

影响外源基因在原核中表达水平因素分析

吴加金 李伍举 王嘉玺

(军事医学科学院基础医学科学院 北京 100850)

摘要 为了预测外源基因在 *E. coli* 中的表达水平,我们分析外源基因在固定载体 pBV220 中表达水平的实验数据,结合有关参考资料,提出预测外源基因表达水平的一组判据:在 AUG 附近的 RNA 二级结构能量, AUG 附近的密码子自适应指数, RBS 和 AUG 距离及 AUG 附近的构象。在表达实验前可利用这些判据对外源基因作改造以获得高效表达。

关键词 基因表达, 二级结构, 计算机

如何实现外源基因在原核细胞或真核细胞中的表达是 DNA 重组技术的关键性环节。自 60 年代以来,对影响外源基因在其表达体系中表达效率的各个因素作了大量实验研究,并已发表几篇归纳性评论^[1~3]。

国内针对外源基因在原核细胞高效表达的关键因素,构造高效表达载体^[4],并在此载体上,成功表达一系列细胞因子的基因^[5~12]。从这些表达研究的结果可以发现,即使选定菌种和有了高效表达载体,要想获得高效表达产物还必须对外源基因作适当改造,如优化 AUG 前后的碱基序列,调整 AUG 的位置,局部改造密码子,使之适合表达菌种的偏性,改造基因 3'端侧翼区基因的调控序列等。但以往对外源基因的改造是定性,较难预测改造效果或有针对性地改造外源基因。为此我们对外源基因在原核细胞中表达实验的部分国内外资料作了研究分析,力图寻找影响外源基因在原核细胞表达的各有关因素的定量指标,作为外源基因改造的判据。实现通过计算机的软件,有针对性地对外源基因作实验前的改造模拟。

1 参量的选择

为实现通过计算机的软件模拟,判断对外源基因改造的效果,必须选择某些参量,它们和外源基因的表达和转录水平有较显著的定量关系。根据对已有实验资料分析,现选择如下几个参量。

1.1 外源基因 5'端 AUG 附近的 RNA 二级结构能量

Smit DE^[13]对核糖体结合部位的二级结构对转录效率的影响作了较系统的定量研究,他通过改变这个区段碱基配对的方式,以碱基突变方式改变二级结构能量,然后实验测定其相对表达水平。从实验结果,他得出在 AUG 前后(从 -20 至 +18 的基因片段)二级结构的稳定性和相对表达水平明显相关。其它一些报道也得出类似结论,国内实现外源基因高效表达的实验研究中,对外源基因改造也都在这区段,因此选用 AUG 前后基因

片段(包括外源基因 5'端和载体中 SD 序列部分)的二级结构能量作为评价外源基因改造的判据。

1.2 AUG 的位置及其附近的构象

乐树云^[14]等对 λ 菌体左臂可能编码的 25 个蛋白质中 11 个不同基因的蛋白合成产率不同的实验事实,并根据 Gold 等^[1]提出位于起始密码子上游第 25 核苷酸残基起至距起始密码子下游的 20 个核苷酸残基(-25~+20)的这区段对 mRNA 翻译的起始有着重要控制作用。因此他们选 -35~+35 区段作出 RNA 二级结构图并分析在这些图形中 AUG 的位置,得出 AUG 在单链区(包括发夹环区,内部环区和外膨胀区),因这些区段 AUG 处于暴露的部位,易被缔合的接受状态,故其蛋白合成效率高。相反若 AUG 处于碱基互补的螺旋区段蛋白产率低。马贤凯等^[15]在研究乙型肝炎病毒核心抗原基因结构对其在大肠杆菌中表达的影响时,仔细研究了 AUG 与核糖体结合位点的距离与构象对表达的影响,得出 AUG 上游的部分回文序列对该基因的表达影响很大,马大龙^[3]和 Gold^[2]的文献中指出 SD 序列到 AUG 的最佳距离为 $9 \pm 3\text{bp}$ 。在同样启动子时,这个距离的不同对 IL-2 表达水平可相差 2~2000 倍,因此可从 RNA 二级结构的图形显示中,判断 AUG 的位置和构象,作为外源基因改造的判据。

1.3 密码子的适配指数

Sharp^[16]提出密码子利用适配指数(Codon adaptation Index),并给出计算公式及相应参数表。

Williams^[17]按照密码子利用偏性对 IL-2 的 mRNA 作突变,使基因中偏性密码子比例从 43% 增加到 85%,获得比其天然 mRNA 序列高 16 倍的表达。根据上面结果,选密码子的适配指数作为另一个判据。

1.4 外源基因 3' 侧翼稳定性

为了外源基因翻译的有效终止,在外源基因的 3' 往往改造为大肠杆菌偏性终止密码子 TAA,达到有效地防止翻译过程中的通读。并有效地删除 3' 端侧翼非编码区对转录的负调控的影响。暂无较定量的判据。

2 实验数据的选取

为使比较分析在相对固定的外界条件下进行,按下述原则选取实验数据:

2.1 载体相同并载体序列已知。为此选属于细胞介素类外源基因在 pBV220 载体内的表达的实验数据^[5~12]。

2.2 根据乐树云^[14]所引用的 Gold 等^[1]的观点。首先按照实验研究报告中关于表达质粒构建原则,用电子计算机的软件-Goldkey 软件系统,构建表达质粒,然后截取其(-25~+20)的基因片段作为分析计算的原始数据。

2.3 表达效率按表达产物占菌体总蛋白的百分数,从实验报告上获取^[5~12]。由于不同实验室可能存在系统误差,因此选用这些数据时力争在选取近似相同的条件。

3 实验数据分析方法

使用 Goldkey 软件中用于 RNA 二级结构能量计算、二级结构图形显示和密码子使用

指数等程序模块,对选中的实验数据作分析。

用 RNA 螺旋区随机折叠的算法^[18]计算 RNA 二级结构能量及其构象,随机计算次数为 100 次。选择其中自由能最低的 RNA 构象。

4 结果

对已发表的细胞介素类的因子在 pBV220 载体质粒中的表达的实验结果的数据,按外源基因 5'端二级结构能量 E,密码子适配指数 CAI,起始密码子 AUG 的位置及构象状态,3'端状态,表达产额及对外源基因改造的侧重点等项目,列于表 1。AUG 附近的二级结构图汇总于图 1。

表 1 细胞介素类的因子在 pBV220 载体中表达水平和有关参数的关系

Table 1 Relation between level of expression of cytokins in vector pBV220 and Some factors

No.	Code	Free Energy (J)	Adaptation Index	Configuration Near AUG	Products/ %	Segments modified	Ref.
1	IL6-O	-43.493	0.0778	In loop	No expression	Cut 3'-terminal, use TAA,	(5)
2	IL6-R	-4.600	0.0447	"	71	Modify 5'-Term. Based on en-	(5)
3	IL6-T	-21.746	0.0441	"	28	ergy Cut 3'-terminal, use TAA	(8)
4	HIL4-L	-27.183	0.0781	"	30~40	Cut 3'-terminal	(12)
5	MIL4-W	-10.873	0.0601	"	25~30	Cut 3'-terminal	(11)
6	HIV1-L	-32.620	0.0631	"	8		(10)
7	GM-O	-20.910	0.1041	"	little		(4)
8	GM-Z	-44.747	0.1596	"	20	Modify 5'-term. adaptation index	(4)
9	NIL-J	-5.437	0.0551	"	18.5	Modify 5'-term. based on adap- tation index energy and use double termination signal	(7)
10	HIL2-Z	-17.146	0.1460	"	20		
11	IFN-Z	-19.237	0.0719	"	24		
12	GM-W	-19.655	0.2239	"	49	Cut 3'-terminal, Modify 5'- term. based on energy and use TAA	(19)

从表 1 的结果可见:

在其它条件相同时,表达产额和 5'端二级结构能量呈明显相关,如表 1 的数据 1, 2, 3, 这三个数据的 3'端条件完全一样,其中第 3 号实验只侧重去 3'端侧翼,对 5'未给予改造而 2 号数据的 3'端和 3 号一样,但又对 5'端作了改造,使其二级结构能提高,从而获得较高表达。

而未作改造(包括 3'和 5'端)即 1 号数据,未能获得有效表达。但从 1 号和 2, 3 号数据比较,3'端非编号区的删除可能作用更大。另外,表 1 的全部数据,除了几个原始的序列外都获得高表达,这些高表达的基因其 3'端侧翼的非编码区都被删除,5'端 AUG 附近的二级结构能量都较高。

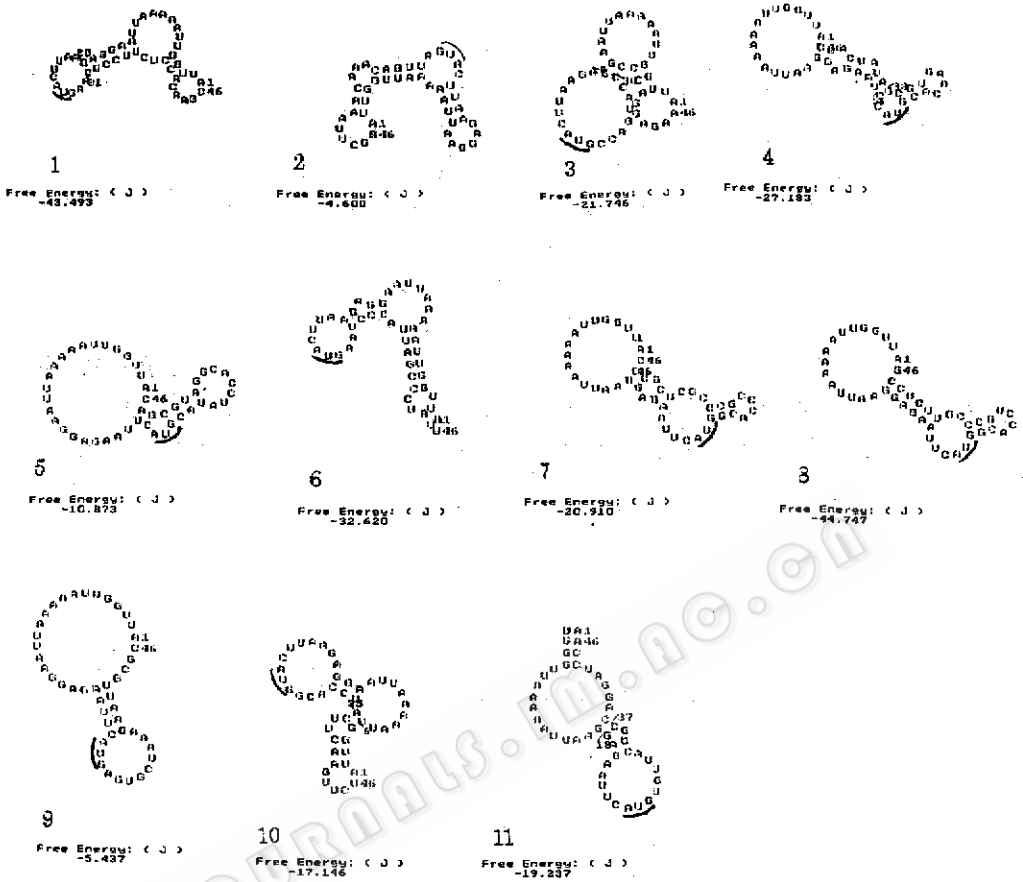


图 1 AUG 附近的二级结构图

Fig.1 Secondary structure map near A U (

在 AUG 附近的密码子偏性也至关重要。

从表 1 的数据 7, 8 比较可见, 8 号数据对 5' 端的密码子偏性作了改造, 其密码子适配指数比未加改造的 7 号高, 从而获得明显高表达。从密码子适配指数分配, 其它的基因其适配指数都较低, 如果在对基因改造时也重视这个因素, 可能会获得更高一些的表达。而 7 号和 8 号比, 因作者只重视密码子偏性, 对二级结构能量未加以考虑, 所以改造后二级结构能量降低, 对高效表达不利。如果在改造密码子偏性时, 也考虑能量因素, 可能会获得更高表达, 例如表 1 的数据 12。

从图 1 可见 AUG 位置都在环区, 符合乐树云^[4]的结论, 并且距核糖体结合部位的距离都在 6~7bp, 因此可把 AUG 位置作为一个因素在改造外源基因时通过二级结构的图形显示加以考察。

5 结论

外源基因在大肠杆菌中表达水平受多种因素影响,有基因内部因素如启动子强度,载体性质,基因的高级结构及载体菌对密码子的适应性等,为了避免上述多种因素同时考虑引起分析的困难,我们只针对同一载体——pBV220 来研究外源基因对表达水平的影响。把与载体有关的因素相对固定,只考虑外源基因与载体结合后的有关因素。尽管如此还有一些外部因素,如:菌株类型,培养基类型与浓度,诱导温度与诱导时间等对表达水平也都有影响,但是本文所选数据都取自公开发表的文献,这些数据是原作者对上述因素优化之后而获得的,因此在表 1 所列数据都是国内在具有双启动子(P_R 、 P_L)的高效表达载体 pBV220 中,在大肠杆菌获得的高效表达结果,但因数据来自不同实验室和不同作者,存在系统误差,并且数据量太少,不足以建立某个参量和表达产额之间的定量关系。但从改造外源基因的判据观点,可明显看出,按 5'端 AUG 附近的二级结构能量、AUG 附近密码子偏性指数, AUG 的位置及附近的构象及 3'端非编码区的负调控序列的删除、强终止密码子的引入等,可做为外源基因改造的判据,实现表达实验前对基因改造计算机模拟。

参 考 文 献

- [1]Gold L., Pribnow D. Schneider T *et al.*, *Ann Rev Microbiol* 1981, **35**:365~403.
- [2]Gold L. *Methods in Enzymology*, 1990, **185**:11.
- [3]马大龙. 国外医学分子生物学分册, 1989, **11**:56.
- [4]张智清,姚立红,侯云德. 病毒学报, 1990, **6**:102.
- [5]任启生,王嘉玺,邹民吉等. 中国免疫学杂志, 1993, **8**:133~135.
- [6]张智清,张颖等. 病毒学报, 1993, **9**:136~142.
- [7]金冬雁,徐荣辉,周圆等. 中国科学(B辑), 1992, **22**(9):951~957.
- [8]田季德,李晨,陈慰峰. 中国免疫学杂志, 1991, **7**:331~336.
- [9]李晨,张智清,田季德等. 中华微生物学和免疫学杂志, 1991, **11**:327~329.
- [10]李满,陈辉等. 生物化学杂志, 1993, **19**:529~532.
- [11]王锦娟,田季德,李晨等. 中国免疫学杂志, 1992, **8**:9~12.
- [12]李晨,张智清,郑蕾等. 中华微生物学和免疫学杂志, 1990, **10**:137~140.
- [13]MaaRTEN H, Smit D E, Duin J V. *Proc Natl Acad Sci USA*, 1990, **87**: 7668~7672.
- [14]乐树云,陆长德,江寿平. 生物化学与生物物理学报, 1985, **17**:664~671.
- [15]马贤凯,丁广治,石成华. 中国科学(B辑), 1987, (1):61~68.
- [16]Paul M S, Li W H. *Nucleic Acids Research*, 1987, **15**:1281~1295.
- [17]Williams D P *et al.* *Nucl Acids Res*, 1988, **16**:10453~10467.
- [18]李伍举,吴加金. 生物物理学报, 1996, **12**: (2)213~218.
- [19]王嘉玺,邹民吉,黄碧莲等. 生物工程学报, 1995, **11**:33~38.

Analysis of the Factors of the Level of Expression of a Heterologous Gene in Prokaryocyte

Wu Jiajin, Li Wuju, Wang Jiaxi

(*Institute of Basic Medical Sciences, Academy of Military Medical Sciences, Beijing 100850*)

Abstract In order to predicting level of heterologous gene expression in *E. coli*, We analyze some experimental data of level of heterolous gene expression in vector pBV220. According to some references, we suggest a set of judgements; the energy of local RNA secondary structures, the local codon adaptation index, the spacing between the RBS and the AUG, configuration in near AUG. We can use those judgements to modify heterologous gene before experiment.

Key words Gene expression, secondary structure, computer