The Effects of Different Cysteine for Glycine Substitutions within α2(I) Chains

EVIDENCE OF DISTINCT STRUCTURAL DOMAINS WITHIN THE TYPE I COLLAGEN TRIPLE HELIX*

(Rceived for publication, August 2, 1990)

Richard J. Wenstrup‡§, Anne W. Shrago-Howard‡¶, Laura W. Lever‡, Charlotte L. Phillips§∥, Peter H. Byers‡**, and Daniel H. Cohn‡‡

From the Departments of ‡Pediatrics and §Medicine, Duke University Medical Center, Durham, North Carolina 27710, the **Department of Pathology and Center for Inherited Diseases, University of Washington, Seattle, Washington 98195, and the ¶Division of Medical Genetics, Cedars-Sinai Medical Center, Los Angeles, California and the Department of Pediatrics, University of California, Los Angeles, California 90048

Affected individuals from two apparently distinct, mild osteogenesis imperfecta families were heterozygous for a G to T transition in the COL1A2 gene that resulted in cysteine for glycine substitutions at position 646 in the α2(I) chain of type I collagen. A child with a moderately severe form of osteogenesis imperfecta was heterozygous for a G to T transition that resulted in a substitution of cysteine for glycine at position 259 in the COL1A2 gene. Type I collagen molecules containing an α2(I) chain with cysteine at position 259 denatured at a lower temperature than molecules containing an α2(I) chain with cysteine at position 646. In contrast to cysteine for glycine substitutions in the α1(I) chain, the severity of the osteogenesis imperfecta phenotype is not directly proportional to the distance of the mutation from the amino-terminal end of the triple helix. These findings could be explained if the type I collagen triple helix contains discontinuous domains that differ in their contributions to maintaining helix stability.

Virtually all forms of the heritable bone disorder osteogenesis imperfecta are due to heterozygosity for mutations in the structural genes for type I collagen (1, 2). Mutations that result only in quantitative decreases in the synthesis of type I procollagen from cultured dermal fibroblasts result in a mild OI phenotype (3-6) identical to OI type I in the clinical classification of Sillence et al. (7). In contrast, heterozygosity for mutations that result in the synthesis and secretion of structurally abnormal type I procollagen molecules usually causes more severe OI variants (8-10). The great majority of mutations in this latter group appear to result in substitutions for triple helical glycine residues in pro-α1(I) or pro-α2(I) chains (11-22). Type I procollagen molecules containing α1(I) or α2(I) chains with substitutions for triple helical glycine residues have increased post-translational modification (primarily excess hydroxylation and glycosylation of Y-position lysyl residues), usually have lower than normal denaturation temperatures, and are often less efficiently secreted than normal molecules (reviewed in Ref. 1). Within the group of OI patients whose cells make structurally abnormal type I collagen, there is a wide range of clinical severity, from lethal OI to mild osseous fragility. Some substitutions for glycine in α1(I) chains exhibit a "position effect" in that the severity of an OI clinical phenotype appears to be related to the closeness of the mutation to the carboxyl-terminal end of the triple helix (1, 20).

We report biochemical and DNA sequence analysis of two different cysteine for glycine substitutions in the triple helical domain of pro-α2(I) chains in three families with nonlethal variants of osteogenesis imperfecta. The more amino-terminal cysteine substitution results in significantly less stable type I collagen molecules than the other substitution and is associated with a more severe clinical phenotype. These data suggest that disruption of specific regions by glycine substitutions differentially affects the stability of the type I collagen triple helix.

MATERIALS AND METHODS

Clinical Description of Affected Individuals

The families have been described more extensively by Cohn and Byers. The families have been described more extensively by Cohn and Byers.

Family A—The proband suffered a congenital hip dislocation. She suffered a fractured tibia at 2 years of age and a fracture of the left humerus at 3 and 8 years of age. An affected sister suffered a hip fracture and within the next 2 years had two femur fractures. Their mother and maternal grandfather were also affected. OI in all individuals was characterized by mild short stature and moderate fracture frequency with little deformity.

Family B—The proband is the only child of an affected father in a large family with a history of OI. At birth, there was marked bowing of the femurs and tibiae, and a skeletal survey showed generalized demineralization. Overall, the severity of OI in this family was comparable to family A, but more variable, with some obligate gene carriers having no history of fractures and others having deforming OI. However, even the most severely affected members of this family were less affected than the proband of family C.

Family C—The proband was the first child of parents who were clinically and biochemically normal. Birth weight was low for gestational age; physical examination at birth revealed a wide anterior

2 Cohn, D. H., and Byers, P. H. (1991) Human Genet., in press.
Cysteine for Glycine in \( \alpha_2(1) \) Collagen Chains

2591

fontanelle and marked shortness and bowing of the upper and lower limbs. Radiographs showed marked generalized demineralization. Two fractures were noted in the neonatal period.

Cell Culture

Cultured dermal fibroblasts were maintained in Dulbecco’s modified Eagle’s medium supplemented with 20% calf serum (Gibco Laboratories) in humidified 5% CO₂ at 37 °C.

Pulse Labeling of Cells with \([\text{H}]\) Proline or \([\text{S}]\) Cysteine

Cells (1 × 10⁵) were plated in 100-mm dishes and allowed to attach and spread overnight. Cells were preincubated for 4 h in RPMI 1640 Select-Amine (Gibco Laboratories) lacking either proline and hydroxyproline (for proline labeling) or cysteine (for cysteine labeling) and supplemented with 50 µg/ml ascorbate. After 4 h, the medium was removed and replaced with fresh medium and either 100 µM of 2,3,4,5-\([\text{H}]\) hydroxyproline (108 Ci/mmol) or 200 µCi of \([\text{S}]\) cysteine (>600 Ci/mmol) (Amersham Corp.) for 16 h. Procollagens were harvested from fibroblast medium into inhibitors as previously described (12) and then dialyzed three times in 2 l of 0.5 N acetic acid prior to lyophilization.

Peptide Digestion of Polypeptides

To produce collagen-sized molecules, polypeptides were digested in 1.0 ml of 50 µg/ml pepsin and 0.5 N acetic acid at 4 °C for 16 h; peptide pepstatin was added, and samples were dialyzed three times for 5 h in 2 liters of 0.5 N acetic acid at 4 °C prior to lyophilization.

Melting Point Experiments

Temperatures were determined using the following adaptation of the methods of Bruckner and Prokop (23). \([\text{H}]\) Proline or \([\text{S}]\) Cysteine-labeled collagens synthesized by cells from affected members of each family were dissolved in 400 nm NaCl and 100 mM Tris (pH 7.4), and 45 µl was added to 0.5 ml microcentrifuge tubes. One tube of each sample was incubated at room temperature, and the other tubes were placed in a Perkin-Elmer Cetus thermal cycler set initially at 50 °C and programmed to increase 1 °C every 12 min. One tube each of \([\text{H}]\) proline- and \([\text{S}]\) cysteine-labeled collagens was removed at the end of each 10-min plateau at 36–42 °C. The sample was cooled at 20 °C for 30 s and then digested with trypsin and chymotrypsin at final concentrations of 100 and 250 µg/ml, respectively, for 2 min. Digestion was stopped by addition of SDS and 2-mercaptoethanol followed by boiling. Digestion products were separated by gel electrophoresis on 5% SDS-polyacrylamide gels (24) and were detected by autoradiography (25) using ENHANCE (Du Pont-New England Nuclear) as the fluor. Quantitation of chains of type I collagen was performed by scanning densitometry using a Pharmacia LKB Biotechnology gel scanner, and peaks were quantitated using LIPS software (Spectrofuge Corp., Durham, NC).

Synthesis of Double-stranded cDNA

**Pro-\( \alpha_2(1) \) cDNA** was synthesized as previously described (26). Total RNA was isolated by the method of Chomczynski and Sacchi (27) from cultured skin fibroblasts which had been preincubated for 72 h in medium supplemented with 100 mM sodium ascorbate, replaced daily. Poly(A) RNA was isolated from total RNA by oligo(dT) chromatinography. Double-stranded cDNA was synthesized by the method of Gubler and Hoffman (28), except that a pro-\( \alpha_2(1) \)-specific primer was employed. Amplification of cDNAs was performed by the method of Gubler and Hoffman (28), except that a pro-\( \alpha_2(1) \)-specific primer was employed. Amplified cDNAs were cloned into pUC18 vectors (Bethesda Research Laboratories). After alkaline denaturation of double-stranded clones (30), cDNA sequence analysis was performed by an adaption of the method of Sanger et al. (32) using a kit (Sequenase, U. S. Biochemical Corp.) according to the manufacturer’s instructions.

**RESULTS**

Cohn and Byers found that pro-\( \alpha_2(1) \) chains synthesized by cells from affected members of families A–C contain a cysteine residue within the triple helical domain, a region from which cysteine is normally excluded. Analysis of cyanogen bromide peptides of \( \alpha_2(1) \) determined that for families A and B, the cysteine substituted for glycine residue 1583, resulting in an alanine (GCT) residue including a valine (GTG) codon at triple helical residue 420.

Pro-\( \alpha_2(1) \) cDNAs were synthesized from total RNA harvested from cultured fibroblasts from an affected member of family B were amplified by the polymerase chain reaction, yielding a single product ~1.3 kilobases in length (data not shown). PCR products were cloned into the EcoRI site of pUC18 containing the EcoRI site of pUC18 and were sequenced in their entirety. Two of seven clones contained a T for G substitution at nucleotide 2341 that resulted in a cysteine residue than the normal glycine codon for triple helical residue 646 (Fig. 2, upper). In addition, all seven clones differed from the published pro-\( \alpha_2(1) \) sequence at nucleotide 1583, resulting in an alanine (GCT) rather than a valine (GTG) codon at triple helical residue 420.

Pro-\( \alpha_2(1) \) cDNAs synthesized from mRNA harvested from cultured fibroblasts from an affected member of family B were amplified by the polymerase chain reaction, yielding a single product ~1.3 kilobases in length (data not shown). PCR products were cloned into the EcoRI site of pUC18 containing the EcoRI site of pUC18 and were sequenced in their entirety. Two of four clones contained a T for G transition at nucleotide 2341 that resulted in a cysteine residue at triple helical residue 646 (data not shown). All four clones also differed from published sequence at nucleotide 1583 as described above.

Pro-\( \alpha_2(1) \) cDNAs from RNA harvested from cells from the

---

**Fig. 1. cDNA cloning strategy for COL1A2 mutations in families A-C.** A, cyanogen bromide peptide map of the triple helical domain of \( \alpha_2(1) \) chains of type 1 collagen. Vertical lines indicate relative locations of PCR primers used to amplify cDNAs from affected members of families A and B, location of PCR primers used to amplify cDNAs from affected members of families A and B, location of PCR primers used to amplify cDNAs for family C.
Cysteine for Glycine in α2(I) Collagen Chains

Fig. 2. Sequence of cDNAs derived from mutant and normal COL1A2 alleles in region of cysteine for glycine substitutions in family A (upper) and family C (lower). The locations of the single nucleotide substitutions are indicated by arrows between the autoradiographs.

proband of family C were amplified by PCR (Fig. 1D), and the products were cloned. Three of seven clones contained a T for G substitution at nucleotide 1180 that resulted in a cysteine for glycine substitution at the codon for triple helical residue 259 (Fig. 2, lower).

To determine whether the position of the two cysteine for glycine substitutions had different effects on the stability of the triple helix, thermal denaturation temperatures of the type I collagen triple helix were assayed by digestion with trypsin and chymotrypsin (Fig. 3). To compare the thermal denaturation of [3H]proline- and [35S]cysteine-labeled proteins, the midpoint temperature (T_m) for helix to coil transition was defined as the resistance of α2(I) chains in type I collagen molecules to protease digestion. For type I collagen molecules containing α2(I) chains with cysteine at position 646, the T_m was 39.5 °C; and for molecules containing α2(I) chains with cysteine at position 259, it was 38 °C. The T_m for type I collagen molecules from control cells labeled with [3H]proline was 40 °C.

DISCUSSION

Cultured dermal fibroblasts from affected members of two families with nonlethal OI synthesized normal and abnormal populations of type I collagen molecules due to heterozygosity for a single nucleotide substitution in COL1A2 that resulted in cysteine for glycine substitutions at different locations in the type I collagen triple helix. In two families, a mild OI phenotype with osseous fragility, mild long bone deformity, and mild short stature was associated with heterozygosity for a cysteine for glycine substitution at position 646 in the triple helix; in a third family, a single proband with severe deformity of long bones and dwarfing had a new dominant mutation that resulted in substitution of cysteine for glycine at position 259. Type I collagen molecules incorporating pro-α2(I) chains with cysteine at position 646 melted 0.5 °C below control type I collagen, and those incorporating pro-α2(I) chains with cysteine at position 259 melted 2.0 °C below control.

T_m values for type I collagen vary slightly from laboratory to laboratory because of differences in experimental procedures, and T_m values of control and OI collagens should only be interpreted as relative values. To improve reproducibility, we programed a thermal cycler to elevate incubation temperature by 1 °C every 12 min; although this produces the same overall temperature elevation rate as was described by Bruckner and Prockop (23), the shape of the temperature curve more closely approximates a series of plateaus than the more continuous elevation generated by a circulating water bath, the method used in previous melting point experiments (12) on type I collagens from OI cell strains. A more important difference is that for the purpose of comparison between proline- and cysteine-labeled molecules, T_m curves in this report only reflect protection of α2(I) chains incorporated into type I collagen molecules. Virtually all melting point studies based on the method of Bruckner and Prockop (23) show loss of protection of α2(I) chains at a lower temperature than that of α1(I) chains, as does our analysis.

Since the type I collagen triple helix winds from its carboxyl- to amino-terminal end, molecules incorporating pro-α chains with substitutions for glycine in the triple helical domain probably assemble normally carboxyl-terminal to the substitution. This model is supported by the consistent finding in many OI cell strains that excess post-translational modifications of abnormal type I collagen molecules occur amino-terminal to the substitution (reviewed in Ref. 1). Either propagation of the triple helix is slowed at the substitution or else the structure of the triple helix amino-terminal to the substitution is abnormal in a way that permits continued post-translational modification. Thus, mutations closest to the carboxyl terminus would alter a greater portion of the triple helix.

Starman et al. (20) proposed that in individuals heterozygous for cysteine for glycine substitutions in the α1(I) chain, severity of the OI clinical phenotype correlates with the position of the substitution along the chain, with the mildest phenotypes resulting from mutations near the amino-terminal end of the triple helix. Arginine for glycine substitutions in α1(I) have a similar position effect, except that the lethal phenotype is seen with more amino-terminal mutations (391 versus 691 for cysteine) (12, 14, 22). However, the more severe OI phenotype associated with cysteine at position 259 in α2(I) chains compared to that associated with cysteine at position 646 suggests that for substitutions in α2(I) chains, the position effect may differ from that in α1(I) or else it may be overridden by disruption of important domains in the triple helix.

Substitutions for glycine residues in α chains of type I collagen can serve as probes for determinants of triple helical stability by examining the decrease in T_m (the temperature of helix to coil transition in type I collagen molecules containing one or more mutant chains).

All but two substitutions for triple helical glycine residues in OI cell strains have resulted in decreased thermal stability of type I collagen molecules. One exception was a substitution of arginine for glycine at position 1012 in the α2(I) chain, the carboxyl-terminal glycine of the triple helix (15). The other was a serine α1(I) glycine at position 844; the limited effect of the latter substitution on thermal stability may be because a serine substituted for glycine results in only minor alterations in the structure of the helix (21).

This study supports a model of the type I collagen triple helix in which some regions contribute to triple helical stabil-
Cysteine for Glycine in α2(I) Collagen Chains

2593

FIG. 3. Thermal denaturation curves (left) of type I collagen from control cells (upper), from affected member of family A (center), and from proband in family C (lower). Proline- or cysteine-labeled procollagens were partially digested with pepsin to produce collagen molecules; and samples from control and OI cell strains were incubated at increasing temperatures, rapidly cooled, and digested with trypsin and chymotrypsin (see "Materials and Methods"). Digestion products were separated by electrophoresis on 5% SDS-polyacrylamide gels and quantified by scanning gel densitometry. ○, denaturation of proline-labeled collagens from control cell strains and products of both normal and abnormal COL1A2 alleles in OI cell strains; □, products of mutant COL1A2 alleles only since cysteine is normally not present in α1(I) and α2(I) chains. Each data point is the average of two different experiments. Complete protection of α2(I) chains was defined as the amount protected after incubation at 22 °C. Autoradiograms of melting point experiments used to generate melting curves.

FIG. 4. Primary structure of α chain type I collagen amino-terminal to cysteine for glycine mutations in chain. Upper, respective locations of COL1A2 mutations in families A–C (arrows) in the type I collagen triple helix. There are 1014 residues in the triple helix; reference numbering of residues in each chain begins at the amino-terminal end of the triple helix. During molecular assembly, the triple helix winds from carboxyl- to amino-terminal end. Lower, residues immediately adjacent and amino-terminal to cysteine substitutions (asterisks) at positions 646 (families A and B) and 259 (family C) in contiguous α1(I) and α2(I) chains of a type I collagen molecule (28, 35). Amino-terminal to position 646 but not position 259 is an abundance of G-P-P and G-P-A triplets (boxed residues) that may provide stability to triple helical structure during assembly of a type I collagen molecule incorporating an α2(I) chain with a cysteine for glycine substitution (33, 34).

The effects on thermal stability of substitutions of cysteine for glycine in the α2(I) chain differ from those in the α1(I) chain. As a general rule, as substitutions in the α1(I) chain move from the carboxyl-terminal end toward the amino terminus, the thermal stability of molecules that contain one or two abnormal chains increases toward normal. Molecules that contain two abnormal chains typically have higher denaturation temperatures than those with one, presumably because the interchain, intramolecular disulfide bond serves to stabilize the molecule. In contrast, there is no similar gradient for substitutions of cysteine for glycine in the α2(I) chain. As a result, mutations in the COL1A2 gene may provide clearer insight than those in the COL1A1 gene into the domain structure of the triple helix and identify those regions which can be destabilized by point mutations.

REFERENCES
1. Byers, P. H. (1989) in The Metabolic Basis of Inherited Diseases (Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds) pp. 2805–2842
Cysteine for Glycine in \( \alpha(2) \) Collagen Chains

2. Prockop D. J., Constantinou, C. D., Combrowski, K. E., Højima, Y., Kadler, K. E., Kuivaniemi, H., Tromp, G., and Vogel, B. E. (1989) *Am. J. Med. Genet.* 34, 60–67

3. Barsh, G. S., David, K. E., and Byers, P. H. (1981) *Proc. Natl. Acad. Sci. U. S. A.* 79, 3838–3842

4. Rowe, D. W., Shapiro, J. R., Poirier, M., and Schlesinger, S. (1985) *J. Clin. Invest.* 76, 604–611

5. Willing, M. C., Cohn, D. H., and Byers, P. H. (1990) *J. Clin. Invest.* 85, 282–290

6. Wenstrup, R. J., Willing, M. C., Starman, B. S., and Byers, P. H. (1990) *Am. J. Hum. Genet.* 46, 975–982

7. Silence, D. O., Senn, A., and Danks, D. M. (1979) *J. Med. Genet.* 16, 101–116

8. bonsen, J., Holbrook, K. A., Gelinas, R. E., Jacob, J., and Byers, P. H. (1985) *J. Biol. Chem.* 260, 1734–1742

9. Bateman, J. F., Mascara, T., Chan, D., and Cole, W. G. (1984) *Biochem. J.* 217, 103–115

10. Wenstrup, R. J., Tsipouras, P. J., and Byers, P. H. (1986) *J. Clin. Invest.* 78, 1449–1455

11. Cohn, D. H., Byers, P. H., Steinmann, B., and Gelinas, R. E. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 6045–6047

12. Bateman, J. F., Chan, D., Walker, I. D., Rogers, J. G., and Cole, W. G. (1987) *J. Biol. Chem.* 262, 7021–7027

13. Vogel, B. C., Minor, R. R., Freund, M., and Prockop, D. J. (1987) *J. Biol. Chem.* 262, 14737–14744

14. Bateman, J. F., Lamande, S. R., Dahl, M., Chan, D., and Cole, W. G. (1988) *J. Biol. Chem.* 263, 11627–11630

15. Wenstrup, R. J., Cohn, D. H., Cohen, T., and Byers, P. H. (1988) *J. Biol. Chem.* 263, 7734–7740

16. Baldwin, C. T., Constantinou, C. D., Dumars, K. D., and Prockop, D. J. (1988) *J. Biol. Chem.* 264, 3002–3006

17. Constantinou, C. D., Nielsen, K. B., and Prockop, D. J. (1989) *J. Clin. Invest.* 85, 574–584

18. Patterson, E., Smiley, E., and Bonadio, J. (1989) *J. Biol. Chem.* 264, 10083–10087

19. Marini, J. C., Grange, D. K., Gottesman, G. S., Lewis, M. B., and Koepplin, D. A. (1989) *J. Biol. Chem.* 264, 11893–11900

20. Starman, B. J., Eye, D. R., Charbonneau, H., Harrylock, M., Weiss, M. A., Weiss, L., Graham, J. M., Jr., and Byers, P. H. (1989) *J. Clin. Invest.* 84, 1206–1214

21. Pack, M., Constantinou, C. D., Kalia, K., Nielsen, K. B., and Prockop, D. J. (1989) *J. Biol. Chem.* 264, 19694–19699

22. Bateman, J. F., Lamande, S. R., Cole, W. J., and Dahl, H.-H. M. (1990) *Ann. N. Y. Acad. Sci.* 580, 324–329

23. Bruckner, P., and Prockop, D. J. (1981) *Anal. Biochem.* 110, 24–26

24. Laskey, R. A., and Mills, A. D. (1975) *Eur. J. Biochem.* 56, 335–341

25. Phillips, C. L., Shrago-Howe, A. W., Pinnell, S. R., and Wenstrup, R. J. (1989) *J. Clin. Invest.* 86, 1223–1228

26. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159

27. Gubler, Y., and Hoffman, B. (1983) *Gene (Amst.)* 276, 263–269

28. de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M.-L., Dickson, L., Weil, D., and Ramirez, F. (1987) *J. Biol. Chem.* 262, 16032–16036

29. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467

30. Zhang, H., School, R., Browse, J., and Somerville, C. (1988) *Nucleic Acids Res.* 16, 1220

31. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467

32. Dölz, R., and Heideman, E. (1986) *Biopolymers* 25, 1069–1080

33. Thakur, S., Vadas, D., Germann, H.-P., and Heidemann, E. (1986) *Biopolymers* 25, 1081–1086

34. Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., and Prockop, D. J. (1983) *Biochemistry* 22, 5213–5223