Functional and structural studies of the hydrophilic acylated surface proteins from *Leishmania donovani*

Sophie Elizabeth McKenna

Doctor of Philosophy

University of York

Chemistry

August 2015

Abstract

Protozoan parasites of the genus *Leishmania* cause a diverse range of tropical diseases referred to as Leishmaniasis that lack effective treatment or licensed vaccines. The hydrophilic acylated surface proteins (HASPs) are present in all human infective *Leishmania* species, are highly immunogenic and their expression is stage-regulated during human infection. HASPs share highly conserved N- and C-terminal domains, but a subset, the HASPBs, have a divergent central domain containing extensive hydrophilic amino acid repeats that vary in number and composition, both within and between *Leishmania* species. HASPs associate with the membrane via an N-terminal dual acylation motif that is myristoylated and palmitoylated.

The focus of this project was the HASPs from *Leishmania donovani*. Extensive crystallisation screening of the proteins in combination with an array of directed crystallisation strategies did not produce crystals. Biochemical and biophysical characterisation supported bioinformatic predicted intrinsic disorder. NMR analysis of HASPA was performed, including resonance assignment of 94% of the backbone nuclei. This required implementation of unlabelling protocols to resolve problems associated with spectral congestion and low sequence complexity. H_N, N, Cα and Cβ chemical shift analysis revealed that HASPA does not contain elements of structural propensity.

Myristoylation of the HASPs is catalysed by N-Myristoyltransferase (NMT). It was shown for the first time that recombinant NMT is able to catalyse recombinant HASP myristoylation *in vitro*. A fluorescence based assay, where a fluorescent CPM-CoA adduct is formed, was implemented to establish the kinetic parameters for this reaction. HASPA myristoylation was also monitored by real-time NMR spectroscopy, which revealed that the residues proximal to the N-terminus experienced the most pronounced changes in chemical shift or resonance intensity. A 2.45 Å resolution crystal structure of *Leishmania major* NMT in complex with 2-oxopentadecyl-CoA was elucidated. The data presented here, particularly IDP classification, will contribute to deciphering the functional role of HASPs.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my supervisors Professor Marek Brzozowski and Professor Tony Wilkinson for the opportunity to undertake my PhD at YSBL. Their continued support and guidance was invaluable throughout the course of my studies. I am indebted to Dr. Michael Plevin for his help and instruction that has allowed this thesis to take its current form. I would like to express my gratitude to Professor Deborah Smith, whose detailed knowledge of *Leishmania* Biology added context to my work. Thanks are also due to Dr. James Brannigan for the work he did on the HASPs before I started.

I would like to thank my colleagues in YSBL, both past and present, for making it a uniquely enjoyable place to work. Jenni, Katie, James, Ben and Dan helped to make everything a combination of fun and awkward. John, Christian, Tim and Catherine were instrumental in keeping me laughing, either with or at me. Last but not least, I would like to thank my family and friends who always seemed confident that I would get here. I guess you were right after all.

Author's Declaration

I declare that I am the sole author of this thesis. The work presented here was neither published before nor used previously to obtain a degree at this or another university with the exception of the Bachelor of Science in Biochemistry obtained by James Chamberlin with the HASPB1 fragment work. I carried out all of the work with the exception of the tasks listed below.

- All NMR spectra, mentioned in Chapter 2 and 3, were collected in collaboration with and processed by Dr. Michael Plevin.
- Cloning and initial expression tests of the full length HASPs, mentioned in Chapter 4, was carried out by Dr. James Brannigan prior to commencement of this thesis.
- Cloning, expression and characterisation of *Leishmania major* NMT, mentioned in Chapter 3, was carried out by Dr. James Brannigan.
- The NMT X -ray crystallographic data discussed in Chapter 3 were processed and the model built by Professor Marek Brzozowski.
- Cloning and initial expression tests of the HASPB1 fragments, mentioned in Chapter 4, was carried out by James Chamberlin.
- The HASPB1 specific antibodies used in Chapter 4 were generated and purified from the source by Mike Hodgkinson.
- The data of ESI-MS mentioned in Chapters 3, 4 and 5 and SEC-MALLS in Chapter 4 were collected by Dr Andrew Leech at the University of York.
- 2-oxopentadecyl-CoA was obtained from Dr. James Brannigan. The synthesis was performed by his collaborators at the Department of Chemistry, Imperial College.
- The pHS1403 vector, encoding T4 Lysozyme C54S C97A, was a gift from Brian Matthews (Addgene plasmid # 18355).

Chapter 1 - Introduction

1.1. Neglected tropical diseases

The bottom billion, a collective term given to the most impoverished and marginalised people in the world who typically survive as subsistence farmers, have severely limited financial means and are stuck in the perpetual cycle of disease, war and little or no education (Hotez et al. 2009). This cycle has wider consequences on socioeconomic advancement and development prospects, which in turn impedes the fight to address the core problems.

The global disease burden is quantified and assessed by the World Health Organisation (WHO) by calculating deaths and disability adjusted life years or DALYs. DALYs are defined as 'the sum of years of potential life lost due to premature mortality and the years of productive life lost due to disability' (WHO. 2013). There are three key factors that perpetuate the disease burden felt by the bottom billion: failure to use current treatments effectively, insufficient or obsolete treatments and poor disease knowledge and understanding (Morel. 2003). A concerted effort is being made by the international community to address these concerns but disproportionate allocation of funding has hampered progress in many cases; it has been estimated that 10% of global health research and development (R&D) expenditure is devoted to diseases that account for 90% of the global disease burden (Global Forum for Health Research, 2002 as reviewed by Morel. 2003).

Neglected tropical diseases (NTDs) make up a substantial proportion of the communicable disease burden felt by the bottom billion (Engels and Savioli. 2006). They are broadly defined as a group of tropical infections that are endemic in the developing world affecting impoverished populations, having been largely overlooked in terms of treatment and control. NTDs encompass the following 17 core diseases: buruli ulcer, chagas disease, cysticercosis/taeniasis, dengue/severe dengue, dracunculiasis, echinococcosis, foodborne trematode infections, human african trypanosomiasis, leishmaniasis, leprosy, lymphatic filariasis, onchocerciasis, rabies, schistosomiasis, soil transmitted helminthiases, trachoma and yaws (WHO. 2013). Notably an extensive proportion of the bottom billion is infected with at least one NTD at any time (Hotez et al. 2009). In combination, NTDs have been shown to be the 6th leading cause of premature deaths and DALYs worldwide (Hotez et al. 2007). Mass drug administration through preventative chemotherapy is being used to control and could feasibly eliminate the NTD burden; however, it is limited in its current scope as it is not effective in every case (Payne and Fitchett. 2010). In 2009 preventative chemotherapy reached 705 million people for at least one NTD (WHO. 2013). Although in recent years an unprecedented amount of money has been pledged by many organisations, governmental and non-governmental, sustained commitment is required to successfully manage the NTD burden (Zhang et al. 2010).

1.2. Introduction to Leishmaniasis

Leishmaniasis refers to a range of tropical diseases caused by flagellated protozoan parasites belonging to the genus *Leishmania*. *Leishmania* parasites belong to the trypanosoma genus of kinetoplastid unicellular parasitic flagellate protozoa. The World Health Organisation (WHO) estimates that leishmaniasis threatens 350 million people globally with 12 million currently infected. Every year 600,000 new cases are reported but due to the social stigma of the disease, many do not report their illness, significantly underplaying the true burden estimated to be 2 million cases (WHO. 2013). Notably 60,000 deaths occur annually due primarily to the visceral form of the disease. Leishmaniasis is prevalent in tropical and sub-tropical regions (shown in Figure 1-1); namely Africa, Central and South America, the Middle East, Central and Southern Asia and increasingly the Mediterranean Basin (Schonian et al. 2008). However, the threat posed by leishmaniasis is set to intensify as global warming increases the environmental range of the sandfly vector.

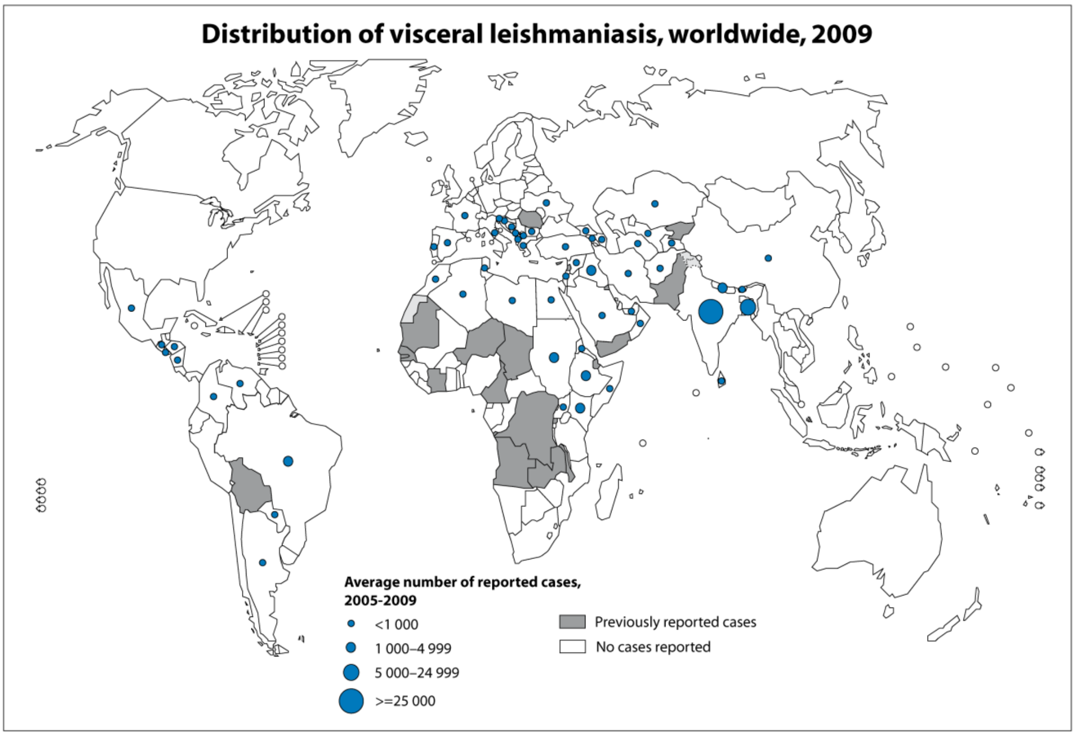


Figure 1-1: Worldwide distribution of visceral leishmaniasis (VL) in 2009 (Source WHO). VL is predominantly found in tropical and sub-tropical regions with 90% of cases in Bangladesh, India, Nepal, Brazil and Sudan.

There are around 20 human infective *Leishmania* species causing three distinct groupings of leishmaniasis; cutaneous, mucocutaneous and visceral leishmaniasis. **Cutaneous leishmaniasis** (CL) is the most common form of the disease with 1.5 million of the aforementioned 2 million cases annually being attributed to it (WHO. 2013). Skin lesions are the most prominent characteristic of CL, forming on exposed areas such as arms, legs and the face (shown in Figure 1-2a). These lesions do heal but leave permanent scars and can cause disability, attracting social stigma and impeding life quality. Species specific immunity is induced after recovery but the parasite can lie dormant triggering latent infections in immunocompromised patients. Notably 90% of all cases occur in Afghanistan, Brazil, Iran, Peru, Saudi Arabia and Syria (WHO. 2013). **Diffuse cutaneous leishmaniasis** (DCL) is characterised by chronic widely spread skin lesions and is commonly confused with lepromatous leprosy due to the similarity in clinical manifestation. Treatment is particularly problematic as relapse is common and species specific immunity is not observed.

Mucocutaneous leishmaniasis (MCL) constitutes facial lesions, leading to destruction of the mucous membranes of the face i.e. nostrils, throat and lips (as shown in Figure 1-2b). 90% of all MCL cases are located in Bolivia, Brazil and Peru (WHO. 2013). Distinct strains of leishmaniasis cause different disease manifestations. In Central and Southern America, CL and MCL are caused by members of the *Leishmania mexicana* and *Leishmania braziliensis* species, while CL is caused by *Leishmania tropica* and *Leishmania major* in the Middle East and Asia.

Visceral leishmaniasis (VL), also known as kala-azar, is the most severe form of the disease as the parasites are internalised into the spleen and liver, which in turn leads to physiological malfunctions and a mortality rate of 95% if the disease remains untreated. It is typified by enlargement of the spleen and liver, high fever, dramatic weight loss and anaemia (shown in Figure 1-2c). The main causative strain is *Leishmania donovani*, which is widespread in Bangladesh, China, Nepal and Sudan. The human infective strains predominant in the Mediterranean Basin and Latin America are *Leishmania infantum* and *Leishmania chagasi,* respectively (WHO. 2013). WHO states that 90% of VL cases are located in Bangladesh, India, Nepal, Brazil and Sudan (as shown in Figure 1-1). **Post kala-azar dermal leishmaniasis** (PKDL) is a complication that occurs after successful treatment of VL. It is typified by hypopigmented skin lesions that can be macular (a rash typified by small, flat red spots), maculopapular (a red rash on the skin covered with small bumps) or nodular non-pigmented lesions that start on the face and spread to other body parts depending on the intensity of the infection (Zijstra et al. 2003). The skins lesions harbour parasites, meaning PDKL acts as a reservoir for anthroponotic disease transmission, where the parasite is transferred by humans to other animals. PKDL is primarily linked to *L. donovani* infection, as the majority of cases are reported in Sudan, Ethiopia, Kenya, India, Nepal and Bangladesh, though *L. infantum* has also been implicated in a minority of cases.

Typically *Leishmania* species cause one form of the disease preferentially, but some species can cause distinct clinical forms of leishmaniasis depending on complex interactions between the parasite and the host immune response. (Cecílio et al. 2014). *Leishmania* isolates from various clinical forms were analysed in Brazil and the parasite species were determined. Three species were isolated from these patients, namely *Leishmania amazonensis*, *Leishmania braziliensis* and *Leishmania chagasi*. Notably, this study showed that *L. amazonensis* was the causative agent of CL, MCL, VL and four cases of PKDL (Barral et al. 1991). Therefore, different Leishmania species may be associated with indistinguishable clinical presentations.

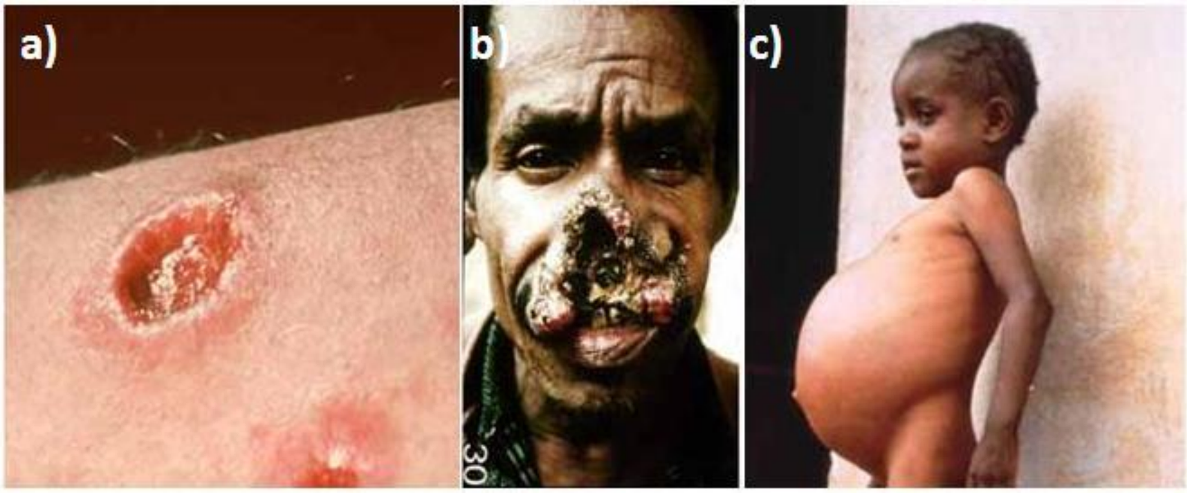


Figure 1-2: Clinical manifestations of the three distinct forms of leishmaniasis. (a) Skin lesions typify cutaneous leishmaniasis (CL). (b) Mucous membrane destruction typifies mucocutaneous leishmaniasis (MCL). (c) Enlargement of the spleen and liver cause swelling of the abdomen typifying visceral leishmaniasis (VL). Taken from The Board of Trustees of the Leland Stanford Junior University. 2007. Extracellular vs. Intracellular Parasitic Mechanisms of Immune Evasion: Leishmania. Available at: http://web.stanford.edu/class/humbio153/ImmuneEvasion/. [Accessed 25 August 15].

1.2.1. Leishmaniasis and HIV/AIDS

Strain-specific immunity prevents reinfection but latent reactivation in times of physiological stress or immunosuppression is a constant concern with leishmaniasis. An emerging epidemic is the co-infection of HIV with visceral leishmaniasis, which has not only medical but diagnostic and epidemiological consequences (Chappuis et al. 2007). It is important to note that cases of HIV co-infection with other forms of leishmaniasis have been reported but as HIV-VL co-infection is of most concern, it will be discussed here. In VL-endemic regions significant proportions of the population have an asymptomatic infection. Concurrent HIV infection considerably escalates the risk of developing active VL by between 100 and 2320 times (WHO. 2013). The two diseases reinforce the progression of the other in the host. Immunosuppression causes the patient to be particularly susceptible to severe and rapid onset VL, while VL accelerates the onset of AIDS by triggering further immunosuppression promoting virus replication. Notably co-infected patients show lessened humoral and cellular responses to *Leishmania*, resulting in a higher rate of treatment failure independent of the drug used, increased probability of relapse, increased parasite load in the blood and bone marrow and ablated sensitivity to serological tests (Chappuis et al. 2007).

Diagnosis of VL in HIV patients is problematic as the clinical manifestations can often be atypical or lacking entirely; for example splenomegaly is less frequent in co-infected patients but is a useful indicator of severe VL. Parasitological diagnosis employs spleen, bone marrow or lymph node aspirates and analyses parasitic load using microscopy or culture. It is considered a highly sensitive detection method but is limited as adequately experienced staff and equipped clinics are required. Unfortunately a spleen aspirate can in 0.1% of patients lead to life threatening haemorrhages so it is not suitable for highly recurrent relapses typical of VL-HIV co-infection (Alvar et al. 2008). Less invasive methods, such as blood tests, can be used in tandem with PCR. This molecular detection method is more sensitive than parasite visualisation techniques, enabling the user to determine the *Leishmania* species involved. PCR can also be used to monitor long term treatment efficacy and relapse reoccurrence. Although it is limited to a clinical setting, this method does prevent false positives when lower trypanosomatids display VL or CL like symptoms.

Serological detection methods are relatively simple and non-invasive but in 40% of co-infected cases there is no detectable level of *Leishmania*-specific antibody (Alvar et al. 2008). To combat poor sensitivity it is standard practice to carry out two tests concurrently, however, false negatives and positives are associated with serological detection methods. In a co-infection context relapse is not detectable using serology as after treatment antibody levels remain elevated for several years. Antigen-based detection methods are more specific than antibody-based systems. A latex agglutination test (KAtex made by Kalon Biological) detects *Leishmania* specific antigen levels in the urine of patients with an active infection. It gives a rapid result and can be used to monitor relapse and treatment efficacy. Further improvement is underway as the urine must be boiled before testing to avoid a false positive result and there are problems with distinguishing a weakly positive result from a weakly negative result.

Highly active antiretroviral therapy (HAART) is the standard treatment for HIV-positive patients, employing a combination of at least three drugs in order to inhibit virus replication and in turn disease progression (WHO. 2013). In the context of co-infected patients HAART intervention reduces morbidity and mortality; whereby the number of new VL cases dropped by 50 to 65% in Europe after HAART introduction (Alvar et al. 2008). HAART intervention alone is not sufficient to prevent relapse, hence a VL re-emergence secondary prophylactic therapy is utilised (antileishmanial drugs are discussed in Section 1.2.3). However, all current antileishmanial regimes are less effective in co-infected patients, leading to high mortality rates due to progressive drug toxicity, complications and concurrent illness. Careful management of treatment (biochemical monitoring is vital) and rapid diagnosis of re-emergence are required to manage this emerging epidemic.

This co-infection phenomenon is rampant in South West Europe among intravenous drug users, whereby 70% of all VL cases in adults are associated with HIV infection (WHO. 2013). It is of grave concern that the spread of HIV infection is facilitating the spread of VL to previously unaffected areas particularly in Southern Europe. HAART intervention (as previously mentioned) has shifted the focal point to where the two diseases cross over geographically and HAART coverage is insufficient (Alvar et al. 2008). In VL endemic areas the issue is particularly pronounced because of poor detection and underestimation of the number of cases due to a lack of facilities. Moreover, there are no co-ordinated detection schemes in place and treatment availability is limited. In Ethiopia it is currently estimated that 1.3 million people are HIV-positive but only 65,000 of these are receiving HAART (Alvar et al. 2008). Moreover, co-infected individuals act as anthroponotic reservoirs perpetuating leishmaniasis in the population.

1.2.2. Leishmaniasis: Control strategies

The impact of leishmaniasis is at present being managed through transmission control and drug intervention. Control strategies include reservoir and vector control and early detection and rapid treatment of infected individuals in endemic areas. Parasite transmission to the female sandfly vector can be zoonotic (wild or domestic animal reservoirs) or anthroponotic (human reservoir) depending on the *Leishmania* strain involved. Zoonotic visceral leishmaniasis caused by *L. infantum* mainly affects young children with domestic dogs implicated as the primary parasite reservoir. Treatment of infected dogs is not considered effective due to the high rate of treatment failure and prospect of relapse. Dog culling has been implemented, particularly in Brazil as a national control programme, but conflicting efficacy reports have called into question the effectiveness of this strategy (Ashford et al. 1998 and Dietze et al. 1997 reviewed by Quinnell and Courtenay. 2009). Replacement of culled dogs with uninfected susceptible dogs in a mean time frame of 4 months has been put forward as an explanation for the limited effectiveness of this strategy (Nunes et al. 2008). A more complete programme of dog culling focusing on entire endemic areas rather than pockets of human habitation, and the use of a possible future canine vaccine would offer more efficacious control. Moreover, better availability of more reliable and rapid diagnostic tests in the field would enable current control programmes to be more effective as use of slow diagnostic methods assures large delays between testing and culling enabling further disease spread. Application of topical insecticides and the use of deltamethrin impregnated collars have demonstrated substantial individual protection against *Leishmania* infection in Europe (Quinnell and Courtenay. 2009). Community-wide implementation of deltamethrin collars one year after introduction in Iran has led to a considerable reduction in infection rates (Gavgani et al. 2002). In a peridomestic setting, collar expense and loss are problematic; though mathematical modelling predicts that collar intervention would be effective even if a proportion of the collars were lost (Quinnell and Courtenay. 2009).

Further work needs to be carried out to establish the effect of sylvatic cases, where wild animals are infected, on *Leishmania* spread; unlike infection in dogs, sylvatic infection is mainly asymptomatic. This makes disease detection in wild animal populations particularly problematic and requires the role of these cases as disease reservoirs (in tandem with or independently of domestic or stray dog cases) to be clarified. Infection rates reported in PCR studies among the red fox population in the Mediterranean were shown to be as high as 40-75%, highlighting the need to evaluate endemic areas as a whole not just the habitable pockets to manage disease transmission (Quinnell and Courtenay. 2009).

Control strategies against the female phlebotomine sandfly (the transmission vector for leishmaniasis) include residual spraying of human dwellings and animal shelters, insecticide impregnated dog collars (as previously mentioned), chemical repellents and insecticide treated bed nets. Residual spraying schemes aim to reduce sandfly bites in and around domestic and peridomestic settings through initial application of DDT (**d**ichloro**d**iphenyl**t**richloroethane) and more recently synthetic pyrethroids. It is important to note that the aforementioned pesticides are primarily used to control the malaria vector but are also efficacious against sandflies. Nationwide schemes to eradicate malaria with DDT spraying in Italy and India have coincided with a reduction in VL cases, but when spraying was halted the disease re-emerged (Alexander and Maroli. 2003). The scope of spraying was shown to be limited in a widespread scheme in Brazil due to low coverage in more remote areas, people being outside in the early evening coinciding with the vector being active, higher concentrations of vectors outside of dwellings and use of synthetic pyrethroids that are short lived compared to DDT (Quinnell and Courtenay. 2009). Therefore, insecticide spraying is more effective in an urban as opposed to a rural setting because of the blanket coverage that can be achieved by spraying every residence and animal shelter compared to spraying dispersed settlements providing weak regional coverage. Fastidious reapplication of insecticide is a limiting factor in efficient control exacerbating the difficulties of remote rural control schemes. Insecticide spraying is also limited by the habits of susceptible populations. In East Africa, most transmission occurs outside the home as many people are subsistence farmers making spraying of dwellings mostly ineffectual in preventing transmission.

Insecticide treated nets (ITNs) have been used to prevent domestic and peridomestic transmission of vector borne diseases such as malaria, leishmaniasis and Japanese encephalitis (Chappius et al. 2007). ITNs function as 'baited traps' whereby the odour of the sleeper attracts the sandfly and the impregnated insecticide kills it. The insecticides used also have deterrent and repellent effects so reducing the number of sandflies inside protected houses even affording relative protection to non-users. To prevent further transmission during the VL epidemic in Eastern Sudan 357,064 ITNs were distributed to 155 villages between May 1999 and March 2001 (Ritmeijer et al. 2007). This emergency intervention saw a reduction of 59% in cases 17-20 months after distribution and an overall drop of 27% in case numbers. Although this study does appear to highlight the positive impact of ITN use on transmission prevention, the authors are critical of the limitations of the data collected. There was much concern that the many variables involved in the study make it difficult to ascertain if the data obtained were truly representative of the situation and impact of ITNs. Original distribution data showed 93.6% bednet coverage in target villages. Two years after distribution 43.7% of the nets were considered intact enough to afford adequate protection (Ritmeijer et al. 2007). There were two clinics in the region involved in VL treatment so it was noted that reported case numbers did not accurately depict the true number of cases as expense of treatment and the distance to the clinic prevented some seeking medical aid. Concurrently health programs run at the time of the outbreak were hypothesised to have led to an increase in cases being reported. In more global terms the variation in sandfly species, habits of local populations, varied environments and type of parasite transmission to the sandfly (zoonotic or anthroponotic) are a few examples of why it is so difficult to ascertain the effectiveness of bednets (Ostyn et al. 2008). The sandfly vector is most active in the early evening so most transmission occurs in this period, rendering the protection afforded by bednets ineffective as they are not being used. Extending the theory of insecticide impregnated materials to curtains and wall cloths could be the next possible step in transmission control. Co-ordinated implementation of the aforementioned control strategies can, and is, being used globally in a multifaceted approach to reduce *Leishmania* transmission.

1.2.3. Leishmaniasis: Treatment options

Early diagnosis (see Section 1.2) and treatment of leishmaniasis is vital to the individual affected and the community at large. Untreated VL patients, when transmission is anthroponotic, act as disease reservoirs and require immediate treatment to control spread of the disease. The current therapeutic repertoire against leishmaniasis is based on drugs developed 70 years ago that require long treatment courses, require administration by injection, have varied toxic side effects leading to patients withdrawing from treatment, are expensive to synthesize and have varied levels of efficacy. The expense of a long treatment cycle has become a limiting factor in the patient's ability to access medical care. Moreover, resistance is an emerging problem, meaning much attention is now being given to the study of *Leishmania* biology to enable more efficacious drugs to be developed.

Pentavalent antimonials were the first class of drugs developed in the 1940s to treat leishmaniasis and are still a first line therapy today. Sodium stibogluconate (Pentostam) and meglumine antimoniate (Glucantime), two pentavalent antimonial compounds, are currently available for clinical use and are manufactured by GlaxoSmithKline and Sanofi-Aventis, respectively (Figure 1-3, compounds 2 and 7). Intravenous or intramuscular administration is required as no oral preparations are available with a typical treatment cycle lasting 28 to 30 days, involving the daily administration of 20 mg per kg of body weight (Alvar et al. 2008). Antimonials are toxic compounds with frequent adverse and in some cases lethal side effects including cardiac arrhythmia, acute pancreatitis, nausea, vomiting, myalgia, abdominal pain, diarrhoea, skin rashes and hepatotoxicity. In some cases the severity of cardiotoxicity and pancreatic inflammation leads to cessation of treatment. Treatment failure due to emerging resistance is rapidly becoming a problem with antimonials particularly in Bihar. Since the 1990s the aforementioned treatment cycle had a cure rate of 36 to 69% in Bihar (Olliaro et al. 2005).

Amphotericin B deoxycholate (Figure 1-3, compound 1) is a polyene anti-fungicide originally isolated from *Streptomyces nodosus* that exhibits anti-leishmanial activity (Monzote. 2009). It selectively binds ergosterol, a sterol present in the parasite cell membrane analogous to mammalian cholesterol. This binding alters the permeability of the cell membrane causing pores to form and leading to ion and intracellular component leakage, which ultimately results in cell death. The efficacy of amphotericin B is high and it is a viable option for patients who have shown resistance to antimonials although toxicity remains a concern. Intravenous administration is required with a standard regime being 1 mg per kg of body weight every other day for 15 doses over 30 days (Olliaro. 2010). Infusion-related side effects such as fever, chills, rigor and thrombophlebitis are common; however, life-threatening conditions including serious cardiac arrhythmias, hepatotoxicity, renal dysfunction, severe hypokalaemia and even death do occur. These severe side effects require a patient to be closely monitored throughout the treatment course, so limiting the availably of treatment to well-equipped hospitals and clinics. The cost of a treatment cycle ($2800 per regime) and the length of the cycle also restrict amphotericin B usage (Chappuis et al. 2007).

In order to reduce toxicity the deoxycholate moiety has been replaced in some formulations with lipids to mask amphotericin B from vulnerable tissues. This has led to improved efficacy and reduced toxicity. There are three commercially available amphotericin B preparations namely AmBisome (Nexstar Pharmaceuticals, Inc.) a liposomal amphotericin B formulation, Abelcet (The Liposome Co., Inc.) an amphotericin B lipid complex preparation and Amphotec (Sequus Pharmaceuticals, Inc.) a colloidal dispersion of cholesteryl sulfate and amphotericin B. AmBisome has been studied most extensively and a typical course involves 5 infusions of 2 mg per kg of body weight. It has been shown to produce high cure rates (>90%) when trialled in conjunction with conventional amphotericin B deoxycholate even when a short course of 5 days is implemented (Sundar et al. 2004). Moreover, the World Health Organisation has worked in conjunction with VL endemic countries to lower the cost of a treatment course of AmBisome to $18 per 50 mg vial as of January 2010 in an effort to make treatment more accessible, equating to a price reduction of 90% (WHO. 2013). In Bihar state, an area where VL is endemic and poverty widespread, antimonial resistance has become an increasing concern and so the implementation of a short treatment course could alleviate much of the burden felt by medical centres, staff, patients and their families.

Pentamidine (Sanofi-Aventis) was originally used to treat African trypanosomiasis but was shown to be effective against *Leishmania* infections (Figure 1-3, compound 4). It requires parenteral administration and a standard treatment regime of 4 mg per kg of body weight per day for 10 to 20 days with breaks between infusions (Olliaro et al. 2005). Low efficacy and high toxicity has led to this drug being abandoned as a viable treatment option. The following side effects have been observed; hypotension, pain at the site of injection, nausea, headache, metallic taste, myalgia, numbness and reversible hypoglycaemia. The latter can in turn lead to irreversible insulin dependent diabetes mellitus in 4 to 12% of cases and even death (Monzote. 2009).

Miltefosine (hexadecylphosphocholine or He-PC) is an alkyl phospholipid compound and was initially developed to treat breast cancer and other solid tumours before it was shown to have anti-parasitic properties (Figure 1-3, compound 3). *In vitro* studies with *L. infantum* and *L. donovani* promastigotes produced IC_{50} values between 0.89 and 2.25 μg/ml (Kuhlencord et al. 1992). The parasite load in the liver, spleen and bone marrow of BALB/c mice was then used to directly compare the efficacy of Miltefosine with Pentostam. Both drugs were equally effective at reducing the parasite load in the liver but Miltefosine treatment led to a reduction of parasite load in the bone marrow and spleen significantly greater than that observed with Pentostam. Miltefosine was already approved for human trials as an anti-cancer drug so clinical studies were set up to determine its efficacy against human visceral leishmaniasis. Initial trials showed it to be highly efficacious and relatively safe when compared with other more toxic treatments. When directly compared with a cure rate of 97% obtained for amphotericin B deoxycholate in a controlled trial Miltefosine treatment reported a cure rate of 94.3% in 299 patients, leading to it being registered for VL treatment (Sundar et al. 2002).

A standard treatment regimen depends on the weight of the patient and entails a daily dose of 50-100 mg for 28 days (Olliaro et al. 2005). In 10 days most patients feel better and this has led to them halting the treatment course early because of the high cost. This is of particular concern in areas where transmission is anthroponotic (e.g. India) as it is likely to lead to drug resistance. Therefore, compliance post-improvement is vital to maintain the longevity of Miltefosine as a viable treatment option. Miltefosine can be administered orally (and topically) so eliminating the need for hospitalisation and in turn reducing related treatment costs. Gastrointestinal complications are the most common side effects with Miltefosine treatment and include vomiting, nausea and diarrhoea in varying degrees of severity. Reversible nephrotoxicity and hepatotoxicity are less common but have also been observed. It is important to note that Miltefosine is teratogenic and cannot be given to pregnant women or those who are about to become pregnant (Sundar and Olliaro. 2007).

Paromomycin (Figure 1-3, compound 6) is an aminoglycoside antibiotic that is effective against a wide range of protozoa and bacteria. It was first shown to have anti-leishmanial activity in the 1960s (Neal. 1968). However, its efficacy against human leishmaniasis was not assessed until the late 1980s. When directly compared with Pentostam for treating VL in a randomised controlled trial in Bihar it was shown to be more efficacious at producing a ‘final cure’ (Jha et al. 1998). Paromomycin was tested at 12, 16 and 20 mg per kg of body weight for 21 days while Pentostam was trialled at 20 mg per kg of body weight for 30 days. Four trial groups each comprising 30 patients were treated with one of the aforementioned regimes. The Paromomycin treated groups observed a ‘final cure’ of 23, 28, and 29 patients (12, 16, and 20 mg/kg/day respectively) 180 days after treatment compared with 19 of the patients given antimony. Intravenous or intramuscular administration is required as the drug is poorly absorbed when taken orally. It can also be applied topically in combination with a variety of supplementary compounds to treat CL. Combination with methylbenzethonium chloride has been shown to be most effective while urea and soft paraffin have been used in some less efficacious formulations (Kim et al. 2009). Pain at the site of injection, reversible ototoxicity and an increase in hepatic transaminase levels are associated side effects of Paromomycin use. The low cost of a course of treatment, which is $15 for a 21 day regime, makes Paromomycin a particularly appealing option (WHO. 2013). However, limited availability restricts widespread usage in endemic areas.

Sitamaquine is an 8-aminoquinoline analogue developed by the Walter Reed Army Institute in conjunction with GlaxoSmithKline for the treatment of VL (Figure 1-3, compound 5). It like Miltefosine can be administered orally. Reversible methemoglobinemia, which causes cyanosis, is a common side effect of Sitamaquine treatment and biochemical monitoring is constantly required throughout a treatment regime. Dyspepsia, vomiting and acute nephritic syndrome (in particular glomerular nephritis) have also been observed after Sitamaquine treatment (Singh et al. 2012). In a Phase 2 trial against *L. chagasi* the efficacy of Sitamaquine was suboptimal after 28 days of treatment (Dietze et al. 2001). Final cure rates were 0%, 17%, 67%, 20% and 0% at doses of 1 mg, 1.5 mg, 2 mg, 2.5 mg and 3.25 mg per kg per day over 28 days respectively. However, efficacy and toxicity require further investigation to better understand how Sitamaquine can fit into the arsenal of anti-leishmanial treatments.

Combinatorial therapy is steadily becoming standard practice to mitigate unfulfilled treatment needs, extend the lifetime of current drugs, reduce the overall cost and shorten treatment schedules. In order for co-administration to be successful, pharmacological and practical considerations must be taken into account. One tried and tested combination is Paromomycin with antimony, which has proved efficacious and relatively innocuous but is limited by increasing resistance to antimonials (Olliaro et al. 2005). In Sudan between 2002 and 2005 the efficacy of Pentostam in combination with Paromomycin was compared with antimony alone (Melaku et al. 2007). A regime of 30 days of Pentostam was replaced with 17 days of combination therapy. Initial cure rates highlighted the comparable efficacy of combination therapy (97%) with monotherapy (92.4%). Data collected also intimated that combination therapy led to a reduced instance of death when directly compared with monotherapy. In order to devise new antimony free combinations *in vitro* and *in vivo* studies were carried out with miltefosine in combination with amphotericin B, sodium stibogluconate, paromomycin and sitamaquine (Seifert and Croft. 2006). Notably *in vitro* and *in vivo* techniques reported conflicting values; for example amphotericin B was effective *in vivo* but suboptimal *in vitro* whereas the converse was true for sodium stibogluconate. Induction of double resistance to multiple drug combinations in *L. donovani* emphasised the ease of double resistance development when treatment is not sufficiently controlled (Garcia-Hernandez et al. 2012). This highly problematic issue could hinder two treatment options instead of one. Clarification of the variables that impede efficacious combinations, such as species specificity, is the focus of this work.

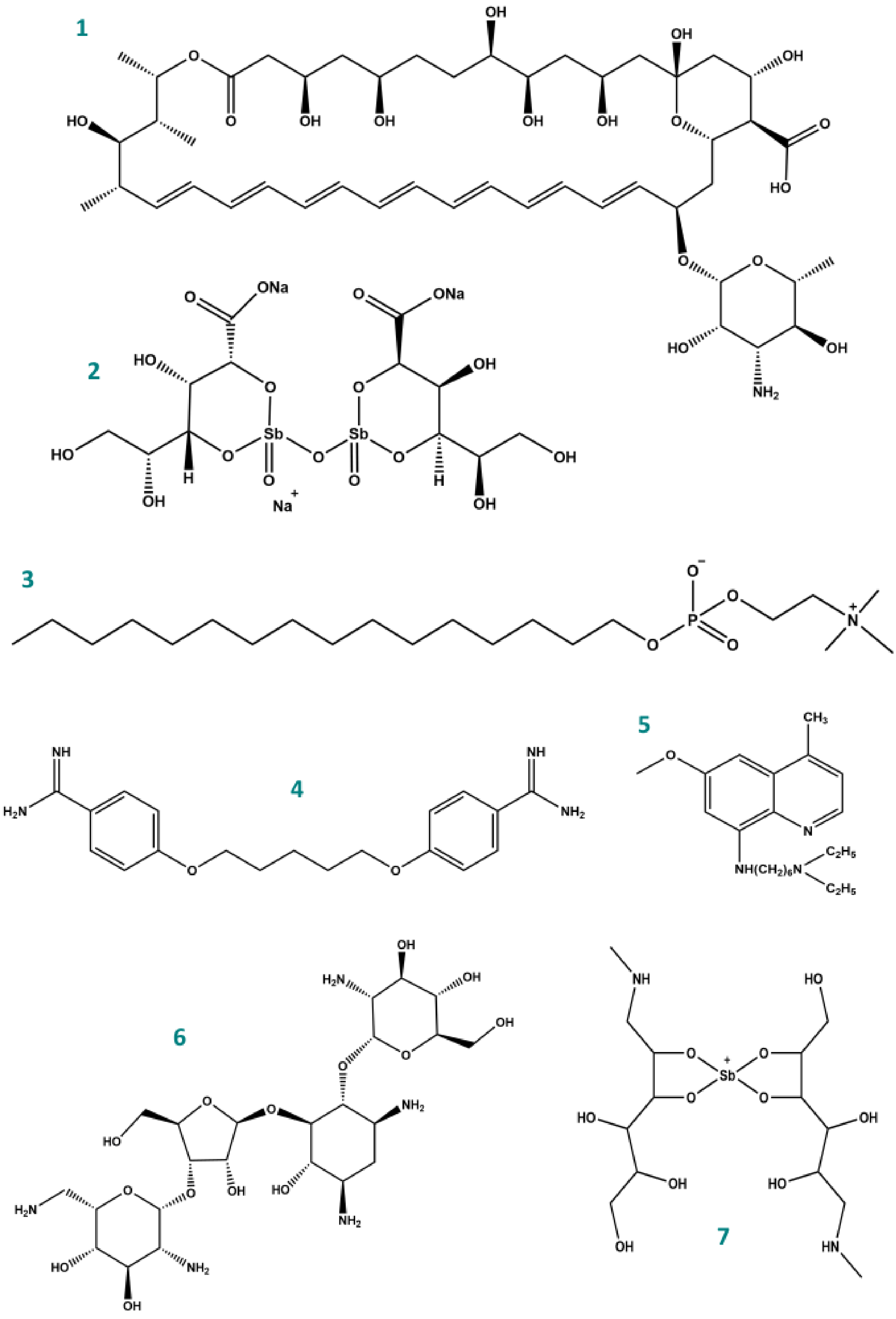


Figure 1-3: Current anti-leishmanial drug compounds. (1) Amphotericin B (2) Sodium stibogluconate (3) Miltefosine (4) Pentamidine (5) Sitamaquine (6) Paromomycin and (7) Meglumine antimoniate.

1.2.4. Vaccine development

There is currently no effective vaccine licensed for the prevention of any form of leishmaniasis licensed for use in humans. Recovery from infection induces lifelong immunity in the host so development of an effective vaccine is achievable, while epidemiological enquiry suggest that a degree of cross-protection between *Leishmania* species is possible (Modabber. 2010). Unfortunately, progress has been limited. 'Leishmanization' is the process of inoculating a patient with live virulent *L. major* causing a self-healing lesion that affords protection against further infection. However, this vaccination programme has been discontinued as in some cases it can cause persistent infection and is particularly hazardous to the health of co-infected patients (Khamesipour et al. 2012).

Attempts to produce a prophylactic vaccine using inactivated parasites have garnered attention over the past several decades. The ease with which parasites can be grown in culture and the low comparative cost associated with production made this an attractive solution but is yet to yield an efficacious product. In Colombia randomised clinical trials were carried out in three doses with a monovalent *L. amazonensis* vaccine to establish safety and immunogenicity (Velez et al. 2005 as reviewed by Noazin et al. 2008). The vaccine was shown to be safe and immunogenic but afforded no protection against infection. This finding is corroborated by results from other studies reviewed by Noazin et al. (2008) and is consistent with an insufficient immune response produced by whole killed parasites in terms of duration and relevance.

Immunochemotherapy combines chemotherapeutic treatment with vaccination to elicit an overall cure in the most challenging cases such as HIV co-infection and difficult to treat diffuse leishmaniasis. An immunochemotherapeutic regime can be used to reduce the dose of antimonials required for a final cure to mitigate the highly toxic side effects associated with treatment. In a controlled trial half the standard dose of Glucantime was combined with a *L. amazonensis* vaccine and achieved 100% cure rate after no more than four treatment cycles (Machado-Pinto et al. 2002). The high incidence of persistent post-kala-azar dermal leishmaniasis in Sudan warranted an initial trial to evaluate the impact of immunochemotherapy on cure rates. An *L. major* vaccine with the Bacille Calmette-Guérin (BCG) adjuvant was tested in combination with Pentostam compared with an antimony only group. Notably 60 days after the start of treatment 87% of the immunochemotherapeutically treated patients were cured compared with 53% in the antimony alone group (Musa et al. 2008).

Sequencing *Leishmania* genomes (*L. major, L. infantum, L. braziliensis, L. mexicana* and *L. donovani* genomes have been sequenced to date) and an improved understanding of *Leishmania* pathogenesis has revealed new lines of inquiry (Kumar and Engwerda. 2014). In particular, genomic analysis has shown a large degree of sequence homology between species, leading to the conclusion that it may be possible to generate broadly effective vaccines against different forms of leishmaniasis. The genetic heterogeneity required to produce the various clinical manifestations is achieved by variations in chromosome number and increased gene expression due to chromosome amplification (Rogers et al. 2011).

Numerous recombinant proteins, either alone or with adjuvant, have also been tested as potential vaccines in preclinical trials. HASPB, a protein from the family of hydrophilic acylated surface proteins that are membrane associated proteins of unknown function expressed at human infective extracellular (metacyclic) and intracellular (amastigote) stages, has been shown to be immunogenic in mice conferring considerable protection to VL challenge without adjuvant and leading to significant control of the parasite burden (70 to 90% protection was reported at day 80 post-infection) in the spleen (Stäger et al. 2000). This is particularly noteworthy because of the persistent nature of parasitic infection in the spleen even after chemotherapeutic treatment. Moreover, HASPB1 immunisation induces production of IL-12p40 (the 40 kDa subunit of IL-12) and IL-12 (heterodimeric cytokine made up of IL-12p40 and IL-12p35) by splenic dendritic cells, which is a key component of the early immune response to *Leishmania* infection as neutralisation of IL-12 leads to elevated parasite numbers in the spleen and liver. Further work by Stäger et al. (2003) showed that natural antibody recognition of HASPB1 led to downstream regulation of complement dependent pathways that in turn lead to IL-4 production. This subsequent production of IL-4 led to a cytokine cascade resulting in CD8+ T cell priming, which is vital to the development of an efficacious *Leishmania* vaccine. The protective capacity of HASPB as a viable vaccine component was further investigated against canine VL (*L. infantum*) with and without Montanide™ ISA 720 (SEPPIC) as an adjuvant (Moreno et al. 2007). Dogs were immunised subcutaneously three times over 3 months with varied *Leishmania* specific antigens and then challenged with 10^8 virulent promastigotes. In the HASPB inoculated groups (either alone or in combination with histone H1 protein) 50% of animals immunised were protected from infection. HASPB1 intervention leads to a marked increase in antigen specific antibody production in both symptomatic and asymptomatic dogs.

Recent development of a recombinant adenovirus based vaccine (Ad5-KH) is of particular interest to this project (Maroof et al. 2012). It encompasses synthetic versions of the *Leishmania* antigens hydrophilic acylated surface protein B (HASPB) and kinetoplastid membrane protein 11 (KMP11). Single vaccination with Ad5-KH inhibited splenic parasite growth by 66% compared with the control after the mice had been infected with *L. donovani* for 21 days prior to immunisation. This project is in the final pre-clinical stages and will be taken further.

1.2.5. Life cycle of *Leishmania*

The *Leishmania* parasite has a digenetic life cycle, colonising the female phlebotomine sandfly and the mammalian host (Figure 1-4). It lives extracellularly in the sandfly midgut as the mammalian infective metacyclic promastigote migrating to the proboscis and when a blood meal is taken the mammalian host is inoculated with parasite. The metacyclic promastigote is defined by a small cell body, a long motile flagellum and its resistance to complement mediated lysis. After inoculation the promastigotes are phagocytosed by the macrophage of the innate immune system. However, when internalised into the phagolysosome, instead of being destroyed, the parasite is able to differentiate into the non-flagellated highly replicative amastigote form and multiply whilst evading the immune system. The change in the extracellular environment, an increase in temperature from 33 to 37°C and decrease in pH to 5.5, facilitates this differentiation and has been shown to elicit the same response *in vivo* (McConville and Handman. 2007). Notably this intracellular differentiation step is a key stage in the life cycle of the parasite and is vital to establish infection in the mammalian host. These infected macrophages will then rupture over time, spreading the infection by releasing the infective parasites.

When a new sandfly feeds on this infected host the tissue damage associated with feeding allows it to ingest infected macrophages from the skin. These macrophages are subsequently lysed in the sandfly midgut, which releases the amastigotes that then transform into rapidly dividing, non-infectious stage procyclic promastigotes due to the ensuing change in environment (decrease in temperature and increase in pH). Subsequent escape from the peritrophic matrix, a semi-permeable chitin, protein and glycoprotein-rich protective structure that encases the blood meal in the midgut of the sandfly, is facilitated primarily by parasite chitinases and differentiation of the parasite into strongly motile long nectomonad promastigotes (Dostalova et al. 2012). The non-infective stage promastigotes undergo a process of attachment to the midgut wall, release and anterior migration to the thoracic midgut and towards the stomodeal valve associated with their differentiation from leptomonad promastigotes into non-dividing metacyclic promastigotes. This process is known as metacyclogenesis. This migration and differentiation of the parasite means that when the next blood meal is taken the sandfly is primed to inoculate a new host. The parasite is able to spread from host to host effectively by constantly adapting to the diverse environments it encounters.

To survive in these immensely disparate environments, the parasite has evolved a sophisticated array of stage-regulated membrane-bound components and secreted factors to aid host colonisation (Corrales et al. 2010). The dense coat of lipophosphoglycan (LPG) expressed on the entire surface, including the flagellum and known as the glycocalyx, is dynamically regulated according to the lifecycle stage. The modified variants of the LPG coat have been shown to confer resistance in both hosts, from complement mediated lysis and the sandfly immune response, while also facilitating vector: parasite interactions such as midgut attachment (Sacks et al. 2000). This survival is analogous to the parasite functioning as a biomolecular Swiss army knife continually adapting protein expression to face the persistent challenges of host establishment. It therefore seems prudent to investigate and characterize these membrane-bound factors to gain a greater insight into parasite infectivity, which can in turn inform and be used to develop more effective intervention strategies.

Metacyclogenesis is a vital developmental process of the *Leishmania* parasite and is required to produce the mammalian infective metacyclic promastigote. Intensive study of this process has identified the genes of the LmcDNA16 locus (Section 1.3) as vital for parasite development in the sandfly and highlighted them as potential targets to disrupt transmission to the mammalian host (Sadlova et al. 2010). This locus, located on chromosome 23, encodes five stage-regulated genes found in all human-infective *Leishmania* species. These genes encode the hydrophilic acylated surface proteins HASPA1, HASPA2 and HASPB and the small hydrophilic ER associated proteins SHERP1 and SHERP2 (Flinn et al. 1992). HASPB protein expression was shown to be strongly upregulated in metacyclic promastigotes (Alce et al. 1999). Notably, protein expression varies between *Leishmania* species, in *L. major* HASPB is initially expressed in metacyclic promastigotes, which continues into the amastigote form. While in *L. mexicana* HASPB expression is greater in the amastigotes rather than the metacyclic promastigotes (Depledge et al. 2010). HASPA1 and HASPA2 expression has currently been studied and published at the mRNA level (Flinn et al. 1992). HASPA1 is expressed in metacyclic promastigotes and continues to be expressed in amastigotes like HASPB. HASPA2 is expressed in procyclic promastigotes after differentiation from amastigotes directly after sandfly uptake of the parasite. This expression continues until the parasite differentiates back into the amastigote form inside the mammalian host. Recent analysis of the stage-regulated expression of the HASPAs showed that HASPA1 expression is predominately observed in amastigotes, while HASPA2 is seen in promastigotes (Doehl et al. unpublished data). SHERP protein expression commences at the late leptomonad stage where the highest expression is observed in metacyclic promastigotes (Knuepfer et al. 2001).

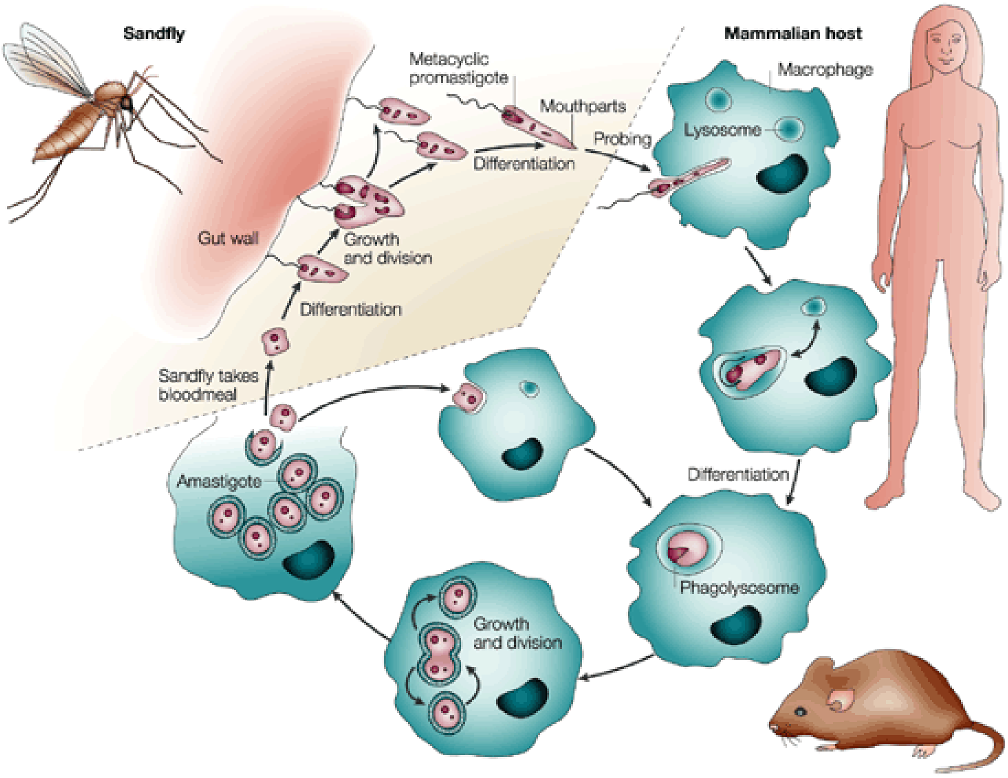


Figure 1-4: The digenetic life cycle of the *Leishmania* parasite (taken from Sacks et al. 2002). When an infected female sandfly takes a blood meal the human host is inoculated with metacyclic promastigotes. Macrophage mediated phagocytosis leads to phagolysosome internalisation. This environmental changes induces differentiation of the parasite into the resistant non-flagellated amastigote form. Multiplication by binary fission follows until the host cell can no longer contain the parasitic burden and ruptures, releasing amastigotes capable of infecting further macrophages. A subsequent bloodmeal transfers the infection to another sandfly. Ingested amastigotes enter the midgut of the sandfly and differentiate into procyclic promastigotes. Gut wall attachment facilitates further multiplication and after 6-9 days they become metacyclic. The metacyclic promastigotes migrate to the pharyngeal valve and are transferred at the next bloodmeal. Taken from Sacks et al. (2002).

1.3 The LmcDNA16 locus and the HASPs

The HASPs, of the LmcDNA16 locus, share highly conserved N- and C-terminal regions (Figure 1-5); demonstrated by Alce et al. (1999) with structural alignment data. There is little variation in the highly related HASPAs, while the HASPBs contain a highly charged repeat region (Flinn et al. 1994). Repeat-containing proteins are a common theme of protozoan parasites and are often involved in the evasion of the host immune system or play a role in virulence, however, this does not imply antigenicity. The repeat region of HASPB, although presently functionally undefined, has been implicated as the cause of HASP immunogenicity (Jensen et al. 1999). The size, number and amino acid composition of these repeats vary between the different *Leishmania* species and within each distinct species. Notably *L. mexicana* HASPB has 3.5 repeats whereas *L. donovani* HASPB1 contains 22 (McKean et al. 1997). It is of interest that during the design of the *haspb* gene used in the Ad5-KH viral vaccine 19 isolate-specific HASPB repeat region variants were identified from VL-causing strains highlighting inter and intra-species variation of the repeats (Maroof et al. 2012).

The focus of the project is visceral leishmaniasis and therefore the *L. donovani* HASPs will be discussed here. The three HASPs are termed HASPA, HASPB1 and HASPB2 and have molecular weights of 9.5, 43.4 and 15.2 kDa respectively. Alignment of these three proteins emphasizes the conserved nature of the N- and C-terminal domains (Figure 1-6). HASPA contains a species non-specific (i.e. the amino acid composition is not conserved) linker region that connects the N- and C-terminal regions. HASPB1 contains 22 imperfect repeats of PKEDGHTQKNDGDG that encompass some amino acids substitutions while HASPB2 has 2.5 repeats. The highly charged nature of these repeats is highlighted in the sequence alignment of the *L. donovani* HASPs (Figure 1-6). It is noteworthy that the *L. donovani* HASP repeats have a lower proline content than the *L. major* repeats (Alce et al. 1999). Analysis of the theoretical isoelectric point (pI) of the three HASPs with the ProtParam tool from ExPASy returned values between 4.7 and 4.9 (Gasteiger et al. 2005). The charged nature of these proteins causes aberrant migration of the proteins on SDS-PAGE and has also been noted in regards to the *L. major* HASPs (Section 4.3.1.).

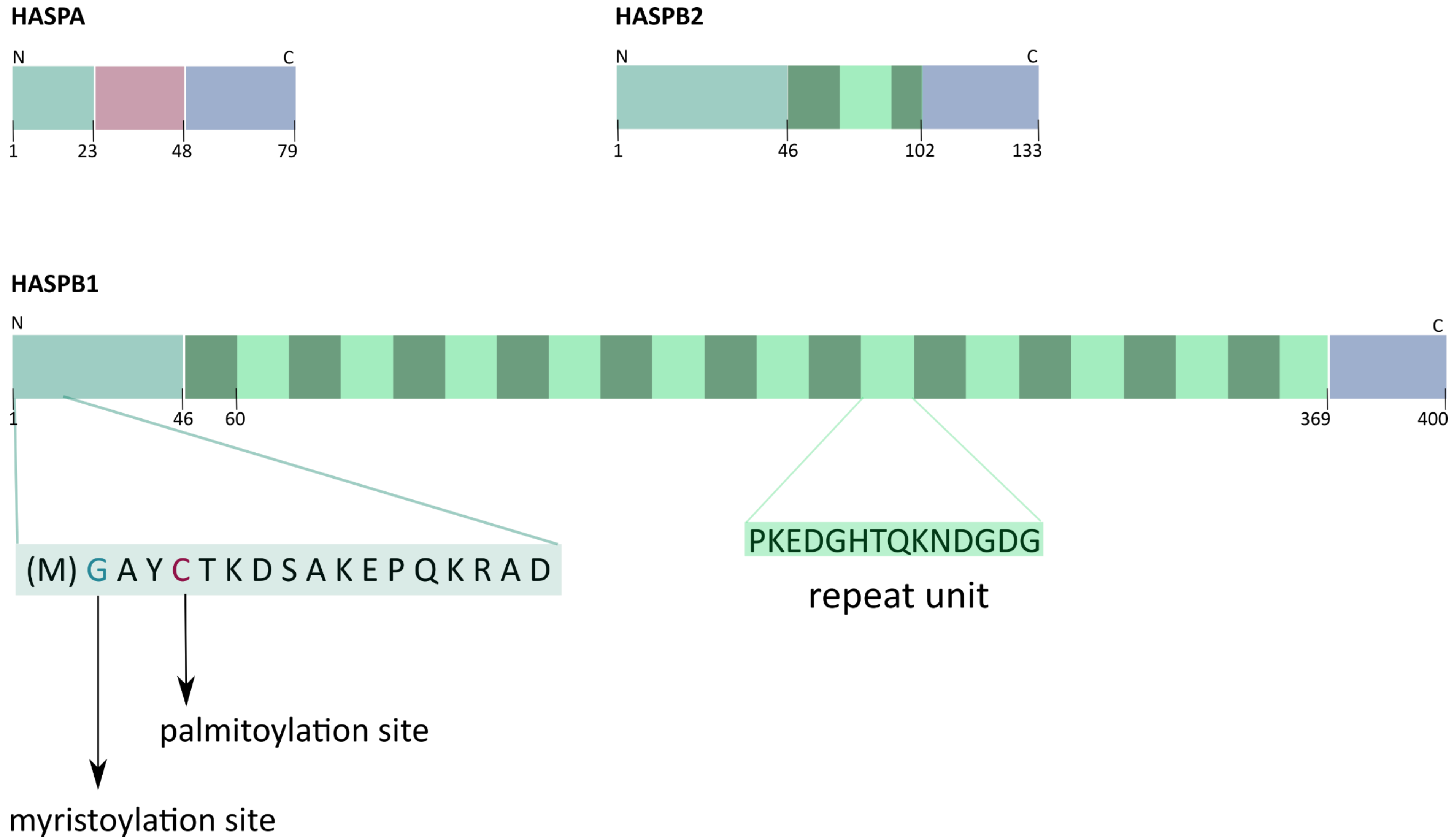

HASPA
N
C
1
23
48
79
HASPB2
N
C
1
46
102
133
HASPB1
N
C
1
46
60
369
400
(M) G A Y C T K D S A K E P Q K R A D
palmitoylation site
myristoylation site
PKEDGHTQKNDGDG
repeat unit

Figure 1-5: Domain composition of the *L. donovani* HASPs. The conserved N- and C-terminal regions are shown in green and blue respectively. The N-terminal region encompasses a dual acylation motif, whereby the N-terminal glycine is firstly myristoylated and then sequentially palmitoylated at the highlighted cysteine. This motif is vital for secretion and is found in all three HASPs. *L. donovani* HASPB1 contains 22 repeats of the PKEDGHTQKNDGDG motif, represented by the striped domain, where HASPB2 contains 2 and a half imperfect repeats. HASPA shares the first 23 amino acids with the HASPBs but the pink domain represents a short linker region, not found in the other HASPs. Residue numbers are included to accurately represent protein sizes.

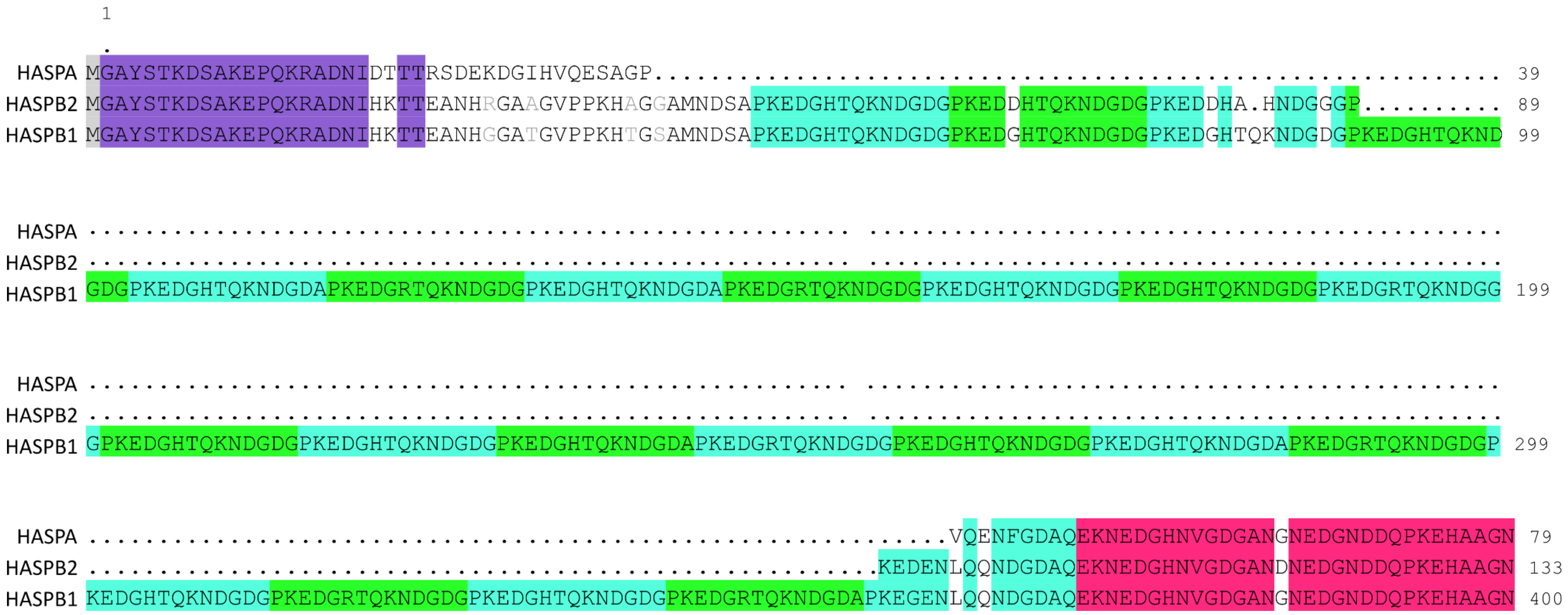


Figure 1-6: Manual sequence alignment of HASPB1, HASPB2 and HASPA from *L. donovani* to emphasize the amino acid composition of the repeats and conserved regions. All three HASPs share high sequence similarity in the first 23 residues, shown in purple, encompassing the dual acylation motif. HASPA then diverts in sequence as it has a unique linker region. The HASPBs share high sequence similarity in the first 46 residues then the repeat units begin. HASPB1 has 22 repeats and HASPB2 has 2.5 imperfect repeats. The repeat unit, of PKEDGHTQKNDGDG, is highlighted in blue and green alternately. The C-terminal region, the last 31 residues, is shared between all three HASPs and shown in pink.

1.3.1. The dual acylation motif and its role in alternative secretion

The HASPs contain no signal peptides, membrane spanning domains or GPI anchor motif for membrane association but they are known to be membrane-associated. Denny et al. (2000) implemented a deletion strategy that pin-pointed the first 18 residues of HASPB as essential for membrane targeting. They were then able to visualise the trafficking of chimeric HASPB18-GFP to the plasma membrane using fluorescence microscopy of fixed procyclic parasites. Further investigation of the membrane association of HASPs by MacLean et al. (2012) demonstrated that full length endogenous HASPB in metacyclic parasites was able to translocate to the external surface of the plasma membrane. Moreover, FRAP (fluorescence recovery after photobleaching) analysis of HASP18-GFP highlighted the bidirectional nature of its association with the inner leaflet of the cell membrane and the flagellum, which is a characteristic of alternative secretion of trypanosomatids parasites (Field et al. 2007). Initial work by Flinn et al. (1994), visualising HASPB at the flagellar pocket and plasma membrane, agreeing with this finding. Although the exact biological function(s) of the HASPs have remained elusive (McKean et al. 2001), MacLean et al. (2012) visualised endogenous HASPB from metacyclic promastigotes throughout and immediately after phagocytosis on the surface of macrophages. Moreover, expression is retained in intracellular amastigotes and HASPB is not detected on the outer membrane of the parasite at this stage. This hints at a two-phase mechanism for HASP action but requires further clarification.

Notably the N-terminal region encompasses a dual acylation motif, whereby it is firstly myristoylated and then palmitoylated (shown in Figure 1-5). This dual acylation of the motif is vital for membrane trafficking and presentation of HASPB, as the HASPs have no classical signal peptide that is required for Golgi-dependent secretion. An alternative secretory mechanism is involved in membrane presentation but the exact mechanism remains elusive (Nickel. 2005). N-myristoylation of HASPB is a co-translational event mediated by N-myristoyltransferase (NMT) a major drug target in current research, as part of a processing cascade that facilitates membrane secretion. It would be of great biological and pharmacological interest to examine this interaction.

1.3.2. The HASP expression profile in relation to parasitic infectivity

HASP expression correlates with parasite infectivity, whereby expression and localisation at the plasma membrane are seen in the metacyclic and amastigote life stages (Flinn et al. 1994). Moreover, it has recently been shown that New World *L.* (*Viannia*) *braziliensis* expresses HASP predominantly at the intracellular amastigote life stage (Depledge et al. 2010). This is of particular importance when you consider the interest in these proteins as vaccine targets because it had previously been reported that *Leishmania* (*Viannia*) parasites had entirely lost the HASP genes from their genomes.

Sadlova et al. (2010) demonstrated that knockdown of the LmcDNA16 locus *in vivo* inhibits metacyclogenesis, the transformation of poorly infective procyclic promastigotes into mammalian host infective metacyclic promastigotes, hinting at the essential role that the HASPs play in parasite differentiation to the metacyclic promastigote stage in the sandfly. Colonisation of the stomodeal valve by *Leishmania* parasites is vital for efficient transmission to the mammalian host. Null and SHERP add back parasites were not able to colonise the stomodeal valve, a small percentage of +SHERP parasites were weakly attached, unlike the wild type and HASPB add back parasites. Null and SHERP add back parasites accumulated at the elongated nectomonad promastigote stage of development and had longer bodies. Wild type and HASPB add back lines had the largest proportion of short bodied metacyclic promastigotes. Taken together these data allows the parasites to be separated into two groups with wild type and HASPB add back lines showing classical *Leishmania* development while null and SHERP add back lines demonstrate impaired development and consequently impaired infectivity. Thus HASP expression is stage specific and the HASPs are a definitive marker for metacyclogenesis and in turn parasite infectivity.

Treatment of BALB/c mice with HASPB without adjuvant leads to significant antibody production, indicating that HASPB is a potent antigen (Stäger et al. 2000 and 2003). Recombinant HASPB1 from *L. donovani* was used to diagnose VL and PKDL in Sudan using ELISA (enzyme-linked immunosorbent assay), which detected 92% (VL) and 93% (PKDL) of cases correctly (Jensen et al. 1999). These findings led to HASPB being explored as a potential vaccine candidate (see Section 1.2.4) and the subsequent Ad5-KH project

(Maroof et al. 2012). Discussion of what is known and not known about the HASPs highlights the importance of further characterisation to enable therapeutic work to progress and determine a functional role *in vivo*.

1.4 Protein Acylation: Palmitoylation and N-Myristoylation

Lipid modifications add a distinct layer of functionality to the proteome and facilitate a variety of processes including protein-protein interactions, subcellular targeting and membrane association. Fatty acylation can be co- or post-translational in nature and a wide variety of lipids can be attached through acyl linkages. Two types of acylation in proteins will be discussed here, namely S-palmitoylation and N-myristoylation that consist of the addition of palmitate and myristate, respectively.

1.4.1. S-Palmitoylation

Protein S-palmitoylation involves the thioester linkage of palmitate, a C16:0 long chain fatty acid, to the thiol group of a cysteine residue and is the most common eukaryotic protein acylation modification. Fatty acids other than palmitic acid are occasionally thioesterified to proteins and these can be saturated, mono- or polyunsaturated but must be C14 or longer. Addition of palmitate increases protein hydrophobicity, which in turn enhances the membrane affinity of otherwise cytosolic proteins. This post-translational modification is used to modulate the functionality of specific proteins and has been implicated in protein sorting and trafficking between distinct intracellular compartments, protein-protein interactions, protein stability and membrane localisation (Smotrys and Linder. 2004). The reversibility of this process adds an additional layer of regulation to target proteins, enabling highly dynamic subcellular trafficking. The rapidity of this turnover is emphasised when you consider that the lifetime of a palmitoyl moiety on a protein is significantly shorter than that of the palmitoylated protein itself (Corvi et al. 2011). This short lifetime allows for successive rounds of palmitate addition and removal, facilitating the wide array of associated processes.

Palmitoylation is predominantly regulated by two groups of enzymes, termed palmitoyl acyltransferases (PATs) and palmitoyl acylthioesterases (see Figure 1-7). PATs catalyse the addition of palmitate to the substrate protein, while the palmitoyl acylthioesterases catalyse palmitate removal. PATs are multi-spanning membrane proteins typified by a conserved DHHC cysteine rich domain. Their active sites face the cytoplasm and they are found on various membranes, mainly the Golgi and the ER but also the plasma membrane, endosomes and organism-specific membrane assemblies (Frénal et al. 2013). The integral membrane nature of the PATs has made their study challenging owing to the difficulties associated with membrane protein purification. Erf2 and Akr1, both yeast proteins, the first identified DHHC domain PATs were shown to palmitoylate the small GTPase Ras2 (Bartels et al. 1999; Roth et al. 2002; Lobo et al. 2002). Subsequently, multiple PATs were found in eukaryotes including yeast, mice, humans and protozoan parasites. The *L. major* genome encodes 20 DHHC domain PATs (Goldston et al. 2014). The large number of distinct PATs is due to their ability to selectively modify diverse targets and their diverse membrane distribution. Notably, PATs specifically accept unmodified or modified proteins and in the case of modified proteins they distinguish between different lipid modifications.

Secreted proteins and peptides are palmitoylated by membrane-bound O-acyltransferases (MBOATs). MBOATs are also multiple membrane-spanning proteins with a conserved histidine and asparagine in the active site. The histidine is surrounded by a patch of hydrophobic residues, while the asparagine is embedded in a hydrophilic region (Chang et al. 2011). The MBOAT family are categorised according to the three distinct functions they perform, namely protein or peptide acylation, acylation of the OH group of a cholesterol or diacylglycerol, and acylation of a lysophopholipid to produce a phospholipid (Matsuda et al. 2008). This functional diversity means that MBOATs are implicated in a range of biological processes, such as membrane lipid remodelling, neutral lipid biosynthesis and embryogenesis. Palmitoylation can also occur spontaneously in an enzyme independent manner. Auto-acylation has been reported for several mitochondrial proteins due to the high local concentration palmitoyl CoA in the organelle (Kostiuk et al. 2008).

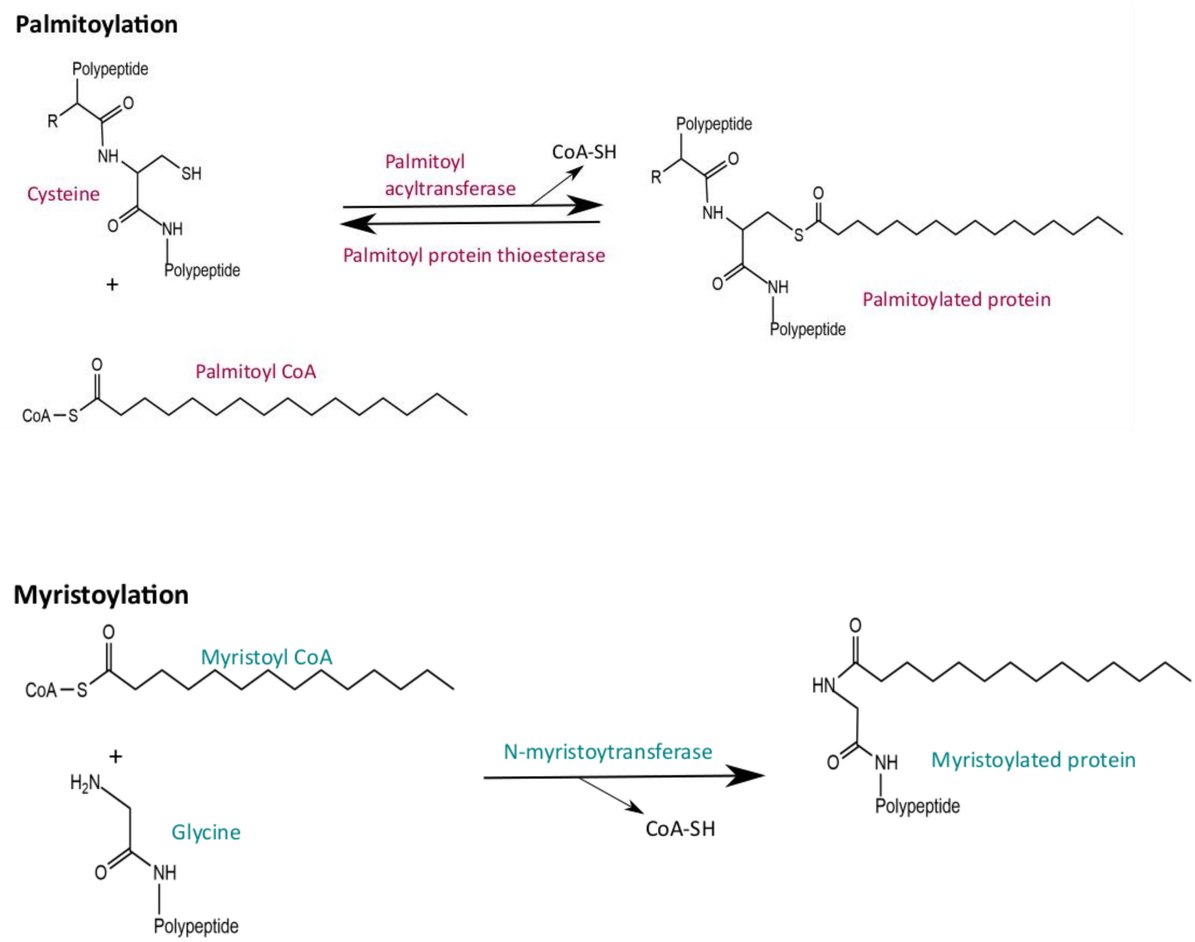


Figure 1-7: Schematic representation of palmitoylation and myristoylation reactions. Palmitoylation: Palmitoyl acyltransferase (PAT) binds palmitoyl coenzyme A and the target protein (see Figure 1-8 for recognition sequence details). It is firstly auto-palmitoylated before transferring the palmitate to the substrate protein producing palmitoylated protein and coenzyme A. This reaction is reversible with palmitate removal catalysed by palmitoyl thioesterases. Myristoylation: N-myristoyl transferase (NMT) sequentially binds myristoyl coenzyme A and the substrate protein at the N-terminal glycine. Transfer of the myristate group to the glycine amine is catalysed leading to release of the myristoylated protein and coenzyme A (CoA). This process is irreversible and is typically a co-translational event but post-translational myristoylation can occur when proteolytic processing generates a new N-terminal glycine.

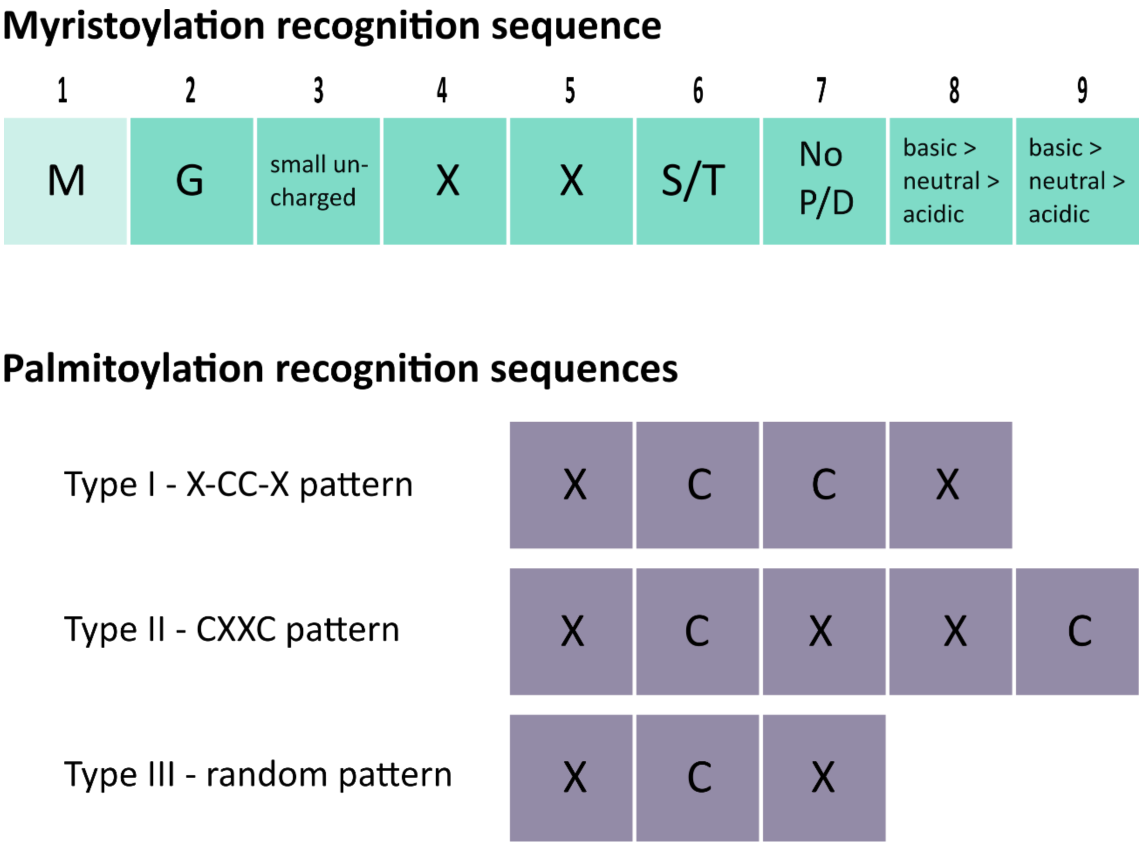
Protein palmitoylation targets are diverse and include peripheral and integral membrane proteins. The role of palmitoylation is a stabilising one for integral membrane proteins, where it is the prevalent lipid modification. Greater modification diversity is found in the peripheral membrane associated proteins, which can be either dual lipidated or singly palmitoylated (Nadolski and Linder. 2007). Myristoylation and prenylation are common modifications that precede palmitoylation. The first modification is always the prerequisite of appropriate palmitoylation. The addition of palmitate strengthens the membrane affinity of the substrate protein (Aicart-Ramos et al. 2011). In cases of dual lipidation, the palmitoylated cysteine is proximal to the other modification site. As there is no single strict palmitoylation consensus sequence other than the requirement for cysteine, *in silico* predictions of the palmitoylome are unreliable. Grouping palmitoylated proteins based on the sequence context of the modification has assisted development of prediction algorithms (Hu et al. 2011). This produced three distinct palmitoylation recognition motifs, namely Type I (XCCX), Type 2 (CXXC) and Type III (random), all with a preference for cysteine (shown in Figure 1-8). The highly variable context of the modification and recognition sites highlights its functional plasticity.

1.4.2. N-Myristoylation

N-myristoylation involves the covalent attachment of myristic acid (C14:0) via an amide bond to the amino-terminal glycine of a nascent eukaryotic or viral polypeptide (see Figure 1-7). This process is irreversible and predominantly co-translational, taking place while the polypeptide is still attached to the ribosome after removal of the terminal methionine by methionine aminopeptidase. Like palmitoylation, addition of the myristoyl moiety increases the hydrophobicity of the target protein facilitating membrane association. However, additional forces are required to facilitate effective membrane binding, such as palmitoylation or a sequence high in basic residues, since myristoylation alone is too weak to mediate irreversible binding. The comparably short myristate chain has a limited capacity for hydrophobic and van der Waals interactions giving rise to the lower membrane affinity (Goldston et al. 2014). Notably, the half-life of a membrane-associated myristoylated peptide is minutes while a palmitoylated peptide is bound for hours (Martin et al. 2011). Myristoylation is associated with diverse biological processes such as signal transduction, protein phosphorylation and dephosphorylation, oncogenesis and viral assembly. N-myristoylation is catalysed by N-myristoyltransferase (NMT), a ubiquitous eukaryotic protein (see Section 1.5).

It was previously thought to be exclusively a co-translational event but myristoylation also occurs post-translationally in apoptotic cells (Martin et al. 2011). It has been implicated in caspase-mediated cleavage of internal glycine residues and subsequent myristoylation of cryptic myristoylation consensus sites. The pro-apoptotic protein Bid was shown to be post-translationally myristoylated by NMT after caspase 8 cleavage and trafficked to the mitochondrial membrane (Zha et al. 2000). This process was shown to be a vital activating step required for Bid-induced release of cytochrome c and cell death.

Although myristoylation is irreversible, membrane association can be dynamically regulated in some target proteins termed ‘myristoyl switches’. These proteins are able to adopt two distinct conformations, one where the myristoyl moiety is buried in the hydrophobic core of the protein and the second where the myristate is free to interact with the membrane. This conformational change can be modulated by ligand binding, proteolysis and electrostatic interactions. Recoverin, a 23 kDa neuronal protein, is a myristoyl switch modulated by calcium binding. In the absence of calcium, the myristoyl group is buried in the hydrophobic core of the N-terminal domain. The binding of two calcium ions to each recoverin molecule induces a conformational change that expels the myristoyl moiety and exposes hydrophobic residues on the surface. NMR analysis highlighted three distinct conformational states and notably a short-lived intermediate state with one calcium ion bound was confirmed (Xu et al. 2011). The dynamic nature of the conformational shifts observed in myristoyl switches enables functional regulation.



Myristoylation recognition sequence

1	2	3	4	5	6	7	8	9
M	G	small un-charged	X	X	S/T	No P/D	basic > neutral > acidic	basic > neutral > acidic

Palmitoylation recognition sequences

Type I - X-CC-X pattern	X	C	C	X	
Type II - CXXC pattern	X	C	X	X	C
Type III - random pattern	X	C	X		

Figure 1-8: General myristoylation and palmitoylation recognition motifs. Myristoylation: Removal of the N-terminal methionine by methionine aminopeptidase prior to myristoylation is required to expose the glycine residue at position 2. The requirements of positions 3 to 9 are indicated above where X denotes any amino acid. Note position 6 does have a slight preference to serine over threonine. Palmitoylation: There are three distinct palmitoylation recognition motifs, Type I (XCCX), Type 2 (CXXC) and Type III (random), located throughout the protein sequence. X represents any amino acids but a preference of neutral > basic > acidic has been documented.

Consensus in the literature dictates that substrate recognition by NMT is dependent on the first 10 residues of a peptide or protein (Figure 1-8). NMT absolutely requires an N-terminal glycine residue for attachment of the lipid moiety. The remainder of the sequence motif does not have such stringent requirements but there are clear preferences for amino acid type. Position 3 shows a preference for small uncharged amino acids, while positions 4 and 5 allow any residue. Serine and to a lesser extent threonine are found at position 6, mediating a stabilising interaction with NMT. The remaining sequence displays residue type preferences but is not rigidly selective (Goldston et al. 2014). Global analysis of myristoylated proteins, crystallographic and biochemical data enabled refinement of the myristoylation motif and prediction of additional myristoylated proteins (Maurer-Stroh et al. 2002). Maurer-Stroh and co-workers expanded the recognition motif to the first 17 residues split into three distinct regions. Region 1 (residues 2 to 7) binds in the active site of the enzyme, region 2 (residues 8 to 11) interacts with the surface of NMT comprising small polar residues and region 3 (residues 12 to 18) is a hydrophilic-rich linker region. This hydrophilic region is thought to obstruct N-terminal folding that would inhibit NMT recognition and interaction. Lastly, it is important to mention that the peptide binding site of NMT is not highly conserved between species. This means that recognition sequences between species vary and there is no strict myristoylation motif for all NMTs.

1.5. N-Myristoyltransferase

N-Myristoyltransferase (NMT) facilitates a diverse array of cellular processes through the transfer of myristate from myristoyl CoA to eukaryotic or viral proteins. Two distinct NMTs catalyse N-myristoylation in vertebrates, namely NMT1 and NMT2, while only one is present in lower eukaryotes. NMTs have been cloned and purified to homogeneity from a wide range of eukaryotic organisms. Analysis of the NCBI Gene database confirms that the NMT encoding gene is present in 53 organisms, including humans, kinetoplastid parasites, yeast and fungi (Martin et al. 2011). The most detailed characterisation has been carried out on *Saccharomyces cerevisiae* NMT and this genetic analysis highlighted the important role that NMT plays in cell viability (Duronio et al. 1989). This characteristic has been subsequently observed in other NMT-expressing organisms. Non-hydrolysable myristate analogue treatment, to inhibit NMT function, of *L. major* promastigotes led to significant cell death (Price et al. 2003). Additionally, NMT is implicated in the development and progression of a range of human diseases, including cancer, epilepsy and bacterial and viral infections (Selvakumar et al. 2007; Selvakumar et al. 2005; Maurer-Stroh and Eisenhaber. 2004). Global assessment of the human proteome to estimate and identify potential targets of myristoylation has highlighted the ubiquity of this modification and clarified its vital role in eukaryotic biology (Thinon et al. 2014).

A search of the RCSB Protein Databank (PDB) reveals that the structure of NMT had been solved for a range of organisms, including *Leishmania major*, *Saccharomyces cerevisiae*, *Homo sapiens, Leishmania donovani, Plasmodium vivax, Aspergillus fumigatus* and *Candida albicans*. These structures show the enzyme to be monomeric with a compact globular structure (Bhatnagar et al. 1998). The NMT fold is the term given to the common structural features of NMT. It consists of a saddle shaped β sheet spanning the core of the protein encircled by several α helices. The structure exhibits pseudo two fold symmetry. Co-crystal structures with myristoyl CoA and peptide analogues bound show that the N-terminal domain binds myristoyl CoA in a bent question mark conformation, while the C-terminal domain binds the peptide substrate (Bhatnagar et al. 1998; Farazi et al. 2001a).

Biochemical and structural analysis confirmed that the myristoyl CoA binding site is highly conserved between species (Johnson et al. 1994; Bhatnagar et al. 1994). However, the peptide binding sites are divergent (see Section 1.4.2). Peptidomimetic inhibitors were developed to specifically target fungal NMTs over the human variants and shown to be particularly selective (Lodge et al. 1998). Unfortunately, the cost of development and the high species specificity, anti-fungal treatments require broad species specificity, meant this work was abandoned. However, selectivity is a desirable quality for the treatment of other infections and human diseases associated with NMT. Transferring the concept of NMT druggability to other species could produce a wealth of highly selective compounds for the treatment of human disease.

1.5.1. The mechanism of N-Myristoyltransferase

The catalytic cycle of NMT follows a sequential, ordered Bi-Bi mechanism (Figure 1-9). Myristoyl CoA binding to the apo-enzyme forms a high affinity myristoyl CoA: NMT complex, with a dissociation constant (K_d) of 15 nM reported for *S. cerevisiae* NMT (Bhatnagar et al. 1994). Formation of this complex induces a conformational change facilitating subsequent peptide binding. Incubation of *S. cerevisiae* NMT with S-(2-oxo)pentadecyl-CoA, a non-hydrolysable myristoyl CoA mimetic, prevented myristoylation of a peptide substrate and confirmed that the peptide binds after myristoyl CoA binds (Rudnick et al. 1991). Formation of a ternary myristoyl CoA: NMT-peptide complex leads to catalysis. The myristate group is transferred to the amino group of the peptide in a nucleophilic addition-elimination reaction. CoA is released first then the myristoylated peptide.

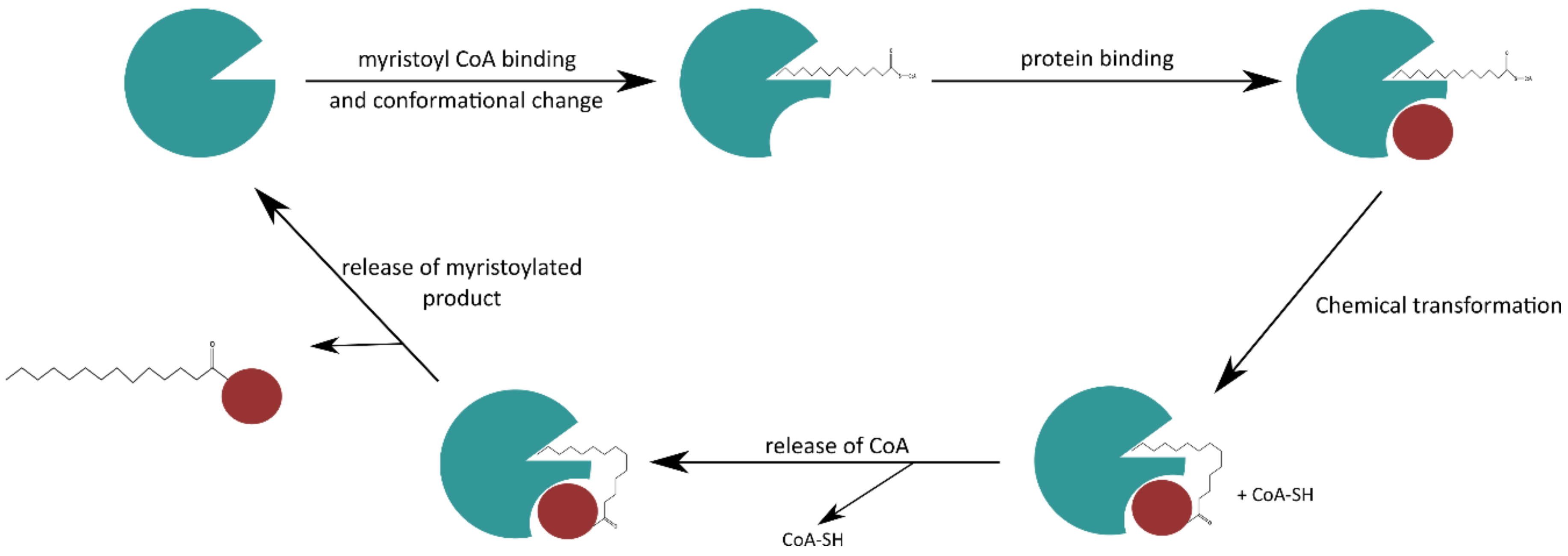


Figure 1-9: The NMT catalytic cycle is a sequential, ordered Bi-Bi reaction mechanism. Myristoyl CoA binds to the apo enzyme, inducing a conformational change that enables protein binding. Nucleophilic substitution of the thioester of myristoyl CoA by the N-terminal glycine amine produces myristoylated protein. CoA is released first followed by the myristoylated substrate. NMT is shown in blue and the protein substrate is shown in red.

1.5.2. NMT of *Leishmania* species

All *Leishmania* species have a single copy of the NMT gene that is constitutively expressed during the parasite lifecycle. 0.75% of *L. major* proteins are predicted to be myristoylated, in line with the predicted values of 0.5-0.8% of the proteomes of other eukaryotes (Mills et al. 2007). Inhibition of NMT was shown to have a detrimental effect on parasite viability, while overexpression induced distinct changes in parasite morphology associated with the subcellular accumulation of lipids that were consequently lethal (Price at al. 2003). Extensive proteomic profiling of the targets of *L. donovani* NMT identified numerous uncharacterized proteins, emphasising the wide ranging effects that NMT has on *Leishmania* biology (Wright et al. 2015). Combined with *in vivo* inhibition by established *in vitro* NMT inhibitors that led to a significant drop in myristoylation of target proteins, Wright and co-workers strengthened the case for the druggability of NMT in *Leishmania*. Moreover, the high sequence identity between NMT of different *Leishmania* species hints that an effective inhibitor against one NMT would be broadly applicable (Goldston et al. 2014).

1.6 Aims of the project

The focus of this project was the hydrophilic acylated surface proteins of *Leishmania donovani*. They have been shown to facilitate metacyclogenesis of the *Leishmania* parasite but their exact function remains unresolved. Elucidation of an X-ray crystallographic structure to determine function was the main aim of this project. Concurrent study of the HASP: NMT interaction, to contribute to the NMT drug development work, was also considered essential.

Chapter 2 - Backbone resonance assignment of HASPA

2.1. Introduction

2.1.1. Intrinsically disordered proteins

The structure-function paradigm, that states a protein must have a defined 3D structure to elicit a specific function, has been a central tenet of biochemistry and structural biology. However, a class of proteins lacking clear 3D structure termed intrinsically disordered proteins (IDPs), are redefining this view. Intrinsic disorder in functional eukaryotic proteins has been shown to be more widespread than initially assumed. It has been estimated that 40% of the human proteome is intrinsically disordered or contains disordered regions of more than 50 residues (Kragelj et al. 2013). This lack of defined structure is thought to convey distinct functionalities over structured proteins, such as promiscuous protein-protein interactions, specific but low affinity binding, flexible linkers, ready accessibility for post-translational modifications and entropic chains. The hydrodynamic radius of an IDP is much larger than that of a folded protein of the same molecular weight giving rise to a larger surface area per amino acid for interactions. Moreover, disordered proteins can potentially be more readily degraded by proteases, meaning they can be effectively regulated by degradation. An additional layer of functional control is added by co- or post-translational modifications enabling significant functional plasticity. IDPs are found as hubs in protein-protein interaction networks such as the cell cycle and signal transduction pathways (Wright and Dyson. 2015). Aberrant IDP regulation has been shown to be associated with various human diseases such as cancer, neurodegenerative diseases and diabetes (Babu et al. 2011).

A rigidly structured protein is characterised by a distinct set of backbone and side chain atom positions. The sum of these values produces an averaged structure that is subject to small amplitude fluctuations about their equilibrium positions. The energy landscape associated with ordered proteins is assumed to be funnelled, exhibiting a kinetically accessible deep free energy minimum associated with a single well-defined tertiary structure (Figure 2-1). Intrinsically disordered proteins lack tertiary structure and exist as highly dynamic heterogeneous ensembles because the backbone and side chain atom positions and dihedral angles vary significantly over time. Their energy landscape (Figure 2-1) is comparatively flat with a vast number of local minima, which leads to a series of interconverting conformers in solution (Papoian. 2008). IDP biological partner or ligand binding events can funnel the free energy landscape to something more similar to an ordered state, producing transient or more long-lived secondary structured elements. Disorder is defined by the dynamic conformational properties of a protein and not necessarily by the presence or absence of local secondary structure.

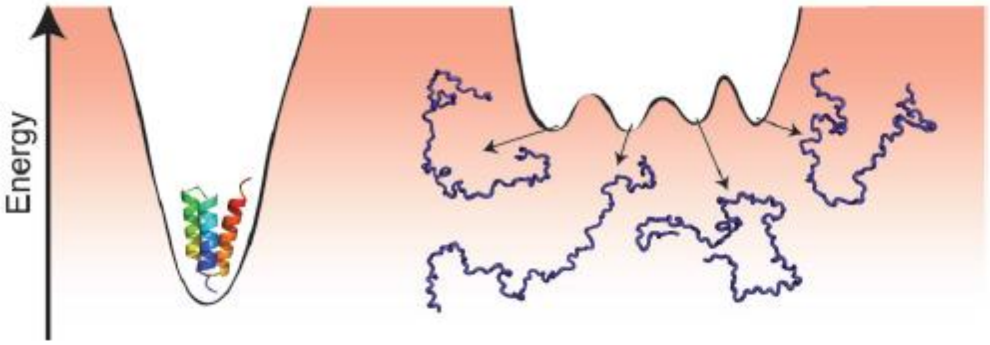


Figure 2-1: Schematic representation of the relationship between protein folding and energy landscape. The energy landscape of a folded protein (left) is funnelled making the stable state a thermodynamic minimum that is also kinetically accessible. IDPs (right) are characterised by a weakly funnelled or rugged landscape, meaning that no one state of order is favoured producing a population of dynamic disordered conformations. This image was taken from Kragelj et al. (2013).

The disordered nature of IDPs is what makes them uniquely functional. Functionality can either be related to the structurally adaptive process of biological partner binding or their ability to sample highly dynamic heterogeneous ensembles alone. Structural plasticity means that IDPs can perform various distinct functions *in vivo*. For example, topological constraints can prevent rigidly structured proteins from forming complexes but IDPs are unhindered by these constraints, allowing them to bind their biological partners and in some cases various partners. Some IDPs are not fully disordered but contain elements that are predisposed to adopt local secondary structure required for biological partner binding, these regions are termed molecular recognition features (MoRFs). The disorder to order transition contains an inherent plasticity as one IDP can adopt multiple MoRFs, this behaviour is termed structural polymorphism. A disordered region at the C-terminal end of p53 can adopt four distinct MoRFs, one α helix, one β strand and two different coils, in order to interact with four distinct biological partners (Oldfield et al. 2008). Structural polymorphism highlights the exceptional plasticity that IDPs are able to bring to cellular processes. However, some IDPs remain disordered even after target binding and form ‘fuzzy’ complexes. Sic1 is an IDP with an N-terminal multiple phosphorylation motif that on binding to Cdc4 regulates yeast cell cycle progression. Elements of disorder and transient order within Sic1 were both present when bound to Cdc4 (Mittag et al. 2010).

Entropic chains are a distinct subset of IDPs that remain disordered and unbound to any biological partners to elicit their function (Tompa. 2005). In this case the ability of IDPs to fluctuate between a series of interconverting conformers is functionally vital, as one rigid structure could not fulfil such a role. Entropic chains are involved in modulating the localisation of attached domains or generating forces to elicit structural changes to a system. This class can be further differentiated into entropic springs, bristles (or spacers), linkers and clocks. The proline-rich PEVK domain of titin was shown to act as an entropic spring in muscle (Linke et al. 2002). The random coil properties of this domain produced various mechanical conformations of different flexibility when analysed by single atom atomic force microscopy. The projection domain of microtubule-associated protein 2 (MAP2) acts as an entropic bristle exerting long range repulsive forces that provide a mechanism for maintaining spacing in the cytoskeleton (Mukhopadhyay and Hoh. 2002). Entropic chains strengthen the case for revision of the structure-function paradigm, as both order and disorder are functional.

The unique functionalities of IDPs have attracted attention as investigation of their dynamic nature could improve our understanding of biological systems and their disease states. Nuclear magnetic resonance (NMR) spectroscopy is a vital tool for analysing IDPs and can provide detailed conformational information (Konrat. 2014). Backbone resonance assignment of an IDP is a prerequisite for analysis of their dynamic nature and any residual structural elements. Specifically, an assignment is informative in cases of ligand or biological partner binding induced folding as residues implicated can be identified when they experience chemical shift and intensity changes associated with the binding event. However, it should be remembered that the plasticity of IDPs means induced folding as well as functional disorder are relevant biological states, meaning that a disorder to order transition will not be observed for all IDPs.

2.1.2. NMR background

2.1.2.1. NMR theory

For the sake of brevity, a detailed discussion of NMR theory is not included in this thesis but can be found in the following reference (Cavanagh et al. 2007). Instead, a more succinct discussion will be provided to ground the work presented here. NMR spectroscopy utilises the magnetic properties of certain atomic nuclei. Nuclei have a positive charge and many possess a net spin, specifically those with an uneven number of protons plus neutrons. This charged movement has a magnetic moment, which in turn produces a magnetic field. Consequently, nuclei can be thought of as a bar magnet that is oriented along the spin rotation axis. This feature is termed nuclear spin and can be aligned in any magnetic field, such as that generated by the magnet of an NMR spectrometer (James. 1998). The applied magnetic field means the nuclei orientation will no longer be random. For a nucleus with spin ½ (such as ^{1}H) two spin states will exist where one aligns with the magnetic field (spin up) and the other opposes it (spin down). NMR spectra arise from the transitions nuclei make between these different energy states and can be observed for certain NMR active nuclei including ^{1}H, ^{13}C, ^{15}N and ^{31}P (Kwan et al. 2011). The resultant resonance differs depending on the type of nucleus under examination, the strength of the applied magnet field and the chemical environment of the nucleus. NMR is a site-resolved spectroscopic technique where the differences between different nuclei in the same residue of a protein are informative of the local environment and subsequently the structural state.

2.1.2.2. Spectrum descriptions

Protein NMR experiments are designed to report the chemical environment of atoms within the context of a protein sequence. This information can be used to determine structure, even that of dynamic structural elements, disorder and the content of any structure i.e. α helical or β sheet. An understanding of experimental features is required to effectively utilise any resultant data. Therefore, a brief description of the NMR experiments used here will follow. Figure 2-2 details the transfer of magnetisation of each experiment described and highlights the resonance frequencies of which atoms are detected.

A 2D ^{1}H, ^{15}N Heteronuclear Single Quantum Correlation (**HSQC**) is a versatile experiment for appraising a protein sample. Recording an HSQC spectrum allows confirmation that a protein is folded, appraisal of the suitability of a protein sample for NMR structural determination and detection of ligand binding or protein-protein interactions. The HSQC experiment involves the transfer of magnetisation from the amide proton to the attached nitrogen via an insensitive nuclei enhanced by a polarization transfer (INEPT) step. The chemical shift is evolved on the nitrogen and after a time delay (t_1), the magnetisation is transferred back to the amide proton via a retro-INEPT step for detection (Figure 2-2). In theory each backbone amide group, with the exception of proline, should yield a peak in the HSQC spectrum (Bodenhausen and Ruben. 1980). In addition, a peak is observed for each indole NH of tryptophan and pairs of peaks are observed for the side chain amide groups of asparagine and glutamine (Kwan et al. 2011).

The **HNCO** is a 3D experiment that typically requires [*U*-^{15}N, ^{13}C] labelling of the sample protein. Magnetisation is passed from the amide bound proton to the ^{15}N, as with the HSQC, then to the carbonyl ^{13}C via the ^{15}NH-^{13}CO J-coupling (Kay et al. 1990). The magnetisation is transferred back via ^{15}N to the amide proton for detection (Figure 2-2). The HNCO spectra is the most sensitive triple resonance experiment and correlates the amide group with the ^{13}C of the carbonyl group of the preceding residue. This experiment also reports resonances for the side chain amides of asparagine and glutamine. The main purpose of the HNCO is to measure the resonance frequency of carbonyl atoms, so these can be correlated to the ^{15}N amide resonances. In addition it is used for sequential assignment with the HN(CA)CO experiment. In this thesis, the 3D HNCO is used in a 2D ^{15}N-H^{N} plane. This appears similar to an HSQC of your target protein but reports amide peaks that are i+1 to the carbonyl group of the preceding residue.

The **HN(CO)CACB** is a 3D experiment that requires ^{15}N, ^{13}C labelling of the sample protein. Initially, magnetisation is passed from i-1 residue $^{1}H\alpha$ and $^{1}H\beta$ to $^{13}C\alpha$ and $^{13}C\beta$, respectively (Grzesiek and Bax. 1992a). Then from $^{13}C\beta$ to $^{13}C\alpha$ where it moves onto ^{13}C of the carbonyl carbon. It passes to the i residue ^{15}NH and then to ^{1}HN for detection (Figure 2-2). This experiment links the Cα/β resonances of the previous residue (i-1) with the amide resonance of the i residue. The chemical shift is not evolved on the carbonyl ^{13}C.

The **HNCACB** is another 3D experiment requiring a ^{15}N, ^{13}C labelling pattern. Magnetisation is passed from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ, respectively, on both the i and i-1 residues then from ^{13}Cβ to ^{13}Cα (Grzesiek and Bax. 1992b). It is transferred to ^{15}N of the i residue from both ^{13}Cα$_{(i)}$ and ^{13}Cα$_{(i-1)}$ then to the amide proton for detection (Figure 2-2). So each amide group will produce four cross peaks for the two sets of Cα and Cβ resonances detected.

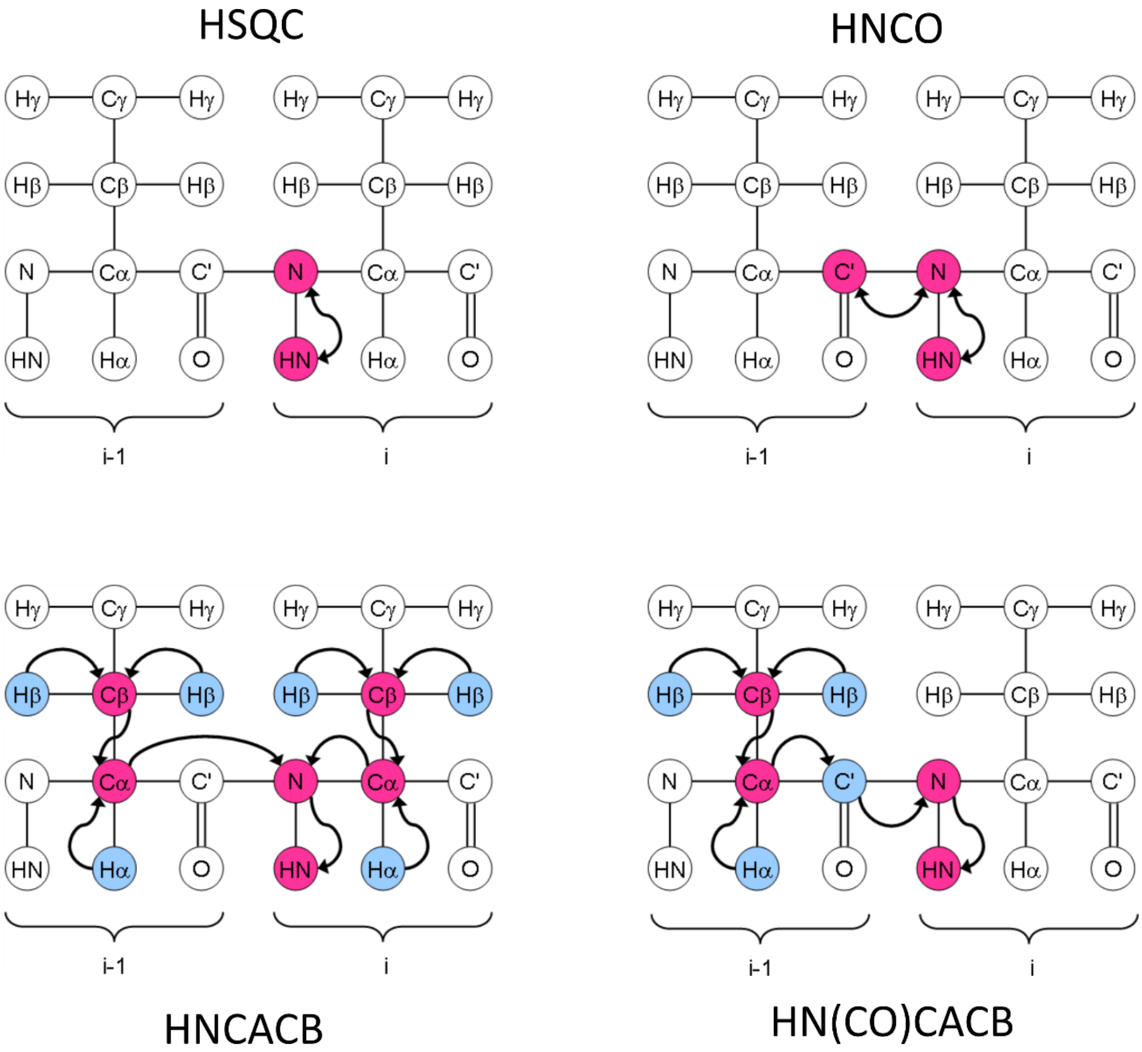


Figure 2-2: Representation of the transfer of magnetisation and atoms detected in the HSQC, HNCO, HNCACB and HN(CO)CACB NMR experiments. The pink atoms are observed in the resultant spectra and the light blue atoms allow magnetisation to flow but are not detected. These images are taken from Dr. Vicky Higman. 2012. Protein NMR: A Practical Guide. Available at: http://www.protein-nmr.org.uk/. [Accessed 15 July 15].

2.1.2.3. Triple resonance backbone assignment

Triple resonance spectra recorded from a uniformly ^{15}N, ^{13}C labelled protein are the basis of backbone resonance assignment. These experiments consist of a 2D ^{1}H, ^{15}N plane expanded into the carbon dimension, by transferring magnetisation over the peptide bond. This process allows for separate spin systems to be connected through bonds and correlated back to their HSQC peaks. Note, in protein NMR a spin system refers to the group of nuclear spins associated with a single residue. A standard suite of 3D experiments includes, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB and CBCA(CO)NH. These spectra usually start and end on a proton to ensure appropriate sensitivity. In some cases the magnetisation is transferred back to the starting proton for collection and this type of experiment is termed 'out and back'. These specific experiments are paired to build up this 'logic puzzle' of disparate resonances. However, the full suite of 3D experiments does not need to be collected on a sample as a full sequential assignment can often be completed with any two complementary spectra. Recording extra pairs of spectra clarifies any assignment ambiguity while adding an extra layer of confidence to the assignment.

Conventional triple resonance assignment is based on pairing the HNCACB and CBCA(CO)NH spectra, although other pairings are available (Figure 2-3). These spectra correlate Cα and Cβ resonances with one amide resonance. The CBCA(CO)NH spectra, also termed HN(CO)CACB, relates the resonance of the i residue amide with the Cα and Cβ resonances of the previous residue (Grzesiek and Bax. 1992a). Two peaks corresponding to the Cα and Cβ are therefore visible for each residue. The HNCACB spectra, links the resonance of the i residue amide with the Cα and Cβ resonances of both the i residue and the previous (i-1) residue (Grzesiek and Bax. 1992b). Notably, the peaks relating to the i-1 residue are weaker. In both experiments the Cα and Cβ resonances are in opposite phase, meaning if one is positive then the other will be negative. This enables simple resolution of any Cα and Cβ ambiguity.

To aid visualisation, the spectra are analysed as 2D strips, allowing strips from the pair of experiments to be correlated with resonances observed in the HSQC, compared or overlaid to facilitate the assignment (Figure 2-3). Four peaks are present in the strip of the HNCACB spectra, the Cα and Cβ resonances of the i residue and those of the i-1 residue, which are typically weaker in intensity. The peaks relating to the previous residue can be identified with the equivalent strip from the CBCA(CO)NH dataset. These strips of peaks can be sequentially linked to adjoining residues based on the positions of the Cα and Cβ resonances from both experiments. Certain residues (serine, threonine, glycine and alanine) have distinctive Cα and Cβ carbon chemical shifts and can be easily identified, which can aid the assignment process. The linked resonances can then be assigned to the protein sequence.

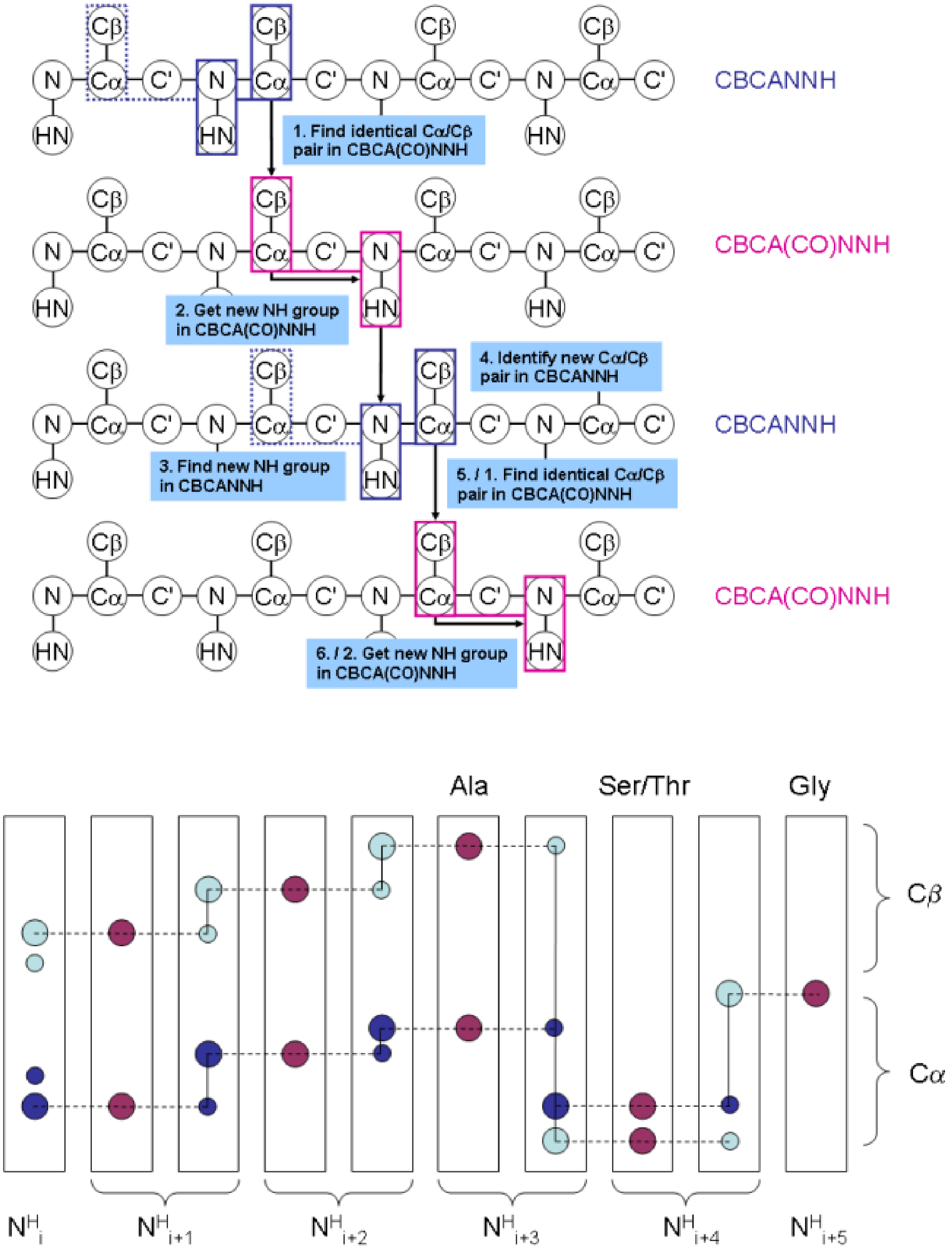


Figure 2-3: Schematic representation of HNCACB and CBCA(CO)NH based triple resonance assignment. Top) the HNCACB (blue) and CBCA(CO)NH (pink) spectra correlate Cα and Cβ resonances with the amide backbone allowing one amide group to be linked to another sequentially. Bottom) representation of the 2D strips used to visualise and link the HNCACB (blue) and CBCA(CO)NH (pink) spectra. To differentiate the resonances of the HNCACB spectra the Cα resonances are shown in dark blue and the Cβ resonances in light blue. The distinctive chemical shifts observed for alanine, glycine, serine and threonine are noted to show how they give sequence context to an assignment. Both images are taken from Dr. Vicky Higman. 2012. Protein NMR: A Practical Guide. Available at: http://www.protein-nmr.org.uk/. [Accessed 15 July 15].

2.1.3. Selective metabolic unlabelling

Proton-detected 2D and 3D NMR experiments are the basis of backbone resonance assignment. Deconvolution of resultant data is key to effective assignment. Spectral congestion, and consequently signal overlap, is a problem when assigning intrinsically disordered proteins because of the limited dispersion of resonances in the ^{1}H dimension, which is still problematic when taken into the ^{13}C dimension. Correlating residue type within the context of a protein sequence with spin systems is difficult as the constant chemical environment gives little variation to the Cα/β chemical shifts. As already mentioned some residues have highly distinctive Cα and Cβ resonances, which can be used to link a resonance to a specific nucleus in a protein. However, in cases of low sequence complexity and/ or signal overlap due to disorder this information is not sufficient for a complete assignment. Specific metabolic precursors can be used to selectively unlabel certain residues in an otherwise uniformly labelled sample rendering them spectrally silent (Rasia et al. 2012). These precursors can be added as a supplement to regular M9 minimal media and combined with standard (^{13}C, ^{15}N) uniform labelling approaches. Thus, producing precise 'NMR invisible' residues in an otherwise uniformly labelled sample. Comparison of these spectra with those from uniformly labelled samples allows the user to assign a residue type to missing resonances. In particular, collection of a ^{1}H, ^{15}N HSQC spectrum on an unlabelled sample renders the selected residue 'NMR invisible'. 2D ^{1}H, ^{15}N HNCO spectrum silences the i and i+1 residue, highlighting the amount of information that can be obtained from this method. This strategy allows the user to control which residues are unlabelled and gives predictable labelling patterns, significantly aiding the backbone assignment of challenging proteins.

2.2. Experimental

2.2.1. Circular dichroism of HASPA and HASPB1

HASPA and HASPB1 were prepared as described in Section 4.2 and evaluated using the Jasco J810 CD Spectrophotometer installed at the Technology Facility, Department of Biology, University of York. The samples were analysed at 0.2 mg/ml in 20 mM sodium phosphate pH 7 with 50 mM sodium chloride in a 1 mm quartz cuvette. The buffer was filtered (pore size: 0.22 µm) before use. Data were collected between 195 and 260 nm.

2.2.2. Isotopic labelling of HASPA and HASPB2

Minimal media were used to express ^{15}N labelled HASPA and HASPB2 and ^{13}C, ^{15}N labelled HASPA. The M9 base solution was prepared (6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl and 1 g $^{15}NH_4Cl$ in 1 l) and autoclaved. ^{15}N ammonium chloride (99%) was obtained from Cambridge Isotope Laboratories. Kanamycin was added at a final concentration of 50 µg/ml and the remaining media components added (see Table 2-1). The vitamin solution (0.22 µM filter sterilised) was obtained from Alex Heyam (University of York) and stored at 4˚C with protection from exposure to light. A 50 ml aliquot of the vitamins solution contains 625 mg of thiamine, 12.5 mg riboflavine and 125 mg of each of the following, pyridoxine, biotine, panthothenate, folic acid, choline chloride, and niacinamide. D-Glucose (*U*-$^{13}C_6$) from Cambridge Isotope Laboratories was used to supplement the minimal media to produce [*U*-^{15}N ^{13}C labelled HASPA].

Component	Volume in 1 L	Component	Volume in 1 L
Glucose (20% w/v)	10 ml	50 mM $ZnSO_4$	1 ml
1M $MgSO_4$	1 ml	0.1 M $FeCl_3$	0.6 ml
0.1 M $CaCl_2$	1 ml	Vitamins	2 ml
0.1 M $MnCl_2$	1 ml		

Table 2-1: Components added to minimal media after base solution autoclaving. When ^{13}C, ^{15}N labelled HASPA was prepared D-Glucose (*U*-^{13}C6) was used instead.

The expression was carried out as follows, 10 ml aliquots of LB media supplemented with kanamycin were inoculated with *E. coli* BL21 (DE3) pET28b_haspa or pET28b_haspb2 glycerol stock. These cultures were grown for 8 hours at 37°C with shaking at 180 rpm and then the optical density at 600 nm (OD_{600}) was measured. In order to inoculate the minimal media with the appropriate amount of culture, an aliquot was taken and centrifuged that, once resuspended, would give a starting OD_{600} of 0.05 in 50 ml minimal media. The 50 ml culture was incubated overnight at 37˚C with shaking at 200 rpm. The OD_{600} was measured again to give the correct starting volume for 1 litre M9 media of overnight culture. The cultures were grown to OD_{600} of 0.8 to 1.0 at 37°C and 180 rpm. Protein expression was induced with addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 1mM and the cultures were transferred to 30°C for 6 hours with shaking at 180 rpm. The cells were harvested by centrifugation for 15 minutes at 5000 xg and the resulting pellets were stored at -20˚C.

2.2.3. Metabolic unlabelling of HASPA

The production of the metabolic unlabelled HASPA samples was similar to ^{15}N, ^{13}C HASPA production (see Section 2.2.2) but the protocol modifications are noted here. At OD_{600} 0.7 the 750 ml culture was split into three autoclaved 500 ml flasks, giving a final volume of 250 ml in each. At this stage the specific metabolic precursors were directly added to the culture in powder form and the flasks returned to 37˚C until an optical density of 1.0 was reached and then induced with IPTG (see Table 2-2). The three cultures were 1) lysine, 2) arginine and 3) isoleucine and valine unlabelled.

Precursor	Amount per litre of culture	Amino acid unlabelled
Arginine	200 mg	Arginine
Lysine	200 mg	Lysine
α-ketobutyrate (CAS number: 600-18-0)	100 mg	Isoleucine
α-ketoisovalerate (CAS number: 3715-29-5)	150 mg	Valine Leucine

Table 2-2: Metabolic precursors used to achieve amino acid specific unlabelling.

2.2.4. Purification of labelled proteins

The stored pellets were thawed and resuspended in 30 ml nickel column buffer A (50 mM Tris pH 7.5, 500 mM sodium chloride, 20 mM imidazole) per 500 ml pellet supplemented with a protease inhibitor tablet (cOmplete, EDTA-free from Roche). The soluble cell suspension was sonicated 6 times on ice for 30 seconds with 1 minute intervals. Lysate clarification was achieved by centrifugation at 38000 xg for 30 minutes. The supernatant was loaded onto a 1 ml HisTrap FF crude column (GE Healthcare) equilibrated with nickel column buffer A and washed with 6 column volumes of the same buffer. The target protein was eluted using a gradient of 0 to 100% nickel column buffer B (50 mM Tris pH 7.5, 500 mM sodium chloride, 500 mM imidazole) over 20 column volumes. SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) was used to identify target protein containing fractions that were concentrated (Vivaspin Sample Concentrator range, 3000 MWCO) to 500 µl for further purification with size exclusion chromatography. A Superdex 75 10/300 column (GE Healthcare) equilibrated with 20 mM Hepes pH 6.5, 50 mM sodium chloride was used to further purify the HASPs. Final sample purity was analysed with SDS PAGE. To improve the purity of HASPB2, an ion exchange step was introduced after the nickel affinity purification step. The HASPB2 sample was diluted with 20 mM Tris pH 7.5 to give a final concentration of 50 mM NaCl and loaded onto a 1 ml HiTrap Q HP column equilibrated with ion exchange buffer A (20 mM Tris pH 7.5, 50 mM NaCl). HASPB2 was eluted with a gradient of 0 to 100% ion exchange buffer B (20 mM Tris pH 7.5, 1 M NaCl) for 20 column volumes. The entire purification schedule was run with the AKTA FPLC system.

The initial ^{1}H, ^{15}N-HSQC experiments with ^{15}N labelled HASPA and HASPB2 were recorded at a concentration of 250 µM in 20 mM sodium cacodylate pH 6.5, 120 mM NaCl. 3D data collection with ^{13}C, ^{15}N HASPA was at 900 µM in 20 mM Hepes pH 6.5, 50 mM NaCl. Directly before data collection D_2O was added to all the samples to give a final concentration of 10%.

2.2.5. Data collection and processing

The spectra presented in this thesis were collected by Dr. Michael Plevin at the Centre for Magnetic Resonance, University of York using the 700 MHz Bruker AV spectrometer at 298K. The preliminary ^{1}H, ^{15}N HSQC experiments were recorded with ^{15}N labelled HASPA and HASPB2. Triple resonance experiments, namely HNCACB and HN(CO)CACB spectra, were collected for the backbone assignment with ^{13}C, ^{15}N labelled HASPA at 0.9 mM. To compare the metabolic unlabelled samples with the uniformly labelled ^{13}C, ^{15}N HASPA two spectra were collected; ^{1}H, ^{15}N HSQC and 2D ^{1}H, ^{15}N HNCO. All NMR data were processed with NMRpipe by Dr. Michael Plevin.

2.2.6. Backbone assignment

The spectra were analysed with CCPN Analysis version 2.4 and the assignment was performed manually with this software (Vranken et al. 2005). Briefly, peaks were picked in the HSQC spectrum above an acceptable user defined contour threshold and amide side chain peaks removed. The 3D data were linked by navigating from a HSQC peak to the NH dimension in the HNCACB and HN(CO)CACB experiments. The Cα and Cβ carbon chemical shifts of the i and i-1 residues were picked apart by comparing the HNCACB and HN(CO)CACB spectra and linked to the HSQC. Alanine, glycine, serine and threonine all have distinctive Cα and Cβ carbon chemical shifts so spin systems associated with these residues were noted. Arginine, lysine, isoleucine and valine residues were defined by the HSQC and 2D HNCO spectra from the metabolic unlabelling work. The spin systems were linked by matching the i Cα and Cβ peaks from the HNCACB with the i-1 Cα and Cβ peaks from the HN(CO)CACB in the context of the HASPA sequence. Backbone phi and psi angles were estimated with TALOS-N to determine if the HASPA sequence contained any secondary structure elements (Shen and Bax. 2013).

2.3. Results

2.3.1. The HASPs are predicted to be intrinsically disordered

Disorder prediction is the first step in defining an intrinsically disordered protein. Extensive analysis of ordered and disordered proteins highlighted their distinct sequence bias. Residues could then be classed as order-promoting, disorder-promoting or neutral. The order-promoting residues are C, W, Y, I, F, V, L, (to a lesser extent H, T, and N), the disorder-promoting residues are D, M, K, R, S, Q, P, and E, and the neutral residues are A and G (Radivojac et al. 2007). Unsurprisingly, hydrophobic residues were shown to be a consistent feature of ordered proteins due to their integral role in core stabilisation. The disorder promoting class of amino acids is populated by charged and polar residues, producing the distinctive sequence composition of IDPs. Negative hydropathicity is also characteristic of IDPs because of the relatively low abundance of hydrophobic residues in their sequences. Bioinformatic analysis to determine sequence motifs found in IDPs indicated a high prevalence of proline and both positive and negative charged motifs (Lise and Jones. 2005). Analysis of IDP sequences highlighted their low sequence complexity when compared with globular proteins (Rezaei-Ghaleh et al. 2012). This is typified by the high occurrence of tandem repeats in IDP sequences. Repeats were shown to be a dominant characteristic of IDPs (39% of all IDPs contain repeats) when compared with all Swiss-Prot (14%), yeast (18%) or human (28%) proteins (Tompa. 2003). Taken in combination, signifiers of disorder from a sequence level are the prevalence of disorder-promoting residues, tandem repeats, highly charged or proline-rich motifs, high hydrophilicity and low sequence complexity.

Examination of the HASP sequences based on the described IDP features above suggests disorder. Analysis of the HASPB1 sequence highlights the prevalence of disorder promoting residues, with charged residues and proline comprising 48% of the entire sequence (Figure 2-4). All three HASP sequences exhibit low sequence complexity indicative of disorder, this is particularly apparent with HASPB1 and HASPB2 due to their repetitive nature. Unsurprisingly, the HASPs are extremely hydrophilic, another feature common in disordered proteins. The grand average of hydropathicity (GRAVY) reported by ProtParam (ExPasy) for HASPB1 is -2.3, where a negative value reports a highly hydrophilic sequence (Gasteiger et al. 2005). More stringent analysis was required to more confidently predict disorder. IUPred, an IDP predictor, estimates the energy of pairwise interactions in a window around a residue assuming that an ordered protein is composed of residues that can form favourable stabilising interactions while an IDP cannot (Dosztányi et al. 2005). A disorder tendency score of 0 predicts a folded protein while a score of 1 indicates intrinsic disorder. It is clear from Figure 2-4 that when the HASP sequences are analysed by IUPred they are all predicted to be significantly disordered with no element of order observed. Confirmation of this predicted disorder was required to accurately classify the HASPs.

HASPB1 sequence

```
001: GAYSTKDSAK EPQKRADNIH KTTEANHGGA TGVPPKHTGS AMNDSAPKED
051: GHTQKNDGDG PKEDGHTQKN DGDGPKEDGH TQKNDGDGPK EDGHTQKNDG
101: DGPKEDGHTQ KNDGDAPKED GRTQKNDGDG PKEDGHTQKN DGDAPKEDGR
151: TQKNDGDGPK EDGHTQKNDG DGPKEDGHTQ KNDGDGPKED GRTQKNDGGG
201: PKEDGHTQKN DGDGPKEDGH TQKNDGDGPK EDGHTQKNDG DAPKEDGRTQ
251: KNDGDGPKED GHTQKNDGDG PKEDGHTQKN DGDAPKEDGR TQKNDGDGPK
301: EDGHTQKNDG DGPKEDGRTQ KNDGDGPKED GHTQKNDGDG PKEDGRTQKN
351: DGDAPKEGEN LQQNDGDAQE KNEDGHNVGD GANGNEDGND DQPKEHAAGN
```

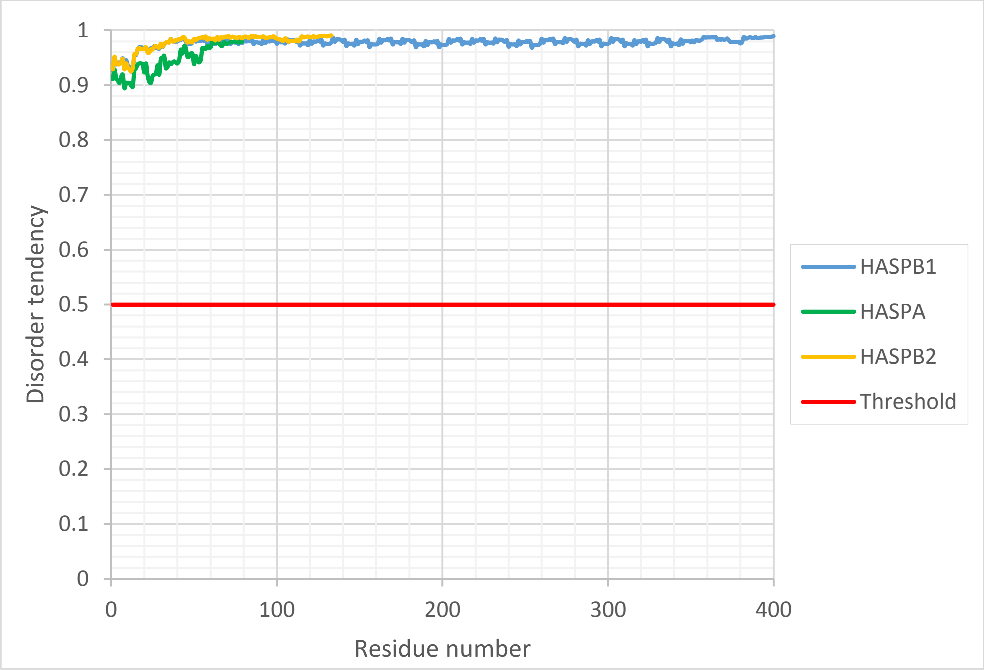


Figure 2-4: Sequence-based analysis of the disordered nature of the HASPs. (Top) Annotated amino acid sequence of HASPB1 emphasizing the prevalence of the disorder promoting proline (pink), basic (blue) and acidic (turquoise) residues. (Bottom) IUPRED disorder prediction of all three HASPs. The disorder threshold is shown in red at 0.5. Anything below this threshold is considered folded, while anything above is not.

2.3.2. Circular Dichroism shows the HASPs are disordered

Circular Dichroism (CD) Spectrophotometry is a technique used to quantify the interaction of a molecule of interest with polarised light. The chiral arrangement of peptides bonds in specific protein secondary structures (e.g. α helices and β sheets) generate characteristic spectra at defined wavelengths. This allows the user to potentially assign secondary structure characteristics of an unknown protein sample based on the spectra obtained.

Analysis of the amino acid composition of the HASPs in Section 2.3.1. revealed a propensity for disorder. However, confirmation of this predicted disorder by biochemical and biophysical methods was required. Circular dichroism (CD) spectroscopy utilises experimentally defined spectroscopic signatures from proteins with known three-dimensional structures to characterise the secondary structure content of sample proteins. To obtain a greater insight into the predicted disorder of the HASPs, CD spectra of HASPA and HASPB1 were collected. Both spectra are typical of random coil conformations with the minimum seen at 195 nm (Figure 2-5). Note data were collected between 195 and 260 nm. The CD signal was considered valid between these points, as the HT voltage was below 600 but saturated at a wavelength lower than 195 nm. DichroWeb is a web-based server that implements different empirical algorithms derived from reference spectra to analyse the secondary structure content of input data (Whitmore and Wallace. 2008). K2d method (Andrade et al. 1993) analysis indicates that HASPA is 92% random coil, confirming this predicted feature.

CD spectra of intrinsically disordered proteins are often shown to be random coil due to their lack of β sheet, α helix, or turn elements. However, this grouping also includes polyproline II (PPII) structures such as collagen. A PPII helix is typified by a high proline content and is defined by phi and psi angles of -75° and 150°, respectively. To distinguish disordered samples from PPII structures Lopes et al. (2014) compared the spectra of native and denatured collagen. They found that the single most characteristic spectroscopic feature distinguishing a (PPII) structure from a disordered structure was a positive peak around 220 nm in PPII samples. Interestingly, this feature is not observed in either HASP spectra strengthening the case for intrinsic disorder classification.

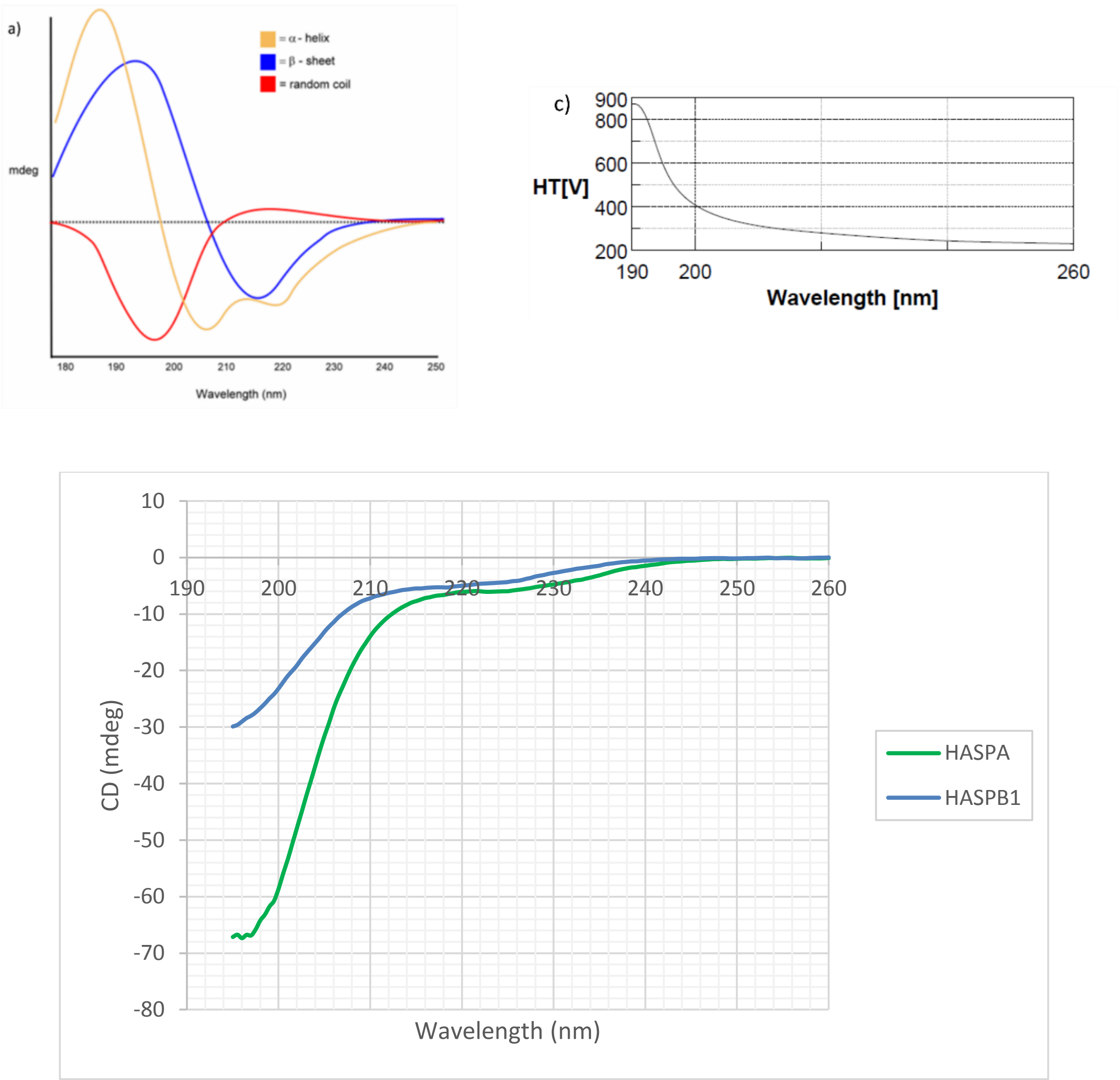
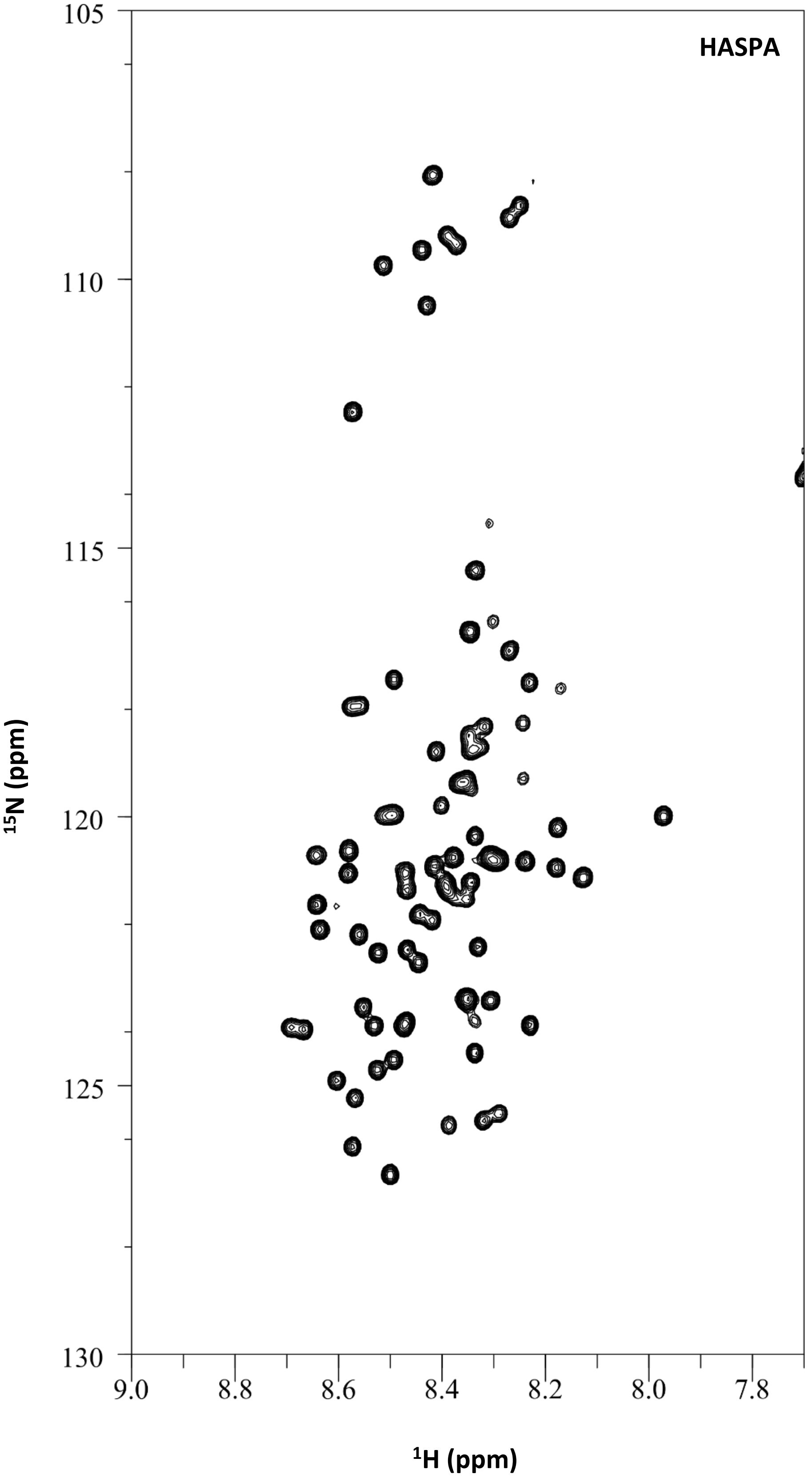


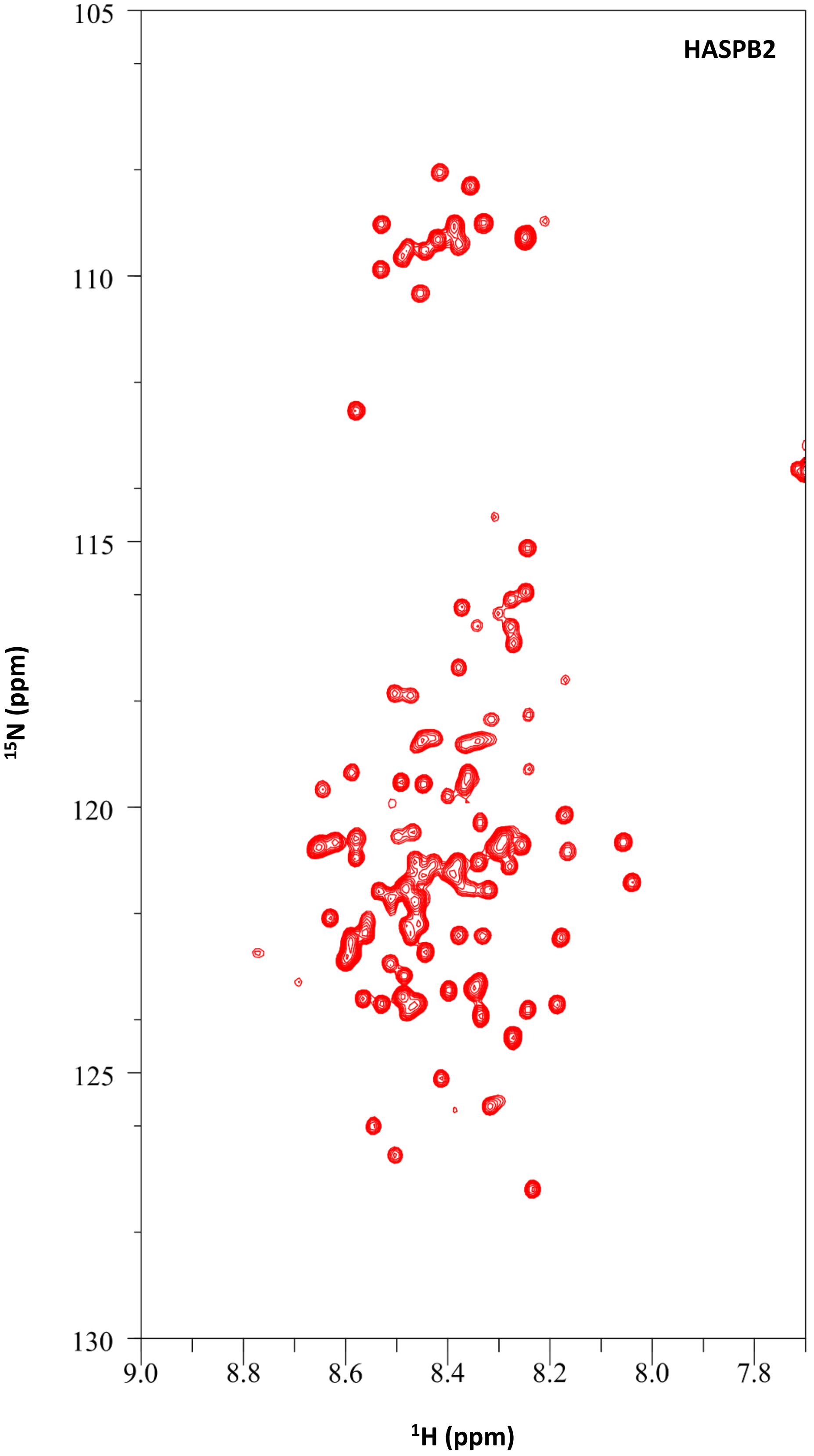
Figure 2-5: CD spectra of HASPB1 and HASPA. a) Example CD spectra of alpha-helical (yellow), beta-sheet (blue) and random coil (red). This image is taken from Dr. Ramy S. Farid. 2006. Circular Dichroism (CD) Spectroscopy. Available at: http://www.proteinchemist.com/cd/cdspec.html. [Accessed 12 August 15]. b) CD spectra of HASPB1 and HASPA. These spectra are indicative of the random coil spectral feature. c) HT voltage trace.

2.3.3. ^{1}H, ^{15}N-HSQC of HASPA and HASPB2

To confirm the predicted intrinsic disorder of the HASPs, uniformly ^{15}N labelled HASPA and HASPB2 were prepared and characterised by NMR spectroscopy. The overlaid ^{1}H, ^{15}N HSQC spectra highlight the limited dispersion of NMR resonances in the ^{1}H dimension around 8.4 ppm (Figure 2-6). This spectral feature is indicative of highly disordered proteins because regions of high backbone flexibility have reduced chemical shift dispersion as they all share the same solvent exposed environment (Kragelj et al. 2013). When samples contain α-helical or β-sheet elements, the peaks are much more widely distributed. Despite the extensive clustering of resonances in the HSQC for both proteins and signal overlap the NMR resonances, are on the whole, well resolved, meaning that backbone assignment should be possible. Glycine has a distinctive nitrogen chemical shift in an HSQC experiment when compared with other residues. Glycine resonances are usually found in the range of 105 to 115 ppm, which occurs because glycine lacks a Cβ. For HASPA, a protein with 10 glycine residues, 9 candidate peaks are visible in the HSQC. The extreme terminal residues are not always clearly defined in HSQC spectra, so it is expected that the N-terminal glycine would not be visible.

Direct visual analysis of the HSQC data shows that HASPA has 73 clearly defined resonances, while it has 78 residues that could be assigned, excluding 3 proline residues and the hexa-histag. This is a promising starting point for assignment. HASPB2 has 127 residues that could be assigned but only 95 clearly defined resonances. This is likely due to the presence of repeats in HASPB2 combined with the limited dispersion of resonances in IDP samples. These residues would be experiencing the same (solvent exposed) chemical environment, producing overlaid but intense resonances for the repeated sequence motif. Therefore, it was decided that HASPA was the best candidate for assignment because it is smaller and the number of observed resonances was closer to the number of residues when compared with the data for HASPB2. Moreover, working with a protein without repeats removes a potential layer of convolution to the assignment.


HASPA
105
110
115
120
125
130
^{15}N (ppm)
9.0
8.8
8.6
8.4
8.2
8.0
7.8
^{1}H (ppm)


HASPB2
105
110
115
120
125
130
15N (ppm)
9.0
8.8
8.6
8.4
8.2
8.0
7.8
1H (ppm)

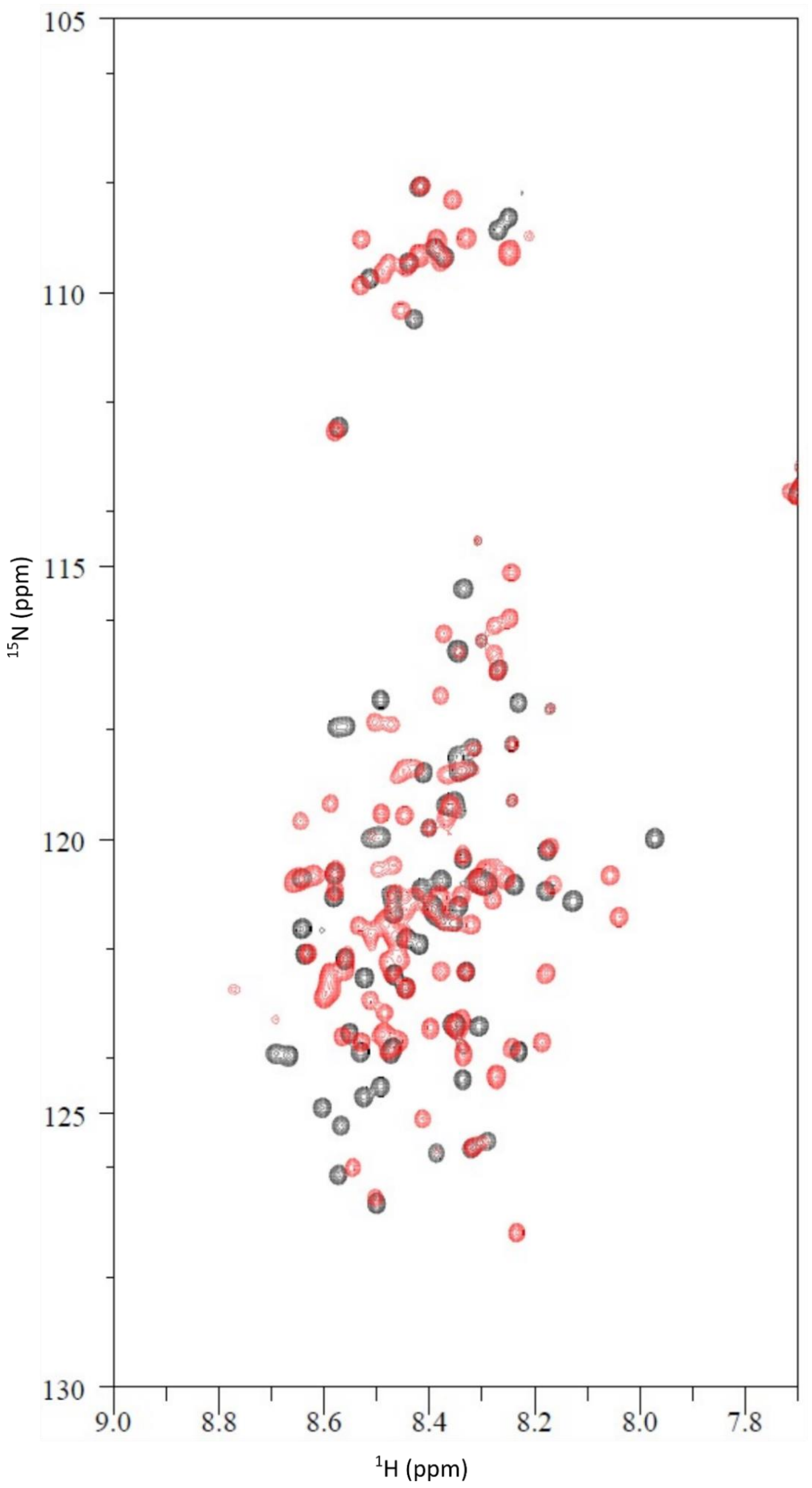


Figure 2-6: ^{1}H, ^{15}N-HSQC of HASPA and HASPB2 recorded at 298K in 20 mM sodium cacodylate pH 6.5, 120 mM sodium chloride. HASPA is shown in black and HASPB2 in red. The limited dispersion of both samples in the proton dimension is indicative of disorder.

2.3.4. Backbone assignment of HASPA

^{15}N, ^{13}C uniformly labelled HASPA was prepared at 0.9 mM in 20 mM Hepes pH 6.5, 50 mM NaCl. Initially, two 3D experimental datasets were collected, HNCACB and HN(CO)CACB spectra, for backbone resonance assignment. These experiments expand the 2D ^{1}H, ^{15}N HSQC into a third ^{13}C dimension to correlate the NH plane from the HSQC spectra with the $^{13}C\alpha$ and $^{13}C\beta$ resonances. Specifically, the HNCACB spectrum reports strong chemical shifts for the Cα and Cβ of the i residue and weakly for the previous (i-1) residue. The CBCA(CO)NH spectrum reports only the Cα and Cβ chemical shifts of the i-1 residue.

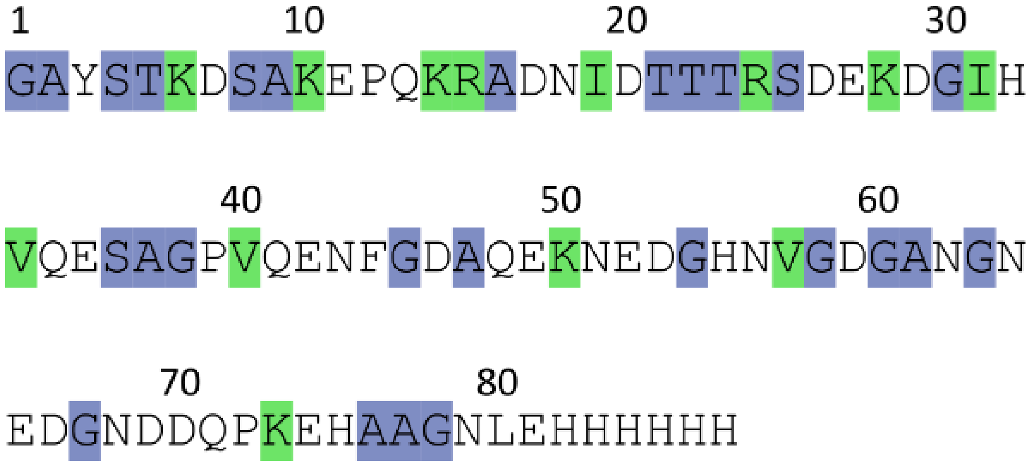


Figure 2-7: Annotated HASPA sequence showing the combined approach implemented for backbone assignment. Residues with distinctive Cα/ Cβ chemical shifts (alanine, glycine, serine and threonine) and those chosen for metabolic unlabelling (valine, isoleucine, lysine and arginine) are highlighted in blue and green, respectively.

To begin, the 2D ^{1}H, ^{15}N HSQC spectra were initialised by picking peaks and creating resonances that are the starting point for spin system groupings. Peaks relating to amide containing side chains were removed from the analysis at this point. The peak list generated from the HSQC spectrum was then used to navigate to 2D strips of the 3D experiments to link the Cα and Cβ carbon chemical shifts to the amide dimension. Alanine, serine, threonine and glycine have distinctive Cα and Cβ chemical shifts, enabling residue type information about specific spin systems to be gained. A note was made of i and i-1 residues that belonged to one of these residue types to give the assignment some sequence context later (Figure 2-7). However, due to low sequence complexity and the disordered nature of HASPA identifying these residues was not sufficient to firmly link any sequentially assigned portions to the sequence. HASPA is disordered in solution, so all residues in the sequence experience the same solvent exposed chemical environment. This means that little variance is observed in Cα and Cβ resonances of the same residue type in different sequence positions and there is little observable difference between residues with similar Cα and Cβ resonances. The metabolic precursor unlabelling strategy (described in Section 2.1.3.) was implemented to obtain residue-specific information to deconvolute the assignment process. This combined approach allowed for specific residues to be identified and their context within the HASPA sequence enabled backbone assignment (Figure 2-7).

A detailed description of the protocol implemented to produce various metabolic unlabelled ^{15}N, ^{13}C HASPA samples can be found in Section 2.2.3. However, it is important to note the simplicity of this protocol. It only requires splitting a pre-induction expression culture and adding the required metabolic precursors. Afterwards the expression protocol was continued as normal. This resulted in three metabolic unlabelled ^{15}N, ^{13}C HASPA samples that were 1) lysine, 2) arginine and 3) isoleucine and valine unlabelled. Isoleucine and valine were unlabelled together to streamline sample production. The precursors used and the labelling pattern they produce in the final sample are outlined in Figure 2-8.

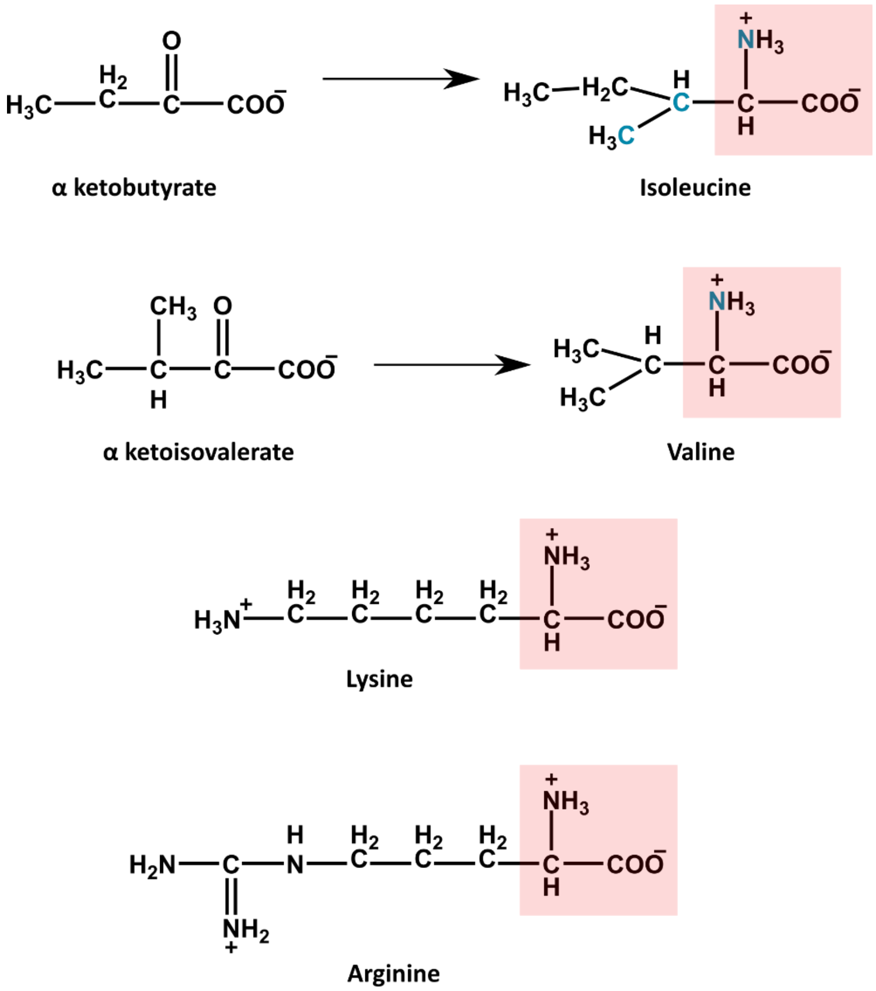


Figure 2-8: Precursors used for metabolic unlabelling of HASPA. Early stage precursors of isoleucine and valine are α-ketobutyrate and α-ketoisovalerate, respectively. While lysine and arginine can be directly added to the growing cultures to unlabel these residues (Hiroaki et al. 2011). Nitrogen and carbon atoms coloured blue denote that these atoms are ^{15}N or ^{13}C labelled by the media.

Lysine and arginine can be directly supplemented into ^{15}N, ^{13}C minimal media to unlabel these specific residues (Hiroaki et al. 2011). In a 2D ^{1}H, ^{15}N HSQC spectrum both these residues will be 'NMR invisible' in an otherwise uniformly labelled sample. Therefore, comparison of these spectra with a [*U*-^{15}N] HASPA ^{1}H, ^{15}N HSQC spectrum allows the investigator to determine which spin systems relate to the 'NMR invisible' residues. HASPA contains 2 arginine and 6 lysine residues that were identified through this unlabelling approach. Moreover, collection of 2D ^{1}H, ^{15}N HNCO spectra on these unlabelled samples renders the i+1 residues 'NMR invisible', supplying the assignment with more residue specific information. Figure 2-9 highlights how the two arginine residues, R15 and R24, were identified when the HSQC spectra were compared.

Moreover, the two i+1 residues were easily distinguished based on the difference between the Cα/ Cβ resonances of serine (residue 25) and alanine (residue 16).

As already mentioned, there are 6 lysine residues in the HASPA sequence. The lysine unlabelled 2D ^{1}H, ^{15}N HSQC spectrum was used to identify these lysine spin systems when compared with the [*U*-^{15}N] HASPA ^{1}H, ^{15}N HSQC spectrum (Figure 2-10). Cross referencing the arginine unlabelled ^{1}H, ^{15}N HSQC with the lysine unlabelled ^{1}H, ^{15}N HNCO allowed firm identification of arginine 15, as it is the i+1 of lysine 14. This produced a vital starting point 'anchor' in the HASPA sequence from which to build the rest of the assignment. However, due to the highly similar Cα and Cβ resonances of aspartic acid and asparagine, identification of residues Asp7, Asp29 and Asn51 was more difficult. The sequence context of the residues was used to determine which spin systems related to the specific aspartic acid or asparagine residues. Glutamic acid has a more distinctive Cα/ Cβ resonance distribution, allowing residue 11 to be assigned directly. No unlabelling related to glutamic acid 74 was observed in the ^{1}H, ^{15}N HNCO. It appeared that the resonance next to Asp7 could be promising as this was unlabelled in this experiment, however, there was no convincing 3D data to confirm this assumption.

Valine and isoleucine residues produced by the metabolic unlabelling strategy are visible in a 2D ^{1}H, ^{15}N HSQC spectra. Enzymatic conversion of α-ketobutyrate and α-ketoisovalerate into isoleucine and valine, respectively, renders the α-amino group of both residues ^{15}N labelled (Rasia et al. 2012; Figure 2-8). However, the remaining carbon atoms are ^{12}C labelled apart from the Cβ and Cγ2 atoms in the isoleucine side chain, which are ^{13}C labelled. This means the i+1 residues could be ascertained from the isoleucine and valine unlabelled 2D ^{1}H, ^{15}N HNCO spectrum when compared with the ^{1}H, ^{15}N HSQC spectrum of [*U*-^{15}N] HASPA (Figure 2-11). There are five residues in the HASPA sequence that are i+1 to either valine or isoleucine. Therefore, this double unlabelled sample significantly contributed to the overall assignment. The high occurrence of glutamines (10.1%) and aspartic acids (13.9%) in the HASPA sequence made assignment of these residues (D20, Q34 and Q41) extremely difficult. Three of these residues are i+1 to isoleucine or valine so were fixed in the sequence and removed from the pool of unassigned spin systems. This made it much easier to identify other non-associated residues and streamlined the rest of the assignment.

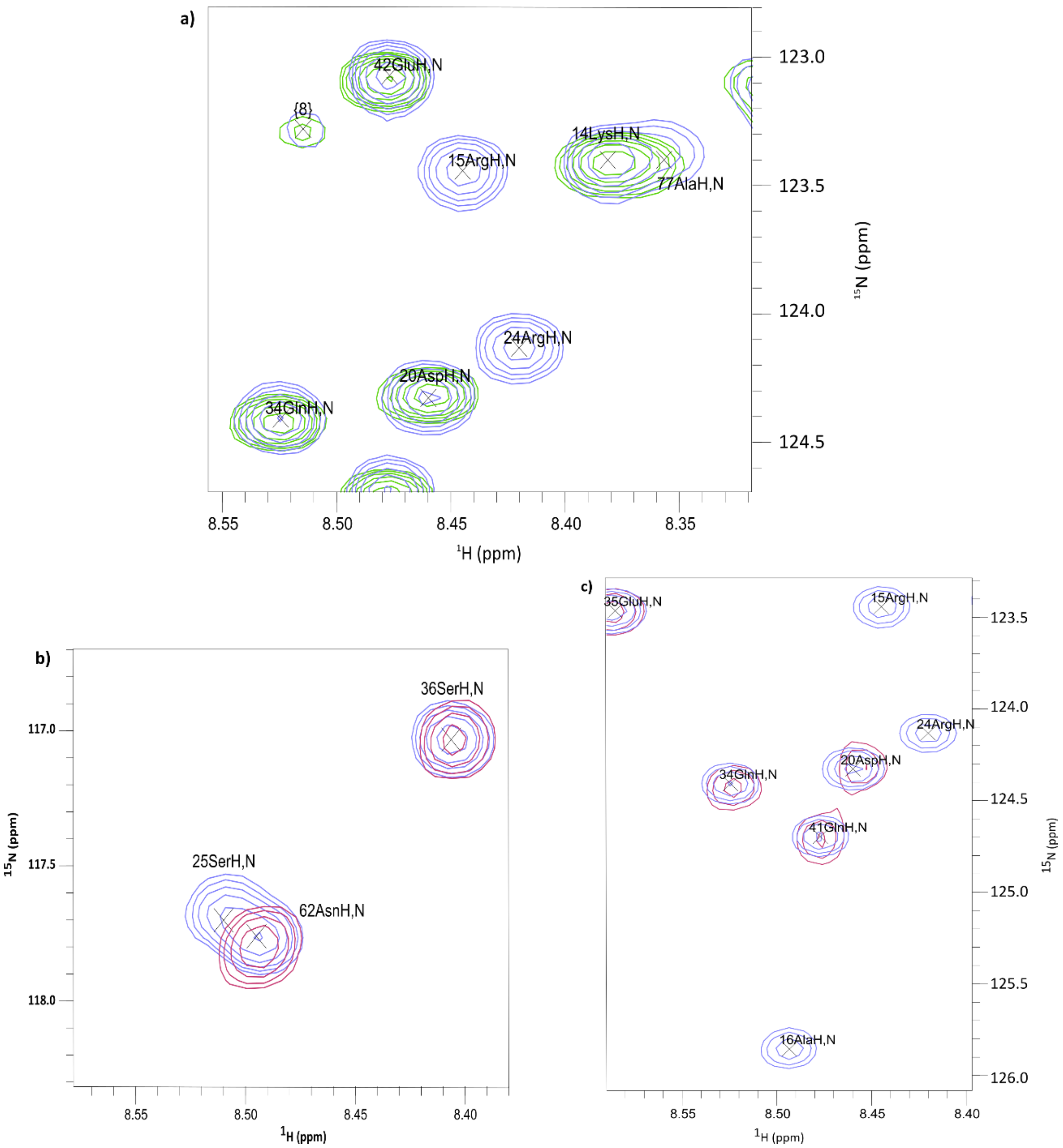


Figure 2-9: Selective arginine unlabelling of HASPA. a) Overlay of ^{1}H, ^{15}N HSQC of uniformly ^{15}N-labelled HASPA (blue) and arginine unlabelled HASPA (green) highlighting R15 and R24. b and c) Overlay of ^{1}H, ^{15}N HSQC of uniformly labelled HASPA (blue) and 2D ^{1}H, ^{15}N HNCO (pink) showing both arginine residues and the corresponding i+1 residues.

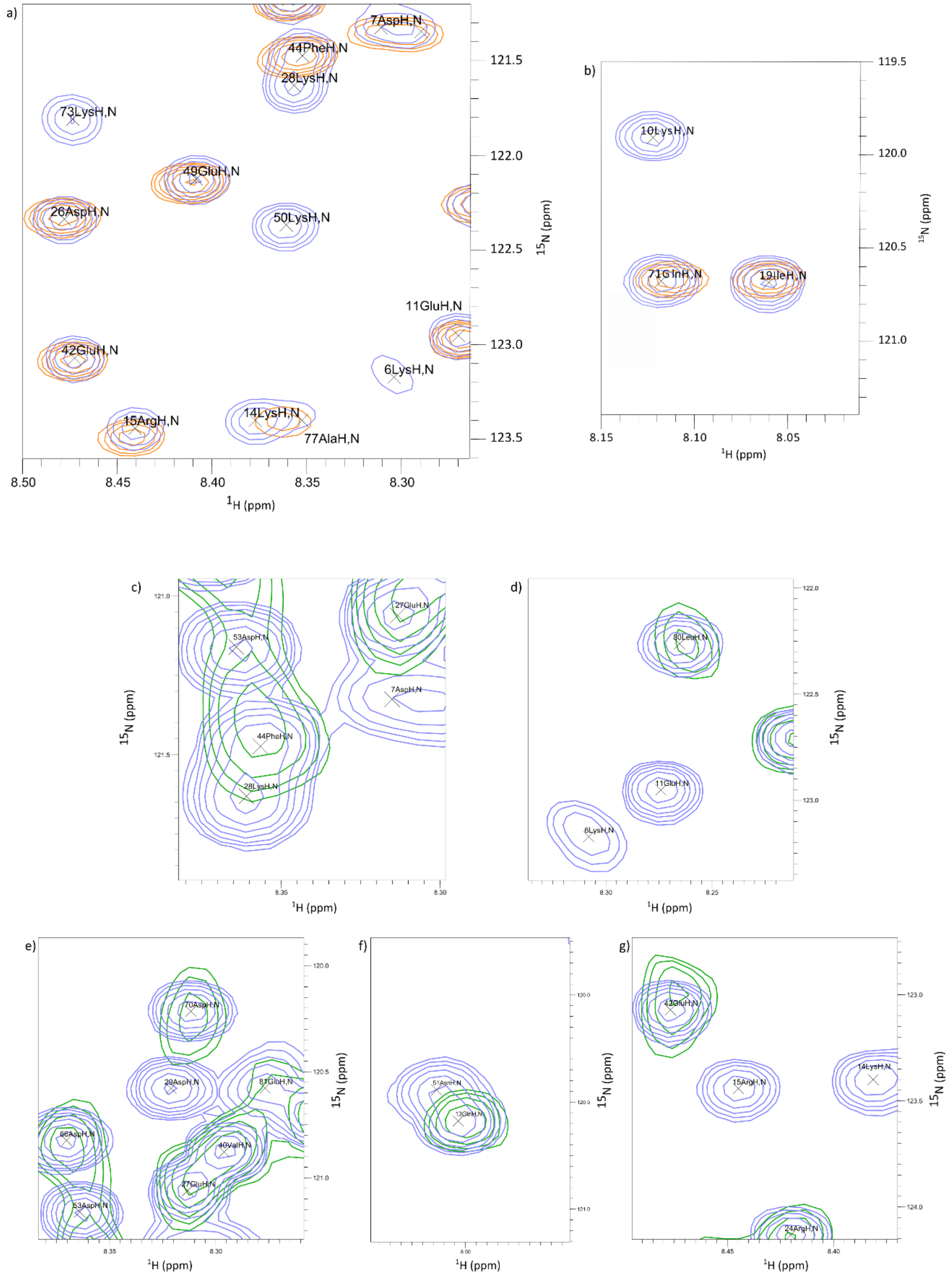


Figure 2-10: Selective lysine unlabelling of HASPA. a) and b) Overlay of ^{1}H, ^{15}N HSQC of uniformly ^{15}N-labelled HASPA (blue) and lysine unlabelled HASPA (orange) highlighting the six lysine residues found in HASPA. c-g) Overlay of ^{1}H, ^{15}N HSQC of uniformly labelled HASPA (blue) and 2D ^{1}H, ^{15}N HNCO (green) showing some lysine residues and the corresponding i+1 residues. There should be 6 i+1 residues but E74 was not assigned because no resonance was observed for this residue.

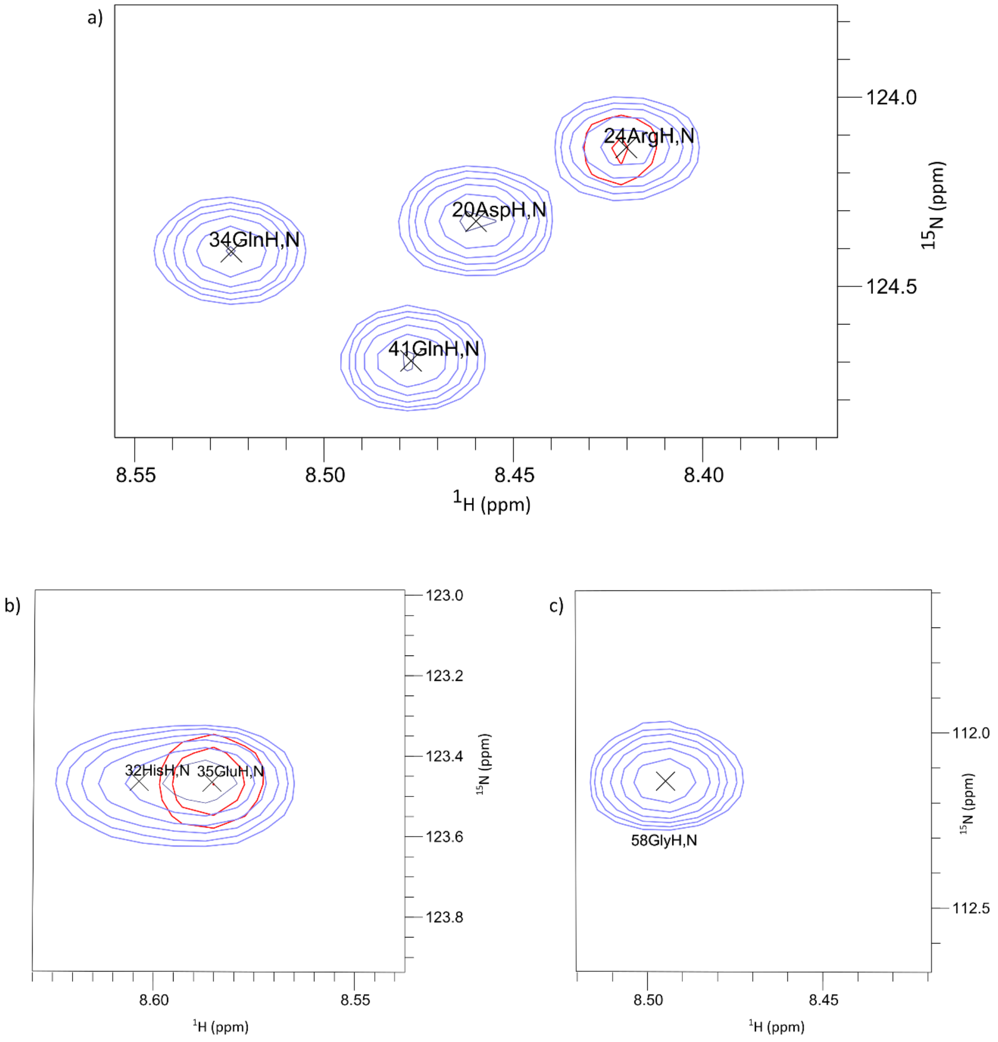


Figure 2-11: Selective isoleucine and valine unlabelling of HASPA. a-c) Overlay of the [*U*-^{15}N] HASPA ^{1}H, ^{15}N HSQC spectrum (blue) with the isoleucine and valine unlabelled 2D ^{1}H, ^{15}N HNCO (red) showing the unlabelled i+1 residues.

This metabolic precursor unlabelling approach gave vital sequence based information on 21 residues in the HASPA sequence. Identified spin systems could be used as anchor points for the remaining assignment, while these spin systems were also removed from the pool of potential linkers in other areas of the sequence. Combining this approach with simply identifying residues based on their unique Cα/ Cβ resonances (8 alanine, 4 serine, 4 threonine and 10 glycine residues) made the backbone resonance assignment of the H_N, N, Cα and Cβ nuclei in HASPA possible (Figure 2-12). There are 78 residues in the HASPA sequence that could have been assigned after excluding the three proline residues and the hexa-histidine tag. The final assignment was of 73 residues, meaning that 94% of the HASPA sequence was successfully assigned, except G1, T5, E74, H75 and A76.

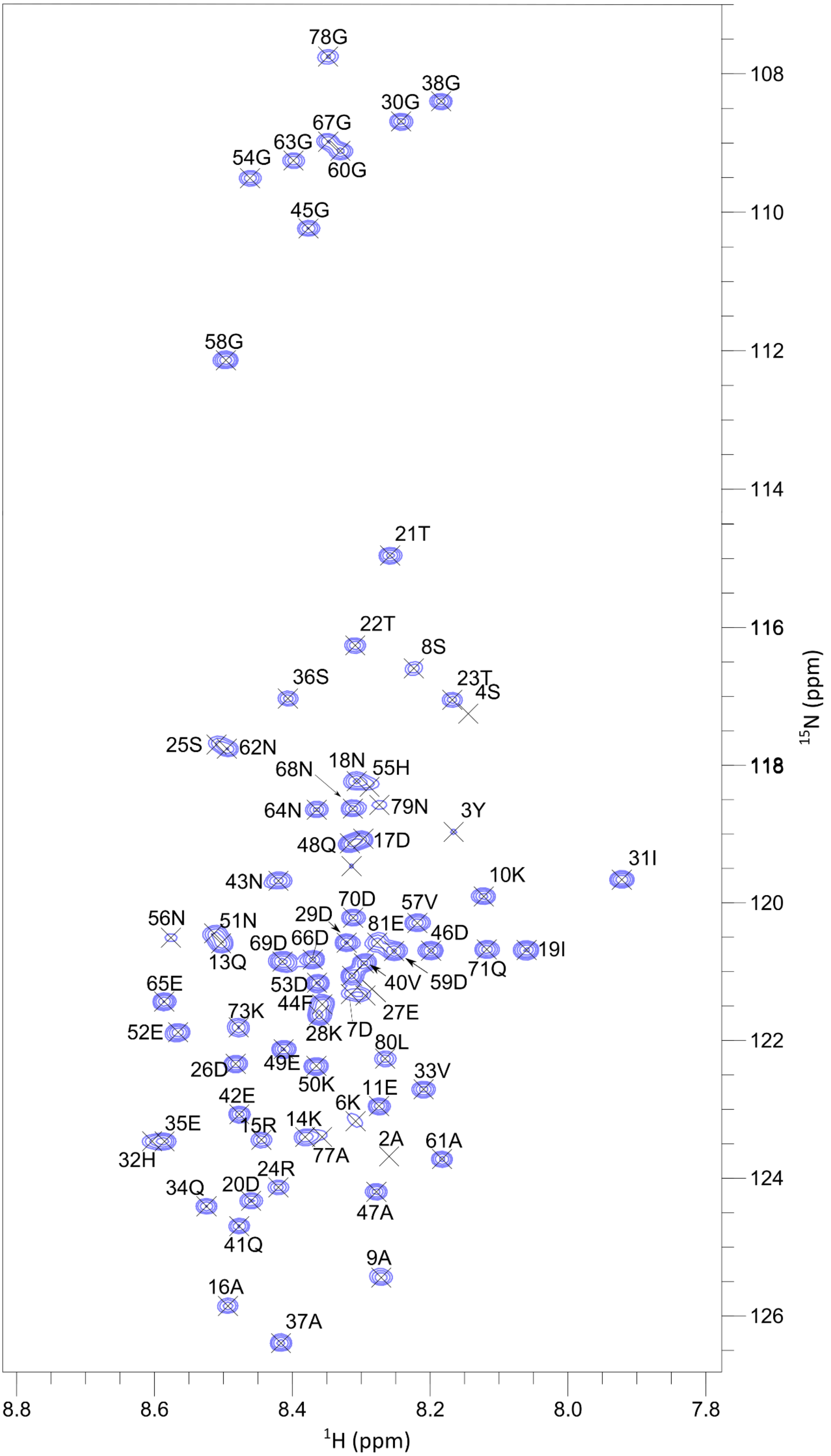


Figure 2-12: Full backbone assignment of HASPA. 73 residues of a possible 78 residues (excluding the hexa-histidine tag and proline residues) were assigned.

2.3.5. Backbone torsion angles of HASPA

Elements of transient secondary structure are a common feature of intrinsically disordered proteins. IDPs exist as a highly dynamic structural ensemble and can in some, but not in all cases, form defined folds when interacting with a binding partner. As already mentioned in Section 2.1., these transiently structured regions are termed molecular recognition features (MoRFs). ANCHOR is a web-based server that predicts binding sites in disordered regions of protein sequence that are likely to undergo a disorder to order transition upon binding (Dosztányi et al. 2009). It was used to highlight potential transiently ordered binding sites within the disordered domain of the erythrocyte binding-like protein from *Plasmodium falciparum* that were later experimentally confirmed (Blanc et al. 2014). Analysis of the HASPA sequence predicts three potential binding sites: 1) residues 1 to 22, 2) residues 25 to 63 and 4) residues 72 to 79 (Figure 2-13). Experimental verification is required to determine if these predictions have any relevance to the disordered nature of HASPA and could contribute to any plausible short-lived structured elements.

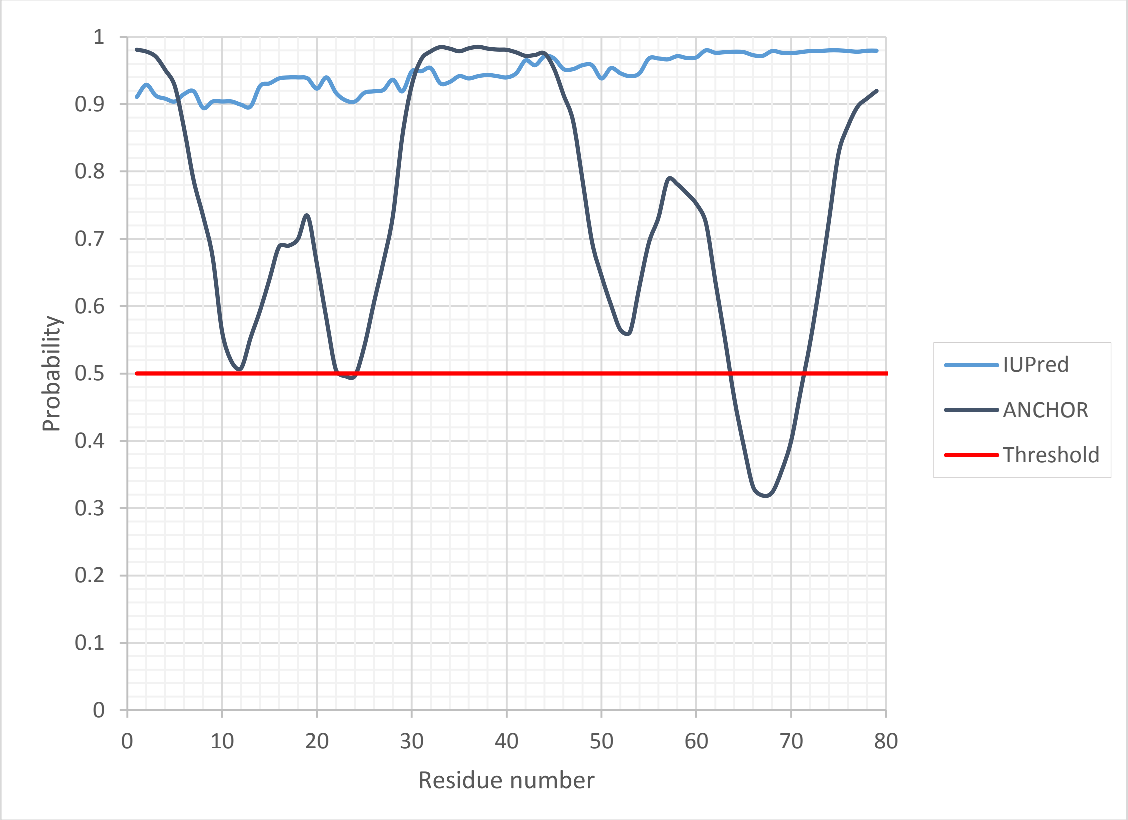
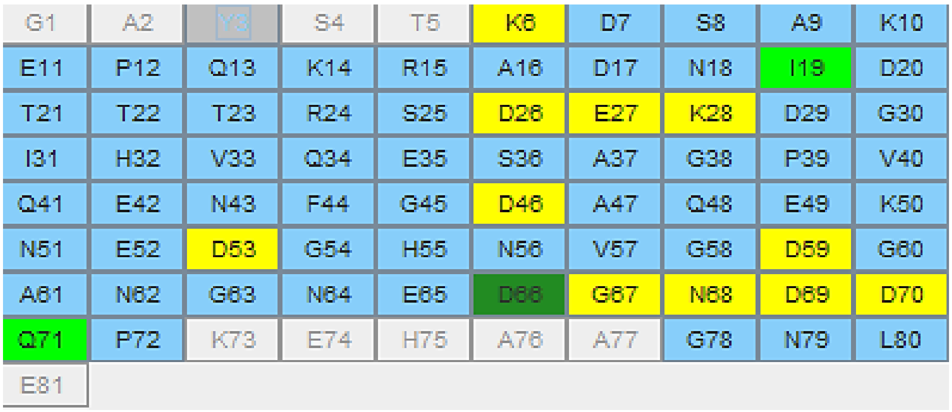


Figure 2-13: ANCHOR and IUPred output for HASPA. ANCHOR analysis of the HASPA sequence predicts three potential binding regions.

These transient ordered states can be experimentally determined. There is an established relationship between chemical shift information and local protein structure. This specifically relates to the dihedral angles of the protein backbone and their relationship to defined structural motifs. Three dihedral backbone angles restrain protein structures, namely phi (ϕ), psi (ψ) and omega (ω). Phi and psi relate to the backbone atoms C'-N-Cα-C' and N-Cα-C'-N, respectively. They share an inherent flexibility, which is dependent on the sequence context and local secondary structure. Omega describes the N-C’ bond geometry and is around 180˚ or 0° in the case of cis-proline. Thus, ϕ controls the C'-C' distance, ψ controls the N-N distance and ω controls the Cα-Cα distance. The advantage of chemical shifts as structural probes is that they allow sequence specific mapping of transient structural elements of IDPs (Kragelj et al. 2013).

TALOS-N, an artificial neural network (ANN) based hybrid system, takes NMR-derived chemical shift information (Cα, Cβ, C’, Hα, HN and N) related to the sequence of the target protein to make quantitative predictions for the protein backbone angles ϕ/ψ (Shen and Bax. 2013). It also provides a measure of the reliability of these predictions and predicts any secondary structure elements. It has been outlined in Section 2.3.3. that HASPA is predominantly intrinsically disordered but identification of any transient secondary structured elements would be interesting and perhaps inform on any unknown functionality. Therefore, the chemical shift information (Cα, Cβ, H_N and N) obtained from the backbone assignment of HASPA was analysed with TALOS-N. It showed the entire assignment to be highly dynamic with no elements of secondary structure predicted. This result suggests that the HASPs do not contain any preformed secondary structure elements that could form the basis of molecular recognition motifs seen in IDPs.



G1	A2	Y3	S4	T5	K6	D7	S8	A9	K10
E11	P12	Q13	K14	R15	A16	D17	N18	I19	D20
T21	T22	T23	R24	S25	D26	E27	K28	D29	G30
I31	H32	V33	Q34	E35	S36	A37	G38	P39	V40
Q41	E42	N43	F44	G45	D46	A47	Q48	E49	K50
N51	E52	D53	G54	H55	N56	V57	G58	D59	G60
A61	N62	G63	N64	E65	D66	G67	N68	D69	D70
Q71	P72	K73	E74	H75	A76	A77	G78	N79	L80
E81									

Figure 2-14: TALOS-N output highlighting the dynamic nature of HASPA. This sequence window displays the classification of each residue in the target protein sequence. The residues are classed as dynamic (blue), ambiguous (yellow), not classified (grey) and a good prediction (green). The majority of the HASPA sequence is considered dynamic. Note residues G1, T5, E74, H75 and A76 were not assigned.

Further analysis of the chemical shift data to define any propensity for transient structural elements within HASPA was performed with ncSPC (neighbour corrected structural propensity calculator). The ncSPC program compares chemical shift data with a chemical shift library of disordered proteins (Tamiola et al. 2010). Deviation from the random coil chemical shift library in the query can indicate the presence of structure. In addition ncSPC can differentiate between global structure and weak sequence specific transient structural elements. It is important to note that the program takes sequence into account as the reference random coil library was generated for every nucleus of each amino acid from the reference IDPs used. Moreover, chemical shifts of specific nuclei have varying reliabilities in reporting certain structured elements. There is a known relationship between ^{13}Cα chemical shifts and reliable α-helical content prediction and ^{13}Cβ chemical shifts with β-sheet content (Weinstock et al. 2008). The contributions of different chemical shifts are weighted by their sensitivity to α-helical or β-sheet structure (Tamiola and Mulder. 2012).

Analysis with ncSPC shows that HASPA has no global defined structure and is consistent with the TALOS-N output (Figure 2-15). It was of particular interest that ncSPC reports structure propensity plots for two slightly different methods, SSP (secondary structure prediction) and ncSPC. The SSP method on which ncSPC is based differs as it uses chemical shifts from known structured elements as a standard to measure structural propensity against (Marsh et al. 2006). However, the graphical output of both methods is the same, enabling direct comparison of the resultant predictions. Scores of 1 and −1 reflect fully formed α helical or β sheet structure, respectively, while a score of 0.5 suggests that 50% of the conformers in the ensemble are α helical at that position. The SSP method detects areas of enhanced β sheet propensity of around 50% in short interspersed regions of HASPA (Figure 2-15a). However, this finding should be treated with caution as its significance would need to be experimentally confirmed before drawing any firm conclusions. While the ncSPC method shows that the propensity falls within the range of the IDP reference set and no structural propensity is observed (Figure 2-15b). In the current literature ncSPC is considered a refinement of SSP because of the use of the IDP random coil reference set (Tamiola and Mulder. 2012). This improved reliability of structure propensity calculations was tested with two experimentally defined systems, namely α synuclein under native and membrane-mimicking conditions and the structural transition of the bacterial PYP (photoactive yellow protein) upon light excitation. In both cases ncSPC was able to detect conformational shifts and highlight known structural propensity. Due to this finding the ncSPC result was taken as the significant output. However, the resultant variance of these methods highlights the importance of experimentally defining any calculated structure propensity.

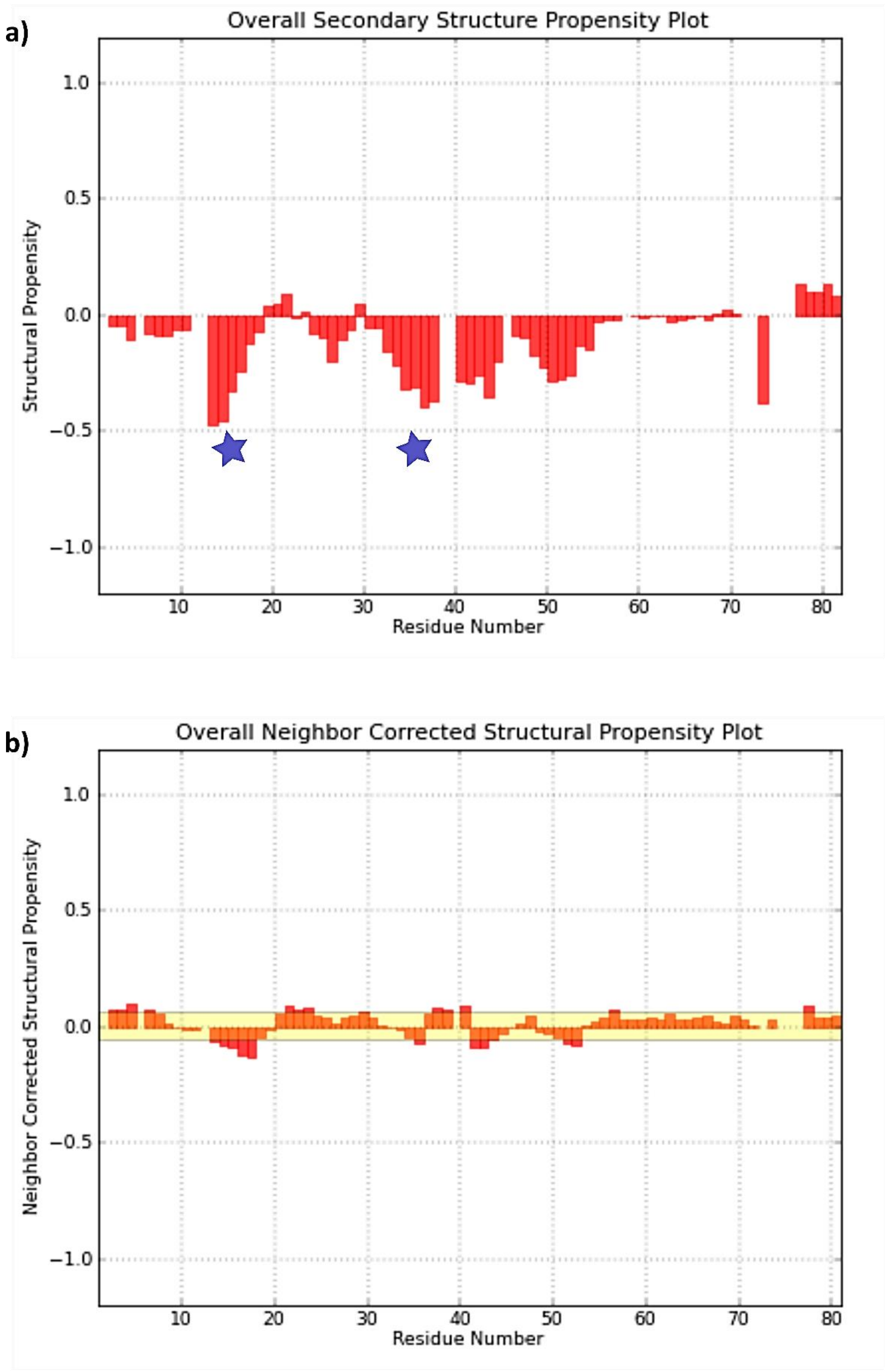


Figure 2-15: Secondary structure propensity plots of HASPA based on the SSP and ncSPC methods. A score of +1 predicts α helical content while -1 predicts β-sheet. a) SSP structural propensity prediction plot shows areas of enhanced β propensity in short interspersed regions of HASPA (denoted by the stars). b) ncSPC structural propensity prediction plot shows that the HASPA chemical shifts adhere closely to those of known IDPs (yellow band) with no elements of predicted structure observed.

2.4. Discussion

2.4.1. The HASPs are intrinsically disordered

Analysis of the HASPs, from their sequence to biophysical characterisation has shown them to be intrinsically disordered. The high occurrence of disorder promoting residues, high hydrophilicity and the low sequence complexity found in all three HASPs was the first indication of disorder. This predicted disorder was supported by CD spectroscopy of both HASPA and HASPB1. Analysis of the raw data with DichroWeb predicted that HASPA is 92% random coil. However, CD is limited when addressing IDPs as it is unable to detect secondary structure elements that are short lived, weakly populated and confined locally within the protein sequence (Rezaei-Ghaleh et al. 2012). Further examination of the dynamic nature and the inherent flexibility of the HASPs was required.

NMR spectroscopy is unique in its capacity to study the dynamic nature of intrinsically disordered proteins in solution (Dyson and Wright. 2004). Backbone assignment with 2D and 3D NMR spectra is the first step in this process and a prerequisite for further study of transient structural elements, biological partner interactions and IDP dynamics. Initial 2D ^{1}H, ^{15}N HSQC spectra of [^{15}N] labelled HASPA and HASPB2 revealed limited resonance dispersion in the proton dimension, a spectral feature indicative of highly disordered proteins. However, the resonances were well resolved suggesting that backbone assignment would be possible. HASPA was chosen as the best candidate because it is the smallest HASP and lacks the repeats seen in the HASPBs.

3D HNCACB and HN(CO)CACB spectra were collected and used to correlate the amide resonances from the 2D ^{1}H, ^{15}N HSQC spectrum to the Cα and Cβ resonances. Spectral congestion can be a severe issue when assigning an IDP because all residues are solvent exposed. This means that there is little dispersion of Cα and Cβ resonances regardless of sequence context. To deconvolute this spectral congestion a metabolic precursor unlabelling strategy (described in Section 2.1.3.) was implemented. This protocol renders specific residues 'NMR invisible' in an otherwise uniformly labelled protein sample and can also be used to gain information of residues precede unlabelled residues. Three unlabelled versions of HASPA were produced that silenced 1) lysine, 2) arginine and 3) valine and isoleucine. This generated additional sequence specific information and considerably aided the backbone assignment of HASPA. In the final assignment 94% of all possible residues were assigned. The completed assignment is the first step in any future studies of HASP dynamics and residual structure.

2.4.2. HASPA does not contain any elements of structural propensity

IDPs sample a diverse array of interconverting conformers in their ‘native' state. This conformational heterogeneity means they cannot be limited to any single averaged structure but are represented as an ensemble (Gibbs and Showalter. 2015). NMR spectroscopy is a vital tool to produce the experimentally derived parameters required for IDP ensemble generation. Information obtained from these ensembles can be used to understand how IDPs carry out their functions, through analysis and comparison of their structural dynamics alone or in complex with biological binding partners. It is an established concept in the IDP field that some proteins undergo disorder to order transitions on biological partner binding, facilitated in most cases by preformed elements of local transient secondary structure that lower the entropic cost of more global folding. Solution state NMR spectroscopy can be used to investigate these local transient states and relate them back to biological function or aggregation behaviour.

Elements of short-lived secondary structure can be predicted and confirmed with experimentally defined parameters. In particular, ^{13}C chemical shifts are highly sensitive to backbone conformation and can inform on short-lived structured states in IDPs, hence their use to determine IDP structural propensity. Chemical shifts obtained from the backbone assignment of a protein can inform on the dihedral (phi and psi) angles. Certain dihedral angles are found in to specific secondary structure elements and can define the random coil state. This makes IDP backbone chemical shifts powerful descriptors of local transient deviations from the random coil state. These deviations can be connected to functional or dysfunctional protein states, involved in adaptive molecular recognition or protein aggregation. The advantage of chemical shifts as structural probes is that they are rooted in the protein sequence and therefore allow transient structural elements to be mapped onto IDPs (Kragelj et al. 2013).

It was shown that conformational ensembles of IDPs derived from chemical shifts alone are consistent with ensembles derived from other NMR experimental parameters such as residual dipolar couplings (Jensen et al. 2010). Therefore, the lack of structural propensity seen in HASPA from chemical shift analysis with TALOS-N and ncSPC can be considered a significant finding. This would suggest that HASPA contains no preformed MoRFs in its ‘native’ state indicating that it does not function as an interaction hub. However, the relationship between intrinsic conformational propensity and the structure adopted by the protein in its bound form remains poorly understood. Notably, the concept that preformed structure is required for continued coupled folding and biological partner binding is not always applicable. Disruption of residual helical structure in PUMA, an IDP of the BCL-2 apoptotic family of proteins, did not inhibit binding to MCL-1 (Rogers et al. 2014). It was previously established that in the unbound form, PUMA has helical propensity of 20% and when bound forms a single continuous α-helix in a surface groove on MCL-1. The helical content of the unbound form of PUMA was reduced by site directed mutagenesis to the helix breaker proline and confirmed by CD. Interestingly, this reduced preformed helical content did not affect MCL-1 binding affinity. The conclusion of this study is that an IDP does not necessarily need to display residual structure resembling the bound form for partner binding to occur. Potentially, the established MoRF view detracts from the functional and structural plasticity that IDPs can adopt. Studying one transient structured state and interaction in isolation does not represent IDPs as multi-layered interaction hubs. Could this be the case for the HASPs?

The lack of structural propensity observed in HASPA suggests entropic chain classification. This grouping of IDPs remain disordered to elicit their functions, extending the structure function paradigm to include functional disorder (van der Lee et al. 2014). Elastin, the main component of elastic fibres and an integral component of the extracellular matrix, is an entropic spring. It is vital for tissue elasticity, providing the distinctive repetitive stretch and elastic recoil properties of vertebrate tissues. Like the HASPs, it is an extracellular intrinsically disordered protein but unlike most IDPs contains a significant number of hydrophobic residues. The high proline and glycine content (42%) of elastin is thought to be the main contributor to disorder (Miao et al. 2003). The specific residue prevalence limits the formation of extended secondary structure, preventing collapse of the hydrophobic sequence into a compact globular structure (Rauscher et al. 2006). This allows the monomer to remain disordered, providing the functionally required conformational flexibility through transiently populated local structural elements. Therefore, elastin can be thought of as a functionally relevant heterogeneous disordered ensemble, where conformational disorder is crucial for driving extensibility and elastic recoil (Muiznieks et al. 2010).

The examples considered throughout this chapter have highlighted the diverse functions that IDPs form over folded proteins. Although the biological function of the HASPs is unknown, intrinsic disorder classification is informative when you consider the characteristic traits of IDPs. Conformational freedom, particularly observed in entropic chains, is highly advantageous and facilitates an array of functions. This can include long range entropic exclusion of other proteins, spacer functions (MAP2) and dynamic springs (elastin and titin). The extended nature of IDPs means they are able to elicit repulsive forces or make longer ranging contacts than globular proteins of the same molecular weight. Reconsidering the functional state of the HASPs as disordered should inform any subsequent biological investigation and help to reframe their potential role. The lack of defined binding partners, other than NMT, could suggest a repulsive type of functionality for the HASPs.

2.5. Summary and future work

The predicted and experimental work presented in this chapter confirm that the HASPs are intrinsically disordered. NMR spectroscopy is a comprehensive source of experimentally derived information on IDPs and was applied to this study. A backbone assignment of HASPA was performed as a starting point for analysis of secondary structure propensity and future study of protein dynamics. Due to the low sequence complexity and disorder of HASPA the assignment was hampered by spectral congestion. Residue-specific metabolic unlabelling was implemented to obtain spin system information that could be directly related to the HASPA sequence. This protocol aided the final 94% backbone resonance assignment of H_N, N, Cα and Cβ in HASPA. NMR-derived chemical shift information from the backbone assignment was used to investigate HASPA structural propensity. There were no regions of deviation from random coil Cα and Cβ chemical shifts, which is consistent with a lack of transient secondary structure. This means that the HASPs do not possess a feature that is common in IDPS that function as interaction hubs, namely MoRFs. Therefore, the HASPs are more similar to IDPs that function as entropic chain, where their lack of order is what makes them functional, such as titin and elastin.

HASPA was studied, therefore, here no inference can be drawn about the presence of MoRFs in the repeats of the HASPBs. However, the sequence similarity in the N- and C-terminal domains of all three HASPs means that findings presented here are transferable to the HASPBs. It would be interesting, although very challenging, to assign HASPB2 to determine if the repeats contained any structural propensity. Moreover, preformed structural elements are not always required for biological partner interaction. However, the current lack of identified biologically relevant HASP binding partners does make the entropic chain classification an interesting concept.

Further experimental validation would be useful to confirm the chemical shift determined lack of transient structure within HASPA. Long range and local structural features of IDPs in solution can be determined with paramagnetic relaxation enhancements, residual dipolar couplings, relaxation studies and nuclear Overhauser enhancement (Salmon et al. 2010). A combination of the aforementioned methods would define and confirm any transient structured states of HASPA. Any information obtained could address any functionally relevant structural organisation.

Chapter 3 - Investigation of the HASP: NMT interaction

3.1. Introduction

The N-terminal 18 residues of *L. major* HASPB were shown to be essential for membrane targeting (Denny et al. 2000). This region comprises a dual acylation motif, where the protein is myristoylated and then sequentially palmitoylated for correct membrane association. The first lipidation step, myristate addition, is catalysed by N-myristoyltransferase (NMT) and is a co-translational event *in vivo*. NMT can myristoylate peptides *in vitro*, requiring an N-terminal glycine residue for modification (Goncalves et al. 2012). However, the catalytic activity of NMT on full length protein substrates has not been established *in vitro*. In order to address this, NMT catalysed myristoylation of the HASPs was investigated *in vitro*. The work presented in this chapter aims to establish if NMT could myristoylated full length protein substrates *in vitro* and define the kinetic parameters of this activity. Moreover, having defined the HASPs as disordered, investigation of the effect and potential structural impact of myristoylation could be functionally relevant.

3.2. Experimental

3.2.1. *Leishmania major* NMT expression and purification

Briefly, aliquots of LB media (5 ml) containing ampicillin and chloramphenicol, at 100 and 30 µg/ml, respectively, were inoculated with a glycerol stock of *E. coli* Rosetta2/plysS with the pET15-MHL plasmid, which was obtained from Dr James Brannigan, and grown overnight at 37°C with shaking at 180 rpm (see Appendix A.1.). It is important to note that pET15-MHL vector is isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible, ampicillin resistant and contains a TEV cleavable N-terminal hexa-histidine tag. LmNMT was expressed with the typical autoinduction cell growth protocol (Studier. 2005).

Each 5 ml aliquot was used to inoculate 20 ml LB media supplemented with the same antibiotics and incubated at 37˚C for an hour with shaking at 180 rpm. The base of the autoinduction protocol is ZY media, which was prepared by adding 10 g tryptone, 5 g yeast extract and 937 ml H_2O per litre required. A 50x stock of 5052 was prepared in a final volume of 100 ml by adding 25 g glycerol, 73 ml dH_2O, 2.5 g glucose and 10 g α-lactose. A 25x M stock was prepared in a final volume of 200 ml by sequentially dissolving 3.6 g Na_2SO_4, 13.4 g NH_4Cl, 17 g KH_2PO_4 and 17.7 g Na_2HPO_4 (all anhydrous) into dH_2O. A 1 M stock of $MgSO_4$ was also prepared and all these components were autoclaved before use. The autoinduction media was assembled according to the ratios outlined in Table 3-1 and also incubated at 37˚C to pre-warm it. After this incubation each 25 ml aliquot was added to 500 ml autoinduction media and returned to 37˚C for 6 hours with shaking at 180 rpm. The cells were then transferred to 18˚C with shaking at 200 rpm overnight (typically between 16 to 18 hours). The cells were harvested by centrifugation for 15 minutes at 5000 xg and the resulting pellets stored at -20˚C.

Component	Volume in 10 ml	Volume in 1 l
1 M $MgSO_4$	10 µl	1 ml
50x 5052	200 µl	20 ml
25x M	400 µl	40 ml
ZY media	9.37 ml	937 ml

Table 3-1: Components required for autoinduction media.

Once thawed, the pellets were resuspended in 30 ml nickel column buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 5 % glycerol) per 500 ml pellet supplemented with 0.1 % Triton X-100, 10 mM $MgCl_2$, DNase and a protease inhibitor tablet (cOmplete, EDTA-free from Roche). The soluble cell suspension was sonicated 6 times on ice for 30 seconds with 1 minute intervals. There was no lysate clarification step as the lysate was loaded directly onto a 1 ml HisTrap FF crude column (GE Healthcare) pre-equilibrated with nickel column buffer A. Ten column volumes of nickel column buffer A washed the column before 100% nickel column buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 250 mM imidazole, 5 % glycerol) was implemented for fifteen column volumes to elute the target protein. SDS-PAGE was used to identify target protein containing fractions that were pooled and diluted at least 20-fold with 20 mM HEPES pH 7.5 containing 1 mM DTT to give a final NaCl concentration of 20 mM. A 1 ml HiTrap Q HP column equilibrated with ion exchange buffer A (20 mM HEPES pH 7.5, 20 mM NaCl, 1 mM DTT) was loaded with the pooled sample and washed for 10 column volumes. Elution was achieved with a gradient of 0 to 100% ion exchange buffer B (20 mM HEPES pH 7.5, 500 mM NaCl, 1 mM DTT) for 20 column volumes. The NMT containing fractions were then concentrated (Vivaspin Sample Concentrator range, 10000 MWCO) to 0.5 ml for size exclusion purification. A Superdex 75 10/300 column (GE Healthcare) equilibrated with 10 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM DTT and used to further purify NMT. Fractions were then analysed with SDS-PAGE to assess final sample purity. All purification columns were run with the AKTA FPLC system.

Subsequently, the hexa-histidine tag was removed by TEV protease (Invitrogen) digestion (1:50 ratio of protease to protein) overnight at 4˚C while the sample was dialysed into 50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM imidazole, 5 % glycerol, 1 mM DTT. The nickel purification step was repeated but the cleaved samples eluted in the nickel column buffer A wash step. No further purification was required but the size exclusion step was run again to ensure the protein was in the correct final buffer.

Myristoyl CoA or non-hydrolysable myristoyl CoA analogue were added directly to NMT after size exclusion purification. The co-factor was chosen based on the downstream application of the protein and was added in 1.5 molar excess of the protein concentration. Myristoyl CoA (CAS Number 3130-72-1) was obtained from Sigma Aldrich and prepared as a 10 mM stock in 50% (v/v) DMSO. The non-hydrolysable myristoyl CoA analogue (PDB code: NHM), 2-oxopentadecyl-CoA, was prepared as a 10 mM stock in 50% DMSO. This compound was obtained from Dr. James Brannigan as it is not commercially available. The synthesis is detailed in the supplementary information of the following paper (Brannigan et al. 2010).

3.2.2. *In vitro* myristoylation of HASPA by NMT

In order to determine if the HASPs could be myristoylated *in vitro* NMT was incubated with a positive control peptide based on the N-terminus of *L. donovani* HASPB2 (GSSSTKD), which is known to be myristoylated *in vitro* or HASPA, as prepared according to the schedule in Section 4.2.1. The reactions were set up in a final volume of 100 µl with NMT size exclusion buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM DTT) as the diluent. A 10 mM stock of the peptide was prepared in the same buffer. For the reaction, 5 µM NMT (complexed with myristoyl CoA) was added to 50 µM HASPA or peptide with 100 µM myristoyl CoA and incubated at 37˚C for 30 minutes. Electrospray ionisation mass spectrometry (ESI-MS) was used to determine the mass of the major components of each sample. The data were collected with an ABI Qstar tandem mass spectrometer. 10 µl of each reaction mix was supplied for analysis, while the peptide was supplied at 1 mM and HASPA at 2 mg/ml in 20 mM sodium cacodylate pH 6.5, 120 mM NaCl for the negative controls. Sample desalting was achieved by ZipTip (Merck Millipore) purification before data were collected.

3.2.3. Kinetic analysis of HASP myristoylation

Kinetics parameters of HASP myristoylation by NMT were obtained using a continuously observed coupled fluorometric assay, termed here the CPM assay, discussed further in Section 3.3.3. 7-diethylamino-3-(4-maleimidophenyl)-4-methylcoumarin (CPM) was obtained from Sigma Aldrich (CAS number: 76877-33-3) and prepared as a 10 mM stock in 100% DMSO. A 4x stock of CPM assay buffer was prepared with 2.5 g KH_2PO_4 and 19.79 g Na_2HPO_4, 8 ml 0.5 M EDTA and 800 µl Triton X-100 in a final volume of 2 litres and adjusted to pH 7.9. All stocks were made in 1x CPM assay buffer (20 mM Na/ K phosphate pH 7.90, 0.5 mM EDTA, 0.1% Triton X-100) supplemented with 2% DMSO, except the CPM stock which had a final concentration of 5% DMSO (Table 3-2). Note, the reagents solutions were prepared directly before the assay was performed.

The assay was carried out in 96-well black polypropylene microplates (Thermo Scientific). Fluorescence readings were performed on a BMG Labtech POLARstar OPTIMA plate reader set at 25˚C. Excitation was set at 405 nm and emission at 460 nm. Each condition was recorded in triplicate with three wells setup per condition. The reagents were prepared and added to the plate as detailed in Table 3-2, except for the NMT stock. This was added to the plate with the syringe pump in the plate reader to start the reaction. The resultant fluorescence intensity was monitored for 2 hours, with a reading every 5 min. All three HASPs were tested between 100 µM to 26 nM. NMT was prepared with histidine tag was removed, as this improved background fluorescence.

Component	Stock concentration	Final concentration	Volume added (µl)
NMT	200 nM	20 nM	10
CPM	32 µM	8 µM	25
Peptide/ HASP	400 µM to 104 nM	100 µM to 26 nM	25
Myristoyl CoA	16 µM	4 µM	25
Buffer	-	-	10

Table 3-2: Scheme for the preparation of the CPM assay.

The V_{max} and K_m were determined using GraphPad Prism version 6.05 for Windows, GraphPad Software, La Jolla California USA. The negative control excluded NMT and the positive control was a saturating concentration of HASP or peptide. CoA (CAS number: 55672-92-9) was titrated, between 5 μM to 5 nM, with CPM to produce a plot that related product formation in μM to fluorescence. The fluorescence values obtained for the HASPs and peptide were was divided by slope from the CoA titration. The normalised values were re-plotted with the rate now expressed in μM s^{-1}.

3.2.4. Real-time observation of HASPA myristoylation

Uniformly ^{15}N labelled HASPA was prepared as described in Section 2.2.2. and was incubated with unlabelled NMT at a ratio of 100:1. In the final sample, HASPA was at 200 μM and NMT at 2 μM in NMT size exclusion buffer (10 mM HEPES pH 7.5, 500 mM NaCl, 0.5 mM DTT). The real-time NMR experiment was run overnight at 298K with 2 fold molar excess of myristoyl CoA (400 μM). A 2D (H,N) HSQC spectrum was recorded with 2 scans/increment and a matrix size of 2048 * 64 points (real and imaginary). Spectral widths of 16 and 32 ppm in 1H and 15N dimensions were used. 71 complete 2D experiments were recorded in total. Directly before data collection D_2O was added to all the samples to give a final concentration of 10% (v/v).

The spectra were collected using the 700 MHz Bruker AV spectrometer at 298K. Raw NMR data were processed with NMRpipe and the processed 2D spectra were analysed with CCPN Analysis (Vranken et al. 2005). Resonances assignments were transferred from the backbone resonance assignment of HASPA presented in Chapter 2. It should be noted that the different buffer conditions meant only a portion of resonance assignments could be transferred with confidence. The subsequent analysis only concerns those peaks that could be confidently assigned from the backbone assignment. Deviations in peak intensity between t=0 and t=175 minutes were calculated with the following equation, $\Delta I = (I_0 - I_{Myr}) / I_0$ (Mulder et al. 1999). I_0 represents the initial peak intensity at t=0 and I_{Myr} denotes the peak intensity at t=175. Chemical shift changes (Δδ), expressed in ppm, of the individual amide pairs, ^1H - ^{15}N, between t=0 and t=175 were calculated with $\Delta\delta = [\Delta\delta^2_H + (\Delta\delta_N/R_{scale})^2]^{½}$. A chemical shift scaling factor, R_{scale}, of 6.5

was used. This R_{scale} value was determined from the ratio of the average variances of backbone nitrogen and proton chemical shifts from residues in proteins deposited with the BioMagResBank (Mulder et al. 1999). $\Delta\delta_H$ and $\Delta\delta_N$ represent the changes in chemical shift for ^{1}H and ^{15}N, respectively.

3.2.5. NMT: HASPA co-crystallisation

HASPA and NMT were dialysed into 20 mM Hepes pH 7.8, 50 mM NaCl before being combined at either 2:1 (475.2 µM HASPA: 237.6 µM NMT) or 5:1 (1.2 mM HASPA: 237.6 µM NMT). NMT was kept constant at 237.6 µM and the hexa-histidine tag was removed before combination. It is also important to note that NMT was complexed with non-hydrolysable myristoyl CoA analogue. The complexes were screened in a 96 well sitting drop vapour diffusion format with drops of 150 nl complex plus 150 nl well solution pipetted by a Mosquito robot (TTP Labtech Ltd). One screen, CSS I+II with 0.1 M Tris pH 7.8, was setup with both complexes (Brzozowski and Walton. 2001). Initial optimisation of promising conditions from the CSS screen focused on testing varied buffer conditions. The crystal shower formed in 0.1 M sodium acetate pH 4.5, 25% polyethylene glycol monomethyl ether 2,000 (PEG 2000 mme) and 200 mM calcium acetate with the 5:1 complex. Refinement of this condition produced single crystals in 0.1 M sodium acetate pH 4.5, 20% PEG 2000 mme and 50 mM calcium acetate. Crystals were retrieved from the drop, vitrified directly in liquid nitrogen and sent for data collection at the Diamond Light Source (DLS).

The CCP4 software suite was used to complete macromolecular structure determination from the NMT X -ray crystallographic data (Winn et al. 2011). The data were processed and scaled by Xia (XDS) and Aimless (Kabsch. 2010; Evans and Murshudov. 2013). The structure was solved by molecular replacement using Molrep (Vagin and Teplyakov. 1997) with 4CGP, *Leishmania major* NMT with myristoyl CoA bound, as the model. Model building, ligand replacement and manual fitting of the electron density map were performed in COOT (Emsley et al. 2010). REFMAC5 was used to carry out refinement of the structural data (Murshudov et al. 2011). All representations of the structure were generated with CCP4mg (McNicholas et al. 2011). It is important to note that at the date of submission this work is still in progress. However, as only minor refinement procedures remains (e.g. solvent modelling, occupancies of some disordered side-chains) the overall quality of the protein model will not change significantly.

3.3. Results

3.3.1. *Leishmania major* NMT expression and purification

It should be noted that although this thesis has examined the HASPs of *L. donovani*, the *L. major* NMT (LmNMT) was used for the myristoylation work presented in this Chapter. The *L. major* protein was preferred because previous experience within YSBL has shown it is readily crystallisable in comparison to *L. donovani* NMT (LdNMT), despite both structures being deposited in the PDB. In addition the final yield from purification is significantly improved over *L. donovani* NMT, an important consideration when carrying out biochemical studies. Clustal Omega sequence alignment of these proteins showed them to be 97.8% identical (Sievers et al. 2011). Lastly, the overall fold and secondary structures of these proteins are highly similar, when superimposed by secondary structure matching they have a Cα RMSD of 0.81 Å, highlighting the lack of any distinct structural features between these two proteins (see Section 3.3.5.).

It is the high sequence similarity between NMTs of different *Leishmania* species that led to the prediction that an efficacious inhibitor against one *Leishmania* NMT could be effective against all (Price et al. 2003). This is pertinent when you consider that analysis and comparison of human and fungal NMTs showed the myristoyl CoA binding site to be highly conserved but that the peptide binding sites had divergent species-specific specificities (Johnson et al. 1994). Moreover, recent analysis of the diverse modes of binding of inhibitors against *L. donovani* NMT were impeded when crystals could not be produced. Notably, the IC_{50} values obtained for LmNMT with the LdNMT inhibitors were comparable. Therefore, the study was continued with *L. major* NMT as a structural model, enabling the desired binding mode assessment of the various inhibitors (Brannigan et al. 2014). This study demonstrates the viability of LmNMT as a structural surrogate for the present work.

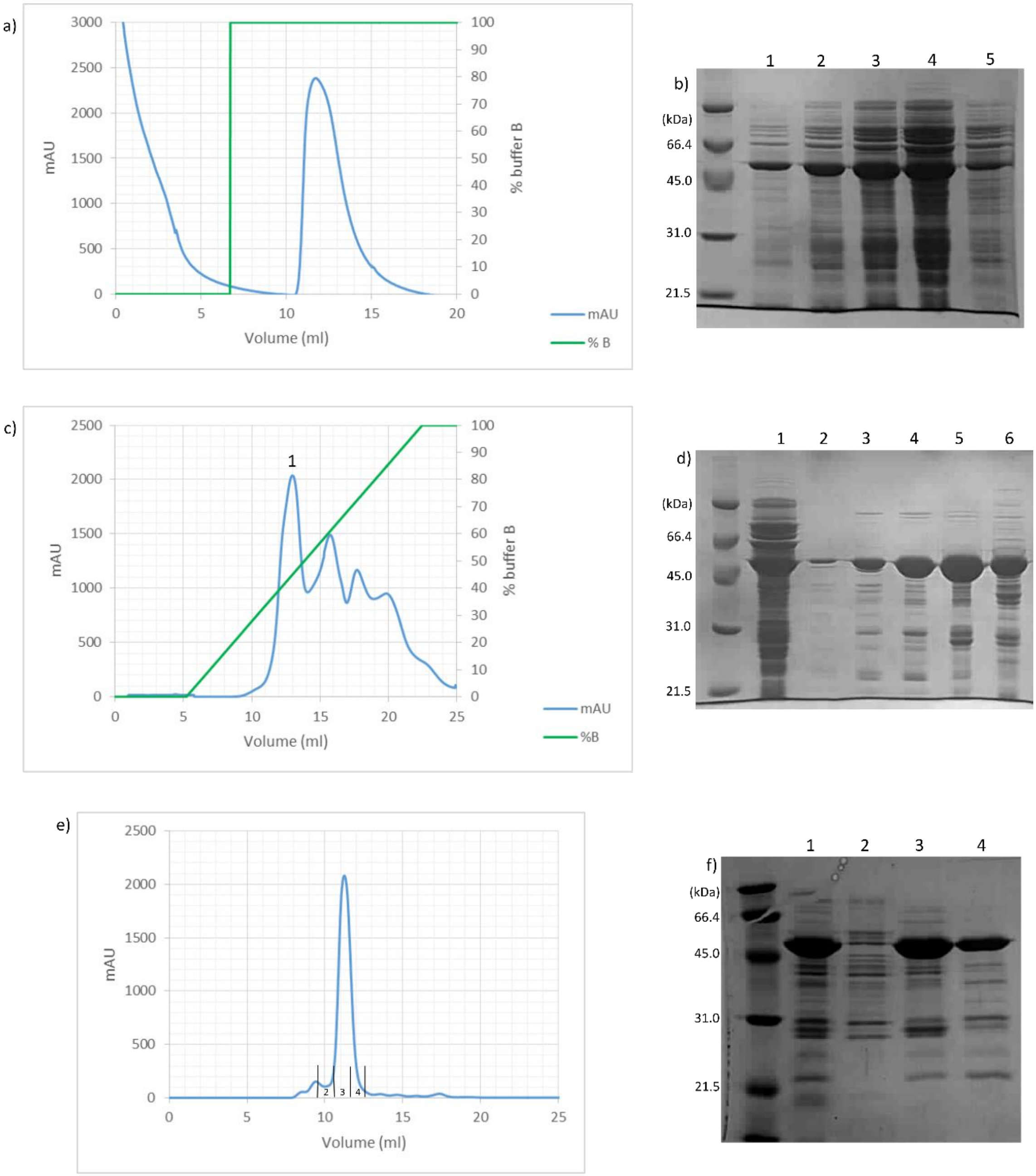


Figure 3-1: Chromatography traces and SDS PAGE analysis of the three step LmNMT purification. a) 1ml HisTrap FF crude column chromatography trace. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 250 mM imidazole) is in green. b) 12% SDS PAGE analysis of resultant fractions from (a). The gel shows resultant fractions (1 to 5). c) 1 ml HiTrapQ chromatography trace. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 500 mM NaCl) is in green. d) 12% SDS PAGE analysis of resultant fractions from (c). The gel shows load (1) and resultant fractions from peak 1 (2 to 6). e) Chromatography trace of Superdex 75 10/300 purification of NMT with fractions from (f) labelled. f) 12% SDS PAGE analysis of (e). The gel shows load (1) and resultant fractions (2 to 4). The molar mass of the NMT construct is 50.5 kDa.

Leishmania major NMT was overexpressed in *E. coli* using the autoinduction method (Studier. 2005). The protein was purified in a three step procedure, encompassing nickel affinity, ion exchange and size exclusion purification (Figure 3-1). The nickel affinity purification step produced NMT in a highly heterogeneous mixture (Figure 3-1a and b). These fractions were pooled and ion exchange purified to enhance contaminant separation (Figure 3-1c and d). NMT containing fractions, identified by SDS-PAGE, were concentrated for size exclusion purification. Adequate separation from remaining contaminants was achieved and confirmed by SDS-PAGE (Figure 3-1e and f). At this stage myristoyl CoA or non-hydrolysable myristoyl CoA analogue was added in a 1.5 molar excess of the protein concentration, depending on the downstream usage of NMT. It should be noted that NMT requires co-factor stabilisation before being concentrated for crystallisation (personal communication Dr James Brannigan). To enhance the final sample purity the N-terminal hexa-histidine tag was removed by TEV digestion. Although this step was optimised, it was only 50% efficient when considering tag removal but did significantly improve sample purity for crystallisation (Figure 3-2). The typical yield was 1 mg of protein per litre of culture after TEV digestion.

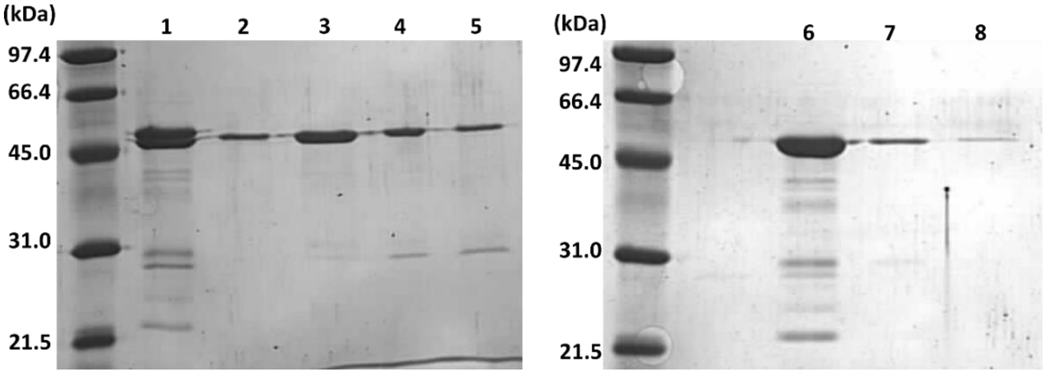


Figure 3-2: SDS-PAGE assessment of TEV digestion of LmNMT. The gels show load (1), hexa-histidine cleaved samples (2 to 5) and hexa-histidine tagged (6 to 8). The gels highlight how TEV digestion enhanced the purity of LmNMT.

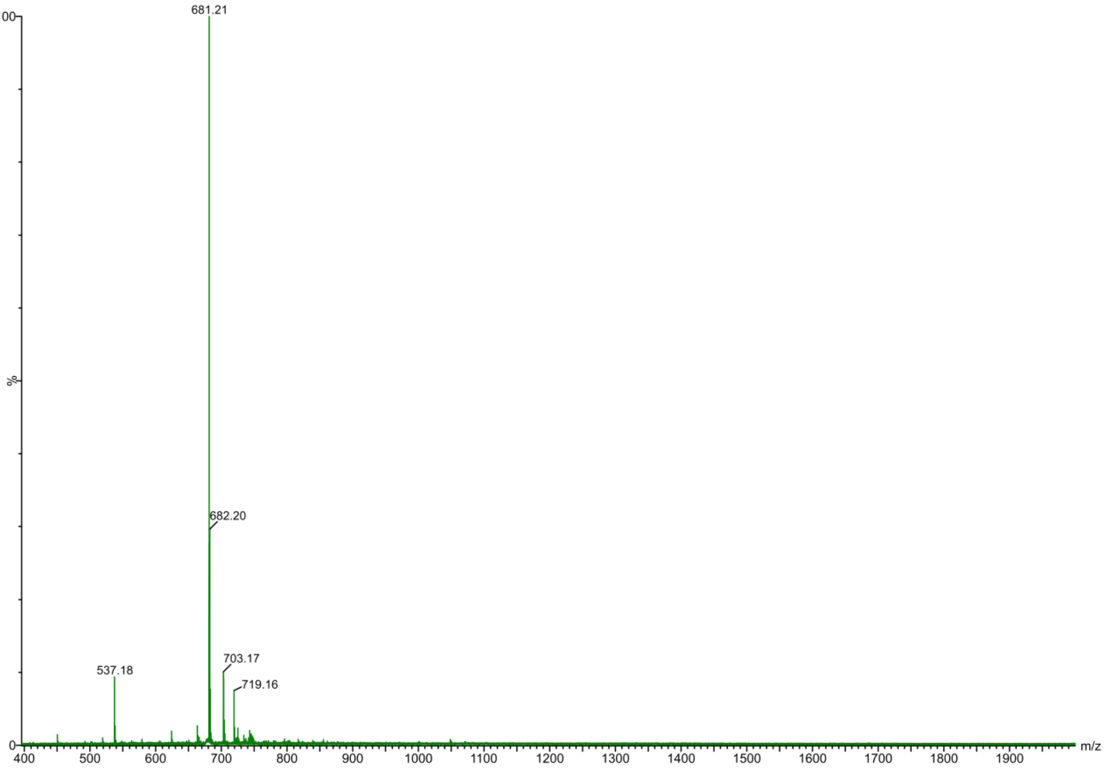
3.3.2. *In vitro* myristoylation of HASPA by NMT

N-myristoylation is a co-translational event that occurs as the nascent polypeptide chain emerges from the ribosome, following removal of the N-terminal methionine by a methionine aminopeptidase (Wilcox et al. 1987). N-myristoyltransferase catalyses this covalent attachment of myristate to the now N-terminal glycine via an amide bond. More recently, it has been shown to occur post-translationally in apoptotic cells when caspase mediated proteolytic cleavage exposes an N-terminal glycine (Zha et al. 2000).

In both co- and post-translation myristoylation events NMT accepts a short peptide as the myristoylation substrate. Therefore, it was unclear if full length proteins could be effectively myristoylated by NMT *in vitro* because of its substrate specificity *in vivo*. The HASPs are a known substrate of NMT *in vivo* but it remained to be seen if they could be myristoylated *in vitro* (Denny et al. 2000).

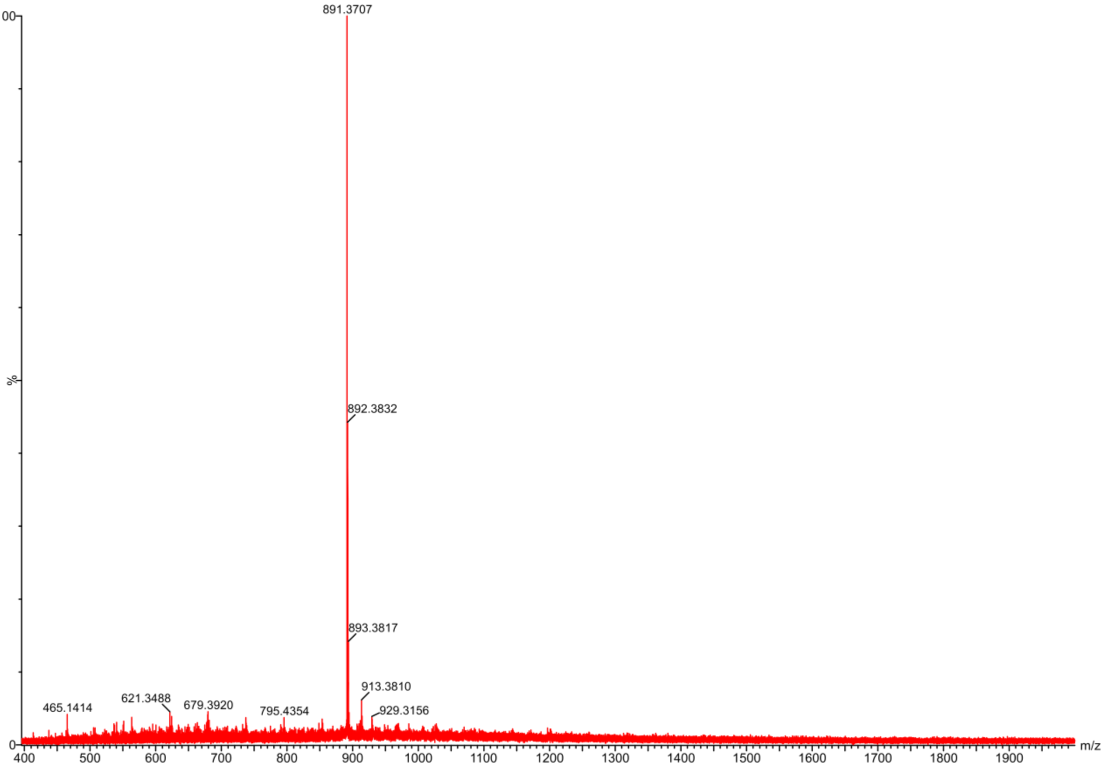
NMT was incubated with a known peptide substrate based on the N-terminus of *L. donovani* HASPB2 (GSSSTKD) with an excess of myristoyl CoA, as a positive control. Electrospray ionisation mass spectrometry (ESI-MS) of the untreated peptide confirmed the expected mass was 681.3 Da (Figure 3-3). Incubation with NMT produced a 210 Da mass increase, consistent with addition of the myristate group (Table 3-3). The molecular weight of HASPA was recorded to confirm N-terminal methionine removal. The expected N-terminal methionine cleaved molecular weight of 9445.6 Da obtained from ProtParam was in line with the ESI-MS reported value of 9445.1 Da (Gasteiger et al. 2005). It should be noted that the data for unmodified HASPA appears quite heterogeneous because it was not desalted before the data were recorded due to time constraints. Incubation of HASPA with NMT with an excess of myristoyl CoA also produced the expected 210 Da increase indicative of myristate addition (Table 3-3). The result demonstrated for the first time that NMT can myristoylate full length HASPA *in vitro* and meant further analysis of this modification and the NMT-HASPA interaction could be undertaken.

a)



681.21
682.20
537.18
703.17
719.16
100
%
0
m/z
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

b)



891.3707
892.3832
893.3817
913.3810
929.3156
465.1414
621.3488
679.3920
795.4354
100
%
0
m/z
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

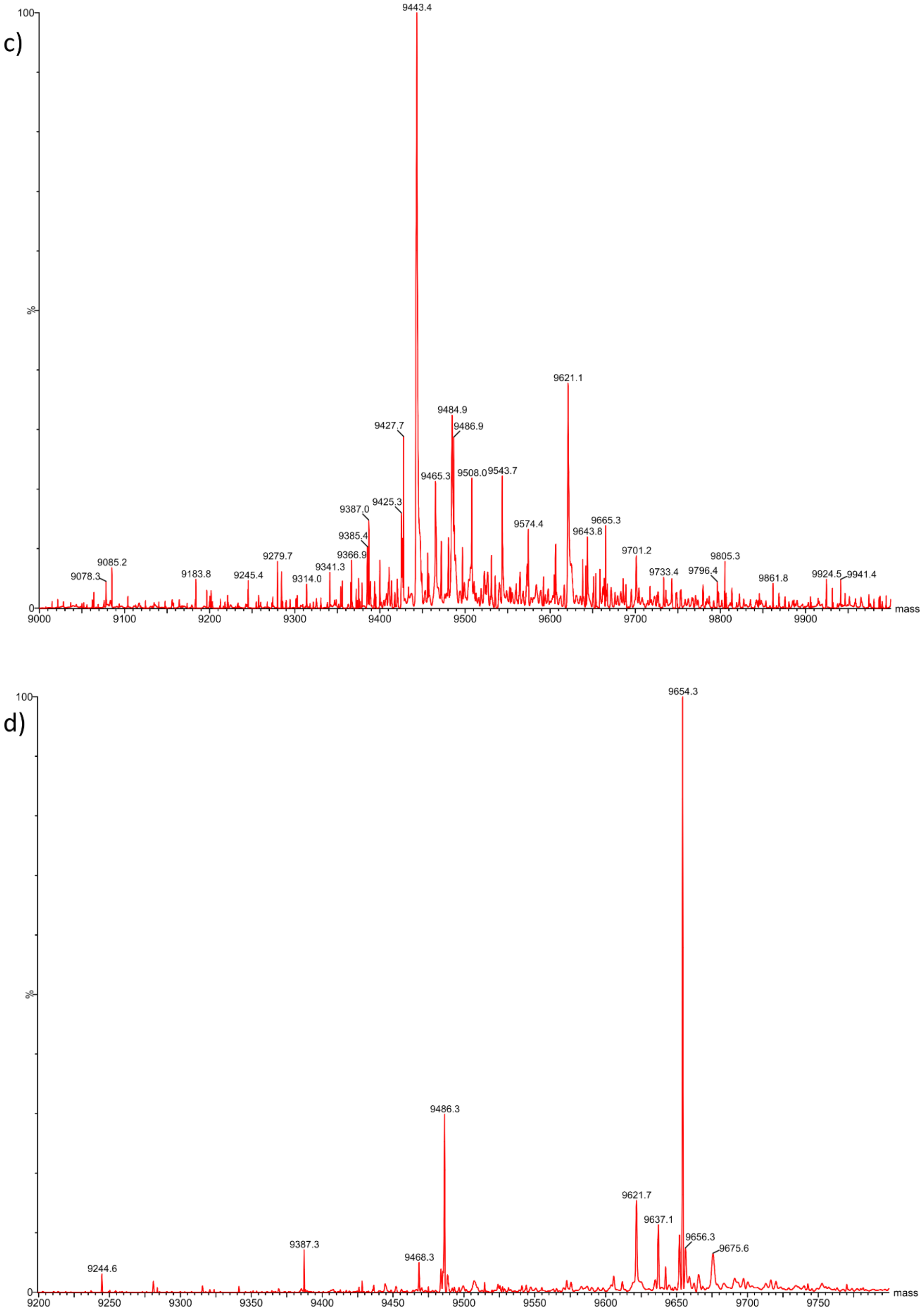


Figure 3-3: Electrospray ionisation mass spectrometry (ESI-MS) traces of untreated and NMT treated peptide and HASPA. a) ESI-MS trace of GSSSTKD peptide. b) ESI-MS trace of NMT treated GSSSTKD peptide. c) ESI-MS trace of N-terminal methionine cleaved HASPA. d) ESI-MS trace of NMT treated HASPA. Note masses do not directly match those reported as this data was scaled after collection giving the values in Table 3-3.

ESI-MS sample	Reported MW (Da)	Theoretical MW (Da)
GSSSTKD HASPB2 based peptide	681.3	680.6
Peptide incubated with NMT	891.5	690.6
HASPA	9445.1	9445.6
HASPA incubated with NMT	9656.1	9645.6

Table 3-3: ESI-MS reported molecular weights of untreated and NMT treated peptide and HASPA with theoretical molecular weights of each species.

3.3.3. Establishing the kinetic parameters of NMT catalysed HASP myristoylation

NMT catalyses the transfer of a myristate from myristoyl CoA to the N-terminal glycine of a substrate peptide. A myristoylated peptide and co-enzyme A are produced from this reaction. Co-enzyme A production can be monitored in real time with the thiol reactive probe, 7-diethylamino-3-(4-maleimido-phenyl)-4-methylcoumarin (CPM). This coumarin derivative contains a thiol-reactive maleimide that quenches coumarin fluorescence in its free form, following reaction with a thiol group the quenching is diminished and coumarin fluorescence can be measured (Chung et al. 2008). This reaction can be coupled to co-enzyme A production during myristoylation (Figure 3-4). Reaction of the maleimide with the CoA thiol generates a fluorescent adduct. The CPM assay is a robust fluorescence based system for continuous reaction monitoring of NMT activity and end-point assays (Goncalves et al. 2012).

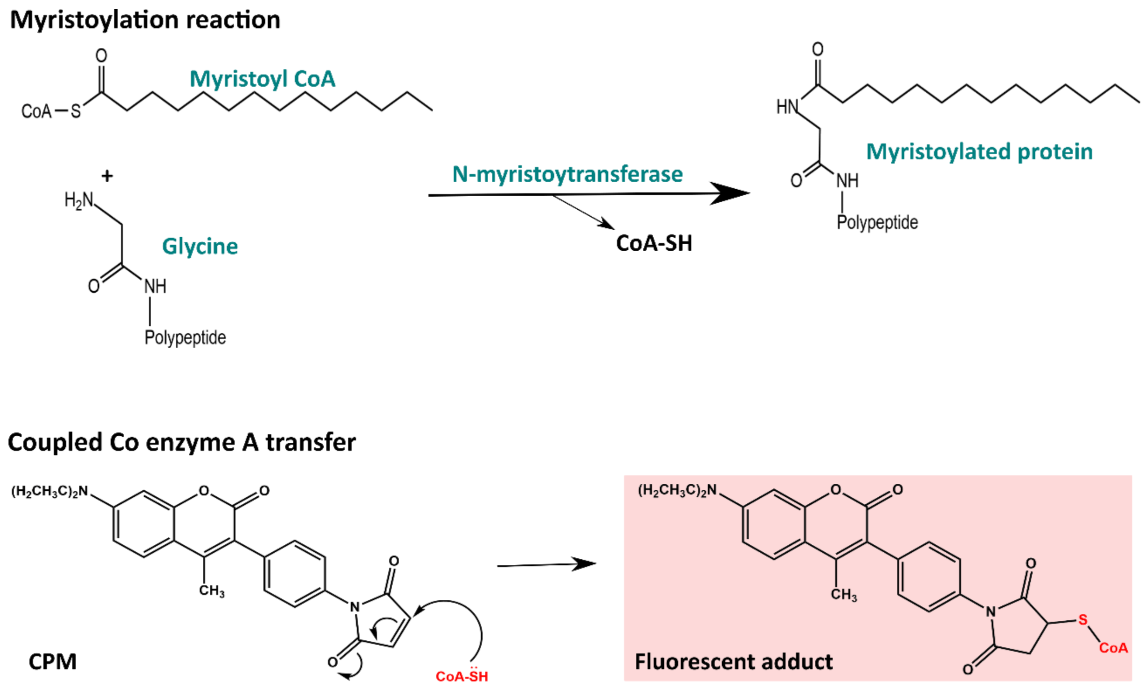


Figure 3-4: The CPM based fluorescence assay used to study the enzymatic activity of NMT. Myristate is transferred from myristoyl CoA by NMT to a free N-terminal glycine residue of a peptide substrate. Co-enzyme A and an N-myristoylated peptide are the products of this reaction. CPM is a thiol-selective dye used to monitor the formation of co-enzyme A via fluorescent adduct production.

The assay was initially tested with the *L. donovani* HASPB2 based peptide GSSSTKD, the positive control for the previous *in vitro* myristoylation work. The CPM setup reported the Michaelis-Menten (K_m) constant as 229.4± 15.3 µM and a V_{max} of 4.67E-04 µM s^{-1} for the peptide (Figure 3-5 and Table 3-4). This K_m value was within the expected µM range for NMTs of *Leishmania* (Brannigan et al. 2010). LdNMT has a K_m value of 31.6 ± 1.0 µM with GSNKSKPK in an excess (4 µM) of myristoyl CoA (personal communication with Dr. Victor Goncalves). This peptide is derived from the N-terminal sequence of the proto-oncogene tyrosine kinase pp60^{src}, a substrate of the NMTs from *Homo sapiens* (Lacal et al. 1988). The data obtained indicated that the assay was robust and could be expanded to define the kinetics of the HASP: NMT interaction.

The Michaelis-Menten (K_m) constant values obtained for all three HASPs were as follows, HASPA = 5.8 ± 0.6 µM, HASPB1 = 16.7± 0.3 µM and HASPB2 = 16.1± 0.7 µM. These values are in agreement with one another and within the µM range expected for NMT (Bhatnagar et al. 1994). The Michaelis-Menten (K_m) constant refers to a molar concentration of substrate at which the enzyme is half saturated under steady state conditions (Copeland. 2013). HASPA reports the lowest K_m value, suggesting that the smallest HASP is the 'best' substrate for LmNMT as it is more efficiently converted to product. The V_{max} is the maximum velocity for an enzyme where the substrate concentration is not limited (Copeland. 2013). The V_{max} values reported for all three HASPs are between 1 to 2 nM s^{-1}, meaning NMT is catalytically more efficient with the HASPs as substrates than the peptide (Table 3-4). The catalytic constant k_{cat} refers to the maximum rate of product formation for an enzyme when it is saturated with substrate. The k_{cat} values reported here for the HASPs are in the order of magnitude 10^{-2} s^{-1}, meaning NMT exhibits a lower than average k_{cat} of 10 s^{-1} for the HASPs (Bar-Even et al. 2011).

The second order rate constant k_{cat}/ K_m relates to the kinetic capacity of an enzyme and is limited to a maximum between 10^8 and 10^9 s^{-1} M^{-1}, equal to the rate of substrate diffusion. It is a particularly effective measure of catalytic capacity because it links substrate affinity to the rate of substrate turnover. Global analysis of the kinetic parameters of thousands of enzymes showed the median k_{cat}/ K_m was in the order of 10^5 s^{-1} M^{-1} (Bar-Even et al. 2011). The k_{cat}/ K_m reported for the HASPs between 10^3 and 10^4 s^{-1} M^{-1}, meaning this reaction has a moderate catalytic efficiency (Table 3-4). Moreover, these values are in line with those reported for *Saccharomyces cerevisiae* NMT (Farazi et al. 2000). Taken together, these kinetic parameters suggest that NMT exhibits moderate catalytic efficiency for the HASPs and are consistent with data reported in the literature for *Leishmania donovani* NMT (Brannigan et al. 2010).

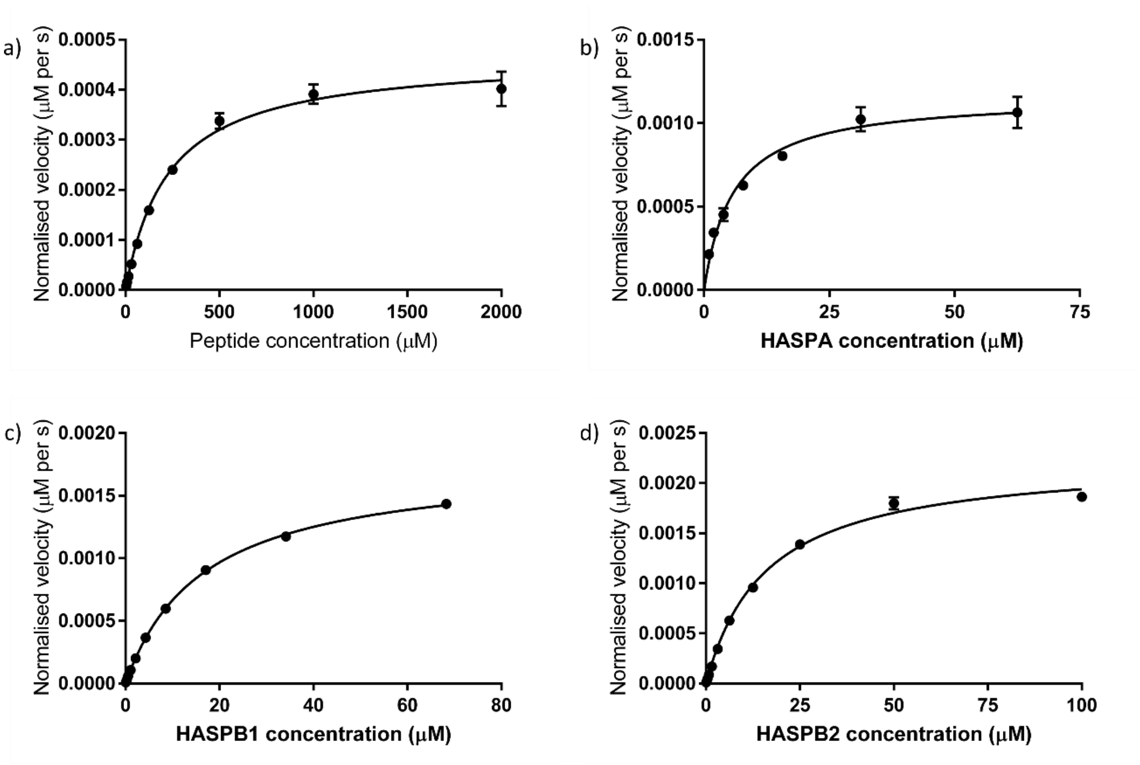


Figure 3-5: Michaelis-Menten plots derived from the CPM assay. The substrate concentration in μM is plotted against the normalised velocity in μM s^{-1}. Michaelis-Menten plot of GSSSTKD peptide (a), HASPA (b), HASPB1 (c) and HASPB2 (d).

	GSSSTKD	HASPA	HASPB1	HASPB2
K_m (μM)	229.4± 15.3	5.8 ± 0.6	16.7± 0.3	16.1± 0.7
Normalised V_{max} (μM s^{-1})	4.67E-04	1.16E-03	1.77E-03	2.25E-03
k_{cat} (s^{-1})	2.34E-02	5.80E-02	8.86E-02	1.13E-01
k_{cat}/ K_m (s^{-1} M^{-1})	1.02E+02	1.84E+04	5.32E+03	7.00E+03

Table 3-4: Michaelis-Menten parameters derived from the CPM assay for the *L. donovani* HASPs and GSSSTKD peptide.

3.3.4. Real-time monitoring of HASPA myristoylation by NMR spectroscopy

Real-time NMR spectroscopy has been used to study the effect and structural impact of post translational modifications on proteins as they occur. A recent study involved the complicated phosphorylation and dephosphorylation processes that modify multiple sites within the intrinsically disordered domain of c-Src (Amata et al. 2013). The effects of various kinases and phosphatases were characterised, allowing certain enzymes to be implicated in the modification of specific residues. This study highlights the role that real-time NMR spectroscopy can play in assessing the structural impact of a modification on a protein and how it can be related back to biological relevance.

Real-time NMR monitoring was used to monitor NMT mediated HASP myristoylation and evaluated any changes or potential structural impact. This entailed recording successive ^{1}H ^{15}N HSQC spectra of [*U*-^{15}N] HASPA incubated with unlabelled NMT at a molar ratio of 100:1 (HASPA:NMT). Spectra were recorded every 5 minutes and the experiment was run overnight at 298K with a 2 fold molar excess of myristoyl CoA, relative to the concentration of HASPA. To ensure that any spectral changes observed were due to myristate addition the NMR sample was subject to ESI-MS analysis after NMT incubation. The expected molecular weight of [*U*- ^{15}N] HASPA was determined to be 9574.7 Da using PROTEIN CALCULATOR v3.4, an online server available at http://protcalc.sourceforge.net/cgi-bin/protcalc. The major peak relating to HASPA after NMT incubation was at 9782.0 Da, in line with the 210 Da mass increase consistent with addition of the myristate group (Figure 3-6).

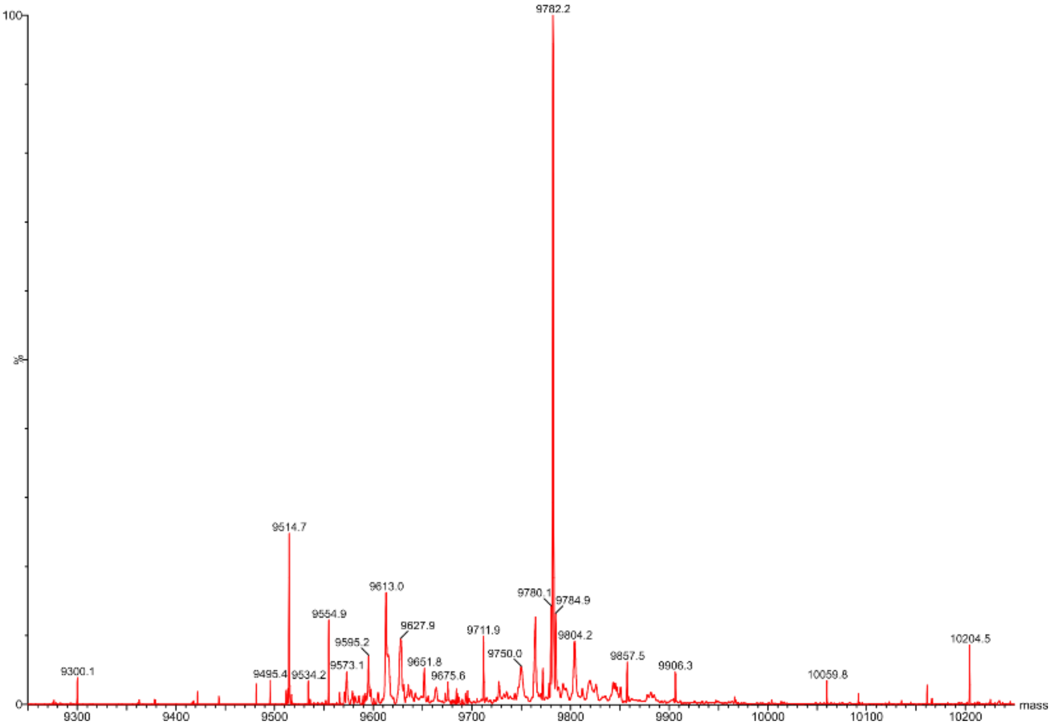
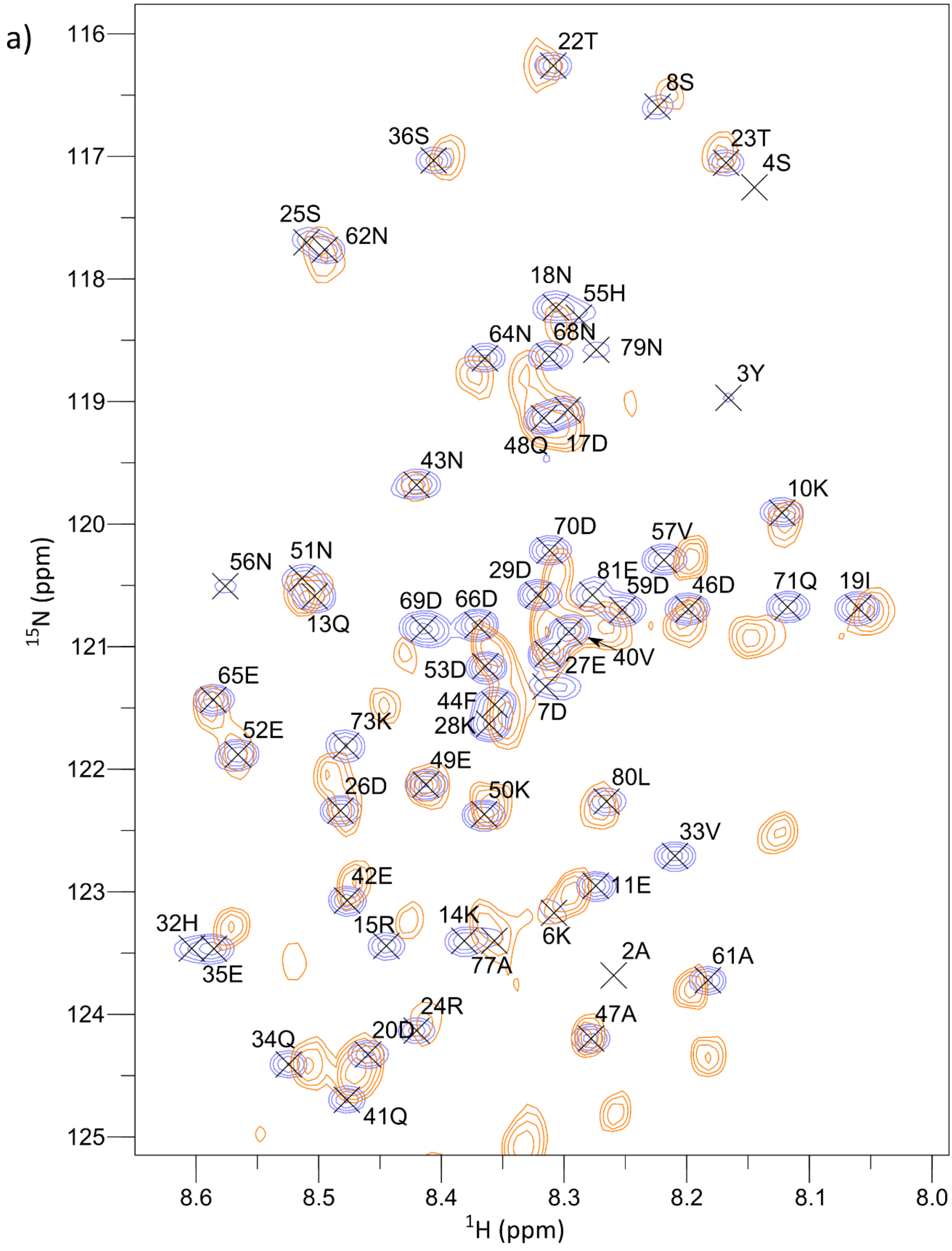


Figure 3-6: ESI-MS of [*U*-^{15}N] HASPA after NMT incubation.

50 of the 73 resonances (69%) that were previously assigned could be confidently transferred onto the t=0 spectrum (Figure 3-7). This partial resonance transfer is due to the difference in the buffer conditions between the two experiments. The assignment was carried out at pH 6.5, while the real-time myristoylation was performed at pH 7.5 to ensure NMT stability. The transfer of assignments between two different buffer conditions means that this data is of limited use, particularly when you consider the limited dispersion of resonances in IDP spectra, and will be treated as such. To achieve a fully comprehensive view of the impact of myristoylation on HASPA it would be beneficial to titrate from pH 7.5 to pH 6.5 to ensure confident assignment transfer or conduct the backbone assignment at pH 7.5.


a)
22T
8S
36S
23T
4S
25S
62N
18N
55H
64N
68N
79N
3Y
48Q
17D
43N
10K
70D
57V
51N
56N
29D
81E
59D
46D
71Q
19I
13Q
69D
66D
40V
53D
27E
65E
7D
44F
28K
52E
73K
49E
26D
50K
80L
33V
42E
11E
32H
15R
14K
6K
35E
77A
2A
61A
24R
47A
20D
34Q
41Q
116
117
118
119
120
121
122
123
124
125
^{15}N (ppm)
8.6
8.5
8.4
8.3
8.2
8.1
8.0
^{1}H (ppm)

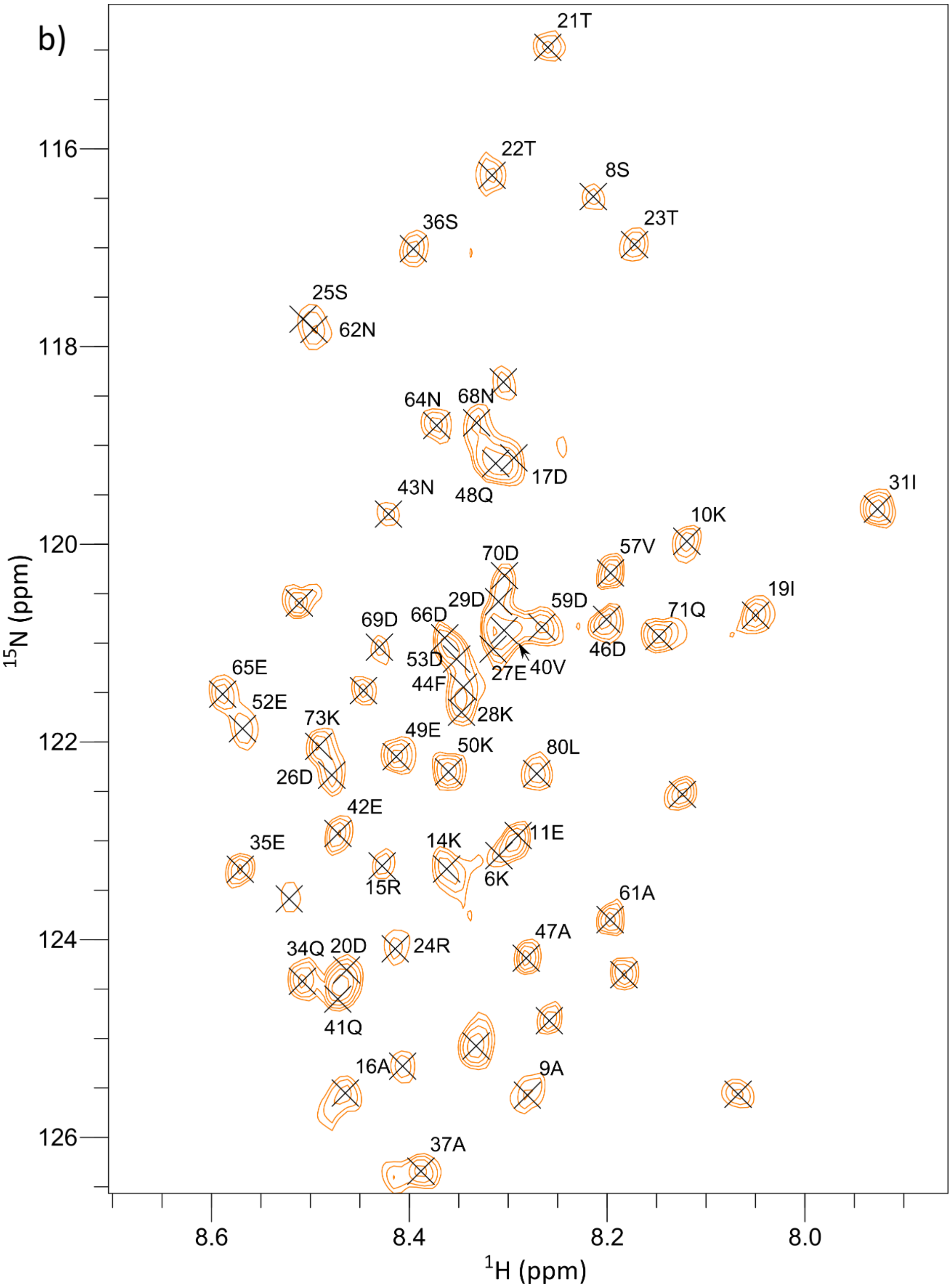


Figure 3-7: Overlay of ^{1}H ^{15}N HSQC spectra of HASPA from backbone resonance assignment and t=0. a) The ^{1}H ^{15}N HSQC spectra from the backbone resonance assignment (blue) with assignments and t=0 (orange) are shown. 50 of the 73 resonances (69%) that were previously assigned could be confidently transferred onto the t=0 spectrum. b) The final resonance transfer onto the t=0 spectrum (orange) showing resonances that could be confidently transferred and those that remained unassigned.

Well-resolved spectra were obtained at each time point and myristoylation was considered complete at 2.5 hours, as determined by the absence of further spectral changes. No detectable secondary structure was induced by NMT-mediated myristoylation of HASPA (Figure 3-8). The ^{1}H ^{15}N HSQC spectrum of [*U*-^{15}N] HASPA remains indicative of a disordered protein with limited dispersion of resonances in the proton dimension. The peaks relating to residues in the N-terminal domain of HASPA, proximal to the N-terminal glycine and site of myristoylation, experienced the most pronounced chemical shift perturbations ($\Delta\delta$) and intensity (ΔI) changes (Figure 3-8a). For example, the peak relating to serine 8 is observed at t=0 minutes but disappears at t=50 minutes (Figure 3-8b) and the peak assigned to lysine 10 gradually disappears between t=0 minutes and t=150 minutes (Figure 3-8c).

Determination of the magnitude of the chemical shift and intensity changes between t=0 minutes and t=175 minutes associated with myristate addition was required for more comprehensive analysis of the system. Notably, chemical shift and intensity changes were only calculated for peaks that could be confidently transferred from the original backbone assignment. The chemical shift changes for the main chain nitrogen and proton resonances between t=0 minutes and t=175 minutes were calculated with $\Delta\delta = [\Delta\delta^2_H + (\Delta\delta_N/R_{scale})^2]^{½}$ to pin-point the impact of myristate addition (Mulder et al. 1999). The first 30 residues, closest to the site of myristoylation, experienced the largest changes in chemical shift (Figure 3-9). The intensity change was calculated with $\Delta I = (I_0 - I_{Myr})/ I_0$ (Mulder et al. 1999). Similarly to the chemical shift changes, the intensity changes were most pronounced in first 20 residues of the N-terminal domain of HASPA (Figure 3-9). The peak for serine 8 disappears entirely over the time course experiment and therefore produces the largest change in intensity of 1, where $\Delta I_{average} = 0.07$. The data presented here shows that myristate addition to HASPA does not induce any global structural changes indicative of a myristoyl switch mechanism. There are very few changes observed in peak position and intensity, indicating that myristate addition has a minor impact on HASPA. The peaks that do experience notable chemical shift or intensity changes are localised to the N-terminus. The small overall chemical shift change indicates that myristoylation has not affected the structure of HASPA. However, comparison of Cα, Cβ and CO chemical shifts would strengthen this conclusion.

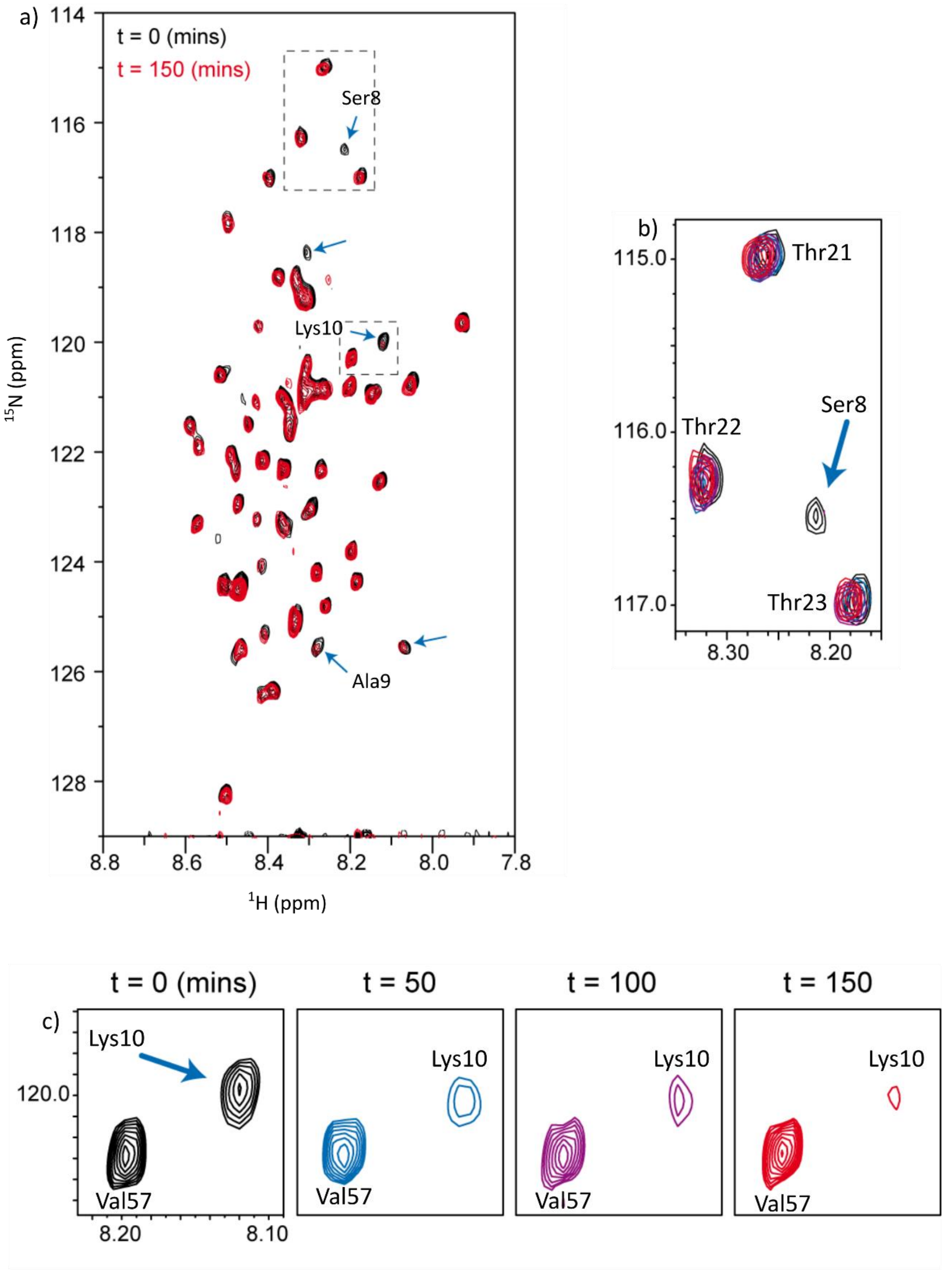
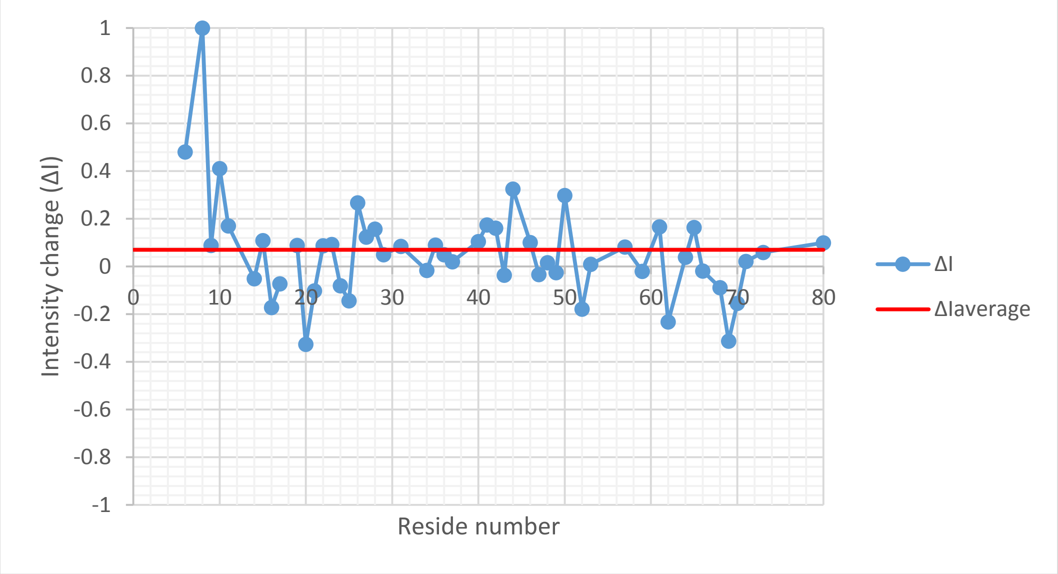


Figure 3-8: Myristoylation of HASPA monitored by real-time NMR spectroscopy. a) Overlay of 2D ^{1}H ^{15}N HSQC spectra of HASPA at 0 (black) and 150 (red) minutes post NMT addition. b) A peak relating to serine 8 is observed at t=0 (black) but is no longer present at t=50 (blue). c) A peak relating to lysine 10 is observed at t=0 (black) but disappears from t= 50 (blue) to t=150 (red). Note, t=100 is shown in purple.



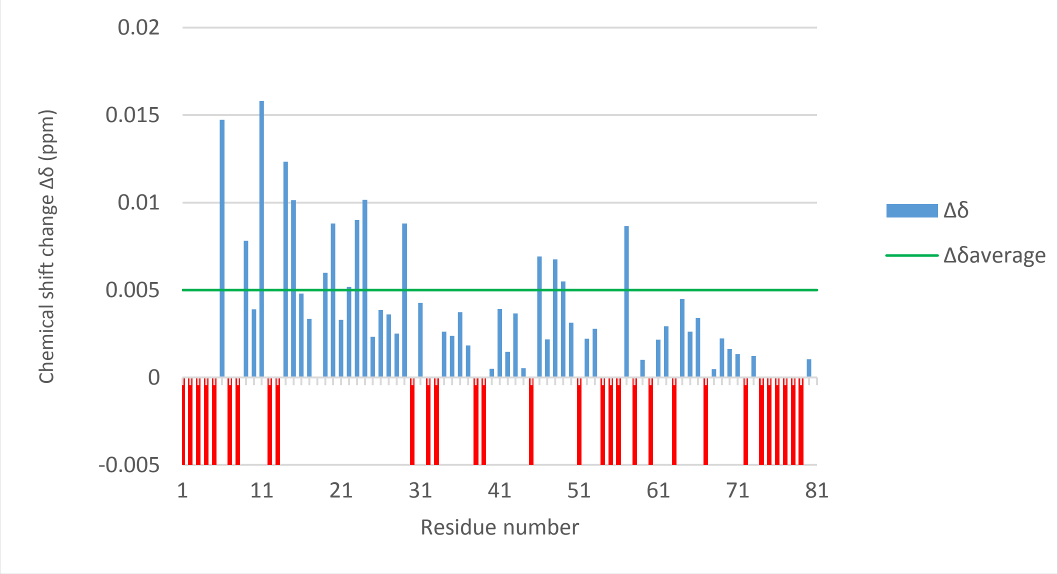


Figure 3-9: Analysis of the peak intensity and chemical shift changes between t=0 and t=175 spectra. Top) Intensity changes between t=0 and t=175 spectra, where $\Delta I_{average}$ = 0.07 and is shown in red. The most prominent intensity change is observed in the first 20 residues, particularly as the peak for S8 disappears. Bottom) Chemical shift changes between t=0 and t=175 spectra, where $\Delta\delta_{average}$ = 0.005 ppm and is shown in green. The most pronounced area of chemical shift change is the N-terminal domain of HASPA. S8 disappears entirely so was not included in this analysis. Note to residues that do not have data are included but shown in red (-0.005 ppm) to frame the available data better.

3.3.5. NMT: HASPA co-crystallisation

Once *in vitro* HASP myristoylation had been established it was decided that co-crystallisation would be explored. There is limited structural data concerned with peptide binding to NMT (Bhatnagar et al. 1998; Farazi et al. 2001a). Therefore, determination of a HASPA: NMT complex would be of great benefit to the field. NMT was prepared for crystallisation with the hexa-histidine tag removed and complexed with non-hydrolysable myristoyl CoA analogue. The proteins were dialysed into 20 mM Hepes pH 7.8, 50 mM NaCl and mixed at 2:1 or 5:1 molar ratios (HASPA: NMT). One crystallisation screen, the CSSI+II screen supplemented with 0.1 M Tris pH 7.8, was prepared with both complexes (Brzozowski and Walton. 2001). The crystal shower shown in Figure 3-10a was obtained from initial buffer optimisation of promising CSS I+II screen conditions. It was further optimised into a cluster of crystals shown in Figure 3-10b. A fragment taken from the cluster diffracted to 2.45 Å, when sent for data collection at the Diamond Light Source.

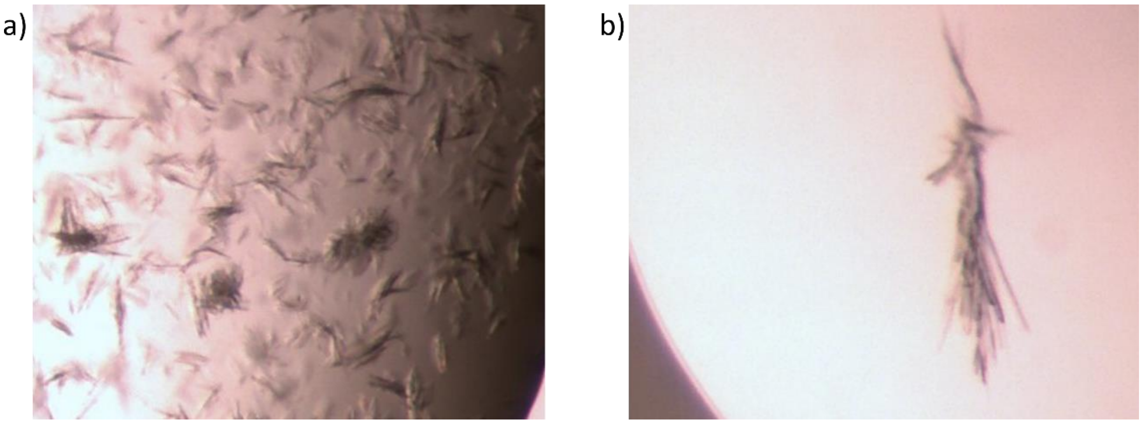


Figure 3-10: Crystals obtained from NMT: HASPA co-crystallisation work. a) Crystal shower from 0.1 M sodium acetate pH 4.5, 25% (w/v) PEG 2000 mme and 200 mM $Ca(C_2H_3O_2)_2$ with the 5:1 complex (HASPA: NMT). b) Optimised crystal (from condition in a) used for structure determination of apo LmNMT. This condition was 0.1 M sodium acetate pH 4.5, 20% (w/v) PEG 2000 mme and 50 mM $Ca(C_2H_3O_2)_2$.

The data were integrated to 2.45 Å and all the related statistics are outlined in Table 3-5. The structure was solved by molecular replacement using 4CGP, *Leishmania major* NMT with myristoyl CoA bound, as the search model (Brannigan et al. 2014). It was clear after molecular replacement that NMT had been crystallised with the non-hydrolysable myristoyl CoA analogue bound but without HASPA bound. There were two NMT molecules in the *P*1 unit cell and therefore the asymmetric unit, which represents a new crystal form for LmNMT. Both chains were well defined in the electron density map, which allowed efficient model building to be undertaken. The density was poorly defined for residues 83 and 84, 239 to 242 and 334 to 337. They were assumed to be disordered and they were removed from the model. Myristoyl CoA (PDB code: MYR) was replaced with 2-oxopentadecyl-CoA (PDB code: NHM) to complete the structure. It is important to note that at the time of submission work on this structure is still in progress, hence the presence of Ramachandran outliers (Table 3-5). The current R_{work}/R_{free} are 0.23/0.31, respectively.

	NMT crystal data
Data collection	
X-ray source	DLS beamline I02
Detector	Pilatus
Temperature	100 K
Cryoprotectant	none
Wavelength (nm)	0.97949
Space group	*P*1
Cell dimensions	
a, *b*, *c* (Å)	53.18 61.47 75.12
α, β, γ (°)	85.6 80.0 78.5
Resolution (Å)	38.4 – 2.45(2.51-2.45)
R_{sym}	0.139(0.822)
$<I / \sigma(I)>$	5.0(1.2)
Completeness (%)	95.7(95.6)
Wilson *B* Factor (A^2)	28.1
Refinement	
Resolution (Å)	38.4 – 2.45
No. reflections	28676
R_{work} / R_{free}	0.23/0.31
R.m.s. deviations	
Bond lengths (Å)	0.020
Bond angles (°)	2.044
B-factors	
Protein	28.3
Ligand/ion	29.8
Water	31.2
Ramachandran plot	
Preferred	91
Allowed	6.2
Outliers	2.8

Table 3-5: Data collection and refinement statistics of LmNMT crystal. All X-ray diffraction data were collected on one crystal only. Values in parentheses relate to the highest-resolution shell.

3.3.6. Structure of *Leishmania major* NMT

A search of the RCSB Protein Databank (PDB) reveals that the structure of NMT has been solved from the following organisms, *Leishmania major*, *Saccharomyces cerevisiae*, *Homo sapiens, Leishmanai danavani, Plasmodium vivax, Aspergillus fumigatus* and *Candida albicans*. In all of these structures NMT is either complexed with myristoyl CoA (PDB code: MYR) or the non-hydrolysable myristoyl CoA analogue 2-oxopentadecyl-CoA

(PDB code: NHM) and in some cases a peptidomimetic inhibitor. These structures show the enzyme to be monomeric with a compact globular structure (Wright et al. 2010).

The NMT fold, the term given to the common structural features of NMT, consists of a saddle shaped β-sheet spanning the core of the protein encircled by three α-helices. This fold encloses the C-terminal portion of the protein chain, meaning the α-carboxylate of the C-terminal leucine residue is at the centre of the protein (Figure 3-11a). The protein has two distinct lobes that display a pseudo two fold symmetry and bind each substrate separately. Co-crystal structures with myristoyl CoA and peptide analogues bound show that the N-terminal domain binds myristoyl CoA in a bent question mark conformation (Figure 3-11a left hand domain), while the C-terminal domain binds the peptide substrate (Figure 3-11a right hand domain) (Bhatnagar et al. 1998; Farazi et al. 2001a).

The asymmetric unit contained two independent NMT molecules with well-defined density in each case. The two chains are highly similar with an RMSD on all Cα positions of 0.44 Å (Figure 3-11b). Chain A is used in the rest of this discussion (Figure 3-11a). This structure of LmNMT in complex with 2-oxopentadecyl-CoA is highly similar to LmNMT in complex with myristoyl CoA (PDB code: 4CGP), reporting RMSD on all Cα positions of 0.91 Å (Brannigan et al. 2014). These differences are due to minor loop movements required to accommodate the different ligands (Figure 3-11c). Overlay of the structure with LdNMT in complex with 2-oxopentadecyl-CoA (PDB code: 2WUU), highlights the high structural similarity between the two species especially when complexed with the same ligand (Brannigan et al. 2010). The RMSD on all Cα positions between the structures is 0.81 Å (Figure 3-11d).

2-oxopentadecyl-CoA (NHM) is a potent inhibitor of N-myristoyltransferase and a non-hydrolysable analogue of myristoyl CoA (Paige et al. 1989). NHM differs from myristoyl CoA as it contains an extra methylene group between the CoA sulphur and the thioester carbonyl of myristoyl CoA (Figure 3-12). NHM and myristoyl CoA bind in the N-terminus containing lobe of NMT in a groove that allows the long aliphatic chain of the myristate to be accommodated by a deep hydrophobic cavity in the protein. While the CoA moiety, the portion of NHM upstream of the sulphur, is more exposed on the surface (Figure 3-12). The moiety folds around its adenine base, giving rise to the distinctive ‘bent question mark’ shape observed in NMT bound forms (Bhatnagar et al. 1998).

The residues implicated in NHM binding in LmNMT are analogous to those involved in the LdNMT structure (Brannigan et al. 2010). The terminal phosphate group of CoA interacts with the side chain of histidine 12 and arginine 179, while also being stabilised by the backbone amides phenylalanine 14 and tryptophan 15 (Figure 3-12). The diphosphate group interacts with the backbone amides of arginine 179, leucine 180 and alanine 181. The carbonyl group of leucine 169 and side chain of arginine 176 contribute to further stabilising interactions with the ligand. NHM is bound in the NMT distinctive 'bent question mark' conformation, where the moiety folds around the adenine base of CoA. Superposition of NHM from this structure and the LdNMT structure highlight this feature (Figure 3-12).

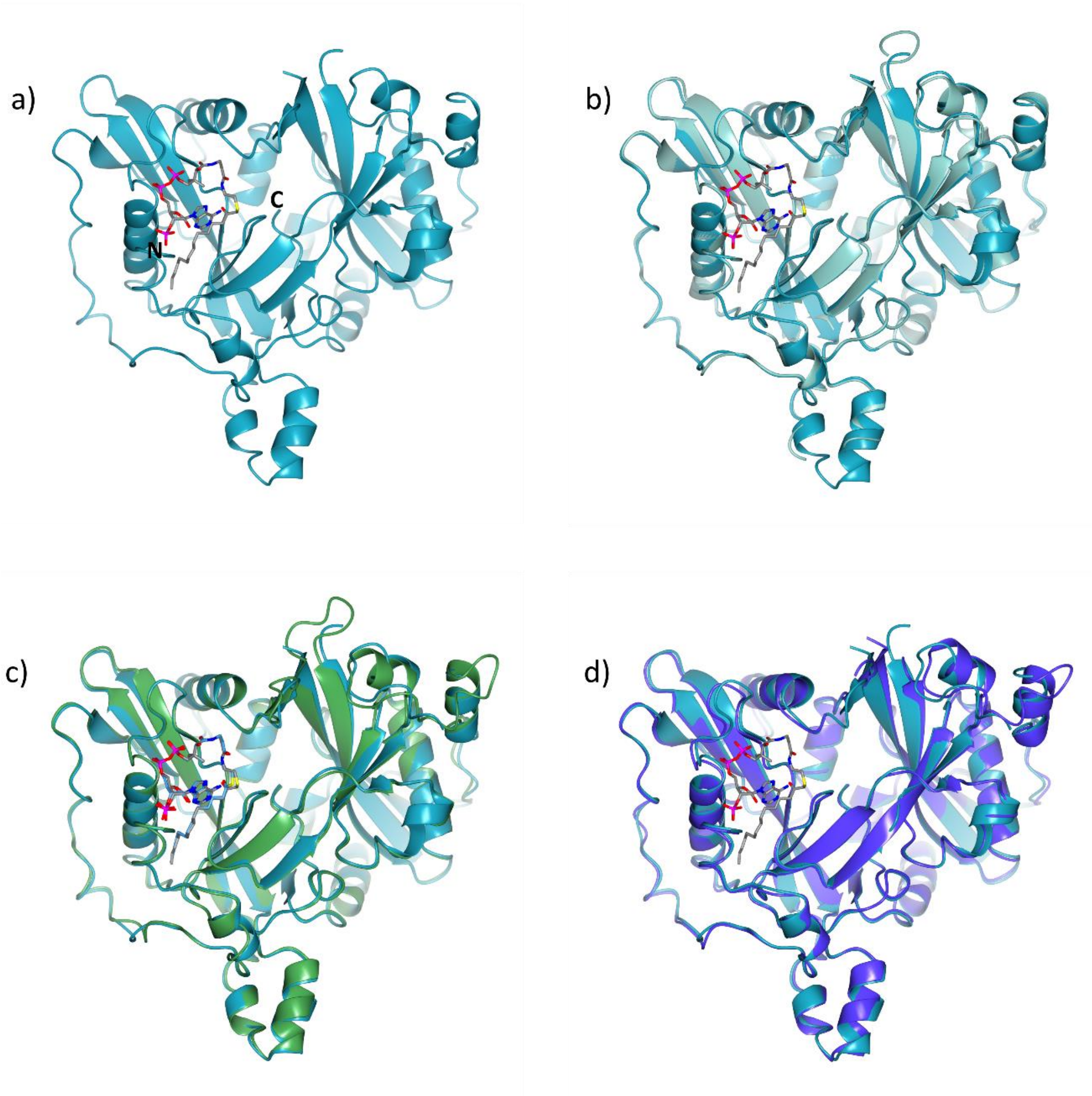


Figure 3-11: Structure of *Leishmania major* NMT in complex with 2-oxopentadecyl-CoA. a) LmNMT (cyan) in complex with 2-oxopentadecyl-CoA (grey), with the N- and C-terminal residues labelled as N and C. b) Overlay of the two molecules of LmNMT from the asymmetric unit in cyan and light green. The RMSD on Cα positions between these two molecules is 0.44 Å. c) Overlay of (a) with LmNMT complexed with myristoyl CoA (PDB code: 4CGP) in green. The RMSD on Cα positions between these two molecules is 0.92 Å. d) Overlay of (a) with LdNMT complexed with 2-oxopentadecyl-CoA (PDB code: 2WUU) in purple. The RMSD on Cα positions between these two molecules is 0.81 Å. This figure and subsequent structures were prepared with CCP4mg (McNicholas et al. 2011).

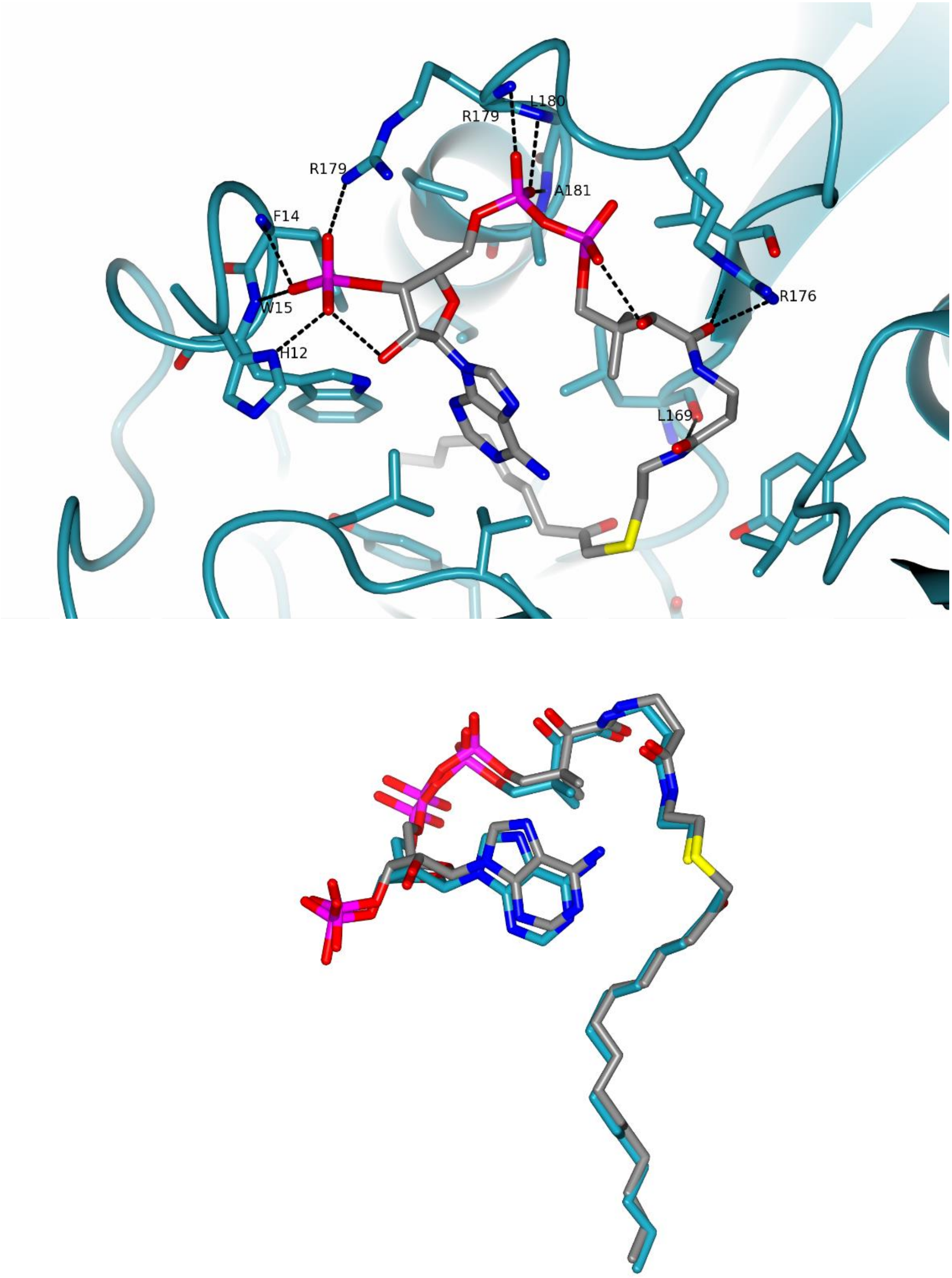


Figure 3-12: Residues implicated in 2-oxopentadecyl-CoA bound to LmNMT. Top) 2-oxopentadecyl-CoA bound to LmNMT, highlighting residues implicated in the interaction. The backbone of the protein is shown in cyan and the carbon atoms of 2-oxopentadecyl-CoA is in grey. The remaining atoms are shown in red (oxygen), blue (nitrogen), yellow (sulphur) and magenta (phosphorus). Polar interactions between the protein and 2-oxopentadecyl-CoA are indicated by dashed lines. Bottom) Overlay of 2-oxopentadecyl-CoA (NHM) bound to NMT in the bent question mark conformation. The carbon atoms are coloured according to the structure; LdNMT (cyan) and LmNMT (grey). 2-oxopentadecyl-CoA (NHM) differs from myristoyl CoA as it contains an extra methylene group between the CoA sulphur and the thioester carbonyl of myristoyl CoA.

3.4. Discussion

The results presented in this Chapter have shown for the first time that NMT is able to catalyse HASPA myristoylation *in vitro*. This was initially verified by directly incubating NMT with HASPA or a known peptide substrate with an excess of myristoyl CoA. ESI-MS showed that both substrates experienced a 210 Da mass increase, indicative of myristate addition, after NMT incubation. The next step to defining this reaction was to obtain kinetic parameters for myristoylation. The CPM assay, a fluorescence based assay where fluorescence is linked to CoA production, was implemented to define the reaction (Goncalves et al. 2012). NMT reported K_m values between 5.8 and 16.7 µM for the three full length HASPs, in line with µM values reported for NMT with various peptide substrates with this method (personal communication with Dr. Victor Goncalves). HASPA, as the smallest HASP, reported the lowest K_m value suggesting that it is the 'best' substrate for LmNMT as it is more efficiently converted to product. All further work to define this interaction used HASPA as the model for all HASPs.

Real-time NMR spectroscopy was used to monitor the effect and potential structural impact of NMT catalysed myristoylation on HASPA. In some cases myristate addition elicits a structural change in the myristoylated protein to protect the hydrophobic moiety from the hydrophilic environment. This is termed a myristoyl switch mechanism and is typified by a dynamic shuttling between myristate exposed and buried forms. This conformational change can be modulated by ligand binding, proteolysis and electrostatic interactions. Recoverin, a 23 kDa neuronal protein, is a myristoyl switch modulated by calcium binding (Ames et al. 1996). The binding of two calcium ions to recoverin induces a conformational change that expels the myristate moiety, exposing the hydrophobic amino acids onto the surface. Recent NMR analysis highlighted three distinct conformational states, notably a short lived intermediate state with one calcium ion bound was confirmed (Xu et al. 2011). The dynamic nature of the conformal shifts observed in myristoyl switches enables functional regulation. Successive ^{1}H ^{15}N HSQC spectra of [U-^{15}N] HASPA incubated with unlabelled NMT showed that the effect of myristate addition, determined by chemical shift and intensity changes of peaks, was localised to residues close to the N-terminal glycine and site of myristoylation. Moreover, addition of myristate did not induce any secondary structure in HASPA that would suggest a myristoyl switch mechanism.

Lastly, attempts to crystallise the HASP: NMT complex produced an apo structure of LmNMT in complex with 2-oxopentadecyl-CoA (NHM) at 2.45 Å. The RCSB Protein Databank (PDB) does not contain a structure of *Leishmania major* NMT in complex with NHM alone, so this represents a unique structure of this protein. The binding of NHM is analogous to that observed in a previously published LdNMT structure (Brannigan et al. 2010). To date there is limited structural information on NMT complexed with peptides (Bhatnagar et al. 1998; Farazi et al. 2001). Therefore, attempting co-crystallisation of the HASP: NMT complex was a viable strategy to enhance our understanding of protein substrate binding to NMT. This is particularly important when you consider that NMT has been explored as a potential drug target against leishmaniasis and enhanced understanding of substrate would further this work (Price et al. 2003; Brannigan et al. 2014). It should be noted that there are published examples of structures obtained by NMR and X-ray crystallography of IDP: binding partner complexes, such as the extended helix formed by PUMA when bound to MCL-1 (Rogers et al. 2014).

3.5. Summary and future work

The work presented in this Chapter has confirmed that the HASPs can be myristoylated by NMT *in vitro*. Kinetic parameters established with the fluorescence based CPM assay were in line with those reported for peptide substrates of *Leishmania donovani* NMT and within the expected µM range (Goncalves et al. 2012). Real-time NMR spectroscopy was used to monitor the effect of myristoylation on HASPA in solution. Successive ^{1}H ^{15}N HSQC spectra of [*U*-^{15}N] HASPA showed that myristate addition did not induce any secondary structure. The impact was modest and localised to residues proximal to the N-terminal glycine and site of myristoylation. Detailed analysis of the effect of myristoylation on assigned peaks was calculated by chemical shift and intensity changes, confirmed the impact on the N-terminal domain. Co-crystallisation efforts with the HASPA: NMT complex produced an apo structure of LmNMT in complex with 2-oxopentadecyl-CoA (NHM) at 2.45 Å.

To continue examination of the HASP: NMT interaction a peptidomimetic substrate based on the HASP N-terminal domain with enhanced binding properties could be designed. This would be an effective strategy to improve co-crystallisation efforts with

NMT and if successful would inform on substrate binding. The work presented here has shown that myristoylated HASPA can be produced. Myristoylation alone is too weak to mediate irreversible membrane binding, hence this modification is usually in found in combination with other stronger mediators of membrane binding (Goldston et al. 2014). However, it would be interesting to explore how myristoylated HASPA interacts with a biological membrane mimetic using NMR spectroscopy. Does it shuttle between bound and unbound as expected? This could be achieved by comparing the diffusion co-efficient of myristoylated HASPA with and without membrane mimetic using diffusion-ordered spectroscopy (DOSY).

Chapter 4 - Crystallisation of the HASPs

4.1. Introduction

This project was conceived as an X-ray crystallography study to decipher functionality from resulting HASP structural data. Various non-standard crystallisation schemes were implemented to induce crystal formation before intrinsic disorder was firmly established. These strategies are outlined here and in Chapter 5. Although unsuccessful, discussion of protein engineering to enhance crystallisability and improve crystal properties is constructive, particularly when you consider the challenging structures still to be solved by X-ray crystallography such as integral membrane proteins, multi-protein assemblies and transient conformational states.

4.1.1. Fab antibody fragments as a non-covalent crystallisation chaperone

Crystallisation is the major rate-limiting step in X-ray crystallographic studies of proteins, with a wide array of strategies in place to circumvent this issue (Derewenda. 2010). In particular, chaperone-assisted crystallography is becoming an increasingly efficacious and versatile approach. The chaperone reduces the conformational freedom of the target protein, stabilising it and increasing likelihood of crystal growth by reducing the entropic cost of lattice formation. Moreover, chaperone-assisted crystallography has led to improvements in diffraction quality, producing higher resolution structures. This strategy also facilitates the stabilisation of alternative but functionally relevant conformations of proteins to be fixed, contributing to a deeper understanding of protein function (Bukowska and Grütter. 2013).

The Fab (fragment antigen binding) fragment of an antibody is a widely used non-covalent crystallisation chaperone that has led to a number of crystal structures, particularly of membrane proteins (Lieberman et al. 2011). A recent example of the scope of Fab mediated crystallisation involves LeuT, a model system for neurotransmitter sodium symporter research, that was crystallised in a range of conformational forms (Krishnamurthy and Gouaux. 2012). LeuT mutants were generated to induce an apo state in the protein, which were used to raise apo state specific antibodies. The Fab-stabilised LeuT mutants were crystallised and revealed two new conformational forms, namely substrate free outward open and inward open forms. This new structural information, taken together with the already determined occluded form, was used to enhance the mechanistic understanding of LeuT mediated transport.

The inherent conformational heterogeneity of IDPs represents a significant challenge for structure determination. However, the Fab co-crystallisation strategy has been successfully implemented to obtain IDP structures, as they are able to capture structural and functionally relevant snapshots of these dynamic proteins. Tau is an intrinsically disordered microtubule-associated neuronal protein that, when misfolded, is implicated in Alzheimer’s disease. It is the main component of the paired helical filaments (PHFs), a defining pathological feature of Alzheimer’s disease (Mudher and Lovestone. 2002). A combined X-ray crystallography and NMR spectroscopy approach did not yield any structural information on the isolated form of tau. This is unsurprising when you consider that tau is an IDP, so exhibits high conformational heterogeneity in solution. Moreover, the PHF form is insoluble. Conformation dependent antibodies that stabilise distinct structural states of tau were generated from native source derived protein (Wischik et al. 1988). Complex formation of the PHF core C-terminal tail of tau with the specific MN423 Fab fragment produced crystals that were used to solve the structure (Sevcik et al. 2007). The MN423 Fab fragment acts as a ‘molecular scaffold’, inducing a folding on binding mechanism on tau and imposing a pre-selected native conformation (Skrabana et al. 2010).

Another example of Fab induced stabilisation and crystallisation of an IDP is the HIV-1 transcriptional activator protein and potential vaccine candidate Tat. It is vital for viral replication and disease pathogenesis. Tat is an IDP that functions as an interaction hub, as it is able to interact with cellular and viral binding partners in both the intracellular and extracellular environments (Foucault et al. 2010). NMR spectroscopy was used to confirm the intrinsic disorder of Tat but also established elements of structural propensity that contribute to the folding on binding required for its wide range of interactions (Shojania and O'Neil. 2006). Monoclonal antibodies were raised against the N-terminal domain of Tat, implicated in viral replication, and used to crystallise two peptides based on this domain (Serrière et al. 2011). Fab’ fragments, generated by pepsin digestion and reduction with 2-mercaptoethanol, were produced from the monoclonal antibodies for complex formation. Direct comparison of the Fab’ co-crystal structures with the Tat: p-TEFb complex shows the N-terminal domain binds in an extended conformation with the conserved and functionally important tryptophan at position 11 exposed (Tahirov et al. 2010). Both structures show that when bound Tat adopts a type I β turn conformation (Figure 4-1). However, the difference in the position of histidine 13 means that Tat is able to form an α helix in the p-TEFb structure but not in the Fab’ bound form. This work highlights the structural plasticity of Tat that enables it to bind to various biological partners by adopting different structural conformations.

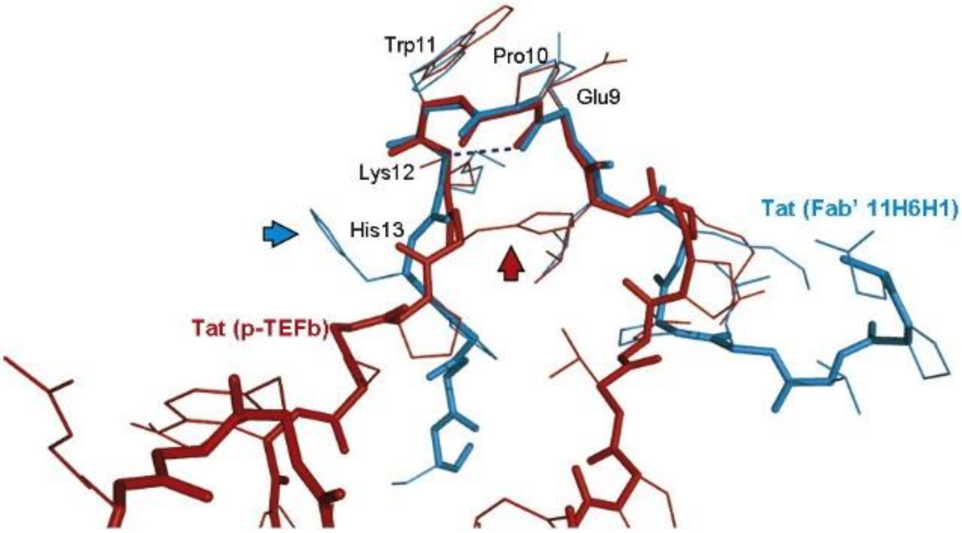


Figure 4-1: Overlay of Tat structures from complexes with p-TEFb and Fab’. The structure of the peptide based on the N-terminal domain of Tat in complex with specific Fab’ (blue) is overlaid with the same residues from the Tat: p-TEFb complex (red). The hydrogen bond, between glutamic acid 9 and lysine 12, required for the β-turn conformation of both peptides is shown by the dashed blue line. The position of the histidine 13 side chain is the major difference between these two structures and is highlighted by blue (Fab’: Tat peptide) and red (Tat: p-TEFb) arrows. This figure was taken from (Serrière et al. 2011).

4.2. Experimental

4.2.1. Expression and purification of the HASPs

Aliquots of LB media (10ml) containing kanamycin and chloramphenicol, at 50 and 30 µg/ml, respectively, were inoculated with glycerol stocks of *E. coli* BL21 (DE3) cells transformed with either pET28b_haspa, pET28b_haspb1 or pET28b_haspb2. Briefly, the pET28b plasmid is isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible, kanamycin resistant and contains a non-cleavable C-terminal hexa-histidine tag. The cells were grown overnight at 37°C with shaking at 180 rpm and each 10 ml aliquot was used to inoculate 500 ml LB media. These cultures were grown to an OD_{600} of 0.8 at 37°C with shaking at 180 rpm. Protein expression was induced with addition of IPTG to 1 mM and the cultures were transferred to 30°C for 4 hours with shaking at 180 rpm. The cells were harvested by centrifugation for 15 minutes at 5000 xg and the resulting pellets stored at -20˚C. Once thawed, the pellets were resuspended in 30 ml nickel column buffer A (50 mM Tris pH 7.5, 500 mM sodium chloride, 20 mM imidazole) per 500 ml pellet supplemented with a protease inhibitor tablet (cOmplete, EDTA-free from Roche). The soluble cell suspension was sonicated 6 times on ice for 30 seconds with 1 minute intervals. Lysate clarification was achieved by centrifugation at 38000 xg for 30 minutes. The supernatant was loaded onto a 1 ml HisTrap FF crude column (GE Healthcare) pre-equilibrated with nickel column buffer A. Five column volumes of nickel column buffer A washed the column before a gradient of 0 to 100% nickel column buffer B (50 mM Tris pH 7.5, 500 mM sodium chloride, 500 mM imidazole) was implemented for 20 column volumes to elute the target protein. SDS-PAGE was used to identify target protein containing fractions that were concentrated (Vivaspin Sample Concentrator range, 10000 MWCO for HASPB1 and 3000 MWCO for HASPA and HASPB2) to 2 ml for size exclusion purification. A Superdex 75 16/60 column (GE Healthcare) equilibrated with 20 mM sodium cacodylate pH 6.5, 120 mM sodium chloride and used to further purify the HASPs. Fractions were then analysed with SDS-PAGE to assess final sample purity. To enhance sample purity HASPB1 and HASPB2 were subject to an ion exchange step after nickel affinity purification. The HASPB samples were diluted with 20 mM Tris pH 7.5 to give a final concentration of 50 mM NaCl and loaded onto a 1 ml HiTrap Q HP column equilibrated with ion exchange buffer A (20 mM Tris pH 7.5, 50 mM NaCl). Elution was achieved with a gradient of 0 to 100% ion exchange buffer B (20 mM Tris pH 7.5, 1 M NaCl) for 20 column volumes. All purification columns were carried out with the AKTA FPLC system.

ESI-MS was used to confirm the mass of each full length HASP. This in-house service was run by Dr. Andrew Leech with the proteins supplied at 2 mg/ml in 20 mM sodium acetate pH 5.5, 15 mM sodium chloride. Data was collected with an ABI Qstar tandem mass spectrometer. The polydispersity and batch consistency of each HASP was analysed with dynamic light scattering (DLS). Measurements were carried out at 20˚C with the protein at 1 mg/ml in size exclusion buffer with the DynaPro Dynamic Light Scattering system (Protein Solutions).

4.2.2. OPA concentration determination

Phthaldialdehyde reagent (OPA) was obtained from Sigma Aldrich (CAS Number 643-79-8). It is important to note that before readings can be taken all reagents and samples must be equilibrated to room temperature. Moreover, the OPA stock must be activated with 2.5 µl β-mercaptoethanol per ml of reagent used. The working range of the assay is a protein concentration of 10 to 500 µg/ml. Bovine serum albumin (BSA) standards were prepared from a 2 mg/ml bought standard stock (Pierce) in this range in the HASP size exclusion buffer (20 mM sodium cacodylate pH 6.5, 120 mM sodium chloride). Serial dilutions of the sample to be tested are made as follows; 1:10, 1:100 and 1:1000 (sample: buffer). This enabled a wide range of possible concentrations to be covered. The standard protocol uses a ratio of 1:10 sample to reagent, so in the 96 well format 200 µl reagent is added to 20 µl sample. To streamline OPA dispensing, a multi-channel pipette was used to add the reagent to the plate and incubated for 1 minute and 30 seconds before the reading was taken. Excitation was set at 355 nm and emission at 460 nm. The 96 well format mean that each reading was taken in quadruplicate to ensure the statistical validity of the setup. The readings were averaged and normalised by a blank reading. Normalised fluorescence of the standards was plotted against concentration and this standard curve was used to determine the unknown sample concentration.

4.2.3. Expression and purification of HASPB1 fragments

The HASPB1 fragments were cloned into the pET-YSBLIC3C vector using ligation independent cloning (Bonsor et al. 2006). This work was outlined in his undergraduate thesis and should be referred to for a detailed description. Four constructs were produced pET-YSBLIC3C_Nt, pET-YSBLIC3C_Ntr, pET-YSBLIC3C_Ct and pET-YSBLIC3C_Ctr. Only two of these constructs (pET-YSBLIC3C_Nt and pET-YSBLIC3C_Ntr) expressed soluble protein when transformed into *E. coli*. This vector contains an N-terminal hexa-histidine tag connected via a HRV 3C protease recognition site for subsequent removal and is kanamycin-resistant.

The expression and purification of the HASPB1 fragments was similar to the full length HASPs (outlined in Section 4.2.1.). However, methodological variations will be mentioned here. Firstly, after IPTG induction the cultures were transferred to 16°C for 18 hours with shaking at 180 rpm. Moreover, the hexa-histidine tag was removed by 3C protease digestion (1:50 ratio of protease to protein) overnight while the sample was dialysed into 50 mM Tris pH 7.5, 500 mM sodium chloride, 20 mM imidazole, 1 mM dithiothreitol (DTT). The nickel purification step was repeated but the cleaved samples eluted in the nickel column buffer A wash step. The ion exchange purification step was not required as everything that bound in the first nickel step bound again, making hexa-histidine tag removal an efficient purification step.

4.2.4. SEC-MALLS analysis of the HASPs and HASPB1 fragments

All experiments were conducted at room temperature on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector installed at the Technology Facility, Department of Biology, University of York. The size exclusion column attached to the system was a Superdex 75 10/300 GL (G.E. Healthcare). For HASPA and HASPB1 the buffer was 40 mM sodium phosphate pH 7.5, 120 mM NaCl and the proteins were supplied at 1 mg/ml. The HASPB1 fragments were analysed in 20 mM Tris pH 7.5, 120 mM sodium chloride and were supplied at approximately 3 mg/ml and encompassed NtR with his-tag, NtR without his-tag, Nt with his-tag and Nt without his-tag. All samples were 0.22 µM filtered before use.

4.2.5. Circular dichroism of Nt fragment

The HASPB1 Nt fragment (without hexa-histidine tag) was evaluated with the Jasco J810 CD Spectrophotometer installed at the Technology Facility, Department of Biology, University of York. The sample was analysed at 0.2 mg/ml in 20 mM sodium phosphate pH 7 with 50 mM sodium chloride in a 1 mm quartz cuvette. The buffer was 0.22 µm filtered before use. Data were collected between 195 and 260 nm. The CD signal was considered valid between these points, as the HT voltage was below 600 but saturated at a wavelength lower than 195 nm.

4.2.6. Fab preparation and purification

5 mg of rabbit polyclonal anti-HASPB1 IgG (a gift from Mike Hodgkinson) was dialysed into 20 mM sodium phosphate pH 7, 10 mM EDTA. This sample was concentrated to 20 mg/ml and then diluted to 10 mg/ml with digestion buffer (sample buffer plus 20 mM cysteine. HCl). Immobilised papain (Thermo Scientific) was used to generate Fab fragments. The papain is supplied as a 50% slurry and was equilibrated with digestion buffer 3 times before use. Between wash steps the buffer was removed by centrifugation at 1200 xg for 5 minutes. For the digestion, an enzyme to substrate ratio of 1:10 was used, so an appropriate volume of immobilised papain was prepared. The concentration of the settled slurry is 250 µg/ml. The reaction was then incubated for 4 hours at 37°C at 750 rpm then stopped by centrifugation and removal of the supernatant.

Fab purification was carried out using Protein A HP column affinity purification. The sample was loaded onto a 1 ml HiTrap Protein A HP column (GE Healthcare) pre-equilibrated with buffer A (20 mM sodium phosphate pH 7). The loaded column was washed with 10 column volumes buffer A and the bound sample was eluted with 5 column volumes buffer B (0.1 M citric acid pH 3). The column was eluted at 0.5 ml per minute and 0.5 ml fractions were collected. To neutralise the eluted fractions, 200 µl 1 M Tris pH 8.8 was added.

4.2.7. HASPB1: Fab complex purification and crystallisation

Due to the polyclonal nature of the Fab sample further purification of the complex was necessary. The samples were complexed at a ratio of 1:10 (with Fab in excess) and left to incubate on ice for 30 minutes. Then loaded onto a Superdex 200 16/60 size exclusion column (GE Healthcare) and purified into 20 mM sodium cacodylate pH 6.5, 120 mM sodium chloride. To ensure high resolution chromatography was carried out at 0.5 ml per minute.

The complexes were screened (CSS I+II with 0.1 M Bis Tris pH 6.5 and Hampton I+II) at 5 mg/ml in 20 mM sodium cacodylate pH 6.5 at 20˚C. Protein desalting was carried out by repeated washing and concentrating of the protein with 20 mM sodium cacodylate pH 6.5 using Vivaspin 500 sample concentrators (10000 MWCO). Screening was in a 96 well sitting drop vapour diffusion format with drops of 100 nl complex plus 100 nl well solution pipetted by a Mosquito robot (TTP Labtech Ltd).

4.3. Results

4.3.1. Purification and characterisation of the HASPs

A two-step purification procedure was optimised for maximal yield and sample purity of the HASPs, the details of which can be found in Section 4.2.1. To enhance the purity of the HASPBs, an extra ion exchange was introduced after nickel affinity purification. Figure 4-2 highlights the purification schedule with HASPA as an example. Firstly, the nickel affinity purification step produced considerably pure eluted sample for this initial step, especially in the case of HASPA (Figure 4-2a and b). These fractions were pooled and concentrated for size exclusion purification. Adequate separation from remaining contaminants was achieved, observed by SDS-PAGE, to allow for sample characterisation (Figure 4-2c and d).

The homogeneity was then assessed by dynamic light scattering (DLS) and ESI-MS for all three HASPs. DLS reported a polydispersity of 19% indicative of a highly monodisperse sample, a feature of a readily crystallisable construct (Figure 4-2e). The molecular weight (MW) obtained by ESI-MS was 9445.1 Daltons (Figure 4-2f), in line with the expected N-terminal methionine cleaved molecular weight of 9445.6 Daltons obtained from ProtParam (Gasteiger et al. 2005). Moreover, no major adducts were noted in this preparation. These positive characterisation results meant that crystallisation trials were started and continued with no modification of the purification schedule. Figure 4-3 shows the SDS-PAGE analysis of the HASPBs post size exclusion purification to emphasise the final sample purity achieved for each protein.

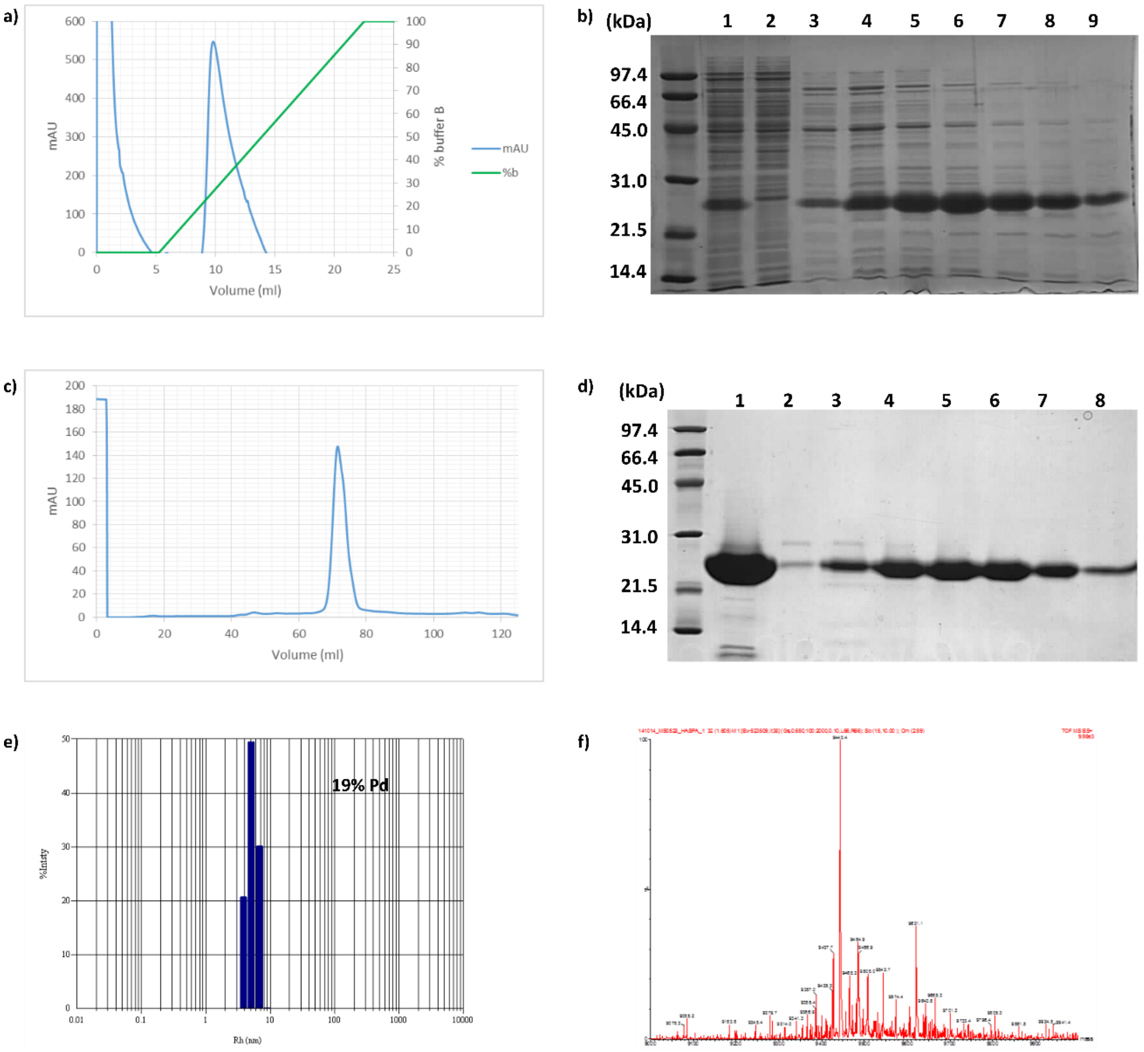


Figure 4-2: Chromatography traces and SDS PAGE analysis of the two step HASPA purification. a) 1ml HisTrap FF crude column chromatography trace. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 500 mM imidazole) is in green. b) 12% SDS PAGE analysis of load and resultant fractions from (a). The gel shows cell lysate load (1), column flow through (2), and resultant fractions (3 to 9). c) Chromatography trace of Superdex 75 16/60 purification of HASPA. d) 12% SDS PAGE analysis of (c). The gel shows load (1) and resultant fractions (2 to 8). The HASPA construct is 9.5 kDa. e) DLS regularisation histogram of an HASPB1, note percentage intensity is plotted against hydrodynamic radius (RH). The polydispersity (Pd) recorded was 19%. f) Electrospray ionisation mass spectrometry (ESI-MS) result of HASPA reporting the expected N-terminal methionine cleaved molecular weight of 9445.1 Da.

HASPA (9.5 kDa), HASPB2 (15.2 kDa) and HASPB1 (43.4 kDa) all run much higher than expected for their respective molecular weights (Figure 4-2 and Figure 4-3). The HASPs migrate aberrantly on SDS-PAGE because of their lack of hydrophobic residues, a feature that has been observed previously (McKean et al. 1997). The molecular weight of each protein was confirmed by ESI-MS to ensure that this migration pattern was not due to unexpected post-translational modifications.

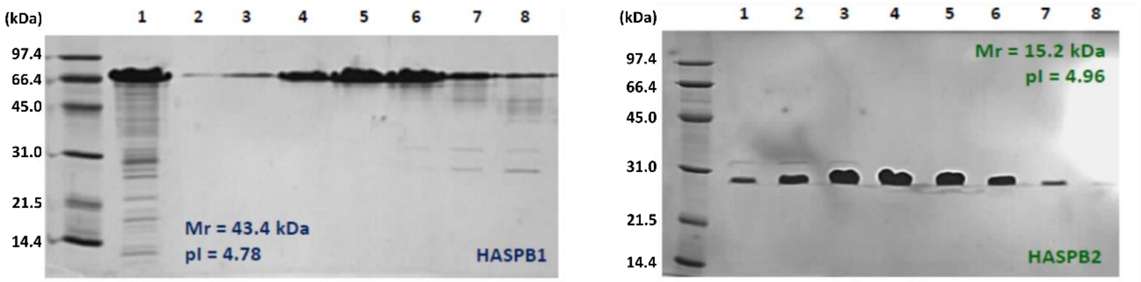


Figure 4-3: 12% SDS-PAGE of the *L. donovani* HASPBs size exclusion fractions, highlighting the homogeneity of the samples produced. The gels show column load (1) and resultant fractions (2 to 8).

4.3.2. Reproducible concentration determination of the HASPs

Accurate and reproducible concentration determination of a protein sample in solution is imperative to facilitate downstream biochemical and biophysical characterisation. Standard concentration determination utilises aromatic amino acid absorbance at 280 nm, resulting in a protein-specific extinction co efficient. Proteins lacking aromatic residues usually contain an acceptable number of hydrophobic residues to be reactive with the Bradford reagent, allowing the concentration to be determined with this method. However, the HASPs only contain one aromatic residue (tyrosine), contributing to a low extinction co-efficient that would produce inaccurate results at 280 nm. They are also highly hydrophilic and had been shown to lack reactivity with Bradford reagent (personal communication with Dr. James Brannigan). Therefore, a reliable and reproducible concentration method was required for atypical sample concentration determination.

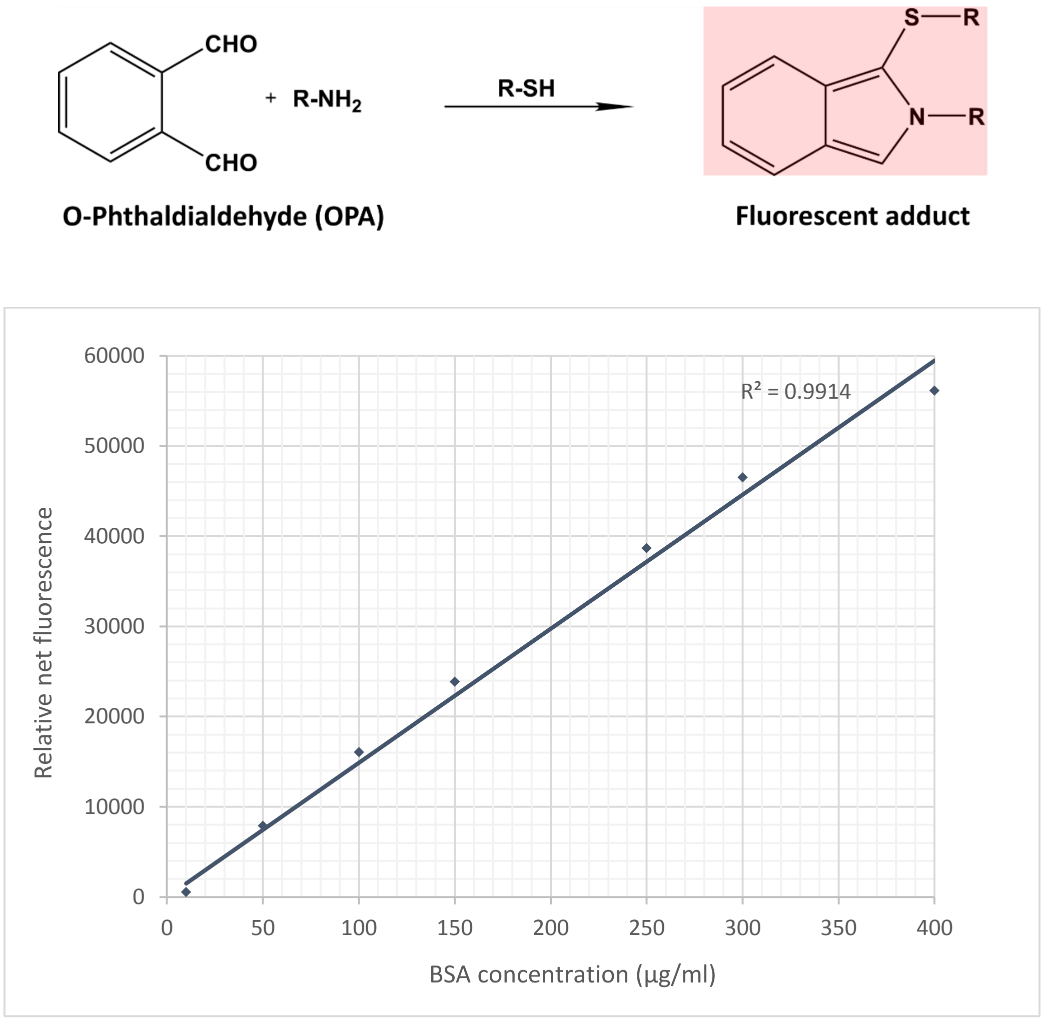


Figure 4-4: OPA concentration determination assay. Top) Reaction of OPA and primary amino groups. O-Phthaldialdehyde, in the presence of reduced sulfhydryl groups, reacts with the primary amino groups to produce a fluorescent adduct. Bottom) An example calibration curve produced with OPA reagent using BSA standards of known concentration. This was used to determine the concentration of the HASPs samples based to the fluorescence readings taken.

OPA is a primary amine-reactive fluorescent detection reagent that can be used as for protein concentration determination or as a post-column detection reagent for amino acid analysis with high-performance liquid chromatography (HPLC). In the presence of 2-mercaptoethanol, OPA reacts with primary amines to form a highly fluorescent product (Benson and Hare. 1975). This reactions yields linear results over a wide range of concentrations (10 to 500 µg/ml) and can be used to generate a standard calibration curve (Figure 4-4). Notably, OPA assay accuracy was consistently determined to be ≥ 85% within 25 to 400 µg/ml (Zhu et al. 2009). A fluorescence based assay using OPA was optimised in a 96 well format, with BSA as the protein standard of known concentration. Buffers with low pH or high molarity can cause decreased fluorescence. Moreover, primary amine containing buffers such as Tris or glycine will react with OPA so should be avoided.

The choice of an appropriate standard is key and can limit the effectiveness of this technique. BSA, obtained as a prepared stock solution at 2 mg/ml from Pierce, has a lysine content of 9.9%, while HASPA, HASPB1 and HASPB2 have lysine contents of 6.9%, 12.7% and 9.2%, respectively. BSA was an appropriate standard for HASPB2 due to their similar lysine contents but led to overestimation of HASPA concentration and underestimation of HASPB1 concentration. The Nt-T4L construct (see Chapter 5) has a lysine content of 8.2% and was used as a more appropriate standard for HASPA concentration determination. Although this technique has limitations it was vital for downstream biochemical and biophysical analysis of the HASPs.

4.3.3. Oligomeric state of the HASPs in solution

Multi-angle laser light scattering (MALLS) is a technique for determining the absolute molar mass and the average molecular weight of particles in solution by detecting how they scatter light (Oliva et al. 2001). Therefore, measuring the scattering and concentration of a sample allows the molecular weight to be calculated. It is coupled with size exclusion chromatography (SEC) to define separate species within a sample. The main advantage of SEC-MALLS over analytical SEC is that the experimentally determined molecular weight is independent of elution order (Ye. 2006). This prevents elution order being used to attribute oligomeric states that are not present. This misleading elution order can be due to the protein interacting with the matrix of the column or a large hydrodynamic radius affecting migration. This technique was applied to the HASPs, HASPA and HASPB1, to analyse their oligomeric state in solution.

It was vital to determine the oligomeric state and propensity to aggregate of the HASPs, particularly to determine the presence of any heterogeneous mixtures that would impede crystallisation. HASPA and HASPB1 were analysed to compare and contrast the role of the repeats on any oligomerisation or aggregation. Both proteins produced well defined single peaks when analysed, indicating that the samples are in one homogenous state in solution (Figure 4-5). Moreover, the reported molecular weight values obtained are close to the expected monomeric values (Table 4-1). Notably, a well resolved peak will report a molecular weight with a ± 10% error, based on empirical analysis of the system by Dr. Andrew Leech. This accounts for the slight difference between the reported and expected values for HASPA. However, the smaller than expected molecular weight value obtained for HASPB1 is most likely due to some sample degradation. Moreover, the HASPB1 is slightly asymmetric compared to the HASPA peak, which would contribute to the slight variance in the reported molecular weight.

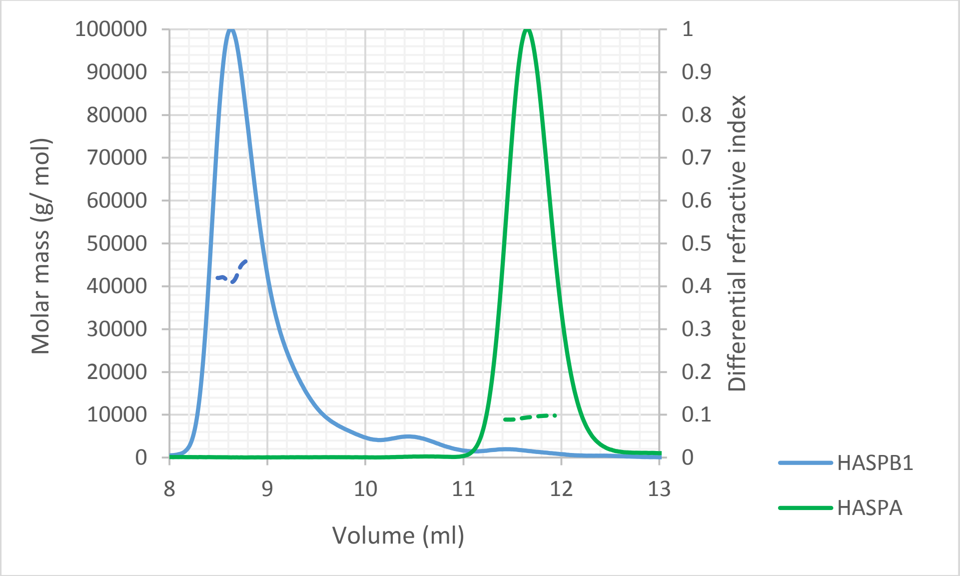


Figure 4-5: SEC-MALLS molecular mass vs volume plot for HASPA and HASPB1. The solid lines report the differential refractive index of HASPA (green) and HASPB1 (blue) against elution volume. The molecular weights were calculated from the dashed peaks (same colours as differential refractive index trace), which correspond to the area of the peak used for this calculation. These plots show the homogeneity of the monomeric forms of both proteins in solution.

Sample	Elution volume (ml)	Expected MW (kDa)	Reported MW (kDa)
HASPA	11.7	9.5	9.4
HASPB1	8.6	43.4	37.1

Table 4-1: SEC-MALLS derived parameters for HASPA and HASPB1.

4.3.4. Crystallisation of the HASPs

Extensive crystallisation screening of all three full length HASPs was carried out before IDP classification and Figure 4-6 shows some typical drops. It would be unnecessary to report an exhaustive list of conditions tested as no crystals were produced. However, the multifaceted approach to crystallise the HASPs encompassed a vast array of non-typical reagents that would be interesting to mention. The HASPs favoured low pH buffers, in particular sodium acetate pH 4.5 was a recurrent theme in initial screening and later work. Cationic detergents, e.g. choline chloride, dodeyltrimethylammonium chloride/ bromide and Girard's reagent T, were tested as additives. It was thought that these positively charged detergents would mimic the membrane environment and stabilise the net negative charge of the HASPs. A further membrane mimetic POC (10 mM 1,2 didodecanoyl-sn-glycero-3-phosphatidylcholine in 20 mM taurodeoxycholic acid) was added directly to the protein before screening. Neither of these approaches produced crystals but did yield interesting semi crystalline material (Figure 4-6). HASPA primarily produced this semi crystalline material, highlighting that the protein was capable of 2D order but this was impossible to tune into 3D order required for proper lattice formation. It was clear that a new strategy was required to coerce the HASPs to crystallise.

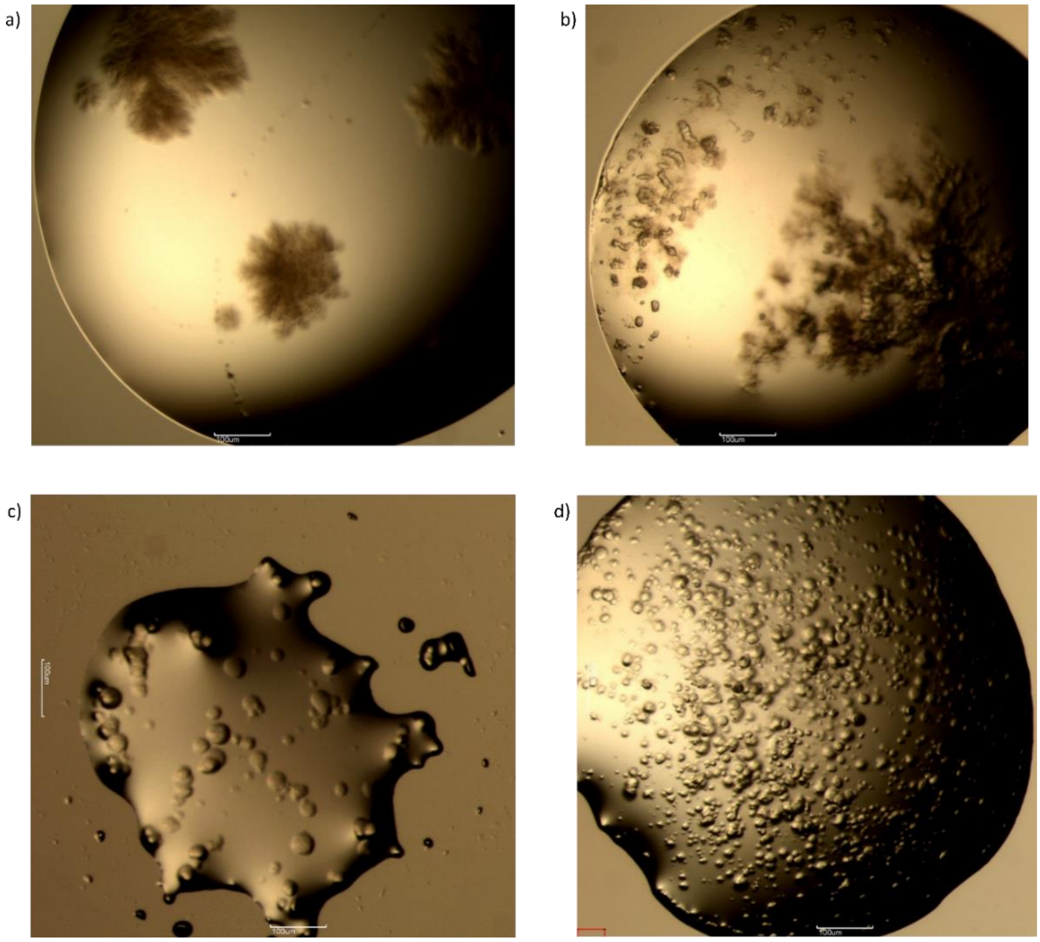


Figure 4-6: Examples of HASP crystallisation screening. All drops relate to HASPA. a) 5 mM dodeyltrimethylammonium bromide, 20% PEG 3350, 0.1 M sodium acetate pH 4.5. b) as (a) but 10 mM dodeyltrimethylammonium bromide. c) 30% ethanol, 0.1 M sodium acetate pH 4.5. d) 50 mM choline chloride, 20% PEG 3350, 0.1M sodium acetate pH 4.5.

4.3.5. Purification and characterisation of the HASPB1 fragments

Truncations of HASPB1 were designed to cover the N- and C-terminal domains and repeats. The amino acid composition of these constructs is shown in Figure 4-7. It should be noted that the cloning procedure will not be discussed here because it was carried out by James Chamberlin as part of his undergraduate studies and was submitted for examination in his thesis. It is important to note that the expression of the Ct and CtR constructs was problematic. No soluble protein was obtained for either construct, as the proteins were degraded while they were expressed. Therefore, all the work outlined here relates to the Nt and NtR constructs.

HASPB1 sequence

(M)GAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTGSAMNDSAPKE **50**
DGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKND **100**
GDGPKEDGHTQKNDGDAPKEDGRTQKNDGDGPKEDGHTQKNDGDAPKEDG **150**
RTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGRTQKNDGG **200**
GPKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDAPKEDGRT **250**
QKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDAPKEDGRTQKNDGDGP **300**
KEDGHTQKNDGDGPKEDGRTQKNDGDGPKEDGHTQKNDGDGPKEDGRTQK **350**
NDGDA**PKEGENLQQNDGDAQEKNEDGHNVGDGANGNEDGNDDQPKEHAAG** **400**
NLEHHHHHH

Nt construct

GPAM**MGAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTGSAMNDSA**

NtR construct

GPAM**MGAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTGSAMNDSAPKEDGHTQKNDGDG**

Ct construct

GPAM**QEKNEDGHNVGDGANGNEDGNDDQPKEHAAGN**

CtR construct

GPAM**PKEGENLQQNDGDAQEKNEDGHNVGDGANGNEDGNDDQPKEHAAGN**

Figure 4-7: Amino acid composition of the HASPB1 fragment constructs. The Nt construct encompasses the N-terminal domain of HASPB1 (blue). The NtR construct is the same as the Nt but with one repeat from HASPB1 (red). The Ct construct comprises the C-terminal domain of HASPB1 (green). The CtR construct is the same with one repeat from HASPB1 (purple).

The two step HASP purification procedure, outlined in Section 4.3.1., was applied to the HASPB1 fragments. Sufficient yield and sample purity of both Nt and NtR was achieved with this protocol, meaning that it was used in all subsequent purifications of these constructs (Figure 4-8). The nickel affinity purification step yielded moderately pure protein (Figure 4-8a to c). The fractions were pooled and concentrated for size exclusion purification. Adequate separation from remaining containments was achieved, observed by SDS-PAGE, to allow for sample characterisation (Figure 4-8 d to f). It is notable that the Nt and NtR fragments also migrated aberrantly on SDS-PAGE like the full length HASPs. ESI-MS was used to confirm the molecular weights of the constructs, which were in line with the expected molecular weights detailed in Table 4-2. However, for brevity this data is not included here. The purified Nt and NtR constructs were taken for biochemical characterisation.

Construct	Description	Molecular weight (Da)	pI
Nt	N-terminal domain of HASPB1	5222.7	8.4
NtR	Nt construct plus one repeat	6702.2	6.4
Ct	N-terminal domain of HASPB1	3689.6	4.2
CtR	Ct construct plus one repeat	5186.1	4.1

Table 4-2: Descriptions and biochemical features of the HASPB1 fragments. The molecular weight and pI values were obtained from ProParam (ExPAsy).

It is interesting to note that the NtR construct degraded rapidly when purified. The bands present underneath the NtR band in Figure 4-8c could be removed by size exclusion purification but were present after storage at 4˚C for 24 hours. It was hypothesised that the exposure of the repeat made the NtR construct sensitive to degradation. Treatment with the serine protease inhibitor PMSF (phenylmethanesulfonyl fluoride) did not prevent degradation. This would lead to an unpredictably heterogeneous sample that would be unlikely to crystallise. Therefore, all crystallisation screening was carried out with the Nt construct with removed hexa-histidine tag but as no crystals were produced so this work will not be discussed here.

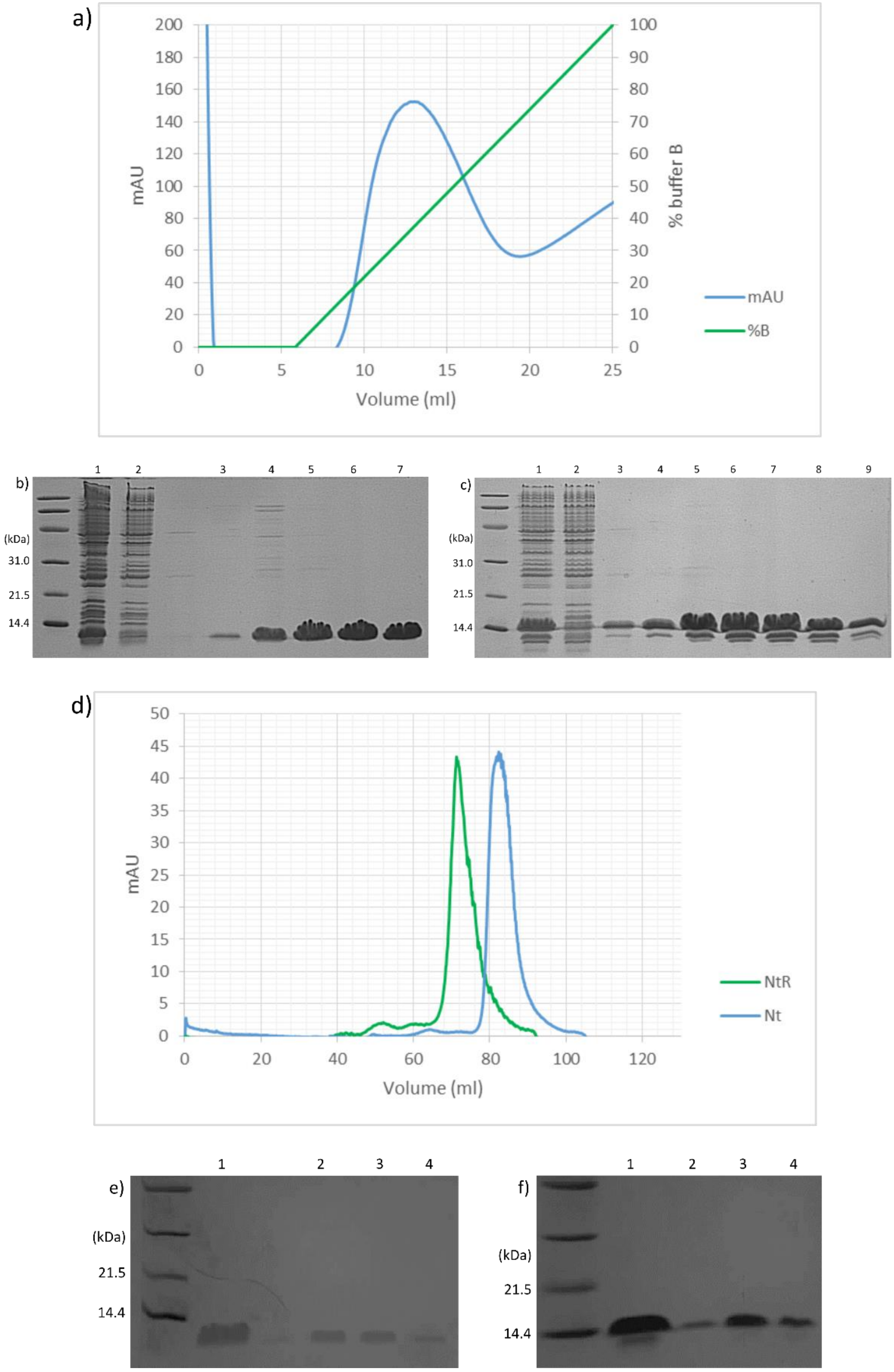


Figure 4-8: Chromatography traces and SDS PAGE analysis of the two step HASPB1 fragment purification. a) 1ml HisTrap FF crude column chromatography trace of the Nt construct. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 500 mM imidazole) is in green. b) 12% SDS PAGE analysis of load and resultant fractions from (a). The gel shows cell lysate load (1), column flow through (2), and resultant fractions (3 to 7). c) same as (b) but for NtR construct. d) Chromatography trace of Superdex 75 16/60 purification of the Nt (blue) and NtR (green) fragments. e) 12% SDS PAGE analysis of Nt size exclusion purification. The gel shows load (1) and resultant fractions (2 to 4). f) same as (e) but for NtR construct.

4.3.6. SEC-MALLS of Nt and NtR fragments

In order to investigate the oligomeric state of the HASPB1 Nt and Ntr fragments *in vitro* SEC-MALLS was carried out. The strength of the light scattering signal produced by a sample is proportional to its molecular weight and concentration (Ye. 2006). Therefore, low molecular weight samples, where the MW is 5000 Da or less, require high concentrations to produce acceptable data. This meant all the HASPB1 fragments were analysed at 3 mg/ml to produce adequate light scattering. As with the full length HASP samples no oligomerisation was observed. The Nt construct generated a symmetrical peak in both untagged and tagged forms, indicating high homogeneity in the sample (Figure 4-9). The reported molecular weight value for the HASPB1 Nt non his-tagged sample is 5070 Da, which relates well with the expected value of 5091.5 Da and is within the 10% error of the system (Section 4.3.3.). Moreover, this slight variance in molecular weight can be attributed to the size of the constructs, because they are the at lower detection limit of the apparatus. This correlation between expected and reported molecular weight is also observed for the Nt his-tagged sample.

However, the NtR construct, in both tagged and untagged forms, produces a broader slightly asymmetric peak, indicative of sample heterogeneity. Considering the sample degradation issues outlined in Section 4.3.5., this is unsurprising. This feature is particularly prominent in the NtR his-tagged sample, where the peak broadens towards later elution time and does not return sharply to the baseline (Figure 4-9 blue peak). The reported molecular weight of this sample is higher than the expected value, which could be due to impurities affecting the estimate (Table 4-3). This could also be attributed to the asymmetric peak producing an inaccurate molecular weight as the peak composition differs from start to end. The sharpness of the peak was improved after his-tag cleavage, highlighting the importance of this step in sample preparation. The molecular weight reported for NtR non-tagged sample was 6610 Da, which was within the acceptable 10% error of the expected 6571 Da.

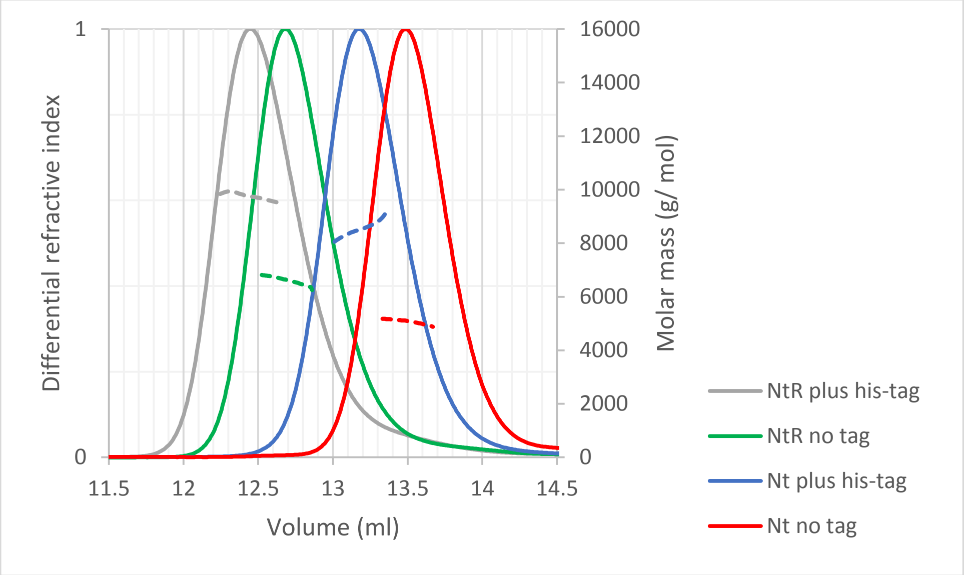


Figure 4-9: SEC-MALLS molecular mass vs elution volume plot generated for the Nt and NtR constructs. Note all the peaks are scaled to the same height to aid comparison. The solid lines report the differential refractive index of Nt plus his-tag (blue), NtR plus his-tag (grey), Nt no his-tag (red) and NtR no his-tag (green) against elution volume. The molecular weights were calculated from the dashed peaks (same colours as differential refractive index trace), which correspond to the area of the peak used for this calculation. These plots show the homogeneity of the monomeric forms of both proteins in solution.

Sample	Elution time (min)	Expected MW (Da)	Reported MW (Da)
NtR plus his-tag	24.9	8717.3	9410
Nt plus his-tag	26.4	7237.8	7773
NtR no tag	25.4	6702.2	6610
Nt no tag	27.0	5091.5	5070

Table 4-3: SEC-MALLS derived parameters of the HASPB1 fragments compared with expected molecular weights.

4.3.7. Circular dichroism of HASPB1 Nt fragment

Circular Dichroism (CD) Spectrophotometry is a technique used to quantify the interaction of a molecule of interest with polarised light. The chiral arrangement of peptides bonds in specific protein secondary structures (e.g. α helices and β sheets) generate characteristic spectra at defined wavelengths. This allows the user to potentially assign secondary structure characteristics of an unknown protein sample based on the spectra obtained.

Data were collected between 195 and 260 nm. The CD signal was considered valid between these points, as the HT voltage was below 600 but saturated at a wavelength lower than 195 nm (Figure 4-10c). It is clear that the Nt construct spectrum is that of a random coil, when compared with a standard set of spectra (Figure 4-10). A typical random coil spectral feature is a negative peak at 195 nm, which is seen here. However, the positive peak around 220 nm usually observed for random coil samples is not observed. There is no positive peak observed in the CD spectra of HASPA and HASPB1 (discussed in Section 2.3.2.), yet these spectra are defined as random coil. This classification means that the Nt fragment lacks discernible secondary structure. Based on this result it was decided that crystallisation trials should be halted with this construct. The result galvanises the narrative of the unstructured nature of the HASPs, especially when taken with the CD spectra of HASPA and HASPB1 presented in Section 2.3.2.

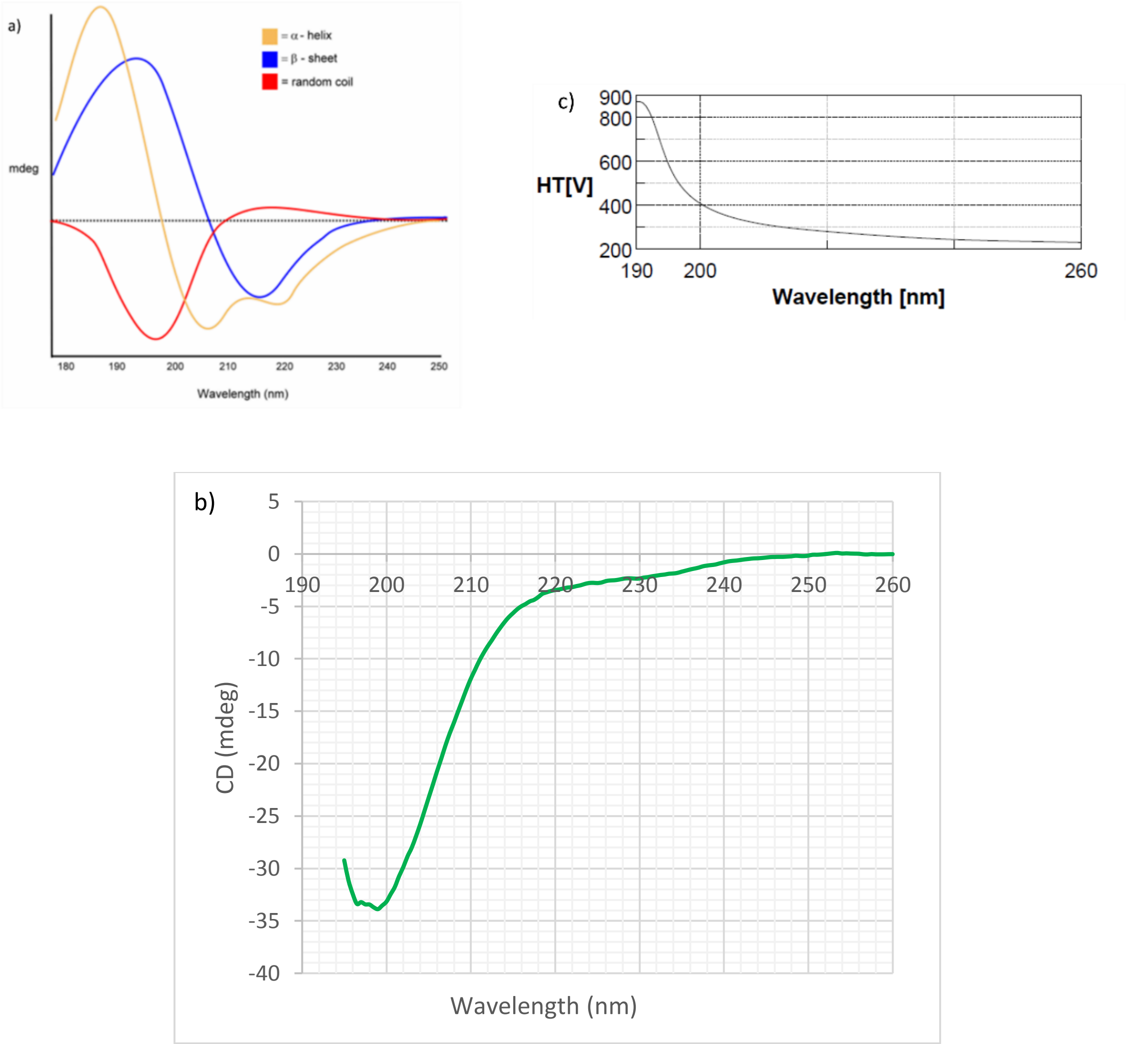


Figure 4-10: CD spectrum of HASPB1 Nt fragment. a) Example CD spectra of alpha-helical (yellow), beta-sheet (blue) and random coil (red). This image is taken from Dr. Ramy S. Farid. 2006. Circular Dichroism (CD) Spectroscopy. Available at: http://www.proteinchemist.com/cd/cdspec.html. [Accessed 12 August 15]. b) CD spectra for the HASPB1 Nt fragment exhibiting the random coil spectral feature. c) HT voltage trace.

4.3.8. Preparation of Fab fragments

In order to stabilise the flexible structure of HASPB1 for crystallisation the idea of complexing the protein with specific antibody was explored. HASPB1 specific rabbit polyclonal antibodies were prepared by Mike Hodgkinson. Preliminary work showed that the HASPB1: antibody complex was too large to enter a native electrophoresis gel, not a good starting point for crystallisation. It was decided that Fab fragments should be prepared from the antibody stock and used for complex formation with HASPB1 as this has shown to be a successful approach for crystallising membrane proteins (Hunte and Michel. 2002). Moreover, removal of the Fc (fragment crystallisable) region in this polyclonal sample would also remove some of the heterogeneity that could impede crystallisation.

Papain, a non-specific thiol-endopeptidase, digestion of Immunoglobulin G (IgG) generates two antigen binding Fab fragments and one Fc fragment. Papain (Thermo Scientific) was obtained immobilised on an agarose resin, enabling greater control of the digestion process and easier removal of the enzyme. Digestion was firstly optimised on a small scale with the immobilised papain. It was found that a ratio of 1:10 (enzyme: substrate) was the most effective protocol for Fab production; note this is also the recommended ratio for effective rabbit IgG digestion (Coulter and Harris. 1983). The experiment was also carried out over a time course (Figure 4-11) with optimal digestion seen after 4 hours at 37°C with shaking at 750 rpm.

Interpretation of the resultant SDS-PAGE gel from IgG digestion can be more complex than initially assumed. IgG and antibody fragments have different migration patterns depending on the pre-treatment of the sample because of the disulphide bonds. Under non-reducing conditions the Fab and Fc fragments migrate to a molecular weight of 45 to 50 kDa. Fab fragments migrate to around 25 kDa and Fc fragments to around 30 kDa with reducing SDS-PAGE. Undigested IgG also has different migration patterns under reducing (heavy chain at 50 kDa and light chain around 25 to 30 kDa) and non-reducing conditions (≥ 100 kDa). Figure 4-11 has been annotated to clarify the bands for future reference.

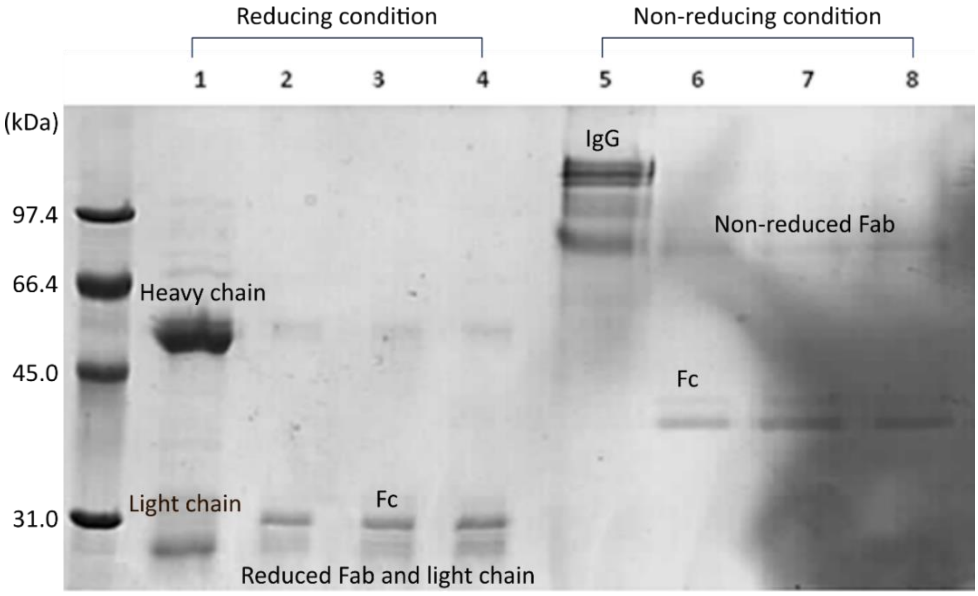


Figure 4-11: Papain digestion of IgG. 10% SDS-PAGE gel showing a time course of 1:10 enzyme to substrate digestion. Lanes 1-4 are under reducing and lanes 5-8 non-reducing conditions. The gels shows undigested IgG (1 and 5), two hours of digestion (2 and 6), three hours of digestion (3 and 7) and four hours of digestion (4 and 8).

Separation of the digested Fab fragments from the undigested IgG and Fc fragments was the vital next step in the preparation procedure. Protein A, derived from *Staphylococcus aureus,* is used for IgG affinity purification because it binds the Fc region of IgG with high specificity and strength (Goding et al. 1978). It can be obtained as a 1 ml or 5 ml HiTrap Protein A HP column (GE healthcare), where protein A is coupled to Sepharose so that the IgG binding regions are free. Notably, one molecule of coupled protein A can bind 5 molecules of IgG. The purification schedule produced the expected two distinct peaks in the resulting chromatograph (Figure 4-12c). SDS-PAGE analysis was used to determine the composition of the peaks and it was shown that the first peak contained purified Fab and the remaining peak undigested IgG and Fc (Figure 4-12). The Fab fractions could then be pooled and dialysed ready for complex formation with HASPB1.

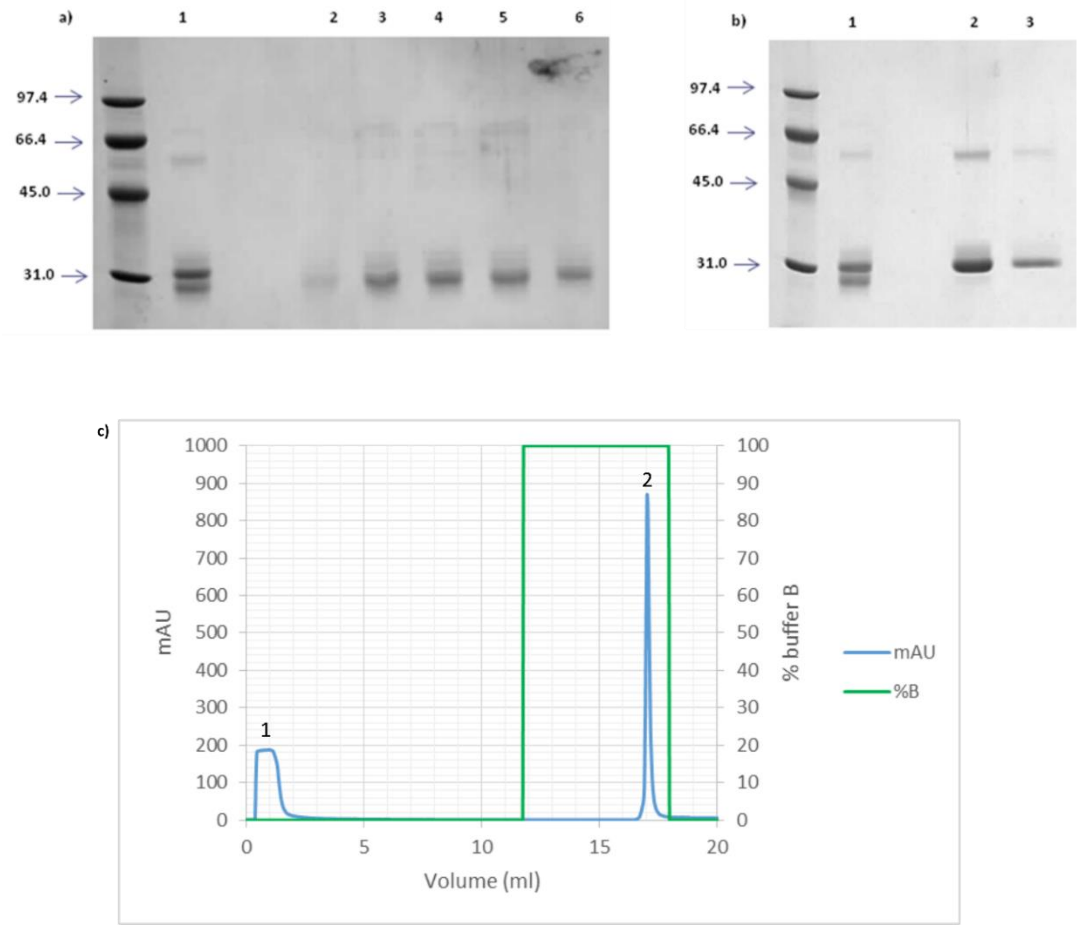


Figure 4-12: Protein A purification of digested antibody sample. SDS-PAGE analysis of Protein A purification, where all lanes are under reducing conditions. a) The gel shows undigested sample (1) and peak 1 fractions lanes (2 to 6). (b) The gel shows undigested sample (1) and peak 2 fractions (2 and 3). c) 1 ml HiTrap Protein A HP column (GE Healthcare) chromatography trace. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 0.1 M citric acid pH 3) is in green.

The samples were complexed at a ratio of 1:10 with Fab in excess but needed to be purified to remove any unbound Fab present. This was achieved by size exclusion purification (Superdex 200 16/60 column) and the resultant chromatograms showed two peaks (Figure 4-13a). It was hypothesised that these peaks could relate to two or more separate complexed species, due to the polyclonal nature of the antibody sample, or one complex with the second peak due to Fab or unbound HASPB1. SDS-PAGE analysis was carried out to determine peak composition. Figure 4-13 highlights that each peak does contain both proteins; however, in order to determine complex formation it was necessary to test the samples on a native gel. It is clear when analysing the native gel data (Figure 4-13c) that neither lane contains any unbound HASPB1. It was concluded that both peaks contained complexes with differing stoichiometries or Fab fragments with different molecular weights. Therefore, both peaks were pooled separately and concentrated for use in crystallisation attempts.

However, due to the extremely small amount of complex produced biochemical characterisation was not possible. SEC-MALS is used to analyse the molecular weight and stoichiometry of a complex but requires at least 200 µl at 2 mg/ml, consuming the entire sample. Native (non-denaturing) mass spectrometry was considered as it requires much less sample and reports the composition, topological arrangements, dynamics, and structural properties of protein complexes (Lorenzen and Duijn. 2001). However, it still requires at least 20 µl and it was decided that this was too sample expensive.

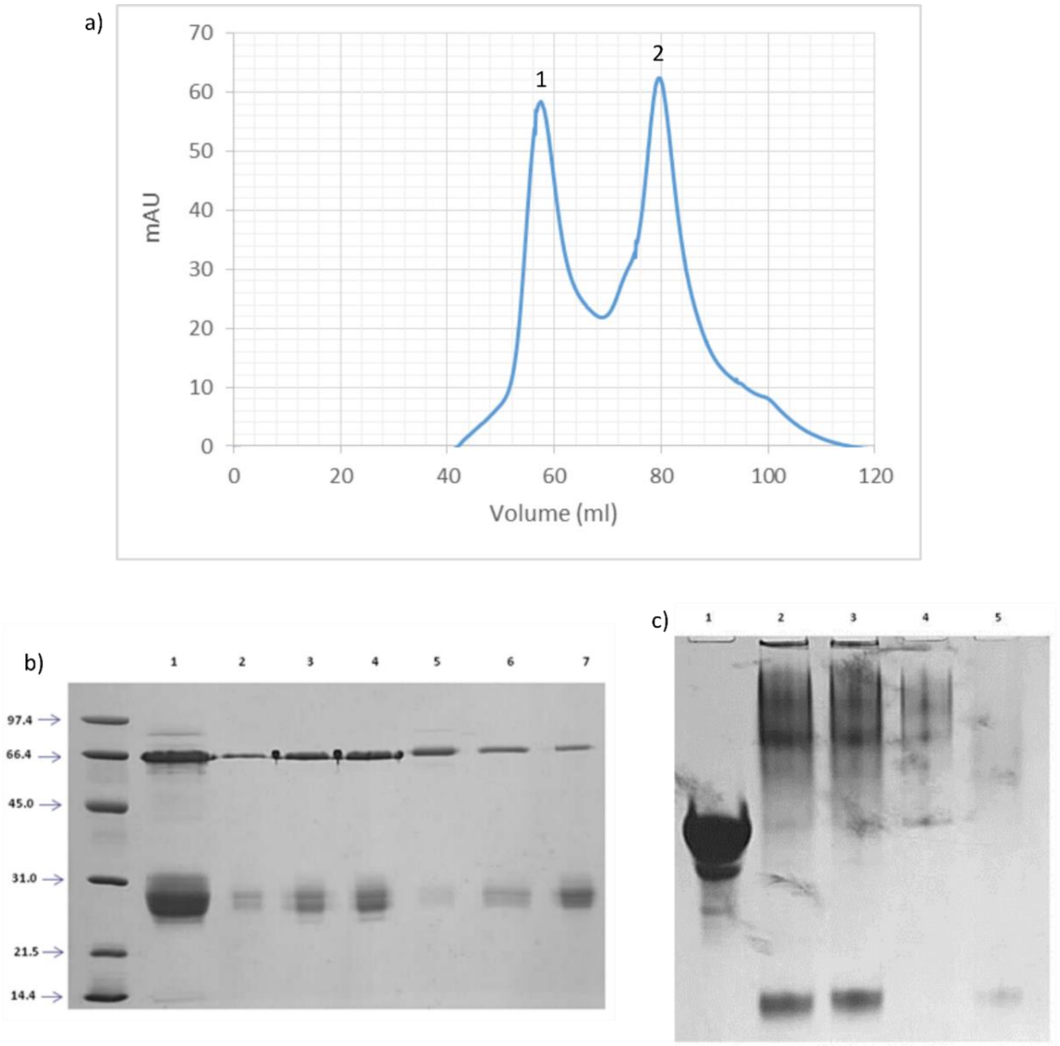


Figure 4-13: Size exclusion purification of the HASPB1: Fab complex. a) Superdex 200 16/60 chromatography trace for the purification of the HASPB1: Fab complex. The UV trace (280 nm) is shown in blue and the peaks taken for SDS-PAGE analysis are numbered. b) SDS-PAGE analysis of size exclusion peaks observed in (a). All lanes are under reducing conditions. The gels shows loaded sample (1), peak 1 fractions (2 to 4) and peak 2 fractions (5 to 7). c) Native interaction gel comparing peaks seen in purification. The gel shows purified HASPB1 (1), loaded complexed sample (2 to 3), peak 1 sample (4) and peak 2 sample (5).

4.3.9. Crystallisation attempts of the HASPB1: Fab complex

After Fab generation and purification the yield of HASPB1: Fab complex was extremely low, approximately 60 µl of the peak 1 complex at 5 mg/ml was obtained from 5 mg starting material. Therefore, two screens were setup; CSS I+II with 0.1 M Bis Tris pH 6.5 and Hampton I+II. Seemingly promising hits were obtained with the peak 1 complex from the Hampton screen then optimised (Figure 4-14). Further improvement of these conditions was unsuccessful and the remaining sample was used. This work was discontinued in favour of other lines of enquiry.

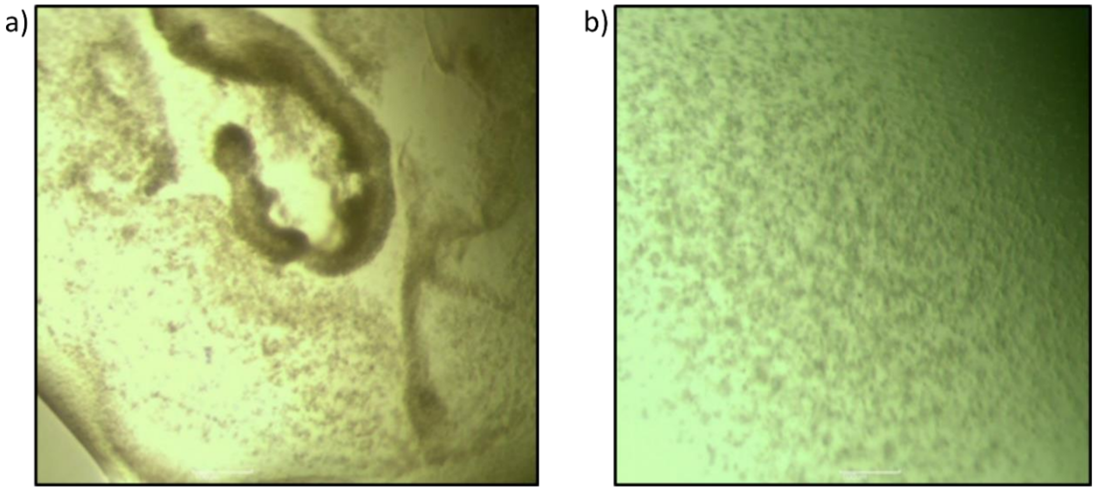


Figure 4-14: Example preciptate obtained from the Hampton I+II screen of the HASPB1: Fab complex. a) Condition D4; 30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M sodium acetate. b) Condition G2; 1.4 M sodium citrate, 0.1 M sodium Hepes pH 7.5

4.4. Discussion

Protein engineering, including the removal of post translational modifications, ligand or partner protein binding, surface entropy reduction, covalent and non-covalent crystallisation chaperones, truncations and flexible domain removal, has been repeatedly implemented to enhance the crystallisability of proteins (Derewenda. 2010). A multidimensional strategy was devised and implemented, presented here and in Chapter 5, to coerce HASP crystallisation but was ultimately unsuccessful. In light of intrinsic disorder classification of the HASPs it is unsurprising that crystallisation attempts did not yield any crystals. The conformational heterogeneity of all IDPs impedes crystal formation and is discussed in more detail in Section 5.4.1.

4.5. Summary and future work

A purification schedule was optimised for maximal yield and sample purity of the HASPs, confirmed by dynamic light scattering and electrospray mass spectrometry. Additional biochemical characterisation with SEC-MALLS showed that the HASPs were monomeric in solution. These favourable biochemical properties meant that crystallisation screening could begin. However, screening of all three full length HASPS showed them to be recalcitrant to crystallisation. It was clear that stabilisation of a single conformation was required to modulate HASP crystallisability. Challenging crystallisation targets often require a multifaceted approach to produce structural data. In this chapter, Fab fragment stabilisation, HASPB1 truncations and additive based stabilisation, were implemented to coerce HASP crystallisation. Unfortunately no crystals were produced during the course of this project. Once intrinsic disorder classification was established (see Chapter 2) it was clear it that the inherent conformational heterogeneity of the HASPs would always impede crystallisation. This classification was further bolstered by circular dichroism of the HASPB1 N-terminal domain construct that showed it to be disordered. Refocussing the project with IDP classification in mind enabled study of the dynamic nature of the HASPs (Chapter 2) and the NMT: HASP interaction (Chapter 3).

It would be beneficial to analyse the HASPB1: Fab complexes with NMR, to monitor any chemical shift or intensity changes indicative of protein-protein interactions. Comparative HSQC spectra of ^{15}N labelled HASPB1 with the two HASPB1: Fab complexes, where HASPB1 was ^{15}N labelled while Fab remained unlabelled, would allow identification of any conformational changes in HASPB1. Change in chemical environment, mediated by Fab binding, would be manifested as changes in peak position and/ or intensity (Bieri et al. 2011). Moreover, this technique would also inform on the scale of any conformational changes based on the region affected by Fab binding. Lastly, all the published structures of Fab stabilised structures are generated with Fab fragments from monoclonal antibodies. If this work was continued then it would be beneficial to work with monoclonal antibodies, reducing the heterogeneity of the system.

Chapter 5 - T4 Lysozyme as a crystallisation chaperone

5.1. Introduction

5.1.1. T4 lysozyme fusion as a crystallisation strategy

Integral membrane proteins are predicted to represent around 30% of known proteins and are involved in many critical biological processes (Wallin and Von Heijne. 1998). It is notable that 40% of all current drug targets are membrane proteins but structure based drug discovery is limited by the lack of available structures (Carpenter et al. 2008). Currently, the Protein Data Bank contains in excess of 100,000 protein structures, 350 of those are unique membrane proteins (Moraes et al. 2014). The problematic nature of membrane protein expression, purification and crystallisation clearly has limited progress but a multifaceted approach to improve their crystallisability is turning the tide. Detergent improvements for purification and crystallisation, application of lipidic cubic phase, stabilisation via ligand binding and covalent or non-covalent fusion to a crystallisation chaperone are strategies that have been successfully implemented (Lieberman et al. 2011). The interest of this chapter is the chimeric covalent fusion strategy, whereby the gene for a chaperone with an intrinsic high propensity to crystallise is inserted into the membrane protein construct. The readily crystallisable protein chosen should also be structurally rigid thus reducing local flexibility of the target protein while increasing the polar surface needed for crystal contacts. Removal of an area of high conformational heterogeneity in the target further improves chimera stability. However, careful construct design is key to success with this strategy but is of course limited when there is no starting structure. Published examples that implement this method successfully are the best starting point to a more comprehensive understanding of its potential and limits.

Heterotrimeric guanine nucleotide binding protein (G protein) coupled receptors (GPCRs) are an extensive and versatile family of transmembrane signalling proteins that are implicated in a wide spectrum of diseases (Pierce et al. 2002). The external stimuli recognised by family members include hormones, neurotransmitters, chemokines, calcium ions, odorants and light. Crystallisation of the GPCRs has been problematic because of their low natural abundance, instability and dynamic nature. The difficulty of GPCR crystallisation is apparent when you consider that rhodopsin is the only wild type receptor to have been crystallised (Palczewski et al. 2000). Point mutations, truncations and the application of fusion partners have been used to circumvent these issues. In particular, fusion with bacteriophage T4 lysozyme (T4L) has enabled the crystallisation of adenosine receptor A_{2a} (Jaakola et al. 2008; Xu et al. 2011), chemokine CXCR4 receptor (Wu et al. 2010), dopamine D3 receptor (Chien et al. 2010), histamine H1 receptor (Shimamura et al. 2011), sphingosine phosphate $S1P_1$ receptor (Hanson et al. 2012), muscarinic acetylcholine M2 (Haga et al. 2012) and M3 (Kruse et al. 2012) receptors, μ-opioid receptor (Manglik et al. 2012), δ-opioid receptor (Granier et al. 2012), κ-opioid receptor (Wu et al. 2012) and the β2 adrenergic receptor (β2AR) (Cherezov et al. 2007; Rosenbaum et al. 2007; Rasmussen et al. 2011; Zou et al. 2012).

The β2 adrenergic receptor (β2AR), involved in the regulation of cardiovascular and pulmonary function by the sympathetic nervous system through adrenalin and noradrenalin recognition, was the starting point for this T4 lysozyme fusion approach (Rosenbaum et al. 2007; Cherezov et al. 2007). It was determined that the highly dynamic and poorly structured third intracellular loop (ICL3), which links the cytoplasmic ends of transmembrane domains 5 and 6, was the main contributor to structural instability (Kobilka and Schertler. 2008). The conformational heterogeneity of ICL3, though vital for its functions of G protein activation and interaction, enabled unobstructed movement of the transmembrane domains. ICL3 was removed and replaced by T4 lysozyme, restricting domain movement and increasing the polar contacts required for crystal growth. This allowed diffraction quality crystals to be grown and the structure of this chimeric protein to be solved (see Figure 5-1).

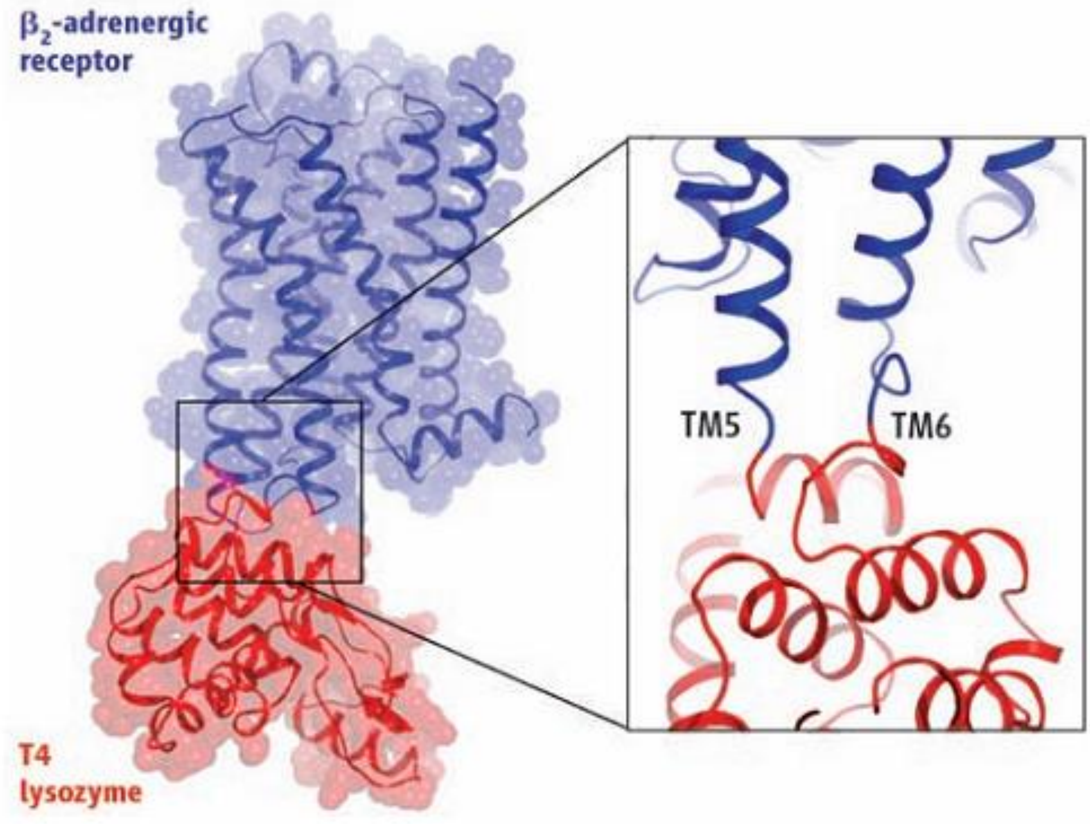


Figure 5-1: Structure of the β2 adrenergic receptor (β2AR): T4 lysozyme fusion protein (taken from Rosenbaum et al. 2007). Removal of the unstructured third intracellular loop (ICL3) and replacement with T4 lysozyme at the cytoplasmic face of transmembrane domains 5 and 6 enabled crystallisation.

The HASPs are not integral membrane proteins but via the N-terminal dual acylation motif they are membrane associated. Moreover, the seemingly promising crystallisation work done in Chapter 4 focused on tailoring the surrounding chemical environment to closely mimic the plasma membrane, a strategy of membrane protein crystallisation. Implementation of a chaperone as a rigid scaffold for crystallisation that masks areas of conformational heterogeneity while promoting lattice formation are attractive qualities lacking in the HASPs. This combination of factors merits the chimera approach as a potential reinvigoration of crystallisation attempts.

5.1.2. Construct design of HASP chimeras

Construct design began with an appraisal of various published examples of effective chaperone proteins (see Table 5-1). The plethora of potential fusions meant this analysis was limited to chimeras with associated structural data to filter out poorly crystallisable chaperones. T4 lysozyme was chosen as the crystallisation scaffold due to the high success rate of this approach, highlighted by the extensive list of published examples (see Section 5.1.1 and Table 5-1). Moreover, T4 lysozyme fusion is not limited in scope to ICL3 replacement but N-terminal fusion enabled crystallisation of a more functionally active form of the β2 adrenergic receptor (Zhou et al. 2012). This functional plasticity adds to the apparent efficacy of this chaperone.

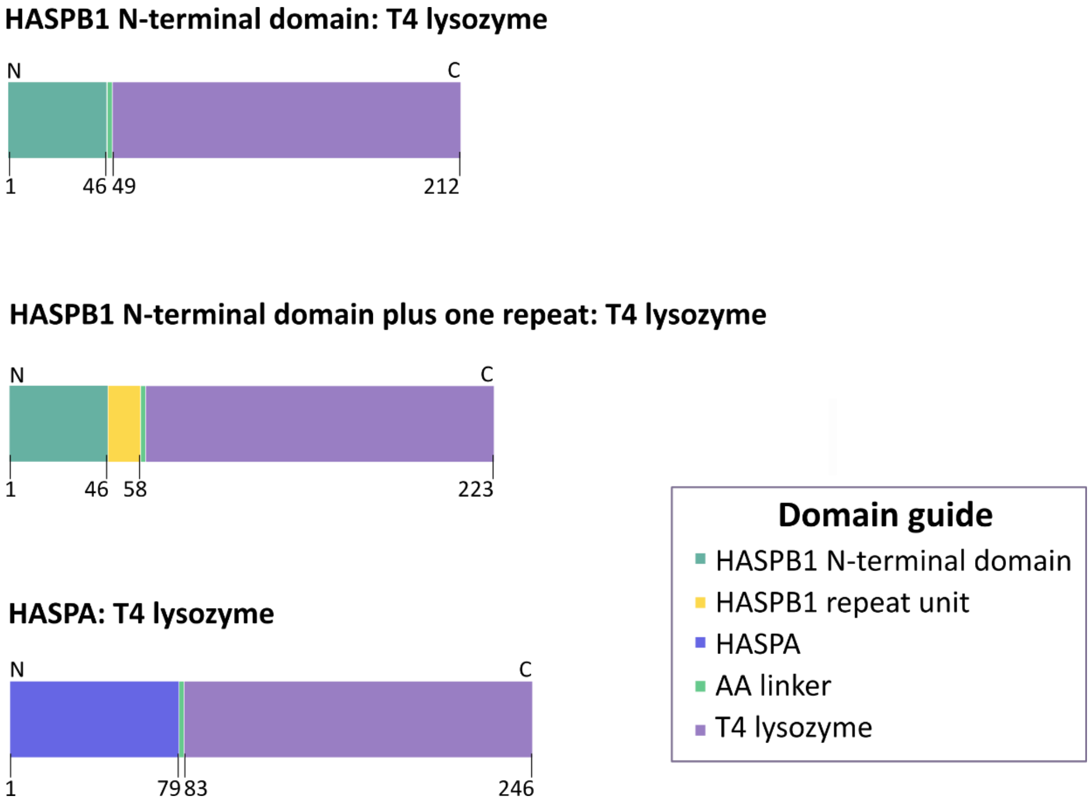


Figure 5-2: Schematic representation of the domain composition of HASP: T4 lysozyme fusion proteins with domain boundaries and residue numbers.

GCPR chimeric constructs sample various domain fusion sites and truncations en route to the final crystallisable chimera (Chun et al. 2012). Consideration of this statement in relation to the HASP sequences led to the design of three constructs that encompass their distinct sequence units (see Section 1.3.). Figure 5-2 shows a schematic representation of the planned constructs; 1) N-terminal domain of HASPB1, 2) N-terminal domain of HASPB1 plus one repeat unit and 3) HASPA (HASP N- and C-terminal domains) all C-terminally fused to T4 lysozyme. To circumvent the inherent flexibility of a two domain fusion protein, which would be highly detrimental to crystallisation attempts, a short linker is used in the literature to maintain structural rigidity throughout the chimera (see Table 5-1). Informed by this literature review I decided to opt for a minimal linker of two alanine residues between the chimeric units.

Chimera	Linker	PDB code	Reference
T4 lysozyme: N-terminus of β2 adrenergic receptor	AA	4GBR	Zou et al 2012
MBP: gp21	AAA	1MG1	Kobe et al 1999
MBP: SarR	AAAEF	1P4X	Liu et al 2001
MBP: MATa1	AAAAA	1MH3 1MH4	Ke and Wolberger 2003
T4 lysozyme: Human histamine H1 receptor	No linker	3RZE	Shimamura et al 2011
T4 lysozyme: human CXCR4 chemokine receptor	GS at both N- and C- terminal ends of T4	3ODU	Wu et al 2010
T4 lysozyme: human β2 adrenergic receptor	No linker at either end	2RH1	Cherezov et al 2007 + Rosenbaum et al 2007
T4 lysozyme: Adenosine receptor A_{2a}	No linker at either end	3QAK	Xu et al 2011
T4 lysozyme: Adenosine receptor A_{2a}	No linker at either end	3EML	Jaakola et al 2008
T4 lysozyme: human dopamine D3 receptor	No linker at either end	3PBL	Chien et al 2010
T4 lysozyme: human β2 adrenergic receptor: Nanobody	AA	3SN6	Rasmussen et al 2011
T4 lysozyme: *E. coli* lactose permease	S (short linker) or SGGSG/GSGGS (longer linker)	No structure	Engel et al 2002
MBP: Islet Amyloid Polypeptide (IAPP or amylin) full length and residues 1 to 22	AAA	3G7V 3G7W	Wiltzius et al 2009
MBP: ZP-N domain of ZP3 (3 crystal forms)	AAA	3D4C 3D4G 3EF7	Monné et al 2008
T4 lysozyme: sphingosine phosphate $S1P_1$ receptor	No linker	3V2W 3V3Y	Hanson et al 2012
T4 lysozyme: human M2 muscarinic acetylcholine receptor	No linker	3UON	Haga et al 2012
T4 lysozyme: rat M3 muscarinic acetylcholine receptor	No linker	4DAJ	Kruse et al 2012
T4 lysozyme: μ-opioid receptor	No linker	4DKL	Manglik et al 2012
T4 lysozyme: δ-opioid receptor	No linker	4EJ4	Granier et al 2012
T4 lysozyme: κ-opioid receptor	No linker	4DJH	Wu et al 2012

Table 5-1: Summary of published crystallisation chaperones in relation to linker composition and length.

5.2. Experimental

5.2.1. SLIC cloning of fusion constructs

SLIC, or sequence and ligase independent cloning does not utilise restriction enzymes or ligases (Li and Elledge. 2007). Primers for the amplification of the destination vector and desired insert are designed to have overlapping homology of around 15 base pairs. Separate treatment of the DNA fragments with T4 polymerase in the absence of dNTPs facilitates 3' exonuclease activity and generates single stranded overhangs (Figure 5-3). When sufficient complementary single stranded 5' overhangs are exposed, approximately 15 base pairs, the reaction can be halted by addition of dCTP. This reverts the activity of T4 polymerase but the limited nucleotide provision maintains the overhangs. The destination vector and PCR product with compatible single strand overlapping ends are mixed together and anneal. Gaps are subsequently repaired upon transformation.

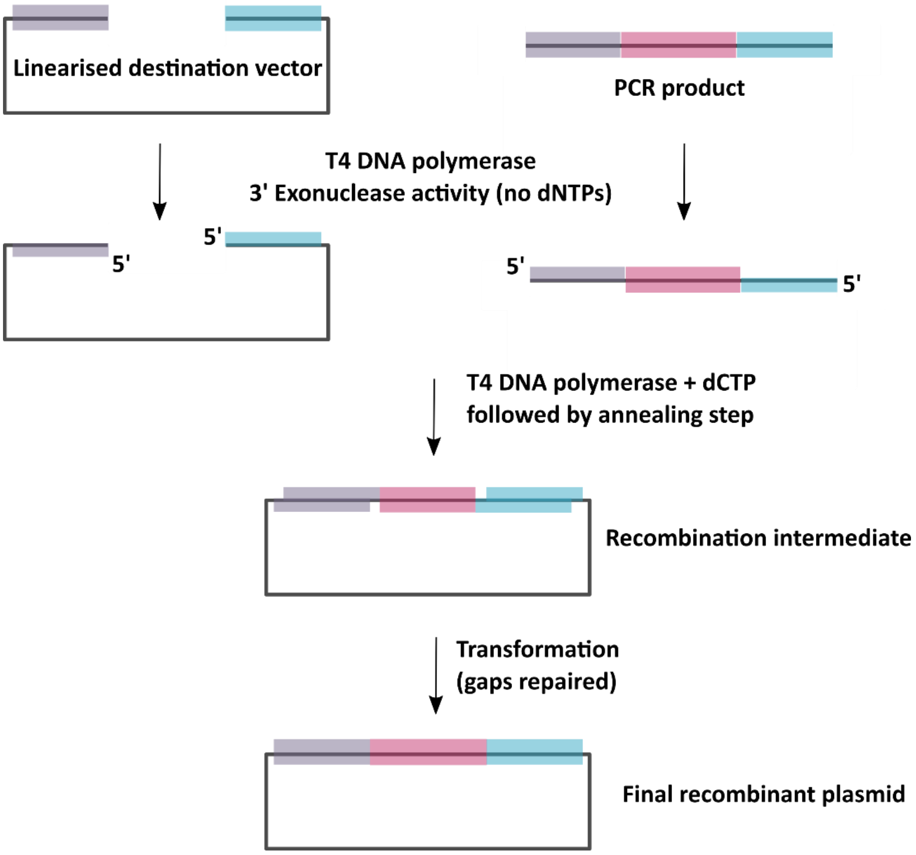


Figure 5-3: Schematic representation of SLIC cloning. Primers are designed to amplify the destination vector and insert (pink) with overlapping homology (shown in blue and purple). T4 polymerase without dNTPs acts as a 3' exonuclease and generates single stranded overhangs of these homologous regions. Addition of dCTP stops the reaction and when mixed the fragments self-anneal. Transformation into *E. coli* repairs the gaps producing the final recombinant plasmid.

The pHS1403 vector, encoding T4 Lysozyme C54S C97A, was a gift from Brian Matthews (Addgene plasmid # 18355), while the destination vectors were pET28b_haspb1 and pET28b_haspa (see Section 4.2.1. and Appendix A.1.). Primers were designed in accordance with the SLIC procedure, the sequences of which are in the Appendix A.1. Tables 3-2 and 3-3 outline reaction conditions and temperature cycles used. PCR products were analysed by agarose gel electrophoresis and treated with 20 U of DpnI (NEB) per 100 µl then incubated at 37°C for 1 hour. The QIAquick PCR Purification Kit (Qiagen) was used for sample purification.

Reagent	Final concentration	Final volume in 20 µl
ddH_2O		11.8 µl
5x Phusion buffer	1x	4 µl
dNTPs (10 mM stock)	200 µM each	0.4 µl
Forward primer (10 µM stock)	0.5 µM	1 µl
Reverse primer (10 µM stock)	0.5 µM	1 µl
Template DNA (50 ng/ µl)	50 ng	1 µl
DMSO	3%	0.6 µl
Phusion DNA Polymerase (2U/ µl)	0.4 U	0.2 µl

Table 5-2: PCR master mix for SLIC cloning.

Step	Temperature (˚C)	Time (s)	No. of cycles
Initial denaturation	98	120	1
Denaturation	98	20	25
Annealing	55	20	
Extension	72	20/ 165*	
Final extension	72	420	1

Table 5-3: SLIC PCR temperature profile. *Extension time differences are due to varied PCR product lengths, 20 seconds for the insert and 165 seconds for the destination vector.

Single strand overhang generation was carried out as follows. 1 µg of the vector and 1 µg of the insert were treated separately with 0.75 U of T4 DNA polymerase (NEB) in NEB buffer 2 plus BSA (see Table 5-4). The reaction mix was incubated at room temperature for 30 minutes. Addition of 1/10 volume of 10 mM dCTP stopped the reaction and the samples were transferred to ice.

Reagent	Final concentration	Final volume in 30 µl
NEB T4 DNA polymerase (3U/ µl)	0.75 U	0.25 µl
BSA (10 mg/ml)	50 µg	0.15 µl
NEB buffer 2 (10x)	1x	3 µl
Sample DNA	1 µg	X µl (to give 1 µg)
ddH_2O	To give a final volume of 30 µl	

Table 5-4: Reaction mix for T4 polymerase mediated single strand overhang generation.

The annealing reaction was set up with a 1:1 molar ratio of vector to insert, using 0.07 pmol of both components. 1 µl 10x T4 DNA Ligase Reaction Buffer (NEB) was added to the reaction and the volume was made up to 10 µl with ddH_2O. The reaction was incubated at 37˚C for 30 minutes. Electro-competent *E. coli* XL-10 were transformed with 5 µl of the desalted annealed product and plated out on LB agar plates containing kanamycin at 50 µg/ ml. Following overnight incubation at 37˚C colonies were picked, grown in 5 ml kanamycin supplemented LB media overnight at 37˚C with shaking at 180 rpm and plasmid DNA purified with the QIAprep Spin Miniprep Kit (Qiagen). Positive clones were identified by EcoRV/ EcoRI double digestion, exploiting a unique EcoRV site in the T4 lysozyme gene. The reaction was set up according to the schedule in Table 5-5 and incubated for 1 hour at 37˚C. A 1% DNA agarose gel was then used to analyse the fragments produced by the double digestion and identify positive clones, which were verified by sequencing.

Reagent	Final concentration	Final volume in 50 µl
Test DNA	0.5 µg	X µl (to give 0.5 µg)
10x NEBuffer EcoRI	1x	5 µl
NEB EcoRI (10 U/µl)	5 U	0.5 µl
NEB EcoRV (20 U/µl)	10 U	0.5 µl
BSA (10 mg/ml)	5 µg	0.5 µl
ddH_2O	To give a final volume of 50 µl	

Table 5-5: EcoRI/ EcoRV double digestion reaction conditions.

5.2.2. Expression of Nt-T4L

Electro-competent *E. coli* BL21 gold (DE3) were transformed with pET28b_NtB1_T4L and plated out on LB agar plates containing kanamycin at 50 µg/ml. After overnight incubation at 37˚C colonies were picked, plasmid DNA purified with the QIAprep Spin Miniprep Kit (Qiagen) and digested with EcoRV and EcoRI to ensure positive clone selection (see Section 5.2.1). A glycerol stock was prepared with a positively transformed clone. Small scale expression tests were carried out to optimise expression conditions. 5ml LB media supplemented with kanamycin was inoculated with the glycerol stock and grown overnight at 37˚C and 180 rpm. This was used as a starter culture to inoculate (2% of the final volume) fresh 5 ml kanamycin supplemented LB media aliquots grown at 37˚C at 180 rpm. Expression was either induced with 0.5 mM or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} 0.6 or OD_{600} of either 0.5 or 1.0 with 1 mM IPTG.

Expression temperature was also a considered factor with cultures grown at 37˚C for 4 hours or 16˚C overnight at 180 rpm in both cases.

Small scale optimisation allowed for streamlined large scale expression. Aliquots of LB media (10ml) containing kanamycin at 50 µg/ ml were inoculated with the *E. coli* BL21 (DE3) pET28b_NtB1_T4L glycerol stock. These cultures were grown overnight at 37°C at 180 rpm and then used to inoculate 500 ml LB media each. The 500 ml cultures were grown to an OD_{600} of 0.8 to 1.0 at 37°C at 180 rpm. Protein expression was induced with 1mM IPTG and the cultures were transferred to 30°C for 4 hours with shaking at 180 rpm. The cells were harvested by centrifugation for 15 minutes at 5000 xg and the resulting pellets were stored at -20˚C.

5.2.3. Purification of Nt-T4L

Pellets were thawed on ice and re-suspended in 30 ml nickel column buffer A (50 mM Tris pH 7.5, 500 mM sodium chloride, 20 mM imidazole) per 500 ml pellet supplemented with a protease inhibitor tablet (cOmplete, EDTA-free from Roche). The cell suspension was sonicated on ice for 30 seconds 6 times with 1 minute intervals. Lysate clarification was achieved by centrifugation at 38000 xg for 30 minutes. The supernatant was loaded onto a pre-equilibrated 1ml HisTrap FF crude column (GE Healthcare) and washed with six column volumes of nickel column buffer A. The target protein was eluted using a gradient of 0 to 100% nickel column buffer B (50 mM Tris pH 7.5, 500 mM sodium chloride, 500 mM imidazole) for 20 column volumes. SDS-PAGE was used to identify target protein containing fractions that were concentrated (Vivaspin Sample Concentrator range, 5000 MWCO) to 2 ml for further purification with size exclusion chromatography. A Superdex 75 16/60 column (GE Healthcare) equilibrated in 20 mM sodium acetate pH 5.5, 300 mM sodium chloride was used to further purify Nt-T4L. Final sample purity was analysed with SDS PAGE. The purification schedule was run with the AKTA FPLC system.

ESI-MS was used to confirm the mass of Nt-T4L. This in house service was run by Dr. Andrew Leech with the protein supplied at 2 mg/ml in 20 mM sodium acetate pH 5.5, 15 mM sodium chloride. Data were collected with an ABI Qstar tandem mass spectrometer. The polydispersity and batch consistency of Nt-T4L was analysed with dynamic light scattering (DLS). Measurements were carried out at 20˚C with the protein at 1 mg/ml in 20 mM sodium acetate pH 5.5, 300 mM sodium chloride with the DynaPro Dynamic Light Scattering system (Protein Solutions).

5.2.4. Crystallisation trials with Nt-T4L

Extensive crystallisation screening and optimisation was carried out with the Nt-T4L construct (Table 5-6). Initially the protein was screened at 12 mg/ml (later 25, 38 and 41.5 mg/ml) in 20 mM sodium acetate pH 5.5 at 20˚C. Protein desalting was carried out by repeated washing and concentrating of the protein with 20 mM sodium acetate pH 5.5 using Vivaspin 500 sample concentrators (5000 MWCO). Screening was in a 96 well sitting drop vapour diffusion format with drops of 150 nl protein plus 150 nl well solution pipetted by a Mosquito robot (TTP Labtech Ltd). Optimisation trays were set up in either a 48 or 24 well hanging drop vapour diffusion format with 1 µl protein plus 1 µl well solution.

[Protein] (mg/ml)	Tray composition
12	Hampton I+II and Index screens
12	Variations of the 'best' lysozyme condition 30% PEG 5000 MME, 1.0 M Sodium chloride, 0.1 M Sodium acetate pH 4.6
25	Hampton I+II, Index, Pact and CSS I+II at pH 4.5 and 6.5
25	Dilution and pH screen (pH 4.5, 6.5 and 8.5) of 'best' conditions from original Hampton screen - 30% PEG 400, 0.2 M CaCl2, 0.1 M Hepes pH 7.5 - 30% PEG 8000, 0.2 M Na acetate, 0.1 M Na cacodylate pH 6.5
25	Optimisation of hits from screening. 2x24 well trays. - Various salts were trialled with PEG 400 and butandiol. - All small PEGs were used to expand the 30% PEG 400, 0.2 M $CaCl_2$, 0.1 M Hepes pH 6.5 'best' condition.
25	The best condition of 30% PEG 400, 0.2 M $CaCl_2$, 0.1 M Hepes pH 6.5 with additives and cationic detergents. - TMAO, SB12, LDAO, βOG, NDSB 201, Isopropanol, DMF, Dioxane, DMSO - Choline chloride, Girard's reagent T, Dodecyltrimethylammonium bromide and chloride. - Note a crystal was tested from the LDAO condition but was found to be salt.
25	The best condition of 30% PEG 400, 0.2 M $CaCl_2$, 0.1 M Hepes pH 6.5 with sugars. - Glucose, lactose, maltose and xylitol.
25	Best condition with varied MPD concentration in place of PEG 400 and different buffers at pH 6.5 were tested.
25	**Set up at 4°C.** Best condition with 10 mM LDAO at pH 6.5 using both Hepes and Bis Tris propane as buffers.
25	Further MPD screening and exploration of LDAO as an additive.
25	Screening of best condition with various polyoxyethlene ethers.
38	JCSG screen
38	Optimisation of JCSG best conditions at a pH range between 4 and 8.5 (0.5 gaps). - 20% Peg 8000, 0.1 M Tris pH 8.5. - 0.1 M KSCN, 30% PEG 2000 MME.
38	Optimisation of precipitant concentrations (25 to 100 mM NaCl and KSCN) of JCSG best conditions at best pHs 4.5, 8.5 and 9.5.
38	Optimisation of buffer for both JCSG best conditions from pH 7.5 to 9.3 (0.2 gaps).
38	Screen of published conditions for monoclinic and triclinic lysozyme crystallisation.
41.5	Extensive screening of the best PACT conditions - 0.2 M NaF/ NaBr/ $NaNO_3$, 20% PEG 3350
41.5	Rescreening of Pact with 2.5 mM chitobiose or 0.2 M maltose to stabilise T4 lysozyme
41.5	Variations of best condition; 30% PEG 2000 mme, 0.1 M citric acid pH 5.5, 0.2 M sodium fluoride

Table 5-6: Summary of crystallisation attempts with Nt-T4L.

5.2.5. Cross seeding with Hen egg white lysozyme

Crystals of hen egg white lysozyme (HEWL), obtained as a lyophilised powder from Sigma, were grown at 100 mg/ml in 20 mM sodium acetate pH 4.6 in a 24 well hanging drop vapour diffusion format with 1 µl protein plus 1 µl well solution at 20˚C. The well solution was 30% (w/v) polyethylene glycol monomethyl ether 5000, 1 M sodium chloride and 0.1 M sodium acetate pH 4.6. A seed stock was created with the subsequent crystal growth using the Seed Bead Kit (Hampton) as instructed. This stock was used for cross seeding of the Nt-T4L crystallisation attempts (Obmolova et al. 2010). A JCSG (Molecular Dimensions) screen was prepared in a 96 well sitting drop vapour diffusion format with drops of 150 nl protein, 100 nl well solution and 50 nl seed stock pipetted by a Mosquito robot (TTP Labtech Ltd). Nt-T4L was used at 25 mg/ml in 20 mM sodium acetate pH 5.5 and the seed stock was used in two dilutions; the original seed stock and a 1/100 dilution with mother liquor.

The ’rational poisoning’ experiments were set up at molar ratios of 1:1 and 1:10 (HEWL: Nt-T4L) with the fusion protein kept constant at 1.6 mM in 20 mM sodium acetate pH 5.5. Pact (Molecular Dimensions) and Index (Hampton) screens were prepared with the above ratios in a 96 well sitting drop vapour diffusion format with drops of 150 nl protein mixture and 150 nl well solution pipetted by a Mosquito robot (TTP Labtech Ltd). Crystals were obtained in various drops and vitrified without cryo-protectant before house testing. Those with unit cell parameters matching known HEWL crystal forms were discarded. One promising crystal grown in 0.2 M sodium malonate pH 7, 20% (w/v) polyethylene glycol (PEG) 3350 at a ratio of 1:1 was sent for data collection at the Diamond Light Source (DLS). As a control, all three samples (1:1, 1:10 and HEWL alone) were set up in a 24 well hanging drop vapour diffusion format (1µl plus 1µl) with the standard HEWL crystal condition described above. No crystal growth was observed in either mixed sample.

5.3. Results

5.3.1. SLIC cloning

The HASPB1 N-terminal domain fused to T4 lysozyme construct was produced according to the schedule in Section 5.2.1. The vector (5373 bps) and insert (495 bps) were amplified and the product sizes were assessed by agarose gel electrophoresis (Figure 5-4 lanes 2 and 3). To identify positive clones an analytical EcoRV/ EcoRI double digest utilised the unique EcoRV site in the T4 lysozyme gene and the one EcoRI site in the pet28b vector. This method gave a clear distinction between positive (two cuts produce two bands) and negative (one cut produces one band) samples (Figure 5-4 lanes 4 and 5). The samples that gave the correct bands were then sent for sequencing, which showed correct incorporation of the lysozyme gene and alanine linker. The final plasmid was termed pET28b_NtB1_T4L and contains a C-terminal non-cleavable his-tag.

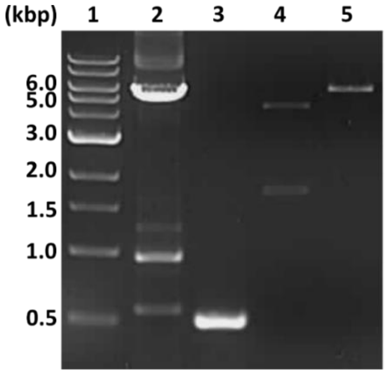


Figure 5-4: 1% agarose gel of SLIC products and final plasmid. 1 kb ladder (lane 1) pET28b_haspb1 - 5373 bps (lane 2), T4 lysozyme insert - 495 bps (lane 3), EcoRI/EcoRV double digestion of pET28b_NtB1_T4L (lane 4), EcoRI/EcoRV double digestion of pET28b_haspb1 (lane 5).

The cloning of the HASPB1 N-terminal domain plus one repeat: T4 lysozyme chimera (NtR-T4L) was much more problematic. The samples that gave the correct bands from the EcoRV/ EcoRI double digest, but the sequencing data revealed a highly variable number of repeats between samples due to mispriming. To prevent this a new procedure was developed to build the repeat back in, eliminating the need for direct amplification from pET28b_haspb1, circumventing the repeat sequence (Figure 5-5). A forward primer was designed to build in the missing element of the repeat fragment and a portion of the N-terminal domain at the 5’ end of the lysozyme insert already produced for the N-terminal domain plus one repeat cloning schedule (see Appendix A.1.). A 45 base pair size difference between these two fragments meant that they could be separated on a 2% agarose DNA gel and standard gel purification carried out. The SLIC protocol was run with the already positively sequenced pet28b N-terminal domain empty vector, the starting point for pET28b_NtB1_T4L production, and the compatible end product of the two PCR steps. Unfortunately after extensive rounds of this procedure no positive clones were identified.

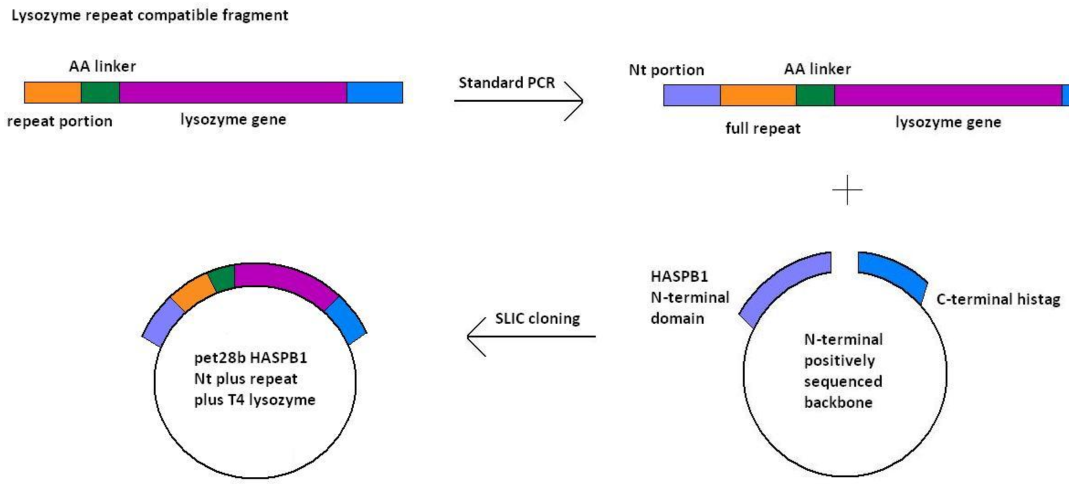


Figure 5-5: Cloning schedule to produce the HASPB1 N-terminal domain plus repeat lysozyme construct. The lysozyme insert with a 5’ portion of the repeat of HASPB1, used in the previous round of SLIC cloning, was the starting point for this procedure. A standard PCR amplification of the lysozyme fragment with the redesigned primer and reverse SLIC primer produced the fragment seen on the right. This fragment is compatible with the empty N-terminal pet28b linearised vector used in the successful pET28b_NtB1_T4L cloning work. The SLIC products can then be annealed and transformed, enabling correct incorporation of the HASPB1 repeat without mispriming issues.

The pET28b HASPA: T4 lysozyme construct was not produced although the procedure was run through multiple times and subjected to extensive troubleshooting. The use of the same vector as a PCR template and final destination vector meant an excessive amount of false positives were seen when double digesting transformed plasmid DNA. I decided to focus my efforts on the Nt-T4L construct and return to the cloning of the remaining constructs if time allowed.

5.3.2. Expression of Nt-T4L

The pET28b_NtB1_T4L plasmid was transformed into BL21 Gold (DE3) cells. To ensure optimal protein production small scale expression tests were carried out to explore the yield of various conditions. The concentration of IPTG (0.5 or 1 mM IPTG), temperature (37˚C or 16˚C), expression time (4 hours or overnight) and point of induction (optical density of 0.5, 0.6 or 1.0) were all varied. Figure 5-6 shows the resultant SDS PAGE analysis of the expression conditions. Lane 5, induction at an optical density of 1.0 with 1 mM IPTG then 37˚C for 4 hours, exhibited the most pronounced expression level. This condition was used for large scale expression, with a slight temperature modification (30˚C instead of 37˚C) to ensure protein stability. It is important to note that some 'leaky' expression was observed in the uninduced negative control sample but this didn't impede cell growth so it was deemed acceptable.

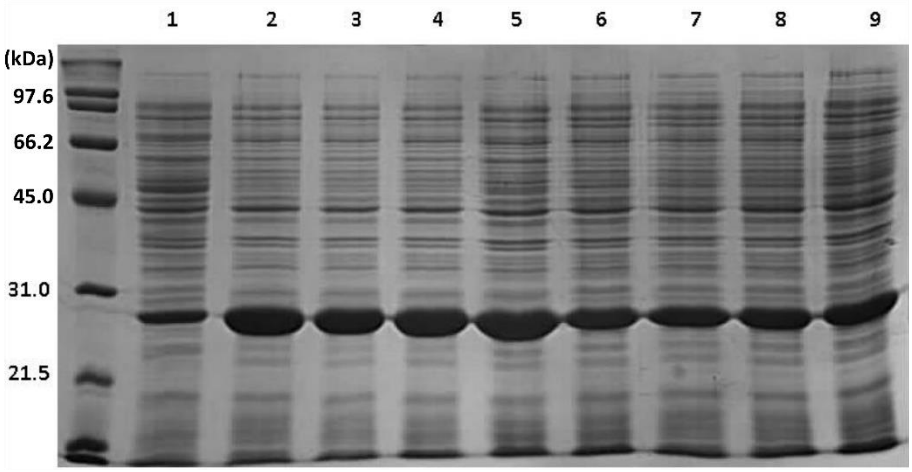


Figure 5-6: Comparison of soluble fractions frim various expression conditions tested for Nt-T4L on 12% SDS PAGE. The gel shows uninduced control (1), induction with 5 mM IPTG (2+6), or with 10 mM IPTG (3+7) at OD 0.6, OD 0.5 (4+8) or OD 1.0 (5+9) induction with 1 mM IPTG. Samples were then incubated at 37˚C for 4 hours (2 to 5) or 16˚C overnight (6 to 9) at 180 rpm. Nt-T4L is 24.5 kDa.

5.3.3. Purification of Nt-T4L

A two-step purification procedure was optimised for maximal yield and sample purity, the details of which can be found in Section 5.2.3. The first step was nickel affinity purification, giving a considerably pure eluted sample for this initial step (see Figure 5-7a and b). These fractions were pooled and concentrated for size exclusion purification (see Figure 5-7c and d). Adequate separation from remaining containments was achieved, observed by SDS PAGE, to allow sample characterisation. The homogeneity and intensity was then assessed by DLS and ESI-MS. DLS reported a polydispersity of 12% indicative of a highly monodisperse sample, a feature of a readily crystallisable construct (Figure 5-7e). The molecular weight obtained by ESI-MS was 24380 Daltons (Figure 5-7f), which was in line with the N-terminal methionine cleaved weight of 24380.5 Daltons obtained from ProtParam (Gasteiger et al. 2005). Moreover, no major adducts were noted in this preparation. These positive characterisation results meant that crystallisation trials were started and continued with no modification of the purification schedule.

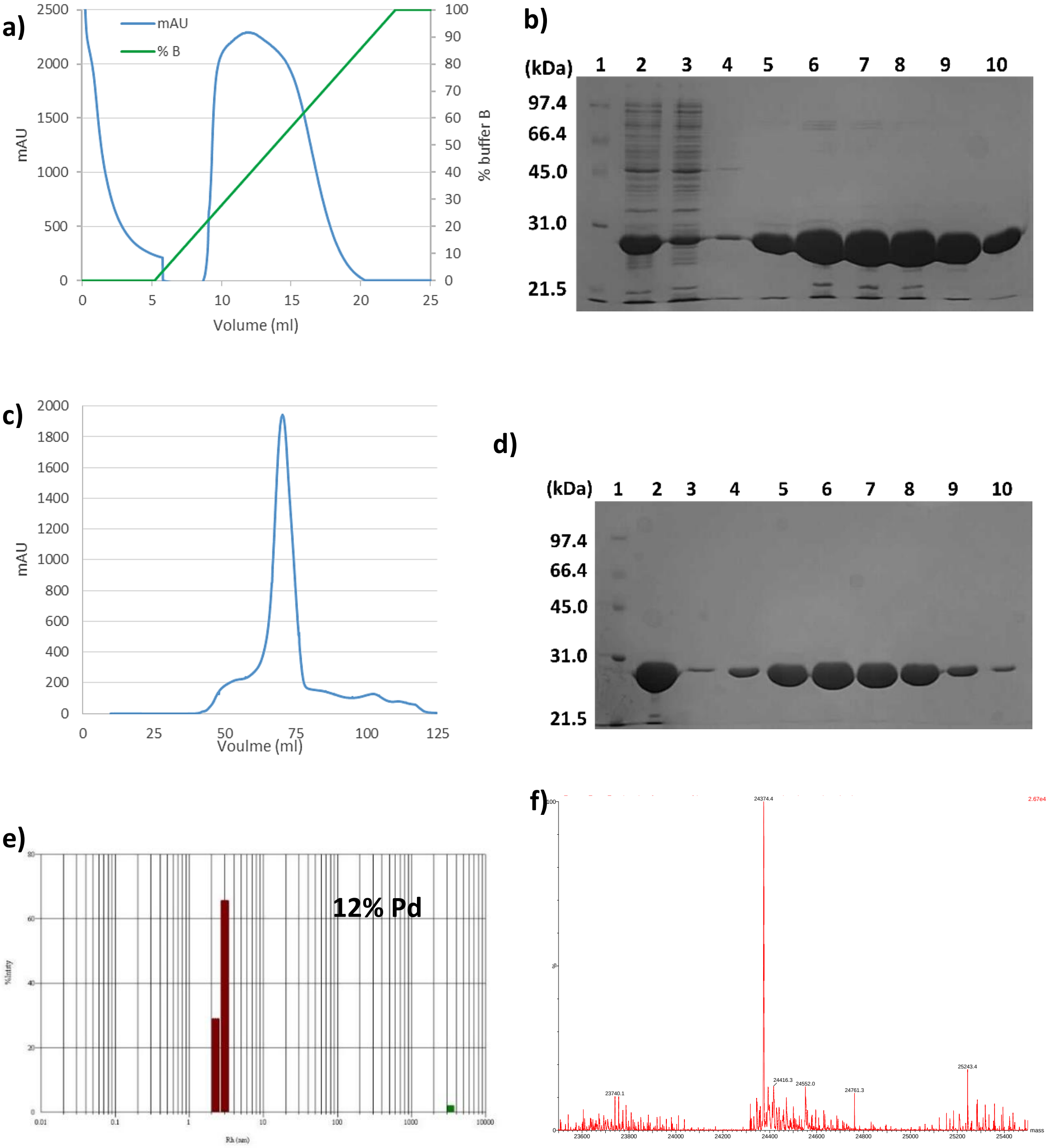


Figure 5-7: Chromatography traces and SDS PAGE analysis of the two step Nt-T4L purification. a) 1ml HisTrap FF crude column chromatography trace. The UV trace (280 nm) is shown in blue and percentage buffer B (where 100% equates to 500 mM imidazole) is in green. b) 12% SDS PAGE analysis of load and resultant fractions from (a). The gel shows low molecular weight ladder (1), cell lysate load (2), column flow through (3) and resultant fractions (4 to 10). c) Chromatography trace of 16/60 Superdex 75 purification of Nt-T4L. d) 12% SDS PAGE analysis of the size exclusion purification of Nt-T4L. The gel shows low molecular weight ladder (1), load (2) and resultant fractions (3 to 10). The Nt-T4L construct is 24.5 kDa. e) DLS regularisation histogram of an Nt-T4L sample, note percentage intensity is plotted against hydrodynamic radius (R_H). The polydispersity (Pd) recorded was 12%. f) Electrospray ionisation mass spectrometry (ESI-MS) result of Nt-T4L reporting the expected molecular weight for this sample of 24380 Da.

5.3.4. Crystallisation trials with Nt-T4L

Classical intensive screening and optimisation at various protein concentrations produced no usable crystals of Nt-T4L. Application of the cross seeding strategy, originally developed to produce crystals of mouse, human and affinity matured Fab: antigen complexes with a seed stock created from the human Fab: antigen complex (Obmolova et al. 2010), was the next step to try to circumvent this barrier. Ideally T4 lysozyme crystals would be the best starting point owing to the high sequence similarity with Nt-T4L. However, the crystallisability and abundance of hen egg white lysozyme (HEWL) as a lyophilised powder, meant this approach could be trialled without further protein production. Screening with the HEWL seed stock at various dilutions produced no workable hits even after considerable optimisation.

A radical approach was taken. The two proteins, Nt-T4L and HEWL, were mixed in two different molar ratios (1:1 and 1:10 HEWL: NT-T4L) and screened as a mixture. The thought process behind this ‘rational poisoning’ strategy (named by Professor Marek Brzozowski) was that as a HEWL crystal began to form in the drop it would be an anchor for Nt-T4L inclusion. Screening with Pact (Molecular Dimensions) and Index (Hampton) produced a multitude of crystals. These were fished and then tested with those displaying unit cell parameters of known HEWL crystal forms discarded. One crystal exhibited unit cell parameters that were not known for HEWL and data were collected at the Diamond Light Source (DLS) (Table 5-7). The crystal diffracted to 1.25 Å (Figure 5-8) and the data belongs to either $P4_1\ 2_1\ 2$ or $P\ 4_3\ 2_1\ 2$. Unfortunately, molecular replacement using either HEWL or T4 lysozyme as the model did not facilitate data deconvolution. It was decided that this data could not be taken any further and crystallisation attempts of Nt-T4L were halted.

	Nt-T4L: HEWL
Data collection	
Diffraction Source	DLS beamline IO4-1
Wavelength (Å)	0.92
Temperature (K)	100
Detector	Pilatus 2M
Rotation range per image (°)	0.1
Total rotation range (°)	180
Crystal data	
Space group	$P4_1\ 2_1\ 2$ or $P\ 4_3\ 2_1\ 2$
a, *b*, *c* (Å)	78.44 78.44 114.50
α, β, γ (°)	90.00 90.00 90.00
Resolution (Å)	27.35 - 1.25 (1.25)
Total reflections	1283555 (59229)
Unique reflections	99094 (4851)
Completeness (%)	100 (100)
Redundancy	13.0 (12.2)
R_{merge}	0.06 (1.37)
R_{pim}	0.03 (0.6)
$CC_{1/2}$	0.99 (0.69)
$I / \sigma I$	19.4 (1.9)

Table 5-7: Crystallographic data table for the HEWL: Nt-T4L crystal. Values in parentheses relate to the highest-resolution shell.

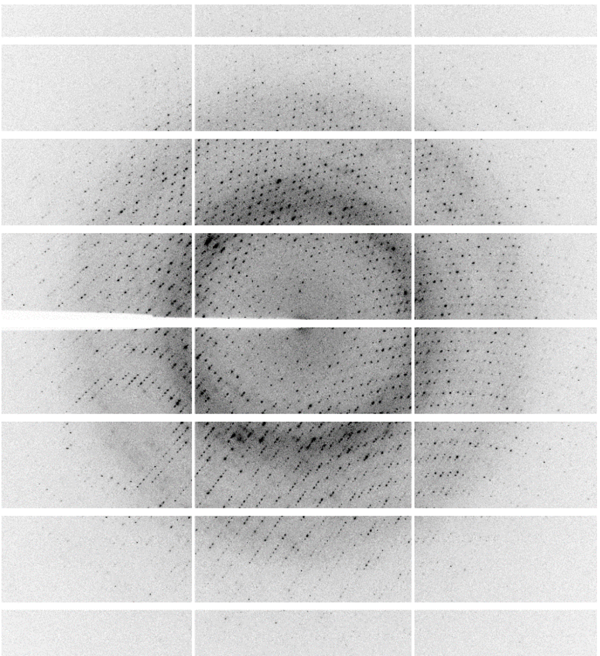


Figure 5-8: Diffraction pattern of the HEWL: Nt-T4L crystal.

5.4. Discussion

5.4.1. Covalent attachment of HASP impedes T4 lysozyme crystallisation

Due to advances in data collection and processing crystallisation has become the rate limiting step in crystallographic studies of proteins and other biological macromolecules. After comprehensive database analysis Price et al. (2009) reported that 35% of all proteins worked on by structural genomics consortia form crystals, while only 12% form crystals of sufficient quality for data collection. Complex and diverse protein targets, such as membrane proteins, intrinsically disordered proteins, transient biological assemblies, highly unstable oligomerisation intermediates and short lived functionally relevant conformations, add a distinct layer of complexity as they are not generally amenable to traditional crystallisation screening. These varied biological systems all share intrinsic structural heterogeneity, which facilitates functional plasticity *in vivo*, inhibiting their propensity to crystallise (Bukowska and Grütter. 2013). Incorporation of highly heterogeneous elements, such as flexible domains or loops or bulky side chains, into a growing crystal greatly increases the entropic cost of the system (Derewenda. 2010). This in turn hinders the formation of the required crystal contacts, making the generation of X-ray diffraction quality crystals a herculean task. Modulating protein crystallisability in a rational manner, via techniques that include removal of post translational modifications, surface entropy reduction, covalent and non-covalent crystallisation chaperones, truncations and flexible domain removal, ligand or partner protein binding, is increasingly used to coerce challenging targets. Stabilisation of a single conformation, enabling crystal growth, is the end goal of these approaches with implementation of a multidimensional strategy prevalent.

In Chapter 4, I focused on Fab fragment stabilisation, truncations and additive based stabilisation of the HASPs, covering some elements of the conformation stabilisation scheme. However, no X-ray diffraction quality crystals were produced, requiring a fresh approach. Exploration of the successful use of T4 lysozyme as a covalent crystallisation chaperone for various GPCRs (see Section 5.1.) led to the idea of adapting this method for the HASPs. In the case of membrane protein attachment, T4 lysozyme not only adds rigidity to the overall structure but increases the hydrophilic surface area necessary for crystal contacts to form. Hydrophilicity is not limited in the hydrophilic acylated surface proteins (HASPs), but the structural rigidity of T4 lysozyme was appealing.

The amino acid sequences of HASPB1 and Nt-T4L were analysed with the disorder prediction server IUPred to assess the stabilising influence of T4 lysozyme attachment (Figure 5-9). IUPred predicts disordered regions of proteins by estimating the capacity of polypeptides to form stabilizing contacts, based on the assumption that globular proteins are stabilised by extensive inter-residue interactions while IDPs are not (Dosztányi et al. 2005). Comparison of the resultant disorder tendency plots emphasises the disordered nature of the native HASP sequence, with the entirety of HASPB1 predicted to be disordered (Figure 5-9). Conversely, the Nt-T4L sequence is predominantly predicted to be ordered, although unsurprisingly the element of disorder observed is due to the N-terminal domain of HASPB1 (the first 50 residues of the chimera). This vast improvement in the predicted disorder of Nt-T4L over native HASPB1 was the underlying idea of this approach. However, the addition of the HASP element prevented T4 lysozyme, a protein chosen as a scaffold because of its ability to crystallise, from forming crystals.

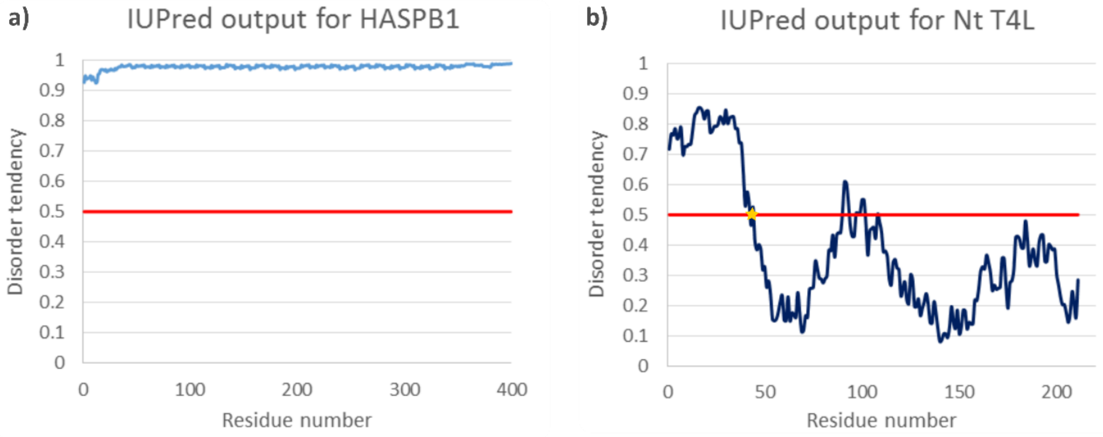


Figure 5-9: IUPred generated disorder tendency plots for (a) HASPB1 and (b) Nt-T4L. The disorder threshold is shown in red at 0.5. Anything below this threshold is considered globular, while anything above is not. The yellow star in plot b at around residue 50 denotes the end of the N-terminal HASB1 domain.

The success of the T4 fusion approach has been to fix already structured elements in one conformation and concurrently remove highly flexible loops that impede crystallisation. This approach relies on the target protein containing some structured domains or being induced into a fixed conformer by ligand/ partner protein binding.

However, this chapter and the previous one have shown that the HASPs do not satisfy the criteria for successful fusion as they do not contain any stable structural elements. Inducing structural rigidity was the preferred outcome for this work but now appears an unlikely goal. Reclassification of the HASPs, in the confines of this project and potentially beyond, as intrinsically disordered seems a pragmatic approach to gain an insight into their nature.

5.5. Summary and future work

T4 lysozyme has been used repeatedly as an effective crystallisation chaperone for many challenging GPCR structures, owing to its inherent structural rigidity and crystallisability. The disrupting influence elicited by the N-terminal domain of HASPB1 on readily crystallisable T4 lysozyme strengthens the case for IDP classification. In isolation this result does not have the same resonance, as challenging crystallisation targets often require multiple approaches to obtain elusive structural data. When taken in combination with the work done in Chapter 5 it is clear that classing the HASPs as intrinsically disordered means that continuation of crystallisation attempts without a proven stabilisation method, such as biological binding partner to trap an expected conformation, is inappropriate. Refocusing the project to biophysically characterise the structural heterogeneity of the HASPs and investigate their interaction with N-myristoyltransferase appeared to be the way forward.

Chapter 6 - Discussion

6.1. Thesis discussion

Protozoan parasites of the genus *Leishmania* cause a wide range of tropical diseases referred to as Leishmaniasis. To date there is no licensed vaccine against any form of Leishmaniasis and the treatment options available are expensive and largely ineffective. The hydrophilic acylated surface proteins (HASPs) are found in all human infective *Leishmania* species, where their expression correlates with parasite infectivity (Alce et al. 1999). The HASPs share highly conserved N- and C- terminal domains, but a subset, the HASPBs, have a divergent central domain containing extensive hydrophilic amino acid repeats that vary in size and composition within and between species (McKean et al. 1997). The N-terminal region encompasses a dual acylation motif that is required to target the protein to the parasite plasma membrane (Denny et al. 2000). This region is myristoylated and then palmitoylated for correct membrane association. While the HASPBs have been shown to be antigenic in the host and can induce a protective immune response without adjuvant, the biological function of these proteins is still unclear.

This project was conceived as an X-ray crystallographic study to decipher functionality from resulting structural data of the HASPs from *Leishmania donovani*. Extensive crystallisation screening of the three full length HASPs did not produce any crystals. Therefore, a wide range of strategies to coerce HASP crystallisation were implemented, namely Fab co-crystallisation, use of cationic detergents and membrane mimetics for stabilisation, generation of HASPB1 fragments and recombinant lysozyme fusion. However, these non-standard crystallisation schemes did not produce crystals of the HASPs. Biochemical and biophysical characterisation experimentally validated predicted intrinsic disorder, meaning continuation of crystallisation screening without a mechanism to induce protein folding in a biologically relevant manner was inadvisable. The project was refocused to biophysically characterise the structural heterogeneity of the HASPs and investigate their interaction with N-myristoyltransferase.

Nuclear magnetic resonance (NMR) spectroscopy is unique in its capacity to study the structural biology of intrinsically disordered proteins in solution (Dyson and Wright. 2004). ^{1}H, ^{15}N-HSQC spectra of HASPA and HASPB2 revealed limited dispersion of resonances in the amide proton dimension, which is a feature indicative of disorder. Although this limited dispersion caused spectral congestion the majority of resonances were well defined, suggesting that backbone resonance assignment would be possible. HASPA, the smallest HASP at 9.5 kDa, was chosen for assignment as it lacks the repeats found in HASPB2 and HASPB1. The disordered nature of HASPA and low sequence complexity required implementation of metabolic precursor unlabelling to deconvolute the assignment process. Specific metabolic precursors can be used to selectively unlabel certain residues in an otherwise uniformly-labelled sample rendering them 'NMR invisible' (Rasia et al. 2012). This strategy enabled a near complete backbone resonance assignment of the H_N, N, Cα and Cβ nuclei in HASPA. NMR derived chemical shift information was used to determine if HASPA contained any elements of structural propensity that could be functionally relevant. This analysis showed that it does not contain elements of structural propensity, suggesting that preformed structure is not functionally required.

N-myristoyltransferase (NMT) catalyses the addition of myristate to the HASPs *in vivo*. It was shown for the first time that recombinant *Leishmania major* NMT is able to catalyse recombinant HASP myristoylation *in vitro*. A fluorescence assay, where a fluorescent CPM-CoA adduct is formed on myristoylation, was implemented to establish the kinetic parameters for this reaction (Goncalves et al. 2012). K_m values between 5.8 and 16.7 µM were reported for the three full length HASPs, in line with values reported for *Leishmania donovani* NMT with various peptide substrates using this method (personal communication with Dr. Victor Goncalves). Real-time NMR spectroscopy was used to monitor the effect and any potential structural impact of NMT-catalysed myristoylation on HASPA. Successive ^1H, ^{15}N HSQC spectra were recorded, which showed that there were no global conformational changes to HASPA in solution, with limited changes in peak position and intensity observed. The residues proximal to the N-terminal glycine and site of myristate addition experienced the most pronounced changes in chemical shift or resonance intensity. A novel 2.45 Å resolution crystal structure of *Leishmania major* NMT in complex with 2-oxopentadecyl-CoA (NHM) was elucidated.

The data presented in this thesis have shown the HASPs to be intrinsically disordered, which should lead to a reframing of the biological questions asked about their function.

The IDP field has shown that disorder as much as order is functionally relevant, particularly when you consider the entropic chains elastin and titin (discussed in Chapter 2). These proteins remain disordered to elicit their functions. Some IDPs are not fully disordered but contain elements that are predisposed to adopt local secondary structure required for biological partner binding termed molecular recognition features (MoRFs). Chemical shift analysis showed that HASPA does not contain elements of transient structure. This means the HASPs do not possess the MoRFs commonly associated with IDPs that function as interaction hubs. This leads to the conclusion that the HASPs could function as entropic chains and that the persistent disorder observed is functionally vital.

The functions revealed for IDPs are increasingly diverse and in some cases unexpected. Recently, the intrinsically disordered domains of the endocytic adaptor proteins Epsin 1 and AP180 were shown to be potent drivers of membrane curvature (Busch et al. 2015). This was unexpected as the prevailing view was that proteins containing distinctive membrane curvature promoting structural motifs were required to direct this process. Busch et al. (2015) went on to decipher how these intrinsically disordered regions (IDRs) facilitate membrane bending. Firstly, they showed that the IDRs crowd membrane surfaces and exert steric pressure that drives membrane curvature. The advantage of IDRs over folded proteins of the same molecular weight is that they have a larger hydrodynamic radius, meaning they are able to exert steric pressure over a larger volume than folded proteins of the same size. This also means that membrane crowding can be achieved with less protein. The IDRs of Epsin 1 and AP180 fulfil their functional role while membrane associated and disordered, exactly like the HASPs. It would be interesting to pursue functional determination of the HASPs with this example in mind.

6.2. Future perspectives and further work

6.2.1. Membrane association of the HASPs

The small hydrophilic endoplasmic reticulum associated protein (SHERP) is found on the same locus as the HASPs (*LmcDNA16*) and its expression is stage regulated (Knuepfer et al. 2001). SHERP is highly disordered in solution with an acidic pI similar to the HASPs. Synchrotron radiation CD was used to show addition of anionic lipids induced predicted helicity of this known membrane associated protein (Moore et al. 2011). This led to elucidation of an NMR structure of SHERP in the presence of 50 mM SDS. The functional implications of this structure and the observation that SHERP associates with V-ATPase *in vivo*, suggests a role in the acidification process vital for parasite development. Acidification is required for parasite development and has been shown to induce development of metacyclic parasite *in vitro* (Bates. 2008). This example highlights the vital importance of emulating *in vivo* conditions, in particular the use of membrane mimetics, to understand the structure and functional role of a protein.

SHERP interacts with the head groups of the membrane to function and fold, whereas the HASPs associate with the membrane via the dual acylation motif. The work presented in this thesis has shown that recombinant HASPs can be myristoylated by NMT *in vitro*. However, myristoylation alone is too weak to mediate irreversible binding, which is why it is usually found in the context of another modification. The comparably short myristate chain has a limited capacity for hydrophobic and van der Waals interactions, giving rise to the lower membrane affinity (Goldston et al. 2014). NMR spectroscopy could be used to determine the transient nature of the interaction between a membrane mimetic and myristoylated HASPA. Does it shuttle between bound and unbound as expected? Comparison of the diffusion co-efficient of myristoylated HASPA with and without membrane mimetic using diffusion-ordered spectroscopy (DOSY) would begin to answer this question. The observation that myristoylated HASPA could interact with a membrane mimetic would lead to the next question. Does membrane association induce any conformational changes in HASPA? ^{1}H, ^{15}N HSQC spectra could then be recorded with the myristoylated protein with and without membrane mimetic to answer this question. Any spectral changes could be linked back to the backbone resonance assignment of HASPA to decipher the impact of membrane association. However, it would also be interesting to establish if HASPA association induced any changes in the membrane, such as those observed for the IDRs of Epsin 1 and AP180. (Busch et al. 2015).

This series of experiments is reliant on a long lasting association between the membrane and myristoylated HASPA, which is limited by the membrane binding affinity of myristate. Production of a dual acylated protein could be key to defining the HASP: membrane interaction. However, the palmitoyl acyl transferase (PAT) that catalyses

HASP palmitoylation *in vivo* is undefined. This is unsurprising when you consider that the *L. major* genome encodes 20 different DHHC domain PATs and that the study of PATs is still in its infancy (Goldston et al. 2014). PATs accept unmodified or modified proteins and in the case of modified proteins they distinguish between different lipid modifications. A well characterised PAT from a different species, such as Erf2 or Akr1 from yeast, would not be an effective substitute because of this highly discriminant substrate recognition. Therefore, an *in vitro* based method could be implemented to produce the dual acylated sample. A method was developed to covalently link saturated C_{16} alkyl groups to cysteine in peptides based on a known palmitoylated protein via a disulphide linkage, mimicking the palmitoylated state of the protein. (Wilkinson et al. 2000). It was shown that C_{16} alkyl group addition to the test peptide enabled membrane association analogous to that observed *in vivo*. Initially, it would be beneficial to determine if full length HASPA could be palmitoylated with this method after NMT mediated myristoylation. Alternatively, if the full length HASPs were destabilised by this method, production of a HASP based peptide that could be myristoylated and then chemically palmitoylated could be required.

6.2.2. Further NMR studies of the HASPs

NMR derived chemical shifts are highly sensitive probes of local protein conformation and have been used in some cases as the sole constraint to determine the structure of proteins (Shen et al. 2008). The relationship between these chemical shifts and the dihedral angles phi and psi is the basis of structural propensity prediction. Analysis of NMR derived chemical shift information for HASPA showed that the protein does not contain any elements of structural propensity. Transient structural elements termed molecular recognition features (MoRFs) are a distinctive characteristic of some IDPs that function as interaction hubs. The absence of transient structural elements in HASPA suggests that it does not function as an interaction hub, leading to the conclusion that the HASPs could function as entropic chains where the persistent disorder observed is functionally vital. It should be noted that preformed structural elements are not always required for biological partner binding. Disruption of residual helical structure in PUMA, an IDP of the BCL-2 apoptotic family of proteins, did not inhibit binding to its biological partner MCL-1 (Rogers et al. 2014), as detailed in Chapter 2.

To confirm entropic chain classification it would be useful to further experimentally validate the lack of transient structure within HASPA. Chemical shifts report local structure and other methods are required to establish long range contacts in an IDP. The current backbone assignment covers the Cα, Cβ, N and HN nuclei of HASPA. To produce more comprehensive chemical shift analysis it would be beneficial to record HNCO and HN(CA)CO experiments, in order to assign the carbonyl 13C’ and include these chemical shifts in the analysis. Long range and local structural features of IDPs in solution can be determined by measuring paramagnetic relaxation enhancement, residual dipolar couplings, NMR relaxation parameters and nuclear Overhauser effect, in particular homonuclear proton NOEs (Salmon et al. 2010). A combination of these methods would define and confirm any transient structured states of HASPA.

In particular, residual dipolar couplings (RDCs) provide structural propensity information in the context of the protein sequence (Mohana-Borges et al. 2004). In a published example the disordered C-terminus of nucleoprotein from Sendai virus that displays helical propensity was examined with this method (Jensen et al. 2008). Analysis of backbone RDCs showed that the protein populates three defined helical conformers, despite its conformational heterogeneity. It was then shown that conformational ensembles of the same protein derived from chemical shifts alone were consistent when cross validated with ensembles derived from RDCs (Jensen et al. 2010).

The NMR analysis conducted here examined HASPA and therefore no conclusions can be drawn on the presence of MoRFs in the repeats found in HASPB1 and HASPB2. It would be interesting to perform a similar NMR analysis of HASPB2 to determine if these repeats contained any structural propensity.

6.3. Concluding remarks

This thesis details the first in-depth characterisation of the intrinsically disordered nature of the HASPs from *Leishmania donovani*. Transient secondary structural elements are a common feature of IDPs, where transient order is a prerequisite of biological partner binding that resembles the final bound form. Analysis of NMR derived chemical shift information showed that they do not contain these short-lived structural elements, hinting that their inherent disorder has a functional role. Moreover, as the HASPs conserved in all human infective *Leishmania* species the role they perform must be significant. This finding should help to refocus efforts to define the biological role of the HASPs, away from identification of biological partner binding and towards a function enabled by disorder. The backbone resonance assignment of HASPA elucidated in this thesis lays the foundation for further in-depth biophysical studies of HASP disorder. Lastly, production of myristoylated HASPA is the first step to studying the membrane associated form of the HASPs, in either the myristoylated or dual acylated form.

Appendix

A.1. DNA sequences and plasmid maps

The sequences and plasmids are listed that relate to the lysozyme fusion work and constructs used in this thesis.

>T4 Lysozyme C54S C97A

TTATAGATTTTTATACGCGTCCCAAGTGCCAGTTCTAAACGTTGTAATGACTCGTTTTGCGCGA
TTAGGTGTTTGATTATACCATCTACTTTTAGCTAAGTTAACTGCTGCTTCATCCCAGCGTTTTTG
TTGAAGCATACGTAAAGAGTTAGTAAATCCTGCCACACCGGTTTCTCCCATTTGGAAAACCATA
TTAATCAATGCAGCGCGACGAACCGCATCAAGAGAATCATAAACCGGTTTTAATTTAGCATTTC
TCAGAATTCCGCGAACAGCAGCATCAACATCCTGATTAAAGAGTTTTTCAGCCTCATCTTTTGT
AATTACACCATTACTATTACGCCCAATAGCTTTATCTAATTCAGATTTAGCAGCATTAAGTGATG
GACTTTTTGTAAGCAAATGACCGATGCCAATAGTGTAATAGCCTTCTGTGTCTTTATAGATTTT
AAGTCTAAGACCTTCATCTATACGTAACATTTCAAATATATTCAT

> HASPA

CCATGGGAGCCTACTCTACGAAGGACTCCGCAAAGGAGCCCCAGAAGCGTGCTGATAACATC
GATACGACCACTCGAAGCGATGAGAAGGACGGCATCCATGTCCAGGAGAGCGCCGGTCCTGT
GCAGGAGAACTTTGGGGATGCGCAGGAGAAGAACGAAGATGGACACAACGTGGGGGATGG
AGCTAACGGCAATGAGGATGGTAACGATGATCAGCCGAAGGAGCACGCTGCCGGCAACCTC
GAGCACCACCACCACCACCACTGA

>HASPB1

CCATGGGAGCCTACTCTACGAAGGACTCCGCAAAGGAGCCCCAGAAGCGTGCTGATAACATC
CATAAAACCACTGAGGCCAATCACGGAGGCGCCACTGGTGTGCCCCCGAAGCACACCGGCAG
TGCGATGAACGACTCTGCCCCGAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGGC
CCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGGCCCGAAGGAGGACGGCCATA
CACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGA
TGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGCCCCTAAGGAGGACGGC

CGTACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACG

GCGATGCCCCTAAGGAGGACGGCCGTACACAGAAAAATGACGGCGATGGCCCTAAGGAGGA

CGGCCATACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAAAAAT

GACGGCGATGGCCCTAAGGAGGACGGCCGTACACAGAAAAATGACGGCGGTGGCCCTAAGG

AGGACGGCCATACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAA

AAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGCCCCT

AAGGAGGACGGCCGTACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACAC

AGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGC

CCCTAAGGAGGACGGCCGTACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCAT

ACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGGCCGTACACAGAAAAATGACGGCG

ATGGCCCTAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGGCCCTAAGGAGGACGG

CCGTACACAGAAAAATGACGGCGATGCCCCTAAGGAGGGTGAGAATCTGCAGCAAAACGAG

GGGATGCGCAGGAGAAGAACGAAGATGGACACAACGTGGGGGATGGAGCTAACGGCAATG

AGGATGGTAACGATGATCAGCCGAAGGAGCACGCTGCCGGCAACCTCGAGCACCACCACCAC

CACCACTGAG

>LmNMT

ATGTCTCGCAATCCATCGAACTCCGACGCTGCGCATGCGTTCTGGAGCACACAGCCCGTACCG
CAGACGGAAGATGAGACGGAGAAAATTGTGTTCGCTGGTCCGATGGACGAGCCAAAGACGG
TAGCCGATATTCCTGAGGAGCCGTACCCGATCGCCAGCACATTTGAGTGGTGGACGCCGAAC
ATGGAGGCGGCCGACGACATTCACGCAATTTACGAGCTTCTTCGGGATAACTACGTCGAAGAC
GACGACAGCATGTTTCGTTTCAACTACTCCGAGGAGTTTCTTCAGTGGGCACTATGCCCACCG
AACTACATCCCGGACTGGCACGTTGCAGTTCGCCGAAAGGCGGATAAGAAGCTGCTGGCCTT
CATTGCCGGCGTTCCCGTGACGTTGCGCATGGGCACTCCCAAGTACATGAAGGTGAAAGCAC
AGGAAAAGGGCGAAGGGGAGGAGGCGGCCAAGTATGATGAACCCCGTCACATCTGCGAAAT
CAACTTTCTCTGTGTCCACAAGCAACTCCGGGAGAAGCGGCTTGCCCCGATTTTGATCAAAGA
GGCGACGCGCCGCGTGAACCGCACCAACGTGTGGCAGGCGGTGTACACGGCTGGGGTGCTG
CTACCCACTCCGTATGCATCAGGGCAGTACTTCCACCGCAGCCTGAACCCCGAGAAGCTTGTG
GAGATCCGCTTCAGCGGCATTCCAGCACAGTACCAAAAGTTTCAGAACCCAATGGCGATGCTG
AAGCGCAACTACCAGCTGCCAAGCGCGCCGAAGAACTCTGGTCTTCGTGAAATGAAGCCGTCT
GACGTTCCGCAGGTGCGGCGGATTCTCATGAACTACCTGGACAGCTTCGATGTAGGTCCCGTC
TTTAGCGATGCCGAGATCAGCCACTACCTCCTTCCACGCGACGGTGTGGTCTTCACCTACGTG
GTGGAAAACGACAAGAAGGTGACCGACTTCTTTTCCTTCTATCGAATTCCGTCGACTGTGATT
GGGAACAGCAATTACAACCTTCTGAACGCAGCTTACGTTCACTACTATGCCGCGACGAGTATA
CCTTTGCATCAACTCATTCTCGACCTTTTGATCGTGGCGCATTCACGCGGCTTCGACGTGTGCA
ATATGGTAGAGATCCTCGACAACCGGTCTTTCGTTGAGCAGCTCAAGTTTGGCGCCGGCGACG
GTCATCTTCGATATTACTTCTACAACTGGGCGTATCCAAAGATCAAGCCTTCTCAGGTTGCCTT
GGTGATGCTGTAG

Name	Primer sequence 5’ to 3’	Template DNA
T4L-A-F	GCCGGCAACGCCGCCAATATATTTGAAATGTTACGTATAGATGAAGGTCTTAG	pHS1403
T4L-A-R	GTGGTGGTGCTCGAGTAGATTTTTATACGCGTCCCAAGTGCC	pHS1403
HASPA-F	CTCGAGCACCACCACCACCACCACTG	pET28b_haspa/ pET28b_haspb1
HASPA-R	GGCGGCGTTGCCGGCAGCGTGC	pET28b_haspa
T4L-Nt-F	GACTCTGCCGCCGCCAATATATTTGAAATGTTACGTATAGATGAAGGTCTTAG	pHS1403
Ntr-R	GGCGGCGGCAGAGTCGTTCATCGCAC	pET28b_haspb1
T4L-Ntr-F	GACGGCGATGGCGCCGCCAATATATTTGAAATGTTACGTATAGATGAAGGTCTTAG	pHS1403
Ntr-R	GCCATCGCCGTCATTTTTCTGTGTATGGCCGTCCTCCTTCGGGGCAGAGTC	pET28b_haspb1
Ntr-R-2	GAACGACTCTGCCCCGAAGGAGGACGGCCATACACAGAAAAATGACGGCGATGGCGCCG	pET28b_NtB1_T4L

Table A-1: List of primers for SLIC cloning of lysozyme fusion constructs.

Plasmid name	Protein	Affinity tag	Antibiotic resistance	Induction
pET28b_haspa	HASPA	C-terminal his-tag	Kanamycin	IPTG
pET28b_haspb1	HASPB1	C-terminal his-tag	Kanamycin	IPTG
pET28b_haspb2	HASPB2	C-terminal his-tag	Kanamycin	IPTG
pET-YSBLIC3C_Nt	HASPB1 N-terminal domain	3C cleavable N-terminal his-tag	Kanamycin	IPTG
pET-YSBLIC3C_Ntr	HASPB1 N-terminal domain plus repeat	3C cleavable N-terminal his-tag	Kanamycin	IPTG
pET-YSBLIC3C_Ct	HASPB1 C-terminal domain	3C cleavable N-terminal his-tag	Kanamycin	IPTG
pET-YSBLIC3C_Ctr	HASPB1 C-terminal domain plus repeat	3C cleavable N-terminal his-tag	Kanamycin	IPTG
pET15-MHL	LmNMT	TEV cleavable N-terminal his-tag	Ampicillin	IPTG
pHS1403	T4 lysozyme C54S C97A	No tag	Ampicillin	-
pET28b_NtB1_T4L	Nt T4L	C-terminal his-tag	Kanamycin	IPTG

Table A-2: List of vectors.

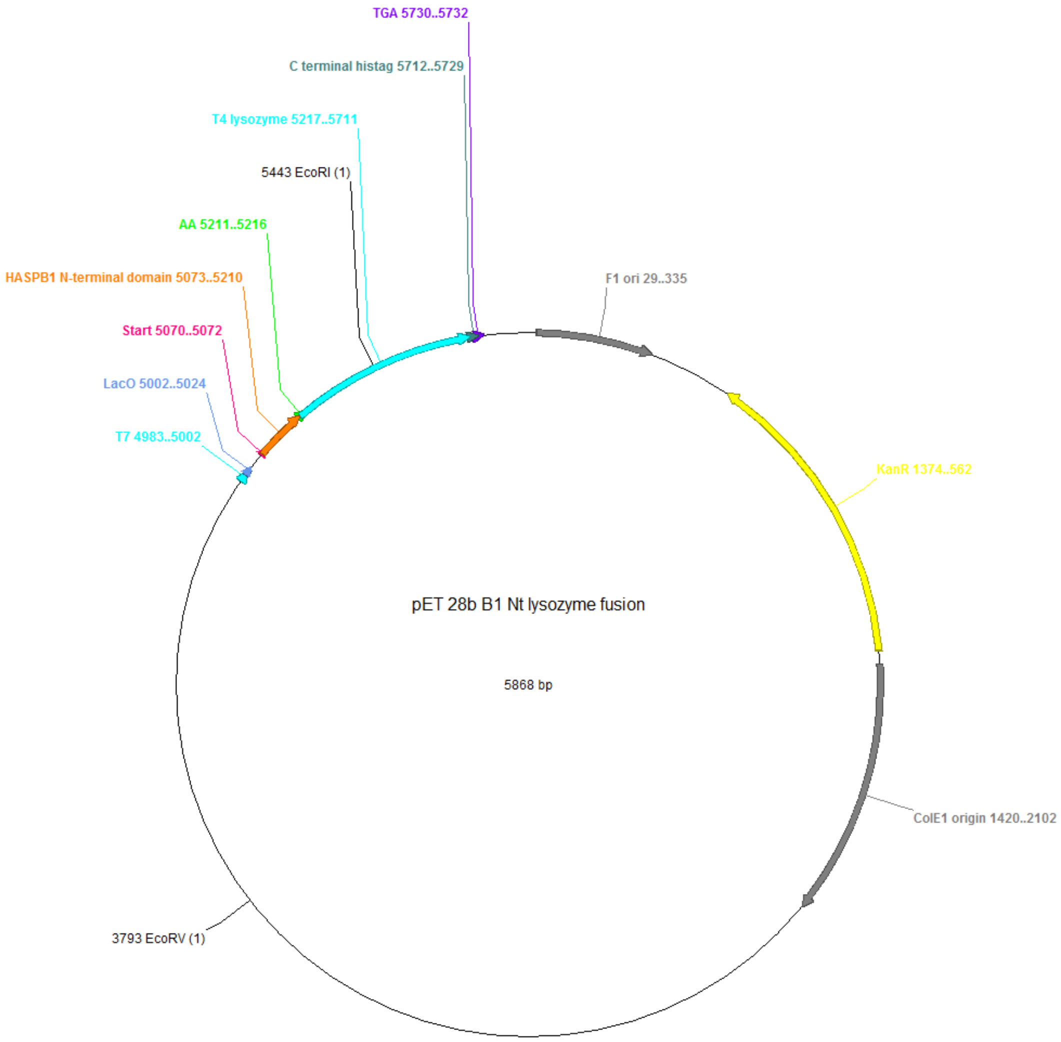


Figure A-1: Plasmid map of pET28b_NtB1_T4L.

A.2. Protein sequences

Full length protein sequences of constructs used during this project. In each case the hexa-histidine tag is underlined including any additional residues incorporated due to the vector.

>HASPA

(M)GAYSTKDSAKEPQKRADNIDTTTRSDEKDGIHVQESAGPVQENFGDAQEKNEDGHNVGDGA
NGNEDGNDDQPKEHAAGNLEHHHHHH

>HASPB1

(M)GAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTGSAMNDSAPKEDGHTQKNDGDG
PKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDAPKED
GRTQKNDGDGPKEDGHTQKNDGDAPKEDGRTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQ
KNDGDGPKEDGRTQKNDGGGPKEDGHTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDG
DAPKEDGRTQKNDGDGPKEDGHTQKNDGDGPKEDGHTQKNDGDAPKEDGRTQKNDGDGPKE
DGHTQKNDGDGPKEDGRTQKNDGDGPKEDGHTQKNDGDGPKEDGRTQKNDGDAPKEGENL
QQNDGDAQEKNEDGHNVGDGANGNEDGNDDQPKEHAAGNLEHHHHHH

>HASPB2

(M)GAYSTKDSAKEPQKRADNIHKTTEANHRGAAGVPPKHAGGAMNDSAPKEDGHTQKNDGD
GPKEDDHTQKNDGDGPKEDDHAHNDGGGPKEDENLQQNDGDAQEKNEDGHNVGDGANDNE
DGNDDQPKEHAAGNLEHHHHHH

> HASPB1 N-terminal fragment

MGSSHHHHHHSSGLEVLFQ/GPAMGAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTG
SAMNDSA

>HASPB1 N-terminal fragment plus one repeat

MGSSHHHHHHSSGLEVLFQ/GPAMGAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTG
SAMNDSAPKEDGHTQKNDGDG

>Nt-T4L

(M)GAYSTKDSAKEPQKRADNIHKTTEANHGGATGVPPKHTGSAMNDSAAANIFEMLRIDEGLRL
KIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNSNGVITKDEAEKLFNQDVDAAVRGILRNAK
LKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNR
AKRVITTFRTGTWDAYKNLLEHHHHHH

>T4 lysozyme C54S C97A

MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNSNGVITKDEAEKLFNQ
DVDAAVRGILRNAKLKPVYDSLDAVRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAV
NLAKSRWYNQTPNRAKRVITTFRTGTWDAYKNL

>LmNMT

MGSSHHHHHHSSGRENLYFQ/GPSNSDAAHAFWSTQPVPQTEDETEKIVFAGPMDEPKTVADIP
EEPYPIASTFEWWTPNMEAADDIHAIYELLRDNYVEDDDSMFRFNYSEEFLQWALCPPNYIPDW
HVAVRRKADKKLLAFIAGVPVTLRMGTPKYMKVKAQEKGEGEEAAKYDEPRHICEINFLCVHKQL
REKRLAPILIKEATRRVNRTNVWQAVYTAGVLLPTPYASGQYFHRSLNPEKLVEIRFSGIPAQYQKF
QNPMAMLKRNYQLPSAPKNSGLREMKPSDVPQVRRILMNYLDSFDVGPVFSDAEISHYLLPRDG
VVFTYVVENDKKVTDFFSFYRIPSTVIGNSNYNLLNAAYVHYYAATSIPLHQLILDLLIVAHSRGFDVC
NMVEILDNRSFVEQLKFGAGDGHLRYYFYNWAYPKIKPSQVALVML

Protein/ construct	ProtParam parameters
HASPB1 with C-terminal his-tag	Number of amino acids: 409 Molecular weight: 43437.1 Theoretical pI: 4.78
HASPB2 with C-terminal his-tag	Number of amino acids: 142 Molecular weight: 15215.5 Theoretical pI: 4.96
HASPA with C-terminal his-tag	Number of amino acids: 88 Molecular weight: 9576.8 Theoretical pI: 4.91
HASPB1 Nt fragment no his-tag	Number of amino acids: 50 Molecular weight: 5091.5 Theoretical pI: 8.36
HASPB1 Ntr fragment no his-tag	Number of amino acids: 64 Molecular weight: 6571.0 Theoretical pI: 6.44
HASPB1 Nt fragment plus his-tag	Number of amino acids: 69 Molecular weight: 7237.8 Theoretical pI: 7.14
HASPB1 Ntr fragment plus his-tag	Number of amino acids: 83 Molecular weight: 8717.3 Theoretical pI: 6.53
HASPB1 Ctr fragment plus his-tag	Number of amino acids: 69 Molecular weight: 7332.5 Theoretical pI: 4.77
HASPB1 Ctr fragment no his-tag	Number of amino acids: 50 Molecular weight: 5186.1 Theoretical pI: 4.05
Nt-T4L fusion	Number of amino acids: 220 Molecular weight: 24511.6 Theoretical pI: 9.59
Leishmania major NMT	Number of amino acids: 438 Molecular weight: 50450.4 Theoretical pI: 6.28

Table A-3: ProtParam generated biochemical parameters for each construct used in this project (Gasteiger et al. 2005). The molecular weight of each construct is reported in Daltons (Da).

A.3. NMR constraints

The following chemical shifts for H^N, N, Cα and Cβ nuclei were obtained from the assignment of HASPA.

Residue	H^N (ppm)	N (ppm)	Cα (ppm)	Cβ (ppm)
1Gly	None	None	None	None
2Ala	8.26	123.65	52.6	19.12
3Tyr	8.17	118.98	57.75	38.64
4Ser	8.15	117.25	None	None
5Thr	None	None	None	None
6Lys	8.31	123.16	56.8	32.91
7Asp	8.32	121.32	54.63	41.22
8Ser	8.22	116.6	58.79	63.79
9Ala	8.27	125.46	52.76	19.17
10Lys	8.12	119.91	56.09	33.14
11Glu	8.27	122.93	54.49	29.69
12Pro	None	None	None	None
13Gln	8.5	120.59	55.89	29.54
14Lys	8.38	123.38	56.22	33.13
15Arg	8.44	123.44	55.7	31.16
16Ala	8.49	125.86	52.59	19.34
17Asp	8.3	119.07	54.32	41.15
18Asn	8.31	118.25	53.35	38.89
19Ile	8.06	120.68	61.29	38.54
20Asp	8.46	124.32	54.26	41.21
21Thr	8.26	114.94	62.05	69.36
22Thr	8.31	116.24	62.71	69.75
23Thr	8.17	117.05	62.13	69.29
24Arg	8.42	124.13	56	31.05
25Ser	8.51	117.7	58.5	63.85
26Asp	8.48	122.34	54.45	41.14
27Glu	8.31	121.04	57.05	30.09
28Lys	8.36	121.62	56.69	32.82
29Asp	8.32	120.58	54.71	41.21
30Gly	8.24	108.67	45.46	None
31Ile	7.92	119.67	61.19	38.56
32His	8.6	123.46	55.4	29.49
33Val	8.21	122.7	62.39	32.85
34Gln	8.52	124.42	55.87	29.52
35Glu	8.59	123.46	56.77	30.36
36Ser	8.41	117.04	58.25	63.91
37Ala	8.42	126.41	52.51	19.53
38Gly	8.18	108.38	44.45	None
39Pro	None	None	None	None
40Val	8.3	120.87	62.43	32.72

41Gln	8.48	124.7	55.65	29.63
42Glu	8.48	123.05	56.57	30.5
43Asn	8.42	119.7	52.98	39.07
44Phe	8.36	121.46	58.02	39.54
45Gly	8.38	110.24	45.43	None
46Asp	8.2	120.69	54.42	41.32
47Ala	8.28	124.19	52.82	19.12
48Gln	8.32	119.14	56.03	29.47
49Glu	8.41	122.11	56.79	30.14
50Lys	8.37	122.37	56.14	33.14
51Asn	8.51	120.45	53.29	39.12
52Glu	8.57	121.87	56.69	30.2
53Asp	8.36	121.03	54.54	41.23
54Gly	8.46	109.49	45.57	None
55His	8.29	118.3	55.69	29.25
56Asn	8.58	120.55	53.3	38.97
57Val	8.22	120.29	62.66	32.58
58Gly	8.49	112.14	45.34	None
59Asp	8.25	120.7	54.51	41.23
60Gly	8.33	109.11	45.57	None
61Ala	8.18	123.7	52.67	19.21
62Asn	8.49	117.77	53.35	39.04
63Gly	8.4	109.25	45.56	None
64Asn	8.36	118.66	53.32	39.07
65Glu	8.59	121.43	56.98	29.99
66Asp	8.37	120.82	54.61	41.2
67Gly	8.35	108.98	45.61	None
68Asn	8.31	118.65	53.34	39.23
69Asp	8.42	120.85	54.56	41.11
70Asp	8.31	120.22	54.09	40.62
71Gln	8.12	120.68	53.86	28.89
72Pro	None	None	None	None
73Lys	8.48	121.82	56.42	33.06
74Glu	None	None	None	None
75His	None	None	None	None
76Ala	None	None	None	None
77Ala	8.36	123.35	None	None
78Gly	8.35	107.73	45.45	None
79Asn	8.27	118.58	53.38	38.8
80Leu	8.26	122.26	55.49	42.13
81Glu	8.28	120.58	56.85	30.04
82His	None	None	None	None
83His	None	None	None	None
84His	None	None	None	None
85His	None	None	None	None
86His	None	None	None	None

87His	None	None	None	None

Table A-4: Chemical shift data for assigned resonances in HASPA.

The following values were used to calculate the intensity and chemical shift changes on HASPA myristoylation by NMT.

Peak	I_0 (height)	I_{myr}	ΔI	ΔH (ppm)	ΔN (ppm)	Δδ (ppm)
6LysH	743849	386643	0.480213	0.000102	0.000115	0.014732
8SerH	618530	0	1	67.46963	321.1428	
9AlaH	754817	688238	0.088205	1.07E-05	5.04E-05	0.007815
10LysH	1030150	606489	0.411261	4E-06	1.12E-05	0.003898
11GluH	1169800	970650	0.170243	7.69E-05	0.000173	0.015816
14LysH	929337	976994	-0.05128	8.28E-07	0.000151	0.012335
15ArgH	644528	574654	0.108411	1.86E-05	8.42E-05	0.010137
16AlaH	873003	1022250	-0.17096	1.06E-06	2.18E-05	0.004786
17AspH	731349	783779	-0.07169	7.02E-06	4.19E-06	0.003348
19IleH	1227560	1119100	0.088354	2.4E-05	1.19E-05	0.005991
20AspH	1119640	1483890	-0.32533	1.07E-05	6.67E-05	0.008795
21ThrH	851778	938530	-0.10185	5.66E-06	5.17E-06	0.003292
22ThrH	865065	789377	0.087494	1.87E-05	8.12E-06	0.005175
23ThrH	817048	741406	0.09258	2.52E-05	5.59E-05	0.009007
24ArgH	596510	644781	-0.08092	1.79E-05	8.51E-05	0.010148
25SerH	430485	492278	-0.14354	2.76E-06	2.62E-06	0.002318
26AspH	914488	670280	0.267043	7.67E-06	7.23E-06	0.003861
27GluH	1029400	902156	0.12361	5.2E-06	7.86E-06	0.003613
28LysH	923335	777601	0.157834	6.25E-06	4.18E-09	0.002501
29AspH	807472	766854	0.050303	4.84E-08	7.73E-05	0.008797
31IleH	1525160	1395370	0.085099	1.08E-05	7.32E-06	0.00426
34GlnH	1271140	1291570	-0.01607	3.8E-06	3E-06	0.002608
35GluH	1194960	1086710	0.090589	5.52E-06	1.16E-07	0.002374
36SerH	844014	801367	0.050529	5.11E-06	8.81E-06	0.00373
37AlaH	1198100	1173440	0.020583	3.6E-07	2.96E-06	0.001823
40ValH	1508560	1350420	0.104828	7.29E-08	1.76E-07	0.000499
41GlnH	1435900	1185200	0.174594	1.16E-05	3.81E-06	0.00392
42GluH	1129700	947310	0.16145	1.52E-07	1.99E-06	0.001464
43AsnH	548549	568389	-0.03617	6.05E-06	7.36E-06	0.003662
44PheH	1038120	700622	0.325105	2.4E-07	4.06E-08	0.00053

46AspH	1197550	1076350	0.101207	1.19E-05	3.59E-05	0.006913
47AlaH	1394010	1440220	-0.03315	3.6E-07	4.4E-06	0.002181
48GlnH	1521570	1497320	0.015937	3.18E-05	1.38E-05	0.006751
49GluH	1309310	1342600	-0.02543	3.6E-07	2.97E-05	0.005487
50LysH	1552800	1088760	0.298841	8.41E-08	9.72E-06	0.00313
52GluH	712732	839873	-0.17839	2.12E-07	4.67E-06	0.002208
53AspH	1150850	1139740	0.009654	0	7.74E-06	0.002782
57ValH	1605070	1473790	0.081791	1.19E-06	7.39E-05	0.008663
59AspH	1205190	1229190	-0.01991	2.12E-07	8.18E-07	0.001015
61AlaH	1402050	1168810	0.166356	3.17E-06	1.49E-06	0.002158
62AsnH	743728	916064	-0.23172	2.5E-06	6.05E-06	0.002924
64AsnH	829973	797962	0.038569	1.79E-05	2.21E-06	0.004483
65GluH	961287	803333	0.164315	1.61E-06	5.27E-06	0.002623
66AspH	1104530	1124520	-0.0181	8.18E-06	3.45E-06	0.00341
68AsnH	796276	866312	-0.08795	2.89E-08	1.95E-07	0.000473
69AspH	601870	789606	-0.31192	6.08E-07	4.35E-06	0.002227
70AspH	779051	899242	-0.15428	1.3E-07	2.55E-06	0.001635
71GlnH	1330830	1302230	0.02149	1.25E-06	5.23E-07	0.001333
73LysH	1145900	1078350	0.058949	4.1E-07	1.11E-06	0.001233
80LeuH	951004	856848	0.099007	1E-08	1.08E-06	0.001042

Table A-5: Table of values used for chemical shift and intensity change calculations in Chapter 3.

Abbreviations

AIDs	Acquired Immunodeficiency Syndrome
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CL	Cutaneous Leishmaniasis
CoA	Co-Enzyme A
CPM	7-Diethylamino-3-(4-Maleimidophenyl)-4-Methylcoumarin
Da	Dalton
DALYs	Disability Adjusted Life Years
DCL	Diffuse Cutaneous Leishmaniasis
DDT	Dichlorodiphenyltrichloroethane
DLS	Dynamic Light Scattering
DMSO	Dimethyl Sulfoxide
DOSY	Diffusion-Ordered Spectroscopy
DTT	Dithiothreitol
E. coli	*Escherichia Coli*
ER	Endoplasmic Reticulum
ESI-MS	Electrospray Ionisation Mass Spectrometry
Fab	Fragment Antigen-Binding
Fc	Fragment Crystallisable
G protein	Heterotrimeric Guanine Nucleotide Binding Protein
GPCR	G Protein Coupled Receptor
GRAVY	Grand Average Of Hydropathicity
HAART	Highly Active Antiretroviral Therapy
HASP	Hydrophilic Acylated Surface Protein
HEWL	Hen Egg White Lysozyme
HIV	Human Immunodeficiency Virus
HPLC	High-Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Correlation
IC_{50}	Half Maximal Inhibitory Concentration
ICL3	Third Intracellular Loop
IDR	Intrinsically Disordered Region
IgG	Immunoglobulin G
INEPT	Insensitive Nuclei Enhanced By Polarization Transfer
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ITN	Insecticide Treated Net
L. braziliensis	*Leishmania Braziliensis*
L. chagasi	*Leishmania Chagasi*
L. donovani	*Leishmania Donovani*
L. infantum	*Leishmania Infantum*
L. major	*Leishmania Major*
L. mexicana	*Leishmania Mexicana*
L. tropica	*Leishmania Tropica*
LB	Lysogeny Broth
LdNMT	*Leishmania Donovani* NMT
LmNMT	*Leishmania Major* NMT

LPG	Lipophosphoglycan
MBCL	Methylbenzethonium Chloride
MBOAT	Membrane Bound O-Acyltransferase
MBP	Maltose Binding Protein
MCL	Mucocutaneous Leishmaniasis
MME	Monomethyl Ether
MoRF	Molecular Recognition Feature
MW	Molecular Weight
MWCO	Molecular Weight Cut-Off
NCBI	National Centre For Biotechnology Information
ncSPC	Neighbour Corrected Structural Propensity Calculator
NEB	New England Bioscience
NMT	N-Myristoyltransferase
NMW	2-Oxopentadecyl-CoA
Nt T4L	HASPB1 N-Terminal Domain: T4 Lysozyme Chimera
NTDs	Neglected Tropical Diseases
OD	Optical Density
OD_{600}	Optical Density at 600 nm
OPA	o-Phthaldialdehyde
PAT	Palmitoyl Acyltransferases
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PHF	Paired Helical Filament
pI	Isoelectric Point
PKDL	Post Kala-Azar Dermal Leishmaniasis
PMSF	Phenylmethanesulfonyl Fluoride
PPII	Polyproline II
PYP	Photoactive Yellow Protein
R&D	Research and Development
RDC	Residual Dipolar Coupling
RMSD	Root Mean Squared Deviation
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC-MALLS	Size Exclusion Chromatography Multi-Angle Laser Light Scatter
SLIC	Sequence And Ligation Independent Cloning
SSP	Secondary Structure Prediction
SV	Stomodeal Valve
T4L	T4 Lysozyme
VL	Visceral Leishmaniasis
WHO	World Health Organisation
β2AR	β2 Adrenergic Receptor

References

Aicart-Ramos C, Valero RA, Rodriguez-Crespo I (2011) Protein palmitoylation and subcellular trafficking. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808:** 2981-2994

Alce TM, Gokool S, McGhie D, Stager S, Smith DF (1999) Expression of hydrophilic surface proteins in infective stages of Leishmania donovani. *Molecular and biochemical parasitology* **102:** 191-196

Alexander B, Maroli M (2003) Control of phlebotomine sandflies. *Medical and veterinary entomology* **17:** 1-18

Alvar J, Aparicio P, Aseffa A, Den Boer M, Canavate C, Dedet JP, Gradoni L, Ter Horst R, Lopez-Velez R, Moreno J (2008) The relationship between leishmaniasis and AIDS: the second 10 years. *Clinical microbiology reviews* **21:** 334-359, table of contents

Amata I, Maffei M, Igea A, Gay M, Vilaseca M, Nebreda AR, Pons M (2013) Multi-phosphorylation of the Intrinsically Disordered Unique Domain of c-Src Studied by In-Cell and Real-Time NMR Spectroscopy. *ChemBioChem* **14:** 1820-1827

Ames JB, Tanaka T, Stryer L, Ikura M (1996) Portrait of a myristoyl switch protein. *Current Opinion in Structural Biology* **6:** 432-438

Andrade MA, Chacón P, Merelo JJ, Morán F (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein engineering* **6:** 383-390

Ashford DA, David JR, Freire M, David R, Sherlock I, Eulalio MC, Sampaio DP, Badaro R (1998) Studies on control of visceral leishmaniasis: impact of dog control on canine and human visceral leishmaniasis in Jacobina, Bahia, Brazil. *The American journal of tropical medicine and hygiene* **59:** 53-57

Babu MM, van der Lee R, de Groot NS, Gsponer J (2011) Intrinsically disordered proteins: regulation and disease. *Current Opinion in Structural Biology* **21:** 432-440

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R (2011) The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50:** 4402-4410

Barral A, Pedral-Sampaio D, Grimaldi G, Momen H, McMahon-Pratt D, de Jesus AR, Almeida R, Badaro R, Barral-Netto M, Carvalho EM, Johnson WD (1991) Leishmaniasis in Bahia, Brazil: Evidence that Leishmania amazonensis Produces a Wide Spectrum of Clinical Disease. *The American journal of tropical medicine and hygiene* **44:** 536-546

Bartels DJ, Mitchell DA, Dong X, Deschenes RJ (1999) Erf2, a Novel Gene Product That Affects the Localization and Palmitoylation of Ras2 in Saccharomyces cerevisiae. *Molecular and Cellular Biology* **19:** 6775-6787

Bates PA (2008) Leishmania sand fly interaction: progress and challenges. *Current opinion in microbiology* **11:** 340-344

Benson JR, Hare PE (1975) O-phthalaldehyde: fluorogenic detection of primary amines in the picomole range. Comparison with fluorescamine and ninhydrin. *Proceedings of the National Academy of Sciences of the United States of America* **72:** 619-622

Bhatnagar RS, Futterer K, Farazi TA, Korolev S, Murray CL, Jackson-Machelski E, Gokel GW, Gordon JI, Waksman G (1998) Structure of N-myristoyltransferase with bound myristoylCoA and peptide substrate analogs. *Nat Struct Mol Biol* **5:** 1091-1097

Bhatnagar RS, Jackson-Machelski E, McWherter CA, Gordon JI (1994) Isothermal titration calorimetric studies of Saccharomyces cerevisiae myristoyl-CoA:protein N-myristoyltransferase. Determinants of binding energy and catalytic discrimination among acyl-CoA and peptide ligands. *Journal of Biological Chemistry* **269:** 11045-11053

Bieri M, Kwan AH, Mobli M, King GF, Mackay JP, Gooley PR (2011) Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS Journal* **278:** 704-715

Blanc M, Coetzer TL, Blackledge M, Haertlein M, Mitchell EP, Forsyth VT, Jensen MR (2014) Intrinsic disorder within the erythrocyte binding-like proteins from Plasmodium falciparum. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1844:** 2306-2314

Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69:** 185-189

Bonsor D, Butz SF, Solomons J, Grant S, Fairlamb IJS, Fogg MJ, Grogan G (2006) Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes. *Organic & Biomolecular Chemistry* **4:** 1252-1260

Brannigan JA, Roberts SM, Bell AS, Hutton JA, Hodgkinson MR, Tate EW, Leatherbarrow RJ, Smith DF, Wilkinson AJ (2014) Diverse modes of binding in structures of Leishmania major N-myristoyltransferase with selective inhibitors. *IUCrJ* **1:** 250-260

Brannigan JA, Smith BA, Yu Z, Brzozowski AM, Hodgkinson MR, Maroof A, Price HP, Meier F, Leatherbarrow RJ, Tate EW, Smith DF, Wilkinson AJ (2010) N-myristoyltransferase from Leishmania donovani: structural and functional characterisation of a potential drug target for visceral leishmaniasis. *Journal of molecular biology* **396:** 985-999

Brzozowski AM, Walton J (2001) Clear strategy screens for macromolecular crystallization. *Journal of Applied Crystallography* **34:** 97-101

Bukowska MA, Grütter MG (2013) New concepts and aids to facilitate crystallization. *Current Opinion in Structural Biology* **23:** 409-416

Busch DJ, Houser JR, Hayden CC, Sherman MB, Lafer EM, Stachowiak JC (2015) Intrinsically disordered proteins drive membrane curvature. *Nat Commun* **6**

Carpenter EP, Beis K, Cameron AD, Iwata S (2008) Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology* **18:** 581-586

Cavanagh J, Fairbrother WJ, Palmer Iii AG, Rance M, Skelton NJ (2007) Chapter 2 - Theoretical description of NMR spectroscopy. In *Protein NMR Spectroscopy (Second Edition)*, Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (eds), pp 29-113. Burlington: Academic Press

Cecílio P, Perez-Cabezas B, Santarém N, Maciel J, Rodrigues V, Cordeiro-Da-Silva A (2014) Deception and Manipulation: the arms of Leishmania, a successful parasite. *Frontiers in Immunology* **5:** 1-16

Chang CY, Sun J, Chang T-Y (2011) Membrane-bound O-acyltransferases (MBOATs). *Front Biol* **6:** 177-182

Chappuis F, Sundar S, Hailu A, Ghalib H, Rijal S, Peeling RW, Alvar J, Boelaert M (2007) Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? *Nature reviews Microbiology* **5:** 873-882

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318:** 1258-1265

Chien EY, Liu W, Zhao Q, Katritch V, Han GW, Hanson MA, Shi L, Newman AH, Javitch JA, Cherezov V, Stevens RC (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330:** 1091-1095

Chun E, Thompson Aaron A, Liu W, Roth Christopher B, Griffith Mark T, Katritch V, Kunken J, Xu F, Cherezov V, Hanson Michael A, Stevens Raymond C (2012) Fusion Partner Toolchest for the Stabilization and Crystallization of G Protein-Coupled Receptors. *Structure* **20:** 967-976

Chung CC, Ohwaki K, Schneeweis JE, Stec E, Varnerin JP, Goudreau PN, Chang A, Cassaday J, Yang L, Yamakawa T, Kornienko O, Hodder P, Inglese J, Ferrer M, Strulovici B, Kusunoki J, Tota MR, Takagi T (2008) A fluorescence-based thiol quantification assay for ultra-high-throughput screening for inhibitors of coenzyme A production. *Assay Drug Dev Technol* **6:** 361-374

Copeland RA (2013) Enzyme Reaction Mechanisms. In *Evaluation of Enzyme Inhibitors in Drug Discovery*, 2nd Edition edn, 2, pp 25-55. John Wiley & Sons, Inc.

Corrales RM, Sereno D, Mathieu-Daude F (2010) Deciphering the Leishmania exoproteome: what we know and what we can learn. *FEMS immunology and medical microbiology* **58:** 27-38

Corvi MM, Berthiaume LG, De Napoli MG (2011) Protein palmitoylation in protozoan parasites. *Frontiers in bioscience (Scholar edition)* **3:** 1067-1079

Coulter A, Harris R (1983) Simplified preparation of rabbit fab fragments. *Journal of Immunological Methods* **59:** 199-203

Denny PW, Gokool S, Russell DG, Field MC, Smith DF (2000) Acylation-dependent protein export in Leishmania. *The Journal of biological chemistry* **275:** 11017-11025

Depledge DP, MacLean LM, Hodgkinson MR, Smith BA, Jackson AP, Ma S, Uliana SR, Smith DF (2010) Leishmania-specific surface antigens show sub-genus sequence variation and immune recognition. *PLoS neglected tropical diseases* **4:** e829

Derewenda ZS (2010) Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallographica Section D: Biological Crystallography* **66:** 604-615

Dietze R, Barros GB, Teixeira L, Harris J, Michelson K, Falqueto A, Corey R (1997) Effect of eliminating seropositive canines on the transmission of visceral leishmaniasis in Brazil. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **25:** 1240-1242

Dietze R, Carvalho SF, Valli LC, Berman J, Brewer T, Milhous W, Sanchez J, Schuster B, Grogl M (2001) Phase 2 trial of WR6026, an orally administered 8-aminoquinoline, in the treatment of visceral leishmaniasis caused by Leishmania chagasi. *The American journal of tropical medicine and hygiene* **65:** 685-689

Dostalova A, Volf P (2012) Leishmania development in sand flies: parasite-vector interactions overview. *Parasites & Vectors* **5:** 276

Dosztányi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21:** 3433-3434

Dosztányi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21:** 3433-3434

Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25:** 2745-2746

Duronio RJ, Towler DA, Heuckeroth RO, Gordon JI (1989) Disruption of the yeast N-myristoyl transferase gene causes recessive lethality. *Science* **243:** 796-800

Dyson HJ, Wright PE (2004) Unfolded Proteins and Protein Folding Studied by NMR. *Chemical Reviews* **104:** 3607-3622

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66:** 486-501

Engel CK, Chen L, Privé GG (2002) Insertion of carrier proteins into hydrophilic loops of the Escherichia coli lactose permease. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1564:** 38-46

Engels D, Savioli L (2006) Reconsidering the underestimated burden caused by neglected tropical diseases. *Trends in parasitology* **22:** 363-366

Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69:** 1204-1214

Farazi TA, Manchester JK, Gordon JI (2000) Transient-State Kinetic Analysis of Saccharomyces cerevisiae MyristoylCoA:Protein N-Myristoyltransferase Reveals that a Step after Chemical Transformation Is Rate Limiting†. *Biochemistry* **39:** 15807-15816

Farazi TA, Manchester JK, Waksman G, Gordon JI (2001b) Pre-Steady-State Kinetic Studies of Saccharomyces cerevisiae MyristoylCoA:Protein N-Myristoyltransferase Mutants Identify Residues Involved in Catalysis†. *Biochemistry* **40:** 9177-9186

Farazi TA, Waksman G, Gordon JI (2001a) Structures of Saccharomyces cerevisiae N-myristoyltransferase with Bound MyristoylCoA and Peptide Provide Insights about Substrate Recognition and Catalysis†. *Biochemistry* **40:** 6335-6343

Farid RS. (2006) Circular Dichroism (CD) Spectroscopy. Vol. 2015.

Field MC, Natesan SK, Gabernet-Castello C, Koumandou VL (2007) Intracellular trafficking in the trypanosomatids. *Traffic* **8:** 629-639

Flinn HM, Rangarajan D, Smith DF (1994) Expression of a hydrophilic surface protein in infective stages of Leishmania major. *Molecular and biochemical parasitology* **65:** 259-270

Flinn HM, Smith DF (1992) Genomic organisation and expression of a differentially-regulated gene family from Leishmania major. *Nucleic acids research* **20:** 755-762

Foucault M, Mayol K, Receveur-Bréchot V, Bussat M-C, Klinguer-Hamour C, Verrier B, Beck A, Haser R, Gouet P, Guillon C (2010) UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. *Proteins: Structure, Function, and Bioinformatics* **78:** 1441-1456

Frénal K, Tay CL, Mueller C, Bushell ES, Jia Y, Graindorge A, Billker O, Rayner JC, Soldati-Favre D (2013) Global Analysis of Apicomplexan Protein S-Acyl Transferases Reveals an Enzyme Essential for Invasion. *Traffic* **14:** 895-911

Frezard F, Demicheli C, Ribeiro RR (2009) Pentavalent antimonials: new perspectives for old drugs. *Molecules* **14:** 2317-2336

Garcia-Hernandez R, Manzano JI, Castanys S, Gamarro F (2012) Leishmania donovani develops resistance to drug combinations. *PLoS neglected tropical diseases* **6:** e1974

Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins M, Appel R, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker J (ed), 52, pp 571-607. Humana Press

Gavgani AS, Hodjati MH, Mohite H, Davies CR (2002) Effect of insecticide-impregnated dog collars on incidence of zoonotic visceral leishmaniasis in Iranian children: a matched-cluster randomised trial. *Lancet* **360:** 374-379

Gibbs EB, Showalter SA (2015) Quantitative Biophysical Characterization of Intrinsically Disordered Proteins. *Biochemistry* **54:** 1314-1326

Goding JW (1978) Use of staphylococcal protein A as an immunological reagent. *J Immunol Methods* **20:** 241-253

Goldston AM, Sharma AI, Paul KS, Engman DM (2014) Acylation in trypanosomatids: an essential process and potential drug target. *Trends in parasitology* **30:** 350-360

Goncalves V, Brannigan JA, Thinon E, Olaleye TO, Serwa R, Lanzarone S, Wilkinson AJ, Tate EW, Leatherbarrow RJ (2012) A fluorescence-based assay for N-myristoyltransferase activity. *Analytical Biochemistry* **421:** 342-344

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic acids research* **38:** W695-699

Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK (2012) Structure of the δ-opioid receptor bound to naltrindole. *Nature* **485:** 400-404

Grzesiek S, Bax A (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114:** 6291-6293

Grzesiek S, Bax A (1992b) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99:** 201-207

Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482:** 547-551

Hanson MA, Roth CB, Jo E, Griffith MT, Scott FL, Reinhart G, Desale H, Clemons B, Cahalan SM, Schuerer SC, Sanna MG, Han GW, Kuhn P, Rosen H, Stevens RC (2012) Crystal Structure of a Lipid G Protein–Coupled Receptor. *Science* **335:** 851-855

Higman V. (2012) Protein NMR: A Practical Guide http://www.protein-nmr.org.uk, Vol. 2015.

Hiroaki H, Umetsu Y, Nabeshima Y-i, Hoshi M, Kohda D (2011) A simplified recipe for assigning amide NMR signals using combinatorial 14N amino acid inverse-labeling. *J Struct Funct Genomics* **12:** 167-174

Hotez PJ, Fenwick A, Savioli L, Molyneux DH (2009) Rescuing the bottom billion through control of neglected tropical diseases. *Lancet* **373:** 1570-1575

Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE, Sachs JD, Savioli L (2007) Control of Neglected Tropical Diseases. *New England Journal of Medicine* **357:** 1018-1027

Hu L-L, Wan S-B, Niu S, Shi X-H, Li H-P, Cai Y-D, Chou K-C (2011) Prediction and analysis of protein palmitoylation sites. *Biochimie* **93:** 489-496

Hunte C, Michel H (2002) Crystallisation of membrane proteins mediated by antibody fragments. *Current Opinion in Structural Biology* **12:** 503-508

Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, IJzerman AP, Stevens RC (2008) The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **322:** 1211-1217

James TL (1998) Fundamentals of NMR. Nuclear Magnetic Resonance. In *Biophysics Textbook Online (BTOL)*, 1, pp 1-31. Bethesda: Biophysical Society

Jeddi F, Piarroux R, Mary C (2011) Antimony resistance in leishmania, focusing on experimental research. *Journal of tropical medicine* **2011:** 695382

Jensen AT, Gasim S, Moller T, Ismail A, Gaafar A, Kemp M, el Hassan AM, Kharazmi A, Alce TM, Smith DF, Theander TG (1999) Serodiagnosis of Leishmania donovani infections: assessment of enzyme-linked immunosorbent assays using recombinant L. donovani gene B protein (GBP) and a peptide sequence of L. donovani GBP. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93:** 157-160

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RWH, Blackledge M (2008) Quantitative Conformational Analysis of Partially Folded Proteins from Residual Dipolar Couplings: Application to the Molecular Recognition Element of Sendai Virus Nucleoprotein. *Journal of the American Chemical Society* **130:** 8055-8061

Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *Journal of the American Chemical Society* **132:** 1270-1272

Jha TK, Olliaro P, Thakur CP, Kanyok TP, Singhania BL, Singh IJ, Singh NK, Akhoury S, Jha S (1998) Randomised controlled trial of aminosidine (paromomycin) v sodium stibogluconate for treating visceral leishmaniasis in North Bihar, India. *BMJ (Clinical research ed)* **316:** 1200-1205

Johnson DR, Bhatnagar RS, Knoll LJ, Gordon JI (1994) Genetic and biochemical studies of protein N-myristoylation. *Annual review of biochemistry* **63:** 869-914

Kabsch W (2010) XDS. *Acta Crystallographica Section D: Biological Crystallography* **66:** 125-132

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89:** 496-514

Ke A, Wolberger C (2003) Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein science* **12:** 306-312

Khamesipour A, Abbasi A, Firooz A, Mohammadi AM, Eskandari SE, Jaafari MR (2012) Treatment of cutaneous lesion of 20 years' duration caused by leishmanization. *Indian journal of dermatology* **57:** 123-125

Kim DH, Chung HJ, Bleys J, Ghohestani RF (2009) Is paromomycin an effective and safe treatment against cutaneous leishmaniasis? A meta-analysis of 14 randomized controlled trials. *PLoS neglected tropical diseases* **3:** e381

Knuepfer E, Stierhof YD, McKean PG, Smith DF (2001) Characterization of a differentially expressed protein that shows an unusual localization to intracellular membranes in Leishmania major. *The Biochemical journal* **356:** 335-344

Kobe B, Center RJ, Kemp BE, Poumbourios P (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 4319-4324

Kobilka B, Schertler GFX (2008) New G-protein-coupled receptor crystal structures: insights and limitations. *Trends in Pharmacological Sciences* **29:** 79-83

Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance* **241:** 74-85

Kostiuk MA, Corvi MM, Keller BO, Plummer G, Prescher JA, Hangauer MJ, Bertozzi CR, Rajaiah G, Falck JR, Berthiaume LG (2008) Identification of palmitoylated mitochondrial proteins using a bio-orthogonal azido-palmitate analogue. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **22:** 721-732

Kragelj J, Ozenne V, Blackledge M, Jensen MR (2013) Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts. *ChemPhysChem* **14:** 3034-3045

Krishnamurthy H, Gouaux E (2012) X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **481:** 469-474

Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae

PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **482:** 552-556

Kuhlencord A, Maniera T, Eibl H, Unger C (1992) Hexadecylphosphocholine: oral treatment of visceral leishmaniasis in mice. *Antimicrobial agents and chemotherapy* **36:** 1630-1634

Kumar R, Engwerda C (2014) Vaccines to prevent leishmaniasis. *Clin Trans Immunol* **3:** e13

Kwan AH, Mobli M, Gooley PR, King GF, Mackay JP (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS Journal* **278:** 687-703

Lacal PM, Pennington Cy Fau - Lacal JC, Lacal JC (1988) Transforming activity of ras proteins translocated to the plasma membrane by a myristoylation sequence from the src gene product. *Oncogene* **6:** 533-537

Li MZ, Elledge SJ (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Meth* **4:** 251-256

Lieberman RL, Culver JA, Entzminger KC, Pai JC, Maynard JA (2011) Crystallization chaperone strategies for membrane proteins. *Methods* **55:** 293-302

Linke WA, Kulke M, Li H, Fujita-Becker S, Neagoe C, Manstein DJ, Gautel M, Fernandez JM (2002) PEVK Domain of Titin: An Entropic Spring with Actin-Binding Properties. *Journal of Structural Biology* **137:** 194-205

Lise S, Jones DT (2005) Sequence patterns associated with disordered regions in proteins. *Proteins: Structure, Function, and Bioinformatics* **58:** 144-150

Liu Y, Manna A, Li R, Martin WE, Murphy RC, Cheung AL, Zhang G (2001) Crystal structure of the SarR protein from Staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 6877-6882

Lobo S, Greentree WK, Linder ME, Deschenes RJ (2002) Identification of a Ras palmitoyltransferase in Saccharomyces cerevisiae. *The Journal of biological chemistry* **277:** 41268-41273

Lodge JK, Jackson-Machelski E, Higgins M, McWherter CA, Sikorski JA, Devadas B, Gordon JI (1998) Genetic and Biochemical Studies Establish That the Fungicidal Effect of a Fully Depeptidized Inhibitor of Cryptococcus neoformans Myristoyl-CoA:ProteinN-Myristoyltransferase (Nmt) Is Nmt-dependent. *Journal of Biological Chemistry* **273:** 12482-12491

Lopes JLS, Miles AJ, Whitmore L, Wallace BA (2014) Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses. *Protein Science* **23:** 1765-1772

Lorenzen K, Duijn Ev (2001) Native Mass Spectrometry as a Tool in Structural Biology.

In *Current Protocols in Protein Science*. John Wiley & Sons, Inc.

Machado-Pinto J, Pinto J, da Costa CA, Genaro O, Marques MJ, Modabber F, Mayrink W (2002) Immunochemotherapy for cutaneous leishmaniasis: a controlled trial using killed Leishmania (Leishmania) amazonensis vaccine plus antimonial. *International journal of dermatology* **41:** 73-78

Maclean LM, O'Toole PJ, Stark M, Marrison J, Seelenmeyer C, Nickel W, Smith DF (2012) Trafficking and release of Leishmania metacyclic HASPB on macrophage invasion. *Cellular microbiology* **14:** 740-761

Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the [micro]-opioid receptor bound to a morphinan antagonist. *Nature* **485:** 321-326

Maroof A, Brown N, Smith B, Hodgkinson MR, Maxwell A, Losch FO, Fritz U, Walden P, Lacey CN, Smith DF, Aebischer T, Kaye PM (2012) Therapeutic vaccination with recombinant adenovirus reduces splenic parasite burden in experimental visceral leishmaniasis. *The Journal of infectious diseases* **205:** 853-863

Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein science* **15:** 2795-2804

Martin DD, Beauchamp E, Berthiaume LG (2011) Post-translational myristoylation: Fat matters in cellular life and death. *Biochimie* **93:** 18-31

Matsuda S, Inoue T, Lee H-C, Kono N, Tanaka F, Gengyo-Ando K, Mitani S, Arai H (2008) Member of the membrane-bound O-acyltransferase (MBOAT) family encodes a lysophospholipid acyltransferase with broad substrate specificity. *Genes to Cells* **13:** 879-888

Maurer-Stroh S, Eisenhaber B, Eisenhaber F (2002) N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences1. *Journal of molecular biology* **317:** 523-540

Maurer-Stroh S, Eisenhaber F (2004) Myristoylation of viral and bacterial proteins. *Trends in Microbiology* **12:** 178-185

McConville MJ, Handman E (2007) The molecular basis of Leishmania pathogenesis. *International journal for parasitology* **37:** 1047-1051

McKean PG, Denny PW, Knuepfer E, Keen JK, Smith DF (2001) Phenotypic changes associated with deletion and overexpression of a stage-regulated gene family in Leishmania. *Cellular microbiology* **3:** 511-523

McKean PG, Trenholme KR, Rangarajan D, Keen JK, Smith DF (1997) Diversity in repeat-containing surface proteins of Leishmania major. *Molecular and biochemical parasitology* **86:** 225-235

McNicholas S, Potterton E, Wilson KS, Noble MEM (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography* **67:** 386-394

Melaku Y, Collin SM, Keus K, Gatluak F, Ritmeijer K, Davidson RN (2007) Treatment of kala-azar in southern Sudan using a 17-day regimen of sodium stibogluconate combined with paromomycin: a retrospective comparison with 30-day sodium stibogluconate monotherapy. *The American journal of tropical medicine and hygiene* **77:** 89-94

Miao M, Bellingham CM, Stahl RJ, Sitarz EE, Lane CJ, Keeley FW (2003) Sequence and Structure Determinants for the Self-aggregation of Recombinant Polypeptides Modeled after Human Elastin. *Journal of Biological Chemistry* **278:** 48553-48562

Mills E, Price HP, Johner A, Emerson JE, Smith DF (2007) Kinetoplastid PPEF phosphatases: Dual acylated proteins expressed in the endomembrane system of Leishmania. *Molecular and biochemical parasitology* **152:** 22-34

Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD (2010) Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **18:** 494-506

Modabber F (2010) Leishmaniasis vaccines: past, present and future. *International journal of antimicrobial agents* **36:** S58-61

Mohana-Borges R, Goto NK, Kroon GJA, Dyson HJ, Wright PE (2004) Structural Characterization of Unfolded States of Apomyoglobin using Residual Dipolar Couplings. *Journal of molecular biology* **340:** 1131-1142

Monne M, Han L, Schwend T, Burendahl S, Jovine L (2008) Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456:** 653-657

Monzote L (2009) Current Treatment of Leishmaniasis: A Review. *The Open Antimicrobial Agents Journal* **1:** 9-19

Moore B, Miles AJ, Guerra-Giraldez C, Simpson P, Iwata M, Wallace BA, Matthews SJ, Smith DF, Brown KA (2011) Structural Basis of Molecular Recognition of the Leishmania Small Hydrophilic Endoplasmic Reticulum-associated Protein (SHERP) at Membrane Surfaces. *Journal of Biological Chemistry* **286:** 9246-9256

Moraes I, Evans G, Sanchez-Weatherby J, Newstead S, Stewart PDS (2014) Membrane protein structure determination — The next generation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838:** 78-87

Morel CM (2003) Neglected diseases: under-funded research and inadequate health interventions. Can we change this reality? *EMBO reports* **4:** S35-38

Moreno J, Nieto J, Masina S, Canavate C, Cruz I, Chicharro C, Carrillo E, Napp S, Reymond C, Kaye PM, Smith DF, Fasel N, Alvar J (2007) Immunization with H1, HASPB1 and MML Leishmania proteins in a vaccine trial against experimental canine leishmaniasis. *Vaccine* **25:** 5290-5300

Mudher A, Lovestone S (2002) Alzheimer's disease – do tauists and baptists finally shake hands? *Trends in Neurosciences* **25:** 22-26

Muiznieks LD, Weiss AS, Keeley FW (2010) Structural disorder and dynamics of elastin. *Biochemistry and Cell Biology* **88:** 239-250

Mukhopadhyay R, Hoh JH (2001) AFM force measurements on microtubule-associated proteins: the projection domain exerts a long-range repulsive force. *FEBS Letters* **505:** 374-378

Mulder FAA, Schipper D, Bott R, Boelens R (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins1. *Journal of molecular biology* **292:** 111-123

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67:** 355-367

Musa AM, Khalil EA, Mahgoub FA, Elgawi SH, Modabber F, Elkadaru AE, Aboud MH, Noazin S, Ghalib HW, El-Hassan AM (2008) Immunochemotherapy of persistent post-kala-azar dermal leishmaniasis: a novel approach to treatment. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102:** 58-63

Nadolski MJ, Linder ME (2007) Protein lipidation. *FEBS Journal* **274:** 5202-5210

Neal RA (1968) The effect of antibiotics of the neomycin group on experimental cutaneous leishmaniasis. *Annals of tropical medicine and parasitology* **62:** 54-62

Nickel W (2005) Unconventional secretory routes: direct protein export across the plasma membrane of mammalian cells. *Traffic* **6:** 607-614

Noazin S, Modabber F, Khamesipour A, Smith PG, Moulton LH, Nasseri K, Sharifi I, Khalil EA, Bernal ID, Antunes CM, Kieny MP, Tanner M (2008) First generation leishmaniasis vaccines: a review of field efficacy trials. *Vaccine* **26:** 6759-6767

Nunes CM, Lima VM, Paula HB, Perri SH, Andrade AM, Dias FE, Burattini MN (2008) Dog culling and replacement in an area endemic for visceral leishmaniasis in Brazil. *Veterinary parasitology* **153:** 19-23

Obmolova G, Malia Tj Fau - Teplyakov A, Teplyakov A Fau - Sweet R, Sweet R Fau - Gilliland GL, Gilliland GL (2010) Promoting crystallization of antibody-antigen complexes via microseed matrix screening. *Acta Crystallographica Section D: Biological Crystallography* **66:** 927–933

Okuda M, Nishimura Y (2015) Real-time and simultaneous monitoring of the phosphorylation and enhanced interaction of p53 and XPC acidic domains with the TFIIH p62 subunit. *Oncogenesis* **4:** 1-11

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9:** S1-S1

Oliva A, Llabrés Ma, Fariña JB (2001) Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. *Journal of Pharmaceutical and Biomedical Analysis* **25:** 833-841

Olliaro PL (2010) Drug combinations for visceral leishmaniasis. *Current opinion in infectious diseases* **23:** 595-602

Olliaro PL, Guerin PJ, Gerstl S, Haaskjold AA, Rottingen JA, Sundar S (2005) Treatment options for visceral leishmaniasis: a systematic review of clinical studies done in India, 1980-2004. *The Lancet infectious diseases* **5:** 763-774

Ostyn B, Vanlerberghe V, Picado A, Dinesh DS, Sundar S, Chappuis F, Rijal S, Dujardin JC, Coosemans M, Boelaert M, Davies C (2008) Vector control by insecticide-treated nets in the fight against visceral leishmaniasis in the Indian subcontinent, what is the evidence? *Tropical medicine & international health : TM & IH* **13:** 1073-1085

Paige La Fau - Zheng GQ, Zheng Gq Fau - DeFrees SA, DeFrees Sa Fau - Cassady JM, Cassady Jm Fau - Geahlen RL, Geahlen RL (1989) S-(2-oxopentadecyl)-CoA, a nonhydrolyzable analogue of myristoyl-CoA, is a potent inhibitor of myristoyl-CoA:protein N-myristoyltransferase. *J Med Chem* **32:** 1665-1667

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong IL, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **289:** 739-745

Papoian GA (2008) Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proceedings of the National Academy of Sciences* **105:** 14237-14238

Payne L, Fitchett JR (2010) Bringing neglected tropical diseases into the spotlight. *Trends in parasitology* **26:** 421-423

Pierce KL, Premont RT, Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nature reviews Molecular cell biology* **3:** 639-650

Price HP, Menon MR, Panethymitaki C, Goulding D, McKean PG, Smith DF (2003) Myristoyl-CoA:protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. *The Journal of biological chemistry* **278:** 7206-7214

Price Ii WN, Chen Y, Handelman SK, Neely H, Manor P, Karlin R, Nair R, Liu J, Baran M, Everett J, Tong SN, Forouhar F, Swaminathan SS, Acton T, Xiao R, Luft JR, Lauricella A, DeTitta GT, Rost B, Montelione GT, Hunt JF (2009) Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotech* **27:** 51-57

Privé GG (2007) Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41:** 388-397

Quinnell RJ, Courtenay O (2009) Transmission, reservoir hosts and control of zoonotic visceral leishmaniasis. *Parasitology* **136:** 1915-1934

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker KA (2007) Intrinsic Disorder and Functional Proteomics. *Biophysical Journal* **92:** 1439-1456

Rasia RM, Brutscher B, Plevin MJ (2012) Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins. *ChemBioChem* **13:** 732-739

Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477:** 549-555

Rauscher S, Baud S, Miao M, Keeley Fred W, Pomès R (2006) Proline and Glycine Control Protein Self-Organization into Elastomeric or Amyloid Fibrils. *Structure* **14:** 1667-1676

Rezaei-Ghaleh N, Blackledge M, Zweckstetter M (2012) Intrinsically Disordered Proteins: From Sequence and Conformational Properties toward Drug Discovery. *ChemBioChem* **13:** 930-950

Ritmeijer K, Davies C, van Zorge R, Wang SJ, Schorscher J, Dongu'du SI, Davidson RN (2007) Evaluation of a mass distribution programme for fine-mesh impregnated bednets against visceral leishmaniasis in eastern Sudan. *Tropical medicine & international health* **12:** 404-414

Rogers JM, Wong CT, Clarke J (2014) Coupled Folding and Binding of the Disordered Protein PUMA Does Not Require Particular Residual Structure. *Journal of the American Chemical Society* **136:** 5197-5200

Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin J-C, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC (2011) Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Research* **21:** 2129-2142

Rosenbaum DM, Cherezov V, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Yao X-J, Weis WI, Stevens RC, Kobilka BK (2007) GPCR Engineering Yields High-Resolution Structural Insights into β2-Adrenergic Receptor Function. *Science* **318:** 1266-1273

Roth AF, Feng Y, Chen L, Davis NG (2002) The yeast DHHC cysteine-rich domain protein Akr1p is a palmitoyl transferase. *The Journal of cell biology* **159:** 23-28

Rudnick DA, McWherter CA, Rocque WJ, Lennon PJ, Getman DP, Gordon JI (1991) Kinetic and structural evidence for a sequential ordered Bi Bi mechanism of catalysis by Saccharomyces cerevisiae myristoyl-CoA:protein N-myristoyltransferase. *Journal of Biological Chemistry* **266:** 9732-9739

Sacks D, Noben-Trauth N (2002) The immunology of susceptibility and resistance to Leishmania major in mice. *Nature reviews Immunology* **2:** 845-858

Sacks DL, Modi G, Rowton E, Späth G, Epstein L, Turco SJ, Beverley SM (2000) The role of phosphoglycans in Leishmania–sand fly interactions. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 406-411

Sadlova J, Price HP, Smith BA, Votypka J, Volf P, Smith DF (2010) The stage-regulated HASPB and SHERP proteins are essential for differentiation of the protozoan parasite Leishmania major in its sand fly vector, Phlebotomus papatasi. *Cellular microbiology* **12:** 1765-1779

Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M (2010) NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **132:** 8407-8418

Schonian G, Mauricio I, Gramiccia M, Canavate C, Boelaert M, Dujardin JC (2008) Leishmaniases in the Mediterranean in the era of molecular epidemiology. *Trends in parasitology* **24:** 135-142

Seifert K, Croft SL (2006) In vitro and in vivo interactions between miltefosine and other antileishmanial drugs. *Antimicrobial agents and chemotherapy* **50:** 73-79

Selvakumar P, Lakshmikuttyamma A, Charavaryamath C, Singh B, Tuchek J, Sharma RK (2005) Expression of myristoyltransferase and its interacting proteins in epilepsy. *Biochemical and Biophysical Research Communications* **335:** 1132-1139

Selvakumar P, Lakshmikuttyamma A, Shrivastav A, Das SB, Dimmock JR, Sharma RK (2007) Potential role of N-myristoyltransferase in cancer. *Progress in Lipid Research* **46:** 1-36

Serrière J, Dugua J-M, Bossus M, Verrier B, Haser R, Gouet P, Guillon C (2011) Fab'-Induced Folding of Antigenic N-Terminal Peptides from Intrinsically Disordered HIV-1 Tat Revealed by X-ray Crystallography. *Journal of molecular biology* **405:** 33-42

Sevcik J, Skrabana R, Dvorsky R, Csokova N, Iqbal K, Novak M (2007) X-ray structure of the PHF core C-terminus: Insight into the folding of the intrinsically disordered protein tau in Alzheimer's disease. *FEBS Letters* **581:** 5872-5878

Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56:** 227-241

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proceedings of the National Academy of Sciences* **105:** 4685-4690

Shi Y, Mowery Ra Fau - Ashley J, Ashley J Fau - Hentz M, Hentz M Fau - Ramirez AJ, Ramirez Aj Fau - Bilgicer B, Bilgicer B Fau - Slunt-Brown H, Slunt-Brown H Fau - Borchelt DR, Borchelt Dr Fau - Shaw BF, Shaw BF (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein science* **21:** 1197–1209

Shimamura T, Shiroishi M, Weyand S, Tsujimoto H, Winter G, Katritch V, Abagyan R, Cherezov V, Liu W, Han GW, Kobayashi T, Stevens RC, Iwata S (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* **475:** 65-70

Shojania S, O'Neil JD (2006) HIV-1 Tat Is a Natively Unfolded Protein: THE SOLUTION CONFORMATION AND DYNAMICS OF REDUCED HIV-1 Tat-(1–72) BY NMR SPECTROSCOPY. *Journal of Biological Chemistry* **281:** 8347-8356

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, Vol. 7, p. 539.

Singh N, Kumar M, Singh RK (2012) Leishmaniasis: current status of available drugs and new potential drug targets. *Asian Pacific journal of tropical medicine* **5:** 485-497

Skrabana R, Dvorsky R, Sevcik J, Novak M (2010) Monoclonal antibody MN423 as a stable mold facilitates structure determination of disordered tau protein. *Journal of Structural Biology* **171:** 74-81

Smotrys JE, Linder ME (2004) Palmitoylation of intracellular signaling proteins: regulation and function. *Annual review of biochemistry* **73:** 559-587

Stager S, Alexander J, Kirby AC, Botto M, Rooijen NV, Smith DF, Brombacher F, Kaye PM (2003) Natural antibodies and complement are endogenous adjuvants for vaccine-induced CD8+ T-cell responses. *Nature medicine* **9:** 1287-1292

Stager S, Smith DF, Kaye PM (2000) Immunization with a recombinant stage-regulated surface protein from Leishmania donovani induces protection against visceral leishmaniasis. *Journal of immunology* **165:** 7064-7071

Stahl SJ, Watts NR, Wingfield PT (2014) Generation and Use of Antibody Fragments for Structural Studies of Proteins Refractory to Crystallization. *Methods in molecular biology* **1131:** 549-561

Stott K, Watson M, Bostock MJ, Mortensen SA, Travers A, Grasser KD, Thomas JO (2014) Structural Insights into the Mechanism of Negative Regulation of Single-box

High Mobility Group Proteins by the Acidic Tail Domain. *Journal of Biological Chemistry* **289:** 29817-29826

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41:** 207-234

Sundar S, Jha TK, Thakur CP, Engel J, Sindermann H, Fischer C, Junge K, Bryceson A, Berman J (2002) Oral miltefosine for Indian visceral leishmaniasis. *The New England journal of medicine* **347:** 1739-1746

Sundar S, Mehta H, Suresh AV, Singh SP, Rai M, Murray HW (2004) Amphotericin B treatment for Indian visceral leishmaniasis: conventional versus lipid formulations. *Clinical infectious diseases* **38:** 377-383

Sundar S, Olliaro PL (2007) Miltefosine in the treatment of leishmaniasis: Clinical evidence for informed clinical risk management. *Therapeutics and clinical risk management* **3:** 733-740

Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* **465:** 747-751

Tamiola K, Acar B, Mulder FAA (2010) Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **132:** 18000-18003

Tamiola K, Mulder FA (2012) Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem Soc Trans*

Tompa P (2003) Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays* **25:** 847-855

Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters* **579:** 3346-3354

Vagin A, Teplyakov A (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30:** 1022-1025

van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114:** 6589-6631

Velez ID, Gilchrist K, Arbelaez MP, Rojas CA, Puerta JA, Antunes CM, Zicker F, Modabber F (2005) Failure of a killed Leishmania amazonensis vaccine against American cutaneous leishmaniasis in Colombia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **99:** 593-598

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59:** 687–696

Wallin E, Von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science* **7:** 1029-1038

Weinstock DS, Narayanan C, Baum J, Levy RM (2008) Correlation between (13)C(α) chemical shifts and helix content of peptide ensembles. *Protein science* **17:** 950-954

Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89:** 392-400

Wilcox C, Hu Js Fau - Olson EN, Olson EN (1987) Acylation of proteins with myristic acid occurs cotranslationally. *Science* **238:** 1275-1278

Wilkinson TA, Yin J, Pidgeon C, Post CB (2000) Alkylation of cysteine-containing peptides to mimic palmitoylation. *The Journal of Peptide Research* **55:** 140-147

Wiltzius JJ, Sievers SA, Sawaya MR, Eisenberg D (2009) Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process. *Protein science* **18:** 1521-1530

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67:** 235-242

Wischik CM, Novak M, Thøgersen HC, Edwards PC, Runswick MJ, Jakes R, Walker JE, Milstein C, Roth M, Klug A (1988) Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* **85:** 4506-4510

Wright M, Heal W, Mann D, Tate EW (2010) Protein myristoylation in health and disease. *J Chem Biol* **3:** 19-35

Wright Megan H, Paape D, Storck Elisabeth M, Serwa Remigiusz A, Smith Deborah F, Tate Edward W (2015) Global Analysis of Protein N-Myristoylation and Exploration of N-Myristoyltransferase as a Drug Target in the Neglected Human Pathogen Leishmania donovani. *Chemistry & Biology* **22:** 342-354

Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews Molecular cell biology* **16:** 18-29

Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330:** 1066-1071

Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang

X-P, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC (2012) Structure of the human [kgr]-opioid receptor in complex with JDTic. *Nature* **485:** 327-332

Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A2A adenosine receptor. *Science* **332:** 322-327

Xu X, Ishima R, Ames JB (2011) Conformational dynamics of recoverin's Ca2+-myristoyl switch probed by 15N NMR relaxation dispersion and chemical shift analysis. *Proteins: Structure, Function, and Bioinformatics* **79:** 1910-1922

Ye H (2006) Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography–multiangle laser light scattering. *Analytical Biochemistry* **356:** 76-85

Zha J, Weiler S, Oh KJ, Wei MC, Korsmeyer SJ (2000) Posttranslational N-myristoylation of BID as a molecular switch for targeting mitochondria and apoptosis. *Science* **290:** 1761-1765

Zhang Y, MacArthur C, Mubila L, Baker S (2010) Control of neglected tropical diseases needs a long-term commitment. *BMC medicine* **8:** 67

Zhu D, Saul A, Huang S, Martin LB, Miller LH, Rausch KM (2009) Use of o-phthalaldehyde assay to determine protein contents of Alhydrogel-based vaccines. *Vaccine* **27:** 6054-6059

Zijlstra EE, Musa AM, Khalil EA, el-Hassan IM, el-Hassan AM (2003) Post-kala-azar dermal leishmaniasis. *The Lancet infectious diseases* **3:** 87-98

Zou Y, Weis WI, Kobilka BK (2012) N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. *PloS one* **7:** e46039