

Supplementary Materials

Application of a novel in silico high throughput screen to identify selective inhibitors for protein–protein interactions

Catherine Joce^a, Joshua A. Stahl^a, Mitesh Shridhar^b, Mark R. Hutchinson^c, Linda R. Watkins^b, Peter O. Fedichev^d, Hang Yin^{a,*}

^a*Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO, 80309, USA.*

^b*Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, Colorado, 80309, USA.*

^c*Discipline of Pharmacology, University of Adelaide, Adelaide, South Australia, Australia.*

^d*Quantum Pharmaceuticals Kosmonavta, Volkova 6A-606, Moscow, Russia*

Experimental Section

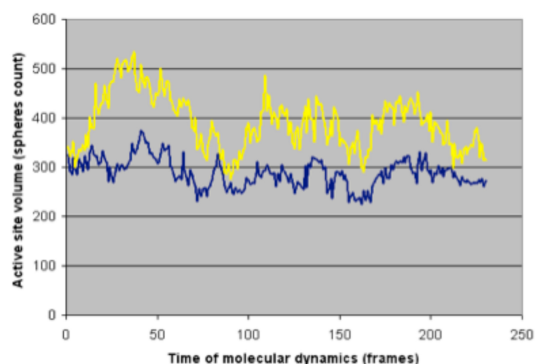
Virtual chemical libraries preparation: The structures of chemical ligands for virtual screening were taken from the 700,000-compound ENAMINE screening collections (enamine.net, Ukraine). To decrease the scale of the necessary computational work, the ENAMINE library was clustered using a combination of Jarvis–Patrick and Li algorithms.^[1-3]

The measure of dissimilarity (“distance”) between the molecules was determined by Tanimoto similarity calculated with Daylight fingerprints of the molecules, which brings together molecules with the largest common substructure.

The clustering parameters were chosen such that a reasonable amount of clusters (86k) was obtained. The compounds representing the cluster centroids were taken for subsequent docking. By the nature of the clustering procedure the cluster centroids normally represent the compounds of the least size and conformational complexity. Therefore using the cluster centroids does not only decrease the size of the compound library for screening, but also decreases the amount of computations required for the docking procedures. An additional filter that matches the molecular volume to the binding site volume (i.e. molecules that are too large to fit into the active site are removed) was also applied.

Docking: The docking procedure involves an extensive conformational search for the positions of the ligand within the target binding site and scoring of the obtained ligand conformations. To account for protein flexibility to a first degree of approximation, we took the protein structure obtained from the protein data bank (pdb) and performed repetitive MD runs of the receptor. These simulations were carried out for a sufficiently long time to obtain a number of the protein

structures with the largest possible volume of the binding pocket. A few top-opened structures were subsequently taken as rigid binding pocket models for docking runs. Top-opened structures were obtained as follows: Firstly, the pdb data was analyzed and missing atoms (mainly hydrogens and sometimes heavy atoms) were added. In the case of additional atoms causing high-energy conformations, the resulting structures were relaxed by performing a long MD run to adjust the conformation. The active site volume was measured every 100 steps (see the Supplementary Figure 1 below depicting two representative active site volumes on the surface of TLR4). After an initial relaxation, the volume of the active site starts oscillating, providing a number of “open” (high volume) and “closed” (low volume) conformations. The most open conformations were then collected and clustered, providing the “top-opened” structures that were used in the following docking simulations.

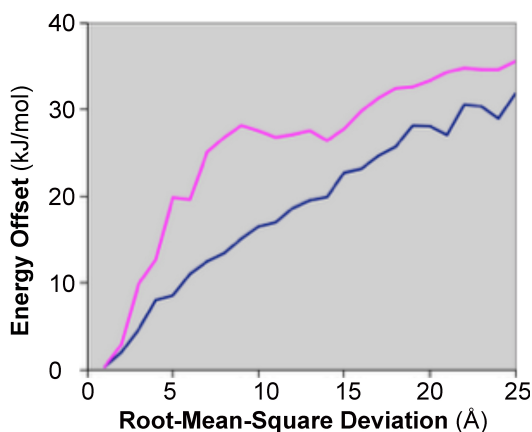


Supplementary Figure 1. Generation of the “top-opened” structures for active sites I (yellow) and II (blue) on the TLR4 surface.

Since the number of possible ligand positions is high, we first fixed the bond lengths and angles to only allow limited freedom for each ligand. The conformational search is provided by QUANTUM implementing a classical docking algorithm.^[4] The scoring of the ligand poses was obtained first by a knowledge-based free energy estimator^[5] at the initial stages of the docking selection and then, at the very end of the docking routine, by a Linear Response (LR) model for a better binding free energy estimation. In our practice, such a two-step routine renders optimal results to gain the computational speed. The calculations were performed using the Amazon EC2 cloud infrastructure. Last, we profiled a “fingerprint” of the identified hits by docking them to an original panel of proteins representing the whole human proteome (*ca.* 500 representative proteins selected from various protein families) and collected IC_{50} data on every protein/small-molecule and protein/protein complex out of this panel.

MD-based binding free energies calculation: The key to elimination of false binders (e.g. increased method selectivity) is a proper inclusion of hydration forces. The scores employed in the docking part of the calculations have been optimized to produce reasonable ligand positions and free energy estimation. To better estimate binding affinities at a high speed, we employed MD-based calculations with a recently developed implicit water model.^[6] Docking gives the positions of the molecules and a free energy estimation. MD-based free energy calculation takes the docked

position and improves the estimation of the free energy. Therefore each of methods ascribes a quantity, the estimate of the free energy, to a certain position of the ligand in the active site.



Supplementary Figure 2. The “docking funnels” (i.e. the energy difference between the conformers plotted as a function of r.m.s.d. from the known crystallographic position of a ligand). Both the free energies obtained from a MD-run (magenta line) and from the docking scoring function (blue line) were used to calculate the energies of thousands of conformers generated by the docking run for more than 300 protein-ligand complexes with known position of the ligand

(corresponding to r.m.s.d. displacement θ). For each of the crystallographic positions the free energy estimate (F_{true}) was calculated. Then, we started a docking run to search for the local energy minima. For each of the local minima (the docking conformer) we took the position and calculated the free energy estimations $F_{\text{calculated}}$. We plotted the energy differences, ($F_{\text{calculated}} - F_{\text{true}}$) as a function of r.m.s.d. of the conformer relative to the crystallographic position. A total of more than 50k conformers were used to obtain the distribution. The resulting energies were averaged over the conformers with similar r.m.s.d. values and plotted on the same graph.

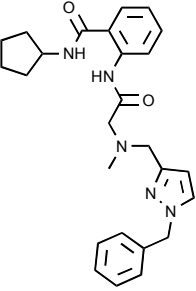
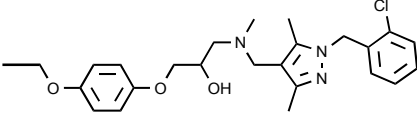
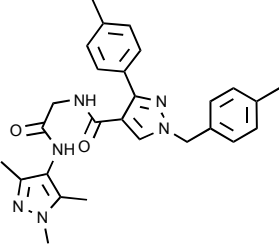
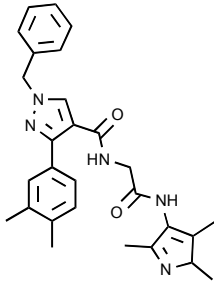
The *selectivity improvement* provided by the combined usage of a simple scoring and a more sophisticated subsequent MD-simulation is illustrated and quantified in Supplementary Figure 2. A good free energy estimate is expected to correlate well with the position: the “true” pose of a ligand should give the minimum value of the free energy estimate. Supplementary Figure 2 shows that the MD-based free energy estimator enhances the “free energy” differences as r.m.s.d. values increases, thereby selecting the crystallographic positions better. This calculation shows that the MD-simulations give free energies estimates with a much steeper docking funnel, i.e. more sensitivity to misplacements of a ligand, than a simple scoring function. Therefore, this MD-based calculation can indeed be used to distinguish similar binding modes and hence obtain much more accurate docking positions. As a result, the combined procedure should produce a lower ratio of false positives and false negatives as compared with a scoring function based method.

Real-time microscopy of TLR4 signaling in RAW264.7 mouse macrophage cell line: TLR4 signaling leads to the simultaneous activation of Akt1.^[7] Given the availability of a RAW264.7 mouse macrophage cell line stably

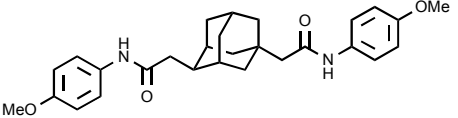
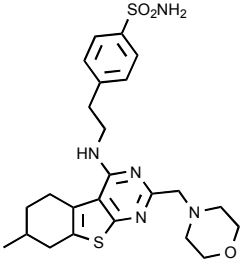
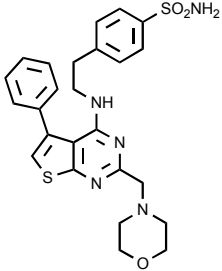
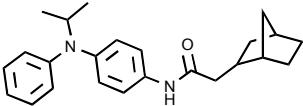
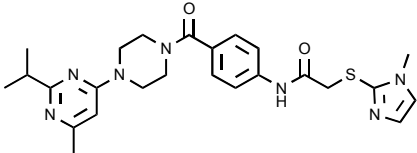
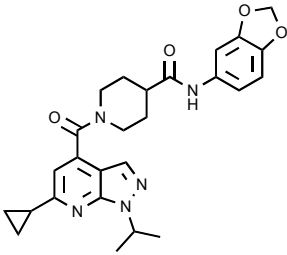
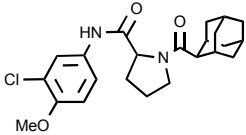
transfected to express green fluorescent protein (GFP)-tagged Akt1,^[8] mobilization and cytosolic clearance of GFP-Akt1 was used as an indicator of TLR4 activation. LPS (*Escherichia coli*; Serotype: 0111:B4) was obtained from Sigma (St. Louis, MO, USA). Compounds **1** and **2** were obtained from Enamine, (Kiev, Ukraine). Cells were grown in growth media supplemented with 10% FBS, 10 × penicillin/streptomycin and 10 × L-glutamine and then were plated at a density of 2×10^5 cells/ml in growth media on 35 mm MatTek Glass Bottom Dishes (Ashland, MA, USA) for 18 h prior to imaging. Just prior to imaging the growth media was removed from the plates, washed twice with 1 ml Hank's buffered saline solution (HBSS) supplemented with 25 mM HEPES buffered to pH 7.4 and replaced with 1 ml warmed conditioned imaging Hanks Buffer media (media was conditioned by a 24 h incubation with RAW264.7 cells). Imaging was carried out on a Nikon inverted microscope (Melville, NY, USA) with a 60× oil immersion objective, GFP/RFP dichroic mirror with corresponding single band excitation and emission filters (Chroma Technology, VT, USA) and CoolSNAP ES camera (Photometrics, Tucson, AZ, USA). A mercury lamp provided excitation. Images were captured every 7.5 s. Baseline fluorescence was captured for 5 frames, following which vehicle or antagonist was added in 200 µl. Imaging continued for a further 20 frames at which time LPS or test agonist (200 µl) was applied and monitored for a further 20 frames. If no visual response was obtained C5a (200 µl) was added to the plates to confirm if the cells were responsive. GFP-Akt1 cytosolic clearance was quantified using ImageJ and expressed as a normalized change in cytoplasmic fluorescence over time.

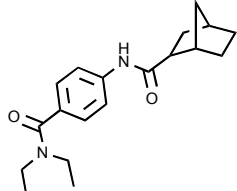
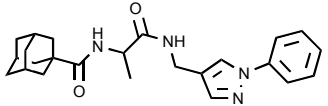
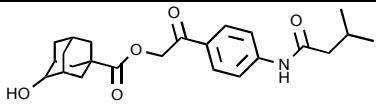
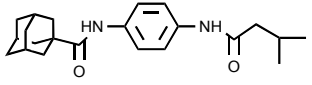
Nitric oxide cell TLR selectivity assay: RAW264.7 cells were grown in DMEM supplemented with 10% FBS, penicillin (100 U/ml), streptomycin (100 mg/ml) and L-glutamine (2 mM). RAW cells were then planted in 96-well plates at 100,000 cells per well and grown for 24 h in the media described previously. After 24 h media was removed and replaced with Macrophage-SFM (Invitrogen, CA, USA). Lanes were doped with the appropriate TLR specific ligands: LPS, poly(I:C), FSL-1, R848, and Pam₃CSK₄ were used to selectively activate TLR4, TLR3, TLR2/6, TLR7 and TLR2/1 respectively. Two lanes for each ligand were prepared: one containing ligand only and the other with the ligand and 300 nM **1**. Plates were then incubated for 24 h. Following incubation 100 µL of media was removed and added to flat black 96-well microfluor plates (Thermo scientific, MA, USA). 10 µL of 2,3-diaminonaphthalene (0.05 mg/ml in 0.62 M HCl) was added to each well and incubated for 15 min then quenched by addition of 3 M NaOH (5 µL). The plate was read on Beckman Coulter DTX880 reader (Beckman Coulter, CA, USA) with excitation at 365 nm and emission at 450 nm. Nitrite (a stable metabolite of nitric oxide) concentration was determined from a nitrite standard curve.

Supplementary Table 1: Identified TLR4 Inhibitors.

Entry	Structure	Compound Identifier
1	 <p>Chemical structure of T5879527: A cyclopentane ring is attached to a benzamide group. This benzamide is linked via its nitrogen to a methylene group, which is further connected to a nitrogen atom. This nitrogen atom is bonded to a methyl group and a methylene group that connects to a 1,2,4-triazole ring. The triazole ring is substituted with a benzyl group.</p>	T5879527
2	 <p>Chemical structure of T5342126: A central carbon atom is bonded to a hydroxyl group, a methyl group, and a methylene group. The methylene group is connected to a nitrogen atom, which is further bonded to a methyl group and a methylene group. This methylene group connects to a 1,2,4-triazole ring. The triazole ring is substituted with a 4-chlorophenylmethyl group.</p>	T5342126
3	 <p>Chemical structure of T5661863: A central carbon atom is bonded to a methyl group, a methylene group, and a nitrogen atom. The nitrogen atom is part of a 1,2,4-triazole ring. The triazole ring is substituted with a 4-methylphenylmethyl group and a methylene group. The methylene group connects to another 1,2,4-triazole ring, which is substituted with two methyl groups.</p>	T5661863
4	 <p>Chemical structure of T5797938: A central carbon atom is bonded to a methyl group, a methylene group, and a nitrogen atom. The nitrogen atom is part of a 1,2,4-triazole ring. The triazole ring is substituted with a 4-methylphenylmethyl group and a methylene group. The methylene group connects to another 1,2,4-triazole ring, which is substituted with two methyl groups.</p>	T5797938

Supplementary Table 2: Identified MD-2 Inhibitors.

Entry	Structure	Compound Identifier
1		T0502-9420
2		T0520-5396
3		T5223761
4		T5246154
5		T5625206
6		T5627443
7		T5926519

8		T6052675
9		T6056555
10		T6071187
11		T6074019

Supplementary References:

1. Li, W. *J. Chem. Inf. Model.* **2006**, *46*, 1919.
2. Willett, P. *Similarity and Clustering in Chemical Information Systems*, **1987**.
3. Jarvis, R. A.; Patrick, E. A. *IEEE Transactions on Computers* **1973**, *22*, 1025.
4. Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1.
5. Muegge, R.; Rarey, M. *Rev. Comput. Chem.* **2001**, *17*, 69.
6. Fedichev, P. O.; Men'shikov, L. I. *J. Struct. Chem.* **2009**, *50*, 97.
7. Dauphinee, S. M.; Karsan, A. *Lab. Invest.* **2006**, *86*, 9.
8. Evans, J. H.; Falke, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16176.