Supporting Information

Origins of the Activity of PAL and LAP Enzyme Inhibitors: Towards Ab Initio Binding Affinity Prediction

Edyta Dyguda, Jolanta Grembecka, W. Andrzej Sokalski, and Jerzy Leszczyński.

Department of Chemistry,Wrocław University of Technology Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

sokalski@pwr.wroc.pl

Methods

The structure of PAL from potato, *Solanum tuberosum* (record P31425 in Swiss-Prot database[1]), was modelled on the basis of homology to *Pseudomonas putida* histidine ammonia-lyase, HAL (PDB access code 1B8F) with the MODELLER program.[2] Six PAL inhibitors[3] (Figure S1) considered in a present study were as follows: 1) (*S*)-phenylalanine with an α-amino group replaced by an aminooxy moiety, namely (*S*)-2-aminooxy-3-phenylpropanoic acid (referred in the following as PheONH_3_S); 2) (*R*)-1-amino-2-phenylethylphosphonic acid (phosphonic analogue of phenylalanine, PheP_R) and 3) its enantiomer (PheP_S); 4) PheP_R with an O2 oxygen atom (designation as in Figure 1 in a corresponding paper) substituted by a hydrogen atom (PhePH_R); 5) PheP_R with O3 replaced by a methyl group (PhePCH_3_R); and 6) PheP_S with an additional methyl group at the Cα position (PhePCH_3a_S). All six structures were modelled in their zwitterionic forms (Figure S1). The PAL inhibitors optimized at the HF/6-31G* level were docked to the enzyme binding pocket, and the entire protein-ligand models soaked in a 5Å thick layer of water molecules were further refined by applying the Discover module of InsightII program.[4] The optimization of protein-ligand complexes was performed by means of the CVFF force field, after adding the corresponding partial charges to all atoms. The main chain atoms of PAL remained frozen, while the protein side chains, ligand and water molecules were free during the optimization procedure. The conjugate gradient algorithm was applied for minimization, which was continued until the RMS derivative was lower than 0.02 kcal/mol *Å. A twin cut-off (17.0 Å, 20.0 Å) was used to calculate the non-bonded interactions at every minimization step. The resultant complexes were employed for calculation of ligand binding energies with CVFF force field.[4] Since the N- and C-termini of polypeptide chain (residues 1–77 and 535–720, respectively) do not participate in the formation of the binding site, they were not included in our modeling procedure. Consequently, enumeration of amino acid residues within the truncated sequence has been changed (the first residue corresponds to the 78th residue of the original PAL sequence). It is remarkable, that our model of PAL active site is in general agreement with recently published the first

X-ray structure of PAL from *Rhodosporidium toruloides*.[5] According to the multiple sequence alignment of phenylalanine ammonia-lyases from different sources, all the amino acid residues of our model are strictly conserved and the three of them (Asn187^{A}, Tyr278^{B}, and Arg^{+}281^{B}) correspond to the residues from the known PAL structure (namely Asn270^{B}, Tyr363^{D}, and Arg^{+}281^{D}), that have been suggested to serve an important function in the substrate binding and subsequent catalysis step.[5]

Six active site residues positioned in the vicinity of a variable part of the inhibitors were subsequently selected to mimic the enzyme environment. The two asparagine (Asn187^{A}, Asn311^{A}; superscripts indicate the corresponding monomer of the PAL homotetramer) and glutamine (Gln275^{B}) residues were represented by the acetamide molecule, arginine (Arg^{+}281^{B}) by methylguanidinium cation, whereas tyrosine residues (Tyr35^{A}, Tyr278^{B}) were modelled as *p*-cresol (Figure S2).

In our subsequent computational approach we assume, that some meaningful conclusions regarding the ligands binding affinity can be drawn on the basis of the stabilization energy analysis. Direct calculation of binding free energy can be attempted, for example, by the free energy perturbation (FEP) method. This can be done mainly by empirical force-field based methods[6] which prevents deeper analysis of the interactions involved. On the other hand, when similar ligands are considered, the relative interaction energy emerges as the decisive factor in the discrimination of particular guest molecules,[7] whereas entropic and solvation contributions can be neglected.

Interaction energy decomposition was performed according to the direct variational-perturbational scheme[8] implemented in the modified version of GAMESS(US).[9] The 6-31G* basis set as well as the full counterpoise correction[10] were applied at each level of theory.

REFERENCES

(1) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M. C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. *Nucleic Acids Res.* **2003**, *31*, 365.

(2) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.

(3) a) Maier, L.; Diel, P. J. *Phosphorus Sulfur* **1994**, *90*, 259; b) Langer, B.; Langer, M.; Retey, J. *Adv. Protein Chem.* **2001**, *58*, 175.

(4) Accelrys Inc., **2000,** Insight II 2000 and Discover version 2.9, San Diego.

(5) Calabrese, J. C.; Jordan, D. B.; Boodhoo, A.; Sariaslani, S.; Vannelli, T. *Biochemistry* **2004**, *43*, 11403.

(6) Gohlke, H.; Klebe, G. *Angew. Chem. Int. Ed.* **2002**, *41*, 2644.

(7) Bohm, H. J.; Klebe, G. *Angew. Chem. Int. Ed.* **1996**, *35*, 2588.

(8) Sokalski, W. A.; Roszak, S.; Pecul, K. *Chem. Phys. Lett.* **1988**, *153*, 153.

(9) Schmidt, M. S.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(10) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

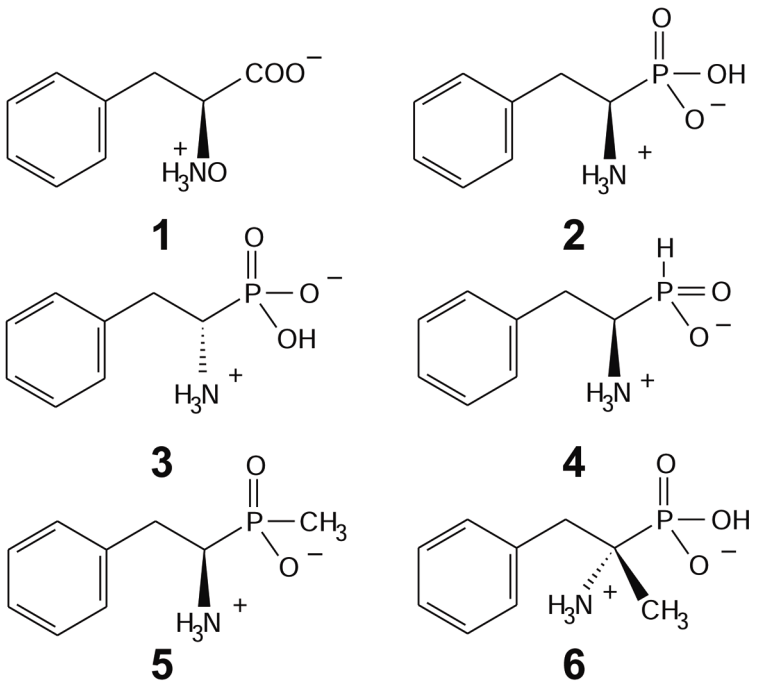

Figure S1. The structures of PAL inhibitors in the order of decreasing inhibitory activity.

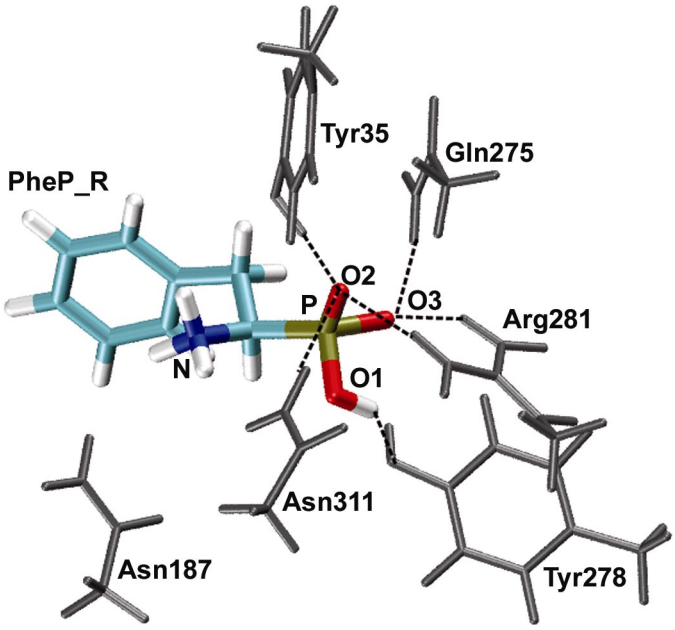

Figure S2. PAL active site model (residues shown in black) and the mode of PheP_R inhibitor binding. The closest atomic contacts with PAL residues are marked as dashed lines.

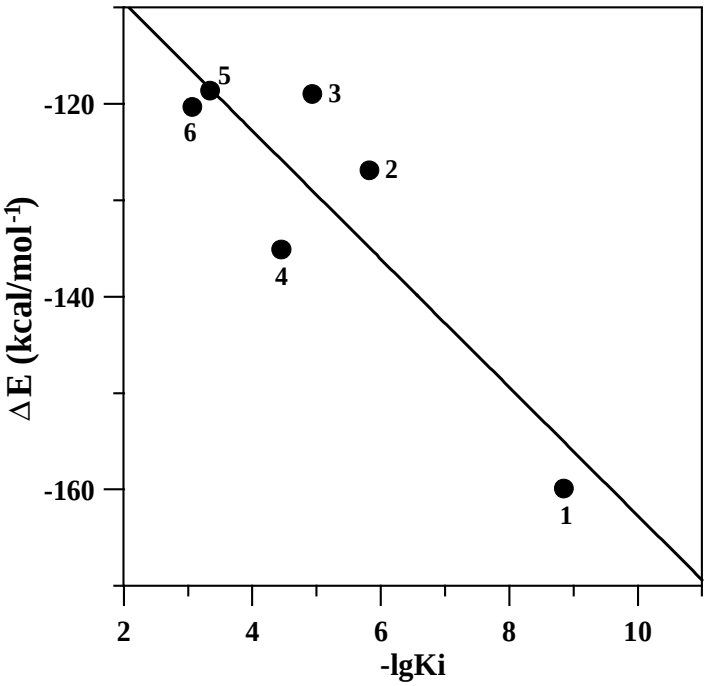

Figure S3. CVFF force field binding energy of PAL inhibitors, calculated using complete PAL molecule, as a function of inhibitory activity (correlation coefficient, R = 0.88).

Table S1. Components of the total interaction energy calculated as the sum of all two-body (inhibitor – active site residue) interactions (A) and, additionally, for the model including four of initial six binding site constituents, namely $Tyr35^{A}$, $Asn187^{A}$, $Tyr278^{B}$, and $Arg^{+}281^{B}$ (B).

Inhibitor	Model	Components of the total interaction energy [kcal mol^{-1}]						
		$E^{(1)}_{EL}$	$E^{(1)}_{EX}$	$E^{(1)}$	$E^{(R)}_{DEL}$	E_{SCF}	$E^{(R)}_{CORR}$	E_{MP2}
$PheONH_3$_S (1)	**A**	-154.907	105.216	-49.691	-46.054	-95.745	-4.727	-100.472
	B	-132.121	87.161	-44.960	-37.757	-82.718	-3.576	-86.293
PheP_R (2)	**A**	-129.762	94.299	-35.463	-44.092	-79.554	-4.019	-83.573
	B	-112.534	80.155	-32.379	-37.970	-70.349	-3.041	-73.390
PheP_S (3)	**A**	-129.774	96.340	-33.434	-42.108	-75.542	-2.970	-78.512
	B	-108.106	72.223	-35.884	-34.949	-70.832	-1.456	-72.289
PhePH_R (4)	**A**	-131.092	111.149	-19.943	-49.705	-69.648	-2.739	-72.387
	B	-126.875	107.111	-19.765	-46.901	-66.665	-1.755	-68.420
$PhePCH_3$_R (5)	**A**	-105.517	80.097	-25.419	-35.786	-61.205	-5.594	-66.799
	B	-82.082	49.633	-32.450	-25.933	-58.382	-4.177	-62.559
$PhePCH_3$a_S (6)	**A**	-127.903	107.107	-20.796	-43.112	-63.908	-5.665	-69.573
	B	-97.938	68.441	-29.496	-31.089	-60.585	-3.948	-64.533