Supporting Information

Significant Enhancement of Docking Sensitivity using Implicit Ligand Sampling

Markus A. Lill, Mengang Xu

Table S1.High-affinity ligands specific for human (hDHFR), Pneumocystis carinii (pcDHFR), and Candida albicans (caDHFR) dihydrofolate reductase. Ligand names and scaffold types are consistent with those in Bowman et al. .[1]

pcDHFR		hDHFR		caDHFR	
Scaffold	Name	Scaffold	Name	Scaffold	Name
D	Methotrexate	D	AMT	N	HH-133
D	029630	D	8	N	HH-136
D	098579	D	9	O	8a
D	098580	D	10	O	8b
D	107146	D	6	O	8d
D	136735	F	1	O	8f
D	136736	F	5	O	8g
D	144698	F	16	O	8h
D	152737	F	17	O	8k
D	169531	F	36	O	8l
E	174121	F	105952	O	8m
F	122870, 19	F	13	O	8r
F	129516, 22	F	14	O	8s
F	132483	F	15	O	8t
F	184692	F	2	O	8u
F	15	K	5-Me-5-dAMT	O	8v
J	21	K	5-Me-5-dMTX	O	8x
K	21	K	5-Me-10-Et-5-dAMT	O	13d
K	13n, 2	K	5-dAMT		Ro 17-3279
K	13q, 43	K	11		

Table S2. Results of multi-linear regression analysis for ligands BMZ, DA2K, C24 and G3X correlating $d_{ij}^{<}$ and $d_{ij}^{>}$ with the RMSD deviation from the native ligand conformation.

Ligand: BMZ	r^2=0.42	c_0=-8.02	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
His57, C_γ	Glu192, C_δ	>	0.16
Tyr60A, C_α	Trp60D, C_α	<	-2.12
Tyr60A, C_α	Ser195, C_α	<	0.65
Tyr60A, C_ξ	Asp189, C_γ	<	-0.51
Trp60D, C_α	Ile174, $C_{\gamma1}$	>	0.14
Trp60D, $C_{\delta2}$	Glu192, C_α	>	0.10
Lys60F, C_α	Ser195, C_α	>	-0.29
Ile174, C_α	Glu192, C_δ	<	0.11
Ile174, $C_{\gamma1}$	Cys191, C_α	>	-0.23
Ile174, $C_{\gamma1}$	Val213, C_β	<	-0.50
Ile174, $C_{\gamma1}$	Gly219, C_α	>	0.39
Asp189, C_α	Val213, C_α	<	1.79
Asp189, C_γ	Ser195, C_α	<	0.46
Asp189, C_γ	Trp215, C_α	>	0.42
Ala190, C_α	Trp215, $C_{\delta2}$	>	-0.60
Ala190, C_α	Gly216, C_α	<	-0.82
Cys191, S_γ	Trp215, $C_{\delta2}$	<	-0.68
Glu192, C_δ	Val213, C_α	<	-0.23
Glu192, C_δ	Gly216, C_α	>	-0.20
Glu192, C_δ	Gly219, C_α	>	0.25
Ser195, O_γ	Gly219, C_α	<	-0.62
Val213, C_α	Gly226, C_α	<	-0.57
Gly216, C_α	Gly219, C_α	<	-0.71

Ligand: DA2K	r^2=0.75	c_0=-1.39	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
Tyr60A, C_α	Trp60D, $C_{\delta2}$	<	0.92
Trp60D, C_α	Glu192, C_δ	<	-0.43

Trp60D, $C_{\delta 2}$	Val213, C_{α}	<	-1.05
Leu99, C_{α}	Gly216, C_{α}	>	-1.54
ILE174, C_{α}	Ala190, C_{α}	>	-1.14
Asp189, C_{γ}	Gly216, C_{α}	<	-2.08
Cys191, C_{α}	Glu192, C_{δ}	>	0.86
Glu192, C_{δ}	Ser195, C_{α}	<	-0.75
Glu192, C_{δ}	Ser195, C_{α}	>	0.85
Glu192, C_{δ}	Trp215, $C_{\delta 2}$	<	-1.18
Ser195, C_{α}	Gly226, C_{α}	>	0.75

Ligand: C24	r^2=0.72	c_0=-9.51	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
Lys60F, C_{α}	Glu192, C_{δ}	>	0.19
Leu99, C_{α}	Gly219, C_{α}	<	-0.31
Asp189, C_{γ}	Ser195, C_{α}	>	0.46
Ala190, C_{β}	Val213, C_{α}	>	-0.92
Cys191, S_{γ}	Ser195, C_{α}	>	0.44
Glu192, C_{δ}	Ser195, C_{α}	<	-0.48
Glu192, C_{δ}	Gly219, C_{α}	>	0.50
Ser195, O_{γ}	Trp215, $C_{\delta 2}$	<	-0.29
Val213, C_{α}	Gly219, C_{α}	<	-0.80

Ligand: G3X	r^2=0.50	c_0=-10.11	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
Asp72, C_{α}	Phe331, C_{α}	<	0.44
Asp72, C_{γ}	Phe331, C_{α}	<	-0.91
Asp72, C_{γ}	Tyr334, C_{ξ}	>	0.17
Trp84, C_{α}	Tyr121, C_{ξ}	<	-0.31
Trp84, C_{α}	Phe331, C_{α}	<	-1.14
Trp84, $C_{\delta 2}$	Gly441, C_{α}	>	0.89
Gly119, C_{α}	Tyr130, C_{α}	>	1.25
Gly119, C_{α}	Tyr130, C_{ξ}	>	-0.69
Gly119, C_{α}	Phe288, C_{α}	>	0.84

Tyr121, C_α	Glu199, C_α	<	-0.92
Tyr121, C_α	Glu199, C_α	>	-0.56
Tyr121, C_α	Phe290, C_γ	<	-1.45
Tyr121, C_α	Phe330, C_α	<	0.62
Tyr121, C_ξ	Phe288, C_α	<	0.85
Tyr121, C_ξ	Phe330, C_γ	<	-0.53
Ser122, C_α	Phe331, C_α	<	1.05
Ser122, O_γ	Glu199, C_α	<	1.12
Tyr130, C_α	Glu199, C_α	<	-0.80
Tyr130, C_α	Glu199, C_δ	>	-0.46
Tyr130, C_ξ	Ser200, O_γ	>	-0.21
Ser200, O_γ	His440, C_γ	>	0.34
Phe288, C_α	Phe290, C_α	>	-1.54
Phe288, C_α	Phe331, C_γ	<	-2.47
Phe290, C_γ	Tyr334, C_ξ	<	-0.11
Phe330, C_γ	Tyr334, C_α	<	-0.27
Tyr334, C_ξ	His440, C_γ	<	-0.23

Table S3. Results of multi-linear regression analysis for ligands BMZ, DA2K, C24 and G3X correlating $d_{ij}^{<}$ and $d_{ij}^{>}$ with scores of binding poses.

Ligand: BMZ	r^2=0.65	c_0=--8.03	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
His57, C_α	Tyr60A, C_ξ	>	0.43
His57, C_α	Trp215, $C_{\delta2}$	<	-0.88
His57, C_γ	Trp60D, $C_{\delta2}$	>	-0.20
His57, C_γ	Glu192, C_δ	<	-1.94
Tyr60A, C_ξ	Ala190, C_β	>	-0.53
Tyr60A, C_ξ	Cys191, C_α	>	0.38
Tyr60A, C_ξ	Val213, C_α	<	-0.37
Trp60D, $C_{\delta2}$	Ser195, C_α	>	0.19
Lys60F, C_α	Asp189, C_γ	>	0.21
Lys60F, C_α	Glu192, C_δ	>	0.10
Lys60F, C_α	Gly226, C_α	>	-0.22
Leu99, C_α	Asp189, C_α	>	0.52
ILE174, C_α	Glu192, C_δ	<	0.26
ILE174, $C_{\gamma1}$	Val213, C_β	<	-0.49
Asp189, C_α	Ala190, C_β	<	1.07
Asp189, C_γ	Val213, C_β	<	1.25
Ala190, C_α	Ser195, C_α	<	-0.83
Ala190, C_α	Gly216, C_α	<	-1.00
Ala190, C_β	Cys191, C_α	>	1.53
Ala190, C_β	Cys191, S_γ	>	2.99
Cys191, C_α	Val213, C_β	<	-1.63
Cys191, S_γ	Glu192, C_α	<	-1.25
Cys191, S_γ	Ser195, O_γ	<	0.56
Cys191, S_γ	Trp215, C_α	>	-0.74
Glu192, C_α	Gly219, C_α	>	0.34
Glu192, C_δ	Ser195, O_γ	<	0.62
Ser195, O_γ	Gly219, C_α	<	-0.96
Ser195, O_γ	Gly226, C_α	<	-0.65

Val213, C_α	Trp215, $C_{\delta 2}$	<	-0.76
Val213, C_β	Gly216, C_α	<	-0.69
Trp215, $C_{\delta 2}$	Gly216, C_α	>	0.52
Gly219, C_α	Gly226, C_α	<	-0.40

Ligand: DA2K	r^2=0.75	c_0=--9.16	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
His57, C_α	Glu192, C_α	>	0.61
Trp60D, $C_{\delta 2}$	Asp189, C_α	>	0.26
Lys60F, C_α	Val213, C_β	<	-0.83
Asp189, C_γ	Val213, C_α	<	-2.53
Glu192, C_α	Gly219, C_α	>	0.83

Ligand: C24	r^2=0.77	c_0=-9.67	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
His57, C_α	Lys60F, C_ε	>	0.18
Trp60D, C_α	Leu99, C_α	<	-1.76
Trp60D, C_α	Leu99, C_γ	>	-0.57
Trp60D, C_α	Ile174, $C_{\gamma 1}$	>	0.39
Trp60D, C_α	Gly219, C_α	>	0.36
Leu99, C_α	Asp189, C_γ	>	1.07
Ala190, C_α	Glu192, C_δ	<	-1.55
Ala190, C_β	Gly219, C_α	>	1.08
Cys191, C_α	Glu192, C_δ	<	1.23

Ligand: G3X	r^2=0.76	c_0=-10.63	
Atom i	Atom j	$d_{ij}^{>}$ or $d_{ij}^{<}$	$c_{ij}^{>/<}$
Asp72, C_α	Asp72, C_γ	<	5.79
Asp72, C_α	Ser122, C_α	<	0.43
Asp72, C_γ	Gly119, C_α	>	0.27
Asp72, C_γ	Tyr121, C_ξ	<	-0.55
Trp84, C_α	Trp84, $C_{\delta 2}$	<	1.64
Trp84, C_α	Gly119, C_α	>	0.34

Trp84, C_{α}	Tyr121, C_{ξ}	>	-1.10
Trp84, C_{α}	Tyr130, C_{ξ}	>	0.49
Trp84, C_{α}	Phe290, C_{α}	>	1.15
Trp84, C_{α}	Phe331, C_{α}	<	-0.98
Trp84, $C_{\delta 2}$	Glu199, C_{δ}	>	-0.75
Gly119, C_{α}	Glu199, C_{δ}	<	-0.48
Tyr121, C_{α}	Phe288, C_{α}	<	0.88
Tyr121, C_{α}	Phe330, C_{γ}	<	-0.50
Tyr121, C_{α}	Trp334, C_{α}	<	-0.34
Tyr121, C_{α}	His440, C_{α}	<	0.50
Tyr121, C_{ξ}	Ser200, O_{γ}	<	-0.45
Tyr121, C_{ξ}	His440, C_{γ}	<	-0.47
Tyr121, C_{ξ}	Gly441, C_{α}	>	-0.74
Ser122, C_{α}	Tyr130, C_{α}	>	0.30
Ser122, O_{γ}	Phe331, C_{α}	>	0.47
Tyr130, C_{α}	Phe290, C_{γ}	>	-0.65
Glu199, C_{α}	Phe290, C_{α}	<	-0.60
Glu199, C_{α}	Tyr334, C_{α}	<	-0.40
Ser200, O_{γ}	Phe330, C_{α}	<	-1.82
Phe288, C_{γ}	Phe331, C_{α}	<	0.66
Phe290, C_{α}	Phe290, C_{γ}	>	7.31
Phe330, C_{α}	Tyr334, C_{α}	<	-1.39
Phe330, C_{γ}	Tyr334, C_{ξ}	>	0.52
Phe330, C_{γ}	Gly441, C_{α}	>	0.37

Table S4. Predicted versus experimental binding affinity for 14 compounds binding to thrombin.

Ligand name or PDB code	Experimental ΔG [kcal/mol]	Predicted ΔG Ligand model [kcal/mol]	Deviation [kcal/mol]	Predicted ΔG Long apo [kcal/mol]	Deviation [kcal/mol]
1ypg	-10.90	-12.46	1.56	-12.75	1.86
1ype	-11.04	-10.18	0.86	-8.71	2.33
1ypj	-9.58	-9.94	0.35	-8.40	1.18
1oyt	-9.87	-8.34	1.53	-9.31	0.55
2cf8	-11.04	-11.35	0.31	-14.05	3.01
2cn0	-10.44	-11.41	0.97	-11.07	0.63
1vzq	-10.16	-10.25	0.09	-7.79	2.37
rac-8 from[2]	-7.10	-6.19	0.91	-5.20	1.90
rac-13a from [3]	-7.77	-9.32	1.56	-9.24	1.48
rac-16 from[4]	-6.86	-5.82	1.04	-7.87	1.01
rac-5 from[5]	-6.31	-7.85	1.54	-7.26	0.95
rac-8 from[5]	-8.25	-6.31	1.93	-6.57	1.67
rac-9 from[5]	-11.33	-10.74	0.59	-11.38	0.05
(+)-7 from[6]	-8.65	-9.14	0.49	-9.70	1.05

Reference List

1. Bowman, A. L.; Lerner, M. G.; Carlson, H. A. Protein flexibility and species specificity in structure-based drug discovery: dihydrofolate reductase as a test system. *J. Am. Chem. Soc.* **2007,** *129*, 3634-3640.

2. Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. Molecular recognition at the thrombin active site: Structure-based design and synthesis of potent and selective thrombin inhibitors and the x-ray crystal structures of two thrombin-inhibitor complexes. *Chemistry & Biology* **1997,** *4*, 287-295.

3. Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. Molecular recognition at the thrombin active site: Structure-based design and synthesis of potent and selective thrombin inhibitors and the x-ray crystal structures of two thrombin-inhibitor complexes. *Chemistry & Biology* **1997,** *4*, 287-295.

4. Obst, U.; Banner, D. W.; Weber, L.; Diederich, F. Molecular recognition at the thrombin active site: Structure-based design and synthesis of potent and selective thrombin inhibitors and the x-ray crystal structures of two thrombin-inhibitor complexes. *Chemistry & Biology* **1997,** *4*, 287-295.

5. Schweizer, E.; Hoffmann-Roder, A.; Olsen, J. A.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Banner, D. W.; Diederich, F. Multipolar interactions in the D pocket of thrombin: large differences between tricyclic imide and lactam inhibitors. *Organic & Biomolecular Chemistry* **2006,** *4*, 2364-2375.

6. Schweizer, E.; Hoffmann-Roder, A.; Scharer, K.; Olsen, J. A.; Fah, C.; Seiler, P.; Obst-Sander, U.; Wagner, B.; Kansy, M.; Diederich, F. A fluorine scan at the catalytic center of thrombin: C-F, C-OH, and C-OMe bioisosterism and fluorine effects on pK(a) and log(D) values. *Chemmedchem* **2006,** *1*, 611-621.