Volume 10 Number 24 1982 Nucleic Acids Research

Duplication followed by deletion accounts for the structure of an Indian deletion β°-thalassemia gene

Richard A.Spritz* and Stuart H.Orkin+

*Dept. Medical Genetics and Pediatrics, Lab. Genetics, Univ. Wisconsin, Madison, WI 53706, and +Dept. Pediatrics, Harvard Medical School and Div. Hematology and Oncology, The Children's Hospital Medical Center and the Sidney Farber Cancer Inst., Boston, MA 02115, USA

Received 4 October 1982; Accepted 15 November 1982

ABSTRACT

Nucleotide sequence analysis of a cloned deletion β-globin gene from a patient with β^{o}-thalassemia demonstrates a 619 nucleotide deletion extending from the 3' third of the second intervening sequence through 209 bases of 3' flanking DNA. However, an additional novel heptanucleotide was identified between the deletion endpoints, suggesting a complex etiology for this rearrangement.

INTRODUCTION

The β^{o}-thalassemias are an heterogeneous group of autosomal recessive anemias characterized by an absence of synthesis of β-globin chains in the erythroid cells of affected individuals. Although most β-thalassemia alleles do not contain large DNA rearrangements (1-3), a deletion of approximately 600 nucleotides has previously been characterized in some β^{o}-thalassemia alleles from Asian Indian individuals (1,2,4). This deletion includes the terminal portion of the second intervening sequence, the third globin coding region, the 3'-untranslated sequence, and approximately the first 150 bases of 3'-flanking DNA (4). We have now determined the complete nucleotide sequence of one of these deletion β^{o}-thalassemia alleles. This β^{o}-thalassemia gene contains a 619 base deletion with addition of a novel heptanucleotide between the deletion endpoints, suggesting a complex etiology for this rearrangement.

MATERIALS AND METHODS

The molecular cloning of an 8.2 kb Eco RI DNA fragment containing the deletion β-globin gene in λgtWES has been described previously (4). This fragment was isolated from the recombinant bacteriophage and subcloned in pBR322. Nucleotide sequence determination was by a modification of the method of Maxam and Gilbert (5). Restriction endonucleases were purchased from New England Biolabs, and α-^{32}[P]-deoxynucleotide triphosphates and

0305-1048/82/1024-8025$2.00/0

γ-^{32}[P]-adenosine triphosphate were from Amersham.

RESULTS

We determined the complete DNA sequence of the thalassemic β-globin gene from the Rsa I site 127 nucleotides 5′ to the cap site of the normal β-globin gene to the Dde I site ordinarily 331 nucleotides 3′ to the polyadenylation site. Comparison of this sequence with the DNA sequence of the intact human β-globin gene (6-9) showed a deletion of 619 nucleotides which extended from 149 bases upstream from the 3′ splice site of the second intervening sequence to 209 bases 3′ to the polyadenylation site (Fig. 1). Between the endpoints of this deletion a novel heptanucleotide, AAGTAGA, was found.

In addition to the deletion and addition noted above, five single nucleotide differences were detected. Four of these, the C to T transition in the third base position of codon two, the C to G and G to T transversions at the 16th and 74th, and a T to C transition at the 666th bases of the second intervening sequence, have been described in several other normal (10) and thalassemic (10-15) β-globin alleles and thus appear to be common DNA sequence polymorphisms. The fifth, a C to G transversion 311 nucleotides 3′ to the polyadenylation site, has been noted only once before (9) and may represent either another polymorphism or an error in the published normal β-globin gene sequence (6). Another polymorphism previously noted in other alleles, the C to T change at position 81 of the second intervening sequence (10-15), was not present in the deletion β^0-thalassemia gene. Accordingly, this gene has a combination of polymorphisms typical of the Asian subtype of type 3 β-globin alleles as proposed by Orkin and colleagues (10,16) which may represent an intermediate type allele.

Figure 1. Nucleotide sequence in the region of the deletion in the β^0-thalassemic and normal β-globin genes. The deletion endpoints are indicated by vertical bars. The nineteen base overlapping direct repeat sequence flanking the 5′ deletion endpoint is indicated by brackets. The 23 base direct repeat sequence in the region of the 3′ deletion endpoint is indicated by over- and underlining. Polymorphic nucleotides are starred.

DISCUSSION

A simple model accounting for the generation of this thalassemic gene is presented in Figure 2. It involves two events, duplication followed by deletion. The duplication of a dodecanucleotide, either GATTCAAGTAGA or ATTCAAGTAGAG in the 3' flanking region of the β-globin gene, may have resulted from sliding of the coding and non-coding DNA strands relative to each other (17) at an overlapping imperfect 23 base direct repeat sequence (AGNAPuAGGNTTNANNTPuGAGGPyT) which spans the duplicated region (Figure 1). Following the duplication, intra- or intermolecular misalignment and simple recombination at one of two alternative positions appears to have produced the 619 base deletion. The novel heptanucleotide which occurs between the deletion endpoints is thus the remnant of half of the leftward dodecanucleotide duplication described above. Other models might also account for this rearrangement; however, the model presented here appears to be the simplest.

Although the immediate stimuli for recombination events are not known, analysis of other known deletions suggests that misalignment between short regions of DNA sequence homology may be an important feature (17-23). However, examination of the sequence of the normal human β-globin gene in the relevant regions demonstrates segments of only limited DNA sequence similarity, approximately 30 bases in length, spanning each end of the region of the deletion. These segments may have facilitated the

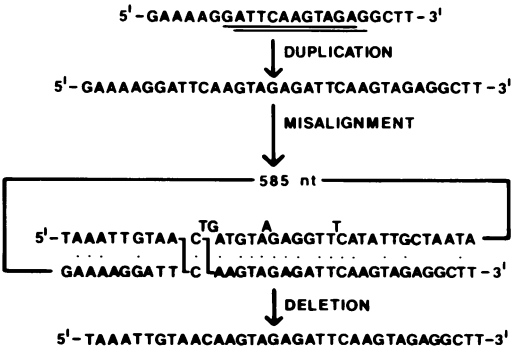

Figure 2. Model for the generation of the deletion β^{o}-thalassemia allele. Although an intramolecular nonhomologous recombinational event is depicted, an intermolecular event is equally possible.

nonhomologous recombination event which produced the deletion. It is intriguing that, prior to the deletion, both deletion endpoints were flanked by direct repeat sequences (Figure 1). The 3' endpoint was flanked by the duplication described above, and the 5' endpoint by an overlapping 19 base direct repeat, (AATAT$^{T}_{C}$TCTGCATATAAAT) normally present in the second intervening sequence of the β-globin gene (6,8). It may be that local strand slippage at these repeats stimulated misalignment and recombination between these sites (24).

Thus, the rearrangement in the 'deletion' β^{o}-thalassemia allele actually appears to be the result of two successive events, a twelve base duplication and a nonhomologous crossover producing a 619 base deletion. This deletion could only have occurred on the background produced by this duplication, therefore these events necessarily proceeded in the order duplication→deletion. Whether these events occurred independently or not cannot be determined at present, and the model presented in Fig. 2 is also consistent with the existence of alleles having this duplication but lacking this deletion. However, this deletion should not be observed in the absence of this duplication. As the deletion can be identified in approximately 20% of Indian β-thalassemia alleles (H. Kazazian, personal communication), a search for β-globin alleles which contain this duplication but lack this deletion should probably focus on the Indian population. Unfortunately, the duplication would not create a novel restriction endonuclease cleavage site, therefore such a search would require direct nucleotide sequence analysis of cloned genes.

ACKNOWLEDGMENTS

This research was supported by grants from the N.I.H. to R.A.S. and S.H.O. and the National Foundation-March of Dimes to S.H.O. In addition, R.A.S. is aided by Basil O'Connor Starter Research Grant No. 5-341 from the March of Dimes Birth Defects Foundation and S.H.O. is the recipient of a Research Career Development Award from the Heart, Lung, and Blood Institute of the N.I.H. We thank H. Kazazian for most helpful discussion and communication of data prior to publication and S. Keller for preparing the manuscript. This is paper number 2589 from the Laboratory of Genetics, University of Wisconsin.

REFERENCES

1. Flavell, R.A., Bernards, R., Kooter, J.M., deBoer, E., Little, P.F.R., Annison, G. and Williamson R. (1979) Nucl. Acids Res. 6, 2749-2760.

2. Orkin, S.H., Old, J.M., Weatherall, D.J. and Nathan, D.G. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 2400-2404.
3. Tam, J.W.O., Kaufman, R.E. and Nienhuis, A.W. (1981) Hemoglobin 5, 209-215.
4. Orkin, S.J., Kolodner, R., Michelson, A. and Husson, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3558-3562.
5. Maxam, A. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
6. Lawn, R.M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980) Cell 21, 647-651.
7. Moschonas, N., deBoer, E., Grosveld, F.G., Dahl, H.H.M., Wright, S., Shewmaker, C.K. and Flavell, R.A. (1981) Nucl. Acids Res. 9, 4391-4401.
8. Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J.T., deRiel, J.K., Manley, J.L., Gefter, M.L., Forget, B.G. and Weissman, S.M. (1981) Proc. Natl. Acad Sci. U.S.A. 78, 2455-2459.
9. Poncz, M., Ballantine, M., Solowiejczyk, D., Barak, I. and Schwartz, E. (1982) J. Biol. Chem. 257, 5994-5996.
10. Orkin, S.H., Kazazian, H.H., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G. and Giardina, P.J.V. (1982) Nature 296, 627-631.
11. Baird, M., Driscoll, C., Schreiner, H., Sciarrata, G., Sansone, G., Niazi, G., Ramirez, F. and Bank, A. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4218-4221.
12. Driscoll, M.C., Baird, M., Bank, A. and Rachmilewitz, E.A. (1981) J. Clin. Invest. 68, 915-919.
13. Orkin, S.H. and Goff, S.C. (1981) J. Biol. Chem. 256, 9782-9784.
14. Antonarakis, S.E., Boehm, C.D., Giardina, P.J.V., and Kazazian, H.H. Jr. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 137-141.
15. Spence, S.E., Pergolizzi, R.G., Donovan-Peluso, M., Kosche, K.A., Dobkin, C.A. and Bank, A. (1982) Nucl. Acids Res. 10, 1283-1294.
16. Antonarakis, S.E., Orkin, S.H., Kazazian, H.H., Goff, S.C., Boehm, C.D., Waber, P.G., Sexton, J.P., Ostrer, H., Fairbanks, V.F. and Chakravarti, A. (1982) Proc. Natl. Acad. Sci. U.S.A., in press.
17. Streisinger, G., Okada, V., Emrich, J., Newton J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.
18. Marotta, C.A., Wilson, J.T., Forget, B.G. and Weissman, S.M. (1977) J. Biol. Chem. 252, 5040-5053.
19. Farabaugh, P.J., Schmeissner, U., Hofer, M. and Miller, J.H. (1978) J. Mol. Biol. 126, 847-863.
20. Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., deRiel, J.K., Forget, B.G., Weissman, S.M., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
21. Orkin, S.H., Goff, S.C. and Hechtman, R.L. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 5041-5045.
22. Schon, E.A., Cleary, M.L., Haynes, J.R. and Lingrel J.B. (1981) Cell 27, 259-269.
23. Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982) Cell 29, 319-328.
24. Radding, C.M. (1978) Ann. Rev. Biochem. 47, 847-880.