

血清抑制基因 Si1 编码蛋白的亚细胞定位研究 *

赵文秀^{1)**} 伍红^{1,2)**} 谭德勇^{1)***}

(¹⁾云南大学生物化学与分子生物学实验室, 昆明 650091; ²⁾西南民族大学生命科学与技术学院, 成都 610041)

摘要 为进一步研究血清抑制相关基因(Si1)的生物学功能, 用绿色荧光蛋白 EGFP 融合的 EGFP/Si1 重组蛋白在 HeLa 细胞中成功表达。结果显示该蛋白定位在细胞核中, 这一结果与生物信息学预测结果相符合。通过构建不同长度的 EGFP/Si1 基因突变体观察其核定位信号所在区段, 发现核定位信号区在 465~531 氨基酸残基之间。Si1 基因编码框中肿瘤相关 1 639 位点的改变, 虽然并不在入核信号相关结构区, 但影响 Si1 基因表达蛋白的细胞核迁移, 进而影响基因功能。

关键词 Si1 基因, 亚细胞定位, 核定位信号, 信号区, 突变

学科分类号 R73, Q756

DOI: 10.3724/SP.J.1206.2008.00758

血清抑制相关基因(Si1)是本实验室通过比较血清培养细胞和血清饥饿细胞的基因表达差异得到的一段在血清饥饿条件下特异表达的 cDNA 序列, 由此序列出发, 通过搜索表达序列标签(EST), 拼接出完整的基因序列, 通过 PCR 分段克隆获得全长 cDNA 序列。该基因 cDNA 全长 5 429 bp, 预测编码框有 791 个氨基酸残基, 共有 21 个外显子 (GenBank 接受号: AY050169)^[1]。研究表明, Si1 一个多态位点的突变与肿瘤有关, 并在大肠癌中有显著的突变频率^[2-3]。同时, Si1 基因还与细胞周期的调控相关(unpublished data), 但其结构以及功能的相关研究至今知之甚少。

与血清抑制相关的基因多与细胞周期、细胞核内转录和调控有关^[4,5], 蛋白质在亚细胞结构中的定位对其功能的研究具有指导性的意义, 为进一步探讨 Si1 基因生物学功能和结构基础, 本研究通过生物信息学分析和 EGFP/Si1 融合蛋白定位分析, 确定了 Si1 蛋白的亚细胞定位。同时, 根据生物信息学分析结果构建不同的 EGFP/Si1 突变体以及与 Si1 基因编码框 1 639 突变位点的突变型 EGFP/Si1, 对 Si1 蛋白细胞核定位相关结构进行了分析。信息学分析结果表明, 在 Si1 基因表达蛋白中存在细胞核定位信号(nuclear location signals, NLS), 实验观察细胞定位情况发现, Si1 表达蛋白确实定位在细胞核, 但 NLS 信号区域与预测不符, 并且发现突变位点虽然并不在 NLS 信号区, 但仍然影响了 Si1

基因表达蛋白的细胞核迁移。

1 材料和方法

1.1 材料

pEGFP-N2 质粒购于 Clontech 公司。细胞转染试剂 Lipofectamine 2000 试剂盒购自 Invitrogen 公司。胶回收试剂盒购自 Promega 公司。U251、HeLa 细胞株由本实验室保存。PMD 18-T VECTOR, T4 连接酶及 *Bam*H I、*Hind*III、*Sac* I、*Eco*R I、*Sal* I、*Eco*O109 I (*Dra* II) 内切酶均购自大连宝生物 (TaKaRa) 公司。TRI REAGENT 购自晶美生物工程有限公司。逆转录试剂盒购自 Bio Basic Inc。其他生化试剂均购自上海生工生物工程有限公司。PCR 扩增试剂由 Sigma 公司提供。PCR 扩增引物由上海生工生物工程有限公司合成。生化试剂由上海生工生物工程有限公司提供。测序由 Invitrogen 公司完成。

1.2 方法

1.2.1 人类血清抑制基因 Si1 的克隆。 Si1 基因编码框(GSF)通过对 U251 细胞的 RT-PCR 分两段

* 国家自然科学基金资助项目(39960030, 30360040, 30760057)和云南省科技计划项目(2008CC003)。

** 共同第一作者。*** 通讯联系人。

Tel: 0871-5032061, E-mail: dytan@ynu.edu.cn

收稿日期: 2009-02-03, 接受日期: 2009-04-22

(GSFA 和 GSFB)获得, 引物序列分别为: GSFA 正义, 5' AAGCTTAATGCCACTCTATGACAATC 3', 反义, 5' TCTCATCCTGAATACATCTGTCCACC 3'; GSFB 正义, 5' GTCATCCCTGTAGCCACCCATCT 3', 反义, 5' GGATCCGGCTGCTTGCCAACTTAGCG-AGG 3'. PCR 反应程序: 94℃ 2 min, 94℃ 30 s, 59℃ 30 s, 72℃ 2 min, 35 个循环, 72℃ 10 min, 通过 *EcoR* I 酶切并拼接得到 Si1 基因完整开放阅读框并克隆入 PM18D-T-VECTOR 载体中, 得到 T-VECTOR/Si1, 连接方法见 TAKARA T4 连接酶说明书, 测序鉴定.

1.2.2 Si1 基因表达蛋白亚细胞定位的生物信息学分析. 用生物信息学软件分析 Si1 基因表达蛋白的亚细胞定位(<http://www.http://www.psорт.org/>), 亚细

胞定位(subcellular localization)专项软件 Gpos-ploc, Psort II prediction, Cello version2, 预测蛋白质的亚细胞定位^[6~8]. SingalIP 3.0 预测信号肽^[9], 细胞核定位信号(nuclear location signals, NLS)预测软件 NUCLEO, NUCPRED, Proloc, PredictNLS 与 Subnuclear compartments prediction system 预测细胞核定位^[10~14].

1.2.3 EGFP/Si1 融合表达载体(GSF)及 EGFP/Si 基因半分子突变体表达载体的构建. 选用酶切位点 *Bam*H I, *Hind*III 将 Si1 基因克隆入 pEGFP-N2 载体中, 获得融合表达载体 EGFP/Si1. 根据预测结果设计不同长度的 Si1 基因突变体(G-TS1~8, GSI5'a~c)并选用相同酶切位点克隆入 pEGFP-N2 载体中, 引物见表 1.

Table 1 Primers of Si1 half-molecular clone

Product number	Primer	Size of PCR product/bp
GSI5'a	5' AAGCTTAATGCCACTCTATGACAATC 3'	1 126
	5' GGATCCCATCCTGAATACATCTGTCCACCAT 3'	
GSI5'b	5' GTCAGAATTCCTTCAATGTCACCC 3'	437
	5' GGATCCCAGAGGTCACCTGTGGTCACTTTGGT 3'	
GSI5'c	5' GTCAGAATTCCTTCAATGTCACCC 3'	818
	5' GGATCCCAGAAGAACTGGACTTGGCTGTGGT 3'	
GSF	5' AGCTTAAATGCCACTCTATGACAATC 3'	2 373
	5' GGATCCGGCTGCTTGCCAACTTAGCGAGG 3'	
G-TS1	5' AAGCTTGATgAAGGTGTCCTTATCTCCCA 3'	778
	5' GGATCCCGGTGCTGGAGGCAGGCGCTGTGAG 3'	
G-TS2	5' AAGCTTTATgCAGGATGAGATTGTGGACTTT 3'	1 256
	5' GGATCCCGGTGCTGGAGGCAGGCGCTGTGAG 3'	
G-TS3	5' AAGCTTCaTGGCCTTTGAAGTCCCTGAGCGG 3'	1 131
	5' GGATCCCGGTGCTGGAGGCAGGCGCTGTGAG 3'	
G-TS4	5' AAGCTTCaTGGCCTTTGAAGTCCCTGAGCGG 3'	152
	5' cgGGATCCCAGAGGTCACCTGTGGTCACTTTGGT 3'	
G-TS5	5' AAGCTTCaTGGCCTTTGAAGTCCCTGAGCGG 3'	353
	5' cgGGATCCTCGGTCGTTTGGAGGTCTTACTGCC 3'	
G-TS6	5' aagggAAGCTTGAiGTCCAGTCAGATTGGAAGTTCT 3'	199
	5' cgGGATCCTCGGTCGTTTGGAGGTCTTACTGCC 3'	
G-TS7	5' aagggAAGCTTCATgTCATCAGAACCACAGAGCA 3'	109
	5' cgGGATCCTCGGTCGTTTGGAGGTCTTACTGCC 3'	
G-TS8	5' aagggAAGCTTGAiGTCCAGTCAGATTGGAAGTTCT 3'	127
	5' cgGGATCCCTCTTTCTCTCCTTCTCTGGTGG 3'	

1.2.4 EGFP/Sit 融合表达载体的构建. 采用 PCR 点突变法获得 Si1 突变体, 设计 4 条引物, 引物序列如表 2. 将待突变位点设计于 2 条互补引物之中, 而另外 2 条引物分别位于待克隆片段的两端. 第一次 PCR 获得两个带有单个突变位点的片段, GSM- I (414 bp)和 GSM- II (138 bp), 两个片段有 20 个碱基的重合序列, 于是利用第二次 PCR 将两个

片段连接起来, 命名为 GSM. PCR 反应体系为: 第一步, 双蒸水 15.0 μl, 10×缓冲液 2.5 μl, 10 mmol/L dNTP 1.0 μl, MgCl₂ 2.0 μl, DNA(GSM I)2.0 μl, DNA (GSM II) 2.0 μl, Taq DNA 聚合酶(5 U/μl) 0.5 μl; 第二步, 双蒸水 15.0 μl, 10×缓冲液 2.5 μl, 10 mmol/L dNTP 1.0 μl, MgCl₂ 2.0 μl, 第一步反应液 25 μl, Taq DNA 聚合酶 0.5 μl. PCR 反应程序为: 第一

步, 94℃ 5 min, 94℃ 1 min, 58℃ 5 min, 72℃ 2 min, 35 个循环, 72℃ 10 min; 第二步, 94℃ 3 min, 94℃ 1 min, 60℃ 2 min, 72℃ 2 min, 35 个循环, 72℃ 10 min. GSM 用 *Sac* I 和 *Sal* I 酶切后克隆入 EGFP/Si 中获得 EGFP/Si1.

Table 2 Primers of Si1 mutation

Product number	Primer	Size of PCR product /bp
GSM- I	5' GCA GAG ACT TGG CCT TTG AA 3' 5' GAG CAC AAG ACC CTC CAG CT 3'	414
GSM- II	5' AGC TGG AGG GTT TTG TGC TC 3' 5' GAA CTG GAC TTG GCT GTG GT 3'	138

1.2.5 融合蛋白在 HeLa 细胞中的表达及定位.

HeLa 细胞用胰蛋白酶消化后, 按 1:3 传代, 加入 1640 培养基, 10% 新生牛血清, 5% CO₂, 37℃ 条件下培养 24 h, 用脂质体转染. 转化用质粒 DNA 为 5 μg 每 25 ml 培养瓶. 转化方法按 Lipofectamine 2000 操作说明进行. 转化后培养 24 h, 在激光共聚焦显微镜 (laser scanning confocal microscope, LSCM) 下观察荧光, 激发光波长 507 nm, 发射光波长 488 nm, 同时加上可见光以观察细胞本身的结构, 图像获得和处理由 Zeiss, LSM, Image Brower 4.0 完成. 由于目的基因接在 EGFP 蛋白基因的 5' 端, 因此 EGFP 蛋白的荧光所在位置即为目的基因表达产物的位置.

1.2.6 DNA 测序. 重组克隆及全部融合表达载体的测序鉴定由 Invitrogen 公司完成.

1.2.7 数据统计. 利用生物学统计软件 Image pro 6.0, 统计转染后 EGFP 融合蛋白发出荧光在细胞中定位的比率, 所得数据为 3 次重复实验统计 3 个视野所得的平均值.

2 结 果

2.1 Si1 基因的克隆与鉴定

采用分段克隆加拼接的方法得到 Si1 基因的编码框的全长序列, RT-PCR 从 HeLa 细胞中分别获得片段 GSF-A 和 GSF-B, 通过酶切位点 *Eco*R I 连接获得 Si1 编码框去终止子序列 GSF(Si1 full). 拼接过程所得 PCR 产物如图 1. 各片段大小与预计片段一致. 然后将 SF 片段克隆到 T 载体 pMD-18 上获得重组质粒 pMD-18-GSF, 经测序确定所得克隆序列是正确的. 各片段电泳如图 1.

(a) M GSF-A GSF-B GSF

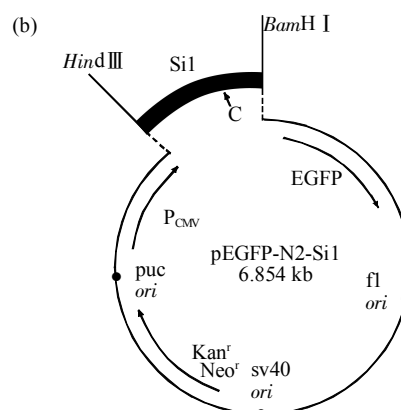
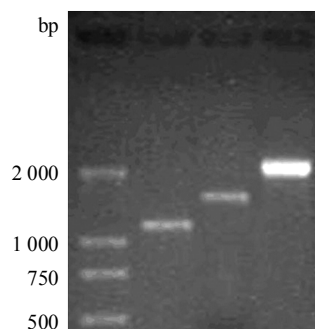


Fig. 1 Clone of Si1 gene

(a) Full frame gene of Si1 gained from RT-PCR, gel electrophoresis. M: Molecular mass marker. (b) Design for pEGFP-N2-Si1 recombinant.

2.2 Si1 蛋白核定位信号的生物信息学 分析

Cello version 2, Psort II prediction 结果如表 3, 预测结果都表明, Si1 基因表达蛋白定位在细胞核内的分值极高, 分别为 0.726% 和 94.1%;

Table 3 Cello version 2, Psort II prediction results

Cello results		
SVM	Localization	Reliability
Amino acid comp	Nuclear	0.576
N-peptide comp	Nuclear	0.623
Partitioned seq	Nuclear	0.726
Physico-chemical	Nuclear	0.563
Neighboring seq	Nuclear	0.386
	Nuclear score	2.874*
Psort II server results		
ER retention motif in the C-terminus	none	
ER membrane retention signals	none	
Peroxisomal targeting signal in the C-terminus	none	
2nd peroxisomal targeting signal	none	
Possible vacuolar targeting motif	none	
N-myristoylation pattern	none	
Transport motif from cell surface to Golgi	none	
Nuclear		Reliability: 0.941

singalIP 3.0 预测无信号肽; 细胞核定位信号 (nuclear location signals, NLS) 专项预测软件 NUCLEO 与 NUCPRED 的细胞核定位数值都为 0.65; Proloc 与 Subnuclear compartments prediction system 的预测也定位在细胞核 (nuclear lamina). PredictNLS 预测出位于编码框 1206~1239 核苷酸 (402~413 位氨基酸残基) 处有一个细胞核定位信号 (图 2). 而 Psort II -K-NN predict 分析出该蛋白质有 6 个可能的核定位信号及 1 个可能的二联体 (bipartite) 核定位信号 (图 3).

MPLYDNQKARSVMNECERHVFARTDADAPPPEDWEEHVNRTGTMAQNKLFNKILKAL
QSDRLARLANEGACNEPVLRRVAVDKCARRVRQALASVSWDTKLIQWLHTTLVETLSLPM
LAAYLDALQTLKGKIPITLIDRMLVSSNTKTGAAGAEALSLLKRPWDPAVGLSHNPKSK
LPGSPLILIASSGPSSSVFPTSRHRFWQSLGSLGKVIPVATHLLNNGSGVGLQCLEH
MIGAVRSKVLIEIHSHPHKPIILIGWNTGALVACHVSMEYVTAVVCLGFLLTVDGPRG
DVDDPLLDMPKPVLFVIGQNSLQCHPEAMEDFREKIRAENSLVVGGADDNLRISKAKKK
SEGLTQSMVDRICIQDEIVDFLTGVLTRAEGHMGSEPRDQAEKKKKPRDVARRDLAFEV
ERGRSPASPAKLPASPSGSEDLSVSSSTSPSKTKVTVTSAQKSSQIGSSQLLKRHV
QRTEAVLTHKQAQVPISSEPPEEGEKEDLRVQLKRHPSSPLPGSKTSKRPIKIVSLISQ
GDTAGGCAPSGQSAPEAAGGKPIITMTLQASAGAKELTGLTTAKSSSEGGVSASPVP
SVVSSSTASALHTLQSRVATSPGSSPLGATSSALLQGLSFLQDISSKTSGLPANPS
PGPAPQATSVKLPMPQSLGAITGTSTIVRTIPVATLSSLGATPGGKPTAIHQLLTNG
GLAKLASSLPGLAQISNQASGLKVPITITLTLRGQPSRITLSPMGGAAPSEESSQVL
PSSSQRLPPAPRCVIC

Fig. 2 Result of NLS analysis

Signal region were framed by grey.

-Results of the <i>k</i> -NN prediction		
<i>k</i> =9/23	95.7%: Nuclear	4.3%: Cytoplasmic
-discrimination of nuclear localization signals		
RRHR(3) at 203		
KKKK(5) at 403		
KKKP(4) at 404		
KKPR(4) at 405		
KRPK(4) at 529		
PTSRRHR(3) at 200		
Bipartite: KRHPSSPLPGSKTSKR at 514		
Content of basic residues: 10.6%		NLS Score: 1.53

Fig. 3 The prediction results of PsortII-K-NN

2.3 构建并鉴定 EGFP/Si1 融合表达载体及 EGFP/Si 基因半分子突变体表达载体

将成功克隆到 pMD18-T 载体的 Si1 基因利用 *Bam*H I, *Hind* III 酶切克隆入 pEGFP-N2 载体中 (图 1b). 利用相同方法构建 Si1 基因不同长度的 EGFP/Si 基因半分子突变体. 构建 Si1 基因 5' 端 GSI5' a, GSI5' b, GSI5' c 以及 Si1 基因 3' 端

G-TS4, G-TS5, G-TS1, G-TS3, G-TS2, 精确定位片段 G-TS6, G-TS7, G-TS8. 上述各片段克隆覆盖肽段位置见图 4, 全部片段均经测序鉴定正确.

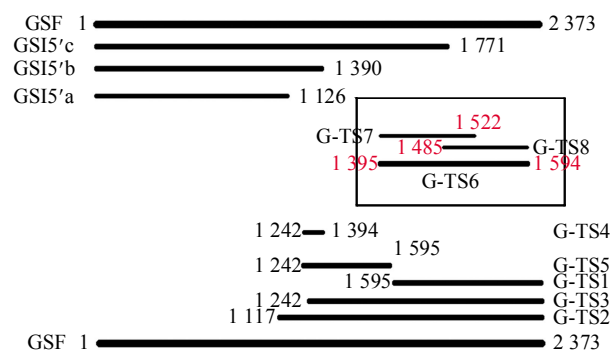


Fig. 4 Different fragments were designed for exploring Si1 NLS

1~2737 is ORF of Si1.

2.4 EGFP/Si1 融合蛋白以及 EGFP/Si 基因半分子突变体在 HeLa 细胞中的亚细胞定位

脂质体转染 EGFP、EGFP/Si1 到 HeLa 细胞中, 培养 24 h 后激光共聚焦显微镜下观察, EGFP 蛋白荧光信号分布在整个细胞中, 而转染了 EGFP/Si1 的细胞其荧光信号主要集中在细胞核内, 细胞核仁清晰可见 (图 5).

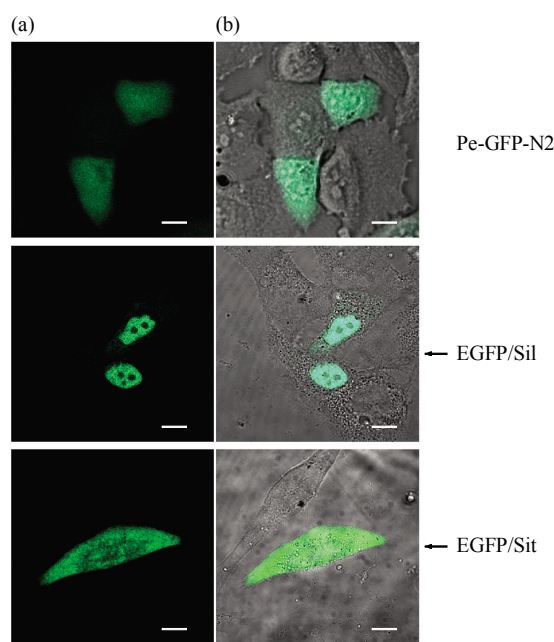


Fig. 5 The Sub-location of Si1 protein

(a) Green fluorescence, (b) Overlap (a) with visible light, the expression of GFP protein was detected by LSCM. The green signals separated in the whole cell that transfected with EGFP vector (control), located only in the nuclear of cell transfected with EGFP/Si1, separated in the whole cell transfected with EGFP/Sit (Si1 mutant). Space bars: 10 μ m.

对 Si1 基因 5'端 GSI5'a, GSI5'b, GSI5'c 以及 Si1 基因 3'端 G-TS4, G-TS5, G-TS1, G-TS3, G-TS2 重组质粒用相同方法转染和观察核定位情况(图 6, 7), 比较分析后确定 Si1 基因的核定位相关

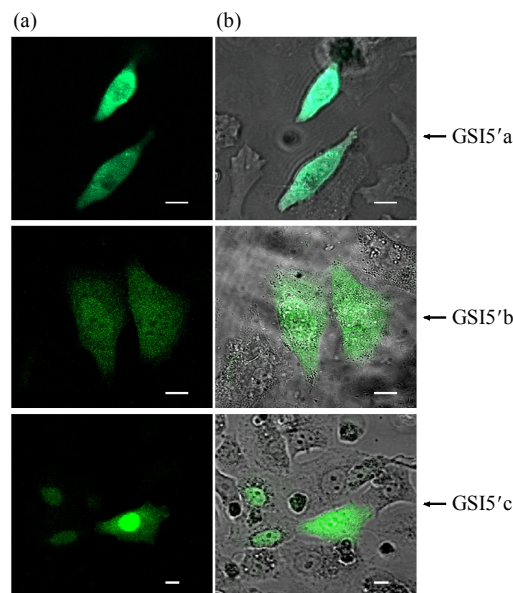


Fig. 6 The location of Si1 protein in 5' terminal

(a) Green fluorescence, (b) Overlap (a) with visible light, the expression of GFP protein was detected by LSCM. Space bars: 10 μ m.

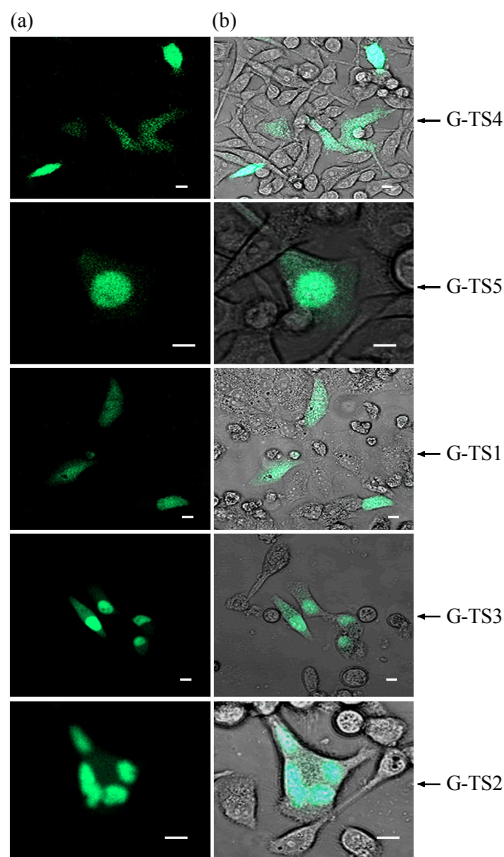


Fig. 7 The location of Si1 protein in 3' terminal

(a) Green fluorescence, (b) Overlap (a) with Visible light, the expression of GFP protein was detected by LSCM. Space bars: 10 μ m.

序列可能在其编码框 1 395~1 595 bp 处. 为进一步精确其核定位序列, 构建了 G-TS6, G-TS7, G-TS8 表达载体进行定位信号分析(图 8), 可见 G-TS7 与 G-TS8 表达蛋白并没有明显入核, G-TS6 是 Si1 表达蛋白入核的关键序列(1 395~1 594 bp).

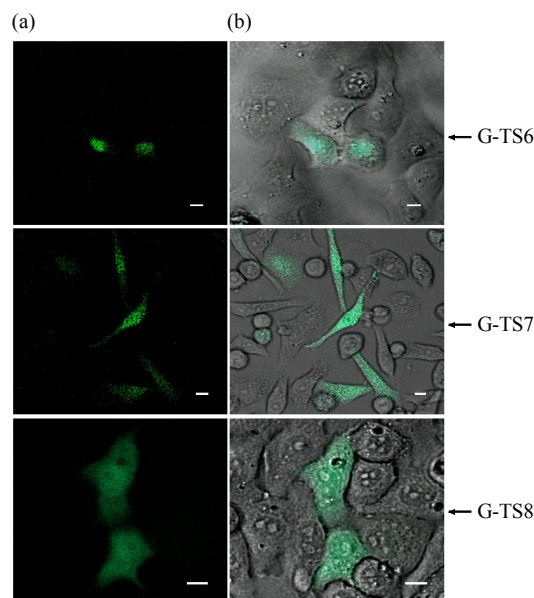


Fig. 8 The location of Si1 supposed NLS protein

(a) Green fluorescence, (b) Overlap (a) with Visible light, the expression of GFP protein was detected by LSCM. Space bars: 10 μ m.

2.5 构建并鉴定 EGFP/Sit 融合基因表达载体

突变位点用 PCR 扩增突变法分两段经过 PCR 连接获得. 凝胶电泳如图 9a. 通过酶切位点 *Sal* I 与 *Sac* I, 由酶切 - 重组的方法插入到 pEGFP-N2 载体中, 获得 GSM(GFP Si1 Mutation)重组子, 测序检测结果如图 9b.

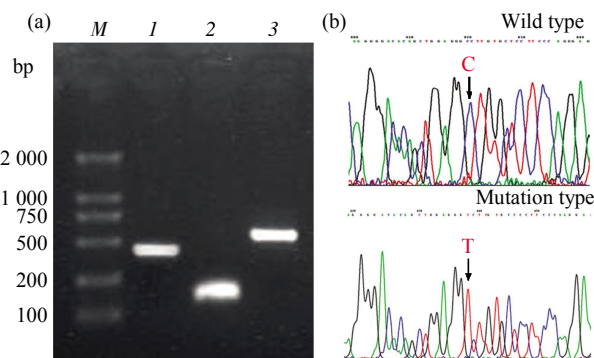


Fig. 9 Construction of Si1 mutation

(a) Gel electrophoresis of GSM- I, GSM- II, GSM. *M*: Molecular mass marker. (b) Sequencing map for Si1 mutant, the arrow point the base C of wild type (top) and the base T of mutated type (bottom). 1: GSM- I; 2: GSM- II; 3: GSM.

2.6 EGFP/Sit 融合蛋白无法定位于细胞核

脂质体转染 EGFP、EGFP/Si1、EGFP/Sit 到 HeLa 细胞中,培养 24 h 后激光共聚焦显微镜下观察(图 5),结果发现,转染了 EGFP/Sit 的细胞与转染了 EGFP 的细胞荧光信号分布在整个细胞中,这与转染了 EGFP/Si1 的细胞荧光主要分布在细胞核中明显不同.上面的研究表明在 Si1 基因编码框 1395~1594 bp(465~531 氨基酸残基)处有一入核功能信号区,而突变点的位置在 Si1 基因编码框 1639bp 的位置,与入核功能信号区不重叠.

2.7 数据统计

通过生物学统计软件 Image pro 6.0,统计转染后 EGFP 融合蛋白发出荧光在细胞中定位的比率,结果为 3 次重复实验选取 3 个视野统计结果的平均值,得到的结果证明,每次转染不同长度的 EGFP/Si1 融合蛋白在 HeLa 细胞中荧光分布基本一致,除去发荧光的死细胞,平均为 $(97.0 \pm 0.8)\%$.

3 讨 论

与血清抑制相关的基因多与细胞周期及细胞核内转录和调控有关^[4,5],我们一方面对在血清抑制条件下获得的 Si1 基因表达蛋白亚细胞定位进行分析,另一方面进行了与 EGFP 蛋白融合并在 HeLa 细胞中表达观察其亚细胞定位. Cello version 2 和 Psort II prediction 的预测结果都表明, Si1 基因表达蛋白定位在细胞核中,融合蛋白在 HeLa 细胞中,正如预测结果定位在细胞核中,这一结果显示 Si1 基因蛋白可能通过与核 DNA 结合,调节 DNA 的复制或基因的表达,进而影响细胞的增殖.其更细节的机制有待进一步研究.

由细胞核定位专项分析软件 NLS predict 分析得出 Si1 基因表达蛋白 402~413 位氨基酸残基(1 206~1 239 位核苷酸)有一核定位信号,而 Psort II 分析出 Si1 基因表达蛋白有 6 个可能的核定位信号及 1 个可能的二联体信号. NLS 一般是由一簇、两簇,或者两簇以上的碱性氨基酸组成^[15],其在序列组成上没有固定模式.通常 NLS 分为两类,即单向靶序列以及两个相互独立却联结作用的二联靶序列^[16].将蛋白质与 GFP 蛋白 5'端融合表达是应用广泛的分析蛋白质亚细胞定位的方法^[17].本研究用 EGFP 蛋白融合定位的实验结果表明,在 Si1 基因的细胞核定位序列位于 1 395~1 594 (465~531 位氨基酸残基)区段,与细胞核定位专项分析软件 NLS predict 预测区段不一致,但与 Psort II -K-

NN predict 的部分区段一致,尤其是二联体信号区是一致的.

G-TS7 和 G-TS8 两段有 13 个氨基酸序列的重叠,但分别表达时都不定位到细胞核中,表明 Si1 蛋白的细胞核定位信号并不只是几个氨基酸确定的,可能与一段序列有关.

前期的研究工作表明, Si1 一个多态位点的突变与肿瘤有关,并在大肠癌中有显著的突变频率^[2,3].这个突变位点位于 Si1 基因编码框 1 639 (546 位氨基酸残基),突变将基因原本的 Ser 变为 Pro,与预测的核定位信号都不重合.构建 Si1 基因突变型与 EGFP 融合蛋白在 HeLa 细胞中的定位结果表明,突变型表达蛋白也为弥散状态,这表明该位点的突变影响了 Si1 基因表达蛋白的细胞核迁移.但该位点不在核定位区(465~531 位氨基酸残基),这表明核定位区以外的个别氨基酸的变异,也会引起蛋白质空间结构的变化,进而影响其蛋白质功能的正常发挥^[18],这可能是导致 Si1 基因在肿瘤中突变频率高的原因之一.本研究的这一发现为下一步研究 Si1 基因表达蛋白的功能及其与肿瘤的相关机制提供了基础.

参 考 文 献

- 1 谭德勇,赖建华,钱伟,等.一个血清抑制基因的克隆.生物化学与生物物理进展,2002,29(5): 816~819
Tan D Y, Lai J H, Qian W, et al. Prog Biochem Biophys, 2002, 29 (5): 816~819
- 2 谭德勇,赖建华,余敏,等. Si-1 基因在肿瘤中的缺失和突变状况初步研究.云南大学学报(自然科学版),2003,25(3): 277~282
Tan D Y, Lai J H, Yu M, et al. J Yunnan University (Natural Sciences Edition), 2003, 25(3): 277~282
- 3 谭德勇,杨克,杨举伦,等. Si-1 基因中一个多态位点的初步分析.云南大学学报(自然科学版),2004,26(5): 450~453
Tan D Y, Yang K, Yang J L, et al. J Yunnan University (Natural Sciences Edition), 2004, 26(5): 450~453
- 4 Chen Y, Zhou G J, Yu M, et al. Cloning and functional analysis of human mTERFL encoding a novel mitochondrial transcription termination factor-like protein. Biochemical and Biophysical Research Communications, 2005, 337(4): 1112~1118
- 5 Cooper S. Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points. Faseb J, 2003, 17(3): 333~340
- 6 Horton P, Nakai K. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. Proc Int Conf Intell Syst Mol Biol, 1997, 5: 147~152
- 7 Shen H B, Chou K C. Gpos-PLoc: an ensemble classifier for predicting subcellular localization of Gram-positive bacterial proteins. Protein Engineering Design & Selection, 2007, 20(1): 39~46

- 8 Yu C S, Chen Y C, Lu C H, *et al.* Prediction of protein subcellular localization. *Proteins*, 2006, **64**(3): 643~651
- 9 Bendtsen J D, Nielsen H, von Heijne G, *et al.* Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 2004, **340**(4): 783~795
- 10 Cokol M, Nair R, Rost B. Finding nuclear localization signals. *EMBO Rep*, 2000, **1**(5): 411~415
- 11 Hawkins J, Davis L, Boden M. Predicting nuclear localization. *J Proteome Res*, 2007, **6**(4): 1402~1409
- 12 Huang W L, Tung C W, Huang H L, *et al.* ProLoc: Prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features. *Biosystems*, 2007, **90**(2): 573~581
- 13 Lei Z, Dai Y. An SVM-based system for predicting protein subnuclear localizations. *BMC Bioinformatics*, 2005, **6**: 291
- 14 Lei Z D, Dai Y. Assessing protein similarity with Gene Ontology and its use in subnuclear localization prediction. *Bmc Bioinformatics*, 2006, **7**: 491
- 15 Dingwall C, Laskey R A. Nuclear targeting sequences—a consensus?. *Trends Biochem Sci*, 1991, **16**(12): 478~481
- 16 Robbins J, Dilworth S M, Laskey R A, *et al.* Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell*, 1991, **64**(3): 615~623
- 17 Milon B, Dhermy D, Pountney D, *et al.* Differential subcellular localization of hZip1 in adherent and non-adherent cells. *FEBS Lett*, 2001, **507**(3): 241~246
- 18 Inoue T, Stuart J, Leno R, *et al.* Nuclear import and export signals in control of the p53-related protein p73. *J Biol Chem*, 2002, **277**(17): 15053~15060

Study of Subcelluar-localization of Serum Inhibited Gene Si1 Protein*

ZHAO Wen-Xiu^{1)**}, WU Hong^{1,2)**}, TAN De-Yong^{1)***}

⁽¹⁾The Laboratory of Biochemistry & Molecular Biology, School of Life Science, Yunnan University, Kunming 650091, China;

⁽²⁾College of Life Science and Technology, Southwest University for Nationalities, Chengdu 610041)

Abstract A serum inhibited gene Si1 (GenBank accession number: AY050169) was previously cloned and identified by differential expression of genes in U251 cells. For the further study of biological function of Si1, prediction procedure was performed to predict its subcelluar-localization. Relative experiments were carried out at the same time. The expression of EGFP/Si1 recombinant in HeLa cells showed Si1 protein located in nuclear which corroborated the prediction results of Psort II, Proloc, Cello version2, Subnuclear compartments prediction system, NUCLEO and NUCPRED. According to the PredictNLS prediction, twelve different fragments of EGFP/Si1 recombinants were constructed to identify precise NLS regulation sequence. Findings proved that the real NLS regulation sequence was not the same as the software predicted(1 206 bp~1 239 bp on Si1 ORF), but located on 1 395 bp~1 594 bp of Si1. A tumor related mutation/EGFP recombinant localization result showed though the mutation site (1 639 bp on Si1 ORF) does not located in NLS regulation sequence, it did affect wildtype Si1 protein divert to nuclear and may affect its natural function in cell, perhaps it is the main reason for highly mutation rate of Si1 in tumor.

Key words Si1 gene, subcelluar-localization, NLS, regulation sequence, mutate

DOI: 10.3724/SP.J.1206.2008.00758

*This work was supported by grants from The National Natural Science Foundation of China (30360040, 39960030, 30760057) and The Technological Programme of Yunnan Province (2008CC003).

**These authors contributed Equally to this work.

***Corresponding author.

Tel: 86-871-5032061, E-mail: dytan@ynu.edu.cn

Received: February 3, 2009 Accepted: April 22, 2009