# Structural study of a mutant type I collagen from a patient with lethal osteogenesis imperfecta containing an intramolecular disulfide bond in the triple-helical domain

Wolfie Traub* and Beat Steinmann

*Structural Chemistry Department, Weizmann Institute of Science, 76100 Rehovot, Israel and Division of Metabolism, Department of Pediatrics, University of Zurich, 8032 Zürich, Switzerland

Received 15 January 1986

We have built molecular models of collagen type I from a patient with lethal osteogenesis imperfecta incorporating one or two mutant $\alpha_1$(I)-chains which contain a cysteine substituting a glycine near the C-terminal end. In either case, the cysteines can only be accommodated with considerable distortion of the native collagen structure, which disrupts inter-chain contacts. The disturbance of the triple helix is limited to a small local region. This suggests that the most important consequence of the mutation is delayed helix formation leading to overmodification and decreased collagen production, rather than the structural abnormality of the folded molecules, which are only marginally unstable.

*Osteogenesis imperfecta* *Collagen type I* *Structural defect* *Cysteine subsitution* *Molecular model* *Triple helix*

## 1. INTRODUCTION

Recently, a patient with perinatal lethal osteogenesis imperfecta has been described in which the abnormalities are attributed to mutant collagen $\alpha_1$(I)-chains [1]. Both $\alpha_1$(I)- and $\alpha_2$(I)-chains of collagen molecules which contained either one or two mutant $\alpha_1$(I)-chains were found to undergo excessive lysyl hydroxylation and hydroxylysyl glycosylation. Such molecules showed a melting temperature of 38°C instead of the normal 41°C. Increased intracellular degradation of newly synthesized collagen, delayed secretion and decreased accumulation of extracellular collagen were observed. Collagen fibril diameters were found to be unusually small.

Biochemical analysis [1] and DNA sequencing [2] have served to locate the mutation near the C-terminal end of the triple-helical domain of the $\alpha_1$(I)-chains, where a cysteine is substituted for a glycine at position 988 due to a point mutation [2]. Collagen molecules containing two such mutant $\alpha_1$(I)-chains were found to have inter-chain disulfide bonds. It was proposed that this mutation would cause at least local distortions in the structure of the collagen molecule, which might account for the observed abnormal physicochemical and biological properties. Therefore, we decided to investigate, with the aid of molecular models, what the nature and extent of the conformational alterations engendered by the mutant sequence might be.

The type I collagen molecule has a 3-stranded (rope-like) or triple-helical structure (fig.1) extending over 1014 amino acid residues with short non-helical or telopeptide regions at either end. Of the 3 polypeptide chains per molecule, the $\alpha_2$(I)-chain differs somewhat in sequence from the two identical $\alpha_1$(I)-chains, but all 3 conform to the polytripeptide sequence $(Gly\text{-}X\text{-}Y)_n$ over the triple-helical domain. The X and Y positions can accommodate any amino acid residues, with only a few exceptions, but substitution at a glycine position of any other residue would be expected to interfere with inter-chain contacts and hydrogen bonding [3].

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/86/$3.50 

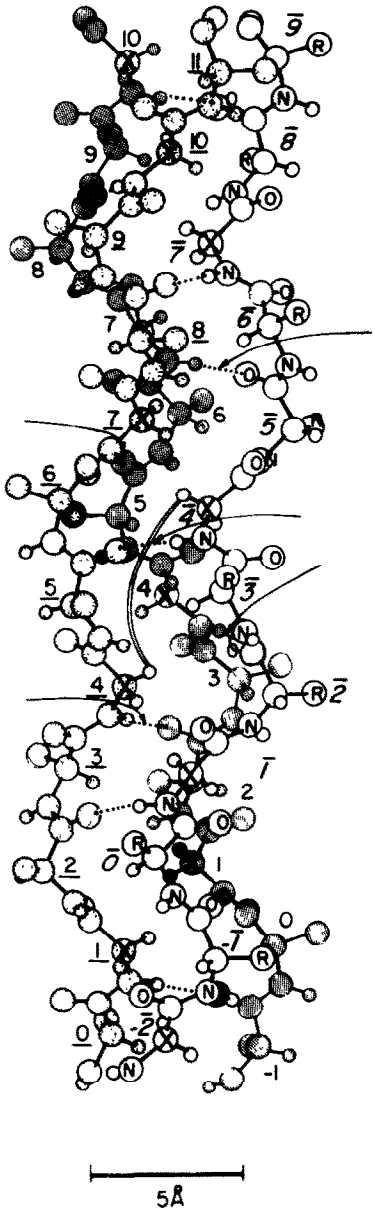


Fig.1. Triple-helical conformation of collagen. In this model the 3 polypeptide chains are staggered by one amino acid residue in the order $\alpha_1$ (stippled), $\alpha_2$ (cross-hatched) and $\alpha_1$ (white). Equivalent amino acid residues on the 3 $\alpha$-chains have the same number at the $\alpha$-carbon atoms. Every third residue (i.e. 1, 4, 7 and 10), marked with an X on the $\alpha$-carbon, can only be glycine. Positions of nitrogens (N), oxygens (O) and side-chains (R) are indicated only in the C-terminal chain; other atoms are carbons (large circles) or hydrogens (small circles). There is one inter-chain hydrogen bond per tripeptide, indicated by ....., linking the NH of glycines to the CO groups of residues −1, 2, 5 and 8. The double line indicates where the disulfide bond between cysteines of the N- and C-terminal $\alpha_1$(I)-chains in the mutant collagen would be. Hydrogen bonds in the vicinity of the disulfide bond are indicated by arrows.

## 2. MATERIALS AND METHODS

We have used two kinds of molecular models in our study; CPK space-filling models designed to a scale of 1.25 per Å, and Labquip ball-and-rod models with a scale of 1.0 cm per Å. We first built the models to conform to the regular triple-helix conformation [4], and then tried to find the least distortions required to accommodate cysteine residues, joined by a disulfide bond, at two equivalent $\alpha_1$(I) glycine positions.

In building models we have considered two situations depending on the position of the $\alpha_2$(I)-chain relative to the two $\alpha_1$(I)-chains. If $\alpha_2$(I) were the middle chain (shown cross-hatched in fig.1), the distance between equivalent glycine $\alpha$-carbon atoms in the $\alpha_1$(I)-chains of the undistorted collagen molecule would be some 6.4 Å. However, this distance would be only 3.7 Å if $\alpha_2$(I) were in the N-terminal or C-terminal position. The $\alpha$-carbon to $\alpha$-carbon distance in a fully extended disulfide bond would be 7.0 Å, and in a number of well-defined protein crystal structures it has been found to range between 4.8 and 6.5 Å [5].

## 3. RESULTS AND DISCUSSION

We have been able to build what appears to be a quite plausible model with the $\alpha_2$(I)-chain in the middle position (figs 1 and 2). In the model the two $\alpha_1$-chains are pushed apart forming a groove to accommodate the two extra carbon atoms and two sulfur atoms of an extended disulfide bond. This involves rupture of the hydrogen bond between the two $\alpha_1$(I)-chains in the immediate vicinity of the mutation, but the adjacent hydrogen bonds, though somewhat distorted, may still remain intact.

If the chain sequence were $\alpha_1$, $\alpha_1$, $\alpha_2$, or $\alpha_2$, $\alpha_1$, $\alpha_1$, an inter-chain disulfide bond could also be

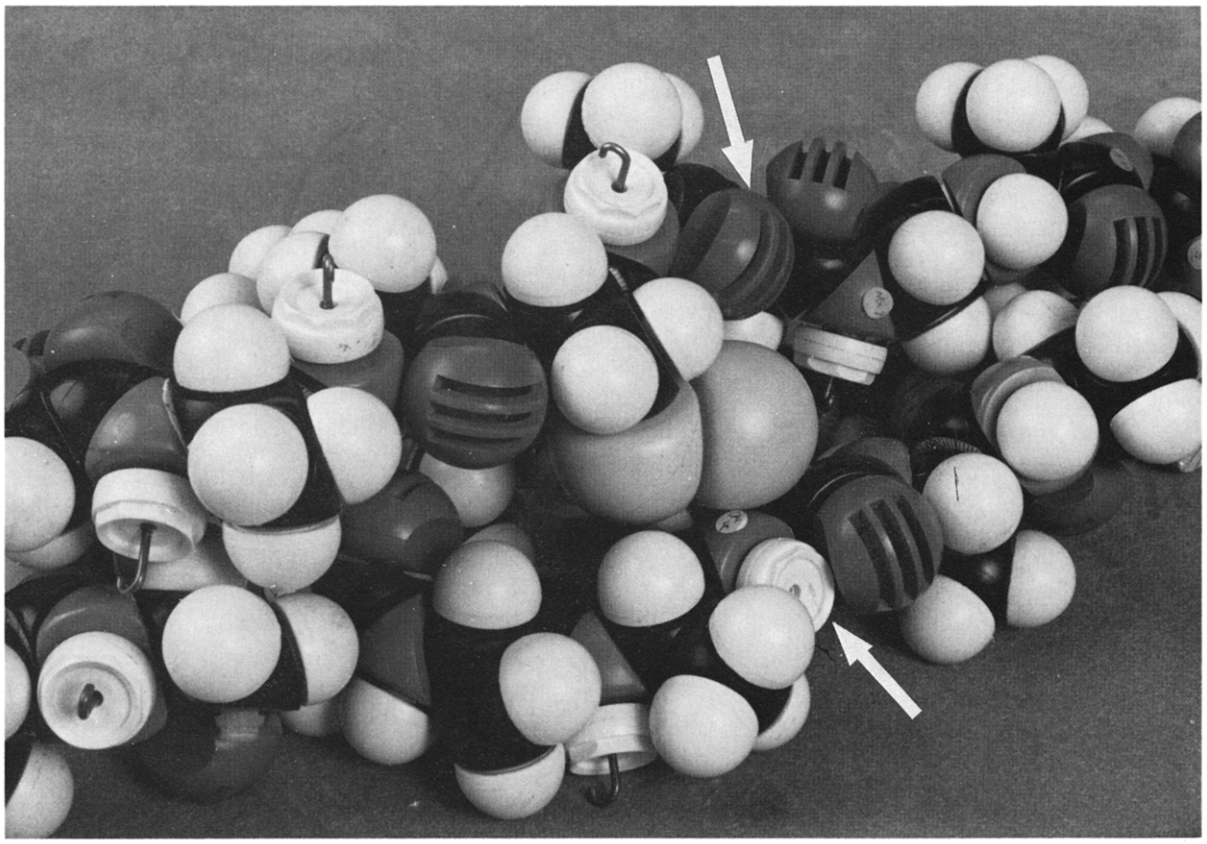

Fig.2. Space-filling model of triple-helical domain of mutant collagen with cysteine residues, substituted for two equivalent glycines in $\alpha_1$(I)-chains, joined by a disulfide bond. In this model the 3 polypeptide chains are staggered by one amino acid residue in the order $\alpha_1$, $\alpha_2$, $\alpha_1$. Arrows indicate NH (residue 4) and CO (residue 5), the components of the hydrogen bond between N- and C-terminal $\alpha_1$(I)-chains disrupted by the disulfide bond. Note the compactness of the collagen triple helix on either side of this 'bubble'.

built into the model, but somewhat less elegantly. Because of the closer distance between the two glycine $\alpha$-carbon atoms in the undistorted triple helix, the two $\alpha_1$(I)-chains would have to be pushed further apart to accommodate a normally extended disulfide bond, with the disruption of probably two or more adjacent inter-chain hydrogen bonds.

Though the sequence of polypeptide chains in the collagen molecule has not been determined unequivocally, circumstantial evidence based on maximisation of stabilizing inter-chain intramolecular interactions favors the sequence $\alpha_1$, $\alpha_2$, $\alpha_1$ [6]. In either case, it is clear that substitution of cysteine for a glycine must cause some disruption of inter-chain contacts and hydrogen bonding. (Even molecules containing only one mutant $\alpha_1$(I)-chain cannot accommodate the cysteine side chain without considerable deviation from the regular collagen triple helix. However, in this case we have been able to maintain all the inter-chain hydrogen bonds in the molecular model, by appreciably distorting the collagen structure.)

This would be expected to delay triple-helix formation, which is known to start at the C-terminal end [7], and consequently allow increased lysyl hydroxylation and hydroxylysyl glycosylation of the bulk of all 3 $\alpha$-chains on the N-terminal side of the mutated residue(s). The model studies show that the disturbance in the triple helix is probably limited to a small local region and that therefore the decrease in melting temperature is rather subtle. However, they do not allow an estimate of the

delay with which helix formation proceeds over the conformationally altered region towards the N-terminus and which allows overmodification. Direct kinetic measurements of helix formation could provide this information.

## ACKNOWLEDGEMENTS

This work was supported in part by US PHS grant DE 06954 to W.T. and Swiss National Science Foundation grant 3.905.083 to B.S.

## REFERENCES

[1] Steinmann, B., Rao, V.H., Vogel, A., Bruckner, P., Gitzelmann, R. and Byers, P.H. (1984) J. Biol. Chem. 259, 11129–11138.
[2] Cohn, D.H., Byers, P.H., Steinmann, B. and Gelinas, R.E. (1985) in press.
[3] Bornstein, P. and Traub, W. (1979) in: The Proteins (Neurath, H. and Hill, R.L. eds) 3rd edn, vol.4, pp.411–632, Academic Press, New York.
[4] Traub, W., Yonath, A. and Segal, D.M. (1969) Nature 221, 914–917.
[5] Brookhaven Protein Data Files 2SGAB (James, M.N.G. and Sielecki, A.R.), 4PTIA (Huber, R., Kukla, D., Ruehlmann, A., Epp, O., Formanek, H., Diesenhofer, J. and Steigemann, W.), 6LYZE (Diamond, R., Phillips, D.C., Blake, C.C.F. and North, A.C.T.) and 3SGBB (Read, R.J., Fujinaga, M., Sielecki, A.R. and James, M.N.G.).
[6] Traub, W. and Fietzek, P.P. (1976) FEBS Lett. 68, 245–249.
[7] Bächinger, H.P., Bruckner, P., Timpl, R. and Engel, J. (1978).