Pharmaceutical Research, Vol. 23, No. 3, March 2006 (© 2006)
DOI: 10.1007/s11095-006-9462-y

Research Paper

Development of a QSAR Model for Binding of Tripeptides and Tripeptidomimetics to the Human Intestinal Di-/Tripeptide Transporter hPEPT1

Rikke Andersen,[1,5] Flemming Steen Jørgensen,[2] Lars Olsen,[2] Jon Våbenø,[1,4] Karina Thorn,[3] Carsten Uhd Nielsen,[1] and Bente Steffansen[1]

Received July 4, 2005; accepted November 11, 2005

Purpose. The aim of this study was to develop a three-dimensional quantitative structure–activity relationship (QSAR) model for binding of tripeptides and tripeptidomimetics to hPEPT1 based on a series of 25 diverse tripeptides.
Methods. VolSurf descriptors were generated and correlated with binding affinities by multivariate data analysis. The affinities for hPEPT1 of the tripeptides and tripeptidomimetics were determined experimentally by use of Caco-2 cell monolayers.
Results. The K_i-values of the 25 tripeptides and tripeptidomimetics ranged from 0.15 to 25 mM and the structural diversity of the compounds was described by VolSurf descriptors. A QSAR model that correlated the VolSurf descriptors of the tripeptides with their experimental binding affinity for hPEPT1 was established.
Conclusion. Structural information on tripeptide properties influencing the binding to hPEPT1 was extracted from the QSAR model. This information may contribute to the drug design process of tripeptides and tripeptidomimetics where hPEPT1 is targeted as an absorptive transporter for improvement of intestinal absorption. To our knowledge, this is the first time a correlation between VolSurf descriptors and binding affinities for hPEPT1 has been reported.

KEY WORDS: hPEPT1; QSAR; tripeptides; tripeptidomimetics; VolSurf.

INTRODUCTION

The human intestinal di-/tripeptide transporter hPEPT1 has been proposed as a potential target for increasing intestinal absorption of di- and tripeptidomimetic drug candidates (1,2). This strategy is based on studies showing that tripeptidomimetics (e.g., β-lactams) (3,4) and dipeptidomimetics (e.g., ACE inhibitors) (5–7) have affinity for hPEPT1 and are translocated across the enterocytic apical membrane by this transporter. hPEPT1 seems to be the sole intestinal transporter for mediating intestinal absorption of dipeptides and tripeptides (1). Furthermore, the 20 naturally occurring amino acids found in proteins transform into 8000 possible tripeptide substrates for hPEPT1, which emphasizes the broad substrate specificity of this transporter.

Rational drug design, in which hPEPT1 may be utilized as an absorptive transporter for di- and tripeptidomimetics, requires detailed knowledge about the structure and function of the transport protein as well as the structural elements of importance for substrate binding. Considering the limited number of crystal structures available for membrane proteins, systematic ligand-based investigations on structural requirements for binding and transport still represent the most important way to gather structural information, which may subsequently be used in drug design of substrates.

Improvement of pharmacokinetic properties has become an integrated part of the drug design process; therefore, the need for *in silico* models correlating molecular structures with pharmacokinetic properties has increased. Several computer-derived models on structural requirements for ligand binding to hPEPT1 have been proposed over the years (8–12). A comprehensive three-dimensional quantitative structure–activity relationship (3D-QSAR) model for dipeptides/-mimetics was published by Gebauer *et al.* (10). The pharmacophore was described by the distance between three important molecular features using the dipeptidomi-

[1] Molecular Biopharmaceutics, The Danish University of Pharmaceutical Sciences, 2-Universitetsparken, DK-2100 Copenhagen, Denmark.
[2] Biostructural Research, The Danish University of Pharmaceutical Sciences, 2-Universitetsparken, DK-2100 Copenhagen, Denmark.
[3] Medicinal Chemistry, The Danish University of Pharmaceutical Sciences, 2-Universitetsparken, DK-2100 Copenhagen, Denmark.
[4] Center for Computational Biology, Washington University School of Medicine, St. Louis, Missouri, USA.
[5] To whom correspondence should be addressed. (e–mail: ra@dfuni.dk)

ABBREVIATIONS: Ψ, *N*- methyl amide bond isostere Ψ [$CONCH_3$]; 3D-QSAR, three-dimensional quantitative structure–activity relationship; CoMSIA, comparative molecular similarity indices analysis; PCA, principal component analysis; PLS, partial least square of latent variables.

metic Alaψ[CS-N]-Pro (thioamide as peptide bond isostere) as reference compound. Gebauer *et al.* used comparative molecular similarity indices analysis (CoMSIA) performed on 79 dipeptides/-mimetics, which resulted in a predictive model from which molecular features important for binding of dipeptides/-mimetics to hPEPT1 could be extracted (10). The features were a hydrogen bond donor site located around the protonated N-terminus, a hydrogen bond acceptor site located around the carbonyl group in the first residue, and an area of high electron density around the deprotonated C-terminus. The identified features were similar to findings in other structural investigations of hPEPT1 substrates (8,13). Based on a series of β-lactam antibiotics, Våbenø *et al.* (14) identified the bioactive conformation of dipeptides/-mimetics recognized by hPEPT1, which was described by three backbone torsion angles (ψ_1~165°, ω_1~180°, and φ_2~280°) (cf. Fig. 1). These angles were in good agreement with the findings of Gebauer *et al.* (10). Furthermore, a correlation was obtained between hPEPT1 binding affinities and the conformational energy penalty paid by the dipeptides/-mimetics when adopting the bioactive conformation.

Although the studies by Gebauer *et al.* (10) and Våbenø *et al.* (14) were extensive, none of the models included tripeptides. Because the region expected to be occupied by the third residue of tripeptides is not represented in these models, it is difficult to extract explicit information about structural requirements for binding of tripeptides/-mimetics to hPEPT1. Consequently, it is not possible to achieve reliable affinity predictions for tripeptides/-mimetics from these models. Thus, there is an apparent need for new computational models that include such compounds. During the preparation of our manuscript, an extension of the dipeptide model by Gebauer *et al.* was published (9). An approach similar to the one used in their dipeptide model was applied on a data set containing dipeptides/-mimetics, tripeptides, and β-lactams, and two pharmacophore models were developed to embrace all β-lactams. Moreover, a 3D-QSAR model based on CoMSIA was established (9).

The aim of the present study was to investigate the affinity for hPEPT1 of a series of diverse tripeptides and subsequently investigate the structure–affinity relationships for tripeptide binding to hPEPT1. Our strategy was to calculate molecular descriptors representing structural properties of the tripeptides relevant for binding to hPEPT1. Subsequently, a correlation between molecular descriptors and binding affinities for hPEPT1 was obtained by multivariate data analysis.

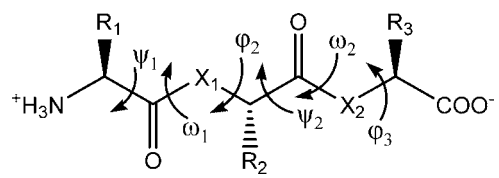

Fig. 1. The general structure of tripeptides. ψ_1, ω_1, φ_2, ψ_2, ω_2, and φ_3 define the torsions of the tripeptide backbone. R_1, R_2, and R_3 are side chains of naturally occurring amino acids in the first, second, and third residue, respectively. X_1 and X_2 are NH or $N(CH_3)$.

MATERIALS AND METHODS

Materials

H-Ala-Ala-Pro-OH (AAP) (purity > 99%), H-Gly-Phe-Phe-OH (GFF) (purity > 99%), H-Gly-Val-Phe-OH (GVF), H-Ser-Ser-Ser-OH (SSS) (purity > 99%), H-Gly-Gly-Leu-OH (GGL) (purity > 99%), H-Ala-Ala-Ala-OH (AAA) (purity > 98%), H-Ser-Gly-Gly-OH (SGG) (purity > 99%), H-Gly-Phe-Gly-OH (GFG) (purity > 98%), H-Gly-Sar-Sar-OH (GΨGΨG), H-Gly-Gly-Gly-OH (GGG) (purity > 99%), H-Gly-Gly-Sar-OH (GGΨG) (purity > 99%), H-Arg-Gly-Asp-OH (RGD) (purity > 98%) (pK_{a1} = 2.1, pK_{a2} = 3.9, pK_{a3} = 9.0, pK_{a4} = 12.5) H-Arg-Lys-Asp-OH (RKD) (purity > 98%) (pK_{a1} = 2.1, pK_{a2} = 3.9, pK_{a3} = 9.0, pK_{a4} = 10.0, pK_{a5} = 12.5), H-Lys-Trp-Lys-OH (KWK) (pK_{a1} = 2.2, pK_{a2} = 9.0, pK_{a3}/pK_{a4} = 10.0) (purity > 99%), and H-Gly-Ala-Asp-OH (GAD) (pK_{a1} = 2.1, pK_{a2} = 3.9, pK_{a3} = 9.6) were purchased from BaChem (Weil am Rhein, Germany). The purity of GVF, GAD, and GΨGΨG was not available from BaChem. The custom syntheses of H-Asp-Arg-Glu-OH (DRE) (pK_{a1} = 2.2, pK_{a2} = 3.9, pK_{a3} = 4.2, pK_{a4} = 9.8, pK_{a5} = 12.5), H-Glu-Glu-Gly-OH (EEG) (pK_{a1} = 2.3, pK_{a2}/pK_{a3} = 4.2, pK_{a4} = 9.7), and H-Asn-Phe-Trp-OH (NFW) as trifluoroacetic acid (TFA) salts were performed by SynPep (Dublin, CA, USA) with HPLC purity > 98%. H-Gly-Glu-Trp-OH (GEW) (pK_{a1} = 2.4, pK_{a2} = 4.2, pK_{a3} = 9.6) was custom synthesized as TFA salt by NeoMPS (Strasbourg, France) with HPLC purity > 99%.

Caco-2 cells were obtained from the American Type Culture Collection (Manassas, VA, USA). Cell culture media and Hanks balanced salt solution (HBSS) were obtained from Life Technologies (Høje Taastrup, Denmark). All HPLC solvents were of analytical grade, and chemicals used in buffer preparations were of laboratory grade. Other commercially available chemicals used were from Sigma (St. Louis, MO, USA). Fmoc-protected amino acids, coupling reagents, and solid supports for the synthesis were also purchased from Sigma. [^{14}C]Gly-Sar with a specific activity of 49.94 mCi/mmol was from New England Nuclear (Boston, MA, USA). H-Lys-Z[NO_2]-Pro-OH was kindly donated by Professor Klaus Neubert, Department of Biochemistry, Martin-Luther-Universität, Halle-Wittenberg, Germany.

Apparatus

The HPLC system for quantitative analyses was a Merck Hitachi LaChrom Elite consisting of an HTA pump (L-2130), a column oven (L-2300), and an autosampler (L-2300) with integrated cooling unit equipped with a Diode Array Detector (L-2450). The system was operated from the software Merck EZChrom Elite v3.1.3. All compounds were eluted on a Waters Spherisorb S5OdS2 reverse-phase column (5 μm, 250 × 4.6 mm). Radioactivity was counted on a Tri-Carb 2110TR Liquid Scintillation Analyzer from Packard (Perkin-Elmer Life and Analytical Sciences, Boston, MA, USA). The transepithelial electrical resistance (TEER) of Caco-2 cell monolayers was measured in a tissue-resistance measurement chamber (Endohm) with a voltohmmeter (EVOM) from World Precision Instruments (Sarasota, FL, USA). The monolayers used had TEERs in the range of 500–700 Ω

cm^{-2}. The shaking plate used for cell culture experiments was a Swip KS 10 Digi from Edmund Bühler (Hechingen, Germany).

The reverse-phase preparative HPLC system used for purification of crude tripeptides consisted of a Gilson 215 liquid handler and a Gilson 322 pump. The system was equipped with a Gilson 155 UV detector and controlled via the Unipoint software. The freeze-dryer was a Christ Alpha 2-4 LSC. ^{1}H and ^{13}C NMR characterization of the synthesized tripeptides was performed on an Oxford 300/75-MHz apparatus and data were recorded on Mercury Plus Varian or Gemini 2000 Varian. Further characterization and purity determinations of the tripeptides were performed using a liquid chromatography/mass spectrometry (LC/MS) system consisting of a Waters 2795 alliance separation module with a Waters 2996 photodiode array detector operated from the MassLynx Spectrometry software.

Tripeptide Synthesis and Purification

Solid-phase peptide synthesis was applied for the synthesis of H-D-Phe-Ala-Ser-OH (D-FAS), H-D-Phe-Ser-Ala-OH (D-FSA), H-D-Val-Ala-Ser-OH (D-VAS), H-D-Val-Ser-Ala-OH (D-VSA), H-D-Tyr-Ala-Ser-OH (D-YAS), and H-D-Tyr-Ser-Ala-OH (D-YSA). The synthesis was performed by standard solid-phase techniques using the 9-fluorenylmethoxycarbonyl (Fmoc) protocol with *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium BF_4^- (TBTU) salt as the coupling reagent and 2-chlorotrityl chloride resin as the solid support (15). The coupling method was based on three equivalents of Fmoc-Xaa-OH/TBTU/diisopropylethylamine (1:0.96:1.33) in *N*,*N*-dimethylformamide (DMF) for 2 h. Deprotection from Fmoc was performed with 20% piperidine in DMF and the amide bond formation was verified by the Kaiser test (16). The final step with concomitant deprotection from the *tert*-butyl side-chain protection group at serine and peptide cleavage from the resin was performed using a mixture of TFA/dichloromethane/H_2O/triisopropylsilane (14:4:1:1).

The tripeptides were purified by reverse-phase preparative HPLC applying a gradient elution of 20–90% acetonitrile with 0.1% TFA in water, and then freeze-dried. Characterization and purity determination was performed by ^{1}H and ^{13}C NMR and LC/MS. The LC/MS purity was 99% for D-FAS, 94% for D-FSA, 90% for D-VAS, 90% for D-VSA, 82% for D-YAS, and 80% for D-YSA.

Biological Investigations

Affinity Measurements

Affinity of tripeptides for hPEPT1 was measured as concentration-dependent inhibition of the apical uptake of [^{14}C]Gly-Sar (20 μM) in Caco-2 cell monolayers using a proton gradient across the monolayer with an extracellular pH of 6.0 (apical) and pH 7.4 (basolateral), as described by Nielsen *et al.* (17). In brief, Caco-2 cells were equilibrated for 15 min with 0.5 mL buffer (pH 6.0, 10 mM MES and 0.05% BSA) on the apical side and 1 mL buffer (pH 7.4, 10 mM HEPES and 0.05% BSA) on the basolateral side. After 15 min, buffers were removed and solutions of 20 μM (0.5 μCi) [^{14}C]Gly-Sar and varying concentrations of tripeptides (0–10 mM) in buffer (pH 6.0) was added to the apical side and buffer (pH 7.4) was added to the basolateral side of monolayers. After 5 min, the uptake of [^{14}C]Gly-Sar was terminated by removal of solutions and washing of cells with ice-cold HBSS buffer. The filter supports were cut out and the amount of [^{14}C]Gly-Sar was measured by liquid scintillation spectrometry. Experiments with Caco-2 cell monolayers were performed using at least three separate monolayers. The aqueous solubility of GFF, GVF, NFW, and GEW was increased by addition of 3% *N*,*N*-dimethylacetamide, which was shown to have no influence on the uptake of [^{14}C]Gly-Sar (data not shown).

Stability on Caco-2 Cells

In affinity studies, tripeptides were applied to the apical side of the Caco-2 cells; hence the tripeptides were exposed to brush border membrane-associated enzymes. The stability of the tripeptides was therefore investigated on the apical side of Caco-2 cell monolayers. To limit transcellular transport, an excess of a high-affinity hPEPT1 substrate or inhibitor was added apically during the stability study. Addition of δ-aminolevulinic acid (25 mM, K_i = 1.5 mM) or Lys-Z[NO_2]-Pro (0.1, 1.5, or 2.0 mM, K_i = 0.01 mM) (18) resulted in 80–99% inhibition of tripeptide uptake depending on the K_i-value of the tripeptide. The stability experiment was initiated by adding 0.5 ml of 0.5 mM tripeptide and 25 mM δ-aminolevulinic acid in buffer (pH 6.0) to the apical side and 1 mL buffer (pH 7.4) to the basolateral side of the Caco-2 cell monolayers. In the cases of SSS, SGG, and GAD, 2.0, 1.5 and 0.1 mM Lys-Z[NO_2]-Pro, respectively, were used to avoid interference of chromatographic peaks of δ-aminolevulinic acid and these tripeptides. Apical samples (20 μl) were taken at different time points (0–60 min), transferred to vials on ice, and analyzed on HPLC-UV. All the developed HPLC methods had a flow rate at 1 mL/min and an autosampler temperature at 4°C. The organic part of the mobile phases consisted of 10–40% acetonitrile and/or methanol and the aqueous part consisted of either 1% phosphoric acid and 10 mM 1-octanesulfonic acid, 20 mM acetate and 0.05% triethylamine or 0.5% TFA. Column temperatures were 25–50°C and the wavelength used in UV detection was 210 or 220 nm. Retention times were from 5.7 to 9.8 min.

The stability was investigated using three separate monolayers.

Analysis of Biological Data

Stability Calculations. Degradation products of tripeptides are primarily combinations of dipeptides and single amino acids. Because both di- and tripeptides are substrates for hPEPT1, the disappearance of tripeptides was monitored on the apical side of the Caco-2 cells. In addition to enzymatic degradation, the percentage of disappeared tripeptide also embraces the loss of tripeptide due to passive para- and transcellular diffusion, although the contribution is

expected to be negligible. Therefore, the enzymatic degradation of tripeptides was determined by following the disappearance of intact tripeptides. The degradation kinetics was considered to be pseudo-first-order, thus percent degraded tripeptide can be calculated as:

$$\% \text{ Disappeared} = 100 - 100\ e^{-k_{obs} \cdot t,}$$

where k_{obs} is the estimated pseudo-first-order rate constant and t is the time.

APA and LRL were excluded from the study because more than 10% of the tripeptides had disappeared from the apical side of the cell monolayers after 5 min (length of affinity assay). This 10% limit was assessed due to the influence degradation products (dipeptides) may have on the interpretation of measured apparent binding affinities of the tripeptides. The percent disappeared tripeptide is expressed as mean values.

Affinity Calculations. The IC_{50}-values were calculated as described by Nielsen *et al.* (17) and the conversion to K_i as described by Cheng and Prusoff (19). K_i-values are expressed as means ± standard error (SE).

Table I. K_i-values of Tripeptides Measured as Inhibition of [^{14}C]Gly-Sar Uptake in Caco-2 Cells and the Degradation of Tripeptides on the Apical Side of the Caco-2 Cells[a]

	Side chain charge (R_1, R_2, R_3)[b]	K_i (mM)	% Disappeared after 5 min
AAP	0, 0, 0	0.15 ± 0.03	9.0
NFW	0, 0, 0	0.19 ± 0.003	5.9
GFF	0, 0, 0	0.19 ± 0.03	9.5
GVF	0, 0, 0	0.22 ± 0.02	4.7
SSS	0, 0, 0	0.24 ± 0.02	NDD[d]
GGL	0, 0, 0	0.29 ± 0.002	2.0
AAA	0, 0, 0	0.35 ± 0.06	8.6
SGG	0, 0, 0	0.37 ± 0.05	3.4
GFG	0, 0, 0	0.39 ± 0.004	5.7
RGD	1, 0, −1	0.70 ± 0.1	1.8
GGΨG	0, 0, 0	0.75 ± 0.2	NDD[d]
GGG	0, 0, 0	0.77 ± 0.04	0.7
RKD	1, 1, −1	1.4 ± 0.1	5.8
GΨGΨG	0, 0, 0	1.6 ± 0.1	NDD[d]
GEW	0, −1, 0	2.5 ± 1.2	1.6
D-FSA	0, 0, 0	2.5 ± 0.3	–[e]
D-YSA	0, 0, 0	2.8 ± 0.4	–[e]
KWK	1, 0, 1	3.7 ± 0.9	6.6
DRE	−1, 1, −1	4.1 ± 2.5	2.1
EEG	−1, −1, 0	4.3 ± 2.1	NDD[d]
GAD	0, 0, −1	6.0 ± 2.4	1.4
D-FAS	0, 0, 0	12[c]	–[e]
D-VSA	0, 0, 0	14[c]	–[e]
D-YAS	0, 0, 0	18[c]	–[e]
D-VAS	0, 0, 0	25[c]	–[e]

[a] K_i-values are given as mean ± SE.
[b] R_1, R_2, and R_3 are the side chains of the first, second, and third residue in the tripeptide. Numbers represent the charge at pH 6.0.
[c] Estimated from the inhibitory effect at 5 mM.
[d] NDD = No degradation detectable.
[e] Stability not determined.

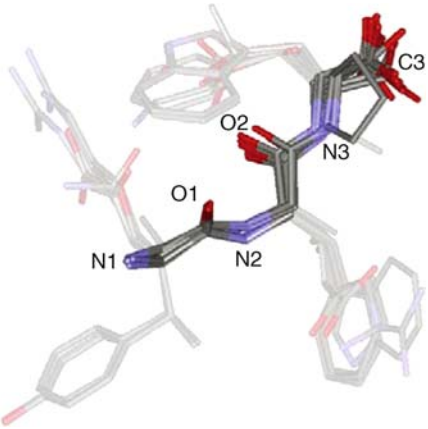

Fig. 2. Superimposition of the 25 tripeptides in a common low-energy conformation. N1 is the nitrogen atom of the N-terminus; N2 and N3 are the nitrogen atoms of the second and third residue, respectively. O1 and O2 are the carbonyl oxygen atoms of the first and second residue, respectively. C3 is the carbon of the carboxylate group of the C-terminus.

Computational Procedures

Molecular Descriptors

VolSurf descriptors were calculated using the VolSurf (v4.2.1) program implemented in SYBYL (v6.9.1). The methodology behind the VolSurf descriptors has been extensively described by Cruciani *et al.* (20). According to standard procedures, the VolSurf descriptors used in this study were based on GRID 3D molecular interaction fields (MIFs) resulting from interaction energies between the ligand and different probes. Initially, VolSurf descriptors were generated for a data set of 8008 tripeptides, using the probes H_2O, DRY, and O. The 8008 tripeptides were all possible tripeptides together with GΨGΨG, GGΨG, D-FAS, D-FSA, D-VAS, D-VSA, D-YAS, and D-YSA. The 25 tripeptides (cf. Table I) were in a common low-energy conformation (Fig. 2), which is described in the supplemental data (Supplementary material available online for authorized users at http://dx.doi.org/10.1007/s11095-006-9462-y). For these 25 tripeptides, additional VolSurf descriptors were calculated using the four probes H_2O, DRY, O, and N1.

The generation of VolSurf descriptors was performed with default settings. VolSurf descriptor definitions can be found in the software manual (32).

Multivariate Data Analysis

The VolSurf descriptors were analyzed by principal component analysis (PCA) and partial least square of latent variables (PLS) in SIMCA-P (v10.0). Default settings were applied. The variables were centered and scaled to unit variance. Cross validation (seven rounds) was performed to test the significance of the model. A further validation of the PLS model was performed by evaluating the statistical significance of the estimated predictive power of the model (21). Thus, PLS models were developed from data sets where descriptors from the original PLS model were left intact, whereas K_i-values, left numerically unchanged, were ran-

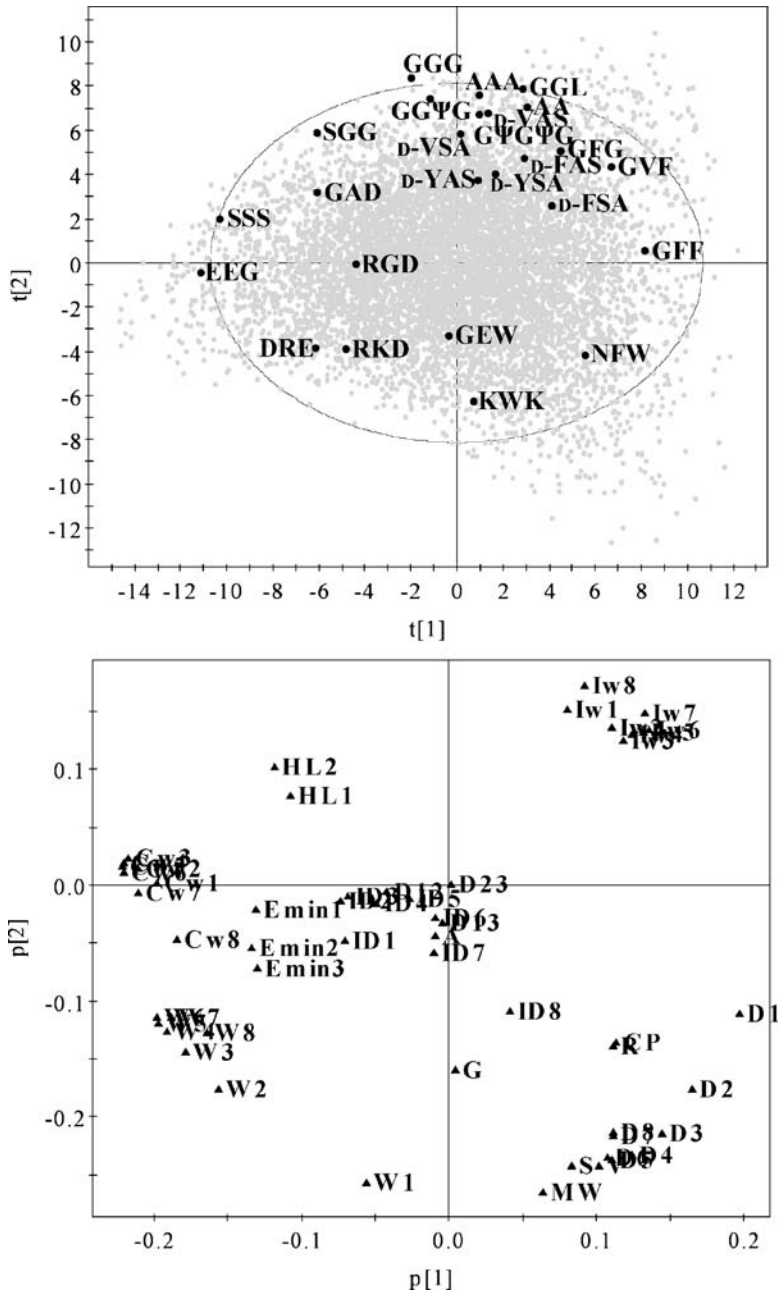

Fig. 3. *Top*: PCA score plot where $t[1]$ and $t[2]$ are the first and second principal components, respectively. The obtained regression coefficient R^2 was 0.78 and the predictive correlation coefficient Q^2 was 0.73. The circle is the 95% confidence interval. The first and second principal components explain 34 and 30% of the variation, respectively. The dots represent all 8000 possible tripeptides constructed from the 20 naturally occurring amino acids and the eight tripeptides/-mimetics included in the study (GGΨG, GΨGΨG, D-XAS, and D-XSA, where X is Y, V, or F). *Bottom*: PCA loading plot of the loadings of the first ($p[1]$) and second ($p[2]$) principal component. Variables important in the description of a tripeptide are located in the same area of the loading plot as the tripeptide in the score plot.

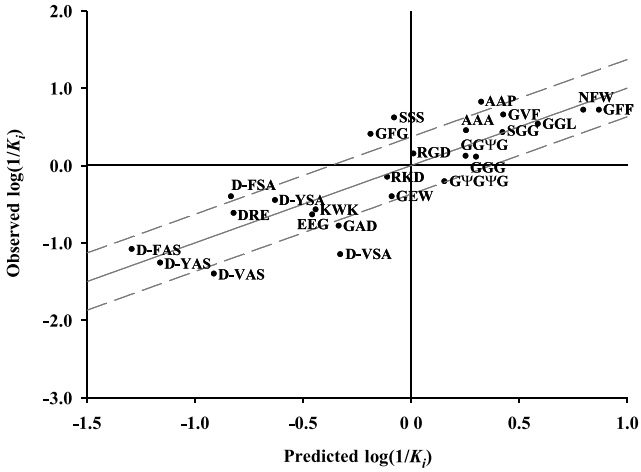

Fig. 4. Relationship between predicted and observed log(1/K_i). Root mean square error (RMSEE) of the fit was 0.37 and is indicated by the dotted lines.

domized. R^2- and Q^2-values of the randomized models were then compared with the values of the original model.

RESULTS

Affinity Measurements

The affinities for hPEPT1 of the 25 tripeptides are given in Table I. The tripeptides for which saturation curves were obtained had affinities for hPEPT1 ranging from 0.15 to approximately 6 mM. Due to the low affinity of D-FAS, D-VSA, D-YAS, and D-VAS, the affinities of these four tripeptides were estimated based on inhibitory effects of [^{14}C]Gly-Sar uptake at a single concentration of the tripeptide and ranged from approximately 12–25 mM. The affinity for hPEPT1 seems to be related to both stereochemistry and charge. The L,L,L-configured tripeptides with neutral side chains had higher affinities than L,L,L-configured tripeptides with charged side chains. The tripeptides with the lowest affinities had D,L,L configuration with neutral side chains. Furthermore, D-Phe or D-Tyr as the first amino acid residue seemed unfavorable for the binding to hPEPT1 when the second residue was a polar serine and the third was a nonpolar alanine, compared to the opposite arrangement. The introduction of an *N*-methyl group in the second peptide bond (GGΨG) of GGG did not affect the affinity notably, whereas *N*-methylation of both peptide bonds (GΨGΨG) slightly decreased the affinity compared to GGG.

Structural Diversity of the Investigated Tripeptides

VolSurf descriptors were generated for all possible natural tripeptides (8000 overall) together with the eight tripeptides/-mimetics included in the study (i.e., GGΨG, GΨGΨG, D-XAS, and D-XSA, where X is Y, V, or F). A PCA model was developed from the set of 8008 tripeptides, which resulted in five principal components. The score and loading plots composed of the first and second principal components are shown in Fig. 3.

The 25 tripeptides were well distributed all over the chemical space spanned by tripeptides (Fig. 3, top). The small tripeptides with neutral side chains were found in the top of the score plot, which was consistent with the descriptors related to molecular size [e.g., molar weight (MW)] and hydrophilic volumes (e.g., W1–8) being found at the bottom of the loadings plot (Fig. 3, bottom). Most

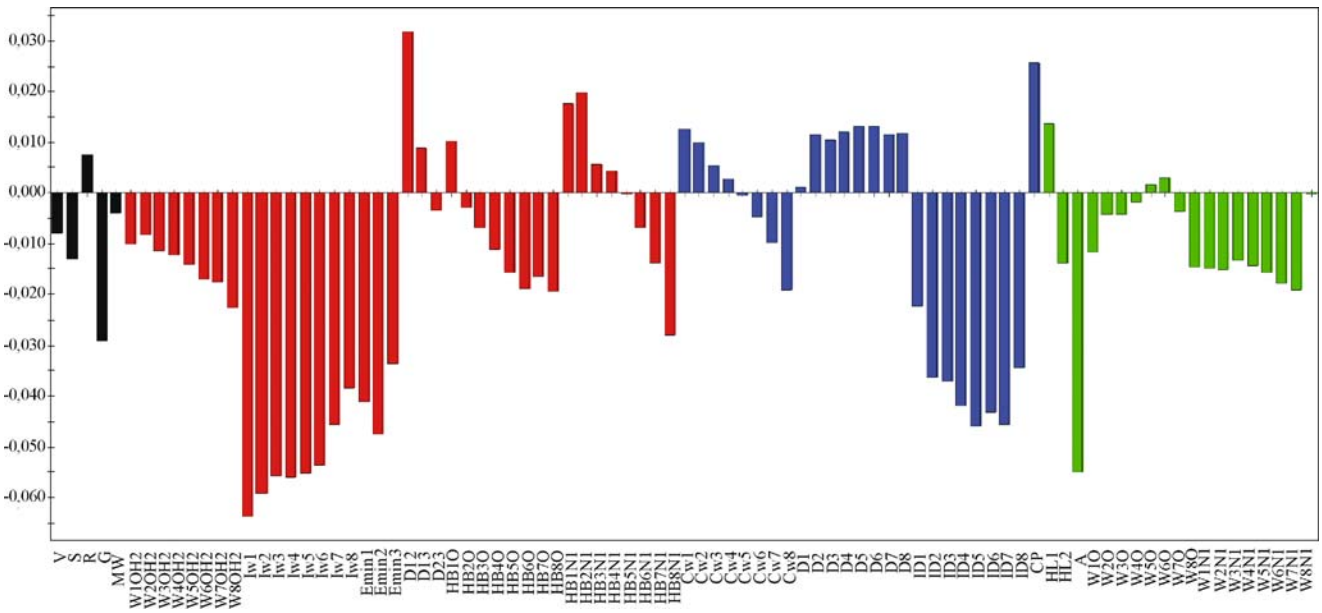

Fig. 5. PLS coefficients plot. Columns with positive values represent variables having a favorable effect on the affinity for hPEPT1, thus an increase of the log(1/K_i) value of the tripeptide. Likewise, columns with negative values represent variables having an unfavorable effect on the affinity for hPEPT1. Black columns are descriptors describing size and shape of the tripeptide; red columns are descriptors enclosing hydrophilic properties; blue columns are descriptors representing hydrophobic properties; and green columns are the balances between different properties in the tripeptide. The *x* axis represents VolSurf descriptors [34]. Values of the *y* axis are the PLS coefficients.

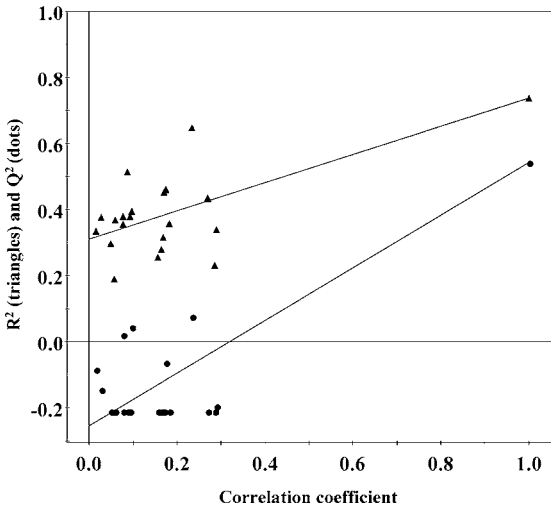

Fig. 6. Validation plot. R^2 and Q^2 were calculated for 20 randomized models. The *y*-axis represents the values of R^2 (triangles) and Q^2 (dots) for every model. The *x*-axis is the correlation coefficient between the original and the randomized K_i-values. Thus, the data point with a correlation coefficient of 1 is the original data. The intercepts of the R^2- and Q^2-regression are 0.31 and −0.26, respectively.

of the charged tripeptides were located at the bottom to the left, probably as a consequence of the hydrophilic volumes being located in that region. Moreover, there was a tendency that the positively charged tripeptides were located at the bottom of the score plot, whereas the negatively charged tripeptides were found to the left in the score plot. A reason for that might be that the negatively charged amino acids (D and E) generally are smaller than the positively charged amino acids (R and K). This is also reflected in the capacity factors (Cw1–8), expressing the ratio between the volume of hydrophilic regions and the molecular surface, which explains why negatively charged amino acids in general have larger capacity factors as seen for EEG. This was further illustrated by the SSS tripeptide, which is very polar and relatively small, being found close to EEG, whereas the nonpolar and relatively large GFF tripeptide was found to the right of the score plot because it had small polar regions per surface area and therefore small capacity factors.

Quantitative Structure–Affinity Relationship

A PLS analysis of the VolSurf descriptors as X-data and $\log(1/K_i)$ as Y-data resulted in a model with two principal components and R^2- and Q^2-values of 0.74 and 0.54, respectively. The first latent variable explained 48% of the variation in the model.

The predicted *vs.* the experimentally determined binding affinities are shown in Fig. 4. It was seen that most of the predicted binding affinities were within the confidence level of the model, with SSS, GFG, D-VAS, and D-VSA deviating most from the model. The influence of each descriptor on the tripeptide–hPEPT1 binding interaction is depicted in Fig. 5.

Further validation of the PLS model was performed by evaluation of the statistical significance of the estimated predictive power of the model (21). R^2- and Q^2-values of the PLS models of randomized data sets are plotted against the correlation coefficient between the K_i-values of the original and randomized data sets in Fig. 6. The intercept of the R^2- and Q^2-regression lines was 0.31 and −0.26, respectively.

DISCUSSION

In the present study, the structure–affinity relationship for tripeptide binding to hPEPT1 was investigated. The binding affinities were determined in Caco-2 cells and VolSurf descriptors were generated for the description of the molecular properties of the tripeptides.

Structural Diversity of Tripeptides

To evaluate how well VolSurf descriptors describe the structural diversity of tripeptides in general, VolSurf descriptors were generated for all possible natural tripeptides. The tripeptides were evenly distributed in the chemical space spanned by the first and second principal component of the performed PCA (Fig. 3). Thus, VolSurf descriptors describe the tripeptides very well with an acceptable number of outliers.

The success of a QSAR model in predicting affinities of different tripeptides depends to a large extent on how well the structural variation of the tripeptides is represented by the tripeptides selected for model development. From the PCA plot, it can be seen that the 25 tripeptides were adequately scattered across the entire chemical space, thus representing a major part of the structural variation of tripeptides. This large degree of structural variation among the 25 tripeptides formed an excellent basis for the development of a comprehensive QSAR model.

Seven tripeptides with side chains that are charged at physiological pH were included in this set of tripeptides. Two tripeptides (GEW and GAD) with a negatively charged side chain in R_2 and R_3, respectively, were studied. In addition, tripeptides with multiple side-chain charges were investigated: RGD and KWK with two charged side chains, a positively charged R_1, and a negatively charged and positively charged R_3, respectively. Furthermore, EEG had two charged side chains: a negatively charged R_1 and R_2. Three charged side chains were present in RKD and DRE with a positively and negatively charged R_1, a positively charged R_2, and a negatively charged R_3, respectively. The charged species described in Table I were the predominant ones at pH 6.0 (>98%) when the estimated pK_a-values of the corresponding amino acids (cf. Materials and Methods) were used to calculate the overall charge. It is also worth noting that RGD, RKD, KWK, and DRE can never attain a classical peptide zwitterionic form in which only the N- and C-termini are charged, regardless of pH.

Our results show that tripeptides with multiple charged side chains bind to hPEPT1 with affinities in the range of 0.70–4.6 mM. In the CoMSIA model by Biegel *et al.* (9), one tripeptide with two positively charged side chains (GHK) was included in the study. This tripeptide had an affinity of 4.1 mM, which is in the same range as our tripeptides with multiple charged side chains (9). A few studies have been

conducted on how substrates with one net charge may bind to and be transported by hPEPT1, based on electrophysiological studies in *Xenopus laevis* oocytes injected with PepT1 mRNA (22,23). It was suggested that dipeptides with one negative net charge are transported along with two protons opposed to the 1:1 ratio of substrates with a neutral net charge (24). After translocation of a negatively charged dipeptide, one proton originating from the proton-binding site and one proton from the protonated anionic side chain is released intracellularly (22,23). Anionic dipeptides are presumably protonated by a histidine residue in the binding pocket (4) and/or by external solution (23). The electrogenic transport of cationic dipeptides is believed to occur with a charged side chain and one proton at the proton-binding site (23). The transport of a cation may occur when a histidine residue of the hPEPT1 binding pocket is deprotonated, thus making an attraction between π electrons of the imidazole ring and the positive charge in the side chain possible (4). As reflected above, the stoichiometric studies and the proposed models on the transport mechanism of charged substrates have mainly focused on dipeptides. As our results show, hPEPT1 can bind tripeptides with more than one net charge at luminal pH (~6.0). This could point toward a transport protein with a binding site that may comprise more than two areas that can accommodate strong electrostatic interactions by, e.g., protonation/deprotonation interplay between histidine residues and a substrate (as described above) or by shielding effects of water-filled pockets within the binding site, as seen in the bacterial periplasmic oligopeptide binding protein OppA (25,26). Moreover, detailed studies are needed to get more solid information on how hPEPT1 handles multiple-charged substrates in the binding and transport process.

Relating Structural Properties to Affinity

A successful QSAR model was achieved by correlation of $\log(1/K_i)$ with VolSurf descriptors. This PLS model had a good correlation and predictability ($R^2 = 0.74$ and $Q^2 = 0.58$). Thus, this model could be used for estimating affinities of the tripeptides with an acceptable result (Fig. 4). A large number of substrates for hPEPT1 have been identified; however, only a small part of these are tripeptides. The reason for this may be the limited commercial availability of such compounds and that the investigations of hPEPT1 affinities only provide usable data when high stability of the tripeptides has been shown in the experimental setting. In the development of the model, we therefore included all 25 tripeptides, which left no room for a subsequent testing of the model with another series of tripeptides. However, in addition to cross-validation, an evaluation of the statistical significance of the estimated predictive power of the model was performed by randomizing the K_i-values of the original data set. The models developed from these randomized data sets all had poorer correlations and lower predictabilities than the originally developed PLS model (cf. Fig. 6). Furthermore, it was suggested by Eriksson *et al.* that for a valid model the intercepts for the R^2-regression and Q^2-regression should not exceed 0.3–0.4 and 0.05, respectively (27). In our study, the intercepts through the R^2- and Q^2-data points of the randomized and original data set did not exceed the proposed limits, which indicates that the developed PLS model is valid.

Due to a broad substrate specificity it would be expected that general surface properties such as shape, electrostatics, hydrogen bonding, and hydrophobicity of the compounds are of major relevance for the description of the interactions between a tripeptide and hPEPT1. VolSurf descriptors are designed to describe such surface properties (20,28). Thus, the use of VolSurf descriptors in describing hPEPT1 substrates seems to be reasonable considering the broad substrate specificity of the transporter. Furthermore, the usefulness of VolSurf descriptors, in combination with Almond descriptors, in describing the ligand–protein interaction in the metabolic degradation of compounds by cytochrome P450 CYP3A4 has also been reported (29).

Although VolSurf descriptors are relatively independent of the initial conformation (20,30) the descriptors were generated for tripeptides in a common low-energy conformation, which is described in the supplemental data. Thus, a possible effect of conformation on VolSurf descriptors was minimized. As shown in Fig. 2, all 25 tripeptides included in our study could be aligned to this common low-energy conformation.

The fact that VolSurf descriptors seem suitable for describing the binding of tripeptides to hPEPT1 may hold a great advantage in the implementation of the model in the drug design process, as VolSurf descriptors are independent of alignment and relatively independent of conformation compared to other molecular descriptors such as Almond- or CoMFA/CoMSIA-based descriptors (20). This means that the affinity predictions performed with the VolSurf-based model would be less sensitive to the initial conformation of the tripeptides/-mimetics and not depend on alignment, hence facilitating the use of the model as an *in silico* prediction tool.

Although affinities give valuable information about the binding interaction between tripeptides/-mimetics and hPEPT1, one should be aware that affinity solely contains information about the ability of a compound to inhibit the uptake of the radiolabeled standard substrate Gly-Sar and not information about whether the compound is translocated itself. However, it should be noted that tripeptides are very likely transported intact across the apical membrane and then partly or completely degraded by peptidases after entering the cytosol of the enterocyte (31).

Molecular Properties Important for Affinity

The VolSurf model holds a large amount of detailed information about structural elements having an impact on the tripeptide–transport protein interaction. The interplay between different descriptors resolves the effect of molecular properties on the affinity to hPEPT1 (cf. Fig. 5). However, to retain clarity we only focus on a few of the descriptors with specific physicochemical properties important for binding of the tripeptides to hPEPT1. The descriptor ID1 represents one of the structural elements shown to have substantial influence. A large value of ID1 should be avoided as it correlates negatively with $\log(1/K_i)$ (cf. Fig. 5). This means that it would be more favorable to have the center of mass of the hydrophobic regions close to the center of mass of the tripeptide (cf. Fig. 7, NFW) than having the hydrophobic region concentrated around one area of the tripeptide (cf.

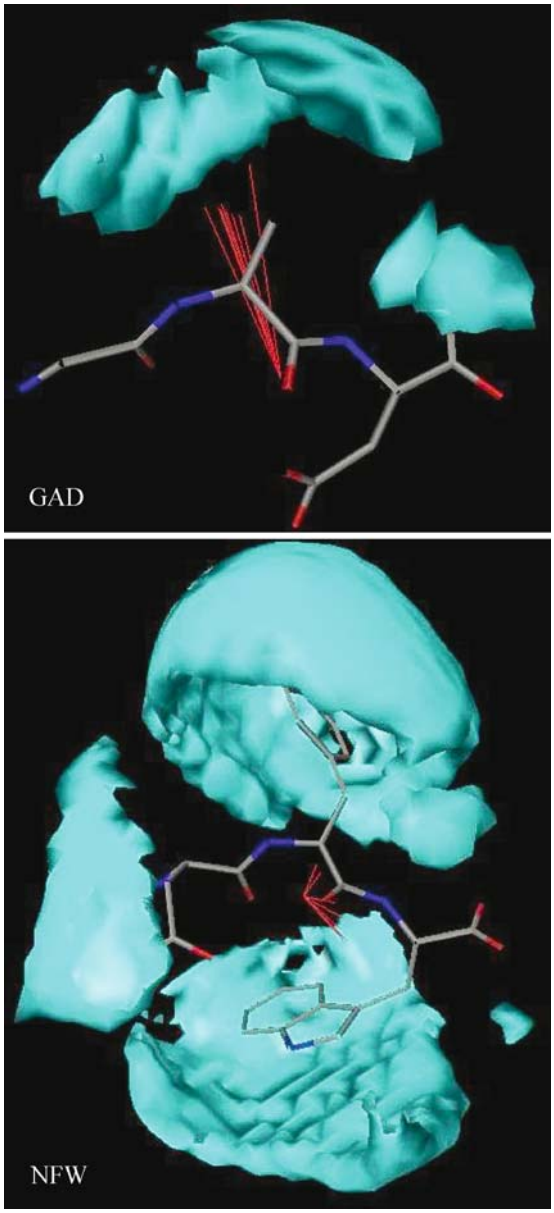

Fig. 7. Visualization of the hydrophobic integy moments (ID1) of the tripeptides GAD and NFW. The integy moments are represented by red lines; vectors going from the center of mass of the tripeptide and the center of the hydrophobic region. The cyan areas are the molecular interaction fields derived with the DRY probe. GAD is a low-affinity compound and has high integy moments, whereas NFW is a high-affinity compound and has low integy moments.

Fig. 7, GAD). To test this hypothesis, the third residue in NFW was changed to glycine and alanine. These changes caused the removal of a strong hydrophobic region and replaced it with a much weaker one. Thus, the hydrophobic region was now mainly centered on one area of the molecule leading to increase in values of ID1 (data not shown), which will decrease the affinity for both NFG and NFA.

Another descriptor, HB1-8N1 (hydrogen bonding capability; the difference between hydrophilic volumes obtained with H_2O probe and N1 probe, respectively), had both a positive and negative influence on the binding affinity depending on the energy level at which it was generated. At lower energy levels (HB1-2N1), the hydrophilic volume of the water probe accounts for weak intermolecular forces (polarizability and dispersion forces), whereas the higher energy levels (HB8N1) account for stronger intermolecular forces (hydrogen bond donor–acceptor regions). This suggests that larger hydrophobic surface areas (non-hydrogen-bonding areas) are better accepted by the transporter than more hydrophilic surface areas (hydrogen-bonding areas). A descriptor with a clear positive effect on affinity is D12. D12 represents the distance between the two lowest local minima of interaction energy obtained by the interaction of the H_2O probe with the tripeptides. Thus, a long distance between the two minima favors high affinity. The lowest energy minima would be obtained around charged functional groups. In a tripeptide with neutral side chains, the C- and N-terminus would become the lowest minima. Because the distance between the termini accounts for one of the longest distances in the tripeptides, neutral side chains are favorable for the binding interaction with hPEPT1. This is in good agreement with the obtained affinities of the tripeptides (cf. Table I). Furthermore, it is also in accordance with the effect of electrostatic properties on affinities for hPEPT1 extracted from the recently reported CoMSIA model for hPEPT1 substrates (9). The region where a negative charge had a positive effect on the affinity for hPEPT1 was located in an area accommodating the C-terminus of a di- or tripeptide. The influence of a positive charge was seen in one area corresponding to the N-terminus of a di- or tripeptide. Because other electrostatic areas were not identified as being influential on the affinity, a positive effect of neutral side chains in the binding interaction of a tripeptide with hPEPT1 must be anticipated.

In summary, our results clearly show that VolSurf descriptors are useful in describing the molecular features important for interaction of tripeptides with hPEPT1. This was expressed by the establishment of a QSAR model correlating structural properties of tripeptides to binding affinities for hPEPT1. This diverse and predictive model may contribute in the future design of tripeptides/-mimetics, e.g., as an *in silico* tool for screening compound libraries to guide synthetic efforts and subsequent *in vitro* studies.

CONCLUSION

In the present study a correlation between molecular properties of tripeptides and their binding interaction with hPEPT1 was achieved based on VolSurf descriptors and binding affinities. The QSAR model was based on a very diverse set of tripeptides embracing multiple-charged as well as neutral tripeptides. These results show that hPEPT1 can bind tripeptides with multiple-charged side chains.

VolSurf descriptors have only to a limited degree been applied successfully in modeling of transporter/enzyme-mediated processes, and to our knowledge this is the first time a correlation between VolSurf descriptors and binding affinities for hPEPT1 has been reported. The structural information on tripeptide properties, which influences the binding to hPEPT1, extracted from our QSAR model may contribute in the drug design process of tripeptides/-mimetics where hPEPT1 is targeted as an absorptive transporter for improvement of intestinal absorption.

ACKNOWLEDGMENTS

The Danish Medicinal Council supported this work (project grant 22-03-0274). Zealand Pharma A/S financially supported this work via the Drug Research Academy, Danish University of Pharmaceutical Sciences, Copenhagen, Denmark. Partial funding was obtained from the BioSim workpackage 15 funded by the European Commission (BioSim NoE LSH-CT-005137). The HPLC system was cofunded by the Hørslev Foundation. We appreciate the support by Tripos Associates and Gabriele Cruciani regarding the access to Sybyl and VolSurf. The technical assistance of Bettina Dinitzen and Susanne Nørskov Sørensen is highly appreciated.

REFERENCES

1. B. Brodin, C. U. Nielsen, B. Steffansen, and S. Frokjaer. Transport of peptidomimetic drugs by the intestinal di/tripeptide transporter, PepT1. *Pharmacol. Toxicol.* **90**:285–296 (2002).
2. C. U. Nielsen, J. Våbenø, R. Andersen, B. Brodin, and B. Steffansen. Recent advances in therapeutic applications of human peptide transporters. *Expert Opin. Ther. Pat.* **15**:153–166 (2005).
3. B. Bretschneider, M. Brandsch, and R. Neubert. Intestinal transport of β-lactam antibiotics: analysis of the affinity at the H+/peptide symporter (PEPT1), the uptake into Caco-2 cell monolayers and the transepithelial flux. *Pharm. Res.* **16**:55–61 (1999).
4. X. Z. Chen, A. Steel, and M. A. Hediger. Functional roles of histidine and tyrosine residues in the H(+)-peptide transporter PepT1. *Biochem. Biophys. Res. Commun.* **272**:726–730 (2000).
5. D. I. Friedman and G. L. Amidon. Passive and carrier-mediated intestinal absorption components of two angiotensin converting enzyme (ACE) inhibitor prodrugs in rats: enalapril and fosinopril. *Pharm. Res.* **6**:1043–1047 (1989).
6. C. Shu, H. Shen, U. Hopfer, and D. E. Smith. Mechanism of intestinal absorption and renal reabsorption of an orally active ACE inhibitor: uptake and transport of fosinopril in cell cultures. *Drug Metab. Dispos.* **29**:1307–1315 (2001).
7. D. T. Thwaites, M. Cavet, B. H. Hirst, and N. L. Simmons. Angiotensin-converting enzyme (ACE) inhibitor transport in human intestinal epithelial (Caco-2) cells. *Br. J. Pharmacol.* **114**:981–986 (1995).
8. P. D. Bailey, C. A. Boyd, J. R. Bronk, I. D. Collier, D. Meredith, K. M. Morgan, and C. S. Temple. How to make drugs orally active: a substrate template for peptide transporter PepT1. *Angew. Chem., Int. Ed. Engl.* **39**:505–508 (2000).
9. A. Biegel, S. Gebauer, B. Hartrodt, M. Brandsch, K. Neubert, and I. Thondorf. Three-dimensional quantitative structure–activity relationship analyses of β-lactam antibiotics and tripeptides as substrates of the mammalian H(+)/peptide cotransporter PEPT1. *J. Med. Chem.* **48**:4410–4419 (2005).
10. S. Gebauer, I. Knutter, B. Hartrodt, M. Brandsch, K. Neubert, and I. Thondorf. Three-dimensional quantitative structure–activity relationship analyses of peptide substrates of the mammalian H+/peptide cotransporter PEPT1. *J. Med. Chem.* **46**:5725–5734 (2003).
11. J. Li and I. J. Hidalgo. Molecular modeling study of structural requirements for the oligopeptide transporter. *J. Drug Target.* **4**:9–17 (1996).
12. P. W. Swaan, B. C. Koops, E. E. Moret, and J. J. Tukker. Mapping the binding site of the small intestinal peptide carrier (PepT1) using comparative molecular field analysis. *Recept. Channels* **6**:189–200 (1998).
13. M. Brandsch, I. Knutter, and F. H. Leibach. The intestinal H+/peptide symporter PEPT1: structure–affinity relationships. *Eur. J. Pharm. Sci.* **21**:53–60 (2004).
14. J. Våbenø, C. U. Nielsen, B. Steffansen, T. Lejon, I. Sylte, F. S. Jørgensen, and K. Luthman. Conformational restrictions in ligand binding to the human intestinal di-/tripeptide transporter: implications for design of hPEPT1 targeted prodrugs. *Bioorg. Med. Chem.* **13**:1977–1988 (2005).
15. R. Knorr, A. Trzeciak, W. Bannwarth, and D. Gillessen. New coupling reagents in peptide chemistry. *Tetrahedron Lett.* **30**:1927–1930 (1989).
16. E. Kaiser, R. L. Colescott, C. D. Bossinger, and P. I. Cook. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **34**:595–598 (1970).
17. C. U. Nielsen, R. Andersen, B. Brodin, S. Frokjaer, M. E. Taub, and B. Steffansen. Dipeptide model prodrugs for the intestinal oligopeptide transporter. Affinity for and transport via hPepT1 in the human intestinal Caco-2 cell line. *J. Control. Release* **76**:129–138 (2001).
18. I. Knutter, S. Theis, B. Hartrodt, I. Born, M. Brandsch, H. Daniel, and K. Neubert. A novel inhibitor of the mammalian peptide transporter PEPT1. *Biochemistry* **40**:4454–4458 (2001).
19. Y. Cheng and W. H. Prusoff. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**:3099–3108 (1973).
20. G. Cruciani, P. Crivori, P. A. Carrupt, and B. Testa. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *J. Mol. Struct., Theochem* **503**:17–30 (2000).
21. H. van der Voet. Comparing the predictive accuracy of models using a simple randomization test. *Chemometr. Intell. Lab.* **25**:313–323 (1994).
22. G. Kottra, A. Stamfort, and H. Daniel. PEPT1 as a paradigm for membrane carriers that mediate electrogenic bidirectional transport of anionic, cationic, and neutral substrates. *J. Biol. Chem.* **277**:32683–32691 (2002).
23. A. Steel, S. Nussberger, M. F. Romero, W. F. Boron, C.-A. R. Boyd, and M. A. Hediger. Stoichiometry and pH dependence of the rabbit proton-dependent oligopeptide transporter PepT1. *J. Physiol.* **498**:563–569 (1997).
24. Y. J. Fei, Y. Kanai, S. Nussberger, V. Ganapathy, F. H. Leibach, M. F. Romero, S. K. Singh, W. F. Boron, and M. A. Hediger. Expression cloning of a mammalian proton-coupled oligopeptide transporter. *Nature* **368**:563–566 (1994).
25. J. E. Ladbury. Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design. *Chem. Biol.* **3**:973–980 (1996).
26. J. R. Tame, S. H. Sleigh, A. J. Wilkinson, and J. E. Ladbury. The role of water in sequence-independent ligand binding by an oligopeptide transporter protein. *Nat. Struct. Biol.* **3**:998–1001 (1996).
27. L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold. *Multi- and Megavariate Data Analysis; Principles and Applications*, Umetrics AB, Umeå, Sweden, 2001.
28. G. Cruciani, M. Pastor, and W. Guba. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **11**:S29–S39 (2000).
29. P. Crivori, I. Zamora, B. Speed, C. Orrenius, and I. Poggesi. Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates. *J. Comput. Aided. Mol. Des.* **18**:155–166 (2004).
30. P. Crivori, G. Cruciani, P. A. Carrupt, and B. Testa. Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **43**:2204–2216 (2000).
31. D. M. Matthews. Intestinal absorption of peptides. *Physiol. Rev.* **55**:537–608 (1975).
32. Molecular Discovery Ltd. VolSurf Manual (VolSurf v4.1.3). http://www.moldiscovery.com/soft_volsurf.php. Accessed 10 Nov 2005.