

Structural Analysis of a Gene Cluster Encoding DFR-like Proteins from Rice Chromosome 4

LEI Hai-Yan^{1,2#}, ZHOU Bo^{1#}, ZHANG Yu¹, HONG Guo-Fan^{1,2*}, HAN Bin^{1*}

(¹ National Center for Gene Research, Shanghai Institutes for Biological Sciences, the Chinese Academy of Sciences, Shanghai 200233, China;

² Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, the Chinese Academy of Sciences, Shanghai 200031, China)

Abstract Sequencing analysis of the 323 kb contig of rice chromosome 4 identified a gene cluster encoding 7 dihydroflavonol-4-reductase (DFR)-like proteins within a 56 kb region. The 7 DFR-like genes were found to be arranged in a tandem array, and all of them comprised 6 exons and 5 introns. Analysis of the predicted amino acid sequences demonstrated that these 7 proteins shared strong similarities with DFR and other enzymes of the phenylpropanoid biosynthesis pathway. RT-PCR revealed the expression pattern of the 7 genes was different in various rice tissues. The structural and functional features of these 7 DFR-like genes and their evolutionary implications are discussed.

Key words gene cluster; rice; *Arabidopsis*; dihydroflavonol-4-reductase (DFR); *Oryza sativa* enzyme of flavonoid biosynthesis (OsEFS)

Flavonoids are secondary metabolites widespread among plants and involved in many plant functions such as UV protection, defense against pathogen attack, legume nodulation and pollen viability^[1,2]. Dihydroflavonol-4-reductase (DFR) catalyses the first common step in the flavonoid biosynthetic pathway leading to anthocyanins and proanthocyanidins. The latter compounds are also known as anthocyanogens and condensed tannins. Some of the flavonoids, the anthocyanidins, anthocyanins and tanins are responsible for the red, purple and brown pigmentation of flowers, fruits, seeds and other plant tissues and organs^[3]. Since these products are not essential for the viability of the plants, flavonoid biosynthesis represents an excellent model system in which to study the regulation of a complex biosynthetic pathway. Thus the genetic control of flavonoid biosynthesis has been studied in several model plants including maize, snapdragon, petunia and *Arabidopsis*^[4,5]. Most of the genes encoding DFR

have been cloned from many plants, such as *Z. mays*, morning glories, *P. hybrida*, grape^[6-9] and *Arabidopsis*^[5] etc., and their sequences are well conserved among plant species.

In this study, the 323 kb contig of rice chromosome 4 was completely sequenced and analyzed. A large gene cluster consisting of seven predicted DFR-like protein genes in a tandem array was found to be located in the 56 kb region of the contig. The deduced protein sequences of these seven genes all shared significant sequence similarities with DFRs and BANYULS^[10] in *Arabidopsis*. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Mutations in the BANYULS gene lead to precocious accumulation of anthocyanins in immature seed coat in *Arabidopsis*. Because DFR and BANYULS both involved in the flavonoid synthesis in plant, we deduced that these seven genes may have similar functions in rice and designated them as OsEFS (*Oryza sativa* enzyme of flavonoid biosynthesis). Each gene of this cluster was named according to their order in the 56 kb fragment. RT-PCR was performed to elucidate the expression pattern of each gene in this cluster. The structural and functional features of these seven DFR-like genes and their evolutionary implications are discussed.

1 Materials and Methods

1.1 Plant materials and growth conditions

Received: March 27, 2002 Accepted: May 21, 2002

This work was supported by grants from the Ministry of Science and Technology, the Chinese Academy of Sciences and the Shanghai Municipal Commission of Sciences and Technology

These authors contribute equally

Present address: Department of Plant Pathology, Ohio State University, Columbus, Ohio, 45210, USA

* Corresponding author: HAN Bin: Tel, 86-21-64825260; Fax, 86-21-64825775; e-mail, bhan@newnetra.ncgr.ac.cn. HONG Guo-Fan: Tel, 86-21-64822885; Fax, 86-21-64825775; e-mail, gfhong@newnetra.ncgr.ac.cn.

Seeds of rice (*Oryza sativa indica Guangluai 4*) were germinated at 37 °C and the seedlings were grown in the light at 30 °C for 3 days for RNA extraction of root and bud. Leaves were collected after 10 days growing. Flowers and young panicles were prepared from 3 month old plants^[11].

1.2 Cloning and sequencing of BAC clones

A BAC (bacterial artificial chromosome) contig, which was anchored on the region from 110.0 cM (centimorgan) to 111.5 cM of chromosome 4, was constructed by using colony hybridization and chromosome walking. The contig consisted of 20 BAC clones which came from two BAC libraries of *Oryza sativa Guangluai 4*, and the genetic markers used as probes were provided by rice genome program (RGP) in Japan^[12,13]. Six tiled BAC clones (BAC H0410G08, H0315F07, H0613A10, B0808H03, H0105C05, H0323C08) with minimum overlaps were sequenced and analyzed. The BAC DNA was purified by cesium chloride gradient centrifugation, and subcloned into pBluescript II plasmid vector (Stratagene) after sonication. Subclones were sequenced at both ends using the DYEnamic™ ET dye terminator kit (Amersham Pharmacia) and analyzed on Megabase100 (Amersham Pharmacia). The sequence data were assembled using PHRED/PHRAP software. Homology searches were performed using the Blast program^[14]. GENSCAN program was used to predict possible genes in this contig^[15].

The nucleotide sequences of the BACs H0410G08 and B0808H03 have been submitted into the EMBL database under the accession numbers AL512546 and AL512545.

1.3 Oligonucleotides

All oligonucleotides used in this study were synthesized by Sangon company, China; except for Oligo dT-adaptor primer which was provided by RNA RCR Kit. The sequences of all oligonucleotides were shown in Table 1.

1.4 RT-PCR

Total RNA of roots, buds, leaves, flowers and young panicles of rice were extracted following the manufacture's instruction (Qiagen RNeasy Plant mini kit). For RT-PCR, 1 µg of DNase-treated total RNA was reverse-transcribed with gene-specific antisense primers and AMV reverse transcriptase using an RNA PCR kit (TaKaRa, Japan), and the entire reaction mixture was used as a template in the subsequent PCR. Each PCR cycle consisted of 94 °C denaturation for 30 s, 60 °C annealing for 30 s, 72 °C extension for 2 min, for 30 cycles.

2 Results

Table 1 Primers for RT-PCR

<i>OsEFS1-3</i>	CAC CAC CAG AGA AAC AGA TG
<i>OsEFS1-5</i>	AGG AAA TCG AGC CGA CTG AC
<i>OsEFS2-3</i>	GTA ACA GCA TTG TCG TCG GC
<i>OsEFS2-5</i>	GCG AAG TAG CCA ACT AAC CG
<i>OsEFS3-3</i>	CGC CCA TCA TCA CCA TCA G
<i>OsEFS3-5</i>	GAC GAT GTC GTC GGA GGT TG
<i>OsEFS4-3</i>	CGA AGG GAG TAG GTT GAA AG
<i>OsEFS4-5</i>	TGT CGG CGG TTG AGA GGA AG
<i>OsEFS5-3</i>	GCA GCA TCA GCT GTT CGA TC
<i>OsEFS5-5</i>	ATC GAC GAC GAC GAT GTC GG
<i>OsEFS6-3</i>	CGG AGC AAC CAT CAA TTC AG
<i>OsEFS6-5</i>	TGT CAG CGG TTG GGA TGA AG
<i>OsEFS7-3</i>	GAA GGA GCA TCA TCA GCG G
<i>OsEFS7-5</i>	ACG AGA TGT CAG CGG TTG AG
Actin-P1	CAT GCT ATC CCT CGT CTC G
Actin-P2	CGC ACT TCA TGA TGG AGT TG

2.1 OsEFS gene cluster was revealed in the 56 kb fragment

Six overlapping BACs representing the 323 kb region of rice chromosome 4 were sequenced and analyzed by using the gene identification software GENSCAN to predict the location of the genes in this contig. The prediction result showed that seven DFR-like genes were clustered in a 56 kb fragment, and the predicted protein sequences of these 7 genes all had high similarities to DFR of rice and BANYULS of *Arabidopsis*. The similarities between 7 OsEFS proteins and BANYULS of *Arabidopsis* are 55.52%—60.18%, and the similarities to DFR in rice are 48.19%—55.56%. One mutator-like transposon was found to be located between *OsEFS3* and *OsEFS4*. These 7 genes all had the same transcription direction and contained 6 exons and 5 introns. Though the length of each gene in genomic level is different from each other, the lengths of their coding regions are almost same. The average G + C content of each gene's coding and noncoding region is similar to other plant genes (Table 2). The predicted seven OsEFS protein sequences had high similarities with each other (Table 3). The relevant features of each deduced protein were shown in Table 4. It is interesting to find that other 6 OsEFS proteins are all

Table 2 Exons and introns of the 7 OsEFS genes

Genes	Length of exons (bp)	Length of introns (bp)
<i>OsEFS1</i>	1 113 (62 %)	1 203 (34 %)
<i>OsEFS2</i>	1 041 (61 %)	1 652 (35 %)
<i>OsEFS3</i>	1 041 (61 %)	559 (40 %)
<i>OsEFS4</i>	1 065 (58 %)	1 994 (34 %)
<i>OsEFS5</i>	1 023 (59 %)	1 422 (33 %)
<i>OsEFS6</i>	1 014 (59 %)	1 016 (36 %)
<i>OsEFS7</i>	1 011 (58 %)	2 376 (41 %)

The GC content of the exons and introns of the 7 *OsEFS* genes are indicated in parentheses.

Table 3 Comparison of the sequences of the 7 *OsEFS*

	<i>OsEFS2</i>	<i>OsEFS3</i>	<i>OsEFS4</i>	<i>OsEFS5</i>	<i>OsEFS6</i>	<i>OsEFS7</i>
<i>OsEFS1</i>	81.15	74.60	62.39	74.04	74.35	62.39
	76.04	65.27	53.73	66.03	68.18	53.73
<i>OsEFS2</i>		74.04	66.87	75.44	74.70	75.52
		65.19	58.36	67.16	67.56	68.06
<i>OsEFS3</i>			65.34	73.59	73.51	72.62
			54.29	62.91	65.18	62.50
<i>OsEFS4</i>				84.59	84.59	73.85
				83.08	83.08	67.39
<i>OsEFS5</i>					82.74	84.23
					75.60	76.79
<i>OsEFS6</i>						87.50
						82.44

The sequence similarity is shown in the upper row and the identity is in the under row.

Table 4 Relevant features of *OsEFS* sequences deduced from the gene cluster

	<i>OsEFS1</i>	<i>OsEFS2</i>	<i>OsEFS3</i>	<i>OsEFS4</i>	<i>OsEFS5</i>	<i>OsEFS6</i>	<i>OsEFS7</i>
AA length	370	346	346	354	340	337	336
Isoelectric point	4.84	4.81	4.89	7.27	5.00	4.98	4.82
Molecular weight (kD)	39.6	36.9	37.6	39.1	37.1	36.3	36.4

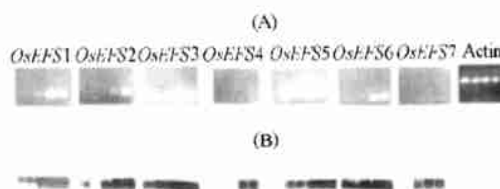
acidic proteins except that the isoelectric point of *OsEFS4* is 7.27. Though the amino acid compositions of 7 proteins are similar, the basic amino acid Arg is higher in *OsEFS4* than in other six proteins. The number of basic amino acid Arg in *OsEFS1-3, 5-7* protein is 17, 11, 12, 17, 13, 12, respectively, however the number in *OsEFS4* is 25.

2.2 Transcript levels of *OsEFS* gene cluster

The expression of the *OsEFS* gene cluster in different tissues of rice was examined by RT-PCR. Total RNAs were extracted from roots, buds, leaves, flowers and young panicles. These RNA were all treated with DNase to remove any contaminated genomic DNA. As a control, amplification by RT-PCR was performed using two primers (actin-P1/actin-P2) specific for the rice actin 1 gene (*ACT1*). The gene-specific primers (*OsEFSs 5 / OsEFSs 3*, see Table 1) were designed according to the GENSCAN prediction result. Each gene's corresponding PCR fragment was sequenced and confirmed that they all came from the transcript of *OsEFS*. The amplification fragments of seven genes were transferred to membranes and hybridized with corresponding genomic DNA.

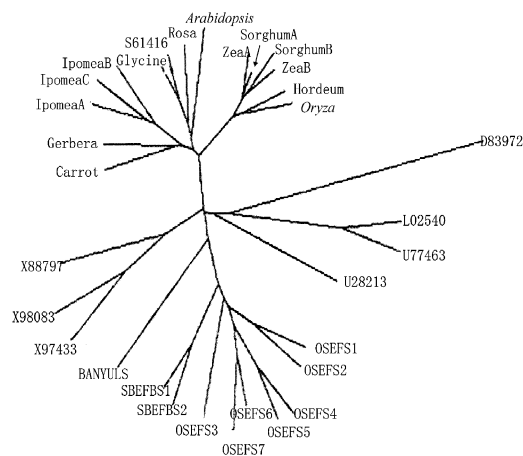
Results of RT-PCR and Southern blot revealed that seven *OsEFSs* transcribed at different levels in five rice tissues [Fig. 1 (A), (B)]. We found that seven genes all had strong expressions in flowers and young panicles and they had different transcription in other three tissues. The expression pattern of seven genes may be due to that these genes involved in the biosynthesis of anthocyanins.

3 Discussion

**Fig. 1** RT-PCR analysis of *OsEFS* transcripts in various tissues

(A) RT-PCR products detected by ethidium bromide staining. (B) RT-PCR products detected by Southern blot hybridization. From left to right is root, immature shoot, mature leaf, panicle and immature seed, respectively. The products of the transcripts of 7 genes were confirmed by sequencing.

We found a large gene cluster encoding seven *OsEFS* genes in the 56 kb fragment of chromosome 4 of *Oryza sativa* ssp. *indica* cv. *Guangluai 4*. Protein sequence analysis demonstrated that *OsEFS* belongs to a superfamily of NADP(H) binding oxidoreductases as defined by Baker *et al.* [16]. Comparison with various databases revealed similarities to known genes encoding the dihydroflavonol reductase (DFR) of rice and many other plant species and to BAN YULS of *Arabidopsis*. A phylogenetic tree was constructed using the CLUSTALX program and is presented in Fig. 2. The tree summarizes the theoretical evolutionary distances among the different NADPH-dependent oxidoreductase superfamily

**Fig. 2** Phylogenetic tree of the members of the superfamily of NADPH-dependent oxidoreductase including 7 *OsEFS* and BAN YULS

The tree was constructed by the program CLUSTALX 1.0.

members and BANYULS and OsEFS. The DFR proteins from several plant species represent a separate cluster from which OsEFSs and BANYULS are excluded. The motif of 13 amino acid residues common to DFRs and thought to define their substrate specificity are not found in BANYULS and seven OsEFS^[10]. However, these 13 conserved amino acids were found in rice DFR^[17]. DFR genes have been cloned from rice^[17], sorghum^[18], barley^[19], maize^[6], *Arabidopsis*^[5] etc.. Though the copy numbers of DFR gene in each species are different, they mostly locate in a single locus. BANYULS^[10] has been cloned from *Arabidopsis* recently and it encodes a DFR-like protein. The BANYULS locates in the chromosome 4 of *Arabidopsis* and the DFR gene of *Arabidopsis* locates in the chromosome 1. Devic *et al.*^[10] thought that BANYULS is not another copy of DFR of *Arabidopsis*, and it may encode the leucoanthocyanidin synthesis. OsEFSs have a closer evolution relation with BANYULS than that of rice DFR. From all these thoughts, we can conclude that OsEFS gene cluster were not a second DFR gene in rice and they may have a closer function with BANYULS.

The predicted protein sequences of seven OsEFS genes showed high similarities to each other. From the phylogenetic tree analysis, we found that seven OsEFS genes formed four evolution branches and the branches were consistent with their localization order in the gene cluster. Members of gene families were often found to be located in the gene cluster in genome. Five peroxidase genes were found in rice chromosome 4^[20]. Three nit genes were located in the 13.8 kb region in *Arabidopsis* genome^[21]. Gene duplication and subsequent divergence are thought to play important roles in evolution of genes^[22]. The putative evolution mechanism of the seven OsEFS genes by duplication was shown in Fig. 3. At least, three gene duplication processes must be involved in generating seven tandem copies from a single copy, and a transposon insertion might occur prior to the second duplication. The insertion of this transposon

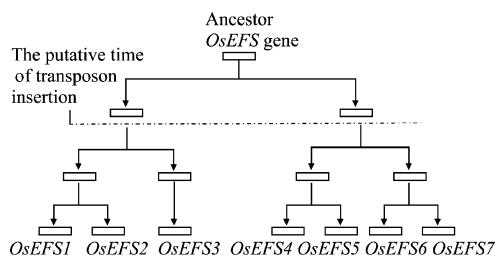


Fig. 3 Schematic representation of the putative evolution mechanism of the 7 OsEFS genes by duplication in rice

led to the phenomenon that previous three genes had higher similarities than other four genes. Transposons were often found in the intergenic regions of gene clusters in the genome^[23].

In summary, we found a large gene cluster encoding seven DFR-like protein in the 56 kb fragment of rice chromosome 4. RT-PCR analysis revealed the expression pattern of these seven genes. We discussed the structural and functional features of these seven genes. The putative evolutionary mechanism of this gene cluster was also discussed in this text.

References

- Lamb CJ, Lawton MA, Dron M, Dixon RA. Signals and transduction mechanisms for activation of plant defenses against microbial attack. *Cell*, 1989, **56**: 215—224
- Long SR. Rhizobium-legume nodulation: Life together in the underground. *Cell*, 1989, **56** (2): 203—214
- Fischer D, Stich K, Britsch L, Grisebach H. Purification and characterization of (+)-dihydroflavonol (3-hydroxy flavanone) 4-reductase from flowers of *Dahlia variabilis*. *Arch Biochem Biophys*, 1988, **264**: 40—47
- Holton TA, Cornish EC. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell*, 1995, **7**: 1071—1083
- Shirley BW, Hanley S, Goodman HM. Effects of ionizing radiation on plant genome: Analysis of two *Arabidopsis* transparent testa mutations. *Plant Cell*, 1992, **4**: 333—347
- Schwarz-Sommer Z, Shepherd N, Tacke E, Gierl A, Rohde W, Leclercq L, Mattes M *et al.* Influence of transposable elements on the structure and function of the A1 gene of *Zea mays*. *EMBO J*, 1987, **6**: 287—294
- Inagaki Y, Johzuka Hisatomi Y, Mori T, Takahashi S, Hayakawa Y, Peyachoknagul S, Ozeki Y *et al.* Genomic organization of the genes encoding dihydroflavonol 4-reductase for flower pigmentation in the Japanese and common morning glories. *Gene*, 1999, **226**: 181—188
- Beld M, Martin C, Huits H, Antoine R, Stuitje, Anton GM Gerats. Flavonoid synthesis in *Petunia hybrida*: Partial characterization of dihydroflavonol 4-reductase genes. *Plant Mol Biol*, 1989, **13**: 491—502
- Sparvoli F, Martin C, Scienza A, Gavazzi G, Tonelli C. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Mol Biol*, 1994, **24**: 743—755
- Devic M, Guilleminot J, Debeaujon I, Bechtold N, Bensaude E, Koornneef M, Pelletier G *et al.* The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J*, 1999, **19**(4): 387—398
- Tao QZ, Zhao HY, Qiu LF, Hong GF. Construction of a full bacterial artificial chromosome (BAC) library of *Oryza sativa* genome. *Cell Res*, 1994, **4**: 127—133
(引自: 细胞研究)
- Hong GF, Qian YM, Yu SL, Hu X, Zhu J, Tao WH, Li W *et al.* A 120 kilobase resolution contig map of the rice genome. *DNA Seq*, 1997, **7**: 319—335
- Sasaki T. The rice genome project in Japan. *Proc Natl Acad Sci USA*, 1998, **95**: 2027—2028
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSFBLAST: A new generation of protein database search programs. *Nucleic Acids Res*, 1997, **25** (17): 3389—3402
- Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*, 1997, **268**: 78—94
- Baker ME, Blasco R. Expansion of the mammalian 3 beta

- hydroxysteroid dehydrogenase/plant dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a bacterial UDP-galactose 4-epimerase, and open reading frames in vaccinia virus and fish lymphocystis disease virus. *FEBS Lett*, 1992, **301**: 89—93
- 17 Nakai K, Inagaki Y, Nagata. Molecular characterization of the gene for dihydroflavonol 4-reductase of Japonica rice varieties. *Plant Biotech*, 1998, **15**: 221—225
- 18 Chen MS, SanMiguel P, AC de-Oliveira, SS Woo, Zhang H, RA Wing, JL Bennetzen. Microcolinearity in sh2-homologous regions of the maize, rice and sorghum genomes. *Proc Natl Acad Sci USA*, 1997, **94**: 3431—3435
- 19 Kristiansen KN, Rohde W. Structure of the *Hordeum vulgare* gene encoding dihydroflavonol-4-reductase and molecular analysis of ant18 mutants blocked in flavonoid synthesis. *Mol Gen Genet*, 1991, **230**: 49—59
- 20 Chen ZH, Zhou B, Han B, Qian YM, Hong GF. Structural analysis of a gene cluster encoding two cationic and three anionic peroxidases from rice chromosome 4. *Acta Biochim Biophys Sin*, 2001, **33** (2): 163—172
(引自: 生物化学与生物物理学报)
- 21 Hillebrand H, Bartling D, Weiler EW. Structural analysis of the *nit2/nit1/nit3* gene cluster encoding nitrilases, enzymes catalyzing the terminal activation step in indole-acetic acid biosynthesis in *Arabidopsis thaliana*. *Plant Mol Biol*, 1998, **36**: 89—99
- 22 Ohno S. *Evolution by Gene Duplication*. New York: Springer, 1970
- 23 Zhou B, Chen ZH, Han B, Hong GF. Identification and structural analysis of a class of potentially transposable solo-LTR in rice genome. *Acta Biochim Biophys Sin*, 2001, **33** (2): 158—162
(引自: 生物化学与生物物理学报)

对水稻 4 号染色体上一个编码二氢黄酮醇还原酶类似蛋白基因簇的结构和进化分析

雷海燕^{1,2#} 周波^{1#} 张宇¹ 洪国藩^{1,2*} 韩斌^{1*}

(¹中国科学院上海生命科学研究院国家基因研究中心, 上海 200233;

²中国科学院上海生命科学研究院生物化学与细胞生物学研究所, 上海 200031)

摘要 通过对水稻 4 号染色体一段 323 kb 的序列测定和分析, 在其中 56 kb 的区域内发现了一个由 7 个编码二氢黄酮醇还原酶 (DFR) 类似蛋白基因组成的基因簇。这 7 个基因在基因簇中串联排列, 每个基因都由 6 个外显子和 5 个内含子组成。这 7 个基因的预测蛋白质序列都和 DFR 以及 BANYULS 蛋白序列类似。DFR 和 BANYULS 都是植物次生代谢类黄酮醇生物合成途径中的结构基因, 它们的缺失或突变都会造成植物花色素合成代谢的不正常。RT-PCR 实验证明这 7 个基因在水稻的 5 个组织中表达不同。文中讨论了这 7 个基因的结构和功能特性以及它们的进化关系。

关键词 基因簇; 水稻; 拟南芥; 二氢黄酮醇还原酶 (DFR); 水稻类黄酮生物合成酶 (OsEFS)

收稿日期: 2002-03-27 接受日期: 2002-05-21

科技部、中国科学院、上海市科委资助项目

并列第一作者

现在地址: Department of Plant Pathology, Ohio State University, Columbus, Ohio, 43210, USA

* 联系人: 韩斌: Tel, 021-64845260; Fax, 021-64825775; e-mail, bhan@newnetra.ncgr.ac.cn. 洪国藩: Tel, 021-64822885; Fax, 021-64825775; e-mail, gfhong@newnetra.ncgr.ac.cn

2002 年《生物化学与生物物理学报》免收版面费的优秀论文名单

- 1 VLDL-受体重复序列的诱变缺失与配体结合力的关系。刘志国 屈伸 冯宁 宗义强 邓耀祖 冯宗忱。2002, **34** (2): 158—164
- 2 酵母 PAPI 与 PHO85 激酶复合物对 YLR190w 的磷酸化。石晓钟 敖世洲。2002, **34**(2): 187—192
- 3 锤头状核酶在 Hep G2. 2. 15 细胞内的抗 HBV 特性。冯英 孔玉英 汪垣 祁国荣。2002, **34**(2): 204—208