Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library

Keisuke Ito1,2,*, Aya Hikida1,*, Shun Kawai3, Vu Thi Tuyet Lan1, Takayasu Motoyama4, Sayuri Kitagawa4, Yuko Yoshikawa1,2, Ryuji Kato3 & Yasuaki Kawarasaki1,2

Peptide uptake systems that involve members of the proton-coupled oligopeptide transporter (POT) family are conserved across all organisms. POT proteins have characteristic substrate multispecificity, with which one transporter can recognize as many as 8,400 types of di/tripeptides and certain peptide-like drugs. Here we characterize the substrate multispecificity of Ptr2p, a major peptide transporter of Saccharomyces cerevisiae, using a dipeptide library. The affinities ($K_i$) of di/tripeptides toward Ptr2p show a wide distribution range from 48 mM to 0.020 mM. This substrate multispecificity indicates that POT family members have an important role in the preferential uptake of vital amino acids. In addition, we successfully establish high performance ligand affinity prediction models (97% accuracy) using our comprehensive dipeptide screening data in conjunction with simple property indices for describing ligand molecules. Our results provide an important clue to the development of highly absorbable peptides and their derivatives including peptide-like drugs.

DOI: 10.1038/ncomms3502

1 Department of Food and Nutritional Sciences, Graduate School of Nutritional and Environmental Sciences, University of Shizuoka, Yada 52-1, Suruga-ku, Shizuoka 422-8526, Japan. 2 School of Food and Nutritional Sciences, University of Shizuoka, Yada 52-1, Suruga-ku, Shizuoka 422-8526, Japan. 3 Department of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, Furocho, Chikusaku, Nagoya 464-8601, Japan. 4 Food Science Research Institute, Research & Development HQ, FujiOil Co., Ltd., Kinunodai 4-3, Tsukubamirai, Ibaraki 300-2497, Japan. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.I. (email: sukeito@u-shizuoka-ken.ac.jp).
Peptide uptake systems are conserved across all organisms from bacteria to higher animals and plants and are important for acquiring nitrogen resources with high efficiency\(^1\)-\(^6\). The molecules involved in these systems belong to the proton-coupled oligopeptide transporter (POT) family, which is also called the peptide transporter (PTR) family\(^7\). POT family members comprise 12 transmembrane domains as a general structure and have characteristic substrate multispecificity, by which one transporter can recognize a variety of substrates\(^8\). In humans, hPEPT1 expressed in the small intestine epithelium is involved in absorbing nutritional peptides\(^9\)-\(^11\), whereas hPEPT2 expressed in the renal tubes is involved in the reabsorption of peptides from primitive urine\(^11\)-\(^13\).

Due to their substrate multispecificity, it is assumed that each of these POT family transporters can recognize as many as 8,400 types of di/tripeptides, which are products generated by protein hydrolysis, at a single substrate-binding site that resides on each transporter and can actively transport these peptides\(^14\). Furthermore, hPEPT1 and hPEPT2 can transport peptide-like drugs such as β-lactam antibiotics, anti-hypertensive drugs and anti-cancer agents\(^15\). This property influences the intestinal absorption of drugs and half-life of drugs in blood. For the yeast S. cerevisiae, Ptr2p is the major transporter involved in the uptake of di/tripeptides\(^17\)-\(^20\). Because the uptake of amino acids in the peptide form is much faster than in the amino-acid form, peptide-based media can improve the growth of yeasts\(^6\),\(^21\). Although Ptr2p is an important protein for the fermentation industry, there is little information on the substrate preferences of Ptr2p\(^22\). As with the examples above, the substrate multispecificity of a POT family protein is of interest in various fields of science including drug development, nutrition and fermentation.

The question of how one substrate-binding site can recognize a variety of substrates is the major focus in peptide transport research. To date, the necessary or important structural characteristics for recognition by POT family transporters have been proposed by determining the affinity of individual substrates for these transporters\(^23\)-\(^25\). However, the entire spectrum of the substrate preference of POT family transporter has not been elucidated. In addition, the crystal structures of two bacterial POT family transporters, PepTso\(^14\) from Shewanella oneidensis and PepTst\(^15\) from Streptococcus thermophilus, were recently determined, which provided important clues for elucidating the substrate multispecificity of the POT family. Because the amino-acid sequence of PepTso is highly homologous to those of hPEPT1 and hPEPT2, PepTso is the best structural model currently available for these eukaryotic peptide transporters. The crystal structure of PepTso provided basic information on the three dimensional configuration of the amino-acid residues at its substrate-binding site. To advance further in designing a detailed pharmacophore map, it is necessary to elucidate the physico-chemical characteristics that determine the affinity of a substrate for a POT family transporter.

This study reports the substrate multispecificity of S. cerevisiae Ptr2p used as a model POT family transporter. Detailed substrate multispecificity of the Ptr2p was characterized by a comprehensive analysis using a dipeptide library and a high-throughput assay system developed by us. By constructing models to predict the dipeptide affinities for Ptr2p, we observed that ligand affinity \(K_i\) values could be predicted in \textit{silico} with high accuracy. For \textit{in silico} model construction, we used a combination of simple ligand property parameters rather than using complex structural information for ligands and receptors. The biological mechanisms and roles of POT family proteins are discussed on the basis of careful examination of the prediction models constructed using different property parameters as descriptive parameters for ligand affinity.

**Figure 1 | Generation of Ptr2p-expressing cells.** (a) Western blot analysis of the membrane fraction prepared from SC-Ptr2p cells (BY4742-ptr2Δ) cells that expressed plasmid-borne Ptr2p) using an anti-FLAG antibody. Arrowhead indicates expressed Ptr2p. (b) Localization of a Ptr2p–GFP fusion protein using confocal fluorescence microscopy in BY4742-ptr2Δ. Scale bar represents 5 µm. (c) Spot assay. Dipeptide uptake capability of SC-Ptr2p cells was analysed. Leucine and histidine, which are required for the growth of strain BY4742 were added to the medium in the form of a dipeptide (His-Leu, 10 mM).
These results indicated that Ptr2p was potent, expressed exogenously and localized on the cell surface.

A fluorescence-based competitive uptake assay system. A fluorescence-based competitive uptake (F-CUP) assay system is a high-throughput assay that determines the competitive-inhibitory activity (IC50) of an analysed substrate versus the uptake of a tracer substrate. An IC50 value can be converted to a Ki value using the Cheng–Prusoff equation26. Functional analysis of Ptr2p peptides showed competitive-inhibitory effects against the uptake of oligopeptides with different chain lengths (Fig. 2d). Only di/tripeptides that included leucine, Leu–Ala (Ki = 0.15 mM) and Ala–Leu (Ki = 0.31 mM), were higher than that of Ala–Ala (Ki = 0.40 mM). The affinity of Leu–Ala was twice that of Ala–Leu. Thus, this demonstrated that in addition to the amino-acid composition, the amino-acid sequence also contributed to the affinity for Ptr2p.

Relationship between Ptr2p affinity and S. cerevisiae growth. The affinity of dipeptides for Ptr2p and their effects on the growth of S. cerevisiae were examined using two different combinations of dipeptides: His–Leu and Leu–His or Leu–Gly and Gly–Leu (Fig. 3). The Ki values of His–Leu, Leu–His, Leu–Gly and Gly–Leu were 0.05, 0.13, 0.36 and 0.60 mM, respectively. Yeast cell growth analysis indicated that those dipeptides with lower Ki values were better nutrients for Ptr2p-expressing yeast despite their identities in terms of their amino-acid composition. The effect for improving cell growth by a high-affinity peptide was also verified using the FGY217 strain, which did not artificially express Ptr2p (Supplementary Fig. S1). The F-CUP assay system combined with growth analysis can also be a useful tool for developing an efficient fermentation medium.

F-CUP assay using a dipeptide library. We performed a comprehensive analysis using a dipeptide library by the F-CUP assay system to characterize the substrate multispecificity of Ptr2p (Fig. 4a). For 338 types of dipeptides that could be synthesized, we calculated the Ki values of 237. Ki values could not be calculated for 100 types of dipeptides due to the limitations of dipeptide synthesis. The highest affinity was observed for Ala–Leu (0.02 mM), followed by Ala–Ala (0.16 mM), Ala–Gly (0.17 mM), Gly–Ala (0.18 mM) and Gly–Leu (0.30 mM). The affinity of Gly–Ala and Gly–Leu were 0.05, 0.13, 0.36 and 0.60 mM, respectively. Yeast cell growth analysis indicated that those dipeptides with lower Ki values were better nutrients for Ptr2p-expressing yeast despite their identities in terms of their amino-acid composition. The effect for improving cell growth by a high-affinity peptide was also verified using the FGY217 strain, which did not artificially expressPtr2p (Supplementary Fig. S1). The F-CUP assay system combined with growth analysis can also be a useful tool for developing an efficient fermentation medium.
Kidipeptides with a high-affinity dipeptide were selected for this study. We incubated the group of high-affinity dipeptides (0.14% yeast synthetic drop-out medium without leucine, 0.67% yeast nitrogen base without amino acids and 2% galactose) and calculated their affinities with $K_i$ values above 0.77 mM. We used these in silico affinity analysis for substrates for which the IC$_{50}$ values were above 1.0 mM due to their poor solubility. Instead, their $K_i$ values were assigned to be greater than 0.77 mM.

The calculated $K_i$ values showed a wide range distribution (Fig. 4b). Trp–Phe exhibited the highest affinity ($K_i = 0.020$ mM). It was also found that their affinities for Ptr2p were 2,400 times higher than those with the lowest affinity: Gly–Gly–Gly with $K_i = 48$ mM. In this study, we selected a group of high-affinity dipeptides with $K_i$ values below 0.077 mM and a group of low-affinity dipeptides with $K_i$ values above 0.77 mM. We used these in an appearance frequency analysis of the amino-acid residues using the WebLogo programme (Fig. 4c). This showed that dipeptides containing aromatic amino acids (namely Phe, Trp and Tyr) and branched-chain amino acids (namely Ile, Leu and Val) frequently appeared in the high-affinity group. On comparing, it was found that the low-affinity group had a high frequency of negatively charged amino acids (that is, Asp and Glu), as well as amino acids that were predicted to influence peptide bond conformation (that is, Gly and Pro). In both groups of dipeptides, amino-acid residues at the amino terminus showed a higher propensity compared with those at the C-terminus, which suggested that an amino-acid residue at the N-terminal had a more significant role in recognition by Ptr2p than those in the C-terminus.

constructed:{
    "text": "Constructing ligand affinity prediction models. To expand the applications of our assay data, we constructed discrimination analysis models to predict ligand affinity in silico (Table 1). Our assay data comprised discrete over-threshold data ($K_i > 0.77$) for several low-affinity dipeptides within continuous $K_i$ data, along with future screening applications for in silico pre-screening; therefore, we selected discrimination analysis models to predict categorical labels for ligand molecules. Compared with conventional ligand prediction models, we selected features that could be simply calculated from the primary sequences of dipeptides as descriptive parameters to construct simple prediction models that utilize fewer parameters.

We examined a total of six prediction models and observed that categorizing dipeptide samples as high, medium or low can predict affinities with extremely high accuracy (average prediction > 84%). By comparing the data set type for whether or not it included intermediate $K_i$ ligands, we observed that 97% prediction accuracy (data set type B, prediction model type M2) can be achieved for objectively screening ligands that would interact with Ptr2p. Even with data set type A modelled by M2, predictions were accurate in the area of "extremely high-affinity samples", which indicated sufficient applicability for ligand screening (Supplementary Fig. S2).

By comparing these prediction model types, we observed that the combined information on amino-acid residues and chemical property information was most effective for obtaining an accurate prediction model. Prediction accuracy slightly increased by adding chemical property parameters that described total ligand molecular properties. However, sufficient prediction accuracy was achieved by converting the primary sequences of dipeptides into a few amino-acid indices. In addition, we observed that even without amino-acid information, total molecular chemical property information could be used as an alternative parameter to maintain similar prediction accuracy.

From the parameter selection process of constructing six discrimination analysis models, the manner of a dipeptide–Ptr2p interaction could be determined. During the modelling process with model M1, index 3 (side-chain contribution to protein stability) at the N-terminal, index 14 (side-chain interaction parameter) and indices 8 and 1 (isoelectric point) were the first four parameters that greatly contributed to the increased prediction accuracy. These results were common physicochemical rules for interactions between dipeptides and Ptr2p, which could be extracted from our comprehensive dipeptide library assay data. This extraction rule by model analysis was only possible with the affinity data with variety, and would not be attainable from partial positive screening data obtained from conventional, limited size assays. In addition to the discriminant analysis models that predicted the categories of ligand affinities, multiple regression models that predicted $K_i$ values directly from molecular properties were observed to provide sufficient accuracy for screening (corrected $R^2$ values > 0.734).

PTR2 gene expression controlled by an N-end rule dipeptide. To examine the effect of amino-acid sequence of dipeptide on PTR2 gene expression, we analysed yeasts grown in YPD media containing dipeptides Ala–Ala or Trp–Ala (Fig. 5). In a direct analysis of gene expression by the FGY217 strain using real time RT–PCR, PTR2 gene expression increased twofold by adding the N-end rule dipeptide Trp–Ala in the YPD medium as compared with adding the non N-end rule dipeptide Ala–Ala.

Discussion
In this study, the substrate multispecificity of Ptr2p, the major peptide transporter of S. cerevisiae, was characterized using an F-CUP assay system (Fig. 4). Although there have been several reports regarding affinity analyses of substrates for POT family members$,^{27–29}$ this is the first study to use a dipeptide library for a comprehensive analysis. From this library assay data, we also successfully constructed in silico ligand affinity prediction models.
for Ptr2p using discriminant analysis models. By analysing the model construction processes, new insights were obtained to further understand the manner of interactions between dipeptides and Ptr2p.

The primary structure of Ptr2p was compared with that of other family members to gain a better understanding of the substrate recognition mechanism of a POT family protein. Although the similarity between the primary structure of Ptr2p

---

**Table 1:** Distribution of affinity values for N-terminal C-terminal residues with high or low-affinity dipeptides using the WebLogo programme (http://weblogo.berkeley.edu/). Blue, red and pink arrowheads indicate Blue, red and pink arrowheads indicate 

---

**Figure 4** | Substrate multispecificity of Ptr2p. (a) Comprehensive analysis using a dipeptide library by the F-Cup assay system. Colours of cells correspond to $K_i$ values. N.T.: not tested. Data are presented as means ± s.d. (n = 3). (b) Distribution of $K_i$ values. (c) Frequency analysis of amino-acid residues with high or low-affinity dipeptides using the WebLogo programme (http://weblogo.berkeley.edu/). Blue, red and pink arrowheads indicate N-end rule amino acids in yeast, essential amino acids and semi-essential amino acids for humans, respectively.
and that of PepTso was 33.7% (Supplementary Fig. S3), three dimensional structures suggested that 12 amino-acid residues comprising the substrate-binding site of PepTso were highly conserved among these family members (Fig. 6a). For Ptr2p, 11 amino-acid residues, excluding Ser100 (corresponds to Arg32 in PepTso)(Supplementary Fig. S3), were functionally similar to those of PepTso. Dipeptides that comprise aromatic amino acids displayed high affinity for Ptr2p, whereas dipeptides that comprise negatively charged amino acids displayed low affinity for Ptr2p (Fig. 4c). The results obtained using individual dipeptides were consistent with our previous results obtained from a comprehensive analysis of a dipeptide library. This was supported by the fact that the substrate-binding sites of these family members comprise highly conserved amino-acid residues.

In S. cerevisiae, Ptr2p expression is regulated by the N-end rule pathway31–33, by which the binding of dipeptides with a certain N-terminus to Ubr1p promotes the degradation of Cup9p, which is a repressor of the PTR2 gene. Thus, the Ptr2p expression level increases when peptides that meet the N-end rule are imported. Cai et al.22 reported that several N-end rule peptides were preferentially recognized by Ptr2p. This is consistent with our present results obtained from a comprehensive analysis of a dipeptide library. The N-terminus of our high-affinity group is

![Table 1 | Prediction results of dipeptide affinity discrimination models with their parameter selection processes.](image-url)

| Data set type | Type A | Type B |
|---------------|--------|--------|
| Ki data (total 337 samples) | 100% | 100% |
| Low Ki (102 samples) | 100% | 100% |
| Medium Ki (135 samples) | 100% | 0% |
| High Ki (100 samples) | 100% | 100% |

**Prediction model type**

| Description and parameter selection results* | Type I parameters (amino-acid indices) | Type II parameters (chemical property indices) | C-terminal amino acid |
|--------------------------------------------|---------------------------------------|-----------------------------------------------|----------------------|
| N-terminal amino acid                      | 1 Isoelectric point                    | 8 Isoelectric point                            | 3 N-terminal amino acid |
|                                            | 2 Normalized van der Waals volume      | 9 Normalized van der Waals volume              | 6 C-terminal amino acid |
|                                            | 3 Side-chain contribution to protein stability | 10 Side-chain contribution to protein stability | 12 Side-chain interaction parameter |
|                                            | 4 Hydropathy index                     | 11 Hydropathy index                            |                       |
|                                            | 5 Normalized frequency of turn         | 12 Normalized frequency of turn                |                       |
|                                            | 6 Polarity                             | 13 Polarity                                    |                       |
|                                            | 7 Side-chain interaction parameter     | 14 Side-chain interaction parameter            |                       |

| Objective parameter | 3 types of labels (low Ki/medium Ki/high Ki) |
|---------------------|---------------------------------------------|
| Low Ki              | 75.5% 77.5% 68.6% 96.1% 100.0% 87.0% |
| Medium Ki           | 73.3% 71.1% 77.0% 95.0% 94.0% 89.6% |
| High Ki             | 79.0% 81.0% 71.0% 95.0% 94.0% 89.6% |

Open circles represent input parameters; dashes, parameters not used for inputs; count numbers, parameter selection order; bold values indicate first four parameters selected to construct a model. Amino-acid indices are from References45–50.
consistent with amino acids that adhered to the N-end rule (Trp, Phe, His, Tyr, Leu, Ile and Lys; Fig. 4c). Interestingly, most of the essential and semi-essential amino acids for humans (Trp, Phe, His, Tyr, Met, Leu, Ile, Val, Lys and Arg) were also the constituent amino acids of high-affinity peptides (Fig. 4c). In general, the biosynthesis of aromatic amino acids requires the expression of multiple enzymes and involves energy-consuming reactions, whereas biosynthesis of acidic amino acids occurs from shorter branches of the TCA cycle. The substrate multispecificity of the POT family of transporter, revealed by analysing Ptr2p, indicates that these family members are involved in the preferential uptake of specific amino acids that impose a biosynthesis burden on organisms (Fig. 4c). Based on this perspective, we propose the following positive feedback model for peptide uptake into S. cerevisiae cells based on the substrate multispecificity of Ptr2p. (ii) Peptide uptake is accurately sensed by the N-end rule pathway via Ubr1p. (iii) Subsequently, Ptr2p expression is promoted after Cup9p degradation. (iv) S. cerevisiae cells can then more efficiently absorb those vital amino acids in their peptide form. Enhanced PTR2 gene expression by a transported dipeptide was experimentally demonstrated using an N-end rule dipeptide, Trp–Ala (Fig. 5). The association between the substrate multispecificity of Ptr2p and the regulatory system for PTR2 gene expression is biologically reasonable.

By examining our prediction models, we observed that a few simple parameters that could be obtained from ligand sequence information could produce high accuracy ligand prediction models. The accuracy and the construction processes indicated that a few physicochemical properties of dipeptides were sufficient for discriminating their affinities for Ptr2p. Our prediction model accuracy strongly suggests that Ptr2p recognition is primarily governed by ‘property combinations’ that characterize the physicochemical properties of ligands rather than their exact sequence motifs. By comparing different types of descriptive parameters that can be derived from the same dipeptide sequence (M1, M2 and M3 comparisons in Table 1), we observed that the best prediction accuracy can be obtained when both ‘amino-acid-specific physicochemical properties’ and ‘total molecular chemical property’ were used as descriptors for dipeptide molecules. However, similar affinity prediction accuracy for Ptr2p was retained even after eliminating the ‘amino acid-specific physicochemical properties’. These results indicate that our affinity prediction models constructed in silico are applicable to pre-screening for medical applications. To examine this possibility, we compared the experimental and predicted binding affinities to Ptr2p for six medical compounds as a trial (Supplementary Table S1). The predicted results matched F-CU assay results for alafoxasil, araphmenine B, valacyclovir and benazepril. Based on the distribution of their molecular sizes, we assumed that the constructed prediction model using dipeptide library affinity data could predict the affinities for molecules that were close in size to those of dipeptides, having an average molecular weight of 256. In other words, it was reasonable that the predicted performance was limited to the variations in the ‘molecular chemical properties’ that existed in the dipeptide library used for model training. Therefore, increasing the ‘molecular information’ to train for wider variations in transporter affinities along with adding descriptive parameters will be the focus of our next investigation to expand our prediction approach. However, despite its limited application, it should be noted that our F-CU assay-derived exhaustive

Figure 5 | PTR2 gene expression controlled by an N-end rule dipeptide.
Pre-cultures of the FGY217 strain were diluted to give OD660 ≈ 0.06 in each test YPD medium (2% peptone, 1% yeast extract, 2% glucose and 10 mM Trp–Ala or Ala–Ala), and then grown at 30 °C for 5 h. PTR2 gene expression was analysed by real time RT–PCR. Data are presented as means ± s.d. (n = 3).

Figure 6 | Substrate recognition shared among POT family members.
(a) Close view of the substrate-binding site structures. Homology models for Ptr2p and hPEPT1 were constructed using the SWISS-MODEL programme (http://swissmodel.expasy.org/) with PepTso (PDB: 2XUT) as a template. Structures of Ptr2p and hPEPT1 were superimposed on the PepTso structure. The PepTso structure is shown as a ribbon model and the amino-acid residues that interact with a substrate peptide in PepTso are shown as a stick model. Blue: PepTso, yellow: Ptr2p, red: hPEPT1. Putative binding substrates are shown as pink and blue circles. (b) Dipeptide affinity for each transporter. Kᵢ values for hPEPT1 and hPEPT2 are from published data. ND: not determined (Kᵢ > 0.77 mM). Yellow column: Ptr2p, red column: hPEPT1, blue column: hPEPT2.
molecular interaction data together with this modelling concept has great potential for affinity predictions of dipeptide-like molecules with high accuracy. Compared to the transported prediction approaches demonstrated by Biegel et al., our affinity prediction approach is sufficient for affinity predictions using only simple and feasibly calculated parameters for molecules without any prior ligand structural information.

The model construction process also guided our understanding of dipeptide interactions with Ptr2p. In M1 models (Table 1), index 3 (side-chain contribution to protein stability) at the N-terminal, index 14 (side-chain interaction parameter) and indices 8 and 1 (isoelectric point) were reproducibly selected in the same order in two models using data set type A or B. Index 3 and index 14 are strongly related to residues with ‘aromatic rings’ and ‘interactive residues’ that can stabilize molecular interactions (Supplementary Table S2). Indices 8 and 1 and some parts of index 14 can be interpreted as ‘effect of charged and polarized residues’. Therefore, these two physicochemical property features can be considered as the main determining factors for Ptr2p ligand recognition. Several substrate characteristics for binding have been elucidated based on the interaction data for many ligands with POT family transporters. For PEPT1, binding models were proposed by Foley et al., Brandsch et al., and Daniel and Kottra. According to these models, the properties of the individual residues at both the N- and C-termini are important. While the interactions between a peptide backbone, side-chains and the binding pocket are not completely understood, all of these models indicate that a bulky, hydrophobic side chain is advantageous for high affinity towards PEPT1. On comparing, acidic amino acids in the N-terminus resulted in a greater reduction in affinity than did the same amino acids in the C-terminus. These models for PEPT1 are consistent with our results based on an in silico analysis of Ptr2p in nature. Therefore, from a general point of view, this also supports the hypothesis that POT family members share a common substrate recognition mechanism.

Most organisms, including yeasts and humans, have both peptide transporters and amino-acid transporters and both types of transporters cooperatively contribute to amino-acid resource uptake. The rates of substrate uptake by peptide transporters are higher than those of amino-acid transporters. Therefore, an important role of peptide transporters is to import amino acids that are found in bulk in the extracellular fluid with high efficiency. To perform this role, peptide transporters must have substrate multispecificity in order to recognize a variety of compounds. Ligand recognition by hydrophobic interactions as well as by $\pi-\pi$ bonds, which are not strictly directional, is suitable for this purpose. It is likely that POT family proteins have evolved so that these transporters have become equipped with such a substrate recognition mechanism (Table 1). The fact that many of the amino-acid residues that are involved in substrate binding are aromatic amino acids supports these characteristics (Fig. 6). An ‘ambiguous’ substrate recognition mechanism, which is primarily based on physicochemical properties with no strict directionality, is the basis for substrate multispecificity and causes POT family members to act as drug transporters to absorb drugs, which are not natural substrates. This ‘ambiguous’ substrate recognition mechanism was also observed in our regression analysis. The concept of affinity prediction for POT family members using combinations of amino-acid indices has a great potential to be extended to other targets for predicting the affinities of as many as 8,000 tripeptides. The concept of using combinations of physicochemical properties for affinity prediction can also be applied to the information of structural properties used in structure-based drug design, which is becoming possible owing to the elucidation of the crystal structures of POT family members. Therefore, the analytical data from this research provides important information for detailed pharmacophore mapping.

**Methods**

**Materials and chemicals.** β-Ala–Lys (AMCA) was purchased from Biotrend (Cologne, Germany). Substrate dipeptides were purchased from AnaSpec (California, USA). The S. cerevisiae strains BY4742 (MATa, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0) and BY4742-ptr2Δ were purchased from Open Biosystems (Alabama, USA).

**Preparation of Ptr2p-expressing cells (SC-Pt2).** PCR was used to isolate the PTR2 gene from the genome of S. cerevisiae FG217 (MATa, ura3-52, lys2A201,
pep4). The gene-specific primers 5’-AACCCGGATTCTAGAACTAGTGGATCC
CCTAGTCTGCACATTCCACGAG-3’ and 5’-AAATGACCTGTTGAAATATAT
AAATTTCCCCACATTCCGTGCTGTCGTCGTAATTATTTTGTTG
GGATCTTAGAC-3’ were used to obtain PCR fragments of the PTR2 gene. These
primers contained gene-specific regions (bold) and homologous regions (Italic). A
pep4
NATURE COMMUNICATIONS | DOI: 10.1038/ncomms3502 ARTICLE
synthetic drop-out medium without uracil, 0.67% yeast nitrogen base without
nitrogen base without amino acids and 2% glucose) at 30°C was grown overnight in culture medium (0.2% yeast synthetic medium, 0.67% yeast
was grown in culture medium (0.2% yeast synthetic medium, 0.67% yeast
N-terminal or the C-terminal side of dipeptides using amino-acid indices (AA
indices, which were calculated by converting each amino acid at either the
three types of Kc samples (high, medium and low); and type B data sets to construct prediction models to discriminate either low Kc samples or high Kc samples. For both prediction concepts, three different prediction models were constructed: M1 prediction models using only type I parameters; M2 prediction models using type I and type II parameters and M3 prediction models using only type II parameters for comparing model performances between the different types of parameters. During the model construction process, a parameter increasing and decreasing method was employed with a threshold of P<0.20 based on F-Test for parameter selection. Model accuracy was evaluated based on the discrimination accuracy of each teaching signal category or as a total.

Confocal microscopy. To express thePtr2–GFP fusion protein, 5’-AAATGACCTGTTGAAATATAT
AAATTTCCCCACATTCCGTGCTGTCGTCGTAATTATTTTGTTG
GGATCTTAGAC-3’ was used as a reverse primer. Expression of thePtr2–GFP fusion protein used the same method as for Ptr2p. The localization of thePtr2p–GFP fusion protein was analysed by detecting GFP fluorescence using an LSM-700 (Carl Zeiss Micro-
Imaging, New York, USA).

Spot assay. SC-Ptr2p cells were grown in pre-culture medium at 30°C for 24 h. Cells were spotted on an assay plate (10 mM His-Leu dipeptide, 0.14% yeast synthetic drop-out medium without histidine, leucine and uracil, 0.67% yeast nitrogen base without amino acids, 2% galactose and 2% agar) and cultured at 30°C for 5 days.

Fluorescence-based Competitive Uptake (F-CUp) assay. After induction, SC-
Ptr2p cells were harvested by centrifugation and re-suspended in F-CUp assay buffer (150 mM NaCl, 50 mM Na-phosphate buffer; pH 6.0). Cells were harvested by centrifugation and re-suspended in F-CUp assay buffer to OD660 = 15. Cells were incubated with 0.05 mM β-Ala–Lys (AMCA) as a tracer substrate and at arbitrary concentration for oligopeptide analysis at 30°C for 60 min. β-Ala–Lys (AMCA) is a dipeptide containing a fluorescent dye. Cells were washed three times with the F-CUp assay buffer. β-Ala–Lys (AMCA) uptake was quantified by whole-cell fluorescence (excitation at 355 nm and emission at 460 nm) using Flexstation III (Molecular Devices, California, USA). Km values were estimated using Lineweaver-Burk plot. The initial velocity of a reaction was determined from the substrate concentration remained the same as the initial condition. The IC50 value was estimated based on the reduced fluorescence by competitive inhibition of the tracer uptake.

Cell growth analysis. For time course analysis of culture turbidity, strain BY4742 was grown overnight in culture medium (0.2% yeast synthetic medium, 0.67% yeast nitrogen base without amino acids and 2% glucose) at 30°C for 24 h. Pre-cultures were diluted to give OD660 = 0.06 in each test medium (10 mM each dipeptide, 0.14% yeast synthetic drop-out medium without leucine, 0.67% yeast nitrogen base without amino acids and 2% glucose) and incubated at 30°C. Culture turbidity was monitored by measuring OD660 using a Biophotorecorder TVS 062CA (Advantec, Tokyo, Japan).

Ligand affinity prediction models and discrimination analysis. Using the data set of dipeptide sequences in conjunction with their affinity data from the F-CUp assay, dipeptide affinity prediction models were constructed using discrimination analysis with PASW version 18, release 18.0.0 (IBM Corporation, Armonk, NY, USA). For objective variables (that is, teaching signals), experimentally determined Kc values were grouped into three categories: low Kc samples with Kc<0.1, N = 102; high Kc samples with Kc>0.77, N = 100; and medium Kc for the remaining samples, N = 1395. For predictor variables (that is, input parameters), dipeptide sequences were converted into two types of parameters. Type I parameters included seven amino-acid indices, which were calculated by converting each amino acid at either the N-terminal or the C-terminal side of dipeptides using amino-acid indices (AA index1, Genome Net Japan, organized by Kyoto University; http://www.genome.jp/dbget-bin/www_bfdn?iaindex1)44. All indices in the database (N = 544, version 9.1, as of August 2006) were previously analysed by hierarchical clustering; seven

Real time RT-PCR. A FGY217 strain pre-culture was diluted to give OD660 = 0.06 in each test YPD medium (2% peptone, 1% yeast extract, 2% glucose and 10 mM Trp–Ala or Ala–Ala), and then grown at 30°C for 5 h. Yeast total RNA was isolated using a NucleoSpin RNA II kit (Machery-Nagel, Düren, Germany) according to the manufacturer’s protocol. The amount of total RNA was quantified by monitoring absorbance at 260 nm. The first strand was synthesized using a PrimeScript RT reagent kit (Takara, Shiga, Japan). Quantitative real-time PCR analysis was done with a Thermal Cycler Dice Real Time System (Takara, Shiga, Japan) using SYBR Premix EX Taq (Takara, Shiga, Japan) and specific primers. The following gene-
specific primers were used: 5’-CACCATGTCCTCAGGTATT-3’ and 5’-CAGAGCATGACACTT-3’ for ACT1; and 5’-GATGACCGCTTGACAT-3’ and 5’-CTGAGACCAGTTTCC-3’ for PTR2. PCR efficiency levels were normalized to those of ACT1 values using the 2-ΔΔCT method. We calculated the fold-change of PTR2 mRNA in YPD medium containing Trp–Ala compared to that containing Ala–Ala.

References
1. Daniel, H., Spanier, B., Kottra, G. & Weitz, D. From bacteria to man: archaic proton-dependent peptide transporters at work. Physiology (Bethesda) 21, 93–102 (2006).
2. Gomolaptitant, K. M. & Saier, Jr. M. H. Evolution of the oligopeptide cotransporter family, from human kidney. Biochim. Biophys. Acta. 176, 628–631 (2012).
3. Steiner, H. Y., Naider, F. & Becker, J. M. The PTR family: a new group of peptide transporters. Mol. Microbiol. 16, 825–834 (1995).
4. Maebuchi, M. et al. Improvement in the intestinal absorption of soy protein by enzymic digestion to oligopeptide in healthy adult men. Food Sci. Technol. Res 13, 45–53 (2007).
5. Matthews, D. M. Intestinal absorption of peptides. Physiol. Rev. 55, 537–608 (1975).
6. Ito, K. et al. Soy peptides enhance heterologous membrane protein productivity during the exponential growth phase of Saccharomyces cerevisiae. Biosci. Biotechnol. Biochem. 76, 628–631 (2012).
7. Biegelsen, A. et al. The renal type H+ pep transport system PEP2: structure-affinity relationships. Am. J. Physiol. 31, 137–156 (2006).
8. Fei, Y. J. et al. Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature 368, 563–566 (1994).
9. Liang, R. et al. Human intestinal H+–peptide cotransporter. Cloning, functional expression, and chromosomal localization. J. Biol. Chem. 270, 6456–6463 (1995).
10. Abidi, S. A. The oligopeptide transporter (Pept-1) in human intestine: biology and function. Gastroenterology 113, 332–340 (1997).
11. Liu, W. et al. Molecular cloning of Pept-2, a new member of the H+–peptide cotransporter family, from human kidney. Biochim. Biophys. Acta. 1235, 461–466 (1995).
12. Boll, M. et al. Expression cloning and functional characterization of the kidney cortex high-affinity proton-coupled peptide transporter. Proc. Natl Acad. Sci. USA 93, 284–289 (1996).
13. Shen, H. et al. Localization of Pept1 and Pept2 proton-coupled oligopeptide transporter mRNA and protein in rat kidney. Am. J. Physiol. 276, F658–F665 (1999).
14. Newsutel, S. et al. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, Pept1 and Pept2. EMBO J. 30, 417–426 (2011).
15. Solcan, N. et al. Alternating access mechanism in the POT family of mammalian oligopeptide proton symporters, PepT1 and PepT2. EMBO J. 31, 4311–4321 (2012).
16. Rubo-Alaina, I. & Daniel, H. Mammalian peptide transporters as targets for drug delivery. Trends Pharmacol. Sci. 23, 434–440 (2002).
17. Homann, O. R., Cai, H., Becker, J. M. & Lindquist, S. Harnessing natural diversity to probe metabolic pathways. PLoS Genet. 1, e80 (2005).
38. Zhang, L., Kaufman, S., Naidier, F. & Becker, J. M. Genomewide screen reveals a wide regulatory network for di/tripeptide utilization in \textit{Saccharomyces cerevisiae}. 
\textit{Mol. Cell. Biol.} 14, 104–115 (1994).

39. Hauser, M., Narita, V., Donhardt, A. M., Naidier, F. & Becker, J. M. Multiplicity and regulation of genes encoding peptide transporters in \textit{Saccharomyces cerevisiae}. 
\textit{Mol. Membr. Biol.} 18, 103–112 (2001).

40. Kaga, C., Okochi, M., Tomita, Y., Kato, R. & Honda, H. Computationally assisted screening and design of cell-interactive peptides by a cell-based assay using peptide arrays and a fuzzy neural network algorithm. 
\textit{Biotechniques} 44, 393–402 (2008).

41. Cai, H., Kauffman, S., Naider, F. & Becker, J. M. Advanced method for high-throughput expression of mutated eukaryotic membrane proteins in \textit{Saccharomyces cerevisiae}. 
\textit{Biochem. Biophys. Res. Commun.} 371, 841–845 (2008).

42. Drew, D. et al. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in \textit{Saccharomyces cerevisiae}. 
\textit{Nat. Protoc.} 3, 784–798 (2008).

43. Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in \textit{Saccharomyces cerevisiae}. 
\textit{Proc. Natl Acad. Sci. USA} 104, 13936–13941 (2007).

44. Kawashima, S. et al. AIndex: amino acid index database, progress report 2008. 
\textit{Nucleic Acids Res.} 36, D202–D205 (2008).

45. Zimmermann, J. M., Eleezer, N. & Simha, R. The characterization of amino acid sequences in proteins by statistical methods. 
\textit{J. Theor. Biol.} 21, 170–201 (1968).

46. Fauchere, J. L., Charton, M., Kier, L. B., Verloop, A. & Pliska, V. Amino acid side chain parameters for correlation studies in biology and pharmacology. 
\textit{Int. J. Pept. Protein Res.} 32, 269–278 (1988).

47. Takano, K. & Yutani, K. A new scale for side-chain contribution to protein stability based on the empirical stability analysis of mutant proteins. 
\textit{Protein. Eng.} 14, 525–528 (2001).

48. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. 
\textit{J. Mol. Biol.} 157, 105–152 (1982).

49. Ito, K. et al. L. Lipscomb, W. N. & Schellman, C. G. The reverse turn as a polypeptide conformation in globular proteins. 
\textit{Proc. Natl Acad. Sci. USA} 70, 538–542 (1973).

50. Krigbaum, W. R. & Komoriya, A. Local interactions as a structure determinant for protein molecules: II. 
\textit{Biochim. Biophys. Acta} 576, 204–248 (1979).

Acknowledgements
This study was supported in part by a Research Grant from the Society for Research on Umami Taste, a Grand-in-Aid for Young Scientists (B) (2370139) and the Fuji Foundation for Protein Research. The authors would like to thank Enago (www.enago.jp) for the English language review.

Author contributions
K.I. established the F-CUp assay system, designed the project, organized the entire research and wrote the manuscript. A.H. performed the exhaustive analysis of dipeptide library analysis using the F-CUp assay system and cell growth analysis. V.T.T.L. and S.K. (Shun Kawai) established the F-CUp assay system, designed the project, organized the entire analysis. Y.K. wrote the manuscript and co-organized the project.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Ito, K. et al. Analysing the substrate multiplicity of a proton-coupled oligopeptide transporter using a dipeptide library. 
\textit{Nat. Commun.} 4:2502 doi: 10.1038/ncomms3502 (2013).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/