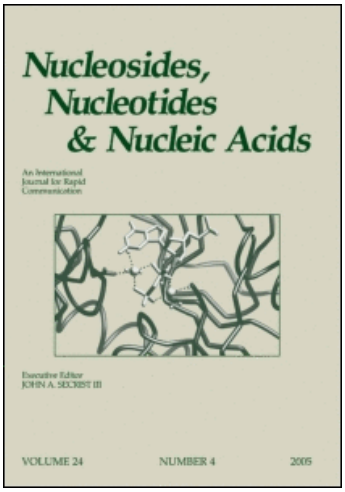

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

The Construction, Cloning, and Expression of Synthetic Genes Coding for Totally Artificial Proteins

H. Itoster[a]; J. Biernat[b]; H. Hasselmanr[b]; B. Hofer[b]; N. Kennedy[b]
[a] MilliGen, Division of Millipore, Bedford, MA, USA [b] Biosyntech, Stresemannstrasse, Federal Republic of Germany

To cite this Article Itoster, H. , Biernat, J. , Hasselmanr, H. , Hofer, B. and Kennedy, N.(1988) 'The Construction, Cloning, and Expression of Synthetic Genes Coding for Totally Artificial Proteins', Nucleosides, Nucleotides and Nucleic Acids, 7: 5, 683 — 686
To link to this Article: DOI: 10.1080/07328318808056309
URL: http://dx.doi.org/10.1080/07328318808056309

NUCLEOSIDES & NUCLEOTIDES, 7(5&6), 683-686 (1988)

THE CONSTRUCTION, CLONING, AND EXPRESSION OF SYNTHETIC GENES CODING FOR TOTALLY ARTIFICIAL PROTEINS

H. Köster*, J. Biernat+, H. Hasselmann+, B. Hofer+ and N. Kennedy+
*MilliGen, Division of Millipore, 75 Wiggins Ave,
Bedford, MA 01730 USA
+Biosyntech, Stresemannstrasse 268-280, 2000 Hamburg 50,
Federal Republic of Germany

Abstract

A series of genes was generated from three interchangeable cassettes, each coding for a specific set of amino acids. The genes were inserted into two different fusion expression vectors and two direct expression vectors. The expression studies demonstrated that proteolytic stability of the proteins is affected by the N-terminal region of the protein.

INTRODUCTION

Automated DNA synthesis has become an integral part of genetic and protein engineering research. Utilizing the genetic code, it also opens the way for the design of completely new proteins.

Several uses can be envisaged for these proteins: (1) as sources of amino acids for individuals with genetic defects in amino acid metabolism; (2) to study the folding rules used to predict secondary protein structure; (3) for production of polymers with specific physico-chemical properties; (4) for production of proteins with particular enzymatic or binding properties.

For the design and expression of totally new proteins, which have not been optimized by the evolutionary process, it has to be considered that today it is difficult if not impossible to predict the tertiary structure of a protein and to define what makes it stable against proteases. As a first step in the exploration of the design of novel proteins, we wished to construct proteins without any specific binding or enzymatic properties. Thus focussing only on the mechanisms of folding up of proteolytically stable secondary/tertiary structures.

RESULTS

Synthetic genes were designed which coded for proteins containing nutritionally significant compositions of amino acids. To facilitate production of a range of proteins, a strategy was selected which utilized segments coding for defined groups of amino acids. Three interchangeable DNA "cassettes", A, B, and H, coding for essential, non-essential, and branched chain amino acids, respectively were constructed. By combining these segments it was possible to produce, amongst others, proteins of Type A, AB, AF^{-}B, and AHB. Type A, consisting solely of essential amino acids, could be a valuable food additive. Type AB, containing all amino acids in optimal proportions, is a complete protein supply. Type AF^{-}B, of the same composition as AB but lacking phenylalanine, has potential as a protein supply for patients suffering from the genetic disease, phenylketonuria. Type AHB, with a high concentration of branched chain amino acids, could serve as a protein for patients with certain forms of liver insufficiency. To construct these three segments, synthetic oligonucleotides were generated and the double stranded sections assembled with the enzymatic "filling in" method. Oligonucleotides were prepared from suitably protected deoxynucleoside ß-cyanoethyl phosphoramidites.

The three cassettes were ligated in a stepwise fashion into pKK233-2 vector. A double digestion allowed insertion of cassette A into the NcoI/HindIII cut pKK233-2. The newly created pKK-A vector and cassette B were then double digested with the next group of restriction enzymes, allowing Segment B to be inserted in the proper orientation (pKK-AB). Segment H was added in a similar fashion (pKK-AHB). The newly constructed plasmids were transformed into E.coli HB101. The cloned gene sequences were confirmed by Maxam-Gilbert sequence analysis.

The various cloned genes were excised from their respective pKK plasmids and inserted into three different expression vectors (pUR300, pGFY221N, and pBI052) for analysis of expression in vitro and in vivo (E.coli minicells). The ß-galactosidase fusion vector pUR292 was modified, to become pUR300, by the addition of a new unique NcoI site thereby allowing all the described synthetic genes to be inserted in phase with ß-galactosidase. Another fusion vector, pGFY221N, was

created from pGFY218L by replacing the PstI site with an NcoI site. The recombinant proteins made from this vector contain an N-terminal leader sequence consisting of six amino acids. The phenylalanine codon was removed from the AB vector, to create AF⁻B, by cassette mutagenesis. The direct expression vector pBIO52 and all its derivatives were created from pGFY221N by removal of the NcoI/EcoI fragment (221bp) containing the six amino acid leader sequence and the upstream located bla promoter. In its place was ligated a chemically synthesized 52 base pair segment which contained an efficient E.coli tranlational initiation region. Genes from this vector were expressed directly.

In the expression experiments using pKK constructs despite efficient transcription, no translation products could be detected in either minicells or the in vitro system. Computer analysis revealed a thermodynamically stable hairpin loop structure of the mRNA that could potentially have masked the ribosome binding site thereby inhibiting translation. Both in vitro and in vivo experiments showed that the artificial fusion proteins could be efficiently expressed as ß-galactosidase fusion proteins in pUR300. In vitro expression of fusion proteins from pGYF221N indicated efficient translation of all gene products except 1-AHB and complete full length expression products. Apparently the six amino acid leader sequence (1-) prevents translation in the AHB gene. In vivo expression indicated variable product stability suggesting proteolytic activity against these proteins since they were stable in vitro. AHB and AF⁻B are unstable or toxic. Direct in vitro expression of proteins from pBIO52 indicated efficient and complete full length expression products. In vivo expression indicated variable product stability suggesting proteolytic activity against these proteins since they were stable in vitro.

DISCUSSION

The introduction of six amino acids at the N-terminus apparently changes the conformation of expressed proteins in a significant way. Depending upon the sequence, this leader sequence can stabilize or destabilize the structure. The presence or absence of phenylalanine also can dramatically affect the stability. In comparing AB and AF⁻B in the direct expression vector, it is noted that AF⁻B is very stable

whereas AB is completely degraded. The presence of phenylalanine at position 3 has apparently a significant effect on the stability of the protein as a whole. These results are an interesting exception to the 'N-end rule' which stresses that the N-terminal amino acid confers proteolytic stability to the protein. However, our results demonstrate that the N-terminal part of the protein plays a signficant role. It may be that it forms a core structure which predetermines during protein biosynthsis the later conformational/structural changes and tertiary structure.

REFERENCES

1. J.Biernat, H. Hasselmann, B. Hofer, N. Kennedy and H. Köster, Protein Engineering, 1, 345 (1987).
2. J. Biernat and H. Köster, Protein Engineering, 1, 353 (1987).
3. B. Hofer, Eur. J. Biochem., 167, 307 (1987).