Secondary and tertiary structural foldings in tRNA

A diagonal plot analysis using the blocked nucleotide scheme

R. MALATHI and N. YATHINDRA*

Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600 025, India

(Received 15 April 1982/Accepted 26 May 1982)

A distance plot obtained using the blocked nucleotide concept, which regards the repeating nucleotide moieties to be made up of two blocks of nearly equal magnitude, has permitted us to visualize the polynucleotide backbone folding in yeast tRNA^{Phe}. The plot clearly manifests medium- and long-range tertiary interactions involving various structural domains. Apart from the well known TΨ-D loop interactions, other long-range interactions associated with the variable loop as well as the D loop are explicitly seen. Most importantly, the plot reveals an approximate two-fold symmetry in the molecule between the domains related to the tertiary interactions in addition to the symmetry between long helical domains. The different patterns on the plot are interpreted in terms of helix–helix, loop–helix and loop–loop interactions.

Distance plots, wherein the interatomic distances of each atom from every other similar atom along a polymer chain are plotted, may be used to describe backbone folding (Phillips, 1970). This procedure has been extensively used for comparison of folding schemes in proteins (Ooi & Nishikawa, 1973; Rossman & Lilijas, 1974; Kuntz, 1975). Such a plot would be relevant for a polynucleotide chain such as tRNA, which is made up of helices, loops and bends, but was not tenable until recently since a distance plot requires the length of the repeating moiety to be nearly constant and independent of the bond rotations which provide the necessary flexibility for the three-dimensional folding. The length of the nucleotide repeat, namely the adjacent phosphate–phosphate separation, although independent of the phosphodiester P–O bond rotations (ω' , ω), is found to vary in tRNA (Quigley *et al.*, 1975; Jack *et al.*, 1976; Sundaralingam *et al.*, 1976) because of the occurrence of different sugar puckers (ψ') and C_{4'}–C_{5'} bond conformations (ψ). The preferred *trans* character of the two C–O bonds in the nucleotide repeat allows us to overcome this by regarding the nucleotide as being made up of two nearly equal blocks (Malathi & Yathindra, 1980, 1981, 1982) as shown in Fig. 1. This further allows a polynucleotide chain to be considered in terms of virtual bonds joining P and C_{4'} atoms (Malathi &

Yathindra, 1980, 1981, 1982; Olson, 1980). Consideration of the polynucleotide backbone in terms of these blocks, which are also independent of the pairs of rotations (ω' , ω) and (ψ' , ψ), permits the construction of a distance plot for polynucleotides. We describe below such a plot obtained for yeast

Fig. 1. Section of a polynucleotide chain showing the two blocks spanning the atoms P–C_{4'} on the 5' side and C_{4'}–P on the 3' side of the repeating nucleotides

The preferred *trans* character of the two C–O bonds renders the two blocks to be of nearly equal magnitude (0.39 nm) and independent of the rotational pairs (ω' , ω) and (ψ' , ψ) which are the principal determinants of polynucleotide chain folding.

Abbreviation used: D, 5,6-dihydrouridine.

* To whom correspondence and reprint requests should be addressed.

tRNA^{Phe}. The plot reveals succinctly the single- and double-stranded helical regions together with a number of patterns due to medium- and long-range tertiary interactions between substructures formed out of helices and loops.

General features of a diagonal plot

A diagonal plot has a mirror symmetry along its diagonal and hence only one-half of it suffices for discussion. Self-interactions between like atoms are represented along the diagonal. Different patterns obtained on the plot represent different structural domains characterized by varieties of short-range and long-range interactions. Single-stranded helical structures representing short-range interactions are indicated by contours close to and parallel to the diagonal. Double-helical regions consisting of anti-parallel segments (3' and 5' strand interactions) can be recognized by contours perpendicular to the diagonal, reminiscent of anti-parallel β -sheets in proteins. Tertiary interactions between residues far from each other in the primary structure appear as patterns far away from the diagonal. A pattern at the origin represents interactions between the two termini of the molecule, suggestive of its folded nature.

Diagonal plot for yeast tRNA^{Phe}

Consideration of the nucleotide in terms of two blocks of equal magnitude allows the atoms P and C_{4'} to be treated alike. This enabled us to obtain a distance map for yeast tRNA^{Phe} with the distances between each of these atoms in a residue from every similar atom along the chain plotted as a square matrix, the chain sequence numbers plotted horizontally and vertically. The co-ordinates are taken from Ladner *et al.* (1977). The various regions of the tRNA backbone are marked along the abscissa to facilitate discussion. Equivalence contours of 1.0, 1.5 and 2.0 nm are shown. The limit is set at 2.0 nm, corresponding to the diameter of an RNA double helix.

Four major domains (marked 1, 2, 3 and 4 in Fig. 2) perpendicular to the diagonal are readily apparent, and are indicative of the four structural domains corresponding to the amino acid (AA), dihydrouridine (D), pseudouridine (T Ψ) and anticodon (AC) arms of the molecule. Correlation of this with the clover-leaf secondary structure representation is obvious. Although the presence of four domains is identified by four distinct patterns on the plot, a detailed study of the shape and nature of the contours within each domain would indicate the way in which the helices and loops and the substructures formed out of them interact to form the domain. A helix-helix interaction would be represented by a

regular pattern, ordered-disordered or disordered-disordered by a distorted pattern. If the interaction is of double-helical type, i.e., residue i of the 5' strand base-paired to residue j of the 3' strand, it is found that the interactions of residues $j \pm 1$ with i and $i \pm 1$ with j are similar to that of i with j itself. This results in a narrow strip of approx. 1.8 ± 0.1 nm contours perpendicular to the diagonal.

A distinct but recurring feature in the plot is that the contours originate as parallel to the diagonal, form closed patterns perpendicular to the diagonal and then retrace their paths to become again parallel to the diagonal. This is due to the occurrence of helical followed by irregular loop structures and vice versa along the chain. The above feature in the plot is reminiscent of the pattern seen in some proteins where α -helix is followed by β -sheet and vice versa (Rossman & Lilijias, 1974).

Inspection of the contours close to and parallel to the diagonal shows marked deviations at residues 10, 34 and 56, indicative of sharp turns known to occur in yeast tRNA^{Phe} (Quigley *et al.*, 1975; Jack *et al.*, 1976; Sundaralingam *et al.*, 1976). A similar deviation found at residue 20 is due to a more gradual turn. These deviations are indicated by horizontal and vertical arrows. The bends occurring near residues 48 and 59 are also suggested by noticeable deviations at these positions on the plot.

A significant observation in Fig. 2 is a small closed pattern (marked 2a) occurring close to and perpendicular to the diagonal. This is suggestive of a loop involving only few residues which are brought into proximity due to a sharp turn at residue 10. This loop, generally referred to as the P10 loop, would not have manifested itself so prominently on a distance plot obtained using P-P separations.

The domain marked 1 at the origin represents the interaction between the 3' and 5' ends of the tRNA molecule, illustrative of its intricate folding. The domain may be approximately divided into three regions with a central narrow strip (2.0 nm contour) representative of the base-paired AA stem. The patterns (1.5 nm contour) just above and below this represent the interaction of the 5' strand of the AA stem with a few residues of the 3' strand of the T Ψ stem (i.e. the residues preceding the 3' strand of the AA stem) and the terminal CCA (i.e. the residues following the 3' strand of the AA stem) respectively. The regularity of these contours suggest that the interacting segments involved in this domain form a section of ordered structure, as is indeed the case (Quigley *et al.*, 1975; Jack *et al.*, 1976; Sundaralingam *et al.*, 1976). A small distorted triangular pattern (1a) occurring in continuity with the upper region in domain 1 is due to the interaction of the T Ψ loop and the 3' strand of the T Ψ stem with the P10 loop. Similarly a small closed pattern (1b) occurs in proximity to the lower region of domain 1

Fig. 2. Diagonal plot for yeast $tRNA^{Phe}$ (monoclinic)

The sequence numbers of the residues from 5' → 3' are represented along the abscissa and from 3' → 5' along the ordinate. Equivalence contours of 1.0 nm (—), 1.5 nm (·-·-·) and 2.0 nm (---) are shown. Contours of 0.6 nm (dark) and 0.8 nm (hatched) are also indicated. The patterns representing various structural domains, the interactions between them and the approximate two-fold symmetry relating them are indicated. The manifestation of the P10 loop and the long-range tertiary interactions associated with the D, TΨ and variable loops are some of the highlights in the plot. The plot thus serves as a good two-dimensional representation of the three-dimensional structure of the molecule.

due to the interaction between the 5' strand of the D stem and a few residues in the D loop with the 3' strand of the AA stem. Both these correspond to long-range tertiary interactions.

Features seen in region 1 recur in regions 2, 3 and 4, particularly in having the central strip indicative of double-helical base-pairing in these regions. The patterns seen above the central region in domains 2, 3 and 4 are distorted to different extents, illustrative of residues in loop conformation preceding the corresponding 3' strand of double-helical stems. On the other hand, the regularity of contours below the double-helical strip in regions 2 and 4 is suggestive of residues occurring in ordered helical conformation following the 3' strands of the D and TΨ stems respectively. A similar region in domain 3 is not as regular, suggesting that the residues following the 3' strand of the AC stem (variable loop) are not highly ordered. A distorted triangular pattern (3a) is seen towards the end of domain 3, further off from the diagonal. This represents medium-range inter-

actions between the 3' strand of the D stem with the continuous stretch of polynucleotide backbone from the 3' strand of the AC stem to the variable loop. Residues in the AC loop are not involved in interaction with any other domains, indicative of their isolation from the rest of the molecule.

The two major grooves are indicated by G1 and G2. The appearance of 0.6 nm contours (darkly shaded portion) in the region marked G2 suggests that this groove is narrower compared with G1 and has been interpreted (Robertus *et al.*, 1974) earlier as being due to a disjointed helix formed out of the 3' D and 5' AC stems.

One of the most significant features of the plot is the occurrence of two similar patterns (5 and 6) far away from the diagonal. While the domain 5 represents the well known TΨ-D loop interaction, domain 6 arises as a result of interactions between the variable loop with the flanking P10 loop and the D arm. Another similar pattern (4a) but of relatively small size is seen in the upper region of domain 4 due

to the interaction of the variable loop with a few residues of the T Ψ loop. The similarity in the shape of these patterns reveals the interaction between similar structural motifs. A careful correlation of the boundary of the domains with the number of residues involved indicates that the tertiary structural patterns mentioned above occur as a result of interactions between a loop and a substructure formed by a loop and a few residues of either the 3' or 5' helical strand, although the predominant interactions are between loops. The size of the area reflects the number of residues involved in such interactions. Thus all the six patterns, i.e. 5, 6, 1a, 1b, 3a and 4a, typify similar tertiary interactions.

It is noteworthy that domains (1a+5) and (3a+6) are related by an approximate two-fold symmetry perpendicular to the plane of Fig. 2 as indicated. This symmetry can be extended approximately to the entire stretch comprising 3 and 6 and 1 (excluding 1b) and 5. The symmetry now relates not only the tertiary interactions due to the substructures mentioned above but also the secondary structural domains, namely the long helices of the AA and T Ψ stems and the D and AC stems. The significance of the latter pseudo-two-fold symmetry in the recognition of tRNA by aminoacyl-tRNA synthetase has been noted earlier (Kim, 1975).

Contours of 0.6 nm (dark) and 0.8 nm (hatched) are also shown in the plot, for they may represent possible metal and solvent binding sites. Such regions occurring close to the diagonal, i.e. near 8–9, 49–50, 58–59 and in the major groove G2, have indeed been identified (Jack *et al.*, 1977; Stout *et al.*, 1978) with different metal binding sites. Similar regions resulting due to tertiary interactions are also indicated in the plot.

Conclusions

The diagonal plot based on the blocked nucleotide scheme offers a complete description of the three-dimensional folding of the polynucleotide chain in yeast tRNA^{Phe}. The various domains, occurring as a result of short-, medium- and long-range interactions between different structural entities such as loops, bends and helices are brought out. The manifestation of the P10 loop, the approximate two-fold symmetry and some of the long-range interactions associated with the variable loop as well

as the D loop are some of the major features explicit in the plot. The diagonal plot offers a convenient procedure for comparing the structures of different tRNAs. For instance, on a diagonal plot obtained for tRNA^{Ser} or tRNA^{Tyr}, which has a long variable loop characterized by a classical helical stem preceded by a loop, a distinct pattern would occur between domains 3 and 4 with other modifications due to tertiary interactions. A comparative study may aid in the better understanding of the functional role of the variable loop.

R.M. thanks the Department of Atomic Energy, Government of India, for a fellowship. This paper is contribution no. 582 from the Department of Crystallography and Biophysics, University of Madras.

References

- Jack, A., Ladner, J. E. & Klug, A. (1976) *J. Mol. Biol.* **108**, 619–650
- Jack, A., Ladner, J. E., Rhodes, D., Brown, R. S. & Klug, A. (1977) *J. Mol. Biol.* **111**, 315–328
- Kim, S. H. (1975) *Nature (London)* **256**, 679–680
- Kuntz, T. D. (1975) *J. Am. Chem. Soc.* **97**, 4362–4366
- Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. C. & Klug, A. (1977) in *Labquip*, Nicholson Molecular Model Kit, England
- Malathi, R. & Yathindra, N. (1980) *Curr. Sci.* **49**, 803–807
- Malathi, R. & Yathindra, N. (1982) *Int. J. Quantum Chem.* **20**, 241–257
- Malathi, R. & Yathindra, N. (1982) *Int. J. Biol. Macromol.* **4**, 18–24
- Olson, W. K. (1980) *Macromolecules* **13**, 721–726
- Ooi, T. & Nishikawa, K. (1973) in *Conformation of Biological Molecules and Polymers* (Bergmann, E. D. & Pullman, B., eds.), pp. 173–187, Academic Press, New York
- Phillips, D. C. (1970) *Biochem. Soc. Symp.* **30**, 11–28
- Quigley, G. J., Seeman, N. C., Wang, A. H. J., Suddath, F. L. & Rich, A. (1975) *Nucleic Acids Res.* **2**, 2329–2341
- Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974) *Nature (London)* **250**, 546–551
- Rossmann, M. G. & Lilijas, A. (1974) *J. Mol. Biol.* **85**, 177–181
- Stout, C. D., Mizuno, H., Rao, S. T., Swaminathan, P., Rubin, J., Brennan, T. & Sundaralingam, M. (1978) *Acta Crystallogr. Ser. B* **34**, 1529–1544
- Sundaralingam, M., Mizuno, H., Stout, C. D., Rao, S. T., Lieberman, M. & Yathindra, N. (1976) *Nucleic Acids Res.* **3**, 2471–2484