Site-directed, recombination-mediated mutagenesis of a complex gene locus

Michelle Craig Barton, Merl F. Hoekstra¹ and Beverly M. Emerson*

Regulatory Biology Laboratory, ¹Molecular Biology and Virology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Received September 4, 1990; Revised and Accepted October 29, 1990

ABSTRACT

We have generated a site-specific 17 bp insertion within a 38 kb chick globin gene cluster by employing the recombination abilities of *Saccharomyces cerevisiae*. This gene cluster contains four β -type globin genes which share a high degree of sequence homology. In this procedure, a small fragment of β^A -globin DNA containing a 17 bp insertion is subcloned into a *URA3*-based yeast integrating vector (YIp). This mutated globin subclone is introduced into cells that carry the 38 kb globin cluster clone on a single-copy, circular vector derived from a yeast artificial chromosome (YAC). Insertion of the 17 bp oligomer is achieved by targeted integration of the YIp subclone. The recombinant contains the normal β^A -globin gene, the mutant gene and YIp vector sequences between the two copies. Excision of the vector sequences and one copy of the duplicated globin sequences by homologous recombination is required for cell survival when exposed to the selective agent 5-fluoroorotic acid, which is toxic to *ura*⁺ yeast cells. Depending upon the point of the cross-over, a *ura*⁻ yeast cell bearing either a wild-type globin gene or a 17 bp insertion mutation will result. By restriction mapping and *in vitro* transcription analysis, the β^A -globin gene containing the 17 bp insert has no nonspecific mutations generated during the recombination and selection procedures. Specific mutations of regulatory regions, including protein-DNA binding sites, can be accurately targeted within extensive DNA clones by this method.

INTRODUCTION

Biological and chemical analysis of large, complex regions of the genome requires the development of unique methods and tools. In order to isolate, map and sequence large DNA fragments, cosmid and yeast artificial chromosome (YAC) vectors have been developed (1,2). The study of the biological function and regulation of large, complete genetic units will also benefit from the availability of DNA clones which contain sequence and functional units that more closely resemble the genomic environment of the cell. For example, the discovery

of 'dominant control regions' (DCR) more than 50 kb away from the human β -globin promoter, which are required for the correct tissue-specific expression of this gene (3,4), underscores the contributions of specific, long-range DNA sequences. Our interest in the transcriptional regulation of the chick β -globin gene led us to pursue studies of a large DNA template containing the entire cluster of β -type globin genes in the physical order and spacing found *in vivo*.

We required a simple mutation within a DNA template for *in vitro* transcription studies with tissue-specific protein extracts. Specifically, we needed a means of distinguishing chick β -globin RNA species generated in an *in vitro* reaction from the endogenous globin RNA found in expressing red cell extracts. S₁ analysis (5) of transcripts copied *de novo* from an insertion-mutated template are larger and migrate more slowly under gel electrophoresis than the wild-type globin RNA endogenous to the red blood cell extract (6). This type of simple insertional mutation and marking of the DNA template is readily accomplished by synthetic DNA oligomer insertion into a unique or low abundance restriction enzyme recognition site in simple, small subclones of DNA.

The feasibility of using a large, complex DNA template for *in vitro* transcription studies or *in vivo* functional manipulations is limited by the difficulty of generating appropriate specific mutations in the DNA. Here, we describe and employ a yeast system capable of performing homologous recombination to alter a specific segment of DNA within complex, nontandem repeated DNA sequences (reviewed in ref. 7). Briefly, a small segment of a larger gene region is altered by simple *in vitro* methods and subcloned into an integrating vector (YIp). This altered segment is targeted to specific DNA sequences within the complex genomic clone by the generation of recombinogenic free ends within the homologous subclone. By this method, we have generated a 17 bp insertional mutation within a 38 kb clone of chicken globin DNA in order to study the transcriptional properties of the entire gene cluster. This method can be employed to specifically insert or delete DNA fragments within complex DNA clones limited in size only by the choice of cloning vector. The ability to manipulate large genomic fragments cloned into cosmid or YAC vectors will expand sequence and mapping analyses currently in progress to encompass mechanistic studies of function and regulation.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Globin gene-specific probes

Plasmids which contained portions of the chicken β -type globin cluster were the generous gift of Dr. J. Nickol. Gene-specific probes were inserts of p β IVS2, a Bgl II-Hind III fragment surrounding the second intron of the β -adult (β^A) gene sequences unique to the β^A gene, and p ϵ IVS an EcoR I-Ban I fragment of the ϵ -embryonic second intron specific for the ϵ -gene. Plasmid, pUC18B2H2, containing the entire ρ -embryonic gene was also employed. All DNA hybridization probes were labeled by random-primed incorporation of $^{32}\text{P}(\alpha)$ -dATP (8).

Selection of Cosmid Clones

High molecular weight genomic DNA was prepared from adult Cornish White Rock cockerel blood (9). The genomic cosmid library was constructed in a pWE15 vector as described (ref. 1, Stratagene Cat. #951401, Stratagene Cloning Systems). The average insert size of this library is 35–42 kb. The pWE15 vector allows excision of the genomic insert by Not I digestion.

Three genome equivalents were plated onto nitrocellulose filters (S&S) overlaid on LB agar containing kanamycin (25 $\mu\text{g}/\text{ml}$) and screened with p β IVS2 insert (5). From 15 positives, three classes were identified. Of these three classes, one class contained 38 kb insert clones identified by restriction mapping and Southern blotting as a complete globin cluster (Fig. 1). This clone, pCos9 β , was employed in further analyses.

Yeast Methods and Media

Yeast were grown in YPD (rich) or synthetic, defined medium (10) at 30°C. Standard methods for growth and strain manipulations were employed throughout this work (10).

Vector Construction

A yeast shuttle vector (pYAC-0.5RC) which would accept fragments of DNA flanked by Not I restriction sites was derived from pYAC-RC (gift of G. Nonet and G. Wahl, ref. 11) by deleting a BamH I-Nco I fragment. Nco I digests pYAC-RC at the 5' end of the *URA3* gene, and in combination with a complete BamH I digestion results in the deletion of 70% of the *URA3* gene, one of the TEL sequences and the inter-TEL stuffer fragment. The removal of the *URA3* gene from pYAC-RC by this deletion paves the way for later implementation of uracil selection of integrants targeted to the globin sequences inserted into this circular plasmid. This plasmid contains a Not I, Sac II, Sal I, Mlu I, Cla I and SnaB I cloning region within the SUP4-o intron. Plasmid pYAC-0.5RC contains TRP1, ARS1, CEN4, amp^R and the colE1 origin.

Globin Gene Cluster Yeast plasmid constructions

The 38kb Not I insert of pCos9 β was subcloned into the Not I site of pYAC-0.5RC, and the yeast strain AB1380 (2) was transformed with the crude ligation mix as lyticase-converted spheroplasts by a modified protocol of Burgers and Percival (12). Transformants were selected on regeneration plates lacking tryptophan and containing limiting adenine for color identification of insertion into the SUP4-o gene (2). Trp⁺ transformants in which the ochre suppressor SUP4 was inactivated (red colonies) were colony purified and tested for the loss of co-suppression at other ochre mutations in AB1380 (2).

Replicas of these colonies grown on nitrocellulose filters (S&S) overlying trp⁻ plates were analyzed by yeast colony hybridization (13) with the β -globin specific probe. DNA was

isolated from β -globin positive colonies (14), analyzed by Southern blotting (5) and sequentially hybridized (15) with the β -, ϵ - and ρ -specific probes and the entire Cos9 β insert. Recombinant targeting and further analyses were performed with one of these globin-containing clones, pYAC5 β .

Targeting vector construction

Plasmid YIplac β A1 which carries the *URA3* selectable marker was constructed by insertion of the EcoR I-Hind III fragment of the chick β -globin gene plasmid p β A1(pUCABC/ Δ 1, ref. 16) into the EcoR I and Hind III sites of the polylinker region of the vector YIplac211 (17). The plasmid p β A1 is pUC18 which contains the portion of the chick adult β -globin promoter and protein coding region from -407 to approximately +3800, a Sac I-BamH I fragment derived from the parental plasmid pCA β G1 (18). The plasmid p β A1 has been modified by insertion of a 17 bp synthetic fragment of DNA encoding BamH I and Xba I restriction sites (insert 'A', ref. 6 and 16) at position +46 in the untranslated leader portion of the β -globin gene. The DNA sequence of insert 'A' is GGGGATCCTCTAGAGTC.

Directed integration was accomplished by creation of free DNA ends within the β^A coding sequences at the Bgl II restriction site. Bgl II linearized YIpbA1 was transformed by the lithium acetate method of Ito *et al.* (19) into pYAC5 β -bearing yeast cells. The changes made in this transformation protocol are as follows. 100 ml of a late-log culture of yeast cells were grown and harvested as described. The TE (10mM Tris-HCl, pH 8.0, 1mM EDTA)-washed yeast cells were resuspended in 5 ml of a 0.1 M lithium acetate-TE solution and shaken gently at 30° for 30 minutes. The suspension was pelleted at 1000 $\times$ g for 5 minutes, and the pellet resuspended in 1.0 ml of TE. 100 μl aliquots of this suspension were incubated at 0° with 50 μg sonicated salmon sperm DNA and 5 to 20 μg of the linearized plasmid DNA for 30 minutes. The cells were heat-shocked at 37° for 5 minutes, vortexed briefly and 1 ml of a 40% polyethylene glycol 8000, 0.1 M lithium acetate, TE solution were added to each sample. Cells were briefly pelleted by a 15 second spin in a microfuge. The pellets were resuspended in 200 μl SOS (1M sorbitol, 38% YEPD, 7mM CaCl₂) and spread on selective media plates. YIplac β A1 transformants were selected on media lacking tryptophan and uracil.

Not I Mini-Cosmid Library Construction

Total DNA isolated from pYAC5 β A1 transformed yeast cells (14) was digested with Not I. This DNA was ligated with Xba I/Not I digested sCos1 vector (gift of K. Lewis and G. Evans), as described (20). The ligation mix was packaged with Gigapack II Gold extract (Stratagene), according to manufacturer's instructions. Transfected DH5 α cells were plated on kanamycin (25 $\mu\text{g}/\text{ml}$) plates. The DNA of kanamycin-resistant colonies was prepared by the alkaline-lysis mini-prep procedure (5). The DNA was digested with EcoR I and analyzed on Southern blots with the β -globin specific probe.

In vitro transcription reactions

Reactions containing 0.5 μg of template DNA were incubated as previously described (6) with HeLa and red blood cell nuclear transcription extracts. Chick red blood cell extracts were prepared from erythrocytes isolated at day 11 of embryonic development as described (6). HeLa cells (gift of Dr. Robert Roeder) were processed according to published procedures (21). Transcripts were hybridized to a 297 bp NcoI-PstI fragment of p β A1 which

includes the 17 bp 'A' insert at +46. This fragment was end-labeled at position +83 relative to the correct transcription start site and employed as probe in S_1 nuclease analysis (5).

RESULTS

Strategy

To study the transcriptional regulation of the β -type globin gene cluster *in vitro*, transcription extracts which are tissue-specific and developmental stage-specific were employed (6). However, this analysis requires a means of distinguishing the synthetic globin template transcripts from the endogenous mRNA present in these extracts. This was accomplished in small subclones of chick β -globin genes by insertion of synthetic oligomers into the untranslated leader portion of the gene at a specific restriction enzyme recognition site (6, 16). Large plasmids, such as the 38 kb clone of the entire globin gene cluster which encompasses four genes of relatively high homology, are difficult to rearrange *in vitro* by standard molecular cloning methods. Our strategy for marking the β^A gene at a specific site exploits the homologous recombination properties of *S. cerevisiae* and is diagrammed in Figure 1.

The strategy for generating this mutation is to 1) create a targeted duplication/deletion in the gene of interest, and 2) select for recombinants at the duplication/deletion that excise the insertion and replace wild-type with mutant sequences. Specifically, a yeast integrating vector bearing a small portion of the β^A gene with the desired mutation, flanked by wild-type sequences, was integrated into the homologous β^A gene on a globin cluster clone by means of homologous recombination between the globin sequences. The YIp vector contains a *URA3* marker for selection at the step of targeted integration, as shown in Figure 1A. The site of integration is determined by restriction enzyme generation of free ends within the mutant globin YIp clone (YIplac β A1). These ends are recombinogenic and mediate a precise integration of the mutation-carrying plasmid. The integration event results in a plasmid carrying two copies of the β^A gene promoter and 5' portion of the transcribed region from -407 to +1039, separated by YIp vector sequences.

Recombination and excision of the *URA3* marker is required for cell survival when the transformants are plated upon the 5-fluoroorotic acid (5-FOA) selective agent. *Ura*⁺ cells cannot grow in the presence of 5-FOA but *ura*⁻ cells can (22). This excision relies upon homologous recombination between the sequences of the β^A gene. On average, the ratio of the segregated recombinants (either wild-type β^A gene or mutated) should be roughly 1.2:1 based upon the amount of DNA homology on either side of the insertion mutant (Figure 1B).

Globin cluster cosmid cloning

Our interest in β -globin gene regulation led us to clone the 38 kb globin gene cluster. The restriction map of this clone agrees with previously published chicken β -type globin clones which, in summary, encompass the entire globin cluster region (23-25). The insert of the pWE15-globin gene cluster cosmid clone was excised with the restriction enzyme Not I and cloned into pYAC-0.5RC. Approximately 200 Trp⁺ transformants were screened for globin cluster sequences by colony hybridization (13) with a β -gene specific probe. DNA from 14 positively hybridizing colonies was digested with Not I restriction enzyme and analyzed on Southern blots. Ten samples strongly hybridized to the radiolabeled pCos9b insert (data not shown). Of these ten,

Figure 1. Paradigm for recombination-mediated mutagenesis. A) The 38 kb insert of the β -type globin cluster clone, pCos9b, is diagrammed with the coding regions of the embryonic (ϵ and ϵ) and adult (β^H and β^A) globin genes designated by filled boxes. Not I (N), EcoR I (R) and BamH I (B) sites are indicated with sites located within the vector enclosed by parentheses (Top). The β^A gene region is enlarged below with Bgl II (G) and Hind III (H) sites also shown. Transformants carrying this globin clone were targeted for recombination with Bgl II-digested YIplac β A1, containing a subcloned region of the β^A gene (solid lines and shaded box) which has a 17 bp mutation in the untranslated leader region ($\blacktriangledown$). The free Bgl II ends are recombinogenic and combine by homology between the β -globin sequences. Transformants which integrate the YIplac β A1 construction are *Ura*⁺. B) These *Ura*⁺ transformants integrate the mutated β^A sequences as well as the YIp vector sequences (- - -), resulting in a duplication of the β^A sequences as shown. Cells which excise the *URA3* gene within the YIp vector are selected by plating on 5-FOA. Loss of the *URA3* gene is mediated by recombination between the homologous, duplicated β^A regions as indicated, resulting in *ura*⁻ cells which carry one of two possible forms of the β -globin cluster clone, wild-type (WT) or mutated (β^A insertion). This clone is unchanged outside the targeted region of mutagenesis.

two appeared to carry fragments which resulted from a gene duplication, and one positive clone deleted 10 kb of the globin cluster insert. Further restriction enzyme and Southern blot analyses of one of the remaining class of full-length, positive plasmids revealed the presence of all of the β -type globin genes, ϵ , ϵ , β^H and β^A . This clone, pYAC5b, encompasses the full-length globin cluster insert (diagrammed in Figure 1A) and was employed in all further analyses.

Analysis of intergenic recombination

We wished to determine the frequency of intergenic recombination within the globin cluster clone for several reasons. First, as noted above, a number of DNA rearrangements are observed during ligation and transformation of the globin cluster clone (3 of 10); second, the stability of the four nontandem globin genes with extensive sequence similarity (e.g., 97% sequence identity between the embryonic ϵ - and ϵ -genes and 89% sequence

identity among the embryonic and adult β -globin genes over a 400 bp region) was of great concern; third, frequent deletions were observed during previous λ genomic cloning of the globin genes (23–25); and fourth, a high frequency of rearrangement is associated with serial passage of cosmid clones in general (26).

A single colony of the transformant pYAC5b was grown by serial passage in tryptophan lacking synthetic media for more than 150 generations. An inoculum of this culture was plated for single colonies. DNA was prepared from 20 of these single colonies, restriction enzyme digested and examined on Southern blots. The results of this analysis with a β -gene specific probe reveal that none of the colonies show evidence of recombination among the areas of high homology within the gene cluster which led to the deletion of the β^A gene (Figure 2). Based upon our estimated ability to detect changes present in 1% of the input DNA by Southern blot analysis, we estimate the frequency of recombination within the cluster to be less than 5×10^{-4} per cell. This low degree of intergenic homologous recombination indicates that few problems of this nature should arise during mutagenesis of the globin genes.

Insert mutagenesis targeting

The integrating plasmid carrying the mutated β -globin gene, YIplac β A1, was digested to completion with Bgl II which cleaves within the β^A globin sequences. This linearized plasmid was transformed into pYAC5 β colonies and plated upon tryptophan and uracil lacking media. Colonies which integrate the YIp plasmid will be able to grow in the absence of exogenous tryptophan and uracil due to the presence of plasmid markers *TRP1* from pYAC5 β and *URA3* from YIplac β A1. The efficiency of this targeting step is low, with recovery of approximately 2 trp+/ura+ colonies per μ g of linearized transforming DNA. Screening of the targeted clones by Southern blotting, reveals an average of 70% of these colonies arise from integration of the mutated globin and YIp vector sequences in the expected sequence-directed manner. The remaining number of trp+/ura+ colonies is composed of transformants which delete the β^A -globin sequences completely (10%) or partially (20%) during the targeting process, while maintaining vector sequences required

Figure 2. Intergenic recombination analysis. pYAC5 β transformed yeast cells were grown by serial passage over 150 generations. Isolated DNA from 20 colonies were digested with EcoR I to completion and analyzed, in two separate experiments, by 0.8% agarose gel electrophoresis and Southern blotting with the β^A gene specific probe, lanes 1–20. Equal fractions of genomic yeast DNA were digested for this analysis. Differences in intensities reflect variations in DNA preparation yields, not in dissimilarities in gene copy. All of the colonies analyzed carried the β^A gene, the sample in lane 8 reveals a specific band of the correct size after prolonged exposure (data not shown).

for growth. Correct integration of the YIplac β A1 plasmid will create a new BamH I site due to the presence of the linker mutation within the globin sequences on the plasmid, Figure 1A. A BamH I restriction analysis and Southern blotting with a β^A gene specific probe for one of the targeted transformants is shown in Figure 3. The expected BamH I fragment size for the wild type β^A globin gene of 4.4 kb is observed (lane 1). Integration of YIplac β A1, as diagramed in Figure 1, leads to the generation of two different BamH I fragments which migrate at 3.4 and 5.9 kb and hybridize to the β^A specific probe (Figure 3, lane 2). No targeted transformants of the 24 examined colonies contained incorrectly integrated YIplac β A1 within the globin cluster or into the yeast genome, nor were intragenic recombinations observed.

Transplacement of mutated globin clones

5-FOA was applied as a selective agent for recombinant cells that could grow by virtue of removing the integrated *URA3* gene and vector sequences during the YIplac β A1 targeting (22). The YIplac β A1-targeted pYAC5 β transformant analyzed in Figure 3, lane 2, was streaked over an entire tryptophan-minus plate. The lawn of cells was then replica-plated onto tryptophan-lacking plates which contained 1 mg/ml 5-FOA. Cells that become ura⁻ survive the presence of the 5-FOA. The expected cross-over

Figure 3. Wild-type, targeted and transplaced globin clones. Representative DNA samples from pYAC5 β (wild-type), pYAC5 β /YIplac β A1 (targeted) and pYAC5 β A1 (transplaced) transformants were digested with BamH I, separated on an 0.8% agarose gel and analyzed by a Southern blot. A β^A -specific probe reveals the expected 4.4 kb BamH I fragment of the wild-type clone (lane 1), the additional BamH I fragment due to integration of the mutated β^A sequences with an added BamH I site within the insertion (lane 2), and two examples of the mutated globin clones (lanes 3 and 4). The mutated clones have lost the 5.9 kb BamH I fragment encompassing the YIp vector sequences and the wild-type duplicated regions of the β^A gene, and retained the 3.4 kb fragment resulting from digestion with BamH I within the 17 bp insertion. Equal amounts, by volume, of genomic yeast DNA were digested for this analysis.

products resulting in the loss of *URA3* sequences are presented in Figure 1B. Homologous recombination between identical β -globin sequences will result in excision of the *URA3* gene, YIp vector sequences and either wild-type or mutated β^A globin sequences.

From plating experiments, we observed a frequency of approximately 10^{-5} *ura*⁻ cells resulting from the 5-FOA selection. This value agrees with our Southern blot analysis after serial passage (see above). Thirty colonies from this selection were picked and replica plated to uracil-lacking media to verify the efficiency of the selection process. None of these colonies were capable of growing without exogenous uracil. As diagrammed in Figure 1B, *ura*⁻ mutants which have excised the *URA3* sequences via homologous recombination will either display a BamH I profile of the wild type gene or maintain the BamH I site carried within the inserted DNA linker sequences. Two of the selected colonies' DNA which have maintained the insertion mutation at the desired position in the β^A gene are shown in Figure 3, lanes 3 and 4. This insertion gives rise to a single 3.4 kb BamH I fragment when hybridized to the intron probe (Bgl II-Hind III fragment) specific for the β^A gene. Of the 30 colonies analyzed, 22 hybridized to the β^A probe. Nine of these 22 cells have excised the *URA3* gene to give rise to a mutated β^A gene. The other 13 display the wild type pattern of BamH I digestion (data not shown). This ratio of wild-type and mutated β^A colonies (1.4:1) does not reveal a particular sequence preference for one specific class of *ura*⁻ plasmids following crossover and excision of YIp vector sequences and one copy of the β^A gene. Under the selective pressure of 5-FOA exposure, the remaining eight colonies (27%) have deleted all or most of both β^A genes to become *ura*⁻ while maintaining the TRP1 sequences of the original pYAC-0.5RC vector.

Bacterial rescue of the mutated plasmid

In order to employ the mutated β^A globin plasmid in our transcription studies, we needed to move the vector to a bacterial cell host for large-scale growth and purification of the DNA template. The pYAC-0.5RC vector is a shuttle vector capable of conferring ampicillin-resistant growth to transformed *E. coli* cells. Attempts to transform the DNA by standard methods (5)

into *E. coli*, even commercially prepared cells capable of transformation frequencies of 1×10^9 transformants per mg DNA (MAX efficiency DH5a competent cells, BRL and SURE cells, Stratagene), resulted in no transformants. Electroporation of both DH5a and SURE cells also failed to produce any ampicillin-resistant colonies which carried the globin cluster clone. The large size of the plasmid insert (38 kb) and the low copy *colE1* origin of replication sequences present in the pYAC-0.5RC vector may combine to sufficiently lower the transformation frequency below detectable levels.

In order to rescue the globin cluster clone into a bacterial host, a mini-cosmid library was constructed with DNA from cells that contain the pYAC5 β A1 plasmid. This mini-library was constructed by Xba I and Not I digestion of the sCos1 vector, ligated with Not I-digested pYAC5 β A1 DNA, and then packaged prior to transfection (20). Twenty kanamycin-resistant colonies were identified and their DNA isolated and analyzed on Southern blots (5), as shown in Figure 4. Only one of these colonies does not contain the globin cluster clone by hybridization to the β^A probe. The DNA of this transformant (lane 4) hybridized to yeast chromosomes which were separated by pulsed-field electrophoresis and Southern blotted (data not shown). One of the insertion mutated globin cluster clones, sCos5 β A1, was grown and purified over two cesium chloride gradients (5) for further analysis.

In vitro transcription analysis

Nuclear transcription extracts were prepared from HeLa tissue culture cells and 11-day chick red blood cells as described (6,21). These extracts were employed in an *in vitro* transcription analysis

Figure 5. In vitro transcription of globin clones. Globin plasmid clones that carry the β^A gene (lanes 2–4, 6–8) and minus DNA template controls (lanes 1,5) were transcribed in HeLa (lanes 1–4) and 11-day chick red blood cell (lanes 5–8) nuclear extracts. The p β A1 subclone employed as the source of the mutated globin sequences produces an RNA species that hybridize to an *S*₁ probe containing the 17 bp mutation to protect a 100 bp fragment (lanes 2 and 6, β A1). Transcription of the wild-type globin cluster clone, pCos9 β , results in a 37 bp fragment (lanes 3 and 7). This fragment is indistinguishable from the endogenous globin RNA protected fragment (lanes 5–8, β), which is abundant in the chick red blood cell extract even in the absence of template DNA (lane 5). The mutated globin cluster clone, pCos5 β A1, is transcribed as expected for the precise insertion of the 17 bp fragment at +46 into the globin gene cluster (lanes 4 and 8).

Figure 4. Mini-cosmid library rescue of DNA. A small cosmid library was constructed as described above. Fifteen of the 20 colonies which grew after transfection by these constructions were analyzed by EcoR I digestion, gel electrophoresis and Southern blotting of mini-alkaline prepared DNA (lanes 1–15). Hybridization with the β^A gene-specific probe demonstrates that all of these transformants contained the mutated globin gene cluster clone except for the DNA of lane 4. The DNA from this colony is a Not I fragment of the yeast genome as revealed by hybridization to yeast chromosomes separated on a pulsed-field electrophoresis gel and analyzed by a Southern blot (data not shown).

of three DNA templates. The p β A1 template, a subclone of the β^A globin gene from -407 to +3483 which contains the 17 bp 'A' insert at +46; the pCos9 β template, the globin cluster clone lacking the 17 bp insert; and the sCos5 β A1 construct which contains the 'A' insert at +46 of the β^A globin gene were transcribed under conditions previously described (6).

The template activity of each of the three β -globin clones is shown in Figure 4 in a HeLa nuclear extract (lanes 1-4) and in an 11-day chick red cell nuclear extract (lanes 5-8). S_1 analysis (5) was carried out with a probe which extends from -214 to +83 of the adult β -globin gene and contains the 'A' insert at +46. Hybridization to correctly initiated *in vitro* synthesized RNAs (6,16) results in a 100 bp hybrid protected from S_1 nuclease if the template contained the 'A' insert, and a 37 bp hybrid if no insert is present. Transcripts of the expected sizes were generated from all of the templates including the gene that was mutagenized by homologous recombination as described above. The results with the red cell extract (lanes 5-8) illustrate the necessity of inserting a linker sequence into the globin gene in order to detect *in vitro* synthesized transcripts. The wild type β^A globin template generates an RNA species identical in size to the endogenous RNA present in the chick red blood cell nuclear extract preparation. This endogenous RNA overwhelms any *in vitro* signal from the wild type β^A globin gene (Figure 5, lane 7). The insertion of 17 bp within the globin untranslated leader region results in an *in vitro* transcript that is readily distinguished from the endogenous RNA signal.

DISCUSSION

As the sequence analysis and physical mapping of larger pieces of higher eukaryotic DNA continues to progress rapidly, this collection of clones will become an important tool in the biological studies of gene regulation. In this paper, we describe a method for the generation of specific mutations, that are difficult to introduce by other procedures, in large, complex pieces of DNA. The availability of such mutations will greatly expand the usefulness of these clones in regulatory studies.

We can take advantage of several special characteristics of yeast as the host for these genetic manipulations: (1) the stable propagation of large pieces of DNA in a single copy, centromeric plasmid; (2) the ability to perform homologous recombination in a predictable, directed manner; (3) the well-characterized growth selection and genetic markers available; (4) the ability to isolate rare *ura*⁻ mutations by selection with 5-FOA. Replacement of a chromosomal gene with a copy of the allele mutated *in vitro* has been successfully employed in a number of instances by yeast geneticists (ref. 27 and reviewed in ref. 7 and 22). Our application of the transplacement procedure exploits these properties by targeting the mutation-bearing YIp to our wild-type clone on a single-copy, circular plasmid rather than to the yeast genome.

Methods that are based upon homologous recombination have been employed in phage-phage (28) recombination studies, plasmid-phage (29) and plasmid-cosmid (30) methods of genetic selection of genomic clones and YAC fragmentation studies (31). Vector-vector recombination experiments employ genetic selection to simplify and facilitate procedures which are otherwise highly labor-intensive, difficult, or based upon infrequent mutations. In our case, application of the 5-FOA selective agent led to the recovery of rare *ura*⁻ mutants and resolution of the targeted DNA clones. The original implementation of transplacement by Schere and Davis (27) successfully isolated rare *ura*⁻ cells that carried the desired segregants by serial passage

in uracil-lacing media for multiple generations. This method of spontaneous mutation/transplacement could be employed in the absence of the selective agent 5-FOA with different auxotrophic markers, but results in a much lower recovery of successful transplacements (0.3%, ref. 27). The method of selected (5-FOA) reversion allowed us to recover clones at a high frequency (73%) that had none of the unpredictable changes which could result from repair of nonequivalent crossover events or sequence mismatches.

The ability to direct the integration site by generation of free ends (32) within the cloned insert is a distinct advantage of this system. We have not yet explored, in this particular system, the size restrictions of flanking homologous DNA involved in the targeting procedure. The mutagenesis of the adult β -globin gene presented here employs 963 bases of homology 5' of the recombinogenic free ends (Bgl II restriction site) and 481 bases to the 3' side. In general, *in vivo* generation of site-directed mutations occurs with greater fidelity than systems which rely upon *in vitro* DNA synthesis. Nonspecific mutations are thought to occur as a consequence of primers annealing to untargeted sites or decreased fidelity of the DNA synthesis reaction performed *in vitro* (reviewed in ref. 33).

A stringent test for the possibility of subtle modifications by nonspecific mutation was applied to the β^A globin gene. *In vitro* transcription of this gene is dependent upon the interaction of multiple factors that bind to DNA in the promoter region (6). Mutation of sequence or perturbation of spacing will lead to diminution or abolishment of the *in vitro* transcription signal. S_1 analysis can detect within a single base the start site of transcription based upon the length of the protected probe (5). Our results indicate that the mutated globin clone has no detectable nonspecific changes within the β^A globin 5' region that adversely affect transcription.

In order to avoid the additional digestion and ligation steps required to recover the mutated clone by the mini-cosmid library method described above, we are constructing a YAC vector that contains the two cos sites of the sCos1 vector. We have placed an Mlu I site between the two cos sequences and deleted this site from the polylinker. The rare occurrence of the Mlu I restriction site in the mammalian genome (1/300,000 bp predicted) will allow recovery of higher eukaryotic clones following targeted mutagenesis via cosmid packaging without the likelihood of cutting within the cloned insert. The design of the vector contains polylinker restriction enzyme sites that encompass most of the mammalian DNA infrequent cutters (11), yeast sequences that allow stable propagation and genetic selection, and cos sites for recovery of the clone. This will permit the rapid and simple manipulation of higher eukaryotic DNA fragments in the 30 to 40 kb size range.

The ability to target specific mutations within these large clones will greatly expand the current genomic mapping and sequencing efforts and increase their usefulness in studies of biological function and mechanisms. For example, the study of all of the gene members of the globin locus in the physical order and spacing found *in vivo* is greatly facilitated by a method of site-specific mutagenesis of large clones. By utilizing these mutated constructs, we can analyze hemoglobin gene switching within the same template by *in vitro* transcription with red blood cell developmental stage-specific transcription extracts. The examination of sequences or elements within extensive loci which are not present in smaller subclones or in the proper cis orientation may illuminate the contributions of distal elements, such as the DCR, to globin gene expression.

ACKNOWLEDGEMENTS

We thank Steven Bodovitz for technical assistance. M.C.B. is a recipient of Anna Fuller Fund fellowship #693. M.F.H. is a Lucille P. Markey Scholar in Biomedical Sciences. B.M.E. is a Pew Scholar in the Biomedical Sciences. This work was supported by grants from the Pew Foundation and the National Institutes of Health (GM38760) to B.M.E. and the Lucille P. Markey Charitable Trust to M.F.H.

REFERENCES

1. Evans, G. and Wahl, G. (1987) *Methods in Enz.*, **152**, 604–610.
2. Burke, D.T., Carle, G.F. and Olsen, M.V. (1987) *Science*, **263**, 806–912.
3. Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987) *Cell*, **51**, 975–985.
4. Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G., and Groudine, M. (1987) *Nucl. Acids Res.*, **15**, 10159–10177.
5. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition. Cold Spring Harbor Press, Cold Spring Harbor.
6. Emerson, B. M., Nickol, J.M. and Fong, T.C. (1989) *Cell*, **54**, 1189–1200.
7. Botstein, D. and Fink, G.R. (1988) *Science*, **240**, 1439–1442.
8. Feinberg, A.P. and Vogelstein, B. (1984) *Anal. Bioch.*, **137**, 266–267.
9. DiLella, A.G. and Woo, S.L.C. (1987) *Meth. in Enz.*, **152**, 199–212.
10. Malone, R.E. and Hoekstra, M.F. (1984) *Genetics*, **107**, 33–48.
11. Marchuk, D. and Collins, F.S. (1988) *Nucl. Acids Res.*, **16**, 7743.
12. Burgers, P.M.J. and Percival, K.J. (1987) *Anal. Bioch.*, **163**, 391–397.
13. Tschumper, G. and Carbon, J. (1983) *Gene*, **23**, 221–232.
14. Hoekstra, M.F. and Malone, R.E. (1986) *Mol. Cell. Biol.*, **6**, 3555–3558.
15. Church, G. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
16. Jackson, P.D., Evans, T., Nickol, J.M. and Felsenfeld, G. (1989) *Genes Dev.*, **3**, 1860–1873.
17. Gietz, R.D. and Sugino, A. (1988) *Gene*, **74**, 527–534.
18. Ginder, G.D., Wood, W.I. and Felsenfeld, G. (1979) *J. Biol. Chem.*, **254**, 8099–8102.
19. Ito, H., Fukuda, Y., Murata, K. and Kumura, A. (1983) *J. Bact.*, **153**, 163–168.
20. Evans, G.A., Lewis K. and Rothenberg, B.E. (1989) *Gene*, **79**, 9–20.
21. Dignam, J.D., Martin, P.L., Shastry, B.S. and Roeder, R.G. (1983) *Meth. in Enz.*, **101**, 582–598.
22. Boeke, J.D., Trueheart, J., Natsoulis, G. and Fink, G.R. (1987). *Meth. in Enz.*, **154**, 164–175.
23. Villeponteau, B. and Martinson, H. (1981). *Nucl. Acids Res.*, **9**, 3731–3746.
24. Dolan, M., Sugarman, B.J., Dodgson, J.B., and Engel, J.D. (1981). *Cell*, **24**, 669–677.
25. Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) *Cell*, **17**, 879–887.
26. Ishiura, M., Hazumi, N., Shinagawa, H., Nakata, A. Uchida, T. and Okada, Y. (1990) *J. Gen. Micro.*, **136**, 69–79.
27. Schere, S. and Davis, R.W. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4951–4955.
28. Carroll, D. and Ajioka, R.S. (1980) *Gene*, **10**, 273–281.
29. Seed, B. (1983) *Nucl. Acids Res.*, **11**, 2427–2445.
30. Poustka, A., Rackwitz, H-R., Frischauf, A-M. Hohn, B., and Lehrach, H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4129–4133.
31. Pavan, W.J., Hieter, P. and Reeves, R.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1300–1304.
32. Orr-Weaver, T. and Szostak, J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4417–4421.
33. Botstein, D. and Shortle, D. (1985) *Science*, **229**, 1193–1201.