 2018; Patra *et al.*, 2020). Nevertheless, another approach to establish a binding energy cut-off consists of performing a docking of a known inhibitor for a known binding pocket of the molecular target. The result obtained of this docking is then taken as the cut-off for the molecules considered as virtual hits in a compound library molecular docking. However, one of the main things to consider for this approach is the molecular docking programs and algorithms used (Macip *et al.*, 2021). Furthermore, to consider a compound as a potential inhibitor, arbitrary inhibition percentages of compounds have been set up between 17.5% to 90% according to different studies and depending of the stage of drug development (Hughes *et al.,* 2011; Mata-Cantero *et al.,* 2021; Tsegay *et al.,* 2021).

However, in practice, the arbitrary binding energy and inhibition percentage cut-offs are influenced by the available budget. It is important to consider in the virtual screening process to select compounds which are affordable, and can be delivered in a reasonable timeframe, and in a suitable format (Willems, 2020). The cost of a screening compound ranges from $2 to $120 depending on the vendor, the number of compounds and the quantity ordered. When the budget is limited, it is recommended

to order over 1000 compounds from a single source vendor. With this tactic, the price will be around $10 or less per compound, with an easier processing of physical compounds and lower shipping costs (Willems, 2020). Therefore, it is suggested to perform virtual screening from single-source vendors compound collection files than from chemical databases such as ZINC20.

For all the models using the BioNet library as ligands, the binding energy cut-off of -7 kcal/mol for all AutoDock 4 models and -5.5 kcal/mol for all AutoDock Vina models was set up. The molecules with a mean of -7 kcal/mol or lower from the 10 poses docked in the molecule in the AutoDock 4 models (or -5.5 kcal/mol or lower for the AutoDock Vina models) were considered as potential inhibitors (“virtual hits”) according to the virtual screening protocols. The established cut-off for percentage of inhibition was set at 60%, that means that all molecules with an inhibition percentage of 60% or above were considered as potential inhibitors according to physical assays.

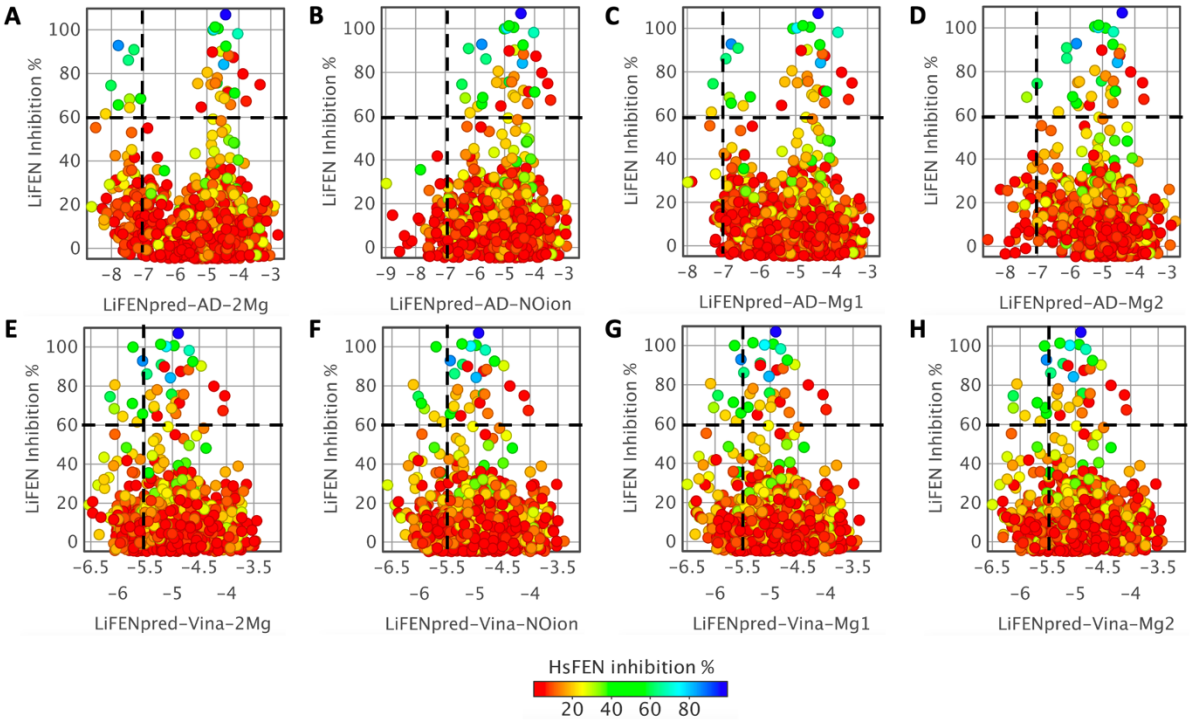


I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	162/1166	13.9%	70/1166	6.9%	14/1166	1.2%	19/1166	1.6%
Hits R&V/ Library	7/1166	0.6%	0/1166	0%	1/1166	0.1%	2/1166	0.2%
Hits R&V/ Real hits	7/21	33.3%	0/21	0%	1/21	4.8%	2/21	9.5%
Hits R&V/ Virtual hits	7/162	4.3%	0/70	0%	1/14	7.1%	2/19	10.5%
	AutoDock Vina (Scoring cut-off = -5.5 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	111/1166	9.5%	99/1166	8.5%	121/1166	20.4%	104/1166	8.9%
Hits R&V/ Library	3/1166	0.3%	3/1166	0.3%	3/1166	0.3%	3//1166	0.3%
Hits R&V/ Real hits	3/21	14.3%	3/21	14.3%	3/21	14.3%	3/21	14.3%
Hits R&V/ Virtual hits	3/111	2.7%	3/99	3%	3/121	2.5%	3/104	2.9%

Figure 4.5 Correlation of predicted docking energies and inhibitory activity in *T. brucei* flap endonuclease 1 with the BioNet library

Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from the BioNet library. The dots are colour coded according to the inhibition percentage of human FEN1. The top row ***(A)*** *to* ***(D)*** *shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row* ***(E)*** *to* ***(H)*** *shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank.* ***(I)*** *shows a numerical summary of the results. Virtual hits corresponds to the hits identified in the virtual screening from the entire library. Library represents the molecules of the entire library. Hits R&V corresponds to the hits identified in the virtual screening and physical assays. Real hits represents the hits identified only on physical assays. The best model predicting the real and virtual hits from the virtually hits is coloured in blue.*

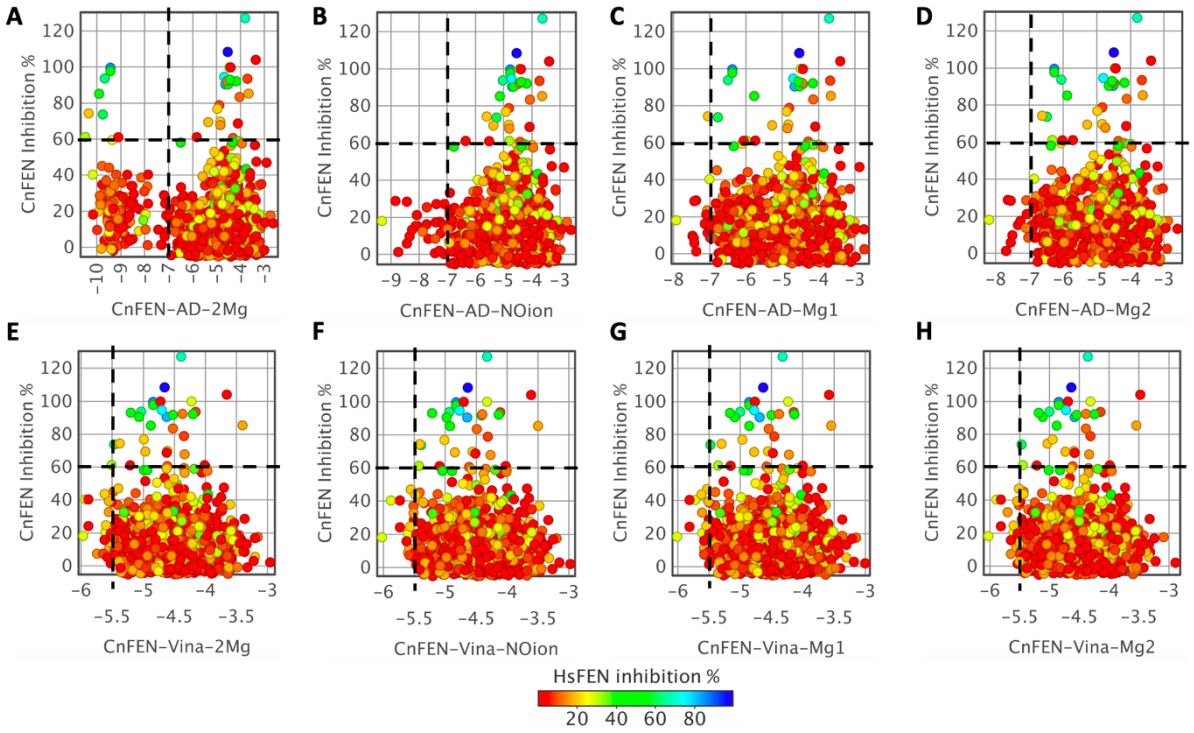


I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	113/1166	9.7%	57/1166	4.9%	20/1166	1.7%	34/1166	2.9 %
Hits R&V/ Library	9/1166	0.8%	0/1166	0%	2/1166	0.2%	2/1166	0.2%
Hits R&V/ Real hits	9/37	24.3%	0/37	0%	2/37	5.4%	2/37	5.4%
Hits R&V/ Virtual hits	9/113	8%	0/57	0%	2/20	10%	2/34	5.9%
	AutoDock Vina (Scoring cut-off = -5.5 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	35/1166	3%	38/1166	3.3%	40/1166	3.4%	39/1166	3.3%
Hits R&V/ Library	9/1166	0.8%	7/1166	0.6%	9/1166	0.8%	10/1166	0.9%
Hits R&V/ Real hits	9/37	24.3%	7/37	18.9%	9/37	24.3%	10/37	27%
Hits R&V/ Virtual hits	9/35	25.7%	7/38	18.4%	9/40	22.5%	10/39	25.6%

Figure 4.6 Correlation of predicted docking energies and inhibitory activity in *L. infantum* flap endonuclease 1 with the BioNet library

*Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from the BioNet library. The dots are colour coded according to the inhibition percentage of human FEN1. The top row **(A)** to **(D)** shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row **(E)** to **(H)** shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank. **(I)** shows a numerical summary of the results as described in Figure 4.5.*



I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	150/1166	12.9%	26/1166	2.2%	15/1166	1.3%	14/1166	12%
Hits R&V/ Library	9/1166	0.8%	0/1166	0%	1/1166	0.1%	0/1166	0%
Hits R&V/ Real hits	9/32	28.1%	0/32	0%	1/32	3.1%	0/32	0%
Hits R&V/ Virtual hits	9/150	6%	0/26	0%	1/15	6.7%	0/14	0%
	AutoDock Vina (Scoring cut-off = -5.5 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	23/1166	19.7%	17/1166	14.6%	22/1166	18.9%	19/1166	16.3%
Hits R&V/ Library	1/1166	0.1%	0/1166	0%	0/1166	0%	0/1166	0%
Hits R&V/ Real hits	1/32	3.1%	0/32	0%	0/32	0%	0/32	0%
Hits R&V/ Virtual hits	1/23	4.3%	0/32	0%	0/22	0%	0/19	0%

Figure 4.7 Correlation of predicted docking energies and inhibitory activity in *C. neoformans* flap endonuclease 1 with the BioNet library

*Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from the BioNet library. The dots are colour coded according to the inhibition percentage of human FEN1. The top row **(A)** to **(D)** shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row **(E)** to **(H)** shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank. **(I)** shows a numerical summary of the results as described in Figure 4.5.*

4.2.5 Selection and preparation of the HitExpansion compound library

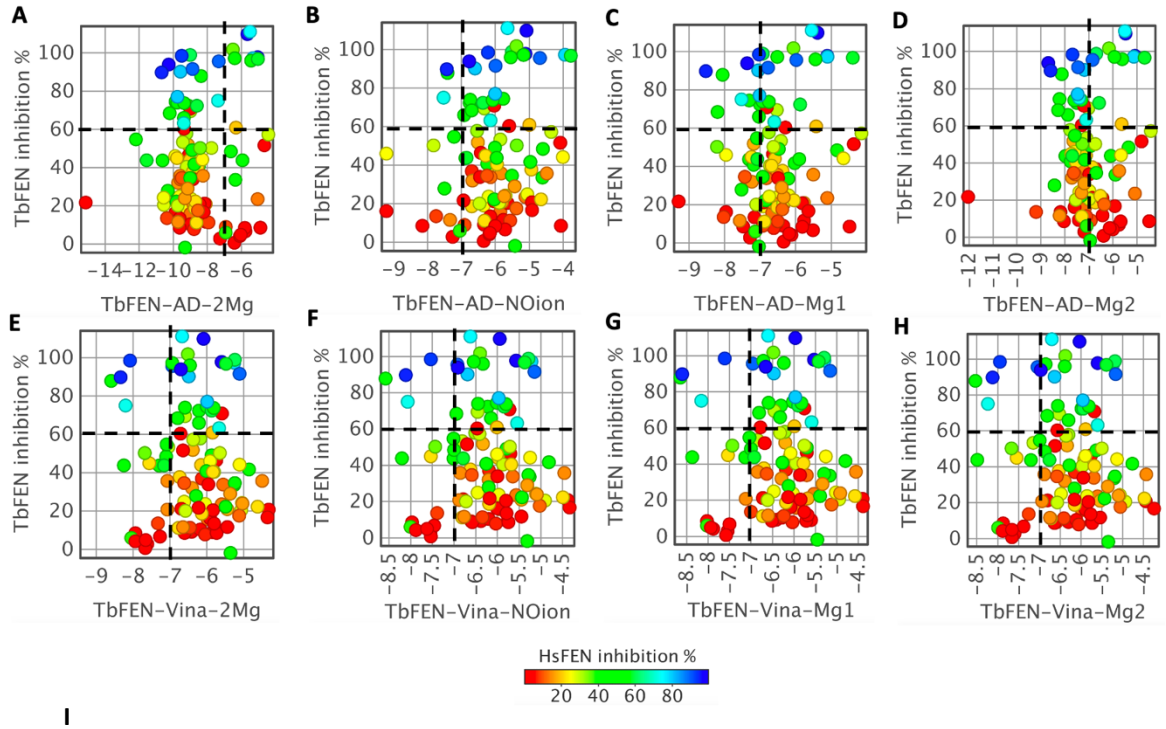
The HitExpansion library is a specific library of compounds designed in Sayers laboratory. This library contains 400 molecules from different chemical suppliers: Enamine (https://enamine.net), Chembridge (https://chembridge.com/screening-compounds/fragments/), Vitas (https://vitasmlab.biz) and Sigma (https://www.sigmaaldrich.com/GB/en/product/sial/dyna001). These molecules had been selected because they contain substructures which have been identified previously as microbial FEN inhibitors in the laboratory. Preparation of the molecules for docking was performed according to Section 2.10.1 from Chapter 2.

4.2.6 Molecular docking results and analysis

Molecular docking of the microbial FENs using the HitExpansion library as ligands was performed according to Section 2.10.3 in Chapter 2. The molecular docking results and analysis was carried out similarly from the Section 4.2.4 using a sdf file containing 400 structures in the library provided by Sayers laboratory.

In silico analysis of the HitExpansion library and microbial FEN docking results

The figures below present an overview showing the range of binding scores obtained for each protein using AutoDock 4 and AutoDock Vina. The percentage of inhibition for each FEN protein is also plotted to allow assessment of each model/algorithms' accuracy. Inspection of the results obtained from each model/algorithms' and identification of potential inhibitors was carried out by establishing arbitrary cut-offs in the same direction as the BioNet library analysis. A binding energy cut-off of -7 kcal/mol and inhibition percentage of 60% were set up in all models.

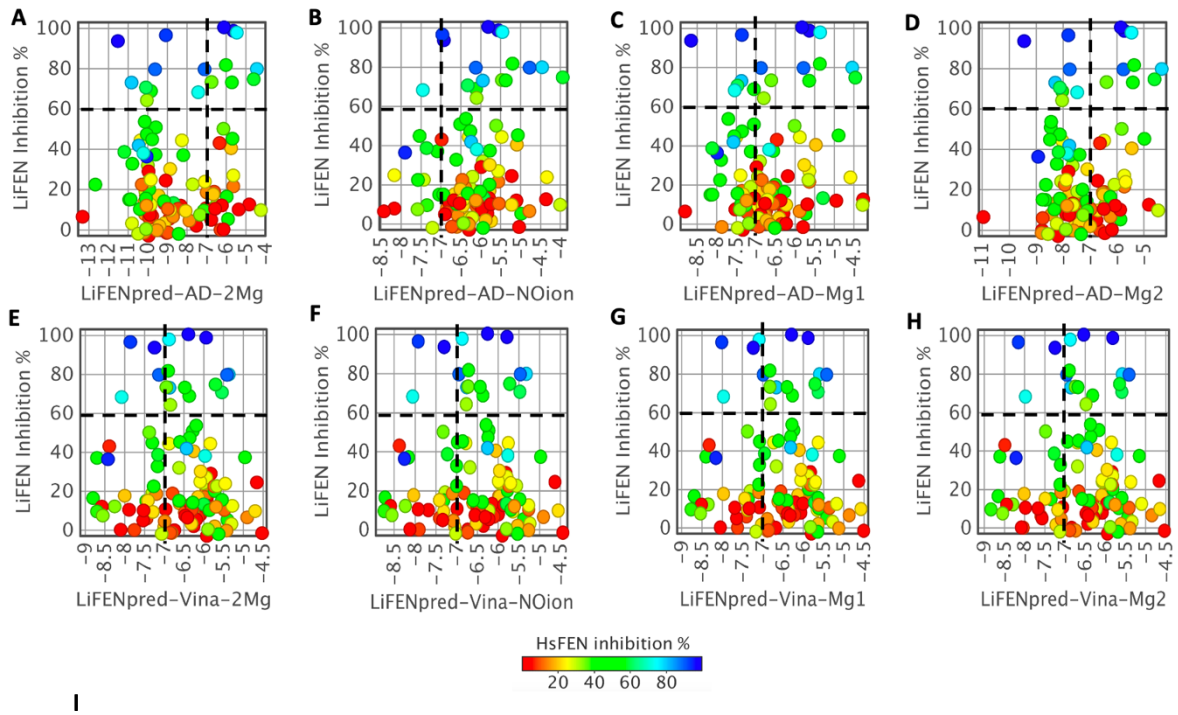


I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	83/126	65.9%	13/126	10.3%	32/126	25.4%	67/126	53.2%
Hits R&V/ Library	21/126	16.7%	3/126	2.4%	11/126	8.7%	18/126	14.3%
Hits R&V/ Real hits	21/30	70%	3/30	10%	11/30	36.7%	18/30	60%
Hits R&V/ Virtual hits	21/83	25.3%	3/13	23.1%	11/32	34.4%	18/67	26.9%
	AutoDock Vina (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	22/126	17.5%	18/126	14.3%	19/126	15.1%	19/126	15.1%
Hits R&V/ Library	4/126	3.2%	4/126	3.2%	4/126	3.2%	5/126	4%
Hits R&V/ Real hits	4/30	13.3%	4/30	13.3%	4/30	13.3%	5/30	16.7%
Hits R&V/ Virtual hits	4/22	18.2%	4/18	22.2%	4/19	21.1%	5/19	26.3%

Figure 4.8 Correlation of predicted docking energies and inhibitory activity in *T. brucei* flap endonuclease 1 with the HitExpansion library

*Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from some molecules of the HitExpansion library (126 from 400 molecules). The dots are colour coded according to the inhibition percentage of human FEN1. The top row **(A)** to **(D)** shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row **(E)** to **(H)** shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank. **(I)** shows a numerical summary of the results as described in Figure 4.5.*

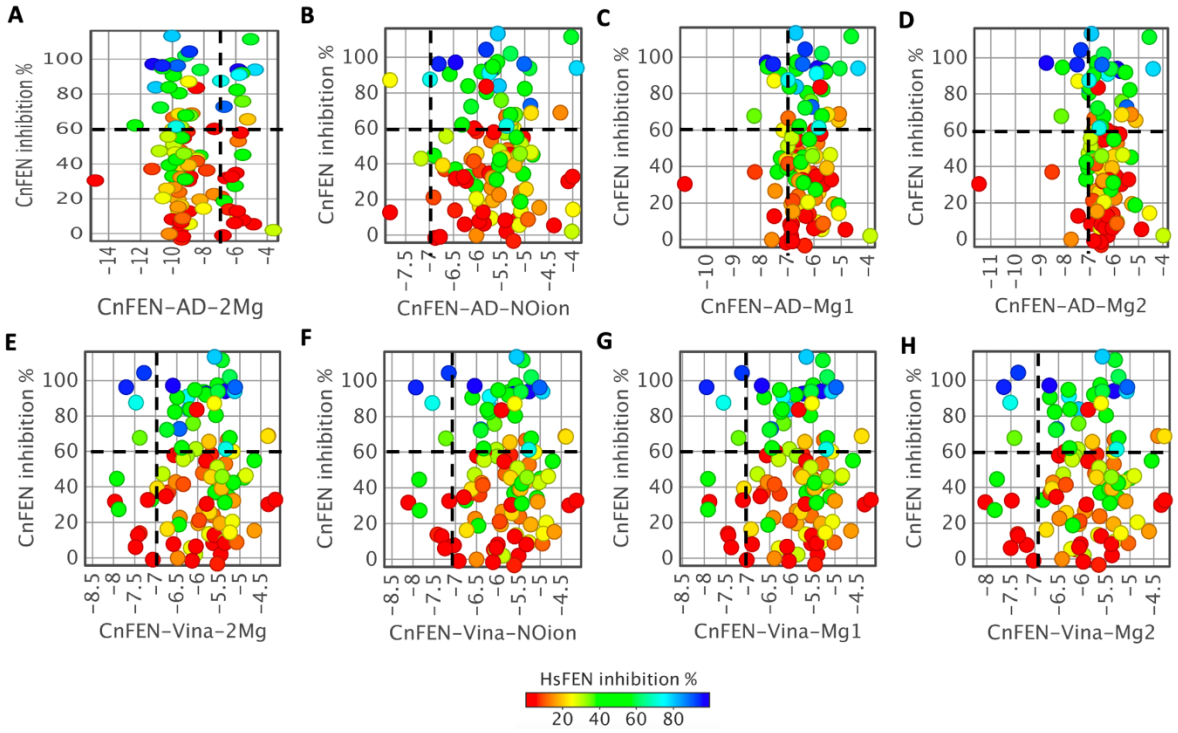


I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	83/126	65.9%	14/126	11.1%	40/126	31.7%	71/126	56.3%
Hits R&V/ Library	9/126	7.1%	1/126	0.8%	6/126	4.8%	8/126	6.3%
Hits R&V/ Real hits	9/17	52.9%	1/17	5.9%	6/17	35.3%	8/17	47.1
Hits R&V/ Virtual hits	9/83	10.8%	1/14	7.1%	6/40	15%	8/71	11.3%
	AutoDock Vina (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	31/126	24.6%	31/126	24.6%	32/126	25.4%	31/126	24.6%
Hits R&V/ Library	4/126	3.2%	3/126	2.4%	5/126	4%	3/126	2.4%
Hits R&V/ Real hits	4/17	23.5%	3/17	17.6%	5/17	29.4%	3/17	17.6%
Hits R&V/ Virtual hits	4/31	12.9%	3/31	9.7%	5/32	15.6%	3/31	9.7%

Figure 4.9 Correlation of predicted docking energies and inhibitory activity in *L. infantum* flap endonuclease 1 with the HitExpansion library

*Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from some molecules of the HitExpansion library (126 from 400 molecules). The dots are colour coded according to the inhibition percentage of human FEN1. The top row **(A)** to **(D)** shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row **(E)** to **(H)** shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank. **(I)** shows a numerical summary of the results as described in Figure 4.5.*



I

	AutoDock 4 (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	80/126	63.5%	4/126	3.2%	18/126	14.3%	18/126	14.3%
Hits R&V/ Library	27/126	21.4%	2/126	1.6%	7/126	5.6%	10/126	7.9%
Hits R&V/ Real hits	27/38	71%	2/38	5.3%	7/38	18.4%	10/38	26.3%
Hits R&V/ Virtual hits	27/80	33.8%	2/4	50%	7/18	38.9%	10/18	55.6%
	AutoDock Vina (Scoring cut-off = -7 kcal/mol)							
	2 Metal ions		NO metal ions		Magnesium 1		Magnesium 2	
Virtual hits/ Library	13/126	10.3%	12/126	9.5%	14/126	11.1%	13/126	10.3%
Hits R&V/ Library	4/126	3.2%	4/126	3.2%	4/126	3.2%	4/126	3.2%
Hits R&V/ Real hits	4/38	10.5%	4/38	10.5%	4/38	10.5%	4/38	10.5%
Hits R&V/ Virtual hits	4/13	30.8%	4/12	33.3%	4/14	28.6%	4/13	30.8%

Figure 4.10. Correlation of predicted docking energies and inhibitory activity in *C. neoformans* flap endonuclease 1 with the HitExpansion library

*Results are presented using Datawarrior to plot the percentage of inhibition and predicted binding energies from some molecules of the HitExpansion library (126 from 400 molecules). The dots are colour coded according to the inhibition percentage of human FEN1. The top row **(A)** to **(D)** shows the results of the AutoDock 4 predictions considering input models containing two, zero, or one metal ion inside site one or two, respectively. Similarly, the bottom row **(E)** to **(H)** shows the same analysis carried out with AutoDock Vina. Arbitrary cut-offs positions for binding affinity and activity are presented as dotted lines. Compounds in the upper left quadrant show agreement between experimentally determined inhibition and predicted binding rank. **(I)** shows a numerical summary of the results as described in Figure 4.5.*

4.2.7 Final analysis of the molecular docking results

Carrying out a general inspection of the results obtained from each one of the models performed, the best model for predicting potential inhibitors from a library corresponds to the AutoDock algorithm using a Mg^{2+} ion in site number 2. This model has been selected because it is the model that has predicted the majority of the molecules which have been previously identified as inhibitors from the physical screening (as in the case of the *Cn*FEN with the HitExpansion library). By selecting the best out of the eight models developed, a summary of the data obtained from the 3 proteins using the BioNet and HitExpansion libraries is shown in Table 6.1.

Table 4.1 Final comparison of hit rate from different microbial FENs and libraries*

	BioNet					
Protein	*Tb*FEN		*Li*FEN		*Cn*FEN	
Model	AD - Magnesium 2		Vina – 2 Metal ions		AD – Magnesium 1	
Virtual hits/ Library	19/1166	1.6%	35/1166	3%	15/1166	1.3%
Hits R&V/ Library	2/1166	0.2%	9/1166	0.8%	1/1166	0.1%
Hits R&V/ Real hits	2/21	9.5%	9/37	24.3%	1/32	3.1%
Hits R&V/ Virtual hits	2/19	10.5%	9/35	25.7%	1/15	6.7%
	HitExpansion					
Protein	*Tb*FEN		*Li*FEN		*Cn*FEN	
Model	AD – Magnesium 1		Vina – Magnesium 1		AD – Magnesium 2	
Virtual hits/ Library	32/126	25.4%	32/126	25.4%	18/126	14.3%
Hits R&V/ Library	11/126	8.7%	5/126	4%	10/126	7.9%
Hits R&V/ Real hits	11/30	36.7%	5/17	29.4%	10/38	26.3%
Hits R&V/ Virtual hits	11/32	34.4%	5/32	15.6 %	10/18	55.6%

**The criteria used to select the best model is based on the hit rate obtained from molecules identified in virtual and real screening versus the molecules identified only in the virtual screening.*

Interestingly, the best model to predict potential inhibitors from the virtual screening, which were identified by physical assays, differs depending on the molecule target and the library of compounds to screen. This could be a matter of interest considering the different factors involved in fragment-based drug design. A couple of those factors are the electronic charges of the metal ions in the active site, and the steric conformation between the small molecules and the residues of the protein surrounding the binding pockets.

Nevertheless, most of the models and algorithms used in this work are equally good at predicting hits as those reported in the literature for large-scale docking projects. For example, a virtual screening targeting the AmpC β-lactamase using a library of 99 million molecules, selecting 44 virtual hits considering the top 0.00001% molecules with the highest binding energy score. The authors experimentally tested these 44 molecules of the virtual hits and found that 5 were actually inhibitors (11%) (Lyu *et al.*, 2019). Furthermore, another virtual screening targeting the A2 adenosine receptor registered a hit rate of 41%, proving that 23 out of 56 virtual hits inhibited the activity of this enzyme. The 56 virtual hits selected on this study were selected within the top 0.05% of molecules with the highest binding energy score from a library of 4 million compounds (Katritch *et al.*, 2010).

Before proceeding to the analysis of the *in silico* results, screening two libraries targeting three microbial FENs, it is important to point out that the results obtained from the predicted protein structures should be taken with precaution. Despite the advances in artificial intelligence and the big impact of prediction algorithms, specially AlphaFold 2, towards structural biology, predicted structures may not be reliable compared to structures obtained from experimental assays and, therefore, should always be used with caution. The use of predicting algorithms face some limitations in terms of elucidating a protein structure with the same conformation as the one obtained experimentally. Those limitations are: (1) difficulty to predict structures of heteromers and multi-domain complexes; (2) cannot predict post-translational modifications, for example, glycosylation, methylation, lipidation, etcetera; (3) Limited predictions for intrinsically disordered protein regions; (4) Low accuracy predicting structural dynamics (presence of water and other molecules that can be found in solution) and effects of point mutations on structural stability (Robinson, 2022).

Analysis of predicted poses for selected inhibitors

The last step of post-docking analysis corresponds to the analysis of predicted poses for selected inhibitors. Once the virtual hits have been identified and selected through the binding energy scores and/or the inhibition percentage of the three proteins, visualization of those hits with their predicted poses was carried out. On this last step, the visualization was performed through the software Pymol using the docked macromolecules in pdb files and the output generated by PyRx in sdf files, as observed in Figures 4.11 to 4.13.

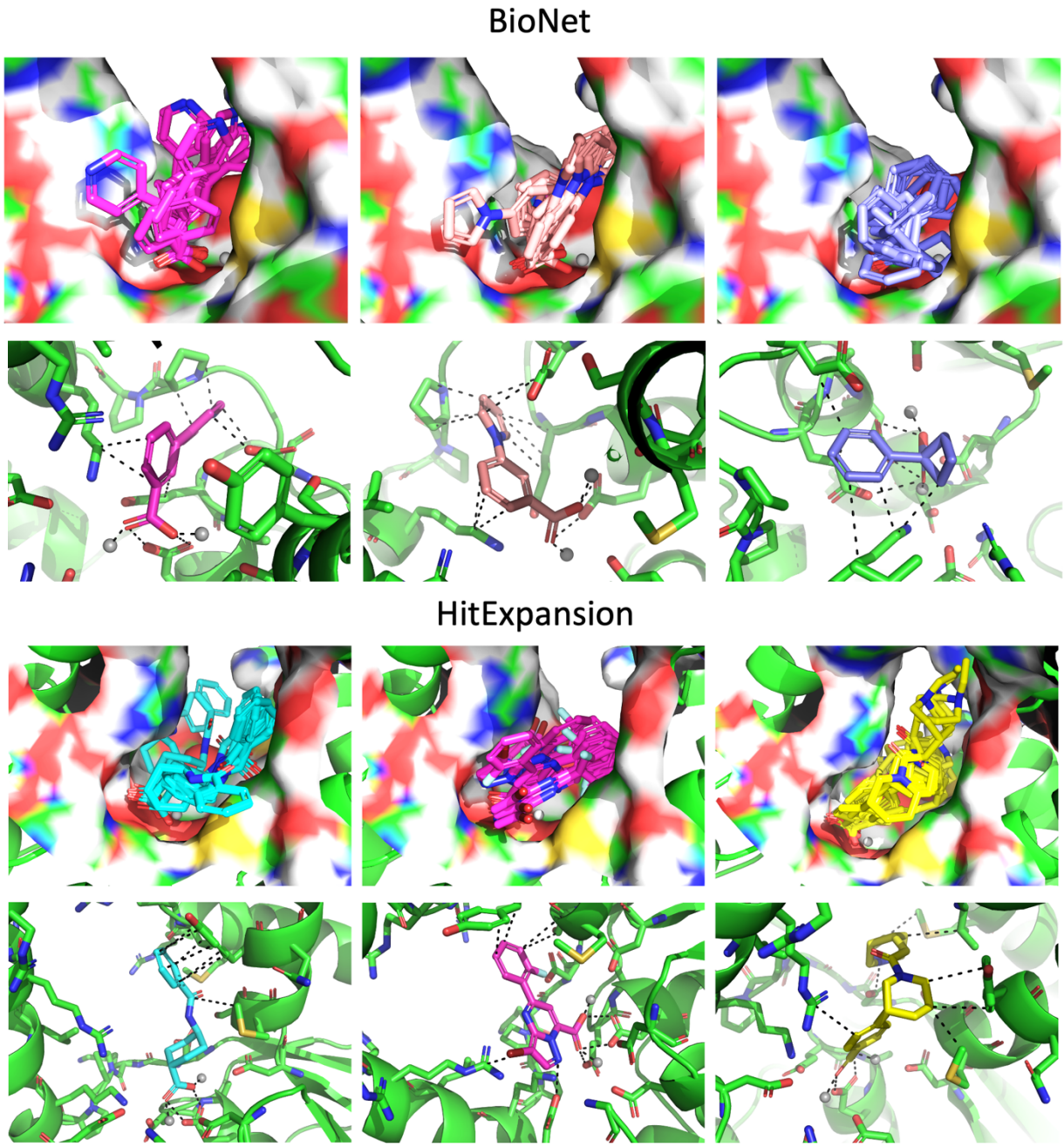


Figure 4.11 Visual analysis of the predicted inhibitors in *T. brucei* FEN

The top 3 molecules with the highest binding energy score are depicted. The top rows depict the 3 molecules of each library and the 10 predicted poses according to the model/algorithm AutoDock 4 with two metal ions in the active site (presented as sticks) and the residues surrounding the compounds as surface. The bottom rows show the 3 molecules of each library and the pose with the highest binding energy score. The residues within a distance of 4 Å of the compound are represented as sticks and dotted lines between the molecule and the residues of the protein correspond to hydrogen bond or Van der Waals contacts. Interactions of the molecules with the magnesium ions (coloured in grey) are also presented as dotted lines. Molecules shown in the figure are FS-2599, PS-4299 and PS-4233, and G1016, G1048 and G1058 for the BioNet and HitExpansion library, respectively.

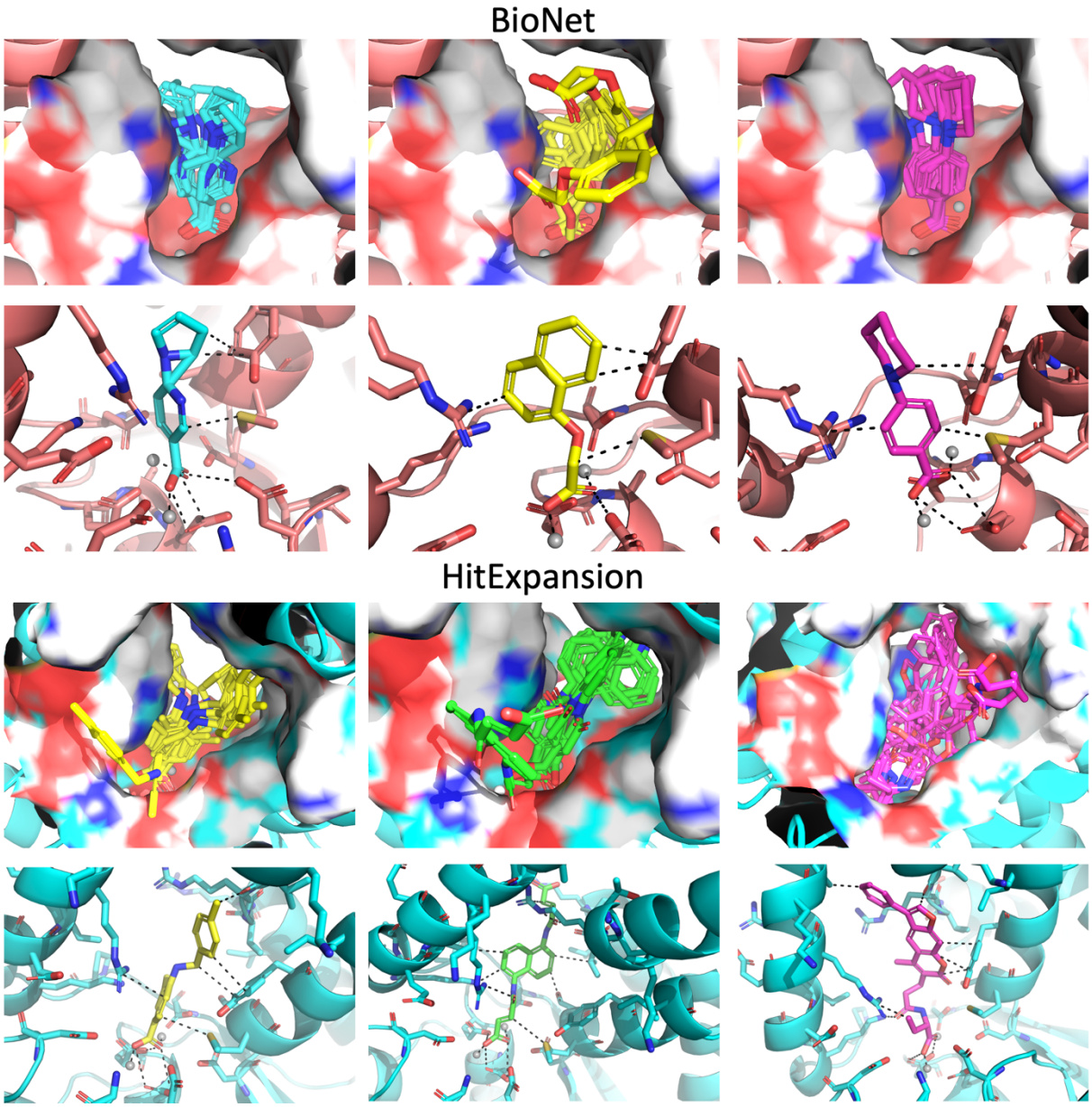


Figure 4.12 Visual analysis of the predicted inhibitors in *L. infantum* FEN

The figure shows the top 3 molecules with the highest binding energy score for each library. The top rows depict the 3 molecules of each library and the 10 predicted poses as in Figure 4.10. The bottom rows show the 3 molecules of each library and the pose with the highest binding energy score. Visual inspection was carried out on the same basis as in Figure 6.10. Molecules shown in the figure are 5Y-0705, 2M-936 and BD-0213 for BioNet library, respectively. G1010, G2079 and G2107 for HitExpansion library.

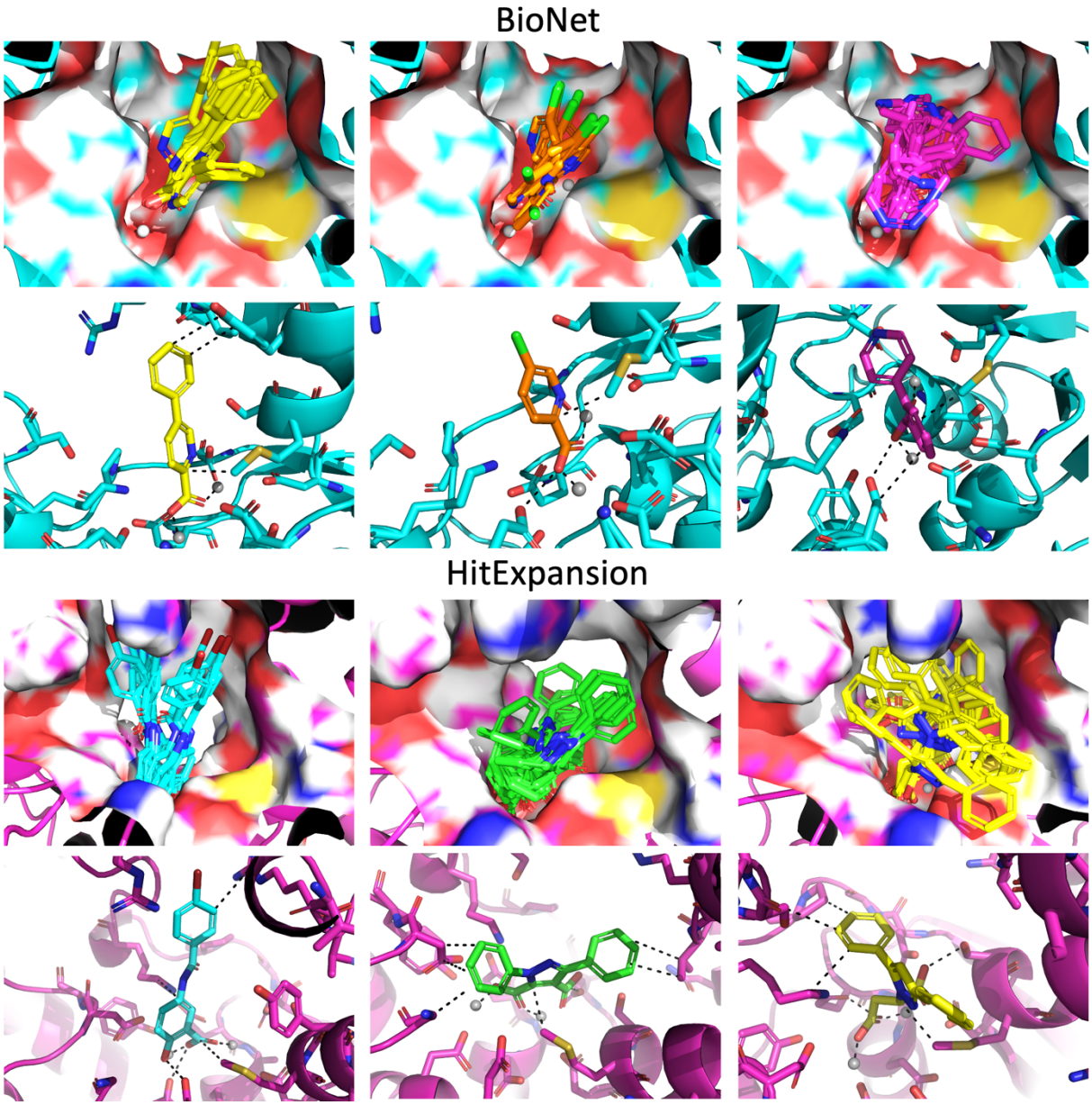


Figure 4.13 Visual analysis of the predicted inhibitors in *C. neoformans* FEN

The figure shows the top 3 molecules with the highest binding energy score for each library. The top rows depict the 3 molecules of each library and the 10 predicted poses as in Figure 4.10. The bottom rows show the 3 molecules of each library and the pose with the highest binding energy score. Visual inspection was carried out on the same basis as in Figure 4.10. Molecules shown in the figure are PS-5576, BB-0607 and PS-5707 for BioNet library, respectively. G1030, G2131 and G2134 for HitExpansion library.

For illustration and with the aim to validate the molecular docking work from this project, the crystallized N-hydroxy urea molecule from 5FV7.pdb was retrieved and later docked to the *Tb*FEN. A comparison of the crystal structure of the N-hydroxy urea molecule bound to the human FEN and a molecular docking of the same molecule targeting a *Tb*FEN is shown in Figure 4.14.

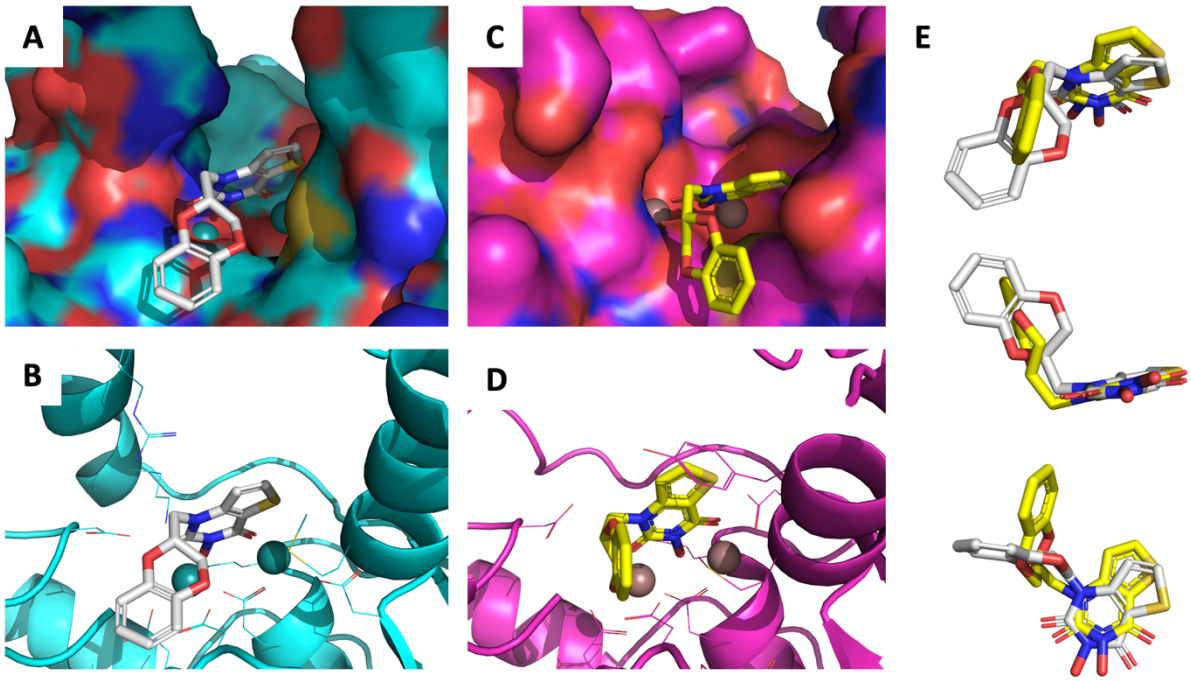


Figure 4.14 Comparison of N-hydroxy urea molecule bound to human and *T. brucei* FENs

Figure A and B show molecular docking result of the N-hydroxy urea molecule with the Tb*FEN. Figure A presents the* Tb*FEN protein as surface (cyan), the magnesium ions (dark cyan) as spheres and N-hydroxy urea molecule (white C atoms) as sticks. Figure B shows the TbFEN protein as cartoon with the residues in close proximity to the docked molecule as lines. Figure C and D show the crystal structure of the N-hydroxy urea molecule (yellow C atoms) bound to the human FEN (5FV7.pdb, magenta) in the presence of two magnesium ions (dark magenta). Figures C and D are presented on the same basis as A and B with the surface removed and side chains within 4 Å shown as lines. Figure E shows three orthogonal views of the N-hydroxy urea docked poses in* Tb*FEN (white) superimposed on the experimental human complex (yellow).*

The N-hydroxy urea derivative shown in Figure 7.1 binds to the active site via the magnesium ions. The docking approach used in this work was able to predict binding of this molecule to the metals despite the lack of visible density in the residues above the active site of the solved human structure (the flexible DNA-binding residues of the helical arch).

In conclusion, *in silico* docking was executed using a variety of models with AutoDock 4 and AutoDock Vina and their performance was evaluated with the aid of data from biophysical inhibitor assays described in Chapter 6.

CHAPTER 5: PROTEIN CRYSTALLIZATION AND STRUCTURAL DETERMINATION OF MICROBIAL FLAP ENDONUCLEASES

5.1 Introduction

Structural biology has become significant in the comprehension of the molecular basis of biological processes during the cell cycle, as well as normal physiology and pathological conditions. Such studies have become a major objective for the discipline of structural biology. The understanding of protein-protein interactions, the identification of enzyme catalytic mechanisms and allosteric regulations, the development of drugs using structural and computational methods, the comprehension of protein flexibility through molecular dynamics simulations, and the prediction of protein structures using homology modelling are just a few examples of the many research fields that use structural data to pursue very different goals (Ronda *et al.*, 2015).

The capacity to understand and manipulate protein function can be improved when the structure of the protein has been elucidated. Obtaining accurate protein structures can significantly improve the comprehension of biochemical mechanisms, leading to the development of new therapies on a faster pace, often enhanced by the use of computer-aided drug design (CADD) (Leelananda & Lindert, 2016). For example, by determining a protein structure, CADD can dramatically reduce the number of small molecules to be screened experimentally, by excluding ligands with a low binding score according to computational predictions (Seffernick & Lindert, 2020).

X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM) are some of the experimental methods used to determine the structures of proteins at resolutions of 3 Å or higher and will likely remain fundamental to structural biology for some time to come (Nwanochie & Uversky, 2019; Würz *et al.*, 2017; Ilari & Savino, 2008) despite the recent advances made in accurate theoretical structure prediction using algorithms such as AlphaFold2 and Phyre2 (Jumper *et al.*, 2021; Kelley *et al.*, 2015). From the methods previously mentioned, X-ray crystallography is the most frequently employed to resolve protein structures to the atomic level (Higgins & Lea, 2017). In order to produce high-quality crystals, the

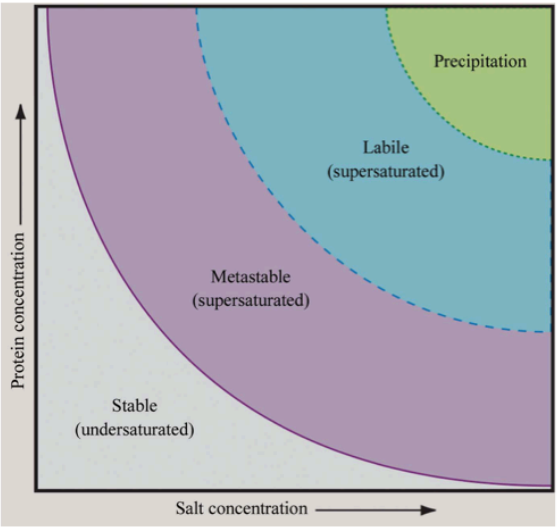


Figure 5.1 Phase diagram for the crystallization of macromolecules. *The solubility diagram divides the undersaturation and saturation regions by the line representing maximum solubility at specific concentrations of a precipitant. Crystals can only be grown from a saturated solution, therefore the way to generate protein crystals is through creating the solution supersaturated in the protein of interest. The figure was reproduced with permission of the International Union of Crystallography from McPherson & Gavira, 2014.*

crystallisation workflow involves the expression and purification of highly homogeneous protein and a series of screening steps, followed by optimization of hit conditions where crystals started to grow (Smyth & Martin, 2000). This involves small and systematic variation of the conditions such as pH, concentration of precipitants, co-factors, temperature and protein concentration.

Crystallization of proteins, nucleic acids, and biological complexes such as viruses depends on the concentration of the macromolecule to crystallized in the solution, which may reach a specific phenomenon known as “supersaturation”. Supersaturation is achieved by addition of mild precipitating agents such as neutral salts or polymers, as well as manipulating different factors such as temperature, ionic strength, and pH (McPherson & Gavira, 2014). Supersaturation is a non-equilibrium state that occurs under certain chemical and physical conditions. In this situation, the concentration of protein in solution exceeds its solubility limit. When a protein approaches its solubility limit, it will usually just precipitate. However, under specific conditions, it can enter in a supersaturated meta-stable phase, and crystals form from this phase (McCoy, 2009).

Many different techniques are used to crystallise macromolecules. From those techniques, vapor diffusion is the most commonly used, either suspended over a larger volume (hanging drop technique) or sitting in a well (sitting drop technique) (McCoy, 2009). Typically, a drop of the protein solution and a drop of the screening solution are mixed. The reservoir is pipetted with the same screening solution. Since the

concentration of screening solution in the protein drop is lower than in the reservoir, water diffuses from the protein drop into the reservoir over time until the concentrations in the drop and the reservoir are equal. Protein crystals can develop in the protein drop as the protein and precipitant concentrations rise (Dessau & Modis, 2011).

Protein crystal formation must be done in a physical setup that enables the researcher to modify the mother liquor characteristics and the protein solubility. Crystallization trials are performed with volumes ranging from microlitres to nanolitres, using plastic multichambered trays for sitting drops and hanging drops, plexiglass buttons for dialysis or microdrops submerged in oil (McPherson & Gavira, 2014).

Furthermore, after the formation of the protein crystal, the crystal is retrieved in a process called “looping”, soaked in a cryoprotectant, and analysed using an X-ray source, which in recent years is more often an X-ray beamline at a synchrotron. The data collected from the X-ray diffraction pattern allows measurement of the intensities of the diffracted waves dispersed from a set of planes that we can see slicing through the crystal in all directions. From these intensities, it is possible to derive the amplitudes of the scattered waves. However, during the experiment the phase information is lost, which describes how the waves are compensated when the researcher combines them to reconstruct an image of the molecule. This is commonly known as the "phase problem" (Taylor, 2010).

In most cases in protein crystallography, the phase problem is solved either by using the atomic coordinates of a structurally similar protein (molecular replacement) or by finding the positions of heavy atoms intrinsic to the protein or added (known as experimental phasing, using methods such as multiple isomorphous replacement (MIR), multiple isomorphous replacement with anomalous scattering (MIRAS), single isomorphous replacement (SIR), single isomorphous replacement with anomalous scattering (SIRAS), multiwavelength anomalous diffraction/dispersion (MAD), single-wavelength anomalous diffraction/dispersion (SAD) or combinations of these; Taylor, 2010).

Crystallization trials for *T. brucei, L. infantum* and *C. neoformans* FENs are covered in this chapter. The enzymes share more than 50% identity with the human FEN1 and

therefore are extremely likely to share the same architecture which is composed of four parts: the N-terminal region, the C-terminal region, a helical arch, and the active site.

The results obtained from the crystal trials will enable us to compare the structure of our pathogen FENs with the human orthologue. The purpose of this experimental work was to determine high-resolution X-ray structures of the relevant proteins to provide a sound base for structure-based drug design. Also, if crystals suitable for soaking experiments can be obtained, these could be used for fragment-based structure screening. Suitable data could greatly improve the likelihood of successful vHTS as well as ligand-based design approaches.

RESULTS AND DISCUSSION

5.2 Crystallization of *Tb*FEN-D183K ΔC-341

*Tb*FEN-D183K ΔC-341 is an inactive "enzyme", as it has been engineered to that one of the active site magnesium ions replaced by the epsilon amine function of lysine. Crystallization trials with purified *Tb*FEN-D183K ΔC-341 were set up in six commercial screens: JCSG, PACT, Morpheus, Proplex (Molecular Dimensions), MPD (NeXtal Biotechnologie), and Natrix (Hampton Research), each containing 96 different reservoir conditions. Initial crystallization screens were set up with 50 µL buffer reservoirs in Swissci 96-Well 3-Drop Plates from Molecular Dimensions. The protein sample and buffer condition were seeded in each well using the Mosquito Crystal robot (TTP Labtech) using 100 nL of protein and 100 nL of buffer. For each condition, the protein was seeded with two DNA substrates (5ov4 and JT) and without DNA substrate.

The initial concentration of *Tb*FEN-D183K ΔC-341 for the crystallization trials without DNA substrate was between 11 to 13 mg/mL. The initial concentration of the protein for the crystallization trials with DNA substrate was determined considering a 1:1 molar ratio of protein:DNA at 0.25 mM.

From the initial crystallization trials of the six commercial screens, it was possible to grow crystals and further optimization of the conditions was carried out. Conditions

from the initial screening where crystallization occurs can be found in Table 5.1. Pictures of some of the crystals grown for this protein are depicted in Figure 5.2.

Table 5.1 Conditions in which *Tb*FEN D183K ΔC-341 crystals were obtained from the six commercial screens and the reagents that integrate these conditions.

Commercial screen	Condition	Reagents*
JCSG	A2	0.1 M sodium citrate pH 5.5, 20% (w/v) PEG 3000
JCSG	A3	0.2 M ammonium citrate dibasic, 20% (w/v) PEG 3350
JCSG	B2	0.1 M MIB pH 5, 25% (w/v) PEG 1500
JCSG	B9	0.1 M citrate pH 5.0, 20% (w/v) PEG 6000
JCSG	C12	10% (v/v) 1M Tris pH 7.6, 10% (w/v) PEG 1000
JCSG	D3	0.1 M MMT pH 6.0, 25% (w/v) PEG 1500
JCSG	E8	1 M ammonium phosphate dibase, 0.1 M sodium actetate 4.5
JCSG	H11	0.2 M $MgCl_2 \cdot 6H_2O$, 0.1 M Bis-Tris pH 5.5, 25% (w/v) PEG 3350
MPD	C9	0.2 M potassium sodium tartrate tetrahydrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350
MPD	F11	1 M succinic acid, 0.1 M HEPES pH 7, 1% (w/v) PEG 2000 MME
MPD	H2	0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 30% (w/v) MPD
PACT	G11	0.2 M sodium citrate tribasic dihydrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350

** MIB: malonic acid, imidazole, and boric acid (2:3:3 molar ratio) pH adjusted with HCl. MMT: malic acid, MES (2-(N-morpholino)ethanesulfonic acid) and Tris (1:2:2 molar ratio) pH adjusted with HCl.*

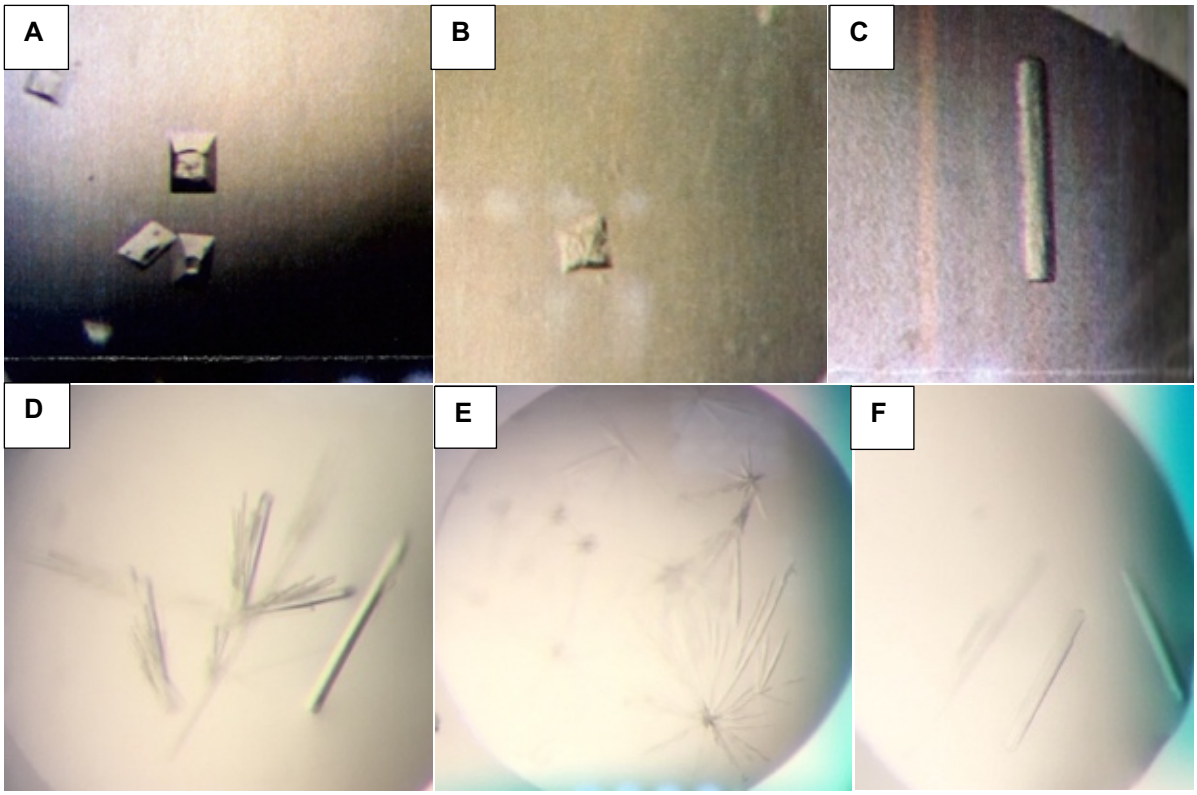


Figure 5.2 *Tb*FEN D183K ΔC-341 crystals from initial screening and optimized conditions

(A) *Initial crystallization trial using the JCSG screen that formed* squared like *crystals. Condition B2 containing 0.1 M MIB pH 5, 25% (w/v) PEG 1500 and JT DNA.* ***(B)*** *Initial crystallization trial using the JCSG screen that formed an irregular shape crystal. Condition D3 containing 0.1 M MMT pH 6.0, 25% (w/v) PEG 1500 and JT DNA.* ***(C)*** *Initial crystallization trial using the JCSG screen forming a* rod-*shape like crystal. Condition C12 containing 10% (v/v) 1M Tris pH 7.6, 10% (w/v) PEG 1000 and JT DNA.* ***(D)*** *Optimization of A3 JCSG screen. Condition C12 containing 0.2 M ammonium citrate dibasic (pH adjusted to 6.5), 22% (w/v) PEG 3350 yielded crystals. Protein sample concentration at 13 mg/mL.* ***(E)*** *Optimization of A3 JCSG screen. Condition D12 containing 0.2 M ammonium citrate dibasic (adjusted to pH 6.5), 24% (w/v) PEG 3350 and 5ov4 DNA.* ***(F)*** *Optimization of A3 JCSG screen. Condition E3 containing 0.2 M ammonium citrate dibasic (adjusted to pH 5), 22% (w/v) PEG 3350 and JT DNA. Protein and DNA substrate were on a 1:1 molar ratio (duplex structure on 5ov4 and flap structure on JT) at a concentration of 0.25 mM in initial screen conditions and 0.3 mM in optimized conditions, except for the condition in Figure 5.2D in which the crystal does not contain DNA substrate.*

5.3 Crystallization of *Li*FEN-WT

5.3.1 Crystallization of *Li*FEN-WT without DNA substrate

Crystallization assays were set up in two commercial screens, JCSG and PACT, each containing 96 different reservoir conditions and distributed according to their pH. The initial concentration for the crystallization trials without substrate was at 16.5 mg/mL. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the two commercial screens, four wells in the high pH condition and four wells in the low pH condition produced microcrystal structures, which are described in Table 5.2.

Table 5.2 Conditions in which crystals were obtained from the two commercial screens and the reagents that integrate these conditions.

Commercial screen	Condition	Reagents*
High pH initial screening		
PACT	C8	0.2 M NH_4Cl, 0.1 M HEPES pH 7, 20% PEG 6000 (w/v)
PACT	C4	0.1 M PCTP pH 7, 25% PEG 1500 (w/v)
PACT	D4	0.1 M MMT pH 7, 25% PEG 1500 (w/v)
PACT	D7	0.2 M NaCl, 0.1 M Tris pH 8, 20% PEG 6000 (w/v)
Low pH initial screening		
PACT	A2	0.1 M SPG pH 5, 25% PEG 1500 (w/v)
JCSG	B9	0.1 M citrate pH 5, 20% PEG 6000 (w/v)
JCSG	C1	0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2, 20% PEG 8000 (w/v)
JCSG	H11	0.2 M $MgCl_2$, 0.1 M Bis-Tris pH 5.5, 25% PEG 3350

** PCTP: sodium propionate, sodium cacodylate trihydrate, Bis-Tris propane: pH 4.0-9.5. MMT: DL-malic acid, MES monohydrate, Tris: pH 4.0-9.0. SPG: succinic acid, sodium phosphate monobasic monohydrate, glycine: pH 4.0-10.0.*

To obtain a higher quality of crystals, conditions C8, A2, C1, and H11 were optimized by varying the concentration of precipitant in 2 percentage and varying the pH by 0.5, considering the initial concentration of precipitant and pH at the middle of the matrix.

The concentration of protein in the optimization of crystal trials was decreased to 12 mg/mL. Pictures of the crystals obtained from the optimization of conditions C1 and C8, as well as the crystals from the initial screening C8 and D4, are visualized in Figures 5.3.

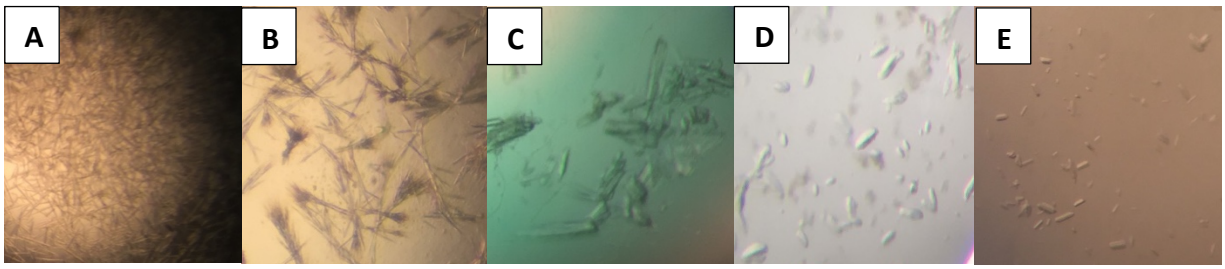


Figure 5.3 *Li*FEN crystals from initial screening and optimized conditions

(A) *Optimization of C1 JCSG screen that formed* Li*FEN needle-like crystals. Condition H2 containing 0.1 M phosphate/citrate pH 4.5, 24% PEG 8000 (w/v). Protein sample concentration at 12.8 mg/mL.* ***(B)*** *Optimization of C1 JCSG screen that formed* Li*FEN needle-like crystals. Condition C8 containing 0.1 M NaCl, 0.05 M, $MgCl_2$, 0.1 M phosphate/citrate pH 5, 22% PEG 8000 (w/v). Protein sample concentration at 12.8 mg/mL.* ***(C)*** *Initial crystallization trial using the PACT screen that formed* Li*FEN rod-shaped crystals. Condition C8 containing 0.2 M NH_4Cl, 0.1 M HEPES pH 7, 20% PEG 6000 (w/v). Protein sample concentration at 16.5 mg/mL.* ***(D)*** *Initial crystallization trial using the PACT screen that formed* Li*FEN rod-shaped crystals. Condition D4 containing 0.1 M MMT pH 7, 25% PEG 1500 (w/v). Protein sample concentration at 16.5 mg/mL* ***(E)*** *Optimization of C8 PACT screen that formed* Li*FEN rod-shaped crystals. Condition G1 containing 0.1 M HEPES pH 6, 22% PEG 6000 (w/v). Protein sample concentration at 11.4 mg/mL.*

5.3.2 Crystallization of *Li*FEN-WT with DNA substrate

Crystallization assays were set up in three commercial screens, JCSG, PACT and Morpheus, each containing 96 reservoir different conditions. The JCSG and PACT screens were according to their pH in two 96 deep-well plates. The DNA substrate used for the trials were 3ov6 and JT. The initial concentration for the crystallisation trials was on a 1:1 protein-DNA molar ratio at a concentration of 0.25 mM and 2:1 protein-DNA molar ratio with a concentration of 0.25 mM of protein and 0.125 mM of

DNA. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the three commercial screens, a high pH buffer condition (with both molar ratios using 3ov6 and JT DNA substrates) and a low pH condition (with both molar ratios using 3ov6 and JT DNA substrates) produced crystal structures, which are described in Table 5.3 and pictures of some of those crystals are shown in Figure 5.4.

Table 5.3 Conditions in which *Li*FEN with DNA substrate crystals were formed

Commercial screen	Condition	DNA substrate	Reagents
High pH initial screening			
PACT	G11	3ov6	0.2 M sodium citrate tribasic dihydrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350
PACT	G11	JT	
Low pH initial screening			
JCSG	E8	3ov6	1 M ammonium phosphate dibasic, 0.1 M sodium acetate pH 4.5
JCSG	E8	JT	

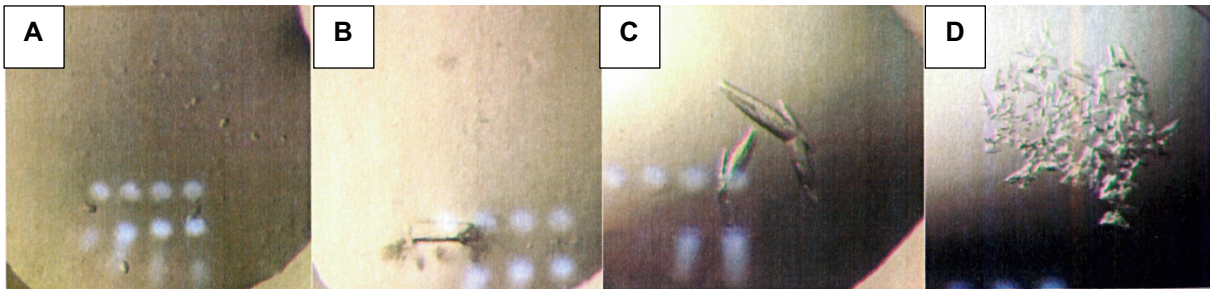


Figure 5.4 *Li*FEN with DNA substrate crystals from initial screening

(A) *&* ***(C)*** *Initial crystallization trial using the JCSG screen. Condition E8 containing 1 M ammonium phosphate dibasic, 0.1 M sodium acetate pH 4.5 (DNA substrates 3ov6 or JT substrate in A and C, respectively).* ***(B)*** *&* ***(D)*** *Initial crystallization trial using the PACT screen. Condition G11 containing 0.2 M sodium citrate tribasic dihydrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350 yielded crystals (DNA substrates 3ov6 or JT substrate in B and D, respectively). Protein and DNA substrate were on a 1:1 molar ratio (duplex structure on 5ov4 and flap structure on JT) at a concentration of 0.25 mM in all conditions.*

5.4 Crystallization of *Cn*FEN protein

5.4.1 Crystallization of *Cn*FEN-WT

Crystallization trials with purified *Cn*FEN-WT (see Section 3.4) were set up in two commercial screens, JCSG and PACT, each containing 96 different reservoir conditions and distributed according to their pH. The initial concentration for the crystallization trials without substrate was at 16 mg/mL. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the two commercial screens, one well in the high pH condition and produced microcrystal structures. Pictures of the microcrystals obtained are presented in Figure 5.5.

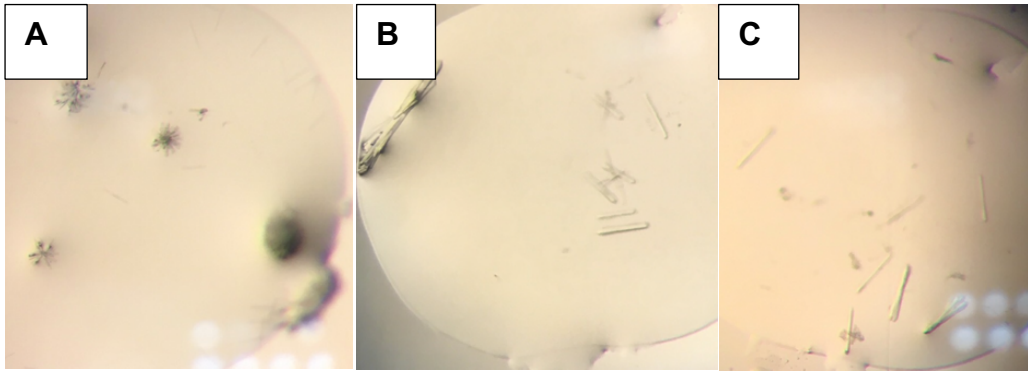


Figure 5.5 *Cn*FEN crystals from initial screening

Initial crystallization trials using PACT screen. Reservoir conditions yielding crystals; 0.2 M potassium sodium tartrate tetrahydrate, 0.1 M Bis-Tris propane pH 7.5, 20% (w/v) PEG 3350. Specific reagents were also added to the conditions as described next. ***(A)*** *50 mM $MgCl_2$.* ***(B)*** *50 mM $MgCl_2$, 100 mM KCl.* ***(C)*** *50 mM $MgCl_2$, 100 mM NaCl.*

5.4.2 Crystallization of *Cn*FEN-FL

Crystallization trials with purified *Cn*FEN-FL were set up in six commercial screens: JCSG, PACT, Morpheus, Proplex (Molecular Dimensions), MPD (NeXtal Biotechnologie), and Natrix (Hampton Research), each containing 96 reservoir different conditions. For each condition, the protein was seeded with two DNA substrates (5ov4 and JT) and without DNA substrate.

The initial concentration of CnFEN-FL for the crystallization trials without DNA substrate was between 11 to 13 mg/mL. The initial concentration of the protein for the crystallization trials with DNA substrate was determined considering a 1:1 molar ratio of protein:DNA at 0.25 mM. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the six commercial screens, it was possible to grow crystals and further optimization of the conditions was carried out. Conditions from the initial screening where crystallization occurred can be found in Table 5.4. Pictures of some of the crystals grown for this protein are depicted in Figure 5.6.

Table 5.4 Conditions in which *Cn*FEN-FL crystals were formed

Commercial screen	Condition	DNA substrate	Reagents
JCSG	A1	None	0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 50% (w/v) PEG 400
		5ov4	
	A11	5ov4	0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% (v/v) MPD
		JT	
	D7	5ov4	0.2 M lithium sulphate, 0.1 M Tris pH 8.5, 40% (v/v) PEG 400
		JT	
	D9	5ov4	0.17 M ammonium sulphate, 25.5% (w/v) PEG 4000, 15% (v/v) glycerol
		JT	
	E9	5ov4	1.6 M magnesium sulphate heptahydrate, 0.1 M MES pH 6.5
		JT	
MPD	C7	5ov4	0.2 M lithium sulphate, 40% (v/v) MPD
		JT	
	C8	5ov4	0.2 M magnesium sulphate, 40% (v/v) MPD
		JT	
	C11	5ov4	0.2 M ammonium sulphate, 40% (v/v) MPD
		JT	
	D7	5ov4	0.2 M ammonium phosphate, 40% (v/v) MPD
		JT	
	H2	5ov4	0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 30% (w/v) MPD
		JT	

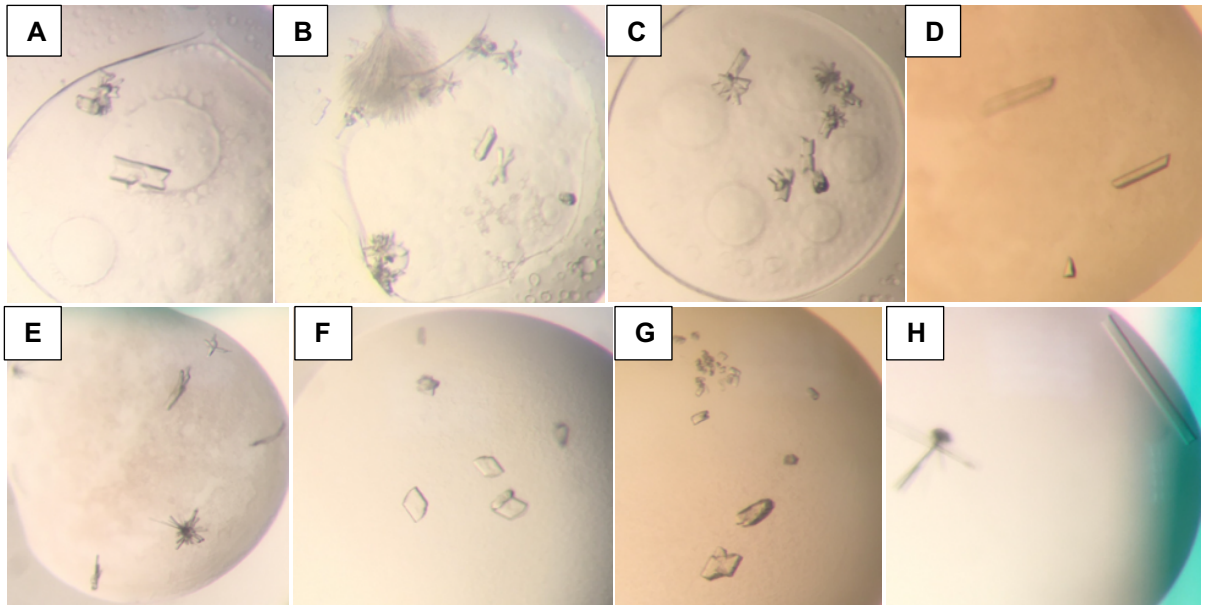


Figure 5.6 *Cn*FEN-FL crystals obtained from optimized conditions

(A) *Optimization of A1 JCSG screen. Condition B3 containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 50% (w/v) PEG 400 and JT DNA substrate.* ***(B)*** *&* ***(C)*** *Optimization of A1 JCSG screen. Condition B4 containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH 5, 50% (w/v) PEG 400 (DNA substrates 5ov4 or JT for pictures B and C, respectively).* ***(D)*** *&* ***(E)*** *Optimization of A11 JCSG screen. Condition E3 containing 0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 7.5, 48% (v/v) MPD (DNA substrates 5ov4 or JT for pictures D and E, respectively).* ***(F)*** *&* ***(G)*** *Optimization of C8 MPD screen. Condition B4 containing 0.22 M magnesium sulphate, 40% (v/v) MPD (DNA substrates 5ov4 or JT for pictures F and G, respectively).* ***(H)*** *Optimization of H2 MPD screen. Condition E9 containing 0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 28% (w/v) MPD and 5ov4 DNA substrate. Protein and DNA substrate were on a 1:1 molar ratio (duplex structure on 5ov4 and flap structure on JT) at a concentration of 0.25 mM in all conditions.*

5.4.3 Crystallization of *Cn*FEN ΔC-414

Crystallization trials with purified *Cn*FEN ΔC-414 were carried out on the same basis as *Cn*FEN-FL (Section 5.4.2). The initial concentration of CnFEN ΔC-414 for the crystallization trials without DNA substrate was between 12 to 14 mg/mL. The initial concentration of the protein for the crystallization trials with DNA substrate was carried out considering a 1:1 molar ratio of protein:DNA at 0.25 mM. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the six commercial screens, it was possible to grow crystals and further optimization of the conditions was carried out. Conditions from the initial screening where crystallization occurred can be found in Table 5.5. Pictures of some of the crystals grown for this protein are depicted in Figure 5.7.

Table 5.5 Conditions in which *Cn*FEN ΔC-414 crystals were formed

Commercial screen	Condition	DNA substrate	Reagents
JCSG	A1	None	0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 50% (w/v) PEG 400
		JT	
	A11	5ov4	0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% (v/v) MPD
		JT	
	D4	5ov4	0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 30% (w/v) PEG 8000
		JT	
	D7	5ov4	0.2 M lithium sulphate, 0.1 M Tris pH 8.5, 40% (w/v) PEG 400
		JT	
MPD	B8	5ov4	0.2 M zinc sulphate, 40% (v/v) MPD
		JT	
	C7	5ov4	0.2 M lithium sulphate, 40% (v/v) MPD
		JT	
	C8	5ov4	0.2 M magnesium sulphate, 40% (v/v) MPD
		JT	
	C11	5ov4	0.2 M ammonium sulphate, 40% (v/v) MPD
		JT	
	D7	5ov4	0.2 M ammonium phosphate, 40% (v/v) MPD
		JT	
	H2	5ov4	0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 30% (w/v) MPD
		JT	
Proplex	G5	JT	1 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.0
	H5	JT	1.2 M potassium sodium tartrate tetrahydrate, 0.1 M Tris pH 8.0

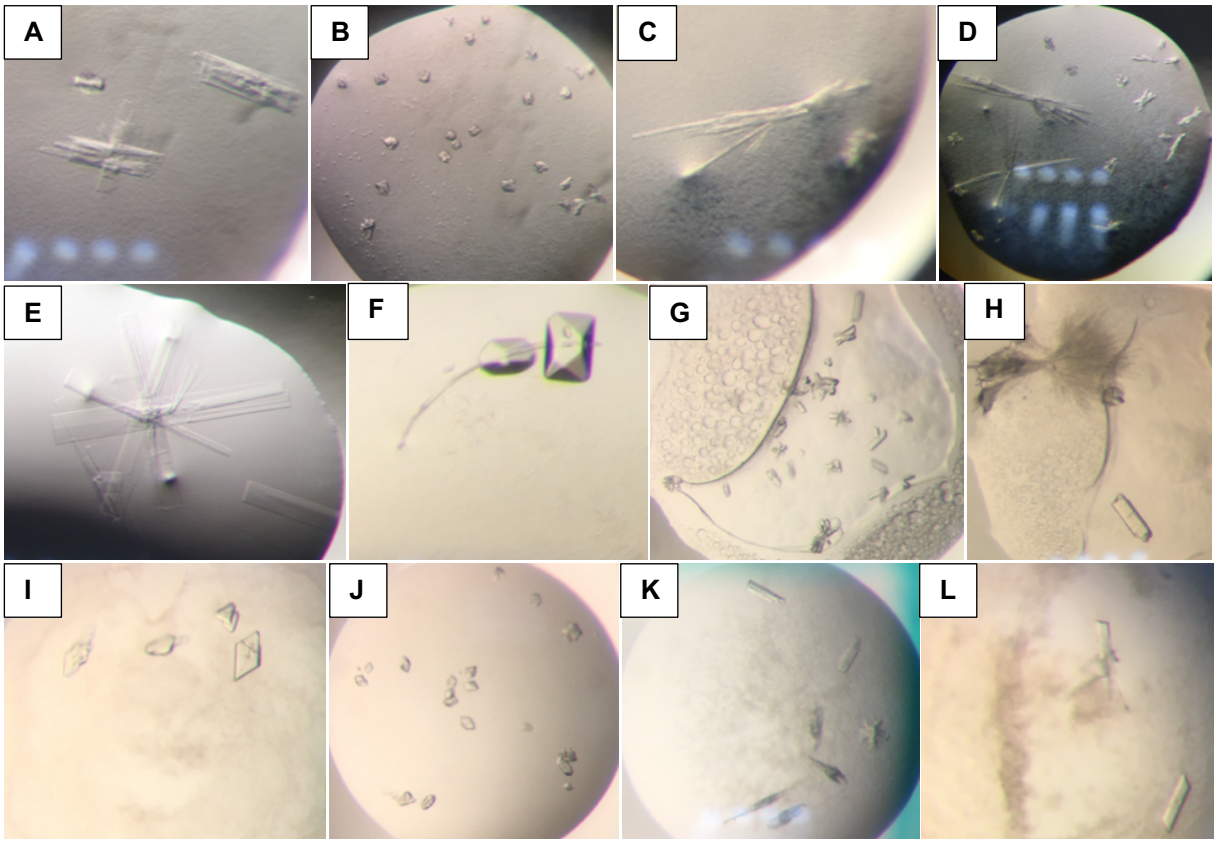


Figure 5.7 *Cn*FEN ΔC-414 crystals from initial screening and optimized conditions

(A) *Initial crystallization trial of MPD screen. Condition H2 containing 0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 30% (w/v) MPD and JT DNA.* ***(B)*** *&* ***(C)*** *Initial crystallization trial of MPD screen. Condition C7 containing 0.2 M lithium sulphate, 40% (v/v) MPD (DNA substrates 5ov4 or JT for pictures B and C, respectively).* ***(D)*** *Initial crystallization trial of MPD screen. Condition C11 containing 0.2 M ammonium sulphate, 40% (v/v) MPD and 5ov4 DNA substrate.* ***(E)*** *Initial crystallization trial of Proplex screen. Condition G5 containing 1 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.0 and JT DNA substrate.* ***(F)*** *Initial crystallization trial of Proplex screen. Condition H5 containing 1.2 M potassium sodium tartrate tetrahydrate, 0.1 M Tris pH 8.0 and JT DNA.* ***(G)*** *&* ***(H)*** *Optimization of A1 JCSG screen. Condition B4 containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH 5, 50% (w/v) PEG 400 (DNA substrates 5ov4 or JT for pictures G and H, respectively).* ***(I)*** *Optimization of A11 JCSG screen. Condition E2 containing 0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 7, 48% (v/v) MPD and 5ov4 DNA substrate.* ***(J)*** *Optimization of C8 MPD screen. Condition C2 containing 0.18 M magnesium sulphate, 42% (v/v) MPD and 5ov4 DNA substrate.* ***(K)*** *Optimization of C8 MPD screen. Condition C1 containing 0.16 M magnesium sulphate, 42% (v/v) MPD and 5ov4 DNA substrate.* ***(L)*** *Optimization of D7 MPD screen. Condition G6 containing 0.26 M ammonium phosphate, 42% (v/v) MPD and 5ov4 DNA substrate. Protein and DNA substrate were on a 1:1 molar ratio (duplex structure on 5ov4 and flap structure on JT) at a concentration of 0.25 mM in all conditions.*

5.4.4 Crystallization of *Cn*FEN ∆*C-405*

Crystallization trials with purified *Cn*FEN ∆C-405 were carried out on the same basis as *Cn*FEN-FL (Section 5.4.2). The initial concentration of CnFEN ∆C-405 for the crystallization trials without DNA substrate was between 11.5 to 13 mg/mL. The initial concentration of the protein for the crystallization trials with DNA substrate was carried out at a 1:1 molar ratio of protein:DNA at 0.25 mM. Protein sample and reservoir buffer were seeded in the crystallization plates at 100 nL each.

From the initial crystallization trials of the six commercial screens, it was possible to grow crystals and further optimization of the conditions was carried out. Conditions from the initial screening where crystallization occurred can be found in Table 5.6. Pictures of some of the crystals grown for this protein are depicted in Figure 5.8.

Table 5.6 Conditions in which *Cn*FEN ∆C-405 crystals were formed

Commercial screen	Condition	DNA substrate	Reagents
JCSG	A1	None	0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 50% (w/v) PEG 400
		5ov4	
		JT	
	A11	5ov4	0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% (v/v) MPD
		JT	
	D4	5ov4	0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 30% PEG 8000
		JT	
	D7	5ov4	0.2 M lithium sulphate, 0.1 M Tris pH 8.5, 40% (v/v) PEG 400
		JT	
MPD	C7	5ov4	0.2 M lithium sulphate, 40% (v/v) MPD
		JT	
	C8	5ov4	0.2 M magnesium sulphate, 40% (v/v) MPD
		JT	
	C11	5ov4	0.2 M ammonium sulphate, 40% (v/v) MPD
		JT	
	D7	5ov4	0.2 M ammonium phosphate, 40% (v/v) MPD
		JT	
	H2	5ov4	0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.5, 30% (w/v) MPD
		JT	
Proplex	G5	JT	1 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.0

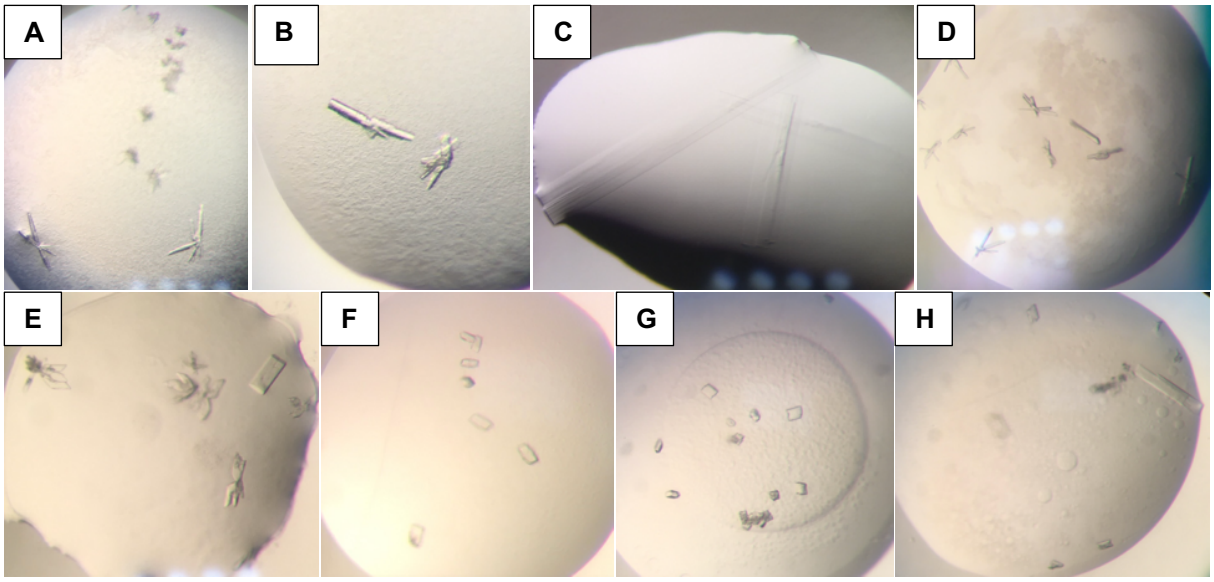


Figure 5.8. *Cn*FEN ΔC-405 crystals from initial screening and optimized conditions

(A) *&* ***(B)*** *Initial crystallization trial of MPD screen. Condition C11 containing* 0.2 M ammonium sulphate, 40% (v/v) MPD *(5ov4 and JT for pictures B and C, respectively).* ***(C)*** *Initial crystallization trial of Proplex screen. Condition G5 containing 1 M ammonium sulphate, 0.1 M HEPES sodium salt pH 7.0 and 5ov4 DNA.* ***(D)*** *Optimization of A11 JCSG screen. Condition F3 containing 0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 7.5, 50% (v/v) MPD and 5ov4 DNA.* ***(E)*** *Optimization of D4 JCSG screen. Condition F12 containing 0.2 M lithium sulphate, 0.1 M sodium acetate pH 6, 30% PEG 8000 and 5ov4 DNA.* ***(F)*** *Optimization of C8 MPD screen. Condition C5 containing 0.24 M magnesium sulphate, 42% (v/v) MPD and 5ov4 DNA.* ***(G)*** *Optimization of H2 MPD screen. Condition F11 containing 0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 8.5, 30% (w/v) MPD and 5ov4 DNA.* ***(H)*** *Optimization of H2 MPD screen. Condition G11 containing 0.5 M ammonium sulphate, 0.1 M HEPES sodium salt pH 8.5, 32% (w/v) MPD and 5ov4 DNA. Protein and DNA substrate were on a 1:1 molar ratio (duplex structure on 5ov4 and flap structure on JT) at a concentration of 0.25 mM in all conditions.*

5.5 X-ray crystal diffraction and data collection

After formation of the protein crystal on the initial screens or optimized conditions, the crystals were retrieved using a loop with specific characteristics according to the size and shape of the crystals, soaked in a cryoprotectant, instantly frozen in liquid nitrogen and properly shipped to the Diamond Light Source for analysis using one of the synchrotrons X-ray beamlines. Crystals of *Li*FEN, *Tb*FEN D183K ΔC-341, *Cn*FEN-WT, *Cn*FEN-FL, *Cn*FEN ΔC414 and *Cn*FEN ΔC405 were shipped to the synchrotron. X-ray

diffraction data collected from the samples is shown in Table 5.7 and in Appendix (Figures 9.6 to 9.11).

Table 5.7 Data collected from X-ray diffraction of microbial FEN crystals

Protein	*Tb*FEN D183K ΔC-341		*Li*FEN		*Cn*FEN ΔC414
Crystal condition*	E3c	C12a	G10a	G1a	E2b
Space group	*P* 4_1 2_1 2	*P* 4_1 2_1 2	*P* 2_1 2_1 2_1	*P* 1 2_1 1	*F* 4 3 2
Cell dimensions *A, B, C* (Å)	100.27 100.27 75.88	101.64 101.64 75.64	48.32 63.57 79.43	60.08 117.19 82.01	254.89 254.89 254.89
α, β, γ (º)	90, 90, 90	90, 90, 90	90, 90, 90	90, 93.66, 90	90, 90, 90
Resolution (Å)	4.65	3.74	1.41	2.82	2.83
R_{meas} Inner	0.064	0.078	0.022	0.138	0.060
R_{meas} Outer	0.925	11.578	1.814	1.558	7.322
I/σ (I)	9.8	5.8	13.0	2.2	10.1
Completeness (%)	100.0	100.0	100.0	99.9	100.0

** Figures 9.6 to 9.11 from Appendix contains further information about the conditions of the crystals formed.*

For further analysis of structural determination and refinement of microbial FENs, it is ideal to have a resolution of 3 Å or better. Taking this into consideration, the quality of the X-ray diffraction data collected from *Tb*FEN D183K ΔC-341 it is not the greatest for further analysis due to the low resolution of the crystals, making this a challenging task for structural determination. For the case of *Cn*FEN ΔC-414, analysis of the diffraction data was attempted but is not presented here as no high-quality model could be built despite processing diffraction results through the same data analysis pipeline due to time limitations and failure of the downstream analysis process (via the Diamond Light Source server). Therefore, analysis of the two data sets obtained from *Li*FEN was performed for structural determination and refinement.

5.6 Structural determination and refinement of *Li*FEN

Structural determination and refinement of *Li*FEN was carried out using the CCP4 suite software (Winn *et al.*, 2011). X-ray diffraction data and preliminary analysis of the structure were downloaded from the Diamond Light Source server and further processing of the structure was performed using the programmes inside the CCP4 such as REFMAC5 (Murshudov *et al.*, 2011), Coot (Emsley *et al.*, 2010) and Buccaneer (Cowtan, 2006).

Primarily, Buccaneer was used for automated building of the protein using the amino acid sequence of the protein and the diffraction data obtained from the preliminary analysis. Afterwards, an iterative process of manual building using Coot and refinement through REFMAC5 was accomplished from the result obtained from Buccaneer. The iterative process was performed until a high verdict score was met. The verdict score is a quantitative value given by the CCP4 software measuring the quality of the structure in terms of different parameters such as R_{work}/R_{free} values, Clashscore, Ramachandran outliers, *B* factor analysis, rotamer outliers, RMS bonds and RMS angles.

For this section, the determination and refinement of the *Li*FEN structure was firstly done from the data set of the crystal that diffracted at 1.41 Å. Subsequently, the data set of the other 2.82 Å crystal was also analysed. A quick overview of the later data set revealed that the model built its structurally similar to the 1.41 Å model. Therefore, the whole process of building and refinement was specifically focused on the structure with the best resolution. Once an acceptable score for the *Li*FEN model had been achieved using the CCP4i2 software, the final model was uploaded to the wwPDB validation server and a report was generated, highlighting the properties of the refined structure (Table 5.8) and validation metrics (Figure 5.9).

Table 5.8 Validation of the final *Li*FEN structure

Resolution (Å)	39.75 – 1.50
No. reflections	39868 / 1930
R_{work}/R_{free}	0.138 / 0.198
Protein	
No. residues	257
No. atoms	4144
Ligand/ion molecules	
Ethylene glycol	3
Mg2+	1
Cl-	2
Water	176
B factors (Å)	
Protein	20.2
Ligand	48.75
Ion	35.3
Water	36.35
Geometry analysis	
Ramachandran outliers	0.00%
Ramachandran favoured	98.80%
Rotamer outliers	0.45%
Clashscore	1.92
RMS bonds	0.0108
RMS angles	1.69

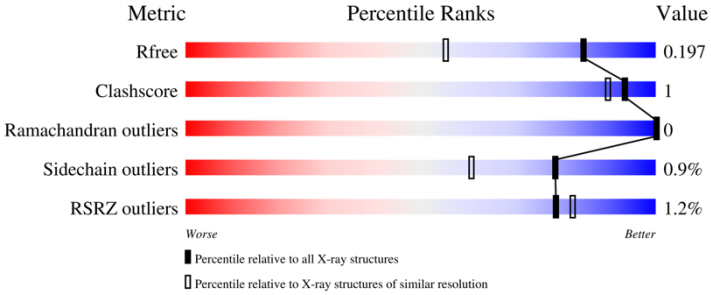


Figure 5.9 Global validation metrics of the *Li*FEN structure.

Validation of the LiFEN structure was carried out using the wwPDB validation server, entering the coordinate and structure factor files generated from the refinement process in the CCP4 software. Percentile scores (ranging between 0-100) for global validation metrics of the entry according to the reported resolution of 1.41 Å.

The model failed to provide a complete structure but did reveal several features that were characteristic of FEN enzymes. As observed in several FEN enzymes, the protein has a helix two-turn helix domain (residues 221 to 258) which is composed of conserved residues including the Gly-Ile-Gly, 244-246, and aspartic acid 237. An active-site consisting of conserved aspartic acid and glutamic acid residues is present on a solvent exposed globular domain form of beta-sheet and alpha-helices. An alignment between the amino acid sequence of *Li*FEN and the protein sequence of the structure elucidated is presented in Figure 5.10. A diagram of the experimental structure is shown in Figure 5.11.

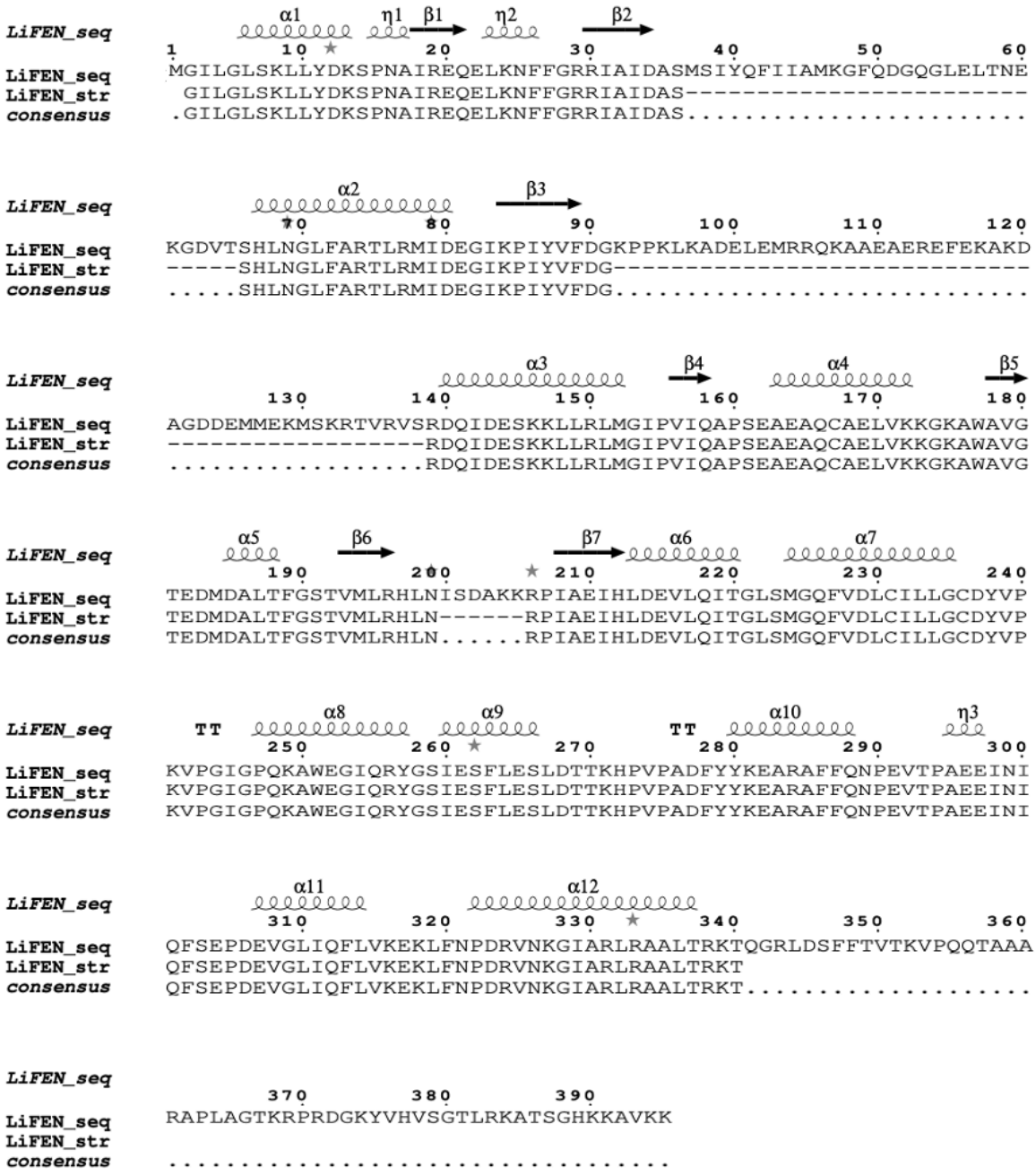


Figure 5.10 Alignment between the *Li*FEN protein sequence and experimental structure

Comparison between the protein sequence of Li*FEN (LiFEN_seq) deposited in the NCBI database and the experimental structure obtained in this project (LiFEN_str), showing the amino acids which can be confidently modelled. Secondary structures are represented as α (alpha-helix), β (beta-strand), η (3_{10}-helix), TT (strict beta-turns). Alternate residues are highlighted as gray stars. Sequence alignment was performed using the online tool MultAlin (Corpet, 1988). Depiction of sequence alignment and secondary structure was carried out with the online tool ESPript 3.0 (Robert & Gouet, 2014).*

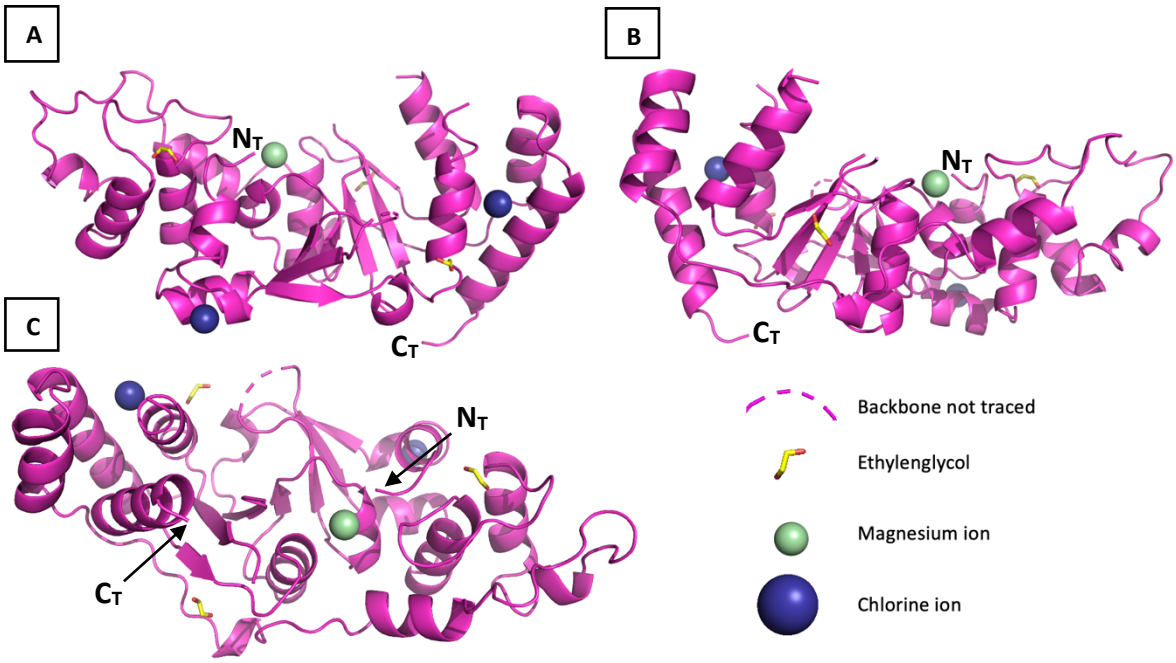


Figure 5.11 Cartoon representation *Li*FEN model

Representation of the final result of the protein model after the refinemet process observed from three orthogonal views. The N and C termini are indicated in each panel ***(A)*** *to* ***(C)****. The model shows a magnesium ion bound in the active site, as well as ethyleneglycol and chlorine ions observed in this 1.4 Å model. Of the 395 amino acid residues of the protein produced, 257 were visible in the electron density map. The regions missing in the model corresponds to 3 domains and one loop (residues 201 to 205). The domains are the helical arch (residues 92 to 138), the hydrophobic wedge next to the helical arch (residues 37 to 65) and the C-terminal (residues 341 to 395).*

In the iterative process of manual building and refinement of the *Li*FEN structure, it was crucial to analyse the conformational space of the structure and correlate the amino acid sequence of the protein with the model which has been built. At the same time, it is important to be able to distinguish between the residues which are part of the model and other molecules which are present and influence the diffraction pattern such as solutes, ions, precipitants and sometimes cryoprotectants. During the refinement process, the observed electron density map was compared with the theoretical map generated from the model built in the previous stage. Examples of the final model are presented in Figure 5.12. An ideal structure model would produce figure in which the atoms placed in a position which correctly predicts the observed density show up as blue contours, observed electron density not accounted for by the current model is

shown in green mesh. Conversely, red mesh indicates presence of excess density in the predicted model.

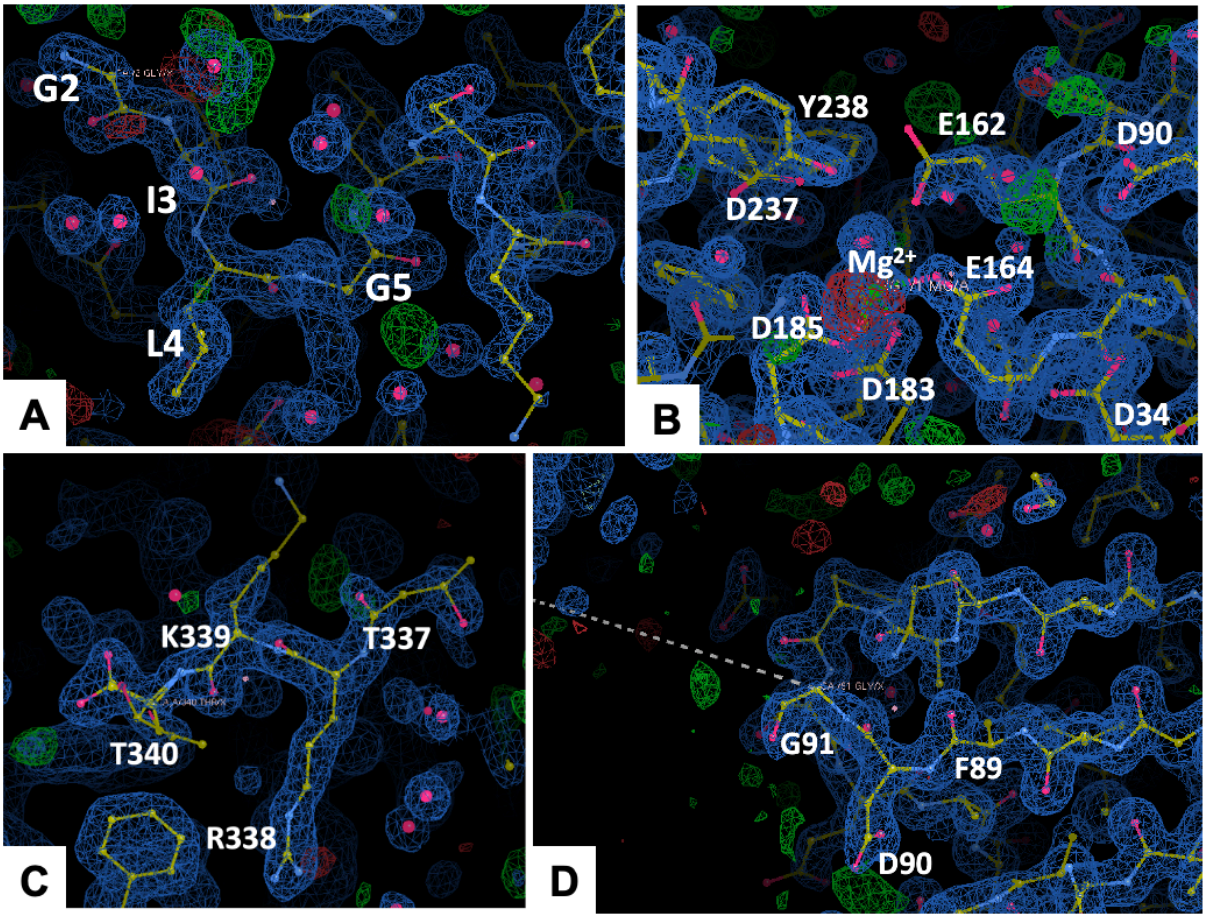


Figure 5.12 Visualization of the *Li*FEN model on Coot

*Pictures depicting the N terminal region **(A)**. The active site **(B)** shows the presence of a magnesium ion (green sphere), and electron density map of the amino acids in this region. The C terminal of the protein and residues labeled on those regions is shown in panel **(C)**. The predicted electron density matches the experimental data well (blue mesh). Some stretches of amino acids could not be built as there was no observable density to fit residues in to. Gly 91 was fitted, but there is no electron density to build the next residues into as presented in panel **(D)**. The electron density level shown in the figure corresponds to 1.5 rmsd.*

5.7 Comparison of *Li*FEN with published FEN structures

From the refined *Li*FEN structure, a comparison of this model has been performed against the human FEN structures which are available in the Protein Data Bank. The human FEN and *Li*FEN proteins share a sequence identity of 54% and similarity of 72% over the first 351 residues of the *Li*FEN. Several human FENs have already been deposited in the PDB database, with a total of 12 protein structures including some with DNA, with an inhibitor, and also in complex with other proteins (Table 5.9).

Table 5.9 Human FEN structures deposited in the PDB database

Structure	Method	PDB code	Reference
FEN1 (WT) in complex with product 5′-flap DNA	X-ray diffraction	3Q8K	Tsutakawa *et al., 2011*
FEN1 (WT) in complex with substrate 5′-flap DNA		3Q8L	
FEN1 (D181A) in complex with substrate 5′-flap DNA		3Q8M	
FEN1 in complex with an N-hydroxyurea compound	X-ray diffraction	5FV7	Exell *et al.,* 2016
FEN1 (D233N) with cleaved product fragment	X-ray diffraction	5K97	Tsutakawa *et al.,* 2017
FEN1 (R100A) with 5′-flap substrate DNA		5KSE	
FEN1 (D86N) with 5′-flap substrate DNA		5UM9	
FEN1 in apo form	X-ray diffraction	5ZOD	Xu *et al.,* 2018
FEN1 (D181A) in complex with DNA		5ZOE	
FEN1 (D181A/R192F) in complex with DNA		5ZOF	
FEN1 (R192F) in complex with DNA		5ZOG	
FEN1-PCNA complex	X-ray diffraction	1UL1	Sakurai *et al.,* 2005
Polymerase δ-FEN1-PCNA toolbelt	Electron microscopy	6TNZ	Lancey *et al.,* 2020

A comparison was done using the refined structure of the *Li*FEN with the structure of the human enzyme which contains a DNA flap substrate and metal ions as a resolution of 2.2 Å (3Q8K.pdb, Tsutakawa *et al.*, 2011). The graphical comparison is shown in Figure 5.13, as can be seen the overall similarity is high as it would be expected of orthologous enzymes.

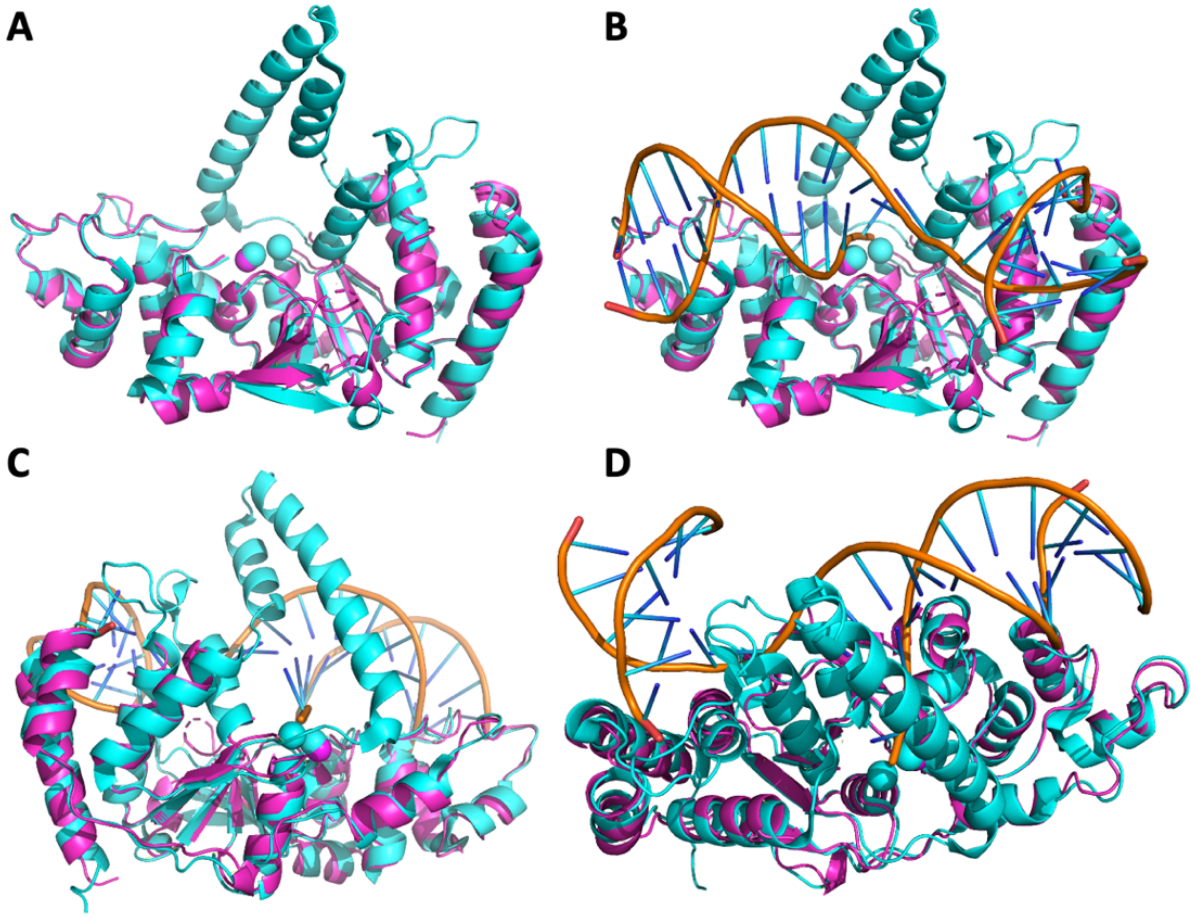


Figure 5.13 Comparison of *Li*FEN and Human FEN structure

(A) *Representation of the model after the refinement process of the* Li*FEN (magenta) aligned with a previously crystallized human FEN structure (PDB code 3Q8K), coloured in cyan.* ***(B)****,* ***(C)*** *and* ***(D)*** *corresponds to the same* Li*FEN-Human FEN alignment from three orthogonal views and DNA processing from the human FEN is also depicted in these 3 panels. Magenta spheres correspond to* Mg^{2+} *ions from the* Li*FEN model and cyan spheres are* Sm^{3+} *ions from the human FEN (RMSD = 0.971 Å of 224 backbone carbon atoms aligned).*

The main structural differences seen in these structures resides in three regions: (1) A large section of the backbone of *Li*FEN could not be built, corresponding to residues 92 to 138 which match to the amino acids 87 to 135 in the human enzyme. (2) A second large missing section from the residues 37 to 65 which correlates to amino acids 37 to 61 in the human enzyme. (3) A small section corresponding to the residues 199 to 210 show significant displacement of the backbone such as Pro207 and Pro203 in the human enzyme, showing a displacement of the C-alpha atoms by 6 Å. Furthermore, 6 residues (200 to 205) could not be built in this region for *Li*FEN, suggesting flexibility

in this region and the probable presence of multiple conformations of this region throughout the crystal. Interestingly, the human structure deposited in the PDB (3Q8K.pdb) shows the equivalent residues in this region (198 to 206) in two independent conformations. These features are summarized in Figure 5.14.

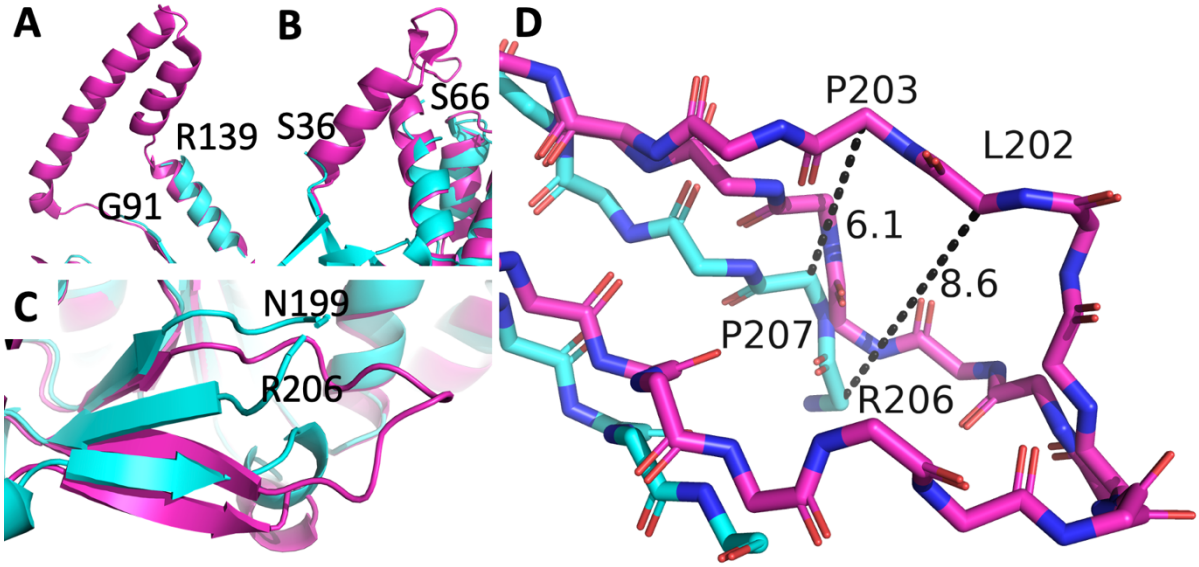


Figure 5.14 Main structural differences between *Li*FEN and human FEN

Major structural differences between Li*FEN model (cyan) and published orthologue (magenta) are shown.* ***(A)*** *shows the missing helical arch feature through which the 5'-single-stranded region of the flap structure is thought to thread.* ***(B)*** *shows the missing large section of the hydrophobic wedge (residues 37 to 65).* ***(C)*** *shows a region of the* Li*FEN protein which is displaced by several Ångstroms* ***(D)****. This region has six unbuilt residues in the* Li*FEN structure (200 to 205). The human structure in this region is present in two conformations as shown in panel D. Example displacements of equivalent residues are shown by the dotted lines in Ångstroms.*

Other FEN structures such as 1UL1 (Sakurai *et al.*, 2005) also show variations of the defined DNA binding arch seen in 3Q8K (Figure 5.14). This structure contains 3 copies of human FEN bound to 3 copies of proliferating cell nuclear antigen (PCNA), which is a DNA clamp important for replication (Querol-Audi *et al.*, 2012). As can be seen in Figure 5.15 the core of all the FENs superimposed very well (*Li*FEN superimposes on X, Y and Z conformers with an RMSD in the range 1.1 to 1.3 Å).

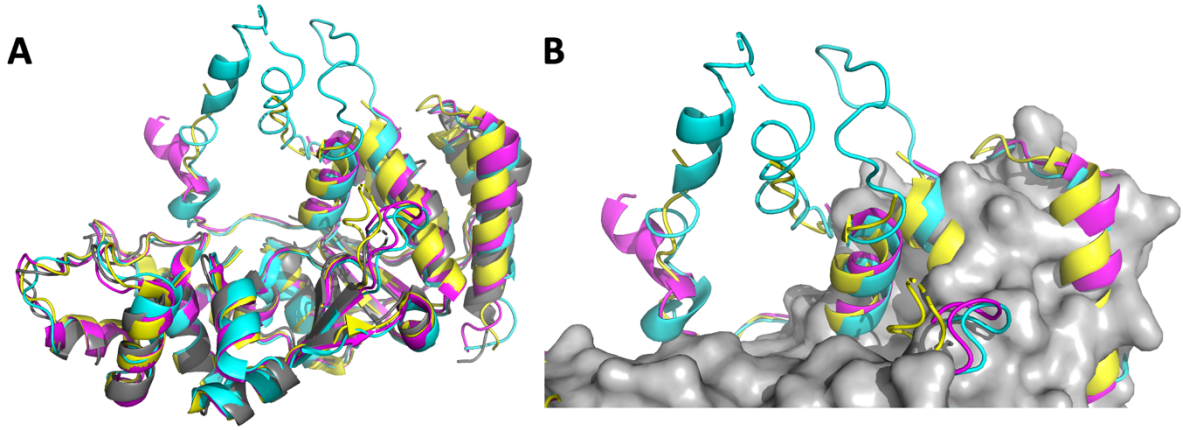


Figure 5.15 Structural comparison of *Li*FEN and human FENs

Comparison between the LiFEN model (gray) with the 3 copies of the human FEN found in the 1UL1 deposited structure (depicted in yellow, magenta, and cyan for the X, Y and Z chains of the 1UL1 structure). ***(A)*** *shows all 3 conformations of human FEN1 reported superimposed on* Li*FEN.* ***(B)*** *shows the missing regions and differences between the proteins with the* Li*FEN model represented as a solid gray surface.*

Evidence for flexibility of the human FEN protein is not limited to that provided by crystallography. A study by Bennet and co-workers demonstrated that these regions were found to be flexible using solution-based methods including nuclear magnetic resonance (NMR) and single-molecule FRET data. These authors suggested that flexibility of the protein is required for the processing of the DNA flap (Bennet *et al.*, 2018).

Flexibility of the arch seems to be important in DNA binding in the more distant homologue from bacterial phage T5 (5HMM.pdb, AlMalki *et al.*, 2016). This viral protein shares 25% of sequence identity with the *Leishmania* protein. However, as it can be seen in Figure 5.16, the structures of the two proteins overlay well. Previous work on the viral protein has demonstrated that the arch can exist in several conformations, including one of the two helices that form the helical arch become into a more relaxed structure.

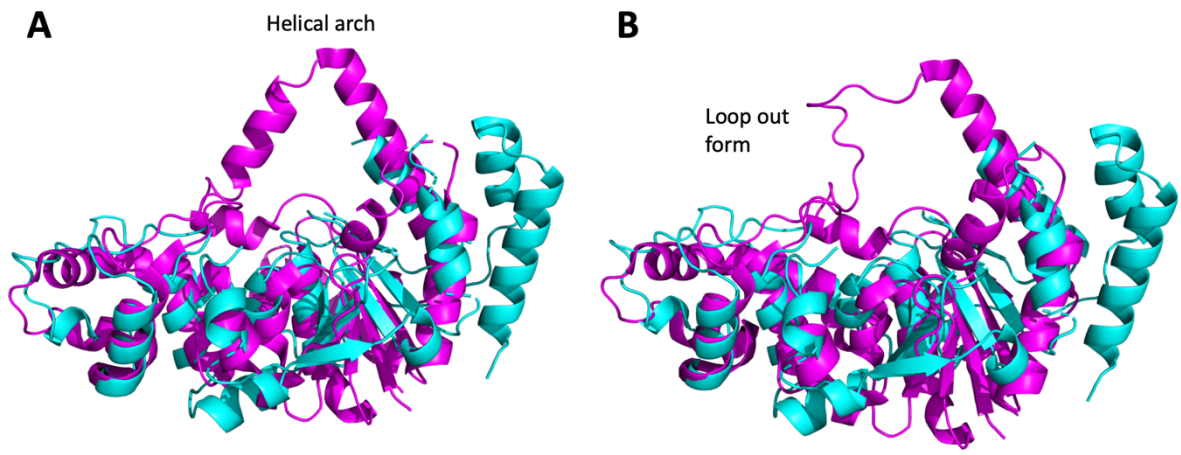


Figure 5.16 Superposition of the *Li*FEN and bacteriophage T5 flap endonuclease structures.

Superposition of Li*FEN (cyan) with two conformers of T5 FEN homologue (cyan, 5HMM.pdb).* ***(A)*** *shows chain A of the T5 protein in the helical arch form. The protein superimposed with an RMSD of 4.4 Å,* ***(B)*** *shows the superposition with chain B of the T5 protein in which one of the helical arch elements has adopted a random coil conformation (RMSD = 5.4 Å).*

One potential caveat here is that when considering the conformations seen in any crystal structure is that crystal packing artifacts could lead to inaccurate conclusions. Figure 5.17 shows the crystal packing around a central molecule of *Li*FEN. The absence of traceable electron density for the missing residues in the *Li*FEN structure suggests that they have adopted a range of different conformations. This raises the question of why no ordered helical arch-like DNA binding region is observed in the *Li*FEN protein? If we superimpose the human FEN protein on one of the molecules in the protein array (Panel B) it is obvious that clashes will be introduced for several regions of the protein. These regions are highlighted in cyan in panel C Figure 5.17 and correspond well with the missing regions of the *Li*FEN model.

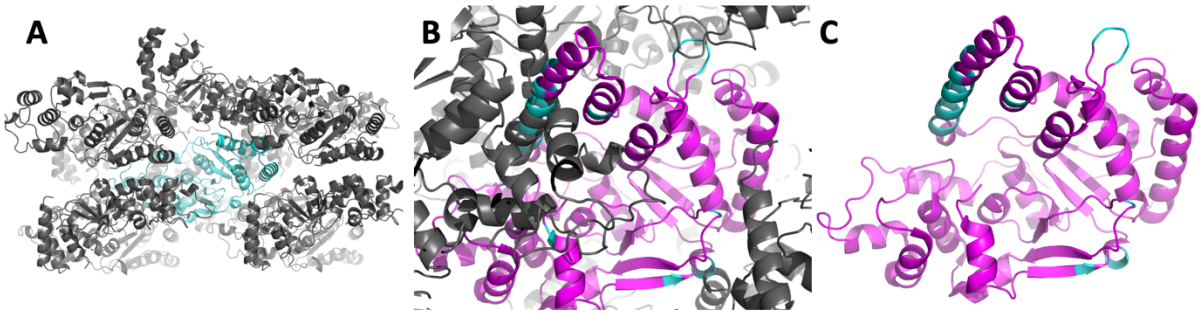


Figure 5.17 Crystal packing of molecules in the *Li*FEN structure

*Symmetry mates (gray) were regenerated for the asymmetric unit in cyan **(A)**. **(B)** shows the human FEN structure (magenta) superimposed on the* Li*FEN asymmetric unit (not shown). **(C)** shows residues that would clash badly with other molecules in the crystal array. These residues are shown in cyan and were selected as they are within 2.5 Å.*

5.8 Structure of the metal-binding active site

The metal-binding active site is essential for the function of the flap endonucleases (Harrington & Lieber, 1994). The activity of FENs requires the presence of divalent metal ions which, coordinated by conserved aspartic and glutamic acids, sets up a process which generates a hydroxide ion which participates in a nucleophilic reaction that cuts the DNA backbone, resulting in cleavage products with a 3′-hydroxyl group and the 5′-phosphate (Syson *et al.*, 2008).

This event has been well characterized in different organisms. For example, in the human FEN, exonuclease 1 (EXO1) from *E. coli* and bacteriophage T5 FEN proteins (Tsutakawa *et al.*, 2011; Anstey-Gilbert *et al.*, 2013; AlMalki *et al.*, 2016). The *Li*FEN structure obtained revealed the presence of one magnesium in the active site (Figure 5.19). A comparison of the *Li*FEN model with the previous homologue proteins mentioned is depicted in Figure 5.18.

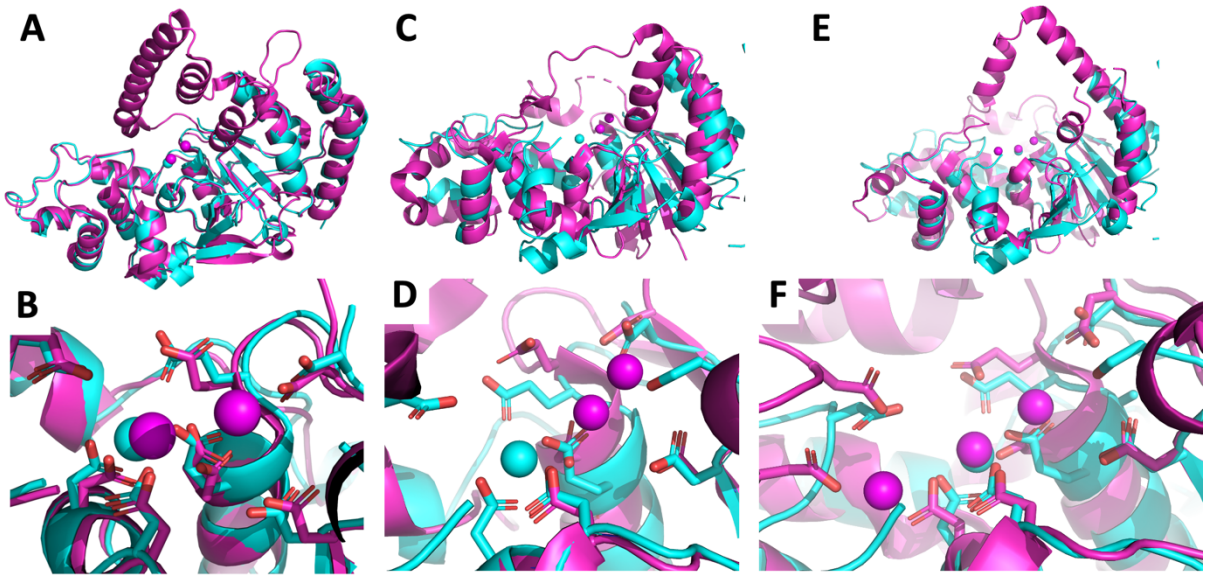
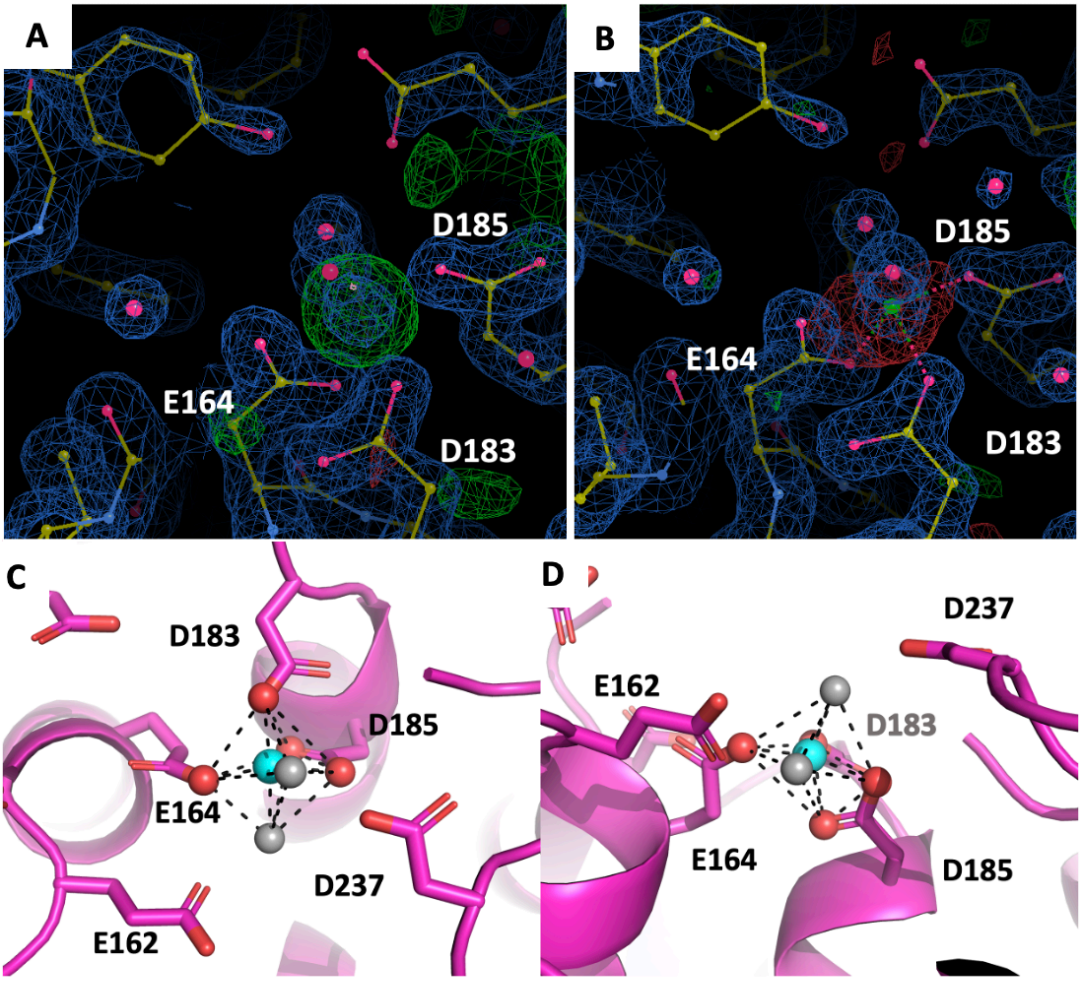


Figure 5.18 Comparison of the active sites in eukaryotic, bacteria, and phage FENs

(A) *and* ***(B)*** *show a general overview and closer view to the active site on the superposition of the* Li*FEN model with the human FEN (3Q8K.pdb). On panel B it is seen clearly two magnesium ions from the human FEN.* ***(C)*** *and* ***(D)*** *show a comparison of the* Li*FEN model with the exonuclease IX (3ZDB.pdb) depicting two metal ions.* ***(E)*** *and* ***(F)*** *depict the superposition of the* Li*FEN model with the T5 FEN (5HMM.pdb) which contains three metal ions in the active site.*

From the *Li*FEN model, one of the expected active-site metal ions was clearly visible showing primarily direct chelation from conserved residues Asp 183, Asp185 (both oxygens) and Glu164 and two water molecules in approximately octahedral geometry. The two water molecules were found to be in hydrogen bonding distance of 2.4 Å and 2.5 Å in Tyr238 and Glu162 side chain residues. In addition, the web server CheckMyMetal (Zheng et al., 2014) was used to validate the presence of the active-site metal ion found in the *Li*FEN model (Figure 5.19).



E

CheckMyMetal(CMM) Home | Report a Problem | Save CMM report (HTML, PDF)

ID	Res.	Metal	Occupancy	B factor (env.)[1]	Ligands	Valence[2]	nVECSUM[3]	Geometry[1,4]	gRMSD(°)[1]	Vacancy[1]	Bidentate	Alt. metal
A:1	MG	Mg	1	26.9 (26.3)	O_6	1.8	0.029	Trigonal Bipyramidal	9.4°	0	1	Co Zn
A:3	CL	Cl	1	30 (33.6)		N/A	N/A	Free	N/A	N/A	N/A	
A:4	CL	Cl	1	49.7 (43.1)		N/A	N/A	Free	N/A	N/A	N/A	

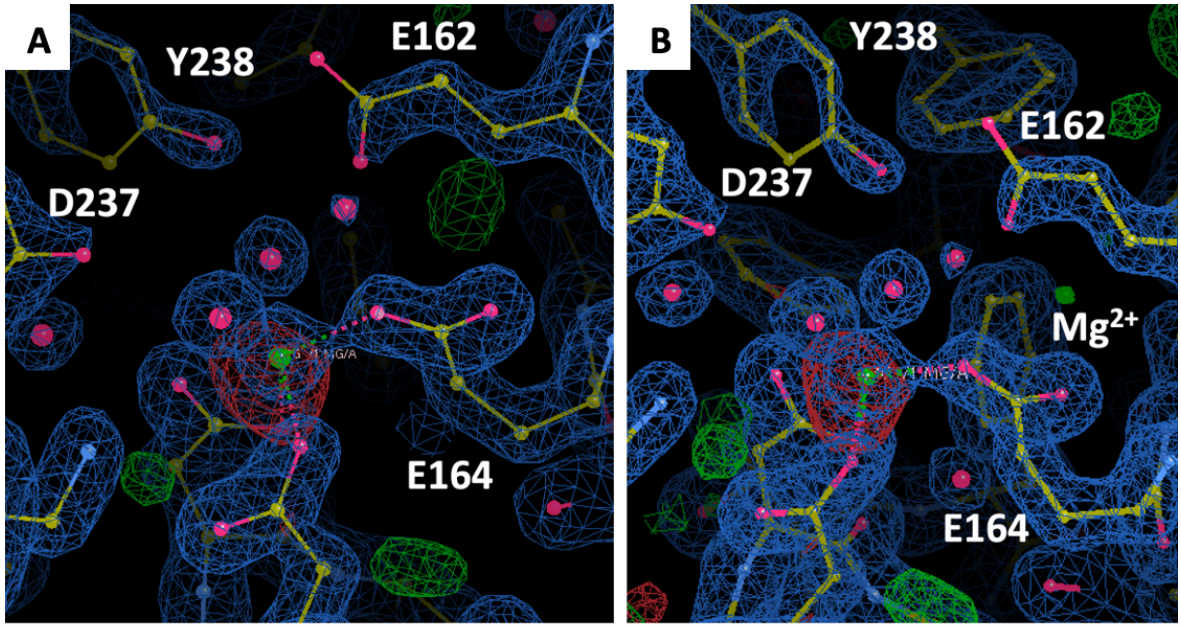
Legend: Not applicable | Outlier | Borderline | Acceptable

Generate a model with alt. metal: Select the metal(s) above and Save

Figure 5.19 Active-site with one metal ion in the *Li*FEN model

(A) *shows the active-site of the protein refined after removing the metal ion model in site 1 viewed using Coot. The green mesh sphere located between residues Glu164, Asp183, and Asp185, strongly suggests that a metal ion should be present at its centre.* ***(B)*** *presents the same active-site with the magnesium ion modelled in.* ***(C)*** *&* ***(D)*** *show two orthogonal views of the active-site of the* Li*FEN model (magenta) with the presence of one magnesium ion (cyan) revealing a distorted octahedral (formally trigonal bipyramidal according to CheckMyMetal) geometry with the chelating conserved residues Glu164, Asp183, Asp185 and two water molecules.* ***(E)*** *presents the report obtained from CheckMyMetal web server for this model shown in Panels B - D.*

However, when the electron density on the active-site was visualize more carefully, it was observable that there was a green sphere-like mesh close to one of the clearly visible metal ions (Figure 5.12 Panel B). Taking into consideration the architecture of published FENs, it was plausible that the area mentioned could correspond to the second metal ion present in eukaryotic organisms. To corroborate this statement, the magnesium was added in the green sphere-like mesh on Coot, followed by a final refinement on REFMAC5 and further validation using wwPDB (not shown in this work) and CheckMyMetal web server was carried out. The *Li*FEN model with two metal ions and the report obtained from CheckMyMetal is shown in Figure 5.20.



E

CheckMyMetal(CMM) Home | Report a Problem | Save CMM report (HTML, PDF)

ID	Res.	Metal	Occupancy	B factor (env.)[1]	Ligands	Valence[2]	nVECSUM[3]	Geometry[1,4]	gRMSD(°)[1]	Vacancy[1]	Bidentate	Alt. metal
A:1	MG	Mg	1	26.4 (25.6)	O_6	1.9	0.031	Trigonal Bipyramidal	9.8°	0	1	Co Zn
A:2	MG	Mg	1	58 (35.9)	O_3	0.8	0.87	Octahedral	23.3°	50%	0	Na Ca
A:3	CL	Cl	1	29.4 (31.8)		N/A	N/A	Free	N/A	N/A	N/A	
A:4	CL	Cl	1	48.3 (39.5)		N/A	N/A	Free	N/A	N/A	N/A	

Legend: Not applicable | Outlier | Borderline | Acceptable

Generate a model with alt. metal: Select the metal(s) above and Save

Figure 5.20 Suspected active-site with two metal ions in the *Li*FEN model

(A) *shows the active-site of the protein from Coot with a green sphere mesh between residues Glu162 and Glu164.* ***(B)*** *presents the same active-site placing a presumable second magnesium ion where the green mesh was before.* ***(C)*** *&* ***(D)*** *show two orthogonal views of the active-site of the* Li*FEN model (magenta) with the presence of two magnesium ions (cyan). A clearly visible magnesium reveals an approximately octahedral geometry with the chelating conserved residues Glu164, Asp183, Asp185 and two water molecules. The second magnesium reveals a deformed trigonal bipyramidal geometry forming chelating bonds with Glu164, three water molecules and Asp90.* ***(E)*** *presents the report obtained from CheckMyMetal web server.*

According to the results obtained from CheckMyMetal (Figure 5.20, Panel E), it is very unlikely that the green sphere-like mesh corresponds to a magnesium ion. Nevertheless, further attempts to solve the specific region of the model using the alternative metal suggested by the web server was performed (sodium was present in the crystallization condition but there was not calcium in it). However, outliers were still identified in the model when the second magnesium was substituted by sodium (data not shown).

5.9 Prediction of *Li*FEN structure and comparison with the experimental model

As the main aim of this chapter was to obtain structures to inform a virtual screening campaign, the observation that the *Li*FEN structure has several missing regions that are important for interacting with DNA was disappointing. However, molecular modelling techniques using AlphaFold2 and Phyre2 (Jumper *et al.*, 2021; Kelley *et al.*, 2015) was used to augment our data. Models generated by AlphaFold2 and Phyre2 can be compared with the experimental data solved at 1.4 Å. As can be seen below (Figure 5.21), the experimental structure of *Li*FEN superimposed on structures generated using Phyre2 and AlphaFold2 with an RMSDs of 0.94 and 0.9 Å respectively, suggesting they provide reasonable accurate predictions.

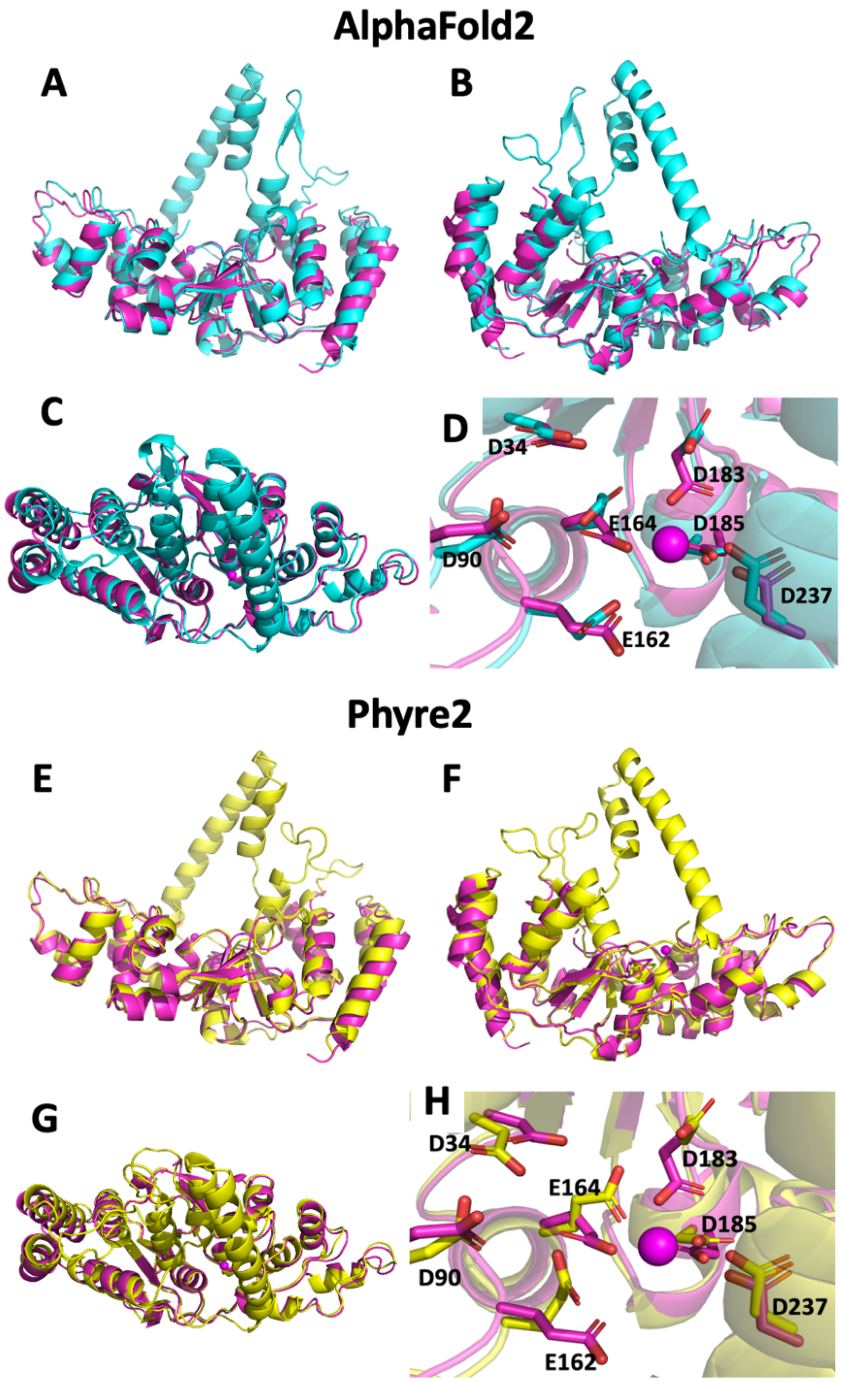


Figure 5.21 Comparison of *Li*FEN determined structure with the predicted structures

(A) *to* ***(D)*** *shows the* Li*FEN predicted by AlphaFold2 (cyan) superimposed with the* Li*FEN model obtained in this work (magenta).* ***(E)*** *to* ***(H)*** *shows the LiFEN predicted by Phyre2 (yellow) superimposed with the* Li*FEN determined model (magenta). Panels A to C and E to G depicts three orthogonal views of the determined* Li*FEN with the predicted models. Panel D and H illustrates the seven conserved residues in the active site of the protein.*

5.10 Conclusion

Several attempts of crystallizing the microbial FENs from the organisms *Trypanosoma brucei*, *Leishmania infantum* and *Cryptococcus neoformans* were performed using different conditions, including over 1000 conditions with native protein and further 2000 or more co-crystallization attempts were made with DNA substrate and protein. However, most of the crystals turned out to be inorganic salts or protein crystals with low-resolution diffraction according to the results obtained from the synchrotron source.

Nevertheless, five crystals of different microbial FENs diffracted with resolutions between 1.4 to 4.5 Å. Furthermore, the crystal structure with a resolution of 1.4 Å (*Li*FEN) was built and refined. The final model shows a clearly defined metal ion binding between the conserved residues Glu164, Asp183 and Asp185 and two water molecules in the active site. The structure solved and refined in this work has a considerably better resolution compared to the first human FEN structure published (1UL1.pdb, Sakurai *et al.*, 2005). A comparison between the later structure mentioned (FEN chain with the best RMSD) and the *Li*FEN, as well as their respective validation metrics are shown in Figure 5.22 below.

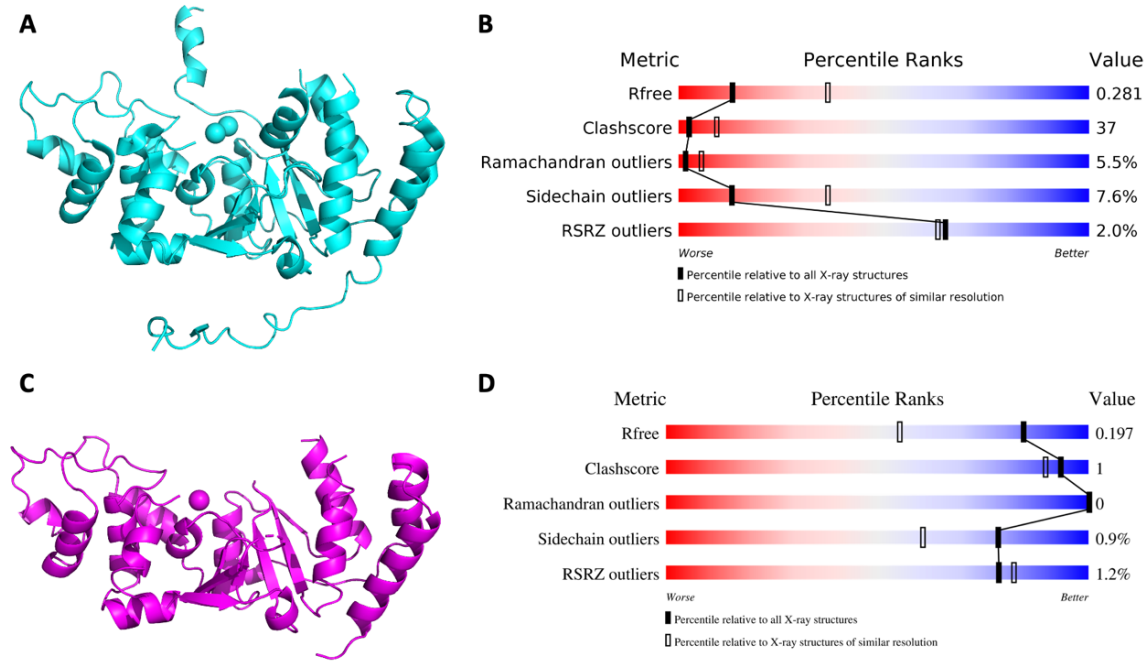


Figure 5.22 Comparison between human FEN and *Li*FEN

(A) *&* ***(B)*** *show the structure and validation metrics of the human FEN (1UL1.pdb), respectively.* ***(C)*** *&* ***(D)*** *show the structure and validation metrics of the* Li*FEN obtained in this work.* ***(E)*** *depicts the aligned structures of LiFEN with the human FEN chain with the best RMSD (1.065 Å). When comparing panels B and D it is important to observe that the* Li*FEN structure presents a better* R_{free} *and fewer outliers than the human FEN.*

The experimental structure agrees well with those models using Phyre2 and AlphaFold2 (Section 5.9). By combining the experimental and predicted data, this work should be useful to inform future structure-based drug design and virtual high-throughput approaches.

CHAPTER 6: SMALL MOLECULE INHIBITOR SCREENING TARGETING MICROBIAL FLAP ENDONUCLEASES

6.1 Introduction

The process of drug discovery is complex and expensive. A high-quality starting point in the form of a potent inhibitor can significantly influence the results in the efforts of small-molecule drug research, both in terms of speed and quality (Jones *et al.*, 2015). Drug discovery normally begins with the selection of appropriate therapeutic targets. My starting point assumes that the flap endonuclease in the target organisms is essential as has been shown in numerous other organisms (Kucherlapati *et al.*, 2002; Diaz *et al*., 1992; Fukushima *et al.*, 2007; DeJesus *et al*., 2017; Baylis *et al.*, 2005). Subsequently, modulators of the target must be identified. These modulators could include enzyme activators and inhibitors, receptor agonists and antagonists, and openers or blockers of ion channels (Symański *et al.*, 2012).

Numerous approaches exist to produce lead compounds. Traditionally, natural products, endogenous ligands, and substrates have been exploited as starting points for optimization and modification, leading to the development of innovative drugs (Jones *et al.*, 2015). More recently approaches such as fragment screening, structure-based design and virtual screening have produced thrilling outcomes (Macarron *et al.*, 2011). A great example of drug development using these approaches is the β-lactamase inhibitor RPX7009 developed by Rempex Pharmaceuticals, later known as vaborbactam. This study was conducted performing an in-depth analysis of possible enzyme-ligand interactions through virtual high-throughput screening (vHTS) and known inhibitor fragmentation approaches, leading to a potential inhibitor from 18 synthesized compounds specifically designed from the previous analyses (Hecker *et al.*, 2015). This compound entered clinical trials in 2014 (NCT02168946 & NCT02166476), and both studies showed promising results. In 2017 the FDA approved the use of this drug in combination with the drug meropenem under the commercial name VABOMERE® for the treatment of complicated urinary tract infections (cUTI) and pyelonephritis (Wunderink *et al.*, 2018; Bhowmick, 2021; Kaye *et al.*, 2018). However, conventional High-Throughput Screening (HTS) is still the most used method in the pharmaceutical industry and has gained recognition over the past

two decades as an important approach for drug discovery due to technology developments that make it possible to screen ever-larger numbers of compounds quickly and effectively (Jones *et al.*, 2015; Murray & Wigglesworth, 2017).

Modern HTS systems are capable of screening at least 1–5 million molecules (Sukuru *et al.*, 2009). However, despite the improvements in cutting-edge technologies, it is frequently preferable to screen a smaller number of compounds. This frequently happens when there is a short supply of reagents or if the purpose of the research is to elucidate the biological function of the target by evaluating a small number of compounds in multiple assays rather than one large screen. Therefore, selecting the suitable libraries in subset screens has a meaningful effect on the outcome of a screening campaign (Petrone *et al.*, 2013). For example, a free database such as Zinc20 (Irwin *et al.*, 2020) provides theoretical access to 230 million compounds which have been offered for sale. Typically, compounds are filtered by properties which are thought to be desirable in each individual screening campaign. For example, predicted solubility, number of rotatable bonds, molecular weight, number of atoms, lipophilicity, solvent accessible surface areas and specific functional groups.

First, compounds are tested in initial screens, which could come from small molecule libraries designed by pharmaceutical and chemical companies to be representative of chemical space (so called diversity or rule of three libraries; Congreve *et al.*, 2003). If a test compound gives a positive or "hit" result in such an assay, a more precise secondary screen is carried out and dose response studies can be performed to allow calculation of IC_{50} values. These initial “hit” compounds are then used as the basis for “hit expansion” and quantitative structure activity relationship (QSAR) studies to identify or develop new and more potent and selective compounds with desirable physicochemical, pharmacokinetic and toxicity properties (Martis *et al.*, 2011). Hit expansion usually consists of obtaining molecules which have similar distribution of functional groups or perhaps contain the original hit molecule as a substructure.

HTS assays must be sensitive enough to identify relatively weakly interacting molecules at the early stages of hit identification. They must also be robust enough so as to minimize false negatives but not have too high false positive discovery rate. False negatives are given a lot of attention since, of course, it is preferable to not miss anything. On the other hand, managing false positives is a more difficult task. If badly

managed, it could lead to difficulties trying to distinguish between real hits and hundreds or even thousands of false hits. In order to prove that hits are not the result of chance, an assay must also be trustworthy in that, when performed again, it will detect the majority of active compounds both times. It quickly becomes clear that an HTS assay must meet some essential requirements in order to maximise its utility in identifying hit molecules. These requirements are established in five components: robustness, reliability, simplicity, affordability, and relevance (Murray & Wigglesworth, 2017).

Robustness and reliability are the most important requirements to be met in an assay, and this can be determined using the Z'-factor parameter. The Z'-factor defines the sensitivity of an assay for detecting hits and the robustness based on the performance of the control wells and its use for so called “single-hit screening” assays (Murray & Wigglesworth, 2017). In terms of quality control in HTS experiments, the Z'-factor is defined by the sample means (μ) and standard deviations (σ) in the negative (1) and the positive (2) controls according to the next equation (Zhang *et al.*, 2020).

$$Z' = 1 - \frac{3(\sigma_1 + \sigma_2)}{\mu_2 - \mu_1}$$

In order to be considered as reliable and robust in a single-hit campaign, a Z’-factor of at least 0.5 needs to be met according to the “industry standard” measure of suitability (Martis *et al.*, 2011). The importance of having an assay with a Z’-factor exceeding 0.5 is due to a desire to keep costs as low as possible, as each compound in the library need only be assayed once, conserving reagents cost and maintaining efficiency. This allows reliable screening of as large a number of compounds as is practical.

With the aim of identifying potential molecules in validated targets with different functions and properties, various HTS technologies have been developed to accomplish this aim, such as scintillation proximity assay (SPA), Förster resonance energy transfer (FRET), fluorescence polarization (FP), homogeneous time resolved fluorescence (HTRF), fluorescence correlation spectroscopy (FCS), fluorescence intensity distribution analysis (FIDA), nuclear magnetic resonance (NMR), and biolayer interferometry (BLI) (Bokhari & Albukhari, 2021; Murray & Wigglesworth, 2017).

My study involved screening of small molecules targeting microbial and human FENs and performing further IC_{50} assays using the FRET technique. FRET is a non-radioactive process where the energy transfer occurs from an excited donor fluorophore to a suitable acceptor fluorophore. The donor fluorophore absorbs energy from incoming light, which is then transmitted to a nearby acceptor molecule (Bokhari & Albukhari, 2021).

The advantage of FRET compared to other methods is the great sensitivity of this method of detection and the accessibility of a wide range of instruments and automated techniques to carry out. These assays can be performed by (a) tracking the fluorescence of the amino acid tryptophan, (b) monitoring the displacement of a known ligand that has been fluorescently labelled (fluorescence anisotropy), or (c) transfer of energy from one binding partner to another (Murray & Wigglesworth, 2017).

For this project, dual-labelled DNA substrates were used, with a 5' – Cy3 (Cyanine-3, a fluorescent yellow compound with an excitation peak at 555 nm and emission peak at 568 nm) and 3' – FAM (Carboxyfluorescein, a fluorescent green compound with an excitation peak at 495 nm and an emission peak at 520 nm; Figure 6.1).

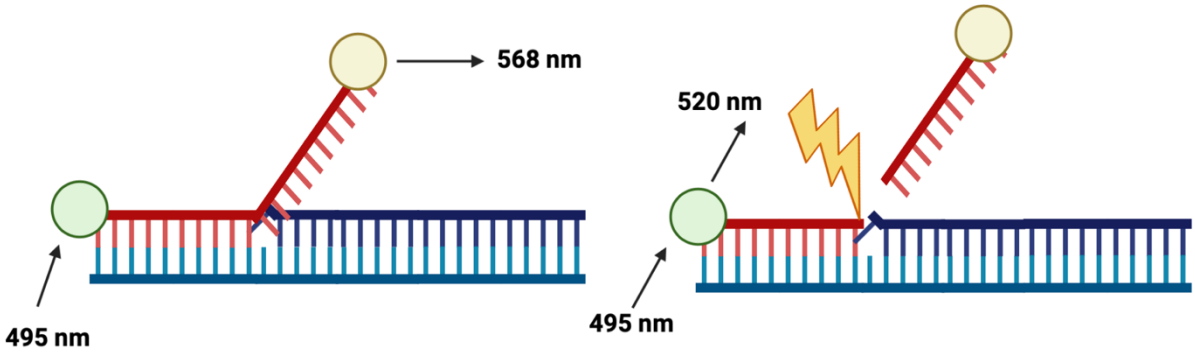


Figure 6.1 Schematic representation of FRET assays using dual-labelled DNA substrate.

The fluorophore 5' – Cy3 is coloured in yellow, 3' – FAM coloured in green and FEN activity presented as a yellow lightning bolt. Flap structures are specifically cleaved by FEN enzymes which results in separation of donor and quencher dyes. This allows real time monitoring of the reaction in a spectrofluorometer. Comparison of uninhibited progress curves with reactions carried out in the presence of candidate inhibitors allows rapid screening.

This chapter will include the results obtained from the initial inhibitor screening and IC_{50} determinations of *Tb*FEN, *Li*FEN, *Cn*FEN and human FEN. From the initial screening, analyses of the results were carried out using the software DataWarrior 5.5.0 (Sander *et al.,* 2015) to compare the percentage of inhibition from each flap endonuclease. To determine the IC_{50} values from the potential hits selected from the initial screening was performed using the software GraphPad Prism 9.4.1 (Transformation of data and non-linear fit analysis log(inhibitor) vs. response – Variable slope (four parameters) was performed using GraphPad Software, San Diego, CA, USA, https://graphpad.com/scientific-software/prism).

6.2 Optimizing FRET assays for testing small-molecule libraries

The concentration of the enzyme affects the rate of the reaction, which means that optimization of the assay needs to be carried out to determine the best enzyme concentration according to the data collected during the assay. *Li*FEN and *Tb*FEN were tested at four different concentrations: 10, 1, 0.1, and 0.01 µg/mL. *Li*FEN and *Tb*FEN at 1 µg/mL were considered to give appropriate fluorescence curves. Under the optimised conditions for each enzyme, the extent of the control (uninhibited) reaction was in the range 40-50% of the maximum change in fluorescence possible for *Li*FEN and *Tb*FEN after ten minutes. The Z' value calculated for the optimized assays were 0.73 for *Li*FEN and 0.72 for *Tb*FEN, suggesting a suitable assay for single-hit screening. These results are shown in Figure 6.2.

*Cn*FEN was tested at two concentrations: 10 and 1 µg/mL; and with two different salts in the reaction buffer: $MgCl_2$ and $MnCl_2$ at four concentrations: 10, 5, 2.5 and 1 mM. *Cn*FEN at 1 µg/mL using the reaction buffer with $MnCl_2$ at 5 mM gave the most appropriate fluorescence curve. Under these conditions, representing an extent of reaction of approximately 60% in 10 minutes. A calculated Z' value of 0.86 was achieved (Figure 6.2).

	*Li*FEN	*Tb*FEN	*Cn*FEN
Neg Mean	-0.2	-0.2	0.27
Neg SD	0.17	0.17	0.03
Pos Mean	4.4	5.0	6.04
Pos SD	0.25	0.31	0.24
Z'	**0.73**	**0.72**	**0.86**

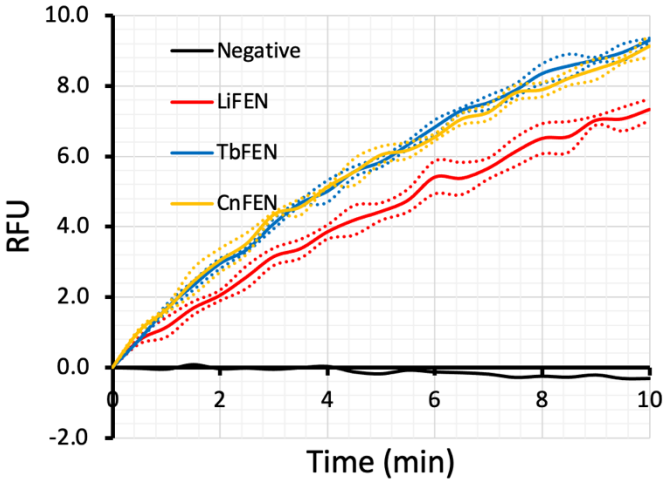


Figure 6.2 Optimization of FRET assay for *Tb*FEN, *Li*FEN and *Cn*FEN.

The concentration of the three proteins considered appropriated for further assays according to the change in RFU was 1 µg/ml. As expected, the negative control, which contains no enzyme, do not show any change in RFU.

6.3 FRET screen of microbial FENs and human FEN against a small-molecule library

The 1166 compound BioNet 2nd generation fragment library was screened at 1 mM concentration against 0.5 µg/ml *Li*FEN, 1 µg/ml *Tb*FEN, 0.75 µg/mL *Cn*FEN and 0.5 µg/ml *Hs*FEN using the optimized FRET assays (Figure 6.3). The top 20 potential inhibitors for each microbial FEN were identified according to the highest percentage of inhibition from each compound (Figure 6.4 – 6.6).

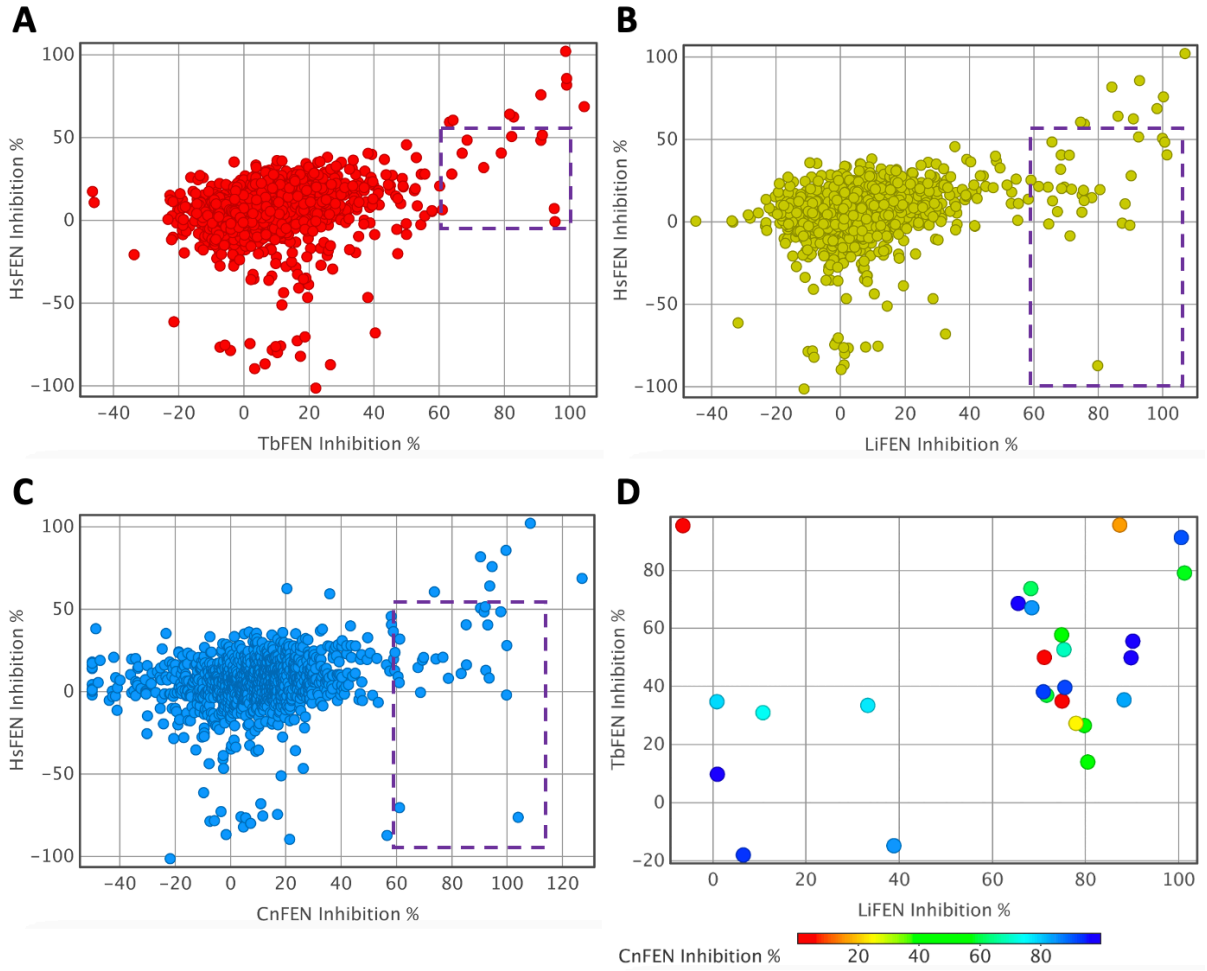


Figure 6.3 Screening FENs against the BioNet fragment library

(A) *Diagram showing percentages of inhibition in* Tb*FEN (X-axis;* $\overline{Z}$' *= 0.64; SD = 0.04) against the human FEN (Y-axis;* $\overline{Z}$' *= 0.62; SD = 0.08).* ***(B)*** *Diagram depicting percentages of inhibition in* Li*FEN (X-axis;* $\overline{Z}$' *= 0.65; SD = 0.07) against human FEN.* ***(C)*** *Diagram showing percentage of inhibition for* Cn*FEN (X-axis;* $\overline{Z}$' *= 0.69; SD = 0.08) against human FEN. Purple dotted squares in (A) (B) and (C) denotes potential hits detected on the FRET assays.* ***(D)*** *Twenty-six potential hits were selected from the 1166-BioNet library considering an inhibition percentage of 65% or higher against any microbial FEN and inhibition percentage of 50% or lower against human FEN. X-axis corresponds to percentage of inhibition in* Tb*FEN. Y-axis corresponds to percentage of inhibition in* Li*FEN. The color range wavelength of the markers represents percentage of inhibition in* Cn*FEN (red → dark blue). Chemical structures of potential hits for* IC_{50} *assays are depicted in Figure 6.11.*

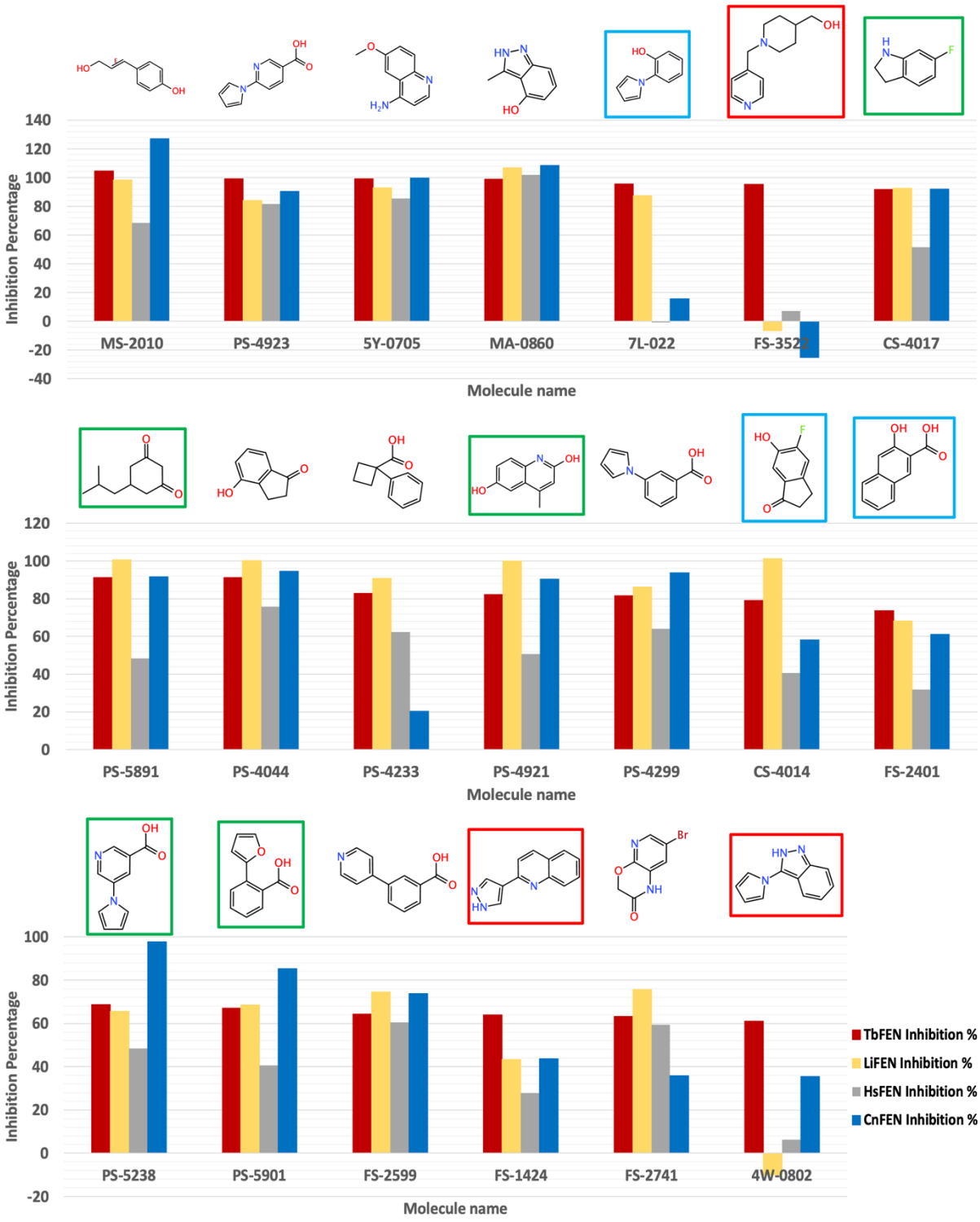


Figure 6.4 Top 20 inhibitors of *Tb*FEN

Red boxed compounds represent molecules inhibiting specifically Tb*FEN. Blue boxed compounds are molecules that inhibit* Tb*FEN and* Li*FEN. Green boxed compounds correspond to molecules that inhibit* Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. The cutoff of 65% inhibition or above against any microbial FENs and cutoff of 55% inhibition or below in the human FEN were considered for the colour labelling.*

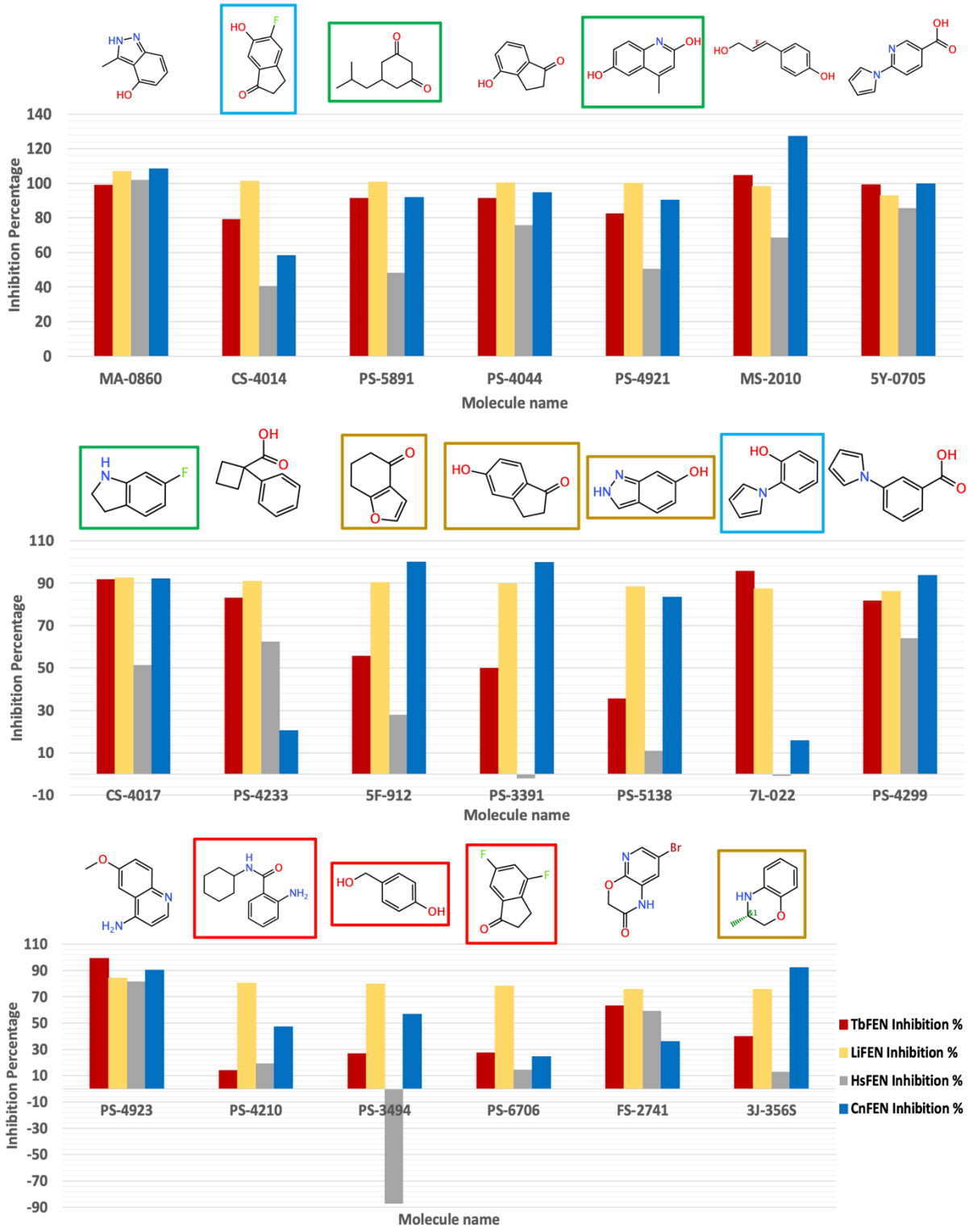


Figure 6.5 Top 20 inhibitors of *Li*FEN

Red boxed compounds represent molecules inhibiting specifically Li*FEN. Blue boxed compounds are molecules that inhibit* Li*FEN and* Tb*FEN. Gold boxed compounds correspond to molecules inhibiting* Li*FEN and* Cn*FEN Green boxed compounds represent molecules that inhibit* Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. The cutoff of 65% inhibition or above against any microbial FENs and cutoff of 55% inhibition or below in the human FEN were considered for the colour labelling.*

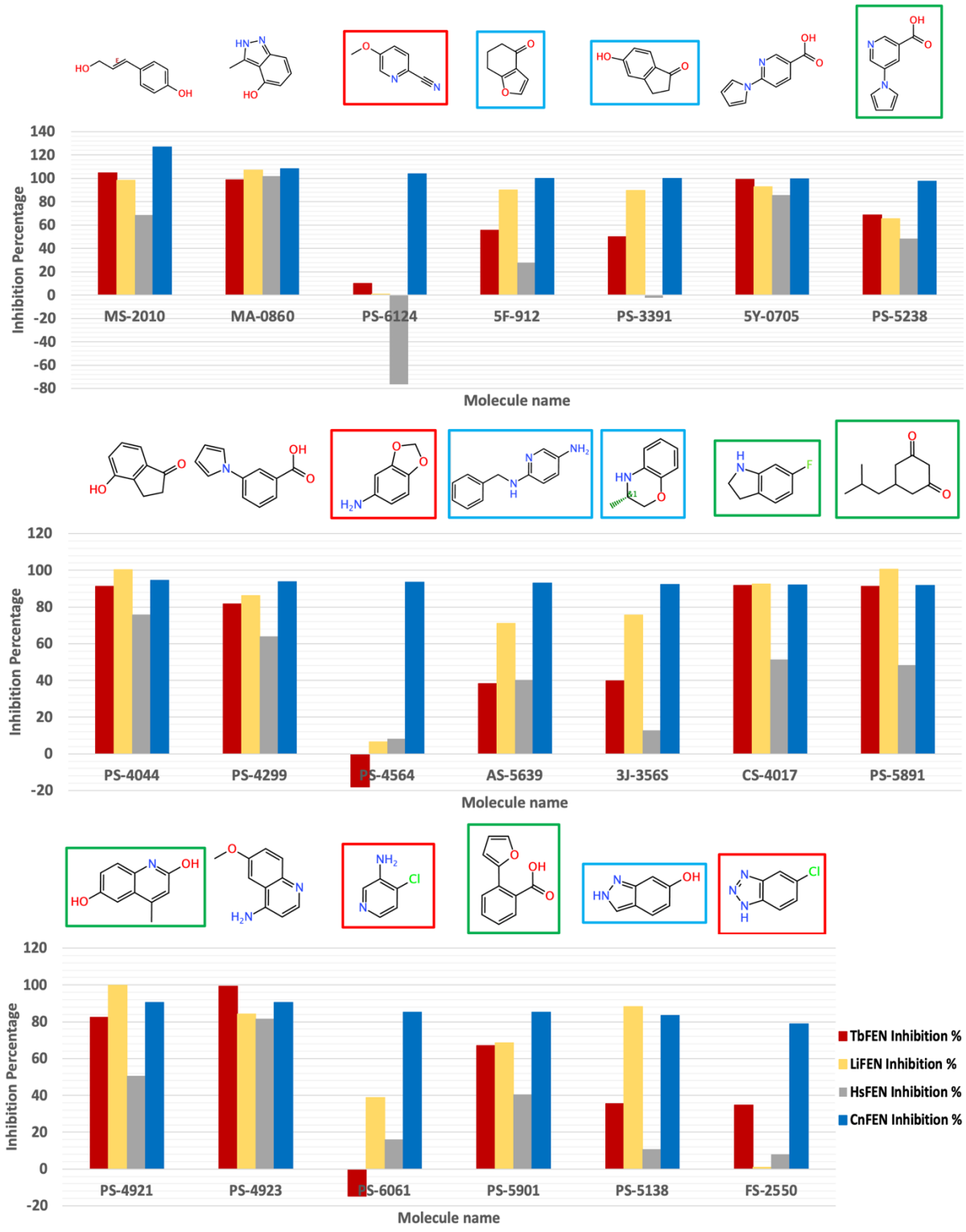


Figure 6.6 Top 20 inhibitors of *Cn*FEN

Red boxed compounds represent molecules inhibiting specifically Cn*FEN. Blue boxed compounds are molecules that inhibit* Cn*FEN and* Li*FEN. Green boxed compounds correspond to molecules that inhibit* Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. The cutoff of 65% inhibition or above against any microbial FENs and cutoff of 55% inhibition or below in the human FEN were considered for the colour labelling.*

6.4 FRET screen of microbial FENs and human FEN against an in-house FEN fragment library

The HitExpansion library consists of a unique collection of 400 molecules with substructures resembling previously identified compounds from several screening campaigns against other FEN enzymes previously carried out in Sheffield. These compounds had activity against homologues from bacteria, *Plasmodium falciparum* and the human enzyme (Sayers laboratory, unpublished results).

Due to time constrictions, a specific subset of the HitExpansion library was screened at 1 mM concentration against 0.5 µg/ml *Li*FEN, 1 µg/ml *Tb*FEN, 0.75 µg/mL *Cn*FEN and 0.5 µg/ml *Hs*FEN using the optimized FRET assays (Figure 6.7). The top 20 potential inhibitors for each microbial FEN were identified according to the highest percentage of inhibition from each compound (Figure 6.8 – 6.10).

The selection of the HitExpansion specific subset library was performed according to the preliminary results obtained from vHTS from Section 4.2. The vHTS was carried out using the AutoDock 4 programme with the default parameters and considering the presence of the two Mg^{2+} ions. An analysis of the docked compounds was done, selecting two batches from the 8 plates conforming the library for physical screening according to the plates containing the compounds with the highest number of virtual hits. From the 400 compounds of the HitExpansion library, 126 molecules were physically screened against the four FENs through FRET assays (Figure 6.7).

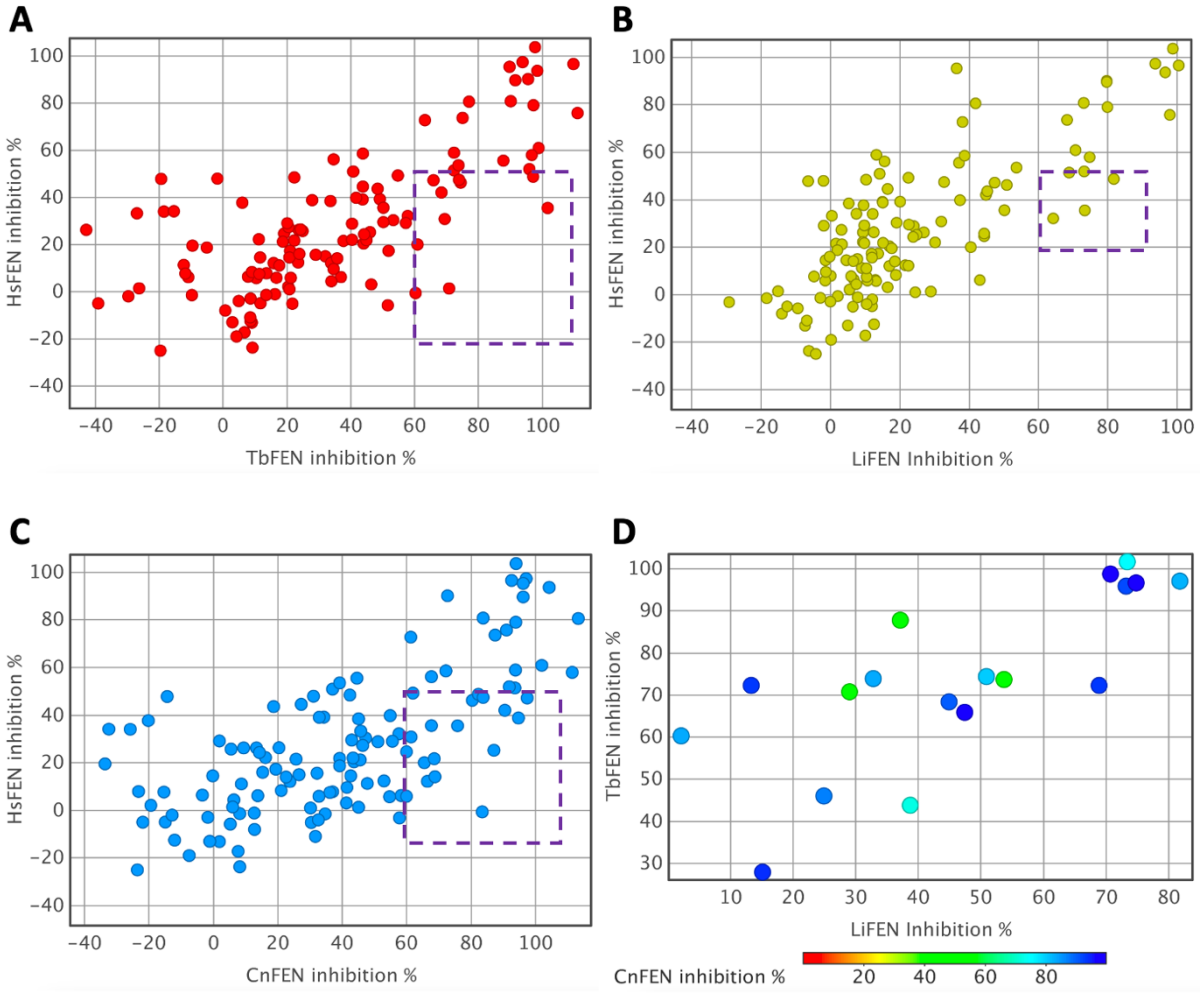


Figure 6.7 Screening FENs against the HitExpansion fragment library subset.

(A) *Diagram showing percentages of inhibition in* Tb*FEN (X-axis;* $\overline{Z}$' *= 0.71; SD = 0.02) against the human FEN (Y-axis;* $\overline{Z}$' *= 0.67; SD = 0.04).* ***(B)*** *Diagram depicting percentages of inhibition in* Li*FEN (X-axis;* $\overline{Z}$' *= 0.64; SD = 0.03) against human FEN.* ***(C)*** *Diagram showing percentage of inhibition for* Cn*FEN (X-axis;* $\overline{Z}$' *= 0.61; SD = 0.02) against human FEN. Purple dotted squares in (A) (B) and (C) denotes potential hits detected on the FRET assays.* ***(D)*** *Eighteen potential hits were selected from the HitExpansion subset library considering an inhibition percentage of 70% or higher against any microbial FEN and inhibition percentage of 61% or lower against human FEN. X-axis corresponds to percentage of inhibition in* Tb*FEN. Y-axis corresponds to percentage of inhibition in* Li*FEN. The color range wavelength of the markers represents percentage of inhibition in* Cn*FEN (red → dark blue). Chemical structures of available potential hits for* IC_{50} *assays are depicted in Figure 6.11.*

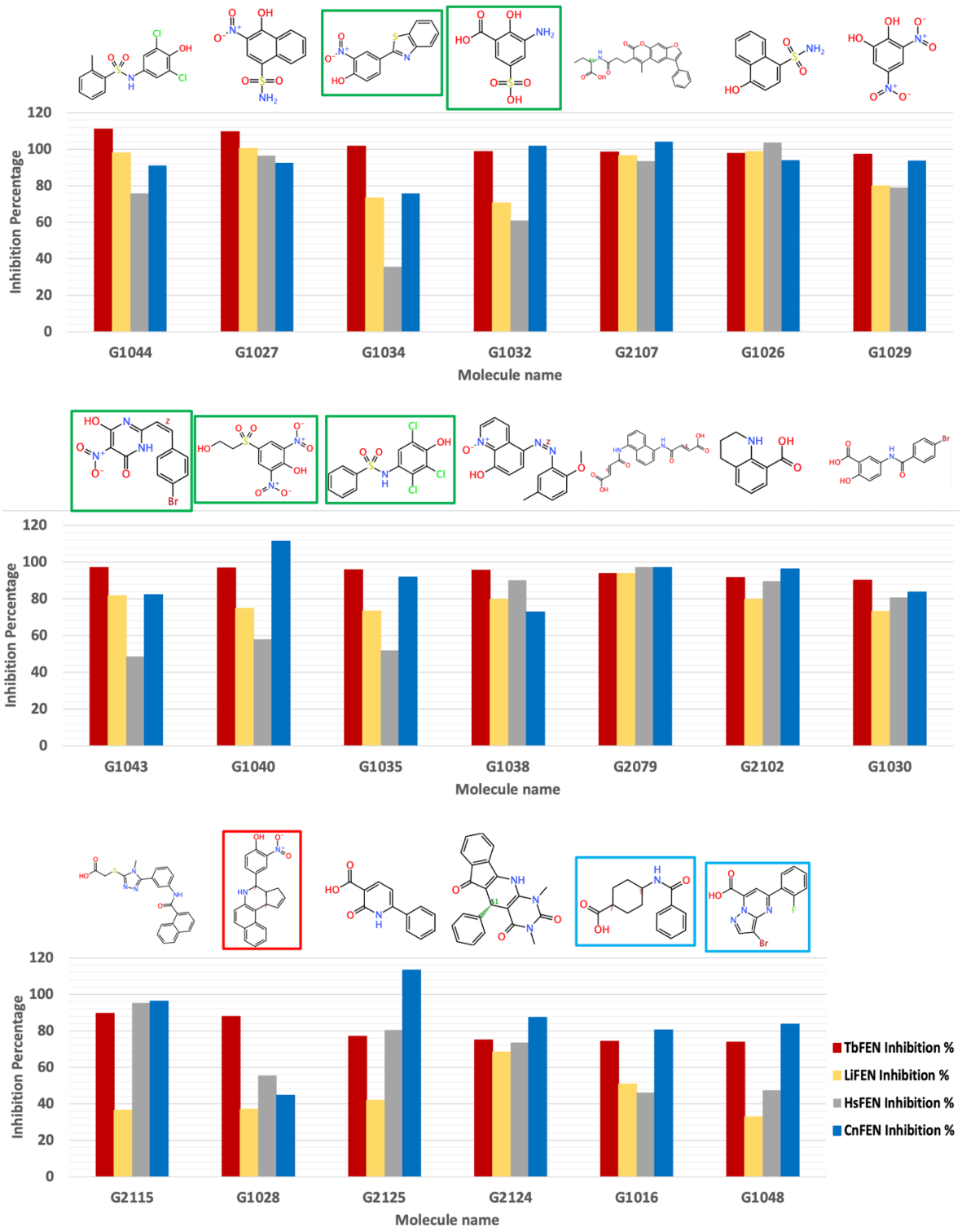


Figure 6.8 Top 20 inhibitor of *Tb*FEN

Red boxed compounds represent molecules inhibiting specifically Tb*FEN. Blue boxed compounds are molecules that inhibit* Tb*FEN and* Cn*FEN. Green boxed compounds correspond to molecules that inhibit* Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. The cutoff of 70% inhibition or above against any microbial FENs and cutoff of 61% inhibition or below in the human FEN were considered for the colour labelling.*

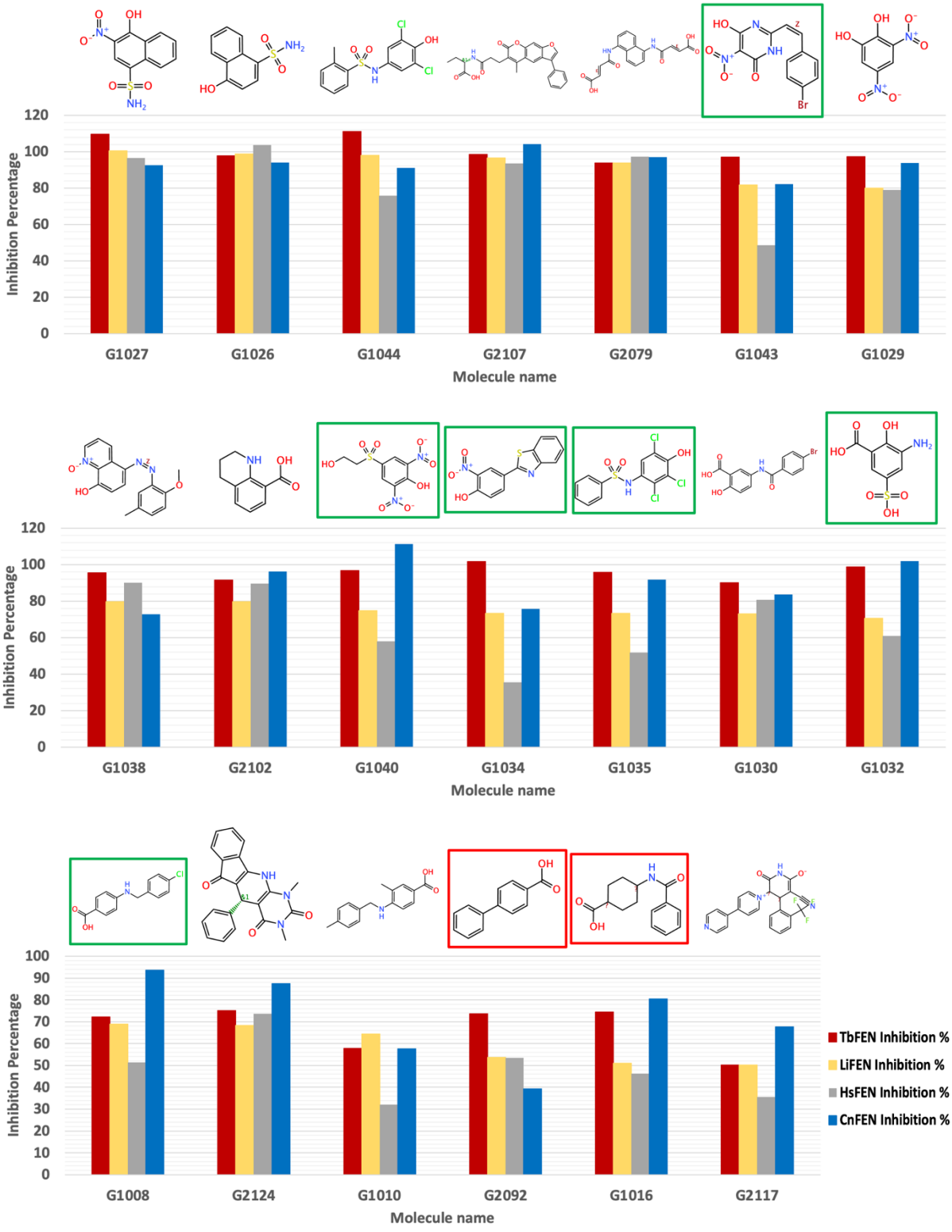


Figure 6.9 Top 20 inhibitors of *Li*FEN

Green boxed compounds represent molecules that inhibit Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. Red boxed compounds represent molecules inhibiting specifically* Tb*FEN. The cutoff of 70% inhibition or above against any microbial FENs and cutoff of 61% inhibition or below in the human FEN were considered for the colour labelling.*

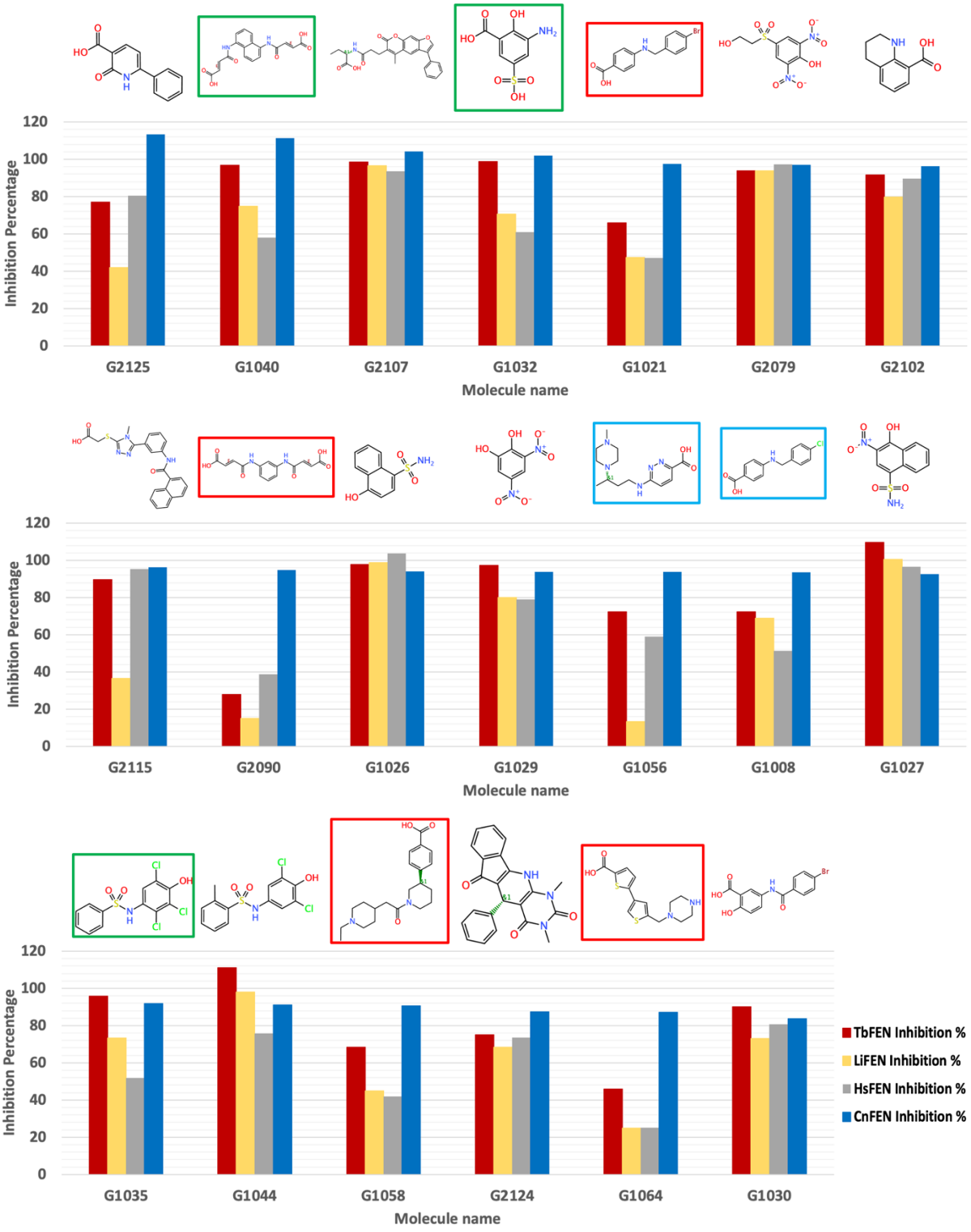


Figure 6.10 Top 20 inhibitors *Cn*FEN

Red boxed compounds represent molecules inhibiting specifically Cn*FEN. Blue boxed compounds are molecules that inhibit* Cn*FEN and* Li*FEN. Green boxed compounds correspond to molecules that inhibit* Tb*FEN,* Li*FEN and* Cn*FEN without inhibiting the human FEN. The cutoff of 70% inhibition or above against any microbial FENs and cutoff of 61% inhibition or below in the human FEN were considered for the colour labelling.*

6.5 Dose-response assays and IC_{50} estimation of selected molecules against microbial FENs and human FEN

According to the results obtained from screening both libraries, 23 of 26 hit compounds from the BioNet library and 13 of the 18 hit compounds from the HitExpansion library were available and selected for dose-response assays (Figure 6.11).

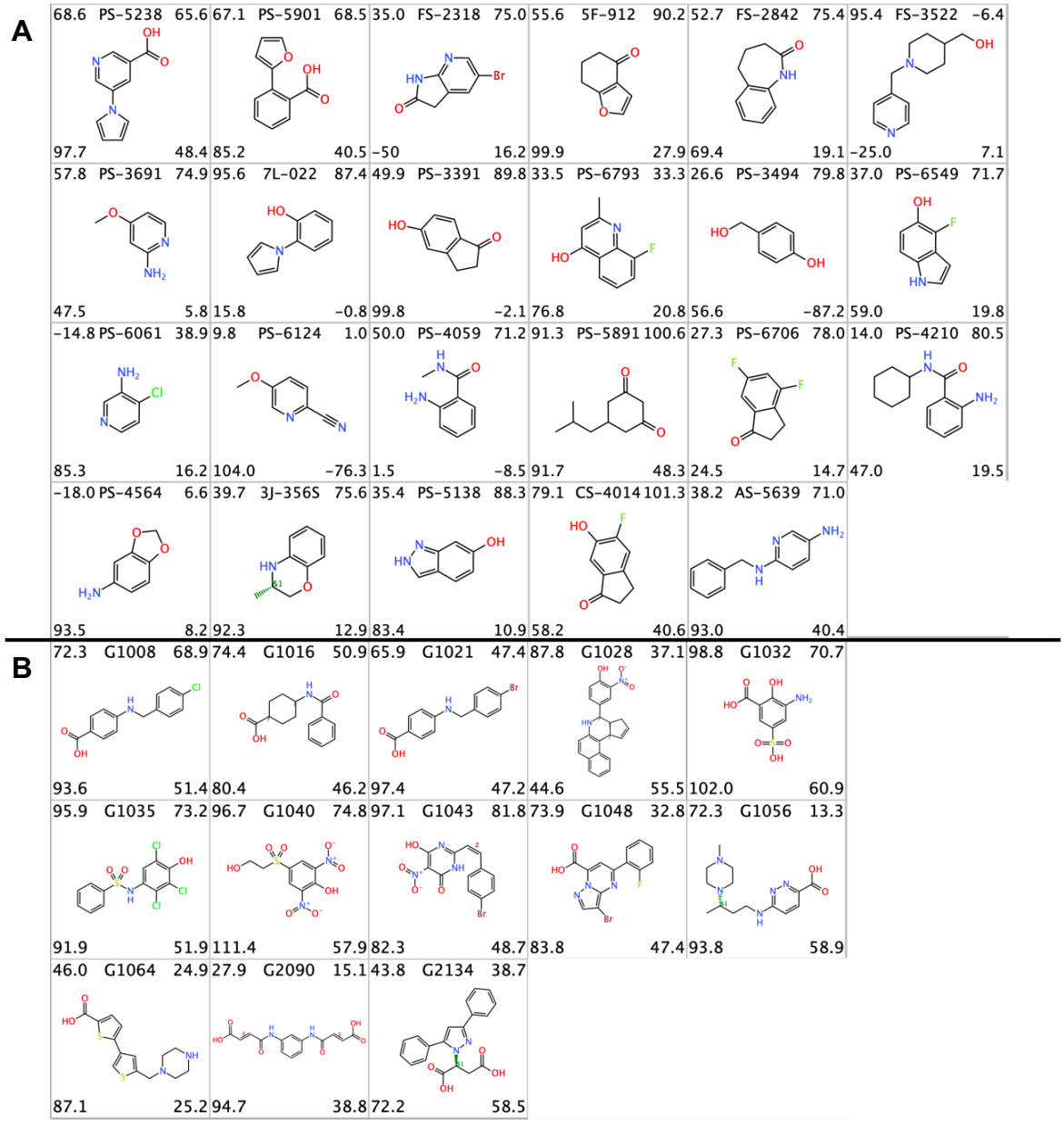


Figure 6.11 Molecules selected from BioNet and HitExpansion libraries for dose-response assays.

*(**A**) Molecules selected from the BioNet library. (**B**) Molecules selected from the HitExpansion library.* Tb*FEN % of inhibition: Top right corner.* Li*FEN % of inhibition: Top left corner.* Cn*FEN % of inhibition: Bottom right corner. Human FEN % of inhibition: Bottom left corner. Molecule name and code name at the top center for the BioNet and HitExpansion libraries, respectively.*

Of the 23 molecules from the BioNet library that were tested in dose-response assays, several proved to be false positives. Only 16 molecules inhibited the activity of at least one microbial FEN at a concentration of 1 mM or less by more than 50% (Figure 6.12 and Figure 6.13)

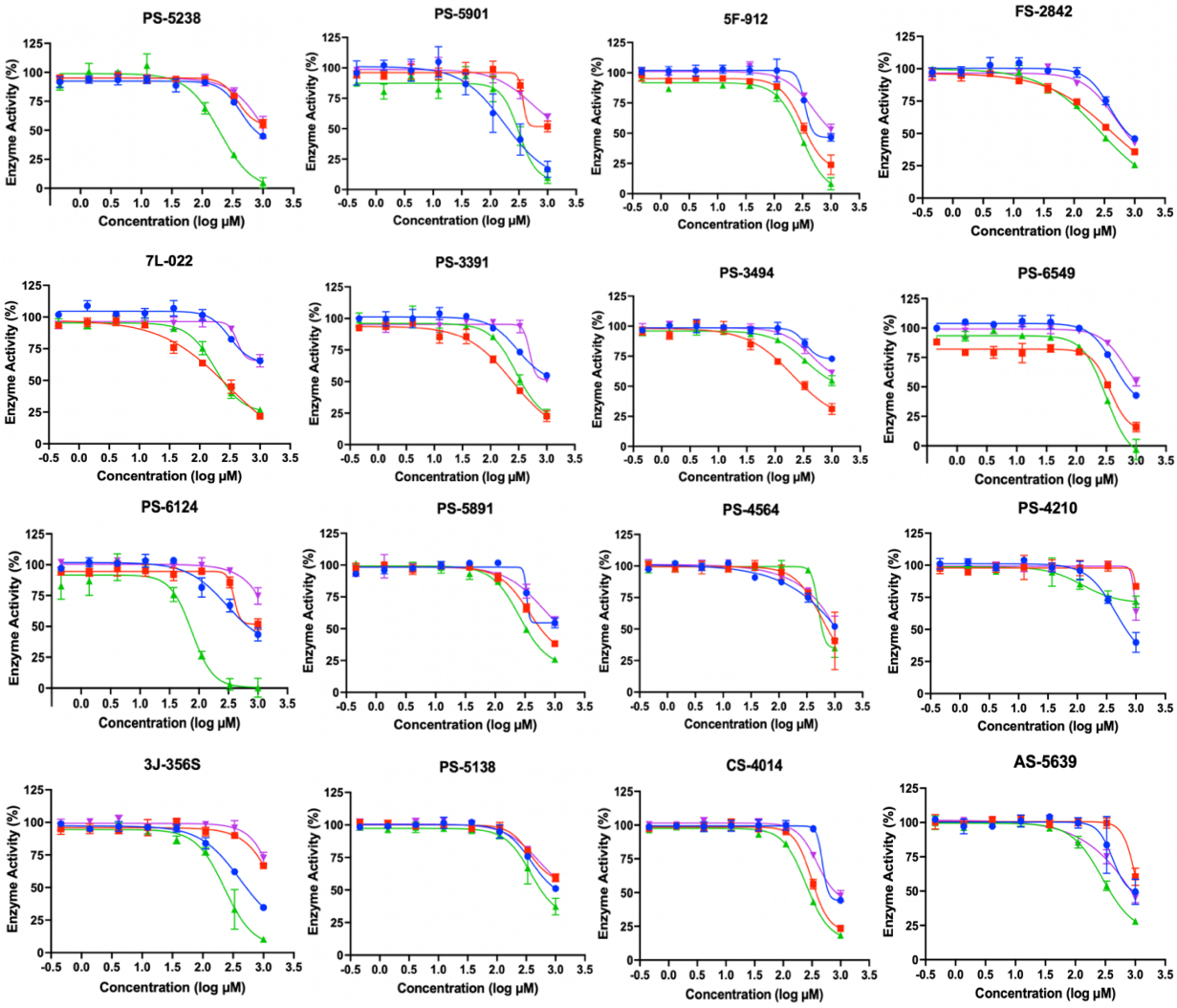


Figure 6.12 Dose-response assays of 16 molecules from the BioNet library

Blue – TbFEN activity %; Red – LiFEN activity %; Green – CnFEN activity %; Pink – Human FEN activity %. Compounds were tested in triplicate at concentrations of 1000, 333, 111, 37, 12, 4.1, 1.4 and 0.5 μM. Calculated mean and SD of the activity of each enzyme and plots were computed using GraphPad Prism 9.4.1.

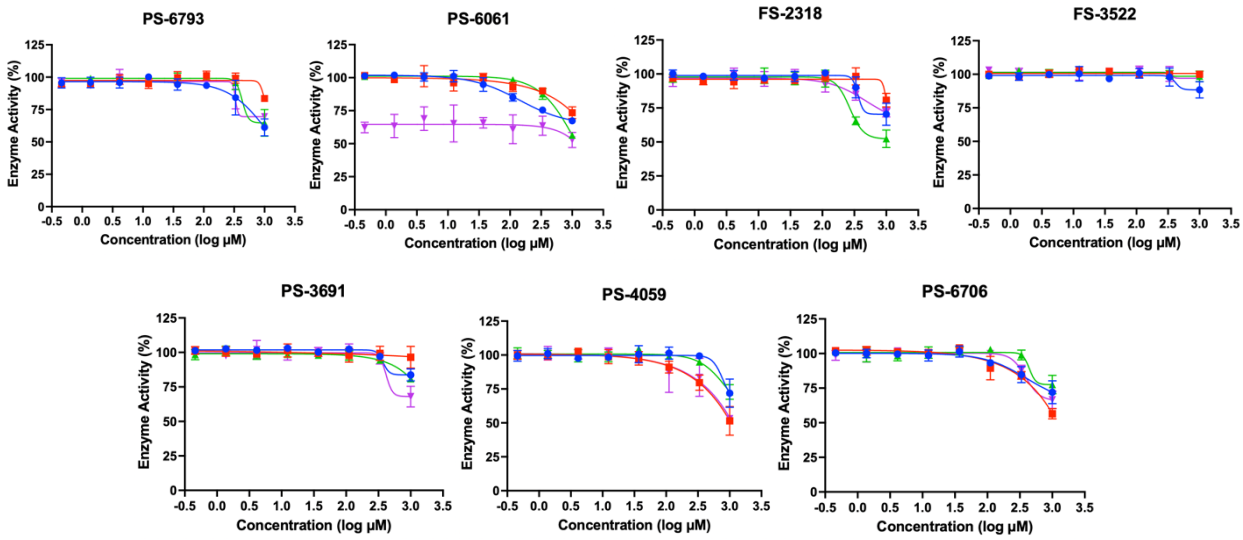


Figure 6.13 Dose-response assays of 9 false-positive molecules from the BioNet library

Blue – TbFEN activity %; Red – LiFEN activity %; Green – CnFEN activity %; Pink – Human FEN activity %. Compounds were tested in triplicate at concentrations of 1000, 333, 111, 37, 12, 4.1, 1.4 and 0.5 µM. Calculated mean and SD of the activity of each enzyme and plots were computed using GraphPad Prism 9.4.1.

Of the 13 molecules from the HitExpansion library, all the molecules showed significant inhibitions at 1mM or less by more than 50% against one or more of the microbial FENs tested (Figure 6.14).

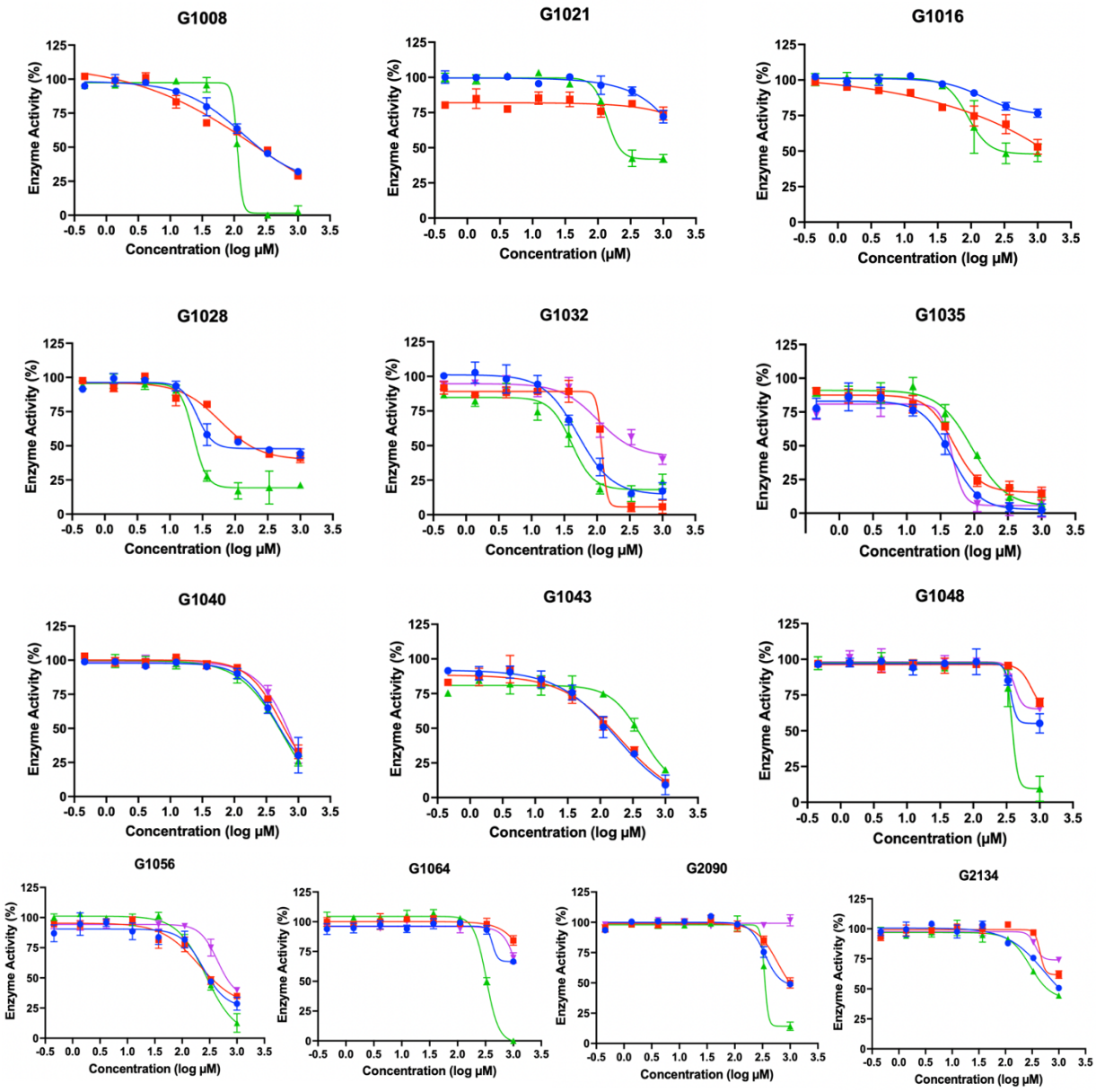


Figure 6.14 Dose-response assays of 11 molecules from the HitExpansion library

Blue – TbFEN activity %; Red – LiFEN activity %; Green – CnFEN activity %; Pink – Human FEN activity %. Compounds were tested in triplicate at concentrations of 1000, 333, 111, 37, 12, 4.1, 1.4 and 0.5 µM. Calculated mean and SD or the activity of each enzyme and plots were computed using GraphPad Prism 9.4.1.

Data collected from the dose-response assays performed on the selected molecules of both libraries was used to estimate the half maximal inhibitory concentration (IC_{50}) against the microbial and human FEN. The IC_{50} is a quantitative measure that shows the necessary concentration of the compound to inhibit a biological process, in this case an enzymatic reaction, by 50%. The coefficient of determination (R^2) was also calculated for each compound tested for each FEN. This will determine how good the models from the collected results are in terms of accuracy during the experimental procedure.

IC_{50} and R^2 for the selected molecules against each FEN were calculated from the non-lineal models previously produced from GraphPad Prism 9.4.1. The IC_{50}'s calculated in Tables 6.1 and 6.2 correspond to the best-fit values for each model. Occasionally, the software is unable to estimate a complete interval when calculating the profile likelihood confidence intervals (95% confidence level) for parameters in non-linear regression. Taking into consideration the last statement, some of the best-fit values should be interpreted with caution.

Table 6.1 Estimations of IC_{50} and R^2 from the 16 molecules of the BioNet library*

	*Tb*FEN		*Li*FEN		*Cn*FEN		Human FEN	
	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2
PS-5238	**428**	**0.96**	**390**	**0.96**	194	0.97	**>1000**	**0.92**
PS-5901	177	0.91	**369**	**0.91**	315	0.94	**567**	**0.91**
5F-912	**344**	**0.96**	313	0.98	306	0.99	488	0.95
FS-2842	**390**	**0.96**	400	0.99	265	0.99	477	0.97
7L-022	**315**	**0.94**	293	0.98	179	0.99	**418**	**0.89**
PS-3391	303	0.96	244	0.98	287	0.96	**497**	**0.96**
PS-3494	**339**	**0.9**	207	0.96	338	0.96	**441**	**0.95**
PS-6549	**420**	**0.98**	**366**	**0.97**	317	0.99	**678**	**0.96**
PS-6124	297	0.94	**392**	**0.93**	75	0.96	**>1000**	**0.83**
PS-5891	**336**	**0.95**	368	0.98	255	0.99	**610**	**0.92**
PS-4564	**>1000**	**0.97**	**574**	**0.88**	**507**	**0.98**	**>1000**	**0.95**
PS-4210	454	0.96	**983**	**0.55**	**123**	**0.86**	**968**	**0.94**
3J-356S	407	0.99	**>1000**	**0.88**	228	0.97	**>1000**	**0.84**
PS-5138	381	0.97	**360**	**0.97**	**374**	**0.96**	**521**	**0.96**
CS-4014	**484**	**0.98**	313	0.99	250	0.99	>1000	0.95
AS-5639	**417**	**0.84**	**>1000**	**0.94**	282	0.99	**>1000**	**0.96**

**Values in bold correspond to the models which were unable to calculate a complete confidence interval, either the upper limit, lower limit, or both.*

Table 6.2 Estimations of IC_{50} and R^2 from the 13 molecules of the HitExpansion library*

	*Tb*FEN		*Li*FEN		*Cn*FEN		Human FEN	
	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2	IC_{50} (µM)	R^2
G1008	132	0.99	**231**	**0.97**	**113**	**0.99**	513[#]	~
G1016	149	0.94	**>1000**	**0.94**	**89**	**0.93**	>1000[#]	~
G1021	**>1000**	**0.87**	**>1000**	**0[&]**	**138**	**0.98**	n.t.[#]	~
G1028	**26**	**0.97**	59	0.97	**23**	**0.99**	>1000[#]	~
G1032	51	0.98	**119**	**0.99**	**41**	**0.96**	**103**	**0.92**
G1035	45	0.97	50	0.98	94	0.98	**47**	**0.97**
G1040	449	0.96	538	0.98	617	0.99	**>1000**	**0.98**
G1043	174	0.97	203	0.96	**433**	**0.94**	n.t.[#]	~
G1048	**366**	**0.92**	**756**	**0.92**	**383**	**0.97**	**417**	**0.89**
G1056	244	0.93	203	0.96	270	0.98	**443**	**0.96**
G1064	**422**	**0.89**	**>1000**	**0.69**	**328**	**0.99**	**>1000**	**0.9**
G2090	348	0.96	**592**	**0.96**	**344**	**0.99**	**>1000**	**0[&]**
G2134	**546**	**0.94**	**445**	**0.92**	289	0.96	**361**	**0.93**

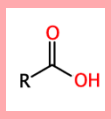
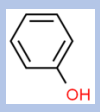
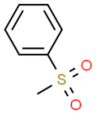
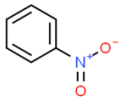
**Values in bold correspond to the models which were unable to calculate a complete confidence interval, either the upper limit, lower limit, or both.*

[#]*Data obtained from previous assays performed in Sayers laboratory. n.t. corresponds to molecule not tested.*

[&]*This model does not behave in a non-lineal regression because the compound did not inhibit enzyme activity in all tested concentrations.*

Observing the chemical structure of the potential inhibitors might give us a clue about the functional groups which interact with the proteins. For example, the molecules G1016 and G1032 possess a carboxylic acid group (coloured in light red in Table 6.3), which is a polar molecule with the characteristic of being a hydrogen-bond donor and acceptor. Furthermore, molecules G1028 and G1035 contains a substructural phenol group, which is also a polar molecule and can produce hydrogen bonds (coloured in light blue in Table 6.3). These functional groups and other chemical substructures have been identified in both BioNet and HitExpansion libraries and compared which ones are most likely to be of special interest for further analysis. This was done by classifying the compounds according to the functional groups and filtering which ones were selected as hits from the initial screening and later on the molecules with low IC_{50} (Table 6.3).

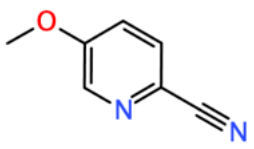
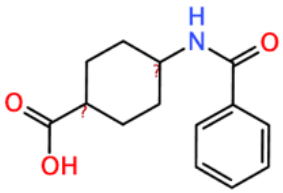
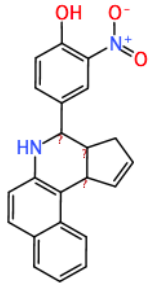
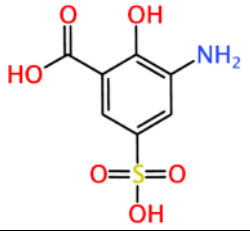
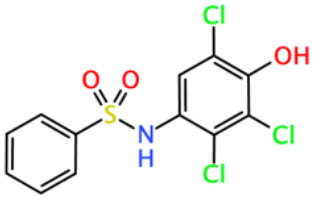
Table 6.3 Potential functional groups which give inhibitory activity to FENs*

Functional Group	Chemical substructure	Hits BN	Molecule name	#BN D.R.	Molecule name	Hits HE	Molecule name	#HE D.R.	Molecule name
Carboxylic acid		2	PS-5238 PS-5901	0	-	9	G1008 G1016 G1021 G1032 G1048 G1056 G1064 G2090 G2134	2	G1016 G1032
Phenol		7	7L-022 PS-3391 PS-3494 PS-6549 PS-4564 3J-356S CS-4014	0	-	4	G1028 G1032 G1035 G1040	3	G1028 G1032 G1035
Keto		9	FS-2318 5F-912 FS-2842 PS-3391 PS-4059 PS-5891 PS-6706 PS-4210 CS-4014	0	-	1	G1016	1	G1016
Pyridine		8	PS-5238 FS-2318 FS-3522 PS-3691 PS-6793 PS-6061 PS-6124 AS-5639	1	PS-6124	0	-	0	-
Pyrrole		3	PS-5238 7L-022 PS-6549	0	-	0	-	0	-
Pyrazole		1	PS-5138	0	-	2	G1048 G2134	0	-
Dimethyl ether		2	PS-3691 PS-6124	1	PS-6124	0	-	0	-
(Methyl-sulphonyl) benzene		0	-	0	-	3	G1032 G1035 G1040	2	G1032 G1035
Nitrobenzene		0	-	0	-	2	G1028 G1040	1	G1028

** Hits BN: 23 potential hits from the BioNet library; #BN D.R.: number of molecules with reproducible dose-response assay giving an IC_{50} lower than 100 µM for at least one microbial FEN; Hits HE: 13 initial hit molecules from the HitExpansion library; #HE D.R.: number of molecules with reproducible dose-response assay giving an IC_{50} lower than 100 µM for at least one microbial FEN. Chemical substructures were retrieved from PubChem (Kim* et al.*, 2021). Full chemical structures of the BioNet and HitExpansion compounds are depicted in Figure 6.11.*

From the IC_{50} calculated for each molecule against the activity of the microbial FENs, molecules from both libraries with an IC_{50} lower than 100 µM for at least one microbial FEN are represented in Table 6.4. This table allows to visualize the chemical structure of the potential inhibitors and predict which functional groups are interacting with the proteins. In addition, the table facilitates the comparison of the IC_{50} between the microbial and human FENs.

Table 6.4 Potential microbial FEN inhibitors of both libraries with an IC_{50} lower than 100 µM

Molecule name	Structure	IC_{50} (µM) *Tb*FEN	IC_{50} (µM) *Li*FEN	IC_{50} (µM) *Cn*FEN	IC_{50} (µM) *Hs*FEN
PS-6124		297	392	75	>1000
G1016		149	>1000	89	>1000
G1028		26	59	23	>1000
G1032		51	119	41	103
G1035		45	50	94	47

In conclusion, the work presented in this chapter has demonstrated that fit-for-purpose FRET based inhibitors screen for selected microbial pathogens FEN enzymes has been established. In addition, some evidence that selected inhibition of individual FEN enzymes is possible as being presented. For example, the compound PS-6124 of the BioNet library, registering a specific inhibitory response to *Cn*FEN with an IC_{50} of 75 μM. At the same time, there is evidence of inhibitory response of all microbial FENs tested but not the human FEN, which is promising for the future development of antimicrobial drugs. The molecule G1028 of the HitExpansion library presents an inhibitory response against TbFEN, LiFEN and CnFEN with IC_{50} of 26, 59 and 23 μM, respectively (Table 6.4).

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1 Expression, purification, and biochemical characterization of microbial FENs

In summary, successful expression and purification of microbial enzymes was carried out in this work, followed by physicochemical characterization of those proteins. Yields of purified protein recovered from the purification steps varied depending on the construct and from the protein purification knowledge and practical skills I developed throughout this project. The yield of purified protein from each microbial enzyme/construct ranged from 0.66 mg (*Cn*FEN WT) to 4.5 mg of protein (*Tb*FEN D183K ΔC341) per g of cell paste. This proved to be sufficient material for biochemical assays in all the constructs and for crystallisation trials. Characterization of microbial FEN activities were carried out using UV Nuclease assays and structure specific substrate assays with fluorescent labelled flap constructs (Chapter 6). Interestingly, magnesium ions were found to be the preferred metal ions for the endonucleases from the human FEN, *Tb*FEN and *Li*FEN enzymes, but the *Cn*FEN appeared to preferred manganese ions.

The flap endonuclease enzymes can use a range of divalent metal ions to support catalysis. Garforth and co-workers (2001) showed that metal ions such as Zn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+}, Mg^{2+} and Mn^{2+} were able to catalyse structure specific hydrolysis of a branched DNA pseudo-Y structure in the FEN homologue bacteriophage T5-D15 exonuclease. Recently, manganese has been identified as the preferred cofactor in vitro for both the *Mycobacterium smegmatis* FenA and DNA polymerase I FEN domain (Uson *et al.*, 2018; Ghosh *et al.*, 2020). Similarly, the *Thermococcus barophilus* Ch5 FEN homologue was examined for cofactor specificity. Lin *et al.* (2022) tested the ability of metal ions Zn^{2+}, Ni^{2+}, Co^{2+}, Ca^{2+}, Cu^{2+}, Mg^{2+} and Mn^{2+} to support catalysis. They found that Mg^{2+} and Mn^{2+} were preferred over Ni^{2+} and Co^{2+} while Cu^{2+}, Zn^{2+} and Ca^{2+} showed little activity (relative activities 100%, ~95%, ~36%, ~30%, ~10%, ~6% and ~5%, respectively). Biological relevance of cofactor preference is not well studied. We cannot be certain whether a metal preference determined *in vitro* will be the same *in vivo*, but given that magnesium ions are far more abundant in cells than manganese or these other metals listed, it seems likely that magnesium or possible manganese

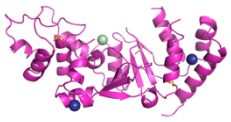
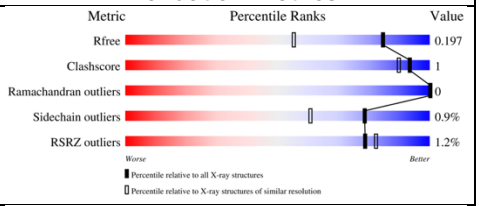
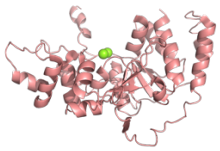
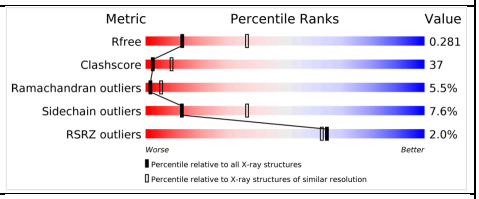
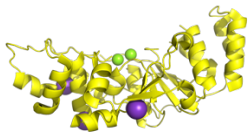
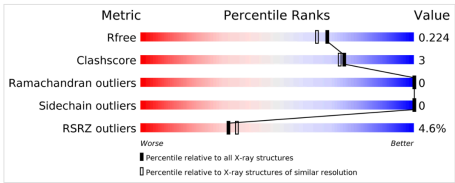
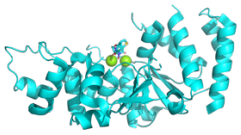
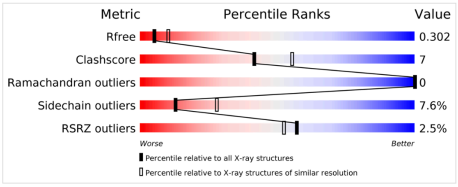
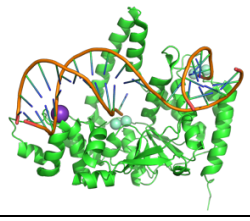
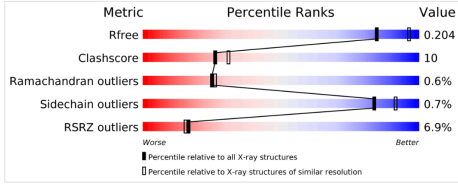
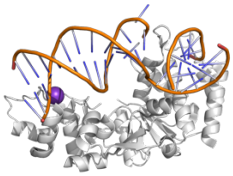
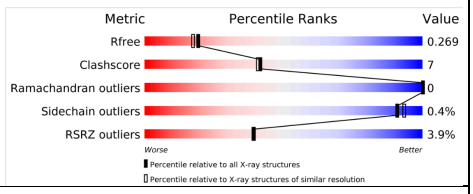
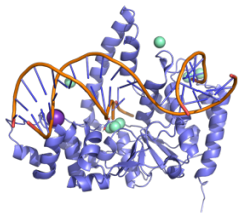
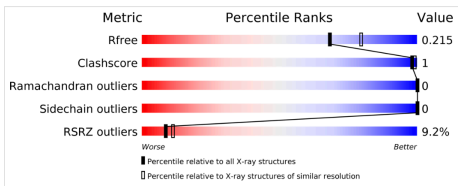
are the likely native cofactors. Concentrations of magnesium in cells (prokaryotic and mammalian) are reported to be in the range of 15 to 25 mM (Maguire & Cowan, 2002; Romani & Scarpa, 1992; Scarpa & Brinley, 1981). Most is likely to be bound to enzymes and, in particular, ATP or nucleotide metabolizing enzymes. This means that the concentration of free magnesium should be as low as around 1mM (Romani, 2013). This is considered high in comparison with the intracellular concentration of manganese, which is approximately 25 μM (Kodama *et al.,* 1991). In addition, the affinity of the metal ions to the active site has been measured in the bacteriophage T5 FEN homologue. Feng *et al.* (2004) examined the enzyme using isothermal titration calorimetry (ITC). Manganese had a higher affinity to the enzyme than magnesium (K_d = 2 and 31 μM, respectively). Considering that there is a thousand-fold excess of magnesium over manganese, it seems likely that magnesium is the biological relevant cofactor. The identity of the biological cofactor is important because magnesium and manganese have different properties, in particular manganese has various oxidation states but magnesium has only one and magnesium ions prefer harder ligands like oxygen (i.e., aspartic acid and glutamic acid) whereas manganese ions favour softer ligands such as nitrogen and sulphur (i.e., glutamine, asparagine, methionine and cysteine). The presence of magnesium or manganese in the active site would be expected to alter the types of small molecules which interact with the metal. For the purposes of the study presented, magnesium was used in the docking work.

7.2 Protein crystallisation and structure determination in microbial FENs

Once the microbial FEN protein was purified, crystallisation trials were performed using 6 different screens under thousands of conditions in an attempt to elucidate the structural conformation of the protein and, therefore, have a better understanding of the molecular mechanism of flap endonucleases in these pathogens. Despite many efforts, only very poorly diffracting co-crystals with DNA were obtained and this work could not be completed due to time constraints. Instead, attention was focussed on optimisation of crystals that showed promising diffraction characteristics. A structure of *Li*FEN diffracting to 1.4 Å was modelled from the experimental data. This structure represents the first flap endonuclease reported from a parasite and was solved at a

relatively good resolution for a FEN enzyme family member. A comparison of the structures obtained for human FEN protein (with and without DNA or inhibitors) is shown in Table 7.1.

Table 7.1 Comparison of *Li*FEN structure with published human FEN structures

Identity	Structure	Experimental data	Validation metrics
*Li*FEN (this work)		**Method:** X-RAY DIFFRACTION **Resolution:** 1.41 Å **R-Value Free:** 0.198 **R-Value Work:** 0.138	
hFEN-PCNA complex (1UL1)[1]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.90 Å **R-Value Free:** 0.284 **R-Value Work:** 0.220	
hFEN in apo form (5ZOD)[3]		**Method:** X-RAY DIFFRACTION **Resolution:** 1.90 Å **R-Value Free:** 0.226 **R-Value Work:** 0.197 **R-Value Observed:** 0.198	
hFEN with HU inhibitor (5FV7)[4]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.84 Å **R-Value Free:** 0.300 **R-Value Work:** 0.234 **R-Value Observed:** 0.237	
hFEN (WT), flap DNA, SM^{3+} and K^{+} (3Q8K)[5]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.20 Å **R-Value Free:** 0.213 **R-Value Work:** 0.168 **R-Value Observed:** 0.170	
hFEN D181A/R192F in complex with DNA (5ZOF)[6]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.25 Å **R-Value Free:** 0.269 **R-Value Work:** 0.235 **R-Value Observed:** 0.237	
hFEN D233N with cleaved product (5K97)[7]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.10 Å **R-Value Free:** 0.219 **R-Value Work:** 0.182 **R-Value Observed:** 0.184	

[1] Sakurai *et al.*, 2005. [2] Lancey *et al.*, 2020. [3] Xu *et al.*, 2018. [4] Exell *et al.*, 2016. [5] Tsutakawa *et al.*, 2011. [6] Xu *et al.*, 2018. [7] Tsutakawa *et al.*, 2017. Data gathered from the RCSB.org web server.

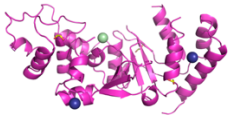
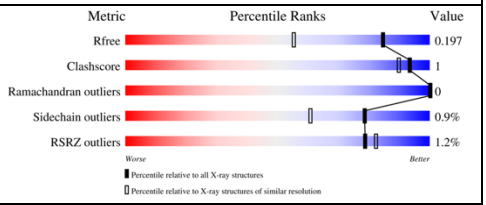
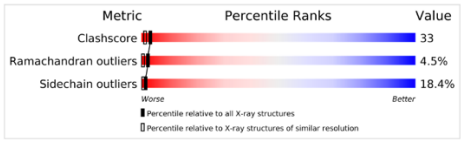
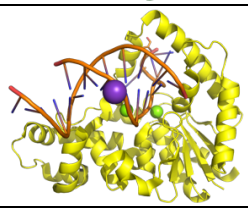
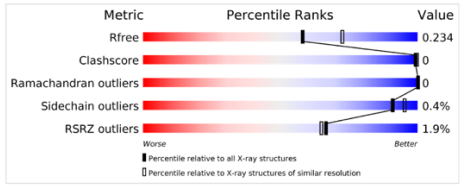
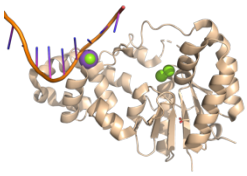
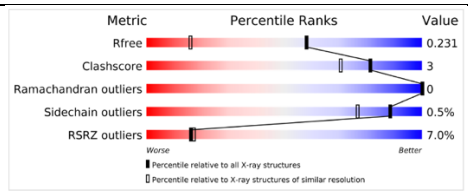
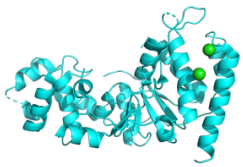
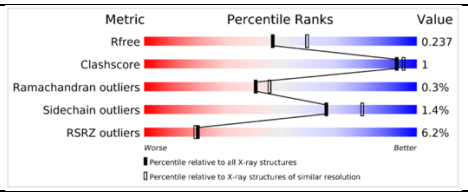
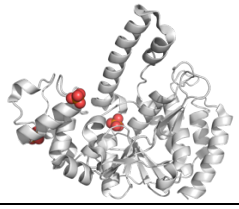
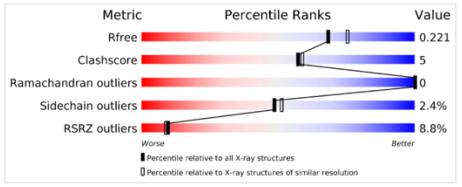
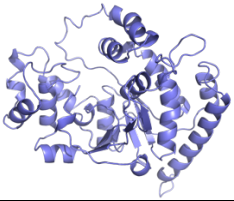
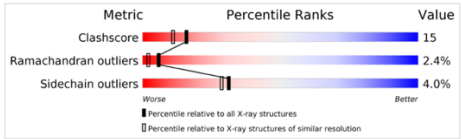
As can be seen, the reported resolutions range from 1.9 – 2.9 Å. If we look at the validation graphic summary for each protein, it appears that *Li*FEN compares quite well with deposited entries (most reported slider-metrics lie to the right in the validation graphic panels). Thus, the *Li*FEN structure appears to be of at least comparable quality and has been determined to a higher resolution (1.4 Å) than any of the human FEN1 structures reported to date.

The reported *Li*FEN structure in this thesis is incomplete, lacking the helical DNA binding regions seen in for example structures such as 3Q8K and 5K97. FEN proteins often have a potassium ion bound in the helix-2-turn-helix (H2TH) motif. The motif itself was clearly observed in my structure, but no convincing electron density for this metal ion was present, even though the crystallisation conditions contained potassium. Interestingly, the structures that show a bound potassium ion all have DNA bound to the enzyme (pdb files 3Q8K, 5ZOF and 5K97). In these structures, the potassium ion is nested at the top of the H2TH domain and interacts directly with a non-bridging phosphate oxygen on the DNA backbone. As there was no DNA in *Li*FEN crystallisation condition, this might explain why the potassium was not seen.

Another difference between published FEN structures and *Li*FEN is that only one divalent metal ion was clearly present in the active site as opposed to the two seen in the majority of reported structures in Table 7.1. Variability in the number of metal ions observed in FEN homologues has been reported. Kim *et al.* (1995) first published a structure containing a FEN domain (also called 5´nuclease) that was from *Thermus aquaticus* DNA polymerase I (or Taq polymerase). It had one metal bound in the active site (1TAQ.pdb). Later, crystal structures of bacteriophage T5 D15 protein (1EXN.pdb) and T4 RNAseH (1TFR.pdb; both of which are flap endonuclease homologues) were reported with two metals in their active sites (Ceska *et al.*, 1996; Mueser *et al.* 1996). Later AlMalki *et al.* (2016) reported three metals bound to the T5 enzyme (5HML.pdb). Prokaryotic enzymes possess additional conserved aspartic acids allowing them to bind up to three metals. An exception is the FEN family member, ExoIX, from E. coli

which lacks three conserved Asp residues compared with other bacterial FENs but binds two metals (Anstey-Gilbert *et al.*, 2013; 3ZDB.pdb).

Table 7.2 Comparison o *Li*FEN with more distant homologues

Identity	Structure	Experimental data	Validation metrics
*Li*FEN (this work)		**Method:** X-RAY DIFFRACTION **Resolution:** 1.41 Å **R-Value Free:** 0.198 **R-Value Work:** 0.138	
M. jannaschii FEN (1A76)[1]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.00 Å **R-Value Free:** 0.279 **R-Value Work:** 0.214 **R-Value Observed:** 0.214	
T5FEN in complex intact substrate and metal ions (5HNK)[2]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.22 Å **R-Value Free:** 0.234 **R-Value Work:** 0.183 **R-Value Observed:** 0.185	
E. coli ExoIX in complex with 5ov4 DNA, 2 Mg^{2+} and K^{+} (3ZDB)[3]		**Method:** X-RAY DIFFRACTION **Resolution:** 1.47 Å **R-Value Free:** 0.231 **R-Value Work:** 0.203 **R-Value Observed:** 0.204	
M. kandleri FEN (4WA8)[4]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.20 Å **R-Value Free:** 0.231 **R-Value Work:** 0.200 **R-Value Observed:** 0.202	
Desulfurococcus amylolyticus FEN (3ORY)[5]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.00 Å **R-Value Free:** 0.225 **R-Value Work:** 0.210 **R-Value Observed:** 0.211	
P. furiosus FEN (1B43)[6]		**Method:** X-RAY DIFFRACTION **Resolution:** 2.00 Å **R-Value Free:** 0.279 **R-Value Work:** 0.231 **R-Value Observed:** 0.231	

[1] Hwang et al., 1998. [2] AlMalki et al., 2016. [3] Anstey-Gilbert et al., 2013. [4] Shah et al., 2015. [5] Mase et al., 2011. [6] Hosfield et al., 1998.

Table 7.2 presents a comparison of LiFEN with structures from a group of more diverse FEN-family members. Again, the structural model refined and presented this thesis compares well with these.

As it has been discussed before, the structure reported in this work of *Li*FEN compares quite well with human FEN1 and homologues protein structures deposited in the RCSB Protein Data Bank database. Furthermore, the experimental *Li*FEN structure from this project was compared with predicted structures of the same protein using the computational tools Phyre2 and AlphaFold 2. The predicted structures share a similar backbone conformation to the experimental *Li*FEN structure, and they give a RMSD score of 0.94 and 0.895 for Phyre2 and Alpha Fold 2, respectively (refer to Figure 5.20 in Section 5.9). As reasonably accurate predictions were obtained from both computational tools for the globular domain of the protein as later in the project determined by X-ray crystallography, this provides some confidence in the overall models used to determine potential inhibitors through *in silico* approaches.

7.3 Physical and virtual screening of small molecule inhibitors targeting microbial FENs

The two last objectives established in this project were to test whether selective inhibitors of pathogen FEN enzymes could be obtained and to build and evaluate *in silico* screening protocols to facilitate physical compound screening. During this work, three fragment libraries containing a total of more than 1500 molecules were screened through virtual computational tools using different models and algorithms. Four different models were set up in these experiments: microbial FENs with two magnesium ions in the active site, no metal ions and only one magnesium ion on each site. In addition, two different algorithms were used in the project: AutoDock 4 and AutoDock Vina. The libraries screened through virtual screening were BioNet and HitExpansion.

Different approaches were performed in this part of the project. In the case of the BioNet library, physical and virtual screening were carried out using all the molecules of the library. Afterwards, the results obtained from the physical and virtual screening

were analysed and determination of the best model for each protein was determined based on arbitrary cut-offs in the binding energy scores and inhibition percentages. With the HitExpansion library, a virtual screening was first assessed, and a subset of molecules were selected for physical screening (a subset of more than 100 molecules from the 400 molecules integrated in the library).

In order to build an *in silico* screening protocol of small molecules targeting microbial FENs, it is necessary to obtain the structural conformation of the proteins. During this project, some experimental and predicted structures of the microbial FENs were used to perform the virtual screening. The experimental structure used in this work was the *T. brucei* FEN previously elucidated in Sayers laboratory. Predicted structures obtained from Phyre2 of *L. infantum* and *C. neoformans* FENs were used for the *in silico* screening. Even though an experimental partial structure of *L. infantum* FEN was obtained and the high accuracy prediction of AlphaFold2 to determine protein structures, all these came late to be included in the project and the Phyre2 *Li*FEN and *Cn*FEN structures were used for the virtual screening. Nevertheless, for future work it has been considered to use a hybrid model of *Li*FEN using experimental data in the regions solved and predicted DNA binding loops for *in silico* experiments.

The analysis of the results obtained from the virtual screening and selecting virtual hits from the experiments is one of the most crucial steps for the identification of potential inhibitors, and this step lies on the researcher in an attempt to establish a valid cut-off in the process to select virtual hits. In this work, to evaluate the models performed and select potential hits from the virtual screening, an arbitrary binding energy cut-off of -7 kcal/mol was set up for the BioNet library with the AutoDock 4 algorithm and the HitExpansion library with both algorithms, while a binding energy of -5.5 kcal/mol was set up for the BioNet library with the AutoDock Vina algorithm.

Establishing an arbitrary binding energy cut-off of the results obtained from virtual screening varies depending on how restrictively or permissively the researcher would like to analyse the results and the available budget for the next stage which consists of performing physical assays from the virtual hits. Typically, our experience of purchasing individual molecules puts the cost in the range of $15 to $200 per compound with an average of ~$100, although this figure varies widely (Sayers,

personal communication). For example, the Drugs for Neglected Diseases initiative estimated an average cost of $300 per compound, whereas Pertusi and collaborators have estimated an average of $47 per molecule, both published in the same year (2017). Because of the high prices of purchasing single compounds, in the early stages of hit-identification it is common practice to use pre-existing collections (libraries), such as those used in this study. The cost per compound is significantly lower when obtained in library form. A recent library of 2726 molecules purchased for hit-identification had an average cost of $3.33 per compound (Sayers, personal communication, October 2022).

It is worth noting at this point that the binding energy results obtained from molecular docking protocols is merely theoretical. This is due to some factors that are not considered in the virtual docking such as the presence of water molecules and the different conformations that the protein fold in the environment, among other factors which are further discussed in the next paragraph. Therefore, the binding energies given by molecular docking protocols should be treated as providing a rank order rather than the Gibbs free energy values, even though there are often given in units of kcal/mol as it was seen in this work (the more negative the score, the strongest the predictive interaction). However, there does appear to be a correlation between predicted and actual Gibbs free energies according to several studies (Lee et al., 2012; Decherchi & Cavalli, 2020; Singh et al., 2021; Kumer et al., 2022).

One caveat of the docking work carried out in the project is that the algorithms only take into account the charges of the metal ions without considering the nature of the metal ion. As it has been mentioned before, magnesium ions are the preferred metal ions for *Li*FEN and *Tb*FEN, while manganese is the preferred metal ion for *Cn*FEN. Another major caveat in AutoDock 4 and AutoDock Vina algorithms is that water is not included, which in some instances might be important in inhibitor binding, such inhibitors would be missed by this process. But perhaps the biggest limitation with both of these algorithms is that the protein is considered as a rigid body. In AutoDock Vina there is the option to incorporate limited side chain flexibility, but in effect this is done by running additional dockings which greatly increases the computational effort required. All these factors will affect the results of the *in silico* screening and the accuracy of real hits to false-positive hit ratio.

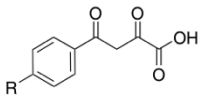
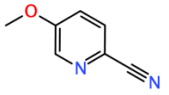
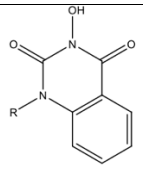
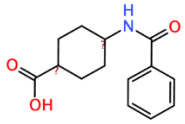
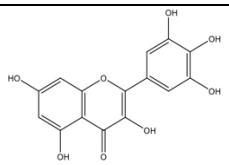
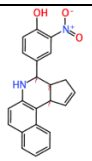
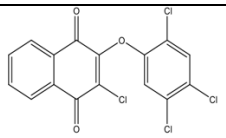
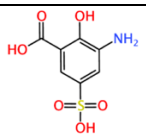
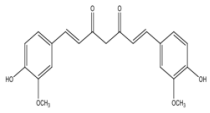
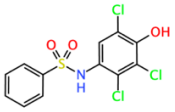
Inhibitors of the human FEN reported in the literature and potential inhibitors of microbial FENs reported in this work, with their respective structure and IC_{50} values are shown in Table 7.3. Inhibitors for the human FEN have been reported. Two studies by Tumey et al., (2004 & 2005) identified inhibitors from a series of chemical products derived from 2,4-diketobutyric acid and N-hydroxyurea compounds. Furthermore, natural products have also been identified as inhibitors such as myricetin and curcumin (Ma *et al.*, 2019; Cheng *et al.*, 2014). Additionally, an inhibitor of the human FEN was identified through virtual screening and machine learning approaches from a library of 53,000 compounds (Deshmukh *et al.*, 2017).

Most of the virtual hits obtained from the *in silico* screening and proved to be inhibitors in the FRET assays share a carboxylic acid group. Interestingly, some of the molecules which were not predicted by docking protocols but did possess inhibitory activity against one or more FEN proteins share a phenolic group. These molecules can be compared to the known inhibitors reported in the human FEN. For example, the molecules G1016 and G1032 reported in this study share the same carboxylic group as the chemical products obtained from 2,4-diketobutyric acid. At the same time, the molecules G1028, G1032 and G1035 from this work have a phenolic functional group as the reported compounds such as myricetin and curcumin.

Myricetin and curcumin belong to a group of compounds called pan-assay interference compounds (PAINS. Nelson *et al.*, 2017; Bisson *et al.*, 2016). This means that they are commonly identified in many programmes and due to their lack of selectivity they are unlikely to be developed further. Indeed, Oates reported that myricetin inhibited the *T. brucei* enzyme with a reported IC_{50} value of 15 μM (Oates, 2016), similar to the value reported in Table 7.3 for the human enzyme.

As we can see in Table 7.3, the compounds G1032 and G1035 inhibited FEN enzymes but not with any apparent selectivity over the human enzyme. In contrast, PS-6124, G1016 and G1028 showed discrimination between FENs of at least 13, 11, and 43-fold respectively.

Table 7.3 Reported inhibitors of human FEN and potential inhibitors of microbial FENs

Inhibitors reported in literature	Structure	IC_{50} value (μM)	Potential inhibitors from this work	Structure	IC_{50} value (μM)[#]
2,4-diketobutyric acid compounds[1]		0.19 *	PS-6124		297 392 75 >1000
N-hydroxy urea compounds[2]		0.01 *	G1016		149 >1000 89 >1000 %
Myricetin[3]		0.69	G1028		26 59 23 >1000 %
JFD00950[4]		5.5	G1032		51 119 41 103
Curcumin[5]		35	G1035		45 50 94 47

[1]*Tumey* et al., *2004;* [2]*Tumey* et al., *2005;* [3]*Ma* et al., *2019;* [4]*Deshmukh* et al., *2017;* [5]*Chen* et al., *2014.*

** For these studies, the molecule with the lowest IC_{50}* is shown in the table.

[#] *Reported IC_{50} values in this work for* Tb*FEN,* Li*FEN,* Cn*FEN and human FEN, respectively.*

% Reported IC_{50} values from Dr. Srdjan Vitovski (University of Sheffield, personal communication).

A more recent study provided the crystal structure of an N-hydroxy urea molecule bound to the human FEN (5FV7.pdb; Exell *et al.*, 2016). Based on the study performed by Tumey *et al.*, (2005), Exell and collaborators analysed four N-hydroxy urea products for activity inhibition and one of those compounds was co-crystallized with a truncated form of human FEN in presence of magnesium metal ions. According to the authors, crystal structure shows the inhibitor in a specific binding pose. This suggests a possible mode of action due to the presence of metal ions in the active site which the inhibitor requires in order to bind to the human FEN.

Docking the G1016, G1028 and G1032 molecules into the microbial FENs also showed that the compounds were predicted to bind to the same pair of metals as it is shown in Figure 7.1. They are predicted to bind via their carboxylate or NO_2^- groups. In the case of G1028 and G1032, both have a neighbouring phenolic hydroxyl group, which would also be substantially negatively charged and may explain why they are slightly more potent than G1016. The docking models do not explain why PS-6124 inhibits as well as it does. It has no acidic oxyanion group and is rather small. It may bind to a small pocket elsewhere in the protein or form a ternary complex with the DNA substrate or product thus stopping the reaction. In the case of G1035, none of the 10 poses provides information that the molecule binds to the metals (Figure 7.1).

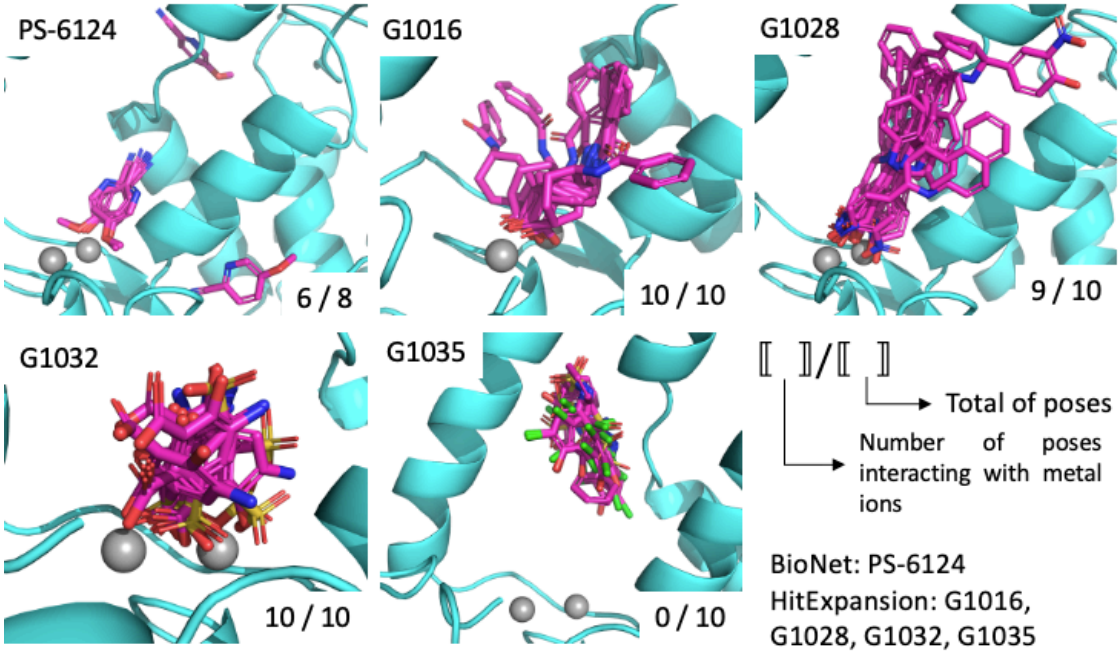


Figure 7.1 Molecules with IC_{50} values >100 μM docked in *Tb*FEN structure

A total of 5 molecules from BioNet and HitExpansion libraries, which registered IC_{50} values lower than 100 μM for at least one microbial enzyme (according to Table 7.3). The molecules were docked against the Tb*FEN and the best poses are visualized using Pymol. The mean binding energy scores for the 5 molecules PS-6124, G1016, G1028, G1032 and G1035 are -3.4, -9.7, -8.4, -8.9 and -5.4 kcal/mol, respectively.*

7.4 Future Work

7.4.1 Structural Biology

The lack of a good co-crystal structure with DNA or of a complete FEN structure could be addressed by trying different DNA substrates or NMR respectively. These substrates could be synthesised with different duplex lengths upstream and downstream of the flap junction and also using different single-stranded flap lengths and sequences in order to try to stabilize crystal contacts to produce better diffraction. Co-crystallisation trials could be set up in the presence of inhibitors (with or without DNA) as these could adopt conformations of the protein which may be more likely to crystallise. NMR is suitable for determining structures of proteins in solution of the size of microbial FENs, but is less suitable for larger complexes.

7.4.2 Characterisation of Inhibitors

The main objective of this work is to identify molecules that can kill the parasite. Therefore, whole-cell assays *on Trypanosoma, Leishmania* and *Cryptococccus* should be carried to see if any of them possess biological activity. A simple assay which is useful to screen out membrane-active compounds can be carried out using haemolysis assays with erythrocytes. Indeed, these were planned but time ran out.

Work needs to be done to check whether the molecules identified interact directly with the protein, protein-substrate or protein-product complexes or only with DNA. The latter type of inhibitor includes DNA intercalators such as ethidium bromide which is part of some commercially available libraries. These molecules would not be suitable for development for obvious reasons. Available techniques to characterise molecular interactions include isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), thermal shift assays (TSA) a variety of nuclear magnetic resonance (NMR) techniques amongst others (Gossert, 2019).

7.4.3 Hit-to-lead Generation

The molecules identified in this study could be used to take the project forward by any combination of hit expansion or medicinal chemistry to obtain molecules which may show improved properties (potency, selectivity over human FEN, and later ADMET

studies). The hit expansion approach could be used by asking the question are they more molecules available with similar arrangement of atoms. For example, the structure of molecules PS-6124, G1016, G1028 and G1032 could be entered into a database like Zinc20 (Irwin *et al.*, 2020). This database contains over 230 million commercially available compounds with links to the suppliers. Similar compounds found on the database can be retrieved in sdf format, which is useful for future molecular docking. Indeed, the in-house Hit Expansion library used in this study was derived by the Sayers laboratory on the basis of other FEN proteins previously studied in this laboratory.

Another approach could use the structures of the initial hits to guide rational design of new molecules as shown in the example in Figure 7.2. One of the major drawbacks of this approach is that, by designing improved candidate molecules, it is possible that it will not be available for purchasing and will require expensive and time-consuming medicinal chemistry.

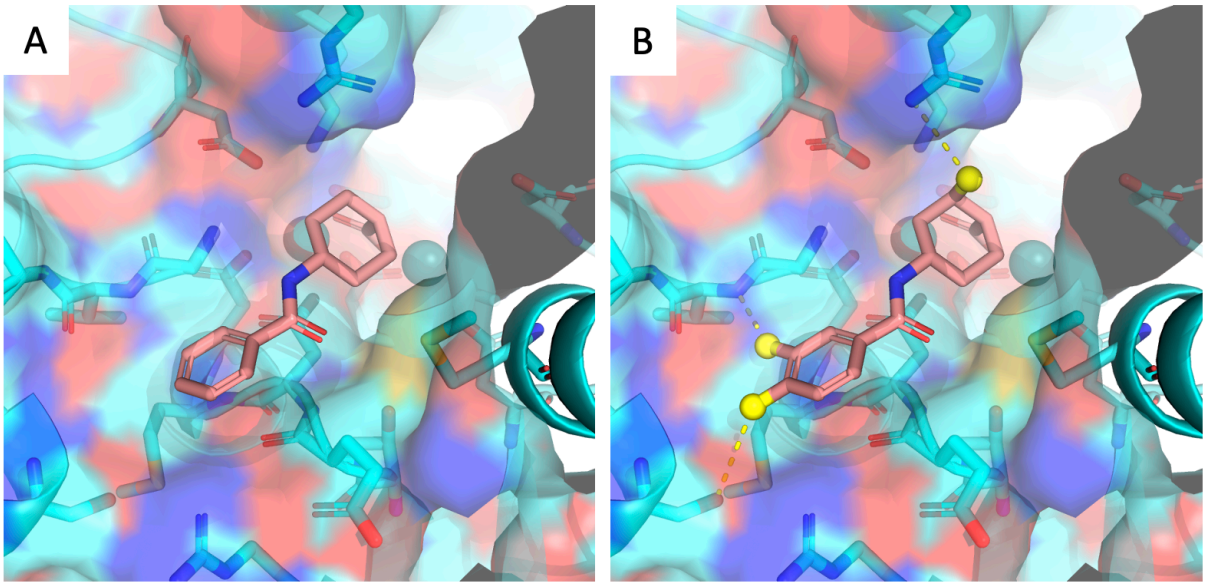


Figure 7.2 Future approach in LBDD to synthesize a molecule with a higher inhibition activity from G1016

(A) *shows the molecule G1016 docked into the* Tb*FEN protein. The protein is shown as cartoon and transparent surface (cyan carbons), the molecule G1016 (pink carbons) and the residues neighbouring G1016 within 5 Å are presented as sticks.* ***(B)*** *shows G1016 with possible additions of oxygen atoms (yellow spheres) to illustrate the principal of ligand-based drug design. Introduction of any of these atoms could increase interaction through potential hydrogen bonds (yellow dotted lines).*

Other potential approaches could include exploring the single-hit assays developed in Chapter 6 to screen much larger libraries such as the 850,000 compound library used by AstraZeneca to identify potential hits via high-throughput screening approach using fluorescent substrates in a polarization assay (McWhirter *et al.*, 2013). Similarly, the virtual screening approach can be further expanded by analysing larger virtual libraries. A recent review (Bender *et al.*, 2021) gave examples of screening campaigns that examined up to a billion compounds. However, this approach needs access to high-performance computational power.

7.4.4 Mode of Action?

Even if we showed that FEN inhibitors kill parasites, this would not prove that the mode of action is by specifically targeting the flap endonuclease. Validation of target engagement *in vitro* is extremely important and can be carried out by a number of methods including cellular thermal shift assay (CETSA) and drug affinity responsive target stability (DARTS). Drug-target validation can also be carried out by genetic knock-downs or target over-expression. In the case of knock-down constructs, these are likely to be more sensitive to the inhibitor whereas over-expression of the target protein will make the cells less sensitive to the molecule. Another approach could be to culture the organism, in gradually increasing concentrations of the inhibitor starting from a sub-lethal dose. If resistance organisms arise, the genome can be sequenced and the mechanism can be determined (Cooper *et al.,* 2020; Chen *et al.,* 2019).

7.4.5 Final Conclusion

The work carried out has resulted in the production of a 1.4 Å-resolution crystal structure of the *Leishmania* flap endonuclease. There is no crystal structure of a eukaryotic pathogen in the current Protein Data Bank (November, 2022). Indeed, the only eukaryotic FEN structure in the database is the human FEN1. This structure validated theoretical models predicted by Phyre2 and AlphaFold 2.

In silico screening approaches were performed in order to identify inhibitors. No one model proved to be successful in identifying real hits. Clearly, more structure information and more sophisticated docking algorithms such as GOLD or Glide are

necessary to further determine the optimal *in silico* algorithm to identify inhibitors (Friesner *et al.*, 2004; Halgren *et al.*, 2004; Jones *et al.*, 1997; Verdonk *et al.*, 2003).

Physical FRET assays and virtual screening were performed to identify potential inhibitors targeting *Trypanosoma, Leishmania* and *Cryptococcus* FEN. Furthermore, data presented from these assays in Chapter 6 suggest that it is possible to selectively inhibit different FENs. The compounds PS-6124, G1016 and G1028 showed discrimination between FENs of at least 13, 11, and 43-fold respectively. Specifically, in terms of selectivity over the human FEN, compound G1028 has IC_{50} values of 23 μM and over 1000 μM against *Cn*FEN and human FEN, respectively, representing selectivity greater than 43-fold. Compound PS-6124 has IC_{50} values of 75 μM and over 1000 μM against *Cn*FEN and human FEN, respectively, representing selectivity greater than 13-fold, which is quite interesting given the small size of the molecule (10 non-hydrogen atoms). Compound G1016 has IC_{50} values of 89 μM and over 1000 μM against *Cn*FEN and human FEN, respectively, representing selectivity greater than 11-fold.

In conclusion, this feasibility study suggests that targeting microbial flap endonucleases may be a worthwhile approach to eventually develop new drugs against neglected diseases caused by *Trypanosoma brucei, Leishmania infantum* and *Cryptococcus neoformans*.

CHAPTER 8: REFERENCES

1. Aguilar, A., Lu, J., Liu, L., Du, D., Bernard, D., McEachern, D., Przybranowski, S., Li, X., Luo, R., Wen, B., Sun, D., Wang, H., Wen, J., Wang, G., Zhai, Y., Guo, M., Yang, D., & Wang, S. (2017). Discovery of 4-((3'R,4'S,5'R)-6''-Chloro-4'-(3-chloro-2-fluorophenyl)-1'-ethyl-2''-oxodispiro[cyclohexane-1,2'-pyrrolidine-3',3''-indoline]-5'-carboxamido)bicyclo[2.2.2]octane -1-carboxylic Acid (AA-115/APG-115): A Potent and Orally Active Murine Double Minute 2 (MDM2) Inhibitor in Clinical Development. Journal of medicinal chemistry, 60(7), 2819-2839. https://doi.org/10.1021/acs.jmedchem.6b01665
2. Alberts, B. (1989). Molecular biology of the cell. Second edition. New York : Garland Pub., [1989] ©1989. https://search.library.wisc.edu/catalog/999598982402121
3. Algasaier, S. I., Exell, J. C., Bennet, I. A., Thompson, M. J., Gotham, V. J., Shaw, S. J., Craggs, T. D., Finger, L. D., & Grasby, J. A. (2016). DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-catalyzed Reaction. The Journal of biological chemistry, 291(15), 8258-8268. https://doi.org/10.1074/jbc.M115.698993
4. Allen, L. M., Hodskinson, M. R. G., & Sayers, J. R. (2009). Active site substitutions delineate distinct classes of eubacterial flap endonuclease. Biochem J, 418(2), 285-292. https://doi.org/10.1042/BJ20081637
5. AlMalki, F. A., Flemming, C. S., Zhang, J., Feng, M., Sedelnikova, S. E., Ceska, T., Rafferty, J. B., Sayers, J. R., & Artymiuk, P. J. (2016). Direct observation of DNA threading in flap endonuclease complexes. Nature Structural &Amp; Molecular Biology, 23, 640. https://doi.org/10.1038/nsmb.3241
6. Almeida, F., Wolf, J. M., & Casadevall, A. (2015). Virulence-Associated Enzymes of Cryptococcus neoformans. Eukaryot Cell, 14(12), 1173-1185. https://doi.org/10.1128/EC.00103-15
7. Alpers, D. H., & Tomkins, G. M. (1965). The Order of Induction and Deinduction of the Enzymes of the Lactose Operon in E. Coli. Proceedings of the National Academy of Sciences of the United States of America, 53, 797-802. https://doi.org/10.1073/pnas.53.4.797

8. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol, 215(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2
9. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389
10. Anderson, A. C. (2003). The Process of Structure-Based Drug Design. Chemistry & Biology, 10(9), 787-797. https://doi.org/https://doi.org/10.1016/j.chembiol.2003.09.002
11. Anstey-Gilbert, C. S., Hemsworth, G. R., Flemming, C. S., Hodskinson, M. R. G., Zhang, J., Sedelnikova, S. E., Stillman, T. J., Sayers, J. R., & Artymiuk, P. J. (2013). The structure of Escherichia coli ExoIX—implications for DNA binding and catalysis in flap endonucleases. Nucleic Acids Research, 41(17), 8357-8367. https://doi.org/10.1093/nar/gkt591
12. Aoyama, Y., & Yoshida, Y. (1978). The 14alpha-demethylation of lanosterol by a reconstituted cytochrome P-450 system from yeast microsomes. Biochem Biophys Res Commun, 85(1), 28-34. https://doi.org/10.1016/s0006-291x(78)80006-0
13. Aparicio, I. M., Scharfstein, J., & Lima, A. P. C. A. (2004). A New Cruzipain-Mediated Pathway of Human Cell Invasion by Trypanosoma cruzi Requires Trypomastigote Membranes. Infection and Immunity, 72(10), 5892-5902. https://doi.org/10.1128/iai.72.10.5892-5902.2004
14. Baginski, M., Sternal, K., Czub, J., & Borowski, E. (2005). Molecular modelling of membrane activity of amphotericin B, a polyene macrolide antifungal antibiotic. Acta Biochim Pol, 52(3), 655-658. https://www.ncbi.nlm.nih.gov/pubmed/16086075
15. Baig, F., Fernando, L. P., Salazar, M. A., Powell, R. R., Bruce, T. F., & Harcum, S. W. (2014). Dynamic transcriptional response of Escherichia coli to inclusion body formation. Biotechnol Bioeng, 111(5), 980-999. https://doi.org/10.1002/bit.25169
16. Bajad, N. G., Rayala, S., Gutti, G., Sharma, A., Singh, M., Kumar, A., & Singh, S. K. (2021). Systematic review on role of structure based drug design (SBDD) in the identification of anti-viral leads against SARS-Cov-2. Curr Res Pharmacol Drug Discov, 2, 100026. https://doi.org/10.1016/j.crphar.2021.100026

17. Baker, N., Glover, L., Munday, J. C., Aguinaga Andres, D., Barrett, M. P., de Koning, H. P., & Horn, D. (2012). Aquaglyceroporin 2 controls susceptibility to melarsoprol and pentamidine in African trypanosomes. Proceedings of the National Academy of Sciences of the United States of America, 109(27), 10996-11001. https://doi.org/10.1073/pnas.1202885109
18. Bakker, B. M., Mensonides, F. I., Teusink, B., van Hoek, P., Michels, P. A., & Westerhoff, H. V. (2000). Compartmentation protects trypanosomes from the dangerous design of glycolysis. Proceedings of the National Academy of Sciences of the United States of America, 97(5), 2087-2092. https://doi.org/10.1073/pnas.030539197
19. Balakrishnan, L., Bambara, R. A. (2013). Flap endonuclease 1. Annual Review of Biochemistry, 82, 119-138. https://doi.org/10.1146/annurev-biochem-072511-122603
20. Bambara, R. A., Murante, R. S., & Henricksen, L. A. (1997). Enzymes and Reactions at the Eukaryotic DNA Replication Fork. Journal of Biological Chemistry, 272(8), 4647-4650. https://doi.org/10.1074/jbc.272.8.4647
21. Barakat, K. (2014). Computer-Aided Drug Design. Journal of Pharmaceutical Care & Health Systems, 01. https://doi.org/10.4172/2376-0419.1000e113
22. Barral, A., Barral-Netto, M., Yong, E. C., Brownell, C. E., Twardzik, D. R., & Reed, S. G. (1993). Transforming growth factor beta as a virulence mechanism for Leishmania braziliensis. Proceedings of the National Academy of Sciences of the United States of America, 90(8), 3442-3446. https://doi.org/10.1073/pnas.90.8.3442
23. Barrenho, E., Miraldo, M., & Smith, P. C. (2019). Does global drug innovation correspond to burden of disease? The neglected diseases in developed and developing countries. Health Econ, 28(1), 123-143. https://doi.org/10.1002/hec.3833
24. Barrett, M. P., Boykin, D. W., Brun, R., & Tidwell, R. R. (2007). Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. British journal of pharmacology, 152(8), 1155-1171. https://doi.org/10.1038/sj.bjp.0707354

25. Batool, M., Ahmad, B., & Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. International Journal of Molecular Sciences, 20(11), 2783. https://www.mdpi.com/1422-0067/20/11/2783
26. Bayliss, C. D., Sweetman, W. A., & Moxon, E. R. (2005). Destabilization of tetranucleotide repeats in Haemophilus influenzae mutants lacking RnaseHI or the Klenow domain of PolI. Nucleic Acids Res, 33(1), 400-408. https://doi.org/10.1093/nar/gki180
27. Becker, J. R., Pons, C., Nguyen, H. D., Costanzo, M., Boone, C., Myers, C. L., & Bielinsky, A. K. (2015). Genetic Interactions Implicating Postreplicative Repair in Okazaki Fragment Processing. PLoS Genet, 11(11), e1005659. https://doi.org/10.1371/journal.pgen.1005659
28. Bekker-Jensen, S., Lukas, C., Kitagawa, R., Melander, F., Kastan, M. B., Bartek, J., & Lukas, J. (2006). Spatial organization of the mammalian genome surveillance machinery in response to DNA strand breaks. J Cell Biol, 173(2), 195-206. https://doi.org/10.1083/jcb.200510130
29. Bello, A. R., Nare, B., Freedman, D., Hardy, L., & Beverley, S. M. (1994). PTR1: a reductase mediating salvage of oxidized pteridines and methotrexate resistance in the protozoan parasite Leishmania major. Proceedings of the National Academy of Sciences of the United States of America, 91(24), 11442-11446. https://doi.org/10.1073/pnas.91.24.11442
30. Ben Salah, A., Ben Messaoud, N., Guedri, E., Zaatour, A., Ben Alaya, N., Bettaieb, J., Gharbi, A., Belhadj Hamida, N., Boukthir, A., Chlif, S., Abdelhamid, K., El Ahmadi, Z., Louzir, H., Mokni, M., Morizot, G., Buffet, P., Smith, P. L., Kopydlowski, K. M., Kreishman-Deitrick, M., . . . Grogl, M. (2013). Topical Paromomycin with or without Gentamicin for Cutaneous Leishmaniasis. New England Journal of Medicine, 368(6), 524-532. https://doi.org/10.1056/NEJMoa1202657
31. Benne, R., Van Den Burg, J., Brakenhoff, J. P. J., Sloof, P., Van Boom, J. H., & Tromp, M. C. (1986). Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell, 46(6), 819-826. https://doi.org/10.1016/0092-8674(86)90063-2
32. Bennet, I. A., Finger, L. D., Baxter, N. J., Ambrose, B., Hounslow, A. M., Thompson, M. J., Exell, J. C., Shahari, N., Craggs, T. D., Waltho, J. P., & Grasby, J. A. (2018). Regional conformational flexibility couples substrate specificity and scissile

phosphate diester selectivity in human flap endonuclease 1. Nucleic Acids Res, 46(11), 5618-5633. https://doi.org/10.1093/nar/gky293

33. Berasain, P., Carmona, C., Frangione, B., Cazzulo, J. J., & Goñi, F. (2003). Specific cleavage sites on human IgG subclasses by cruzipain, the major cysteine proteinase from Trypanosoma cruzi. Molecular and Biochemical Parasitology, 130(1), 23-29. https://doi.org/https://doi.org/10.1016/S0166-6851(03)00139-7
34. Bergström, F., & Lindmark, B. (2019). Accelerated drug discovery by rapid candidate drug identification. Drug Discov Today, 24(6), 1237-1241. https://doi.org/10.1016/j.drudis.2019.03.026
35. Bethony, J. M., Cole, R. N., Guo, X., Kamhawi, S., Lightowlers, M. W., Loukas, A., Petri, W., Reed, S., Valenzuela, J. G., & Hotez, P. J. (2011). Vaccines to combat the neglected tropical diseases. Immunol Rev, 239(1), 237-270. https://doi.org/10.1111/j.1600-065X.2010.00976.x
36. Bhowmick, T. (2021). Clinical Outcomes of Patient Subgroups in the TANGO II Study. Infect Dis Ther, 10(1), 35-46. https://doi.org/10.1007/s40121-021-00405-x
37. Bian, S., Ni, W., Zhu, M., Zhang, X., Qiang, Y., Zhang, J., Ni, Z., Shen, Y., Qiu, S., Song, Q., Xiao, M., & Zheng, W. (2022). Flap endonuclease 1 Facilitated Hepatocellular Carcinoma Progression by Enhancing USP7/MDM2-mediated P53 Inactivation. Int J Biol Sci, 18(3), 1022-1038. https://doi.org/10.7150/ijbs.68179
38. Biggins, J. B., Prudent, J. R., Marshall, D. J., Ruppen, M., & Thorson, J. S. (2000). A continuous assay for DNA cleavage: the application of "break lights" to enediynes, iron-dependent agents, and nucleases. Proceedings of the National Academy of Sciences of the United States of America, 97(25), 13537-13542. https://doi.org/10.1073/pnas.240460997
39. Bina-Stein, M., & Tritton, T. R. (1976). Aurintricarboxylic acid is a nonspecific enzyme inhibitor. Mol Pharmacol, 12(1), 191-193. https://www.ncbi.nlm.nih.gov/pubmed/1256444
40. Biter, A. B., Weltje, S., Hudspeth, E. M., Seid, C. A., McAtee, C. P., Chen, W. H., Pollet, J. B., Strych, U., Hotez, P. J., & Bottazzi, M. E. (2018). Characterization and Stability of Trypanosoma cruzi 24-C4 (Tc24-C4), a Candidate Antigen for a Therapeutic Vaccine Against Chagas Disease. J Pharm Sci, 107(5), 1468-1473. https://doi.org/10.1016/j.xphs.2017.12.014

41. Blaney, J. (2012). A very short history of structure-based design: how did we get here and where do we need to go? J Comput Aided Mol Des, 26(1), 13-14. https://doi.org/10.1007/s10822-011-9518-x
42. Blaza, J. N., Bridges, H. R., Aragão, D., Dunn, E. A., Heikal, A., Cook, G. M., Nakatani, Y., & Hirst, J. (2017). The mechanism of catalysis by type-II NADH:quinone oxidoreductases [Article]. Scientific Reports, 7, 40165. https://doi.org/10.1038/srep40165. https://www.nature.com/articles/srep40165#supplementary-information
43. Boggs, J. (2017). Characterizing the role of DNA polymerase gamma in the yeast Cryptococcus Neoformans Western Carolina University]. https://libres.uncg.edu/ir/wcu/listing.aspx?id=22715
44. Bokhari, F. F., & Albukhari, A. (2021). Design and Implementation of High Throughput Screening Assays for Drug Discoveries. In K. S. Shailendra (Ed.), High-Throughput Screening for Drug Discovery (pp. Ch. 4). IntechOpen. https://doi.org/10.5772/intechopen.98733
45. Bonaldo, M. C., d'Escoffier, L. N., Salles, J. M., & Goldenberg, S. (1991). Characterization and expression of proteases during Trypanosoma cruzi metacyclogenesis. Experimental Parasitology, 73(1), 44-51. https://doi.org/https://doi.org/10.1016/0014-4894(91)90006-I
46. Bonardi, A., Vermelho, A. B., da Silva Cardoso, V., de Souza Pereira, M. C., da Silva Lara, L., Selleri, S., Gratteri, P., Supuran, C. T., & Nocentini, A. (2019). N-Nitrosulfonamides as Carbonic Anhydrase Inhibitors: A Promising Chemotype for Targeting Chagas Disease and Leishmaniasis. ACS Medicinal Chemistry Letters, 10(4), 413-418. https://doi.org/10.1021/acsmedchemlett.8b00430
47. Boucher, H. W., Groll, A. H., Chiou, C. C., & Walsh, T. J. (2004). Newer systemic antifungal agents : pharmacokinetics, safety and efficacy. Drugs, 64(18), 1997-2020. https://doi.org/10.2165/00003495-200464180-00001
48. Boucinha, C., Caetano, A. R., Santos, H. L., Helaers, R., Vikkula, M., Branquinha, M. H., Dos Santos, A. L. S., Grellier, P., Morelli, K. A., & d'Avila-Levy, C. M. (2020). Analysing ambiguities in trypanosomatids taxonomy by barcoding. Mem Inst Oswaldo Cruz, 115, e200504. https://doi.org/10.1590/0074-02760200504

49. Boyle, P. M., & Silver, P. A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab Eng, 14(3), 223-232. https://doi.org/10.1016/j.ymben.2011.10.003
50. Brand, S., Norcross, N. R., Thompson, S., Harrison, J. R., Smith, V. C., Robinson, D. A., Torrie, L. S., McElroy, S. P., Hallyburton, I., Norval, S., Scullion, P., Stojanovski, L., Simeons, F. R. C., van Aalten, D., Frearson, J. A., Brenk, R., Fairlamb, A. H., Ferguson, M. A. J., Wyatt, P. G., . . . Read, K. D. (2014). Lead optimization of a pyrazole sulfonamide series of Trypanosoma brucei N-myristoyltransferase inhibitors: identification and evaluation of CNS penetrant compounds as potential treatments for stage 2 human African trypanosomiasis. Journal of medicinal chemistry, 57(23), 9855-9869. https://doi.org/10.1021/jm500809c
51. Brannigan, J. A., Smith, B. A., Yu, Z., Brzozowski, A. M., Hodgkinson, M. R., Maroof, A., Price, H. P., Meier, F., Leatherbarrow, R. J., Tate, E. W., Smith, D. F., & Wilkinson, A. J. (2010). N-Myristoyltransferase from Leishmania donovani: Structural and Functional Characterisation of a Potential Drug Target for Visceral Leishmaniasis. Journal of Molecular Biology, 396(4), 985-999. https://doi.org/https://doi.org/10.1016/j.jmb.2009.12.032
52. Branzei, D., & Foiani, M. (2005). The DNA damage response during DNA replication. Current Opinion in Cell Biology, 17(6), 568-575. https://doi.org/https://doi.org/10.1016/j.ceb.2005.09.003
53. Bratton, E. W., El Husseini, N., Chastain, C. A., Lee, M. S., Poole, C., Sturmer, T., Juliano, J. J., Weber, D. J., & Perfect, J. R. (2012). Comparison and temporal trends of three groups with cryptococcosis: HIV-infected, solid organ transplant, and HIV-negative/non-transplant. PLOS ONE, 7(8), e43582. https://doi.org/10.1371/journal.pone.0043582
54. Bressi, J. C., Verlinde, C. L. M. J., Aronov, A. M., Shaw, M. L., Shin, S. S., Nguyen, L. N., Suresh, S., Buckner, F. S., Van Voorhis, W. C., Kuntz, I. D., Hol, W. G. J., & Gelb, M. H. (2001). Adenosine Analogues as Selective Inhibitors of Glyceraldehyde-3-phosphate Dehydrogenase of Trypanosomatidae via Structure-Based Drug Design. Journal of medicinal chemistry, 44(13), 2080-2093. https://doi.org/10.1021/jm000472o

55. Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., Seigneuric, R., & Neiers, F. (2016). A self-inducible heterologous protein expression system in Escherichia coli. Scientific Reports, 6(1), 33037. https://doi.org/10.1038/srep33037
56. Brook, I. (1987). Approval of Zidovudine (AZT) for Acquired Immunodeficiency Syndrome: A Challenge to the Medical and Pharmaceutical Communities. JAMA, 258(11), 1517-1517. https://doi.org/10.1001/jama.1987.03400110099035
57. Brown, D. G., & Bostrom, J. (2018). Where Do Recent Small Molecule Clinical Development Candidates Come From? Journal of medicinal chemistry, 61(21), 9442-9468. https://doi.org/10.1021/acs.jmedchem.8b00675
58. Bruckner, D. A., & Labarca, J. A. (2015). Leishmania and Trypanosoma. In Manual of Clinical Microbiology (11 ed., pp. 2357-2368). ASM Press.
59. Brylinski, M., & Skolnick, J. (2009). FINDSITE: a threading-based approach to ligand homology modeling. PLoS Comput Biol, 5(6), e1000405. https://doi.org/10.1371/journal.pcbi.1000405
60. Burgers, P. M. J. (2009). Polymerase dynamics at the eukaryotic DNA replication fork. The Journal of biological chemistry, 284(7), 4041-4045. https://doi.org/10.1074/jbc.R800062200
61. Burgers, P. M. J., & Kunkel, T. A. (2017). Eukaryotic DNA Replication Fork. Annual Review of Biochemistry, 86(1), 417-438. https://doi.org/10.1146/annurev-biochem-061516-044709
62. Butts, A., Koselny, K., Chabrier-Rosello, Y., Semighini, C. P., Brown, J. C., Wang, X., Annadurai, S., DiDone, L., Tabroff, J., Childers, W. E., Jr., Abou-Gharbia, M., Wellington, M., Cardenas, M. E., Madhani, H. D., Heitman, J., & Krysan, D. J. (2014). Estrogen receptor antagonists are anti-cryptococcal agents that directly bind EF hand proteins and synergize with fluconazole in vivo. mBio, 5(1), e00765-00713. https://doi.org/10.1128/mBio.00765-13
63. Calderano, S. G., Godoy, P. D. d. M., da Cunha, J. P. C., & Elias, M. C. (2011). Trypanosome Prereplication Machinery: A Potential New Target for an Old Problem. Enzyme Research, 2011. https://doi.org/10.4061/2011/518258
64. Callens, M., & Hannaert, V. (1995). The rational design of trypanocidal drugs: selective inhibition of the glyceraldehyde-3-phosphate dehydrogenase in Trypanosomatidae. Annals of Tropical Medicine & Parasitology, 89(sup1), 23-30. https://doi.org/10.1080/00034983.1995.11813011

65. Cameron, P., McGachy, A., Anderson, M., Paul, A., Coombs, G. H., Mottram, J. C., Alexander, J., & Plevin, R. (2004). Inhibition of Lipopolysaccharide-Induced Macrophage IL-12 Production by Leishmania mexicana Amastigotes: The Role of Cysteine Peptidases and the NF-κB Signaling Pathway. The Journal of Immunology, 173(5), 3297-3304. https://doi.org/10.4049/jimmunol.173.5.3297
66. Capasso, C., & Supuran, C. T. (2015). Bacterial, fungal and protozoan carbonic anhydrases as drug targets. Expert Opinion on Therapeutic Targets, 19(12), 1689-1704. https://doi.org/10.1517/14728222.2015.1067685
67. Casadevall, A., Cleare, W., Feldmesser, M., Glatman-Freedman, A., Goldman, D. L., Kozel, T. R., Lendvai, N., Mukherjee, J., Pirofski, L. A., Rivera, J., Rosas, A. L., Scharff, M. D., Valadon, P., Westin, K., & Zhong, Z. (1998). Characterization of a murine monoclonal antibody to Cryptococcus neoformans polysaccharide that is a candidate for human therapeutic studies. Antimicrobial agents and chemotherapy, 42(6), 1437-1446. https://doi.org/10.1128/AAC.42.6.1437
68. Casgrain, P.-A., Martel, C., McMaster, W. R., Mottram, J. C., Olivier, M., & Descoteaux, A. (2016). Cysteine Peptidase B Regulates Leishmania mexicana Virulence through the Modulation of GP63 Expression. PLoS pathogens, 12(5), e1005658-e1005658. https://doi.org/10.1371/journal.ppat.1005658
69. Centers for Disease and Control and Prevention (CDC). (2020). Parasites - African Trypanosomiasis (also known as Sleeping Sickness). In https://www.cdc.gov/parasites/sleepingsickness/biology.html
70. Ceska, T. A., Sayers, J. R., Stier, G., & Suck, D. (1996). A helical arch allowing single-stranded DNA to thread through T5 5'-exonuclease. Nature, 382(6586), 90-93. https://doi.org/10.1038/382090a0
71. Champoux, J. J. (2001). DNA Topoisomerases: Structure, Function, and Mechanism. Annual Review of Biochemistry, 70(1), 369-413. https://doi.org/10.1146/annurev.biochem.70.1.369
72. Chang, W. C., Tzao, C., Hsu, H. H., Lee, S. C., Huang, K. L., Tung, H. J., & Chen, C. Y. (2006). Pulmonary cryptococcosis: comparison of clinical and radiographic characteristics in immunocompetent and immunocompromised patients. Chest, 129(2), 333-340. https://doi.org/10.1378/chest.129.2.333
73. Chatelain, E. (2015). Chagas disease drug discovery: toward a new era. J Biomol Screen, 20(1), 22-35. https://doi.org/10.1177/1087057114550585

74. Chen, B., Zhang, Y., Wang, Y., Rao, J., Jiang, X., & Xu, Z. (2014). Curcumin inhibits proliferation of breast cancer cells through Nrf2-mediated down-regulation of Fen1 expression. The Journal of Steroid Biochemistry and Molecular Biology, 143, 11-18. https://doi.org/10.1016/j.jsbmb.2014.01.009
75. Chen, J., Varma, A., Diaz, M. R., Litvintseva, A. P., Wollenberg, K. K., & Kwon-Chung, K. J. (2008). Cryptococcus neoformans strains and infection in apparently immunocompetent patients, China. Emerg Infect Dis, 14(5), 755-762. https://doi.org/10.3201/eid1405.071312
76. Chen, S. C., Meyer, W., & Sorrell, T. C. (2014). Cryptococcus gattii infections. Clin Microbiol Rev, 27(4), 980-1024. https://doi.org/10.1128/CMR.00126-13
77. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr, 66(Pt 1), 12-21. https://doi.org/10.1107/S0907444909042073
78. Chen, X., He, G., Wang, S., Lin, S., Chen, J., & Zhang, W. (2019). Evaluation of Whole-Genome Sequence Method to Diagnose Resistance of 13 Anti-tuberculosis Drugs and Characterize Resistance Genes in Clinical Multi-Drug Resistance Mycobacterium tuberculosis Isolates From China. Front Microbiol, 10, 1741. https://doi.org/10.3389/fmicb.2019.01741
79. Chen, Y. C., Chang, S. C., Shih, C. C., Hung, C. C., Luhbd, K. T., Pan, Y. S., & Hsieh, W. C. (2000). Clinical features and in vitro susceptibilities of two varieties of Cryptococcus neoformans in Taiwan. Diagn Microbiol Infect Dis, 36(3), 175-183. https://doi.org/10.1016/s0732-8893(99)00137-6
80. Cheng, L., Pettersen, D., Ohlsson, B., Schell, P., Karle, M., Evertsson, E., Pahlen, S., Jonforsen, M., Plowright, A. T., Bostrom, J., Fex, T., Thelin, A., Hilgendorf, C., Xue, Y., Wahlund, G., Lindberg, W., Larsson, L. O., & Gustafsson, D. (2014). Discovery of the Fibrinolysis Inhibitor AZD6564, Acting via Interference of a Protein-Protein Interaction. ACS Med Chem Lett, 5(5), 538-543. https://doi.org/10.1021/ml400526d
81. Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: a problem-centric review. The AAPS journal, 14(1), 133-141. https://doi.org/10.1208/s12248-012-9322-0

82. Clark, R. A., Greer, D., Atkinson, W., Valainis, G. T., & Hyslop, N. (1990). Spectrum of Cryptococcus neoformans infection in 68 patients infected with human immunodeficiency virus. Rev Infect Dis, 12(5), 768-777. https://doi.org/10.1093/clinids/12.5.768
83. Clayton, C., Häusler, T., & Blattner, J. (1995). Protein trafficking in kinetoplastid protozoa. Microbiological Reviews, 59(3), 325. http://mmbr.asm.org/content/59/3/325.abstract
84. Cohen, J. P., Silva, L., Cohen, A., Awatin, J., & Sturgeon, R. (2016). Progress Report on Neglected Tropical Disease Drug Donation Programs. Clin Ther, 38(5), 1193-1204. https://doi.org/10.1016/j.clinthera.2016.02.031
85. Congreve, M., Carr, R., Murray, C., & Jhoti, H. (2003). A 'rule of three' for fragment-based lead discovery? Drug Discov Today, 8(19), 876-877. https://doi.org/10.1016/s1359-6446(03)02831-9
86. Cooper, A. L., Low, A. J., Koziol, A. G., Thomas, M. C., Leclair, D., Tamber, S., Wong, A., Blais, B. W., & Carrillo, C. D. (2020). Systematic Evaluation of Whole Genome Sequence-Based Predictions of Salmonella Serotype and Antimicrobial Resistance. Front Microbiol, 11, 549. https://doi.org/10.3389/fmicb.2020.00549
87. Cooper, C. D. O., & Marsden, B. D. (2017). N- and C-Terminal Truncations to Enhance Protein Solubility and Crystallization: Predicting Protein Domain Boundaries with Bioinformatics Tools. Methods Mol Biol, 1586, 11-31. https://doi.org/10.1007/978-1-4939-6887-9_2
88. Cordeiro, R. A., Evangelista, A. J. J., Serpa, R., Marques, F. J. F., Melo, C. V. S., Oliveira, J. S., Franco, J. D. S., Alencar, L. P., Bandeira, T., Brilhante, R. S. N., Sidrim, J. J. C., & Rocha, M. F. G. (2016). Inhibition of heat-shock protein 90 enhances the susceptibility to antifungals and reduces the virulence of Cryptococcus neoformans/Cryptococcus gattii species complex. Microbiology (Reading), 162(2), 309-317. https://doi.org/10.1099/mic.0.000222
89. Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res, 16(22), 10881-10890. https://doi.org/10.1093/nar/16.22.10881
90. Correddu, D., Montano Lopez, J. J., Angermayr, S. A., Middleditch, M. J., Payne, L. S., & Leung, I. K. H. (2020). Effect of consecutive rare codons on the recombinant production of human proteins in Escherichia coli. IUBMB Life, 72(2), 266-274. https://doi.org/10.1002/iub.2162

91. Cowan, J. A. (1998). Metal Activation of Enzymes in Nucleic Acid Biochemistry. Chemical Reviews, 98(3), 1067-1088. https://doi.org/10.1021/cr960436q
92. Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr, 62(Pt 9), 1002-1011. https://doi.org/10.1107/S0907444906022116
93. Dale, G. E., Oefner, C., & D'Arcy, A. (2003). The protein as a variable in protein crystallization. J Struct Biol, 142(1), 88-97. https://doi.org/10.1016/s1047-8477(03)00041-8
94. Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. Methods Mol Biol, 1263, 243-250. https://doi.org/10.1007/978-1-4939-2269-7_19
95. Das, B. B., Ganguly, A., & Majumder, H. K. (2008). DNA Topoisomerases of Leishmania: The Potential Targets for Anti-Leishmanial Therapy. In H. K. Majumder (Ed.), Drug Targets in Kinetoplastid Parasites (pp. 103-115). Springer New York. https://doi.org/10.1007/978-0-387-77570-8_9
96. de la Cruz, J. J., Villanueva-Lizama, L., Dzul-Huchim, V., Ramirez-Sierra, M. J., Martinez-Vega, P., Rosado-Vallado, M., Ortega-Lopez, J., Flores-Pucheta, C. I., Gillespie, P., Zhan, B., Bottazzi, M. E., Hotez, P. J., & Dumonteil, E. (2019). Production of recombinant TSA-1 and evaluation of its potential for the immuno-therapeutic control of Trypanosoma cruzi infection in mice. Hum Vaccin Immunother, 15(1), 210-219. https://doi.org/10.1080/21645515.2018.1520581
97. De Luca, L., Ferro, S., Buemi, M. R., Monforte, A.-M., Gitto, R., Schirmeister, T., Maes, L., Rescifina, A., & Micale, N. (2018). Discovery of benzimidazole-based Leishmania mexicana cysteine protease CPB2.8ΔCTE inhibitors as potential therapeutics for leishmaniasis. Chemical Biology & Drug Design, 92(3), 1585-1596. https://doi.org/10.1111/cbdd.13326
98. De Souza Leao, S., Lang, T., Prina, E., Hellio, R., & Antoine, J. C. (1995). Intracellular Leishmania amazonensis amastigotes internalize and degrade MHC class II molecules of their host cells. Journal of Cell Science, 108(10), 3219-3231. https://jcs.biologists.org/content/joces/108/10/3219.full.pdf
99. Decherchi, S., & Cavalli, A. (2020). Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. Chem Rev, 120(23), 12788-12833. https://doi.org/10.1021/acs.chemrev.0c00534

100. Decuypere, S., Vanaerschot, M., Brunker, K., Imamura, H., Muller, S., Khanal, B., Rijal, S., Dujardin, J. C., & Coombs, G. H. (2012). Molecular mechanisms of drug resistance in natural Leishmania populations vary with genetic background. PLoS Negl Trop Dis, 6(2), e1514. https://doi.org/10.1371/journal.pntd.0001514

101. DeJesus, M. A., Gerrick, E. R., Xu, W., Park, S. W., Long, J. E., Boutte, C. C., Rubin, E. J., Schnappinger, D., Ehrt, S., Fortune, S. M., Sassetti, C. M., & Ioerger, T. R. (2017). Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. mBio, 8(1). https://doi.org/10.1128/mBio.02133-16

102. Del Nery, E., Juliano, M. A., Lima, A. P. C. A., Scharfstein, J., & Juliano, L. (1997). Kininogenase Activity by the Major Cysteinyl Proteinase (Cruzipain) from Trypanosoma cruzi. Journal of Biological Chemistry, 272(41), 25713-25718. https://doi.org/10.1074/jbc.272.41.25713

103. Delespaux, V., & de Koning, H. P. (2007). Drugs and drug resistance in African trypanosomiasis. Drug Resist Updat, 10(1-2), 30-50. https://doi.org/10.1016/j.drup.2007.02.004

104. Derewenda, Z. S. (2010). Application of protein engineering to enhance crystallizability and improve crystal properties. Acta Crystallogr D Biol Crystallogr, 66(Pt 5), 604-615. https://doi.org/10.1107/S090744491000644X

105. Deshmukh, A. L., Chandra, S., Singh, D. K., Siddiqi, M. I., & Banerjee, D. (2017). Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening. Molecular BioSystems, 13(8), 1630-1639. https://doi.org/10.1039/C7MB00118E

106. Dessau, M. A., & Modis, Y. (2011). Protein crystallization for X-ray crystallography. J Vis Exp(47). https://doi.org/10.3791/2285

107. Diaz, A., Lacks, S. A., & Lopez, P. (1992). The 5' to 3' exonuclease activity of DNA polymerase I is essential for Streptococcus pneumoniae. Mol Microbiol, 6(20), 3009-3019. https://doi.org/10.1111/j.1365-2958.1992.tb01759.x

108. DiMasi, J. A., & Grabowski, H. G. (2007). Economics of new oncology drug development. J Clin Oncol, 25(2), 209-216. https://doi.org/10.1200/JCO.2006.09.0803

109. DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health

Economics, 47, 20-33. https://doi.org/https://doi.org/10.1016/j.jhealeco.2016.01.012

110. do Monte-Neto, R. L., Coelho, A. C., Raymond, F., Legare, D., Corbeil, J., Melo, M. N., Frezard, F., & Ouellette, M. (2011). Gene expression profiling and molecular characterization of antimony resistance in Leishmania amazonensis. PLoS Negl Trop Dis, 5(5), e1167. https://doi.org/10.1371/journal.pntd.0001167

111. Docampo, R., & Moreno, S. N. (1984). Free radical metabolites in the mode of action of chemotherapeutic agents and phagocytic cells on Trypanosoma cruzi. Rev Infect Dis, 6(2), 223-238. https://doi.org/10.1093/clinids/6.2.223

112. Dorjsuren, D., Kim, D., Maloney, D. J., Wilson, D. M., 3rd, & Simeonov, A. (2011). Complementary non-radioactive assays for investigation of human flap endonuclease 1 activity. Nucleic Acids Res, 39(2), e11. https://doi.org/10.1093/nar/gkq1082

113. dos Santos, R. N., Ferreira, L. G., & Andricopulo, A. D. (2018). Practices in Molecular Docking and Structure-Based Virtual Screening. In M. Gore & U. B. Jagtap (Eds.), Computational Drug Discovery and Design (pp. 31-50). Springer New York. https://doi.org/10.1007/978-1-4939-7756-7_3

114. Dovrat, D., Stodola, J. L., Burgers, P. M., & Aharoni, A. (2014). Sequential switching of binding partners on PCNA during in vitro Okazaki fragment maturation. Proceedings of the National Academy of Sciences of the United States of America, 111(39), 14118-14123. https://doi.org/10.1073/pnas.1321349111

115. Doyle, P. S., Zhou, Y. M., Hsieh, I., Greenbaum, D. C., McKerrow, J. H., & Engel, J. C. (2011). The Trypanosoma cruzi protease cruzain mediates immune evasion. PLoS pathogens, 7(9), e1002139-e1002139. https://doi.org/10.1371/journal.ppat.1002139

116. Dreiman, G. H. S., Bictash, M., Fish, P. V., Griffin, L., & Svensson, F. (2021). Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding. SLAS Discov, 26(2), 257-262. https://doi.org/10.1177/2472555220949495

117. Duarte, M. C., Lage, D. P., Martins, V. T., Chavez-Fumagalli, M. A., Roatt, B. M., Menezes-Souza, D., Goulart, L. R., Soto, M., Tavares, C. A., & Coelho, E. A. (2016). Recent updates and perspectives on approaches for the development of vaccines against visceral leishmaniasis. Rev Soc Bras Med Trop, 49(4), 398-407. https://doi.org/10.1590/0037-8682-0120-2016

118. Dumonteil, E., Bottazzi, M. E., Zhan, B., Heffernan, M. J., Jones, K., Valenzuela, J. G., Kamhawi, S., Ortega, J., Rosales, S. P. d. L., Lee, B. Y., Bacon, K. M., Fleischer, B., Slingsby, B. T., Cravioto, M. B., Tapia-Conyer, R., & Hotez, P. J. (2012). Accelerating the development of a therapeutic vaccine for human Chagas disease: rationale and prospects AU - Dumonteil, Eric. Expert Review of Vaccines, 11(9), 1043-1055. https://doi.org/10.1586/erv.12.85

119. Dumonteil, E., Escobedo-Ortegon, J., Reyes-Rodriguez, N., Arjona-Torres, A., & Ramirez-Sierra, M. J. (2004). Immunotherapy of Trypanosoma cruzi infection with DNA vaccines in mice. Infect Immun, 72(1), 46-53. https://doi.org/10.1128/IAI.72.1.46-53.2004

120. Dumonteil, E., & Herrera, C. (2021). The Case for the Development of a Chagas Disease Vaccine: Why? How? When? Trop Med Infect Dis, 6(1). https://doi.org/10.3390/tropicalmed6010016

121. Dumonteil, E., Herrera, C., Tu, W., Goff, K., Fahlberg, M., Haupt, E., Kaur, A., Marx, P. A., Ortega-Lopez, J., Hotez, P. J., & Bottazzi, M. E. (2020). Safety and immunogenicity of a recombinant vaccine against Trypanosoma cruzi in Rhesus macaques. Vaccine, 38(29), 4584-4591. https://doi.org/10.1016/j.vaccine.2020.05.010

122. Eder, J., Sedrani, R., & Wiesmann, C. (2014). The discovery of first-in-class drugs: origins and evolution. Nat Rev Drug Discov, 13(8), 577-587. https://doi.org/10.1038/nrd4336

123. Eisenman, H. C., Casadevall, A., & McClelland, E. E. (2007). New insights on the pathogenesis of invasive Cryptococcus neoformans infection. Curr Infect Dis Rep, 9(6), 457-464. https://doi.org/10.1007/s11908-007-0070-8

124. Elena, C., Ravasi, P., Castelli, M. E., Peiru, S., & Menzella, H. G. (2014). Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. Front Microbiol, 5, 21. https://doi.org/10.3389/fmicb.2014.00021

125. Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr, 66(Pt 4), 486-501. https://doi.org/10.1107/S0907444910007493

126. Endy, D. (2005). Foundations for engineering biology. Nature, 438(7067), 449-453. https://doi.org/10.1038/nature04342

127. Engels, D., & Zhou, X. N. (2020). Neglected tropical diseases: an effective global response to local poverty-related disease priorities. Infect Dis Poverty, 9(1), 10. https://doi.org/10.1186/s40249-020-0630-9

128. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., & Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery [Review Article]. Nature Reviews Drug Discovery, 15, 605. https://doi.org/10.1038/nrd.2016.109

129. Ertl, P. (2014). Intuitive ordering of scaffolds and scaffold similarity searching using scaffold keys. J Chem Inf Model, 54(6), 1617-1622. https://doi.org/10.1021/ci5001983

130. Eschenfeldt, W. H., Maltseva, N., Stols, L., Donnelly, M. I., Gu, M., Nocek, B., Tan, K., Kim, Y., & Joachimiak, A. (2010). Cleavable C-terminal His-tag vectors for structure determination. J Struct Funct Genomics, 11(1), 31-39. https://doi.org/10.1007/s10969-010-9082-y

131. Evaluation, I. f. H. M. a. (2013). The Global Burden of Disease: Generating Evidence, Guiding Policy. IHME. https://www.healthdata.org/sites/default/files/files/policy_report/2013/GBD_GeneratingEvidence/IHME_GBD_GeneratingEvidence_FullReport.pdf

132. Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des, 15(5), 411-428. https://doi.org/10.1023/a:1011115820450

133. Exell, J. C., Thompson, M. J., Finger, L. D., Shaw, S. J., Debreczeni, J., Ward, T. A., McWhirter, C., Siöberg, C. L. B., Molina, D. M., Abbott, W. M., Jones, C. D., Nissink, J. W. M., Durant, S. T., & Grasby, J. A. (2016). Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. Nature Chemical Biology, 12, 815. https://doi.org/10.1038/nchembio.2148

134. Farabaugh, P. J. (2000). Translational frameshifting: implications for the mechanism of translational frame maintenance. Prog Nucleic Acid Res Mol Biol, 64, 131-170. https://doi.org/10.1016/s0079-6603(00)64004-7

135. Farooqui, A. A. (1980). Purification of enzymes by heparin-sepharose affinity chromatography. J Chromatogr, 184(3), 335-345. https://doi.org/10.1016/s0021-9673(00)89004-4

136. Faust, G., Stand, A., & Weuster-Botz, D. (2015). IPTG can replace lactose in auto-induction media to enhance protein expression in batch-cultured Escherichia coli. Engineering in Life Sciences, 15(8), 824-829. https://doi.org/https://doi.org/10.1002/elsc.201500011

137. Feng, M., Patel, D., Dervan, J. J., Ceska, T., Suck, D., Haq, I., & Sayers, J. R. (2004). Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol, 11(5), 450-456. https://doi.org/10.1038/nsmb754

138. Feng, T., & Barakat, K. (2018). Molecular Dynamics Simulation and Prediction of Druggable Binding Sites. In M. Gore & U. B. Jagtap (Eds.), Computational Drug Discovery and Design (pp. 87-103). Springer New York. https://doi.org/10.1007/978-1-4939-7756-7_6

139. Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. Molecules (Basel, Switzerland), 20(7), 13384-13421. https://doi.org/10.3390/molecules200713384

140. Ferreira, L. L. G., & Andricopulo, A. D. (2018). Chemoinformatics Strategies for Leishmaniasis Drug Discovery [Review]. Frontiers in Pharmacology, 9(1278). https://doi.org/10.3389/fphar.2018.01278

141. Finger, L. D., Atack, J. M., Tsutakawa, S., Classen, S., Tainer, J., Grasby, J., & Shen, B. (2012). The Wonders of Flap Endonucleases: Structure, Function, Mechanism and Regulation. In S. MacNeill (Ed.), The Eukaryotic Replisome: a Guide to Protein Structure and Function (pp. 301-326). Springer Netherlands. https://doi.org/10.1007/978-94-007-4572-8_16

142. Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc, 11(5), 905-919. https://doi.org/10.1038/nprot.2016.051

143. Franco, J. R., Cecchi, G., Paone, M., Diarra, A., Grout, L., Kadima Ebeja, A., Simarro, P. P., Zhao, W., & Argaw, D. (2022). The elimination of human African trypanosomiasis: Achievements in relation to WHO road map targets for 2020. PLoS Negl Trop Dis, 16(1), e0010047. https://doi.org/10.1371/journal.pntd.0010047

144. Freedland, S. J., Pantuck, A. J., Paik, S. H., Zisman, A., Graeber, T. G., Eisenberg, D., McBride, W. H., Nguyen, D., Tso, C. L., & Belldegrun, A. S. (2003). Heterogeneity of molecular targets on clonal cancer lines derived from a novel

hormone-refractory prostate cancer tumor system. Prostate, 55(4), 299-307. https://doi.org/10.1002/pros.10226

145. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of medicinal chemistry, 47(7), 1739-1749. https://doi.org/10.1021/jm0306430

146. Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L. S., Lane, D. P., Abbondandolo, A., & Dogliotti, E. (1996). Two Pathways for Base Excision Repair in Mammalian Cells. Journal of Biological Chemistry, 271(16), 9573-9578. https://doi.org/10.1074/jbc.271.16.9573

147. Fukushima, S., Itaya, M., Kato, H., Ogasawara, N., & Yoshikawa, H. (2007). Reassessment of the in vivo functions of DNA polymerase I and RNase H in bacterial cell growth. Journal of bacteriology, 189(23), 8575-8583. https://doi.org/10.1128/JB.00653-07

148. Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Soding, J., Lupas, A. N., & Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. Curr Protoc Bioinformatics, 72(1), e108. https://doi.org/10.1002/cpbi.108

149. Garforth, S. J., Patel, D., Feng, M., & Sayers, J. R. (2001). Unusually wide co-factor tolerance in a metalloenzyme; divalent metal ions modulate endo-exonuclease activity in T5 exonuclease. Nucleic Acids Research, 29(13), 2772-2779. https://doi.org/10.1093/nar/29.13.2772

150. Garg, P., & Burgers, P. M. J. (2005). DNA Polymerases that Propagate the Eukaryotic DNA Replication Fork. Critical Reviews in Biochemistry and Molecular Biology, 40(2), 115-128. https://doi.org/10.1080/10409230590935433

151. Garvey, E. P., Sharp, A. D., Warn, P. A., Yates, C. M., & Schotzinger, R. J. (2018). The novel fungal CYP51 inhibitor VT-1598 is efficacious alone and in combination with liposomal amphotericin B in a murine model of cryptococcal meningitis. J Antimicrob Chemother, 73(10), 2815-2822. https://doi.org/10.1093/jac/dky242

152. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy

Server. In J. M. Walker (Ed.), The Proteomics Protocols Handbook (pp. 571-607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571

153. Gerstein, A. C., Fu, M. S., Mukaremera, L., Li, Z., Ormerod, K. L., Fraser, J. A., Berman, J., & Nielsen, K. (2015). Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. mBio, 6(5), e01340-01315. https://doi.org/10.1128/mBio.01340-15

154. Ghosh, S., Goldgur, Y., & Shuman, S. (2020). Mycobacterial DNA polymerase I: activities and crystal structures of the POL domain as apoenzyme and in complex with a DNA primer-template and of the full-length FEN/EXO-POL enzyme. Nucleic Acids Res, 48(6), 3165-3180. https://doi.org/10.1093/nar/gkaa075

155. Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in Escherichia coli. Protein J, 32(6), 419-425. https://doi.org/10.1007/s10930-013-9502-5

156. Gordenin, D. A., Kunkel, T. A., & Resnick, M. A. (1997). Repeat expansion — all in flap? Nature Genetics, 16(2), 116-118. https://doi.org/10.1038/ng0697-116

157. Gossert, A. D. (2019). Assessing molecular interactions with biophysical methods using the validation cross. Biochem Soc Trans, 47(1), 63-76. https://doi.org/10.1042/BST20180271

158. Grenon, M., Gilbert, C., & Lowndes, N. F. (2001). Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. Nat Cell Biol, 3(9), 844-847. https://doi.org/10.1038/ncb0901-844

159. Gupta, S. K., & Shukla, P. (2016). Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications. Crit Rev Biotechnol, 36(6), 1089-1098. https://doi.org/10.3109/07388551.2015.1084264

160. Gutteridge, W. E. (1985). Existing chemotherapy and its limitations. Br Med Bull, 41(2), 162-168. https://doi.org/10.1093/oxfordjournals.bmb.a072044

161. Hai, T. P., Van, A. D., Ngan, N. T. T., Nhat, L. T. H., Lan, N. P. H., Vinh Chau, N. V., Thwaites, G. E., Krysan, D., & Day, J. N. (2019). The combination of tamoxifen with amphotericin B, but not with fluconazole, has synergistic activity against the majority of clinical isolates of Cryptococcus neoformans. Mycoses, 62(9), 818-825. https://doi.org/10.1111/myc.12955

162. Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. Chem Biol Drug Des, 69(2), 146-148. https://doi.org/10.1111/j.1747-0285.2007.00483.x

163. Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model, 49(2), 377-389. https://doi.org/10.1021/ci800324m

164. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of medicinal chemistry, 47(7), 1750-1759. https://doi.org/10.1021/jm030644s

165. Harrington, J. J., & Lieber, M. R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. The EMBO journal, 13(5), 1235-1246. https://www.ncbi.nlm.nih.gov/pubmed/8131753

166. Hartshorn, M. J., Murray, C. W., Cleasby, A., Frederickson, M., Tickle, I. J., & Jhoti, H. (2005). Fragment-based lead discovery using X-ray crystallography. Journal of medicinal chemistry, 48(2), 403-413. https://doi.org/10.1021/jm0495778

167. He, L., Luo, L., Zhu, H., Yang, H., Zhang, Y., Wu, H., Sun, H., Jiang, F., Kathera, C. S., Liu, L., Zhuang, Z., Chen, H., Pan, F., Hu, Z., Zhang, J., & Guo, Z. (2017). FEN1 promotes tumor progression and confers cisplatin resistance in non-small-cell lung cancer. Molecular Oncology, 11(6), 640-654. https://doi.org/10.1002/1878-0261.12058

168. He, L., Zhang, Y., Sun, H., Jiang, F., Yang, H., Wu, H., Zhou, T., Hu, S., Kathera, C. S., Wang, X., Chen, H., Li, H., Shen, B., Zhu, Y., & Guo, Z. (2016). Targeting DNA Flap Endonuclease 1 to Impede Breast Cancer Progression. EBioMedicine, 14, 32-43. https://doi.org/10.1016/j.ebiom.2016.11.012

169. Hecker, S. J., Reddy, K. R., Totrov, M., Hirst, G. C., Lomovskaya, O., Griffith, D. C., King, P., Tsivkovski, R., Sun, D., Sabet, M., Tarazi, Z., Clifton, M. C., Atkins, K., Raymond, A., Potts, K. T., Abendroth, J., Boyer, S. H., Loutit, J. S., Morgan, E. E., . . . Dudley, M. N. (2015). Discovery of a Cyclic Boronic Acid beta-Lactamase Inhibitor (RPX7009) with Utility vs Class A Serine Carbapenemases. Journal of medicinal chemistry, 58(9), 3682-3692. https://doi.org/10.1021/acs.jmedchem.5b00127

170. Heinemann, M., & Panke, S. (2006). Synthetic biology--putting engineering into biology. Bioinformatics, 22(22), 2790-2799. https://doi.org/10.1093/bioinformatics/btl469

171. Heinrich, R., Montero, F., Klipp, E., Waddell, T. G., & Meléndez-Hevia, E. (1997). Theoretical Approaches to the Evolutionary Optimization of Glycolysis. European Journal of Biochemistry, 243(1-2), 191-201. https://doi.org/10.1111/j.1432-1033.1997.0191a.x

172. Henrich, S., Salo-Ahen, O. M., Huang, B., Rippmann, F. F., Cruciani, G., & Wade, R. C. (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. J Mol Recognit, 23(2), 209-219. https://doi.org/10.1002/jmr.984

173. Henricksen, L. A., Veeraraghavan, J., Chafin, D. R., & Bambara, R. A. (2002). DNA Ligase I Competes with FEN1 to Expand Repetitive DNA Sequences in Vitro. Journal of Biological Chemistry, 277(25), 22361-22369. https://doi.org/10.1074/jbc.M201765200

174. Higgins, M. K., & Lea, S. M. (2017). On the state of crystallography at the dawn of the electron microscopy revolution. Curr Opin Struct Biol, 46, 95-101. https://doi.org/10.1016/j.sbi.2017.06.005

175. HIV/AIDS, U. T. J. U. N. P. o. (2022). Global HIV & AIDS statistics - Fact sheet. In (pp. 6).

176. Hollingsworth, T. D. (2018). Counting Down the 2020 Goals for 9 Neglected Tropical Diseases: What Have We Learned From Quantitative Analysis and Transmission Modeling? Clinical Infectious Diseases, 66(suppl_4), S237-S244. https://doi.org/10.1093/cid/ciy284

177. Honarparvar, B., Govender, T., Maguire, G. E., Soliman, M. E., & Kruger, H. G. (2014). Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity. Chem Rev, 114(1), 493-537. https://doi.org/10.1021/cr300314q

178. Hopfner, K. P. (2014). ATP puts the brake on DNA double-strand break repair: a new study shows that ATP switches the Mre11-Rad50-Nbs1 repair factor between signaling and processing of DNA ends. BioEssays, 36(12), 1170-1178. https://doi.org/10.1002/bies.201400102

179. Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A., & Shen, B. (1998). Newly Discovered Archaebacterial Flap Endonucleases Show a Structure-Specific Mechanism for DNA Substrate Binding and Catalysis Resembling Human Flap Endonuclease-1. Journal of Biological Chemistry, 273(42), 27154-27161. https://doi.org/10.1074/jbc.273.42.27154

180. Hosfield, D. J., Mol, C. D., Shen, B., & Tainer, J. A. (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. Cell, 95(1), 135-146. https://doi.org/10.1016/s0092-8674(00)81789-4

181. Hotez, P. J. (2013). Forgotten people, forgoten diseases: the neglected tropical diseases and their impact on global health and development (2nd edition ed.). ASM Press.

182. Hotez, P. J. (2020). NTDs in the 2020s: An epic struggle of effective control tools versus the Anthropocene. PLoS Negl Trop Dis, 14(9), e0007872. https://doi.org/10.1371/journal.pntd.0007872

183. Hotez, P. J., Alvarado, M., Basáñez, M.-G., Bolliger, I., Bourne, R., Boussinesq, M., Brooker, S. J., Brown, A. S., Buckle, G., Budke, C. M., Carabin, H., Coffeng, L. E., Fèvre, E. M., Fürst, T., Halasa, Y. A., Jasrasaria, R., Johns, N. E., Keiser, J., King, C. H., . . . Naghavi, M. (2014). The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. PLOS Neglected Tropical Diseases, 8(7), e2865. https://doi.org/10.1371/journal.pntd.0002865

184. Hotez, P. J., Pecoul, B., Rijal, S., Boehme, C., Aksoy, S., Malecela, M., Tapia-Conyer, R., & Reeder, J. C. (2016). Eliminating the Neglected Tropical Diseases: Translational Science and New Technologies. PLOS Neglected Tropical Diseases, 10(3), e0003895. https://doi.org/10.1371/journal.pntd.0003895

185. Hu, G., Cheng, P. Y., Sham, A., Perfect, J. R., & Kronstad, J. W. (2008). Metabolic adaptation in Cryptococcus neoformans during early murine pulmonary infection. Mol Microbiol, 69(6), 1456-1475. https://doi.org/10.1111/j.1365-2958.2008.06374.x

186. Hu, Y., Stumpfe, D., & Bajorath, J. (2017). Recent Advances in Scaffold Hopping. Journal of medicinal chemistry, 60(4), 1238-1246. https://doi.org/10.1021/acs.jmedchem.6b01437

187. Huggins, C. F., Chafin, D. R., Aoyagi, S., Henricksen, L. A., Bambara, R. A., & Hayes, J. J. (2002). Flap Endonuclease 1 Efficiently Cleaves Base Excision Repair and DNA Replication Intermediates Assembled into Nucleosomes. Molecular Cell, 10(5), 1201-1211. https://doi.org/https://doi.org/10.1016/S1097-2765(02)00736-0

188. Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. British journal of pharmacology, 162(6), 1239-1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

189. Hussein, H. A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., & Camproux, A. C. (2015). PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res, 43(W1), W436-442. https://doi.org/10.1093/nar/gkv462

190. Hwang, K. Y., Baek, K., Kim, H. Y., & Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nat Struct Biol, 5(8), 707-713. https://doi.org/10.1038/1406

191. Ilari, A., & Savino, C. (2008). Protein structure determination by x-ray crystallography. Methods Mol Biol, 452, 63-87. https://doi.org/10.1007/978-1-60327-159-2_3

192. Illuzzi, J. L., & Wilson, D. M., 3rd. (2012). Base excision repair: contribution to tumorigenesis and target in anticancer treatment paradigms. Curr Med Chem, 19(23), 3922-3936. https://doi.org/10.2174/092986712802002581

193. initiative, D. f. N. D. (2017). NTD Drug Discovery Booster. In. Annual meeting of the Pharmaceutical Society of Japan, Sendai.

194. initiative, D. f. N. D. (2020). 2019 DNDi Annual Report https://dndi.org/wp-content/uploads/2020/07/DNDi-AnnualReport-2019.pdf

195. Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. J Chem Inf Model, 60(12), 6065-6073. https://doi.org/10.1021/acs.jcim.0c00675

196. Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature, 461(7267), 1071-1078. https://doi.org/10.1038/nature08467

197. Jacobs, R. T., Nare, B., Wring, S. A., Orr, M. D., Chen, D., Sligar, J. M., Jenks, M. X., Noe, R. A., Bowling, T. S., Mercer, L. T., Rewerts, C., Gaukel, E., Owens,

J., Parham, R., Randolph, R., Beaudet, B., Bacchi, C. J., Yarlett, N., Plattner, J. J., . . . Don, R. (2011). SCYX-7158, an Orally-Active Benzoxaborole for the Treatment of Stage 2 Human African Trypanosomiasis. PLOS Neglected Tropical Diseases, 5(6), e1151. https://doi.org/10.1371/journal.pntd.0001151

198. Jain, A. N., & Nicholls, A. (2008). Recommendations for evaluation of computational methods. J Comput Aided Mol Des, 22(3-4), 133-139. https://doi.org/10.1007/s10822-008-9196-5

199. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P., & Herschlag, D. (2020). How to measure and evaluate binding affinities. Elife, 9. https://doi.org/10.7554/eLife.57264

200. Jennings, F. W., & Urquhart, G. M. (1983). The use of the 2 substituted 5-nitroimidazole, fexinidazole (Hoe 239) in the treatment of chronicT. brucei infections in mice [journal article]. Zeitschrift für Parasitenkunde, 69(5), 577-581. https://doi.org/10.1007/bf00926669

201. Jha, T. K., Sundar, S., Thakur, C. P., Bachmann, P., Karbwang, J., Fischer, C., Voss, A., & Berman, J. (1999). Miltefosine, an oral agent, for the treatment of Indian visceral leishmaniasis. N Engl J Med, 341(24), 1795-1800. https://doi.org/10.1056/NEJM199912093412403

202. Jin, L., Wang, W., & Fang, G. (2014). Targeting protein-protein interaction by small molecules. Annu Rev Pharmacol Toxicol, 54, 435-456. https://doi.org/10.1146/annurev-pharmtox-011613-140028

203. Johns, B. A., Kawasuji, T., Weatherhead, J. G., Taishi, T., Temelkoff, D. P., Yoshida, H., Akiyama, T., Taoda, Y., Murai, H., Kiyama, R., Fuji, M., Tanimoto, N., Jeffrey, J., Foster, S. A., Yoshinaga, T., Seki, T., Kobayashi, M., Sato, A., Johnson, M. N., . . . Fujiwara, T. (2013). Carbamoyl pyridone HIV-1 integrase inhibitors 3. A diastereomeric approach to chiral nonracemic tricyclic ring systems and the discovery of dolutegravir (S/GSK1349572) and (S/GSK1265744). Journal of medicinal chemistry, 56(14), 5901-5916. https://doi.org/10.1021/jm400645w

204. Johnston, E. A., Teague, J., & Graham, J. P. (2015). Challenges and opportunities associated with neglected tropical disease and water, sanitation and hygiene intersectoral integration programs. BMC Public Health, 15, 547. https://doi.org/10.1186/s12889-015-1838-7

205. Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices11Edited by G. Von Heijne. Journal of Molecular Biology, 292(2), 195-202. https://doi.org/https://doi.org/10.1006/jmbi.1999.3091

206. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. J Mol Biol, 267(3), 727-748. https://doi.org/10.1006/jmbi.1996.0897

207. Jones, P., McElroy, S., Morrison, A., & Pannifer, A. (2015). The importance of triaging in determining the quality of output from high-throughput screening. Future Med Chem, 7(14), 1847-1852. https://doi.org/10.4155/fmc.15.121

208. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

209. Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol, 6(5), 494-500. https://doi.org/10.1016/0958-1669(95)80082-4

210. Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach. Chemico-biological interactions, 171(2), 165-176. https://doi.org/10.1016/j.cbi.2006.12.006

211. Kapust, R. B., Routzahn, K. M., & Waugh, D. S. (2002). Processive degradation of nascent polypeptides, triggered by tandem AGA codons, limits the accumulation of recombinant tobacco etch virus protease in Escherichia coli BL21(DE3). Protein Expr Purif, 24(1), 61-70. https://doi.org/10.1006/prep.2001.1545

212. Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., Sugimoto, K., & Shirahige, K. (2003). S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature, 424(6952), 1078-1083. https://doi.org/10.1038/nature01900

213. Katritch, V., Jaakola, V. P., Lane, J. R., Lin, J., Ijzerman, A. P., Yeager, M., Kufareva, I., Stevens, R. C., & Abagyan, R. (2010). Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. Journal of medicinal chemistry, 53(4), 1799-1809. https://doi.org/10.1021/jm901647p

214. Kawabata, T. (2010). Detection of multiscale pockets on protein surfaces using mathematical morphology. Proteins, 78(5), 1195-1211. https://doi.org/10.1002/prot.22639

215. Kaye, K. S., Bhowmick, T., Metallidis, S., Bleasdale, S. C., Sagan, O. S., Stus, V., Vazquez, J., Zaitsev, V., Bidair, M., Chorvat, E., Dragoescu, P. O., Fedosiuk, E., Horcajada, J. P., Murta, C., Sarychev, Y., Stoev, V., Morgan, E., Fusaro, K., Griffith, D., . . . Giamarellos-Bourboulis, E. J. (2018). Effect of Meropenem-Vaborbactam vs Piperacillin-Tazobactam on Clinical Cure or Improvement and Microbial Eradication in Complicated Urinary Tract Infection: The TANGO I Randomized Clinical Trial. JAMA, 319(8), 788-799. https://doi.org/10.1001/jama.2018.0438

216. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10, 845. https://doi.org/10.1038/nprot.2015.053

217. Kim, K., Biade, S., & Matsumoto, Y. (1998). Involvement of Flap Endonuclease 1 in Base Excision DNA Repair. Journal of Biological Chemistry, 273(15), 8842-8848. https://doi.org/10.1074/jbc.273.15.8842

218. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res, 49(D1), D1388-D1395. https://doi.org/10.1093/nar/gkaa971

219. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., & Steitz, T. A. (1995). Crystal structure of Thermus aquaticus DNA polymerase. Nature, 376(6541), 612-616. https://doi.org/10.1038/376612a0

220. Kinch, M. S., & Hoyer, D. (2015). A history of drug development in four acts. Drug Discov Today, 20(10), 1163-1168. https://doi.org/10.1016/j.drudis.2015.04.003

221. Kingsnorth, A. N. (1986). The chemotherapeutic potential of polyamine antimetabolites. Ann R Coll Surg Engl, 68(2), 76-81. https://www.ncbi.nlm.nih.gov/pubmed/3082276

222. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov, 3(11), 935-949. https://doi.org/10.1038/nrd1549

223. Klenow, H., & Henningsen, I. (1970). Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. Proceedings of the National Academy of Sciences of the United States of America, 65(1), 168-175. https://doi.org/10.1073/pnas.65.1.168

224. Klenow, H., & Overgaard-Hansen, K. (1970). Proteolytic cleavage of DNA polymerase from Escherichia Coli B into an exonuclease unit and a polymerase unit. FEBS Lett, 6(1), 25-27. https://doi.org/10.1016/0014-5793(70)80032-1

225. Kodama, H., Shimojo, N., & Suzuki, K. T. (1991). Distribution of manganese in rat pancreas and identification of its primary binding protein as pro-carboxypeptidase B. Biochem J, 278 (Pt 3), 857-862. https://doi.org/10.1042/bj2780857

226. Koorneef, L. L., van den Heuvel, J. K., Kroon, J., Boon, M. R., t Hoen, P. A. C., Hettne, K. M., van de Velde, N. M., Kolenbrander, K. B., Streefland, T. C. M., Mol, I. M., Sips, H. C. M., Kielbasa, S. M., Mei, H., Belanoff, J. K., Pereira, A. M., Oosterveer, M. H., Hunt, H., Rensen, P. C. N., & Meijer, O. C. (2018). Selective Glucocorticoid Receptor Modulation Prevents and Reverses Nonalcoholic Fatty Liver Disease in Male Mice. Endocrinology, 159(12), 3925-3936. https://doi.org/10.1210/en.2018-00671

227. Koselny, K., Green, J., DiDone, L., Halterman, J. P., Fothergill, A. W., Wiederhold, N. P., Patterson, T. F., Cushion, M. T., Rappelye, C., Wellington, M., & Krysan, D. J. (2016). The Celecoxib Derivative AR-12 Has Broad-Spectrum Antifungal Activity In Vitro and Improves the Activity of Fluconazole in a Murine Model of Cryptococcosis. Antimicrobial agents and chemotherapy, 60(12), 7115-7127. https://doi.org/10.1128/AAC.01061-16

228. Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., Xia, B., Beglov, D., & Vajda, S. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc, 10(5), 733-755. https://doi.org/10.1038/nprot.2015.043

229. Kucherlapati, M., Yang, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M. F., Fan, K., Russell, R., Brown, A. M., Kneitz, B., Edelmann, W., Kolodner, R. D., Lipkin, M., & Kucherlapati, R. (2002). Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression. Proceedings of the National Academy of

Sciences of the United States of America, 99(15), 9924-9929. https://doi.org/10.1073/pnas.152321699

230. Kumar, A., Voet, A., & Zhang, K. Y. (2012). Fragment based drug design: from experimental to computational approaches. Curr Med Chem, 19(30), 5128-5147. https://doi.org/10.2174/092986712803530467

231. Kumer, A., Chakma, U., Rana, M. M., Chandro, A., Akash, S., Elseehy, M. M., Albogami, S., & El-Shehawi, A. M. (2022). Investigation of the New Inhibitors by Sulfadiazine and Modified Derivatives of alpha-D-glucopyranoside for White Spot Syndrome Virus Disease of Shrimp by In Silico: Quantum Calculations, Molecular Docking, ADMET and Molecular Dynamics Study. Molecules (Basel, Switzerland), 27(12). https://doi.org/10.3390/molecules27123694

232. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. J Mol Biol, 161(2), 269-288. https://doi.org/10.1016/0022-2836(82)90153-x

233. Kwon, S., Bae, H., Jo, J., & Yoon, S. (2019). Comprehensive ensemble in QSAR prediction for drug discovery. BMC Bioinformatics, 20(1), 521. https://doi.org/10.1186/s12859-019-3135-4

234. Kwon-Chung, K. J., Fraser, J. A., Doering, T. L., Wang, Z., Janbon, G., Idnurm, A., & Bahn, Y.-S. (2014). Cryptococcus neoformans and Cryptococcus gattii, the etiologic agents of cryptococcosis. Cold Spring Harbor perspectives in medicine, 4(7), a019760-a019760. https://doi.org/10.1101/cshperspect.a019760

235. Lam, J. S., Seligson, D. B., Yu, H., Li, A., Eeva, M., Pantuck, A. J., Zeng, G., Horvath, S., & Belldegrun, A. S. (2006). Flap endonuclease 1 is overexpressed in prostate cancer and is associated with a high Gleason score. BJU Int, 98(2), 445-451. https://doi.org/10.1111/j.1464-410X.2006.06224.x

236. Lancey, C., Tehseen, M., Raducanu, V. S., Rashid, F., Merino, N., Ragan, T. J., Savva, C. G., Zaher, M. S., Shirbini, A., Blanco, F. J., Hamdan, S. M., & De Biasio, A. (2020). Structure of the processive human Pol delta holoenzyme. Nat Commun, 11(1), 1109. https://doi.org/10.1038/s41467-020-14898-6

237. Larsen, E., Gran, C., Saether, B. E., Seeberg, E., & Klungland, A. (2003). Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. Molecular and cellular biology, 23(15), 5346-5353. https://doi.org/10.1128/MCB.23.15.5346-5353.2003

238. Larsen, R. A., Pappas, P. G., Perfect, J., Aberg, J. A., Casadevall, A., Cloud, G. A., James, R., Filler, S., & Dismukes, W. E. (2005). Phase I evaluation of the safety and pharmacokinetics of murine-derived anticryptococcal antibody 18B7 in subjects with treated cryptococcal meningitis. Antimicrobial agents and chemotherapy, 49(3), 952-958. https://doi.org/10.1128/AAC.49.3.952-958.2005

239. Laurie, A. T., & Jackson, R. M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics, 21(9), 1908-1916. https://doi.org/10.1093/bioinformatics/bti315

240. Lavecchia, A., & Giovanni, C. D. (2013). Virtual Screening Strategies in Drug Discovery: A Critical Review. Current Medicinal Chemistry, 20(23), 2839-2860. https://doi.org/http://dx.doi.org/10.2174/09298673113209990001

241. Lazera, M. S., Salmito Cavalcanti, M. A., Londero, A. T., Trilles, L., Nishikawa, M. M., & Wanke, B. (2000). Possible primary ecological niche of Cryptococcus neoformans. Med Mycol, 38(5), 379-383. https://doi.org/10.1080/mmy.38.5.379.383

242. Lazo-Porras, M., Prutsky, G. J., Barrionuevo, P., Tapia, J. C., Ugarte-Gil, C., Ponce, O. J., Acuña-Villaorduña, A., Domecq, J. P., De la Cruz-Luque, C., Prokop, L. J., & Málaga, G. (2020). World Health Organization (WHO) antibiotic regimen against other regimens for the treatment of leprosy: a systematic review and meta-analysis. BMC Infect Dis, 20(1), 62. https://doi.org/10.1186/s12879-019-4665-0

243. Leder, K., & Weller, P. L. (2015). Antiparasitic Agents. In Manual of Clinical Microbiology (11 ed., pp. 2529-2549). ASM Press.

244. Lee, H. S., Jo, S., Lim, H. S., & Im, W. (2012). Application of binding free energy calculations to prediction of binding modes and affinities of MDM2 and MDMX inhibitors. J Chem Inf Model, 52(7), 1821-1832. https://doi.org/10.1021/ci3000997

245. Lee, J., Shin, M.-K., Ryu, D.-K., Kim, S., & Ryu, W.-S. (2010). Insertion and Deletion Mutagenesis by Overlap Extension PCR. In J. Braman (Ed.), In Vitro Mutagenesis Protocols: Third Edition (pp. 137-146). Humana Press. https://doi.org/10.1007/978-1-60761-652-8_10

246. Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. Beilstein J Org Chem, 12, 2694-2718. https://doi.org/10.3762/bjoc.12.267

247. Leonard, E., Nielsen, D., Solomon, K., & Prather, K. J. (2008). Engineering microbes with synthetic biology frameworks. Trends Biotechnol, 26(12), 674-681. https://doi.org/10.1016/j.tibtech.2008.08.003

248. Li, L., & Zou, L. (2005). Sensing, signaling, and responding to DNA damage: Organization of the checkpoint pathways in mammalian cells. Journal of Cellular Biochemistry, 94(2), 298-306. https://doi.org/10.1002/jcb.20355

249. Lidani, K. C. F., Andrade, F.A., Tizzot, M.R.P.A., Costa-Ribeiro M.C.V., Beltrame M.H., Messias-Reason, I.J. (2017). Visceral Leishmaniasis and Natural Infection Rates of Leishmania in Lutzomyia longipalpis in Latin America. The Epidemiology and Ecology of Leishmaniasis, IntechOpen. https://doi.org/10.5772/65787

250. Lieber, M. R. (1997). The FEN-1 family of structure-specific nucleases in eukaryotic dna replication, recombination and repair. BioEssays, 19(3), 233-240. https://doi.org/10.1002/bies.950190309

251. Lin, T., Zhang, L., Jiang, D., Wu, L., Chen, K., Li, L., Jin, C., Li, Z., & Oger, P. (2022). Biochemical characterization and mutational analysis of a novel flap endonuclease 1 from Thermococcus barophilus Ch5. Int J Biochem Cell Biol, 143, 106154. https://doi.org/10.1016/j.biocel.2021.106154

252. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature, 362(6422), 709-715. https://doi.org/10.1038/362709a0

253. Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: principles, applications and recent advances. Current topics in medicinal chemistry, 14(16), 1923-1938. https://doi.org/10.2174/1568026614666140929124445

254. Lipinski, C. A. (2016). Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. Advanced Drug Delivery Reviews, 101, 34-41. https://doi.org/https://doi.org/10.1016/j.addr.2016.04.029

255. Lo, Y. C., Gui, R., Honda, H., & Torres, J. Z. (2016). Quantitative Methods in System-Based Drug Discovery. Complex Systems, Sustainability and Innovation. https://doi.org/10.5772/65833

256. Lockhart, S. R., Fothergill, A. W., Iqbal, N., Bolden, C. B., Grossman, N. T., Garvey, E. P., Brand, S. R., Hoekstra, W. J., Schotzinger, R. J., Ottinger, E.,

Patterson, T. F., & Wiederhold, N. P. (2016). The Investigational Fungal Cyp51 Inhibitor VT-1129 Demonstrates Potent In Vitro Activity against Cryptococcus neoformans and Cryptococcus gattii. Antimicrobial agents and chemotherapy, 60(4), 2528-2531. https://doi.org/10.1128/AAC.02770-15

257. Loiseau, P. M., Cojean, S., & Schrével, J. (2011). Sitamaquine as a putative antileishmanial drug candidate: from the mechanism of action to the risk of drug resistance. Parasite (Paris, France), 18(2), 115-119. https://doi.org/10.1051/parasite/2011182115

258. Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., Newlon, C. S., & Foiani, M. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. Nature, 412(6846), 557-561. https://doi.org/10.1038/35087613

259. Lotun, D. P., Cochard, C., Vieira, F. R. J., & Bernardes, J. S. (2019). 2dSS: a web server for protein secondary structure visualization. bioRxiv, 649426. https://doi.org/10.1101/649426

260. Lovett, S. T. (2007). Polymerase Switching in DNA Replication. Molecular Cell, 27(4), 523-526. https://doi.org/https://doi.org/10.1016/j.molcel.2007.08.003

261. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., AlMazroa, M. A., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Barker-Collo, S., Bartels, D. H., . . . Murray, C. J. L. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet, 380(9859), 2095-2128. https://doi.org/10.1016/S0140-6736(12)61728-0

262. Lozano Terol, G., Gallego-Jara, J., Sola Martinez, R. A., Martinez Vivancos, A., Canovas Diaz, M., & de Diego Puente, T. (2021). Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21. Front Microbiol, 12, 682001. https://doi.org/10.3389/fmicb.2021.682001

263. Lukeš, J., Wheeler, R., Jirsova, D., David, V., & Archibald, J. M. (2018). Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. IUBMB Life, 70(12), 1267-1274. https://doi.org/10.1002/iub.1894

264. Lundquist, R. C., & Olivera, B. M. (1982). Transient generation of displaced single-stranded DNA during nick translation. Cell, 31(1), 53-60. https://doi.org/10.1016/0092-8674(82)90404-4

265. Lyamichev, V., Brow, M. A., & Dahlberg, J. E. (1993). Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science, 260(5109), 778. https://doi.org/10.1126/science.7683443

266. Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., O'Meara, M. J., Che, T., Algaa, E., Tolmachova, K., Tolmachev, A. A., Shoichet, B. K., Roth, B. L., & Irwin, J. J. (2019). Ultra-large library docking for discovering new chemotypes. Nature, 566(7743), 224-229. https://doi.org/10.1038/s41586-019-0917-9

267. Ma, L., Cao, X., Wang, H., Lu, K., Wang, Y., Tu, C., Dai, Y., Meng, Y., Li, Y., Yu, P., Man, S., & Diao, A. (2019). Discovery of Myricetin as a Potent Inhibitor of Human Flap Endonuclease 1, Which Potentially Can Be Used as Sensitizing Agent against HT-29 Human Colon Cancer Cells. Journal of Agricultural and Food Chemistry. https://doi.org/10.1021/acs.jafc.8b05447

268. 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(3), 188-195. https://doi.org/10.1038/nrd3368

269. Macip, G., Garcia-Segura, P., Mestres-Truyol, J., Saldivar-Espinoza, B., Ojeda-Montes, M. J., Gimeno, A., Cereto-Massague, A., Garcia-Vallve, S., & Pujadas, G. (2022). Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. Med Res Rev, 42(2), 744-769. https://doi.org/10.1002/med.21862

270. MacLean, L., Price, H., & O'Toole, P. (2016). Exploring the Leishmania Hydrophilic Acylated Surface Protein B (HASPB) Export Pathway by Live Cell Imaging Methods. In A. Pompa & F. De Marchis (Eds.), Unconventional Protein Secretion: Methods and Protocols (pp. 191-203). Springer New York. https://doi.org/10.1007/978-1-4939-3804-9_13

271. Maguire, M. E., & Cowan, J. A. (2002). Magnesium chemistry and biochemistry. Biometals, 15(3), 203-210. https://doi.org/10.1023/a:1016058229972

272. Mahmoudi Gomari, M., Saraygord-Afshari, N., Farsimadan, M., Rostami, N., Aghamiri, S., & Farajollahi, M. M. (2020). Opportunities and challenges of the tag-

assisted protein purification techniques: Applications in the pharmaceutical industry. Biotechnology Advances, 45, 107653. https://doi.org/https://doi.org/10.1016/j.biotechadv.2020.107653

273. Mahmoudzadeh-Niknam, H., & McKerrow, J. H. (2004). Leishmania tropica: cysteine proteases are essential for growth and pathogenicity. Experimental Parasitology, 106(3), 158-163. https://doi.org/https://doi.org/10.1016/j.exppara.2004.03.005
274. Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. Front Chem, 8, 343. https://doi.org/10.3389/fchem.2020.00343
275. Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M., & Striedner, G. (2013). Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free Escherichia coli expression systems under production conditions. Appl Environ Microbiol, 79(12), 3802-3812. https://doi.org/10.1128/AEM.00365-13
276. Malygin, A., Baranovskaya, O., Ivanov, A., & Karpova, G. (2003). Expression and purification of human ribosomal proteins S3, S5, S10, S19, and S26. Protein Expr Purif, 28(1), 57-62. https://doi.org/10.1016/s1046-5928(02)00652-6
277. Marreiros, B. C., Sena, F. V., Sousa, F. M., Batista, A. P., & Pereira, M. M. (2016). Type II NADH:quinone oxidoreductase family: phylogenetic distribution, structural diversity and evolutionary divergences. Environmental Microbiology, 18(12), 4697-4709. https://doi.org/10.1111/1462-2920.13352
278. Marschall, L., Sagmeister, P., & Herwig, C. (2017). Tunable recombinant protein expression in E. coli: promoter systems and genetic constraints. Appl Microbiol Biotechnol, 101(2), 501-512. https://doi.org/10.1007/s00253-016-8045-z
279. Martis, E. A., Radhakrishnan, R., & Badve, R. R. (2011). High-Throughput Screening: The Hits and Leads of Drug Discovery- An Overview.
280. Mase, T., Kubota, K., Miyazono, K., Kawarabayasi, Y., & Tanokura, M. (2011). Structure of flap endonuclease 1 from the hyperthermophilic archaeon Desulfurococcus amylolyticus. Acta crystallographica. Section F, Structural biology and crystallization communications, 67(Pt 2), 209-213. https://doi.org/10.1107/S1744309110053030

281. Mata-Cantero, L., Xie, S. C., Garcia, M., Coyle, J., Fernandez, R., Cabrera, A. C., Gillett, D. L., Crespo, B., Gamo, F. J., Fernandez, E., Tilley, L., & Gomez-Lorenzo, M. G. (2021). High Throughput Screening to Identify Selective and Nonpeptidomimetic Proteasome Inhibitors As Antimalarials. ACS Infect Dis, 7(6), 1818-1832. https://doi.org/10.1021/acsinfecdis.1c00110

282. Mathieu, C., Gonzalez Salgado, A., Wirdnam, C., Meier, S., Grotemeyer, M. S., Inbar, E., Maser, P., Zilberstein, D., Sigel, E., Butikofer, P., & Rentsch, D. (2014). Trypanosoma brucei eflornithine transporter AAT6 is a low-affinity low-selective transporter for neutral amino acids. Biochem J, 463(1), 9-18. https://doi.org/10.1042/BJ20140719

283. Matsumoto, Y., Kim, K., & Bogenhagen, D. F. (1994). Proliferating cell nuclear antigen-dependent abasic site repair in Xenopus laevis oocytes: an alternative pathway of base excision DNA repair. Molecular and cellular biology, 14(9), 6187-6197. https://doi.org/10.1128/mcb.14.9.6187

284. Matsuzaki, Y., Adachi, N., & Koyama, H. (2002). Vertebrate cells lacking FEN-1 endonuclease are viable but hypersensitive to methylating agents and H2O2. Nucleic Acids Res, 30(14), 3273-3277. https://doi.org/10.1093/nar/gkf440

285. Matthews, R. C., Rigg, G., Hodgetts, S., Carter, T., Chapman, C., Gregory, C., Illidge, C., & Burnie, J. (2003). Preclinical assessment of the efficacy of mycograb, a human recombinant antibody against fungal HSP90. Antimicrobial agents and chemotherapy, 47(7), 2208-2216. https://doi.org/10.1128/AAC.47.7.2208-2216.2003

286. McCoy, A. J. (2009). Protein Crystallography with Coffee. In. In BCA Protein Crystallography Summer School 2010.

287. McDaniel, R., & Weiss, R. (2005). Advances in synthetic biology: on the path from prototypes to applications. Curr Opin Biotechnol, 16(4), 476-483. https://doi.org/10.1016/j.copbio.2005.07.002

288. McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., & Brown, F. K. (2003). Gaussian docking functions. Biopolymers, 68(1), 76-90. https://doi.org/10.1002/bip.10207

289. McGowan, C. H., & Russell, P. (2004). The DNA damage response: sensing and signaling. Current Opinion in Cell Biology, 16(6), 629-633. https://doi.org/https://doi.org/10.1016/j.ceb.2004.09.005

290. McManus, K. J., Barrett, I. J., Nouhi, Y., & Hieter, P. (2009). Specific synthetic lethal killing of RAD54B-deficient human colorectal cancer cells by FEN1 silencing. Proceedings of the National Academy of Sciences of the United States of America, 106(9), 3276-3281. https://doi.org/10.1073/pnas.0813414106

291. McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. Acta crystallographica. Section F, Structural biology communications, 70(Pt 1), 2-20. https://doi.org/10.1107/S2053230X13033141

292. McWhirter, C., Tonge, M., Plant, H., Hardern, I., Nissink, W., & Durant, S. T. (2013). Development of a high-throughput fluorescence polarization DNA cleavage assay for the identification of FEN1 inhibitors. J Biomol Screen, 18(5), 567-575. https://doi.org/10.1177/1087057113476551

293. Meng, H., Wang, J., Xiong, Z., Xu, F., Zhao, G., & Wang, Y. (2013). Quantitative design of regulatory elements based on high-precision strength prediction using artificial neural network. PLOS ONE, 8(4), e60288. https://doi.org/10.1371/journal.pone.0060288

294. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des, 7(2), 146-157. https://doi.org/10.2174/157340911795677602

295. Mitsuyama, J., Nomura, N., Hashimoto, K., Yamada, E., Nishikawa, H., Kaeriyama, M., Kimura, A., Todo, Y., & Narita, H. (2008). In vitro and in vivo antifungal activities of T-2307, a novel arylamidine. Antimicrobial agents and chemotherapy, 52(4), 1318-1324. https://doi.org/10.1128/AAC.01159-07

296. Molyneux, D. (2013). Neglected tropical diseases. Community eye health, 26(82), 21-24. https://www.ncbi.nlm.nih.gov/pubmed/24023397

297. Moran, M. G., J.; Chapman, N.; Abela-Oversteegen, L.; Whittall, C.; Howard, R.; Farrell, P.; Halliday, D.; Hirst, C. (2014). G-FINDER 2014. Neglected Disease Research and Development: Emerging Trends (Policy Cures ed.). Policy Cures. https://policy-cures-website-assets.s3.ap-southeast-2.amazonaws.com/wp-content/uploads/2022/06/09014258/Y7-ND-G-FINDER-report-2014.pdf

298. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry,

19(14), 1639-1662. https://doi.org/https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

299. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem, 30(16), 2785-2791. https://doi.org/10.1002/jcc.21256

300. Mottram, J. C., Souza, A. E., Hutchison, J. E., Carter, R., Frame, M. J., & Coombs, G. H. (1996). Evidence from disruption of the lmcpb gene array of Leishmania mexicana that cysteine proteinases are virulence factors. Proceedings of the National Academy of Sciences of the United States of America, 93(12), 6008-6013. https://doi.org/10.1073/pnas.93.12.6008

301. Moura Barbosa, A. J., & Del Rio, A. (2012). Freely accessible databases of commercial compounds for high- throughput virtual screenings. Current topics in medicinal chemistry, 12(8), 866-877. https://doi.org/10.2174/156802612800166710

302. Mueser, T. C., Nossal, N. G., & Hyde, C. C. (1996). Structure of bacteriophage T4 RNase H, a 5' to 3' RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. Cell, 85(7), 1101-1112. https://doi.org/10.1016/s0092-8674(00)81310-0

303. Muftuoglu, M., Wong, H. K., Imam, S. Z., Wilson, D. M., Bohr, V. A., & Opresko, P. L. (2006). Telomere Repeat Binding Factor 2 Interacts with Base Excision Repair Proteins and Stimulates DNA Synthesis by DNA Polymerase β. Cancer Research, 66(1), 113-124. https://doi.org/10.1158/0008-5472.can-05-2742

304. Mukherjee, A., Boisvert, S., Monte-Neto, R. L., Coelho, A. C., Raymond, F., Mukhopadhyay, R., Corbeil, J., & Ouellette, M. (2013). Telomeric gene deletion and intrachromosomal amplification in antimony-resistant Leishmania. Mol Microbiol, 88(1), 189-202. https://doi.org/10.1111/mmi.12178

305. Mukherjee, A., Padmanabhan, P. K., Singh, S., Roy, G., Girard, I., Chatterjee, M., Ouellette, M., & Madhubala, R. (2007). Role of ABC transporter MRPA, gamma-glutamylcysteine synthetase and ornithine decarboxylase in natural antimony-resistant isolates of Leishmania donovani. J Antimicrob Chemother, 59(2), 204-211. https://doi.org/10.1093/jac/dkl494

306. Mukhopadhyay, R., Mukherjee, S., Mukherjee, B., Naskar, K., Mondal, D., Decuypere, S., Ostyn, B., Prajapati, V. K., Sundar, S., Dujardin, J. C., & Roy, S. (2011). Characterisation of antimony-resistant Leishmania donovani isolates: biochemical and biophysical studies and interaction with host cells. Int J Parasitol, 41(13-14), 1311-1321. https://doi.org/10.1016/j.ijpara.2011.07.013

307. Mullard, A. (2022). 2021 FDA approvals. Nat Rev Drug Discov, 21(2), 83-88. https://doi.org/10.1038/d41573-022-00001-9

308. Murray, C. W., Verdonk, M. L., & Rees, D. C. (2012). Experiences in fragment-based drug discovery. Trends in Pharmacological Sciences, 33(5), 224-232. https://doi.org/https://doi.org/10.1016/j.tips.2012.02.006

309. Murray, D., & Wigglesworth, M. (2017). Chapter 1 HTS Methods: Assay Design and Optimisation. In High Throughput Screening Methods: Evolution and Refinement (pp. 1-15). The Royal Society of Chemistry. https://doi.org/10.1039/9781782626770-00001

310. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr, 67(Pt 4), 355-367. https://doi.org/10.1107/S0907444911001314

311. Murta, S. M., Ropert, C., Alves, R. O., Gazzinelli, R. T., & Romanha, A. J. (1999). In-vivo treatment with benznidazole enhances phagocytosis, parasite destruction and cytokine release by macrophages during infection with a drug-susceptible but not with a derived drug-resistant Trypansoma cruzi population. Parasite Immunol, 21(10), 535-544. https://doi.org/10.1046/j.1365-3024.1999.00251.x

312. Nare, B., Hardy, L. W., & Beverley, S. M. (1997). The Roles of Pteridine Reductase 1 and Dihydrofolate Reductase-Thymidylate Synthase in Pteridine Metabolism in the Protozoan Parasite Leishmania major. Journal of Biological Chemistry, 272(21), 13883-13891. https://doi.org/10.1074/jbc.272.21.13883

313. Nelson, K. M., Dahlin, J. L., Bisson, J., Graham, J., Pauli, G. F., & Walters, M. A. (2017). The Essential Medicinal Chemistry of Curcumin. Journal of medicinal chemistry, 60(5), 1620-1637. https://doi.org/10.1021/acs.jmedchem.6b00975

314. Newman, G. (1895). Leprosy in the British Islands (Publication Number 2) University of Edinburgh].

https://era.ed.ac.uk/bitstream/handle/1842/33388/NewmanG_1895_v2redux.pdf?sequence=2&isAllowed=y

315. Nguyen, N. T., Nguyen, T. H., Pham, T. N. H., Huy, N. T., Bay, M. V., Pham, M. Q., Nam, P. C., Vu, V. V., & Ngo, S. T. (2020). Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. J Chem Inf Model, 60(1), 204-211. https://doi.org/10.1021/acs.jcim.9b00778

316. Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. J., & Kunkel, T. A. (2008). Division of labor at the eukaryotic replication fork. Molecular Cell, 30(2), 137-144. https://doi.org/10.1016/j.molcel.2008.02.022

317. Nielsen, K., Vedula, P., Smith, K. D., Meya, D. B., Garvey, E. P., Hoekstra, W. J., Schotzinger, R. J., & Boulware, D. R. (2017). Activity of VT-1129 against Cryptococcus neoformans clinical isolates with high fluconazole MICs. Med Mycol, 55(4), 453-456. https://doi.org/10.1093/mmy/myw089

318. Nikolova, T., Christmann, M., & Kaina, B. (2009). FEN1 is overexpressed in testis, lung and brain tumors. Anticancer Res, 29(7), 2453-2459. https://www.ncbi.nlm.nih.gov/pubmed/19596913

319. Nishikawa, H., Fukuda, Y., Mitsuyama, J., Tashiro, M., Tanaka, A., Takazono, T., Saijo, T., Yamamoto, K., Nakamura, S., Imamura, Y., Miyazaki, T., Kakeya, H., Yamamoto, Y., Yanagihara, K., Mukae, H., Kohno, S., & Izumikawa, K. (2017). In vitro and in vivo antifungal activities of T-2307, a novel arylamidine, against Cryptococcus gattii: an emerging fungal pathogen. J Antimicrob Chemother, 72(6), 1709-1713. https://doi.org/10.1093/jac/dkx020

320. Nisius, B., Sha, F., & Gohlke, H. (2012). Structure-based computational analysis of protein binding sites for function and druggability prediction. J Biotechnol, 159(3), 123-134. https://doi.org/10.1016/j.jbiotec.2011.12.005

321. Nouvellet, P., Dumonteil, E., & Gourbiere, S. (2013). The improbable transmission of Trypanosoma cruzi to human: the missing link in the dynamics and control of Chagas disease. PLoS Negl Trop Dis, 7(11), e2505. https://doi.org/10.1371/journal.pntd.0002505

322. Nugent, T., & Jones, D. T. (2009). Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics, 10(1), 159. https://doi.org/10.1186/1471-2105-10-159

323. Nwanochie, E., & Uversky, V. N. (2019). Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. Int J Mol Sci, 20(17). https://doi.org/10.3390/ijms20174186

324. O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. J Cheminform, 3, 33. https://doi.org/10.1186/1758-2946-3-33

325. O'Connor, C. D., & Timmis, K. N. (1987). Highly repressible expression system for cloning genes that specify potentially toxic proteins. Journal of bacteriology, 169(10), 4457-4462. https://doi.org/10.1128/jb.169.10.4457-4462.1987

326. Oates, S. (2016). Molecular Characterisation of the Trypanosoma brucei Flap endonuclease University of Sheffield]. https://etheses.whiterose.ac.uk/14274/

327. Opperdoes, F. R., & Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: The glycosome. FEBS Letters, 80(2), 360-364. https://doi.org/10.1016/0014-5793(77)80476-6

328. Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H., & Walter, J. C. (2006). Localization of MCM2-7, Cdc45, and GINS to the Site of DNA Unwinding during Eukaryotic DNA Replication. Molecular Cell, 21(4), 581-587. https://doi.org/https://doi.org/10.1016/j.molcel.2006.01.030

329. Pan, P., Vermelho, A. B., Capaci Rodrigues, G., Scozzafava, A., Tolvanen, M. E. E., Parkkila, S., Capasso, C., & Supuran, C. T. (2013). Cloning, Characterization, and Sulfonamide and Thiol Inhibition Studies of an α-Carbonic Anhydrase from Trypanosoma cruzi, the Causative Agent of Chagas Disease. Journal of medicinal chemistry, 56(4), 1761-1771. https://doi.org/10.1021/jm4000616

330. Panda, H., Jaiswal, A. S., Corsino, P. E., Armas, M. L., Law, B. K., & Narayan, S. (2009). Amino acid Asp181 of 5'-flap endonuclease 1 is a useful target for chemotherapeutic development. Biochemistry, 48(42), 9952-9958. https://doi.org/10.1021/bi9010754

331. Park, B. J., Lockhart, S. R., Brandt, M. E., & Chiller, T. M. (2011). Public Health Importance of Cryptococcal Disease: Epidemiology, Burden, and Control. In Cryptococcus (pp. 585-593). https://doi.org/https://doi.org/10.1128/9781555816858.ch43

332. Park, B. J., Wannemuehler, K. A., Marston, B. J., Govender, N., Pappas, P. G., & Chiller, T. M. (2009). Estimation of the current global burden of cryptococcal

meningitis among persons living with HIV/AIDS. AIDS, 23(4), 525-530. https://doi.org/10.1097/QAD.0b013e328322ffac

333. Patel, N., Atack, J. M., Finger, L. D., Exell, J. C., Thompson, P., Tsutakawa, S., Tainer, J. A., Williams, D. M., & Grasby, J. A. (2012). Flap endonucleases pass 5′-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5′-ends. Nucleic Acids Research, 40(10), 4507-4519. https://doi.org/10.1093/nar/gks051

334. Patra, S., Saravanan, P., Das, B., Subramanian, V., & Patra, S. (2020). Scaffold-based Screening and Molecular Dynamics Simulation Study to Identify Two Structurally Related Phenolic Compounds as Potent MMP1 Inhibitors. Comb Chem High Throughput Screen, 23(8), 757-774. https://doi.org/10.2174/1386207323666200428114216

335. Payne, D. J., Gwynn, M. N., Holmes, D. J., & Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov, 6(1), 29-40. https://doi.org/10.1038/nrd2201

336. Pedrique, B., Strub-Wourgaft, N., Some, C., Olliaro, P., Trouiller, P., Ford, N., Pecoul, B., & Bradol, J. H. (2013). The drug and vaccine landscape for neglected diseases (2000-11): a systematic assessment. Lancet Glob Health, 1(6), e371-379. https://doi.org/10.1016/S2214-109X(13)70078-0

337. Perez-Victoria, J. M., Bavchvarov, B. I., Torrecillas, I. R., Martinez-Garcia, M., Lopez-Martin, C., Campillo, M., Castanys, S., & Gamarro, F. (2011). Sitamaquine overcomes ABC-mediated resistance to miltefosine and antimony in Leishmania. Antimicrobial agents and chemotherapy, 55(8), 3838-3844. https://doi.org/10.1128/AAC.00065-11

338. Perfect, J. R., Dismukes, W. E., Dromer, F., Goldman, D. L., Graybill, J. R., Hamill, R. J., Harrison, T. S., Larsen, R. A., Lortholary, O., Nguyen, M. H., Pappas, P. G., Powderly, W. G., Singh, N., Sobel, J. D., & Sorrell, T. C. (2010). Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the infectious diseases society of america. Clin Infect Dis, 50(3), 291-322. https://doi.org/10.1086/649858

339. Pertusi, D. A., Moura, M. E., Jeffryes, J. G., Prabhu, S., Walters Biggs, B., & Tyo, K. E. J. (2017). Predicting novel substrates for enzymes with minimal

experimental effort with active learning. Metab Eng, 44, 171-181. https://doi.org/10.1016/j.ymben.2017.09.016

340. Petrone, P. M., Wassermann, A. M., Lounkine, E., Kutchukian, P., Simms, B., Jenkins, J., Selzer, P., & Glick, M. (2013). Biodiversity of small molecules--a new perspective in screening set selection. Drug Discov Today, 18(13-14), 674-680. https://doi.org/10.1016/j.drudis.2013.02.005

341. Petruševski, A. B. (2013). History of infectious diseases development in the Old and the Middle Ages with the emphasis on the plague and leprosy. Vojnosanit Pregl, 70(7), 704-708. https://doi.org/10.2298/vsp1307706p

342. Pfaller, M. A., Huband, M. D., Flamm, R. K., Bien, P. A., & Castanheira, M. (2019). In Vitro Activity of APX001A (Manogepix) and Comparator Agents against 1,706 Fungal Isolates Collected during an International Surveillance Program in 2017. Antimicrobial agents and chemotherapy, 63(8). https://doi.org/10.1128/AAC.00840-19

343. Ponce, I., Aldunate, C., Valenzuela, L., Sepulveda, S., Garrido, G., Kemmerling, U., Cabrera, G., & Galanti, N. (2017). A Flap Endonuclease (TcFEN1) Is Involved in Trypanosoma cruzi Cell Proliferation, DNA Repair, and Parasite Survival. J Cell Biochem, 118(7), 1722-1732. https://doi.org/10.1002/jcb.25830

344. Porter-Goff, M. E., & Rhind, N. (2009). The role of MRN in the S-phase DNA damage checkpoint is independent of its Ctp1-dependent roles in double-strand break repair and checkpoint signaling. Molecular biology of the cell, 20(7), 2096-2107. https://doi.org/10.1091/mbc.e08-09-0986

345. Prasad, R., Kumar, R., Jaiswal, B. P., & Singh, U. K. (2004). Miltefosine: an oral drug for visceral leishmaniasis. Indian J Pediatr, 71(2), 143-144. https://doi.org/10.1007/BF02723096

346. Purkait, B., Kumar, A., Nandi, N., Sardar, A. H., Das, S., Kumar, S., Pandey, K., Ravidas, V., Kumar, M., De, T., Singh, D., & Das, P. (2012). Mechanism of amphotericin B resistance in clinical isolates of Leishmania donovani. Antimicrobial agents and chemotherapy, 56(2), 1031-1041. https://doi.org/10.1128/AAC.00030-11

347. Pursell, Z. F., Isoz, I., Lundström, E.-B., Johansson, E., & Kunkel, T. A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication.

Science (New York, N.Y.), 317(5834), 127-130. https://doi.org/10.1126/science.1144067

348. Querol-Audi, J., Yan, C., Xu, X., Tsutakawa, S. E., Tsai, M. S., Tainer, J. A., Cooper, P. K., Nogales, E., & Ivanov, I. (2012). Repair complexes of FEN1 endonuclease, DNA, and Rad9-Hus1-Rad1 are distinguished from their PCNA counterparts by functionally important stability. Proceedings of the National Academy of Sciences of the United States of America, 109(22), 8528-8533. https://doi.org/10.1073/pnas.1121116109

349. Rai, S., Bhaskar, Goel, S. K., Nath Dwivedi, U., Sundar, S., & Goyal, N. (2013). Role of efflux pumps and intracellular thiols in natural antimony resistant isolates of Leishmania donovani. PLOS ONE, 8(9), e74862. https://doi.org/10.1371/journal.pone.0074862

350. Rajasingham, R., Smith, R. M., Park, B. J., Jarvis, J. N., Govender, N. P., Chiller, T. M., Denning, D. W., Loyse, A., & Boulware, D. R. (2017). Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. Lancet Infect Dis, 17(8), 873-881. https://doi.org/10.1016/S1473-3099(17)30243-8

351. Rao, H. G., Rosenfeld, A., & Wetmur, J. G. (1998). Methanococcus jannaschii flap endonuclease: expression, purification, and substrate requirements. Journal of bacteriology, 180(20), 5406-5412. https://www.ncbi.nlm.nih.gov/pubmed/9765572

352. Rarey, M., Kramer, B., & Lengauer, T. (1997). Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. J Comput Aided Mol Des, 11(4), 369-384. https://doi.org/10.1023/a:1007913026166

353. Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. J Mol Biol, 261(3), 470-489. https://doi.org/10.1006/jmbi.1996.0477

354. Ravu, R. R., Chen, Y. L., Jacob, M. R., Pan, X., Agarwal, A. K., Khan, S. I., Heitman, J., Clark, A. M., & Li, X. C. (2013). Synthesis and antifungal activities of miltefosine analogs. Bioorg Med Chem Lett, 23(17), 4828-4831. https://doi.org/10.1016/j.bmcl.2013.06.096

355. Reagan, M. S., Pittenger, C., Siede, W., & Friedberg, E. C. (1995). Characterization of a mutant strain of Saccharomyces cerevisiae with a deletion of the RAD27 gene, a structural homolog of the RAD2 nucleotide excision repair gene.

Journal of bacteriology, 177(2), 364-371. https://doi.org/10.1128/jb.177.2.364-371.1995

356. Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., & Sastry, G. N. (2007). Virtual Screening in Drug Discovery - A Computational Perspective. Current Protein & Peptide Science, 8(4), 329-351. https://doi.org/http://dx.doi.org/10.2174/138920307781369427

357. Remaut, E., Stanssens, P., & Fiers, W. (1981). Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. Gene, 15(1), 81-93. https://doi.org/10.1016/0378-1119(81)90106-2

358. Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res, 42(Web Server issue), W320-324. https://doi.org/10.1093/nar/gku316

359. Roberts, A. J., Torrie, L. S., Wyllie, S., & Fairlamb, A. H. (2014). Biochemical and genetic characterization of Trypanosoma cruzi N-myristoyltransferase. Biochem J, 459(2), 323-332. https://doi.org/10.1042/BJ20131033

360. Robertson, A. B., Klungland, A., Rognes, T., & Leiros, I. (2009). DNA Repair in Mammalian Cells [journal article]. Cellular and Molecular Life Sciences, 66(6), 981-993. https://doi.org/10.1007/s00018-009-8736-z

361. Robinson, S. L. (2022). Artificial intelligence for microbial biotechnology: beyond the hype. Microb Biotechnol, 15(1), 65-69. https://doi.org/10.1111/1751-7915.13943

362. Rodero, L., Mellado, E., Rodriguez, A. C., Salve, A., Guelfand, L., Cahn, P., Cuenca-Estrella, M., Davel, G., & Rodriguez-Tudela, J. L. (2003). G484S amino acid substitution in lanosterol 14-alpha demethylase (ERG11) is related to fluconazole resistance in a recurrent Cryptococcus neoformans clinical isolate. Antimicrobial agents and chemotherapy, 47(11), 3653-3656. https://doi.org/10.1128/AAC.47.11.3653-3656.2003

363. Romani, A., & Scarpa, A. (1992). Regulation of cell magnesium. Arch Biochem Biophys, 298(1), 1-12. https://doi.org/10.1016/0003-9861(92)90086-c

364. Romani, A. M. P. (2013). Magnesium in Eukaryotes. In R. H. Kretsinger, V. N. Uversky, & E. A. Permyakov (Eds.), Encyclopedia of Metalloproteins (pp. 1255-1264). Springer New York. https://doi.org/10.1007/978-1-4614-1533-6_260

365. Ronda, L., Bruno, S., Bettati, S., Storici, P., & Mozzarelli, A. (2015). From protein structure to function via single crystal optical spectroscopy. Front Mol Biosci, 2, 12. https://doi.org/10.3389/fmolb.2015.00012

366. Rothwell, D., & Hickson, I. (1997). Base Excision Repair of DNA Damage. In: Chapman and Hall, New York.

367. Roy, A., Yang, J., & Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res, 40(Web Server issue), W471-477. https://doi.org/10.1093/nar/gks372

368. Sagemark, J., Kraulis, P., & Weigelt, J. (2010). A software tool to accelerate design of protein constructs for recombinant expression. Protein Expr Purif, 72(2), 175-178. https://doi.org/10.1016/j.pep.2010.03.020

369. Saharia, A., Guittat, L., Crocker, S., Lim, A., Steffen, M., Kulkarni, S., & Stewart, S. A. (2008). Flap endonuclease 1 contributes to telomere stability. Current biology : CB, 18(7), 496-500. https://doi.org/10.1016/j.cub.2008.02.071

370. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., & Hakoshima, T. (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. EMBO J, 24(4), 683-693. https://doi.org/10.1038/sj.emboj.7600519

371. Sander, T., Freyss, J., von Korff, M., & Rufener, C. (2015). DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. Journal of Chemical Information and Modeling, 55(2), 460-473. https://doi.org/10.1021/ci500588j

372. Santacroce, L., Del Prete, R., Charitos, I. A., & Bottalico, L. (2021). Mycobacterium leprae: A historical study on the origins of leprosy and its social stigma. Infez Med, 29(4), 623-632. https://doi.org/10.53854/liim-2904-18

373. Sarilla, S., Habib, S. Y., Tollefsen, D. M., Friedman, D. B., Arnett, D. R., & Verhamme, I. M. (2010). Glycosaminoglycan-binding properties and kinetic characterization of human heparin cofactor II expressed in Escherichia coli. Anal Biochem, 406(2), 166-175. https://doi.org/10.1016/j.ab.2010.07.024

374. Sayers, J. R., & Eckstein, F. (1990). Properties of overexpressed phage T5 D15 exonuclease. Similarities with Escherichia coli DNA polymerase I 5'-3' exonuclease. The Journal of biological chemistry, 265(30), 18311-18317. https://www.ncbi.nlm.nih.gov/pubmed/2211703

375. Scarpa, A., & Brinley, F. J. (1981). In situ measurements of free cytosolic magnesium ions. Fed Proc, 40(12), 2646-2652. https://www.ncbi.nlm.nih.gov/pubmed/7286245

376. Schabert, V. F., Watson, C., Joseph, G. J., Iversen, P., Burudpakdee, C., & Harrison, D. J. (2013). Costs of tumor necrosis factor blockers per treated patient using real-world drug data in a managed care population. J Manag Care Pharm, 19(8), 621-630. https://doi.org/10.18553/jmcp.2013.19.8.621

377. Schaduangrat, N., Lampa, S., Simeon, S., Gleeson, M. P., Spjuth, O., & Nantasenamat, C. (2020). Towards reproducible computational drug discovery. J Cheminform, 12(1), 9. https://doi.org/10.1186/s13321-020-0408-x

378. Scharfstein, J., Schmitz, V., Morandi, V., Capella, M. M., Lima, A. P., Morrot, A., Juliano, L., & Müller-Esterl, W. (2000). Host cell invasion by Trypanosoma cruzi is potentiated by activation of bradykinin B(2) receptors. The Journal of experimental medicine, 192(9), 1289-1300. https://doi.org/10.1084/jem.192.9.1289

379. Schmidtke, P., & Barril, X. (2010). Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. Journal of medicinal chemistry, 53(15), 5858-5867. https://doi.org/10.1021/jm100574m

380. Scior, T., Bender, A., Tresadern, G., Medina-Franco, J. L., Martinez-Mayorga, K., Langer, T., Cuanalo-Contreras, K., & Agrafiotis, D. K. (2012). Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model, 52(4), 867-881. https://doi.org/10.1021/ci200528d

381. Secor, E. W., Le Bras, J., & Clain, J. (2015). Mechanisms of Resistance to Antiparasitic Agents. In Manual of Clinical Microbiology (11 ed., pp. 2550-2562). ASM Press.

382. Seffernick, J. T., & Lindert, S. (2020). Hybrid methods for combined experimental and computational determination of protein structure. J Chem Phys, 153(24), 240901. https://doi.org/10.1063/5.0026025

383. Setlow, P., & Kornberg, A. (1972). Deoxyribonucleic Acid Polymerase: Two Distinct Enzymes in One Polypeptide: II. A PROTEOLYTIC FRAGMENT CONTAINING THE 5' → 3' EXONUCLEASE FUNCTION. RESTORATION OF INTACT ENZYME FUNCTIONS FROM THE TWO PROTEOLYTIC FRAGMENTS. Journal of Biological Chemistry, 247(1), 232-240. http://www.jbc.org/content/247/1/232.abstract

384. Shah, S., Dunten, P., Stiteler, A., Park, C. K., & Horton, N. C. (2015). Structure and specificity of FEN-1 from Methanopyrus kandleri. Proteins, 83(1), 188-194. https://doi.org/10.1002/prot.24704

385. Sharma, S., Doherty, K. M., & Brosh, J. R., M. . (2005). DNA Helicases as Targets for Anti-Cancer Drugs. Current Medicinal Chemistry - Anti-Cancer Agents, 5(3), 183-199. https://doi.org/http://dx.doi.org/10.2174/1568011053765985

386. Sharma, S., Otterlei, M., Sommers, J. A., Driscoll, H. C., Dianov, G. L., Kao, H.-I., Bambara, R. A., & Brosh, R. M., Jr. (2004). WRN helicase and FEN-1 form a complex upon replication arrest and together process branchmigrating DNA structures associated with the replication fork. Molecular biology of the cell, 15(2), 734-750. https://doi.org/10.1091/mbc.e03-08-0567

387. Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. Cold Spring Harbor perspectives in medicine, 1(1), a006841. https://doi.org/10.1101/cshperspect.a006841

388. Shaw, K. J., Schell, W. A., Covel, J., Duboc, G., Giamberardino, C., Kapoor, M., Moloney, M., Soltow, Q. A., Tenor, J. L., Toffaletti, D. L., Trzoss, M., Webb, P., & Perfect, J. R. (2018). In Vitro and In Vivo Evaluation of APX001A/APX001 and Other Gwt1 Inhibitors against Cryptococcus. Antimicrobial agents and chemotherapy, 62(8). https://doi.org/10.1128/AAC.00523-18

389. Shaw, S. J., Finger, L. D., & Grasby, J. A. (2017). Human Exonuclease 1 Threads 5′-Flap Substrates through Its Helical Arch. Biochemistry, 56(29), 3704-3707. https://doi.org/10.1021/acs.biochem.7b00507

390. Sheng, C., Miao, Z., Ji, H., Yao, J., Wang, W., Che, X., Dong, G., Lu, J., Guo, W., & Zhang, W. (2009). Three-dimensional model of lanosterol 14 alpha-demethylase from Cryptococcus neoformans: active-site characterization and insights into azole binding. Antimicrobial agents and chemotherapy, 53(8), 3487-3495. https://doi.org/10.1128/AAC.01630-08

391. Shibata, T., Takahashi, T., Yamada, E., Kimura, A., Nishikawa, H., Hayakawa, H., Nomura, N., & Mitsuyama, J. (2012). T-2307 causes collapse of mitochondrial membrane potential in yeast. Antimicrobial agents and chemotherapy, 56(11), 5892-5897. https://doi.org/10.1128/AAC.05954-11

392. Shibata, Y., & Nakamura, T. (2002). Defective Flap Endonuclease 1 Activity in Mammalian Cells Is Associated with Impaired DNA Repair and Prolonged S Phase

Delay. Journal of Biological Chemistry, 277(1), 746-754. https://doi.org/10.1074/jbc.M109461200

393. Shiue, E., & Prather, K. L. J. (2012). Synthetic biology devices as tools for metabolic engineering. Biochemical Engineering Journal, 65, 82-89. https://doi.org/https://doi.org/10.1016/j.bej.2012.04.006

394. Sindermann, H., Croft, S. L., Engel, K. R., Bommer, W., Eibl, H. J., Unger, C., & Engel, J. (2004). Miltefosine (Impavido): the first oral treatment against leishmaniasis. Med Microbiol Immunol, 193(4), 173-180. https://doi.org/10.1007/s00430-003-0201-2

395. Singh, N., Chaput, L., & Villoutreix, B. O. (2021). Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. Brief Bioinform, 22(2), 1790-1818. https://doi.org/10.1093/bib/bbaa034

396. Sionov, E., Chang, Y. C., Garraffo, H. M., Dolan, M. A., Ghannoum, M. A., & Kwon-Chung, K. J. (2012). Identification of a Cryptococcus neoformans cytochrome P450 lanosterol 14alpha-demethylase (Erg11) residue critical for differential susceptibility between fluconazole/voriconazole and itraconazole/posaconazole. Antimicrobial agents and chemotherapy, 56(3), 1162-1169. https://doi.org/10.1128/AAC.05502-11

397. Sionov, E., Chang, Y. C., Garraffo, H. M., & Kwon-Chung, K. J. (2009). Heteroresistance to fluconazole in Cryptococcus neoformans is intrinsic and associated with virulence. Antimicrobial agents and chemotherapy, 53(7), 2804-2815. https://doi.org/10.1128/AAC.00295-09

398. Sionov, E., Lee, H., Chang, Y. C., & Kwon-Chung, K. J. (2010). Cryptococcus neoformans overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. PLoS pathogens, 6(4), e1000848. https://doi.org/10.1371/journal.ppat.1000848

399. Smyth, M. S., & Martin, J. H. (2000). x ray crystallography. Mol Pathol, 53(1), 8-14. https://doi.org/10.1136/mp.53.1.8

400. Sorrell, T. C., & Ellis, D. H. (1997). Ecology of Cryptococcus neoformans. Rev Iberoam Micol, 14(2), 42-43. https://www.ncbi.nlm.nih.gov/pubmed/16854168

401. Spadari, C. C., Vila, T., Rozental, S., & Ishida, K. (2018). Miltefosine Has a Postantifungal Effect and Induces Apoptosis in Cryptococcus Yeasts. Antimicrobial agents and chemotherapy, 62(8). https://doi.org/10.1128/AAC.00312-18

402. Spadari, C. C., Wirth, F., Lopes, L. B., & Ishida, K. (2020). New Approaches for Cryptococcosis Treatment. Microorganisms, 8(4). https://doi.org/10.3390/microorganisms8040613

403. Spigelman, M., & Rubini, M. (2016). Paleomicrobiology of Leprosy. Microbiol Spectr, 4(4). https://doi.org/10.1128/microbiolspec.PoH-0009-2015

404. Spiro, C., Pelletier, R., Rolfsmeier, M. L., Dixon, M. J., Lahue, R. S., Gupta, G., Park, M. S., Chen, X., Mariappan, S. V. S., & McMurray, C. T. (1999). Inhibition of FEN-1 Processing by DNA Secondary Structure at Trinucleotide Repeats. Molecular Cell, 4(6), 1079-1085. https://doi.org/https://doi.org/10.1016/S1097-2765(00)80236-1

405. Srivastava, M., Suri, C., Singh, M., Mathur, R., & Asthana, S. (2018). Molecular dynamics simulation reveals the possible druggable hot-spots of USP7. Oncotarget, 9(76), 34289-34305. https://doi.org/10.18632/oncotarget.26136

406. Starai, V. J., & Escalante-Semerena, J. C. (2004). Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in Salmonella enterica. J Mol Biol, 340(5), 1005-1012. https://doi.org/10.1016/j.jmb.2004.05.010

407. Stauch, B., Simon, B., Basile, T., Schneider, G., Malek, N. P., Kalesse, M., & Carlomagno, T. (2010). Elucidation of the structure and intermolecular interactions of a reversible cyclic-peptide inhibitor of the proteasome by NMR spectroscopy and molecular modeling. Angew Chem Int Ed Engl, 49(23), 3934-3938. https://doi.org/10.1002/anie.201000140

408. Stevanović, S., Perdih, A., Senćanski, M., Glišić, S., Duarte, M., Tomás, A. M., Sena, F. V., Sousa, F. M., Pereira, M. M., & Solmajer, T. (2018). In Silico Discovery of a Substituted 6-Methoxy-quinalidine with Leishmanicidal Activity in Leishmania infantum. Molecules (Basel, Switzerland), 23(4), 772. https://doi.org/10.3390/molecules23040772

409. Stevens, J. R., Noyes, H. A., Dover, G. A., & Gibson, W. C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, Trypanosoma brucei and T. cruzi. Parasitology, 118(1), 107-116. http://www.genomics.liv.ac.uk/tryps/papers/057STEVENS.pdf

410. Stewart, M. L., Burchmore, R. J., Clucas, C., Hertz-Fowler, C., Brooks, K., Tait, A., Macleod, A., Turner, C. M., De Koning, H. P., Wong, P. E., & Barrett, M. P.

(2010). Multiple genetic mechanisms lead to loss of functional TbAT1 expression in drug-resistant trypanosomes. Eukaryot Cell, 9(2), 336-343. https://doi.org/10.1128/EC.00200-09

411. Stodola, J. L., & Burgers, P. M. (2016). Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. Nat Struct Mol Biol, 23(5), 402-408. https://doi.org/10.1038/nsmb.3207

412. Strijbis, K., & Distel, B. (2010). Intracellular acetyl unit transport in fungal carbon metabolism. Eukaryot Cell, 9(12), 1809-1815. https://doi.org/10.1128/EC.00172-10

413. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif, 41(1), 207-234. https://doi.org/10.1016/j.pep.2005.01.016

414. Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol, 185, 60-89. https://doi.org/10.1016/0076-6879(90)85008-c

415. Studier, P. M. (1986). Use of Bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol., 189, 113-130. https://ci.nii.ac.jp/naid/80002931300/en/

416. Sukuru, S. C., Jenkins, J. L., Beckwith, R. E., Scheiber, J., Bender, A., Mikhailov, D., Davies, J. W., & Glick, M. (2009). Plate-based diversity selection based on empirical HTS data to enhance the number of hits and their chemical diversity. J Biomol Screen, 14(6), 690-699. https://doi.org/10.1177/1087057109335678

417. Sundar, S. (2015). Leishmaniasis. In Harrison's Principles of Internal Medicine (19 ed., pp. 1387-1394). McGraw-Hill Education.

418. Swinney, D. C., & Anthony, J. (2011). How were new medicines discovered? Nature Reviews Drug Discovery, 10, 507. https://doi.org/10.1038/nrd3480. https://www.nature.com/articles/nrd3480#supplementary-information

419. Symington, L. S. (1998). Homologous recombination is required for the viability of rad27 mutants. Nucleic Acids Res, 26(24), 5589-5595. https://doi.org/10.1093/nar/26.24.5589

420. Syrjänen, L., Vermelho, A. B., de Almeida Rodrigues, I., Corte-Real, S., Salonen, T., Pan, P., Vullo, D., Parkkila, S., Capasso, C., & Supuran, C. T. (2013). Cloning, Characterization, and Inhibition Studies of a β-Carbonic Anhydrase from

Leishmania donovani chagasi, the Protozoan Parasite Responsible for Leishmaniasis. Journal of medicinal chemistry, 56(18), 7372-7381. https://doi.org/10.1021/jm400939k

421. Syson, K., Tomlinson, C., Chapados, B. R., Sayers, J. R., Tainer, J. A., Williams, N. H., & Grasby, J. A. (2008). Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. The Journal of biological chemistry, 283(42), 28741-28746. https://doi.org/10.1074/jbc.M801264200

422. Szilágyi, K., Flachner, B., Hajdu, I., Szaszko, M., Dobi, K., Lorincz, Z., Cseh, S., & Dorman, G. (2021). Rapid Identification of Potential Drug Candidates from Multi-Million Compounds' Repositories. Combination of 2D Similarity Search with 3D Ligand/Structure Based Methods and In Vitro Screening. Molecules (Basel, Switzerland), 26(18). https://doi.org/10.3390/molecules26185593

423. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol, 38(7), 3022-3027. https://doi.org/10.1093/molbev/msab120

424. Taylor, G. L. (2010). Introduction to phasing. Acta Crystallogr D Biol Crystallogr, 66(Pt 4), 325-338. https://doi.org/10.1107/S0907444910006694

425. Tercero, J. A., & Diffley, J. F. X. (2001). Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. Nature, 412(6846), 553-557. https://doi.org/10.1038/35087607

426. Theuretzbacher, U., Ihle, F., & Derendorf, H. (2006). Pharmacokinetic/pharmacodynamic profile of voriconazole. Clin Pharmacokinet, 45(7), 649-663. https://doi.org/10.2165/00003088-200645070-00002

427. Torreele, E., Bourdin Trunz, B., Tweats, D., Kaiser, M., Brun, R., Mazué, G., Bray, M. A., & Pécoul, B. (2010). Fexinidazole – A New Oral Nitroimidazole Drug Candidate Entering Clinical Development for the Treatment of Sleeping Sickness. PLOS Neglected Tropical Diseases, 4(12), e923. https://doi.org/10.1371/journal.pntd.0000923

428. Totrov, M., & Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins, Suppl 1, 215-220. https://doi.org/10.1002/(sici)1097-0134(1997)1+<215::aid-prot29>3.3.co;2-i

429. Tran-Nguyen, V. K., Bret, G., & Rognan, D. (2021). True Accuracy of Fast Scoring Functions to Predict High-Throughput Screening Data from Docking

Poses: The Simpler the Better. J Chem Inf Model, 61(6), 2788-2797. https://doi.org/10.1021/acs.jcim.1c00292

430. Tripathi, K., Mor, V., Bairwa, N. K., Del Poeta, M., & Mohanty, B. K. (2012). Hydroxyurea treatment inhibits proliferation of Cryptococcus neoformans in mice. Front Microbiol, 3, 187. https://doi.org/10.3389/fmicb.2012.00187

431. Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Front Bioeng Biotechnol, 7, 420. https://doi.org/10.3389/fbioe.2019.00420

432. Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 31(2), 455-461. https://doi.org/10.1002/jcc.21334

433. Tsegay, K. B., Adeyemi, C. M., Gniffke, E. P., Sather, D. N., Walker, J. K., & Smith, S. E. P. (2021). A repurposed drug screen identifies compounds that inhibit the binding of the COVID-19 spike protein to ACE2. bioRxiv. https://doi.org/10.1101/2021.04.08.439071

434. Tsukahara, K., Hata, K., Nakamoto, K., Sagane, K., Watanabe, N. A., Kuromitsu, J., Kai, J., Tsuchiya, M., Ohba, F., Jigami, Y., Yoshimatsu, K., & Nagasu, T. (2003). Medicinal genetics approach towards identifying the molecular target of a novel inhibitor of fungal cell wall assembly. Mol Microbiol, 48(4), 1029-1042. https://doi.org/10.1046/j.1365-2958.2003.03481.x

435. Tsutakawa, Susan E., Classen, S., Chapados, Brian R., Arvai, A. S., Finger, L. D., Guenther, G., Tomlinson, Christopher G., Thompson, P., Sarker, Altaf H., Shen, B., Cooper, Priscilla K., Grasby, Jane A., & Tainer, John A. (2011). Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. Cell, 145(2), 198-211. https://doi.org/10.1016/j.cell.2011.03.004

436. Tsutakawa, S. E., Thompson, M. J., Arvai, A. S., Neil, A. J., Shaw, S. J., Algasaier, S. I., Kim, J. C., Finger, L. D., Jardine, E., Gotham, V. J. B., Sarker, A. H., Her, M. Z., Rashid, F., Hamdan, S. M., Mirkin, S. M., Grasby, J. A., & Tainer, J. A. (2017). Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. Nat Commun, 8, 15855. https://doi.org/10.1038/ncomms15855

437. Tulchinsky, T. H., & Varavikova, E. A. (2014). Chapter 4 - Communicable Diseases. In T. H. Tulchinsky & E. A. Varavikova (Eds.), The New Public Health (Third Edition) (pp. 149-236). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-415766-8.00004-5

438. Tumey, L. N., Bom, D., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., Murphy, S., Dent, T., Leventhal, C., Bailey, A., Harrington, J., & Bennani, Y. L. (2005). The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. Bioorganic & Medicinal Chemistry Letters, 15(2), 277-281. https://doi.org/10.1016/j.bmcl.2004.10.086

439. Tumey, L. N., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., Murphy, S., Bailey, A., Dent, T., Leventhal, C., Harrington, J., & Bennani, Y. L. (2004). The identification and optimization of 2,4-diketobutyric acids as flap endonuclease 1 inhibitors. Bioorg Med Chem Lett, 14(19), 4915-4918. https://doi.org/10.1016/j.bmcl.2004.07.028

440. Umemura, M., Okamoto, M., Nakayama, K., Sagane, K., Tsukahara, K., Hata, K., & Jigami, Y. (2003). GWT1 gene is required for inositol acylation of glycosylphosphatidylinositol anchors in yeast. The Journal of biological chemistry, 278(26), 23639-23647. https://doi.org/10.1074/jbc.M301044200

441. Uniting to Combat, N. (2012). Neglected tropical diseases burden map (2009/2010). In.

442. Uson, M. L., Carl, A., Goldgur, Y., & Shuman, S. (2018). Crystal structure and mutational analysis of Mycobacterium smegmatis FenA highlight active site amino acids and three metal ions essential for flap endonuclease and 5' exonuclease activities. Nucleic Acids Res, 46(8), 4164-4175. https://doi.org/10.1093/nar/gky238

443. Uzcanga, G., Lara, E., Gutiérrez, F., Beaty, D., Beske, T., Teran, R., Navarro, J.-C., Pasero, P., Benítez, W., & Poveda, A. (2017). Nuclear DNA replication and repair in parasites of the genus Leishmania: Exploiting differences to develop innovative therapeutic approaches. Critical Reviews in Microbiology, 43(2), 156-177. https://doi.org/10.1080/1040841X.2016.1188758

444. Van der Auwera, G., & Dujardin, J.-C. (2015). Species Typing in Dermal Leishmaniasis. Clinical Microbiology Reviews, 28(2), 265. https://doi.org/10.1128/CMR.00104-14

445. van Pel, D. M., Barrett, I. J., Shimizu, Y., Sajesh, B. V., Guppy, B. J., Pfeifer, T., McManus, K. J., & Hieter, P. (2013). An evolutionarily conserved synthetic lethal interaction network identifies FEN1 as a broad-spectrum target for anticancer therapeutic development. PLoS Genet, 9(1), e1003254. https://doi.org/10.1371/journal.pgen.1003254

446. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., . . . Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res, 50(D1), D439-D444. https://doi.org/10.1093/nar/gkab1061

447. Vecchiarelli, A., Pericolini, E., Gabrielli, E., Kenno, S., Perito, S., Cenci, E., & Monari, C. (2013). Elucidating the immunological function of the Cryptococcus neoformans capsule. Future Microbiol, 8(9), 1107-1116. https://doi.org/10.2217/fmb.13.84

448. Vechi, H. T., Theodoro, R. C., de Oliveira, A. L., Gomes, R., Soares, R. D. A., Freire, M. G., & Bay, M. B. (2019). Invasive fungal infection by Cryptococcus neoformans var. grubii with bone marrow and meningeal involvement in a HIV-infected patient: a case report. BMC Infect Dis, 19(1), 220. https://doi.org/10.1186/s12879-019-3831-8

449. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. Proteins, 52(4), 609-623. https://doi.org/10.1002/prot.10465

450. Verdun, R. E., & Karlseder, J. (2006). The DNA Damage Machinery and Homologous Recombination Pathway Act Consecutively to Protect Human Telomeres. Cell, 127(4), 709-720. https://doi.org/10.1016/j.cell.2006.09.034

451. Vermes, A., Guchelaar, H. J., & Dankert, J. (2000). Flucytosine: a review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions. J Antimicrob Chemother, 46(2), 171-179. https://doi.org/10.1093/jac/46.2.171

452. Vincent, I. M., Creek, D., Watson, D. G., Kamleh, M. A., Woods, D. J., Wong, P. E., Burchmore, R. J., & Barrett, M. P. (2010). A molecular mechanism for eflornithine resistance in African trypanosomes. PLoS pathogens, 6(11), e1001204. https://doi.org/10.1371/journal.ppat.1001204

453. Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics, 28(15), 2074-2075. https://doi.org/10.1093/bioinformatics/bts310

454. Volpedo, G., Huston, R. H., Holcomb, E. A., Pacheco-Fernandez, T., Gannavaram, S., Bhattacharya, P., Nakhasi, H. L., & Satoskar, A. R. (2021). From infection to vaccination: reviewing the global burden, history of vaccine development, and recurring challenges in global leishmaniasis protection. Expert Rev Vaccines, 20(11), 1431-1446. https://doi.org/10.1080/14760584.2021.1969231

455. von Stosch, M., Portela, R. M. C., & Varsakelis, C. (2021). A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice. Current Opinion in Chemical Engineering, 33, 100692. https://doi.org/https://doi.org/10.1016/j.coche.2021.100692

456. Waga, S., & Stillman, B. (1994). Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. Nature, 369(6477), 207-212. https://doi.org/10.1038/369207a0

457. Walter, J., & Newport, J. (2000). Initiation of Eukaryotic DNA Replication: Origin Unwinding and Sequential Chromatin Association of Cdc45, RPA, and DNA Polymerase α. Molecular Cell, 5(4), 617-627. https://doi.org/https://doi.org/10.1016/S1097-2765(00)80241-5

458. Ward, T. A., McHugh, P. J., & Durant, S. T. (2017). Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. PLOS ONE, 12(6), e0179278. https://doi.org/10.1371/journal.pone.0179278

459. Wells, J. A., & McClendon, C. L. (2007). Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature, 450(7172), 1001-1009. https://doi.org/10.1038/nature06526

460. Wiederhold, N. P., Patterson, H. P., Tran, B. H., Yates, C. M., Schotzinger, R. J., & Garvey, E. P. (2018). Fungal-specific Cyp51 inhibitor VT-1598 demonstrates in vitro activity against Candida and Cryptococcus species, endemic fungi, including Coccidioides species, Aspergillus species and Rhizopus arrhizus. J Antimicrob Chemother, 73(2), 404-408. https://doi.org/10.1093/jac/dkx410

461. 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. Proceedings of the National Academy of Sciences, 105(13), 5022. https://doi.org/10.1073/pnas.0711014105

462. Willems, H., De Cesco, S., & Svensson, F. (2020). Computational Chemistry on a Budget: Supporting Drug Discovery with Limited Resources. Journal of medicinal chemistry, 63(18), 10158-10169. https://doi.org/10.1021/acs.jmedchem.9b02126

463. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr, 67(Pt 4), 235-242. https://doi.org/10.1107/S0907444910045749

464. World Health, O. (2018). Guidelines for the diagnosis, prevention and management of cryptococcal disease in HIV-infected adults, adolescents and children (WHO Press ed.) https://www.who.int/publications/i/item/9789241550277

465. World Health, O. (2020). Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021-2030 https://www.who.int/publications/i/item/9789240010352

466. Wu, F., Cheng, L., Yu, Q., Zhang, L., Li, H., & Wang, C. (2018). Purification and Functional Characterization of the C-Terminal Domain of the beta-Actin-Binding Protein AIM1 In Vitro. Molecules (Basel, Switzerland), 23(12). https://doi.org/10.3390/molecules23123281

467. Wunderink, R. G., Giamarellos-Bourboulis, E. J., Rahav, G., Mathers, A. J., Bassetti, M., Vazquez, J., Cornely, O. A., Solomkin, J., Bhowmick, T., Bishara, J., Daikos, G. L., Felton, T., Furst, M. J. L., Kwak, E. J., Menichetti, F., Oren, I., Alexander, E. L., Griffith, D., Lomovskaya, O., . . . Kaye, K. S. (2018). Effect and Safety of Meropenem-Vaborbactam versus Best-Available Therapy in Patients with Carbapenem-Resistant Enterobacteriaceae Infections: The TANGO II Randomized Clinical Trial. Infect Dis Ther, 7(4), 439-455. https://doi.org/10.1007/s40121-018-0214-1

468. Würz, J. M., Kazemi, S., Schmidt, E., Bagaria, A., & Guntert, P. (2017). NMR-based automated protein structure determination. Arch Biochem Biophys, 628, 24-32. https://doi.org/10.1016/j.abb.2017.02.011

469. Xin, X., Wen, T., Gong, L. B., Deng, M. M., Hou, K. Z., Xu, L., Shi, S., Qu, X. J., Liu, Y. P., Che, X. F., & Teng, Y. E. (2020). Inhibition of FEN1 Increases Arsenic Trioxide-Induced ROS Accumulation and Cell Death: Novel Therapeutic Potential for Triple Negative Breast Cancer. Front Oncol, 10, 425. https://doi.org/10.3389/fonc.2020.00425

470. Xu, H., Shi, R., Han, W., Cheng, J., Xu, X., Cheng, K., Wang, L., Tian, B., Zheng, L., Shen, B., Hua, Y., & Zhao, Y. (2018). Structural basis of 5' flap recognition and protein-protein interactions of human flap endonuclease 1. Nucleic Acids Research, 46(21), 11315-11325. https://doi.org/10.1093/nar/gky911

471. Xu, J., Onyewu, C., Yoell, H. J., Ali, R. Y., Vilgalys, R. J., & Mitchell, T. G. (2001). Dynamic and heterogeneous mutations to fluconazole resistance in Cryptococcus neoformans. Antimicrobial agents and chemotherapy, 45(2), 420-427. https://doi.org/10.1128/AAC.45.2.420-427.2001

472. Xu, L., Qu, J. L., Song, N., Zhang, L. Y., Zeng, X., Che, X. F., Hou, K. Z., Shi, S., Feng, Z. Y., Qu, X. J., Liu, Y. P., & Teng, Y. E. (2020). Biological and clinical significance of flap endonuclease-1 in triple-negative breast cancer: Support of metastasis and a poor prognosis. Oncol Rep, 44(6), 2443-2454. https://doi.org/10.3892/or.2020.7812

473. Yan, X., C., L., Liu, Z., Hagler, A. T., Gu, Q., & Xu, J. (2016). Chemical Structure Similarity Search for Ligand-based Virtual Screening: Methods and Computational Resources. Current Drug Targets, 17(14), 1580-1585. https://doi.org/http://dx.doi.org/10.2174/1389450116666151102095555

474. Yang, F., Hu, Z., & Guo, Z. (2022). Small-Molecule Inhibitors Targeting FEN1 for Cancer Therapy. Biomolecules, 12(7). https://doi.org/10.3390/biom12071007

475. Yang, X., Pistolozzi, M., & Lin, Z. (2018). New trends in aggregating tags for therapeutic protein purification. Biotechnol Lett, 40(5), 745-753. https://doi.org/10.1007/s10529-018-2543-2

476. Yoshinaga, T., Kobayashi, M., Seki, T., Miki, S., Wakasa-Morimoto, C., Suyama-Kagitani, A., Kawauchi-Miki, S., Taishi, T., Kawasuji, T., Johns, B. A., Underwood, M. R., Garvey, E. P., Sato, A., & Fujiwara, T. (2015). Antiviral

characteristics of GSK1265744, an HIV integrase inhibitor dosed orally or by long-acting injection. Antimicrobial agents and chemotherapy, 59(1), 397-406. https://doi.org/10.1128/AAC.03909-14

477. Younis, B. M., Osman, M., Khalil, E. A. G., Santoro, F., Furini, S., Wiggins, R., Keding, A., Carraro, M., Musa, A. E. A., Abdarahaman, M. A. A., Mandefield, L., Bland, M., Aebischer, T., Gabe, R., Layton, A. M., Lacey, C. J. N., Kaye, P. M., & Musa, A. M. (2021). Safety and immunogenicity of ChAd63-KH vaccine in post-kala-azar dermal leishmaniasis patients in Sudan. Mol Ther, 29(7), 2366-2377. https://doi.org/10.1016/j.ymthe.2021.03.020

478. Yu, J., Zhou, Y., Tanaka, I., & Yao, M. (2010). Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics, 26(1), 46-52. https://doi.org/10.1093/bioinformatics/btp599

479. Zhang, X. D., Wang, D., Sun, S., & Zhang, H. (2020). Issues Of Z-factor and an approach to avoid them for quality control in high-throughput screening studies. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa1049

480. Zhang, Y., Cooper, B., Gui, X., Sherer, R., & Cao, Q. (2019). Clinical diversity of invasive cryptococcosis in AIDS patients from central China: report of two cases with review of literature. BMC Infect Dis, 19(1), 1065. https://doi.org/10.1186/s12879-019-4698-4

481. Zhao, M., Lepak, A. J., Marchillo, K., Vanhecker, J., Sanchez, H., Ambrose, P. G., & Andes, D. R. (2019). APX001 Pharmacokinetic/Pharmacodynamic Target Determination against Aspergillus fumigatus in an In Vivo Model of Invasive Pulmonary Aspergillosis. Antimicrobial agents and chemotherapy, 63(4). https://doi.org/10.1128/AAC.02372-18

482. Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Muller, P., Sheldrick, G. M., & Minor, W. (2014). Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Nat Protoc, 9(1), 156-170. https://doi.org/10.1038/nprot.2013.172

483. Zheng, L., Jia, J., Finger, L. D., Guo, Z., Zer, C., & Shen, B. (2011). Functional regulation of FEN1 nuclease and its link to cancer. Nucleic Acids Res, 39(3), 781-794. https://doi.org/10.1093/nar/gkq884

484. Zheng, L., Zhou, M., Chai, Q., Parrish, J., Xue, D., Patrick, S. M., Turchi, J. J., Yannone, S. M., Chen, D., & Shen, B. (2005). Novel function of the flap

endonuclease 1 complex in processing stalled DNA replication forks. EMBO reports, 6(1), 83-89. https://doi.org/10.1038/sj.embor.7400313

485. Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol, 430(15), 2237-2243. https://doi.org/10.1016/j.jmb.2017.12.007

486. Zoete, V., Grosdidier, A., & Michielin, O. (2009). Docking, virtual high throughput screening and in silico fragment-based drug design. J Cell Mol Med, 13(2), 238-248. https://doi.org/10.1111/j.1582-4934.2008.00665.x

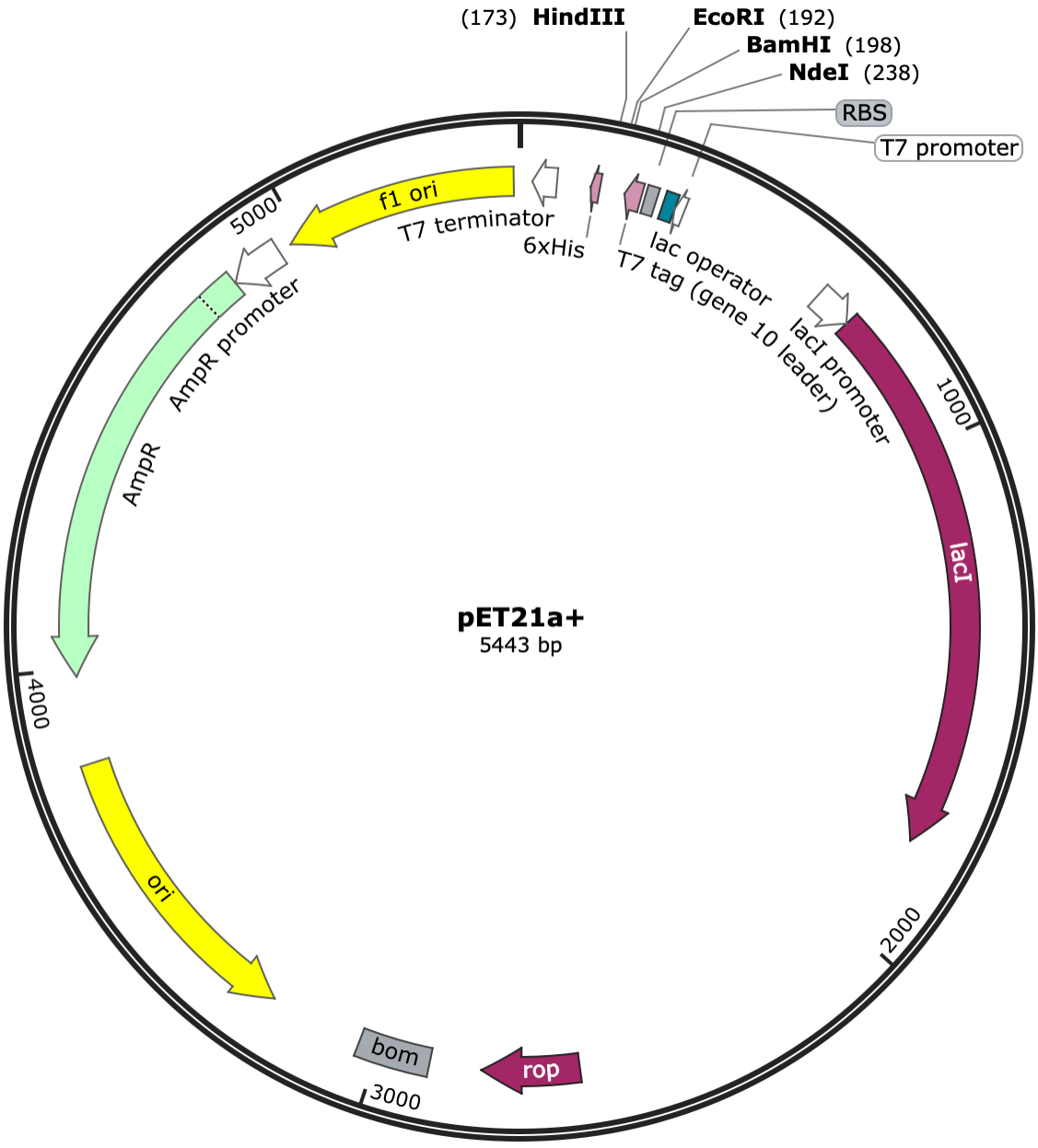


Figure 9.1 Vector map of pET21a+

Vector map was created using SnapGene® Viewer 6.1.2.

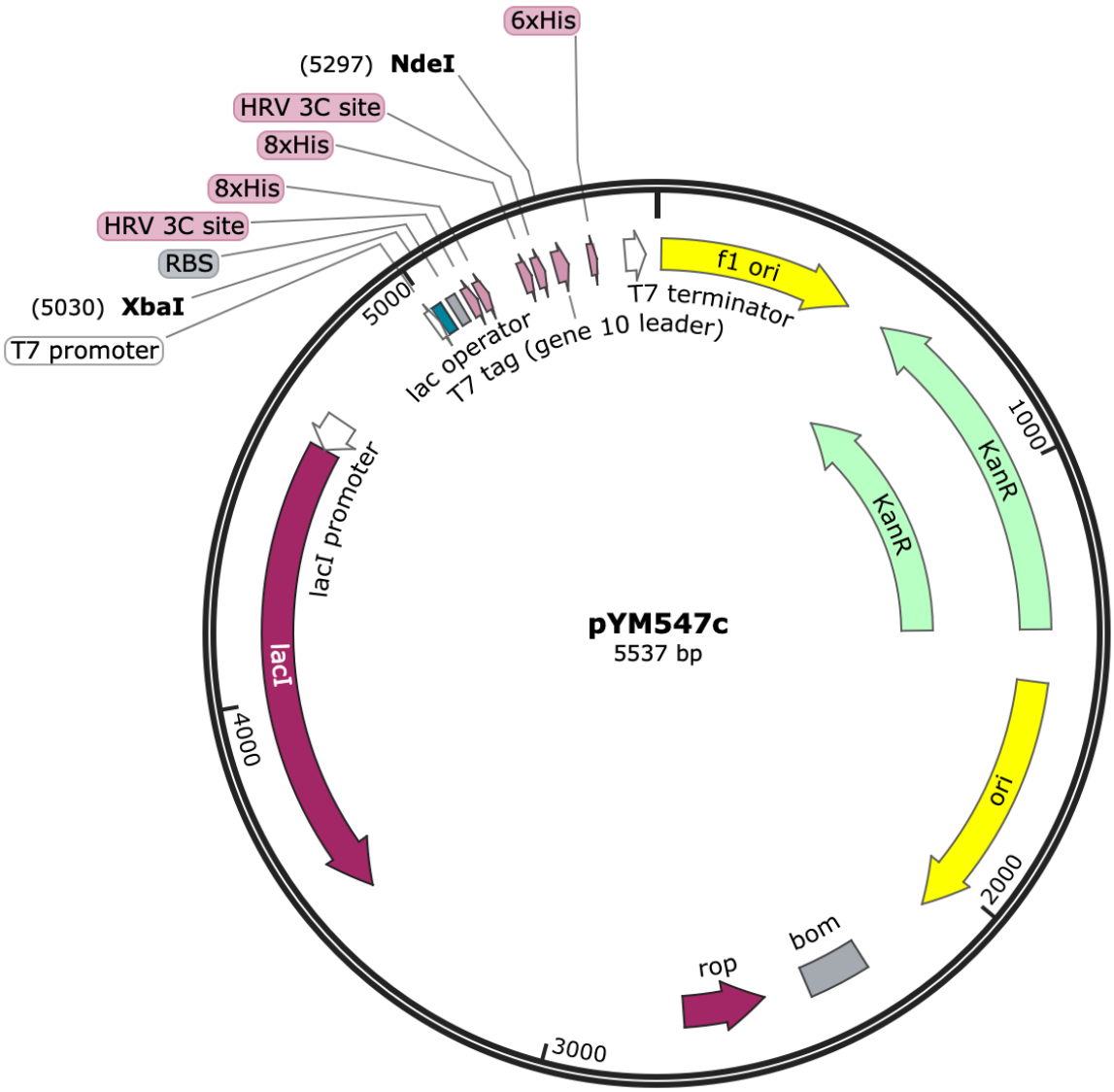


Figure 9.2 Vector map of pYM547c

Vector map was created using SnapGene® Viewer 6.1.2.

Table 9.1 Physicochemical properties of WT and mutants, FL or truncated forms, of *Hs*FEN, *Tb*FEN, *Li*FEN, *Cn*FEN

Proteins		Length	Theoretical pI	Mw (Da)	Ext. coefficient ($m^{-1}cm^{-1}$)	Abs 0.1% (=1g/L)[2]	A_{280} absorbers		
							Tyrosines (Y)	Tryptophans (W)	Cysteines (C)
***Hs*FEN**	WT	380	8.80	42592.98	22920	0.538	8	2	6
***Tb*FEN**	WT	393	7.66	44362.98	25900	0.584	10	2	3
***Tb*FEN**	D183K FL	393	8.49	44244.87	25900	0.585	10	2	3
***Tb*FEN**	D183K ΔC	341	5.56	38654.43	25900	0.670	10	2	3
***Li*FEN**	WT	395	9.12	44201.05	22920	0.519	8	2	3
***Cn*FEN**	WT	453	6.56	50022.50	25440	0.509	6	3	8
***Cn*FEN FL**	His-tag	511	6.80	55731.45	25440	0.456	6	3	8
	Cleaved His-tag	462	6.36	50983.58	25440	0.499	6	3	8
***Cn*FEN ΔC-414**	His-tag	472	6.01	51623.68	25440	0.493	6	3	8
	Cleaved His-tag	423	5.36	46875.81	25440	0.543	6	3	8
***Cn*FEN ΔC-405**	His-tag	463	6.01	50574.53	25440	0.503	6	3	8
	Cleaved His-tag	414	5.36	45826.66	25440	0.555	6	3	8

[1] All parameters were analysed using the ExPASy protein parameters (ProtParam) tool, based on the amino acid sequences (Gasteiger et al., 2005).

[2] Abs 0.1% was calculated assuming that formation of disulphide bonds between the thiol groups of cysteine residues formed are red

Table 9.2 Protein mass estimated using Mass spectrometry for the *Li*FEN protein

Protein	Predicted Mass (Da)[1]	Precise Mass (Da)[2]
Wild type Full length LiFEN	43941.69	43942.1

[1] Predicted mass analysed with ExPASy protein parameters (ProtParam) tool, based on the aminoacid sequence of the protein.
[2] Precise mass estimated by spectrometry.

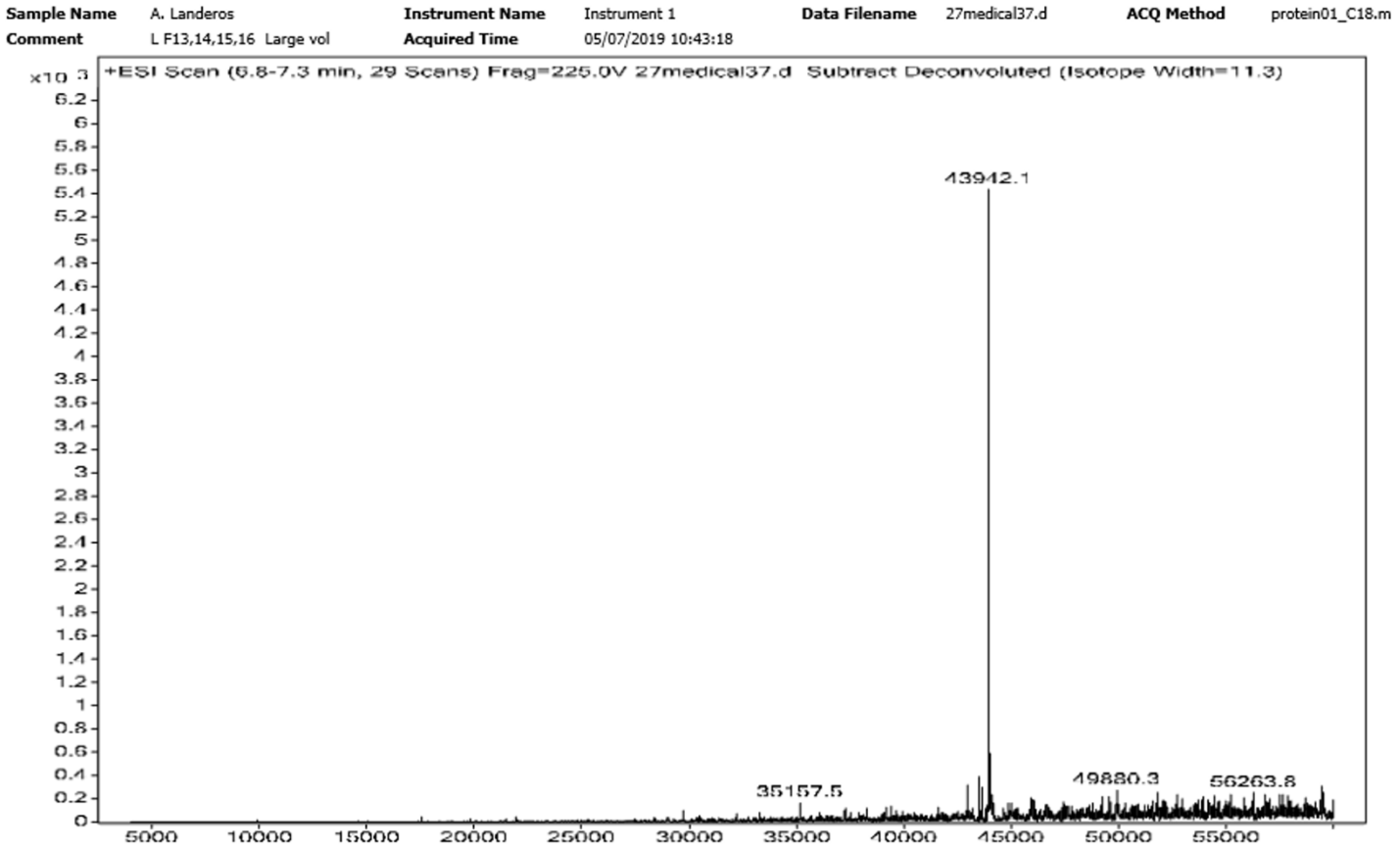


Figure 9.3 Mass spectrum of purified *Li*FEN (Section 2.3.3, 2.3.4 & 2.6.3)

Table 9.3 Protein mass estimated using Mass spectrometry for the *Cn*FEN-WT protein

Protein	Predicted Mass (Da)[1]	Precise Mass (Da)[2]
Wild type Full length CnFEN	50022.50	50022.46

[1] Predicted mass analysed with ExPASy protein parameters (ProtParam) tool, based on the aminoacid sequence of the protein.
[2] Precise mass estimated by spectrometry.

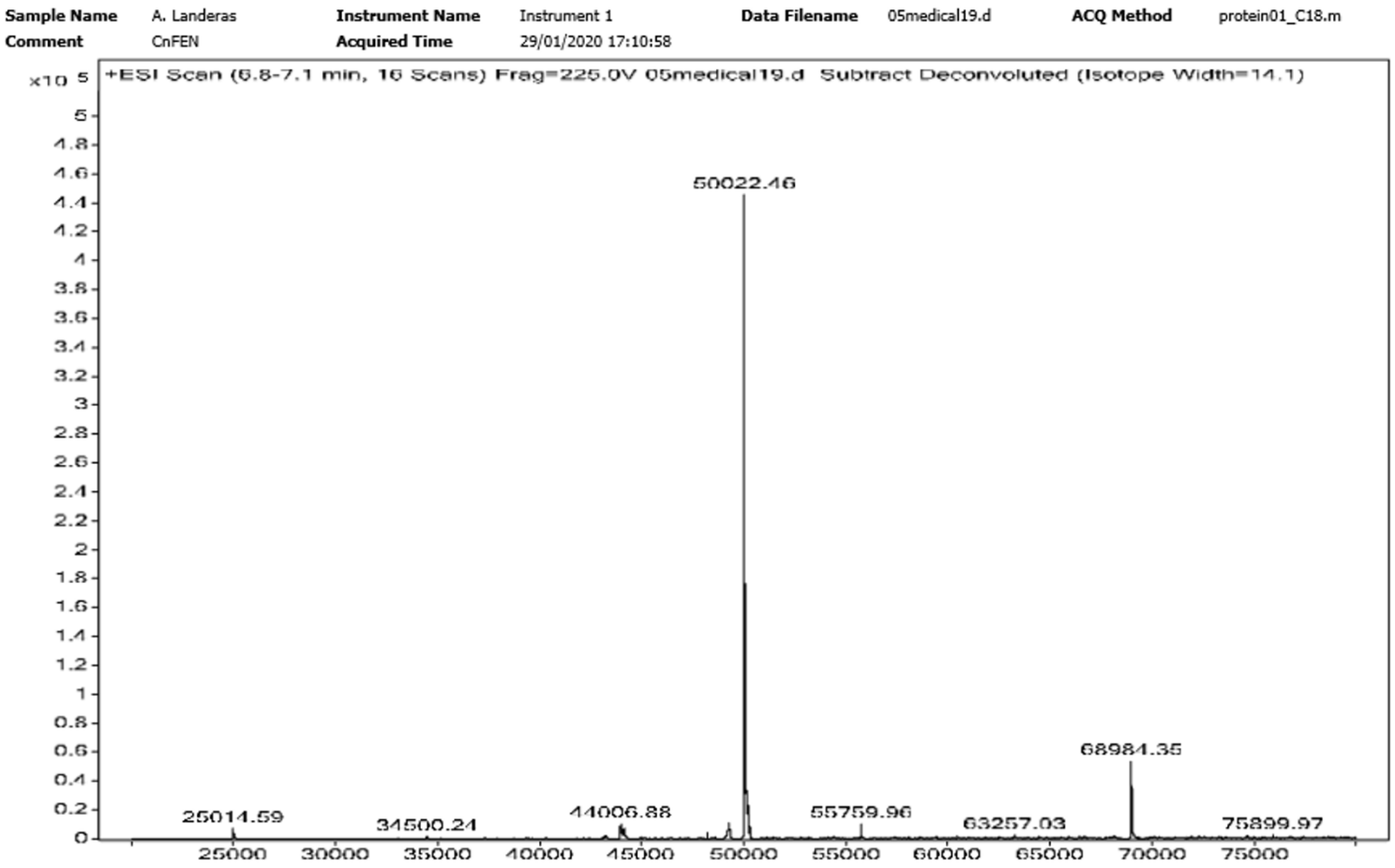


Figure 9.4 Mass spectrum of purified *Cn*FEN-WT (Section 2.3.3, 2.3.4 & 2.6.3)

Table 9.4 Protein mass estimated using Mass spectrometry for the *Cn*FEN-FL cleaved His-tag protein

Protein	Predicted Mass (Da)[1]	Precise Mass (Da)[2]
Full length cleaved His-tag CnFEN	50983.58	50984.90

[1] Predicted mass analysed with ExPASy protein parameters (ProtParam) tool, based on the aminoacid sequence of the protein.
[2] Precise mass estimated by spectrometry.

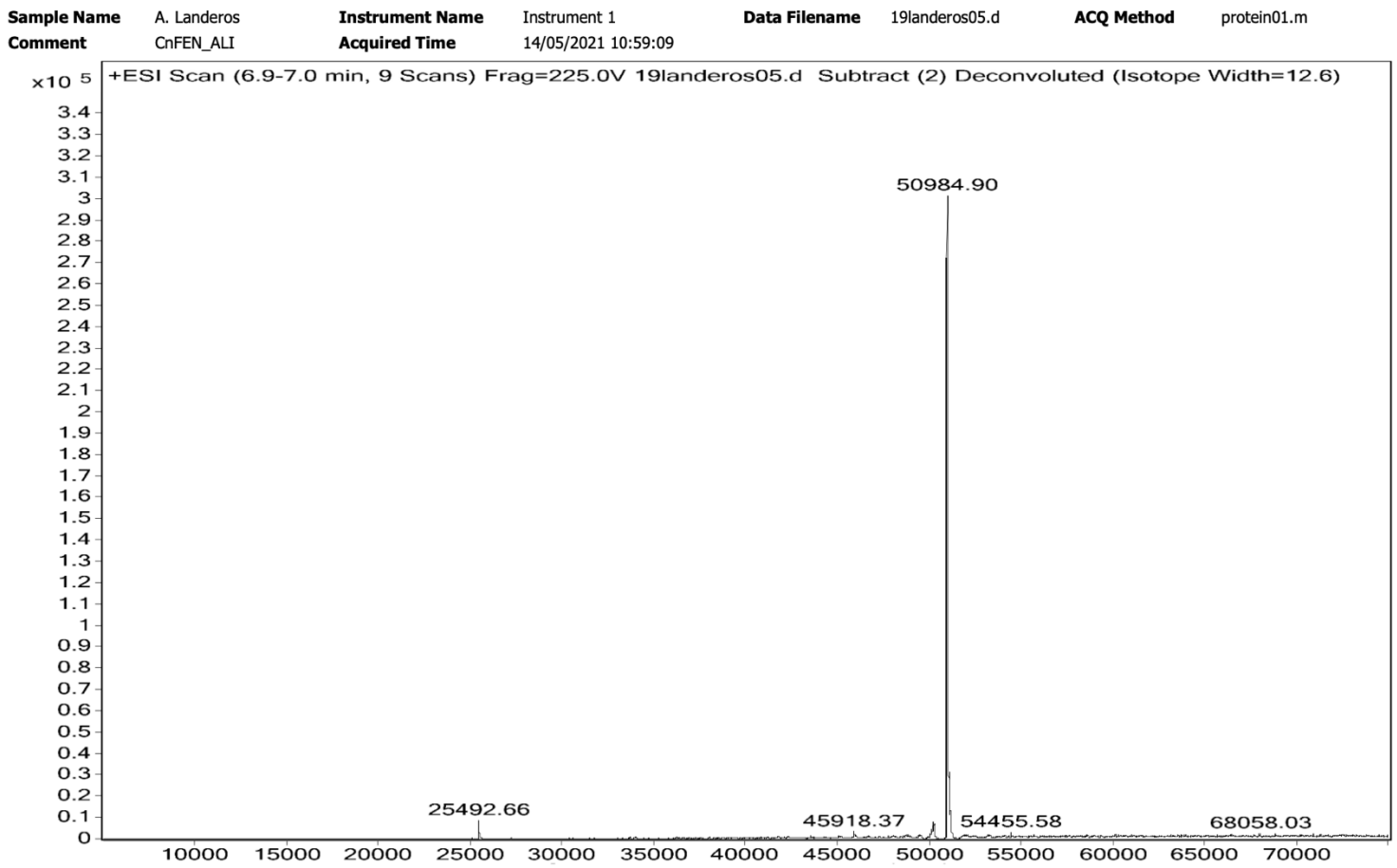


Figure 9.5 Mass spectrum of purified *Cn*FEN-FL cleaved His-tag (Section 2.3.3, 2.3.4 & 2.6.3)

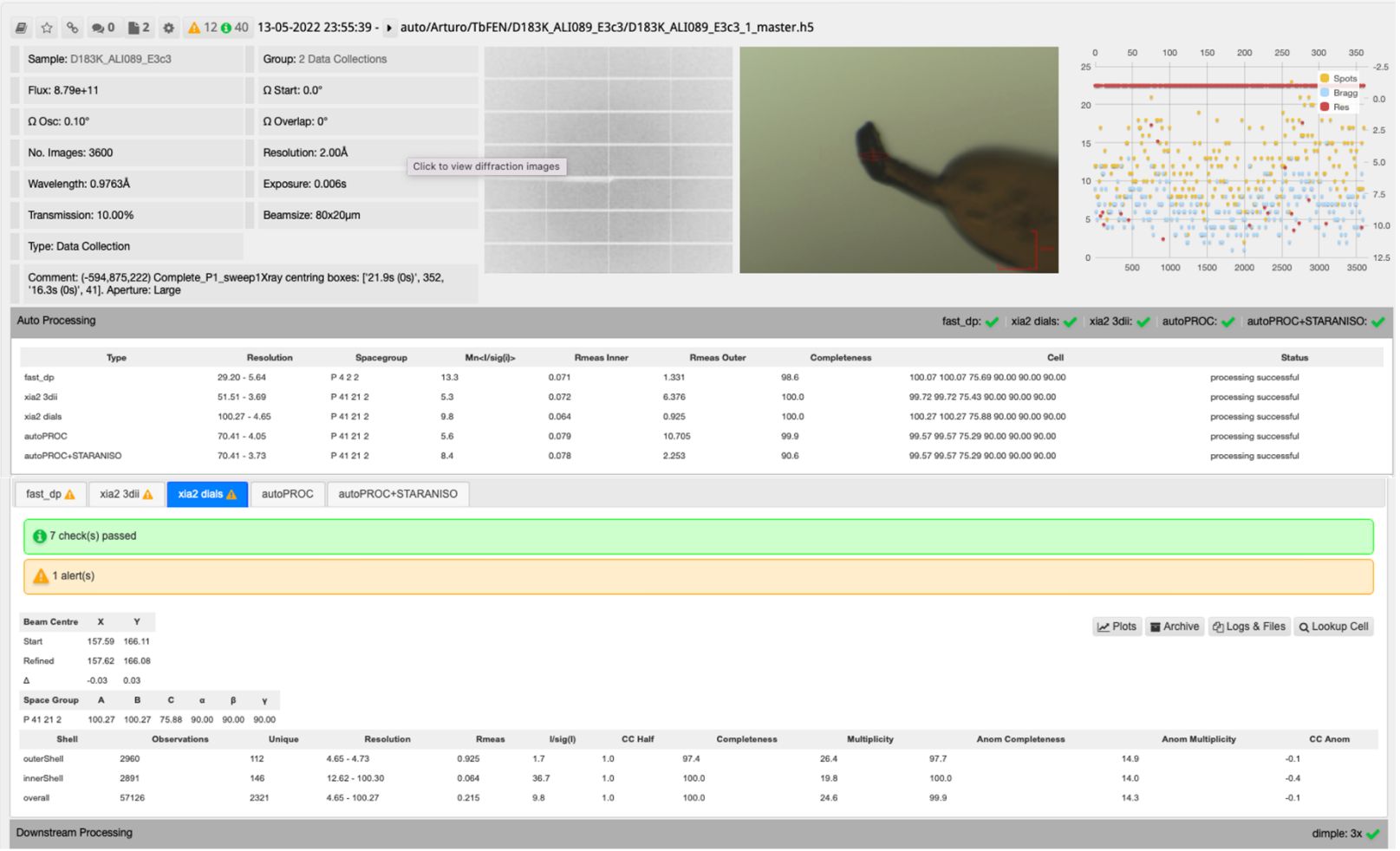


Figure 9.6 X-ray diffraction data set for *Tb*FEN-D183K ΔC-341 crystal E3c

Resolution of the X-ray diffraction of 4.65 Å according to xia2 DIALS. Crystal growing condition: 0.05 M $MgCl_2$, 0.2 M Citric acid pH 6, 18% (w/v) PEG 3350. DNA substrate: JT.

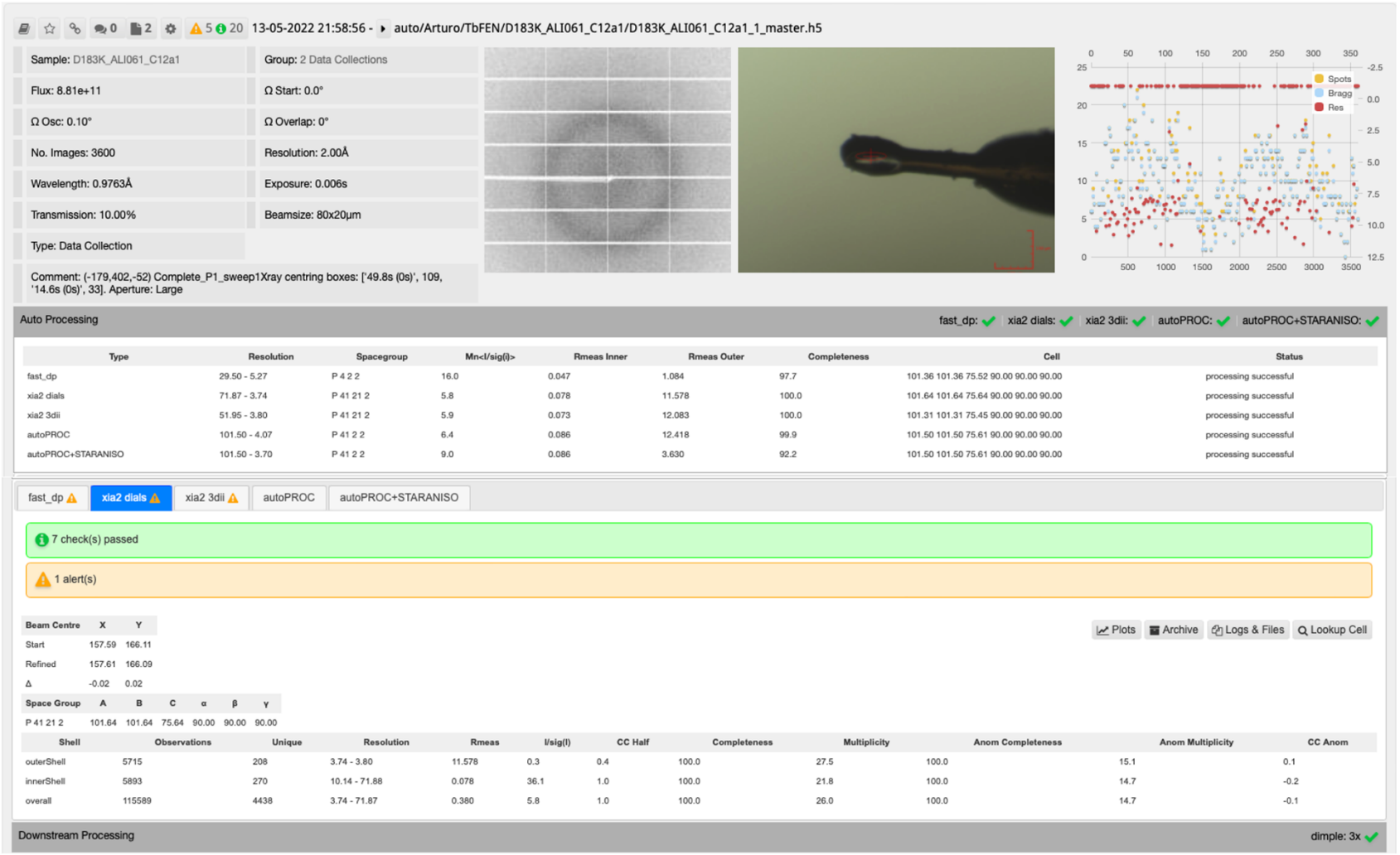


Figure 9.7 X-ray diffraction data set for *Tb*FEN-D183K ΔC-341 crystal C12a

Resolution of the X-ray diffraction of 3.74 Å according to xia2 DIALS. Crystal growing condition: 0.05 M $CaCl_2$, 0.2 M Citric acid pH 6.5, 22% (w/v) PEG 3350.

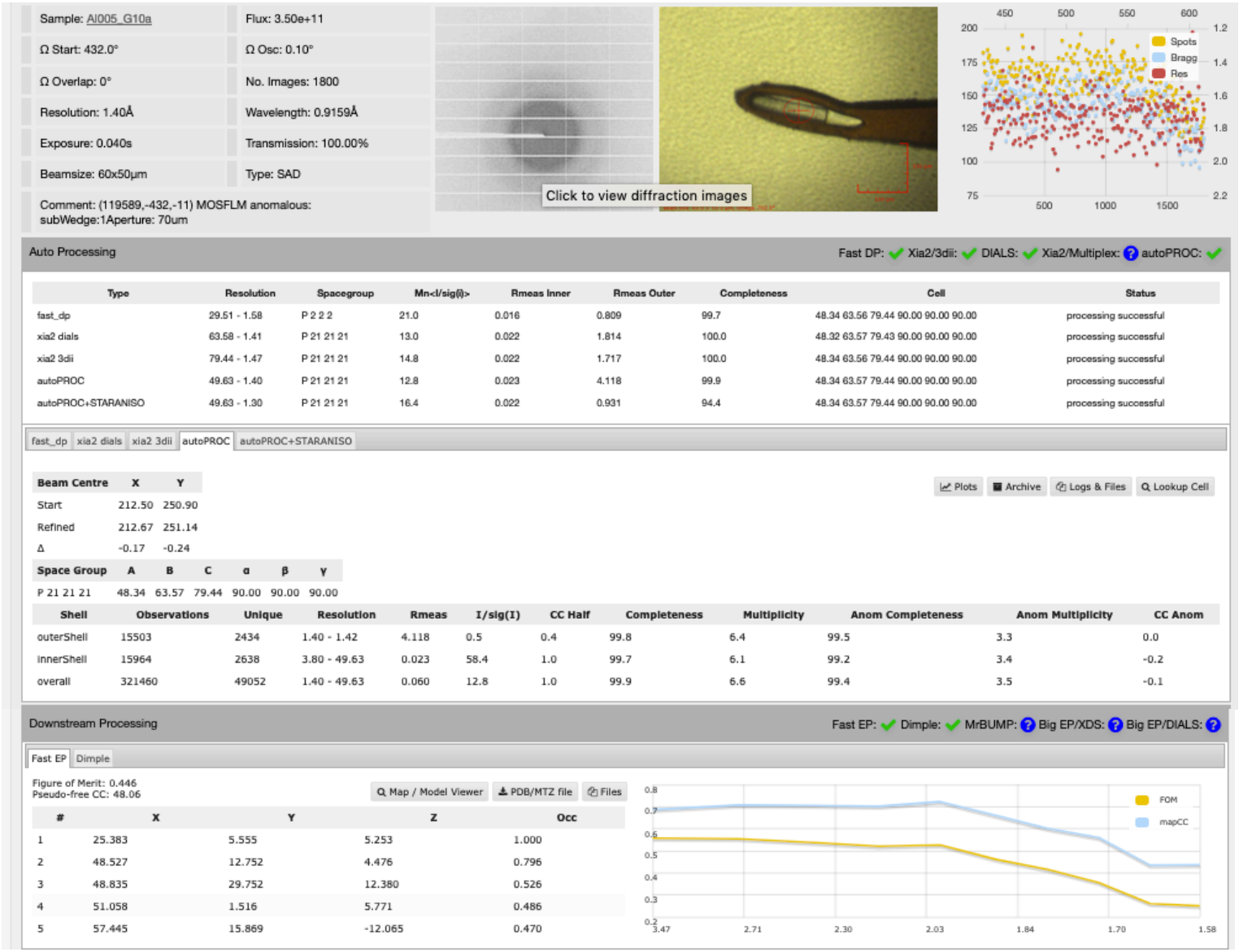


Figure 9.8 X-ray diffraction data set for *Li*FEN crystal G10

Resolution of the X-ray diffraction of 1.41 Å according to xia2 DIALS. Crystal growing condition: 0.05 M NaCl, 0.1 M $MgCl_2$, 0.1 M Bis-Tris pH 6.5, 27% (w/v) PEG 3350.

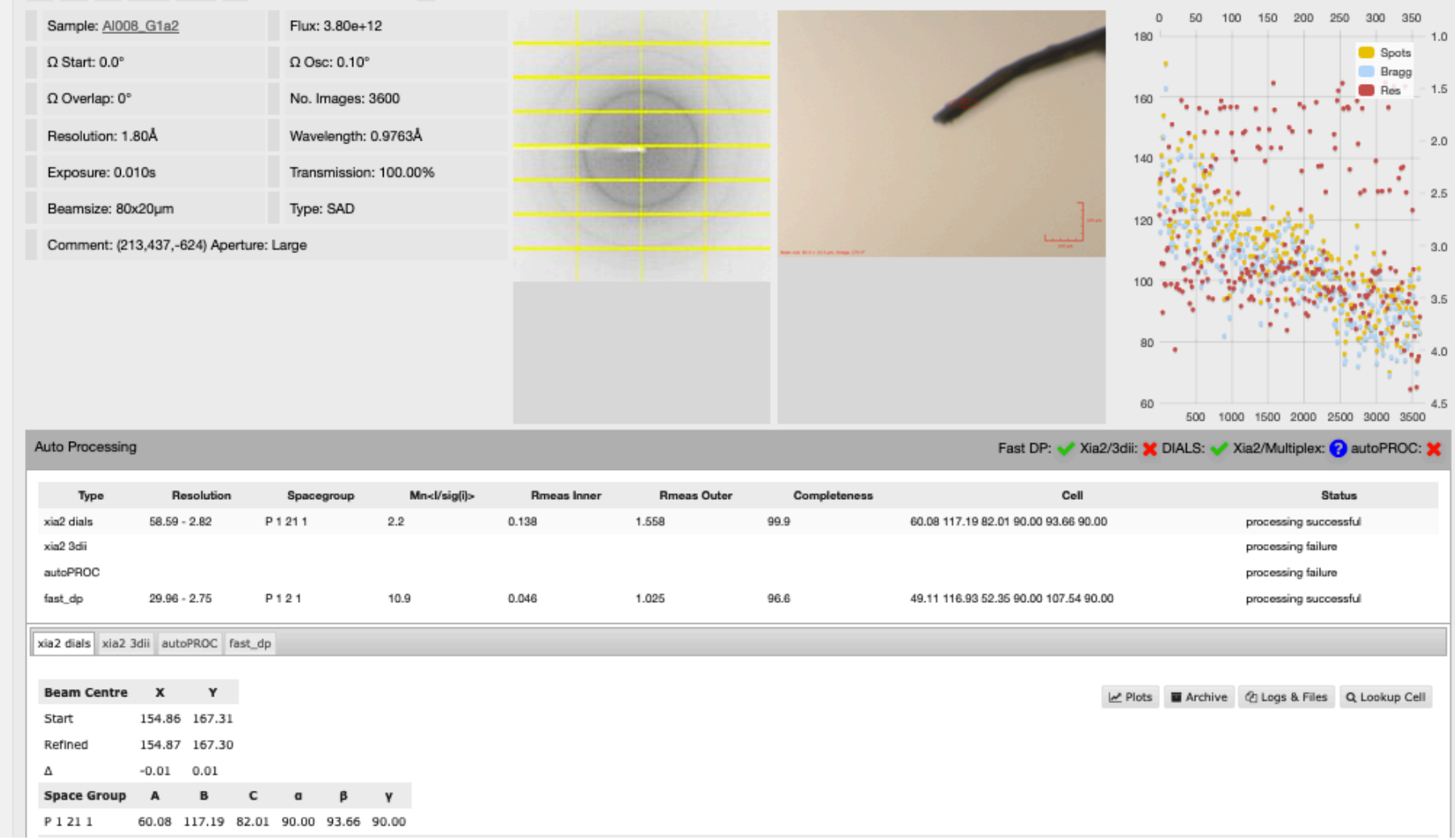


Figure 9.9 X-ray diffraction data set for *Li*FEN crystal G1

Resolution of the X-ray diffraction of 2.8 Å. Crystal growing condition: 0.1 M HEPES pH 6, 22% PEG 6000 (w/v).

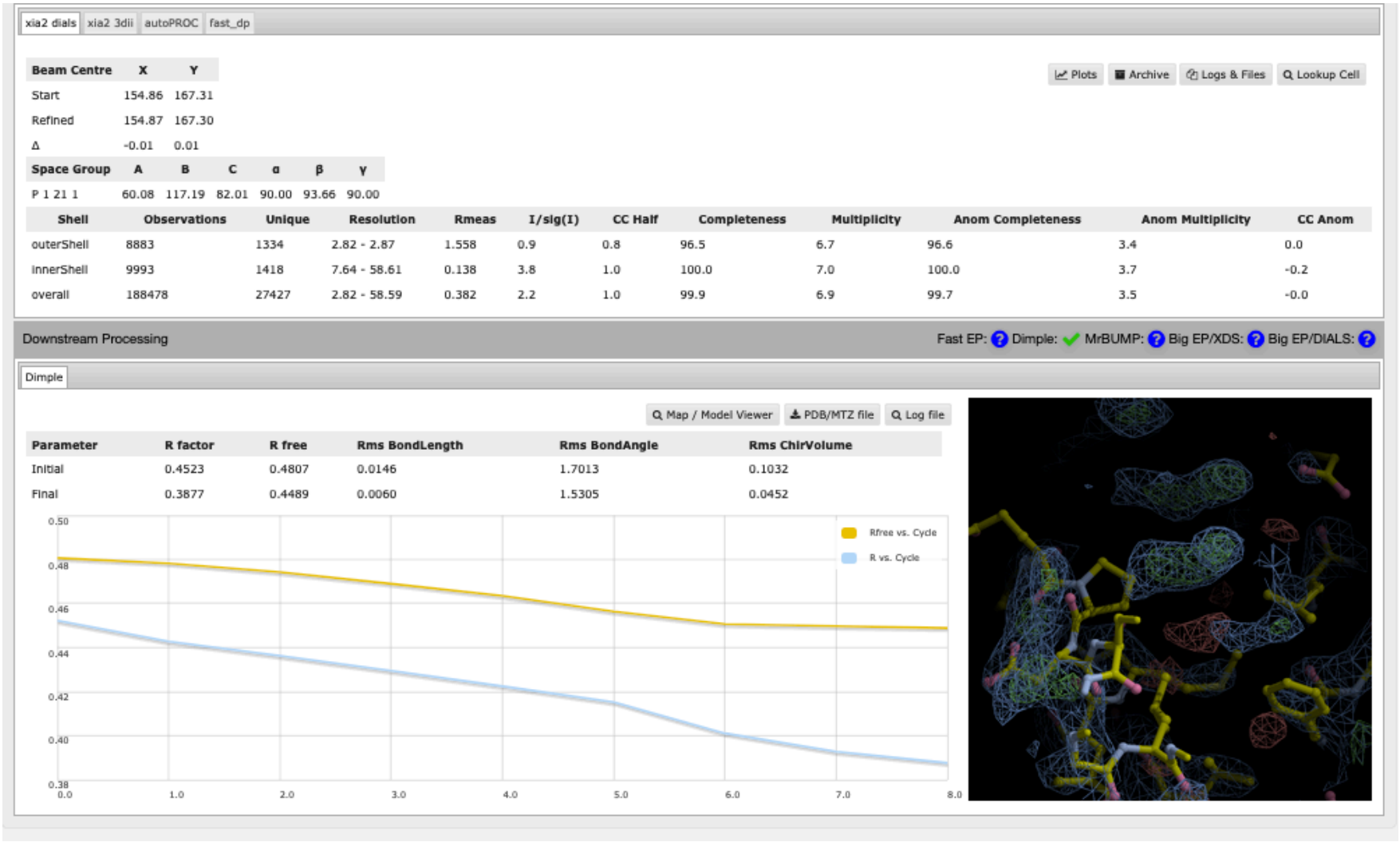


Figure 9.10 X-ray diffraction data set for *Li*FEN crystal G1 (continue)

Resolution of the X-ray diffraction of 2.8 Å. Crystal growing condition: 0.1 M HEPES pH 6, 22% PEG 6000 (w/v).

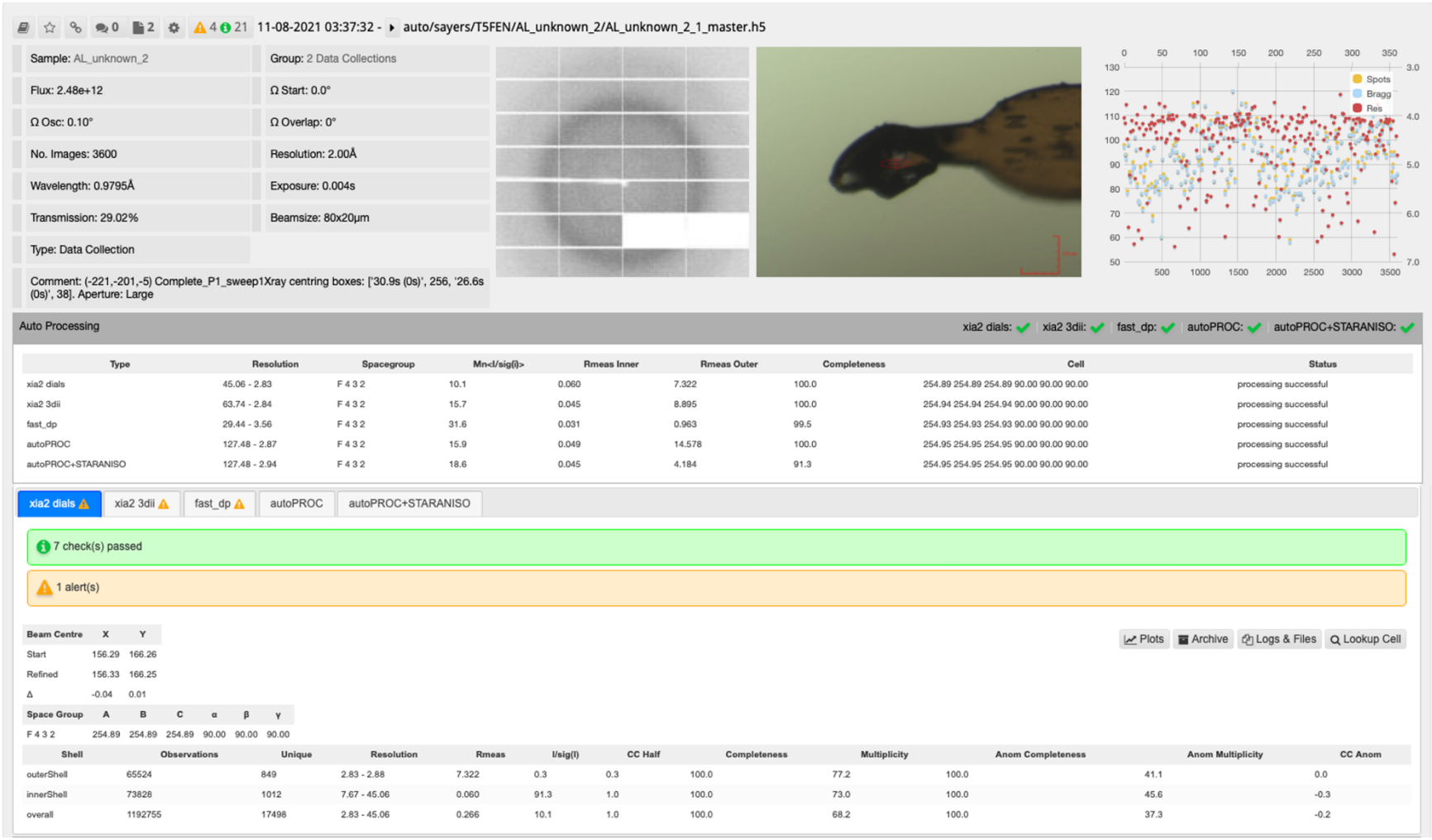


Figure 9.11 X-ray diffraction data set for *Cn*FEN ΔC-414 crystal E2b

Resolution of the X-ray diffraction of 2.83 Å. Crystal growing condition: 0.05 M CaCl2, 0.1 M Sodium acetate pH 5, 10% MPD (v/v). DNA substrate: 5ov4.