

Mimicking direct protein–protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer: a pharmacophore-directed virtual screening approach

Victor Giménez-Oya · Óscar Villacañas ·
Xavier Fernández-Busquets · Jaime Rubio-Martinez ·
Santiago Imperial

Received: 15 September 2008 / Accepted: 2 December 2008 / Published online: 7 February 2009
© Springer-Verlag 2009

Abstract The 2C-methylerythritol 4-phosphate (MEP) pathway for the biosynthesis of isopentenyl pyrophosphate and its isomer dimethylallyl pyrophosphate, which are the precursors of isoprenoids, is present in plants, in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including pathogenic agents. However, the MEP pathway is absent from fungi and animals, which have exclusively the mevalonic acid pathway. Given the characteristics of the

MEP pathway, its enzymes represent potential targets for the generation of selective antibacterial, antimalarial and herbicidal molecules. We have focussed on the enzyme 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth reaction step of the MEP pathway. A molecular dynamics simulation was carried out on the CMK dimer complex, and protein–protein interactions analysed, considering also water-mediated interactions between monomers. In order to find small molecules that bind to CMK and disrupt dimer formation, interactions observed in the dynamics trajectory were used to model a pharmacophore used in database searches. Using an intensity-fading matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry approach, one compound was found to interact with CMK. The data presented here indicate that a virtual screening approach can be used to identify candidate molecules that disrupt the CMK–CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial, and herbicidal compounds.

V. Giménez-Oya and Ó. Villacañas contributed equally to this work.

V. Giménez-Oya · S. Imperial (✉)
Department de Bioquímica i Biologia Molecular,
Universitat de Barcelona,
Avda Diagonal 645,
08028 Barcelona, Spain
e-mail: simperial@ub.edu

Ó. Villacañas · J. Rubio-Martinez
Department de Química Física, Institut de Recerca en Química
Teòrica i Computacional (IQTCUB), Universitat de Barcelona,
Martí i Franquès 1,
08028 Barcelona, Spain

X. Fernández-Busquets
Biomolecular Interactions Team, Nanobioengineering Group,
Institute for Bioengineering of Catalonia,
Barcelona Science Park, Baldiri Reixac 10–12,
08028 Barcelona, Spain

X. Fernández-Busquets · S. Imperial
Nanoscience and Nanotechnology Institute,
Universitat de Barcelona,
Barcelona, Spain

Present address:

Ó. Villacañas
Intelligent Pharma,
S.L. Barcelona Science Park, Baldiri Reixac 10–12,
08028 Barcelona, Spain

Keywords CMK-MEP · MVA ·
Solvent-mediated interactions · Protein–protein interactions ·
Molecular dynamics · Drug design ·
Intensity-fading MALDI-TOF mass spectrometry

Introduction

Isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) [1, 2] are the universal five-carbon precursors of one of the largest families of natural products—the isoprenoids. Isoprenoids include hopane triterpenes, ubiquinones and menaquinones in bacteria; carotenoids, plastoquinones, mono-, sesqui-, di and tri- terpenes and the

prenyl side chains of chlorophylls in plants; and quinones, dolichols, steroids and retinoids in mammals [3], and account for more than 30,000 naturally occurring molecules of both primary and secondary metabolism.

After the discovery of the mevalonic acid (MVA) pathway in yeast and animals, it was assumed that IPP was synthesised from acetyl-CoA via MVA and then isomerised to DMAPP in all organisms [2–5]. However, an alternative, MVA-independent, route for the biosynthesis of IPP was recently identified by labelling experiments in bacteria [6, 7] and plants [8]. It is named the MEP pathway, after what is currently considered its first committed precursor, 2C-methylerythritol 4-phosphate (for recent reviews see [9] and [10]).

Experimental evidence has shown that most organisms only use one of the two pathways for the biosynthesis of their isoprenoid precursors. Thus, the MEP pathway is essential in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including the causal agents for diverse and serious human diseases like leprosy, bacterial meningitis, various gastrointestinal and sexually transmitted infections, tuberculosis, and certain types of pneumonia. The MEP pathway is absent in archaeobacteria, fungi and animals, which synthesise their isoprenoids exclusively through operation of the MVA pathway, whereas plants use both pathways for isoprenoid biosynthesis, although they are localised in different compartments [9–16].

Given the absence of the MEP pathway in mammals, its enzymes represent potential targets for the development of selective antibacterial, antimalarial and herbicidal compounds [10, 15–18]. An example of this new class of chemotherapeutic agents is fosmidomycin, a natural antibiotic originally isolated from *Streptomyces lavendulae*, which inhibits the second enzyme of the MEP pathway, deoxyxylulose 5-phosphate reductoisomerase [16]. Fosmidomycin and its derivatives have been shown to inhibit bacterial growth [19] and to be effective against *Plasmodium* infections [17, 20–22]. A combined therapy of fosmidomycin and clydamycin, a chemical that inhibits the prokaryote-like translation machinery of the *P. falciparum* apicoplast, proved to be highly efficient against malaria, with a 3-day regime resulting in an overall cure rate of 95% [23]. These results are very promising and validate the usefulness of MEP pathway enzymes as drug targets.

In this work we have studied the enzyme 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth step in the MEP pathway, i.e. the ATP-dependent phosphorylation of the intermediate CDP-ME at the C-2 hydroxyl group to yield CDP-MEP. The enzyme from *Escherichia coli* has been cloned, overexpressed, purified, and crystallised. X-ray structure studies showed that CMK is organised as a homodimer. The asymmetric unit consists of two subunits, A and B, of

approximate total mass 62 kDa, which assemble with C2 symmetry to form an extended homodimer, but only 4% of the total surface area of the two monomers is involved in dimer formation. Despite such a small area of interaction, a dimer is observed in gel filtration experiments that can be detected in matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) [24] and in analytical ultracentrifugation assays [25].

Disruption of protein–protein interactions through small molecules that target hotspots in the contact regions between protein surfaces [26] is an attractive therapeutic approach that can be very useful for drug design [27–30]. Thus, destabilising the CMK complex could represent a new strategy for the development of antibacterial and antimalarial agents. Although few direct protein–protein interactions are observed in the crystal structure of CMK, several water-mediated interactions seem to be important for dimer stabilisation. Solvent-mediated protein–protein interactions play an important role in molecular recognition [31–35], and are being increasingly considered when modelling protein complexes [36]. Here, we have carried out molecular dynamics (MD) simulations on the CMK dimer in order to gain insight into such protein–protein interactions. With the objective of studying the intermolecular interactions of the complex in more detail, both direct and solvent-mediated protein–protein interactions were monitored throughout the dynamics trajectory. A pharmacophore was designed based on the position of atoms involved in the interactions, and used in a database search for small molecules capable of disrupting the CMK dimer. The association between candidate compounds and CMK was empirically checked by intensity-fading MALDI-TOF MS [37, 38]. This approach provides a quick method for the detection of complex formation between the protein and a ligand based on the reduction of ion intensities of the mixture, compared to the intensity of both protein and ligand alone. One hit provided by the pharmacophore-guided database search turned out to decrease dimer formation.

Materials and methods

System design

All minimisations and MD simulations were performed with the AMBER 7 suite of programs [39] and all-atom *parm99* and *gaff* force fields [40, 41]. For organic small molecules, atom types were assigned with the antechamber module of AMBER7, and atomic charges were calculated following the BCC-AM1 method [42]. Starting coordinates of the CMK complex were downloaded from the Protein Data Bank (PDB) [24], identification number PDBid=1oj4. The complex consists of two CMK proteins (each monomer comprising the full sequence, from residue 1 to residue 283), two

molecules of adenosine 5'-(β , γ -imido)triphosphate (AMP-PNP), two molecules of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME), 192 water molecules and 1 chloride ion. The structure was resolved by X-ray diffraction at a resolution of 2.01 Å. Some side-chains presented two possible conformations. Because they were not involved in protein–protein close contacts, only the conformations marked in the PDB file as 'A' were considered further. Selenomethionine residues were replaced by methionines. Hydrogen atoms were placed where necessary with the LEaP module of AMBER7. The system was neutralised with Na⁺ ions (31 ion atoms were required) placed at the most negative electrostatic potential points following a grid-based box-shaped procedure for mapping electrostatic potential surface (as implemented in LEaP).

Molecular dynamics simulations

A first minimisation step was carried out in vacuo with a distance-dependent dielectric constant of 4r. Then, a rectangular box of TIP3P water [43] was added, with a minimum distance between the protein and the edges of the box along the Cartesian axis of 10 Å. Water molecules closer than 1.8 Å to protein/CDP-ME/AMP-PNP were removed. The system was then relaxed by a two-step procedure. First, water molecule positions were optimised and, second, the whole system was optimised. MD simulations were performed using the particle mesh Ewald (PME) summation method [44] to compute long-range electrostatic energy. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm [45]. The time step was set to 1 fs and the non-bonded pair list was updated every 25 ps, applying a cutoff value of 9 Å. Temperature and pressure regulation was simulated using Berendsen's algorithm [46], with time coupling constants of 1.0 and 0.2 ps, respectively, unless specified otherwise. The system was heated gradually to 300°K increasing 30°K each 10 ps, at constant volume. Then, a two-step pressure equilibration process followed. A temperature coupling constant of 1.0 ps was applied during the first 20 ps and increased to 2.0 ps for a further 40 ps. Up to this point, ions and protein atoms had been restrained under harmonic potentials with a force constant values of 1 and 5 kcal mol⁻¹ Å⁻¹, respectively. Next, restraints were reduced gradually for 60 ps. Finally, simulations were carried out in the NVT ensemble with no restraints. The last 1,000 ps with stable total energy were considered as the production run for further analysis.

Energy analysis

Lennard-Jones (LJ) and Coulombic energies between every side chain of the first monomer protein and the entire second monomer protein, and vice versa, were computed

for snapshots every picosecond throughout the 1,000 ps of production time with the ANAL module of AMBER 7. These calculations were also performed after adding water molecules to one of the interacting partners.

Solvent-mediated interaction analysis

Water molecules that established hydrogen bonds simultaneously with both proteins throughout the last 200 ps were analysed with a program developed by our group. Hydrogen bond distances were monitored and occupancy was defined as the percentage of time for which there existed at least one hydrogen bond to each monomer with a distance value lower than 2.5 Å.

Design of interaction models from MD trajectory

During the last 200 ps of production time, the positions of those atoms in one monomer that effectively interacted with its partner monomer were monitored. If several atoms contributed to the same effective interaction, the centre of masses of the mean positions was calculated and treated as a single interaction point. Maximum and minimum distances between every pair of points were used to construct the model for database searches, and mean positions of points as Cartesian coordinates were used for the docking procedure. Because all points must be referred to the same reference system, each snapshot was superimposed onto the backbone heavy atoms. An interaction feature was assigned to every point: hydrogen donor/acceptor, hydrophobic.

Database search and docking procedure

A selected interaction model was introduced as input for the compound search. Catalyst (Accelrys, San Diego, CA) software together with the Available Chemical Database (ACD), National Cancer Institute (NCI) database, ChemDiv database and Specs database were used. Matches were filtered and molecules with unusual bonds or with a molecular mass higher than 750 Da were not selected. Hits were saved as *mol2* format files containing all the conformations given by Catalyst. Each conformation was docked to the CMK monomer as follows. Pharmacophoric features were automatically assigned to the ligands, and distances between pharmacophoric points were calculated for all conformations. Next, for each conformation, we tested whether a combination that satisfied the interaction model existed and, in such cases, the ligand pharmacophoric points were superimposed on those of the interaction model, optimising the root mean square deviation (RMSD). At this point, the CMK monomer structure was introduced into the model to detect van der Waals (vdW) clashes between the ligand and the protein. In order to avoid bad

vdW contacts, a slight translational movement of the ligand was allowed. Atoms were not allowed to be closer than half of the sum of their vdW radii. If there were no forbidden vdW contacts, the dissociation constant was evaluated with X-Score [40].

A visual structure analysis was carried out on ligands with best docking, defined as those with best X-Score and RMSD consensus values. The compounds selected were finally classified according to their structure.

Preparation of recombinant CMK

Recombinant CMK from *E. coli* was prepared as previously described [47]. Briefly, M15[pREP4] *E. coli* cells were transformed with the pQE30-*ispE* construct containing the coding region of the *ispE* gene of *E. coli*. Transformed cells were grown at 37°C in LB medium supplemented with ampicillin (100 µg/mL) and kanamycin (50 µg/mL) until an OD₆₀₀ of 0.3–0.4 was reached. Induction was then performed with 0.4 mM IPTG for 3 h at 30°C. Cells were harvested by centrifugation (3,000 g, 10 min) and resuspended in 40 mM Tris–HCl buffer, pH 8, containing 0.5 mM EDTA, lysozyme (1 mg/mL) and complete-mini EDTA-free (Roche) (one tablet/10 mL buffer). Cells were incubated at 4°C for 20 min and sonicated (5×30 s bursts with 1 min cooling at 0°C). The extract was centrifuged at 11,000 g for 30 min at 4°C; the supernatant was recovered and supplemented with protamine sulphate (1.25 mg/mL), incubated at room temperature for 15 min, and then centrifuged at 13,000 g for 20 min at 4°C.

Hi-trap chelating columns (GE Healthcare, Little Chalfont, UK) were used for the purification of recombinant CMK. The resin was washed with 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, and 10 mM imidazole. His6-CMK was eluted with an imidazole gradient (10–500 mM) in 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl. The fractions containing His6-CMK were 95% pure as judged from SDS–polyacrylamide gel electrophoresis and Coomassie brilliant blue R-250 staining and appeared as a major band corresponding to an apparent molecular mass of 31 kDa. CMK preparations were desalted by gel filtration on disposable Sephadex G-25 columns (PD10) (GE Healthcare) equilibrated in 40 mM Tris–HCl buffer pH 8.0 containing 0.1M NaCl, distributed into aliquots, snap-frozen with liquid N₂ and stored at –80°C until use. Protein concentration was measured by the method of Bradford [48] using bovine serum albumin as a standard.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS was used to study the formation of a complex between one of the selected compounds in the database search, 2,4,6-trichloro-N-[5,6-

dimethyl-1-(2-methylbenzyl)-1H-1,3-benzimidazol-4-yl] benzenesulfonamide (Bionet Research, Key Organics, Camelford, UK), Compound 1 (C1), and purified CMK from *E. coli*. In these experiments, a direct protocol could be used because both C1 and CMK are soluble in the concentration range used. This protocol also requires that the complex remains intact in solution at least in part during the process of MALDI-TOF MS determination [37]. Samples were prepared as follows: a 1 mM C1 solution prepared in water was mixed (1:1, v:v) with different dilutions in the purification buffer of recombinant CMK ranging from 0.002 mM to 0.2 mM. Then, the samples were mixed (1:1, v:v) with a 10 mg/mL solution of α -ciano-4-hydroxycinnamic acid (CHCA) (Aldrich; Sigma-Aldrich, St. Louis, MO) matrix solution in acetonitrile/deionised water (1:1, v:v). One micro-litre of the mixture was overlaid onto the MALDI-TOF plate and dried using the dried droplet method [49]. MALDI-TOF mass spectra were recorded in a 4700 Proteomics Analyzer instrument (Applied Biosystems, Foster City, CA). Acquisition of mass spectra was performed in the MS reflector positive-ion mode. Typical parameters were set to source and grid voltages 20 and 14 kV, respectively, power laser from 5,200 to 5,800, signal/noise threshold 5, and a noise window width of 50.

Drawings

Figures depicting complex structures were constructed with Chimera [50] or VMD [51].

Results and discussion

Interactions deduced from X-ray crystal structure

The X-ray crystal structure of CMK [24] shows a homodimeric complex, where the two monomers are apparently positioned in an inverted position. The protein monomers interact with one another through two small interfaces: A and B (see Fig. 1), in which similar interactions are established. Residues of monomer 1 (chain A in PDB original file) found in interface A are the same as those of monomer 2 (chain B in PDB original file) in interface B, and vice versa. In the experimental structure of the complex, three direct protein–protein electrostatic interactions are detected. Two of them are observed in interface A, an intermolecular salt bridge between Arg²¹ and Asp⁸⁰ (hereafter, monomer 1 and monomer 2, respectively) and a hydrogen bond between Ala²²(carbonyl) and Gly⁸⁷(amine). While the former is also observed in interface B (the third protein–protein electrostatic interaction), the latter interaction is detected in interface B not as a direct protein–protein interaction but as a solvent-mediated interaction (Table 1),

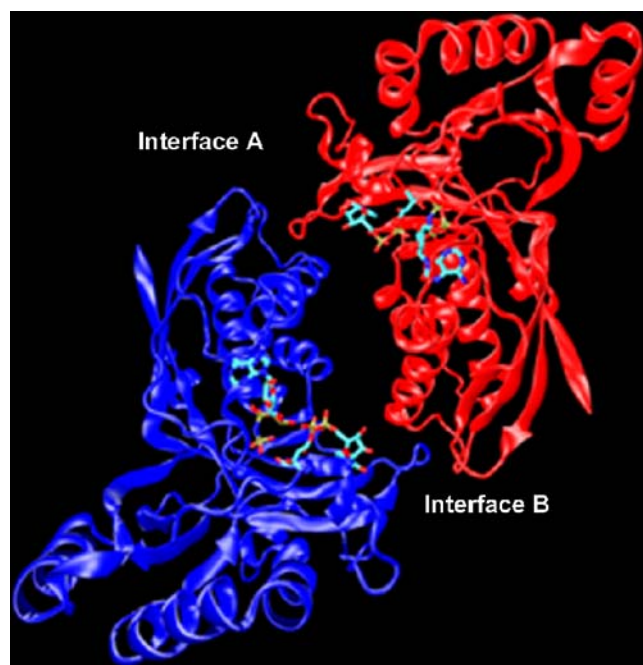


Fig. 1 X-ray structure of *Escherichia coli* 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK) dimer. Monomers are represented in cartoon mode; and adenosine 5'-(β , γ -imido) triphosphate (AMP-PNP) and 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME) molecules are in licorice mode as implemented in VMD visualisation software [51]. According to the reference used in the text, in the protein data bank (PDB) the first monomer is coloured blue and the second red

although with very distorted hydrogen bond angles between water molecule and protein atoms. A more relaxed geometry is found in the solvent-mediated interaction between Ser⁸⁸(carbonyl) and Ala²²(carbonyl) (Table 1). Table 1 lists all solvent-mediated hydrogen bonding networks between monomers.

Interactions from dynamics trajectory

Hydrogen bonds and salt bridges

The scenario obtained throughout the MD production time was slightly different from that suggested by the crystal structure. Our analysis revealed only one direct monomer–monomer electrostatic interaction, a salt bridge between Arg²¹ and Asp⁸⁰ in interaction interface A [mean minimum hydrogen bond distance was 1.81 Å, with a root mean square (RMS) of 0.12 Å]. Although the interaction in the other contact region (interaction interface B) was not of the protein–protein type, a more detailed study revealed that a water bridge joined both side chains through hydrogen bonds (Fig. 2). The hydrogen bond distances were analysed at every snapshot extracted from the last 200 ps of the production time, and the respective percentages of time with values below 2.5 Å (occupancy) are noted in Table 2. The analysis showed that a water bridge (always the same water molecule) between Asp⁸⁰ and Arg²¹ established hydrogen bonds during 96% and 99% of the time, respectively. As far as the interaction between Ala²² and Gly⁸⁷ backbones is concerned (detected in the crystal structure), it was observed as water-mediated in both interfaces. In interface A, a water molecule acted as the centre of a hydrogen bond network and established favourable interactions with Ala²²(carbonyl) of monomer 1 and the backbones of Gly⁴⁹, Pro⁸⁵, Gly⁸⁷, and Ser⁸⁸ of monomer 2. The occupancies of hydrogen bond distances were 83% and 97%, respectively (Table 2). In this latter case, the water molecule involved was labile and, therefore, different molecules were observed to play the role of mediating the interaction. However, the equivalent interaction in the other interface (B) was mediated by a water bridge, with respective hydrogen bond occupancy values of 65% and 99% (Table 2).

Table 1 Distances and angles of hydrogen bonds detected in the crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK). Distance 1 and angle 1 are calculated between water and residues of monomer 1, Distance 2 and angle 2 are calculated between water and residues of monomer 2. When a protein amino group is involved, both carbon-bonded atoms were used to calculate angles (carbon of the N-terminal side/carbon of C-terminal side)

Residues of monomer 1	Residues of monomer 2	Distance 1 (Å)	Angle 1 (degrees)	Distance 2 (Å)	Angle 2 (degrees)	Interface
Ala ²² (amine)	Thr ⁸⁶ (side chain)	3.3	83/87	3.4	117	A
Ala ²² (carbonyl)	Ser ⁸⁸ (carbonyl)	3.2	124	2.7	138	A
Gly ⁴⁹ (carbonyl)	Ala ²² (carbonyl)	2.8	121	3.1	131	B
Lys ⁷⁶ (side chain)	Leu ¹³⁶ (carbonyl)	2.9	103	3.0	157	B
Asp ⁸⁰ (side chain)	Glu ¹⁵⁸ (side chain)	3.5	117	3.4	116	B
Ser ⁸⁸ (amine)	Ala ²² (carbonyl)	3.2	93/102	2.1	141	B
Ser ⁸⁸ (carbonyl)	Ala ²² (carbonyl)	2.8	124	2.1	141	B
Gly ⁸⁷ (amine)	Ala ²² (carbonyl)	3.6	131/88	2.1	141	B
Gly ⁸⁷ (amine)	Ala ²² (carbonyl)	2.8	118/119	3.7	88	B

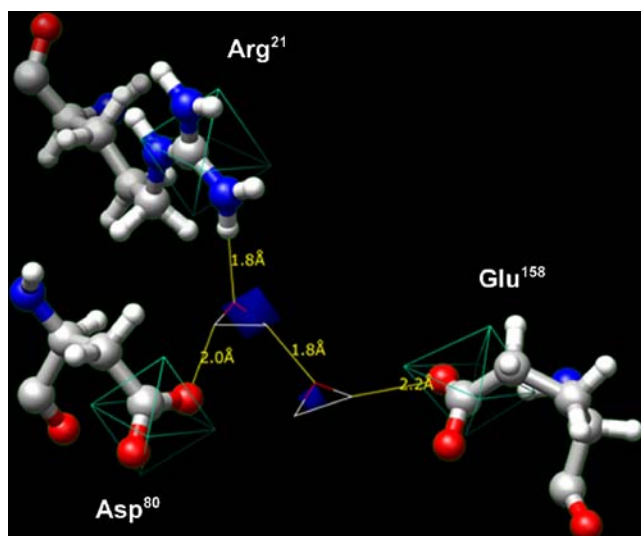


Fig. 2 Water-mediated interactions between Asp⁸⁰ of monomer 1 and Arg²¹ and Glu¹⁵⁸ of monomer 2 in interface B. Occupational analysis was carried out with Chimera software [50]. Grid volumes in which an oxygen atom of a water molecule is found at least 70% (approximately) of the time throughout the last 200 ps of the production time are coloured *solid blue*. Carbon atoms of carboxylic (Asp⁸⁰ and Glu¹⁵⁸) and guanidinium (Arg²¹) groups, and the oxygen atom of the water bridge between Asp⁸⁰ and Arg²¹, were used to superimpose snapshots before analysis. Cutoff and grid spacing values were set to 7 and 1.8 Å, respectively. For validation, volumes comprising the carbon atoms mentioned above 100% of the same time are outlined in *light blue*. As an example, the first snapshot is represented in the figure

Summarising, the intermolecular protein–protein hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein–protein interactions in the crystal were observed in the dynamics trajectory as mediated by the solvent through hydrogen bond networks.

The role of water molecules in mediating interactions between both monomers is also observed in the crystal structure. The CMK homodimer complex appears to overcome its low protein–protein interaction surface by water-mediated interactions. Further interactions of this kind were resolved from the trajectory and will be detailed in the following section.

Protein–protein solvent-mediated interactions

In addition to those described above, other stable water-mediated protein–protein interactions were found to be stable in the last 200 ps of production time (Table 2). In interaction interface A, three further water-mediated interactions were elucidated, two of which involving both the side chain and the carbonyl group of Asp²³ in monomer 1. The side chain established a solvent-mediated interaction with the Lys⁷⁶(side chain) of monomer 2, whereas the carbonyl group did the same with Gly⁴⁹(carbonyl) and Arg⁷²(side chain) of monomer 2. Finally, the side chain of Glu¹⁵⁸ in monomer 1 interacted with Lys⁷⁶(side chain) and Asp⁸⁰(side chain/carbonyl) of monomer 2 through hydrogen bonds established with water molecules. In interface B, one additional solvent-mediated interaction was observed between Ala⁷⁹(carbonyl) and Ala²²(amine). Similarities and differences between the crystal structure and the dynamics trajectory can be deduced by comparing values from Tables 1 and 2. The addition of temperature to the system changed the water-mediated monomer–monomer interaction pattern.

The interaction involving Glu¹⁵⁸ and Asp⁸⁰(side chain)/Lys⁷⁶(side chain) in interface A was not detected in interface B. However, some evidence led us to carry out a spatial occupancy analysis on the oxygen atoms of the water molecules (Chimera software [50]). The results revealed

Table 2 Percentage of hydrogen bond (HB) distances under 2.5 Å in protein–protein solvent-mediated interactions throughout the last 200 ps of the production time

Residues of monomer 1	Residues of monomer 2	HB 1 (%) ^a	HB 2 (%) ^b	Interface
Ala ²² (amine)	Ala ⁷⁹ (carbonyl)	19	94	A
Ala ²² (carbonyl)	Pro ⁸⁵ (carbonyl), Gly ⁸⁷ (amine), Ser ⁸⁸ (carbonyl/amine) and Gly ⁴⁹ (carbonyl)	83	97	A
Asp ²³ (side chain)	Lys ⁷⁶ (side chain)	98	97	A
Asp ²³ (carbonyl)	Gly ⁴⁹ (carbonyl), Arg ⁷² (side chain)	94	99	A
Glu ¹⁵⁸ (side chain)	Lys ⁷⁶ (side chain), Asp ⁸⁰ (carbonyl/side chain)	99	99	A
Pro ⁸⁵ (carbonyl), Thr ⁸⁶ (carbonyl), Ser ⁸⁸ (amine)	Ala ²² (carbonyl)	65	99	B
Asp ⁸⁰ (side chain)	Arg ²¹ (side chain)	96	99	B
Ala ⁷⁹ (carbonyl)	Ala ²² (amine)	62	97	B

^a Hydrogen bonds between water and monomer 1

^b Hydrogen bonds between water and monomer 2

that this interaction was described as mediated by two water molecules (Fig. 2), which most of the time filled optimal sites for establishing a hydrogen bond network between the residues. There was also evidence of a further example of a double water-mediated interaction pattern between Ala²²(amine) and Ala⁷⁹(carbonyl) in interface A. However, occupational analysis did not show a clear picture of two oxygen atoms located in optimal positions (data not shown). In this last case, a single water-mediated interaction pattern was observed 19% of the time. It is likely that both patterns exchanged with one another several times throughout the trajectory.

On the whole, the results extracted from dynamics analysis show the importance of water-mediated intermolecular interactions in the CMK dimer complex. It has been argued that bound water should be considered as an integral part of the tertiary structure of proteins [52], a proof of which is the conservation of buried water molecules in homologous proteins [53]. Here, in the CMK complex, most water molecules that were involved in intermolecular solvent-mediated interactions were highly labile. Because of the small monomer interaction surfaces, these water molecules are highly exposed, that is, they are not buried in protein cavities, and have high exchange rates. The existence of the hydrogen bond network was detected for most of the 200 ps time, suggesting that the role of water molecules is crucial for binding and recognition, although they are not tightly bound to the proteins. The computation of the protein–protein interaction surface throughout the 20 snapshots extracted from the production time (carried out with MolSurf as implemented in AMBER7), gave a value of 355 Å². When considering water molecules involved in solvent-mediated interactions as an integral part of monomers, the surface increased to 422 Å², a 19% increase. Thus, taking into account water molecules significantly raised the interaction surface.

Van der Waals interactions

Van der Waals (vdW) interactions were analysed in terms of energy, as described in **Materials and Methods**. Interaction vdW energy between every side chain of one monomer and the entire counterpart protein is depicted in Fig. 3. The results revealed two main groups of amino acid residues with favourable interaction energies; the first group is composed of Arg²¹ [with a mean LJ energy value, considering both cases, of -4.7 kcal mol⁻¹], Ala²² (-2.6 kcal mol⁻¹), Asp²³ (-4.7 kcal mol⁻¹), and Tyr²⁵ (-1.4 kcal mol⁻¹). The second group is formed by Met⁷⁵ (-1.9 kcal mol⁻¹), Lys⁷⁶ (-3.0 kcal mol⁻¹), Ala⁷⁹ (-2.3 kcal mol⁻¹), Leu⁸⁴ (-2.4 kcal mol⁻¹) and Thr⁸⁶ (-1.2 kcal mol⁻¹). In fact, these two groups interacted with one another when considered in different monomers.

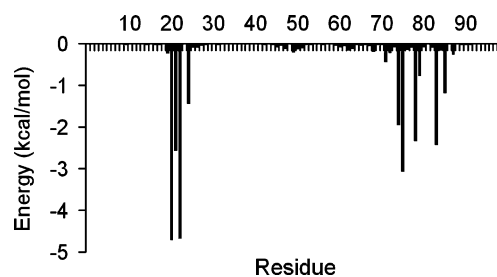


Fig. 3 Van der Waals (vdW) interaction energy between each individual side chain of one monomer and the entire partner protein. Mean values accounting for the same residue in both monomers are represented

Electrostatic interactions

Protein–protein electrostatic interaction energies throughout the production time were also analysed (Fig. 4a). Arginine and lysine residues had negative favourable values, while aspartic acid and glutamic acid residues had complex destabilising values. A single CMK monomer has 26 positively charged residues and 36 negatively charged residues. Therefore, one monomer has a high net negative charge that is translated into an unfavourable energy when analysing the electrostatic contribution of glutamic and aspartic acid residues. The net electrostatic interaction energy between monomers had a mean value of -27 kcal mol⁻¹ (100 snapshots were analysed), but it was strongly favoured by the presence of intervening water molecules—the mean electrostatic interaction energy was computed to be $-5,537$ kcal mol⁻¹ when water molecules were incorporated in the calculation. The inclusion of ions and/or CDP-ME/AMP-PNP did not change the effect of the water molecules (data not shown). In order to see the effect of the solvent on residue electrostatic contribution, water molecules were included in a further energy computation. As seen in Fig. 4b, the electrostatic energies of glutamic acid and aspartic acid residues decreased to negative values. Thus, water molecules are essential in explaining how protein–protein electrostatic interactions favour complex formation.

Database search for compounds disrupting protein–protein interactions

Atoms involved in protein–protein interactions were monitored throughout the last 200 ps of the dynamics trajectory to extract interaction models, as described in **Materials and Methods**. One of those models, depicted in Fig. 5, was introduced as a hypothesis in Catalyst (Accelrys). It is worth mentioning that one of the points of the model was deduced from a water molecule involved in a solvent-mediated interaction between monomers (between Asp²³ and Lys⁷⁶

Fig. 4 **a** Coulombic interaction energy between each individual side chain of one monomer and the entire partner protein. **b** *White bars* Same as **a** for glutamic acid and aspartic acid residues but including water molecules in the partner monomer part, *black bars* same as **a** for comparison. Mean values accounting for same residue in both monomers are represented

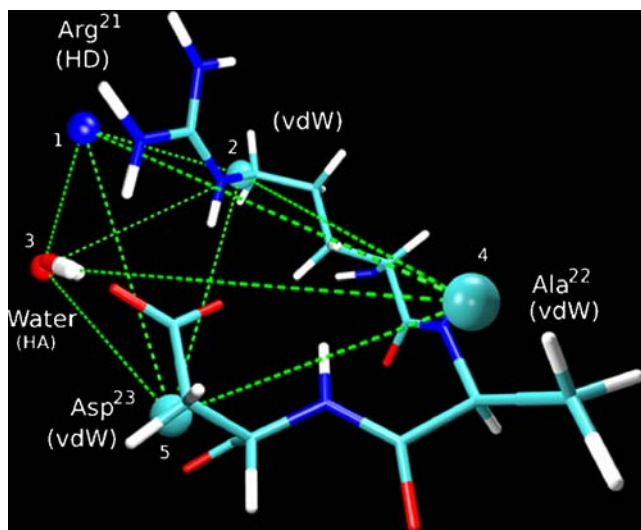
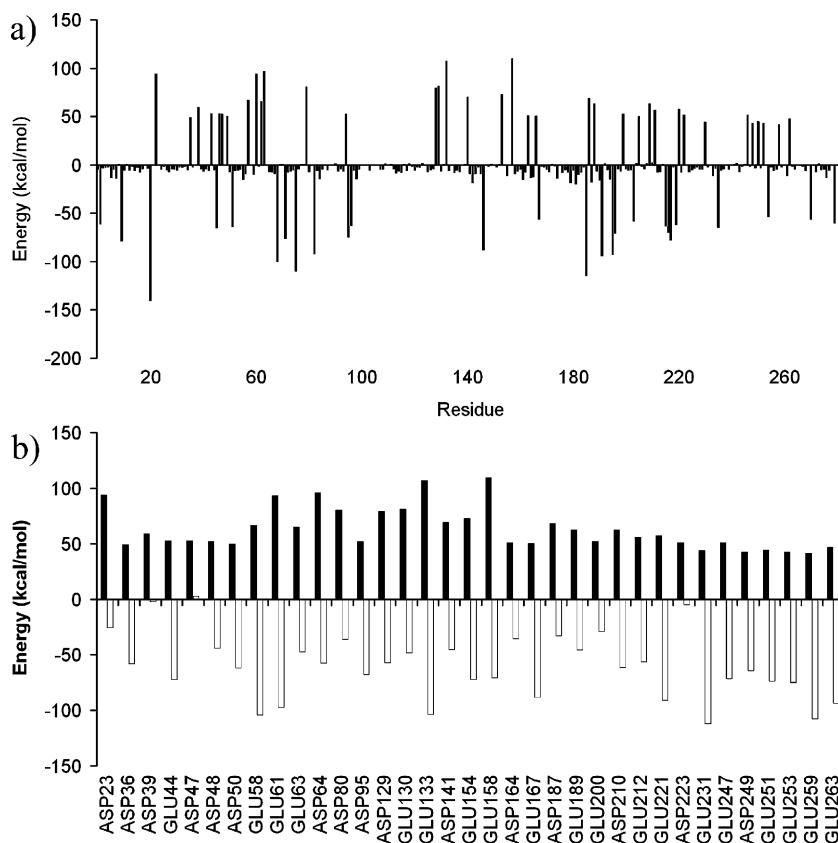


Fig. 5 Pharmacophoric points forming the interaction model used in the database search. Points are derived from residues 21–23 of monomer 2 and a water molecule. Arg²¹ provided a hydrogen donor (HD) and a vdW pharmacophoric point; Ala²² and Asp²³ provided vdW pharmacophoric points; the water molecule provided a hydrogen acceptor point. Distances between pair of points are: 1–2: 2.644–3.099 Å; 1–3: 3.004–4.837 Å; 1–4: 6.965–8.907 Å; 1–5: 3.982–4.795 Å; 2–3: 3.737–5.753 Å; 2–4: 5.958–7.235 Å; 2–5: 3.755–4.644 Å; 3–4: 8.715–11.321 Å; 3–5: 3.111–5.839 Å; 4–5: 5.354–6.696 Å

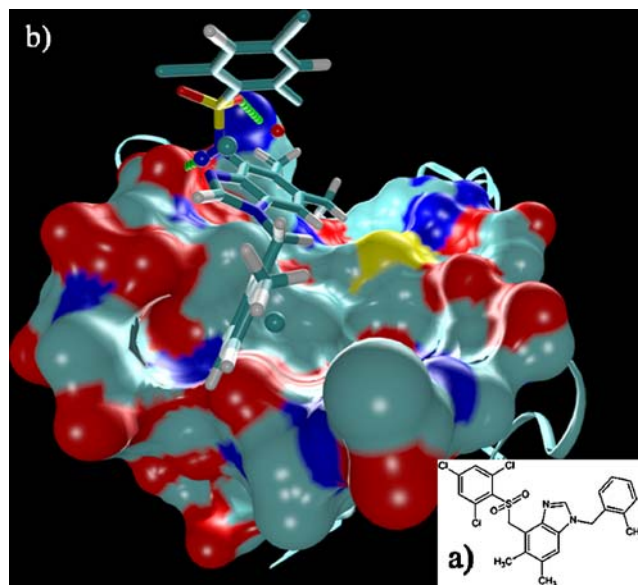


Fig. 6 **a** 2-D structure of compound 1. **b** Docking of C1 on CMK surface. CMK atoms closer than 5 Å were represented as a *molecular surface* using MSMS [51]. The remaining protein atoms were depicted following a *ribbon pattern*. Pharmacophoric points are represented as *spheres*

in interface A). Points marked *vdW* were modified to *hydrophobic* when defining the pharmacophoric hypothesis in Catalyst. The search procedure was carried out in the Available Chemical Database (ACD; 214 hits), Specs (512 hits) and ChemDiv databases (835 hits).

Hits were docked on CMK protein, as described in [Materials and Methods](#), and poses were evaluated with the scoring function X-Score [40].

Final selection of compounds

A visual structure analysis was carried out on the 15 ligands with best docking from each database. Ranking was based on X-Score consensus values and the final selection took into account the compound structure.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS is a rapid and sensitive approach for detecting the formation of soluble complexes between proteins and small molecules [38]. This methodology has the advantages of low cost and rapid performance together with the capability of checking the affinity properties of the analysed compounds. In this approach, detection of the complex is based on a reduction in the molecular ion intensities of the partners in a MALDI mass spectrum compared to the mass spectra of the partners alone. In our study, we tested complex formation between one of the final selected compounds (C1) and CMK. Docking of C1 onto CMK is shown in Fig. 6. RMSD after docking between pharmacophoric features on compound

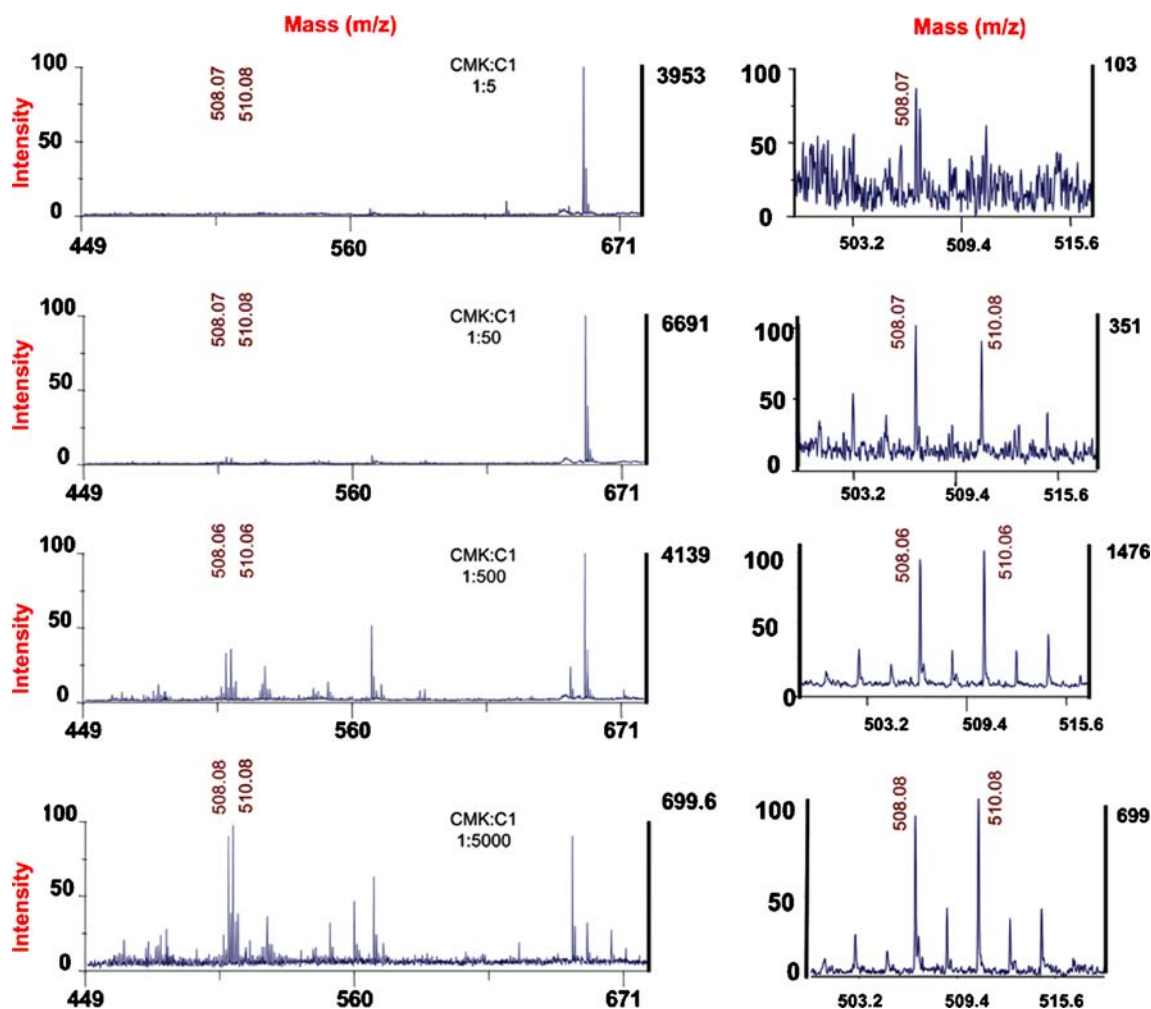


Fig. 7 Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) of the compound 1 (C1) molecule in the presence of increasing CMK concentrations. From *top to bottom*: 0.2, 0.02, 0.002, and 0.0002 mM CMK, respectively; *x-axis* mass(*m/z*), *y-axis* relative intensity (*left*) and ion intensity (*right*). Peaks at *m/z* 508

and 510 correspond to C1. A detailed view is shown on the *right panel*. Note that in order to visualize C1 *m/z* 508 and 510 peaks, spectra are presented at different intensity scales. The additional peaks correspond to adducts, corresponding to the α -cyano-4-hydroxycinnamic acid used as a matrix

and pharmacophore from Fig. 5 was 1.65 Å, with a pK_d estimated value (X-Score) of 5.43. Because the complex is formed between a small molecule and an enzyme, the reduction of the small-molecule ion intensity can be associated directly to formation of the complex.

The mass spectrum for C1 was acquired by adding 1 µL 1 mM C1 to 1 µL CHCA matrix and by depositing 1 µL of the mixture onto a MALDI-TOF plate. The MALDI-TOF mass spectrum of the C1+CHCA matrix showed two major peaks at m/z 508 and 510. The m/z 508 peak correlates with C1 molecular mass. In the subsequent series of experiments, 1 µL 1 mM C1 was mixed with different concentrations of CMK and the corresponding MALDI-TOF mass spectrum was acquired. As shown in Fig. 7, all C1+CMK+CHCA mass spectra obtained showed C1 peaks at m/z 508 and 510. The intensities of these peaks showed a reduction (fading) in the presence of CMK that was CMK concentration-dependent. At the highest concentration assayed (0.2 mM, Fig. 7, top), the intensity-fading of the m/z 508 and 510 peaks was almost complete, making the peak intensities similar to the baseline noise of the spectrum. These results demonstrate the formation of a complex between CMK and C1.

Conclusions

This paper describes a MD study carried out on a solvated dimer complex of CMK. The intermolecular hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein–protein interactions in the crystal structure were observed in the dynamics trajectory as being mediated by solvent through hydrogen bonds, a pattern already seen in the crystal structure of the complex. A deeper analysis of the solvent-mediated interactions indicates that four water-mediated protein–protein interactions remain along the MD trajectory, with one of them, in interface B, following a double-water motif. These “wet spots” [31], which likely are interfacial contact residues that interact through a water molecule, and double-water bonds, play an important role in protein–protein interaction between contact surfaces, as has been described in other protein models [32–34].

The study of intermolecular electrostatic interaction energies highlights the importance of water structure. Thus, the contribution of negatively charged residues was unfavourable when considering only protein atoms, but became instead stabilising when water molecules were included in the computation. These results indicate that water molecules are essential in explaining how protein–protein electrostatic interactions favour complex formation and, together with the observations of solvent-mediated

intermolecular interactions, reveal the importance of water for protein–protein interactions in the CMK dimer.

According to intermolecular interactions deduced from MD trajectory, some CMK residues could be hotspots for complex stabilisation, and thus useful targets for protein–protein interaction disruption. By monitoring the positions of Arg²¹, Ala²², Asp²³ and a water molecule during the MD trajectory, a pharmacophore model was constructed where the first three residues of monomer 1 interacted favourably with monomer 2. The water molecules established a hydrogen bond network between Asp²³ of monomer 1 and Lys⁷⁶ of monomer 2. Compounds that matched the pharmacophore model were searched for in the ACD, Specs and ChemDiv databases. After the docking procedure and visual analysis, four compounds were empirically tested and one of them exhibited binding to recombinant CMK in a concentration-dependent manner according to intensity-fading MALDI-TOF MS.

The data presented here indicate that a virtual screening approach can be used to identify candidate molecules that disrupt the CMK–CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial and herbicidal compounds.

Acknowledgements The excellent technical assistance of Drs. Eliandre de Oliveira and Maria Antonia Odena Caballol, from the Proteomics Platform, Barcelona Science Park, in the intensity-fading MALDI-TOF mass spectrometry experiments is gratefully acknowledged. This work was financed in part with grants from the University of Barcelona (ACES-UB), the Spanish Ministerio de Ciencia y Tecnología (CTQ2006-06588/BQU, BIO2002-04419-C02-02 and BIO2008-01184) and the Generalitat de Catalunya (2005SGR00914).

References

1. Sacchettini JC, Poulter CD (1997) *Science* 277(5333):1788–1789
2. Chappell J (1995) *Annu Rev Plant Physiol Plant Mol Biol* 46:521–547
3. McGarvey DJ, Croteau R (1995) *Plant Cell* 7(7):1015–1026
4. Croteau R, Kutchan T, Lewis N (2000) Natural products (secondary metabolites). In: Buchanan B, Grissem W, Jones R (eds) *Biochemistry and molecular biology of plants*. American Society of Plant Biologists, Rockville, MD, pp 1250–1268
5. Chappell J (2002) *Curr Opin Plant Biol* 5(2):151–157
6. Flesch G, Rohmer M (1988) *Eur J Biochem* 175(2):405–411
7. Rohmer M, Knani M, Simonin P, Sutter B, Sahn H (1993) *Biochem J* 295(Pt 2):517–524
8. Schwarz MK (1994) *Terpenbiosynthese in Ginkgo biloba: Eine überraschende Geschichte*. PhD Thesis. Eidgenössische Technische Hochschule, Zurich
9. Rodríguez-Concepción M, Boronat A (2002) *Plant Physiol* 130(3):1079–1089
10. Rohmer M, Grosdemange-Billiard C, Seemann M, Tritsch D (2004) *Curr Opin Investig Drugs* 5(2):154–162
11. Boucher Y, Doolittle WF (2000) *Mol Microbiol* 37(4):703–716
12. Lichtenthaler HK, Schwender J, Disch A, Rohmer M (1997) *FEBS Lett* 400:271–274

13. Eisenreich W, Schwarz M, Cartayrade A, Arigoni D, Zenk MH, Bacher A (1998) *Chem Biol* 5(9):R221–R233
14. Lichtenthaler HK (1999) *Annu Rev Plant Physiol Plant Mol Biol* 50:47–65
15. Rohmer M (1999) *Nat Prod Rep* 16(5):565–574 (Oct)
16. Kuzuyama T, Shimizu T, Takahashi S, Seto H (1998) *Tetrahedron Lett* 39(43):7913–7916
17. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D, Beck E (1999) *Science* 285(5433):1573–1576
18. Lichtenthaler HK (2000) *Biochem Soc Trans* 28:785–789
19. Testa CA, Brown MJ (2003) *Curr Pharm Biotechnol* 4(4):248–259
20. Wiesner J, Borrmann S, Jomaa H (2003) *Parasitol Res* 90(2):S71–S76
21. Missinou MA, Borrmann S, Schindler A, Issifou S, Adegnikaa AA, Matsiegui PB, Binder R, Lell B, Wiesner J, Baranek T, Jomaa H, Kreamsner PG (2002) *Lancet* 360(9349):1941–1942
22. Zeidler JG, Schwender J, Müller C, Wiesner J, Weidemeyer C, Beck E, Jomaa H, Lichtenthaler HK (1998) *Z Naturforsch* 53 (c):980–986
23. Borrmann S, Issifou S, Esser G, Adegnikaa AA, Ramharter M, Matsiegui PB, Oyakhirome S, Mawili-Mboumba DP, Missinou MA, Kun JF, Jomaa H, Kreamsner PG (2004) *J Infect Dis* 190 (9):1534–1540
24. Miallau L, Alphey MS, Kemp LE, Leonard GA, McSweeney SM, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2003) *Proc Natl Acad Sci USA* 100(16):9173–9178
25. Gabrielsen M, Bond CS, Hallyburton I, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2004) *J Biol Chem* 279 (50):52753–52761
26. Wells JA, McClendon CL (2007) *Nature* 450(7172):1001–1009
27. Sperandio O, Miteva MA, Segers K, Nicolaes GA, Villoutreix BO (2008) *Open Biochem J* 2:29–37
28. Smrcka AV, Lehmann DM, Dessal AL (2008) *Comb Chem High Throughput Screen* 11(5):382–395
29. Fletcher S, Hamilton AD (2007) *Curr Top Med Chem* (10):922–7
30. Whitty A, Kumaravel G (2006) *Nat Chem Biol* 2(3):112–118
31. Samsonov S, Teyra J, Pisabarro MT (2008) *Proteins* 73(2):515–525
32. Jiang L, Kuhlman B, Kortemme T, Baker D (2005) *Proteins* 58 (4):893–904
33. Furukawa Y, Morishima I (2001) *J Biol Chem* 276(16):12983–12990
34. Langhorst U, Backmann J, Loris R, Steyaert J (2000) *Biochemistry* 39(22):6586–6593
35. Janin J (1999) *Structure* 7(12):R277–R279
36. van Dijk AD, Bonvin AM (2006) *Bioinformatics* 22(19):2340–2347
37. Yanes O, Villanueva J, Querol E, Aviles FX (2005) *Mol Cell Proteomics* 4(10):1602–1613
38. Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX (2003) *Anal Chem* 75(14):3385–3395
39. Case DA, Pearlman DA, Caldwell JW, Cheatham TEI, Wang J, Ross WS, Simmerling C, Darden T, Merz KM, Stanton RV, Cheng A, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer R, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner P, Kollman PA (2002) AMBER7. University of California, San Francisco, CA
40. Wang R, Lai L, Wang S (2002) *J Comput Aided Mol Des* 16:11
41. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) *J Comput Chem* 25:1157
42. Jakalian A, Bush BL, Jack DB, Bayly CL (2000) *J Comp Chem* 21(2):132–146
43. Jorgensen WL, Chandrasekhar J, Madura JD (1983) *J Chem Phys* 79:926
44. Darden T, York D, Pedersen L (1993) *J Chem Phys* 98:10089
45. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) *J Comp Phys* 23:327
46. Berendsen HJC, Postman JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* 81:3684
47. Bernal C, Mendez E, Terencio J, Boronat A, Imperial S (2005) *Anal Biochem* 340(2):245–251
48. Bradford MM (1976) *Anal Biochem* 72:248–254
49. Karas M, Hillenkamp F (1988) *Anal Chem* 60(20):2299–2301
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) *J Comput Chem* 25(13):1605–1612
51. Humphrey W, Dalke A, Schulten K (1996) *J Mol Graph* 14(1):33–38,27–38
52. Levy Y, Onuchic JN (2006) *Annu Rev Biophys Biomol Struct* 35:389–415
53. Sreenivasan U, Axelsen PH (1992) *Biochemistry* 31(51):12785–12791