

# ChEMBL

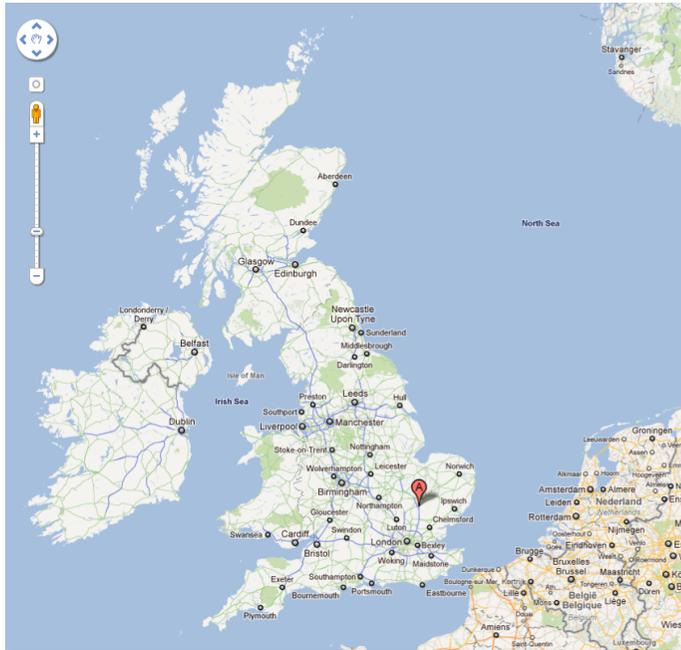
## 5<sup>th</sup> Meeting on US Government Chemical Databases and Open Chemistry

26<sup>th</sup> August 2011, Frederick, MD

John P. Overington – Computational Chemical Biology  
[jpo@ebi.ac.uk](mailto:jpo@ebi.ac.uk)



# The EMBL-EBI



European Bioinformatics Institute | Homepage | EBI

http://www.ebi.ac.uk/

EMBL-EBI European Bioinformatics Institute

Databases Tools Research Training Industry About Us Help Site Index

Explore the EBI:

Examples: [ROA1\\_HUMAN](#), [tpi1](#), [Sulston...](#)

[Help](#) | [Feedback](#)

### Data Resources and Tools

- ENNA
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDBe
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules
- Gene Expression
- Protein Expression
- Molecular Interactions
- Reactions & Pathways
- Protein Families
- Enzymes
- Literature
- Taxonomy
- Ontologies
- Patent Resources
- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads
- Web Services

### Latest News

**European Bioinformatics Institute to lead UK PubMed Central in 2011-2016**

Posted: **Jul 28, 2011**

The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) has been awarded the contract to run and lead the development of [UK PubMed Central](#) (UKPMC), the free online literature resource for life science researchers.

View press release [HTML](#) | [PDF](#).

### Research Highlights

**Michael Ashburner wins prestigious computational biology award**

Posted: **Jul 21, 2011**

The International Society for Computational Biology (ISCB) has presented Michael Ashburner with this year's Accomplishment by a Senior Scientist Award, acknowledging his outstanding contributions to the field of computational biology. "His work is now seen as a landmark and an achievement in technology," says Alfonso Valencia, chair of the ISCB awards committee. Michael was joint-director of EMBL-EBI with Graham Cameron.

[Read more about this research highlight](#) | [Watch Michael Ashburner's ISMB / ECCB 2011 keynote on Vimeo](#)

### Events & Training

[Next Generation Sequencing Workshop](#)  
**18-20 October 2011**  
Registration closes 30 September 2011

[EMBL-EBI Events Year Planner 2012](#)  
Please [register to receive your free copy of the EMBL-EBI Events Year Planner](#)

[Conference: Functional Genomics and Systems Biology 2011](#)  
**29 November - 1 December 2011**  
Abstract deadline: 3 October 2011  
Registration closes 17 October 2011

[EBI Open Day](#)  
**1 November, 2011**  
Registration closes 1 October 2011

[EMBO Practical Course: Computational structural biology - from data to structure to function](#)  
**14-18 November 2011**  
Registration closes 30 September 2011

# 'The Special Relationship'



# What Is the ChEMBL Data?

2432

*J. Med. Chem.* 2002, 45, 2432–2453

## Design of Selective Thrombin Inhibitors Based on the (R)-Phe-Pro-Arg Sequence

John C. Danilewicz,<sup>1,2</sup> Stuart M. Abel,<sup>3</sup> Alan D. Brown,<sup>2</sup> Paul V. Fish,<sup>4\*</sup> Edward Hawkeswood,<sup>3</sup> Stephen J. Holland,<sup>1</sup> Keith James,<sup>1</sup> Andrew B. McElroy,<sup>1</sup> John Overington,<sup>1</sup> Michael J. Powling,<sup>1</sup> and David J. Rance<sup>5</sup>

Departments of Discovery Chemistry, Drug Metabolism, Discovery Biology, and Molecular Informatics Structure and Design, Pfizer Global Research and Development, Sandwich, Kent CT13 9NJ, United Kingdom

Received December 21, 2001

Potent and selective inhibitors of thrombin were sought based on the (R)-Phe-Pro-Arg sequence. The objective was to generate similar binding interactions to those achieved by potent competitive inhibitors of the argatroban type, so eliminating the need for covalent interaction with the catalytic serine function, as utilized by aldehyde and boronic acid type inhibitors. Improving the S<sub>1</sub> subsite interaction by substitution of arginine with a 4-alkoxybenzamide residue provided potent lead 2 ( $K_i = 0.37$  nM). Though an amide bond, which H-bonds to the active site, is lost, modeling indicated that a new H-bond is generated between the alkoxy oxygen atom and the catalytic Ser-195 hydroxyl group. Substitution of the benzamide system by 1-amidinopiperidine then gave compound 4, which provided a further gain in selectivity over trypsin. However, previous work had shown that these compounds were likely to be too lipophilic (Log *D* +0.4 and +0.2, respectively) and to suffer rapid hepatic extraction, presumably via biliary elimination. Accordingly, both proved short-acting when administered intravenously to rats and showed poor activity when given intraduodenally. The aim was then to reduce lipophilicity below a log *D* of -1.2, which in a previously reported series had been effective in preventing rapid clearance. It was anticipated that compounds of this type would rely on the cation selective paracellular route of absorption from the gastrointestinal tract. Potent polar analogues with selectivity >1000 over trypsin were obtained. The best in vivo activity was shown by compound 12. However, in the final analysis, its oral bioavailability proved poor, relative to analogues with similar physicochemical properties derived from argatroban, consistent with the hypothesis that molecular shape is an additional important determinant of paracellular absorption.

### Introduction

The search for potent selective and orally active thrombin inhibitors has gathered momentum in recent years.<sup>1</sup> Thrombin is the last in a cascade of trypsin-like plasma serine proteases, which by catalyzing the conversion of fibrinogen to fibrin, activation of FXIII and inducing platelet aggregation is a key enzyme in haemostasis and thrombus formation. The inhibition of a single enzyme in the cascade, and in particular thrombin, has been an attractive goal in that it could also provide superior antithrombotic therapy by increasing efficacy and safety as compared to heparin and the coumarins. Additionally, by keeping molecular size small, the opportunity exists for obtaining oral activity.<sup>1,2</sup>

Two small molecular weight inhibitor types are emerging as structure–activity relationships are explored. The first is of the argatroban<sup>3</sup> and NAPAP<sup>4</sup> type (Chart 1), where lipophilic groups on either side of the

basic P<sub>1</sub> side chain pack together to interact with the hydrophobic S<sub>2</sub> site.<sup>5–7</sup> Napsagatran (Ro 46-6240), developed by Hilpert et al.,<sup>8</sup> though having a more complex P<sub>1</sub> residue, can nevertheless be viewed as belonging to this group. The only interaction with the catalytic serine residue is via a hydrogen bond to the carboxylate function in both argatroban and napsagatran. Unfortunately, none of these compounds is orally active due to either poor absorption from the gastrointestinal tract and/or rapid clearance via the bile.<sup>4,9</sup>

A second inhibitor type is based on the substrate-derived irreversible chloromethyl ketone inhibitor PPACK and includes compounds such as DuP-714<sup>10,11</sup> and efegatran (GYKI-14 766).<sup>12</sup> These compounds interact covalently with the hydroxyl group of the catalytic serine residue. The neighboring proline ring and (R)-Phe side chain cooperate to fill the S<sub>2</sub> site in a similar fashion as the two distal lipophilic groups of the first series.<sup>5</sup> Though oral activity has been claimed for these compounds, we were concerned that high enzyme selectivity might not be obtainable when substantial affinity is derived by interacting covalently with the ubiquitous active site serine function. In the case of aldehyde type inhibitors, there is also the potential problem of achieving adequate optical and chemical stability.

\* To whom correspondence should be addressed. Tel: 44 1304 644589. Fax: 44 1304 656433. E-mail: paul.fish@sandwich.pfizer.com.

<sup>1</sup> Senior author.

<sup>2</sup> Department of Discovery Chemistry.

<sup>3</sup> Department of Drug Metabolism.

<sup>4</sup> Department of Discovery Biology.

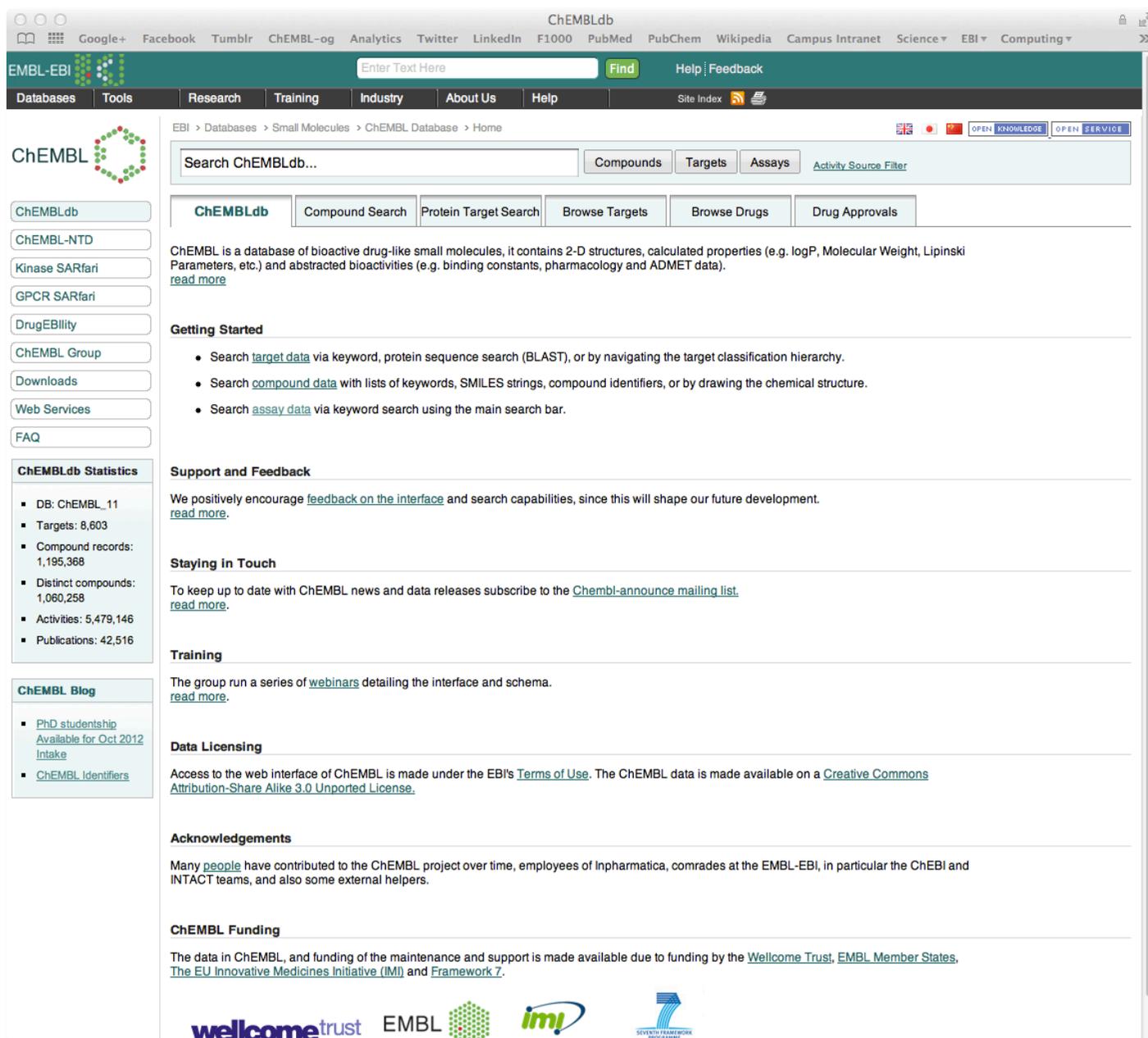
<sup>5</sup> Molecular Informatics Structure and Design.

# What Is the ChEMBL Data?

## Design of Novel Thrombin Inhibitors

510, 7.39 (d, 1H), 7.36–7.40 (m, 4H), 8.82 (s, 2H), 9.04 (s, 2H). Anal. (C<sub>24</sub>H<sub>26</sub>N<sub>2</sub>O<sub>5</sub>) Calcd: C, 67.11%; H, 5.71%; N, 6.18%.  
 3-[2-[[4-(benzylamino)-5-carboxy-1H-pyridin-2-yl-aminio]propionic acid Ethyl Ester] (28b). Starting from 2-amino-3-amino-benzoic acid and 3-pyridin-2-ylamino-propionic acid ethyl ester, the title compound was prepared via the procedure described for the synthesis of intermediate 28a. mp 86–87 °C. <sup>1</sup>H NMR (DMF-d<sub>2</sub>, δ): 1.11 (t, 3H), 1.55 (d, 2H), 2.30 (q, 3H), 3.07 (q, 2H), 4.18 (t, 2H), 6.82 (d, 1H), 7.08 (d, 1H), 7.21 (m, 1H), 7.31 (d, 1H), 7.30 (d, 1H), 7.52 (d, 1H), 8.37 (q, 1H), 8.44 (dd, 1H). Anal. (C<sub>24</sub>H<sub>26</sub>N<sub>2</sub>O<sub>5</sub>) Calcd: C, 67.11%; H, 5.71%; N, 6.18%.  
 3-[2-[[4-(Carbamimididyl-phenylamino)methyl]-1-methyl-5-*H*-benzimidazole-5-carboxyl]pyridin-2-yl-aminio]propionic Acid Ethyl Ester. Hydrochloride (89). Compound 28d (16.7 g, 44.1 mmol) was hydrogenated (Pd on charcoal, 100% in 200 mL of methanol at room temperature to afford 3-[2-[[4-(1-methyl-5-*H*-benzimidazole-5-carboxyl)pyridin-2-yl-aminio]propionic acid ethyl ester] (10.0 g, 65%) after purification by chromatography (eluent: dichloromethane/methanol 9:1). The intermediate (2.1 g, 6.1 mmol) was added to a solution of 4-cyano-phenylboronic acid and indazole (7.3 mmol) in 50 mL of THF and refluxed for 24 h. The reaction mixture was concentrated *in vacuo*, and the residue was dissolved in 50 mL of glacial acetic acid and refluxed for 1 h. The solution was diluted with 100 mL of water and neutralized with concentrated aqueous sodium hydroxide. Extracted with ethyl acetate and purification via chromatography (silica gel, dichloromethane/ethyl acetate 7:3) (14-cyano-phenylboronic acid ethyl ester) (1.8 g, 43%) was converted into the title compound following the protocol of the Pfizer reaction described above. mp 174 (s, 1H). LMS: (BR): (M + H)<sup>+</sup> = 530, (M + H)<sup>+</sup> = 549, (M + H)<sup>+</sup> = 568, (M + H)<sup>+</sup> = 587, (M + H)<sup>+</sup> = 606, (M + H)<sup>+</sup> = 625, (M + H)<sup>+</sup> = 644, (M + H)<sup>+</sup> = 663, (M + H)<sup>+</sup> = 682, (M + H)<sup>+</sup> = 701, (M + H)<sup>+</sup> = 720, (M + H)<sup>+</sup> = 739, (M + H)<sup>+</sup> = 758, (M + H)<sup>+</sup> = 777, (M + H)<sup>+</sup> = 796, (M + H)<sup>+</sup> = 815, (M + H)<sup>+</sup> = 834, (M + H)<sup>+</sup> = 853, (M + H)<sup>+</sup> = 872, (M + H)<sup>+</sup> = 891, (M + H)<sup>+</sup> = 910, (M + H)<sup>+</sup> = 929, (M + H)<sup>+</sup> = 948, (M + H)<sup>+</sup> = 967, (M + H)<sup>+</sup> = 986, (M + H)<sup>+</sup> = 1005, (M + H)<sup>+</sup> = 1024, (M + H)<sup>+</sup> = 1043, (M + H)<sup>+</sup> = 1062, (M + H)<sup>+</sup> = 1081, (M + H)<sup>+</sup> = 1100, (M + H)<sup>+</sup> = 1119, (M + H)<sup>+</sup> = 1138, (M + H)<sup>+</sup> = 1157, (M + H)<sup>+</sup> = 1176, (M + H)<sup>+</sup> = 1195, (M + H)<sup>+</sup> = 1214, (M + H)<sup>+</sup> = 1233, (M + H)<sup>+</sup> = 1252, (M + H)<sup>+</sup> = 1271, (M + H)<sup>+</sup> = 1290, (M + H)<sup>+</sup> = 1309, (M + H)<sup>+</sup> = 1328, (M + H)<sup>+</sup> = 1347, (M + H)<sup>+</sup> = 1366, (M + H)<sup>+</sup> = 1385, (M + H)<sup>+</sup> = 1404, (M + H)<sup>+</sup> = 1423, (M + H)<sup>+</sup> = 1442, (M + H)<sup>+</sup> = 1461, (M + H)<sup>+</sup> = 1480, (M + H)<sup>+</sup> = 1499, (M + H)<sup>+</sup> = 1518, (M + H)<sup>+</sup> = 1537, (M + H)<sup>+</sup> = 1556, (M + H)<sup>+</sup> = 1575, (M + H)<sup>+</sup> = 1594, (M + H)<sup>+</sup> = 1613, (M + H)<sup>+</sup> = 1632, (M + H)<sup>+</sup> = 1651, (M + H)<sup>+</sup> = 1670, (M + H)<sup>+</sup> = 1689, (M + H)<sup>+</sup> = 1708, (M + H)<sup>+</sup> = 1727, (M + H)<sup>+</sup> = 1746, (M + H)<sup>+</sup> = 1765, (M + H)<sup>+</sup> = 1784, (M + H)<sup>+</sup> = 1803, (M + H)<sup>+</sup> = 1822, (M + H)<sup>+</sup> = 1841, (M + H)<sup>+</sup> = 1860, (M + H)<sup>+</sup> = 1879, (M + H)<sup>+</sup> = 1898, (M + H)<sup>+</sup> = 1917, (M + H)<sup>+</sup> = 1936, (M + H)<sup>+</sup> = 1955, (M + H)<sup>+</sup> = 1974, (M + H)<sup>+</sup> = 1993, (M + H)<sup>+</sup> = 2012, (M + H)<sup>+</sup> = 2031, (M + H)<sup>+</sup> = 2050, (M + H)<sup>+</sup> = 2069, (M + H)<sup>+</sup> = 2088, (M + H)<sup>+</sup> = 2107, (M + H)<sup>+</sup> = 2126, (M + H)<sup>+</sup> = 2145, (M + H)<sup>+</sup> = 2164, (M + H)<sup>+</sup> = 2183, (M + H)<sup>+</sup> = 2202, (M + H)<sup>+</sup> = 2221, (M + H)<sup>+</sup> = 2240, (M + H)<sup>+</sup> = 2259, (M + H)<sup>+</sup> = 2278, (M + H)<sup>+</sup> = 2297, (M + H)<sup>+</sup> = 2316, (M + H)<sup>+</sup> = 2335, (M + H)<sup>+</sup> = 2354, (M + H)<sup>+</sup> = 2373, (M + H)<sup>+</sup> = 2392, (M + H)<sup>+</sup> = 2411, (M + H)<sup>+</sup> = 2430, (M + H)<sup>+</sup> = 2449, (M + H)<sup>+</sup> = 2468, (M + H)<sup>+</sup> = 2487, (M + H)<sup>+</sup> = 2506, (M + H)<sup>+</sup> = 2525, (M + H)<sup>+</sup> = 2544, (M + H)<sup>+</sup> = 2563, (M + H)<sup>+</sup> = 2582, (M + H)<sup>+</sup> = 2601, (M + H)<sup>+</sup> = 2620, (M + H)<sup>+</sup> = 2639, (M + H)<sup>+</sup> = 2658, (M + H)<sup>+</sup> = 2677, (M + H)<sup>+</sup> = 2696, (M + H)<sup>+</sup> = 2715, (M + H)<sup>+</sup> = 2734, (M + H)<sup>+</sup> = 2753, (M + H)<sup>+</sup> = 2772, (M + H)<sup>+</sup> = 2791, (M + H)<sup>+</sup> = 2810, (M + H)<sup>+</sup> = 2829, (M + H)<sup>+</sup> = 2848, (M + H)<sup>+</sup> = 2867, (M + H)<sup>+</sup> = 2886, (M + H)<sup>+</sup> = 2905, (M + H)<sup>+</sup> = 2924, (M + H)<sup>+</sup> = 2943, (M + H)<sup>+</sup> = 2962, (M + H)<sup>+</sup> = 2981, (M + H)<sup>+</sup> = 3000, (M + H)<sup>+</sup> = 3019, (M + H)<sup>+</sup> = 3038, (M + H)<sup>+</sup> = 3057, (M + H)<sup>+</sup> = 3076, (M + H)<sup>+</sup> = 3095, (M + H)<sup>+</sup> = 3114, (M + H)<sup>+</sup> = 3133, (M + H)<sup>+</sup> = 3152, (M + H)<sup>+</sup> = 3171, (M + H)<sup>+</sup> = 3190, (M + H)<sup>+</sup> = 3209, (M + H)<sup>+</sup> = 3228, (M + H)<sup>+</sup> = 3247, (M + H)<sup>+</sup> = 3266, (M + H)<sup>+</sup> = 3285, (M + H)<sup>+</sup> = 3304, (M + H)<sup>+</sup> = 3323, (M + H)<sup>+</sup> = 3342, (M + H)<sup>+</sup> = 3361, (M + H)<sup>+</sup> = 3380, (M + H)<sup>+</sup> = 3399, (M + H)<sup>+</sup> = 3418, (M + H)<sup>+</sup> = 3437, (M + H)<sup>+</sup> = 3456, (M + H)<sup>+</sup> = 3475, (M + H)<sup>+</sup> = 3494, (M + H)<sup>+</sup> = 3513, (M + H)<sup>+</sup> = 3532, (M + H)<sup>+</sup> = 3551, (M + H)<sup>+</sup> = 3570, (M + H)<sup>+</sup> = 3589, (M + H)<sup>+</sup> = 3608, (M + H)<sup>+</sup> = 3627, (M + H)<sup>+</sup> = 3646, (M + H)<sup>+</sup> = 3665, (M + H)<sup>+</sup> = 3684, (M + H)<sup>+</sup> = 3703, (M + H)<sup>+</sup> = 3722, (M + H)<sup>+</sup> = 3741, (M + H)<sup>+</sup> = 3760, (M + H)<sup>+</sup> = 3779, (M + H)<sup>+</sup> = 3798, (M + H)<sup>+</sup> = 3817, (M + H)<sup>+</sup> = 3836, (M + H)<sup>+</sup> = 3855, (M + H)<sup>+</sup> = 3874, (M + H)<sup>+</sup> = 3893, (M + H)<sup>+</sup> = 3912, (M + H)<sup>+</sup> = 3931, (M + H)<sup>+</sup> = 3950, (M + H)<sup>+</sup> = 3969, (M + H)<sup>+</sup> = 3988, (M + H)<sup>+</sup> = 4007, (M + H)<sup>+</sup> = 4026, (M + H)<sup>+</sup> = 4045, (M + H)<sup>+</sup> = 4064, (M + H)<sup>+</sup> = 4083, (M + H)<sup>+</sup> = 4102, (M + H)<sup>+</sup> = 4121, (M + H)<sup>+</sup> = 4140, (M + H)<sup>+</sup> = 4159, (M + H)<sup>+</sup> = 4178, (M + H)<sup>+</sup> = 4197, (M + H)<sup>+</sup> = 4216, (M + H)<sup>+</sup> = 4235, (M + H)<sup>+</sup> = 4254, (M + H)<sup>+</sup> = 4273, (M + H)<sup>+</sup> = 4292, (M + H)<sup>+</sup> = 4311, (M + H)<sup>+</sup> = 4330, (M + H)<sup>+</sup> = 4349, (M + H)<sup>+</sup> = 4368, (M + H)<sup>+</sup> = 4387, (M + H)<sup>+</sup> = 4406, (M + H)<sup>+</sup> = 4425, (M + H)<sup>+</sup> = 4444, (M + H)<sup>+</sup> = 4463, (M + H)<sup>+</sup> = 4482, (M + H)<sup>+</sup> = 4501, (M + H)<sup>+</sup> = 4520, (M + H)<sup>+</sup> = 4539, (M + H)<sup>+</sup> = 4558, (M + H)<sup>+</sup> = 4577, (M + H)<sup>+</sup> = 4596, (M + H)<sup>+</sup> = 4615, (M + H)<sup>+</sup> = 4634, (M + H)<sup>+</sup> = 4653, (M + H)<sup>+</sup> = 4672, (M + H)<sup>+</sup> = 4691, (M + H)<sup>+</sup> = 4710, (M + H)<sup>+</sup> = 4729, (M + H)<sup>+</sup> = 4748, (M + H)<sup>+</sup> = 4767, (M + H)<sup>+</sup> = 4786, (M + H)<sup>+</sup> = 4805, (M + H)<sup>+</sup> = 4824, (M + H)<sup>+</sup> = 4843, (M + H)<sup>+</sup> = 4862, (M + H)<sup>+</sup> = 4881, (M + H)<sup>+</sup> = 4900, (M + H)<sup>+</sup> = 4919, (M + H)<sup>+</sup> = 4938, (M + H)<sup>+</sup> = 4957, (M + H)<sup>+</sup> = 4976, (M + H)<sup>+</sup> = 4995, (M + H)<sup>+</sup> = 5014, (M + H)<sup>+</sup> = 5033, (M + H)<sup>+</sup> = 5052, (M + H)<sup>+</sup> = 5071, (M + H)<sup>+</sup> = 5090, (M + H)<sup>+</sup> = 5109, (M + H)<sup>+</sup> = 5128, (M + H)<sup>+</sup> = 5147, (M + H)<sup>+</sup> = 5166, (M + H)<sup>+</sup> = 5185, (M + H)<sup>+</sup> = 5204, (M + H)<sup>+</sup> = 5223, (M + H)<sup>+</sup> = 5242, (M + H)<sup>+</sup> = 5261, (M + H)<sup>+</sup> = 5280, (M + H)<sup>+</sup> = 5299, (M + H)<sup>+</sup> = 5318, (M + H)<sup>+</sup> = 5337, (M + H)<sup>+</sup> = 5356, (M + H)<sup>+</sup> = 5375, (M + H)<sup>+</sup> = 5394, (M + H)<sup>+</sup> = 5413, (M + H)<sup>+</sup> = 5432, (M + H)<sup>+</sup> = 5451, (M + H)<sup>+</sup> = 5470, (M + H)<sup>+</sup> = 5489, (M + H)<sup>+</sup> = 5508, (M + H)<sup>+</sup> = 5527, (M + H)<sup>+</sup> = 5546, (M + H)<sup>+</sup> = 5565, (M + H)<sup>+</sup> = 5584, (M + H)<sup>+</sup> = 5603, (M + H)<sup>+</sup> = 5622, (M + H)<sup>+</sup> = 5641, (M + H)<sup>+</sup> = 5660, (M + H)<sup>+</sup> = 5679, (M + H)<sup>+</sup> = 5698, (M + H)<sup>+</sup> = 5717, (M + H)<sup>+</sup> = 5736, (M + H)<sup>+</sup> = 5755, (M + H)<sup>+</sup> = 5774, (M + H)<sup>+</sup> = 5793, (M + H)<sup>+</sup> = 5812, (M + H)<sup>+</sup> = 5831, (M + H)<sup>+</sup> = 5850, (M + H)<sup>+</sup> = 5869, (M + H)<sup>+</sup> = 5888, (M + H)<sup>+</sup> = 5907, (M + H)<sup>+</sup> = 5926, (M + H)<sup>+</sup> = 5945, (M + H)<sup>+</sup> = 5964, (M + H)<sup>+</sup> = 5983, (M + H)<sup>+</sup> = 6002, (M + H)<sup>+</sup> = 6021, (M + H)<sup>+</sup> = 6040, (M + H)<sup>+</sup> = 6059, (M + H)<sup>+</sup> = 6078, (M + H)<sup>+</sup> = 6097, (M + H)<sup>+</sup> = 6116, (M + H)<sup>+</sup> = 6135, (M + H)<sup>+</sup> = 6154, (M + H)<sup>+</sup> = 6173, (M + H)<sup>+</sup> = 6192, (M + H)<sup>+</sup> = 6211, (M + H)<sup>+</sup> = 6230, (M + H)<sup>+</sup> = 6249, (M + H)<sup>+</sup> = 6268, (M + H)<sup>+</sup> = 6287, (M + H)<sup>+</sup> = 6306, (M + H)<sup>+</sup> = 6325, (M + H)<sup>+</sup> = 6344, (M + H)<sup>+</sup> = 6363, (M + H)<sup>+</sup> = 6382, (M + H)<sup>+</sup> = 6401, (M + H)<sup>+</sup> = 6420, (M + H)<sup>+</sup> = 6439, (M + H)<sup>+</sup> = 6458, (M + H)<sup>+</sup> = 6477, (M + H)<sup>+</sup> = 6496, (M + H)<sup>+</sup> = 6515, (M + H)<sup>+</sup> = 6534, (M + H)<sup>+</sup> = 6553, (M + H)<sup>+</sup> = 6572, (M + H)<sup>+</sup> = 6591, (M + H)<sup>+</sup> = 6610, (M + H)<sup>+</sup> = 6629, (M + H)<sup>+</sup> = 6648, (M + H)<sup>+</sup> = 6667, (M + H)<sup>+</sup> = 6686, (M + H)<sup>+</sup> = 6705, (M + H)<sup>+</sup> = 6724, (M + H)<sup>+</sup> = 6743, (M + H)<sup>+</sup> = 6762, (M + H)<sup>+</sup> = 6781, (M + H)<sup>+</sup> = 6800, (M + H)<sup>+</sup> = 6819, (M + H)<sup>+</sup> = 6838, (M + H)<sup>+</sup> = 6857, (M + H)<sup>+</sup> = 6876, (M + H)<sup>+</sup> = 6895, (M + H)<sup>+</sup> = 6914, (M + H)<sup>+</sup> = 6933, (M + H)<sup>+</sup> = 6952, (M + H)<sup>+</sup> = 6971, (M + H)<sup>+</sup> = 6990, (M + H)<sup>+</sup> = 7009, (M + H)<sup>+</sup> = 7028, (M + H)<sup>+</sup> = 7047, (M + H)<sup>+</sup> = 7066, (M + H)<sup>+</sup> = 7085, (M + H)<sup>+</sup> = 7104, (M + H)<sup>+</sup> = 7123, (M + H)<sup>+</sup> = 7142, (M + H)<sup>+</sup> = 7161, (M + H)<sup>+</sup> = 7180, (M + H)<sup>+</sup> = 7199, (M + H)<sup>+</sup> = 7218, (M + H)<sup>+</sup> = 7237, (M + H)<sup>+</sup> = 7256, (M + H)<sup>+</sup> = 7275, (M + H)<sup>+</sup> = 7294, (M + H)<sup>+</sup> = 7313, (M + H)<sup>+</sup> = 7332, (M + H)<sup>+</sup> = 7351, (M + H)<sup>+</sup> = 7370, (M + H)<sup>+</sup> = 7389, (M + H)<sup>+</sup> = 7408, (M + H)<sup>+</sup> = 7427, (M + H)<sup>+</sup> = 7446, (M + H)<sup>+</sup> = 7465, (M + H)<sup>+</sup> = 7484, (M + H)<sup>+</sup> = 7503, (M + H)<sup>+</sup> = 7522, (M + H)<sup>+</sup> = 7541, (M + H)<sup>+</sup> = 7560, (M + H)<sup>+</sup> = 7579, (M + H)<sup>+</sup> = 7598, (M + H)<sup>+</sup> = 7617, (M + H)<sup>+</sup> = 7636, (M + H)<sup>+</sup> = 7655, (M + H)<sup>+</sup> = 7674, (M + H)<sup>+</sup> = 7693, (M + H)<sup>+</sup> = 7712, (M + H)<sup>+</sup> = 7731, (M + H)<sup>+</sup> = 7750, (M + H)<sup>+</sup> = 7769, (M + H)<sup>+</sup> = 7788, (M + H)<sup>+</sup> = 7807, (M + H)<sup>+</sup> = 7826, (M + H)<sup>+</sup> = 7845, (M + H)<sup>+</sup> = 7864, (M + H)<sup>+</sup> = 7883, (M + H)<sup>+</sup> = 7902, (M + H)<sup>+</sup> = 7921, (M + H)<sup>+</sup> = 7940, (M + H)<sup>+</sup> = 7959, (M + H)<sup>+</sup> = 7978, (M + H)<sup>+</sup> = 7997, (M + H)<sup>+</sup> = 8016, (M + H)<sup>+</sup> = 8035, (M + H)<sup>+</sup> = 8054, (M + H)<sup>+</sup> = 8073, (M + H)<sup>+</sup> = 8092, (M + H)<sup>+</sup> = 8111, (M + H)<sup>+</sup> = 8130, (M + H)<sup>+</sup> = 8149, (M + H)<sup>+</sup> = 8168, (M + H)<sup>+</sup> = 8187, (M + H)<sup>+</sup> = 8206, (M + H)<sup>+</sup> = 8225, (M + H)<sup>+</sup> = 8244, (M + H)<sup>+</sup> = 8263, (M + H)<sup>+</sup> = 8282, (M + H)<sup>+</sup> = 8301, (M + H)<sup>+</sup> = 8320, (M + H)<sup>+</sup> = 8339, (M + H)<sup>+</sup> = 8358, (M + H)<sup>+</sup> = 8377, (M + H)<sup>+</sup> = 8396, (M + H)<sup>+</sup> = 8415, (M + H)<sup>+</sup> = 8434, (M + H)<sup>+</sup> = 8453, (M + H)<sup>+</sup> = 8472, (M + H)<sup>+</sup> = 8491, (M + H)<sup>+</sup> = 8510, (M + H)<sup>+</sup> = 8529, (M + H)<sup>+</sup> = 8548, (M + H)<sup>+</sup> = 8567, (M + H)<sup>+</sup> = 8586, (M + H)<sup>+</sup> = 8605, (M + H)<sup>+</sup> = 8624, (M + H)<sup>+</sup> = 8643, (M + H)<sup>+</sup> = 8662, (M + H)<sup>+</sup> = 8681, (M + H)<sup>+</sup> = 8700, (M + H)<sup>+</sup> = 8719, (M + H)<sup>+</sup> = 8738, (M + H)<sup>+</sup> = 8757, (M + H)<sup>+</sup> = 8776, (M + H)<sup>+</sup> = 8795, (M + H)<sup>+</sup> = 8814, (M + H)<sup>+</sup> = 8833, (M + H)<sup>+</sup> = 8852, (M + H)<sup>+</sup> = 8871, (M + H)<sup>+</sup> = 8890, (M + H)<sup>+</sup> = 8909, (M + H)<sup>+</sup> = 8928, (M + H)<sup>+</sup> = 8947, (M + H)<sup>+</sup> = 8966, (M + H)<sup>+</sup> = 8985, (M + H)<sup>+</sup> = 9004, (M + H)<sup>+</sup> = 9023, (M + H)<sup>+</sup> = 9042, (M + H)<sup>+</sup> = 9061, (M + H)<sup>+</sup> = 9080, (M + H)<sup>+</sup> = 9099, (M + H)<sup>+</sup> = 9118, (M + H)<sup>+</sup> = 9137, (M + H)<sup>+</sup> = 9156, (M + H)<sup>+</sup> = 9175, (M + H)<sup>+</sup> = 9194, (M + H)<sup>+</sup> = 9213, (M + H)<sup>+</sup> = 9232, (M + H)<sup>+</sup> = 9251, (M + H)<sup>+</sup> = 9270, (M + H)<sup>+</sup> = 9289, (M + H)<sup>+</sup> = 9308, (M + H)<sup>+</sup> = 9327, (M + H)<sup>+</sup> = 9346, (M + H)<sup>+</sup> = 9365, (M + H)<sup>+</sup> = 9384, (M + H)<sup>+</sup> = 9403, (M + H)<sup>+</sup> = 9422, (M + H)<sup>+</sup> = 9441, (M + H)<sup>+</sup> = 9460, (M + H)<sup>+</sup> = 9479, (M + H)<sup>+</sup> = 9498, (M + H)<sup>+</sup> = 9517, (M + H)<sup>+</sup> = 9536, (M + H)<sup>+</sup> = 9555, (M + H)<sup>+</sup> = 9574, (M + H)<sup>+</sup> = 9593, (M + H)<sup>+</sup> = 9612, (M + H)<sup>+</sup> = 9631, (M + H)<sup>+</sup> = 9650, (M + H)<sup>+</sup> = 9669, (M + H)<sup>+</sup> = 9688, (M + H)<sup>+</sup> = 9707, (M + H)<sup>+</sup> = 9726, (M + H)<sup>+</sup> = 9745, (M + H)<sup>+</sup> = 9764, (M + H)<sup>+</sup> = 9783, (M + H)<sup>+</sup> = 9802, (M + H)<sup>+</sup> = 9821, (M + H)<sup>+</sup> = 9840, (M + H)<sup>+</sup> = 9859, (M + H)<sup>+</sup> = 9878, (M + H)<sup>+</sup> = 9897, (M + H)<sup>+</sup> = 9916, (M + H)<sup>+</sup> = 9935, (M + H)<sup>+</sup> = 9954, (M + H)<sup>+</sup> = 9973, (M + H)<sup>+</sup> = 9992, (M + H)<sup>+</sup> = 10011, (M + H)<sup>+</sup> = 10030, (M + H)<sup>+</sup> = 10049, (M + H)<sup>+</sup> = 10068, (M + H)<sup>+</sup> = 10087, (M + H)<sup>+</sup> = 10106, (M + H)<sup>+</sup> = 10125, (M + H)<sup>+</sup> = 10144, (M + H)<sup>+</sup> = 10163, (M + H)<sup>+</sup> = 10182, (M + H)<sup>+</sup> = 10201, (M + H)<sup>+</sup> = 10220, (M + H)<sup>+</sup> = 10239, (M + H)<sup>+</sup> = 10258, (M + H)<sup>+</sup> = 10277, (M + H)<sup>+</sup> = 10296, (M + H)<sup>+</sup> = 10315, (M + H)<sup>+</sup> = 10334, (M + H)<sup>+</sup> = 10353, (M + H)<sup>+</sup> = 10372, (M + H)<sup>+</sup> = 10391, (M + H)<sup>+</sup> = 10410, (M + H)<sup>+</sup> = 10429, (M + H)<sup>+</sup> = 10448, (M + H)<sup>+</sup> = 10467, (M + H)<sup>+</sup> = 10486, (M + H)<sup>+</sup> = 10505, (M + H)<sup>+</sup> = 10524, (M + H)<sup>+</sup> = 10543, (M + H)<sup>+</sup> = 10562, (M + H)<sup>+</sup> = 10581, (M + H)<sup>+</sup> = 10600, (M + H)<sup>+</sup> = 10619, (M + H)<sup>+</sup> = 10638, (M + H)<sup>+</sup> = 10657, (M + H)<sup>+</sup> = 10676, (M + H)<sup>+</sup> = 10695, (M + H)<sup>+</sup> = 10714, (M + H)<sup>+</sup> = 10733, (M + H)<sup>+</sup> = 10752, (M + H)<sup>+</sup> = 10771, (M + H)<sup>+</sup> = 10790, (M + H)<sup>+</sup> = 10809, (M + H)<sup>+</sup> = 10828, (M + H)<sup>+</sup> = 10847, (M + H)<sup>+</sup> = 10866, (M + H)<sup>+</sup> = 10885, (M + H)<sup>+</sup> = 10904, (M + H)<sup>+</sup> = 10923, (M + H)<sup>+</sup> = 10942, (M + H)<sup>+</sup> = 10961, (M + H)<sup>+</sup> = 10980, (M + H)<sup>+</sup> = 11000, (M + H)<sup>+</sup> = 11019, (M + H)<sup>+</sup> = 11038, (M + H)<sup>+</sup> = 11057, (M + H)<sup>+</sup> = 11076, (M + H)<sup>+</sup> = 11095, (M + H)<sup>+</sup> = 11114, (M + H)<sup>+</sup> = 11133, (M + H)<sup>+</sup> = 11152, (M + H)<sup>+</sup> = 11171, (M + H)<sup>+</sup> = 11190, (M + H)<sup>+</sup> = 11209, (M + H)<sup>+</sup> = 11228, (M + H)<sup>+</sup> = 11247, (M + H)<sup>+</sup> = 11266, (M + H)<sup>+</sup> = 11285, (M + H)<sup>+</sup> = 11304, (M + H)<sup>+</sup> = 11323, (M + H)<sup>+</sup> = 11342, (M + H)<sup>+</sup> = 11361, (M + H)<sup>+</sup> = 11380, (M + H)<sup>+</sup> = 11399, (M + H)<sup>+</sup> = 11418, (M + H)<sup>+</sup> = 11437, (M + H)<sup>+</sup> = 11456, (M + H)<sup>+</sup> = 11475, (M + H)<sup>+</sup> = 11494, (M + H)<sup>+</sup> = 11513, (M + H)<sup>+</sup> = 11532, (M + H)<sup>+</sup> = 11551, (M + H)<sup>+</sup> = 11570, (M + H)<sup>+</sup> = 11589, (M + H)<sup>+</sup> = 11608, (M + H)<sup>+</sup> = 11627, (M + H)<sup>+</sup> = 11646, (M + H)<sup>+</sup> = 11665, (M + H)<sup>+</sup> = 11684, (M + H)<sup>+</sup> = 11703, (M + H)<sup>+</sup> = 11722, (M + H)<sup>+</sup> = 11741, (M + H)<sup>+</sup> = 11760, (M + H)<sup>+</sup> = 11779, (M + H)<sup>+</sup> = 11798, (M + H)<sup>+</sup> = 11817, (M + H)<sup>+</sup> = 11836, (M + H)<sup>+</sup> = 11855, (M + H)<sup>+</sup> = 11874, (M + H)<sup>+</sup> = 11893, (M + H)<sup>+</sup> = 11912, (M + H)<sup>+</sup> = 11931, (M + H)<sup>+</sup> = 11950, (M + H)<sup>+</sup> = 11969, (M + H)<sup>+</sup> = 11988, (M + H)<sup>+</sup> = 12007, (M + H)<sup>+</sup> = 12026, (M + H)<sup>+</sup> = 12045, (M + H)<sup>+</sup> = 12064, (M + H)<sup>+</sup> = 12083, (M + H)<sup>+</sup> = 12102, (M + H)<sup>+</sup> = 12121, (M + H)<sup>+</sup> = 12140, (M + H)<sup>+</sup> = 12159, (M + H)<sup>+</sup> = 12178, (M + H)<sup>+</sup> = 12197, (M + H)<sup>+</sup> = 12216, (M + H)<sup>+</sup> = 12235, (M + H)<sup>+</sup> = 12254, (M + H)<sup>+</sup> = 12273, (M + H)<sup>+</sup> = 12292, (M + H)<sup>+</sup> = 12311, (M + H)<sup>+</sup> = 12330, (M + H)<sup>+</sup> = 12349, (M + H)<sup>+</sup> = 12368, (M + H)<sup>+</sup> = 12387, (M + H)<sup>+</sup> = 12406, (M + H)<sup>+</sup> = 12425, (M + H)<sup>+</sup> = 12444, (M + H)<sup>+</sup> = 12463, (M + H)<sup>+</sup> = 12482, (M + H)<sup>+</sup> = 12501, (M + H)<sup>+</sup> = 12520, (M + H)<sup>+</sup> = 12539, (M + H)<sup>+</sup> = 12558, (M + H)<sup>+</sup> = 12577, (M + H)<sup>+</sup> = 12596, (M + H)<sup>+</sup> = 12615, (M + H)<sup>+</sup> = 12634, (M + H)<sup>+</sup> = 12653, (M + H)<sup>+</sup> = 12672, (M + H)<sup>+</sup> = 12691, (M + H)<sup>+</sup> = 12710, (M + H)<sup>+</sup> = 12729, (M + H)<sup>+</sup> = 12748, (M + H)<sup>+</sup> = 12767, (M + H)<sup>+</sup> = 12786, (M + H)<sup>+</sup> = 12805, (M + H)<sup>+</sup> = 12824, (M + H)<sup>+</sup> = 12843, (M + H)<sup>+</sup> = 12862, (M + H)<sup>+</sup> = 12881, (M + H)<sup>+</sup> = 12900, (M + H)<sup>+</sup> = 12919, (M + H)<sup>+</sup> = 12938, (M + H)<sup>+</sup> = 12957, (M + H)<sup>+</sup> = 12976, (M + H)<sup>+</sup> = 12995, (M + H)<sup>+</sup> = 13014, (M + H)<sup>+</sup> = 13033, (M + H)<sup>+</sup> = 13052, (M + H)<sup>+</sup> = 1

# ChEMBL – [www.ebi.ac.uk/chembl](http://www.ebi.ac.uk/chembl)



The screenshot shows the ChEMBL website homepage. At the top, there is a navigation bar with links to various services like Google+, Facebook, and Twitter. Below this is a search bar with the text "Enter Text Here" and a "Find" button. The main content area is divided into several sections:

- ChEMBL Statistics:** A sidebar on the left lists statistics: DB: ChEMBL\_11, Targets: 8,603, Compound records: 1,195,368, Distinct compounds: 1,060,258, Activities: 5,479,146, and Publications: 42,516.
- Getting Started:** A central section with a search bar and navigation buttons (Compounds, Targets, Assays, Activity Source Filter). Below the search bar are buttons for "ChEMBLdb", "Compound Search", "Protein Target Search", "Browse Targets", "Browse Drugs", and "Drug Approvals".
- Support and Feedback:** A section encouraging feedback on the interface and search capabilities.
- Staying in Touch:** A section for subscribing to the ChEMBL-announce mailing list.
- Training:** A section for webinars detailing the interface and schema.
- Data Licensing:** A section explaining that ChEMBL data is available under a Creative Commons Attribution-Share Alike 3.0 Unported License.
- Acknowledgements:** A section thanking contributors to the ChEMBL project.
- ChEMBL Funding:** A section mentioning funding from the Wellcome Trust, EMBL Member States, and the EU Innovative Medicines Initiative (IMI) and Framework 7.

At the bottom of the page, there are logos for Wellcome Trust, EMBL, IMI, and the Seventh Framework Programme.

# Compound Searching

EMBL-EBI  All Databases

Databases Tools EBI Groups Training Industry About Us Help   

EBI > Databases > Small Molecules > ChEMBL Database > Compound Search

Search:



C  
N  
O  
S  
F  
Cl  
Br  
I  
P  
X



**List Search**

SMILES Search  Compound ID Search  Keyword Search

Please enter a list of Compound IDs, keywords, or SMILES separated by newlines

**ChEMBL** 

- ChEMBL Home
- Search Compounds
- Search Targets
- Help
- Acknowledgements
- Contact ChEMBL
- ChEMBL Research
- ChEMBL Blog

**ChEMBL Basic Statistics**

Targets: 7,330  
Compound records: 680,293  
Distinct compounds: 565,245  
Activities: 2,705,136  
Publications: 35,821

ChEMBL database version: ChEMBL\_04  
ChEMBL web version: 1.2.2

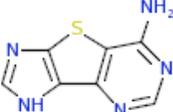
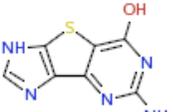
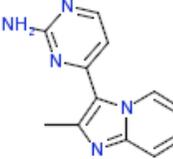
**ChEMBL Blog**

- [Role of open chemical data in aiding drug discovery and design](#)
- [Postdoc position available - ChEMBL and malaria](#)

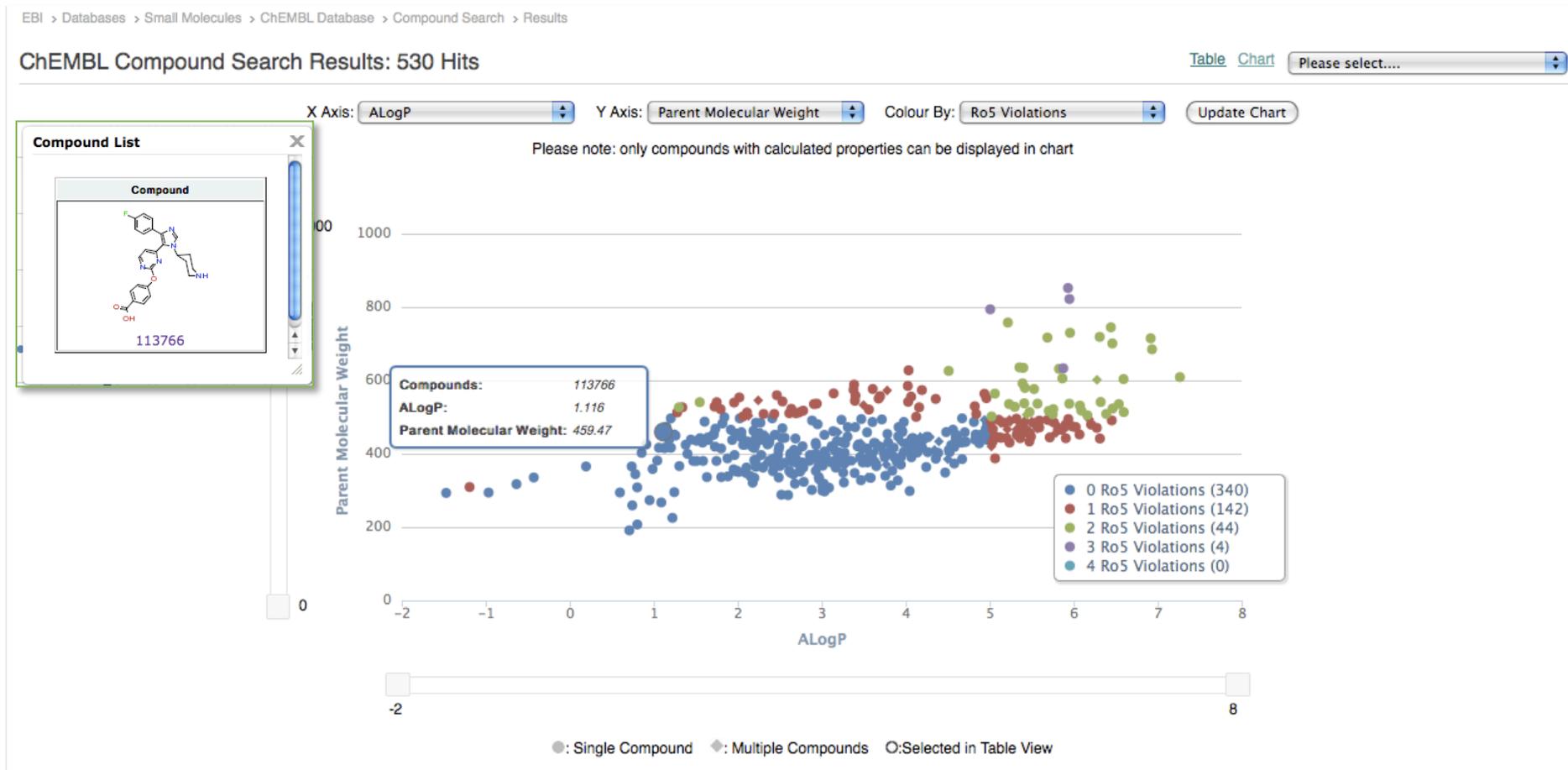
# Chart Views of Data

EBI > Databases > Small Molecules > ChEMBL Database > Compound Search > Results

ChEMBL Compound Search Results: 530 Hits 1 2 3 4 5 6 [Next] [End] Please select....

<input type="checkbox"/>	Compound	Synonyms	Parent Mol Weight	ALogP	PSA	HBA	HBD	#Ro5 Vio.
<input type="checkbox"/>	 <a href="#">340853</a>		191.2	0.71	108.7	4	2	0
<input type="checkbox"/>	 <a href="#">341229</a>		207.2	0.81	129	5	3	0
<input type="checkbox"/>	 <a href="#">268499</a>		225.3	1.22	69.1	4	1	0

# Chart Views of Data



# Target Subset Data

ChEMBL

http://www.ebi.ac.uk/chembl/index.php/target/browser/classification

Analytics ChEMBL Log Twitter LinkedIn Facebook Wikipedia Import to Mendeley Flickr! Doodle jpo iDisk ChEMBL EBI Google Science

EMBL-EBI EB-eye Search All Databases Enter Text Here Go Reset Advanced Search Give us feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

EBI > Databases > Small Molecules > ChEMBL Database > Target Search > Target Classification Hierarchy

Search ChEMBLdb... Compounds Targets Assays

ChEMBLdb Compound Search Protein Target Search **Browse Targets**

Browse  Protein Target Tree  Taxonomy Tree

Click arrows to navigate tree

- Enzyme (2443)
- Membrane receptor (554)
- Ion channel (338)
- Transporter (136)
- Transcription Factor (103)
- Cytosolic other (102)
- Secreted (58)
- Structural (29)
- Surface antigen (26)
- Membrane other (16)
- Adhesion (14)
- Nuclear other (13)

Category	Count
Enzyme	2443
Membrane receptor	554
Ion channel	338
Transporter	136
Transcription Factor	103
Cytosolic other	102
Secreted	58
Structural	29
Surface antigen	26
Membrane other	16
Adhesion	14
Nuclear other	13

ChEMBL

ChEMBLdb

- Search Compounds
- Search Targets

ChEMBL-NTD

Kinase SARfari

ChEMBL Group

Downloads

FAQ

**ChEMBLdb Statistics**

- Targets: 8,054
- Compound records: 726,872
- Distinct compounds: 600,625
- Activities: 2,925,588
- Publications: 38,029
- DB: ChEMBL\_06

**ChEMBL Blog**

- Lecture: Development and Applications of Computational Chemogenomics
- Books: Germ Stories

# Organism Subset Data

The screenshot shows the ChEMBL website interface. The browser address bar displays <http://www.ebi.ac.uk/chembl/index.php/target/browser/organism/>. The page title is "ChEMBL".

The navigation bar includes links for Analytics, ChEMBL Log, Twitter, LinkedIn, Facebook, Wikipedia, Import to Mendeley, Flickr!, Doodle, jpo iDisk, ChEMBL, EBI, Google, and Science. The search bar contains "All Databases" and "Enter Text Here".

The main content area shows the "Target Organism Hierarchy" page. The search bar contains "Search ChEMBLdb...". The navigation tabs include "ChEMBLdb", "Compound Search", "Protein Target Search", and "Browse Targets".

The "Browse" section shows "Protein Target Tree" and "Taxonomy Tree" options. The "Taxonomy Tree" is selected, and the "Click arrows to navigate tree" instruction is displayed.

The taxonomy tree shows the following hierarchy:

- Eukaryotes (5448)
  - Mammalia (5076)
  - Apicomplexa (69)
    - Plasmodium (49)
    - Eimeria (7)
    - Other (7)
    - Cryptosporidium (6)
  - Viridiplantae (68)
  - Kinetoplastida (57)
  - Arthropoda (52)
  - Aves (35)
  - Nematoda (26)
  - Eukaryotes (other) (19)
  - Teleostei (13)
  - Amphibia (12)
  - Lepidosauria (11)
  - Platyhelminthes (7)
  - Echinodermata (3)
- Bacteria (614)
- Unclassified (561)
- Fungi (229)
- Viruses (211)
- Archaea (6)

The pie chart visualizes the distribution of targets across four categories:

- Plasmodium (red)
- Eimeria (green)
- Other (purple)
- Cryptosporidium (blue)

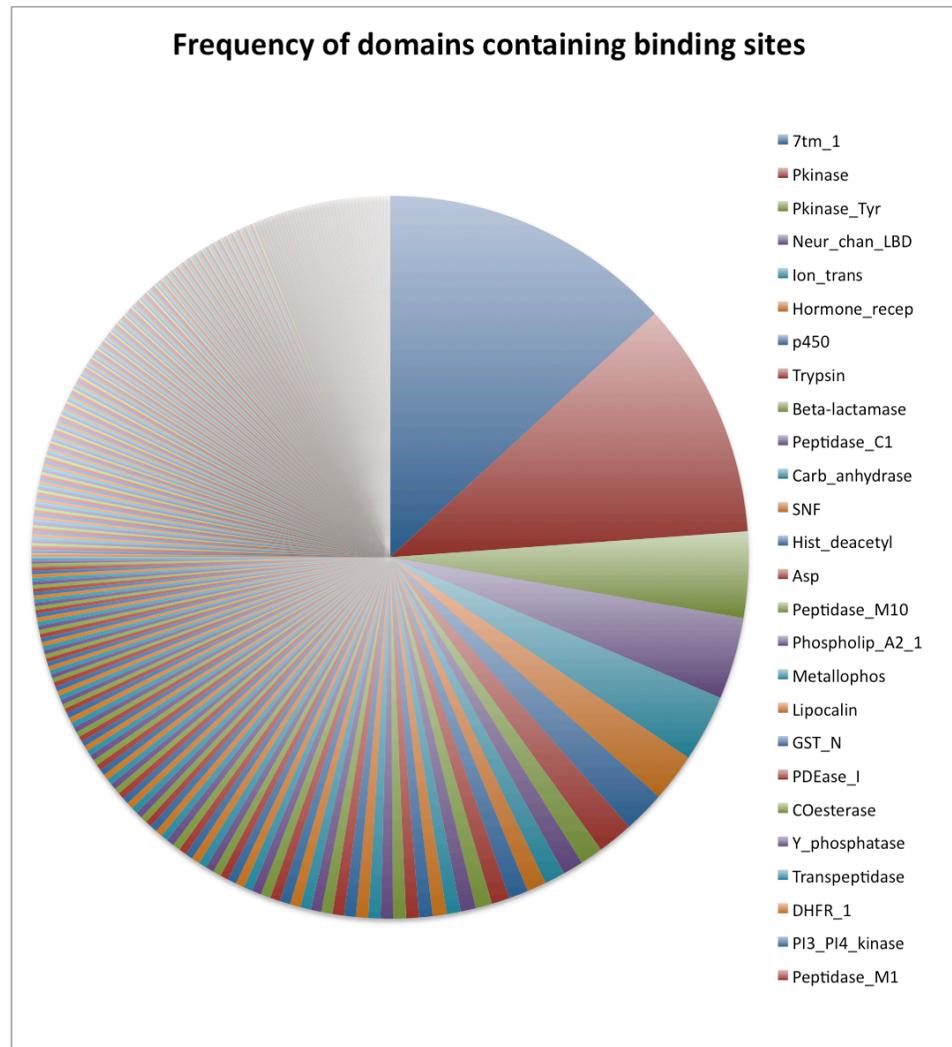
The ChEMBLdb Statistics section shows the following data:

- Targets: 8,054
- Compound records: 726,872
- Distinct compounds: 600,625
- Activities: 2,925,588
- Publications: 38,029
- DB: ChEMBL\_06

The ChEMBL Blog section shows the following links:

- Lecture: [Development and Applications of Computational Chemogenomics](#)
- Books: [Germ Stories](#)

# Domain-level Assignment of Binding



## Depleted and Enriched Pfam Domains

Neur_chan_memb	-1.63
zf-C4	-0.94
ANF_receptor	-0.88
SH2	-0.83
Pkinase_C	-0.70
fn3	-0.53
SH3_1	-0.51
Lig_chan	-0.50
C2	-0.50
C1_1	-0.50
Guanylate_cyc	-0.46
HATPase_c	-0.46
I-set	-0.44
adh_short	-0.39
PH	-0.39
Ank	-0.39
.....	
Metallophos	0.35
Phospholip_A2_1	0.38
Peptidase_M10	0.41
Asp	0.45
SNF	0.48
Hist_deacetyl	0.48
Carb_anhydrase	0.50
Peptidase_C1	0.51
Trypsin	0.51
Beta-lactamase	0.57
p450	1.00
Hormone_recep	1.19
lon_trans	1.66
Neur_chan_LBD	2.02
Pkinase_Tyr	2.12
Pkinase	5.87
7tm_1	7.30

# Drug Approvals

ChEMBLdb

EMBL-EBI  Enter Text Here  Help | Feedback

Databases Tools Research Training Industry About Us Help Site Index

EBI > Databases > Small Molecules > ChEMBL Database > Home

ChEMBL

ChEMBLdb  
ChEMBL-NTD  
Kinase SARfari  
GPCR SARfari  
DrugEBility  
ChEMBL Group  
Downloads  
Web Services  
FAQ

ChEMBLdb Statistics

- DB: ChEMBL\_11
- Targets: 8,603
- Compound records: 1,195,368
- Distinct compounds: 1,060,258
- Activities: 5,479,146
- Publications: 42,516

ChEMBL Blog

- PhD studentship Available for Oct 2012 Intake
- ChEMBL Identifiers

Search ChEMBLdb...    [Activity Source Filter](#)

ChEMBLdb Compound Search Protein Target Search Browse Targets Browse Drugs **Drug Approvals**

The following table lists US Food and Drug Administration (FDA) drug approvals for New Molecular Entities (NMEs) in 2009, 2010, and 2011. Links are to the corresponding Drug Approval Monographs on the ChEMBL-og.

### Browse Drug Approvals

Generic Name	Trade Names	ATC Code	Date of Approval	Drug Monograph	Icon
Brentuximab	Adcetris	<a href="#">L01XC12</a>	19-Aug-2011	<a href="http://chembl.blogspot.com/2011/08/new-drug-approvals-2011-pt-xxv.html">http://chembl.blogspot.com/2011/08/new-drug-approvals-2011-pt-xxv.html</a>	
Vemurafenib	Zelboraf	<a href="#">L01XE15</a>	17-Aug-2011	<a href="http://chembl.blogspot.com/2011/08/new-drug-approvals-2011-pt-xxiv.html">http://chembl.blogspot.com/2011/08/new-drug-approvals-2011-pt-xxiv.html</a>	
Ticagrelor	Brilinta	<a href="#">B01AC24</a>	20-Jul-2011	<a href="http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxiii.html">http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxiii.html</a>	
Indacaterol Maleate	Arcapta	<a href="#">R03AC18</a>	02-Jul-2011	<a href="http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxii.html">http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxii.html</a>	
Rivaroxaban	Xarelto	<a href="#">B01AX06</a>	01-Jul-2011	<a href="http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxi.html">http://chembl.blogspot.com/2011/07/new-drug-approvals-2011-pt-xxi.html</a>	
Azficel-T	laViv	Not Assigned	21-Jun-2011	<a href="http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xx-azficel-t.html">http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xx-azficel-t.html</a>	
Belatacept	Nulojix	<a href="#">L04AA28</a>	15-Jun-2011	<a href="http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xix.html">http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xix.html</a>	
Ezogabine	Potiga	<a href="#">N03AX21</a>	10-Jun-2011	<a href="http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xviii.html">http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xviii.html</a>	
Fidaxomicin	Dificid	Not Assigned	27-May-2011	<a href="http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xvii.html">http://chembl.blogspot.com/2011/06/new-drug-approvals-2011-pt-xvii.html</a>	
Telaprevir	Incivek	Not Assigned	23-May-2011	<a href="http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xvi.html">http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xvi.html</a>	
Rilpivirine	Edurant	Not Assigned	20-May-2011	<a href="http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xv.html">http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xv.html</a>	
Boceprevir	Victrelis	Not Assigned	13-May-2011	<a href="http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xiv.html">http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xiv.html</a>	
Linagliptin	Tradjenta	<a href="#">A10BH05</a>	02-May-2011	<a href="http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xiii.html">http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-pt-xiii.html</a>	
Abiraterone	Zytiga	Not Assigned	28-Apr-2011	<a href="http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-part-xi.html">http://chembl.blogspot.com/2011/05/new-drug-approvals-2011-part-xi.html</a>	

View 1 - 69 of 69

Terms of Use | EBI Funding | Contact EBI | © European Bioinformatics Institute 2011. EBI is an Outstation of the European Molecular Biology Laboratory.

# 'Annotation' of the *S. Mansoni* Genome

nature

Vol 460 | 16 July 2009 | doi:10.1038/nature08160

## ARTICLES

### The genome of the blood fluke *Schistosoma mansoni*

Matthew Berriman<sup>1</sup>, Brian J. Haas<sup>3†</sup>, Philip T. LoVerde<sup>4</sup>, R. Alan Wilson<sup>5</sup>, Gary P. Dillon<sup>5</sup>, Gustavo C. Cerqueira<sup>6,7,8</sup>, Susan T. Mashiyama<sup>9,10</sup>, Bissan Al-Lazikani<sup>11</sup>, Luiza F. Andrade<sup>12</sup>, Peter D. Ashton<sup>4</sup>, Martin A. Aslett<sup>1</sup>, Daniella C. Bartholomeu<sup>3†</sup>, Gaëlle Blandin<sup>3</sup>, Conor R. Caffrey<sup>9</sup>, Avril Coghlan<sup>13</sup>, Richard Coulson<sup>2</sup>, Tim A. Day<sup>14</sup>, Art Delcher<sup>7</sup>, Ricardo DeMarco<sup>5,15,16</sup>, Appolinaire Djikeng<sup>3</sup>, Tina Eyre<sup>1</sup>, John A. Gamble<sup>1</sup>, Elodie Ghedin<sup>3†</sup>, Yong Gu<sup>1</sup>, Christiane Hertz-Fowler<sup>1</sup>, Hirohisa Hirai<sup>17</sup>, Yuriko Hirai<sup>17</sup>, Robin Houston<sup>1</sup>, Alasdair Ivens<sup>1†</sup>, David A. Johnston<sup>18†</sup>, Daniela Lacerda<sup>3†</sup>, Camila D. Macedo<sup>6,8</sup>, Paul McVeigh<sup>14</sup>, Zemin Ning<sup>1</sup>, Guilherme Oliveira<sup>12</sup>, John P. Overington<sup>2</sup>, Julian Parkhill<sup>1</sup>, Mihaela Pertea<sup>7</sup>, Raymond J. Pierce<sup>19</sup>, Anna V. Protasio<sup>1</sup>, Michael A. Quail<sup>1</sup>, Marie-Adèle Rajandream<sup>1</sup>, Jane Rogers<sup>1†</sup>, Mohammed Sajid<sup>9†</sup>, Steven L. Salzberg<sup>7,8</sup>, Mario Stanke<sup>20</sup>, Adrian R. Tivey<sup>1</sup>, Owen White<sup>3†</sup>, David L. Williams<sup>21†</sup>, Jennifer Wortman<sup>3†</sup>, Wenjie Wu<sup>4†</sup>, Mostafa Zamanian<sup>14</sup>, Adhemar Zerlotini<sup>11</sup>, Claire M. Fraser-Liggett<sup>3†</sup>, Barclay G. Barrell<sup>1</sup> & Najib M. El-Sayed<sup>3,6,7,8</sup>

*Schistosoma mansoni* is responsible for the neglected tropical disease schistosomiasis that affects 210 million people in 76 countries. Here we present analysis of the 363 megabase nuclear genome of the blood fluke. It encodes at least 11,809 genes, with an unusual intron size distribution, and new families of micro-exon genes that undergo frequent alternative splicing. As the first sequenced flatworm, and a representative of the Lophotrochozoa, it offers insights into early events in the evolution of the animals, including the development of a body pattern with bilateral symmetry, and the development of tissues into organs. Our analysis has been informed by the need to find new drug targets. The deficits in lipid metabolism that make schistosomes dependent on the host are revealed, and the identification of membrane receptors, ion channels and more than 300 proteases provide new insights into the biology of the life cycle and new targets. Bioinformatics approaches have identified 11,809 genes that are essential for the chemogenomic screen has pinpointed schistosome proteins for which existing drugs might provide an invaluable resource for the research community to develop much needed interventions for the treatment and eradication of this important and neglected disease.

disease that ranks with malaria as one of the most common causes of morbidity affecting 210 million people in 76 countries, despite strenuous efforts to control the spread of the genus *Schistosoma*. The diversity of the species and their complex life cycles and complex phenotypes in different environments. *Schistosoma mansoni*, which occurs across much of Africa, South America, the Caribbean, Brazil, Venezuela and the Middle East, dwells in the human small intestine and the intestinal wall that either

pass to the gut lumen and are voided in the faeces, or travel to the liver where they trigger immune-mediated granuloma formation and peri-portal fibrosis<sup>2</sup>. Approximately 280,000 deaths per annum are attributable to schistosomiasis in sub-Saharan Africa alone<sup>3</sup>. However, the disease is better known for its chronicity and debilitating morbidity<sup>4</sup>. A single drug, praziquantel, is almost exclusively used to treat the infection but this does not prevent reinfection, and with the large-scale control programmes in place, there is concern about the development of drug resistance. Indeed, resistance can be selected for in the laboratory and there are reports of increased drug tolerance in the field<sup>5</sup>.

Genomics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. <sup>2</sup>The Institute for Genomic Sciences and Technology, Rockville, Maryland 20850, USA. <sup>3</sup>Department of Biochemistry and Pathology, Mail Code 7760, University of Maryland, USA. <sup>4</sup>Department of Biology, University of York, PO Box 373, York YO10 5YW, UK. <sup>5</sup>Department of Cell Biology and Anatomical Biology, and <sup>6</sup>Maryland Pathogen Research Institute, University of Maryland, College Park, Maryland 20742, USA. <sup>7</sup>Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, California Institute for Quantitative Biomedical Sciences, San Francisco, California 94158-2330, USA. <sup>8</sup>Cancer Research UK Centre for Cancer Therapeutics, The Institute for Cancer Research, Belmont, Surrey SM2 5NG, UK. <sup>9</sup>Centro de Pesquisas René Rachou (CPRR)—FIOCRUZ, Av. Augusto de Lima s/n, 41280-000, Belo Horizonte, Minas Gerais, Brazil. <sup>10</sup>Department of Microbiology, University College Cork, Western Road, Cork, Ireland. <sup>11</sup>Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA. <sup>12</sup>Instituto de Física de São Carlos, Universidade de São Paulo, Brazil. <sup>13</sup>Primate Research Institute, Kyoto University, Inuyama, Aichi 486-8602, Japan. <sup>14</sup>Department of Biology, University of Cambridge, Cambridge CB2 3RQ, UK. <sup>15</sup>Inserm, U 547, Université Lille 2, Institut Pasteur de Lille, Pasteur Institute, Lille, France. <sup>16</sup>Department of Biology, University of Cambridge, Cambridge CB2 3RQ, UK. <sup>17</sup>Institut für Bioinformatik, Universität Göttingen, Goldschmidtstraße 1, Göttingen 37077, Germany. <sup>18</sup>Department of



Nature 460, 352-358 (2009)

Table 1 | *S. mansoni* genes that match a human gene with marketed drugs

Gene identifier	Protein description	Potential drugs
Smp_005210	Histone deacetylase 1 (HDAC1)	Vorinostat†
Smp_009030	Ribonucleoside-diphosphate reductase, α subunit, putative	Fludarabine phosphate†
Smp_012930	Inosine-5-monophosphate dehydrogenase, putative	Mycophenolate mofetil†, mycophenolic acid†, ribavirin†
Smp_015020	Na <sup>+</sup> /K <sup>+</sup> -ATPase α subunit (SNaK1)	Digoxin†, digitoxin†, acetyldigitoxin†, deslanoside†
Smp_016780*	Tubulin α chain, putative	Vinblastine†, colchicine†, vincristine†
Smp_022960	Aldehyde dehydrogenase, putative	Disulfiram†
Smp_026560	Calmodulin, putative	Bepiridil†
Smp_030730*	Tubulin β chain, putative	Colchicine†, vinblastine†, vincristine†, albendazole†, mebendazole†, paclitaxel†, thiabendazole†, vinorelbine†, docetaxel†
Smp_040130	Cyclophilin (p17.7)	Cyclosporine†
Smp_040790	Cyclophilin B	Cyclosporine†
Smp_044440	Alcohol dehydrogenase, putative	Fomepizole†
Smp_048430	Thioredoxin glutathione reductase	Auranofin†
Smp_050390	Aldehyde dehydrogenase, putative	Disulfiram†
Smp_053220	Aldo-keto reductase, putative	Tolrestat†
Smp_055890	Ribonucleoside-diphosphate reductase small chain, putative	Hydroxyurea†, gemcitabine†
Smp_065120	Deoxyhypusine synthase, putative	Ciclopirox†
Smp_069160	Cyclophilin, putative	Cyclosporine†
Smp_079230	Immunophilin FK506 binding protein FKBP12, putative	Pimecrolimus†, temsirolimus†, sirolimus†, tacrolimus†
Smp_093280	Histone deacetylase 3 (HDAC3)	Vorinostat†
Smp_094810	Cyclophilin E	Cyclosporine†
Smp_121920	Vesicular amine transporter, putative	Rauwolfia serpentina†, reserpine†, deserpidine†, rescinnamine†, alseroxylon†
Smp_135460	Bitransferase dihydrofolate reductase-thymidylate synthase, putative	Pemetrexed†, flucytosine†, flouxuridine†, capecitabine†, fluorouracil†
Smp_136300	Tyrosine kinase 5	Dasatinib†
Smp_147050	ATP synthase α subunit vacuolar, putative	Tiludronate†
Smp_171580	Aromatic amino acid decarboxylase, putative	Carbidopa†
Smp_173280	Cyclophilin, putative	Cyclosporine†

Gene identifier is the genome project systematic name for each gene. It corresponds to the locus tag in the DDBJ/EMBL/GenBank record and to the main accession numbers for GeneDB.

\* There are several copies of tubulin (α, Smp\_027920, Smp\_090120 and Smp\_103140; β, Smp\_192110, Smp\_079960, Smp\_079970, Smp\_078040 and Smp\_035760).

† The potential drugs are classified according to the confidence with which the efficacy of the drug in human can be attributed to the target.

‡ Direct and clear evidence that this interaction is primarily responsible for the therapeutic action of the drug.

§ Direct and clear evidence that this interaction represents one mechanism for the drug, other targets/mechanisms may also exist.

¶ Indirect or inferred evidence of the association of the drug, target and therapeutic action, although the exact mechanism is still speculative.

Map established intervention points on basis of known pharmacology/ medicinal chemistry

# Errors/Differences in Published Bioactivities

## Global mapping of small molecule binding to homologous proteins



Felix A. Krüger, John P. Overington

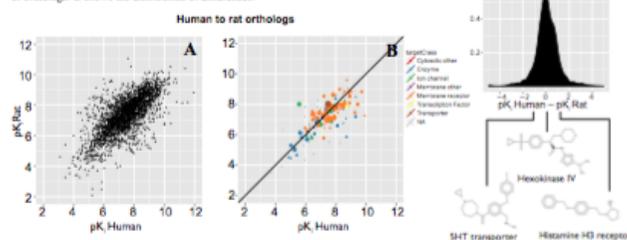
European Bioinformatics Institute, Wellcome Trust Genome Campus, CB10 1SD, UK  
fkrueger@ebi.ac.uk

### Context

We present a large scale study of small molecule binding to related protein targets. Pharmacological data and homology information was integrated and small molecule binding to protein targets was compared for pairs of human to rat orthologs. Our results indicate that small molecule binding between the human and rat species is largely conserved, but not entirely. Proteins exhibiting different pharmacology in the two species were identified using statistical testing. A homology model of the Histamine H3 receptor was used to elucidate the molecular mechanism underlying this species specific pharmacology. Our study of the robustness of small molecule binding across species confirms a longstanding critical assumption in the field of medicinal and biological chemistry - that of the utility of model organisms to study human biology.

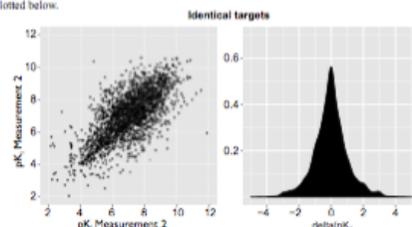
### Comparing small molecule binding for human and rat orthologs

Using the ChEMBL data base [1], differences in binding affinity have been compared between rat and human orthologs. Pairs of human to rat orthologs were identified using the Compara Gene Tree [2] and the Ensembl data base [3]. A shows the raw data and B the aggregate affinity of all compounds tested against a given pair of orthologs. C shows the distribution of differences.



### Inter-assay variation

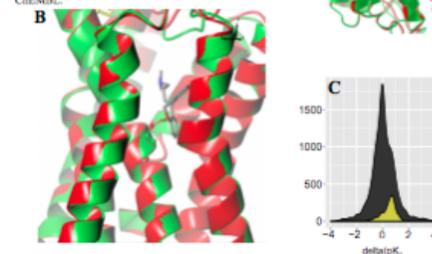
Much of the difference observed when comparing binding affinities for human and rat targets is due to high noise levels in the data. The measurements we compare in this study are aggregated from the literature and have been taken at different times in different assay systems and different laboratories. To control for this variation, we compared measurements taken for identical targets and ligands. To reflect the composition of the human vs. rat distribution, we randomly selected 1500 pairs of measurements taken against human proteins and 1500 measurements taken against rat proteins. The resulting distribution is plotted below.



### Species specific pharmacology of HRH3

Using Mann-Whitney U tests, we identified 14 pairs of human to rat orthologs for which distributions of observed differences in small molecule binding were significantly different from the observed inter-assay variations. For some of the listed targets, we found further evidence for species specific pharmacology in the literature, most prominently the Histamine H3 receptor (HRH3). We used available structures of G-protein coupled receptors [4] and the MODELLER software [5] to create a homology model of the HRH3 and analysed residues near the binding site of Doxepin (this ligand was taken from the crystal structure of the closest homolog, HRH1) to get an idea of the molecular mechanism underlying the species specific pharmacology of the human and rat HRH3. Between the human and rat receptor, we identified a substitution Thr119Ala in the fourth transmembrane helix, which is ~3 Å from the ligand and likely the cause of the observed differences.

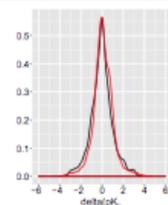
Superpositions of homology models of the human (green) and rat (red) HRH3 are shown below. A shows the full receptor and highlights the overlap of the transmembrane helices; the calculated RMSE between the two models is 1.09 Å. B highlights the Threonine residue at position 119 in the human receptor, which is substituted for an Alanine in the rat receptor. Its proximity to the superposed ligand Doxepin suggests it plays a crucial role in ligand binding. We propose this is why on average, small molecules bind the human receptor with 0.5 log units higher affinity. C shows the distribution of observed differences for HRH3 ligands (yellow) compared to all interactions retrieved from ChEMBL.



### Conservation of small molecule binding

The high similarity between the distribution of differences observed for identical targets and the same distribution for orthologs suggests that small molecule binding between the human and rat species is globally conserved. Both distributions can quite accurately be described by a Laplace distribution, with a position parameter  $\mu = 0$  and a scale parameter  $b = 1.3$ .

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$



### Summary

We used large scale data integration to show that small molecule binding is globally conserved between human and rat orthologs. Using the distribution of inter-assay differences as a control, we were able to identify pairs of targets which do not follow the global rule of conserved pharmacology among them the HRH3. For this receptor, we propose that a substitution from Threonine to Alanine between the human and rat receptor is the cause of species specific pharmacology.

### References

- [1] ChEMBL version 10, [ftp://ebl.ac.uk/pub/databases/chembl/ChEMBLdb/release/chembl\\_10](http://ebl.ac.uk/pub/databases/chembl/ChEMBLdb/release/chembl_10)
- [2] A. J. Vilella, J. Severin, A. Ureta-Vidal, et al., "EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates", *Genome Res.* 19, pp.227-235, 2009
- [3] Ensembl version 38, <http://www.ensembl.org>
- [4] for full references see PDB entries 3p8l, 2rh1, 2v84, 3rya, 3ry8, 3d6a, 3zcc.
- [5] Sali A., Poterro L., Yuan F. Evaluation of comparative protein modeling by MODELLER. *Protein* 23, pp318-326, 1995

# Targets of Launched Drugs

Nat. Rev. Drug Disc., 5, pp. 993-996 (2006)

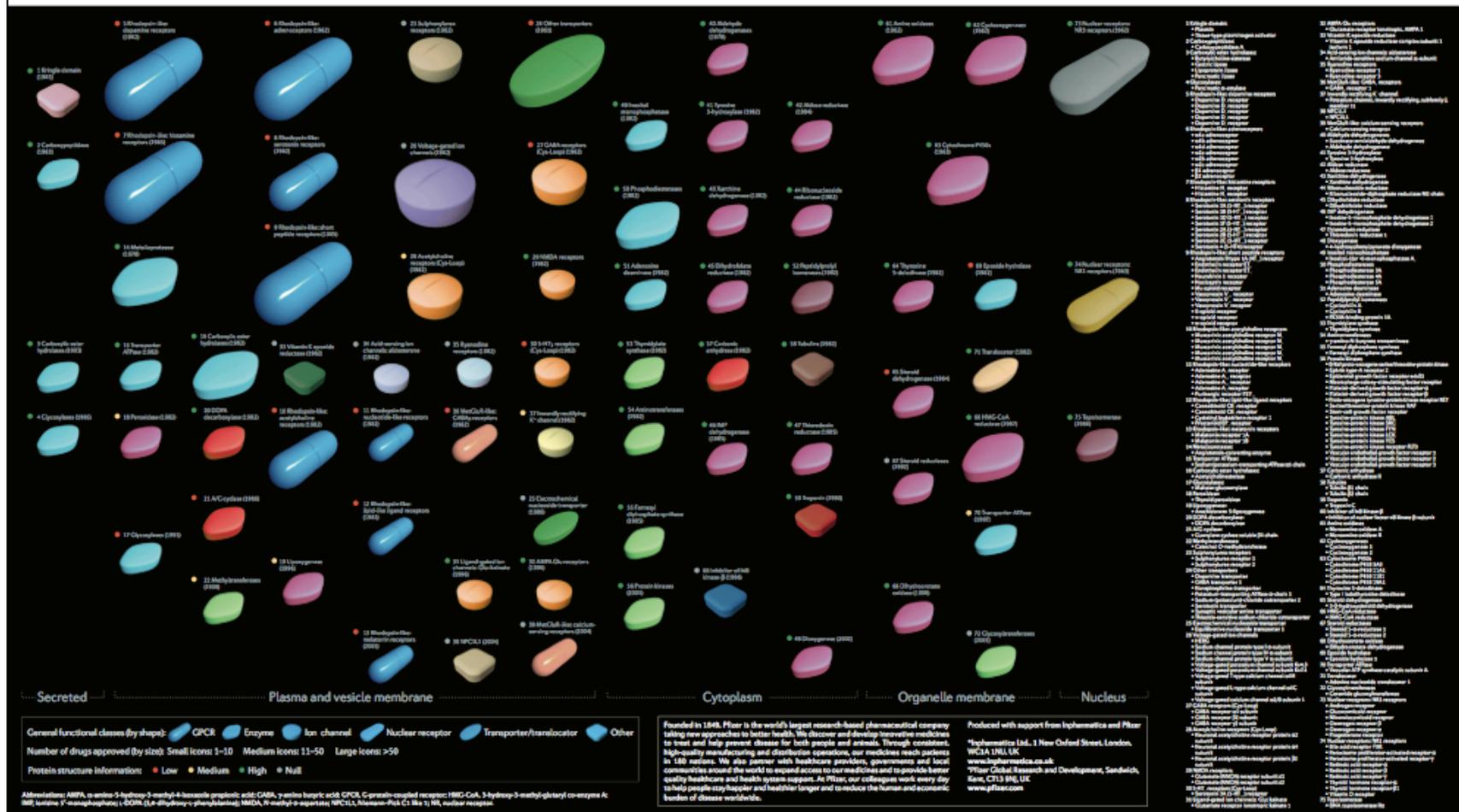
**nature**  
REVIEWS  
DRUG  
DISCOVERY

## The molecular pharmacopoeia

The human targets of FDA-approved oral drugs

John P. Overington\*, Bissan Al Lazikani\* and Andrew L. Hopkins<sup>1</sup>

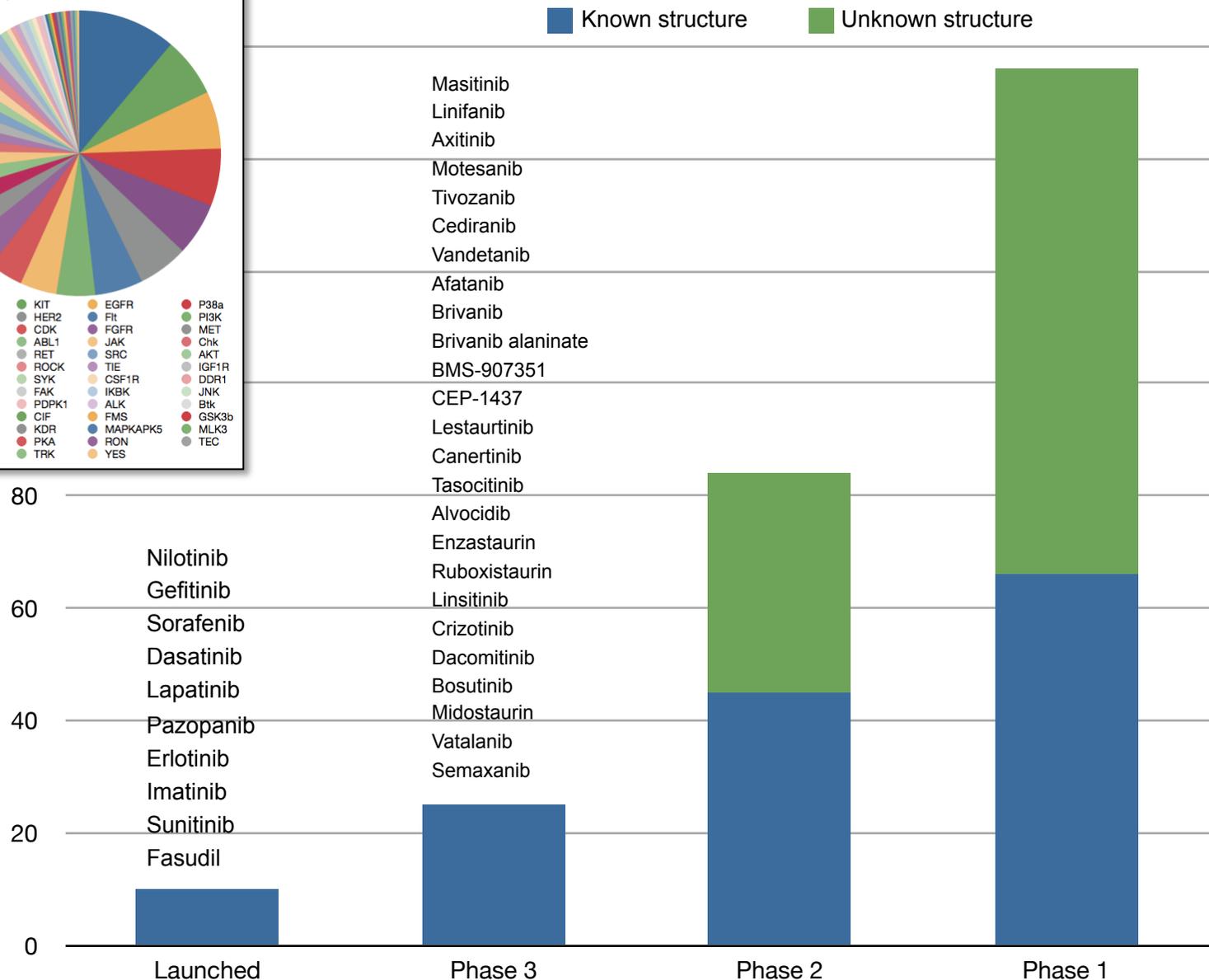
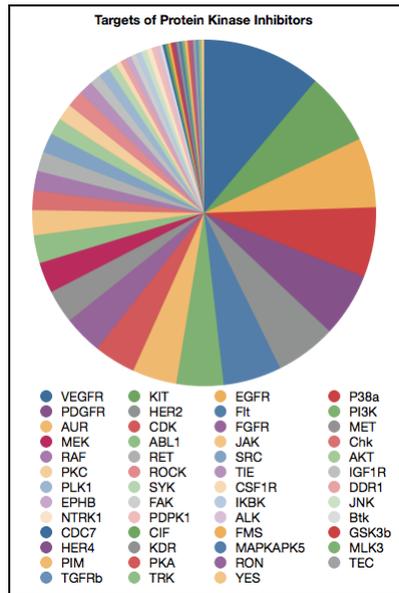
The molecular pharmacopoeia of 186 human protein drug targets for FDA-approved oral drugs. The drug targets are grouped into target subfamilies where there are multiple related drug targets. The size of the icons represents the number of drugs approved for that target or target subfamily. The horizontal axis illustrates the cellular location of a drug target. The vertical axis illustrates an approximate timeline depicting when the first drug for a target or target subfamily was approved (older drug targets are at the top of the chart). The dates next to the targets illustrate the year in which the first USAN (United States Adopted Name) was assigned for the first drug against that target or any target in the subfamily (which usually occurs in the late stages of clinical development). The availability of protein structural information for the target or target subfamily is illustrated by the coloured dot next to the target name. The shapes of the icons represent the general functional classes of drug targets and related groups within a functional class are coloured the same.



# ChEMBL REST Web Services

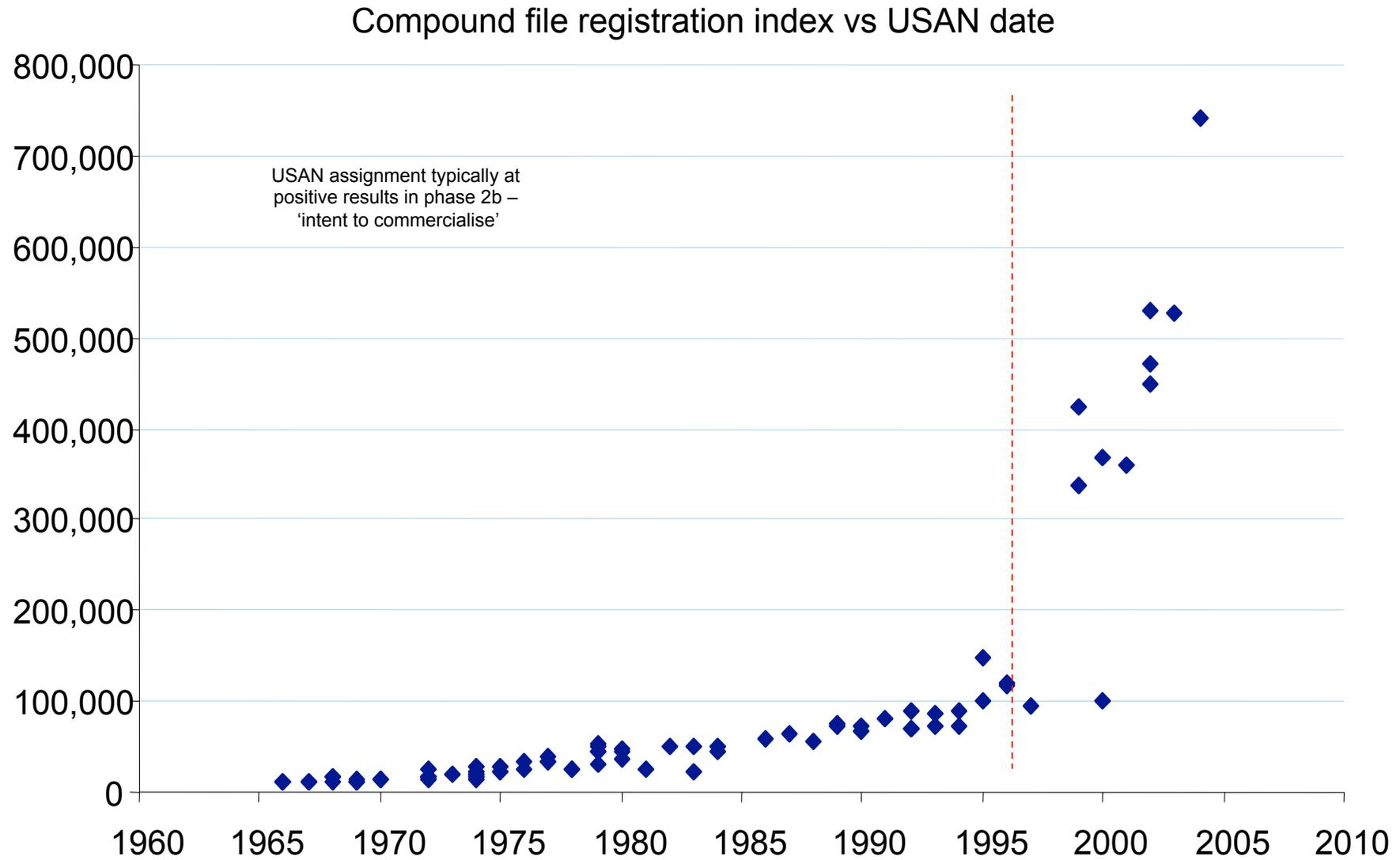
- Examples
  - <https://www.ebi.ac.uk/chemblws/compounds/CHEMBL1>
  - <https://www.ebi.ac.uk/chemblws/compounds/stdinchikey/QFFGLORLPOAEC-SNVBAGLBSA-N>
- Documentation
  - <https://www.ebi.ac.uk/chembl/db/index.php/ws>
- Currently testing search based Web Services
  - Keyword-based searches across the database
  - BLAST searches (find similar targets)
  - Chemical structure searches (find similar compounds)

# Kinase Inhibitors in Clinical Development

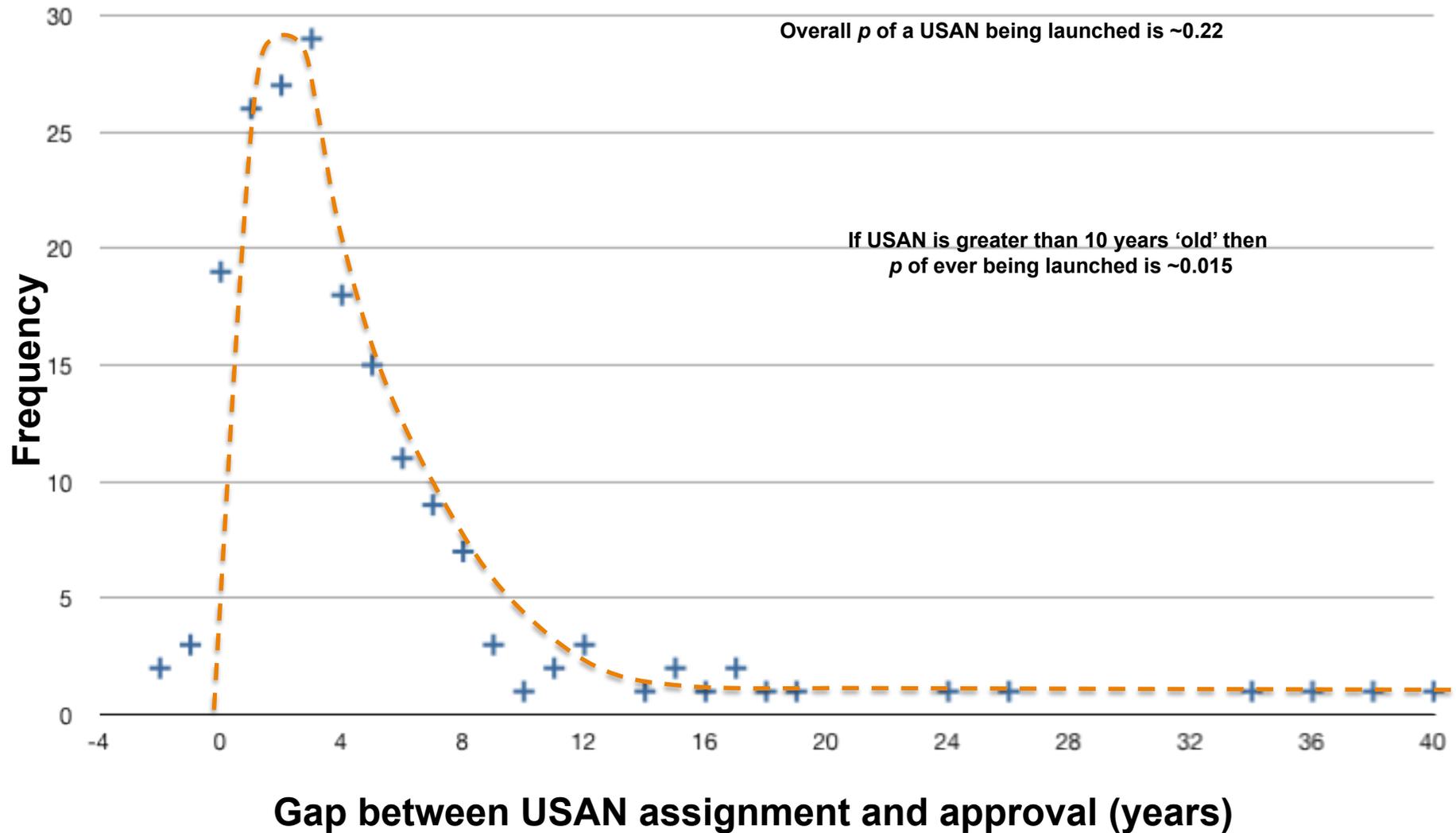


Total of 329 clinical phase kinase inhibitors – August 2011

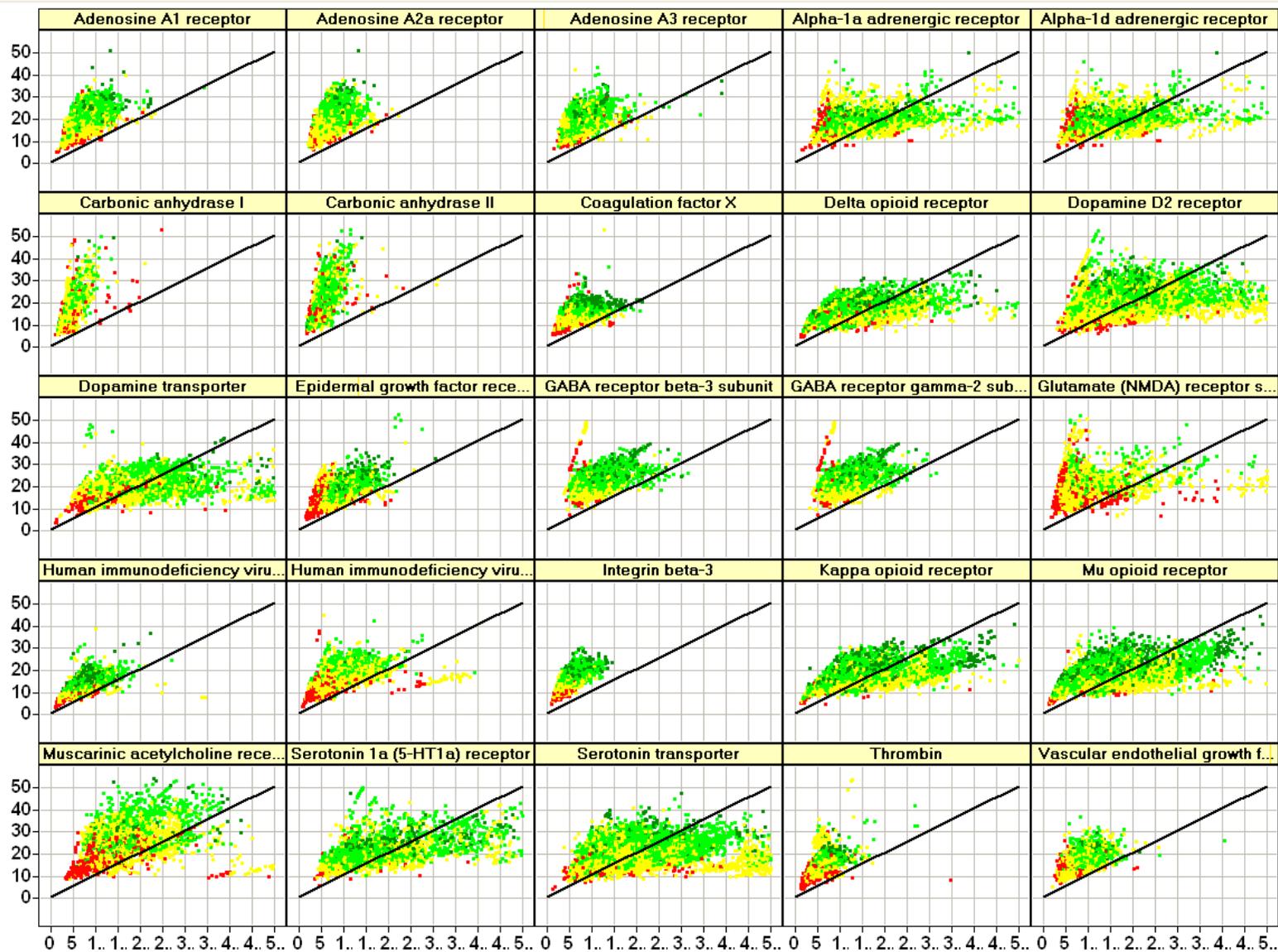
# Pharma Industry Productivity



# Insights Into Attrition/Approval



# Visualisation of Ligand Efficiency Space



# Acknowledgements

## EMBL-EBI

Anne Hersey  
Anna Gaulton  
Mark Davies  
Louisa Bellis  
Kazuyoshi Ikeda  
Patricia Bento  
Jon Chambers  
Shaun McGlinchey  
Yvonne Light

## Inpharmatica

Rick Fagan  
Dave Michalovich  
Edith Chan  
Chris Carline  
John Bouquiere  
David Gee  
Richard Cox  
Scott Dann  
Ian Carruthers  
Bart Ailey



## Institute of Cancer Research (ICR)

Bissan Al-Lazikani  
Mark Halling-Brown

## University of Dundee

Andrew Hopkins  
Richard Bickerton

## Bayer-Schering

Ursula Egner  
Judith Günther  
Jörg Wichard

## Pfizer

Eric Fauman  
Lee Harland  
Jerry Lanfear  
Chris Lipinski

## Merck

Brad Sherbourne

## University Illinois at Chicago

Cele Abad-Zapatero



**Funding for Computational Chemical Biology Group at EMBL is from The Wellcome Trust, The European Commission, The Innovative Medicines Initiative, EMBL Member Nations, EPSRC, BBSRC, MRC, GlaxoSmithKline, Pfizer & Sanofi-Aventis**

The ChEMBL-og - Open Data For Drug Discovery: ChEMBL Identifiers

http://chembl.blogspot.com/2011/08/chembl-identifiers.html

Google

Follow Share Report Abuse Next Blog» john.p.overington@gmail.com New Post Design Sign Out

# The ChEMBL-og - Open Data For Drug Discovery



The news, progress, whereabouts, and ephemera from the Computational Chemical Biology group (ChEMBL) at the EMBL-EBI.

[ChEMBL database](#) [ChEMBL-NTD](#) [GPCR SARfari](#) [Kinase SARfari](#)

MONDAY, 22 AUGUST 2011

## ChEMBL Identifiers



A few notes about the use and format of identifiers in ChEMBL:

Each of the major entity types within ChEMBL (documents, assays, compounds and targets) are assigned unique ChEMBL identifiers, which take the form of a 'ChEMBL' prefix followed immediately by an integer (e.g., ChEMBL25 is the compound aspirin, ChEMBL210 is the human beta-2 adrenergic receptor "target"). There is no distinction between the format of the identifier for different types of entities, but a given ChEMBL identifier will only ever be assigned to a single entity (i.e., ChEMBL25 will only ever be used for the compound aspirin and never for an assay, document or target). A lookup table is provided in the database, to resolve which identifiers correspond to which entity types.

ChEMBL identifiers are stable with respect to the entities they represent. For compounds (with known/defined structures), ChEMBL identifiers represent distinct compound structures, as defined by the standard InChI, e.g., ChEMBL25 represents: InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12). Therefore, two compounds reported in different papers but having the same standard InChI will be assigned the same ChEMBL ID.

### SUBSCRIBE TO THE CHEMBL-OG

Posts

Comments

### ABOUT CHEMBL



This informal blog (the ChEMBL-og) covers the activity and interests of the Computational Chemical Biology Group at the EMBL-EBI in Hinxton. Our interests cover Drug Discovery, Computational Chemical Biology, Chemogenomics, Chemoinformatics, Bioinformatics, Structural Biology, Open Data, Knowledge Management, and Data Integration and include...

- [ChEMBLdb](#) - a drug discovery database (previously known as [StARLite](#), [CandiStore](#) and [DrugStore](#)). Data-mining/KDD approaches designed to improve drug discovery efficiency.
- [SARfari](#) - a sequence, structure, SAR integration platform.