Design of New Plasmepsin Inhibitors: A Virtual High Throughput Screening Approach on the EGEE Grid

Vinod Kasam,*,[†],[‡] Marc Zimmermann,*,[†] Astrid Maaβ,[†] Horst Schwichtenberg,[†] Antje Wolf,[†] Nicolas Jacq,[‡] Vincent Breton,[‡] and Martin Hofmann-Apitius[†]

Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), 53754 Sankt Augustin, Germany, and LPC Clermont-Ferrand, Campus des Cézeaux, 63177 Aubière Cedex, France

Received October 20, 2006

Though different species of the genus *Plasmodium* may be responsible for malaria, the variant caused by *P. falciparum* is often very dangerous and even fatal if untreated. Hemoglobin degradation is one of the key metabolic processes for the survival of the *Plasmodium* parasite in its host. Plasmepsins, a family of aspartic proteases encoded by the *Plasmodium* genome, play a prominent role in host hemoglobin cleavage. In this paper we demonstrate the use of virtual screening, in particular molecular docking, employed at a very large scale to identify novel inhibitors for plasmepsins II and IV. A large grid infrastructure, the EGEE grid, was used to address the problem of large computation resources required for docking hundreds of thousands of chemical compounds on different plasmepsin targets of *P. falciparum*. A large compound library of about 1 million chemical compounds was docked on 5 different targets of plasmepsins using two different docking software, namely FlexX and AutoDock. Several strategies were employed to analyze the results of this virtual screening approach including docking scores, ideal binding modes, and interactions to key residues of the protein. Three different classes of structures with thiourea, diphenylurea, and guanidino scaffolds were identified to be promising hits. While the identification of diphenylurea compounds is in accordance with the literature and thus provides a sort of “positive control”, the identification of novel compounds with a guanidino scaffold proves that high throughput docking can be effectively used to identify novel potential inhibitors of *P. falciparum* plasmepsins. Thus, with the work presented here, we do not only demonstrate the relevance of computational grids in drug discovery but also identify several promising small molecules which have the potential to serve as candidate inhibitors for *P. falciparum* plasmepsins. With the use of the EGEE grid infrastructure for the virtual screening campaign against the malaria causing parasite *P. falciparum* we have demonstrated that resource sharing on an eScience infrastructure such as EGEE provides a new model for doing collaborative research to fight diseases of the poor.

MOTIVATION

The real motivation to start the WISDOM[1] project comes from the fact that every year millions of people worldwide are getting affected by tropical and protozoan diseases and are even dying if these diseases remain untreated. One such major neglected disease is malaria. Though several national and international efforts like MVI, MMV, and RBM[2−4] are in action to combat this deadly disease, the number of cases and deaths from malaria increases in many parts of the world so that new complementary efforts are needed. As conventional drug discovery processes involving physical high throughput screening are extremely costly, diseases which affect mostly poor people like malaria are often neglected by the pharmaceutical industries and the developed countries.[5] Therefore, we developed a computational screening approach, based on high throughput molecular docking and employed it at large scale on the EGEE grid infrastructure.[6] Our goals for this project were threefold: the biological goal was to identify novel candidate molecules to combat malaria, the biomedical informatics goal was to establish a distributed virtual screening workflow, and the grid computing goal was to demonstrate the usefulness of the grid approach for large scale virtual biomedical experimentation.

INTRODUCTION

Malaria is a dreadful disease affecting 300 million people and killing 1−1.5 million people every year. Malaria is caused by a protozoan parasite, belonging to the genus *Plasmodium*.[7−9] There are several species of *Plasmodium* infecting cattle, birds, and humans. The four species *P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale* are in particular considered important as these species infect humans.[10] One of the main causes for the comeback of malaria is that the most widely used drug against malaria, chloroquine, has been rendered useless by drug resistance in much of the world. New antimalarial drugs are presently available, but the potential emergence of resistance,[11] the difficulty to synthesize these drugs at a large scale, and their cost[12] make it of utmost importance to keep searching for new drugs.

Hemoglobin metabolism is one the key metabolic processes for the survival of the parasite in its human blood stages. There are several proteases involved in the degrada-

* Corresponding author phone: +33-47340-5324; fax: +33-47326-4598; e-mail: kasam@clermont.in2p3.fr, kasam@scai.fraunhofer.de (V.K.); phone: +49-224114-2276; fax: +49-224114-2656; e-mail: marc.zimmermann@scai.fraunhofer.de (M.Z.).

[†] Fraunhofer Institute for Algorithms and Scientific Computing (SCAI).

[‡] LPC Clermont-Ferrand.

10.1021/ci600451t CCC: $37.00
Published on Web 08/30/2007

tion of host erythrocyte hemoglobin inside the food vacuole of the parasite. Plasmepsins, the aspartic proteases of *Plasmodium* species, are responsible for the initial cleavage of hemoglobin (Hb) and later followed by other proteases. There are ten different plasmepsins encoded by ten different genes in *P. falciparum* (Plm I, II, IV, V, VI, VII, VIII, IX, X, and HAP). High levels of sequence homology are observed between different plasmepsins (65−70%).[13−15] At the same time, they share only 35% sequence homology with its nearest human aspartic protease, Cathepsin D.[16] This together with the presence of sound X-crystallographic data make plasmepsins ideal targets for malaria and rational drug design approaches, respectively. Though several peptidic and non-peptidic inhibitors have been described as inhibitors for plasmepsins, none of them were effective in killing the parasite in cell culture. This is due to the fact that large size compounds cannot easily penetrate the food vacuole where hemoglobin degradation occurs.[17,18] Here we present novel small molecules potentially able to serve as lead compounds for the development of new inhibitors of *P. falciparum* plasmepsins.

Advance in combinatorial chemistry has paved the way for synthesizing millions of different chemical compounds. Thus, there are millions of chemical compounds available in the labs and also in 2D, 3D electronic databases, but it is very expensive to physically test such a high number of compounds in the experimental labs by high throughput screening (HTS). Besides the significant costs, the hit rate in HTS is on average quite low: approximately 1 out of 100 000 compounds when screened on targets such as enzymes.[19] An alternative to physical HTS is high throughput virtual screening by molecular docking, a technique which can screen millions of compounds rapidly, reliably, and cost effectively. There is a steadily increasing number of success stories published by different scientific groups all over the world.[20−22] The general screening techniques are random screening and generation of focused libraries based on pharmacophore with subsequent screening of the resulting virtual compound. Although both approaches have their specific advantages and disadvantages, they are widely in use.[23] For our WISDOM project, we applied a random screening approach.

Screening millions of chemical compounds in the computer brings along a high storage complexity which means a computational data challenge on its own. Screening each compound, depending on structural complexity, can take from a few minutes to hours on a standard PC, which means screening all compounds in a large virtual compound library can take years of computation time on a single machine. This problem can be addressed by distributing the work on a large computational grid of thousands of computers, reducing the time for screening large virtual compound libraries to days.[24−26] So far, in silico drug discovery on gridlike infrastructures has successfully been deployed for the search for new drugs against smallpox,[27] anthrax,[28] and cancer.[29,30] The current aim of the WISDOM project (WISDOM being an acronym standing for **W**ide **I**n **S**ilico **D**ocking **O**n **M**alaria) is to identify novel inhibitors for plasmepsins II and IV and to demonstrate the relevance of computational grids for large scale virtual drug discovery approaches.

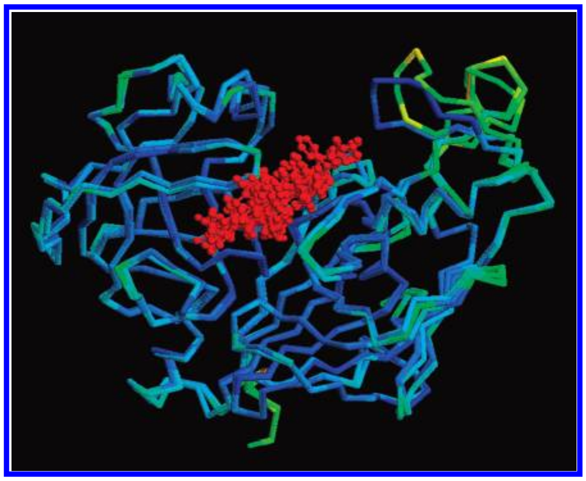


Figure 1. Screen shot of 5 plasmepsin structures superimposed. The different protein structures used as templates share high sequence similarity. Variations exist in the loop regions and are highlighted by the color yellow. The cocrystallized ligands are represented as balls and sticks by the color red. The picture has been generated using Rasmol.

WISDOM

Target Selection. Several target proteins play a pivotal role in the life cycle of *P. falciparum*. These targets are well studied and have been validated for their significance for parasite survival as well as their "drugability".[31,32] However, there is considerable concern in malaria so far as the available drugs focus on a limited number of biological targets. Therefore, there is unanimity that substantial scientific effort should be devoted to the development of additional novel targets, in the hope that subsequently developed drugs would not demonstrate cross-resistance with presently known antimalarials. Moreover, with the sequencing of the *Plasmodium* genome many new potential targets came into light. These potential antimalarial drug targets can be broadly classified into three categories, and each category has many individual targets. The three categories are (i) targets involved in hemoglobin degradation (proteases like plasmepsins, falcipains), (ii) targets involved in metabolism, and (iii) targets engaged in membrane transport and signaling.[33,34]

The significance of plasmepsins in the *Plasmodium* life cycle and the presence of X-ray crystal structure data make plasmepsins ideal targets for antimalaria therapy and new approaches in rational drug design, respectively. We therefore chose plasmepsins as the target of choice for the current study.

Target Preparation. The structures of *P. falciparum* plasmepsins were retrieved from PDB.[35] The 3D coordinates used in the current study are 1lee (crystal structure of plasmepsin from *P. falciparum* in complex with inhibitor RS367), 1lf2 (crystal structure of plasmepsin from *P. falciparum* in complex with inhibitor EH58), and 1ls5 (crystal structure of plasmepsin IV from *P. falciparum* in complex with pepstatin A). In the first step, the C-alpha carbons of all the plasmepsins are superimposed on 1lee. Therefore, in order to transfer and compare binding modes between different receptor structures, all the water molecules and cocrystallized ligands are removed. The superimposed C α carbon traces of all the protein structures are represented in Figure 1. The active site comprises all atoms within 6.5 Å of the cocrystallized ligands as well as residues of known

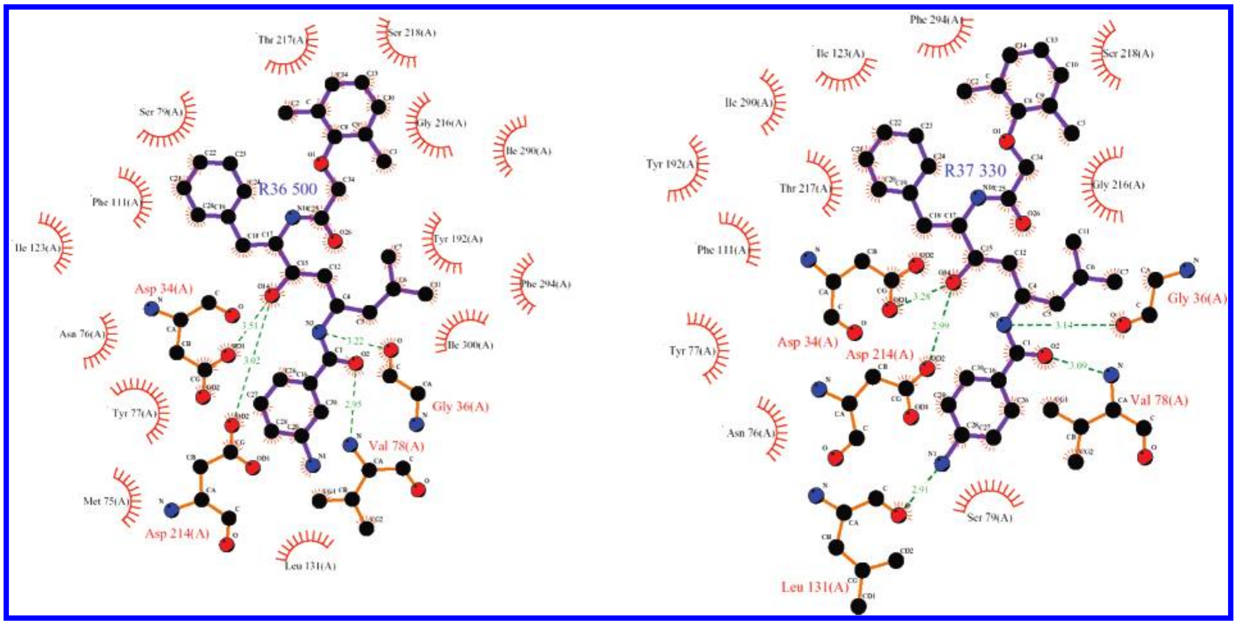


Figure 2. Ligand plots of target structures 1lee (left) and 1lf2 (right). The ligands are represented in a ball and stick model in CPK color. Hydrogen bonding between the ligand and the active site residues are indicated by green dotted lines, and the hydrophobic environment is indicated by a crecent shape. The plots are obtained from www.pdb.org.

relevance. The charges of the ionizable groups are chosen to be consistent with acidic conditions (pH 5). The side chains of lysine and arginine residues are protonated, and, in addition, the side chain of histidine is protonated (for comparability reasons, since the docking tool AutoDock regards all histidine residues as protonated). The carboxylic groups of glutamic acid and aspartic acid are deprotonated.

Description of 1lee and 1lf2 Active Sites. The active site description is of great importance as it gives information on the binding mode of cocrystallized inhibitors and it also serves as template when defining the active site and validating the compounds, so we describe it in detail. Each crystal has one monomer per asymmetric unit. These complexes disclose key hydrogen bonds between the inhibitor and the active site residues, particularly with the flap residues Val78 and Ser79, the catalytic dyad Asp34 and Asp214, and the residues Ser218 and Gly36 that are in closeness to the catalytic dyad.[36,37] The ligand plots of the target structures 1lee and 1lf2 are shown in Figure 2. As the resolutions for 1lf3 2.7 Å and 1ls5 2.8 Å are clearly suboptimal, the structural details are not discussed in detail.

Target Scenarios. Different scenarios have been prepared for docking based on the inclusion of crystal water molecules in the active site. Several tests have been done to check which crystal water molecules are influencing the docking scores (energy) and poses in terms of RMSD values (Root-Mean-Square Deviations). The different proteins and their original residue numbers for crystal water molecules used in the docking procedure with FlexX are given in Table 1.

Compound Database Selection. Compound libraries used in the virtual screening experiments should be filtered first to remove unsuitable compounds that would not reach and pass the clinical trials due to undesired and toxic properties. A very popular method to evaluate the drug likeness of a candidate structure is the so-called Lipinski "Rule-of-five". The Compound library used for WISDOM was obtained from the ZINC database.[38,39] The ZINC database is a collection of 3.3 million chemical compounds from different vendors. We have chosen to use the ZINC library because ZINC is an open source database, the structures have already been filtered according to the Lipinski rules, and the data are available in different file formats (Sybyl mol2 format, sdf, and smiles). So, basically, ZINC provides virtual compounds ready for virtual screening. A total of 1 000 000 compounds were downloaded from the ZINC database, including about 500 000 ChemBridge compounds and about 500 000 additional druglike compounds from other vendors. As AutoDock requires the pdbqs file format, all the compounds were first converted into pdbq format using ADT tools.[40]

Table 1. Different Plasmepsin Structures and the Corresponding Crystal Water Molecules with Their Original Residue Numbers Used during the FlexX Data Challenge

protein	water molecules
1lee	protein structure 1lee without any crystal water molecule
1lee_h2	protein structure 1lee with two crystal water molecules (residues 1256, 1270)
1lee_h3	protein structure 1lee with three crystal water molecules (residues 1065, 1125, 1247)
1lf2	protein structure 1lf2 without any crystal water molecule
1lf2_h	protein structure 1lf2 with one crystal water molecule (residue 585)
1lf3	protein structure 1lf3 without any crystal water molecules
1ls5_a	protein structure 1ls5, chain A without any crystal water molecules
1ls5_b	protein structure 1ls5 chain B without any crystal water molecules

Docking Software. Two different docking software have been used in the current study, FlexX[41,42] and AutoDock.[43,44] FlexX is an extremely fast, robust, and highly configurable computer program for predicting protein−ligand interactions. Standard parameter settings are used except for two cases ("place particles"[45] and "maximum overlap volume"). These two parameters were subject to deliberate variation with FlexX, Table 2.

The second docking tool, AutoDock, is a suite of automated docking tools. AutoDock parameter sets used in

Table 2. Parameter Sets Used during the FlexX Data Challenge[a]

parameter sets	place particles	maximum overlap volume
parameter set 1	yes	2.5
parameter set 2	yes	5
parameter set 3	no	2.5
parameter set 3	no	5

[a] Place particles and max overlap volume are the two variations used in FlexX software.

WISDOM project are (1) GALS (Genetic Algorithm Local Search with Solis-Wets (SW)) and (2) GALS (Genetic Algorithm Local Search with pseudo Solis-Wets (pSW)).[46]

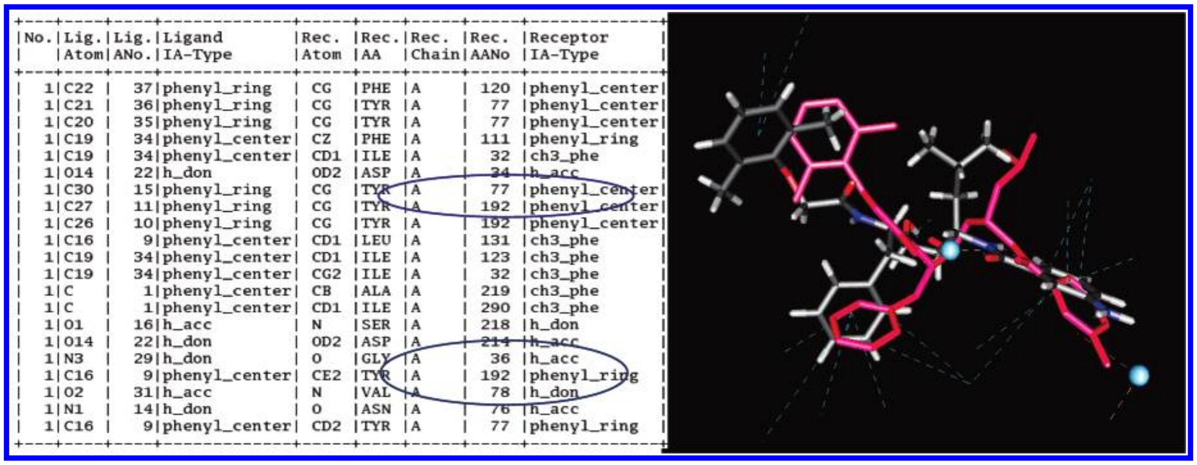
VIRTUAL DOCKING PROCESS

Redocking, Cross-Docking, and Docking under Different Parameter Sets. Direct docking and redocking experiments are performed between the target structures and their respective cocrystallized ligands on different parameter sets for FlexX. Validation was done by comparing the interaction information between the ligand and target to the ligand plot information obtained from the Brookhaven protein database. Despite the large ligand size (>15 ligand components and >12 rotatable bonds) the RMSD values in redocking experiments for 1lee and 1lf2 were convincing, about 2.5 Å for the top ranking solutions. The scores and RMSD values of the redocking are shown in Table 3. In Figure 3 the binding mode of the ligand R36 in the redocking experiment with target 1lee is displayed. The RMSD values for 1lf3 and 1ls5 are marginal and >3 Å in all the parameter sets, see Table 3. This may be due to the suboptimal resolution of the X-ray crystal structure, very large ligand size, numerous ligand components (cocrystalized ligand of 1lf3 has about 24 fragments), and consequently a high number of rotatable bonds. In the redocking experiment for llf3, we observed the rotation of a ligand, while maintaining the essential contacts to the catalytic dyad: this may be due to particular crystallization conditions which have such an influence, thus the calculated position is not necessarily wrong. The resulting binding modes for 1lee and 1lf2 were well in concordance with the crystal structure. The binding mode of the best ranked solution displayed good interactions with catalytic residues of the targets (Asp214, Asp34) and also with other residues of relevance. This comparison revealed that the ligand in the protein structures 1lee and 1lf2 has found all the significant interactions responsible for the activity of the protein. From parameter set 1 it became also clear that direct docking (targets without any crystal water) performed well both in terms of scoring and RMSD values. From interaction information it is observed that protein structure 1lee without any crystal water molecules forms hydrogen bonds with both the catalytic residues Asp214 and Asp34 as well as with flap residues Val78 and Ser218. Remarkably, for the protein structure 1lee with crystal water molecules, all ligands failed to form interactions with the key residues.

Table 3. Overview of Docking Scores for the Best Ranking Solutions and Their Corresponding RMSD Values in Direct Docking and Redocking Experiments Obtained for Four Different Parameter Sets with FlexX[a]

		1		2		3		4	
target	ligand	total score	RMS-value	total score	RMS-value	total score	RMS- value	total score	RMS-value
1lee	1lee (R36)	−21.128	2.34	−15.348	9.66	−28.075	9.82	−25.914	2.09
1lee_h2	1lee (R36)	−20.079	8.51	−14.806	9.51	−25.959	2.09	−25.959	2.09
1lee_h3	1lee (R36)	−26.197	9.80	−19.664	8.92	−26.431	9.81	−26.039	2.09
1lf2	1lf2 (R37)	−24.319	4.93	−23.401	3.26	−22.134	9.86	−24.672	9.37
1lf2_h	1lf2 (R37)	−19.563	10.03	−27.984	4.81	−22.962	2.77	−24.672	9.37
1lf3	1lf3 (E58)	−20.928	8.97	−19.461	13.78	−13.941	7.60	−16.921	12.46
1ls5_a	1ls5 pepstatin A	−23.219	10.87	−22.35	11.72	−27.677	3.22	−33.698	3.13
1ls5_b	1ls5 pepstatin A	−23.105	10.52	−20.708	12.20	−30.313	4.92	−24.86	11.71

[a] The numbers 1, 2, 3, and 4 correspond to parameter sets 1, 2, 3, and 4, respectively, see Table 2. Units for the total score and the RMS value are kJ/mol and Å, respectively.



```
+----+----+----+--------------+-----+----+-----+-----+---------------+
|No.|Lig.|Lig.|Ligand        |Rec. |Rec.|Rec. |Rec. |Receptor       |
|   |Atom|ANo.|IA-Type       |Atom |AA  |Chain|AANo |IA-Type        |
+----+----+----+--------------+-----+----+-----+-----+---------------+
|   1|C22 |   37|phenyl_ring   | CG  |PHE |A    |  120 |phenyl_center|
|   1|C21 |   36|phenyl_ring   | CG  |TYR |A    |   77 |phenyl_center|
|   1|C20 |   35|phenyl_ring   | CG  |TYR |A    |   77 |phenyl_center|
|   1|C19 |   34|phenyl_center| CZ  |PHE |A    |  111 |phenyl_ring  |
|   1|C19 |   34|phenyl_center| CD1 |ILE |A    |   32 |ch3_phe      |
|   1|O14 |   22|h_don        | OD2 |ASP |A    |   34 |h_acc        |
|   1|C30 |   15|phenyl_ring   | CG  |TYR |A    |   77 |phenyl_center|
|   1|C27 |   11|phenyl_ring   | CG  |TYR |A    |  192 |phenyl_center|
|   1|C26 |   10|phenyl_ring   | CG  |TYR |A    |  192 |phenyl_center|
|   1|C16 |    9|phenyl_center| CD1 |LEU |A    |  131 |ch3_phe      |
|   1|C19 |   34|phenyl_center| CD1 |ILE |A    |  123 |ch3_phe      |
|   1|C19 |   34|phenyl_center| CG2 |ILE |A    |   32 |ch3_phe      |
|   1|C   |    1|phenyl_center| CB  |ALA |A    |  219 |ch3_phe      |
|   1|C   |    1|phenyl_center| CD1 |ILE |A    |  290 |ch3_phe      |
|   1|O1  |   16|h_acc        | N   |SER |A    |  218 |h_don        |
|   1|O14 |   22|h_don        | OD2 |ASP |A    |  214 |h_acc        |
|   1|N3  |   29|h_don        | O   |GLY |A    |   36 |h_acc        |
|   1|C16 |    9|phenyl_center| CE2 |TYR |A    |  192 |phenyl_ring  |
|   1|O2  |   31|h_acc        | N   |VAL |A    |   78 |h_don        |
|   1|N1  |   14|h_don        | O   |ASN |A    |   76 |h_acc        |
|   1|C16 |    9|phenyl_center| CD2 |TYR |A    |   77 |phenyl_ring  |
+----+----+----+--------------+-----+----+-----+-----+---------------+
```

Figure 3. Redocking of ligand (R36) into target structure 1lee. The docking solution is represented in CPK colors, while the binding mode of R36 before docking is shown by the color red. The RMSD value between the two binding modes is 2.3 Å. Interactions to the key residues are indicated by blue circles. The figure is generated by using FlexV.

DEPLOYMENT STRATEGY

Several test runs were launched before the full deployment on the EGEE grid. The major purpose of performing test runs is to cover technical as well as modeling aspects: the system credibility in terms of modeling, performance, setting, and tuning of parameters and analyzing software parameter influence on docking results are checked. As the X-ray resolution of target structures 1lee (1.8 Å) and 1lf2 (1.9 Å) were well suited for our purposes, major test runs were performed on 1lee and 1lf2. Test runs were performed with a combination of known compounds (Walter Reed compounds found to have micromolar inhibitions in vitro),[17] 20 000 ZINC compounds, 400 ZINC compounds, and the FlexX 200 standard data set.

Test Run 1. Twenty thousand chemical compounds were selected from the ZINC database by using two constraints, the presence of a phenyl ring in all the structures and a molecular weight above 400 Da. For the target structure 1lee about 1300 compounds achieved better docking scores than the original ligand (R36). Likewise for the target structure 1lf2 about 7000 compounds achieved better docking scores than the cocrystallized ligand (R37). The binding modes of the top 20 compounds were visualized manually and found to be satisfying.

Test Run 2. The FlexX 200 data set is a highly diverse data set of 200 structurally known protein−ligand complexes obtained from BiosolveIT GmbH, the commercial vendor of FlexX. A docking experiment was performed on this data set. The results were exactly in accordance with the published data,[47] indicating that the overall settings chosen for the random docking are satisfactory.

Test Run 3. The main reason for this test run was to check the numerical stability of the results obtained with different computers. Moreover, we wanted to test the grid with a smaller data set before we started the main data challenge. Thus, this docking experiment was performed on the EGEE grid. A compound library of 400 compounds was prepared based on the results of test run 1. Different docking scenarios were prepared (a scenario corresponds to a specific combination of target parameters and software parameters), and the above compounds were docked into the target structures 1lee and 1lf2. As the calculations were performed on different grid computing elements, we were able to check the influence of the hardware on the docking scores. As a result, no significant change in scoring was observed. This confirmed that in our experimental setup the influence of the hardware on the docking results is negligible, reflecting also the fact that the grid is quite homogeneous with respect to processor types.

Final Deployment. Based on the results from direct docking, redocking, cross-docking, and test runs, different parameters were prepared for both targets and docking software. Finally, large scale computations have been conducted for 8 receptor scenarios (5+3 see Table 3) for FlexX and 10 (5+5) receptor scenarios for AutoDock with 1 million compounds for each receptor scenario on the EGEE grid infrastructure. The final data challenge has been performed between July 11, 2005 and August 30, 2005. The data challenge witnessed 42 million docking experiments on more than 1700 computers distributed in Europe and Asia. The description of grid implementation is out of scope of this paper. The details about grid implementation and deployment of the docking jobs can be found in these latest publications.[48,49]

For the current project, a supercomputer or a computer cluster with hundreds of computers will be able to perform the same application. Since we are working on neglected diseases like malaria, we want to utilize the free resources. EGEE is one of the free computing resources provider for neglected diseases. More details on the advantages of the EGEE infrastructure can be found at http://public.eu-egee.org. A description of the computational grid infrastructure, EGEE, and the supercomputer is also made available in the Supporting Information. Besides that one of our aims of the project is also to demonstrate the capability of a computational grid executing the biomedical application like virtual screening at large scale.

ANALYSIS STRATEGY

Summary of the Output. We encountered problems with the parametrization of AutoDock, and the results have not been convincing (high internal energies, for more accurate reasons for omitting the discussion on AutoDock results, see the Supporting Information). Hence we will focus the subsequent analysis solely on the results obtained using FlexX. The outputs of the docking results in FlexX are log files. The log files are converted into CSV file format (comma separated files). These CSV files in turn serve as input for VS explorer, a Java based prototype software for analyzing virtual and high throughput screening results developed at SCAI (Fraunhofer Institute for Algorithms and Scientific Computing, Germany). Three different forms of results are saved and analyzed from each docking assay: (i) docking scores of the ten best solutions after clustering, (ii) interaction information between protein and ligands of the ten best solutions, and (iii) binding modes of the ten best solutions. Moreover the ranking process is the integral part of the docking software. FlexX have a postprocessing optimization of the docking solution and clustering. Clustering in FlexX is based on RMSD, angle, and distance deviation. Default values of FlexX are used as clustering cutoffs. The overall filtering process we employed is shown in Figure 4.

Clustering and Match Information. Usually, result analysis concentrates on the best ranking solution only or on the 5 best solutions based on docking score. But when exhaustive analysis is done for all predictions, there is a smoothening of score observed from the best solution to the next best solution in many cases. Moreover, the binding modes are very often nearly the same. To address this problem, result analysis of the ten best solutions after clustering is done: this allows screening diverse binding modes and identifying compounds with interactions to key residues of the protein even if the score is not optimal.

Strategies in Result Analysis. The aim of result analysis is reducing the number of false positives and finally identifying a few hundred promising compounds that can be tested in experimental laboratories. Two strategies are employed for result analysis: statistical analysis based on scoring values (for more details on statistical analysis refer to the Supporting Information) and analysis of interactions

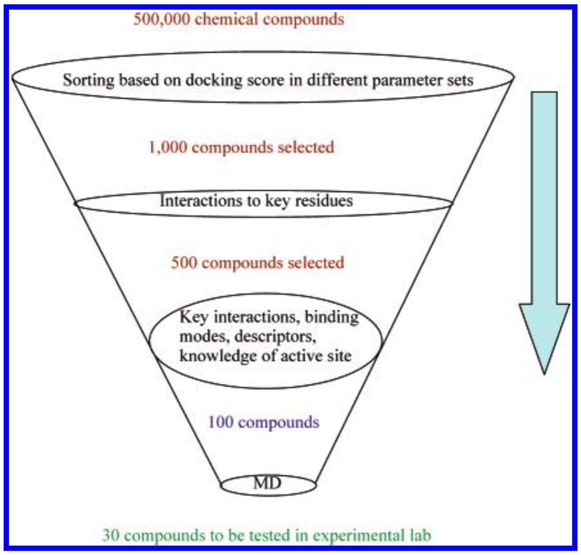

500,000 chemical compounds
Sorting based on docking score in different parameter sets
1,000 compounds selected
Interactions to key residues
500 compounds selected
Key interactions, binding modes, descriptors, knowledge of active site
100 compounds
MD
30 compounds to be tested in experimental lab

Figure 4. Representation of overall filtering process. The process starts with 500 000 compounds, and at different stages various filters are applied as indicated in the figure to reduce the false positives and finally to identify the best possible hits which are more likely to be leads.

between the ligand and key residues of the target protein (match-information).

Data Validation. The first step in the result analysis is checking whether the docking run was correct.

As the grid jobs are submitted in chunks of compounds, a set of 14 known reference compounds was submitted every time. When the scores obtained for the reference compounds were not coherent with the determined reference scores, the chunk was discarded from further analysis.

Result Analysis Based on Match Information. FlexX provides interaction information between the protein and ligand for each docking experiment. The information of catalytic residues and other residues of relevance is obtained from the literature and ligand plots (see Figure 2). Thus only those compounds with interactions to relevant residues are extracted. The following residues are considered to be significant: ASP-214, ASP-34, VAL-78, SER-79, SER-218, and GLY-36.

RESULTS AND DISCUSSION

Top Scoring Compounds. As evident from the redocking experiments, the target 1lee without any inclusion of crystal water performed well both in terms of RMSD values and docking scores, so we restricted the final result analysis to target structure 1lee, i.e., without any crystal water molecules. Figure 5 represents the top ten compounds based on the score from parameter set 1. Visualization of the docking solutions shows that, although some compounds possess high scores, they are quite far away from the center of the binding pocket. As we are looking for competitive inhibitors, the ideal

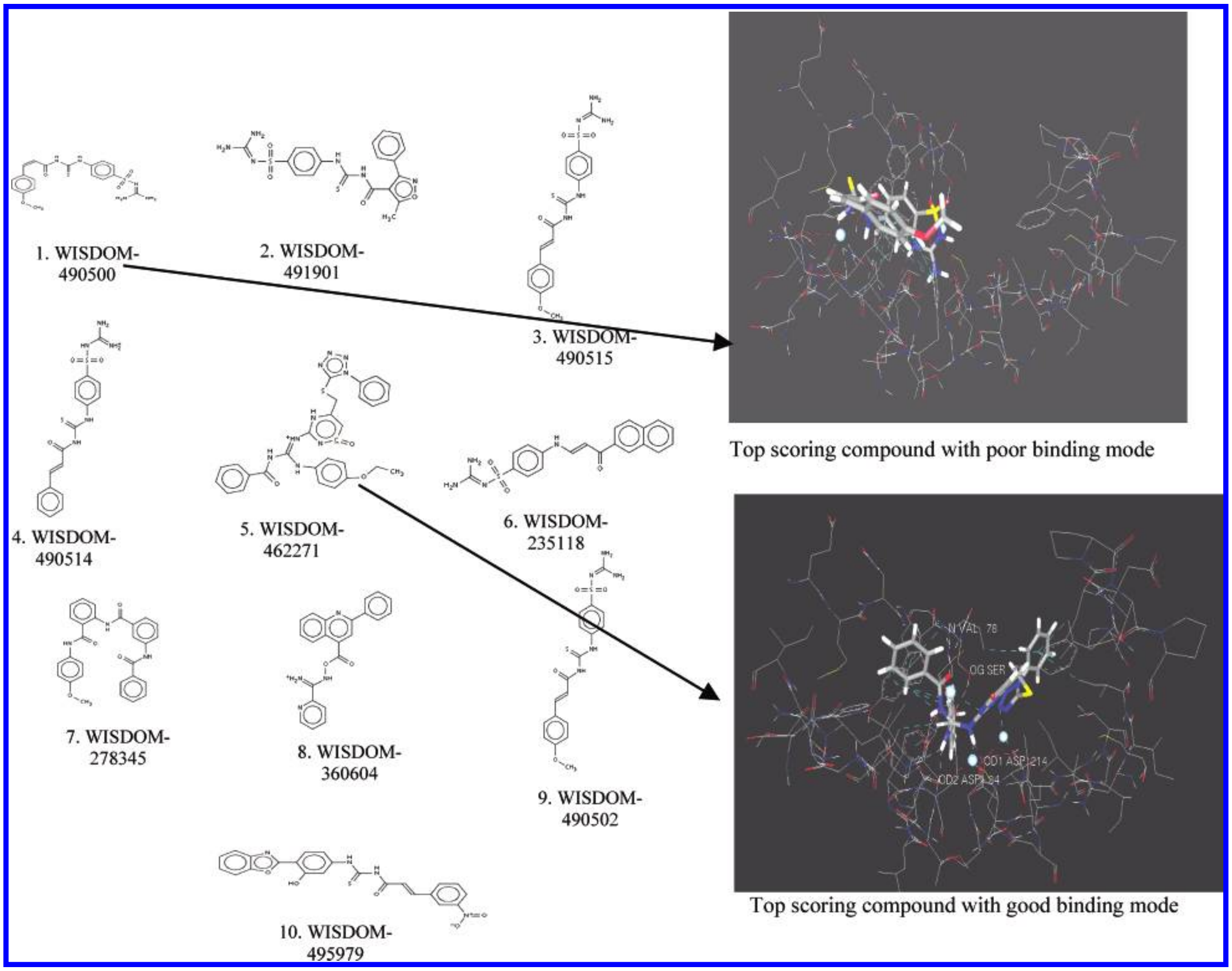
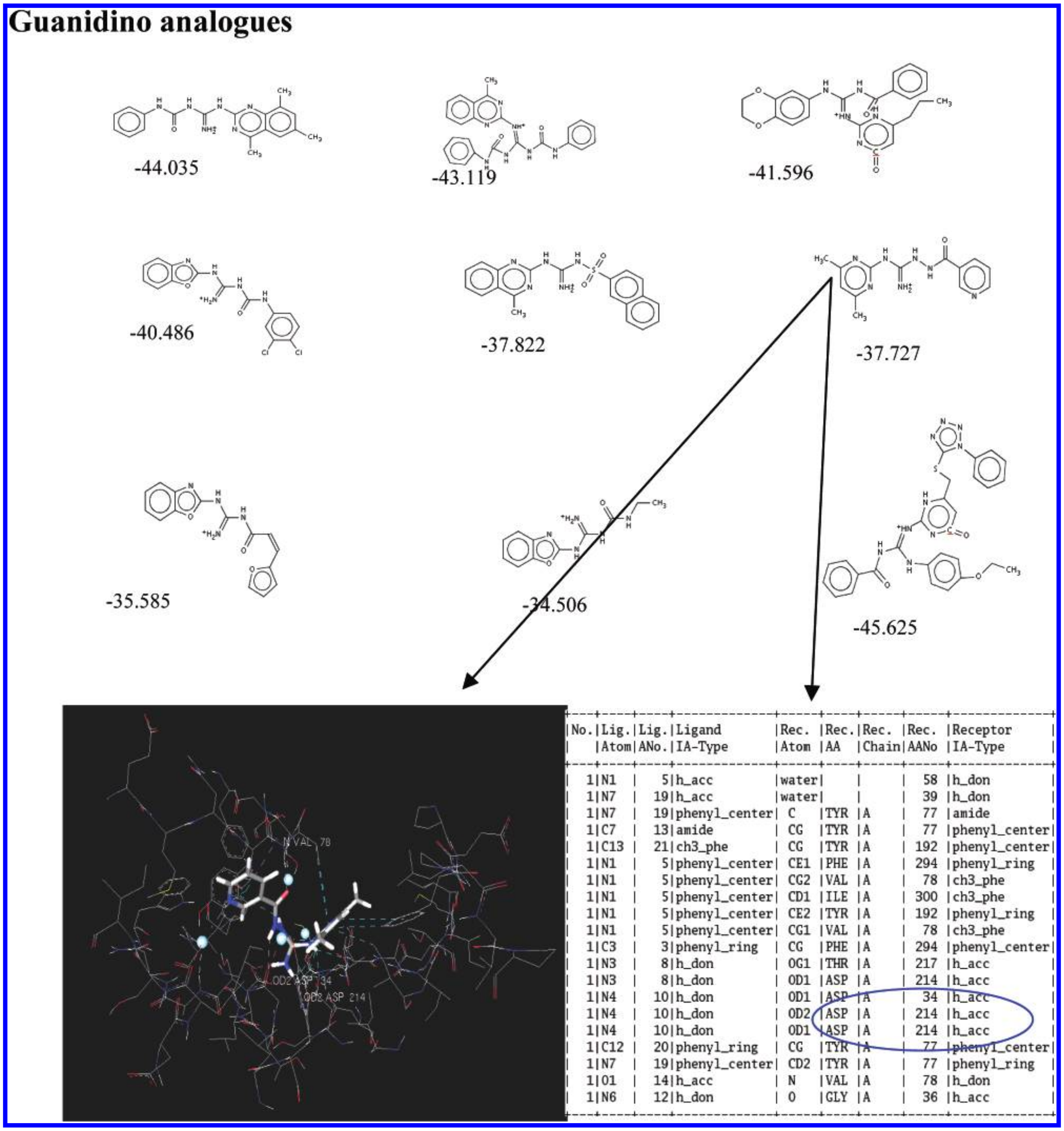


Figure 5. Representation of the top scoring compounds in parameter set 1. Top scoring compounds with poor binding mode and good binding modes inside the active site of the protein (pdb id: 1lee) are indicated with arrows.



```
+---+----+----+--------------+----+----+-----+-----+--------------+
|No.|Lig.|Lig.|Ligand        |Rec. |Rec.|Rec. |Rec. |Receptor     |
|   |Atom|ANo.|IA-Type       |Atom |AA  |Chain|AANo |IA-Type      |
+---+----+----+--------------+----+----+-----+-----+--------------+
|  1|N1  |   5|h_acc         |water|    |     |   58 |h_don        |
|  1|N7  |  19|h_acc         |water|    |     |   39 |h_don        |
|  1|N7  |  19|phenyl_center| C    |TYR |A    |   77 |amide        |
|  1|C7  |  13|amide         | CG   |TYR |A    |   77 |phenyl_center|
|  1|C13 |  21|ch3_phe       | CG   |TYR |A    |  192 |phenyl_center|
|  1|N1  |   5|phenyl_center| CE1  |PHE |A    |  294 |phenyl_ring  |
|  1|N1  |   5|phenyl_center| CG2  |VAL |A    |   78 |ch3_phe      |
|  1|N1  |   5|phenyl_center| CD1  |ILE |A    |  300 |ch3_phe      |
|  1|N1  |   5|phenyl_center| CE2  |TYR |A    |  192 |phenyl_ring  |
|  1|N1  |   5|phenyl_center| CG1  |VAL |A    |   78 |ch3_phe      |
|  1|C3  |   3|phenyl_ring   | CG   |PHE |A    |  294 |phenyl_center|
|  1|N3  |   8|h_don         | OG1  |THR |A    |  217 |h_acc        |
|  1|N3  |   8|h_don         | OD1  |ASP |A    |  214 |h_acc        |
|  1|N4  |  10|h_don         | OD1  |ASP |A    |   34 |h_acc        |
|  1|N4  |  10|h_don         | OD2  |ASP |A    |  214 |h_acc        |
|  1|N4  |  10|h_don         | OD1  |ASP |A    |  214 |h_acc        |
|  1|C12 |  20|phenyl_ring   | CG   |TYR |A    |   77 |phenyl_center|
|  1|N7  |  19|phenyl_center| CD2  |TYR |A    |   77 |phenyl_ring  |
|  1|O1  |  14|h_acc         | N    |VAL |A    |   78 |h_don        |
|  1|N6  |  12|h_don         | O    |GLY |A    |   36 |h_acc        |
+---+----+----+--------------+----+----+-----+-----+--------------+
```

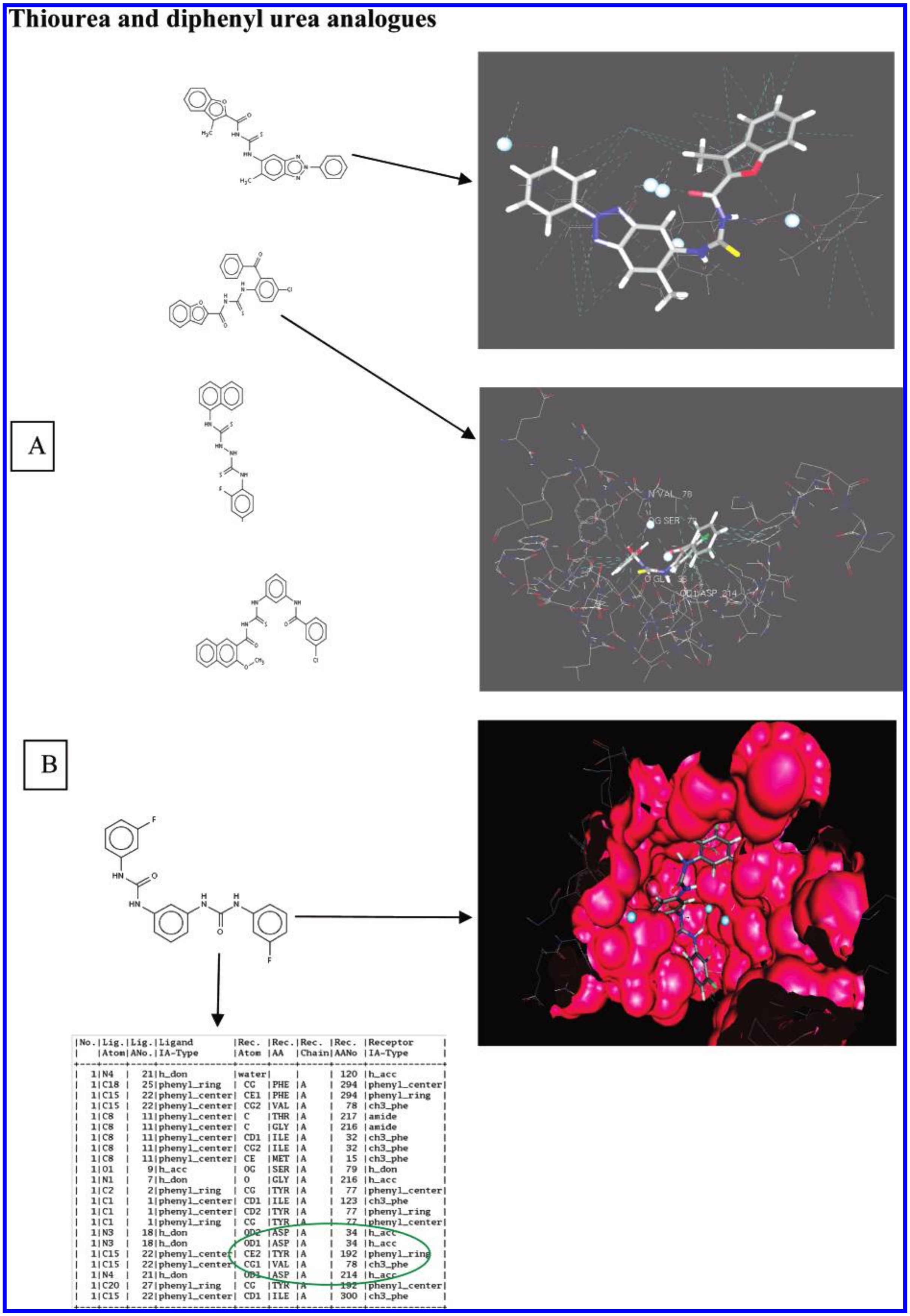
Figure 6. Representation of top scoring guanidino analogues and their respective docking score (kJ/Mol) with FlexX indicated below the compound. One compound with an ideal binding mode and interactions to key residues is displayed. The interactions between the compounds and the key amino acids of the protein are highlighted by the blue circle.

compounds will be the ones which are well within the binding pocket. Consequently, compounds which are remote from the binding pocket are rejected. An example of a top scoring compound with poor binding mode is represented in Figure 5.

Special attention has been given to all individual complexes for the top 1000 compounds from all four parameter sets. As observed with parameter 1, some of the top scoring compounds from various other parameter sets failed to form the expected interactions to key residues of the protein, and the binding mode was not convincing. So we employed several filters for the final selection of the compounds. (Filtering criteria is indicated in Figure 4). After undergoing the filtering procedure, 100 compounds have been selected for reranking by molecular dynamics. Most of the compounds selected are thiourea, diphenyl urea, and guanidino analogues.

Another interesting observation is that diphenylurea analogues are already known to be micromolar inhibitors for plasmepsin (Walter Reed compounds).[17] This suggests that the overall approach is sensible for the discovery of inhibitors for plasmepsins. Figure 7 (B) represents a diphenylurea analogue inside the active site, and the interactions to key residues of the target are highlighted in a circle. A close inspection revealed that this compound displays a similar binding mode as cocrystallized ligand (R36) and forms the expected key interactions.

The other group of compounds gathers thiourea analogues. There is always a consensus in placing the core group



```
|No.|Lig.|Lig.|Ligand        |Rec. |Rec.|Rec. |Rec. |Receptor     |
|   |Atom|ANo.|IA-Type       |Atom |AA  |Chain|AANo |IA-Type      |
+---+----+----+--------------+-----+----+-----+-----+-------------+
|  1|N4  |  21|h_don         |water|    |     | 120 |h_acc        |
|  1|C18 |  25|phenyl_ring   | CG  |PHE |A    | 294 |phenyl_center|
|  1|C15 |  22|phenyl_center| CE1  |PHE |A    | 294 |phenyl_ring  |
|  1|C15 |  22|phenyl_center| CG2  |VAL |A    |  78 |ch3_phe      |
|  1|C8  |  11|phenyl_center| C    |THR |A    | 217 |amide        |
|  1|C8  |  11|phenyl_center| C    |GLY |A    | 216 |amide        |
|  1|C8  |  11|phenyl_center| CD1  |ILE |A    |  32 |ch3_phe      |
|  1|C8  |  11|phenyl_center| CG2  |ILE |A    |  32 |ch3_phe      |
|  1|C8  |  11|phenyl_center| CE   |MET |A    |  15 |ch3_phe      |
|  1|O1  |   9|h_acc         | OG  |SER |A    |  79 |h_don        |
|  1|N1  |   7|h_don         | O   |GLY |A    | 216 |h_acc        |
|  1|C2  |   2|phenyl_ring   | CG  |TYR |A    |  77 |phenyl_center|
|  1|C1  |   1|phenyl_center| CD1  |ILE |A    | 123 |ch3_phe      |
|  1|C1  |   1|phenyl_center| CD2  |TYR |A    |  77 |phenyl_ring  |
|  1|C1  |   1|phenyl_ring   | CG  |TYR |A    |  77 |phenyl_center|
|  1|N3  |  18|h_don         | OD2 |ASP |A    |  34 |h_acc        |
|  1|N3  |  18|h_don         | OD1 |ASP |A    |  34 |h_acc        |
|  1|C15 |  22|phenyl_center| CE2  |TYR |A    | 192 |phenyl_ring  |
|  1|C15 |  22|phenyl_center| CG1  |VAL |A    |  78 |ch3_phe      |
|  1|N4  |  21|h_don         | OD1 |ASP |A    | 214 |h_acc        |
|  1|C20 |  27|phenyl_ring   | CG  |TYR |A    | 192 |phenyl_center|
|  1|C15 |  22|phenyl_center| CD1  |ILE |A    | 300 |ch3_phe      |
+---+----+----+--------------+-----+----+-----+-----+-------------+
```

Figure 7. (A) Top scoring thiourea analogues. The binding mode of one of the compounds inside the active site is displayed. (B) A compound from diphenyl urea analogues is displayed. The binding mode and interaction information are shown. Highlighted by the green circle are compound interactions to one of the catalytic residues (ASP 34) of the target protein. The figure was generated by using FlexV. Interactions are indicated by dotted lines.

(thiourea) with the sulfur atom positioning itself toward the flap residue Val78 and the two nitrogen atoms making interactions to the catalytic Asp214 or Asp34 or both. This observation is well in concordance with the binding modes of the Walter Reed compounds as described in ref 17 as well as with the binding modes of the cocrystallized ligands.

The most exciting observation from the current study is the identification of the guanidino analogues. These compounds are very promising as they obeyed all the filtering criteria we employed in finding the hits. Figure 6 represents the guanidino analogues with their respective docking scores in kJ/mol. The binding mode of one compound is shown in the active site of protein 1lee. The interactions to key residues are highlighted by circles. Like the thiourea compounds, there was a consensus observed in the binding modes of guanidino compounds: the deprotonated nitrogen atom positioning itself toward the flap residue Val78 and the adjacent nitrogen atoms making interactions to catalytic residues Asp214 and/or Asp34. Guanidino analogues are likely to be a novel class of compounds, as they have not yet been reported as inhibitors for plasmepsins. Additionally, chemically diverse compounds have also been identified as hits, including thiazole analogues. About 18 different chemical descriptors are calculated for all the finally selected 100 compounds to reduce the late stage attrition rates. Almost all the compounds possess acceptable chemical descriptor values.

CONCLUSION AND PERSPECTIVES

The eScience paradigm is based on the observation that an increasing number of scientific problem solving approaches require large computational efforts. One of the key features of eScience is that it supports scientific work through mediating collaboration between individual researchers in virtual organizations and that it enables large scale experimentation through the sharing of resources.

The WISDOM project we present in this paper is one example for the successful utilization of this paradigm in the area of computational life sciences. Making use of the world's largest scientific compute infrastructure, the EGEE grid, we have realized a large virtual screening project aiming at the identification of new, potential candidate molecules against the plasmepsin family of proteases encoded by *Plasmodium falciparum* the malaria causing protozoan parasite.

Besides the demonstration that a global eScience production infrastructure such as EGEE enables a new dimension of in silico experimentation in the area of computational life sciences we were aiming at identifying real new candidate inhibitor molecules that could be proposed for further drug development.

The first reports on WISDOM and the antimalaria virtual screening approach taken in this project elicited a strong response in the scientific community working on *Plasmodium falciparum* and antimalarial drugs. With the new class of potential inhibitors, the guanidino group of compounds, we hope to have established a new class of chemical entities with inhibitory activity against *Plasmodium falciparum* plasmepsins. A strong support for their putative activity is that most of the so far known antimalarial drugs likewise contain basic groups. The virtual screening approach taken by us could be subject to criticism as alternative strategies such as pharmacophore or similarity searches do exist. However, the fact that we were able to point to a new class of potential inhibitors after using a selection of publicly available “virtual” compounds (the ZINC database of compounds) and the fact that we could identify candidate inhibitors that fall into the already well-established inhibitor classes of thiourea and diphenyl urea analogues speak for the route we have taken. Future work will focus on expanding our collaboration of scientists interested in finding new cures for malaria, and, of course, we aim at involving experts from the pharmaceutical industry to make use of their expertise in drug development for further evaluation of the guanidino analogues.

WISDOM may serve as a template for more efforts to combat diseases of the poor driven by public and nonprofit research organizations. The majority of the EGEE infrastructure is financed by public sources, and based on this infrastructure follow-up projects aimed at finding new inhibitors, e.g., against Asian bird flu, have already been started. The challenge we face now is to expand the workflow from in silico screening to in silico confirmation of the docking results by molecular dynamics simulation and from this point to experimental validation in the drug development laboratory. Moreover, we have to turn this experimental workflow into a sustainable pipeline for the generation of new drug candidates against the major neglected diseases. Grid computing and eScience have the potential to bundle national and international funding efforts and to enable a new type of interaction between computer scientists, grid experts, cheminformatics and modeling experts, drug development experts, and researchers involved in epidemiological aspects of neglected diseases.

WISDOM project has successfully addressed the subject of reproducibility which is a key issue in distributed projects. By utilizing the same procedure, docking software, and chemical compounds from the ZINC database we successfully performed large scale virtual screening against four different proteins implicated in malaria (WISDOM-II).[50] Provided the appropriate software licenses and being a member of the EGEE by registering with any of the virtual organizations (for example BIOMED virtual organization to access to EGEE grid infrastructure) one can certainly reproduce the work. Our future works hold in reranking of the compounds by molecular dynamics and experimental testing of the best compounds.

ACKNOWLEDGMENT

EGEE is a project funded by the European Union under contract INFSO-RI-508833. Auvergrid is a project funded by Conseil Régional d'Auvergne. The authors thank particularly the Biomed Task Force members, the EIS team, the JRA2 members, Lydia Maigne, and Cheick O. Thiam. The authors acknowledge the support of BiosolveIT which made freely available up to 1000 FlexX licenses. The BioinfoGRID project is cofunded by the European Commission under contract INFSO-RI-026808. The following institutes contributed computing resources to the data challenge: ASCC (Taipei); IPP-BAS, IMBM-BAS, and IPP-ISTF (Bulgaria); CYFRONET (Poland); ICI (Romania); CEA-DAPNIA, CGG, IN2P3-CC, IN2P3-LAL, IN2P3-LAPP, and IN2P3-LPC (France); SCAI (Germany); INFN (Italy); NIKHEF, SARA, and Virtual Laboratory for e-

Science (Netherlands); IMPB RAS (Russia); UCY (Cyprus); AUTH FORTH-ICS and HELLASGRID (Greece); RBI (Croatia); TAU (Israel); CESGA, CIEMAT, CNB-UAM, IFCA, INTA, PIC, and UPV-GryCAP (Spain); BHAM, University of Bristol, IC, Lancaster University, MANHEP, University of Oxford, RAL, and University of Glasgow (United Kingdom).

Supporting Information Available: Plot comparing scores of the AutoDock and FlexX (Figure 1), histogram representation comparing scores of the AutoDock and FlexX (Figure 2), description of computational grids and EGEE (section 2), technical details of AutoDock docking setup (section 3), detailed description of ranking procedures (section 4), plot displaying score correlation between parameter sets 1 and 2 for 1lee (Figure 3), histogram plot of solution 1 for 1lee in parameter sets 1 and 2 (Figure 4), all statistics of target structures 1lee, 1lee_h2, and 1lee_h3 (Table 1), interesting compounds against plasmepsin in Table 2 (Tables 2−4), all the Software, python scripts, and Perl scripts used in this project and necessary to reproduce the work (appendixes). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Breton, V.; Jacq, N.; Kasam, V; Hofmann-Apitius, M. Grid added value to address malaria. *IEEE Trans. Inf. Technol. Biomed*. **2007**, in press.

(2) Malaria Vaccine Initiative. Bethesda, MD, U.S.A. http://www.malariavaccine.org (accessed Apr 7, 2005).

(3) Curing Malaria Together. Medicines for Malaria Venture. http://www.mmv.org (accessed Apr 7, 2005).

(4) Roll back Malaria Partnership. http://www.rbm.who.int (accessed Apr 7, 2005).

(5) Saffron Media Pvt Ltd. Mumbai, India. http://www.pharmabiz.com (accessed Apr 11, 2005).

(6) Gagliardi, F.; Jones, B.; Grey, F.; Bégin, M. E.; Heikkurinen, F. Building an infrastructure for scientific Grid computing: status and goals of the EGEE project. *Philos. Trans. R. Soc. London, Ser. A* **2005**, *363*, 1729−1742.

(7) Breman, J. G.; Alilio, M. S.; Mills, A. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. *Am. J. Trop. Med. Hyg.* **2004**, *71S*, 1−15.

(8) Malaria Site. All about malaria. http://www.malariasite.com (accessed Apr 15, 2005)

(9) Centers for Disease Control and Measures. http://www.cdc.gov (accessed Apr 22, 2005)

(10) Weisner, J.; Ortmann, R.; Jomaa, H.; Schlitzer, M. New Antimalarial Drugs. *Angew. Chem. Int. Ed.* **2003**, *42*, 5274 − 5293.

(11) White, N. J. Antimalarial drug resistance. *J. Clin. Invest.* **2004**, *113*, 1084−1092.

(12) Robert, A.; Benoit-Vical, F.; Dechy-Cabaret, O.; Meunier, B. From Classical Antimalarial Drugs to New Compounds Based on the Mechanism of Action. *Pure Appl. Chem.* **2001**, *73*, 1173−1188.

(13) Coombs, G. H.; Goldberg, D. E.; Klemba, M.; Berry, C.; Kay, J.; Mottram, J. C. Aspartic Proteases of Plasmodium falciparum and other parasitic protozoa as drug targets. *Trends Parasitol.* **2001**, *17*, 532−537.

(14) Francis, S. E.; Sullivan, D. J., Jr.; Goldberg, D. E. Hemoglobin Metabolism in the Malaria Parasite Plasmodium falciparum. *Annu. Rev. Microbiol.* **1997**, *51*, 97−123.

(15) Banerjee, R.; Liu, J.; Beatty, W.; Pelosof, L.; Klemba, M.; Goldberg, D. E. Four Plasmepsins are Active in the Plasmodium falciparum Food Vacuole, Including a Protease with an Active-site Histidine. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 990−995.

(16) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. Structure and Inhibition of Plasmepsin II, a Hemoglobin-Degrading Enzyme from Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034−10039.

(17) Jiang, S.; Prigge, S. T.; Wei, L.; Gao, Y. E.; Hudson, T. H; Gerena, L.; Dame, J. B.; Kyle, D. E. New Class of Small Nonpeptidyl Compounds Blocks Plasmodium falciparum Development In Vitro by Inhibiting Plasmepsins. *Antimicrob. Agents Chemother.* **2001**, *45*, 2577−2584.

(18) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. Potent Inhibitors of the Plasmodium falciparum Enzymes Plasmepsin I and II Devoid of Cathepsin D Inhibitory Activity. *J. Med. Chem.* **2004**, *47*, 110−122.

(19) Spencer, R. W. High Throughput Virtual Screening of Historic Collections on the File size, Biological targets and File diversity. *Biotechnol. Bioeng.* **1998**, *61*, 61−67.

(20) Gruneberg, S.; Wendt, B.; Klebe, K. Subnanomolar Inhibitors From Computer Screening: A Model Study Using Human Carbonic anhydrase II. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 389−393.

(21) Doman, T.; N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly. D. T.; Shoichet, B. K. Molecular Docking and High Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase 1B. *J. Med. Chem.* **2002**, *45*, 2213−2221.

(22) Baxter, C. A.; Murray, C. W.; Waszkowycz, B.; Li, J.; Sykes, R. A.; Bone, R. G. A.; Perkins, T. D. J.; Wylie, W. New Approach to Molecular Docking and its Application to Virtual Screening of Chemical Databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 254−262.

(23) Waszkowycz, B.; Perkins, T. D. J.; Sykes, R. A.; Li, J. Large Scale Virtual Screening for Discovering Leads in the Postgenomic era. *IBM Syst. J.* **2001**, *40*, 360−376.

(24) Buyya, R.; Branson, K.; Giddy, J.; Abramson, D. The Virtual Laboratory. A Toolset to Enable Distributed Molecular Modeling for Drug Design on the WorldWide Grid. *Concurrency Computat. Pract. Exper.* **2003**, *15*, 1−25.

(25) Jacq, N.; Salzemann, J.; Legré, Y.; Reichstadt, M.; Jacq, F.; Zimmermann, M.; Maass, M.; Sridhar, M.; Kasam, V. K.; Schwichtenberg, H.; Hofmann, M.; Breton, V. Demonstration of In Silico Docking at a Large Scale on Grid Infrastructure. *Stud. Health Technol. Informatics* **2006**, *120*, 155−157.

(26) Foster, I.; Kesselman, C. Computational Grids. In *The Grid: Blueprint for a New Computing Infrastructure*, 1st ed.; Penrose, D. E. M., Eds.; Morgan Kaufmann Publishers: San Fransisco, CA, U.S.A., 1999; pp 15−25.

(27) The Smallpox Reasearch Grid. http://grid.org/projects/smallpox/index.htm (accessed May 25, 2005).

(28) The Anthrax Research Project. http://grid.org/projects/anthrax/index.htm (accessed May 25, 2005).

(29) United Devices Cancer Research Project. http://grid.org/projects/cancer/index.htm (accessed May 25, 2005).

(30) Richards, W. G. Virtual Screening Grid Computing: The Screensaver Project. *Nat. Rev. Drug. Discovery* **2002**, *1*, 551−555.

(31) Mehlin, C. Structure based drug discovery for Plasmodium falciparum. *Comb. Chem. High Throughput Screening* **2005**, *8*, 5−14.

(32) Drug Targets for P. falciparum. Compiled by Yeh, I. http://plasmocyc.stanford.edu/target.html (accessed Mar 25, 2005).

(33) National Institute of Allergy and Infectious Diseases. National Institutes of Health. http://www.niaid.nih.gov (accessed Apr 28, 2005).

(34) Pattanaik, P.; Raman, J.; Balaram, H. Perspectives in drug design against malaria. *Curr. Top. Med. Chem.* **2002**, *2*, 483−505.

(35) The Brookhoven Protein Database. www.pdb.org (accessed Mar 25, 2005).

(36) Asojo, O. A.; Afonina, E.; Gulnik, S. V.; Yu, B.; Erickson, J. W.; Randad, R.; Medjahed, D.; Silva, A. M. Structures of Ser205 Mutant Plasmepsin II from Plasmodium falciparum at 1.8 Å in Complex with the Inhibitors rs367 and rs370. *Acta Crystallogr.* **2002**, *D58*, 2001−2008.

(37) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Majer, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. Structure and Inhibition of Plasmepsin II, a Hemoglobin-Degrading Enzyme from Plasmodium falciparum. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034−10039.

(38) A free database for virtual screening. ZINC is not commercial. UCSF, University of California: San Francisco, CA. http://blaster.docking.org/zinc/ (accessed Jun 25, 2005).

(39) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.

(40) *ADT Tools, Version 2.0*; Molecular Graphics Lab, The Scripps Research Institute: La Jolla, CA, 2005.

(41) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(42) *FlexX, Version 2.0*; BioSolveIT Gmbh: Sankt Augustin, Germany, 2005.

(43) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(44) *AutoDock, Version 3.0.5*; Molecular Graphics Laboratory, The Scripps Research Institute: LA Jolla, CA, 2005.

(45) Rarey, M.; Kramer, B.; Lengauer, T. The Particle Concept: Placing discrete Water Molecules During Protein-Ligand Docking Predictions. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 17−28.

(46) Herrera, F.; Lozano, M.; Molina, D. Continuous Scatter Search: An Analysis of the Integration of Some Combination Methods and Improvement Strategies. *Eur. J. Oper. Res.* **2006**, *169*, 450−476.

(47) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FlexX Incremental Construction Algorithm for Protein Ligand Docking. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 228−241.

(48) Jacq, N.; Breton, V.; Chen, H.-Y.; Ho, L.-Y.; Hofmann, M.; Lee, H.-C.; Legre, Y.; Lin, S. C.; Maaβ, A.; Medernach, E.; Merelli, I.; Milanesi, L.; Rastelli, G.; Reichstadt, M.; Salzemann, J.; Schwichtenberg, H.; Sridhar, M.; Kasam, V.; Wu, Y.-T.; Zimmermann, M. Virtual Screening on Large Scale Grids. *Parallel Comput.* **2007**, *33*, 289−301.

(49) Jacq, N.; Salzemann, J.; Jacq, F.; Legré, Y.; Medernach, E.; Montagnat, J.; Maaβ, A.; Reichstadt, M.; Schwichtenberg, H.; Sridhar, M.; Kasam, V.; Zimmermann, M.; Hofmann, M.; Breton, V. Grid-enabled Virtual Screening against malaria. *J. Grid Comput.* **2007**, in press.

(50) Kasam, V.; Salzemann, J.; Jacq, N.; Mass, A.; Breton, V. Large Scale Deployment of Molecular Docking Application on Computational Grid infrastructures for Combating Malaria. In *Cluster Computing and the Grid, 2007. CCGRID 2007.* Proceedings of the Seventh IEEE International Symposium, Rio de Janeiro, Brazil, May 14−17; IEEE Computer Society: Washington, DC, U.S.A., 2007; pp 691−700, 10.1109/CCGRID.2007.66.

CI600451T