Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 3612–3615, April 1994
Genetics

Targeted gene duplication and disruption for analyzing quantitative genetic traits in mice

(homologous recombination/double-strand gap repair/gene dosage/angiotensinogen gene/essential hypertension)

O. SMITHIES AND H.-S. KIM

Department of Pathology, University of North Carolina, Chapel Hill, NC 27599-7525

Contributed by O. Smithies, January 7, 1994

ABSTRACT **Experimental analysis of complex quantitative genetic traits, such as essential hypertension, should be greatly facilitated by being able to manipulate the expression of a gene in living animals without altering the nucleotide sequence, chromosomal location, or regulatory elements of the gene. To explore this possibility, we have used targeted gene disruption and duplication to generate mice that are genetically identical [(129 × C57BL6)F_1] except for having one, two, or three functional copies of the gene coding for angiotensinogen. The two-copy animals have two normal copies of the angiotensinogen gene; the one-copy and three-copy animals have one normal copy with the other either disrupted or duplicated by gene targeting. The duplicated pair of genes was generated by a special form of gap-repair gene targeting that tandemly duplicates the whole of a gene together with 5′ and 3′ flanking regions. We find progressively and significantly higher levels of the gene product in the animals having increasing numbers of gene copies: the one-copy animals have steady-state plasma angiotensinogen levels ≈35% of normal ($P < 0.0001$), and the three-copy animals have levels ≈124% of normal ($P < 0.004$). Detailed information about regulatory sequences is not required for this type of experiment; nor is it necessary to have DNA clones or targeting constructs that cover the whole of the target gene. Varying gene copy numbers by targeting consequently offers a promising approach to quantitative genetics.**

Essential hypertension is characterized by elevated blood pressures without ascertained cause. Genetic and environmental factors are significant in its etiology, with the former being predominant (reviewed in ref. 1). Its mode of inheritance is generally agreed to be polygenic, but specific genes have not been definitively identified. Indeed, different combinations of genetic variations may be operating in different individuals. One way of finding genes important in the etiology of a disease is the "candidate gene" approach. A candidate gene is chosen on the basis of nongenetic criteria, such as physiologic relevance, and genetic linkage studies are then carried out to look for cosegregation of polymorphic forms of the candidate gene and the disease. Using this approach, Jeunemaitre *et al.* (2) found evidence of genetic linkage between certain variants of the human angiotensinogen gene (*AGT*) and hypertension. [Angiotensinogen (AGT) is an essential precursor of a potent vasopressive octapeptide, angiotensin II.] The specific variants proved to be associated with an ≈20% higher level of AGT in the plasma, and the authors advanced the hypothesis that genetically determined elevated AGT levels might be predisposing for hypertension.

Our present experiments were initiated as the first step in testing this hypothesis in mice. Specifically, we have generated mice having genetically determined gradations in their steady-state plasma AGT levels, including levels close to those seen by Jeunemaitre *et al.* (2) in their hypertensive patients. The graded levels were achieved by using gene targeting to derive mice that are genetically identical except for having one, two, or three functional copies of the *Agt* gene. Conventional gene targeting was used to produce the one-copy mice. A special form of gap-repair gene targeting, which we describe here, was used to produce the mice with three gene copies.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Cloning the *Agt* Gene. Genomic DNA was isolated from E14TG2a (3) mouse embryonic stem (ES) cells and partially digested with *Mbo* I. Approximately 15-kilobase (kb) fragments prepared from this digest were used to make a library in λ phage Charon 35 (4). Clones containing portions of the *Agt* gene were isolated by screening with a probe made by PCR amplification of exon 2 of the mouse gene (5). An 8-kb *Hin*dIII fragment of DNA (see Fig. 1*A*) extending 5′ of the transcriptional start of the gene was isolated from one of these clones. A 1.8-kb fragment extending 3′ of the transcriptional stop was synthesized by PCR amplification using ES cell genomic DNA as the template with primers designed from the sequence of the mouse *Agt* gene (5). These two fragments were then used to make the gap-repair targeting construct described below.

Cell Culture and Electroporation. Line E14TG2a ES cells (3) were used for targeting and were cultured on feeder cells as described (6). Electroporation (7) in the presence of the linearized gap-repair targeting construct at a concentration of 2–5 nM used a 1-sec discharge from a 200-μF capacitor charged to 300 V. Linearization of the plasmid construct was achieved with the restriction enzyme *Not* I. Selection was with G418 and ganciclovir (a gift from Syntex, Palo Alto, CA) as described (6).

PCR Assay for Recombinants. Potential recombinants were identified by a recombinant fragment assay for gene targeting (8) that is based on PCR. PCR primer 1 is specific to the target locus; it is from a part of intron 3 of the mouse *Agt* gene that is within the 8-kb gap of the targeting construct; its sequence is 5′-ACCACTCACGGAAGCTGCAT-3′. PCR primer 2 is from the neomycin-resistance gene; its sequence is 5′-ACGCGTCACCTTAATATGCG-3′. Cell lysates from individual ES colonies surviving selection with G418 and ganciclovir were prepared (8) and amplified by PCR using primers 1 and 2. Amplification of a 2-kb fragment predicted from the sequences of the *Agt* and neomycin-resistance genes was taken as indicating a potential recombinant.

Southern Blot Analysis. Potentially recombinant colonies were expanded to a level sufficient for Southern blot analysis, which was carried out conventionally with probes specific for exon 2 or 5 of the *Agt* gene.

Abbreviations: AGT, angiotensinogen; *Agt*, mouse angiotensinogen gene; ES, embryonic stem.

Plasma AGT Levels. Plasma AGT levels were determined as angiotensin I equivalents by a radioimmunoassay for angiotensin I scaled down from the manufacturer's protocol (DuPont/NEN). Plasma samples (1 μl) had been incubated for 1 hr at 37°C with 4 μl of a renin-containing kidney extract, prepared by homogenization of 14 adult mouse kidneys in 5 ml of 0.05 M sodium phosphate buffer (pH 7.4) followed by centrifugation for 1 hr at 100,000 × *g* (9).

RESULTS AND DISCUSSION

Targeting Strategy for Gene Duplication. Fig. 1 illustrates with the mouse *Agt* gene how gap-repair targeting with an O-type (insertional) targeting construct can create a complete tandem duplication of a gene. Targeting constructs to be used in this way must include two regions of homology, one from upstream (5′) and a second from downstream (3′) of the target gene locus. To maximize the probability that both copies of the tandemly duplicated gene will be fully functional, the 5′ and 3′ flanking regions of homology should extend far enough so that all known or suspected regulatory sequences can be expected to lie between their respective 5′ and 3′ ends. Spanning sufficient DNA to include all regulatory sequences is facilitated by noting that (*i*) the flanking sequences do not need to contain any coding sequences and (*ii*) they can be separated by a large gap because of the gap repair described below. Duplication is therefore possible with genes that have not been completely cloned or that are not available on a single DNA fragment.

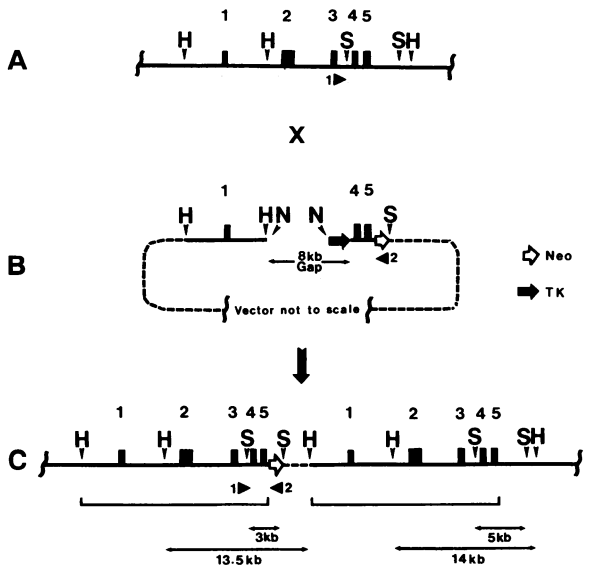

FIG. 1. Duplication of the *Agt* gene by gap-repair gene targeting. (*A*) The structure of the target locus coding for AGT (adapted from ref. 5). The line represents mouse genomic DNA sequences, with the five exons of the gene shown as numbered black rectangles above the line. *Hin*dIII and *Sac* I sites (H and S) used in Southern blot analyses (see Fig. 2) are shown. (*B*) The linearized gap-repair targeting construct used to create the tandem duplication. The two letters N represent a *Not* I site that was used to linearize the construct. An 8-kb gap that is in the targeting construct relative to the target *Agt* gene is shown; it spans exons 2 and 3. Copies of the neomycin-resistance (Neo) and herpes simplex thymidine kinase (TK) genes from pMC1Neo polA and pMC1TK polA (Stratagene) are indicated by open and solid arrows. The interrupted line represents vector plasmid DNA (not to scale in this part of the figure). (*C*) The chromosome resulting from the homologous recombination that mediates the gap-repair gene targeting. The 8-kb gap has been repaired by using chromosomal information, and the *Agt* gene together with 5′ and 3′ flanking regions has been tandemly duplicated. Two horizontal square brackets indicate the duplicated region. Arrowheads 1 and 2 are primers for PCR (also indicated in *A* and *B*). Genomic DNA fragments that should hybridize to exon 2- or 5-specific probes after successful gap-repair targeting are indicated by the double-headed horizontal arrows; their predicted sizes are given in kilobases.

The extent of the duplication in this type of experiment is determined by the distance in genomic DNA between the 5′ end of the upstream fragment of genomic DNA included in the construct and the 3′ end of the downstream region. There is no obvious theoretical upper limit to the length of such duplications, nor need the duplicated region include a gene.

In the present gap-repair construct (Fig. 1*B*) the 5′ region of homology (8 kb) extends about 3 kb 5′ to the start of transcription of the mouse *Agt* gene; the 3′ region (1.8 kb) extends about 200 bp 3′ to the poly(A) addition site and includes a 3′ polypyrimidine tract. A gap of about 8 kb is present in the targeting construct relative to the target *Agt* gene; it spans exons 2 and 3. This gap is repaired (filled in) during the homologous recombination that mediates gene targeting. Such repairs, referred to as double-strand gap repairs, use chromosomal information to supply the missing sequences, as first described in yeast by Orr-Weaver *et al.* (10) and in mammalian gene targeting systems by Valancius and Smithies (11).

The gap in the construct serves three practical functions. First, it allows the size of the construct to be less than the gene that is to be duplicated; for example, in other experiments we have duplicated 25 kb of DNA with a construct having a total length of 15 kb and a gap of over 16 kb. Second, it allows a target-specific primer (primer 1 in Fig. 1) to be chosen from genomic sequences that occur in the gap but not in the incoming DNA (11); this primer is needed for detecting recombinants by PCR (8). Third, it provides a suitable position within the construct for including a copy of the herpes simplex thymidine kinase gene required for negative selection with ganciclovir (12); this negatively selectable gene is lost during the targeting. A positively selectable gene must be included in the construct between the two regions of homology. In the present construct we used the neomycin-resistance gene, which allows positive selection with G418; this positively selectable gene (shown by the open arrow in Fig. 1*C*) is retained as a 1.1-kb insertion between the duplicated genes after the targeting, as is the 3-kb vector plasmid [pBluescript KS(+), Stratagene] shown by a broken line in the figure.

Frequency of Gap-Repair Targeting. The targeting construct, linearized with the restriction enzyme *Not* I, was introduced into E14TG2a ES cells by electroporation. Colonies surviving positive/negative selection (12) with G418 and ganciclovir were assayed by PCR amplification with primers 1 and 2 to identify potential recombinants (8). Previous studies by Valancius and Smithies (11) showed that gaps of up to 2.5 kb in the targeting construct have essentially no negative effects on targeting frequencies compared to those seen with equivalent plasmids having a double-strand break but no gap. Targeting with the construct illustrated in Fig. 1*B* was also achieved at a frequency not significantly different from that usually seen in gene targeting, despite the presence of an 8-kb gap. Thus in two independent electroporations, gap-repair targeting was detected in 1 out of 72 and in 10 out of 130 colonies surviving G418 and ganciclovir selection, as judged by the PCR assay for recombinants.

Southern Blot Analysis. Southern blots were made with DNA prepared from the 11 colonies judged recombinant by the PCR assay. All gave the patterns expected from targeting accompanied by repair of the 8-kb gap and tandem duplication of the *Agt* gene plus its 5′ and 3′ flanking regions.

Fig. 2 illustrates Southern blots of DNA from normal untargeted ES cells (N) and from a gap-repair-targeted colony (T). Digests were with *Sac* I or *Hin*dIII; hybridization was to probes specific for exon 5 or 2, respectively, of the *Agt* gene. Comparison of the sizes of the hybridizing bands with those

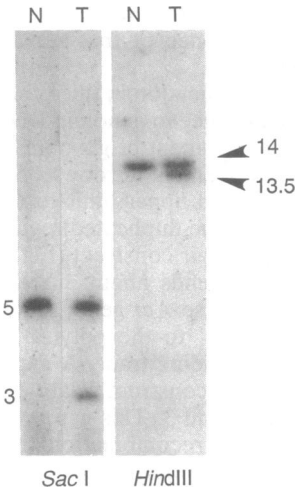

FIG. 2. Southern blot analysis of digests of genomic DNA from normal (lanes N) and gap-repair-targeted (lanes T) ES cell colonies. Digests with *Sac* I or *Hin*dIII were hybridized to probes specific for exon 5 or exon 2, respectively. Fragment sizes in kilobases are given for the bands that demonstrate successful targeting (see Fig. 1 for their predicted sizes).

expected after successful gene duplication shows complete agreement (see Fig. 1*C* for expected sizes). The targeted ES cells have a normal *Agt* gene on one chromosome and a pair of tandemly duplicated genes on its homologue; the relative darknesses of the autoradiographic bands obtained with DNA from the targeted cells can consequently be used to check conclusions drawn from sizes. Thus the 3-kb *Sac* I band and the 13.5-kb *Hin*dIII band, which come *only* from the tandemly duplicated pair of *Agt* genes, should be—and indeed are—approximately half as dark as the 5-kb *Sac* I and 14-kb *Hin*dIII bands, which come from *both* the singleton *Agt* gene and the duplicated pair of genes. These data therefore establish that gap-repair gene targeting was achieved as illustrated in Fig. 1 and created a tandem gene duplication at the target locus.

Generation of Mice. ES cells having the planned tandem duplication of the *Agt* gene were injected into blastocysts in a conventional manner to generate chimeras, and male chimeras were mated with inbred females of strain C57BL6. Pups receiving the ES cell genome were identified by their coat color (agouti). These agouti pups are all genetically identical [(129 × C57BL6)F_1] except for their sexes and their genotypes at the *Agt* locus. As expected, about half (35/86) of them inherited the duplicated *Agt* gene and consequently have three copies of the *Agt* gene (we designate them 2/1, since they have two *Agt* genes on one chromosome and one on its homologue); the remainder are wild type (1/1) and have two copies.

In a second set of experiments, to be described elsewhere, we used conventional gene targeting to disrupt the *Agt* gene, again in E14TG2a ES cells. From these modified cells we generated chimeras, which were also mated with C57BL6 females. The agouti offspring from this second set of experiments are identical [(129 × C57BL6)F_1] to those from the first set of experiments except for their *Agt* genotypes; about half have only one functional copy of the *Agt* gene and are accordingly designated 1/0. We therefore have available males and females that are genetically identical except for having one, two, or three copies of the *Agt* gene.

Plasma AGT Levels Versus Gene Copy Number. Plasma samples were collected from three age-matched males and three age-matched females of the three *Agt* genotypes (1/0, 1/1, and 2/1) and AGT levels were determined. Three to six replicate measurements were made with every sample and were averaged. The average values were then expressed as percentages of the mean plasma AGT level for the six normal (1/1) animals to facilitate pooling of data. Note that these plasma AGT levels are steady-state levels, such as would normally be measured in human subjects; no attempt has been made to factor out the relative contributions to these values of production, utilization, clearance, etc.

The means of the individual percent AGT levels were then calculated for the six animals of each of the three genotypes. Fig. 3 displays these means, together with their standard deviations and the probabilities that the means for the 1/0 and 2/1 animals differ significantly from that of the 1/1 animals. The data show that the 1/0 animals have ≈35% of the level of normal 1/1 animals and that this difference is highly significant ($P < 0.0001$ by the two-sample t test). The 2/1 animals have ≈124% of the level of 1/1 animals, and again the difference is highly significant ($P < 0.004$). The conclusion is therefore clear—an increasing number of functional copies of the *Agt* gene progressively and significantly increases the level of its product. We stress that the increased AGT level in the three-copy animals was achieved without altering the nucleotide sequences of either of the duplicated gene copies or their known regulatory elements and without changing their chromosomal location. We therefore expect that the duplicated *Agt* genes in these animals will have retained their normal patterns of tissue expression and regulation. This can be tested when 2/0 animals (see below) become available.

The extent of the changes in the steady-state plasma AGT level as the gene copy number varies requires comment. The one-copy animals have considerably less (35%) than half the plasma AGT levels of the two-copy animals. The level (124%) in the three-copy animals is also less than the value (150%) expected from strict proportionality to gene dosage. This deviation from proportionality is not of great practical importance at this stage of our investigations, since determining the effects on a downstream quantitative variable (in this case blood pressure) of increasing or decreasing the level of a specific gene product is not dependent on achieving some particular change in the level. In fact, the increase (24%) in

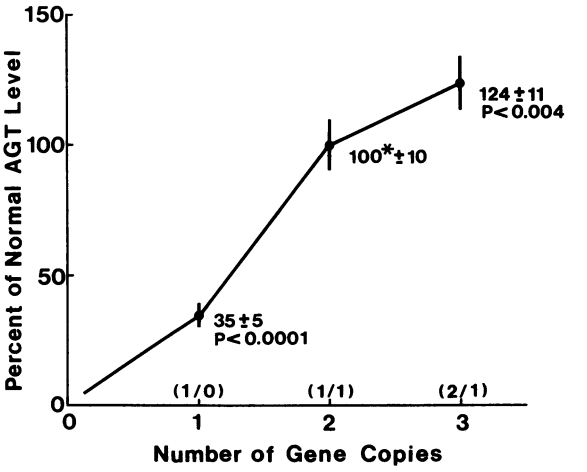

FIG. 3. Plasma AGT levels, presented as percentages of the mean AGT values for normal (1/1) animals, versus the number of functional copies of the *Agt* gene. The points on the graph show the mean percentages, together with their standard deviations (also indicated by error bars), for the six animals of each genotype that were tested. (*, The mean of all the 1/1 animals is assumed to be 100%.) The *Agt* genotypes of the animals, (1/0), (1/1), and (2/1), are shown. The genetic background of all the animals is identical: (129 × C57BL6)F_1. The probabilities, P, (calculated by the two-sample t test) that the means for the 1/0 and the 2/1 animals differ significantly from that of the 1/1 animals are displayed.

plasma AGT levels in the three-copy animals is close to that (20%) observed by Jeunemaitre *et al.* (2) in their hypertensive patients. We thus anticipate being able to use our present animals to test the hypothesis that genetically determined elevations in plasma AGT levels in the range seen in humans are predisposing to hypertension.

Nevertheless, the reasons for deviation from proportionality are of interest. A likely explanation of the relatively reduced plasma AGT levels in the one-copy animals is that the amount of AGT synthesized from one gene is insufficient to maintain a half-normal steady-state plasma level in the face of the prevailing level of utilization and clearance. At least two factors could contribute to the nonlinear increase in AGT in the three-copy animals: utilization may be increased when AGT levels are above normal, and/or the pair of duplicated genes may not function at the same level as two singleton genes. The effectiveness of the duplicated *Agt* genes can be rigorously tested, when the relevant animals become available, by comparing the AGT levels in normal two-copy animals (1/1) with the levels in two-copy animals (designated 2/0) that are genetically identical except for having one chromosome with the *Agt* gene duplicated and one chromosome with the *Agt* gene disrupted.

General Comments. In assessing the relative effects of different genetic manipulations on a phenotype, particularly if the effects are subtle and only quantitative, it is highly advantageous to be able to keep all other genetic variables constant. In our current breeding system, this was achieved readily and rapidly by mating inbred strain 129-derived chimeras with inbred strain C57BL6 animals to obtain genetically uniform and vigorous (129 × C57BL6)F_1 animals that differed only in their *Agt* genotypes. Other breeding schemes can achieve constant genetic backgrounds and an even greater span of gene copy numbers, although at the expense of hybrid vigor. For example, mating strain 129-derived male chimeras to strain 129 females and using their offspring for further breeding allows the production of inbred 129 animals having 0, 1, 2, 3, or 4 copies of the *Agt* gene. The 2/0 animals required for the tests discussed above can be produced by mating strain 129 three-copy (2/1) and one-copy (1/0) animals.

An important consideration when exploring the effects of gene dosage changes on a complex quantitative phenotype is that the changes may have effects that are small or not detectable. Past studies of humans and other animals heterozygous (1/0) for a wide range of null mutations illustrate that frequently the phenotypic effects of halving the number of functional gene copies are subtle—possibly because the gene product is normally in large excess, or because compensatory changes occur in some related system. Indeed, much of the difficulty of deciphering the intricacies of complex quantitative genetic traits stems from such possibilities. Our scheme does not remove these possibilities. Rather it provides genetically uniform animals useful both for accurately assessing their magnitude and for determining their relevance under a variety of conditions. For example, the achieved variations in level of expression may prove to have inconsequential effects under some environmental conditions but be important under others. And study of the compensations induced by a single genetic change is likely to indicate what other genetic systems merit study. The combined effects of changes in the levels of expression of two independent genes can then be studied.

Being able to genetically increase or decrease the expression of a gene, over ranges that are close to physiologic, without deliberately changing the chromosomal location or sequence of the gene and its regulatory elements appears therefore to offer a promising approach to analyzing complex quantitative genetic traits.

We thank Chris Best, Tom Coffman, John Hagaman, Kimberly Kluckman, John Krege, Raju Kucherlapati, and Nobuyo Maeda for helping and advising us. The work was supported by Grants GM20069 and HL49277 from the National Institutes of Health to O.S.

1. Ward, R. (1990) in *Hypertension: Pathophysiology, Diagnosis, and Management*, eds. Laragh, J. H. & Brenner, B. M. (Raven, New York), pp. 81–100.
2. Jeunemaitre, X., Soubrier, F., Kotelevtsev, Y. V., Lifton, R. P., Williams, C. S., Charru, A., Hunt, S. C., Hopkins, P. N., Williams, R. R., Lalouel, J.-M. & Corvol, P. (1992) *Cell* **71**, 169–180.
3. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. (1987) *Nature (London)* **326**, 292–295.
4. Loenen, W. A. M. & Blattner, F. R. (1983) *Gene* **26**, 171–179.
5. Clouston, W. M., Evans, B. A., Haralambidis, J. & Richards, R. I. (1988) *Genomics* **2**, 240–248.
6. Koller, B. H., Kim, H.-S., Latour, A. M., Brigman, K., Boucher, R. C., Jr., Scambler, P., Wainwright, B. & Smithies, O. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10730–10734.
7. Boggs, S. S., Gregg, R. G., Borenstein, N. & Smithies, O. (1986) *Exp. Hematol.* **14**, 988–994.
8. Kim, H.-S. & Smithies, O. (1988) *Nucleic Acids Res.* **16**, 8887–8903.
9. Iwao, H., Nakamura, N., Kim, S., Ikemoto, F., Yamamoto, K. & Schlager, G. (1984) *Jpn. Circ. J.* **48**, 1270–1279.
10. Orr-Weaver, T. L., Szostak, J. W. & Rothstein, R. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6354–6358.
11. Valancius, V. & Smithies, O. (1991) *Mol. Cell. Biol.* **11**, 4389–4397.
12. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. (1988) *Nature (London)* **336**, 348–352.