

Discovery of Inhibitors of the Pentein Superfamily Protein Dimethylarginine Dimethylaminohydrolase (DDAH), by Virtual Screening and Hit Analysis.

Basil Hartzoulakis,^{#a} Sharon Rossiter,^{§a} Herpreet Gill,^b Bernard O'Hara,^c Emily Steinke,^a Paul J. Gane,^a Ramon Hurtado-Guerrero,^b James M Leiper,^b Patrick Vallance,^b Judith Murray Rust,^c David L. Selwood^{1a*}

Wolfson Institute for Biomedical Research,^a University College London, Gower Street, London WC1E 6BT, UK, Division of Medicine, BHF Laboratories,^b Rayne Building, University College London, 5 University Street, London WC1E 6JF, UK, School of Crystallography,^c Birkbeck, Malet Street, London WC1E 7HX, UK.

This is where the receipt/accepted dates will go; Received Month XX, 2000; Accepted Month XX, 2000 [BMCL RECEIPT]

Abstract— An efficient process for the discovery of inhibitors of DDAH enzymes, without the requirement for high throughput screening, is described. Physicochemical filtering of a 308,000 compound library according to drug likeness followed by reciprocal nearest neighbor selection produced a representative subset of 35,000 compounds. Virtual screening on a dual processor PC using FlexX, followed by biological screening identified two hit series. Similarity searches of commercial databases and chemical re-synthesis of pure compounds resulted in SR445 as an inhibitor of *Pseudomonas aeruginosa* DDAH at 2 μ M.

The pentein superfamily of proteins (Figure 1a) is characterized by a propeller like arrangement of five $\alpha\beta\beta\alpha\beta$ units forming a narrow channel with a central negatively charged core.^{1, 2} This elegant structure is ideal for recognition of the guanidinopropyl side chain of the amino acid arginine: its main biological function in the guanidine modifying enzyme (GME) branch of the superfamily. GMEs contain a conserved triad of His, Asp, and Cys amino acids which serve to catalyze a range of hydrolysis and transferase reactions on the arginine or, in the case of DDAH, the methyl arginine side chain (Figure 1b). The biological functions of this family are very diverse,² not only allowing metabolic reactions on simple arginine derivatives but also, in the case of the protein arginine deiminases (PADs), able to function as molecular switches for transcription by citrullination of arginine side chains on histones.³ Mammalian DDAH functions to control ADMA levels (a cardiovascular risk factor) and can modulate the nitric oxide pathway with its pivotal role in endothelial cell function.^{4, 5} Significantly DDAH may also participate in

protein regulation and DDAH 2 has been shown to increase transcription of the cancer promoting cytokine VEGF⁶ (Figure 1). *Pseudomonas aeruginosa* (Pa) is a pathogen linked to chronic lung infections in cystic fibrosis⁷ and PaDDAH is a possible factor involved in pathogenicity.⁸ The discovery of inhibitors of DDAH and the wider pentein family is therefore of considerable interest (Figure 2) and both competitive **1**, **2** and irreversible **3.4.5** inhibitors have been reported.^{3, 9-12}

We have previously described substrate based inhibitors⁹ of mammalian DDAH such as **1** (SR291), however we required access to inhibitors of better potency and different chemical structural types. The availability of the PaDDAH crystal structure, allowed us to consider virtual screening as an approach to the discovery of new inhibitors. Virtual screening has many potential advantages: lower cost, higher speed and lower resource requirements make it ideally suited to an academic research group. We therefore initiated an approach based on physicochemical filtering, virtual screening and hit analysis to identify inhibitors of PaDDAH utilizing the crystal structure.¹³

*Corresponding author: d.selwood@ucl.ac.uk [#]Current address, NCE discovery Ltd, 418 Science Park, Milton Road, Cambridge, CB4 0PZ, UK. [§]Current address, School of Life Sciences, University of Hertfordshire, College Lane, Hatfield, Herts AL10 9AB, UK

Starting with a database of 308,000 commercially available compounds,¹⁴ we initially employed a series of filters to remove about 48,000 non-druglike^{15,16} compounds and reactive molecules (Figure 3). This first step is divided into a structural filter^{17,18} (see supplementary data) which removes non-drug-like compounds and those with undesirable atoms/functional groups and a properties filter^{19, 20} that passes only lead-like compounds. All filtering was implemented with scripts written in the Sybyl programming language (SPL). After this first step the number of compounds in the databases was reduced to ~260k. A reciprocal nearest neighbor²¹ (RNN) packing algorithm was then utilized to generate a diverse, subset of compounds such that clusters of similar compounds in the original database were represented by small groups of compounds (not singletons) in the subset. In this way a cluster containing actives is better represented in the subset.²² The subset of 35,000 two dimensional structures was then converted into 3D using Concord²³ and minimized with the MMFF94 force field. In the final step the MMFF94 charges were replaced with formal charges for input to the docking program FlexX. The PaDDAH structure (pdb: 1h70) was manipulated by restoring the Cys249, adding hydrogens and Kollman atom charges then minimized using the Amber Force Field (version 7.0). Citrulline was left in place during the minimization to prevent collapse of the active site and was then removed. Virtual docking was conducted using FlexX.²⁴ FlexX cuts ligands into fragments at rotatable bonds and places a base fragment into the active site, the ligand is then reformed and the energy calculated using Bohm's function. The dimethylarginine binding pocket site was defined accurately using the Sybyl/Flexx interface (Leu18, Asp 60, Phe63, Asp66, Glu65, Arg85, Glu114, Arg132, Leu161, His162, Ile243, C249).

The top 2,000 successfully docked compounds were ranked with the CScore scoring function. The highest scoring 200 molecules were visually inspected to ensure that the docking algorithm had placed them in the binding pocket with at least one hydrogen bond between the protein and the ligand. Of these 109 structures, 90 were available for supply and these were screened, using a simple colorimetric assay (see supporting information). Three actives were identified, compounds **6**, **7** and **8** (Figure 4) representing two chemical scaffolds.

About 30% of the original screening set exhibited poor solubility precluding accurate IC₅₀ determination. The activity around the active scaffolds was probed by the purchase of 20 additional analogues identified using

similarity searches on the active hits based on Unity fingerprints and a Tanimoto score of >80%.

Re-synthesis to provide pure authentic samples was a key part of the hit analysis and the general synthesis²⁵ is shown for compounds **9** and **10** in scheme 1. If required the indole carbaldehydes were N-alkylated followed by condensation with the substituted barbituric acids²⁶ to provide the target molecules.

The indolythiobarbituric acid scaffold proved the most interesting; the compounds resulting from the similarity searches, **9** and **10** (SR445) demonstrated much improved activity in the re-synthesised samples with **10** having an IC₅₀ of 2 µM. This is the most potent PaDDAH inhibitor identified to date and is structurally distinct from the other inhibitors of both PaDDAH and mammalian DDAH (Figure 2). The predicted binding mode of **10** (Figure 5), shows insertion into the active site of DDAH with the indole moiety taking the place of the arginine side-chain.

This study highlights the advantages of virtual screening, allowing academic laboratories with limited manpower and resources to discover protein ligands for biological studies. Pre-filtering the database produced a representative set amenable to screening on a dual processor PC. We cannot yet consider the docking process to be very precise.²⁷ Rather docking enriches the final biological screening set to increase the chances of obtaining an active. The recent disclosures of the crystal structures of bovine and human DDAH^{5, 28} should allow similar methodology to be applied to the discovery of inhibitors of the mammalian and human isoforms.

Acknowledgements

This work was funded by the Wellcome Trust (Grant 065612), and UCL Biomedica.

Supplementary data

Details of computational methods, chemistry syntheses and biological screening.

References and Notes

Reference List

1. Teichmann, S. A.; Murzin, A. G.; Chothia, C. *Curr.Opin.Struct.Biol.* **2001**, *11*, 354-363.
2. Shirai, H.; Mokrab, Y.; Mizuguchi, K. *Proteins* **2006**, *64*, 1010-1023.
3. Luo, Y.; Knuckley, B.; Lee, Y. H.; Stallcup, M. R.; Thompson, P. R. *J.Am.Chem.Soc.* **2006**, *128*, 1092-1093.

4. Knipp, M. *Chembiochem* **2006**, *7*, 879-889.
5. Leiper, J.; Nandi, M.; Torondel, B.; Murray-Rust, J.; Malaki, M.; O'Hara, B.; Rossiter, S.; Anthony, S.; Madhani, M.; Selwood, D.; Smith, C.; Wojciak-Stothard, B.; Rudiger, A.; Stidwill, R.; McDonald, N. Q.; Vallance, P. *Nat.Med.* **2007**, *13*, 198-203.
6. Hasegawa, K.; Wakino, S.; Tanaka, T.; Kimoto, M.; Tatematsu, S.; Kanda, T.; Yoshioka, K.; Homma, K.; Sugano, N.; Kurabayashi, M.; Saruta, T.; Hayashi, K. *Arterioscler.Thromb.Vasc.Biol.* **2006**, *26*, 1488-1494.
7. Agarwal, G.; Kapil, A.; Kabra, S. K.; Das, B. K.; Dwivedi, S. N. *BMC.Microbiol.* **2005**, *5*, 43.
8. Stone, E. M.; Person, M. D.; Costello, N. J.; Fast, W. *Biochemistry* **2005**, *44*, 7069-7078.
9. Rossiter, S.; Smith, C. L.; Malaki, M.; Nandi, M.; Gill, H.; Leiper, J. M.; Vallance, P.; Selwood, D. L. *J.Med.Chem.* **2005**, *48*, 4670-4678.
10. Vallance, P.; Bush, H. D.; Mok, B. J.; Hurtado-Guerrero, R.; Gill, H.; Rossiter, S.; Wilden, J. D.; Caddick, S. *Chem.Commun.(Camb.)* **2005**, 5563-5565.
11. Stone, E. M.; Schaller, T. H.; Bianchi, H.; Person, M. D.; Fast, W. *Biochemistry* **2005**, *44*, 13744-13752.
12. Knipp, M.; Braun, O.; Vasak, M. *J.Am.Chem.Soc.* **2005**, *127*, 2372-2373.
13. Murray-Rust, J.; Leiper, J.; McAlister, M.; Phelan, J.; Tilley, S.; Santa, M. J.; Vallance, P.; McDonald, N. *Nat.Struct.Biol.* **2001**, *8*, 679-683.
14. The *Asinex database was chosen as the compounds were easily available and each was provided with analytical data.* **2007**.
15. Lipinski, C. A. *J.Pharmacol.Toxicol.Methods* **2000**, *44*, 235-249.
16. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J.Med.Chem.* **2002**, *45*, 2615-2623.
17. Muegge, I.; Heald, S. L.; Brittelli, D. *J.Med.Chem.* **2001**, *44*, 1841-1846.
18. Rishton, G. M. *Drug Discov.Today* **2003**, *8*, 86-96.
19. Clark, D. E.; Pickett, S. D. *Drug Discov.Today* **2000**, *5*, 49-58.
20. Walters, W. P.; Ajay; Murcko, M. A. *Curr.Opin.Chem.Biol.* **1999**, *3*, 384-387.
21. Ward, J. H. *J.Am.Stat.Soc.* **1963**, *58*, 236-244.
22. Leach, A. R.; Bradshaw, J.; Green, D. V.; Hann, M. M.; Delany, J. J., III. *J.Chem.Inf.Comput.Sci.* **1999**, *39*, 1161-1172.
23. Pearlman, R. S. "Concord" distributed by Tripos Inc., St Louis, Missouri, 63144, USA. 2006. Ref Type: Generic
24. Rarey, M.; Wefing, S.; Lengauer, T. *J.Comput.Aided Mol.Des* **1996**, *10*, 41-54.
25. Nightingale, D.; Alexander, C. H. *J.Am.Chem.Soc.* **1936**, *58*, 794-796.
26. Dox, A. W.; Plaisance, G. P. *J.Am.Chem.Soc.* **1916**, *38*, 2164-2166.
27. Marsden, P. M.; Puvanendrapillai, D.; Mitchell, J. B.; Glen, R. C. *Org.Biomol.Chem.* **2004**, *2*, 3267-3273.
28. Frey, D.; Braun, O.; Briand, C.; Vasak, M.; Grutter, M. G. *Structure.* **2006**, *14*, 901-911.

Figure and Scheme Legends

Figure 1. (a) Ribbon diagrams of pentein family proteins PaDDAH and PAD4. (b) DDAH hydrolyses methylated arginines including ADMA a cardiovascular risk factor. DDAH2 can increase the expression of VEGF through phosphorylation of its transcription factor SP1.

Figure 2. Inhibitors of DDAH and PAD4.

Figure 3. Filtering, virtual screening, biological testing and hit analysis lead to the discovery of micromolar inhibitors of DDAH.

Figure 4. Structures and activity (500 μ M) of screening hits (blue) and similarity search hits (black). IC₅₀s are of re-synthesized pure samples. *Insoluble in buffer.

Figure 5. Top and side views of predicted binding mode of **10**.

Scheme 1. General synthesis of the indolyl barbiturates.

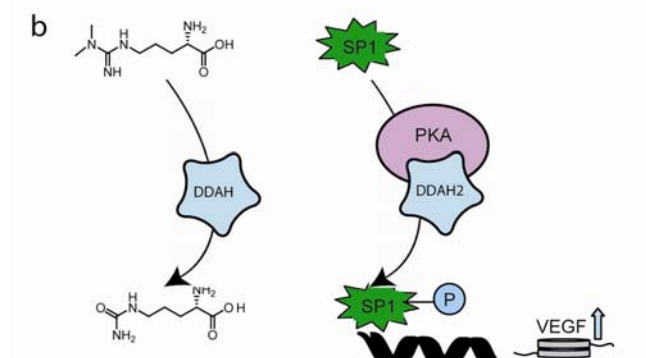
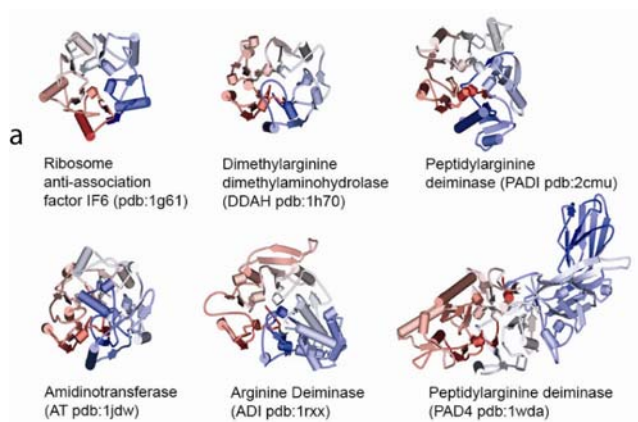


Figure 1. (a) Ribbon diagrams of peptin family proteins PaDDAH and PAD4. (b) DDAH hydrolyses methylated arginines including ADMA a cardiovascular risk factor. DDAH2 can increase the expression of VEGF through phosphorylation of its transcription factor SP1.

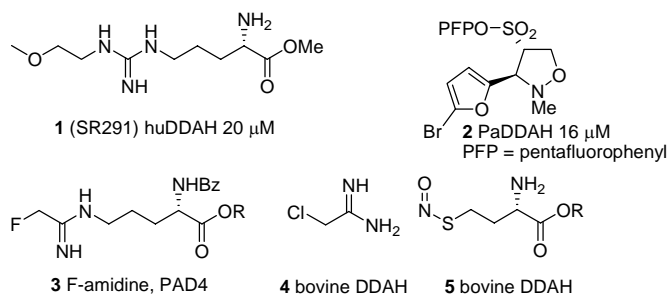


Figure 2. Inhibitors of DDAH and PAD4

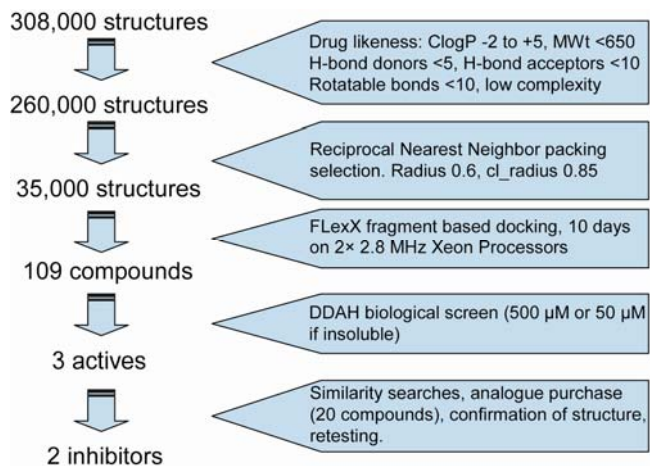


Figure 3. Filtering, virtual screening, biological testing and hit analysis lead to the discovery of micromolar inhibitors of DDAH.

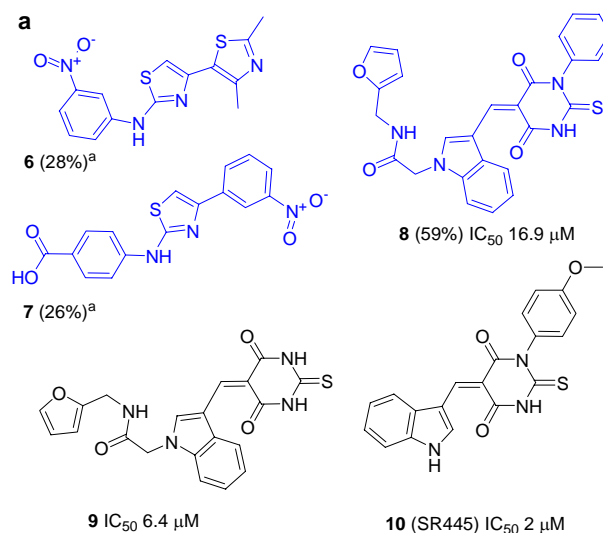
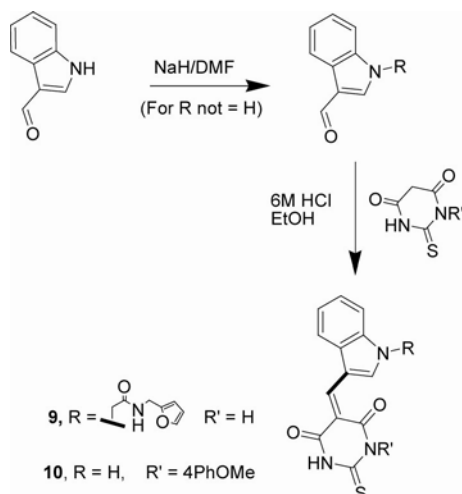


Figure 4. Structures and activity (500 μ M) of screening hits (blue) and similarity search hits (black). IC₅₀s are of re-synthesized pure samples. ^aInsoluble in buffer.



Scheme 1. General synthesis of the indolyl barbiturates.

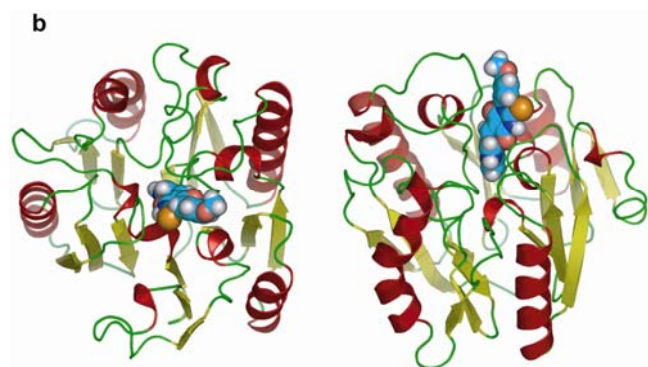


Figure 5. Top and side views of predicted binding mode of **10**.