Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 9622–9625, September 1994
Genetics

A cluster of mutations within a short triplet repeat in the C1 inhibitor gene

(slipped mispairing/DNA deletion/DNA duplication)

JOHN J. BISSLER*†, MARCO CICARDI‡, VIRGINIA H. DONALDSON*§, PAUL A. GATENBY¶, FRED S. ROSEN‖, ALBERT L. SHEFFER**, AND ALVIN E. DAVIS III*

Departments of *Pediatrics and §Medicine, University of Cincinnati, Children's Hospital Research Foundation, Cincinnati, OH 45229; ‡Clinica Medica, Via Pace 9, Milan, Italy 20122; ¶Department of Clinical Immunology, Royal Prince Alfred Hospital, University of Sydney, Camperdown, NSW 2050, Australia; and ‖Center for Blood Research and **Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

Communicated by Elkan Blout, June 27, 1994

ABSTRACT **Mutations in the C1 inhibitor gene that result in low functional levels of C1 inhibitor protein cause hereditary angioneurotic edema. This disease is characterized by episodic edema leading to considerable morbidity and death. Among 60 unreported kindred with the disease, four patients were discovered to have mutations clustered within a 12-bp segment of exon 5 from nucleotide 8449 to nucleotide 8460. This short segment of DNA contains three direct repeats of the triplet CAA and is immediately preceded by a similar adenosine-rich sequence (CAAGAACAC). These triplet repeats make this region susceptible to mutation by a slipped mispairing mechanism. There are two other short triplet repeat elements in the coding region for this gene, but they have not become mutated in any kindred examined. This suggests that the apparent enhanced mutation rate in this region of exon 5 may be influenced by DNA structural characteristics.**

Hereditary angioneurotic edema is characterized by episodic nonpitting edema of the skin and mucous membranes. Considerable morbidity and mortality result from this edema when it leads to a functional bowel obstruction from intestinal edema or asphyxia from tracheobronchial tree involvement (1). The disease is transmitted as an autosomal-dominant trait; patients express only a single normal allele at the C1 inhibitor locus. The mutant allele may encode a dysfunctional protein or may not be expressed.

There are two well-recognized mechanisms for the mutations within the C1 inhibitor gene. Large deletions and duplications have been attributed to unequal crossing-over between the numerous intronic *Alu* repetitive elements (2). These repetitive elements (3) are densely represented in introns 3, 4, 6, and 7. A second mechanism frequently involves the inhibitor's reactive center codon (CGC) (4). Spontaneous deamination of the methylated cytosine in the CpG dinucleotide results in a transition to thymine. In addition, the CpG dinucleotide appears to lead to other single-base mutations that are not restricted to mutations predicted by methylation and deamination. These point mutations may result from non-Watson–Crick base pairing, such as cis instead of trans base rotation with respect to the sugar moiety, keto–enol or amino–imino base tautomerization, and protonization or ionization (5).

A third mechanism for mutagenesis in the C1 inhibitor gene recently has been proposed (6). DNA polymerase chain elongation rate and fidelity are affected by primary nucleotide sequence and secondary structure. Although the canonical right-handed double helix is the predominant form of DNA in the eukaryotic nucleus, under physiological conditions nonrandom regions of DNA can adopt a number of alternate conformations (7). These alternate secondary structures, such as cruciform structures, are an interruption in the interstrand base pairing by a region of intrastrand base pairing and present an obstacle to high-fidelity replication. This process is referred to as DNA-directed mutagenesis. Within exon eight in the C1 inhibitor gene, there is a cluster of mutations (8) surrounding a sequence with the potential to form a cruciform structure. These mutations may be a direct result of difficult passage of the DNA polymerase through the region during DNA replication. In this paper, we report a second cluster of mutations in exon 5 of the C1 inhibitor gene. These mutations appear to be due to slipped mispairing within a short triplet repeat. This region is unique in its replication infidelity because two other comparable triplet repeats in exons of the C1 inhibitor gene are devoid of mutation in 60 kindred.

METHODS

Patients. Deoxyribonucleic acid from patients with hereditary angioneurotic edema was isolated from peripheral blood leukocytes by a previously reported method (9).

Sequencing. The exon five region of the C1 inhibitor gene was amplified by PCR. The primer pair used in amplification consisted of a 5′ primer corresponding to the adjacent sequence of the fourth intron (TTGGATCCGACTCATGCCTCCCTTTCTCAACATACCCC) and a complementary primer from the adjacent fifth intron (TTGGATCCATGGGAAGACTAACTGGCTAAGGGCACCCT). These primers allowed examination of the intron–exon junctions. PCR was carried out with 0.5 μg of genomic DNA, 25 pmol of each of the primers, 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris (pH 8.0), 0.1% bovine serum albumin, and 0.5 μl of *Taq* polymerase (Perkin–Elmer/Cetus) for 30 cycles (94°C for 1 min, 50°C for 2 min, and 72°C for 3 min). Amplified fragments were purified by isopropyl alcohol precipitation.

For the sequencing reactions, the 5′ oligonucleotide was 5′ end-labeled with [γ-^{32}P]ATP (New England Nuclear) using T4 polynucleotide kinase (Boehringer Mannheim). One picomole was added to a reaction mixture containing buffer, template, deoxynucleotides, and one dideoxynucleotide (BRL). The reaction was thermal cycled in the same fashion as for the initial PCR reaction. The mixtures were denatured in formamide and resolved on an 8 M urea/5% acrylamide (5% *N,N′*-methylenebisacrylamide) gel in 0.09 M Tris/0.09 M boric acid/0.001 M EDTA buffer. The gel was dried and exposed to Kodak XAR-5 film overnight.

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.

†To whom reprint requests should be addressed at: Children's Hospital Research Foundation #5, 3333 Burnet Avenue, Cincinnati, OH 45229-3039.

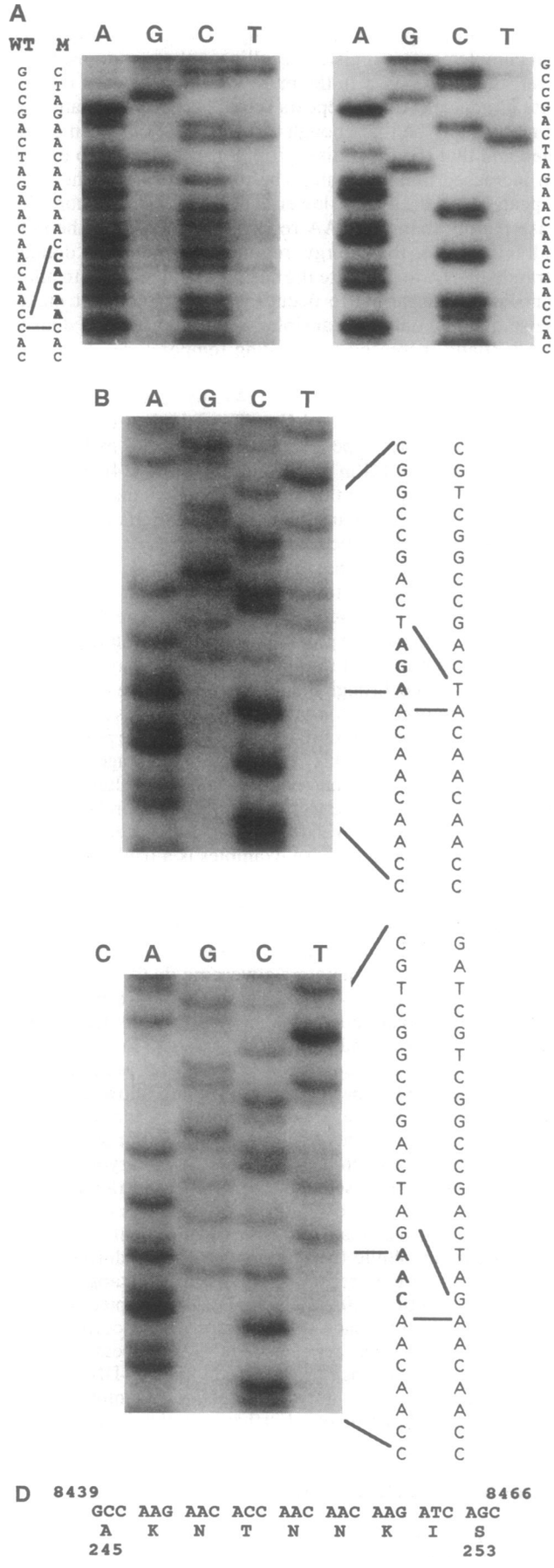

FIG. 1. Duplex DNA sequence from patients. Exon 5 was amplified inclusive of the intron–exon junctions. The reaction products were precipitated with isopropyl alcohol and sequenced by unidirectional PCR in the presence of dideoxynucleotides. Because both normal and mutant alleles are present in the genome, both alleles are present in the PCR product. Sequencing identifies the frameshift. To the side of the sequence are the normal (WT) and mutant (M) allele sequences. The mutant sequence is always to the right of the normal sequence. Boldfaced letters indicate the mutations. (*A*) In Bo, AACAC is boldfaced in the mutant sequence as it is inserted. The normal sequence through this region is shown in *Right*. (*B*) The AGA in the normal sequence is boldfaced in Cr and is deleted in the mutant sequence. (*C*) In Le, the CAA is boldfaced in the normal sequence and is deleted in the mutant. (*D*) Translated sequence from this region of exon 5.

RESULTS

Three unrelated patients, in addition to one previously described patient (10), were found to have mutations within exon 5 of the C1 inhibitor gene. No other mutations were identified in the coding sequence. These mutations were clustered in the segment from nucleotide 8449 to nucleotide 8460 (nucleotides numbered according to genomic DNA; ref. 3). This segment of DNA contains three repeats of the triplet CAA (nucleotides 8449–8457). In addition, the nine nucleotides preceding the repeat are similar (CAAGAACAC). DNA sequence analysis revealed one patient with an insertion (Fig. 1*A*). Patient Bo demonstrated an insertion of five nucleotides (AACAC) between nucleotide 8448 and nucleotide 8449. The frameshift caused by this insertion results in the introduction of a termination codon 75 nucleotides downstream (nucleotides 8523–8525).

Two different deletions were observed in the other patients (Fig. 1 *B* and *C*). One copy of the repeated CAA triplet was deleted in patient Le (Fig. 1*C*). This results in the deletion of one of the two Asn residues at positions 249 and 250. In patient Cr the second and third nucleotides of the codon for Lys-251 and the first nucleotide of the codon for Ile-252 have been deleted (Fig. 1*B*). This results in the deletion of Lys-251, and thus is identical, at the amino acid level, to the previously reported C1 inhibitor Ta, which has a deletion of the codon for Lys-251 (AAG) (10).

DISCUSSION

The patients described here were deficient in C1 inhibitor for a variety of genetic reasons. The insertion in patient Bo leads to a frameshift mutation, probably resulting in a truncated peptide product because of a downstream stop codon. It is not known if any mutant protein is secreted. However, based on knowledge of the three-dimensional structures of protease inhibitors related to C1 inhibitor, it is very unlikely that a stable protein product would result. Patients Ta (10) and Cr have a deletion of Lys-251, although the mutations resulting in the deletion are different. Patient Le has a deletion of the preceding amino acid, Asn-250. These mutations result in variable serum C1 inhibitor antigenic levels. Patients Bo and Le have low antigenic levels (type I hereditary angioneurotic edema) while Cr has normal levels (type II hereditary angioneurotic edema). It would be expected that Bo would have low levels because only one allele encodes a viable inhibitor. Since Cr has the same mutation at the amino acid level as C1 inhibitor Ta, it is not surprising that it too is present in plasma as a dysfunctional protein. Previous analysis of C1 inhibitor Ta shows that this deletion results in glycosylation of Asn-250. These studies suggested that this protein is dysfunctional because of an alteration of a protease contact site due either to the glycosylation or to the amino acid deletion itself. At this time, it is unclear why patient Le has low inhibitor levels considering the relatively small change that is made in its primary amino acid structure. This low level may be due to

alterations in protein folding resulting in decreased stability and altered clearance kinetics. This instability could result in enhanced degradation or in a molecule that is retained intracellularly. One mechanism that traps a serine protease inhibitor like C1 inhibitor intracellulary is the loop–sheet polymerization of Z α_1-antitrypsin (11). A C1 inhibitor reactive-center loop mutant (C1 inhibitor Mo) undergoes a similar polymerization in plasma but is present at relatively normal levels (12). Polymerization kinetics may dictate whether secretion can actually occur.

A mechanism for these mutations can be postulated based on the nucleotide sequence characteristics. A factor that may predispose to the clustering of mutations in this region is the existence of the nine-nucleotide triplet repeat that extends from nucleotide 8449 through nucleotide 8457. C1 inhibitor has three examples of such triplet repeats within its exons. One such repeat is within the untranslated region of exon 8, 65 nucleotides 3′ to the codon for the carboxyl-terminal alanine. There are no mutations identified at this site. The other two examples of triplet repeats both occur in exon 5. The first repeat, CAGCAGCAG, extends from nucleotide 8371 to nucleotide 8379. No mutations have been identified within or near this repeat in the 60 kindred thus far examined or among those reported in the literature. The second triplet is the CAACAACAA repeat involved in the mutations presented here. It is 69 bp downstream from the CAG repeat.

The mutations described here very likely have resulted from slipped mispairing (13). This mechanism involves complementary base pairing between the temporarily melted template and copied strands. During reannealing, the two strands slip out of correct register and again anneal so that the replication fork can continue. If the copied strand slips 5′ and reanneals, a duplication results, or if it slips 3′ on the template strand, a deletion occurs (Fig. 2). After the first round of replication, the mutation is in only one strand. These mutations can be made permanent by another round of replication or by DNA repair mechanisms. Because of the repetitive

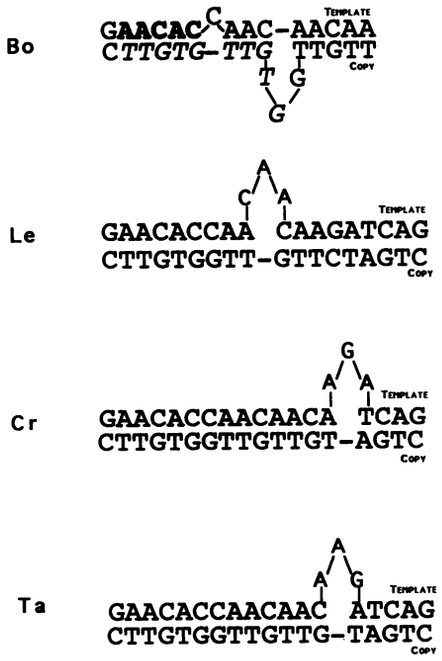

FIG. 2. Template and copied strands are drawn to illustrate the specific slipped mispairing alignment. These are all simple mechanisms except Bo in whom possibly guanine was copied and then slipped mispaired to the 5′ complementary cytosine. Boldfaced letters illustrate the twice-copied sequence in Bo. Italic letters represent the duplicated sequence. The looped-out letters in Cr, Le, and Ta represent deleted portions.

sequence in the affected region of the C1 inhibitor gene, the copied and template strands easily could come into slipped but complementary register, resulting in the mutations. However, the discrepancy in the mutation rates between the two sets of adjacent triplet repeats requires further explanation. It may simply be that enough mutants have not been examined and that the CAG direct triplet repeat may also exhibit an increased mutation rate. The mutation rate of the CAG repeat, however, may be lower because it contains more C·G base pairing than the CAA repeat and, therefore, theoretically increases the energy required to melt the double-stranded DNA to generate the single-stranded DNA to allow for slipped mispairing to occur. While this concept makes sense, it could only be plausible if the increased C·G density substantially increased the melting temperature. This does not seem to be the case. The melting temperatures for the 9-bp repeats differ by only 6°C [24°C for CAA vs. 30°C for CAG by $t_m = 2(A + T) + 4(G + T)$ method]. Furthermore, examples of triplet repeat expansion of CGG in fragile X disease (13) and CTG in myotonic dystrophy (14), Kennedy disease (15), and Huntington disease (16) likely involve slipped mispairing and are relatively C+G rich compared to the CAA repeat described here.

Another mechanism that may be at play in producing these mutations involves the interaction of the DNA polymerase and the DNA template secondary structure. The DNA polymerase holoenzyme, the polymerase and associated enzymes that allow the growth of the replication fork, may encounter a difficult passage through the region of sequence responsible for these mutations. Previous literature describes an increase in mutations adjacent to sequences that cause the polymerase to pause (17). Perhaps this repeat is more mutagenic based purely on the primary nucleotide structure. Alternatively, these pause sites may be alternative DNA secondary structures. One such non-Watson–Crick base-paired structure that may be involved in these examples is a triple helix (18). A triple helix consists of a segment of DNA that is roughly in a double helix conformation with an up- or downstream segment of single-stranded DNA from the same DNA molecule bound in the major groove. Clearly, such a curious structure has strict parameters governing its formation (19). First, such alternative secondary structures form only to "absorb" the negative superhelical tension, or underwinding, of prokaryotic and eukaryotic DNA. This tension is increased during the replication process and requires helicases to adjust the superhelical tension and allow the replication fork to advance. These structures cannot form in random DNA but require oligopurine–oligopyrimidine tracts. Triple helices have been shown to form in eukaryotic DNA. Examination of the 3′ sequence of exon 5 and the adjacent intron reveal oligopurine–oligopyrimidine tracts (8475–8722 and 8896–8950) that may fulfill the requirements for triple helix formation. While the polymerase pauses during replication of this triple repeat, perhaps to "untangle" this secondary structure (20), the template and copied strands dissociate and come back into register over incorrect, but complementary, direct repeats. The CAG repeat may be outside a critical distance from the polymerase–DNA pause site and therefore may not participate in the mutagenesis exhibited by the CAA repeat. Further mutation identification together with basic mutagenesis investigation using *in vitro* model systems may more clearly define the mutational mechanisms that have affected the C1 inhibitor gene.

We thank Gregory Grabowski, M.D., Nancy Lesley, M.D., and Richard R. Sinden, Ph.D., for their generous discussion and constructive suggestions. This work was supported by United States Public Health Service Grant HD22082.

1. Donaldson, V. H. & Evans, R. R. (1961) *J. Lab. Clin. Med.* **58,** 812.
2. Stoppa-Lyonnet, D., DuPonchel, C., Meo, T., Laurent, J., Carter, P. E., Arala-Chaves, M., Cohen, J. H. M., Dewald, G., Goetz, J., Hauptmann, G., Lagrue, G., Lesavre, P., Lopez-Trascasa, M., Misiano, G., Moraine, C., Sobel, A., Spath, P. J. & Tosi, M. (1991) *Am. J. Hum. Genet.* **49,** 1055–1062.
3. Carter, P. E., DuPonchel, C., Tosi, M. & Fothergill, J. E. (1991) *Eur. J. Biochem.* **197,** 301–308.
4. Cooper, D. N. & Youssoufian, H. (1988) *Hum. Genet.* **78,** 151–155.
5. Sinden, R. R. & Wells, R. O. (1991) *Curr. Opin. Biotechnol.* **3,** 612–622.
6. Bissler, J. J., Donaldson, V. H. & Davis, A. E., III (1994) *Hum. Genet.* **93,** 265–269.
7. Wells, R. D. (1988) *J. Biol. Chem.* **263,** 1095–1098.
8. Verpy, E., Laurent, J., Späth, P., Meo, T. & Tosi, M. (1991) *Complement Inflammation* **8,** 238–239 (abstr.).
9. Davis, L. G. & Battey, J. F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York).
10. Parad, R. B., Kramer, J., Strunk, R. C., Rosen, F. S. & Davis, A. E., III (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6786–6790.
11. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. (1992) *Nature (London)* **357,** 605–607.
12. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31,** 77–84.
13. Caskey, C. T., Pizzuti, A., Fu, Y.-H., Fenwick, R. G., Jr., & Nelson, D. (1992) *Science* **256,** 784–789.
14. Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, P. S., Shaw, D. J. & Housman, D. E. (1992) *Cell* **68,** 799–808.
15. LaSpada, E. M., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. (1991) *Nature (London)* **352,** 77–79.
16. The Huntington's Disease Collaborative Research Group (1993) *Cell* **72,** 971–983.
17. Papanicolaou, C. & Ripley, L. S. (1991) *J. Mol. Biol.* **221,** 805–821.
18. Wells, R. D. & Sinden, R. R. (1993) in *Genome Analysis*, Vol. 7, *Genome Rearrangement and Stability*, eds. Warren, S. T. & Davies, K. A. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 107–138.
19. Wells, R. D., Collier, D. A., Harvey, J. C., Shimizu, M. & Wohlrob, F. (1988) *FASEB J.* **2,** 2939–2979.
20. Baran, N., Lopidot, A. & Monor, H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 507–511.