Supporting Information

Template selection. hPepT1 is a member of the proton-dependent oligopeptide transporter (POT) family, which includes 16 atomic structures in the PDB. To identify the best templates for modeling hPepT1, we used the HHpred server,[1] as well as analyzed the hPepT1 entry (2.A.17.4.9) in the Transporter Classification Database (TCDB)[2] and the POT family structures Orientations of Membrane Protein (OPM) database.[3] The most suitable templates were prioritized based on the following considerations: (i) sequence similarity to hPepT1 (i) uniqueness of the conformation including whether it binds a ligand, and (iii) quality of the x-ray structure. Therefore five structures of three different proteins ($PepT_{So}$, GkPOT, and $PepT_{St}$), sharing sequence identity of 22%-34% with hPepT1, were selected as modeling templates.[4-9] In particular, the two structures in occluded and unbound inward-open conformations (PDB IDs 2XUT and 4APS respectively) were solved at low resolution (3.62 Å and 3.3 Å respectively). It is important to note that such resolution is not ideal for structure based drug discovery studies. However, since our goal is to sample intermediate conformations of the transport cycle, we chose to include them in our study. 2XUT in particular is the only structure available in an occluded conformation and we hypothesized that could increase our chances to discover substrates. This hypothesis was later confirmed since aspartame, shown to be a substrate by experimental tests, was selected from the screening against 2XUT (Table S1).

PepT1-template alignment. The initial alignment between hPepT1, $PepT_{So}$, GkPOT, and $PepT_{St}$ was obtained using the Promals3D server,[10] and was subsequently refined by based on previously published alignments of SLC15 family members.[11] For example, $PepT_{So}$ consists of 14 TMs with two additional helices, HA and HB, that are localized in periphery of the transporter far away from the binding site. Thus, HA and HB were removed from the alignment, as well as three loops distant from the binding site (i.e., loops H3-H4, H5-H6, and H9-H10).

Homology modeling. We used MODELLER-9-v12 and 9-v14[12] with the 'automodel' class of to build 100 initial homology models (Table 1; Models 2-4), which were ranked using the statistical potential Z-DOPE.[13] Due to the lower quality of the occluded conformation template (PDB: 2XUT), we generated 500 models to sample more conformations (Table 1; Model 1). The inward-open conformation models (Table 1; Models 3, 4 and 5) contained non-protein atoms from the substrate alafosfalin, Ala-Phe and Ala-Ala-Ala derived from the corresponding coordinates in the templates. The Z-DOPE score of ranged between -0.18 and -0.049 for the inward-open conformations (Models 2-4). The Z-DOPE score of the occluded model (Model 1) was +0.12; however, the corresponding template structure exhibited similar Z-DOPE score (i.e., +0.024), which likely resulted from a lower resolution structure (i.e., 3.6 Å), and the model is unlikely to score better than the template it is based on. It is worth noting that the enrichment of known ligands among decoy compounds that we previously used to assess homology models for structure-based virtual screening[14-16] did not give satisfying results for this study. The enrichment values were not satisfactory likely because of two main reasons: First, hPepT1's ligands are particularly challenging for molecular docking. hPepT1 ligands are polar and flexible, and they form interactions with water molecules as well as with the binding site residues. Second, as was shown in multiple studies of hPepT1 homologs, the binding site of hPepT1 is likely to be highly flexible, adopting different conformations depending on the ligands it binds. Therefore, during enrichment calculations, ligands are docked against a single conformation that binds only a small subset of the ligands; however, the entire set of ligands is not expected to score well compared to non-ligands, leading to poor enrichment scores.

Selecting relevant docking programs. We used different freely available docking methods for the different virtual screenings. The usage of the methods has evolved throughout the project based on the results of each screen. Our main considerations were: (i) Speed and efficiency. For example, FRED is significantly faster than AutoDock Vina and is thus, more appropriate for screening large datasets such as ZINC lead-like library; (ii) Flexibility in incorporating constraints. We have generated hypotheses regarding the mechanism of binding and thus, added interaction constraints (with R27

and Y31) in the docking protocol. The incorporation of constraints was easier in FRED. (iii) The relevance of the computational method. It has been previously suggested that different methods can be more accurate for different targets. Our docking studies on related targets (e.g., amino acid transporters[15, 17]) suggested that FRED accurately captures interactions involving chemically similar ligands (i.e., di- and tri-peptides vs. amino acids) with transporters.

Docking with AutoDock Vina. Autodock Vina[18] was used for the preliminary screenings of the FDA-approved drug library (6,719 molecules) against the occluded and inward-open Models 1 and 3, respectively. The box enclosing the binding site of the inward-facing model (Model 3) was generated based on the coordinates of the ligand alafosfalin in the template structure. For Model 1 we first structurally aligned to the ligand-bound models and the box enclosing the binding site was derived from the aligned ligands.

We hypothesized that the occluded conformation and the alafosfalin-bound structure would increase our chances to identify substrate-like compounds. However, because Model 1 was based on a template structure of the lowest quality, we considered only compounds predicted to interact with R27 and Y31 in both conformations for visual inspection (i.e., total of 276 compounds). Within this subset, known transported drugs such as Cephadroxil were identified, which increased our confidence in our strategy. We selected three compounds for experimental testing from this subset (Table S1).

Docking with FRED. OpenEye FRED[20] was used for the following screenings against Models 3, and 5 to screen larger libraries. The models were prepared with the MAKE_RECEPTOR utility of FRED. The box enclosing each binding site was generated based on the coordinates of the corresponding ligand in the template structures (alafosfalin, Ala-Ala-Ala for Model 3 and Model 5, respectively). Docking was run using hydrogen bonds constraints on the hydroxyl group of Y31. The docking poses were ranked by the Chemgauss4 scoring function, which is defined by smoothed Gaussian potentials describing the complementarity (by shape and chemical properties) between the ligands and the binding site. The ZINC Leads-Now set[21] that contained 2,268,809 compounds of molecular weights between 250 and 350 Dalton, 7 or fewer rotatable bonds, and xlogP value of 3.5 or lower was screened against the two models of hPepT1. The top scoring ligands of each screen were visualized.

Chemical Similarity Calculations. The chemical novelty of the hits was evaluated with the Tanimoto coefficient (Tc) calculated relying on Daylight fingerprints, compared to a known ligands from the literature.[22, 23] Tc values of < 0.5 suggest that the molecule is a chemically novel hPepT1 ligand. Specifically, the Tc of compounds 3, 7 10 and 39 were Tc of 0.4, 0.43, 0.38 and 0.4 respectively.

Docking with QM-Polarized Ligand Docking (QPLD). The five newly confirmed hPepT1 ligands (Fig. 3*B*) and four known ligands (i.e., the di-peptide Ala-Phe, the tri-peptide Ala-Ala-Ala, valacyclovir, and cephalexin) were re-docked using the QM-Polarized Ligand Docking implemented in the Schrödinger suite[24] against Models 4 and 5. First, the ligands were prepared with the LigPrep.[25] The receptor was prepared with the Protein Preparation Wizard implemented in Maestro and an implicit membrane was then set up with Prime. The docking grid of the receptor was generated with Glide, [26, 27] with the box enclosing the coordinates of the ligands bound to the prokaryotic templates used to model hPepT1. The docking calculations were performed using hydrogen bonds constraints on the hydroxyl group of Y31. After the initial Glide docking, performed in Single Precision mode, the partial charges of the docked ligands were calculated with Jaguar. Finally, the ligands with the new charges were redocked in Single Precision.

Free energy estimation with Molecular Mechanics Generalized Born Surface Area solvation We estimated the binding energies of the four new inhibitors with the MMGBSA method, implemented in the Schrödinger suite. We docked the new hits as well as known ligands (i.e. Ala-Phe, Ala-Ala-Ala, valacyclovir, and cephalexin) in two ligand-bound x-ray structures of $PepT_{St}$ in inward open

conformations (PDB IDs 4D2C and 4D2D) and the two models of hPepT1 based on those templates (Models 4 and 5 in Table 1).

Supplemental Figures

Figure S1

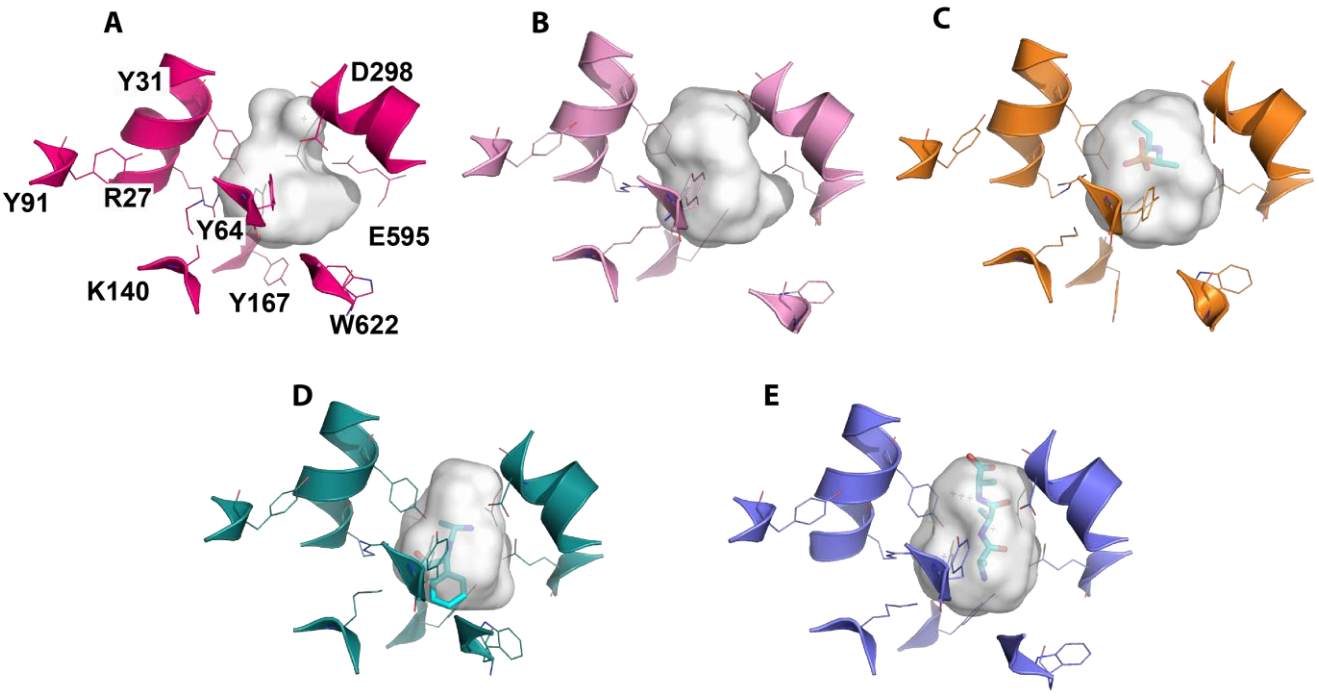

Fig. S1. Binding sites of the hPepT1 models in various conformations. The binding sites of each model are represented in white surface. The various conformation include: an occluded conformation (Model 1, Table 1) (*A*), an unbound inward-open conformation (Model 2) (*B*), three inward-open conformations bound to alafosfalin (Model 3), (*C*) and the di-peptides Ala-Phe (Model 4) (*D*), and Ala-Ala-Ala (Model 5) (*E*). The ligands are shown in sticks and the key residues interacting with the ligands are shown in lines.

Figure S2

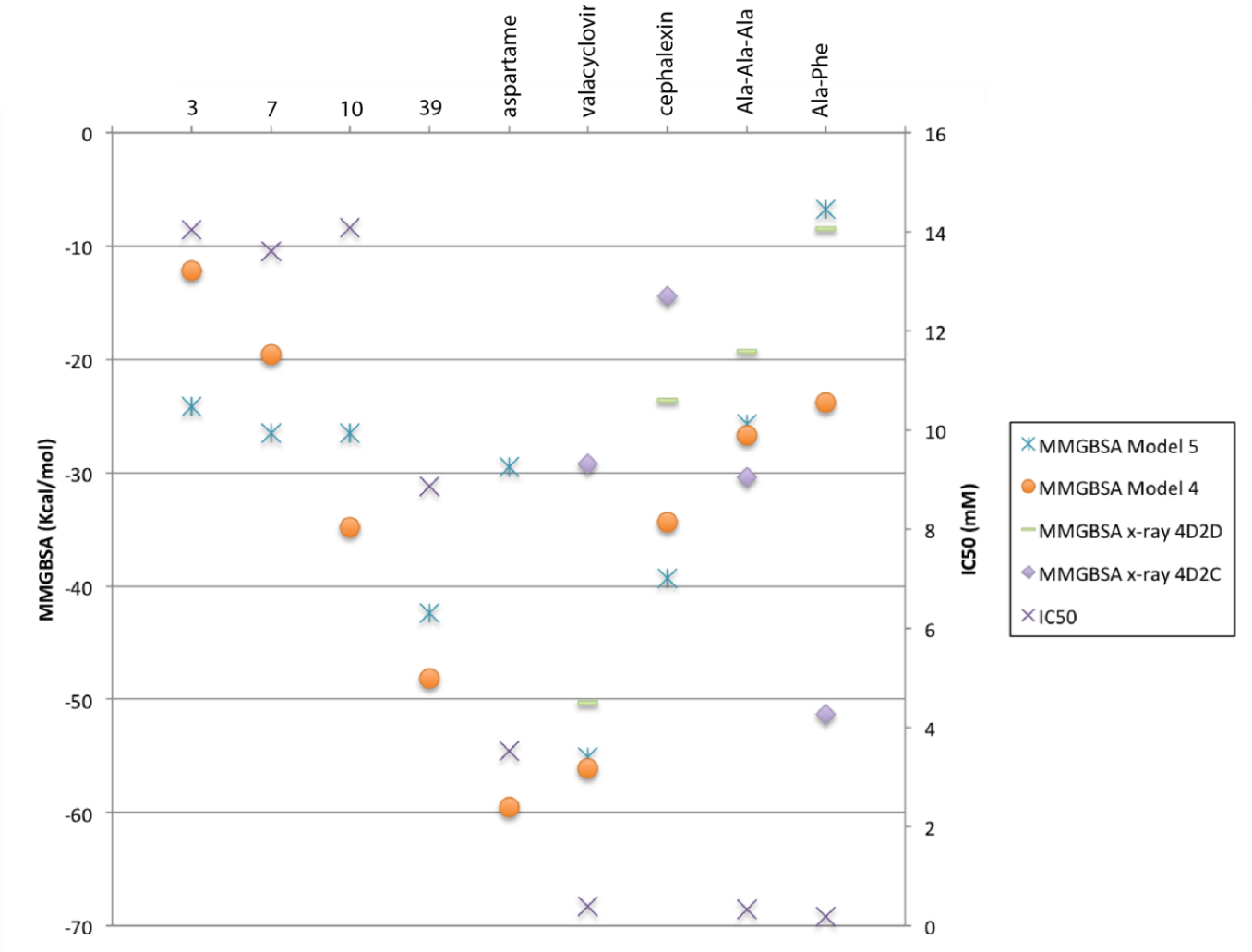

Fig. S2. MMGBSA calculations and experimental IC_{50}s values. Binding energies of our five hits and four known ligands predicted by MMGBSA calculations (Kcal/mol) in two inward-open models (Models 4 and 5) and their corresponding templates, and experimental IC_{50} values (mM).

Supplemental Table

Table S1

Compound name[a]	Rank[b]	Library[c]	Docking Program[d]	Model[e]
Aspartame	238 1420	FDA	AutoDock Vina	1 3
Betamipron	270 1717	FDA	AutoDock Vina	1 3
Mimosine	3649 1979	FDA	AutoDock Vina	1 3
ZINC71817303	16	Lead-like	OpenEye FRED	3
ZINC65585007	232	Lead-like	OpenEye FRED	3
ZINC71817322	2	Lead-like	OpenEye FRED	3
ZINC22242499	8	Lead-like	OpenEye FRED	3
ZINC72271520	23	Lead-like	OpenEye FRED	3
ZINC00143440	26	Lead-like	OpenEye FRED	3
ZINC30766047	210	Lead-like	OpenEye FRED	3
ZINC00205111	15	Lead-like	OpenEye FRED	3
ZINC0097001624	188	Lead-like	OpenEye FRED	5
ZINC0084673463	132	Lead-like	OpenEye FRED	5
ZINC0071281267	218	Lead-like	OpenEye FRED	5
ZINC0064969323	291	Lead-like	OpenEye FRED	5
ZINC0012296468	8	Lead-like	OpenEye FRED	5
ZINC00146165	43	Lead-like	OpenEye FRED	5
ZINC0072307150	5	Lead-like	OpenEye FRED	5
3	3	Lead-like	OpenEye FRED	5
7	7	Lead-like	OpenEye FRED	5
10	10	Lead-like	OpenEye FRED	5
39	39	Lead-like	OpenEye FRED	5

Table S1 : Compounds experimentally tested

[a]*Compound name* refers to the name or ID of the compounds tested
[b]*Rank* defines the rank of the compound in the screen
[c]*Library* refers to the virtual library screened; both the FDA and Lead-like libraries were downloaded from the ZINC database
[d]*Docking program* refers to the program used to conduct the virtual screening

[e]*Model* refers to the Model used to conduct the virtual screening

References

1. Soding, J.; Biegert, A.; Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* **2005,** *33*, (Web Server issue), W244-8.
2. Saier, M. H., Jr.; Yen, M. R.; Noto, K.; Tamang, D. G.; Elkan, C. The Transporter Classification Database: recent advances. *Nucleic acids research* **2009,** *37*, (Database issue), D274-8.
3. Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research* **2012,** *40*, (Database issue), D370-6.
4. Doki, S.; Kato, H. E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H.; Newstead, S.; Ishitani, R.; Nureki, O. Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, (28), 11343-8.
5. Guettou, F.; Quistgaard, E. M.; Tresaugues, L.; Moberg, P.; Jegerschold, C.; Zhu, L.; Jong, A. J.; Nordlund, P.; Low, C. Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO reports* **2013,** *14*, (9), 804-10.
6. Newstead, S.; Drew, D.; Cameron, A. D.; Postis, V. L.; Xia, X.; Fowler, P. W.; Ingram, J. C.; Carpenter, E. P.; Sansom, M. S.; McPherson, M. J.; Baldwin, S. A.; Iwata, S. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO journal* **2011,** *30*, (2), 417-26.
7. Solcan, N.; Kwok, J.; Fowler, P. W.; Cameron, A. D.; Drew, D.; Iwata, S.; Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO journal* **2012,** *31*, (16), 3411-21.
8. Lyons, J. A.; Parker, J. L.; Solcan, N.; Brinth, A.; Li, D.; Shah, S. T.; Caffrey, M.; Newstead, S. Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO reports* **2014,** *15*, (8), 886-93.
9. Solcan, N.; Kwok, J.; Fowler, P. W.; Cameron, A. D.; Drew, D.; Iwata, S.; Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *The EMBO journal* **2012,** *31*, (16), 3411-21.
10. Pei, J.; Kim, B. H.; Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research* **2008,** *36*, (7), 2295-300.
11. Newstead, S.; Drew, D.; Cameron, A. D.; Postis, V. L.; Xia, X.; Fowler, P. W.; Ingram, J. C.; Carpenter, E. P.; Sansom, M. S.; McPherson, M. J.; Baldwin, S. A.; Iwata, S. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO Journal* **2011,** *30*, (2), 417-26.
12. Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **1993,** *234*, (3), 779-815.
13. Shen, M. Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **2006,** *15*, (11), 2507-24.
14. Colas, C.; Ung, P. M.; Schlessinger, A. SLC Transporters: Structure, Function, and Drug Discovery. *MedChemComm* **2016,** *7*, (6), 1069-1081.
15. Colas, C.; Grewer, C.; Otte, N. J.; Gameiro, A.; Albers, T.; Singh, K.; Shere, H.; Bonomi, M.; Holst, J.; Schlessinger, A. Ligand Discovery for the Alanine-Serine-Cysteine Transporter (ASCT2, SLC1A5) from Homology Modeling and Virtual Screening. *PLoS computational biology* **2015,** *11*, (10), e1004477.
16. Geier, E. G.; Schlessinger, A.; Fan, H.; Gable, J. E.; Irwin, J. J.; Sali, A.; Giacomini, K. M. Structure-based ligand discovery for the Large-neutral Amino Acid Transporter 1, LAT-1. *Proceedings of the National Academy of Sciences of the United States of America* **2013,** *110*, (14), 5480-5.

17. Augustyn, E.; Finke, K.; Zur, A. A.; Hansen, L.; Heeren, N.; Chien, H. C.; Lin, L.; Giacomini, K. M.; Colas, C.; Schlessinger, A.; Thomas, A. A. LAT-1 activity of meta-substituted phenylalanine and tyrosine analogs. *Bioorganic & medicinal chemistry letters* **2016,** *26*, (11), 2616-2621.
18. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* **2010,** *31*, (2), 455-61.
19. Ekins, S.; Johnston, J. S.; Bahadduri, P.; D'Souza, V. M.; Ray, A.; Chang, C.; Swaan, P. W. In vitro and pharmacophore-based discovery of novel hPEPT1 inhibitors. *Pharmaceutical research* **2005,** *22*, (4), 512-7.
20. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011,** *51*, (3), 578-596.
21. Carr, R. A.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug discovery today* **2005,** *10*, (14), 987-92.
22. Daniel, H.; Kottra, G. The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflugers Archiv : European journal of physiology* **2004,** *447*, (5), 610-8.
23. Smith, D. E.; Clemencon, B.; Hediger, M. A. Proton-coupled oligopeptide transporter family SLC15: Physiological, pharmacological and pathological implications. *Molecular aspects of medicine* **2013,** *34*, (2-3), 323-36.
24. Colas, C.; Schlessinger, A.; Pajor, A. M. Mapping functionally important residues in the Na+/dicarboxylate cotransporter, NaDC1. *Biochemistry* **2017**.
25. Schrödinger, LLC. LigPrep. **2017**.
26. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry* **2004,** *47*, (7), 1750-9.
27. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* **2004,** *47*, (7), 1739-49.