Supporting Information

MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist

Gabriele Cruciani *[¤], *Emanuele Carosati* [¤], *Benoit De Boeck* [•1], *Kantharaj Ethirajulu*[•2], *Claire Mackie*[•2], *Trevor Howe* [•3], *Riccardo Vianello* [♣]

[¤] Laboratory for Chemometrics and Cheminformatics, Chemistry Department, University of Perugia, Via Elce di Sotto, 10, I-06123, Perugia – Italy [•1] Medicinal Chemistry Department, [•2] ADME Department, [•3] Molecular Informatics Department, Johnson and Johnson Pharmaceutical Research and Development, A division of Janssen Pharmaceutica, Turnhoutseweg 30, 2340 Beerse, Belgium

[♣] Molecular Discovery Ltd, 215 Marsh Road, Pinner, Middlesex, HA5 5NE London – UK

Table of Contents:

Table S1. Compounds used in all the examples, in SMILES format.

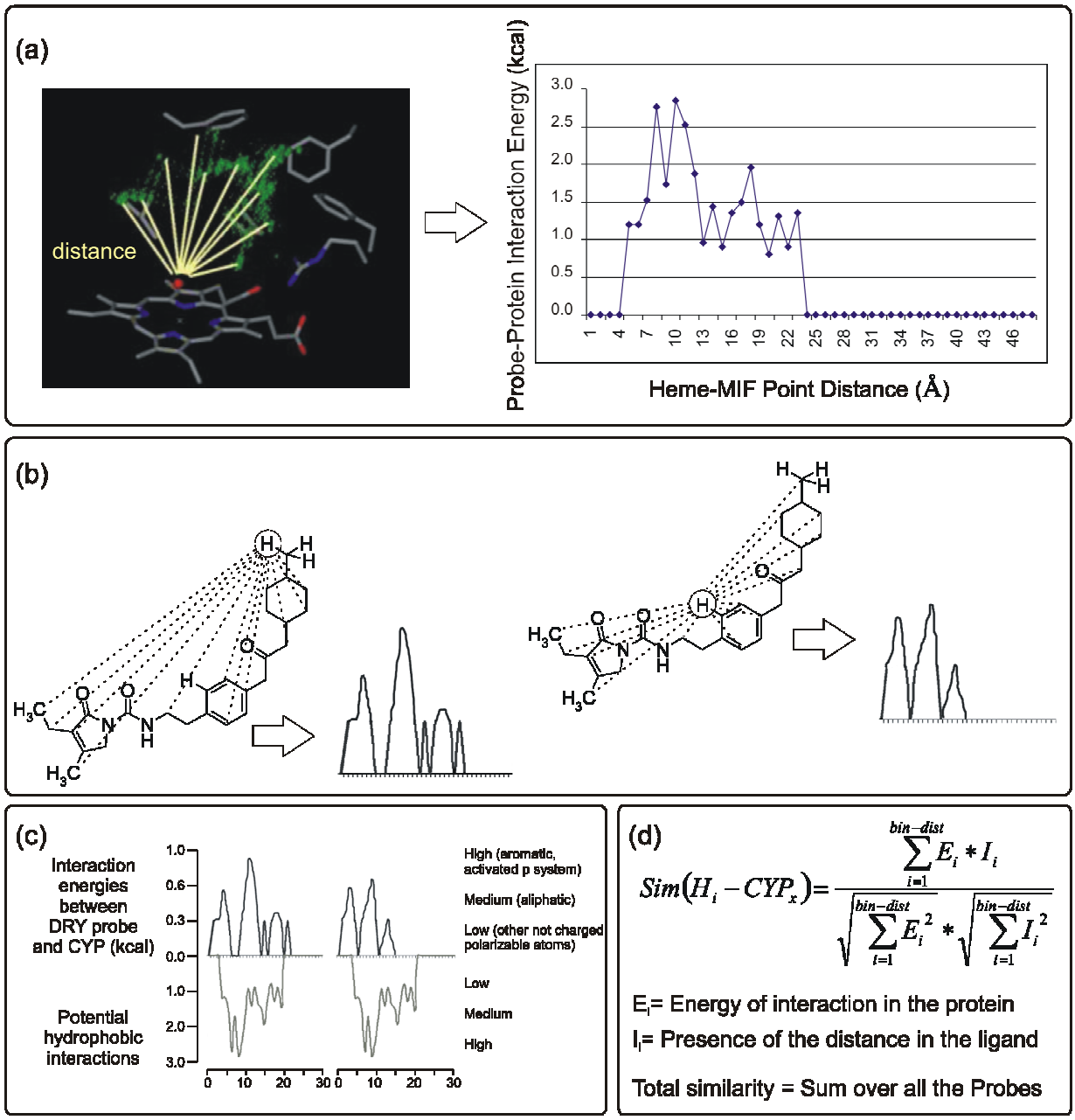
Table S2. Flow-chart calculation for the **E*i*** component.

Table S3. Flow-chart calculation for the **R*i*** component and $\mathbf{P_{SM}}$ values.

Table S1. Compounds used in all the examples, in SMILES format: all calculations were run with the reactivity component ON and the results were analyzed looking at the "average ranking".

Figure	Compound	SMILES	CYP
3	Eugenol	c1cc(c(cc1CC=C)OC)O	2D6
7	Pyrimidin-5-one derivative	Clc1cc2c(cc1)N[C@H]1[C@H]2CN(CC1)CCc1c(nc2n(c1=O)C(=C(S2)C)C)C	3A4; 2C9
8	EGF-receptor inhibitor	N(c1ccccc1)c1c(cc2c(c1)C(NC2=O)=O)Nc1ccccc1	3A4
9	(+)-*cis*-Diltiazem	COc1ccc(cc1)[C@@H]1Sc2ccccc2N(CCN(C)C)C(=O)[C@@H]1OC(=O)C	3A4
10	Inhibitor cholesterol absorption	N1([C@@H]([C@@H](C1=O)CCCc1ccccc1)c1ccc(cc1)OC)c1ccc(cc1)OC	3A4
11a	Bunitrolol	CC(C)(C)NC[C@@H](O)COc1ccccc1C#N	2D6
11b	Desmethylbupranolol	C(NC[C@H](COc1c(cccc1)Cl)O)(C)(C)C	2D6
11c	N-ethyl-amphetamine	CCN[C@@H](C)Cc1ccccc1	2D6
11d	N-3-propynyl-amphetamine	C#CN[C@@H](C)Cc1ccccc1	2D6

Table S2. Flow-chart calculation for the **E*i*** component.



(a) The Molecular Interaction Fields (MIF) are calculated for a CYP-enzyme, using the hydrophobic probe. The 3D-MIF representation is encoded into a correlogram reporting the interaction energies and the distances from the heme-iron moiety. The correlogram represents the fingerprint of the CYP-hydrophobic interaction from a point of view of the heme-iron atom.

(b) Among many possibilities two atoms of the same CYP-ligand are highlighted: starting from these atoms, a set of distances are calculated between the two atoms and the hydrophobic part of the molecule. Two correlograms are produced.

(c) The protein-correlogram and the ligand-correlogram are compared.

(d) The comparison yields a similarity value obtained from this equation

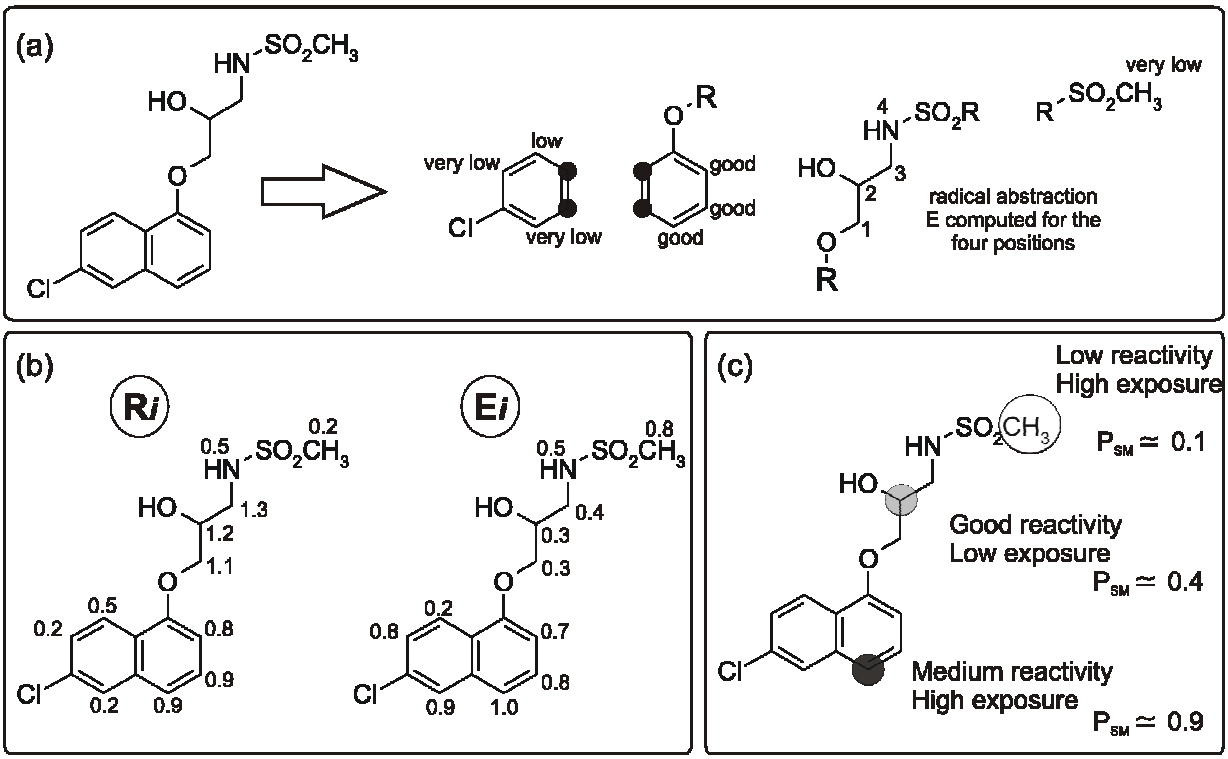
$$Sim(H_i - CYP_x) = \frac{\sum_{i=1}^{bin-dist} E_i * I_i}{\sqrt{\sum_{i=1}^{bin-dist} E_i^{\,2}} * \sqrt{\sum_{i=1}^{bin-dist} I_i^{\,2}}}$$

Finally, the **E*i*** component is the overall summation on each probe-CYP similarity

$$E_i = \sum_{p=1}^{n} Sim(H_i)_p$$

where n = number of probes used.

Table S3. Flow-chart calculation for the **R*i*** component and $\mathbf{P_{SM}}$ values.



(a) The ligand substrate is dissected in fragments for which the atom reactivity (toward radical abstraction) is known and reported in an internal database. The reactivity of different hydrogen atoms in chlorobenzene are known, and are easily assigned. A similar case is represented by the activated methoxybenzene, or by the strongly-inactivated methylsulphonyl groups. Part of the molecule is not recognized as an available fragment, therefore AM1 calculations are performed on all the fragment atomic positions.

(b) The atom reactivity values retrieved from a database of fragments, or computed on-the-fly, are then scaled from 0.0 (no reactivity) to 1.5 (high reactivity). A similar scale reporting the **E*i*** component is then used to compute $\mathbf{P_{SM}}$ values.

(c) The probability of site of metabolism is calculated as a product of the **R*i*** and **E*i*** components. To be site of metabolism, an atom should possess both non neglecting accessibility and reactivity components with the heme.