



SHENGWU XINXIXUE JISHU  
ZAI SHUIDAO GONGNENG JIYIN  
YANJIU ZHONG DE YINGYONG

# 生物信息技术 在水稻功能基因研究中的 应用

江绍玫 著

中国环境科学出版社

# 生物信息学技术在水稻功能基因 研究中的应用

江绍玫 著

中国环境科学出版社·北京

**图书在版编目（CIP）数据**

生物信息学技术在水稻功能基因研究中的应用/江绍玫著.  
北京：中国环境科学出版社，2008.12

ISBN 978-7-80209-869-5

I. 生… II. 江… III. 生物信息论—应用—水稻—  
基因—研究 IV. S511.032

中国版本图书馆 CIP 数据核字（2008）第 178766 号

责任编辑 张维平  
封面设计 龙文视觉

---

出版发行 中国环境科学出版社  
(100062 北京崇文区广渠门内大街 16 号)  
网 址: <http://www.cesp.cn>  
联系电话: 010-67112765 (总编室)  
发行热线: 010-67125803

印 刷 北京中科印刷有限公司  
经 销 各地新华书店  
版 次 2008 年 12 月第 1 版  
印 次 2008 年 12 月第 1 次印刷  
开 本 787×960 1/16  
印 张 5  
字 数 105 千字  
定 价 16.00 元

---

【版权所有。未经许可请勿翻印、转载，侵权必究】  
如有缺页、破损、倒装等印装质量问题，请寄回本社更换

# 目 录

第 1 章 水稻谷蛋白研究进展 .....	1
1.1 水稻贮藏蛋白 .....	1
1.2 水稻谷蛋白的生物合成 .....	3
1.3 水稻谷蛋白基因的克隆及其结构与表达分析 .....	5
1.4 谷蛋白与水稻育种 .....	9
1.4.1 水稻谷蛋白突变体的形成途径 .....	9
1.4.2 水稻低谷蛋白突变体 LGC-1 .....	10
第 2 章 水稻功能基因的电子克隆 .....	13
2.1 水稻功能基因电子克隆的方法 .....	13
2.1.1 利用 EST 数据库信息 .....	13
2.1.2 利用基因组信息 .....	15
2.2 与水稻功能基因电子克隆相关的生物信息资源 .....	16
2.3 电子克隆中的一些常见问题和对策 .....	18
2.4 水稻电子克隆技术的展望 .....	18
第 3 章 水稻谷蛋白基因 <i>GluB-4</i> 的克隆与表达分析 .....	20
3.1 材料与方法 .....	21
3.1.1 试验材料 .....	21
3.1.2 试验方法 .....	23
3.2 结果与分析 .....	28
3.2.1 全长 <i>GluB-4</i> 的克隆与序列分析 .....	28
3.2.2 <i>GluB-4</i> mRNA 的组织特异性表达 .....	35
3.2.3 <i>GluB-4</i> mRNA 的时间特异性表达 .....	36
3.2.4 Southern 杂交分析 .....	36
3.3 讨论 .....	37

第 4 章 LGC-1 低谷蛋白突变相关基因的 <i>GluB-5</i> 克隆与序列分析 .....	39
4.1 材料与方法 .....	40
4.1.1 试验材料 .....	40
4.1.2 试验方法 .....	42
4.2 结果与分析 .....	43
4.2.1 已知谷蛋白基因在 LGC-1 和 NM 中表达无差异 .....	43
4.2.2 利用生物信息学技术克隆到 1 个新谷蛋白基因 <i>GluB-5</i> .....	44
4.2.3 新谷蛋白基因 <i>GluB-5</i> 在 LGC-1 中表达被抑制 .....	45
4.2.4 新谷蛋白基因 <i>GluB-5</i> RNA 在 LGC-1 中发生了编辑 .....	45
4.2.5 <i>GluB-5</i> 在水稻基因组中数多拷贝基因 .....	58
4.3 讨论 .....	58
参考文献 .....	62
缩略语 .....	72

## 第 1 章

# 水稻谷蛋白研究进展

水稻是世界上最重要的农作物之一，全世界有 40% 的人口以大米为主食。水稻种子蛋白质含量占粒重的 7%~10%，主要以贮藏蛋白的形式存在于胚乳中。因此，大米是人们（尤其是亚洲人）蛋白质消费的重要来源。谷蛋白是水稻种子主要贮藏蛋白，也是稻米中能被人体消化的主要蛋白成分。随着经济的发展和生活水平的提高，人们对大米的蛋白品质也提出了更高的要求。因而，改善谷蛋白含量，培育适合不同消费群体的新品种，一直是水稻育种的一个重要目标。

### 1.1 水稻贮藏蛋白

水稻种子蛋白质主要以贮藏蛋白的形式存在于胚乳。水稻贮藏蛋白按溶解性可分为清蛋白（溶于水）、球蛋白（溶于稀盐）、谷蛋白（溶于稀酸、稀碱）和醇溶蛋白（溶于 60% 乙醇）。其中谷蛋白占米粒总蛋白的 60%~80%，醇溶蛋白占 5%~10%（Cagampang, 1966; Padhye, 1979; Resurreccion, 1983）。实验早已证实，种子贮藏蛋白存积于蛋白体（protein body, PB）中。对玉米、大麦、小麦等大部分禾谷类种子胚乳中单一蛋白体的分离与特性分析表明，存积于其中的蛋白质主要是醇溶蛋白。于是，长期以来，醇溶蛋白被认为是禾谷类的主要贮藏蛋白（Burr, 1976; Camero, 1980; Pernollet, 1983）。然而，后来的实验还证实，水稻胚乳中却存在两种蛋白体，PB-I 与 PB-II，它们主要由醇溶蛋白和谷蛋白组成，这与其他禾谷类种子有着显著区别。

对水稻蛋白体的形态报道始于 20 世纪 30 年代（Santos, 1933）。60 年代电子显微镜应用于水稻胚乳研究后，极大地加快了这一研究领域的发展（Buttrose, 1962）。不久便在发育的水稻胚乳中发现了球形蛋白体，并用蔗糖密度梯度离心将它们分离出来。据观察，水稻种子贮藏蛋白全部储存于这些胚乳蛋白体中。这

些蛋白体用柠檬酸铅和醋酸双氧铀染色后，呈现明暗交替的片层结构，类似于木料的同心层（Mitsuda, 1967）。

由于分离技术的滞后，最初无法将淀粉粒和蛋白体区分开来，曾有过水稻胚乳中存在三种蛋白体的报道（Bechtel, 1978）。不过，随着研究的深入，以及水-多聚物双相系统（aqueous polymer two-phase system, APTP system）的建立，终于成功地从水稻成熟胚乳中将蛋白体与淀粉粒分离开来（APTP 系统是根据各粒子不同表面特性进行样品分离，蛋白体和淀粉粒经 APTP 处理后有非常显著的行为差异），进而通过蔗糖梯度离心分离出仅有的两种蛋白体，即 PB-I 与 PB-II（Ogawa, 1975；Tanaka, 1980）。

经 APTP 系统处理和蔗糖梯度离心得到的蛋白体形态研究显示，PB-I 为片层结构，直径 1~2  $\mu\text{m}$  的球形体；PB-II 形状不规则，着色较均一，无片层结构，直径 2~3  $\mu\text{m}$ 。在数量上，胚乳细胞中 PB-II 明显多于 PB-I（图 1-1）。两种蛋白体主要富集于亚糊粉层，而很少出现在大米粒的中心区域，糊粉细胞中则未曾发现这两种蛋白体。用胃蛋白酶消化 PB-I、PB-II，结果显示，PB-I 稳定，而 PB-II 常断裂。因此表明，PB-I 抗酶解，而 PB-II 则易酶解。SDS-PAGE 分析表明，PB-II 由分子量为 22 kDa、26 kDa、37 kDa、38 kDa、39 kDa 的多肽成分组成；PB-I 由分子量为 10 kDa、13 kDa、16 kDa、57 kDa 多肽组成。溶解度逐步分离表明，22 kDa、37 kDa、38 kDa、39 kDa、57 kDa 多肽为谷蛋白（只溶于稀酸或稀碱），13 kDa 为醇溶蛋白，10 kDa、16 kDa、26 kDa 为球蛋白（溶于稀盐）。由此推论，PB-I 是醇溶蛋白储存体，PB-II 为谷蛋白储存体。PB-II 中的谷蛋白分为分子量 22 kDa 和 37~39 kDa 两种类型（Juliano, 1976；Tanaka, 1980）。



图 1-1 电镜下观察到的水稻蛋白体

Fig.1-1 Rice protein bodies under electronic microscope

其实, 水稻蛋白体的上述分离效果还不够理想。因为, 透射电子显微镜下观察结果显示, PB-II 组分较单一, 而 PB-I 中则混杂有 PB-II。只有通过蔗糖梯度离心和胃蛋白酶处理相结合, 才能有效分离并纯化发育期水稻种子中 PB-I。进一步的 SDS-PAGE 分析显示, PB-I 中含有多组多肽链。最大的一组分子量为 13 kDa, 二组较小的为 10 kDa 和 16 kDa。13 kDa 组又可分为分子大小有微小差异的 13a (大组分) 和 13b (小组分)。通过光密度仪测定 SDS-PAGE 胶中各色带相对密度, 还可读出 PB-I 和 PB-II 中每条多肽链的相对含量。多方实验证实, PB-I 与 PB-II 在密度、形状、蛋白质组成及合成过程上均存在显著差异 (Sugimoto, 1980; Oparka, 1982; Ogawa, 1987)。

在发育中的水稻胚乳细胞存在两种显著不同的糙面内质网 (ER), 潴泡糙面内质网 (C-ER) 和蛋白体糙面内质网 (PB-ER)。前者谷蛋白 mRNA 的富集量高于醇溶蛋白, 后者主要富含醇溶蛋白 mRNA。谷蛋白和醇溶蛋白 mRNA 分布于不同的糙面内质网可以为两种蛋白合成后定位于各自的蛋白体提供一个高效机制 (Bechtel, 1980; Yamagata, 1982; Yamagata, 1986; Li, 1993; Muench, 1997; Muench, 1998; Muench, 2000)。

## 1.2 水稻谷蛋白的生物合成

研究表明, 水稻胚乳先在与糙面内质网结合的多核糖体上合成谷蛋白前驱体, 之后在信号肽的引导下进入糙面内质网腔形成谷蛋白前体 (谷蛋白前驱体信号肽被切除), 进而通过高尔基体的作用输送到 PB-II 的液泡状前体中, 谷蛋白前体在其中被水解成酸 ( $\alpha$ )、碱 ( $\beta$ ) 两种亚基 (分子量分别为 37~39 kDa 和 22~23 kDa), 酸碱亚基共价交联在一起, 此为水稻成熟谷蛋白的存在形式 (图 1-2)。

采用 SDS-PAGE 和体内同位素标记等方法, 分析水稻种子发育期贮藏蛋白的合成过程。结果显示, 22~23 kDa 和 37~39 kDa 组分总量与 57 kDa 量相当接近; 在发育期间, 57 kDa 多肽的合成早于谷蛋白各亚基。短期标记 (从开花后到谷穗发育期的 4~12 d 用  $^{14}\text{C}$  Leucine 处理 2 h) 谷蛋白迅速累积时期的谷穗或淀粉胚乳后发现, 57 kDa 多肽被标记上, 而谷蛋白各亚基却未被标记。体内脉冲追踪结果显示, 57 kDa 多肽的放射性转移到了 22~23 kDa 和 37~39 kDa 多肽中。体外蛋白质合成表明, 在发育的水稻种子中不存在直接编码 22~23 kDa 和 37~39 kDa 多肽的 mRNA, 却有编码 57 kDa 多肽的 mRNA, 而且 57 kDa 谷蛋白前体的合成早于谷蛋白各亚基, 该多肽在体内相对较稳定 (Yamagata, 1982)。综上所述, 可得出如下结论: 在水稻种子发育期, 胚乳先合成单链的 57 kDa 多



肽，随后通过水解酶裂解为成熟的 22~23 kDa 和 37~39 kDa 多肽；57 kDa 前体的水解断裂发生在翻译后，而不是共翻译过程；前体的翻译后裂解不同于信号肽的去除，它要求有一个高度组织性的系统 (Blobel, 1975; Sacker, 1986; Luthe, 1983; Krishnan, 1986a; Krishnan, 1986b; Furuta, 1986)。

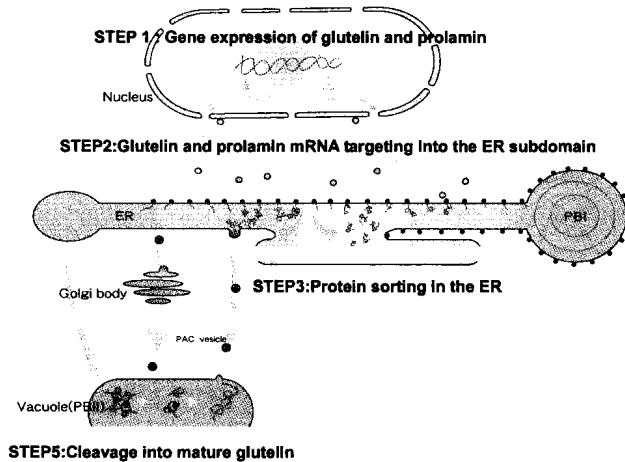


图 1-2 水稻种子贮藏蛋白生物合成途径

Fig.1-2 Biosynthesis pathway of seed storage proteins in rice

在发育期水稻胚乳中，谷蛋白的合成和储存完全不同于玉米种子。后者前体多肽仅较成熟肽重 1.1~2.0 kDa，在合成好的多肽通过内质网膜时 N 末端的一段额外肽行使着信号肽的功能 (Bur, 1976; Brinegar, 1982; Kirihara, 1988)。研究还表明，水稻谷蛋白的生物合成及储存系统与豌豆和大豆球蛋白非常相似，后者的成熟多肽由一个 20 kDa 亚基和一个 40 kDa 亚基组成，它们均为 60 kDa 前体的裂解产物。水稻谷蛋白和豆球蛋白在氨基酸序列上也存在高度的同源性 (Croy, 1980; Tumer, 1981; Zhao, 1983; Domoney, 1987; Fukazawa, 1985; Naito, 1988)。

水稻谷蛋白是禾谷类植物贮藏蛋白中首次被报道生物合成过程中需经过翻译后加工的。不过，脉冲标记显示，储存于水稻 PB-I 中的蛋白质 (10 kDa、13 kDa、16 kDa) 是以成熟多肽的形式合成，并不经过翻译后修饰 (Yamagata, 1982; Yamagata, 1986a; Yamagata, 1986b; Kim, 1988; Zhou, 1990; Li, 1993; Zhou, 1993)。

水稻谷蛋白的合成时期与其他贮藏蛋白各组分存在一定的差异，这可以通过

SDS-PAGE 分析水稻种子发育时期胚乳蛋白得到证实。一般情况下, 22~23 kDa、37~39 kDa 及 26 kDa 多肽在开花后 5 d 出现, 10~16 kDa 多肽在开花后 10 d 出现; 相反, 57 kDa 和 76 kDa 多肽在整个种子发育时期水平相当稳定。采用透射电镜观察和蔗糖梯度分离不同发育时期水稻胚乳, 结果都表明, PB- II 的形成速度快于 PB- I。

水稻胚乳中谷蛋白含量和种类随水稻品种不同而异。有些谷蛋白是所有水稻共有的, 有些是籼稻、粳稻或爪哇稻特有的, 还有些是某一地方品种特有。因此, 谷蛋白 2-D 电泳图谱可以作为判断水稻品种及各品种之间关系的重要指标之一。如籼稻中只含谷蛋白 $\alpha_{5b}$ 和 $\alpha_8$ , 不存在 $\alpha_3$ 或 $\alpha_{5a}$ ; 相反, 粳稻和爪哇稻含谷蛋白 $\alpha_3$ 或 $\alpha_{5a}$ , 而不含 $\alpha_{5b}$ 和 $\alpha_8$ 等(Kagawa, 1988; Huebner, 1990; Hirano, 1991a; Saruyama, 1992; Komatsu, 1993; Abe, 1996; Qu, 2001c)。

有趣的是, 存在于高秆水稻品种 Norin29 及其半矮秆近等基因系 SC-TN1 胚中的半矮秆相关蛋白 SRP1 和 SRP2 分别与水稻谷蛋白 $\alpha_{5a}$ 和 $\alpha_{5b}$ 亚基在氨基酸序列, 等电点及结构上都非常一致。在这两个品种中,  $\alpha_{5a}$ 和 $\alpha_{5b}$ 不仅在胚乳中表达, 还在胚和幼叶中表达。但是, 在种子萌发期, 胚中的谷蛋白 $\alpha$ 亚基含量上升, 而胚乳中谷蛋白所有 $\alpha$ 亚基含量都降低。这一结果表明, 编码同一谷蛋白亚基的基因在胚和胚乳中的表达调控机制却显著不同; 胚乳中的谷蛋白一直被看作是种子萌发所需氮和氨基酸的来源, 而胚中谷蛋白 $\alpha_{5a}$ 和 $\alpha_{5b}$ 亚基很可能行使着不同于贮藏蛋白的功能(Hirano, 1991b)。

### 1.3 水稻谷蛋白基因的克隆及其结构与表达分析

已有研究报道, 水稻谷蛋白由一多基因家族编码, 分 *GluA* 和 *GluB* 两个亚族, 单倍体各由约 5 个成员组成。它们的表达有着时空特异性, 个别基因还存在时间差异性表达。自 1986 年以来, 陆续有关于谷蛋白基因克隆的报道, 迄今为止, 已有 8 种谷蛋白基因(或 cDNA)得到了克隆并被测序。根据基因核苷酸序列的推导, 水稻谷蛋白 57 kDa 前体由 496~499 个氨基酸残基组成, 其中包括信号肽(24 aa)、酸性亚基(278~284 aa)和碱性亚基(191~197aa)。

pREE61 cDNA (*GluA-1*) 克隆是首次被报道的谷蛋白基因(Takaiwa, 1986), 之后 TypeI 基因(*GluA-1*)、TypeII 基因(*GluA-2*) (Takaiwa, 1987a; ), pREEK1cDNA (*GluB-1*) (Takaiwa, 1987b), pREE79cDNA (*GluA-3*)、*GluA-3*、*GluA-4* (Takaiwa, 1991a), *GluB-1*、*GluB-2*、*GluB-3* (Takaiwa, 1991b), pG22cDNA (*GluA-3*) (Satoh, 1994) 和 *Gt1* (*GluA-2*)、*Gt2* (*GluA-1*)、*Gt3* (*GluA-3*) (Okita, 1989),  $\lambda$ RG21cDNA

(Masumura, 1989) 相继被克隆报道。后来证明, 它们有些属于同一基因, 只是克隆时所用材料不同, 致使序列间有细微差异。

谷蛋白前体 cDNA 克隆, pRG206 cDNA 是为了阐述水稻谷蛋白与其它种子球蛋白的结构关系和进化过程, 以一段已知谷蛋白氨基酸序列对应的简并寡核苷酸序列为探针, 筛选水稻种子发育时期胚乳的 cDNA 文库获得的 (Zhao, 1983; Higuchi, 1987)。由 pRG206 cDNA 核苷酸序列推导的氨基酸序列和大豆球蛋白及豌豆球蛋白氨基酸序列之间有 32% 的同源性; 水稻谷蛋白和 11S 大豆球蛋白的合成及翻译后的加工特点、成熟肽的存在方式等方面也存在很多的相似性 (Yamagata, 1982; Lycett, 1984; Marco, 1984; Lycett, 1984; Marco, 1984; Wen, 1985; Momma, 1985a; Momma, 1985b; Negoro, 1985; Fukazawa, 1985; Negoro, 1985)。于是推断, 水稻谷蛋白基因与大豆球蛋白基因起源于同一祖先, 然后通过趋异进化成现在的基因形式 (Higuchi, 1987)。

水稻谷蛋白基因, *Gt1*、*Gt2*、*Gt3* 是以  $^{32}\text{P}$  标记的 cDNA 克隆 pG22 (插入 1.8 kb 谷蛋白 cDNA 全长序列) 为探针, 从几十个水稻基因组文库中 ( $\lambda$ 噬菌体为载体) 筛选到的。当时认为, 水稻谷蛋白就是由这三个基因亚族编码, 后来证实它们分别对应于 *GluA* 亚族的 *GluA-2*、*GluA-1* 和 *GluA-3*。将这三个基因的相关 DNA 片段亚克隆, 并作进一步序列分析。结果表明: *Gt1* 和 *Gt2* 在 5'-端 900 bp 左右相当保守, 相互有 87% 的同源性; *Gt3* 则与其它两个基因表现出很大的异源性, 特别在非编码区; 三个基因启动子近侧至翻译起点 (-125~-1 bp) 都相当保守, 而且三个小内元均位于编码片段相同的位置; 由这些基因编码的氨基酸序列的多态性易于解释谷蛋白各组分 2-D 电泳所显示的分子大小和电荷多相性。

以 *Gt1* 和 *Gt3* 为特异性探针, 采用 Northern 杂交分别对水稻种子不同发育时期总 RNA 进行分析。结果表明, 谷蛋白 mRNA 的转录始于开花后的 5 d (与 Yamagata 的谷蛋白检测结果相同), 但在种子发育的后期, *Gt1* 和 *Gt2* 与 *Gt3* 转录有很大不同。*Gt1* 转录积累贯穿种子整个发育期; *Gt2* 的转录积累方式与 *Gt1* 相似; *Gt3* 的转录则呈现时间差异性: 最大转录积累于开花后 5~10 d, 在种子成熟期稳定下降 (Liu, 1995)。由此可见, 在水稻种子发育期, 谷蛋白基因家族各成员的表达不是同步的, 而是受着基因差异性表达调控。

*Gt1*、*Gt2*、*Gt3* 推导的氨基酸序列和豆科贮藏蛋白氨基酸序列之间存在很高的同源性, 进一步证实谷蛋白与豆球蛋白起源于同一基因。更值得注意的是, 谷蛋白与豆 11S 球蛋白一样, 编码序列存在非保守区域, 这些区域似乎更易发生进化性变化, 在插入和缺失积累的情况下尤其如此 (比较不同来源的 11S 蛋白得出)。 *Gt1*、*Gt2*、*Gt3* 对应的氨基酸序列相互比较表明, 突变主要集中在三个多

肽区域,其中之一与豆 11S 贮藏蛋白高突变区域相一致,这可能是蛋白质修饰的一个潜在靶区域 (Okita, 1989)。

Takaiwa 用 pREE61 cDNA 作探针,在低严谨条件下从水稻胚乳 cDNA 文库筛选到一种新型的 cDNA——pREE79。然后以 pREE79 cDNA 作探针,筛选 mini-genomic 文库 (以  $\lambda$ gt11 为载体的基因组 DNA EcoRI 2-6 kb 片段和  $\lambda$ Ongc 为载体的基因组 DNA XbaI 2.5~6 kb 的片段分别构建的)。在  $\lambda$ gt11 文库中分离到 4.4 kb、2.8 kb、0.8 kb 三种 EcoRI 片段;在  $\lambda$ Ongc 文库中分离到的都是 XbaI 2.9 kb 的插入片段。将 EcoRI 2.8 kb、0.8 kb 和 XbaI 2.9 kb 导入 pUC118 进行亚克隆后测序。结果发现,曾报道的 *GluA-2* 位于 4.4 kb 片段中, pREE79 cDNA 则位于 2.8 kb 片段中, 2.8 kb EcoRI 和 2.9 kb XbaI 片段重叠的核苷酸序列完全一样,于是把这个基因命名为 *GluA-3*。它的核苷酸序列除多了三个内元外,其它与 pREE79 cDNA 完全一样,由此证明 *GluA-3* 基因是有功能活性的。0.8 kb 片段与谷蛋白基因编码区域有同源性 (与 *GluA-3* 有 88% 的同源性,与 *GluA-1*、*GluA-2* 只有 73% 的同源性),两个内元位于与其它谷蛋白基因相同的位置 (与 pREE79 cDNA 序列比较得出),于是将它命名为 *GluA-4*。由于 *GluA-4* 在 1 174 bp~1 176 bp 处有一个 TAG (终止密码),因而它是假基因 (Takaiwa, 1991a)。

Takaiwa 等首次分离到了水稻谷蛋白 *GluB* 亚族基因的三个成员——*GluB-1*、*GluB-2*、*GluB-3*。根据序列分析表明,它们之间外元的同源性为 80%~88%,而与 *GluA* 亚族基因外元的同源性为 60%~65%,表明两个亚族基因外元有着明显的区别。在 *GluB-3* 的 740 bp~742 bp 处有一个 TAG (终止密码),故 *GluB-3* 与 *GluA-4* 一样是假基因。 $\lambda$ RG21 cDNA 与 pREE61 cDNA 核苷酸序列的同源性为 67.6%,故属于 *GluB* 亚族基因 cDNA,其基因尚未被克隆,暂被定为 *GluB-4* cDNA (Takaiwa, 1991b)。

分析比较已克隆的 8 种谷蛋白基因 (或 cDNA) 和燕麦 11S 球蛋白的核苷酸序列,不仅可以看出水稻谷蛋白各基因的共性,还得出了水稻谷蛋白与燕麦 11S 球蛋白基因之间的系统发育关系。所有谷蛋白基因的内元与外元连接点都有严格保守的 GT/AG 序列,连接区附近的序列彼此很相似,且与单子叶植物内元共有序列相似;谷蛋白基因的内元较短 (77 bp~129 bp),并富含 A+T 序列 (内元含 68.5%~79.8%,外元含 54%~57%);所有谷蛋白均含 24 个氨基酸组成的信号肽,信号肽与成熟肽酸性亚基间连接均为 Ala-Gln,酸性亚基和碱性亚基连接处均为 Asn-Gly;在谷蛋白中存在 5 个高变区域,分别位于信号肽,酸性亚基 N-末端、C-末端、中部和碱性亚基 C-末端 (Hanley, 1988)。

Northern 杂交分析表明, *GluA-3* 的 mRNA 水平在种子发育早期上升,于开

花后 10 d 达到峰值, 之后便下降。这不同于谷蛋白其它基因的表达方式。除 *GluA-3* 外的其他谷蛋白基因在开花 6 d 后才能检测到 mRNA, 然后不断上升, 于开花后 14~16 d 达到峰值, 随后降至一个低水平。种子贮藏蛋白的这种差异表达, 在豌豆球蛋白 (Domoney, 1987),  $\beta$  伴大豆球蛋白 (Naito, 1988) 和玉米醇溶蛋白 (Kirihara, 1988) 已有报道。因而, *GluA-3* 被认为是研究种子贮藏蛋白特异性表达非常有价值的系统。对 *GluA* 亚族各成员 5'-端区域核苷酸序列的比较显示, 时间差异性表达大概是由位于 -150 bp 以外的区域序列决定的, 因为这一区域的核苷酸序列同源性大为降低。

水稻谷蛋白基因的组织特异性和发育时期特异性表达在启动子区域都有相应的元件控制。*Gt3* 启动子曾被导入烟草中, 结果发现它能指导报告基因在转基因烟草中呈现时间和种子特异性表达 (Leisy, 1990)。部分缺失实验进一步证明, *Gt3* 启动子 5'-侧的 -945 bp~-726 bp (Domain I) 和 -346 bp~-263 bp (Domain II) 对 *Gt3* 正常表达是必要的, 而且存在于这两个区域的顺式作用元件在控制 *Gt3* 的表达上起着不同的作用。Domain I 中的调节因子含有胚乳特异性表达的所有信息; 而 domain II 中的顺式作用元件则负责时间和数量特异性表达 (Zhao, 1994)。*GluA-2* 也曾被导入烟草中, 结果发现, 在 *GluA-2* 启动子控制下, GUS 报告基因于转基因烟草种子成熟胚乳中产生特异性表达; 在转基因系统中, *Gt3* 的表达早于 *GluA-2* 的相对时间与水稻细胞相一致。这表明, 谷蛋白时间特异性表达在异源的烟草系统中被严格地保留下来 (Takaiwa, 1990)。

转基因实验还证实, 水稻谷蛋白启动子区域的某些基序对保证基因的一定表达水平必不可少。如谷蛋白基因 -73 bp~-61 bp 与 -212 bp~-200 bp 的 AACA (AACAAACTCTATC) 基序和 -165 bp~-158 bp 与 -96 bp~-92 bp 的 GCN4 (ATGAG/CTCAT) 基序对高水平种子特异性表达是必需的; -437 bp~-317 bp (121 bp) 序列对 *GluA-3* 表达的数量调节十分重要 (Yoshihara, 1996a; Yoshihara, 1996b); 5'-侧 -245 bp 中的 AACA、GCN4、ACGT 序列对 *GluB-1* 高水平胚乳特异性表达不可或缺 (Wu, 1998a; Wu, 1998b; Woshida, 1999)。

目前, 在水稻种子已发现两类转录因子与谷蛋白的胚乳特异性表达有关。一类是转录因子 OSMYB5, 它通过与 AACA 基序特异性结合调节谷蛋白的胚乳特异性表达 (Zhao, 1994; Gubler, 1997; Suzuki, 1997; Suzuki, 1998); 另一类是转录因子 RISBZ1, 它通过与谷蛋白和醇溶蛋白基因 5'-侧序区均存在的 GCN4 基序结合调节水稻整个贮藏蛋白基因的胚乳特异性表达 (Nakase, 1996; Onodera, 2001)。

## 1.4 谷蛋白与水稻育种

### 1.4.1 水稻谷蛋白突变体的形成途径

为明晰人体对水稻蛋白的消化利用情况, Tanaka 曾对人体粪便作过详尽分析。在电镜下, 从食用大米者的新鲜粪便中可观察到水稻 PB-I 状的蛋白颗粒 (fecies protein particles, FPP), 但未见 PB-II 状蛋白颗粒; 进一步的分析结果显示, FPP 在重量及成分上与食入体内的大米 PB-I 相当 (Tanaka, 1975a; Tanaka, 1975b; Tanaka, 1980; Resurreccion, 1993; )。由此可见, PB-II 能被人胃蛋白酶降解, 而 PB-I 却不能被消化。也就是说, 醇溶蛋白因无法被胃蛋白酶降解而成为人体不能消化的蛋白质成分, 而谷蛋白不仅是水稻种子主要贮藏蛋白, 也是大米中能为人体消化吸收的主要蛋白成分。由此, 围绕改进水稻贮藏蛋白品质形成了两种育种目标: 一是培育高谷蛋白含量品种, 以增加其营养价值; 二是创造低谷蛋白含量品种, 以适应蛋白摄入被限制的肾脏病人等的需要。

高谷蛋白含量品种可通过转基因、化学诱变或常规杂交育种获得 (Shin, 1977; Higashi, 1974; Higashi, 1976; Kataoka, 1978; Satoh, 1979; Kambayashi, 1984; Schaeffer, 1987; Ogawa, 1989; Hiei, 1994; Iida, 1997; Satoh, 1997; Katsube, 1999)。

用 MNU (甲基亚硝基脲) 处理水稻受精卵细胞是获得贮藏蛋白突变体的一种有效方法。由此法得到的突变体大致有 4 种类型: A 型 (10/13a-L), 低含量 10 kDa 多肽; B 型 (10/16-H), 高含量 10 kDa 多肽; C 型 (13b-L), 低含量 13 kDa 多肽; D 型 (57-H), 高含量 57 kDa 多肽。其中 C 型和 D 型突变几率较高, C 型除醇溶蛋白降低外, 谷蛋白酸碱亚基含量明显高于野生型 (Kumamaru, 1988)。

转基因等现代生物工程技术是获得高谷蛋白水稻品种的另一有效途径。大豆球蛋白基因导入水稻后, 70%新合成的大豆球蛋白可以以盐溶蛋白的形式积累, 并且像在发育期的大豆种子中一样形成 7S 的三聚体和 11S 的六聚体; 30%大豆球蛋白通过特定连接与谷蛋白形成杂合型寡聚体。但不管以何种形式存在的大豆球蛋白, 最终全部被分拣贮存在 PB-II 中。这样, 转基因水稻植株种子可消化蛋白含量被提高 20%左右 (Katsube, 1999)。

低谷蛋白水稻品种都是通过物理或化学诱变形成的突变体 (Kumamaru, 1987; Kumamaru, 1988; Iida, 1993; Fukuoka, 1996; Iida, 1997; Qu, 1997; Qu, 2001; Takemoto, 2002)。

目前获得的低谷蛋白突变体主要由 $\gamma$ -射线、乙烯亚胺(EI)或甲基亚硝基脲(MNU)等物化条件诱变而来。在这些突变体中,它们有各自的蛋白特征,与野生型相比,分别缺失着谷蛋白 $\alpha$ -1、 $\alpha$ -2、 $\alpha$ -3和其他亚基等,而且各自遗传规律有所不同。

#### 1.4.2 水稻低谷蛋白突变体 LGC-1

Iida 等利用 $\gamma$ -射线,0.2%乙烯亚胺和甲基磺酸乙酯分别处理普通水稻栽培品种日本优(Nihonmasari, NM),共获得1433个突变系。在乙烯亚胺处理组中,筛选到一个低育、半矮、早黄的突变系 NM67。采用 SDS-PAGE 对 NM67 种子全蛋白进行分析,结果发现,与野生型 NM 相比, NM67 的 37 kDa~39 kDa 和 22 kDa~23 kDa 多肽色带明显变薄,且缺失了 37 kDa~39 kDa 四条色带中分子量最大的那条;醇溶蛋白组分的 13 kDa 多肽和清蛋白-球蛋白组分的 16 kDa、26 kDa 多肽色带密度变大。将全蛋白 SDS-PAGE 图谱进行扫描,可看出 NM67 种子中各蛋白亚基的相对含量发生了明显变化,其中谷蛋白含量下降了一倍多,而醇溶蛋白却增加了一倍多(表 1-1)。2-DE 结果还显示,在众多的谷蛋白 37 kDa~39 kDa 亚基中, NM67 中缺少了分子量最大的亚基。

表 1-1 NM67 与日本优糙米中主要多肽成分的百分含量(%)

Table 1-1 Percentage of major polypeptides contained in brown rice of Nihonmasari (NM) and NM67

分子量/kDa	76	57	37~39	26	22~23	16	13	总数
NM	4.8	8.8	29.0	8.0	23.0	8.8	17.8	100
NM67	4.0	7.6	13.0	16.5	7.3	10.5	41.1	100

将 NM 与 NM67 回交,  $BC_1F_1$  种子全部表现为低谷蛋白(高醇溶蛋白)型,表明突变基因具显性特征。在  $F_2$  种子中,低谷蛋白性状总是与高醇溶蛋白相伴,且低谷蛋白(高醇溶蛋白)突变型与野生型性状比例约为 3:1。这一结果进一步说明, NM67 低谷蛋白(高醇溶蛋白)性状由单个显性基因控制。

NM67 是以低育、半矮、早黄筛选到的。分析  $F_1$ 、 $F_2$  植株表型发现,早黄性状由一隐性基因控制。在  $F_2$  中,所有早黄植株都是半矮和部分不育的。NM67 的半矮低育被看成是控制早黄性状基因的多相性表达。 $F_2$  植株的基因型通过对  $F_3$  种子的全蛋白分析得以确定。这个后代分析实验显示,低谷蛋白(高醇溶蛋白)纯合型、杂合型与正常蛋白纯合型分离比例在绿色和黄色植株中都是 1:2:1。

这充分表明, 控制低谷蛋白(高醇溶蛋白)基因与控制早黄基因是独立的, 而非连锁。

NM67 与 NM 后代的遗传分析过程中, 在  $F_2$  中筛选到 41 株低谷蛋白(高醇溶蛋白)纯合型植株, 其中 34 株属绿叶杂合型, 7 株绿叶纯合型。低谷蛋白纯合型兼具绿叶纯合型植株被定名为 LGC-1 (low glutelin content-1) (表 1-2)。

表 1-2 NM67 与日本优杂交后代  $F_3$  中低谷蛋白和早黄性状分离比例  
Table 1-2 Segregation of low glutelin trait and early yellowing in  $F_3$  of the cross between NM67 and Nihonmasari

叶型 Leaf type	低谷蛋白纯合型 Homozygotes of low glutelin	低谷蛋白杂合型 Heterozygotes of low glutelin	正常蛋白纯合型 Homozygotes of normal type	总数 Total	$\chi^2$ (1:2:1)
正常 Normal	41	81	49	171	1.21
早黄 Early yellowing	14	28	10	52	0.92
总数 Total	55	109	59	223	0.26
$\chi^2$ (1:2:1)	0.01	0.03	2.04	0.89	

LGC-1 一方面克服了 NM67 早黄、低育、半矮、总蛋白高的缺陷; 另一方面保留了低谷蛋白(高醇溶蛋白)的优势, 且种子蛋白总量与 NM 无差别(表 1-3) (Iida, 1993)。

由于 LGC-1 稻米谷蛋白含量低于粒重 4%, 因此它被作为低蛋白食品用于肾脏病患者的辅助治疗。临床实验表明, 以 LGC-1 为主食的患者蛋白质摄入量 and 血清肌酸酐含量明显降低, 它改善患者病情的效果与低蛋白代用品相同。因此, LGC-1 被认为是对肾病患者(尤其以大米为主食者)非常有效的食疗辅助品。从 1994 年起, 日本各地广泛试种 LGC-1。目前, LGC-1 稻米已成为肾脏病人和肾功能衰退的糖尿病人专用食品。此外, LGC-1 稻米较低蛋白代用品要便宜得多, 可以大为降低患者的治疗费用, 因而深受广大肾病患者的欢迎 (Mochizuki, 2000)。

关于 LGC-1 低谷蛋白突变的分子机理, Miyahara 曾做过部分研究。他分别利用水稻 *GluA* 和 *GluB* 亚族基因的 cDNA 克隆为探针, 对 LGC-1 及其野生型 Nihonmasari 基因组 DNA 进行 Southern 杂交分析。结果显示, LGC-1 和 NM 的条带大致相同; Northern 杂交结果则表明, LGC-1 的 *GluA* 转录本略微减少, 而 *GluB* 转录本却只相当于 NM 的 1/10 左右。进一步对 LGC-1 的蛋白质进行 2-D



电泳分析还发现，在众多的谷蛋白酸性亚基中，缺失了 2 个 *GluB* 亚族基因的编码产物 (Miyahara, 1999)。这些结果根本无法清楚解释 LGC-1 的突变现象。因而，有待进一步的研究来揭示 LGC-1 低谷蛋白突变的分子机理。

目前，国内有关水稻谷蛋白的研究鲜见报道，对谷蛋白品质改造的水稻育种更是处于起步阶段，我国急需加强对这一领域人力、物力资源的投入。

表 1-3 NM67、LGC-1 与日本优特性比较

Table 1-3 Comparison of characteristics among NM67, LGC-1 and Nihonmasari

	特性 Characteristics	NM67	LGC-1	日本优 (Nihonmasari)
	株高/cm Plant height	71.8	91.3	96.1
	茎长/cm Culm length	50.9	69.1	73.4
	蛋白含量/% Protein content	10.5	7.6	7.5
A	每株穗数 Number of panicles in a plant	7.3	12.9	13.1
B	每穗花数 Number of flowers in a panicle	76.9	88.5	88.2
C	每穗可育种子数 Number of fertile seeds in a panicle	27.3	81.3	81.5
D	种子育性/% Seed fertility (C/B)	43.9	92.1	92.4
E	千粒重/g Weight of 1 000 grains	17.1	21.4	22.2
	每株产量/g Estimate of yield per plant (A×B×D×E) /100 000	4.2	21.4	23.7