Affinity prediction for substrates of the peptide transporter PepT1†

Patrick D. Bailey,*[a] C. A. Richard Boyd,[b] Ian D. Collier,‡[c] John P. George,[c] George L. Kellett,[d] David Meredith,[b] Keith M. Morgan,[c] Rachel Pettecrew[a] and Richard A. Price[a]

Received (in Cambridge, UK) 25th August 2005, Accepted 27th October 2005
First published as an Advance Article on the web 30th November 2005
DOI: 10.1039/b511996k

A quantitative method has been developed for determining the affinity of substrates for the peptide transporter PepT1, allowing oral availability of drugs *via* PepT1 to be estimated.

PepT1 is an essential protein expressed in the upper intestine of all mammals, and it facilitates the active uptake into the bloodstream of di- and tripeptides derived from food.[1] Of particular importance medicinally is the ability of PepT1 to transport certain hydrophilic drugs into the circulatory system, facilitating the oral bioavailability of pharmaceuticals that might otherwise need to be administered intravenously. Examples of such drugs include the penicillin and cephalosporin antibiotics, and the antihypertensive drugs captopril and enalapril,[2] whilst the transport of the non-peptidic drugs acyclovir and L-dopa have been improved by converting them into hydrolysable derivatives that are recognised by PepT1.[3,4]

Although PepT1 is able to transport an amazingly diverse range of substrates, there are clearly key structural features necessary for rapid transport (Fig. 1a). Thus, di- and tripeptides are generally transported efficiently, whilst amino acids and tetrapeptides (or larger peptides) are not; many β-lactam antibiotics are rapidly transported, but β-amino acid dipeptides are poor substrates; alanyl anilides have high affinity, whilst dipeptides of the stereochemistry L-Xaa-D-Xaa do not.

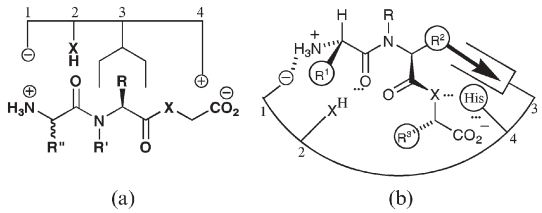


Fig. 1 (a) '2D' features of PepT1; (b) PepT1 substrate template.

[a]*School of Chemistry, Faraday Building, The University of Manchester, Sackville Street, Manchester, UK M60 1QD. E-mail: pat.bailey@manchester.ac.uk; Fax: +44 (0)161 306 4541; Tel: +44 (0)161 306 4448*
[b]*Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford, UK OX1 3QX. E-mail: richard.boyd@anat.ox.ac.uk; Fax: +44 (0)1865 272 420; Tel: +44 (0)1865 272 169*
[c]*School of EPS – Chemistry, Perkin Building, Heriot-Watt University, Edinburgh, UK EH14 4AS. E-mail: k.m.morgan@hw.ac.uk; Fax: +44 (0)131 451 3180; Tel: +44 (0)131 451 8020*
[d]*Department of Biology, University of York, PO Box 373, York, UK YO10 5YW. E-mail: glk1@york.ac.uk; Fax: +44 (0)1904 328 505; Tel: +44 (0)1904 328 500*
† Electronic supplementary information (ESI) available: Data input for Table 1. See DOI: 10.1039/b511996k
‡ In memory of Ian Collier (1960–2002).

In 2000, we published a template model for predicting whether substrates would have low, medium or high affinity for PepT1 (Fig. 1b).[5] Since then, Thondorf's group have used computer modelling to refine the analysis for dipeptides[6a] and tripeptides/β-lactams,[6b] but the range of substrates was limited, and access to the parametized software is required for use by others. Herein we propose a simple, quantitative predictor for the binding of substrates to PepT1, based on our template model.

Our analysis was carried out in 3 stages.

(a) Defining the configuration of the first peptide bond, and the conformation of the first residue, based on recent studies.[7]

(b) Examination of the K_i/K_m (defined as K) values for a range of substrates (Table 1), with particular attention to those with unexpectedly high or low binding (see Fig. 2).

(c) Using the template model in Fig. 1b, parameters were associated with each recognition feature, and these were optimised empirically to achieve a good linear correlation against the experimental log K values.[8]

Each of the factors that contribute to the binding are summarised below, with their parameter range in brackets.[8]

(1) An N-terminal $-NH_3^+$ group ($+2 \rightarrow -2$).
(2) Stereochemistry at C_α of R^1 ($+1 \rightarrow -1$).[7a]
(3) Planar backbone from N-terminal C_α to R^2 ($+1 \rightarrow 0$).[7b]
(4) A hydrogen bond to the first peptide C=O ($+1 \rightarrow 0$).
(5) No alkylation of N^2-amide NH ($0 \rightarrow -1$).
(6) Stereochemistry at C_α of R^2 ($+2 \rightarrow -2$).
(7) Hydrophobic pocket for R^2, possessing a strong directional vector as indicated ($+2$ per aryl group $\rightarrow -1$).
(8) A carboxylate binding site ($+2 \rightarrow 0$).
(9) For larger substrates: (a) space for side-chain R^3 ($0 \rightarrow -2$).
(b) an alternative (tripeptide) carboxylate binding site ($+2 \rightarrow 0$).
(10) Size: $M_r < 300$ (-1), $M_r > 300$ (-2).

The aggregate of these terms generated a total binding parameter T (see Table 2 for examples); scaling this by a factor of 2.8 gave calculated binding energies in kJ mol^{-1}, which are plotted against log K in Fig. 3.

The model effectively identifies four key binding sites (factors 1, 4, 7, and 8/9b), whilst the 3D layout of the template is defined by the other stereochemical features. It is the aggregate effect of all of these factors that governs whether substrates have high, medium, or low affinity for PepT1.

The linear correlation in Fig. 3 indicates that our model allows the prediction of ΔG within ± 4 kJ mol^{-1}; given the standard errors in determining K experimentally, our model generally predicts binding to within ± 2 kJ mol^{-1} (*i.e.* K within a factor of 3), and is applicable across a wide range of tissue and cell systems. It is noteworthy that we used a very diverse range of substrates in order to thoroughly explore the structural requirements for

Table 1 Calculation of the binding paramter (T) for 50 potential substrates of PepT1 with structural diversity, taken from a wide range of tissue types (see refs)

Entry	Substrate	K/mM	log$K$	ΔG/kJ mol^{-1}	T	Est. ΔG	Ref.
1	L-Dopa–Phe	0.03	−1.52	26.0	9	25.2	4
2	Ala–NH–C_6H_4–(4Ph)	0.03	−1.52	26.0	8	22.4	10
3	Gly-Ala	0.032	−1.49	25.8	8	22.4	11*a*
4	Gly-Val	0.032	−1.49	25.8	8	22.4	12
5	Ala-Ala	0.08	−1.10	23.5	8	22.4	12
6	Phe-Tyr	0.1	−1.00	23.0	9	25.2	13
7	Asp-Ala	0.12	−0.92	22.5	8	22.4	11*a*
8	Ala-Ala-Ala	0.16	−0.80	21.8	8	22.4	13
9	Ser-Ala	0.21	−0.68	21.1	8	22.4	11*a*
10	Phe-Ala	0.21	−0.68	21.1	8	22.4	11*a*
11	Arg-Ala	0.22	−0.66	21.0	8	22.4	11*a*
12	Phe-Pro	0.23	−0.64	20.9	7	19.6	11*a*
13	Ala-Pro	0.25	−0.60	20.7	7	19.6	11*a*
14	Ala-(*trans*-thio)-Pro	0.3	−0.52	20.2	6	16.8	14
15	Ala–NH–C_6H_4–(4Me)	0.34	−0.47	19.9	6	16.8	10
16	$H_2NCH_2COC_2H_4CO_2H$	0.4	−0.40	19.5	7	19.6	19
17	Val-Lys	0.64	−0.19	18.3	7	19.6	15
18	L-Loracarbef	0.7	−0.15	18.1	7	19.6	16
19	Val–Acyclovir	0.74	−0.13	18.0	7	19.6	17
20	Phe-Tyr–NH_2	0.9	−0.05	17.5	7	19.6	13
21	Ala-Ala-D-Ala	0.99	0.00	17.3	6	16.8	13
22	Enalapril	1.1	0.04	17.0	5	14.0	18
23	H_2N–$(CH_2)_4$–CO_2H	1.14	0.06	16.9	6	16.8	19
24	D-Phe-Ala	1.14	0.06	16.9	6	16.8	20
25	Pro-Ala	1.26	0.10	16.7	6	16.8	11*a*
26	D-Phe-Gly	1.7	0.23	15.9	6	16.8	11*a*
27	D-Loracarbef	1.8	0.26	15.8	5	14.0	16
28	Ac–Phe	2	0.30	15.5	5	14.0	13
29	D-Amoxycillin	2	0.30	15.5	5	14.0	16
30	D-Phe-Glu	2.15	0.33	15.3	5	14.0	20
31	Ala–NH–Ph	2.9	0.46	14.6	5	14.0	10
32	D-Ala-Ala-Ala	3.04	0.48	14.5	6	16.8	13
33	4-$(H_2NCH_2)C_6H_4CO_2H$	3.1	0.49	14.4	5	14.0	21
34	Captopril	4	0.60	13.8	4	11.2	18
35	3-H_2N–$C_6H_4CH_2CO_2H$	6	0.78	12.8	4	11.2	11*b*
36	Ala-D-Ala-Ala	6.43	0.81	12.6	4	11.2	13
37	4-H_2N–$C_6H_4CH_2CO_2H$	6.5	0.81	12.6	4	11.2	22
38	Ac–Phe–Tyr	8.4	0.92	11.9	5	14.0	13
39	Ac–Phe-Tyr–NH_2	10	1.00	11.5	3	8.4	13
40	4-H_2N–$C_6H_4CO_2H$	10.6	1.03	11.3	4	11.2	11*b*
41	Ala–NH–CH_2Ph	14.1	1.15	10.6	4	11.2	10
42	D-Phe-L-Pro	21	1.32	9.6	5	8.4	11*a*
43	Ac–Phe–NH_2	22	1.34	9.5	3	8.4	13
44	Ala-D-Phe-Ala	22.7	1.36	9.4	3	8.4	23
45	Phe–NH_2	50	1.70	7.5	2	5.6	13
46	Phe	100	2.00	5.7	2	5.6	13
47	H_2N–$(CH_2)_3CO_2H$	>50	—	—	2	5.6	19
48	L-Dopa	>100	—	—	2	5.6	4
49	Acyclovir	>100	—	—	0	0	17
50	*Cyclo*(Gly-Gly)	>100	—	—	−1 (−2.8)		24

binding—*e.g.* 7 classes of compounds, but only a small number of peptides, to avoid biasing the analysis to natural substrates. We also attempted to use computer modelling to provide a quantitative prediction of binding, but this was less successful than our manual approach.[9a]

It is clear that our manual approach can only be approximate, but the surprisingly good correlation is probably helped by the nature of the binding of substrates to transporter proteins in general, and to PepT1 in particular. Substrates generally bind relatively weakly to transporters because rapid release is necessary after transport—*e.g.* high affinity substrates for PepT1 bind in the 0.1–1 mM range, whereas nM binding to enzymes is common. For

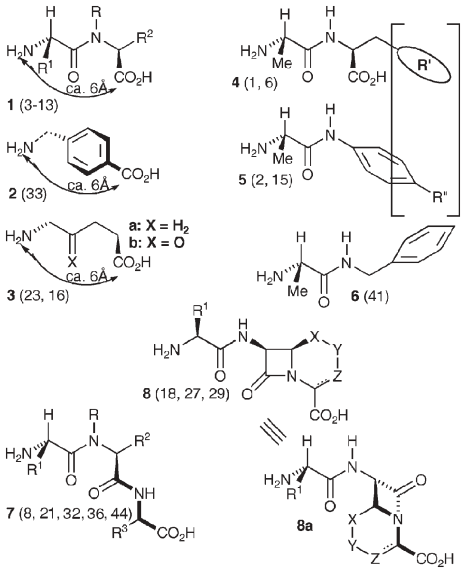


Fig. 2 Selected substrates that were used to determine the binding parameters. Numbers in brackets refer to entries in Table 1. Substrates **1–3** explore the required amino–carboxylate distance, and amide bond features; substrates **4–6** explore the hydrophobic pocket; substrates **7** and **8** explore tripeptides/β-lactams, with **8a** being the bound conformation for β-lactams.

Table 2 Examples of T calculation for PepT1 substrates

Entry	Substrate	1	2	3	4	5	6	7	8	9	10	T
5	Ala-Ala	2	1	1	1	0	2	0	2	—	−1	8
15	Ala–NH–C_6H_4(4Me)	2	1	1	1	0	0	2	0	—	−1	6
25	Pro-Ala	0	1	1	1	0	2	0	2	—	−1	6
35	3-H_2N–$C_6H_4CH_2CO_2H$	0	0	1	0	0	2	0	2	—	−1	4
45	Phe–NH_2	2	1	1	1	0	—	—	—	—	−1	4

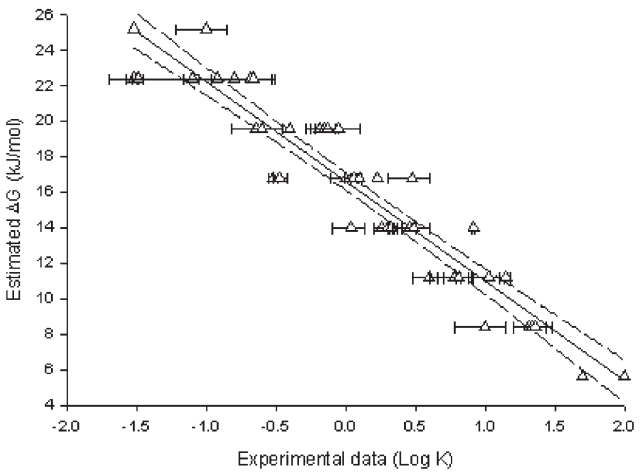


Fig. 3 Linear correlation of ΔG (y-axis, calc., kJ mol^{-1}) against log K.

PepT1, the structural diversity of di- and tripeptide substrates is accommodated by utilizing several low energy interactions, which means that the overall binding is only slightly affected by each factor; similarly, the calculated binding is relatively tolerant of inaccuracies in any one of the binding factors. It is also important to note that substrates with medium/high affinity are almost

always transported by PepT1,[9b] so medium/high affinity substrates are almost certain to be orally absorbed.

There are several interesting and important observations concerning the structures of the substrates, and the estimation of their binding to PepT1, and three are highlighted here.

(a) Most substrates that can be mapped onto a di- or tripeptide skeleton, with free amino and carboxylic acid groups, will bind at least quite well; however, mismatched stereochemistry seriously diminishes this (*e.g.* compare entries 8, 21, 32, 36), whereas optimal use of the R^2 binding pocket can offset the lack of a carboxylic acid group (*e.g.* entries 2, 15).

(b) It is intriguing that β-lactams (entries 18, 27, 29) apparently have the 'wrong' stereochemistry at C_α of residue 2 (derived from D-amino acids), but the 4-membered ring allows them to be mapped onto the template structure of Fig. 1b by a 180° rotation (see structure **8a**), so that the configuration is now appropriate for binding and transport.

(c) There is evidence from work with PepT2 that substrates with large groups at R^1 would not be transported;[25] in contrast, we have attached extremely large groups at R^2 of dipeptide analogues (*e.g.* steroidal esters linked to Asp/Ser), and still observed transport—this may be important for drug delivery.[26]

In conclusion, this paper describes a simple quantitative analysis for the binding of substrates to PepT1, which is a good indicator of transport *via* PepT1, and this should be of value in the design of medicinally important compounds that have good oral bioavailablity *via* the PepT1 mechanism.

Notes and references

1 (*a*) H. Daniel, *Annu. Rev. Physiol.*, 2004, **66**, 361; (*b*) T. Terada and K. Inui, *Curr. Drug Metab.*, 2004, **5**, 85; (*c*) D. Meredith and C. A. R. Boyd, *Cell. Mol. Life Sci.*, 2000, **57**, 754.

2 G. L. Amidon and H. J. Lee, *Annu. Rev. Pharmacol. Toxicol.*, 1994, **34**, 321.

3 P. V. Balimane, I. Tamai, A. Guo, T. Nakanishi, H. Kitada, F. H. Leibach, A. Tsuji and P. J. Sinko, *Biochem. Biophys. Res. Commun.*, 1998, **250**, 246.

4 I. Tamai, T. Nakanishi, H. Nakahara, Y. Sai, V. Ganapathy, F. H. Leibach and A. Tsuji, *J. Pharm. Sci.*, 1998, **87**, 1542.

5 P. D. Bailey, C. A. R. Boyd, J. R. Bronk, I. D. Collier, D. Meredith, K. M. Morgan and C. S. Temple, *Angew. Chem., Int. Ed.*, 2000, **39**, 505.

6 (*a*) S. Gebauer, I. Knutter, B. Hartrodt, M. Brandsch, K. Neubert and I. Thondorf, *J. Med. Chem.*, 2003, **46**, 5725; (*b*) A. Biegel, S. Gebauer, B. Hartrodt, M. Brandsch, K. Neubert and I. Thondorf, *J. Med. Chem.*, 2005, **48**, 4410.

7 We have recently confirmed the stereochemistry of the first residue, and the geometry of the first amide bond, for PepT1 substrates—see: (*a*) P. D. Bailey, C. A. R. Boyd, I. D. Collier, J. G. George, G. L. Kellett, D. Meredith, K. M. Morgan, R. Pettecrew, R. A. Price and R. G. Pritchard, *Chem. Commun.*, 2005, 5352; (*b*) P. D. Bailey, C. A. R. Boyd, I. D. Collier, G. L. Kellett, D. Meredith, K. M. Morgan, R. Pettecrew and R. A. Price, *Org. Biomol. Chem.*, 2005, **3**, 4038.

8 The factors were assigned integer values, once the substrate was aligned as closely as possible with the template structure (Fig. 1b). The values were optimised purely empirically, and we found that coarse 'quantization' of factors gave good correlations with binding. The parameters assigned to each factor were as follows: factor 1, 1° amines (+2), 2° and/or aryl amines (0), others (−2); factor 2, template stereochemistry at α-C^1 (+1), planar (0), incorrect stereochemistry (−1); factor 3, planar from α-C^1 → side-chain of 2nd residue (+1); factor 4, any C=X (where X is heteroatom with lone pair) at the 1st peptide bond (+1); factor 5, alkylation of N^2-amide nitrogen (−1); factor 6, template stereochemistry at α-C^2 (+2), planar (0), incorrect stereochemistry (−2); factor 7, eacharyl ring in pocket (+2), each charge in pocket (−1); factor 8, carboxylate within H-bonding distance of histidine in template (+2); factor 9a, template stereochemistry at α-C^3 (0), planar (−1), incorrect stereochemistry (−2); factor 9b, carboxylate within H-bonding distance of histidine in template (+2); factor 10, $M_r < 300$ (−1), $M_r > 300$ (−2). Manual optimisation of the aggregates (T) led to a good linear correlation against experimental log K, with 1 M binding standardized at $T = 0$. This indicated that T was an estimate of ΔG, and the gradient allowed us to determine that a scaling factor of 2.8 would convert T (number) into ΔG (kJ mol^{-1}).

9 (*a*) We found that commercial pharmacophore based modelling was reasonably successful for 'standard' substrates, but the prediction of activity was unreliable for structures outwith the initial training set, with high affinity often attributed to poor substrates; (*b*) Note that, although there is not a linear correlation, medium/high affinity substrates are almost always transported by PepT1 (see ref. 1); there is only one report of a non-transported inhibitor for PepT1,[11a] and one report for PepT2[25].

10 V. Borner, Y. J. Fei, B. Hartrodt, V. Ganapathy, F. H. Leibach, K. Neubert and M. Brandsch, *Eur. J. Biochem.*, 1998, **255**, 698.

11 (*a*) K_i from oocytes expressing PepT1—see: D. Meredith, C. A. R. Boyd, J. R. Bronk, P. D. Bailey, K. M. Morgan, I. D. Collier and C. S. Temple, *J. Physiol.*, 1998, **512**, 629; (*b*) K_i from reconstituted BBMV—see: C. S. Temple, P. D. Bailey, J. R. Bronk and C. A. R. Boyd, *J. Physiol.*, 1996, **494**, 795.

12 H. Daniel, E. L. Morse and S. A. Adibi, *J. Biol. Chem.*, 1992, **267**, 9565.

13 D. Meredith, C. S. Temple, N. Guha, C. J. Sword, C. A. R. Boyd, I. D. Collier, K. M. Morgan and P. D. Bailey, *Eur. J. Biochem.*, 2000, **267**, 3723.

14 M. Brandsch, F. Thunecke, G. Kullertz, M. Schutkowski, G. Fischer and K. Neubert, *J. Biol. Chem.*, 1998, **273**, 3861.

15 R. Wootton and R. Hazelwood, *Biochem. Soc. Trans.*, 1989, **17**, 691.

16 U. Wenzel, D. T. Thwaites and H. Daniel, *Br. J. Pharmacol.*, 1995, **116**, 3021.

17 M. E. Ganapathy, W. Huang, H. Wang, V. Ganapathy and F. H. Leibach, *Biochem. Biophys. Res. Commun.*, 1998, **246**, 470.

18 C. S. Temple and C. A. R. Boyd, *Biochim. Biophys. Acta*, 1998, **1373**, 277.

19 F. Döring, J. Will, S. Amasheh, W. Clauss, H. Ahlbrecht and H. Daniel, *J. Biol. Chem.*, 1998, **273**, 23211.

20 C. S. Temple, J. R. Bronk, P. D. Bailey and C. A. R. Boyd, *Pflügers Arch.*, 1995, **430**, 825.

21 C. S. Temple, I. D. Collier, J. R. Bronk, P. D. Bailey and C. A. R. Boyd, *J. Physiol.*, 1998, **511**, 26P.

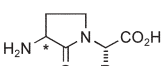
22 C. S. Temple, A. K. Stewart, D. Meredith, N. A. Lister, K. M. Morgan, I. D. Collier, R. D. Vaughan-Jones, C. A. R. Boyd, P. D. Bailey and J. R. Bronk, *J. Biol. Chem.*, 1998, **273**, 20.

23 C. J. Sword, C. A. R. Boyd and C. S. Temple, *J. Physiol.*, 1999, **515**, 169P.

24 T. Terada, K. Sawada, M. Irie, H. Saito, Y. Hashimoto and K. Inui, *Pflügers Arch.*, 2000, **440**, 679.

25 S. Theis, I. Knütter, B. Hartrodt, M. Brandsch, G. Kottra, K. Neubert and H. Daniel, *J. Biol. Chem.*, 2002, **277**, 7287.

26 As expected, our quantitative model is a good predictor of the binding of new substrates, as evidenced by the derivatives shown below, which were chosen to probe details of stereochemistry at residue 1 (ref. 7a), and at the 1st peptide bond (ref. 7b).

Substrate		K_i/mM	T	K_{calc}	Ref
	* = *L*, R = H	2.25	5	3.6	7*a*
	* = *D*, R = H	0.37	7	0.38	7*a*
	* = *L*, R = Me	1.04	5	3.6	7*a*
	* = *D*, R = Me	0.32	7	0.38	7*a*
Gly–Sar–Ala		0.54	7	0.38	7*a*
Ala–Sar–Ala		1.19	7	0.38	7*a*
D-Ala–Sar–Ala		5.88	5	3.6	7*a*
Aib–Sar–Ala		4.73	5	3.6	7*a*
trans-Ala-Ψ[CS-N]-Pro		0.36	7	0.38	7*b*
cis-Ala-Ψ(CS-N)-Pro		>4	5	3.6	7*b*
Ala-Ser(OBn)		0.23	7	0.38	7*b*
Ala-Ψ[CS-N]-Ser(OBn)		0.29	7	0.38	7*b*