PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Host-Vector Systems

S. M. Kingsman and A. J. Kingsman

Phil. Trans. R. Soc. Lond. B 1989 **324**, 477-485
doi: 10.1098/rstb.1989.0061

Phil. Trans. R. Soc. Lond. B **324**, 477–485 (1989)
Printed in Great Britain

Host–vector systems

By S. M. Kingsman[1] and A. J. Kingsman[1,2]

[1] *Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.*
[2] *British Biotechnology Ltd, Watlington Road, Cowley, Oxford OX4 5LY, U.K.*

[Plate 1]

In 1980 it was only possible to express foreign genes in bacteria and a few easily cultured animal cells. During the subsequent eight years specialized vectors have been developed to allow the genetic manipulation of a wide range of both prokaryotes and eukaryotes. One of the major goals of biotechnology in 1980 was to use host cells as 'factories' for the production of proteins that were only available in minute quantities from natural sources. This has already lead to a new generation of pharmaceutical products. Advances in our understanding of host–vector systems have defined new goals. The basic concepts of expression vector design will be illustrated. Some of the new goals are discussed with particular reference to the exploitation of novel host–vector systems to develop vaccines and anti-viral agents against AIDS.

Introduction

Biotechnology is the exploitation of the biochemical potential of living organisms for medical, agricultural, industrial and social purposes. It is an ancient science but in the last 10–20 years the scope of biotechnology has been dramatically extended by the emergence of a new technology known as 'genetic engineering'. This is the modification of the genetic potential of a cell or an entire organism by artificially introducing foreign genes, that is genes or gene combinations that are not naturally found in that particular cell or organism.

There are four general steps in the process of genetic engineering. These are: (1) *gene cloning*, which is the purification of a specific gene from a complex genome; (2) *gene manipulation*, which involves rearranging the components of the gene and/or making specific alterations (this latter process is called site-directed mutagenesis); (3) *gene transfer*, which involves returning the gene to a living cell; and (4) *gene expression*, which involves 'forcing' the gene to function in the new host cell and to produce the end product, which is in most cases an active protein. The last two stages, namely gene transfer and gene expression, depend upon the use of highly specialized DNA molecules known as gene-expression vectors.

Most organisms have some unique features; this means that they can only be genetically engineered by using specialized gene expression vectors that accommodate their special features. The major limiting step in genetic engineering is therefore to understand enough of the basic genetics and biochemistry of an organism to design a vector that will be capable of introducing, maintaining and expressing new genes in that organism. In 1980, at the time of the Spinks report there was a good understanding of the biochemistry and genetics of the bacterium *Escherichia coli*, a simple prokaryotic organism, and of some simple viruses such as SV40 that infect higher (eukaryotic) cells. This knowledge allowed *E. coli* and animal cells to be engineered but no other system, not even the simple eukaryote yeast, could be modified to

express foreign genes. In 1988 the number of hosts, both prokaryotic and eukaryotic, that can be genetically engineered and the number of specialized gene expression vectors have dramatically increased (table 1). This is leading to a spectacular extension of the scope of biotechnology and the goals and possibilities are becoming increasingly broad. In this brief review we shall consider some aspects of this expansion in the range and sophistication of host–vector systems and attempt to highlight important future areas in biotechnology. We shall concentrate upon eukaryotic systems where there have been some major advances in the last eight years.

TABLE 1. HOST–VECTOR SYSTEMS FOR FOREIGN GENE EXPRESSION

(Further details and supporting references can be found in Kingsman & Kingsman (1988*a*).)

host	vector
E. coli *Bacillus* *Streptomyces* *Pseudomonas* *Salmonella* etc.	plasmids, phages
cultured animal cells yeast plants *Drosophila* insect cells farm animals differentiated human cells	plasmids, amplicons, replicons; viruses e.g. retroviruses, SV40 vaccinia, baculovirus; transposons; minichromosomes

BASIC FEATURES OF EUKARYOTIC GENE EXPRESSION VECTORS

There are two main classes of expression vectors: these are (1) autonomously replicating vectors and (2) integrative vectors.

The autonomously replicating vectors are usually circular, but more recently linear vectors have been constructed (see below). They contain specialized DNA signals that enable them to be replicated and stably maintained during cell division. These signals are called replication-maintenance signals. They are often derived from endogenous plasmids, e.g. the 2 μm circle in yeast (see, for example, Som *et al.* 1986) or from viruses such as bovine papillomavirus (Mecsas & Sugden 1987).

Integrative vectors contain specialized signals that facilitate their incorporation into the host-cell chromosomes where they are simply replicated and maintained as passengers in the chromosome. These signals are called integration-maintenance signals and they include chromosomal genes that target to the chromosome by homologous recombination as in yeast (Rothstein 1985), the long terminal repeats of retroviruses (Varmus 1988*a*) and the border sequences of the *Agrobacterium tumefaciens* Ti plasmid that is used to engineer plant cells (Klee *et al.* 1987).

An essential component of both these types of vector is the 'expression cassette'. This is the group of specialized DNA signals that 'force' the expression of the foreign gene. These signals are not universal and must be chosen to function efficiently in each different specific cell type. The key component of the expression cassette is the promoter. This is placed 'behind' (upstream of) a protein-coding region and is responsible for driving the production of mRNA. A generalized gene-expression vector is shown in figure 1.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

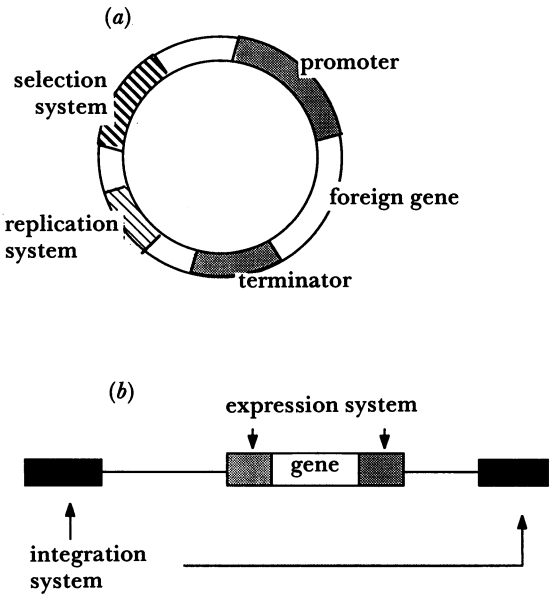

FIGURE 1. Generalized expression vectors: (*a*) autonomous vectors; (*b*) integration vectors.

The initial goal of genetic engineering was to convert cells into 'factories' for producing large quantities of natural proteins. For example, the insertion of the coding region for human interferon into the expression cassette of a yeast expression vector converts baker's yeast into a factor for producing human interferon (reviewed in Kingsman & Kingsman 1983). Many different coding regions have now been inserted into yeast, ranging from hirudin from the leech to acetylcholine receptors from the giant squid (reviewed in Kingsman *et al.* 1985, 1987). Many natural proteins are more complex than interferon and they must be modified inside the cell, e.g. by proteases, by the addition of carbohydrate or by other chemical modifications, before they are biologically functional. In these cases it is often essential to use mammalian cells. Many cells cannot perform all the types of modifications and therefore new host–vector systems must be developed. For example, a complex blood protein such as Factor XIII can be produced by using a vector that causes it to be secreted into the milk of genetically engineered mice (Gordon *et al.* 1987).

The amount of product and the economics of a process can be significantly influenced by manipulating the basic components of expression vectors. In addition new goals can be contemplated as the design of expression vectors becomes more sophisticated. Some of the advances in vector systems and their applications are discussed below.

ADVANCES IN EXPRESSION VECTOR DESIGN

In 1980 the basic structure of a promoter from a eukaryotic gene was not known. Today we know that it is composed of multiple DNA elements that function by interacting with many different proteins (reviewed by Ptashne 1988). The protein–DNA interactions serve to control the level of expression, the timing of expression (e.g. during the development of an organism) and the cell specificity for expression. Many of these features are in fact controlled by functionally independent components of the promoter which can therefore be viewed as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

composed of multiple modules. It is now possible to use this information to design artificial promoters that combine different numbers and types of modules. It is also possible to 'mix and match' modules from different promoters. In a typical example a bacterial regulatory system, the *lacO/lac* repressor system, has been inserted into the SV40 promoter to render it easily inducible by the simple chemical IPTG (Brown *et al.* 1987). In the future these artificial promoters will be incorporated into vectors to allow increased yields of foreign product and to optimize the production process. It will be possible to design expression systems that fit with optimized fermentation conditions by switching promoters on and off at precise times and by controlling them via standard fermentation parameters, e.g. nutrient supply or limitation. This is '*designer fermentation*'. In complex organisms (plants and animals) gene expression will be targeted only to specific tissues or developmental stages. This is the new area of '*genetic programming*', which is only made possible by using advanced promoter design.

In the future it will be important to express multiple foreign proteins in a single cell. There are several reasons for this: (1) modifying enzymes can be introduced to optimize production of a biologically active protein; (2) protein complexes can be assembled, e.g. to construct complex receptors or heterosubunit enzymes; and (3) multistep biochemical syntheses can be achieved. This latter is the new area of '*metabolic design*'. For example, it will be possible to assemble the pathways for producing plant secondary metabolites in much simpler organisms such as yeast. Such small chemicals have wide applications ranging from food technology to medicine. Initially the simplest systems will result in cytoplasmic synthesis but ultimately organelles will be engineered to confine substrates and products and optimize flux through the pathway.

The first steps towards metabolic design are already being taken by designing new replication and integration systems that allow multiple genes to be maintained in a single cell. For example, it is possible to assemble linear artificial chromosomes *in vitro* using the cloned functional elements, namely the telomeres, centromere and replication origins (Murray & Szostak 1983). Simple vectors cannot carry many additional genes but because chromosomes are normally large these artificial chromosomes can accommodate large stretches of additional DNA that could carry multiple genes (Burke *et al.* 1987). This technology is becoming established in yeast but may soon be possible in mamalian cells. Novel integration vectors are also being developed; one example is a transposon vector for yeast (Boeke *et al.* 1988). Transposons are pieces of DNA that disperse themselves in multiple copies throughout the genome via their specialized DNA signals and by means of proteins encoded by transposon genes (reviewed in Kingsman *et al.* 1988*b*). The transposon DNA signals and genes can be manipulated to produce vectors that can disperse additional genes throughout the chromosomes where they are stably maintained; by repeated rounds of transposition multistep pathway genes can be assembled in one genome. Metabolic design can be further optimized by using the artificial modular promoters described above so that the levels and timing of different pathway enzymes can be controlled to optimize the pathway for production.

In addition to advances in the design of promoters, and replication and integration systems there are other 'modules' that can be added to vectors to optimize and extend the potential of genetically engineered organisms. Many of these involve modifying the protein product itself; for example, a targeting signal can be added to direct the protein outside the cell (i.e. to the secreted), or to direct it to a particular subcellular organelle. Additional protein modules can be added to help purify the protein from the cell, e.g. the addition of an affinity-binding

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

domain that can be subsequently removed from the product by cleavage. As new modules are discovered they can be added to expression vectors and so increase their versatility (reviewed in Kingsman & Kingsman 1988*a*; Bebbington & Hentschel 1985).

Host–vector systems and aids

The importance of advances in host–vector systems design can be illustrated further by considering a single application, namely the use of biotechnology to combat the AIDS pandemic.

In 1980 the clinical picture of AIDS was barely described and it was not until 1983/4 that the causative agent human immunodeficiency virus (HIV) was identified and characterized (reviewed by Fauci 1988). Today HIV threatens to devastate the populations of large sections of Africa and major cities on all continents, and if unchecked poses a global threat. New host–vector systems are having impact in two key areas: the production of vaccines and the production of anti-viral agents.

It is generally believed that any vaccine against AIDS will need to be generated by recombinant DNA techniques. This is largely because the virus is dangerous to grow in large quantities and the extent of any physical or chemical treatment required to inactivate the virus to produce a vaccine would probably destroy the desired immunogenicity. Also problems of the variability of the virus, its natural weak immunogenicity and possible detrimental effects of parts of the viral proteins are best tackled by producing defined portions of viral proteins, and that is only feasible by genetic engineering. In addition it may be possible to enhance the immunogenicity by improving the antigen presentation by constructing novel recombinant proteins.

Viral antigens have been produced in a wide variety of host–vector systems (table 2). These generate antigens in a range of different forms for presentation to the immune system, ranging from simple monomers to high-molecular-weight particulate structures.

Table 2. Host–vector systems for the production of candidate HIV vaccines

(For further details see Adams *et al.* (1987*b*), Zagury *et al.* (1988), Zuckerman (1988) and Koff & Hoth (1988).)

host	vector	
E. coli	plasmid	monomeric antigens
insect cells	baculoviruses	monomeric antigens
animal cells	*Vaccinia* virus	a replicating antigen-presentation system
	DHFR amplicon	monomeric antigens
yeast	Ty–VLPs	particulate high-density antigen presentation

It is thought that antigen presentation will be critically important in producing an effective vaccine. One approach to achieving effective presentation is shown in figure 2, which depicts a novel yeast vector system for producing polyvalent particulate antigens. This is based on the discovery of a yeast protein called p1 encoded by the *TYA* gene of the Ty transposon. The p1 protein has the unusual property of assembling spontaneously inside the yeast cell into a highly ordered particle called a Ty virus-like particle (Ty-VLP) that is composed of about 300–400 copies of p1 (Adams *et al.* 1987*a*; Kingsman & Kingsman 1988*b*; Kingsman *et al.*

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

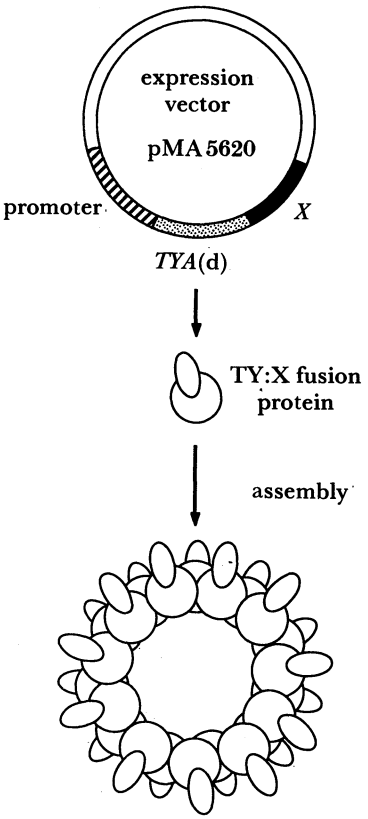

FIGURE 2. A particulate antigen-presentation system. The coding sequence for any antigen (*X*) is fused to the end of the *TYA* gene to produce a p–X fusion protein. This assembles into a virus-like particle in yeast cells (see text for further details).

1988). The coding region for any domain of any HIV protein can be fused to the end of the *TYA* gene to produce a p1 fusion protein with additional HIV amino acids. The size of the addition can range from 3 kDa to at least 45 kDa. The p1–HIV fusion protein still assembles into VLPs and these fill the yeast cell. They are easily harvested and purified and they display the antigens all over the surface of the particle (figure 3, plate 1). This polyvalent particulate structure is highly effective in presenting antigens to the immune system (Adams *et al.* 1987*b*; Kingsman & Kingsman 1988*c*). By using this host–vector system to make many different hybrid HIV: Ty–VLPs and by using different combinations of antigens it may be possible to induce effective immunity against HIV infection.

Novel host–vector systems are also being used to help develop anti-viral agents. The central problem is to identify a process that is unique to the virus so that basic cellular mechanisms are not damaged at the same time. One approach is to use knowledge about the virus life cycle to develop a rapid screen for compounds that specifically inhibit key steps in virus replication. One virus-specific step is at the point of the expression of viral genes, which depends on the TAT protein encoded by the virus and on a sequence called TAR in the virus genome (reviewed by Varmus 1988*b*). Specialized vectors can be constructed that fuse TAR to an easily monitered gene called a reporter gene and then stably incorporate the hybrid gene into an assay cell. Superinfection of the cell by HIV provides TAT and the activity of the reporter gene can be measured. When anti-viral chemicals are added that block HIV replication there is no

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

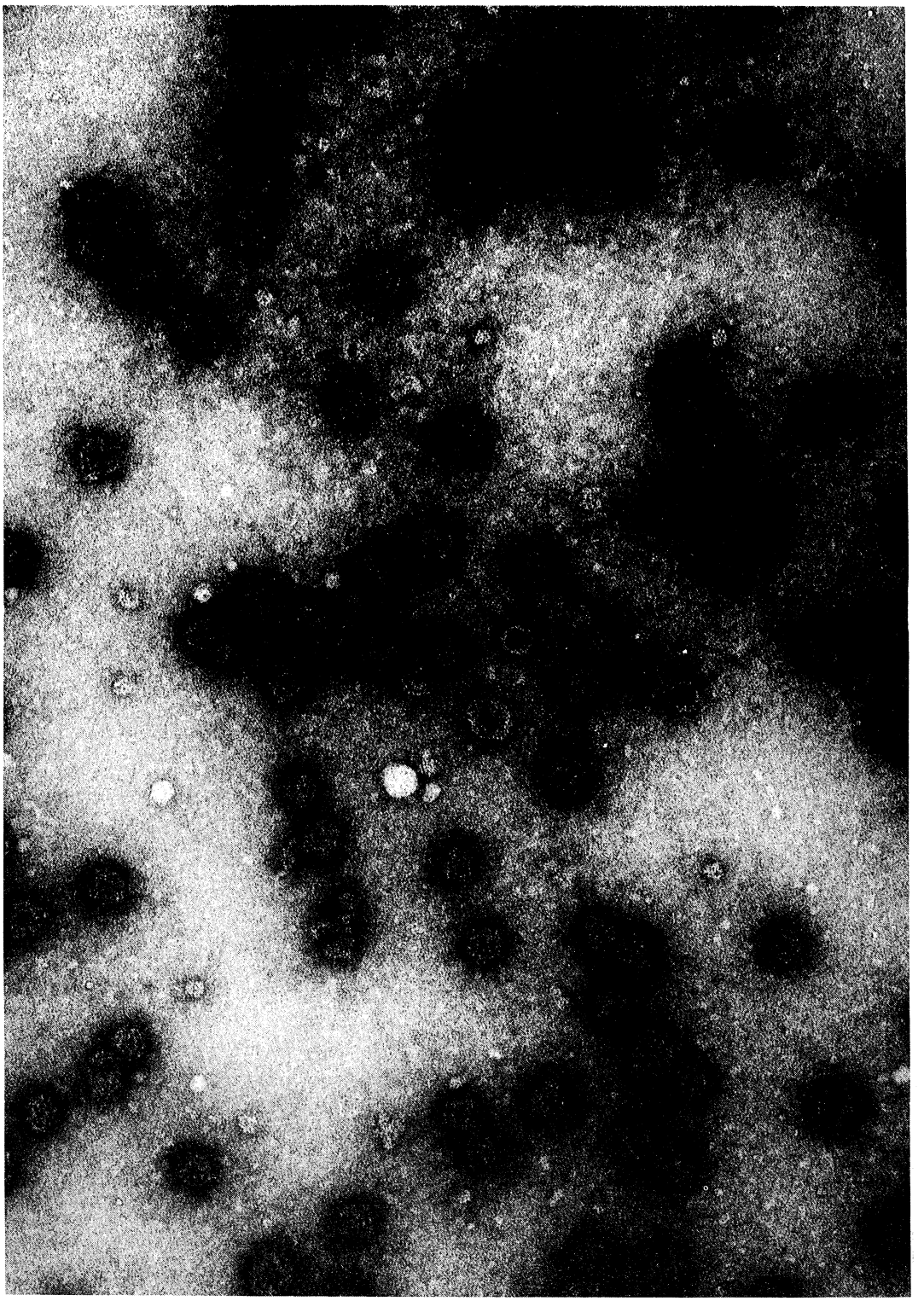

FIGURE 3. Electron micrograph of hybrid HIV: Ty-VLPs purified from yeast. (Magn. × 72900.)

(*Facing p.* 482)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TAT produced and therefore no reporter-gene activity (see Felber & Pavlakis 1988). This type of host–vector system for so-called '*smart screening*' rapidly accelerates the search for candidate anti-viral drugs.

An alternative approach is to design a drug specifically to inactivate a viral protein. This is called '*rational design*' and depends on understanding both the structure and function of the viral protein. Specialized host–vector systems are essential for rational design to allow the production of authentic viral proteins for structural studies and to assay biological activity. A general procedure for producing large quantities of active authentic viral proteins is based on the yeast Ty–VLP vector system and is outlined in figure 4. In this case the viral protein is linked to the p1 carrier particle via a short stretch of amino acids that contain the recognition site for the highly specific protease Factor Xa (Nagai & Thoresen 1984). The particles are easily purified, then treatment with Factor Xa liberates the additional viral protein and the carrier p1 particle is easily removed. This vector system has been used to produce substantially pure, biologically active TAT and to identify the key site of TAT action in the cell nucleus (Braddock *et al.* 1989).

These are just a few examples of the way in which the development of host–vector systems is having a major impact on approaches for vaccination and antiviral therapy for AIDS.

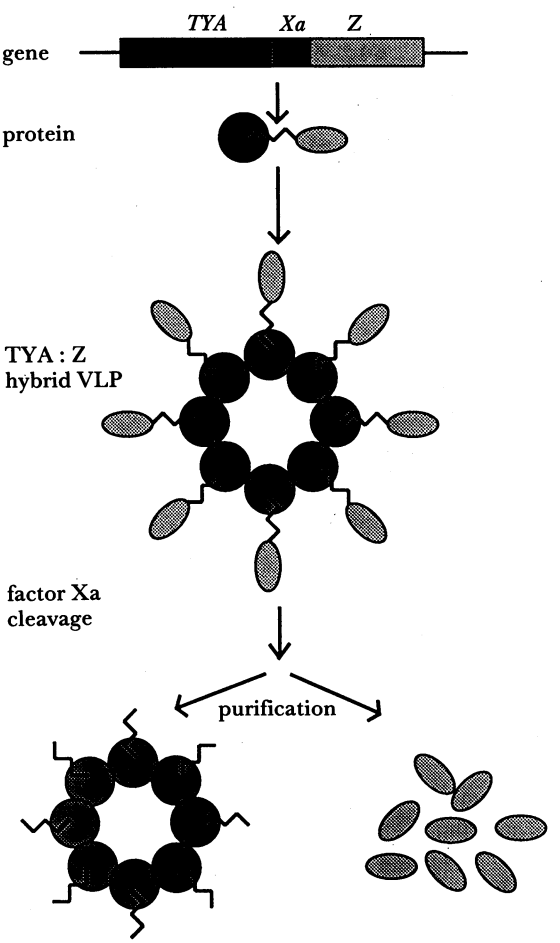

FIGURE 4. A general host–vector system for the production of authentic viral proteins. See text.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The future

The last eight years have demonstrated that as our basic understanding of biochemistry and genetics has increased so has our ability to genetically engineer cells. There has been a vast increase in the number of host–vector systems and in the sophistication of gene-expression technology. This has allowed us to achieve some of the goals outlined in the Spinks report but it has also broadened our horizons to incude possibilities not suspected in 1980. In table 3 we have outlined the areas where advances in host–vector systems will be of key importance in the future. The key areas are metabolic design, genetic programming, designer fermentation, 'smart screening', rational drug design and intracellular intervention.

Table 3. Future impact of advanced host–vector systems

protein production
second-generation products
custom fermentation
increased product
genetic programming
custom plants and animals
metabolic design
gene therapy
intracellular intervention
molecular virology, molecular oncology
improved drug development
mechanism-based screening
rational design
novel approaches to solve old and new problems
hepatitis
HIV
environment
food production

The technology will provide new solutions to old problems and prepare us to tackle problems as yet unsuspected. This is exemplified by the AIDS pandemic where it is clear that host–vector systems and related technologies offer hope for combatting a disease that 20 years earlier would have been an inevitable disaster. Finally we must remember that we are in the discovery business, and as long as we continue to probe the fundamental properties of living organisms then in the next eight years we will have generated even greater opportunities in biotechnology.

We thank the following for their support of research into host–vector systems and eukaryotic molecular genetics in our laboratory: AFRC, British Biotechnology Ltd, Celltech, Delta Biotechnology, Glaxo plc, MRC, SERC and the Royal Society.

References

Adams, S. E., Mellor, J., Gull, K., Tuite, M. F., Kingsman, A. J. & Kingsman, S. M. 1987*a* The functions and relationships of Ty–VLP proteins in yeast reflect those of mammalian retroviral proteins. *Cell* **49**, 111–198.

Adams, S. E., Dawson, K. M., Gull, K., Kingsman, S. M. & Kingsman, A. J. 1987*b* The expression of hybrid HIV: Ty virus like particles in yeast. *Nature, Lond.* **329**, 68–70.

Bebbington, C. & Hentschel, C. 1985 The expression of recombinant DNA products in mammalian cells. *TIBTECH* **3**, 314–317.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Boeke, J. D., Xu, H. & Fink, G. R. 1988 A general method for the chromosomal amplification of genes in yeast. *Science, Wash.* **239**, 280–282.

Braddock, M., Chambers, A., Wilson, W., Adams, S. E., Esnouf, M. P., Kingsman, A. J. & Kingsman, S. M. 1989 HIV LTR activation in *Xenopus* oocytes. *Cell* (Submitted.)

Browne, M., Figge, J., Hansen, U., Wright, C., Jeang, K.-T., Khoury, G., Livingston, D. M. & Roberts, T. M. 1987 *Lac* repressor can regulate expression from a hybrid SV40 early promoter containing a *lac* operator in animal cells. *Cell* **49**, 603–612.

Burke, D. T., Carle, G. F. & Olsen, M. V. 1987 Cloning of large segments of exogenous DNA in yeast by means of artificial chromosome vectors. *Science, Wash.* **236**, 806–813.

Fauci, A. S. 1988 The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science, Wash.* **239**, 617–622.

Felber, B. & Pavlakis, G. N. 1988 A quantitative bioassay for HIV-1 based on trans-activation. *Science, Wash.* **239**, 184–187.

Gordon, K., Lee, E., Vitale, J. A., Smith, A. E., Westphal, H. & Hennighausen, L. 1987 Production of human tissue plasminogen activator in transgenic mouse milk. *Biotechnol.* **5**, 1183–1187.

Kingsman, S. M. & Kingsman, A. J. 1983 The production of interferon in bacteria and yeast. In *SGM Symposium vol. 23* (*Interferons*) (ed. D. C. Burke & A. Morris), pp 211–234. Cambridge University Press.

Kingsman, S. M., Kingsman, A. J., Dobson, M. J., Mellor, J. & Roberts, M. J. 1985 Heterologous gene expression in *Saccharomyces cerevisiae*. In *Biotechnology and genetic engineering reviews* (ed. G. E. Russell), vol. 3, pp. 377–416. Newcastle upon Tyne: Intercept.

Kingsman, S. M., Kingsman, A. J. & Mellor, J. 1987 The production of mammalian proteins in yeast. *TIBTECH* **5**, 53–57.

Kingsman, S. M. & Kingsman, A. J. 1988*a* *Genetic engineering: an introduction to gene analysis and exploitation in eukaryotes*. Oxford: Blackwell Scientific Publishing.

Kingsman, S. M. & Kingsman, A. J. 1988*b* Ty, a retroelement moving forward. *Cell* **53**, 333–335.

Kingsman, S. M. & Kingsman, A. J. 1988*c* Polyvalent recombinant antigens: a new vaccine strategy. *Vaccine* **7**, 304–307.

Kingsman, A. J., Adams, S. E., Malim, M. H., Rathjen, P., Fulton, A. M. & Kingsman, S. M. 1988 *The yeast retrotransposon Ty and related elements. Society for General Microbiology Symposium vol.* 43 (ed. A. J. Kingsman, K. F. Chater & S. M. Kingsman), pp. 223–247. Cambridge University Press.

Klee, H., Horsch, R. & Rogers, S. 1987 *Agrobacterium* mediated plant transformation and its further applications in plant biology. *A. Rev. Pl. Physiol.* **38**, 467–486.

Koff, W. C. & Hoth, D. F. 1988 Development and testing of AIDS vaccines. *Science, Wash.* **241**, 426–432.

Mecsas, J. & Sugden, B. 1987 Replication of plasmids derived from bovine papilloma virus type 1 and Epstein–Barr virus in cells in culture. *A. Rev. Cell Biol.* **3**, 87–108.

Murray, A. W. & Szostak, J. 1983 Construction of artificial chromosomes in yeast. *Nature, Lond.* **305**, 189–193.

Nagai, K. & Thoresen, H. C. 1984 Generation of β-globin by sequence specific proteolysis of a hydrid protein produced in *Escherichia coli*. *Nature, Lond.* **309**, 810–812.

Ptashne, M. 1988 How eukaryotic transcription activators work. *Nature, Lond.* **335**, 683–689.

Rothstein, R. 1985 Cloning in yeast. In *DNA cloning volume II* (ed. D. M. Glover), pp. 45–66. Oxford: IRL Press.

Som, T., Armstrong, K. A., Volkert, F. C. & Broach, J. 1988 Autoregulation of a 2 μm circle gene expression provides a model for maintenance of stable plasmid copy levels. *Cell* **52**, 27–37.

Varmus, H. 1988*a* Retroviruses. *Science, Wash*, **240**, 1427–1435.

Varmus, H. 1986*b* Regulation of HIV and HTLV gene expression. *Genes Devt* **2**, 1055–1062.

Zagury, D., Bernard, J., Cheynier, R., Desportes, I., Leonard, R., Fouchard, M. *et al.* 1988 A group-specific anamnestic reaction against HIV-1 induced by a candidate vaccine against AIDS. *Nature, Lond.* **332**, 728–731.

Zuckerman, A. J. 1988 Prospects for vaccines against HIV. *Br. med. J.* **297**, 86–88.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

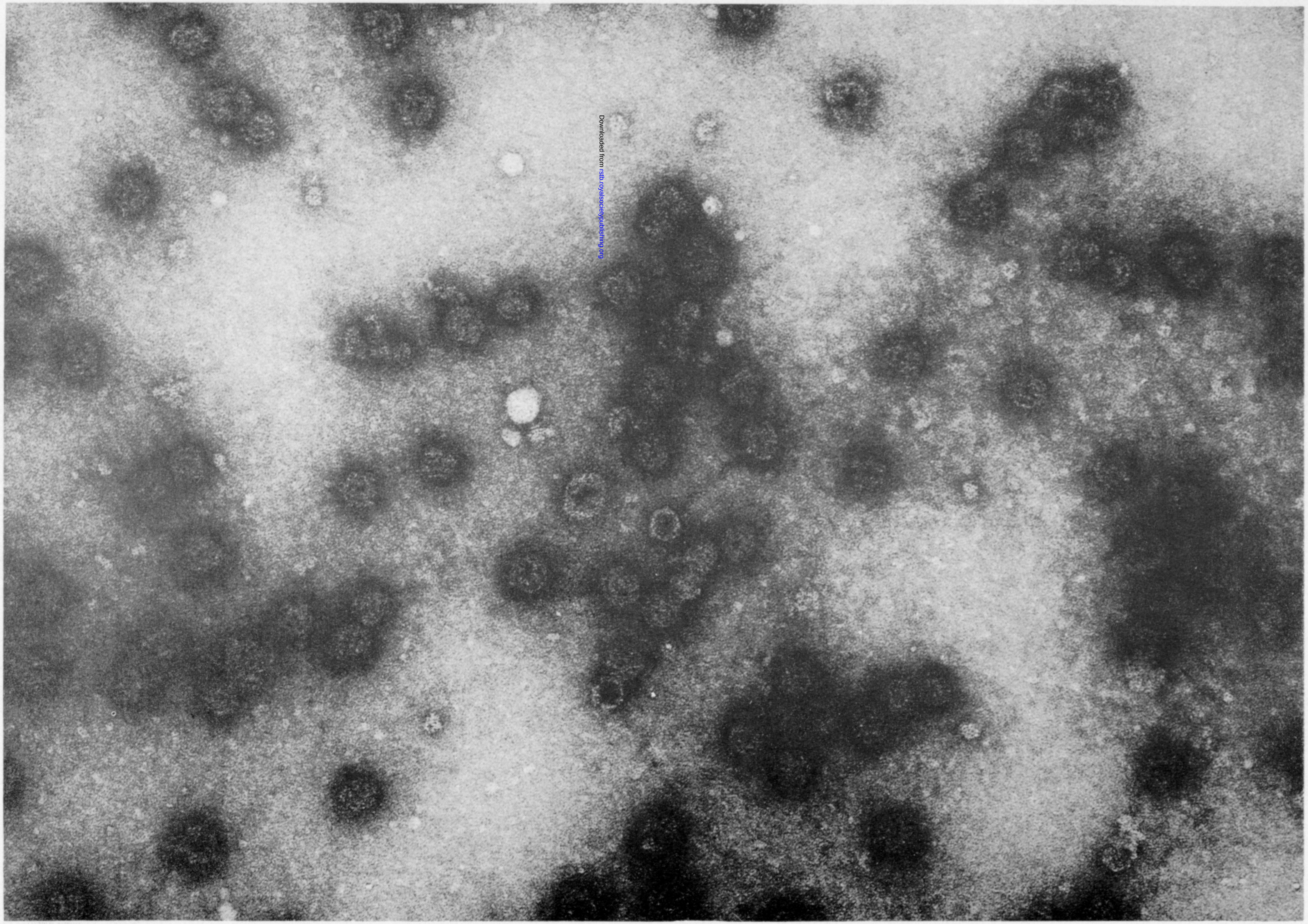

FIGURE 3. Electron micrograph of hybrid HIV : Ty–VLPs purified from yeast. (Magn. × 72 900.)