Changes in collagen stability and folding in lethal perinatal osteogenesis imperfecta

The effect of $\alpha 1(I)$ -chain glycine-to-arginine substitutions

Anthony T. BAKER,*† John A. M. RAMSHAW,* Danny CHAN,† William G. COLE† and John F. BATEMAN†§

*C.S.I.R.O. Division of Biotechnology, 343 Royal Parade, Parkville, Vic. 3052, Australia, and †Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Parkville, Vic. 3052, Australia

The effect of glycine-to-arginine mutations in the $\alpha 1(I)$ -chain on collagen triple-helix structure in lethal perinatal osteogenesis imperfecta was studied by determination of the helix denaturation temperature and by computerized molecular modelling. Arginine substitutions at glycine residues 391 and 667 resulted in similar small decreases in helix stability. Molecular modelling suggested that the glycine-to-arginine-391 mutant resulted in only a relatively small localized disruption to the helix structure. Thus the glycine-to-arginine substitutions may lead to only a small structural abnormality of the collagen helix, and it is most likely that the over-modification of lysine, poor secretion, increased degradation and other functional sequelae result from a kinetic defect in collagen helix formation resulting from the mutation.

INTRODUCTION

Mutations in the type I collagen *COL1A1* and *COL1A2* genes result in human 'brittle-bone disease', lethal perinatal osteogenesis imperfecta type II. Although large gene rearrangements have been defined (Barsh & Byers, 1981; Barsh *et al.*, 1985; Chu *et al.*, 1985; Byers *et al.*, 1988; Willing *et al.*, 1988), more commonly, subtle structural mutations of the pro $\alpha 1(I)$ - and pro $\alpha 2(I)$ -chains lead to the disease (Steinmann *et al.*, 1984; Cohn *et al.*, 1986; Vogel *et al.*, 1987; Bateman *et al.*, 1987*a,b*, 1988*a*). The unifying feature of these different mutations is that they all lead to abnormally high extents of post-translational modification of lysine in the collagen α -chains (Bonadio & Byers, 1985; Bateman *et al.*, 1986).

The widely accepted explanation for this phenomenon is that the mutations delay helix formation and thus increase the exposure of the α -chains to the modifying enzymes (Bonadio & Byers, 1985). In the helical region of the α -chains, glycine occurs as every third residue and is a structural prerequisite for normal triple-helix folding in the interstitial collagens (Bornstein & Traub, 1979). Mutations leading to substitution of these glycine residues would be expected to have a significant effect on triple-helix formation and the properties of the resulting collagen (Bonadio & Byers, 1985).

Glycine-to-cysteine substitutions of the $\alpha 1(I)$ -chain have been described (Steinmann *et al.*, 1984; Cohn *et al.*, 1986; Vogel *et al.*, 1987). These result in over-hydroxylation of the collagen and a decrease in the thermal stability of the molecule. Previously, in a study of the structural effects associated with the glycine-to-cysteine-998 substitution, Traub & Steinmann (1986) examined space-filling molecular models of the predicted structure of the mutant collagen. These studies suggested that a glycine-to-cysteine mutation caused only a limited local disturbance to the helix. However, these glycine-to-

cysteine mutations introduce disulphide bonds into the helix, and thus may represent a special subset of the eight possible substitutions that can arise from a single base change.

We have recently described in two patients with osteogenesis imperfecta type II glycine-to-arginine mutations at positions 391 (Bateman *et al.*, 1987*b*) and 667 (Bateman *et al.*, 1987*a*, 1988*a*) of the $\alpha 1(I)$ -chain. [The latter mutation was wrongly reported to be at position 664 in Bateman *et al.* (1988*a*).] In the present study the effect of these glycine-to-arginine substitutions on collagen structure was examined by helix-stability measurements and molecular modelling (Levitt, 1974).

EXPERIMENTAL

Human fibroblast culture and collagen labelling

Dermal fibroblasts from controls and from two patients with lethal perinatal osteogenesis imperfecta type II were grown as previously described (Bateman *et al.*, 1984, 1986). The two patients were heterozygous for the mutation and produced both normal and mutant type I collagen with increased extents of post-translational modification of lysine. The mutations have been fully characterized as glycine-to-arginine point mutations at amino acid residues 391 (Bateman *et al.*, 1987*b*) and 667 (Bateman *et al.*, 1987*a*, 1988*a*) of the $\alpha 1(I)$ -chain. Procollagens were labelled in cell cultures for 18 h with 10 μ Ci of [5- 3 H]proline/ml in Dulbecco's modified Eagle's medium containing 0.15 mM-sodium ascorbate and 10 % (v/v) foetal-calf serum as previously described (Bateman *et al.*, 1984, 1986). Procollagen in the medium fractions was purified by precipitation with $(\text{NH}_4)_2\text{SO}_4$ at 25 % saturation and subjected to limited digestion with pepsin followed by freeze-drying.

† Present address: Department of Chemistry, University of Technology, Sydney, N.S.W. 2007, Australia.

§ To whom correspondence should be addressed.

Thermal stability of collagen

The freeze-dried [^3H]proline-labelled collagens were redissolved in 0.4 M-NaCl/0.1 M-Tris/HCl buffer, pH 7.4, at 4 °C. Under these conditions the pepsin is irreversibly inactivated. The samples were warmed stepwise (1 °C/10 min) from 34 °C to 43 °C. At 0.5 °C intervals samples were taken and digested with a mixture of trypsin and chymotrypsin (Bruckner & Prockop 1981) to determine the temperature at which the helix was susceptible to degradation by the proteinases, which indicates the temperature at which it was thermally denatured. The proportion of collagen that was resistant to proteolysis, through retaining its helical structure, was determined by SDS/polyacrylamide-gel electrophoresis of the samples. Collagen α -chain bands were identified by fluorography and were quantified by excision and scintillation counting of radioactivity (Bateman *et al.*, 1988b).

Molecular model building and energy refinement

Molecular models of (Gly-Pro-Hyp) $_{10}$ and a 30-residue fragment of human collagen, residues 376–405 (Bernard *et al.*, 1983; Bateman *et al.*, 1987b; de Wet *et al.*, 1987), were constructed on an Evans and Sutherland PS-300 graphics system by use of the FRODO program (Jones, 1982). Co-ordinates for (Gly-Pro-Hyp) $_{10}$ were determined from X-ray fibre diffraction of kangaroo tail tendon by using this as a collagen model structure for a least-squares refinement (Fraser *et al.*, 1979). The co-ordinates for the human fragment were based on those for (Gly-Pro-Hyp) $_{10}$ with replacements to give the observed amino acid sequence (Bernard *et al.*, 1983; Bateman *et al.*, 1987b; de Wet *et al.*, 1987) made by using the SAM sub-routine in the FRODO program. In positions where replacements were made, the position was first changed to glycine, and then changed to the required residue. This procedure avoided the positions of the side-chain atoms of the previous side chain influencing those of the new side chain. In all cases the initial configurations of the side chains were determined solely by the placement given by the FRODO program, and no operator-based changes were made. The human collagen fragments were constructed with the alternative chain orders $\alpha 2\alpha 1\alpha 1$ and $\alpha 1\alpha 2\alpha 1$.

Refinement of molecular models was performed by using the energy-refinement method (EREF) of Levitt (1974). Typically, 30 cycles of refinement were used, although in some cases up to 60 cycles were required to allow convergence towards a minimum. Each model was refined independently, and was not based on a previously refined structure.

The EREF procedure with (Gly-Pro-Hyp) $_{10}$ indicated that this model structure was very close to an energy minimum. Within the central six triplets of the structure the shifts in C_{α} co-ordinates were all less than 0.007 nm. The greatest shifts in the structure occurred in proline rings, where a slight buckling was observed. Since the ends of each fragment were not fixed, a slight 'fraying' can occur, so the co-ordinates were only examined in this central portion of the fragment. The extent of this 'fraying' varied slightly between repeat refinements, so the calculated energy values obtained from the refinement could not be rigorously compared.

In the models with the osteogenesis-imperfecta substitution glycine-to-arginine-391 the inserted arginine

side chain was sometimes in a sterically constrained position, which prevented the refinement from proceeding. In these cases the substituted side chain was moved manually by rotation of the side chain at the β -carbon atom, so that it was external to the triple-helix backbones and not interfering with other residues. In addition, with certain of the osteogenesis-imperfecta models, the adjacent proline residues were changed from the *trans* to *cis* configuration. In these cases the conformation of the chain was also adjusted to bring bond lengths and angles and non-bonded contacts to acceptable values before the refinement process.

RESULTS AND DISCUSSION

The thermal stability of the type I and type III collagen synthesized by the two patients with the glycine-to-arginine mutations was estimated by measuring its resistance to proteinase digestion after heating to different temperatures (Fig. 1a). In both mutants the type I collagen 'melting' temperature was approx. 39.8 °C

Fig. 1. Thermal denaturation of type I and III collagens

Pepsin-digested [^3H]proline-labelled collagens from the medium of control and osteogenesis-imperfecta fibroblasts were warmed slowly from 34 °C to 43 °C, and at 0.5 °C intervals samples were digested with trypsin and chymotrypsin. The degree of helicity was determined by the relative proportion of α -chains resistant to proteolysis resolved by SDS/polyacrylamide-gel electrophoresis (see the Experimental section for details). (a) Type I collagen from control fibroblasts (○), osteogenesis-imperfecta glycine-to-arginine-391 fibroblasts (△) and osteogenesis-imperfecta glycine-to-arginine-667 fibroblasts (□). (b) Type III collagen from control fibroblasts (○), osteogenesis-imperfecta glycine-to-arginine-391 fibroblasts (△) and osteogenesis-imperfecta glycine-to-arginine-667 fibroblasts (○).

Fig. 2. Molecular models showing the proposed structural changes associated with the $\alpha 1(I)$ -chain-glycine-to-arginine-391 substitution

Molecular models were calculated as described in the Experimental section by EREF refinement of the native and mutant $\alpha 1(I)$ -chain collagen sequences. The models illustrated have the α -chain orders $\alpha 2\alpha 1\alpha 1$ and include the mutant chain ($\alpha 1^*$) in the following chain associations: (a) $\alpha 2\alpha 1^*\alpha 1$; (b) $\alpha 2\alpha 1\alpha 1^*$; (c) $\alpha 2\alpha 1^*\alpha 1^*$. The mutant chain positions are superimposed on the normal sequence. Where changes occur the mutant chain $C_{(\alpha)}$ atoms were displaced away from the axis of the helix. The glycine-to-arginine substitutions are indicated by arrows. Colour prints are available from the authors on request.

compared with 41.3 °C for type I collagen from control cells. In contrast, the type III collagen from the osteogenesis-imperfecta cases had a 'melting' temperature of 39.3 °C, indistinguishable from control type III collagen (Fig. 1b). The lowering of the 'melting' temperature in the glycine-to-arginine mutants was less than that reported in other cases of osteogenesis imperfecta type II (Steinmann *et al.*, 1984; Bonadio & Byers, 1985; Vogel *et al.*, 1987).

As both the glycine-to-arginine-391 and glycine-to-arginine-667 patients were heterozygotes, the 'melting' curves (Fig. 1) represented an average helix stability of the mixture of normal and mutant collagens. It has not been possible to resolve the mutant and normal forms of the $\alpha 1(I)$ -chains by SDS/polyacrylamide-gel electrophoresis for either of these two cases (Bateman *et al.*, 1987a,b). The helix 'melting' temperature of the mutant $\alpha 1(I)$ -chain cannot therefore be determined separately from that of the normal $\alpha 1(I)$ -chain produced by the cells. The determined 'melting' temperatures for the

mutant collagens may therefore represent an overestimate due to lowered yield of the mutant forms. However, the relatively small lowering of the 'melting' temperature resulting from the glycine-to-arginine substitutions may reflect the nature of the mutations, with the glycine-to-cysteine or other mutations previously examined leading to more thermal instability in the collagen helix.

As well as being adjacent to proline residues, both the observed glycine-to-arginine mutations are in regions of sequence that are proline-rich. This may lead to a stabilizing effect and minimize the decrease in thermal stability of the helix. In a comprehensive study of many cases of lethal perinatal osteogenesis imperfecta Bonadio & Byers (1985) concluded that the position of the mutation in the helix, estimated from the localization of over-hydroxylation, was important in determining the degree of thermal instability of the helix. Although the exact mutations were not determined in these cases, they suggested that the closer the mutation was to the C-terminus of the helix, the greater was the loss in helix stability. Our studies suggest that the localization of the mutation did not effect the degree of destabilization, since the two mutants are widely separated at amino acid sequence positions 391 and 667 of the $\alpha 1(I)$ -chain. The local sequence motifs adjoining the mutation may play a more fundamental role in determining the ease of helix denaturation.

However, the lower 'melting' temperature in both these cases suggests that the glycine-to-arginine mutations cause some instability in the structure of the helix. The possible nature and extent of any structural change was studied by molecular modelling with the use of a computer graphics system.

For examination of possible changes in structure introduced by amino acid substitutions, computer graphics and energy-refinement procedures have the advantages over the use of space-filling models (Traub & Steinmann, 1986) of providing values for the shifts in co-ordinates, of providing semi-quantitative comparisons between various structures, and the graphics representations of the structures allow the predicted changes to be seen readily.

To examine the possible changes associated with a glycine-to-arginine substitution, a segment of ten triplets of human collagen centred around the glycine-to-arginine-391 site was constructed by using two chain orders. The $\alpha 2\alpha 1\alpha 1$ configuration has been supported by analysis of staining patterns of segment-long-spacing aggregates (Bender *et al.*, 1982), whereas the $\alpha 1\alpha 2\alpha 1$ configuration has been supported by consideration of side-chain interactions (Traub & Fietzek, 1976).

The refinement data (Fig. 2) indicated that it was possible to accommodate the glycine-to-arginine substitution within the collagen conformation, with little distortion of the packing. Calculation of the differences in $C_{(\alpha)}$ positions between the refined structures of the substituted fragments and the refined structure of the unsubstituted fragment for both chain orders (Fig. 3) indicated that the disruption of the chain was limited to only two or three triplets, with a maximum shift in $C_{(\alpha)}$ positions of about 0.06 nm. In the single-substitution cases, only one of the other two chains showed appreciable movement, this being the one preceding the substituted chain in all cases. In the double substitution, all chains show movement.

Energy calculations for a series of models showed that

Fig. 3. Calculated displacements of $C_{(\alpha)}$ in the glycine-to-arginine-391 mutant

$C_{(\alpha)}$ displacements were calculated from the EREF refined structures of normal and mutant collagens (Fig. 2). (a) $\alpha_2\alpha_1^*\alpha_1$; (b) $\alpha_2\alpha_1\alpha_1^*$; (c) $\alpha_2\alpha_1^*\alpha_1^*$. $\circ$, Normal α_1 -chains; $\bullet$, mutant α_1^* -chains; $\triangle$, α_2 -chains. The amino acid sequences of the regions of the α -chains flanking the mutation are shown.

there were no significant differences between the various models, particularly when the 'fraying' was taken into account. The differences between the bond angle, bond length and bond torsion energies were small among the various models, with the non-bonded energy term being the most important in providing the total energy differences. Examination of covalent bond lengths and angles indicated that there were no significant differences from the mean values. However, the substitutions apparently led to local disruption of the hydrogen-bonding pattern, which would lower the helical stability, consistent with the observed lowered 'melting' temperature.

The modelling studies indicate that the glycine-to-arginine substitution probably leads to only a minor localized disruption of the triple-helix structure. These results are in agreement with the prediction from space-filling models of a glycine-to-cysteine substitution (Traub & Steinmann, 1986). It is likely that the glycine substitutions cause a kinetic effect on the collagen folding that leads to over-modification of the collagen α -chains, rather than the substitution leading to a major disruption of the collagen folding. However, the model-building studies have been performed on pre-folded molecules, and it is possible that energy barriers during folding could lead to an alternative packing with a major conformational change.

The rate of collagen helix formation is dependent on the rate of *cis*- to *trans*-proline isomerization (Bachinger *et al.*, 1978). Since both substitutions have adjacent proline residues, models including *cis*-proline conformations were also examined. Although chain packing without undue strain was possible, these models did not appear to have any advantages when compared with those with proline in the *trans* conformation.

This work was supported, in part, by grants from the National Health and Medical Research Council of Australia and the Royal Children's Hospital Research Foundation.

REFERENCES

- Bachinger, H. P., Bruckner, P., Timpl, R. & Engel, J. (1978) *Eur. J. Biochem.* **90**, 605–613
- Barsh, G. S. & Byers, P. H. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5142–5146
- Barsh, G. S., Roush, C. L., Bonadio, J., Byers, P. H. & Gelinas, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2870–2874
- Bateman, J. F., Mascara, T., Chan, D. & Cole, W. G. (1984) *Biochem. J.* **217**, 103–115
- Bateman, J. F., Chan, D., Mascara, T., Rogers, J. G. & Cole, W. G. (1986) *Biochem. J.* **240**, 699–708
- Bateman, J. F., Mascara, T., Chan, D. & Cole, W. G. (1987a) *J. Biol. Chem.* **262**, 4445–4451
- Bateman, J. F., Chan, D., Walker, I. D., Rogers, J. G. & Cole, W. G. (1987b) *J. Biol. Chem.* **262**, 7021–7027
- Bateman, J. F., Lamande, S., Dahl, H. H.-M., Chan, D. & Cole, W. G. (1988a) *J. Biol. Chem.* **263**, 11627–11630
- Bateman, J. F., Harley, V., Chan, D. & Cole, W. G. (1988b) *Anal. Biochem.* **168**, 171–176
- Bender, E., Silver, F. H., Hayashi, K. & Trelstad, R. L. (1982) *J. Biol. Chem.* **257**, 9653–9657
- Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22**, 5213–5223
- Bonadio, J. & Byers, P. H. (1985) *Nature (London)* **316**, 363–366
- Bornstein, P. & Traub, W. (1979) in *The Proteins* (Neurath, H. & Hill, R. L., eds.), vol. 4, pp. 411–632, Academic Press, New York
- Bruckner, P. & Prockop, D. J. (1981) *Anal. Biochem.* **110**, 360–368
- Byers, P. H., Starman, B. J., Cohn, D. H. & Horwitz, A. L. (1988) *J. Biol. Chem.* **263**, 7855–7861
- Chu, M.-L., Gargiulo, V., Williams, C. J. & Ramirez, F. (1985) *J. Biol. Chem.* **260**, 691–694
- Cohn, D. H., Byers, P. H., Steinmann, B. & Gelinas, R. E. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6045–6047
- de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M.-L., Dickson, L., Weil, D. & Ramirez, F. (1987) *J. Biol. Chem.* **262**, 16032–16036

- Fraser, R. D. B., MacRae, T. M. & Suzuki, E. (1979) *J. Mol. Biol.* **129**, 463–481
- Jones, A. (1982) in *Crystallographic Computing* (Sayre, D., ed.), pp. 303–317, Oxford University Press, Oxford
- Levitt, M. (1974) *J. Mol. Biol.* **82**, 393–420
- Steinmann, B., Rao, V. H., Vogel, B. E., Bruckner, P., Gitzelmann, R. & Byers, P. H. (1984) *J. Biol. Chem.* **259**, 11129–11138
- Traub, W. & Fietzek, P. P. (1976) *FEBS Lett.* **68**, 245–249
- Traub, W. & Steinmann, B. (1986) *FEBS Lett.* **198**, 213–216
- Vogel, B. E., Minor, R. R., Freund, M. & Prockop, D. J. (1987) *J. Biol. Chem.* **262**, 14737–14744
- Willing, M. C., Cohn, D. H., Starman, B. J., Holbrook, K. A., Greenberg, C. R. & Byers, P. H. (1988) *J. Biol. Chem.* **263**, 8398–8404

Received 28 November 1988/6 February 1989; accepted 22 February 1989