



Contents lists available at ScienceDirect

# Bioorganic & Medicinal Chemistry

journal homepage: [www.elsevier.com/locate/bmc](http://www.elsevier.com/locate/bmc)



## Elucidation of common pharmacophores from analysis of targeted metabolites transported by the multispecific drug transporter—Organic anion transporter1 (Oat1)

Valentina L. Kouznetsova<sup>a,b,f</sup>, Igor F. Tsigelny<sup>c,g</sup>, Megha A. Nagle<sup>a</sup>, Sanjay K. Nigam<sup>a,d,e,\*</sup>

<sup>a</sup> Department of Medicine, Division of Nephrology and Hypertension, University of California, San Diego, La Jolla, CA 92093, USA

<sup>b</sup> Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

<sup>c</sup> Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA

<sup>d</sup> Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA

<sup>e</sup> Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

<sup>f</sup> Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA

<sup>g</sup> San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA

### ARTICLE INFO

#### Article history:

Received 30 March 2011

Revised 22 April 2011

Accepted 23 April 2011

Available online 28 April 2011

#### Keywords:

Metabolite  
Molecular database screening  
Molecular fingerprint  
Molecular flexible alignment  
Organic anion transporter  
Pharmacophore consensus

### ABSTRACT

Organic anion transporter 1 (Oat1), first identified as NKT, is a multispecific transporter responsible for the handling of drugs and toxins in the kidney and choroid plexus, but its normal physiological role appears to be in small molecule metabolite regulation. Metabolites transported by Oat1 and which are altered in the blood and urine of the murine Oat1 knockout, may serve as templates for further drug design. This may lead to better tissue targeting of drugs or design of Oat1 inhibitors that prolong the half-life of current drugs. Due to the multispecificity of the transporter, 19 of known targeted metabolites have different chemical structures and properties that make constructing a common pharmacophore model difficult. Here we propose an approach that clustered the metabolites into four distinct groups which allowed for the construction of a consensus pharmacophore for each cluster. The screening of commercial molecular databases determined the top candidates whose interaction with Oat1 was confirmed in an experimental model of organic anion transport. Thus, these candidate selections represent potential molecules for further drug design.

© 2011 Elsevier Ltd. All rights reserved.

### 1. Introduction

The FDA Guidance for Industry<sup>1</sup> pays special attention to the design of drugs and drug–drug interactions related to transporters involved in metabolism. Among the transporters involved in the disposition of drugs, the organic anion transporters (Oats) have been the subject of numerous studies. The Oats are members of the well-characterized solute carrier (Slc) subfamily of drug transporters and facilitate the uptake and excretion of diverse array of endogenous metabolites/compounds, drugs, and toxins.<sup>5,6</sup> Oat1 (Slc22a6), originally identified as NKT, is expressed in a number of barrier epithelial tissues, including the proximal tubule of the kidney, where it is found in the basolateral membrane and is the rate-limiting step in the movement of drugs (i.e., diuretics, antibiotics, and selected anionic compounds), endogenous metabolites, and toxins from the blood and into the urine.<sup>2–4</sup>

Since the protein structures for these transporters have yet to be resolved, combined computational and experimental studies

of their specific molecular features are important for drug development. In the case of Oat1, the development of possible Oat1-related drugs requires an analysis of common features of molecules that are known to be transported by these proteins and their conformational space that makes possible pharmacophore modeling. A well-defined pharmacophore model of Oat1-transported metabolites could lead to design of new drugs aimed at targeting uptake and elimination by a particular organ such as the kidney. For example, inhibitors that specifically target Oat1 could increase the plasma half-life of drugs cleared by Oat1 (e.g., antibiotics, antivirals, and NSAIDs) or it could help reduce drug–drug interactions potentially toxic to kidney.

In a previous targeted metabolomics study, mass-spectrometric quantitative analysis of plasma and urine from Oat1-knockout mice was performed to identify endogenous metabolites transported by Oat1.<sup>7</sup> These global analyses revealed that the metabolite composition of plasma and urine in the knockout (KO) mice is indeed significantly different from that of wild-type (WT) mice.<sup>7</sup> Importantly, the observed differences were markedly skewed in the direction expected for endogenous substrates (i.e., increased in concentration in plasma and decreased in urine in KO animals).<sup>7</sup> Here, we have

\* Corresponding author. Tel.: +1 858 822 3482; fax: +1 858 822 3483.

E-mail address: [snigam@ucsd.edu](mailto:snigam@ucsd.edu) (S.K. Nigam).

used this list of Oat1-transported endogenous metabolites as the basis of pharmacophore hypotheses generation. These hypotheses were used to in silico screen 3D databases of NCI and commercial compounds to find potential Oat1-specific compounds.

## 2. Methods

### 2.1. Pharmacophore development and database screening

For the elucidation of possible fits to the pharmacophore hypotheses based on the set of targeted metabolites,<sup>7</sup> the following programs and modules from the program package MOE 2009.10 (Molecular Operating Environment, Chemical Computing Group (CCG), Montreal, Canada) were used: (1) database viewer for the creation of new databases, calculation of metabolite fingerprints, and database searching and browsing; (2) conformations program for database conformational import; (3) QuaSAR (quantitative structural activity relationship) program for cluster composition; (4) simulations program for flexible alignment of composed clusters; and (5) Pharmacophore Query Editor for pharmacophore development. The Open NCI Database (<http://cactvs.nci.nih.gov/download/nci/>) with 3D structures of over 250,000 compounds and CCG databases containing 3D structures of around 650,000 commercially available compounds were utilized. Because metabolites are small molecules, only those compounds with a molecular weight  $MW \leq 600$  g/mol were selected and all allowed conformations were found with conformational import module. These compounds were subjected to extensive conformational searches and all of those which had conformations within the limits of general molecular energy were used for further screening.

### 2.2. Pharmacophore target validation

Using pharmacophore-based screening of the 3D structure databases, a set of compounds with reasonable fits to the pharmacophore models were selected. Four groups of compounds were selected according to 4 metabolite clusters, each of which was then divided into 2 subgroups—"exact fit" (compounds that fit the selected features of the pharmacophore and geometrically do not extend over its boundaries) and "tail" (compounds that fit the selected features of the pharmacophore but have a portion of their structure that extends beyond the 3D profile of the pharmacophore). Representative compounds from each cluster (tail and exact fit) were purchased for subsequent analysis.

The following compounds were used: Cluster 1, exact fit—Chemical Block A3979/0169481 (MW 179.216), tail—Bionet 6H-305S (MW 325.148); Cluster 2, exact fit—Labotest LT00454450 (MW 187.121), tail—Bionet 8X-0800 (MW 295.319); Cluster 3, exact fit—Labotest LT00005591 (MW 270.475) and Labotest LT00847268 (MW 203.173), tail—Labotest LT00127613 (MW 269.297); Cluster 4, exact fit—Chembridge 5848164 (MW 237.303), tail—Chembridge 9004707 (MW 272.301). Compounds were dissolved in Barth's buffer (pH 7.4) to make up stock of 50 mM solution of each compound and then stored at  $-20^{\circ}\text{C}$ .

As described by Truong et al.<sup>8</sup> and Ahn et al.,<sup>9</sup> *Xenopus* oocyte cells were harvested, defolliculated with collagenase–trypsin inhibitor and microinjected with 23 nl/oocyte of 1  $\mu\text{g}/\lambda$  of mOat1 cRNA, which was transcribed by using mMessage mMachine in vitro transcription kit, from Ambion, Austin, TX. Capped RNA was synthesized using Image clone ID 4163278 for Slc22a6 (mOat1) from previously linearized plasmid DNA of mOat1, by using NotI restriction enzyme. *Xenopus* oocytes were allowed to rest for 2 days in solution containing 5% horse serum with gentamycin (0.05 mg/ml) in Barth's buffer before binding interaction assay was carried out as published previously.<sup>8,9</sup>

Each compound was tested at six different concentrations ranging from 10 mM to 0.1  $\mu\text{M}$  in the presence of a fluorescent tracer specific for mOat1 transporter, the 6-carboxyfluorescein (6CF). Experimental group of 20–25 oocytes/well, at each of the six different concentrations were tested against 30  $\mu\text{M}$  concentration of 6CF and incubated for 1 h at room temperature. After that the plate was placed on ice-water bath and each well with *Xenopus* oocytes, washed 3–4 times with ice-cold Barth's buffer and lysed overnight with 1 M NaOH to measure the tracer uptake using fluorometer (Polar Star plate reader, BMG Labtechnologies, Durham, NC). Tracer inhibition in the *Xenopus* oocyte cells by the selected compounds was calculated as percentage of controls.

## 3. Results

### 3.1. Clusters elucidation and flexible alignment

If one has a basis a set of compounds, the usual way to proceed with the design of pharmacophore hypotheses is the superimposition of the compounds using points specified in the molecular fingerprints of the compounds. Each fingerprint represents a set of features derived from the structure of a molecule. Fingerprints allow similarity searching and the clustering of a set of molecules. If a set contains a number of compounds with significantly different fingerprints, then the first step towards pharmacophore elucidation is clusterization of this set to identify smaller subsets of compounds having similar steric, electronic, and other fingerprint features (see, e.g., Chen et al.<sup>10</sup>).

In this study a set of 19 targeted metabolites with altered concentrations in the plasma and urine of Oat1-knockout mice were used (Table 1).<sup>7</sup> From these, a database of 3D structures and their conformers was created for further use with the MOE programs. Each of the metabolites studied was characterized by a number of molecular fingerprints. We selected and calculated 10 molecular fingerprints for each targeted metabolite. A fingerprint is a list of values, which characterize a molecule and in this study included (1) the list version of MACCS Structural Keys, which indicates the presence of 166 structural keys; (2) its bit-packed version; an eigenvalue spectrum of 3D shape made from the (3) heavy atoms and (4) the hydrophobic heavy atoms of a molecule; a three-point pharmacophore based on eight atom types calculated from the 2D molecular graph (5) and (6) from a 3D conformation; (7) a two-point pharmacophore based on six atom types calculated from the 3D conformation and (8) from the 2D molecular graph; a three-point pharmacophore based on six atom types calculated from the 3D conformation (9) and (10) from the 2D molecular graph.

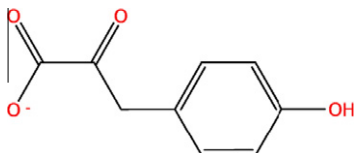
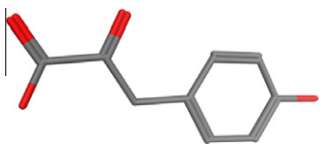
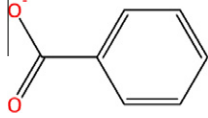
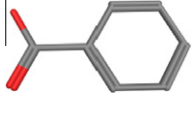
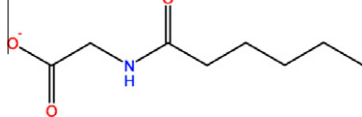
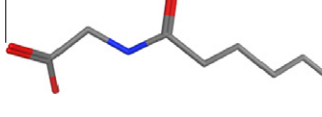
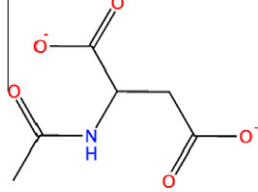
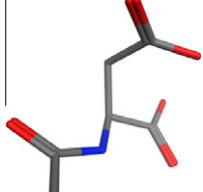
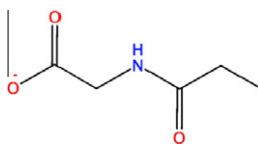
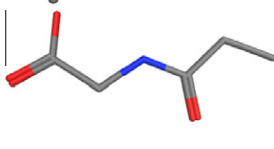
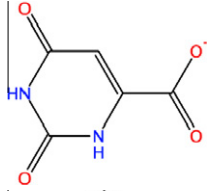
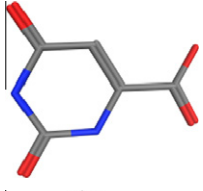
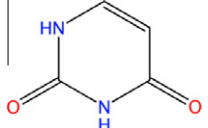
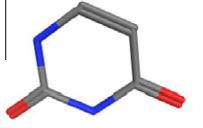
To combine the metabolites with similar chemical properties and geometry, QuaSAR-Cluster module in the MOE programs was employed.<sup>11</sup> This module is based on multidimensional weighted nonparametric ranking by principal component analysis (PCA) and can be applied to various three-dimension-structure sets. The metabolites were separated into clusters, using the following fingerprints: 3D shapes of all heavy atoms or only hydrophobic atoms, and three-point pharmacophore based on six atoms (ESshape3D, ESshape3D\_HYD, and TAT); as the weighted vector. A three-point pharmacophore based on eight atom types calculated from a 3D conformation (piDAPH3) was ultimately used and four clusters were obtained (Fig. 1). Flexible alignment of the molecules included in each cluster was then performed using a Flexible Alignment application from Simulations module.

According to Paul Labute,<sup>12</sup> 'an alignment is good if (a) the strain energy of each molecule is small; (b) they have a similar shape; (c) their aromatic atoms overlap; (d) their donors overlap and (e) their acceptors overlap.' A metric for feature comparison

**Table 1**  
Targeted metabolites<sup>7</sup>

#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	2D structure	3D structure
1	2-Hydroxy-3-methylvalerate—acid	C <sub>6</sub> H <sub>12</sub> O <sub>3</sub>	131.149740		
2	2-Hydroxyisovalerate	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub>	118.131100		
3	2-Oxo-3-methylvalerate—α-keto-β-methylvaleric acid	C <sub>6</sub> H <sub>9</sub> O <sub>3</sub> <sup>−</sup>	130.141800		
4	2-Oxoglutarate—α-ketoglutarate	C <sub>5</sub> H <sub>4</sub> O <sub>3</sub> <sup>2−</sup>	144.082260		
5	2-Oxoisocaproate	C <sub>6</sub> H <sub>9</sub> O <sub>3</sub> <sup>−</sup>	129.133860		
6	3-Hydroxybutyrate	C <sub>4</sub> H <sub>7</sub> O <sub>3</sub> <sup>−</sup>	103.096580		
7	3-Hydroxyisobutyrate	C <sub>4</sub> H <sub>7</sub> O <sub>3</sub> <sup>−</sup>	103.096580		
8	3-Hydroxypropionate	C <sub>3</sub> H <sub>5</sub> O <sub>3</sub> <sup>−</sup>	89.070000		
9	3-Hydroxyvalerate—β-hydroxyvaleric acid	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub>	118.131100		
10	3-Methylcrotonylglycine	C <sub>7</sub> H <sub>11</sub> NO <sub>3</sub>	157.167140		
11	4-Hydroxyphenylacetate	C <sub>8</sub> H <sub>8</sub> O <sub>3</sub>	152.147320		
12	4-Hydroxyphenyllactate	C <sub>9</sub> H <sub>9</sub> O <sub>4</sub> <sup>−</sup>	181.165360		

Table 1 (continued)

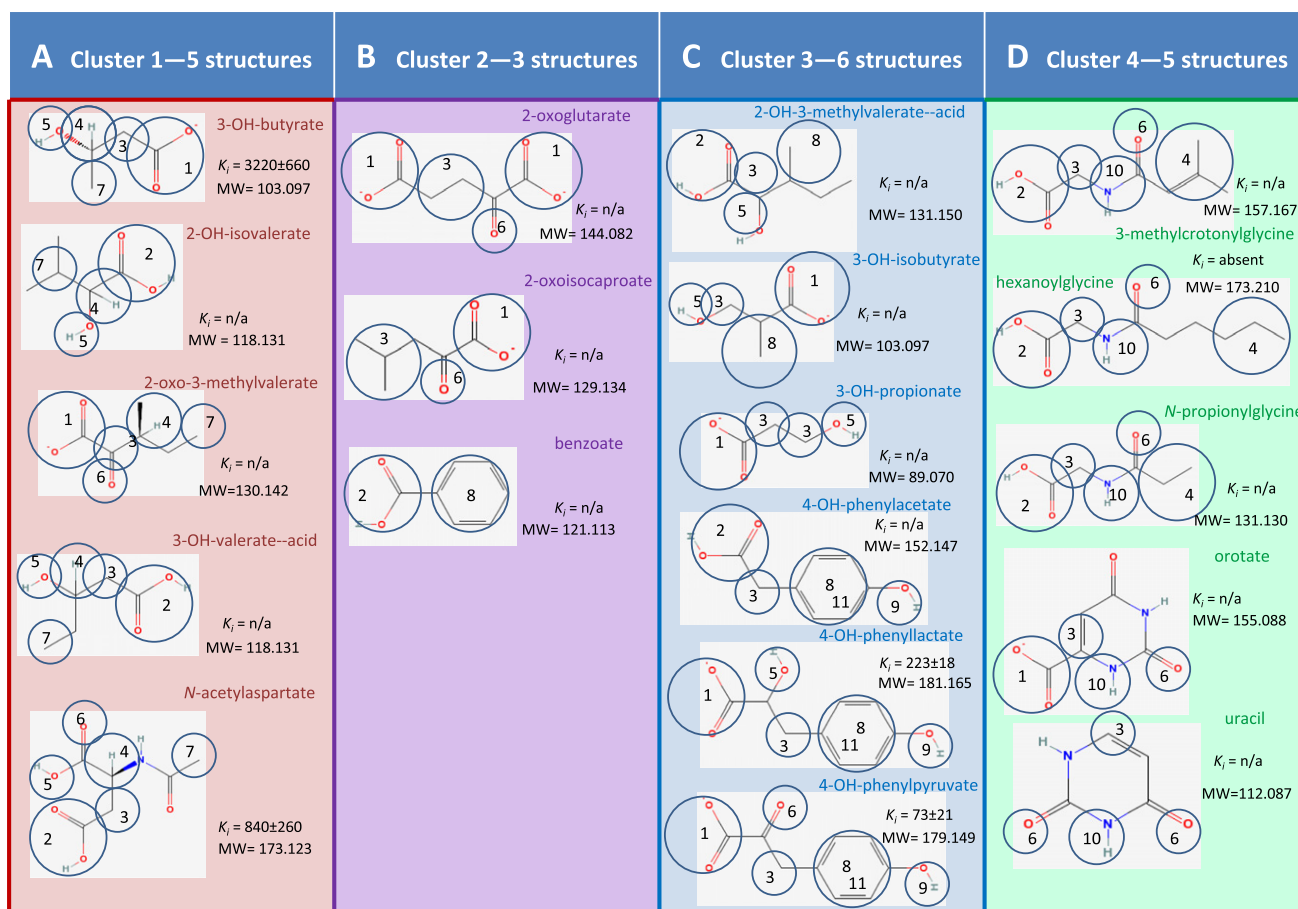
#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	2D structure	3D structure
13	4-Hydroxyphenylpyruvate	$C_9H_7O_4^-$	179.149480		
14	Benzoate	$C_7H_5O_3^-$	121.113400		
15	Hexanoylglycine	$C_8H_{15}NO_3$	173.209600		
16	N-Acetylaspartate	$C_6H_7NO_5^{2-}$	173.123480		
17	N-Propionylglycine	$C_5H_9NO_3$	131.129860		
18	Orotate	$C_4H_4N_2O_4^-$	155.088320		
19	Uracil	$C_4H_4N_2O_2$	112.086760		

is a feature probability density of a given conformation of a molecule.<sup>12</sup> Based on this notion, the alignment databases for each cluster with their corresponding scores were created. The score *S* is sum of the average potential of the molecules, *U* represents an internal strain, while feature similarity is represented by function *F*. The lower the score, the better the alignment. Score data together with cluster alignment images are shown in Table 2. For cluster 1, the alignment 1 has the lowest total score. For cluster 2 the first two alignments have the same total scores, but similarity for alignment 1 is a little better; so it was selected. The alignment 1 for cluster 3 has the lowest total score as well the lowest similarity score. We can see that all double oxygen (O2) groups align together. For cluster 4, alignment 4 was selected. Although alignments 1, 2, and 3 have lower scores and higher similarity, there are nitrogen atoms spread around the rings and in alignment 3 nitrogen donors were combined with hydrophobic carbons and hydrogen-bond oxygens are not aligned well; alignment 5 has all these drawbacks and had the worst all three scores. The constructed flexible alignments are shown in Fig. 2.

### 3.2. Pharmacophore development

As a next step, a pharmacophore hypothesis for each of four clusters was generated using the Pharmacophore Query Editor module implemented in MOE. According to Peter Gund, a pharmacophore is 'a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity'.<sup>13</sup>

Because each aligned cluster includes a set of somewhat different structures, a consensus hypothesis using the Pharmacophore Consensus application was obtained for each cluster. The Pharmacophore Consensus application calculated a query composed of all features of all molecules with an indication of the proportion of shared features, which in this case included anionic (negative ionizable), hydrogen-bond donor or acceptor, cationic (positive ionizable), double oxygen, and hydrophobe. Existing consensus structures aid in deciding how many features to use for further employment of the pharmacophore for compound searches in databases. This decision is usually based on the maximum number of



**Figure 1.** Features used for pharmacophore design. Circled: 1—doubled oxygen; 2—anionic group (in the media with the pH around 7, which is typical for human body, COOH group is ionized to carboxylate ion ( $\text{COO}^-$ ); that is why in both cases the  $\text{COO}^-$  and  $\text{COOH}$  we use them in the pharmacophore hypotheses always as the feature double oxygen); 3, 4, 7, and 8—hydrophobic centers; 5 and 9—H-bond donors; 6 and 10—H-bond acceptors; 11—aromatic ring. MW—molecular weight, g/M;  $K_i$ —inhibition coefficient,  $\mu\text{M}$ .

features having maximum fits for all superimposed molecules. After analysis of consensus pharmacophore hypotheses, 9 features for the first cluster, 7 features for the second, and 5 features each for the third and the fourth cluster were selected. Because the resolved structure for Oat1 (Slc22a6) does not exist, a ligand-based approach for pharmacophore development was used that employed conformational analysis of known ligands and did not take into consideration possible ligand–transporter interactions.

Our analysis identified the common functional group features (see Fig. 1). The main feature of pharmacophore hypotheses for all clusters is either a  $\text{COO}^-$  group—a double oxygen (O2) center (points 1)—or a COOH group—an anionic group (points 2), both of which could be an acceptor. At pH  $\sim 7$ , which is biologically relevant, the COOH group is ionized to carboxylate ion ( $\text{COO}^-$ ); this allows for use of both the  $\text{COO}^-$  and COOH in the pharmacophore hypotheses always as the double oxygen feature. Each cluster also includes at least two hydrophobic centers (points 3 and 4), hydrogen-bond (H-bond) donors (points 5) and H-bond acceptors (points 6).

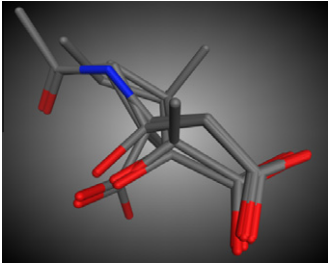
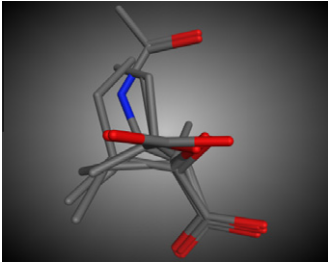
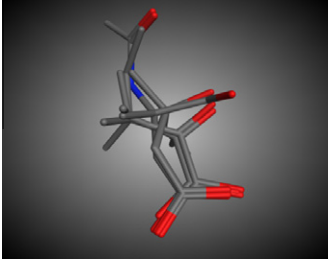
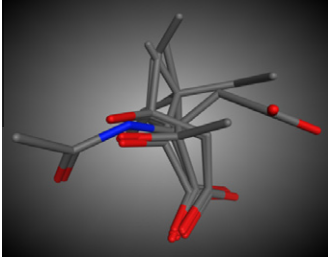
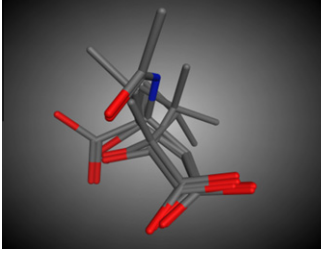
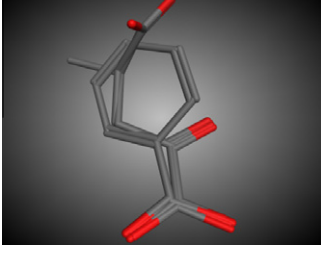
For the cluster 1 containing 5 targeted metabolite structures (indicated in Fig. 1A by black circles), the following applies: 5 of them contain a double oxygen (O2) center or anionic group (points 1 or 2) that could be also an acceptor. Each structure also includes at least 2 hydrophobic centers (points 3 and 4). Three of metabolite molecules of cluster 1 contain H-bond donors (points 5), and 2 of them contain H-bond acceptors (points 6); each of structures has an additional hydrophobic center (point 7). The consensus phar-

macophore hypothesis consists of 5 main features: (F1) O2|Ani|Acc (points 1 and 2 in Fig. 1A, red sphere in Fig. 3A); (F2) O2|Acc|Ani|Don (points 5 and 6, pink sphere); (F3) hydrophobic center (points 4, green sphere); and (F4) Hyd|Don (points 3 and 7, purple sphere). To specify the directions of possible interactions with the Oat1 multispecific drug transporter, we included for F1 4 projected acceptors (Acc2)—F5–F9 (cyan spheres) making total 9 features, and applied a constraint that two belong to the same atoms.

The cluster 2 contains 3 structures and is simpler consisting of two features (indicated in Fig. 1B by black circles): either O2 or anionic group (points 1 or 2) that could also be an acceptor and a hydrophobic center (points 3); 2 of the structures have H-bond acceptor (points 6). The consensus pharmacophore hypothesis consists of 3 main features: (F1) O2 (points 1 and 2 in Fig. 1B, red spheres in Fig. 3B); (F2) Hyd (points 3, green sphere); and (F3) Acc (points 6, cyan sphere). To specify the directions of possible molecular interactions, we include for F1 4 projected acceptors (Acc2)—F4–F7 (cyan spheres), making total 7 features, and we applied a constraint that two belong to the same atoms.

The cluster 3 contains 6 structures and has the following features (indicated in Fig. 1C by black circles): either O2 or anionic group (points 1 or 2) that could be also an acceptor and hydrophobic centers (points 3 and 8); 5 of the structures include H-bond donors (points 5) and 3 of them—H-bond donors (points 9) (all donors 5 and 9 also could be represented as H-bond acceptors); 3 of the members have aromatic rings (points 11), which were not taken into consideration, using only their hydrophobic features

**Table 2**  
Flexible alignment analysis of metabolites

Cluster	Alignment	Image	U	F	S
1	1		−63.1706	−68.6994	−131.8700
1	2		−61.3114	−69.6110	−130.9224
1	3		−62.4237	−68.1157	−130.5194
1	4		−62.2058	−68.2948	−130.5007
1	5		−61.3571	−69.1425	−130.4996
2	1		−50.3771	−66.2615	−116.6386

(continued on next page)

Table 2 (continued)

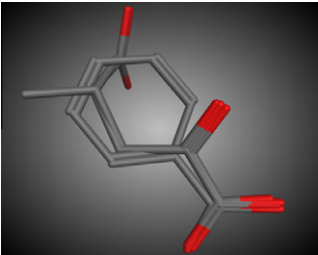
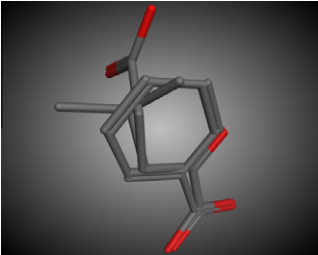
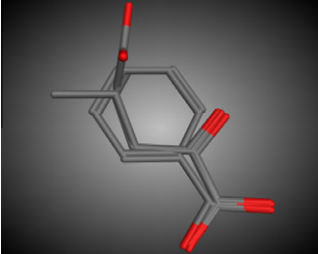
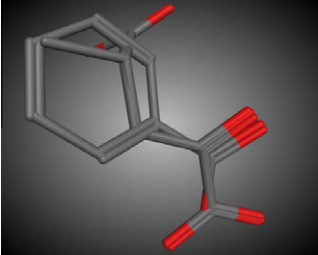
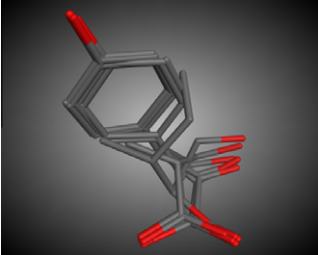
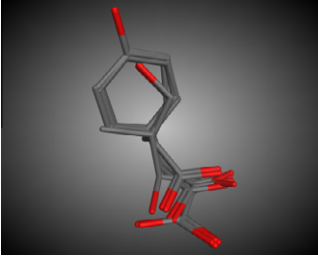
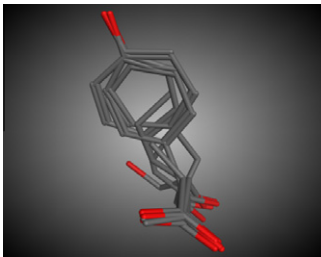
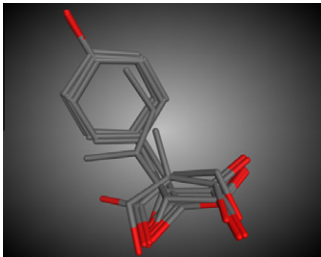
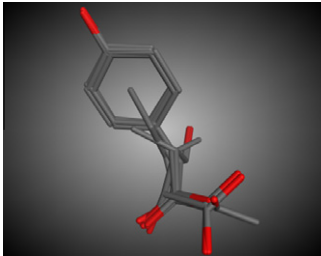

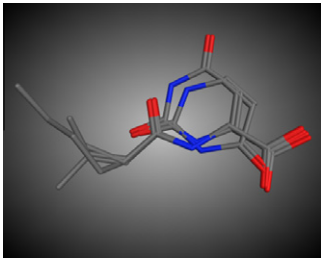
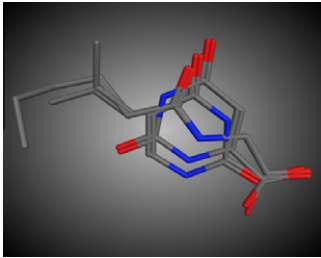
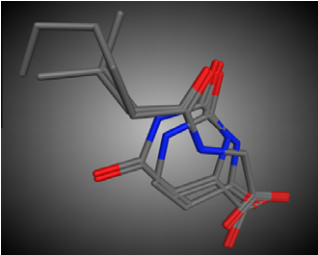

Cluster	Alignment	Image	U	F	S
2	2		−50.3803	−66.2583	−116.6386
2	3		−50.5530	−65.2380	−115.7910
2	4		−50.5534	−65.2377	−115.7910
2	5		−49.7750	−62.3577	−112.1328
3	1		−42.6422	−65.6408	−108.2830
3	2		−44.0517	−61.9684	−106.0201

Table 2 (continued)

Cluster	Alignment	Image	U	F	S
3	3		–42.6829	–63.2378	–105.9207
3	4		–42.8749	–62.7915	–105.6664
3	5		–44.3928	–60.9445	–105.3374
4	1		–91.5169	–63.9051	–155.4220
4	2		–91.7140	–63.5931	–155.3071
4	3		–91.1619	–63.8385	–155.0004

(continued on next page)

Table 2 (continued)

Cluster	Alignment	Image	U	F	S
4	4		−91.2368	−63.5285	−154.7652
4	5		−90.6560	−63.68354	−154.3423

The score S is sum of average potential of the molecules U that represents an internal strain and a feature similarity function F.

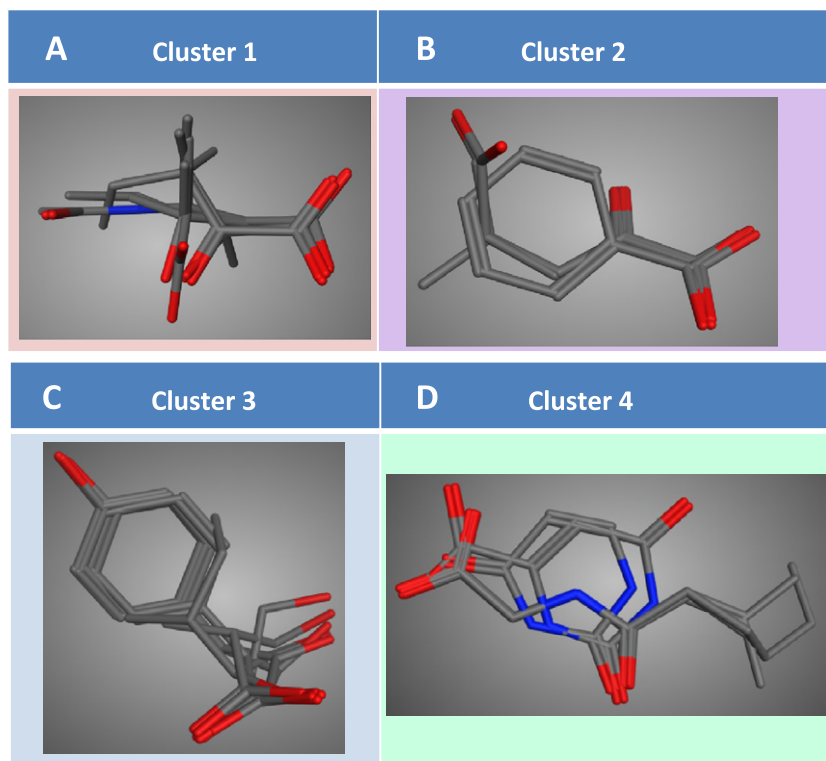


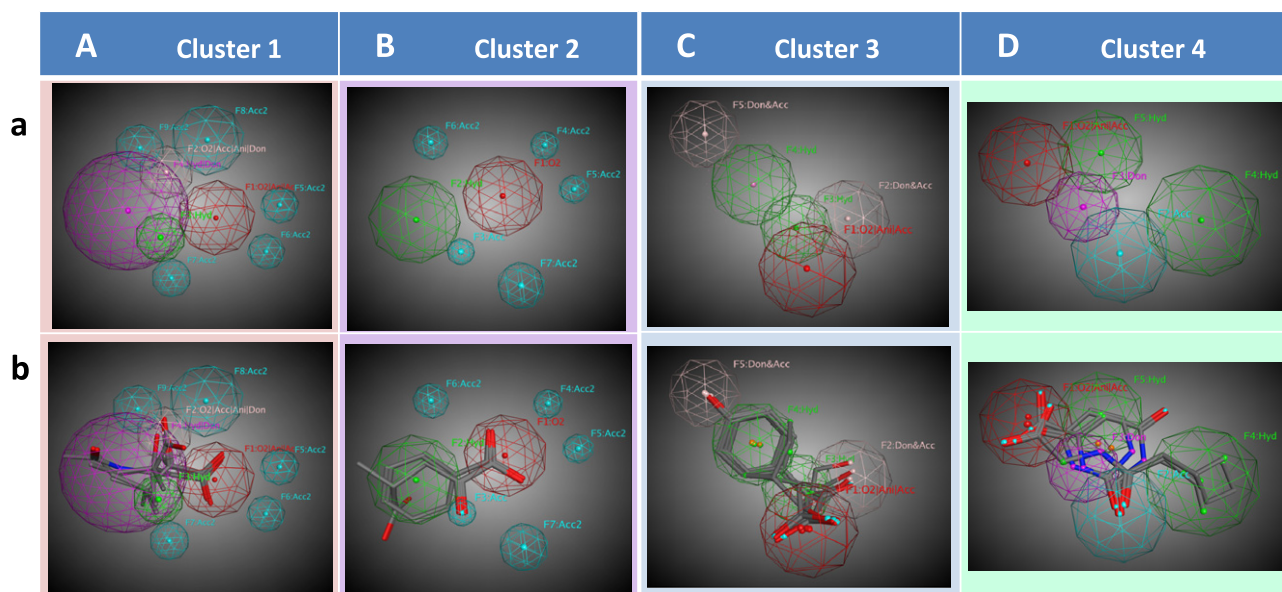
Figure 2. Superimposition and flexible alignments of metabolites for each cluster: red—oxygen, blue—nitrogen, gray—carbon atoms.

(points 8). The consensus pharmacophore hypothesis consists of 5 main features: (F1) O2|Ani|Acc (points 1 and 2 in Fig. 1C, red sphere in Fig. 3C); (F2) Don&Acc (points 5, pink sphere); (F3) Hyd (points 3, green sphere); (F4) Hyd (points 8, green sphere); (F5) Don&Acc (points 9, pink sphere).

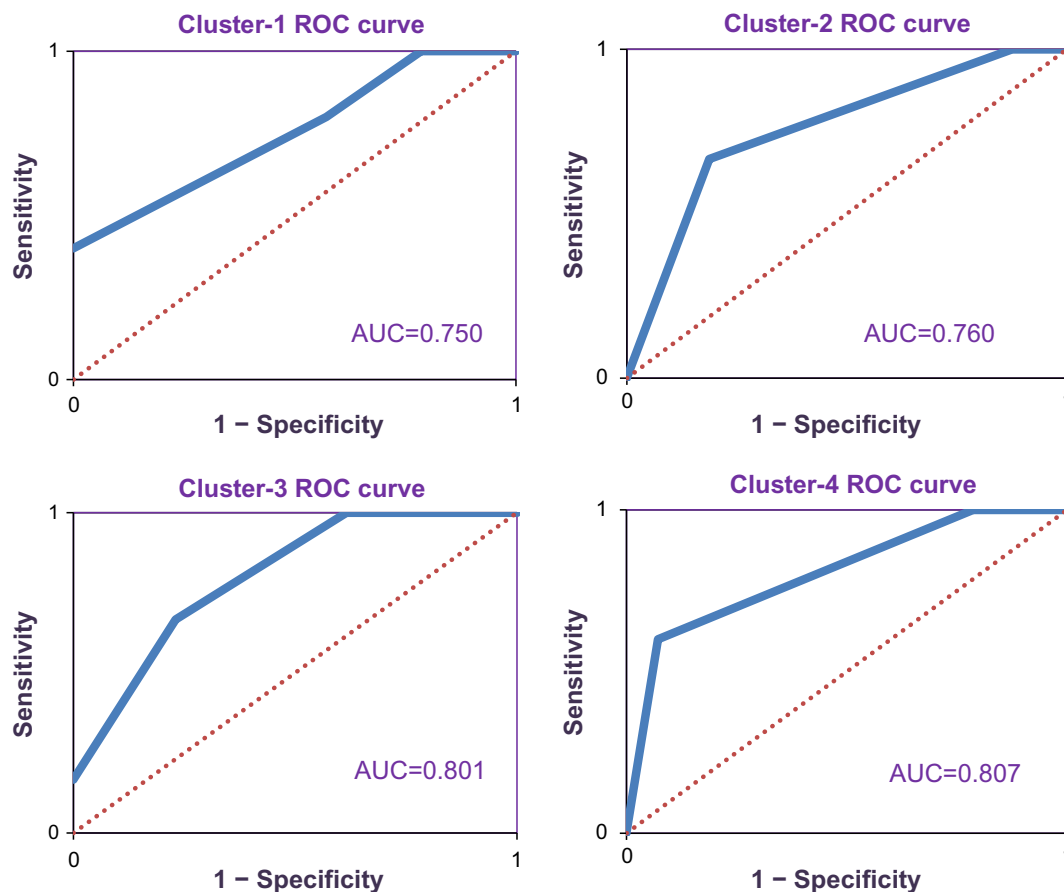
The cluster 4 consists of 5 structures, each of which includes (indicated in Fig. 1D by black circles) either O2 or anionic group (points 1 or 2) that could be also an acceptor, hydrophobic centers (points 3), H-bond acceptors (points 6), and H-bond donors (points 10); 3 of the members have additional hydrophobic groups (points

4). The resulting consensus pharmacophore hypothesis consists of 5 main features: (F1) O2|Ani|Acc (points 1 and 2 in Fig. 1D, red sphere in Fig. 3D); (F2) Acc (points 6, cyan sphere); (F3) Don (points 10, green sphere); (F4) Hyd (points 3); and (F5) Hyd (points 4, green sphere).

We validated developed pharmacophore performance, generating confusion matrices and plotting receiver operating characteristic (ROC) curves. The ROC curve is a plot of sensitivity, or true-positive rate (TPR) versus 1 – specificity, or 1 – true-negative rate (TNR). Both rates are calculated by Eqs. (1) and (2) correspondingly.<sup>14,15</sup>



**Figure 3.** Pharmacophore consensus hypotheses for each cluster (A–D). a—pharmacophore hypotheses, b—pharmacophore hypotheses aligned with superimposed metabolites. Red spheres—O2|Ani|Acc (points 1 and 2 in Fig. 1A–D), pink spheres—O2|Acc|Ani|Don (points 5 and 6 in Fig. 1A–D and 5 and 9 in Fig. 1C), green spheres—hydrophobic centers (points 3 in Fig. 1B and C, 4 in Fig. 1A and D, and 8 in Fig. 1B and C), purple spheres—Hyd|Don (points 3 and 7 in Fig. 1A and D) and cyan spheres—acceptors (points 6 in Fig. 1B and D) and projected acceptors (see text for details).



**Figure 4.** Receiver operating characteristic (ROC) curves for the pharmacophore performance on the set of studied metabolites. Red dotted line is a random guess. AUC is an area under curve that characterizes the accuracy of prediction.

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

where TP is the number of true positives, that is a number of correctly captured compounds; TN—number of true negatives, for example, the number of correctly rejected compounds; FP—number of false positives, for example, the number of mistakenly captured

**Table 3**

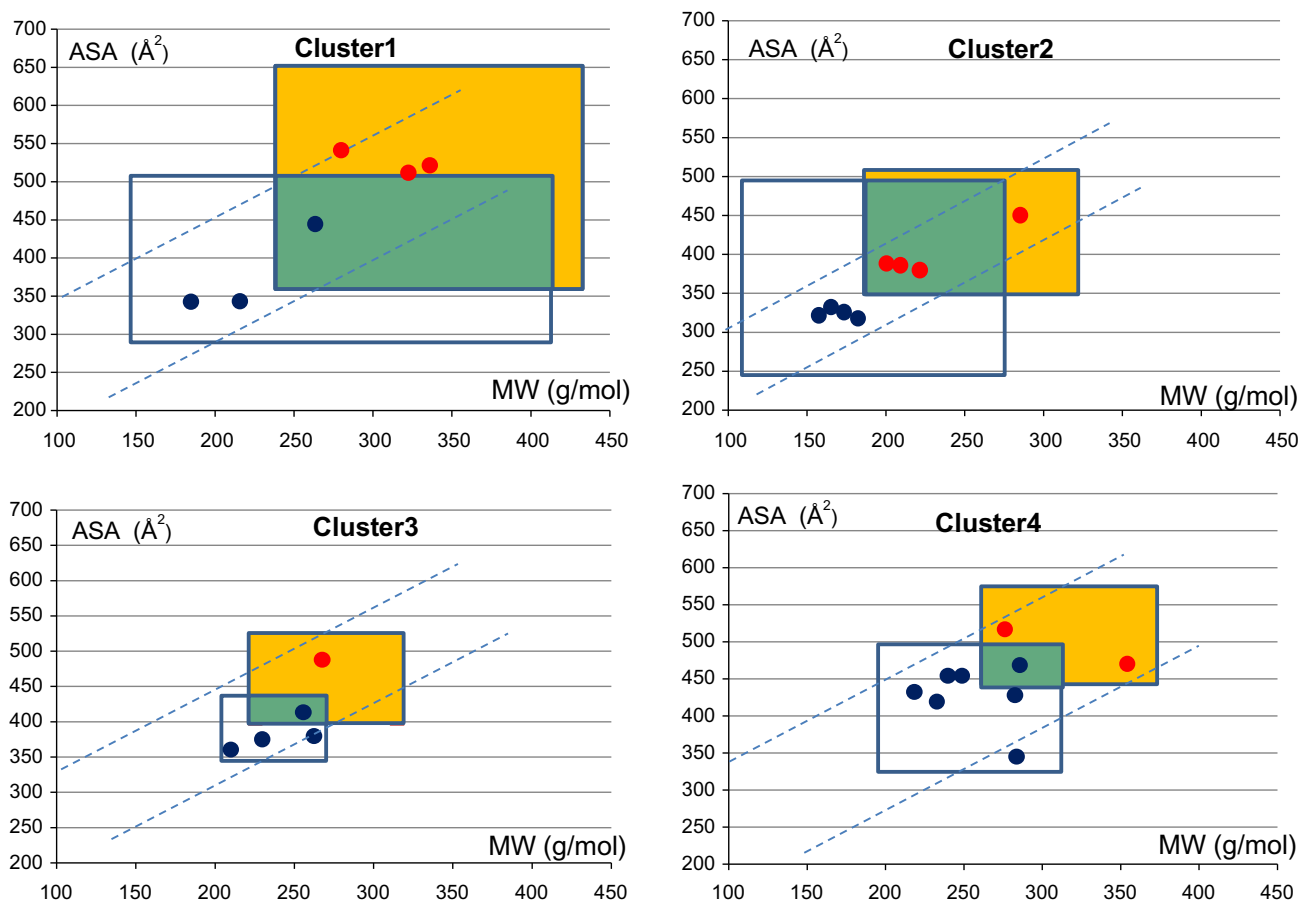
Summary of database search (653,214 compounds): (a) exact-fit molecules; (b) molecules extending outside the pharmacophore ('tails')

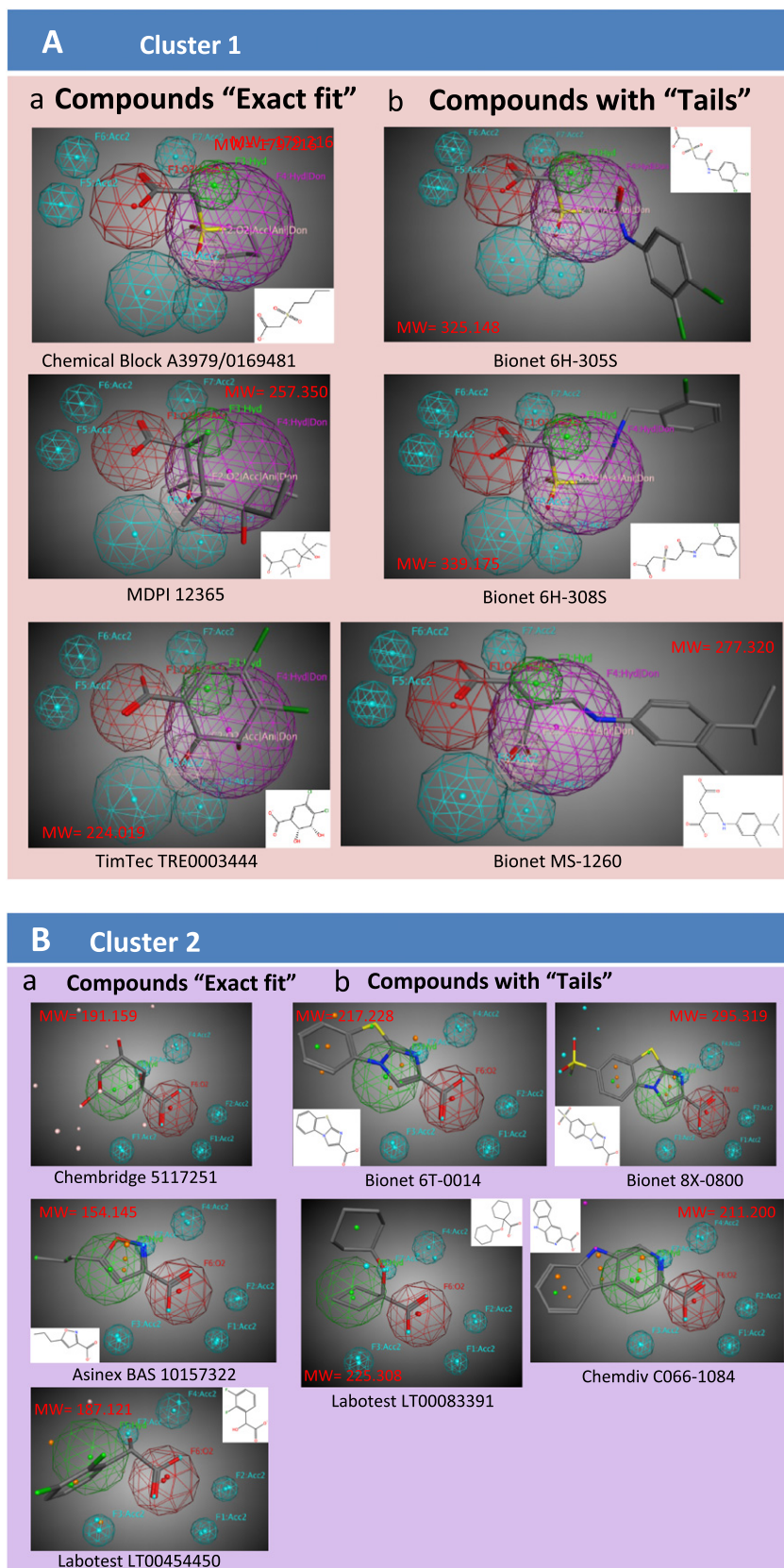
Pharmacophore hypothesis	Points	Hits	Exact-fit hits acquired	Exact-fit hits recommended	Exact-fit hits tested
<b>(A)</b>					
1 (cluster # 1)	9 of 9	4296	14	3	1
2 (cluster # 2)	7 of 7	954	16	3	1
3 (cluster # 3)	5 of 5	56	7	4	2
4 (cluster # 4)	5 of 5	7407	31	7	1
Total		12,713	58	17	5
<b>(B)</b>					
			Tail hits acquired	Tail hits recommended	Tail hits tested
1 (cluster # 1)	9 of 9	4296	27	3	1
2 (cluster # 2)	7 of 7	954	25	4	1
3 (cluster # 3)	5 of 5	56	3	1	1
4 (cluster # 4)	5 of 5	7407	11	2	1
Total		12,713	46	10	4

**Table 4**

Analysis of molecular weight (MW) and solvent accessible surface (ASA) of molecules that fit pharmacophore hypothesis

	MW_MAX_EF	MW_min_EF	MW_MAX_T	MW_min_T	ASA_MAX_EF	ASA_min_EF	ASA_MAX_T	ASA_min_T
Cluster 1	411.822	145.134	430.864	236.252	505.488	291.194	653.782	360.896
Cluster 2	274.299	111.080	322.650	187.219	489.292	243.441	508.802	348.623
Cluster 3	270.475	203.173	314.749	225.250	431.279	341.592	510.030	404.792
Cluster 4	313.172	194.221	372.771	264.345	502.898	329.861	578.862	450.862
Global	411.842	111.080	430.864	187.219	505.488	243.441	653.782	348.623

**Figure 5.** Selection of compounds representative for each cluster of metabolites. Compounds: fitting in pharmacophore boundaries (transparent rectangles); extending beyond the boundaries (yellow rectangles); zone containing both fitting and extending compounds (green). Blue points—parameters of selected fit compounds, red—extended compounds.



**Figure 6.** Results for pharmacophore search of commercial compound database incorporated in MOE (CCG, Montreal, Canada) for clusters 1–4 (A–D). Pharmacophore superimposed with (a) ‘exact fit’ compounds; (b) compounds with structures extending outside the pharmacophore—‘tails’.

compounds; and FN—number of false negative hits, for example, the number of missed compounds.

ROC curves along with confusion matrices allow estimating the performance of classifiers.<sup>15</sup> The ROC curve encapsulates all infor-

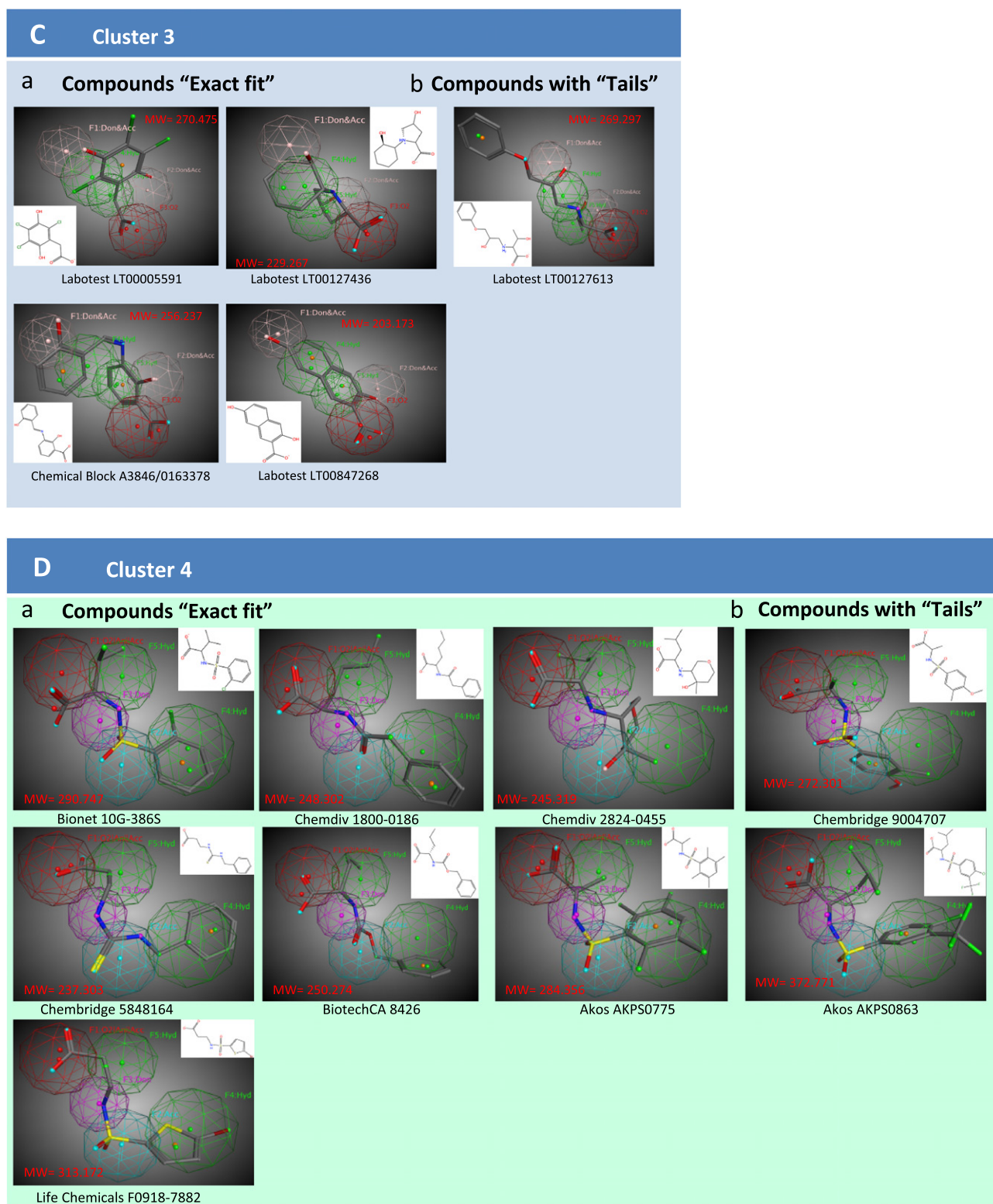
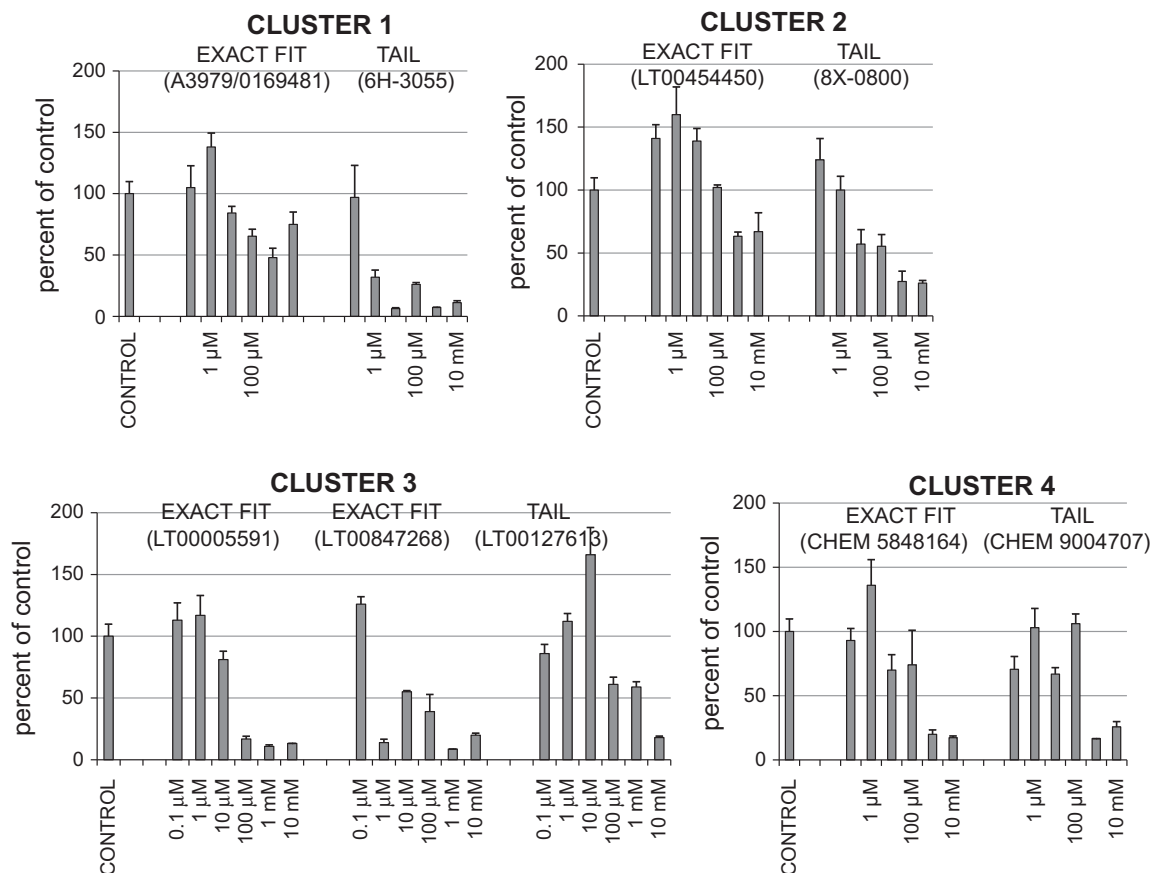


Fig. 6 (continued)

mation contained in the confusion matrix and provides clear visualization. The point (0, 1) is the perfect classifier that distinguishes correctly between classes, the point (0, 0) is the classifier that predicts all negative, the point (1, 1) is the classifier that predicts all positive, and the point (1, 0) is the classifier that predicts all incorrectly. The quantitative parameter derived from ROC curve is an

area under curve (AUC) that represents the accuracy of prediction. Closer AUC is to 1, the higher is prediction. The AUC value equal to 0.5 represents a random selection (red dotted lines in Fig. 4).

We prepared a 3D conformational database containing 3D metabolite structures used in this study and screened it with each of four designed pharmacophores. The thresholds were set through



**Figure 7.** In vitro validation of pharmacophore model by *Xenopus* oocyte assay. Compounds belonging to Clusters 1, 2, 3, and 4 (exact fit and tails) were tested at 6 different concentrations against 30  $\mu$ M of 6-carboxyfluorescein (6CF). Clusters 1 and 2 showed strong inhibition of 6CF with tail compounds, but moderate inhibition with exact-fit compounds. In contrast, cluster 3 showed extremely strong binding in two tested compounds when compared with tail compound which showed moderate inhibition. In cluster 4 both tail and exact-fit compounds showed variable binding in Oat1 microinjected *Xenopus* oocytes. A3979/0169481—Chemical Block A3979/0169481, 6H-305S—Bionet 6H-305S, 8X-0800—Bionet 8X-0800, LT00454450, LT00005591, LT00847268, and LT00127613—Labotest LT00454450, LT00005591, LT00847268, and LT00127613 correspondingly, CHEM 5848164—Chembridge 5848164, and CHEM 9004707—Chembridge 5848164.

number of pharmacophore features. The resulting ROC curves are shown in Figure 4. One can see that all the pharmacophores show AUC between 0.8 and 0.75 that represents good accuracy of prediction.

### 3.3. Molecular database screening

Each of four designed pharmacophore hypotheses were used to search a commercial compound database provided by CCG (Montreal, Canada), which is composed of 653,214 compounds. Screening the database resulted in 12,713 hits (see Table 3A and B), which were filtered using molecular weight (MW) and water-accessible soluble area (ASA). There are 2 categories of molecules that fit the pharmacophore hypotheses: (1) those that fit it exactly and (2) those that have structures extended beyond the pharmacophore boundaries (see Table 3).

Analysis of MWs and ASAs of both categories of molecules show that there exists a fuzzy category that includes both types of molecules with exact fit but which, for example, have aromatic rings or heavier atoms, and are heavier than that MW boundary selected for the 'fit' molecules and molecules with simple long tails that are lighter. We defined the following selections based on MW: (a) we excluded all molecules that have MW less than minimum MW for molecules extended beyond the pharmacophore boundaries (MW\_min\_T); (b) we set a handicap between minimum MW for extended molecules (MW\_min\_T) and maximum MW for exact-fit molecules (MW\_MAX\_EF); and (c) we included those that

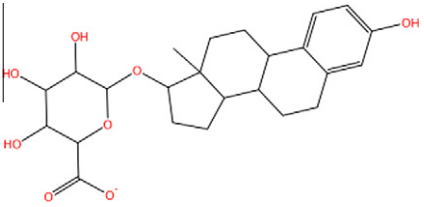
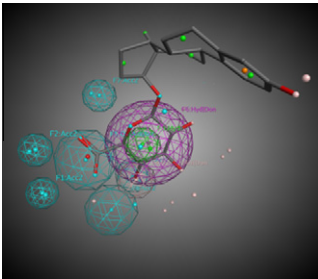
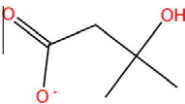
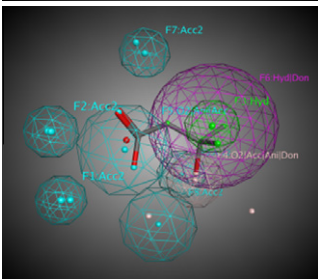
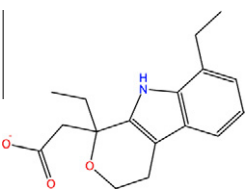
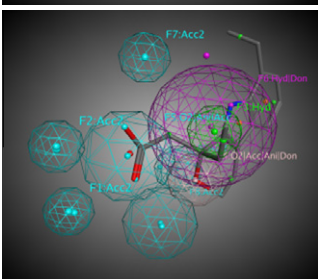
fall in the same MW range for exact-fit and extended compounds (see Table 4). Then we performed the same procedure for ASA. Figure 5 shows the two-dimensional plots with MW on axis *x* and ASA on axis *y*, with the points corresponding to the parameters of the selected compounds. We also selected the compounds within the possible linear increase region for combination of MW and ASA to find out whether this combination would be useful for further selection of compounds from databases. As a result 58 exact-fit compounds and 46 extended compounds were obtained as shown in Table 3. For experiments we chose the top-scoring compounds: 17 for exact fit and 10 for tails (see Table 3 and Fig. 6).

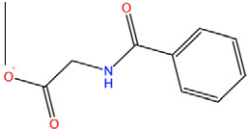
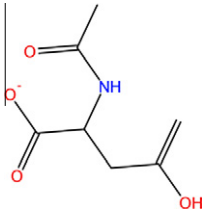
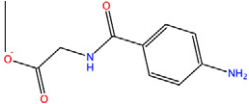
### 3.4. Pharmacophore validation by *Xenopus* oocyte assay

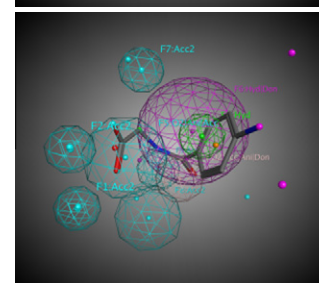
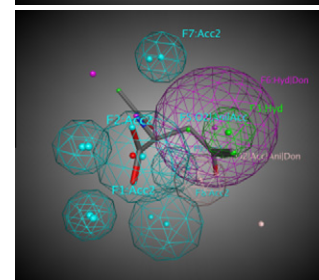
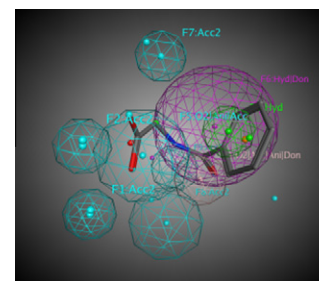
All the compounds that were selected based on the pharmacophore models designed from the Oat1-related metabolites were assayed in vitro in mouse Oat1 microinjected *Xenopus* oocytes. The inhibition of 6CF uptake, a fluorescent Oat1-specific tracer, by the selected compounds was determined. Overall, most of these compounds showed strong inhibition of 6CF uptake and some of the compounds showed affinities at concentrations as low as 1  $\mu$ M.

From the pharmacophore hypothesis based on cluster 1 (further called cluster-1-based compounds), an exact-fit compound from Chemical Block A3979/0169481 (MW 179.216) did not show the expected affinity towards mOat1 in contrast to the compound from the tail group from Bionet 6H-305S (MW 325.148), which exhibited extremely strong binding even at concentrations as low as

**Table 5**  
Results of screening of Oat1-related substrates<sup>16</sup> database with the pharmacophore of cluster 1 (9 features of 9)

#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	Transporter	System	$K_i$	RMSD, fit	2D structure	3D structure
1	17 $\beta$ -Estradiol-17-( $\beta$ -D-glucuronide)–E217G	C <sub>24</sub> H <sub>32</sub> O <sub>8</sub>	448.5061	rOat1	LLC-PK1	>300	0.9977 tail		
2	3-Hydroxybutyrate <sup>+</sup>	C <sub>4</sub> H <sub>7</sub> O <sub>3</sub> <sup>–</sup>	103.0966	mOat1	<i>X. laevis</i>	3320±660	0.5872 exact fit		
3	Etodolac	C <sub>17</sub> H <sub>21</sub> NO <sub>3</sub>	287.3535	rOAT1	LLC-PK1	>100	0.6848 exact fit		

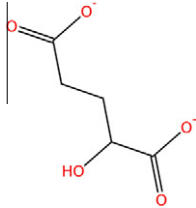
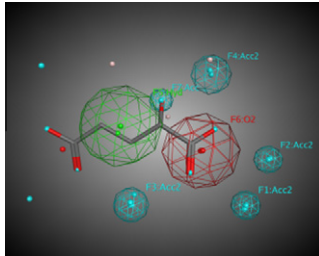
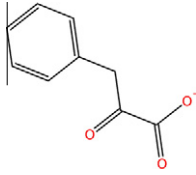
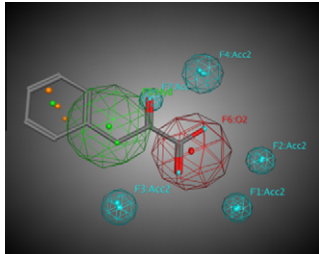
4	Hippurate	$C_9H_8NO_3^-$	178.1647	hOAT1	HEK293	66.0±13.6	0.9421 exact fit	
5	N-Acetylaspartate <sup>*</sup>	$C_6H_7NO_5^{2-}$	173.1235	mOat1	<i>X. laevis</i>	840±260	0.7068 exact fit	
6	p-Aminohippuric acid	$C_9H_{10}N_2O_3$	194.1873	hOAT1	S2	6.02	0.904 exact fit	



<sup>\*</sup> Metabolite, cluster 1.

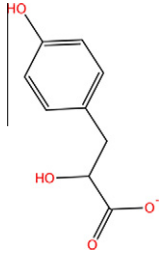
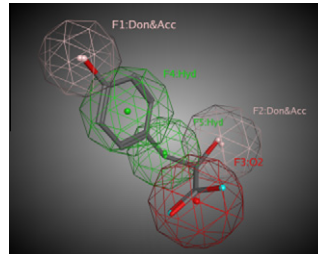
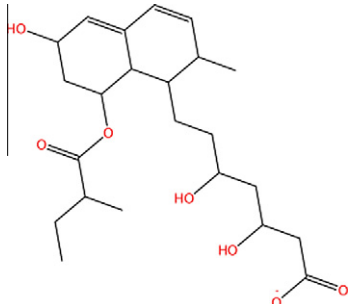
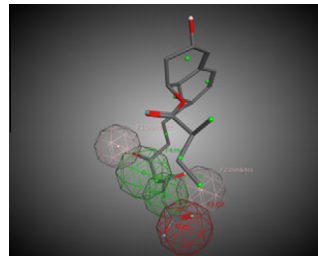
**Table 6**  
Results of screening of Oat1-related substrates<sup>16</sup> database with the pharmacophore of cluster 2 (7 features of 7)

#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	Transporter	System	$K_i$	RSMD, fit	2D structure	3D structure
1	17 $\beta$ -Estradiol-17-( $\beta$ -D-glucuronide)—E217G	C <sub>24</sub> H <sub>32</sub> O <sub>8</sub>	448.5061	rOat1	LLC-PK1	>300	0.2657 tail		
2	3-Hydroxyglutarate*	C <sub>5</sub> H <sub>8</sub> O <sub>5</sub>	148.114	hOAT1	HEK293	98 ± 30	0.2519 exact fit		
3	4-Hydroxyphenyl-lactate*	C <sub>9</sub> H <sub>9</sub> O <sub>4</sub> <sup>-</sup>	181.1654	mOat1	<i>X. laevis</i>	223 ± 18	0.2956 tail		
4	4-Hydroxyphenyl-pyruvate*	C <sub>9</sub> H <sub>7</sub> O <sub>4</sub> <sup>-</sup>	179.1495	mOat1	<i>X. laevis</i>	73 ± 21	0.2281 tail		
5	Citrinin	C <sub>13</sub> H <sub>14</sub> O <sub>5</sub>	250.2473	hOAT1 rOat1	S2 S2	3080	0.5888 tail		

6	L-2-Hydroxyglutarate	$C_5H_6O^{5-}$	146.0981	mOat1	<i>X. laevis</i>	$840 \pm 260$	0.2276 exact fit		
7	Phenylpyruvate	$C_9H_8O_3$	163.1588	mOat1	<i>X. laevis</i>	$79 \pm 26$	0.2281 exact fit		

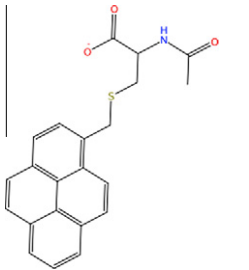
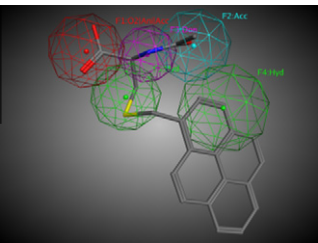
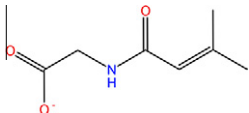
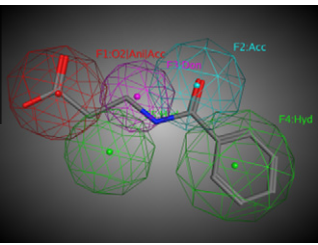
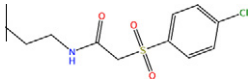
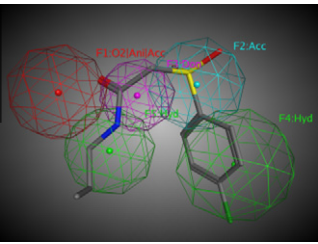
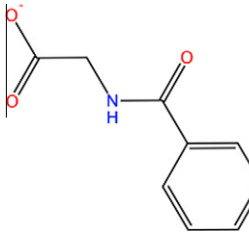
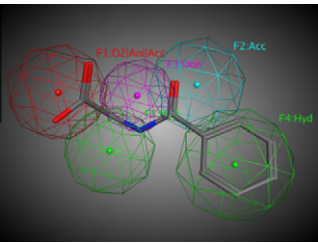
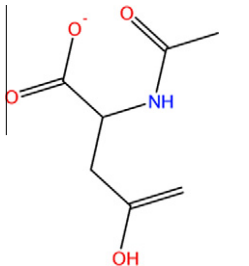
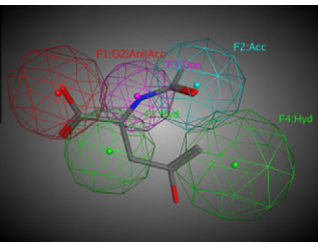
\* Metabolite, cluster 3.

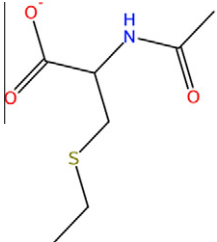
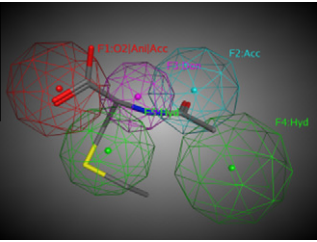
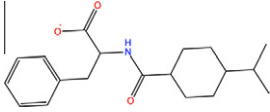
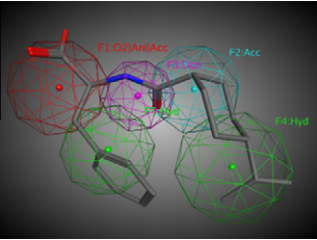
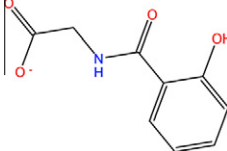
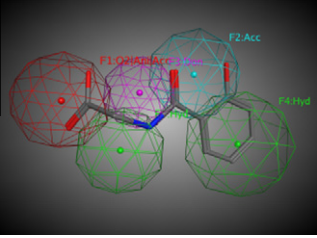
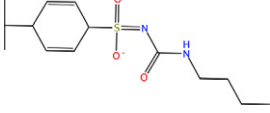
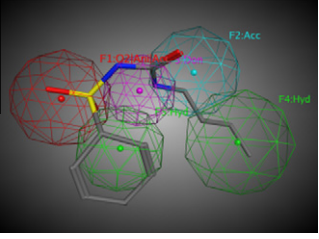
**Table 7**  
Results of screening of Oat1-related substrates<sup>16</sup> database with the pharmacophore of cluster 3 (5 features of 5)

#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	Transporter	System	$K_i$	RSMD, fit	2D structure	3D structure
1	4-Hydroxyphenyl-lactate*	$C_9H_9O_4^-$	181.1654	mOat1	<i>X. laevis</i>	$223 \pm 18$	0.1661 exact fit		
2	Pravastatin	$C_{23}H_{36}O_7$	424.5277	rOat1	LLC-PK1	52	1.2174 tail		

\* Metabolite, cluster 3.

**Table 8**  
Results of screening of Oat1-related substrates<sup>16</sup> database with the pharmacophore of cluster 4 (5 features of 5)

#	Metabolite, targeted	Molecular formula	Molecular weight (g/mol)	Trans-porter	System	$K_i$	RSMD, fit	2D structure	3D structure
1	1-Methylpyrenyl mercapturic acid	$C_{22}H_{17}NSO_3^-$	376.456	hOAT1	HEK293	14.5	0.5557 tail		
2	Betamipron	$C_{10}H_{11}NO_3$	193.1992	hOAT1	CHO S2	$23.6 \pm 3.10$	1.1314 exact fit		
3	Chlorpropamide	$C_{10}H_{13}ClN_2O_3S$	276.7398	rOat1	<i>X. laevis</i>	39.5	1.4377exact fit		
4	Hippurate	$C_9H_8NO_3^-$	178.1647	hOAT1 mkOat1 mOat1 rOat1	HeLa OK HEK293 HEK293 <i>X. laevis</i> LLC-PK1	$18.8 \pm 4.7$ $55.6 \pm 1.8$ $27.5 \pm 4.3$	1.1831exact fit		
5	N-Acetylaspartate	$C_6H_7NO_5^{2-}$	173.1235	mOat1	<i>X. laevis</i>	$840 \pm 260$	0.9475exact fit		

6	<i>N</i> -Acetyl-S-ethyl-L-cystein	C <sub>7</sub> H <sub>13</sub> NO <sub>3</sub> S	191.248	rOat1	<i>X. laevis</i>	1100 ± 500	1.1559exact fit	 
7	Nateglinide	C <sub>19</sub> H <sub>27</sub> NO <sub>3</sub>	317.4226	rOat1	<i>X. laevis</i>	9.2	1.6375exact fit	 
8	Salicylurate	C <sub>9</sub> H <sub>9</sub> NO <sub>4</sub>	195.1721	rOat1	<i>X. laevis</i>	11	1.2624 exact fit	 
9	Tolbutamide	C <sub>12</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub> S	270.3479	rOat1	<i>X. laevis</i>	55.5	1.1519 tail	 

1  $\mu\text{M}$  (see Fig. 7, cluster 1). Similar results were obtained with cluster 2; the exact-fit compound from Labotest LT00454450 (MW 187.121) did not show the expected binding affinity (Fig. 7, cluster 2). This is in contrast to the compound from Bionet 8X-0800 from the tail category, which showed some binding affinity at concentrations as low as 1 mM.

Two cluster-3-based compounds were tested in the exact-fit set and were obtained from Labotest (LT00005591 and LT00847268). Both of the compounds from the exact-fit set in the cluster-3-based, in contrast to cluster-1-based and cluster-2-based compounds, showed extremely strong mOat1 affinity in *Xenopus* oocytes. As shown in Fig. 7, Cluster 3, both compounds inhibited the uptake of 6CF (30  $\mu\text{M}$ ) at a concentration as low as 100  $\mu\text{M}$  for LT00005591 and between 1 and 10  $\mu\text{M}$  for LT00847268. One compound was tested in the tail group, for cluster-3-based, from Labotest LT00127613 (MW 269.297), which exhibited some binding affinity at around 100  $\mu\text{M}$  concentration. Compounds in cluster 4 in the exact-fit group were from Chembridge 5848164 (MW 237.303) and showed strong affinity towards mOat1 at 1 mM concentration similar to tail compound, from Chembridge 9004707 at concentration of 1 mM (see Fig. 7, cluster 4).

### 3.5. Screening of Oat1-related substrates database and selected fits of existing drugs

To study Oat1–drug interaction, we examined the known Oat1-related substrates<sup>16</sup> and created the 3D-conformation database, using the MOE Database Viewer and Conformation Import module of Conformations program. The database contains 3763 entries for 91 substrates. As a result of screening of this database with our pharmacophore hypotheses, 20 substrates were selected (see Tables 5–8) that belong to the following classes: (1) drugs, (2) toxins, and (3) metabolites including drug metabolites. Such results support the validity of the pharmacophore hypotheses and pave the road to possible design of lead compounds using structure-activity correspondences.

## 4. Discussion

Oat1 is one of the major multispecific drug, toxin, and metabolite transporters found in most barrier epithelia, including the kidney and choroid plexus. Knockout mice in which the Oat1 gene has been deleted accumulate a range of metabolites of biological and medical interest (such as uric acid and 3-hydroxybutyrate). We reasoned that drugs (e.g.,  $\beta$ -lactam antibiotics, HIV antivirals, NSAIDs, or diuretics) and toxins (e.g., mercurials) may 'hijack' Oat1-mediated metabolite transport processes. We therefore attempted to use the chemical structures of metabolites accumulating in the Oat1 knockout mouse to define the essential features of Oat1 substrates by our pharmacophore strategy. Conceivably, these pharmacophores could be used to screen chemical libraries to identify novel Oat1 drugs, for example inhibitors to extend drug half-lives or as templates for designing drug targeting Oat1-expressing tissues.

We defined the fingerprint features of Oat1-transported metabolites that can be used for construction of consensus pharmacophores and performed database screening. Furthermore, the power of ligand-based pharmacophore design, using the clusterization of these compounds (i.e., metabolites identified by targeted methods) was then used to separate distinct consensus groups. We found many compounds—including other metabolites, drugs, and toxins—already known to bind Oat1 based on previously published results in the literature fit reasonably with our pharmacophore models. The pharmacophores were then used to screen commercial

databases. Importantly, we were able to then demonstrate that several compounds selected by this screen inhibited binding of prototypical Oat1 substrates in a wet lab assay of Oat1 function.

## 5. Conclusions

We refined the approach for the ligand-based pharmacophore design in the case of basic compounds having various structures (i.e., blood-plasma and urine metabolites), and defined a set of molecular fingerprints that helped us to clusterize them. As a result of the study of targeted metabolites, 4 clusters were obtained with specific electron and geometrical features for each cluster. The best flexible alignments were also determined for each cluster, which allowed for the development of 4 consensus pharmacophore hypotheses. For each pharmacophore hypothesis, commercial databases provided by CCG (more than 600,000 compounds) were screened. The experimental results support the 4 pharmacophore hypotheses created for targeted metabolites of Oat1. As a result, several compounds were selected for each cluster, most of which showed a strong binding affinity towards the Oat1 transporter in *in vitro* experiments and could be used for refinement of pharmacophore hypotheses and further drug design. The designed pharmacophore hypotheses were also validated by screening of a database developed from the known OAT1-binding compound data that has been previously published.

The use of this computational method based on the targeted metabolite data has provided additional insight to the ligand-based approach that employed conformational analysis of known ligands. This approach has implications for developing novel compounds structurally similar to endogenous metabolites and for devised strategies for drug design.

## Acknowledgments

We thank the Chemical Computing Group for their generous support and responsiveness and acknowledge the editorial contributions of Dr. Kevin Bush from University of California, San Diego.

This work was supported, in whole or in part, by S.K.N. 5R01GM088824-03, 5R01NS062156-03, 5R01DK079784-04.

## References and notes

- Food and Drug Administration. Guidance for Industry: Drug Interaction Studies—Study Design, Data Analysis, and Implications for Dosing and Labeling [Internet]. U.S. Department of Health and Human Services (September 2006, updated April 30, 2009). <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072101.pdf>.
- Lopez-Nieto, C. E.; You, G.; Bush, K. T.; Barros, E. J.; Beier, D. R.; Nigam, S. K. *J. Biol. Chem.* **1997**, 272, 6471.
- Pavlova, A.; Sakurai, H.; Leclercq, B.; Beier, D. R.; Yu, A. S.; Nigam, S. K. *Am. J. Physiol. Renal Physiol.* **2000**, 278, F635.
- Miyazaki, H.; Sekine, T.; Endou, H. *Trends Pharmacol. Sci.* **2004**, 25, 654.
- Nigam, S. K.; Bush, K. T.; Bhatnagar, V. *Nat. Clin. Pract. Nephrol.* **2007**, 3, 443.
- Ahn, S. Y.; Nigam, S. K. *Mol. Pharmacol.* **2009**, 76, 481.
- Eraly, S. A.; Vallon, V.; Vaughn, D. A.; Gangoiti, J. A.; Richter, K.; Nagle, M.; Monte, J. C.; Rieg, T.; Truong, D. M.; Long, J. M.; Barshop, B. A.; Kaler, G.; Nigam, S. K. *J. Biol. Chem.* **2006**, 281, 5072.
- Truong, D. M.; Kaler, G.; Khandelwal, A.; Swaan, P. W.; Nigam, S. K. *J. Biol. Chem.* **2008**, 283, 8654.
- Ahn, S. Y.; Eraly, S. A.; Tsigelny, I.; Nigam, S. K. *J. Biol. Chem.* **2009**, 284, 31422.
- Chen, H.; Engkvist, O.; Blomberg, N. *Methods Mol. Biol.* **2011**, 685, 135.
- Labute, P. Quasar-Cluster: A Different View of Molecular Clustering. Electronic Publication: <http://www.chemcomp.com/J/cluster.htm> (Chemical Computing Group, Inc., Montreal, Canada), 1998.
- Labute, P.; Williams, C. J. *Med. Chem.* **2001**, 44, 1483.
- Gund, P. *Prog. Mol. Subcell. Biol.* **1977**, 5, 117.
- Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. *J. Comput. Aided Mol. Des.* **2008**, 22, 213.
- Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. *J. Med. Chem.* **2005**, 48, 2534.
- VanWert, A. L.; Gionfriddo, M. R.; Sweet, D. H. *Biopharm. Drug Dispos.* **2010**, 31, 1.