Modeling and solution structure probing of the HIV-1 TAR stem-loop

Andrew D. Critchley, I. Haneef,*† Diane J. Cousens,‡ and Peter G. Stockley

Department of Genetics and †Department of Biochemistry and Molecular Biology, University of Leeds, Leeds, UK; ‡The Wellcome Foundation Ltd., Beckenham, Kent, UK

We present a model for the three-dimensional structure of the HIV TAR stem-loop,[1] based on a modeling algorithm[2] which makes use of the known X-ray coordinates of tRNAs to generate a model structure, which has then been tested experimentally in solution by enzymatic and chemical structure probing[3–5] of ribo-oligonucleotides encompassing the TAR sequence. The modeling suggested that the structure of TAR was similar to that of the anti-codon loop of $tRNA^{Asp}$, having a loop of just three single-stranded residues with a mismatched adenine excluded from the helical stem on the 3′ side of the loop.[6] The structural probing is consistent with such a structure for the loop, and reveals an unusual structure around the 5′ uridine-rich bulge, which is the binding target for the transactivator protein Tat.[7] These data may be useful in understanding the interaction of TAR with the Tat protein and may aid in the design of anti-AIDS drugs. The coordinates of the model are available on request.

Keywords: HIV TAR, molecular model, RNA structure probing

INTRODUCTION

Human immunodeficiency virus-1 (HIV-1), the causative agent of the acquired immunodeficiency syndrome (AIDS)[8–10] encodes a number of novel regulatory proteins.[11,12] One of these is Tat, a *trans*-activator, which results in greatly increased expression of viral RNA and is essential for viral replication.[12] Tat protein appears both to increase the rate of transcription initiation and the efficiency of transcriptional elongation.[11,12] The *cis*-acting target sequence of Tat (TAR) is located downstream of the transcriptional initiation site.[13] The results of a large number of mutational studies, together with studies of TAR-Tat complexes *in vitro* are consistent with TAR forming a stable stem-loop (Figure 1a) having a six-nucleotide single-stranded loop, with a pyrimidine-rich single-stranded bulge on the 5′ stem. Functional analysis of this structure suggests that Tat binds to the region around the pyrimidine bulge but cannot itself produce transactivation, which requires that further cellular proteins bind at the loop region.[11]

Previously we have used molecular modeling of RNA stem-loop structures, based on homologous structural elements in the known database of tRNA crystal structures, to produce a three-dimensional model of the translational operator of the RNA bacteriophage MS2.[2] The value of such models lies in providing an essential framework for testing hypotheses, and can be used to suggest experimental tests of the structure. Chemical and enzymatic structural probes have been used extensively to examine RNA structures.[4] The most comprehensive experiments have used tRNAs whose X-ray crystal structures have been determined in order to examine the structure in solution.[5] This work has defined the usefulness of such studies, and our approach is based heavily on it. With the MS2 translational operator detailed, quantitative chemical and enzymatic probing has allowed us to "improve" our initial model, and has confirmed that the modeling algorithm produces reasonable starting structures for the design of further experiments.[2] In this paper, we report the application of this approach to the TAR stem-loop.

MATERIALS AND METHODS

Molecular modeling

The model of the 12-mer RNA fragment encompassing the top of the TAR stem-loop was derived by superimposing the structures of the phosphodiester backbones of appropriate loops in known tRNA structures[2] in the Brookhaven Protein Data Bank.[14] We used the interactive graphics program FRODO[15] and the program, MNYFIT (Haneef, unpublished) which derives the optimum superposition of a number of related molecular structures by a least-squares method that gives equal weight to each of the structures. Figure 1a shows the proposed secondary structure for the residues 27 to 38 of TAR. The base-paired stem was modeled using a template structure generated from available 3bp stems in the database as described,[2,16] followed by base substitution to the TAR sequence using the sugar-phosphate backbone as the frame. The bases were built with ideal geometries and preferred conformations that maximize hydrogen-bonding capacity of the complementary bases. Iterative least-squares

Color Plates for this article are on page 124.

Address reprint requests to Dr. Peter G. Stockley at the Department of Genetics, University of Leeds, Leeds LS2 9JT, UK.

Received 4 August 1992, revised 23 September 1992; accepted 29 September 1992

*Deceased.

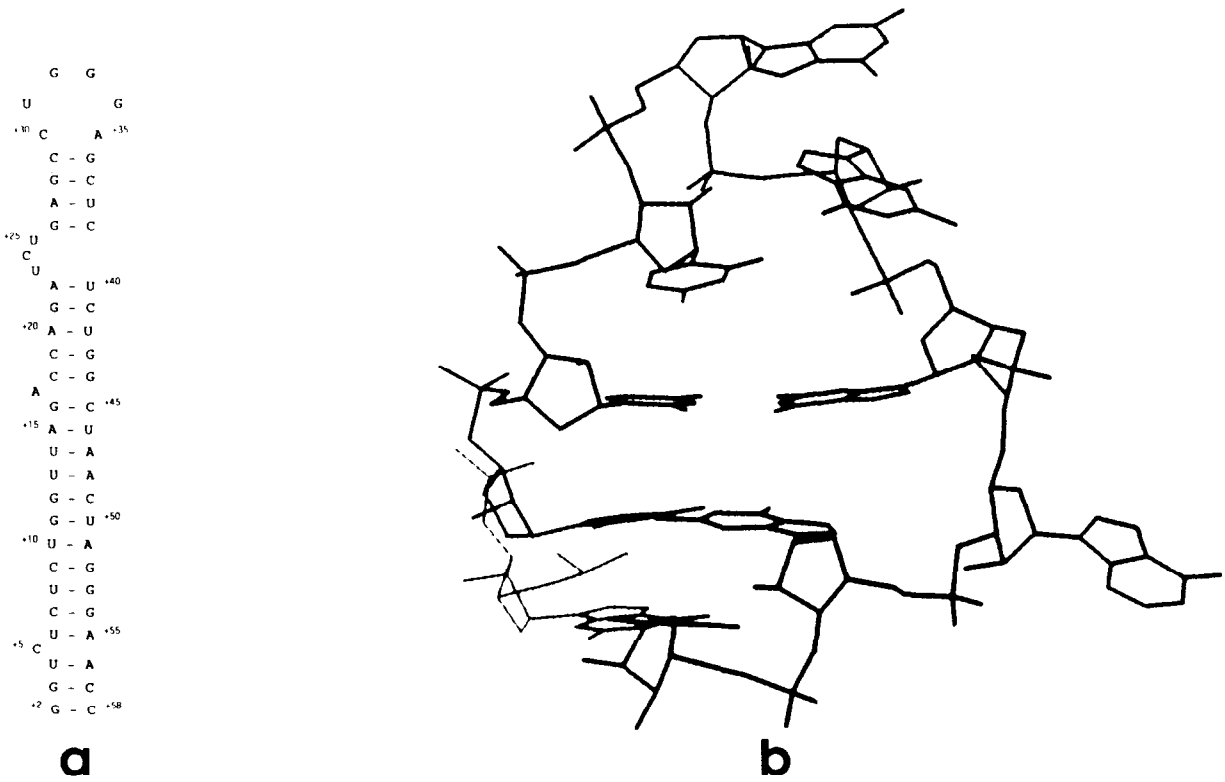

Figure 1. (a) Proposed secondary structure for the HIV TAR stem-loop region.[1] (b) Complete three-dimensional stick-model of the TAR stem-loop after insertion of the base sequence and energy minimization.

fitting and distance geometry averaging methods[17,18] were used to construct the sugar-phosphate backbone for the loop (residues 30 to 35) as described previously.[2] The result of these operations was a backbone with a conformation very similar to that seen in the known structure of the anti-codon loop of $tRNA^{Asp}$ (Brookhaven code 4TNA), although the tRNA has one extra nucleotide in the loop. Base addition was followed by manual intervention to improve base stacking and hydrogen bonding, and the resulting model regularized using the SHAKED algorithm[2] employing upper and lower bounds for the interatomic distances found in known tRNA structures.

Preparation of plasmids

The plasmid pSP64TAR was created by annealing together two oligonucleotides encompassing the TAR sequence from +20 to +42 (see Figure 1a), having a 9bp overlap and unique restriction sites at each end. The annealed DNA was then treated with Klenow polymerase to generate fully double-stranded molecules which were ligated to form concatemers, digested with EcoRI and HindIII to produce unit length molecules, gel purified and then cloned[19] into pSP64. The resulting transcript contains the region of the TAR sequence specified above with both 5′ and 3′ non-complementary six-base extensions. The plasmid pGEM3Z′BI, containing the entire TAR fragment +2 to +58 generated from synthetic oligonucleotides, was used to subclone a sequence encompassing TAR into pSP65, to form the plasmid pSP65TAR.

Preparation of RNA

In vitro run-off transcription reactions were performed on plasmids linearised with either EcoRI or HindIII, using SP6 RNA polymerase. Typical transcription reactions were performed in 50 μl and contained 10 mM dithiothreitol, 1 mM rNTPs, 5 μg bovine serum albumin, 2 μg of linearised template, 40 mM Tris · HCl (pH 7.9), 6 mM $MgCl_2$, 2 mM spermidine-$(HCl)_3$, and 15 units of SP6 RNA polymerase. Incubations were carried out for 2 hours at 37°C.[20] RNA was labeled at the 5′terminus with $\gamma[^{32}P]$ATP and T4 polynucleotide kinase (Pharmacia),[21] and purified on 1.5-mm thick 20% (w/v) polyacrylamide gels containing 7 M urea, 0.1 M Tris-Borate pH 8.1, 2 mM EDTA.[22] The sequences of the RNA runoffs were confirmed by the sequence-specific enzymatic digestion procedure.[23]

Chemical and enzymatic structural analysis

Modification of 5′-labeled RNA with diethylpyrocarbonate, DEPC (Sigma) under "native" conditions was performed in 200 μl structure buffer (30 mM Tris · Cl pH 7.4, 10 mM

$MgCl_2$, and 200 mM KCl) containing ≈50,000 cpm 5′[^{32}P]RNA, 10 μg tRNA carrier (Sigma), and 5% (v/v) DEPC at 20°C for 60 minutes. Modification under denaturing conditions was performed in 200 μl D1 buffer (50 mM sodium cacodylate pH 6.8, 1 mM EDTA) containing ≈50,000 cpm 5′[^{32}P]RNA, 10 μg tRNA carrier, and 0.5% (v/v) DEPC at 90°C for 10 minutes. All reactions were stopped by addition of 50 μl 1.5 M sodium acetate pH 6.0 and 750 μl ethanol. The recovered RNA pellets were then reprecipitated from 200 μl 0.3 M sodium acetate pH 6.0 and dried briefly under vacuum. Cleavage of RNA at the DEPC-modified sites with aniline (Sigma) was performed as described by Peattie,[24] and the products separated on 20% (w/v) polyacrylamide/7 M urea gels.[22] Aliquots of 5′[^{32}P]RNA were digested with either RNase V1 (Pharmacia), which cleaves 5′ to double-stranded or structured bases, RNase T2 (3′ single-strand specific), RNase T1 (3′ single-strand G specific), RNase CL3 (3′ single-strand C≫A>U specific), RNase U2 (3′ single-strand A>G specific), RNase *B. cereus* (3′ single-strand U and C specific) or RNase *Phy.M* (3′ single-strand A and U specific) at 24°C for 30 minutes. The latter enzymes were all from BRL. Reactions were performed in 4 μl of structure buffer containing 2 μg tRNA carrier and were stopped by addition of 5 μl of denaturing loading buffer (10 M urea, 1.5 mM EDTA, 0.05% (w/v) xylene cyanol, and 0.05%(w/v) bromophenol blue). The oligonucleotides produced were separated on 20% (w/v) polyacrylamide/7 M urea gels[22] and located by autoradiography. In order to identify each residue, control lanes containing enzymatic sequence and formamide ladders were run alongside experimental samples.[23]

RESULTS

The modeling algorithm used here is based on distance geometry techniques,[25] and was first developed for modeling of drugs to enzyme active sites and loops in protein structures. However, RNA stem-loops present very similar modeling problems.[2] As far as possible, the modeling is based on homologous structures determined experimentally, which in the case of RNA is restricted, at atomic resolutions, to tRNAs, although a number of NMR structures have recently been reported.[26] In this way, it is relatively straightforward to model stems,[16] including those containing mismatched residues. The completion of the molecule by modeling of the single-stranded loops is then an ideal problem for the techniques of distance geometry; with the loop needing to begin and end at defined positions and to include a defined number of phosphodiester steps. The algorithm automatically scans a database of known tRNA loop backbones from which it selects a number of possible solutions. From this it is then possible to derive the "best" backbone structure to which the bases are then added. Consideration can then be given to maximizing hydrogen bonding and base-stacking, both of which are apparently key features of known structures, and the resultant model energy minimized.[2]

The modeling was restricted to the top portion of the TAR stem-loop (residues 27 to 38) shown in Figure 1a. The algorithm immediately suggested that an extra base-pair could be made between the 5′ C(+30) of the loop and the penultimate G(+34), producing a three-base single-stranded loop and a mismatched adenine on the 3′ leg of the stem. An almost identical structure has been observed[6] for the anti-codon loop of $tRNA^{Asp}$ (Color Plate 1). The completion of the model by addition of the bases and subsequent energy minimization, was as described above, and produced the structure shown in Figure 1b. An obvious drawback of modeling from preexisting structures is that the resultant model can be heavily biased in the direction of the input structures. Our model of the TAR stem-loop is particularly open to this charge since its backbone structure is so close to that of $tRNA^{Asp}$. We have therefore conducted a series of chemical and enzymatic structure probing experiments[4] aimed at testing the major features of the model.

We produced SP6 RNA polymerase transcripts encompassing the top half of the TAR stem-loop (from plasmid pSP64TAR) or the whole TAR sequence (from plasmid pSP65TAR). Radioactively end-labeled oligonucleotides were then treated with a variety of enzymatic and chemical probes under buffer conditions expected to preserve both secondary and tertiary structure,[4] and the extent of cleavage/modification at various sites was determined by separating the products on polyacrylamide sequencing gels followed by autoradiography.

Our previous work with the MS2 operator has shown that the reagent diethyl pyrocarbonate (DEPC) which modifies N_7 positions[27] can be used as a structural probe of the conformation of single-stranded adenines. Previously, this reagent had been used to determine adenines involved in Watson–Crick base-pairing; the N_7 position being sterically hindered in such interactions and hence the base unreactive. However, we have been able to show that certain adenines exhibit apparent hyperreactivity towards DEPC. Hyperreactivity occurs, we believe, in single-stranded *A* positions (such as intercalation sites) where neighboring functional groups can participate in the modification reaction.[27] Treatment of TAR oligonucleotides with DEPC suggests that the 3′ mismatched A (+35) is modified no more extensively than control single-stranded adenines in the sequence (data not shown), consistent with its modeled position of being extruded from the stem and in contact with the solvent rather than intercalated between the neighboring C-G base-pairs.

Transcripts were then probed enzymatically using the single-strand, base-specific ribonucleases, T1, *B. cereus*, *Phy.M*, CL3, together with the base-paired or structure-specific ribonuclease from cobra venom V1 and the single-strand specific ribonuclease T2. The results of some of these experiments are shown in Figure 2, and a full summary is given in Table 1 and Figure 3. In the loop region (+29 to +36) RNase V1 cleaves at residues C29, C30, and U31, the latter residue being the most readily cleaved. These data are consistent with stacking of the bases at the 5′ side of the loop. The next position which is cleaved by V1 is G36, implying that the residues between U31 and A35 are neither stacked nor base-paired. The most striking feature of the enzymatic digestions is seen with RNases T1 and T2, which are specific for single-stranded guanines and any single-stranded residue respectively. In each case the enzymes cleave strongly at G32, less strongly at G33 and barely, if at all, at G34. U31 was clearly recognized by RNase *B. cereus* which also appeared to cleave weakly at C30. The other

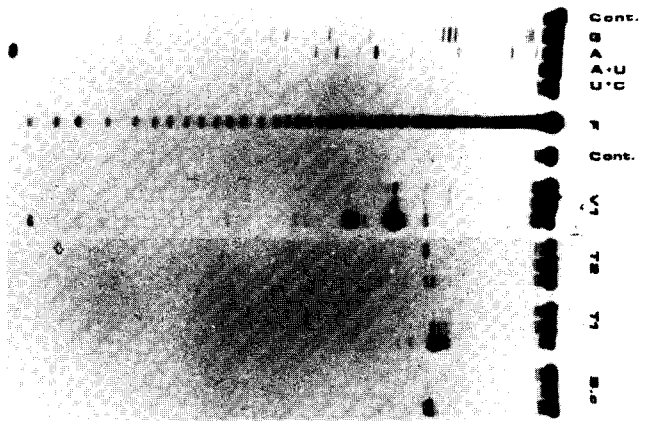

Figure 2. Enzymatic probing of RNAs encompassing TAR. Samples were: Cont., incubation controls; G, RNase T1; A, RNase U2; A + U, RNase Phy.M; *U + C, RNase* B. cereus; *F, formamide ladder; V1, T2, T1, and B.c indicate lanes containing RNase digestions in structure buffer with (from left to right); 0.005, 0.01, and 0.1 units; 0.1, 0.5, and 1.0 units; 0.001, 0.01, and 0.1 units; 0.001, 0.01, and 0.1 units, respectively.*

sequence-specific enzymes cleave so poorly in this region that little clear information can be gleaned.

The bulge region comprises residues +20 to +26 on the 5′ side of the stem and residues +38 to +43 on the 3′ side. RNase V1 cleaves at all the residues on the 3′ side of the stem and at all residues except A22 and U23 on the 5′ side.

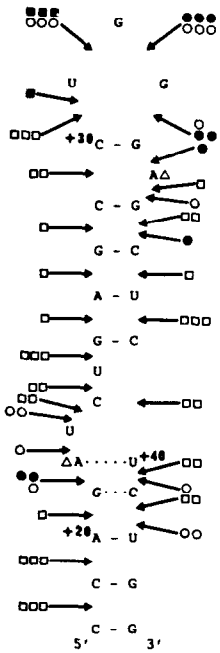

Figure 3. Summary of the structure probing experiments and the implied secondary structure of TAR in solution. Symbols are □, *RNase V1;* ●, *RNase T1;* ○, *RNase T2; and* ■, *RNase* B. cereus. *The extent of cleavage is indicated approximately by the numbers of each symbol. Nucleotides modified by DEPC are indicated by* △.

The implication, namely that A22 and U23 are single-stranded, is supported by the T2 and *B. cereus* cleavages which occur at these sites. However, the single-strand specific nucleases also produce cleavages at G21, a result apparently inconsistent with the V1 cleavage 5′ to this residue (see below).

DISCUSSION

The results of both chemical and enzymatic structure probing with oligonucleotides encompassing the TAR stem-loop are largely consistent with the structure for the loop region proposed in the model of Figure 1b. DEPC modification of A35 supports the suggestion that this residue is single-stranded, while the extent of modification suggests that it is not intercalated.[27] V1 cleavages at C30 and U31, together with the differential cleavages by T1 and T2 at G32, G33, and G34, are also consistent with formation of the C-G base pair (+30/+34). The cleavages at U31 are interesting because both T2 and the *B. cereus* enzyme cleave here readily, consistent with a single-stranded residue, but it is also recognized by V1, suggesting a structured conformation, probably stacked onto the base pair below.

The uridine-rich bulge region showed a number of cleavages which suggest that the secondary structure of Figure 1a is not an accurate description of the conformation in solution. In particular, the structure probing suggests that the potential base pairs between A22 and U40, and G21 and C41 are not made, or are at least transient. V1 cleavages across every position on the 3′ leg of the bulge suggest that although U40 and C41 may be single-stranded, they are nevertheless structured. The same argument can be used for U25, C24, and G21 on the 5′ leg of the bulge, although A22 and U23 are clearly not recognized and presumably have a conformation distinct from the helix. A22 was also modified by DEPC (data not shown) consistent with a single-stranded conformation. A summary of the structure probing data and the implied TAR secondary structure in solution is shown in Figure 3.

Although the results described here suggest that the structure in Figure 1b is a reasonable representation of the solution structure of the loop region of short ribo-oligonucleotides containing the TAR sequence, it cannot take into account long-range interactions which might occur in natural HIV RNAs, and therefore, should be used only as a starting point in experiments aimed at understanding interaction with Tat or drug-binding. The principle behind enzymatic structure probing is that only single cleavages, which are dependent on the initial conformation, are produced in each molecule. We have assumed this to be the case here given the low levels of cleavage in experiments used to generate the data. Enzyme processivity could also be a source of unexpected cleavages. Again, we do not think that this was a significant problem here, although it might explain the apparent weak cleavage by RNase *B. cereus* at C30 and G21, and the weak cleavage by RNase T1 at G36.

Recently Colvin and Garcia-Blanco reported[28] the results of an even more extensive set of structure probing experiments on TAR. The data obtained for the residues in the loop are entirely consistent with the structure probing results presented here, although these authors propose a different base pair interaction across the loop, namely C30 to G33.

Table 1. Summary of the solution structure probing data. Positions which were uncleaved are indicated by −, the approximate extent of cleavage by +. Data are summaries taken from several experiments and exposures of autoradiographs such as that shown in Figure 2. Nucleotide sequences are according to Feng and Holland,[1] the modeled region of Figure 1b corresponds to the sequence from +29 to +36

Base	RNase V1	RNase T2	RNase T1	RNase *B.cereus*
U +42	+ +	−	−	−
C	+ +	+ +	−	−
U +40	+ +	+	−	−
C	+ + +	−	−	−
U	+	−	−	−
C	+ +	−	−	−
G	+	−	+	−
A +35	−	+	−	−
G	−	−	+	−
G	−	+	+ +	−
G	−	+ + +	+ + +	−
U	+ + +	+ + +	−	+ + +
C +30	+ +	−	−	+
C	+	−	−	−
G	+	−	−	−
A	+	−	−	−
G	+ + +	−	−	−
U +25	+ +	−	−	−
C	+ +	−	−	−
U	−	+ +	−	+
A	−	+	−	+
G	+	+	+ +	−
A +20	+ + +	−	−	−
C	+ + +	−	−	−
C	−	−	−	−
A	−	−	−	−
G	−	−	+	−
A +15	−	−	−	−
U	+ + +	−	−	−
U	+ + +	−	−	−
G	−	−	−	−
G	−	−	−	−
U +10	−	−	−	−
C	+ +	−	−	−
U	+ +	−	−	−
C	+	−	−	−
U	+	−	−	−
C +5	+	−	−	−
U	+ + +	−	−	−
G	+	−	−	−
G +2	+ + +	−	−	−

Both sets of data, however, suggest a defined, sequence-dependent tertiary conformation for the loop. We and others are attempting to determine the structure of TAR oligonucleotides in solution directly by NMR. Until such time as these experiments can produce an experimental structure for TAR, the model would appear to be a useful tool for studies on the functional role of TAR.

The coordinates of the model are available on request.

ACKNOWLEDGMENTS

This work was supported by SERC grants GR/D 86331 and GR/E 95750 to PGS; by SERC Grant GR/E76001, establishing a Centre for Molecular Recognition in Biological Systems at the University of Leeds; and by The Wellcome Foundation. We thank Simon Talbot for his preliminary work on the DEPC reactivity of TAR. We thank Prof.

A.C.T. North for helpful comments during the revision of this manuscript.

REFERENCES

1 Feng, S. and Holland, E.C. *Nature* 1988, **334,** 165–167
2 Haneef, I., Talbot, S.J., and Stockley, P.G. *J. Mol. Graphics* 1989, **7,** 186–195
3 Mougel, M., Eyermann, F., Westhof, E., Romby, P., Bezancon, A.E., Ebel, J.-P., Ehresmann, B., and Ehresmann, C. *J. Mol. Biol.* 1987, **198,** 91–107
4 Ehresmann, C., Boudin, F., Mougel, M., Romby, P., Ebel, J.-P., and Ehresmann, B. *Nuc. Acids Res.* 1987, **15,** 9109–9128
5 Vlassov, V.V., Giege, R., and Ebel, J.-P., *Eur. J. Biochem.* 1981, **119,** 51–59
6 Moras, D., Comarmond, M.B., Fischer, J., Weiss, R., Thierry, J.C., Ebel, J.P., and Giege, R. *Nature* 1980, **288,** 669
7 Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M., Skinner, M.A., and Valerio, R. *PNAS* 1989, **86,** 6925–6929
8 Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Charmaret, S., Gruest, J., Dauget, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rosenbaum, W., and Montagnier, L. *Science* 1983, **220,** 868–871
9 Gallo, R.C., Salahuddin, S., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, T. *Science* 1984, **224,** 500–503
10 Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. *Science* 1984, **224,** 497–500
11 Cullen, B.R. *Cell* 1989, **63,** 655–657
12 Rosen, C.A. and Pavlakis, G.N. *AIDS* 1990, **4,** 499–509
13 Rosen, C., Sodroski, J., and Haseltine, W. *Cell* 1985, **41,** 813–823
14 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. *J. Mol. Biol.* 1977, **122,** 535
15 Jones, T.A. *J. Appl. Crystallogr.* 1978, **11,** 268
16 Subbarao, N. and Haneef, I. *Prot. Eng.* 1991, **4,** 877–884
17 Haneef, I. and Sutcliffe, M.J. *Inf. Quart. Protein Crystallogr.* 1986, **18,** 11
18 Sutcliffe, M.J., Haneef, I., Carney, D., and Blundell, T.L. *Prot. Eng.* 1987, **1,** 377
19 Maniatis, T., Fritsch, E.P., and Sambrook, J. *Molecular Cloning. A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982
20 Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. *Nuc. Acids Res.* 1984, **12,** 7035–7056
21 Siberklang, M., Gillum, A.M., and Rajbhandary, U.L. *Nuc. Acids Res.* 1977, **4,** 2527–2538
22 Sanger, F.G. and Coulson, A.R. *FEBS Lett.* 1979, **87,** 107–110
23 Donis-Keller, H. *Nuc. Acids Res.* 1980, **8,** 3133–3141
24 Peattie, D.A. *PNAS* 1979, **76,** 1760–1764
25 Haneef, I., Foundling, S., Cooper, J., and Blundell, T. unpublished
26 Cheong, C., Varani, G., and Tinoco, I. Jr. *Nature* 1990, **346,** 680–682
27 Talbot, S.J., Medina, G., Fishwick, C.W.G., Haneef, I., and Stockley, P.G. *FEBS Lett.* 1991, **283,** 159–164
28 Colvin, R.A. and Garcia-Blanco, M.A. *J. Virol.* 1992, **66,** 930–935