Chapter 6

General discussion

A better understanding of the biology of *Plasmodium* parasites is crucial for the identification of novel drug targets to treat malaria. Our understanding of the molecular cues that drive *Plasmodium* development and differentiation, and how it can be interfered with has been limited by the fact that only a small proportion of genes have been assigned a function experimentally.

Signature tagged mutagenesis has been used extensively in bacterial pathogens to identify virulence genes by parallel phenotyping of pools of individually tagged mutants [91]. The main aim of this dissertation was the development of such strategies for the study of *P. berghei* biology.

A three-step strategy was developed that involved (1) parallel transfection of barcoded *Plasmo*GEM vectors, followed by (2) a propagation step during which small blood samples were collected from day 4 to 8 post-transfection, and finally (3) calculation of the fitness of each population of mutants within the pool, through sequencing of their barcodes. Optimisation of each step was followed by validation of the method with protein kinases chosen as test genes. Using this strategy, complex and defined pools of KO mutants were reproducibly generated in a single mouse. Barcode sequencing enabled the measurement of the daily fitness of each mutant and how it evolved on the course of infection. For the first time in malaria research, hundreds of mutants can now be screened for quantitative growth phenotypes, which is a much more powerful approach than the traditional bimodal essential/redundant classification.

One way of using STM is in drop out screens, where disruption of essential genes produces lethal phenotypes and its consequent loss from the study pool. However, *in vivo* systems have a high selection pressure. While this promotes competition and reduces false positives it also prevented the survival of mutants with a fitness lower than 0.45 of the WT-like references. Despite the absence of fitness data for these highly attenuated or non-viable mutants, knowing which genes are essential for parasite development is highly valuable for drug discovery research. Coupling the STM/bar-seq strategy with the newly developed conditional systems such as DiCre mediated gene excision [202], FLP/FRT mediated conditional mutagenesis [64] or protein destabilisation domain strategies [61] will be crucial to validate targetability of this group of highly attenuated or non-viable mutants.

In some pilot experiments, I asked whether mutants for which barcodes were either missing at the end of the experiments (i.e. where likely essential genes were disrupted), or were inconsistent, could be observed more abundantly when mutants with normal values of fitness were excluded from the transfection. After this, some of the cases that had previously

been considered “inconsistent”, because no growth measurements had been obtained for at least two out of three replicates, were rescued. However, the absence of growth references prevented the calculation of fitness values for these mutants.

Although the throughput and sensitivity of the STM approach are unparalleled in *Plasmodium*, some caveats need to be considered. As with any gene knock out experiment if a targeting vector fails to integrate, the consequent absence of a given mutant in the final pool might be mistaken as gene essentiality. Furthermore, while genetic screens can generate new hypotheses, validation and follow-up experiments still require the independent generation of clonal lines. Importantly, this technology is of limited utility to study mosquito stages. Random fertilisation between the different genotypes present in a pool generates heterozygous (i.e. double mutant) zygotes that are not detectable by the barcoding method. The fact that barcodes are amplified through the same PCR reaction makes it impossible to pinpoint the origin of each barcode. One approach to overcome this limitation is to generate pools of mutants in a background where production of one of the gametes is impaired, for instance a male defective mutant like the *map2* KO, and then supply the cultures with the missing set of gametes. This would ensure that the “male” barcode is always the same. However, these experiments would have to be analysed as genetic interaction experiments. Beyond the zygote stage, meiosis complicates this issue further due to unpredictable cross-over events that may either repair disrupted loci and generate WT alleles or create mutants with several genes deleted.

It will be possible, however, to use barcode counting for the identification of genes required for gametocyte maturation and fertility by asking which barcodes are present among the asexual population in an infection, but absent from the gametocyte population. Finally, it is important to remember that the STM study design makes it impossible to detect essential genes whose function is compensated by other mutants in the pool. However, this is unlikely to be a frequent event.

The new parallel transfection approach coupled with bar-seq detection of the mutants within a pool proved to be more efficient and sensitive than previous strategies and allowed the detection of nine new mutants including *gsk3*, a gene previously targeted for drug development [169]. This finding was validated with a cloned mutant and its phenotypic analysis revealed a strong reticulocyte preference. In normal culturing conditions only normocytes are used as hosts for the parasites. If the *gsk3* KO phenotype seen in *P. berghei*

extends to *P. falciparum*, the absence of reticulocytes in culture would be the reason why the mutant could not be generated. Understanding what determines this host cell specificity (i.e. normocytes versus reticulocyte) would be of great relevance for drug development research. For instance, if drugs are directed against factors like GSK3, although this would be highly effective in culture, it would be of little use *in vivo* since *P. falciparum* parasites can also invade reticulocytes.

The properties of the STM approach have enabled large scale genetic screens with a dramatic reduction in workload and also a sharp decrease in the animal usage. Previously, in order to check a set of genes for redundancy one would need at least as many animals as genes. With the strategy presented here, I was able to screen a list of 50 genes in a single mouse. For a new project that will systematically screen 1000 metabolic enzymes, the vector pool size has been scaled up to ~100 vectors. Such large scale screens are now possible because data from each experiment are directly comparable and can be analysed together. This clear increase in scale offers new opportunities for reverse genetic screening that were previously unachievable for malaria parasites. I have demonstrated this here for the study of genetic interactions at scale by screening 41 ePK genes in six different mutant lines for growth phenotypes. This screen generated fitness data for 258 double or triple mutants and revealed an important interaction between *cdpk4* and *pkg*.

Biological circuits are protected by a certain degree of redundancy to ensure robustness of biological processes. However, some genes are central to several networks and cannot be disrupted. These are referred to as "network hubs" [100]. The study of these genes often requires the use of hypomorphic alleles or conditional strategies. The negative interaction between *pkg* and *cdpk4* was exactly one such example revealed by the use of a hypomorphic allele of the hub *pkg*.

Chemical genetics is another potential application for the barcoding approach. This is performed by testing barcoded mutants for sensitivity to sub-lethal concentrations of chemicals. These mutants can be the result of loss-of-function, point mutations (resistant or hypomorphic alleles) or even over-expression of genes. The principle applied here is similar to that of genetic interactions and greatly relies on the detection of unexpected fitness phenotypes of mutants in the presence and absence of chemicals. In yeast, hierarchical clustering of data generated by such screens revealed a number of genes associated with multidrug resistance when overexpressed. These genes, if disrupted, made the strains sensitive

to diverse compounds [203]. This kind of approach has also the potential to group together genes of unknown function based on the likely mode of action of a chemical. This would be highly relevant for malaria parasites as half of the protein coding genes lack annotation.

Progress in recent years in the field of genetics has suggested that gene essentiality is contextual, i.e. that a phenotype results from the interplay between the genotype and the environment in which it is expressed. The availability of a high throughput approach now offers the possibility of investigating not only gene-gene interactions, but also gene-environment interactions. For instance, in collaboration with Maria Mota and Liliana Mâncio (IMM, Lisbon) we performed some preliminary experiments that looked at the impact of the nutritional status of the host in the fitness of a set of mutants.

Other examples of environment-gene interactions include for example the use of different strains of receptor mice where genes controlling, for instance, cell receptors or immunological factors, could be either disrupted or over-expressed. Such experiments would be highly relevant for the study of host-pathogen interactions.

Plasmodium species diverged from their last common ancestor 67.8 million years ago [204]. It can therefore be expected that significant differences have emerged between the rodent parasites and those infecting humans. In fact, in some aspects the differences are obvious, for example the duration of the asexual cycle is 24 h in *P. berghei* and 48 h in *P. falciparum*, and the time required for the liver stage development is 48 h for the first and close to two weeks for the second. As a result, direct translation of findings between species can be limited in some cases. For this reason transfer of the STM approach to *P. falciparum* would be of great clinical relevance. The recent development of efficient site specific genome editing technologies for *P. falciparum* parasites such as zinc finger nucleases (ZFNs) [205] and CRISPR/Cas9 [206] have substantially increased the efficiency and ease with which genes can be targeted in these parasites. On the other hand, *P. knowlesi* parasites have recently been adapted to culturing conditions and arise now as a promising model to study human malaria [207]. These parasites are 1,000 fold more amenable to genetic manipulation than *P. falciparum* and, similarly to *P. berghei* parasites can be transfected with linear DNA [207].

As a result, the same parallel transfection of barcoded vectors could potentially be applied to these parasites. Furthermore, the fact that these parasites can be propagated in culture would certainly bring a new dimension to malaria reverse genetics. The disadvantage

of this system is the absence of host selection pressure and natural conditions such as reticulocytes as mentioned above. This will likely influence the rate of false positives, for example due to the absence of clearance of degenerated parasites that carry a barcode, and false negatives, i.e. if in natural conditions a given mutant would be viable, which will need to be controlled for. In my opinion this will be the next great challenge in the *Plasmodium* biology field.

In conclusion, the kind of high throughput genetic approach described here provides the basis for future screenings on large subsets of parasite genes. These will greatly contribute to the understanding of the molecular biology of *Plasmodium* parasites and will certainly take malaria research closer to eradication of this disease.