

# 생물정보를 이용하여 바실러스 서브틸리스에서 새로운 Small RNA를 예측하는 방법

이 상 수  
배재대학교 생명공학과

## Searching Method for New Small RNA in *Bacillus subtilis* Using Bioinformation

Sang Soo Lee  
Department of Life Science and Technology, Pai Chai University

### 요 약

바실러스 서브틸리스 유전체에서 여러 환경의 적응에 이용할 것으로 보이는 새로운 small RNA를 찾는 시도로 다음과 같은 방식으로 유전체를 검색하였다. 첫째로 유전자들 사이에 존재하는 DNA 서열을 대상으로 전사인자들인 PerR, OhrR, Fur, Zur의 인식서열을 조사하였고, 둘째로 인자 비의존성 전사종결 부위를 조사하였다. 이들 조사에서 전사인자의 위치와 전사종결부위가 300 bp 내외에서 가까이 존재하는 후보 DNA 서열을 대상으로 전사인자 부위에서 전사촉진자의 서열이 발견되는지를 조사하였다. 이 결과 PerR 5개, Fur 2개, OhrR, Zur에서 각각 1개의 새로운 small RNA로 추정되는 부위가 발견되었다.

### Abstract

In order to find novel sRNA in *Bacillus subtilis* which would be used to adapt to several

conditions, we searched the whole genome of *Bacillus subtilis* using the following procedure. At first, the locations of recognition sequence of transcription factors such as PerR, OhrR, Fur and Zur were searched in the intergenic region of *Bacillus subtilis* genome and the locations of rho independent transcription terminator sites were also determined. Based on the information of these locations, the sRNA candidates were chosen by close locations (less than 300 bp) between the recognition site of transcription factors and rho independent transcription terminator site. Then transcription promoter sites were searched in the region of previously identified sRNA candidates and 5 PerR, 1 OhrR, 1 Fur and 1 Zur regulated good sRNA candidates were found.

## I. 서 론

최근에 작은 non-coding RNA가 많은 주목을 받고 생명계 전반에 걸쳐 조절자로서 인식되고 있다. 진핵세포의 siRNA와 miRNA와는 달리 원핵세포의 small RNA는 그 크기와 구조면에서 다양성을 보이고 있다. small RNA(sRNA)는 일반적으로 번역되지 않으며 50-250 뉴클레오타이드 정도이다. sRNA는 1970년대 이후 알려졌으나, 최근의 체계적인 유전체 전반에 관한 탐색을 통하여 새로운 sRNA를 암호화하는 많은 유전자들이 발견되었다(1,2). 이러한 sRNA를 암호화하는 유전자들은 실험적으로나 계산에 의해 발견하기 매우 어렵다. 더욱이 이들의 유전자들은 그 크기가 매우 작아 생화학 및 유전학적 방법으로 찾아내기가 어렵다. 실제로 2000년까지 지난 30년 동안에 발견된 sRNA의 개수는 고작 10개로 우연히 대장균에서 발견된 것뿐이다(3).

지난 몇 해 동안 대장균에서 새로운 sRNA를 발견하려는 체계적인 연구가 진행되었고 이들의 연구를 통해 sRNA라 생각되는 유전자들을 예측하였다. 기계적 학습을 통한 시도에 의해 562개의 새로운 nontranslated RNA를 예측하였고, 총 유전체 배열 방식을 통해 317개의 기능을 모르는 새로운 전사물을 예측하였다(4). Argaman은 유전자들 사이의 서열에서 전사 신호를 찾는 방식으로 24개의 후보자를 찾아냈다(5). 실험적 검증에서 14개가 RNA로 발현되고 이중 두개가 *gevB* 및 *rprA*라는 sRNA로 확인되었다(6,7). 또한 Wassarman등에 의해 시도된 보존된 서열과 DNA array data에서 60개의 sRNA를 예측하였다(8). 이것 들을 Northern blot으로 확인한 결과 이중 17개가 sRNA로 밝혀졌다. 이외에도 Chen에 의해 개발된 알고리즘은 전사 신호에 의존하는 것으로 227개의 후보자를 예측하

고 이중 7개가 발현됨을 확인하였다(9). 이와 같은 시도로부터 55개의 알려진 sRNA를 암호화하는 유전자를 알게 되었다.

반면에 대장균이외에서 sRNA를 발견하려는 시도에서 여러 개의 대장균에서 발견된 것과 유사한 sRNA를 발견하였다. 이들은 대부분 그람 음성균이 주종을 이룬다. 본 연구에서는 그람 양성균의 대표 균주인 *Bacillus subtilis*에서 sRNA를 예측하려는 시도로 이 유전체의 서열에서 유전자 사이의 서열을 대상으로 전사인자 인식 DNA 서열, 전사종결서열, 전사신호서열을 모두 갖춘 후보자를 찾아내고자 하였다. 특히 관심을 갖고 탐색한 전사인자는 PerR, OhrR, Fur, Zur등의 산소활성종 및 철과 아연의 금속이온의 농도에 따른 전사 조절인자를 대상으로 이와 관련된 sRNA를 찾고자 하였다.

## II. 방 법

### 전사인자 인식서열의 탐색

*Bacillus subtilis*에서 전사조절인자인 PerR, OhrR, Fur 및 Zur가 인식하는 서열을 *Bacillus subtilis* 유전체의 유전자 사이의 DNA 서열을 대상으로 pattern search하였다(표 1). 이를 위하여 Subtilist World-Wide Web Server의 home page인 <http://genolist.pasteur.fr/SubtiList/>을 이용하였다. 이 pattern search에서는 mismatch의 개수를 4개 까지 허용하여 조사하였다. 이 결과 인식 DNA 서열이 14개인 PerR과 Fur는 총인식개수가 9686과 8018로 제일 많았고 인식서열이 15개인 OhrR은 1785의 총 인식 개수를 보였다. 반면에 인식서열의 개수가 9개인 Zur는 총인식개수가 불과 20개에 지나지 않았다.

표 1. *Bacillus subtilis* 전사조절인자인 PerR, OhrR, Fur 및 Zur가 인식하는 서열의 개수

전사인자	인식염기서열	총인식 개수	인식 서열에서 mismatch의 개수				
			0	1	2	3	4
PerR	ttataatnattataa	9686	8	22	154	1256	8246
OhrR	tacaattaaattgta	1785	1	2	14	181	1587
Fur	tgataatnattatca	8018	8	34	94	1004	6878
Zur	aaatcgtaanattacgatt	20	0	4	6	4	6

### 전사종결자의 위치 및 방향에 대한 정보

sRNA는 그 크기가 매우 작지만 그 기능을 수행하기 위해서는 전사가 일어나야한다. 따라서 전사가 종결되는 전사종결자가 반듯이 sRNA의 유전자의 '3 말단에 존재해야한다. 이 같은 정보를 바탕으로 *Bacillus subtilis*의 유전체에서 전사종결위치를 찾는 인터넷 사이트를 이용하였다. 이 사이트는 <http://transterm.cbcb.umd.edu/>로 박테리아 유전체에서 rho-independent terminator를 예측하여 자료를 제공한다. 이 프로그램에서 제공받은 자료는 전사종결자의 위치 방향 전사종결정도를 예측할 수 있어 매우 유용하게 이용할 수 있었다. 이 전사 종결자의 위치가 유전자 서열의 reading frame '3 말단에 위치하지 않을 경우나 그 전사종결 방향이 반대로 향할 경우, 또한 유전자 사이에 두 개 이상의 전사종결자가 위치할 경우 등 확실히 전사종결자의 기능이 정해지지 않은 것을 찾았다. 이 결과 총 243개의 전사종결자를 발견하였다(data not shown).

### 전사인자 인식서열과 전사종결자 서열의 비교를 통한 sRNA의 위치 예측

전사인자 인식서열과 전사종결자의 위치가 300-400 뉴클레오타이드 이상 떨어져 있지 않고 가까이 있을 경우 이부분의 DNA 서열이 전사인자의 조절을 받는 전사가 일어날 것으로 생각된다. 일반적으로 전사인자 인식서열은 mismatch가 3개 정도까지 허용하여 이 근처에 전사종결자가 위치하는지를 탐색하였다. 이와 같이 탐색된 sRNA 후보 유전자 서열을 대상으로 전사촉진자의 공통염기서열인 ttgaca(-35 부분)과 tataat(-10 부분)을 찾아보았고 이들 사이의 뉴클레오타이드 간격은 16-18정도로 하여 조사하였다.

## III. 결과 및 고찰

전사인자 인식부위와 전사종결자의 위치가 근접한 후보 sRNA는 아래 표2와 같다(표 2.)

표 2. 후보 sRNA의 전사인자 인식부위 및 전사종결자 DNA 위치

전사인자	전사인자 인식부위 및 전사종결자 DNA위치 (전사인자/전사종결자)
PerR	<b>(559273/559084)</b> , (1347679/1347511), (1690287/1690208), <b>(2091698/2091520)</b> , (2268992/2269150), (2280677/2280558), (2677790/2677661), (2724257/2724259), <b>(2981515/2981620)</b> , (3359860/3359958), (3386999/3386778), <b>(3572116/3572082)</b> , <b>(3822486/3822173)</b> , (4034954/4034733), (4117575/4117524),
OhrR	(999350/999534), <b>(2319395/2319301)</b>
Fur	(502307/502411), (662637/662835), (972267/972167), (1216445/1216600), <b>(1482841/1482921)</b> , <b>(2091698/2091520)</b> , (2315624/2315423), (2328932/232246), (3359843/3359958)
Zur	<b>(2477131/2476930)</b>

표 2에서와 같이 PerR이 15개, OhrR이 2개 Fur가 9개, Zur는 1개의 sRNA 후보자를 확인하였다. 이중에서 PerR의 전사인자 인식위치 2091698은 Fur의 전사인자 인식위치와 공통으로 나타나는 경우가 있는데 이는 PerR의 인식부위와 Fur의 인식부위가 유사하여 일어난 것으로 생각된다. 한 후보자 이들 후보자를 대상으로 전사촉진자의 서열을 확인한 결과 뚜렷한 전사촉진자의 서열이 확인되는 후보자는 PerR 4개, OhrR 1개, Fur 2개, Zur 1개를 아래에 서열을 표시하였다.

표 3. 전사촉진자, 전사인자 인식부위 및 전사종결자의 위치가 확인된 sRNA 유전자 서열

굵은 글씨는 sRNA 유전자를 표시하며, 이탤릭체는 전사촉진자 서열, 밑금 친 부분은 전사인자 인식 DNA 서열을 나타낸다.

-----  
 PerR(559273), Terminator(559084, -100) transcript length: 194 nt  
 gtttatcttcttcataaaagagatatgagagttg**ataacatt**tagatgaaccatttt**ataa**tt**catc**ta**atc**tacttaaatgaacgaacaggtagtgaaaaaa  
**agagtcagttatctatattgaacaaacattaataattatatacatccgcattgtaacagtagtatactgtaaacaaaccttgagtttaaaccttttc**  
**tattacctgtagcatctccctactcgaacgttagtagggtttttattg**  
 -----

Per R / Fur(2091698), Terminator(2091520, +/- 100) transcript length: 204 nt  
 gtggcacaact**ttgtc**actgaaatacaata**ataatcaat**atcagaaattctgctgatgcaatggccgaccatttaacagggatatcatcaatccagtc**tttt**  
**attgggctg**tt**aac**aa**caatc**gttcagcgggctaggagcaggatggatca**ttta**atctggttcgacgaggtt**ctca**acggga**attc**aggt**taa**  
**caaaaaggatctggcagatccaagatccttttt**atttaaacagccttttctttga  
 -----

Per R(2981515), Terminator(2981620, +/- 100) transcript length: 82 nt  
acccttgcctttgaaagaacatacaaaattttataagcttattaagttaagcct**tatttcattata**aacaataatcagaca  
ttgttaaatgtttacc**tataat**ttttttccattc**gtgattcccg**aaaaaacgctttatt**gcgggtttg**cataaa  
aaaaagcggcttaaggccgctttttta

-----  
PerR(3572116), Terminator(3572082, -81) transcript length: 72 nt  
ggtttaattccgctg**taacacat**gtgtaatgctttcgg**ttatgattaataa**agtaattgaccccttt**tataatgaatacact**ttttggcacgattccagttac  
**cggtccttttt**gtccataataaaagagaattccagttaca

-----  
PerR(3822486), Terminator(3822173, -100) transcript length: 305 nt  
atggaa**ataaac**caaggat**tttactcat**tttaaacggcggaacagccagattatgattcagcctgactaaccttatatcttctgacat**ttcccg**ttttat  
**at**ttctctcatcagaatgggtacaa**tttactctatccaccg**tatt**gcagtcacagc**ccgctcttttcggcgt**g**ccggag**gtc**gagcgggtcagctg  
**ccgctcat**gtt**gtcactc**ggcg**ttgcca**atagctt**gtcttattctgc**agaaacgatcggatgaatccg**tgacag**tagtttctc**taaacctctc**ctctcttggc  
**ggagagg**tttttacttacaaggtt

-----  
OhrR(2319395), Terminator(2319253, -100) transcript length: 150 nt  
aagagattagaatc**ttgatata**atttat**tacaatataatagga**aacactcatataatc**gctggatag**gcacgcaagttctaccgggaccgtaaatg**tcg**cact  
**atgggtgagcaatgga**accgacg**gtacgg**ttttgtgatatcagatt**gctctcttattg**agcgggcaatgcttttttatt

-----  
Fur(1482841) Terminator(1482921,+100 )  
atatcaaaaaacgatt**gacattgatactgagaatcattatcatt**aattatagagagaagctactct**gttcccaacccctctatc**agatcaagatccgaaaaa  
**actggcgctacccccgcaag**ttttttttgttcaagcataggacaat

-----  
Zur(2477131), Terminator(2476930, +/- 85) transcript length: 224 nt  
tgaagcggagccttatt**tgaca**attt**gtgatgacgtgtata**taaaacgta**atcattacg**ttttataaggagatg**taattatgagag**taaatg**tactt**agcttt  
**gcactga**aacgggagaccgaa**ttacatcaca**caaaaaataaacgcaccaatccagaccg**ttggagctg**aaaaatattcaccg**ctt**aaagaaatata  
**ccctcaccg**tgaacaaaaataataaaagaccgggaatcc**cttacggc**ccg**ctcttttt**acgtaattgactgcctgtaaaaga

-----  
이와 같은 sRNA 후보자 서열을 이용하여 target 유전자를 탐색하였다. 탐색 프로그램은 인터넷에 올라와 있는 사이트로 주소는 <http://snowwhite.wellesely.edu/targetRNA>이다. 이 탐색의 결과 PerR의 3572116 위치에 있는 sRNA는 target으로 *katE*를 찾아냈으며, 같은 PerR의 3822173 위치에 있는 sRNA는 *sodF* 유전자를 target하고 있었다. 이는 아마도 PerR에 의해 주로 조절되어 발현되는 *katA*와 *sodA*의 작용으로 *katE*와 *sodF*의 유전자 발현을 억제하는 것으로 생각된다. OhrR(2319395)의 target으로는 *cysK*, *katX*가 탐색되는데 이는 OhrR

의 작용에 cysteine이 관여하는 것과 밀접한 관련이 있을 것으로 생각되며 *katX* 또한 PerR의 마찬가지로 주 조절단백질의 발현으로 필요성이 없어진 *katX*의 억제로 조절이 일어나는 것으로 볼 수 있다. Zur(2477131)의 경우 target으로 *rpsL*(s12)로 리보솜을 구성하는 유전자이다. 이는 아연의 경우 리보솜 단백질이 그 조절에 관여한다는 연구 등이 발표되어 이와 매우 관련이 있는 것으로 생각된다. 이렇듯 위의 방법으로 탐색된 sRNA 유전자들이 매우 그럴듯한 예측으로 생각되고 후속의 실험 등에 의해 이의 증명이 필요한 실정이다.

## V. 참고문헌

1. Storz, G (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260-1263.
2. Wassarman, K.M., Zhang, A. and Storz, G. (1999) Small RNAs in *Escherichia coli*. *Trends Microbiol.*, **7**, 37-45.
3. Hershberg, R., Altuvia, S. and Margalit, H. (2003) A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 1813-1820.
4. Carter, R. J., Dubchak, I. and Holbrook, S. R. (2001) A computational approach to identify genes for functional RNAs in genomic sequences. *Nucleic Acids Res.*, **29**, 3928-3938.
5. Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H. and Altuvia, S. (2001) Novel small RNA-encoding genes in intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**, 941-950.
6. Urbanowski, M., Stauffer, L. T. and Stauffer, G. V. (2000) The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport system in *Escherichia coli*. *Mol. Microbiol.* **37**, 856-868.
7. Majdalani, N., Chen, S., Murrow, J., St John, K and Gottesman, S. (2001) Regulation of RpoS by a novel small RNA: the characterization of *RprA*. *Mol. Microbiol.* **39**, 1382-1394.
8. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. and Gottesman, S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637-1651.
9. Chens, S., Lesnik, E. A., Hall, T. A., Sampath, R., Griffey, R. H., Ecker, D. J. and Blyn, L. B. (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *Biosystems*, **65**, 157-177.