	Substrate	M.Wt.	Ki	Log(Ki)	F1	F2	F3	F4	F5	F6	F7	F8	F9a	F9b	F10	Total	Ki	Error	Ref	tissue	Ki or Km?	Log(K)	+sem	-sem	Substrate
1	L-Dopa-Phe	344	0.03	-1.52	2	1	1	1	0	2	2	2			-2	9	0.03	nd	1	oocytes	Ki	-1.52	nd	nd	L-Dopa-Phe
2	Ala-NH-C6H4-(4Ph)	240	0.03	-1.52	2	1	1	1	0	0	4	0			-1	8	0.03	0.00	2	caco-2	Ki	-1.52	0.04	-0.05	Ala-NH-C6H4-(4Ph)
3	Gly-Ala	146	0.032	-1.49	2	1	1	1	0	2	0	2			-1	8	0.032	0.00	5	oocytes	Ki	-1.49	0.04	-0.04	Gly-Ala
4	Gly-Val	174	0.032	-1.49	2	1	1	1	0	2	0	2			-1	8	0.032	0.00	3	bbmv	Ki	-1.49	0.01	-0.01	Gly-Val
5	Ala-Ala	160	0.08	-1.10	2	1	1	1	0	2	0	2			-1	8	0.08	0.01	3	bbmv	Ki	-1.10	0.05	-0.06	Ala-Ala
6	Phe-Tyr	328	0.1	-1.00	2	1	1	1	0	2	2	2			-2	9	0.1	0.04	4	oocytes	Ki	-1.00	0.15	-0.22	Phe-Tyr
7	Asp-Ala	204	0.12	-0.92	2	1	1	1	0	2	0	2			-1	8	0.12	nd	5	oocytes	Ki	-0.92	nd	nd	Asp-Ala
8	Ala-Ala-Ala	231	0.16	-0.80	2	1	1	1	0	2	0	0	0	2	-1	8	0.16	0.14	4	oocytes	Ki	-0.80	0.27	-0.90	Ala-Ala-Ala
9	Ser-Ala	176	0.21	-0.68	2	1	1	1	0	2	0	2			-1	8	0.21	nd	5	oocytes	Ki	-0.68	nd	nd	Ser-Ala
10	Phe-Ala	236	0.21	-0.68	2	1	1	1	0	2	0	2			-1	8	0.21	0.10	5	oocytes	Ki	-0.68	0.17	-0.28	Phe-Ala
11	Arg-Ala	245	0.22	-0.66	2	1	1	1	0	2	0	2			-1	8	0.22	nd	5	oocytes	Ki	-0.66	nd	nd	Arg-Ala
12	Phe-Pro	262	0.23	-0.64	2	1	1	1	-1	2	0	2			-1	7	0.23	nd	5	oocytes	Ki	-0.64	nd	nd	Phe-Pro
13	Ala-Pro	186	0.25	-0.60	2	1	1	1	-1	2	0	2			-1	7	0.25	0.10	5	oocytes	Ki	-0.60	0.15	-0.22	Ala-Pro
14	Ala-(trans-thio)-Pro	202	0.3	-0.52	2	1	1	0	-1	2	0	2			-1	6	0.3	0.02	6	caco-2	Ki	-0.52	0.03	-0.03	Ala-(trans-thio)-Pro
15	Ala-NH-C6H4-(4Me)	178	0.34	-0.47	2	1	1	1	0	0	2	0			-1	6	0.34	0.04	2	caco-2	Ki	-0.47	0.05	-0.05	Ala-NH-C6H4-(4Me)
16	H2N-CH2-(C=O)-C2H4-CO2	131	0.4	-0.40	2	1	0	1	0	2	0	2			-1	7	0.4	nd	11	P.pastoris	Ki	-0.40	nd	nd	H2N-CH2-(C=O)-C2H4-CO2H
17	Val-Lys	245	0.64	-0.19	2	1	1	1	0	2	-1	2			-1	7	0.64	nd	7	bbmv	Ki	-0.19	nd	nd	Val-Lys
18	L-Loracarbef	349	0.7	-0.15	2	1	1	1	0	2	0	0	0	2	-2	7	0.7	0.14	8	caco-2	Ki	-0.15	0.08	-0.10	L-Loracarbef
19	Val-Acyclovir	324	0.74	-0.13	2	1	1	1	1	0	2	0			-1	7	0.74	0.14	13	HeLa	Ki	-0.13	0.08	-0.09	Val-Acyclovir
20	Phe-Tyr-NH2	327	0.9	-0.05	2	1	1	1	0	2	2	0			-2	7	0.9	0.38	4	oocytes	Ki	-0.05	0.15	-0.24	Phe-Tyr-NH2
21	Ala-Ala-D-Ala	231	0.99	0.00	2	1	1	1	0	2	0	0	-2	2	-1	6	0.99	0.22	4	oocytes	Ki	0.00	0.09	-0.11	Ala-Ala-D-Ala
22	Enalapril	376	1.1	0.04	0	1	1	1	0	2	0	2			-2	5	1.1	0.30	10	bbmv	Ki	0.04	0.10	-0.14	Enalapril
23	H2N-(CH2)4-CO2H	117	1.14	0.06	2	1	0	0	0	2	0	2			-1	6	1.14	0.06	11	P.pastoris	Ki	0.06	0.02	-0.02	H2N-(CH2)4-CO2H
24	D-Phe-Ala	236	1.14	0.06	2	-1	1	1	0	2	0	2			-1	6	1.14	0.16	12	bbmv	Km	0.06	0.06	-0.07	D-Phe-Ala
25	Pro-Ala	186	1.26	0.10	0	1	1	1	0	2	0	2			-1	6	1.26	nd	5	oocytes	Ki	0.10	nd	nd	Pro-Ala
26	D-Phe-Gly	222	1.7	0.23	2	-1	1	1	0	2	0	2			-1	6	1.7	nd	5	oocytes	Ki	0.23	nd	nd	D-Phe-L-Gly
27	D-Loracarbef	349	1.8	0.26	2	-1	1	1	0	2	0	0	0	2	-2	5	1.8	0.25	8	caco-2	Ki	0.26	0.06	-0.06	D-Loracarbef
28	Ac-Phe	207	2	0.30	-2	0	1	1	0	2	2	2			-1	5	2	0.37	4	oocytes	Ki	0.30	0.07	-0.09	Ac-Phe
29	D-Amoxycillin	365	2	0.30	2	-1	1	1	0	2	0	0	0	2	-2	5	2	nd	8	caco-2	Ki	0.30	nd	nd	D-Amoxycillin
30	D-Phe-Glu	294	2.15	0.33	2	-1	1	1	0	2	-1	2			-1	5	2.15	0.10	12	bbmv	Km	0.33	0.02	-0.02	D-Phe-Glu
31	Ala-NH-Ph	164	2.9	0.46	2	1	1	1	0	0	1	0			-1	5	2.9	0.30	2	caco-2	Ki	0.46	0.04	-0.05	Ala-NH-Ph
32	D-Ala-Ala-Ala	231	3.04	0.48	2	-1	1	1	0	2	0	0	0	2	-1	6	3.04	0.96	4	oocytes	Ki	0.48	0.12	-0.17	D-Ala-Ala-Ala
33	4-H2N-CH2-C6H4-CO2H	151	3.1	0.49	2	1	1	0	0	0	0	2			-1	5	3.1	0.90	9	oocytes	Ki	0.49	0.11	-0.15	4-H2N-CH2-C6H4-CO2H
34	Captopril	217	4	0.60	0	0	1	1	-1	2	0	2			-1	4	4	1.00	10	bbmv	Ki	0.60	0.10	-0.12	Captopril
35	3-H2N-C6H4-CH2-CO2H	151	6	0.78	0	0	1	0	0	2	0	2			-1	4	6	2.00	5	bbmv	Ki	0.78	0.12	-0.18	3-H2N-C6H4-CH2-CO2H
36	Ala-D-Ala-Ala	231	6.43	0.81	2	1	1	1	0	-2	0	0	0	2	-1	4	6.43	1.83	4	oocytes	Ki	0.81	0.11	-0.15	Ala-D-Ala-Ala
37	4-H2N-C6H4-CH2-CO2H	151	6.5	0.81	0	0	1	0	0	2	0	2			-1	4	6.5	0.30	14	bbmv	Ki	0.81	0.02	-0.02	4-H2N-C6H4-CH2-CO2H
38	Ac-Phe-Tyr	370	8.4	0.92	-2	1	1	1	0	2	2	0	0	2	-2	5	8.4	0.11	4	oocytes	Ki	0.92	0.01	-0.01	Ac-Phe-Tyr
39	Ac-Phe-Tyr-NH2	369	10	1.00	-2	1	1	1	0	2	2	0	0	0	-2	3	10	4.01	4	oocytes	Ki	1.00	0.15	-0.22	Ac-Phe-Tyr-NH2
40	4-H2N-C6H4-CO2H	137	10.6	1.03	0	0	1	0	0	2	0	2			-1	4	10.6	3.04	5	bbmv	Ki	1.03	0.11	-0.15	4-H2N-C6H4-CO2H
41	Ala-NH-CH2-(Ph)	178	14.1	1.15	2	1	1	1	0	0	0	0			-1	4	14.1	1.10	2	caco-2	Ki	1.15	0.03	-0.04	Ala-NH-CH2-(Ph)
42	D-Phe-L-Pro	262	21	1.32	2	-1	1	1	-1	2	0	2			-1	5	21	n.d.	5	oocytes	Ki	1.32	nd	nd	D-Phe-L-Pro
43	Ac-Phe-NH2	206	22	1.34	-2	0	1	1	0	2	2	0			-1	3	22	5.64	4	oocytes	Ki	1.34	0.10	-0.13	Ac-Phe-NH2
44	Ala-D-Phe-Ala	307	22.7	1.36	2	1	1	1	0	-2	0	0	0	2	-2	3	22.7	7.10	15	oocytes	Ki	1.36	0.12	-0.16	Ala-D-Phe-Ala
45	Phe-NH2	164	50	1.70	2	1	-1	1	0	0	0	0			-1	2	50	nd	4	oocytes	Ki	1.70	nd	nd	Phe-NH2
46	Phe	165	100	2.00	2	1	-1	1	0						-1	2	100	nd	4	oocytes	Ki	2.00	nd	nd	Phe
47	H2N-(CH2)3-CO2H	103	>50		2	1	0	0	0						-1	2	>50		11	P.pastoris	Ki				H2N-(CH2)3-CO2H
48	L-Dopa	197	>100		2	1	0	0	0						-1	2	>100		1	oocytes	Ki				L-Dopa
49	Acyclovir	225	>100		-2	0	0	1	0	0	2				-1	0	>100		13	HeLa	Ki				Acyclovir
50	cyclo(Gly-Gly)	128	>100		-2	1	0	1	0						-1	-1	>100		16	LLC cells	Ki				cyclo(Gly-Gly)

References for Table 1 (ESI)

The following references relate to the Table in the ESI – they are provided for ease of location, and are identical to those in the paper (but numbered differently).

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

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

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

4 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.

5 New data presented in this paper, determined using one of two methods: 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 reconsituted BBMV – see C.S. Temple, P.D. Bailey, J.R. Bronk and C.A.R. Boyd, *J. Physiol.*, 1996, **494**, 795.

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

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

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

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

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

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

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

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

14 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.

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

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

Notes for determining "T" in Table 1 (ESI)

a) Factor 2: for the stereochemistry at residue 1, L-residues normally adopt the template conformation. However, note that it is the spacial arrangement that is important, so that the constrained D-analogues presented in reference 26 of the paper possess the correct 3D arrangement for tight binding, whereas the L-isomers do not; α,α-disubstituted residues at position 1 also show reduced binding because the "D" side-chain has an unfavourable interaction (P.D. Bailey, C.A.R. Boyd, I.D. Collier, J.G. George, G.L. Kellett, D. Meredith, K.M. Morgan, R. Pettecrew, R.G. Pritchard and R.A. Price, *Chem. Commun.*, 2005, DOI: 10.1039/b510697d).

b) Factor 4: for the first amide bond, the presence of a C=O or C=S has been explored, both contributing to improved binding (parameter of +1); other potential changes at this position have not been probed.

c) Factor 7: the hydrophobic pocket confers significantly improved binding when aromatic residues can be located therein – more precise measures of

hydrophobicity failed to provide better correlation. Note that the 4-toluidide (entry 15) reaches the hydrophobic pocket by virtue of the methyl group, and such 4-substituted analogues show high binding (ref. 13 above). However, the unsubstituted anilide (entry 31) binds much less well as it fails to extend properly into the pocket (see Figure 2 of the paper), whilst the benzyl analogue (entry 40) also binds poorly as the aryl group is misaligned. Charged residues in the pocket lead to reduced binding (incurring a penalty of -1 in the calculation of T). Heteroaryl groups have not screened, but a binding parameter of +1 for such groups might be expected to generate reasonable correlations.

d) There is no preconception about the mode of binding in the analyses – that alignment which generates the highest T value is assumed to correspond to the mode of binding. A good set of examples to illustrate this are the Phe analogues (entries 28, 43, 45, 46) and the Phe-Tyr analogues (entries 6, 20, 38, 39), which show T values ranging from 2–9.

e) Finally, it must be recognized that substrates with structures significantly different from the (admittedly wide) range used in this analysis cannot be predicted with great confidence. For example, there may be size limits associated with side chains R^1, R^2 and R^3, but we have not yet reached those with substrates up to indolylmethyl (i.e. the side chain on tryptophan), or significantly bigger (e.g. steroidal) in the case of R^2.