Making high fidelity expression libraries

ArticleInfo		
ArticleID	:	3621
ArticleDOI	:	10.1186/gb-2000-1-2-reports0044
ArticleCitationID	:	reports0044
ArticleSequenceNumber	:	18
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2000–4–1 Received : 2000–4–1 OnlineDate : 2000–5–11
ArticleCopyright	:	BioMed Central Ltd2000
ArticleGrants	:	

ArticleContext	:	130591122

Rachel Brem

Abstract

Error-free protein plus nucleic acid libraries have been generated by fusing mRNA to expressed protein and selecting for full-length proteins.

Significance and context

Many new biochemistry experiments involve the expression of DNA or RNA libraries of thousands of genes at once. A technical problem arises when some sequences of the library contain errors, such as frameshifts, deletions, or misplaced start or stop codons. Here, Cho *et al*. have developed a new technique to eliminate such mistakes. For each sequence in an RNA library, they translate the encoded protein and then covalently attach the protein to its RNA. From a pool of these protein-RNA fusions, they select only full-length proteins, guaranteeing, in theory, an error-free library. This new technique is important not only because it can improve the quality of nucleic acid libraries, but also because the final RNA-protein fusions can be used in further selection experiments for protein structure or function. The technique may turn out to be a viable cell-free alternative to phage display.

Key results

Cho *et al*. use a new procedure to generate full-length proteins that are coupled to their mRNAs. When the authors make a library of random 20 amino-acid peptides using their new procedure, 88% of them are perfect, compared with 34% for a control library that has no specific selection of full-length proteins. The authors then use their basic idea to make several different test libraries. For each library, they want proteins that are long (80-300 amino acids), but their error-free fusions are only 20 amino acids long. So they must ligate the short error-free RNA cassettes into longer sequences and then translate those sequences; the ligation introduces some error, however. The final libraries of long sequences are between 60 and 100% error-free.

Methodological innovations

The technique of Cho *et al.* is as follows. First, the authors synthesize cassettes of RNA coding for sequences of interest flanked by the code for protein purification tags at the amino and carboxyl termini. At the end of each RNA sequence is an adduct of the antibiotic puromycin. Next, the authors add ribosomes and free amino acids. Ribosomes translate the RNA into protein, then stall at the puromycin at the end of the RNA and the protein becomes covalently attached to its message. The covalent protein-RNA fusions are then purified against affinity columns corresponding to the new protein's amino- and carboxy-terminal tags. Proteins in fusions that are selected should have amino and carboxyl termini and, therefore, should be full length.

Reporter's comments

The new technology of Cho *et al.* is exciting and innovative, but it seems as though a few problems still need to be worked out. It would be interesting to check whether the final expressed proteins are functional, or at least structurally intact, on the RNA fusions. And now that Cho *et al.* have developed the main library strategy, they will probably need to go back to improve the fidelity of their ligation step and/or the chemical synthesis of their initial sequences.

Table of links

Journal of Molecular Biology

References

1. Cho G, Keefe AD, Liu R, Wilson DS, Szostak JW: Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. J Mol Biol. 2000, 297: 309-319. 0022-2836

This PDF file was created after publication.