Bioorganic & Medicinal Chemistry 19 (2011) 4544–4551

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Fragmental modeling of hPepT2 and analysis of its binding features by docking studies and pharmacophore mapping

Alessandro Pedretti [a], Laura De Luca [b], Cristina Marconi [a], Luca Regazzoni [a], Giancarlo Aldini [a], Giulio Vistoli [a,*]

[a] Dipartimento di Scienze Farmaceutiche 'Pietro Pratesi', Facoltà di Farmacia, Università degli Studi di Milano, Via Mangiagalli, 25, I-20133 Milano, Italy
[b] Dipartimento Farmaco-Chimico, Facoltà di Farmacia, Università di Messina, Viale Annunziata, I-98168 Messina, Italy

ARTICLE INFO

Article history:
Received 28 September 2010
Revised 1 June 2011
Accepted 8 June 2011
Available online 16 June 2011

Keywords:
hPepT2
Drug transporters
Pharmacokinetic prediction
Fragmental homology modeling
Molecular docking
Pharmacophore mapping

ABSTRACT

Over the last years, considerable progress has been made for the identification and characterization of drug transporters, and several modeling studies have been undertaken to predict their effects on ADME profiling. Thus, this study was focused on the peptide transporter hPepT2, which influences the regional pharmacokinetics in brain, the reabsorption from renal tubular fluid and the pulmonary delivery. A reliable model for hPepT2 was generated by fragments based on the resolved structure of the homologue lactose permease LacY and the structure is made available as Supplementary data. The interaction capacities of such a model were explored by docking a set of 75 known ligands. Docking results underlined the predilection of hPepT2 for highly hydrophobic ligands and the key role of ionic interactions elicited by both charged termini. The docking results were further verified developing a pharmacophore model which clarified the key features required for an optimal hPepT2 affinity and confirmed the main factors governing the hPepT2/hPepT1 selectivity. The soundness of the docking results and the agreement with the pharmacophore mapping afford an encouraging validation for the proposed hPepT2 model and suggest that it can be conveniently exploited to design peptide-like molecules with an improved affinity for this transporter.

1. Introduction

More than half of drug candidates fail during clinical trials due to an unsuitable pharmacokinetic profile. For this reason, the recent strategies in drug discovery and development give attention to ADMET profiling of new molecules with the clear aim to select (and develop) only drug-like compounds with an optimal pharmacokinetic profile.[1] Such an early pharmacokinetic screening requires the assessment of useful and relevant molecular properties that allow a reliable prediction of drug-likeness also for large molecular libraries. As a result, much research effort has been invested in developing in silico tools to predict physicochemical properties relevant to pharmacokinetic profile, for example, aqueous solubility and lipophilicity.[2]

Over the last two decades, significant progress has been made in understanding the remarkable role of transporters in influencing the pharmacokinetic profile of drugs.[3,4] Consequently, computational analyses of the molecular recognition between drugs and transporters have attracted increasing interest.[5] Such predictions can involve several computational approaches including 3D-QSAR methods, pharmacophore mapping and docking analyses,[6] although the last ones appear markedly more difficult since the three-dimensional structure of such transporters is seldom available and has to be generated by homology techniques.[7]

Among the major transporters involved in ADMET profiling, the superfamily of Proton-Coupled Oligopeptide Transporters (POTs) includes four members in humans (PepT1, PepT2, PHT1 and PHT2),[8–10] and is characterized by the ability to transport small peptides and peptidomimetics across biological membranes. Such transporters show unique profiles of expression which are clearly related to their functions in vivo.[11] Specifically, hPepT2 (gene *SLC15A2*), on which this study is focused, is expressed in the *pars recta* of proximal tubule of kidney, lung, nasal mucosa and mammary gland as well as in brain, where it is predominantly localized into astrocytes, subependymal cells, ependymal cells and epithelial cells of the choroid plexus.[12] Studies on Pept2 knockout mice revealed that the major effects of PepT2 are: (1) effluxing substrates from cerebrospinal fluid (CSF) into choroid plexus thus affecting regional pharmacokinetics in brain, and (2) reabsorbing substrates from renal tubular fluid into proximal tubules thus influencing systemic pharmacokinetics and exposure.[13] Moreover, the significant expression of hPepT2 in pneumocytes, bronchial epithelium, and endothelium of the lung small arteries renders this transporter an interesting target for the pulmonary delivery of peptides and peptidomimetics.[14] hPepT1 and hPepT2 show quite similar

* Corresponding author.
E-mail address: giulio.vistoli@unimi.it (G. Vistoli).

0968-0896/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2011.06.027

structural requirements for an optimal transport even though a careful analysis of the binding data reveals that hPepT2 transports its substrates with a 10–15-fold higher affinity than hPepT1 and appears more selective than hPepT1, the requirements for substrate recognition being narrower and more specific. Besides di- and tripeptides, hPepT2 recognizes β-lactams, ACE inhibitors, valacyclovir, sartans, and some peptidomimetics, whereas single amino acids and tetrapeptides are not recognized by both transporters. Globally, the structural requirements for hPepT2 can be summarized as follows: (i) compounds in L-configuration show higher affinities than those in D-configuration; (ii) high affinity substrates feature marked hydrophobicity; (iii) charged termini are surely beneficial but not mandatory for the affinity.[15] Such qualitative criteria have been translated in QSAR models mainly using 3D approaches.[16] Nonetheless, full-length homology models for hPepT2 and docking simulations with the aim to analyze the substrate recognition at an atomic level are not yet reported in the literature.

hPepT2 is a 729 residues protein whose transmembrane bundle is composed by 12 membrane-spanning helices, with a large extracellular loop between the transmembrane regions 9 and 10, while amino and carboxyl termini face the cytosol.[17] The few data about the hPepT2 binding site derive from the analysis of its genetic variants and evidence that Arg57 is required for the transport, since its mutation abolishes the activity,[18] presumably because it contacts the substrate carboxyl terminus. Given its transmembrane nature, the hPepT2 structure cannot be easily resolved, although it can be modelled considering its known homology with both the crystallized bacterial lactose permease LacY[19] and the recently resolved bacterial transporter from *Shewanella oneidensis*.[20]

In a recent study, we have generated a reliable model for the human peptide transporter hPepT1 and we have explored its interaction capacities by docking simulations and pharmacophore analyses.[21] In the present study, we focused the attention on hPepT2 using a computational strategy very similar to that exploited by the first study. More in detail, the 3D model of hPepT2 was generated by fragments[22] based on the resolved structure of the homologue lactose permease LacY.[23] The reliability of such a model was assessed by docking a dataset of 75 known hPepT2 substrates. Docking results were then substantiated by pharmacophore mapping which allows the recognition of key features for the affinity. Finally, the main factors governing the hPepT1/hPepT2 selectivity were examined by comparing docking and pharmacophoric results of this study with those of the previous one. Globally, these two computational studies should allow a reliable prediction of the likelihood with which oligopeptides and peptidomimetics can be recognized by such carriers. This information may be of particular relevance for hydrophilic molecules which cannot cross the biological membranes through passive permeation.

2. Results

2.1. Analysis of the hPepT2 homology model

The homology model of hPepT2, whose quality was evaluated as detailed under Methods, is made available (as pdb file) as Supplementary data so that other researchers can use this structure for their own simulations or compare it with other generated models. Due to the known similarity between hPepT1 and hPepT2, only a brief description for the hPepT2 model is given here. Further structural details can be found in the previous study.[21] Figure 1 depicts the tube structure of the hPepT2 model, coloured by segments, and shows the typical folding of such transporters with 12 transmembrane segments (TM1-12) and a very large extracellular loop (EL5), which covers the extracellular side and may modulate the channel entrance.

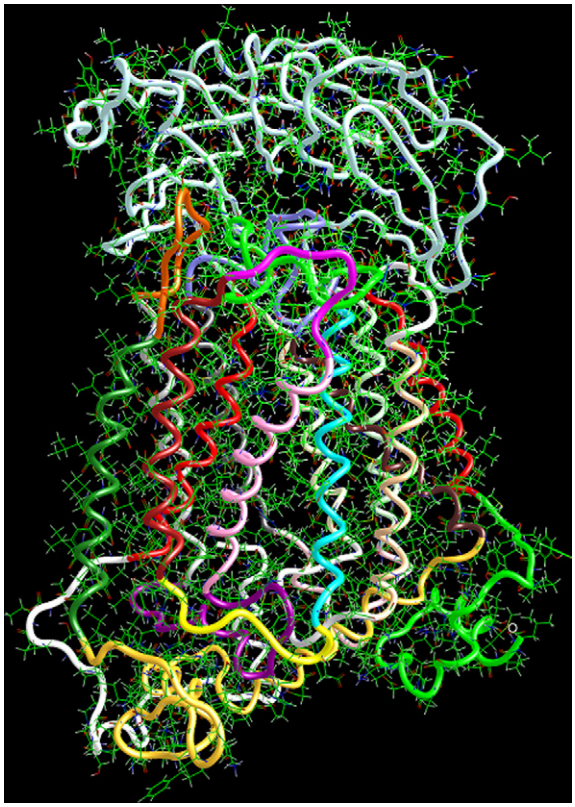

Figure 1. Tube structure of the hPepT2 protein. In azure is well visible the large EL5 loop which fully covers the extracellular side.

The transmembrane bundle assumes an elliptical truncated conic shape with the cytoplasmic side larger than the extracellular one. This particular shape is due to the TM segments, which are far from being parallel, and some segments appear markedly staggered with an angle of 30° in respect to the adjacent helices, thus distorting the global arrangement of the TM domain. Table S1 (Supplementary data) helps to clarify the correct disposition of each TM segment. As already observed for hPepT1, the arrangement of TM segments does not agree the numerical order, but it is possible to recognize a more internal group of helices (namely, TM1, TM4, TM5, TM7, TM10), which line the central pore and bear the key residues involved in substrate recognition, and a second more external set of TM segments (TM2, TM3, TM6, TM8, TM11, TM12), which delimit the TM bundle and contact phospholipidic molecules. Interestingly, the two groups of TM segments also differ for their lipophilic properties. As reported by Table S1, the helices facing the central pore are, on average, more polar, while the more external segments include some markedly lipophilic helices (e.g., TM6, TM8, TM12). The significant correlation ($r^2 = 0.94$) between the distances compiled in Table S1 (Supplementary data) and the corresponding values as computed for hPepT1 (taken from previous study[21]) confirms that the topology of the TM bundle is highly conserved within the POT superfamily.

Apart from the large EL5 segment, the other extracellular loops are short arc-shape 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 shows a β propeller-like fold which is characterized by two β strands packed in a semi-circular array. Since such a motif (also called WD repeat) is typical of segments involved in protein–protein interactions,[24] one can suppose the also the EL5 loop may modulate the pore opening interacting with other proteins and peptides. The cytoplasmic loops are short hydrophilic arc-shape

segments apart from the longer CL3 loop which shows a mixed αβ topology and covers a large part of the intracellular side. Finally, both terminal domains face the cytosol and reveal predominant helical motifs mainly stabilized by polar interactions.

2.2. Docking results

A global analysis of the obtained complexes for the 75 analyzed hPepT2 substrates (as generated by minimizing FRED-based docking poses and compiled in Table 1) reveals a significant homogeneity among them, allegedly because almost all complexes are tethered by two ion-pairs which involve both ligand charged termini, namely the N-terminus with Glu622 and the C-terminus with Arg57 (as schematized by Fig. 2). Hence, the distance between Glu622 and Arg57 (10.5 Å) plays a key role in determining the ligand recognition by hPepT2. Such a distance can well explain why single residues (too small) and tetrapeptides (too large) are not recognized by hPepT2, and suggests that, as a trend, dipeptides should bind the transporter assuming extended conformations, while tripeptides have to assume more folded geometries.

Figure 2 and Table S2 (Supplementary data) show the major residues involved in ligand binding and allow some relevant considerations to be drawn. Firstly, the binding cavity is completely lined by apolar residues including both alkyl and aryl side chains and this can clearly explain why hPepT2 prefers highly hydrophobic substrates. Secondly, the ligand peptide bond is never involved in pivotal contacts with hPepT2 residues and indeed hPept2 can recognize derivatives without central amide functions. Finally,

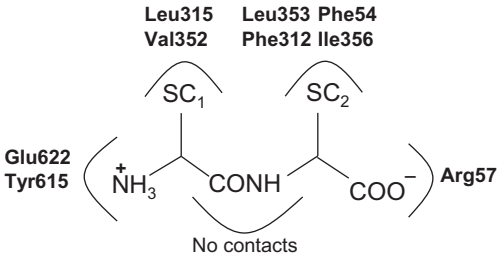

Figure 2. Bi-dimensional scheme illustrating the key residues involved in ligand recognition (SC = side chain).

the amino group normally interacts with Glu622, Tyr615 and Thr619, while the carboxyl terminus interacts constantly with Arg57, an ion-pair shared by all docked compounds.

A more careful analysis of the putative complexes (as summarized by Table S2) allows the recognition of a set of possible binding modes for dipeptides, depending on polarity of their side chains. With two alkyl side chains, the dipeptide beneficially contacts a large set of aliphatic residues flanking the binding pocket (Leu315, Leu316, Leu325, Val352, Leu353, Ile356 and Val623). When the dipeptide bears aromatic moieties, it can elicit π–π interactions with many surrounding aromatic residues (Phe54, Phe58, Phe312, Trp313 and Trp323). Conversely, when the substrate includes polar side-chains, it cannot stabilize significant polar contacts since the binding pocket presents almost only hydrophobic residues. More in detail, a polar side-chain in first

Table 1
Affinity values, docking scores and in silico predictions (K_i values are expressed in mM and En_EL scores in kcal/mol)

Compound	K_i	pK_i	En_EL	Pred pK_i by Eq. 1	Pred K_i by HypoGen	Compound	K_i	pK_i	En_EL	Pred pK_i by Eq. 1	Pred K_i by HypoGen
Ala-Ala	0.0060	2.22	−95.24	1.74	0.05	Gly-Gly	0.0540	1.27	−98.84	1.88	1.6
Ala-Ala-Ala	0.0180	1.74	−93.49	1.68	0.039	Gly-His	0.0390	1.41	−99.55	1.90	1
Ala-Ala-Asp	0.1600	0.80	−97.27	1.82	0.052	Gly-His-Lys	3.1000	−0.49	−55.85	0.28	4.6
Ala-ATAA	0.1900	0.72	−64.03	0.59	0.1	Ile-Tyr	0.0080	2.10	−99.01	1.88	0.028
Ala-D-Ala	0.2700	0.57	−55.28	0.26	0.061	Ile-Val-Tyr	0.0140	1.85	−104.35	2.08	0.017
Ala-Gly	0.0070	2.15	−98.94	1.88	0.44	Leu-Gly-Gly	0.0180	1.74	−94.09	1.70	0.014
Ala-Lys	0.0230	1.64	−103.57	2.05	0.82	Leu-Pro	0.0100	2.00	−83.95	1.33	0.039
Ala-Nle	0.0310	1.51	−91.50	1.61	0.011	Levulinic acid	0.2300	0.64	−97.17	1.82	2.9
Ala-Pro	0.0120	1.92	−81.66	1.24	0.13	LysAc-Ala	0.0160	1.80	−98.77	1.88	0.45
Ala-Ser	0.0060	2.22	−95.86	1.77	0.54	Lys-Ala	0.0410	1.39	−97.60	1.83	0.045
Ala-Val-Leu	0.0120	1.92	−104.41	2.08	0.017	LysBz-Ala	0.0060	2.22	−97.33	1.82	0.52
Amoxicillin	0.4300	0.37	−101.98	1.99	0.35	LyzZNO$_2$-Ala	0.0007	3.15	−127.60	2.94	0.25
Ampicillin	1.3000	−0.11	−45.67	−0.09	0.12	Metampicillin	0.7300	0.14	−31.43	−0.62	2.8
Asp-Asp	0.1500	0.82	−70.78	0.84	1.8	Met-Met	0.0030	2.52	−92.40	1.64	0.0072
Ataa	0.5500	0.26	−50.47	0.08	3.2	Met-Met-Met	0.0020	2.70	−103.23	2.04	0.0016
Benzylpenicillin	11.0000	−1.04	−32.17	−0.59	2.2	Oxacillin	3.3000	−0.52	−28.61	−0.73	1.7
Cefaclor	0.0290	1.54	−103.04	2.03	0.12	Phe-Ala	0.0160	1.80	−95.30	1.75	0.087
Cefadroxil	0.0030	2.52	−103.88	2.06	0.044	Phe-ATAA	0.2400	0.62	−72.13	0.89	0.44
Cefalotin	8.3000	−0.92	−32.37	−0.59	2.1	Phe-Gly	0.0140	1.85	−107.94	2.22	0.011
Cefamandole	2.8000	−0.45	−41.54	−0.25	0.51	Phe-Phe	0.0370	1.43	−107.00	2.18	0.014
Cefapirin	11.1000	−1.05	−32.48	−0.58		Pro-Asp	2.1000	−0.32	−51.02	0.10	2.7
Cefepime	11.0000	−1.04	−46.79	−0.05	1.6	Pro-Glu	2.6000	−0.41	−16.09	−1.19	3.3
Cefixime	2.6000	−0.41	−37.74	−0.39	0.63	Pro-Gly-Gly	11.0000	−1.04	−27.43	−0.77	7.0
Cefmetazole	4.3000	−0.63	−34.37	−0.51	2	Pro-Phe-Lys	0.9000	0.05	−70.07	0.81	0.27
Cefodizime	9.0000	−0.95	−57.22	0.33	2.8	Ser-Ala	0.0070	2.15	−98.14	1.85	0.043
Cefotaxime	20.0000	−1.30	−9.24	−1.44	3.5	Trp-Ala	0.0040	2.40	−102.50	2.01	0.0032
Cefpodoxime	31.0000	−1.49	−2.18	−1.71	3.1	Trp-Gly-Tyr	0.0017	2.77	−90.76	1.58	0.009
Cefuroxime	12.6000	−1.10	−34.46	−0.51	1.3	Trp-Trp	0.0008	3.10	−100.70	1.95	0.0003
Cephalexin	0.0750	1.12	−88.37	1.49	0.034	Trp-Trp-Trp	0.0003	3.52	−121.41	2.71	0.0002
Cephaloridine	8.5000	−0.93	−17.64	−1.13	2.7	Tyr-Ala	0.0044	2.36	−107.00	2.18	0.0018
Ciclacillin	0.0440	1.36	−88.56	1.50	0.041	Tyr-Phe	0.0090	2.05	−103.44	2.05	0.0023
Cloxacillin	0.9500	0.02	−32.18	−0.59		Val-Ala-Leu	0.0090	2.05	−103.31	2.04	0.0018
Cys-Gly	0.0290	1.54	−97.02	1.81	0.44	Val-ATAA	0.1100	0.96	−68.75	0.76	0.27
D-Ala-Ala	0.1300	0.89	−97.33	1.82	0.043	Val-Phe	0.0038	2.42	−100.14	1.93	0.066
Dicloxacillin	0.4200	0.38	−32.58	−0.58	0.26	Val-Pro	0.0390	1.41	−81.68	1.24	0.16
Dmet-Met-Met	0.0060	2.22	−105.64	2.13	0.019	Val-Pro-Pro	0.0230	1.64	−97.39	1.82	0.94
Glu-Ala	0.0300	1.52	−71.69	0.87	0.45	Val-Tyr	0.0025	2.60	−100.3988	1.94	0.95
Gly-Ala	0.0035	2.46	−99.23	1.89	0.0011						

position (SC_1) can stabilizes H-bonds with Tyr615, while a polar side-chain in second position (SC_2) does not have suitable residues close enough to elicit polar interactions. The first two residues of the docked tripeptides interact with the same hPepT2 residues already described for dipeptides, while the third residue contacts mainly Phe54, when its side chain is aromatic or Ile356 and Val50 with alkyl side chains.

Although the described binding modes cover almost all docked peptides, some substrates deserve a specific description. As a rule, residues in D configuration weaken the precise network of polar interactions which involve both charged termini, thus justifying the marked stereospecificity of hPepT2. As already observed for hPepT1, the proline as first residue has a detrimental effect on ligand affinity probably because the intracyclic amino group is less accessible for the mentioned polar contacts. The remarkable affinity of LyzZNO2-Ala derivative can be explained considering that the phenyl ring generates π–π stacking with Phe312, while the nitro function realizes an H-bond with Ser626. This last residue is significantly far from the other key hPepT2 residues (distance Ser626-Tyr615 ≈ 15 Å) and thus it cannot interact with unmodified polar dipeptides. The notable affinity of Nε-modified lysine peptides indicates that this lateral pocket can accommodate quite large and polar moieties and suggests the opportunity to exploit such modified dipeptides as carriers to transport polar molecules.[25] Finally, substrates with ionized side-chains show in general modest affinity values because they cannot stabilize additional polar interactions but detrimentally interfere with the pivotal contacts elicited by the charge termini.

The putative complexes for the β-lactams appear more heterogeneous and mainly differ in their capability to interact with Tyr615 and Glu622. Indeed, all β-lactams possess a carboxyl group able to contact Arg57, but not all derivatives have an amino group which can conveniently mimic the peptidic N-terminus. Although both charged termini are not mandatory for ligand affinity, it is easily understandable that the compounds lacking in a suitable amino group show dramatically poorer affinity values, since they can at least elicit H-bonds with Tyr615 and/or Glu622, but loose the key ion-pair with Glu622. This lack of a protonated function clearly impacts on computed docking scores and, indeed, the interaction energy average as computed for peptides is markedly lower than that derived for β-lactams (−86.82 kcal/mol vs −40.37 kcal/mol).

Exemplarily, Figure 3 depicts the putative complex for Cephamandole showing that Glu622 and Tyr615 stabilize an H-bond with the hydroxy function and Tyr615 also elicits π–π stacking with the ligand phenyl ring. Like all docked β-lactams, this substrate conserves the key ionic contacts between Arg57 and carboxyl group. The cephalosporanic ring contacts a rich set of apolar residues (Val50, Val51, Phe54, Phe312 and Leu353), while the tetrazole ring stabilizes hydrophobic interactions with Ile46, Ile49 and Val50 plus an H-bond with Asn192. This polar contact characterizes the cephalosporins which bear bulky and polar groups in 3 on cephalosporanic ring. Finally, the cephalosporins bearing positively charged rings (e.g., cefepime) can stabilize ionic interactions with Glu53 which normally interacts with Arg57.

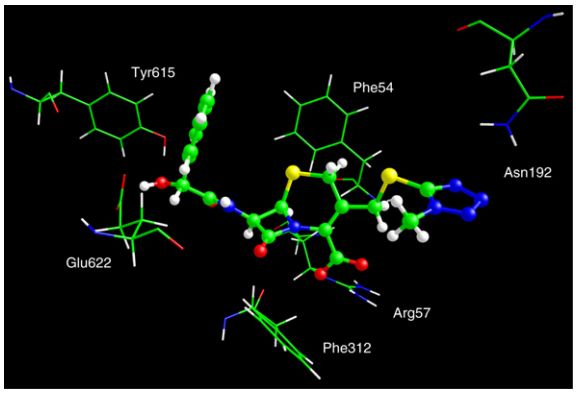

Figure 3. Main interactions stabilizing the putative complex for Cephamandole. One can observe the pivotal ion-pair with Arg57, while Tyr615 and Glu622 elicit H-bonds with the hydroxy function.

With a view to developing predictive relationships for the hPepT2 affinity values, many scoring functions as computed by FRED and VEGA programs were tested along with a set of ligand-based descriptors (e.g., accessible surface, polar surface, apolar surface, volume, molecular weight and number of rotors). Eq. 1 reports the truly remarkable relation between affinity values (expressed as pK_i, mM) and the electrostatic interaction energies (En_EL, as computed by VEGA using a distance dependent dielectric constant). The robustness of this equation is confirmed by its statistics and the fact that it includes only one independent variable.

$$pK_i = -0.037(\pm 0.0021)En\_EL - 1.787(\pm 0.18) \quad (1)$$

$$n = 75;\ r^2 = 0.80;\ q^2 = 0.80;\ SE = 0.45;\ F = 293.16;\ p < 0.0001$$

To further assess the predictive power of Eq. 1, the dataset was randomly subdivided in training ($n = 50$) and test ($n = 25$) sets. Eq. 2 reports the correlation between affinity values and energy scores as computed for the training set only. The reliability of Eq. 1 is further confirmed by the fact that Eq. 1 and Eq. 2 have quite identical slopes and intercepts suggesting that the correlation between affinities and energy scores is stable and reasonably independent on the number of included ligands. The relation between experimental and predicted pK_i values (Eq. 3), as computed by Eq. 2 for the external test set, affords an encouraging validation for the predictive power of Eq. 1. Eq. 3 is indeed a satisfactory one given its statistics, its slope close to 45° ($\alpha = 47.1°$) and the intercept close to zero:

$$pK_i = -0.035(\pm 0.0025)\ En\_EL - 1.785(\pm 0.20) \quad (2)$$

$$n = 50;\ r^2 = 0.80;\ q^2 = 0.80;\ SE = 0.43;\ F = 194.81;\ p < 0.0001$$

$$pK_{iexp} = 1.077\ pK_{ipred} - 0.273 \quad (3)$$

$$n = 25;\ r^2 = 0.83;\ SE = 0.53;\ F = 119.02;\ p < 0.0001$$

A more in-depth analysis of the pK_i values, as predicted by Eq. 1 for the entire dataset, reveals that for only six substrates the difference between experimental and predicted pK_i values is greater than 1, whereas for 51 substrates out of 75 the difference is below 0.5. Among the six poorly predicted ligands, there are four β-lactams and the levulinic acid thus to suggest that Eq. 1 is principally optimized to predict the affinity of peptidic ligands. As a trend, also the marked affinity of highly apolar substrates is not always suitably predicted and this can be explained considering that the electrostatic score cannot properly account for the hydrophobic contacts. However, it should be noted that the introduction of lipophilicity descriptors does not significantly improve the goodness of Eq. 1. The analysis of all residuals shows that there is no correlation with the pK_i values suggesting that Eq. 1 is equally predictive for good and poor hPepT2 substrates.

2.3. Pharmacophore mapping

With a view to confirming the docking results, enriching the knowledge about the hPepT2 binding modes, the HypoRefine

module within Catalyst program package was exploited to derive an automated SAR pharmacophore model for hPepT2 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 0.0004 to 31 mM). The 10 hypotheses, which showed the best correlation between predicted and experimental affinity values, are compiled in Table S3 (Supplementary data) 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 which depend on weight cost, error cost and configuration cost, as computed by the HypoGen module during hypothesis generation. The cost difference should be greater than 60 to represent a true correlation data.[26]

The top-ranked pharmacophore model (Hypo1, Fig. 4A) had the best predictive power and statistical significance and was characterized by the highest cost difference (74.864), the lowest RMS (1.174), and the best correlation coefficient (0.922). These values emphasize the predictive power of the 3D-QSAR pharmacophore as evidenced by the predicted pK_i values for the entire dataset (Table 1), and suggest that it is very unlikely that it comes about by chance. The correlation between experimental and predicted affinities is good ($r^2 = 0.64$) and, even if fairly poorer than that obtained by docking simulations, it confirms the reliability and predictive power of the proposed pharmacophore which is in line with previously published models.[27]

Figure 4A shows that the selected 3D hypothesis consists of two hydrophobic region (H1–H2), one hydrogen bond acceptors (A), one hydrogen bond donor (D), and nine excluded volume sites (E1–E9) in a specific three-dimensional orientation. Furthermore, with the aim to compare the pharmacophore model with docking results, all ligands, in their bound conformation, were mapped onto the pharmacophoric model (as shown in Fig. 4B), whose key elements confirm the precise role of each hPepT2 residue.

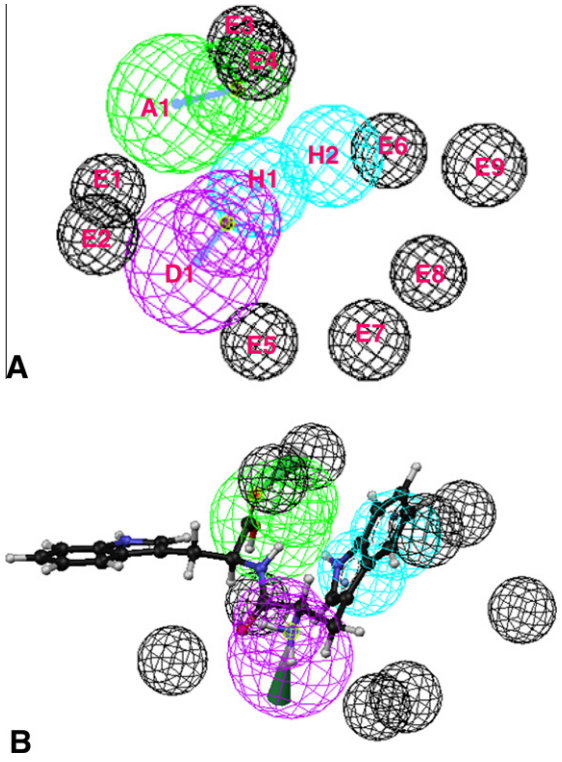

Figure 4. Pharmacophore model for hPepT1 ligands. (A) Pharmacophore features as derived by HypoRefine approach (legends as follows: hydrogen bond acceptor, green A1–A2; hydrogen bond donor, purple D1–D2; hydrophobic region, azure H1–H2; excluded volumes, black E1–E9). (B) Pharmacophore model mapped on docking pose of Trp-Trp dipeptide.

Thus, Figure 4B depicts the pharmacophoric regions mapped on docking pose of most affinitive dipeptide Trp-Trp. It is worth observing that ammonium head occupies the H-bond donor region which overlaps Glu622, Tyr615, the carboxylate maps the H-bond acceptor regions corresponding to Arg57, and the two hydrophobic regions correspond to C-terminal side chain, which can contact a large set of hydrophobic residues. Finally, the excluded volumes represent the position of other amino acids not involved in ligand recognition but able to create steric clashes with the less affinitive compounds. Interestingly, such excluded volumes fully surround the binding cavity thus explaining why tetrapeptides are not recognized by hPepT2.

2.4. hPepT1/hPepT2. selectivity

The comparison of here described results with those reported by the previous study for hPepT1[21] allows the major differences between such peptide transporters to be clarified. The first key difference concerns the interactions stabilized by charged termini. Indeed, while the contacts elicited by N-terminus are very similar in the two transporters (an H-bond with a conserved tyrosine residue and salt bridges with negatively charged residues), the interactions stabilized by ligand's carboxyl group are markedly different, since it realizes only weak H-bonds with a set of backbone atoms in hPepT1, whereas it stabilizes a pivotal ion-pair with Arg57 in hPepT2 as also confirmed by mutational analyses.[18] Such a difference can explain the greater heterogeneity observed for the docked complexes for hPepT1, since the involvement of the sole ligand's amino terminus in strong interactions permits a marked variability in the accommodation of the carboxyl groups. Finally, the involvement of both charged termini in key salt bridges affords more stable complexes with hPepT2 as demonstrated by the observation that the here exploited electrostatic interaction score affords an average value for hPepT2 vastly lower than for hPepT1 by considering the same set of ligands (−74.05 kcal/mol vs −43.56 kcal/mol, as taken from previous study). This finding can explain why hPepT1 is a low-affinity and high-capacity transporter, whereas hPepT2 is a high-affinity and low-capacity transporter for the same substrates.[15]

A second significant difference concerns the apolarity of the binding site. Both hPepT1 and hPepT2 possess a vastly apolar cavity thus preferring hydrophobic substrates. Nonetheless, the pocket of hPepT1 includes some important polar residues which can conveniently contact polar substrates, while hPepT2 has no polar residues in its cavity, and peptides with polar side-chains may detrimentally hamper the pivotal ionic contacts stabilized by the charged termini. Such a difference can explain why highly hydrophobic ligands (e.g., Trp-Trp-Trp or Lys(Z)-Lys(Z)) show an exceptional selectivity for hPepT2, while peptides containing charged residues (especially in the last position) show higher binding affinities for hPepT1.

A third relevant difference involves the contacts stabilized by the central peptide bonds. Indeed, while amide functions are often involved in H-bonds in the hPepT1 cavity, they are never involved in clear contacts in hPepT2, but probably they favor a ligand conformation in which the charged termini are distant enough to stabilize the key ionic interactions. As previously discussed, whereas hPepT1 permits a certain degree of ligand flexibility since only the ligand N-terminus elicits strong contacts, the implication of both termini in hPepT2 constrains the ligand to assume a well-defined conformation to contact both key hPepT2 residues. The conformational factors are well illustrated, for example, by the difference between the dipeptide Gly-Gly ($K_i = 0.054$ mM) and the 5-amino pentanoic acid ($K_i = 23$ mM, ligand not included in

dataset) for which a strong intramolecolar ion-pair may prevent a suitable extended geometry.

These considerations are further confirmed by the comparison between the pharmacophore model here developed for hPepT2 and that previously proposed for hPepT1. The main difference involves the H-bond acceptor regions which are more defined in hPepT2 constantly corresponding to carboxylate, while in hPepT1 they were larger and involved also the ligand peptide bonds which conversely are never engaged by hPepT2 interactions. Again, the hydrophobic regions are more extended in hPepT2 and can encompass both side chains thus to emphasize the more beneficial role of apolar contacts in hPepT2 recognition. Finally, the excluded regions are more homogeneously arranged around the bound substrate in hPepT2, whereas they appear mainly focused on C-terminus in hPepT1. This may indicate that hPepT1 can recognize substrates with N-terminal side chain largely modified whereas hPepT2 prefers less bulky substrates even if it can bind, for instance, Nε-modified lysine containing dipeptides.

3. Conclusions

The study describes the generation of a reliable homology model for the human transporter hPepT2 which is consistent with structural and functional data available in literature. The agreement between docking simulations and pharmacophoric mapping affords a convincing validation for the here proposed model and its binding modes and allows the main differences between hPepT1 and hPepT2 to be revealed. The exploited combined approach (i.e., docking simulations plus pharmacophore mapping) can find fertile applications to predict the affinity of potential hPepT2 substrates and to screen peptidomimetic libraries unveiling which molecules should be recognized by hPepT2. More intriguingly, this model can be exploited to design compounds in which a not transported molecule is combined with a suitable peptidomimetic moiety in order to obtain a prodrug carrier which is recognized by hPepT2.

It should be noted that the affinity of a molecule to a transporter is not a safe indicator for the actual transport since some compounds (e.g., Trp-Trp-Trp tripeptide and some sartans) show remarkable hPepT2 affinities but are too large to be transported and behave as competitive inhibitors.[28] Nonetheless, the affinity is a prerequisite for the transport and the ability to conveniently predict in silico the affinity values for novel compounds should allow molecules with very modest computed affinities to be excluded from further experimental testing.

4. Computational details

4.1. Generation of the hPepT2 model

The amino acidic sequence of the human PepT2 transporter was retrieved from Swiss-Prot database (entry code Q16348, S15A2_HUMAN). The hPepT2 model was generated by fragments with a strategy involving the following steps: (1) the subdivision of the amino acidic sequence in 25 fragments as predicted by TMPRED[29] (namely 12 TM segments, TM1-12, 6 extracellular loops, EL1-6, 5 cytoplasmic loops, CL1-5, and 2 terminal segments, NT and CT), (2) the homology modeling of these segments separately by LOMETS[30] (http://zhang.lab.ccmb.med.umich.edu/LOMETS/), a local meta-threading server for protein structure prediction, and (3) the assembly of fragments using the structure of lactose permease of *E. coli* as a final template (LacY, PDB Id: 1PV6) the use of which is justified by the known homology (similarity = 54.6%) with the hPepT2 as illustrated by the pair-wise alignment (Fig. S1, Supplementary data).

For each segment, LOMETS was able to produce several realistic models and the best structure has been chosen considering what result better fulfilled the following major conditions: (a) the predicted secondary structure from the sequence alignment; (b) the lack of not predicted gaps; (c) the prediction score (ZSCORE) as calculated by LOMETS program; (d) the helix conformation of 12 transmembrane segments (TM1-TM12) with characteristic slight bend of helices containing proline and glycine residues; (e) the global 'U' shape for the loops in which the two ends are close enough to join to TM segments.

Table S4 (Supplementary data) reports the templates used by LOMETS to generate the best model for each fragment. While avoiding a systematic description of the used templates, Table S4 allows some relevant considerations to be drawn: (1) the bacterial lactose permease is never used as template indicating that the reported homology concerns the global topology rather then the local folding; (2) all TM segments are conceivably predicted using templates which possess exclusive (e.g., plantaricin, catenin) or largely predominant helical folding (3) a large number of TM segments is modelled using templates characterized by coiled-coils structures; this may indicate that such helical motifs often play key roles in molecular recognition processes and, indeed, the so modelled TM segments are involved in substrate recognition (TM1, TM7, TM8, TM9); (3) more external TM segments are modelled using proteins which are characterized by highly hydrophobic surfaces (e.g., AKAP proteins); (4) the templates used for the loops are more heterogeneous including enzymes, receptors and regulatory peptides; (5) the longest segment (EL5) is predicted using Trm8, a yeast tRNA m7G methylation complex, which shows a mixed αβ topology. Interestingly, some used templates for TM segments were already exploited in our previous studies suggesting that they can be seen as general templates for transmembrane helical segments.

The assembly of predicted fragments was performed superimposing the backbone of a fragment with that of the correspondent segment in the structure of lactose permease and manually connecting the adjacent segments using VEGA software. In detail, the superimposition involved the Cα atoms of transmembrane helices only, since 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 MD simulations (while the transmembrane bundle remains constrained during the molecular dynamics, as later described).

With the backbone completed, side chains and hydrogen atoms were added using VEGA.[31] According physiological pH, Arg, Lys, Glu and Asp residues were preserved ionized, while His residues were considered neutral by default. After a careful scrutiny of obtained structure to avoid unphysical conditions, the hPepT2 model underwent an initial minimization until RMS gradient was equal to 1 to discard high-energy interactions, followed by a local minimization until RMS = 0.05, where all atoms were kept fixed except for atoms included within a 5.0 Å sphere around the manually connected bonds (at the fragment ends). Finally the model was optimized by a final minimization made up by 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.

In order to gain a better relaxation and a more correct arrangement of the whole hPepT2 model a molecular dynamics 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/10 iterations), (2) starting equilibration of 2500 ps, where the transmembrane backbone was kept fixed, and (3) equilibration of 7500 ps, in which the transmembrane backbone was harmonically restrained with decreasing harmonic force constants. More in details, harmonic force constant value was equal to 1

(1000 kJ mol^{-1} nm^{-2}) at the beginning of simulation and then was divided into two every 1.5 ns (then 5 MD were performed with 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 also with 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 stereochemical quality of the model was evaluated using PROCHECK[32] and MolProbity,[33] and the obtained results are collected in Table S5.

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 cut-off of 10 Å and the pair list was updated every 20 iterations; Newton's equation was integrated, using r-RESPA method, every 4 fs for long-range electrostatic forces, 2 fs for short-range non bonded forces, and 1 fs for bonded forces; a frame was stored every 5 ps, yielding 2000 frames. All calculations were carried out on an 8-node Tyan-VX50 system. All minimizations were performed using the conjugated gradients algorithm. The package Namd 2.51[34] was used with the force-field CHARMm v22 and Gasteiger's atomic charges.

4.2. Ligands set-up and docking analyses

The set of 75 known PepT2 ligands (as compiled in Table 1) was compiled from literature data[15,35] and can be structurally subdivided in peptides, peptidomimetics and β-lactams. The affinities of such ligands as measured in SKTP cells (Table 1) range from 0.0001–31 mM.

The ligands were modelled in their ionised forms since they may be involved in ligand recognition. The ligand structure was built using VEGA software, and the overall geometry and the atomic charges were optimized using MOPAC6.0. Their conformational profile was explored by a MonteCarlo 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 ones; in detail, two geometries were considered as non-redundant if they differed by more than 60 degrees 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 15 Å side box around Arg57 whose key role in substrate recognition is confirmed by the analysis of the single nucleotide polymorphisms.[18] The obtained complexes were then refined focusing the minimization on the atoms inside a 10 Å sphere around the bound ligand. The minimized complexes were finally used to recalculate FRED docking scores as well as the interaction energies as computed by VEGA. In the predictive analyses, the goodness of the computed relationships was assessed by considering the correlation coefficient (r^2), leave-one-out cross-validated r^2 (q^2), standard error (SE), Fisher test (F) and significance value (p). In details, Eq. 1 was generated iteratively by maximizing the r^2 value and its predictive power was evaluated by subdividing in training ($n = 50$) and test ($n = 25$) sets. Though randomly subdivided, the Tanimoto comparison of the chemical fingerprints for substrates included in the two sets indicates that they have a similar average variability (0.26 vs 0.28) suggesting that they equally cover the chemical space of PepT2 substrates.

4.3. Pharmacophore generation

All structures were generated using the 2D/3D editor sketcher in the Catalyst 4.10 software package (Catalyst, version 4.10; Accelrys Inc., San Diego, CA, 2006) 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/receptor interactions. These are hydrogen bond donors, hydrogen bond acceptors, aromatic ring, hydrophobic (aliphatic or aromatic) groups, and positively and negative ionisable groups. The interfeature spacing penalty was reduced 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. Specifically, we select compounds with different degree of activity, spanned five orders of magnitude, making this a good data set for HypoRefine module. The uncertainty value of compounds 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.

Acknowledgment

Financial support for this research by MiUR is gratefully acknowledged.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.06.027.

References and notes

1. Wang, J. *Curr. Pharm. Des.* **2009**, *15*, 2195.
2. Vistoli, G.; Pedretti, A.; Testa, B. *Drug Discovery Today* **2008**, *13*, 285.
3. Shitara, Y.; Horie, T.; Sugiyama, Y. *Eur. J. Pharm. Sci.* **2006**, *27*, 425.
4. Oostendorp, R. L.; Beijnen, J. H.; Schellens, J. *Cancer Treat. Rev.* **2009**, *35*, 137.
5. Ekins, S.; Ecker, G. F.; Chiba, P.; Swaan, P. W. *Xenobiotica* **2007**, *37*, 1152.
6. Winiwarter, S.; Hilgendorf, C. *Curr. Opin. Drug Discov Devel.* **2008**, *11*, 95.
7. Chang, C.; Swaan, P. W. *Eur. J. Pharm. Sci.* **2006**, *27*, 411.
8. Terada, T.; Inui, K. *Curr. Drug Metab.* **2004**, *5*, 85.
9. Daniel, H.; Kottra, G. *Pflugers Arch.* **2004**, *447*, 610.
10. Brandsch, M.; Knütter, I.; Bosse-Doenecke, E. *J. Pharm. Pharmacol.* **2008**, *60*, 543.
11. Terada, T.; Inui, K. *Biochem. Pharmacol.* **2007**, *73*, 440.
12. Kamal, M. A.; Keep, R. F.; Smith, D. E. *Drug Metab. Pharmacokinet.* **2008**, *23*, 236.
13. Kamal, M. A.; Jiang, H.; Hu, Y.; Keep, R. F.; Smith, D. E. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2009**, *296*, R986.
14. Groneberg, D. A.; Nickolaus, M.; Springer, J.; Döring, F.; Daniel, H.; Fischer, A. *Am. J. Pathol.* **2001**, *158*, 707.
15. Biegel, A.; Knütter, I.; Hartrodt, B.; Gebauer, S.; Theis, S.; Luckner, P.; Kottra, G.; Rastetter, M.; Zebisch, K.; Thondorf, I.; Daniel, H.; Neubert, K.; Brandsch, M. *Amino Acids* **2006**, *31*, 137.
16. Biegel, A.; Gebauer, S.; Brandsch, M.; Neubert, K.; Thondorf, I. *J. Med. Chem.* **2006**, *49*, 4286.
17. Meredith, D.; Price, R. A. *J. Membr. Biol.* **2006**, *213*, 79.
18. Terada, T.; Irie, M.; Okuda, M.; Inui, K. *Biochem. Biophys. Res. Commun.* **2004**, *316*, 416.
19. Abramson, J.; Smirnova, I.; Kasho, V.; Verner, G.; Kaback, H. R.; Iwata, S. *Science* **2003**, *301*, 610.
20. Newstead, S.; Drew, D.; Cameron, A. D.; Postis, V. L.; Xia, X.; Fowler, P. W.; Ingram, J. C.; Carpenter, E. P.; Sansom, M. S.; McPherson, M. J.; Baldwin, S. A.; Iwata, S. *EMBO J.* **2011**, *30*, 417.
21. Pedretti, A.; De Luca, L.; Marconi, C.; Negrisoli, G.; Aldini, G.; Vistoli, G. *ChemMedChem* **2008**, *3*, 1913.
22. Pedretti, A.; Villa, M.; Pallavicini, M.; Valoti, E.; Vistoli, G. *J. Med. Chem.* **2006**, *49*, 3077.
23. Kasho, V. N.; Smirnova, I. N.; Kaback, H. R. *J. Mol. Biol.* **2006**, *358*, 1060.
24. Smith, T. F. *Subcell. Biochem.* **2008**, *48*, 20.
25. Zimmermann, M.; Stan, A. C. *J. Neurosurg.* **2009**, *112*, 1005.

26. Sakkiah, S.; Thangapandian, S.; John, S.; Kwon, Y. J.; Lee, K. W. *Eur. J. Med. Chem.* **2010**, *45*, 2132.
27. Swaan, P. W.; Bensman, T.; Bahadduri, P. M.; Hall, M. W.; Sarkar, A.; Bao, S.; Khantwal, C. M.; Ekins, S.; Knoell, D. L. *Am. J. Respir. Cell Mol. Biol.* **2008**, *39*, 536.
28. Brandsch, M. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5*, 887.
29. Hofmann, K.; Stoffel, W. *Biol. Chem. Hoppe Seyler* **1993**, *374*, 166.
30. Wu, S.; Zhang, Y. *Nucleic Acids Res.* **2007**, *35*, 3375.
31. Pedretti, A.; Villa, L.; Vistoli, G. *J. Mol. Graph.* **2002**, *21*, 47.
32. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S. *J. Appl. Crystallogr.* **1993**, *26*, 283.
33. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12.
34. Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comp. Phys.* **1999**, *151*, 283.
35. Luckner, P.; Brandsch, M. *Eur. J. Pharm. Biopharm.* **2005**, *59*, 17.