

# 小干扰 RNA 的合理设计

张美红,周克元

关键词:RNAi;siRNA;shRNA;设计;基因沉默

中图分类号:Q522 文献标识码:A 文章编号:1000-8578(2006)11-0837-03

## 0 引言

大量实验表明,针对同一靶基因的不同 siRNA 序列具有不同的沉默效率,为了能够运用 RNAi 技术更有效地沉默靶基因,需要在 RNAi 的第一步,也是其成功与否的关键环节— siRNA 序列的设计方面进行更多的改进。本文将在以下几个方面探讨 siRNA 序列设计方面的最新研究进展,作为应用 RNAi 技术时的参考。

## 1 siRNA 序列的设计

RNAi 最终要通过 siRNA 片段与靶基因结合并使之降解,因此,确保高度同源于靶基因而绝无与其他基因同源的 siRNA 序列,是决定 RNAi 特异性的关键所在,也是 siRNA 设计的基本原则。从具有不同沉默效率的 siRNA 序列中筛选出高效的 siRNA 序列,需要经过严密的设计和不断的实验检验<sup>[1]</sup>。

### 1.1 siRNA 序列在靶基因中的位置

从靶基因起始密码子 AUG 下游 50~100 个核苷酸开始搜寻理想的 siRNA 序列,越靠近靶基因的 3' 端,其基因沉默效果可能越好<sup>[2]</sup>。以前的研究表明:不要以 5' 非翻译区(5' UTR)和 3' 非翻译区(3' UTR)及起始密码子附近序列作为设计 siRNA 的模板,这些区域含有调节蛋白结合位点(如翻译起始复合物),调节蛋白可与 RISC 竞争结合 siRNA 序列,降低 RNAi 效应<sup>[3]</sup>;也要避开外显子与外显子的交界区域和单核苷酸多形性(SNP)区域。最近, Yokota 等发现在 HCV(丙型肝炎病毒)基因组中, 5' UTR 是一个高度保守区,使之成为 siRNA 理想的靶点。运用 RNAi 技术作用于 5' UTR 或 3' UTR

序列,同样可引起靶基因沉默<sup>[4,5]</sup>,这些发现使基因功能分析领域扩展到 5' UTR 和 3' UTR 内。

### 1.2 siRNA 序列的起始碱基与长度

SiRNA 序列最好为 AA(N n)UU(N 代表任意碱基;n 为碱基数目,在 19~29 nt 之间), NA(N n)UU 和 NA(N n)NN 序列也可以。以前的研究表明,典型 siRNA 的三大结构特征是:长度为 21~23 nt;siRNA 双链的 3' 端各有两个突出碱基;siRNA 双链的 5' 端有磷酸基团。以 AA-N19 标准设计的 siRNA,在哺乳动物细胞中 70%~80% 可产生 RNAi 作用。但是一些具有较小脱靶效应的 siRNA 序列不是以 AA 为起始的,所以现在不推荐以“AA 为起始”作为选择标准之一<sup>[6]</sup>。最新研究表明<sup>[7,8]</sup>, 27 nt 或 29 nt 的 siRNA 与 21 nt siRNA 相比:(1)其抑制活性可提高数倍以上;(2)不易于诱导干扰素反应和激活 PKR;(3)一些基因对 21 nt siRNA 不敏感,但是可以被 27 nt siRNA 有效的抑制;(4)与 21 nt SiRNA 相比,27 nt SiRNA 对靶基因的最大抑制率可在相对低的浓度下得到。另外,在选择 siRNA 序列时,要考虑到所用启动子的类型。U6 启动子要求转录产物的第一个碱基是 G,正反义链均如此;H1 启动子转录产物的第一位碱基为 U、G、C 均不影响基因的沉默效果<sup>[9]</sup>。

### 1.3 siRNA 3' 端突出碱基的选择

当 siRNA 的 3' 端突出碱基为 UU 时,其基因抑制效率最高;但是 3' 端的突出碱基不能为 G,因为 RNase 会降解以 G 为末尾的 RNA 单链。Elbashir 等<sup>[2]</sup>建议用 dTdT 取代 3' 端的 2 个碱基突出,增强 siRNA 双链复合体的稳定性。需要注意的是,由于双链 siRNA 中的正义链不参与对靶 mRNA 的识别,所以正义链的 3' 端突出碱基可以用 dTdT 替代,但是反义链 3' 端突出碱基必需与靶 mRNA 序列相同的。

### 1.4 siRNA 序列中 G/C 含量的选择

在 siRNA 序列中,G/C 含量的选择比确定以 AA 为起始的序列更重要。具有均衡碱基含量(即 G/C 含量在 30%~70%) 的序列可以作为 siRNA

收稿日期:2006-03-06;修回日期:2006-05-18

基金项目:广东省自然科学基金资助项目(31962);广东省重点学科资助项目(GX9307)

作者单位:524023 广东湛江,广东医学院生物化学与分子生物学研究所

通讯作者:周克元

作者简介:张美红(1980-),女,硕士在读,主要从事基因治疗研究

候选序列;而具有较低 G/C 含量(30%~52%)的 siRNA 序列,其沉默基因效果较好<sup>[10]</sup>。

1.5 siRNA 中应避免连续的单一碱基和反向重复序列<sup>[10,11]</sup>

因为连续 2 个以上的 G 和 C 可以降低双链 RNA 的内在稳定性,从而抑制 RNAi 作用<sup>[12]</sup>;连续 3 个以上的 U 和 A 可能终止由 RNA Polymerase III 介导的转录<sup>[13]</sup>。siRNA 序列中的重复序列或回文结构可能形成发夹状结构,这种结构的存在可以降低 siRNA 的有效浓度和沉默效率。

1.6 siRNA 双链的内在稳定性

siRNA 反义链 5' 端具有较低的热稳定性,即反义链 5' 端的第一个碱基为 A 或 U; siRNA 正义链的 5' 端第一个碱基为 G 或 C。因为 siRNA 解旋可能从富含 A/U 的反义链 5' 端有效的起始,有利于反义链与 RISC 的结合,而抑制正义链与 RISC 的结合<sup>[14]</sup>。

1.7 siRNA 正义链的碱基偏爱性

以 3' 端具有 2 个碱基突出的 19 nt siRNA 为例,正义链的第一位和第十九位碱基为 A;正义链的第十位碱基为 U;正义链的第十三位碱基不为 G;正义链的第十九位碱基不为 G 或 C 时,siRNA 序列具有较高的基因沉默效率<sup>[10]</sup>。

1.8 siRNA 序列的特异性

为了确保候选 siRNA 序列只沉默单一的靶基因,候选 siRNA 序列应在美国 NCBI (National Center for Biotechnology Information) 的 EST 或 Unigene 数据库进行 BLAST 同源性比对。当 siRNA 序列只同源于靶基因时,才可以用于下一步的基因功能研究。

1.9 可能影响 siRNA 效能的其他因素

RNA 结合蛋白的结合情况;mRNA 的二、三级结构;mRNA 的丰度、半衰期;以及所表达蛋白质的半衰期等等<sup>[15]</sup>也可能影响 siRNA 的效能。有研究表明,与反义技术或核酶技术不同,在 RNAi 过程中,靶 mRNA 的二级结构对基因沉默的效果没有太大影响<sup>[16]</sup>。为了使 siRNA 具有最大的靶向特异性,在条件允许的情况下应该尽可能的全面考虑。

1.10 设计的 siRNA 序列数目

目前没有发现 siRNA 的干扰效率与 mRNA 中具体位置的关系,为确保对靶 mRNA 的有效抑制,针对每一个靶基因应该设计 4 条以上 siRNA 序列,进行化学合成后,通过实验筛选出沉默效率最高的 siRNA 序列进行下一步的基因功能研究。

需要注意的是,有的 siRNA 序列在设计时没有遵守这些指导方针,实验证明也具有很高的干扰效

应,这表明,我们只能根据上述指导方针设计出理论上具有较高沉默效应的 siRNA 序列,siRNA 的最终活性需要用实验来验证<sup>[17]</sup>。

## 2 shRNA 序列的设计

短发夹 RNA (small hairpin RNA, shRNA) 可在细胞中被加工成 siRNA。设计 shRNA 时,shRNA 中两个反向重复序列的设计原则与 siRNA 序列的设计相同,需要重点考虑的是 shRNA 环序列的长度及组成,而且环序列不能与靶基因内的其它序列具有同源性。

2.1 shRNA 环序列的长度

一般来讲,设计 shRNA 时选择 3~10 nt 长度的发夹环均可。Brummelkamp 等<sup>[18]</sup>构建了具有 5 nt、7 nt、9 nt 环的 shRNA,比较它们在 MCF-7 细胞中抑制 CHD1 基因的效果:9 nt 环的 shRNA 抑制率达 90%;7 nt 环的 shRNA 只有中等的抑制率;5 nt 环的 shRNA 则无抑制活性。但是 Siolas 等<sup>[8]</sup>报道,环序列的长度对 RNAi 效率无明显影响。

2.2 shRNA 环序列的组成

研究表明,当 shRNA 发夹环的组成为“UU-GAUAUCCG”和“UUCAAGAGA”时具有较高的抑制效率。当发夹环前面的序列是以“UU”终止时,应该选用“CCACACC”环序列<sup>[12]</sup>。环序列中有 2 个 U 对产生有效的基因抑制很重要<sup>[18]</sup>;但环序列中不能有连续 3 个以上的 U<sup>[12]</sup>,因为这可能导致 shRNA 转录的提前终止。

## 3 siRNA 阴性对照序列的设计

所有的 RNAi 试验均应设立阴性对照,siRNA 阴性对照序列的合理设计与 siRNA 序列的设计同样重要。因为有效的对照可以充分证明 siRNA 只对靶基因产生特异性基因沉默,从而增强实验的可信度。阴性对照 siRNA 包括碱基错配或混乱序列的 siRNA。在实验中最好设计两条 siRNA 对照序列。

3.1 碱基错配的阴性对照 siRNA

最初的研究证明,在 siRNA 双链内,一个碱基的突变就足以阻断 RNAi 过程。反义链内具有 1~2 个碱基错配的阴性对照 siRNA 可用于鉴别 RNAi 通路和 micro RNA 通路,因为后者是由碱基的非精确配对引起。最近的研究表明:siRNA 允许其分子中心部位存在单个碱基突变,siRNA 完全失活需要大于 3 个碱基的突变<sup>[19]</sup>,在 Mangeot<sup>[20]</sup>的研究中,对照 siRNA 序列中含有 6 个精确定位的突变碱基,也证明了这点。Kim 等<sup>[7]</sup>的研究表明,在 27 nt 的

siRNA 中,有 3 个碱基错配的对照 siRNA 序列可以在不同的浓度下阻断针对靶基因的 RNAi 过程,而单个碱基错配的 siRNA 仅在相对高的浓度下才可阻断 RNAi 过程,所以多碱基错配比单碱基错配的 siRNA 阴性对照序列具有更高的实际应用价值。此外,对于 RNAi 效应而言,位于 siRNA 序列中部的碱基错配比位于 siRNA 序列尾部的碱基错配更有效。

### 3.2 混乱序列的阴性对照 siRNA

来源于 siRNA 序列的混乱序列 siRNA,在 siRNA 实验中是一个阴性对照,作为一个信息控制可反映 RNAi 的特异性。混乱序列 siRNA 与 siRNA 序列相比具有相同的核苷酸组成,但是核苷酸顺序被打乱,同样要保证它和靶 mRNA 没有同源性。

## 4 展望

运用 RNAi 技术不但可以迅速、准确的剖析基因功能,分析基因之间错综复杂的联系和相互作用,还为人们预防和治疗癌症和病毒疾病提供了新的思路。具有最大靶向特异性的 siRNA 在 RNAi 过程中的运用,不仅可以提高对靶基因的抑制效率,而且可以避免非特异性反应。随着 RNAi 技术在应用中所遇困难的不断解决、新的更高效的 RNAi 表达体系的不断涌现, RNAi 的应用范围将不断扩大,必将为生命科学研究和临床治疗带来新的革命。

### 参考文献:

- [1] McManus M T, Petersen C P, Haines B B, et al. Gene silencing using micro-RNA designed hairpins[J]. *RNA*, 2002, 8 (6): 842-850.
- [2] Elbashir S M, Harborth J, Weber K, et al. Analysis of gene function in somatic mammalian cells using small interfering RNAs[J]. *Methods*, 2002, 26(2): 199-213.
- [3] Elbashir S M, Harborth J, Lendeckel W, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells[J]. *Nature*, 2001, 411(6836): 494-498.
- [4] Doench J G, Petersen C P, Sharp P A. siRNAs can function as miRNAs[J]. *Genes Dev*, 2003, 17(4): 438-442.
- [5] Eckmann C R, Kraemer B, Wickens M, et al. GLD-3, a bi-caudal-C homolog that inhibits FBF to control germline sex determination in *C. elegans*[J]. *Dev Cell*, 2002, 3(5): 697-710.
- [6] Naito Y, Yamada T, Ui-Tei K, et al. siDirect; highly effective, target-specific siRNA design software for mammalian RNA interference[J]. *Nucleic Acids Res*, 2004, 32 (Web Server issue): W124-129.
- [7] Kim D H, Behlke M A, Rose S D, et al. Synthetic dsRNA Dicer substrate enhance RNAi potency and efficiency[J]. *Nat Biotechnol*, 2005, 23(2): 222-226.
- [8] Siolas D, Lerner C, Burchard J, et al. Synthetic shRNAs as potent RNAi triggers[J]. *Nat Biotechnol*, 2005, 23(2): 227-231.
- [9] Scherr M, Morgan MA, Eder M. Gene silencing mediated by small interfering RNAs in mammalian cells[J]. *Curr Med Chem*, 2003, 10(3):245-256.
- [10] Reynolds A, Leake D, Boese Q, et al. Rational siRNA design for RNA interference[J]. *Nat Biotechnol*, 2004, 22(3): 326-330.
- [11] Shirane D, Sugao K, Namiki S, et al. Enzymatic production of RNAi libraries from cDNAs[J]. *Nat Genetics*, 2004, 36 (2): 190-196.
- [12] Wang L, Mu F Y. A Web-based design center for vector-based siRNA and siRNA cassette[J]. *Bioinformatics*, 2004, 20 (11): 1818-1820.
- [13] Kasahara H, Aoki H. Gene silencing using adenoviral RNAi vector in vascular smooth muscle cells and cardiomyocytes[J]. *Methods Mol Med*, 2005, 112: 155-172.
- [14] Khvorova A, Reynolds A, Jayasena S D. Functional siRNAs and miRNAs exhibit strand bias[J]. *Cell*, 2003, 115(2): 209-216.
- [15] Kretschmer-Kazemi F R, Schzakiel G. The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides[J]. *Nucleic Acids Res*, 2003, 31(15): 4417-4424.
- [16] Ui-Tei K, Naito Y, Takahashi F, et al. Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference[J]. *Nucleic Acids Res*, 2004, 32(3): 936-948.
- [17] Bitko V, Musiyenko A, Shulyayeva O, et al. Inhibition of respiratory viruses by nasally administered siRNA[J]. *Nat Med*, 2005, 11(1): 50-55.
- [18] Brummelkamp T R, Bernards R, Agami R. A system for stable expression of short interfering RNAs in mammalian cells[J]. *Science*, 2002, 96(5567): 550-553.
- [19] Jacque J M, Triques K, Stevenson M. Modulation of HIV - 1 replication by RNA interference [J]. *Nature*, 2002, 418 (6896): 435-438.
- [20] Mangeot P E, Cosset F L, Colas P, et al. A universal transgene silencing method based on RNA interference[J]. *Nucleic Acids Res*, 2004, 32(12): e102.

[编辑:刘红武]