J Comput Aided Mol Des (2011) 25:371–393
DOI 10.1007/s10822-011-9427-z

Molecular dynamics and docking simulations as a proof of high flexibility in *E. coli* FabH and its relevance for accurate inhibitor modeling

Yunierkis Pérez-Castillo · Matheus Froeyen · Miguel Ángel Cabrera-Pérez · Ann Nowé

Received: 26 January 2011 / Accepted: 9 April 2011 / Published online: 23 April 2011

Abstract Bacterial β-ketoacyl-acyl carrier protein synthase III (FabH) has become an attractive target for the development of new antibacterial agents which can overcome the increased resistance of these pathogens to antibiotics in clinical use. Despite several efforts have been dedicated to find inhibitors for this enzyme, it is not a straightforward task, mainly due its high flexibility which makes difficult the structure-based design of FabH inhibitors. Here, we present for the first time a Molecular Dynamics (MD) study of the *E. colil* FabH enzyme to explore its conformational space. We compare the flexibility of this enzyme for the unliganded protein and an enzyme-inhibitor complex and find a correspondence between our modeling results and the experimental evidence previously reported for this enzyme. Furthermore, through a 100 ns MD simulation of the unliganded enzyme we extract useful information related to the concerted motions that take place along the principal components of displacement. We also establish a relation between the presence of water molecules in the oxyanion hole with the conformational stability of structural important loops. Representative conformations of the binding pocket along the whole trajectory of the unliganded protein are selected through cluster analysis and we find that they contain a conformational diversity which is not provided by the X-ray structures of ecFabH. As a proof of this last hypothesis, we use a set of 10 FabH inhibitors and show that they cannot be correctly modeled in any available X-ray structure, while by using our set of conformations extracted from the MD simulations, this task can be accomplish. Finally, we show the ability of short MD simulations for the refinement of the docking binding poses and for MM-PBSA calculations to predict stable protein-inhibitor complexes in this enzyme.

Keywords FabH · Molecular dynamics · Molecular docking · Multiple receptor conformations docking

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9427-z) contains supplementary material, which is available to authorized users.

Y. Pérez-Castillo (✉) · M. Froeyen
Laboratory for Medicinal Chemistry, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium
e-mail: yunierkis@gmail.com

Y. Pérez-Castillo · M. Á. Cabrera-Pérez
Molecular Simulations and Drug Design Department, Centro de Bioactivos Químicos, Universidad Central de Las Villas, Carretera a Camajuaní km 5 ½, Santa Clara, Cuba

Y. Pérez-Castillo · A. Nowé
Computational Modeling Lab (CoMo), Department of Computer Sciences, Faculty of Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

M. Á. Cabrera-Pérez
Engineering Department, Pharmacy and Pharmaceutical Technology Area, Faculty of Pharmacy, University Miguel Hernandez, 03550 Alicante, Spain

Introduction

Bacterial diseases have not yet been overcome. More than one-third of the world population is likely infected by bacterial pathogens and two million fatalities occur per year from bacterial infections [1]. As a direct consequence of antibiotics overuse, resistance to them has increased over the past two decades with almost every human pathogen and every class of antimicrobials in clinical use acquiring resistance [2–4]. This situation becomes more critical if we take into account that only three new classes

of antibiotics have been introduced during the last 40 years: oxazolidinones, cyclic lipopeptides and mutilins [5–9]. Furthermore, all classes of antibiotics currently in use are directed against a small subset of essential bacterial targets. As a consequence, new and innovative strategies are necessary for the development of new broad-spectrum antibacterial agents acting against a wider range of molecular targets essentials for bacteria survival.

One of the most attractive biochemical pathways to be used as the target for new antibacterial agents is fatty acid biosynthesis (FAS). This pathway has been demonstrated to be essential for the bacterial survival [10], and differs considerably from the human FAS pathway. While in humans fatty acid synthesis occurs via a homodimeric multifunctional enzyme [11, 12], in bacteria the pathway is composed of various discrete enzymes and each one can be considered a putative molecular target. These features make the type II FAS pathway a potential target for new antimicrobial agents [13].

A key enzyme in this pathway is the β-ketoacyl-acyl carrier protein synthase III (FabH), which is the one responsible for the first pathway reaction and plays an important regulatory role [14, 15]. FabH has also been demonstrated to be essential for organism survival and it is present in a wide number of important human pathogens [14, 16–22]. At the same time, FabH has no close homologue in humans and its selectivity for acetyl-CoA over acyl-ACP diverges from the other condensing enzymes involved in fatty acid synthesis in bacteria [14, 23]. Furthermore, some chemical compounds have been shown to inhibit FabH from diverse microorganisms, including multi-drug resistant strains [18, 24–35]. These facts, along with the high conservation in the sequence of the enzyme among several microorganisms, support the idea that FabH can be used as an effective molecular target for the development of new broad-spectrum antimicrobial agents [36].

On the other hand, the rapid development of structural elucidation techniques such as X-ray crystallography and NMR, as well as the increase in computational power that allows a more accurate and reliable modeling of molecular process, has made the structure-based molecular modeling methodologies a key component of the drug development process. Among the structure-based design tools, molecular dynamics and molecular docking are widely used. At the same time that molecular docking is applied for virtual screening campaigns and the prediction of the binding mode of ligands to receptors, molecular dynamics is becoming an essential tool that can be used for the preparation of the protein receptor before docking, to optimize its structure and account for protein flexibility; for the refinement of docked complexes, to include solvent effects and account for induced fit; to calculate binding free energies and to provide an accurate ranking of the potential ligands. Several examples are reported in literature of successful application of the combinations of these two types of techniques to drug design and have been reviewed elsewhere [37–40].

In this sense, the availability of different crystal structures of FabH [25, 35, 41–47] make the structure-based design of new inhibitors for this enzyme now possible. Nevertheless, it is not a straightforward task, mainly due to the observed high flexibility of this enzyme and the limitations of X-ray crystallography to provide information about the dynamic evolution of a protein. There are several experimental results supporting that this enzyme shows high flexibility and that it can exist as an ensemble of ordered and catalytically impaired conformations. In this sense, it was observed in a crystal structure of *E. coli* FabH (ecFabH) solved by Qiu et al. [41] that four loops involved in the formation of the dimer interface were highly disordered and could not be solved, while Al-hamadsheh et al. [44] proposed a catalytic model, based on kinetic and crystallographic studies, in which a structurally disordered apo-ecFabH dimer orders on binding of the substrate or inhibitor to one monomer and afterwards the same binding process takes place in the second subunit at a much slower rate. Such catalytic or inhibition processes where large conformational changes take place, can't be modeled based on a single structure and often the study of the time dependent evolution of the system is essential. In this sense, molecular dynamics simulations have been shown to be useful for obtaining a time-dependent description of macromolecular structure and can be used to overcome the limitations associated with the lack of flexibility information in the X-ray solved structures [48–50].

Here we present, for the first time, a molecular dynamics study of the *E. coli* FabH enzyme and the use of protein conformations extracted from the simulation for the accurate prediction of the binding mode of a series of benzoylaminobenzoic acid inhibitors. Molecular dynamics simulations are accomplished for two systems, the unliganded enzyme and the complex of the enzyme with a dichlorobenzyloxy-indole-carboxylic acid inhibitor. Using the data provided by the simulations, we compare the evolution of both systems and analyze, among other factors, protein flexibility and conformational changes, mainly in the binding pocket, that can considerably influence the topology of the binding pocket and hence the binding or not of inhibitors to the enzyme. After demonstrating that the apo-enzyme exhibits higher flexibility than the inhibitor-protein complex, we extend the MD simulations of the first system up to 100 ns and we use Principal Components Analysis to examine which concerted motions are taking place in the enzyme along the main axes of motion. Taking

into account the information derived from the PCA, we implement a methodology for clustering the matrix of correlations in atomic positions using functional criteria with the aim of investigating whether the correlated motions along the major eigenvectors also take place when all degrees of freedom of the system are considered.

We also carry out a clustering procedure to select representative conformations of the binding pocket and study their usefulness to correctly predict the binding mode of a series of benzoylaminobenzoic acid inhibitors which can't be modeled in any of the available ecFabH crystal structures available to date as we show. Furthermore, we propose a molecular docking and rescoring procedure capable of discriminating between the correct and wrong poses of these inhibitors in the previously extracted receptor conformations. Finally, the obtained complexes from the docking procedure are subject to a short (1 ns) molecular dynamics simulation to refine their structures, account for induced fit and evaluate their stability through the calculation of the binding free energy.

Materials and methods

Molecular dynamics simulations

For molecular dynamics simulations, the crystal structure of the ecFabH in complex with a dichlorobenzyloxy-indole-carboxylic acid inhibitor (pdb code 1MZS) was selected as the starting structure since it is the only crystal structure of the ecFabH solved in complex with an inhibitor [35]. The AMBER 10 suite was used for all molecular dynamics simulations in this study and the AMBER ff03.r1 force field was used for the protein parameterization [51]. For ligands, hydrogen atoms and AM1-BCC charges were added using ANTECHAMBER. The PARMCHECK program was used to obtain GAFF force field parameters for the ligands.

The systems were solvated inside a truncated octahedral box using TIP3P water molecules, a buffering distance to the edges of the box of 10 Å was used. The net negative charge of the apo-protein and ligand–protein complexes was neutralized by the addition of Na+ ions. All these steps were carried out using the LEAP module of the AMBER suite.

Afterwards, solvent was relaxed by means of a minimization of the system with protein and ligand restrained for 200 steps using a steepest descent algorithm. This step is aimed to eliminate close contacts between the solvent and the solute. This step was followed by another similar minimization of the system as a whole without restrained atoms to eliminate possible remaining bad contacts in the system.

The next step, the first equilibration phase, consisted of system heating from 100 to 300 K over 10 ps of simulation time at constant volume using periodic boundary conditions and no pressure control, only bonds involving hydrogen atoms were restrained using the SHAKE algorithm. The final equilibration step aimed to equilibrate the density in the system through a 10 ps simulation at constant pressure and isotropic position scaling. Further equilibration of the system for 1 ns was also carried out using the production run parameters.

Production MD simulations were run in the NTP ensemble at 300 K. Pressure was regulated using isotropic position scaling and a regulation time of 2 ps was choosen. Only bonds involving hydrogen atoms were constrained using the SHAKE algorithm. MD trajectories post-processing which include RMSD, temperature factors, correlation and covariance matrixes calculations were carried out using the PTRAJ module of the AMBERTOOLS suite.

To keep track of the presence of water molecules in the oxyanion hole, we first examined the available ordered X-ray structures of the enzyme and found that a water molecule or acetyl group could be located at distances ranging from 3 to 4 Å relative to the amide nitrogens of G306 and C112. Afterwards, we used the watershell command in PTRAJ to calculate how many water molecules are found at a distance lower than 4 Å from these two atoms separately ($N_{wat}(G)$ and $N_{wat}(C)$) and considering both atoms together ($N_{wat}(G,C)$). The number of water molecules in the proximity of the oxyanion hole was calculated as $N_{wat}(OXY) = N_{wat}(C) + N_{wat}(G) - N_{wat}(G,C)$ for all the MD snapshots under analysis. Next, we calculated the cumulative occupancy of the water molecules in the oxyanion hole for each snapshot n as the number of snapshots containing a water molecule in the oxyanion hole of either the A or B subunit divided by n.

Correlation analysis and functional-based clustering of the correlation matrix

Principal Components Analysis (PCA) of MD trajectories is based on the diagonalization of the positional covariance matrix. For a system with N atoms, the covariance matrix consists of 3 N × 3 N elements and this diagonalization process leads to an orthogonal set of unitary vectors describing the directions of maximum variation in the observed conformational distribution. Each eigenvector is associated with it eigenvalue that determines the amplitude of the motions along each principal axis of motion. The sum of all 3 N eigenvalues can be considered as the total conformational variance of the system and then the relative contribution of each eigenvector to the overall system flexibility can be calculated as the ratio between its associated eigenvalue and the total system flexibility. Furthermore, the

MD trajectory can be projected onto these eigenvectors and the concerted motions along each one can be visualized. In this work, we calculated the covariance matrix of the 634 alpha carbons in the enzyme and it was diagonalized in order to obtain the eigenvectors and eigenvalues describing the motion of the system. The visualization of the motions along each eigenvector was accomplished with the VMD molecular package and the IED plugin [52, 53]. We also examined the matrix of correlations in atomic positions. In this matrix a value between −1 and 1 is assigned to each pair of atoms under consideration that indicates whether they are completely correlated (1), anticorrelated (−1) or no correlation exists (0). The information provided in this matrix can be useful for defining regions of the protein that move in the same direction along the whole MD simulation in the cartesian space. Since the visual inspection of this matrix can only reveal limited information, we developed a methodology for evaluating the degree of correlation between structural and functional relevant regions of the enzyme. This methodology is based on the clustering of this matrix and the whole process consists of:

- Calculate the correlation matrix for the N alpha carbons of the enzyme using PTRAJ.
- Transform the correlation matrix to a distance matrix using the transformation:

$$D(i,j) = 1 - \mathrm{abs}(C(i,j))$$

- Cluster the distance matrix into a number of clusters going from 1 to N.
- Define functional relevant regions of the protein.
- For each pair of regions, find a cluster containing at least one residue of each one. This search is carried out moving from the leaves of the clustering tree to the root until a partition of the data containing at least one residue of each region is found. If more than one cluster is found to meet this condition at a certain level of the tree, then the cluster with the larger number of members is selected.
- Examine the clustering tree to find the distance at which the selected cluster is formed.

The methodology was implemented in MATLAB [54].

Ligand preparation for docking

For docking studies, we selected the ten benzoylaminobenzoic acid inhibitors published by Nie et al. [55] that were assayed against *E. coli*. Three dimensional conformers for the compounds were generated using the OMEGA software. A maximum of 50,000 conformations per molecule were generated using an energy window of 10 kcal/mol. All rotatable bonds were considered during the torsion search using the Merck Molecular Force Field (MMFF) and duplicate conformers were discarded based on a RMS value of 0.5 Å. A maximum number of 300 conformers were saved for each compound. Afterwards, AM1-BCC charges were added to each conformer using the MOLCHARGE programs that is part of the QUACPAC package [56].

Molecular docking protocol

Molecular docking calculations were performed for all the receptor conformations selected using the clustering procedure of the molecular dynamics trajectory following the same protocol. To avoid differences in the setup of the systems and their preparation, all molecular docking steps that involve receptor preparation, grid generation and docking were automated using a perl script.

Rigid docking was carried out for each ligand conformer using the DOCK v. 6.3 package [57]. A maximum of 5,000 orientations per ligand was explored allowing a maximum of two bumps between the ligand and the receptor. Bumps were defined as any pair of atoms closer than the 80% of the sum of their Van der Waals radii. The energy grid-based scoring function was selected for poses quality evaluation. The five poses with the lowest score for each ligand conformer were saved, allowing for a maximum of 1,500 saved poses for each of the ten compounds.

For each receptor conformation, all poses of the ligands lying inside the cavity and having a grid-based score lower than zero were selected for analysis and further rescoring. Selection of such ligand conformations was automatically carried out by selecting every pose having at least one atom at a distance lower than 3 Å of the catalytic residue HIS244 at the bottom of the binding channel. For a more accurate treatment of the electrostatic interactions, the poses selected in the previous step were rescored using the PBSA scoring function implemented in DOCK that uses the ZAP library for solving the Poisson–Boltzmann equation. Using this approach we can explore, compare and evaluate the binding of all the compounds in the receptor conformations extracted from the molecular dynamics simulation.

Molecular graphics images were produced using the UCSF Chimera package [58].

MM-PBSA/GBSA calculations

For the final refinement and MM-PBSA/GBSA calculations of the predicted complexes, 1 ns MD simulations were run for each molecular complex. The last 400 snapshots corresponding to 200 ps of simulation time were selected for MM-PBSA/GBSA calculations and clustering analysis.

The MM-PBSA/GBSA approach is used as the method for evaluating in a more accurate way the free energy of binding which can be calculated using the equation:

$$\Delta G_{bind} = G_{com} - (G_{rec} + G_{lig})$$

where com, rec, and lig stand for complex, receptor, and ligand, respectively. The free energy of each of these was estimated as a sum of the four terms:

$$G = \langle E_{MM} \rangle + \langle G_{psolv} \rangle + \langle G_{npsol} \rangle - T\langle S \rangle$$

where E_{MM} is the molecular mechanics component, G_{psol} is the polar contribution to the solvation energy, G_{npsol} is the nonpolar solvation energy, T the temperature and S the entropy. The calculation of the polar solvation free energies was carried out solving the Poisson-Boltzman equation using the *pbsa* program and the generalized Born (GB) approach implemented in AMBER [59, 60].

For PB and GB calculations an internal dielectric constant of 1 and an external of 80 were used. The solvent probe radius was set to 1.4 Å and the radii for PB calculations were taken from the topology files. The nonpolar solvation free energy was calculated as $0.0072\ kcal \cdot mol^{-1}\ \mathring{A}^{-2} * SASA$. A total of 1,000 iterations were run for the linear PB equation. For calculating the free energies of binding using the GB approach, an Onufriev's model was used [61].

Results and discussion

10 ns MD simulation of Apo and Holo FabH

The first part of our research aims to explore the flexibility of the ecFabH enzyme. For this purpose, we selected two systems: an unliganded enzyme and a complex of it with a dichlorobenzyloxy-indole-carboxylic acid inhibitor. The starting protein conformation for both simulations was the same (pdb code 1MZS). System preparation was carried out as shown in the "Materials and methods" section. The production MD simulation for each system was run for 10 ns after system equilibration stages.

The first experimental evidence about the high flexibility of the FabH enzyme was reported by Qiu et al. [41] through a crystal structure of the ecFabH where four loop regions constituting important parts of the dimer interface and also the substrate binding channel is disordered and the functional groups of the active site are perturbed relative to their positions in the ligand complexes, thus suggesting a catalytically impaired conformation for the protein. In Fig. 1 the structure of the ecFabH enzyme is represented. The dichlorobenzyloxy-indole-carboxylic acid inhibitor in the binding pocket of subunit A is represented as spheres while no ligand is present in the B subunit. The four loops disordered in one of the structure of Qiu et al. are highlighted using thicker ribbons in both monomers and they include residues 84–86, 146–152, 185–217 and 305–307. As can be seen from Fig. 1, such loops on each monomer make extensive contacts in the dimer interface and also are part of the binding pocket. Aside these loops, two helices that include residues forming the binding pocket show high flexibility and some side chains in these regions are disordered in this ecFabH structure: Ca1 and Ca2 which are also indicated in Fig. 1 (for a complete nomenclature of protein secondary structure see [41, 43]).

Based on kinetic studies developed by Alhamadsheh et al. [44], it has also been demonstrated that FabH catalytic and inhibition mechanisms can involve transitions from disordered to ordered conformations. One of the most interesting results derived from such studies is that the binding of an inhibitor in one subunit of the dimer can stabilize the conformation of the second active site making ligand or substrate binding to occur at a much slower rate on it. Taking these findings into account, we built our protein-inhibitor complex system with the ligand occupying only one of the catalytic pockets since, in this way, we can compare the flexibility of both monomers in the complex as well as this system and the unliganded one. Main conclusions that can be derived from these available experimental data of FabH are that it exhibits high flexibility and that there exists some kind of asymmetry between the functioning of each monomer.

In Fig. 2 is shown the backbone RMSD for both systems during the 10 ns of each simulation relative to the starting structure which is the same for both simulations. As can be

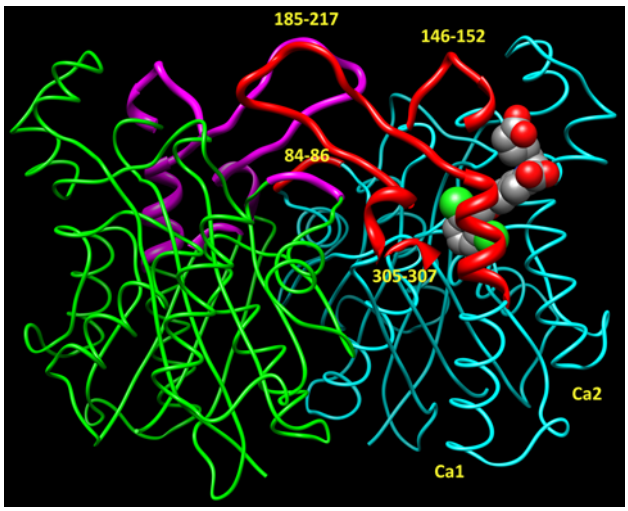

Fig. 1 Overall structure of the ecFabH enzyme. The dichlorobenzyloxy-indole-carboxylic acid inhibitor in the binding pocket of subunit A is represented as spheres while no ligand is present in the B subunit. The four loops disordered in one of the structure of Qiu et al. are highlighted using thicker ribbons in both chains

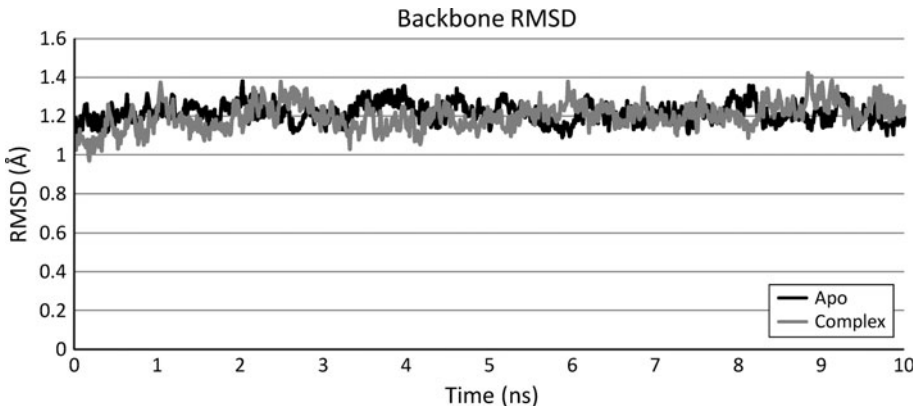

Fig. 2 Backbone's RMSD during the 10 ns simulation for the unliganded (*purple*) and complex (*red*) systems

seen, both trajectories are stable during the 10 ns production run with RMSD values oscillating between 1 and 1.4 Å after the initial equilibration stages. Although the behavior of the RMSD of the protein backbone in the two systems is similar, there are differences, which are remarkable in some cases, in the flexibility between both simulations as shown by the values of the temperature factors calculated for both MD trajectories. The temperature factors for the backbone of both subunits in the two systems as well as of the starting crystallographic structure are represented in Fig. 3.

From Fig. 3 we can observe that the behavior of the temperature factors for all the systems is similar. The same regions are among the most flexible on each system and correspond to the most flexible zones in the experimental X-ray structure. The values of the backbone temperature factor calculated from the MD trajectories as well as those from the starting X-ray structure are reported in the Supporting Information Section (TS1)**.** It can be observed that there exist peaks of the temperature factor associated to specific regions in both systems. Among these regions we can identify those including residues 47–48, 101–102, 147–152, 183–187, 196–202, 208–209, 229–233, 261–262, 304 and 307 with zones 147–152, 183–187, 196–202 and 208–209 containing the most flexible residues during the MD simulations which are also included in the experimentally highly disordered loops 146–152 and 185–217. This last finding is one of the results showing the agreement between the models here developed and the available experimental data for this enzyme.

For a more detailed analysis of the protein flexibility and with the aim to compare the behavior of both systems along the 10 ns that the MD simulations were carried out, we made a deeper exploration of two regions: the four experimental highly disordered loops and the binding pocket. The residues forming the binding pocket were selected as those with at least one atom lying at a distance shorter than 5 Å from the ligand in the starting crystallographic structure of the complex. In Fig. 4 is represented the molecular surface of the cavity defined using this distance-based criterion and the bound ligand. As can be seen, in this way we can obtain a good definition of the binding pocket. Shorter distances yield an incomplete pocket, while increasing it causes residues from the protein core and surface lying away from the binding site to be included. Special attention has to be paid to common residues in the binding pocket (28 residues) and the four loops (46 residues). In this sense, it is worth noting that 39.29% of the residues in the binding pocket are also included in the four disordered loops while 23.91% of the residues in these loops belong to the binding pocket. That means, as was demonstrated by Qiu et al. [41], that the high flexibility of

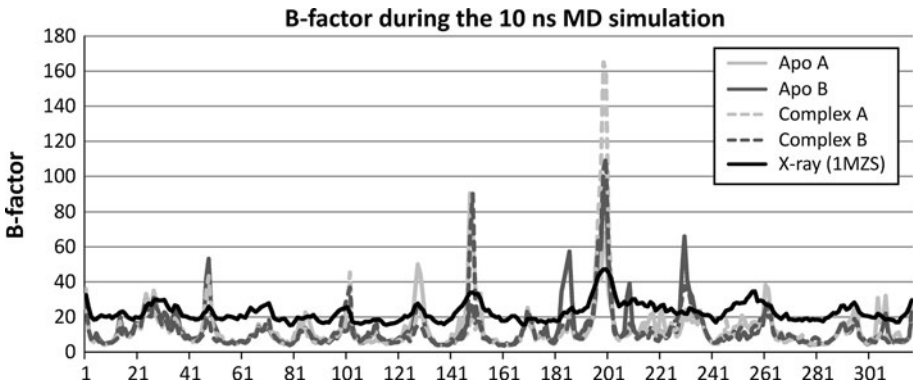

Fig. 3 Backbone temperature factors derived from the simulations of the protein-inhibitor complex (*dashed gray lines*) and the ligand free protein (*solid gray lines*) and the experimental ones from the complex (*solid black line*, pdb code 1MZS)

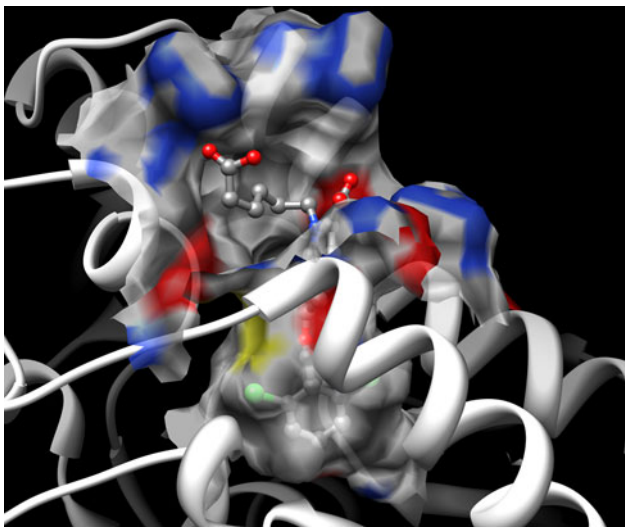

Fig. 4 Molecular surface representation of the binding pocket selected as any residue located at a distance lower than 5 Å from the crystallographic position of the ligand

the loops has a great impact not only on the overall enzyme conformational changes but also in the structure of the binding pocket.

In Fig. 5 are represented the temperature factors for the backbone of the four loops 84–86, 146–152, 185–217 and 305–307 in the two systems under study while in Fig. 6 are plotted those of the backbone of residues in the binding pocket. Vertical grid lines indicate the beginning of each loop while shadows indicate residues in common between the loops and the binding pocket. As can be seen from the plots, there are inter-system and inter-chain differences in the flexibility of residues in both the loops and the binding pocket. As we can note from Fig. 5, residues in the loops and also included in the binding pocket are, in general, not among the most flexible aminoacids in the loops. Furthermore, the higher flexibility is observed in residues 196–202 (sequence RVNPENS) which are located in the furthest region of the large loop 185–217 away from the binding pocket. In the dimer, these amino acids in one subunit make contacts with their analogue residues in the second subunit and they have been proposed to stabilize the dimer.

From Fig. 6 it can be seen that most residues are more flexible in the unliganded protein than in the enzyme-inhibitor complex. Significant peaks of the temperature factor in the ligand-free protein can be observed for residues: W32, R36, R151, G152, G209, F304, G305 and G306. The first four residues are located in the entrance of the binding pocket and have a high influence on it topology as can be seen when comparing, for instance, the structures of the ecFabH solved in complex with a dichlorobenzyloxy-indole-carboxylic acid inhibitor by Daines et al. [35] and the one solved by Gajiwala et al. [25]. On the other hand, residues F304, G305 and G306 are located at the bottom of the binding pocket and form the "oxyanion hole" reported to stabilize the acetyl group of the acetylated intermediate formed after the first halve reaction. Conformational changes in the binding pockets of the free protein will be more deeply discussed later for the 100 ns MD simulation.

If we make a chain by chain analysis of the temperature factors of residues in the four experimentally disordered loops of both systems, we find that 76.09% of the amino acids in the loops of chain A (the one having the bound ligand in the complex) are more flexible in the unliganded protein than in the complex. The same analysis for chain B, that contains no ligand in both systems, yields that 65.21% of the residues are also more flexible in the ligand-free

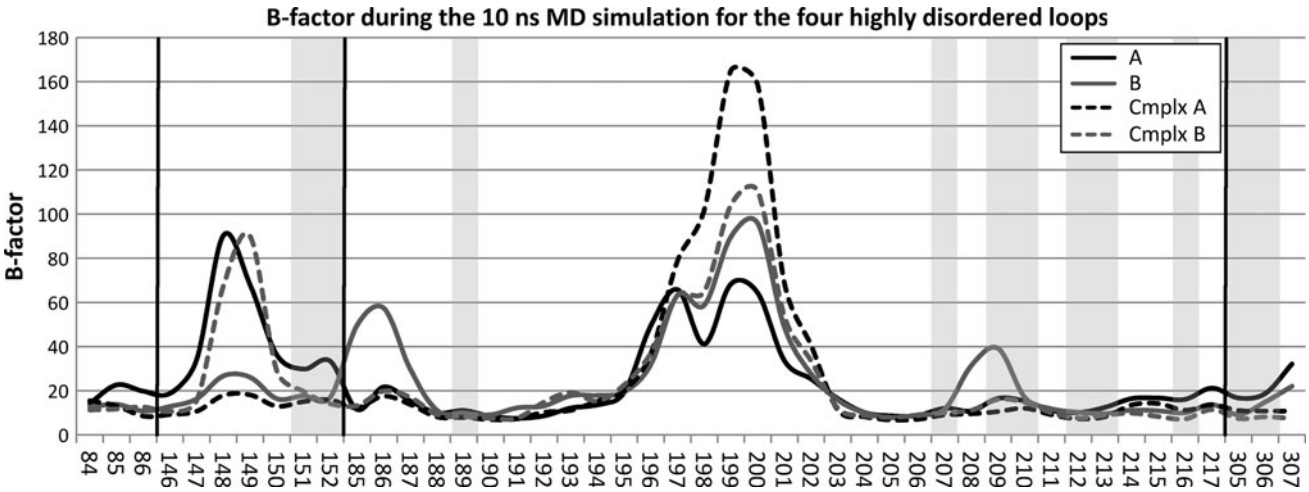

Fig. 5 Temperature factors of the residues backbone in the four experimentally disordered loops. *Vertical lines* indicate the boundaries of each loop while residues that belong to both the loops and the binding pocket are highlighted using a shadow

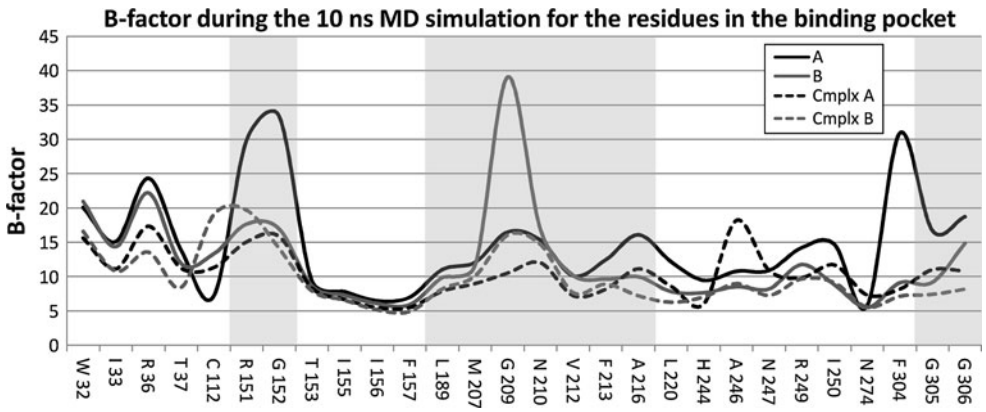

Fig. 6 Temperature factors of the backbone of residues in the binding pocket. Residues that belong to both the loops and the binding pocket are highlighted using a shadow

enzyme. Although these intersystem differences in temperature factors is more pronounced in the subunit containing the bound inhibitor than in the ligand free one, the slightly higher flexibility of the loops of the latter chain in the unliganded protein suggest a certain degree of stabilization in the second subunit as a consequence of ligand binding in the first one as proposed by Alhamadsheh et al. [44]. This stabilizing effect is more remarkable if we analyze the residues in the binding pocket as well as those common to it and the loops as we show next.

If now we only consider common residues in the loops and the binding pocket we can see that all the 11 common residues in both regions are more flexible in the free protein for chain A and the same behavior is observed for 10 out of the 11 aminoacids in chain B. The temperature factors of these residues during the 10 ns simulations are shown in Table 1. The same analysis for the residues in the binding pocket reveals that 85.71 and 82.14% of the aminoacids forming the cavity are more flexible in the unliganded system than in the complex (see TS2 and TS3 in the Supporting Information section for the complete T-factors of the loops and pockets).

Table 1 Flexibility of common residues in the experimentally disordered loops and the binding pocket

Residue	Apo A	Complex A	Apo B	Complex B
R151	29.81	15.04	17.65	19.70
G152	33.36	15.84	16.97	14.10
L189	11.02	7.88	9.87	8.21
M207	12.07	8.96	11.64	10.08
G209	16.49	10.52	39.09	16.05
N210	15.30	12.01	16.92	14.70
V212	10.11	7.24	10.22	7.66
F213	12.44	8.17	9.63	8.89
A216	16.08	11.12	9.99	7.19
G305	16.71	11.01	9.17	7.38
G306	18.73	10.79	14.87	8.17

All these results relative to protein flexibility such as the higher flexibility in the free protein system relative to the enzyme-inhibitor complex in the four highly flexible loops and the binding pocket agree with previous experimental evidence. This comparison of protein flexibility between both systems is summarized in Fig. 7 which represents the enzyme backbone temperature factor for the unliganded protein (left) and the enzyme-inhibitor complex (right). Ribbon representation is used to highlight the four disordered loops while the rest of the protein is represented using the backbone trace and the bound ligand in the A subunit of the complex is shown using a stick and surface representation. It can be seen that region 196–202, the furthest part of the loop 185–217, is the most flexible in both systems while the rest of this loop is more flexible in either chain A or B of the ligand-free protein. This same behavior is observed when the loops 84–86 are compared. In the case of the zone 146–152, it is much more flexible in chain A of the free protein than in the same chain of the complex; in chain B it shows slightly higher temperature factors in the complex although in the free protein it is among the most flexible regions of the system. Finally, the 305–307 small loop at the bottom of the binding pocket show higher flexibility in both chains of the unliganded protein. Until here, we have qualitatively compared the flexibility of ligand-free ecFabH and a complex of this enzyme with an inhibitor occupying only one of the binding pockets. We have also shown that the results derived from our simulations agree with some of the observed experimental facts regarding FabH flexibility such as the higher flexibility of the enzyme in the absence of a bound ligand and the stabilization of the ligand free subunit as a consequence of the binding of the inhibitor to the other one. In the light of these results and to get a better understanding of the conformational changes that ecFabH

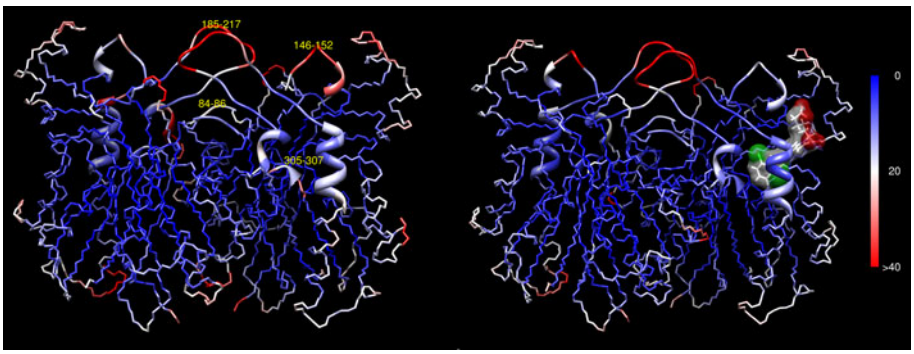

Fig. 7 Comparison of protein backbone flexibility along 10 ns between the unliganded (*left*) and holo (*right*) ecFabH enzyme. The flexibility scale goes from *blue* (low) to *red* (high)

enzyme undergoes, we extended the simulation of the unliganded protein system to 100 ns.

100 ns simulation of Apo-ecFabH

As in both 10 ns simulations, the ligand-free ecFabH remains stable along the 100 ns simulation as shown in the plot of the protein backbone RMS available in the Supporting Information section (FS1). The calculation of the temperature factors for this 100 ns trajectory also yields similar results to the 10 ns simulation with the most flexible residues being: 47, 148–151, 195–203, 208–209, 217–218, 220–222, and 224–236; and aminoacids 148–151, 195–203, 208–209 and 217 located in the highly flexible loops 146–52 and 185–217. All backbone temperature factors calculated from this trajectory are given in the Supporting Information section (Table TS4; Fig. FS2). From the analysis of the backbone temperature factors we can see that the behavior of this measure of protein flexibility is almost the same in both monomers except for the 195–203 loop which is more flexible in the B subunit, and the 208–236 region that shows higher temperature factors in the A chain. This latter zone comprises the whole Ca1 helix which is the most flexible region, not considering the four disordered loops, in the unliganded FabH structure solved by Qiu et al. [41]. The inter-chain differences in the flexibility of the most flexible zones of the enzyme, particularly when comparing the Ca1 helices, confirms the existing asymmetry in FabH function.

Since one of the objectives of this work is to study the conformational changes that take place in the binding pocket, we also examined the temperature factors for whole residues (backbone and side chains together) in this region. This analysis reveals that the binding pocket of the A subunit is more flexible than that of the B one and that, in both chains, the more flexible residues are: W32, R36, C112, R151, G209, N210, L220, H244 and R249. Furthermore, the major differences in the flexibility of both pockets can be found in residues R36, C112, G209, N210, L220 and R249. All of these residues but L220 and the catalytic H244 and C112, which are located at the bottom of the cavity, are at the entrance of the pocket and influence the size and shape of the upper part of the cavity as we will see later. The observed high flexibility in the binding pockets and the different flexibility levels between them, evidence that MD simulations provide unique structural information that is not available in the ecFabH structures solved to date. A plot of the temperature factors of the residues in the binding pocket is shown in the Supporting Information section (FS3).

For a better understanding of the conformational changes that take place in the enzyme we used Principal Components Analysis (PCA). This approach is based on the diagonalization of the covariance matrix and the projection of the resulting eigenvectors onto the structure of the protein. Following this procedure, we can obtain a set of orthogonal axes of movement (eigenvectors), each one representing one direction of the atomic fluctuations in the system. At the same time, each eigenvector is associated with one eigenvalue that gives the amplitude of the fluctuations along each principal component while the sum of all eigenvalues can be considered as the total conformational variance observed during the simulation. One of the advantages of this approach is that it can be used to reproduce most of the motions that occur at longer time scales through the analysis of nanosecond MD simulations as has been demonstrated elsewhere [62–65]. The PCA of the MD trajectory was carried out as described in the "Materials and methods" section.

Here, we analyzed the covariance matrix of the 634 alpha carbons in the enzyme which gives 1902 modes. In Table 2 is shown a resume of the PCA analysis we carried out, that includes the eigenvalues associated to selected eigenvectors, the percentage of conformational variance explained by each eigenvector and the cumulative explained variance. As can be seen, the two more significant eigenvectors account together for 39% of the overall system motion while, as the number of considered eigenvectors increases, the percentage of overall flexibility explained by each component rapidly decreases and after the first 10 principal axes of movement they provide less

than 1% to the overall system conformational fluctuations individually. More interestingly we find that with four components we can explain more than 50% of system flexibility and 70% of it can be explained considering the first 20 principal components.

For a better understanding of the meaning of these principal modes of motion and how they influence the dynamics of our system, we projected our MD trajectory onto them. The projection of eigenvectors 1 and 2 which are the ones individually explaining more than 10% of system flexibility as well as the cumulative projection of eigenvectors 1–4 and 1–20 that explain 39 and 70% of overall conformational variance on the system respectively are represented in Fig. 8. The color scale goes from low flexibility regions along the selected eigenvectors to high.

From Fig. 8 it can be seen that the first eigenvector accounting for 26.97% of total system conformational variance is mainly distributed among the large 185–217 flexible loops of the two subunits of the dimer and helix Ca1 of the chain A. Along the second largest axis of movement of our system, most conformational changes occur at loop 185–217 of the B subunit and both 146–152 loops and Ca1 helices. If we analyze the conformational changes along the first two eigenvectors that account for 39% of system flexibility it can be seen that, although they are more evenly distributed, again most flexibility variance is concentrated in experimental highly flexible regions such as the 185–217, 146–152 and 305–307 loops and Ca1 helices in both monomers. Finally, when taking into account the first 20 principal components which agglomerate 70% of total protein flexibility, the analysis yields similar results as when analyzing the first two components but also includes among the most flexible regions in both subunits the Ca2 helices and the "cap" of the binding pocket formed by the residues 14–40 that contains functionally important aminoacids such as W32 and R36. All these regions shown to concentrate most of the system conformational variance, have been previously shown to be highly flexible which indicates that through the use of MD simulations we can reproduce conformational changes taking place in ecFabH.

Table 2 Resume of the PCA analysis of the 100 ns MD simulation of free ecFabH

Eigenvector	Eigenvalue	% of motion	Cumm. % of motion
1	147.43	26.97	26.97
2	65.73	12.02	39.00
3	45.28	8.28	47.28
4	18.82	3.44	50.72
5	14.16	2.59	53.31
6	12.37	2.26	55.57
7	11.51	2.11	57.68
8	9.08	1.66	59.34
9	7.63	1.40	60.74
10	6.94	1.27	62.01
15	4.08	0.75	66.52
20	2.98	0.55	70.03
All	546.64		100

Once we have shown through PCA of our MD trajectory that the regions where the major motions in the system take place are in agreement with previous experimental results and since there is no experimental data related to the kind of motions these highly flexible regions experience, we can identify through the visual inspection of the projection of the MD trajectory onto the eigenvectors which are the concerted motions that occur in this enzyme along the principal axes of motion. The analysis of the movement along the first eigenvector shows that it mainly consists of the movement of both 185–217 loops with a tendency to approach to each other and to the 146–152 loop of the other monomer. The correlated motions along this axis also include the movement of the 84–86 loops closer to the 185–217 loops and the displacement of the 146–152 loop of the B subunit away from the binding pocket. All these conformational changes pull up the whole Ca1 helix of the A subunit decreasing the distance between this helix and the 146–152 loop. This last conformational change of the Ca1 helix of chain A has a high impact on the topology of the binding pocket since the Ca1 helix contains residues V212, F213, A216 and L220 which are included in the cavity. This observed displacement disrupts not only the position of their side chains but also their backbones. Furthermore, the approximation of the cap of this helix to the loop 146–152 reduces the size of the binding tunnel in its upper region making it narrower. It is interesting to note that most of the conformational changes observed along this first eigenvector take place mainly between the 20 and 50 ns of simulation time, and after 50 ns the conformational changes associated with this direction of motion are considerably reduced (see Fig. 9a for the evolution of the projection of the trajectory into the first PC).

The projection of the trajectory onto the next important principal axis of motion (eigenvector 2) is related to a less wide movement of both 185–217 loops. The relative position of these two loops along this mode range from protein conformations where they are close, to others where they move apart. In contrast to the first principal axis of motions, along the second one, the tendency of the 185–217 loops is to approach during the first 35 ns; afterwards they move in opposite directions until the 80 ns of simulation time, to stay relatively stable during the rest of the simulation in conformations similar to the starting structure. We also found that the projection of the MD trajectory onto this mode reflects a concerted motion in the

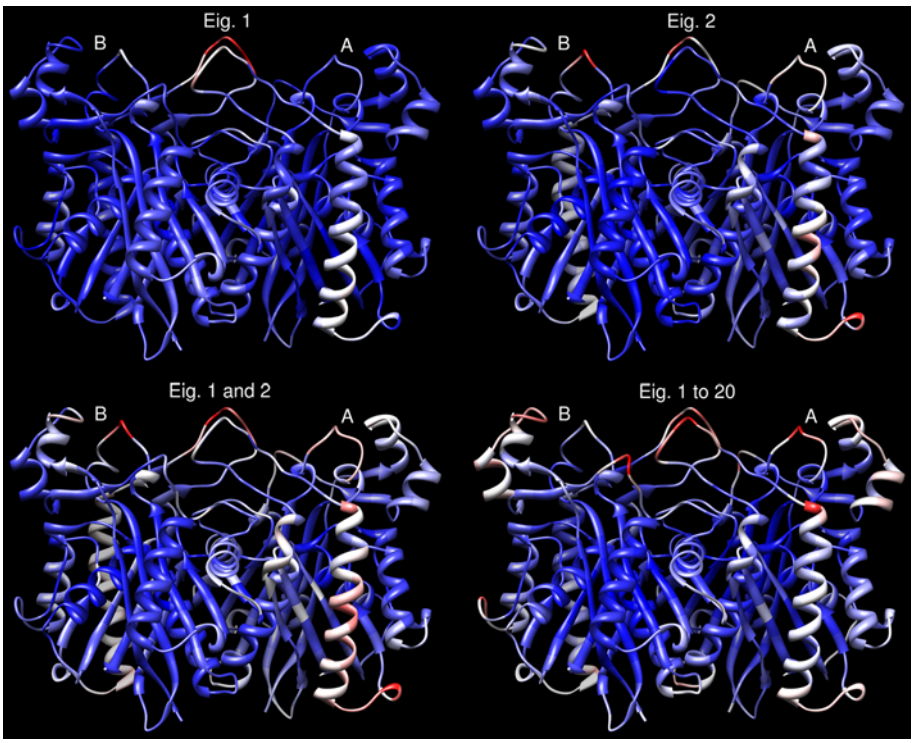

Fig. 8 RMS fluctuations of the unliganded ec-FabH alpha carbons along principal components calculated from the MD simulation. Here we represent system conformational fluctuations along modes 1 (Eig. 1), 2 (Eig. 2), 1 to 4 (Eig. 1 to 4) and 1 to 20 (Eig. 1 to 20). The flexibility scale goes from *blue* (low) to *red* (high)

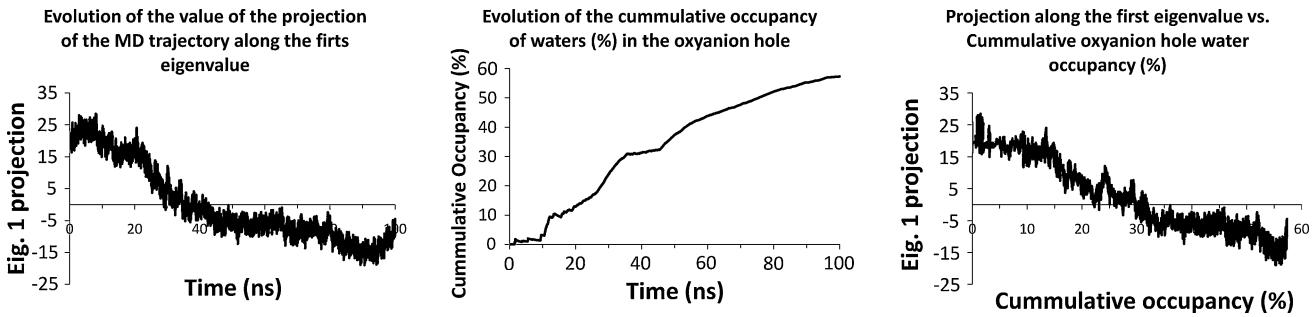

Fig. 9 Relationship between the cumulative occupancy of solvent molecules in the oxyanion hole and the conformations fluctuations of ecFabH along the first PC. **a** Time-dependent evolution of the projection of the trajectory along the first PC. **b** Cumulative occupancy of water molecules in the oxyanion hole. **c** Relationship between the conformational fluctuations of the enzyme along the first PC and the cumulative occupancy of water molecules in the oxyanion hole

same direction of the 146–152 and 185–217 loops of the B subunit. Furthermore, this mode also provides information about the two Ca1 helices and shows a bending in Ca1A and, with lower amplitude, in Ca1B. Albeit less pronounced, a pull up motion in Ca1B similar to that observed along the first mode for Ca1A can also be identified when analyzing this second eigenvector. The stability of the value of the projection onto the first two major eigenvectors by the end of the trajectory indicate that, with our MD simulation, we can identify the principal motions that should take place in the large 185–217 loops at longer time scales.

An interesting result derived from the PCA of the MD trajectory is that the movement of the enzyme along the first principal component concentrated in the same loops found by Qiu et al. [41] to be extremely flexible on their highly disordered ligand-free ecFabH structure. They noticed that, when comparing this disordered ecFabH structure with others which were crystallized following the same protocol and having the same crystal packing environment, the only visible difference between the ordered and disordered structures is the absence of any electronic density in the oxyanion hole. This region of the enzyme encompasses the backbone amide groups of C112 and G306, the cavity formed by G5 and the side chain of F87 from the other monomer. The role of this region is to stabilize the developing negative charge during enzyme catalysis [66]. Based on their findings, Qiu et al. proposed the hypothesis that the presence of a water molecule or acetyl group in the oxyanion hole near the amide nitrogens

of C112 and G306 is critical in the stabilization of FabH structure.

To evaluate this hypothesis, we decided to take advantage of the agreement between the flexibility of the system along the first PC of movement and the high flexibility found by Qiu et al. in four large loops, and to track the presence of water molecules in the oxyanion hole of the unliganded ecFabH during the 100 ns MD simulation following the procedure described in the "Materials and methods" section. In Fig. 9 are shown the plots of the evolution of the value of the projection of the 100 ns trajectory along the first PC (**A**), the evolution of the cumulative occupancy of water molecules in either chain A or B oxyanion holes (**B**) and the value of the projection of the trajectory into the first PC versus the occupancy **(C)**.

As can be seen from Fig. 9a and previously discussed, the major conformational fluctuations along the first PC occur during the first 50 ns of the MD simulation and after that time the variations in the protein conformational changes along the first PC, which can be interpreted as 27% of the total enzyme conformational variance which is also concentrated in these 4 loops, are decreased. On the other hand, the evolution of the cumulative occupancy of water molecules in the oxyanion hole gradually increases as the simulation advances and reach about 40% after 50 ns (Fig. 9b). This lead us to relate the presence of water molecules in the oxyanion hole to the highly flexible loops observed by Qiu et al. As shown in Fig. 9c, higher values of cumulative occupancy of water molecules in the oxyanion hole correspond to smaller changes in the movement of the highly flexible loops. In our opinion, these results are a demonstration of the importance of the presence of a solvent molecule in the oxyanion hole of ecFabH for its conformational stabilization, mainly of loops 84–86, 146–152 and 185–217 that have a high importance in dimer stability and ligand binding.

The results of the PCA of the MD trajectory show concerted motions of important regions of the enzyme that have been previously identified to be highly flexible and to have a high relevance for the enzyme mechanism. Particularly interesting is that, by examining the first two principal axes of motions that account together for 39% of the observed conformational variance during the 100 ns MD simulation, we could identify through visual inspection correlated motions involving the experimentally highly flexible loops 185–217, 146–152 and 84–84 and both Ca1 helices. To further determine whether such concerted motions take place only along the principal axis of motion or also when all degrees of freedom of the system are taking into account, we explored the correlation matrix of the alpha carbons along the trajectory. With this aim, we developed a methodology for clustering the correlation matrix of motions that takes into account functional information of the protein. This clustering procedure is described in the "Materials and methods" section and for the analysis of our MD trajectory we selected 7 regions in both monomers: loops 84–86, 146–152, 185–217 and 305–307; Ca1; Ca2 and the binding pocket. In Table 3 are given the merging distances at which at least one residue from each of the two regions is found on the same cluster. The metric used for building the clustering tree was the average distance. Lower distances indicate higher correlations.

Although the analysis of Table 3 reveals no strong correlation between any pair of the considered zones, useful information can be extracted from it. In this sense, poor intra and inter-chain correlations can be found between some of the enzyme zones used for this analysis (cells with horizontal bars background in Table 3). This is the case for each pocket and Ca1 helix with the 185–217 loop of the chain they belong to. In both cases, the correlated residues are contiguous aminoacids in the Ca1 helix and the 185–217 loops. The same levels of weak correlations are found between residues 182–185 and 306–308 of the B subunit with the first zone containing the first residue of the large 185–217 loop and the last region including the catalytically important aminoacid G306. A network of residues formed by 188–192(A), 207–208(A) and 86–88(B) containing residues from the A subunit binding pocket and 185–217 loop as well as part of the loop 84–86 from the B monomer that extends from the A subunit pocket to the dimer interface, also show weak correlation. On the other hand, the finding of some correlation between part of the 185–217 loop of the B subunit (residues 202–204) with the 146–152 loop in the same monomer is in agreement with the concerted movements observed along the first eigenvector. Finally, the highest correlation among the selected regions of the enzyme is found between the loops 84–86 of the A subunit and loop 185–217 of B. This finding is also in agreement with the concerted motions provided by the PCA.

These results show that, although concerted motions can be identified from the PCA of the MD trajectory, only small groups of weakly correlated residues can be found when analyzing the correlation matrix of atomic positions. These results point to a highly disordered enzyme where large concerted movements can only be found when the principal axes of motion explaining part of overall flexibility are explored. These findings also support the observed asymmetry between the functioning of both monomers. It is also worth noting that the highest correlated motions between the structural important regions under analysis, are found for residues of both chains located at the dimer interface. This observation reflects the importance of this region for enzyme stability despite its high flexibility. On the other hand, the lack of correlation in

Table 3 Matrix of merging distances of clusters containing at least one residue from each of the regions under investigation: P. A/B, binding pocket; Ca1 A/B, Ca1 helix; Ca2 A/B, Ca2 helix; 185 A/B, 185–217 loop; 146 A/B, 146–152 loop; 84 A/B, 84–86 loop; 305 A/B, 305–307 loop

	P. A	P. B	Ca1 A	Ca1 B	Ca2 A	Ca2 B	185 A	185 B	146 A	146 B	84 A	84 B	305 A	305 B
P. A	**0.00**	0.59	0.70	0.74	0.82	0.86	0.49	0.59	0.59	0.63	0.59	0.49	0.69	0.77
P. B	0.59	**0.00**	0.64	0.87	0.88	0.54	0.64	0.49	0.83	0.54	0.51	0.62	0.78	0.67
Ca1 A	0.70	0.64	**0.00**	0.87	0.88	0.86	0.45	0.70	0.87	0.70	0.70	0.86	0.87	0.77
Ca1 B	0.74	0.87	0.87	**0.00**	0.88	0.87	0.78	0.45	0.74	0.87	0.87	0.78	0.78	0.87
Ca2 A	0.82	0.88	0.88	0.88	**0.00**	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
Ca2 B	0.86	0.54	0.86	0.87	0.88	**0.00**	0.86	0.86	0.87	0.86	0.86	0.86	0.87	0.86
185 A	0.49	0.64	0.45	0.78	0.88	0.86	**0.00**	0.67	0.83	0.70	0.70	0.49	0.59	0.77
185 B	0.59	0.49	0.70	0.45	0.88	0.86	0.67	**0.00**	0.83	0.44	0.34	0.76	0.76	0.49
146 A	0.59	0.83	0.87	0.74	0.88	0.87	0.83	0.83	**0.00**	0.87	0.87	0.83	0.83	0.87
146 B	0.63	0.54	0.70	0.87	0.88	0.86	0.70	0.44	0.87	**0.00**	0.63	0.86	0.87	0.77
84 A	0.59	0.51	0.70	0.87	0.88	0.86	0.70	0.34	0.87	0.63	**0.00**	0.86	0.87	0.77
84 B	0.49	0.62	0.86	0.78	0.88	0.86	0.49	0.76	0.83	0.86	0.86	**0.00**	0.69	0.86
305 A	0.69	0.78	0.87	0.78	0.88	0.87	0.59	0.76	0.83	0.87	0.87	0.69	**0.00**	0.87
305 B	0.77	0.67	0.77	0.87	0.88	0.86	0.77	0.49	0.87	0.77	0.77	0.86	0.87	**0.00**

Bold characters are used to highlight the main diagonal of the matrix

the movement of residues belonging to both binding pockets is a reflection of the existing non-cooperative function in the free-ligand form of ecFabH.

Selection of representative binding pocket conformations and molecular docking

One of the objectives of this research is to evaluate the value of the combination of MD simulations and molecular docking for the accurate prediction of the binding mode of inhibitors to ecFabH. There are X-ray data for two kinds of FabH inhibitors: dichlorobenzyloxy-indole-carboxylic acids and benzoylaminobenzoic acids. The starting structure for the MD simulations corresponds to the complex of ecFabH with one of the inhibitors in the first class and then they can be modeled without the need for any protein flexibility information. For that reason, we selected the ten benzoylaminobenzoic acid inhibitors published by Nie et al. [32] which were shown to be broad spectrum FabH inhibitors and were assayed against *E. coli*. The X-ray structure of one of these inhibitors in complex with FabH from *E. feacalis* (efFabH) has been solved (pdb 3IL6) and the key interactions stabilizing the complex were found to be with the catalytic H244 and N274 through the acid moiety as well as a stacking interaction between F213 and the benzene ring of the phenoxy group [25]. Since no crystal structure of these inhibitors in complex with ecFabH is available, we will evaluate the value of a set of representative conformations of the binding pocket extracted from the MD simulation of the free enzyme and the available structures of ecFabH to reproduce the right orientation of the ten benzoylaminobenzoic acid inhibitors inside the cavity. The structure of these inhibitors is shown in Fig. 10.

As can be seen from Fig. 10, all compounds used in this study share the same scaffold with the only difference being the substituent at the para-position of the benzoylamino group. Taking this into account, the correctness of the orientation of the ligands is evaluated based on the RMSD of the common heavy atoms of the compounds relative to the co-crystallized X-ray structure of one of the inhibitors with efFabH (pdb 3IL6, compound **1**). We also examine whether the above mentioned key protein–ligand interactions can be reproduced in the predicted complexes of the inhibitors and ecFabH.

The first issue to be solved to accomplish this step is the selection of a set of conformations of the binding pocket of the enzyme. This can be done in several ways: selection of *n* snapshots every regular time steps, selection of receptor low energy conformations and by means of clustering analysis among other strategies. Since we are interested in obtaining a diverse set of representative binding pocket conformations and at the same time minimizing the number of receptor conformations to save calculation time during docking calculations, we selected cluster analysis for this selection step. This analysis was carried out using the PTRAJ tool distributed within the AmberTools package.

For cluster analysis we followed the recommendations given by the developers of the clustering algorithms implemented in PTRAJ [67]. On the other hand, we have shown that there are differences in the flexibility of both binding pockets; for that reason the cluster analysis was carried out separately for the binding pocket of each subunit. We analyzed a number of clusters going from two to twenty

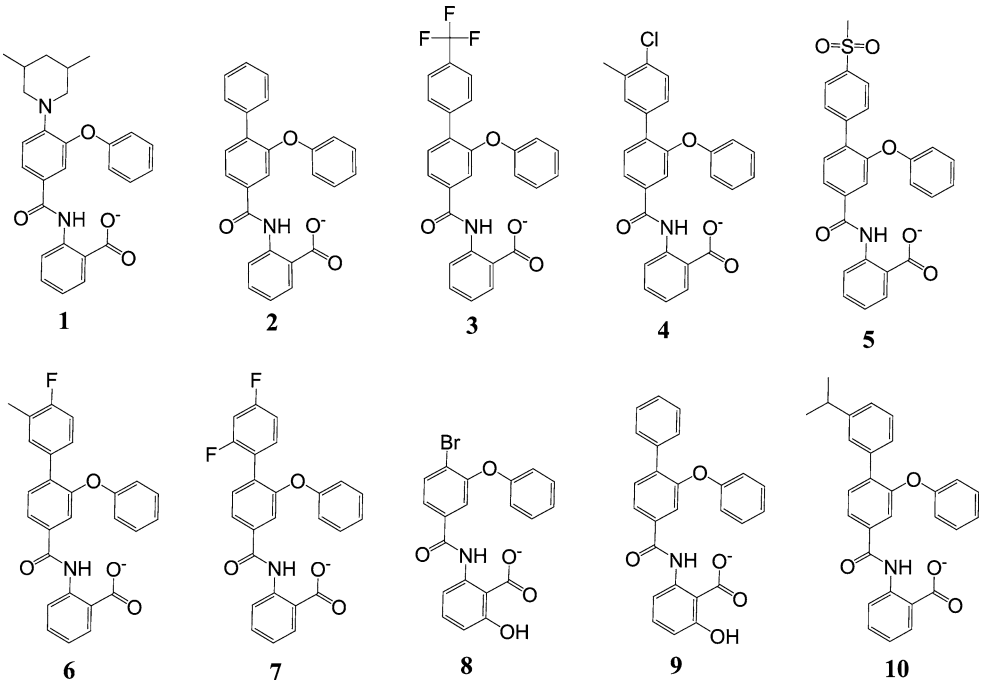

Fig. 10 Ten benzoylaminobenzoic acid inhibitors used for docking calculations

for each active site and studied the behavior of the Critical Distance (CC), DBI, pSF and SSR/SST statistics. The theoretically optimal number of clusters is obtained when the slope of the CC curve drastically decreases after a new partition, DBI has a minimum, pSF show a maximum and the SSR/SST plot reaches a plateau. Of course, all these conditions are difficult to meet at the same time in a real MD trajectory clustering problem and a balance between all of them should be established. Besides all the above mentioned statistical criteria, we also need to take into account that the more clusters we select, more complex and less efficient will be the docking process. Based on these considerations we decided to select 8 and 7 representative conformations of the binding pocket in chains A and B respectively as shown in Fig. 11. The vertical lines on the plots are used to highlight the selected number of clusters.

The stereo view of the selected representative conformations of the pockets of subunits A and B is represented in Fig. 12. Among the residues forming the binding pocket, we can identify a set of aminoacids which are located in the entrance of the cavity. These residues are: W32, R36, R151, G209, N210 and R249, all of them influence the size and shape of the cavity and are among the most flexible aminoacids in both binding pockets as previously stated. Particularly interesting is that in three of the representative conformations of the A subunit pocket, R36 completely (c0A and c1A) or partially (c4A) folds over the entrance of the cavity fully or partly blocking the access to the pocket.

To obtain conformations of the compounds under study in agreement with the experimentally solved orientation in efFabH, there should be enough space for the stacking interaction of the compounds with F213. In Fig. 12 it can be seen that this inhibitor interaction should not be possible to model in the receptor conformations c2A, c2B and c4B since R249 folds over F213 and prevents the stacking of the compounds phenoxy moiety. On the other hand, W32 serves as a kind of cap of the binding pocket and adopts different conformations that range from rotamers oriented to the interior of the cavity which restrict its size, to others pointing to the solvent that open an additional subpocket that can be occupied by the ligands. Another highly flexible residue during the MD simulations in the ligand binding pocket is the catalytic H244; the representative snapshots of both cavities show this aminoacid in two preferred orientations: the one found in the crystallographic structures of ecFabH solved to date (c0A, c1A, c0B, c1B, c2B and c3B) and a different one where H244 is shifted towards E302 increasing the size of the pocket in the bottom of the cavity (c2A, c3A, c4A, c5A, c6A, c7A, c4B, c5B and c6B). This shifted conformation, in all cases, is stabilized by a network of hydrogen bonds between the side chains of E302 and H244.

As we previously analyzed, there is higher flexibility in the chain A Ca1 helix than in Ca1 of chain B. The analysis of the representative conformations of the binding pocket in the A subunit is consistent with this finding since we can observe a high diversity of conformations in the backbone of G209, N210, F213, V212, A216 and L220 which is not seen in the B subunit and a set of selected X-ray structures as we will show below.

To evaluate the structural diversity of the binding pocket conformations provided by the MD simulations, we also

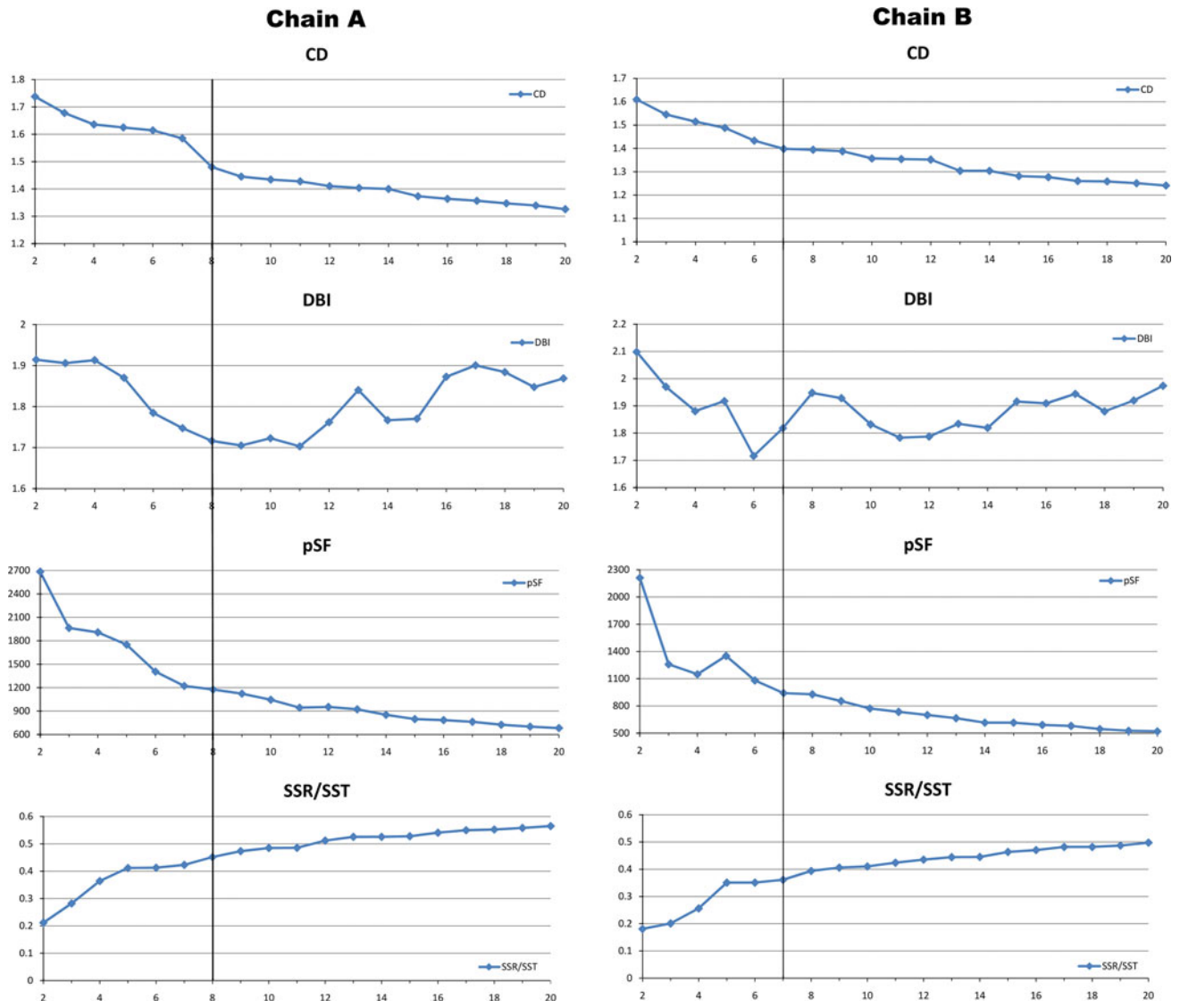

Fig. 11 Metrics used for the selection of the optimal number of clusters of each pocket. The *vertical lines* indicate the selected number of clusters on each chain

analyzed the ecFabH crystallographic structures available in the Protein Data Bank. For this analysis, we only used those structures with all residues in the active site solved. The superposition of all these FabH structures is provided in Fig. 12 (bottom). The visual inspection of the crystal structures of ecFabH reveals that they are not adequate to correctly model the inhibitors we selected for this study; in 1EBL and 3IL9 R36 folds over the entrance of the binding pocket blocking part of the entrance, the same effect can be observed for one of the two rotamers of N210 found in 2GYO; in 1HND, 1HNJ, 2EFT and 2GYO, W32 points to the entrance of the binding pocket and reduces the size of the upper subpocket where the para-substituent of the benzoylamino group should bind and R249 folds over F213 blocking it interaction with the phenyl ring of the benzoylamino moiety. In the case of the 1MZS, R249 also blocks the interaction with F213. Additionally, the visual inspection of the representative conformations of the binding pocket reveals different size, shape and relative residue orientations along the 100 ns simulation of ecFabH and shows a more diverse set of protein conformations than provided by the available crystallographic structures for this enzyme.

Despite there is no crystallographic structure of the inhibitors used in this study in complex with ecFabH, the high conservation of the residues in the binding pocket in efFabH and ecFabH strongly suggests a similar binding mode of these inhibitors to ecFabH. This inter-species extrapolation of the binding mode of ligands from one protein to a highly conserved one in a different species is often used in the context of homology modeling, and can be directly corroborated by analyzing the X-ray structure of

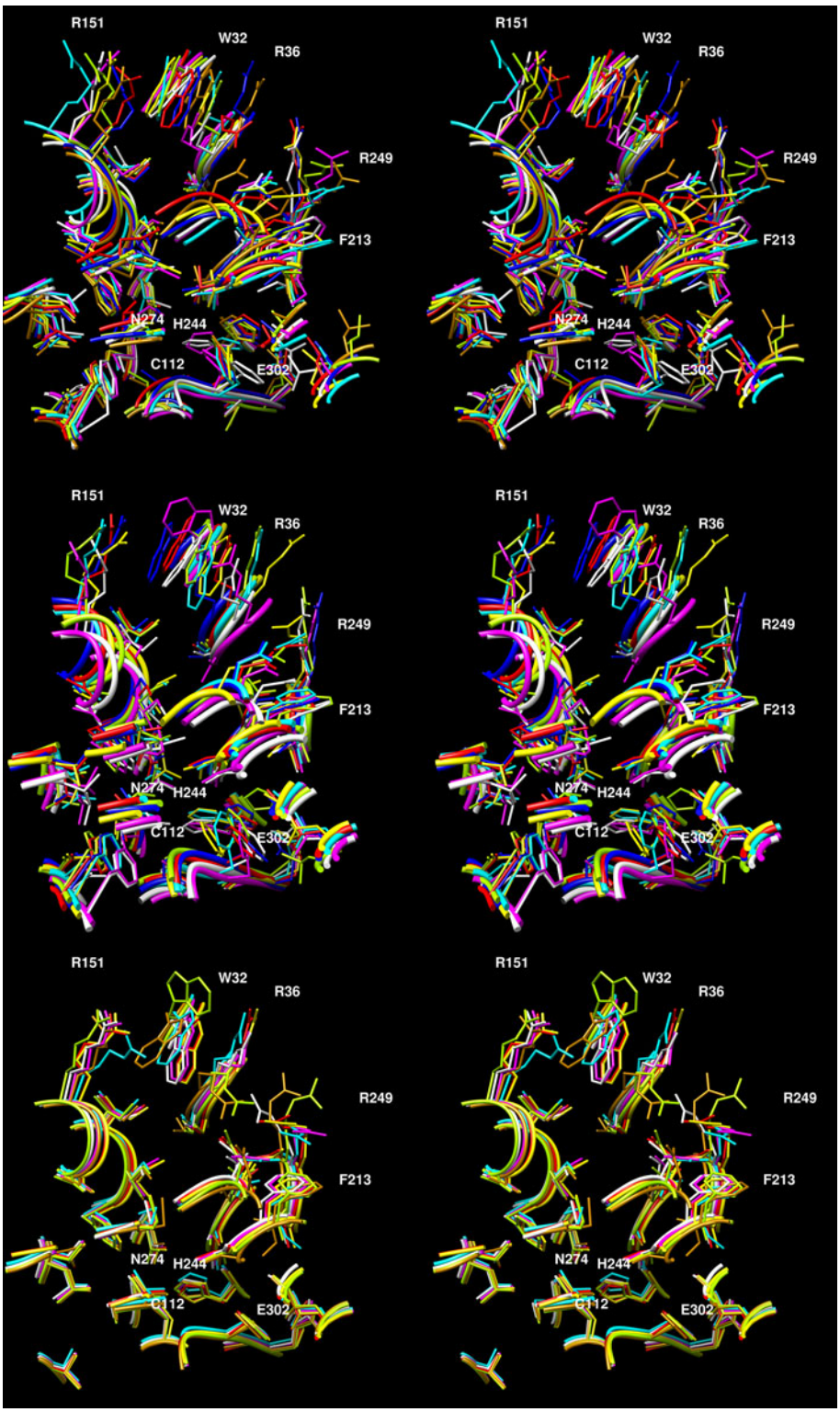

Fig. 12 Stereo view of the selected 8 and 7 representative conformations of the binding pocket of subunits A (*top*) and B (*center*) as well as the 7 crystallographic structures of ecFabH with no unsolved residues in the binding pocket. The color scheme is the same for both subunits A and B: c0 *white*, c1 *purple*, c2 *cyan*, c3 *yellow*, c4 *red*, c5 *blue*, c6 *green* and c7 *brown*. For the X-ray structures the correspondence is: 1HND *white*, 1HNJ *purple*, 1MZS *cyan*, 2EFT *yellow*, 2GYO *red*, 3IL9 *green* and 1EBL *brown*

homologous proteins in complex with the same ligand [68–70]. When the alignment of the residues in the binding pocket in ecFabH and efFabH is analyzed, we find that most residues are conserved or mutated by analogous residues (i.e. aliphatic by aliphatic, H-bond donor/acceptor by H-bond donor/acceptor, etc.). More interestingly, if we analyze only the residues in the efFabH binding pocket making direct contacts with the ligand in the X-ray structure, we find that among all these 18 residues, 13 are conserved in ecFabH and the remaining 5 aminoacids are

mutated in a way that does not affect the size or chemical environment of the binding pocket. These mutated residues are, in ecFabH and efFabH respectively: G152S (here the inhibitor makes contacts with the backbone atoms), I157 V, I158L, N210R and V211I. These facts evidence that the binding mode of the inhibitors used in this study should be conserved in ecFabH and efFabH.

To test our hypothesis that the MD simulation provides useful and unique structural information for the accurate modeling of FabH inhibitors, we carried out molecular docking calculations of the selected 10 FabH inhibitors against the 15 representative binding pocket conformations selected using cluster analysis of the MD trajectory, in addition to the 7 available ecFabH crystallographic structures.

Molecular docking involves two main steps: the orientation of the molecules inside the cavity and the scoring of each pose to evaluate the quality of the obtained binding mode. A good docking protocol should be able to find correct orientations of the ligand inside the binding pocket and give them the lowest scoring values for a successfully discrimination between poses. In the particular case of our modeling studies we should expect that, at least for most of the ligands, the conformation with the lowest scoring value among all considered receptor conformations will resemble the overall orientation of compound **1** co-crystallized with efFabH (pdb 3IL6). This is evaluated by measuring the RMSD of the common core of the inhibitors relative to the experimental orientation of compound **1** in efFabH as well as by means of the visual inspection of the predicted binding mode to determine whether the key interaction observed in the X-ray structure are reproduced [25]. All calculations were run using the protocol described in the "Materials and methods" section and a summary of the results of the docking calculations is given in the Supporting Information Tables TS5 and TS6. These results are summarized in Fig. 13 where we plot the RMSD of the common core of each compound relative to the X-ray structure of **1** (pdb 3IL6) against its DOCK PBSA scoring value. In this figure we only considered the lower scored pose of each compound on each receptor conformation.

A quick survey of the conformations of the ligands obtained after docking these ten compounds against the 22 binding pocket conformations under study (15 from the MD simulations and 7 X-ray ecFabH structures), shows that they are grouped in four main sets: (1) conformations resembling the experimental orientation of **1** inside the pocket (acid function pointing towards N274 and H244, phenyl of the benzoylamino moiety above F213 and para-substitution of the benzoylamino group located near W32, R36 and R151); (2) conformations with the phenyl ring, the para-substituent and the carboxylate group twisted (rotated about 180 degrees relative to the X-ray position); (3) conformations with only the acid moiety twisted; conformations with only the phenyl and para- benzoylamino group twisted and (4) conformations with the acid group located at the entrance of the cavity and the phenyl and para- benzoylamino group occupying the bottom of the binding tunnel.

The results of the molecular docking against the crystallographic structures show that poses of the compounds inside the binding pocket can be obtained for only three out of the seven available X-ray receptor conformations. If we do not take into consideration compound **8**, which is the one with the smaller substituent (Br) in the para-position of the benzoylamino core, conformations of the ligands inside the cavity can only be found in the receptor corresponding to pdb structure 1MZS. It can also be seen that the best conformation per compound obtained using the crystallographic structures of ecFabH is either oriented with the acid pointing to the entrance of the binding tunnel and the phenyl and para- benzoylamino substituents oriented to the base of the cleft; or with these last two moieties twisted. The RMSD values between 2.6–5.3 Å are a reflection of the inability to find correctly oriented poses of the compounds under study on this X-ray receptor conformation. An exception here is compound **10**, which lowest scored conformation shows a RMSD value of 1.9 Å relative to the X-ray reference structure. This compound is correctly oriented at the bottom of the binding cavity but the position of R249 prevents the phenyl of the benzoylamino moiety to interact with F213. After rescoring the conformations of the ligands found inside the binding pocket, these results are not improved since there are no close to experimental binding poses predicted for them in the X-ray ecFabH structures. These results confirm our initial supposition that the available crystallographic structures of ecFabH can't be used to correctly model the ten benzoylaminobenzoic acid inhibitors selected to carry out this research.

From Fig. 13 it can be seen that, for every compound, the lowest PBSA scoring value as well as the closest to experimental binding pose are achieved on complexes derived from the MD snapshots. Furthermore, the modeling process is considerably more accurate when the MD-derived receptor conformations are used than when the ecFabH X-ray structures are considered since docking poses combining low scoring and RMSD values are only obtained when the modeling takes place on this first group of receptor structures.

The analysis of the binding poses and their scoring values obtained when the compounds are docked onto the 15 receptor conformations selected from the MD simulation, show that there is an increased diversity in the prediction of the binding modes than when calculations are carried out for the crystallographic structures of the enzyme. In this sense, we can find orientations of almost all

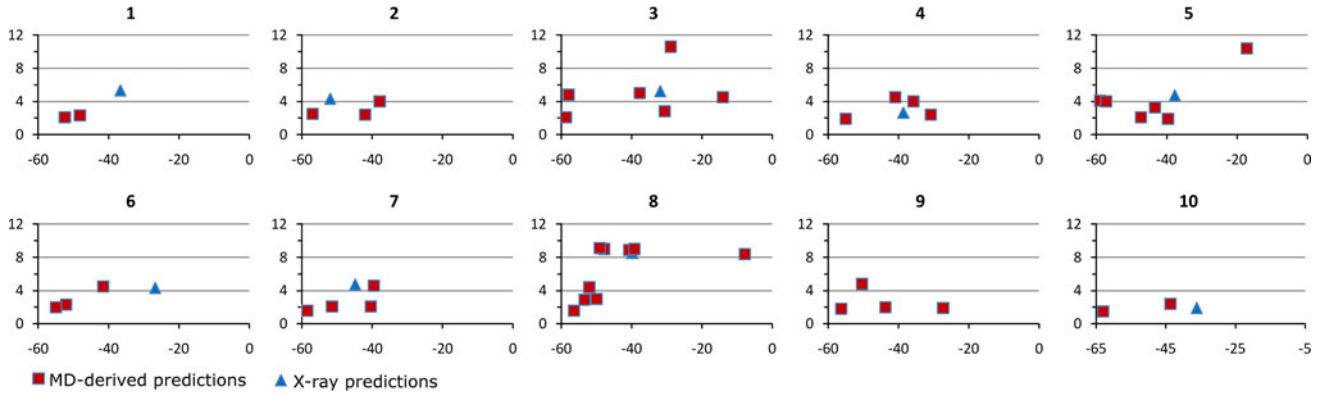

Fig. 13 RMSD between the lowest scored pose of each compound on each receptor conformation and the X-ray structure of compound **1** in complex with efFabH, versus its DOCK PBSA scoring value. *Square* and *triangular* markers are used to represent the predicted binding poses on the MD extracted receptor conformations and the ecFabH X-ray structures, respectively

ligands resembling the X-ray orientation of **1** in receptor conformations c1B and c5B and diverse inhibitor poses mainly in receptors c2A, c3B and c4B. The visual inspection of the disposition of the residues in the binding pocket of receptors c1B and c5B, purple and red respectively in Fig. 12 (center), show that these two conformations comply with two structural features necessary for predicting the correct orientation of these ligands inside the cavity: R249 is oriented towards the solvent and hence the stacking interaction with F213 can take place and W32 partly moves away from the entrance of the cavity and allows the para-substitution on the benzoylamino group to fill the space created by this movement. More interesting, when considering the best global pose for each ligand after the docking procedure using the grid-based scoring function, we found that for eight out of the ten compounds, the best predicted pose corresponds to a low RMSD orientation of the ligand in the cavity which is achieved in the c5B receptor conformation. After rescoring with the DOCK PBSA scoring function, the previous results are improved and for nine out of the ten compounds the best scored pose is correctly oriented within the cavity with RMSD values relative to the crystallographic structure of compound **1** below 2 Å (**4, 6, 7, 8, 9** and **10**) or in the range 2.1–2.5 Å (**1, 2** and **3**). A superposition of the best predicted binding pose of all compounds but **5** after the rescoring process is shown in Fig. 14a.

The next set of conformations with better scoring values, either with the grid-based and PBSA functions, are those obtained for compounds **3, 4, 5, 7, 8 and 9** docked in the receptor c2A. In this receptor conformation, although there is enough space in the binding tunnel to accommodate the ligand, R249 folds over F213 and no orientations of the inhibitor in agreement with the experimental pose can be found. This is the cause of the high RMSD values (2.9–4.8 Å) observed for the predicted binding modes on this receptor conformation. The best poses after the rescoring process obtained for compounds **3, 4, 5, 7, 8 and 9** in the receptor conformation c2A are shown in Fig. 14b.

It is also noteworthy the existing high differences in the final scoring values between the predicted binding poses on the MD-derived and X-ray receptor conformations. To get more insight into the possible reasons for these differences, we examined the components of the PBSA scoring function. This scoring function is composed of three main terms: the Van der Waals (VDW) interactions (modeled using a 6–12 Leonard-Jones potential), the electrostatic component and an approximation of the hydrophobic effect through the surface area. These last two terms are calculated using the ZAP Tool Kit from Openeye which is a solver of the Poisson–Boltzmann equation to account for the hydrophilic effect on electrostatic screening and the exposed surface area of the complex as an approximation of the hydrophobic effect. In Fig. 15 the final scores are plotted as well as each of its components for the nine compounds for which a solution can be found simultaneously on both c5B and 1MZS receptors.

From Fig. 15 it can be seen that for seven out of the nine compounds (**1**, **3**, **4, 5**, **6**, **7** and **10**) there exists a high difference in the score of binding to these two ecFabH conformations. Although the major differences among all the scoring function components are observed for the Van der Waals portion of it, the electrostatic interactions are equally or more favorable in the c5B MD-derived ecFabH conformation; while there are no significant differences in the contribution of the surface area term. In the case of the electrostatic component, we should note that in this approach the effect of the solvent is implicitly taken into account during the scoring process. This explains the positive values of the electrostatic components of the binding energies since it is a common issue that the desolvation penalty of removing water from around charges often outweighs the newly established charge-charge interactions. On the other hand, in both ecFabH conformations, the

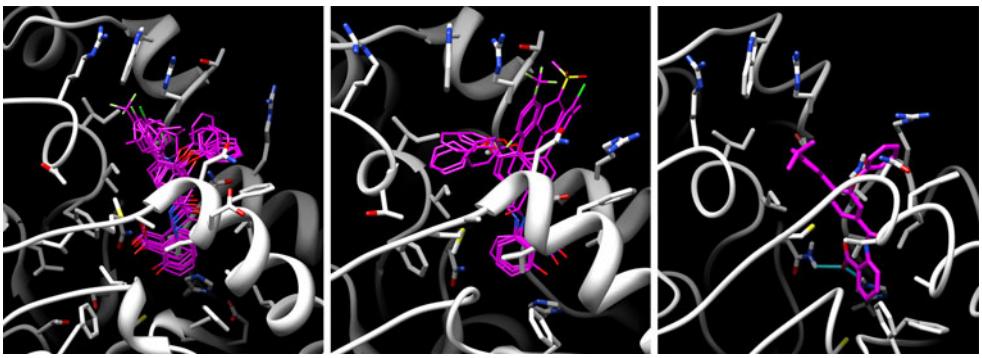

Fig. 14 Predicted binding mode of all compounds but 5 in the c5B receptor (*left*) as well as the wrong predicted poses if some chemicals in c2A (*center*) and the final predicted pose of compound 10 after the 1 ns MD simulation. H-bonds of compound 10 with H244 and N274 are drawn with *cyan* pseudobonds

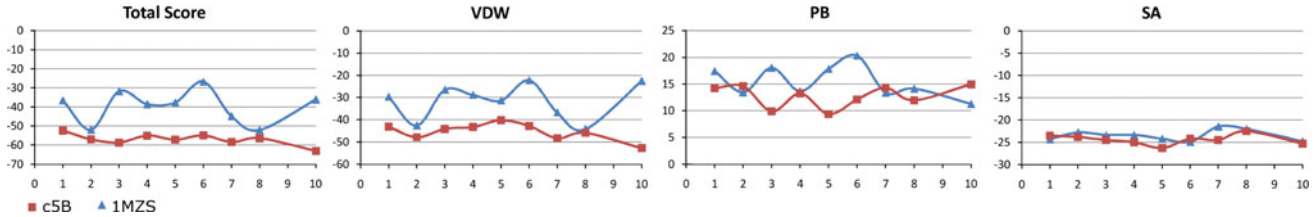

Fig. 15 Total DOCK PBSA scores and its components (VDW: Van der Waals; PB: Poisson-Boltzmann electrostatic component; SA: hydrophobic effect calculated through the surface area) for the predicted binding poses in receptor conformations c5B (*squares*, from the MD) and 1MZS (*triangles*, from PDB)

predicted poses of the ligands fill the binding pocket space which explains the similar results obtained for the surface area component.

To illustrate these observed differences we can analyze the predicted binding mode of compound **10** in both receptor conformations. The orientation of this ligand inside the cavity is very similar in both receptor conformations with a RMSD of 1.9 and 1.5 Å relative to the reference X-ray ligand in 1MZS and c5B respectively and, more interesting, the RMSD between these two predicted poses is 1.3 Å, being the only difference between them the orientation of the phenyl ring of the benzoylamino moiety. Despite these two conformations are very similar; there exists a high difference on their scores given by the VDW component of the scoring function. It is also necessary to highlight that the binding pocket conformations of these two receptor structures are very similar with the major differences found for the arginine residues at the entrance of the pocket, mainly for R249. This difference in the VDW component can be explained on the base that in the X-ray structure of ecFabH, R249 folds over F213 and blocks the formation of the stacking interaction between this last residue and the ligand; this interaction is reproduced for the c5B MD-derived receptor conformation.

Although the analysis of the best global pose for each ligand under study except compound **5**, show that they are correctly oriented inside the cavity, we decided to run an additional 1 ns MD simulation of the predicted complexes to refine the poses and evaluate the stability of their binding to ecFabH through MM-PBSA calculations. One advantage of this last modeling step is that we can reproduce to some extend the induced fit effect due to ligand binding and then obtain more accurate and near to reality binding poses. The calculation of the free energies of binding for these compounds was carried out using the last 400 trajectory snapshots corresponding to the last 200 ps of these 1 ns MD simulations. To propose a final binding mode for each compound, we considered the above mentioned 400 snapshots as a cluster and the centroid of it was selected based on the RMS of the ligand. In Table 4 are given the obtained free energies of binding, the RMSD of the ligands relative to the X-ray structure after the docking calculations and after the MD refinement, as well as the experimental minimum inhibitory concentration (MIC, taken from [32]) for these compounds. In Fig. 14c is shown the refined binding mode of compound **10** which is the one showing the better $\Delta G_{binding}$ using both the PB and GB approaches.

It can be seen that, not only for compound **10** but in all cases, the stacking interaction between F213 and the benzene ring of the phenoxy group is better defined, W32 rotates for a stronger interaction with the para-substituent of the benzoylamino group and H244, N274 and the ligand approach allowing for H-bonds interactions between the carboxylate moiety of the ligand and these two catalytic

Table 4 Free energy of binding obtained using both the MM_PBSA and MM_GBSA approaches, RMSD relative to the experimental structure of compound **1** in complex with efFabH before and after the MD refinement, as well as the experimental minimum inhibitory concentration (MIC) for each compound

Comp.	PB Tot	GB Tot	RMSD after docking (Å)	RMSD after MD refinement (Å)	MIC
1	−28.44	−35.95	2.1	1.6	5.6
2	−28.04	−31.92	2.5	2.7	1.4
3	−24.16	−31.94	2.1	1.6	0.7
4	−25.08	−33.98	1.9	1.6	0.7
6	−22.41	−33.62	2.0	2.0	0.7
7	−37.01	−40.23	1.6	1.4	0.7
8	−23.46	−33.4	1.6	1.2	2.8
9	−28.73	−36.36	1.8	1.4	1.4
10	−37.13	−40.88	1.5	0.8	2.8

MIC values are taken from [32]

residues. As a result of this refinement step, the ligands' core RMSD relative to the experimental orientation of compound **1** in complex with efFabH is reduced for all compounds but **2**, confirming the utility of this short MD refinement for increasing the accuracy of the predicted bonding modes.

Although no correlation can be found between the MIC values, which are low for all the 9 compounds used for docking, and the PB or GB free energies of binding, the low values of GB and PB indicate strong binding as well as stability in the predicted complexes. This lack of correlation can be explained on the base that the MIC measures the activity of the compounds against the whole microorganism and the free energy calculations are intended for estimating the stability of the ligand-receptor complex. Moreover, MIC determination is carried out using discrete compound concentrations and hence makes this measure of activity not optimal for correlation analysis.

Since here we have shown the advantages of the combination of MD and docking simulations to accurately model ecFabH inhibitors over a X-ray structures based strategy, this methodology will be extended to a broader set of FabH inhibitors with determined enzymatic inhibitory activity in future works.

Concluding remarks

Here, we have used MD simulations and molecular docking to make an exhaustive exploration of the conformational space of *E. coli* FabH. We compared the behavior of a complex between FabH and a dichlorobenzyloxy-indolecarboxylic acid inhibitor and the apo-enzyme through 10 ns of simulation and found higher flexibility in the latter system which agrees with previously experimental results suggesting the stabilization of the protein upon ligand binding. The simulation of the unliganded protein was then extended up to 100 ns and we have shown through the PCA of the trajectory that 70% of system conformational variance is concentrated in regions experimentally shown to be highly flexible such as: the loops 185–217, 146–152 and 305–307, both Ca1 and Ca2 helices and the "cap" of the binding pocket formed by the residues 14–40 that contains catalytically important residues such as W32 and R36. By using the PCA of the 100 ns MD trajectory and the analysis of the occupancy of the solvent in the oxyanion hole during our simulation, we also found that the presence of water molecules in the oxyanion hole of the enzyme is closed related to the conformational stability of loop regions of this protein previously described as being highly flexible.

This 100 ns simulation was subjected to a clustering process to find a representative set of binding pocket conformations that contain relevant information related to its flexibility. This set of conformations of the pocket was compared with the available X-ray structures of ecFabH and we saw that they contain structural relevant information which is not present in the crystals. The importance of the information related to the binding pocket flexibility obtained through the MD simulations was shown by docking ten FabH inhibitors into the available X-ray structures of the receptor and into the set of representative binding pocket conformations extracted from the trajectory. The results of this step show that the binding mode of the selected inhibitors could only be correctly predicted in receptor conformations extracted from the MD simulations, highlighting the relevance of our approach to accurately model the binding of FabH inhibitors. Furthermore, the best binding poses resulting from the docking process were subjected to additional refinement through a short MD simulation and the free energy of binding was evaluated using MM-PBSA calculations. These predictions based on complex refinement proved the stability of the predicted complexes and reproduced conformational changes that take place in the receptor as a consequence of ligand binding. Undoubtedly, the results here shown, besides being in agreement with the observed experimental behavior of this enzyme and provide new structural

information which is not available in the X-ray structures solved to date, will allow for a more accurate modeling of ecFabH inhibitors and will serve to set up protocols for the study of FabH inhibitors at different levels of accuracy ranging from virtual screening of libraries of chemical compounds to their highly accurate modeling.

Supporting information available

Supporting information includes the tables relative to: backbone temperature factors calculated from the 10 ns MD trajectories as well as those from the starting X-ray structure (TS1); backbone temperature factor calculated from the MD 10 ns trajectories for the four experimentally highly flexible loops (TS2) and for the residues in the binding pocket (TS3); backbone temperature factors calculated from the MD simulation for 100 ns of the free ecFabH (TS4); the lower GRID and PBSA scores for every compound after the docking them on each receptor conformation (TS5) as well as the RMSD of these conformations relative to the X-ray structure of compound **1**.

Also included in this section are the plots of: the backbone RMSD along the 100 ns simulation of the unliganded ecFabH (FS1); backbone temperature factors derived from the 100 ns simulation of the protein (FS2) and the calculated temperature factors of residues in the binding pocket of each subunit along the 100 ns simulation (FS3).

Acknowledgments We thank the Vlaams Supercomputer Centrum (VSC) for providing access to the computational resources needed for accomplishing all calculations. Pérez-Castillo Y. thanks the Flemish Interuniversity Council (VLIR) for financial support through the project: "Strengthening research and PhD formation in Computer Sciences and its applications" in the framework of the VLIR-UCLV collaborative program. Cabrera-Pérez M. A. and Pérez-Castillo Y. thank the Spanish Agency of International Cooperation for the Development (AECID) for financial support through the project "Montaje de un laboratorio de química computacional, con fines académicos y científicos, para el diseño racional de nuevos candidatos a fármacos en enfermedades de alto impacto social" (D/024153/09).

References

1. Monaghan RL, Barrett JF (2006) Antibacterial drug discovery–then, now and the genomics future. Biochem Pharmacol 71(7): 901–909
2. Levy SB, Marshall B (2004) Antibacterial resistance worldwide: causes, challenges and responses. Nat Med 10(12 Suppl):S122–S129
3. Nikaido H (2009) Multidrug resistance in bacteria. Annu Rev Biochem 78(1):119–146
4. Livermore DM (2004) The need for new antibiotics. Clin Microbiol Infect 10(s4):1–9
5. Pucci MJ (2006) Use of genomics to select antibacterial targets. Biochem Pharmacol 71(7):1066–1072
6. Conly J, Johnston B (2005) Where are all the new antibiotics? The new antibiotic paradox. Can J Infect Dis Med Microbiol 16(3):159–160
7. Barbachyn MR, Ford CW (2003) Oxazolidinone structure-activity relationships leading to linezolid. Angew Chem Int Ed Engl 42(18):2010–2023
8. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science 325(5944):1089–1093
9. Kern WV (2006) Daptomycin: first in a new class of antibiotics for complicated skin and soft-tissue infections. Int J Clin Pract 60(3):370–378
10. Heath RJ, Rock CO (2004) Fatty acid biosynthesis as a target for novel antibacterial. Curr Opin Investig Drugs 5(2):146–153
11. Jayakumar A, Tai MH, Huang WY, al-Feel W, Hsu M, Abu-Elheiga L, Chirala SS, Wakil SJ (1995) Human fatty acid synthase: properties and molecular cloning. Proc Natl Acad Sci USA 92(19):8695–8699
12. Jayakumar A, Huang WY, Raetz B, Chirala SS, Wakil SJ (1996) Cloning and expression of the multifunctional human fatty acid synthase and its subdomains in *Escherichia coli*. Proc Natl Acad Sci USA 93(25):14509–14514
13. Wright HT, Reynolds KA (2007) Antibacterial targets in fatty acid biosynthesis. Curr Opin Microbiol 10(5):447–453
14. Tsay JT, Oh W, Larson TJ, Jackowski S, Rock CO (1992) Isolation and characterization of the beta-ketoacyl-acyl carrier protein synthase III gene (fabH) from *Escherichia coli* K-12. J Biol Chem 267(10):6807–6814
15. Heath RJ, Rock CO (1996) Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in *Escherichia coli*. J Biol Chem 271(4):1833–1836
16. Lai CY, Cronan JE (2003) Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. J Biol Chem 278(51):51494–51503
17. Revill WP, Bibb MJ, Scheu AK, Kieser HJ, Hopwood DA (2001) Beta-ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in Streptomyces coelicolor A3(2). J Bacteriol 183(11):3526–3530
18. Khandekar SS, Gentry DR, Van Aller GS, Warren P, Xiang H, Silverman C, Doyle ML, Chambers PA, Konstantinidis AK, Brandt M, Daines RA, Lonsdale JT (2001) Identification, substrate specificity, and inhibition of the Streptococcus pneumoniae beta-ketoacyl-acyl carrier protein synthase III (FabH). J Biol Chem 276(32):30024–30030
19. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269(5223):496–512
20. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393(6685):537–544
21. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K (2001) Whole genome sequencing of meticillin-resistant Staphylococcus aureus. Lancet 357(9264):1225–1240

22. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV (2000) Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. Nature 406(6799):959–964
23. Heath RJ, Rock CO (1996) Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in *Escherichia coli*. J Biol Chem 271(18):10996–11000
24. Alhamadsheh MM, Waters NC, Sachdeva S, Lee P, Reynolds KA (2008) Synthesis and biological evaluation of novel sulfonyl-naphthalene-1, 4-diols as FabH inhibitors. Bioorg Med Chem Lett 18(24):6402–6405
25. Gajiwala KS, Margosiak S, Lu J, Cortez J, Su Y, Nie Z, Appelt K (2009) Crystal structures of bacterial FabH suggest a molecular basis for the substrate specificity of the enzyme. FEBS Lett 583(17):2939–2946
26. He X, Reeve AM, Desai UR, Kellogg GE, Reynolds KA (2004) 1, 2-dithiole-3-ones as potent inhibitors of the bacterial 3-ketoacyl acyl carrier protein synthase III (FabH). Antimicrob Agents Chemother 48(8):3093–3102
27. Li HQ, Luo Y, Lv PC, Shi L, Liu CH, Zhu HL (2010) Design and synthesis of novel deoxybenzoin derivatives as FabH inhibitors and anti-inflammatory agents. Bioorg Med Chem Lett 20(6): 2025–2028
28. Li HQ, Shi L, Li QS, Liu PG, Luo Y, Zhao J, Zhu HL (2009) Synthesis of C(7) modified chrysin derivatives designing to inhibit beta-ketoacyl-acyl carrier protein synthase III (FabH) as antibiotics. Bioorg Med Chem 17(17):6264–6269
29. LV PC, Wang KR, Yang Y, Mao WJ, Chen J, Xiong J, Zhu HL (2009) Design, synthesis and biological evaluation of novel thiazole derivatives as potent FabH inhibitors. Bioorg Med Chem Lett 19(23):6750–6754
30. Senior SJ, Illarionov PA, Gurcha SS, Campbell IB, Schaeffer ML, Minnikin DE, Besra GS (2004) Acetylene-based analogues of thiolactomycin, active against Mycobacterium tuberculosis mtFabH fatty acid condensing enzyme. Bioorg Med Chem Lett 14(2):373–376
31. Senior SJ, Illarionov PA, Gurcha SS, Campbell IB, Schaeffer ML, Minnikin DE, Besra GS (2003) Biphenyl-based analogues of thiolactomycin, active against Mycobacterium tuberculosis mtFabH fatty acid condensing enzyme. Bioorg Med Chem Lett 13(21):3685–3688
32. Nie Z, Perretta C, Lu J, Su Y, Margosiak S, Gajiwala KS, Cortez J, Nikulin V, Yager KM, Appelt K, Chu S (2005) Structure-based design, synthesis, and study of potent inhibitors of beta-ketoacyl-acyl carrier protein synthase III as potential antimicrobial agents. J Med Chem 48(5):1596–1609
33. Alhamadsheh MM, Waters NC, Huddler DP, Kreishman-Deitrick M, Florova G, Reynolds KA (2007) Synthesis and biological evaluation of thiazolidine-2-one 1, 1-dioxide as inhibitors of *Escherichia coli* beta-ketoacyl-ACP-synthase III (FabH). Bioorg Med Chem Lett 17(4):879–883
34. He X, Reynolds KA (2002) Purification, characterization, and identification of novel inhibitors of the beta-ketoacyl-acyl carrier protein synthase III (FabH) from Staphylococcus aureus. Antimicrob Agents Chemother 46(5):1310–1318
35. Daines RA, Pendrak I, Sham K, Van Aller GS, Konstantinidis AK, Lonsdale JT, Janson CA, Qiu X, Brandt M, Khandekar SS, Silverman C, Head MS (2003) First X-ray cocrystal structure of a bacterial FabH condensing enzyme and a small molecule inhibitor achieved using rational design and homology modeling. J Med Chem 46(1):5–8
36. Perez-Castillo Y, Perez MA (2008) Bacterial beta-ketoacyl-acyl carrier protein synthase III (FabH): an attractive target for the design of new broad-spectrum antimicrobial agents. Mini Rev Med Chem 8(1):36–45
37. Cornell W, Nam K (2009) Steroid hormone binding receptors: application of homology modeling, induced fit docking, and molecular dynamics to study structure-function relationships. Curr Top Med Chem 9(9):844–853
38. Rosales-Hernandez MC, Bermudez-Lugo J, Garcia J, Trujillo-Ferrara J, Correa-Basurto J (2009) Molecular modeling applied to anti-cancer drug development. Anticancer Agents Med Chem 9(2):230–238
39. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. Med Res Rev 26(5):531–568
40. Talele TT, Khedkar SA, Rigby AC (2009) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem 10(1):127–141
41. Qiu X, Janson CA, Smith WW, Head M, Lonsdale J, Konstantinidis AK (2001) Refined structures of beta-ketoacyl-acyl carrier protein synthase III. J Mol Biol 307(1):341–356
42. Scarsdale JN, Kazanina G, He X, Reynolds KA, Wright HT (2001) Crystal structure of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III. J Biol Chem 276(23): 20516–20522
43. Qiu X, Janson CA, Konstantinidis AK, Nwagwu S, Silverman C, Smith WW, Khandekar S, Lonsdale J, Abdel-Meguid SS (1999) Crystal structure of beta-ketoacyl-acyl carrier protein synthase III. A key condensing enzyme in bacterial fatty acid biosynthesis. J Biol Chem 274(51):36465–36471
44. Alhamadsheh MM, Musayev F, Komissarov AA, Sachdeva S, Wright HT, Scarsdale N, Florova G, Reynolds KA (2007) Alkyl-CoA disulfides as inhibitors and mechanistic probes for FabH enzymes. Chem Biol 14(5):513–524
45. Musayev F, Sachdeva S, Scarsdale JN, Reynolds KA, Wright HT (2005) Crystal structure of a substrate complex of Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III (FabH) with lauroyl-coenzyme A. J Mol Biol 346(5):1313–1321
46. Qiu X, Choudhry AE, Janson CA, Grooms M, Daines RA, Lonsdale JT, Khandekar SS (2005) Crystal structure and substrate specificity of the beta-ketoacyl-acyl carrier protein synthase III (FabH) from Staphylococcus aureus. Protein Sci 14(8):2087–2094
47. Sachdeva S, Musayev FN, Alhamadsheh MM, Scarsdale JN, Wright HT, Reynolds KA (2008) Separate entrance and exit portals for ligand traffic in Mycobacterium tuberculosis FabH. Chem Biol 15(4):402–412
48. Marco E, Gago F (2007) Overcoming the inadequacies or limitations of experimental structures as drug targets by using computational modeling tools and molecular dynamics simulations. Chem Med Chem 2(10):1388–1401
49. Durrant JD, de Oliveira CA, McCammon JA (2010) Including receptor flexibility and induced fit effects into the design of MMP-2 inhibitors. J Mol Recognit 23(2):173–182
50. Liu M, Yuan M, Luo M, Bu X, Luo HB, Hu X (2010) Binding of curcumin with glyoxalase I: molecular docking, molecular dynamics simulations, and kinetics analysis. Biophys Chem 147(1–2):28–34
51. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA (2008) AMBER 10. University of California, San Francisco
52. Mongan J (2004) J Comput Aided Mol Des 18(6):433–436

53. Humphrey W, Dalke A, Schulten K (1996) VMD–visual molecular dynamics. J Mol Graph 14:33–38
54. MATLAB (2009) R2009a edn. The MathWorks Inc
55. OpenEye Scientific Software I (2008) OMEGA, 2.3.2 edn. OpenEye Scientific Software, Inc., Santa Fe
56. OpenEye Scientific Software I (2008) QUACPAC, 1.3.1 edn. OpenEye Scientific Software, Inc., Santa Fe
57. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. RNA 15(6):1219–1230
58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
59. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688
60. Feig M, Im W, Brooks CL 3rd (2004) Implicit solvation based on generalized Born theory in different dielectric environments. J Chem Phys 120(2):903–911
61. Onufriev A, Case DA, Bashford D (2002) Effective born radii in the generalized born approximation: the importance of being perfect. J Comput Chem 23(14):1297–1304
62. Li L, Uversky VN, Dunker AK, Meroueh SO (2007) A computational investigation of allostery in the catabolite activator protein. J Am Chem Soc 129(50):15668–15676
63. Lange OF, Grubmuller H (2006) J Phys Chem B 110(45):22842
64. Teotico DG, Frazier ML, Ding F, Dokholyan NV, Temple BRS, Redinbo MR (2008) Active nuclear receptors exhibit highly correlated AF-2 domain motions. PLoS Comput Biol 4(7): e1000111
65. Bradley MJ, Chivers PT, Baker NA (2008) Molecular dynamics simulation of the *Escherichia coli* NikR protein: equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. J Mol Biol 378(5):1155–1173
66. Davies C, Heath RJ, White SW, Rock CO (2000) The 1.8 A crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. Structure 8(2):185–195
67. Shao J, Tanner SW, Thompson N, Cheatham TE (2007) Clustering molecular dynamics trajectories: 1. characterizing the performance of different clustering algorithms. J Chem Theory Comput 3(6):2312–2334
68. Mitchell JB (2001) The relationship between the sequence identities of alpha helical proteins in the PDB and the molecular similarities of their ligands. J Chem Inf Comput Sci 41(6):1617–1622
69. Nobeli I, Spriggs RV, George RA, Thornton JM (2005) A ligand-centric analysis of the diversity and evolution of protein-ligand relationships in *E. coli*. J Mol Biol 347(2):415–436
70. Kalinina O, Gelfand M, Russell R (2009) Combining specificity determining and conserved residues improves functional site prediction. BMC Bioinformatics 10(1):174