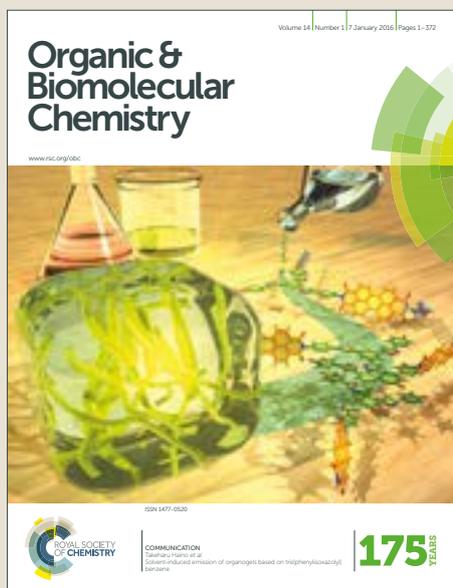


Organic & Biomolecular Chemistry

Accepted Manuscript



This article can be cited before page numbers have been issued, to do this please use: N. J. Tatum, J. Liebeschuetz, J. C. Cole, R. Frita, A. Herledan, A. Baulard, N. Willand and E. Pohl, *Org. Biomol. Chem.*, 2017, DOI: 10.1039/C7OB00910K.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [author guidelines](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the ethical guidelines, outlined in our [author and reviewer resource centre](#), still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.



Organic and Biomolecular Chemistry

ARTICLE

New active leads for Tuberculosis booster drugs by structure-based drug discovery

Natalie J. Tatum^a, John W. Liebeschuetz^b, Jason C. Cole^c, Rosangela Frita^d, Adrien Herledan^e, Alain R. Baulard^d, Nicolas Willand^e and Ehmke Pohl^{a,f,g*}Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The transcriptional repressor EthR from *Mycobacterium tuberculosis*, a member of the TetR family of prokaryotic homodimeric transcription factors, controls the expression of the mycobacterial mono-oxygenase EthA. EthA is responsible for the bio-activation of the second-line tuberculosis pro-drug ethionamide, and consequently EthR inhibitors boost drug efficacy. Here, we present a comprehensive in-silico structure-based screening protocol that led to the identification of a number of novel scaffolds of EthR inhibitors in subsequent biophysical screening by thermal shift assay. Growth inhibition assays demonstrated five of the twenty biophysical hits were capable of boosting ethionamide activity in vitro, with the best novel scaffold displaying an EC50 of 34 μM . In addition, the co-crystal structures of EthR with four new ligands at resolution ranging from 2.1 to 1.4 Å confirm the binding and inactivation mode, and will enable future lead development.

Introduction

Tuberculosis remains one of the premier causes of mortality and illness, infecting about one third of the world's population and with more than 8 million new cases of disease leading to more than 1.5 million deaths per year¹. The treatment of Tuberculosis is hampered by the pathogen's ability to remain in a latent, non-replicating state.² The problem is exacerbated by the rapid increase in multi-drug resistant (MDR) *Mycobacterium tuberculosis* strains that are characterized by showing resistance to the first line drugs isoniazid and rifampicin.³ Furthermore, MDR *M. tuberculosis* strains that are also resistant to fluoroquinolone and to an injectable second-line antibiotic are classified as extensively drug resistant (XDR).⁴ Whereas first line drugs such as isoniazid, rifampicin,

pyrazinamide and ethambutol - all developed and in use since the 1950s and 60s - have been and still are very successful in the DOTS (Directly Observed Treatment Short-course) strategy introduced in the 1990s, infections with MDR and XDR TB require much more extensive treatment with second-line antibiotics including fluoroquinolones, D-cycloserine, ethionamide or other antibiotics for 2-4 years. These drugs, however, are generally less effective and show more often serious side effects.⁵ For all of these reasons there have been renewed efforts in pharmaceutical industry and public research institutes to discover and develop new antibiotics.⁶ While target and structure-based approaches such as the Tuberculosis Structural Genomics Consortium (TBSGC, <http://www.webtb.org>) have contributed great functional knowledge and structural detail about key proteins and their interaction partners,^{7,8} high-throughput assays based on whole-cells have in general been more successful in identifying promising lead compounds.⁹ Combining all efforts has led to more than a dozen compounds, either based on known drugs or in some cases new chemical entities that are currently in clinical trials.^{10,11} However, in spite of these encouraging recent advances there is an urgent need to optimize existing treatments and develop novel therapeutics.^{5,6} One promising strategy developed over the last years focuses on the transcriptional regulator EthR, a member of the TetR family involved in ethionamide resistance in *M. tuberculosis*.^{12,13} Ethionamide is administered as a pro-drug, which is converted into its active NAD adduct inside the mycobacteria by the Bayer-Villiger monooxygenase EthA. The NAD-adducts inhibit the 2-trans-enoyl-acyl carrier protein reductase InhA (Fig. 1), thus interfering with

^a Department of Chemistry, Durham University, South Road, Durham DH1 3LE, United Kingdom

^b Skilos, Cheminformatics, Cambridge, UK.

^c Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, United Kingdom.

^d Université Lille, CNRS, Institut Pasteur de Lille, U1019- UMR8204 = CIL, F-59000 Lille, France.

^e Université Lille, Inserm, Institut Pasteur de Lille, U1177 - Drugs and Molecules for Living Systems, F-59000 Lille

^f Department of Biosciences, Durham University, Durham DH1 3LE, United Kingdom

^g Biophysical Sciences Institute, Durham University, Durham DH1 3LE, United Kingdom

*Current address: Northern Institute for Cancer Research, Paul O'Gorman Building, Medical School, Newcastle University, Framlington Place, Newcastle Upon Tyne, NE2 4HH, United Kingdom.

Electronic Supplementary Information (ESI) available: [Computational and experimental detail, list of compounds, crystallographic data]. See DOI: 10.1039/x0xx00000x

ARTICLE

Journal Name

mycolic acid synthesis essential to maintain the pathogen's cell wall.¹⁴

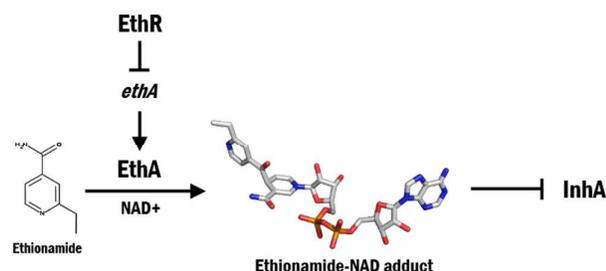


Fig. 1: Schematic drawing showing the action of the transcriptional regulator EthR on the *ethA* gene. When bound to DNA EthR limits the expression of EthA which is responsible for the bioactivation of the prodrug forming the active ethionamide-NAD adducts. These adducts inhibit the 2-trans-enoyl-acyl carrier protein InhA.

Based on the discovery that the transcriptional regulator EthR limits the expression of EthA, and inhibiting the regulator increases the drug action, EthR has been validated as a suitable target for the development of compounds that boost the *in vitro* and *in vivo* efficacy of ethionamide.¹⁵ The crystal structures of EthR solved in 2004, fortuitously both in ligand-bound form, adopt an *inactive* conformation where the DNA-binding domains are orientated such to be incompatible to DNA recognition.^{16,17} Consecutive structure-based design studies taking advantage of the hydrophobic binding pocket led to a series of 1,2,4-oxadiazole inhibitors^{15,18,19} while high-throughput screening resulted in the discovery of N-phenoxy acetamide derivatives.²⁰ In addition, fragment-based approaches were also applied, and through growing, merging and linking strategies led to low nanomolar affinity ligands of EthR and low nanomolar boosters of ethionamide activity *in vitro*.²¹ Moreover, Nikiforov et al. showed that hydrophobic piperidinyl amide compounds, which were able to inhibit the repressor, filled the entire binding pocket.²²

In order to take advantage of the wealth of information hidden in the multiple crystal structures of EthR published so far, a high-throughput *in-silico* screening strategy was employed starting with the ZINC database of available chemical compounds.²⁴ In the past, *ensemble docking* has been suggested where each compound is computationally docked in the different protein conformations, which can either be derived from a number of crystal structures or – albeit less successfully – by molecular dynamics simulation.^{25,26} However, firstly the computational costs of this method scales with the number of structures and therefore quickly becomes unfeasible, and secondly compounds are docked into potentially very similar target structures producing a large amount of redundant results. Therefore, a target-specific decoy set of molecules was developed to select the most suitable EthR crystal structure available in the PDB and to optimize the virtual screening parameters. In addition, a pre-docking protocol tailored to the unique requirements of the protein target was created using the pipeline software KNIME, which reduced the number of compounds in the drug-like subset of

the ZINC database for virtual screening. After high-throughput docking using GOLD,²⁷ further detailed post-screening filtering – tailored to identify new chemical scaffolds with good starting physicochemical properties for development – resulted in a set of 284 compounds, further reduced to 85 potential hits for biophysical analysis by careful manual inspection. Protein-ligand binding was measured by thermal shift assay with 20% of virtual screening hits showing a significant increase of thermal protein stability. The co-crystal structures of four compounds, three of which comprise new chemical entities for EthR inhibition, were determined to resolutions ranging from 1.4 to 2.1 Å. These structures showed that in all cases the ligand causes the protein to adopt its inactive conformation, confirming the inhibition mechanism. Furthermore, detailed analysis of bound compounds led to the identification of certain essential binding motifs and allowed a direct comparison of computational docking poses and experimental binding mode. These results represent the starting point for further lead development of tuberculosis *booster* drugs.

Results

Virtual Screening Optimised for EthR – At the time the project was started, 12 co-crystal structures of EthR were available in PDB,²⁸ all crystallised in the inactive conformation incompatible with DNA binding.^{15-17,29-31} Despite close similarities in protein arrangement in the crystal structures (RMSDs of all C α -atoms in each monomer ranging from 0.2–0.4), distinct differences can be observed in the exact ligand positions as shown in Fig. 2.

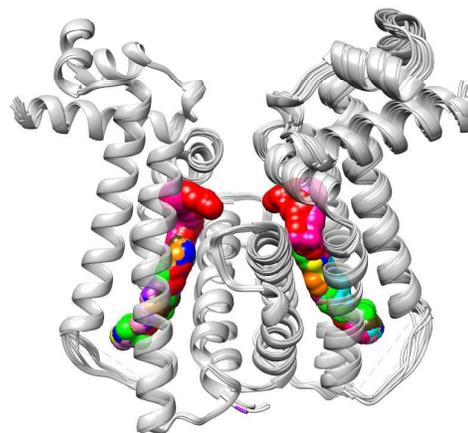


Fig. 2: Superposition of the EthR co-crystal structures used in this study shown in grey and the different ligands shown as CPK model with each color representing one independent structure. Note that the left-hand side monomer was used to calculate the transformation matrix applied to the EthR dimer.

Cross-docking was performed to identify the most suitable receptor structure. All co-crystallised synthetic ligands were docked into all available crystal structures. It is noteworthy that in one case (PDB:

1U9N) the key residue Asn179, which forms a highly conserved hydrogen bond to the ligands, needed to be manually curated into its correct hydrogen bonding pattern. The 25 *genetic algorithm* runs using 1U9N resulted in 150 correct solutions compared to as few as less than ten in other cases. Consequently, coordinates from this crystal structure were chosen for the virtual screening.

Compounds were sourced from the *Drugs Now* subset of the ZINC database, a free resource of commercially available chemical compounds comprising of 6.06 million compounds (as of 2014 when work was undertaken).²⁴ For all compounds in the subset, twenty chemical and physicochemical descriptors were calculated (Fig. 3) of which several were used for pre-docking filtering using the KNIME pipeline software,³² the boundaries of which were based on the detailed knowledge of the binding pocket. The most stringent filter, excluding approximately 50% of all compounds, restricted the compound volume to 200–700 Å³. This decision was taken in spite of the fact that the binding pocket extends up to 1100 Å² in 1U9N to focus on smaller, more varied scaffolds which could be devolved and further modified in the downstream medicinal chemistry efforts. Overall pre-docking filtering reduced the ligand set down to approximately 1.3 million compounds (Fig. 3).

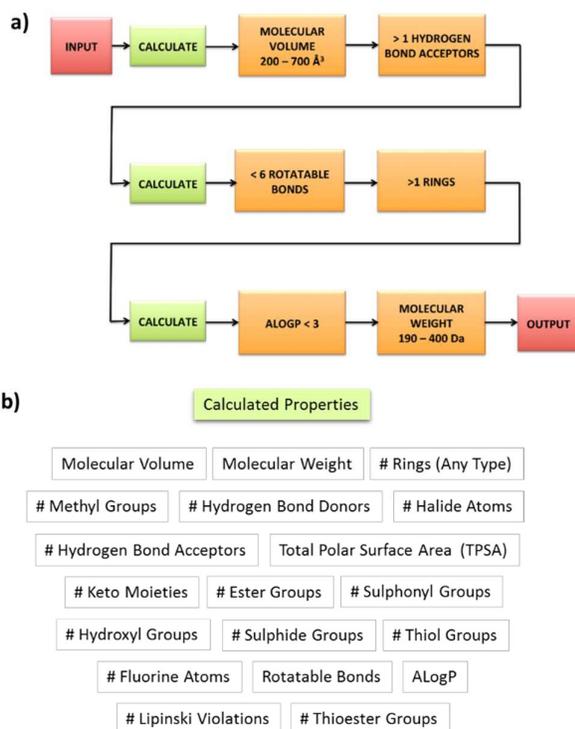


Fig. 3: Panel a) shows a schematic of the KNIME pipeline used to pre-filter compounds for docking. Red boxes indicated input/output nodes; green boxes represent calculation nodes; orange boxes denote filter nodes with the parameters shown. Panel b) shows the chemical descriptors calculated upon the *Drugs Now* subset, some of which were used in pre- and/or post-docking filtering as described in the materials and methods section.

Rhodanines, a class of compounds previously identified as *pan-assay interference (PAINs)* compounds that are well known for producing false-positives in drug discovery attempts,³³ were removed and remaining structures were sorted into 23 groups according to chemical ring-system types (see SI: Figure S1 and Table S1). Since the resulting cluster sizes range from 8,805 (thiadiazoles) to 240,174 (pyrroles), a varying percentage from each cluster was taken forward to maintain chemical diversity. The final compound cohort for virtual screening comprised of 409,201 structures.

Parameters for virtual docking were optimized by constructing a decoy set of compounds structurally similar to one of the most efficacious inhibitors (BDM41906), chosen because crystal structures of this compound alone and in complex with EthR have been previously determined.^{18,19} A KNIME pipeline was used to select 350 compounds from the ZINC database and create a decoy set, applying a set of physicochemical parameters including molecular weight and logP,³⁴ with the set used to systematically test the docking parameters for screening. The best results as measured by highest enrichment of the known ligands within the decoy set of 350 compounds were obtained using default genetic algorithm parameters implemented in GOLD for 10 runs, maximum ligand flexibility and employing the empirical ChemPLP scoring function.³⁵ Using the decoy data set, the search efficiency in GOLD could be reduced to 20% with a modest search space of radius 10 Å, while generating a diverse population of five resultant poses per ligand. The decoy set of 350 molecules seeded with 9 known actives under these conditions gave enrichment factors of 4.0 - 4.4 at a 10% cut-off.

All optimisation experiments were performed on a four processor Intel Xeon Desktop, however for expediency, the virtual screening itself was conducted on the Darwin Supercomputer at the University of Cambridge High Performance Computing facility. With 409,201 potential ligands in the library and only six compounds that failed to generate any docking results, 2,045,975 poses were obtained.

Selection of Virtual Screening Hits – The highest scoring pose for each chemical entity was used, and further only the top scoring 10% of solutions were carried forward. All descriptors calculated by KNIME in pre-filtering were carried forward through screening into post-docking filtering. In addition, novel descriptors were specifically defined for this binding pocket, such as a requirement for a chemically reasonable hydrogen bond geometry with a hydrogen bond angle of > 120° (based on the co-crystal structure between the ligand and Asn179 as seen for example in PDB entry 3G1M). Additionally, parameters of the scoring function such as clash and torsion terms were used to filter ligands. The range for each of the ten parameters used to filter all binding poses of the 40,919 compounds are summarized in Table 1. This process resulted in 284 potential ligands, further evaluated using interactive graphics and MOGUL, which uses information from the Cambridge Structural Database (CSD) to assess ligand geometries.³⁶ Binding poses with highly unusual or unlikely torsion angle and ring geometries were manually removed from the hit set intended for

ARTICLE

Journal Name

biophysical screening. In addition, chemically similar hits were removed in order to reduce redundancies. These steps resulted in 85 commercially available compounds. The example shown in Fig. 4 exhibits chemically reasonable ligand geometry and exemplifies key interactions of the ligands with the hydrophobic pocket lined by Phe110, Phe114, Trp138 and Phe184. In addition, the binding poses also show the possible hydrogen bond with the side chain of Asn179. The 85 hits were carried forward to biophysical characterization.

Table 1: Details of the parameters used in post-docking filtering to derive 284 poses for manual inspection.

Parameter	Range	No. of ligands
CHEMPLP score	10%	40919
Number of hydrogen bonds with angle $\geq 120^\circ$	1 - 6	21949
Total polar surface area	75 \AA^2 – 110 \AA^2	15033
AlogP	-3 - 1	5390
Number of hydrogen bonds formed protein:ligand	1 - 7	25941
Occluded donor atoms	0 - 1	33603
Occluded polar atoms	0 - 1	33221
Ligand clash count	0 - 14	32399
CHEMPLP ligand torsion score	0.027 – 3.0	28248
CHEMPLP ligand clash score	0 – 2.0	31950
Internal energy correction	0 - 3.0	20529

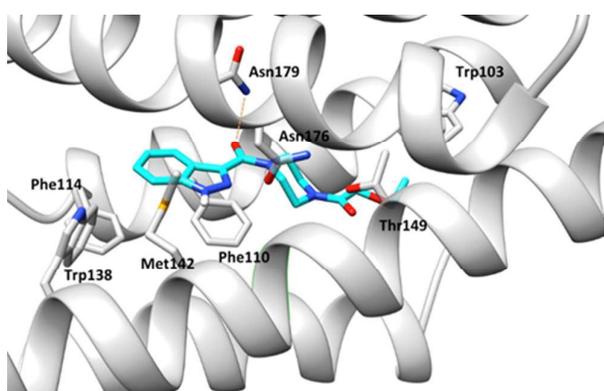


Fig. 4: Close-up of the ligand binding pocket with the protein shown in a grey ribbon representation and compound 64 shown in stick representation. Key residues are labelled.

Identification of Biophysical Hits by Thermal Shift Assay – Compounds were sourced from companies ChemBridge, Enamine

and Ambinter. All compounds were quality-checked by LC-MS and or NMR by the company at >95% purity. The complete list is given in the ESI. Analytical data (LC-MS/NMR) for the twenty top compounds summarized in Table 2 is also presented in the ESI.

Thermal shift assays, also known as Differential Scanning Fluorimetry (DSF) or ThermoFluor[®] assay have been widely used as a fast and easy initial biophysical test to verify protein-ligand binding.³⁷⁻³⁹ In almost all cases, ligand binding to a protein target causes an increase in thermal stability, although in some cases a decrease has also been observed.³⁹ Importantly, TSAs have been employed successfully to identify EthR inhibitors from high-throughput screens²⁰ and to verify EthR-ligand binding of designed compounds.^{22,40} TSAs were performed on all 85 available compounds using an optimized protocol in a 96-well-plate format (for a complete list see the SI table S2). A typical example of an experimental melting curve is shown in Fig. 5a where the addition of 80 μM of compound 25 leads to a melting temperature (T_m) of 66 $^\circ\text{C}$, which represents a remarkable increase of 9.2 $^\circ\text{C}$ indicating strong binding. Furthermore, the concentration-dependent change of T_m in particular supports the notion of a specific protein-ligand interaction (Fig. 5b). In summary, while one compound was responsible for a decrease of T_m , 19 out of 85 compounds cause a significant increase in T_m . Table 2 lists the T_m of each hit.

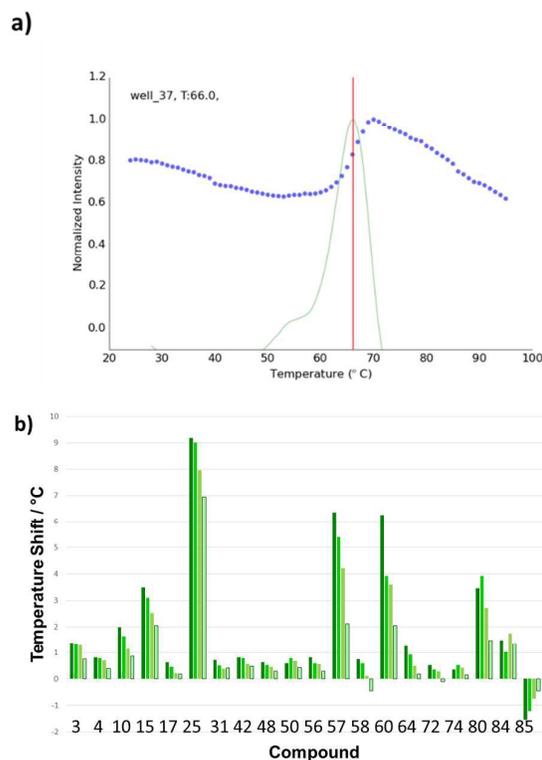


Fig. 5: Thermal shift assays. Panel a) shows a representative example of the experimental fluorescence signal with compound 25 shown in blue, the first derivative in green and the T_m as a red line. Data were analysed using the program NAMI developed in-house³⁹. b) Concentration dependent-thermal shifts of the best 20

Journal Name

ARTICLE

compounds at concentrations of 400, 320, 160 and 80 μM , respectively, depicted from left to right in dark green to light green. Note, that 85 led to a decrease in thermal stability.

Table 2. Chemical structures and melting temperature of the 20 potential leads identified by TSAs.

No	Structure	ΔT_m /°C
3		1.4
4		0.8
10		2.0
15		3.5
17		0.6

25		9.2
31		0.7
42		0.8
48		0.6
50		0.6
56		0.8
57		6.3

ARTICLE

Journal Name

58		0.8
60		6.3
64		1.3
72		0.5
74		0.4
80		3.5
84		1.5

85		-1.6
----	--	------

Growth Inhibition Assays Confirm Several Hits as ETH Boosters – Fifteen of the best hit compounds were tested in a growth inhibition assay based on fluorescent GFP *M. tuberculosis* H37Rv strains.²³ Boosting activity in the presence of ETH at 0.1 $\mu\text{g/mL}$ (approximately 20 times lower than the MIC of ETH alone) was assessed, with seven compounds demonstrating some boosting effect (table 3). The identified boosters include compound 42 and compound 25 for which we have solved the co-crystal structures.

Table 3: Summary of the growth inhibition of *M. tuberculosis* HR37v GFP strains.

No of compound	EC ₅₀ [μM]	No of compound	EC ₅₀ [μM]
3	n/a	48	34
4	n/a	50	> 90
10	n/a	57	n/a
15	> 90	60	> 90
25	0.76	64	n/a
31	n/a	72	n/a
42	> 90	84	> 90
48	34	85	n/a
50	> 90		

Crystal Structures of EthR with Novel Ligands - Crystallization experiments were performed with the 20 compounds identified through the TSA screen to determine the exact binding mode and verify the mechanism of EthR binding and inhibition. Four co-crystal structures at high resolution ranging from 1.4-2.1 \AA are presented here (see SI Table S3 for crystallographic data). As expected all crystal structures show the same overall fold of the EthR-homodimer in the *inactive* conformation with RMSDs of all C α -atoms in one protein monomer ranging from 0.2-0.3 \AA . As in all previous co-crystal structure determined so far, ligand binding appears to lock the protein structure in a conformation where the putative DNA-recognition helices are in a relative orientation that makes their binding to the major groove as seen in other members of the TetR family physically impossible.¹³

Structure of EthR with compound 25 – The structure of EthR with N-butyl-4-methyl-1-piperidine carboxamide was determined at a resolution of 1.4 \AA (Fig. 6a), which represents the highest resolution for any EthR-ligand complex reported so far. However, this

structure represents a serendipitous crystallization event since after solving the crystal structure it was discovered that this compound was ordered/delivered by mistake and represents a much smaller piperidine than originally planned. In fact, with a molecular weight of 193 g/mol this compound would have been excluded in the initial pre-screening filtering. Nevertheless, the compound contains the same basic building blocks and retains the key protein-ligand interactions, namely the hydrogen bond to Asn179 (2.8 Å), a second hydrogen bond to Asn176 (2.8 Å) and hydrophobic interactions with Phe110 and Trp145. Importantly, the carboxamide-piperidine moiety occupies the same position as in the EthR-co-crystal structure of compound BDM31369 (PDB entry 3Q0V). These results show that compound 25 represents an ideal starting point for a fragment-growing approach to hit optimization.

Structure of EthR with compound 10 - The second compound co-crystallised with EthR, represents a novel scaffold with a central carbonyl function flanked by a substituted 1,2-oxazole moiety and a substituted 1H, 4H, 5H, 6H-pyrrolo group. The structure was also determined at very high resolution showing unambiguously that this comparatively large ligand can adopt two distinct orientations, each with the occupancy of 50% in the binding site (Fig. 6b). The orientation shown in blue (A) maintains the key hydrogen bond to Asn179 whereas in the second position (B) shown in yellow the carbonyl-function of the ligand forms a hydrogen bond to Thr149. Surprisingly, no other potential hydrogen bond donor/acceptor of the ligand is involved in protein-ligand interactions and there is no space for any water-mediated contacts. Instead both crystallographic binding modes appear to be dominated by hydrophobic interactions. In orientation A the aliphatic chain of the 5-propyl-1,2-oxazole is perfectly located in the hydrophobic end of the binding pocket composed of residues Leu183, Phe184 and Trp138. In contrast, the key interaction in orientation B appears to involve hydrophobic stacking of the pyrrolo ring system with Phe110. In summary, the crystal structure shows the promiscuity of the EthR binding pocket able to accommodate multiple ligand binding poses.

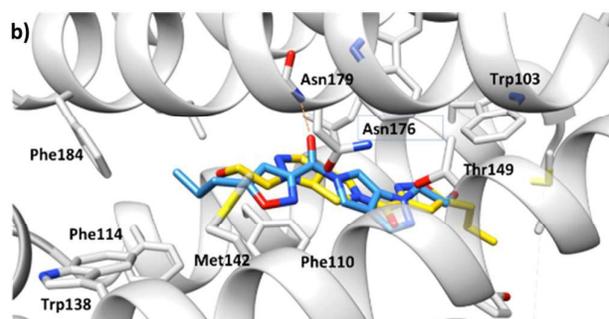
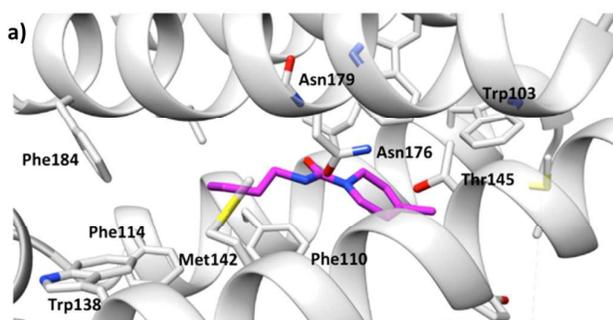


Fig. 6: Panel **a**) Close-up of the ligand binding pocket of EthR shown in grey ribbon representation co-crystallised with compound 25 with Carbon bonds shown in magenta, nitrogen in blue, oxygen in red. **b**) Close-up of the ligand-binding pocket of EthR co-crystallised with compound 10. The two binding modes are depicted in blue and yellow, respectively.

Structure of EthR with compound 42 - The structure of EthR in complex with 1-(3-{5-[(1R,2R)-2-methylcyclopropyl] furan-2-yl}propanoyl)piperidine-4-carboxamide was solved to a resolution of 1.95 Å (Fig. 7a). This ligand represents the second novel chemical scaffold. It is noteworthy that the carbonyl group of the linker does not form a hydrogen bond to either Asn176 or Asn179 as one may have expected. Presumably this is due to the relatively long linker region combined with two ring systems in the hydrophobic pocket. In this position the furan ring maintains an optimal position for hydrophobic stacking contacts with Phe110 while the 2-methylcyclopropyl ring is surrounded by Phe114, Leu183, and Phe184 respectively. This particular binding mode underlines the importance of hydrophobic interactions that can compensate for the cost of losing one hydrogen bond. In addition, this structure reveals quite important side chain flexibility upon ligand binding presumably by an induced fit mechanism.⁴¹

Structure of EthR with compound 57 - The structure of EthR in complex with N-[(2R)-2-hydroxy-2-(4-nitrophenyl) ethyl]-2H-1,3-benzodioxole-5-carboxamide was determined at a resolution of 2.1 Å (Fig. 7b). This compound induces a particularly high and strongly concentration-dependent thermostabilisation of EthR. The ligand forms a number of specific hydrogen bonds, which include residues Thr149 and Asn176 via its alcohol oxygen in the linker region, and Asn179 via the carboxamide carbonyl function. The 1,3-benzodioxole moiety of the ligand is located as expected in the hydrophobic deep end of the binding site with VdW contacts to the hydrophobic side chains of residues Leu183, Phe184, and Phe114, respectively.

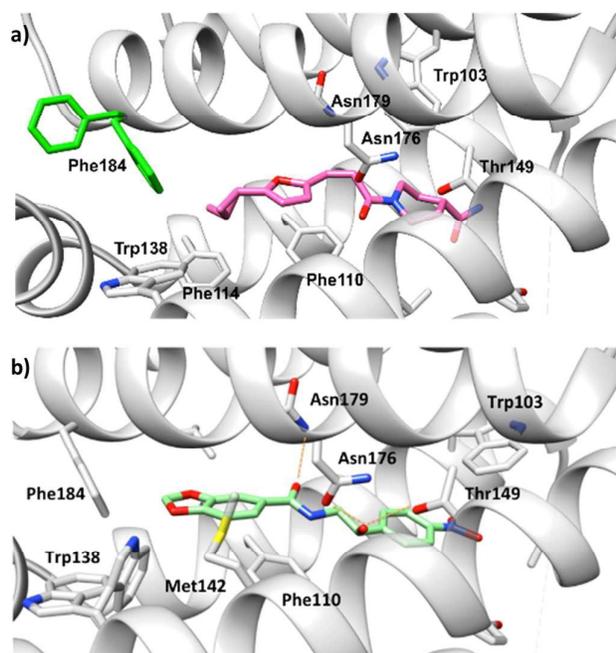


Fig. 7: a) Close-up of the ligand-binding pocket of EthR co-crystallised with compound 42. Phe184 is observable in two orientations in the electron density, both are modelled and shown in green. b) Close-up of the ligand-binding pocket of EthR co-crystallised with compound 57.

Taken together the four crystal structures presented here reveal the crucial binding motifs of ligands, provide important information about the flexibility of the binding pocket and provide the basis of ligand optimization based on new chemical scaffolds.

Discussions

Since the early 1980s structure-based drug discovery and design has become an important tool in pharmaceutical and academic research institutions.⁴²⁻⁴⁴ For more and more pharmaceutical targets, structural information at atomic resolution based on one or – increasingly – several X-ray crystal structures are now available. The ever-growing number of relevant PDB entries has been greatly supported by the various structural genomics endeavours that had drug discovery and design at the core of their aims.⁷ The development of efficient and practical methods to optimally exploit the wealth of structural information is therefore crucial. In this work we present a computational strategy for filtering the 6.06 million compounds in the *DrugsNow* subset of the ZINC database to obtain a chemically diverse set of 409,201 compounds tailored to the transcriptional regulator EthR. After selecting the most suitable protein crystal structure by employing a *cross-docking* experiment with a decoy data set, all compounds were docked into the binding site. It is noteworthy that one of the original crystal structures fortuitously co-crystallised with hexadecyl octanoate proved most receptive to docking presumably due to the fact the binding site geometry is not biased through an induced fit by an unnatural ligand. Remarkably only six compounds failed to produce any

computationally feasible binding pose validating the pro-docking filtering strategy. A careful post-docking filtering, which included clustering the results into 23 different ring chemistries to ensure diversity, a shortlist of 284 compounds with five binding poses each was obtained. A visual analysis of all binding poses and ligand geometries was used in combination with an automatic pipeline implementing interactive graphics and tools of GOLDMine, which led to a final subset of 85 potential ligands. These *in silico* hits were then evaluated for their capacity to interact with EthR. Twenty compounds produced a significant shift in thermal protein stability, which provides clear evidence of their binding to EthR¹⁸. As docking scores and binding assays generally show poor correlation, the straightforward biophysical characterisation of all compounds by thermal shift assays proved essential.

The co-crystal structures of these four compounds confirm the mode of action. Binding of the inhibitor locks the EthR dimer into a conformation that is incompatible with DNA binding of the two putative DNA-recognition helices in the major grooves of the DNA operator. In addition, the high-resolution structures of these new chemical scaffolds can be evaluated in the light of the docking results, which now suggest new directions for the consecutive medicinal chemistry approach.

Compound 10 – This compound clearly adopts two distinct binding modes in the crystal structure as described above. Overall both binding modes correspond approximately to two of the predicted docking poses. Our more detailed comparison shows that binding pose 1 is more closely related to the crystallographic mode A with the same hydrogen bonding pattern, whereas 3 and 4 closely resemble the second crystallographic position (Fig. 8). These results clearly suggest that the scoring function alone is not a sensitive predictor of the best binding mode, but that the multiple poses obtained by the GA docking algorithm may indicate promiscuity in the pocket. Furthermore, the docking poses show that there is additional free space in the binding pocket, which opens an obvious perspective for further evolution of this ligand. Standard *O*-alkylation of the free alcohol function could advantageously complete the structure with a methyl, ethyl or isopropyl group, filling the hydrophobic space of the binding pocket while maintaining the key hydrogen bond to Asn179 revealed by the crystallographic binding mode 2. Extending the size on this side of the ligand is very likely to render binding mode B energetically less favourable.

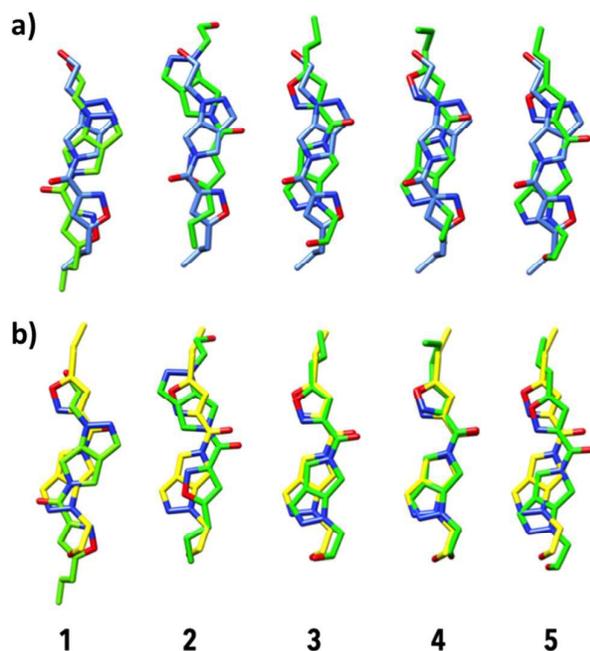


Fig. 8: Superposition of the top five binding poses of compound 10 (in green) with the two binding modes found in the crystal structure depicted in **a)** blue and **b)** yellow respectively.

Compound 42 – In contrast to the previous ligand, compound 42 adopts a well-defined binding mode, however, somewhat surprisingly, the carbonyl oxygen of the central amide linker does not form a hydrogen bond to either Asn179 or Asn176 as predicted by all of the binding poses and despite at least one ligand-protein hydrogen bond being an important criteria the filtering process. Docking scores are in the same range as seen for compound 10.

In the crystal structure the methylcyclopropyl group appears to fill the hydrophobic end of the pocket as supported by four out of the five binding modes (Fig. 9). In order to maintain these interactions while enabling the hydrogen bond of the amide carbonyl oxygen to Asn179 (or Asn176) the propanoyl linker needed to be shortened, for instance by coupling the piperidine with the corresponding activated acyl chloride.

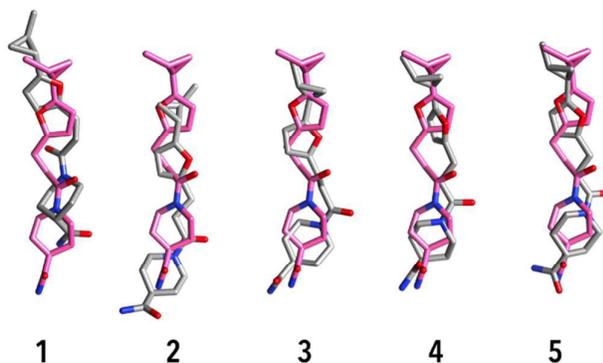


Fig. 9: Superposition of the best five binding poses of compound 42 (grey) with the binding mode in the crystal structure (magenta).

Compound 57 – This compound resulted in the highest shift in melting temperature with a ΔT_m of 6.3°C. Two of its binding poses correspond to the position found in the crystal structure with pose 3 being almost identical (Fig. 10).

The 1,3-benzodioxole moiety seems to fill the hydrophobic end of the binding pocket almost perfectly while the amide group is almost co-planar to Phe110 providing optimal pi-stacking interactions. The carbonyl amide and the free alcohol group of the linker form a network of hydrogen bonds to Asn179, Asn176 and Thr149. It is likely that these strong ligand-protein interactions are responsible for the important thermostabilization of the complex as revealed by thermal shift assays.

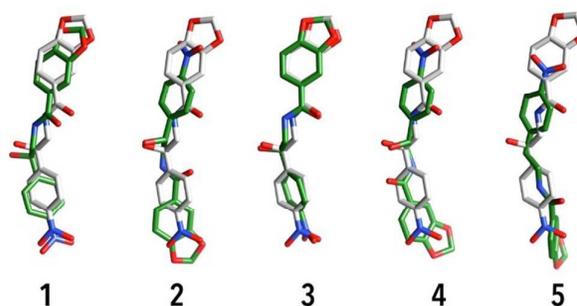


Fig. 10: Superposition of the best five binding poses of compound 57 (grey) with the binding mode in the crystal structure (green).

Biological Activity - Of the fifteen compounds tested five showed modest boosting ability in addition to their *in-vitro* binding. One compound, No. 48 demonstrated a promising EC_{50} value of 34 μM , while No. 25 resulted in an EC_{50} value of 760 nM. Compound 25 is of particular interest due to its relatively small size which makes it very amenable for elaboration in a medicinal chemistry program. Though not identified by the original virtual screening process, this compound was subject to the same vigorous biophysical, structural and biological tests and immediately recognized to be a perfect starting point for further development. Comparing the results of thermal shift assays and biological activity shows limited correlation, presumably due to differing compound uptake by the pathogen. This underlines again the importance of biological assays early in the drug discovery pipeline but with such assays in hand, the qualities of binding, penetrance and cellular activity can be optimised in tandem.

Conclusions

In this work, we have developed a robust and versatile strategy for *in-silico* hit discovery based on the ZINC database and exploiting the wealth of information in multiple crystal structures. Employing tailored pre- and post-docking filtering

pipelines that ensured high chemical diversity we identified 20 new potential candidates for hit optimization. Importantly, this pipeline can easily be adapted to other cases where new chemical scaffolds are required for structurally well-characterized target proteins. Given the ever-increasing number of high-resolution crystal structures in the PDB, this approach will continue to become ever more important. In addition, we determined four new co-crystal structures of EthR-ligand complexes to resolution of up to 1.4 Å, confirming the inhibition mechanism through the opening of the DNA-binding HTH jaw. The structures reveal the mode of ligand to protein binding in atomic detail. Compound 25 presented a relatively small molecule with a molecular weight of less than 200 Da and could thus serve as a perfect starting point for a ligand-growing strategy. Compound 10 presented a dual binding mode in the crystal structure highlighting the challenges for rational drug design in the case of largely hydrophobic pockets. Compounds 42 and 57 are probably the most promising hits with well-defined binding modes, and in the case of 57 exploiting all possible hydrogen bonds in the binding pocket. Finally, biological assays demonstrated that six compounds have a modest yet significant boosting effect, allowing the future development of cell-penetrant, biologically active EthR inhibitors.

Acknowledgements

We gratefully acknowledge funding from the Biophysical Sciences Institute and from the French ANR-14-CE14-0027. NJT is grateful to the CDC for a pre-doctoral fellowship. We thank E. Dickinson and I. Edwards for their excellent technical support, the University of Cambridge High Performance Computing for access to their supercomputer, and the Diamond Light Source and their staff for building and maintaining world leading facilities.

References

1. WHO, *World Health Organisation: Global Tuberculosis Report*. 2015.
2. Gengenbacher, M.; Kaufmann, S. H., Mycobacterium tuberculosis: success through dormancy. *FEMS Microbiol Rev* **2012**, *36*, 514-32.
3. Jain, A.; Mondal, R., Extensively drug-resistant tuberculosis: current challenges and threats. *FEMS Immunol Med Microbiol* **2008**, *53*, 145-50.
4. Velayati, A. A.; Masjedi, M. R.; Farnia, P.; Tabarsi, P.; Ghanavi, J.; Ziazarifi, A. H.; Hoffner, S. E., Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* **2009**, *136*, 420-5.
5. Caminero, J. A.; Sotgiu, G.; Zumla, A.; Migliori, G. B., Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet Infect Dis* **2010**, *10*, 621-9.
6. Wejse, C., Tuberculosis elimination in the post Millennium Development Goals era. *Int J Infect Dis* **2015**, *32*, 152-5.
7. Terwilliger, T. C.; Park, M. S.; Waldo, G. S.; Berendzen, J.; Hung, L. W.; Kim, C. Y.; Smith, C. V.; Sacchettini, J. C.; Bellinzoni, M.; Bossi, R.; De Rossi, E.; Mattevi, A.; Milano, A.; Riccardi, G.; Rizzi, M.; Roberts, M. M.; Coker, A. R.; Fossati, G.; Mascagni, P.; Coates, A. R.; Wood, S. P.; Goulding, C. W.; Apostol, M. I.; Anderson, D. H.; Gill, H. S.; Eisenberg, D. S.; Taneja, B.; Mande, S.; Pohl, E.; Lamzin, V.; Tucker, P.; Wilmanns, M.; Colovos, C.; Meyer-Klaucke, W.; Munro, A. W.; McLean, K. J.; Marshall, K. R.; Leys, D.; Yang, J. K.; Yoon, H. J.; Lee, B. I.; Lee, M. G.; Kwak, J. E.; Han, B. W.; Lee, J. Y.; Baek, S. H.; Suh, S. W.; Komen, M. M.; Arcus, V. L.; Baker, E. N.; Lott, J. S.; Jacobs, W., Jr.; Alber, T.; Rupp, B., The TB structural genomics consortium: a resource for Mycobacterium tuberculosis biology. *Tuberculosis (Edinb)* **2003**, *83*, 223-49.
8. Chim, N.; Habel, J. E.; Johnston, J. M.; Krieger, I.; Miallau, L.; Sankaranarayanan, R.; Morse, R. P.; Bruning, J.; Swanson, S.; Kim, H.; Kim, C. Y.; Li, H.; Bulloch, E. M.; Payne, R. J.; Manos-Turvey, A.; Hung, L. W.; Baker, E. N.; Lott, J. S.; James, M. N.; Terwilliger, T. C.; Eisenberg, D. S.; Sacchettini, J. C.; Goulding, C. W., The TB Structural Genomics Consortium: a decade of progress. *Tuberculosis (Edinb)* **2011**, *91*, 155-72.
9. Cole, S. T.; Riccardi, G., New tuberculosis drugs on the horizon. *Curr Opin Microbiol* **2011**, *14*, 570-6.
10. Villemagne, B.; Crauste, C.; Flipo, M.; Baulard, A. R.; Deprez, B.; Willand, N., Tuberculosis: the drug development pipeline at a glance. *Eur J Med Chem* **2012**, *51*, 1-16.
11. Zumla, A. I.; Gillespie, S. H.; Hoelscher, M.; Philips, P. P.; Cole, S. T.; Abubakar, I.; McHugh, T. D.; Schito, M.; Maeurer, M.; Nunn, A. J., New antituberculosis drugs, regimens, and adjunct therapies: needs, advances, and future prospects. *Lancet Infect Dis* **2014**, *14*, 327-40.
12. Engohang-Ndong, J.; Baillat, D.; Aumercier, M.; Bellefontaine, F.; Besra, G. S.; Loch, C.; Baulard, A. R., EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Mol Microbiol* **2004**, *51*, 175-88.
13. Deng, W.; Li, C.; Xie, J., The underlying mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cell Signal* **2013**, *25*, 1608-13.
14. Dessen, A.; Quemard, A.; Blanchard, J. S.; Jacobs, W. R., Jr.; Sacchettini, J. C., Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. *Science* **1995**, *267*, 1638-41.
15. Willand, N.; Dirie, B.; Carette, X.; Bifani, P.; Singhal, A.; Desroses, M.; Leroux, F.; Willery, E.; Mathys, V.; Deprez-Poulain, R.; Delcroix, G.; Frenois, F.; Aumercier, M.; Loch, C.; Villeret, V.; Deprez, B.; Baulard, A. R., Synthetic EthR inhibitors boost antituberculous activity of ethionamide. *Nat Med* **2009**, *15*, 537-44.
16. Frenois, F.; Engohang-Ndong, J.; Loch, C.; Baulard, A. R.; Villeret, V., Structure of EthR in a ligand bound conformation

reveals therapeutic perspectives against tuberculosis. *Mol Cell* **2004**, *16*, 301-7.

17. Dover, L. G.; Corsino, P. E.; Daniels, I. R.; Cocklin, S. L.; Tatituri, V.; Besra, G. S.; Futterer, K., Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. *J Mol Biol* **2004**, *340*, 1095-105.

18. Flipo, M.; Desroses, M.; Lecat-Guillet, N.; Villemagne, B.; Blondiaux, N.; Leroux, F.; Piveteau, C.; Mathys, V.; Flament, M. P.; Siepmann, J.; Villeret, V.; Wohlkonig, A.; Wintjens, R.; Soror, S. H.; Christophe, T.; Jeon, H. K.; Loch, C.; Brodin, P.; Deprez, B.; Baulard, A. R.; Willand, N., Ethionamide boosters. 2. Combining bioisosteric replacement and structure-based drug design to solve pharmacokinetic issues in a series of potent 1,2,4-oxadiazole EthR inhibitors. *J Med Chem* **2012**, *55*, 68-83.

19. Tatum, N. J.; Villemagne, B.; Willand, N.; Deprez, B.; Liebeschuetz, J. W.; Baulard, A. R.; Pohl, E., Structural and docking studies of potent ethionamide boosters. *Acta Crystallogr C* **2013**, *69*, 1243-50.

20. Flipo, M.; Willand, N.; Lecat-Guillet, N.; Hounsou, C.; Desroses, M.; Leroux, F.; Lens, Z.; Villeret, V.; Wohlkonig, A.; Wintjens, R.; Christophe, T.; Kyoung Jeon, H.; Loch, C.; Brodin, P.; Baulard, A. R.; Deprez, B., Discovery of novel N-phenylphenoxyacetamide derivatives as EthR inhibitors and ethionamide boosters by combining high-throughput screening and synthesis. *J Med Chem* **2012**, *55*, 6391-402.

21. Villemagne, B.; Flipo, M.; Blondiaux, N.; Crauste, C.; Malaquin, S.; Leroux, F.; Piveteau, C.; Villeret, V.; Brodin, P.; Villoutreix, B. O.; Sperandio, O.; Soror, S. H.; Wohlkonig, A.; Wintjens, R.; Deprez, B.; Baulard, A. R.; Willand, N., Ligand efficiency driven design of new inhibitors of Mycobacterium tuberculosis transcriptional repressor EthR using fragment growing, merging, and linking approaches. *J Med Chem* **2014**, *57*, 4876-88.

22. Nikiforov, P. O.; Surade, S.; Blaszczyk, M.; Delorme, V.; Brodin, P.; Baulard, A. R.; Blundell, T. L.; Abell, C., A fragment merging approach towards the development of small molecule inhibitors of Mycobacterium tuberculosis EthR for use as ethionamide boosters. *Org Biomol Chem* **2016**, *14*, 2318-26.

23. Blondiaux, N.; Moune, M.; Desroses, M.; Frita, R.; Flipo, M.; Mathys, M.; Soetaert, K.; Kiass, M.; Delorme, V.; Djaout, K.; Trebosc, V.; Kemmer, C.; Wintjens, R.; Wohlkönig, A.; Antoine, R.; Huot, L.; Hot, D.; Coscolla, M.; Feldmann, J.; Gagneux, S.; Loch, C.; Brodin, P.; Gitzinger, M.; Déprez, B.; Willand, N.; Baulard, A. R., Reversion of antibiotic resistance in Mycobacterium tuberculosis by spiroisoxazoline SMARt-420. *Science* **2017**, *355*, 1206-11.

24. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G., ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* **2012**, *52*, 1757-68.

25. Craig, I. R.; Essex, J. W.; Spiegel, K., Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. *J Chem Inf Model* **2010**, *50*, 511-24.

26. Korb, O.; Olsson, T. S.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C., Potential and limitations of ensemble docking. *J Chem Inf Model* **2012**, *52*, 1262-74.

27. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609-23.

28. Berman, H. M.; Kleywegt, G. J.; Nakamura, H.; Markley, J. L., The Protein Data Bank archive as an open data resource. *J Comput Aided Mol Des* **2014**, *28*, 1009-14.

29. Willand, N.; Desroses, M.; Toto, P.; Dirie, B.; Lens, Z.; Villeret, V.; Rucktooa, P.; Loch, C.; Baulard, A.; Deprez, B., Exploring drug target flexibility using in situ click chemistry: application to a mycobacterial transcriptional regulator. *ACS Chem Biol* **2010**, *5*, 1007-13.

30. Flipo, M.; Desroses, M.; Lecat-Guillet, N.; Dirie, B.; Carette, X.; Leroux, F.; Piveteau, C.; Demirkaya, F.; Lens, Z.; Rucktooa, P.; Villeret, V.; Christophe, T.; Jeon, H. K.; Loch, C.; Brodin, P.; Deprez, B.; Baulard, A. R.; Willand, N., Ethionamide boosters: synthesis, biological activity, and structure-activity relationships of a series of 1,2,4-oxadiazole EthR inhibitors. *J Med Chem* **2011**, *54*, 2994-3010.

31. Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G., Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J Mol Biol* **2003**, *326*, 607-20.

32. Beisken, S.; Meinel, T.; Wiswedel, B.; de Figueiredo, L. F.; Berthold, M.; Steinbeck, C., KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics* **2013**, *14*, 257.

33. Baell, J. B.; Holloway, G. A., New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* **2010**, *53*, 2719-40.

34. Wallach, I.; Lilien, R., Virtual decoy sets for molecular docking benchmarks. *J Chem Inf Model* **2011**, *51*, 196-202.

35. Liebeschuetz, J. W.; Cole, J. C.; Korb, O., Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J Comput Aided Mol Des* **2012**, *26*, 737-48.

36. Cottrell, S. J.; Olsson, T. S.; Taylor, R.; Cole, J. C.; Liebeschuetz, J. W., Validating and understanding ring conformations using small molecule crystallographic data. *J Chem Inf Model* **2012**, *52*, 956-62.

37. Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. A.; Lane, P.; Salemme, F. R., High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **2001**, *6*, 429-40.

38. Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J., Jr.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H.; Hui, R.; Marsden, B. D.; Nordlund, P.; Sundstrom, M.; Weigelt, J.; Edwards, A. M., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **2006**, *103*, 15835-40.

39. Groftehauge, M. K.; Hajizadeh, N. R.; Swann, M. J.; Pohl, E., Protein-ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr D Biol Crystallogr* **2015**, *71*, 36-44.

40. Surade, S.; Ty, N.; Hengrung, N.; Lechartier, B.; Cole, S. T.; Abell, C.; Blundell, T. L., A structure-guided fragment-based approach for

ARTICLE

Journal Name

the discovery of allosteric inhibitors targeting the lipophilic binding site of transcription factor EthR. *Biochem J* **2014**, *458*, 387-94.

41. Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C., Protein Binding Pocket Dynamics. *Acc Chem Res* **2016**, *49*, 809-15.

42. Hol, W. G. J., Protein Crystallography and Computer-Graphics toward Rational Drug Design. *Angew Chem Int Edit* **1986**, *25*, 767-778.

43. Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angewandte Chemie-International Edition* **2003**, *42*, 2718-2736.

44. Brown, D. G.; Shotton, E. J., Diamond: shedding light on structure-based drug discovery. *Philos T R Soc A* **2015**, 373.

Three new chemical scaffolds for the inhibition of the transcriptional regulator EthR from *M. tuberculosis* have been identified and verified by biophysical and biological assays.

