Special Collection

Ligand- and Structure-Based Approaches of *Escherichia coli* FabI Inhibition by Triclosan Derivatives: From Chemical Similarity to Protein Dynamics Influence

Thales Kronenberger+,*[b, c] Philipe de Oliveira Fernades+,[a, d] Isabella Drumond Franco,[a] Antti Poso,[b, c] and Vinícius Gonçalves Maltarollo*[a]

Enoyl-acyl carrier protein reductase (FabI) is the limiting step to complete the elongation cycle in type II fatty acid synthase (FAS) systems and is a relevant target for antibacterial drugs. *E. coli* FabI has been employed as a model to develop new inhibitors against FAS, especially triclosan and diphenyl ether derivatives. Chemical similarity models (CSM) were used to understand which features were relevant for FabI inhibition. Exhaustive screening of different CSM parameter combinations featured chemical groups, such as the hydroxy group, as relevant to distinguish between active/decoy compounds.

Those chemical features can interact with the catalytic Tyr156. Further molecular dynamics simulation of FabI revealed the ionization state as a relevant for ligand stability. Also, our models point the balance between potency and the occupancy of the hydrophobic pocket. This work discusses the strengths and weak points of each technique, highlighting the importance of complementarity among approaches to elucidate *Ec*FabI inhibitor's binding mode and offers insights for future drug discovery.

Introduction

The increase of antimicrobial resistance has become a global healthcare problem, rendering obsolete many antibiotic therapies.[1] There is, therefore, an urgent medical need for new antibacterial drugs, especially with novel mechanisms of action that would display minimal cross-resistance with currently used treatments. The need for new drug targets leads to the use of

[a] P. de Oliveira Fernades,+ I. Drumond Franco,
    Prof. Dr. V. Gonçalves Maltarollo
    Departamento de Produtos Farmacêuticos, Faculdade de Farmácia
    Universidade Federal de Minas Gerais
    Av. Antônio Carlos, 6627 Pampulha, Belo Horizonte, MG, 31270-901 (Brazil)
    E-mail: maltarollo@ufmg.br

[b] Dr. T. Kronenberger,+ Prof. Dr. A. Poso
    Department of Medical Oncology and Pneumology, Internal Medicine VIII
    University Hospital of Tübingen
    Otfried-Müller-Strasse 14, 72076 Tübingen (Germany)
    E-mail: thales.kronenberger@uni-tuebingen.de

[c] Dr. T. Kronenberger,+ Prof. Dr. A. Poso
    School of Pharmacy
    University of Eastern Finland
    Faculty of Health Sciences, Kuopio 70211 (Finland)

[d] P. de Oliveira Fernades+
    Departamento de Química, Instituto de Ciências Exatas
    Universidade Federal de Minas Gerais
    Av. Antônio Carlos, 6627 – Pampulha, Belo Horizonte, MG, 31270-901 (Brazil)

[+] These authors equally contributed to this work

    Supporting information for this article is available on the WWW under https://doi.org/10.1002/cmdc.201900415

    This article belongs to the Special Collection "BrazMedChem 2019: Medicinal Chemistry in Central and South America.

bacterial fatty acid synthase type II as a possible biological source (FAS-II). Prokaryote FAS-II system is based on individual enzymes, and it is different from the multifunctional fatty acid synthase type I system found in eukaryotes, therefore providing good prospects for a selective inhibition.[2] However, the use of FASII system as a source of antibiotic targets for Gram-positive pathogens have been challenged in recent years. FASII pathway is considered essential for *Staphylococcus aureus*, while recent evidence also shows that pathogens could incorporate exogenous fatty acids, therefore bypassing the FASII inhibition.[3–5] It has been shown also that enoyl-acyl carrier protein reductase (FabI) is essential for *in vivo* survival of *S. aureus*, as FabI inhibitors supported animal recovery in mice infections models.[6]

FAS-II last step is the fatty acid elongation cycle, which is commonly catalysed by enoyl-ACP reductases and, specifically, FabI catalyses the reduction of the trans-2-enoyl-ACP double bond.[7] Most FabI enzymes catalyse the substrate reduction through an ordered bi-bi mechanism, in which the binding of the cofactor NAD(P)H precedes the substrate binding[7] (Figure 1A). Among them, FabI is the major isoform in pathogens such as *S. aureus*,[8] *Escherichia coli*[9] and *Mycobacterium tuberculosis*[10] (also called InhA). FabI is a target for the broad-spectrum antibacterial triclosan (TCL, compound 1), as determined from resistant *E. coli* strains with a mutation in this gene.[11] Triclosan was further characterized as a reversible inhibitor of *E. coli* FabI[12] and it has a consistent use despite the intravenous toxicity and spread resistance.[13] TCL binding to the active site of *Ec*FabI reorganizes the amino-acid residues on the loop (residues 191–205) into a slow binding mechanism with long residence time.[14]

TCL inhibitory mechanism can be used as an example for the entire class of competitive FabI inhibitors, where essential
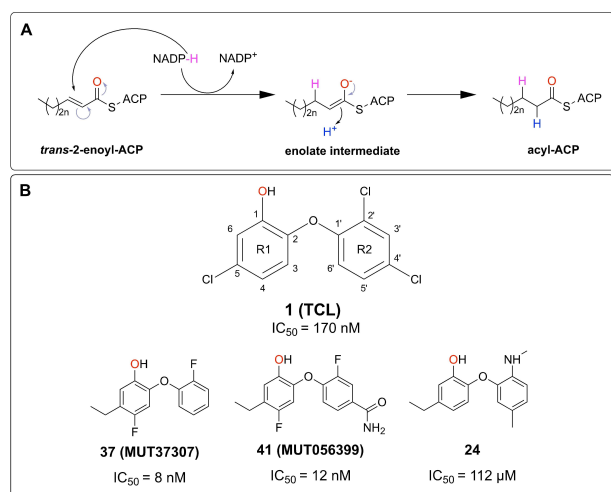
Figure 1. A) proposed enzymatic mechanism for FabI catalyses, the double bond reduction occurs by conjugate addition of a hydride ion from NAD(P)H leading to a rearrangement that generates the enolate intermediate, that attacks the proton from the active side residues B) Gold standard compound triclosan (1) employed as scaffold for the development of the studied series followed by the three most potent compounds generated after SAR studies and employed in our model generation. TCL image highlights the numbering system employed here to discuss the ring substituents.

features, such as the presence of an aromatic ring and phenolic hydroxyl group can be modified to improve the biological activity. The aromatic ring can be involved in π-stacking interactions with the aromatic nicotinamide moiety of NAD$^+$, while the hydroxyl group interacts with Tyr156 and the ribose moiety of NAD$^+$ in the active site.[12]

FabI belongs to an enzymatic class known to present substrate promiscuity.[15] Additionally, FabI active site is covered by a highly flexible loop, which has been suggested to reduce the efficacy of docking as a tool for drug discovery.[16] Successful drug discovery projects against FabI/InhA relied on either the combination of docking with other computational technique (molecular dynamics or quantum chemical/mechanical calculations) or on the extensive synthesis of analogues.[17,18] In accordance with the ever-increasing ligand information, ligand-based approaches could be employed[16,18] and are often employed in parallel to structure-based drug design approaches to reach similar and complementary findings.[19]

Despite the clinical success of the InhA isoniazid inhibitor[20] and the FabI inhibitors that underwent clinical trials, none of them has made it yet to market.[21–23] Recently, however combinatorial treatment approach with daptomycin has been proposed, which can breathe new air into the field of FASII inhibitor development. Daptomycin is a potent last-resource antibiotic acting directly on the gram-positive bacteria's membrane. Also, its resistance mechanism has been attributed to the release of decoy lipids from the pathogen that would interfere with the compound's membrane interaction.[24] In this context, *S. aureus*, when infecting host tissues, may depend upon endogenous phospholipid biosynthesis to generate the decoys lipids for release.[25] Additionally, AFN-1252, which is a potent FabI inhibitor, was shown to efficiently block daptomycin-induced phospholipid decoy production. Interestingly,

AFN-1252 resistant isolates, were still vulnerable to the double treatment with AFN-1252 and daptomycin, which provides a niche for a double-punch antibacterial targeting, ultimately showing that FASII inhibition is still on the vogue for drug discovery.

In the last years, we observed a great advance in the discovery and development of FabI inhibitors, some of which reached clinical tests. That scenario reflects the clinical and commercial interest of the development of new FabI inhibitors as new antibacterial. In this sense, there is a broad range of computer-aided drug design (CADD) methods, which can use a set of validated bioactive compounds to get insights into the protein function and help to design more potent/specific agents. The integration of different CADD techniques is essential to overcome the inherent limitations of each of them and, at the same time, efficiently use the available computational resources. Specific compound datasets with well-characterised activities against targets with available structural information are excellent models for employing both ligand-based and structure-based computational drug discovery, respectively.

Therefore, with the aim of better understanding the influence of specific chemical features on the FabI inhibition by diphenyl ether derivatives, within this work, we did an exhaustive study employing a variety of chemical similarity models. Further structural interpretation of the identified features related to the inhibition mechanism of FabI was also validated by molecular modelling, by the means of molecular dynamics simulation (Figure 2). Altogether, the influence of chemical features from known inhibitors and their interactions within *Ec*FabI binding site residues enabled helpful insights to drug discovery. Ultimately, this work discusses the strengths and weak points of each technique, highlighting the importance of complementarity among different approaches to elucidate inhibitor's binding mode and further compound development.

Results and Discussion

Chemical Similarity Models

Initially, the first three chemical similarity models built using all features of template molecules presented AUC values above 0.9
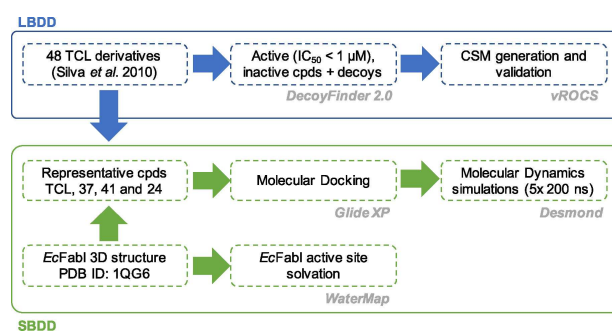

Figure 2. Flowchart illustrating the combination of different techniques from ligand-based drug design (LBDD) and structure-based drug design (SBDD) approaches employed together in this work.

Table 1. Results of validation metrics for the first three models.

Template	AUC	Confidence levels		Enrichment factor at		
		−95%	+95%	0.5%	1.0%	2.0%
37	0.991	0.982	0.997	116.3	81.9	45.7
41	0.995	0.991	0.998	129.7	85.9	49.9
48	0.937	0.875	0.984	107.2	75.7	39.6

(Table 1), indicating a satisfactory result for the unedited models. CSM derived from compound 41 has better statistical values when compared to its counterparts (Table 1). Enrichment values also indicated the better ability of the model derived from compound 41 to distinguish between active and decoys (Table 1). Figure 3 (A) enumerates the chemical features evaluated along with the different 3D chemical similarity models.

At the following step, 445 new CSMs were built by excluding systematically the features followed by evaluating the ability of them to distinguish active molecules from decoys in a ROC curve validation (Figure 3B). The calculated AUC values range from 0.762 to 0.996. The main reason for this result could be the high similarity between the models, the template and the dataset molecules. Seven models presented the AUC value equal to 0.996 and all of them were derived from the

compound 41 as a template, indicating the highest predictability of CSMs built with this template. Similar results could be observed on the enrichment factors (on the top 1 and 5% ranked poses, Supporting Information, Figure 1).

This first analysis indicated that both H-bond acceptor and donor from the amide function at the side chain of the template 41 are relevant to distinguish actives from decoys. Following, the presence of the hydrophobic group, H-bond donor from phenolic hydroxyl and H-bond acceptor from phenoxy group is related to increased AUC values. In contrast, the presence of an H-bond acceptor from pyridinic N atom decreased the AUC value, suggesting a negative influence on the classification ability of the biological activity. Table 2 presents a summary of the statistical parameters evaluated in the validation of the ten best ROC models.

The similarity values calculated from the seven models with the highest AUC values were also employed in a confusion matrix calculation (Table 2). The Tc value of the first decoys screened by CSM was employed as a cut-off to classify the compounds as active (TC higher than cut-off) or inactive. After this step, five models with the highest MCC values (models 253 and 301–304, Supporting Information, Table S1) were chosen to perform a weight applying on features. Weights equal to 2, 3 and 5 times in single, double and triple combinations were performed to build new 443 models. The validation of CSMs with feature weighting indicated that 93 models have AUC equals to 0.996. Then, we ranked it according to its enrichment factor at 0.5% and choose the top five models to construct a confusion matrix (Table 2)

In general terms, this step did not produce statistically better results, just a slight increase of the early enrichment factors (at 0.5% of the dataset) and MCC values but decreased true positive rate. However, those models were also employed in statistical analysis to interpret the importance of related features. From the comparison of the AUC values of all models, it is possible to note that some chemical features, namely HF, HBA1, HBD1, HBA2 and R2, were significant only when used with increased weight (Figure 3C). Surprisingly, R1 was not relevant for the differentiation between actives and decoys, despite its proximity to biologically relevant residues, as discussed in later sections. The fine-tuning of weight parameters has an overall positive impact (see Figure 3C), which is statistically significant. Furthermore, the best model (302) was also validated without decoys, in other words, only using the experimentally inactive compounds and, as expected, the calculated metrics decreased but all values suggest acceptable predictability (Supporting Information, Figure 2 and Table S2). Finally, for future potential virtual screening purposes, model
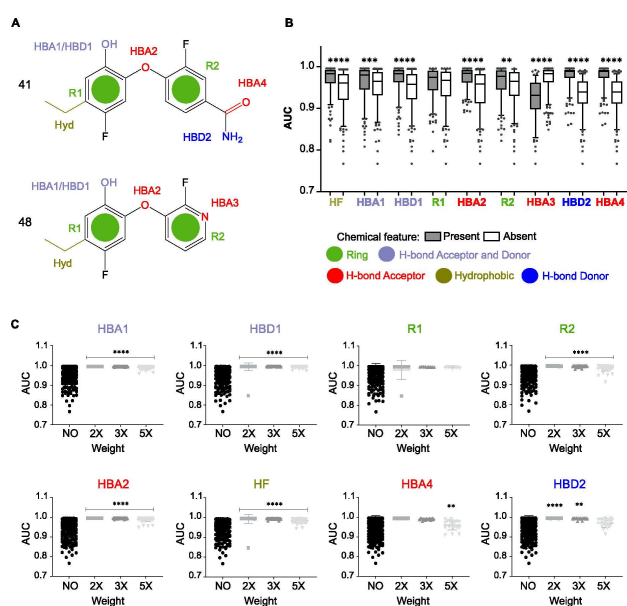

Figure 3. Overview of the chemical similarity queries and validation results. A) 2D representation of the chemical features evaluated along with the different 3D chemical similarity models. Compounds 41 and 48 (the most active of the series) were employed as a template for the model generation and features were consistently numbered for comparison purposes. B) Models with systematically screened chemical features. Models in the presence (grey) and absence (white boxes) of specific chemical features were compared in terms of AUC. Boxes represent the distribution around 5–95 quantiles, while dots are models with values considered as outliers in this interval. C) Models with differently weighted chemical features were compared to models without this feature in terms of AUC. Statistical significance between models and the control without the chemical feature was evaluated using Mann-Whitney non-parametric model and the p-value is represented as follow: ****<0.0001, ***<0.001, **<0.01.

Table 2. Summary of the statistical parameters evaluated in the validation of the ten best ROC models followed by the statistical parameters evaluated in the validation of the ten best ROCS varying the weight of chemical features. Information regarding all calculated models is provided in the Supporting Information (Table S1). Data from the best model is underlined. AUC: area under the ROC curve. E. (0.5%) stands for the enrichment factor on 0.5% (data for 1% and 2% enrichment is provided with full models as Supporting Information) TC1FP*: TanimotoCombo value for the first decoy predicted as false positive. TPR: true positive rate. TNR: true negative rate. MCC: Matthews correlation coefficient. ACC: accuracy.

Model	AUC	E. (0.5%)	Tc Cut-off	TPR	TNR	ACC	TC1FP*	MCC
Best ROC models (fixed weight)								
253	0.996	138.7	1.33	0.83	0.99	0.99	0.81	0.80
301	0.996	141.3	1.25	0.9	0.99	0.99	0.81	0.81
302	0.996	133.7	1.2	0.97	0.99	0.99	0.83	0.83
303	0.996	137.0	1.16	0.93	0.99	0.99	0.79	0.79
304	0.996	130.6	1.17	0.9	0.99	0.99	0.79	0.79
313	0.996	133.1	1.21	0.93	0.99	0.99	0.78	0.78
317	0.996	128.6	1.34	0.93	0.99	0.99	0.79	0.79
Best ROC models (weight variation)								
589	0.99	155.1	1.73	0.83	1	0.99	0.85	0.84
875	0.996	154.8	2.59	0.57	1	0.99	0.69	0.71
612	0.996	154.6	1.73	0.83	1	0.99	0.85	0.84
824	0.996	151.2	2.09	0.7	1	0.99	0.76	0.76
505	0.996	149.2	1.91	0.87	0.99	0.99	0.79	0.79

302 shows the best balance between validation parameters as AUC, TPR, TNR, ACC and MCC, as well as the highest true positive rate value. The use of generated CSM in the discovery of new FabI inhibitors could provide new chemotypes since its usage in the search of new bioactive compounds is based on the arrangement of features into space (and not a search for a specific functional group).

Molecular Modelling

Molecular docking studies followed by molecular dynamics simulation of selected analogues (compounds 41, 27, 24 and TCL) were conducted to complement the interpretation of the chemical features in a structural context. Triclosan and compound 41 were simulated independently in both ionization states, the charged and neutral phenolic group, due to the calculated pKa values (8.06 and 8.26 respectively). Accordingly, compounds 37 and 24 have calculated pKa in the order of 9.17 and 10.88 (Supporting Information, Figure 3), respectively, which discouraged the simulation of charged states, since they would represent less than 2% of the total sample, for the most conservative measurement. Comparison between experimentally determined pKa values and our predictions showed adequate correlation (Supporting Information, Figure 3), which incentivized the use of the predictions determined in solution to derive further conclusions.

Due to the highly flexible loops covering FabI's active site, as already reported by other groups and corroborated by high B-factor values,[16] (Figure 4A and Supporting Information, Figure 4), further molecular modelling was based not only in the interpretation of docking poses (Figure 4B shows the redocking result), but rather discussed the interactions along molecular dynamics simulations. On one hand, TCL and compound 41 ionized forms kept similar poses as the initial states (Figure 4C–F), however, compounds 24 and 37 heavily relied on hydro-

phobic interactions to maintain binding stability (Figure 4G,H). Furthermore, calculated poses with Glide were corroborated by Surflex Docking[26] and GOLD 5.1[27] software (Supporting Information, Figure 5) except for compound 24. Molecular docking on different EcFabI structures led to similar docking results, where compounds TCL, 41 and 37 presented classical hydrogen bond interactions with Tyr156 and NAD⁺ (Supporting Information Figure 6–7) and similar orientation to the co-crystallized ligand.

Triclosan's binding mode reproduced commonly described interactions such as hydrogen bond between the hydroxyl moiety and Tyr156/NAD⁺ and the π-π interaction of the ring 1 with both Tyr146 (T-shaped) and the NAD⁺ cofactor (π-stacking) (Figure 5A–C), but not an ionic interaction with Lys163 (Figure 5B). However, only anionic states of triclosan (Figure 5D–E) and compound 41 (Figure F) maintained a hydrogen bond with Tyr156, while the neutral forms drifted after few nanoseconds of simulation (Supporting Information's Figure 4 and 8 highlights the conformational changes FabI underwent due to the compound displacement in the simulations with the compound neutral states by the root-mean-square deviation and fluctuation, respectively).

On the other hand, compound 24 (inactive) preferred the interaction with Tyr146 after the minimization and did not remain stable within the binding pocket. This pattern of recognition of the phenolic group by the twin tyrosine residues is suggested to be responsible for the activity since changes from hydroxyl to methoxy groups are known to lead inactive compounds on enzymatic assay.[28] Also, the ionized form of Compound 37 (the most active of the series), stablished a stable ionic interaction between its phenolate and the Lys163 (Figure 5B), which was seldom observed in the TCL/41 anionic counterparts. Chang et al., 2013 hypothesized that the oxyanion forms of diphenyl-ethers contribute to the high affinity binding to SaFabI, where the charge of Lys164 (homologue to Lys163 in
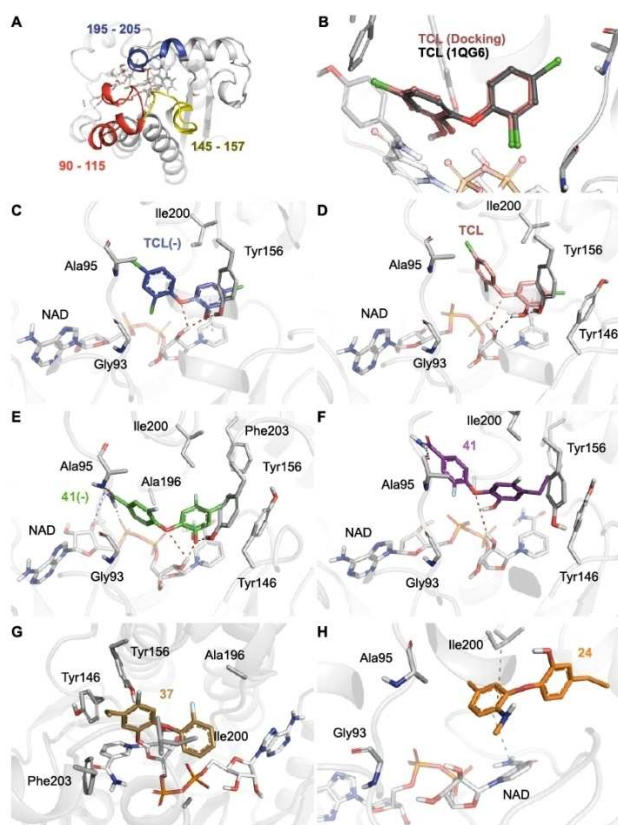
Figure 4. A) The three main loops of *Ec*FabI (PDB ID: 1QG6), residues 90–115 (in red) encompassing the flexible lid region, residues 145–157 (dark yellow) where the active site twin tyrosine residues are presented, and residues 195–205 (blue, 191–200 is also annotated as substrate binding loop or SBL), which represents an hydrophobic region that covers the active site. B) redocking of TCL within the active site. Representative structure of the last frames from the molecular dynamics simulations of with different compounds: charged (C) and neutral triclosan (D), charged (E) and neutral (F) compound 41, neutral compound 37 (G) and 24 (H). Interactions are represented by dashed lines as follow: the π-π interactions are coloured in light blue, hydrophobic in dark yellow, hydrogen bond acceptors in red and hydrogen bond donor interactions in dark blue. FabI's residues are coloured according to the atom types of the interacting amino-acid residues (protein's carbon, light grey; nitrogen, blue; oxygen, red; phosphorus, orange).

Figure 5. Summary of hydrogen bond interactions and π-π interactions frequencies with main residues within the active site. (A) Summary of the hydrogen bond interaction between the hydroxyl group and residue Tyr156 and also between the Ala95 and both HBA4/HBD2 chemical features. B) ionic interaction profile with Lys163. C) Summary of the π-π interactions among the aromatic rings within the active site: Phe203, Tyr146 and 156. D) Hydrogen bond interaction distance between HBA1 and Tyr156 hydroxyl group along the simulation for the TCL in both ionization states, neutral (in red) and charged (in blue), followed by an exemplary snapshot of the MD simulation (E), same analyses were performed for both ionization states of compound 41 (F), with charged form in green and neutral form in purple. Each dark coloured line represents the average distance of the five independent simulations and the respective light coloured to represent the observed standard deviation. G) Hydrogen bond interaction between Ala95 main-chain hydrogen and the compound 41 amide's oxygen, followed by a representative snapshot from the simulation (H). Residues are coloured according to the atom types of the interacting amino-acid residues (protein's carbon, light grey; nitrogen, blue; oxygen, red; phosphorus, orange), polar interactions are represented by dashed yellow.

E. coli) would assist in the stabilization of the hydrogen bond network with the Tyr157 and the oxidized cofactor $NADP^+$.[29]

Clearly, the ionized forms showed consistent interactions with Tyr156 and NAD^+ when compared to the respective neutral states or neutral compounds, which suggests a direct relation between biological activity and the phenolic group pKa values. This can be further corroborated by the lesser stability at the binding site of the neutral ones. The equilibrium between the charged and neutral species in solution may be influenced by the better binding of ionized compounds, which would be trapped in the FabI active site, despite its lower amount in the biological environment. Since both the ionization process and protein binding are driven by its respective equilibrium constants, the retained charged inhibitors within the protein would shift the ionization equilibrium in solution by reducing its product.
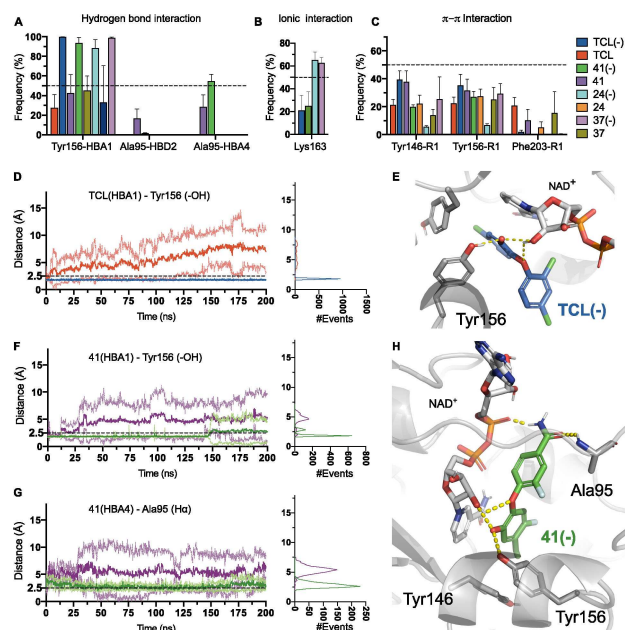
During simulations ring 1 featured π-π interactions with both Tyr146 and Tyr156 (Figure 5C). Charged states of active compounds have closer interaction with Tyr156, where simulations of compound 24 lead to unstable interactions. Compound 24 inactivity could be explained by the predominance of unionized form caused by the substitution of halogens (electron withdrawing groups) for alkyl groups (electron donor groups) and its respective absence of these interactions (Figure 5G). This observation could be related to the stabilization of H-bonds between compounds and Tyr156 and NAD^+ due the hydrogen bonding dependence of suitable distance and angle between the donor, the hydrogen atom and the acceptor.

On the other side of the molecule, the TCL's chlorine substituents at ring 2 (position 2', Figure 1B for reference) were initially suggested to form halogen bonds with the main chain of Ala95 (Figure 4C–D). The amide of the compound ionized 41 (HBA4/HBD2) interacts both with the main chain of Ala95 and phosphate group of NAD^+ (Figure 4E–F and Figure 5C, H), however the same was not observed for the neutral state. We
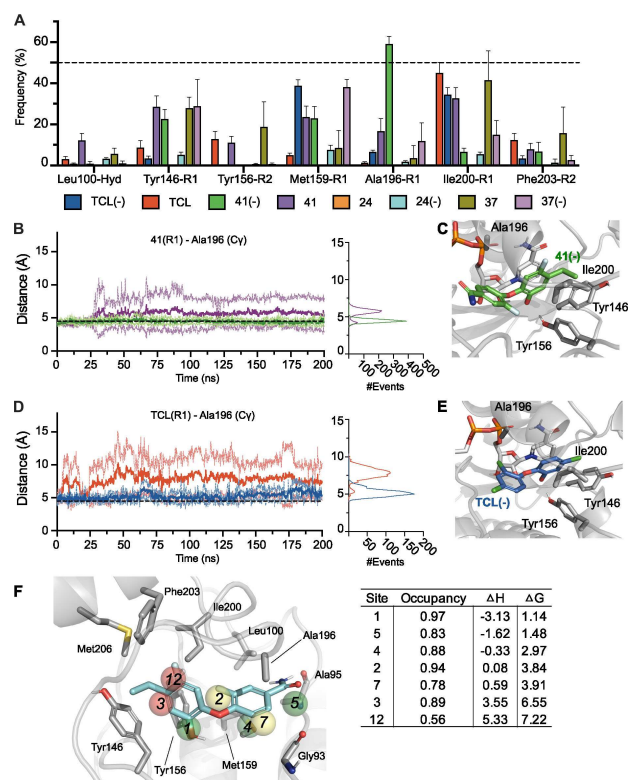
Figure 6. Hydrophobic interactions along with the molecular dynamic simulation. A) most frequent hydrophobic interactions are described for each analysed compound and different moieties. Each bar represents the standard error of five independent simulations. Specifically, the interaction between residues Ala196 and the ethyl moiety of compound 41, which remains stable along the simulation (B for distance along the simulation and C highlighting a representative snapshot) with higher frequency for the deprotonation ionized state (green, and with the ring (R1) of TCL (D and E – for representative snapshot). Each deep coloured line represents the average of five independent simulations, while the respective lighter tone represents the standard deviation. F) Superimposed *Ec*FabI side-chains, compound 41 and conserved water molecules are presented coloured spheres, according to their free energy value (ΔG), also described in the adjacent table. The table contains the thermodynamic parameters for the solvation of the seven hydration sites, within the compound pocket. Green spheres represent regions where stable water molecules could be placed and therefore less likely to contributed to free energy gain upon ligand binding, alternatively, red and orange spheres represent regions where ligand occupancy could contribute to binding affinity by enthalpic energy gain. Occupancy is calculated from the number of water-oxygen atoms found occupying a given hydration site during the 2 ns of molecular dynamics simulation, enthalpic energy (ΔH) and free energy value (ΔG) are given in kcal/mol.

hypothesize, based on chemical similarity, that the pyridinic nitrogen atom of the compounds 46 to 48 could form a direct hydrogen bond with the main chain atoms of Gly93 or Ala95. Interestingly, the vicinal substituents of the N-pyridine (ring 2, position 2′, see Figure 1B) have a striking effect over the activity, ranging from highly active compounds, such as 48 (−F) and 46 (−NH₂), towards completely inactive (47, =O). Experimental evidence from Gly93Val mutants, which are TCL resistant, corroborates the importance of this interaction to the structure stabilization.[30]

Initial attempts to modify the ring 1 chlorine towards more hydrophobic moieties improved both enzyme affinity and

microbicide activity,[28] which were initially attributed to similarity with the natural hydrophobic substrate. Modelling results suggested that the hydrophobicity of the ethyl substituent at ring 1, in the compounds 37 and 41 (Figure 4C–E) can occupy the hydrophobic back-pocket composed by Ile200, Phe203 and Met206 (Figure 6A and F).

Slow binding inhibitor association to the protein can be described in one- or two-step process one-step processes. Briefly, an one-step inhibitor association accounting for slow formation of the enzyme-inhibitor complex (EI) by overcoming the energetic barrier of the transition state (TS). Alternatively, a special type of two-step process involves rapid formation of EI, which is then followed by a slow induced-fit in the structure of complex (EI*), which for FabI would be related to the SBL stabilization. FabI is suggested to have a special kind of two-step mechanism that is kinetically indistinguishable from the one-step mechanism,[29,31] in which the free energy of EI* is much lower than that of EI and the initial formation of EI cannot be detected at low inhibitor concentrations.

The increase in residence time can be described in terms of either by the stabilisation of the EI/EI* or the disruption of the TS. It has been shown that hydrophobic 5-substituents can enhance residence time and affinity of diphenyl-ethers in both *Sa*FabI[29] and *Bp*FabI1[31] by both of those mechanisms. We could suggest that this residence time and affinity gain rise from the displacement of high energy water molecules (Figure 6F, sites 3 and 12) by the hydrophobic moieties in those compounds.

Interestingly, the phenolate ring and the ethyl substituent occupies the region near the Phe203, Tyr156 and Ile200, in which we observed hydration sites with high free energy, suggesting that occupying those sites would contribute to ligand binding energy (Figure 6F). This pocket was suggested to be responsible for the fatty acid orientation during the catalysis.[32] Complementarily, compound 11 (Supporting Information, Table S3) has a hydroxyethyl moiety that reduces the local hydrophobicity of the ligand at this pocket, which could explain its inactivity. Longer acyl chains were suggested to address those pockets, however, there is lack of refined structural information on that region since residues between Lys201–Arg218 could not be solved in the respective crystal structures.[33] The work of Vani and Palermo *et al.*,[34,35] discussed the importance of paired aromatic rings for substrate specificity in lipid processing enzymes, by acting as gates that can dynamically change between a substrate accepting state towards a the catalytic closed state. Similarly, in our systems we could observe that Tyr146 and Phe203 assume a "closed" gate conformation upon the inhibitor binding (Supporting Information, Figure 9), without, however transitioning towards the open one. We hypothesize that, to fully understand the particular mechanism of Tyr146/Phe203 gating and its transitions, longer simulations with the natural ligand would be required, which are intended for future works.

On the Comparison Between CSM and MD Results

Chemical similarity models (CSMs) have a range of applications and can be employed in early drug discovery pipeline with low computational costs, enabling the screening of very large databases. However, CSMs often do not take into account the contributions of protein interactions. We here employed an exhaustive screening of parameter combinations to construct reliable CSMs, which were used to classify a series of TCL derivatives according to the *Ec*FabI inhibition capability.

Hawkins *et al.* pointed out that the assignment of different weights to the chemical features (represented by different colours), therefore changing the combo score, could improve the performance in virtual screening.[36] Additionally, models' enrichment can either be over[37] or underestimated[38,39] by inappropriate decoy selection.[40] Specifically, DUD-E's decoy set was shown to overestimate AUC values when used to validate machine learning-based models.[41] Herein, we followed the idea from Huang *et al.* (2006), which states that a decoy compound must be physiochemically similar to the known active compound, but structurally dissimilar. As CSMs calculate the similarity between a compound and a query defined by the molecular volume and some chemical features distributed in specific positions on space, we assumed that dissimilar compounds in decoy set may not interfere in the model's evaluation. Accordingly, the lack of chemical diversity in active compounds set is a major limiting step for the model training and can become a source of bias to the model.[40] Then, for this work, all analysed subsets (active, inactive and decoys compounds) were similar in terms of physicochemical properties to avoid bias. Our intention was not to compare different decoy generation methods, but given the presented metrics, to use the models' specificity to evaluate statistically significant chemical features and develop a structure-activity relationship. Similarly to our employed strategy, other works have previously employed DUD datasets in the validation of chemical similarity models.[43,44]

CSMs without the HBA1/HBD1 chemical feature were rendered unselective towards *Ec*FabI enzyme activity, which is not surprising since phenolic hydroxyl moieties fitting this region are responsible for the interaction with Tyr156 and NAD$^+$. Simulations of the compound 24, as an example of an inactive compound, showed the instability of this interaction. This observation agrees with data in the literature that indicate *Ec*FabI Tyr156Phe mutants have decreased affinity for triclosan.[45] Comparatively, InhA (the respective homologous from *Mycobacterium tuberculosis*) is poorly inhibited by TCL, mainly due to the presence of Phe156, instead of a tyrosine. Chemical similarity models were limited by the ionization state provided considering only the major species, presumably, being unable to distinguish between the hydrogen acceptor or donor character of the hydroxyl (HBA1/HBD1) and classifying both as significantly relevant. However, our data strongly indicate that the ionization state of compounds has an important effect on the ligand stability in longer time scale simulations. It is essential to highlight that crystal structures can provide a static idea of the inhibitor's binding modes offering snapshots of the different protein states, which, especially in highly flexible complexes prone to induced-fit effects, can be insufficient to explain the inhibitory mechanism.

The current hypothesis for the ACP substrate binding to FabI, relies on FabI-ACP complex models generated via MD simulations. This model pointed out the importance of the flexible charged side-chains on FabI's substrate binding loop, such as Lys201, Arg204 and Lys205.[46] Additionally, the model shows the crotonyl thioester of ACP near to Tyr146 hydroxyl group, which highlights the importance of this residue for substrate reduction. Interestingly, in our simulations, Tyr156 is responsible for the main interactions, with Tyr146 playing a support role for the binding.

Yang and collaborators (2017) have already analysed the structural potential of triclosan derivatives by interpreting 50 ns long MD simulations with several TCL derivatives.[47] They have shown that the loop covering the active site can assume structured conformation upon ligand binding. Specifically, the residues Leu195–Ser198 stabilized this loop as an α-helix, by the compounds 17 and 18, while the compounds 22 or 23 (for a detailed description of chemical structures see Supporting Information, Table S3) led to the formation of both 3–10 and α-helices. Additionally, in the same work, it was reported the absence of π-π interaction between the inactive compound 23 and the twin tyrosine residues, which partially agrees with our observations regarding the stabilisation importance (Figure 5B). According to Yang,[47] mostly hydrophobic residues favourably contributed to the binding energy, namely Leu100, Met159, Ala196, Ala197, Ile200, Phe203. This amino-acid set agrees with our CSM's suggestion that hydrophobic chemical features are relevant to discern differences in the biological activity. Complementarily, we also observed that compounds with hydrophobic substituent (37 and 41) occupied back-pocket composed by Ala196 and Ile200, maintaining van der Waals contacts, which we suggest increased the binding stability in longer simulations (Figure 6A). Additionally, these residues have been previously characterized on an enzymatic level with mutants, showing that exchanges of Met159Thr and Phe203Leu significantly reduced the enzyme affinity towards triclosan, while the replacement of Ala197Met had no impact.[45]

Simulations of *Ec*FabI with compound 41 showed a unique preference towards Ala196 for hydrophobic interactions (Figure 6B), which remains to be biochemically exploited for this particular enzyme. Interestingly, studies with FabI1 Enoyl-ACP Reductase from *Burkholderia pseudomallei* have shown that α-helix 6 (the so-called substrate binding-loop, represented by the residues Thr194–Gly199, original numbering from *Bp*FabI1), becomes ordered and closed upon ligand binding.[31] In our *Ec*FabI studies, also a flexible loop (comprised by the residues 191 to 205) covers the active site, as mentioned previously, which contains the key residues related to the inhibitor interaction (Figure 4A), and were kept ordered during most of the simulations with the exception of compound 24 and the neutral forms of compound 41 and TCL (Figure 4C–H and Supporting information, Figure 4, showing the comparison between the ligand-free and bound systems' RMSD values).

Kinetic and structural studies with the compound 41 related the effect of 5-ethyl substituents with the increases in the

residence time, which presumably is related to the stabilization of the ground state's protein conformation.[31] Additionally, three-dimensional structures of *Bp*FabI1 co-crystallised with different TCL analogues showed that Phe203 rotates toward Leu207 displacing Ala197 (equivalent to the Ala196 of *Ec*FabI), upon ligand binding, which creates a hydrophobic pocket that accommodates the 5-ethyl group. Complementarily, in *Ec*FabI, the Ile200 and Phe203 amino-acids compose the beginning of the short loop (near the α-helix-8) and end, which surrounds the ethyl moiety in our simulations. In this study, we observed that interactions with Ile200 and Met159 are also relevant, especially when the most active compound 37 is concerned, which could be explained by this pocket's remodelling.

Interestingly, chemical features originated from the 4'-amide group (Figure 1B for orientation), originally 4'-chlorine group on TCL, had divergent interpretations on CSMs. Both HBD2 and HBA4 are relevant for the obtained models to discriminate the actives from decoys, which could be explained by the hydrogen bond interactions with Ala95. However, while HBD2 contributed to the model selectivity, HBA4 only showed significant changes when highly weighted. We have observed in our simulations that HBD2 can interact with NAD^+'s phosphate (Figure 5G) or indirectly with Gly93 by a conserved water network, both pertinent for compound interaction. HBA4, represented by the amide's oxygen atom, showed stable interaction with Ala95's Hα in our 41(−) simulations, and also superimposes with interesting hydration sites (Figure 5F), which leads to the suggestion that HBA4 can contribute towards binding affinity. CSM ignoring the induced-fit effects of compound binding and solvation effects are the main limitation of the technique and, therefore it is natural that divergent points may arise when compared with MD results.

Conclusions

Our models highlighted the importance of ethyl moieties (position 5) for the occupancy in the hydrophobic pocket containing Ile200, Phe203 and Ala196, as well the interaction of hydrogen acceptor/donor groups (position 4) with Ala95 by a halogen bond/hydrogen bond, which greatly contributed to the compound stability. Furthermore, results from both ligand- and structure-based strategies partially converged indicating that simple and fast techniques such as three-dimensional similarity models could be successfully employed in the identification of potential new *Ec*FabI inhibitors in virtual screening simulations. As molecular docking followed by molecular dynamics analyses were employed successfully to explain differences of activity of representative compounds of dataset, those two techniques could be used in the refinement of the hit selection of future virtual screening campaigns by analysing the main ligand-target known interactions as well as its stability along simulation time, mainly due to the importance of induced-fit effects for this class of enzyme. The design of new antibacterial targeting FabI, especially where diphenyl-ethers are concerned, can benefit from understanding the structure and dynamics of transition state, since many available drugs act as TS mimetics.

Experimental Section

Dataset

Forty-eight diphenyl ether derivatives with experimental values of an inhibitory concentration of *Ec*FabI (IC_{50}) were employed in this study[28] (Figure 1B displays the representative dataset compounds). Aiming the validation of the constructed models, the compounds were classified either as active ($IC_{50} < 1.0\ \mu M$) or inactive (all ones).[48] The structures of all dataset compounds, as well as its IC_{50} values, are listed in Supporting Information, Table S3. The 3D structures of all compounds were generated with Discovery Studio 2017[49] and had their ionization states calculated using QUACPAC 1.7.0.2.[50] Subsequently, 30 lowest energy conformers were calculated for each compound in the dataset using OMEGA 2.5.1.4[51] and, for the three most active compounds, the lowest energy conformation was determined as further for the following steps.

Construction and Validation of 3D Chemical Similarity Models

Three-dimensional chemical similarity models (CSMs) were generated using the structure of the three most active compounds as a template. Then, the similarity values between CSMs and all dataset compounds were calculated by Tanimoto Coefficient (Tc) related to the molecular shape and chemical features (H-bond donors and acceptors [HBD and HBA], rings [R], hydrophobic groups [HF], anions [A−] and cations [C+]). The initial models included all features present in the template molecules. Aiming to validate the constructed models, we also generate decoys (putative inactive molecules with high physicochemical similarity and structural dissimilarity) in a 1:36 (active: decoys) proportion using the DecoyFinder 2.0 software.[52] The 3D structures of the decoys were prepared by adjusting the ionization state and charges using QUACPAC 1.7.0.2[53] and AM1BCC method,[54] with the *fixpka* option and the *AM1bccspt* force-field, followed by conformer generation using OMEGA 2.5.1.4,[55] where 30 conformers were generated and grouped with real inactive compounds in a file named decoys. A validation run with each generated model (see below) to select and score a set of active molecules and a set of decoy molecules, suggesting confidence levels for future ROCS runs against compounds with unknown activity. The calculated values of Tc for all dataset (active and decoys) were then employed to generate a ROC curve and, consecutively, to calculate the area under the curve (AUC) and enrichment factors at 0.5, 1 and 2% of the screened dataset as validation metrics.

Then, we exhaustively constructed CSMs by excluding each feature and its combinations. All generated models in this step were validated according to the AUC values and enrichment factors and these values were employed in statistical and hypothesis analysis aiming to evaluate the importance of chemical features in active/inactive classification ability of the models. Afterwards, the groups with higher impact in the analysed metrics were used to generate other series of CSMs by varying systematically its weight on Tc calculations. The CSMs generated at this step were also submitted to statistical and hypothesis analyses. At the final step of model generation and validations, a confusion matrix was built to the models with the highest AUC values aiming to calculate the rates of true positives and negatives, accuracy, F1-score and Matthews correlation coefficient (MCC).[56,57] All CSM generations, as well as its validations, were performed with ROCS 3.2.1.4 software and its graphical user interface vROCS.[36] The statistical analysis consisted of a normality test, analysis of groups by boxplot and non-parametric Mean-Whitney hypothesis test[58] performed with Graph-Pad software (v8.1, La Jolla, California, USA). All statistical data

referring to the chemical similarity models are available online in the Zenodo repository (under the code 10.5281/zenodo.3257327).

Molecular Docking

*Ec*FabI proteins co-crystallized with TCL (wild-type PDB ID: 1QG6[12]) was selected for docking simulations of relevant compounds, based on the completeness (no missing residues in the final structure), resolution (1.9 Å) and similarity of the co-crystallized ligand with our compound series. This PDB structure was prepared by adjusting ionization states of amino-acid residues and fixing missing side-chain atoms (PrepWiz, Maestro v2017.4). The grid employed in the docking was generated by Maestro using default settings for van der Waals radius (1.0) and partial charge cut-off (0.25). Molecular docking was performed in a grid encompassing residues around 20 Å from the centroid of the co-crystallized ligand, using the default settings of the Glide program (Glide v7.7, Maestro v2017.4) in extra-precision mode, with at least five poses selected for further visual inspection.[59] Amino-acid residues were considered rigid and both structural water molecules and cofactor NAD^+ were maintained in the active site during the calculation. Structural waters were kept when they made at least two hydrogen bonds, namely water 666 and 667 (numbering based on the PDB 1QG6), interacting among themselves and with Gly93 and with the oxygen from the NAD's phosphate, respectively. The employed docking protocol was evaluated with TCL redocking, which showed a difference in the RMSD (root mean square deviation) values smaller than 1 Å when compared to the experimental binding mode and were able to reproduce interactions with Tyr146 and Tyr156 (Figure 3B). Relevant compounds for docking were selected based on the enzymatic activity: TCL as a positive control, 37 and 41 as active compounds, and 24 inactive one. All ligands were drawn using Maestro and prepared by adding hydrogen atoms according to physiological pH (7.4), followed by the calculation of the atomic charges with the force-field OPLS3e (LigPrep, default settings). Hydroxyl group of different compounds underwent density functional theory-based pKa prediction using Jaguar pKa (Maestro v2018.1) using five initial conformations.[60] Jaguar uses a combination of correlated *ab initio* quantum chemistry to calculate microscopic pKa (*i.e.* on the atomic level), a self-consistent reaction field (SCRF) continuum treatment of solvation and empirical corrections, the latter is employed to repair deficiencies in both solvation models. Calculations were run with the QM method DFT B3LYP/6-31G** level of theory. From calculated pKa values, we calculated the percentage of ionized and neutral species of compounds in pH of simulation using the Henderson-Hasselbach equation.

Molecular Dynamics Simulation

Chosen docking poses for each compound underwent molecular dynamics simulation to evaluate ligand stability within the active site and analyse its interactions. For TCL and compound 41, simulations with their anionic states were also performed according to pKa prediction results. MD simulation was carried out using Desmond[61] with the OPLS3e force-field. This force-field has a better performance representing ligand properties and therefore is suitable to deal with the chemical diversity derived from the virtual screenings.[62] Also, along this force-field represent the halogen bonds by an off-atom charge site, which is suitable for the ligands of this series. The simulated system encompassed the protein-ligand complex, a predefined water model (TIP3P[63]) as explicit solvent and counter-ions (Na^+ or Cl^- adjusted to neutralize the overall system charge, around 4–5 Na^+ atoms). The system was treated in a cubic box with periodic boundary conditions specifying the shape and the size of the box as 13 Å distance from the box edges to any atom of the protein (totalizing around 45,000 atoms between protein, ligand, solvent and

ions). We used a time step of 1 fs, the short-range coulombic interactions were treated using a cut-off value of 9.0 Å using the short-range method, while the smooth Particle Mesh Ewald method (PME) handled long-range coulombic interactions.[64]

Initially, the relaxation of the system was performed using Steepest Descent and the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithms in a hybrid manner. The simulation was performed under the NPT ensemble for 5 ns implementing the Berendsen thermostat and barostat methods. A constant temperature of 310 K was maintained throughout the simulation using the Nose-Hoover thermostat algorithm and Martyna-Tobias-Klein Barostat algorithm to maintain 1 atm of pressure, respectively. After minimization and relaxation of the system, we proceeded with the production step of 200 ns. All MD simulations were performed at least in five independent runs with randomly generated seeds. The representative structure was selected by clustering the structures from the RSMD values, using 1 Å as a cut-off (Supporting Information, Figure 4 represents the variation of the RMSD values along with the simulation). All trajectories from MD simulations are available online in the Zenodo repository (10.5281/zenodo.3257327).

Interactions and distances were determined using the Simulation Event Analysis pipeline implemented in Maestro (Maestro 2018v1). The current geometric criteria for protein-ligand H-bond is distance of 2.5 Å between the donor and acceptor atoms (D—H···A); a donor angle of $\geq 120°$ between the donor-hydrogen-acceptor atoms (D—H···A); and an acceptor angle of $\geq 90°$ between the hydrogen-acceptor-bonded atom atoms (H···A—X). Similarly, protein-water or water-ligand H-bond are: a distance of 2.8 Å between the donor and acceptor atoms (D—H···A); a donor angle of $\geq 110°$ between the donor-hydrogen-acceptor atoms (D—H···A); and an acceptor angle of $\geq 90°$ between the hydrogen-acceptor-bonded atom atoms (H···A—X). Non-specific hydrophobic interactions are defined by hydrophobic side-chain within 3.6 Å of a ligand's aromatic or aliphatic carbons and π-π interactions required two aromatic groups stacked face-to-face or face-to-edge, within 4.5 Å of distance.

WaterMap Calculations

WaterMap calculations were carried out to understand the solvation impact within the inhibitor site of FabI. All WaterMap calculations were run in with default settings. Briefly, a 2 ns molecular dynamics simulation of the *Ec*FabI active site without the compound, but with NAD^+, was performed using the Desmond molecular dynamic engine with the OPLS3e force field. The binding site was defined to include all protein residues within 5 Å distance of any TCL atom, those amino-acids were restrained throughout the simulation. Water molecules from the simulation were then clustered into distinct hydration sites. Enthalpy values for each hydration site can then be obtained by averaging over the non-bonded interaction for each water molecule in the cluster. Entropy values were calculated using numerical integration of local expansion of the entropy in terms of spatial and orientational correlation functions.

[1] S. Santajit, N. Indrawattana, *BioMed Res. Int.* **2016**, *2016*, DOI 10.1155/2016/2475067.

[2] D. J. Payne, P. V. Warren, D. J. Holmes, Y. Ji, J. T. Lonsdale, *Drug Discovery Today* **2001**, *6*, 537–544.

[3] S. Brinster, G. Lamberet, B. Staels, P. Trieu-Cuot, A. Gruss, C. Poyart, *Nature* **2009**, *458*, 83–86.

[4] P. C. Delekta, J. C. Shook, T. A. Lydic, M. H. Mulks, N. D. Hammer, *J. Bacteriol.* **2018**, DOI 10.1128/JB.00728-17.

[5] J. B. Parsons, C. O. Rock, *Curr. Opin. Microbiol.* **2011**, *14*, 544–549.

[6] W. Balemans, N. Lounis, R. Gilissen, J. Guillemont, K. Simmen, K. Andries, A. Koul, *Nature* **2010**, *463*, E3, discussion E4.

[7] S. W. White, J. Zheng, Y.-M. Zhang, C. O. Rock, *Annu. Rev. Biochem.* **2005**, *74*, 791–831.

[8] R. J. Heath, J. Li, G. E. Roland, C. O. Rock, *J. Biol. Chem.* **2000**, *275*, 4654–4659.

[9] R. J. Heath, C. O. Rock, *J. Biol. Chem.* **1995**, *270*, 26538–26542.

[10] A. Banerjee, E. Dubnau, A. Quemard, V. Balasubramanian, K. S. Um, T. Wilson, D. Collins, G. de Lisle, W. R. Jacobs, *Science* **1994**, *263*, 227–230.

[11] L. M. McMurry, M. Oethinger, S. B. Levy, *Nature* **1998**, *394*, 531–532.

[12] W. H. Ward, G. A. Holdgate, S. Rowsell, E. G. McLean, R. A. Pauptit, E. Clayton, W. W. Nichols, J. G. Colls, C. A. Minshull, D. A. Jude, *Biochemistry* **1999**, *38*, 12514–12525.

[13] F. L. Lyman, T. Furia, *IMS Ind Med Surg* **1969**, *38*, 64–71.

[14] H. Lu, K. England, C. am Ende, J. J. Truglio, S. Luckner, B. G. Reddy, N. L. Marlenee, S. E. Knudson, D. L. Knudson, R. A. Bowen, *ACS Chem. Biol.* **2009**, *4*, 221–231.

[15] G. S. Freund, T. E. O'Brien, L. Vinson, D. A. Carlin, A. Yao, W. S. Mak, I. Tagkopoulos, M. T. Facciotti, D. J. Tantillo, J. B. Siegel, *ACS Chem. Biol.* **2017**, *12*, 2465–2473.

[16] K. E. Hevener, S. Mehboob, P.-C. Su, K. Truong, T. Boci, J. Deng, M. Ghassemi, J. L. Cook, M. E. Johnson, *J. Med. Chem.* **2012**, *55*, 268–279.

[17] J.-L. Stigliani, V. Bernardes-Génisson, J. Bernadou, G. Pratviel, *Org. Biomol. Chem.* **2012**, *10*, 6341–6349.

[18] V. A. Morde, M. S. Shaikh, R. R. S. Pissurlenkar, E. C. Coutinho, *Mol. Diversity* **2009**, *13*, 501–517.

[19] R. V. C. Guido, G. Oliva, A. D. Andricopulo, *Comb. Chem. High Throughput Screening* **2011**, *14*, 830–839.

[20] A. Quémard, J. C. Sacchettini, A. Dessen, C. Vilcheze, R. Bittman, W. R. Jacobs, J. S. Blanchard, *Biochemistry* **1995**, *34*, 8235–8241.

[21] S. Escaich, L. Prouvensier, M. Saccomani, L. Durant, M. Oxoby, V. Gerusz, F. Moreau, V. Vongsouthi, K. Maher, I. Morrissey, *Antimicrob. Agents Chemother.* **2011**, *55*, 4692–4697.

[22] B. Hafkin, N. Kaplan, B. Murphy, *Antimicrob. Agents Chemother.* **2016**, *60*, 1695–1701.

[23] C. Mouzé-Soulama, V. Gerusz, A. Denis, *Novel Drug Combination*, **2015**, WO2015028104 A1.

[24] W. R. Miller, A. S. Bayer, C. A. Arias, *Cold Spring Harbor Perspect. Med.* **2016**, *6*, a026997.

[25] C. J. E. Pee, V. Pader, E. V. K. Ledger, A. M. Edwards, *Antimicrob. Agents Chemother.* **2019**, *63*, DOI 10.1128/AAC.02105-18.

[26] A. N. Jain, *J. Med. Chem.* **2003**, *46*, 499–511.

[27] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748.

[28] V. Gerusz, A. Denis, F. Faivre, Y. Bonvin, M. Oxoby, S. Briet, G. LeFralliec, C. Oliveira, N. Desroy, C. Raymond, *J. Med. Chem.* **2012**, *55*, 9914–9928.

[29] A. Chang, J. Schiebel, W. Yu, G. R. Bommineni, P. Pan, M. V. Baxter, A. Khanna, C. A. Sotriffer, C. Kisker, P. J. Tonge, *Biochemistry* **2013**, *52*, 4217–4228.

[30] N. Jiten Singh, D. Shin, H. M. Lee, H. T. Kim, H.-J. Chang, J. M. Cho, K. S. Kim, S. Ro, *J. Struct. Biol.* **2011**, *174*, 173–179.

[31] C. Neckles, S. Eltschkner, J. E. Cummings, M. Hirschbeck, F. Daryaee, G. R. Bommineni, Z. Zhang, L. Spagnuolo, W. Yu, S. Davoodi, *Biochemistry* **2017**, *56*, 1865–1878.

[32] J. Schiebel, A. Chang, B. Merget, G. R. Bommineni, W. Yu, L. A. Spagnuolo, M. V. Baxter, M. Tareilus, P. J. Tonge, C. Kisker, *Biochemistry* **2015**, *54*, 1943–1955.

[33] J. Schiebel, A. Chang, S. Shah, Y. Lu, L. Liu, P. Pan, M. W. Hirschbeck, M. Tareilus, S. Eltschkner, W. Yu, *J. Biol. Chem.* **2014**, *289*, 15987–16005.

[34] S. Vanni, L. Riccardi, G. Palermo, M. De Vivo, *Acc. Chem. Res.* **2019**, DOI 10.1021/acs.accounts.9b00134.

[35] G. Palermo, I. Bauer, P. Campomanes, A. Cavalli, A. Armirotti, S. Girotto, U. Rothlisberger, M. D. Vivo, *PLoS Comput. Biol.* **2015**, *11*, e1004231.

[36] P. C. D. Hawkins, A. G. Skillman, A. Nicholls, *J. Med. Chem.* **2007**, *50*, 74–82.

[37] C. Bissantz, P. Bernard, M. Hibert, D. Rognan, *Proteins* **2003**, *50*, 5–25.

[38] M. L. Verdonk, V. Berdini, M. J. Hartshorn, W. T. M. Mooij, C. W. Murray, R. D. Taylor, P. Watson, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.

[39] A. C. Good, T. I. Oprea, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169–178.

[40] M. Réau, F. Langenfeld, J.-F. Zagury, N. Lagarde, M. Montes, *Front. Pharmacol.* **2018**, *9*, DOI 10.3389/fphar.2018.00011.

[41] L. Chen, A. Cruz, S. Ramsey, C. J. Dickson, J. S. Duca, V. Hornak, D. R. Koes, T. Kurtzman, **2019**, DOI 10.26434/chemrxiv.7886165.v1.

[42] N. Huang, B. K. Shoichet, J. J. Irwin, *J. Med. Chem.* **2006**, *49*, 6789–6801.

[43] M. J. Vainio, J. S. Puranen, M. S. Johnson, *J. Chem. Inf. Model.* **2009**, *49*, 492–502.

[44] L. A. C. Vaz de Lima, A. S. Nascimento, *Eur. J. Med. Chem.* **2013**, *59*, 296–303.

[45] S. Sivaraman, J. Zwahlen, A. F. Bell, L. Hedstrom, P. J. Tonge, *Biochemistry* **2003**, *42*, 4406–4413.

[46] S. Rafi, P. Novichenok, S. Kolappan, C. F. Stratton, R. Rawat, C. Kisker, C. Simmerling, P. J. Tonge, *J. Biol. Chem.* **2006**, *281*, 39285–39293.

[47] X. Yang, J. Lu, M. Ying, J. Mu, P. Li, Y. Liu, *J. Mol. Model.* **2017**, *23*, 25.

[48] J. Blaney, A. M. Davis, in *The Handbook of Medicinal Chemistry*, **2014**, pp. 96–121.

[49] *Discovery Studio Visualizer*, Dassault Systèmes BIOVIA, San Diego, **n.d.**

[50] *QUACPAC*, OpenEye Scientific Software, Santa Fé, NM, **2016**.

[51] P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, M. T. Stahl, *J. Chem. Inf. Model.* **2010**, *50*, 572–584.

[52] A. Cereto-Massagué, L. Guasch, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallvé, *Bioinformatics* **2012**, *28*, 1661–1662.

[53] OpenEye Scientific Software, **2016**.

[54] A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623–1641.

[55] P. C. D. Hawkins, A. G. Skillman, G. L. Warren, B. A. Ellingson, M. T. Stahl, **2010**.

[56] B. W. Matthews, *Biochim. Biophys. Acta* **1975**, *405*, 442–451.

[57] D. Powers, *Journal of Machine Learning Technologies* **2011**, *2*, 37–63.

[58] H. B. Mann, D. R. Whitney, *Ann. Math. Stat.* **1947**, *18*, 50–60.

[59] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz, *J. Med. Chem.* **2006**, *49*, 6177–6196.

[60] J. J. Klicić, R. A. Friesner, S. Y. Liu, W. G. Guida, *J. Phys. Chem. A* **2002**, *106*, 7, 1327–1335.

[61] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, ACM, New York, NY, USA, **2006**.

[62] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, *J. Chem. Theory Comput.* **2016**, *12*, 281–296.

[63] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.

[64] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.