Amphotericin B resistance in Leishmania: A polyomics approach.

1. Summary

Leishmania is a protozoan parasite that causes a range of diseases across large parts of the tropics and subtropics. It is one the group of communicable diseases labelled neglected tropical diseases (NTD) targeted for control by the WHO public health programs. These diseases have a major impact on the health and economies of some of the poorest countries of the world. Treatment for Leishmaniasis is limited to a few drugs, with Amphotericin B being the most effective at present. Drug resistance is an ever present threat to effective treatment and understanding the mechanism by which the parasite becomes resistant to drugs is crucial to the development of new treatments.

We used a polyomics approach to comparing resistant and amphotericin B sensitive strains of the parasite to study the resistance mechanism. This include genomics, proteomics and metabolomics. By combining the three approaches a single mutation was identified in the gene CYP51 that codes for demethylase in the steroid biosynthesis pathway.

The discovery that this mutation was not in the active site of the enzyme for the effected reaction but in the enzyme immediately upstream has major implications to both our understanding of resistance mechanisms and to our understanding of the interactions between enzymes in the sterol biosynthesis pathway.

2. Introduction

Leishmaniasis is a group of diseases caused by the protozoan parasite of the genus Leishmania . The three main forms of the didease are cutaneous, causing skin lesions, mucocutaneous that attacks the mucus membranes of the nose and mouth and visceral Leishmaniasis that which effects the liver and spleen and has a very high fatality rate. The protozoa is transmitted by the phlebotomine sandfly and is prevalent across Asia, Africa, South and Central America and Southern Europe causing an estimated 900 000 to 1.3 million new infections and 20 000 to 30 000 deaths annually (WHO). It effects the poorest populations of the world and is associated with malnutrition, displacement and compromised immune system (AIDs).

The mainstay of treatment for visceral Leishmaniasis and in areas where resistance to older drugs occurs is Amphotericin B .This is given as a single shot of intravenous drug in the form a liposome which reduces its toxic effects on the patient. Amphotericin B is a polyene amphipathic molecule that has a hydrophobic side and a hydrophilic side. It's mechanism of action is that it binds to ergosterol molecules found in the parasite cell membrane(Kamiński, 2014). This causes disruption of the membrane with the Amphotericin B molecules self-organising into pores in the membrane as the hydrophilic side molecule cluster together away from the lipid bilayer. These pores allow potassium ions to leak out of cells causing ionic imbalance resulting in cell death. Amphotericin B binds with a much higher affinity to ergosterol than cholesterol making it more toxic for the parasite than for humans

The increasing emergence of resistant strains of Leishmania threaten our ability to treat these diseases. The limited range of drugs effective available is a severe impedance to effective treatment and attempts to control this disease. Understanding the mechanism of drug resistance is a vital developing new and effective drugs in the future.
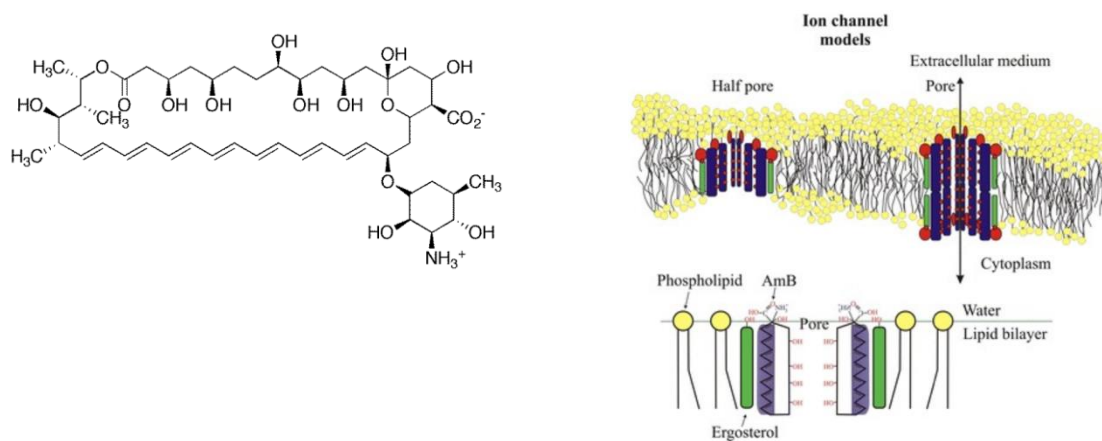
Figure 2.1.a Amphotericin B.
Figure 2.1.b Amphotericin binding ergosterol in the membrane   (Kamiński, 2014)

In this assignment we take a multi-omics approach to studying the resistance mechanism. Genomics was used to identify genome wide Single Nucleotide Polymorphisms (SNP) in wildtype and resistant strains of the protozoa. Proteomics was used in an attempt to identify the key genes that had been up or down regulated by these changes. The results from metabolomics of was to identify molecules that would indicate which pathways had been modulated to either confirm or dismiss various candidate genes. We integrated the data from three approaches to narrow down the search and identify the SNP responsible for resistance. By using the power of combining all three approaches with knowledge of amphotericin B mechanism we were able narrow the mutation down to a single SNP.

3.   Materials and method

3.1   Data Sets

 Leishmania Mexicana laboratory strain M379 progenitor was split into two groups, both were passaged over a period of 2 months. One group was exposed to increasing concentrations of Amphotericin B. After 16 passages the IC_{50} value showed a 13 fold increase in the resistance induced population. This gave us two distinct cell lines, L.mexicana M379 Wild Type and L.mexicana M379 Amphotericin B-Resistant, from now on referred to as WT and AmpB.

3.2   Genomics

Sequencing . The DNA for each group was sequenced on the Illumina (Solexa) platform using paired read technology

Quality control and alignment of raw reads data. The two fastq files were quality checked using the FastQC   software (Andrews). Trim Galore (Trim_Galore!) was then run for paired end reads to remove the adaptors and poor quality reads with a phred score of less than 20. FastQC was run again to check for improved quality.

Align reads to Genome Bowtie2_build was used to build an index for the reference genome from the fasta files obtained from the TryTrpDB (Aslett et al., 2010). Bowtie was then run to align the paired end reads for both the wildtype(WT) and resistant (AmpB) data to give two BAM files.

Samtools was then used to sort the files and remove duplicates. Picard AddorReplace was used to assign read groups to the reads in theBAM file ready for input into the GATK programs.

Variant Calling and Quality Filtering. The GAtk software(Van der Auwera et al., 2013) was used to call and filter the variants.

GATK UnifiedGenotyper was run on the sorted BAM to call variants on the genome with the, producing a raw VCF file. This was run four times once on each file to call SNPs and once to call indels.

GATK VariantFiltration is used to filter the raw data to produce a more robust set of variants. The following iltering parameters used were used :

--filterExpression "QD < 2.0" filterName "QDFilter"
 --filterExpression "MQ < 40.0" --filterName "MQFilter"
--filterExpression "FS > 60.0" --filterName "FSFilter"
--filterExpression "MQRankSum < -12.5" --filterName "MQRankSumFilter"
--filterExpression "ReadPosRankSum < -8.0" --filterName "ReadPosFilter"
--filterExpression "MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)" --filterName "HARD_TO_VALIDATE" --filterExpression "QUAL < 30.0 || DP < 6 || DP > 5000" --filterName "QualFilter" -l INFO

SNPEff (Cingolani et al., 2012) was used to annotate the VCF files .a variant annotation and effect prediction tool. It annotates and predicts the effects of genetic variants (such as amino acid changes).

SNPSIFT to produce VCF file for only the SNPs that pass the filters and that produce a missense variant.

VCFTools used to combine these two into a file with the chromosome, position and base change of each of SNP showing which SNPs are present in both strains and which are unique.

3.3   Metabolomics

Chromatography/ Mass Spectroscopy was used to collect the time resolved spectral data for each biological sample. The data for whole metabolite (cellular and extra-cellular) WT and AmpB strains was processed separately. This workflow involves quality control, peak picking , filtering and annotation and applying statistical analysis to the data. The IDEOM software (Creek et al., 2012) was used to collate and analyse the results.

4. Results

4.1   Genomics

The number of variants called for each stain are shown in table 4.1 below for before and after filtering and then those only that will cause an amino acid substitution in a putative coding region.

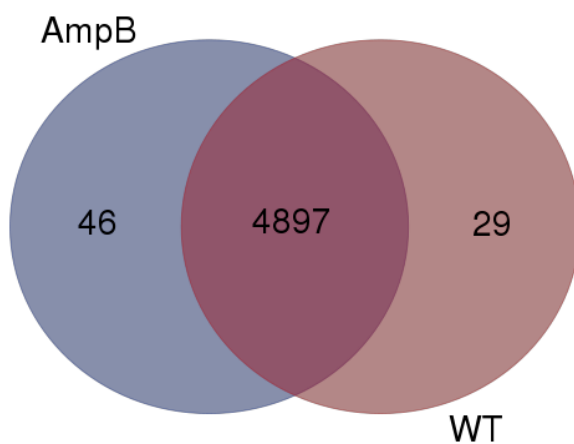| SNP frequency	WildType total	Missense	AmpB total	Missense
Pre filtering	29 515	5580	29 347	5565
Filtered	23 185		23 249	-
Unique: Filtered missense SNPs	(Shared missense SNPs 4897)	29	-	46

Table 4.1 SNP Frequency

Figure 4.1 . Showing the final number of SNPs called for each strain after filtering and annotating for missense mutations.

4.2   Proteomics

The results from the DIGE, were that 37 spots showed modulation with greater than 2.5 fold change. Of these 28 were of sufficient strength and resolution to be cut from the gel for MS analyse resulting in the identification of all 28 proteins. (No results available).

4.3   Metabolomics

Chromatography mass spectroscopy was used to collect raw data for the complete small molecule complement of the samples. The results of the processed data was analysed using the IDEOM spread sheet. Table 4.2. This was sorted on fold-change to group the most up regulated, and down regulated molecules together. This showed the steroid biosynthesis pathway to be over-represented. The volcano plot shows there were relatively few statistically significant high fold change molecules.

Mass	RT	FORMULA	Isomer	name	Confidence	Map	KEGGid#	Groups	Wtcells	Rescells	ttest: Wtcells	ttest: Rescells	Mean: Wtcells	Mean: Rescells	SD: Wtcells
410.35	5.013	C29H46O	8	[ST dimethyl(3:0)] 4,4-dimethylcholesta-8,11,24-	6	Lipids: Sterol lipids	C11455	Rescells	1.00	152.01	NA	0.000446	720.9619	109596.3	NA
356.27	5.21	C24H36O2	6	[FA (24:6)] 4,8,12,15,19,21-tetracosahexaenoic acid	7	Lipids: Fatty Acyls	0	Rescells	0.00	46.61	NA	0.002543	0	46614.95	NA
438.3	5.219	C25H42O6	1	[PR] 3beta-(3-methyl-butanoyloxy)-villanovane-	5	Lipids: Prenols	0	Rescells/Resuperna xal	0.00	31.52	NA	0.002036	0	31520.41	NA
426.35	4.935	C29H46O2	3	[ST methyl(4:0)] (5Z,7E)-(1S,3R)-24-methyl-	5	Lipids: Sterol lipids	0	Rescells	0.00	21.68	NA	0.000671	0	21681.64	NA
374.28	5.205	C24H38O3	36	3-Oxo-5beta-cholanate	5	0	C03070	Rescells	0.00	21.10	NA	0.008501	0	21102.6	NA
567.56	5.015	C36H73NO3	1	[SP (18:0)] N-(octadecanoyl)-sphingosine	5	Lipids: Sphingolipids	0	Wtcells	1.00	0.05	NA	476931.7	25747.08	189811.9	
169.05	20.16	C4H12NO4P	8	Phosphodimethylethanolamine	8	Lipid Metabolism	C13482	Rescells	1.00	0.05	1	0.044158	59961.6	3235.164	33918.5
394.32	4.983	C28H42O	3	Ergosta-5,7,22,24(28)-tetraen-3beta-ol	6	Lipid Metabolism	C05440	Wtcells	1.00	0.05	1	NA	33706.94	1779.861	9331.028
410.32	5.001	C28H42O2	3	[ST hydroxy,methyl(4:0)] (22E)-(8S)-3-hydroxy-22-	5	Lipids: Sterol lipids	0	Wtcells	1.00	0.05	1	NA	24683.3	1149.355	11467.7
539.53	4.947	C34H69NO3	1	[SP (16:0)] N-(hexadecanoyl)-sphingosine	7	Lipids: Sphingolipids	0	Wtcells	1.00	0.04	1	NA	1473996	64438.19	651671.4
155.03	19.26	C3H10NO4P	3	N-Methylethanolamine phosphate	6	Lipid Metabolism	C01210	Wtcells/Rescells	1.00	0.01	1	0.013076	636197.5	7651.121	237012.4

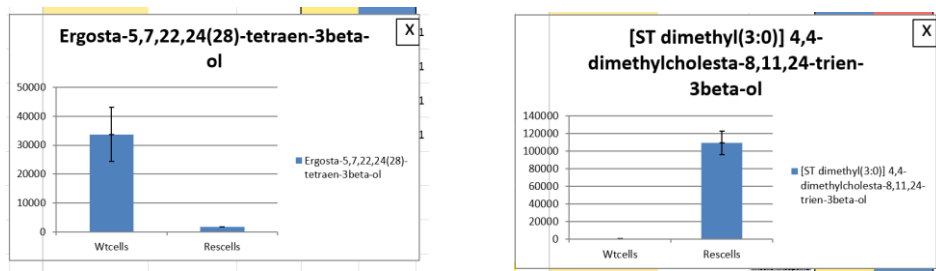Table 4.2 Showing the results on IDEOM , filtered to show biggest fold changes

Figure 4.2a . Results for two most statistically significant molecules with the highest fold change.
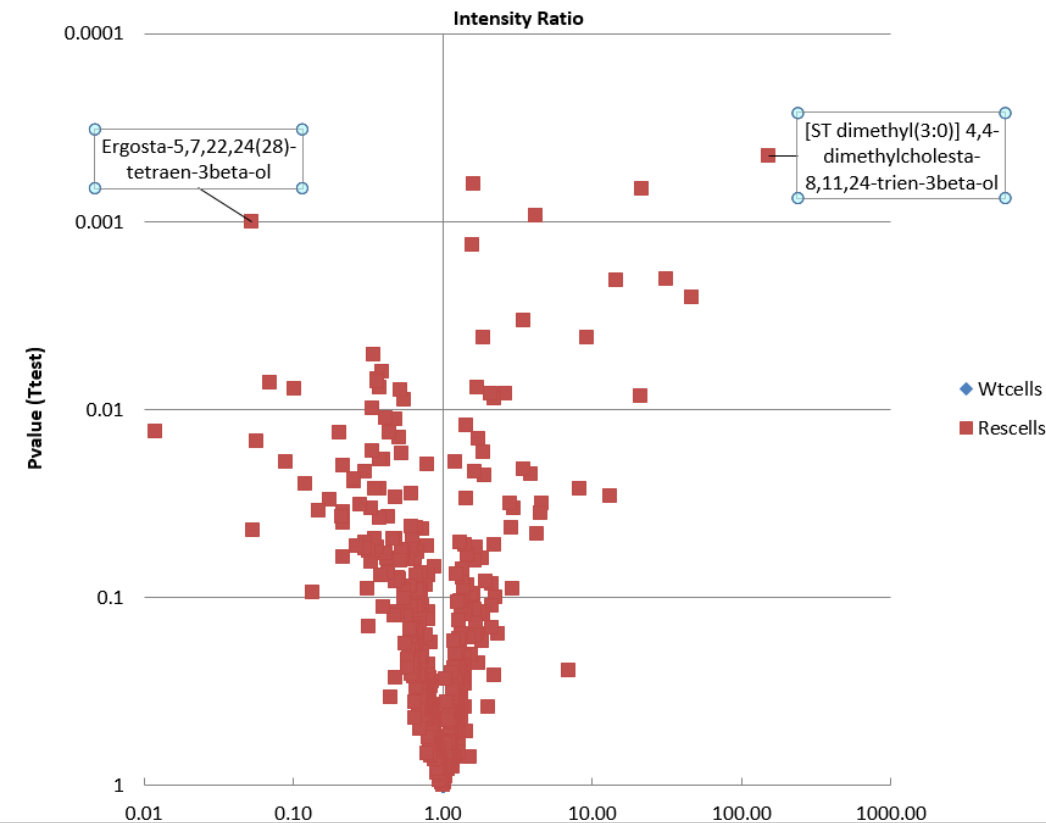

Figure 4.2.Volcano Plot of the metabolic data , plotted Intensity Ratio (fold-change) against P-value.

The facility on IDEOM to search KEGG(Kanehisa et al., 2006) pathways via Pathos (Leader et al., 2011)was used and the two molecules with the biggest fold change were highlighted in the pathway. The steroid biosynthesis pathway as most significant. Figure4.3.
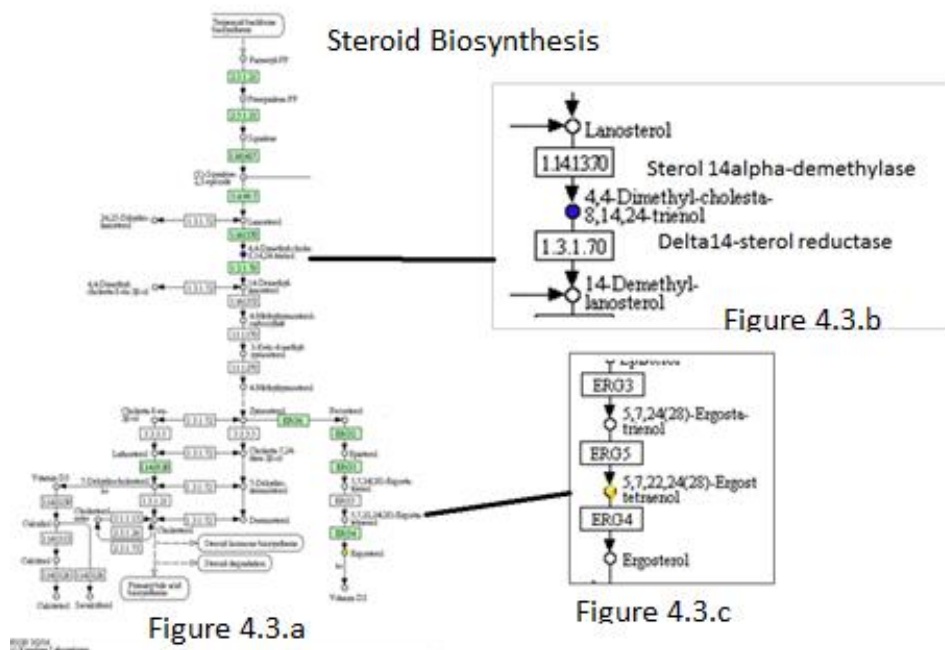

Figure 4.3 . Kegg Pathway

4.4   Intergation

The metabolomics results indicates that the steroid biosynthesis pathway to be key. The TryTrpDB resource (Aslett et al., 2010) was then used to obtain the chromosome position for all the Leishmania Mexicana enzymes in the steroid biosynthesis pathway. A script was then used to compare this to the final list of candidate SNPs produced above. This narrowed it down to a list of six genes. The most significant was in CYP51 that codes for the protein 14 alpha- demethylase DM. The Kegg reaction shows the involvement of both the enzyme 4DM and the dimethyl-cholesta molecule (Figure 4.3 b)

5.   Discussion

The results from the genomic experiment alone produced candidate SNPs that were present in only the resistant strain , too many to be processed manually. By integrating the metabolomics information this list was narrowed down to one SNP in the gene CYP51 that codes for the enzyme in the steroid biosynthesis pathway. Including prior knowledge of the proposed action of Amphotericin B on the lipid membrane further strengthens the case for this candidate.

A gene exchange experiment was done where the wildtype Cyp51 gene was substituted back into the resistant strain. The results show this shows amphotericin B resistance returning to the same level as the wildtype strain. This strongly backs up the claim that the CYP51 mutation induces resistance.
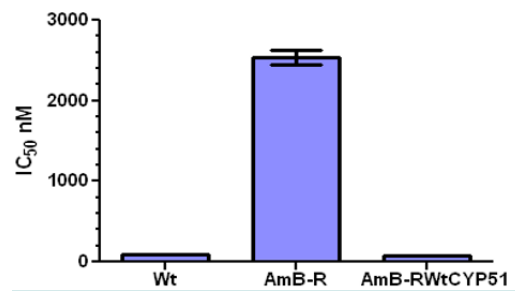
Figure 5.1 The results of the gene exchange experiment (Lecture Notes, K.Crouch, M.Mudalair,J.Wilkes)

The results from the proteomic analysis are uncertain, they could indicate a lack of expression modulation of the proteins of interest or may be due to the difficulties in characterisation and identification of membrane proteins due to their low abundance and hydrophobic nature. Further investigating is necessary to confirm this either way such as using a Western Blot protocol suitable for membrane proteins.

The CYP51 enzyme 14alpha-demethylase catalyses the reaction of lanosterol into 4,4-dimethyl-5alpha-cholesta-8,14,24-trien-3beta as seen in figure. It is the product of this reaction that is up-regulated, not as would be expected if the active site was disrupted the substrate. The enzyme C14 sterol reductase (LmM.31.2320) which acts on this cholesta molecule as a substrate does not contain any SNPs, even in the pre-filtered variants.

The regulation of a single gene such is unlikely since *Leishmania* has an unusual mechanism of transcriptional control (Clayton and Shapira 2007). Protein-coding genes in transcription units are co-transcribed by RNA polymerase, messenger RNA levels are regulated by RNA stability, rather than the activity of promoters, precluding the up-regulation of gene expression through increased RNA polymerase II activity.

The SNP identified has substituted a thymine in the reference genome for adenine in the resistant strain, causing an amino acid substitution from asparagine (N 176) to Isoleucine. Asparagine is a neutral polar side chain often involved in forming hydrogen bonds with a hydropathy index of -3.5, whereas isoleucine is a nonpolar side chain with a hydopathy index of 4.5(the most hydrophobic side chain). The hydropathy index(Kyte and Doolittle, 1982) is a measure of the degree hydrophobicity and has been used to predict the structure of membrane proteins. This change in property maybe very important as the hydrophobic effect is one of the main driving forces of protein folding in both soluble and membrane proteins.

CYP51 is an essential enzyme for the synthesis of the sterols and is present across all biological kingdoms. The primary sequence shows low identity across kingdoms , 26% between Leishmania infantum and human ,(Lepesheva and Waterman, 2011) as compared to However the three dimensionally structure is very strong similarity 84% between Leishmania infantum and Human and 92% with T. brucie (CATH).

The sequence comparison shows that key residues are conserved across all speicies Figure 5.2. As expected the active site and those that are key to structure. The ASN 176 is not conserved across the animal kingdoms but is conserved within the family Trypanosomatidae, Figure 5.3. Further investigation is needed to ascertain if it is conserved in all ergosterol membrane species such as fungi which imply that it is important to the ergosterol pathway.
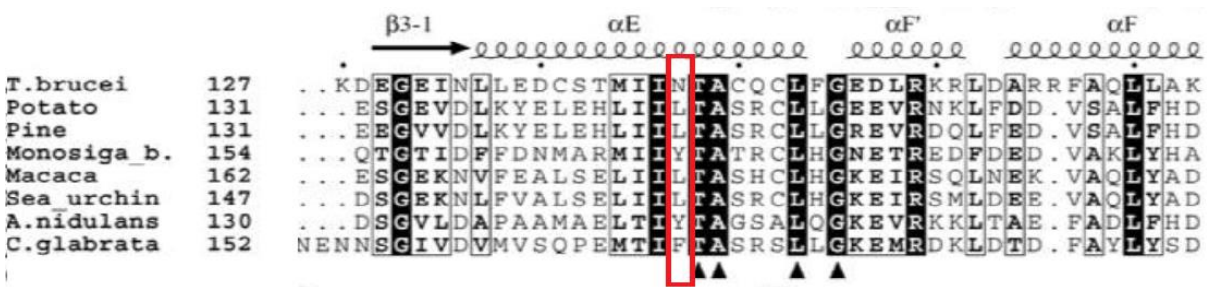
Figure 5.2. Multiple sequence alignment of the CYP51 family and displays two protozoan (T. brucei and M. brevicollis), two animal (monkey and sea urchin), two plant (potato and pine tree) and two fungal (A.nidulans (filamentous) ,C.glabrata (yeast)). (Lepesheva and Waterman, 2011)
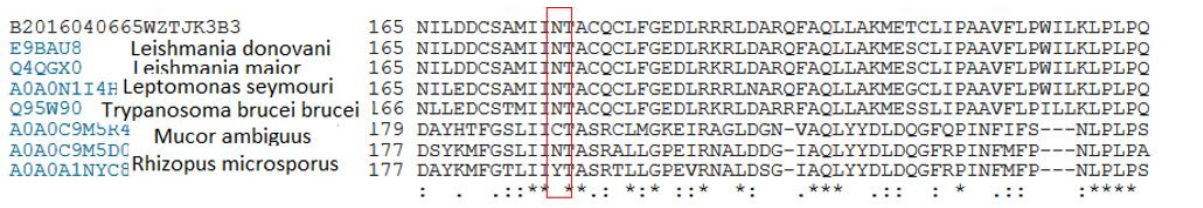

Figure 5.3. Alignment of members of the Trypanosomatidae family (top 4) and three fungi(bottom 3) (Uniprot, ClustalΩ)

That an amino acid substitution in one enzyme appears to affect the action of the next enzyme in the pathway would suggest that there is some form of physical interaction between the enzymes. The crystal structure of sterol 14-alpha demethylase from Leishmania infantum (Hargrove et al., 2011) was used to investigate the position of the substitution. Asparagine 176 sits in a narrow groove less than 12 Å wide .This indicates that a protein-protein interaction could form via binding a helix in this groove. Another factor that might indicate whether this is a biding site is the polar character of asparagine, it would be expected that the residues exposed to the lipid bilayer would be of a hydrophobic nature. Further investigation into the conservation and properties of the amino acids lining this groove is needed both for this a-helix and the two either side.
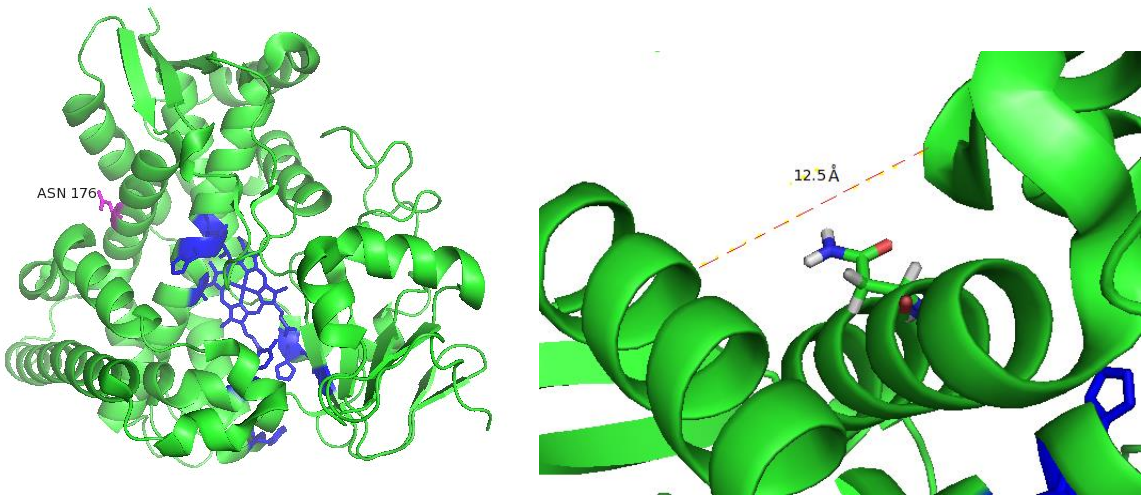

Figure 5.4. The 3.D structure of 14alpha-demethylase. The active site and heam are highlighted in blue and ASN 176 in pink. 5.4.b shows the groove containing ASN 176 (produced on Pymol)

Investigating the structure of sterol reductase (Li et al., 2015) from Methylomicrobium alcaliphilum may reveal an α-helix with complementary side chains to the Asp176 groove, such as the external α2 helix. This is the only 3.D structure this enzyme available at present. Clustal Ω shows 27% primary sequence identity with L.mexicana with 144 identical residues and a further 144 that are similar. A thorough comparison of the primary sequence would   need to establish which regions are likely to have structural homology.

Further investigation is needed to establish the nature of the interaction between these two enzymes and how it influences their function and if they are part of a bigger interactome that would enhance the efficiency of the sterol biosynthesis pathway. This may lead a better understanding of the sterol biosynthesis pathways in the Trypanosomatidae family and to the potential to find new targets for drug development.

The power of integrating omics data has been demonstrated. If we had used the SNP data alone to look for a mechanism we could well have missed this one in CYP51 as we would expect the active site or regulatory element to be involved. By combining information from the different omics experiments we have discovered that Leishmania gains resistance to amphotericin B by interrupting the ergosterol pathway and so altering the composition of its membrane. The mechanism by which this is achieved is speculative but appears to point towards a close interaction of the two enzymes involved.

References

Andrews Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w [1118] ; iso-2; iso-3. Fly (Austin) *6*, 80–92.

Creek, D.J., Jankevics, A., Burgess, K.E.V., Breitling, R., and Barrett, M.P. (2012). IDEOM. Bioinformatics *28*, 1048–1049.

Hargrove, T.Y., Wawrzak, Z., Liu, J., Nes, W.D., Waterman, M.R., and Lepesheva, G.I. (2011). Substrate Preferences and Catalytic Parameters Determined by Structural Characteristics of Sterol 14α-Demethylase (CYP51) from Leishmania infantum. J. Biol. Chem. *286*, 26838.

Kamiński, D.M. (2014). Recent progress in the study of the interactions of amphotericin B with cholesterol and ergosterol in lipid environments. Eur. Biophys. J. *43*, 453.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. *34*, D354–D357.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

Leader, D.P., Burgess, K., Creek, D., and Barrett, M.P. (2011). Pathos: A web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry. Rapid Commun. Mass Spectrom. *25*, 3422–3426.

Lepesheva, G.I., and Waterman, M.R. (2011). Structural basis for conservation in the CYP51 family. Biochim. Biophys. Acta *1814*, 88.

Li, X., Roberti, R., and Blobel, G. (2015). Structure of an integral membrane sterol reductase from Methylomicrobium alcaliphilum. Nature *517*, 104–107.

Trim_Galore! Babraham Bioinformatics - Trim Galore!

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al *43*, 11.10.1–33.