Expression systems for production of heterologous proteins

Meena Rai* and Harish Padh[#,†]

*Biochemistry Department, M.S. University of Baroda, Vadodara 390 002, India
[#]B.V. Patel Pharmaceutical Education and Research Development Centre, Thaltej–Gandhinagar Highway, Thaltej, Ahmedabad 380 054, India

With the advent of our ability to clone and express a foreign gene in the heterologous host, came a remarkable capability to make almost any protein in abundant quantity to be used as therapeutic or diagnostic agents. It quickly led to the realization that proteins made in different hosts are different in many ways, particularly in their post-translation modifications. In this review a variety of available expression host systems are evaluated for heterologous production of proteins. Factors affecting the stability and expression of heterologous genes are also discussed. Eventual objective of producing a desired protein in an economical heterologous host is influenced by a variety of factors discussed in this review. Subsequent to the production, stabilization and formulation of proteins will pose significant hurdles in utilizing the natural biological catalysts and other proteins for therapeutic and industrial purposes.

ADVANCES in genetic engineering have made possible the production of therapeutics and vaccines for human and animals in the form of recombinant proteins[1,2]. These biotechnology-derived recombinant proteins form a new class of drugs for many ailments like genetic disorders, cancer, hypertension and AIDS for which we have no better treatment or cure. Unlike chemical drugs, biologicals are our own molecules and hence more compatible with biological systems. At present there are more than 100 biotechnology-derived therapeutics and vaccines approved by US FDA for medical use and over 1000 additional drugs and vaccines are in various phases of clinical trials. In addition, use of DNA, proteins and enzymes in diagnostics is increasing exponentially. Industrial uses of enzymes in food, textile, leather, detergent, medicinal chemistry sectors are also increasing rapidly. The growing need of therapeutic and other applications of enzymes and proteins could only be met by heterologous synthesis of recombinant proteins[1,2]. Table 1 indicates key therapeutic recombinant biotech products with their recent market sales and names of key manufacturers. Table 2 shows how rapidly many new therapeutic products have recently entered the market and the packed pipeline of product development and clinical trials. Many of these products will be approved in the near future, adding to the growing biotech product line. Indian share in consumption of biotech products is very insignificant, totalling only about 1.2 billion USD[1], most of which is by trading of imported goods (Table 3). Recently new companies like Shanta Biotech and Bharat Biotech have come up in Hyderabad, producing hepatitis B vaccine and with few other products in the pipeline. Shanta Biotech is using yeast as an expression system. Similarly few pharma companies like Torrent, Cadila Pharma and Zydus Cadila have also initiated work in production of recombinant therapeutic proteins.

In this review we outline steps involved in the production of heterologous proteins and then evaluate in detail available expression systems and factors affecting heterologous protein expression.

Heterologous production of proteins

Protein over-expression refers to the directed synthesis of large amounts of desired proteins. The heterologous production of proteins and enzymes involves two major steps:

(1) Introduction of foreign DNA into the host cells. This step has three major considerations. (a) Identification and isolation of the DNA to be introduced; (b) Identification of the vector and construction of recombinant vector; (c) Identification of the suitable expression system to receive rDNA.

(2) Factors affecting the expression of foreign DNA for protein synthesis in the chosen expression system.

Points (1a) and (1b) are topics in themselves and will not be dealt with in this review. Briefly, at present a variety of vectors are available to ferry DNA in and out of cells: plasmids, lambda phage, cosmids, phagmids, artificial chromosomes from bacteria, yeast or human origin (BAC, YAC and HAC[3], respectively). The vectors could either be integrating (becomes part of the host's chromosomes) or extrachromosomal. They could be in copies varying from one to several hundreds.

In general, expression vectors have the following attributes (Figure 1):

[†]For correspondence. (e-mail: perd@wilnetonline.net)

Table 1. Key therapeutic products

Product	1998 sales (billion USD)	Manufacturer
H-Insulin	3.0	Eli Lily, Novo Nordisk Hoechst, Yamanouchi
G-Factors	2.2	Eli Lily, Novo Nordisk Genentech, Pharmcia
Blood factors	5.5	Amgen, J & J, Sankyo Chugai, Sandoz
mABs	0.5	J & J, Cytogen
Interferons	3.8	Schering-Plough, Roche Daiichi, Wellcome

Table 2. Therapeutic pipelines

Year	Approved	Pipeline
1994	17	–
1995	33	450
1996	79	700
1998	> 100	1200

Table 3. Biotechnology sales in India*

	Biotech sales (Rs in crores)	
Category	1995	2000
Healthcare	1959	3532
Agriculture	154	385
Industrial products	570	1500
Others	30	130
Total	2793	5547
USD in million	635	1260

*Courtesy of Dr P. K.Ghosh, DBT.

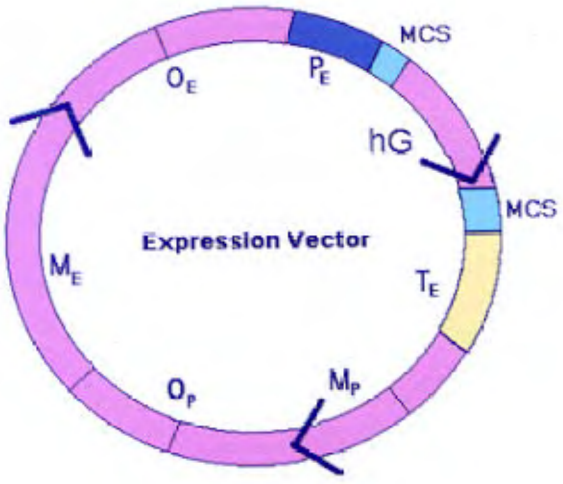

Figure 1. Design of typical shuttle vector for heterologous gene expression. MCS = multiple cloning sequence; O_p = bacterial origin of replication; M_p = marker gene for selection in bacteria; M_E = marker gene for selection in eukaryotes; O_E = eukaryotic origin of replication; P_E = eukaryotic promoter sequence; T_E = eukaryotic terminator sequence; hG = heterologous gene for expression.

- Ori: Sequences that allow their autonomous replication within the cell.
- Promoter: A tightly regulated promoter, i.e. one which can be switched on and off easily, is desirable.

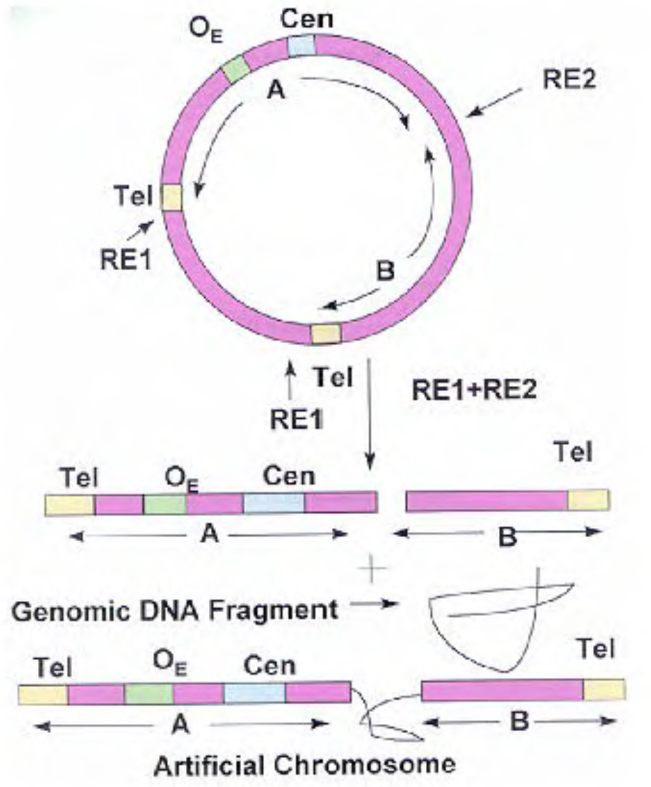

Figure 2. Design of an artificial chromosome. RE1 and RE2, Restriction enzymes cleavage sequences; O_E, Eukaryotic origin of replication; Cen, Respective centromeric sequence; Tel, Respective telomeric sequence; A and B, Fragments generated by cleavage of the vector by RE1 and RE2. Cleavage of the vector by a mixture of RE1 and RE2 generates fragments A and B and another fragment which is discarded. A and B fragments have Tel sequences at one end. Larger fragments of targeted genomic DNA to be cloned can be bracketed by ligation with fragments A and B as shown. Since the construct has centromeric and telomeric sequences, it can replicate as a chromosome in the appropriate eukaryote.

- Selection marker(s): Sequences encoding a selectable marker that assures maintenance of the vector in the host.
- Terminator: A strong transcriptional terminator should be used with a strong promoter to ensure that the RNA polymerase disengages and does not continue to transcribe downstream genes.
- Polylinker: To simplify the insertion of the heterologous gene in the correct orientation within the vector.

In addition, artificial chromosomes have centromeric and telomeric sequences which are host-specific[3] (see Figure 2). These attributes permit artificial chromosome

vectors to be faithfully replicated and distributed to daughter cells during cell division. Use of artificial chromosomes in commercial production of heterologous proteins remains unexplored.

Available expression system

Prokaryotic and eukaryotic systems are the two general categories of expression systems. Prokaryotic systems are generally easier to handle and are satisfactory for most purposes. However, there are serious limitations in using prokaryotic cells for the production of eukaryotic proteins. For example, many of the eukaryotic proteins undergo a variety of post-translational modifications like proper folding, glycosylation, phosphorylation, formation of disulphide bridges, etc. There is no universal expression system for heterologous proteins. All expression systems have some advantages as well as some disadvantages that should be considered in selecting which one to use. Choosing the best one requires evaluating the options – from yield to glycosylation, to proper folding, to economics of scale-up.

Bacterial system

E. coli is by far the most widely employed host, provided post-translational modifications of the product are not essential. Its popularity is due to the vast body of knowledge about its genetics, physiology and complete genomic sequence which greatly facilitates gene cloning and cultivation[4,5]. High growth rates combined with the ability to express high levels of heterologous proteins, i.e. strains producing up to 30% of their total protein as the expressed gene product, result in high volumetric productivity. Furthermore, *E. coli* can grow rapidly to high densities in simple and inexpensive media. Strains used for recombinant production have been genetically manipulated so that they are generally regarded as safe for large-scale fermentation. Purification has been greatly simplified by producing recombinant fusion proteins which can be affinity-purified, e.g. glutathione-S-transferase and maltose-binding fusion proteins[6]. However, expression in bacteria does have some serious disadvantages. It poses significant problems in post-translational modifications of proteins. Common bacterial expression systems such as *E. coli* have no capacity to glycosylate proteins in either N- or O-linked conformation. Although other bacterial strains such as *Neisseria meningirulls* have recently been shown to O-glycosylate some of their endogenous proteins, the trisaccharide added is different from O-linked sugars found in eukaryotes. Protein expressed in large amounts often precipitates into insoluble aggregates called inclusion bodies, from which it can only be recovered in an active form by solubilization in denaturing agents followed by careful renaturation[7]. Lysis to recover the cytoplasmic proteins often results in the release of endotoxins, which must be removed from the final product. Currently, strategies to secrete the target proteins by translocation into the periplasmic space or to release the target proteins by linking to existing excretory systems are being developed[8]. Additionally, the efficiency of expression will also depend on differences of codon utilization by bacteria[9]. At times the original sequence of the heterologous gene has to be modified to reflect the codon usage by the chosen expression system. *E. coli* has toxic cell wall pyrogens and hence products need to be tested more extensively before use.

Yeast

Yeast is the favoured alternative host for expression of foreign proteins for research, industrial or medical use[10]. As a food organism, it is highly acceptable for the production of pharmaceutical proteins. In contrast, *E. coli* has toxic cell wall pyrogens and mammalian cells may contain oncogenic or viral DNA. Compared to mammalian cells, yeast can be grown relatively rapidly (doubling time 90 min) on simple media and to high cell density and its genetics is more advanced than any other eukaryote, so that it can be manipulated as readily. Added advantages are the availability of complete genomic sequence, the nuclear stable high copy plasmids and ability to secrete the target protein[11]. *Saccharomyces* strains have high copy stably-inherited plasmid of 6.3 kb known as 2-micron plasmid, which codes for 4 genes *FLP*, *REP1*, *REP2* and *D*. It also contains an open reading frame (ORF), STB locus (required in *cis* for stabilization) and two 599 bp inverted repeat sequences. *FLP* encodes a site-specific recombinase which promotes flipping about the *FLP* recombination targets (FRT) within the inverted repeats, so that cells contain two forms of 2-micron plasmids, A and B. The simple 2-micron shuttle vectors contain the 2-micron ORI-STB, a yeast selectable marker and bacterial plasmid sequence (*ori* and selection markers) and are used in host strains which supply REP1 and REP2 proteins.

These lower eukaryotic systems are able to glycosylate the target proteins, but it has been shown that both N- and O-linked oligosaccharide structures are however, significantly different from their mammalian counterparts[12,13]. Hypermannosylation (addition of a large number of mannose residues to the core oligosaccharide) is a common feature in yeast, hindering proper folding and therefore the activity of the protein. At the moment yeast provides a good compromise between bacteria on one side and mammalian cell lines on the other.

Insect

The baculoviruses have emerged as a popular system for overproducing recombinant proteins in eukaryotic

cells[14,15]. Several factors have contributed to their popularity. Being eukaryotes, they use many of the protein modifications, processing, and transport systems present in higher eukaryotic cells[16]. They use a helper-independent virus that can be propagated to high titers in insect cells adapted for growth in suspension cultures, making it possible to obtain large amounts of recombinant proteins with relative ease. Expressed proteins are usually expressed in the proper cellular compartment, i.e. membrane proteins are usually localized to the membrane, nuclear proteins to the nucleus and secreted proteins secreted into the medium. Majority of the overproduced proteins remain soluble in insect cells. Viral genome is large (130 kb) and thus can accommodate large fragments of foreign DNA. Baculoviruses are non-infectious to vertebrates and their promoters have been shown to be inactive in mammalian cells, which gives them a possible advantage over other systems when expressing oncogenes or potentially toxic proteins. Also the process development time is short. Expression using baculoviral vectors also has some limitations. Since baculoviruses infect invertebrates, it is possible that the processing of proteins produced by vertebrates is different and this seems to be the case for some post-translational modifications, e.g. internal proteolytic cleavages at arginine- or lysine-rich sequences are highly inefficient. The glycosylation capability is generally limited to producing only high mannose type and not processed to complex type oligosaccharides containing fucose, galactose and sialic acid.

Mammalian cells[17,18]

Ideally, proteins requiring mammalian post-translational modifications should be expressed in mammalian cells. If product authenticity is absolutely essential for clinical efficacy, then despite the many shortcomings, a mammalian host is the only choice, as it offers the greatest degree of product fidelity. It should, however, be noted that oligosaccharide processing is species- and cell type-dependent among mammalian cells. Differences in glycosylation pattern are reported in rodent cell lines and human tissues. Even the use of human cell line is not perfect, since the transformation event required in most cases to produce a stable cell line may itself result in altered glycosylation profiles. Also mammalian expression techniques are time consuming and much more difficult to perform on a large scale. Complex nutrient requirement and low product concentration have meant that the end product must be highly value-added for this approach to be commercially viable.

Dictyostelium discoideum

Recently, the cellular slime mold *Dictyostelium discoideum* has been developed as an alternative eukaryotic system for expressing recombinant proteins[19,20]. Circular plasmids are common in prokaryotes, but only a few eukaryotes have been identified and studied for having circular nuclear plasmids and *Dictyostelium* is one of them[21,22]. The cellular slime molds have families of similar plasmids which are found in the nuclei of different species. They are organized differently from the yeast 2-micron like plasmids. *Dictyostelium* plasmids are packaged in a nucleosomal structure similar to the chromatin organization of higher eukaryotes. The best characterized family is that of the Ddp2-like plasmids. These plasmids are small (4.4–5.8 kb), encode ORF and contain an inverted repeat. The product of ORF is required in *trans* for plasmid maintenance, while an approximately 600 bp fragment, presumably containing the origin of replication is required in *cis*. The second best characterized plasmid family is Ddp1, found in wild type isolates like NC4 and V12. Ddp1 is a 13.7 kb plasmid which is present at an estimated copy number of 50–100 per cell and encodes at least 5 growth-specific and 5 development-specific transcripts. Although Ddp1 is larger and highly transcribed than the plasmids of the Ddp2-like family, none of the known transcripts seems to be essential for its replication. However, while approximately 1.2 kb region appeared to carry all the elements necessary for extra chromosomal replication, in the absence of selection, this fragment is quickly lost from the sub-population. In addition, deletion of plasmid sequences outside the region results in reduction of plasmid copy number. Therefore at least some of the Ddp1-encoded genes are required for the long-term maintenance of the plasmid at its normal copy number.

The development of reliable transformation systems for *D. discoideum* has provided the possibility of expressing heterologous genes in this microbe[23,24]. *Dictyostelium* is a simple eukaryotic micro-organism with a haploid genome of 5×10^7 bp and a lifecycle that alternates between single-celled and multicellular stages. They grow to a high cell density without the serum factors or special aeration needed by animal cell cultures. There is no cell wall and the high copy number plasmid vectors allow the expression of protein in cell-associated, membrane-attached or secreted form under the control of regulatable promoters[25]. The cells of *Dictyostelium* can carry out both O- as well as N-glycosylation[26,27]. The major advantages of this system include a very simple and cheap growth medium and the potential for large-scale production of proteins.

Factors affecting intracellular expression

Having the target DNA in an appropriate vector in the expression system of choice is the first step in optimizing production of the heterologous proteins. Within a given system the transcription and translation processes leading to the heterologous protein production are a complex set

of reactions. Each process is carried out and controlled by several enzymes/factors. In recent years we have learnt that the following few key steps or reactions are critical in determining the ultimate outcome.

Initiation of transcription

Gene expression is most frequently regulated at the level of transcription and it is generally assumed that the steady-state mRNA level is a primary determinant of the final yield of a foreign protein. The mRNA level is determined both by the rate of initiation and the rate of turnover. In most cases the yield of a foreign protein expressed using a host promoter has been much lower than the yield of the homologous protein, using the same promoter (~ 50%) (refs 28, 29). Many factors could account for these differences, e.g. downstream activating sequences (DAS) and upstream activating sequences (UAS)[30]. If DAS are characterized, it may be possible to incorporate them into upstream promoter fragments, in order to create more efficient expression vectors.

Alternatively, if DAS prove to be strongly position-dependent, they could be placed within an intron which could be excised prior to translation. If neither of these options works, then maximal transcription will only be possible using fusion proteins.

RNA elongation

The elongation of transcripts is not thought normally to affect the overall rate of transcription, but the yield of full length transcripts could be affected by fortuitous sequences in foreign genes which cause pausing or termination. These could either act in the same way as natural host terminator or else by a different mechanism, e.g. in yeast, though not widely recognized, this problem could be a very common reason for low yields or complete failure of expression of foreign genes. At present the only solution is to increase the AT/GC of the offending section of genes by chemical synthesis.

RNA stability[31,32]

There is evidence that subtle changes in mRNA sequence affect the stability of mRNA, with low mRNA stability being a primary factor in poor yields of foreign proteins. Where mRNA instability is diagnosed as a problem, overall yield might be improved by (i) using a more powerful promoter, (ii) using a promoter with more rapid induction kinetics or (iii) chemically synthesizing the gene with altered codons or deleting the 3′ untranslated region in the hope that the instability determinant will be removed. Degradation of mRNA is also more pronounced under adverse growth conditions.

Gene dosage

Since the target gene is often incorporated into a plasmid vector system, gene dosage is dependent on plasmid copy number. As can be expected, an increase in copy number results in concomitantly higher recombinant protein productivity, but not indefinitely. Plasmid copy number is affected by plasmid and host genetics and also by cultivation conditions such as growth rates, media and temperature[33].

Initiation of translation[34]

Translational efficiency is a function of either translational initiation or elongation rate. Translational efficiency is controlled primarily by the rate of initiation. Initiation in eukaryotes is thought to follow a scanning mechanism, whereby the 40 S ribosomal subunit plus co-factors bind the 5′ cap of the mRNA and then migrate down the untranslated leader scanning for the first AUG codon. Any part of this process, which is affected by the structure of the leader and the AUG content, could limit the initiation rate. AUG is recognized efficiently as initiation codon only when it is in the right context and optimal content is found to be GCC(A or G)CC**AUGG**. The purines (A or G) three bases before AUG and G immediately following it are found to be the most important, influencing translation to the tune of 10-fold. The following factors have also been found to be important for prokaryotes: (i) the ribosome-binding nucleotide sequence or Shine–Dalgarno (S–D) sequence; (ii) the distance between the initiation codon and S–D, and (iii) the secondary structure of mRNA.

Translational elongation

Translational elongation does not effect the yield or quality of polypeptide normally, but it can become limiting with very high mRNA levels. Codon usage is considered a potential factor affecting product yield. Despite the degeneracy of the genetic code, a non-random codon usage is found in most organisms. The codon usage of most genes reflects the nucleotide composition of the genome; highly-expressed genes show a strong bias towards a subset of codons[35]. This major codon bias, which can vary greatly between organisms, is thought to be a growth optimization strategy such that only a subset of tRNAs and aminoacyl-tRNA synthetases are needed at high concentration for efficient translation of highly-expressed genes at fast growth rates[36]. Rare codons, for which the cognate tRNA is less abundant, are translated at a slower rate, but this will not normally affect the level of product from an mRNA, since initiation is usually rate-limiting[37]. A ribosome finishing translation of one mRNA molecule is most likely to initiate translation of a different mRNA

species, unless the original species comprises a large proportion of the total mRNA. Thus, the overall rate of translation of an mRNA is not usually affected by a slower elongation rate unless ribosomes become limiting, which would affect all transcripts in the cell. In contrast to the normal situation, there is evidence that codon usage may affect both the yield and quality of a protein when a gene is transcribed to very high levels. With very high levels of mRNA containing rare codons, aminoacyl-tRNAs may become limiting, increasing the probability of mistranslation[38], which is the incorporation of an amino acid which does not correspond to the codon being translated, and possibly causing ribosomes to drop off. Thus the codon content of a foreign gene may influence the yield of protein where the mRNA is produced at very high levels. This may more likely occur on growth in minimal medium, when the cell produces a wide variety of biosynthetic enzymes encoded by genes containing rare codons. The effect on product quality has been difficult to measure, but requires further attention since it has further implications for therapeutic proteins. Proteins containing amino acid mis-incorporation are difficult to separate and may affect the activity and antigenicity of the product. Since small genes are now frequently synthesized chemically, they may be easily and perhaps profitably engineered to contain optimal codons for high-level expression. mRNA secondary structure, in addition to codon usage may affect translational elongation[39].

Polypeptide folding

During or following translation, the polypeptide must fold so as to adopt its functionally-active conformation. Since many denatured proteins can be refolded *in vitro*, it appears that the information for correct folding is contained in the primary polypeptide structure[40]. However, folding comprises rate-limiting steps during which some molecules may aggregate, particular!y at high rates of synthesis and at higher temperatures. There is evidence that certain heat shock proteins act as molecular chaperones in preventing the formation and accumulation of unfolded aggregates, while accelerating the folding reactions. Due to the intrinsic nature of polypeptide folding and low specificity of chaperones, it is very unlikely that foreign cytosolic proteins will accumulate in non-native conformations, but when fragments of proteins or fusion proteins are expressed, normal folding domains may be perturbed, resulting in an insoluble product. Insoluble proteins can often be renatured *in vitro*, though the techniques for this can be complex and unpredictable[41]. In contrast to intracellular proteins, naturally secreted proteins encounter an abnormal environment in the cytoplasm; disulphide bond formation is not favoured and glycosylation cannot occur. In *E. coli*, foreign proteins are frequently insoluble, but low temperature has been found to increase solubility in some cases[42]. This may be due to a decreased translation rate or to the fact that hydrophobic interactions, such as those occurring in aggregates, become less favourable. A dramatic increase in the yield of active, soluble protein is observed on reducing the rate of induction[43].

Post-translational processing

Prokaryotic expression systems are generally useful for producing heterologous proteins from cloned eukaryotic cDNA. In some cases however, eukaryotic proteins that have been synthesized in bacteria are either unstable or lack biological activity. The inability of prokaryotic organisms to produce authentic versions of proteins is for the most part due to the absence of appropriate mechanisms for generating certain post-translational modifications. In eukaryotes there are a number of modifications that may occur at the post-translational stage, after protein synthesis is complete.

Amino-terminal modifications of polypeptides: These are the most common processing events and occur on most cytosolic proteins[44]. Two types of events normally occur – removal of the N-terminal Met residue, catalysed by Met aminopeptidase (MAP) and acetylation of the N-terminal residue, catalysed by N-acetyltransferase (NAT). Both enzymes are associated with ribosomes and act on nascent polypeptide. In most cases the structure of the N-terminus should not affect the biological activity of a protein, but there may be exceptions, e.g. the response of haemoglobin to physiological modifiers involves the N-terminus and correct folding of alpha and beta globins is therefore advantageous. Similarly N-acetylation of melanocyte-stimulating hormone is required for full biological activity.

Disulphide bond formation: In eukaryotes, formation of disulphide bond (cys-s-s-cys) occurs in the lumen of RER and is mediated by an enzyme called disulphide isomerase[45]. Disulphide bond is confined to secretory proteins and exoplasmic membrane proteins. This is important in stabilization of tertiary structure. An improperly folded protein is unstable and lacks activity.

Proteolytic cleavage of a precursor form: This is required in some cases. Selected segments of amino acid sequences are removed to yield a functional protein[46].

Glycosylation: This is the most extensive of all the post-translational modifications and has important function in secretion, antigenicity and clearance of glycoproteins[47]. Oligosaccharides can attach to proteins in three ways: (i) Via an N-glycosidic bond to the R-group of an Asn-residue within the consensus sequence Asn-X-Ser/

Thr (N-glycosylation)[48]. All mature N-linked glycan structures have a common core of Man.GlcNAc, which can form part of simple oligomannose structures or be extensively modified by other residues such as fucose, galactose and sialic acid. Hybrid structures also exist where one or more arms of the glycan are modified and the remaining arms contain only mannose. (ii) Via an O-glycosidic bond to the R group of the Ser or Thr (O-glycosylation). O-linked glycosylation is extensive in structural proteins such as proteoglycans. Small glycan structures can also be O-linked to the side chain of hydroxylysine or hydroxyproline. (iii) Carbohydrates are also components of the glycophosphotidylinositol anchor used to secure some proteins to the cell membrane. The presence of these consensus sequences by no means guarantees their glycosylation. They show varying degrees of occupancy with oligosaccharides (macroheterogeneity) depending on their position within the protein and its conformation, the host cell type used for expression and its physiological status. These three factors also determine the extent of variation in the type of sugar residues found within each oligosaccharide (microheterogeneity). Glycosylation is both organism- and cell type-specific and therefore expression of a protein in a heterologous system will almost certainly result in a product with different modification from the native protein. This may affect the function and immunogenicity of the protein[49,50].

Modification of amino acid within proteins: Modifications of this type include phosphorylation, acetylation, sulphation, acylation (carboxylation, myristylation and palmitylation)[51]. Of these modifications, prokaryotic host cells are least likely to carry out either proper glycosylation or additions to specific amino acids within the heterologous protein. Moreover, no single eukaryotic host cell system is capable of performing all the possible post-translational modifications for every potential heterologous protein. Therefore, if a particular protein requires a specific set of modifications, then it may be necessary to examine different eukaryotic expression systems to find the one that can produce a biologically authentic product.

Stability of intracellular proteins

So far processes affecting the rate of synthesis of proteins have been considered, but the ultimate yield is equally affected by the rate of degradation[52]. Yields might logically be improved by the following measures: (i) fusion to a stable protein[53,54]; (ii) secretion to segregate the product from intracellular proteases[55]; (iii) using a more rapidly induced promoter; (iv) using additional protease inhibitors to minimize degradation during extraction; (v) inducing at lower temperature; and (vi) harvesting cells in the exponential growth phase.

Stability of plasmid

Plasmid instability is a major problem in continuous and large-scale fermentation, since these cultures go through many generations. The resulting effects are lower productivity and increased production cost, because of the build-up of non-productive plasmid-free cells. Plasmid instability is categorized as segregational instability[56,57] and structural instability[58,59]. Segregational instability is the loss of plasmid from one of the daughter cells during division because of defective partitioning. Structural instability is attributed to deletions, insertions and rearrangements in the plasmid structure, resulting in the loss of the desired gene function. Plasmid stability is influenced by the vector and host genotypes; the same plasmid in different hosts exhibits different degrees of stability and vice versa. The origin and size of foreign DNA have been observed to affect the plasmid stability. Plasmid stability is also a function of physiological parameters that affect the growth rate of the host cell, which include pH, temperature, aeration rate, medium components and heterologous protein accumulation. Mathematically structured and unstructured kinetic models of plasmid stability have been developed which are ultimately useful for the design of recombinant processes.

Summary and conclusion

The eventual objective of producing a desired protein in an economical heterologous host is influenced by a variety of factors. However, maximizing production of heterologous proteins for commercial application is still an art. We have begun to understand factors influencing the eventual production. These factors, described in detail in this article are varied and at times poorly understood. Largely the approach remains empirical. However, our collective experience will permit us to rationalize our approach in designing heterologous production of commercially important proteins in a variety of expression systems. Subsequent to production, stabilization and formulation of proteins will pose significant hurdles in utilizing the natural biological catalysts and other proteins for therapeutic and industrial purposes.

1. Ghosh, P. K., *Aust. Biotechnol.*, 1998, **3**, 214–222.
2. Padh, H., *BioPharma (USA)*, 1999, **12**, 18–19.
3. Ikeno, M. *et al.*, *Nat. Biotechnol.*, 1998, **16**, 431–439.
4. Shatzman, A. R. and Rosenberg, M., *Methods Enzymol.*, 1987, **152**, 661–673.
5. Casadaban, M., Martinez-Arias, A., Shapira, S. and Chou, J., *Methods Enzymol.*, 1983, **100**, 293–308.
6. Maina, C. V. *et al.*, *Gene*, 1988, **74**, 365–373.
7. Marston, F. A. O. and Hartley, D. L., *Methods Enzymol.*, 1990, **182**, 264–276.
8. Robinson, M. *et al.*, *Nucleic Acids Res.*, 1984, **12**, 6663–6671.

9. Schein, C. H., *Bio/Technology*, 1989, **7**, 1141–1148.
10. Hitzeman, R. A., Hagie, F. F., Levine, H. L., Goeddel, D. W., Ammerer, G. and Hall, B. D., *Nature*, 1981, **293**, 717–732.
11. Hitzeman, R. A. *et al.*, *Methods Enzymol.*, 1990, **185**, 421–440.
12. Kukuruzinska, M. A., Bergh, M. L. E. and Jackson, B. L., *Annu. Rev. Biochem.*, 1987, **56**, 915–944.
13. Kornfeld, R. and Kornfeld, S., *Annu. Rev. Biochem.*, 1985, **54**, 631–664.
14. Kitt, P. A. and Possee, R. D., *BioTechniques*, 1993, **14**, 810–817.
15. O'Reilly, D. R., Miller, L. K. and Lucknow, V. A., *Baculovirus Expression System – A Laboratory Manual*, Oxford University Press, New York, 1992.
16. Matsuura, Y., Possee, R. D. and Bishop, D. H. L., *J. Gen. Virol.*, 1987, **68**, 1233–1250.
17. Kaufman, R. J., *Methods Enzymol.*, 1990, **185**, 537–566.
18. Kaufman, R. J., *Methods Enzymol.*, 1990, **185**, 487–511.
19. Dittrich, W., Williams, K. L. and Slade, M. B., *Biotechnology (NY)*, 1994, **12**, 614–618.
20. Tiltscher, H. and Storr, M., *Appl. Microbiol. Biotechnol.*, 1993, **40**, 246–250.
21. Ashktorab, H. and Welker, D. L., *Gene*, 1988, **65**, 41–49.
22. Ahern, K. G., Howard, P. K. and Firtel, R. A., *Nucleic Acids Res.*, 1988, **16**, 6825–6837.
23. Firtel, R. A., Silan, C., Ward, T. E., Howard, P., Metz, B. A., Nellen, W. and Jacobson, A., *Mol. Cell. Biol.*, 1985, **5**, 3241–3250.
24. Hughes, J. E., Podgorski, G. J. and Welker, D. L., *Plasmid*, 1992, **28**, 37.
25. Manstein, D. J., Schuster, H. P., Morandins, P. and Hint, D. M., *Gene*, 1995, **162**, 129–134.
26. Jung, E. and Williams, K. L., *Biotechnol. Appl. Biochem.*, 1997, **25**, 3–8.
27. Jung, E., Gooley, A. A., Packer, N. H., Slade, M. B., Williams, K. L. and Dittrich, W., *Biochemistry*, 1997, **36**, 4034–4040.
28. Mellor, J., Dobson, M., Roberts, N. A., Kingsman, A. J. and Kingsman, S. M., *Gene*, 1985, **33**, 215–226.
29. Chen, C. Y., Oppermann, H. and Mitgeenan, R. A., *Nucleic Acids Res.*, 1984, **12**, 8951–8970.
30. Purvis, I. J., Betlany, A. J. E., Loughlin, L. and Brown, A. J. P., *Nucleic Acids Res.*, 1987, **15**, 7963–7974.
31. Humphrey, T., Sadhale, P., Platt, T. and Proudfoot, N., *EMBO J.*, 1991, **10**, 3503–3511.
32. Romanos, M. A. *et al.*, *Nucleic Acids Res.*, 1991, **19**, 1461–1467.
33. Hughes, J. E. and Welker, D. L., *Plasmid*, 1989, **22**, 215–223.
34. Kozak, M., *J. Cell Biol.*, 1989, **108**, 229–241.
35. Bennetzen, J. L. and Hall, B. D., *J. Biol. Chem.*, 1982, **257**, 3026–3031.
36. Kurland, C. G., *Trends Biochem. Sci.*, 1987, **12**, 126–128.
37. Pedersen, S., *EMBO J.*, 1984, **3**, 2895–2898.
38. Scorer, C. A., Carrier, M. J. and Rosenberger, R. F., *Nucleic Acids Res.*, 1991, **19**, 3511–3516.
39. Baim, S. B., Pietras, D. F., Eustice, D. C. and Sherman, F., *Mol. Cell. Biol.*, 1985, **5**, 1839–1846.
40. Gething, M. J. and Sambrook, J., *Nature*, 1992, **355**, 33–45.
41. Marston, F. A. O., *Biochem. J.*, 1986, **240**, 1–12.
42. Schein, C. H. and Noteborn, M. H. N., *Bio/Technology*, 1988, **6**, 291–294.
43. Kopetzki, E., Schumacher, G. and Buckel, P., *Mol. Gen. Genet.*, 1989, **216**, 149–155.
44. Kendall, R. L., Yamada, R. and Bradshaw, R. A., *Methods Enzymol.*, 1990, **185**, 398–407.
45. Freedman, R. B., *Cell*, 1989, **57**, 1069–1072.
46. Thim, L. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 6766–6770.
47. Rademacher, T. W., Parekh, R. B. and Dwek, R. A., *Annu. Rev. Biochem.*, 1988, **57**, 785–838.
48. Komfeld, R. and Komfeld, S., *Annu. Rev. Biochem.*, 1985, **54**, 631–664.
49. Parekh, R. B. *et al.*, *Biochemistry*, 1989, **28**, 7670–7679.
50. Kukuruzinska, M. A., Bergh, M. L. E. and Jackson, B. L., *Annu. Rev. Biochem.*, 1987, **56**, 915–944.
51. Tames, G. and Olson, E. N., *Biochemistry*, 1990, **29**, 2623–2634.
52. Dice, F., *FASEB J.*, 1987, **1**, 349–357.
53. Lees, N., Cozzjkorto, J., Wainwright, N. and Testa, D., *Nucleic Acids Res.*, 1984, **12**, 6797–6812.
54. Cousens, L. S. *et al.*, *Gene*, 1987, **61**, 265–275.
55. Itoh, Y. and Fujisawa, Y., *Biochem. Biophys. Res. Commun.*, 1986, **141**, 942–948.
56. Cashmore, A. M., Aibury, M. S., Hadfield, C. and Meacock, P. M., *Mol. Gen. Genet.*, 1986, **203**, 154–162.
57. Scott, J. R., *Microbiol. Rev.*, 1984, **48**, 1–23.
58. Novick, R. P., *Microbiol. Rev.*, 1987, **51**, 381–395.
59. Ahern, K. G., Howard, P. K. and Firtel, R. A., *Nucleic Acids Res.*, 1988, **16**, 6825–6837.

ACKNOWLEDGEMENTS. We thank the Council of Scientific and Industrial Research (project 38(973)99-EMR-II) for supporting research work in H.P.'s laboratory and award of Senior Research Fellowship to M.R.

Received 5 July 2000; revised accepted 15 December 2000