agreement with the data for the Great Red Spot and the White Oval BC on Jupiter lends support to our application of these simple models to the Great Dark Spot of Neptune for which there is less direct velocity data.

Finally, we point out that it would be of great interest to use the models we have presented here to perform fits to the "brown barge" type vortices on Jupiter (*18*), whose aspect ratios have been observed to vary by as much as 10% over a 15-day period, and for which the vorticity can be directly measured from velocity vectors.

REFERENCES AND NOTES

1. B. A. Smith *et al.*, *Science* **246**, 1422 (1989).
2. Recall that geostrophy is characterized by the balance between the pressure gradient and the Coriolis force. Quasigeostrophy describes the time evolution of a flow in geostrophic balance.
3. P. S. Marcus, *Nature* **331**, 693 (1988).
4. J. Sommeria *et al.*, *ibid.*, p. 689.
5. J. Pedlosky, *Geophysical Fluid Dynamics*, (Springer-Verlag, New York, 1979).
6. The modeling of a presumably internally heated, compressible, continuously stratified and baroclinic atmosphere with essentially two-dimensional incompressible and adiabatic models is supported in part by the normal mode analysis of R. K. Achterberg and A. P. Ingersoll [*J. Atmos. Sci.* **46**, 2448 (1989)].
7. S. Kida, *J. Phys. Soc. Japan* **50**, 3517 (1981).
8. L. M. Polvani and J. Wisdom, in *Topological Aspects of Fluid Motion*, H. K. Moffat Ed. (Cambridge University Press, Cambridge, 1990), and also L. M. Polvani and J. Wisdom, *Phys. Fluids A* **2**, 123 (1990).
9. N. J. Zabusky, M. H. Hughes, K. V. Roberts, *J. Comp. Phys.* **30**, 96 (1979).
10. R. Bulirsch and J. Stoer, *Numerische Math.* **8**, 1 (1966).
11. J. A. Nelder and R. Mead, *Comp. J.* **7**, 308 (1965).
12. T. E. Dowling and A. P. Ingersoll, *J. Atmos. Sci.* **45**, 1380 (1988).
13. As we have pointed out, the vorticity inside the Great Red Spot is not uniform [see (*12*) and J. L. Mitchel, R. F. Beebe, A. P. Ingersoll, G. W. Garneau, *J. Geophys. Rev.* **86**, 8751 (1981)]. The vorticity is small near the center and is maximum in an annular ring with peak values near 5×10^{-5} s^{-1}. We estimate the area weighted mean vorticity to be roughly $3 \pm 1 \times 10^{-5}$ s^{-1}.
14. S. S. Limaye, *Icarus* **65**, 335 (1986).
15. For a complete description of the linear and nonlinear stability analysis of the Kida solutions see D. G. Dritschel, *J. Fluid Mech.* **206**, 193 (1990) and also S. P. Meacham, G. R. Flierl, U. Send, *Dynam. Atmos. Oceans* **14**, 333 (1990).
16. L. M. Polvani, N. J. Zabusky, G. R. Flierl, *J. Fluid Mech.* **205**, 215 (1989). See also D. G. Dritschel, *Comp. Phys. Rep.* **10**, 77 (1990).
17. For background on chaotic behavior in Hamiltonian systems see, for example, the articles in *Hamiltonian Dynamical Systems*, R. S. MacKay and J. D. Meiss, Eds. (Adam Hilger, Bristol, 1987).
18. A. Hatzes, D. D. Wenkert, A. P. Ingersoll, G. E. Danielson, *J. Geophys. Res.* **86**, 8745 (1981).

19. We thank G. Flierl and P. Marcus for interesting conversations. The shape oscillations of the Great Dark Spot were originally observed by H. Hammel and L. Sromovski. J.W. gratefully acknowledges support from the NASA Planetary Geology and Geophysics Program (NAGW-706) and the NSF Presidential Young Investigator Program (AST-8857365). L.M.P. gratefully acknowledges support from the Atmospheric Science Division of the NSF under grant 8911459-ATM. E.D. and A.P.I. acknowledge the support of the Voyager Project, and the NASA Planetary Atmospheres Program (NAGW-58).

Structure of Ribonuclease H Phased at 2 Å Resolution by MAD Analysis of the Selenomethionyl Protein

WEI YANG, WAYNE A. HENDRICKSON, ROBERT J. CROUCH, YOSHINORI SATOW

Ribonuclease H digests the RNA strand of duplex RNA·DNA hybrids into oligonucleotides. This activity is indispensable for retroviral infection and is involved in bacterial replication. The ribonuclease H from *Escherichia coli* is homologous with the retroviral proteins. The crystal structure of the *E. coli* enzyme reveals a distinctive α-β tertiary fold. Analysis of the molecular model implicates a carboxyl triad in the catalytic mechanism and suggests a likely mode for the binding of RNA·DNA substrates. The structure was determined by the method of multiwavelength anomalous diffraction (MAD) with the use of synchrotron data from a crystal of the recombinant selenomethionyl protein.

RIBONUCLEASES H (RNASE'S H) CONSTITUTE A FAMILY OF enzymes that hydrolyze RNA molecules only when hybridized with complementary DNA strands (*1*). Although this activity is distributed broadly, the biological role of RNase H is poorly characterized, except in the instances of *Escherichia coli* and retroviruses. RNase H participates in DNA replication in *E. coli*; it helps to specify the origin of genomic replication by suppressing initiation at origins other than the locus oriC (*2, 3*); along with the 5′ → 3′ exonuclease of DNA polymerase I, it removes RNA primers from the Okazaki fragments of lagging strand synthesis (*4*); and it defines the origin of replication for ColE1-type plasmids by specific cleavage of an RNA preprimer (*5*). However, these do not appear to be vitally important activities since *rnh*$^-$ mutants of *E. coli* survive except in conjunction with certain other genetic defects (*3, 6, 7*). By contrast, RNase H activity is absolutely indispensable for retroviral replication. The RNase H of retroviruses is a component of reverse transcriptase (RT) (*8, 9*). Briefly stated, during reverse transcription the polymerase moiety of the transcriptase uses the genomic RNA as a template for synthesis of an RNA·DNA intermediate; RNase H then removes the RNA to free the complementary DNA strand which serves as the template for plus strand synthesis; and finally the resulting DNA duplex can be integrated into the host genome.

W. Yang and W. A. Hendrickson are in the Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032 where W.A.H. is also an investigator with the Howard Hughes Medical Institute. R. J. Crouch is at the Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892. Y. Satow is at the Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan, and also at the Photon Factory, National Laboratory for High Energy Physics, 1-1 Oho, Tsukuba-shi, Ibaraki 305, Japan.

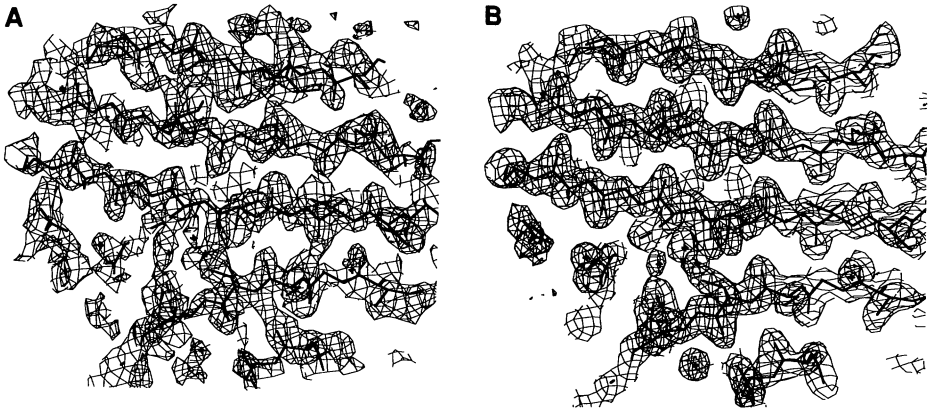


Fig. 1. Electron density distribution in the β sheet. (**A**) MAD phased map at 2.2 Å resolution. (**B**) Refined $2|F_{obs}| - |F_{calc}|$ map at 2.0 Å resolution. Contours are drawn at the level of 0.9 σ. The same point of view (perpendicular to the sheet) and density clipping are used in each case to emphasize main-chain features. The model refined at 2 Å resolution is superimposed in both maps. Good density definition of carbonyl groups is evident in both maps.

Significantly, RNase H cleavages at the origin of plus strand synthesis are species specific (*10*). This crucial role in infection by retroviruses makes RNase H an obvious target for drugs against retroviral diseases.

We have undertaken a crystallographic study of RNase H in order to elucidate the structural basis for its action. We chose the *E. coli* enzyme because of its essential similarity with the retroviral enzymes and because of its relative simplicity and an available expression system (*11*). In common are the requirement for RNA·DNA hybrids as substrates and divalent cations (usually Mg^{2+}) for catalysis, very limited base sequence specificity, endonucleolytic action (*12*), and the release of 5′ phosphate and 3′ hydroxyl products (*9, 13*). The main distinction is that, whereas *E. coli* RNase H is a 17,580-dalton monomer, RNase H enzymes of retroviruses are integrated into much larger RT molecules, some of which are dimeric (*9*). However, sequence comparisons and mutational studies localize the RNase H portion to the COOH-terminal third of RT polypeptide chains (*14–16*). There is significant similarity in the RNase H sequences from microbes, retroviruses and retrotransposon elements (*17*). Site-directed mutations of absolutely conserved residues, done both in Moloney murine leukemia virus (MMLV) RNase H and in the *E. coli* enzyme, show that Asp^{10}, Glu^{48}, and Asp^{70} (*E. coli* numbering) are crucial for activity (*18, 19*).

The central problem in a crystal structure determination is that of evaluating phase angles for the diffracted x-ray amplitudes. For biological macromolecules, the usual method for solving this problem de novo has been isomorphous replacement with heavy atoms. We and others have developed (*20–21*) and applied (*22–24*) the alternative method of multiwavelength anomalous diffraction (MAD) in which the scattering strength at a few special sites is changed physically by change of x-ray wavelength rather than chemically by atomic substitution. MAD phasing has the advantages of perfect isomorphism and an algebraically exact analysis. The method does require that appropriate diffraction centers be present and that the x-ray source be tunable as at synchrotrons. Selenomethionyl proteins provide an attractive vehicle for MAD phasing (*25*).

We have previously described the expression and crystallization of *E. coli* selenomethionyl RNase H (*26*). We now describe the structure of RNase H and the MAD phasing method by which the structure was determined. First, we report the measurement and analysis of synchrotron data recorded on imaging phosphor plates. Next we present the resulting atomic structure in comparison with related proteins. Finally we discuss implications of these results for substrate binding and catalytic action. An independent study of *E. coli* RNase H in another crystal form has also been reported (*27*).

Synchrotron measurements. Selenomethionyl RNase H protein was expressed in an auxotrophic strain of *E. coli* grown with selenomethionine replacing methionine in the medium (*26*). The purified *E. coli* RNase H with complete selenomethionine substitution was fully active, and it could be crystallized by vapor diffusion in hanging droplets after microseeding from a natural RNase H crystal (*26*). Type I RNase H crystals are in space group $P2_12_12_1$ with unit cell parameters of $a = 41.79$ Å, $b = 86.34$ Å, and $c = 36.31$ Å. The mother liquor of these crystals is at pH 8.8 and contains EDTA, DTT (dithiothreitol), and ammonium sulfate as well as the PEG (polyethylene glycol) precipitant (*26*). A crystal with extreme dimensions of 0.62 mm by 0.26 mm by 0.12 mm was freshly mounted from a droplet into a 0.7-mm glass capillary immediately before the data collection, with the *b* axis (of medium length) roughly along the capillary axis. All measurements used in this analysis were made from this single crystal.

The MAD diffraction data were collected at beamline 14 A (*28*) of the Photon Factory (KEK, Japan) with radiation from a superconducting vertical wiggler in the storage ring that was operated at 2.5 GeV with electron currents from 270 mA to 150 mA. The monochromatic beam was achieved as described for streptavidin (*23*). The slit after the monochromator was set at 2 mm so as to assure a uniform 0.5-mm beam with satisfactory energy resolution. The crystal was mounted on the goniometer in a dry gas stream at ~5°C and aligned so that mirror-related Bijvoet mates would be recorded simultaneously. X-ray absorption spectra, measured by

Table 1. Multiwavelength data statistics. The data were recorded on Fuji HR and HR-III plates (201 by 252 mm) at a crystal-to-plate distance of approximately 185 mm. Thirty-two frames were collected in oscillations (16°/min) of 3.1° about the *b* axis (at $\chi = 0°$) with 0.2° overlap between successive photographs. Four more frames were collected by oscillations of 3.5° about the *a* axis (at $\chi = -90°$) in order to record the blind cone region. About four to six consecutive frames of diffraction data at one wavelength were taken before the recording of the equivalent diffraction data at the next wavelength. After each change of the monochromator setting angle for a desired wavelength, optical elements were realigned and optimized. The exposure time (number of scans) for each frame was adjusted to approximately uniform radiation dosage, which was monitored by an ion chamber, and to compensate for the differing sensitivity of different types of imaging plates. The digitized images were reduced to integrated intensities with DENZO (*33*). Only the results for data within Bragg spacing of 2.0 Å are included here. Unique reflections were defined by point group symmetry *P222* (not *Pmmm*) to distinguish Bijvoet mates.

$$R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$$

Wavelength (Å)	Imaging plates (No.)	Measurements (No.)	Unique reflections* (No.)	R_{sym}*
$\lambda_2 = 0.9802$	35	61710	17062 (99.5)	0.045 (97.9)
$\lambda_3 = 0.9795$	35	62353	17069 (99.6)	0.042 (99.1)
$\lambda_4 = 0.9000$	36	69549	17086 (99.7)	0.040 (98.6)

*Completeness; numbers in parentheses are percentages.

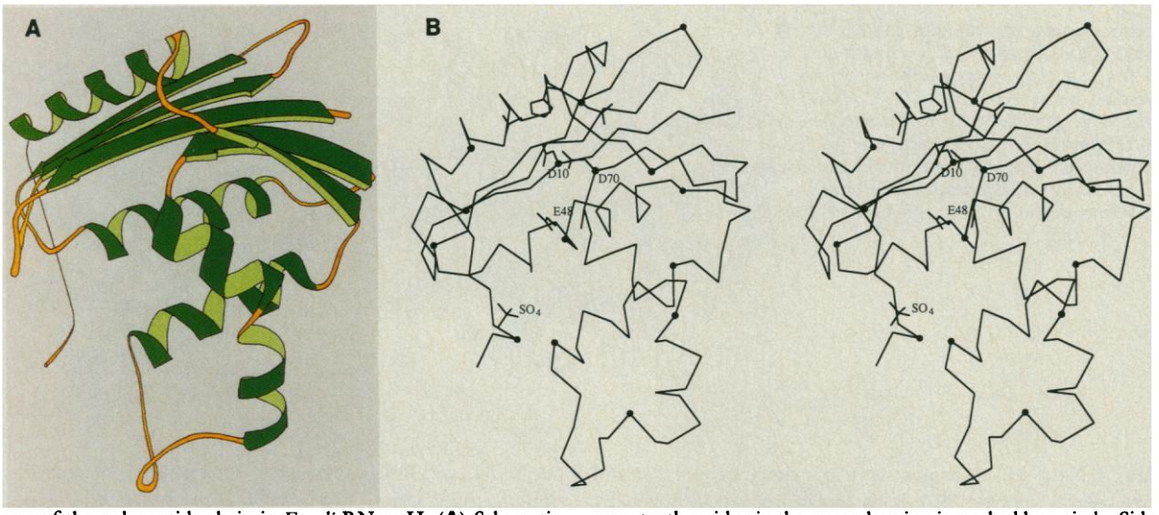


Fig. 2. Course of the polypeptide chain in *E. coli* RNase H. (**A**) Schematic ribbon drawing. (**B**) Stereodrawing of the α-carbon backbone from residue 4 to 152. The point of view in both cases is along Z′ after a 25° counterclockwise rotation about Y and a 12° clockwise about X. The Cα of every tenth residue in the stereodrawing is marked by a circle. Side chains are drawn for the seven residues common to all bacterial and retroviral RNase's H and those for the carboxyl triad are labeled. The bound sulfate ion is also shown.

fluorescence with a scintillation detector, were used to select monochromator setting angles for the diffraction measurements. These angles corresponded to nominal wavelengths of 0.9802 Å at the inflection point, 0.9795 Å at the peak value, and 0.9000 Å at a remote point. The diffraction data were collected on imaging phosphor plates (*29*) by oscillation photography with procedures established in an earlier study (*30*). Imaging plates were scanned on a laser reader system (*31*), and digitized images were stored on magnetic tape. The complete set of x-ray diffraction data at three wavelengths was collected in 108 exposures over a period of 12.8 hours during which the crystal was irradiated for a total of 2.4 hours.

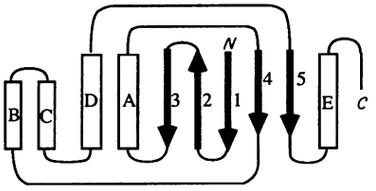
The x-ray fluorescence spectra were reduced to anomalous scattering factors as described (*22*). In brief, the experimental spectra near the edge were spliced into calculated atomic absorption spectra after least-squares fitting to the theoretical values remote from the edge. These composite spectra were then converted to the f'' and f' components of anomalous scattering by the absorptive and dispersive transformations. No pleiochroism as evident in the spectra measured with polarization along the three cell axes, but macroscopic isotropy is not unexpected for multiple sites in an orthorhombic system. The energy resolution achieved here with the 2-mm slit was slightly better than that with streptavidin (*23, 32*), and hence similar intrinsic anisotropy is expected.

Diffraction data to 1.7 Å Bragg spacings were reduced to integrated intensities with the program DENZO (*33*). No correction was made for the small effect of absorption at these wavelengths. Reflections at each wavelength were brought to a common scale with the use of SCALEPACK to apply a scale factor and an isotropic B value for each imaging plate. Only the 2 Å data were processed at the scaling stage. Symmetry equivalent reflections were kept separate after this scaling (Table 1).

Structure analysis by the MAD phasing method. Procedures implemented in the MADSYS programs (*34*) developed in our earlier applications of MAD phasing (*21–24*) were adapted and enhanced. First, the data from each wavelength and orientation were sorted by LEXICO without merging of replicates but with pairing of mirror-related Bijvoet mates. Next, parameterized local scaling was performed by ANOSCL to reduce systematic errors between Bijvoet mates and then by WVLSCL to bring the data from all wavelengths to a common and approximately absolute scale. Anomalous diffraction ratios were then consistent with expectation from the composition and scattering factors (Table 2). Each scaled reflection set, consisting of Bijvoet mates at the three wavelengths, was then fitted by MADLSQ with the use of the experimentally determined f' and f'' values.

The algebraic analysis executed by MADLSQ reduced the diffraction information to its essentials: structure factor amplitudes $|^{\circ}F_T|$ due to the wavelength invariant "normal" scattering from the total structure, amplitudes $|^{\circ}F_A|$ from just the anomalous centers, and the associated phase differences $\Delta\varphi = {}^{\circ}\varphi_T - {}^{\circ}\varphi_A$. There were an average of 2.2 replicate determinations per unique reflection set, and we developed a new program MERGIT to merge equivalent reflections

Fig. 3. Schematic diagram of the folding topology and nomenclature for *E. coli* RNase H.



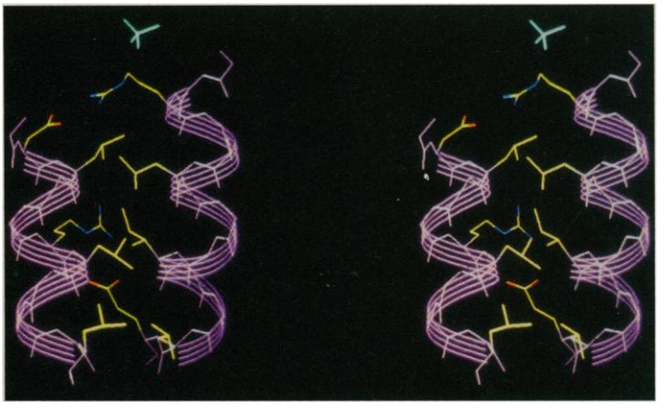


Fig. 4. Stereodrawing of interactions between parallel helices A and D. Helix backbones are represented by ribbons. Bonds between side chain carbon atoms are drawn in yellow, those to oxygen atoms are in red, and those to nitrogens are in blue.

Table 2. Anomalous diffraction differences and scattering factors for selenomethionyl RNase H. Bijvoet differences ratios are given in diagonal elements with centric values in parenthesis, and dispersive difference ratios are given as off-diagonal elements. Observed Bijvoet and dispersive ratios are reported as [rms ($\Delta F \pm h$)]/[rms ($|F|$)] and [rms ($\Delta F_{\Delta\lambda}$)]/[rms ($|F|$)], respectively. Expected values were computed as $q(2f'')$ for Bijvoet ratios and as $q|f'(\lambda_i) - f'(\lambda_j)|$ for dispersive ratios where $q = (N_A/2N_P)^{1/2}/Z_{eff}$ with specific values of $N_A = 3$ (assuming a disordered NH_2-terminus), $N_P = 1237$ total nonhydrogen protein atoms, and $Z_{eff} = 6.7e$ as the effective average normal scattering at zero angle from nonhydrogen atoms. Scattering factors at $\lambda = 0.9000$ Å are theoretical values and those at the edge are the experimental values derived from the fluorescence spectra.

Wavelength (Å)	Observed ratio ($\infty > d > 3.0$ Å)			Observed ratio ($3.0 > d > 2.0$ Å)			Expected ratio ($d = \infty$)			Scattering factors	
	0.9000	0.9795	0.9802	0.9000	0.9795	0.9802	0.9000	0.9795	0.9802	f''	f'
0.9000	0.033 (0.018)	0.044	0.057	0.060 (0.031)	0.071	0.086	0.034 (0.000)	0.026	0.035	3.284	−1.622
0.9795		0.043 (0.015)	0.032		0.076 (0.028)	0.056		0.047 (0.000)	0.009	4.504	−6.702
0.9802			0.026 (0.017)			0.047 (0.030)			0.028 (0.000)	2.662	−8.422

with eliminations and weighted averaging based on refinement statistics, primarily the estimated accuracy of $|^{\circ}F_A|$ and $\Delta\varphi$ values. The merged data included 9342 reflections (99.8 percent of expected total 2 Å data) with $\Delta\varphi$ evaluations for 8896 (95.0 percent). For 5072 duplicate pairs meeting the cut-off criteria, agreement R values were 0.044 and 0.377, respectively, for $|^{\circ}F_T|$ and $|^{\circ}F_A|$ and the average discrepancy between $\Delta\varphi$ values was 38.6°. The 7273 $|^{\circ}F_A|$ values smaller than 250 (maximal $|^{\circ}F_A|$ for four selenium atoms is 504) were used in a Patterson synthesis from which two selenium sites were immediately identified on Harker sections. The best MULTAN solution confirmed these sites at 6.7 Å apart. Two additional sites (2.6 Å apart), presumed to be from a third selenomethionine residue, were found in difference Fourier maps. Positions and B values for these four sites (occupancies fixed at 1.0, 1.0, 0.62, and 0.38) were refined by ASLSQ to $R = 0.334$ against 7307 $|^{\circ}F_A|$ data. Finally, MADFAZ was used to produce phases and figure-of-merit weights for $^{\circ}F_T$ coefficients ($\bar{m} = 0.78$) from the merged MADLSQ data and calculated selenium phases.

Fourier syntheses computed with MAD phases based on the enantiomorphic selenium alternatives revealed the correct hand for the structure as the one with physically reasonable electron density. After examination of maps at various resolutions, we settled on a 2.2 Å resolution synthesis for interpretation because of its excellent carbonyl definition. A portion of this map is shown in Fig. 1 together with the same portion after refinement at 2.0 Å resolution. We traced the polypeptide chain in a minimap plotted on transparent sheets and read out coordinates for all α-carbon and most carbonyl oxygens for residues 5 to 152. A complete model was then built into the electron density map with the display program FRODO (*35*). The agreement factor R for this starting model was 0.476 for data from 5.0 to 2.2 Å spacings, and refinement with PROLSQ (*36*) for 21 cycles brought the R-value down to 0.318 for these data. After another 15 cycles of refinement with tightly restrained individual isotropic B values, the R-value came to 0.293 for data from 5.0 to 2.0 Å with stereochemical ideality typified by 0.021 Å deviation from ideal bond lengths. Several iterations of model rebuilding and further refinement ensued. Electron densities for the NH_2-terminal residue and for the last three at the COOH-terminus remained unclear after this refinement; presumably these residues are disordered. The present model comprises 1214 protein atoms from residues 2 to 152, including alternate conformations for Met^{142}, and 86 water sites. It has been refined to $R = 0.198$ against 8646 data having $|F| > 4\sigma$ in the range of 10 to 2 Å spacings, and it has an rms bond ideality of 0.015 Å and an average $|\Delta B|$ discrepancy of 0.75 Å^2 for bonded main-chain atoms. Coordinates for the model described above have been deposited in the Protein Data Bank (entry name 1RNH), and the associated diffraction data will be deposited after processing to 1.7 Å spacings.

Overall structure. The tertiary folding of *E. coli* RNase H (Fig. 2) is that of an α-β protein. An analysis of the hydrogen-bonding pattern (*37*) in the refined model places 41 percent of the 155 residues in helices and 28 percent in β strands. This can be compared with values of 25 percent α and 41 percent β found by circular dichroism (*19*). The RNase H fold bears no substantial similarity to other known structures, including other nucleases or ricin, with which homology has been suggested (*38*).

A mixed, five-stranded β sheet is a central feature of the RNase H structure. The first three β strands (residues 4 to 13, 18 to 27, and 32 to 42 for strands 1, 2, and 3, respectively) follow one another directly in an antiparallel course. Strands 4 and 5 (residues 64 to 69

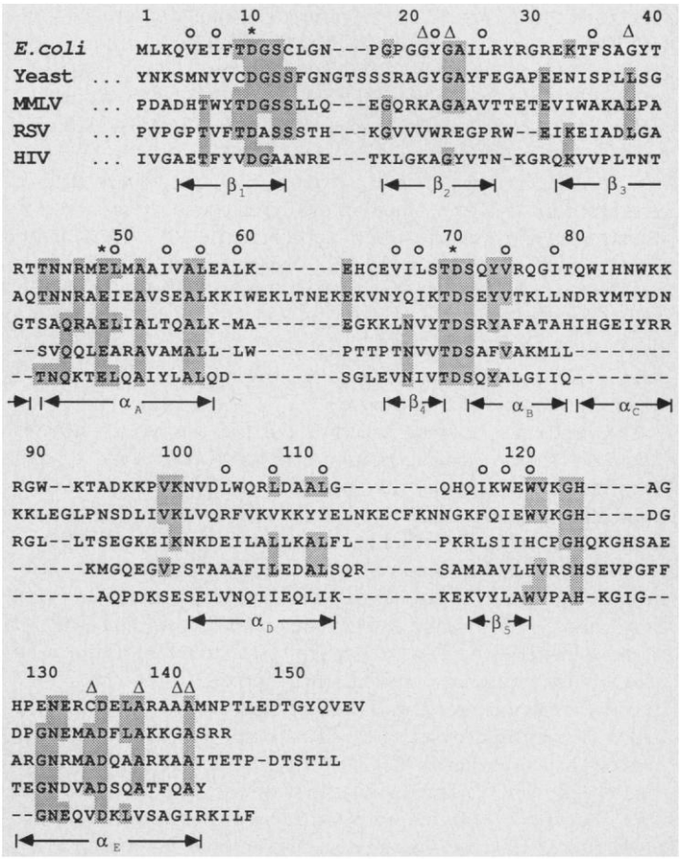


Fig. 5. Alignment of selected RNase H amino acid sequences. These sequences are from the bacterium *Escherichia coli* (*11*), the yeast *Saccharomyces cerevisiae* (*7*), and the retroviruses Moloney murine leukemia virus (MMLV), Rous sarcoma virus (RSV) and human immunodeficiency virus (HIV) (*44*). A topologically constrained alignment has been made which avoids insertions and deletions within α helices and β strands as designated for *E. coli* RNase H. Residues from the hydrophobic core of RNase H are marked by (O) and those at the E-helix to β-sheet contact are indicated by (Δ). Positions of identity in 3, 4, or 5 of the aligned sequences are shaded.

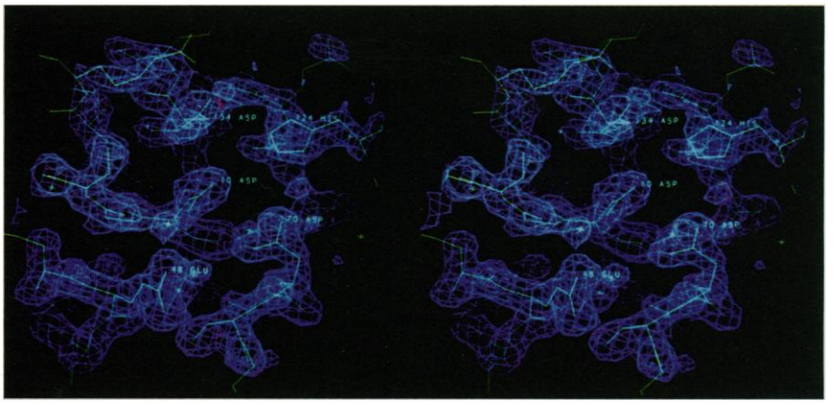


Fig. 6. Stereodrawing of the catalytic site structure. Electron density contours are drawn at 2.3 σ in a fragment ΔF map computed in two passes with residues 9 to 13 and 47 to 51 (main chain only for 47 and 50) and waters 209 and 283 omitted in one synthesis and residues 69 to 72, 124 and 130 to 138 (main chain only for 135 and 136) and waters 269 and 272 omitted in the second synthesis. The refined model is superimposed on the density with five of the seven conserved residues labeled. The carboxylates of Glu^{48}, which originates from helix A, and of Asp^{10} and Asp^{70}, which are at the point of divergence of β strands 1 and 4, form a triangle. An oxygen from Glu^{48} is at one vertex ~5 Å from counterparts of Asp^{10} and Asp^{70}, which are in turn paired head-to-head at a separation of ~4 Å. The other carboxyl oxygen of Glu^{48} is hydrogen-bonded with the hydroxyl group of Ser^{71}. Side chain atoms of His^{124} and Asp^{134} are within 4 Å of the Asp^{10} carboxylate, and the side chain of Asn^{130} (obscured) is just over 6 Å from carboxylate oxygens of both Asp^{70} and Asp^{10} with an intervening water molecule.

and 115 to 120, respectively) are juxtaposed in parallel with the first strand, but intervening helical elements separate these successive strands. Thus, the strand order in the sheet is 32145 (↓ ↑ ↓ ↓ ↓). There also are five helices A, B, C, D, and E, named again in order of occurrence (Fig. 3). Helix A (residues 43 to 58), between strands 3 and 4, is central. Helices B, C, and D (residues 71 to 80, 81 to 88, and 100 to 112, respectively) intervene between strands 4 and 5. One could consider the B and C helices to be a single, sharply kinked (57.5°) helix. They are separated at Gln^{80} which has a conformation ($\varphi = -123°$ and $\psi = -44°$) like that at residue 110 in myohemerythin (*39*). Helix D runs approximately parallel (30.3°) with helix A and is in contact with a closest approach of 9.5 Å between helix axes occurring near the helix centers (residues 53 and 107). Whereas helices A, B, C, and D are all on one side of the β sheet, helix E (residues 127 to 142) is alone on the other side where it crosses β strands 1, 2, and 3 at an angle of approximately 40°. This helix-to-sheet contact is very close with axes crossing at from 4.2 to 5.3 Å by virtue of small contact residues (Gly^{21}, Gly^{23}, Gly^{38}, Cys^{133}, Ala^{137}, Ala^{140}, Ala^{141}). Helices A and D also contact the sheet (especially closely at highly conserved Ala^{51} and Ala^{55}), but here the helices and strands are more nearly parallel. An extended COOH-terminal appendage completes the structure.

The interaction between helices A and D is particularly notable (Fig. 4). On the outside, doubly hydrogen-bonded salt bridges (Arg^{46} to Asp^{102}, and Glu^{57} to Arg^{106}) tie the two helices together. Internally, there is a ladder of leucine and isoleucine interactions (Leu^{49} to Leu^{103}, Ile^{53} to Leu^{107}, and Leu^{56} to Leu^{111}). These hydrophobic residues join with others (Leu^{59} immediately after helix A, Val^{74} and Ile^{78} of helix B, Val^{5} and Ile^{7} of strand 1, Tyr^{22}, Leu^{26}, and Tyr^{28} of strand 2, Phe^{35} of strand 3, Val^{65} and Leu^{67} of strand 4, and Ile^{116}, Trp^{118} and Trp^{120} of strand 5) to form an extensive hydrophobic core. This core is protected not only by the direct electrostatic interactions between helices A and D and by the hydrogen-bonded network of the sheet, but also by hydrogen bonds between helices A and B (Glu^{48} to Ser^{71}), B and D (Tyr^{73} both to Asn^{100} and Trp^{104}), and C and D (Lys^{86} to Asp^{108}). Aromatics Trp^{118} and Trp^{120} of the core are partially exposed. A more extensive patch of partially exposed aromatics is formed by Tyr^{73}, Trp^{81}, Trp^{85}, Trp^{90}, and Trp^{104}.

For the most part, the connecting loops between elements of secondary structure also have well-defined structure. Most prominent of these is the protruding handle at the extreme end of the molecule. This consists of residues 90 to 99, which connect helices C and D, and it features a contorted β hairpin with a type I turn at residues 92 to 95 and two additional hydrogen bonds connecting ministrands 90 to 92 and 96 to 98. Two additional loops probably have significance in substrate binding. One connects strands 1 and 2 and includes a cis peptide at Pro^{17}, and the other is between strand 5 and helix E and includes the conserved His^{124}. These loops protrude out from the molecule and define the rim of a concave surface. Although atomic mobilities (B values) in the loops are generally higher than in regular structure, only the loop between strands 2 and 3 and the extreme termini are ill-defined.

Comparison with retroviral ribonuclease H. Sequence comparisons suggest that a structural similarity underlies the functional relation between *E. coli* and retroviral RNase H. The Doolittle alignment (*17*) has identical amino acids at nine positions in 13 retroviral RNase H sequences, and seven of these nine are also preserved in *E. coli* RNase H. Moreover, when extended to include retrotransposons and other retrovirus-like genetic elements, four residues (Asp^{10}, Glu^{48}, Asp^{70}, and Asp^{134}) are invariant in all 26 sequences. We have now made a topologically restricted alignment of representative sequences in which gaps are not permitted within elements of secondary structure in *E. coli* RNase H (Fig. 5).

Our restricted alignment has a 28 percent identity in the comparison of *E. coli* RNase H with the yeast enzyme, 27 percent with MMLV, 19 percent with RSV, and 24 percent with HIV. Other comparisons among this set range from 18 to 25 percent. While such percentages are not large [in contrast, the *E. coli* and *Salmonella* proteins are 93 percent identical (*7*)], the case for a homologous family is strong. This hypothesis is strengthened by the spatial distribution of conserved residues in the *E. coli* structure. First, all seven of the residues that are identical among all retroviral and bacterial sequences (Asp^{10}, Glu^{48}, Asp^{70}, Ser^{71}, His^{124}, Asn^{130}, and Asp^{134}) are contiguous in a surface concavity surrounding the essential carboxyl triad (Fig. 2). Second, substitutions for residues that form the hydrophobic core and the contact between helix E and β sheet in *E. coli* RNase H are generally conservative. Greatest variability is found in the loops and the COOH-terminal extension.

Given the overall similarity of RNase H molecules, the differentiating distinctions become of interest. The yeast enzyme is distinguished by several insertions relative to the others. A more profound difference, however, is that between *E. coli* RNase H and the HIV and RSV molecules. The latter have major deletions that can be accommodated without disruption of the core only by the replacement of helix C and the β hairpin handle with a new connection between helix B and helix D. Coincidently, both HIV and RSV RT molecules are dimeric, and HIV RNase H activity requires intact folding of some polymerase region (*16*). In contrast, MMLV and *E. coli* RNase's H are very much alike except for a five-residue lengthening of the connection between strand 5 and helix E, and isolated RNase H constructs from MMLV are active (*14*).

Implications for substrate binding and catalysis. The RNase H analyzed here was crystallized in the absence of nucleotides or divalent cations, and thus we have no direct identification of the catalytic site. Nevertheless, this site can be securely located in the three-dimensional structure with reference to conserved residues and the results of site-directed mutagenesis. As mentioned, above, all seven residues that are absolutely conserved in known retroviral and bacterial RNase H sequences are clustered together in a depression on the *E. coli* RNase H surface. Kanaya *et al.* (*19*) have shown that *E. coli* RNase H is exquisitely sensitive to mutations of three of these residues—Asp^{10}, Glu^{48}, and Asp^{70}. Amidation of any of these carboxylates (Asp→Asn or Glu→Gln) eliminated activity, and single methylene shifts in side chain length (Asp→Glu or Glu→Asp) reduced k_{cat} to 1 to 8 percent that of the wild type without change of K_m. This indicates an essential role in catalysis, not merely in binding. Although disparate in the sequence, these key residues are at the heart of the conserved depression (Fig. 6) with carboxylates deployed in a triangle of 4 to 5 Å on an edge. Functional groups of the other fully conserved residues are each within 6 Å of the carboxyl triad (within 4 Å for all but Asn^{130}). Mutations made at His^{124} and Asn^{130} left the enzyme active, but with reduced substrate binding affinity and a lower catalytic rate; changes at Ser^{71} and Asp^{134} were quite benign (*19*).

While the evidence at this point is insufficient to specify a mechanistic role for the carboxyl triad in RNase H, we can limit the possibilities with reference to other enzymes. (i) The mechanism of pancreatic and microbial ribonucleases can be excluded since the cyclic phosphate intermediate cannot yield 5′ phosphate products (Table 3). (ii) Although apposed carboxyl groups are responsible for the hydrolytic action of lysozyme and aspartyl proteases, a carboxylate must become protonated to activate a water molecule for attack. This is unlikely for any of the RNase H carboxylates at the optimum of ~pH 8.0 (*12*). (iii) Another alternative is provided by deoxyribonuclease I wherein carboxyl groups coordinate a metal ion (Ca^{2+}) and a Glu-His-hydroxyl relay attacks the P–O3′ bond, but geometric placements and the mutation results preclude such a critical role for histidine in RNase H.

Discounting a direct catalytic role for His^{124}, RNase H is like ribozymes in having only anionic functional groups and requiring divalent cations (Table 3). Fortunately, a suitable structural precedent for such reactions is available from studies on the 3′→5′ exonuclease of DNA polymerase I (*40*). Here nucleolytic activity requires two divalent metal ions and four carboxylates. In the proposed mechanism a hydroxyl ion, activated by a cation, attacks to form a metal-stabilized pentacovalent phosphorus intermediate that cleaves and inverts. Our results lend support for this kind of mechanism, which applies equally for RNA cleavage. Clearly, substrates cannot bind without intervening counterions. There are no divalent cations in our structure (the carboxyl triad interacts with Lys^{87} from a symmetry mate), but crystals are destroyed when soaked with Mg^{2+}, and a new type of crystal grows in the presence of free Mg^{2+}. We expect conformational changes in the catalytic center upon metal and substrate binding to be limited to side chains since all implicated residues emanate from core elements of this single-domain structure. Whether more than one Mg^{2+} site is involved remains to be seen.

Since a sulfate ion binds in the place of a nucleotide phosphate in RNase A and acts as an inhibitor (*41*), we expected that a sulfate ion might mark the active site in type I RNase H crystals. Type I crystals are grown in the presence of 60 mM sulfate, and type II crystals grown with NaCl as the ionic medium can be converted to type I by addition of phosphate (*26*). Indeed, a sulfate ion is bound with full occupancy and low atomic mobility ($B \sim 14$ Å^2), but not at the catalytic site. Instead, it is located at the NH_2-termini of helices A and D (Fig. 4) in a site 13.7 Å away from the center of the carboxyl triad (Fig. 2). Besides apparent stabilization by the helix dipoles, this sulfate ion is coordinated by four protein side chains and three water molecules. Two of the protein ligands are quite conservative; Thr^{43} often goes to Ser and Asn^{45} is always replaced by Gln in retroviruses. Phosphoribosyl linkages can be modeled to replace water ligands, and we suggest that this sulfate position marks a strong substrate binding site. It appears that sulfate or phosphate binding can be accomplished without appreciable conformational change; we have solved the type II crystal structure at 4 Å resolution and find that a rigid body transformation relates the two polymorphs.

Table 3. Attributes of diverse ribonucleases. A number of different ribonucleases have been characterized, and crystallographic analyses have been carried out on pancreatic ribonuclease A (*41*) and on several microbial ribonucleases (*45*). The microbial substrate specificities shown here are for barnase (Ba) and RNase T1 (T1). In keeping with the marked distinctions in catalytic parameters, there is no structural resemblance whatever between RNase H and either RNase A or the microbial structures. Judging simply by the similarity in products and the catalytic requirements for acidic groups and divalent cations. RNase H may be somewhat similar in action to the ribozymes RNase P and the *Tetrahymena* self-splicing RNA (*46*); ss, single stranded; ds, double stranded.

Type	Substrates	Products	Catalytic groups
Pancreatic A	s-s RNA (C, U)	5′ OH, 3′ PO_4	His, His, Lys
Microbial	s-s RNA	5′ OH, 3′ PO_4	His, Glu, Arg
Bacterial (Ba)	(N)		
Fungal (Tl)	(G)		
Ribonuclease H	d-s RNA:DNA	5′ PO_4, 3′ OH	Asp, Asp, Glu Mg^{2+}
Ribozymes	s-s RNA		
E. coli RNase P	pre-tRNA	5′ PO_4, 3′ OH	RNA, Mg^{2+}
Tetrahymena	pre-rRNA	5′ RNA, 3′ OH	

Model complex with an RNA·DNA hybrid helix. The very specificity of RNase H for RNA·DNA hybrids argues for heteroduplex substrates for this enzyme. While both the nucleic acid and the protein might change conformation on binding, modeling the association from free components should provide useful insight. As a mock substrate, we have constructed a pure hybrid decamer from an A-type mixed hybrid (*42*). Examination of the RNase H structure reveals a surface (Fig. 7A) that has features complementary to this hypothetical substrate; (i) the relief of this surface has two groove-like depressions separated by a slight central ridge; (ii) this putative catalytic site and the sulfate binding site are in these separate depressions; and (iii) the electrostatic potential surface (Fig. 7B), calculated with the sulfate removed and without counterions for the carboxylates, is strongly electronegative at the carboxyl triad but is generally positive in and along the outer rim of the depressions.

We have docked the hybrid model onto RNase H by simply superimposing a phosphate from the DNA strand onto the sulfate position and pivoting around this fixed point to bring the RNA strand close to the catalytic site. This places each strand along a depression with the minor groove of the hybrid helix fitted on the ridge (Fig. 7, C to E). The interaction spans about four nucleotides on each strand, but two base pairs intervene between the sulfate site and the putative scissile O3′–P bond located near Asp^{10}, Glu^{48} is in the minor groove of this model. Cations, which must accompany substrate binding, have not been modeled. This rigid body docking leaves a few clashes, mostly involving the strand 1 to 2 loop and Gln^{72}, which might be relieved by conformational adjustment.

While this model is too crude to permit detailed analysis, it does suggest that specificity for hybrid molecules could be generated by specific deoxyribose and 2′-hydroxyl interactions along the respec-

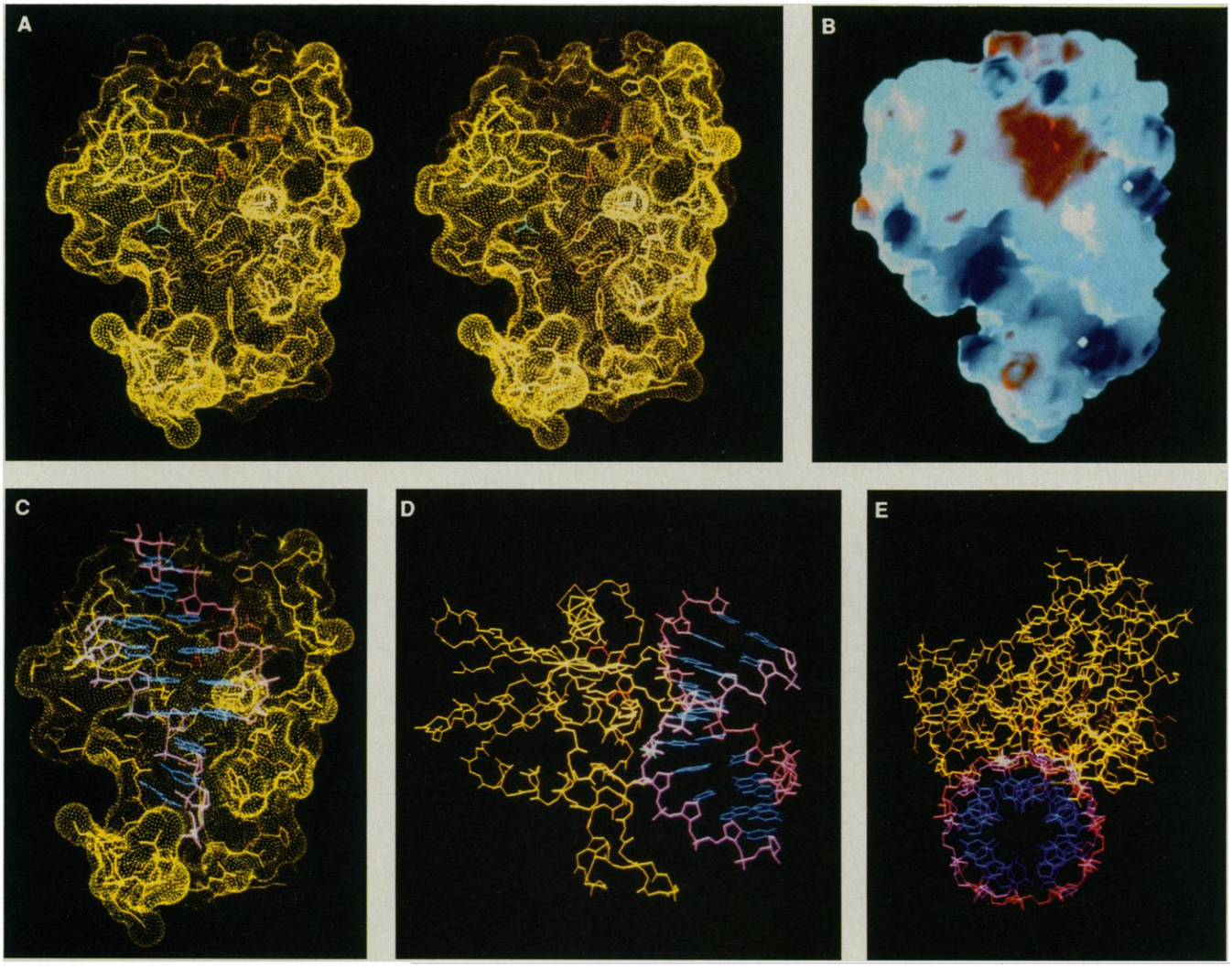


Fig. 7. Substrate binding surface of RNase H and a model-built complex with an A-type RNA·DNA hybrid oligonucleotide. (**A**) Stereoview of the molecular surface in contact with a probe sphere of 1.4 Å radius as drawn by QUANTA (Polygen). Both the surface and most of the atoms are shown in yellow, but side chains of the carboxyl triad are drawn in red and the sulfate ion is in green. The rear surface is clipped away. (**B**) Electrostatic potential surface computed with DELPHI (*47*) and displayed with AAK (A. Nicholls) at the level of the solvent accessible surface with a probe radius of 1.4 Å. A pH of 8.0 was assumed and full charges were used for side chain groups with partial charges for main chain atoms. The sulfate ion was omitted. Blue represents positive potential, red negative, and white neutral. (**C**) Interaction between an RNA·DNA hybrid decamer (*40*) and the molecular contact surface shown in (A). The phosphoribose backbone of the DNA is in lavender (left) and that of RNA is in pink (right); base pairs are blue. Nucleic acid atoms nearest to the viewer are clipped away to emphasize the direct interactions. P6 of the DNA strand is superimposed on the sulfate sulfur atom, and the phosphodiester bond between bases 7 and 8 of the RNA strand are positioned near the carboxyl triad shown in red. P6–P9 of the DNA strand fits into the lower groove, P7–P10 of the RNA strand is in the upper groove, and the minor groove of the nucleic acid straddles the ridge between these depressions on the protein surface. (**D**) Side view of the modeled complex with protein backbone atoms in yellow and with the nucleic acid colored as in (C). The loops between strands 1 and 2 and between strand 5 and the E helix protrude to embrace the substrate. (**E**) End view of the hypothetical complex with all atoms shown. Coloring is as in (D).

tive DNA and RNA depressions. RNA-specific binding interactions might also exist at the catalytic site, but the suggested catalytic mechanism suits both DNA and RNA. Interactions with the DNA strand are consistent with the observed preference of *E. coli* RNase H for cleavage at the phosphodiester bond between the sixth and seventh bases on the RNA strand of hybrid oligonucleotides (*43*). The handle region that seems to be deleted in HIV RNase H is not a major factor in the proposed substrate interactions.

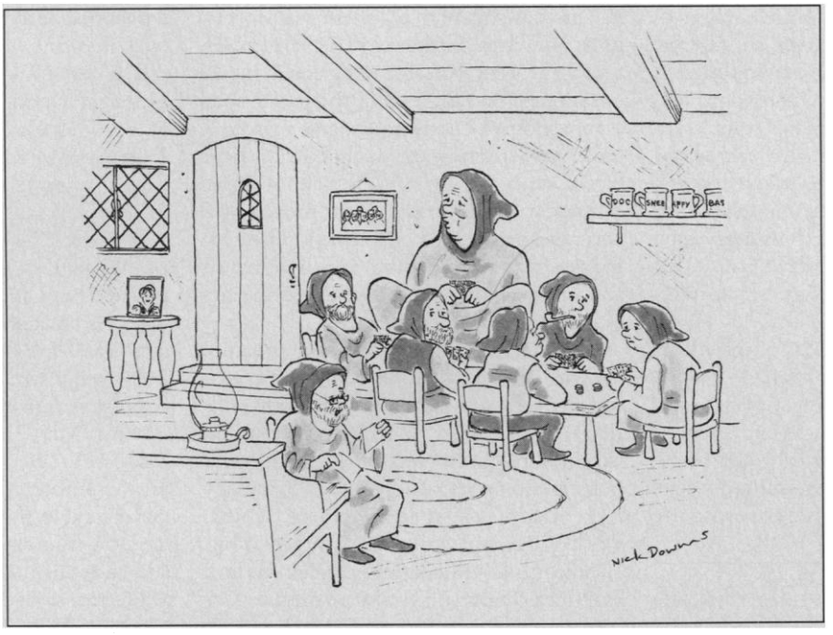
Our results have multiple implications. First, the structure generates hypotheses that can be tested by structure-function studies on cation and nucleic acid complexes and on site-directed mutants. Second, this structure should simplify the structural analysis of reverse transcriptases. Moreover, the *E. coli* RNase H model itself helps to define targets for anti-retroviral drugs. Finally, the determination of this structure directly from the diffraction data of a single crystal demonstrates the power of the MAD method and suggests that recombinant selenomethionyl proteins can serve in general as phasing vehicles in protein crystallography.

REFERENCES AND NOTES

1. R. J. Crouch and M.-L. Dirksen, in *Nucleases*, S. M. Linn and R. J. Roberts, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), p. 211.
2. B. de Massy, O. Fayet, and T. Kogoma, *J. Mol. Biol.* **178**, 227 (1984).
3. T. Horiuchi, H. Maki, M. Sekiguchi, *Mol. Gen. Genet.* **195**, 17 (1984).
4. T. Kitani, K.-y. Yoda, T. Ogawa, T. Okazaki, *J. Mol. Biol.* **184**, 45 (1985).
5. T. Itoh and J. Tomizawa, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2450 (1980); J.

Tomizawa, *Cell* **38**, 861 (1984); S. Dasgupta, H. Masukata, J. Tomizawa, *ibid.* **51**, 1113 (1987).
6. T. A. Torrey, T. Atlug, T. Kogoma, *Mol. Gen. Genet.* **196**, 350 (1984).
7. M. Itaya, D. McKelvin, S. K. Chatterjee, R. J. Crouch, *ibid.* in press.
8. K. Moelling *et al.*, *Nature New Biol.* **234**, 240 (1971); W. Keller and R. J. Crouch, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3360 (1972).
9. S. P. Goff, *J. AIDS* **3**, 817 (1990).
10. G. Luo, L. Sharmeen, J. Taylor, *J. Virol.* **64**, 592 (1990).
11. S. Kanaya and R. J. Crouch, *J. Biol. Chem.* **258**, 1276 (1983).
12. M. S. Krug and S. L. Berger, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3539 (1989); F. Oyama, R. Kikuchi, R. J. Crouch, T. Uchida, *J. Biol. Chem.* **264**, 18808 (1989).
13. H. I. Miller, A. D. Riggs, G. N. Gill, *J. Biol. Chem.* **248**, 2621 (1973); I. B. Berkower, J. Leis, J. Hurwitz, *ibid.* **249**, 5914 (1973).
14. N. Tanese and S. P. Goff, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1777 (1988).
15. M. S. Johnson, M. A. McLure, D.-F. Feng, J. Gray, R. F. Doolittle, *ibid.* **83**, 7648 (1986); J. G. Levin *et al.*, *J. Virol.* **62**, 4376 (1987).
16. V. R. Prasad and S. P. Goff, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3104 (1989); A. Hizi, S. H. Hughes, M. Shaharabany, *Virology* **175**, 575 (1990); J. Hansen, T. Schulze, W. Mellert, K. Moelling, *EMBO J.* **7**, 239 (1988).
17. R. F. Doolittle, D.-F. Feng, M. S. Johnson, M. A. McClure, *Q. Rev. Biol.* **64**, 1 (1989).
18. R. Repaske, J. W. Harley, M. F. Kavilick, R. R. O'Neil, J. B. Austin, *J. Virol.* **63**, 1460 (1989).
19. S. Kanaya *et al.*, *J. Biol. Chem.* **265**, 4615 (1990).
20. J. Karle, *Int. J. Quant. Chem. Symp.* **7**, 357 (1980); R. Kahn *et al.*, *FEBS Lett.* **179**, 133 (1985).
21. W. A. Hendrickson, *Trans. Am. Crystallogr. Assoc.* **21**, 11 (1985).
22. ———, J. L. Smith, R. P. Phizackerley, E. A. Merritt, *Proteins* **4**, 77 (1988).
23. W. A. Hendrickson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2190 (1989).
24. J. M Guss *et al.*, *Science* **241**, 806 (1988); B. J. Graves *et al.*, *Biochemistry* **29**, 2679 (1990); H. M. K. Murthy, W. A. Hendrickson, W. H. Orme-Johnson, E. A. Merritt, R. P. Phizackerley, *J. Biol. Chem.* **263**, 18430 (1988).
25. W. A. Hendrickson, J. R. Horton, D. M. LeMaster, *EMBO J.* **9**, 1665 (1990).
26. W. Yang, W. A. Hendrickson, E. T. Kalman, R. J. Crouch, *J. Biol. Chem.* **265**, 13533 (1990).
27. S. Kanaya *et al.*, *ibid.* **264**, 11546 (1989).
28. Y. Satow and Y. Iitaka, *Rev. Sci. Instr.* **60**, 2390 (1989).
29. J. Miyahara, K. Takasashi, Y. Amemiya, N. Kamiya, Y. Satow, *Nucl. Instr. Methods A* **246**, 572 (1986).
30. C. M. Ogata, W. A. Hendrickson, L. Gao, Y. Satow, Y. Amemiya, in preparation.
31. Y. Amemiya *et al.*, *Nucl. Instr. Methods A* **266**, 645 (1988).
32. E. Fanchon and W. A. Hendrickson, *Acta Cryst. A*, in press.
33. DENZO was written and supplied to us by Z. Otwinowski (Yale University). This program reduces diffraction data measured by rotation photography and is adapted for imaging plates. It corrects for background, refines alignment parameters, performs profile fitted integrations, and applies the Lorentz and polarization factor corrections. The system includes a SCALEPACK routine for merging the data from multiple films or plates.
34. MADSYS is a system of programs for phase determination from multiwavelength data. Included among its programs are LEXICO to format data for the system, ANOSCL and WVLSCL for local scaling of Bijvoet and dispersive data, respectively, MADLSQ to solve the phasing equations, MERGIT to merge redundant evaluations, ASLSQ to refine the structure of anomalous scattering centers, and MADFAZ to produce phases and figure of merit weights.
35. T. A. Jones, *J. Appl. Cryst.* **11**, 268 (1978).
36. J. H. Konnert, *Acta Cryst.* **A 32**, 614 (1976); W. A. Hendrickson and J. H. Konnert, in *Computing in Crystallography*, R. Diamond, S. Ramaseshan, K. Venkatesan, Eds. (Indian Academy of Science, Bangalore, 1980), p. 13.01.
37. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
38. M. P. Ready, B. J. Katzin, J. D. Robertus, *Proteins* **3**, 53 (1988).
39. S. Sheriff, W. A. Hendrickson, J. L. Smith, *J. Mol. Biol.* **197**, 273 (1987).
40. V. Derbyshire *et al.*, *Science* **240**, 199 (1988); P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988).
41. G. Kartha, J. Bello, D. Harker, *Nature* **213**, 862 (1967); F. M. Richards *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **36**, 35 (1971); C. H. Carlisle, R. A. Palmer, S. K. Mazumdar, B. A. Gorinsky, D. G. R. Yeates, *J. Mol. Biol.* **85**, 1 (1974); A. Wlodawer, R. Bott, L. Sjolin, *J. Biol. Chem.* **257**, 1325 (1982).
42. A. H.-J. Wang *et al.*, *Nature* **299**, 601 (1982).
43. H. Inoue, Y. Hayas, S. Iwai, E. Ohtsuka, *FEBS Lett.* **215**, 327 (1987).
44. T. M. Shinnick, R. A. Lerner, J. G. Sutcliffe, *Nature* **293**, 543 (1981); D. Schwartz, R. Tizard, W. Gilbert, *Cell* **32**, 853 (1983); L. Ratner *et al.*, *Nature* **313**, 277 (1985).
45. C. Hill *et al.*, *Trends Biochem. Sci.* **8**, 364 (1983).
46. J. A. McSwiggen and T. R. Cech, *Science* **244**, 679 (1989); C. Guerrier-Takada, K. Haydock, L. Allen, S. Altman, *Biochemistry* **25**, 1509 (1986).
47. M. Gilson, K. Sharp, B. Honig, *J. Comp. Chem.* **9**, 327 (1987).
48. We thank M. Cuff for help in data collection, C. Ogata for advice in data reduction and refinement, A. Pähler for help with computing, A. Nicholls for electrostatic calculations, A. Telesnitsky and S. Goff for discussions about retroviruses, Z. Otwinowski (Yale) for advice in the use of DENZO, A. Wang (Illinois) for atomic coordinates, and T. Steitz (Yale), and T. Cech (Colorado) for discussions about catalytic mechanisms. Experiments at the Photon Factory were performed under the auspices of proposal No. 87-152. Supported in part by NIH grant GM 34102.

19 July 1990; accepted 20 August 1990



" What are all these bills from Genentech, Inc.?"