that achieves the best performance, i.e., the highest correlation coefficients (results not shown for other peak lengths).

Number of motif candidates. In the first step of PeakRegressor, we use MEME to find over-represented DNA motifs in the peak sequences. This step results in 800 motif candidates for STAT1 and 880 for RNA Polymerase II. Given the large number of motif candidates, we empirically observe the presence of similar motifs in the set of motif candidates. We may wonder if this redundancy could affect the prediction performance of PeakRegressor. However, we show that this is not the case.

PeakRegressor uses a regression method called L1-norm log linear regression. In contrast with other regression methods, L1-norm log linear regression achieves its best prediction performance by removing redundant or uninformative motifs from the regression model. Therefore, the removal of redundant motifs is automatically performed when using L1-norm log linear regression. Table 2 shows the set of motifs that achieve the best correlation coefficient for STAT1. We can see that some motifs are similar. For example, the motifs A[CT]**TTC[TC][TG]GGAA, TT[CA]C[TAG][GA]GAA** [GA]T, A[TA]**TTCC[CT][GA]GAA**[AC]T[CG][AC], and **TT-[CA][TC][GA]GGAA**[AG] are short, similar motifs containing the STAT1 binding motif. In other experiments, we find that the prediction performance worsens when similar motifs are removed (results not shown). Hence, although the motifs appear similar and redundant, they actually contain complementary information for the prediction performance.

Moreover, the motif weights computed by PeakRegressor are all different (resp. 0.56, 0.55, 0.48, and 0.47). Hence, while other approaches, such as motif clustering, would consider all these motifs to be equally important, PeakRegressor is able to detect the relative importance of each motif and compute the corresponding weight. This is explained by the noisy nature of the DNA motifs found by MEME in step 1. For a given binding motif, PeakRegressor needs to use all the noisy PSSM approximations to achieve the best prediction performance. This is an important property of PeakRegressor, especially when the number of noisy motifs is very large.

Candidate motifs and their potential rSNPs

Single or composite motifs found in the PeakRegressor system may reflect actual transcription factor binding sites. If a single nucleotide polymorphism (SNP) occurs within the sites, regulatory control of neighboring gene transcription will be perturbed, thus leading to genetic diseases in some cases [10]. Therefore, true binding sites may have SNPs less frequently than the non-binding sites. As an important verification, we check the number of known SNPs to be found within the STAT1 positions presented by PeakRegressor by using dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/). We find that 0.36% (147 for 40,395 bp) of mapped positions with 10 STAT1 motifs in Table 2 on the peak sequences contains SNPs, while as much as 0.53% (17,852 for 3,344,439 bp) of all positions contains SNPs on the peak sequences. The statistical difference between the above two ratios is highly significant such as $p<3.7^{-7}$ according to the hypergeometric distribution. These sites are possible candidates of rSNPs because the slight change within the motif may affect the change of gene expression level and might cause diseases.

Materials and Methods

PeakRegressor

PeakRegressor is a system to find TFBMs that are statistically important for transcription factor binding signals, by taking ChIP-Seq data as input, and outputs a list of TFBM candidates. In contrast with previous approaches, PeakRegressor uses the peak scores (provided by [9]) as a surrogate for the binding affinities. We argue that the peak scores provide more accurate approximations of the binding affinities than the methods based on transcription factor occupancy [4,5]. Therefore, using the peak scores lead to better identification of functional TFBMs. In addition, PeakRegressor identifies not only primary TFBM candidates but also secondary motifs that may often synergistically strengthen or weaken the binding. The workflow is summarized in Figure 2.

Step 1. First, we define the peak sequences as the 200-bp genomic regions centered around the peaks. Then, we sort the peak sequences according to their ascending scores. We group the peak sequences into clusters such that each cluster contains 200 peaks of consecutive scores. Then, we apply MEME (http://meme.sdsc.edu/) to each peak sequence cluster. For each sequence cluster, MEME is parameterized in ZOOPS mode to find 10 motifs of lengths $8-20$.

This strategy has two advantages. First, it allows us to identify motifs that may be associated with a given binding affinity level. If a cluster contains only low (resp. high) binding affinity peaks, the corresponding sequences may contain weak (resp. strong) binding motifs, i.e., motifs that are specific to low (resp. high) binding affinity. Second, it reduces computational time by parallelizing MEME computations.

Step 2. In order to predict the binding affinity of the peaks, we need to represent each peak as a vector in the motif space. Let seq^i be the DNA sequence of peak i. Let $\text{seq}^i_{j,\ell}$ be the ℓ-length sub-sequence of seq^i, starting from position j. Let S^d be the PSSM of motif d. Let ℓ_i be the length of seq^i and ℓ_d be the length of motif d. We represent peak i as vector $x_i \in R^D$, such that

$$x_{id} = \max_{j=1\ldots\ell_i-\ell_d+1} f(\text{seq}^i_{j,\ell_d}, S^d) - \max(S^d)$$

for $d=1\ldots D$. The quantity $f(\text{seq}^i_{j,\ell_d}, S^d)$ is a sum of log-odd scores, representing how well motif d matches sub-sequence seq^i_{j,ℓ_d}. Hence, the first term of the sum, x_{id}, corresponds to the best match when we slide motif d along sequence seq^i. The term $\max(S^d)$ is the maximum score achievable by any sequence matching with the motif d. Therefore, we always have $x_{id} \leq 0$, with $x_{id}=0$ for the best possible match.

Next, we want all the x_{id} to be positive for interpretability purpose. So we simply shift their values by substracting the lowest component: $x_{id} \leftarrow x_{id} - a$, where a is the minimum value of the original x_{id}. Finally, we normalize each data vector by dividing it with its euclidean norm: $x_i \leftarrow x_i / \|x_i\|^2$.

Step 3. Quantities y_i to be fitted are the log values of the peak enrichment scores, as given by PeakSeq [9]. We can now solve the regression problem defined by (x_i, y_i) pairs for $i=1\ldots N$. Linear regression is a simple and popular approach, but is prone to overfitting. Hence, we choose to regularize the model with L1-norm, i.e., we want to minimize the sum of squared errors and the L1-norm of the regression coefficient vector:

$$\min_{b \in R^D} \beta\|b\| + \sum_{i=1}^{N} (b^T x_i - y_i)^2 \qquad (1)$$

where $\beta>0$ is a user-defined regularization coefficient. The L1-norm log linear regression is able to remove redundant or uninformative features, and to select a small number of features that best explain the fitted quantity [11]. In our case, the features correspond to DNA motifs and hence, the result of this step is a set

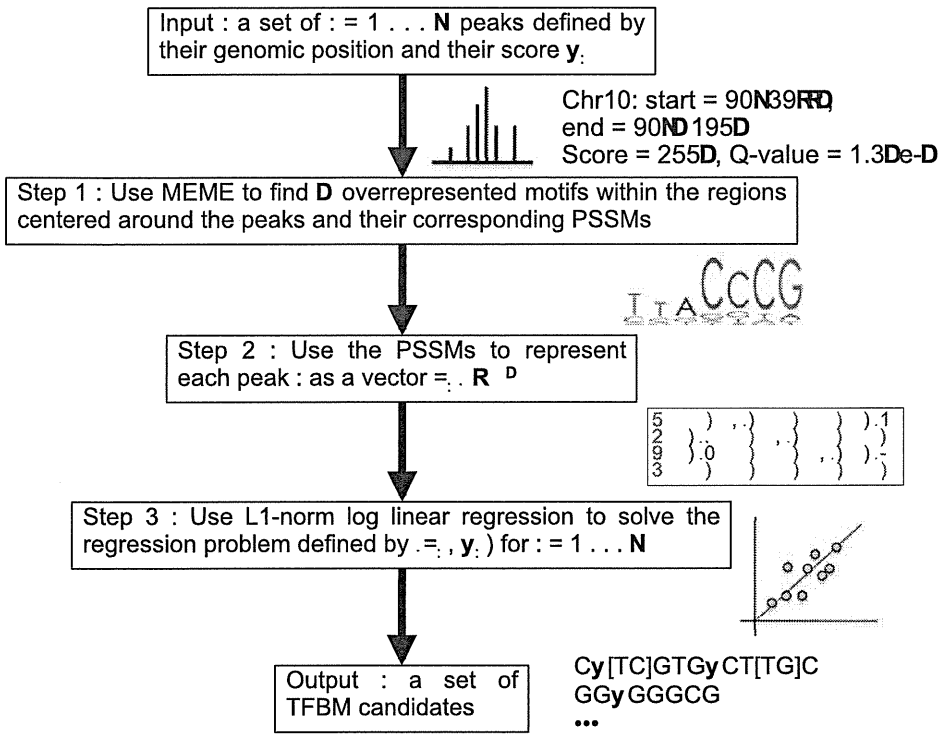


Figure 2. Schematic view of the workflow of PeakRegressor. PeakRegressor takes ChIP-Seq data as input and outputs a list of TFBM candidates and their weights that give the best regression accuracies.
doi:10.1371/journal.pone.0011881.g002

of motifs that best explain the binding signal values from ChIP-Seq dataset. We use Lasso, a popular algorithm for solving L1-norm log linear regression. Lasso is available as part of the LARS package for R (http://www-stat.stanford.edu/~hastie/Papers/LARS/).

Other regression methods

In this section, we present alternatives to the L1-norm log linear regression: linear least squares regression, ridge regression, partial least squares regression, and principal component regression. All these regression methods are used in the following way. Once a regression model is fitted to the peak dataset, we rank the regression coefficients with respect to their absolute values. Using this ranking, the top motifs are the potential TFBMs.

Linear least squares regression. The linear least squares regression is the simplest regression approach. It fits a linear model to the dataset by minimizing the sum of squared errors $\sum_{i=1}^{N}(y_i - b^T x_i)$. Its difference with the L1-norm log linear regression (equation 1) is the absence of a regularization term. Therefore, the linear least squares regression is more prone to overfitting when the regression problem contains more dimensions than samples.

Ridge regression. The ridge regression [12] minimizes $\|b\|^2 + \sum_{i=1}^{N}(y_i - b^T x_i)$, where the regularization term is $\|b\|^2 = \sum_{d=1}^{D} b_d^2$, i.e., the Euclidean norm of b. It is quite similar to the L1-norm log linear regression, and their main difference lies in the regularization term. The ridge regression seeks a solution with a low Euclidean norm. Although the Euclidean norm is a protection against overfitting, it does not favor sparse solutions (i.e., solutions with many motifs) as the L1-norm log linear regression does [11].

Partial least squares regression and principal component regression. The partial least squares regression [13] and the principal component regression are two approaches of the same idea; they perform linear regression using the low-dimensional data matrix Z instead of the initial data matrix X. This approach avoids overfitting problems. Therefore, the partial least squares regression and the principal component regression have been widely used in problems containing several dimensions (i.e., motifs) and few samples (i.e., peaks).

In the principal component regression, the low-dimensional data matrix Z contains the most information about the initial data matrix X (according to the singular value decomposition of X). In the partial least squares regression, the low-dimensional data matrix Z is calculated using both the initial data matrix X and the peak score vector y. In both cases, linear regression is performed using Z instead of the initial data matrix X. Both partial least squares regression and principal component regression are available as part of the PLS package for R (http://mevik.net/work/software/pls.html). Once the regression coefficients have been computed in the low-dimensional space, they are mapped back in the original motif space. Then, these coefficients can be used to identify potential binding motifs.

Input ChIP-Seq datasets

The ChIP-Seq dataset we used is provided by [9] and is publicly available (http://www.camda2009.org/). The dataset provides various information about each peak, including the peak score, the peak center (for STAT1), and the Q-value that reflects the significance of the peak. The Q-values are derived from the P-values. First, they compute the P-values that reflect the significance of peak enrichment in the number of DNA tags, compared to control samples. These P-values are computed using the binomial distribution. Then, to account for multiple hypothesis testing, the Q-values are derived from the P-values. See [9] for more details.

For STAT1, we use 200-bp windows around the peak centers to define the peak sequences. For RNA Polymerase II, the peak centers are not available and thus, we use the peak start and peak end coordinates to define the peaks. When the length of the resulting sequence is less than 200 bp, we enlarge it in both directions in order to reach 200 bp length. When the length is more than 4000 bp, we trim it in both directions in order to reach 4000 bp length. As a result, all the RNA Polymerase II peak sequence lengths lie between 200 and 4000 bp.

Evaluation of prediction performance

PeakRegressor predicts the peak scores and therefore, we have two different values for each peak. The "true" peak score is the score provided by [9], and is derived from the frequency of reads of ChIP-Seq data. The predicted score is computed by PeakRegressor using the peak sequence information. Ideally, the predicted score should be equal to the true score. We use correlation coefficients to evaluate the prediction quality of PeakRegressor.

Experimental protocol

For L1-norm log linear regression and ridge regression, we have to set the regularization parameter β. First, we define $\beta = 2^i$ for $i \in [-25, 25]$. Then for each value of β, we perform a 30-fold cross-validation. In each fold, we split the dataset into a training set and a test set, with a $90\% - 10\%$ ratio. The optimal value for β is the one which corresponds to the lowest prediction error on the test set. All the results of L1-norm log linear regression and ridge regression are averaged over the 30-fold cross-validation.

For partial least squares regression and principal component regression, the experiments were limited by the slowness of both methods. First we have to set the number of components K used for regression. We tried $K = 1...10$, and performed a 30-fold cross-validation for each value of K. In each fold, we split the dataset into 50% for training and 50% for testing. All the results of partial least squares regression and principal component regression are averaged over the 30-fold cross-validation.

Acknowledgments

The authors thank the anonymous CAMDA reviewers for their helpful comments.

Author Contributions

Conceived and designed the experiments: JFP WF. Performed the experiments: JFP HH TT. Analyzed the data: JFP HC WF. Wrote the paper: JFP HC WF.

References

1. Bussemaker HJ, Li H, Siggia ED (2001) Regulatory element detection using correlation with expression. In: RECOMB '01: Proceedings of the fifth annual international conference on Computational biology. New York, NY, USA: ACM. 86 p. doi:http://doi.acm.org/10.1145/369133.369174.
2. Conlon EM, Liu XS, Lieb JD, Liu JS (2003) Integrating regulatory motif discovery and genome-wide expression analysis. PNAS.
3. Das D, Pellegrini M, Gray JW (2009) A primer on regression methods for decoding cis-regulatory logic. PLoS Comput Biol 5: e1000269.
4. Foat BC, Morozov AV, Bussemaker HJ (2006) Statistical mechanical modeling of genome-wide transcription factor occupancy data by matrixreduce. Bioinformatics 22: e141–e149.
5. Gao F, Foat BC, Bussemaker HJ (2004) Defining transcriptional networks through integrative modeling of mrna expression and transcription factor binding data. BMC Bioinformatics.
6. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, et al. (2007) Genome-wide profiles of stat1 dna association using chromatin immunoprecipitation and massively parallel sequencing. Nat Meth 4: 651–657.
7. Butler JE, Kadonaga JT (2002) The rna polymerase ii core promoter: a key component in the regulation of gene expression. Genes Dev 16: 2583–2592.
8. Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. The Annals of Statistics.
9. Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, et al. (2009) Peakseq enables systematic scoring of chip-seq experiments relative to controls. Nat Biotech 27: 66–75.
10. Ameur A, Rada-Iglesias A, Komorowski J, Wadelius C (2009) Identification of candidate regulatory snps by combination of transcription-factor-binding site prediction, snp genotyping and haplochip. Nucleic acids research 37.
11. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Statist Soc Ser B 58: 267–288.
12. Bishop CM (2006) Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc.
13. Frank IE, Friedman JH (1993) A statistical view of some chemometrics regression tools. Technometrics.

Anal. Chem. **2010**, *82*, 5768–5776

Sensitive and Convenient Yeast Reporter Assay for High-Throughput Analysis by Using a Secretory Luciferase from *Cypridina noctiluca*

Yuki Tochigi,[†] Natsuko Sato,[†] Takehiko Sahara,[†] Chun Wu,[‡] Shinya Saito,[§] Tsutomu Irie,[§] Wataru Fujibuchi,[∥] Takako Goda,[†] Ryoichi Yamaji,[○] Masahiro Ogawa,[○] Yoshihiro Ohmiya,[◆,⊥] and Satoru Ohgiya*[,†,▽]

Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-17-2-1 Tsukisamu-Higashi, Toyohira-ku, Sapporo 062-8517, Japan, Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba 305-8566, Japan, Health Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-8-31 Midorigaoka, Ikeda 563-8577, Japan, ATTO Corporation, 1-5-32 Yushima, Bunkyo-ku, Tokyo 113-0034, Japan, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan, and Laboratory for Nutrition Chemistry, Division of Applied Life Sciences, Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

The yeast reporter assay has been widely used in various applications such as detection of endocrine disruptors and analysis of protein−protein interactions by the yeast two-hybrid system. The molecular characteristics of the reporter enzyme are critical determinants for this assay. We herein report the establishment of a novel yeast reporter assay using a secretory luciferase, *Cypridina noctiluca* luciferase (CLuc), as an alternative to the conventional β-galactosidase. The CLuc reporter assay in yeast is more sensitive and convenient than the conventional assay. A yeast high-throughput reporter assay was established with a laboratory automation system, and the transcriptional activity of hundreds of yeast promoter fragments was comprehensively determined. Our results indicate that the yeast CLuc reporter assay is a promising tool for large-scale and sensitive analysis in the development of new drugs and in various fields of biotechnology research.

The yeast reporter assay has been widely used in various applications, including determination of promoter activities,[1] identification of *cis*-elements,[2−4] analysis of chemicals (i.e., yeast estrogen screen (YES) assay),[5,6] and detection of protein−protein interactions (two-hybrid system).[7] Relative to mammalian cultured cell-based reporter assays, the yeast reporter assay is suitable for analyzing the effects of chemicals and for the two-hybrid system because of its high performance and low cost for large-scale screening. For the yeast reporter system, the concentrations of chemicals or the binding affinities of interacting proteins are determined by the amount of protein or the activity of reporter enzyme produced. The gene for the reporter enzyme is located downstream of a promoter specifically designed for the phenomena. Therefore, the characteristics of the reporter enzyme are critical for the performance of the reporter assay.

The *Escherichia coli* enzyme β-galactosidase (β-Gal) is the most popular reporter enzyme used in the yeast reporter assays.[8] However, the β-Gal assay is not highly sensitive because its enzymatic activity is usually determined by a colorimetric method. This low sensitivity also prevents reduction of the scale of the assay. Furthermore, the procedure comprises multiple laborious steps because the enzyme is intracellularly produced. Yeast cells in liquid culture must first be harvested and disrupted with one or a combination of cell wall-degrading enzymes, beads, and/or detergent-containing reagents to expose the intracellular protein for enzymatic assay. The solution must then be centrifuged again to remove cell debris. Consequently, the β-Gal reporter assay is rarely applied for high-throughput analysis. Other reporter pro-

* To whom correspondence should be addressed. E-mail: s.ohigya@aist.go.jp.
[†] Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 2-17-2-1 Tsukisamu-Higashi, Toyohira-ku, Sapporo 062-8517, Japan.
[‡] Health Research Institute.
[§] ATTO Corporation.
[∥] Computational Biology Research Center.
[⊥] Graduate School of Medicine, Hokkaido University.
[▽] Graduate School of Life Sciences, Hokkaido University.
[○] Osaka Prefecture University.
[◆] Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba 305-8566, Japan.

(1) Funk, M.; Neidenthal, R.; Mumberg, D.; Brinkmann, K.; Rönicke, V.; Henkel, T. *Methods Enzymol.* **2002**, *350*, 248–257.
(2) Kuroda, S.; Otaka, S.; Fujisawa, Y. *J. Biol. Chem.* **1994**, *269*, 6153–6162.
(3) Hiraishi, H.; Miyake, T.; Ono, B. *Curr. Genet.* **2008**, *53*, 225–234.
(4) Schmitt, M.; Schwanewilm, P.; Ludwig, J.; Lichitenberg-Fraté, H. *Appl. Environ. Microbiol.* **2006**, *72*, 1515–1522.
(5) Routledge, E. J.; Sumpter, J. P. *Environ. Toxicol. Chem.* **1996**, *15*, 241–248.
(6) Bovee, T. F.; Helsdingen, R. J.; Koks, P. D.; Kuiper, H. A.; Hoogenboom, R. L.; Keijer, J. *Gene* **2004**, *21*, 187–200.
(7) Fields, S.; Song, O. *Nature* **1989**, *340*, 245–246.
(8) Rupp, S. *Methods Enzymol.* **2002**, *350*, 112–131.

10.1021/ac100832b
Published on Web 05/28/2010

teins such as yeast enhanced green fluorescent protein (yEGFP),[9,10] firefly and *Renilla* luciferases,[11] PHO5 (acid phosphatase),[12] and MEL1 (α-galactosidase)[13] have been proposed but have not gained wide acceptance thus far. The demand for compatibility with high-throughput analysis as well as high sensitivity is increasing in current applications of the yeast reporter assay. However, conventional yeast reporter systems are not fully compatible with highly sensitive high-throughput assays. A novel reporter enzyme with outstanding characteristics will have a decisive impact on various applications.

The Japanese ostracod, *Cypridina noctiluca*, secretes luciferase (CLuc) and *Cypridina* luciferin. CLuc catalyzes oxidation of *Cypridina* luciferin resulting in emission of blue bioluminescence.[14] Nakajima et al.[15] cloned CLuc cDNA and developed a CLuc reporter assay in mammalian cell lines. In addition, Kanjou et al.[16] used CLuc for the screening of yeast secretion mutants. We have developed a new and effective synthetic method of *Cypridina* luciferin to provide the substrate for CLuc assay.[17] These luciferase assays are quite convenient because luminescence can be measured simply by obtaining aliquots of the culture medium.

We herein report the establishment of a quantitative CLuc reporter assay in yeast, which provides results consistent with those obtained by the conventional assay and the advantageous features of a secreted protein with luminescence. The CLuc reporter assay overcomes the limitations of existing bioassay systems and the conventional yeast two-hybrid system. A yeast high-throughput reporter assay was developed with an existing laboratory automation system and applied to comprehensive determination of the basal transcriptional activity of hundreds of yeast promoter fragments.

MATERIALS AND METHODS

Reagents. Restriction endonucleases, other DNA-modifying enzymes, and reagent kits were obtained from Takara Bio (Kyoto, Japan), Toyobo (Osaka, Japan), New England Biolabs (Beverly, MA), and Roche (Mannheim, Germany). Yeast nitrogen base without amino acids and ammonium sulfate, and other ingredients of the culture media were purchased from Becton, Dickinson and Company (Sparks, MD). *Saccharomyces cerevisiae* strain YPH500 (MATα, *ura3*, *lys2*, *ade2*, *trp1*, *his3*, *leu2*) was purchased from Invitrogen (Carlsbad, CA). All other reagents were of the highest available grade. Synthetic *Cypridina* luciferin was dissolved in 5 mM HCl in ethanol and stored at −80 °C until use. The lyophilized *Cypridina* luciferin and its stock solution in acidic ethanol were stable at −80 °C for more than 1 year and for at least 1 month, respectively.

Construction of mCLuc DNA and Reporter Plasmids. In our primary experiments, we used a variant of native CLuc cDNA (accession number AB262361) similar to the reported cDNA[15] (accession number AB159608), and a chimeric αCLuc cDNA[16] in which the native signal sequence (from the first amino acid to the 18th amino acid) was replaced with yeast α-factor prepro-signal sequence (from the first amino acid to the 89th amino acid)[18] for study of basic assay conditions. We then designed an appropriate DNA sequence for expression of the reporter protein in *S. cerevisiae*. First, the amino acid sequence of native CLuc (accession number AB262361) was reverse-translated using optimal codons for *S. cerevisiae*, according to Akashi,[19] to generate a new CLuc-encoding DNA sequence composed of optimal codons for *S. cerevisiae*. We then searched for putative *cis*-elements in the codon-optimized CLuc-encoding DNA sequence by using the yeast promoter database (SCPD; http://rulai.cshl.org/dbsd/index.html) (Figure 1a). When a *cis*-element was found in the DNA sequence, one or more bases in the *cis*-element were replaced with other nucleotides to remove the putative *cis*-element without changing the amino acids encoded. This process was used for optimal removal of putative *cis*-elements throughout the sequence. In the following step, the second database (Gene2 Promoter, provided by Genomatix, http://www.genomatix.de/index.html) was employed. Other putative *cis*-elements were identified in the SCPD-processed DNA sequence by using the second database and were similarly removed. The resultant DNA sequence was further subjected to another cycle of putative *cis*-element screening using the SCPD and Gene2 promoter databases. The cycle was repeated until none of the putative *cis*-elements in the processed DNA sequence could be removed without changing the encoded amino acids. The final CLuc-encoding DNA is referred to as mCLuc DNA (accession number AB259056) (see Figure 1a and b); mCLuc DNA encodes the native CLuc protein.

To construct the entire mCLuc DNA, 29 oligonucleotides were synthesized (see Supporting Information (SI) Table S-1) and used to generate the complete sequence of mCLuc DNA by repetitive PCR. The nucleotide sequence of the resultant PCR product was confirmed. A reporter plasmid was generated as a single copy plasmid derived from pUG35 (generously provided by Prof. J. H. Hegemann; http://mips.gsf.de/proj/yeast/info/tools/hegemann/gfp.html). The reporter plasmid, named pCLY, includes a centromere origin (*CEN6/ARSH4*), a *URA3* marker, two sets of SV40 poly(A) terminator, a multicloning site, mCLuc DNA, and CYC1 terminator (Figure 1c). The reporter plasmid pCLY and synthetic *Cypridina* luciferin are now available from ATTO (Tokyo, Japan).

Preparation of Yeast Promoters and Reporter Plasmids. In the experiments for comparison of the five constitutive promoters and in the experiment for the two-hybrid system, we defined a promoter as a sequence located between the initiation codon of the gene and the proximal end of the open reading frame of the adjacent upstream gene in the yeast genome.

A *TDH3* (*YGR192C*) gene promoter (approximately 650 bp), a region between the *TDH3* and *PDX1* genes (the adjacent

(9) Niedenthal, R. K.; Riles, L.; Johnston, M.; Hegemann, J. H. *Yeast* **1996**, *12*, 773–786.
(10) Li, J.; Wang, S.; VanDusen, W. J.; Schultz, L. D.; George, H. A.; Herber, W. K.; Chae, H. J.; Bentley, W. E.; Rao, G. *Biotechnol. Bioeng.* **2000**, *70*, 187–196.
(11) McNabb, D. S.; Reed, R.; Marciniak, R. A. *Eukaryotic Cell* **2005**, *4*, 1539–1549.
(12) Harashima, S.; Kaneko, Y. *J. Biosci. Bioeng.* **2001**, *91*, 325–338.
(13) Aho, S.; Arffman, A.; Pummi, T.; Uitto, J. *Anal. Biochem.* **1997**, *253*, 270–272.
(14) Kishi, T.; Goto, T.; Hirata, Y.; Shimomura, O.; Johnson, F. H. *Tetrahedron Lett.* **1966**, *7*, 3427–3436.
(15) Nakajima, Y.; Kobayashi, K.; Yamagishi, K.; Enomoto, T.; Ohmiya, Y. *Biosci. Biotechnol. Biochem.* **2004**, *68*, 565–567.
(16) Kanjou, N.; Nagao, A.; Ohmiya, Y.; Ohgiya, S. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 429–434.
(17) Wu, C.; Kawasaki, K.; Ohgiya, S.; Ohmiya, Y. *Tetrahedron Lett.* **2006**, *47*, 753–756.
(18) Brake, A. J.; Julius, D. J.; Thorner, J. *Mol. Cell. Biol.* **1983**, *3*, 1440–1450.
(19) Akashi, H. *Genetics* **2003**, *164*, 1291–1303.

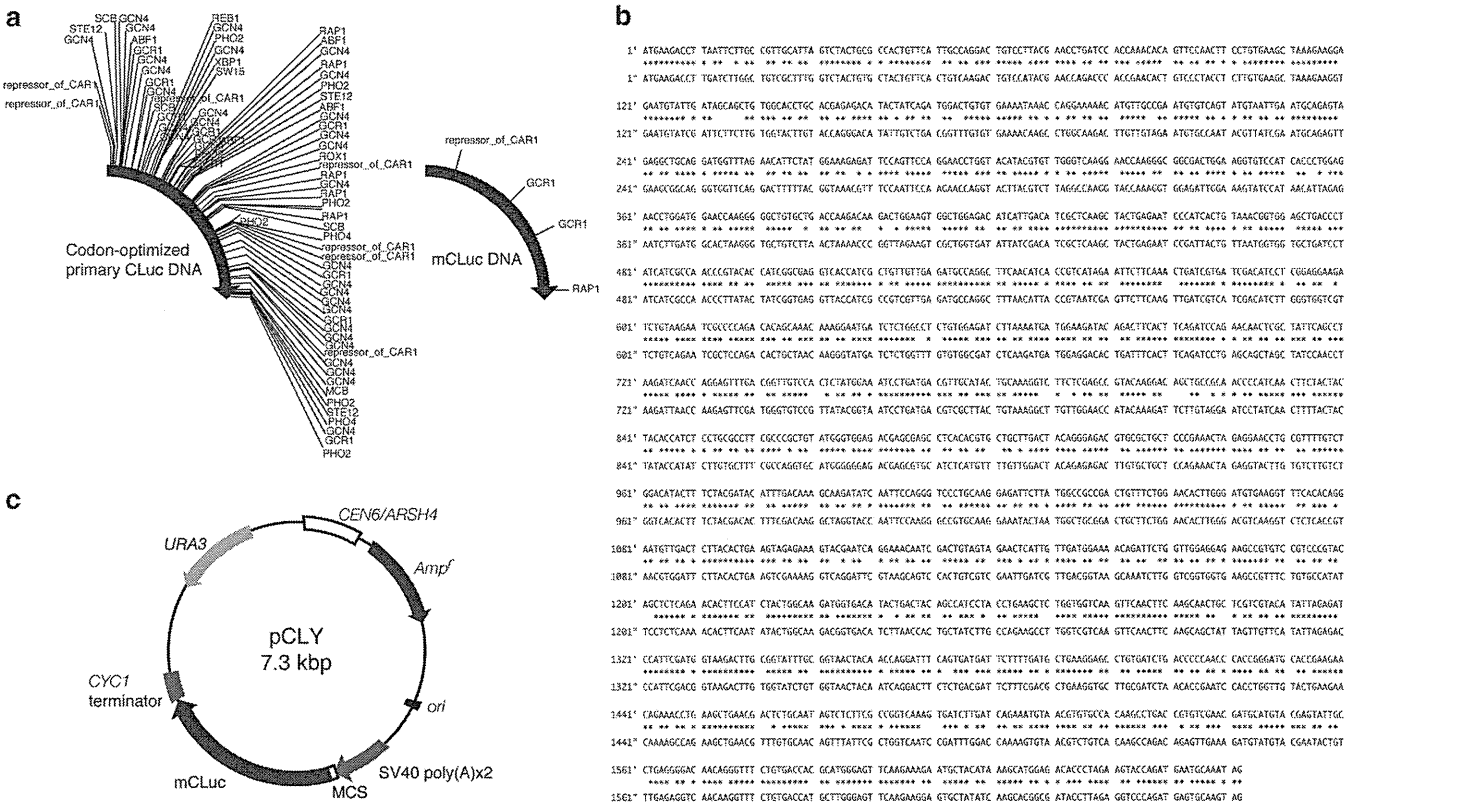


Figure 1. Construction of a reporter plasmid pCLY. (a) Putative *cis*-elements in CLuc-coding DNA: Left, putative *cis*-elements existing in the primary CLuc-coding DNA designed by reverse-translation with optimal codons for *S. cerevisiae*; right, putative *cis*-elements existing in the final CLuc-coding DNA, mCLuc DNA, after removal of most of the putative *cis*-elements by nucleotide substitution without changing encoded amino acids. (b) Comparison of the nucleotide sequence of a variant of native CLuc cDNA (upper, Accession number AB262361) with that of mCLuc DNA (lower, accession number AB259056). Amino acid sequences encoded by these DNAs are the same. (c) Structure of pCLY: *ori*, replication origin for *E. coli*; MCS, multicloning site.

upstream gene of the *TDH3* gene), was amplified by PCR with the following primers and *S. cerevisiae* S288C genomic DNA:

5′TDH3_*Bam*HI: GGGTGGATCCCGAGTTTATCATTATCAA-TAC

3′TDH3: TCGAAACTAAGTTCTTGGTG

An *ACT1* (*YFL039C*) gene promoter (approximately 670 bp), a region between the *ACT1* and *YPT1* genes, was amplified by PCR with the following primers:

5′ACT1_SpeI: CCCATTACTAGTACAAGCGCGCCTCTACC

3′ACT1: TGTTAATTCAGTAAATTTTCGATCTTG

An *ADH1* (*YOL086C*) gene promoter (approximately 1.1 kb), a region between the *ADH1* and *YOL085C* genes, was amplified by PCR with the following primers:

5′ADH1_*Bam*HI: CCCAGGATCCGTATACTAGAAGAAT-GAGCC

3′ADH1: TGTATATGAGATAGTTGATTGTATG

A *CYC1* (*YJR048*) gene promoter (approximately 950 bp), a region between the *CYC1* and *ANB1* genes, was amplified by PCR with the following primers:

5′CYC1_*Bam*HI: CCCAGGATCCGTTTTAGTGTGTGAATGAAA

3′CYC1: TATTAATTTAGTGTGTGTATTTGTG

A *TEF1* (*YPR080W*) gene promoter (approximately 580 bp), a region between the *TEF1* and *MRL1* genes, was amplified by PCR with the following primers:

5′TEF1_*Bam*HI: CCCAGGATCCACAATGCATACTTTGTACG

3′TEF1: TTTGTAATTAAAACTTAGATTAGATTGC

A *GAL1* (*YBR020W*) gene promoter (approximately 450 bp), a region between the *GAL1* and *GAL10* genes, was amplified by PCR with the following primers:

5′GAL1_SpeI: CCCAACTAGTACGGATTAGAAGCCGCCGAG

3′ GAL1: GGTTTTTTCTCCTTGACGTTAAAG

These fragments were ligated with the digested pCLY and digested pGALuRA (see below), and the resultant reporter plasmids were used for transformation of yeast.

Measurement of Promoter Activities by the CLuc Assay. Synthetic liquid media (0.17% yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, 2% glucose supplemented with appropriate amino acids and bases except for uracil) were inoculated with at least three colonies of transformants together (pooled transformants) to reduce variation in CLuc expression among the colonies. The pH of the culture medium was maintained at 6 by adding potassium phosphate (pH 6.0) to a final concentration of 0.2 M because CLuc is unstable at low pH. Transformants were precultured at 30 °C until they reached the stationary phase, inoculated in fresh media, and further cultured at 30 °C overnight. A volume of 5 mL of fresh media was provided in a tube for a midscale assay and 1 mL was provided in a well of a 96-well deep-well plate for small-scale and high-throughput assays. At the early log phase, aliquots of each culture solution were transferred to transparent and black microplates to measure the optical density at 600 nm (OD_{600}) and relative luminescence unit (RLU), respectively. CLuc activity in 20 μL

of yeast culture was measured by a luminometer equipped with an injector (model LB960, Berthold, Bad Wildbad, Germany) after adding 80 μL of 2.5 μM synthetic *Cypridina* luciferin diluted with 0.3 M sodium ascorbate and 25 mM Na_2SO_3 in 200 mM Tris-HCl buffer (pH 7.5). The luminescence intensity (expressed in RLU) was measured for 5 s with a 2 s delay after the addition of *Cypridina* luciferin. The OD_{600} of 200 μL of yeast culture was measured in a transparent microplate by a microplate reader (Sunrise remote, TECAN, Männedorf, Switzerland). The OD_{600} of 3 mL of yeast culture was measured in a 1 cm path cuvette by a spectrometer for midscale analysis. The transcriptional activity of each promoter was defined as RLU divided by OD_{600}. For quantitative analysis, the RLU and OD_{600} were measured when the OD_{600} was between 0.1 and 0.35 as measured by the microplate reader using 200 μL of yeast culture in a well, according to the result shown in Figure 2a, or approximately between 0.2 and 0.8 measured with a 1 cm path cuvette.

Precision of CLuc Bioluminescence Measurement. Intra- and interday precision of CLuc bioluminescence measurements was determined using synthetic *Cypridina* luciferin and CLuc solution in the CLuc secreted luciferase reporter assay kit (ATTO).

Correlation of CLuc Activity with Optical Density of Yeast Culture at 600 nm. Synthetic liquid media were inoculated with pooled transformants in a well. Four cultures of these pooled transformants were incubated at 30 °C until they reached the stationary phase. The transformants in the preculture were inoculated in fresh media and further cultured at 30 °C overnight. At different times, aliquots of each culture solution were transferred to transparent and black microplates to measure OD_{600} and RLU, respectively, for determination of the linear correlation between OD_{600} and RLU.

Construction of a Reporter Vector Carrying β-Gal cDNA. To construct a reporter plasmid using β-Gal, β-Gal cDNA was prepared by PCR using pJM133 DNA[20] and the primers described below:

LacZ_5′*Sma*I_F: GGGTCCCGGGATGACCGGTTCCGGAGCTTG

LacZ_3′XbaI_R: CCCTGTCTAGATTACGCGAAATACGGGCAG

The isolated β-Gal DNA fragment was used to construct a reporter vector, which was almost identical to pCLY except that mCLuc DNA was replaced with β-Gal DNA. The resultant plasmid was named pGALuRA. *ACT1*, *ADH1*, *CYC1*, *TDH3*, and *TEF1* promoters were introduced into pGALuRA, as described for construction of the pCLY reporter plasmid. *S. cerevisiae* YPH500 was transformed with the resultant reporter plasmids, and the transformants were precultured as described above. The volume of the final culture was 20 mL. At the early log phase, the transformant cells were collected and disrupted with CelLytic (Sigma, St. Louis, MO) in the presence of zirconia beads. Cellular extracts were prepared by centrifugation. The concentration of the protein of the cellular extract was quantified using a modification of the method of Lowry et al.[21] for normalization. β-Gal activity was measured according to the standard method.[22] For β-Gal activity, the transcriptional activity of each promoter was defined as β-Gal activity divided by protein concentration.

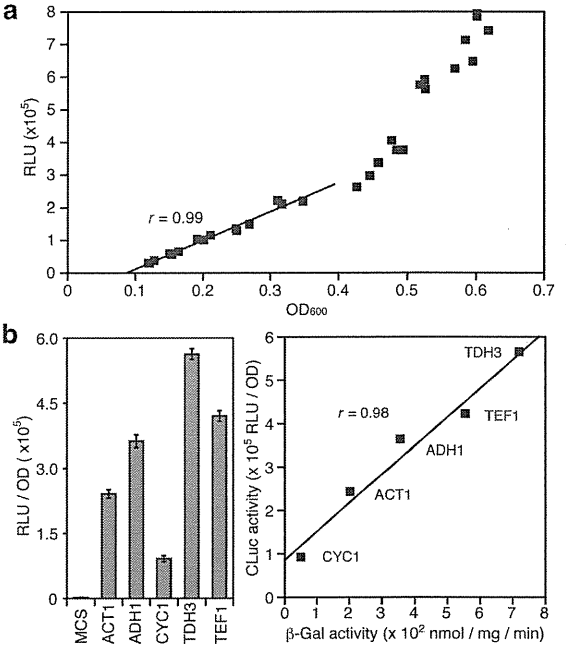


Figure 2. Validation of the CLuc assay by comparison with the conventional assay. (a) Correlation of OD_{600} measured with 200 μL of yeast culture by a microplate reader and CLuc activity of yeast transformants. Cultures of four yeast pooled transformants carrying pCLY-tdh3, pCLY with the constitutive TDH3 promoter, were sampled at different time points. A scatter plot of CLuc activity versus OD_{600} of all data from four pooled transformants is shown. A positive correlation coefficient (r = 0.99) was observed at a range of 0.1−0.35 of OD_{600} measured with 200 μL of yeast culture by a microplate reader, indicating that CLuc activity is proportional to OD_{600} in the early log phase. An increase in CLuc activity at the middle-late log phase would be due to the accumulation of CLuc in yeast culture. (b) Left, normalized CLuc activity of yeast cultures transformed with pCLY carrying one of five promoters: *ACT1*, *ADH1*, *CYC1*, *TDH3*, and *TEF1*. The normalized CLuc activity expressed as RLU/OD_{600} is the average ± SE (standard error) for three pooled transformants, and the data shown are representative of two independent experiments. MCS, pCLY with no promoter. Right, correlation of data between the β-Gal and CLuc assays. β-Gal activity of yeast transformed with pGALuRA carrying one of five promoters was measured and normalized with the protein concentration of the cell extracts. The scatter plot of the normalized β-Gal activity in the conventional assay and the normalized CLuc activity in the CLuc assay for the five promoters is shown. A positive correlation coefficient (r = 0.98) was observed, indicating that the CLuc assay provides data consistent with the conventional assay.

Transcriptional Activities of Truncated TDH3 and GAL1 Promoters. The data for the *TDH3* promoter (−698) and the *GAL1* promoter (−451) were obtained with the transformants carrying pCLY-tdh3 and pCLY-gal1 as described above. The *TDH3* promoter (−471), *TDH3* promoter (−432), *TDH3* promoter (−411), and *TDH3* promoter (−295) were obtained by PCR using the 3′TDH3 primer and one of the forward primers TDH3−471_F, TDH3−432_F, TDH3−411_F, and TDH3−295_F:

(20) McKinney, J. D.; Chang, F.; Heintz, N.; Cross, F. R. *Genes Dev.* **1993**, *7*, 833–843.

(21) Lowry, O. H.; Rosebrough, N. J.; Farr, A. L.; Randall, R. J. *J. Biol. Chem.* **1951**, *193*, 265–275.

(22) Adams, A.; Gottschling, D. E.; Kaiser, C. A.; Stearns, T. In *Methods in Yeast Genetics*; Cold Spring Harbor Laboratory Press: New York, 1998; pp 123–127.

TDH3−471_F: AAGAGGATCCAATGGAGCCCGCTTTTTAAG
TDH3−432_F: AAGAGGATCCAGAATCCCAGCACCAAAATA
TDH3−411_F: AAGAGGATCCTGTTTTCTTCACCAACCATC
TDH3−295_F: AAGAGGATCCGGAGTAAATGATGACACAAG

These truncated *TDH3* promoters, namely, *TDH3* promoter (−471), *TDH3* promoter (−432), *TDH3* promoter (−411), and *TDH3* promoter (−295) were designed to examine the regulatory function of a fermentable carbon source-dependent upstream activation sequence (UAS1*1; −486 bp to −474 bp), a fermentable carbon source-dependent upstream activation sequence (UAS1*2; −448 bp to −436 bp), a fermentable carbon source-dependent upstream repression sequence (URS; −431 bp to −419 bp), and a nonfermentable carbon source-dependent upstream activation sequence (UAS2; −305 bp to−297 bp), respectively.

Similarly, the *GAL1* promoter (−396) and the *GAL1* promoter (−288) were obtained by PCR using 3′GAL1 primer and one of the forward primers GAL1−396_F and GAL1−288_F, respectively:

3′ GAL1: GGTTTTTTCTCCTTGACGTTAAAG
GAL1−396_F: AAGAGGATCCTGCGTCCTCGTCTTCACCGG
GAL1−288_F: AAGAGGATCCGAAAAATTGGCAGTAACCTG

The *GAL1* promoter (−451) has four GAL4-binding domains, which are crucial for the regulation of expression.[23−25] Thus, truncated *GAL1* promoters were produced and used for the CLuc reporter assay. *GAL1* promoter (−396) and *GAL1* promoter (−288) were designed to examine the regulatory function of GAL4-binding domains located between −451 bp and −397 bp and between −349 bp and −332 bp, respectively.

The amplified DNA fragments of the truncated *TDH3* promoters or *GAL1* promoters were ligated with the digested pCLY, and the resultant reporter plasmids were used for transformation of yeast. The transcriptional activity of each truncated promoter was determined as described above. For the induction of *GAL1* promoter, 2% galactose was used in the final medium as a carbon source instead of 2% glucose to induce *GAL1* promoter.

Detection of 17β-Estradiol Based on CLuc Assay. A pCLY-based reporter plasmid for the detection of 17β-estrariol (E2) was constructed. A DNA fragment containing estrogen responsive element (ERE) was amplified by PCR with the *CYC1* promoter fragment and the following primers (EREs are underlined):

5′-3xERE-CYC1_F: GAGACTAGTTAGGTCACTGTGACCTG-AGCTTAGGTCACTGTGACCTGAGCTTAGGTCACTGTGACCT-GATGCATGCATGTGCTCTGTATGTAT

5′-2xERE-CYC1_F: GAGACTAGTTAGGTCACTGTGACCTG-AGCTTAGGTCACTGTGACCTGATGCATGCATGTGCTCTGTA-TGTAT

5′-1xERE-CYC1_F: GAGACTAGTTAGGTCACTGTGACCT-GATGCATGCATGTGCTCTGTATGTAT

3′-CYC1p_R: GGGTATTAATTTAGTGTGTGTATTTGTG

The resultant fragments, which contain EREs upstream of the *CYC1* basal promoter (−143 bp from the initiation codon) as described Bovee et al.[6] were cloned into pCLY. The human estrogen receptor α (ERα) cDNA (GenBank accession: NM_000125) was amplified by PCR from human brain (cerebral cortex) Marathon-ready cDNA (Clontech Laboratories, Inc., San Jose, CA) as a template and cloned into a yeast expression vector pLTex321sV5H driven by the HSP12 promoter (unpublished plasmid). The yeast strain YPH500—carrying one of the reporter plasmids and the ERα expression plasmid—was cultured until the stationary phase and was then subjected to a downshift of the culture temperature from 30 to 10 °C to produce ERα protein under the HSP12 promoter, which is activated at low temperature. Ninety-six hours later, the culture medium was replaced with a fresh medium at the same cell density, and aliquots were transferred into a 96-well deep-well plate. E2 dissolved into ethanol at various concentrations or ethanol alone as a solvent control was added to the yeast culture. Four hours later, CLuc activity was measured as described above.

Yeast Two-Hybrid System Based on CLuc Assay. The CLuc and β-Gal assays were compared using the ProQuest two-hybrid system (Invitrogen) as a model system, because the kit included three useful pairs of interactive proteins, which showed different strengths of interaction. First, the *URA3* gene of MaV203, a yeast strain included in the kit, was disrupted by the G418-resistance gene using a pUG6-based disruption cassette[26] (http://mips.gsf.de/proj/yeast/info/tools/hegemann/loxp_kanmx.html) to yield the *ura3*$^-$ genotype. Primers used for the disruption of *URA3* are as follows:

URA3ko-pUG6f: CCTGTTGCTGCCAAGCTATTTAATATCAT-GCACGAAAAGCAAACAAACTTGCATAGGCCACTAGTGGATC-TG

URA3ko-pUG6r: TTAGTTTTGCTGGCCGCATCTTCTCAAAT-ATGCTTCCCAGCCTGCTTTTCCAGCTGAAGCTTCGTACGC

Since the *LacZ* gene was already integrated into the genome of MaV203, the *LacZ* gene in the genome was further disrupted by the *HIS5* gene derived from *Schizosaccharomyces pombe*. The DNA fragment for the gene disruption was amplified by PCR with pFA6a-His3MX6[27] and the following primers:

5′-LacZ-His3_F: CGTGACCTATCCCATTACGGTCAATCCGC-CGTTTGTTCCCACCGGATCCCCGGGTTAATTAA,

3′-LacZ-His3_R: GTATCGCCAAAATCACCGCCGTAAGCCG-ACCACGGGTTGCCGGAATTCGAGCTCGTTTAAAC.

For the CLuc reporter assay, the resultant strain MaV203 (*ura3*$^-$, *lacZ*$^-$) was transformed with pCLY-gal1, the expression plasmid for Krev1, and one of the three different plasmids: an expression plasmid for RalGDS-*wt*, an expression plasmid for RalGDS-*m1*, or an expression plasmid for RalGDS-*m2*. Krev1 strongly interacts with RalGDS-*wt*, weakly interacts with RalGDS-*m1*, and does not interact with RalGDS-*m2*, according to the manufacturer's protocol. After cultivation, the CLuc activities of the three pooled transformants were determined. For the β-Gal reporter assay, the reporter plasmid pGALuRA-gal1 was transfected to MaV203 (*ura3*$^-$, *lacZ*$^-$) instead of pCLY-gal1. The β-Gal activity of the cell lysates was determined according to the manufacturer's protocol. The disruption of the *Lac Z* gene in the genome in MaV203 (*ura3*$^-$, *lacZ*$^-$) was confirmed by the measurement of the β-Gal activity of the cell lysates from MaV203 (*ura3*$^-$, *lacZ*$^-$) transformed with expression plasmids for Krev and RalGDS-*wt* (data not shown).

(23) Giniger, E.; Varnum, S. M.; Ptashne, M. *Cell* **1985**, *40*, 767–74.
(24) Johnston, M.; Davis, R. W. *Mol. Cell. Biol.* **1984**, *4*, 1440–1448.
(25) Yocum, R. R.; Hanley, S.; West, R. Jr.; Ptashne, M. *Mol. Cell. Biol.* **1984**, *4*, 1985–1998.
(26) Johnson, M.; Riles, L.; Hegemann, J. H. *Methods Enzymol.* **2002**, *350*, 290–315.
(27) Longtine, M. S.; McKenzie, A.; Demarini, D. J.; Shah, N. G.; Wach, A.; Brachat, A.; Philippsen, P.; Pringle, J. R. *Yeast.* **1998**, *14*, 953–961.

Preparation of Yeast 1-kb-Promoter Fragments. We selected 586 genes from the yeast genome by referring to our microarray data[28] and other available microarray data[29] (unpublished). Because it was difficult to define the exact promoter region for each gene, we decided to use a 1-kb-region upstream from the initiation codon of each gene as a promoter fragment. These fragments are referred to as 1-kb-promoter fragments and were prepared from yeast genome DNA by two-step PCR.

Primers for the first PCR amplification of the 1-kb-promoter fragments were designed (SI Table S-2) and synthesized. For the first PCR, the 1-kb-promoter fragments were amplified with a pair of specific primers in which nucleotide sequences of 10 bp adjacent to the ends of digested pCLY were added at the 5′-termini. The resultant products were then subjected to the second PCR to extend the homologous regions to 50 bp by using common primers as follows:

A: GCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGCTAGAACTAGTG

B: GCGGAGGATGCTGCGAATAAAACAGCAGTAAAAATTGAAGGAAATCTCAT

After the second PCR, all PCR products included the 50-bp extension homologous to both ends of the digested pCLY.

High-Throughput Assay of Yeast 1-kb-Promoter Fragments. The *S. cerevisiae* strain YPH500 was transformed with the digested pCLY and a 1-kb-promoter fragment. A circular reporter plasmid was spontaneously generated in a yeast cell by in vivo recombination.[30] We carried out these procedures in a 96-well microplate configured to handle more than 500 samples. After the transformation step, an aliquot of the transformation solution in the 96-well microplate was directly added to fresh liquid media in the corresponding well of a 96-well deep-well plate at a 100-fold dilution. After cultivation for a few days, an aliquot of this culture solution was then transferred to 1 mL of fresh culture media in the corresponding well of another 96-well deep-well plate, and the cultivation was continued until it reached the stationary phase. The second preculture is optional but preferable for obtaining reproducible results. The second preculture was again inoculated in fresh media at a 100-fold dilution in the corresponding well of a new 96-well deep-well plate. CLuc activities and OD_{600} were measured as described above. All procedures, including PCR, yeast transformation, yeast culture, CLuc assay, and determination of OD_{600}, were consecutively performed in a 96-well format with a dispensing robot (BIOMEK2000, Beckman, Fullerton, CA, USA). The half-life of *Cypridina* luciferin in the assay buffer was longer than 48 h;[31] therefore, the diluted working solution of *Cypridina* luciferin was sufficiently stable during bioluminescence measurement in a 96-well format (approximately 15 min are required for 96 samples).

To confirm the in vivo recombination events, a plasmid was isolated from each yeast transformant, and the length of the incorporated promoter fragment in the plasmid was determined by PCR. Promoter activity data were collected only for the 519 transformants shown to contain the promoter fragment.

RESULTS

Construction of a Yeast Reporter Plasmid Using CLuc. We designed and constructed synthetic DNA coding for CLuc (mCLuc DNA, Accession number AB259056) by referring to yeast optimal codons and yeast promoter databases. The number of putative *cis*-elements in the DNA encoding CLuc was minimized by repetitive procedures using databases as described materials and methods (Figure 1a). Consequently, the mCLuc DNA mostly consists of optimal codons for yeast and a minimal number of putative *cis*-elements, with the encoded amino acid sequence being identical to that of the native CLuc (accession number, AB262361) (Figure 1b). Finally, we constructed a single copy reporter plasmid containing mCLuc DNA-pCLY (Figure 1c).

Establishment of Standard Protocol for CLuc Reporter Assay with pCLY. Basic assay conditions were selected using the native CLuc cDNA (accession number AB262361) or the chimeric αCLuc cDNA.[16] Luminescence from CLuc could be measured without the removal of yeast cells by centrifugation. The luminescence intensity of the yeast culture containing the cells was slightly higher than, but correlated with, that of the yeast culture supernatant (data not shown). CLuc in the extracellular matrix of yeast cells could be responsible for producing the increased luminescence.

Quantitative assay conditions were established using the reporter plasmid pCLY carrying mCLuc DNA. We found that the optical density of the yeast culture at 600 nm (OD_{600}) could be used for normalization of the luminescence from CLuc because the RLU was proportional to the OD_{600} of the culture during the early log phase for the yeast transformant carrying pCLY with the constitutive *TDH3* promoter (pCLY-tdh3) (Figure 2a). Therefore, it is recommended that for quantitative assays, the luminescence of 20 μL of yeast culture be measured when the OD_{600} of 200 μL of yeast culture measured by a microplate reader is between 0.1 and 0.35 (Figure 2a). Yeast culture at a wider growth phase could be used for the qualitative assay. The standard protocol for the CLuc reporter assay is illustrated in SI Figure S-1.

Validation of the CLuc Assay. CLuc activity in the yeast culture was compared with β-Gal activity of the conventional assay by using the same promoters. For the experiments, the reporter plasmid carrying β-Gal DNA, designated pGALuRA, was constructed, and five constitutive promoters, *ACT1*, *ADH1*, *CYC1*, *TDH3*, and *TEF1*, were introduced into pCLY and pGALuRA. Figure 2b clearly shows that the transcriptional activities of the five promoters determined by the CLuc reporter assay exhibit a linear correlation with those determined by the conventional β-Gal assay.

We then tested the CLuc reporter assay for its ability to identify *cis*-elements in the *TDH3* and *GAL1* promoters. The truncated *TDH3* promoters were designed to examine the effects of *cis*-elements on transcriptional activity. The result obtained by the CLuc reporter assay (Figure 3a) was consistent with previously reported data.[2] The *GAL1* promoter was similarly analyzed. The *GAL1* promoter has four GAL4-binding domains, which are important for the regulation of expression.[23−25] Thus, truncated *GAL1* promoters were produced and used for the CLuc reporter assay. The result obtained by the CLuc reporter assay (Figure 3b) was also consistent with previously reported results.[23−25]

(28) Sahara, T.; Goda, T.; Ohgiya, S. *J. Biol. Chem.* **2002**, *73*, 113–9.
(29) Gasch, A. P.; Spellman, P. T.; Kao, C. M.; Carmel-Harel, O.; Eisen, M. B.; Storz, G.; Botstein, D.; Brown, P. O. *Mol. Biol. Cell* **2000**, *11*, 4241–4257.
(30) Orr-Weaver, T. L.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4417–4421.
(31) Wu, C.; Kawasaki, K.; Ogawa, Y.; Yoshida, Y.; Ohgiya, S.; Ohmiya, Y. *Anal. Chem.* **2007**, *79*, 1634–1638.

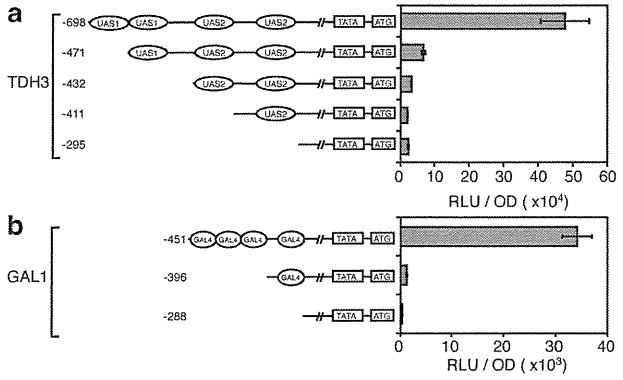


Figure 3. Promoter truncation assay with CLuc. (a) Identification of *cis*-elements in the *TDH3* promoter by the CLuc assay. The *TDH3* promoter and four truncated *TDH3* promoters were introduced into pCLY. CLuc activity of yeast cultures transformed with pCLY carrying one of the five promoters was measured. Normalized CLuc activity expressed as RLU/OD_{600} is the average ± SE for three pooled transformants, and data shown are representative of two independent experiments. (b) Identification of *cis*-elements in the *GAL1* promoter by the CLuc assay. The *GAL1* promoter and two truncated *GAL1* promoters were introduced into pCLY. CLuc activity of yeast cultures transformed with pCLY carrying one of the three promoters was measured. Normalized CLuc activity expressed as RLU/OD_{600} is the average ± SE for three pooled transformants, and data shown are representative of two independent experiments.

The coefficients of variation (CVs) for intra- and interday precision of bioluminescence measurement of CLuc activity were 2.8% (n = 10) and 5.1% (n = 3), respectively. The CVs for intra- and interday precision of the entire CLuc reporter assay in yeast including all steps from preculture to bioluminescence measurement were 7.6% (n = 5) and 10.7% (n = 3), respectively. These data clearly indicated good reproducibility of the CLuc reporter assay in yeast.

Applications of CLuc Reporter Assay. The usefulness of the CLuc assay was compared with that of two conventional reporter bioassays: assay of estrogenic compounds and the two-hybrid system. In the estrogen assay, CLuc activity was clearly increased in a dose-dependent manner in the presence of ERα (Figure 4a). On the other hand, CLuc activity was undetectable in the absence of ERα (mock) even at high concentrations of E2. This indicates that the increased CLuc activity is dependent on E2 via the expressed human estrogen receptor. In this experiment, we examined several reporter plasmids containing various numbers of EREs. Although dose–response curves were similar in all constructs, the basal level of CLuc activity (background) was found to be increased in a manner depending upon the number of EREs. We sequentially measured the CLuc activity of the same culture at time points 4, 8, 12, and 24 h after the addition of E2. Essentially, the same results were observed at all time points (data not shown). The ability to perform multiple sampling from a small volume of culture in a time-course experiment is a major advantage of the CLuc assay over conventional assays such as the β-Gal assay.

The yeast two-hybrid system is one of the major applications of the yeast reporter assay. We thus incorporated the CLuc assay into the conventional yeast two-hybrid system. We used the commercial two-hybrid system containing negative and positive controls to compare the conventional β-Gal assay with the CLuc assay. The commercial product contained three pairs of proteins, including a noninteracting protein pair, a weakly interacting protein pairs, and a strongly interacting protein pair. pCLY or

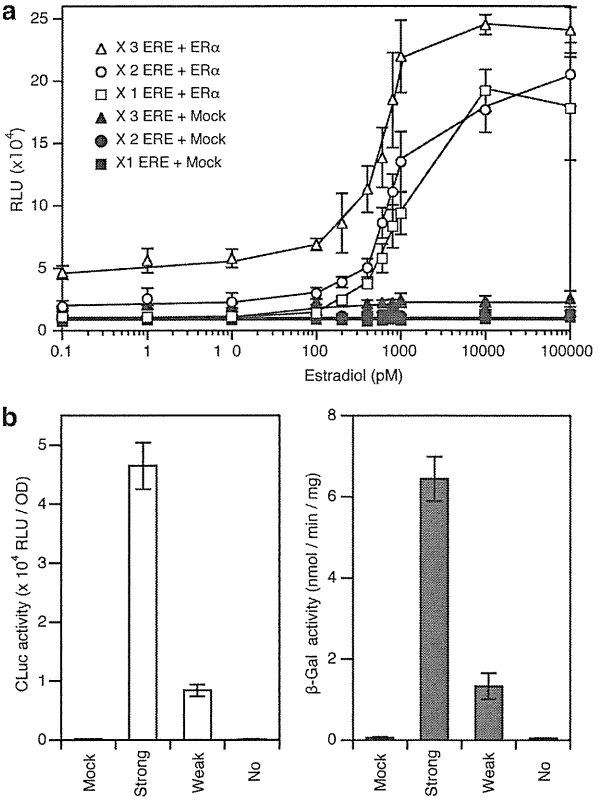


Figure 4. Application of the CLuc assay for the detection of E2 and analysis of protein–protein interactions. (a) The detection of E2 by the CLuc assay. The *CYC1* basal promoter containing EREs was introduced into pCLY (pCLY-X 1, X 2, and X 3 ERE). Various final concentrations of E2 were added to the yeast cultures transformed with pCLY containing EREs and an expression plasmid carrying ERα (ERα) or no cDNA (mock). Aliquots of the yeast cultures were taken 4 h after the addition of E2 and CLuc activity of the cultures was measured. CLuc activity expressed as RLU is the average ±SE for three pooled transformants, and the data shown are representative of three independent experiments. (b) Comparison of the CLuc and β-Gal assays in the yeast two-hybrid system. The *GAL1* promoter was introduced into pCLY (pCLY-gal1) and pGALuRA (pGALuRA-gal1). The yeast strain MaV203 (*ura3*⁻, *lacZ*⁻) was transformed with one of the reporter plasmids and a pair of expression plasmids for the production of a pair of proteins with different affinities: no interactions, weak interactions and strong interactions (see Materials and Methods). The relative values of the averages of normalized CLuc activity in RLU/OD_{600} and normalized β-Gal activity in mmol/min/mg protein in three pooled transformants are indicated. Data shown are representative of three independent experiments.

pGALuRA carrying the same *GAL1* promoter (pCLY-gal1 or pGALuRA-gal1) and two plasmids for the expression of a pair of interactive proteins were introduced into a genetically engineered MaV203 (*ura3*⁻, *lacZ*⁻). Figure 4b reveals that the CLuc-based two-hybrid system produced results consistent with the conventional β-Gal-based two-hybrid system.

Application of the CLuc Assay to High-Throughput Analysis. The CLuc reporter assay was applied to high-throughput analysis. The basal transcriptional activities of more than 500 yeast promoter fragments were determined using the CLuc assay. Five hundred and eighty-six DNA fragments consisting of a 1-kb-promoter fragments and flanking regions for recombination were prepared and used for the transformation of yeast cells together with digested pCLY. A circular plasmid consisting of pCLY and a promoter was spontaneously formed in yeast by in vivo recom-

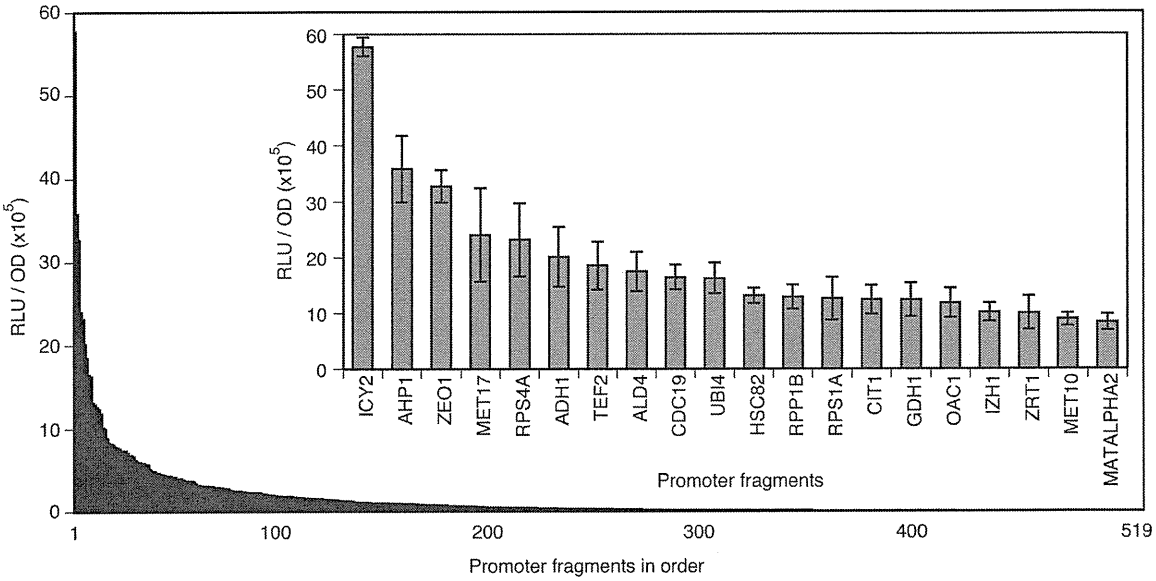


Figure 5. Basal transcriptional activity of 519 yeast promoter fragments. Yeast promoter fragments were introduced into pCLY (see Materials and Methods). The basal transcriptional activity of 519 promoter fragments is indicated in the order of normalized CLuc activity. The average of normalized CLuc activity of each promoter fragment expressed as RLU/OD_{600} for three independent experiments is indicated. Inset: normalized CLuc activity of the top 20 promoter fragments expressed as RLU/OD_{600} is indicated as the average ±SE for three independent experiments.

bination. The success rate of the in vivo recombination was 89% (519/586). Figure 5 and SI Table S-2 show the distribution of basal transcriptional activity of 519 1-kb-promoter fragments determined by the CLuc assay. The 1-kb-promoter fragment of *ICY2* (*YPL250c*) showed the highest level of basal transcriptional activity among the 1-kb promoters tested. The relatively low transcriptional activities (>1500) shown in SI Table S-2 were still higher than a background level (around 200) obtained by a promoterless reporter plasmid, indicating that these promoters exhibited low but significant transcriptional activities.

DISCUSSION

The yeast reporter assay has been used in various applications. In conventional reporter assays, β-Gal is used as the de facto standard for the reporter enzyme regardless of its disadvantages. We herein propose a novel yeast reporter enzyme, which we have designated CLuc. Although CLuc has been expressed in cultured cells,[15] the secretion of CLuc is more pronounced in the yeast reporter assay because yeast is cultured in suspension and its cell wall is difficult to disrupt. Therefore, the assay procedure could be dramatically improved by the replacement of the β-Gal reporter enzyme with CLuc (SI Figure S-1). In addition, CLuc has advantages over firefly and *Renilla* luciferases: CLuc is quite stable at neutral pH and the ATP cofactor is not required for the enzymatic reaction.[32] Moreover, because luminescence could be conveniently measured in aliquots of yeast culture in CLuc assay, a laboratory automation system using a dispensing robot could be employed to increase the number of samples. This enabled us to develop a high-throughput assay. Considering these advantages, we believe that CLuc is one of the best possible reporter enzymes for the yeast high-throughput reporter assay.

We previously reported a study on the secretory pathway for CLuc in yeast using αCLuc cDNA, a mature CLuc cDNA with the yeast α-factor prepro-signal sequence, as the monitoring enzyme.[16] After qualitative experiments, we constructed a synthetic CLuc DNA suitable for a quantitative reporter assay in yeast. Gene expression efficiency in *S. cerevisiae* depends, at least in part, upon the codon usage of a target DNA.[19,33] Indeed, 44% of the codons of the native CLuc cDNA (Accession number AB262361) are considered nonoptimal for *S. cerevisiae* (data not shown). It is also known that *cis*-elements, which regulate the initiation of transcription, may even function apart from the basal promoter region. To restrain possible transcriptional activation at the reporter gene region, known *cis*-element motifs are removed from reporter genes in some examples of commercial products. Therefore, we designed a new DNA encoding CLuc protein, designated mCLuc DNA, whose sequence predominantly includes as many optimal codons of *S. cerevisiae* as possible and otherwise contains the minimal number of known *cis*-element motifs in *S. cerevisiae* (see Figure 1a and b). Although we did not determine the direct effect of these modifications of the nucleotide sequence of CLuc-encoding DNA on the reporter assay, our results indicate that our pCLY system carrying mCLuc DNA provides a functional yeast reporter assay.

The consistency of data obtained from the CLuc reporter assay and the conventional β-Gal reporter assay clearly validates the CLuc reporter assay (Figures 2b, 3, and 4b). Furthermore, the CLuc reporter assay possesses vast superiority to the conventional reporter assay in terms of analysis time required, convenience, throughput (see SI Figure S-1b), and sensitivity.

Another major advantage of this assay over the conventional β-Gal reporter assay is that multiple determinations may be carried out using a single sample. Since only a small aliquot (220 μL) is needed for the CLuc assay and measurement of OD_{600} for normalization, samples for the CLuc assay can be repeatedly taken from the same culture medium even in the case of individual wells of a 96-well deep-well plate (1 mL). Conse-

(32) Shimomura, O. In *Bioluminescence*; World Scientific Publishing: Hackensack, NJ, 2006; pp 62–71.

(33) Bennetzen, J. L.; Hall, B. D. *J. Biol. Chem.* **1982**, *257*, 3026–3031.

quently, it is very easy to achieve a precise time-course determination with low fluctuation by using the same sample throughout the experiment.

Reporter assays for the detection of detrimental chemicals have been established using a combination of the expression of a receptor, especially a nuclear receptor. In the present work, we also established a CLuc-based estrogen reporter assay. Bovee et al.[6] reported that yeast enhanced green fluorescence protein (yEGFP) is a better reporter enzyme than β-Gal. The results in Figure 4a show a high correlation with the reported result[6] using yEGFP. In general, a background level of fluorescence in yeast culture is high due to the existence of fluorescent compounds in yeast culture medium and scattering by yeast cells. Thus, the CLuc-based assay is superior to the conventional β-Gal-based assay and is expected to be similar to or more sensitive than a GFP-based assay, depending on the individual applications. We expect that CLuc will improve other existing bioassays, such as the arylhydrocarbon receptor (AhR) assay,[34] in terms of sensitivity and the number of samples that can be handled.

The yeast two-hybrid system is a major application of the yeast reporter assay. In the conventional system, the affinity of two proteins is visualized as the intensity of blue color derived from the β-Gal reaction. However, this reaction often requires lengthy periods of time for color development, ranging from 15 min to 24 h, according to the manufacturer's protocol. Thus, we attempted to introduce the CLuc assay into the yeast two-hybrid system. Figure 4b clearly indicates that in the yeast two-hybrid system, CLuc was again confirmed to be consistent with the conventional assays using β-Gal. Moreover, the CLuc assay dramatically reduces the time required for the enzymatic reaction (5 s versus 8 h) and dispenses with the laborious procedures required for the β-Gal assay. These results reveal that existing assays can be performed faster and more conveniently by introducing CLuc as the reporter enzyme.

The CLuc reporter assay eliminates the steps of centrifugation, cell lysis, and preparation of supernatant after centrifugation. These steps are difficult to automate. In contrast, the CLuc system enables the development of automation-friendly dispensing steps. In addition, the sensitivity of the assay allows the culture size to be scaled-down from a tube (5 mL) to a 1 mL well of a 96-well deep-well plate. Therefore, we validated the high-throughput capability by investigating more than 500 yeast promoter fragments. The 586 yeast promoters were chosen on the basis of their expected expression performance in microarray data.[28,29] Because the exact promoter regions of all the promoters have not been identified, we used 1-kb DNA fragments upstream of the open reading frames for these genes. These were designated as 1-kb-promoter fragments. In this experiment, we aimed to assess the efficiency of simultaneous plasmid formation and transformation by in vivo recombination and reproducibility of the CLuc reporter assay in a high-throughput format. All procedures from the preparation of 1-kb-promoter fragments by PCR through the measurement of CLuc activity were performed in the 96-well format. A dispensing robot was employed for transfer of aliquots. The success rate of the in vivo recombination was determined to be 89% (519/586) by verifying the size of the 1-kb-promoter fragments in the spontaneously produced circular reporter plasmids. The in vivo recombination-based transformation method will be useful at least when many types of promoters are screened in a high-throughput format. The relatively small standard error of the mean (SE) in three independent experiments in the high activity group provide evidence of the reproducibility of the CLuc reporter assay coupled with the in vivo recombination-based transformation method (see Figure 5 inset and SI Table S-2). This indicates that the CLuc reporter assay can provide reproducible data even in a high-throughput format. Strong promoters may be useful in constitutive or inducible expression systems for foreign proteins in yeast. In order to analyze the characteristics of yeast promoters, microarray experiments are considered to provide useful data. However, microarray data will not exactly reflect transcriptional activity of the promoters because microarray data simply indicate steady-state levels of mRNAs. In contrast, the reporter assay directly determines the amount of protein produced by the promoter function. Thus, the reporter assay would be a powerful tool for the characterization of yeast promoters. However, the conventional reporter assays are not suitable for high-throughput analysis due to the nature of the reporter enzymes used. CLuc is a much more adaptable enzyme for high-throughput reporter assays in yeast. In addition, the large-scale yeast promoter library we constructed will identify useful characteristics of promoters for yeast biotechnology applications. Indeed, in this experiment, we found that the 1-kb-promoter fragment of *ICY2* (*YPL250c*) showed the highest basal transcriptional activity among the 519 kinds of 1-kb-promoter fragments tested. Because the gene product is functionally unknown, the basal transcriptional ability of the gene has not been of interest thus far. The 1-kb-promoter fragments showing high transcriptional activity and the full promoter regions may be used for the production of proteins in yeast in the future. The assessment of transcriptional activity of the 519 promoters under various inducible conditions is now underway.

In conclusion, the CLuc reporter assay in yeast provide results consistent with the conventional β-Gal reporter assay as well as additional outstanding merits of compatibility with an automated high-throughput format, suitability for time-course experiments, reduced culture volume requirements and faster detection. The CLuc assay can be applied to various yeast reporter assay systems, including large-scale analyses, in the development of new drugs and in various fields of biotechnology.

ACKNOWLEDGMENT

We thank Dr. Hegemann for providing pUG6 and pUG35, Y. Nakajima of AIST, and H. Kubota and T. Enomoto of ATTO Corporation for their valuable discussions.

SUPPORTING INFORMATION AVAILABLE

Figure S-1, Table S-1, Table S-2. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review March 31, 2010. Accepted May 17, 2010.

AC100832B

(34) Kawanishi, M.; Sakamoto, M.; Ito, A.; Kishi, K.; Yagi, T. *Mutat. Res.* **2003**, *540*, 99–105.

Cancers **2010**, *2*, 1717-1730; doi:10.3390/cancers2031717

OPEN ACCESS

cancers

ISSN 2072-6694

www.mdpi.com/journal/cancers

Review

5-FU Metabolism in Cancer and Orally-Administrable 5-FU Drugs

Koh Miura [1,*], Makoto Kinouchi [1], Kazuyuki Ishida [2], Wataru Fujibuchi [3], Takeshi Naitoh [1], Hitoshi Ogawa [1], Toshinori Ando [1], Nobuki Yazaki [1], Kazuhiro Watanabe [1], Sho Haneda [1], Chikashi Shibata [1] and Iwao Sasaki [1]

[1] Department of Surgery, Tohoku University Graduate School of Medicine, Sendai, Japan; E-Mails: kinouchi@surg1.med.tohoku.ac.jp (M.K.); hogawa@surg1.med.tohoku.ac.jp (H.O.); ando@surg1.med.tohoku.ac.jp (T.A.); n_yazaki@surg1.med.tohoku.ac.jp (N.Y.); k-wata@surg1.med.tohoku.ac.jp (K.W.); sho@surg1.med.tohoku.ac.jp (S.H.); cshibata@surg1.med.tohoku.ac.jp (C.S.); isasaki@surg1.med.tohoku.ac.jp (I.S.)

[2] Department of Pathology, Tohoku University Hospital, Sendai, Japan; E-Mail: musubi@patholo2.med.tohoku.ac.jp

[3] Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; E-Mail: w.fujibuchi@aist.go.jp

* Author to whom correspondence should be addressed; E-Mail: k-miura@surg1.med.tohoku.ac.jp; Tel.: +81-22-717-7205; Fax: +81-22-717-7209.

Received: 23 August 2010; in revised form: 14 September 2010 / Accepted: 15 September 2010 / Published: 17 September 2010

Abstract: 5-Fluorouracil (5-FU) is a key anticancer drug that for its broad antitumor activity, as well as for its synergism with other anticancer drugs, has been used to treat various types of malignancies. In chemotherapeutic regimens, 5-FU has been combined with oxaliplatin, irinotecan and other drugs as a continuous intravenous infusion. Recent clinical chemotherapy studies have shown that several of the regimens with oral 5-FU drugs are not inferior compared to those involving continuous 5-FU infusion chemotherapy, and it is probable that in some regimens continuous 5-FU infusion can be replaced by oral 5-FU drugs. Historically, both the pharmaceutical industry and academia in Japan have been involved in the development of oral 5-FU drugs, and this review will focus on the current knowledge of 5-FU anabolism and catabolism, and the available information about the various orally-administrable 5-FU drugs, including UFT, S-1 and capecitabine. Clinical studies comparing the efficacy and adverse events of S-1 and capecitabine have been

reported, and the accumulated results should be utilized to optimize the treatment of cancer patients. On the other hand, it is essential to elucidate the pharmacokinetic mechanism of each of the newly-developed drugs, to correctly select the drugs for each patient in the clinical setting, and to further develop optimized drug derivatives.

Keywords: 5-FU metabolism; cell death; colon cancer; oral 5-FU drugs

1. Introduction

Since its introduction more than 50 years ago, 5-fluorouracil (5-FU) has become a key anticancer drug that has been used to treat various types of malignancies for its broad antitumor activity, as well as its synergism with other anticancer drugs. In 1957, Heidelberger *et al.* [1] reported the development of 5-FU, but several important findings had preceded their work. For example, in 1954 Rutman *et al.* [2] showed that uracil was incorporated into rat hepatomas more rapidly than normal tissues; and in 1956 Handschumacher *et al.* reported the tumor-inhibitory activity by 6-azauracil [3]. In recent chemotherapeutic regimens, the continuous intravenous infusion of 5-FU has been combined with oxaliplatin, irinotecan and other drugs. The continuous 5-FU infusion is based on an official report published in the US in 1964 [4], showing that 5-FU is a time-dependent antimetabolite. The meta-analysis of more than 1,200 colorectal cancer patients in six randomized clinical trials, which showed the efficacy of continuous 5-FU infusion compared with bolus 5-FU administration [5], also supported the importance of continuous 5-FU infusion. Based on these results, continuous 5-FU infusion regimens, such as FOLFOX or FOLFIRI, have been established and are widely utilized. On the other hand, recent clinical studies have shown that several of the chemotherapeutic regimens with oral 5-FU drugs are not inferior to those with continuous 5-FU infusion chemotherapy, and in some regimens it may be possible to replace continuous 5-FU infusion chemotherapies with oral 5-FU drugs. Historically, both the pharmaceutical industry and academia in Japan have contributed to the development of oral 5-FU drugs. This review will summarize the current knowledge about 5-FU metabolism, and the information about orally-administrable 5-FU drugs.

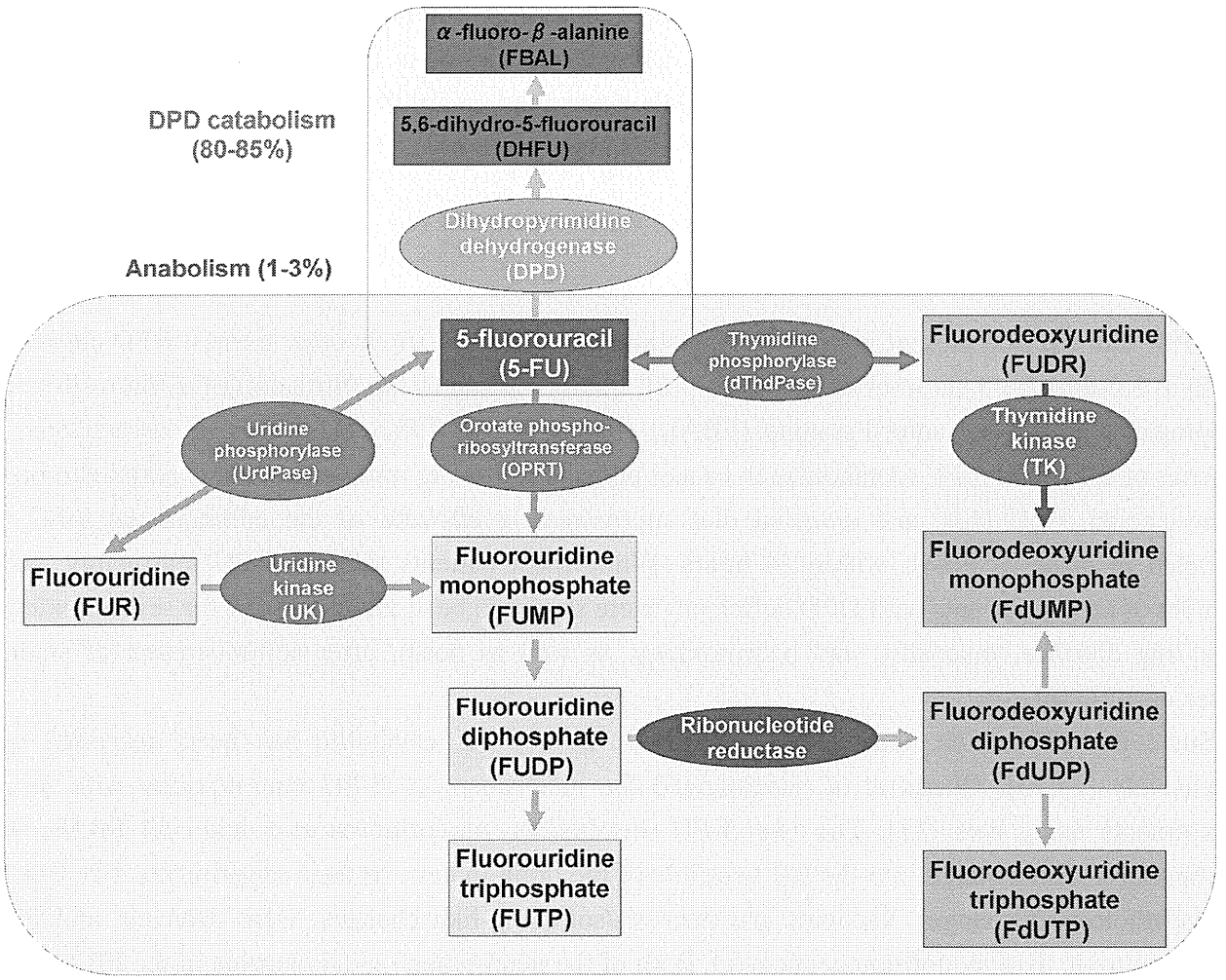
2. 5-FU Metabolism

It has been demonstrated that 80% to 85% of 5-FU is catabolized to inactive metabolites by dihydropyrimidine dehydrogenase (DPD), and only 1 to 3% of the original dose of 5-FU mediates the cytotoxic effects on tumor cells and normal tissues through anabolic actions [6], thereby inhibiting DNA synthesis and RNA processing and function (Figure 1). The 5-FU metabolite, fluorodeoxyuridine monophosphate (FdUMP), forms a ternary complex with thymidylate synthase (TS) and 5,10-methylene tetrahydrofolate (CH2THF), thereby inhibiting the synthesis of DNA.

2.1. 5-FU Anabolism

The chemotherapeutic compound 5-FU is a uracil analogue with a fluorine atom at the C-5 position. After intravenous administration of 5-FU, it rapidly enters cells using the same transport mechanism as uracil [7]. The processing mechanism of 5-FU in cells is as diverse as that of normal pyrimidines, and the current understanding of the metabolism is summarized in Figure 1. First, 5-FU is converted to the following active metabolites: 1) fluorouridine triphosphate (FUTP), which is incorporated into RNA instead of uridine triphosphate (UTP); 2) fluorodeoxyuridine triphosphate (FdUTP), which is incorporated into DNA instead of deoxythymidine triphosphate (dTTP); and 3) FdUMP, which inhibits the activity of TS in the ternary complex, as described in the previous section. FUTP causes alterations in RNA processing and function, and FdUTP and FdUMP cause DNA damage; both of these processes affect RNA and DNA and cause cell death.

Figure 1. 5-FU anabolism and catabolism.



As mentioned, a US report published in 1964 demonstrated 5-FU to be a time-dependent antimetabolite [4]. The main mechanism of 5-FU activation is conversion to fluorouridine monophosphate (FUMP), either directly by orotate phosphoribosyltransferase (OPRT) with phosphoribosyl pyrophosphate as a cofactor, or indirectly via fluorouridine (FUR) through the

sequential action of uridine phosphorylase (UrdPase) and uridine kinase (UK) [8]. The other 5-FU activation pathway involves thymidine phosphorylase (dThdPase), which catalyzes the conversion of 5-FU to fluorodeoxyuridine (FUDR), and FUDR is then phosphorylated by thymidine kinase (TK) to FdUMP. In this series of reactions, the phosphorylation reaction by the UrdPase requires ribose-1-phosphate as a cofactor, eventually synthesizing FUMP. In contrast, the phosphorylation reaction by dThdPase requires deoxyribose-1-phosphate as a cofactor, eventually leading to the synthesis of FdUMP. FUMP is further phosphorylated to fluorouridine diphosphate (FUDP), which is either further phosphorylated to the active metabolite FUTP, or converted to fluorodeoxyuridine diphosphate (FdUDP) by ribonucleotide reductase [8]. FdUDP is then either further phosphorylated to FdUTP, or dephosphorylated to FdUMP. Both FdUTP and FdUMP cause DNA damage.

The conversion of 5-FU to FdUMP in the gastrointestinal (GI) tract and bone marrow elicits GI toxicity and myelotoxicity, respectively. In 1979, an *in vivo* mouse study by Houghton *et al.* indicated that GI toxicity was caused by the incorporation of fluorinated pyrimidines, mainly FdUMP [9]. In 1984, Schuetz *et al.* analyzed the myelotoxicity of 5-FU using CF-1 mouse bone marrow cells under 5-FU exposure *in vitro* [10], and demonstrated that 5-FU incorporation into DNA was closely associated with toxicity and inhibition of DNA synthesis with FdUMP [10]. Interestingly, the meta-analysis of six randomized clinical trials performed in 1998 showed that the grade 3 or 4 hematologic toxicity was more frequent in patients assigned to bolus 5-FU infusion rather than in those assigned to continuous 5-FU infusion [11].

2.2. 5-FU Catabolism

DPD is an enzyme present in the liver, intestinal mucosa and various other tissues. DPD catabolizes 5-FU to 5,6-dihydro-5-fluorouracil (DHFU) [12], finally leading to the formation of α-fluoro-β-ureido-propionic acid and α-fluoro-β-alanine (FBAL) (Figure 1). In 1987, Heggie *et al.* investigated the kinetics of 5-FU and 5-FU metabolites in cancer patients following intravenous bolus administration of radio-labeled 5-FU [13], and revealed that approximately 60–90% of the administered 5-FU was excreted in urine as FBAL within 24 hours. While most patients tolerate 5-FU reasonably well, a number of cancer patients with DPD deficiency were shown to be at increased risk for severe toxicities, including diarrhea, mucositis, and neurotoxicity, as well as death, after administration of standard doses of 5-FU [6].

Since the 1970s, the neurotoxicity of FBAL as a 5-FU catabolite has been discussed quite extensively [14,15]. Okeda *et al.* investigated the mechanism of 5-FU neurotoxicity with *in vivo* experiments using cats [15]. The two 5-FU metabolites, monofluoroacetic acid and FBAL, were continuously administered into the left ventricle of the brain in cats. In their experiments, two types of neuropathological changes, vacuoles and necrosis/softening-like changes, were detected, and FBAL was more toxic than monofluoroacetic acid. Both of the neuropathological changes in the FBAL group were similar to those found in patients following orally-administered 5-FU [15].

The cardiotoxicity of 5-FU has also been attributed to FBAL. Matsubara *et al.* investigated the mechanism of cardiotoxicity for 5-FU and its derivatives using *in vivo* experiments with anesthetized open-chest guinea pigs [16], and proposed that the formation of fluoroacetate, an inhibitor of aconitase, from 5-FU via FBAL, caused cardiotoxicity during chemotherapy [16]. As described in later

publications, FBAL is also the main cause of hand-foot syndrome (HFS) acquired in cancer patients during 5-FU-based chemotherapy. In the 1998 meta-analysis HFS was more frequent in the continuous 5-FU infusion group than in the bolus 5-FU infusion group [5].

2.3. Ternary Complex

FdUMP forms a stable ternary complex with TS and CH2THF [17]. TS catalyzes the reductive methylation of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) with the reduced folate CH2THF. The ternary complex blocks the access of dUMP to the nucleotide-binding site of TS by competition with FdUMP, which results in pool imbalances of deoxynucleotides, especially an increased level of deoxyuridine triphosphate (dUTP); leading to DNA damage. Depletion of dTMP results in the subsequent depletion of dTTP, which perturbs the levels of the other deoxynucleotides [18]. The pool imbalances of deoxynucleotides severely disrupt DNA synthesis and repair, again resulting in DNA damage [19]. As a result, the inhibition of TS results in the accumulation of dUMP, which leads to the increased levels of dUTP [20]. Thymidylate can be salvaged from thymidine through the action of TK, and this salvage pathway can also represent a mechanism of resistance to 5-FU [21]. Despite this information about the activity of 5-FU, the molecular mechanisms downstream of TS inhibition still have to be confirmed [8]. In addition, the clinical significance of TS needs to be demonstrated. In 2008, Showalter *et al.* investigated the connection between TS expression and 5-FU with a thorough literature survey, and in contrast to previous predictions, they found no connection between TS and the patient response to 5-FU [22]. To discuss this matter, we must remember that the influence of TS activity on 5-FU metabolism may change depending on the administration routes of 5-FU drugs, types of 5-FU drugs, the effects of LV, and other factors.

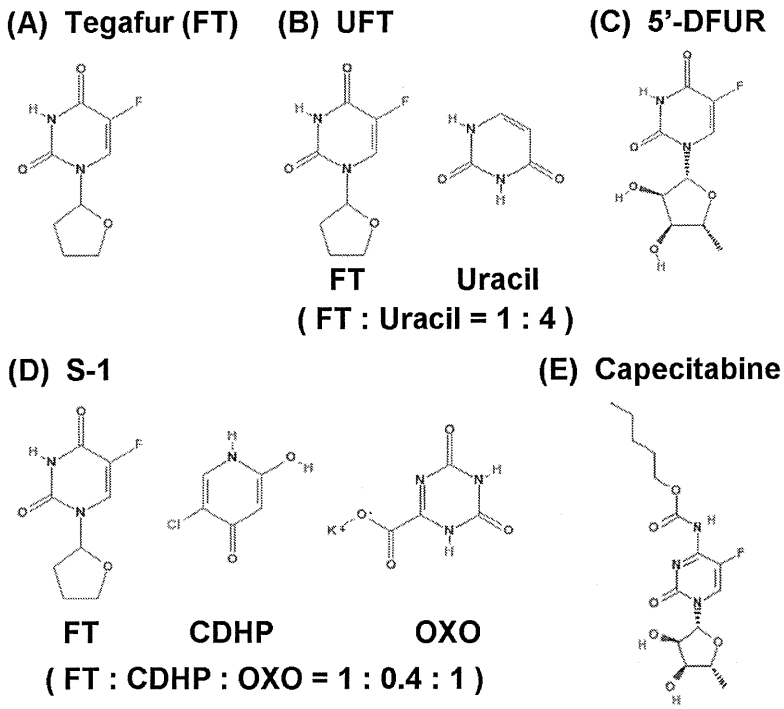
3. Oral 5-FU Drugs

As described in the "Introduction" section, 5-FU is a key anticancer drug for the treatment of various malignancies, and continuous 5-FU infusion regimens have been frequently used because of the apparent time-dependent effects of the drug. However, recent studies have shown that the continuous 5-FU infusion chemotherapies can be replaced with orally-administrable 5-FU drugs in some regimens, without any significant changes in either efficacy or adverse events [23,24]. In addition, oral administration of drugs allows several types of iatrogenic issues to be avoided. For the continuous infusion regimens such as FOLFOX or FOLFIRI, the implantation of a central venous port is required, but complications such as pneumothorax, hemothorax, or disconnection of the devices can occur. Furthermore, catheter-related infection or thrombosis is a serious problem for cancer patients [25,26]. The cost and benefit balance with the use of the central venous port system has been discussed [27], and recent clinical studies revealed that patients prefer oral administration rather than continuous infusion procedures. As such, orally-administered 5-FU regimens are likely to become more common in the clinical setting. Some fluoropyrimidines such as BOF-A2 (Emitefur) and Galocitabine (Ro 09-1390) are under development but not clinically available. In this section, we summarize the information currently available about orally-administrable 5-FU drugs (Table 1 and Figure 2).

Table 1. Orally-administrable 5-FU drugs.

Drug name	Structure (Composition)	Concept	Developer	Refs.
Tegafur	1-(2-Tetrahydrofuryl)-5-fluorouracil	Prodrug	National Institute for Organic Syntheses (Latvia)	[28]
UFT	FT:Uracil = 1:4	Prodrug, DPD inhibitor	Osaka University (Japan)	[30]
5'-DFUR	5'-Deoxy-5-fluorouridine	Prodrug	Hoffmann-La Roche (Switzerland); Nippon Roche Research Center (Japan)	[38,39]
S-1	FT:CDHP:OXO = 1:0.4:1	DPD inhibitor, OPRT inhibitor	Taiho Pharmaceuticals (Japan)	[40]
Capecitabine	N4-Pentyloxycarbonyl-5'-deoxy-5-fluorocytidine	Prodrug	Nippon Roche Research Center (Japan)	[44]

Figure 2. Structures of oral 5-FU drugs. (**A**) Tegafur; (**B**) UFT; (**C**) 5'-DFUR; (**D**) S-1; (**E**) Capecitabine.



3.1. Tegafur

1-(2-Tetrahydrofuryl)-5-fluorouracil (tegafur, FT, FT-207, Futrafur, Ftorafur, *etc.*) was developed as a 5-FU prodrug in the Soviet Union during the Cold War (as reported in 1967 by Giller *et al.* in a Russian record [28]). In 1970, the drug was introduced to Taiho Pharmaceuticals (Japan). Utilizing the benefits of FT, including: 1) its excellent absorbability from the GI tract and 2) its slight conversion to 5-FU in the GI tract, the development of orally-administrable FT was attempted, accomplished and reported in 1977 [29,30]. FT was shown to be gradually converted to 5-FU via cytochrome p450 enzymes in hepatic microsomes [31].

3.2. UFT

UFT consists of uracil and FT. Uracil competes with 5-FU for DPD activity [32,33], resulting in a prolonged 5-FU half-life. To optimize the molecular ratio of FT and uracil, Fujii *et al.*, at the Institute for Protein Research (Osaka University, Japan), analyzed *in vivo* rat models administered with the combination of drugs, and revealed the optimal molar ratio to be 1:4 [34], which led to the introduction of UFT in 1985. In 1978, Fujii *et al.* also reported that the antitumor activity of FT on sarcoma-180 and AH-130 tumors was enhanced by oral administration of uracil, deoxyuridine or uridine [30], and this enhancement of the antitumor activity of FT increased with uracil, which caused a more extensive enhancement than did deoxyuridine or uridine. Furthermore, biochemical modulation of 5-FU had been investigated [35] using methotrexate, trimetrexate, interferon-α, leucovorin (LV) [36], and *N*-(phosphonacetyl)-L-aspartic acid. The addition of LV to UFT regimens increases the available reduced folates, and thereby stabilizes the binding of FdUMP to TS, eventually inhibiting DNA synthesis. In 1997, Rustum *et al.* showed that LV increased the antitumor activity of UFT in the rat [32]; and Ichikura *et al.* showed that UFT with LV enhanced the inhibition of TS activity in gastric cancer patients [37]. In fact, the combination of 5-FU-based drugs with LV has been regarded as one of the standard treatments for colorectal cancer. These results eventually led to the development of S-1.

3.3. 5'-DFUR

In 1979, Cook *et al.* at Hoffmann-La Roche (Switzerland) [38] and Ishitsuka *et al.* in 1980 at the Nippon Roche Research Center (Japan) [39] reported the development of 5'-deoxy-5-fluorouridine (5'-DFUR, doxyfluridine, 5'-fluoro-5'-deoxyuridine, Ro 21-9738, Furtulon, *etc.*). The compound 5'-DFUR is parenterally and orally effective, and its activity was better than that of other fluorinated pyrimidines available at that time. A subline of L1210 leukemia cells was resistant to 5'-DFUR, and Ishitsuka *et al.* revealed that its resistance to 5'-DFUR was due to the lack of the UrdPase [39]. This is because 5'-DFUR is considered to be a depot form of 5-FU, which can be promptly activated by UrdPase [39]. Capecitabine (see below) was developed as the next generation of 5'-DFUR.

3.4. S-1

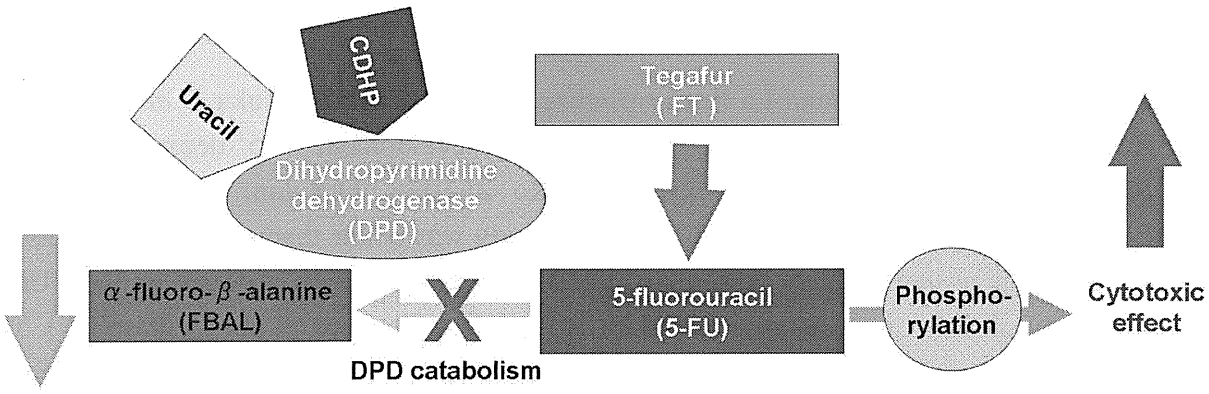
After the development of UFT, Shirasaka *et al.* focused on the development of a novel oral FT-based fluoropyrimidine agent. They developed the next-generation drug, S-1, which both enhances the anticancer activity of 5-FU and reduces its GI toxicity [40]. The development of S-1 was based on two important findings: 1) 5-chloro-2,4-dihydroxypyridine (CDHP, Gimeracil, gimestat, *etc.*) is a DPD inhibitor, and 2) potassium oxonate (OXO) is an OPRT inhibitor (Figure 3).

Tatsumi *et al.* at Otsuka and Taiho Pharmaceuticals (Japan) investigated about 30 compounds for their inhibitory effects of DPD, mainly focusing on pyrimidines, barbituric acid and pyridine derivatives [41]; and in 1987 they reported that 3-cyano-2,6-dihydroxypyrimidine (CNDP) and CDHP were the strongest inhibitors of DPD [41]. Next, Shirasaka *et al.* [42] investigated the possibility of decreasing the GI toxicity of 5-FU without reducing its antitumor activity in rats. OXO localizes in the GI mucosa and selectively inhibits the OPRT, which inhibits 5-FU phosphorylation to FUMP, limiting GI toxicity effects (diarrhea, nausea and vomiting) [42]. In 1993, they reported that OXO inhibited the

phosphorylation of 5-FU to FUMP catalyzed by pyrimidine phosphoribosyl-transferase, in a different manner from allopurinol. With experiments using Yoshida sarcoma-bearing rats, OXO was found to inhibit the formation of FUMP from 5-FU, with its subsequent incorporation into the RNA fractions of the small and large intestine, but not of the tumor and bone marrow tissues. This selective inhibition of 5-FU phosphorylation in the GI tract was due to the much higher concentrations of OXO in GI tissues than in other tissues and in the blood [42].

Based on these findings, CDHP and FT were simultaneously given orally to Yoshida sarcoma-bearing rats in various molar ratios, and then OXO was given orally during consecutive administration of the FT-CDHP mixture to find out the best condition to protect the animals from body weight loss without affecting the high antitumor efficacy of the FT-CDHP mixture [40]. Shirasaka *et al.* finally proposed a suitable formulation of the FT-based anticancer drug, called S-1, consisting of FT, CDHP and OXO at a 1:0.4:1 molar ratio and showed that it had tumor-selective cytotoxicity. S-1 is designed to reduce the GI toxicity of 5-FU; and in 2005 Muneoka *et al.* also reported that S-1 may be administered safely to patients with 5-FU-induced cardiotoxicity in whom FBAL is related to adverse events [43]. Recently, a combination granule version of S-1 has become commercially available.

Figure 3. The metabolism of S-1.



3.5. Capecitabine

Capecitabine (N4-pentyloxycarbonyl-5′-deoxy-5-fluorocytidine, Xeloda™, Ro 09-1978, *etc.*) is an oral fluoropyrimidine carbamate [44], which is selectively converted to 5-FU in tumors through a cascade of three enzymes: (1) carboxylesterase, which is almost exclusively located in the liver and hepatoma, but not in other tumors and normal tissues; (2) cytidine deaminase, which is located in the liver and various types of solid tumors, and 3) dThdPase, which is more concentrated in various types of tumor tissues than in normal tissues (Figure 4).

Miwa *et al.* investigated the tissue localization of the three enzymes in humans [44], and these unique tissue localization patterns enabled the design of capecitabine. Oral capecitabine passes intact through the intestinal tract, but is converted first by carboxylesterase to 5'-deoxy-5-fluorocytidine (5'-DFCR) in the liver, then by cytidine deaminase to 5'-DFUR in the liver and tumor tissues, and finally by dThdPase to 5-FU in tumors. To design the optimized fluoropyrimidine carbamate, a series of N4-alkoxylcarbonyl derivatives were screened for hydrolysis to 5'-DFCR, specifically by carboxylesterase [45]. During the screening process, derivatives having an N4-alkoxylcarbonyl moiety