Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of the dimerization-initiation site of genomic HIV-1 RNA: preliminary crystallographic results

M. Yusupov,† P. Walter, R. Marquet, C. Ehresmann, B. Ehresmann and P. Dumas*

Institut de Biologie Moléculaire et Cellulaire du CNRS, UPR900215, rue René Descartes, F67084 Strasbourg CEDEX, France

† Present address: Center for Molecular Biology of RNA, Sinsheimer Laboratories, University of California, Santa Cruz, CA 95064, USA.

Correspondence e-mail: dumas@ibmc.u-strasbg.fr

Received 20 January 1998
Accepted 19 June 1998

The genomic RNA of all retroviruses is encapsidated in virions as a dimer of single-stranded chains held together near their 5′-end. For HIV-1, the initial site of dimerization has been shown to be a hairpin with a nine-residue loop containing a self-complementary sequence of six residues. This structure is proposed to promote dimerization by loop–loop interaction and formation of a so-called 'kissing complex'. A 23-nucleotide RNA strand containing the loop enclosed by a seven base-pair stem has been synthesized. This oligomer was crystallized by the vapour-diffusion method at 310 K, pH 6.5, with methyl-pentanediol as the precipitant agent in the presence of $MgCl_2$, KCl and spermine. Quasi-complete diffraction data were obtained at 2.7 Å resolution with a conventional X-ray source and at 2.3 Å resolution on a synchrotron beamline. The space group is $P3_121$ or its enantiomorph $P3_221$, with cell parameters $a = b = 60.1$, $c = 65.9$ Å at ambient temperature, or $a = b = 59.0$, $c = 64.3$ Å in a nitrogen-gas stream. There are two oligomers per asymmetric unit as determined from absorbance measurements of a dissolved crystal whose volume was carefully determined. In some cases, either perfectly or partially twinned crystals were obtained. Perfect twinning is detected by an apparent hexagonal symmetry and yields unusable crystallographic data, whilst partial twinning yields usable data after adequate processing. Structure solution is under way by searching for heavy-atom derivatives and systematically substituting bromo- or iodo-uridines for uridines.

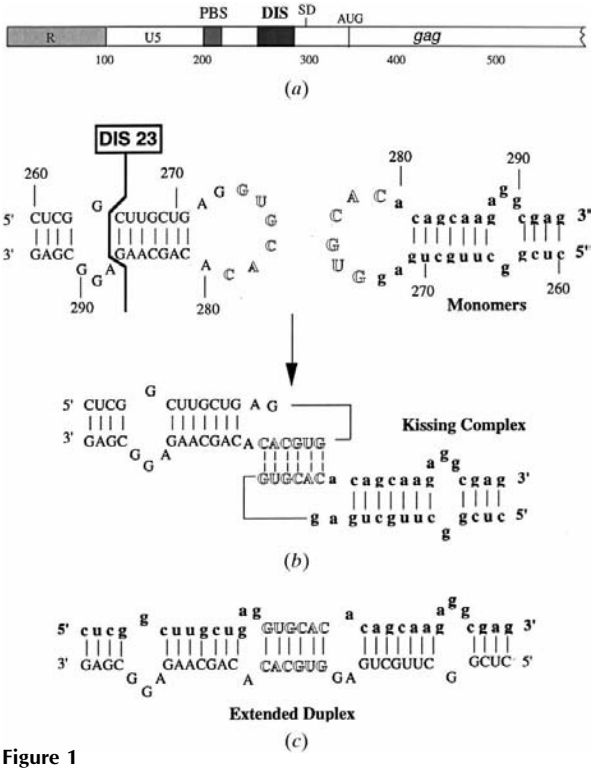

Figure 1
(*a*) Localization of the dimerization-initiation site (DIS) within the HIV-1 genomic RNA. (*b*) Secondary structure of the 'kissing complex'. (*c*) Secondary structure of the extended duplex.

1. Introduction

The genome of all retroviruses in general, and of HIV in particular, consists of an RNA that needs to be copied into a proviral DNA by a viral reverse transcriptase (RT) (Baltimore, 1970; Temin & Mizutani, 1970). This step is a prerequisite to stable integration of the viral genome into the genome of the infected cells which, in turn, will transcribe it into new viral RNA and propagate the infection. The genomic RNA is encapsidated into virions, not simply as a single chain, but as a dimeric species maintained, presumably, by non-covalent links. The first evidence of dimerization came from electron-microscopy studies (Bender *et al.*, 1978; Kung *et al.*, 1975; Maisel *et al.*, 1978; Murti *et al.*, 1981). It was further shown that the two RNA molecules are associated near their 5′ ends at a specific site (Bender & Davidson, 1976; Kung *et al.*, 1976) loosely named the dimer-linkage structure (DLS).

The exact structure and localization of the DLS depend on the particular retrovirus. Its accurate localization in HIV-1 (Baudin *et al.*, 1993; Marquet *et al.*, 1991, 1994) ultimately allowed us to prove that the initial and key site of interaction in the HIV-1 Mal isolate corresponds to a short palindromic sequence making a loop within a hairpin located upstream of the splice donor site (Fig. 1) (Skripkin *et al.*, 1994). This finding was readily confirmed by others (Laughrea & Jetté, 1994; Muriaux, Fossé *et al.*, 1996) on another HIV-1 subtype (Lai isolate). We have termed this site the dimerization-initiation site (DIS) and shown that any mutation affecting the palindromic nature of the loop sequence impaired dimerization of large RNA fragments containing the mutated DIS (Paillart *et al.*, 1994). Recently, we have proposed a molecular model of the DIS based on mutagenesis and chemical probing (Paillart *et al.*, 1997). The

results of this work are in agreement with two RNA molecules engaged in a so-called 'kissing complex' due to an interaction through loop–loop pairing (Fig. 1*b*).

Such viral-RNA dimerization is of great functional importance, as efficient viral replication and fitness depend on it (for a general review, see Paillart, Marquet *et al.*, 1996). Firstly, dimerization of the viral RNA acts as a signal for its encapsidation (Berkhout & van Wamel, 1996; McBride & Panganiban, 1996; Paillart, Berthoux *et al.*, 1996; for a review, see Rein, 1994). Secondly, by facilitating RNA-strand transfer during reverse transcription, dimerization allows efficient genetic recombination between the two copies of the RNA (Hu & Temin, 1990). The latter event increases the viral genomic variability resulting from the error-prone copying mechanism by the viral RT. It is now widely accepted that such genomic variability is a key feature, allowing the virus to escape the selection pressure from drugs directed against various viral components. It thus appears appropriate to try to prevent the RNA-dimerization step for therapeutic goals. We have, therefore, initiated the determination of the three-dimensional structure of the DIS with the goal of designing inhibitors of dimer formation. Here, we report the crystallization and preliminary crystallographic results concerning a 23-nucleotide RNA strand comprising a seven base-pair stem enclosing the loop bearing the palindrome (Fig. 1*b*).

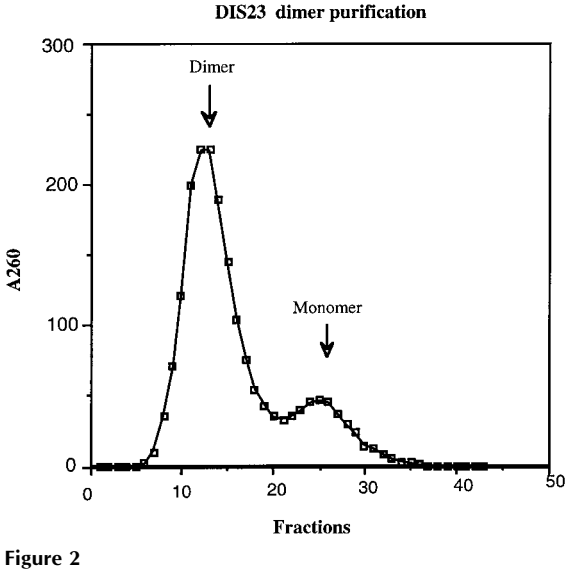

Figure 2
Purification of DIS23 RNA dimer by gel filtration using a Superdex 75 column (24 ml). 10.0 optical density units of RNA in dimerization buffer (see §2) was loaded on the column. Flow rate was 1 ml min^{-1} and temperature was 293 K.

2. Materials and methods

2.1. RNA synthesis and RNA dimer formation

Oligoribonucleotides were chemically synthesized and purified initially by GENSET (Paris, France) and subsequently in our laboratory on a 392 DNA/RNA synthesizer (Applied Biosystems). Lyophilized DIS RNA was solubilized in water to a final concentration of 0.4 mg ml^{-1}, heated at 363 K for 2 min, cooled on ice and incubated at 310 K for 30 min in the dimerization buffer (50 m*M* sodium cacodylate pH 7.0, 300 m*M* KCl, 5 m*M* $MgCl_2$) as described (Marquet *et al.*, 1991). Samples were concentrated to 10 mg ml^{-1} on a Centricon 10 (Amicon) at 293 K. Dimer formation was confirmed with RNA at 0.05 mg ml^{-1} on native 20% acrylamide gel in $TBM_{0.5}$ buffer [45 m*M* Tris borate pH 8.3, 0.5 m*M* $Mg(OAc)_2$]. Electrophoresis was run in the same buffer.

2.2. Purification of RNA dimers

For analytical or preparative purification, 0.05 or 10.0 optical density (OD) units, respectively, of DIS RNA sample was injected onto a 24 ml Superdex 75 column (Pharmacia, Sweden) and eluted at 293 K with the dimerization buffer. For the purpose of crystallization, the dimer peak was collected, concentrated to 10 mg ml^{-1} on a Centricon 10 and filtered through a 0.22 µm filter (Millipore).

2.3. Preliminary screening of crystallization conditions

NATRIX solutions (Hampton Research, USA) containing MPD, PEG and ammonium sulfate were chosen for preliminary screening of crystallization conditions. DIS RNA of 35, 27 and 23 nucleotides (nt) in length, corresponding to 13, nine and seven base pairs in the stem, respectively, were tested. Small volumes of 1 µl of RNA and reservoir solutions were mixed and allowed to reach equilibrium at 293 K. Two conditions, namely, (i) 35% MPD, 50 m*M* sodium cacodylate pH 6.5, 5 m*M* $MgCl_2$ and 20 m*M* NaCl and (ii) 30% PEG 8000, 100 m*M* sodium cacodylate pH 6.5, 200 m*M* ammonium sulfate, gave similar microcrystal clusters for the three DIS RNA lengths. With 20 m*M* spermine, however, small flat individual crystals, 100 µm in diameter, were obtained for the 23 nt long DIS RNA (DIS23). As crystals in MPD were thicker than those in PEG, we focused on MPD conditions, *i.e.* the ion concentration ($MgCl_2$, KCl and spermine) and the MPD concentration in the reservoir were systematically varied.

2.4. Crystallographic data

The first crystallographic data with conventional sources were recorded in different laboratories and, thus, on various installations: a Siemens area detector and generator and a MAR Research image plate and Rigaku generator. More recently, data were collected on our own installation comprising a Nonius generator with MacScience double-mirror focusing optics and a DIP2000 image-plate detector. All generators were operated at 45 kV and 90 or 100 mA. On our own installation, we found that data collection could be performed without any additional nickel filter. This gave a significant 20% increase in intensity and, although the remaining $K\beta$ radiation was present, the relative smallness of the unit cell allowed the processing software to perfectly separate the two sets of spots. Data were also collected on a CCD area detector at the D2AM beamline of the ESRF (Grenoble, France). Processing was first performed with *XDS* (Kabsch, 1988*a*,*b*) and, for the more recent data, with *DENZO* (Otwinowski & Minor, 1996). A significant difference between our own laboratory data and those obtained with an X-ray beam of better monochromaticity (neither continuous background nor $K\beta$ radiation) concerns the intensities of the systematically absent reflections ($0, 0, l \neq 3n$) which were measured as weak, but not extinct (up to 10σ), while they were exactly zero on average on a synchrotron beamline. This is interpreted as a specific problem for these reflections for two reasons. Firstly, they lie on the same line as the non-extinct ($0, 0, l \neq 3n$) reflections and, secondly, these non-extinct reflections are very strong. As a consequence, they are much more affected than the 'average' reflections by both the streaks due to the remaining continuous background and the 'ghosts' due to the remaining $K\beta$ radiation.

3. Results

3.1. Crystallization of purified DIS23 dimer

In order to improve the quality of crystals, dimeric DIS23 RNA was purified by gel

filtration on a Superdex 75 column (Fig. 2). The purified DIS23 dimer was stable following rechromatography under the same ionic conditions. Temperature was very important for resolution limits: crystals grown at 293 and 303 K usually diffracted, at most, to 5 Å resolution, while crystals grown at 310 K usually diffracted to 2.8 Å resolution or better (in one case, however, a crystal grown at 293 K diffracted to 2.8 Å resolution).

Screening for the best conditions led to the following procedure, currently in use. Purified DIS23 dimer at 10 mg ml^{-1} is diluted and complemented with spermine, methyl-pentanediol (MPD) and magnesium, thus resulting in DIS RNA dimer at 5.0 mg ml^{-1} in a 'crystallization buffer' of 25 m*M* sodium cacodylate pH 7.0, 150 m*M* KCl, 5 m*M* $MgCl_2$, 5 m*M* spermine and 1%(*v*/*v*) MPD. Crystals are usually obtained within 2–4 d in sitting drops at 310 K, by equilibrating 5–6 µl of DIS23 in the crystallization buffer against 0.5 ml of 50%(*v*/*v*) MPD in the dimerization buffer. Different morphologies are obtained, ranging from characteristic truncated triangular flat plates (Fig. 3*a*) to more elongated prisms (Fig. 3*c*) to irregularly shaped crystals (Figs. 3*b* and 3*d*). Despite this, all these crystals belong to the same trigonal space group and have the same cell parameters (see below). Much less often, flat parallelepipeds are also obtained (Fig. 3*e*) with monoclinic lattice (not discussed further here).

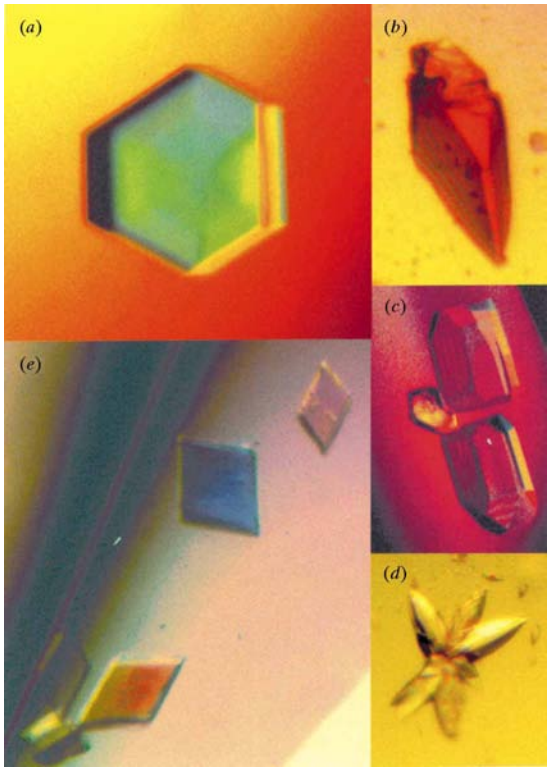

Figure 3
Crystals from purified DIS RNA dimer (23 nucleotides). (*a*) Typical crystal showing the threefold symmetry. One such crystal was used for the asymmetric unit content determination. The maximum dimension was 0.5 mm. (*b*), (*c*) and (*d*) Other morphologies obtained for the trigonal DIS crystals. (*e*) Less frequent crystal form with different space group. These crystals are very thin plates of small size, typically 0.1–0.2 mm.

3.2. Crystallographic characteristics and existence of twinning by merohedry

The limit of resolution is 2.3 Å on a synchrotron beamline, but only 2.7 Å on a conventional source. At ambient temperature, there is a significant and rapid loss of diffraction power after 6–7 h under the X-ray beam. Trials at liquid-nitrogen temperature to overcome this problem, first made by flash freezing in the nitrogen-gas stream, usually led to poor results owing to loss of resolution and large or twinned spots. However, using liquid ethane cooled at liquid-nitrogen temperature for flash freezing (Rodgers, 1997, and references therein), followed by transfer under the nitrogen-gas stream, solved all problems except for a frequent, but not systematic, increase of mosaicity (from 0.45° for the minimum value of the mosaicity to 1.2° in the worse case).

The space group is $P3_121$ or its enantiomorph $P3_221$, with cell parameters $a = b = 60.1$, $c = 65.9$ Å at ambient temperature, or $a = b = 59.0$, $c = 64.3$ Å at 120 K. In one exceptional case, however, a seemingly related hexagonal crystal form ($P6_422$ or $P6_222$; $a = b = 59.6$, $c = 65.4$ Å) was obtained. Although we had suspected that perfect twinning by merohedry (see Yeates, 1997, for an excellent review) could be responsible for such extra twofold symmetry, statistical tests *à la* Wilson on values of $\langle F^2\rangle/\langle F\rangle^2$ (Stanley, 1972) did not clearly denote twinning. Furthermore, no defects whatsoever could be detected upon visual inspection of the crystals used for data collection. More recently, however, other crystals with sixfold symmetry were obtained and all attempts at using these data remained inconclusive. Reconsidering the twinning possibility, a slightly improved statistical method taking experimental errors into consideration (Dumas *et al.*, 1999) clearly denoted twinning. This result led us to scrutinize our data for partial twinning that can be 'untwinned' and subsequently be used, as opposed to data corrupted by perfect twinning. At the time of writing, one crystal, soaked with $Ru(NH_3)_6$, has been found presenting such partial twinning. The percentage of twinning (23%) was readily obtained by the method described by Yeates (Yeates, 1997, and references therein). This value was confirmed by a completely different method based upon maximization of the correlation between the 'untwinned' data and those of a twinning-free reference data set (Dumas *et al.*, 1999). The final clue to the correctness of such processing came from the immediate finding, using the program *LOCHVAT* (Dumas, 1994*a*,*b*), of the same substitution sites in this 'untwinned' crystal and in a twinning-free $Ru(NH_3)_6$ derivative crystal.

3.3. Experimental determination of the asymmetric unit content

Calculation of the volume occupied by RNA (based on an RNA buoyant density of 1.4) leads to approximately 75 or 50% of the unit cell being filled by solvent if there are one or two DIS23 molecules, respectively, in the asymmetric unit. This uncertainty was corrected experimentally by determining the number of molecules contained in one crystal whose geometric shape allowed an accurate determination of its volume (Fig. 3*a*). To that end, the optical density (OD) of the dissolved crystal was measured before and after RNAase T1 hydrolysis (310 K, overnight). The ratio of the OD after and before hydrolysis was found to be as high as 1.56, in agreement with complete hydrolysis. From the 23-nucleotide sequence and the published values for the extinction coefficients of the ribonucleotides (Fasman, 1975), it was concluded that there are two DIS23 molecules per asymmetric unit (experimental value: 2.06) which, in turn, implies that ∼50% of the unit cell is filled by bulk solvent.

Calculation of a self-rotation function with values of κ equal and close to 180°, in order to determine the direction of a plausible non-crystallographic twofold axis of rotation, did not yield interpretable results (not shown). This is probably due to the difficulty of locating one twofold axis when helices are present.

3.4. Present state of solution structure

Initial attempts at using the available model (Paillart *et al.*, 1997) for solving the

structure by molecular replacement were not conclusive. In particular, no solutions could be found that were consistent with the self-rotation function. For solution of the structure independently of any model, we are testing various heavy-atom derivatives and also replacing a uridine by its bromo derivative in order to use MAD phasing (Hendrickson, 1991). With respect to heavy-atom derivatives, so far only the use of $Ru(NH_3)_6$ has yielded usable substitution. Each uridine in turn has been replaced by bromouridine, but only replacement of U266 and U270 allowed us to grow crystals. Crystals obtained after the replacement of U267 by bromouridine grew with difficulty and were extremely sensitive to irradiation and, therefore, unusable. Both crystals with BrU266 and BrU270 suffer from significant lack of isomorphism (-0.5 Å on *a* and *b* for BrU266 and -1 Å on *c* for BrU270) at room temperature. Until the recent use of flash cooling by ethane, the experimental data were of insufficient quality as the isomorphous and anomalous differences appeared to be within the limit of experimental noise. Work is also in progress to obtain iodo, instead of bromo, derivative crystals, as the isomorphous signal is more favourable for the former.

4. Discussion

Although previous solution studies (Paillart, Skripkin *et al.*, 1996) strongly indicate that the DIS structure corresponds to a loop–loop interaction, thus making a 'kissing complex' (Fig. 1*b*), one cannot fully reject the possibility that the X-ray structure might reveal an alternative structure, such as an extended duplex (Fig. 1*c*). First, such extended duplex conformations for the HIV-1 DIS were observed *in vitro*, albeit only under non-physiological conditions including high temperature or the absence of divalent cations (Laughrea & Jetté, 1996, 1997; Muriaux, Fossé *et al.*, 1996). This mechanism may be mediated *in vivo* by the basic viral nucleocapsid protein (Muriaux, Rocquigny *et al.*, 1996). Second, several RNA molecules susceptible to adopting a hairpin structure have shown different conformations by NMR and X-ray crystallography. The NMR structures showed the hairpin conformations (Cheong *et al.*, 1990; Jucker & Pardi, 1995), while the crystal structures of identical or related molecules showed an extended helical duplex with central residues, corresponding to the hairpin loop in the NMR structures, engaged in non-canonical interactions (Holbrook *et al.*, 1991; Lietzke *et al.*, 1996). These conflicting results were systematically examined by Kanyo *et al.* (1996). Interestingly, an NMR study of the DIS is also under investigation. The availability of both structures will provide a better insight into the potential functionally important dynamic behaviour of the HIV-1 DIS in particular, and of RNA loop–loop interactions in general.

This work was supported by grants from the 'Agence Nationale de Recherche sur le SIDA' (ANRS) and by GENSET, Paris (No. 90 10). We are extremely grateful to E. Ennifar as a 'DEA student' for his valuable participation. We thank A.-C. Dock-Bregeon and A. Munishkin for useful discussions, G. Lacroix for the asymmetric unit content determination, P. Guerne for oligonucleotide synthesis and M. Gaine for excellent technical help. We are also grateful to the staff of the D2AM beamline at ESRF (Grenoble), R. Fourme and W. Sheppard at LURE (Orsay), J. M. Rondeau (now at Novartis, Basel), C. Zelwer at CBM (Orléans) and, especially, to D. Moras (and to many people in his laboratory) at IGBMC (Illkirch-Strasbourg) for their friendly support and for kindly making their X-ray equipment available to us during the early period of this work.

References

Baltimore, D. (1970). *Nature (London)*, **226**, 209–211.
Baudin, F., Marquet, R., Isel, C., Darlix, J.-L., Ehresmann, B. & Ehresmann, C. (1993). *J. Mol. Biol.* **229**, 382–397.
Bender, W., Chien, Y. H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B. & Davidson, N. (1978). *J. Virol.* **25**, 888–896.
Bender, W. & Davidson, N. (1976). *Cell*, **7**, 595–607.
Berkhout, B. & van Wamel, J. L. (1996). *J. Virol.* **70**, 6723–6732.
Cheong, C., Varani, G. & Tinoco I., Jr (1990). *Nature (London)*, **346**, 680–682.
Dumas, P. (1994*a*). *Acta Cryst.* A**50**, 526–537.
Dumas, P. (1994*b*). *Acta Cryst.* A**50**, 537–546.
Dumas, P., Ennifar, E. & Walter, P. (1999). Submitted.
Fasman, G. D. (1975). In *Nucleic Acids*, 3rd ed., Vol. 1. Cleveland: CRC Press.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Holbrook, S. R., Cheong, C., Tinoco, I. Jr, & Kim, S.-H. (1991). *Nature (London)*, **353**, 579–581.
Hu, W. S. & Temin, H. M. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 1556–1560.
Jucker, F. M. & Pardi, A. (1995). *Biochemistry*, **34**, 14416–14427.
Kabsch, W. (1988*a*). *J. Appl. Cryst.* **21**, 67–71.
Kabsch, W. (1988*b*). *J. Appl. Cryst.* **21**, 916–924.
Kanyo, J. E., Duhamel, J. & Lu, P. (1996). *Nucleic Acids Res.* **24**, 4015–4022.
Kung, H. J., Bailey, J. M., Davidson, N., Nicholson, M. O. & McAllister, R. M. (1975). *J. Virol.* **16**, 397–411.
Kung, H. J., Hu, S., Bender, W., Bailey, J. M., Davidson, N., Nicolson, M. O. & McAllister, R. M. (1976). *Cell*, **7**, 609–620.
Laughrea, M. & Jetté, L. (1994). *Biochemistry*, **33**, 13464–13474.
Laughrea, M. & Jetté, L. (1996). *Biochemistry*, **35**, 1589–1598.
Laughrea, M. & Jetté, L. (1997). *Biochemistry*, **36**, 9501–9508.
Lietzke, S. E., Barnes, C. L., Berglund, J. A. & Kundrot, C. E. (1996). *Structure*, **4**, 917–930.
McBride, M. S. & Panganiban, A. T. (1996). *J. Virol.* **70**, 2963–2973.
Maisel, J., Bender, W., Hu, S., Duesberg, P. H. & Davidson, N. (1978). *J. Virol.* **25**, 384–394.
Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C. & Ehresmann, B. (1991). *Nucleic Acids Res.* **19**, 2349–2357.
Marquet, R., Paillart, J.-C., Skripkin, E., Ehresmann, C. & Ehresmann, B. (1994). *Nucleic Acids Res.* **22**, 145–151.
Muriaux, D., Fossé, P. & Paoletti, J. (1996). *Biochemistry*, **35**, 5075–5082.
Muriaux, D., Rocquigny, H. D., Roques, B. P. & Paoletti, J. (1996). *J. Biol. Chem.* **271**, 33686–33692.
Murti, K. G., Bondurant, M. & Tereba, A. (1981). *J. Virol.* **37**, 411–419.
Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.
Paillart, J.-C., Berthoux, L., Ottmann, M., Darlix, J.-L., Marquet, R., Ehresmann, C. & Ehresmann, B. (1996). *J. Virol.* **70**, 8348–8354.
Paillart, J.-C., Marquet, R., Skripkin, E., Ehresmann, B. & Ehresmann, C. (1994). *J. Biol. Chem.* **269**, 27486–27493.
Paillart, J.-C., Marquet, R., Skripkin, E., Ehresmann, C. & Ehresmann, B. (1996). *Biochimie*, **78**, 639–653.
Paillart, J.-C., Skripkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 5572–5577.
Paillart, J.-C., Westhof, E., Ehresmann, C., Ehresmann, B. & Marquet, R. (1997). *J. Mol. Biol.* **270**, 36–49.
Rein, A. (1994). *Arch. Virol.* **9**, 513–522.
Rodgers, D. W. (1997). *Methods Enzymol.* **276**, 183–203.
Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B. & Ehresmann, C. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4945–4949.
Stanley, E. (1972). *J. Appl. Cryst.* **5**, 191–194.
Temin, H. M. & Mizutani, S. (1970). *Nature (London)*, **226**, 211–213.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.