DOI: 10.1002/cmdc.200900078

Discovery of Plasmepsin Inhibitors by Fragment-Based Docking and Consensus Scoring

Ran Friedman and Amedeo Caflisch*[a]

Plasmepsins (PMs) are essential proteases of the plasmodia parasites and are therefore promising targets for developing drugs against malaria. We have discovered six inhibitors of PM II by high-throughput fragment-based docking of a diversity set of ~40 000 molecules, and consensus scoring with force field energy functions. Using the common scaffold of the three most active inhibitors (IC_{50} = 2–5 μM), another seven inhibitors were identified by substructure search. Furthermore, these 13 inhibitors belong to at least three different classes of compounds. The in silico approach was very effective since a total of 13 active compounds were discovered by testing only 59 molecules in an enzymatic assay. This hit rate is about one to two orders of magnitude higher than those reported for medium- and high-throughput screening techniques in vitro. Interestingly, one of the inhibitors identified by docking was halofantrine, an antimalarial drug of unknown mechanism. Explicit water molecular dynamics simulations were used to discriminate between two putative binding modes of halofantrine in PM II.

Introduction

There is an urgent need to develop potent antimalarial agents due to emerging drug resistance.[1] Many efforts are aimed at inhibition of plasmodial proteases, which are involved in metabolism and host cell invasion.[2] Plasmepsins (PMs) are pepstatin-like aspartic proteases unique to plasmodia. Ten PMs have been identified in the genome of *P. falciparum*, the plasmodium species that causes the most fatal form of malaria in humans, and four of these ten are located in the food vacuole, an acidic lysosome-like organelle in which hemoglobin degradation takes place. Inhibitors of PMs are fatal to the parasite,[3] which indicates that PMs are relevant drug targets (see Reference [3] for a comprehensive list of known PM II inhibitors). Furthermore, several small-molecule inhibitors of retroviral and human aspartic proteases, namely HIV protease[4] and renin,[5] are effective and safe agents, which provides additional support to the relevance of PMs as drug targets. Several computational studies of PMs have been reported in the past three years. The computational approaches used for PM inhibitor development range from modeling of compounds in the binding site[6,7] to homology modeling of PM, docking, and molecular dynamics (MD) simulations.[8,9] On the other hand, high-throughput docking into the active site of PMs followed by experimental validation has yet to be described in the literature.

Herein, we report the discovery of novel PM inhibitors by high-throughput fragment-based docking of 40 000 molecules obtained by 2D structural clustering of an initial library of 4.6 million compounds. A similar in silico fragment-based approach had been used previously to identify novel inhibitors of the aspartic protease β-secretase,[10,11] two kinases,[12,13] and a flaviviral serine protease.[14] The major difference between our previous in silico screening campaigns and the one reported here is the final scoring of the poses. In fact, poses were sorted in the past by a single energy function, while consensus scoring, precisely the median rank of four force field based energy functions, is used here. The four energy functions are: the binding free energy approximated by the linear interaction energy with continuum electrostatics (LIECE) method,[15] the CHARMm (Accelrys, Inc., USA)[16] electrostatic interaction energy, the CHARMm van der Waals efficiency, and the TAFF[17] interaction energy. Only 59 molecules were tested in an enzymatic assay, and 13 of them displayed inhibitory activity (IC_{50} < 100 μM). Furthermore, four of these 13 compounds have an IC_{50} value between 2 and 5 μM. Remarkably, one of these four inhibitors is the antimalarial drug halofantrine, whose mechanism of action is still not clear although halofantrine was discovered four decades ago and has been on the market since 1988. The binding mode of halofantrine in PM II was further studied by multiple explicit water MD simulations.

Results and Discussion

Validation of the LIECE model for PM II

The three-parameter LIECE model for PM II was fitted to 20 peptidic inhibitors (see Experimental Section). The values of the LIECE parameters (see Equation 1) are $\alpha = 0.20 \pm 0.04$, $\beta = 0.05 \pm 0.01$ and $\Delta G_{tr,rot} = 6.0 \pm 2.5$ kcal mol^{-1}, where the standard deviation is calculated using the 20 LIECE models obtained by the leave-one-out (LOO) procedure. The free energy of binding calculated by LIECE has a root mean square error of 1.0 kcal mol^{-1} for the 20 known inhibitors used for fitting, with

[a] *Dr. R. Friedman, Prof. A. Caflisch*
Department of Biochemistry, University of Zürich
Winterthurerstrasse 190, CH-8057 Zürich (Switzerland)
Fax: (+41) 44 635 68 62
E-mail: caflisch@bioc.uzh.ch

a maximal unsigned error of 1.7 kcal mol^{-1} (Table 1). A cross-validated correlation coefficient q^2 of 0.60 is obtained by the LOO procedure. To provide further evidence against chance correlation in the LIECE model, a scrambling test of the data points was performed. The experimental values of binding affinity were randomly permuted 10000 times resulting in 10000 random data sets. In this way, 10000 random models were generated by fitting to the random data sets. A scatter plot of the correlation coefficient values versus the LOO cross-validated q^2 values for the random models and the true LIECE model is shown in Figure 1. The statistically most significant model (highest LOO cross-validated q^2) is the LIECE model fitted to the original (i.e., unpermuted) data points (triangle in Figure 1), which provides strong evidence that the LIECE model of PM II does not suffer of chance correlation.

Table 1. Calculated and experimental binding free energies of the 20 known PM II inhibitors used for fitting of the three parameter LIECE model.[a]

Inhibitor[b]	ΔE_{vdW}	ΔE_{Coul}	ΔG_{solv}	ΔG_{LIECE}	ΔG_{exp}[c]	Error
4	−65.5	11.0	−131.7	−13.0	−14.5	1.5
5	−58.2	13.3	−130.2	−11.3	−12.1	0.8
6	−69.5	−221.7	117.5	−13.0	−14.2	1.2
9	−60.4	−223.7	110.1	−11.6	−12.6	1.0
10	−63.5	−199.5	77.3	−12.6	−11.4	−1.3
11	−63.3	−214.1	100.8	−12.2	−11.8	−0.4
12	−53.0	−194.8	84.6	−9.9	−10.9	1.0
13	−69.8	−213.0	104.4	−13.3	−12.2	−1.0
14	−51.0	−213.7	79.9	−10.6	−10.0	−0.7
20	−53.3	−489.9	406.2	−8.7	−7.5	−1.3
21	−54.9	−484.9	397.7	−9.2	−7.5	−1.7
22	−51.9	−493.3	412.3	−8.3	−9.5	1.1
23	−52.5	−493.4	411.8	−8.5	−9.4	0.9
24	−59.1	−492.8	425.3	−9.2	−8.5	−0.6
25	−60.8	−497.8	425.9	−9.7	−10.3	0.6
26	−64.0	−500.3	427.5	−10.4	−9.8	−0.6
27	−58.3	−447.7	385.4	−8.7	−10.0	1.3
29	−57.5	−453.8	381.2	−9.1	−8.4	−0.6
30	−59.2	−202.0	98.1	−10.9	−10.6	−0.3
31	−60.9	−200.5	96.2	−11.3	−10.3	−1.0

[a] The model is $\Delta G_{LIECE} = 0.20\,\Delta E_{vdW} + 0.05\,(\Delta E_{Coul} + \Delta G_{solv}) + 6.0$ kcal mol^{-1}. [b] Numbering of inhibitors is as in Reference [3]. Note that the numbers of the known inhibitors are not in boldface to distinguish them from those discovered in this study. [c] Experimental values of binding free energies are derived from K_i values listed in Reference [3], using $\Delta G = -RT\ln(K_i)$. All energy values are in kcal mol^{-1}.

Validation of consensus scoring

To check if consensus scoring is able to identify known inhibitors, the median rank was calculated for the 20 peptidic inhibitors used for the LIECE model fitting. Ten of the 20 inhibitors have the top-ten median ranks among the 40000 compounds. Of the remaining ten inhibitors, eight have a median rank below 200, while inhibitors **4** and **5** have median ranks of 12902 and 25457, respectively. The 18 inhibitors in the first 200 ranks score well in three of the four energy functions and poorly in the van der Waals efficiency, mainly because of their size. The molecular weight of the 20 peptidic inhibitors is 627 ± 78 Da, compared with 333 ± 99 Da for the 40000 docked compounds. Furthermore, inhibitors **4** and **5** have very unfavorable CHARMm electrostatic interaction energy as they have a negative charge (carboxyl group) and the net charge of PM II is −17 electronic units. The good ranks of the known inhibitors indicate that consensus scoring is appropriate for high-throughput docking.

The 40000 docked compounds (which include six active molecules) were ranked together with the 20 peptidic inhibitors and the seven inhibitors identified by substructure search to compare the predictive power of individual energy functions and consensus scoring. The receiver operating characteristic (ROC) plot (Figure 2) shows that consensus scoring yields the highest enrichment factor when considering the first 1000 compounds only, while the TAFF energy function is the best when a larger set of compounds is considered. One should note that only the top-ranking 500 compounds were examined visually in this work (see below), which is a posteriori justification for the use of consensus scoring. Moreover, the LIECE model and the electrostatic interaction energy are slightly worse than consensus scoring, while the van der Waals efficiency is close to random selection, probably because

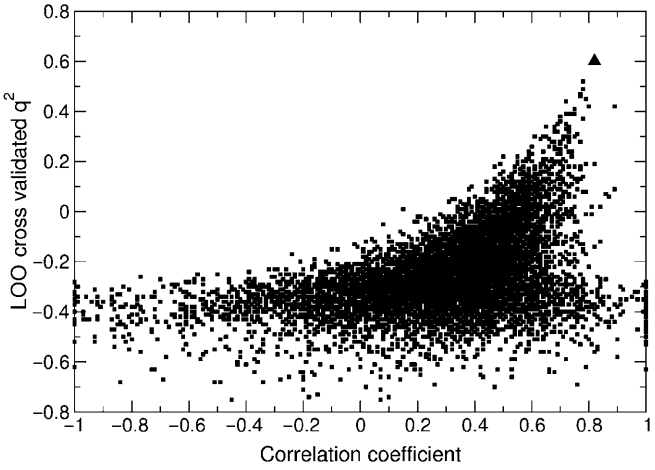


Figure 1. The LIECE model does not suffer of chance correlation. A total of 10000 random permutations of the ΔG_{exp} values were performed for the data presented in Table 1. For each random permutation the correlation coefficient obtained by fitting and the LOO cross-validated q^2 were calculated (circles). The model with the best fit to the data is the true LIECE model (triangle). Note that a few random models with correlation coefficient higher than 0.9 are present, but they have a poor LOO cross-validated q^2 because they fit the data by chance.

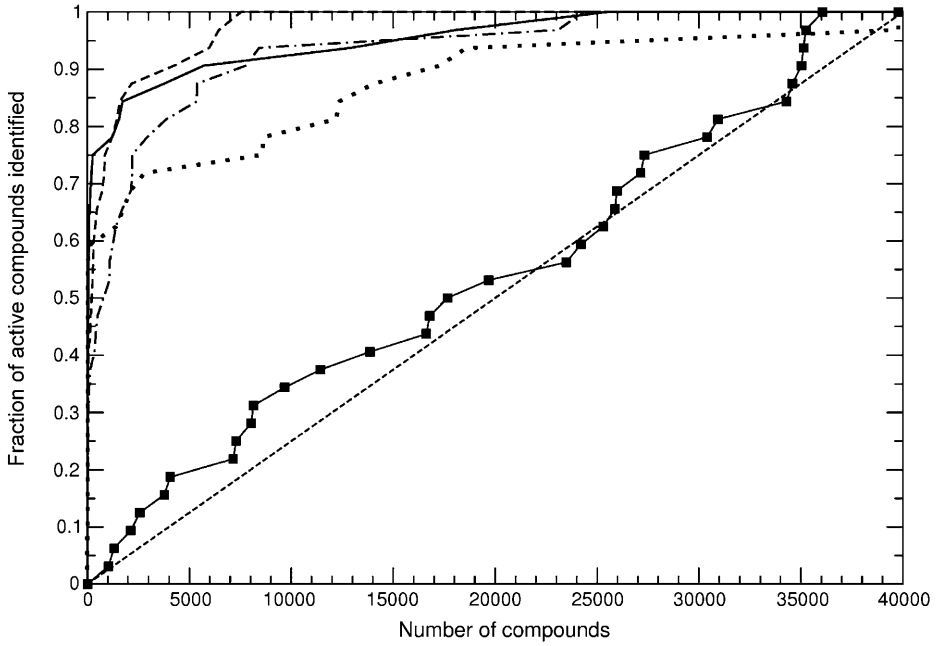


Figure 2. ROC plot of the individual energy terms and consensus scoring used in this study. A total of 33 active compounds, 20 peptidic inhibitors and the 13 compounds identified in this study, were ranked together with the ~40 000 docked compounds. Consensus scoring (——) is the preferred method if one examines up to approximately the first 1000 compounds. LIECE, –•–•; Electrostatic interaction energy, •••••; van der Waals efficiency, ——■——; TAFF, – – –; Consensus scoring, ——; Random model, -----.

of the larger molecular weight of the peptidic inhibitors than most of the compounds in the library.

Inhibitor discovery by fragment-based docking and consensus scoring

Upon sorting the compounds according to median rank and visual inspection of the poses of the top 500 compounds, 19 of them were tested in a fluorogenic assay[18] with the enzyme in solution. Six of these 19 compounds are active at concentrations below 100 μM against PM I, PM II, and PM IV (compounds **1**–**6** in Table 2; Figure 3). Therefore, the in silico screening approach is much more effective than medium- and high-throughput screening procedures in vitro, for which hit rates varying between 0.1% and 5% have been reported.[19]

Three of the six active compounds have an IC_{50} value ≤ 5 μM for PM I and PM II. Notably, these three compounds were identified in the database of the National Cancer Institute (USA), i.e., they do not originate from a malaria-related collection of compounds. Five of the six inhibitors have a basic amino group that, according to the predicted binding modes, can be either partially solvated or involved in electrostatic interactions with the catalytic dyad. The binding mode of compound **1** is shown in Figure 4. The positively charged tertiary amino group participates in electrostatic interactions with the two carboxy groups of the Asp side chains in the catalytic dyad. Furthermore, the hydroxy group of compound **1** is involved in a hydrogen bond with the Ser 79 side chain located at the tip of the flap that covers the central part of the substrate-binding site. Compound **6** has a primary amino group that also forms a salt bridge with Asp 34 in the most favorable pose. On the other hand, the partial solvation of the bulky tertiary amino group of compounds **2**–**4** and lack of direct electrostatic interactions with the catalytic aspartates might be due to the steric requirements of the former or the whole compound.

Remarkably, compound **3** is the antimalarial drug halofantrine,[20] whose mechanism of action has not been fully elucidated although it was discovered more than 40 years ago. It has been suggested that halofantrine reacts with hematin, a metabolite of hemoglobin degradation, which is toxic to plasmodia. Indeed, halofantrine covalently binds to hematin in vitro.[21] Furthermore, quinine, an antimalarial drug with some similarity to halofantrine, is capable of inhibiting the dimerization of hematin to the nontoxic dimer

Table 2. Experimental validation of the in silico screening.[a]

Compound	IC_{50} [μM][b]			ΔG_{LIECE}[c]
	PM I	PM II	PM III	[kcal mol^{-1}]
Pepstatin	0.001	0.0002	0.002	
Inhibitors identified by high throughput docking				
1	4	5	21	−9.1
2	22	20	>100	−9.4
3	3	2	6	−9.0
4	2	2	5	−8.4
5	22	18	13	−9.9
6	37	56	52	−8.7
Inhibitors identified by substructure search				
7	23	17	92	−8.1
8	3	5	9	−8.9
9	23	19	65	−7.7
10	24	95	100	−16.6
11	71	32	>100	−6.2
12	28	62	39	−7.6
13	41	24	80	−6.3

[a] The IC_{50} values (IC_{50} = concentration of inhibitor at which 50% of the maximum initial rate is observed) were determined for the 13 compounds shown in Figure 3. [b] The IC_{50} values represent the average of two repetitions of a fluorogenic proteolysis assay.[18] The estimated error for this enzymatic assay is ±50%. The values measured for pepstatin using the same assay in parallel are given as a reference. [c] The free energy of binding to PM II was calculated according to the LIECE model (see Equation (1) and caption of Table 1). Note that ΔG_{LIECE} systematically overestimates the affinity of these inhibitors except for compounds **11** and **13**.

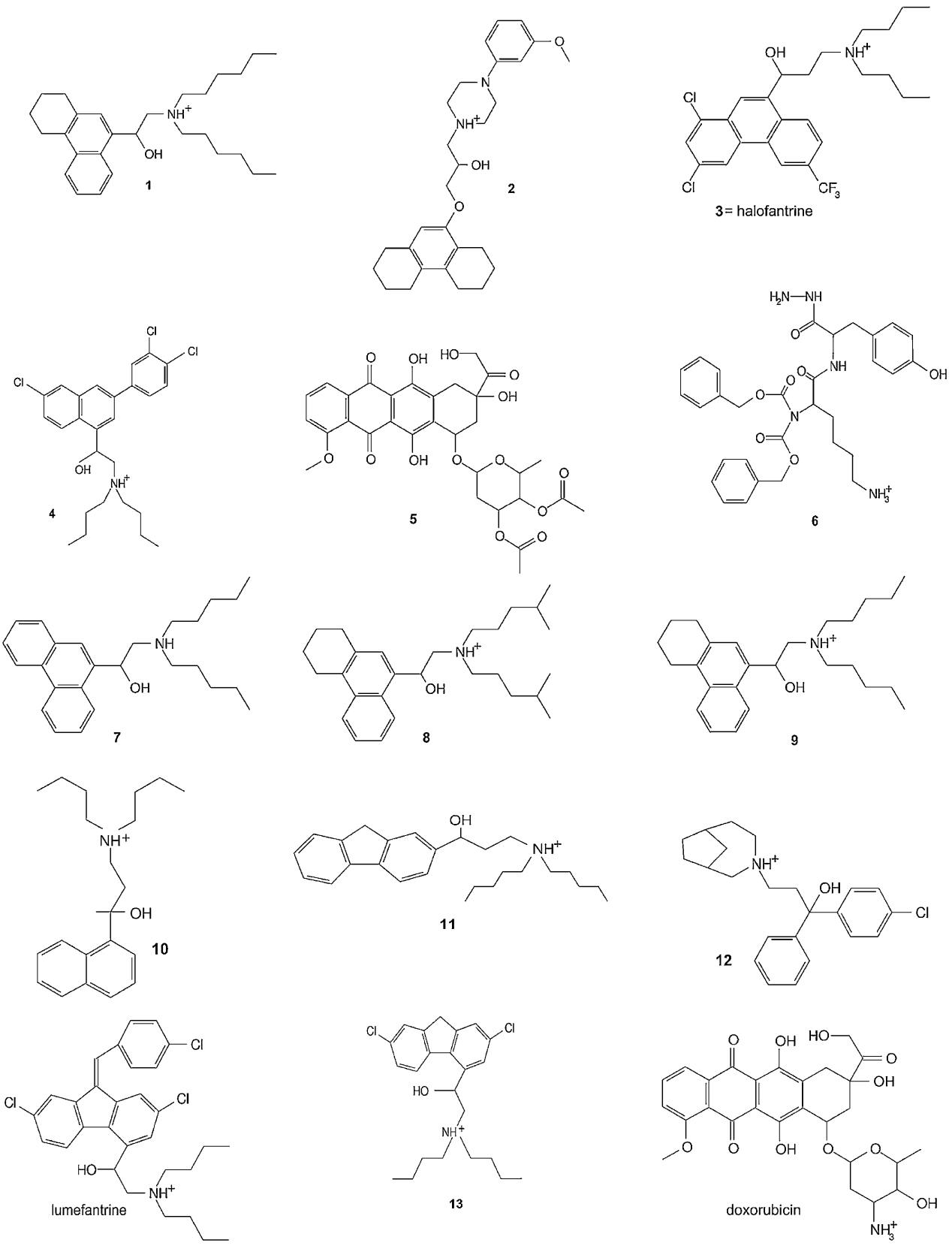
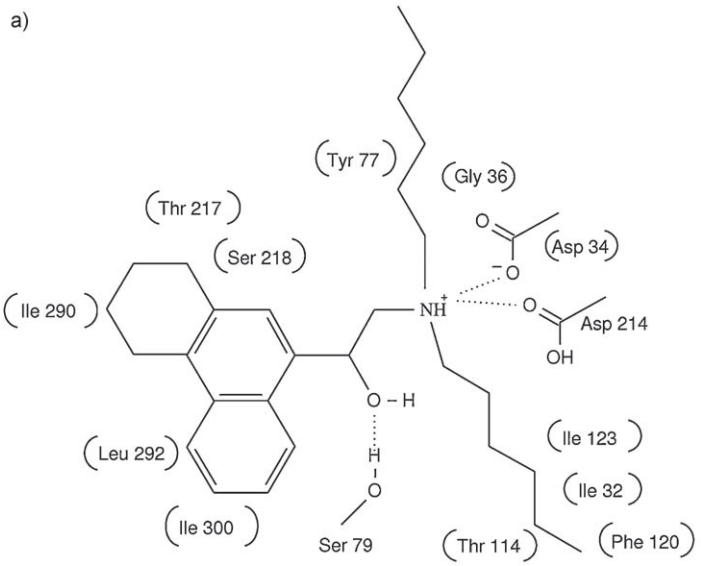
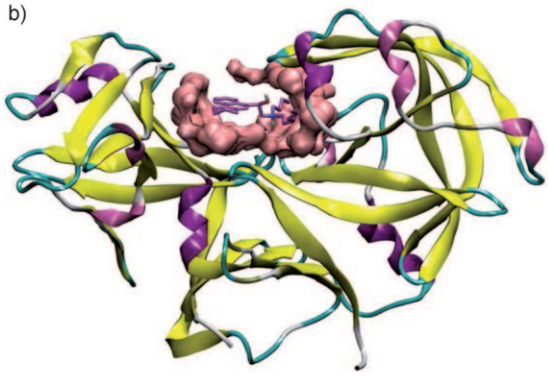


Figure 3. The 13 PM inhibitors discovered in this study. The structures of lumefantrine and doxorubicin are also shown as they share the same scaffold with compounds **13** and **5**, respectively.





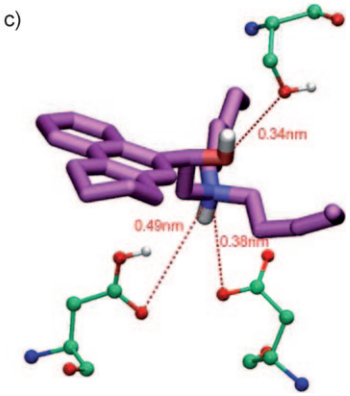


Figure 4. The binding mode of compound **1** obtained by docking. a) A diagram of the binding mode showing electrostatic interactions with Asp 214 and Asp 34, a weak hydrogen bond with Ser 79 of the flap and van der Waals interactions with 12 residues (shown in parentheses). b) The binding pocket and its location in the protein, and c) the interactions of the inhibitor with the catalytic aspartates and Ser 79.

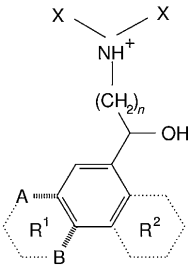


Figure 5. The common scaffold of compounds **1**, **3**, and **4** (see Figure 3). The substituent X stands for an alkyl chain, *n* can be 1 or 2, R^1 and R^2 are aliphatic or aromatic hydrocarbon rings. Groups A and B are adducts that replace R^1, as in compound **4**.

β-hematin by plasmodia.[22] However, stereoisomers of quinine did not significantly inhibit β-hematin formation but showed antiplasmodial activity at μM concentrations,[22] suggesting that a different mechanism of action is involved. Therefore, the docking results imply that halofantrine acts as PM inhibitor in addition to its interference of hematin dimerization.

Substructure search and structure–activity relationships

Compounds **1**, **3** and **4**, with IC_{50} values ≤ 5 μM, share a common substructure (Figure 5), which was used to search in the Zinc library for further derivatives. In total 40 molecules were retrieved and tested in the enzymatic assay and seven of them are active against PM II (compounds **7**–**13** in Table 2; Figure 3). Compounds **7**–**10** are similar to compounds **1**, **3**, and **4** but the affinity for PM II of compound **10** is between 5 and 50 times poorer than the affinity of compounds **1**, **3**, **4** and **7**–**9**. These values suggest that molecules with a three-ring system are more potent than those with two rings. Furthermore, the longer the alkyl chain, the higher the affinity, which can be explained by additional hydrophobic contacts with the protein (Figure 4). On the other hand, halogen substituents on the ring system (Cl and CF_3, compounds **3** and **4**) favor binding even if the alkyl chain consists of only four carbon atoms. Instead of three 6-membered rings, compounds **11** and **13** have a 5-membered ring fused to two 6-membered rings. Interestingly, a similar substructure is shared between these compounds and an antimalarial drug with an unknown mode of action, lumefantrine (benflumetol).[23] Lumefantrine was docked into PM II with SEED and FFLD, and the LIECE binding free energy was calculated as -8.2 kcal mol^{-1}. Visual inspection of the 2D structure of the 13 inhibitors of PM II indicates that they can be clustered into five different classes of compounds, three of which contain a single member, namely compounds **5**, **6**, and **12** (Figure 3). Grouping of the 13 inhibitors by 2D structural similarity using DAIM and a Tanimoto threshold of 0.950 and 0.975 yields 3 and 6 clusters, respectively. The similarity coefficients, as calculated using the DAIM fingerprints, are shown in Figure 6.

MD simulation study of the putative binding modes of halofantrine

Explicit water MD simulations were carried out to gain more information on the binding of halofantrine (**3**) in the catalytic site of PM II. Two putative binding modes were studied (Figure 7). In one binding mode (Amino_In) the amino group of halofantrine is located near Asp 34, and the hydroxy moiety is hydrogen bonded to Asp 214. It was generated manually according to the binding of the amino group and alkyl chains in

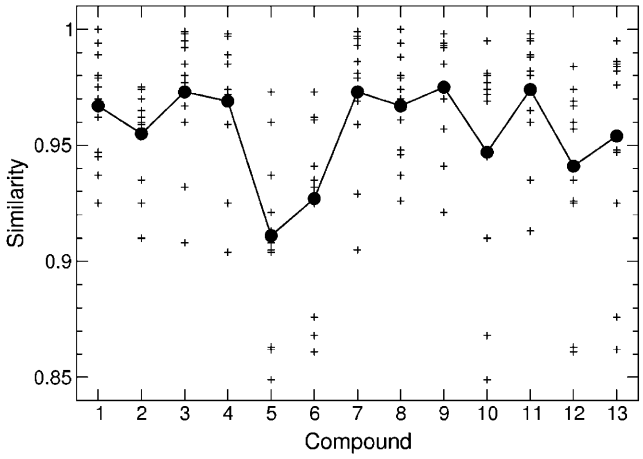


Figure 6. Similarity between active compounds. Similarity coefficients were calculated with DAIM, based on the DAIM fingerprints.[24] For each compound the average similarity coefficient, i.e., the similarity value averaged over the 12 other compounds, is shown with a filled circle (●) and the circles are connected by a solid line to guide the eye. Individual data points are shown by crosses.

compound **1**. The other binding mode (Amino_Out), in which the amino group is partially exposed to the bulk and is not in contact with the catalytic dyad, was generated automatically by high-throughput docking, as detailed in the Experimental Section. The protein appears to be stable in both MD runs, as indicated by the low Cα root mean square deviation (RMSD) and the small number of different structural clusters (Table 3).

The distance between the catalytic dyad oxygen atoms (Asp 34 Oδ–Asp 214 Oδ) is larger in the presence of halofantrine than in the complex with the EH58 inhibitor (PDB: 1LF3), and more so in the Amino_In conformation (Figure 8), in which the amino moiety of halofantrine is hydrogen bonded to Asp 34. In the Amino_Out conformation, it can be hydrogen bonded to the hydroxy group of Tyr 192 or the carbonyl of Gly 36, but it is more often solvent exposed (Table 4). In both binding modes, the hydroxy of Asp 214 acts as the donor in a hydrogen bond with the inhibitor hydroxy group. The mobility of halofantrine within the binding site can be assessed by calculating its RMSD after fitting of the protein atoms (Figure 9, top) and comparing with the intrinsic flexibility of the inhibitor (Figure 9, bottom). The time series of RMSD reveal a significant difference in the stability of the two putative binding modes. In the Amino_In conformation, the RMSD of the inhibitor reaches values of 0.7–0.8 nm upon overlapping the protein backbone. The high RMSD values indicate that in the Amino_In binding mode halofantrine changes its orientation with respect to the tip of the flap. Such displacement is not observed in the run started from the Amino_Out conformation. Finally, in one of the two Amino_In simulations, the hydrogen bond between the amino group of the inhibitor and Asp 34 is irreversibly lost after 43 ns, leading to partial unbinding. Taken together, the MD simulations indicate that the binding mode Amino_Out (generated by automatic docking) is more stable than Amino_In (obtained by manual docking) although it is not possible to definitively discard either.

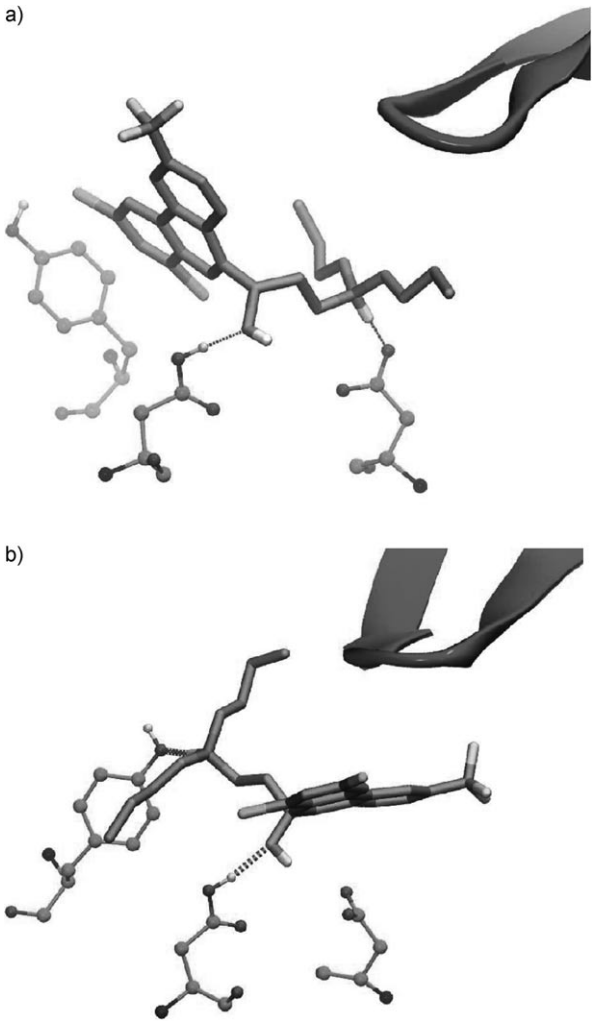


Figure 7. The two putative binding modes of compound **3**. a) Amino_In: The nitrogen of the tertiary amino group in halofantrine is located near the charged Asp 34, while the hydroxy group is hydrogen bonded to Asp 214. The Tyr 192 side chain does not interact with halofantrine. The flap is shown as a ribbon. The snapshot was taken at 3.6 ns. b) Amino_Out: The amino nitrogen of halofantrine is hydrogen bonded to Tyr 192 and the hydroxy to Asp 214. The snapshot was taken at 30 ns.

Conclusions

To identify inhibitors of PM II, 4.6 million compounds were clustered according to 2D structural similarity resulting in

Table 3. Summary of the explicit water MD simulations[a] of the complex of PM II with halofantrine.

Starting structure	Simulations	Cα RMSD[b] [nm]	Protein clusters	Halofantrine clusters[c]
Amino_Out	2	0.15	8	4
Amino_In	2	0.15	8	2

[a] Each simulation was run for 50 ns. [b] Average value of the Cα root mean square deviation between pairs of MD conformations. [c] Structural clustering was performed with the Jarvis-Patrick algorithm[58] as described in Reference [56].

Table 4. Hydrogen bonds between the protein and halofantrine (HF).[a]

Donor	Acceptor	Amino_In	Amino_Out
HF–NH	Asp 34	92	0
HF–NH	Gly 36 or Tyr 192	0	24
HF–OH	Asp 34	0	68
Asp 214	HF-OH	94	91

[a] The percentage of simulation frames in which a certain hydrogen bond (including hydrogen bonds mediated by a single water molecule) is present is given for the simulations Amino_In and Amino_Out.

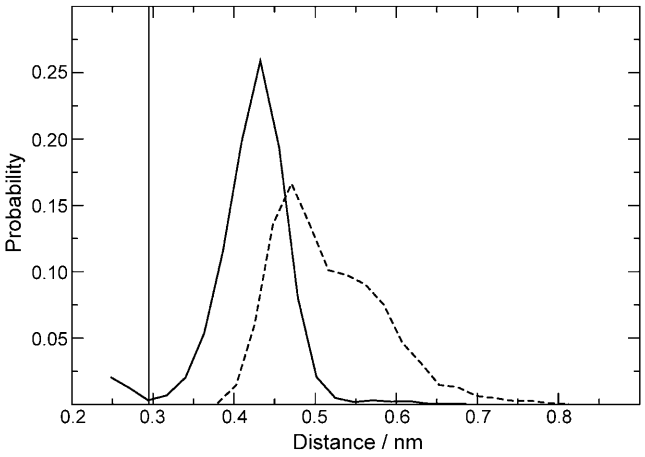


Figure 8. Probability distribution of the minimal distance between the oxygens of the catalytic dyads; Amino_In, -----; Amino_Out, ——. The vertical line is the value of the distance in the crystal structure 1LF3 ($d = 0.295$ nm).

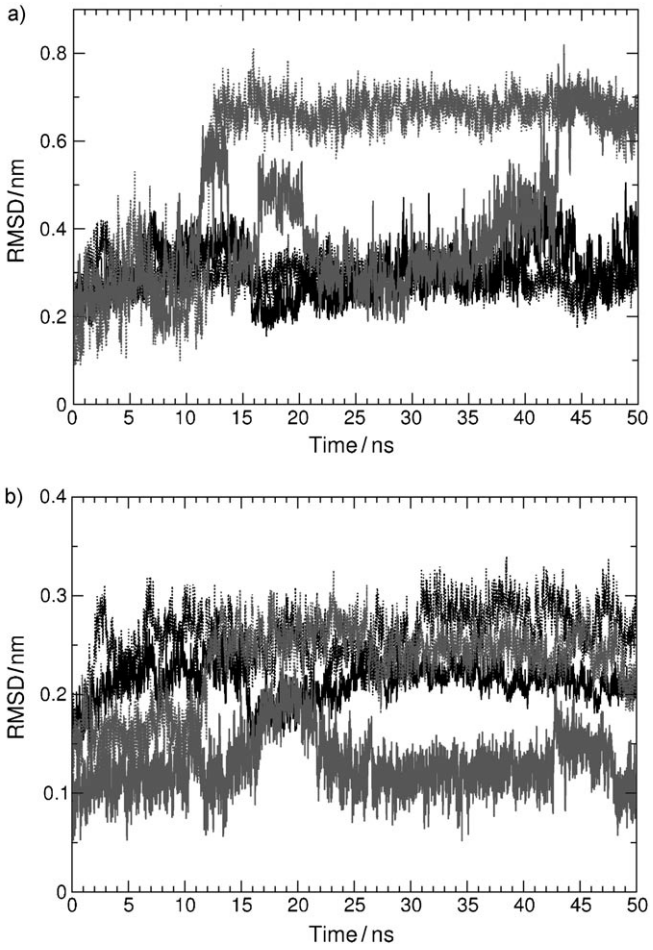


Figure 9. Time series of RMSD of halofantrine during explicit water MD simulations. Heavy atoms RMSD are calculated after fitting the Cα atoms of a) PM II (displacement from active site) or b) the atoms of halofantrine (intrinsic flexibility of halofantrine). The four lines correspond to the two MD runs started from each the Amino_In (grey) and the Amino_Out (black) binding modes.

~40 000 molecules, which were then used for fragment-based docking. Docking into the PM II active site was followed by consensus scoring using four force field based energy functions. A total of 19 compounds were tested in an enzymatic assay, and three of them showed single-digit micromolar inhibitory activity. One of these three inhibitors was halofantrine, an antimalarial drug discovered more than 40 years ago whose mechanism of action has not yet been elucidated. Four 50 ns explicit water MD simulations were performed to assess the relative stability of two putative binding modes of halofantrine—with the tertiary amino group of halofantrine involved in electrostatic interactions with the catalytic dyad or pointing towards the solvent—but the analysis of the MD trajectories did not provide definitive evidence for either of the binding modes. A substructure search performed using the common scaffold of the three inhibitors as a query yielded another seven active compounds.

In conclusion, the fragment-based docking approach identified 13 inhibitors of medium to low micromolar potency against PM II. These 13 compounds belong to at least three different classes of compounds, and were identified through experimental testing of only 59 molecules. This success rate is much higher than the 0.1–5% hit rates reported in medium- to high-throughput screening campaigns in vitro. The in silico identification of halofantrine and a close analogue of lumefantrine as PM II inhibitors provides further evidence that PMs are interesting targets for the development of small-molecule antimalarial drugs.

Experimental Section

A full list of abbreviations can be found at the end of the Experimental Section.

Computational methods

The in silico screening was carried out by a fragment-based procedure for high-throughput docking.[24–26] The poses were scored by the median value of the ranks of four energy functions. Three of these functions consisted of intermolecular energy contributions, and the fourth was the binding free energy evaluated by the LIECE approach.[15]

Preparation of the PM II structure

The structure of PM II in complex with the hydroxymethylamine inhibitor EH58 (PDB: 1LF3,[27]) was downloaded from the protein

data bank.[28] This structure was chosen because of its slightly larger substrate-binding site, in particular at the S_1' subpocket, which has to accommodate the large P_1' group of EH58. In fact, EH58 cannot be docked into structures with a smaller S_1' subpocket, e.g., the complex of PM II with the protease inhibitor pepstatin (PDB: 1LS5[27]). The catalytic dyad was modeled with protonated Asp 214 and negatively charged Asp 34, according to a previous MD study.[29] All Arg and Lys side chains were positively charged, while Glu and Asp side chains (except Asp 214) were negatively charged. The inhibitor and all water molecules were removed. Hydrogen positions were optimized by energy minimization with the computer program CHARMM,[30,31] using 100 steps of steepest descent and 10 000 steps of conjugate gradient minimization. The distance-dependent dielectric function $\varepsilon(r) = r$ was used in the calculation of the electrostatic energy where r is the interatomic distance in Å. Nonbonding interactions were truncated with a switching function at 14 Å.

Preparation of the library of compounds

First, redundant entries were removed from the Zinc library (version 5),[32] which contains ~4.6 million molecules. Then the compounds were clustered according to their 2D structural features by the leader algorithm[33] as implemented in DAIM[24] and using a Tanimoto coefficient of 0.995 as the similarity threshold (note that a high threshold value was used to weed out only the compounds that were very similar to each other as the docking approach employed in this study is very efficient). The compound with the highest similarity to all other molecules in the same cluster was used for docking into the binding site. Each representative compound was automatically decomposed into molecular fragments by DAIM, which was also used to select the three fragments for docking by FFLD.[26,34] This clustering yielded ~40 000 representative molecules upon removal of small molecules that could not be decomposed into at least three fragments by DAIM. All input fragments were minimized with CHARMM prior to docking, by 2000 steps of steepest descents and 10 000 steps of conjugate gradients. The distance dependent dielectric function $\varepsilon(r) = 4r$ was used in the calculation of electrostatic interactions during minimization.

High-throughput fragment-based docking

High-throughput fragment-based docking was carried out using the methods described in Reference [11]. Briefly, fragments were docked to the binding site using SEED,[25,35] which takes into account electrostatic solvation effects by a generalized Born approach based on numerical evaluation of Born radii.[36] For each of the 40 000 compounds, the most favorable poses of three of its fragments were used as "anchors" by the program FFLD[26,34] for automatic docking into the binding site. The binding modes were clustered by using the leader algorithm with a similarity cutoff of 0.6.[25,37] A maximal number of 30 different poses per molecule was allowed. The total number of docked compounds and poses was approximately 40 000 and 750 000, respectively. Each pose was minimized within the rigid protein using CHARMM by 500 steps of steepest descent and 10 000 steps of conjugate gradient using the dielectric function $\varepsilon(r) = 4r$.

Consensus scoring

The following force field based energy functions were used for consensus scoring:

1) The binding free energy evaluated by a three-parameter LIECE model (see next section).
2) The CHARMm[16] protein–compound electrostatic interaction energy with $\varepsilon(r) = 4r$ and cutoff at 14 Å.
3) The CHARMm protein–compound van der Waals interaction energy (cutoff at 14 Å) divided by the molecular weight of the compound (van der Waals efficiency).
4) the TAFF[17] protein–compound interaction energy.

Electrostatic solvation energies (for LIECE) were calculated by numerical solution of the finite difference Poisson equation using the PBEQ module[38] in CHARMM with an initial grid size of 1.0 Å, and focusing to a final grid size of 0.4 Å. CHARMm and TAFF interaction energies were calculated after ligand minimization, using the corresponding force field, in the rigid structure of PM II. Note that the TAFF force field differs significantly from CHARMm as it was developed to reproduce molecular structures by use of bonded and van der Waals interactions only, i.e., neglecting electrostatic interactions.

Consensus scoring is generally preferable to the use of a single scoring function.[39] It has been reported that the median rank is more suitable than the average rank in consensus scoring because the former is less sensitive to outliers.[40] Here, a median rank for each pose was calculated after ranking all poses independently by each of the four energy functions. As an example, consider a certain pose that is ranked 14th, 24th, 26th, and 27 000th by the four functions: Its median rank is 25 (while its average rank of 6766 is clearly affected by the very poor ranking of only one of the four functions). The 750 000 poses were sorted according to their median rank and the top-scoring poses were examined visually. Since the protein structure was kept rigid during docking it is very difficult to devise rules for the automatic evaluation of the significance of the poses. It was not possible, for example, to use the number of hydrogen bonds between the compound and the protein as a filter since protein relaxation might promote formation of additional intermolecular hydrogen bonds. Moreover, it is not simple to treat water molecules at the protein surface, some of which can be involved in intermolecular hydrogen bonds.[18] Visual examination was used to discard high-scoring compounds, which nevertheless did not seem to be properly docked. For example, a few compounds were discarded because they did not form hydrogen bonds with the catalytic dyad, or because they had a charged group located in a mainly hydrophobic subpocket.

LIECE model for PM II

LIECE uses only single-point energy evaluation on minimized structures[15] and is therefore an efficient approximation of the original linear interaction energy method,[41] which requires computationally more expensive MD simulations.[42] Here, the free energy of binding is approximated by a three-parameter LIECE model

$$\Delta G_{\text{bind}} = \alpha \Delta E_{\text{vdW}} + \beta(\Delta E_{\text{Coul}} + \Delta G_{\text{solv}}) + \Delta G_{\text{tr,rot}} \quad (1)$$

where ΔE_{vdW} is the protein–ligand van der Waals interaction energy, ΔE_{Coul} is the Coulombic interaction energy calculated in vacuo, and ΔG_{solv} is the electrostatic contribution to the free energy of solvation. The parameters α and β are dimensionless, while $\Delta G_{\text{tr,rot}}$ accounts for the loss of translational and rotational degrees of freedom upon binding and is assumed to be independent of the bound ligand. The three parameters α, β, and $\Delta G_{\text{tr,rot}}$ are determined by calculating ΔE_{vdW}, ΔE_{Coul}, and ΔG_{solv} for a series of compounds of known binding affinities and fitting according to Equa-

tion (1). A set of 20 known peptidic inhibitors of PM II was used for fitting, i.e., to derive the LIECE model for PM II. The 20 inhibitors (numbered as in Reference [3]) are nine statines (compounds 4–6 and 9–14), nine hydroxyethylamines (compounds 20–27 and 29), and two hydroxypropylamines (compounds 30 and 31). The coordinates of compounds 4, 11, 30, and 31 were taken from crystal structures of bound PM II (PDB: 1XE5, 1W6H, 1LEE, and 1LF2,[43] respectively). The coordinates of these structures were superimposed on that of the protein with EH58 bound (PDB: 1LF3), as used for docking. The coordinates of compound 4 were modified manually to account for the binding of compounds 5 and 6. Similarly, models of the complexes of PM II with compounds 9–12, 13 and 14 were generated using compound 11 as a template; the binding mode of the hydroxymethylamine inhibitor EH58 (PDB: 1LF3,[27]) was used as a template for all other compounds where crystal structures were not available. The coordinates of the inhibitors were refined by 100 steps of steepest descents and 10000 steps of conjugate gradients energy minimization. A distance-dependent dielectric function $\varepsilon(r) = 4r$ was used in the calculation of electrostatic interactions during minimization. Upon minimization, ΔE_{Coul} was calculated with vacuum dielectric ($\varepsilon = 1$), without long-range truncation. The electrostatic solvation ΔG_{solv} was determined by solving the finite difference Poisson equation with $\varepsilon_{solvent} = 78.5$ and $\varepsilon_{solute} = 1$ using the PBEQ module[38] in CHARMM with an initial grid size of 1.0 Å and focusing to a final grid size of 0.3 Å.

MD simulations

Explicit-water MD simulations were carried out to discriminate between two halofantrine binding modes obtained by automatic (mode Amino_Out) and manual docking (mode Amino_In). Two 50 ns runs differing in the initial random assignment of atomic velocities were started from each binding mode (Table 3). The four MD simulations were performed using GROMACS (version 3.3.3),[44,45] with the OPLS all atom force field[46] and TIP4P model of water.[47] Parameters for halofantrine were derived as follows; first, its geometry was optimized by quantum mechanical ab initio approach with the 6-31Gd basis set, which is compatible with the force field.[46] The partial charges were then determined by fitting them to the electrostatic potential around the molecule using the Kollman-Singh method.[48] The calculation was performed with GAMESS.[49] The topology for halofantrine was generated automatically by the topology builder mktop (Federal University of Rio de Janeiro, Brazil), which assigns OPLS atom types (as used in GROMACS) according to the chemical moieties of the molecules, e.g., amide, ketone and alcohol.

The protein–halofantrine complex was immersed in a truncated octahedral box containing pre-equilibrated TIP4P waters.[47] Water molecules were removed from the box if the distance between any solute atom and any atom of the water was less than their sum of van der Waals radii. The edges of the box extended to at least 1.3 nm from the solute. Sodium and chloride ions were added randomly by replacing water molecules in order to neutralize the charge of the system and maintain a salt concentration of 0.1 M. Cation parameters derived by Åqvist[50] were used for Na^+. Parameters for Cl^- were taken from Reference [51]. Before each MD simulation, internal strain was relaxed by energy minimization until the maximal force on individual atoms was smaller than 100 kJ mol^{-1} nm^{-1}. After the minimization, a restrained MD run was performed for 100 ps. During the restrained simulations, protein heavy atoms were fixed to their initial positions with a force constant of 1000 kJ mol^{-1} nm^{-2}. The restraints were released, and the system was equilibrated for 1 ns before data was collected for analysis. During the MD runs, the LINCS algorithm[52] was used to constrain the lengths of bonds, while water molecules were kept rigid by use of the SETTLE algorithm.[53] The time step for the simulations was 2 fs. The temperature and pressure were coupled to an external bath with Berendsen's coupling algorithm[54] ($P_{ref} = 1$ bar, $\tau_P = 0.5$ ps; $T_{ref} = 300$ K; $\tau_T = 0.1$ ps). Van der Waals forces were truncated at 1.0 nm with a plain cutoff. Long-range electrostatic forces were treated using the particle mesh Ewald method.[55]

Structural analysis and clustering

Structural analysis was carried out by GROMACS utilities. Clustering was performed as detailed in Reference [56]. The figures with protein structures were generated by visual molecular dynamics (VMD).[57]

Hydrogen-bond analysis

Hydrogen bonds between two residues were calculated with the program g_hbond, which is available in GROMACS. The criteria for a hydrogen bond were a donor (nitrogen or oxygen) to acceptor (oxygen) distance $\leq$ 0.35 nm and an acceptor atom–donor hydrogen angle $\leq 30^\circ$, which are the default parameters in GROMACS. In addition, protein donor–acceptor pairs separated by a single water molecule were also taken into account as hydrogen bonds.

Enzymatic assay

A previously described fluorogenic assay[18] was used for all compounds tested in this work.

Abbreviations

CHARMM, chemistry at Harvard macromolecular mechanics; GAMESS, general atomic and molecular electronic structure system; LIECE, linear interaction energy with continuum electrostatics; MD, molecular dynamics; OPLS, optimized potentials for liquid simulations; PM, plasmepsin; ROC, receiver operating characteristic.

Acknowledgements

We thank Dr. Solange Meyer and Dr. Christoph Binkert (Actelion) for performing the enzymatic assays and very interesting discussions. We thank Dr. Danzhi Huang for critical reading of the manuscript. We acknowledge the Developmental Therapeutics Program of the National Institute of Health, USA for supplying us with some of the compounds used in this study. This work was partially supported by the Forschungskredit Program of the University of Zürich (R.F.)

Keywords: antiplasmodial agents · drug discovery · high-throughput docking · malaria · molecular dynamics

[1] T. J. Egan, C. H. Kaschula, *Curr. Opin. Infect. Dis.* **2007**, *20*, 598–604.
[2] P. J. Rosenthal, *Emerging Infect. Dis.* **1998**, *4*, 49–57.
[3] K. Ersmark, B. Samuelsson, A. Hallberg, *Med. Res. Rev.* **2006**, *26*, 626–666.
[4] S. M. Hammer, J. J. Eron, P. Reiss, R. T. Schooley, M. A. Thompson, S. Walmsley, P. Cahn, M. A. Fischl, J. M. Gatell, M. S. Hirsch, D. M. Jacobsen,

J. S. Montaner, D. D. Richman, P. G. Yeni, P. A. Volberding, *JAMA J. Am. Med. Assoc.* **2008**, *300*, 555–570.
[5] K. K. Gaddam, S. Oparil, *Curr. Opin. Nephrol. Hypertens.* **2008**, *17*, 484–490.
[6] F. Hof, A. Schütz, C. Fäh, S. Meyer, D. Bur, J. Liu, D. E. Goldberg, F. Diederich, *Angew. Chem.* **2006**, *118*, 2193–2196; *Angew. Chem. Int. Ed.* **2006**, 45, 2138–2141.
[7] M. Zürcher, T. Gottschalk, S. Meyer, D. Bur, F. Diederich, *ChemMedChem* **2008**, *3*, 237–240.
[8] S. Bjelic, M. Nervall, H. Gutierrez-de Teran, K. Ersmark, A. Hallberg, J. Åqvist, *Cell. Mol. Life Sci.* **2007**, *64*, 2285–2305.
[9] T. Luksch, N. S. Chan, S. Brass, C. A. Sotriffer, G. Klebe, W. E. Diederich, *ChemMedChem* **2008**, *3*, 1323–1336.
[10] D. Huang, U. Luthi, P. Kolb, K. Edler, M. Cecchini, S. Audetat, A. Barberis, A. Caflisch, *J. Med. Chem.* **2005**, *48*, 5108–5111.
[11] D. Huang, U. Luthi, P. Kolb, M. Cecchini, A. Barberis, A. Caflisch, *J. Am. Chem. Soc.* **2006**, *128*, 5436–5443.
[12] P. Kolb, D. Huang, F. Dey, A. Caflisch, *J. Med. Chem.* **2008**, *51*, 1179–1188.
[13] P. Kolb, C. Berset, D. Huang, A. Caflisch, *Proteins* **2008**, *73*, 11–18.
[14] D. Ekonomiuk, X.-C. Su, K. Ozawa, C. Bodenreider, S. P. Lim, Z. Yin, T. H. Keller, D. Beer, V. Patel, G. Otting, A. Caflisch, D. Huang, *PloS Neglected Trop. Dis.* **2009**, *3*, e356.
[15] D. Huang, A. Caflisch, *J. Med. Chem.* **2004**, *47*, 5791–5797.
[16] F. Momany, R. Rone, *J. Comput. Chem.* **1992**, *13*, 888–900.
[17] M. Clark, R. D. I. Cramer, N. Van Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982–1012.
[18] C. Boss, O. Corminboeuf, C. Grisostomi, S. Meyer, A. F. Jones, L. Prade, C. Binkert, W. Fischli, T. Weller, D. Bur, *ChemMedChem* **2006**, *1*, 1341–1345.
[19] G. Siegal, E. AB, J. Schultz, *Drug Discovery Today* **2007**, *12*, 1032–1039.
[20] J. Rinehart, J. Arnold, C. J. Canfield, *Am. J. Trop. Med. Hyg.* **1976**, *25*, 769–774.
[21] K. A. de Villiers, H. M. Marques, T. J. Egan, *J. Inorg. Biochem.* **2008**, *102*, 1660–1667.
[22] D. C. Warhurst, J. C. Craig, I. S. Adagu, D. J. Meyer, S. Y. Lee, *Malar. J. Malaria J.* **2003**, *2*, 26.
[23] P. L. Olliaro, P. I. Trigg, *Bull. W. H. O.* **1995**, *73*, 565–571.
[24] P. Kolb, A. Caflisch, *J. Med. Chem.* **2006**, *49*, 7384–7392.
[25] N. Majeux, M. Scarsi, J. Apostolakis, C. Ehrhardt, A. Caflisch, *Proteins* **1999**, *37*, 88–105.
[26] N. Budin, N. Majeux, A. Caflisch, *Biol. Chem.* **2001**, *382*, 1365–1372.
[27] O. A. Asojo, S. V. Gulnik, E. Afonina, B. Yu, J. A. Ellman, T. S. Haque, A. M. Silva, *J. Mol. Biol.* **2003**, *327*, 173–181.
[28] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
[29] R. Friedman, A. Caflisch, *FEBS Lett.* **2007**, *581*, 4120–4124.
[30] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187–217.
[31] B. R. Brooks, C. L. I. Brooks, A. D. J. MacKerell, L. Nilsson, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. Dinner, S. Fischer, J. Gao, M. Hodoscek, K. Kuczera, T. Lazaridis, J. Ma, E. Paci, R. W. Pastor, C. B. Post, M. Schaefer, B. Tidor, R. W. Venable, H. L. Woodcock, X. Wu, M. Karplus, *J. Comput. Chem.* **2009**; DOI: 10.1002/jcc.21287.
[32] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
[33] J. A. Hartigan, *Clustering Algorithms*, Wiley, New York, **1975**.
[34] M. Cecchini, P. Kolb, N. Majeux, A. Caflisch, *J. Comput. Chem.* **2004**, *25*, 412–422.
[35] N. Majeux, M. Scarsi, A. Caflisch, *Proteins* **2001**, *42*, 256–268.
[36] M. Scarsi, J. Apostolakis, A. Caflisch, *J. Phys. Chem. A* **1997**, *101*, 8098–8106.
[37] S. K. Kearsley, G. M. Smith, *Tetrahedron Comput. Methodol.* **1990**, *3*, 615–633.
[38] W. Im, D. Beglov, B. Roux, *Comput. Phys. Commun.* **1998**, *111*, 59–75.
[39] P. S. Charifson, J. J. Corkery, M. A. Murcko, W. P. Walters, *J. Med. Chem.* **1999**, *42*, 5100–5109.
[40] A. E. Klon, M. Glick, J. W. Davies, *J. Med. Chem.* **2004**, *47*, 4356–4359.
[41] J. Åqvist, C. Medina, J. E. Samuelsson, *Protein Eng.* **1994**, *7*, 385–391.
[42] J. Åqvist, V. B. Luzhkov, B. O. Brandsdal, *Acc. Chem. Res.* **2002**, *35*, 358–365.
[43] O. A. Asojo, A. Afotina, S. V. Gulnik, B. Yu, J. W. Erickson, R. Randad, D. Medjahed, A. M. Silva, *Acta Crystallogr. Sect. D* **2002**, *58*, 2001–2008.
[44] H. J. C. Berendsen, D. van der Spoel, R. Vandrunen, *Comput. Phys. Commun.* **1995**, *91*, 43–56.
[45] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.
[46] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
[47] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[48] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129–145.
[49] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. Elbert, M. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347–1363.
[50] J. Åqvist, *J. Chem. Phys.* **1990**, *94*, 8021–8024.
[51] J. Chandrasekhar, D. C. Spellmeyer, W. L. Jorgensen, *J. Am. Chem. Soc.* **1984**, *106*, 903–910.
[52] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463–1472.
[53] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952–962.
[54] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684–3690.
[55] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
[56] R. Friedman, A. Caflisch, *Proteins* **2008**, *73*, 814–827.
[57] a) W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33–38; b) http://www.ks.uiuc.edu/Research/vmd/.
[58] R. A. Jarvis, E. A. Patrick, *IEEE Trans. Comput.* **1973**, *C22*, 1025–1034.

Received: February 26, 2009
Revised: May 4, 2009
Published online on May 26, 2009

