



Three-dimensional quantitative structure–activity relationship analyses of substrates of the human proton-coupled amino acid transporter 1 (hPAT1)

Iris Thondorf^{a,*}, Valerie Voigt^b, Sarah Schäfer^a, Sabine Gebauer^a, Katja Zebisch^b, Linda Laug^b, Matthias Brandsch^b

^a Institute of Biochemistry and Biotechnology, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Str. 3, D-06099 Halle, Germany

^b Biozentrum, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Str. 3, D-06099 Halle, Germany

ARTICLE INFO

Article history:

Received 16 May 2011

Revised 22 August 2011

Accepted 28 August 2011

Available online 5 September 2011

Keywords:

Amino acid transport

hPAT1

3D QSAR

CoMSIA

Virtual screening

ABSTRACT

The proton-coupled amino acid transporter hPAT1 has recently gained much interest due to its ability to transport small drugs thereby allowing their oral administration. A three-dimensional quantitative structure–activity relationship (3D QSAR) study has been performed on its natural and synthetic substrates employing comparative molecular similarity indices analysis (CoMSIA) to investigate the structural requirements for substrates and to derive a predictive model that may be used for the design of new pro-drugs. The cross-validated CoMSIA models have been derived from a training set of 40 compounds and the predictive ability of the resulting models has been evaluated against a test set of 10 compounds. Despite the relatively narrow range of binding affinities (K_i values) reliable statistical models with good predictive power have been obtained. The best CoMSIA model in terms of a proper balance of all statistical terms and the overall contribution of individual properties has been obtained by considering steric, hydrophobic, hydrogen bond donor and acceptor descriptors ($q^2_{cv} = 0.683$, $r^2 = 0.958$ and $r^2_{pred} = 0.666$). The 3D QSAR model provides insight in the interactions between substrates and hPAT1 on the molecular level and allows the prediction of affinity constants of new compounds. A pharmacophore model has been generated from the training set by means of the MOE (molecular operating environment) program. This model has been used as a query for virtual screening to retrieve potential new substrates from the small-molecule, 'lead-like' databases of MOE. The affinities of the compounds were predicted and 11 compounds were identified as possible high-affinity substrates. Two selected compounds strongly inhibited the hPAT mediated L-[³H]proline uptake into Caco-2 cells constitutively expressing the transport protein.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The proton-coupled amino acid transporter hPAT1 was characterized on a molecular level in 2002 and was assigned to family 36 of the solute carriers (SLC36A1).¹ Its cDNA was isolated first from rat brain² and subsequently from mouse intestine¹ and human Caco-2 cells.³ hPAT1 is expressed in the apical membrane of intestinal epithelial cells, but hPAT1 mRNA was also detected in brain, colon, liver and lung.^{3,4} The transporter was finally shown to be the protein entity of the H⁺/amino acid symport system functionally first described more than 10 years ago.^{5–11} The system acts as a H⁺-dependent, Na⁺-independent, and rheogenic transporter. Compared to the isoform transporter hPAT2, it can be classified as a low affinity, high capacity transporter.¹²

The primary substrates for hPAT1 are amino acids such as glycine, L-proline, and L-alanine.³ hPAT1 appears to represent the major route by which small, unbranched, neutral amino and imino

acids are absorbed after intestinal protein digestion.^{3,8} An unusual feature of the hPAT system is that it also transports D-amino acids such as D-alanine, D-serine, and D-cycloserine with affinities similar or even higher than those for glycine or L-proline.^{8,11,13} PAT1 proteins also transport GABA and important osmolytes such as betaine and taurine.^{3,14} Recently, hPAT1 gained much interest because of its ability to transport drugs such as inhibitors of collagen biosynthesis (e.g., 3,4-dehydro-D,L-proline, L-azetidine-2-carboxylic acid), the anticraving agent 3-amino-1-propanesulfonic acid, the antihyperglycemic creatine analog β-guanidinopropionic acid and the GABA derivative vigabatrin, thereby allowing their oral administration.^{15–17} Therefore, several studies have focused on the carrier's substrate specificity.^{18,19}

Thus, substrates should possess a free carboxylic group which can be replaced by a sulfonic acid group at the expense of a moderate loss in affinity.[†] Substitutions at the amino group are well tolerated. The distance between the two functional groups can be up to

* Corresponding author. Tel./fax: +49 345 552 4862.

E-mail address: iris.thondorf@biochemtech.uni-halle.de (I. Thondorf).

[†] In the following the expression 'substrate' is used for molecules with $K_i < 30$ mM. A few of them might be inhibitors.

three methylene units. The transporter seems to be only moderately stereoselective. Four- to six-membered heterocyclic amino acids are accepted by the protein while the side chain size of acyclic amino acids is an affinity limiting factor. Aromatic α -amino acids display low affinity. In contrast, tryptophan and its derivatives tryptamine and serotonin bind to hPAT1 with potencies similar to the prototype substrates but inhibit the transport function of the carrier protein.²⁰

Although many experimental data regarding the substrate specificity of hPAT1 are available, its molecular basis is still poorly understood. Three-dimensional quantitative structure–activity relationship (3D QSAR) studies have been developed as a tool to understand and explain the mechanism of drug action at the molecular level since it relates its three-dimensional structure with its affinity to the receptor. In this study, we report on the comparative molecular similarity indices analysis (CoMSIA)²¹ using data of 50 substrates of hPAT1 to derive a mathematical model for structure–activity relationships and to explore the key structural features determining the binding. Recently reported 3D QSAR models for the interaction of hPEPT1 and hPEPT2 with di- and tripeptides as well as β -lactam antibiotics have demonstrated the feasibility of this approach for transport proteins.^{22–24}

2. Results and discussion

2.1. Selection of the training set, conformational analysis and molecular alignment

A training set and a test set consisting of 40 and 10, respectively, natural and nonnatural amino acids as well as their derivatives have been chosen for the CoMSIA studies. Their affinity data have been determined in our group in competition assays using L-[³H]proline as the reference substrate at Caco-2 cells as described previously.¹⁵ The measured K_i values (Table 1) range from 1.2 to >30 mM (higher values extrapolated beyond the measurement range). We are well aware that it would be desirable to use a training set which covers K_i values in the range of three or more orders of magnitude, but so far no substrates of hPAT1 with K_i values below 1 mM have been identified.

A crucial step in a 3D QSAR analysis is the alignment of molecules in space. The biologically active conformation of the compounds should be aligned in a manner representing a similar binding mode. This is usually done by overlay of common structural features using the lowest energy conformer of each structure or, if available, by using the bioactive conformation found in the X-ray crystal structure as template. Unfortunately, X-ray crystallographic structural information about hPAT1 and binding of its substrates is not available and it might not be available in the near future. Moreover, since the active compounds in our training and test sets share only two common recognition elements (the amino/imino nitrogen atoms and the carboxyl groups) we had to resort to a different way for the alignment. Thus, the lowest energy conformer of L-proline was used as a template since L-proline is both conformationally rigid and displays a high affinity towards hPAT1. Rigid α -amino acids such as L-azetidine-2-carboxylic acid were fitted onto the template using the amino/imino nitrogen, carboxyl carbon and C_α atoms. In the case of nonnatural ω -amino acids a centroid of the backbone carbon atoms was used instead of the C_α atom for the rms fitting procedure. Flexible compounds were first submitted to a systematic conformational search, the resulting structures were minimized and the conformer with a low rms value and low steric energy was chosen for the alignment. The alignment of the training set is shown in Figure 1.

2.2. CoMSIA studies

Using the five CoMSIA descriptors (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor) as independent

variables and the pK_i values as dependent variables, partial least squares (PLS) regression analyses were carried out. First, the optimum number of components was determined with the SAMPLS method.²⁵ Leave-one-out (LOO) cross validation was then used to evaluate the internal predictive ability of the models. To further assess the stability and the robustness of the models 20 runs of the 'leave-five-out' and 'leave-20%-out' cross-validation were performed. The bootstrapping procedure²⁶ was used to validate each model. Finally, an external validation was performed on a test set comprising 10 compounds (Table 1, compounds 41–50).

We have generated CoMSIA models using the five descriptors in different combinations. The statistical parameters of CoMSIA along with the contribution of the descriptor variables are summarized in Table 2 for the best five models in terms of high q^2_{cv} and low S_{PRESS} values. All five 3D QSAR models are characterized by meaningful correlative and predictive properties. The q^2_{cv} values are all well above 0.5,²⁷ the r^2_{ncv} values above 0.9 and the standard errors of estimate (SEE) below 0.2 which suggest their significance in terms of explaining the observed biological activity. It is, however, obvious that neglecting the hydrophobic descriptor results in somewhat less significant statistics (Table 2, model SEA). The same is true when the steric contribution is omitted (data not shown). The models generated with steric, (electrostatic), hydrophobic and H-bond acceptor descriptors (models SEHA and SHA in Table 2) have higher q^2_{cv} , q^2_{L50} and $q^2_{L20\%O}$ values than the rest of the models but have lower r^2_{ncv} , bootstrapped q^2 and F values. The SEE is also somewhat higher. Incorporation of all the fields and the combination of steric, hydrophobic, H-bond donor and acceptor descriptors yield models with a proper balance of all the statistical terms (models SEHDA and SHDA in Table 2). They also show a good external predictive ability r^2_{PRED} of 0.708 and 0.699. The comparison of the individual components of both models reveals that the electrostatic field in the SEHDA model accounts for only 8% of the total contribution. Therefore, we have selected the model with steric, hydrophobic, H-bond donor and acceptor fields (SHDA) as the best model to explain the biological data. The linear regression plot for this model is shown in Figure 2. It indicates that the measured and predicted affinities vary less than a 0.5 logarithmic unit for the training set,²⁸ the highest residual being 0.38 logarithmic units in the case of L-alanine 12, while in the test set the highest deviation has been observed for taurine 46 (0.44 logarithmic units, Table 1).

2.3. Graphical interpretation of the results

The CoMSIA steric fields are shown in Figure 3. The green contours represent regions of high steric tolerance while the yellow contours denote regions of unfavorable steric effects. The sterically favored green contours enclose side chains of high affinity guanidino derivatives (2, K_i = 1.3 mM, 3, K_i = 1.5 mM and 13, K_i = 3.2 mM) and parts of the pyrrolidine ring of D-proline (1, K_i = 1.2 mM) as well as the methyl substituents at secondary amino groups, for example, in betaine (6, K_i = 1.9 mM), sarcosine (7, K_i = 1.9 mM), and N-methyl-L-alanine (8, K_i = 2.2 mM). The latter is consistent with the observation that among the amino acid substrates N-alkylated amino acids possess particular high affinities. The sterically unfavored yellow region is occupied by the side chains of larger amino acids such as L-leucine, L-cysteine, L-norvaline, and L-isoleucine which possess K_i values well above 35 mM. To illustrate the steric effect on the binding of substrates Figure 3 shows exemplarily the positioning of the high affinity substrate D-pipecolic acid (11, K_i = 2.5 mM) within the favored field and of the medium affinity substrate L-pipecolic acid (21, K_i = 11 mM) with molecule parts in the unfavored yellow region.

In the hydrophobic property maps, shown in Figure 4, the orange contour indicates areas where increasing hydrophobicity enhances the affinity for the carrier. This region partly overlaps with

Table 1Structures and binding affinities of the training set (**1–40**) and the test set compounds (**41–50**) used for the CoMSIA studies (K_i in mM)

No	Compound	Structure	K_i	Ref.	$pK_{i,act}$	$pK_{i,pred.}$	Residual
1	D-Proline		1.2	15	−0.15	−0.35	0.20
2	β-Guanidinopropionic acid		1.3	16	−0.11	−0.15	0.04
3	Guanidinoacetic acid		1.5	16	−0.15	−0.07	−0.07
4	L-Proline		1.6	15	−0.20	−0.53	0.33
5	L-Azetidine-2-carboxylic acid		1.8	15	−0.26	−0.33	0.08
6	Betaine		1.9	19	−0.28	−0.12	−0.16
7	Sarcosine		1.9	19	−0.28	−0.34	0.06
8	N-Methyl-L-alanine		2.2	15	−0.34	−0.35	0.01
9	Isonipecotic acid		2.3	^a	−0.36	−0.71	0.35
10	β-Alanine		2.4	19	−0.38	−0.36	−0.02
11	D-Pipecolic acid		2.5	19	−0.40	−0.33	−0.07
12	L-Alanine		2.8	19	−0.45	−0.83	0.38
13	Creatine		3.2	^a	−0.51	−0.38	−0.13
14	cis-4-Hydroxy-D-proline		3.3	15	−0.52	−0.36	−0.16
15	γ-Aminobutyric acid		4.0	20	−0.60	−0.55	−0.06
16	D-Cycloserine		4.4	19	−0.64	−0.60	−0.04
17	Homotaurine		7.1	15	−0.85	−0.94	0.08
18	L-Cycloserine		7.4	19	−0.87	−0.83	−0.04
19	L-Thiaproline		7.9	19	−0.90	−0.70	−0.19

(continued on next page)

Table 1 (continued)

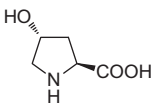
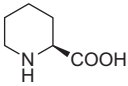
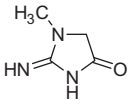
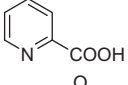
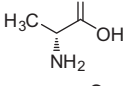
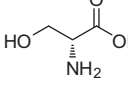
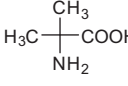
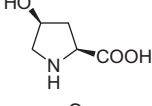
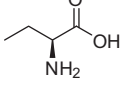
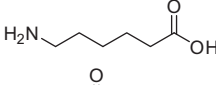
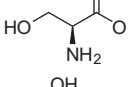
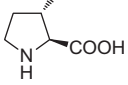
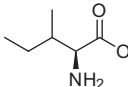


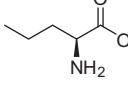
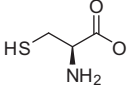

No	Compound	Structure	K_i	Ref.	$pK_{i,act}$	$pK_{i,pred.}$	Residual
20	<i>trans</i> -4-Hydroxy-L-proline		9.0	15	−0.95	−0.95	−0.01
21	L-Pipecolic acid		11	15	−1.04	−0.95	−0.09
22	Creatinine		15	^a	−1.18	−1.26	0.09
23	Picolinic acid		18	^a	−1.26	−1.18	−0.07
24	D-Alanine		19	^a	−1.28	−1.11	−0.16
25	D-Serine		22	^a	−1.34	−1.41	0.07
26	α -Aminoisobutyric acid		23	19	−1.36	−1.49	0.13
27	<i>cis</i> -4-Hydroxy-L-proline		30	15	−1.48	−1.39	−0.08
28	L- α -Aminobutyric acid		35	19	−1.54	−1.29	−0.25
29	Aminohexanoic acid		35	^a	−1.54	−1.38	−0.16
30	L-Serine		35	^a	−1.54	−1.58	0.04
31	<i>trans</i> -3-Hydroxy-L-proline		37	^a	−1.57	−1.61	0.04
32	L-Isoleucine		37	^a	−1.57	−1.65	0.08
33	Pyrrolidine		38	^a	−1.58	−1.45	−0.13
34	Thiazolidine		44	^a	−1.64	−1.54	−0.11
35	L-Norvaline		46	^a	−1.66	−1.51	−0.15
36	L-Cysteine		46	^a	−1.66	−1.60	−0.06
37	Piperidine		54	^a	−1.73	−1.87	0.14

Table 1 (continued)

No	Compound	Structure	K_i	Ref.	$pK_{i,act}$	$pK_{i,pred.}$	Residual
38	γ -Hydroxybutyric acid		65	^a	−1.81	−1.95	0.14
39	L-Leucine		100	^a	−2.00	−2.06	0.06
40	D-Valine		100	^a	−2.00	−1.92	−0.08
41	N-Methylaminoisobutyric acid		2.2	19	−0.34	−0.54	−0.20
42	(1R,2S,5S)-3-Azabicyclo[3.1.0] hexane-2-carboxylic acid		2.7	19	−0.43	−0.66	−0.23
43	R-Vigabatrin		4.8	^a	−0.68	−0.61	0.07
44	Glycine		5.1	15	−0.71	−0.52	0.19
45	Isoguvacine		5.8	^a	−0.76	−0.74	0.03
46	Taurine		7.8	19	−0.89	−0.45	0.44
47	m-Aminobenzoic acid		12	^a	−1.08	−1.33	−0.25
48	L-Oxo-thiazolidine-4-carboxylic acid		46	^a	−1.66	−1.91	−0.25
49	L-Prolinol		50	^a	−1.70	−1.30	0.40
50	D-Norvaline		100	^a	−2.00	−1.71	0.29

^a This work.

the sterically favored area shown in Figure 3. The orange contour encloses small side chains of L-amino acids and the methylene groups of heterocyclic L-amino acids as shown exemplarily for the prototype substrate L-proline (**4**, $K_i = 1.6$ mM) in Figure 4 on the left. A blue contour situated near the C-terminus of the molecular alignment (see Fig. 1) denotes areas where increasing hydrophilicity is beneficial for affinity. It is occupied by parts of the carboxylic groups of some high-affinity substrates such as guanidinoacetic acid (**3**, $K_i = 1.5$ mM, Fig. 4b), betaine (**6**, $K_i = 1.9$ mM) and isonipecotic acid (**9**, $K_i = 2.3$ mM) as well as the hydrophilic part of the isoxazoline ring of L-cycloserine (**18**, $K_i = 7.4$ mM).

The H-bond donor and acceptor fields shown in Figures 5 and 6 describe the spatial situation for hydrogen bonding interactions between the substrate and the carrier protein. The purple contour in Figure 5 marks an area where the presence of a hydrogen-bond donor decreases affinity, most probably due to a lack of the

corresponding acceptor group in the protein. In contrast, the field colored in cyan represents a region where an acceptor group at the protein forms the counterpart to the substrate's donor group. The occurrence of these fields can be easily interpreted by the fact that the amino acids having a high affinity to PAT1 are either imino acids with the imino protons pointing to the cyan field or N-methylated amino acids where a methyl group points into the direction of the violet isopleth. This is illustrated in Figure 5 for L-azetidine-2-carboxylic acid (**5**, $K_i = 1.8$ mM) and N-methyl-L-alanine (**8**, $K_i = 2.2$ mM), respectively, both high affinity substrates of hPAT1.

The green–blue contour in Figure 6 denotes the presence of hydrogen bond donors at the receptor site while the region colored in red indicates that acceptor groups in the substrates have no counterparts at the receptor site. One or both oxygen atoms of the carboxylic group of most substrates point to either of the two acceptor sites as shown in Figure 6 for β -alanine (**10**,

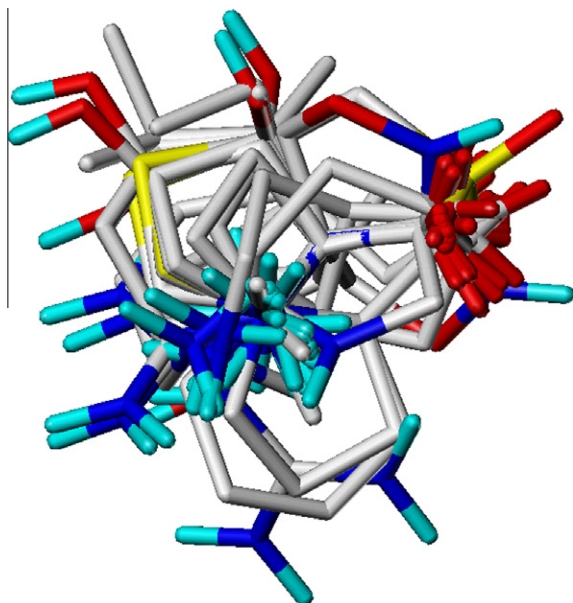


Figure 1. Alignment of the compounds of the training set. Nonpolar hydrogen atoms have been omitted for clarity.

Table 2
Summary of the CoMSIA results for the different models

	SEHDA ^a	SHDA	SEHA	SEA	SHA
q_{cv}^2 ^b	0.651	0.683	0.704	0.619	0.695
SEP ^c	0.387	0.369	0.357	0.405	0.357
N ^d	6	6	6	6	5
q_{L50}^2 ^e	0.638	0.632	0.651	0.588	0.670
SEP _{L50} ^f	0.395	0.393	0.387	0.421	0.372
$q_{L20\%O}^2$ ^g	0.584	0.598	0.614	0.528	0.621
SEP _{L20\%O} ^h	0.414	0.411	0.412	0.454	0.399
r_{ncv}^2 ⁱ	0.959	0.958	0.939	0.921	0.930
SEE ^j	0.133	0.135	0.162	0.184	0.171
F value	127.518	124.112	84.535	64.372	89.829
r_{pred}^2 ^k	0.708	0.699	0.597	0.554	0.587
q_{bs}^2 ^l	0.970	0.973	0.958	0.946	0.946
SEE _{bs} ^l	0.112	0.106	0.131	0.148	0.147
Contribution (%)					
S ^a	15.7	19.5	20.5	43.0	22.8
E	8.0	—	11.4	19.7	—
H	34.6	30.7	45.6	—	49.0
D	23.4	24.4	—	—	—
A	18.3	25.3	22.5	37.3	28.2

^a S: steric, E: electrostatic, H: hydrophobic, D: H-bond donor, A: H-bond acceptor.

^b Cross-validated q^2 by leave-one-out method.

^c Standard error of prediction by leave-one-out method.

^d Number of components.

^e Cross-validated q^2 by the leave-five-out method.

^f Standard error of prediction by the leave-five-out method.

^g Cross-validated q^2 by the leave-20%-out method.

^h Standard error of prediction by the leave-20%-out method.

ⁱ Noncross-validated r^2 .

^j Standard error of estimate.

^k Predictive r^2 .

^l From 100 bootstrapping runs.

$K_i = 2.4$ mM) thus highlighting the observation that a complete lack of the carboxylic group results in almost a complete loss in the affinity to the transporter (cf., **41** and **42** in Table 1). The red property field is generated by hydroxyl groups of low affinity compounds such as L-serine (**30**, $K_i = 35$ mM), *cis*-4-hydroxy-L-proline (**27**, $K_i = 30$ mM), *trans*-3-hydroxy-L-proline (**31**, $K_i = 37$ mM) as well as by γ -hydroxybutyric acid (**38**, $K_i = 65$ mM), respectively.

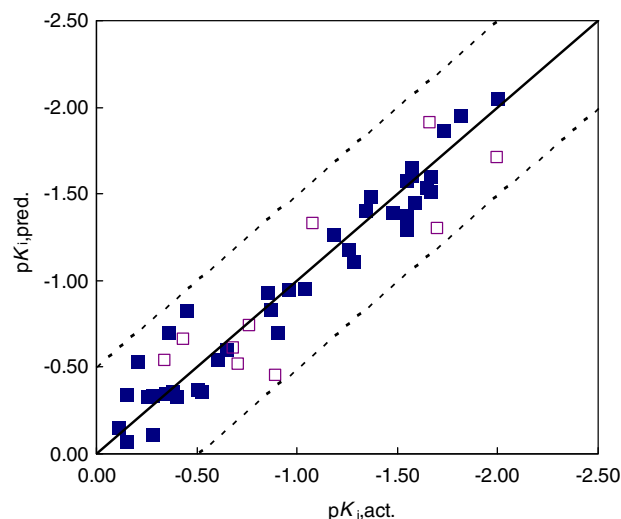


Figure 2. Predicted versus measured activity of molecules in the training set (■) and the test set (□). Dotted lines denote deviations of 0.5 logarithmic units.

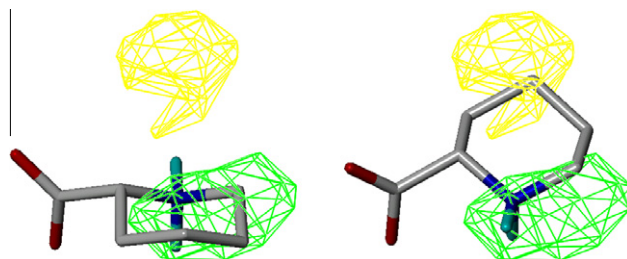


Figure 3. CoMSIA stdev*coeff contour plots for steric properties: D-pipecolic acid (**11**, left) and L-pipecolic acid (**21**, right). Sterically favorable regions are colored in green and disfavored regions are in yellow.

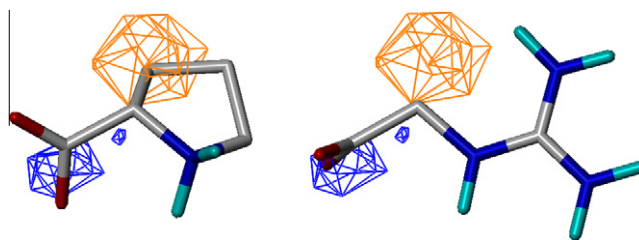


Figure 4. CoMSIA stdev*coeff contour plots for hydrophobic properties: L-proline (**4**, left) and guanidinoacetic acid (**3**, right). The orange contour denotes an area favorable for hydrophobic groups and the blue contour indicates an area where hydrophilic groups are favored.

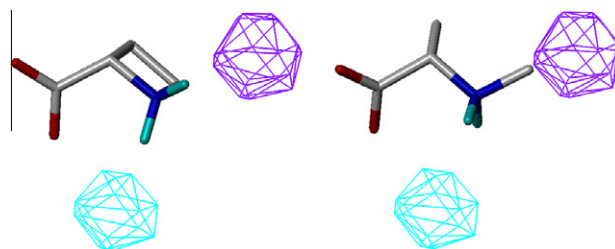


Figure 5. CoMSIA stdev*coeff contour plots for hydrogen bond donor properties: L-azetidine-2-carboxylic acid (**5**, left) and N-methyl-L-alanine (**8**, right). The cyan contour denotes a region where hydrogen-bond acceptor groups of the substrate binding site of the carrier are situated to form hydrogen bonds with the ligand. The violet isopleth characterizes a region that is unfavorable for hydrogen bonding.

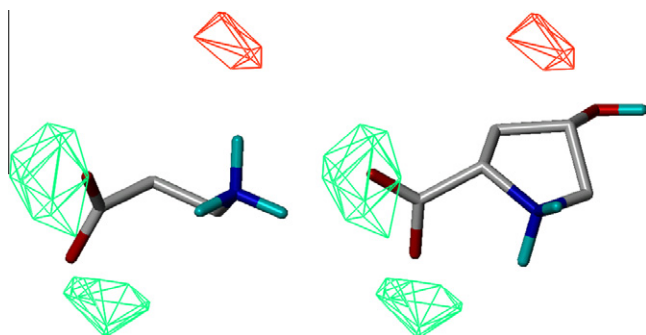


Figure 6. CoMSIA stdev*coeff contour plots for hydrogen bond acceptor properties: β -alanine (**10**, left) and *cis*-4-hydroxy-L-proline (**27**, right). The green–blue isopleths mark regions where hydrogen-bond donor groups of the substrate binding site are situated to form hydrogen bonds with the ligand. The red isopleth characterizes a region that is unfavorable for hydrogen bonding.

2.4. Virtual screening

The alignment of the training set molecules has been converted to the MOE (molecular operating environment²⁹) program and a pharmacophore model (Fig. 7) has been generated with a cation & donor center, an O2 centroid and a hydrophobic centroid. The cation & donor center defines the position of the protonated nitrogen atom, the hydrophobic centroid coincides with the sterically favored and the hydrophobic field obtained by the CoMSIA analyses and the O2 centroid reflects the necessity of a carboxylic group for the affinity towards the transporter. Since amino acids with longer side chains do not display a high affinity to hPAT1 we have added an excluded volume at the ends of the side chains of, for example, L-leucine and L-valine to the pharmacophore model. The pharmacophore pattern was then used to browse the small-molecule ('lead-like') databases distributed with MOE. From these multi-conformer databases comprising approximately 650,000 molecules we have obtained 1269 compounds that matched all of the features present in the pharmacophore model. Since the substrates of hPAT1 are low-molecular weight compounds large compounds (MW >155 Da) were excluded from the dataset. The remaining 30 compounds were converted into a Sybyl³⁰ database and their affinity towards hPAT1 was predicted using the model established by the CoMSIA analysis of the training

set. From this dataset 11 top compounds were predicted to display an affinity to the carrier which is comparable or even higher than the substrates known to date (cf., Table 3 and Supplementary data). They are mostly characterized either by a guanidinium group incorporated in a cyclic ring system (**51**, **52**, **57**) or by a tertiary, protonable nitrogen which is part of a heterocycle (**53–56**, **58**, **59**). It is interesting to note that of the compounds investigated within the training and test sets, respectively, none exhibited the latter structural pattern.

The most prominent compound obtained from virtual screening was arecaidine **55**, an alkaloid of the areca nut palm. Another compound which is easily available is 3-azetidinecarboxylic acid **61**. Therefore, we have decided to test the interaction of these compounds with hPAT1 in greater detail. hPAT1 mediated L-[³H]proline uptake in Caco-2 cells was measured in the presence of increasing concentrations (0.316–31.6 mM) of arecaidine and 3-azetidinecarboxylic acid. In addition, we have studied the effect of arecoline, the methylester of arecaidine which also occurs in the areca nut palm. The effects on the L-[³H]proline uptake are shown in Figure 8. From these curves K_i values of 6.1 mM (arecaidine), 2.5 mM (3-azetidinecarboxylic acid) and > 35 mM (arecoline) were derived. These experimental results correspond well with the results of the predictions (residuals about 0.3 logarithmic units).

3. Conclusion

In summary, we have established a 3D QSAR model to rationalize and understand the binding affinities of substrates to the human proton-coupled amino acid transporter hPAT1. The model has been validated for its stability and robustness by group validation and bootstrapping techniques and for its predictive ability by using an external test set. By the combination of four CoMSIA contour maps, that is, steric, hydrophobic/hydrophilic, hydrogen bond donor, and hydrogen bond acceptor, we were able to explain the affinity differences of structurally similar compounds. Computational screening of 'lead-like' molecule databases using a pharmacophore model derived from the CoMSIA contour maps have allowed the identification of a new scaffold for possible substrates of hPAT1 consisting of a N-substituted nitrogen heterocycle. The affinity data towards hPAT1 measured for two compounds obtained from the virtual screening show a good correlation between predicted and experimental data.

4. Methods

4.1. Data set for analysis

Fifty molecules selected for the present study have been taken from our laboratory where their affinity data were measured using the same assay.^{15,16,19,20} The structures of the compounds and their biological data are given in Table 1. The CoMSIA models were generated using a training set of 40 molecules (No. **1–40** in Table 1). The predictive power of the resulting models has been evaluated using a test set of 10 compounds (No. **41–50** in Table 1). The affinity constants K_i cover an evenly spread range of binding affinities from 1.2 mM to 100 mM. For the CoMSIA studies, $\log 1/K_i$ values have been used.

All computational studies have been performed with the Sybyl 8.1 software.³⁰ The molecular structures were constructed either manually from the fragment database of Sybyl or by the modification of crystal structures, respectively. In order to adapt the structures to pH 6, the pH which was adjusted for the determination of the affinity constants, N-terminal amino groups were considered to be protonated and all carboxylic groups to be deprotonated. Atomic point charges were calculated with the Gasteiger–Hückel

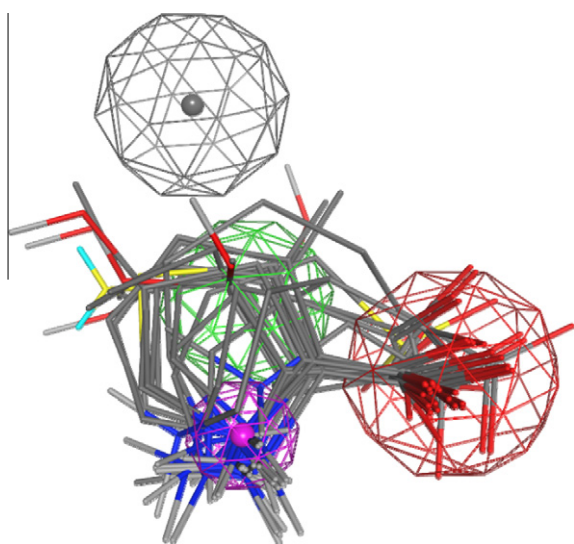
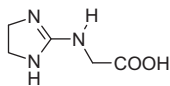
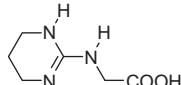
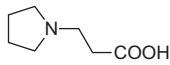
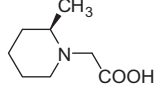
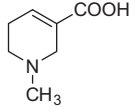
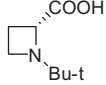
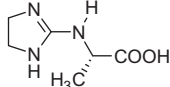
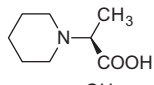
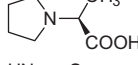
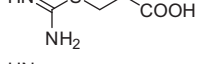
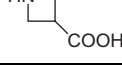


Figure 7. Pharmacophore model used for virtual screening consisting of a cation and donor center (violet), the centroid of the carboxyl oxygen atoms (red), a hydrophobic centroid (green) and an excluded volume (gray).

Table 3
Possible substrates of hPAT1 obtained by virtual screening

No	MOE Compound identifier	Chemical name	Structure	$K_{i, \text{pred}}$ [mM]
51	ASINEX BAS 03182941	<i>N</i> -(4,5-Dihydro-1 <i>H</i> -imidazol-2-yl)-glycine		0.6
52	ASINEX BAS 03182945	<i>N</i> -(1,4,5,6-Tetrahydro-2-pyrimidinyl)-glycine		0.8
53	ASINEX BAS 07870488	3-(<i>N</i> -Pyrrolidinyl)-propionic acid		1.7
54	ASINEX BAS 09000050	(<i>R</i>)-2-(2-Methylpiperidin-1-yl)-acetic acid		1.8
55	CHEMBRIDGE 5132242	1,2,5,6-Tetrahydro-1-methyl-3-pyridinecarboxylic acid (arecaidine)		3.0
56	ASINEX BAS 05338789	(<i>R</i>)-1-(1,1-Dimethylethyl)-2-azetidine carboxylic acid		3.4
57	ASINEX BAS 06502532	(<i>S</i>)- <i>N</i> -(4,5-Dihydro-1 <i>H</i> -imidazol-2-yl)-alanine		4.2
58	ASINEX BAS 05947733	(<i>S</i>)-2-(Piperidin-1-yl)-propionic acid		4.4
59	CHEMICAL BLOCK A2494/0106048	(<i>S</i>)-2-(<i>N</i> -Pyrrolidinyl)-propionic acid		4.5
60	BIONET MS-0192	3-[(Aminoiminomethyl)-thio]-propanoic acid		5.0
61	BIONET MS-0192	3-Azetidinecarboxylic acid		5.3

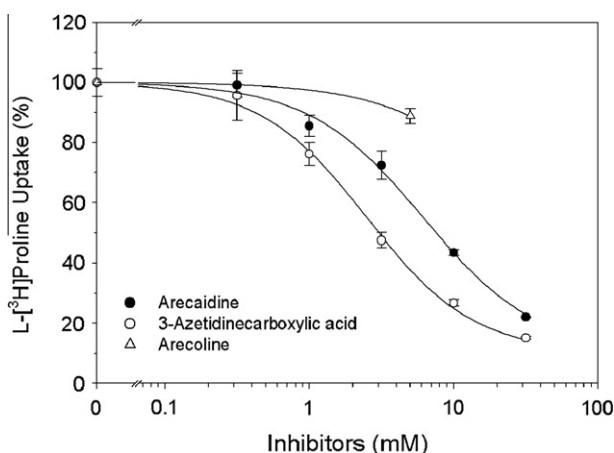


Figure 8. Inhibition of L -[^3H]proline uptake in Caco-2 cells. Uptake of L -[^3H]proline (10 nM; pH 6.0, readjusted if necessary) was measured at increasing concentrations of unlabeled arecaidine, 3-azetidinecarboxylic acid and arecoline. L -[^3H]Proline uptake measured in the absence of inhibitors (550.6 ± 19.9 fmol/10 min per mg protein) was taken as 100%. Values are mean \pm S.E.M, $n = 4$.

method.³¹ Each structure was geometry-optimized using the Tripos force field³² with a dielectric constant of $\epsilon = 80$ to mimic the

aqueous environment until the rms gradient was less than 0.001 kcal/(mol Å). Molecules with rotatable single bonds were submitted to a systematic conformational search by rotating the bonds from 0° to 359° in 30° increments. The resulting structures were subsequently minimized as described before and stored in conformer databases.

The molecular alignment has been performed using the minimum energy conformation of L -proline as template. Rigid molecules and all structures from the conformer databases were manually superimposed onto this template by rms fitting using the FIT command of Sybyl. Where possible, the amino nitrogen, the carboxyl carbon and the C_α atom were superimposed. Otherwise, a centroid was generated from the backbone atoms and used instead of the C_α atom for the fitting procedure. From the conformer databases the structure with the lowest rms value was chosen. The superimposed structures were stored in a molecule database for subsequent CoMSIA analysis.

4.2. PLS and CoMSIA analyses

To derive 3D QSAR models, different combinations of the five CoMSIA descriptors (steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor) were used as independent variables and the pK_i as the dependent variable. The descriptors were scaled to each other by the standard scaling option. The physicochemical

properties were examined using a sp^3 carbon atom with a 1 Å radius and a charge, hydrophobicity, and hydrogen bond property of +1, respectively, on a regularly spaced grid of 1 Å. The lattice dimensions were selected with a sufficiently large margin (>4 Å) to enclose all aligned molecules. The value of the attenuation factor α was set to 0.3.

Regression analyses have been performed using the Partial Least Squares (PLS)³³ analyses as implemented in the Sybyl 8.1 package. Initially, to obtain the optimum number of components cross-validated runs were done using a quicker PLS algorithm, the SAMPLS method (Sample-distance Partial Least Squares).²⁶ The optimum number of components was taken as the number to increase q_{cv}^2 by ~5% from the model with one fewer component rather than the default Sybyl estimate which is based on the highest q_{cv}^2 value. Subsequent cross-validations through the leave-one-out, leave-five-out and leave-20%-out procedures were conducted using this number of components. The L50 and L20%O procedures were repeated 20 times for each combination of descriptor variables to obtain statistically reliable results. To improve the signal-to-noise ratio and to reduce the computation time all crossvalidation runs were performed selecting a 2.0 kcal/mol energy column filter. Final PLS runs without cross-validation were then performed to estimate r^2 . Additionally, a bootstrapping analysis²⁷ was performed for 100 runs in order to estimate the confidence limits for the parameters. The results of the PLS runs are summarized in Table 2. To validate the derived CoMSIA model and to assist selection among various CoMSIA models, biological activities of an external test set of ten compounds (Table 1) with known affinity data were predicted using the models derived from the training set. From these predicted activities the predictive r^2 value, which is usually considered as the most appropriate parameter to evaluate the predictive power of the CoMSIA model, was calculated using the equation:

$$r_{pred}^2 = 1 - (S_{PRESS}/SD)$$

where S_{PRESS} is the sum of the squared deviations between the observed and predicted activities of the test set molecules and SD is the sum of the squared deviations between the biological activities of the test set and mean activities of the training set molecules.

The CoMSIA fields shown in Figures 3–6 have been contoured by their actual values chosen as follows: 0.002 and –0.0055 for favorable and unfavorable steric fields, 0.01 and –0.0055 for hydrophobic fields, 0.0062 and –0.007 for favorable and unfavorable donor fields as well as 0.0055 and –0.004 for favorable and unfavorable acceptor fields, respectively.

4.3. Virtual screening

For the virtual screening the conformer database with the molecules aligned in Sybyl was first converted into the MOE (molecular operating environment) program. The alignment was then refined using the flexible alignment option of MOE with the refine existing alignment command. By a pharmacophore query based on the prototype substrate L-proline the pharmacophore features cation & donor, an O2 centroid as well as a hydrophobic centroid were identified. Evaluation of these pharmacophoric elements against the active compounds of the database have yielded a pharmacophore consensus of 100% for cation & donor, 100% for the O2 centroid and 59% for the hydrophobic centroid. In order to restrict the virtual screening to small molecules without long side chains, we added an excluded volume in the region of the side chains of the low-affinity amino acids such as L-isoleucine, L-norvaline and L-leucine as a sphere with a radius of 1.4 Å to the pharmacophore model. Using the MOE pharmacophore search module, the eight multi-conformational lead-like databases distributed with MOE

were then screened for compounds with the desired features. The dataset of the virtual screening was reduced by applying a threshold of 155 Da for the molecular weight and subsequently exported to the Sybyl program. Finally, we predicted the affinity of the compounds by means of the CoMSIA model derived from the training set of the compounds.

4.4. Cell culture and L-[³H]proline uptake measurements

Caco-2 cells (passages 20–42) obtained from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany) were cultured in minimum essential medium (PAA, Cölbe, Germany) supplemented with 10% fetal bovine serum (Biochrom, Berlin, Germany), 1% nonessential amino acid solution (PAA, Cölbe, Germany) and gentamicin (50 µg/ml) as described earlier.^{15,16,19,20} Cells at 80–90% density were released using trypsin/EDTA (0.05 %/0.02%) and seeded in 35 mm disposable petri dishes (Sarstedt, Nümbrecht, Germany). With a starting cell density of 0.8×10^6 cells per dish, the cultures reached confluence within 24 h. Medium was replaced every second day and the day before the uptake experiment took place. Uptake experiments were performed on the 7th day after seeding. L-[³H]Proline (spec. radioactivity 50 Ci/mmol; Perkin Elmer, Massachusetts, USA) uptake was measured as described previously,¹⁵ using uptake buffer (1 ml) containing 25 mM MES/Tris (pH 6.0) with 140 mM NaCl, 5.4 mM KCl, 1.8 mM CaCl₂, 0.8 mM MgSO₄, 5 mM glucose, 10 nM L-[³H]proline and concentrations of unlabeled compounds. Arecaidine was purchased from AlfaAesar (Karlsruhe, Germany), 3-azetidinecarboxylic acid, arecoline and all substances and chemicals for buffer solutions were bought from Sigma–Aldrich (Taufkirchen, Germany). After a 10 min incubation time, cells were quickly washed four times with ice-cold buffer, solubilized and prepared for liquid scintillation spectrometry. Protein was determined according to the procedure of Bradford. Results are given as means ± S.E.M. Nonlinear regression analysis and calculation of inhibition constants (K_i) from IC₅₀ values were done as described.^{15,16}

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (grant BR 2430/4-3).

Supplementary data

Supplementary data (survey of the compounds obtained by virtual screening) associated with this article can be found, in the online version, at [doi:10.1016/j.bmc.2011.08.058](https://doi.org/10.1016/j.bmc.2011.08.058).

References and notes

- Boll, M.; Foltz, M.; Rubio-Aliaga, I.; Daniel, H. *J. Biol. Chem.* **2002**, 277, 22966.
- Sagné, C.; Agulhon, C.; Ravassard, P.; Darmon, M.; Hamon, M.; El Mestikawy, S.; Gasnier, B.; Giros, B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 7206.
- Chen, Z.; Fei, Y. J.; Anderson, C. M.; Wake, K. A.; Miyauchi, S.; Huang, W.; Thwaites, D. T.; Ganapathy, V. *J. Physiol.* **2003**, 546, 349.
- Anderson, C. M.; Grenade, D. S.; Boll, M.; Foltz, M.; Wake, K. A.; Kennedy, D. J.; Munck, L. K.; Miyauchi, S.; Taylor, P. M.; Campbell, F. C.; Munck, B. G.; Daniel, H.; Ganapathy, V.; Thwaites, D. T. *Gastroenterology* **2004**, 127, 1410.
- Boll, M.; Foltz, M.; Rubio-Aliaga, I.; Daniel, H. *Genomics* **2003**, 82, 47.
- Thwaites, D. T.; McEwan, G. T. A.; Cook, M. J.; Hirst, B. H.; Simmons, N. L. *FEBS Lett.* **1993**, 333, 78.
- Thwaites, D. T.; McEwan, G. T. A.; Brown, C. D. A.; Hirst, B. H.; Simmons, N. L. *J. Biol. Chem.* **1993**, 268, 18438.
- Thwaites, D. T.; McEwan, G. T. A.; Simmons, N. L. *J. Membr. Biol.* **1995**, 145, 245.
- Thwaites, D. T.; McEwan, G. T. A.; Hirst, B. H.; Simmons, N. L. *Biochim. Biophys. Acta* **1995**, 1234, 111.
- Thwaites, D. T.; Stevens, B. C. *Exp. Physiol.* **1999**, 84, 275.
- Ranaldi, G.; Islam, K.; Sambuy, Y. *Antimicrob. Agents Chemother.* **1994**, 38, 1239.
- Thwaites, D. T.; Anderson, C. M. H. *Biochim. Biophys. Acta* **2007**, 1768, 179.

13. Thwaites, D. T.; Armstrong, G.; Hirst, B. H.; Simmons, N. L. *Brit. J. Pharmacol.* **1995**, *115*, 761.
14. Thwaites, D. T.; Basterfield, L.; McCleave, P. M. J.; Carter, S. M.; Simmons, N. L. *Brit. J. Pharmacol.* **2000**, *129*, 457.
15. Metzner, L.; Kalbitz, J.; Brandsch, M. *J. Pharmacol. Exp. Ther.* **2004**, *309*, 28.
16. Metzner, L.; Dorn, M.; Markwardt, F.; Brandsch, M. *Mol. Pharmaceutics* **2009**, *6*, 106.
17. Abbot, E. L.; Grenade, D. S.; Kennedy, D. J.; Gatfield, K. M.; Thwaites, D. T. *Brit. J. Pharmacol.* **2006**, *147*, 298.
18. Boll, M.; Foltz, M.; Anderson, C. M. H.; Oechsler, C.; Kottra, G.; Thwaites, D. T.; Daniel, H. *Mol. Membr. Biol.* **2003**, *20*, 261.
19. Metzner, L.; Neubert, K.; Brandsch, M. *Amino Acids* **2006**, *31*, 111.
20. Metzner, L.; Kottra, G.; Neubert, K.; Daniel, H.; Brandsch, M. *FASEB J.* **2005**, *19*, 1468.
21. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
22. Gebauer, S.; Knütter, I.; Hartrodt, B.; Brandsch, M.; Neubert, K.; Thondorf, I. *J. Med. Chem.* **2003**, *46*, 5725.
23. Biegel, A.; Gebauer, S.; Hartrodt, B.; Brandsch, M.; Neubert, K.; Thondorf, I. *J. Med. Chem.* **2005**, *48*, 4410.
24. Biegel, A.; Gebauer, S.; Brandsch, M.; Neubert, K.; Thondorf, I. *J. Med. Chem.* **2006**, *49*, 4286.
25. Bush, B. L.; Nachbar, R. B. *J. Comp. Aided Mol. Des.* **1993**, *7*, 587.
26. Cramer, R. D.; Bunce, G. D.; Patterson, D. E.; Frank, I. E. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18.
27. Correlations with a crossvalidated $q^2 \geq 0.5$ are considered of predictive utility: Cramer, R. D.; Depriest, S. A.; Patterson, D. E.; Hecht, P. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; p 443.
28. A deviation of less than 1 logarithmic unit indicates a sufficient correlation: Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.
29. Molecular Operating Environment (MOE) 2010.10; Chemical Computing Group: Quebec, Canada, 2010.
30. SYBYL 8.1, Tripos International, 1669 S. Hanley Road, St. Louis, MO 63144.
31. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
32. Clark, M.; Cramer, R. D.; van Opdenbosch, N. *J. Comp. Chem.* **1989**, *10*, 982.
33. Stähle, L.; Wold, S. *J. Chemom.* **1987**, *1*, 185.