# METHOD FOR OBTAINING D.N.A. R.N.A. PEPTIDES POLYPEPTIDES OR PROTEINS OF A D.N.A. RECOMBINANT TECHNIQUE

The present invention has as its object a process for obtaining DNA, RNA, peptides, polypeptides or proteins, through use of transformed host cells containing genes capable of expressing these RNAs, peptides, polypeptides, or proteins; that is to say, by utilisation of recombinant DNA technique.

The invention aims in particular at the production of stochastic genes or fragments of stochastic genes in a fashion to permit obtaining simultaneously, after transcription and translation of these genes, a very large number (on the order of at least 10'000) of completely new proteins, in the presence of host cells (bacterial or eucaryotic) containing these genes respectively capable of expressing these proteins, and to carry out therafter a selection or screen among the said clones, in order to determine which of them produce proteins with desired properties, for exemple structural, enzymatic, catalytic, antigenic, pharmacologic, or properties of liganding, and more generally, chemical, biochemical, biological, properties.

The invention also has as its aim procedures to obtain, sequences of DNA or RNA with utilizable properties notably chemical, biochemical, or biological properties.

It is clear, therefore, that the invention is open to a very large number of applications in very many areas of science, industry and medicine.

The process for production of peptides or polypeptides according to the invention is characterized in that one produces simultaneously, in the same medium, genes which

are at least partially composed of synthetic stochastic polynucleotides, that one introduces the genes thus obtained into host cells, that one cultivates simultaneously the independent clones of the transformed host cells containing these genes in such a manner so as to clone the stochastic genes and to obtain the production of the proteins expressed by each of these stochastic genes, that one carries out selection and/or screening of the clones of transformed host cells in a manner to identify those clones producing peptides or polypeptides having at least one desired activity, that one thereafter isolates the clones thus identified and that one cultivates them to produce at least one peptide or polypeptide having the said property.

In a first mode of carrying out this process, stochastic genes are produced by stochastic copolymerization of the four kinds of deoxyphosphonucleotides, A,C,G and T from the two ends of an initially linearized expression vector , followed by formation of cohesive ends in such a fashion as to form a stochastic first strand of DNA constituted by a molecule of expression vector possessing two stochastic sequences whose 3' ends are complementary, followed by the synthesis of the second strand of the stochastic DNA.

In a second mode of carring out this process, stochastic genes are produced by copolymerization of oligonucleotides without cohesive ends, in a manner to form fragments of stochastic DNA, followed by ligation of these fragments to a previously linearized expression vector.

The expression vector can be a plasmid, notably a bacterial plasmid. Excellent results have been obtained using the plasmid pUC8 (see Vieira & Joachim, Gene, 19, (1982), 259-268) as the expression vector.
The expression vector can also be viral DNA or a hybrid of plasmid and viral DNA.

The host cells can be prokaryotic cells such as HB 101 (American Type Culture Collection No. 23724) and C 600 (American Type Culture Collection No. 33694), or eukaryotic cells.

When utilizing the procedure according to the second mode mentioned above, it is possible to utilize oligonucleotides which form a group of palindromic octamers.

Particularly good results are obtained by utilizing the following group of palindromic octamers:

5' GGAATTCC 3'
5' GGTCGACC 3'
5' CAAGCTTG 3'
5' CCATATGG 3'
5' CATCGATG 3'

It is also possible to use oligonucleotides which form a group of palindromic heptamers.

Very good results are obtained utilizing the following group of palindromic heptamers:

5' XTCGCGA 3'
5' XCTGCAG 3'
5' RGGTACC 3'

where X = A,G,C or T, and R = A or T

According to a method to utilize these procedures which is particularly advantageous, one isolates and purifies the transforming DNA of the plasmids from a culture of independent clones of the transformed host cells obtained by following the procedures above, then the purified DNA is cut by at least one restriction enzyme corresponding to specific enzymatic cutting site present in the palindromic octamers or heptamers but absent from the expression vector which was utilized; this cutting is followed by inactivation of the restriction enzyme, then one simultaneously treats the ensemble of linearized stochastic DNA fragments thus obtained with T4 DNA ligase, in such a

manner to create a new ensemble of DNA containig new stochastic sequences,this new ensemble can therefore contain a number of stochastic genes larger than the number of genes in the initial ensemble. One then utilizes this new ensemble of transforming DNA to transform the host cells and clone these genes, and finally utilizes screening and/or selection and isolates the new clones of transformed host cells and finally these are cultivated to produce at least one peptide or polypeptide, for example, a new protein.

The property serving as the criterion for selection of the clones of host cells can be the capacity of the peptides or polypeptides,produced by a given clone, to catalyse a given chemical reaction.

For instance, for the production of several peptides and/or polypeptides, the said property can be the capacity to catalyse a sequence of reactions leading from an initial group of chemical compounds to at least one target compound.

With the aim of producing an ensemble constituted by a plurality of peptides and polypeptides which are reflexively autocatalytic, the said property can be the capacity to catalyse the synthesis of the same ensemble from amino acids and/or oligopeptides in an appropriate milieu.

The said property can also be the capacity to modifiy selectively the biological or chemical properties of a given compound, for example the capacity to selectively modify the catalytic activity of a polypeptide.

The said property can also be the capacity to stimulate, inhibit, or modify at least one biological function of at least one biologically active compound, chosen, for example, among the hormones, neurotransmitters, adhesion

factors, growth factors and specific regulators of DNA replication and/or transcription and/or translation of RNA.

The said property can equally be the capacity of the peptide or polypeptide to bind to a given ligand.

The invention also has as its object the use of the peptide or polypeptide obtained by the process specified above, for the detection and/or the titration of a ligand.

According to a particularly advantageous mode of carrying out the invention, the criterion for selection of the clones of transformed host cells is the capacity of these peptides or polypeptides to simulate or modify the effects of a biologically active molecule, for example,a protein, and screening and/or selection for clones of transformed host cells producing at least one peptide or polypeptide having this property, is carried out by preparing antibodies against the active molecule, then utilizing these antibodies after their purification, to identify the clones containing this peptide or polypeptide, then by cultivating the clones thus identified, separating and purifying the peptide or polypeptide produced by these clones, and finally by submitting the peptide or polypeptide to an in vitro assay to verify that it has the capacity to simulate or modify the effects of the said molecule.

According to another mode of carrying out the process according to the invention, the property serving as the criterion of selection is that of having at least one epitope similar to one of the epitopes of a given antigen.

The invention carries over to obtaining polypeptides by the process specified above and utilizable as chemotherapeutically active substances.

In particular, in the case where the said antigen is EGF, the invention permits obtaining polypeptides usable for chemotherapeutic treatment of epitheliomas.

According to a variant of the procedure , one identifies and isolates the clones of transformed host cells producing peptides or polypeptides having the property desired, by affinity chromatography against antibodies corresponding to a protein expressed by the natural part of the DNA hybrid.

For example, in the case where the natural part of the hybrid DNA contains a gene expressing β-galactosidase, one can advantageously identify and isolates the said clones of transformed host cells by affinity chromatography against anti-β-galactosidase antibodies.
After expression and purification of hybrid peptides or polypeptides, one can separate and isolate their novel parts.

The invention also applies to a use of the process specified above for the preparation of a vaccine; the application is characterized by the fact that antibodies against the pathogenic agent are isolated, for exemple antibodies formed after injection of the pathogenic agent in the body of an animal capable of forming antibodies against this agent, and these antibodies are used to identify the clones producing at least one protein having at least one epitope similar to one of the epitopes of the pathogenic agent, the transformed host cell corresponding to these clones are cultured to produce these proteins, this protein is isolated and purified from the clones of cells , then this protein is used for the production of a vaccine against the pathogenic agent.

For example in order to prepare an anti-HBV vaccine, one can extract and purify at least one capside protein of the HBV virus, inject this protein into an animal capable of forming antibodies against this protein having at least one

epitope similar to one of the epitopes of the HVB virus, then cultivate the clones of transformed host cells corresponding to these clones in a manner to produce this protein, isolate and purify the protein from culture of these clones of cells and utilize the protein for the production af an anti-HBV vaccine.

According to an advantageous mode of carrying out the process according to the invention, the host cells consist in bacteria such as Escherichia coli whose genome contains neither the natural gene expressing β -galactosidase, nor the EBG gene, that is to say, $Z^-$,$EBG^-$ E coli. The transformed cells are cultured in the presence of X gal and the indicator IPTG in the medium , and cells positive for β-galactosidase functions are detected; thereafter, the transforming DNA is transplanted into an appropriate clone of host cells for large scale culture to produce at least one peptide or polypeptide.

The property serving as the criterion for selection of the transformed host cells can also be the capacity of the polypeptides or proteins produced by the culture of these clones to bind to a given compound.

This compound can be in particular chosen advantageousely among peptides, polypeptides and proteins, notably among proteins regulating the transcription activity of DNA.

On the other hand , the said compound can also be chosen among DNA and RNA sequences.

The invention has also as its object those proteins which are obtained in the case where the property serving as criterion of selection of the clones of transformed host cells consist in the capacity of these proteins to bind to regulatory proteins controlling transcription activity of the DNA, or else to DNA and RNA sequences.

The invention has, in addition, as an object the use of a protein which is obtained in the first particular case above mentioned, as a cis regulatory sequence controlling replication or transcription of a neighboring DNA sequence.

On the other hand, the aim of the invention also includes utilization of proteins obtained in the second case mentioned to modify the properties of transcription or replication of a sequence of DNA, in a cel containing the sequence of DNA , and expressing this prote .

The invention has as its object as well a process of production of DNA, characterized by simultaneous production in the same medium, of genes at least partially composed of stochastic synthetic polynucleotides, in that the genes thus obtained are introduced into host cells to produce an ensemble of transformed host cells, in that screening and/or selection on this ensemble is carried out to indentify those host cells containing in their genome stochastic sequences of DNA having at least one desired property, and finally, in that the DNA from the clones of host cells thus identified is isolated.

The invention further has as its object a procedure to produce RNA, characterized by simultaneous production in the same medium, of genes at least partially composed of stochastic synthetic polynucleotides, in that the genes thus obtained are introduced into host cells to produce an ensemble of transformed host cells, in that the host cells so produced are cultivated simultaneously, and screening and/or selection of this ensemble is carried out in a manner to identify those host cells containing stochastic sequences of RNA having at least one desired property, and in that the RNA is isolated from the host cells thus identified.

The said property can be the capacity to bind a given compound, which might be for example a peptide or

polypeptide or protein, or also the capacity to catalyse a given chemical reaction, or the capacity to be a transfer RNA.

Now the process according to the invention will be described in more details, as well as some of its applications, with reference to non limitative embodiments.

First, we shall describe particularly useful procedures to carry out the synthesis of stochastic genes, and the introduction of those genes in bacteria to produce clones of transformed bacteria.

I) Direct synthesis on an expression vector.

a) Linearization of the vector

30 micrograms, that is, approximately $10^{13}$ molecules of the pUC8 expression vector are linearized by incubation for 2 hours at 37°C with 100 units of the PstI restriction enzyme in a volume of 300 1 of the appropriate standard buffer. The linearized vector is treated with phenol-chloroform then precipitated in ethanol, taken up in volume of 30 1 and loaded onto a 0,8% agarose gel in standard TEB buffer. After migration in a field of 3V/cm for three hours, the linearized vector is electro-eluted, precipitated in ethanol, and taken up in 30 1 of water.

b) Stochastic synthesis using the enzyme Terminal Transferase (TdT)

30 ug of the linearized vector are reacted for 2 hours at 37°C with 30 units of TdT in 300 1 of the appropriate buffer, in the presence of 1mM dGTP, 1mM dCTP, 0.3 mM dTTP and 1 mM dATP. The lower concentration of dTTP is chosen in order to reduce the frequency of "stop" codons in the corresponding messenger RNA. A similar result, although somewhat less favorable, can be obtained by utilizing a

lower concentration for dATP than for the other desoxynucleotide triphosphates. The progress of the polymerization on the 3' extremity of the PstI sites is followed by analysis on a gel of aliquotes taken during the course of the reaction.

When the reaction attains or passes a mean value of 300 nucleotides added per 3' extremity, it is stopped and the free nucleotides are separated from the polymer by differential precipitation or by passage over a column containing a molecular sieve such as Biogel P60 (Registered Trade Mark). After concentration by precipitation in ethanol, the polymeres are subjected to a further polymerization with TdT, first in the presence of dATP, then in the presence of dTTP. These last two reactions are separated by a filtration on a gel and are carried out for short intervals (30 seconds to 3 minutes) in order to add sequentially 10-30 A followed by 10-30 T to the 3' ends of the polymers.

c) Synthesis of the second strand of stochastic DNA

Each molecule of vector possesses, at the end of the preceding operation, two stochastic sequences whose 3' ends are complementary. The mixture of polymers is therefore incubated in conditions favoring hybridization of the complementary extremities (150mM NaCl, 10mM Tris-HCl, pH 7.6, 1mM EDTA at 65° for 10 minutes, followed by lowering the temperature to 22°C at a rate of 3 to 4° C per hour . The hybridized polymers are then reacted with 60 units of the large fragment (Klenow) of polymerase 1, in the presence of the four nucleotide triphosphates (200mM) at 4°C for two hours. This step accomplishes the synthesis of the second strand from the 3' ends of the hybrid polymers. The molecules which result from this direct synthesis starting from linearized vector are thereafter utilized to transform competent cells.

d) Transformation of competent clones

100 to 200 ml of competent HB101 of C600 at a concentration of $10^{10}$ cells/ml, are incubated with the stochastic DNA preparation (from above) in the presence of 6mM $CaCl_2$, 6 mM Tris-HCl pH8, 6 mM $MgCl_2$ for 30 minutes at 0°C. A temperature shock of 3 minutes at 37°C is imposed on the mixture, followed by the addition of 400 to 800 ml of NZY culture medium, without antibiotics. The transformed culture is incubated at 37° for 60 minutes, then diluted to 10 litres by addition of NZY medium containing 40 µg/ml of ampicillin. After 3-5 hours of incubation at 37°C, the amplified culture is centrifuged, and the pellet of transformed cells is lyophilysed and stored at -70°C. Such a culture contains $3x10^7$ to $10^8$ independent transformants, each containing a unique stochastic gene inserted into the expression vector.

II) Synthesis of stochastic genes starting from oligonucleotides without cohesive ends.

This procedure is based on the fact that polymerization of judiciously chosen palindromic oligonucleotides permits construction of stochastic genes which have no "stop" codon in any of the six possible reading frames, while at the same time assuring a balanced representation of triplets specifying all amino acids. Further, and to avoid a repetition of sequence motifs in the proteins which result, the oligonucleotides can contain a number of bases which is not a multiple of three. The example which follows describes the use of one of the possible combinations which fulfil these criteria:

a) Choice of a group of octamers

The group of oligonucleotides following:

5' GGAATTCC 3'
5' GGTCGACC 3'
5' CAAGCTTG 3'

5' CCATATGG 3'
5' CATCGATG 3'

is composed of 5 palindromes (thus self complementary sequences) where it is easy to verify that their stochastic polymerization does not generate any "stop" codons, and specifies all the amino acids.

Obviously, one can utilize other group of palindromic octamers which do not generate any "stop" codons and specify all the amino acids found in polypeptides. Clearly, it is also possible to utilize non palindromic groups of octamers, or other oligomers, under the condition that their complements forming double stranded DNA are also used.

b) Assembly of a stochastic gene from a group of octamers.

A mixture containing 5 $\mu g$ each of the oligonucleotides indicated above (previously phosphorylated at the 5' position by a standard procedure) is reacted in a 100 ul volume containing 1 mM ATP, 10% polyethyleneglycol, and 100 units of T4 DNA ligase in the appropriate buffer at 13°C for six hours. This step carries out the stochastic polymerization of the oligomers in the double stranded state and without cohesive ends. The resulting polymers are isolated by passage over a molecular sieve (Biogel P60) recovering those with 20 to 100 oligomers. After concentration, this fraction is again submitted to catalysis or polymerization by T4 DNA ligase under the conditions described above. Thereafter, as described above, those polymers which have assembled at least 100 oligomers are isolated.

c) Preparation of the host plasmid

The pUC8 expression vector is linearized by SmaI enzyme in the appropriate buffer, as described above. The vector

linearized by SMa 1 does not have cohesive ends. Thus the linearized vector is treated by calf intestine alkaline phosphatase(CIP) at a level of one unit per microgram of vector in the appropriate buffer, at 37°C for 30 minutes . The CIP enzyme is thereafter inactivated by two successive extractions with phenol-chloroform. The linearized and dephosphorylated vector is precipitated in ethanol, then redissolved in water at 1 mg/ml.

d) Ligation of stochastic genes to the vector

Equimolar quantities of vector and polymers are mixed and incubated in the presence of 1000 units of T4 DNA ligase, 1 mM ATP, 10% polyethylene glycol, in the appropriate buffer, for 12 hours at 13°C. This step ligates the stochastic polymers in the expression vector and forms double stranded circular molecules which are, therefore, capable of transforming.

Transformation of competent clones.

Transformation of competent clones is carried out in the manner previously described.

III) Assembly of stochastic genes starting from a group of heptamers.

This procedure differs from that just discussed in that it utilizes palindromic heptamers which have variable cohesive ends, in place of stochastic sequences containing a smaller number of identical motifs.

a) Choice of a group of heptamers

It is possible, as an example, to use the following three palindromic heptamers:

5' XTCGCGA 3'

5' XCTGCAG 3'

5' RGGTACC 3'

where X = A,G,C, or T and R=A or T , and where polymerization cannot generate any "stop" codons and forms triplets specifying all the amino acids.
Clearly it is possible to use other groups of heptamers fulfilling these same conditions.

b) Polymerization of a group of heptamers

This polymerization is carried out exactly in the fashion described above for octamers.

c) Elimination of cohesive extremities

The polymers thus obtained have one unpaired base on their two 5' extremities. Thus, it is necessary to add the complementary base to the corresponding 3' extremities. This is carried out as follows: 10 micrograms of the double stranded polymers are reacted with 10 units of the Klenow enzyme, in the presence of the four deoxynucleotide-phosphates (200 mM) in a volume of 100 $\mu$l, at 4°C, for 60 minutes. The enzyme is inactivated by phenol choloroform extraction, and the polymers are cleansed of the residual free nucleotides by differential precipitation. The polymers are then ligated to the host plasmid (previously linearized and dephosphorylated) by following the procedures described above.

It is to be noted that the two last procedures which were described utilize palindromic octamers or heptamers which constitute specific sites of certain restriction enzymes. These sites are absent, for the most part, from the pUC8 expression vector. Thus, it is possible to augment considerably the complexity of an initial preparation of stochastic genes by proceeding in the following way: the plasmid DNA derived from the culture of $10^7$ independent transformants obtained by one of the two last procedures

described above, is isolated. After this DNA is purified, it is partially digested by ClaI restriction enzyme (procedure II) or by the Pst 1 restriction enzyme (procedure III). After inactivation of the enzyme, partially digested DNA is treated with T4 DNA ligase, which has the effect of creating a very large number of new sequences, while conserving the fundamental properties of the initial sequences. This new ensemble of stochastic sequences can then be used to transform competent cells.
In addition, the stochastic genes cloned by procedure II and III can be excised intact from the pUC8 expression vector by utilizing restriction sites belonging to the cloning vector and not represented in the stochastic DNA sequences.

Recombination within the stochastic genes generated by the two procedures just described, which result from the internal homology due to the recurrent molecular motifs, is an important additional method to achieve <u>in vivo</u> mutagenesis of the coding sequences. This results in an augmentation of the number of new genes which can be examined .

Finally, for all the procedures to generate novel synthetic genes, it is possible to use a number of common techniques to modify genes <u>in vivo</u> or <u>in vitro</u>, such as a change of reading frame, inversion of sequences with respect to their promotor, point mutations, or utilization of host cells expressing one or several suppressor tRNAs.

In considering the above description, it is clear that it is possible to construct, in vitro, an extremely large number (for example greater than a billion) different genes, by enzymatic polymerization of nucleotides or of oligonucleotides. This polymerization is carried out in a stochastic manner, as determined by the respective concentrations of the nucleotides or oligonucleotides present in the reaction mixture.

As indicated above, two methods can be utilized to clone such genes (or coding sequences): the polymerization can be carried out directly on a cloning expression vector, which was previously linearized; or it is possible to proceed sequentially to the polymerization then the ligation of the polymers to the expression vector.

In the two cases, the next step is transformation or transfection of competent bacterial cells (or cells in culture). This step constitutes cloning the stochastic genes in living cells where they are indefinitely propagated and expressed.

Clearly, in addition to the procedures which were described above, it is feasible to use all other methods which are appropriate for the synthesis of stochastic sequences. In particular, it is possible to carry out polymerization, by biochemical means, of single stranded oligomers of DNA or RNA obtained by chemical synthesis, then treat these segments of DNA or RNA by established procedures to generate double stranded DNA (cDNA) in order to clone such genes.

Screening or selection of clones of transformed host cells

The further step of the procedure according to the invention consists in examining the transformed or transfected cells by selection or screening, in order to isolate one or several cells whose transforming or transfecting DNA leads to the synthesis of a transcription product (RNA) or translation product (protein) having desired property.These properties can be, for example, enzymatic, functional, or structural.

One of the most important aspects of the process, according to the invention, is that it permits the simultaneous screening or selection of an exploitable product (RNA or protein) and the gene which produces that product. In

addition, the DNA synthetized and cloned as described, can be selected or screened in order to isolate sequences of DNA constituting products in themselves, having exploitable biochemical properties.

We shall now describe, as non-limitating examples, preferred procedures for screening or selection of clones of transformed cells such that the novel proteins are of interest from the point of view of industrial or medical applications.

One of these procedures rests in the idea of producing, or obtaining, polyclonal or monoclonal antibodies, by established techniques, directed against a protein or another type of molecule of biochemical or medical interest, where that molecule is, or has been rendered, immunogenic, and thereafter using these antibodies as probes to identify among the very large number of clones transformed by stochastic genes, those whose protein react with these antibodies. This reaction is a result of a structural homology which exists between the polypeptide synthesized by the stochastic gene and the initial molecule. It is possible in this way to isolate numbers of novel proteins which behave as epitopes or antigenic determinants on the initial molecule. Such novel proteins are liable to simulate, stimulate, modulate, or block the effect of the initial molecule. It will be clear that this means of selection or screening may itself have very many pharmacologic and biochemical applications.Below we describe, as a non limiting example, this first mode of operation in a concrete case :

EGF, (epidermal growth factor) is a small protein present in the blood, whose role is to stimulate the growth of epithelial cells. This effect is obtained by the interaction of EGF with a specific receptor situated in the membrane of epithelial cells. Antibodies directed against EGF are prepared by injecting animals with EGF coupled to

KLH (keyhole limpet hemocyanin) to augment the immunogenicity of the EGF. The anti-EGF antibodies of the immunized animals are purified, for example, by passage over an affinity column, where the ligand is EGF or a synthetic peptide corresponding to a fragment of EGF. The purified anti-EGF antibodies are then used as probes to screen a large number of bacterial clones lysed by cloroform, and on a solid support. The anti-EGF antibodies bind those stochastic peptides or proteins whose epitopes resemble those of the initial antigen. The clones containing such peptides or proteins are shown by autoradiography after incubation of the solid support with radioactive protein A, or after incubation with a radioactive anti-antibody antibody.

These steps identify those clones, each of which contains one protein (and its gene) reacting with the screening antibody. It is feasible to screen among a very large number of colonies of bacterial cells or viral plaques (for example, on the order of 1'000'000) and it is feasible to detect extremely small quantities, on the order of 1 nanogram, of protein product. Thereafter, the identified clones are cultured and the proteins so detected are purified in conventional ways. These proteins are tested in vitro in cultures of epithelial cells to determine if they inhibit, simulate, or modulate the effects of EGF on these cultures. Among the proteins so obtained, some may be utilized for the chemotherapeutic treatment of epitheliomes. The activities of the proteins thus obtained can be improved by mutation of the DNA coding for the proteins, in ways analogous to those described above.A variant of this procedure consists in purifying these stochastic peptides, polypeptides or proteins, which can be used as vaccines or more generally to confer an immunity against a pathogenic agent or to exercise other effects on the immunological system, for example, to create a tolerance or diminish hypersensivity with respect to a given antigen, in particular due to binding of these

peptides, polypeptides or proteins with the antibodies directed against this antigen. It is clear that it is possible to use such peptides, polypeptides or proteins in vitro as well as in vivo.

More precisely, in the ensemble of novel proteins which react with the antibodies against a given antigen X, each has at least one epitope in common with X, thus the ensemble has an ensemble of epitopes in common with X. This permits utilization of the ensemble or sub-ensemble as a vaccine to confer immunity against X. It is, for example, easy to purify one or several of the capsid proteins of the hepatitis B virus. These proteins can then be injected into an animal, for example, a rabbit, and the antibodies corresponding to the initial antigen can be recovered by affinity column purification. These antibodies may be used, as described above, to identify clones producing at least one protein having an epitope resembling at least one of the epitopes of the initial antigen. After purification, these proteins are used as antigens (either alone or in combination) with the aim of conferring protection against hepatitis B. The final production of the vaccine does not require further access to the initial pathogenic agent.

Note that, during the description of the procedures above, a number of means to achieve selection or screening have been described. All these procedure may require the purification of a particular protein from a transformed clone. These protein purifications can be carried out by established procedures and utilize, in particular, the techniques of gel chromatography, by ion exchange, and by affinity chromatography . In addition, the proteins generated by the stochastic genes can have been cloned in the form of hybrid proteins having, for example, a sequence of the $\beta$ - galactosidase enzyme which permits affinity chromatography against anti-$\beta$-galactosidase antibodies, and allows the subsequent cleavage of the hybrid part (that is to say, allowing separation of the novel part and the

bacterial part of the hybrid protein.Below we describe the principles and procedures for selection of peptides or polypeptides and the corresponding genes , according to a second method of screening or selection based on the detection of the capacity of these peptides or polypeptides to catalyse a specific reaction.

As a concrete and non limiting example, screening or selection in the particular case of proteins capable of catalysing the cleavage of lactose, normally a function fulfilled by enzyme β -galactosidase ( β-gal) will be described.

As above described, the first step of the process consists in generating a very large ensemble of expression vectors, each expressing a distinct novel protein. To be concrete, for example, one may choose the pUC8 expression vector with cloning of stochastic sequences of DNA in the PstI restriction site. The plasmids thus obtained are then introduced in a clone of E coli from whose genome the natural gene for β -galactosidase, Z, and a second gene EBG, unrelated to the first but able to mutate towards β -gal function, have both been eliminated by known genetic methods. Such host cells ($Z^-$, $EBG^-$) are not able by themselves to catalyse lactose hydrolysis, and as a consequence to use lactose as carbon source for growth. This permits utilization of such host clones for screening or selection for β -gal function.

A convenient biological assay to analyse transformed E coli clones for those which have novel genes expressing a β-gal function consists in the culture of bacteria transformed as described in petri dishes containing X-gal in the medium . In this case , all bacterial colonies expressing a β -gal function are visualized as blue colonies. By using such a biological assay, it is possible to detect even weak catalytic activity. The specific activity of characteristic

enzymes ranges from 10 and 10,000 product molecules per second.

Supposing that a protein synthesized by a stochastic gene has a weak specific activity, on the order of one molecule per 100 seconds, it remains possible to detect such catalytic activity. In a petri dish containing X-gal in the medium, and in the presence of the non metabolizable inducer IPTG (isopropyl-D-thiogalactoside) visualization of a blue region requires cleavage of about $10^{10}$ to $10^{11}$ molecules of X-gal par square millimeter. A bacterial colony expressing a weak enzyme and occupying a surface area of 1mm2 has about $10^{7}$ to $10^{8}$ cells. If each cell has only one copy of the weak enzyme , each cell would need to catalyse cleavage of between 10'000 and 100 of x-gal to be detected, which would require between 2.7 and 270 hours. Since under selective conditions it is possible to amplify the number of copies of each plasmid per cell to 5 to 20 copies per cell, or even to 100 or 1000, and because up to 10% of the protein of the cell can be specified by the new gene, the duration needed to detect a colony blue in the case of 100 enzyme molecules of weak activity per cell is on the order of 0.27 to 2.7 hours.

As a consequence of these facts, screening a very large number of independant bacterial colonies, each expressing a different novel gene, and using the capacity to express a β-gal function as the selection criterion, is fully feasible. It is possible to carry out screening of about 2000 colonies in one Petri dish of 10 cm diameter. Thus, about 20 million colonies can be screened on a sheet of X gal agar of 1 square meter.

It is to be noted that bacterial colonies which appear blue on X gal Petri dishes might be false positives due to a mutation in the bacterial genome which confers on it the capacity to metabolize lactose, or for other reasons than those which result from a catalytic activity of the novel

protein expressed by the cells of the colony. Such false positives can be directly eliminated by purifiying the DNA of the expression vector from the positive colony, and retransforming $Z^-$,$EBG^-E$ coli host cells. If the β-gal activity is due to the novel protein coded by the new gene in the expression vector, all those cells transformed by that vector will exhibit β-gal function. In contrast, if the initial blue colony is due to a mutation in the genome of the host cell, it is a rare event and independent of the transformation, thus the number of cells of the new clone of transformed E coli capable of expressing gal function will be small or zero.

The power of mass simultaneous purification of all the expression vectors from all the positive clones (blue) followed by retransformation of naive bacteria should be stressed. Suppose that the aim is to carry out a screening to select proteins having a catalytic function, and that the probability that a new peptide or polypeptide carries out this function at least weakly is $10^{-6}$, while the probability that a clone of the E coli bacterial host undergoes a mutation rendering it capable of carrying out the same function is $10^{-5}$, then it can be calculated that among 20 million transformed bacteria which are screened, 20 positive clones will be attributable to the novel genes in expression vectors which each carries, while 200 positive clones will be the result of genomic mutation. Mass purification of the expression vectors from the total of 220 positive bacterial clones followed by retransformation of naive bacteria with the mixture of these expression vectors will produce a large number of positive clones consisting of all those bacteria transformed with the 20 expression vectors which code for the novel proteins having the desired function, and a very small number of bacterial clones resulting from genomic mutations and containing the 200 expression vectors which are not of interest . A small number of cycles of purification of expression vectors from positive bacterial

colonies, followed by such retransformation, allows the detection of very rare expression vectors truly positive for a desired catalytic activity, despite a high background rate of mutations in the host cells for the same function.

Following screening operations of this type, it is possible to purify the new protein by established techniques. The production of that protein in large quantity is made possible by the fact that identification of the useful protein occurs together with simultaneous identification of the gene coding for the same protein. Consequently, either the same expression vector can be used, or the novel gene can be transplanted into a more appropriated expression vector for its synthesis and isolation in large quantity.

It is feasible to apply this method of screening for any enzymatic function for which an appropriate biological assay exists. For such screenings, it is not necessary that the enzymatic function which is sought be useful to the host cell. It is possible to carry out screenings not only for an enzymatic function but for any other desired property for which it is possible to establish an appropriate biological assay. It is thus feasible to carry out , even in the simple case of β-gal function visualized on an X-gal Petri plate, a screening of on the order of 100 million, or even a billion novel genes for a catalytic activity or any other desired property.

Selection of transformed host cells.

On the other hand, it is possible to use selection techniques for any property, catalytic or otherwise, where the presence or absence of the property can be rendered essential for the survival of the host cells containing the expression vectors which code for the novel genes, or also can be used to select for those viruses coding and expressing the desired novel gene. As a non-limiting, but concrete example, selection for β-galactosidase function

shall be described. An appropriate clone of $Z^-EBG^-$ E. coli is not able to grow on lactose as the sole carbon source. Thus, after carrying out the first step described above, it is possible to culture a very large number of host cells transformed by the expression vectors coding for the novel genes, under selective conditions, either by progressive diminution of other sources of carbon, or utilization of lactose alone from the start. During the course of such selection, in vivo mutagenesis by recombination, or by explicity recovering the expression vectors and mutagenizing their novel genes in vitro by various mutagens, or by any other common technique, permits adaptive improvements in the capacty to fulfil the desired catalytic function. When both selection techniques and convenient bioassay techniques exist at the same time, as in the present case, it is possible to use selection techniques initially to enrich the representation of host bacteria expressing the β-gal function, then carry out a screening on a Petri plate on X-gal medium to establish efficiently which are the positive cells . In the absence of convenient bioassays, application of progessively stricter selection is the easiest route to purify one or a small number of distinct host cells whose expression vectors code for the proteins catalysing the desired reaction.

It is possible to utilize these techniques to find novel proteins having a large variety of structural and functional characteristics beyond the capacity to catalyse a specific reaction. For example, it is possible to carry out a screen or select for novel proteins which bind to cis-regulatory sites on the DNA and thereby block the expression of one of the host cell's functions, or block transcription of the DNA, or stimulate transcription.

For example, in the case of E. Coli, a clone mutant in the repressor of the lactose operon (i-) expresses β-gal function constitutively due to the fact the lactose

operator is not repressed. All cells of this type produce blue clones on Petri plates containing X-gal medium. It is possible to transform such host cells with expression vectors synthesizing novel proteins and carry out a screen on X-gal Petri plates in order to detect those clones which are not blue. Among those, some represent the case where the new protein binds to the lactose operator and represses the synthesis of β-gal. It is then feasible to mass isolate such plasmids, retransform, isolate those clones which do not produce β-gal, and thereafter carry out a detailed verification.

As mentioned above, the process can be utilized in order to create then isolate, not only exploitable proteins, but also RNA and DNA as products in themselves, having exploitable properties. This results from the fact that, on one hand, the procedure consists in creating stochastic sequences of DNA which may interact directly with other cellular or biochemical constituents, and on the other hand, these sequences cloned in expression vectors are transcribed into RNA which are themselves capable of multiple biochemical interactions.

<u>An example of the use of the procedure to create and select for a DNA which is useful in itself.</u>

This example illustrates selection for a useful DNA, and the purification and study of the mechanism of action of regulatory proteins which bind to the DNA. Consider a preparation of the oestradiol receptor, a protein obtained by standard techniques. In the presence of oestradiol, a steroid sexual hormone, the receptor changes conformation and binds tightly to certain specific sequences in the genomic DNA, thus affecting the transcription of genes implicated in sexual differentiation and the control of fertility. By incubating a mixture containing oestradiol, its receptor, and a large number of different stochastic DNA sequences inserted in their vectors, followed by

filtration of the mixture across a nitro cellulose membrane, one has a direct selection for those stochastic DNA sequences binding to the oestrogen-receptor complex, where only those DNAs bound to a protein are retained by the membrane. After washing and elution, the DNA liberated from the membrane is utilized as such to transform bacteria. After culture of the transformed bacteria, the vectors which they contain are again purified and one or several cycles of incubation, filtration and transformation are carried out as described above. These procedures allow the isolation of stochastic sequences of DNA having an elevated affinity for the oestradiol-receptor complex. Such sequences are open to numerous diagnostic and pharmacologic applications, in particular, for developping synthetic oestrogens for the control of fertility and treatment of sterility.

Creation and selection of an RNA useful in itself

Let there be a large number of stochastic DNA sequences, produced as has been described and cloned in an expression vector. It follows that the RNA transcribed from these sequences in the transformed host cells can be useful products themselves. As a non limiting example, it is possible to select a stochastic gene coding for a suppressor transfer RNA (tRNA) by the following procedure : A large number ($\geqslant 10^8$) of stochastic sequences are transformed into competent bacterial hosts carrying a "nonsense" mutation in the arg E. gene. These transformed bacteria are plated on minimal medium without arginine and with the selective antibiotic for that plasmid (ampicillin if the vector is pUC8). Only those transformed bacteria which have become capable of synthesizing arginine will be able to grow. This phenotype can result either from a back mutation of the host genome, or the presence in the cell of a suppressor. It is easy to test each transformed colony to determine if the arg+ phenotype is or is not due to the presence of the stochastic gene in its vector; it suffices to purify the plasmid from this colony and verify that it

confers an Arg+ phenotype on all arg E cells transformed by it.

Selection of proteins capable of catalysing a sequence of reactions

Below we describe another means of selection, open to independent applications, based on the principle of simultaneous and parallel selection of a certain number of novel proteins capable of catalysing a connected sequence of reactions.

The basic idea of this method is the following : given an initial ensemble of chemical compounds considered as building blocks or elements of construction from which it is hoped to synthetize one or several desired chemical compounds by means of a catalysed sequence of chemical reactions, there exists a very large number of reaction routes which can be partially or completely substituted for one another, which are all thermodynamically possible, and which lead from the set of building blocks to the desired target compound(s). Efficient synthesis of a target compound is favored if each step of at least one reaction pathway leading from the building block compounds to the target compound is comprised of reactions each of which is catalysed. On the other hand, it is relatively less important to determine which among the many independent or partially independent reaction pathways is catalysed. In the previous description, we have shown how it is possible to obtain a very large number of host cells each of which expresses a distinct novel protein.

Each of these novel proteins is a candidate to catalyse one or another of the possible reactions, in the set of all the possible reactions leading from the ensemble of building blocks to the target compound. If a sufficiently large number of stochastic proteins is present in a reaction mixture containing the building block compounds, such that

a sufficiently large number of the possible reactions are catalysed, there is a high probability that one connected sequence of reactions leading from the set of building block compound to the target compound will be catalysed by a subset of the novel proteins. It is clear that this procedure can be extended to the catalysis not only of one, but of several target compounds simultaneously.

Based on this principle it is possible to proceed as follows in order to select in parallel a set of novel proteins catalysing a desired sequence of chemical reactions :

1. Specify the desired set of compounds constituting the building blocks, utilizing preferentially a reasonably large number of distinct chemical species in order to increase the number of potential concurrent reactions leading to the desired target compound.

2.Using an appropriate volume of reaction medium, add a very large number of novel stochastic proteins isolated from transformed or transfected cells synthesizing these proteins. Carry out an assay to determine if the target compound is formed. If it is, confirm that this formation requires the presence of the mixture of novel proteins. If so, then the mixture should contain a subset of proteins catalysing one or several reaction pathways leading from the building block set to the target compound. Purify and divide the initial ensemble of clones which synthesize the set of novel stochastic proteins, the subset which is required to catalyse the sequence of reactions leading to the target compound.

More precisely, as a non limiting example, below we describe selection of novel proteins capable of catalysing the synthesis of a specific small peptide, in particular, a pentapeptide, starting from a building block set constituted of smaller peptides and amino acids. All

peptides are constituted by a linear sequence of 20 different types of amino acides, oriented from the amino to the carboxy terminus. Any peptide can be formed in a single step by the terminal condensation of two smaller peptides (or of two amino acids), or by hydrolysis of a larger peptide. A peptide with M residues can thus be formed by M-1 condensation reactions. The number of reactions, R, by which a set of peptides having length 1,2,3,...M residues can be interconverted is larger than the number of possible molecular species, T. This can be expressed as $R/T \cong M - 2$. Thus, starting from a given ensemble of peptides, a very large number of independent or partially independent reaction pathways lead to the synthesis of a specific target peptide. Choose a pentapeptide whose presence can be determined conveniently by some common assay technique for example HPLC (liquid phase high pressure chromatography), or paper chromatography. Formation of a peptide bond requires energy in a dilute aqueous medium, but if the peptides participating in the condensation reactions are adequately concentrated, formation of peptide bonds is thermodynamically favored over hydrolysis and occurs efficiently in the presence of an appropriate enzymatic catalyst, for example pepsin or trypsin, without requiring the presence of ATP or other high energy compounds. Such a reaction mixture of small peptides whose amino acids are marked radioactively to act as tracers with $^{3}H$, $^{14}C$, $^{35}S$, constituting the building block set can be used at sufficiently high concentrations to lead to condensation reactions.

For example, it is feasible to proceed as follows : 15mg of each amino acid and small peptides having 2 to 4 amino acids, chosen to constitute the building block set, are dissolved in a volume of 0,25ml to 1.0ml of a 0,1M pH 7,6 phosphate buffer. A large number of novel proteins, generated and isolated as described above are purified from their bacterial or other host cells. The mixture of these novel proteins is dissolved to a final concentration on the

order of 0,8 to 1,0 mg/ml in the same buffer. 0,25 ml to 0,5 ml of the protein mixture is added to the mixture of building blocks. This is incubated at 25°C to 40°C for 1 to 40 hours. Aliquots of 8 µl are removed at regular intervals, the first is used as a "blank" and taken before addition of the mixture of novel proteins. These aliquots are analysed by chromatography using n-butanol-acetic acid-pyridine-water (30:6:20:24 by volume) as the solvent. The chromatogram is dried and analysed by ninhydrin or autoradiography (with or without intensifying screens). Because the compound constituting the building block set are radioactively marked, the target compound will be radioactive and it will have specific activity high enough to permit detection at the level of 1-10 ng. In place of standard chromatographic analysis, it is possible to use HPLC (high pressure liquid chromatography) which is faster and simpler to carry out. More generally, all the usual analytic procedures can be employed. Consequently it is possible to detect a yield of the target compound of less than one part per million by weight compared to the compounds used as initial building blocks.

If the pentapeptide is formed in the conditions described above, but not when an extract is utilized which is derived from host cells transformed by an expression vector containing no stochastic genes, the formation of the pentapeptide is not the result of bacterial contaminants and thus requires the presence of a subset of the novel proteins in the reaction mixture.

The following step consists in the separation of the particular subset of cells which contain expression vectors with the novel proteins catalysing the sequence of reactions leading to the target pentapeptide. As an example, if the number of reactions forming this sequence is 5, there are about 5 novel proteins which catalyse the necessary reactions. If the clone bank of bacteria containing the expression vectors which code for the novel

genes has a number of distinct novel genes which is on the order of 1'000'000 all these expression vectors are isolated en masse and retransformed into 100 distinct sets of $10^8$ bacteria at a ratio of vectors to bacteria which is sufficiently low that, on average, the number of bacteria in each set which are transformed is about half the number of initial genes, i.e. about 500'000. Thus, the probability that any given one of the 100 sets of bacteria contains the entire set of 5 critical novel proteins is $(1/2)^5 = 1/32$. Among the 100 initial sets of bacteria, about 3 will contain the 5 critical transformants. In each of these sets, the total number of new genes is only 500'000 rather than 1'000'000. By successive repetitions the total number of which is about 20 in the present case, this procedure isolates the 5 critical novel genes. Following this, mutagenesis and selection on this set of 5 stochastic genes allows improvement of the necessary catalytic functions. In a case where it is necessary to catalyse a sequence of 20 reactions and 20 genes coding novel proteins need to be isolated in parallel, it suffices to adjust the multiplicity of transformation such that each set of $10^8$ bacteria receives 80% of the $10^6$ stochastic genes, and to use 200 such sets of bacteria. The probability that all 20 novel proteins are found in one set is $0.8^{20} \cong 0{,}015$. Thus, about 2 among the 200 sets will have the 20 novel genes which are needed to catalyse the formation of the target compound. The number of cycles required for isolation of the 20 novel genes is on the order of 30.

The principles and procedures described above generalize from the case of peptides to numerous areas of chemistry in which chemical reactions take place in aqueous medium, in temperature, pH, and concentration conditions which permit general enzymatic function. In each case it is necessary to make use of an assay method to detect the formation of the desired target compound(s). It is also necessary to chose a sufficiently large number of building block compounds to

augment the number of reaction sequences which lead to the target compound.

The concrete example which was given for the synthesis of a target pentapeptide can also be generalized as follows:

The procedure as described, generates among other products, stochastic peptides and proteins. These peptides or proteins can act, catalytically or in other ways, on other compounds. They can equally constitute the substrates on which they act. Thus, it is possible to select ( or screen) for the capacity of such stochastic peptides or proteins to interact among themselves and thereby modify the conformation, the structure or the function of some among them. Similary, it is possible to select (or screen) for the capacity of these peptides and proteins to catalyse among themselves, hydrolysis, condensation, transpeptidation or other reactions modifying the peptides. For example, the hydrolysis of a given stochastic peptide by at least one member of the set of stochastic peptides and proteins can be followed and measured by radioactive marking of the given protein followed by an incubation with a mixture of the stochastic proteins in the presence of ions such as Mg,Ca,Zn, Fe and ATP or GTP. The apperance of radioactive fragments of the marked protein is then measured as described. The stochastic protein(s) which catalyse this reaction can again be isolated, along with the gene(s) producing them, by sequential diminution of the library of transformed clones, as described above.

An extension of the procedure consists in the selection of an ensemble of stochastic peptides and polypeptides capable of catalysing a set of reactions leading from the initial building blocks (amino acids and small peptides) to some of the peptides or polypeptides of the set. It is therefore also possible to select an ensemble capable of catalysing its own synthesis; such a reflexively autocatalytic set can be established in a chemostat where the products of the

reactions are constantly diluted, but where the concentration of the building blocks is maintained constant. Alternatively, synthesis of such a set is aided by enclosing the complex set of peptides in liposomes by standard techniques. In a hypertonic aqueous environment surrounding such liposomes, condensation reaction forming larger peptides lowers the osmotic pressure inside the liposomes, drives water molecules produced by the condensation reactions out of the liposomes, hence favors synthesis of larger polymers. Existence of such an autocatalytic ensemble can be verified by two dimensional gel electrophoresis and by HPLC, showing the synthesis of a stable distribution of peptides and polypeptides. The appropriate reaction volume depends on the number of molecular species used, and the concentrations necessary to favor the formation of peptide bonds over their hydrolysis. The distribution of molecular species of an autocatalytic ensemble is free to vary or change due to the emergence of variant autocatalytic ensembles. The peptides and polypeptides which constitute an autocatalytic set may have certain elements in common with the large initial ensemble (constituted of coded peptides and polypeptides as given by our procedure) but can also contain peptides and polypeptides which are not coded by the ensemble of stochastic genes coding for the initial ensemble.

The set of stochastic genes whose products are necessary to establish such an autocatalytic set can be isolated as has been described, by sequential diminution of the library of transformed clones. In addition, an autocatalytic set can contain coded peptides initially coded by the stochastic genes and synthesized continuously in the autocatalytic set. To isolate this coded subset of peptides and proteins, the autocatytic set can be used to obtain, through immunization in an animal, polyclonal sera recognizing a very large number of the constituents of the autocatalytic set.

These sera can be utilized to screen the library of stochastic genes to find those genes expressing proteins able to combine with the antibodies present in the sera.

This set of stochastic genes expresses a large number of coded stochastic proteins which persist in the autocatalytic set. The remainder of the coded constituents of such an autocatalytic set can be isolated by serial diminution of the library of stochastic genes, from which the subset detected by immunological methods has first been subtracted.

Such autocatalytic sets of peptides and proteins, obtained as noted, may find a number of practical applications.

* C L M S *

## CLAIMS

1. Process for the production of peptides or polypeptides by microbiological means, characterized in that genes which are at least partially composed of stochastic synthetic polynucleotides are produced simultaneously in a common milieu, that the genes thus obtained are introduced into host cells, that the independent clones of the transformed host cells containing these genes are simultaneously cultivated so as to clone the stochastic genes and lead to the production of proteins expressed by each of these stochastic genes, that screening and/or selection is carried out on such clones of transformed host cells to identify those clones producing peptides or polypeptides having at least one specified property, that the clones so identified are isolated, then grown in a manner so as to produce at least one peptide or polypeptide having the said property.

2. Process according to claim 1, characterized by the fact that the genes are produced by stochastic copolymerization of the four types of deoxyphosphonucleotides A, C, G, and T, starting from the two extremities of an expression vector which was previously linearized, then by formation of cohesive extremities to create a first strand of stochastic DNA constituted of a molecule of expression vector possessing two stochastic sequences whose 3' extremities are complementary, followed by synthesis of the second strand of the stochastic DNA.

3. Process according to claim 1, characterized by the fact that the genes are produced by stochastic copolymerization of double stranded oligonucleotides which do not have cohesive ends, in a manner so as to form fragments of stochastic DNA, followed by ligation of these fragments in an expression vector which was previously linearized.

* C L M S *

4. Process according to claim 2 or 3, characterized by the fact that the expression vector is a plasmid.

5. Process according to claim 4, characterized by the fact that the expression vector is pUC8.

6. Process according to claim 2 or claim 3, characterized by the fact that expression vector is a fragment of viral DNA.

7. Process according to claim 2 or claim 3, characterized by the fact that the expression vector is a hybrid of plasmid and viral DNA.

8. Process according to claim 1 to 6, characterized by the fact that the host cells are prokaryotic cells.

9. Process according to claims 1 to 7 characterized by the fact that the host cells are eukaryotic cells.

10. Process according to claim 8, characterized by the fact that the cells are chosen among HB101 and C600.

11. Process according to claim 3, characterized by the fact that the oligonucleotides form a group of palindromic octamers.

12. Process according to claim 11 characterized by the fact that the group of palindromic octamers is the following group :

5' GGAATTCC 3'
5' GGTCGACC 3'
5' CAAGCTTG 3'
5' CCATATGG 3'
5' CATCGATG 3'

13.Process according to claim 3, characterized by the fact that the oligonucleotides form a group of palindromic heptamers.

14. Process according to claim 13, characterized by the fact that the group of palindromic heptamers is the following group :

5' XTCGCGA 3'
5' XCTGCAG 3'
5' RGGTACC 3'

where X = A, G, C, or T, and R = A or T.

15.Process according to claim 4 and one of the claim 12 or 14, characterized by the fact that one first isolates and purifies the transforming DNA derived from a culture of independent clones of transformed host cells obtained by proceeding in the manner specified in claim 11 or in claim 13, then that one cuts the purified DNA by means of at least one restriction enzyme which coresponds to a specific restriction site present in these palindromic octamers or heptamers, but absent from the expression vector being utilized, that one thereafter simultaneously treats the ensemble of linearized stochastic DNA fragments so obtained by T4 DNA ligase in such a manner as to create a new ensemble of DNA containing new stochastic sequences, and that one uses this new ensemble of transforming DNA to transform host cells and clone such genes, and finally that one carries out screening and/or selection and isolates the new clones of transformed cells and that one grows these so as to produce at least one peptide or polypeptide having a desired property.

16.Process according to claim 1, characterized by the fact that the said property is the capacity to catalyse a given chemical reaction.

* C L M S *

17.Process according to claim 1 for the production of several peptides and/or polypeptides, characterized by the fact that the said property is the capacity to catalyse a sequence of reactions leading from a given group of initial chemical compounds to at least one target compound.

18.Process according to claim 16 for the production of an ensemble consisting of more than one peptide an/or polypeptide which is reflexively autocatalytic, characterized by the fact that the said property is the capacity to catalyse the synthesis of the ensemble itself starting from amino acids and/or oligopeptides.

19.Process according to claim 1, characterized by the fact that the said property is the capacity to modify selectively the chemical and/or biological properties of a given compound.

20.Process according to claim 19, characterized by the fact that the said property is the capacity to modify selectively the catalytic activity of a polypeptide.

21.Process according to claim 19, characterized by the fact that the said property is the capacity to simulate or modify at least one biological function of at least one biologically active compound.

22.Process according to claim 21, characterized by the fact that the said biologically active compound is chosen among the hormones, neurotransmitters, adhesion or growth factors, and specific regulators of replication and/or transcription of DNA, and/or translation of RNA.

23.Process according to claim 1, characterized by the fact that the said property is the capacity to bind to a given ligand.

24.Utilization of a peptide or polypeptide obtained by the process according to claim 23 for the detection and/or titration of the ligand.

25.Process according to claim 1, characterized by the fact that the property is to have at least one epitope similar to one of the epitopes of a given antigen.

26.Process according to claim 19 and 25, characterized by the fact that the said property is the capacity to simulate or modify the effects of a biologically active molecule, and that screening and/or selection of the clones of transformed host cells producing at least one peptide or polypeptide having this property is carried out by preparing antibodies against that molecule, and utilizing these antibodies so obtained to identify those clones containing those peptides or polypeptides, then by growing the clones thus identified and separating and purifying the peptide or polypeptide produced by these clones, and finally by submitting these peptide(s) or polypeptide(s) to an assay in vitro to verify that it has in fact the capacity to simulate or modify the effects of the said molecule.

27.Peptides or polypeptides obtained by the process according to claim 1 or claim 26, utilizable as active substances having a pharmacologic and/or chemotherapeutic action.

28.Peptides or polypeptides, obtained by the process according to claim 25, utilizable to diminish, in vitro or in vivo, the concentration of free antibodies, specific for the said antigen, by formation of bonds between these paptides or polypeptides and these antibodies.

29.Peptides or polypeptides, according to claims 27 or 28, utilizable as agents to suppress immunological hypersensitivity.

30.Peptides or polypeptides, obtained by the process according to claim25, utilizable as agents to create tolerance with respect to the said antigen.

31.Process according to claim 25, characterized by the fact that the antigen is EGF.

32.Peptides or polypeptides obtained by the process according to claim 31, utilizable for the chemotherapeutic treatment of epitheliomas.

33.Process according to claim 1, characterized by the fact that clones of transformed host cells producing peptides or polypeptides having the desired property are identified and isolated by affinity chromatography on antibodies corresponding to a protein expressed by the natural fragment of the DNA hybrid.

34.Process according to claim 33, characterized by the fact that the natural fragment of the DNA hybrid contains a gene expressing β -galactosidase, and that one identifies and isolates the said clones of transformed cell hosts by affinity chromatography with anti-β-galactosidase antibodies.

35.Process according to claim 1, or claim 34, characterized by the fact that after expression and purification of the hybrid peptides or polypeptides, the novel fragments are separated and isolated.

36.Application of the process according to claim 25 or claim 26, for the preparation of a vaccine, characterized by the fact that antibodies against a pathogenic agent are obtained and used to identify those clones producing at least one protein having at least one epitope similar to one of the epitopes of the pathogenic agent, that the corresponding clones of transformed host cells are grown in such a manner as to produce this protein, that the protein

* C L M S *

is isolated and purified from the cultures of clones of cells and that this protein is used for the production of a vaccine against the pathogenic agent.

37.Application according to claim 36, for the preparation of an anti-HVB vaccine, characterized by the fact that at least one capsid protein of the HBV virus is extracted and purified, that this protein is injected into the body of an animal capable of forming antibodies against this protein, that these antibodies are recovered and purified, that these antibodies are used to identify those clones producing at least one protein having at least one epitope similar to one of the epitopes of the HBV virus, that the clones of transformed host cells corresponding to these clones are grown in a manner to produce this protein, that this protein is isolated and purified from these cultures of host cells, and that this protein is used for the production of an anti-HBV vaccine.

38.Process according to claim 1, characterized by the fact that the host cells consist in bacteria of Escherichia coli type, whose genome contains neither the natural β-galactosidase gene, nor the EBG gene, that is $Z^-$,$EGB^-$ E coli, and that these transformed host cells are cultured in an X-gal medium also containing the inducer lPTG, that clones which are positive for β-galactosidase function are detected in the culture milieu, that thereafter this DNA is transplanted to a clone of host cells appropriate for industrial production of at least one peptide, polypeptide or protein with β-galactosidase function.

39.Process according to claim 1, characterized by the fact that the said property is the capacity to bind to a given compound.

40.Process according to claim 39, characterized by the fact that the said compound is chosen among the petpides,polypeptides and proteins.

* C L M S *

41.Process according to claim 40, characterized by the fact that the said proteins are proteins regulating the transcription activity or replication of DNA.

42.Process according to claim 39, characterized by the fact that the said compound is chosen among the sequences of DNA and RNA.

43.Proteins obtained by the process according to claim 40 or claim 42.

44.Process for the production of DNA, characterized by the fact that, in the same milieu, genes which are at least partially composed of stochastic synthetic polynucleotides are produced, that the genes so produced are introduced into host cells an a manner to produce an ensemble of transformed host cells, that these are grown so as to produce independent clones of the host cells so produced, that screening and/or selection is carried out on this ensemble to identify those host cells which contain those stochastic sequences of DNA having at least one desired property, and that such DNA is isolated from the identified cultures of the host cells.

45.Process according to claim 44, characterized by the fact that the said property is the capacity to bind to given compound.

46.Process according to claim 45, characterized by the fact that the said compound is chosen among the peptides, polypeptides and proteins.

47.Process according to claim 45, characterized by the fact that the said compound is a compound regulating the transcription activity or the replication of DNA.

48.Process according to claim 47, characterized by the fact that the said compound is a regulatory protein controlling the transcription or replication of DNA.

49.Utilization of a sequence of DNA obtained by the process according to claim 46, or claim 47, as a cis-regulatory sequence of replication or transcription of a neighboring sequence of DNA.

50.Process according to claim 42, characterized by the fact that the proteins obtained have the capacity to modify the transcription activity, the replication, or the stability of DNA.

51.Utilization of a protein obtained by the process according to claim 48, to modify the transcription, replication or stability of a sequence of DNA in a cell containing this sequence of DNA and expressing this protein.

52.Process for the production of RNA, characterized by the fact that, in the same milieu, genes which are at least partially composed of synthetic stochastic polynucleotides are produced simultaneously, that the genes thus obtained are introduced in host cells in a manner to produce an ensemble of transformed host cells, that the independent clones of transformed host cells so produced are grown simultaneously, that a screening and/or selection is carried out on this ensemble in a manner to identify those host cells containing stochastic sequences of RNA having at least one desired property, and that the RNA is isolated from the cultures of host cells so identified.

53.Process according to claim 52, characterized by the fact that the said property is the capacity to bind to a given compound.

* C L M S *

54.Process according to claim 52, characterized by the fact that the said property is the capacity to catalyse a given chemical reaction.

55.Process according to claim 52, characterized by the fact that the said property is to be a transfer RNA.