REPLICATION OF DEOXYRIBONUCLEIC ACID (DNA) WITH RESPECT TO GENE TECHNOLOGY

By KARL ESSER and BIRGITT OESER

Lehrstuhl für Allgemeine Botanik, Ruhr-Universität Bochum, Postfach 102148, D-4630 Bochum 1, Federal Republic of Germany.
Telephone: (0234) 700 2211; Telex: 0825860

SYNOPSIS

Nucleic acids do not only carry the genetic information, but are also the only substances being able of self-replication. Molecular cloning, an essential tool in biotechnology, requires among other things, an understanding of the mechanisms of replication which at present is fairly well known. After an introduction to the general principle, the status of art on replication procedure and its implication for biotechnology are dealt with.

Introduction

With exception of some viruses, the genetic information of all living beings consists of deoxyribonucleic acid (DNA). Nucleic acids are the only molecules which are able to replicate autonomously, *i.e.*, an identical carrier of genetic information is produced by copying a pre-existing DNA molecule. Thus replication is unrenouncable for continuous propagation of cells, and this provides the prerequisite for the expression of genetic information into a messenger (ribonucleic acid, RNA) and translation into a protein.

Thus replication is not only instrumental in each living cell, but must also take place as an initial step in molecular cloning, the basic procedure of modern gene technology.

From this follows that gene cloning particularly requires vigorous replication of the foreign DNA in a host cell. Only if this problem is successfully solved may one consider obtaining its product as a basis for scaling up for biotechnological exploitation.

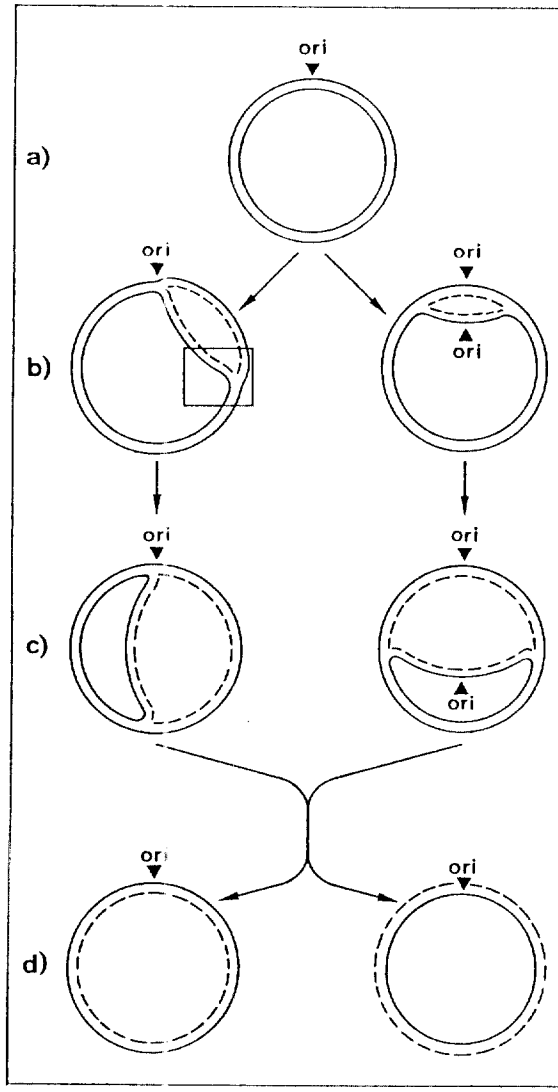
Mainly from experiments with the bacterium *Escherichia coli*, the molecular events of replication are understood, at least partially. It was found that each unit of replication (termed 'replicon' by Jacob and Brenner[1]) starts with a specific nucleotide sequence, which serves as a physiological signal for initiation, and which ends with a termination sequence. It is now generally accepted that this initiation sequence should be called 'origin of replication' (ori). In contrast to terminators which consist of defined linear DNA sequences,[2] origins of replication have complex structures. They have not only a specific structure in prokaryotes and eukaryotes, but also even show species' specificity in some eukaryotes. Thus research has focused more attention on oris than on terminators. In addition, terminators are neither present, nor required, for replication of circular DNA molecules. It is therefore understandable that, in a state of art paper written for biotechnologists, after the general principle of replication are briefly dealt with, the present knowledge of the identification and the structure of replication origins will be summarized, because the right choice of oris is essential in constructing vectors for either prokaryotic or eukaryotic cloning in biotechnology.[3,4]

The principle of gene cloning[5-8] is purposely neglected in this paper because it is supposed that it is familiar to the reader.

General Principle

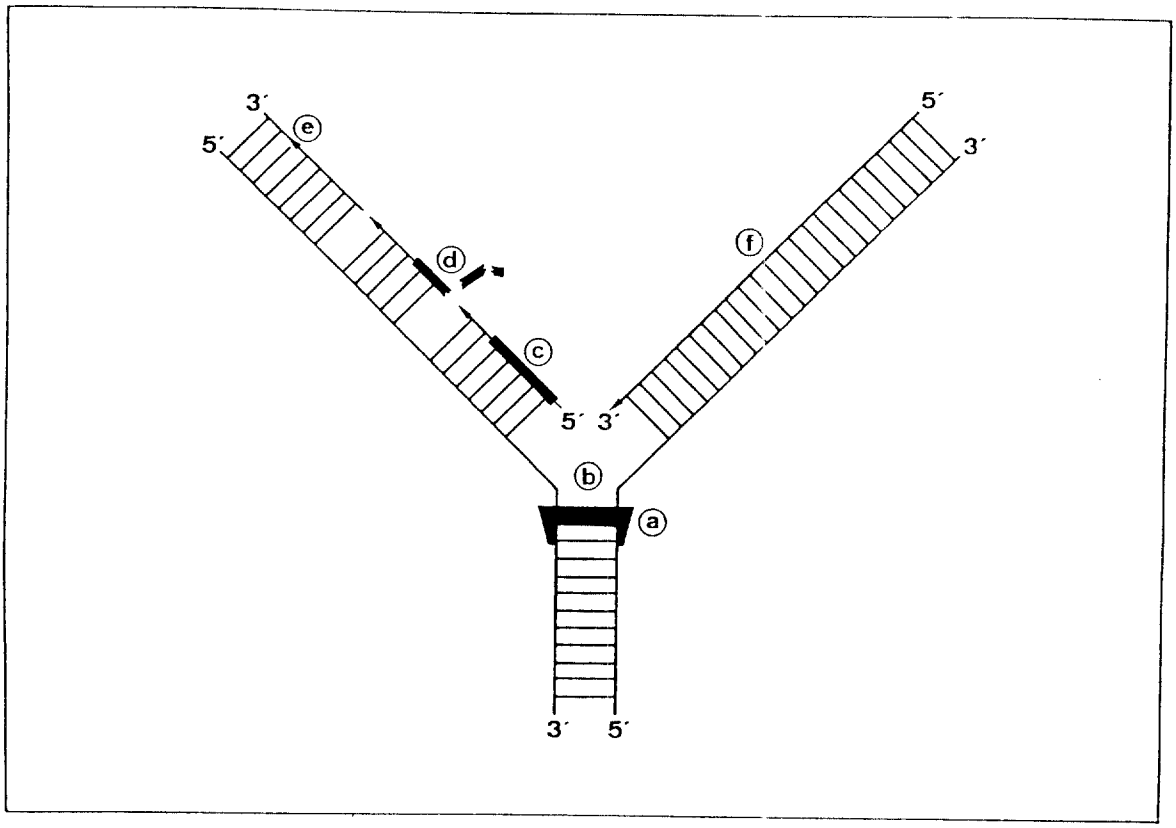
The general principle of replication,[9-11] as shown in Figure 1, lies in the formation of a complementary strand for each strand of the DNA helix. At first this requires a separation of the two strands: this is initiated at the origin of replication and comparable to the opening of a zip fastener. It becomes visible as replication loop.

Within a Y-like structure, called a replication fork, continuous addition of nucleotides takes place, as shown in Figure 2. The replication procedure comes to an end as soon as the



(a) Double stranded DNA molecule. The origin of replication is indicated. Replication starts with an opening of the two parental strands (replication loop) (b,c). This is followed alternatively either by an 'unidirectional (d) or bi-directoral (e)' synthesis of two new DNA strands (dashed lines). Eventually both procedures result in two new DNA molecules (f), each having a parental and a daughter strand. The rectangle in (b) encloses the so-called replication fork which is described in detail in Figure 2.[12]

Figure 1: Scheme of replication of a circular DNA molecule.



The general direction of replication is downstream from the 5′ end to the 3′ end of each DNA strand. From this follows that the replication procedure differs between the two strands of the fork. It is continuous at the right arm of the fork and discontinuous at the left arm. The molecular details are naturally alike and are easier to explain in using the right arm as an example. (a) The unwinding of the two strands characterized by 5′ and 3′ starts at the origin of replication. It is caused by an enzymatic 'swivel' which at the same time separates the two strands. (b) At the beginning of the fork the single stranded DNA regions are stabilized by DNA binding proteins. (c) On the left arm of the fork short RNA sequences (thick line) are synthesised. The 3′ end of each RNA primer serves as adhesive point for the polymerisation of DNA nucleotides caused by a complex DNA polymerase. (d) Subsequently the RNA primer is eliminated by an exonuclease. (e) Eventually the small DNA segments are ligated. For the sake of clarity some other enzymes or proteins respectively involved in the procedure of replication are not mentioned. The same holds true for differences in details between *E. coli* and other organisms. (f) Since on the right arm of the fork the replication is continuous a single prime event is sufficient for the initiation of a continuous DNA polymerisation which naturally does not require the action of ligases.

Figure 2: Simplified scheme of the molecular events taking place at the replication fork of a double stranded DNA molecule.[13-15] (Redrawn and altered from Knippers.[5])

replication fork reaches the ori from the other end. The two new molecules thus have each one parental and one newly synthetised strand.

In addition to this simple example of replication (un-directional replication) bi-directional replication was observed in many other cases. Again starting from the origin of replication, two replication forks become instrumental, but in the opposite direction (Figure 1c). The replication procedure ends as soon as the two forks meet each other (see Figure 1e). In linear DNA molecules both uni-directional and bi-directional replication may take place. However, in both cases specific terminator sequences determine the end-point of replication.

Identification and Isolation of Replication Origins

As is evident from the foregoing a replication origin is the unequivocal prerequisite for the start of replication. Therefore the identification and isolation of an ori is essential for molecular cloning. This may be achieved in two ways, indirectly and directly.

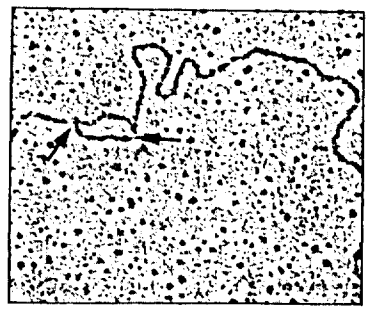
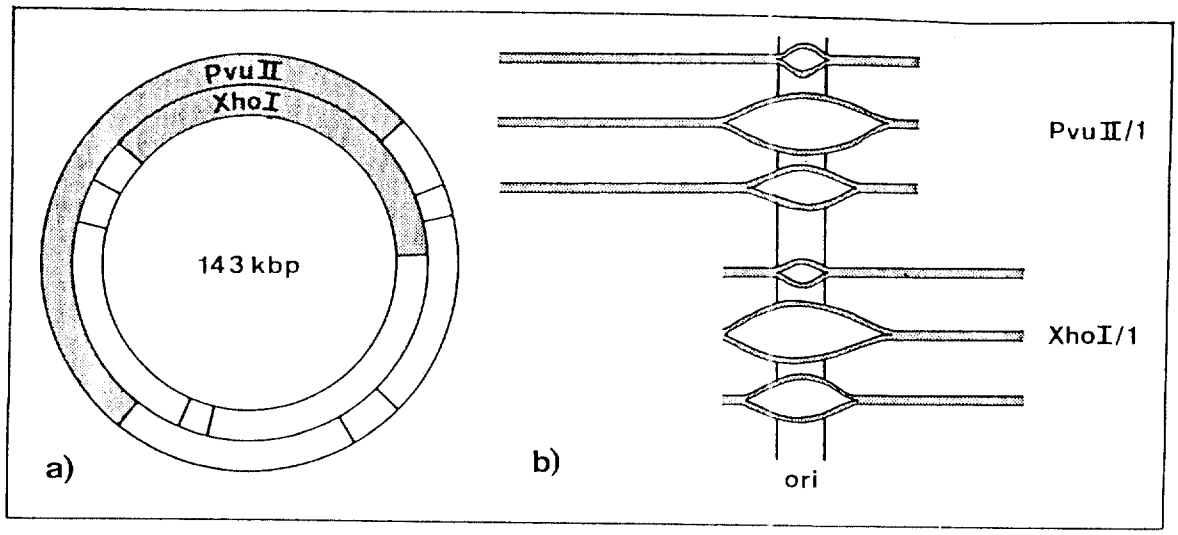


Figure 3: Electron micrograph of a replicating DNA molecule of *E. coli:* The replication loops are indicated by arrows. (Courtesy R. Eichenlaub, Hamburg).

Indirect method

The indirect method makes use of the so-called shot-gun cloning.[12] This consists of digesting DNA molecules by an



(a) The physical map shows the restriction sites of two endonucleases, PvUII and XhoI respectively.

(b) In electron micrographs of replicating DNA, digested with either one of the restriction enzyme, the various fragments showing replication loops may be identified according to their length. Thus it is possible to have the same origin as a common sequence of two different restriction fragments. In both cases there is naturally a difference in length of the two flanking regions. In the same fragments, in the present case PvuII/1 and XhoI/1, the replication procedure is in the various electron micrographs in different stages of development indicated by different length of the loop. In order to locate rather precisely the site of origin it is necessary to compare a number of different micrographs and to look for a region common to all loops. The accuracy of the localisation naturally depends on the number of samples assayed.

Figure 4: Identification of the site of a replication origin (ori) within the chloroplast DNA of *Euglena gracilis* having a length of 143 kpb. (From data of Ravel-Chapuis *et al.*[17]).

endonuclease and subsequent cloning into vectors being devoid of an ori. After introduction into the host, those vectors which replicate must have obtained a piece of DNA containing an ori. The identification of this ori on its original DNA is accomplished by an isolation of the replication vector and a subsequent identification of the segment of foreign DNA containing the ori by DNA/DNA-hybridisation. This technique, as exemplified later in another context, is sufficient for practical purposes if only a functional ori is needed. It does however not allow any conclusion of its site of location within the 'parental' DNA.

Direct method

In contrast, with the direct technique, which is more laborious than the indirect, one will obtain both identification and location of the ori.[16] The mere identification of oris is a rather simple procedure. The replication loops as diagrammed in Figure 1 may be easily seen in electron micrographs of replicating DNA (Figure 3). However the isolation of the ori from the DNA molecule is somewhat cumbersome because the sites of its location need to be determined. This requires the production of a physical map based on the restriction sites of various endonucleases and its correlation with the electron micrographs as exemplified by Figure 4.

Problems from both methods

Naturally one must be aware that with both techniques – the direct and the indirect method – a rather coarse isolation of an ori is obtained, because at both sides it is framed by unspecific DNA strands of different length. In many cases of practical applications this is unimportant. But sometimes if cloning of rather long DNA strands is desired, an origin of minimal size is required. This is obtained by shortening a DNA sequence in which an ori is identified by either one of the two techniques by a further endonuclease treatment. But in order to avoid a destruction of the ori an understanding of its specific structure is helpful.

Structure of Replication Origins

The knowledge of the structure of replication origins stems from DNA sequencing data and may be summarized as follows:

(1): All oris are predominantly located in AT-rich DNA segments.[18-21]

(2): Another specific feature of the oris is the formation of stem loop structures, as a result of short complementing of DNA sequences. Although the function of these loops is not yet understood, there seems to be a difference between prokaryotic and eukaryotic loops, irrespective from which eukaryotic DNA species (nuclear, mitochondrial, chloroplast) the ori is isolated (Figure 5).[12, 24-31]

(3): So far analysed there is indication that an ori consists of two domains: one is essential for replication and the other regulates the efficiency of replication.[19,30,32-34]

(4): Ori of related descent have so-called consensus sequences of various length. As may be seen from Figure 6, the consensus sequence of the various taxonomic groups do not only differ in length but also in structure.[19,23,29,35]

(5): Sometimes in the vicinity of an ori a promotor sequence (necessary for protein synthesis) was observed.[14,36,37]

It becomes evident from these criteria that our knowledge of the structure of origins of replication is rather meagre. From this follows that all manipulations performed with oris in molecular cloning are strictly empirical and have to be adapted to each specific case. One of the main handicaps is that there are no definite ideas about the length of an ori. For example, the minimal known length of the ori of the *E. coli* chromosome is 245 bp,[21] in contrast to 11 bp and 14 bp respectively of two putative origins of *Saccharomyces cerevisiae* chromosomal DNA.[32,38]

(a): Ori of the *E. coli* chromosome (Lewin[17]).
(b): Putative origin of chromosomal *S. cerevisiae* DNA (Feldmann *et al.*[22]).
(c): Ori of mitochondrial *S. cerevisiae* DNA (Bernardi[23]).
(d): Putative origin of *Nicotiana tabacum* chloroplast DNA (Ohtani *et al.*[24]).

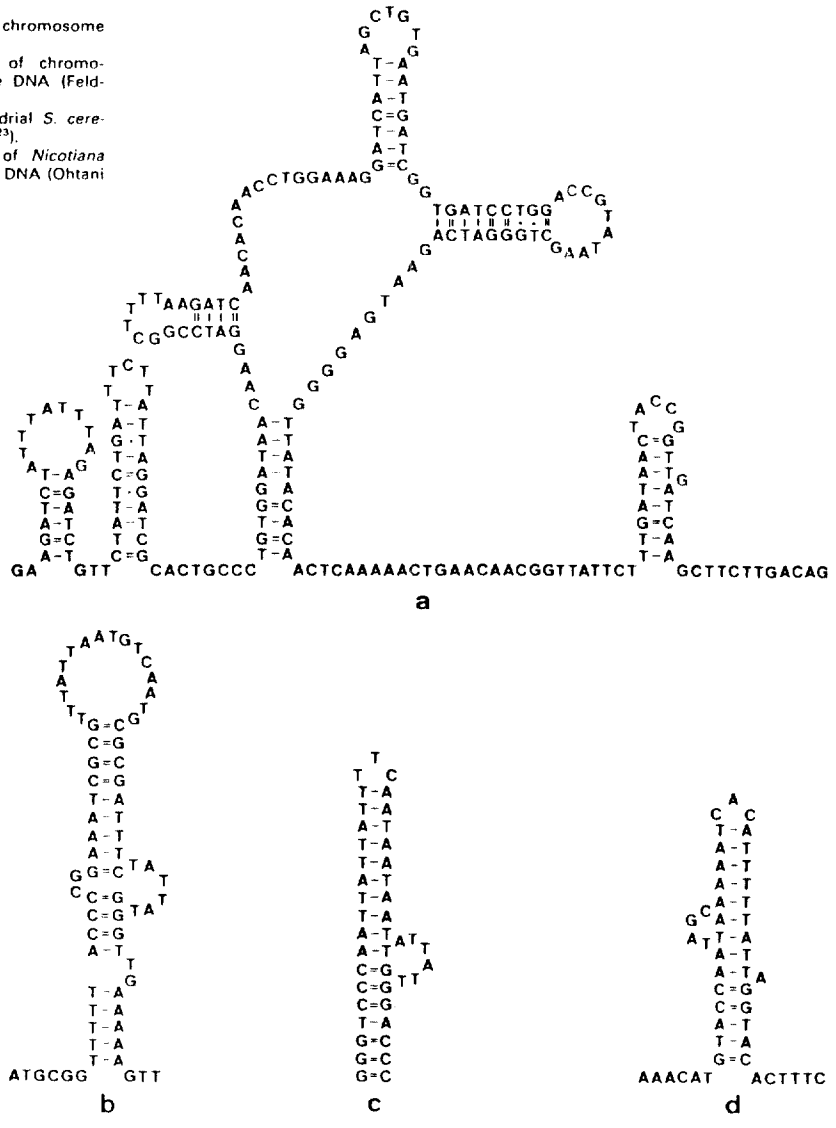


Figure 5: Examples of DNA secondary structures (stem loops) within prokaryotic and eukaryotic origins of replication.

(a): GATCT.ttTATtTAgaGATCTTgTTcTATTgtGATCTcTTATTAGGATCGcc.tgttcTGTGGATAA.tcggatccg.tatttAaGATCAa..g.tTgg.aagGATCacta.cTGTGA—ATGATcggTGATCctggtccGtATaagCTGgGATCacAAtga.gggTTATaCACAgctCaAAAa.c..acaacgGTT–TtcTTtGGATAACTccGgttgatCcaagctttt.agCa.agTTATCCAT
(b): attaatg.ttgta.GACATAaat..tac— up to 60 bp —aaatatttt.....caaacCCCcc.CCCc— up to 30 bp —GcCAAACCCCAAAACCCcAAAAaCa

(a) Consensus sequence of the minimal origin of the entero-bacteriae *E. coli, Salmonella typhimurium, Enterobacter aerogenes, Klebsiella pneumonia, Erwinia carotovora* and the related *Vibrio harveyi* (Zyskind *et al.*[35]).
(b) Consensus sequence near a mitochondrial origin of replication of mammals (human, *Xenopus laevis*, bovine, mouse, rat) (Wong *et al.*[29]).

Capital letters mean nucleotides are found in all sequences. Small letters mean the presence of nucleotides in at least 50% of the compared sequences. The sign '.' means that none of the four nucleotides is mainly present in the sequence. To maximise sequence alignment, gaps (–) are introduced.

Figure 6: Examples of consensus sequences within the origin of replication of various taxonomic groups.

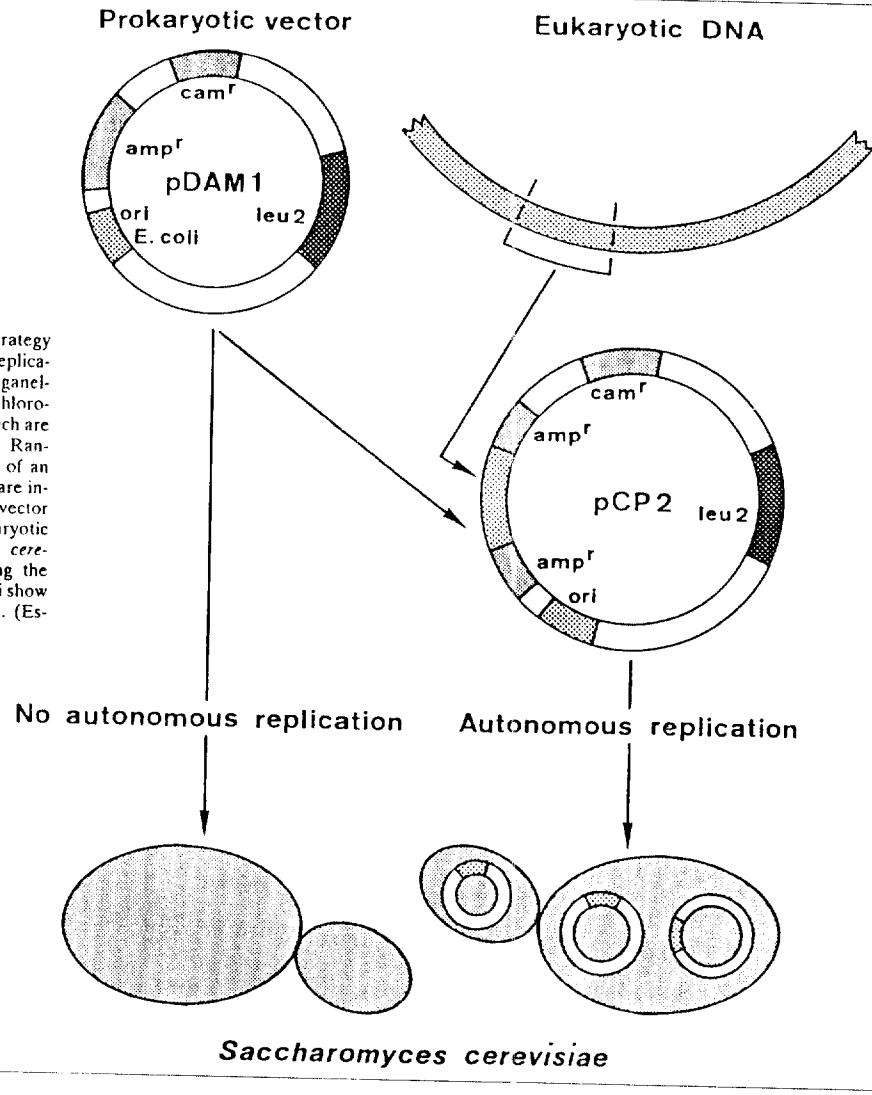


Figure 7: Generalized strategy for screening of origins of replication from either nuclear, organellar (mitochondrial or chloroplast) or plasmid DNA which are functional in *S. cerevisiae*. Random restriction fragments of an appropriate DNA species are inserted into a selectable vector which possesses no eukaryotic ori. After transfer into *S. cerevisiae* only cells containing the vector with a eukaryotic ori show a high transformation rate. (Esser *et al.*[43]).

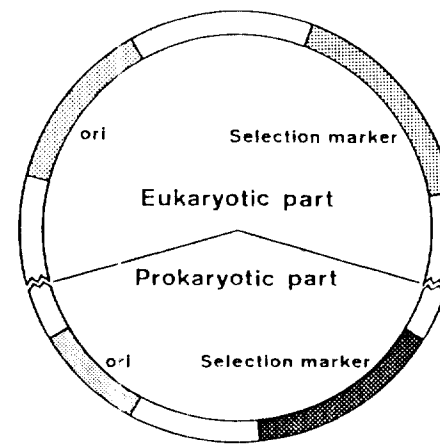


Figure 8: Generalised scheme of a vector for eukaryotic cloning. It consists of eukaryotic and prokaryotic parts represented by different hatching. The latter part is optimal but enables the vector to be efficiently propagated and selected in a host of *E. coli*. By analogy to the prokaryotic part the eukaryotic part is composed of three functional domains:

(1) a selective marker (genes for an appropriate antibiotic resistance or for complementation in an essential biosynthetic pathway).
(2) a sequence inducing autonomous replication which serves as a start signal for the DNA polymerase of the host and
(3) a region for the insertion of genes to be transferred. It is advantageous if host-specific transcriptional and translational signals are present in this third region. (Esser *et al.*[43])

Implications for Biotechnology

At present there are strong efforts to include eukaryotes into genetic engineering not only with respect to clone eukaryotic genes in bacteria but much more to use eukaryotes as host. The latter point has a strong implication for biotechnology because many secondary metabolites (*e.g.*, antibiotics) are produced with eukaryotic filamentous fungi. As briefly mentioned previously there are striking structural differences between prokaryotic and eukaryotic oris. These are also reflected in their function, in that in general a prokaryotic ori does not function in a eukaryotic cell.[39] There are other reasons for establishing eukaryotic cloning systems:

(1): *E. coli*, as the classical host until now, produces toxic substances which may interfere with cloned gene products.

(2): It is known that in the non-toxic *Bacillus subtilis* transformants are rather unstable.

(3): For filamentous fungi, *e.g.*, *Cephalosporium acremonium*, optimized production conditions are already known.

These and some other reasons (for details see Esser and Meinhardt,[40] Esser *et al.*[41]) have promoted the idea of establishing eukaryotic host vector systems.

There are two main attempts along these lines: a very specific one and a more general one. The first attempt uses the yeast *S. cerevisiae*, an organism also very well accessible to classical genetics and well adapted for biotechnological production. It was found that the vectors, such as *pDAM1*[42] (see upper part of Figure 7), consisting mainly of *E. coli* DNA (ori, two marker genes well suited to serve as cloning sites) and having as eukaryotic marker the leu$^+$ gene (prototrophy for the amino acid leucin) was only able to replicate autonomously in yeast if replication origin of yeast DNA was inserted (homologous cloning). In a second attempt this system was extended to heterologous cloning in that via shot-gun experiments eukaryotic oris from DNA sources of other organisms such as organellar DNA were selected (see compilations in Refs. 43, 44, 45, 46).

In Figure 7 this is exemplified for the *pCP2* vector having as eukaryotic origin a segment of mitochondrial DNA of a filamentous fungus, *Cephalosporium*.[47] Because of the presence of the eukaryotic origin the newly constructed vector was not only able to replicate but was also able to express in the yeast host. Subsequent experiments showed that in using this strategy it is possible to screen for stable vectors suitable for heterologous cloning.[48]

However, there is one handicap which needs to be mentioned. During the extension of heterologous cloning in yeast, it became evident that a DNA segment which had the function of an ori in yeast did not necessarily have this function in its organism of descent or in other organisms.[49-51] This relative capacity of originating replication has led to the agreement to use the expression 'putative origin of replication' unless the general capacity of an ori is established. In the yeast literature the term 'ars' (autonomous replicating sequence) is used instead of putative ori. Again another example for the great amount of empirism still involved in molecular cloning yet.

It is not surprising that mitochondrial DNA and also chloroplast DNA may be successfully used for eukaryotic cloning because according to current theories they both are homologous to eubacteria and cyanobacteria respectively.[52] In following these lines it should also now be possible to use organellar DNA for homologous gene cloning for each organism. This certainly has the advantage that there is, *a priori*, no incompatibility with the host DNA if transformation is undertaken with vectors, the functional unit of which originates from the host itself. Furthermore, as briefly indicated above, the advantage of established cultivation and production procedures of an eukaryote may be fully exploited.

For practical purposes it is advisable to perform molecular cloning in eukaryotes with hybrid vectors having a small bacterial part in order to be rapidly replicated in a prokaryote in order to gain enough material for its eukaryotic application. The structure of a vector of this kind is diagrammed in Figure 8. Thus the 'vector of the future' is a universal pro- and eukaryotic vector being able to serve both groups if the host's own DNA is provided as origin of replication.

In Conclusion

Despite the fact that molecular cloning has not only gained great popularity and that it is at present widely applied in both fundamental and industrial research, in contrast to Mendelian genetics, there is a tremendous lack of molecular details, especially in eukaryotes. From this follows that experimental results are not at all predictable, but rather still depend to a wide extent on the empirical 'trial and error system'.

References

[1] Jacob, F., and Brenner, S. *C. R. Acad. Sci. (Paris)*, 1963, **256**, 298.
[2] Walmsley, R.L. *et al. Nature*, 1984, **310**, 157.
[3] Esser, K., and Lang-Hinrichs, C. *Adv. Biochem. Engineering/Biotech.*, 1983, **26**, 143.
[4] Pühler, A., and Timmis, K.N. (Eds). *'Advances in Molecular Genetics'*, 1984. (Berlin: Springer Verlag).
[5] Knippers, R. *'Molekulare Genetik'*, 1983, 3rd Edition. (Stuttgart: Thieme).
[6] Kaudewitz, F. *'Genetik'*, 1983. (Stuttgart: Ulmer).
[7] Fincham, J.R.S. *'Genetics'*, 1983. (London: R.S.G. Wright).
[8] Winnacker, E.-L. *'Gene und Klone'*, 1984. (Weinheim: Springer Verlag).
[9] Kornberg, A. *'DNA Replication'*, 1980. (San Francisco: Freeman).
[10] Herzog, P. *BIUZ*, 1981, **11**, 144.
[11] Kornberg, A. *'Supplement to DNA Replication' (Ref 9), 1982. (San Francisco: Freeman).*
[12] Lewin, B. *'Genes'*, 1983. (New York: John Wiley & Sons).
[13] de Pamphilis, M.L., and Wasserman, P.M. *Ann. Rev. Biochem.*, 1980, **49**, 627.
[14] Nossal, N.G. *Ann. Rev. Biochem.*, 1983, **52**, 581.
[15] Kornberg, A. *TIBS*, 1984, **100**, 122.
[16] Kolter, R., and Helinski, D.R. *Ann. Rev. Genet.*, 1979, **13**, 355.
[17] Ravel-Chapuis, P., *et al. Nature*, 1982, **300**, 78.
[18] Wackernagel, W. *Progress Bot.*, 1980, **42**, 154.
[19] Broach, J.R., *et al.* in *Cold Spring Harbor, Symposium on Quantitative Biology'*, 1982, **47**, 1165.
[20] Pritchard, A.E., and Cummings, D.J. in *'Mechanisms of DNA Replication and Recombination'*, 1983, p. 595. (New York: A.L. Liss).
[21] Buhk, H.-J., and Messer, W. *Gene*, 1983, **24**, 265.
[22] Feldman, H., *et al. Nucl. Acid Res.*, 1981, **9**, 2949.
[23] Bernardi, G. *TIBS*, 1982, **7**, 404.
[24] Ohanti, T., *et al. Molec. Gen. Genet.*, 1984, **195**, 1.
[25] Nagl, W. *Progr. Bot.*, 1979, **41**, 161.
[26] Hartley, J.L., and Danelson, J.E. *Nature*, 1980, **286**, 860.
[27] Lazarus, C.M., and Küntzel, H. *Current Genet.*, 1981, **4**, 99.
[28] Anderson, S., *et al. J. mol. Biol.*, 1982, **156**, 683.
[29] Wong, J.F.L., *et al. Nuc. Acid Res.*, 1983, **11**, 4977.
[30] Tikhomirova, L.P. *et al. Molec. Gen. Biol.*, 1983, **189**, 479.
[31] Osrewacz, H.D. *Thesis*, Ruhr-University, Bochem, 1984.
[32] Kearsey, S. *Cell*, 1984, **37**, 299.
[33] Montiel, J.F., *et al. Nuc. Acid Res.*, 1984, **12**, 1049.
[34] Rawlins, D.R., *et al. Cell*, 1984, **37**, 309.
[35] Zyskind, J.W., *et al. PNAS*, 1983, **80**, 1164.
[36] Baldacci, G., and Bernardi, G. *'The Replication of Yeast Mitochondrial DNA'* in *'Mitochondria'*, 1983. (Berlin: Walter de Gruyter).
[37] Osinga, K.A., *et al. Nuc. Acid Res.*, 1984, **12**, 1889.

[38] Celniker, S.E., *et al. Mol. Cell Biol.*, 1984, **4**, 2455.
[39] Stinchcomb, D.T. *et al. PNAS*, 1980, **77**, 4559.
[40] Esser, K., and Meinhardt, F. *'Genetics of Strain Improvement in Filamentatious Fungi with respect to Biotechnology'* in Vanek, Z., and Hestalch, (Eds). *'Overproduction of Microbial Metabolites'*. In press.
[41] Esser, K., *et al. Current Genet.*, 1983, **7**, 239.
[42] Beach, D., *et al. Nature*, 1980, **284**, 185.
[43] Esser, K. *et al. 'Plasmids of Eukaryotes: Fundamentals and Application'*, 1985. (Berlin: Springer Verlag).
[44] André, D., *et al. Plant Cell Rep.*, 1983, **2**, 175.
[45] Kawamura, M., *et al. Gene*, 1983, **24**, 157.
[46] Rao, T.R., and Reddy, C.A. *Biochem. Biophys. Res. Commun.*, 1984, **118**, 821.
[47] Tudzynski, P., and Esser, K. *Current Genet.*, 1982, **6**, 153.
[48] Tudzynski, P. *Habilitationschrift*, 1983. (Ruhr-Universität Bochum, Abteilung für Biologie).
[49] Beach, D., and Nurse, P. *Nature*, 1981, **290**, 140.
[50] Das, S., and Hollenburg, C.P. *Current Genet.*, 1982, **6**, 123.
[51] Hirth, P., *et al. PNAS*, 1982, **79**, 7356.
[52] Gray, M.W. *Bioscience*, 1983, **33**, 693.

Received 15 March, 1985.