Durham Research Online

Deposited in DRO:
02 May 2019

Version of attached file:
Published Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:

Further information on publisher’s website:
https://doi.org/10.1021/acs.jpcllett.9b00741

Publisher’s copyright statement:
This is an open access article published under a Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:
- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.
Relative Binding Energies Predict Crystallographic Binding Modes of Ethionamide Booster Lead Compounds

Natalie J. Tatum,‡ Fernanda Duarte,§ Shina C. L. Kamerlin,¶ and Ehmke Pohl*†∥

†Department of Chemistry, Durham University, South Road, Durham DH1 3LE, U.K.
‡Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, U.K.
§Department of Chemistry - BMC, Uppsala University, BMC Box 576, S-751 23 Uppsala, Sweden
¶Department of Biosciences, Durham University, Durham DH1 3LE, U.K.
∥Biophysical Sciences Institute, Durham University, Durham DH1 3LE, U.K.

Supporting Information

ABSTRACT: Transcriptional repressor EthR from Mycobacterium tuberculosis is a valuable target for antibiotic booster drugs. We previously reported a virtual screening campaign to identify EthR inhibitors for development. Two ligand binding orientations were often proposed, though only the top scoring pose was utilized for filtering of the large data set. We obtained biophysically validated hits, some of which yielded complex crystal structures. In some cases, the crystallized binding mode and top scoring mode agree, while for others an alternate ligand binding orientation was found. In this contribution, we combine rigid docking, molecular dynamics simulations, and the linear interaction energy method to calculate binding free energies and derive relative binding energies for a number of EthR inhibitors in both modes. This strategy allowed us to correctly predict the most favorable orientation. Therefore, this widely applicable approach will be suitable to triage multiple binding modes within EthR and other potential drug targets with similar characteristics.

EthR (Figure 1a) is a well-studied transcriptional repressor from Mycobacterium tuberculosis that represses the expression of the Bayer–Villager mono-oxygenase protein EthA, the in vivo activator of antibiotic pro-drug ethionamide. The homodimeric protein belongs to the family of TetR repressors comprising an N-terminal DNA-binding domain and a C-terminal ligand binding domain. Importantly, EthR ligands inhibiting DNA-binding have been shown to increase ethA expression and significantly boost ethionamide efficacy in vivo models.1–4,17 Consequently, EthR inhibitors are under intense investigation as booster codrugs.5–8,12–15 Several approaches have been used to explore novel chemotypes for binding the lipophilic EthR ligand site (Figure 1b), including traditional high-throughput library screening,6 fragment-based drug discovery,13,16 and our previously reported virtual screening (VS) program based on the GOLD software.18 We reported our screen of over 400 000 ligands against EthR, a cohort filtered from an initial starting library in excess of six million compounds derived from the Drugs Now subset of the ZINC database.18 Our predocking filtering strategy included restricting compound volume to 200–700 Å³ and the requirement of at least one H-bond acceptor, compatible with the size and nature of the EthR binding channel, as well as tailoring physiochemical properties to largely lipophilic binding site.18 In most cases, rigid receptor docking (RRD) yielded two distinct orientations of binding to EthR; however, only the top scoring pose was used for further filtering. Subsequent in vitro assays yielded several biophysically and structurally confirmed hits, though the in crystallo binding mode in some cases was different in orientation to the top scoring virtual screening pose. For one compound we observed both orientations within the crystallized complex, each modeled with 50% occupancy. Importantly, because of the design of the VS, each compound represented a different scaffold, as confirmed in the structural analysis of EthR complexes performed by Tanina and colleagues in 2018.19

In recent years, both RRD and induced-fit docking (IFD) protocols have become reasonably proficient at reproducing crystallographic binding modes, as evaluated by the D3R Grand Challenges (Drug Design Data Resource).21–23 RRD and IFD have been applied in virtual screening and structure-based drug design to discover and develop novel ligands for a wide variety of medically relevant targets. EthR poses a challenge to RRD due to the composition of the binding site: it is highly lipophilic with two asparagine residues (Asn 176 and Asn 179, respectively) located in the center of the pocket. The side chains of these two residues are typically involved in hydrogen bonding and “anchoring” the compound, allowing for the proposed flipping of binding modes.

Received: March 14, 2019
Accepted: April 9, 2019
Published: April 9, 2019
While automated ligand docking approaches can be combined with molecular dynamics (MD) simulations, explicit MD simulations themselves are not yet feasible for screening libraries containing millions of compounds. The combined RDD−MD approach presented here refines the binding poses in explicit solvent (often absent in RRD), and following RDD−MD with the calculation of relative ligand binding energies quantifies the binding event in solution to prioritize favorably binding ligands.

In this contribution we combine RRD, MD simulations, and the linear interaction energy (LIE) method to calculate the relative free energies of binding for four inhibitors in their two proposed orientations.

From our previous virtual screening, we identified compounds 3, 10, and 85 as promising EthR ligands, and successfully cocry stallized EthR with 10. The structures of these compounds are given in the supplementary material in Chart S1. Here we report in addition the cocry stallized structures of EthR with 3 and 85, both at a resolution of 1.8 Å.

Compound 3 has two similar poses where the methyl-oxopyrazine moiety is exposed to solvent (Figure 2a). Key interactions between compound 3 and the EthR pocket include a hydrogen bond between the R159 side chain and the oxopyrazine carbonyl oxygen (2.6 Å, only present in one pose), an OH···N hydrogen bond from the hydroxyl oxygen of Y148 to a pyrazine nitrogen (2.9−3.3 Å), and a CH···O interaction to L87 (2.8 Å). A close-up of these interaction patterns can be seen in Figure S1. The B-factors of the ligand itself are approximately twice the magnitude of the average B-factor for the surrounding protein, an indication of increased ligand flexibility.

No hydrogen bond interactions are observed in the crystal structure of EthR cocry stallized with compound 85 (Figure 2b). In this case binding appears to be driven by van der Waals interactions (Figure S2). A close contact is formed between the central carbonyl on the pyrazolopyrimidine core to the sulfur atom of M142 (2.5 Å). Further details on data collection and refinement statistics are presented in Table S1.

These three crystal structures highlight the differences between the virtual screening pose used for ligand selection and the experimentally observed binding mode. Specifically, the VS provided two orientations for compound 3, the top scoring pose of which matches the crystal structure. Conversely, for compound 85, the VS protocol yielded only one orientation, which does not match the crystal structure. Finally, two orientations were suggested for compound 10 (Figure 2c) and both are observed in crystallo. We used these cases and the crystal structure of the EthR-BDM31343 complex (PDB 3TP0, 1.9 Å resolution, Figure 2d) to study alternate binding conformations. BDM31343 presents an

Figure 2. Ligand binding pocket of (a) EthR-compound 3, ligand in plum; (b) EthR-compound 85, ligand in green; (c) EthR-compound 10, ligand A in yellow and ligand B in blue; and (d) EthR-BDM31343, ligand in dark red. Hydrogen bonds are highlighted as dotted magenta lines, and close contacts are in blue. Only key residues are shown for clarity.
unambiguous single binding mode, and docking studies undertaken in 2013\textsuperscript{11} demonstrated that docking of this compound (under the VS protocol by which we identified compounds 10, 3, and 85\textsuperscript{11,12}) produced five poses of one orientation, in agreement with the crystal structure. We hypothesized that had we utilized relative binding energies in selecting which orientation from virtual screening was energetically favorable, we would have correctly predicted the experimental binding mode. This suggests that a more quantitative approach is required to evaluate the likelihood of observing different poses. Here we used the calculated binding energies in explicit solvent as the criteria to determine the most favorable pose and compared these results to our crystallographic data. It was found that the MD–LIE approach can confidently identify and rank correct binding modes.

For each EthR–ligand complex, MD–LIE calculations in explicit solvent (24 Å sphere radius) were carried out employing the highest scoring VS pose ("pose 1") and the highest-scoring alternate pose for the opposite binding orientation ("pose 2"). For compounds 85 and BDM31343 where no second orientation was proposed by docking, the alternate binding mode within the linear tube-like EthR binding channel was created manually through an approximately 180° rotation, maintaining the hydrogen bonding pattern and the hydrophobic interactions on either side of the amide bond. Each simulation was performed in quadruplicate employing different starting velocities (see the Supporting Information for details).\textsuperscript{20} Following equilibration, the complexes were unrestrained for an initial 10 ns production MD run, extended to 20 ns when required, leading to a total of 40–80 ns of unrestrained MD per system. Each compound was also independently simulated in explicit solvent. Further details are provided in the Supporting Information. All four replicates were used to calculate a relative binding energy, given in Table 1. Ligand positions per pose can be seen in Figure S4.

Per-residue contributions to binding energy (Figure 3 and Table S2) were calculated for each complex. These calculations allowed us to confirm the relevance of the interactions observed in the crystal structure and identify any exploitable interactions for ligand development that were not captured by the crystal structure. This was particularly relevant for compounds 3 and 85 which were indicated to be more flexible and of lower occupancy, respectively, in the EthR binding site in crystallo. While BDM31343 has been reported to have an EC50 of 3.3 \times 10^{-6} mol, compounds 3, 10, and 85 displayed no antitubercular booster activity.\textsuperscript{18} Although the compounds appear quite similar in terms of molecular weight and size, our lead compounds display significantly more polar surfaces compared to BDM31343, which may be responsible for a reduced uptake through the very hydrophobic \textit{M. tuberculosis} cell wall. The detailed energetic characterization of protein–ligand interactions reported here will be used to further optimize our lead compounds.

Figure 3 shows the electrostatic and van der Waals contribution of the relevant binding site residues for each ligand (contributions for compound 10 are deconvoluted as 10-A and 10-B). Negative binding energy contributions denote a favorable interaction. As anticipated, given the highly lipophilic nature of the binding channel, binding is mostly driven by van der Waals interactions (shown in blue). The common largest contributions arise from residues F110, W103, W145, and W207. Electrostatic interactions are mostly due to residues N176 and N179, which are the two key hydrogen bonding residues of the EthR binding channel. Compound 3 adopts a different binding position than the other presented ligands; it interacts more strongly with L90, M102, W103, and V152 and weakly with F110, W145, N176, and N179. Binding by compound 85 is dominated by van der Waals interactions; compound 85 makes electrostatic interactions with only N179 and E180. Clear differences can be seen between poses A and B of compound 10; the alternate positions of the carbonyl oxygen forms an HB interaction with N176 in pose A, while in pose B this interaction occurs preferentially with N179. Similarly, only pose A of compound 10 interacts with Y148 and T149 and the binding energy includes a contribution from M142; pose B interacts favorably with E180. Compound 3 also benefits from the strong electrostatic contributions by G106 due to Ca-H···O and Ca-H···N hydrogen bonds. A Ca-H···S weak hydrogen bond from the sulfur of M142 to compound 10-A is responsible for a M142 contribution to binding energy, which is more pronounced for compound 85 where a short Ca-H···S contact by the M142 sulfur can be observed in the crystal structure and the van der Waals energy contribution ranges from ~1.60 to ~2.49 kcal mol\textsuperscript{-1}. M142 is also involved in BDM31343 binding: the M142 sulfur interacts with the cyano-group of the ligand. BDM31343 makes more favorable electrostatic interactions than the other ligands, and van der Waals interactions are weaker in energy contribution. Overall, analysis of the energy contributions to ligand binding

<table>
<thead>
<tr>
<th>compound</th>
<th>virtual screening pose</th>
<th>binding mode in crystallo</th>
<th>ligand in water (kcal mol\textsuperscript{-1})</th>
<th>ligand in protein (kcal mol\textsuperscript{-1})</th>
<th>β</th>
<th>averageΔG\textsubscript{bind} (kcal mol\textsuperscript{-1} ± SE)</th>
</tr>
</thead>
</table>
corroborate the interactions inferred from the crystal structures, emphasizing the lipophilicity-driven nature of the binding to EthR. This data can be used to quantitatively evaluate the binding event and help inform design decisions—at which positions to try to strengthen hydrogen bond acceptors, for example, or to exploit “weak” hydrogen bonding opportunities such as those seen via G106 with compound 3 or M142 with BDM31343 and compound 85. Given the evidence of weak binding by 85, both compounds 3 and 10-A provide the best starting points for further development.

Figure 4a shows the overlay of the crystal structure of compound 3, with each of the two simulated poses (taken as an average structure at the 10 ns time point). Pose 1 (the highest-scoring VS mode) is a match for the crystallographically observed binding mode and is also the most energetically favorable binder as compared to pose 2 (Table 1, \( \Delta\Delta G = 2.3 \pm 0.6 \text{ kcal mol}^{-1} \)). For compound 85, the results of the virtual screening suggested only one pose (pose 1). However, it is pose 2 which is energetically more favorable (\( \Delta\Delta G = 2.0 \pm 1.2 \text{ kcal mol}^{-1} \), Table 1). In the case of compound 10, the dual occupancy found in the crystal structure is in line with our MD–LIE free energy calculations, which demonstrates that the binding free energy for these poses is almost identical (\( \Delta\Delta G = 0.1 \pm 0.6 \text{ kcal mol}^{-1} \)), Figure 4c). Finally, for BDM31343 (Figure 4d) pose 1 is energetically favorable (\( \Delta\Delta G = 1.3 \pm 1.1 \text{ kcal mol}^{-1} \)), and it corresponds to the pose found in the crystal structure. Therefore, we can conclude that both virtual screening and the MD–LIE estimate correctly predict the binding mode of compounds 3 and 10 and BDM31343. However, it is only the MD–LIE approach that correctly identified the experimentally observed binding mode of compound 85. Therefore, we have demonstrated that the MD–LIE approach allows one to accurately determine and rank potential poses from the in silico screening output.

Figure 3. Mean contribution of energy to the overall binding, per-residue of the binding site, for both (a) electrostatic and (b) van der Waals energies for compounds 3, 10, and 85 and BDM31343. Energy contributions over the course of the simulation are given in Figure S3. Only values exceeding \( \pm 1 \text{ kcal mol}^{-1} \) are shown. Graphs prepared in Graphpad Prism 7.
In summary, this contribution has addressed a key issue posed to computer-aided drug discovery on the EthR target. In this system, the cylindrical, hydrophobic binding site may give rise to multiple likely binding modes which may be difficult to infer only by docking score rank. Determination of the most favorable pose(s) requires crystallographic confirmation, which may be difficult to achieve. The inclusion of the MD−LIE approach to calculate the relative binding energy calculations is an established practice for ranking small cohorts of screening outputs. Though still not viable for routine screening at scale,31,32 its application on the EthR systems has the potential to accelerate the identification and evaluation of novel ligands. Further studies would be required to determine if MD−LIE is sensitive enough for smaller fragment-sized molecules with weak binding affinities, or if indeed a one-simulation end-point method such as MM/GBSA or MM/PBSA could be a higher-throughput method with sufficient accuracy. MD−LIE or an equivalent method could be used to prioritize outputs from crystallographic fragment screening,33 where multiple, milli-molar-affinity binding events across a protein surface are frequently observed.34

Recent literature has demonstrated that multiple binding modes within series can be commonplace35 and often are unapparent unless crystal structures are obtained. This is also true of EthR, for example with compound BDM14500.5 In the absence of a crystal structure unapparent unless crystal structures are obtained. This is also candidates are prioritized. hits toward active cell-penetrant lead compounds, and channels will heal them or important to exploit forward, because of the hydrophobicity of the EthR binding structure strategies by characterizing the protein to the binding event will enable medicinal chemistry design is required to determine if our approach is applicable to targets beyond EthR. Furthermore, per-residue energy contributions to the binding event will enable medicinal chemistry design strategies by characterizing the protein−ligand interaction and highlighting where gains could be made (or unfavorable interactions are occurring).36

Ultimately, the pilot experiment presented here suggests a role for relative binding energies and per-residue contributions in not just ranking ligand binders but also confirming ligand binding modes. Furthermore, we have contributed two more crystal structures to the growing data on EthR complexes, which are a valuable resource for further study.19 Moving forward, because of the hydrophobicity of the EthR binding channel it will be all the more important to exploit unconventional binding modes and chemistries to optimize hits towards active cell-penetrant lead compounds, and inclusion of MD−LIE in this pipeline will ensure the best candidates are prioritized.

# ASSOCIATED CONTENT

## Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcllett.9b00741.

Experimental Methods and supplementary chart and tables (PDF)

# AUTHOR INFORMATION

## Corresponding Author

*E-mail: ehmke.pohl@durham.ac.uk.

**ORCID**

Fernanda Duarte: 0000-0002-6062-8209

Shina C. L. Kamerlin: 0000-0002-3190-1173

Ehmke Pohl: 0000-0002-9949-4471

## Notes

The authors declare no competing financial interest.

The crystal structure data for EthR with compound 3 and compound 85 are deposited in the PDB under respective accession codes 6R1P and 6R1S.

# ACKNOWLEDGMENTS

This work was conducted with the resources of the Swedish National Infrastructure for Computing (SNIC). N.J.T. gratefully acknowledges support from the Cambridge Crystallographic Data Centre and the EPSRC (EP/J500215/1) for a predoctoral fellowship and an EMBO Short Term Fellowship (ASTF 123-2015) which funded travel and subsistence for this work to be carried out. This work was also supported by the Swedish Foundation for International Co-operation in Research and Higher Education (STINT). We are grateful to Ian Edwards and Erin Dickinson for their excellent technical support in protein production and crystallization and the Diamond Light Source for beamtime and their excellent assistance on beamlines I03 and I04-1, respectively.

# REFERENCES


