STRUCTURE OF THE CRYSTALLINE COMPLEX BETWEEN RIBONUCLEASE A AND D(PA)$_4$

ALEXANDER MCPHERSON, GARY BRAYER, AND ROBERT MORRISON
Department of Biochemistry, University of California, Riverside, California 92521

ABSTRACT Crystals of a complex formed between ribonuclease A and d(pA)$_4$ were grown and their structure determined by a combination of multiple isomorphous replacement (MIR) and molecular replacement techniques. The known structure of ribonuclease A in the correct orientation in the unit cell yielded a conventional crystallographic R factor of 0.32 at 2.8 Å resolution when refined as a rigid body. Difference Fourier syntheses permitted determination of the disposition of the DNA in the unit cell. Refinement of both protein and DNA by constrained-restrained least squares procedures resulted in an R factor of 0.22 at 2.5 Å resolution. The structure of the crystalline complex is comprised of four ordered oligomers of d(pA)$_4$ associated with each molecule of RNAse. If the sites of interaction between protein and d(pA)$_4$ fragments are mapped on the surface of the protein, they describe an essentially continuous path into and through the active site, across the surface of the enzyme and finally into the basic amino acid cluster on the opposite side of the protein.

Felsenfeld and his colleagues showed that pancreatic ribonuclease, because it binds single-stranded DNA preferentially, is an effective DNA helix destabilizing, or unwinding, protein (1). These studies were extended by Jensen and von Hippel (2), and by Record et al. (3), who demonstrated that RNAse binds to both double- and single-stranded DNA, apparently by two distinctly different mechanisms, though both use predominantly electrostatic interactions. They further showed that when RNAse is bound to single-stranded DNA, it covers or protects 8–12 bases along the polynucleotide chain; this binding proceeds through the formation of at least seven ion pairs (2, 3). The protection size predicted by the data and the probable involvement of seven lysine and arginine residues forming salt bridges to nucleotide phosphate groups are not readily explained by the current concept of the enzyme's nucleic acid binding site. The active site cleft alone is limited in extent and could not easily accommodate more than a total of four to five nucleotides.

The many studies delineating the features of the active site cleft and the amino acid residues responsible for substrate binding and catalysis, both in solution and in the various crystals that have been analyzed, have been reviewed by Anfinsen and White (4), Richards and Wyckoff (5), and by Blackburn and Moore (6). The structure of the protein has been solved independently by three different laboratories (7–9) and refined using both x-ray and neutron diffraction data (10, 11). The various complexes formed between ribonuclease and a range of substrate analogues and inhibitors as seen in the crystal by difference Fourier analysis have been recently reviewed (12). In no case, however, has a substrate analogue greater in size than two nucleosides with an intervening phosphodiester linkage, such as UpA or CpG, been investigated by x-ray analysis while bound to the active site of the enzyme. Such a study involving longer oligonucleotides is needed to allow confident description of the path taken by a long, continuous single strand of RNA or DNA when it is engaged by the protein. Dinucleotides alone are insufficient to allow deduction of the amino acid residues involved in the extended binding site, and they do not provide an adequate explanation for the protection size and number of ion pairs formed.

We became interested in pancreatic ribonuclease as a result of our studies on the gene 5 DNA unwinding protein from bacteriophage fd. RNAse afforded us the opportunity to study a second single-stranded DNA-binding protein. From studies of these proteins it is our expectation that the precise atomic interactions by which this class of proteins recognizes and binds to single strands of DNA will become apparent. We believe there will be similar mechanisms involved and similar sets of interactions and structures used. This is evidenced, for example, by our recent finding of significant topological equivalence between the gene 5 protein and pancreatic RNAse (13). To achieve these objectives, however, it is essential to observe protein-nucleic acid complexes directly, by Fourier analyses. Up to now we have had only limited success in obtaining isomorphous crystalline complexes of the gene 5 protein either by diffusion of oligonucleotides into preformed native crystals (14) or by cocrystallizing the gene 5 protein with oligonucleotides (15). We have had more success in obtaining such complexes using RNAse as the binding protein.

SOLUTION OF THE STRUCTURE

As described elsewhere (16) we have obtained large single crystals of both ribonuclease A and B complexed with

$d(pA)_4$ and with $d(pA)_6$. These crystalline complexes do not involve the association of a single tetramer of $d(pA)_4$ with the protein, but multiple copies of the oligonucleotide are bound to each enzyme molecule. We have, more recently, also obtained crystals of the protein complexed with the oligomers $d(pT)_4$ and d(pCTTC). Except for a few percent change in cell dimensions, these are isomorphous with the crystals grown from $d(pA)_4$. Given the pattern established by this sampling of possible lengths and sequences of oligonucleotides, we believe that an extensive investigation of subtle binding variations as a function of length and sequence is possible from the root structure we present here.

Crystals of pancreatic RNAse complexed with $d(pA)_4$, like those shown in Fig. 1, are of space group $P2_12_12_1$ and have cell dimensions $a = 44.56$, $b = 75.30$, and $c = 44.60$. A diffraction pattern from one of these crystals is shown in Fig. 2. Crystals grown using $d(pA)_6$ have dimensions $a = 44.50$, $b = 75.30$, $c = 43.40$. Analogous complexes grown using RNAse B are of tetragonal space group $P4_12_12$ (or $P4_32_12$) and have cell dimensions of $a = b = 44.40$, $c = 156.5$. From the cell dimensions and symmetry it is clear that the two crystal forms are closely related in terms of packing, which further suggests similarity in the structure of the protein-nucleic acid complexes. The crystals are all grown from unbuffered polyethylene glycol 4000 (PEG 4000) solutions by the method of vapor diffusion (17) at 4°C and appear after two days to two weeks. The concentration of PEG varied from 16% wt/vol in the case of the $d(pA)_4$ crystals to 28% in the case of the $d(pA)_6$ crystals. These latter crystals were also more difficult to grow.

The details of the structure determination and relevant statistics regarding both the MIR procedure and refinement of the structure will be presented elsewhere. Here we shall summarize the general approach and call attention to certain unanticipated problems that were encountered and solved. Our procedure was to produce a low-resolution MIR map of the protein using as few heavy atom derivatives as possible. We then determined the orientation and position of the known protein molecule in the electron density map by inspection, placed the atomic structure of the protein in the unit cell in this optimal orientation, and refined the protein against the x-ray data (treating it strictly as a rigid body). Finally we calculated structure factors and difference Fourier maps to locate the $d(pA)_4$ oligomers. The complex was then refined by least-squares methods. The application of these procedures was not so straightforward as one might have expected.

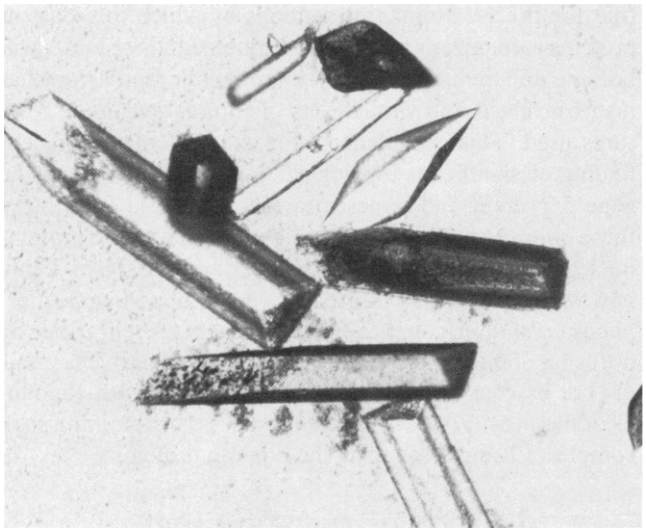

FIGURE 1 A low-power light microscope photograph of crystals of the complex between ribonuclease A and $d(pA)_4$ grown from 17% PEG 4000 at 4°C and low ionic strength by the vapor diffusion technique.

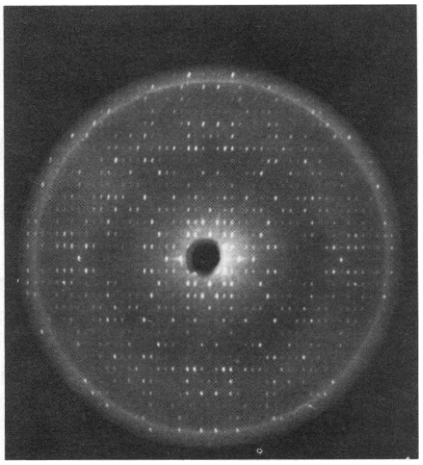

FIGURE 2 A precession x-ray diffraction photograph from a crystal of the complex formed between ribonuclease A and $d(pA)_4$. The precession angle is 14° and the x-ray source was an Eliott rotating anode operated at 40 kV and 40 mA with a focal spot size of 200 μ^2. The reciprocal lattice level shown is 0kl.

We found it necessary to collect data from five separate heavy atom derivative compounds to obtain a native Fourier map that eventually could be interpreted in terms of the known structure of the protein molecule. These heavy atom compounds included K_2PtCl_4, UNO_3, $PtBr_2(NH_3)_2$, mercury acetate, and K_2PtCl_6. The data were collected to a resolution of 3.5 Å when crystal stability permitted, which was not very often. The native crystals were extremely stable, but except for mercury acetate, all of the crystals derivitized with heavy atoms deteriorated rapidly. UNO_3 could be collected only to 5.0 Å resolution; mercury acetate was only weakly substituted, and all of the platinum-containing derivatives decayed rapidly and were nonisomorphous beyond low resolution. The most interpretable map was that computed at 4 Å resolution. The average difference between those phases derived from MIR data and the phases calculated from the structure at the end of the 2.5 Å refinement was 64.4° to 3.5 Å, the limit of the MIR data.

After several futile attempts using incorrect positions and orientations (the wrong placements were obvious when

the rigid body structures were refined at high resolution) we were able to place the structure correctly in the unit cell. This was recognized by (*a*) attainment of a conventional crystallographic R factor of 0.33 first at 5.0 Å resolution using the protein alone; (*b*) the ability of calculated phases to return the heavy atom sites clearly in difference Fourier syntheses; and finally, (*c*) the attainment of an R factor of 0.32 at 2.8 Å after refinement, again using the unaltered protein structure of Wlodawer and Sjolin (10) as the model. The electron density map was complicated to interpret considering the amount of information known, but we did not know during the course of our investigation precisely how much nucleic acid was present in the asymmetric unit of the crystal. We proceeded on the assumption that no more than two $d(pA)_4$ oligomers per protein molecule were present, an assumption that turned out to be wrong. We know now that there are four ordered tetranucleotides present per RNAse molecule and possibly disordered oligomers in the unit cell as well. Given the substantially greater x-ray scattering power of the nucleic acid at low resolution and generally high thermal parameters, we can see now in retrospect that the low-resolution (4.0 Å) electron density map we had calculated was dominated by images of the DNA fragments. This complicated matters considerably; not only did we not know the correct number of $d(pA)_4$ oligomers present, but also the oligomers in many cases contacted more than one protein molecule in the crystal lattice. The entire unit cell was extensively cross-linked, and the apparent envelopes seen in the map were not representative of the features of the protein.

Following correct orientation of the protein molecule and refinement of the protein as a rigid body at 2.8 Å resolution, difference Fourier maps were calculated to determine the disposition of the $d(pA)_4$ oligomers. The packing of the ribonuclease molecules in the orthorhombic unit cell is shown in Fig. 3. Once again the interpretation of the maps was not straightforward, partly because we did not know the number of oligomers present, partly because only some of the $d(pA)_4$ was clearly recognizable as such; the rest appeared partially resolved and fragmented. Given the R factor at this stage, R = 0.32, we might have expected better. However, if the amount of nucleic acid as a percentage of nonsolvent weight in the unit cell is calculated, it comprises almost 25% of the total. Considering nucleic acid's greater x-ray scattering power, it is perhaps not so surprising that the image was less than ideal. It should also be remembered that the RNAse coordinates of Wlodawer and Sjolin's refinement (10) provided the phasing model. The differences between that model and the one we eventually found in this unit cell could not at this stage be taken into account. Although the changes are quite small (according to the results of our refinement the root mean square change in α-carbon position is 0.36 Å), and confined primarily to side chain atoms and flexible loops, they are significant in terms of the phases. This was shown by calculation of difference Fourier maps at the end of refinement using phases based on the refined protein alone, which produced perceptibly improved images of the $d(pA)_4$ oligomers as compared with the starting model.

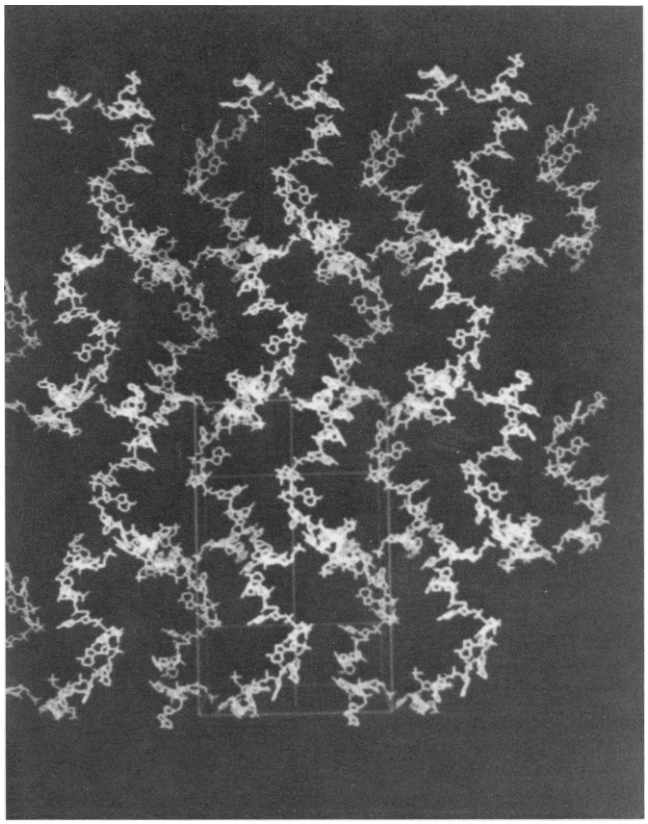

FIGURE 3 A stereo diagram showing the packing of the protein molecules in the $P2_12_12_1$ unit cell of the crystals used in this study. Apparent in this figure is the scarcity of direct protein-protein contacts in the lattice. Thus the protein-DNA interactions appear to supply the majority of stabilizing cross links that are responsible for crystal formation. Please refer to the color figure section at the back of this book.

From the initial difference Fourier maps, two $d(pA)_4$ oligomers could be delineated with confidence. Although portions of others could also be seen, the connectivity was not clear, and they were not initially included in the phasing model. The coordinates of the previously refined ribonuclease structure (10) were placed appropriately in the unit cell and refined as a rigid body along with two tetramers of DNA derived from the difference map. These were fitted into the unit cell using the graphics program FRODO (written by A. Jones [18] and modified by B. Bush) running on an Evans and Sutherland multipicture system. The procedure used for the rigid body refinement as well as all subsequent refinement described here was the restrained-constrained least-squares procedure CORELS written by J. Sussman (19, 20). This approach has been applied successfully in the past to both protein and nucleic acid structures independently, but so far as we know, not to complexes of the two. In the initial cycles only the $d(pA)_4$

tetramers were permitted to vary according to the restrained bonds joining the phosphate and nucleoside portions of each oligomer. Thus there were eight constrained, or structurally fixed, groups defining each $d(pA)_4$ fragment. When no significant change was observed in the coordinates of the two oligomers included in the model, difference Fourier maps were calculated and portions of the third and fourth tetramers were introduced into the refinement. The dihedral angles of the amino acid side chains were allowed to refine in concert with the $d(pA)_4$ oligomers; new difference Fouriers were computed; and the third and fourth tetramers were gradually completed. In the final stages of the refinement, all amino acids were treated as constrained groups and allowed to refine simultaneously with the constrained DNA while the group temperature factors were refined as well. During the course of refinement, geometrical constraints were stringently maintained and all shifts were damped to 0.10–0.50 of their calculated values. Nearly 200 cycles of refinement were required to reach the final R value at 2.5 Å of 0.22. The approach to refinement we used here was perhaps overly conservative, but the statistics describing the ideality of the final geometrical parameters suggest that the structure of the complex we eventually obtained is realistic and does not deviate from the high accuracy of the starting protein model (10).

DESCRIPTION OF THE STRUCTURE

The asymmetric unit of the crystalline complex of ribonuclease A with $d(pA)_4$ is composed of one protein molecule associated with four tetramers of $d(pA)_4$. The presence of isolated peaks in difference Fourier maps indicate there may be additional, disordered fragments of DNA present in the solvent regions. The complex of protein plus all oligonucleotides is shown in Figs. 4–6, and the protein with each of the four $d(pA)_4$ oligomers in turn in Figs. 7–10.

Two of the DNA fragments may be considered entirely bound by a single protein molecule with no more than their 3′ terminal nucleotides engaged in any intermolecular interaction. They are essentially intramolecular oligomers. The two tetramers appearing less clearly in the difference

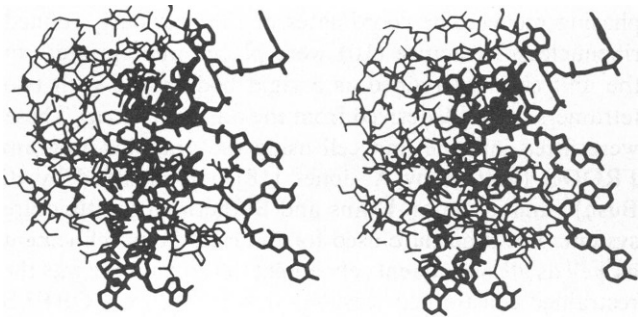

FIGURE 4 Stereo diagram of the asymmetric unit of the crystalline complex between ribonuclease A and $d(pA)_4$ consisting of one protein molecule associated with four individual $d(pA)_4$ oligomers. The bonds of the oligonucleotides have been overdrawn for clarity.

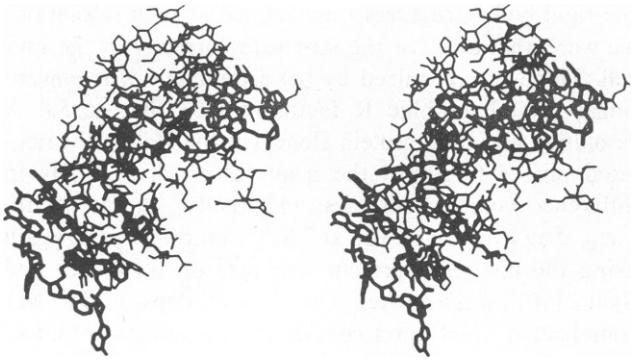

FIGURE 5 Stereo diagram of the ribonuclease A protein molecule associated with the four different $d(pA)_4$ oligonucleotides that comprise the asymmetric unit of the crystal. Running more or less from the top of the figure to the bottom in a consistent 3′ to 5′ direction, the four fragments of DNA can be seen to trace out an essentially continuous path through the active site cleft of the enzyme and over the surface of the protein. Electrostatic linkages are made between phosphates on the oligonucleotides and lysines 7, 41, 66, 37, 31, 98 as well as with arginines 85, 39, and possibly 33. Interactions involving the ribose and base moieties occur in the active site cleft.

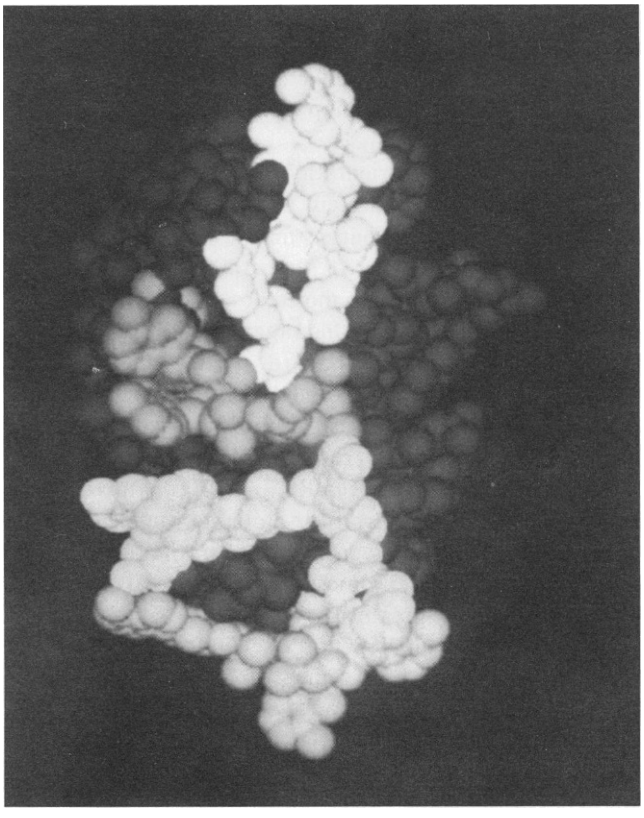

FIGURE 6 Stereo photograph of a van der Waals sphere representation of the RNAse protein molecule associated with four oligomers of $d(pA)_4$ as seen in a similar orientation as in Fig. 5. Please refer to the color figure section at the back of this book.

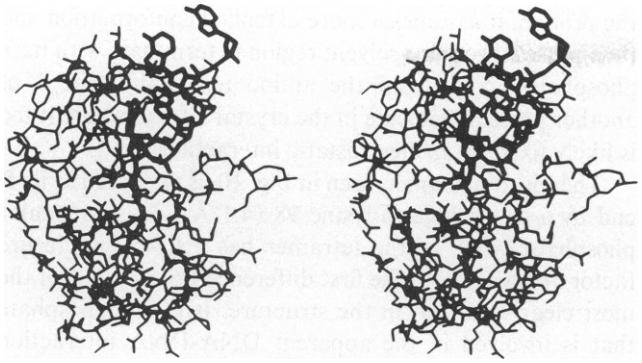

FIGURE 7 The RNAse protein is seen here associated with an oligomer of $d(pA)_4$ that enters the active site of the enzyme 3′ to 5′, occupies the known purine binding site with the 5′ terminal adenosine and is fixed immediately adjacent to histidines 12 and 119, which are catalytic residues, by a salt bridge to lysine 41 at the catalytic center. The disposition of the 5′ nucleotide portion of this $d(pA)_4$ oligomer is similar to if not the same as the purine residues of dinucleotides previously visualized in difference Fourier studies of RNAse.

Fouriers and requiring cautious reconstruction during the refinement process, are, in part, intermolecularly bound.

The two intramolecular oligomers were resolved in difference Fourier maps, refined without difficulty, and are characterized by relatively low temperature factors. One of the $d(pA)_4$ tetramers, seen isolated in Fig. 7, lies in the active site cleft of the enzyme, as might have been predicted from the previous x-ray diffraction studies involving dinucleotide substrate analogues (21–25). The 5′ terminal phosphate of this tetramer is at the active center

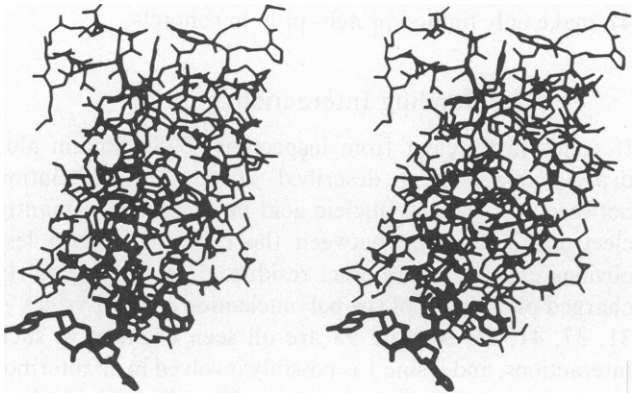

FIGURE 8 The tetramer of $d(pA)_4$ farthest from the active site of the enzyme is shown here associated with RNAse in the absence of all other DNA tetramers. It binds securely through its 5′ terminal and penultimate phosphate groups to the strongly electropositive anion binding site formed by the clustering of lysines 31, 91, and 37 as well as arginine 33 on the underside of the protein molecule. The nucleotides forming the 3′ half of the oligomer follow the contours of the protein and end near the 5′ phosphate of the last $d(pA)_4$ segment in the set (that shown in Fig. 10) bound by this same RNAse molecule. The 3′ nucleotide of the tetramer seen in Fig. 7, but from a symmetry equivalent $d(pA)_4$ in the lattice, is also near this phosphate creating an apparent DNA-DNA interaction involving two 3′ terminal bases and a 5′ terminal phosphate. Thus the crystal apparently uses DNA-DNA, protein-DNA and protein-protein interactions to maintain its integrity.

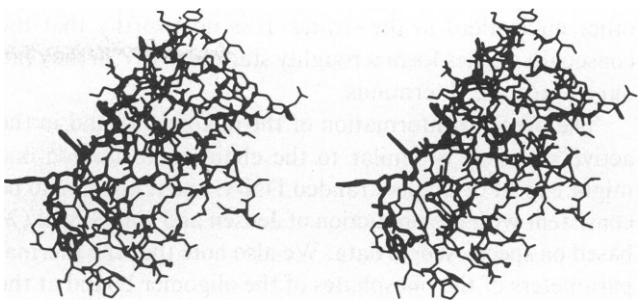

FIGURE 9 A third tetramer of $d(pA)_4$ is seen here bound to the RNAse molecule in the absence of all others. It extends from a point very close to the pyrimidine binding site, though it does not appear to occupy this site, over the surface of the protein and into a solvent region of the crystal running 3′–5′. The most 3′ phosphate group could be salt-bridged to lysine 66 and others to arginines 85 and 39. After passing through solvent, the 5′ phosphate may be salt-bridged to lysine 1 of another protein molecule in the lattice.

only 3.8 Å from histidine 12, 4.2 Å from the amino group of lysine 41 and 3.7 Å from histidine 119, the generally accepted catalytically important residues in the enzyme. The base of residue 1 is stacked against the surface of the protein, as previously observed for other complexes, and the base is in an anti conformation. The second phosphate from the active center on this tetramer is 4.7 Å from lysine 7, to which it is apparently salt-bridged. The ribose and base are again packed closely against the protein surface. The third phosphate from the active center cannot be seen to make any obvious salt bridges to a lysine or arginine. The 3′ nucleotide is not bound in the crystal by the same protein molecule as the others in its strand, but makes limited contact with another $d(pA)_4$ tetramer in the lattice. While the first nucleotide at the 5′ end of the oligomer is in a fully extended conformation, the same is less true of the

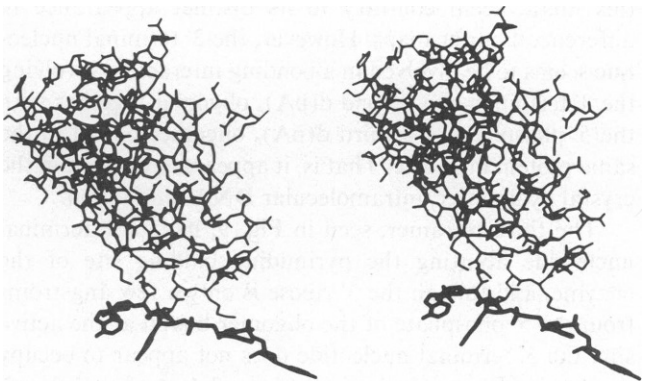

FIGURE 10 In this stereo diagram, the RNAse molecule is seen associated with the last of the four $d(pA)_4$ oligomers. This tetramer is secured at its 5′ terminal phosphate by a salt bridge to lysine 98 and extends over the surface of the protein, possibly making an electrostatic interaction with arginine 39, and then passes into a large interstitial solvent region. The 3′ terminal nucleotide appears to be in contact with another protein molecule in the crystal lattice. It is the 5′ terminal phosphate of this tetramer, bound by lysine 98, that is associated as well with the 3′ bases of two other tetramers in the crystal.

other nucleotides in the strand. It is noteworthy that the consecutive bases form a roughly stacked array as they fan out toward the 3′ terminus.

The overall conformation of the oligomer bound in the active site cleft is similar to the chain conformation one might expect of single-stranded DNA, and it appears to be consistent with the prediction of Jensen and von Hippel (2) based on spectroscopic data. We also note that the thermal parameters of the phosphates of the oligomer bound at the active site are about the same as the overall temperature factor for the protein itself, suggesting that immobilization of the nucleic acid occurs.

The second tetramer to be found entirely by a single protein molecule is not found at the active site, or even near the active site, but on the opposite side of the protein molecule, as seen in Fig. 8. This mode of binding is not unexpected, however, and might have been predicted from the results of Matthew and Richards (26, 27). In calculating the electrostatic field in the volume of space around a molecule of RNAse under low ionic strength conditions, they discovered the presence of a very electropositive cavity in the midst of the cluster of basic amino acids comprised of lysine 31, 37, and 91 along with arginine 33. Indeed, it is in the center of this cluster that the bulk of the second $d(pA)_4$ oligomer is bound, and it seems clear that at least the 5′ terminus as well as the penultimate phosphate of this oligomer form electrostatic linkages to amino acid side chains, with the most likely participants being lysines 31 (3.1 Å, 91 (3.2 Å), and probably 37 (~6 Å).

The phosphates of this second tetramer are tucked closely against the protein surface while the bases of all of the nucleotides are exposed to the solvent. The entire oligomer is in an extended conformation. Figs. 4, 5, 6, and 8 show that the 3′ terminal nucleotides of this oligomer extend into a solvent region between protein molecules and this might seem contrary to its distinct appearance in difference Fourier maps. However, the 3′ terminal nucleotide seems to be involved in a bonding interaction involving the 3′ terminus of a second $d(pA)_4$ oligomer together with the 5′ phosphate of a third $d(pA)_4$ oligomer bound to the same protein molecule. That is, it appears to be fixed in the crystal by an inter/intramolecular DNA interaction.

The third tetramer, seen in Fig. 9, has its 3′ terminal nucleotide abutting the pyrimidine binding site of the enzyme, and though the 3′ ribose is only a few ångströms from the 5′ phosphate of the oligomer bound at the active site, the 3′ terminal nucleotide does not appear to occupy the pyrimidine site. Lysine 66 is 3.4 Å from the 3′ phosphate of this oligomer. Arginine 85, which appears to shift the position of its side chain toward this tetramer, is likely to aid in positioning the 3′ end. The $d(pA)_4$ strand runs 3′–5′ over the surface of the protein passing near arginine 39 which could form electrostatic linkages with phosphates of this oligomer. The two nucleotides comprising the 3′ end of the tetramer are somewhat compressed and the bases stacked. As the DNA crosses the surface of the protein, it assumes a more extended conformation and then passes through a solvent region to terminate with its 5′ phosphate ~5 Å from the amino terminal lysine 1 of another protein molecule in the crystal lattice, where it too is likely fixed by an electrostatic interaction.

The fourth tetramer, seen in Fig. 10, is anchored at its 5′ end by a salt bridge to lysine 98 (4.1 Å). This particular phosphate group of the tetramer has a low temperature factor and was, from the first difference Fourier, one of the most clearly defined in the structure. It is this phosphate that is involved in the apparent DNA-DNA interaction noted above. From the 5′ terminus the oligomer passes in more or less extended conformation up the side of the protein very near to arginine 39, where it could form another ion pair and then extends out into a large solvent cavity, becoming the most ill-defined of the $d(pA)_4$ oligomers. The 3′ phosphate group lies less than 4 Å from lysine 91 of another molecule in the unit cell.

While it appears reasonably safe at this stage to describe in detail the first two tetramers, those bound most securely to a single protein molecule, some of the characteristics of the second pair of $d(pA)_4$ tetramers should be regarded with caution. Their elusive images in the difference electron density maps suggest some degree of disorder. The positions of the bases in the solvent regions are unresolved in the difference maps and it would be unwise at this stage to attempt a detailed description of their conformations and the interactions they make. Even so, these two intermolecular oligomers appear important to the integrity of the crystal, because they supply the electrostatic cross-links that bind the protein molecules, which (as seen in Fig. 4), make only limited protein-protein contacts.

The Bonding Interactions

It seems fairly clear from inspection of the protein plus $d(pA)_4$ arrangement described above that association between RNAse and nucleic acid involves predominantly electrostatic bonding between the lysine, and to a less obvious extent the arginine, residues, and the negatively charged phosphates of the polynucleotide chain. Lysines 7, 31, 37, 41, 66, 91, and 98 are all seen engaged in such interactions, and lysine 1 is possibly involved in an intermolecular salt cross-bridge. Arginines 39 and 85 are involved in such ion pairs as well. The protein-nucleic acid interactions are not limited entirely to electrostatic bonds. In this structure we can see close contact of two nucleotides with amino acid residues in the active site cleft and these involve ribose and adenine moieties. From other studies, we can be certain that at least one additional ribose and base are bound by the pyrimidine binding site which is, in this complex, unoccupied. The failure of the DNA fragment to use this binding mode apparently is a consequence of the absence of a pyrimidine in the oligonucleotides. Combination of the two kinds of chemical interactions in the proportions seen here agrees with the data and predictions

of Jensen and von Hippel (2), who described just such an arrangement from their solution studies, for the case of an extended single strand of DNA. They suggested that the bulk of the binding would derive from electrostatic bonds and that the specificity for the enzyme would be provided by a limited set of interactions directly at the active site involving the sugar and base portions of the nucleic acid.

Binding of a Single Strand of DNA

A question that must be addressed is whether the complex seen here in the crystal resembles the complex between protein and single-stranded DNA that forms in solution, or whether it is primarily a consequence of nonspecific binding and the influence of lattice interactions. Clearly, the crystalline complex between RNAse and $d(pA)_4$ is not an exact replica of the single-stranded DNA interaction. The nucleic acid chain is not continuous but composed of four segments; the pyrimidine binding site is unfilled because there are no pyrimidine residues; and some of the observed interactions occur between protein molecules and reflect lattice interactions. Yet, there are a number of reasons to believe that the complex we observe here does mimic the complex formed with single stranded nucleic acid in solution. If those electrostatic interactions that form between the $d(pA)_4$ tetramers and only a single RNAse molecule are considered, and if the nucleotides involved in intermolecular interactions are ignored, then the series of nucleotides remaining trace out a near-continuous path running from the 3′ terminus of the oligomer bound in the active site cleft, through the active site, over the surface of the protein, and finally into the electropositive cluster on the back side of the protein. This is shown in Fig. 11. The path would require between 11 and 13 total nucleotides and would use all of the interactions described previously. The distances between phosphate groups would be stereochemically reasonable and no dramatic turns or bends would be required. The course of the nucleotides would be 3′ to 5′ over the entire path so no reversal in polarity occurs for any oligomer.

While the model described here is consistent with the binding of a continuous single strand of DNA, one may ask if there is any evidence that this path is in fact the one actually used by the DNA. We believe that there is such evidence. Jensen and von Hippel (2) have shown the protection, or site-covering size, to be 8–12 nucleotides long when the protein binds to single-stranded DNA. Clearly the active site cleft alone is inadequate to explain this feature. The model presented here, based on the $d(pA)_4$ fragments, requires ~12 nucleotides along the polynucleotide chain. These would be in more or less extended conformation, as predicted from Jensen and von Hippel's data. Record et al. (3) determined by salt titration that seven to eight ion pairs are formed between the protein and the DNA over its course. The model presented here in fact contains interactions between six lysine residues and

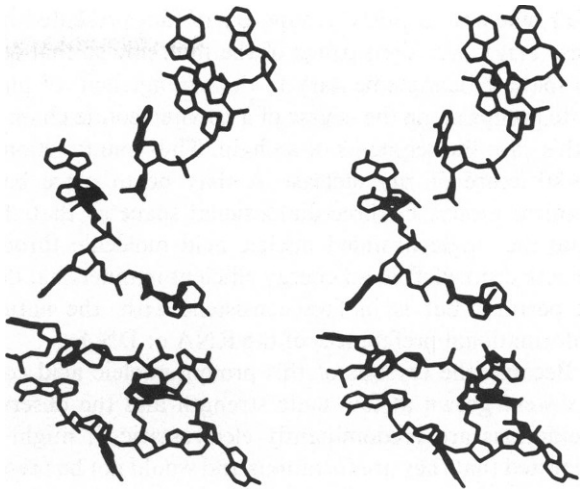

FIGURE 11 Stereo diagram of the four oligomers of $d(pA)_4$ showing their relative disposition in space in the absence of the protein molecule. If the binding interactions between the DNA fragments and the protein are mapped on the surface of the ribonuclease, they trace out a near continuous path from the 3′ terminus of the first to the 5′ terminus of the last. The path is particularly evident if those nucleotides are ignored that are responsible for intermolecular cross-links, and only those considered that are reasonably grouped with a single protein molecule.

possibly two to three arginine side chains, numbers consistent with their results. The distribution of the kinds of interactions in this complex is also consistent with the distribution involving phosphates and ribosides predicted by Jensen and von Hippel (2) and are consistent with the findings of Karpel et al. (28) using nucleic acid analogues lacking heterocyclic bases.

As further evidence that this may resemble a specific complex, the first tetramer involved occupies the active site cleft and uses some interactions previously seen in x-ray diffraction studies of dinucleotide complexes and many of those predicted on the basis of solution studies. These include, for example, the involvement of lysines 7, 41, and 66 with phosphate groups. The thermal parameters for this oligomer, as well as for the elements of all the other oligomers bound directly by the RNAse, are as low as those for the protein. The 5′ phosphate group is fixed by lysine 41 with histidines 12 and 119 nearby, as would be expected from previous studies. No substantial changes in the protein were required to accommodate these interactions and the backbone conformation of the protein is virtually unchanged. At the 5′ end of the path is another oligomer that was consistently clear and is apparently firmly bound by the protein molecule. This oligomer has its 5′ and penultimate phosphate nearly buried in the center of the electropositive cavity lying on the back side of the RNAse molecule and predicted by Matthew and Richards to be a strong anion binding site (26, 27).

The distribution of basic amino acid residues on the protein capable of forming electrostatic interactions with a negatively charged polynucleotide chain is not random, as is evidenced by this complex between RNAse and $d(pA)_4$.

The lysine and arginine groups are in fact presented in a linear array over the surface of the molecule so that they are spatially complementary to the arrangement of phosphate groups along the course of a polynucleotide chain, or in this case the segments of a chain. Thus one function of the structure of ribonuclease A may be to place basic chemical groups in three-dimensional space so that they guide the single-stranded nucleic acid molecule through the active site cleft in an energy efficient manner that does not perturb, but is in fact consistent with, the natural conformational preferences of the RNA or DNA.

Because the crystals of this protein-nucleic acid complex were grown at low ionic strength and the observed interactions are predominantly electrostatic, it might be suggested that they are fortuitous and would not be present at the higher salt concentrations that would be found under more physiological conditions. As Record et al. (3) have argued, however, it is the mixing entropy of released bound ions displaced by the structurally constrained nucleic acid that drives the formation of the RNAse and DNA complex. The energy change derived from the Coulombic interactions is not of primary importance. Record's interpretation of Jensen and von Hippel's data in this regard is convincing. The interactions seen in this complex would not necessarily be absent or appreciably different under conditions of more physiologically relevant ionic environment.

Finally, if the $d(pA)_4$ oligomer bound in the active site cleft and the oligomer that appears to emerge from the pyrimidine binding site are considered alone as an essentially continuous strand of eight nucleotides, they seem to describe an arc, seen in Fig. 12, of roughly constant radius. This arc is similar to that assumed by a single strand of nucleic acid when a part of a DNA or RNA helical duplex. Further, if the remaining two oligomers are considered in the same way, they too form a similar kind of arc, seen in Fig. 13, though not quite so well defined as in the first case. Perhaps these arcs, resembling helical turns, simply reflect the conformational affinities of nucleic acid strands, or perhaps they imply certain mechanistic features of the protein-nucleic acid complex as it forms under physiological conditions. We hope the significance of these arcs will become clearer as these studies progress.

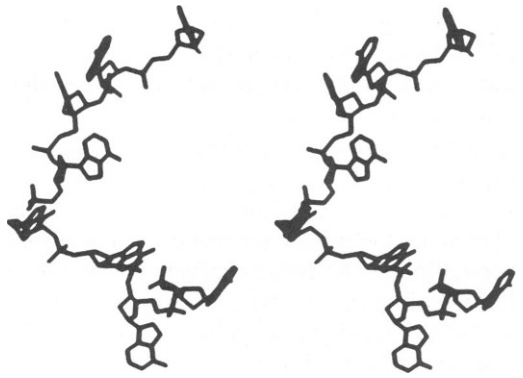

FIGURE 12 Stereo diagram of the $d(pA)_4$ oligomers seen in Figs. 7 and 9 that more or less use the active site cleft of the ribonuclease molecule. Evident here is the rather smooth and continuous arc formed by the eight nucleotides that resembles a turn of helix in native duplex DNA.

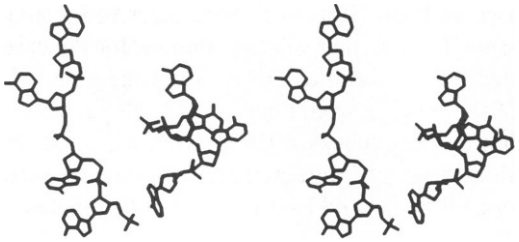

FIGURE 13 Stereo diagram of the $d(pA)_4$ oligomers seen in Figs. 8 and 10 that use the electropositive anion binding site on the side of the protein molecule away from the active site. Curiously, these two oligomers also form a rather smooth and continuous arc of nucleotides similar to that created by the other pair of $d(pA)_4$ tetramers seen in Fig. 12.

The authors would like to thank Drs. Duilio Cascio and Roger Williams for invaluable assistance with computer programs and the members of David Eisenberg's laboratory at the University of California, Los Angeles, particularly Dr. Ward Smith and Tom Terwilliger, for their assistance as well. We would also like to acknowledge the help of Gregory DeLozier and Marie Greene with computer graphics problems; to thank Dr. Alex Wlodawer for his helpful comments; and Drs. Joel Sussman and Steve Holbrook for their advice regarding the use of the refinement program.

This research was supported by a grant (GM-27838) from the United States Public Health Service. G. Brayer was a postdoctoral fellow of the Medical Research Council of Canada and gratefully acknowledges their support.

Received for publication 8 May 1985 and in revised form 17 June 1985.

REFERENCES

1. Felsenfeld, G., G. Sandeen, and P. H. von Hippel. 1963. The destabilizing effect of ribonuclease on the helical DNA structure. *Proc. Natl. Acad. Sci. USA.* 50:644–651.
2. Jensen, D. E., and P. H. von Hippel. 1976. DNA melting proteins. *J. Biol. Chem.* 251:7198–7214.
3. Record, M. T., T. M. Lohman, and P. de Haseth. 1976. Ion effect on ligand-nucleic acid interactions. *J. Mol. Biol.* 107:145–158.
4. Anfinsen, C. B., and F. H. White, Jr. 1961. The ribonucleases: occurrence, structure, and properties. *In* The Enzymes. P. D. Boyer, editor. Academic Press, Inc., New York. 2nd Ed. 5:95–122.
5. Richards, F. M., and H. W. Wyckoff. 1971. Bovine pancreatic ribonuclease. *In* The Enzymes. P. D. Boyer, editor. Academic Press, Inc., New York. 3rd Ed. 4:709.
6. Blackburn, P., and S. Moore. 1982. Pancreatic ribonuclease. *In* The Enzymes. P. D. Boyer, editor. Academic Press, Inc., New York. 3rd Ed. 15.
7. Wyckoff, H. W., D. Tsernoglou, A. W. Hanson, J. R. Knox, B. Lee, and F. M. Richards. 1970. The three-dimensional structure of ribonuclease-S: interpretation of an electron density map at a nominal resolution of 2 Å. *J. Biol. Chem.* 245:305–328.
8. Kartha, G., J. Bello, and D. Harker. 1967. Tertiary structure of ribonuclease. *Nature* (*Lond.*). 213:862–865.

9. Carlisle, C. H., R. A. Palmer, S. K. Mazumdar, B. A. Gorinsky, and D. G. R. Yeates. 1974. The structure of ribonuclease at 2.5 Å resolution. *J. Mol. Biol.* 85:1–18.
10. Wlodawer, A., and L. Sjolin. 1983. Structure of ribonuclease A: results of joint neutron and x-ray refinement at 2.0 Å resolution. *Biochemistry.* 22:2720–2728.
11. Borkakoti, N., R. A. Palmer, I. Haneef, and D. S. Moss. 1983. Specificity of pancreatic ribonuclease-A: an x-ray study of a protein-nucleotide complex. *J. Mol. Biol.* 169:743–756.
12. Wlodawer, A. 1984. Structure of bovine pancreatic ribonuclease by x-ray and neutron diffraction. *In* Biological Macromolecules and Assemblies. Nucleic Acids and Interactive Proteins. F. A. Jurnak, and A. McPherson, editors. John Wiley & Sons, Inc., New York. 2:393–440.
13. Brayer, G. D., and A. McPherson. 1985. Topological equivalence of the gene 5 DNA binding protein and pancreatic ribonuclease. *J. Biolmolec. Struct. Dynam.* In press.
14. Brayer, G. D., and A. McPherson. 1984. Mechanism of DNA binding to the gene 5 protein of bacteriophage fd. *Biochemistry.* 23:340–349.
15. McPherson, A., A. H. J. Wang, F. A. Jurnak, I. Molineux, F. Kolpak, and A. Rich. 1980. X-ray diffraction studies on crystalline complexes of the gene 5 DNA-unwinding protein with deoxyoligonucleotides. *J. Biol. Chem.* 255:3174–3177.
16. Brayer, G. D., and A. McPherson. 1981. Preliminary diffraction data for crystals of ribonuclease A and B and their complexes with deoxy$(pA)_4$ and deoxy$(pA)_6$. *J. Biol. Chem.* 257:3359–3361.
17. McPherson, A. 1982. The Preparation and Analysis of Protein Crystals. John Wiley & Sons, Inc., New York. 94–97.
18. Jones, T. A. 1982. FRODO: A graphics fitting program for macromolecules. *In* Computational Crystallography. D. Sayre, editor. Oxford University Press, New York. 303–317.
19. Sussman, J. L. 1983. Protein model building by the use of a constrained-restrained least squares procedure. *J. Appl. Crystal.* 16:144–150.
20. Sussman, J. L. 1983. Application of refinement constraints and restraints to proteins and nucleic acids. *Proc. Interntl. Sum. Sch. Crystallogr. Comput., Kyoto.*
21. Wyckoff, H. W., W. Carlson, and S. Wodak. 1977. The structure of nucleic acid-protein complexes as evidenced by dinucleotide complexes with RNAse-S. *In* Nucleic Acid-Protein Recognition. H. J. Vogel, editor. Academic Press, Inc., New York. 569–580.
22. Pavlovsky, A. G., S. N. Borisova, V. V. Borisov, I. V. Antonov, and M. Y. Karpeisky. 1978. The structure of the complex of ribonuclease S with fluoride analogue of UpA at 2.5 Å resolution. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 92:258–262.
23. Wodak, S., M. Y. Liu, and H. W. Wyckoff. 1977. The structure of cytidilyl (2′, 5′) adenosine when bound to pancreatic ribonuclease S. *J. Mol. Biol.* 116:855–875.
24. Wlodawer, A., M. Miller, and L. Sjolin. 1983. Active site of RNAse: neutron diffraction study of a complex with uridine vanadate, a transition-state anlog. *Proc. Natl. Acad. Sci. USA.* 80:3628–3631.
25. Borkakoti, N. 1983. Enzyme specificity: base recognition and hydrolysis of RNA by ribonuclease A. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 162:367–373.
26. Matthew, J. B., and F. M. Richards. 1982. Anion binding and pH-dependent electrostatic effects on ribonuclease. *Biochemistry.* 21:4989–4999.
27. Richards, F. M. 1982. Anion binding and pH-dependent electrostatic effects on ribonuclease. *Brookhaven Symp. Biol.* 32.
28. Karpel, R. L., V. A. Yrttimaa, and G. L. Patel. 1981. A helix destabilizing protein substrate devoid of heterocyclic bases. *Biochem. Biophys. Res. Commun.* 100:760–768.

DISCUSSION

Session Chairman: Lee Makowski
Scribes: Martha Briggs and William E. Royer, Jr.

DURKIN: I need some clarification about the virtual strands. Are they like one strand of a double helix?

McPHERSON: No. They are extended single strands, but they are definitely helical in character. They form beautiful helices in the crystal.

DURKIN: Are the nucleotides that you eliminated to find the virtual strand involved in interactions with neighboring molecules in the crystal?

McPHERSON: The termini of nucleotides are those principally involved in interactions with other molecules. The majority of the intermolecular nucleotides are situated in solvent regions and are not making contact with the protein molecule complexed with the virtual strand, but are on the back side of another molecule in the lattice.

DURKIN: So they are not making base pairs with nucleotides from a neighboring helix?

McPHERSON: No, although there are DNA–DNA interactions in these crystals. The complex is not maintained strictly by protein-DNA contacts. There are substantial protein-protein contacts, and there are also some DNA–DNA interactions. These involve, for example, a 5′ phosphate group that is bound to lysine 91, plus two bases from two other tetramers. We think it is the amino groups of the adenines that are hydrogen-bonding to the phosphate groups at the 5′ of another DNA tetramer. Thus protein-protein, nucleic acid-nucleic acid, and protein-nucleic acid interactions all hold the crystal together.

MAKOWSKI: What interactions are there between the protein and the bases?

McPHERSON: The bases appear to be engaged only at the active site, and those are essentially the same interactions that were observed by crystallographers working with dinucleotide complexes of ribonuclease. A pyrimidine is inserted into the deep pocket of the P site on the ribonuclease, and the adenosine base is packed against the surface of the protein at the A site. Those are the only two bases that are closely bound by protein. The other bases are turned away from the protein, with the phosphates tucked in close to the surface, in contact with the lysines and arginines. It is primarily an electrostatic complex, with little dependence on hydrophobic interactions.

WICKSTROM: Have you tried crystallizing ribonuclease with the dodecanucleotide to make sure that your virtual DNA strand is not an artifact?

McPHERSON: We're attempting that. We've set up many crystal trials. We can vary the sequence to any one of many sequences. We can also vary the length. For example, we've already solved this complex with $d(pA)_6$, so we know what the hexamer does. The problem with using a dodecamer is that to get it to bind like the virtual strand we need a pyrimidine in the sequence, to fill the pyrimidine site. The adenines won't occupy the pyrimidine site; we'd have to use $d(pT)_{12}$, for example. We've attempted crystallization with oligomers of eight and 12 bases, as well as the tetramers and hexamers, and we've obtained microcrystals in nearly every case, but we have not yet obtained crystals suitable for diffraction from the long oligomers.

WICKSTROM: Do crystals of ribonuclease with the hexamer resemble those you got with the tetramer?

McPHERSON: They were isomorphous, in the sense that the unit cell and symmetry were the same. There was an 18% residual difference between the $d(pA)_6$ and the $d(pA)_4$ crystals. The protein is in the identical position. That's what provided phasing, which is why we could solve other complexes very rapidly; we just inserted the protein in its known position. When we did that with $d(pA)_6$, the protein was essentially the same, but many of the DNA-oligomer interactions were quite different. There was a much more complex arrangement. With 24 nucleotides associated with each protein molecule there were more intermolecular nucleotides. The helical filaments look quite different: there are long-period helical filaments, but they are not nearly so pronounced as in the $d(pA)_4$ complex. When we use longer oligomers we get a substantial modification of the DNA distribution in the unit cell. This is in part a result of the fact that the protein only allows the nucleic acid a certain amount of space in the unit cell, and the nucleic acid must accommodate itself to the available space as well as to the interactions on the protein.

MOORE: Have the base-stacking interactions along the strand been lost?

McPHERSON: Yes. A lot of them have been lost because the DNA must fit into the available space allowed by the crystal. The protein determines the basic cell size and symmetry.

MOORE: So you would expect that base stacking in a real substrate would also be lost?

McPHERSON: Yes, especially at the active site. We did idealize the base orientations in the virtual strand. These are not the dispositions observed in the complex; we imposed the positions to improve its appearance. We observed some stacking in the crystalline complex, but not nearly as much as one might expect.

GLITZ: The primary structures of about 40 pancreatic ribonucleases are now known, mostly from Jaap Beintema. Have you looked at any of those sequences to see how much conservation there is of the positively charged residues you see interacting with your oligomer?

McPHERSON: That analysis was done some time ago, but not by us, and I can't tell you the exact results. You're correct about the number of sequences. How much preservation of those residues there is, is a good question.

MAKOWSKI: One of your paper's referees asks: Why do you consider an oligodeoxynucleotide containing purines to be a good model for the catalytically productive interaction of oligoribonucleotides with ribonuclease, given the fact that the enzyme is known to be specific for pyrimidine units in RNA?

McPHERSON: The referee is asking whether the DNA single-strand is like the RNA single-strand that the protein actually degrades and acts upon catalytically. Almost all spectral studies and competition studies have shown that DNA binds in almost exactly the same way as a single-stranded RNA chain. A number of studies in the literature have shown quite convincingly that that is true. The most extensive study was by von Hippel and Jensen, who did a tremendous amount of work on the binding of pancreatic ribonuclease to single-stranded DNA and showed that the interactions were basically the same.

MAKOWSKI: Another question from the referee: Can you assume that the conformational preferences of RNA and DNA oligonucleotides are sufficiently similar that the oligodeoxynucleotide used in this study is, in fact, a good mimic of the interaction of the enzyme with the RNA?

McPHERSON: Why are we using poly dA instead of pyrimidines when ribonuclease is known to be specific for pyrimidines? That is a perfectly legitimate criticism. We started with pA because that is what we happened to have on hand when we tried to grow our crystals. We got beautiful crystals and, needless to say, we could not resist going ahead and solving the structure. The fair criticism is that ribonuclease doesn't work that well with pA, so we then employed tetramers of deoxy pT in the crystallization attempts and, though it was more difficult than with pA, we obtained beautiful crystals with those as well. Because, with respect to the protein, these crystals were essentially isomorphous with the $d(pA)_4$ complex, we inserted the protein, and within two days of obtaining the native data we had a map that showed the positions of the $d(pT)_4$ tetramers. That structure has just been refined by Stan Kozolak in our laboratory. The structure appears to have three ordered $d(pT)_4$'s in the structure, rather than four $d(pA)_4$'s. One of the d(pT)4's extends into the binding cleft of the protein, just as one $d(pA)_4$ does, except that it is offset by one nucleotide in the 5′ direction because the 5′ terminal pyrimidine is now in the pyrimidine binding site. So the pT goes all the way through the active site, with the 5′ terminal phosphate sticking out. Now we can see two situations: the $d(pT)_4$ complex, where the pyrimidine binding site is filled, and the $d(pA)_4$ complex, where the 5′ phosphate is held by lysine 41 and terminates in the active site. It appears that the phosphate at the site of catalysis may not be at exactly the same site in the two complexes; it is displaced by an ängström or two, which is a large difference. The other phosphates are superimposable in the two complexes.

RAGHAVENDRA: What is the backbone conformation of the oligonucleotide? Does it conform to any particular conformation, like A, B, C, or Z?

McPHERSON: No. The backbone is in a very extended conformation. It is not similar to any of the known forms of DNA. However, the conformation of single-stranded DNA is very difficult to define because it has so many conformational possibilities available to it. A refinement program has not yet been written that will properly refine single-stranded DNA simultaneously with protein. We're using Joel Sussman's CORELS program, which is about as good as there is. But there are so many energy minima available to single-stranded DNA that unless you

start very close to the correct conformation from your electron density maps, I'm not sure that you're going to end up in exactly the right conformation.

MAKOWSKI: How far apart are the phosphates?

McPHERSON: They vary from ~ 5 to 6.5 A.

CHANDRASEKARAN: It is very difficult to say what kind of conformation single-stranded DNA would take on. If the phosphorus-phosphorus distance is ~ 6 A, then it can be almost any structure, and you are finishing with ~ 72 A for 12 nucleotides. That means the bases could be apart by as much as 6 A, if you follow a helical path. This might not be the energetically most favorable conformation, but it is interacting with a protein. Until you have the high-resolution data, it is going to be difficult to see more about the structure. However, it is encouraging to know the kind of interactions that it can have, particularly the arginines and lysines interacting with the phosphates.

BURNETT: How well can you see the DNA? Do the bases have high temperature factors?

McPHERSON: The nucleotides that occur at the active site have thermal parameters no higher than the average for the protein. The phosphate groups that are directly bound by lysine or arginine have equally reasonable temperature factors. The nucleotides that extend through solvent are very poorly ordered, and have high temperature factors of ~ 50–60. Thus, portions of the DNA that are making bonds with the protein or with other DNA have fairly low temperature factors, and are clear on the difference map; the intermolecular portions of DNA are very poorly defined.

BURNETT: Is that why the bases in the solvent region were drawn with idealized stacking?

McPHERSON: That's right.

HENDRICKSON: In the $d(pA)_6$ complex, are there pieces that span the junctures between nucleotides that occur in the $d(pA)_4$ complex?

McPHERSON: No, but in the $d(pT)_4$ structure there are some.

HENDRICKSON: In the active site tetramer, or in other tetramers as well?

McPHERSON: Certainly in the active site, because the tetramer is advanced one position along the chain, so the gap that occurs in the $d(pA)_4$ complex between the first tetramer and the second tetramer is spanned by the first tetramer of the $d(pT)_4$ complex. We also see a gap jump near lysine 91. We don't see gap fills in the $d(pA)_6$ structure.

BAYLEY: Is there evidence from solution studies for the affinity of these small oligomers for ribonuclease? Can you show the stoichiometric binding of up to four oligomers?

McPHERSON: To my knowledge, no one has studied that.

BAYLEY: Could you hazard a guess as to the relative affinities of the oligonucleotides?

McPHERSON: The ones that go to the active site probably have very good affinity, because they make a lot of lysine-phosphate contacts and they have the hydrophobic interactions at the active site. My guess is that the affinities for the others would not be very high. I would, by the way, encourage anyone who sees this as an attractive system for their own techniques to apply them. We're doing crystallography, but little more than that on these systems.