3 RESULTS

3.1 BIOINFORMATIC ANALYSIS OF *LEISHMANIA MEXICANA* PROMASTIGOTE AND AMASTIGOTE PROTEOME

3.1.1 PROTEOME ANALYSIS

In the Diploma thesis 'Proteome comparison of *L. mexicana* amastigotes versus promastigotes'[4] a novel approach of purifying amastigotes from their intracellular habitats in bmdm was developed and established. Parasite material of promastigotes and purified amastigotes, was lysed in 2-dimensional (2D) lysis buffer and therein soluble proteins were separated by 2D gel electrophoresis (2-DE). The gel was Coomassie-stained and all visible spots, 391 and 390 from promastigotes and amastigotes respectively, were excised. The 2D-lysis buffer insoluble fraction was washed and solubilized in Laemmli-buffer and separated on an 1-dimensional SDS PAGE (1-DE). After Coomassie-staining the lanes of the 1-DE were cut in 96 slices and these were processed and analysed by nano liquid chromatography electrospray ionization tandem mass spectrometry (nano-LC ESI-MS/MS). In the early stages of this PhD, an additional amastigotes sample was obtained and separated by 2D gel electrophoresis. The spot pattern of the gels showed a high similarity (see FIGURE 6). This was confirmed by a comparison of the spot pattern using 'PDQuest' (Biorad, Hercules, CA, USA). In addition, 100 protein spots that were not identified on the left-hand gel (FIGURE 6) were subsequently excised from the right-hand gel, and analysed. These were processed and analysed as described in the Diploma thesis, i.e. by Matrix-assisted laser desorption/ionization tandem mass spectrometry (MALDI MS/MS). In total, 391 spots from promastigotes and 490 spots from amastigotes were analysed (see FIGURE 6 and FIGURE 7) with 74 and 49% success in spot identification, respectively. Overall 509 leishmanial proteins were identified (Paape, D. *et. al.*, 2008), see FIGURE 8.

[4] Daniel Paape 'Ein Vergleich der Proteome von *Leishmania mexicana* Promastigoten versus Amastigoten' Diploma thesis, Freie Universität Berlin, Department Biology, Chemistry and Pharmacy in 2005.

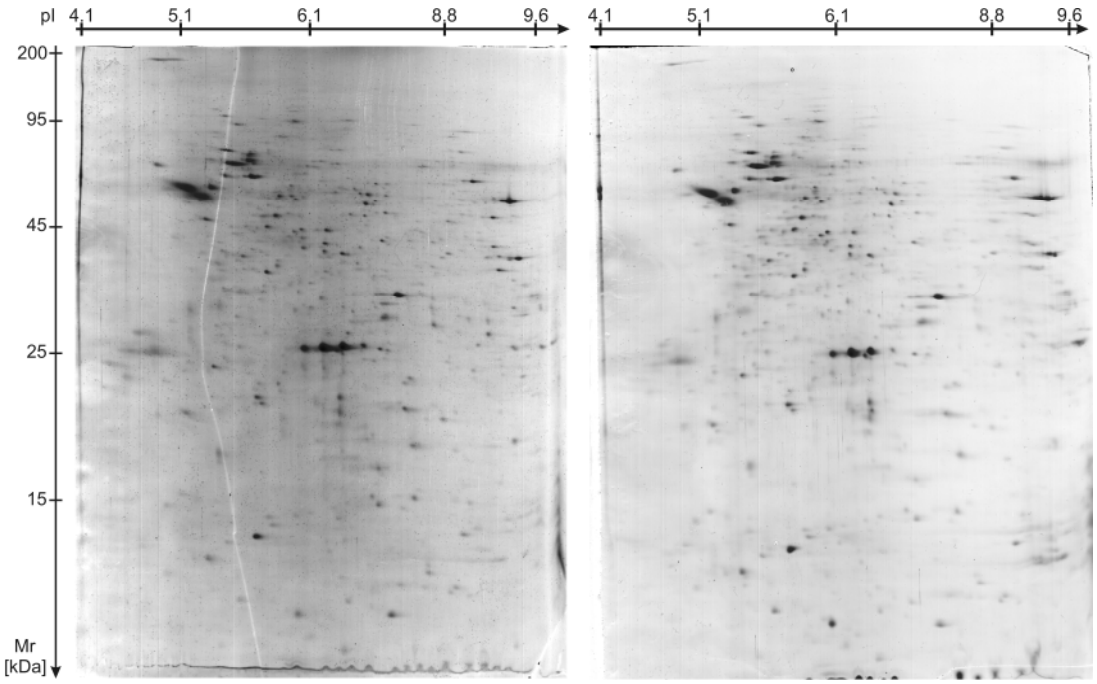


FIGURE 6. Proteome separated by 2-DE of *L. mexicana::DsRed* amastigotes

Two individual samples of amastigotes isolated from 24 h infected bmdm are shown. All visible spots from the gel depicted on the left were excised, processed and analysed by MALDI MS/MS. After comparison of the two gels with PDQuest, protein spots, not identified from the left gel were excised from the right gel and analysed as before. In total 490 spots were processed and analysed.

The MALDI MS/MS and the LC-ESI MS/MS generated datasets were largely complementary as only approximately 10% of identified proteins were in both sets, see FIGURE 8.

The genome of Leishmania contains many tandem arrays of identical or highly similar genes encoding for proteins. Hence peptide mass fingerprint and MS/MS spectra derived identifications could not always be assigned to an individual gene copy, instead they were assigned to the respective gene product array.

Compared with the genome sequence-predicted-proteome, a conservative estimate of coverage counting only distinct proteins identified, was ~5.5% and 2.5% for promastigotes and amastigotes, respectively. Total coverage was ~6% (see suppl. Table 1 for a comprehensive list of all identified proteins). Only five spots in the amastigote sample corresponded to mouse proteins, indicating that contamination with host cell material was <2.3% (5 of 214). Mouse-derived peptides were assigned to vimentin and vacuolar ATPase A and B subunits,

whereas another set matched HSP70 chaperones (most likely BiP), and another to an unnamed mouse protein.

In addition amastigotes were also purified from BALB/c lesions and the proteomes of amastigotes from either 24 h infected bmdm, or obtained from BALB/c lesions, were highly similar as analysis by PDquest revealed, see FIGURE 7.

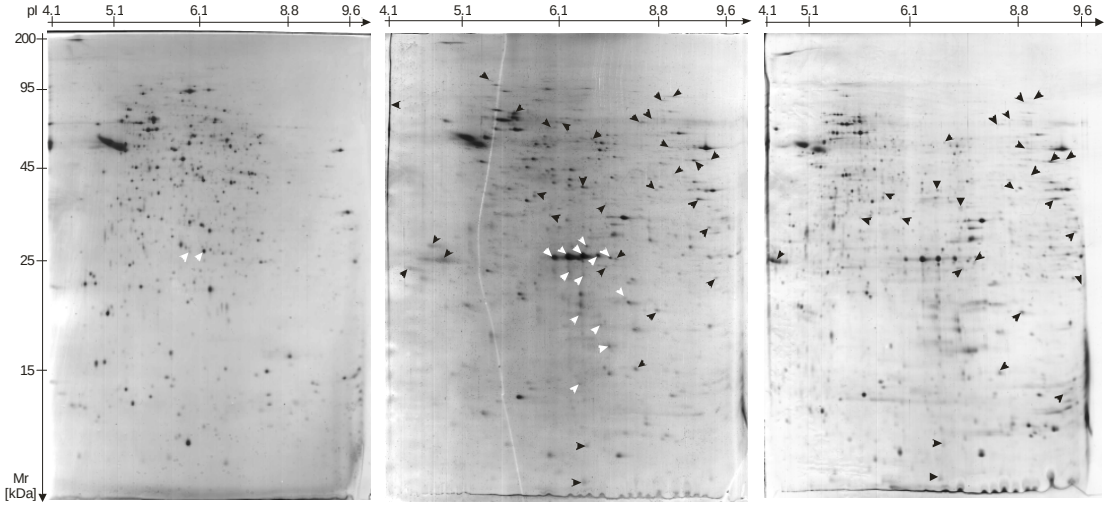


FIGURE 7. Comparison of 2-DE-separated proteome of *L. mexicana::DsRed* pro- and amastigotes.

Left) promastigotes; **middle)** amastigotes isolated from 24 h infected bmdm; **right)** amastigotes isolated from BALB/c mice lesions. *White arrowheads* point at spots identified as DsRed, and *black arrowheads* in the middle panel point at 31, of a total of 45 protein spots, of which the proteins were only detectable in amastigotes; the remaining 14 proteins spots were identified in the second gel of bmdm isolated amastigotes. *Black arrowheads* in the proteome of lesion amastigotes point at 26 spots assigned by PDQuest as being identical to marked spots in the 2-DE gel of amastigotes isolated from MΦ.

3.1.2 GENERAL CHARACTERISTICS OF PROMASTIGOTE AND AMASTIGOTE PROTEOME

Bioinformatic analyses based on the *L. major* reference genome and the thereof derived proteome were applied to reveal general characteristics of the identified protein sets. These included calculating the theoretical isoelectric point (pI) and molecular weights (MW) for all identified proteins, and plotting these values for every protein to obtain a theoretical 2-DE gel. In addition the bias in codon usage was investigated. Furthermore, datasets of pro- and amastigotes were classified and compared according to their likely function and predicted subcellular localisation.

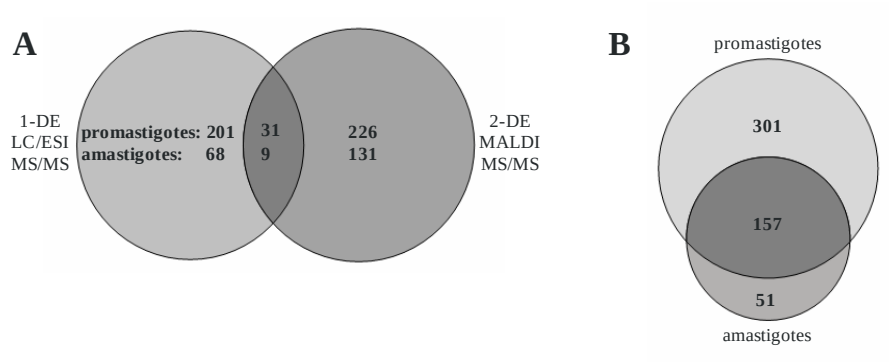


FIGURE 8. Complementary protein sets identified by MALDI MS/MS and nano-LC ESI-MS/MS.

A) Proteins identified by 2D gel electrophoresis and subsequent MALDI MS/MS and/or 1D gel electrophoresis with subsequent LC ESI-MS/MS in promastigotes and amastigotes, respectively. In total, 509 leishmanial proteins were identified (458 and 208 for promastigotes and amastigotes, respectively). B) Number of identified proteins in different life cycle stages identified by both mass spectrometric approaches.

Visual inspection of the 2-DE gels of promastigotes and amastigotes suggests that relatively more basic proteins were present in the amastigote proteome. Therefore, the theoretical MW and pI based on the *L. major* reference genome/proteome were calculated and plotted, see FIGURE 9. The reference proteome shows a bimodal distribution with 50% of the proteins encoding for acidic or basic proteins. By plotting the genome, abundance data are not considered and every gene has the same weight. Abundance data are contained in the proteome data, and differences detected. When 2-DE identified proteins of promastigotes and amastigotes were plotted a significant shift towards proteins with higher pI values was observed for the latter. In the dataset of promastigotes, 80% of the identified proteins were acidic whereas in the amastigote dataset only 59% were acidic, this is significantly different ($p < 0.0001$; Fisher exact test).

Next, translational bias in codon usage was investigated. Codon bias is the preferred usage of a certain codon for a particular amino acid. Translational bias is if this codon usage bias is linked to translation (Sharp, P. M. *et. al.*, 1986; Sharp, P. M. and Li, W. H. 1987). At ~ 6% coverage only abundant proteins were identified and a preferred usage of codons should be evident. In order to find evidence for translational bias in *Leishmania* spp. codon usage was analysed based

on our dataset. To this end a codon adaptation index (CAI) was calculated according to Carbone et al. (Carbone, A. *et. al.*, 2003). Their algorithm iteratively computes a weight for every single codon and an integrated measure for the whole of the respective ORF, its CAI. Hence for all ORFs excluding pseudogenes of the *L. major* reference genome, CAIs were computed and these were ranked on a scale of 0 to 8292 (0 indicates the least and 8292 indicating the highest translational bias) corresponding to the number of ORFs. Subsequently ranks of genes encoding proteins detected in our pro- or amastigote derived datasets were plotted, see FIGURE 10.

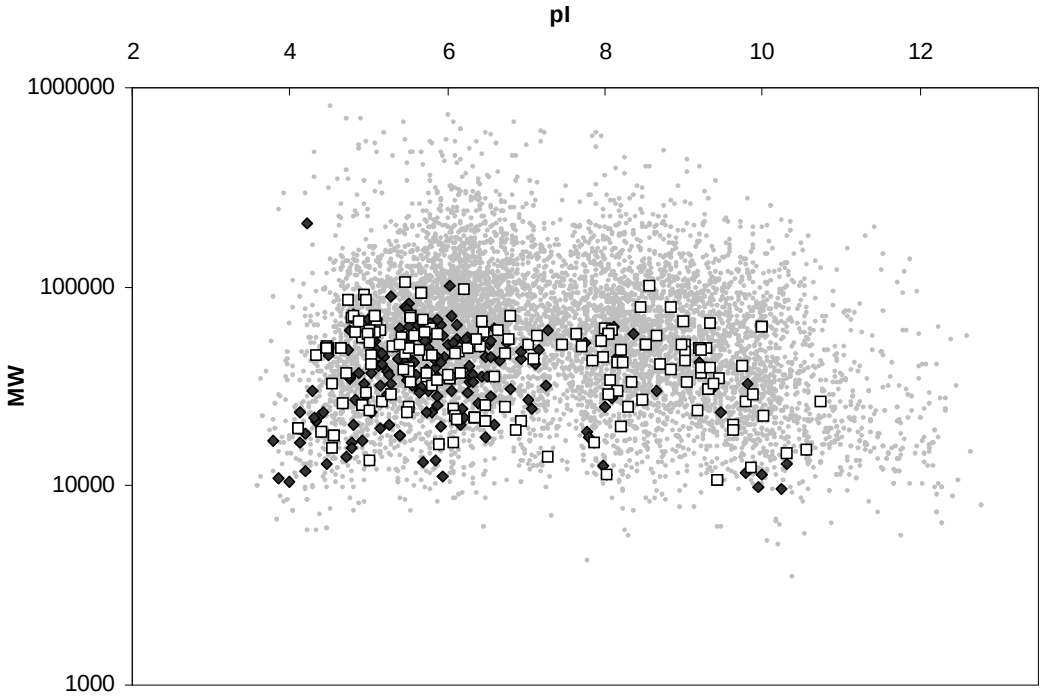


FIGURE 9. Amastigotes contain more proteins with basic pI.

A virtual 2-DE gel image is shown: *L. major* (Friedlin) reference genome ORFs (*gray dots*), ORFs corresponding to proteins identified in *L. mexicana* promastigotes (*black diamonds*), and ORFs corresponding to proteins identified in amastigotes (*white squares*). The genome shows a biphasic distribution with an average pI of 7.312 and a median of 6.9817, and it encodes for 50% acidic (pI < 7.0) proteins. The majority, 80% of protein homologues identified in *L. mexicana* promastigotes were acidic in contrast to only 59% in amastigotes; this is significantly different $p < 0.0001$ (Fisher's exact test).

The vast majority of the respective genes have high CAI values. This resulted in skewed median ranks of 7704 and 7958 (25–75th percentile ranks of 6732–8099 and 7304–8163, respectively) for pro- and amastigote datasets, indicating highly biased codon usage in genes encoding abundant proteins. This is in comparison the median rank of 4146.5 of the genome

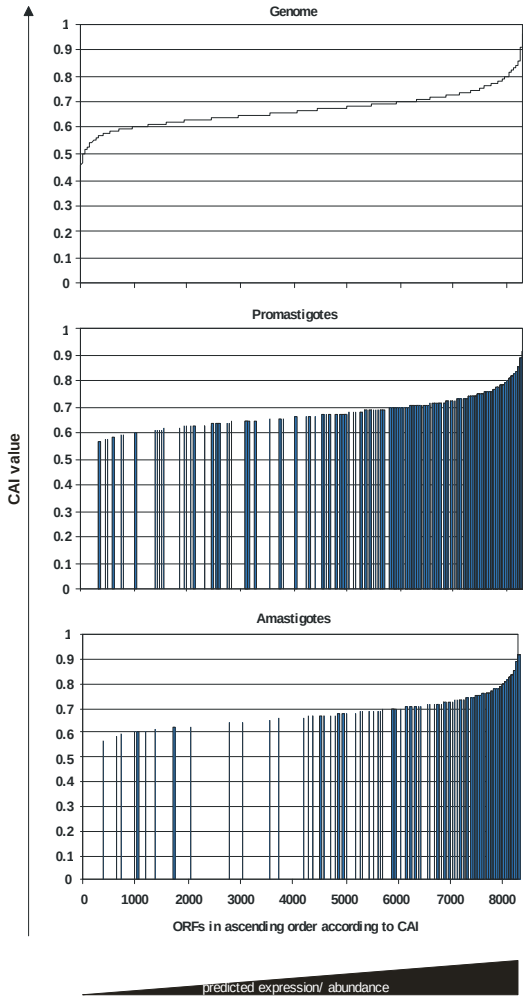


FIGURE 10. Codon bias in abundantly expressed ORFs.

Top) CAIs of all ORFs of the *L. major* (Friedlin) reference genome and of ORFs corresponding to identified proteins were plotted. *x axis*, 8292 ORFs of *L. major* in ascending order according to CAI value, *y axis*. Middle) CAI values of ORFs corresponding to proteins identified in the promastigotes or (bottom) amastigote proteome are shown; each *bar* represents one ORF. The median CAI value of the genome was 0.66308865; for ORFs constituting the promastigote and amastigote proteome the median CAI value was 0.762384 and 0.790777, respectively. The distribution of observed *versus* non-observed ranks was significantly different in promastigotes and amastigotes (Mann-Whitney test, $p < 0.0001$).

Furthermore, the likely subcellular localisation was analysed in the datasets and compared with the genome predicted proteins. Proteins were classified by Proteome Analyst 2.5, a sequence based predictive algorithm (Szafron, D. *et. al.*, 2004). It classifies proteins from eukaryotic organisms into nine classes (cytoplasm, nucleus, mitochondrion, endoplasmic reticulum, extracellular, Golgi apparatus, plasma membrane, peroxisome/glycosome, and lysosome). Proteins predicted to occur in more than one location were classified as ambiguous and proteins in which prediction was not possible (44% of the L. *major* ORFs) were omitted for clarity. This revealed that in this study, identified proteins predicted to be membrane-bound or secreted were under represented (see FIGURE 11). The former were expected to be under represented due to limitation in current sample preparation and mass spectrometric techniques. The latter were also expected to be under represented due to their low abundance within cells (Chevallet, M. *et. al.*, 2007) and due to lysis of the PV membrane during FACS as demonstrated through analytical EM pictures (Figure 2). However, secreted proteins would likely remain in the supernatant after pelleting of parasites from the collected FACS fraction.

As a result, soluble proteins with predicted cytoplasmic and mitochondrial localisation were significantly over represented. The frequency of proteins with predicted Golgi apparatus or endoplasmic reticulum localisation was as the frequency encoded in the genome. There was no significant under representation for nuclear localised proteins but a trend was evident.

Of note, the major surface protein of promastigotes, the glycosylphosphatidyl-inositol (GPI) anchored gp63, which makes up 1% of total protein in promastigotes (personal communication Dr T. Aebischer), was detected in multiple fractions but only by the same single peptide in the 1-DE-separated material (not shown). Amastigotes express a non-GPI anchored form gp63 which is localized in the lumen of lysosomal structures (Medina-Acosta, E. *et. al.*, 1989). This form was indeed detected solely in amastigote samples, see TABLE 5.

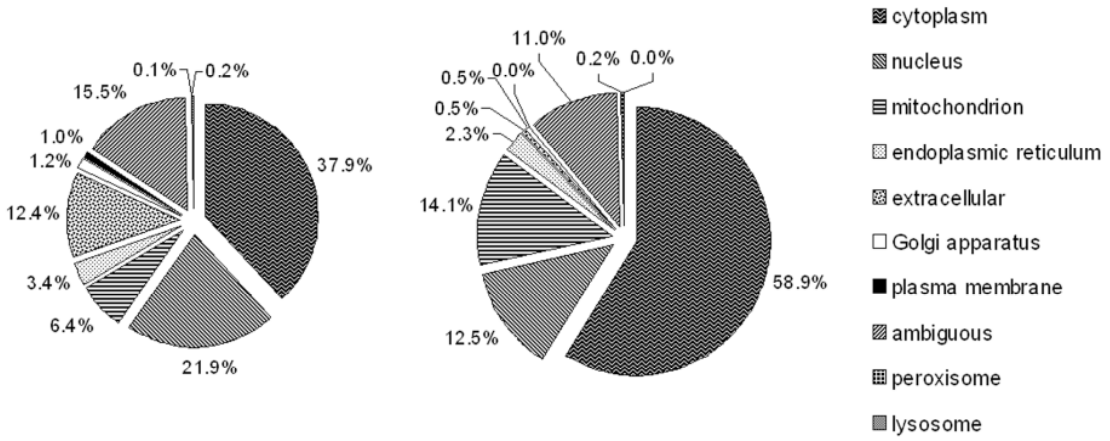


FIGURE 11. Pie charts of predicted subcellular localization of *L. major* reference genome ORFs *(left)* and ORFs corresponding to identified proteins from this study (*right*).

Proteins were classified by Proteome Analyst 2.5. The program classifies proteins into nine classes (cytoplasm, nucleus, mitochondrion, endoplasmic reticulum, extracellular, Golgi apparatus, plasma membrane, peroxisome/glycosome, and lysosome). Proteins predicted to occur in more than one location were classified as ambiguous. Proteins for which prediction was not possible (44% of the *L. major* ORFs) were omitted for clarity. Cytoplasmic and mitochondrial proteins were over represented ($p < 0.0001$). A trend for the under representation of nuclear localised proteins was apparent ($p = 0.1411$). Proteins localised in the Golgi apparatus ($p = 0.4739$) or endoplasmic reticulum (0.9439) were as their expected frequency in the genome. Proteins with extracellular ($p < 0.0001$) or plasma membrane ($p < 0.05$) localization were under represented. Statistical analyses were performed with χ^2–test.

In addition, identified proteins were classified into functional groups (carbohydrate metabolism, fatty acid metabolism, proteins of the respiratory chain/ energy metabolism, protein biosynthesis, ribosomal proteins, protein degradation, amino acid metabolism, cytoskeleton, chaperones/stress induced proteins, purine and pyrimidine metabolism, histones, antioxiants, hypothetical proteins and others) and the representation of these classes in the two life cycle stages was compared to the theoretical frequency in the genome (based on Gene DB dataset version 5.1; see FIGURE 12).

An under representation was observed for predicted hypothetical proteins in both experimental datasets (11.5% for amastigote and 17.1% for promastigote datasets versus 65.1% in the genome; $p < 0.0001$, χ^2–analysis) suggesting that most gene products in this class were present in low copy numbers per parasite cell. This is

supported by the CAI analysis which predicts translational bias. Hypothetical proteins showed a median CAI rank of 3323.5, which is significantly different from the median rank of 4146.5 in the genome ($p < 0.0001$, χ^2 analysis; 25–75th percentile ranks of 1614–5242 and 2073–6219 for hypothetical proteins and genome, respectively).

Omitting hypothetical proteins from the functional analysis, revealed that proteins with a role in protein degradation, amino acid metabolism, and purine/pyrimidine metabolism were under represented in the amastigote dataset as were proteins with predicted but not classified functions (Figure 12).

In comparison to their theoretical frequency in the genome (without considering hypothetical proteins) proteins with predicted but not classified functions were under represented ($p < 0.0001$) in both lifecycle stages. The statistical analyses were performed with χ^2–test. Proteins with a predicted role in protein degradation (trend), purine and pyrimidine metabolism (trend) were under represented only in amastigotes. Whereas proteins involved amino acid metabolism ($p < 0.005$) were over represented only in promastigotes compared to their frequency in the genome. Proteins with a predicted role in aerobic respiration/energy metabolism (trend) were identified more frequently only in amastigotes. Proteins involved in carbohydrate metabolism (both life cycle stages $p < 0.0001$), protein biosynthesis (both $p < 0.005$), ribosomal proteins (both $p < 0.0001$), cytoskeleton (amastigotes $p < 0.0005$, promastigotes $p < 0.05$), chaperones / stress induced proteins (both $p < 0.0001$), histones (both $p < 0.0001$) and antioxidants (trend in amastigotes, promastigotes $p < 0.005$) were over represented in both life cycle stages. Chaperones were also significantly more frequent in amastigotes ($p < 0.05$) when compared to the frequency in promastigotes.

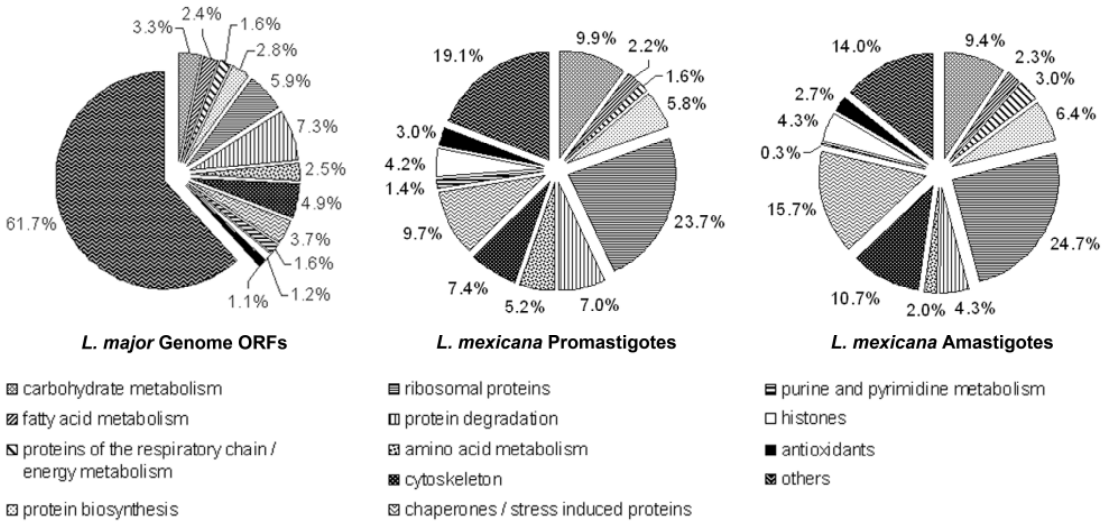


Figure 12. Predicted functional classification of *L. major* reference genome ORFs (*left*) and ORFs corresponding to proteins identified in *L. mexicana* promastigotes (*middle*) and amastigotes (*right*).

Functional classification was performed according to the Sanger database and "KEGG2." Classifications are shown below figures.

3.1.3 PROTEINS MORE ABUNDANT IN AMASTIGOTES

A series of prominent spots in the 25-kDa range (FIGURE 7; pI 5.6 and higher, white arrowheads left and middle panels) and numerous other spots, (e.g. black arrowheads in middle and right panels) were more abundant or solely detectable in amastigote samples. The white arrowheads denoting spots of a series in the range of 25 kDa and smaller, were identified as DsRed, the fluorescent marker protein used to purify amastigotes from infected cells as mentioned before (Paape, D. *et. al.*, 2008). Higher expression of DsRed reflects properties of the expression system driving 7-10-fold increased synthesis in amastigotes (Misslitz, A. *et. al.*, 2000). Due to the increased synthesis in amastigotes purified from bmdms (FIGURE 7, middle panel) various breakdown products of DsRed were detected with a molecular weight below 25 kDa.

In total 51 protein species were detected only in amastigotes, either in the 1D gel electrophoresis or 2-DE approach (suppl. Tables 1, 2). The variation of the LC-ESI MS/MS approach is higher due to the complex sample composition. In contrast, samples are much less complex in 2-DE combined with MALDI MS/MS, hence less variation in data acquisition. Because samples of the 1-DE were only analysed once by LC-ESI MS/MS, a comparison of promastigote and amastigote samples

was not performed. Only identifications of the 2-DE analyses were considered for analysis of differential protein expression. This revealed that 45 protein spots, corresponding to 34 proteins, were only detectable in the amastigote probes (indicated in FIGURE 7 by black arrowheads; see also TABLE 5). It was possible that these proteins were expressed in promastigotes but escape detection because of low abundance. However, the proteins listed in TABLE 5 represent a differentially expressed set of proteins and included cysteine proteinase B locus products, which are known to be more abundant in amastigotes (Mottram, J. C. *et. al.*, 1997; Souza, A. E. *et. al.*, 1992). The differentially expressed set also contained proteins involved in fatty acid and energy metabolism; proteins with predicted roles in RNA and protein processing; and chaperones. In addition among the differentially expressed protein species were seven hypothetical proteins with unknown function, see TABLE 5.

Previous comparative proteome studies of promastigotes and axenic amastigotes of *Leismania* spp., including *L. mexicana,* identified several protein species of HSPs, β-tubulin, and the eEF-1α subunit interpreted of proteolytic processing as stage specific. This study also observed stage specific protein species for eEF-1α (suppl. FIGURE 2, spots 162 and 216) as well as HSP (suppl. Fig. 2, spot 84) in amastigotes, but these were not identical to those previously reported from axenic amastigote.

TABLE 5. Proteins identified solely in amastigotes

Sanger ID	NCBI Acc.-No.	Spot. No.[a]	Annotation
LmjF08.0320		245	mitochondrial associated ribonuclease
LmjF09.0100		10	hypothetical protein, conserved
LmjF18.0580		169	peroxisomal enoyl-coa hydratase
LmjF19.0710		213, SSP_8307	**glycosomal malate dehydrogenase**
LmjF21.1710		296	**COX6 cytochrome C oxidase**
LmjF21.1760		112	centromere/microtubule binding protein cbf5
LmjF22.1540		SSP_4801	alanyl-tRNA synthetase
LmjF23.0370		287, SSP_7006	hypothetical protein, conserved
LmjF23.0760		164	mitochondrial RNA binding protein
LmjF24.1210		SSP_9004	translation factor SUI1
LmjF25.2130[b]		160, SSP_8408	**succinyl-CoA synthetase**
LmjF25.2140[b]		160, SSP_8408	**succinyl-CoA synthetase**
LmjF26.1550		97	**trifunctional enzyme alpha subunit**
LmjF27.1110		267, SSP_8207	mitochondrial RNA binding protein 1
LmjF27.1220		122, SSP_6603	hypothetical protein, conserved
LmjF27.1300		101, 102	hypothetical protein, conserved
LmjF28.0490		107	propionyl-coa carboxylase beta chain
LmjF29.0760		2	lipophosphoglycan biosynthetic protein
LmjF31.1630[b]		175	**putative 3-ketoacyl-coa thiolase-like**
LmjF31.1640[b]		113, 175, SSP_8502	**thiolase protein-like**
LmjF31.1900[b]		289	ubiquitin-fusion protein
LmjF31.2030[b]		289	ubiquitin-fusion protein
LmjF31.2250		70	**3,2-trans-enoyl-CoA isomerase**
LmjF32.0840		40	hypothetical protein, conserved
LmjF33.0830		98	**2,4-dienoyl-coa reductase**
LmjF33.1630		268	Cyclophilin
LmjF33.2390		SSP_4808	heat shock protein
LmjF34.3670		312	vacuolar ATP synthase
LmjF35.1300		SSP_4106	ubiquitin-conjugating enzyme E2
LmjF36.1160		SSP_6414	hypothetical protein, conserved
LmjF36.3100		242	**ATP synthase**
LmjF36.3780		SSP_5302	hypothetical protein, conserved
LmjF36.4360		151	proteasome regulatory ATPase subunit
	gi|68223993	SSP_5402	RNA helicase
	gi|14348750[b]	223, 225, 226, SSP_0103	CPB2 protein
	gi|1749812[b]	226, SSP_0103	CPB1
	gi|886681[b]	223, 225, 226, SSP_0103	CPb19
	gi|1730100[b]	223, 225, 226, SSP_0103	Cysteine proteinase B precursor
	gi|2780176[b]	SSP_0103	cystein proteinase
	gi|461905[b]	SSP_0103	Cysteine proteinase 2 precursor
	gi|9542[b]	SSP_0103	cysteine proteinase
	gi|51317309	41	Leishmanolysin
	gi|146077037	90	stomatin-like protein

[a] 45 protein spots identified in 2-DE analysis represented 34 proteins solely identified in amastigotes. Spot numbers refer to numbers indicated in supplementary figures 1 - 3 representing gel images and tables detailing identification results. GeneDB Systematic IDs and NCBI accession numbers are indicated. Proteins involved in carbohydrate, fatty acid, or energy metabolism are highlighted in **bold**.

[b] Proteins that could not be assigned based on identified peptides to a single ORF but to tandem arrays of ORFs. In this case all corresponding IDs or accession numbers are indicated but were counted as one distinct identification only.

3.1.4 COMMON SIGNATURES IN 3'-UNTRANSLATED MRNA SEQUENCES OF DIFFERENTIALLY EXPRESSED ORFS

As mentioned above, genes in trypanosomatids are organized in constitutively transcribed polycistronic units (Leifso, K. *et. al.*, 2007). Regulatory sequences in 3'-untranslated regions (UTRs) are known to impact gene expression levels in trypanosomatids (Clayton, C. E. 2002). By adapting a nucleotide counting algorithm, such regulatory elements were recently identified in genes differentially expressed in trypanosomes (Mayho, M. *et. al.*, 2006). It was assumed that the higher expression of proteins in amastigotes (TABLE 5) was partially due to the presence of similar regulatory elements in specific sequences in the 3'-UTR of mRNA. Therefore, intergenic sequences 3' of the respective genes were analysed by the nucleotide counting algorithm. Loci encoding proteins found to be more abundant in amastigotes by quantitative analysis using isotope-coded protein labeling were also included[5]. The analysis was performed on a total set of 51 intergenic sequences from *L. major* (see suppl. Table 2 for the respective loci). A random selection of intergenic regions of 30% of all loci in the genome served as a control set. The oligonucleotide sequences AAGAGAA and/or TCTCCTTT were found within predicted 3'-UTRs in 21 (see TABLE 6) of the 51 test sequences; this was significantly different from the control 3'-UTR sequence set ($p < 0.0001$ and $p = 0.0012$ for TCTCCTTT and AAGAGAA, respectively). Furthermore a comparative analysis of the respective 3'-UTR of homologous genes in *L. infantum* indicated a high degree of conservation of the signatures between these species. After the motifs were identified, it was further verified whether known experimentally analysed stage-specific gene, such as *L. infantum* amastin (LinJ34.0840; (McNicoll, F. *et. al.*, 2005), contain either motif. The 3'- UTR of amastin was scanned for these motifs. And the 3'-UTR the pyrimidine-rich motif was also found. In the amastin 3'-UTR, the motif TCTCCTTT starts 1587 bp downstream of the stop codon in a region shown to regulate amastigote-specific expression at a translational level in response to the temperature shift when switching to mammalian hosts.

[5] Anja Freiwald ,Etablierung der Markierungsmethode ICPL für Proteomics am Mikroorganismus *Leishmania mexicana*'. Masters of Science thesis, Technische Fachhochschule Berlin in 2006

Two sequence motifs in the 3-UTR of loci of in amastigotes higher expressed proteins were identified. These motifs could be linked to the regulation of their expression as presence of one of the two motifs in the 3' UTR of amastin an experientially known amastigotes specific gene indicates.

TABLE 6. Sequence motifs in 3'-UTR characteristic of differentially expressed ORFs

Below are the GeneDB IDs (http://www.genedb.org/) of *L. major* loci that encode gene products homologous to the 51 differentially expressed proteins detected in *L. mexicana* amastigotes and in which the predicted 3'-UTR contained one or both of two sequence motifs (denoted by x) that were significantly more frequently present when compared with a large control set of 30% of the genomic loci. For a list of all 51 loci in *L. major* and 45 homologous loci in *L. infantum* analysed, see suppl. Table 2. In 67% and 64% of cases, respectively, the motifs TCTCCTTT and AAGAGAA were at the same relative position within the predicted 3'-UTRs. In **bold**) loci for which proteins were found by isotope coded protein labelling to be more abundant in amastigotes[6].

L. major				*L. infantum*		
Sanger ID	Annotation	Motif		Sanger ID	Motif	
		TCTCCTTT[a]	AAGAGAA[b]		TCTCCTTT	AAGAGAA
LmjF09.0100	Hypothetical protein, conserved	x		LinJ09_V3.0120		
LmjF19.0710	Glycosomal malate dehydrogenase		x	LinJ19_V3.0710		x
LmjF23.0370	Hypothetical protein, conserved	x	x	LinJ23_V3.0420	x	x
LmjF23.0760	Mitochondrial RNA binding protein	x		LinJ23_V3.0930	x	
LmjF24.1210	Translation factor SUI1		x	LinJ24_V3.1240		x
LmjF25.1420	**GTP-binding protein**	**x**	**x**	**LinJ25_V3.1460**	**x**	**x**
LmjF25.2130	Cuccinyl-CoA synthetase α subunit			LinJ25_V3.2220	x	
LmjF26.1550	Trifunctional enzyme α subunit	x	x	LinJ26_V3.1530	x	x
LmjF27.1110	Mitochondrial RNA binding protein 1	x		LinJ27_V3.0980	x	
LmjF27.1220	Hypothetical protein, conserved	x	x	LinJ27_V3.1100		x
LmjF27.1300	Hypothetical protein, conserved	x	x	LinJ27_V3.1220	x	x
LmjF28.2770[c]	**Heat-shock protein hsp70**	**x**	**x**	**LinJ28_V3.3060**		
LmjF28.2780[c]	**Heat-shock protein hsp70**	**x**		**LinJ28_V3.2960**		
LmjF32.0840	Hypothetical protein, conserved		x	LinJ32_V3.0890	x	x
LmjF33.0830	2,4-dienoyl-coa reductase	x		LinJ33_V3.0870	x	
LmjF33.2390	Heat shock protein	x		LinJ33_V3.2520	x	
LmjF34.2580	**Hypothetical protein, conserved**	**x**	**x**	**LinJ34_V3.2410**	**x**	**x**
LmjF35.1300	UYbiquitin-conjugating enzyme E2		x	LinJ35_V3.1310		x
LmjF36.1160	Hypothetical protein, conserved	x	x	LinJ36_V3.1220	x	
LmjF36.3100	ATP synthase			LinJ36_V3.3250		x
LmjF36.3780	Hypothetical protein, conserved		x	LinJ36_V3.3970		x
LmjF36.3990	**hs1vu complex proteolytic subunit-like**	**x**	**x**	**LinJ36_V3.4180**	**x**	
LmjF36.4360	Proteasome-regulatory ATPase subunit	x		LinJ36_V3.4570	x	x

[a] $p < 0.0001$ by X^2 test for the indicated motif,
[b] $p = 0.0012$ by X^2 test for the indicated motif,
[c] belong to the same gene array.

[6] Anja Freiwald ,Etablierung der Markierungsmethode ICPL für Proteomics am Mikroorganismus *Leishmania mexicana*'. Masters of Science thesis, Technische Fachhochschule Berlin in 2006

3.2 SECRETED PROTEINS OF *LEISHMANIA MEXICANA* AMASTIGOTES

3.2.1 ANALYSIS OF *L. MEXICANA* AMASTIGOTE PROTEOME DATASET

Analytical electron micrographs showed that FACS not only enriched amastigotes but also fragmented the parasitophorous vacuole (PV) membrane (Figure 2). Shear forces during sorting were likely responsible for rupture of the PV. Therefore, it was hypothesised that not only parasites but also luminal and membrane proteins of the PV, as well as material of lysed parasites, were contained in the very same fraction/volume as the sorted parasites. The luminal fraction of the PV would in addition also contain proteins secreted by amastigotes. Therefore, the fluorescence based method to purify amastigotes from *in vitro* or *in vivo* infected host cells can be applied again. Parasites would be removed, from the sort supernatant, by centrifugation leaving the supernatant likely containing the above mentioned proteins.

Bone marrow derived MΦ were infected with axenic amastigotes for 3 h and amastigotes/phagosomes were purified as shown in Figure 2. After sorting, intact parasites were separated by low centrifugal force centrifugation from proteins in the supernatant. The latter assumed to represent parasite secreted proteins, luminal proteins and membrane fragments of the PV, were precipitated as described by Caldwell *et al.* (Caldwell, R. B. and Lattemann, C. T. 2004). Pelleted proteins were reduced, alkylated and trypsin digested. Peptides were then separated by strong cation exchange chromatography. Fractions (in total 22) were analysed using an automated micro-LC-MS/MS system (consised of a 1200 series micro-WPS autosampler and a binary micro-pump (Agilent Technologies, Palo Alto, CA) coupled to an LTQ-Orbitrap hybrid mass spectrometer (ThermoFisher, San Jose, CA)[7].

Protein identities were assigned by searching against a comprehensive *Leishmania* spp. database (Material and Methods). In total 1764 distinct *Leishmania* proteins were identified of which 1689 were identified by two or more peptides (suppl. Table 3) Identifications were compared with the *L. mexicana* promastigote and amastigote proteomes, presented in the previous chapter and with a recently

[7] Obtaining phagosomes and processing thereof as well as acquiring mass spectrometric data was done by Martin Barrios-Llerena. Analysis of data was part of this PhD

published dataset (Rosenzweig, D. *et. al.*, 2008). The proteome study presented in the previous chapter, and the study of Rosenzweig and colleagues, were until then the most comprehensive. Comparison with these two datasets revealed that 763 proteins were newly identified (FIGURE 13), but failed to identify 92 and 670 proteins of the *L. mexicana* (Paape, D. *et. al.*, 2008) and *L. infantum* (Rosenzweig, D. *et. al.*, 2008) datasets, respectively.

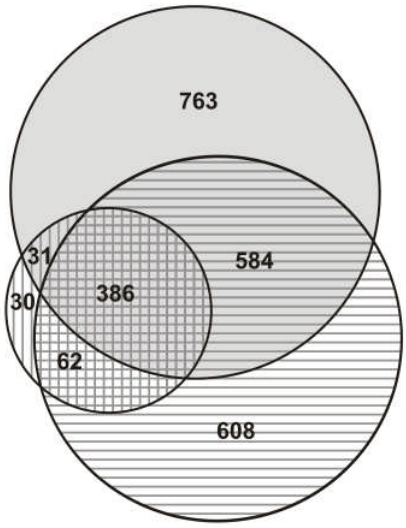


FIGURE 13. Comparison of identified proteins with two published datasets.

Grey shaded circle contains proteins identified in this study, in total 1764. Vertical hatched circle contains identifications, 509 in total, of Paape et al. (Paape, D. *et. al.*, 2008). Horizontal hatched circle contains assigned *L. major* orthologues; in total 1640 of 1697 retrievable *L. infantum* identifications, originally reported by Rosenzweig *et al.* (Rosenzweig, D. *et. al.*, 2008).

Comparison to other Leishmania proteome studies (Bente, M. *et. al.*, 2003; Dea-Ayuela, M. A. *et. al.*, 2006; Drummelsmith, J. *et. al.*, 2003; El Fakhry, Y. *et. al.*, 2002; Foucher, A. L. *et. al.*, 2006; Gongora, R. *et. al.*, 2003; McNicoll, F. *et. al.*, 2006; Nugent, P. G. *et. al.*, 2004; Silverman, J. M. *et. al.*, 2008) revealed that 741 were not reported so far.

As before proteins with a predicted membrane localisation were under represented when compared to their expected frequency in the genome. The number of identified proteins with predicted trans-membrane (TM) domains per 1000 identified proteins, was calculated and plotted as a function of the number of predicted TM domains (FIGURE 14). Between the three studies (Paape *et al.*,

Rosenzweig *et al.* and the study presented here) no significant differences were detected by χ^2-analysis. The putative membrane proteins were identified with similar proportions as in the first proteome study and in the dataset of Rosenzweig *et al.* but again under represented indicating a technical bias.

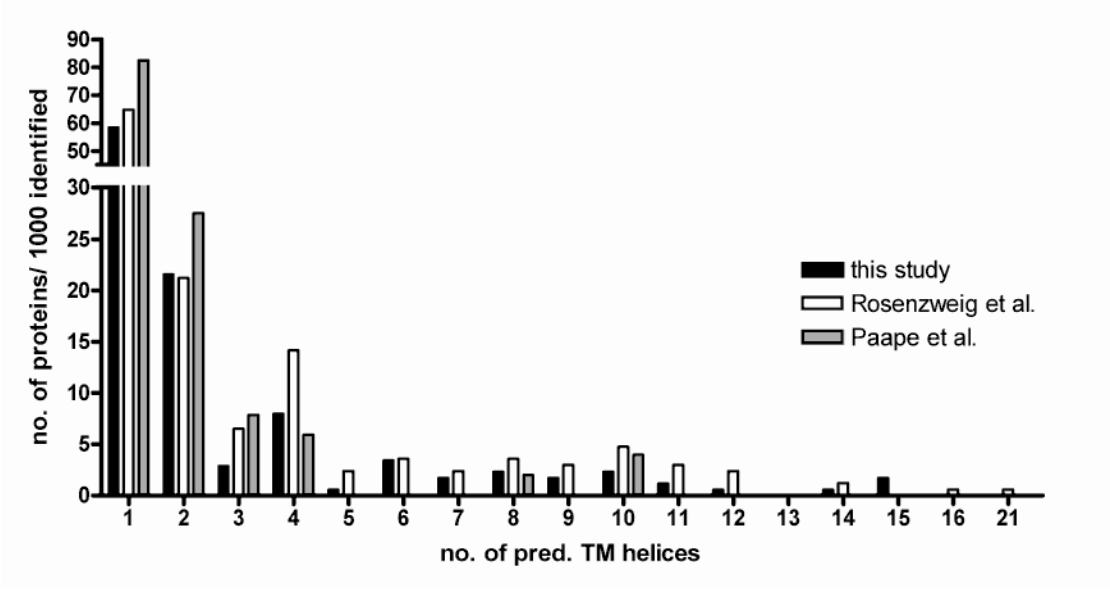


FIGURE 14. Distribution of identified proteins containing one or more predicted trans-membrane helices (TM).

No significant difference between datasets (χ^2-analysis).

Similar to the analysis of the *L. mexicana* pro- and amastigote dataset, identified proteins were classified into the 14 functional groups (FIGURE 15).

Hypothetical proteins of the amastigote proteome were again under represented compared to their frequency in the genome, with a frequency of only 43.8 % in amastigotes compared to 65.1 % in the *L. major* genome ($p < 0.0001$). Because hypothetical proteins account for nearly half of the identified proteins in the amastigotes and almost two thirds in the *L. major* genome, they were omitted for the following analyses but included in FIGURE 15.

Compared to the theoretical frequency in the genome without hypothetical proteins, proteins involved in carbohydrate metabolism ($p < 0.0001$), energy metabolism ($p < 0.05$) protein degradation ($p = 0.0068$), purine and pyrimidine metabolism (trend), chaperones ($p < 0.05$), protein biosynthesis ($p < 0.005$) and amino acid metabolism ($p = 0.0069$) were over represented. Cytoskeletal proteins

(p = 0.0047), histones (trend) as well as predicted proteins with not classified function (p = 0.0002) were under represented in amastigotes.

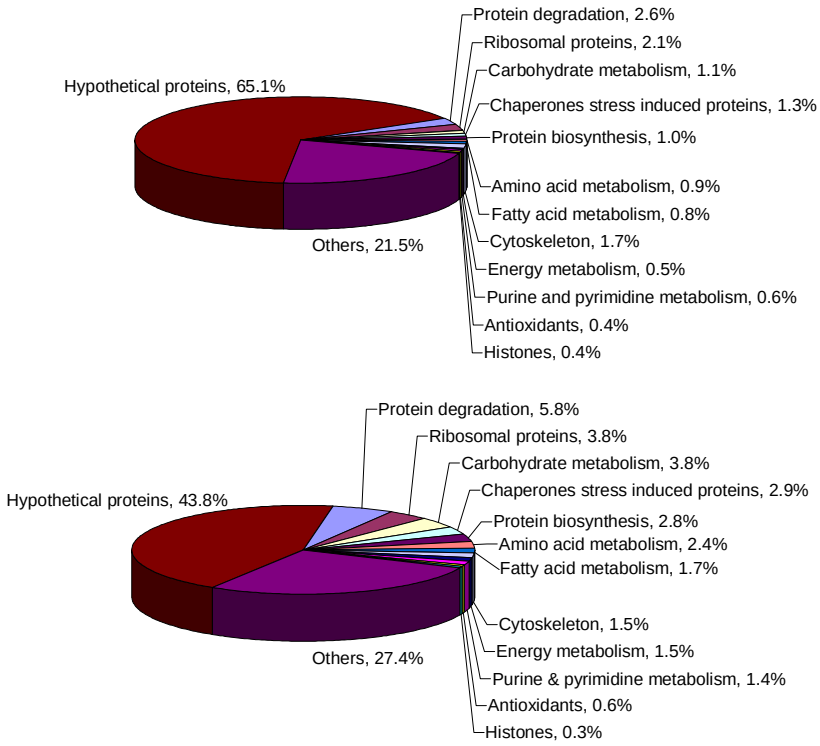


FIGURE 15. Representation of predicted functional classification.

Top) predicted proteins, encoded by *L. major* reference genome; **Bottom)** identified in *L. mexicana* amastigotes proteome. Prediction of functional classification was performed using GeneDB database (http://www.genedb.org/genedb/leish/index.jsp) and 'KEGG2' (www.genome.jp/kegg/kegg2.html).

An under representation of a protein class may be related to technical bias (e.g. low abundance) such as noted for membrane proteins.

Based on MS/MS data a label-free method can be used to calculate relative approximate quantities can expressed as an exponentially modified protein abundance index (emPAI; (Ishihama, Y. *et. al.*, 2005). This index takes into account the number of observed peptides and the number of observable peptides for a given protein under given MS/MS analysis parameters. Consequently the emPAI for every identified protein was calculated. The resulting ranking covered a range of approximately 2 orders of magnitude of relative abundance. More than two thirds of hypothetical proteins were in the lower half of the ranked emPAIs, which reflected their low abundance, hence under representation was expected. This

corroborates the conclusion from the CAI analysis (see above). The emPAI confirm that hypothetical proteins are on average of low abundance. However, among the 50 most abundant proteins detected in *L. mexicana* amastigotes were three homologues of *L. major* hypothetical proteins, LmjF34.2580, 30.3310 and 30.2845. The predicted translational bias is further supported by the fact that 8 out of 10 hypothetical proteins with the highest emPAI having also a high CAI value, hence predicted to be more abundant. That means the emPAI confirmed the CAI values, thus the translational bias in *Leishmania*.

3.2.2 PROTEOME DERIVED PREDICTIONS FOR AMASTIGOTE METABOLISM

Recently, metabolic pathway knowledge was integrated with Leishmania genome information to create LeishCyc as a resource (Doyle, M. A. *et. al.*, 2009). It represents metabolic potential of *L. major*, as predicted from the genome, in an interactive diagram and allows integration of additional data such as relative abundance of proteins with metabolic functions or metabolites. A diagram, visualising enzymes present in the proteome of amastigotes purified from 3 h infected bmdms and indicating their abundance using a colour code reflecting emPAI values was generated FIGURE 16. This representation immediately revealed that enzymes of the carbohydrate (1) and energy (2), metabolism were abound and enzymes involved in the β-oxidation of fatty acids (3 in FIGURE 16) were also amongst the most abundant proteins identified. Moreover, the pathway diagram showed an obvious absence of enzymes involved in fatty acid elongation (4), the synthesis of lipophilic compounds including glycoinositolphospholipid (GIPL) (5), GPI-anchor (6) and lipophosphoglycan (LPG) synthesis (7) as well as ubiquinone synthesis (8). This may be a direct consequence of still inadequate detection of proteins strongly attracted or integral to membranes by current MS/MS approaches.

Glycolytic and gluconeogenic enzymes are mainly the same as they carry out reversible reactions. Here, it was noted that emPAI values for gluconeogenic enzymes carrying out irreversible reactions, phosphoenolpyruvate carboxykinase and pyruvate phosphate dikinase generating phosphoenolpyruvate from oxalacetate and pyruvate, respectively, as well as fructose-1,6-bisphosphatase generating fructose-6-phosphate, had by the majority comparably higher emPAI

values than the glycolytic enzymes catalysing the reverse reactions. Further, enzymes such as phosphomannose isomerase and mannose-1-phosphate guanylyltrans-ferase utilising fructose-6-phosphate as a substrate for anabolic processes to generate GDP-mannose. GDP-mannose is the donor for the biosynthesis of several glycoconjugates in *Leishmania,* such as GPI anchor, LPGs and GIPLs (FIGURE 17). This shows the advantage of proteomics, which can in combination with the calculated relative abundance predict prevailing metabolic pathways. This is vital, in order to identify new candidates for drug development.

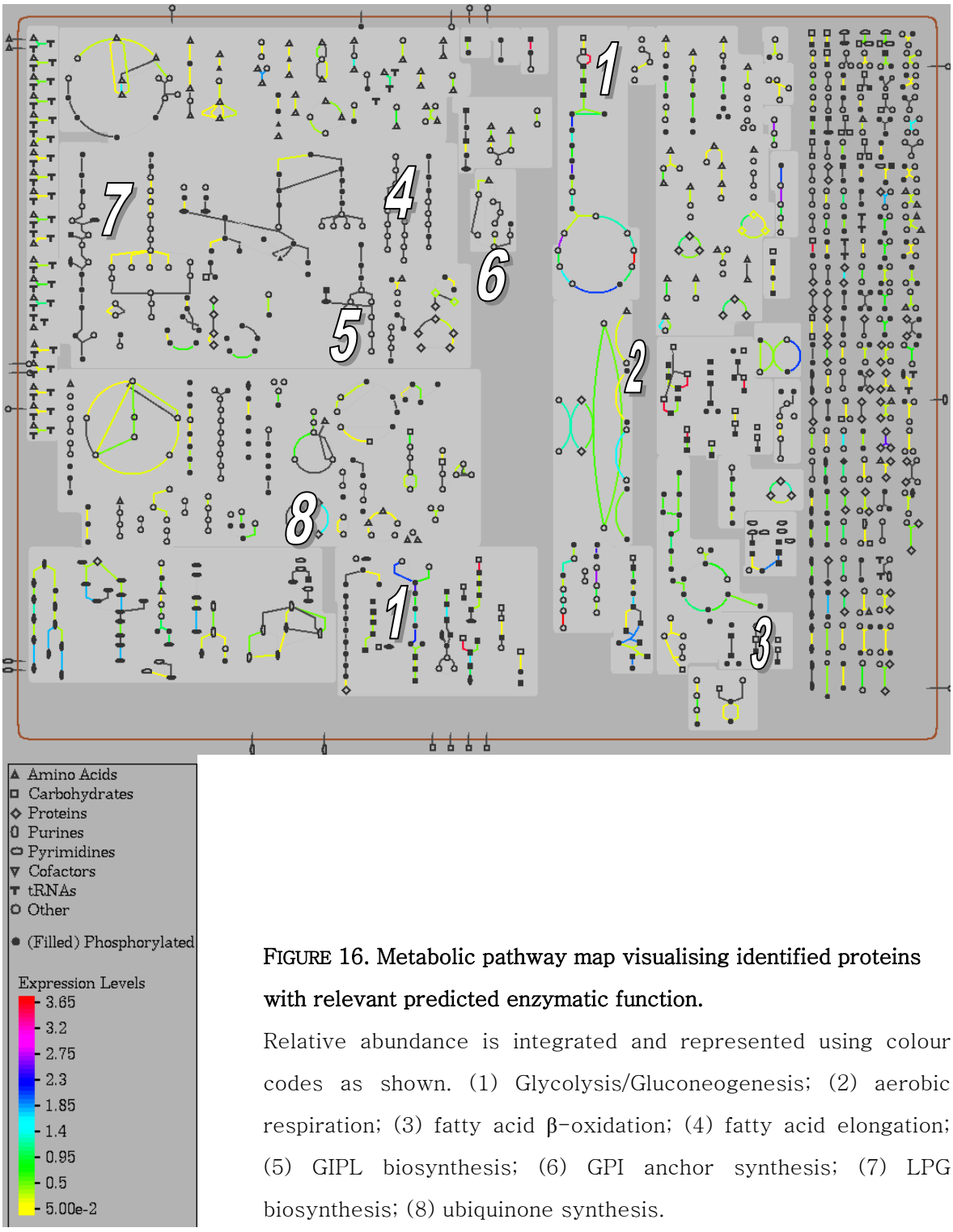


FIGURE 16. Metabolic pathway map visualising identified proteins with relevant predicted enzymatic function.

Relative abundance is integrated and represented using colour codes as shown. (1) Glycolysis/Gluconeogenesis; (2) aerobic respiration; (3) fatty acid β-oxidation; (4) fatty acid elongation; (5) GIPL biosynthesis; (6) GPI anchor synthesis; (7) LPG biosynthesis; (8) ubiquinone synthesis.

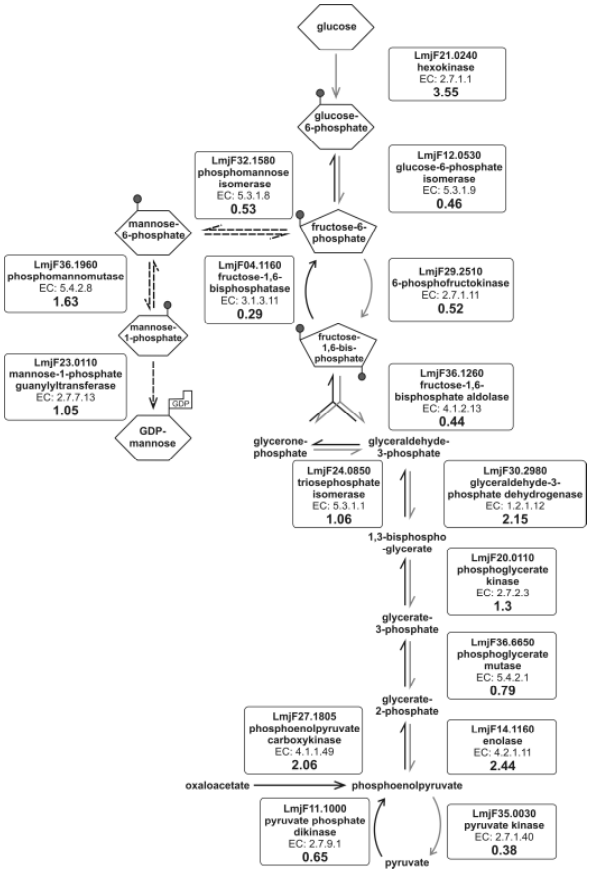


FIGURE 17. Comparative analysis of gluconeogenic and glycolytic pathways.

Enzyme names are indicated as are Sanger ID, EC number and emPAI values (**bold**). Grey and black arrows denote catabolic glycolysis and anabolic gluconeogenesis, respectively. Dashed arrows denote synthesis of GDP-mannose. Oxaloacetate, of the citric acid cycle, or pyruvate, are gluconeogenic precursors of fructose-6-phosphate which may be utilised by phosphomannose isomerase (LmjF32.1580) and converted into mannose-6-phosphate.

3.2.3 GENOME DISTRIBUTION OF ORFS ENCODING ABUNDANT PROTEINS

In trypanosomatids, to which *Leishmania* spp. *Trypanosoma* brucei and *T. cruzi* belong, genes are organized in polycistronic units that seem to be constitutively transcribed (Leifso, K. *et. al.*, 2007). *Leishmania major* has 133 polycistronic units (Ivens, A. C. *et. al.*, 2005) and it was hypothesised that variation in expression levels of these units could contribute to differential protein representation, i.e. abundant proteins may cluster to particular polycistronic units.

A tool[8] was developed to explore this hypothesis. The tool visualised the chromosomal localisation of ORFs corresponding to identified proteins. Datasets from both proteome studies carried out here were combined and loci of

[8] Visualisation tool was programmed in JAVA by Benjamin Arndt under the supervision of Klaus Peter Pleissner, both Max Planck Institute for Infection Biology, Charitéplatz 1, 10117 Berlin,Germany

amastigotes and promastigotes were visualised on the chromosome and their distribution analysed (FIGURE 18). This revealed that identified ORF are distributed over all polycistronic units. Loci encoding proteins identified do not cluster in specific units. However, there were units, or regions within units which did not contain any identified ORF, as highlighted in FIGURE 18. These regions mainly contained genes encoding hypothetical proteins (highlighted regions in FIGURE 18) and / or proteins involved in phosphoglycan biosynthesis (chr. 2), surface antigen proteins (chr. 12), proteophosphoglycans (chr. 35) as well as amastin and tuzin-like proteins (chr. 34). This also reflects the technical bias mentioned previously.

FIGURE 18. Screen shot of proteome-to-genome visualization tool.

Chromosomes 1–36 are plotted with the predicted 133 polycistronic clusters indicated as arrows in light green and light gray. Black marks show ends of clusters. Boxes localise individual loci of identified proteins. **Purple)** ORFs corresponding to proteins identified in both life cycle stages; **Blue**) ORFs corresponding to proteins identified only in amastigotes; **Red**) ORFs corresponding to proteins identified only in promastigotes. Red encircled areas do not contain any identified ORF, see text for details.

3.2.4 BIOINFORMATIC ANALYSIS OF POTENTIALLY SECRETED PROTEINS

It was hypothesised that the supernatant of FACS sorted material contained phagosomal, luminal, membranous, as well as those proteins secreted by the amastigote. The latter (which are a focus of this thesis) were probably not detected in the first proteome study. Therefore, the dataset was analysed in order to identify putatively secreted proteins. Firstly a putative mitochondrial localisation was excluded. Then the presence of a signal peptide and the absence of TM-domains and / or GPI-anchor addition sites was analysed. For the localisation prediction TargetP 1.1 (Emanuelsson, O. *et. al.*, 2000). TargetP assigns a reliability class (1 to 5) for its prediction which certainty. A reliability class of 3 means, within a trainings set of proteins with known localisation, 83% were predicted correctly.

Predicted mitochondrial localised proteins were excluded. By TargetP predicted secreted, all for which a prediction was not possible as well as all with a reliably class > 3, hence the prediction unlikely to be correct, a signal peptide prediction was performed with SignalP 3.0 (Nielsen, H. and Krogh, A. 1998; Bendtsen, J. D. *et. al.*, 2004). If a signal peptide was predicted it was verified that neither a trans-membrane domain (predicted with TMHMM 2.0 (Krogh, A. *et. al.*, 2001) nor a GPI-anchor could be predicted. Proteins predicted to have a GPI anchor were retrieved from GeneDB (http://www.genedb.org/genedb/leish). In total 143 proteins fulfilled the criteria (Signal peptide present, TM-domain and no GPI-anchor addition site absent) see suppl. TABLE 4. Of these 67 had never been reported before. The proportions of these putatively secreted proteins in relation to all proteins identified were plotted and compared to previously published datasets, see FIGURE 19. The proportion of signal peptide positive proteins was significantly higher in this study ($p = 0.0042$, χ^2-analysis) compared to 92 of 1697 predicted secreted in the *L. infantum* dataset reported by Rosenzweig *et al.*

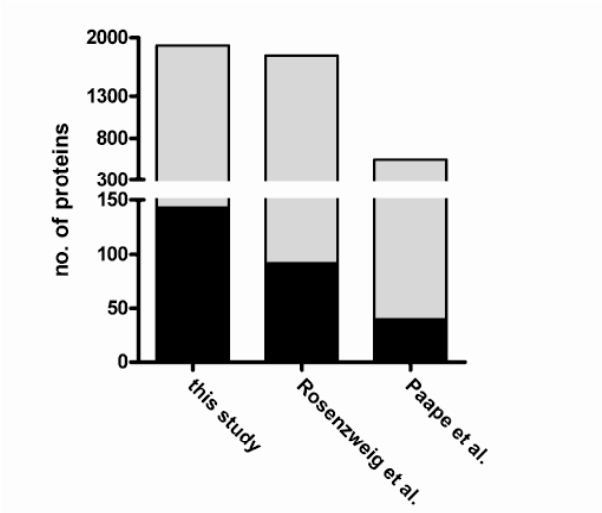


FIGURE 19. Comparative analysis of predicted secreted proteins detected in three proteomic studies.

Grey bars) reflect total number of proteins identified in the respective study; **black bars**) predicted secreted proteins. There is a significant difference between this study and Rosenzweig's data (Rosenzweig, D. *et. al.*, 2008) in identification of possibly secreted proteins $p = 0.0042$ (χ^2-analysis).

This analysis was performed because proteins secreted by parasites are considered good vaccine antigen candidates. Furthermore, a good vaccine candidate has minor or no homology with host proteins. This means that secreted proteins of unknown function are predestined vaccine antigen candidates. In addition protein abundance is recognised to contribute to immunogenicity (Overath, P. and Aebischer, T. 1999). Hence, potentially secreted proteins were ranked according to their relative abundance, i.e. their emPAI, and the 20 most abundant hypothetical proteins were listed in, TABLE 7 (suppl. TABLE 4). This list may therefore hold many novel vaccine candidates.

TABLE 7. The 20 most abundant hypothetical proteins putatively secreted.

Sanger accession no.	Gene name	Coverage (%)	Peptide matches	emPAI
LmjF08.1100	hypothetical protein, conserved	24.7	34	1.29
LmjF24.2110	hypothetical protein, conserved	30.5	41	0.97
LmjF36.4460	hypothetical protein, conserved	21.9	12	0.73
LmjF24.2160	hypothetical protein, conserved	16.8	13	0.68
LmjF30.1040	hypothetical protein, conserved	18.8	11	0.52
LmjF16.0520	hypothetical protein, conserved	13.4	11	0.31
LmjF18.0010	hypothetical protein, conserved	8.4	17	0.25
LmjF29.2100	hypothetical protein, conserved	7.2	14	0.24
LmjF21.0310	hypothetical protein, conserved	8.4	4	0.24
LmjF29.0320	hypothetical protein, conserved	6.9	8	0.23
LmjF36.0590	hypothetical protein, conserved	9.1	2	0.23
LmjF06.1240	hypothetical protein, conserved	10.3	6	0.21
LmjF36.4860	hypothetical protein, conserved	7.7	3	0.20
LmjF16.1065	hypothetical protein, conserved	6.0	10	0.19
LmjF36.5460	hypothetical protein, conserved	10.3	10	0.19
LmjF09.0840	hypothetical protein, conserved	7.7	5	0.19
LmjF14.0440	hypothetical protein, conserved	11.1	6	0.18
LmjF05.0490	hypothetical protein, conserved	6.9	2	0.18
LmjF36.4670	hypothetical protein, conserved	7.0	2	0.18
LmjF14.1330	hypothetical protein, conserved	8.0	4	0.17

3.3 RELEVANCE OF 2,4-DIENOYL-COA-REDUCTASE OF *LEISHMANIA* FOR VIRULENCE

3.3.1 BIOINFORMATIC ANALYSIS OF 2,4 DIEOYL-COA-REDUCTASE

Comparison of the *L. mexicana* promastigote and amastigote proteome revealed that several proteins involved in fatty acid metabolism were solely detected in amastigotes (see TABLE 5). Among these were several which were involved in catabolic fatty acid metabolism, namely LmjF26.1550, trifunctional enzyme alpha subunit; LmjF31.1630, putative 3-ketoacyl-CoA thiolase-like; LmjF31.1640, thiolase protein-like; LmjF31.2250, 3,2-trans-enoyl-CoA isomerase and LmjF33.0830, 2,4-dienoyl-coa reductase. BLAST analysis revealed, that the latter protein was most closely related to homologues of other kinetoplastids, followed by the NADPH-dependent 2,4-dienoyl-CoA reductase of *Corynebacterium lipophiloflavum* but has no homology with mammalian (i.e. human, mouse, rat) proteins of similar function. The other mentioned enzymes involved in fatty acid metabolism were homologous to both, prokaryotic and mammalian proteins of the same activity/function. This led to the hypothesis, that DECR is a potential drug target and that it is maybe possible to develop a specific compound to inhibit only DECR of *Leishmania* spp., because it is only homologous to prokaryotic and not to mammalian proteins. Therefore, the relevance of DECR of *L. major* and *L. mexicana* was going to be investigated, and *decr*-deficient parasites were to be constructed and analysed for their virulence and pathogenicity.

The *decr* gene of *L. mexicana* was amplified with primers designed based on the *L. major* LmjF33.0830 sequence. The product was sequenced and a nucleotide and protein NCBI-BLAST search was performed. This revealed that the *L. mexicana* nucleotide and protein sequence is related closest to the gene and protein of *L. infantum*, (TABLE 8 and TABLE 9 for details). Furthermore, the protein sequence of *L. mexicana* is more closely related to *fadh* (the gene encoding for 2,4-dienoyl-CoA reductase in bacteria), of *Marinobacter aquaeolei* than to *Corynebacterium lipophiloflavum*. Blast analysis also identified the conserved domain 'cd02930' in the *L. major* and *L. mexicana* sequences. This domain denotes a 2,4-dienoyl-CoA reductase FMN-binding domain.

TABLE 8: NCBI-Blast result of nucleotide sequence of *L. mexicana* 2,4 Dienoyl-CoA Reductase

Accession	Description	Total score	Query coverage	E value	Max ident	Identities
XM_001468128.1	2,4-dienoyl-CoA reductase [Leishmania infantum JPCM5]	3302	100%	0	93%	2055/2188
XM_001685787.1	2,4-dienoyl-CoA reductase [Leishmania major strain Friedlin]	3225	100%	0	93%	2042/2189
XM_001567821.1	2,4-dienoyl-CoA reductase [Leishmania braziliensis MHOM/BR/75/M2904]	2302	99%	0	85%	1898/2211

Hubbard and colleagues (Hubbard, P. A. *et. al.*, 2003) solved the structure and determined the reaction mechanism of the *Escherichia coli* FadH 2,4-dienoyl-CoA reductase, the only solved structure for a bacterial DECR. Since the leishmanial DECRs orthologues are homologous to bacterial DECRs, all leishmanial sequences, *Corynebacterium lipophiloflavum* and *Marinobacter aquaeolei* were aligned to the *E. coli* sequence, see FIGURE 20.

TABLE 9: NCBI-Blast result of protein sequence of *L. mexicana* 2,4 Dienoyl-CoA Reductase

Accession	Description	Total Score	E value	Max ident	Identities	Positives
XP_001468165.1	2,4-dienoyl-CoA reductase [Leishmania infantum]	1390	0	94%	688/729	710/729
XP_001685839.1	2,4-dienoyl-CoA reductase [Leishmania major strain Friedlin]	1372	0	93%	679/729	703/729
XP_001567871.1	2,4-dienoyl-CoA reductase [Leishmania braziliensis MHOM/BR/75/M2904]	1274	0	84%	617/730	671/730
XP_814813.1	2,4-dienoyl-CoA reductase [Trypanosoma cruzi strain CL Brener]	722	0	52%	383/731	480/731
YP_959342.1	2,4-dienoyl-CoA reductase [Marinobacter aquaeolei VT8]	634	1.00E-179	45%	335/734	456/734
ZP_03699253.1	NADH:flavin oxidoreductase/NADH oxidase [Lutiella nitroferrum 2002]	626	3.00E-177	47%	342/725	454/725

The alignment revealed that all residues involved in the coordination of flavin adenine dinucleotide (FAD), iron (to form iron sulphur cluster) as well as flavin mononucleotide (FMN), were conserved within all aligned sequences. Residues involved in substrate binding as well as the catalytic residues, tyrosine-167 and histidine-253 (*E. coli* numbering marked with a black chevron in FIGURE 20) were also conserved. Hubbard *et al.* unfortunately were only able to position NADP(H) approximately, due to insufficient electron density. Most, if not, all residues necessary for NADPH binding are closer to the C-terminus. A search for a domain, with 331 C-terminal amino acids of *L. mexicana's* and *L. major* DECR, in the Pfam-database (http://pfam.sanger.ac.uk/) revealed a similarity to the

'Pyr_redox_2 (PF07992)' family. This family includes both, class I and class II oxidoreductases and also NADH oxidases and peroxidases. The identified domain consisted of two domains; a smaller NADH binding domain within a larger FAD binding domain. Within the C-terminal 331 amino acids, 10 of the 11 residues involved in FAD coordination were located. Unfortunately, no residues necessary for NADP(H) binding could be assigned, due to insufficient sequence homology to proteins with existing crystal structure within the 'PF07992' family.

A 3D model of *L. mexicana* and *L. major* DECR was also calculated. This was performed with the online tool 'Swiss-model' (http://swissmodel.expasy.org/). It compares a query sequence against a database of proteins with solved structures. If the homology of the query sequence is sufficient, a structure based on the homologous template is modelled. The prediction tool, Swiss-model, assigned the FADH structure of *E. coli* as a template. This is because the *E. coil* protein sequence had the highest homology, among proteins with a solved structure, to the DECR sequence of *L. mexicana* and *L. major*.

Later the models of *L. mexicana* and *L. major* DECR were re-calculated and energy minimisation was performed. *In silico* structure prediction revealed a remarkable overall similarity to FADH. All relevant regions, involved in FAD and FMN coordination, formation of an iron-sulphur cluster as well as the active site were structurally conserved, as the overlay of the *E. coli* structure with the *L. mexicana* DECR model shown in FIGURE 21 shows. The protein backbone involved in NADPH binding of the modelled leishmanial DECR and the *E. coli* structure fit also very well (FIGURE 21). Basically the backbones of both proteins fit very well in the area where cofactors bind and differ only in regions at the periphery (FIGURE 21 and FIGURE 22). Differences resulted mainly from insertions into the DECR sequence of *Leishmania* (see alignment FIGURE 20). These insertions in the *L. mexicana* DECR sequence localise to loops and helices in the model. Differences in the model compared to *E. coli* structure based on these insertions were highlighted in yellow (FIGURE 22). An accurate prediction of the secondary structure for these insertions is not possible without experimental data. However the models of *L. mexicana* and *L. major* (not shown) are highly similar to the *E. coli* structure and the inserted regions should not interfere with its predicted function.

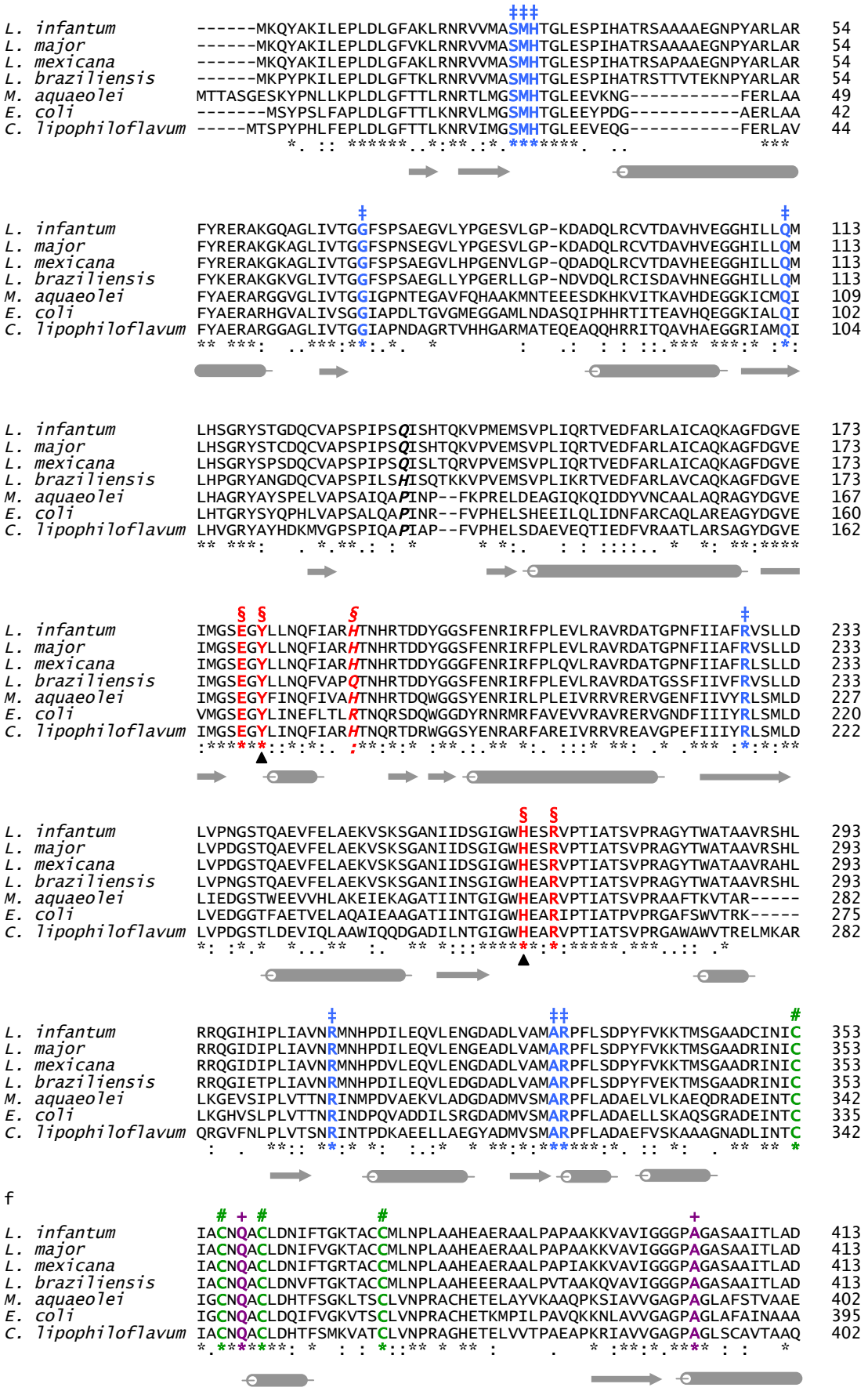


FIGURE 20 continued overleaf

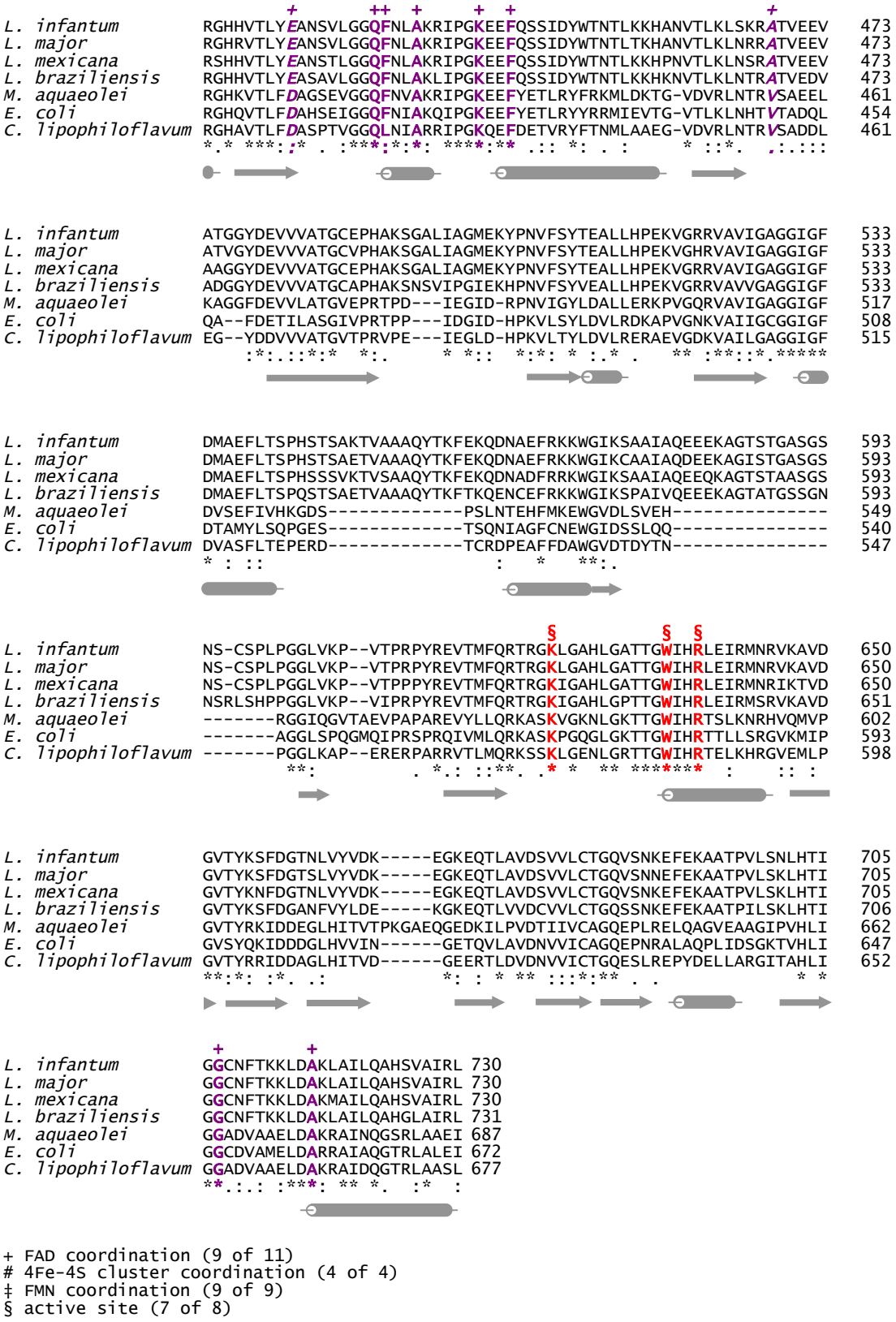


Figure 20. ClustalW protein sequence alignment of putative Leishmania DECR.

The alignment of DECR of the three sequenced *Leishmania* strains (*L. major*, *L infantum* and *L. braziliensis*) as well as *L. mexicana*, translated from nucleotide sequence, with FADH of three prokaryotes *Marinobacter aquaeolei* (YP_959342), *Corynebacterium*

lipophiloflavum (ZP_03978730) and *E. coli*. (AP_003630).
Amino acid residues involved in substrate binding as well as co-factor coordination were identified based on a crystal structure of *E. coli* DECR (Hubbard, P. A. *et. al.*, 2003). These residues are conserved as revealed by alignment. Involved residues are highlighted. Bold purple – FAD coordination; bold orange – 4 FE-4S; bold blue – FMN coordination and bold red – substrate binding. Catalytic residues are denoted with a black triangle. Residues for coordination of NADPH could not be assigned by Hubbard *et al*. Italicised residues are not identical to *E. coli*. DECR. The corresponding secondary structure elements of *E. coli* DECR is included below each alignment block, arrows denote β-sheet; tubes denote α-helix. At the end of each line for each sequence is the last residue number indicated. '*' denotes residue is identical in all sequences, ':' denotes conserved substitutions and '.' denotes semi-conserved substitutions have occurred.

The genome of *L. major* encodes also for a DECR-like protein, LmjF06.0930, which was not identified in the comparative proteome study. NCBI-Blast analysis revealed that only 5 out of the 11 residues involved in the FAD coordination were conserved and the conserved glutamine-340 (*E. coli* numbering) is deleted. Glutamine-340 is probably involved in the electron transfer from FAD to the iron sulphur cluster (Hubbard, P. A. *et. al.*, 2003). Only 7 out of 9 residues for FMN coordination were conserved and only 5 out of 9 residues in the active site. Among the non conserved residues in the active site is tyrosine-167, one of the two catalytic residues, which is substituted by a serine. Although all cysteines required to form an iron sulphur cluster were present, the model predicted that the steric orientation does not allow for the formation of an iron sulphur cluster. This suggests that the DECR-like protein is not functional whereas, DECR is very likely functional.

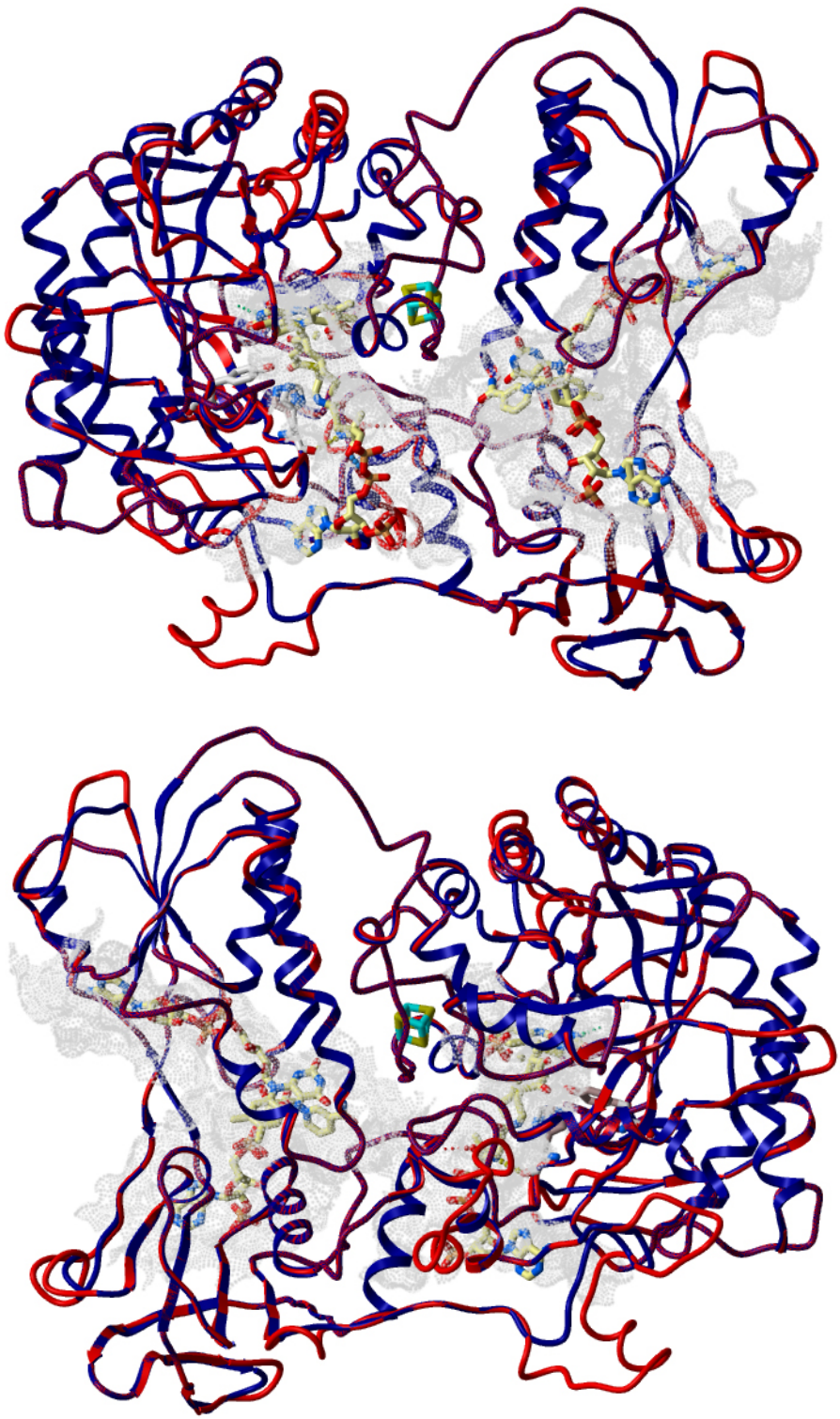


FIGURE 21. Overlay of *E. coli* FADH structure and the modelled DECR of *L. mexicana* Blue) *E. coli* FADH structure solved by Hubbard *et al.* Red) The *L. mexicana* DECR model, calculated based on the *E. coli* structure. Top and bottom showing the same overlay but turned 180°. The model also shows cofactors as well as the substrate (ball and stick models). Semi-solid mesh light grey represents the contact area of the ligands and the protein.

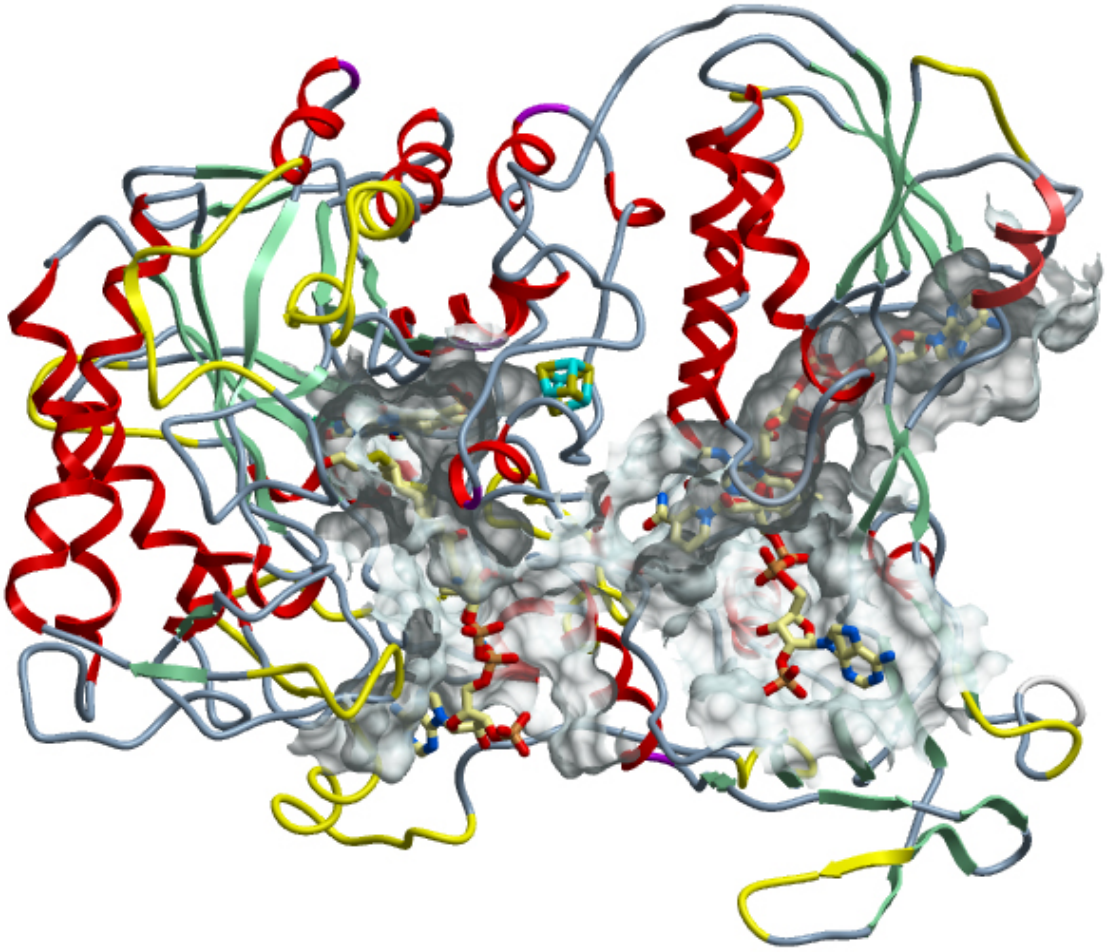


FIGURE 22. Model of *L. mexicana* DECR with differences to *E. coli* FADH structure highlighted

Differences in the *L. mexicana* compared to the DECR structure of *E. coli* are highlighted in yellow. These are mainly loop regions which result from insertions into the leishmanial DECR sequence (see alignment Figure 20). Solid mesh grey represents the contact area of the ligands and the protein.

The overlay models clearly show that backbone features for co-factor coordination are conserved in the leishmanial DECR. The catalytic residues tyrosine-180 and histidine-266 are in close vicinity to each other and to the substrate (FIGURE 23). Sterically, these two residues are in the same orientation as in the *E. coli* DECR. Therefore the nucleophilic attack of FMN on the substrate (Hubbard, P. A. *et. al.*, 2003) can occur and tyrosine-166 can provide a proton to reduce the substrate completely. Due to the close proximity of histidine-266 to

tyrosine-166 it can act to stabilise the phenolate intermediate (Hubbard, P. A. *et. al.*, 2003) by providing a hydrogen bond to tyrosine-166.

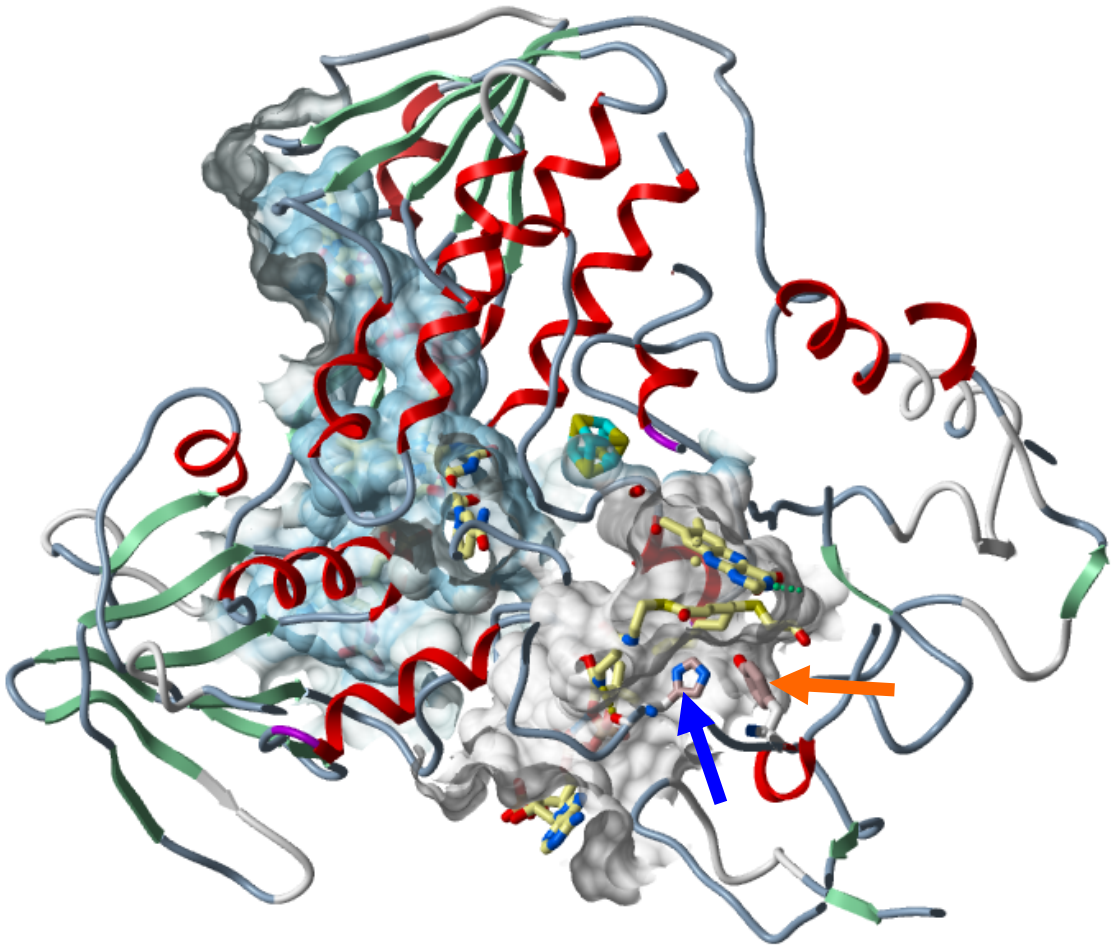


FIGURE 23. Active site of *L. mexicana* DECR model.

Catalytic residues tyrosine-180 (marked with orange arrow) and histidine-266 (marked with blue arrow) are in close vicinity to each other and to the substrate (ball and stick model). Slab mode was used to expose the catalytic residues.

3.3.2 FUNCTIONAL ANALYSIS OF 2,4-DIENOYL-COA-REDUCTASE

In order to validate the bioinformatic predictions the biological function was examined by an enzyme assay using recombinant proteins as well as by a complementation assay of DECR deficient yeast mutants (*sps19*Δ *Saccharomyces cerevisiae*; (Gurvitz, A. *et. al.*, 1997).

The open reading frame of *L. mexicana* and *L. major* was amplified from genomic DNA (gDNA) with primers complementary to the 5' and 3' end of *decr*, and cloned into pET28a (kindly provided by Balázs Szöőr). The pET28a vector is a bacterial

expression vector with an N-terminal 6 × histidine tag. The histidine tag allows purification of the tagged protein on a Ni^{2+} column.

For the amplification of plasmids the *E. coli* Xl-1 blue was generally used and recombinant protein was expressed in *E. coli* BL-21.

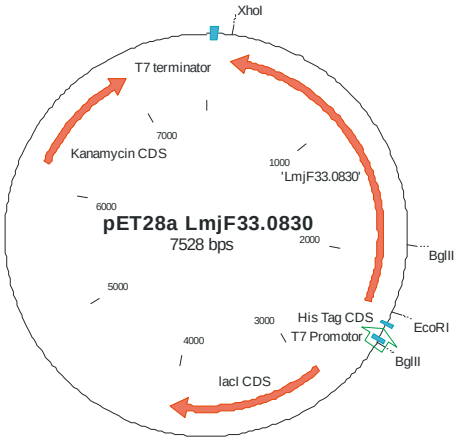


Figure 24. Expression vector pET28a carrying *decr* open reading frame.

Amplified open reading frame was cloned into the vector between the *EcoR*I and *Xho*I restriction sites. Correct targeting was confirmed by *Bgl*II restriction digest and subsequent sequencing of plasmids. The diagram shows the vector carrying the *decr* of *L. major*, the construct of *L. mexicana* is similar. Red arrows denote coding sequences, blue boxes denote regulatory elements and the His-tag, green open arrow denotes transcription start site for *decr*.

Firstly, induction conditions were determined by growing 25 ml cultures at 37 °C with different concentrations (0.1, 0.4, 0.8 mM) of Isopropyl β-D-1-thiogalactopyranoside (IPTG). Best induction was obtained at 0.4 mM IPTG for 5 h, as verified by SDS-PAGE.

Subsequently, a purification protocol was established. Bacteria were harvested, lysed and the soluble fraction was applied onto a Ni^{2+} charged column. To determine optimal purification conditions, the column was rinsed with a buffer containing different concentrations (100 mM, 200 mM, 300 mM, 500 mM and 1 mM) of Imidazole. Aliquots were taken and analysed by SDS-PAGE. The determined purification conditions for recombinant 6 × His-tagged DECR were elution with 500 mM Imidazole preceded by two wash steps with 100 mM and 200 mM Imidazole (for details see Material and Methods) was found to be optimal.

Having established the purification protocol, the biological function was investigated in an enzyme assay using the recombinant enzymes. As a positive

control FadH, the *E. coli* equivalent of DECR, was also expressed and purified (expression plasmid pQE-80L carrying a His6-tagged *fadh* was kindly provided by Horst Schulz). The enzyme assay was performed by Dr Aner Gurvitz, Vienna. For *E. coli* FadH, activity could be measured but the recombinant leishmanial DECR were not active in this assay.

In parallel, a complementation assay of *sps19* deficient *S. cerevisiae* (Gurvitz, A. *et. al.*, 1997) was performed. The *sps19* gene encodes for DECR in *S. cerevisiae* which localises to peroxisomes. Therefore, the open reading frame of leishmanial *decr* was C-terminally extended, by PCR, to encode three amino acids, serine, lysine and leucine (SKL), the consensus peroxisomal targeting signal 1. The targeting signal is sufficient to direct proteins into the peroxisome. The extended *Leishmania* open reading frames were cloned into a yeast expression vector, carrying the *cta1* (encodes peroxisomal catalase A) promoter, and terminator and were transformed into *sps19*Δ cells. Yeast *sps19* served as a positive control and as negative control *cta1* was used. Cells were plated onto a petroselinic acid containing plate and assessed for growth (FIGURE 25). The formation of transparent zones in the opaque agar around regions of cell growth would indicate the utilisation of pertoselineate as a fatty acid substrate. Yeast cells transformed with the leishmanial constructs did not show clearing as noted with *sps19* transformed mutants. Hence leishmanial DECR showed no complementation of the deletion mutant. The complementation assay was also performed by Dr Aner Gurvitz.

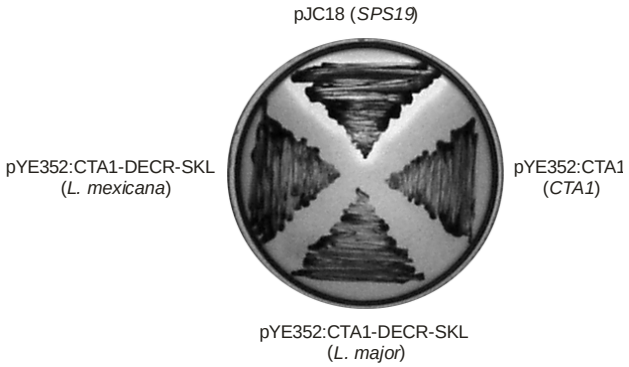


FIGURE 25. Complementation assay of leishmanial DECR in 2,4 Dienoyl-CoA reductase deficient *S. cerevisiae* cells.

Plate assay for the ability of leishmanial 2,4-Dienoyl-CoA reductase to restore the competence of BJ1991*sps19*Δ cells to degrade petroselineate as the sole carbon source. Clearing around plated cells indicates a 2,4-dienolyl-CoA-reductase activity. BJ1991*sps19*Δ cells were transformed with indicated constructs; clockwise from top: pJC18 expression plasmid carrying *sps19* promoter, *sps19* and *sps19* terminator; pYE352:CTA1 expression plasmid carrying *cta1* (peroxisomal catalyse A) promoter, *cta1p* and *cta1* terminator; pYE352:CTA1-DECR-SKL *cta1* was substituted with *decr* of *L. major* or *L. mexicana.*

3.3.3 2,4-DIENOYL-COA-REDUCTASE DEFICIENT *LEISHMANIA*

L. major and L. infantum do not encode a functional RNAi machinery unlike *Trypanosoma brucei*, in which gene knock down by RNA interference is well established. Only *L. braziliensis,* the most divergent species of the three sequenced Leishmania strains, encodes for orthologues of *T. brucei*'s RNAi machinery (Peacock, C. S. *et. al.*, 2007). To investigate if *decr* is an essential gene or contributes to virulence in *L. mexicana* and *L. major*, *decr* deficient parasites had to be generated.

Sequence data for the *L. mexicana decr* locus were not available at the time. Therefore the *decr* locus had to be characterised. It was necessary to identify if *decr* in *L. mexicana,* was also a single copy gene, as in the other three sequenced strains. Southern blot analysis was performed with restriction enzymes chosen based on the sequence of the *decr* open reading frame. During this *L. mexicana* genome sequence data became available. By blasting the *L. mexicana decr* sequence against these sequence fragments the gene sequence could be verified, as well as 612 nucleotides upstream and 1491 nucleotides downstream. Thus, with

respect to the southern blot analysis, some restriction enzyme sites cut within the *decr* locus could be predicted (FIGURE 26). Furthermore, because the primers were designed according to the *L. major* sequence, blast against the *L. mexicana* sequence fragments also revealed that two nucleotides were different in the 3'-end of the gene between *L. major* and *L. mexicana*. The changes are in the wobble base pair position of coding triplets, hence these still encode the same amino acid histidine-724 and serine-725 at the C-terminus of the protein.

The southern blot analysis showed the expected fragment sizes indicating that *decr* is a single copy gene in the *L. mexicana* isolate used here.

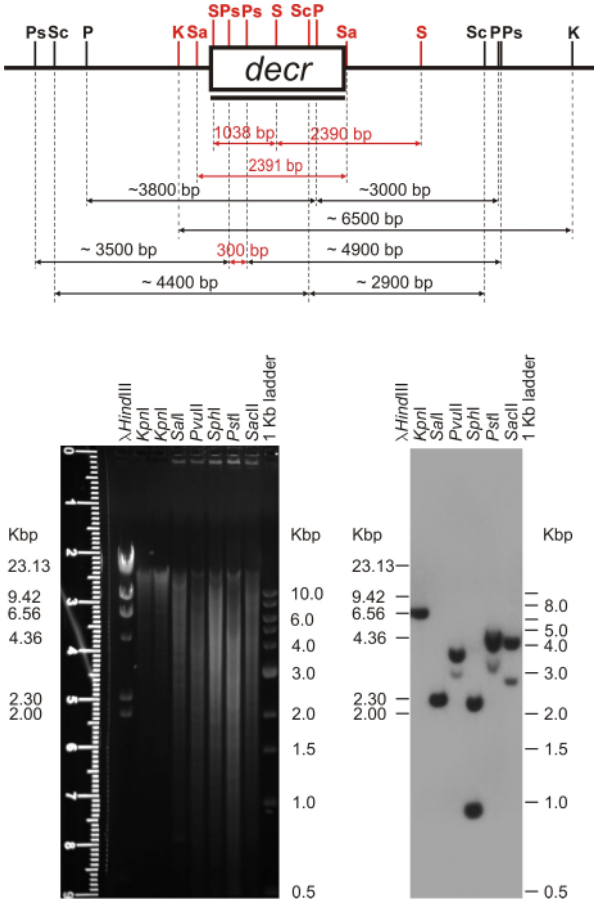


FIGURE 26. Southern blot analysis of *decr* locus of *L. mexicana*.

Top) Physical map of the *decr* locus. In red are restriction enzymes where restriction sites were known, in black sites were determined from developed Southern blot analysis. K – *Kpn*I; Sa – *Sal*I; P – *Pvu*I; S – *Sph*I; Ps – *Pst*I; Sa – *Sac*II. **Bottom left)** gDNA of *L. mexicana* was digested by the indicated restriction enzymes and separated on a 0.8% agarose gel, stained with Ethidium bromide, and imaged before transfer to nylon membrane for hybridisation. **Bottom right)** Results of hybridisation

with *decr* probe (black line under *decr* gene in top diagram). Measurements, to identify fragment sizes, were taken from x-ray film.

3.3.3.1 Generation of constructs to obtain decr*-deficient* Leishmania

Gene deletions in *Leishmania* spp. are accomplished by replacing the gene of interest with a selectabel resistance gene by homologous recombination. Homologous recombination is a rare event, in which genetic material is exchanged between similar or identical strands/ fragments of DNA. The resistance gene is therefore ligated between sequences representing the flanking regions of the gene of interest. For diploid organisms, such as *Leishmania* spp., two rounds of homologous recombination with two different resistance genes are necessary to replace both alleles of a gene of interest.

In order to generate *decr* deletion mutants, the 5'- and 3'-flanking regions of *decr* of *L. mexicana* and *L. major* were ligated into the pUC18 vector together with either the hygromycin or neomycin resistance gene. Flanking regions were amplified from gDNA and gene replacement vectors were constructed as outlined in FIGURE 27. Every cloning step was verified by restriction digest and subsequently the respective part was sequenced.

Papadopoulou and Dumas studied the dependence between frequency of recombination and length of homologous sequence. Highest recombination frequency was obtained if the homologous regions to either side of the resistance gene were ≥ 1Kb. The frequency of recombination was significantly decreased but crossover still possible when homologous sequences were < 1Kb (Papadopoulou, B. and Dumas, C. 1997). Therefore, it was initially attempted to amplify fragments of more than 1 Kb of the 3' flanking region of *L. major*, but fragments larger than ~560 bps could not be amplified for the *L. major* 3' flanking region. Therefore an artificial *Hpa*I (generates a blunt end), preceding the *Kpn*I, restriction site was introduced by PCR. *Hpa*I recognises the palindromic sequence 'GTTAAC', and cuts between the T and the A generating a blunt end. The introduced *Hpa*I restriction site is composed of the parasite sequence 'GTT' complemented with the trinucleotide 'AAC'. Later digestion with *Hpa*I generated a blunt end fragment used for homologous recombination that finished with parasite sequence.

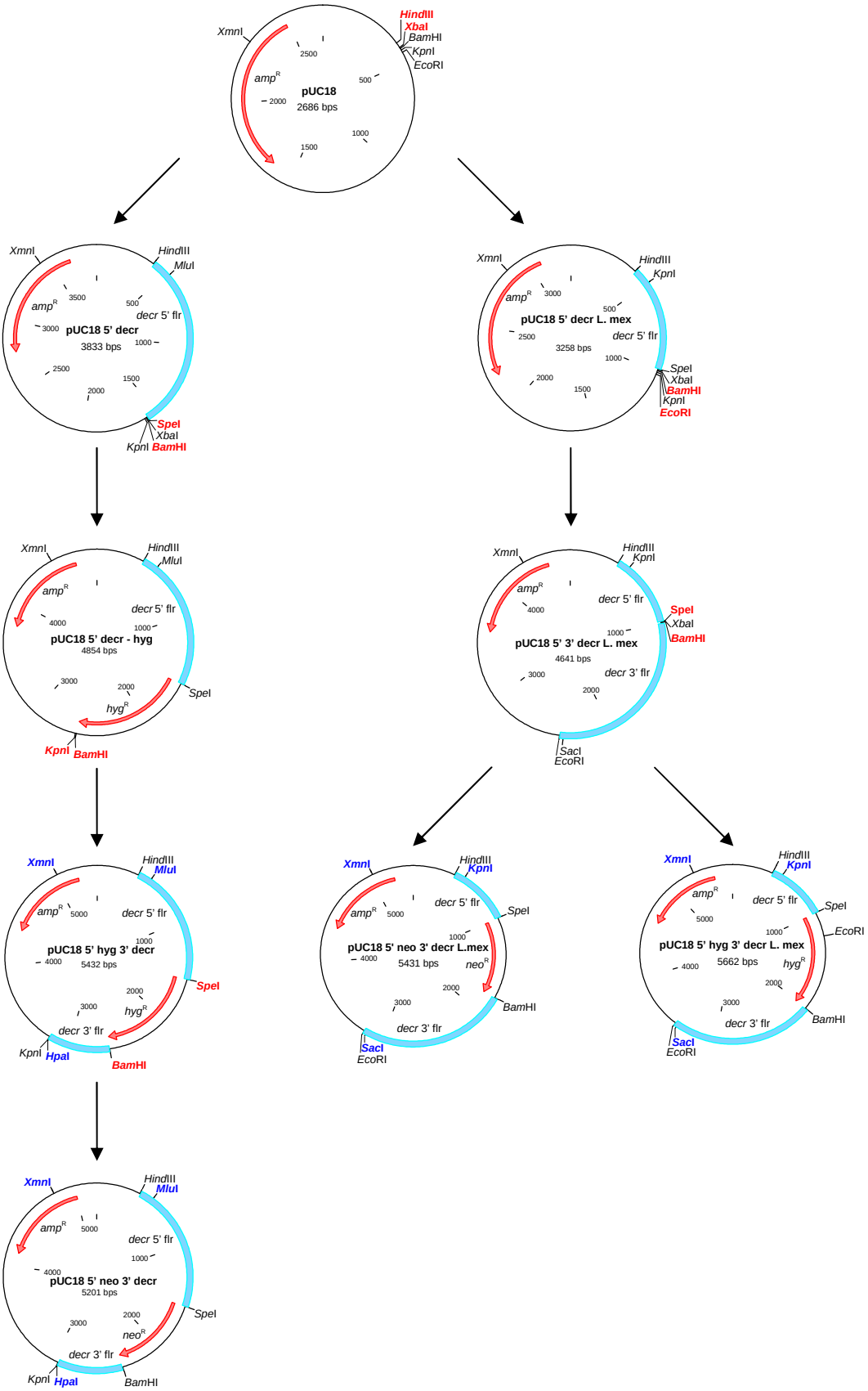


FIGURE 27. Cloning strategy of vectors for targeted gene replacement of *decr* in *L. major* and *L. mexicana*

The left and right hand side of the diagram shows cloning strategies pursued for *L. major* and *L. mexicana*, respectively. In bold red are restriction endonucleases used to digest the vector and the respective PCR product. Ligation resulted in the next vector, shown below. In bold blue are restriction endonucleases used to digest the vector to obtain DNA fragment for parasite transfections. Flanking region – flr, ampicillin resistance gene – amp^R; hygromycin resistance gene – hyg^R; neomycin resistance gene – neo^R, red arrows denote a coding sequence, blue bars denote a non-coding sequence

3.3.3.2 *Targeting of* decr

In order to obtain fragments for homologous recombination, flanking regions together with the resistance gene were excised from the generated vectors. Vectors were digested with *Mlu*I/*Hpa*I/*Xmn*I and *Kpn*I/*Sac*I/*Xmn*I for *L. major* and *L. mexicana*, respectively. Digestion with *Xmn*I allowed the separation and subsequent purification of the fragment of interest from the vector backbone (FIGURE 27).

FIGURE 28 outlines the targeting strategy of the *decr* locus of *L. major* and *L. mexicana*. Promastigotes (2×10^7) in logarithmic growth phase were transfected with 5 µg DNA of construct, containing the hygromycin resistance gene, by electroporation. After 24 h culture in SDM-medium, a selection pressure was applied by supplementing the medium with 16 µg/ml and 32 µg/ml Hygromycin B for *L. major* and *L. mexicana*, respectively. In parallel, parasites were also transfected with the construct containing the neomycin resistance gene. The medium in that case, was supplemented with 16 µg/ml G418 for both strains. After 2-3 weeks, cultures in 24 well plates were analysed for proliferating parasites. These were passaged three times until a line was considered stable. Genomic DNA was obtained and homologous recombination was verified by PCR. TABLE 10 summarises results for single allele replacements.

Ommen and colleagues targeted simultaneously two genes in *L. donovani* by transfecting both targeting constructs at the same time (Ommen, G. *et. al.*, 2009). Based on that study it was attempted to target both *decr* loci in *L. major* and *L. mexicana* simultaneously. Of the 25 drug resistant clones obtained, three were screened but correct targeting could not be verified, see TABLE 10. The remaining

clones were not screened by PCR because single allele replacements were confirmed in the mean time.

Mutants with a single *decr* allele replacement were subjected to a second transfection with the construct containing another resistance gene. The selection procedure was the same as above, only this time Hygromycin B and G418 were both supplemented to selection medium, 16 µg/ml Hygromycin B and G418 for *L. major* and 32 µg/ml Hygromycin B as well as 16 µg/ml G418 for *L. mexicana*. TABLE 11 summarises recombination and screening results of targeting of the second allele. The second homologous recombination was also validated by PCR, and a southern blot analysis (FIGURE 29).

In order to screen whether the resistance gene had targeted the right locus, primers were used which were located further up- or downstream of the flanking regions present in targeting vectors. Sequences up- or downstream of the flanking regions amplified for *L. mexicana* were not known at the time of designing the constructs to replace *decr*. The synteny at the *decr* locus is conserved between the three sequenced *Leishmania* species, *L. major*, *L. infantum* and *L. braziliensis*, (Peacock, C. S. *et. al.*, 2007) and therefore probably in all *Leishmania* species. Therefore, primers especially for *L. mexicana*, but also for *L. major*, were designed based on areas of conserved sequences in the gene immediately upstream or downstream of *decr* in *L. major*, *L. infantum* and *L. braziliensis*. These primers were used in conjunction with primers located within the *decr,* hygromycin or neomycin resistance gene.

By the time the *decr* deficient parasites were obtained, the Sanger consortium made sequence contigs of *L. mexicana* available online. Blast search with the *L. mexicana decr* sequence identified the contig containing the *decr* locus. First the primer binding sites in the immediate up- or downstream gene could be verified as well as restriction sites could be identified to perform a Southern blot analysis. Restriction sites for *Age*I and *Xmn*I were not within the amplified flanking regions of *decr* (FIGURE 28, bottom), hence were useful to verify the correct targeting by Southern blot analysis. The restriction sites were also present in *L. major* (FIGURE 28, top). A probe was generated by using the 5' and 3' amplified flanking regions as templates for *L. major* and *L. mexicana*, respectively. This allowed a

simultaneous detection of wt *decr*, as well as integration events of the hygromycin and neomycin resistance gene.

TABLE 10. Results of homologous recombination of selection marker genes to replace *decr* in *Leishmania*.

Strain	Resistance gene transfected	Positive wells	Screened	Right targeted
L. major 173	*hyg*	33 / 0 [a)]	5	3
L. major 173	*neo*	38 / 6 [a)]	4	2
L. major 173	*hyg & neo*	25 [b)]	3	0
L. mexicana	*hyg*	9 / 0 [a)]	3	3
L. mexicana	*neo*	0 / 0 [a)]		
L. mexicana	*hyg & neo*	1 [b)]	1 /0	0

[a)] number of wells containing proliferating parasites transfected with replacement construct / number of well with proliferating parasites of mock transfected

[b)] no mock transfection performed

TABLE 11. Result of homologous recombination to replace second *decr* allele.

Strain and clone	Positive wells	Screened	Right targeted
L. major 173 *decr* $^{hyg}/^{+}$ cl. 6	12	4	2
L. major 173*decr* $^{hyg}/^{+}$ cl. 12	26	12	0
L. major 173 *decr* $^{hyg}/^{+}$ cl. 19	14	7	3
L. major 173 *decr* $^{neo}/^{+}$ cl. 16	48	12	3
L. mexicana decr $^{hyg}/^{+}$ cl. 3	1	1	0
L. mexicana decr $^{hyg}/^{+}$ cl. 7	2	2	0
L. mexicana decr $^{hyg}/^{+}$ cl. 9	1	1	0

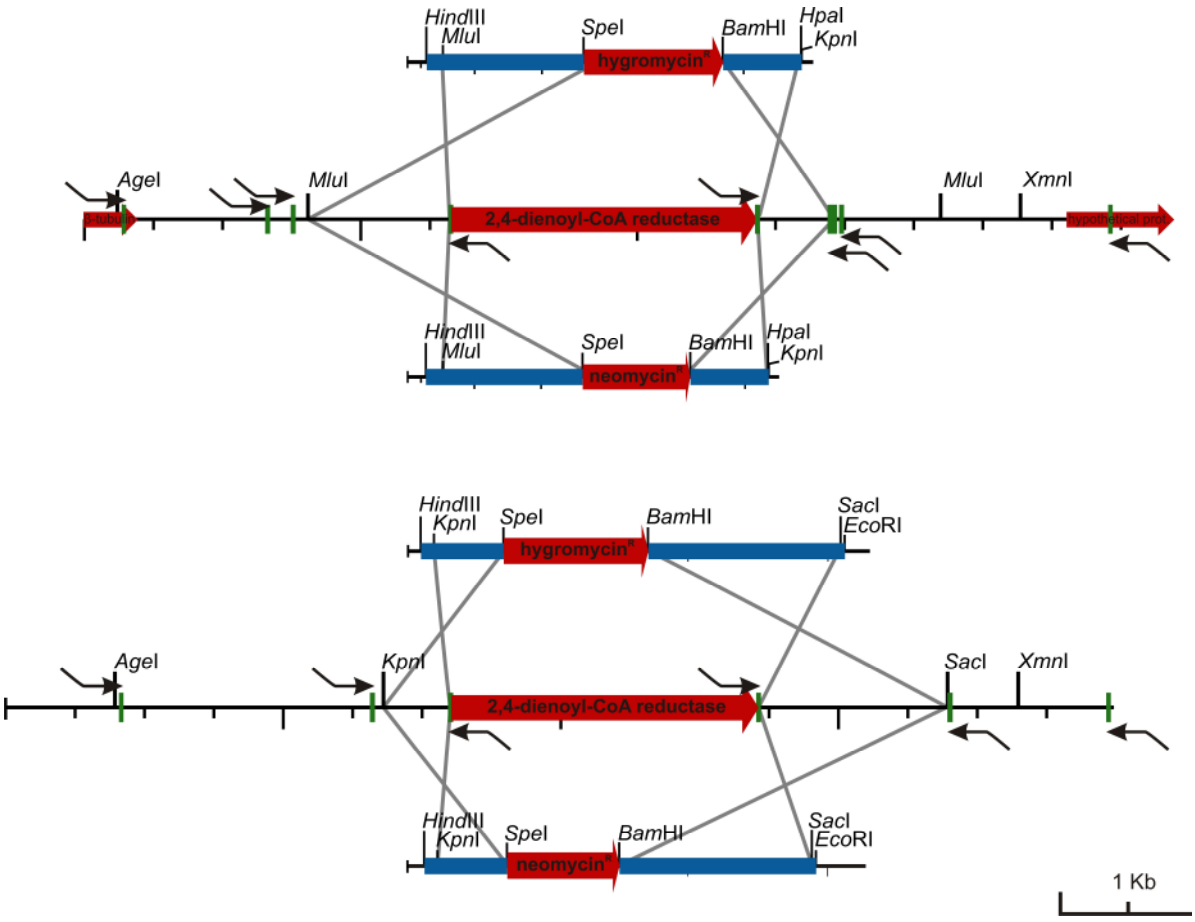


FIGURE 28. Gene replacement strategy and *decr* locus of *L. major (top)* and *L. mexicana (bottom)*.

Schematic representation of *decr* locus with homologous section of targeting constructs above and below. Homologous recombination must occur at sites of identical sequence, blue boxes, and replace in a two step process both alleles of *decr* on either chromosome, either with hygromycin or neomycin resistance gene. Red arrows represent CDS; green boxes represent primer positions on the chromosome, and black arrows above or below denote direction of amplification. Outmost primer sites, either side of *decr* were used for screening of correct targeting in conjunction with primers within *decr*, hygromycin or neomycin resistance gene.

Relevant restriction enzyme recognition sites are indicated. Grey lines indicate areas of homologous recombination.

Digestion with *Age*I and *Xmn*I of *L. major* would result in a single band of 6522 bp (respective size is indicated by a red bar, FIGURE 29) for *decr* wt locus. An additional band would appear for correct replacement of one *decr* allele either at 5367 bp for *decr* $^{+/hyg}$ (respective size is indicated by blue bar) or 5122 bp for

decr $^{+/neo}$ (respective size is indicated by green bar). For mutants with correct replacement of both alleles *decr* $^{hyg/neo}$ only the latter two bands, and not the 6522 bp band, would be detected. FIGURE 29 depicts the respective Southern blot analysis of *L. major* 173 wt, three mutants with single allele replacement (*L. major* 173 *decr* $^{hyg/+}$ cl.6, cl. 19 and *L. major* 173 *decr* $^{neo/+}$ cl. 16) and of two mutants with double allele replacement, each derived from an independent single allele replacement clone. A single major fragment at ~6500 bp (red bar) was detected in wt gDNA, whereas the two expected fragments at ~6500 bp and either 5300 bp (blue bar) for the *decr* locus correctly targeted with the hygromycin resistance gene (clone 6 and 19), or at ~5100 bp (green bar) for the single allele replacement with the neomycin resistance gene (clone 16), were detected. Also, as expected the intensity of both bands is identical, because the probe applied hybridised at equal amounts to the fragments. For the *decr*-deficient mutants the wt band was not detected anymore and only one band at ~ 5300 bp (blue bar) was detected. Separation of the hygromycin and neomycin resistance gene fragments (5367 bp and 5122 bp, respectively) was attempted but separation on agarose gel was not good enough to separate bands. The probe also hybridised with fragments of the 1 Kb size marker and with fragments in all strains in the range between 23 and 10 Kb. The latter two fragments could represent incompletely digested gDNA or unspecific hybridisation, as for the size marker. Unspecific hybridisation might occur, because for probe generation Klenow fragment was used which produced DIG-labelled fragments of various lengths. Shorter fragments are to a lesser extent specific and might have not been removed from the membrane in the wash procedure.

The Southern blot analysis in FIGURE 29 verified that three and six independent single and double allele replacement clones, respectively, were generated.

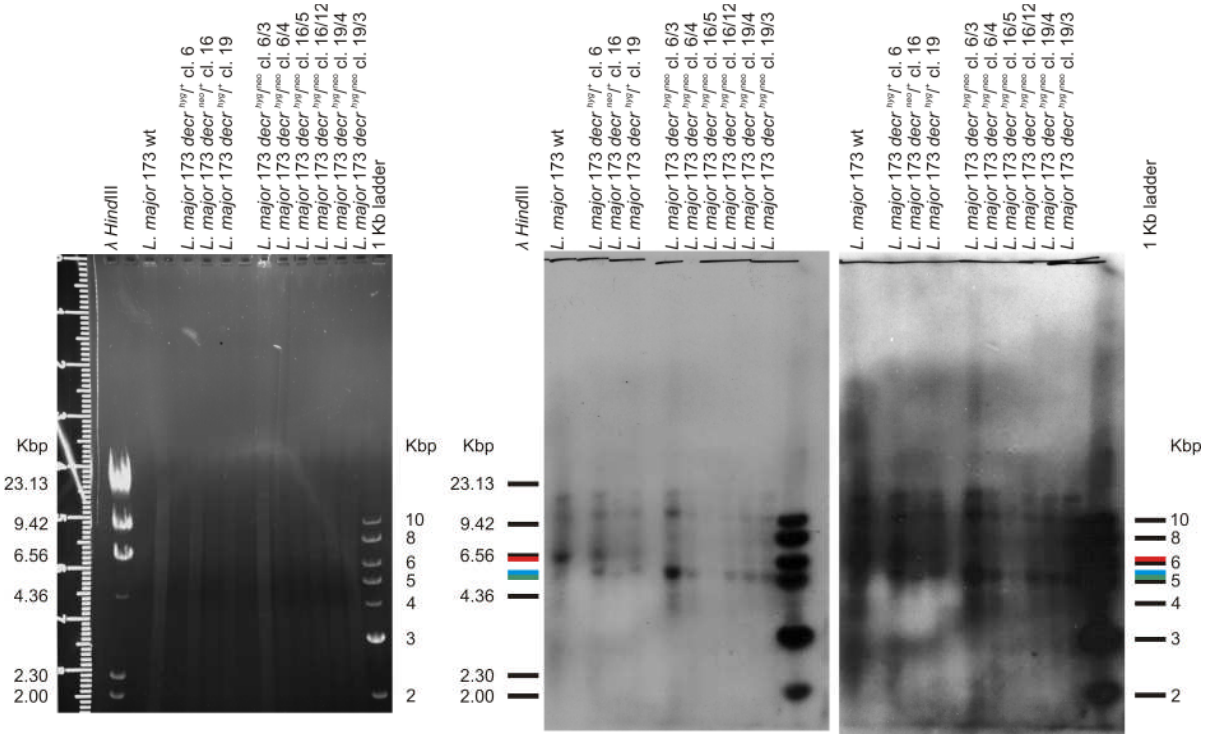


FIGURE 29. Southern blot analysis of *L. major* 173 wt parasites as well as parasites with single *decr* gene replacement and with double *decr* allele replacement.

After to verification of the correct targeting by PCR a Southern blot analysis was performed. Genomic DNA of wt, single and double allele replacement clones was digested with *Age*I and *Xmn*I. Restriction (sites are indicated in FIGURE 28) digest resulted in expected fragments of 6500 bp (red bar next to developed membrane) for the wt, and in 5300 bp (blue bar) and 5100 bp (green bar) for correct targeted hygromycin and neomycin resistant gene, respectively. **Left)** Ethidium bromide stained 0.8 % agarose gel before transfer of fragments to nylon membrane. **Middle)** short-time exposed X-ray film, **Right)** longer exposed x-ray film. The DIG labelled 5' flanking region was used as probe. Black bars denote respective Kbps of size marker.

For *L. mexicana* a Southern blot analysis after the targeting of the second chromosome was also performed. Genomic DNA of wt, single (clones 3 and 9) and double allele replacement clones (derived from clones 3 and 9, respectively) was digested with *Age*I and *Xmn*I and separated on an agarose gel and subsequently transferred onto a membrane.

Digestion of the wt gDNA would result in a fragment of 6501 bp (respective size is indicated by a red bar, Figure 30) and for targeting of the *decr* locus with the hygromycin resistance gene an additional fragment of 5346 bp (respective sizes is indicated by a blue bar) was expected. Genomic DNA of clones deficient for *decr* digested with *Age*I and *Xmn*I would result in two bands at 5346 bp and 5115 bp (respective size is indicated by a blue and green bar, respectively), if correct replacement occurred. Hybridisation with labelled probe (complementary to amplified 3' flanking region) resulted in a major band at ~6500 bp (red bar) for the wt gDNA, and a less intense band at ~5500 bp (blue bar; Figure 30). The probe also detected two unspecific fragments of smaller size at ~2800 bp and ~1800 bp. Two bands of similar intensity were detected for clones with single allele replacement (*decr* $^{+/hyg}$), at the expected size of ~6500 (red bar) and ~5500 bp (blue bar). The fragment of ~5500 bp is slightly more intense than the fragment of ~6500 bp. This indicated correct targeting of the hygromycin resistance gene, because the major single band of the wt is the combined signal of both *decr* allele fragments. Due to the correct replacement of one allele by the smaller hygromycin resistance gene, the hygromycin resistance gene targeted locus fragment is of a smaller size and separated on the agarose gel. The higher intensity of the ~5500 bp fragment is due to the combined signal of the unspecific hybridisation of the probe in that region (observed in the wt sample) and the hygromycin resistance gene targeted locus fragment. Digested DNA of mutants with a single allele replacement and subjected to a second homologous recombination did not result in the expected disappearance of the ~6500 bp (red bar) fragment, denoting the presence of the wt gene. In these clones an additional band at ~23 Kbp was apparent. This indicates that the neomycin resistance gene replacement construct has integrated at a different locus in the genome, probably multiple times according to the signal intensity.

The failure to generate fully *decr*-deficient *L. mexicana* may indicate that *decr* is an essential gene in *L. mexicana*.

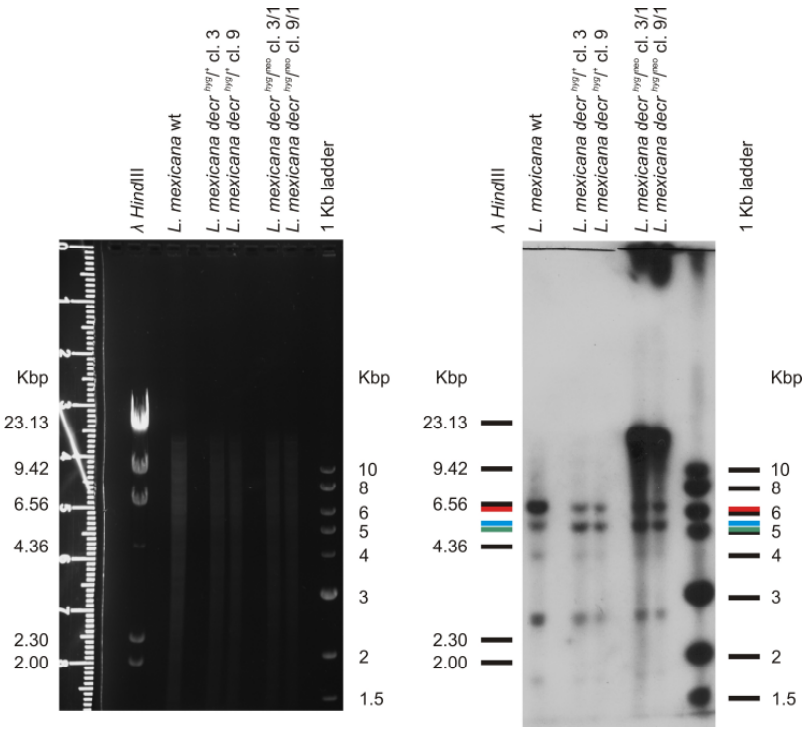


FIGURE 30. Southern blot analysis of *L. mexicana* wt parasites as well as parasites with single and double *decr* allele replacement

Genomic DNA of *L. mexicana* wt, clones targeted with hygromycin resistance gene cl. 3 and 9 as well as clones targeted with neomycin resistance gene was digested with *Age*I and *Xmn*I. The expected fragment size for the *L. mexicana* wt is 6501 bp (red bars in right panel) and for the correctly targeted hygromycin and neomycin resistance genes 5346 bp (blue bar) and 5115 bp (green bar), respectively. The probe was the DIG labelled 3' flanking region. See text for details.

3.3.3.3 Growth characteristics of Leishmania major *173* decr*-deficient mutant promastigotes*

The growth of *L. major* 173 wt and of the mutants was investigated by turbidimetry, measuring absorbance at 570 nm. Parasites were grown in SDM medium and experiments were performed in triplicate with four individual biological replicates. No significant difference between the wt and the mutants was observed (FIGURE 31; one experiment is shown as an example). Promastigotes utilise glucose for their energy metabolism, and fatty acid mainly for anabolic reactions (Hart, D. T. and Coombs, G. H. 1982). As the SDM medium contains 1.9 g/l glucose, no difference in growth was expected. So far axenic amastigotes of *L.*

major can not be grown *in vitro*. Therefore promastigotes were used to investigate whether the *decr*-deficent parasites grew less well compared to wt in a medium containing a limited amount of glucose (0.5 g/l) but 2 mmol linoleic acid compared to medium containing either 3 g/l glucose or 2 mmol oleic acid and 0.5 g/l glucose. Since *decr* degrades linoleic acid the hypothesis was, the wt will grow in all three different media compositions whereas the *decr* deficient parasites only grow in medium containing either glucose or oleic acid. McCarthy-Burke and colleagues described the usage of Roswell Park Memorial Institute (RPMI) medium for growing promastigotes (McCarthy-Burke, C. *et. al.*, 1991). Therefore commercial glucose free RPMI medium was used and supplemented with the respective components. In order to adapt the parasites to this RPMI medium, they were cultured for one passage in RPMI/10% FCS medium containing 2g/l glucose. Parasites grew in the RPMI/10% FCS medium until they reached logarithmic growth phase and were subsequently transferred into the selective RPMI medium containing in addition 10% FCS, inoculated at a cell density of 5×10^4 parasites/ml. A volume of 200 µl was placed per well of a 96 well plate. No growth was observed over a period of 10 days for neither medium.

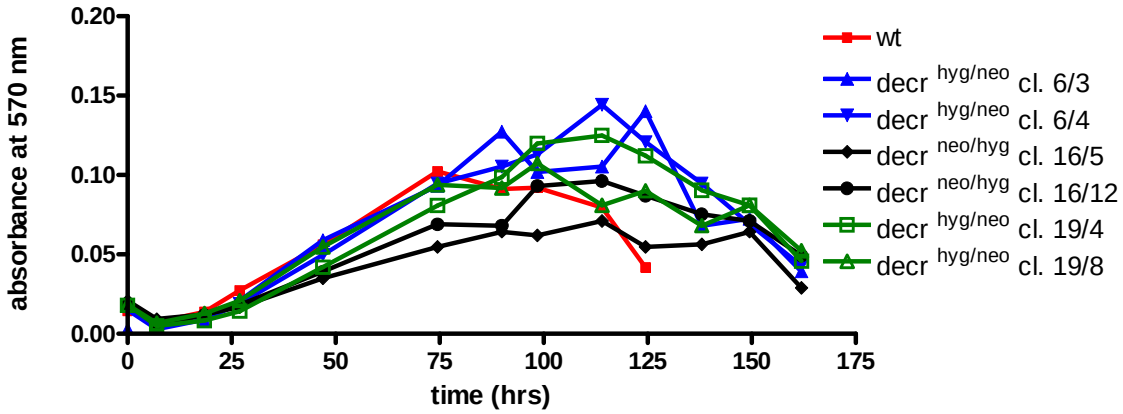


FIGURE 31. Growth curve of *L. major* 173 wt and of *decr* null mutants clones.

Parasites were inoculated into SDM medium at a density of 5×10^4 parsites/ml. One experiment is shown of four individual biological replicates. Each experiment was performed in triplicates and averaged values were normalised against averaged medium control values. *Decr*-deficient clones and wt reaching logarithmic growth phase at the same time, there were slight differences in time needed to reach maximum density and when parasites started dying (noticeable as a decline of curve). *L. major* wt parasites started dying under these conditions (96 well plate) first, in tissue culture flasks the opposite was observed, null mutants died first.

3.3.3.4 Analysis of virulence and pathogenicity of decr*-deficent* L. major

To determine potential effects of *decr*-deficiency on virulence, BALB/c mice that are susceptible to *Leishmania* spp. infection, were infected.

For the following infection experiments, equal parasite numbers of the *decr*-deficient clones, derived from the same single allele replacement clone, were pooled. The parasites were pooled in order to keep mice numbers low and in case the parasites had become attenuated, due to prolonged culture and/or non-specific loss of virulence during two rounds of transfection and subsequent selection. Five BALB/c mice per group were infected with 2×10^6 *Leishmania major* 173 wt or with each of the three *decr*-deficient clone pools, in the right hind footpad.

When parasites are injected into the foot, macrophages are attracted, parasites are phagocytosed and start to proliferate. Over time more and more macrophages migrate to the site resulting in lesion formation, the foot is swelling. By measuring the footpad size the progression of the lesion can be monitored. Therefore, the size of the infected and uninfected footpad was measured over a period of 14 weeks and the difference in thickness was plotted over time. After 7 weeks small lesions between toes of two mice infected with the wt strain appeared and after 8 weeks three out of five mice infected with wt strain showed swelling whereas the three groups infected with the different pools of *decr*-deficient parasites showed no swelling. After 9 weeks all mice challenged with the wt strain had swollen feet (Figure 32). One group of BALB/c mice, infected with the *L. major* 173 *decr* $^{hyg/neo}$ cl. 6 derived pool, developed delayed small lesions starting 11 weeks after infection. Progression of lesion development, however, was much slower compared to the wt infected group. The other two groups infected with *L. major* 173 *decr* $^{neo/hyg}$ cl. 16 and cl 19 derived pools, developed no lesions.

This indicated that deletion of *decr* affected *L. major* 173 virulence significantly.

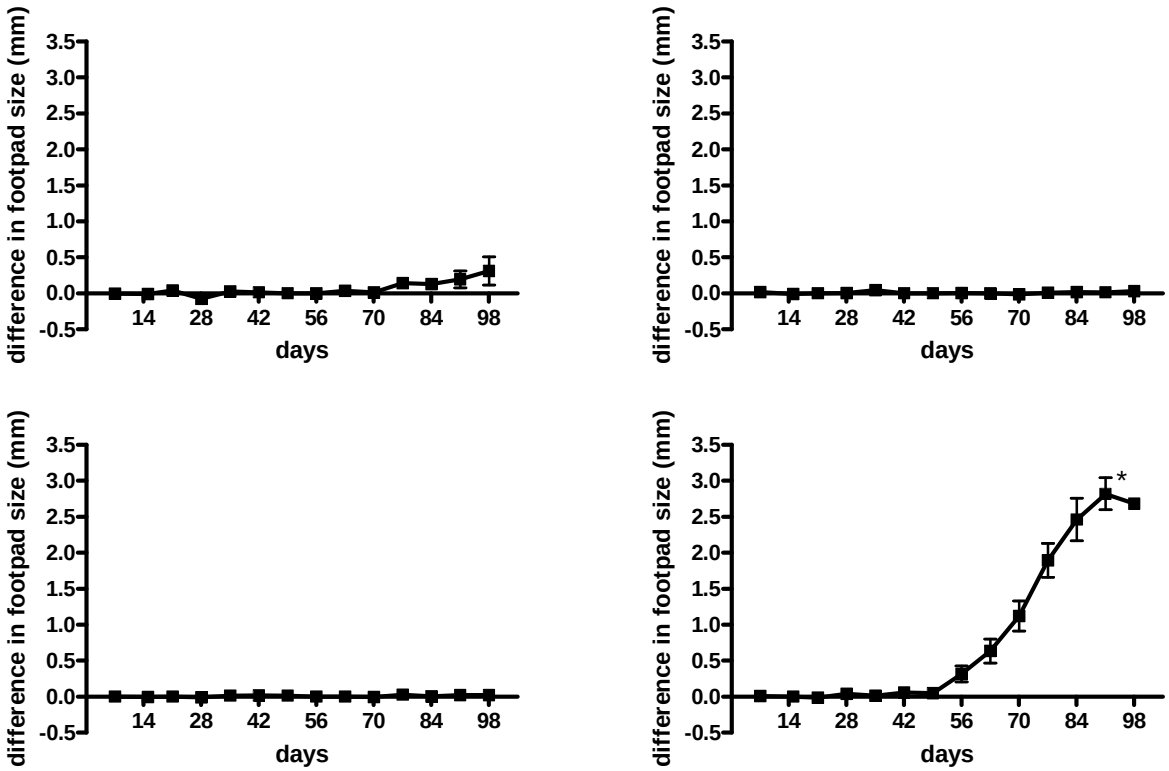


Figure 32. Infection of BALB/c mice with *L. major* 173 wt or *L. major* 173 *decr* deficient mutant promastigotes.

Mice were challenged with 2×10^6 *L. major* 173 promastigotes in the right hind footpad. Only *L. major* 173 *decr* $^{hyg/neo}$ derived from cl. 6 was able to cause lesion in BALB/c mice, but progression is much slower compared to *L. major* 173 wt. Top left) *L. major* 173 *decr* $^{hyg/neo}$ cl. 6 derived pool; Top right) *L. major* 173 *decr* $^{neo/hyg}$ cl. 16 derived pool; Bottom left) *L. major* 173 *decr* $^{hyg/neo}$ cl. 19 derived pool; Bottom right) *L. major* 173 wt. '*' three mice were sacrificed for ethical reasons. The differences between the *decr*-deficient *L. major* mutants and wt were significant ($p < 0.0001$ for all, general linear model; wt against *decr* $^{hyg/neo}$ cl. 6 derived pool $F_{1,8}=161.97$; wt against *decr* $^{neo/hyg}$ cl. 16 derived pool $F_{1,8}=1216.17$; wt against *decr* $^{hyg/neo}$ cl. 19 derived pool $F_{1,8}=1266.36$). The area under each individual curve was analysed, of 5 mice per group. The infected and uninfected feet were measured and the difference between them was plotted over time.

3.4 HOST CELL FATTY ACID METABOLISM UPON *LEISHMANIA MEXICANA* INFECTION

3.4.1 TRANSCRIPTIONAL CHANGES IN HOST GENES INVOLVED IN LIPID METABOLISM

Parasites reside in the PV, a lysosome like compartment. Cells in the periphery such as macrophages are supplied with fatty acids by LDL and as mentioned in chapter 1.2.1 this is taken up in a receptor mediated manner. The receptor is conveyed back to the plasma membrane and LDL is delivered via early endosomes, late endosomes to lysosomes. There it is degraded and FFA, FC and phospholipids are released. The lipids are exported into the cytosol, where FFA and FC are used immediately or esterified and stored in lipid droplets in the cytosol. Cholesterol can also be exported and conveyed back via the blood stream to the liver, in the form of HDL. Furthermore, macrophages use cholesterol for their own membrane to maintain membrane permeability and fluidity. Free fatty acids are stored, esterified to phospholipids or cholesterol. Studies have shown that linoleic and oleic acid derived from LDL is mainly incorporated into phospholipids and CE (Chapkin, R. S. *et. al.*, 1988; Groener, J. E. *et. al.*, 1996; Li, Q. *et. al.*, 1995). It was also shown that linoleic acid can not be converted to arachidonic acid in murine macrophages (Chapkin, R. S. *et. al.*, 1988).

It is assumed that the parasite resides in the food chain of the cell itself and sequesters at least a part of the fatty acids for its own energy metabolism. The question is whether this sequestration of fatty acids by the parasite does affect the macrophage?

Several transcription factors are activated either by the presence (PPAR-γ, LXR-α) or absence (SREBP) of particular lipids. Thus, it was reasoned that by analysing the relative change of the mRNA amount of target genes of these transcription factors, effects on lipid homeostasis within the host cell can be detected, see Figure 4.

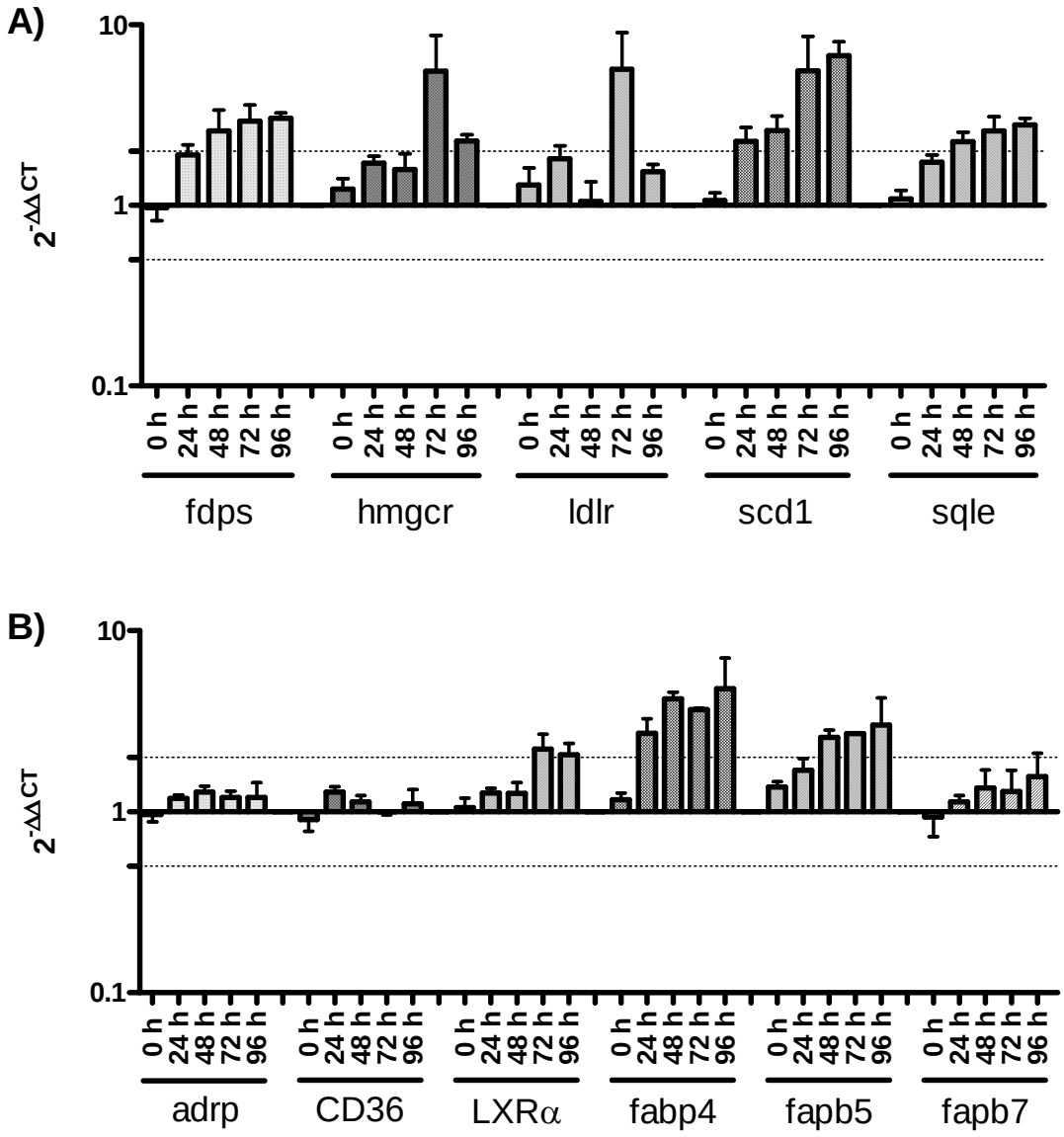
Therefore, bmdms were infected with *L. mexicana::dsred*14 amastigotes at a MOI of 5 and total RNA was obtained at different time points after infection (0, 24, 48, 72 and 96 h). Subsequently, cDNA was generated of the total RNA and qRT-PCR was performed. Upregulation or downregulation of genes of interest was determined by the comparative cycle threshold (C_T) method (Livak, K. J. and

Schmittgen, T. D. 2001). The C_T value reflects the cycle number at which the fluorescence generated within a reaction crosses a threshold. The threshold is placed in a region of exponential amplification across all samples. Furthermore, the C_T value is different for every gene dependent on its initial amount in the reaction.

In order to determine the relative change in gene expression upon infection, C_T values of genes of interest of infected cells were compared to uninfected cells at the respective time point. Prior to this, the C_T values of both the infected and uninfected cells were normalised to an endogenous housekeeping gene, hypoxanthine guanine phosphoribosyl transferase 1 (*hprt*). *Hprt* was chosen, because a study investigated the variation of 13 different housekeeping genes in five different types of normal and tumor tissue and revealed that *hprt* was the most accurate single normalisation gene (de Kok, J. B. *et. al.*, 2005). Not all samples could be analysed in one run qRT-PCR therefore to normalise between different runs of one experiment an inter-run calibrator was used. For this purpose, throughout all runs of one experiment the same sample (0 h uninfected) was used as template as well as the same primer mixtures for housekeeping genes (*hprt* and *γ-actin*). To keep variation between time points low, for each target gene all samples (uninfected and infected as well as all time points) were analysed in the same run.

Additionally, for hprt two primer pairs were used and the average between the C_T values of each pair was used for normalisation. This was done to compensate for different amplification efficiencies of each primer pair. The amplification efficiency of all primer pairs was determined by generating standard curves. Primers were used when amplification efficiency was between 1.96 and 2.17 otherwise new sets were designed and ordered. Furthermore, the comparative C_T method assumes an amplification efficiency of 100% which corresponds to two and is also the base for the exponential amplification ($2^{-\Delta\Delta C_T}$). Here, due to only minor variations in the amplification efficiencies, the base for the exponential amplification was left at two.

Results are representative of two or three biological replicates and each was performed in duplicate. In case two-fold change was detected after two experiments, a third experiment was performed.



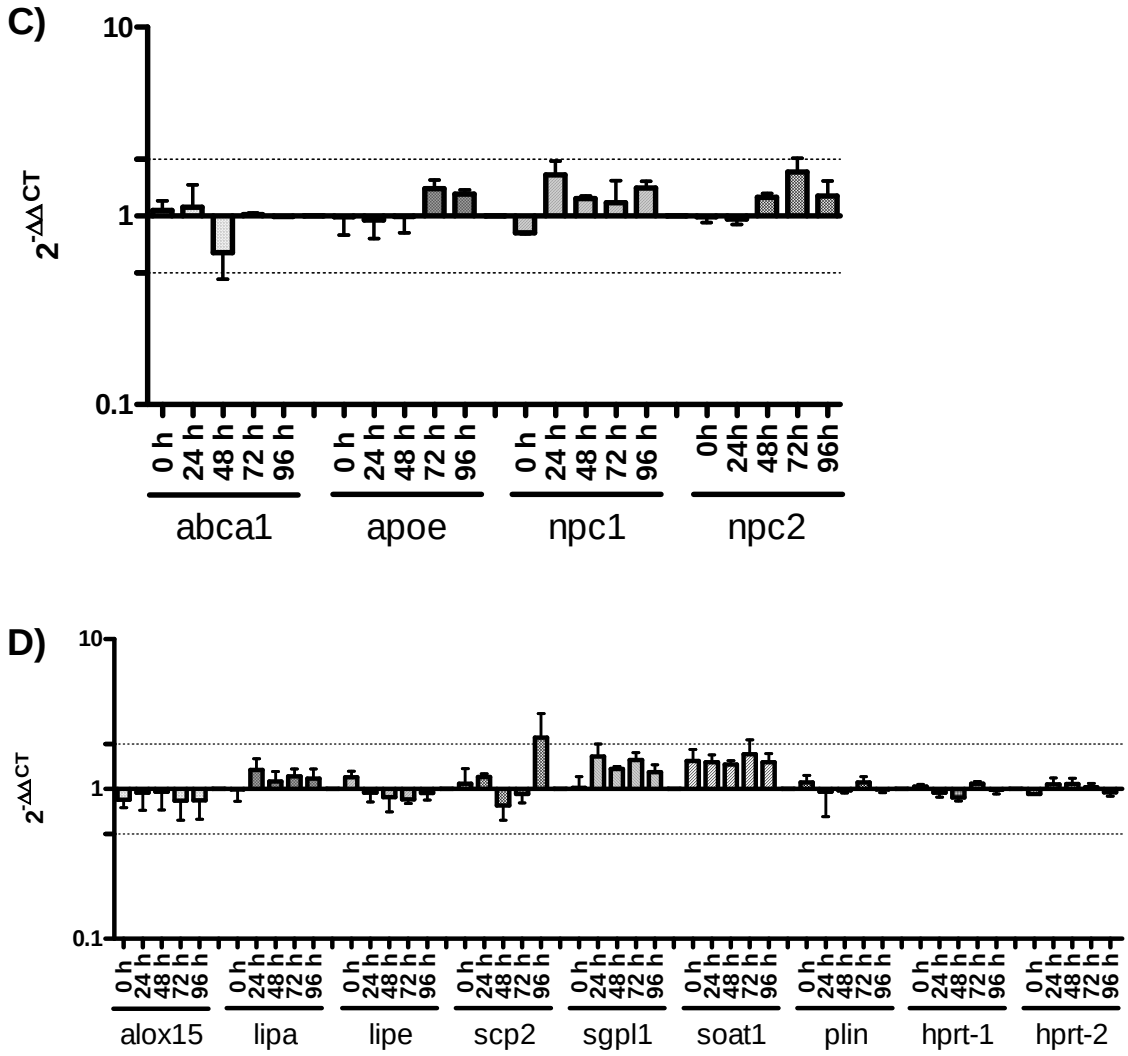


FIGURE 33. Change of mRNA levels in macrophages upon *L. mexicana* amastigotes infection.
Relative mRNA level changes between infected and uninfected bmdm of genes regulated by A) SREBP, B) PPARγ, C) LXRα and D) other transcription factors are shown. A value of one corresponds to no change in gene expression. Dotted lines show either two fold up- or downregulation.

Upon *L. mexicana* infection transcriptional changes occur within the bmdm. All genes involved in cholesterol synthesis (*fdps*, *hmgcr*, *scd1* and *sqle*), under transcriptional control of SREBP were upregulated 24 – 48 h post infection. In comparison, the LDL receptor (*ldlr*), also regulated by SREBP, and involved in LDL uptake, was not consistently upregulated. There is a trend for an upregulation of ldlr at 24 and 96 h post infection and only at the 72 h time point was it more than two-fold upregulated.

Several genes under the transcriptional control of PPARγ were more than two-fold upregulated such as the transcription factor LXRα as well as some genes involved in transport of fatty acids. Namely, fabp4 and fabp5 were upregulated at 24 h and 48 h, respectively, post infection. LXRα was upregulated after 72 h of infection. All mentioned genes, controlled by either PPARγ as well as SREBP except *ldlr* remained upregulated throughout the remaining time course. The transcription of other PPARγ regulated genes was not altered, i.e. *CD36* (uptake), *adrp* (storage) *fabp7* (transport).

Also several genes under the control of LXRα were analysed. LXRα is activated by oxysterols, but none of the analysed genes *abca1* (export), *apoe* (acceptor and transport), *npc1* and *npc2* (cholesterol transporter) were significantly upregulated.

Further, mRNA levels of genes not under the control of any of the three transcription factors mentioned above but either involved with processing of LDL and CE (*lipa*, *lipe* and *soat1*) or involved in other functions (*alox15*, *plin1*, *scp2* and *sgpl1*) were analysed. In this set of genes no two-fold up- or downregulation was detected but there was a trend in an upregulation for sgpl1 and soat1 observed. The former is involved in sphingolipid metabolism, whereas soat1 is involved in lipid storage. It esterifies FC and FFA to form CE which is later stored in lipid droplets. At 96 h post infection, Scp2 was more than two-fold upregulated, but in a third experiment this could not be confirmed, hence the large standard deviation. Scp2 is a soluble protein that binds and transports cholesterol as well as phospholipids. More experiments have do be performed in order to obtain unambiguous results.

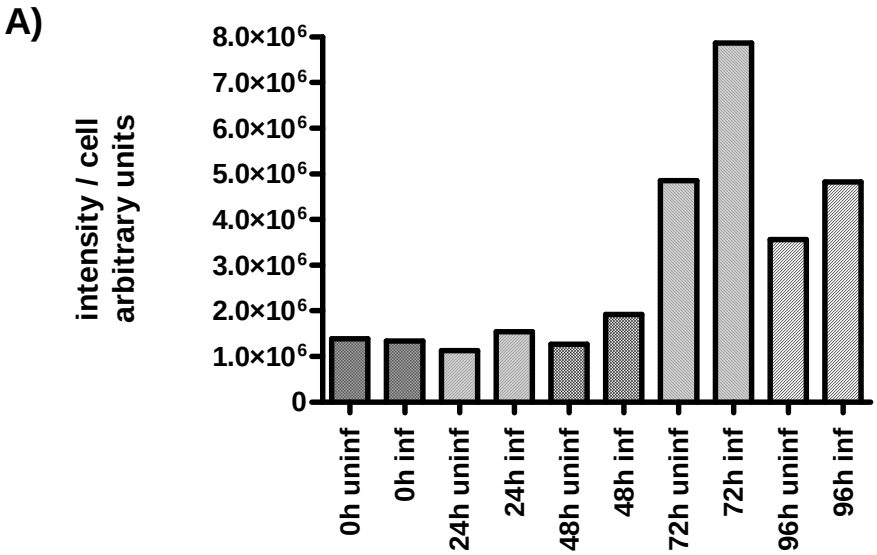
These results indicate that lipid homeostasis in macrophages is affected by infection. The activation of SREBP and failure in activation of LXRα indicates a depletion of FC in the cell, hence activation of SREBP. The mainly unaltered expression of LXRα genes indicated that the oxysterol pool is not affected and remained at basal level. Partial activation of PPARγ genes indicates a surplus of PPARγ ligands.

3.4.2 FATTY ACID BINDING PROTEINS

Furthermore, it was investigated whether upregulated mRNA levels of the fatty acid binding proteins 4 and 5 led to higher expression of the respective protein

and also whether localisation within the cell upon infection changes. Therefore, bmdm infected with the same MOI as for the qRT-PCR investigation, and uninfected cells were analysed by confocal microscopy after antibody staining of FABP4, FABP5 and FABP7. Images were recorded as stacks of serial z-sections with a magnification of 63x. Volocity (ver. 5.2.0, Improvision, Coventry, UK) was used for analysing images. Brightness and contrast were adjusted and z-stacks were projected into a single 2D-layer, the extended focus.

In order to analyse the amount of protein present, the overall intensity, of the respective channel, of the image was calculated and expressed as intensity per cell. The cell number per image was determined by counting the nuclei. The fluorescence data did not indicate significant changes of protein expression in the infected macrophages. Although as a trend fatty acid binding proteins were more abundant in amastigote-infected MΦ, this was obvious at almost all time points, (FIGURE 34), for results of FABP4, 5 and 7.



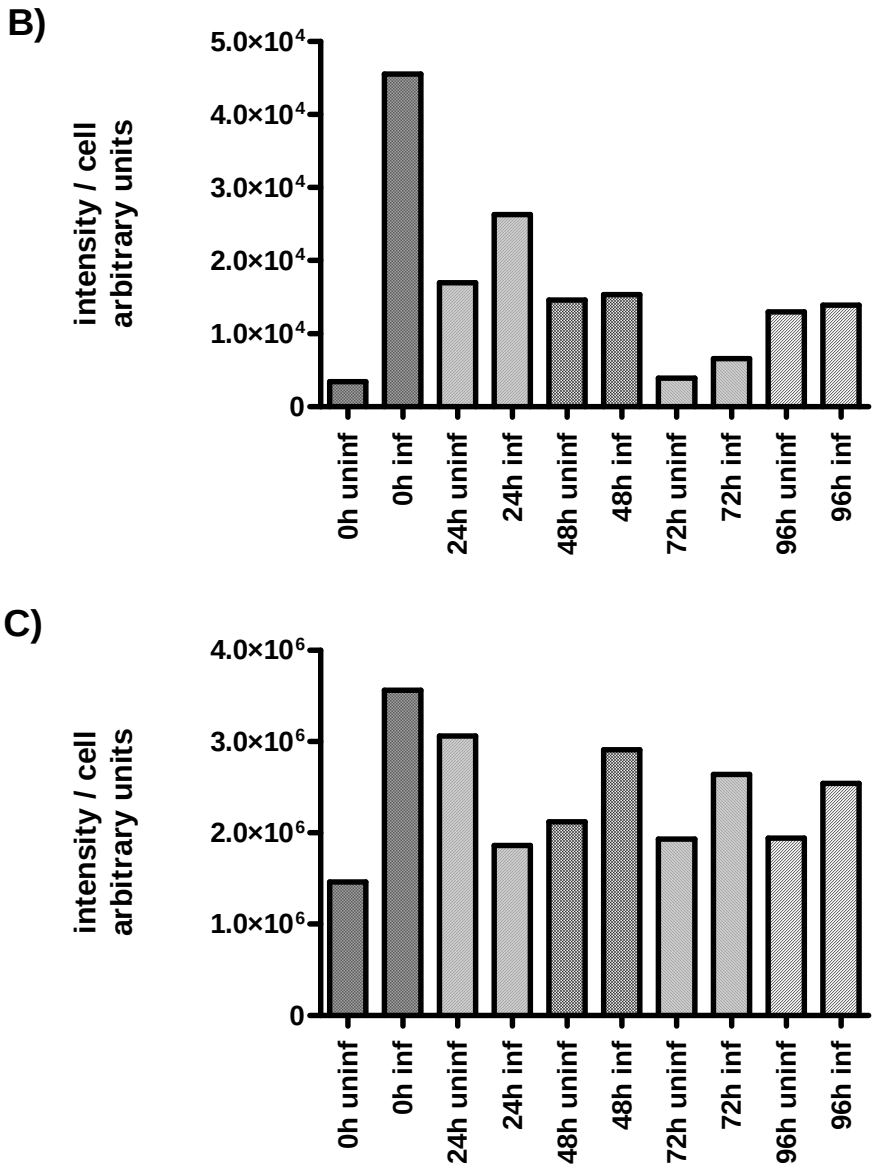


Figure 34. Protein amount of fatty acid binding protein 4, 5 and 7 in with *L. mexicana* amastigotes infected and uninfected bmdm.

Shown is the intensity per cell for the respective sample of **A)** FABP4, **B)** FABP5 and **C)** FABP7. Total intensity was measured and divided by the number of cells (nuclei) present in section. In general > 100 cells were analysed per time point and condition.

It was reported that FABP4 affects transcriptional activity of PPARγ by shuttling its ligands into the nucleus (Adida, A. and Spener, F. 2006). It appears that all FABPs can translocate into the nucleus under certain conditions (Furuhashi, M. and Hotamisligil, G. S. 2008).

Therefore, the redistribution of fatty acid binding proteins was investigated. Translocation of fatty acid binding proteins into the nucleus could prove that fatty

acids were imported and would have lead to the partial upregulation of PPARγ genes. Images were manually analysed but no indication of translocation into the nucleus for all FABP4, FABP5 and FABP7 was observed.

3.4.3 FATE OF CELLULAR CHOLESTEROL

The upregulation of genes involved in cholesterol synthesis and under the control of SREBP led to the question, what happens to cellular cholesterol? Therefore the fate of free cholesterol was investigated by a microscopic and a biochemical approach.

Free cholesterol was stained with filipin, a polyene macrolide antibiotic which binds specifically to cholesterol and can be excited by UV-light in the range of 340 – 380 nm. A series of images of infected and uninfected bmdm macrophages after 0, 24, 48, 72 and 96 h post infection was recorded. The images revealed that FC was around the amastigote within the PV as early as the amastigotes infected the macrophages and it remained throughout the time course (FIGURE 35). Filipin staining of axenic amastigotes alone did not result in the characteristic filipin halo around the parasite.

This finding agreed with the result of experiments where LDL containing ^{3}H-labelled cholesterol in form of cholesteryl linoleate was taken up by macrophages and its localisation was investigated. Bone marrow derived macrophages were infected with *L. mexicana* amastigotes (MOI of 5) or beads (MOI of 5) and 96 h post infection, infected as well as uninfected cells incubated for 1 h in medium containing 10% LDL-depleted plasma and ^{3}H-labelled cholesteryl linoleate (see also flow chart FIGURE 36). Aliquots were taken before (S1; FIGURE 36) and after incubation (S2). Labelled cholesterol containing medium was removed and cells were washed with normal medium and incubated for 2.5 h at 34 °C. An aliquot (S3) was withdrawn after the 2.5 h incubation and cells were washed in cold DMEM medium. Next, cells were lysed in DMEM/0.008% SDS, transferred into a 1.5 ml tube and centrifuged at 1100 ×g. Supernatant was withdrawn (S4) and kept and the pellet (P1, FIGURE 36) was washed three times with cold PBS. Subsequently aliquots of all samples were spotted in duplicate, except pellet fraction onto a filter mat. The mat was dried and a wax scintillant

was melted onto the filter mat and subsequently evaluated in a scintillation counter (1450 MicroBeta TriLux, Wallac).

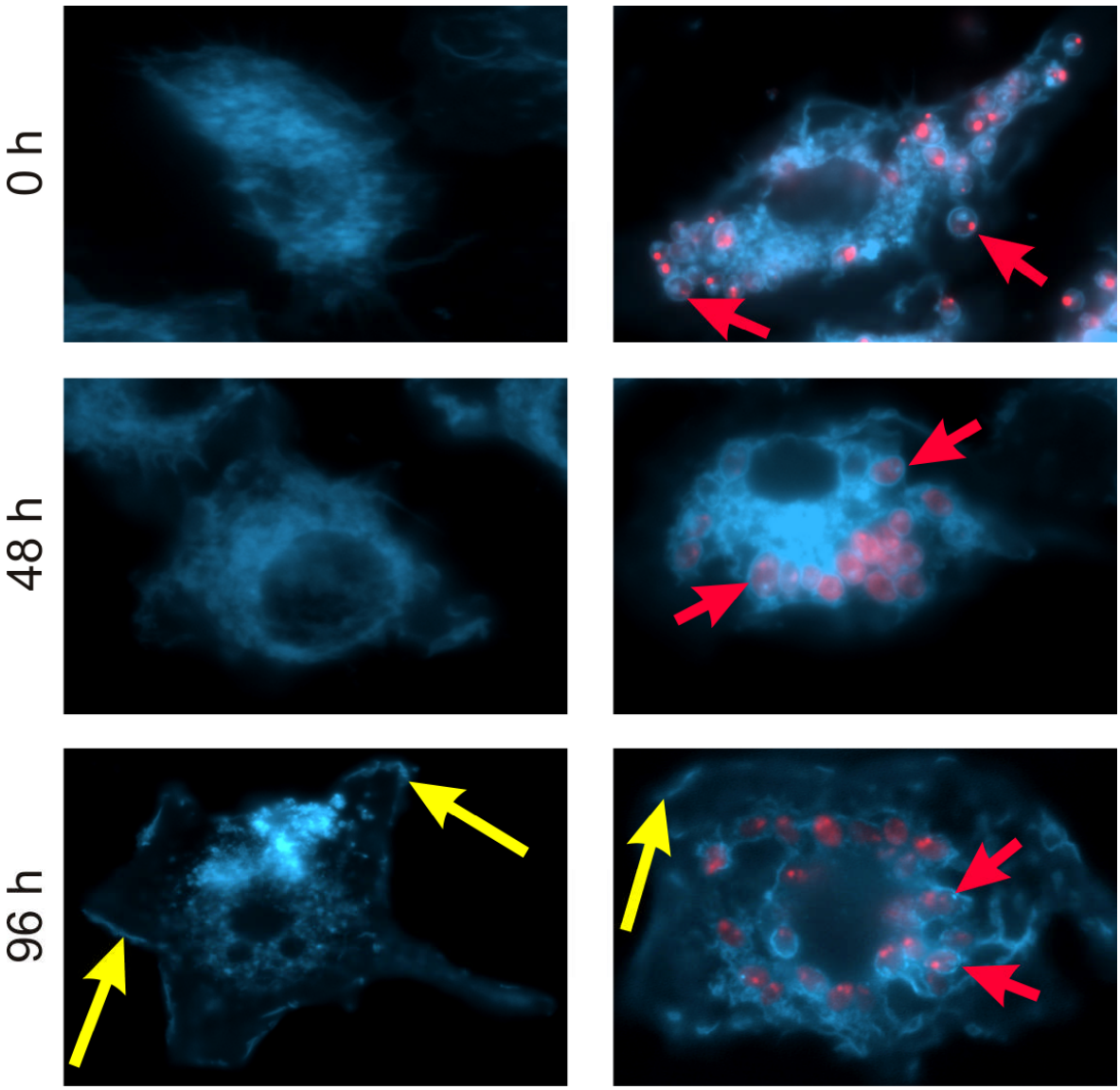


FIGURE 35. Distribution of free cholesterol in uninfected and *L. mexicana*::*DsRed* infected MΦ.

Left hand side, uninfected MΦ; right hand side, *L. mexicana*::*DsRed* amastigotes (MOI 5) infected MΦ. Distribution of free cholesterol was analysed over a time course of 96 h, with time points every 24 h. Only 0 h, 48 h and 96 h time points are shown. Cells were fixed and subsequently stained with filipin. Throughout the time course amastigotes were surrounded by free cholesterol (red arrow). After 96 h relatively less free cholesterol was present at the plasma membrane (yellow arrows). Staining of axenic amastigotes did not result in the characteristic halo around the parasites.

Images were taken at 100× magnification, numerical aperture 1.4, oil immersion.

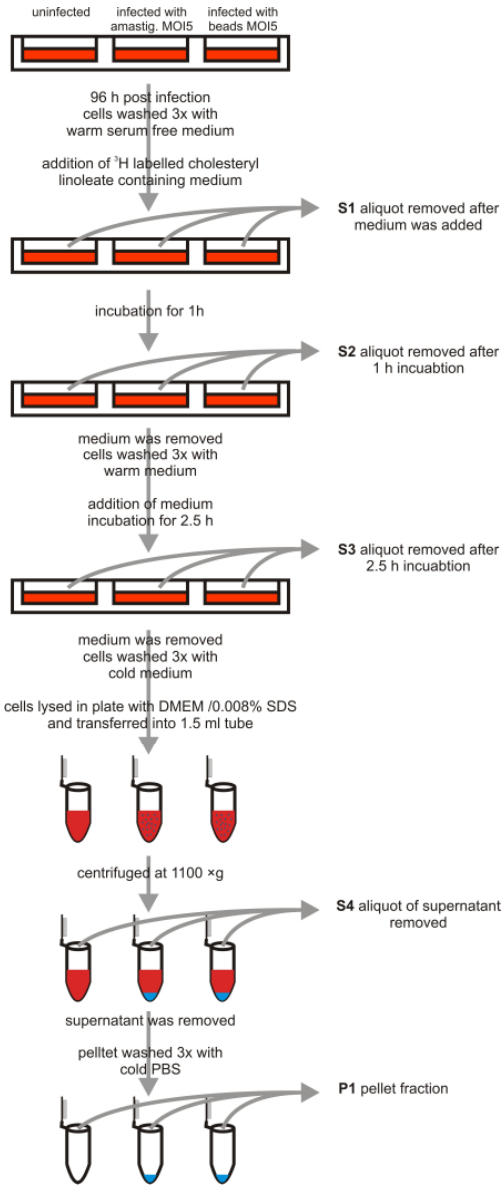


FIGURE 36. Flow chart indicating origin of aliquots in anlysis of the fate of cholesterol in *L. mexicana* amastigote infected bmdm compared to beads and non infected bmdms.

The fate of cholesterol was investigated in bmdm either infected with *L. mexicana* amastigotes or beads at an MOI of 5 and in uninfected control bmdms. Cells were incubated with cholesteryl-linoleate in which the cholesterol was ^{3}H-labelled. Aliquots of supernatants were withdrawn at indicated steps in order to determine uptake (difference in counts per minute of S1 and S2), export from cell (S3), export into the cytosol or translocation to the plasma membrane (S4) as well as retention of cholesterol in the parasite or bead containing compartment (P1).

There were no reliable data obtained for aliquots S1 – S3. Therefore the exact amount of labelled cholesterol taken up could not be determined. Also the proportions which were exported during the three hours incubation could not be

determined. In a separate experiment the overall uptake of labelled cholesterol was investigated, this revealed no differences in uptake between uninfected, amastigotes or beads infected bmdm.

In order to determine the retained proportion of cholesterol or cholesteryl-linoleate in the PV the ratio of pellet (P1) and total uptake (sum of counts per minute (cpm) of cell lysate supernatant (S4) and pellet (P1)) was calculated. In the P1 sample of uninfected bmdm only counts at background level were measured, hence no cholesterol was retained in that fraction. Infection of bmdm with amastigotes and beads resulted in proportions of retained cholesterol within the compartment they were residing of 32% and 13%, respectively, see TABLE 12. Thus, amastigote infected bmdm retained more than twice the amount of cholesterol in the PV than bead containing cells and this was a significant difference (general linear model, $F_{1,4} = 20.13$, $p = 0.011$). TABLE 12 represents the combined results of three independent experiments.

Taking the microscopic and the metabolic analysis together revealed that cholesterol is retained in the amastigote containing PV to a higher degree than in the bead containing compartment.

TABLE 12. Uptake and retention of cholesterol in uninfected bmdm and amastigotes or beads exposed bmdm

	uninfected bmdm	bmdms infected with **amastigotes** (MOI 5)	bmdm infected with **beads** (MOI 5)
total activity cell (cpm)	1135 ± 195	1199 ±32	1570 ± 325
activity pellet (cpm)	33 ± 18	386 ± 81	209 ±55
cholesterol proportion retained in pellet	**0.03 ± 0.01**	**0.32 ± 0.07***	**0.13 ± 0.03**

*) significant difference, $p = 0.011$, in cholesterol retention in compartments harbouring amastigotes compared to beads, 96 h post infection (analysed with general linear model, $F_{1,4} = 20.13$; n=3)

4 DISCUSSION

4.1. PROTEOME ANALYSES OF *L. MEXICANA*

Leishmaniases are widespread diseases, endemic mostly in so-called third world countries. Disease is caused by unicellular, obligate intracellular parasites of the genus *Leishmania*, transmitted by the sandfly vector (Alexander, J. *et. al.*, 1999). Currently, no vaccine is available and only a handful of drugs (Croft, S. L. *et. al.*, 2006; Reithinger, R. *et. al.*, 2007). Due to emergence of resistant parasites, severe side effects and / or cost, new drugs or vaccines have to be developed. Therefore, identifying proteins expressed in amastigotes that contribute to virulence allowing the parasites to thrive in host cells was proposed here as a way to discover new drug targets or vaccine antigen candidates.

Two ways are possible to identify those proteins either by direct proteomic or via a transcriptomics approach. Kinetoplastids show special features with respect to their gene regulation. Genes are organised in long arrays of polycistronic units which are transcribed constitutively into RNA precursors (Leifso, K. *et. al.*, 2007; Martinez-Calvillo, S. *et. al.*, 2003). Subsequent processing of these transcribed polycistronic units by *trans* splicing and polyadenylation leads to individual RNA molecules (Liang, X. H. *et. al.*, 2003). In *L. major* polycistronic units are organised in 133 clusters distributed across 36 chromosomes. These clusters vary in length and alternate between DNA strands of chromosomes (see light grey and light green arrows in Figure 18; and (Ivens, A. C. *et. al.*, 2005). Transcription of clusters starts bidirectionally within the divergent strandswitch regions and terminates within convergent strandswitch regions (Martinez-Calvillo, S. *et. al.*, 2003; Martinez-Calvillo, S. *et. al.*, 2004). Thus, gene regulation must occur at the post trancriptonal level (Clayton, C. E. 2002) and involves often regions in the 3' UTR of mRNA that determine stability and translational efficiency (Charest, H. and Matlashewski, G. 1994; Wu, Y. *et. al.*, 2000). For these reasons mRNA levels do not reflect protein levels in kinetoplastids, hence the proteome of the intracellular and extracellular *L. mexicana* was analysed directly. The aim was to identify proteins differentially expressed in the intracellular lifecycle stage, hypothesising that these proteins were good vaccine candidates as well as potential drug targets.

The insufficient depletion of host cell material had so far hampered proteomic analysis of intracellular *Leishmania* spp. Axenic amastigotes were used instead to mimic the intracellular state. Axenic amastigotes are morphologically like lesion amastigotes but their transcription pattern is similar to that of promastigotes, whereas lesion derived amastigotes are less similar to promastigotes/ axenic amastigotes (Holzer, T. R. *et. al.*, 2006). Although axenic amastigotes display a number of biochemical markers of lesion amastigotes they fail to synthesise others e.g. the secreted amastigote-specific proteophosphoglycan in the case of *L. mexicana* (personal communication T. Aebischer). Here, a method was developed and established which combines classical purification. i.e. density gradient centrifugation with fluorescent activated cell sorting (FACS) in order to purify DsRed expressing *L. mexicana* from infected model host cells, bmdm, or mouse lesions (Paape, D. *et. al.*, 2008). By using DsRed-expressing parasites, the harvested sucrose gradient fraction could be stained with 4',6-diamidino-2-phenylindole (DAPI) which created a "*dump*" fluorescence channel. This allowed sorting for DsRed-positive and DAPI-negative events. Thereby nuclei, and nuclear material is depleted, the major host cell contaminants of density gradient purified parasties. Sample purity was dramatically improved but, clearly, at the expense of parasite yield. The development of this novel purification method was already discussed in the diploma thesis[9] and shall therefore not be repeated here.

Samples purified by the above mentioned purification method resulted in < 3% contamination with host proteins which is far better than what was accomplished by techniques without a dump channel strategy, *e.g.* with salmonella (Becker, D. *et. al.*, 2006). The few host proteins detected, were likely to be components of the PV membrane, marking the habitat of *Leishmania* spp. Inspection of purified parasites by electron microscopy confirmed purity of the material and revealed the presence of phagosomal membranes that were co-purified see Figure 2. Hence, depletion of proteins not associated with the PV membrane/ habitat was maybe even more efficient. This led to the hypothesis that a proteomic analysis of phagosomal components could also be possible by exploiting this purification

[9] Daniel Paape 'Ein Vergleich der Proteome von *Leishmania mexicana* Promastigoten versus Amastigoten' Diploma thesis, Freie Universität Berlin, Department Biology, Chemistry and Pharmacy in 2005.

approach. This was later done and is referred to as 'second' proteomic analysis hereafter.

In the first proteome study (Paape, D. *et. al.*, 2008) amastigotes were purified from bmdms in order to characterise their proteome and compare it with that of cultured promastigotes. The bmdm were infected with axenic amastigotes and purified 24 h after infection, arguably enough time to complete the transformation. Parasite material was solubilised in 2D-lysis buffer and soluble proteins were separated by 2-DE. Material insoluble in 2D-lysis buffer was solubilised in Laemmli-buffer and separated by SDS-PAGE. Visible spots of 2-DE gels were excised and analysed by MALDI, whereas SDS-PAGE derived material was analysed by nano LC-ESI MS/MS. Comparison to a 2-DE separated proteome of BALB/c lesion derived amastigotes revealed remarkable similarities, suggesting that bmdm purified material reflected fully transformed amastigotes Figure 7.

The second proteome study was performed based on the assumption that the PV harbouring the parasite remains intact until particle sorting. Lysis then occurred during sorting possibly due to the shear forces generated. It was further reasoned that not only PV membrane fragments but also PV luminal proteins of host cell or parasite origin remain in the very same fraction. The latter would be composed of parasite secreted proteins as well as material from the lysed parasite. For this second proteome analysis, amastigote material was again purified from bmdm but intact parasites were pelleted and proteins in the supernatant were precipitated. The proteome of this sort supernatant was analysed by gel free peptide separation techniques combined with LC-ESI MS/MS. The dataset was split into *L. mexicana*-derived and host cell-derived datasets. The *L. mexicana* derived dataset was analysed as part of this PhD.

In the first and second proteome analysis products of 509 (Figure 8) and 1764 (Figure 13) distinguishable, parasite ORFs were identified, respectively. These datasets confirmed the presence of the majority of proteins reported previously for promastigotes or axenic amastigotes of *L. infantum*, *L. donovani*, *L. mexicana*, *L. major*, *L.* (*Viannia*) *guyanensis* or *L. panamensis* (Bente, M. *et. al.*, 2003; Dea-Ayuela, M. A. *et. al.*, 2006; Drummelsmith, J. *et. al.*, 2003; El Fakhry, Y. *et. al.*, 2002; Foucher, A. L. *et. al.*, 2006; Gongora, R. *et. al.*, 2003; McNicoll, F. *et. al.*,

2006; Nugent, P. G. *et. al.*, 2004) and together with a recent report on *L. donovani* (Rosenzweig, D. *et. al.*, 2008) extended knowledge significantly. The latter combined with the two proteome studies described here are the most comprehensive. To date, approximately 2500 proteins have been identified by mass spectrometry. These represent ~ 30% of proteome predicted ORFs of the *L. major* reference genome. Thus, coverage is now in the range of other model organisms such as *Bacillus subtilis* where also ~ 30% of the predicted proteins have been identified (Wolff, S. *et. al.*, 2007).

Only 92 of the 509 proteins which were identified by the gel based approach could not be detected by the gel free proteomic approach. However, the gel free proteomic approach complemented the previously published datasets by more than 740 proteins Figure 13. Due to the different approaches in our two proteome studies, a direct comparison was not possible with respect to abundance. In the gel based study, differential gene expression could only be assigned by the presence or absence of protein spots in the 2-DE gels which led very likely to an underestimation of differentially regulated / expressed proteins. In contrast, the gel free proteome study, MS/MS output data from MASCOT could be used for quantification in the very same dataset. Using the approach developed by Ishihama and colleagues (Ishihama, Y. *et. al.*, 2005), a label free measure, emPAI, was used to calculate the relative abundance of *L. mexicana* amastigote proteins. This is a very cost effective and simple method and could have the potential to generate a comprehensive differential proteome of the *Leishmania* life cycle stages. In contrast, label-based quantification methods compare two or more samples, each with a different label and have to be mixed upon labelling. The label is either incorporated metabolically *in vivo* in ideally all proteins as in 'stable isotope labelling with amino acids in cell culture' (SILAC), or a chemical reagent is used to label proteins/ peptides *in vitro*. Examples for the latter are 'isobaric tag for relative and absolute quantitation' (ITRAQ) or 'isotope-coded protein label' (ICPL). Samples analysed with the SILAC approach are distinguished by incorporating stable isotopes of amino acids (e.g. ^{15}N or ^{13}C labelled of arginines and lysines or essential amino acids) and mixed with samples containing the standard isotopes (Ong, S. E. *et. al.*, 2002; Mann, M. 2006). SILAC is accurate and straightforward to

use. However, due to usage of stable isotopes and a complex media composition for *Leishmania* spp. it is expensive. Another drawback of the metabolic labelling is, that with only 6 dalton mass difference between the labelled and non-labelled sample e.g. arginine, access to a sensitive mass spectrometer is mandatory and not always possible. Other labelling methods label free amines either by isobaric labelling of peptides (ITRAQ) or non-isobaric labelling of proteins (ICPL). The former acquires quantification data only after fragmentation of the peptide, hence it requires more sample and is time consuming (Schneider, L. V. and Hall, M. P. 2005).

Here, as well as in the published protein studies, membrane bound proteins are clearly under represented. This is due to a technical bias, association with lipids hampers the isolation and solubilisation of the proteins in buffers suitable for mass spectrometry, as well as efficient generation of positively charged peptide ions by electrospray ionization (Mirza, S. P. *et. al.*, 2007). Furthermore, an underrepresentation of secreted proteins was also obvious in published datasets as well as in the first proteome study performed and discussed here. Secreted proteins are in general of low abundance and therefore are missed due to the technical bias of mass spectrometry, unless they were enriched. (Silverman, J. M. *et. al.*, 2008). Enriching of proteins secreted by *L. mexicana* as well as host cell-derived luminal proteins was tried, by precipitating the sort supernatant in our second proteome study. Moreover, as in all (Bente, M. *et. al.*, 2003; Dea-Ayuela, M. A. *et. al.*, 2006; Drummelsmith, J. *et. al.*, 2003; El Fakhry, Y. *et. al.*, 2002; Foucher, A. L. *et. al.*, 2006; Gongora, R. *et. al.*, 2003; McNicoll, F. *et. al.*, 2006; Nugent, P. G. *et. al.*, 2004; Rosenzweig, D. *et. al.*, 2008) proteome studies to date predicted hypothetical proteins were also under represented in our two proteome studies.

The dynamic range of current mass spectrometric techniques is not broad and is dependant on the mass spectrometer used (Makarov, A. *et. al.*, 2006). In order to increase the likelihood of identifying low abundant proteins, pre-fractionation of the sample has to be performed (Simpson, K. L. *et. al.*, 2009).

In the first proteome study where proteomes of promastigotes and amastigotes were compared, this was achieved by 2-DE of 2D-lysis buffer soluble material

and subsequent separation of insoluble material by SDS-PAGE. In 2-DE protein spots were visualised by Coomassie staining of the gel, clearly relying on the sensitivity of the Coomassie stain. The SDS PAGE derived material was later analysed by nano LC-ESI MS/MS. Here protein identifications were obtained by fragmentation of a selected peptide (precursor mass). Due to time constraints and a constant flow rate only a few precursor masses were selected for fragmentation and subsequent identification. For identification, selected precursor masses were always those with the highest intensity (the most prominent) at the time. Hence only abundant proteins were identified, despite pre-fractionation. This knowledge was applied in order to investigate if *Leishmania* spp. have a translational bias in codon usage.

Translation efficiency correlates with the frequency of a certain codon to be used to encode a certain amino acid. This translation efficiency can be expressed in a numerical value the CAI. If a translational bias is present expression levels can be predicted based on the CAI (Sharp, P. M. and Li, W. H. 1987). Clear evidence for a biased codon usage was found, as the majority of identifications had high CAI values (Figure 10). As mentioned previously, hypothetical proteins were under represented. A reason for this could be that this class is in general of low abundance. Indeed hypothetical proteins of *Leishmania* spp. have a lower median CAI rank than the overall median CAI rank of the genome. Furthermore, for this class emPAI values were on average lower than the values of proteins with predictable or known functions. This provides now experimental evidence that they are mostly of low abundance. However, the emPAI analysis also identified notable exceptions. As among the 50 most abundant proteins were three hypothetical proteins.

Taking the mainly post-transcriptional regulation of gene expression into account (Clayton, C. E. 2002), the existent translational bias may not be surprising, but it has not been tested on a similar scale before, and it seems not to be a facet of gene regulation in parasites in general (Cutter, A. D. *et. al.*, 2006) and lastly, it is not just a consequence of GC bias (Carbone, A. *et. al.*, 2005).

The knowledge of the existing codon usage bias and its link to efficiency of translation can be applied to transgenes to regulate expression levels in these

organisms. This is especially true for *Leishmania* as this species is considered as a protein expression platform (Breitling, R. *et. al.*, 2002).

Experimental and theoretical analyses revealed a bias of basic proteins in the amastigote derived datasets (Figure 9). This was also confirmed by the second proteome study. To prove that this is not by coincidence a published proteome (Rosenzweig, D. *et. al.*, 2008), of similar scale but reflecting a mixed proteome of both life cycle stages, was analysed in the same way. The study of Rosenzweig and colleagues reflects a mixed proteome because *Leishmania infantum* promastigotes were transformed to axenic amastigotes, and proteomes at different transformation time points (0, 2.5, 5, 10, 15, 24 and 144 h) were generated and compared. As expected in this study there were almost equal amounts of basic and acidic protein species identified which were of similar bimodal distribution ratios as in the genome of *L. infantum*.

Amastigotes have to adapt to an acidic environment (Antoine, J. C. *et. al.*, 1998) where they actively maintain a neutral intracellular pH (Glaser, T. A. *et. al.*, 1988). It has also been shown that they possess a comparatively low overall negative surface charge (Pimenta, P. F. and de Souza, W. 1983). An evolution towards the expression of basic proteins, in the intracellular life cycle stage, would be consistent with their habitat in order to buffer protons and a change in total charge; although the latter may be mainly due to a reduced polyanionic phosphoglycan synthesis. There is evidence for an evolution of such kind: the proteome of *Helicobacter pylori*, a Gram-negative bacterium that colonizes the human stomach, a highly acidic environment, exhibits a dominance of basic proteins (Jungblut, P. R. *et. al.*, 2000). It also appears to be a general characteristic of organisms with an intracellular habit to shift to higher average pI. A study has shown by analysing 1784 predicted proteomes of phylogenetically diverse organisms, that in host associated organisms pI bias is shifted towards higher values (Kiraga, J. *et. al.*, 2007).

As mentioned above, in kinetoplastids gene regulation occurs at the post-transcriptional level by modulating translation efficiency or RNA stability. RNA-binding proteins are implicated to be involved in this by recognising sequence

motifs in the 3'-UTR (Clayton, C. E. 2002; El Sayed, N. M. *et. al.*, 2005). Such regulatory motifs have been identified in *T. brucei* by using an oligonucleotide counting algorithm (Mayho, M. *et. al.*, 2006). It was assumed that the proteins identified solely in amastigotes were, in part, due to an increased translation efficiency of binding of RNA-binding proteins to specific sequence motifs. The same approach was pursued here and two specific short sequences within the 3'UTR of several of the respective ORFs were identified. Furthermore, these sequence motifs seem to be conserved between the *Leishmania* spp. as an analysis of the respective *L. infantum* orthologous revealed (Table 6). Furthermore, these motifs were present in the 3'-UTR in an experimentally defined regulatory region of amastin a *L. infantum* amastigote specific protein (McNicoll, F. *et. al.*, 2005). This region containing the motif is responsible for an increase in translation because the abundance of mRNA encoding a reporter gene was not affected by the presence or absence of this region, but protein levels were. It is thought that this region in amastigotes belongs to a large family of degenerate retroposons (Bringaud, F. *et. al.*, 2007). A study analysed the mRNA levels of promastigotes and lesion derived amastigotes by microarray and found that for some genes (288 genes, 3.5 % of genome) mRNA levels in amastigotes were upregulated (Holzer, T. R. *et. al.*, 2006). These upregulated transcripts did not correlate with proteins identified as more highly expressed in amastigotes and containing a sequence motif. Therefore the motifs identified here may rather affect translation efficiency than stability.

To test the influence on protein expression in amastigotes of these 3'-UTRs, 3'-UTR sequence of transcripts of proteins in Table 6, containing and devoid of these motifs could be fused to a fluorescent reporter gene. Influence on expression in amastigotes as well as promastigotes could be compared by flow cytometry. Also possible would be fusing the 3'-UTR of a protein, known not to be upregulated in amastigotes, to a fluorescence reporter protein and inserting each of these motifs, or in combination, into the 3'-UTR and comparing the generated fluorescence signal to a fused unaltered 3'-UTR. If these motifs have an influence on translation efficiency the fluorescence reporter protein would have increased expression, and a more intense fluorescence signal would be expected.

If the influence of these motifs on gene expression would be validated, a search for these motifs genome wide could identify more potentially in amastigotes upregulated proteins.

Extension of our knowledge of differentially expressed proteins in promastigotes and amastigotes could be used in the future to identify similarities and differences within the 3'-UTRs of these transcripts and more motifs could be identified and might help to understand this intriguing way of gene regulation in these organisms. Interestingly, among the amastigote upregulated proteins were predicted RNA-binding proteins Table 5. This correlates with their involvement in increasing translation efficiency or mRNA stability of amastigote transcripts at the identification of 3'-UTR sequence motifs. It has been reported that the RNA binding protein TbDRBD3 increased stability of several transcripts in *T. brucei* (Estevez, A. M. 2008). In addition mitochondrial RNA-binding proteins were among the upregulated genes. This could indicate that this involvement is also the case in mitochondrial gene expression.

The phagosomal membrane of sucrose gradient harvested parasites was intact, but not after the FACS sorting Figure 2. This led to the second proteome study with the aim to identify proteins secreted by the parasite as well as host cell-derived luminal proteins of the amastigote habitat. It was assumed that the sort supernatant also contained membranous host proteins of PV as well as proteins of lysed parasites. As mentioned previously, intact parasites were removed from the FACS sort supernatant. Proteins in this supernatant were precipitated and analysed by gel free mass spectrometry. Thus 143 proteins (Figure 19 and suppl. TABLE 4), of which 67 have not been identified before, based on the sequence of the respective *L. major* ORF possess a signal peptide but no transmembrane domain and/or additional GPI anchor sequence. So far only a few secreted parasite proteins have been identified, e.g. secreted acid phosphatase, HASPB, chitinase, *Leishmania* EF-1alpha and *Leishmania* LmSIR2 protein (Ilg, T. *et. al.*, 1991; Stegmayer, C. *et. al.*, 2005; Shakarian, A. M. and Dwyer, D. M. 1998; Nandan, D. *et. al.*, 2002; Sereno, D. *et. al.*, 2005). Although most secreted proteins identified to date possess a signal peptide, there clearly exists signal independent secretion

mechanisms e.g. HASPB. Recently, signal independent secretion mechanisms have been extended by the proposed secretion via exosome-like microvesicles detected in *L. donovani* promastigotes supernatant (Silverman, J. M. *et. al.*, 2008). Silverman and colleagues assigned 151 of 358 identified proteins as actively secreted. They cultured promastigotes of *L. donovani* in serum free medium and the supernatant was taken after centrifugation. Care was taken to remove all parasites and to control lysis of parasites. Despite a gentle sample processing, the latter is difficult to avoid. Of the 151 proteins attributed to exosome excretion only 26 were not identified in our gel free proteome study carried out here. Interestingly, those identified by us and in the Silverman *et al.* were of high abundance. Those we did not identify were of very low abundance and reproducibility of identification between secretion experiments was reportedly low for these proteins (Silverman, J. M. *et. al.*, 2008). Thus, parasites have probably developed multiple mechanisms of protein secretion either via the classical, amino terminal signal sequence, the non classical, without a signal sequence, as well as the use of exosome-like microvesicles.

However, lysis of parasites during sampling could not be avoided and apart from secreted also intracellular leishmanial proteins were identified, see below. Proteins present in the lumen of the PV are contained in the sort supernatant as the detection of host-cell derived luminal proteins as cathepsins (Henry, J. A. *et. al.*, 1990) and hexosaminidase B (Schutt, F. *et. al.*, 2002) indicated (Barrios-Llerena *et al.* manuscript in preparation).

In both proteome studies, identified proteins were classified according to their predicted function. In the first proteome study proteins with a predicted role in purine/pyrimidine metabolism, protein degradation, as well as proteins with predicted but not classified function, were under represented in amastigotes. Proteins involved in carbohydrate, aerobic respiration/energy metabolism, protein biosynthesis, ribosomal, cytoskeleton, chaperones, histones and antioxidants were over represented in amastigotes Figure 12. The second proteome study showed proteins involved in carbohydrate, aerobic respiration / energy metabolism, protein degradation and purine / pyrimidine metabolism, chaperones, protein biosynthesis and amino acid metabolism were over represented. Cytoskeletal

proteins, and proteins with a predicted function but not classified, were under represented Figure 15. This is reflecting the different sample processing: gel based separation and subsequent gel free analysis of the urea insoluble material in the first study, and a complete gel free analysis in the second proteome study. In addition different mass spectrometers and techniques were used as well as different sample origins. For instance in the first study, the whole organisms were investigated whereas in the second proteome study intact parasites, as well as cell debris of lysed parasites were pelleted, and the precipitated FACS sort supernatant analysed. This is reflected by the over- and under representation of cytoskeletal proteins in proteome analysis of whole parasite and precipitated sort supernatant, respectively. In the latter sample, cytoskeletal proteins were most likely depleted by the removal of intact cells and debris. The over representation of carbohydrate metabolic enzymes in the second study is due to the lysis of cells. This class was also identified at an increased frequency in intact amastigotes compared to their occurrence in the genome. Hence proteins of this class are very abundant and contribute significantly to the pool of proteins in the FACS sort supernatant.

Although gene expression in *Leishmania* is regulated mostly post-transcriptionally, microarray analyses showed that steady state RNA levels vary between different transcripts over a range of about 2-3 orders of magnitude (Holzer, T. R. *et. al.*, 2006). As mentioned previously genes in *Leishmania* spp. are organised in polycistronic units (Ivens, A. C. *et. al.*, 2005). A tool was developed that displays proteome data onto the genome map.

Visualisation of identified ORFs made it immediately obvious that regions within the polycistronic units were devoid of any of the identified ORFs (Figure 18). In these devoid regions are mainly genes encoding hypothetical proteins, promastigote specific membrane proteins (surface antigen 2, chr. 15) as well as other membrane bound proteins (proteophosphoglycans, chr. 35; amastin, chr 34 (Wu, Y. *et. al.*, 2000; tuzin-like, chr. 34) or proteins involved in phosphoglycan biosynthesis. The proteins involved in phosphoglycan biosynthesis are involved in the biosynthesis of lipophilic compounds, as mentioned earlier, these reactions

take place mainly in or at membranes (McConville, M. J. *et. al.*, 2002). As discussed previously, hypothetical as well as membrane and membrane associated proteins were under represented. Therefore it is not expected to identify ORFs in these regions as they contain respective genes encoding for low abundant or mass spectrometric incompatible proteins. However, this suggests that these classes of proteins are kept in the genome as syntenic blocks and are probably regulated similarly.

The PV harbouring the parasites contains limited amounts of glucose (Burchmore, R. J. *et. al.*, 2003; Naderer, T. *et. al.*, 2006) and it is thought that amino acids which are present at high levels are the major carbon source for the parasites (Lorenz, M. C. *et. al.*, 2004). Key enzymes for the glyoxylate pathway are absent in *Leishmania* spp. (Ivens, A. C. *et. al.*, 2005) and labelling experiments in *L. major* promastigotes have further revealed that fatty acids are not glucogenic substrates (Naderer, T. *et. al.*, 2006). The absence of the glyoxylate pathway in *Leishmania* prevents fatty acids from entering the gluconeogenic pathway, as the β-oxidation pathway generated acetyl-CoA cannot be converted to two molecules of malate. The two malate molecules are necessary, one as a precursor for gluconeogenesis and the other to maintain the citric acid cycle.

Here, further evidence was found that in amastigotes gluconeogenesis is dominant over glycolysis. Enzymes carrying out irreversible reactions in gluconeogenesis are dominant over their respective counterparts in glycolysis (Figure 17). *De novo* generated fructose-6-phosphate is probably used immediately by phosphomannose isomerise, and converted into mannose-6-phosphate, a precursor for mannan and glycolipids, such as glycosylphosphatidylinositol (GPI) and lipophosphoglycan (LPG) (Garami, A. and Ilg, T. 2001). The upregulated gluconeogenesis in amastigotes is in agreement with published data (Naderer, T. *et. al.*, 2006). The disruption of this pathway in *L. major*, by deletion of fructose-1,6-bisphosphatase, led to attenuation of the parasites (Naderer, T. *et. al.*, 2006). Despite gluconeogenesis, additional import of hexoses is still essential because glucose transporter deficient *L. mexicana* were not viable inside macrophages (Burchmore, R. J. *et. al.*, 2003).

Amastigotes upregulate the fatty acid metabolism as they transform after isolation from lesion to promastigotes (Hart, D. T. and Coombs, G. H. 1982) and it has been demonstrated that lesion amastigotes scavenge lipids from their host cell (McConville, M. J. *et. al.*, 2007) in order to generate ATP (Opperdoes, F. R. and Coombs, G. H. 2007). Consistent with these observations, several enzymes of the fatty acid metabolism were among those differentially expressed proteins in amastigotes, as were enzymes linked to aerobic respiration / energy metabolism protein synthesis and chaperones / stress response. Among the catabolic fatty acid enzymes were the 2,4 dienoyl-CoA reductase and *trans*-enoyl-CoA isomerase. A switch towards a fatty acid metabolism in amastigotes is further supported as proteins of tricarboxylic acid cycle, e.g. succinyl-CoA synthetase and malate dehydrogenase, were also more abundant as well as cytochrome c oxidase as part of the respiratory chain. In addition our data agree with data published for *L. donovani axenic* amastigotes, where, for example, 2,4 dienoyl-CoA reductase and *trans*-enoyl-CoA isomerise, as well as cytochrome c oxidase, were also upregulated (Rosenzweig, D. *et. al.*, 2008).

Enzymes are particularly informative with respect to metabolic adaptation to intracellular life which, for *Leishmania*, is not yet fully understood. Recently, an interactive tool, LeishCyc, to visualise known or predicted metabolic pathways in *Leishmania*, has been developed (Doyle, M. A. *et. al.*, 2009). It is based on BioCyc (Karp, P. D. *et. al.*, 2005) and allows the analysis of gene expression, proteomic, or metabolomic data by displaying the complete metabolic map, colour coded according to provided data, e.g. protein abundance. By using LeishCyc it was confirmed that the glycolytic/gluconeogeneic, as well as the catabolic fatty acid, pathways were prominent but this analysis also highlighted the absence of pathways involved in synthesis of lipophilic compounds in the dataset FIGURE 16, for reasons discussed earlier. In the future complementing abundance with kinetic data may be used to model metabolic fluxes. Such models could contribute to identify essential genes and potential drug targets, and refine "Systems Biology" approaches in which the *Leishmania* metabolism is already analysed in silico (Chavali, A. K. *et. al.*, 2008).

In summary, the proteome studies performed here identified a differentially represented set of proteins in amastigotes and confirmed current views on metabolic adaptations to life inside the host cell. In total more than 1800 leishmanial proteins were identified, which represents ~ 22% of the predicted proteome. 143 predicted secreted proteins, of which 67 were novel, were also identified.

Bioinformatic analyses revealed evidence for translational bias in codon usage in *Leishmania*, evolved to increase translation efficiency of RNA transcripts. They also revealed sequence motifs in the 3'UTR which may be associated, with translational control of gene expression.

The proteins identified here included all *L. mexicana* homologues of vaccine antigens that have been tested to date with promising results in diverse models of experimental leishmaniases (suppl TABLE 1).

These datasets provide a unique resource for selecting novel vaccine candidates. The second proteome study also provides vital protein abundance data. This is an important parameter for choosing vaccine candidates (Aebischer, T. *et. al.*, 2000; Sabarth, N. *et. al.*, 2002). This aspect is currently being exploited within the lab group with very promising initial results.

Among the upregulated proteins in amastigotes were several involved in the catabolic fatty acid metabolism. This subset was investigated more closely, in order to identify a potential virulence factor.

4.2. 2,4 DIENOYL CoA REDUCTASE AND ITS RELEVANCE FOR VIRULENCE AND PATHOGENICITY

Proteomic analyses of promastigotes and amastigotes identified that a subset of the differentially expressed proteins was involved in catabolic fatty acid metabolism. This confirmed data from Hart and Coombs that lesion derived amastigotes utilise more triglycerides and non-esterified fatty acids compared to promastigotes (Hart, D. T. and Coombs, G. H. 1982). Of these enzymes, three were predicted to be involved in the ß-oxidation pathway, based on homology. Two other enzymes were involved in auxiliary reactions of the ß-oxidation pathway, processing of unsaturated fatty acids.

Blast analysis revealed all except one were homologous to mammalian proteins with the same metabolic function. The protein not homologous to mammalian proteins was 2,4 dienoyl CoA reductase (DECR). Up to 45% of the *L. mexicana* DECR sequence is identical to the sequence of *M. aquaeolei*, and 47% of *L. major* DECR sequence is identical to *C. lipophiloflavum*. *Leishmania* has DECR and other genes, probably acquired from bacteria through lateral gene transfer (Opperdoes, F. R. and Michels, P. A. 2007; El Sayed, N. M. *et. al.*, 2005; Ivens, A. C. *et. al.*, 2005; Peacock, C. S. *et. al.*, 2007). So far only one structure of a bacterial DECR has been resolved. Hubbard *et al.* (Hubbard, P. A. *et. al.*, 2003) identified the structure of *E. coli* DECR and assigned the residues necessary to coordinate all cofactors except NADPH and proposed a reaction mechanism. The prokaryotic DECR catalyses the reaction as a monomer, whereas the eukaryotic DECR reduces its substrate as a homo tetramer (Dommes, V. and Kunau, W. H. 1984). Furthermore the prokaryotic DECR contains the prosthetic groups flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) and an iron-sulphur cluster (4Fe-4S). Hubbard and colleagues proposed that the electrons supplied by NADPH are transferred onto FAD and in a two step mechanism via 4Fe-4S cluster onto FMN. The transfer of electrons to the 4Fe-4S cluster is either direct or possibly via glutamine-340. Glutamate-165, fully reduced FMN and the catalytic residues tyrosine-167 and histidine-253 are involved in the reaction mechanism. The glutamate-165 stabilises the carbonyl oxygen of the thioester. This allows the reduction of the substrate by FMN and the catalytic residues (Hubbard, P. A. *et. al.*, 2003), resulting in trans-Δ^2-enoyl-CoA which can directly enter the β-oxidation pathway (Dommes, V. and Kunau, W. H. 1984).

In eukaryotic DECR, the electrons are transferred directly from the NADPH on to the substrate (Alphey, M. S. *et. al.*, 2005), and no further coenzymes are involved. This results in trans-Δ^3-enoyl-CoA, which has to be converted to trans-Δ^2-enoyl-CoA by enoyl-CoA-isomerase before it can enter the β-oxidation pathway again (Dommes, V. and Kunau, W. H. 1984). In the eukaryotic DECR the carbonyl oxygen of the thioester is stabilised, during the reduction, by tyrosine-199 and asparagine-148 (Alphey, M. S. *et. al.*, 2005).

Alignment and model prediction revealed that residues necessary for the coordination of the prosthetic groups as well as the catalytic residues were conserved in the leishmanial DECR. The alignment also showed that leishmanial DECR have several insertions compared to the bacterial DECR proteins. A predicted model of the leishmanial DECR showed remarkable similarity to the *E. coli* DECR structure. On the basis of structural information of the template both target and template sequences could be aligned. The insertion resulted in loops and helices (highlighted in yellow in Figure 22). A prediction of secondary structure elements without experimental data will not be accurate or possible. Therefore, insertion sequences predicted to be a loop may be orientated in the native protein in a different way. Loops are in general very flexible and could interfere with substrate or cofactor binding, but based on the model calculated here, it is unlikely that these loops interfere with substrate binding or cofactor coordination.

Because no residues could be assigned which are necessary to coordinate NADPH, there is a possibility, although unlikely that substitutions occurred in the C-terminus. This would prevent binding and/or the electron transport from NADPH. Nevertheless, based on the alignment data as well as the predicted model, it is assumed that leishmanial DECR is a functional enzyme.

The biological function of the leishmanial DECR has to be validated. This was addressed in an enzyme assay (Gurvitz, A. *et. al.*, 1999; Kunau, W. H. and Dommes, P. 1978) and complementation of DECR deficient *S. cerevisiae (Gurvitz, A. et. al., 1999)*. In both cases the reductase activity could not be verified. For the complementation assay leishmanial DECR was extended by the tri-peptide SKL in order to direct it into the peroxisome. The peroxisome has a pH 8.2 ± 0.2 (van Roermund, C. W. *et. al.*, 2004) whereas the intracellular pH of amastigotes is ~ pH 7. This intracellular amastigote pH is also maintained even inside the PV with a pH 4 - 5.5 (Glaser, T. A. *et. al.*, 1988). It is possible that leishmanial DECR is not functional at pH 8, as enzymes are sensitive to pH changes. Complementations of yeast DECR null mutants with DECR of *E. coli* were not performed. Therefore it is unclear whether this model is not suitable for DECR of bacterial origins in general. It has however been shown that the human mitochondrial DECR containing a C-

terminal SKL tripeptide, could restore growth of DECR deficient yeast mutants (Gurvitz, A. *et. al.*, 1999). Therefore the complementation of an *E. coli* mutant devoid of DECR could validate its biological function.

The enzyme assay also did not indicate an activity of leishmanial DECR. It could be possible, although unlikely, that leishmanial DECR relies on different co-factors which were not present or can not be synthesised by *E. coli*. Another explanation could be that the leishmanial DECR in order to function has to be post-translationally modified. *E. coli* is not able to glycosylate proteins or form disulfide bonds (Lee, S. Y. 1996). The complementation in *E. coli*, would in that case also not be possible.

Due to the availability of the *L. major* devoid of DECR, lysates of mutants and the wt could be compared. DECR is upregulated in the amastigote form, therefore the likelihood of detecting DECR activity is probably higher in lysates of that life cycle stage. However, it is not possible to obtain axenic amastigotes of *L. major*.

In order to test whether leishmanial DECR is a virulence gene and contributes to pathogenicity, *L. major* and *L. mexicana* deficient for DECR had to be generated.

Replacement of both *decr* alleles was possible in *L. major* 173 (Figure 29). This resulted in significantly reduced virulence (Figure 32). No lesion development was observed for two of the three *decr* deficient clone pools and only one clone pool led to a delayed lesion formation and had a decreased progression compared to wt.

Decreased virulence was either due to repeated passaging or the transfection process. The selection process after transfection could have led to the selection of non-virulent mutants.

It is known that *L. major* loses virulence during repeated passaging *in vitro* (da Silva, R. and Sacks, D. L. 1987). For that reason the parental *L. major* 173 wt used to generate the deletion mutant was passaged over the time period taken to generate the *decr* deficient mutant until the time of infection. Virulent parasites can be re-selected by *in vivo* passage in mice (da Silva, R. and Sacks, D. L. 1987). Loss of virulence due to manipulations as well as prolonged culturing was addressed by choosing several independent clones after the first selection round

for further transfections and subsequent infection. However, the clone which caused a delayed lesion development has to be isolated and used for re-infection and compared to infection with isolated *L. major* 173 wt. If the loss in virulence was independent of deficiency of *decr*, then the isolated clone would have the same virulence as the wt and onset and progression of lesion development would be similar.

If the decreased virulence is due to the lack of *decr*, then virulence should only be restored by ectopic expression of *decr* in *L. major* mutants devoid of *decr*. This would prove that DECR is a virulence factor and involved in the pathogenicity of *Leishmania*. In addition targeted mutations of the catalytic residues (tyrosine-180 and or histidine-266; (Hubbard, P. A. *et. al.*, 2003) and subsequent ectopic expression of *decr* should not restore virulence and at the same time indicate biological function of leishmanial DECR.

L. mexicana double allele replacements of *decr* were not obtained, only one allele could be replaced. Targeting of the second allele in *L. mexicana* was not successful and Southern blot analysis revealed integration of replacement construct at a different locus FIGURE 30. This could be an indication that *decr* is essential in extracellular *L. mexicana*. In that case, prior to the second targeting, parasites have to be complemented with an extracellular copy of *decr* as it has been done for the N-Myristoyltransferase, which was shown to be an essential gene for *L. major* (Price, H. P. *et. al.*, 2003). Two other reasons could also explain why the second targeting was not successful. Firstly, the frequency of recombination was not as high in *L. mexicana* as in *L. major*, see Table 10. This means more *L. mexicana* promastigotes have to be transfected. Secondly, the southern blot analysis of *L. mexicana* revealed that the probe complementary to the 3' flanking region of *decr* also hybridised to an apparently similar sequences the genome, see FIGURE 30. This could have led to false integration. The latter is unlikely but possible, as the correct single allele replacement was verified for three clones out of nine (Table 10).

These data indicate that *decr* in *L. major* is not essential but contributes to virulence, whereas *decr* in *L. mexicana* seems to be an essential gene, this makes *decr* to a possible target for an anti-parasitic drug.

Interestingly, during a screen for proteins that are required in the maintenance of the peroxisome in *S. cerevisiae*, mutants were identified which could not grow on oleate containing medium plates even in the presence of glucose as a carbon source (Lockshon, D. *et. al.*, 2007). Lockshon and co-workers concluded that the peroxisome is involved in catabolism of fatty acids, which have been removed from membrane phospholipids. Impairment of this degradation in *S. cerevisiae* affected the membrane integrity. It could be possible that *Leishmania*, apart from using linoleate as a carbon source, may also or even exclusively need the DECR to maintain its membrane integrity inside the host.

Finding drugs that are efficacious and safe is a multi-faceted process, but what is a good drug target? The target gene should be directly causing disease or contributing to disease progression (Hajduk, P. J. *et. al.*, 2005). The ligand binding site should be known in order to determine whether the binding site can bind to small drug molecules with high affinity and specificity (Hajduk, P. J. *et. al.*, 2005). For that purpose it is important to know the structure of the target. In the case of enzymes it is beneficial if the reaction catalysed is either specific to the pathogen or catalysis mechanism is different. This is important to minimise side effects as well as increasing specificity of the drug. In addition, to screen for compounds a functional assay is vital.

The potential for an anti-parasitic drug has been shown for N-myristoyltransferase (Price, H. P. *et. al.*, 2003). Generation of deficient *L. major* parasites was not possible, and parasites had to be complemented with a plasmid coding for N-myristoyltransferase in order to replace the chromosomal alleles. N-myristoyltransferase is also essential in *T. brucei* as its downregulation by RNA interference revealed (Price, H. P. *et. al.*, 2003). It was also reported that compounds were able to selectively inhibit the N-myristoyltransferase of *T. brucei* but so far none of the tested compounds was effective in targeting *L. major*

N-myristoyltransferase (Panethymitaki, C. *et. al.*, 2006). These compounds were also able to inhibit growth of *T. brucei* in culture (Bowyer, P. W. *et. al.*, 2008). Currently there is a screen underway, which includes more compounds, to identify *Leishmania* effective compounds (personal communication with D. Smith).

Validation of the contribution of DECR towards virulence, by the above mentioned ectopic expression of DECR in the deficient clones, could lead to the development of a compound, targeted at DECR, especially at the substrate binding site and the catalytic residues. Compounds could be mimicking the substrate and could form a covalent bond to the catalytic residues, thus inactivating the enzyme.

The binding pocket of the human DECR is shaped differently to the bacterial / leishmanial form, and different residues are involved in coordinating the substrate and the CoA moiety (Alphey, M. S. *et. al.*, 2005; Hubbard, P. A. *et. al.*, 2003). Therefore another possibility is to develop a compound target to those residues in the binding pocket (FIGURE 37 highlighted light blue), which are involved in substrate coordination in order to block access of the substrate to the pocket. In all circumstances, interference with the activity of human DECR should be avoided as it would likely result in severe side effects. Human DECR is an essential enzyme and deficiency of DECR is fatal (Roe, C. R. *et. al.*, 1990).

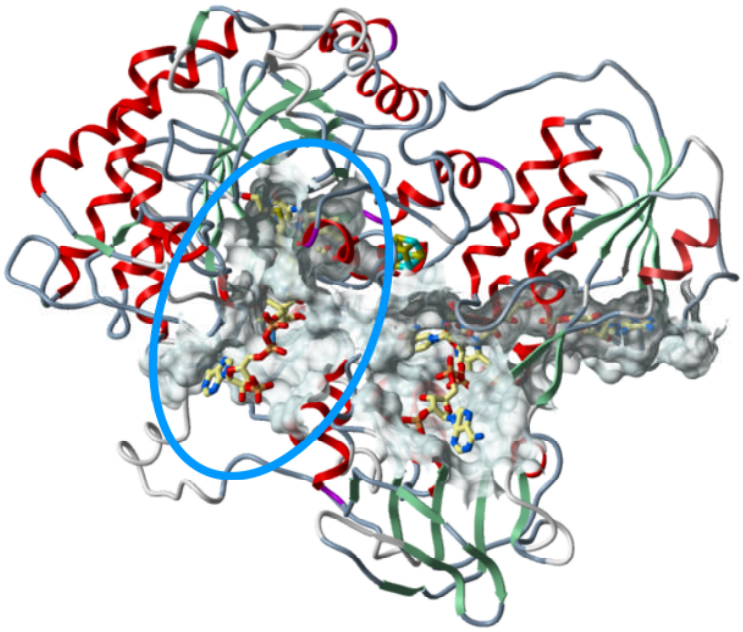


FIGURE 37. Model of leishmanial DECR highlighting substrate binding site.
Substrate binding site is encircled in light blue.

Compounds targeted at cofactor binding sites are unfavourable, because they are very likely of structural similarity to host proteins and would lead to side effects. Nevertheless if a compound is identified which specifically inhibits DECR, the next challenge will be delivering it to the PV as well as ensuring the compound's stability. The compound has to be stable and should not be modified due to the harsh conditions i.e. acidic pH in the PV.

In summary, a potential new virulence factor 2,4-dienoyl CoA reductase has been identified. So far enzyme assays to verify its biological function - reducing polyunsaturated fatty acids, such as linoleate - have failed. However alignment of the protein sequence to a functional bacterial enzyme, and a structural model, support the functional prediction. *Leishmania major* devoid of *decr* were generated and these were less virulent in susceptible BALB/c mice. It further seems to be an essential gene for *L. mexicana*. These data suggest that DECR enzyme may be an appropriate target for the development of novel anti-parasitic drugs.

4.3 CHANGES IN HOST CELL FATTY ACID METABOLISM UPON *L. MEXICANA* INFECTION

Leishmania amastigotes within the PV rely on amino acids as a carbon source and maybe fatty acids as an energy source. It was hypothesised here that *Leishmania* sequesters fatty acids from the host PV in order to thrive and that this changes the homeostasis of the macrophage. To find evidence for this hypothesis a qRT-PCR screen was developed that reveals changes in lipid homeostasis in infected cells.

Several mammalian transcription factors sense the presence (PPARγ or LXRα; (Desvergne, B. and Wahli, W. 1999; Edwards, P. A. *et. al.*, 2002; Tontonoz, P. and Spiegelman, B. M. 2008) or absence (SREBP; (Ikonen, E. 2008) of lipid compounds. Therefore, changes in mRNA abundance of target genes of these transcription factors were analysed. Transcript amount were compared between uninfected MΦ and MOI 5 infected MΦ over a period of 96 h, at 24 h time points.

Changes in mRNA levels may be related to phagocytosis. However, transcriptional changes due to phagocytosis are known to occur mostly within the first two hours post uptake. The gene expression profile of uninfected MΦ and those infected with

inert latex beads were very similar 24 h post infection (Gregory, D. J. *et. al.*, 2008). Gregory and colleagues identified a statistically significant change in gene expression for only 15 out of 12,488 probe sets upon phagocytosis of latex beads. None of the genes investigated here were among those. Thus the changes in transcript abundance over the time course observed here were most likely due to the presence of *L. mexicana* amastigotes.

The most dramatic changes detected here, were for SREBP regulated target genes (*hmgcr*, *fdps*, *scd1* and *sqle*) which senses the absence of cholesterol (Ikonen, E. 2008). Genes involved in cholesterol uptake (ldlr; Goldstein, J. L. and Brown, M. S. 1990), transport (fabp4; Furuhashi, M. and Hotamisligil, G. S. 2008) and biosynthesis (fdps, hmgcr and sqle; (Horton, J. D. *et. al.*, 1998; Murphy, C. *et. al.*, 2006) were upregulated as early as 24 h after infection. These data are consistent with a dataset published recently, where the transcriptome of uninfected MΦ and MΦ infected with *L. amazonensis* amastigotes purified from mouse lesions, were compared 24 h post infection (Fortea, J. *et. al.*, 2009). Interestingly, so far this upregulation in sterol-biosynthesis-involved-genes has only been observed for *L. mexicana* and the closely related strain *L. amazonensis*. This might be because this subspecies resides and multiplies within a communal PV whereas *L. major*, *L. donovani* and *L. chagasi* do not. Gregory and colleagues compared *L. major* or *L. donovani* infected MΦ to inert latex bead harbouring macrophages and no changes in gene expression were detected (Gregory, D. J. *et. al.*, 2008). Differences between species or experimental approach may account for these contradicting findings. Firstly, macrophages were infected with promastigotes and not amastigotes (Gregory, D. J. *et. al.*, 2008). Secondly, synchronous uptake was not attempted since 16 hours were allowed for infection, and 24 h post infection cells were harvested. Thirdly, Gregory *et al.* used a strict 2.5 fold change threshold for an up- or downregulation. Fortea *et al.* used a 5% significance level to assign up- or downregulation, whereas a 2-fold change was considered significant here. The 24 h time period is not sufficient to transform from promastigotes to amastigotes which is only complete after approximately 48 h – 120 h (Antoine, J. C. *et. al.*, 1998). Rodriguez *et al.* also used promastigotes of the genus *L. chagasi* and investigated the change in gene expression 1, 4 and 24 h post infection. Activation

of anti-inflammatory and down regulation of pro-inflammatory genes involved in immune responses was observed. This was maybe biased as they used BALB/c bmdms, and it is known that this mouse strain is susceptible for *Leishmania* infection (Handman, E. *et. al.*, 1979) and generates a Th2-type (anti-inflammatory) immune response (Heinzel, F. P. *et. al.*, 1991; Scott, P. *et. al.*, 1988). The upregulation of genes encoding for proteins involved in the sterol biosynthesis, lipid uptake and encoding for fatty acid binding proteins seems to be attributed either to amastigotes or *L. mexicana* and closely related strains, and it is not characteristic of the host cell genetic background, as this was observed in C57BL/6 (in this study) and BALB/c (Fortea, J. *et. al.*, 2009) mouse strains.
It has been shown that *Leishmania* parasites use precursors of the host cholesterol biosynthesis in order to synthesise ergosterol, a sterol predominant in trypanosomatids membranes. Cholesterol sequestration could be relevant for ergosterol synthesis although no conversion from cholesterol to ergosterol has been shown yet. It is rather synthesised from acetate (Roberts, C. W. *et. al.*, 2003) and leucine (Ginger, M. L. *et. al.*, 1999). But it is known that more than 10% of the sterol content of leishmanial membrane, is cholesterol which is probably host cell derived (Roberts, C. W. *et. al.*, 2003).

Dendritic cells infected with *L. mexicana* amastigotes also upregulated proteins involved in cholesterol biosynthesis (*scd2* and *fdps*) 24 h post of infection (Aebischer, T. *et. al.*, 2005). In this study microarray data were also available for *lipa*, *npc2* and *adrp* which were downregualted after 24 h infection. This is different to macrophages in which expression in these genes was not affected. It seems that *Leishmania* has different effects on the lipid metabolism in DC than in Mϕ, but depletion of cholesterol seems to occur in both cell types. In DC the mRNA level of sterol carrier protein 2 (*scp2*) was upregulated 24 h post infection, whereas it was not changed in MΦ.

Genes regulated by PPARγ or LXRα were in general not significantly changed, apart from fatty acid binding protein 4 (Adipocyte lipid-binding protein) and fabp5 (epidermal fabp; Figure 33). These will be, together with fabp7 (brain lipid binding

protein), discussed separately. The study by Gregory *et al.* also did not detect changes in expression of these groups of genes, although they detected a downregulation of fabp7 and abca1 in *L. major*, but not in *L. donovani*, promastigote infected macrophages compared to inert latex beads. It can be concluded that the oxysterol pool in the cell is not altered upon infection, as LXRα genes remain unchanged in their expression profile. Partial activation of PPARγ genes indicates a surplus of PPARγ ligands. Free fatty acids are either esterified with FC or with glycerol, to form phospholipids. This surplus of free fatty acids is maybe due the impaired esterification to cholesterol as cholesterol is depleted. There are also indications that these fatty acids are maybe derived from the catabolism of sphingolipids, as discussed later. Free fatty acids are imported into the nucleus and activate PPARγ genes which results in upregulation of its target genes (Desvergne, B. and Wahli, W. 1999; Tontonoz, P. and Spiegelman, B. M. 2008).

No changes in gene expression was detected for genes encoding proteins involved in hydrolysis of neutral lipids (lipa, lipe; Brasaemle, D. L. 2007; Lian, X. *et. al.*, 2005), lipid storage (plin; (Brasaemle, D. L. 2007), lipid transport (scp2; (Frolov, A. *et. al.*, 1996; Maxfield, F. R. and Mondal, M. 2006) and generation of lipid signalling molecules (alox15; (Middleton, M. K. *et. al.*, 2006). These genes are not under the transcriptional control of the above discussed transcription factors. For two genes of this group a trend in upregulation was detected, i.e. proteins involved in the esterification of FFA and FC (soat1; (Welch, C. L. *et. al.*, 1996) and involved in the sphingolipid metabolism (sgpl1; (Bandhuvula, P. and Saba, J. D. 2007; Johnson, K. R. *et. al.*, 2003).
These results indicate that the lipid homeostasis is altered in infected MΦ. The activation of SREBP and failure in activation of LXRα indicates a depletion of FC in the cell, hence activation of SREBP.

Fortea et al. hypothesised the sequestration of cholesterol in PV membrane is maybe beneficial, especially for *L. mexicana* which live in communal PV and expand over the course of infection. Fortea et al. suggested that retained

cholesterol perhaps plays a role in PV lipid rafts formation in order to rearrange the PV membrane, or controlling fusion and fission events (Fortea, J. *et. al.*, 2009). This could ensure a constant supply of nutrients, and incorporation of proton pumps to maintain an acidic pH. It was reported that 80% latex bead containing compartments were positive for flotilin-1, a protein enriched in lipid rafts, but only 20% of *L. donovani* containing phagosomes were positive for flotilin-1 (Dermine, J. F. *et. al.*, 2001), which would not support Forteas suggestion. Staining of infected macrophages with filipin also did not support Fortea and colleagues' hypothesis of lipid raft formation on the parasites harbouring PV. Filipin staining resulted in a characteristic halo around the parasites but no filipin positive structures could be detected on the PV FIGURE 35. Although staining of axenic amastigotes did not result in the characteristic halo after filipin staining it can not be completely excluded that filipin had stained ergosterol. Furthermore, a statistically significant retention of LDL-derived cholesterol in parasite harbouring compartments compared to IgG coated latex bead containing compartment was identified here (Table 12).

It was reported that infection of macrophages with *L. donovani* disrupts lipid rafts on the plasma membrane and thereby impairs antigen presentation (Chakraborty, D. *et. al.*, 2005). The ability to activate a T-cell hybridoma was restored after treatment of infected macrophages with exogenous cholesterol liposomes (Chakraborty, D. *et. al.*, 2005). Data of this study are in agreement with data of Chakraborty and colleagues. Delivery of LDL derived cholesterol to the plasma membrane is very fast with a half time of approximately 45 min after internalisation and remained there for more than 6 h in Chinese hamster ovary cells (Brasaemle, D. L. and Attie, A. D. 1990). Brasaemle and Attie further reported a lag phase of 30 min, which was due to chilling of the cells to 4 °C to promote binding of LDL but no uptake prior to the shift to 37 °C. These times are well in agreement with the applied incubation time here, 1 h pulse and 2.5 h chase. Lange *et al.* reported that approximately 85% of LDL-derived free cholesterol was associated with the plasma membrane at all times during a 2 h chase experiment in rat hepatoma cells (Lange, Y. *et. al.*, 1997).

This means *L. mexicana* amastigotes harbouring compartments retained or accumulated LDL-derived cholesterol and this was not an observation due to the chosen time point. *Leishmania mexicana* also induced the formation of large vacuoles (Antoine, J. C. *et. al.*, 1998) and thereby display characteristics of Niemann-Pick-cells. Increased lipid content and enlarged late endosomes / early phagosomes are also characteristic of lipid storage diseases such as Niemann-Pick type C. It is characterised by lysosomal storage of cholesterol and gangliosides which result from defects in intracellular lipid trafficking (Wojtanik, K. M. and Liscum, L. 2003). The mechanism, of how exactly the cholesterol is accumulated in the lysosomes of Niemann-Pick-cells is still a matter of debate.

Most if not all, newly synthesised cholesterol is transported to the sphingolipid rich domains, lipid rafts, of the plasma membrane. Later it is delivered to late endosomes/ lysosomes. This was identified in Niemann-Pick-cells where newly synthesised cholesterol several hours after its initial synthesis accumulates in late endosomes/ lysosomes (Chang, T. Y. *et. al.*, 2006). In the late endosomes/ lysosome it enters the pool of LDL-derived cholesterol and is indistinguishable from the LDL-derived cholesterol.

The retention or accumulation of cholesterol in the PV may also be beneficial for the parasites for other reasons. When lipid rafts are disrupted antigen presentation is impaired (Chakraborty, D. *et. al.*, 2005) and signal transduction of e.g. CD40, a co-stimulatory molecule which binds CD40 ligand (CD40L) of T-cells (lymphocytes which play a role in the cell-mediated immunity) is altered. Functional rafts and CD40L binding to CD40 leads to an intense crosslinking of CD40. This in turn leads to a strong stimulation of the macrophage and the production of the pro-inflammatory cytokine IL-12. On the other hand, infection of macrophages with *L. major* led to the disruption of lipid rafts and due to this disruption, CD40/CD40L crosslinking was impaired resulting in a weak stimulation. Weak stimulation triggered a different signal cascade in the macrophage and the anti-inflammatory cytokine, IL-10, was produced (Rub, A. *et. al.*, 2009). Downregulation of CD40 mediated IL-12 production upon phagocytosis of *L. mexicana* amastigotes has already been reported (Weinheber, N. *et. al.*, 1998). Resistance to *Leishmania* infection is associated with production of IL-12 and

interferon-γ (Afonso, L. C. *et. al.*, 1994; Sypek, J. P. *et. al.*, 1993) which results in the production of reactive nitrogen species and clearance of *Leishmania* (Louis, J. *et. al.*, 1998). Susceptibility to *Leishmania* is associated with the production or responsiveness to IL-10 (Belkaid, Y. *et. al.*, 2001; Kane, M. M. and Mosser, D. M. 2001). IL-10 only contributes to susceptibility but the extent depends also on the parasites species. BALB/c mice deficient for IL-10, are unable to control infection to *L. mexicana* and *L. amazonensis* (Padigel, U. M. *et. al.*, 2003), whereas they can control infection with *L. major* (Kane, M. M. and Mosser, D. M. 2001). Deficiency of IL-10 in C57B/6 mice infected with *L. mexicana* on the other hand promoted healing (Jones, D. E. *et. al.*, 2002).

Interestingly, for macrophages differentiated from human monocytes in the presence of macrophage colony stimulating factor (M-CSF) and IL-10, a 30% higher uptake in LDL was observed compared to monocyte derived macrophages differentiated only with M-CSF (Zhao, B. *et. al.*, 2006). In another study IL-10 led to an enhanced oxLDL accumulation in THP-1 macrophages compared to control cells incubated without IL-10 (Halvorsen, B. *et. al.*, 2005).

In summary, retention of cholesterol by the parasites leads to disruption of rafts on the plasma membrane. This disruption leads to an impaired antigen presentation and an altered signalling cascade in the MΦ. The production of IL-12 decreased whereas IL-10 production increased. The latter prevents the activation of the MΦ and may even result in an increase in LDL-uptake.

Lipid rafts are detergent-resistant membrane domains enriched in cholesterol and sphingolipids (Brown, D. A. and London, E. 1997). As a consequence of cholesterol depletion and the lipid raft disruption, the plasma membrane contains an unbalanced ratio of cholesterol and sphingolipids. The trend observed here in the upregulated gene expression of sphingosine-1-phosphate lyase (sgpl1), as it catalyses the final step of sphingolipid catabolism, could represent the removal of excess sphingolipids from the cell. During the catabolism of sphingolipids FFA are released, which could contribute to the partial activation of PPARγ. As mentioned previously, transcript levels of two of the three investigated intracellular lipid chaperones or fatty acid binding proteins were identified to be upregulated in

macrophages upon *Leishmania* infection, namely fabp4 and fabp5. A trend in the upregulation of protein level of fabp4 was detected 48 h after infection throughout the remaining time course, whereas this was not the case for fabp5. For fabp7 no upregulation in gene expression was detected, but a similar effect as for fabp4, a trend of increased protein level 48 h post infection compared to uninfected MΦ.

An upregulation of fabp4 gene expression was also described in murine macrophages infected with *Streptococcus pyogenes*, the causative agent of a variety of clinical manifestations, including infections of the skin and upper respiratory tract, bacteremia, and occasionally sepsis and septic shock (Goldmann, O. *et. al.*, 2007). According to the gene expression profile Goldmann and co-workers discovered that *S. pyogenes* infected macrophages were in between classical and alternatively activated macrophages. It was also reported that fabp4 deficient mice in the presence of IL-4 upregulate CD36, abca1 and LXRα which, combined, promoted uptake of modified LDL and increased cholesterol export. In contrast, only small amounts of mRNA for these genes were detected when fabp4 was present and the macrophages were stimulated with IL-4 (Makowski, L. *et. al.*, 2005). This is in agreement with our data - LXRα target genes (e.g. abca1, apoe, npc1 and npc2) as well as CD36 were not upregulated and uptake of ^{3}H-labelled LDL was not altered in *L. mexicana* infected MΦ compared to beads exposed or uninfected MΦ. Consistent with our data was the upregulation of fabp4 and fabp5 in *L. amazonensis* infected BALB/c bmdm 24 h post infection (Fortea, J. *et. al.*, 2009). Fortea *et al.* also reported that mRNA levels of fabp4 were upregulated to a higher extent than of fabp5, as it was found here. A possible biological function for the upregulation of fatty acid binding proteins also at the protein level, could be that these proteins compensate the lack of cholesterol and bind FFA and transport a fraction to the peroxisomes (Furuhashi, M. and Hotamisligil, G. S. 2008) for degradation, which would result in acetate generation (Reddy, J. K. and Mannaerts, G. P. 1994) a precursor for cholesterol (Ikonen, E. 2008).

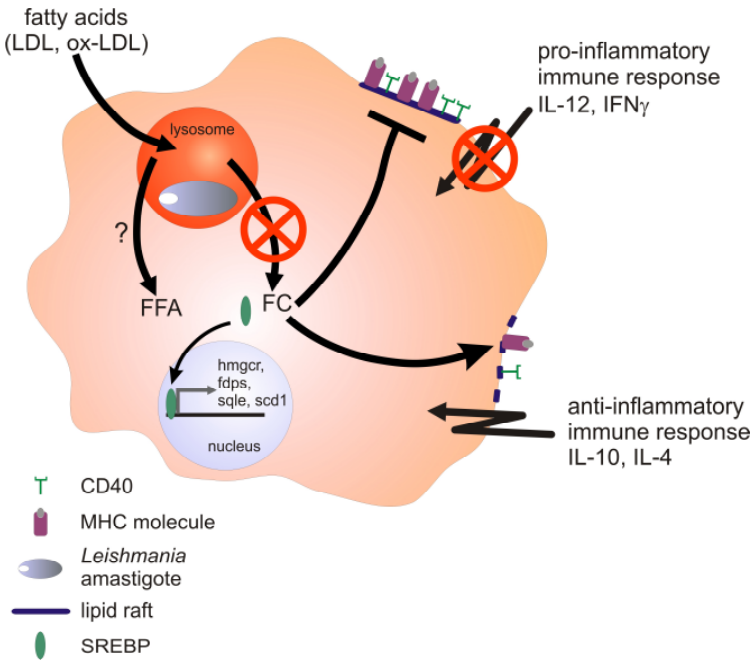


FIGURE 38. Altered lipid homeostasis and immune function of *Leishmania* infected MΦ.

Leishmania spp. reside in the PV and retain LDL-derived cholesterol. This affects the integrity of lipid rafts on the host cells, which leads to an impaired antigen presentation. Furthermore, due to the disruption of lipid rafts the signalling in the host cell is altered. This leads to an anti-inflammatory immune response instead of a pro-inflammatory immune response. The MΦ senses a lack of cholesterol and the transcription factor SREBP becomes activated and induces the expression of genes encoding proteins involved in cholesterol biosynthesis. The fate of FFA could not be determined.

In summary, *Leishmania* infection altered the lipid homeostasis in the macrophage host cell. *Leishmania* infection led to retention of cholesterol in the PV. This retention is detrimental, rafts are disrupted which could lead to an impaired antigen presentation (Chakraborty, D. *et. al.*, 2005) and an altered signalling in the macrophage. The affected signalling could downregulate IL-12 (Rub, A. *et. al.*, 2009; Weinheber, N. *et. al.*, 1998) and upregulate IL-10 (Rub, A. *et. al.*, 2009) which induces an anti-inflammatory response and prevents the killing of the parasites. IL-10 possibly stimulates the LDL-uptake and allows the parasite to thrive within the PV. Inactivation of the macrophage seems to occur as early as parasites are transmitted during a blood meal. It has been shown that saliva of *P. papatasi* and *L. longipalpis*, *Leishmania* transmitting sandflies, can induce IL-4 in epidermal cells and IL-10 in macrophages, respectively (Norsworthy, N. B. *et. al.*, 2004; Monteiro, M. C. *et. al.*, 2007; Mbow, M. L. *et. al.*, 1998). Recently it has been

shown that proteophosphoglycans regurgitated during transmission by *L. longipalpis* enhance an alternative activation of macrophages (Rogers, M. *et. al.*, 2009).

It appears that during the transmission of the parasite the sandflies initiate an inactivation of the Mϕ and the parasite has developed a mechanism to maintain the inactivity of Mϕ.