Title
Crystal structure of RNase A complexed with d(pA)4

Journal
Journal of Molecular Biology, 189(2)

ISSN
0022-2836

Authors
McPherson, Alexander
Brayer, Gary D
Morrison, Robert D

Publication Date
1986-05-01

DOI
10.1016/0022-2836(86)90512-7

J. Mol. Biol. (1984) **189**, 305–327

Crystal Structure of RNase A Complexed with $d(pA)_4$

Alexander McPherson†, Gary D. Brayer‡ and Robert D. Morrison

Department of Biochemistry
University of California
Riverside, CA 92521, U.S.A.

(Received 23 August 1985, and in revised form 31 December 1985)

Co-crystals of pancreatic RNase A complexed with oligomers of $d(pA)_4$ were grown from polyethylene glycol 4000 at low ionic strength and the X-ray diffraction data were collected to 2·5 Å resolution. From a series of heavy-atom derivatives a multiple isomorphous replacement-phased electron density map of the RNase–$d(pA)_4$ complex was calculated to 3·5 Å. By inspection, the disposition of the known structure of RNase in the unit cell was determined and this was confirmed by calculation of a standard crystallographic residual, R. Refinement of the protein alone in the unit cell as a strictly rigid body yielded an R factor of 0·32 at 2·8 Å resolution. From difference Fourier syntheses DNA fragments were elucidated and incorporated into a model of the complex. The entire asymmetric unit was refined using a restrained–constrained least-squares procedure (CORELS) interspersed with difference Fourier syntheses. At the present time the crystal structure has been refined to an overall R value of 0·215 at 2·5 Å resolution.

The asymmetric unit of the complex crystals contains four oligomers of $d(pA)_4$ associated with each molecule of RNase. In addition, there may also be partially ordered fragments of DNA at low occupancy present in the unit cell, but these have not, at this time, been incorporated into the model. One tetramer of $d(pA)_4$ is entirely bound by a single protein molecule and occupies a portion of the active site cleft, filling the purine binding site and the phosphate site at the catalytic center with its 5′ nucleotide. Two other tetramers are partly intermolecular. One passes from near the pyrimidine binding site over the surface of the protein toward arginine 39 and into a solvent region. A third tetramer is anchored at its 5′ terminus by a salt link to lysine 98, passes near arginine 39 and then through a solvent region to terminate with its 3′ end near the surface of another protein molecule in the lattice. The fourth tetramer of $d(pA)_4$ is bound at its 5′ end on the opposite side of the protein from the active site in an electropositive anion trap that includes lysines 31 and 91 as well as arginine 33. There may be a DNA–DNA interaction involving the 5′ phosphate of one tetramer and the 3′ bases of two other tetramers and this may help to stabilize the crystalline complex.

If the sites of interaction between the protein and the $d(pA)_4$ fragments are mapped on the surface of the protein, they describe a nearly 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. Such a virtual DNA strand could explain the observation that, when RNase binds to a long single strand of nucleic acid, it can cover or protect 11 to 12 bases along the polynycleotide chain. The path would also account for the observation that at least seven individual electrostatic interactions are involved in the binding of RNase to single-stranded nucleic acids.

1. Introduction

Pancreatic ribonuclease A (RNase) is a protein of molecular weight 13,800, having a sequence of 124 amino acids that is responsible, in higher animals, for the enzymic degradation of RNA (Anfinsen & White, 1961; Richards & Wyckoff, 1971; Smyth *et al.*, 1963). Though predominantly associated with digestive processes, it is widely distributed among most tissues in mammals (Smyth *et al.*, 1963; Sierakowska & Shugar, 1977). It exists as well in a glycosylated form known as RNase B, which has a carbohydrate chain of known structure (Liang *et al.*,

† Author to whom all correspondence should be addressed.

‡ Present address: Department of Biochemistry, Medical School, University of British Columbia, Vancouver, British Columbia, Canada.

0022-2836/86/100305-23 $03.00/0

1980), covalently linked to asparagine 34 (Plummer *et al.*, 1968). The enzyme has been shown to cleave RNA following uridine or cytidine by a 2′–3′ cyclic monophosphate intermediate to yield fragments having terminal pyrimidines with 3′ phosphate groups (Usher *et al.*, 1972). A number of amino acid residues have been identified at the active site that appear essential for enzymic activity, and these include histidines 12 and 119 and lysine 41. The chemical, physical and enzymic properties of RNase A have been extensively reviewed (Anfinsen & White, 1961; Richards & Wyckoff, 1971; Blackburn & Moore, 1982; Wlodawer, 1984).

The three-dimensional structure of pancreatic RNase has been determined independently by three separate laboratories using X-ray diffraction analysis (Wyckoff *et al.*, 1970; Kartha *et al.*, 1967; Carlisle *et al.*, 1974) and independently refined to high resolution by two laboratories (Wlodawer & Sjolin, 1983; Borkakoti *et al.*, 1983). In one case, the refinement incorporated X-ray and neutron diffraction data to provide co-ordinates for hydrogen atoms as well (Wlodawer & Sjolin, 1983). The results of diffraction analyses both completed and in progress have also been reviewed (Wlodawer, 1984). It is fair to say that the atomic structure of RNase A is among the most precise and thoroughly studied of all of the known protein structures.

RNase A (and RNase B) bind to DNA with an affinity comparable to that for RNA (Walz, 1971; Sekine *et al.*, 1969). Because of the absence of 2′ hydroxyl groups on the nucleotides, obligatory for catalysis, the DNA is not cleaved but forms a stable complex with the enzyme. Although RNase A binds to the native, double-stranded form with a stabilizing effect, it preferentially complexes with single-stranded regions of DNA to produce destabilization and ultimate unwinding of the duplex form (Felsenfeld *et al.*, 1963; Jensen & von Hippel, 1976; von Hippel *et al.*, 1977). Hence RNase A is a primitive kind of DNA unwinding protein, or helix destabilizing protein, though less sophisticated than the gene *5* protein from bacteriophage fd (Alberts *et al.*, 1972) or the gene *32* protein from bacteriophage T4 (Alberts *et al.*, 1968), which are highly co-operative in their binding.

Although the structures of native RNase A and of its complexes with several dinucleotides are known from crystallographic studies (Wlodawer, 1984; Wyckoff *et al.*, 1977; Pavlovsky *et al.*, 1978; Wodak *et al.*, 1977; Wlodawer *et al.*, 1983; Borkakoti, 1983; Borkakoti *et al.*, 1982, 1983) and though its active site has been extensively investigated by an array of techniques, certain characteristics remain a puzzle. One aspect of RNase A binding that has never been satisfactorily explained is that the protein can cover or protect from 8 to 12 nucleotides along a single polynucleotide strand (Jensen & von Hippel, 1976; Record *et al.*, 1976). A second observation is that the protein can bind to double helical DNA with little perturbation of the native DNA conformation (Jensen & von Hippel, 1976). The active site of the enzyme, given its shape and limited extent, could not readily account for either of these properties. In addition to presenting a perplexing problem, the possibility of forming complexes between RNase A and DNA presented an attractive opportunity for the study of protein–nucleic acid complexes. Such a multi-component system would, we believed, allow visualization of an array of chemical interactions whose properties might be representative of the general class of such complexes.

To this end we attempted to obtain crystalline specimens of RNases A and B complexed with fragments of DNA. As previously reported (Brayer & McPherson, 1981), we were successful in our efforts and obtained large single crystals of RNase A complexed with $d(pA)_4$ and with $d(pA)_6$, as well as crystals of RNase B complexed with $d(pA)_4$ and $d(pA)_6$. Crystals of RNases A and B complexed with other oligonucleotides such as $d(pT)_4$ have also been grown and these are currently under investigation.

Chemical analysis of thoroughly washed crystals of RNase A–$d(pA)_4$ indicated the presence of between two and three oligomers of $d(pA)_4$ for each molecule of the protein (Brayer & McPherson, 1981). The volume of the unit cell, however, is consistent, depending on solvent content, with up to four or five oligomers of $d(pA)_4$ per protein molecule. Because of the uncertainty of the stoichiometry, no assumption was made as to the nature of the crystalline complex during the structure analysis.

We describe here the solution, to 2·5 Å resolution, of the structure of the RNase A plus $d(pA)_4$ crystals and its subsequent refinement by crystallographic least-squares techniques. On the basis of these results, we suggest how the structure might provide an explanation for some observations regarding nucleic acid binding.

2. Materials and Methods

The general approach to the solution of this structure was to obtain a multiple isomorphous replacement-phased electron density map at sufficient resolution and of adequate quality to allow accurate placement of the known structure of RNase A in the unit cell. The placement was then refined, with the protein molecule treated as a strictly rigid body. Difference Fourier syntheses were utilized to obtain the positions of the nucleic acid fragments and finally the entire crystal structure was refined by constrained–restrained crystallographic least-squares procedures (CORELS). The various stages of the structure solution are described below.

(a) *Crystallization*

RNase A (bovine) was purchased from Sigma Biochemicals (type V, 5 × recrystallized) and dissolved to a concentration of 20 mg/ml in distilled water. No further purification of the material was carried out and no other salts or small molecules were added to the solution. On sodium dodecyl sulfate/polyacrylamide gradient gels (5% to 20%), a major band corresponding to the correct molecular weight for RNase A was observed with perhaps

5% of the total material appearing in other protein bands. The d(pA)$_4$ was purchased from Collaborative Research of Waltham, MA, and was not purified further. Crystallization was carried out by combining 8 μl of the protein solution with d(pA)$_4$, so that the final volume was 10 μl and the molar ratio of oligomer to protein was 5 : 1. This was mixed in the well of a glass depression plate (Corning no. 7720) with 10 μl of 16% to 20% polyethylene glycol 4000. The depression plate containing 9 samples was sealed in a plastic box and allowed to equilibrate against 25 ml of 16% to 20% polyethylene glycol 4000 at 4°C. Crystals generally appeared in 1 to 2 weeks with good reproducibility and yield. The general procedures for vapor diffusion crystallization are described by McPherson (1982).

The crystals were examined by precession photography and, as reported by Brayer & McPherson (1981), are of space group $P2_12_12_1$ with $a = 44{\cdot}4$ Å, $b = 75{\cdot}3$ Å and $c = 44{\cdot}6$ Å. Although it was certain that there could be but one protein molecule in the asymmetric unit of the crystal, it was not clear how many d(pA)$_4$ oligomers were associated with each RNase molecule. This number remained in question until the final stages of the structure determination.

(b) *Data collection*

X-ray diffraction data for the preliminary survey of crystals and evaluation of potential heavy-atom derivatives were obtained photographically using Buerger precession cameras with GX6 and GX20 Elliott rotating anode generators as sources. These were operated at 40 kV and 40 mA with a focal spot size of 200 μm^2. Diffraction data employed in the structure determination itself were collected on an Enraf-Nonius CAD-4 automated diffractometer with an extended counter arm and helium-filled path. Data were collected on a native crystal to 2·5 Å Bragg spacings using an ω scan of 0·60° width. The generator was operated at 40 kV and 32 mA with a fine focus X-ray tube and the scan rate was 1·0°/min. Data were collected at 20°C and an entire 2·5 Å set of 3485 Friedel pairs was obtained from a single crystal. Backgrounds on both sides of the scan were collected for every reflection, and the conventional R factor for symmetry-related reflections was 0·035 on $|F|$ for all measurements above 1 estimated standard deviation. An empirical absorption correction curve was obtained by measuring a phi-independent reflection at 72 consecutive intervals of 5°. Three standard reflections were collected every 1·5 h of X-ray exposure time to monitor crystal deterioration.

The integrated intensities were corrected for Lorentz polarization effects, absorption, and backgrounds were subtracted. The backgrounds for each reflection were taken as the local average of backgrounds for the 24 nearest neighbors in reciprocal space. The intensities were merged and converted to structure amplitudes. Standard deviations were determined from counting statistics and from merging residuals. The structure amplitudes were placed on an absolute scale using the program ORESTES written by Thiessen & Levy (1973), which employs a least-squares Wilson procedure. Only the protein was assumed to be present in the unit cell at this stage.

Diffraction data for the heavy-atom derivative crystals were obtained, reduced and corrected as described for the native crystals. In general, only data from 30·0 Å to either 5·0 Å or 3·5 Å Bragg spacing were collected, since high resolution m.i.r.† was not intended. In all cases Friedel pairs were collected at $\pm 2\theta$ to provide a measure of anomalous dispersion differences. It was observed that decay, negligible in the native case, was very severe in all of the heavy-atom derivative crystals. Those containing uranyl ions yielded only 5·0 Å data before declining in average intensity by nearly 50%. Platinum-containing derivatives generally yielded 3·5 Å data but only if a 40% decline in the intensity standards was allowed. The mercury acetate derivative was most stable, but was also the least substituted.

The derivative data sets were scaled to the native structure amplitudes using a Fourier–Bessel procedure programmed by L. Weissman at UCLA, as modified by D. Cascio and R. Williams of this laboratory. The R_{sym} and R factors for derivative to native scaling are shown in Table 1. Before applying this procedure, all structure amplitudes less than 3·0 times their estimated standard deviation were eliminated thus decreasing the 2·5 Å data set to 3363 independent reflections.

(c) *Isomorphous replacement*

Patterson difference syntheses were calculated for each derivative at 30·0 to 5·0 Å, 30·0 to 4·0 Å and 10·0 to 4·0 Å resolution, where possible. Those using 10·0 to 4·0 Å data only were found to be least ambiguous in their interpretation though correct heavy-atom vectors were found to be consistently present in all maps. The difference Patterson maps for K_2PtCl_4 and $PtBr_2(NH_3)_2$ were found to be nearly identical and interpretable in terms of a single major site. The major site of the UNO_3 derivative was also found independently from its 5·0 Å Patterson map while the minor sites were not. No anomalous difference Patterson map was of any value in confirming the solutions.

Single isomorphous replacement phases to 5·0 Å resolution were calculated using each of the 2 platinum derivatives and the uranyl derivative separately and difference Fourier syntheses computed for each of the other derivatives not included in the phase calculations. The uranyl-based phases were successful in unambiguously identifying the major platinum sites for both K_2PtCl_4 and $PtBr_2(NH_3)_2$ at those positions deduced from the difference Patterson maps. Similarly, the platinum based phases calculated from either K_2PtCl_4 or $PtBr_2(NH_3)_2$, demonstrated in difference Fourier maps the major site for the uranyl compound deduced from its difference Patterson map. In addition, the minor sites for the uranyl derivative were determined from these difference Fourier maps.

Repeating the difference Fourier analyses with the heavy-atom configurations reversed for each derivative but the sign of the anomalous dispersion components constant, and using the magnitude of the relevant difference peaks as criteria (Blundell & Johnson, 1976), the correct heavy-atom configurations for each derivative were found.

The m.i.r. phases were computed using the K_2PtCl_4, $PtBr_2(NH_3)_2$ and uranyl derivatives to 5·0 Å resolution and difference Fourier maps calculated for all other potential heavy-atom derivatives. In this way, the major sites for the high concentration K_2PtCl_4 and for $HgAc_2$, and the minor sites for all derivatives were determined. Sites for these latter derivatives were tested for validity by their capacity for yield s.i.r. phases that produced

† Abbreviations used: m.i.r., multiple isomorphous replacement; s.i.r., single isomorphous replacement; e.s.d., estimated standard deviation; Ac, acetate.

Table 1
Scaling statistics for heavy-atom derivatives

Heavy-atom compound	Crystal	R_{sym}†	No. reflections observed	$>3\sigma$	R_{fit}‡
Native	1	0·047	3547	3210	...
K_2PtCl_4-I	1	0·034	1441	1345	0·203
$PtBr_2(NH_3)_2$	1	0·045	1387	1222	0·164
K_2PtCl_4-II	1	0·047	1212	1122	0·244
	2	0·053	648	597	...
	3	0·092	1492	1213	...
UNO_3	1	0·035	674	627	0·173
$HgAc_2$	1	0·054	1972	987	0·132
	2	0·108	1559	605	...
K_2PtCl_6	1	0·045	1010	936	0·169
	2	0·042	953	889	...
	3	0·046	394	335	...
p-Chloromercuribenzoate	1	0·035	1634	1530	0·090

† $R_{sym} = \dfrac{\sum_{hkl} \| |F^+| - |F^-| \|}{\sum_{hkl} \dfrac{|F^+| + |F^-|}{2}}$.

‡ $R_{fit} = \dfrac{\sum_{hkl} \| |F_{HA}| - |F_{NAT}| \|}{\sum_{hkl} (|F_{HA}| + |F_{NAT}|)/2}$.

Table 2
Refined parameters for heavy-atoms used in m.i.r. phase calculations

	A	*B*	*x*	*y*	*z*
K_2PtCl_4-I					
Site 1	1·00	56·60	0·1057	0·8446	0·8240
$PtBr_2(NH_3)_2$					
Site 1	0·95	86·52	0·0954	0·8448	0·7983
UNO_3					
Site 1	0·54	40·79	−0·0068	0·4422	0·3069
Site 2	0·13	30·23	0·7546	0·1743	0·7237
Site 3	0·25	0·00	0·4072	0·1664	0·5987
K_2PtCl_4-II					
Site 1	0·91	44·28	0·8972	0·3393	0·6686
Site 2	0·27	0·71	0·4420	0·4344	0·6952
Site 3	0·29	0·00	0·7925	0·1602	0·6655
K_2PtCl_6					
Site 1	0·22	1·99	0·7942	0·1477	0·6658
$HgAc_2$					
Site 1	0·21	6·61	0·2705	0·3224	0·8043

correct peaks in difference Fourier maps of the K_2PtCl_4 and uranyl derivatives.

The individual heavy-atom derivatives were refined by minimizing the difference between the calculated and observed intensity difference for each derivative. The refined co-ordinates, thermal factors and occupancies for all sites are shown in Table 2. This was essentially refinement against an origin-removed Patterson map, as suggested by Rossmann (1960) and programmed in this case by T. Terwilliger at UCLA (Terwilliger & Eisenberg, 1984). This method assumes a considerably more stringent error estimate, E_j, for non-centric reflections, and uniformly yields a lower figure of merit than the Blow–Crick procedure more commonly used (Blow & Crick, 1959). The figure-of-merit distributions for these data are shown in Table 3.

The refinement statistics for the 4·0 Å data eventually used to produce the m.i.r. phases for the native electron density map are shown in Tables 3 to 5. Derivatives produced from high and low concentrations of K_2PtCl_4 and the $PtBr_2(NH_3)_2$ derivative gave the same major site of substitution, although the high concentration K_2PtCl_4 was also accompanied by minor sites. The uranyl derivative was obtained in such an irreproducible manner that only one 5·0 Å data set was included in the phasing.

Table 3
Figure of merit with resolution

									Total
Resolution (Å)	12·27	8·48	6·84	5·89	5·25	4·78	4·42	4·12	
No. of observations	73	100	84	105	111	165	202	204	1044
Mean figure of merit	0·60	0·67	0·62	0·67	0·61	0·57	0·48	0·47	0·56
Centric f_m	0·681	0·760	0·697	0·759	0·717	0·574	0·057	0·583	0·645
Non-centric f_m	0·569	0·661	0·648	0·672	0·615	0·591	0·500	0·438	0·528

f_m, figure of merit.

Table 4
Centric R *factor*

									Total
Resolution (Å)	12·27	8·48	6·84	5·89	5·25	4·78	4·42	4·12	
K_2PtCl_4-I	0·51	0·52	0·59	0·57	0·43	0·70	0·63	0·77	0·57
$PtBr_2(NH_3)_2$	0·57	0·53	0·49	0·43	0·62	0·72	0·69	0·78	0·62
UNO_3	0·70	0·85	0·93	0·44	0·59	0·73	0·00	0·00	0·71
K_2PtCl_4-II	0·61	0·49	0·42	0·77	0·58	0·48	0·71	0·72	0·59
K_2PtCl_6	0·74	0·67	0·86	0·69	0·83	0·58	0·73	0·70	0·71
$HgAc_2$	0·65	0·70	0·60	0·76	0·65	0·67	0·79	0·69	0·69

Table 5
r.m.s. (FH)/r.m.s (E)

									Total
Resolution (Å)	12·27	8·48	6·84	5·89	5·25	4·78	4·42	4·12	
K_2PtCl_4-I	0·63	1·04	1·50	1·15	1·07	1·01	0·71	0·62	0·88
$PtBr_2(NH_3)_2$	0·80	1·15	1·27	1·28	0·89	0·72	0·52	0·37	0·78
UNO_3	0·67	1·14	1·30	1·34	1·07	1·55	0·00	0·00	1·06
K_2PtCl_4-II	0·75	1·12	1·54	1·31	0·96	0·96	0·94	1·01	1·01
K_2PtCl_6	0·63	0·72	0·74	0·63	0·60	0·57	0·64	0·55	0·61
$HgAc_2$	0·48	0·50	0·61	0·62	0·68	0·49	0·47	0·40	0·50

r.m.s., root-mean-square; FH, heavy-atom structure factor; E, lack of closure.

K_2PtCl_6 was weakly substituted and $HgAc_2$ was also of low substitution. Phases based on $HgAc_2$ were adequate, however, in demonstrating the major Pt sites in difference Fourier maps of K_2PtCl_4 and $PtBr_2(NH_3)_2$.

The overall figure of merit for the 1465 reflections incorporated in the 3·5 Å native electron density map was 0·56, and for the centric reflections alone it was 0·68. The set of heavy-atom derivatives used to produce the m.i.r. phases was hardly ideal. Although the uranyl derivative phased well, it extended only to 5·0 Å resolution. Both the high and low concentrations of K_2PtCl_4 showed indications of substantial non-isomorphism and this was principally why the electron density map was calculated at no more than 3·5 Å resolution. K_2PtCl_6 was weakly substituted. $HgAc_2$, which was isomorphous, was also only weakly substituted. $PtBr_2(NH_3)_2$ was probably the best overall derivative. Error estimates from the heavy-atom refinement procedure indicated that the anomalous dispersion data from all derivatives were useful to about 5·0 Å resolution but was at best marginal beyond that point.

At the conclusion of refinement of the RNase–d(pA)$_4$ structure, which is described below, calculated phases from the model ($R = 0·22$ at 2·5 Å resolution) were compared with the m.i.r. phases that had been utilized in computing the electron density maps. The overall average phase difference between m.i.r. and calculated phases was 63·6° for the 1155 acentric reflections and the proportion of centric reflections for which the sign of the phase was correct was 62·5%.

The native electron density map was calculated at a variety of resolutions and it gradually became clear that for our purposes the map computed to 4·0 Å was most interpretable. The map we examined was calculated using the fast Fourier algorithm programmed by TenEyck *et al.* (1976) on a 1 Å grid spacing. The maps were automatically contoured and listed on a Printronix printer with a program written by D. Cascio and R. Williams. The map sections were transferred to acetate transparencies, mounted on Plexiglas and reviewed on a light box in the conventional manner. The scale of the maps was 2·46 Å/cm and a backbone model of RNase A on the same scale, produced with a Byron Bender (Rubin & Richardson, 1972), was used as a superposition aide.

The position and orientation of the RNase A molecule in the electron density maps was not clear by inspection. In retrospect, we know that this was because the strongest and most prevalent features at this resolution were the DNA components, while the protein molecule, in terms of contrast, was subdued. In addition, the DNA fragments link crystallographically equivalent RNase A molecules in the unit cell and this further complicated the choice of an orientation for the protein. The most dense features of the protein, the 3 helices for example, could not be discriminated readily, again because of the prominent DNA fragments.

A number of possible placements for the protein in the unit cell were identified by superimposing overlay images of the RNase structure on the map and examining for coincidence. This was done by choosing 12 or more points on the map and assigning to them α-carbon co-ordinates from the molecular overlays. The protein molecule was then rotated and translated into the intended orientation using a least-squares procedure (Rossmann & Argos, 1975). The RNase atomic co-ordinates were provided by A. Wlodawer and were based on a combined X-ray and neutron diffraction refinement (Wlodawer & Sjolin, 1983). The validity of the trial was determined by computing structure factors from the model and at 5·0 Å resolution quantifying their correlation with those observed by calculation of a conventional R factor. In every case, the orientation of the model was incrementally adjusted about its center of mass by a rigid-body refinement procedure to minimize the value of the R factor. This was accomplished using the program CORELS written by J. Sussman (1983, 1985). Incorrect placements for the protein generally resulted in convergence of orientation shifts at R values ranging

from 0·49 to 0·58. Approximately 10 incorrect placements were attempted before success was achieved.

A correct solution for the placement of the protein structure in the unit cell was obtained when: (1) the assumption was made that the major Pt site was at methionine 29, as it was found to be (Carlisle *et al.*, 1974; Wyckoff *et al.*, 1967) in the native structure; (2) the density corresponding to the N-terminal helix was finally recognized; and (3) the helix composed of residues 50 to 60 was also correctly identified. The helix composed of residues 26 to 33 was not recognizable because of local artifacts and non-isomorphism created by substitution of a heavy-atom at position 29, near its midpoint.

The correct placement of the molecule yielded an initial R factor, at 5·0 Å resolution, of 0·56, but the rotational parameters for the rigid body continued to change and when convergence at 5·0 Å was achieved the R value was 0·33. Inclusion of progressively higher resolution reflections with continued refinement of the protein as a rigid body yielded a final R value of 0·32 at 2·8 Å. We would like to note that the orientation of the RNase molecule underwent a rotation of nearly 22° as the R factor declined from 0·56 to 0·32, attesting to the substantial radius of convergence of the procedure in this particular case at least.

Difference Fourier syntheses with coefficients F_0-F_c and phases ϕ_c were calculated at a variety of resolutions. The most interpretable at this stage was the 3·5 Å map. Initially, 2 of the $d(pA)_4$ fragments were quite visible and portions of others, those most firmly bound by the protein, were observed in the difference Fouriers as well. The oligomers that could be clearly discerned were incorporated into the structure factor calculations, while the segments of those partially seen were entered as mononucleotides, dinucleotides or trinucleotides. As refinement progressed with inclusion of more DNA and higher resolution reflections, additional difference Fourier maps were computed. These rounds of refinement and difference Fourier analysis were continued until the R factor was minimized, and all of the DNA fragments associated with an assymmetric unit were reconstructed. Because portions of the $d(pA)_4$ oligomers have relatively high thermal parameters, a series of maps was generally computed at various resolutions and with a variety of phasing models to achieve consistency. Fig. 1 is a composite F_0-F_c difference Fourier map showing the density at low resolution (4·0 Å) arising from the $d(pA)_4$ oligomers with respect to the protein molecule.

(d) *Refinement of the structure*

The structure of the RNase–$d(pA)_4$ complex was refined at all stages, ultimately using 3363 reflections to 2·5 Å resolution, by constrained–restrained least-squares procedures. Throughout only reflections greater than 3 times their estimated standard error were included and unitary weights were used. Some refinement using weights based on the e.s.d. was carried out but it did not appear to yield results significantly different from those obtained by using unitary weights. The program employed was CORELS written by Sussman (1983, 1985), a version of October 1983. All calculations were performed on a UCR VAX 750 or URC VAX 780 running under VMS. Some relevant model and refinement parameters are shown in Table 6. The approach taken was conservative, with the initial assumption that little or no significant change would be found for the polypeptide backbone of the protein though amino acid side-chains would adjust to the presence of the nucleic acid. It was only with reluctance that conformation change was allowed, and indeed, it appears that little alteration was necessary.

The course of the structure refinement is summarized in Table 7. Initially, the protein was maintained as a strictly rigid body and only DNA oligomers were allowed to refine. Each nucleotide of a $d(pA)_4$ oligomer, or the

Figure 1. F_0-F_c difference Fourier maps were calculated of the RNase–$d(pA)_4$ complex crystals with each of the DNA oligomers omitted, leaving in each case the protein and 3 remaining $d(pA)_4$ strands. The drawing here is a composite difference Fourier synthesis made by combining those maps and shows the regions of density corresponding to each of the 4 $d(pA)_4$ oligomers. The resolution of the map was 4·0 Å. An image of the polypeptide backbone of RNase in the correct orientation is shown superimposed on the difference density.

Table 6
CORELS refinement parameters

No. reflections	3363
Unitary weights all reflections	
Resolution	10·0 to 2·5 Å
No. of parameters	1867
No. of atoms	1443
No. of protein atoms	950
No. of DNA atoms	493
No. of DNA constrained groups	32
No. of protein constrained groups	124

portion included in the refinement, was considered as 2 constrained (ideal) groups, the nucleoside and the phosphate. These were joined by restrained linkages. The degree of restraint was varied during the course of refinement to reconcile minimization of the conventional R factor with ideal bond lengths and angles. The procedure is described in detail by Sussman (1983, 1985) and has been ussed in the refinement of a number of proteins and nucleic acid molecules (Sussman, 1985). A note of caution is perhaps in order here regarding this technique. Because the individual nucleotide components are maintained ideal, all errors and deviations from ideality tend to be compensated or accounted for by adjustments to the restrained bond lengths and angles. Thus, it may appear that certain geometrical parameters, particularly the torsion angles and bonds along the polyphosphate backbone, have disproportionately large errors when compared with the overall geometrical parameters.

While the refinement of the orientation, structure and disposition of the protein molecules in the unit cell was relatively straightforward, the same was not true of the d(pA)$_4$ oligomers except in two cases. Initial difference Fourier syntheses calculated on the basis of the rigid protein model alone (R factor = 0·32 at 2·8 Å resolution) clearly demonstrated the presence of 2 d(pA)$_4$ tetramers bound to the protein molecule. In these 2 cases all 4 nucleotides in the DNA tetramer could be identified with reasonable certainty. In addition to these 2 oligomers, however, other strong, extended difference electron density peaks appeared in the $F_o - F_c$ map at the surface of the protein and in solvent regions adjacent to the RNase molecule.

Before interpreting the additional peaks, the 2 d(pA)$_4$ oligomers that could be clearly seen were incorporated into the phasing model and refinement at 2·8 Å continued until convergence was reached ($R = 0{\cdot}25$ at 2·8 Å resolution). Again a series of difference Fourier maps were calculated as a function of resolution. Many peaks seen previously were, in this map, enhanced compared to the original $F_o - F_c$ Fourier. An examination of the difference density peaks showed that while some were distant from any protein molecule, those that were persistently highest in magnitude were without exception at the immediate surface of a protein molecule, and invariably adjacent to lysine or arginine residues.

By successive rounds of constrained–restrained least-squares refinement at 2·5 Å resolution, alternating with difference Fourier syntheses, peaks were eventually linked in a manner consistent with d(pA)$_4$ oligomers. These oligomers were then incorporated into the model of the asymmetric unit and refined collectively as described above. The final R factors for the refinement as a function of resolution are shown in Table 8. The final overall residual was $R = 0{\cdot}215$ at 2·5 Å resolution. The deviations from ideal geometry are shown in Tables 9 and 10.

The final rounds of refinement included restrained rigid-body refinement of the individual amino acids, treated as ideal constrained groups along with their side-chain dihedral angles, as well as DNA restrained parameters. The geometrical constraints were rigorously maintained in all cycles of computation and shifts were initially damped to 0·10 to 0·25 for many cycles before relaxation to 0·5 to 0·66 of their full shift values. For all refinement cycles the geometry was permitted most freedom on the first cycle and then quadratically restrained on subsequent cycles. This approach was intended to prevent occurrence of dramatic changes in protein conformation and to maintain the DNA fragments as close as possible to positions consistent with the difference Fourier syntheses.

The course of refinement and the current model of the asymmetric unit was monitored using the program FRODO written by Jones (1982), running on the MPS Evans and Sutherland graphics system in the Chemistry

Table 7
Stages of constrained–restrained least-squares refinement (CORELS)

1 Refinement of RNase as a rigid body alone in the unit cell, no DNA present at 5·0 Å resolution. Final R factor, 0·33

2 Refinement of RNase as a rigid body alone in the unit cell, no DNA present at 2·8 Å resolution. Final R factor, 0·32

3 Refinement of constrained group's (PO_4, deoxyribose, adenine) rigid and dihedral parameters of 2 tetramers of DNA with protein as a rigid body at 3·5 Å

4 Refinement of constrained group's rigid and dihedral parameters for 4 tetramers of DNA with protein as a rigid body at 3·0 Å resolution

5 Refinement of constrained group's rigid and dihedral parameters for 4 tetramers of DNA and the side-chain dihedral angles of the protein at 2·8 Å resolution. R factor at this stage, 0·255

6 Alternated refinement of constrained DNA groups, rigid and dihedral with rigid and dihedral parameters of the individual amino acids treated as unique constrained groups at 2·5 Å resolution

7 Simultaneously refined all constrained groups, both rigid and dihedral parameters, of both protein and 4 tetramers of DNA at 2·5 Å resolution. R factor at this stage, 0·243

8 Refined all constrained group parameters of protein and DNA and the thermal parameters of all constrained groups at 2·5 Å resolution, final R factor, 0·223

Table 8
Distribution of R *factor*† *as a function of resolution*

D(min)	No. reflections	Average e.s.d.	$F_o - F_c$	Shell	Sphere
5·00	415	18·76	97·33	0·236	0·236
4·00	606	19·59	105·47	0·185	0·202
3·40	666	25·85	88·84	0·190	0·197
2·90	647	35·98	74·73	0·229	0·204
2·65	649	19·27	67·40	0·301	0·216
2·50	380	15·40	68·00	0·362	0·224
Overall ∞ to 2·5 Å. $R = 0{\cdot}223$ for 3363 reflections					

† $R = \dfrac{\sum_{hkl} |F_o| - |F_c|}{\sum_{hkl} (|F_o| + |F_c|)/2}$

Table 9
Deviations from geometrical ideality of restrained linkages

Restraint type		No. of restraints	Average deviation from ideal	r.m.s. deviation from ideal
C–N	Bond length	123	0·023	0·029
CA–N	Ang (CA–C–N)	123	0·071	0·092
C–CA	Ang (C–C–CA)	123	0·053	0·065
O–N	Ang (O–C–N)	123	0·052	0·075
O–CA	Tors ang omega	123	0·130	0·162
N–C	Ang (N–CA–C, tau)	124	0·061	0·074
N–CD	Bond length (Pro)	4	0·012	0·014
N–CG	Ang (N–CD–CG: Pro)	4	0·091	0·096
CA–CD	Ang (CA–N–CD: Pro)	4	0·032	0·037
C–CD	Ang (C(i–1)–N–CD: Pro)	4	0·074	0·091
S–S	Bond length disulfide	4	0·032	0·039
C–S	Ang (C–S–S, disulfide)	8	0·078	0·102
CA–CA	Tors ang omega	123	0·058	0·082
ODUM–N	Planar peptide lingage	123	0·127	0·157
ODUM–CA	Planar peptide linkage	123	0·105	0·131
O5′–C5′	DNA: bond length	16	0·097	0·268
P–C′	DNA: ang (P–O5′–C5′)	16	0·152	0·257
O5′–C4′	DNA: ang (O5′–C5′–C4′)	16	0·106	0·153
O3′–P	DNA: bond length	12	0·040	0·085
C3′–P	DNA: ang (C3′–O3′–P)	12	0·212	0·163
O3′–O1P	DNA: ang (O3′–P–O5′)	12	0·133	0·151
O3′–O2P	DNA: ang (O3′–P–O1P)	12	0·074	0·089
O1′–C4′	DNA: ang (O3′–P–O2P)	12	0·064	0·102
O1′–C5′	DNA: bond length	16	0·018	0·031
O1′–C5′	DNA: ang (O1′–C4′–C5′)	16	0·057	0·071
O1′–C3′	DNA: ang (O1′–C4′–C3′)	16	0·090	0·108
C1′–C4′	DNA: ang (C1′–O1′–C4′)	16	0·114	0·145
All restraints		1318	0·077	0·112

r.m.s., root-mean-square; Ang, angle; Tors ang, torsion angle.

Department at UCLA and using the programs and PS300-based graphics system assembled by the Computer Science Department at the University of North Carolina. Most illustrations were generated in our laboratory using a PDP 11/34 interfaced to an Advanced Electronics Design raster graphics device coupled to a Calcomp plotter.

3. Results

There are no major conformational changes involving the polypeptide backbone of the RNase molecule but there are numerous changes in the orientations of amino acid side-chains, particularly those on the surface of the protein and in contact with DNA. The average change in position from the native structure for all main-chain atoms is 0·24 Å and the root-mean-square change is 0·38 Å. Some individual alterations will be discussed below. The arrangement of the protein molecules alone in the crystal unit cell with no DNA present is shown in Figure 2.

Table 10
Deviation from ideality of non-bonded contacts

Type of contact		No. of restraints	Average deviation from ideal	r.m.s. deviation from ideal
O(N)–O(N)	Bump	254	0·347	4·64
C–N	Bump	113	0·710	0·811
C–O	Bump	238	0·629	0·696
C–C	Bump	125	0·927	0·976
All contracts		730	0·595	0·706

r.m.s., root-mean-square.

In the crystal, protein–protein intermolecular contacts appear to be rather spare along all three crystallographic directions and particularly in the Y and Z directions. The direction in which a significant number of contacts is apparent is the X direction, and these involve chiefly the loop formed by residues 62 to 72 and the irregular loop of polypeptide between residues 16 and 24.

Figure 3 (see p. 314) shows the structure of an asymmetric unit of the crystal consisting of one RNase A molecule plus four bound tetramers of $d(pA)_4$. In addition to protein–protein contacts, there are numerous contacts formed between asymmetric units that involve the DNA tetramers. That is, there are numerous instances in which a $d(pA)_4$ fragment is bound chiefly by one protein molecule but extends through a solvent region to make contact with another protein molecule in the lattice. Thus the $d(pA)_4$ fragments serve as bridges or crosslinkers to maintain the integrity of the crystal. Portions of the $d(pA)_4$ oligomers also approach one another at certain points in the crystal and form DNA–DNA interactions.

From the contents of the crystallographic unit cell, some useful parameters can be calculated for the crystals. The fraction of the crystallographic

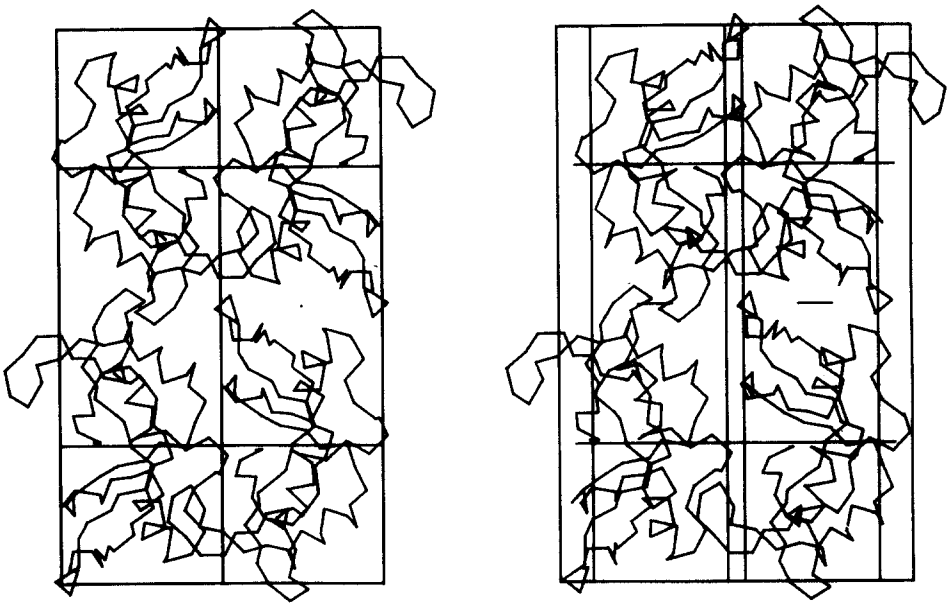

Figure 2. The packing of 4 RNase molecules in a single unit cell of the $P2_12_12_1$ crystal form as shown by the α-carbon backbones of the proteins. The stereo diagram is viewed along the crystallographic *c* axis.

unit cell occupied by protein is:

$$V_{\text{prot}} = \frac{(1{\cdot}66)(v_{\text{prot}})(M_{\text{prot}})}{V_{\text{unit cell}}},$$

and a corresponding fraction for DNA is:

$$V_{\text{DNA}} = \frac{(1{\cdot}66)(v_{\text{DNA}})(M_{\text{DNA}})}{V_{\text{unit cell}}},$$

where the partial specific volume $v = 0{\cdot}07$ and $0{\cdot}55$ for protein and DNA, respectively (Jensen & von Hippel, 1976). The volume of the unit cell $V_{\text{unit cell}} = 149{,}279$ Å^3, the total amount of protein in the unit cell is $4 \times 13{,}700 = 54{,}800$ M_r and, assuming four fully occupied tetramers of d(pA)$_4$, the total amount of DNA is $4 \times 4 \times 320 = 20{,}480$ M_r. The fraction of protein in the unit cells is, therefore, $V_{\text{prot}} = 0{\cdot}426$, $V_{\text{DNA}} = 0{\cdot}125$ and the total mass of protein plus DNA occupies 0·552 of the unit cell volume. The overall volume-to-mass ratio for these crystals (Matthews, 1968) can be calculated to be $V_m = 1{\cdot}98$ Å^3/dalton.

The environment of each of the heavy-atom sites used in the calculation of the m.i.r.-phased electron density maps was searched to identify the amino acid residues or DNA groups to which they were bound. Results of the examinations are shown in Table 11. The major platinum binding site for

Table 11

Environment of heavy-atom substitution sites

Heavy-atom derivative	*X*	*Y*	*Z*	Atoms of closest amino acids		Distance (Å)
K_2PtCl_4	0·1057	0·8446	0·8240	Met29	SD	2·64
	0·7925	0·1602	0·6655	Ala6	O	2·53
				Glu9	CG	1·21
	0·4420	0·4344	0·6952	Glu86	OE2	3·41
				Ribose	C4′	3·51
K_2PtCl_6	0·7942	0·1477	0·6658	Ala5	O	2·57
				Ala6	CA	2·65
				Glu9	CG	0·82
$PtBr_2(NH_3)_2$	0·0954	0·8448	0·7983	Met29	SD	2·25
UNO_3	0·9940	0·4422	0·3069	Glu111	OE1	1·65
				Adenine	N6	2·20
				PO_4	O1	2·73
	0·7546	0·1743	0·7237	Phe8	C	1·37
				Glu9	N	1·55
				Glu9	OE1	3·86
	0·4072	0·1664	0·5987	PO_4	O1	4·06
$HgAc_2$	0·2705	0·3224	0·8043	Tyr76	O	2·42
				His105	CE1	3·06
				PO_4	O1	2·54

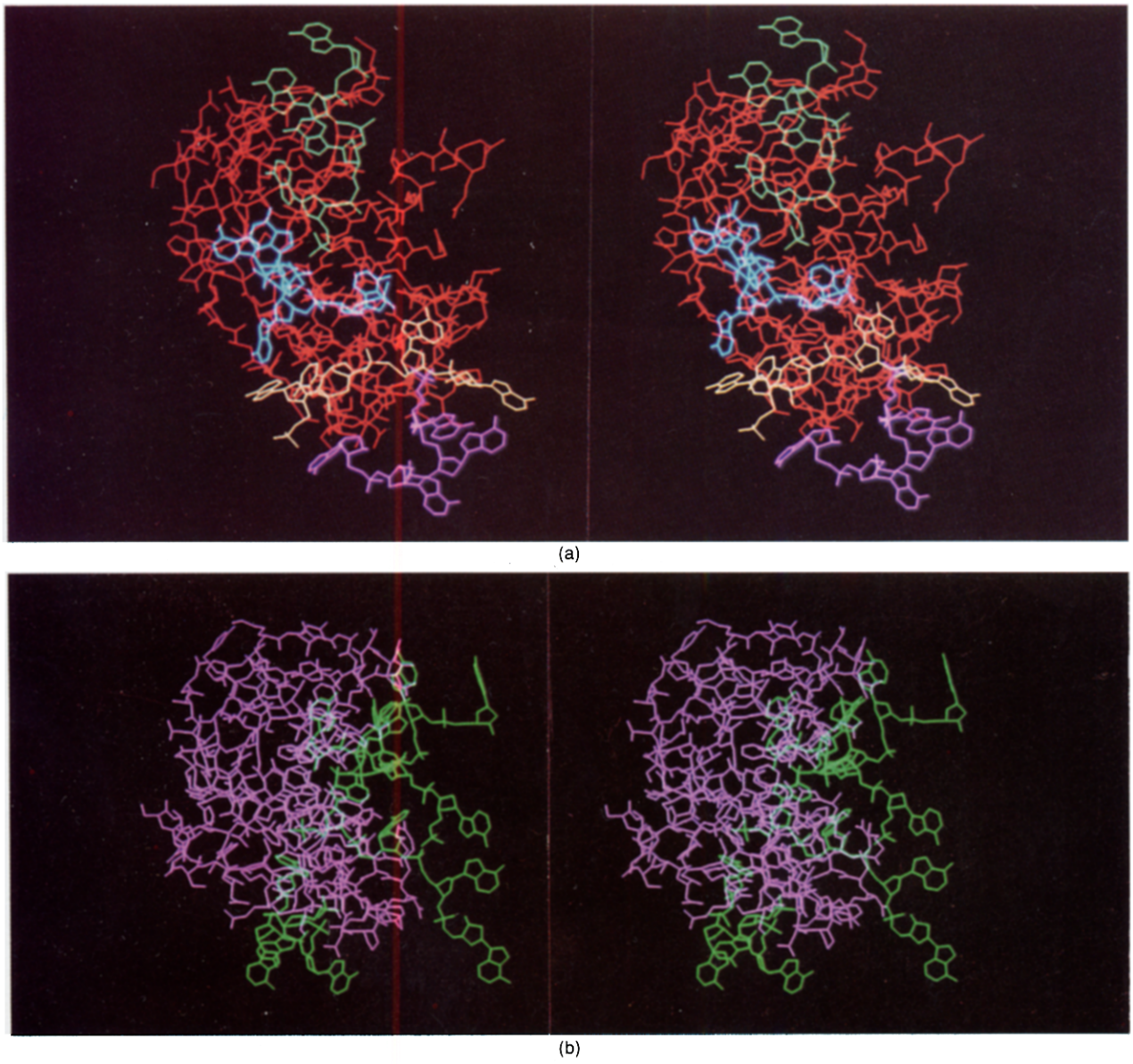

Figure 3. In stereo drawings (a) and (b), RNase is seen associated with the 4 tetramers of $d(pA)_4$ that together comprise an asymmetric unit of the crystal. Running more or less from the bottom of (a) to the top in a consistent 5′ to 3′ direction, the 4 oligomers of DNA, which are referred to in the text as 5′-A, B, C, D-3′ (violet, yellow, blue and green, respectively), can be seen to trace out a virtually continuous path through the active site cleft of the enzyme and over the surface of the protein (in red). Electrostatic linkages are made between phosphates on the oligonucleotides and lysines, 7, 31, 41, 66 and 98 and possibly 37, as well as with arginines 39 and 85 and probably 33. More specific interactions involving the ribose and base moieties occur in the active site cleft. In a second orientation (b), the protein is shown in violet and all oligomers of DNA in green. Note the protrusion of some of the tetramers into solvent regions.

K_2PtCl_4 and $PtBr_2(NH_3)_2$ was the sulfur group of methionine 29, as was found to be the case for K_2PtCl_4 in the native structure (Carlisle *et al.*, 1974; Wyckoff *et al.*, 1967). UNO_3 is apparently bound by glutamic acids 111 and 9, with some interaction with DNA phosphate groups as well. Mercury acetate is bound by histidine 105.

(a) *Description of the structure*

(i) *The $d(pA)_4$ oligomers*

There are associated with each RNase A protein molecule four tetramers of $d(pA)_4$ that are well-ordered and consistent with the Fourier and difference Fourier maps. Two of these, as noted previously, are particularly well fixed and certain. What appear to be disordered fragments of others are barely discernible within solvent regions in the unit cell that cannot, at this stage of refinement, be confidently included in the model.

The RNase A protein molecule and four tetramers of $d(pA)_4$ are shown in Figure 4(a) and (b) (see p. 315). The four tetramers will be referred to in the text below as A, B, C, D and the nucleotides within each tetramer as 1, 2, 3, 4 from 5′ to 3′; thus tetramer A is composed of 5′-A1-A2-

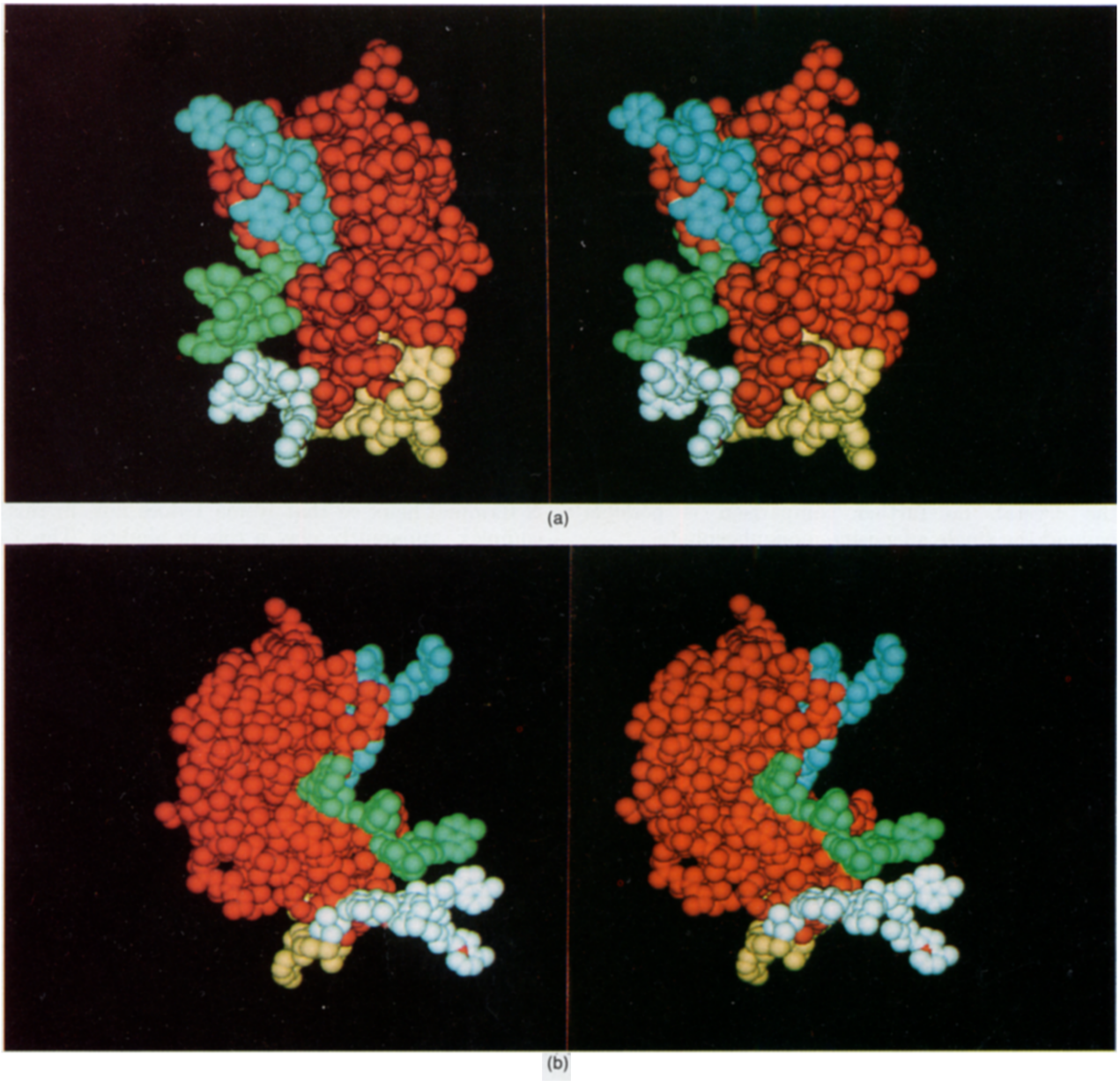

Figure 4. In stereo photographs (a) and (b), RNase is again seen associated with the 4 oligomers of d(pA)$_4$, but as van der Waals' sphere representations. With some exceptions, the nucleic acid follows the surface contours of the protein and is generally in close contact. The exceptional residues protrude into or cross solvent regions in the crystal to make contact with adjacent molecules in the lattice. The protein is shown in red and the 4 deoxyoligomers A, B, C, D are in this Figure colored yellow, white, green and blue, respectively.

A3-A4-3′, etc. The individual tetramers A, B, C and D are defined in Figures 5, 6, 7 and 8. The two tetramers seen most clearly in the initial, as well as subsequent, difference Fourier syntheses are those seen in Figures 5 and 8, with D the best defined.

DNA tetramer D is bound at the enzyme active site in a manner consistent with studies on protein–dinucleotide complexes in the crystal (Wlodawer, 1984; Wyckoff *et al.*, 1977; Pavlovsky *et al.*, 1978; Wodak *et al.*, 1977; Wlodawer *et al.*, 1983; Borkakoti, 1983; Borkakoti *et al.*, 1983). The 5′ phosphate group lies 4·1 Å (see Table 12) from the ε-amino group of lysine 41 and 3·5 Å from histidine 12, two of the amino acid residues implicated directly in catalysis. The position of the phosphate group at this stage of refinement appears to be the same as that found in a variety of other crystalline enzyme–ligand complexes (Richards & Wyckoff, 1971; Wlodawer, 1984). The 5′ adenosine residue, A1, occupies essentially the same locus as the adenosine moiety in previous protein–nucleotide and dinucleotide crystallographic studies, with the base in the *anti* conformation and stacked against the protein surface. We have not, at this time, made a quantitative comparison of the similarity between this nucleotide and those investigated in other substrate–complex studies. Because pseudo substrates of length greater than two nucleotides

Table 12
Possible salt bridges between phosphate groups and basic side-chains of amino acids (NZ or NH_2 nitrogen to closest oxygen of PO_4)

Amino acid	Phosphate group	Distance (Å)
Lysine 7	D2	4·76
Lysine 31	A1	3·87
Arginine 39	C1	4·85
Lysine 41	D1	4·19
Lysine 41	D2	5·2
Lysine 61†	C2	5·8
Lysine 66	C3	4·76
Lysine 66	C4	3·36
Arginine 85	C4	4·38
Lysine 91†	B4	3·80
Lysine 98	B1	4·11

† Intermolecular.

have not been observed bound to RNases A or S in the crystal, no further comparison is possible beyond the single adenosine monophosphate.

From the definition and intensity of the phosphate groups of tetramer D in difference Fourier maps and the modest values in Table 13 for the temperature factors of these groups, it seems that phosphate D1, D2 and D3 are immobilized by binding to the protein and that phosphate D4, at the 3′ end of the oligomer, is at least constrained. Thus, influence of the enzyme appears to extend three to four nucleotides away from the catalytic center (the 5′ phosphate) in the 5′ → 3′ direction. Nucleotides D2 and D3 are also packed against the enzyme surface. As may be seen in Figure 5, nucleotide D1 is in an extended conformation linking the 5′ and penultimate phosphate (D1 and D2). Subsequent nucleotides, D2, D3 and D4 are in less extended conformations, and all of the bases have similar plane normals, like a hand of cards. This conformation might be that adopted by a long single polynucleotide strand, since it permits a substantial degree of base stacking.

Oligomer D, as it extends from the catalytic center is also engaged by lysine 7 with its ε-amino group less than 5 Å from phosphate A2. The 3′ nucleotide D4 is found well away from the N-terminal helix so that lysine 1 does not, in this complex, apparently play a role in binding. The 3′ nucleotide of tetramer D appears to participate in interactions with $d(pA)_4$ oligomers in the unit cell associated with other asymmetric units in the crystal lattice and it makes contact also with another protein molecule in the lattice. So the

Table 13
Thermal parameters for $d(pA)_4$ constituent constrained groups

Tetramer	5′ PO_4	R1	B1	PO_4	R2	B2	PO_4	R3	B3	PO_4	R4	B4 3′
D	23†	86	80	10†	149	9	22	60	61	30	69	65
C	24†	90	67	25†	51	82	31†	89	104	41†	79	56
B	56†	106	99	33†	82	127	52	187	165	45†	40	74
A	37†	73	72	23†	222	132	43	70	107	42	65	92

† Probable electrostatic interaction with a lysine or an arginine residue.

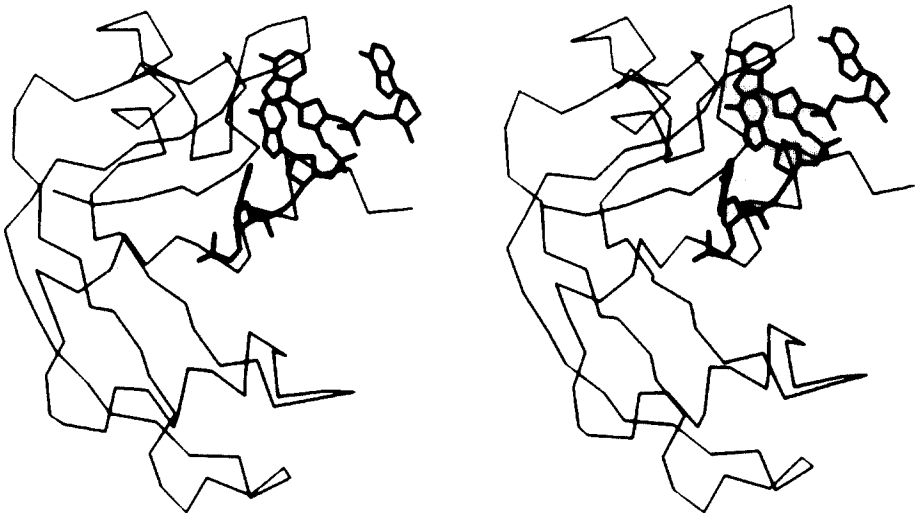

Figure 5. In this stereo diagram, RNase is seen associated with the $d(pA)_4$ oligomer referred to as D and the individual nucleotides running 5′ to 3′ as D1, D2, D3 and D4. The 5′-terminal phosphate of nucleotide D1 is fixed adjacent to histidine 12 and near histidine 119, which are catalytic residues, by a salt bridge to lysine 41 (and possibly lysine 66) at the catalytic center. The disposition of the 5′ nucleotide (D1) 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 complexes.

position of the 3′ nucleotide, D4, may be determined by lattice interactions that cause it to assume a location different from that which it would occupy if free in solution

Tetramers B and C are more intermolecular in that they bind chiefly to one RNase molecule but extend through solvent and make contact with another in the unit cell. From Table 13, the groups comprising tetramers B and C, particularly the phosphates, have generally higher thermal parameters than the equivalent groups in tetramers A and D. There is no simple way to discriminate between real thermal motion and statistical disorder about a point, but it seems probable that in the case of the nucleic acid oligomers described here, the thermal values reflect the statistical disorder of the nucleotides. If the negative charges on the phosphates and the positive charges on the protein are largely delocalized, as suggested by Record *et al.* (1976), then this might be expected.

Unlike tetramer D, d(pA)$_4$ oligomer C, which is seen in Figure 6, makes nearly all its contacts with the protein through the phosphate groups, while the bases project away from the polypeptide. The 3′ nucleotide, C4, is not in the pyrimidine binding site observed for dinucleotide–protein complexes such as CpA and UpA (see Wlodawer, 1984). This is consistent with the specificity of RNase A, which cleaves only after pyrimidine residues. Presumably, if a purine could occupy the pyrimidine site, then cleavage would occur after any nucleotide, whether it be pyrimidine or not. Borkakoti (1983) has, however, reported that purine bases can occupy the pyrimidine site under some circumstances; here it apparently does not. As with d(pA)$_4$ oligomer D, the bases of C2, C3 and C4 near the pyrimidine site are more or less parallel to one another and this permits stacking to take place. Whether this stacking occurs in solution or when RNase cleaves a long polynucleotide chain we cannot be certain but such a stacking would stabilize the protein–nucleic acid complex and compensate for the otherwise unobstructed exposure to solvent. The 2′ hydroxyl group of the 3′-terminal ribose, C4, of tetramer C is less than 4·0 Å from the 5′ phosphate, D1, at the catalytic center. This is, again, in spite of the fact that 3′ purine nucleotide, C4, does not occupy the pyrimidine binding site, but packs against the surface of the protein. The ε-amino group of lysine 66 is less than 4·0 Å from the most 3′ phosphate, C4. The possible salt bridge between lysine 66 and the 3′ phosphate, C4, could aid in fixing the DNA at the active site. It seems clear from the structure, refined at 2·5 Å resolution at least, that lysines 7, 41 and 66 are most crucial in the configuration of the nucleotides at the active site. We noted that arginine 85, in the refined structure, has also moved its side-chain so that the guanidinium group of arginine 85 is also close to the PO_4 group nearest the 3′ terminus, C4, as well as to the PO_4 of C3, and may be engaged in a salt bridge with either.

While nucleotides C4 and C3 are somewhat compressed and stacked, nucleotides 2 and 1 are more extended. Because of the relatively high thermal factors associated with several of the nucleotides in this tetramer, however, it is not clear how much significance should be attached to the precise backbone conformation seen in Figure 6. Phosphate C2 of this tetramer, situated primarily in a solvent cavity, probably makes an intermolecular

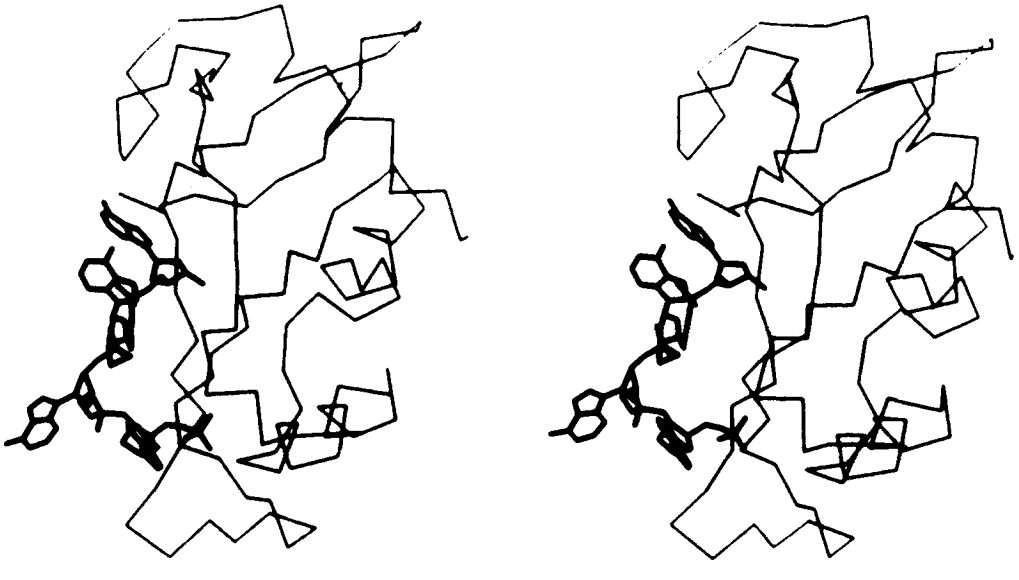

Figure 6. A second tetramer of d(pA)$_4$ is seen in this stereo diagram bound to RNase in the absence of all others. This oligomer is referred to in the text as tetramer C and the individual nucleotides running 5′ to 3′ as C1, C2, C3 and C4. The oligomer extends from the 3′-C4 nucleotide lying very close to, but not occupying, the pyrimidine binding site over the surface of the protein and into a solvent region. The 5′ phosphate group could form a salt bridge with lysine 1 of another molecule in the lattice, while phosphates of nucleotides C4, C3 and C2 apparently form ion pairs with lysine 66, arginine 85 and arginine 39, respectively. Phosphate C2 is also likely to form a salt bridge with lysine 61 of another protein molecule in the lattice.

Figure 7. RNase is seen in this stereo diagram associated with a third $d(pA)_4$ oligomer referred to in the text as tetramer B and the individual nucleotides running 5′ to 3′ as B1, B2, B3 and B4. This tetramer is secured at the 5′-terminal phosphate of nucleotide B1 by a salt bridge to lysine 98. It extends over the surface of the protein, possibly making an electrostatic interaction with arginine 39 at B2 and then it passes into a large interstitial solvent region. The 3′ nucleotide B4 is in contact with another protein molecule in the crystal lattice and the phosphate of B4 forms a salt bridge with lysine 91 of that molecule. It is the 5′-terminal phosphate of this tetramer B, bound by lysine 98, that appears to be associated as well with the 3′ bases of 2 other tetramers in the crystal.

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. This oligomer is referred to in the text as tetramer A and the individual nucleotides running 5′ to 3′ as A1, A2, A3 and A4. Tetramer A binds securely through its 5′-terminal and penultimate phosphate groups (A1 and A2) to the strongly electropositive anion binding site formed by the clustering of lysines 31, 37 and 91, 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, B1, of tetramer B, also bound to this same protein molecule. The 3′ nucleotide of a tetramer D, i.e. D4, but from a symmetry-equivalent $d(pA)_4$ oligomer in the lattice is also near the phosphate of B1, which creates an apparent DNA–DNA interaction involving 2, 3′-terminal bases (A4 and D4) and a 5′-terminal phosphate (B1).

ion pair with lysine 61 of another protein molecule in the lattice at a distance of 5·8 Å away. After extending over the protein surface, the 5′ phosphate group of tetramer C is found 4·8 Å from the guanidinium group of arginine 39.

The 5′-terminal phosphate of tetramer B seen in Figure 7 is found adjacent to lysine 98, where it is fixed by a salt bridge. Arginine 39, allowing for the mobility and reach of its side-chain, is within a distance consistent with an electrostatic linkage to phosphate B2. The thermal factor for the 5′-terminal phosphate is low, and indeed, its position was from the beginning always a major positive feature in the difference Fourier maps. Tetramer B extends from the protein surface through a solvent cavity and terminates with its 3′ nucleotide B4 against another protein molecule in the lattice. The phosphate of B4 at the 3′ end is 3·8 Å from the ε-amino group of lysine 91 of a different protein molecule in the crystal lattice.

The fourth d(pA)$_4$ tetramer, A, which is seen in Figure 8, was also clearly visualized in initial and subsequent difference Fourier maps and, as seen in Table 13, is also characterized by reasonably low thermal parameters. The 5′-terminal phosphate of nucleotide A1 is found on the opposite side of the protein from the active site and among the polypeptide strands formed by amino acid segments 26 to 39, 95 to 100 and 90 to 95, and near the disulfide bridge formed by Cys40 and Cys95. It is located at the center of a cluster of basic amino acid residues that includes Lys31, Lys91 and Arg33, where it is almost certainly fixed by electrostatic interactions. The distance between phosphate A1 and the ε-amino group of lysine 31, according to Table 12, is 3·87 Å. Tetramer A is in a more or less extended conformation as it emerges from the electropositive cavity (Matthew & Richards, 1982; Richards, 1982) beneath this domain of the protein. Phosphate A2 is contained in the same electropositive cluster as A1, but nucleotides A3 and A4 trace a path around the 90 to 95 β-loop of the protein toward the side of RNase bearing the active site cleft. All of the nucleotides in this oligomer are well-represented in difference Fourier maps and have reasonably low thermal parameters. The 3′ nucleoside, A4, of tetramer A is quite close to the 5′ phosphate of tetramer B, which is in turn bound by lysine 98. In addition, the 3′ nucleoside, D4, of a tetramer D that is bound by another protein molecule in the lattice is also close to both the 5′ phosphate of tetramer B and the 3′ nucleoside of tetramer A. Thus, a constellation of two 3′ adenosines and a 5′ phosphate is formed near the surface of one protein molecule, where it is anchored to the protein through the phosphate by a lysine interaction. The details of this arrangement are not entirely clear in the difference Fourier maps at this resolution and stage of refinement, but the observation seems to support an interaction involving the amino groups of two 3′ adenine bases (A4 and D4) with the doubly negative 5′-terminal phosphate group, B1, of tetramer B. Whatever the nature of the interaction, however, it seems that the convergence of the three DNA fragments at this point is likely to represent another of the crosslinks responsible for the formation of the crystal and the nucleic acid network it contains.

If the d(pA)$_4$ oligomer bound in the active site cleft and the oligomer that emerges from the pyrimidine binding site are considered alone as a virtually continuous strand of eight nucleotides, they describe an arc. The remaining two oligomers form a similar kind of arc, though not quite so well-defined as in the first case. These arcs, resembling helical turns, may reflect the conformational affinities of nucleic acid strands, or they may imply some mechanistic feature of the protein–nucleic acid complex as it forms under physiological conditions. The four oligomers of d(pA)$_4$ in the absence of protein are seen in Figure 9.

Figure 10 (see p. 320) shows the disposition of the d(pA)$_4$ tetramers alone in the unit cell. The individual DNA oligomers are not isolated from one another nor is their arrangement solely a consequence of the packing of the protein molecules. The DNA segments form a crosslinked and virtually contiguous network of strands and interactions that display conformational characteristics that might be ascribed to extended single-stranded nucleic acids. Multiple segments of d(pA)$_4$ align themselves more or less end-to-end into smooth helical arcs, and there are instances of successively stacked bases. This network of DNA was one of the more unexpected features to come from the crystal packing analysis. Figure 11 shows how an individual RNase molecule is embedded in the DNA network within the crystal. Apparent here are the extensive contacts that exist between

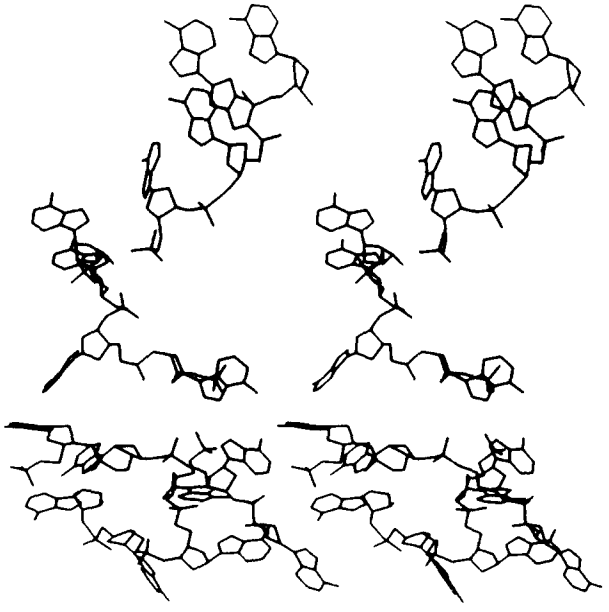

Figure 9. Stereo diagram of the 4 oligomers of d(pA)$_4$, with tetramer A at the bottom and D at the top, showing their relative disposition in space in the absence of the protein molecule.

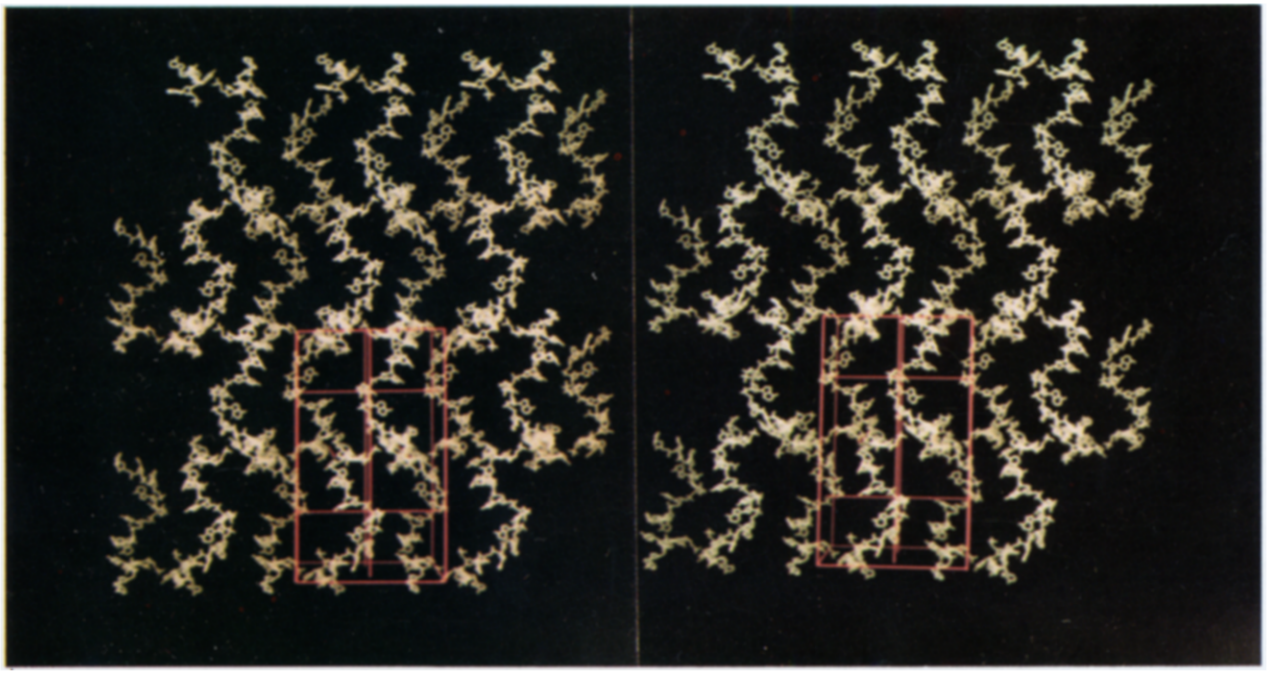

Figure 10. The contents of 6 unit cells of the crystal of RNase–$d(pA)_4$ with the protein molecules omitted. The individual segments of DNA can be seen to form a network of virtually continuous strands of nucleic acid that run through the crystal and appear to display some degree of helical character. The protein molecules and the DNA strands crosslink one another in the crystal primarily through electrostatic interactions involving the negatively charged phosphate groups and positively charged lysine amino and argine guanidinium groups on the protein. The outline of a single unit cell of the crystal is shown in red at the bottom, and the view is along the crystallographic *c* axis.

any one protein molecule and the strands of the DNA net.

In addition to the four tetramers described above, several isolated peaks occurred in the difference electron density maps. These peaks were occasionally close to tetramers A, B, C and D and could reflect alternate orientations for some portions of these oligomers. On the other hand, the solvent regions of the crystal may contain disordered DNA fragments. These fragments might make non-specific electrostatic, or for that matter hydrophobic, interactions with the protein by employing a phosphate or nucleotide anywhere along the DNA oligomer.

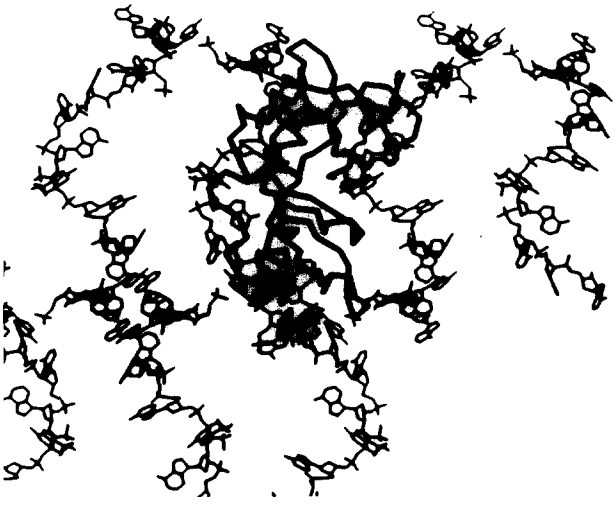

Figure 11. In this drawing, a single RNase molecule, represented only by its α-carbon backbone, is inserted into the 3-dimensional network of DNA strands as found in the protein–nucleic acid complex crystal. Apparent here are the extensive contacts between a single protein molecule and the DNA both at and distant from the active site of the enzyme.

(b) *Protein–nucleic acid interface*

An analysis of the crystal structure of RNase plus $d(pA)_4$ was carried out with the objective of delineating the amino acid environment of each nucleotide and thereby identifying the interactions responsible for formation and maintenance of the complex. Table 14 is a summary of all interactions between protein and nucleic acid that are less than 3·2 Å and which defines the interfaces between the two types of molecules. One must be cautious in accepting individual values too literally at this point, however, since the resolution of the refinement is only 2·5 Å.

Examination of the structure clearly shows the importance of salt bridges between protein and $d(pA)_4$ tetramers. Table 12 contains the distances from ε-amino groups of lysines and guanidinium groups of arginines to DNA phosphates. No attempt was made to intervene in the refinement process to accentuate these interactions and they are otherwise unbiased by energy or binding considerations. In addition to salt bridges involving lysine and arginine groups, we sought to determine if direct hydrogen bonds between phosphate

Table 14

DNA to amino acid distances less than 3·2 Å

Group	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	D1	D2	D3	D4
PO_4	Lys31 Pro93 Cys95	 Asn94			Thr87 Gly88 Lys98	Gly60	†Glu60 †Tyr76	Asp53† Val54† Gln55† Val57† Gln60†	None				Lys41 Val43 Lys66 His119	His119		
Ribose	Thr36 Asn94 Cys95 Ala96	Thr70† Ser89 Ala96		Lys104†	Gly88 Glu111†		†Gln60 †Met79	Asp53† Val54† Gln55† Ala56† Val57†		Thr78†	Lys66	Lys66 Asp121	His119			
Base	Lys31 Pro93 Cys95	 Asn27 Cys95 Ala96 Tyr97 Lys98	Gln69†	Thr100			†Val57 †Ser75 †Met79 †Ile106	Met13† Val47† Glu49† Asp53† Val54†		†Thr78 †Asn103	Lys66† Thr78 Met79 Ser80 Ile81 Ala102† Asn103†	Ser18† Glu49† Lys66 Ser80† Asp121	Asn67 His119	Asn71 Ala109 Val118 His119	Gln69 Thr87† Ser89	Asn24† Asn27†

† Intermolecular contact.

oxygens and amino acid side-chains were present. Asparagine 94 appears to be associated with a phosphate (which could be either A1 or A2 at the active site) and asparagine 113 may form an intermolecular bond with phosphate C2. Aside from these two, however, it is not clear that any other amino acid side-chains are hydrogen-bonded to phosphate oxygens unless through a water intermediary.

(c) *The protein structure*

The structure of the protein has not, at this resolution, undergone any substantial conformational change as a consequence of complex formation, although some side-chains have necessarily adjusted to the presence of the $d(pA)_4$ tetramers. Several of the more flexible loops of protein also appear to have modified their positions. The histogram of Figure 12 shows the shift in α-carbon co-ordinates for all amino acids. Comparable histograms of side-chain atoms alone and main-chain atoms alone were not appreciably different, so that Figure 12 is representative of overall changes. It should be noted that clusters of fairly large changes are present in the histogram, implying concerted movements by individual segments. As one might expect the N-terminal four amino acids that are in extended conformation fall into this category, as do two small loops 61 to 71 and 90 to 96, which were observed to be flexible in the native structure. The tip of the β-loop formed by residues 111 to 115 also shows significant movement, as do the carboxyl-terminal amino acids.

The greatest change in this protein model, when compared with the native enzyme, begins at serine 16, is centered around residue 21 and extends over the entire helix 26 to 35 and the returning β-strand 40 to 48. The major alteration in position, however, is clearly the flexible sweeping loop of extended chain that includes at its center the bond between serine 21 and 22 where subtilisin cleaves RNase A to make RNase S (Richards & Vithayathil, 1959). This segment of chain was found to be flexible and poorly defined in the native structure (Richards & Wyckoff, 1971; Kartha *et al.*, 1967; Carlisle *et al.*, 1974).

The apparent discontinuity at residue 21 and broad changes in the structure following raised suspicions that the protein might have been inadvertently cleaved by contaminant or invasive proteases during the crystallization procedure. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis of washed and redissolved crystals, along with RNase A and RNase S standards convinced us, however, that the crystals did contain uncleaved RNase A. A more likely explanation for this movement is that lattice interactions are responsible. As seen in Figure 2, the flexible loop containing serine 21 is involved in one of the few clear protein–protein contacts in the crystal. We suspect, therefore, that the movement here can be attributed to that source.

Figure 13 is a histogram of the average thermal factor for each amino acid in the structure. In the refinement procedure we utilized (CORELS) a separate *B* factor was applied for the main-chain atoms as a group and for the side-chain atoms as a group. The refinement procedure has a fundamental weakness in that no continuity is imposed on the contiguous groups along the chain, thus leading to some sudden and unlikely discontinuities and, we believe, also anomalous values for some side-chain *B* factors. Nevertheless, the histogram in Figure 13 has a number of interesting features that are likely to be trustworthy.

Terminal amino acids 1 to 4 have high values consistent with a lack of restraint on their positions. Residues 21 to 44, also as indicated by Figure 12,

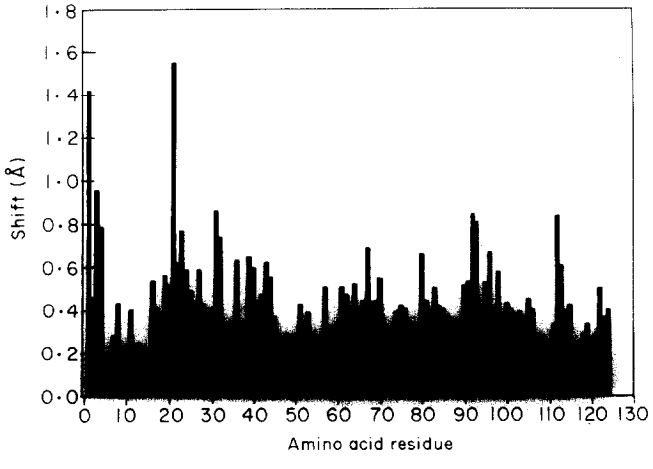

Figure 12. A histogram of the shift in *x*, *y*, *z* position of the α-carbon of each amino acid in the RNase molecule in the refined $d(pA)_4$ complex from its position in the refined native structure (Wlodawer & Sjolin, 1983). The 2 structures were fitted to one another by a least-squares procedure (Rossmann & Argos, 1975) that minimized the sum of the distances squared.

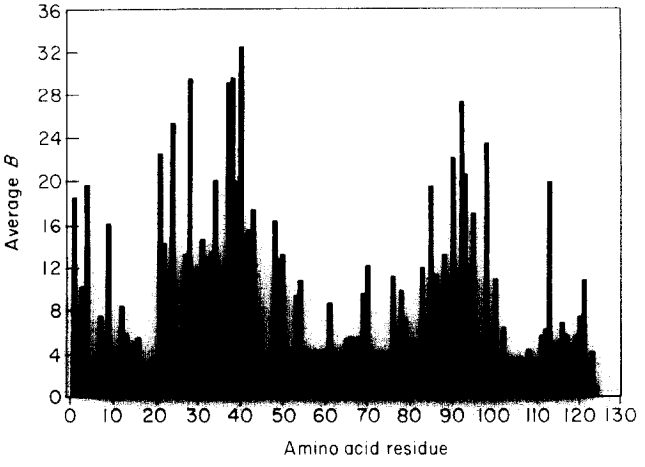

Figure 13. A histogram of the thermal factor *B* for the 124 amino acids of ribonuclease A in the $d(pA)_4$ complex where the assigned *B* value is the average of the values for the main-chain and side-chain atoms of each residue.

appear to be quite mobile in the crystal and contain the highest B values in the protein portion of the structure. However, threonine 45, phenylalanine 46 and valine 47 at the active site, have among the lowest. The small loop 61 to 76, although seen to move substantially in Figure 12, has modest B values, suggesting that this loop may be stabilized or rigidified by the nucleic acid. A region of substantial motion is centered on proline 93 in a fairly flexible small loop that was observed to be mobile in the native structure as well. Asparagine 113 is observed to have an unusually high B factor. It is at the tip of a β-loop and is probably disordered.

4. Discussion

(a) *Protein–DNA bonding interactions*

The structure of the crystalline complex between RNase A and d(pA)$_4$ described here uses both electrostatic interactions between phosphates and basic amino acids as well as interactions between the protein and the constituent nucleosides to produce and maintain the integrity of the complex. Over the extended path outlined by the d(pA)$_4$ oligomers, the electrostatic interactions clearly predominate. It is only at the active center that the latter, presumably more specific, kinds of associations are in evidence. Record *et al.* (1976) conducted titrations of RNase with cations and showed the importance of electrostatic interactions with nucleic acid. They found their data to be consistent with seven ion pairs when single-stranded DNA or RNA is bound to RNase A. Karpel *et al.* (1981), using oligonucleotides devoid of heterocyclic bases, also provided evidence that electrostatic interactions predominate in complex formation with base contacts adding only slightly to stability.

Jensen & von Hippel (1976) showed that for double-stranded DNA complex formation is accompanied by about four ion pairs and that the protein can cover or protect about four nucleotide base-pairs. When bound to single-stranded DNA, by an apparently different mechanism, RNase can cover more than 11 nucleotides. Jensen & von Hippel (1976) showed that when RNase binds to single-stranded DNA it does so with an affinity that is roughly two orders of magnitude higher than for native DNA, and that RNase A binding to single-stranded DNA does not appreciably perturb either the conformation of the sugar–phosphate backbone or the relative orientations of vicinal nucleotide bases. RNase does not apparently alter the stacking of bases along the single chain, except for the bases that actually articulate with the active site.

In accordance with the above data, we observed in this crystalline complex seven lysine and two arginine residues, and possibly one more of each that are sufficiently near phosphate groups on one of the four oligomers to assume that a salt bridge, or ion pair, is formed. Lysines 7 and 41 bind tetramer D, lysines 66 and 61 and arginines 39 and 85 bind to tetramer C, lysines 98 and 91 and possibly arginine 39 bind to tetramer B, and lysine 31 and arginine 33 bind to tetramer A. There are enough electrostatic interactions, therefore, to explain the solution data regarding ion pair formation. From the Figure it is evident that the bases of the DNA are frequently arranged in parallel arrays, though not necessarily stacked in the manner found in native DNA. In general, the bases, again consistent with the data of Jensen & von Hippel (1976), Karpel *et al.* (1981), Record *et al.* (1976) and von Hippel *et al.* (1977), protrude away from the protein surface and do not contribute significantly to the binding. Exceptions to this are bases directly at the active site of the enzyme. In this complex there is direct interaction of at least two adenosine nucleosides (nucleotides 1 and 2 of tetramer D) and possibly a third (D3) with the protein surface. The pyrimidine binding site is unoccupied by a base though the 3′ nucleoside of tetramer C is crowded against the binding pocket. There is a vast amount of crystallographic difference Fourier results, however, to show that the pyrimidine binding site would be filled were a pyrimidine base rather than only adenosine available in the complex (see Wlodawer, 1984). Thus, we can surmise that, for an extended and continuous polynucleotide, there are at least three nucleosides that are bound by RNase A and that involve ribose and base interactions.

In the complex the ion-pair interactions are arrayed more or less sequentially along the course followed by oligomers, with each tetramer of d(pA)$_4$ generally making at least two salt bridges. The predominance of lysines here is in keeping with the contention of Record *et al.* (1976) that the charge interactions are likely to be delocalized.

Matthew & Richards (1982) and Richards (1982) calculated the electric field potential in a three-dimensional volume about a single RNase molecule. They found that at low ionic strength there are five locations that have a significantly positive potential and that would probably serve as anion binding sites. One of these sites is much larger than any of the others and persists even at higher ionic strengths. This site lies on the opposite side of the molecule, in the cluster of basic residues lysine 31, lysine 37, arginine 33 and lysine 91. The phosphates of nucleotides 1 and 2 of tetramer A fall within this network of positive charges. Other electropositive sites discovered by Matthew & Richards (1982) are also occupied in the complex RNase with d(pA)$_4$ by some phosphate of an oligomer strand.

(b) *The active site cleft*

Difference Fourier analyses have been conducted of various crystalline complexes of RNase A and RNase S complexed with mononucleotides and dinucleotides, and these have been recently reviewed by Wlodawer (1984). There is general agreement on the positions of the 5′ pyrimidine, the phosphate group at the active center, and the purine that occupies the nucleotide binding site on the 3′ side of the catalytic center, frequently

referred to as the P, PO_4 and A sites, respectively. Evidence from these studies, and that provided by other techniques applied to the enzyme in solution (see Richards & Wyckoff, 1971; Blackburn & Moore, 1982), indicate that the phosphate is fixed by lysine 41 and possibly lysine 66 while catalysis is mediated by histidines 12 and 119. The studies further show that a dinucleotide occupying the P and A sites tends to be in a fully extended conformation with both bases in close contact with the protein surface.

In the RNase–$d(pA)_4$ complex we describe here, the pyrimidine site is free, though obstructed by the 3′ nucleotide, C4, of tetramer C, which is forced against the P site as if trying to enter. The 5′ phosphate and 5′ adenosine of tetramer I (nucleotide 1A) is, however, in the phosphate and A site, as would be expected. The phosphate group is quite clear in difference Fourier maps and is 4·1 Å from the ε-amino group of lysine 41, which moved during refinement only slightly from its native position. That there is a salt bridge between the amino group of lysine 41 and the 5′-terminal phosphate of tetramer A seems certain. The closest amino acid side-chain to the phosphate is histidine 12 which is 3·5 Å away and also has moved very little from its native position. Histidine 119 in the refined complex is not particularly close to the phosphate at the active center. It does not extend toward that phosphate in this structure but is turned away and is obstructed by the 5′-terminal ribose group. If anything, histidine 119 is here closer to the penultimate phosphate between nucleosides D1 and D2 than the 5′-terminal phosphate. Consistent with previous reports (Borkakoti *et al.*, 1982), description of histidine 119 was not unambiguous. In difference electron density maps there were strong positive density peaks in the immediate vicinity of the histidine 119 side-chain and some were compatible with alternative positions for the imidizole group that could place it closer to the 5′ phosphate. Recalling that the level of resolution of this refinement is 2·5 Å and that the area immediately surrounding histidine 119 is the center of numerous difference electron density changes arising from the intrusion of a $d(pA)_4$ oligomer, it may be best to defer conclusions regarding the disposition of histidine 119 until a higher level of resolution and refinement is achieved.

(c) *Implications for the RNase–RNA complex*

Although the specificity of RNase A is for RNA with respect to enzymic activity, it is known that it binds to both double and single-stranded DNA with comparable affinity. Walz (1971) demonstrated that RNase has the same, if not slightly higher, affinity for deoxyribonucleotides as for ribonucleotides. Sekine *et al.* (1969) showed that denatured DNA competitively inhibits the hydrolysis of substrate analogues and RNA itself. Difference Fourier experiments have shown directly the binding of deoxynucleotides at the active site of the enzyme (see Wlodawer, 1984) and binding constants have been determined for some DNA components (Richards & Wyckoff, 1971). In addition, extensive experiments by Jensen & von Hippel (1976) and von Hippel *et al.* (1977) confirmed and delineated the binding characteristics of ribonuclease for both double and single-stranded DNA. All the available evidence tends to support the assertion that single-stranded DNA binds to RNase by the same mechanism as RNA, and that it occupies the active site of the enzyme in an essentially identical manner. Thus, it is fair to assume that the RNase–$d(pA)_4$ complex reported here is similar, particularly at the catalytic center, to the complex with RNA.

The crystalline complex between RNase and $d(pA)_4$ is clearly not an exact model of the complex formed between the protein and single-stranded DNA or RNA. The nucleic acid strand is segmented and at points discontinuous, the pyrimidine binding site is not properly utilized, and several interactions occur between different protein molecules in the unit cell but involve a single $d(pA)_4$ oligomer. Nevertheless, there is cause to believe that the RNase–$d(pA)_4$ complex does reflect the complex formed with single-stranded nucleic acid in solution. If those interactions, electrostatic and otherwise, that form between the $d(pA)_4$ tetramers and only a single RNase molecule are considered, and if the intermolecular (lattice) interactions are ignored, then the series of nucleotides trace out a nearly continuous path over the enzyme surface. The $d(pA)_4$ tetramers A and D plus three nucleotides of C and two from B form a chain, seen in Figure 14 (see p. 325), that runs from the 5′ PO_4 of tetramer A, fixed by the electropositive anion trap on the back side of the protein, around and over the surface of the protein (B and C) and finally into and through the active site cleft of the enzyme to terminate at the 3′ hydroxyl of tetramer D.

This path utilizes 12 nucleotides and all of the non-intermolecular interactions between RNase and $d(pA)_4$ observed in the crystals. The distances between phosphate groups are stereochemically reasonable and no distortions in conformation are necessary. The course of the nucleotides is 5′ to 3′ over the entire path, so no reversal in polarity would be required for any oligomer. In no case was significant translation applied to any nucleotide nor to any tetramer in order to create the chain, seen in Figure 14, from the complex of four $d(pA)_4$ oligomers. Only rotations and minor adjustments to conformations were required.

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 the polynucleotide chain, or in this case the segments of

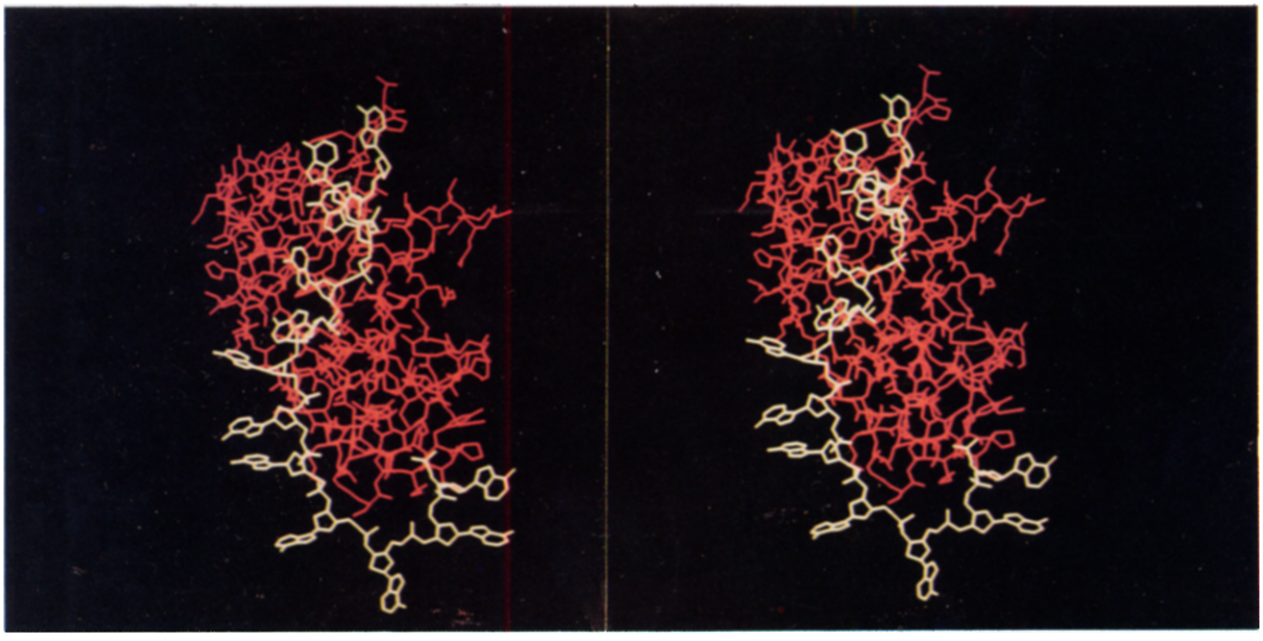

Figure 14. The binding interactions between the DNA oligomers and the protein trace out a virtually continuous path from the 5′-terminal phosphate, A1, of tetramer A to the 3′-terminal nucleotide (D4) of tetramer D. The path is particularly evident if those nucleotides are ignored that are responsible for intermolecular interactions (B2, B3, B4 and C1) and only those that are reasonably associated with a unique protein molecule (A1–A4, B1, C2–C4, D1–D4) are considered. If this is done then the oligomers form a more or less continuous strand of nucleic acid (shown here in yellow) that runs from the electropositive anion trap on the underside of the protein (shown in red), over the surface of the enzyme, into and finally through the active site cleft. The chain length here is 12 nucleotides from 5′ to 3′ (A1 to D4).

that chain. Thus, one function of the structure of RNase 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 many are fortuitous and those that would be found under more physiological conditions would not be present at higher salt concentrations. As Record *et al.* (1976) have argued, however, it is the mixing entropy of bound ions released and displaced by the structurally constrained nucleic acid that drives the formation of the RNase–DNA complex. The free-energy change derived from the Coulombic interactions are not of primary importance. Record's interpretation of the data of Jensen & von Hippel (1976) and von Hippel *et al.* (1977) 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.

(d) *Implications for crystal growth*

The structure of the RNase–d(pA)$_4$ crystals has some significance, we believe, for the growth of macromolecular crystals in general. There are rather few protein–protein contacts in the crystal and these do not appear sufficient to explain the rapid growth, high degree of order and large size of the crystals grown for this analysis. The cohesive forces must, therefore, be supplemented in part at least by the d(pA)$_4$ oligomers. This is suggested as well by the network of DNA strands, seen in Figure 10, that are present independent of the protein. At neutral pH, RNase A carries three to four excess positive charges, but it contains a total of 13 basic amino acid residues (4 arginines plus 8 lysines plus the N-terminal amino group). Thus, the protein is a large cation with the positive charge distributed selectively over its surface. The d(pA)$_4$ oligomers on the other hand each bear a possible five negative charges at neutral pH. When the ionic strength of the mother liquor is maintained low, as it was in this study, it is inevitable that association will occur between the two charged species, the protein and the d(pA)$_4$. In agreement with this expectation, we find that the majority of chemical interactions involved are indeed ion pairs between basic amino acid side-chains and phosphate groups. Hydrophobic and hydrogen bonding interactions appear to be restricted almost entirely to the residues bound at the catalytic center.

Within the crystal, the protein molecules are essentially congealed into the symmetrical unit cell aggregate by the d(pA)$_4$ oligomers, which serve as electrostatic crosslinking agents. One might see the general value of such non-covalent crosslinkers in

the growth of macromolecular crystals. Crystallization proceeds by the trial and error addition of molecules associating in various ways with the growing aggregate, readjusting all the while until the optimal set of bonds is made and the maximum decrease in energy of the system achieved. Because the linking agents, the $d(pA)_4$, are not covalent bond-forming molecules, but form only electrostatic bridges that are reversibly made and broken, the association of molecules is significantly enhanced without sacrificing the freedom to reorient, to readjust and to approach the lowest energy minimum.

This mechanism of crystal formation may explain why other, probably disordered molecules of $d(pA)_4$, may be present in the solvent regions of these crystals and why the occupancies and order of the $d(pA)_4$ molecules actually seen may be somewhat less than full. These probably represent some of the intermediate states of association that have not yet, or cannot, find a satisfactory mode of binding but have been frozen into the crystal. At the same time, the crystals we describe here suggest a mechanism for protein crystal growth that might have a general application. That is, for protein molecules of predominantly negative or positive character, i.e. having p*I* values significantly different from neutrality, an appropriate electrostatic linking agent may be found. For example, molecules such as polyamines of varying chain length might be useful for acidic proteins. Similarly, polysulfates or other polyanions of fixed and specific lengths might be useful for the crystallization of more basic proteins. When using such precipitants as polyethylene glycol the interactions would be enhanced by maintaining low ionic strength.

This research was supported by a grant (GM-27838) from the United States Public Health Service. G.B. was a postdoctoral fellow of the Medical Research Council of Canada and gratefully acknowledges their support. The authors 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 programming assistance. We also acknowledge the help of Gregory DeLozier and Marie Greene for their assistance with problems in computer graphics; Dr Alex Wlodawer for his helpful comments and Drs Joel Sussman and Steve Holbrook for their advice regarding the use of the refinement program. Finally, we thank Doug Schiff and the members of the Graphics Group of the Computer Science Department at the University of North Carolina for their unselfish assistance and advice and their exceptional facilities.

References

Alberts, B. M., Amodio, F. J., Jenkins, M., Gutmann, E. D. & Ferris, F. L. (1968), *Cold Spring Harbor Symp. Quant. Biol.* **33**, 289–305.

Alberts, B., Frey, L. & Delius, H. (1972). *J. Mol. Biol.* **68**, 139–152.

Anfinsen, C. B. & White, F. H. Jr (1961). *The Enzymes* (Boyer, P. D., ed.), 2nd edit., vol. 5, Academic Press, New York.

Blackburn, P. & Moore, S. (1982). *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 15, Academic Press, New York.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Crystallogr.* **12**, 794.

Blundell, T. L. & Johnson, L. N. (1976). In *Protein Crystallography*, Academic Press, New York.

Borkakoti, N. (1983). *FEBS Letters*, **162**, 367–373.

Borkakoti, N., Moss, D. A. & Palmer, R. A. (1982). *Acta Crystallogr. sect. B*, **38**, 2210–2217.

Borkakoti, N., Palmer, R. A., Maneef, I. & Moss, D. S. (1983). *J. Mol. Biol.* **169**, 743–756.

Brayer, G. D. & McPherson, A. (1981). *J. Biol. Chem.* **257**, 3359–3361.

Carlisle, C. H., Palmer, R. A., Mazumdar, S. K., Gorinsky, B. A. & Yeates, D. G. R. (1974). *J. Mol. Biol.* **85**, 1–18.

Felsenfeld, G., Sandeen, G. & von Hippel, P. H. (1963), *Proc. Nat. Acad. Sci., U.S.A.* **50**, 644–651.

Jensen, D. E. & von Hippel, P. H. (1976). *J. Biol. Chem.* **251**, 7198–7214.

Jones, T. A. (1982). In *Computational Crystallography* (Sayer, D., ed.), pp. 307–317, Oxford University Press, New York.

Karpel, R. L., Yrttimaa, V. A. & Patel, G. L. (1981). *Biochem. Biophys. Res. Commun.* **100**, 760–768.

Kartha, G., Bello, J. & Harker, D. (1967). *Nature (London)*, **213**, 862–865.

Liang, C.-J., Yamashita, K. & Kobata, A. (1980). *J. Biochem. (Tokyo)*, **85**, 51–58.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Matthew, J. B. & Richards, F. M. (1982). *Biochemistry*, **21**, 4989–4999.

McPherson, A. (1982). *The Preparation and Analysis of Protein Crystals*, John Wiley and Sons, New York.

Pavlovsky, A. G., Borisova, S. N., Broisov, V. V., Antonov, I. V. & Karpeisky, M. Y. (1978). *FEBS Letters*, **92**, 258–262.

Plummer, T. H., Tarentino, A. & Maley, F. (1968). *J. Biol. Chem.* **243**, 5158–5164.

Record, M. T., Lohman, T. M. & de Haseth, P. (1976). *J. Mol. Biol.* **107**, 145–158.

Richards, F. M. (1982). *Brookhaven Symp. Biol.* no. 32.

Richards, F. M. & Vithayathil, P. J. (1959). *J. Biol. Chem.* **234**, 1459–1465.

Richards, F. M. & Wyckoff, H. W. (1971). *The Enzymes* (Boyer, P. D., ed.), 3rd edit., vol. 4, pp. 709, Academic Press, New York.

Rossmann, M. (1960). *Acta Crystallogr.* **13**, 221–226.

Rossmann, M. G. & Argos, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.

Rubin, B. & Richardson, D. (1972). *Biopolymers*, **11**, 2381.

Sekine, H., Nakano, E. & Sakaguchi, K. (1969). *Biochim. Biophys. Acta*, **174**, 202–210.

Sierakowska, H. & Shugar, D. (1977). *Progr. Nucl. Acid Res. Mol. Biol.* **20**, 59.

Smyth, D. G., Stein, W. H. & Moore, S. (1963). *J. Biol. Chem.* **283**, 227.

Sussman, J. L. (1983). *J. Appl. Crystallogr.* **16**, 144–150.

Sussman, J. L. (1985). *Methods in Enzymology*, vol. 115, *Diffraction Methods for Biological Macromolecules* (Wyckoff, H. W., Hirs, C. H. W. & Timasheff, eds), pp. 271–303, Academic Press, N.Y.

TenEyck, L., Weaver, L. H. & Matthews, B. W. (1976). *Acta Crystallogr. sect. A*, **32**, 349.

Terwilliger, T. & Eisenberg, D. (1984). *Acta Crystallogr. sect. A*, **39**, 813–816.

Thiessen, W. E. & Levy, H. A. (1973). *J. Appl. Crystallogr.* **6**, 309–346.

Usher, D. A., Erenrich, E. S. & Eckstein, F. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 115–118.

von Hippel, P. H., Jensen, D. E., Kelly, R. C. & McGhee, J. D. (1977). In *Nucleic Acid–Protein Recognition* (Vogel, H. J., ed.), pp. 66–90, Academic Press, New York.

Walz, F. G. (1971). *Biochemistry*, **10**, 2156–2162.

Wlodawer, A. (1984). In *Biological Macromolecules and Assemblies, Nucleic Acids and Interactive Proteins* (Jurnak, F. A. & McPherson, A., eds), pp. 393–440, John Wiley Co., New York.

Wlodawer, A. & Sjolin, L. (1983). *Biochemistry*, **22**, 2720–2728.

Wlodawer, A., Miller, M. & Sjolin, L. (1983). *Proc. Nat. Acad. Sci., U. S. A.* **80**, 3628–3631.

Wodak, S., Ieu, M. Y. & Wyckoff, H. W. (1977). *J. Mol. Biol.* **116**, 855–875.

Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Tsernoglou, D., Johnson, L. N. & Richards, F. M. (1967). *J. Biol. Chem.* **242**, 3984.

Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970). *J. Biol. Chem.* **245**, 305.

Wyckoff, H. W., Carlson, W. & Wodak, S. (1977). In *Nucleic Acid–Protein Recognition* (Vogel, H. J., ed.), pp. 569–580, Academic Press, New York.

Edited by R. Huber