# 人 *HGLP* cDNA 的克隆和原核表达

周海军 黄晓玮 周 严 李光涛 阴 彬 马艳玲 彭小忠 袁建刚* 强伯勤

(中国医学科学院基础医学研究所, 中国协和医科大学基础医学院, 中国人类基因组北方研究中心, 北京 100005. * 联系人, E-mail: yuanjg@mail.east.net.cn)

**摘要** 克隆了人 *HGLP* cDNA, 编码蛋白具有 G 蛋白偶联受体类似的 7 个跨膜区, 具有视紫红质样 G 蛋白偶联受体超家族的基序(motif). Northern 杂交分析发现 *HGLP* 广泛表达. 将 *HGLP* 的 N, C 端非跨膜区片段克隆到 pET30a$^+$表达载体中, 在 *E. coli* 中获得了高效表达.

**关键词** G 蛋白偶联受体 *HGLP* cDNA 原核表达

G 蛋白偶联受体是最大的膜受体超家族之一, 其典型结构是: 具有 7 个跨膜 α螺旋, 糖基化的 N 端位于质膜外, 形成配体结合结构域; 带有多个磷酸化位点(丝氨酸, 苏氨酸)的 C 端位于细胞内, 第 5, 6 螺旋间的环(loop)与 G 蛋白结合, 从而传递受体信号, 在细胞的很多信号传导途径中具有重要功能[1]. G 蛋白偶联受体除了在生理活动中扮演着重要的角色外, 它与疾病的关系也很密切, G 蛋白偶联受体的突变可以造成多种疾病. 而且许多药物是以 G 蛋白偶联受体为靶子而发挥作用.

我们通过对人胎脑 cDNA 文库规模化测序直接分离、克隆了一条全新的 cDNA, 其推测的 ORF 编码一个具有 555 个氨基酸, 分子量为 63.3 ku 的蛋白质, 其 mRNA 在多种组织中表达. 新蛋白具有 7 个跨膜区, 这类似于 G 蛋白偶联受体的结构. Motif 分析将其归为视紫红质样的 G 蛋白偶联受体超家族. 据此我们将新克隆的 cDNA 命名为 *HGLP* (human G-protein coupled recepter like protein gene). 为了进一步研究人 *HGLP* 的功能, 我们将该基因的 N, C 端编码区克隆到 pET30a$^+$载体中进行了原核表达.

## 1 材料和方法

(ⅰ) *HGLP* cDNA 的筛选与序列分析. (1) cDNA 克隆的筛选. 我室在构建λZAP Express 胎脑文库时, 随机挑取 895 个克隆, 5′测序 150 ~ 300 bp, 然后进行序列分析, 挑取具有全新 EST, 有同源 EST 但无全长 cDNA 或是其他物种有同源蛋白而无人全长 cDNA 的克隆共 362 个, 进行克隆环化. 经 5′, 3′测序、序列分析后挑出一部分全长可能性较大、与 GenBank 库比较为新序列的克隆, 进行全序列测定. 其中新基因 *HGLP* 的 cDNA 就是这样发现和克隆的. (2) cDNA 序列的网上分析. DNA 和蛋白质的同源性比较使用 Blast 软件[2]与 NCBI, EBI, DDBJ 的数据比较 (http://www.ncbi.nlm.nih.gov). 穿膜区的预测采用 Hofmann 的方法[3]. 基序(motif)分析运用 motif Finder 程序(http://www.motif.genome.ad.jp).

(ⅱ) *HGLP* 基因组织表达谱分析. 以 *HGLP* 5′端 510 bp 编码区为杂交探针, 对成人 8 种组织进行 Northern blot(购自 Clonetech)杂交. 操作步骤按产品说明书中的方法进行.

(ⅲ) HGLP 蛋白 N, C 端的融合表达. (1) HGLP 蛋白 N, C 端编码区表达载体的构建. 根据 *HGLP* cDNA 的序列设计引物 Nu 和 Nl 扩增其 N 端编码区, Cu 和 Cl 扩增其 C 端编码区(N 端上游引物 Nu: 5′GG GGT ACC GGA AAG ATC CTC TTC AGA AAT ACC 3′, 下游引物 Nl: 5′CCC AAG CTT GGG ATA GTC TTC AAG TGT GAG G 3′; C 端上游引物 Cu: 5′GA AGA TCT TTC ACC ATT GTC TGA GGA AG 3′, 下游引物 Cl: 5′GG GGT ACC TCC CAT TCC TTA CTC CAT TTT GG3′). 引物分别含有 *Kpn* Ⅰ和 *Hind*Ⅲ, *Bgl*Ⅱ和 *Kpn* Ⅰ酶切位点. 以含全长 *HGLP* cDNA 的阳性克隆为模板, 用高保真的耐热 DNA 聚合酶 (Expand$^{TM}$ high fidelity PCR system, 购自德国 Boehringer Mannheim 公司)扩增. PCR 产物 HGLP-N, HGLP-C 分别连入 pQE40/41 载体, pQE40/41 含 DHFR 蛋白, 将 DHFR 及 DHFR-HGLP-N, DHFR-HGLP-C 重组序列切下克隆到 pET30a$^+$原核表达载体中. (2) 重组蛋白的表达及表达产物的 SDS-PAGE 电泳分析. 阳性质粒转入 BL21 菌后用 1 mmol/L IPTG 诱导表达, 12% 的 SDS-PAGE 检测表达蛋白, 考马斯亮蓝染色.

## 2 结果

### 2.1 cDNA 阳性克隆的筛选及其核苷酸序列与编码氨基酸的分析

从λZAP Express 胎脑文库随机挑取的克隆, 测序、序列分析后在 GenBank 库中比较发现了一个新序列的克隆, 其 cDNA 序列及其推测的氨基酸序列见图 1. 所测 cDNA 长度为 2286 bp, 编码一长 555 个氨

```
   1 AGA GGG TGA AAT GCT TTC GGT AGG CAC TCC ACG GCT GTG AAG ATG GCG GCG GCT GCG TGG CTT CAG GTG TTG  72
   1                                                       M   A   A   A   A   W   L   Q   V   L  10
  73 CCT GTC ATT CTT CTG CTT CTG GGA GCT CAC CCG TCA CCA CTG TCG TTT TTC AGT GCG GGA CCG GCA ACC GTA 144
  11   P   V   I   L   L   L   L   G   A   H   P   S   P   L   S   F   F   S   A   G   P   A   T   V  34
 145 GCT GCT GCC GAC CGG TCC AAA TGG CAC ATT CCG ATA CCG TCG GGG AAA AAT TAT TTT AGT TTT GGA AAG ATC 216
  35   A   A   A   D   R   S   K   W   H   I   P   I   P   S   G   K   N   Y   F   S   F   G   K   I  58
 217 CTC TTC AGA AAT ACC ACT ATC TTC CTG AAG TTT GAT GGA GAA CCT TGT GAC CTG TCT TTG AAT ATA ACC TGG 288
  59   L   F   R   N   T   T   I   F   L   K   F   D   G   E   P   C   D   L   S   L   N   I   T   W  82
 289 TAT CTG AAA AGC GCT GAT TGT TAC AAT GAA ATC TAT AAC TTC AAG GCA GAA GAA GTA GAG TTG TAT TTG GAA 360
  83   Y   L   K   S   A   D   C   Y   N   E   I   Y   N   F   K   A   E   E   V   E   L   Y   L   E 106
 361 AAA CTT AAG GAA AAA AGA GGC TTG TCG GGA AAT ATC CAA ACA TCA TCA AAA TTG TTC CAG AAC TGC AGT GAA 432
 107   K   L   K   E   K   R   G   L   S   G   N   I   Q   T   S   S   K   L   F   Q   N   C   S   E 130
 433 CTC TTT AAA ACA CAG ACC TTT TCT GGA GAT TTT ATG CAT CGA CTG CCT CTT TTA GGA GAA AAA CAG GAG GCT 504
 131   L   F   K   T   Q   T   F   S   G   D   F   M   H   R   L   P   L   L   G   E   K   Q   E   A 154
 505 AAG GAG AAT GGA ACA AAC CTT ACC TTT ATT GGA GAC AAA ACC GCA ATG CAT GAA CCA TTG CAA ACT TGG CAA 576
 155   K   E   N   G   T   N   L   T   F   I   G   D   K   T   A   M   H   E   P   L   Q   T   W   Q 178
 577 GAT GCA CCA TAC ATT TTT ATT GTA CAT ATT GGC ATT TCA TCC TCA AAG GAA TCA TCA AAA GAA AAT TCA CTG 648
 179   D   A   P   Y   I   F   I   V   H   I   G   I   S   S   S   K   E   S   S   K   E   N   S   L 202
 649 AGT AAT CTT TTT ACC ATG ACT GTT GAA GTG AAG GGT CCC TAT GAA TAC CTC ACA CTT GAA GAC TAT CCC TTG 720
 203   S   N   L   F   T   M   T   V   E   V   K   G   P   Y   E   Y   L   T   L   E   D   Y   P   L 226
 721 ATG ATT TTT TTC ATG GTG ATG TGT ATT GTA TAT GTC CTG TTT GGT GTT CTG TGG CTG GCA TGG TCT GCC TGC 792
 227   M   I   F   F   M   V   M   C   I   V   Y   V   L   F   G   V   L   W   L   A   W   S   A   C 250
 793 TAC TGG AGA GAT CTC CTG AGA ATT CAG TTT TGG ATT GGT GCT GTC ATC TTC CTG GGA ATG CTT GAG AAA GCT 864
 251   Y   W   R   D   L   L   R   I   Q   F   W   I   G   A   V   I   F   L   G   M   L   E   K   A 274
 865 GTC TTC TAT GCG GAA TTT CAG AAT ATC CGA TAC AAA GGA GAA TCT GTC CAG GGT GCT TTG ATC CTT GCA GAG 936
 275   V   F   Y   A   E   F   Q   N   I   R   Y   K   G   E   S   V   Q   G   A   L   I   L   A   E 298
 937 CTG CTT TCA GCA GTG AAA CGC TCA CTG GCT CGA ACC CTG GTC AGC ATA GTC AGT CTG GGA TAT GGC ATC GTC 1008
 299   L   L   S   A   V   K   R   S   L   A   R   T   L   V   S   I   V   S   L   G   Y   G   I   V 322
1009 AAG CCA CGC CTT GGA GTC ACT CTT CAT AAG GTT GTA GTA GCA GGA GCC CTC TAT CTT TTG TTC TCT GGC ATG 1080
 323   K   P   R   L   G   V   T   L   H   K   V   V   V   A   G   A   L   Y   L   L   F   S   G   M 346
1081 GAA GGG GTC CTC AGA GTT ACT GGG GCC CAG ACT GAT CTT GCT TCC TTG GCC TTT ATC CCC TTG GCT TTC CTA 1152
 347   E   G   V   L   R   V   T   G   A   Q   T   D   L   A   S   L   A   F   I   P   L   A   F   L 370
1153 GAC ACT GCC TTG TGC TGG TGG ATA TTT ATT AGC CTG ACT CAA ACA ATG AAG CTA TTA AAA CTT CGG AGG AAC 1224
 371   D   T   A   L   C   W   W   I   F   I   S   L   T   Q   T   M   K   L   L   K   L   R   R   N 394
1225 ATT GTA AAA CTC TCT TTG TAT CGG CAT TTC ACC AAC ACG CTT ATT TTG GCA GTG GCA GCA TCC ATT GTG TTT 1296
 395   I   V   K   L   S   L   Y   R   H   F   T   N   T   L   I   L   A   V   A   A   S   I   V   F 418
1297 ATC ATC TGG ACA ACC ATG AAG TTC AGA ATA GTG ACA TGT CAG TCG GAC TGG CGG GAG CTG TGG GTA GAC GAT 1368
 419   I   I   W   T   T   M   K   F   R   I   V   T   C   Q   S   D   W   R   E   L   W   V   D   D 442
1369 GCC ATC TGG CGC TTG CTG TTC TCC ATG ATC CTC TTT GTC ATC ATG GTT CTC TGG CGA CCA TCT GCA AAC AAC 1440
 443   A   I   W   R   L   L   F   S   M   I   L   F   V   I   M   V   L   W   R   P   S   A   N   N 466
1441 CAG AGG TTT GCC TTT TCA CCA TTG TCT GAG GAA GAG GAG GAG GAT GAA CAA AAG GAG CCT ATG CTG AAA GAA 1512
 467   Q   R   F   A   F   S   P   L   S   E   E   E   E   E   D   E   Q   K   E   P   M   L   K   E 490
1513 AGC TTT GAA GGA ATG AAA ATG AGA AGT ACC AAA CAA GAA CCC AAT GGA AAT AGT AAA GTT AAC AAA GCA CAG 1584
 491   S   F   E   G   M   K   M   R   S   T   K   Q   E   P   N   G   N   S   K   V   N   K   A   Q 514
1585 GAA GAT GAT TTG AAG TGG GTA GAA GAG AAT GTT CCT TCT TCT GTG ACA GAT GTA GCA CTT CCA GCC CTT CTG 1656
 515   E   D   D   L   K   W   V   E   E   N   V   P   S   S   V   T   D   V   A   L   P   A   L   L 538
1657 GAT TCA GAT GAG GAA CGA ATG ATC ACA CAC TTT GAA AGG TCC AAA ATG GAG TAA GGA ATG GGA AGA TTT GCA 1728
 539   D   S   D   E   E   R   M   I   T   H   F   E   R   S   K   M   E                           555
1729 GTT AAA GAT GGC TAC CAT CAG GGA AGA GAT CAG CAT CTG TGT CAG TCT TCT GTA CGG CTC CAT GGG ATT AAA 1800
1801 GGA AGC AAT GAC ATC CTG ATC TGT TCC TTG ATC TTT GGG CAT TGG AGT TGG CGA GAG GTG TCA GAA CAA AGA 1872
1873 GAA CAT CTT ACT GAA AAC AAG TTC ATA AGA TGA GAA AAA TCT ACG AGC TTC TTA TTT ACA ACA CTG CTG CCC 1944
1945 CCT TTC CTC CCA GAC TCT GAC ATG GAT GTT CAT GCA ACT TAA GTG TGT TGT TCC TGA ACT TTC TGT AAT GTT 2016
2017 TCA TTT TTT AAA TCT GAC AAA CTA AAA AGT TTA ACG TCT TCT AAA AGA TTG TCA TCA ACA CCA TAA TAT GTA 2088
2089 ATC TCC AGG AGC AAC TGC CTG TAA TTT TTA TTT ATT TAG GGA GTT ACA TAG GTG ATG GGG GAA ATT GTT AAC 2160
2161 TAC CTT TCA TTT TCC TGG GAA GTC AAG GTT ACA TCT TGC AGA GGT TGT TTT GAG AAA AAA GGG CCC TTC TGA 2232
2233 GTT AAG GAG CCA TAG TTC TAT CAA TGA TCA AAA GAA AAA AAA AAA AAA AAA AAA                       2286
```

图 1 *HGLP* cDNA 序列及推测的氨基酸序列图

下画线部分表示 *HGLP* 具有的视紫红质样的 G 蛋白偶联受体超家族基序

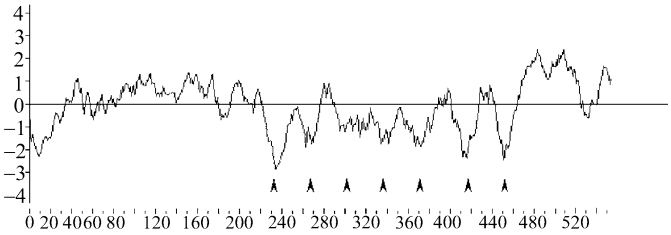



图 2 HGLP 蛋白亲疏水性示意图
箭头所指为推测的穿膜区

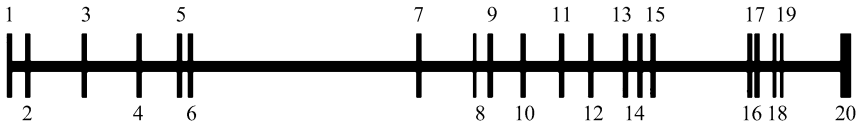



图 3 *HGLP* 基因组结构示意图

基酸、分子量为 63.3 ku 的蛋白(GenBank 编号为 AF132733). 根据可读框的搜寻, 确定位于 43 ~ 45 bp 的 ATG 是翻译起始点, 1705 ~ 1707 bp 的 TAA 是翻译终止密码子. 在可读框 5′上游有一个框内终止密码. 亲疏水性分析发现 *HGLP* 有 7 个穿膜区(图 2). 比较蛋白的 motif, 发现我们所克隆的 cDNA 编码的蛋白序列中具有许多视紫红质样的 G 蛋白偶联受体超家族的 motif (图 1), 将其归类为视紫红质样的 G 蛋白偶联受体超家族.

### 2.2 *HGLP* 的染色体定位

随着人类基因组计划的发展, 人基因组序列草图可由公共数据库获得. 利用相应软件, 将 *HGLP* 与人基因组序列比较发现 *HGLP* 基因定位于人染色体 15q15 的大约 62 kb 的范围里, 具有 20 个外显子(图 3).

### 2.3 组织表达谱分析

多组织 Northern 杂交显示 *HGLP* 只有一个转录本, 大小约为 3 kb. 说明我们克隆的 cDNA 并不全, 但根据框内 5′终止密码子可判定编码区已全. 该基因在成人 8 种组织中均有表达, 其中以胰腺、肺和胎盘较多(图 4).

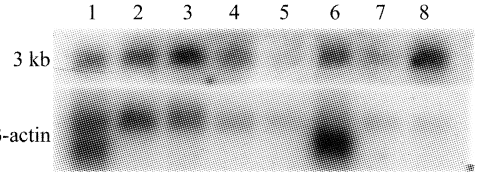



图 4 *HGLP* 多组织 Northern blot
1 ~ 8 依次示成人组织: 心、脑、胎盘、肺、肝、骨骼肌、肾、胰腺

### 2.4 原核表达

将含 DHFR, DHFR-HGLP-N 和 DHFR-HGLP-C 片段的 pET30a$^+$阳性质粒转入 BL21 菌中, IPTG 诱导表达融合蛋白. 12%SDS-PAGE 结果如图 5. 第 6, 8 道可见有大量的外源蛋白. 参照蛋白分子量标准, 外源蛋白大小与 HGLP 的 N, C 端多肽分子量(19.6 和 9.8 ku)加上 DHFR 蛋白分子量(26 ku)之和(45.6 和 35.8 ku)大致相同, 可见 IPTG 诱导后表达的外源蛋白就是我们的目的蛋白: *HGLP* 的 N, C 端融合蛋白.

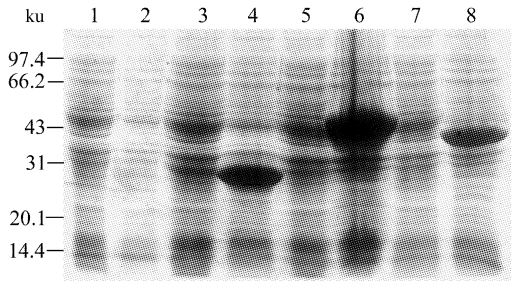



图 5 HGLP N, C 端融合蛋白 12%SDS-PAGE 图
1, 2 示转入 pET30a$^+$质粒的 BL21 菌诱导前后; 3, 4 示 DHFR 阳性质粒诱导前后; 5, 6 示 DHFR-HGLP-N 阳性质粒诱导前后; 7, 8 示 DHFR-HGLP-C 阳性质粒诱导前后

## 3 讨论

HGLP 蛋白具有 7 个跨膜区, 这类似于 G 蛋白偶联受体的结构. 虽然对于 G 蛋白偶联受体整个家族来讲, 氨基酸序列几乎没有什么保守性, 但通过寻找进化发育上相关的基序(motif), 可以将新的成员归类. Motif 分析将 HGLP 归为视紫红质样的 G 蛋白偶

联受体超家族, 推测 HGLP 是一个 G 蛋白偶联受体.

G 蛋白偶联受体在细胞功能中发挥重要作用. 它可以介导人体多种激素的信号传递: 如肾上腺皮质激素 (epinephrine)、卵泡刺激素 (FSH)、甲状旁腺素 (parathyroid hormone)、甲状腺刺激激素(TSH)和垂体加压素(vasopressin)等[4, 5]. 它可以介导人体的嗅觉、味觉和视觉[6~8]. 因为嗅觉受体、视紫红质及 3 种色觉视锥细胞受体都是属于 G 蛋白偶联受体. 此外, 许多神经肽类和神经递质的受体也是G蛋白偶联受体[9,10]. 推测有的 G 蛋白偶联受体还与发育有关[11]. 这些 G 蛋白偶联受体和相应配体结合后, 激活 G 蛋白, 然后经第 2 信使传递信息, 引起细胞内相应的生理反应.

G 蛋白偶联受体除了在生理活动中扮演着重要的角色外, 它与疾病的关系也很密切. G 蛋白偶联受体的突变可以造成多种疾病如尿崩症[12], 最近研究表明 HIV-1 病毒侵入 T 细胞[13]和药物的成瘾性[14]也与 G 蛋白偶联受体有关. 由于测序技术的发展, 新发现的G蛋白偶联受体超家族的新成员越来越多, 而对于新克隆的 G 蛋白偶联受体来讲, 其中有许多是配体和功能并不明确的孤儿G蛋白偶联受体(orphan GPCR), 确定孤儿G蛋白偶联受体的性质并寻找与其偶联的配体是目前研究的热点.

近年发展的反向药理学方法已成功应用于孤儿G蛋白偶联受体的研究. 即将推测的孤儿G蛋白偶联受体在哺乳动物细胞或其他合适的系统中表达, 在组织提取物、多肽文库或化合物中筛选相应的配体, 利用此配体探讨孤儿 G 蛋白偶联受体的生理、病理功能并研究配体的药理价值. 降钙素基因相关肽受体(CGRP)、过敏毒素 C3a 和 GPR10 受体等就是通过反向药理学方法鉴定的[15].

*HGLP* 是一个来自胎脑文库、编码蛋白结构类似G蛋白偶联受体的基因. 人脑中存在多种G蛋白偶联受体, 发挥着广泛的作用. *HGLP* 在多种组织中广泛分布, 提示 *HGLP* 不仅在人脑中, 而且在其他系统发挥重要作用. 反向药理学方法将是 *HGLP* 鉴定和配体寻找的有力工具.

## 参 考 文 献


1 Strader C D, Fong T M, Graziano M P, et al. The family of G-protein coupled receptors. FASEB J, 1995, 9: 745 ~ 754

2 Altschul S F, Madden T L, Schaffer A A, et al. Gapped BLAST and PST-BLAST: a new generation of database searching programs. Nucleic Aci Res, 1997, 25: 3389 ~ 3402

3 Nakai K, Kanehisa M. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics, 1992, 14: 897 ~ 911

4 Bisello A, Greenberg Z, Behar V, et al. Role of glycosylation in expression and function of the human parathyroid hormone/ parathyroid hormone-related protein receptor. Biochemistry, 1996, 35(49): 15890 ~ 15895

5 Ingram C D, Ciobanu R, Coculescu I, et al. Vasopression neurotransmission and the control of circadian rhythms in the suprachiasmatic neucleus. Prog Brain Res, 1998, 119: 351 ~ 364

6 Lundstron I, Svensson S. Biosensing with G-protein coupled receptor systems. Biosens Bioelectron, 1998, 15(16): 689 ~ 695

7 Ruiz-Avila L, Mclaughlin S K, Wildman D, et al. Coupled of bitter receptor to phosphodiesterase through transduction in taste receptor cells. Nature, 1995, 376(6535): 80 ~ 85

8 Shichida Y, Imai H. Visual pigment: G-protein coupled receptor for light signals. Cell Mol Life Sci, 1998, 54(12): 1299 ~ 1315

9 Wieland H A, Eckard C P, Doods H N, et al. Probing of the neuropeptides Y-Y1-receptors interaction with anti-receptor antibodies. Eur J Biochem, 1998, 255(3): 595 ~ 603

10 Vallano M L. Development aspects of NMDA receptor function. Curr Rev Neurobiol, 1998, 12(3): 177 ~ 204

11 Ciko S. Sequence and tissue distribution of a novel G-protein coupled receptor expressed prominently in human placenta. Biochem Biophys Res Commun, 1999, 256(2): 352 ~ 356

12 Birnbaumer M. Mutation and disease of G-protein coupled receptors. J Recep Sig Trans Res, 1995, 15: 131 ~ 160

13 Unntmaz D, Kewalramani V N, Littman D R. G-protein coupled receptor in HIV and SIV entry: new perspectives on lentivirus-host interactions and on the utility of animal models. Semi Immun, 1998, 10: 225 ~ 236

14 Roth B L, Willins D L, Kroeze W K. G-protein coupled receptor (GPCR) trafficking in the central nervous system: relevan- ce for drugs of abuse. Drug Alcohol Depend, 1998, 51(1-2): 73 ~ 85

15 Wilson S, Bergsma D J, Chambers J K, et al. Orphan G-protein coupled receptor: the next generation of drug target? Brit J Pham, 1998, 125: 1387 ~ 1392


(2000-07-06 收稿, 2000-11-13 收修改稿)