Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# On the packing structure of collagen: response to Okuyama *et al.*'s comment on *Microfibrillar structure of type I collagen in situ*

**Joseph. P. R. O. Orgel**

BioCAT and μCoSM Centres: Pritzker Institute of Biomedical Science and Engineering, Illinois Institute of Technology, 3440 S. Dearborn Avenue, Chicago, IL 60616, USA, and CSRRI and Department of Biological, Chemical and Physical Sciences, Illinois Institute of Technology, 3101 S. Dearborn Ave, Chicago, IL 60616, USA

Correspondence e-mail: orgel@iit.edu

Received 15 June 2009
Accepted 15 June 2009

A response is published to the comment by Okuyama *et al.* [(2009) *Acta Cryst.* D**65**, 1007–1008] on *Microfibrillar structure of type I collagen in situ*.

## 1. Introduction

It is disappointing to us that Okuyama *et al.* (2009) chose to largely ignore the most important and substantially supported aspects of our study, namely collagen's molecular packing structure. Instead, by either misunderstanding or through selective attention, they present minor flaws in the coordinate file 1y0f as if they are serious blows to the overall study.

## 2. The first experimentally determined (low-resolution) packing structure of collagen

The purpose of Orgel *et al.* (2006) was to determine the relative spatial arrangement of the five collagen molecules in the unit cell of natively crystalline rat-tail tendon without a dependency on experimentally biased models. This was an essential first step before more detailed structural models could build upon, improve or surpass the initial work. The electron-density map, constructed from experimentally determined phases and observed amplitudes, is clearly and prominently shown and compared with the low-resolution and coordinate based models [see *Supporting Methods* published as supporting information (SI) in Orgel *et al.* (2006)] and $2F_o - F_c$ electron-density map in the paper, and all show good agreement. Hence, at the resolution of the study (5.16 Å axial and 11.1 Å equatorial) we stand by its conclusions.

As a byproduct of the final steps in our attempt to exhaustively test the accuracy of the experimental results (SI Table 3, *Supporting Methods* of Orgel, 2006), the coordinates contained in 1y0f and 1ygv were reached by fitting high-resolution collagen-like peptide structural data into our low-resolution electron-density map, essentially a molecular envelope. This approach is analogous with 'docking' fragments of a high-resolution structure into low-resolution molecular envelopes derived from cryo-electron microscopy or SAXS data (Henderson, 2004; Petoukhov & Svergun, 2007). These represent credible attempts to establish the context in which these detailed, but incomplete, pieces of the puzzle fit together. No-one should confuse the resulting small-scale features of those fragments within the low-resolution structures with those derived by high-resolution single-crystal crystallography or multidimensional NMR. In our case, the low-resolution molecular envelope details the gross arrangement of the collagen molecules, and is not suitable for the study of the specific helical conformation, without further higher resolution equatorial data.

In communicating the coordinate files to the RCSB database, it was our hope that these would provide useful starting points for subsequent studies. At the same time, our caution and transparency in submitting both the 'rigid' (1ygv) and 'relaxed' (1y0f) models and only the C$\alpha$ atoms in both should communicate clearly that the

coordinates are derived from low-resolution data and should be handled appropriately. This point is further made by the fiber diffraction specific annotations within the files and the substantial SI material contributed with the original publication showing what was done and how.

## 3. Specific issues

### 3.1. Completeness

Okuyama *et al.* (2009) misinterpret the information within the 1y0f and 1ygv coordinate files. By mistaking the resolution of the study as isotropic, they assume that 5% represents the completeness of the whole data set. This is despite the fact that in both Orgel *et al.* (2006) and the RCSB coordinate files the resolution is clearly shown to be of anisotropic resolution (5.16 Å axial and 11.1 Å equatorial). Both the publication and coordinate files discuss the number of observed and utilized reflections and the completeness of the refinement data set is actually around 95%.

### 3.2. Chain sequence

The chain sequences were mostly right. The discrepancies between the coordinate file sequence [linked to earlier studies (Orgel *et al.*, 2000) when the sequence at the end of the $\alpha$2 sequence was uncertain] and the updated Uniprot data are a small percentage of the whole molecule and do not effect chain registration *etc.* The comment that nine residues are missing from the C-terminus of the $\alpha$2 sequence seems to be incorrect as we understand the rat $\alpha$2 C-terminal region to be shorter than that of other species and the other telopeptide differences were trivial, but we thank Okuyama *et al.* for bringing these to our attention.

More importantly however, it should be noted that given the resolution of the study and given that only C$\alpha$ positions were reported, these errors are of little or no significance; any mammalian type I collagen sequence would have sufficed for the purpose of model refinement. In our case, after repeating the refinement of the molecular packing model with the corrected sequences, we found no change in the molecular trace, only trivial changes in the specific peptide chain position and no significant change in the *R* factors (or *b*/*q* factors). The small reduction in *R* factor with the corrected sequence indicates that the refinement method is fundamentally sound. We have uploaded the sequence corrected files as referenced under RCSB codes 3hqv and 3hr2.

### 3.3. Chain arrangement

The peptide chain registration, the position of the whole helices relative to the electron density, cross-linking locations and telopeptide conformations were based on the alignment shown in Orgel *et al.* (2000) and Orgel *et al.* (2001), which were referenced in Orgel *et al.* (2006). Here, the heavy atoms in isomorphous derivatives serve as markers of key sequence elements (*e.g.* the Tyr residues in the telopeptides). These features are in no way dependent on the 1ygv or 1y0f models; they were determined independently of them. Rather, the models were constructed to include these experimentally observed features.

### 3.4. Residue occupancy *versus* temperature factor

Okuyama *et al.* raise an important concern, but the regional calculation of temperature factor and lattice distortions were, in fact, discussed in Orgel *et al.* (2006): the temperature factor was assessed as 190 Å$^2$ for the molecule overall. The use of the '*q* factor' was clearly stated in the publication and what its relation is to the overall temperature factor. It does not refer to the residue 'absence' in our study. In the refinement of the coordinate models, we chose to use the *q* factor as a more parsimonious approach because both *q* and *b* factors are approximations and either parameter has roughly equivalent effects at this resolution and we did not refine >3000 parameters at the same time (see SI *Supporting Methods*). What is more, the low-resolution pre-refinement model used only a handful of regional (along the D-period/crystallite *c* axis) temperature factors and the fit of the sequence to the data was good (initial model in SI *Supporting Methods* and SI Fig. 12).

### 3.5. Data-to-parameter ratio

In the *Supporting Methods* to Orgel *et al.* (2006) it is clearly explained that there was an approximately tenfold excess of data to parameters in the refinement of the 1ygv coordinates and how this was achieved. For instance, rather than refining the individual position of 3300 amino-acid residues, the molecular refinement involved

> ... defining 46 regions of the collagen triple helix that are relatively straight, as individual rigid bodies of different lengths, connected by short sections (average length »6 aa) of triple helix that were not constrained, the latter acting as hinges for the refinement of the straight sections. This greatly restrained the degrees of freedom involved in the molecular refinement ...

The final coordinates in 1y0f did not have this degree of constraint, but the molecular trace does not deviate significantly from that of 1ygv. The significance of this last step was that only the stereochemistry of the bonds and the experimental electron density constrained the fit, allowing for some insight into how disassociated the peptide chains might be from the triple-helix in some parts of the molecular packing structure. This is seen in the varying diameter of the electron-density 'tubes' showing the outline of the collagen molecules.

### 3.6. The collagen structure, a model to be handled with care

The coordinates we have contributed currently represent the best known alignment of collagen sequence to the three-dimensional packing structure of collagen molecules *in situ*, despite their known deficiencies. They are not, and were never intended to be a direct contribution to our understanding of collagen's triple-helical symmetry as Okuyama *et al.* appear to believe. However, we fully agree with Okuyama *et al.*'s conclusion that the coordinates provided in Orgel *et al.* (2006) should be used with care and with due consideration of their intrinsic limitations.

## References


Henderson, R. (2004). *Q. Rev. Biophys.* **37**, 3–13.

Okuyama, K., Bachingera, H. P., Mizuno, K., Boudko, S., Engel, J., Berisio, R. & Vitagliano, L. (2009). *Acta Cryst.* D**65**, 1007–1008.

Orgel, J. P., Irving, T. C., Miller, A. & Wess, T. J. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 9001–9005.

Orgel, J. P., Miller, A., Irving, T. C., Fischetti, R. F., Hammersley, A. P. & Wess, T. J. (2001). *Structure*, **9**, 1061–1069.

Orgel, J., Wess, T. & Miller, A. (2000). *Structure Fold. Des.* **8**, 137–142.

Petoukhov, M. & Svergun, D. (2007). *Curr Op. Struct. Biol.* **17**, 562–571.