Rapid Microarray-Based Method for Monitoring of All Currently Known Single-Nucleotide Polymorphisms Associated with Parasite Resistance to Antimalaria Drugs

Andreas Crameri,†‡§ Jutta Marfurt,†‡ Kefas Mugittu,‡ Nicolas Maire,† Atilia Regös,¶ Jean Yves Coppee,‖ Odile Sismeiro,‖ Richard Burki,‖‖ Eric Huber,†† Daniel Laubscher,†† Odile Puijalon,‡ Blaise Genton,†‡‡ Ingrid Felger,† and Hans-Peter Beck*†

Swiss Tropical Institute, Socinstrasse 57 Basel, Switzerland; Ifakara Health Research and Development Centre, P.O. Box 53, Ifakara, Tanzania; Plate-forme puces a ADN, Genopole/Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France; and Unité d’Immunologie Moléculaire des Parasites, CNRS URA 2581, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Received 12 June 2007/Returned for modification 8 August 2007/Accepted 23 August 2007

Parasite drug resistance is partly conferred by single-nucleotide polymorphisms (SNPs), and monitoring them has been proposed as an alternative to monitoring drug resistance. Therefore, techniques are required to facilitate analyses of multiple SNPs on an epidemiological scale. We report a rapid and affordable microarray technique for application in epidemiological studies of malaria drug resistance. We have designed a multiwell microarray that is used in conjunction with PCR-amplified target genes implicated in the drug resistance of malaria with subsequent one-tube minisequencing using two fluorochromes. The drug-resistance-associated genes pfRhfr, pfHps, pfTert, pfmdr1, and pfATPase were amplified and analyzed for cultured Plasmodium falciparum strains and from field samples. We obtained a specificity of 94%, and comparison of field sample results to those of restriction fragment length polymorphism (RFLP) typing resulted in an overall consistency of >90%, except for pfRhfr51, for which most discrepancies were due to false determinations by RFLP of mixed infections. The system is sufficiently sensitive to assay parasites in clinical malaria cases and in most asymptomatic cases, and it allows high throughput with minimal hands-on time. The cost for the assay has been calculated as 0.27 euros/SNP (US$0.33), which is below the cost incurred with other systems. Due to the simplicity of the approach, newly identified SNPs can be incorporated rapidly. Such a monitoring system also makes it possible to identify the reemergence of drug-susceptible parasites once a drug has been withdrawn.

Parasite resistance to antimalarial drugs has become a major public health concern for areas in which malaria is endemic, and it has become a threat to malaria control programs (16). Therefore, the monitoring of antimalarial drug efficacy has become an integral part of national health systems. The efficacy of first-line antimalarial drugs currently is monitored primarily by in vivo methods. Such investigations pose major problems in terms of recruitment, costs, and adherence to follow-up visits. In areas in which malaria is highly endemic, efficacy studies are confounded by new infections during the follow-up period (23). In addition, such a monitoring system does not allow the determination of the efficacy of drugs that have been discontinued as a result of decreased efficacy and the efficacy of which has been reported to reemerge once drug selection pressure had ceased (14, 25).

Resistance of the malaria parasite Plasmodium falciparum to antimalarial drugs often is conferred by single-nucleotide polymorphisms (SNPs) or gene duplications. Resistance to sulfadoxine-pyrimethamine is conferred by point mutations in the dihydrofolate reductase (dfr5) gene at codons A16V, N51I, C59R, S108N/T, and I164L. Resistance is further augmented by mutations in the dihydropteroate synthase (dhs5) gene (436Phe, 437Gly, 540Glu, 581Gly, and 613Ser/Thr) (7). SNPs in genes encoding putative transporter molecules, such as multidrug resistance gene 1 (mdr1) or the chloroquine resistance transporter (crt), have been implicated in resistance to 4-aminopyrimidines (4, 20) as well as in gene amplification of mdr1 (18, 21). Recently, mutations in the plasmoidal ATPase6 gene have been associated with decreased susceptibility to artesinins (11).

The monitoring of parasite drug resistance has the potential to become a tool for long-term surveillance and for developing predictive models of malaria drug resistance (5, 17). For this purpose, a technique is required that facilitates...
parallel analysis of multiple SNPs that is affordable and applicable on an epidemiological scale. A number of standard methods exist for SNP analysis, mostly based on PCR-restriction fragment length polymorphism (RFLP) of selected loci or on sequence-specific amplification or hybridization (19). Unfortunately, most of these techniques do not allow the analysis of many SNPs on an epidemiological scale, and consequently many previous studies have analyzed only a few SNPs deemed primary predictors of resistance. Little attention has been paid to mutations that are not directly associated with resistance but that are considered to have modulating or compensatory effects.

Here, we present a microarray-based system to determine all known SNPs in drug-resistance-associated *P. falciparum* genes. In relation to previously used techniques, costs are significantly lower, and large numbers of samples can be analyzed in a reasonably short period of time. We have shown that this technique can be transferred and run in laboratories with minimal infrastructure (J. Marfurt, unpublished data, and K. Mugittu, unpublished data). The system also is flexible and amenable to many other applications requiring SNP analyses.

### MATERIALS AND METHODS

Our method is based on parallel PCR amplification of the target sequences, followed by primer-extension-mediated minisequencing using fluorochrome-labeled deoxyinosine triphosphates (ddNTPs). Subsequent base calling occurs on a microarray upon sequence-specific hybridization. The flow chart in Fig. 1 depicts schematically the principle of the parallel SNP analysis system.

**Analyzed material.** To establish and evaluate the technique, cultural material from strains 3D7 and K1 and samples collected during drug efficacy studies using molecular analyses of drug resistance markers in Papua New Guinea have been used. Ethical clearance has been granted for this study by the Papua New Guinea Medical Research Council.

**Blood samples and DNA preparation.** Blood samples were collected either in EDTA Microtainer tubes (BD Biosciences, Allschwil, Switzerland) or on Isocode sticks (Schleicher and Schuell, Dassel, Germany). Plasma was separated from blood samples in EDTA by centrifugation, and red blood cell pellets were stored frozen until use. DNA from cultures and field samples was extracted from 50- to 100-μl blood pellets using QIAamp DNA blood kits (Qiagen, Hombrechtikon, Switzerland) according to the manufacturer's instructions. DNA was eluted from Isocode sticks according to the manufacturer’s recommendations.

**DNA amplification.** We analyzed the following 36 polymorphisms in five genes at 32 SNP sites: *pfmdr1* codons N86Y, Y184F, S1034C, N1042D, and D1246Y; *pfcrt* codons C725E, K767T, H970Q, T1252A, S163R, Q271E, N326D/S; *pfmdrl* codons S353N, A443G, K540E, A670V, and H691P; and *pfatp6* codons S535R, Q574P, A623E, N683K, and S769N. Oligonucleotides for amplification, extension, and arraying are shown in Table S1 in the supplemental material. To cover all SNP sites, we performed 10 PCRs with the amplification primers listed in Table S1 in the supplemental material. The amplification reaction mixture contained 1× PCR buffer with MgCl2 in a final concentration of 5 mM, 0.2 mM deoxyinosine triphosphates, and 0.2 μM of each primer. Reactions were carried out in 50 μl containing 2.5 μl DNA and 2.5 μl of a 1:10 dilution of the pooled PCR product was digested with 2 U shrimp alkaline phosphatase (SAP) (Amersham Biosciences, Freiburg, Germany) in a reaction volume of 12 μl for 1 h at 37°C. SAP was inactivated by incubating samples for 15 min at 90°C.

Since most microarray scanners support only dual-fluorescence measures simultaneously, a strategy of two parallel reactions had to be applied. Two primer extension reactions were carried out per sample. The reaction mixtures differed in their combinations of Cy3- and Cy5-labeled ddNTPs (Perkin Elmer, Courbevoie, France). Extension primers were added as shown in Table S2 in the supplemental material. Thus, it was possible to detect all possible SNP permutations by using only two fluorochromes. All primer extension reactions for one sample were carried out in two aliquots of 20 μl containing 1× Sequenase buffer, extension primer mix 1 or 2, ddNTP mix 1 or 2 (see Table S2 in the supplemental material), and 2 U Thermo Sequenase (Terminol; Solis). The concentration of ddNTPs was 0.25 μM, and primers were diluted to a concentration of 0.25 μM each. The extension reaction was carried out for 90 s.

**Chip hybridization.** Twenty-three microarrays of the pooled and denatured primer extension reaction mixture were hybridized to a well of a microarray glass slide, and 6 μl 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) was added. Hybridization was carried out in a humid chamber at 50°C for 60 to 90 min. After hybridization, the slide was washed at room temperature in 2× SSC plus 0.1% SDS for 20 min, followed by another wash with 2× SSC for 20 min, and then a final wash with 2× SSC plus 2% ethanol for 2 min. The slides were dried with compressed air and stored in the dark until scanned.

**Data acquisition.** Hybridized slides were scanned at 635 and 532 nm using an Axon 4100A fluorescence scanner (Buchler Biotech AG, Basel, Switzerland). Cy3 and Cy5 images were acquired using microarray analysis software (version 6.0) software. This software generates data points using pixel intensity after background subtraction. We developed software for further analysis of raw data that produces an output determining whether an infection is wild type, mutant, or mixed, and it also determines the dominant genotype in the latter case. Each signal was classified either as wild type, mutant, or mixed based on the expression intensities of the scanned image. The grouping was done according to the following algorithm: fluorescence intensities below 9,000 (Cy3) or 10,000 (Cy5) U (mean intensities minus background) were regarded as negative. For measures above these cutoff values, we considered the ratio of Cy5 intensity to Cy3 intensity to discriminate between wild-type, mutant, or mixed infection.

To determine an optimal algorithm to translate the output of the GenePix Pro software into predictions about the genotypes present in analyzed samples, we used two singly infected blood samples that previously were sequenced at 29 SNP sites. Sequence data showed that the samples from 3 of 29 SNP sites differed from the published strain 3D7 (GenBank accession number AY356960). Either single or mixed samples were analyzed using the chip in various proportions (1:2, 1:4, 1:8, and 1:16). With this approach, we could empirically determine the following threshold values: for Cy5/Cy3 ratios below 0.7, the sample was classified according to whether the wild type or mutant was labeled with Cy3. Ratios between 0.7 and 2.4 were assigned to mixed, and it also determines the dominant genotype in the latter case. Each signal was classified either as wild type, mutant, or mixed based on the expression intensities of the scanned image. The grouping was done according to the following algorithm: fluorescence intensities below 9,000 (Cy3) or 10,000 (Cy5) U (mean intensities minus background) were regarded as negative. For measures above these cutoff values, we considered the ratio of Cy5 intensity to Cy3 intensity to discriminate between wild-type, mutant, or mixed infection.

To estimate the above-mentioned threshold parameters and to determine the predictive accuracy of our method, we used five of the four identical but independently processed microarrays to estimate the threshold value to distinguish positive from negative signals so that the results would match the sequence data as closely as possible. The fourth microarray then was used to apply this algorithm to determine the predictive accuracy of the method. This procedure was repeated four times in all possible combinations. Finally, we applied this algorithm to
samples that were genotyped by sequencing and PCR-RFLP to determine the sensitivity and specificity of our method.

Sequencing. PCR products were purified by size-selective polyethylene glycol precipitation and directly sequenced using one of the respective nested PCR primers. Cycle sequencing (25 cycles of 96°C for 30 s, 50°C for 15 s, and 60°C for 4 min) was performed using the ABI PRISM Dye Terminator cycle sequencing ready reaction kit (Perkin Elmer), and sequences were analyzed using an ABI PRISM 310 genetic analyzer and the ABI PRISM software.

RESULTS

Amplification and sensitivity. Ten PCR fragments were required to cover all SNP locations. Amplicon sizes were 637 bp for pfdhfr; 686 bp for pfdhps; 799 and 526 bp for pfmdr1; five fragments of 630, 548, 476, 304, and 200 bp for pfcrt; and 798 bp for pfATPase6. To increase sensitivity, we applied a nested
PCR protocol and reamplified all PCR products. For assay validation, DNA was extracted from parasitized erythrocytes added in tenfold dilutions to noninfected human blood. Thus, the amount of template per PCR corresponded to the range of 1 to 100,000 parasites. All samples were subjected to a primer extension reaction and were hybridized onto the microarray. Figure 2 shows signals obtained from the dilution series at selected SNP sites. Except for \textit{pfmdr1} codon 1246, we were able to detect 1,000 parasites, which represents an approximate parasitemia of 15 parasites/\mu{l} and is mostly sufficient for clinical cases. Eight of 13 extension primers tested gave a signal with 10 parasites per reaction, and six primers were positive for 1 parasite only.

\textbf{Mixed infections.} Since naturally occurring blood samples often contain multiple \textit{P. falciparum} strains, we tested whether the presence of two different templates would decrease the sensitivity of detection. We mixed 3D7 parasites and K1 parasites, which differ in their genetic profile in \textit{pfldhfr} 59 and 108. While one parasite strain was kept at 1\% parasitemia, the other strain was serially diluted from 4 to 0.00125\%. When the dilution of K1 was kept constant, 3D7 gave acceptable results even at the lowest dilutions, except for \textit{pfldhfr} 59, with an endpoint at 0.0075\%. When 3D7 was kept constant, similar results were obtained with decreasing template concentrations of K1. The exception was a loss of the K1 \textit{pfldhfr} 59 signal at the dilution of 0.06\%. Figure 3 shows the signals for both strains at both SNP positions. This experiment showed that the dynamic range for quantification of signals is small.

\textbf{Specificity.} We have used 12 different culture strains from which we directly sequenced the genes \textit{pfmdr1}, \textit{pfcrt}, \textit{pfldhps}, and \textit{pfldhfr}, comprising 16 different SNP sites. Of these 192 SNPs analyzed, we failed to detect 1 SNP in one strain. A mixed signal was produced for four SNPs, and seven SNPs gave discrepant results compared to those from sequencing, three of which were at codon 86 of \textit{pfmdr1}. Based on the results for these 16 SNPs, we estimated an overall specificity of 94\% compared to the gold standard of sequencing. After that, we determined the precision of base calling for naturally infected blood samples. We compared data for samples that were previously analyzed by RFLP and sequencing to data from microarray analyses. Thirty-six PCR-positive samples from field studies in Papua New Guinea were analyzed by PCR-RFLP (7,
10, 15) for *pfmdr1* 86; *pfcrt* 76; *pfdfhr* 51, 59, and 108; and *pfdfhs* 437, 540, and 581. The mean multiplicity of infection for these samples, determined by genotyping the polymorphic *msp2* (merozoite surface protein 2) locus (8), was 1.56 (range, 1 to 4; 95% confidence interval, 1.29 to 1.84). Only 10 of these samples were positive by microscopy, and densities were between 80 and 9,440 asexual parasites per μl blood, with a geometric mean density of 1,731/μl. Table 1 summarizes the concordance between microarray and PCR-RFLP analyses.

In addition, 12 of these samples harboring a single clone infection were sequenced. There was excellent agreement between the microarray data and sequencing results (data not shown).

**Costs.** Because we developed this microarray system to monitor parasite drug resistance to antimalarial drugs in resource-constrained countries, it was essential to keep costs as low as possible so that the system can be used routinely for drug resistance monitoring. The cost calculation included plastic material for DNA preparation, PCRs, primer extension with fluorochromes, and microarray production. However, it does not take into account acquisition, maintenance, and amortization of equipment, nor does it take into account labor costs. We calculated a price of 0.27 euros (US$0.33) per SNP for determining 32 SNP sites per sample and simultaneously analyzing 12 samples on one slide.

**DISCUSSION**

Monitoring parasite resistance to antimalarial drugs has become an essential part of the malaria control programs in countries in which malaria is endemic. Common standards have been in vivo efficacy trials at health facilities (26), which

---

**FIG. 3.** Detection of SNPs in mixed parasite infections. The upper panel depicts arbitrary fluorescence values obtained when strain K1 was mixed with various dilutions of the 3D7 strain. K1 and 3D7 differ at codons 59 and 108 in the *dhfr* gene. The lower panel shows arbitrary fluorescence values obtained when strain 3D7 was mixed with various dilutions of the K1 strain.
are time-consuming and labor-intensive. These studies are hampered, particularly in areas in which malaria is highly endemic, by the frequent reoccurrence of parasites from new inoculations. This leads to an underestimation of drug efficacy (23).

In order to circumvent these problems, systematic molecular monitoring of parasite-resistance-associated SNPs has been widely promoted and used to complement in vivo efficacy studies (3, 24). While current systems work well with small sample sets, unfortunately they are limited in the number of samples and SNPs that can be analyzed simultaneously. Furthermore, for epidemiological monitoring they become expensive, both in terms of equipment and running costs.

Therefore, large studies analyzing multiple SNPs of multiple genes in parallel have never been performed because of high costs and labor intensiveness. Here, we report a novel method that permits the simultaneous analysis of many SNPs in hundreds of samples in a very short time period (approximately 15 h for four 96-well plates) with significantly reduced costs. The microarray system was shown to be fast and accurate. In particular, the low detection limit of 10 to 100 parasites and the suitability for samples containing multiple infections represent added advantages over many competing systems. The significantly reduced cost per SNP compares favorably to the costs of other systems. In resource-restricted countries such as the Sub-Saharan countries in which parasite resistance to antimalarial drugs is a major concern (2), only a low-cost system permits molecular monitoring of drug resistance.

The most critical factors influencing this technique seem to be the quality of DNA and the need for a large enough amount of template for low-density cases. Although only small amounts are needed (less than 50 µl of blood), it is crucial that the material is of good quality; otherwise, some PCRs might fail.

In contrast to the analysis of diploid organisms, the analysis of *P. falciparum* infections poses an additional challenge because multiple infections commonly are found, leading to a highly skewed distribution of different templates within a blood sample (9). In addition, PCR amplification might favor the dominant templates. Therefore, it was important to ensure that minor template populations can be detected. We therefore designed an elaborate algorithm to determine the detection threshold for genotype calling. Evidently, however, some low-density infections may be missed for some individuals, whatever the threshold used. Whether this is important in the epidemiological assessments of resistance remains to be seen, because it is not clear to what degree these low-density infections contribute to disease and transmission.

It has been shown that a synergistic action of transmembrane transporters is involved in parasite resistance to antimalarial drugs. In addition to *pfcr* , another transporter involved in chloroquine resistance (*pfmdr1*), the homologue to the human P-glycoprotein, seems to contribute to resistance against chloroquine, the most commonly used drug against malaria (13). *pfmdr1* also has been shown to modulate resistance to mefloquine and related drugs (22). To date, however, no clear association has been shown between individual SNPs and parasitological failure of a given drug. Hence, it is possible that the parallel analysis of all SNPs in several genes might identify certain haplotypes suspected to be involved in drug resistance. With the prospect of analyzing all known drug-resistance-associated SNPs at once, elucidation of the genetic background of drug failure becomes feasible. This underscores the need for linking individual SNPs to haplotypes, because interactions between SNPs from different loci are likely to account for the phenotypic effect. However, current algorithms and techniques are not yet able to generate true haplotypes of unlinked loci in samples containing multiple infections of *P. falciparum*. In Tanzania, for instance, the mean multiplicity of infection for infections in children is five concurrent infections per child (1), considerably complicating or preventing the determination of haplotypes of individual *P. falciparum* clones. Although we use a nested PCR strategy, in most cases we were able to determine the most dominant haplotype with our array in a semiquantitative manner. Since parasite density is a correlate of malaria symptoms, the most dominant haplotype within a multiple-clone infection is likely to represent the clone actually causing clinical malaria.

We have now used our microarray system successfully for

### Table 1. SNP analysis of 36 field samples from Papua New Guinea and agreement between data obtained by microarray and PCR-RFLP methods

<table>
<thead>
<tr>
<th>Locus and codon</th>
<th>Total no. of samples</th>
<th>No. of samples with undetermined results by RFLP (n = 1)</th>
<th>No. of samples with undetermined results by microarray (n = 0)</th>
<th>Concordant&lt;sup&gt;b&lt;/sup&gt; (86.8% agreement [250 nt])</th>
<th>Mixed RFLP/single array&lt;sup&gt;a&lt;/sup&gt; (8.3% agreement [25 nt])</th>
<th>Mixed array/single RFLP&lt;sup&gt;c&lt;/sup&gt; (2.8% agreement [8 nt])</th>
<th>Single RFLP/single alternative RFLP&lt;sup&gt;d&lt;/sup&gt; (1.7% agreement [8 nt])</th>
<th>% Agreement (χ test; n = 288 nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>mdr1</em> 86</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>100 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>crt</em> 76</td>
<td>35</td>
<td>35</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>80 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhfr</em> 51</td>
<td>36</td>
<td>36</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>63.9 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhfr</em> 59</td>
<td>36</td>
<td>36</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>80.6 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhfr</em> 108</td>
<td>36</td>
<td>36</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>80.6 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhps</em> 437</td>
<td>36</td>
<td>36</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>94.4 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhps</em> 540</td>
<td>36</td>
<td>36</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>97.2 (χ test; n = 288 nt)</td>
</tr>
<tr>
<td><em>dhps</em> 581</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>100 (χ test; n = 288 nt)</td>
</tr>
</tbody>
</table>

<sup>a</sup> nt, nucleotides.<br><sup>b</sup> Samples were considered concordant if both the RFLP and microarray gave identical results.<br><sup>c</sup> Samples showing a single infection on the microarray but multiple infections by RFLP. Note that the determination of single infections by RFLP requires complete restriction digestion.<br><sup>d</sup> Samples showing multiple infections on the microarray but only a single infection by RFLP.
drug resistance monitoring in several sites for more than 3 years in Tanzania (K. Mugittu, unpublished), Papua New Guinea (J. Marfurt, unpublished), and the Solomon Islands (B. Genton, unpublished data). This demonstrates that standardized and comparable data can be produced at an affordable price. The flexibility of the system facilitates prompt inclusion of newly identified point mutations associated with parasite resistance. However, it needs to be emphasized that this technique allows only the analysis of already validated SNP, as does any other technology. The development of compatible assays for the detection of gene duplication or amplification on microarrays is necessary, as it has been shown that amplification of pfmdr might play an important role in modulating resistance against chloroquine and probably also against artemisinin derivatives (6, 21).

In conclusion, this method offers an unmatched capacity to provide evidence-based data on the dynamics of parasite resistance to antimalarial drugs in a cost-effective way. This platform also can be widely adapted and applied with ease to other genotyping tasks requiring highly parallel multiple-SNP analyses.

ACKNOWLEDGMENTS

We acknowledge the technical support of Lucy Kind (FHBB) and Marie Ballif (STI).

This project received financial support from the Roche Research Foundation, Freie Akademische Gesellschaft Basel, the European Community project ResMalChip (grant number QLK2-CT-2002-01503), and the Swiss National Foundation (grant number 3100-A0-103968).

None of the funding agencies were involved in study design; collection, analysis, and interpretation of data; writing of the paper; and the decision to submit it for publication.

REFERENCES


Downloaded from http://jcm.asm.org/ on December 29, 2017 by guest