DOI: 10.1002/cmdc.200800184

Modeling of the Intestinal Peptide Transporter hPepT1 and Analysis of Its Transport Capacities by Docking and Pharmacophore Mapping

Alessandro Pedretti,[a] Laura De Luca,[b] Cristina Marconi,[a] Gianpaolo Negrisoli,[c] Giancarlo Aldini,[a] and Giulio Vistoli*[a]

An early pharmacokinetic screen for peptidomimetic drugs should have the ability to predict molecules with high affinity for intestinal transporters, as peptide-like derivatives are seldom absorbed passively. Hence, the first objective of this study was to generate a reliable model for the structure of the hPepT1 protein, which is the main intestinal transporter involved in the absorption of both dietary peptides and peptidomimetics. The modeling was based on the resolved structure of the homologous bacterial lactose permease LacY using a fragmental strategy. The interaction capacities of the hPepT1 model were explored by docking a set of 50 known ligands. Despite the known predilection of hPepT1 for hydrophobic ligands, docking results unveiled the key role of the polar interactions stabilized by charged termini, especially concerning the ammonium head group. The docking results were further verified by developing a pharmacophore model that confirmed the key features required for optimal hPepT1 affinity. The consistency of the docking results and the agreement with the pharmacophore model afford an encouraging validation for the proposed model and suggest that it can be exploited to design peptide-like molecules with an improved affinity for such a transporter.

Introduction

Peptidomimetic drugs have attracted increasing interest and intense research efforts over the past 20 years, especially because it was shown that the severe limitations, which hampered the oral delivery of peptide-like molecules, could be successfully addressed.[1] Among the barriers affecting peptide bioavailability after oral administration is intestinal absorption, which plays a crucial role for hydrophilic peptides in particular, as they are rarely absorbed by passive permeation. This problem might be overcome by using prodrug strategies or by increasing the affinity for intestinal carriers.[2]

The more recent strategies in medicinal chemistry involve pharmacokinetic characterization of new molecules as soon as possible in the development pipeline, with the clear aim to select (and develop) only drug-like compounds with optimal pharmacokinetic behavior.[3] Such early pharmacokinetic analysis requires the acquisition of as many useful and relevant molecular properties as possible to allow reliable predictions of drug-likeness, also for large molecular libraries.[4] Because natural peptides are actively absorbed by intestinal carriers, a valuable pharmacokinetic screen for peptidomimetic compounds should predict which molecules have a significant likelihood of mimicking the natural peptides that are actively absorbed via intestinal carriers. For these hydrophilic peptide-like molecules, this information is even more important than the classical physicochemical descriptors (such as log *P*, PSA, and aqueous solubility) commonly used in pharmacokinetic screening.

Among the intestinal carriers,[5] the apical proton-dependent peptide transporters play a key role in the absorption of digested dietary proteins as well as peptidomimetic drugs such as β-lactam antibiotics and ACE inhibitors.[6] The opportunity to exploit such transporters in drug delivery has spurred several functional characterization studies, and in 1994 the first intestinal transporter, PepT1, was cloned from rabbit.[7] Afterward, a wealth of data on various transported substrates was accumulated, allowing rationalization of the structural requirements for optimal transport.

The cloned transporter belongs to the POT (proton-dependent oligopeptide transporter) superfamily, which includes proteins from a wide species range, from yeasts to humans.[8] In humans, two members of the POT superfamily possess transport activity, namely hPepT1 (*SLC15a1*) and hPepT2 (*SLC15a2*). Specifically, hPepT1 is expressed in the small intestine, in the proximal tubules of the kidney, as well as in pancreatic, liver, and renal cells. This transporter is a H^+/peptide co-transporter, which exploits the proton gradient between the intestinal lumen (pH ~5.5–6) and cells (pH ~7). Such a gradient is maintained by the activity of a Na^+/H^+ exchanger, which, in turn, is fueled by the basolateral Na^+/K^+ ATPase. Moreover, anionic,

[a] *Dr. A. Pedretti, Dr. C. Marconi, Prof. G. Aldini, Dr. G. Vistoli*
Istituto di Chimica Farmaceutica e Tossicologica "Pietro Pratesi"
Facoltà di Farmacia, Università degli Studi di Milano
Via Mangiagalli 25, 20133 Milano (Italy)
Fax: (+39) 02 503119359
E-mail: giulio.vistoli@unimi.it

[b] *Dr. L. De Luca*
Dipartimento Farmaco-Chimico, Facoltà di Farmacia
Università di Messina, Viale Annunziata, 98168 Messina (Italy)

[c] *Dr. G. Negrisoli*
Flamma S.p.A., Via Bedeschi 22, Chignolo d'Isola (BG) (Italy)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.200800184.

cationic, and neutral molecules are transported with very similar efficacy, suggesting that the number of co-transported protons can vary depending on the substrate's charge.[9]

The large number of affinity values for hPepT1 afforded a comprehensive understanding of the structure–activity relationships (SAR) for such a transporter.[10] The structural requirements that a given molecule must fulfill for optimal transport can be summarized as follows: 1) although no systematic studies have analyzed the effects of substrate size on hPepT1 activity, it is reasonable that the optimal length corresponds to that of di- and tripeptides, while tetrapeptides and single amino acids are not well recognized; 2) transport is markedly stereospecific: natural amino acids are properly transported, whereas the introduction of D isomers vastly decreases affinity; 3) the terminal groups are not mandatory for transport, even if modifications of the carboxylate group are better tolerated than those of the amino group; 4) central peptide bonds are not required for optimal bioactivity, and several substrates without peptide bonds show significant affinity for hPepT1; and 5) hydrophobic side chains are preferred. Such qualitative observations have been translated into QSAR models mainly through the use of 3D approaches.[11,12] Similarly, pharmacophore models have allowed the differentiation between substrates and inhibitors.[13] However, full-length homology models for hPepT1 and docking simulations with the aim to analyze substrate recognition at the atomic level have not yet been reported.[14]

The hPepT1 transporter is a 708-residue protein with a transmembrane bundle composed of 12 membrane-spanning helices. It has a large extracellular loop between transmembrane (TM) regions 9 and 10, and both termini face the cytosol.[15] Because hPepT1 is a transmembrane protein, its structure cannot be easily resolved. However, the hPepT1 structure can be modeled considering its homology with the recently crystallized bacterial lactose permease LacY.[16] We therefore set out to generate a 3D model of full-length hPepT1 by homology techniques. Although it was possible to model such a protein using LacY as the unique template, we exploited the fragmental approach, as recently proposed by some of us.[17–20] Such a method allows the generation of consistent models for any transmembrane protein, accounting for local homologies and avoiding the construction of models that lose their structural peculiarities in being forced to conform with the structure of a unique global template.

The model obtained was tested by docking a representative dataset of 50 known hPepT1 substrates. The objectives of a docking analysis are to investigate ligand recognition, to rationalize (and predict) the various affinities for the docked substrates, and to afford an indirect validation for the proposed hPepT1 model. The docking results were further assessed by comparing them with a purposely developed pharmacophore model, which confirmed the key features required for optimal transport.

Results and Discussion

Homology model analysis

Figure 1 shows the ribbon structure of the hPepT1 model, colored by segments, which unveils the typical folding of such transporters with 12 TM segments (TM1–TM12) and a very large extracellular loop (EL5), which fully covers the extracellular side. The transmembrane bundle assumes an elliptical truncated cone shape with the cytoplasmic side larger than the extracellular side. This particular shape is due to the TM segments, which are far from being parallel, and some segments appear markedly staggered at an angle of 30° with respect to the adjacent helices; this distorts the global arrangement of the TM domain. In the bacterial template,[21] the large extracellular loop is placed between TM6 and TM7 and allows the TM bundle to be divided in two specular six-helix domains, whereas in hPepT1 the large extracellular loop between TM9 and TM10 does not allows such a clear-cut subdivision of the TM bundle, which appears more homogeneously arranged.

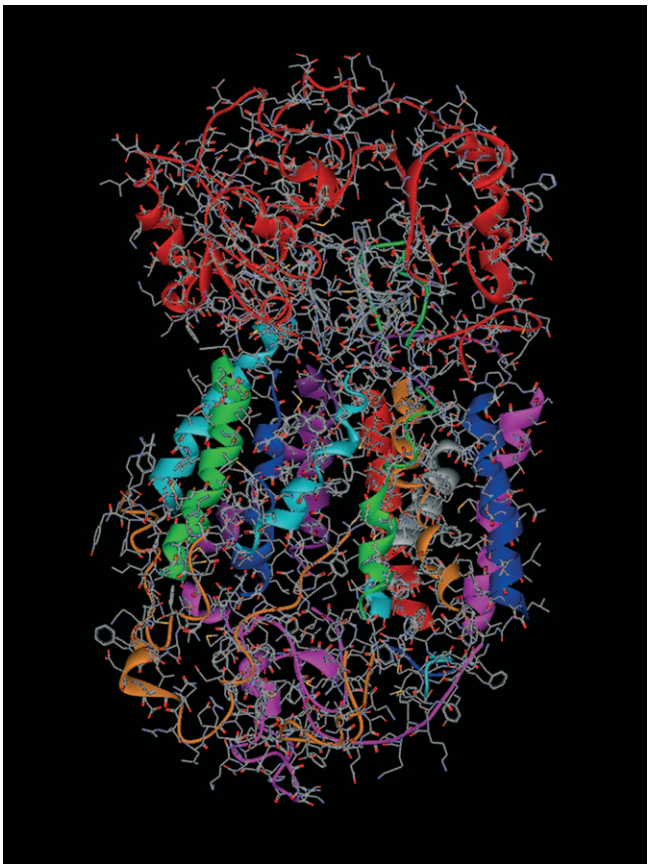

Figure 1. Ribbon structure of the hPepT1 protein model. Readily visible in red is the very large EL5 loop, which fully covers the extracellular side.

Figure S1 and Table S1 in the Supporting Information clarify the correct disposition of TM segments. Specifically, their arrangement does not pertain to numerical order, but it is possible to recognize a more internal group of helices (TM1, TM4,

TM5, TM7, and TM10), which line the central pore and bear the key residues involved in substrate recognition. A second and more external set of TM segments (TM2, TM3, TM6, TM8, TM11, and TM12) define the boundary of the TM bundle and contact phospholipid molecules. Interestingly, the two groups of TM segments also differ in their lipophilic properties. Indeed, the helices that face the central pore are more hydrophilic and were predicted mainly using coiled-coil-containing templates,[22] whereas the more external segments are markedly lipophilic and were modeled using templates characterized by highly hydrophobic surfaces (as described below in the Experimental Section).

Apart from the large EL5 segment, the other extracellular loops are short arc-shaped segments inserted between the TM bundle and the EL5 loop. Their folding is mainly stabilized by polar interactions between neighboring charged residues. Conversely, the large EL5 loop fully covers the extracellular side of TM bundle and consists of two large domains connected by two hinge loops. Both domains include a three-layer ($\alpha\beta$) sandwich. The hinges may confer flexibility to two domains, which could assume closed or open conformations to modulate the accessibility of the binding cavity. Compelling evidence for such putative flexibility is that the template used here (sucrose phosphatase) can also assume two different conformations, as revealed by its crystal structures.[23] Notably, the template is a metalloenzyme, which recognizes some sugar substrates with remarkable selectivity. This suggests that EL5 may also selectively bind sugars or metal ions involved in modulatory effects on hPePT1, as reported by recent experimental studies.[24,25] Apart from CL3, the cytoplasmic loops (CL) are short hydrophilic segments, whose folding appears stabilized by polar interactions. The CL3 segment assumes a more complex conformation, as it begins with an extended region and ends with two parallel helices connected by a short bridge. Globally, these segments are rich in cationic residues, which anchor the protein to phospholipids or contact other intracellular proteins.

Finally, both terminal domains are hydrophilic segments that face the cytosol. The N-terminal domain is a very short segment, whereas the C-terminal domain is an extended arc-shaped segment that lines the intracellular side of the transporter. The folding is stabilized by polar interactions, and the cationic residues have the same roles as described above for CL loops.

Docking results

A bird's-eye analysis of the docking results reveals palpable heterogeneity among the binding modes that the substrates can assume in the hPepT1 cavity. This finding is easily explained considering that no functional group is truly mandatory for bioactivity, as evidenced by SAR, and thus the binding mode is influenced by the physicochemical properties of the substrate's side chains. Nevertheless, a more in-depth analysis of the computed complexes allows identification of hPepT1 residues that are most frequently involved in ligand recognition, as compiled in Table S2 (Supporting Information), which lists the residues involved as well as the ligand-interacting moieties for each substrate.

Overall, the key residues can be subdivided into three groups: 1) residues that preferentially interact with the ligand's charged N terminus, 2) residues that can interact with the ligand's charged C terminus, and 3) residues that primarily contact the side chains. Remarkably, residues of the first two groups mainly undergo polar interactions, whereas those of last group can be involved in both polar and hydrophobic interactions, depending on side chain polarity, even if the apolar contacts are clearly more frequent.

Analysis of the relative frequency of each monitored residue (as listed in Table S2) reveals that only one residue, Tyr588 (TM10), is always involved in ligand recognition. In particular, compounds with the highest affinity (namely those with $pK_i > 0$) are able to stabilize strong interactions between Tyr588 and the ligand's ammonium head group, whereas derivatives with lower affinity cannot realize such an interaction; at best, they contact Tyr588 through the side chains or peptide bonds.

Figure 2 summarizes the most frequent interactions as collected in Table S2. Specifically, the N-terminal ammonium head group probably plays the most critical role, as it interacts with

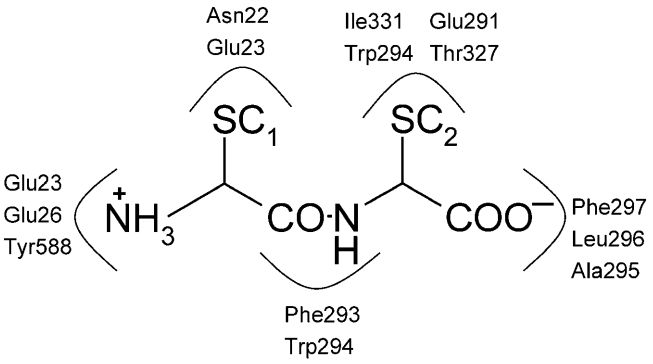

Figure 2. Two-dimensional scheme illustrating the most frequent residues involved in ligand recognition (SC = side chain).

Tyr588 and also forms ion pairs with Glu23 (TM1) and/or Glu26 (TM1). Conversely, the C-terminal carboxylate group appears less involved in ligand recognition, as it stabilizes only H bonds with the backbone of Ala295 (TM7), Leu296 (TM7), and Phe297 (TM7) without forming strong ionic interactions. The pattern of residues that interact with the side chains is clearly more heterogeneous, thus justifying the ability of hPepT1 to interact with structurally diverse substrates. Nevertheless, it is possible to recognize a set of residues involved mainly in the interaction with the N-terminal side chain (SC_1) such as Asn22 (TM1), Glu23, and Phe293, whereas the C-terminal side chain (SC_2) contacts mostly Trp294, Ile331 (TM8), and Ile335 (TM8). Notably, Tyr588 can also contact the side chains to form apolar interactions or undergo π–π stacking.

Although apolar substrates are clearly preferred, docking results suggest that polar side chains can be accommodated both at SC_1, where they can contact Asn22 and Glu23, reinforcing the interactions generated by the N terminus, and at SC_2,

where they can interact with Glu291 (TM7) and Thr327 (EL4). Finally, the central peptide bond can stabilize H bonds with backbone atoms of Phe293 and Trp294. Such interactions can be hindered by bulky side chains, and thus one can conclude that the contacts of peptide groups could partially counterbalance the decreased interactions stabilized by small side chains.

These preliminary observations are in remarkable agreement with known SAR and mutational analyses. Indeed, the different pattern of interactions realized by the charged termini can clearly explain why modifications on the C terminus are better tolerated than those on the N terminus. Again, the scarce interactions realized by peptide bonds can clarify why this function is not mandatory and several substrates without amide groups are successfully transported. Finally, the precise arrangement of the interactions formed by the charged termini (especially the ammonium head group) justify the stereospecificity of the interaction between peptides and hPepT1, as proven by the inability of D-Tyr(OBz)-Ala to form strong H bonds between Tyr588 and the N-terminal ammonium head group. Moreover, results of mutagenesis studies confirm the critical role of Glu26, Phe293, Trp294, and Tyr588, the mutations of which decrease affinity and block transport.[26,27]

As an informative example of the binding mode of the tripeptides, Figure 3 depicts the complex formed with Met-Met-Met, which has the highest affinity among the docked tripeptides. Specifically, Figure 3 shows that the hPepT1 binding cavity can successfully accommodate such a tripeptide and realizes a vast pattern of relevant interactions involving all the ligand functional groups. Indeed, 1) the ammonium head group forms ion pairs with Glu23 and Glu26 in addition to the critical H bond with Tyr588; 2) the C-terminal carboxylate group forms H bonds with the backbone atoms of Tyr588 and Gln587 (TM10) instead of Ala295, Leu296, and Phe297, as is the case with dipeptides; 3) the peptide bonds stabilize H bonds with the backbone atoms of Phe293, Trp294, Leu296, and Phe297; and 4) the side chains form hydrophobic contacts with Cys25 (TM1), Phe293, Trp294, Leu296, Phe297, plus a set of aliphatic residues (such as Ile331, Ile335, and Leu591) not shown in Figure 3 for clarity.

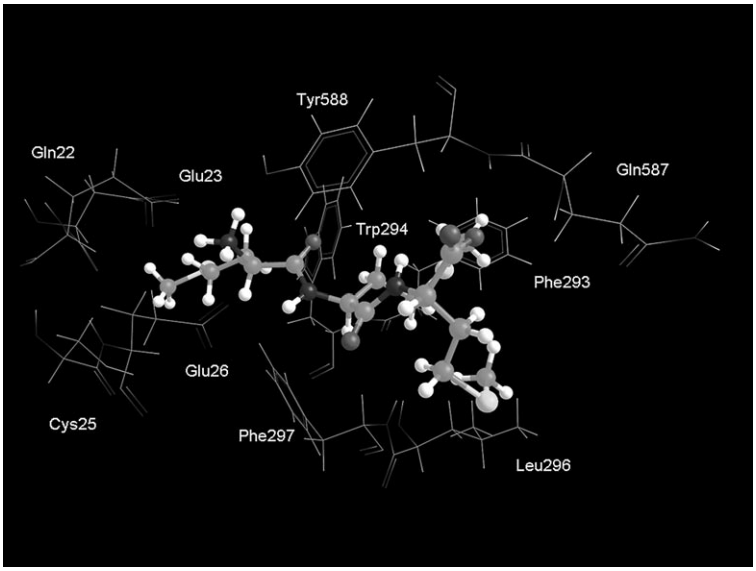

Figure 3. Three-dimensional representation of the complex realized by the highest-affinity tripeptide (Met-Met-Met) in the binding cavity of hPepT1. The main interactions involve: 1) the N terminus with Tyr588, Glu23, and Glu26; 2) the C terminus with Gln587; 3) the side chains with Phe293, Trp294, Leu296, and Phe297.

Taken together, the docking results for tripeptides suggest that the interaction pattern realized by the N-terminal ammonium head group is constant and independent of substrate length, whereas the interactions realized by the carboxylate terminus can change depending on ligand size, even when always involved in H bonds with backbone atoms. Curiously, the peptide bonds in the tripeptides are involved in H bonds with residues that usually interact with the carboxylate group of dipeptide substrates.

Even avoiding a systematic description of all computed complexes, some docking results deserve a deeper analysis. The Ala-Ala dipeptide clearly shows the relevance of H bonds realized by peptide bonds which can successfully compensate for the limited contacts elicited by small side chains. Again, the ligand with highest affinity, namely Tyr(OBz)-Ala, underscores the positive role of π–π stacking that the ligand forms with Phe293, Trp294, Phe297, and Tyr588, while the ether function forms H bonds with Asn22. Finally, the modest affinity of the Gly-containing peptides can be justified by the absence of side chains, and thus interactions which could further stabilize the complexes.

As anticipated, the main feature that characterizes the complexes with compounds of lower affinity (i.e. $pK_i < 0$), is their inability to stabilize strong H bonds between Tyr588 and a given ligand's ammonium head group. The reasons for such a failure are as follows: 1) limited accessibility of the ammonium head group, which is shielded by surrounding bulky groups, as is the case with amoxicillin, where Tyr588 is hindered by the phenoxy ring, with which it forms π–π stacking interactions and thus cannot approach the N terminus; 2) insertion of the ammonium head group in a ring, as is the case with Pro-containing peptides: in such complexes, the protonated group is buried by a cyclic moiety and cannot interact with Tyr588; 3) the lack of an N terminus as is the case for several docked β-lactams: here, Tyr588 forms H bonds with the peptide bonds or carboxylate functions. Conceivably, the ligands without an ammonium terminus are characterized by the lowest affinity values.

Another ligand property that appears to influence hPepT1 affinity is the distance between the two charged termini. In particular, Table 1 suggests that the compounds with highest affinity are those with a distance of about 6 Å between the N and C termini. This value is in agreement with previous COMSIA models,[11] which suggested a distance range of 5.2–5.6 Å between the charged termini to elicit optimal affinity for hPepT1. Conversely, ligands with a greater distance, as with some antibiotics, or with a very small distance,

Table 1. Affinity values, docking results, and physicochemical properties used in Eq. 1.

Compound	K_i [mM]	Zapbind [kcal mol^{-1}]	Distance N+COO− [Å]	Int_ Tyr588	Pred. K_i [mM][a]
Tyr(OBz)-Ala	0.01	−30.64	6.02	1	0.014
Phe(Bz)-Ala	0.02	−30.57	6.01	1	0.026
Val-Phe	0.05	−24.86	4.92	1	0.36
Ala-Pro	0.06	−19.12	5.59	1	0.015
Val-Pro	0.06	−22.24	5.68	1	0.021
Val-ATAA	0.07	−25.38	8.17	1	0.25
Tyr-Ala	0.09	−24.40	3.52	1	0.035
Tyr-Phe	0.09	−31.69	5.36	1	0.25
Ala-Nle	0.09	−22.96	5.87	1	0.035
D-Tyr(OBz)-Ala	0.09	−31.47	8.08	0	0.015
Met-Met-Met	0.1	−37.49	5.25	1	0.13
Val-Pro-Pro	0.1	−30.19	7.55	1	0.47
Val-Tyr	0.1	−34.66	5.11	1	0.21
Ala-Phe-Pro	0.11	−32.32	8.78	1	0.15
Leu-Ala-Arg	0.11	44.71	8.40	1	1.1
Phe-Ala	0.11	−25.06	5.73	1	0.035
Ala-Ala	0.11	−19.05	4.82	1	0.21
Ala-Lys	0.22	75.88	6.15	1	0.3
Glu-Ala	0.25	−20.24	3.59	1	0.024
Lys-Asp	0.33	−25.46	5.23	1	0.49
Asp-Asp	0.41	−10.30	5.81	1	0.37
Ciclacillin	0.50	−29.59	8.33	1	1.2
Ala-ATAA	0.51	−23.51	7.72	0	0.29
Phe-ATAA	0.95	−29.10	8.55	1	0.67
Gly-His	1.00	43.06	3.50	1	1.3
Pro-Pro	1.2	−20.64	5.64	0	1.6
Lys-Glu	1.3	−26.16	5.43	0	1.4
Pro-Phe-Lys	2	38.58	9.38	1	4.6
Cloxacillin	2.5	−63.79	0.00	0	2.2
(2*S*)-1-anilino-1-oxopropan-2-amine	2.9	16.23	0.00	1	2.4
Gly-His-Lys	4.1	40.34	9.45	1	18
ATAA	4.9	−16.87	3.25	0	9.3
Cefadroxil	7.5	−33.01	8.45	0	3.6
Pro-Asp	9.8	−20.15	2.70	0	9
Cefixime	12	−48.90	0.00	0	9.4
Oxacillin	12	−63.98	0.00	0	5.6
Metampicillin	13	−52.37	0.00	0	3.7
Ampicillin	14	−31.31	8.45	0	4.6
Pro-Gly-Gly	16	−23.45	8.57	0	27
Pro-Glu	20	−19.45	7.18	0	15
Penicillin V	21	−54.44	0.00	0	4.2
Cefodizime	22	−43.52	0.00	0	7.2
Amoxicillin	25	−32.10	8.47	0	5.3
Leu-Gly-Gly	25	−26.40	8.04	0	10
Benzylpenicillin	40	−48.36	0.00	0	3
Cefotaxime	50	−59.83	0.00	0	0.5
Piperacillin	61	−53.03	0.00	0	3.7
Cefepime	70	−41.67	0.00	0	58
Cephaloridine	100	−37.40	0.00	0	11
Cefpodoxime	110	−57.78	0.00	0	11

[a] K_i value predicted by HypoGen.

as found in some ligands in which the charged moieties are involved in intramolecular ion pairs, show lower affinity as they are probably unable to generate the rich pattern of polar interactions illustrated in Figure 2. Thus, the intermediate affinity of 2-aminothiazole-4-acetic acid (ATAA) derivatives[28] can be explained considering that such a constrained moiety prevents the formation of intramolecular ion pairs even if the two charged groups are arranged too far to fully optimize their key polar interactions.

Table 1 lists the 50 docked substrates with their affinity values (expressed as K_i, mM) as well as the values of the docking score and properties exploited in the predictive equation 1. Notably, all score functions were preliminarily tested as computed by FRED software. Table 1 lists only Zapbind scores, because they afforded the best correlation with the experimental affinities and were thus inserted in Eq. 1. In particular, such a function accounts for the ionic interactions being based on a combination of surface contact terms and Poisson–Boltzmann energy approximations (as computed by ZAP module).[29]

$$\begin{aligned} p\mathit{K}_i = &-0.235\ (\pm 0.0366)\ \Delta\,\text{distance} - 7.912 \times 10^{-3}\ (\pm 0.00268)\ \text{Zapbind} \\ &+ 1.534\ (\pm 0.167)\ \text{Int\_Tyr588} - 0.489\ (\pm 0.194) \qquad (1) \\ &n = 50,\ r^2 = 0.85,\ s = 0.45,\ F = 89.92 \end{aligned}$$

The affinity prediction afforded by Eq. 1 proved successful as demonstrated by the regression parameters and Figure 4. The parameters included in Eq. 1 allow a deeper analysis of the

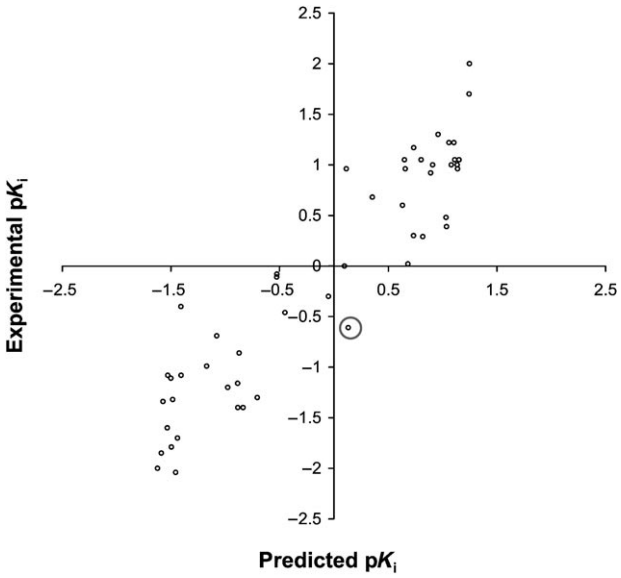

Figure 4. Correlations between experimental affinities (expressed as pK_i) and affinity values as predicted by Eq. 1.

main factors that influence hPepT1 affinity. More specifically, Eq. 1 indicates that the affinity is clearly related to the arrangement of charged groups expressed by the difference (Δ distance) between the distance values compiled in Table 1 and the optimal distance (6.02 Å) as evidenced by the ligand with highest affinity. This reflects the significant role of the contacts elicited by the charged groups in hPepT1 recognition, although compounds without such charged moieties also show a detectable affinity for the transporter. Despite the known predilection of hPepT1 for hydrophobic residues, the role of the Zapbind score further emphasizes the relevance of the polar interactions formed by charged groups on the ligand. Fi-

nally, given the vastly beneficial role of the interaction between Tyr588 and the ammonium head group, we introduced a binary descriptor that is assigned a value of 1 for substrates that form such a reinforced H bond, and a value of 0 otherwise. Conceivably, such a descriptor significantly enhances the predictive power of Eq. 1. The parameters included in Eq. 1 underscore the relevance of polar interactions, whereas the role of hydrophobic contacts appears completely ignored. Interestingly, the insertion of the apolar area descriptor (ASA) in Eq. 1 slightly increases its regression parameters, thus confirming the beneficial role of apolar contacts. The modest improvement in r^2 (0.85 vs. 0.87), however, does not justify the inclusion of this descriptor, as confirmed by the decrease in F (89.92 vs. 67.75).

A more in-depth analysis of Figure 4 confirms the goodness of the affinity predictions; considering the classification in ligands with high affinity ($pK_i > 0$) from those of lower affinity ($pK_i < 0$) as defined by Cartesian axes, it is apparent that Eq. 1 is able to successfully discriminate among the docked ligands. Only one derivative is incorrectly predicted, giving a false positive (indicated by the circle). Specifically, the poorly predicted compound is the tripeptide Gly-His-Lys, which, in fact, gave a good docking pose (as exemplified by its H bond with Tyr588).

The soundness of the obtained complexes and their remarkable agreement with both SAR and mutagenesis studies afford an encouraging validation of the proposed hPepT1 model and its computed binding modes. Moreover, Eq. 1 suggests that these results can be successfully used to predict the ability of a molecule to interact with hPepT1, clearly distinguishing between high- and low-affinity derivatives.

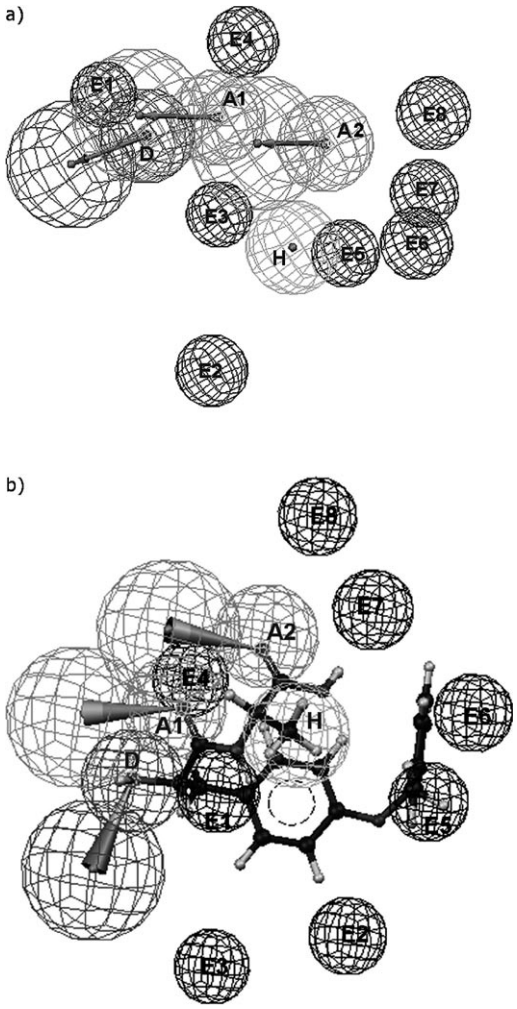

Figure 5. Pharmacophore model for hPepT1 ligands. a) Pharmacophore features as derived by the HypoRefine approach (H bond acceptor, A1 A2; H bond donor, D; hydrophobic region, H; excluded volumes, E1–E8). b) Pharmacophore model mapped on the docking pose of Tyr(OBz)-Ala.

Pharmacophore mapping

With a view to confirm the docking results, enriching our knowledge about the hPepT1 binding modes, HypoRefine was exploited to derive an automated SAR pharmacophore model for hPepT1 ligands by using a training set of 25 compounds taken from the docked dataset. The compounds of the training set were suitably selected to have a broad range of affinity values (expressed as K_i, from 0.01 to 70 mM). The 10 hypotheses, which showed the best correlation between estimated and measured affinity values, are compiled in Table S3 (Supporting Information) as well as the results of statistical significance and predictive ability.

The quality of the generated pharmacophore hypotheses was evaluated by considering the cost functions as computed by the HypoGen module during hypothesis generation. The top-ranked pharmacophore model (Hypo1) had the best predictive power and statistical significance and was characterized by the highest cost difference (65.316), the lowest RMS (0.748), and the best correlation coefficient (0.939). These values emphasize the great predictability of the 3D-QSAR pharmacophore as evidenced by predicted pK_i values compiled in Table 1, and confirm that it did not come about by chance.

Figure 5a shows the selected 3D hypothesis which consists of one hydrophobic region (H), two hydrogen bond acceptors (A1 A2), one hydrogen bond donor (D), and eight excluded volume sites (E1–E8) in a specific three-dimensional orientation. To check the predictive power of this model, the selected hypothesis was used to predict the affinity values for compounds also not included in the training set (as seen in Table 1). Globally, the correlation between experimental and predicted affinities is good ($r^2 = 0.78$) and quite similar to that obtained by docking simulations. This result further confirms the goodness and predictive power of the developed pharmacophore model. Furthermore, with the aim of comparing such a pharmacophore model with the docking results, all ligands, in their bound conformation, were mapped onto the hypothesis. Although the described heterogeneity among the docking results does not allow a quantitative comparison of all docking complexes with pharmacophore, the compounds with highest affinity give a precise mapping onto the pharmacophore model (as shown in Figure 5b), the key elements of which confirm the precise role of each hPepT1 residue.

Thus, Figure 5b depicts the pharmacophore regions mapped on the docking pose of the highest-affinity derivative, Tyr(OBz)-Ala. Notably, the ammonium head group occupies the

H bond donor region which overlaps Tyr588, Asp23, and Asp26; the carboxylate terminus and the peptide bond map to the H bond acceptor regions corresponding to the backbone of Phe293, Trp294, Leu296, and Phe297; and the hydrophobic region corresponds to the C-terminal side chain, suggesting that a hydrophobic residue is particularly beneficial at such a position. Finally, the excluded volumes represent the position of other amino acids not involved in ligand recognition but that are able to create steric clashes with the lower-affinity compounds. Interestingly, such excluded volumes are mostly located near the C terminus, whereas the N terminus appears less sterically constrained. Indeed, the Tyr(OBz)-Ala dipeptide confirms that the modification of the N-terminal side chain with bulky moieties can increase affinity.

Conclusions

This study describes the generation of a reliable 3D model for hPepT1 that is consistent with published structural and functional information. In particular, the model confirms the global arrangement of the transmembrane helices, which define a central vestibule (termed "bubble"), where the substrates can cross the membrane. The proposed model also affords a more detailed description of the relative position of TM helices, refuting the implicit assumption that helices adjacent in sequence are also physically adjacent,[30] as the TM helices are subdivided into two concentric rings, and their arrangement does not agree with numerical order (see Figure S1). Again, docking results confirm that the key residues involved in ligand recognition are mainly located in TM1, TM8, and TM10, whose critical roles were already evidenced by mutational analyses. Finally, the proposed model emphasizes the relevance of the large EL5 segment, which fully covers the extracellular side of the transporter and probably acts as a modulator of vestibule gating.

Moreover, the docking results confirm that the model can be successfully used to predict ligand affinity for the transporter. The clear agreement between two different computational techniques (docking simulations and pharmacophore models) yields further validation for the proposed binding modes, suggesting that the described model can be used to design novel peptidomimetics with improved affinity for hPepT1. Therefore, the obtained model can find important application in optimizing the structure of potential hPepT1 substrates and in screening peptidomimetic libraries to unveil molecules that may be absorbed via hPepT1 transport. More intriguingly, this model can be exploited to design compounds in which a poorly absorbed drug is combined with a suitable peptidomimetic moiety to generate a prodrug carrier that is recognized by hPepT1, which also transports the drug.

The basic idea is to adapt the structure of a drug molecule to the substrate features required by hPepT1 for optimal recognition. The prodrug should possess sufficient solubility in the gastrointestinal lumen, be absorbed across the intestinal brush border membrane via hPepT1, and then be hydrolyzed in the mucosal cell, blood, or liver. The design of prodrugs of poorly permeable analogues that target the PepT1 transporter has been demonstrated to be a successful strategy to improve oral bioavailability.[31] In addition, recent reports on the expression of hPepT1 in some cancer cell lines describe the possibility of exploiting such transporters for anticancer therapy.[32]

Experimental Section

Generation of the hPepT1 model

The amino acid sequence of the human PepT1 transporter was retrieved from the Swiss-Prot database (entry code P46059 S15 A1_HUMAN). The model of the human transporter was generated by fragments with a strategy that involves the following steps: 1) the fragmentation of the primary sequence into 25 segments (namely 12 TM segments TM1–TM12, six extracellular loops EL1–EL6, five cytoplasmic loops CL1–CL5, and two terminal segments NT and CT, as compiled in Table S4, Supporting Information); 2) the homology modeling of these segments separately using Fugue;[33] and 3) the assembly of fragments using the structure of the *E. coli* lactose permease LacY (PDB code: 1PV6) as a final template, the use of which is justified by its known homology as illustrated by pairwise alignments between the sequences of LacY and hPepT1 (Figure S2, Supporting Information).

For each segment, Fugue is able to produce several realistic models, and the best structure was chosen considering the following major conditions: 1) the predicted secondary structure from the sequence alignment using ClustalX (see Figure S2); b) the lack of unpredicted gaps; c) the prediction score (ZSCORE) as calculated by the Fugue program; d) the helical conformation of the transmembrane segments with a slight characteristic bend in helices containing proline and glycine residues; and e) the global "U" shape of the loops in which the two ends are close enough to join to TM segments.

Table S4 (Supporting Information) lists the templates used by Fugue to generate the selected model for each fragment. Even avoiding a systematic analysis, Table S4 allows some relevant considerations: 1) the bacterial LacY is never used as a fragmental template, indicating that the homology concerns the global topology rather than local folding; 2) several TM segments are modeled using templates characterized by coiled-coil structures,[22] and this can be explained by considering that such helical motifs often play key roles in molecular recognition processes and indeed the modeled TM segments are involved in substrate recognition (TM1, TM7, TM8, TM9); 3) more external TM segments are modeled using AKAP proteins, which are characterized by highly hydrophobic surfaces;[34] 4) the templates used for the loops are surely more heterogeneous, including enzymes, receptors, and regulatory peptides; 5) the longest segment (EL5) is predicted by using a bacterial sucrose phosphatase that shows mixed α/β topology.[23] The use of such a template for the EL5 loop can be justified by sequence alignment, which confirms an interesting homology (identity: 11.3%, similarity: 30.9%).

The assembly of predicted fragments was performed by superimposing the backbone of a fragment with that of the corresponding segment in the structure of LacY and manually connecting the adjacent segments using VEGA software.[35] Specifically, the superimposition involved the Cα atoms of transmembrane helices only, as the loop arrangements, which were, however, defined considering the corresponding segment of experimental template, are clearly defined by the position of TMs, and their conformation was further relaxed by subsequent molecular dynamics (MD) simulations (while

the TM bundle remains harmonically constrained during MD, as described below).

With the backbone completed, side chains and hydrogen atoms were added using VEGA. According to physiological pH, Arg, Lys, Glu, and Asp residues were kept ionized, while His residues were considered neutral by default. After a careful scrutiny of the obtained structure to avoid unphysical conditions, the model underwent an initial minimization until the RMS gradient was equal to 1 to discard high-energy interactions, followed by a local minimization until RMS = 0.05, at which all atoms were kept fixed except for those within a sphere of 5.0 Å radius around the manually connected bonds (at the fragment ends). The model was optimized by a final minimization composed of two phases: first, a minimization without constraints until RMS = 0.1 kcal mol^{-1} $Å^{-1}$, and then a second minimization with backbone fixed until RMS = 0.01 kcal mol^{-1} $Å^{-1}$ to preserve the predicted structure.

To gain a better relaxation and a more accurate arrangement of the hPepT1 model, an MD equilibration was performed in vacuo. The simulations were carried out in three phases: 1) heating from 0 to 300 K over 3000 iterations (3 ps, i.e., 1 K per 10 iterations); 2) starting equilibration for 2500 ps, during which the transmembrane backbone was kept fixed; and 3) equilibration for 7500 ps, during which the transmembrane backbone was harmonically restrained with decreasing harmonic force constants. In detail, the harmonic force constant was equal to 1 (1000 kJ mol^{-1} nm^{-2}) at the beginning of the simulation and was then divided in two every 1.5 ns (then five MD simulations were performed with the harmonic force constant equal to 1, 0.5, 0.25, 0.12, and 0.06). Globally, the MD simulations lasted 10 ns, and the helices were correctly preserved with the harmonic force constant equal to 0.06. The last frame was used for the trimer assembly after a final minimization until RMS = 0.01 (with harmonic force constant equal to 0.06).

The MD simulations had the following general characteristics: constant temperature at 300 ± 10 K by means of Langevin's algorithm; Lennard–Jones (L–J) interactions were calculated with a cutoff of 10 Å, and the pair list was updated every 20 iterations; Newton's equation was integrated with the r-RESPA method every 4 fs for long-range electrostatic forces: 2 fs for short-range nonbonded forces, and 1 fs for bonded forces; a frame was stored every 5 ps, yielding 2000 frames. All calculations were carried out on an eight-node Tyan-VX50 system. All minimizations were performed using the conjugated gradients algorithm. The package Namd 2.51[36] was used with the force field CHARMM v22 and Gasteiger's atomic charges.

Ligand setup and docking analyses

The set of 50 hPepT1 ligands was compiled from published data.[11,37] Table 1 lists the modeled substrates that can be structurally subdivided into peptides and β-lactam antibiotics. The affinities of such ligands (expressed as K_i and listed in Table S2, Supporting Information) range from 0.01 to 110 mM, and are equally divided among substrates with negative and positive pK_i values.

The ligands were modeled in their ionized form, as they may be involved in ligand recognition. The ligand structure was built using VEGA software, and the overall geometry and atomic charges were optimized using MOPAC 6.0. Their conformational profile was explored by a Monte Carlo procedure (as implemented in VEGA) which generated 1000 conformers by randomly rotating the rotors. All geometries so obtained were optimized and clustered according to similarity to discard redundant conformers; in detail, two geometries were considered as nonredundant if they differed by $>60^\circ$ in at least one torsion angle.

The docking and scoring procedure involved extensive rigid-body sampling with the OpenEye Scientific Software package FRED (OpenEye Scientific Software, Santa Fe, NM, USA). Briefly, the FRED-based sampling was performed in a side box of 8 Å around some known residues involved in ligand recognition (i.e. Glu26, Trp294, and Tyr588). The obtained complexes were then refined focusing the minimization on the atoms inside a sphere of 10 Å radius around the bound ligand. The minimized complexes were finally used to recalculate the docking scores.

Pharmacophore generation

All structures were generated using the 2D/3D editor sketcher in the Catalyst 4.10 software package (Catalyst v 4.10, 2006, Accelrys Inc., San Diego, CA, USA) and submitted to energy minimization and conformational analysis (maximum number of conformers: 250, generation type: best quality, energy range: 10 kcal mol^{-1}).

Catalyst provides a dictionary of chemical features found to be important in drug–enzyme and drug–receptor interactions. These are hydrogen bond donors, hydrogen bond acceptors, aromatic ring, hydrophobic (aliphatic or aromatic) groups, and positively and negatively charged groups. The inter-feature spacing penalty was decreased from its default value to 100 pm. No constraint on the minimum and maximum number of each type of feature in the reported pharmacophores was applied.

The pharmacophore model was generated by HypoRefine 4.10 using a training set of 25 compounds.[38] Specifically, we select compounds with various degrees of activity that spans five orders of magnitude, making this a good data set for the HypoRefine module. The uncertainty value of compound activity, which represents the ratio range of uncertainty in the activity value based on the expected statistical variability of biological data collection, was set to 3.

On the basis of the atom types in the molecules of the training set, five chemical feature types were used in the HypoGen run: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic (H), hydrophobic aliphatic (Z), and hydrophobic aromatic (Y) groups.

Acknowledgements

Financial support for this research by MiUR is gratefully acknowledged.

Keywords: homology modeling · hPepT1 · intestinal absorption · molecular docking · pharmacophore mapping

[1] E. Lipka, J. Crison, G. L. Amidon, *J. Controlled Release* **1996**, *39*, 121–129.
[2] D. M. Oh, H. K. Han, G. L. Amidon, *Pharm. Biotechnol.* **1999**, *12*, 59–88.
[3] G. Vistoli, A. Pedretti, B. Testa, *Drug Discovery Today* **2008**, *13*, 285–294.
[4] U. Norinder, C. A. Bergström, *ChemMedChem* **2006**, *1*, 920–937.
[5] H. Daniel, *Annu. Rev. Physiol.* **2004**, *66*, 361–384.
[6] B. Brodin, C. U. Nielsen, B. Steffansen, S. Frøkjaer, *Pharmacol. Toxicol.* **2002**, *90*, 285–296.
[7] Y. J. Fei, Y. Kanai, S. Nussberger, V. Ganapathy, F. H. Leibach, M. F. Romero, S. K. Singh, W. F. Boron, M. A. Hediger, *Nature* **1994**, *368*, 563–566.

[8] H. Daniel, B. Spanier, G. Kottra, D. Weitz, *Physiology (Bethesda)* **2006**, *21*, 93–102.
[9] Y. Sai, *Drug Delivery Syst. Drug Metab. Pharmacokinet.* **2005**, *20*, 91–99.
[10] M. Brandsch, I. Knütter, F. H. Leibach, *Eur. J. Pharm. Sci.* **2004**, *21*, 53–60.
[11] S. Gebauer, I. Knütter, B. Hartrodt, M. Brandsch, K. Neubert, I. Thondorf, *J. Med. Chem.* **2003**, *46*, 5725–5734.
[12] A. Biegel, S. Gebauer, B. Hartrodt, M. Brandsch, K. Neubert, I. Thondorf, *J. Med. Chem.* **2005**, *48*, 4410–4419.
[13] B. S. Vig, T. R. Stouch, J. K. Timoszyk, Y. Quan, D. A. Wall, R. L. Smith, T. N. Faria, *J. Med. Chem.* **2006**, *49*, 3636–3644.
[14] S. Ekins, G. F. Ecker, P. Chiba, P. W. Swaan, *Xenobiotica* **2007**, *37*, 1152–1170.
[15] R. Liang, Y. J. Fei, P. D. Prasad, S. Ramamoorthy, H. Han, T. L. Yang-Feng, M. A. Hediger, V. Ganapathy, F. H. Leibach, *J. Biol. Chem.* **1995**, *270*, 6456–6463.
[16] V. N. Kasho, I. N. Smirnova, H. R. Kaback, *J. Mol. Biol.* **2006**, *358*, 1060–1070.
[17] A. Pedretti, M. Villa, M. Pallavicini, E. Valoti, G. Vistoli, *J. Med. Chem.* **2006**, *49*, 3077–3085.
[18] A. Pedretti, G. Vistoli, *Bioorg. Med. Chem.* **2007**, *15*, 3054–3064.
[19] A. Pedretti, L. De Luca, C. Sciarrillo, G. Vistoli, *ChemMedChem* **2008**, *3*, 79–90.
[20] A. Pedretti, C. Marconi, C. Bolchi, L. Fumagalli, R. Ferrara, M. Pallavicini, E. Valoti, G. Vistoli, *Biochem. Biophys. Res. Commun.* **2008**, *369*, 648–653.
[21] J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, S. Iwata, *Science* **2003**, *301*, 610–615.
[22] A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37–78.
[23] S. Fieulaine, J. E. Lunn, F. Borel, J. L. Ferrer, *Plant Cell* **2005**, *17*, 2049–2058.
[24] V. M. D'Souza, D. J. Buckley, A. R. Buckley, G. M. Pauletti, *J. Pharm. Sci.* **2003**, *92*, 594–603.
[25] M. Okamura, T. Terada, T. Katsura, H. Saito, K. Inui, *Pharm. Res.* **2003**, *20*, 1389–1393.
[26] D. Meredith, R. A. Price, *J. Membr. Biol.* **2006**, *213*, 79–88.
[27] A. A. Kulkarni, I. S. Haworth, T. Uchiyama, V. H. Lee, *J. Biol. Chem.* **2003**, *278*, 51833–51840.
[28] A. Biegel, S. Gebauer, B. Hartrodt, I. Knütter, K. Neubert, M. Brandsch, I. Thondorf, *Eur. J. Pharm. Sci.* **2007**, *32*, 69–76.
[29] J. A. Grant, B. T. Pickup, A. Nicholls, *J. Comput. Chem.* **2001**, *22*, 608–640.
[30] M. B. Bolger, I. S. Haworth, A. K. Yeung, D. Ann, H. von Grafenstein, S. Hamm-Alvarez, C. T. Okamoto, K. J. Kim, S. K. Basu, S. Wu, V. H. Lee, *J. Pharm. Sci.* **1998**, *87*, 1286–1291.
[31] G. M. Friedrichsen, W. Chen, M. Begtrup, C. P. Lee, P. L. Smith, R. T. Borchardt, *Eur. J. Pharm. Sci.* **2002**, *16*, 1–13.
[32] D. E. Gonzalez, K. M. Y. Covitz, W. Sadee, R. J. Mrsny, *Cancer Res.* **1998**, *58*, 519–525.
[33] J. Shi, T. L. Blundell, K. Mizuguchi, *J. Mol. Biol.* **2001**, *310*, 243–257.
[34] P. Banky, M. Roy, M. G. Newlon, D. Morikis, N. M. Haste, S. S. Taylor, P. A. Jennings, *J. Mol. Biol.* **2003**, *330*, 1117–1129.
[35] A. Pedretti, L. Villa, G. Vistoli, *J. Mol. Graph. Model.* **2002**, *21*, 47–49.
[36] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283–312.
[37] S. B. Larsen, F. S. Jørgensen, L. Olsen, *J. Chem. Inf. Comput. Sci.* **2008**, *48*, 233–241.
[38] H. Li, J. Sutter, R. Hoffmann, in *Pharmacophore Perception, Development, and Use in Drug Design*, International University Line: La Jolla, **2000**, 172 – 189.

Received: June 17, 2008
Revised: August 30, 2008
Published online on October 31, 2008

