FAST COMMUNICATIONS

Contributions intended for this section should be submitted to any of the Co-editors of Acta Crystallographica *or* Journal of Applied Crystallography. *In the letter accompanying the submission authors should state why rapid publication is essential. The paper should not exceed two printed pages (about 2000 words or eight pages of double-spaced typescript including tables and figures) and figures should be clearly lettered. If the paper is available on 5.25" IBM PC compatible or 3.5" Apple /Macintosh diskettes it would be helpful if these could be sent with the manuscript together with details of the word-processing package used. Papers not judged suitable for this section will be considered for publication in the appropriate section of* Acta Crystallographica *or in* Journal of Applied Crystallography.

Crystallographic Analyses of an Active HIV-1 Ribonuclease H Domain Show Structural Features that Distinguish it from the Inactive Form

By D. Chattopadhyay [1], B. C. Finzel [1], S. H. Munson [1], D. B. Evans [1], S. K. Sharma [1], N. A. Strakalaitis [1], D. P. Brunner [1], F. M. Eckenrode [1], Z. Dauter [2], Ch. Betzel [2] and H. M. Einspahr [1]

[1] *The Upjohn Company, Kalamazoo, Michigan 49001, USA*

[2] *EMBL Outstation, c/o DESY, Notkestrasse 85, 2000 Hamburg 52, Germany*

(*Received 4 January* 1993; *accepted 5 March* 1993)

Abstract. An active recombinant preparation of the carboxy-terminal ribonuclease H (RNase H) domain of HIV-1 reverse transcriptase has produced crystals of several different forms, including a trigonal prism form ($P3_1$; $a = b = 52.03$, $c = 113.9$ Å with two molecules per asymmetric unit) and a hexagonal tablet form ($P6_222$ or $P6_422$; $a = b = 93.5$, $c = 74.1$ Å with one molecule per asymmetric unit). The former appears to be isomorphous with crystals of a similar, but inactive, version of the enzyme that was used for a prior crystal structure determination [Davies, Hostomska, Hostomsky, Jordan & Matthews (1991). *Science*, **252**, 88–95]. We have also obtained a structure solution for this crystal form and have refined it with 2.8 Å resolution data ($R = 0.216$). We report here details of our crystallization studies and some initial structural results that verify that the preparation of active HIV-1 RNase H yields a protein that is not just enzymatically, but also structurally, distinguishable from the inactive form. Evidence suggests that region 538–542, which may be involved in the catalytic site and which is disordered in both molecules in the prior structure determination, is ordered in the crystal structure of the active enzyme, although the ordering may include more than one conformation for this loop. It should also be noted that, in the crystal structure of the trigonal form, RNase H monomers associate to form noncrystallographic twofold-symmetric dimers by fusing five-stranded mixed β sheets into a single ten-stranded dimerwide sheet, an assembly that was not remarked upon by previous investigators.

The reverse transcriptase (RT) of human immunodeficiency virus type 1 (HIV-1) plays a key role in the life cycle of the virus by copying the single-stranded RNA genome of the virus into double-stranded DNA within the infected cells. Successful reverse transcription is a concerted functioning of RNA- and DNA-directed DNA polymerase activities, together with a ribonuclease H (RNase H) activity (Furfine & Reardon, 1991).

The RNase H activity is an integral part of HIV-1 RT function, acting as an endonuclease as well as a 3′-5′ exonuclease (Schatz, Mous & LeGrice, 1990). It is responsible for the removal of the tRNA that is utilized as primer for initiation of DNA strand synthesis. It is also responsible for partial digestion and removal of the RNA template to permit synthesis of the second DNA strand (Mitra, Goff, Gilboa & Baltimore, 1979; Mitra, Chow, Champoux & Baltimore, 1982; Finston & Champoux, 1984; Resnick, Omer & Faras, 1984; Rattray & Champoux, 1987; Furfine & Reardon, 1991). As might be expected, RNase H activity has been found to be essential for replication of HIV-1 (Tanese & Goff, 1988; Repaske, Hartley, Kavlick, O'Neill & Austin, 1989). Inhibitors of this RNase H activity may therefore have potential as drugs effective against the HIV-1 virus and acquired immunodeficiency syndrome (AIDS). So far as is known, most of the efforts to design chemotherapeutic agents against RT have targeted the RNA-directed DNA polymerase activity, while the other potentially unique target, the RNase H function, has not yet been explored.

From mutagenesis studies (Larder, Purifoy, Powell & Darby, 1987; Schatz, Cromme, Grueninger-Leitch & LeGrice, 1989; Mizrahi, Usdin, Harington & Dudding, 1990), sequence similarities (Johnson, McClure, Feng, Gray & Doolittle, 1986), and monoclonal antibody

probing studies (Hansen, Schulze, Mellert & Moelling, 1988), the RNase H activity of HIV-1 RT has been assigned to a 15 kDa domain at its C-terminal extreme. Although some mutations in HIV-1 RT influence both the polymerase and RNase H functions (Hizi, Hughes & Shaharabany, 1990), certain residues in the C-terminal domain of the enzyme can be mutated to affect the RNase H activity without altering the DNA polymerase activities (Mizrahi et al., 1990; Schatz et al., 1989). Prior to this, RNase H activities of the avian myeloblastosis and murine moloney leukemia viruses were shown to be associated with isolatable domains capable of functioning independently of the remainder of the RT enzymes of these viruses (Brewer & Wells, 1974; Tanese & Goff, 1988). These findings, together with the fact that, during its processing to the 51 kDa form, the 66 kDa monomer of HIV-1 RT loses its C-terminal 15 kDa RNase H segment by proteolysis at the Phe440-Tyr441 peptide bond (Chattopadhyay et al., 1992; Goff, 1990), would seem to suggest a relevance for an isolated HIV-1 RNase H domain.

While existence of a 15 kDa C-terminal RT fragment has been demonstrated in HIV-1 viral lysates (Hansen et al., 1988), the enzymatic role of the isolated RNase H domain in vivo remains to be established. The understanding of its functions is further complicated by the fact that recombinant versions of the protein expressed in bacteria by two different groups were found to be enzymatically inactive (Becerra et al., 1990; Hostomska, Matthews, Davies, Nodes & Hostomsky, 1991), except in association with a recombinant 51 kDa polymerase domain of HIV-1 RT (Hostomsky, Hostomska, Hudson, Moomaw & Nodes, 1991). The latter RNase H protein, which could be crystallized in trigonal form ($P3_1$; $a = b = 51.9$, $c = 114.9$ Å; two molecules per asymmetric unit; Hostomsky et al., 1991), is the basis of a crystal structure determination at 2.4 Å resolution (Davies, Hostomska, Hostomsky, Jordan & Matthews, 1991). A significant aspect of the structural results for this inactive RNase H protein is that, in both molecules in the asymmetric unit, a loop region of five or six residues containing His539, which is an invariant residue in other RNase H sequences (Doolittle, Feng, Johnson & McClure, 1989) and which has been shown to be important to RNase H activity (Schatz et al., 1989; Tisdale, Schulze, Larder & Moelling, 1991), is apparently disordered and could not be modeled.

We report here crystallization of another version of the HIV-1 RNase H, one which has been shown to be enzymatically active in two different in vitro assays (Evans, Brawn, Deibel, Tarpley & Sharma, 1991). This active RNase H protein comprises residues 427 through 560 of the RT sequence, to which is attached a genetically engineered N-terminal 11-residue peptide sequence designed to bind metal ions. The protein is expressed in E. coli cells by conventional recombinant methods and is purified by extraction from cell lysates on a metal-affinity column containing an iminodiacetate Sepharose resin loaded with Ni^{2+}.

Crystals of this active RNase H were grown in hanging-drop and sitting-drop vapor-diffusion experiments under conditions similar to those used by Davies et al. (1991). Drops containing a 1:1 mixture (by volume) of 8.5–10 mg ml^{-1} protein solution and a precipitant solution consisting of 20–23% (w/v) PEG 6000 in 0.1 M sodium citrate and 0.15 M sodium tartrate, pH 5.2, when equilibrated against reservoirs containing the precipitant solution, produce crystals of a trigonal-prism morphology at room temperature. At 277 K, crystals of this form grow better at somewhat lower concentrations of PEG 6000. While the crystals described by Davies et al. (1991) were grown from PEG 8000, we found that use of this precipitant very often resulted in poorly formed crystals. Concentration of the tartrate salt in the precipitant solution also seems to be critical, particularly for the 277 K experiments. Crystals appear within a few days and continue to grow for two to three weeks. Preliminary X-ray diffraction data show these crystals to be isomorphous with those reported by Davies et al. (1991), with space group $P3_1$, $a = b = 52.03$, $c = 113.9$ Å, and two molecules per asymmetric unit. Synchrotron X-ray data collected from specimens of this crystal form at the EMBL Outstation at DESY in Hamburg initially showed scattering to 2.4 Å resolution.

A hexagonal crystal form of HIV-1 RNase H has also been obtained in hanging-drop vapor-diffusion experiments. For this form, 4 μl drops containing a 1:1 mixture of protein solution (8.5 mg ml^{-1}) and 4–11% (w/v) PEG 6000 in 50 mM Tris HCl, 2 mM NaN_3, pH 9.5, are equilibrated at room temperature against a reservoir containing 1 ml of the latter solution. Crystals in shapes ranging from hexagonal tablets to hexagonal prisms appear within a week and grow to a maximum dimension of 0.3 mm on a hexagonal edge in about three weeks. The average size of crystals obtained in these experiments increases when the protein is preincubated with deoxyadenosine monophosphate (10 mM) and magnesium chloride (17 mM) at 277 K. X-ray diffraction from specimens of this crystal form show symmetries and extinction patterns consistent with space group $P6_222$ or its enantiomer, with $a = b = 93.5$, $c = 74.1$ Å and one molecule per asymmetric unit. Synchrotron X-ray diffraction by specimens of this crystal form has been observed to extend to 3 Å resolution.

We have also identified two other crystal forms of this recombinant HIV-1 RNase H. Crystals of a square-prism form grow in hanging-drop vapor-diffusion experiments when drops containing 1 μl protein solution (8.5 mg ml^{-1}) mixed with 1 μl 12% (w/v) PEG 6000 in 50 mM Tris HCl, 2 mM NaN_3, pH 9.5, are equilibrated against the precipitant at room temperature. Small

square-prism crystals appear in about a week. Unit-cell parameters (a = 63.0, b = 97.4, c = 114.4 Å) were measured by analysis of the limited X-ray scattering from a very small specimen (0.05 × 0.05 × 0.1 mm). The space group is either $P2_12_12$ or $P2_12_12_1$ and there are apparently four molecules in the asymmetric unit. Finally, we occasionally find crystals in the form of tiny cubes less than 0.01 mm on an edge growing under a wide variety of conditions, some similar to and some very different from those producing the other three crystal forms. So far, we have not found an example of this last form large enough to permit X-ray diffraction study.

Our structure determination efforts have to this point concentrated exclusively on the trigonal form. In solving the structure, we have applied essentially the same methods as those that resulted in the successful analysis of the inactive form (Davies *et al.*, 1991), in which the single isomorphous replacement method (Terwilliger & Eisenberg, 1983) was used with data from native and uranyl pentafluoride derivative crystals. A native data set was collected at room temperature from one crystal on the synchrotron beamline X31 at the EMBL Outstation at the DORIS storage ring, DESY, Hamburg. The DORIS ring was operated in main user mode with 4.37 GeV and 20–45 mA. Rotation images were recorded at a wavelength of 1.009 Å on an image plate scanner and processed with a version of the *MOSCO* integration package (Leslie, 1991). 19 985 observations of 8145 unique reflections were combined in good agreement (R_{sym} = 0.072). This data set is 95% complete with an effective resolution of 2.8 Å. For preparation of derivative crystals, a fresh sample of $K_3UO_2F_5$ reagent was synthesized by the method of Baker (1879). Data from derivative crystals were collected on a Siemens area-detector diffractometer mounted on a rotating-anode (Cu $K\alpha$) source. Phases calculated from isomorphous and anomalous differences observed in native data and data from three crystals of the $K_3UO_2F_5$ derivative gave rise to an interpretable electron density map. The construction of a molecular model into this density was made simpler by the availability of atomic coordinates from the homologous *E. coli* enzyme (Yang, Hendrickson, Crouch & Satow, 1990) and the published account of the HIV-1 RNase H domain (Davies *et al.*, 1991). Refinement of the initial model against the 2.8 Å data has been completed by application of restrained least-squares techniques (Hendrickson & Konnert, 1980). The crystallographic R value is 0.216; the root-mean-square difference between model bond lengths and the values to which they were restrained is 0.026 Å. Only 24 water molecule sites have been explicitly modeled. Atomic coordinates have been deposited with the Protein Data Bank (PDB) (Bernstein *et al.*, 1977).*
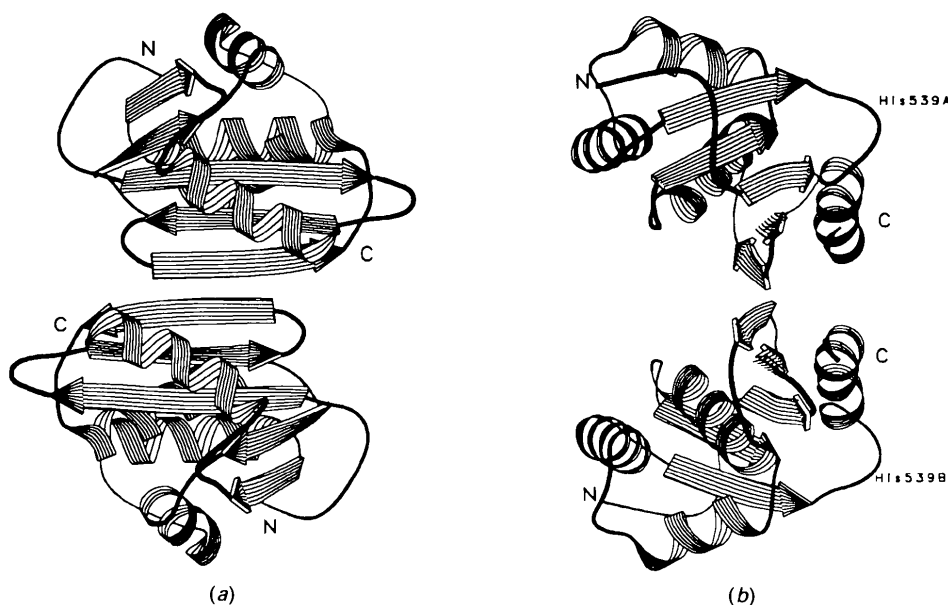
Fig. 1. Stylized ribbon representations of the two HIV-1 RNase H domains found in the crystallographic asymmetric unit. The two domains are related by a noncrystallographic twofold axis that is normal to the figure in view (a) and horizontal in the plane of the page in view (b). The two monomers interact at the outermost strand ($\beta3$) of the five-stranded β sheet to form a single large sheet that spans both domains. The position of His539 is indicated.

The crystal structure of active HIV-1 RNase H closely resembles that reported by Davies *et al.* (1991). The root-mean-square difference in positions of corresponding C^α atoms is 0.43 Å, and that of all corresponding atoms is 0.86 Å. The magnitudes of these differences reflect the essential similarity of two structures that were determined independently and refined with different protocols. They are also consistent with the magnitudes of differences observed for other independently determined structures (Ohlendorf, Treharne, Weber, Wendoloski & Salemme, 1991). Backbone atoms in the two independent molecules differ in position by only 0.49 Å, r.m.s., in the present structure.

As was pointed out by Davies *et al.* (1991), the two unique molecules are related by noncrystallographic dyads. One of these dyads is of special significance, however, because the two molecules it relates form a dimer (Fig. 1) in which five-stranded mixed β sheets in the two monomers are joined seamlessly to form a dimer-wide ten-stranded sheet in a manner reminiscent of some of the legume lectins [for example, pea lectin (Einspahr, Parks, Suguna, Subramanian & Suddath, 1986)]. Four hydrogen bonds typical of antiparallel β-structure involving residues 464 and 466 from outer β strands of each monomer highlight the interaction (Fig. 1), though a variety of other interactions between monomers related by this dyad appear to help in stabilizing the dimer. The significance of this doubled form of the enzyme to the seemingly solitary role it plays in the HIV-1 RT heterodimer is not immediately apparent, but its existence is an interesting anomaly worthy of note, especially as it was not remarked upon by previous investigators.

There is one significant difference between crystal structures of the active and inactive forms of the RNase H domain. The difference is in the extent of ordering in one important loop near the active site. Upon examination of the first electron density maps, we were encouraged by the presence of density in the vicinity of residues 538–542 that indicated some ordering of this loop that is disordered in the inactive form. This loop may be important to enzyme function, since it includes an invariant histidine residue (His539) (Doolittle *et al.*, 1989) shown by site-specific mutagenesis to be important to catalysis (Tisdale *et al.*, 1991). The density has persisted throughout the course of refinement, and has at various times led us to models ranging in completeness from partial backbone to complete with side chains. At one point about midway through the refinement, we substituted the model of the inactive form (PDB code 1HRH) for our own model and performed some additional refinement against our diffraction data. The 1HRH model contains no atomic positions for residues 538–542 in either molecule. This model gave an initial R value of 0.32 and was reduced without manual intervention to 0.26. In difference maps examined following this experiment, the largest density

feature by far was that in the vicinity of the disordered loop (Fig. 2). The existence of this density, arising as the result of a model that has no bias toward our own interpretation, dispelled any apprehensions that the density was a residual artifact of some previous model. Because crystals of the active and inactive forms of the enzyme are so nearly isomorphous, intermolecular crystal packing of molecules cannot explain this difference. It is also unlikely that residues added at the N-terminus to facilitate purification are directly responsible, since these residues are disordered and at least 15 Å from the loops in the crystal structure.

The structure of the 538–542 loop in HIV-1 RNase H bears little resemblance to the analogous loop in the *E. coli* homologue (Yang *et al.*, 1990). The disposition of the side chain of His539, suggested in the density for molecule *A*, is significantly displaced from its position in the *E. coli* enzyme, possibly several ångströms further above the cluster of conserved carboxylates that comprise the active site. The loops are involved in no intermolecular contacts arising from crystal symmetry that would favor this conformation over others.

The final refined model for one of the molecules includes a provisional complete model for the 538–542 loop, but only a partial model (residues 538 and 542) has been accepted for the other. We caution, further, that
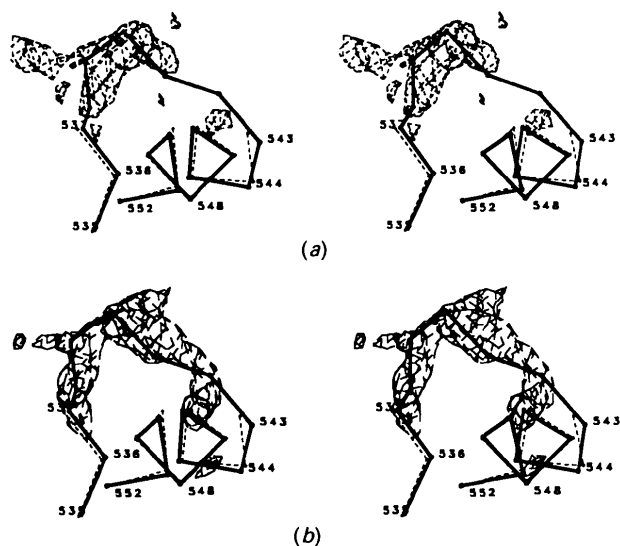

(a)

(b)

Fig. 2. Stereoscopic depiction of the C^α backbone with $+3\sigma$ difference density in the region of the loop between residues 538 and 542. (a) The density is excerpted from an $F_o - F_c$, α_c difference map. The Davies *et al.* (1991) model of the inactive RNase H domain (indicated by dashed lines), which lacks residues 538–542, was used in the calculation of model phases and structure-factor amplitudes, and the observed structure-factor amplitudes used were those of the active RNase H domain. (b) A similar difference map calculated from the final model of the active form with atoms from residues 538–542 omitted. The positive density suggests the presence of ordered atoms not accounted for by the model. All density is for molecule *A*, for which a provisional complete model for the loop has been created.

these provisional models are of limited utility. Despite our best efforts, we have been unable to construct a model that both fits the density and conforms with all the geometrical restraints that we expect of an acceptable model. One interpretation is that the density we see does not represent a single discrete conformation, but arises from a combination of conformational states. Such discrete disorder is well known in protein structures and has in many cases been resolved and modeled (Smith, Hendrickson, Honzatko & Sheriff, 1986), but not at the limited resolution (2.8 Å) of the diffraction data in this case. Either higher resolution data or data from an inhibited complex will be required to pinpoint the exact conformation of this critical loop. We are currently pursuing strategies along both of these lines.

In light of the evidence that the active preparation of the HIV-1 RNase H is not only enzymatically, but also structurally, distinct from inactive preparations, it may be advisable to make use of active material (Evans *et al.*, 1991) in future structural studies of enzyme-inhibitor complexes.

References

BAKER, H. (1879). *J. Chem. Soc.* **35**, 760-769.

BECERRA, S. P., CLORE, G. M., GRONENBORN, A. M., KARLSTROEM, A. R., STAHL, S. J., WILSON, S. H. & WINGFIELD, P. T. (1990). *FEBS Lett.* **270**, 76-80.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.

BREWER, L. C. & WELLS, R. D. (1974). *J. Virol.* **14**, 1494-1502.

CHATTOPADHYAY, D., EVANS, D. B., DEIBEL, M. R. JR, VOSTERS, A. F., ECKENRODE, F. M., EINSPAHR, H. M., HUI, J. O., TOMASSELLI, A. G., ZURCHER-NEELY, H. A., HEINRIKSON, R. L. & SHARMA, S. K. (1992). *J. Biol. Chem.* **267**, 14227-14232.

DAVIES, J. F. II, HOSTOMSKA, Z., HOSTOMSKY, Z., JORDAN, S. R. & MATTHEWS, D. A. (1991). *Science*, **252**, 88-95.

DOOLITTLE, R. F., FENG, D.-F., JOHNSON, M. S. & MCCLURE, M. A. (1989). *Q. Rev. Biol.* **64**, 1-30.

EINSPAHR, H., PARKS, E. H., SUGUNA, K., SUBRAMANIAN, E. & SUDDATH, F. L. (1986). *J. Biol. Chem.* **261**, 16518-16527.

EVANS, D. B., BRAWN, K., DEIBEL, M. R. JR, TARPLEY, W. G. & SHARMA, S. K. (1991). *J. Biol. Chem.* **266**, 20583-20585.

FINSTON, W. I. & CHAMPOUX, J. J. (1984). *J. Virol.* **51**, 26-33.

FURFINE, E. S. & REARDON, J. E. (1991). *J. Biol. Chem.* **266**, 406-412.

GOFF, S. P. (1990). *J. Acquired Immunodefic. Syndr.* **3**, 817-831.

HANSEN, J., SCHULZE, T., MELLERT, W. & MOELLING, K. (1988). *EMBO J.* **7**, 239-243.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Biomolecular Structure, Function, and Evolution*, edited by R. SRINIVASAN, pp. 43-57. Oxford: Pergamon Press.

HIZI, A., HUGHES, S. & SHAHARABANY, M. (1990). *J. Virol.* **175**, 575-580.

HOSTOMSKA, Z., MATTHEWS, D. A., DAVIES, J. F. II, NODES, B. R. & HOSTOMSKY, Z. (1991). *J. Biol. Chem.* **266**, 14697-14702.

HOSTOMSKY, Z., HOSTOMSKA, Z., HUDSON, G. O., MOOMAW, E. W. & NODES, B. R. (1991). *Proc. Natl Acad. Sci USA*, **88**, 1148-1152.

JOHNSON, M. S., MCCLURE, M. A., FENG, D. F., GRAY, J. & DOOLITTLE, R. F. (1986). *Proc. Natl Acad. Sci USA*, **83**, 7648-7652.

LARDER, B., PURIFOY, D., POWELL, K. & DARBY, G. (1987). *EMBO J.* **6**, 3133-3137.

LESLIE, A. G. W. (1991). Crystallographic Computing 5: From Chemistry to Biology, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 50-61. Oxford Univ. Press.

MITRA, S. W., CHOW, M., CHAMPOUX, J. & BALTIMORE, D. (1982). *J. Biol. Chem.* **257**, 5983-5986.

MITRA, S. W., GOFF, S., GILBOA, E. & BALTIMORE, D. (1979). *Proc. Natl Acad. Sci USA*, **76**, 4355-4359.

MIZRAHI, V., USDIN, M. T., HARINGTON, A. & DUDDING, L. R. (1990). *Nucleic Acids Res.* **18**, 5359-5363.

OHLENDORF, D. H., TREHARNE, A., WEBER, P. C., WENDOLOSKI, J. J. & SALEMME, F. R. (1991). Am. Crystallogr. Assoc. Annu. Meet., Toledo, Ohio, Series 2, Vol. 19, p. 36, Abstract G09.

RATTRAY, A. J. & CHAMPOUX, J. J. (1987). *J. Virol.* **61**, 2843-2851.

REPASKE, R., HARTLEY, J. W., KAVLICK, M. F., O'NEILL, R. R. & AUSTIN, J. B. (1989). *J. Virol.* **63**, 1460-1464.

RESNICK, R., OMER, C. A. & FARAS, A. J. (1984). *J. Virol.* **51**, 813-821.

SCHATZ, O., CROMME, F. V., GRUENINGER-LEITCH, F. & LEGRICE, S. F. J. (1989). *FEBS Lett.* **257**, 311-314.

SCHATZ, O., MOUS, J. & LEGRICE, S. F. J. (1990). *EMBO J.* **9**, 1171-1176.

SMITH, J. L., HENDRICKSON, W. A., HONZATKO, R. B. & SHERIFF, S. (1986). *Biochemistry*, **25**, 5018-5027.

TANESE, N. & GOFF, S. P. (1988). *Proc. Natl Acad. Sci USA*, **85**, 1777-1781.

TERWILLIGER, T. C. & EISENBERG, D. (1983). *Acta Cryst.* **A39**, 813-817.

TISDALE, M., SCHULZE, T., LARDER, B. A. & MOELLING, K. (1991). *J. Virol.* **72**, 59-66.

YANG, W., HENDRICKSON, W. A., CROUCH, R. J. & SATOW, Y. (1990). *Science*, **249**, 1398-1405.

International Union of Crystallography

Acta Cryst. (1993). D**49**, 427

Executive Secretary

It is with deep regret that the death of J. N. King is announced. Jim joined the Union as the first Executive Secretary in 1969 and gave loyal service until his untimely death on April 12. He had known about his illness for about fifteen months but had continued working with remarkable fortitude almost until the end. A full obituary will appear in *Acta Crystallographica*, Section A, in due course.