CHEMBIOCHEM

Supporting Information

Biofragments: An Approach towards Predicting Protein Function Using Biologically Related Fragments and its Application to *Mycobacterium tuberculosis* CYP126

Sean A. Hudson,[a, d] Ellene H. Mashalidis,[a, e] Andreas Bender,[b] Kirsty J. McLean,[c] Andrew W. Munro,[c] and Chris Abell*[a]

cbic_201300697_sm_miscellaneous_information.pdf

Supporting Information
Table of Contents

Supplementary Figures and Tables

Table S1. PDB CYP substrates by organism*

Taxonomy	Organism	CYP ligands
Bacteria	*Bacillus megaterium*	5
	Bacillus subtilis	2
	Citrobacter braakii	1
	Escherichia coli	4
	Mycobacterium tuberculosis	8
	Novosphingobium aromaticivorans	1
	Pseudomonas putida	13
	Saccharopolyspora erythraea	4
	Sorangium cellulosum	2
	Streptomyces avermitilis	1
	Streptomyces coelicolor	6
	Streptomyces griseolus	1
	Streptomyces sp. TP-A0274	2
	Streptomyces venezuelae	3
	Synechocystis sp.	1
Eukaryota	*Arabidopsis thaliana*	3
	Danio rerio	1
	Fusarium oxysporum	1
	Homo sapiens	16
	Oryctolagus cuniculus	3
	Parthenium argentatum	1
	Rattus norvegicus	6
	Saccharomyces cerevisiae	5

*A few substrates appear across multiple organisms

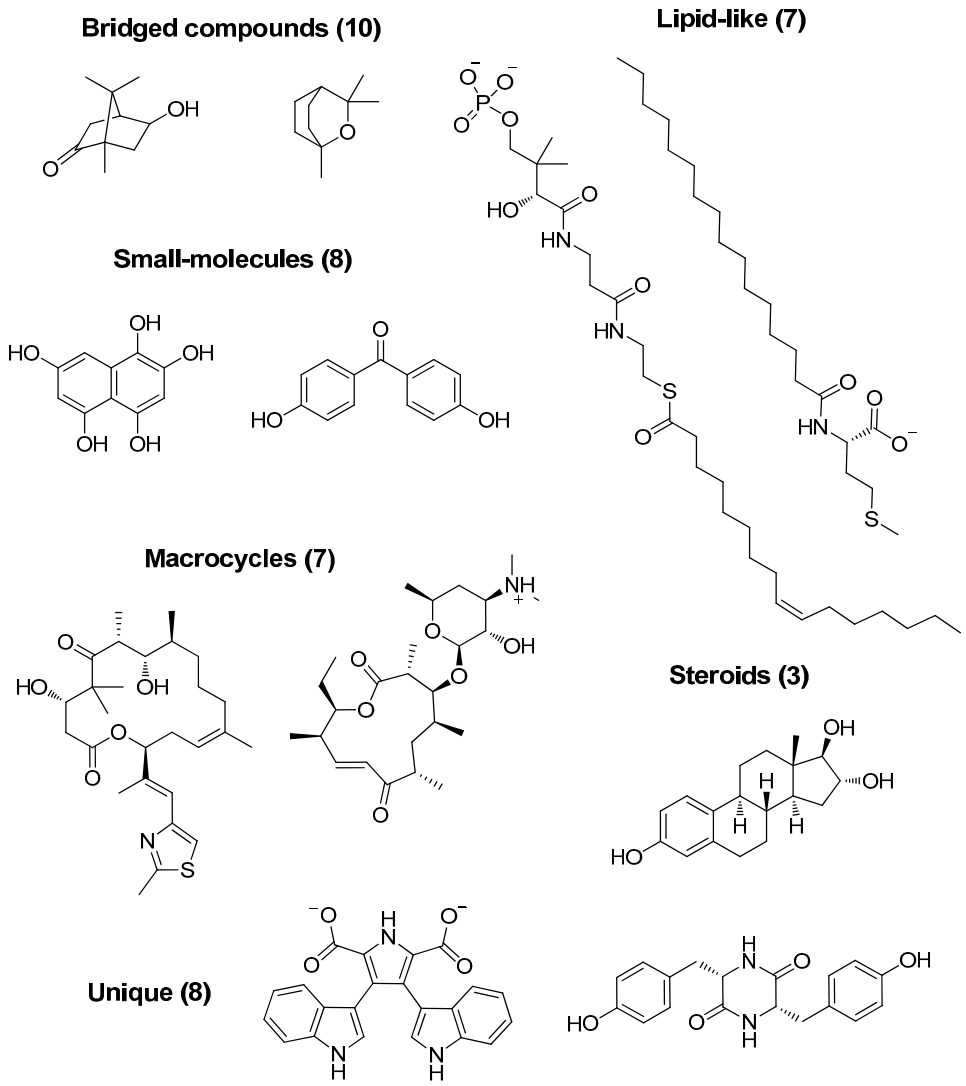

Figure S1. By-eye structural grouping of PDB bacterial CYP substrates and substrate-like/mimetic ligands. The number of members within each grouping is indicated in brackets. Note the prevalence of structural motifs that are not small flat heterocyclic molecules.

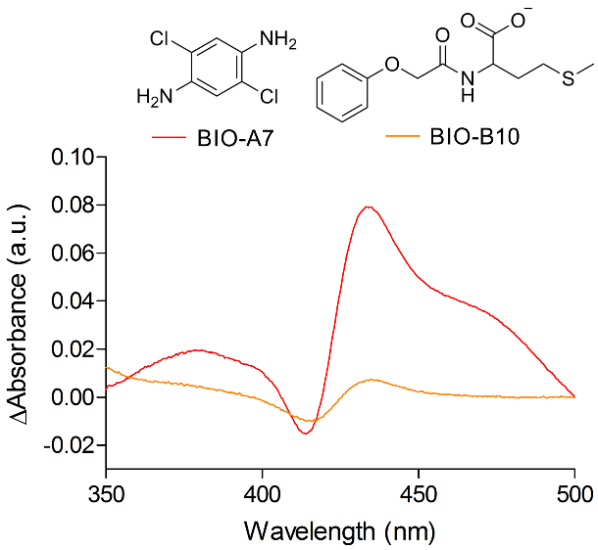

Figure S2. Absorbance difference spectra for CYP126 (5 µM) with Biofragment hits BIO-A7 and B10 (1 mM) inducing a type-II red shift in the heme Soret absorbance band.

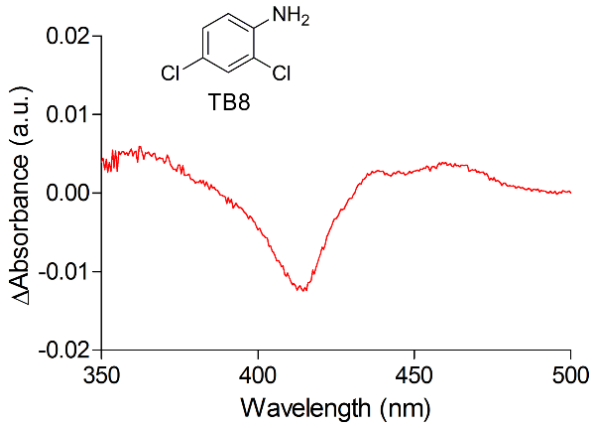

Figure S3. Absorbance difference spectra for CYP126 (5 µM) with TB8 (1 mM) inducing a weak type-II red shift in the heme absorbance.

Experimental Methods

Materials and libraries

All reagents and chemicals were of analytical grade and were obtained from Sigma-Aldrich (Dorset, England) unless otherwise specified. The traditional drug-like fragment library was a first-generation Rule of Three[1] compliant set of 1250 purchased from Maybridge (Cornwall, UK). No known target-specific bias was placed on the fragments selected for incorporation. The vast majority of the commercial fragments contain benzene and/or heterocyclic aromatic subunits. The fragments have a calculated aqueous solubility of >1 mM according to the vendor. The CYP Biofragment library was designed as described in the main text. Individual member compounds were purchased from: Sigma-Aldrich, Specs (Delft, The Netherlands), Ambinter (Paris, France), ChemBridge Corporation (San Diego, CA, USA), MP Biomedicals (Illkirch, France) or Bachem (Weil am Rhein, Germany). The library was stored as a 100 mM stock in DMSO-d_6 at -80 ºC in a WebSeal 96 glass vial rack (500 µL) sealed with a silicone/PTFE mat (Thermo Fisher Scientific, Rockwood, TN, USA).

In silico analysis

All chemical structure database/library collation, filtering, physicochemical property calculations/predictions, similarity coefficient comparisons (using connectivity molecular fingerprints (ECFP6)) and chemical diversity analysis was performed in Pipeline Pilot[2] 6.1.5.0 Student Edition (Accelrys, San Diego, USA), unless otherwise described. Comparison plots/graphs were constructed with either Pipeline Pilot or GraphPad Prism 5.01 (GraphPad Software, San Diego, USA).

Virtual TB metabolome

A virtual TB metabolome was constructed by extracting all small-molecules from the *Mtb* H37Rv pathways in the KEGG PATHWAY database (http://www.genome.jp/kegg/pathway.html, extraction date: 22/7/12). There were 110 biochemical pathways annotated for *Mtb*, comprising a final dataset of 3856 metabolites.

Expression and purification of CYP126

A pET15b/*CYP126* expression vector was constructed with the wild-type *Mtb* H37Rv CYP126 gene cloned between the NdeI and BamHI restriction sites using standard molecular biology protocols.[3] Recombinant N-terminal His_6-tagged CYP126 was expressed and purified from this vector using the protocol previously described for His_6-tagged CYP121.[4] The mass of purified CYP126 was confirmed on an Applied Biosystems QSTAR nanoESI QTOF mass spectrometer (Applied Biosystems, CA, USA): expected 47986 Da without the initial methionine; observed 47986.

NMR spectroscopy screening

All CYP Biofragments were initially filtered through an STD[5] experiment using only a protein-containing sample. Samples (200 µL) comprising 1 mM CYP Biofragment with 15 µM CYP126 were prepared in 100 mM potassium phosphate buffer at pH 7.5 with 100 mM potassium chloride, 10% (v/v) D_2O, 20 µM deuterated 3-trimethylsilylpropanoate (TSP-d_4) and 2.5% (v/v) DMSO-d_6. The samples were pipetted into 3-mm NMR capillaries (Hilgenberg GmbH, Malsfeld, Germany) and loaded into 528-PP-8 NMR tubes (Wilmad-LabGlass, NJ, USA). STD[5] 1D ^{1}H NMR spectra were acquired at 278 K on a Bruker DRX 700 MHz NMR spectrometer (Bruker, MA, USA) equipped with a 5 mm triple resonance inverse (TXI) cryoprobe with *z*-gradients and an auto sampler. STD experiments utilised a 40 ms selective Gaussian 180° pulse at a frequency alternating between on-resonance (0.9 or 0.4 ppm for the drug-like fragment screening and Biofragment screening, respectively) and off-resonance (40 ppm) after every scan. The primary STD screen identified 30 first-pass hits which were then analyzed more rigorously using the suite of 1D ^{1}H NMR experiments: CPMG[6], STD[5] and WaterLOGSY[7], acquired as above including with relevant controls without protein. Ketoconazole (250 µM) was added to samples with an additional 2.5% (v/v) DMSO-d_6 for CPMG displacement experiments. CPMG experiments employed a relaxation delay of 100-400 ms. WaterLOGSY experiments employed a 20 ms selective Gaussian 180° shaped pulse at the water signal frequency and an NOE mixing time of 1 s. Water signal suppression was achieved using a W5 Watergate gradient spin-echo pulse sequence.[8] All spectra were processed using TopSpin 3.0 (Bruker, Coventry, UK) and the CPMG and WaterLOGSY spectra were scaled relative to the 20 µM TSP peak intensity at 0 ppm. All pulse sequences were provided by Dr Glyn Williams (Astex Therapeutics, UK).

Electronic spectroscopy

Heme absorbance shift assays were performed as previously described for azole antifungals binding to CYP121.[9] Absorption spectra were recorded at room temperature on a Cary 400 UV-Vis spectrophotometer (Varian, CA, USA). Individual samples (200 µL) comprised 5 µM CYP126 in 50 mM Tris-HCl buffer at pH 7.5 with 1 mM EDTA, 5% (v/v) DMSO as co-solvent, and CYP Biofragment or TB metabolite at a single-point concentration (1 mM) for screening, or at varying concentrations (0-1000 µM) for complete titrations, as required to achieve the maximal heme absorbance band shift consistent with ligand saturation. Plots and curve fitting for K_D determination were done with GraphPad Prism 5.01 (GraphPad Software, San Diego, USA). It should be noted that although the top 35 commercially available TB metabolites from searches were purchased, only 23 were found to be viable for screening by heme absorbance shift, i.e. not too inherently absorptive at 350-450 nm or insoluble at 1 mM in aqueous buffer.

References

[1] M. Congreve, R. Carr, C. Murray, H. Jhoti, *Drug Discov. Today* **2003**, *8*, 876.

[2] W. A. Warr, *J. Comput. Aided Mol. Des.* **2012**, *26*, 801.

[3] J. Sambrook, D. W. Russell, *Molecular cloning: a laboratory manual*, 3rd ed., Cold Spring Harbor Laboratory Press, New York, **2001**.

[4] S. A. Hudson, K. J. McLean, S. Surade, Y. Q. Yang, D. Leys, A. Ciulli, A. W. Munro, C. Abell, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 9311.

[5] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784.

[6] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, *J. Am. Chem. Soc.* **1997**, *119*, 12257.

[7] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, M. Sundstrom, *J. Biomol. NMR* **2000**, *18*, 65.

[8] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661.

[9] a) K. J. McLean, P. Carroll, D. G. Lewis, A. J. Dunford, H. E. Seward, R. Neeli, M. R. Cheesman, L. Marsollier, P. Douglas, W. E. Smith, et al., *J. Biol. Chem.* **2008**, *283*, 33406; b) K. J. McLean, M. R. Cheesman, S. L. Rivers, A. Richmond, D. Leys, S. K. Chapman, G. A. Reid, N. C. Price, S. M. Kelly, J. Clarkson, et al., *J. Inorg. Biochem.* **2002**, *91*, 527.