Pramāṇa, Vol. 9, No. 4, October 1977, pp. 339–347, © Printed in India.

Fourier transform of the collagen triple-helical structure and its significance*

M BANSAL
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012

MS received 4 May 1977; revised 14 July 1977

Abstract. The Fourier transforms of the collagen molecular structure have been calculated taking into consideration various side chain atoms, as well as the presence of bound water molecules. There is no significant change in the calculated intensity distribution on including the side chain atoms of non-imino-acid residues. Taking into account the presence of about two bound water molecules per tripeptide unit, the agreement with the observed x-ray pattern is slightly improved. Fourier transforms have also been calculated for the detailed molecular geometries proposed from other laboratories. It is found that there are no major differences between them, as compared to our structure, either in the positions of peak intensity or in the intensity distribution. Hence it is not possible to judge the relative merits of the various molecular geometries for the collagen triple helix from a comparison of the calculated transforms with the meagre data available from its x-ray fibre pattern. It is also concluded that the collagen molecular structure should be regarded as a somewhat flexible chain structure, capable of adapting itself to the requirements of the different side groups which occur in each local region.

Keywords. Collagen molecular structure; Fourier transforms; water-bound collagen triple helix.

1. Introduction

The molecular conformation of collagen has been determined primarily from an interpretation of its high-angle x-ray diffraction pattern, which was first characterised by Astbury (1938) as being different from the α-helical and the β-pleated sheet patterns. The x-ray pattern of an unstretched collagen fibre is rather diffuse and poor in detail. Its principal features are a strong meridional arc at about 2·9 Å and weaker arcs at 4·0 Å and 9·5 Å. There are also strong equatorial reflections with spacings corresponding to approximately 6 Å and 12 Å (which depend on the humidity), and a diffuse distribution of intensity around 4·5 Å, mainly near the equator. This pattern is greatly improved in orientation and detail if the fibre is kept stretched during the x-ray exposure (Cowan *et al* 1953; Ramachandran and Ambady 1954).

The general distribution of intensity in the collagen pattern is of the type given by helical structures. An interpretation of this pattern in terms of the helix diffraction theory of Cochran *et al* (1952) led to the conclusion that the collagen helix has a unit height of approximately 3 Å, the number of units per turn being close to 10/3, which corresponds to a unit twist of 108°. A more accurate measurement later of the

*Contribution No. 102 from the Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012

spacings of both stretched and unstretched collagen fibres indicated that the helical parameters for the moist unstretched fibres are: unit height=2·91 ± 0·01 Å and unit twist = 110° ± 2° (Lakshmanan *et al* 1962; Ramachandran 1967).

The variation of intensity along the layer lines also contains much quantitative information, but it is not possible to directly interpret this into structural details. Hence the problem was approached in an indirect fashion. A detailed structural model was built based on various stereochemical properties of the polypeptide chains and the intensity distribution expected for it was calculated using the helix diffraction theory of Cochran *et al* (1952) or its generalisation, as developed by Ramachandran (1960). The calculated intensity distribution was then compared with the observed one (Bradbury *et al* 1958; Rich and Crick 1961; Lakshmanan *et al* 1962; Ramachandran 1967; Yonath and Traub 1969).

In most of the calculations done so far, the observed intensities have been compared with the intensities calculated for the repeating polypeptide sequences of the type (glycine-alanine-alanine)$_n$ or (glycine-imino acid-imino acid)$_n$. Rich and Crick (1961) have carried out a comparison using the intensities calculated for the above two types of sequences added up in the ratio 2:1, to correspond to the known imino acid composition of collagen. Such a calculation assumes that the imino-acid residues are clumped together in certain regions of the collagen molecule, while they are now known to be distributed quite randomly along the whole length of the collagen chains (Piez 1976). Also, though most of the diffracted intensity is undoubtedly due to the polypeptide backbone and the atoms of the rigid pyrrolidine rings, the atoms of the side chains of other amino-acid residues, as well as bound water molecules, must also be included for an accurate comparison. Therefore, the effect of introducing the various side chain atoms, on the calculated intensity distribution for the backbone structure as proposed by the Madras group (Ramachandran 1967), has been studied in detail. The influence of bound water molecules has also been looked into. The intensity distributions for the detailed atomic coordinates as proposed by other groups of workers (Rich and Crick 1961; Yonath and Traub 1969) have also been calculated, and compared with that for the Madras structure, as well as with the observed diffraction pattern.

2. Details of the calculations

The intensities for the various layer lines have been calculated using the formulae given by Ramachandran and Venkatachalam (1970) for the cylindrically averaged Fourier transform of a helical structure. The reciprocal spacings (Z) of the layers on which diffraction occurs are given by

$$Z = Z_0(m \pm qt)$$

where q and m are integers, and $Z_0 = 1/h$, h and t being the helical parameters, unit height and unit twist respectively (Ramachandran 1960).

Bessel functions corresponding to the two lowest values for q were used in the calculations for each layer line. The atomic scattering factors were calculated using Moore's relation (Moore 1963).

The distribution of amino-acid residues among the locations 1, 2 and 3 of the triplets-Gly-R_2-R_3- in the complete α1(I) chain of rat-calf skin is now available (Piez 1976). It is found that only the amino acids Gly, Ala, Pro and Hyp occur in substantial amounts, with glycine being confined to the first position. Hence the side chain atoms of these amino acids were explicitly taken into account. Since, except for a single glycyl residue at location 2 (and none at 3) all other residues have side chains containing at least a C^β atom, this atom was given a weight unity. The imino acids proline and hydroxyproline constitute approximately one-third the total number of residues at locations 2 and 3 respectively, hence the side chain atoms of the pyrrolidine rings were weighted accordingly.

Among the amino-acid residues in locations 2 and 3, 50% contain side chain atoms beyond C^β, while about 40% contain atoms beyond C^γ. These C^γ and C^δ atoms can take up various possible orientations about the bonds $C^\alpha - C^\beta$ and $C^\beta - C^\gamma$ respectively. Stereochemical studies have shown that, at both the locations 2 and 3, the C^γ atom can occur with a considerable degree of freedom in the staggered positions II and III, corresponding to the dihedral χ^1 being 180° and −60° respectively, while it is appreciably constrained if it occurs at the position I (Bansal 1977). Hence only the positions II and III have been considered for the C^γ atom. Since the C^δ atom can occur freely only at the position II (corresponding to the C^δ atom being *trans* to the atom C^α, about the bond C^β—C^γ) only this position was considered for the C^δ atom. Side chain atoms beyond C^δ are relatively rare, and since they can take up several possible orientations, they have not been considered. Similarly, side chains containing atoms other than C^γ and C^δ i.e. O^γ, S^δ, etc. (as in serine, threonine or methionine) are present only in very small numbers and so were also not considered.

Firmly bound water molecules were taken to be present in the positions specified for them by Ramachandran and Chandrasekaran (1968) in the 1-bonded model from Madras. In this water-bridged structure, there are two water molecules per tripeptide, but in the regions where a prolyl residue occurs at location 2, or a C^γ atom occurs at position II in the side group R_3, only one of these water molecules can be firmly bound to the polypeptide chains. Hence the water molecules were also suitably weighted.

The experimental data of Yonath and Traub (1969) for unstretched sheep submucosa collagen were used for comparison with the calculated intensity distribution.

3. Results

3.1. *Effect of including various side chain atoms*

In order to study the effect of the presence of various side chain atoms on the calculated intensities, the Fourier transform for the collagen structure was calculated taking into consideration the following side chains at locations 2 and 3 in the repeating triplet sequence $(\text{Gly-}R_2\text{-}R_3)_n$:

(a) only C^β atoms,
(b) C^β atoms, C^γ and C^δ atoms of the proline ring at location 2 and C^γ, $(OH)^\gamma$, C^δ of the hydroxyproline residue at location 3
(c) as in (b) but with C^γ atoms of non-imino-acid residues also included,

(d) as in (c), with C^{δ} atoms in position II for each of the C^{γ} atoms considered above, also taken into consideration.

The coordinates of the backbone atoms and the hydroxyproline side chain at location 3 were taken from Bansal *et al* (1975).

The pyrrolidine ring at location 2 was assumed to have the geometry given in Ramachandran and Bansal (1976), while the other side chain atoms were fixed in staggered orientations as described in section 2.

The Fourier transforms calculated, taking into considerations the various side chain atoms fixed in appropriate positions, are given in figure 1.

3.2. *Effect of including bound water molecules*

Taking into account the presence of firmly bound water molecules, there are some noticeable differences in the calculated intensity distribution on some of the layer

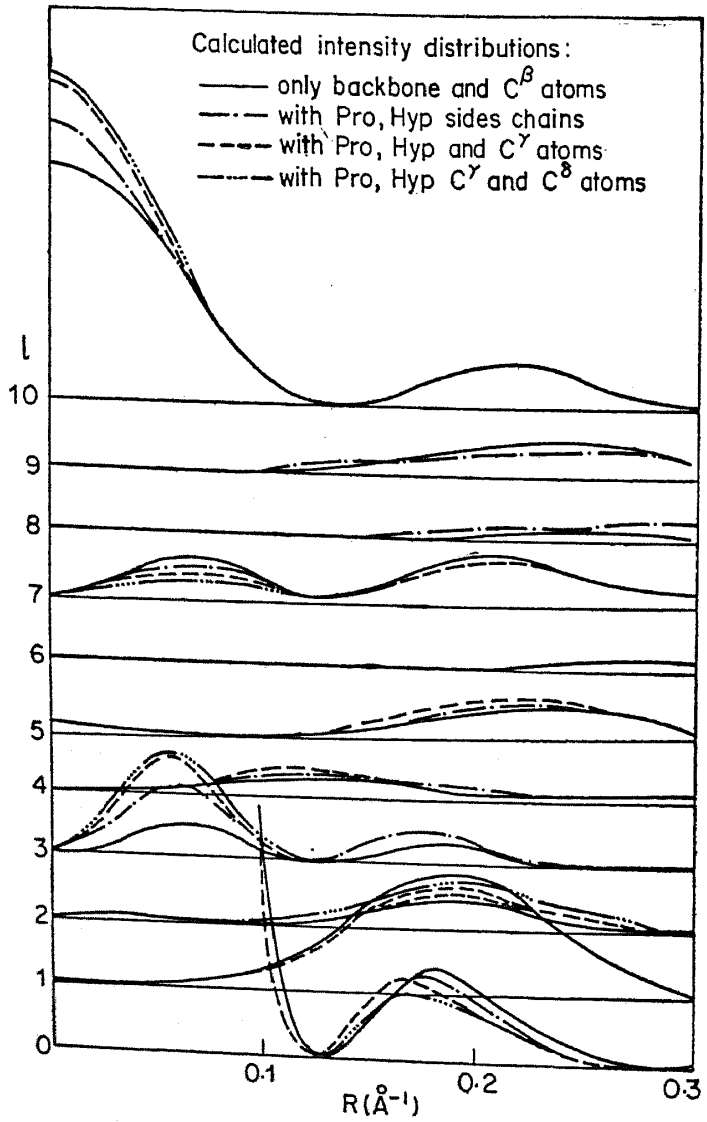

Figure 1. Calculated intensity distribution along the layer lines for the collagen triple-helical structure. The various side chain atoms have also been taken into consideration.

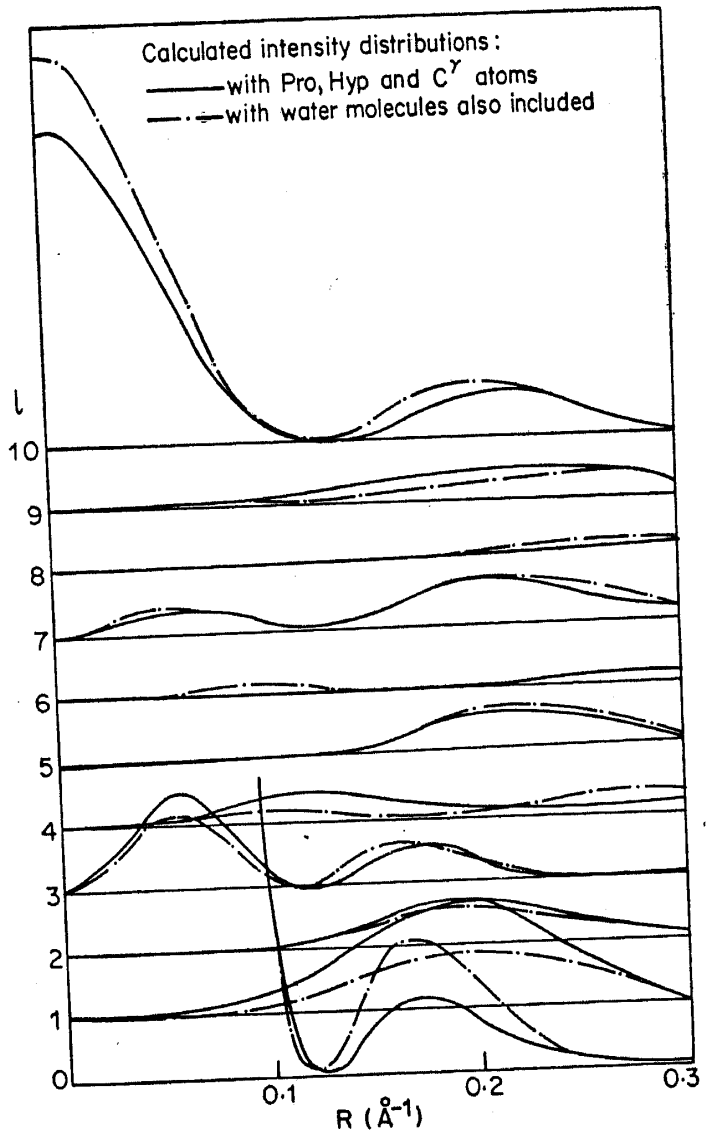

Figure 2. Intensity distribution for the collagen structure showing the effect of including bound water molecules in the calculations.

lines. The intensity maxima on the equator (corresponding to $R \simeq 0.18$ Å^{-1}) and the meridional arc on the tenth layer are considerably more intense while that on the 1st, 3rd and 4th layers are now much less intense as shown in figure 2.

3.3. *Comparison of the calculated transforms for different molecular geometries*

The intensity distribution along various layer lines has also been calculated using the detailed atomic coordinates as given by other groups of workers (Rich and Crick 1961; Yonath and Traub 1969). The side chain atoms of the pyrolidine rings were taken to be present in the positions given by these workers, while the C^{γ} and C^{δ} atoms of non-imino-acid side groups were fixed in the positions described above.

As in the case of our model structure, for the Yonath-Traub and Rich-Crick (collagen II) molecular geometries also, there were no significant differences on including the C^{γ} and C^{δ} atoms of side chains other than those of proline and hydroxy-

proline. For purposes of comparison, the calculated intensity distribution for the different molecular geometries of the backbone, and including the side chain atoms of imino acids and the C^{β} and C^{γ} atoms of other side groups, have been plotted in figure 3 on the same scale. It is clearly seen from this figure that the calculated intensities for the various molecular geometries are not appreciably different—in spite of the fact that the positions of the individual atoms differ considerably. The positions of the intensity maxima as well as the relative heights of the maxima are quite similar. Hence the agreement with the observed data is almost equally good for the different molecular geometries, except that, for the Rich-Crick model the meridional arc on the 10th layer is relatively weak as compared to the intensity maxima on the 3rd and 7th layers.

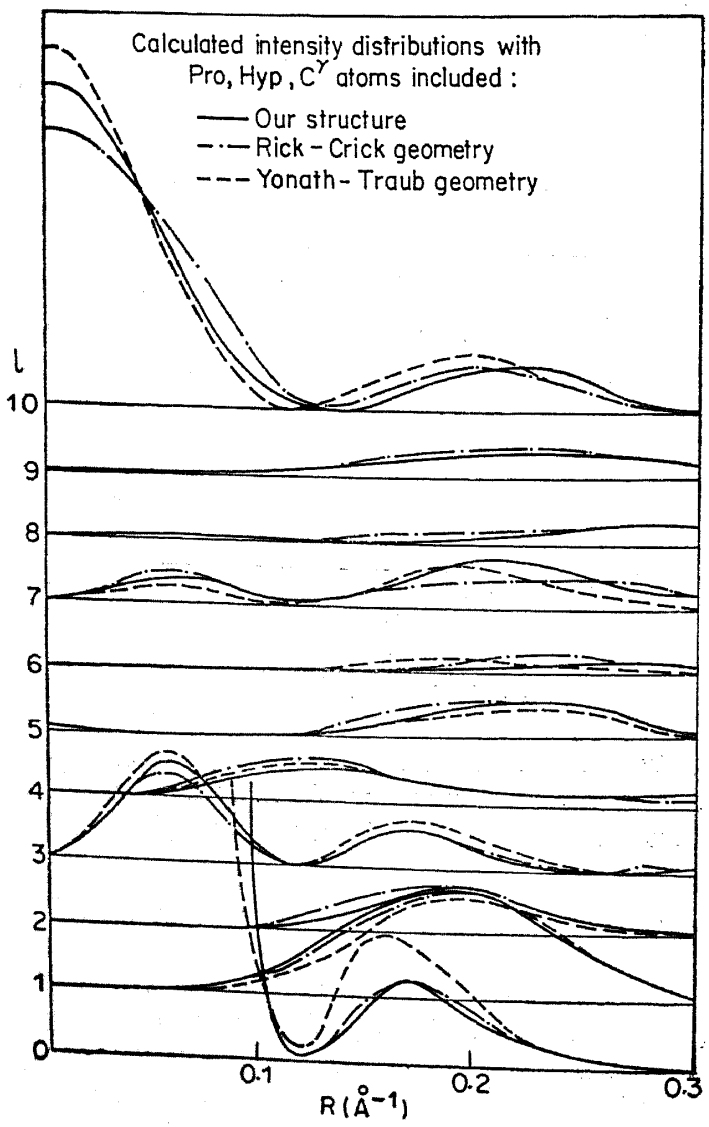

Figure 3. Calculated intensity distribution for the detailed molecular geometries of the triple helix, as proposed by different groups of workers. All the side chain atoms of the imino acids and C^{γ} atoms of other residues have been included in each case.

4. Discussion of the results

It is obvious from the results of the preceding section that the data from the observed x-ray pattern of collagen are too meagre for a meaningful quantitative comparison with the calculated intensity distribution to be carried out. Hence only a qualitative comparison has been attempted here.

As shown in figure 1, there is no significant change in the calculated intensity on including the side chain atoms of non-imino-acid residues. The agreement with the observed data is fairly good on the equator, the 7th and the 10th layer lines. There is a slight discrepancy in the position of the first maximum on the 3rd layer line—it is closer to the meridian than the observed reflection. However, this reflection has a sharpness which suggests that the Fourier transform is sampled at this point owing to partial crystallinity arising from packing of the molecules (Rich and Crick 1961). It is also found that the observed diffuse reflection on the fourth layer occurs at the same distance from the meridian, in agreement with the calculated intensity maxima on this layer line.

Taking into consideration the presence of water molecules, hydrogen-bonded to the polypeptide chains in the manner proposed by Ramachandran and Chandrasekaran (1968), the agreement between the calculated and the observed intensity on the equator and the 10th layer line is slightly improved. However, the calculated intensity for the 103 reflection is very much reduced relative to the intensity of the meridional arc on the 10th layer.

Yonath and Traub (1969), in their proposed model for the synthetic polymer (Gly-Pro-Pro)$_n$, worked out possible locations for water molecules so as to get a good agreement between the observed x-ray pattern and the calculated intensity transform. However, water molecules cannot be accommodated in the positions specified by these workers, if bulky side groups occur at locations 2 and 3 in the collagen triplet sequence. The orientations of these water molecules are also not consistent with the infrared dichroism data of Suzuki and Fraser (1974). Hence, the positions of bound water molecules in collagen need to be examined in greater detail so as to simultaneously satisfy the observed infrared dichroism and x-ray data, as well as being stereochemically satisfactory. The water molecules in the positions specified by Ramachandran and Chandrasekaran (1968) satisfy all these criteria reasonably well.

The calculated intensity transforms, for the detailed atomic coordinates given by different workers, are found to be quite similar and it is not possible to discriminate between them by a comparison with the observed x-ray pattern. This is particularly so, because the two discrete and most intense reflections (i.e. the first order reflections on the equator and the 3rd layer) cannot be directly compared with the calculated intensities. The reason for the observed discrepancy in the third layer reflection has already been given above. On the equator, as can be seen from figures 1 to 3 the calculated intensity curve rises very sharply when the value of R is below 0·1 Å^{-1} and hence the calculated intensity for the 100 reflection (corresponding to $R \simeq 0{\cdot}088$ Å^{-1}), is extremely sensitive to the value of R. Since the spacings of the discrete reflections on the equator are highly dependent on the water content of the collagen sample, it is very difficult to measure them accurately and also allow for the water content in the inter-protofibrillar space. Therefore it is not possible to make a meaningful comparison of the calculated and the observed intensity for these reflections.

It may also be mentioned that recently single crystals have been obtained for the

collagen model compound $(Pro\text{-}Pro\text{-}Gly)_{10}$ and a detailed structure analysis has been carried out by a method of model-building and two-dimensional least squares refinement (Okuyama *et al* 1976). The effect of including amorphous water between the chains of the triple-helical structure was also considered. However even from this study it has not been possible to arrive at a unique structure for the decatripeptide.

5. Conclusion

From the above calculations it is seen that the side chain atoms of non-imino-acid residues, even if they are assumed to have certain fixed orientations with respect to the backbone, do not contribute significantly to the calculated intensity distribution. Hence the major contributions are from the atoms of the polypeptide backbones and the rigid pyrolidine rings.

The presence of bound water molecules considerably alters the calculated intensity transforms, however the agreement with the observed x-ray pattern is only slightly improved.

Finally, it is concluded that, it is not possible to make any definite deductions about the relative merits of the detailed molecular models of collagen from a comparison of the intensity data for the prominent layer lines (with $l=0$, 3, 7 and 10) with the calculated intensity distributions, as has been done by some previous workers (Yonath and Traub 1969).

Even a quantitative comparison of the available experimental data with the calculated transforms gives only a qualitative indication of the general correctness of the molecular model considered, since the calculated intensity distribution is found to be very similar, even for molecular models with fairly large differences in the atomic positions. In view of this finding, and the observation that most of the spots in the x-ray pattern of collagen are quite diffuse, it is concluded that it is not correct to regard the collagen molecule as being a rigidly symmetrical structure with an exact set of parameters throughout the whole of its length. It is quite possible that the actual structure varies from one region to another, in order to satisfy the stereochemical and hydrogen-bonding requirements of different groups of amino acids which occur together in each particular region.

Acknowledgements

The Author is very grateful to G N Ramachandran for advice and discussions and the CSIR, India, for scholarship. The work was supported by SERC (DST), India.

References

Astbury W T 1938 *Trans. Faraday Soc.* **34** 378
Bansal M 1977 *Int. J. Pept. Prot. Res.* **9** 224
Bansal M, Ramakrishnan C and Ramachandran G N 1975 *Proc. Indian Acad. Sci.* **A82** 152
Bradbury E M, Burge R E, Randall J T and Wilkinson G R 1958 *Disc. Faraday Soc.* **25** 173
Cochran W, Crick F H C and Vand V 1952 *Acta Cryst.* **5** 581

Cowan P M, North A C T and Randall J T 1953 in *Nature and Structure of Collagen* ed. J T Randall (London: Butterworths) p. 241

Lakshmanan B R, Ramakrishnan C, Sasisekharan V and Thathachari Y T 1962 in *Collagen* ed N Ramanathan (New York: Wiley-Interscience) p 117

Moore F H 1963 *Acta Cryst.* **16** 1169

Okuyama K, Tanaka N, Ashida T and Kakudo M 1976 *Bull. Chem. Soc. Japan* **49** 1805

Piez K A 1976 in *Biochemistry of Collagen* eds G N Ramachandran and A H Reddi (New York: Plenum Press) Ch. 1

Ramachandran G N 1960 *Proc. Indian Acad. Sci.* **A52** 240

Ramachandran G N 1967 in *Treatise on Collagen* Vol. 1 ed. G N Ramachandran (New York: Academic Press) Chapter 3

Ramachandran G N and Ambady G K 1954 *Curr. Sci.* **23** 349

Ramachandran G N and Chandrasekaran R 1968 *Biopolymers* **6** 1649

Ramachandran G N and Venkatachalam C M 1970 *Z. Kristallogr.* **132** 152

Ramachandran G N and Bansal M 1976 *Curr. Sci.* **45** 647

Rich A and Crick F H C 1961 *J. Mol. Biol.* **3** 483

Suzuki E and Fraser R D B 1974 in *Peptides, Polypetides and Proteins* eds E R Blout, F A Bovey, M Goodman and N Lotan (New Yok: Wiley-Interscience) p. 449

Yonath A and Traub W 1969 *J. Mol. Biol.* **43** 461