

특 집 : 포스트-게놈 시대의 단백질 발현 시스템(II)

초고속 단백질생산기술

Technical Developments in High-throughput Protein Expression

성 백 린

연세대학교 공과대학 생명공학과

현재까지 단백질의 발현 및 생산은 단백질의 구조 기능연구에 주된 bottleneck으로 남아있다. 따라서 human genome으로 유래되는 새로운 drug target의 validation, 치료용 단백질의 창출과 개발 및 신약 screening 방법을 개발하기 위해서는 많은 단백질을 초고속으로 생산하는 기술의 개발이 시급하다 (그림 1).

대략 35,000개 이상으로 알려져 있는 human genome의 기능연구가 향후 post-genome 연구의 주요한 challenge로 남아 있다. 현재 유전체 서열연구에 의한 BLAST search로 추정결과 약 60% 정도의 유전자에 대해서는 기능이 유추되어 있다. 그러나 나머지 40%에 대해서는 기능을 전혀 모르고 있는 실정이다[1].

이제는 유전자 서열에 의존한 기능의 annotation을 넘어서 구조를 바탕으로한 기능연구가 더 효율적이다. 그 이유로서는 단백질의 구조는 단순아미노산 서열보다 더 많이 보존되어 있고 따라서 구조적 유사성이 기능의 유추에 더 좋은 parameter 이기 때문이다.

초고속단백질생산기술(High-throughput protein expression; HT protein expression)은 microarray용 단백질 생산 및 면역학적 연구에 의한 기능연구까지 확대될 수 있다. 즉 어떠한 유전자의 기능도 이의 최종 산물인 단백질의 기능연구 및 확인으로 완성되기 때문이다.

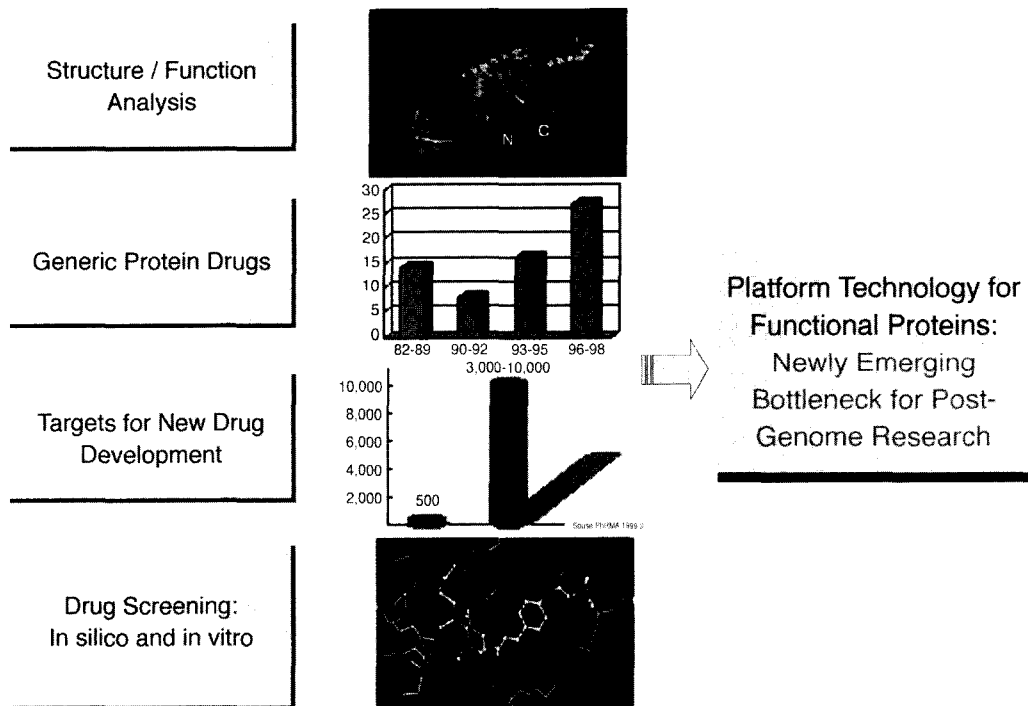


그림 1. 초고속 단백질 생산기술과 post-genome 연구영역

High-throughput cloning

7. 높은 restriction enzyme/DNA ligase에 의한 'cut and paste' 방법은 automation이 어렵다. 즉 각각의 expression vector 및 유전자의 특징적인 cleavage site를 이용하고 정제하는 것은 automation에 큰 어려움을 주고 있다.

이를 극복하기 위하여 Invitrogen에서 Topo cloning 방법이 개발되었다[2]. Vaccinia virus DNA topoisomerase-1를 이용하여 PCR product를 직접 vector에 연결하는 방법이다. 이에 대한 PCR 후에 blunting 또는 cutback 등의 처리가 필요없이 약 5분 간의 incubation으로 cloning을 할 수 있고 이는 automation이 가능하다[3].

Expression systems:

가장 단순한 system으로서 *Escherichia coli* host가 가장 적합한 HT-expression system으로 개발이 가능하며 몇가지 방법들이 보고되어 있다[4,5,6] 그러나 단점으로는 복잡한 단백질 특히 glycosylation등이 필요한 단백질의 expression에는 적절하지 않다. 아울러 대부분 활성이 없는 insoluble한 침전으로 발현되는 것이 단점으로 지적되고 있다. 이러한 단점을 보완하기 위하여 yeast, mammalian, insect cell system들이 HT-format으로 개발되고 있으나 아직 많은 기술적 문제들을 안고 있다.

이를 극복하기 위한 일반적인 방법으로 Cell-free expression systems이 개발되고 있다.

이들의 장점으로는 i) 단백질 발현시 toxic 효과 때문에 발현이 어려운 단백질의 생산이 가능하고, ii) cloning이 필요없이 직접 PCR transcript를 이용하여 발현이 가능하며[9], iii) Isotope 또는 seleno-methionine등으로 label하는 기술을 이용함으로써 단백질의 구조 분석에 용이하게 사용될 수 있다 [7,8]. 이 기술은 일반적으로 단백질의 발현yield가 낮은 것이 단점으로 지적되어왔으나 최근 이를 극복한 기술이 소개되었고 96 well format으로의 expression 및 solubility 검색이 가능하게 되었다[10].

Expression vectors and Recombination cloning

현재 단백질 발현기술의 문제점으로는 i) 각각의 유전자를 수 많은 expression vector에 sub-cloning하는 것은 labor-intensive할 뿐만 아니라 많은 시간이 소요되는 것과, ii) PCR에 의한 cloning은 과정중에 많은 DNA의 mutation을 야기하기 때문에 cloning 후 DNA sequencing을 확인해야 하는 번거로움이 존재한다.

이를 극복하기 위한 연구로서는 첫째 Gateway cloning (Invitrogen)을 들 수 있다[11]. 즉 lambda recombinase를 이용하여 specific site(att site)에 integration 및 excision을 할 수 있으며 원하는 유전자를 specific 한 entry vector의 2개의 att site 사이에 cloning한다. 이후 complementary att site를 지나는 destination vector와 *in vitro*에서 mix하면 expression clone이 형성된다. 이 방법은 현재 Harvard의 Institute of Proteomics에서 FLEXGene repository에서 사용중이며 production-ready cDNA clone의 생산에 이용되고 있다[12].

또 다른 방법으로는 Cre-loxP system(Creator system; Clontech)을 들 수 있는데 이 방법은 Mammalian Gene Collection Project에 사용중이며 NIH와 NCI의 공동협력으로 full-length clone 확보방안으로 진행중에 있다[13]. 그러나 recombination site의 transfer 중에 다른 유전자와의 recombination 문제가 남아있으며 따라서 이후 구조/기능 연구에 혼동이 발생할 수도 있다. 이러한 문제를 해결하는 방안으로는 multiple system vector를 들 수 있다. 즉 pTriEx-1 vector나 pBEV vector를 이용하여 *E. coli* 및 insect cell에 one-step cloning이 가능하다[14]. 이 경우 recombinant baculovirus는 insect 세포의 transfection과 mammalian 세포의 transduction 모두에 사용이 가능하다. 뿐만 아니라 promotor의 사용에 따라서 여러 종의 insect, mammalian 세포에서의 발현 및 viral vector를 이용한 stable expression도 가능하다[15].

Protein Expression

전통적인 shaker flask는 비효율적이며 이 대신 deep-well block의 사용이 고안되었다. 이는 기본적으로 DNA 정제를 위하여 고안되었던 장치로서 여러 format(96, 48, 및 24 well)과 volume(1, 2, 5ml)으로 사용이 가능하다. 이 경우 AirPore tape(Qiagen)을 이용하여 공기의 공급과 아울러 sterility를 유지하여 flask culture 조건을 유지할 수 있다. 사용하는 shaker로는 microtiter format, vortex mixing, gas flow 및 온도조절 기능을 갖는 HiGro shaker (GeneMachines사)를 사용할 수 있는데 실제로 flask culture와 비교 시 비슷한 성장 curve를 보이는 것으로 보고되고 있다. 최근에 Micro-bioprocessor (Flowmetrix사)를 이용하여 non-invasive sensor 기술에 의해 pH, O.D., 및 용존산소를 측정하는 기술이 개발되었고 이를 이용하여 다양한 배양조건에서 HT-expression 조건을 검색할 수 있다[16].

Protein Purification

주로 tagging 방법을 이용하는데 일반적으로 6개의

표 1. 초고속 단백질 발현기술

기술영역	개발기술	개발회사	참고문헌
High-throughput cloning	- Topo cloning	Invitrogen	2
Recombination cloning	- Gateway cloning - Cre-loxP system(Creator system)	Invitrogen Clontech	11, 13
High-throughput expression	- Fusion to solubilizing partners; GST, TRX, MBP, NusA - Cisperone	Invitrogen, Promega, Sigma-Aldrich, Clontech Protheon	17, 18
Cell Growth	- AirPore tape - HiGro shaker - Micro-bioprocessor	Qiagen GeneMachines Flowmetrix	
Protein Purification	- Ni-NTA magnetic agarose bead 96-well format - SwellGel : Ni-chelating disc	BioRobot Qiagen Pierce	19, 20
Screening for Expression	GFP fusion; <i>in vivo</i> fluorescence as indicator of solubility	Stratagene	23

Histidine으로 이루어진 tag을 이용한다. C-terminal에 tagging 하는 경우 발현된 단백질을 정제 중 truncated form을 배제하는데 도움이 되며 N-terminal tagging의 경우 optimal expression을 위해 codon usage를 최적화가 가능하다. 이뿐 아니라 MBP[17] 또는 *Schistosoma japonicum* GST[18]을 soluble expression 및 정제에 이용되고 있다. Liquid handling robot을 이용하여 single-step으로 정제가 가능하다. 예를 들어 His-tag에 대한 high affinity Ni-NTA magnetic agarose bead를 이용하는 96-well format (BioRobot; Qiagen)이 이용된다 [19]. SwellGel (Pierce) Ni-chelating disc의 경우 더 binding capacity가 크며 one-step으로 단백질의 정제가 가능하다. SwellGel glutathione disc 이용하여 GST fusion protein을 정제할 수 있다. 이 경우 96-well format 으로 구입하여 liquid handling robot으로 automation 방법으로 통상 1ug ~ 6mg 정도의 정제된 단백질을 얻을수 있다[20]. 좀 더 큰 scale로 96X65ml의 format으로도 정제가 가능하며 이 경우 원심분리, sonication, 정제 모두를 robotics 방법으로 하여 10mg까지 정제가 가능함이 보고되었다[6].

Screening for Expression

일단 원하는 단백질을 수용성으로 발현 가능여부에 따라 향후 구조결정 및 활성연구가 가능한지 결정된다. 수용성 발현이 어려운 단백질의 경우 어떤 특정부위의 아미노산을 mutation[21] 하거나 deletion[22] 함으로써 수용성을 크게 증가시킬 수 있다. Reporter 유전자를 이용한 solubility assay가 개발되어 있으며 GFP fusion 단백질을 이용하여 *in vivo* fluorescence를 solubility의 척도로 사용할 수 있다[23]. 같은 목적으로 CAT, beta-galactosidase도 사용이 가능하다. 그러나 fusion 자체에 의해 단백질의 solubility가 영향을 받을 수 있음

생물산업

이 이 기술의 단점으로 지적되고 있다.

결론

역사적으로 볼때 각각의 단백질의 특성에 입각하여 그에 맞는 expression system이 적용되어왔다. 이에는 우연적 발견과 실험당사자의 개개적 skill에 의존도가 매우 높아 기술 자체가 매우 점진적으로 개선되어 왔음이 사실이다. 그러나 이러한 방법은 HT-expression에는 적용될 수 없으며 따라서 유전자 분석과 같은 scale로 강력한 일반적인 방법이 개발되어 단순화되고 안정한 automation 기술로의 개발이 매우 시급하다.

Human genome 연구의 진행에는 instrumentation - parallel processing, miniaturization, 그리고 automation 기술 등 -이 크게 공헌하였다. 물론 단백질의 경우 다양성에 기인하는 많은 어려움이 존재하나 비슷한 개념들이 단백질의 발현, 생산의 상용기술의 개발에 적용될 수 있다.

그러나 아직까지의 HT-expression 연구들은 주로 pilot study에 불과하며 사용한 단백질의 수나 종류에서 볼때 아주 제한적인 결과에 불과하다. 따라서 full spectrum의 human 유전자의 발현연구에 단지 guide 역할만 할 뿐이며 아직 다양한 단백질의 구조결정, 기능연구의 need를 맞추지 못하고 있다. Ultra HTS 방법과 micro-crystallization 방법을 동원한다 하여도 아직 신약개발에 필요한 단백질의 공급은 challenge로 남아있다. 향후 HT-expression이 단백질발현 연구실의 단순 service 차원이 아니라 일종의 연구 infra로 정착되어야만 될 것이다.

참고 문헌

1. J.C. Venter et al., The sequence of the human genome. *Science* **291**(2001), 1304-1351.

2. J.A. Heyman et al., Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* **9**(1999), 383-392.
3. M. Nasoff et al., High-throughput expression of fusion proteins. *Methods Enzymol.* **328**(2000), 515-529.
4. K. Bussow et al., A human cDNA library for high-throughput protein expression screening. *Genomics* **65** (2000), 1-8.
5. M. Larsson et al., High-throughput protein expression of cDNA products as a tool in functional genomics. *J. Biotechnol.* **80**(2000), 143-157.
6. S.A. Lesley, High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expres. Purif.* **22**(2001), 159-164.
7. D-M. Kim et al., A highly efficient cell-free protein synthesis system from *E. coli*. *Eur. J. Biochem.* **239** (1996), 881-996.
8. T. Kigawa et al., Selenomethionine incorporation in to a protein by cell-free synthesis. *J. Struct. Funct. Genomics* **2** (2002), 29-35.
9. K.A. Martemyanov et al., Direct expression of PCR products in a cell-free transcription/translation system: synthesis of antibacterial peptide cecropin. *FEBS Lett.* **414** (1997), 268-270.
- 10 S. Yokayama et al., Structural genomics in Japan. *Nat. Struct. Biol.* **7** Suppl. (2000), 943-945.
- 11 J.L. Hartley et al., DNA cloning using in-vitro site-specific recombination. *Genome Res.* **10**(2000), 1788-1795.
- 12 L. Brizuela et al., FLEXGene repository: from sequenced genes to gene repositories for high-throughput functional biology and proteomics. *Mol. Biochem. Parasitol.* **118** (2001), 155-165.
- 13 Q. Liu et al., The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**(1989), 1300-1309.
14. R. Novy et al., pTriEx multisystem vector for protein expression in *E. coli*, mammalian and insect cells. *InNovations* **10**(1999), 1-5.
15. R.V. Merrihew et al., Chromosomal integration of transduced recombinant baculovirus in mammalian cells. *J. Virol.* **75**(2001), 903-909.
16. Y. Kostov et al., Low-cost microbioreactor for high-throughput bioprocessing. *Biotechnol. Bioeng.* **72**(2000), 346-352.
17. C.V. Maina et al., An *E. coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **74**(1988), 365-373.
18. D.B. Smith and K.S. Johnson, Single-step purification of peptidase expressed in *E. coli* as fusions with glutathione S-transferase. *Gene* **67**(1988), 31-40.
19. T. Lanio et al., Automated purification of His6-tagged proteins allows exhaustive screening of libraries generated by random mutagenesis. *BioTechniques* **29**(2000), 338-342.
20. U. Romer et al., Significantly higher yields from automated protein purification procedures. *QIAGEN News* **5**(2001), 12.
21. R. Wetzel et al., Mutations in human interferon affecting inclusion body formation identified by a general immunochemical screen. *BioTechnology* **9**(1991), 731-737.
22. J.A. Barwell et al., Overexpression, purification, and crystallization of the DNA binding and dimerization domains of the Epstein-Barr virus nuclear antigen 1. *J. Biol. Chem.* **270**(1995), 20556-20559.
23. G.S. Waldo et al., Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* **17**(1999), 691-695.