X-RAY CRYSTALLOGRAPHIC STUDIES OF DNA-DRUG AND DNA-PROTEIN INTERACTIONS

Thesis by

Philip Pjura

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1987

(Submitted November 10, 1986)

Acknowledgements

There are always more people to thank than can ever be thanked enough. First, I would like to thank my advisor, Professor Richard E. Dickerson, for his continuous support and encouragement during my stay in his group, and for patiently putting up with more from me than he ever had need to; and Mary Lou Kopka, my "co-advisor," for her advice, friendship, and moral support. Also, I would like to thank the rest of my friends in the group for their support and friendship, in particular, Horace R. Drew III and Ben N. Conner, my guides into the world of crystallography during the early years. I would like to thank my professional collaborators with whom I worked on many projects during my stay, in particular, Professor Richard M. Wing, who led me through my first crystal structure; Professor Doug Rees, for his patient advice and help with all things crystallographic and not; and Dr. Udo Heinemann, with whom I built my first diffractometer, and who showed me what was expected of me when I became a postdoc. I would like to thank my many friends at Caltech, especially Ed Schlesinger, Jerry Siu, and the whole Marks House gang, for making me feel like part of a community during my first year at Caltech, and for helping me through the "exile years" in Westwood. And finally, I would like to thank my parents, who still haven't the slightest idea what it is I do for a living.

I would like to thank the National Institute of Health and the National Science Foundation for their financial support.

ABSTRACT

I. Cisplatin-12mer

Cisplatin (cis-diamminodichloroplatinum[II]) is a widely used antineoplastic agent, which is believed to work by means of covalent interaction with DNA. Complexes of this compound were made with the B-DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G by diffusion of the drug into pregrown DNA crystals, and a structure determined to 2.6 Angstrom resolution by molecular replacement. Cisplatin was found to bind with partial occupancy at three discrete sites: G16 (61%), G4 (30%) and G10 (22%), in each case by means of a single covalent bond from the metal to guanine N7 in the major groove. The square plane of the metal complex ligands is rotated out of the plane of the guanine base, with one of the ligands that is cis to the guanine N7, presumably an amine, in a position to make a hydrogen bond with guanine O6; the long metal-O6 distance precludes the possiblilty of a direct metal-O6 bond. The DNA structure itself is essentially undisturbed by the metal binding; the only change is a slight motion of the bound guanines outward into the major groove toward the metals, resulting in a slight opening up of the groove but without pulling the base-pairs out of the helix stack. The structure shows that it is not possible for a direct N7-to-N7 crosslink between two adjacent bases by the metal to exist in an intact B-DNA double-helix. It is suggested that the observed structure is a primary mode of binding of the drug, which then could become this postulated active form upon disruption of the DNA duplex.

II. Lac headpiece-operator complex

The lac repressor protein of E. coli controls expression of the genes necessary for lactose utilization by the organism. It is a tetramer of four identical subunits of 355 amino acids each, each of which is divided into a 51 amino acid N-terminal DNA-binding, or "headpiece" domain and a 300 residue C-terminal regulatory region, or "core." In the presence of 1M Tris.HCl pH 7.5 and 30% glycerol, the headpiece can be isolated intact by proteolytic digestion and is believed to retain its specific DNA binding properties for the lac operator site. A large-scale purification scheme for the headpiece protein has been developed, using a specific protease-affinity column to eliminate all residual proteolytic activity from the prep, thus making it possible to isolate large quantities of the protease-sensitive fragment in stable form for crystallization trials. Attempts to crystallize the protein by itself resulted only in fibrous microcrystals. But cocrystallization trials with a 21 base-pair lac operator DNA oligomer yielded what appear to be small cocrystals, too small to characterize, under conditions in which neither the protein nor the DNA by themselves would crystallize.

III. Hoechst 33258-Complex with DNA 12mer

Hoechst 33258 is a widely used histological stain that forms a fluorescent complex with DNA, showing strong preference for AT-rich regions. Crystals were grown of a 1:1 complex of the dye with the B-DNA 12mer C-G-C-G-A-A-T-T-C-G-C-G, and a structure was determined to 2.2 Angstrom resolution. The compound was found to bind noncovalently

in the minor groove, with its aromatic phenol and two benzimidazole rings spanning three bases of the A-A-T-T region, and its aliphatic piperazine ring binding in the adjacent C-G-C-G region. Three hydrogen bond contacts were found to bridge between adjacent base-pairs, the bridging resembling that by water molecules in the native structure. This bridging of base pairs was achieved via three amine groups on the dye. These all undoubtedly help to stabilize the interaction, but the shape of the compound allows a good hydrogen bonding interaction of normal length through only one of these contacts. The piperazine ring, because of its orientation perpendicular to the other rings, is unable to bind into the narrower minor groove of the AT region and must bind in the wider CG minor groove. In a second conformation of the dye, also seen in the structure, the piperazine ring points out of the minor groove because of a rotation of the benzimidazole ring to which it is attached around its bond to the other benzimidazole. In this orientation, the piperazine ring no longer makes contact with the DNA, and therefore the drug is not restricted to binding at a site containing GC base-pairs. As a result, it could bind within a region of contiguous AT base-pairs without the problem of steric clash due to the piperazine ring.

Table of Contents

X-RAY CRYSTALLOGRAPHIC STUDIES OF DNA-DRUG AND DNA-PROTEIN INTERACTIONS

Introduction

X-ray crystallography must rank as the only science in which the limiting factor is the availability of suitable subject matter. To study a subject by crystallographic techniques, one must be able first to crystallize it; frequently, such a study ends here, as only a few of the many things that one would like to be able to study in this way are amenable to crystallization. As a result, crystallographers study what they can crystallize, for it is only in a few fortunate cases that one is able to crystallize what one would like to study.

For many years, this was the case in the study of nucleic acids by crystallographic techniques. The early work in this area was done entirely by diffraction analysis of drawn natural fibers, because none of the species of nucleic acids that it is possible to isolate from natural sources, except for a few of the transfer RNA molecules, can be crystallized. From this work, it was at least possible to get cylindrically averaged data on long stretches of these molecules at the molecular level, and based on this information, the now classical A and B forms of DNA were first recognized (2, 5). Nevertheless, the fruitfulness of this work can be measured by the fact that the structural basis of DNA as a self-complementary double helix was first deduced by Watson and Crick (8) largely from these data . This brilliant insight fused the then disparate fields of biochemistry and genetics into the grand synthesis of molecular biology and changed our understanding of the basis of life forever.

In recent years, this state of affairs for nucleic acid crystallography has all changed. Thanks largely to the work of Khorana and coworkers (3), it has become possible to make defined sequences of nucleic acids synthetically, in quantities large enough and of sufficient purity to make growth of single crystals of many of these sequences a reality. As a result, these molecules have now come under the scrutiny of high resolution crystallographic analysis, and the structures of many of these compounds have been determined at the atomic level. As proof of the great potential for new insights into nucleic acid structure that this has heralded, the first such structure of a DNA sequence, that of the hexamer C-G-C-G-C-G by Rich and coworkers (7), astounded the world by revealing the existence of a previously unsuspected conformation of DNA, the left-handed Z form, the implications of which are still being debated with respect to the process of gene regulation in the living cell. Since then, single crystal structures of the classical A and B forms of the early fiber studies have been also been determined (1, 9), and much has been learned from these studies about the architecture of DNA at the atomic level.

What these structures provide us with are snapshots of specific instances of DNA molecules, as frozen out in the crystalline state. Such pictures define what the molecule does in each of these instances, showing explicitly how a particular set of conditions affect the sequence, and how the sequence can respond to these conditions. When studied as a series of sequences, they show how sequence and structure interact, providing a basis for understanding

how a seemingly uniform molecule can in fact be rich in structural detail, which makes different sequences become different structures. And, finally, when studied in complex with other molecules which interact with it, they can define the nature of these interactions, and how structure affects function.

This work describes the results of three such studies of nucleic acids and what these studies have told us. The first of these is of a complex of the antitumor drug cisplatin with the B-DNA 12mer C-G-C-G-A-A-T-T-C-G-C-G and might be thought of as "taking advantage of a good thing." The complex was originally used as a phasing derivative in the solution of the native structure by multiple isomorphous replacement techniques (9). Since the compound is such an important antitumor agent, a study of this complex as an independent structure was undertaken in order to see how the drug interacted with DNA; in particular, the hope was to see what makes cisplatin such an effective antitumor drug, while its <u>trans</u> isomer is therapeutically inert. The results were at first rather disappointing in that they showed virtually no effect on the DNA from binding of the drug. But nonetheless, the details of the structure did place constraints on what sort of complexes the drug would be able to form with DNA in the double-helical state, and because of this we termed this complex the "primary mode of binding" of the drug to DNA in the native form (10).

The second study involved a continuation of an earlier project on a complex of the lac repressor protein and its DNA binding site and would definitely have to be thought of as "shooting the moon." The lac operon is the most widely studied DNA regulatory system, and a

structure of this repressor and operator in complex would tie together the enormous body of biochemical and genetic information which is known about it into a precise mechanism for gene regulation in living systems. Begun in this group by John Rosenberg even before DNA synthesis had become practical for crystallographic use, the original goal was to cocrystallize the entire lac repressor molecule with a DNA restriction fragment containing its consensus operator binding site. Unfortunately, the original project had to be abandoned after several years of effort failed to produce suitable crystals of the complex. It was revived, however, because of two subsequent developments: first, the ability to isolate a small proteolytic fragment of the large, unstable repressor called the lac headpiece, as a small, stable protein which still possesses the specific DNA binding activity of the whole repressor; and second, the advent of the new DNA synthesis techniques, which made it possible to synthesize the 21 base pair operator site as a pure, synthetic DNA oligomer. Unfortunately, however, this second attempt was also unsuccessful, and also for the same reason as before, But this time the project shows some promise for success with additional effort.

The final work consisted of the study of a complex of the B 12mer with the DNA fluorochrome and anthelmintic Hoechst 33258. Hoechst is a member of a class of DNA binding molecules, many of which show promise as antitumor drugs, which bind to DNA by means of noncovalent, nonintercalative interaction with the minor groove. In addition, Hoechst is structurally similar in several key respects to the antitumor antibiotic netropsin, the structure of whose complex with

the same 12mer had previously been determined in our lab (4). For this reason, a structural study of the complex of Hoechst bound to the DNA oligomer also was tried, in the hope that this would be "going for a sure thing." Fortunately in this instance, it was. The structure of the complex gave a wealth of information on how this molecule acts as a minor groove binder, confirming many of the details that were deduced from the netropsin complex of how these molecules bind to DNA, and providing additional insights into how its unique structure influences its ability to bind to DNA (6).

References

1. Conner, B. N., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1982) Nature 295, 294-299.
2. Fuller, W., Wilkins, M. H. F., Wilson, H. R., Hamilton, L. A. and Arnott, S. (1965) J. Mol. Biol. 12, 60-80.
3. Khorana, H. G., Agarwal, K. L., Buechi, H., Caruthers, M. H., Gupta, N. K., Kleppe, K., Kumar, A., Ohtsuka, E., RayBhandary, U. L., van de Sande, J. H., Sgaramella, V., Terao, T., Weber, H. and Yamada, T. (1972) J. Mol. Biol. 72, 209-222 et seq.
4. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) J. Mol. Biol. 183, 553-563.
5. Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. and Hamilton, L. D. (1960) J. Mol. Biol. 2, 38-64.
6. Pjura, P., Grzeskowiak, K. and Dickerson, R. E. (1986) manuscript submitted.
7. Wang, A. H-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.
8. Watson, J. D. and Crick, F. H. C. (1953) Nature 171, 737.
9. Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 287, 755-758.
10. Wing, R. M., Pjura, P., Drew, H. R. and Dickerson, R. E. (1984) EMBO J. 3, 1201-1206.

CHAPTER 1

THE PRIMARY MODE OF BINDING OF CISPLATIN TO A B-DNA DODECAMER

Introduction

The story of the antitumor compound cisplatin is a typical example of serendipity in scientific research. Discovered by accident during a study of the effects of electric fields on bacterial growth (7), in which strange distortions of normal binary fission were found to be caused by hydrolysis products of the supposedly "inert" platinum electrodes being used, it has proven to be one of the most effective antineoplastic agents known, curing several forms of testicular and ovarian cancer, and in conjunction with other agents providing effective treatment of a wide range of cancers (4). From very early on, the activity of cisplatin was thought to result from interaction with DNA. Current opinion favors the idea that the drug's active lesion is an intrastrand crosslink between the N7 atoms of two adjacent guanine bases in DNA, interfering with DNA replication and transcription (5). Several structures of the compound complexed with nucleosides and nucleotides have shown the details of such an interaction (1, 2, 3), and recently a structure of cisplatin complexed with a pGpG dimer has demonstrated that this complex can occur in an intact strand of DNA (8). Nevertheless, the reason why such a lesion should make cisplatin so effective an antineoplastic agent *in vivo* has remained obscure. Efforts to capitalize on this insight to produce more effective forms of the drug have not yet produced agents that are any more effective than the parent compound.

Our involvement with cisplatin also began in a rather roundabout way, when, during the structural solution of the B-DNA dodecamer

C-G-C-G-A-A-T-T-C-G-C-G, cisplatin was used to produce a heavy atom derivative for multiple isomorphous phase analysis of the native structure (9). Because of the importance of the compound as an antineoplastic agent, we decided to solve the complex of the drug with the B-form 12mer as an independent structure, so that we could study at first hand the details of its interaction with a segment of double-helical DNA. This was done by diffusing the drug into pre-grown crystals of the dodecamer, collecting diffraction data on the crystals and solving the structure by molecular replacement. In this way, we would see how the drug interacts with DNA in the crystal, in which the helices are effectively locked into the double-helical B conformation.

What we found largely confirmed the results of the other crystallographic and solution studies, as well as added a few new insights into limits on its possible interaction with DNA in the double-helical form. As expected from previous work, the compound bound to the DNA in our crystals solely to N7 atoms of three guanine bases in the duplex. As predicted by the work of Goodgame et al. (3), one corner of the square planar complex around Pt was occupied by the guanine N7 atom and another by a water molecule bridging between Pt and the guanine O6. (Waters presumably replace chlorines in the aquated form of the complex.) Such a Pt-O6 bridge requires tilting the metal coordination plane some 60 degrees out of the plane of guanine. Sequence effects also were seen on occupancy of the metal at different sites, with differences not only between guanines adjacent to different bases in the sequence but also between two guanine sites on

symmetrically related bases on the palindromic oligomer, suggesting that local conformational effects of the helix played a role in determining binding as well. Finally, in agreement with results from solution studies, we found that at the low levels of binding seen in this structure, binding of the metal produced no major distortions of the helix at all, with a slight motion of the bound base-pairs toward the major groove being the only visible effect of the drug on the DNA. We interpreted our results as representing the probable primary mode of binding of cisplatin with B-DNA in the cell, with this monodentate binding setting the stage for the formation of later bidentate binding on some occasion in which the helix would be disrupted, as in replication or repair. A full report of the cisplatin structure analysis is given in the following paper.

References

1. Cramer, R. E., Dahlstrom, P. L., Seu, M. J. T., Norton, T. and Kshiwagi, M. (1980) Inorg. Chem. 19, 148-154.
2. Gellert, R. W. and Bau, R. (1975) J. Amer. Chem. Soc. 97, 7379-7380.
3. Goodgame, D. M. L., Jeeves, I., Phillips, F. L. and Skapski, A. C. (1975) Biochim. Biophys. Acta 378, 153-157.
4. Javadpour, N. (1985) Urology 25, 155-160.
5. Lippard, S. J. (1982) Science 218, 1075-1082.
6. Prestayko, A. W., Crooke, S. T. and Carter, S. K., eds. Cisplatin: Current Status and New Developments (Academic Press, New York, 1980).
7. Rosenberg, B., Van Camp, L. and Krigas, T. (1965) Nature 205, 698-699.
8. Sherman, S. E., Gibson, D., Wang, A. H-J. and Lippard, S. J. (1985) Science 230, 412-417.
9. Wing, R. M., Drew, H. , Takano, T., Broca, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) Nature 287, 755-758.

The EMBO Journal vol.3 no.5 pp.1201 – 1206, 1984

The primary mode of binding of cisplatin to a B-DNA dodecamer: C-G-C-G-A-A-T-T-C-G-C-G

Richard M. Wing[1], Philip Pjura, Horace R. Drew[2] and Richard E. Dickerson*

Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90024, USA

[1]Present address: Department of Chemistry, University of California, Riverside, CA 92521, USA
[2]Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
*To whom reprint requests shold be sent
Communicated by J.H. Miller

When cisplatin [*cis*-diamminodichloroplatinum(II)] is diffused into pre-grown crystals of the B-DNA double-helical dodecamer C-G-C-G-A-A-T-T-C-G-C-G, it binds preferentially to the N7 positions of guanines, with what probably is an aquo bridge between Pt and the adjacent O6 atom of the same guanine. The entire guanine ring moves slightly toward the platinum site, into the major groove. Only three of the eight potential cisplatin binding sites on guanines actually are occupied, and this differential reactivity can be explained in terms of the relative freedom of motion of guanines toward the major groove. This shift of guanines upon ligation may weaken the glycosyl bond and assist in the depurination that leads to mismatch SOS repair and G.C to T.A transversion.
Key words: cisplatin/B-DNA/X-ray structure

Introduction

Cisplatin [*cis*-diamminodichloroplatinum(II)] is one of the most effective and extensively studied inorganic antitumor drugs. In contrast, for reasons that are not fully understood, its transplatin isomer (*trans*-$Pt(NH_3)_2Cl_2$) exhibits little antineoplastic activity (Rosenberg *et al.*, 1969; Prestayako *et al.*, 1980). Cisplatin is believed to interfere with DNA replication and transcription in a manner similar to that of alkylating agents (Harder and Rosenberg, 1970; Howle and Gale, 1970; Taylor *et al.*, 1976). In solution under normal physiological conditions, cisplatin binds most strongly to the N7 position of guanine, with lesser attraction for adenine and cytosine and none for thymine or uridine (Mansy *et al.*, 1973, 1978).

Single crystal X-ray structure analyses have been carried out previously for complexes of cisplatin or of *cis* $Pt(en)Cl_2$ (en = ethylenediamine) with inosine monophosphate or with guanosine (Goodgame *et al.*, 1975; Gellert and Bau, 1975; Cramer *et al.*, 1980). In each case the platinum atom bridges the N7 positions of two guanine rings, with *cis* ligation to Pt. This has led to the suggestion that cisplatin might act *in vivo* by making interstrand or intrastrand cross-links between guanines. Interstrand cross-linking (Horacek and Drobnik, 1971; Roberts and Pascoe, 1972; Zwelling *et al.*, 1979) would necessitate unwinding of the double helix, since a cisplatin group bound to guanine in the bottom of the major groove would be inaccessible to another strand from the same or a different helix, for the second bond. Intrastrand cross-linking could be achieved between adjacent guanines along a polynucleotide chain (Cohen *et al.*, 1980; Lippard, 1982), but only at the price of destacking the bases and tipping the N7 positions toward the metal site in the major groove.

Alkylating agents such as benzo[a]pyrene, 2-acetylaminofluorene and aflatoxin B1 have another important effect on DNA, in addition to simple alkylation and cross-linking. They are potent mutagens, introducing G.C to T.A transversions in a parallel fashion (Foster *et al.*, 1983). The common feature of these alkylation adducts seems to be a weakening of the glycosyl bond leading to depurination, followed by insertion of adenine opposite the depurinated site by the SOS repair system of the cell (Witkin, 1976, 1982). Miller (1983) has suggested that cisplatin might also mimic this property of alkylating agents, leading to mutagenesis with transversion as well as cross-linking. In support of this idea, Brouwer *et al.* (1981) have found that cisplatin strongly favors G.C to T.A transversions among 650 nonsense mutations produced in the *lacI* gene of *Escherichia coli*, supporting the idea that what is occurring is depurination followed by bypass repair. $Pt(dien)Cl^-$ appears to strengthen the glycosyl bond when it complexes with the N7 position of guanine (Johnson, 1982; Johnson *et al.*, 1982). However, as these authors themselves point out, 'The enhanced biological activities of *cis*-PDD compared with *trans*-PDD and [Pt(dien)Cl]Cl are a consequence of different platinum-DNA adducts formed by these compounds *in vivo*'. For example, cisplatin is a potent mutagen, with excision repair and daughter strand gap repair, and has clinically useful antitumor activity; $Pt(dien)Cl^-$, in contrast, is only a very weak mutagen, shows no evidence of repair, and has no antitumor activity. These observations provide strong circumstantial support for the idea that mutagen activity involves as an early step the weakening and rupture of the glycosyl bond, and subsequent depurination.

Regardless of whether the ultimate outcome is cross-linking or mutagenesis, the initial biological interaction of cisplatin with double-helical DNA almost certainly is binding to guanine N7 with loss of one of the chloro (or rather, aquo) ligands. As Lippard (1982) has pointed out, aqueous solution of cisplatin is followed by displacement of chlorides by water molecules. Displacement of one water and ligation to N7 then follows, but it is unclear whether a second interaction occurs with the adjacent O6 under physiologically reasonable conditions (Chu *et al.*, 1978). The distance between Pt and O6 is too long for a direct Pt-O bond, but too short to accommodate an intermediate bridging water molecule unless the square planar Pt complex is rotated about the Pt-N7 bond, and out of the plane of the guanine rings (Goodgame, 1975).

In lieu of co-crystallized complexes of cisplatin with G-G-containing DNA oligomers, thus far unattainable, we have solved the crystal structure of the diffusion complex of cisplatin with the B-DNA double-helical dodecamer of sequence C-G-C-G-A-A-T-T-C-G-C-G. Attempts to achieve complete substitution or 100% binding to the major site led to destruction of crystal order as reflected in degradation of the X-ray pattern, an observation that may support the hypothesis that cisplatin binding ultimately deforms and

Table 1. Binding of cisplatin to C-G-C-G-A-A-T-T-C-G-C-G as a function of concentration

Data set		Native	Pt1	Pt2	Pt3
Cell dimensions (Å)	*a*	24.87	24.36	24.33	24.16
	b	40.39	40.05	40.08	39.93
	c	66.20	66.13	66.26	66.12
Resolution, *d* (Å)		1.9	2.5	2.2	2.6
Residual error, R.	All data:	23.9%	20.0%	27.0%	16.6%
	Two-sigma data:	17.8%	–	–	11.2%
Percent site occupancy	G16	–	20%	38%	61%
	G4	–	10%	17%	30%
	G10	–	–	13%	22%
Pt-N7 bond length, G16 (Å)		–	2.51	2.43	2.16
Apparent guanine shift (Å)		–	0.51	0.43	0.16
Occupancy ratios:	G16/G4		2.0	2.24	2.03
	G16/G10		–	2.92	2.77

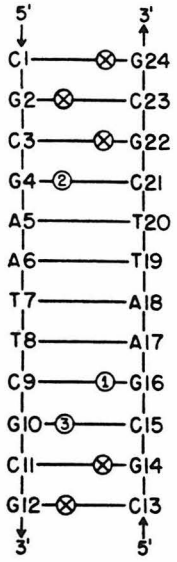


Fig. 1. Unrolled ladder diagram of the major groove of the B-DNA C-G-C-G-A-A-T-T-C-G-C-G dodecamer, showing the eight potential cisplatin binding N7 sites as circles. Numbers 1 – 3 indicate observed binding sites in order of decreasing occupancy by cisplatin complexes. No binding was observed at the five positions marked by circled X.

destroys the double helix. However, lower levels of substitution yield a clear indication of the cisplatin binding sites, and a pattern of selective binding to guanines that provides information about the primary binding steps in the interaction of cisplatin with double-helical DNA. The most highly substituted data set has been refined to completion, and two other sets have been refined to the point where meaningful structure comparisons can be made with the primary set.

Results

The three data sets, labeled Pt1, Pt2 and Pt3 in order of increasing cisplatin substitution, are listed in Table I. In set Pt2 the emphasis was on resolution of the data; in Pt3, on maximal substitution. The Pt3 set was chosen for complete refinement (see Materials and methods), to a final zero-sigma residual error or R factor of 16.6%, and a two-sigma R factor of only 11.2%, the lowest values obtained in any of our oligonucleotide refinements. The double-helical B-DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G has eight guanines that could serve as binding sites for cisplatin, as indicated by circles on the rungs of the DNA ladder in Figure 1. No more than three of these were occupied by cisplatin even in the most highly substituted crystal, and always in the order of decreasing affinity: G16, G4, G10. As Table I indicates, in the most highly substituted Pt3 set, site G16 exhibits 61% occupancy, with 30% at G4 and 22% at G10. Such incomplete site occupancy is not without precedent in other crystal structure analyses of complexes of cisplatin with organic bases. Goodgame (1975) reacted cisplatin with inosine monophosphate prior to crystallization, but still obtained a non-stoichiometric compound with only 0.56 platinum atoms per two inosine bases. The relative site occupancies in the cisplatin complex of C-G-C-G-A-A-T-T-C-G-C-G listed in Table I can be given a straightforward structural explanation, as described below.

Figure 2 shows a minor groove stereo pair drawing of the refined Pt3 structure, with cisplatin groups depicted at sites G4, G16 and G10 from top to bottom. The three ligands drawn around each platinum atom were located from difference electron density maps and then refined as described in Materials and methods. No restraints were applied to tie the ligands to the platinums during refinement, and the fact that the PtL_3 complexes remained intact, with reasonable Pt-L bond lengths of ~2.0 Å as listed in Table II, indicates that even at less than complete substitution, real information about the structures of the complexes is being obtained. Platinum positions may be regarded as secure; ligand positions are reasonable and suggestive.

A close up view of sites G16 and G10 in the bottom half of the helix is provided by the stereo drawings of Figure 3. Bearing in mind the provisional nature of the ligand sites, it appears that the platinum atom is indeed bridged by a ligand, probably water, to the adjacent O6 atom on the same guanine ring, and that the square plane of Pt ligands is rotated out of

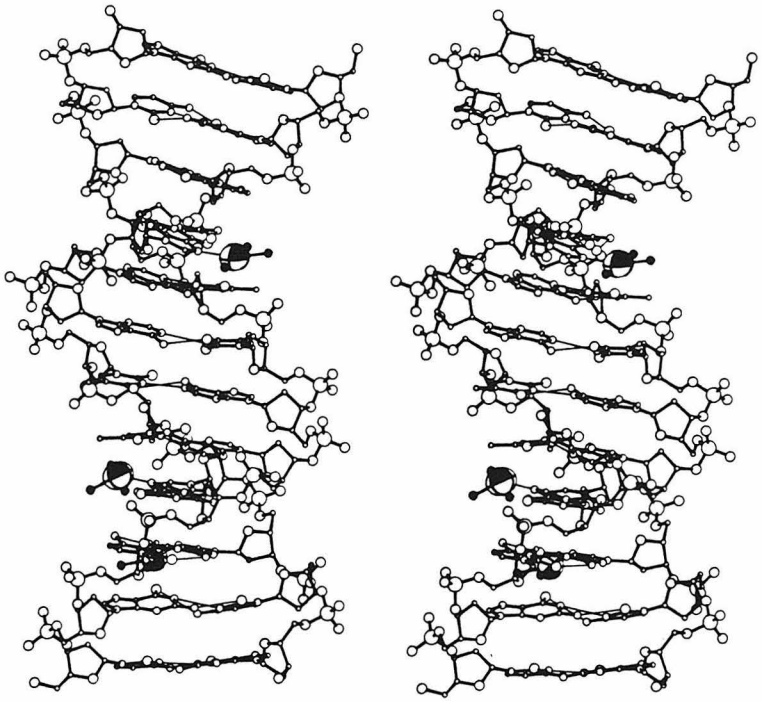


Fig. 2. Major groove view of the cisplatin complex of C-G-C-G-A-A-T-T-C-G-C-G as refined from data set Pt3. Base pair C1.G24 is at the top, and G12.C13 is at the bottom. Pt atoms at guanines G4, G16 and G10 (from top to bottom) are represented by their anisotropic thermal ellipsoids. The three smaller crossed spheres around each Pt are ligand sites obtained as described in Materials and methods. Other atoms are P, O, N and C in order of descending size.

Table II. Cisplatin-DNA complex geometry

	G16	G4	G10
Percent occupancy	61%	30%	22%
Bond lengths (Å)			
Pt-N7	2.16	2.31	2.23
Pt-L1	2.02	1.75	1.83
Pt-L2	1.81	2.25	2.10
Pt-L3	2.00	2.23	2.13
Bond angles (°)			
L1-Pt-L2	166	175	173
L1-Pt-L3	98	95	105
L2-Pt-L3	79	85	82

the guanine plane as in Figure 4. So far this only confirms expectations from previous crystal structure analyses with bases and nucleotides. However, Figure 3 also indicates that base pairs in the cisplatin complex are shifted relative to their positions in the native structure. At the most highly substituted G16 site, the guanine ring moves towards the platinum, bringing the hydrogen-bonded cytosine with it. Cytosine C15 just below it in Figure 3a moves in the same direction, but adenine A17 just above it moves in a different direction along the long axis of the base pair, indicating that what is being observed is not a rigid-body shift of the entire helix, but a local deformation. At the less highly substituted G10 site (top rear of Figure 3a), the guanine appears to pivot so the five-membered ring again moves toward its bound Pt site.

Discussion

An X-ray crystal structure analysis gives the averaged molecular structure over the entire crystal. If substitution were complete, one would observe an image of the pure cisplatin/DNA complex. If there were no substitution, one would see the image of the parent DNA molecule. For partial substitution, what is observed is a composite image of both structures, with the cisplatin-bound image at each site weighted according to the degree of substitution at that site. If the binding of cisplatin involves a displacement of the guanine ring, then for partial occupancy sites the apparent position of the ring will be intermediate between that of the parent helix and that of the 1:1 cisplatin complex. This seems the best explanation of a curious observation that arises when the three structures with different levels of cisplatin substitution are compared: Pt1, Pt2, and Pt3 (Table I). The apparent Pt-N7 bond length of 2.51 Å in the low-substitution Pt1 structure is far too long, but the bond becomes shorter in Pt2 and Pt3 as the degree of substitution increases. Linear extrapolation of a plot of bond length *versus* percent substitution leads to a Pt-N7 distance of 1.8 Å at 100% substitution, overshooting slightly the expected 1.97–2.02 Å (Goodgame *et al.*, 1975; Gellert and Bau, 1975; Cramer *et al.*, 1980), but

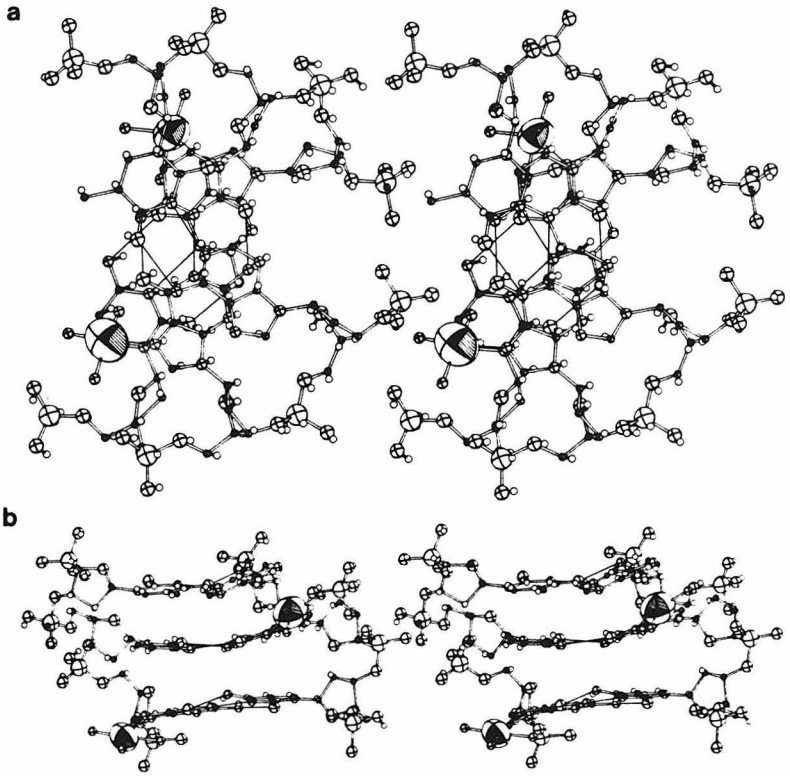


Fig. 3. Plan view (**a**) and major groove edge view (**b**) of cisplatin sites G16 and G10 and their immediate surroundings. Base pair T8.A17 is nearest the viewer in (**a**) and at the top in (**b**). Base pair G10.C15 is farthest from the viewer in (**a**) and at the bottom in (**b**). Platinum and ligand atoms are displayed as in Figure 2, and other crossed spheres are P, O, N and C in order of descending size. The small, uncrossed sphere beside each atom of the DNA helix indicates the location of the equivalent atom in the native dodecamer structure without cisplatin. (**a**) Shows particularly clearly that binding of cisplatin to the N7 of guanine G16, the most highly substituted site, pulls the entire base pair into the major groove in the direction of the Pt atom. A similar but smaller effect is visible at the more weakly substituted G10 site.

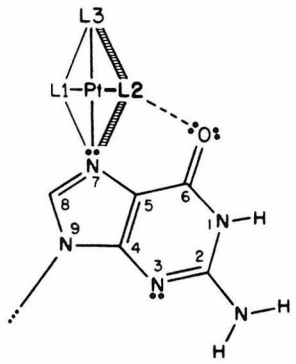


Fig. 4. Ligation geometry suggested by the Pt3 structure analysis. The N7 of guanine supplies one of the four square planar Pt ligands. A second ligand L2, most probably a water molecule, bridges the Pt and the guanine O6. Ligands L1 and L3 probably are the ammonia molecules of the original cisplatin complex. Since the Pt atom is symmetrically located relative to the geometry of the five-membered ring of guanine, the separation between Pt and O6 is too short to permit a bridging ligand unless the square planar Pt complex is rotated out of the plane of the guanine. This same geometry was suggested by Goodgame *et al.* (1975) as the initial mode of binding of cisplatin to double-stranded DNA.

indicating a substitution-dependent effect on apparent bond length.

This observation and the atom shift evidence in Figure 3 both indicate that when cisplatin binds, the guanine ring moves out of the base pair stack of the double helix into the major groove. Since the apparent shift in the composition image is 0.5 Å for 61% occupancy, one might expect a full-occupancy real shift of nearly an Ångstrom. This could be enough to destabilize the helix, disrupt the crystal, and perhaps even to weaken and break the glycosyl bond that attaches the guanine ring to the deoxyribose. Hence the present structure analysis, even though it is concerned directly only with the initial steps of cisplatin-DNA interaction, also provides support for the idea that depurination and mutagenesis are involved in the overall process.

The radically different degrees of substitution at the eight guanine sites along the double helix at first are surprising, but they have a simple explanation in terms of local guanine environment. The occupancy trends in Table I can be factored into two components: (i) increased reactivity of guanines the farther they are from the ends of the helix, and (ii) greater reactivity in the bottom half of the molecule than the top. The occupancy ratio between equivalent sites four base pairs in from each end, G16/G4, is 2.0–2.2 in all three data sets of Table I, and that between the sites four and three base pairs in

from the same end, G16/G10, is 2.8–2.9. An occupancy ratio of 2.3 corresponds at 298 K to a free energy difference of 0.5 kcal/mol. Why should symmetrically equivalent positions on the two helix strands exhibit a difference in binding energy of a little less than 0.5 kcal, and why should the binding energy at the fourth base pair from the end differ from that at the third base pair by a slightly greater amount?

The increased reactivity of guanines the farther they are from the ends of the helix, probably is explained by intermolecular interactions in the crystal. As can be seen from Figure 3 of Wing *et al.* (1980), the crystals contain columns of helices with overlapping ends. The minor grooves of two successive helices are interlocked, with the first two base pairs of one helix hydrogen-bonded to the first two base pairs of the next helix via N3...H-N2 and N2-H...N3 bonds between guanines. Hence the first two base pairs in from each end of the helix are immobilized on the minor groove side, and cannot shift toward the major groove to accommodate cisplatin binding. The third base pair is somewhat less constrained, and the fourth pair is freest of all. The reactivity of a particular guanine appears to be related directly to its freedom from constraints, and its ability to move toward the potential cisplatin site.

Why should the extent of cisplatin binding, at the same distance in from the end of the helix, differ between the two ends? The molecules after crystallization have an overall 19° bend in helix axis, and the sharpest bending occurs at the upper end of the molecule in Figure 2, where a spermine molecule bridges the major groove in the parent DNA structure. The molecule appears to close down its major groove slightly around the spermine. In the cisplatin complex the spermine molecule is displaced by the diffused-in cisplatin, but the asymmetric bending at the two ends of the molecule remains. In effect, the top four base pairs are inclined to the left in Figure 2 by rolling the fourth base pair against its neighbors along its long axis (Dickerson *et al.*, 1983). This constricts the major groove, and probably makes it difficult for guanines G22 and G4 to slip toward the major groove, hence decreasing their affinity for cisplatin.

In summary, the initial steps of binding of cisplatin to this B-DNA double helix involve ligation to the N7 position of those guanines that are most free to move slightly into the major groove. A second, stabilizing interaction appears to form between the platinum and the adjacent O6 atom on the same guanine, most probably involving a water molecule as a bridge. This may be the basis for the difference in reactivity of the *cis* and *trans* isomers. The destruction of crystal order that always accompanies high substitution makes it clear that this is not the entire story; the binding that we see here is only the preface to a more serious rearrangement. This may involve distortion of the helix during formation of Pt cross-links between pairs of guanines. The shift of guanine positions upon interaction with Pt suggests that it may also involve straining the glycosy bond and depurination, whch could lead to mutagenesis during the repair process. What is needed at present is a structure analysis of a cisplatin complex containing the G-G sequence, in which the complex is formed in solution, purified, and crystallized. Efforts in this direction are continuing.

Materials and methods

Native C-G-C-G-A-A-T-T-C-G-C-G crystals grown by vapor diffusion as described previously (Wing *et al.*, 1980; Drew *et al.*, 1981) were equilibrated with solid cisplatin or with 25–55% saturated solutions of cisplatin in 40:60 water/MPD (2-methyl-2,4-pentanediol) at 4°C for periods of a few days to several weeks. The derivatized crystals then were back-soaked in fresh platinum-free MPD/water solution for 4–24 h to remove any non-covalently bound or interstitial platinum. Attempts to obtain high occupancy of platinum binding sites, e.g., a Pt to DNA site ratio close to 1:1, invariably led to destruction of the X-ray pattern from the outside inward, and to increased sensitivity of the crystals to X-ray damage. Although backsoaking was found to aid the situation, all cisplatin-substituted crystals remained very sensitive to X-ray irradiation; the useful data collection lifetime fell from ~200 h for the native dodecamer to 25 h for the highest level of Pt substitution.

Three data sets were collected at different levels of substitution, as listed in Table I. Pt1, the least substituted but most ideally isomorphous set, was used for phase analysis of the parent or native structure. The higher substituted of the remaining sets, Pt3, was selected for complete Jack-Levitt restrained energy refinement as described by Fratini *et al.* (1982), leading to a final residual error or crystallographic R factor of 16.6% for all data or 11.2% for those reflections above the two-sigma confidence level. 128 solvent peaks were added gradually during refinement, using the strategy described by Drew and Dickerson (1981) and by Kopka *et al.* (1983).

Initial ligand positions around the Pt sites were obtained from difference maps in which the DNA, ordered solvent, and Pt atoms were subtracted out. These ligands, roughly in a square planar array around each Pt site, then were refined in the Jack-Levitt procedure without restraints, i.e., independently of any specified connection to the Pt atoms. The fact that the Pt-ligand positions remained around 2.0 Å (Table II) is evidence that the sites are real and that the refinement is meaningful. At the very end of 61 cycles of Jack-Levitt refinement, one more test of ligand geometry was made: the DNA atoms were held fixed, ligand atoms were eliminated, and Pt atoms were subjected to full-matrix, anisotropic least squares refinement. New ligand positions then were obtained from difference maps, and refined isotropically along with the anisotropic Pt. This refinement was well-behaved for the G16 cisplatin complex with 61% occupancy, but ligand positions wandered away from the Pt positions for the two less-substituted sites. The final coordinate set, which has been deposited with the Brookhaven Protein Data Bank for general distribution, is a composite set: anistropic data for the three Pt atoms along with their isotropic equivalents, full-matrix refined positions for the ligands of site G16, and best Jack-Levitt refined positions for the ligands of G4 and G10, and for the DNA. The original X-ray diffraction data also have been deposited.

The Pt2 data set with intermediate substitution was refined only to the point where no further motion of atoms in the DNA was occurring, and it was judged that no further comparative structure information would result by continuing. The main value of this was its Pt-N apparent bond distance.

Acknowledgements

We would like to thank Douglas C.Rees for help with anisotropic refinement of the Pt groups, Lillian Casler for preparation of Figures, Tsunehiro Takano for assistance with refinement programs, and Mary L.Kopka for helpful comments during the preparation of this manuscript. This work was carried out with the support of NIH grant GM-30543 and NSF grant PCM82-02775.

References

Brouwer,J., van de Putte,P., Fichtinger-Schepman,A.M.J. and Reedijk,J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7010-7014.

Chu,G.Y.H., Mansy,S., Duncan,R.E. and Tobias,R.S. (1978) *J. Am. Chem. Soc.*, **100**, 593-606.

Cohen,G.L., Ledner,J.A., Bauer,W.R., Ushay,H.M., Caravana,C. and Lippard,S.J. (1980) *J. Am. Chem. Soc.*, **102**, 2487-2488.

Cramer,R.E., Dahlstrom,P.L., Seu,M.J.T., Norton,T. and Kashiwagi,M. (1980) *J. Inorg. Chem.*, **19**, 148-154.

Dickerson,R.E., Kopka,M.L. and Pjura,P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7099-7103.

Drew,H.R. and Dickerson,R.E. (1981) *J. Mol. Biol.*, **151**, 535-556.

Drew,H.R., Wing,R.M., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2179-2183.

Foster,P.L., Eisenstadt,E. and Miller,J.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2695-2698.

Fratini,A.V., Kopka,M.L., Drew,H.R. and Dickerson,R.E. (1982) *J. Biol. Chem.*, **257**, 14686-14707.

Gellert,R.W. and Bau,R. (1975) *J. Am. Chem. Soc.*, **97**, 7379-7380.

Goodgame,D.M.L., Jeeves,I., Phillips,F.L. and Skapski,A.C. (1975) *Biochim. Biophys. Acta*, **378**, 153-157.

Harder,H.C. and Rosenberg,B. (1970) *Int. J. Cancer*, **6**, 207-216.

Horacek,P. and Drobnik,J. (1971) *Bicohim. Biophys. Acta*, **254**, 341-347.

Howle,J.A. and Gale,G.R. (1970) *Biochem. Pharmacol.*, **19**, 2757-2762.

Johnson,N.P. (1982) *Biochem. Biophys. Res. Commun.*, **104**, 1394-1400.
Johnson,N.P., Macquet,J.P., Weibers,J.L. and Monsarrat,B. (1982) *Nucleic Acids Res.*, **10**, 5255-5271.
Kopka,M.L., Fratini,A.V., Drew,H.R. and Dickerson,R.E. (1983) *J. Mol. Biol.*, **163**, 129-146.
Lippard,S.J. (1982) *Science (Wash.)*, **218**, 1075-1082.
Mansy,S., Rosenberg,B. and Thomson,A.J. (1973) *J. Am. Chem. Soc.*, **95**, 1633-1640.
Mansy,S., Chu,G.Y.H., Duncan,R.E. and Tobias,R.S. (1978) *J. Am. Chem. Soc.*, **100**, 607-616.
Miller,J.H. (1983) *Annu. Rev. Genet.*, **17**, 215-238.
Prestayko,A.W., Crooke,S.T. and Carter,S.K., eds. (1980) *Cisplatin: Current Status and New Developments*, published by Academic Press, NY.
Roberts,J.J. and Pascoe,J.M. (1972) *Nature*, **235**, 282-284.
Rosenberg,B., van Camp,L., Trosko,J.E. and Mansour,V.H. (1969) *Nature*, **222**, 385-386.
Taylor,D.M., Tew,K.D. and Jones,J.D. (1976) *Eur. J. Cancer*, **12**, 249-254.
Wing,R., Drew,H., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1980) *Nature*, **287**, 755-758.
Witkin,E.M. (1976) *Bacteriol. Rev.*, **40**, 869-907.
Witkin,E.M. (1982) *Biochimie*, **64**, 549-555.
Zwelling,L.A., Anderson,T. and Kohn,K.W. (1979) *Cancer Res.*, **39**, 365-369.

Received on 5 December 1983; revised on 17 February 1984

Note added in proof

Since submission of this manuscript, Rubin, Sabat and Sundaralingam (*Nucleic Acids Res.*, **11**, 6571-6586, 1983) have published an account of the very same binding of cisplatin to guanines 15 and 18 of $tRNA^{Phe}$: direct platinum ligation to N7, and ligand bridging to O6.

CHAPTER 2

CRYSTALLOGRAPHIC STUDIES OF THE LAC REPRESSOR DNA-BINDING DOMAIN

1. Introduction

It is never easy to write about one's failures. This chapter chronicles efforts made toward the crystallographic analysis of the DNA-binding domain of the lac repressor protein, both by itself and in complex with a twenty-one base pair DNA oligomer spanning the repressor's consensus binding site. This project was begun in 1974 in this group by John Rosenberg, with the original goal of isolating, crystallizing and solving the structure of the entire repressor molecule, in complex with a DNA restriction fragment containing its binding site. At that time, Rosenberg *et al.*(52) were successful in developing a procedure for isolation of the protein in pure, stable form but were not successful in producing crystals of suitable quality for crystallographic structure analysis. My goals in this work were essentially the same: to isolate the DNA binding domain in pure form and then to crystallize it, both alone and in complex with its operator site. And likewise, I was successful at the former, but not the latter, task.

The lac repressor/operator system of *E. coli* was the first example of a regulatory system identified and studied in a living organism at the genetic and molecular level. First delineated by Jacob and Monod (26), as described below, it became the paradigm of gene regulation in a living system, and although other regulatory systems in both cells and viruses have since been identified and studied in great detail, the lac system remains the most extensively studied and well understood regulatory system of all. In one area of

study, however, it has recently been eclipsed by several other repressor/operator systems, and that is in the area of three-dimensional structural analysis. In the past few years, x-ray crystallographic studies have revealed at the atomic level the structures of the lambda (47) and cro (4) repressors of phage lambda, and of the catabolite gene activator protein (36) and the trp repressor (54) of E. coli. In addition, the low resolution structure of the phage 434 repressor (3) in complex with its DNA binding site has been solved, and work has begun on the solution of high resolution crystals of the lambda repressor/operator complex (28). The structures solved to date have revealed a number of interesting details about each of these proteins, in particular that there appears to be a conserved structural motif present in all of them consisting of a helix-turn-helix configuration that is believed to represent the specific DNA binding site of these molecules (43, 44, 60, 62). Comparison of amino acid sequences of these repressors with that of lac as well as of a number of other sequence-specific DNA binding proteins has shown that this structure appears to be conserved among all of them and probably is a fundamental structural mechanism of DNA recognition (61). Because of the enormous body of biochemical and genetic information available on the lac system, a structural study of the lac protein and DNA would add immensely to our understanding of gene regulation at the molecular level.

This second attempt was prompted by two developments that came about after the original work had been abandoned: the availability of a synthetic twenty-one base pair lac operator fragment, and the

discovery of a procedure for isolating the amino-terminal DNA binding region, or headpiece, of the lac repressor as a stable proteolytic fragment. The former was made possible by the development of liquid phase phosphotriester chemistry for the synthesis of DNA fragments of defined sequence in pure form (23), and large quantities of the twenty-one base pair consensus DNA binding site were synthesized for this purpose by Dr. Peter Dembek in our lab. The latter is described below and made it possible to go from working on the large and unstable tetrameric complex that is the lac repressor, to a small, stable monomeric peptide that nevertheless possesses all of the DNA binding activity of the intact protein. In light of the great successes that were being achieved at the time with other small repressor proteins, it was felt that these two factors made the proposed headpiece/operator project, which is in fact an entirely different project from the original one, seem very feasible.

This report describes the outcome of these two steps: the isolation of the headpiece, and attempts at crystallization of it and its DNA complex. The first step required the development from the isolation procedures then in use of a protocol that would allow the fragment to be isolated in high concentration and in very stable form, two specific requirements for the growth of macromolecules. How this was done is detailed below. The second step involved a number of trial screens for conditions under which the protein and the complex would yield crystals suitable for x-ray structural analysis. The specific conditions that were tried, the reason why they were chosen, and their results, are reported. Finally, an attempt at analyzing the

results and at making suggestions for future efforts on this system will be made, for the benefit of future workers on this project.

2. Background

In their study of the regulation of lactose metabolism in E. coli, Jacob and Monod (26) proposed a system by which the synthesis of the enzymes necessary for this process could be controlled by the presence of the metabolized sugar. Their scheme postulated the existence of two distinct types of genes in the organism: structural genes, called z, y and a, which code for the synthesis of beta-galactosidase, beta-galactoside permease and beta-galactoside transacetylase, the three enzymes necessary for lactose utilization; and a new class of genes, which they called regulatory genes, whose function is to see that the enzymes coded for by the structual genes are produced only when there is lactose present to be metabolized. Through analysis of the types of mutants of lactose metabolism regulation that could be isolated, they determined that there must be controlling this system two of these regulatory genes, which they called i and o and to which they assigned the following functions. In the absence of lactose, the product of the i gene, called a repressor, interacts with the o, or operator gene in such a way that production of the three structural genes is prevented. In the presence of the sugar, however, the repressor would not interact with the operator, and the structural genes would start to be synthesized. They speculated that the repressor must be a protein that would interact specifically with a section of the E. coli DNA within the operator gene and in so doing would block the synthesis of the structural genes only when lactose was not present. But when the sugar was present, it

would somehow stop this interaction and free the genes coding for the enzymes for synthesis.

This model, for what is now termed an inducible system, has since been shown to be the correct mechanism of control of this enzyme system, called the lac operon (8). The current picture is that it consists of a single stretch of DNA containing the above five genes along with another, called the promotor or p, that is the binding site for RNA polymerase, in the order i-p-o-z-y-a (38). The i gene codes for the lac repressor (21), a 155,000 dalton tetrameric complex of four identical subunits (51) that binds tightly and specifically to the operator region in the absence of its specific inducer, the lactose isomer allolactose (27), but not in its presence, and thereby inhibits the attachment of RNA polymerase at the promoter site. This accordingly prevents readout of the three structural genes (22). The lac operator has been sequenced (20), and the actual base pairs within the gene to which the repressor binds have been determined by DNA digestion studies (20, 56) and by repressor dissociation kinetics of various fragments of the operator sequence (6). This has shown the operator to consist of a pseudopalindrome of 21 base pairs, in which the outer twelve base pairs form a palindrome separated by a nine base-pair nonpalindromic region (6):
A-A-T-T-G-T-G-A-G-C-G-G-A-T-A-A-C-A-A-T-T. The monomeric repressor protein has also been sequenced (16), and both genetic mapping analysis (2, 55) and study of the effects of partial proteolysis on binding activity (17, 48) have shown the protein to be divided into two distinct functional regions: an amino terminal region, consisting

of amino acid residues 1 to 50, which are necessary for DNA binding activity, and a carboxy terminal region, residues 60 to 360, which binds the inducer molecule and regulates DNA binding activity, as well as forms the tetameric complex (66). In addition, it has been shown that these two functional regions are in fact distinct structural regions of the protein as well, with the smaller and more mobile amino terminal region, or "headpiece," being attached to the carboxy terminal segment, or "core," by an extremely flexible "hinge" region consisting of residues 51 through 59 (64). The repressor has been implicated by hybridization (18) and by direct visualization studies (45, 59) to have 222 molecular symmetry, with the core regions of each monomer forming the inter-subunit contacts and holding the headpieces outward in pairs on opposite ends of the cluster. DNA binding to this complex has been shown to occur through pseudosymmetric binding of two headpieces to opposite ends of the operator binding sequence (18, 30). Electron microscopic studies reveal a complex in which the repressor just touches the operator at one edge of the tetramer and in fact does so on only one side of the operator (1, 13, 24).

How this specific repressor-operator interaction occurs has been a point of major interest, and research has shown that lac repressor in fact has two modes of DNA binding: a tight, specific binding to the operator site itself, and a less tight, nonspecific binding to DNA in general (32). Specific binding has a dissociation constant K_{diss} of $3.5 \times 10^{13}\ M^{-1}$ (7), and is believed to occur through sequence-specific interactions within the major groove of the operator site (33, 34, 40). Nonspecific binding, on the other hand, has a

K_{diss} of approximately 10^3 less and seems to occur on either the major or the minor groove side (12, 68), since DNA in which the major groove is filled with blocking groups exhibits the same nonspecific binding affinity as normal DNA of similar sequence (50).

What is intriguing is that both of these types of binding are performed by what appears to be the same binding site on the headpiece. Numberous experiments have attempted to probe the structure of bound and unbound repressor and operator, and from them have come the following generalizations: First of all, amino acid residues 1 through 20 are important for binding activity, and, in particular, residues 7, 12, and 17, all tyrosines, are believed to interact directly with the DNA in both types of binding (11, 15). Secondly, operator DNA is partially unwound, by either 40 or 90 degrees, on binding to repressor (65), while non-operator DNA is not (50); but in general, no major structural reorganization of the DNA occurs in either type of binding (12, 25). Finally, although both of the headpiece binding sites (taken as pairs of headpieces in the intact repressor complex) and the operator binding sites should be symmetrical, and hence binding should be symmetrical as well, evidence suggests that binding to the operator by intact repressor is not symmetrical but in fact is stronger on one side of the operator region than the other (42).

Lac Headpiece

Studies of the interaction between lac repressor and lac operator made a significant methodological advance with the discovery by

Geisler and Weber (19) of a protocol for the isolation of the lac headpiece as a stable proteolytic fragment of the whole molecule. Under most conditions, the headpiece region is very quickly cleaved from the repressor and then degraded into several fragments under even mild proteolytic conditions, starting in the residue 51 to 59 "tether" region connecting the two domains (17); as a result, the headpiece cannot be isolated intact. Geisler and Weber found, however, that in a buffer system consisting of 1 M Tris·HCl pH 7.5 and 30% glycerol, proteolysis stops after this first cut, and the headpiece is resistant to further degradation. They developed a procedure for isolating the headpiece in homogeneous form and showed by circular dichroism that it possessed a stable secondary structure. Furthermore, they found that it was active in DNA binding in filter binding assays, though it apparently had lost the ability to distinguish operator from nonoperator DNA. This made it possible to shift from study of a complex of four chains with 360 amino acids each to the analysis of a single 51 to 59 amino acid monomer, simplifying the problem greatly. It was soon shown by a number of techniques, such as DNA protection studies with dimethyl sulfate (41), circular dichroism studies of operator vs nonoperator binding (14) and by nuclear magnetic resonance studies of headpiece binding to operator-containing DNA fragments (9), that the headpiece still did in fact discriminate between operator and nonoperator binding, and that the apparent loss of specific binding seen in the filter binding assay was due to the greatly reduced affinity of the headpiece monomer, K_d 5 x 10^7 M^{-1} (14), versus the tight binding due to two headpieces in the intact repressor, K_d = 3.5

x 10^{13} M^{-1} (7). Hence, the headpiece is a suitable system for the study of both specific and nonspecific binding.

Analysis of the headpiece by a number of methods has revealed the following information about it and its interaction with DNA. First of all, the protein does make both specific and nonspecific interactions with DNA, and in a manner that is virtually, but not precisely, the same as when it is part of the intact repressor. In both, specific contacts are made by residues Tyr 7, Tyr 17, Tyr 47 and His 29, as shown by the effects of iodination on repressor binding (15), by NMR effects seen on headpiece binding to poly d(AT) (11, 25) and to a synthetic operator fragment (53), by UV titration of operator-containing DNA (29) and by fluorescence and circular dichroism (14, 57). Furthermore, these contacts are sensitive to ionic strength and disappear in salt concentrations of 120 mM or higher (9, 29, 57).

The protein by itself has been found to form a stable, though flexible, structure in solution. NMR studies of the headpiece have shown that it possesses the same overall structure, both when attached to the repressor and as a separate entity (10, 49, 64). Further, thermal melting curves have shown that this structure is stabilized by salt, with PO_4 ion being a more effective stabilizer than NaCl (58, 63, 67). Unfolding and renaturation of the protein, both by temperature and by change in pH, is a smooth and continuous process, with no specific melting temperature, indicating that the protein possesses a flexible structure unlike those of most globular proteins (49, 58, 63, 67), though circular dichroism studies have indicated

that the protein does in fact possess an extensive secondary structure (63). Recently, the headpiece has been studied by 2D NOE NMR techniques, and the resonances of all hydrogen atoms have been assigned. Based on this information, it was possible to determine the precise secondary structure and all close contacts occurring within the molecule, and by combining this information with the necessary molecular constraints required of a protein, a structure of the fragment in solution has been deduced (31, 69, 70, 71). In this model, an illustration of which is given in Reference 69, residues 6 to 15, 17 to 25 and 35 to 45, respectively, form three alpha helices, with bends at residue 16 and in in the region 26 to 34. A hydrophobic core is formed by residues Val 19, Val 20, Val 23, Val 24, Leu 6, Leu 45, Ala 10, Ala 41, Met 42 and Tyr 47. Tyrosine 7, Tyrosine 17 and His 29 point to one side of the structure, as expected from their known involvement in DNA binding. Also, as expected from sequence homology studies (61), helices one and two form the helix-turn-helix motif seen in the other repressor structures, and the overall structure is homologous to that of the lambda repressor (31).

3. Lac Headpiece Isolation

a. Earlier Protocols

At the time that this project was begun, there were several protocols in the literature for the isolation of the headpiece protein. Each had its own drawbacks in the preparation of material for crystallization purposes, but all taken together provided a basis for the protocol that was finally adopted. These procedures, their advantages and disadvantages will be discussed in turn, to show the steps leading to the strategy that was finally settled on, and its results.

The first of these was the original isolation procedure of Geisler and Weber (19). This consisted of digestion of repressor with protease in 1 M Tris.HCl pH 7.5, 30% glycerol buffer, followed by two sizing column steps, a Sephadex G-150 step and a Sephadex G-25 step. The strategy was to separate the headpiece and tether protein from the repressor, core and protease during the first step, and then to separate the headpiece from the tether during the second. The procedure has several problems. The first of these is the rather crude separation achieved by the choice of sizing resins, in that the headpiece is entirely included by the first and entirely excluded by the second, so that no separation of material of similar molecular weight, such as partially digested headpiece protein or similar molecular weight contamination in the repressor or protease preparations, is achieved. The second problem is the lack of any way to distinguish active headpiece from material that is intact but

inactive because of denaturation. The third problem is the production of a very dilute product, which ideally should be as concentrated as possible for use in crystallization trials. And the last difficulty is the lack of any scrupulous method of removing residual protease activity, critical for material that must be maintained in solution for long periods of time during crystallization.

An entirely different approach was taken by Buck et al. (11). Their procedure involved separation by use of the fact that active headpiece binds to phosphocellulose. Core protein and proteases such as trypsin and chymotrypsin are not retained under these conditions. Their protocol used a phosphate buffer gradient on a column of Whatman P-11 resin as their first and only separating step. Following this, they used a procedure, first developed by Otsuka and Price (46) to stabilize DNase I preparations, in which the material is passed through a Sepharose column to which lima bean trypsin inhibitor protein has been linked covalently, making a specific protease affinity column. As used by Otsuka and Price, the column made stable for days DNase I preps, which by conventional purification methods lost activity in solution after only a few hours because of residual protease activity. Using this strategy, the final product of Buck et al. then represents a fraction of stable, purified binding activity, consisting mainly of the headpiece protein. The problem with this scheme is the lack of a size separation step, so that any residual repressor protein or any other contaminants that bind to the ion-exchange resin under these conditions also would be present in their final product. And, as with the procedure of Geisler and Weber (19),

the product would be dilute as well.

A third procedure was used by Arndt et al. (5), which combines many of the strengths of the other two into a more lengthy and rigorous purification scheme. Following digestion, the material was first separated on phosphocellulose with a step gradient, giving a concentrated fraction containing the binding activity. This then was fractionated on Sephadex G-75 to separate out the headpiece from other size contamination present, and, finally, the material was concentrated by loading and eluting it on another phosphocellulose column. The result was a highly purified protein possessing full binding activity.

The only drawback was the lack of a step to remove the residual protease activity.

The protocol that was finally settled on takes advantage of the best features of all these methods to produce a product that is of high purity, possesses full DNA binding activity as assayed by its ability to bind to phosphocellulose, is highly concentrated, and is scrupulously protease-free. In this procedure, the proteolytic digest is first passed through a lima bean trypsin inhibitor Sepharose affinity column while still protected in Tris/glycerol buffer, in order to remove the bulk of the active protease before this protection is lost. Then it is fractionated on phosphocellulose in a step gradient, to remove most of the nonbinding protein present, followed by passage through a second lima bean trypsin inhibitor column to remove the last traces of protease activity. The material is then size-fractionated on Sephadex G-50 to remove any residual repressor

and any other higher or lower molecular weight material that might have bound to the ion exchange column, and finally is concentrated by passage through a second phosphocellulose column. The results are described below.

b. Experimental: Repressor Isolation

Lac repressor is isolated from the overproducing strain BMH74-14 (i^{ql}, Reference 39). This strain contains the gene for the repressor protein coupled to a lysogenic phage lambda expression vector system. Induction of the phage by temperature shock causes the cells to express the gene for the repressor and to produce the protein in high quantities. The cells are grown in LB broth, which consists of 10 g Bacto-tryptone, 5 g Bacto-yeast extract (both Difco) and 10 g NaCl per liter, adjusted to pH 7.5 with NaOH (35). The bacteria are grown in 100 liters of medium starting from a two-liter overnight starter culture. The cells are initially grown at the permissive temperature of 32°C, under which the vector is not expressed, to an A_{600} of 1.2 OD. The culture is then heated quickly to 45°C for fifteen minutes, which causes induction of the vector, resulting in production of the protein at high levels in the cells, followed by growth at 35°C for four hours to allow the repressor gene to be expressed. Cells are isolated by centrifugation and then quick-frozen with liquid nitrogen and stored at -70°C until use. A 100-liter fermenter run typically yields 300 grams of wet cell paste.

The repressor protein is isolated by the method of Rosenberg et al. (52) with minor modifications. The 300-gram frozen cell pellet

is broken into small pieces and lysed by addition to 200 ml of CSB buffer (0.2 M Tris.HCl, 0.2 M KCl, 10 mM $MgAc_2$, 0.3 mM dithoithreitol (DTT), 5% (w/w) glucose, 0.1 mM orthonitrophenylfucose (ONPF) and 0.05 mg/ml phenylmethane sulfonyl fluoride (PMSF), pH 7.75 at 4^oC) containing 10 mg of DNase I and 10 mg of hen egg white lysozyme, at 4^oC. The lysate is further broken up by mixing for brief intervals in a blender, then allowed to stand for an hour and a half, with additional blending every thirty minutes; the mixture is thin and homogeneous at the end of this period. The pH is readjusted to 7.5 with KOH, and the solution spun for two hours at 15,000 rpm at 4^oC in a Beckman Type 19 rotor. The clear, greenish supernatant is decanted off, and solid ammonium sulfate added to it to 33% saturation. The mixture is allowed to precipitate overnight at 4^oC with gentle stirring.

The precipitate is spun for two hours at 8,000 rpm at 4^oC in a Sorvall GS-3 rotor. The pelleted material is resuspended in 100 ml of 0.075 KPG (75 mM KPO_4, 0.3 mM DTT, 0.1 mM ethylenediamine tetraacetic acid (EDTA), 5% glucose (w/w) and 0.05 mg/ml PMSF, pH 7.4 at 4^oC) and dialyzed overnight against 6 liters of 0.075 KPG; all at 4^oC.

The slightly cloudy dialysate is diluted as necessary with 0.0 KPG (0.075 KPG buffer without KPO_4) until its conductivity matches that of 0.075 KPG, and then spun for 3.5 hours at 35,000 rpm at 4^oC in a Beckman Type 35 rotor. The clear supernatant is loaded onto a 400 ml bed volume column of P-11 phosphocellulose ion exchange resin (Whatman), which is first equilibrated by washing with several volumes of 0.075 KPG. After loading, the column is washed with 0.075 KPG

until the A_{280} is reduced to 0.1 AU, and then the repressor is eluted with a 1500 ml linear gradient of 0.075 to 0.4 KPG buffer. The column is monitored by UV absorption, and the peak fractions assayed for isopropyl thiogalactoside (IPTG) binding activity (51), before pooling based on the latter assay. The protein is then precipitated by addition of 0.1 KPG saturated with ammonium sulfate to 40% saturation, allowed to sit for one hour at $4^{o}C$ with gentle stirring, and then pelleted for one hour at 8,000 rpm at $4^{o}C$ in a Sorvall GS-3 rotor. The pelleted material is resuspended in 1 M Tris.HCl, pH 7.5, 30% (vol/vol) glycerol, 1 mM EDTA and 0.3 mM DTT and dialyzed overnight against one liter of the same, all at $4^{o}C$. The final product is aliquotted into 1.8 ml portions and quick-frozen in liquid nitrogen, for storage until needed. Typical yields are 300 mg of final product at approximately 15 mg/ml.

c. Lac Headpiece Isolation

Lac headpiece is isolated by a combined protocol of Buck et al. (11) and of Arndt et al. (5). Alpha-chymotrypsin at approximately 7 mg/ml in 1 M Tris.HCl pH 7.5, 30% (vol/vol) glycerol, 1 mM EDTA, 0.3 mM DTT and 1 mM toslylysine chloromethyl ketone (TLCK) is preincubated for twenty minutes at $20^{o}C$, in order to eliminate any tryptic activity present in the protease. It is then added to lac repressor protein, isolated as described above, in the same final buffer plus 0.25 mM TLCK, to a level of 1% (wt chymotrypsin/wt repressor). The solution is allowed to incubate for three hours at $20^{o}C$ with slow end-over-end mixing in a polypropylene tube. At time, PMSF is added to 1 mM, and

the digest allowed to incubate for an additional twenty minutes.

The digest is passed through a 9 ml volume of lima bean trypsin inhibitor-Sepharose, prepared as described by Otsuka and Price, in Tris/glycerol buffer at 20°C. The eluted peak fractions are cooled to 4°C and diluted by slow addition of 12.5 volumes of 0.05 K buffer (50 mM KPO_4 pH 7.4, 1 mM EDTA, 0.1 mM DTT, 0.1 mM PMSF). The sample is then loaded onto a small P-11 column (about 1.5 times the starting volume of protein), which is pre-equilibrated with the same 1:12.5 mixed buffer. The column is washed with about five more volumes of the buffer, then eluted with 0.6 K buffer (0.6 M KPO_4 pH 7.4, the rest as above), and the peak fractions of the step gradient pooled. This peak is passed through a 2 ml volume lima bean trypsin inhibitor-Sepharose and the peak isolated; all at 4°C.

The peak is next loaded onto a 2.5 by 100 cm column of Sephadex G-50 Fine (Pharmacia), which has been equilibrated with 0.06 K buffer at 4°C, and fractionated. This yields three peaks: a leading peak, containing the fully excluded material; a trailing peak of the fully included matter; and a middle peak, centered in the sizing range of the resin, containing lac headpiece (Figure 1). The pooled fractions of this peak are loaded onto a 15 ml column of P-11 equilibrated with 0.06 K buffer, also at 4°C, washed with several more volumes of the buffer, and then eluted with 0.6 K buffer in a step gradient. The final material is quick-frozen in liquid nitrogen and stored in lN_2.

To concentrate the material further and to put it into a buffer suitable for crystallization, the protein solution is dialyzed overnight against 1 liter of 100 mM NaCl, 10 mM Na cacodylate pH 7.5,

0.1 mM EDTA, 0.1 mM DTT and 0.1 mM PMSF at 4°C, in Spectrapor 6, 1,000 MWCO dialysis tubing (Spectrapor Medical Industries), and then dialyzed against the same buffer containing 25% (w/vol) polyethylene glycol (PEG) 20,000, until the desired concentration is reached. The final material is dialyzed against one more liter of the original buffer without PEG and finally aliquotted into 100 microliter portions and quick-frozen in liquid nitrogen until use. Typical yields from a 300 mg repressor prep are 10-15 mg of final headpiece protein at 6 to 7 mg/ml.

The course of the purification can be seen in the gel shown in Figure 2. The original repressor is shown in lane 1, and all subsequent steps are shown in proportional volume loadings, except as noted. Lane 2 shows the material after digestion, which appears to be essentially complete. Column 3 shows the digest after passage through the lima bean trypsin inhibitor-Sepharose column, where it can be seen that none of the headpiece protein has been retained.

Lanes 4 and 5 show the load/wash and high-salt elution peaks, respectively, of the first P-11 column step. The core and most of the other contaminant peaks are seen to have been removed by this step, although some high molecular weight contaminants still remain. The headpiece itself also seems to have been separated in good yield, although some apparently has washed through; this may be inactive material, or simply some that was not retained on the column under the conditions necessary to pass the core.

Lanes 6 and 7 show the fully excluded volume and the headpiece fractions, respectively, of the Sephadex G-50 column; these

correspond to peaks 1 and 2 of Figure 1. While some headpiece has eluted with the excluded volume, due to the width of the peak on this resin, virtually all has come through as the second peak, and all of the high molecular weight contaminants have been removed.

Lane 8 shows the high-salt elution peak of the second P-11 column. This shows that the activity of the fraction isolated has been preserved, and that the yield is still high; this was also confirmed by monitoring the UV absorbance of the load/wash step.

Lane 9 shows the final, concentrated material after dialysis against PEG 20,000. Once again, no substantial losses are seen.

The activity and stability of the isolated headpiece protein are demonstrated by the following results. First, active headpiece is known to bind to P-11 cellulose phosphate under the conditions used in the prep, whereas inactive headpiece does not (5); therefore, the material, at least as far at the final dialysis step, is by this criterion, active headpiece. Secondly, stability was confirmed by incubation of a 2.5 times proportional volume load of the product at $4^{o}C$ for 14 days in the final storage buffer; this sample was then run on the gel shown in Figure 2 as lane 10. The band is at the expected intensity on the gel, and furthermore, no lower molecular weight bands representing partially digested headpiece are seen on this gel system, which is able to resolve material as small as 1,400 dalton as a sharp band.

4. Crystallization Trials

a. Background

The growth of large, single crystals of a material of suitable quality for diffraction analysis is the essential first step in any x-ray crystallographic study. While this is a rather obvious point, it is a crucial one for the course of any such analysis, because for very many compounds that have been taken up for study by crystallographers over the years, the analysis has ended at this stage, often after many months or even years of effort, because of a failure to produce the necessary crystals. The variable and seemingly random nature by which some macromolecules will produce crystals and others will not is a well-known phenomenon and has shrouded the crystallization process with an air of mystery and mumbo-jumbo that it probably does not deserve. For this reason, a discussion of the theory and practice of crystallization will be given, followed by a description of which methods were tried on the two molecules being studied and what results were obtained. This will serve two purposes: first, to explain the reason behind the particular systems that were tried, and, second, to serve as a guide to others who might work on this project as to what has been tried and the results of these trials.

The crystallization of a macromolecule is the process of causing a solution of that molecule to go into the solid state in the form of a stable, repeating lattice of the molecules in space. This process has been discussed from both a theoretical and a practical point of

view by McPherson (37), who considers it in terms of the thermodynamics of a transition from the solution to the crystalline state. In this view, whether a macromolecule is in solution or not depends on the overall free energy of the system under a given set of conditions; it is only when these conditions favor interactions between solute molecules over solute-solvent interactions that the transition to the solid state will occur. Whether this transition results in well-ordered crystals or simply amorphous precipitate depends on the nature of these solute-solute interactions, and McPherson discusses them in terms of the effects of counterions, ligands and various other parameters, as they may affect the overall free energy of the crystal lattice. These considerations favor the crystalline state as that of minimum free energy; but it is up to the nature of the components of the system to determine whether this minimum can actually be achieved.

Another way to consider this problem is in terms of how well the components of the system are able to form a repeating lattice. As in a brick wall, the most stable lattice can be formed from molecules which, like bricks, are sufficiently rigid and uniform in shape to form a regular, repeating structure. This implies that the macromolecules must be able to interact with each other in a regular, stable manner so that they form an array that possesses precise translational repeats along three orthogonal axes in space. Therefore, macromolecules which do not possess a rigid, regular structure, or that are not capable of forming regular contacts that are conducive to the establishment of such an array, will not be able

to form good crystals. For this reason, it is probably not a coincidence that the proteins that give the most stable, high-resolution crystal forms are proteins such as the various proteases, many of their macromolecular inhibitors, and a host of other largely secretory proteins that possess very rigid, usually highly crosslinked structures. Conversely, a number of other proteins, such as recA protein (D Goodsell, personal communication), which otherwise might serve as very stable building blocks for crystal lattices, do not do so in practice because they naturally interact to form complexes that do not possess a suitable translational repeat. As will be discussed, this requirement of a rigid, stable structure for lattice formation is probably a major reason for the failure of the lac headpiece protein itself to crystallize, while in complex with its DNA operator site it appeared to give quite different results.

From a practical point of view, finding the right crystallization conditions for a macromolecule generally takes on the strategy of starting with a general set of conditions under which the molecule will come out of solution, and then refining these to see if they can be made to bring about the desired crystal growth. The first step is to find a precipitant, or some condition under which the molecule will come out of solution. This can be as simple as concentration of the solution by dialysis or even by evaporation, but the usual procedure is to use a precipitating agent. As set forth by McPherson, these fall into two general categories: organic precipitants, ranging from mild solvents such as 2-methyl-2,4-pentanediol (MPD), to such heavy-duty organic solvents as acetone or toluene; and salts, usually

polyvalent, always very concentrated, with ammonium sulfate, ammonium phosphate and ammonium citrate being three very popular and successful examples. Their purpose is to promote interaction between molecules by lowering the dielectric constant of the medium and/or by helping to crosslink the molecules together. The ideal precipitant must achieve these objectives while not denaturing the macromolecule in the process. The general procedure for testing them is to perform an initial precipitation trial of the macromolecule with a number of them, and, based on the results, to select as many as can be tested in further crystallization trials.

Once a series of precipitants has been selected, the next step is to choose a crystallization method. Crystallization requires that some parameter or parameters of the system can be varied in a controlled manner, so that the macromolecule can be brought to the proper state of precipitation to allow crystallization to occur preferentially over amorphous precipitation. Again, a number of methods exist, ranging from crude bucket methods of adding precipitant to large amounts of material, to a number of very sophisticated microscale techniques that can be used when only small amounts of the material are available (which is almost always the case). These latter methods generally fall into one of two categories of how the conditions are varied: namely, by vapor phase equilibration or by dialysis. The former methods all work by equilibrating a small sample of the macromolecule solution, usually a drop, against a larger reservoir of solvent at a different concentration, all in a sealed system. Over time, the activity of the macromolecule-containing

solution becomes equal to that of the solvent reservoir, and, by judicious control of the starting activities of the two solutions, the sample can be taken very slowly through the desired activity range for crystallization to occur. Taking such forms as the spot plate and the hanging drop, these have been very successful in producing crystals of compounds in a number of cases. Their advantages are that they are easy to set up and maintain, and in suitable configurations allow for easy examination of the solutions during crystallization. Their main disadvantage is that they allow only volatile components of the system to be varied, leaving salts and other nonvolatile components of the system unchanged. In addition, volume change is a necessary result of equilibration, with the result that the concentrations of all of the nonvolatile components of the system change over the course of the crystallization, making control of the individual components impossible.

A different technique that overcomes these disvantages is microdialysis, in which the macromolecule solution is dialyzed in a small cell against a larger reservoir of precipitant. Again, by judicious choice of the two solutions, very precise control of the conditions in the sample can be achieved. The main advantage is that any of the components small enough to pass through the membrane can be varied at will, while the volume of the sample can be maintained constant. This feature allows precise control of these conditions during the course of the crystallization. The disadvantages are that setup is in general more difficult, and that examination of the sample is usually much more difficult than with other methods — no small

consideration when one is searching for microcrystals.

The general wisdom is that once a set of crystallization conditions has been found for a material, any of these techniques can be used to produce crystals by suitable adjustment of the conditions. However, in practice, it turns out that certain techniques often produce bigger or better crystals, or produce them faster or more repeatably than do others. And so, it is usually necessary to try a series of techniques with a given set of conditions until the best ones for the particular compound can be found.

Once a precipitant and a methodology have been selected, the final step is to vary the system parameters until crystallization is achieved. Besides the two major parameters already discussed, there are two others that need to be varied: the presence of counterions and other ligands in the solution, and the solution pH. The former can include monovalent, divalent and higher anions and cations, which may stabilize the structure, aid in forming crosslinks between adjacent molecules in the crystal lattice, and so on; stabilizers, such as reducing agents, antimicrobials, and chelating agents, which help to maintain the integrity of the macromolecule; various ligands that the macromolecule may bind to specifically, whose presence may also stabilize the compound's structure; and, frequently, other components that have no obvious role in the crystallization process at all, but which may help to promote, or even be essential for, crystallization to occur. The latter parameter, the solution pH, is almost always a critical variable for crytallization, as most macromolecules will crystallize only over a certain range of pH

values, often because they frequently are active and/or stable only over this range. Ideally, the range of stability and the range of crystallizability should overlap, but occasionally this is not the case.

As the range of possible combinations of these variables is essentially infinite, the screening process for the right combination of parameters necessary to produce crystals can be endless or at least frequently seems to be. The one saving grace of this is that, in general, crystallization can be achieved in steps, by a series of successive refinements on a promising starting set of precipitating agents and a method of crystallization by systematic variation of the various parameters. The nightmare case of there being one and only one set of conditions under which the macromolecule gives crystals, all others giving no result, seems to be more the exception than the rule. Nevertheless, the right conditions may be sufficiently stringent that they will be found only after long and arduous searching. The best strategy is to try as many initial conditions as possible, and then based on the results to refine the system parameters of those giving the most promising signs until suitable crystals are produced. The important thing, and the one that probably is the reason that at least some macromolecules are never successfully crystallized, is to keep an open mind on which sets of conditions are pursued and not to prejudice the choice of these conditions prematurely.

b. Headpiece Crystallization Trials

As has been described, the lac headpiece is a small protein, consisting of a loose association of three helical units that probably are held together by rather weak interhelical interactions only. The observations indicating that it can be continuously denatured and then renatured into what seems to be a well-defined structure by a variety of methods suggests that suitable crystallization conditions would be those that are mild and stabilizing. Therefore, high-salt conditions, which have been found to be necessary to preserve headpiece structural integrity in solution, or ones involving only mildly denaturing organic precipitants, would seem to be the best options to explore for crystallization.

A series of mild organic precipitants and salts were tested for their ability to precipitate the protein; the results of this screen are listed in Table 1. All of the organic precipitants tried, which were MPD, several different sizes of polyethylene glycol (PEG) and ethanol, were able to cause precipitation, either as amorphous precipitate or as a fine oil; with one of the PEGs, PEG 600, the precipitate in fact appeared microcrystalline. Of the salts tried, all of the trivalent salts induced precipitation, while none of the monovalent or divalent salts would do so. Based on these results, MPD, PEG, ammonium sulfate, ammonium phosphate and ammonium citrate were chosen for use in subsequent trials.

Crystallization trials were carried out using three microtechniques: vapor diffusion in spot plates; vapor diffusion in hanging drops; and microdialysis in Zeppezauer tubes. The various

setups tried and the specific conditions employed, along with the results of each of these trials, are described in detail in Table 2. The results of these screens can be summarized as follows:

1) MPD: The protein either oiled out of solution or precipitated under all conditions tried. Several trials at pH 7.0 using various salts as counterions occasionally gave what looked like microcrystals, but nothing that was reproducible. This series was not extensively pursued.

2) PEG 600: The protein either remained in solution or formed oils in the initial pH screens. This series was also not extensively pursued.

3) Ammonium Sulfate: Trials at pH 5 gave nothing but oils and were not pursued further. Trials at pH 7 and 9 gave fibrous microcrystals after first oiling out, and a number of screens in this pH range were tried. Using both microdialysis and vapor diffusion techniques, these latter trials gave only oils along with several varieties of patchy, multinucleated microcrystals, both using the salt alone and in conjunction with a series of other monovalent, divalent, and trivalent salts.

4) Ammonium Phosphate: In the initial pH screens, trials with the protein using the salt alone gave oils over the entire pH range 5 to 9. A number of screens were then tried using the salt in conjunction with other salts, as with the ammonium sulfate series; these trials gave only oils or multinucleated fibers with all conditions tried.

5) Ammonium Citrate: Again, initial screens at pH 5 produced only oils, and so this pH range was not extensively pursued. At pH 7 and 9, trials with the salt alone gave precipitates, but in conjunction with a number of other salts, a several of these trials gave rather similar results consisting of the appearance of numerous microcrystals along with a characteristic "skinning over" of the surface of the spots in vapor diffusion trials. These were the most promising results achieved with lac headpiece, and this series was the most extensively pursued of all. However, no large crystals ever resulted using all of the conditions tried.

c. Headpiece/Operator Crystallization Trials

Trials were also carried out on the protein complexed with a 21 base-pair synthetic operator site of sequence A-A-T-T-G-T-G-A-G-C-G-G-A-T-A-A-C-A-A-T-T. This sequence is a pseudopalindrome, in which the first and last six base-pairs form a perfect palindrome, and the central nine a partial palindrome between bases 8 and 14 and 10 and 12 (numbering from left to right). As described above, numerous studies have indicated that the headpieces make contact with the operator at the two palindromic ends, so that there would be two headpieces bound per operator. Headpiece binding is known to be abolished by high salt, and so crystallization trials were designed that used the organic precipitants MPD and PEG in the presence of low to moderate amounts of salt exclusively.

The results of these trials are shown in Table 3. The initial results with PEG were encouraging, but further trials with it were

suspended after crystals were found on the first screen with MPD. These crystals were rectangular, and characteristically twinned, with two crystals growing end-on-end and inclined at about ten degrees to each other with respect to the long axis of each, making them look like boomerangs. Each half would light and extinguish as a single crystal and was usually brightly colored under polarized light, probably due to birefringence. The largest of these measured up to 0.3 mm long but less than 0.05 mm wide. In addition to this form, a second, much smaller but apparently untwinned form of the crystals was also seen. These were characteristically cubic chunks, which also would light and extinguish singly and which were also colored under polarized light. Unfortunately, the largest of these were less than 0.05 mm on edge, and no larger ones could be produced.

Attempts were made to take x-ray photographs of these crystals to see if they were ordered or not. The crystals were extremely fragile, and it was not possible to mount them without causing at least some damage during the mounting and drying procedures. The largest of the twinned form of the crystals were mounted in glass capillaries, and several of these that could be mounted without too extensive damage were photographed on an Eliott GX-6 rotating anode x-ray generator producing copper radiation. Unfortunately, all of these crystals melted before they could be photographed, despite efforts to cool them using a cold stream device. Efforts to produce more of these crystals failed before the project was finally abandoned.

5. Conclusions

This chapter has been written with two main objectives in mind. The first is to describe the work that has been done in an attempt to carry out a structural analysis of the lac headpiece, both by itself and in complex with its DNA binding site. The second is to act as a bridge between the previous stage of this work, involving the original attempts to study the whole repressor, and the next stage, which will involve the structural solution of the headpiece in complex with some variant of its operator site. For this reason, a few points will be discussed about the results achieved here and a few suggestions made for future efforts, as a guide to the next generation of workers on this project.

First, the work has shown that it is possible to isolate the headpiece protein as a stable species in a form pure enough and in quantities sufficient for use in crystallization trials. The prep, however, is long and tedious, and the quantities of material isolated in the end are smaller than one would ideally like to achieve, in fact, so much so that this was very much the limiting step in the number of crystallization trials that could be carried out with the protein. Much of this is simply a consequence of the fact that the headpiece is such a small part of the whole repressor, with the result that even the maximum theoretical yield of lac headpiece is only one seventh of that by weight of the starting material. When losses inevitable in each step of the purification are included, the final yield is considerably less than that. A few changes in the isolation

protocol might lead to better results. One would be the inclusion of the final sizing step in the repressor isolation protocol. The crude repressor used as starting material here still contains some contaminants; while these are removed in subsequent steps, some traces, which might affect attempts at crystallization, may still remain in the final headpiece product.

Another would be the use of a different protease for digestion of the repressor. Chymotrypsin was chosen here for two reasons. The first is that it makes only one cut in the repressor, giving a homogeneous headpiece product. Other proteases, such as trypsin, make multiple cuts in the tether region, giving mixed products that are difficult to separate. The second is that chymotrypsin can be removed completely from the prep by means of the lima bean affinity column used; other proteases that also give single cuts, such as clostripain and papain, cannot, making it more difficult to assure that the final product is in fact stable and protease-free. Unfortunately, for some reason chymotrypsin seems to lead to smaller yields of headpiece than do these other enzymes, possibly due to some additional proteolysis of the headpiece peculiar to this enzyme (F. Buck, personal communication). And so, the finding of some other protease for use in the digestion step, preferably one for which an affinity column or some other means of efficiently removing it could be used, might result in higher yields of the headpiece protein.

A third improvement might be to circumvent the proteolysis step altogether by constructing a plasmid vector containing the gene for direct synthesis of the headpiece. Since the lac headpiece is the N

terminus of the repressor, the protein synthesized by this gene would fold itself correctly during synthesis and would be active in DNA binding. By cloning multiple copies of the gene into a high-efficiency expression vector and inserting the vector into a suitable host, it would be possible to isolate the protein in high yield, using a much faster combined prep. The technology to do so by genetic engineering techniques is well developed, and the potential benefits of being able to produce the protein by this shortcut method are obvious.

The second point is the question of whether the headpiece protein can be crystallized, and how this might be done. The crystallization screens described all yielded negative results, though some of the precipitants tried, particularly ammonium citrate, gave preliminary results that were rather encouraging. The problem seems to be that the protein by itself does not possess the rigid, compact structure of proteins that do crystallize well, such as the various proteases and inhibitors, many of the metabolic enzymes such as the cytochromes, and so on, which can pack together to form tight lattices because of this internal rigidity. Instead, it is a fairly loose association of three helices, held together by only a few weak hydrophobic contacts, which can unfold and refold very easily, making the structure very unstable and the protein a poor candidate for crystallization. What is needed is something that will stabilize the headpiece structure enough to allow it to form a tight lattice. Some of the counterions tried seemed to give better results in forming microcrystals than did others, and further effort in crystallization seems in order. One

promising counterion would be Tris buffer, which is the stabilizing agent of the protein in the isolation procedure. Geisler and Weber (19) showed that the protective effect of their Tris/glycerol buffer system against proteolysis was due specifically to the presence of the Tris cation. It is possible that in a similar system, using one of the organic precipitants such as MPD in place of glycerol, this stabilization might make the protein rigid enough to allow it to form a crystal. This counterion was never tried with the headpiece because of known problems of Tris buffer's crystallizing out in systems using organic precipitants (M. L. Kopka, personal communication). However, it has been used in other crystallization systems successfully (Reference 37) and might be pursued at some point in the future.

A third and final point, probably the one of greatest interest, is that of cocrystallization of the protein with some form of its DNA binding site. Preliminary results with the 21mer operator using MPD have been extremely encouraging, and we hope that some refinement of conditions with this system would produce crystals large enough and stable enough for diffraction analysis. Another approach would be to vary the sequence of the DNA binding site, with the aim of making an oligomer which would pack to form a strong lattice, allowing the protein to "go along for the ride" on the nucleic acid. Impressive results have been achieved using this approach by Pabo and coworkers with the lambda headpiece. They found that either a 17 or a 23 base-pair operator site in complex with their protein gave only microcrystals, while a 20mer gave large crystals which diffracted to better than 2.5 Angstrom resolution. Their analysis of these crystals

showed that the DNA helices in it pack end-on-end to form continuous helices throughout the crystal, and that at 20 base-pairs the repeat is in phase with the ten base-pairs per turn of B-DNA, allowing the repeat to match the crystal lattice (Reference 28). The 21mer, by this criterion, would be one base-pair too long for ideal B-form DNA; therefore, the use of a sequence one base-pair shorter might yield better results.

Another problem with the 21mer being used is the location of the headpiece binding sites on the helix. As described above, the headpieces are thought to bind to the six base pairs of either end of the 21mer. Therefore, the presence of the headpiece protein bound to these sites on the DNA might interfere with the establishment of this end-on-end packing in the crystal. In addition, the fact that there are actually two headpiece binding sites per DNA helix makes the combined structure a complex between three independent components, which is quite complicated. A possible solution to this would be to use a shorter DNA sequence, containing only one headpiece binding site, which also could form the end-on-end stacking arrangement described. One such sequence might be A-C-A-A-T-T-G-T-G-C, which is a ten base-pair palindrome containing the A-A-T-T-G-T headpiece binding site as its center, surrounded by a two base overlapping end. This overlap could base-pair with adjacent helices, to form continuous ten base pair per turn helical repeats in a crystal lattice, as in the 20mer used by Pabo and coworkers (28). The basic unit would then contain one headpiece and one DNA binding site, making the structure an ideal size for crystallographic analysis.

References

1. Abermann, R., Bahl, C. P., Marians, K. J., Salpeter, M. M. and Wu, R. (1976) J. Mol. Biol. 100, 109-114.
2. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Mueller-Hill, B., Pfahl, M. and Schmitz, A. (1972) Nature 237, 322-327.
3. Anderson, J. E., Ptashne, M. and Harrison, S. C. (1985) Nature 316, 596-601.
4. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (1981) Nature 290, 754-758.
5. Arndt, K. T., Boschelli, F., Lu, P. and Miller, J. H. (1981) Biochem. 20, 6109-6118.
6. Bahl, C. P., Wu, R., Strawinsky, J. and Narang, S. A. (1977) Proc. Natl. Acad. Sci. USA 74, 966-970.
7. Barkley, M., Riggs, A., Jobe, A. and Bourgeois, S. (1975) Biochem. 14, 1700-1712.
8. Beckwith, J. and Zipser, D., eds. The Lactose Operon (Cold Spring Harbor Laboratory, New York, 1970).
9. Buck, F., Hahn, K-D., Zemann, W., Rueterjans, H., Sadler, J. R., Beyreuther, K., Kaptein, R., Scheek, R. and Hull, W. E. (1983) Eur. J. Biochem. 132, 321-327.
10. Buck, F., Ruterjans, H. and Beyreuther, K. (1978) FEBS Letts. 96, 335-338.
11. Buck, F., Rueterjans, H., Kaptein, R. and Beyreuther, K. (1980) Proc. Natl. Acad. Sci. USA 77, 5145-5148.

12. Butler, A., Revzin, A. and von Hippel, P. (1977) Biochem. 16, 4757-4768.

13. Chao, M., Gralla, J. and Martinson, H. (1980) Biochem. 19, 3254-3260.

14. Culard, F., Schnarr, M. and Maurizot, J. C. (1982) EMBO J. 1, 1405-1409.

15. Fanning, T. G. (1975) Biochem. 14, 2512-2520.

16. Farabaugh, P. (1978) Nature 274, 765-769.

17. Files, J. and Weber, K. (1976) J. Biol. Chem. 251, 3386-3391.

18. Geisler, N. and Weber, K. (1976) Proc. Natl. Acad. Sci. USA 73, 3103-3106.

19. Geisler, N. and Weber, K. (1977) Biochem. 16, 938-943.

20. Gilbert, W. and Maxam, A. (1973) Proc. Natl. Acad. Sci. USA 70, 3581-3584.

21. Gilbert, W. and Mueller-Hill, B. (1966) Proc. Natl. Acad. Sci. USA 56, 1891-1898.

22. Gilbert, W. and Mueller-Hill, B. (1967) Proc. Natl. Acad. Sci. USA 58, 2415-2421.

23. Heyneker, H. L., Schine, J., Goodman, H. M., Boyer, H. W., Rosenberg, J. R., Dickerson, R. E., Narang, S. H., Itakura, K., Lin, S-y. and Riggs, A. D. (1976) Nature 263, 748-752.

24. Hirsch, J. and Schlief, R. (1976) J. Mol. Biol. 108, 471-490.

25. Hogan, M., Wemmer, D., Bray, R. P., Wade-Jardetzky, N. and Jardetzky, O. (1979) FEBS Letts. 124, 202-203.

26. Jacob, F. and Monod, J. (1961) J. Mol. Biol. 3, 318-356.

27. Jobe, A. and Bourgeois, S. (1972) J. Mol. Biol. 69, 397-408.

28. Jordan, S. R. , Whitcombe, T. V., Berg, J. M. and Pabo, C. O. (1985) Science 230, 1383-1385.

29. Jovin, T. M., Geisler, N. and Weber, K. (1977) Nature 269, 668-672.

30. Kania, J. and Brown, D. T. (1976) Proc. Natl. Acad. Sci. USA 73, 3529-3533.

31. Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. and van Gunsteren, W. F. (1985) J. Mol. Biol. 182, 179-182.

32. Lin, S-y. and Riggs, A. (1970) Nature 228, 1184-1186.

33. Lin, S-y. and Riggs, A. (1971) Biochem. Biophys. Res. Comm. 45, 1542-1547.

34. Lin, S-y. and Riggs, A. (1975) Biochem. Biophys. Res. Comm. 62, 704-710.

35. Maniatis, T., Fritsch, E. F. and Sambrook, J. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York, 1982).

36. McKay, D. B. and Steitz, T. A. (1981) Nature 290, 744-749.

37. McPherson, A. (1976) Methods Biochem. Anal. 23, 249-345.

38. Miller, J. H., Ippen, K., Scaife, J. G. and Beckwith, J. R. (1968) J. Mol. Biol. 38, 413-420.

39. Mueller-Hill, B., Fanning, T., Geisler, N., Gho, D., Kania, J., Kathman, P., Meisner, G., Schlotman, M., Schmitz, A., Triesch, I. and Beyreuther, K. Symposium on Protein-Ligand Interactions

(Walter de Gruyter, Berlin, 1974).

40. Ogata, R. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 4973-4976.

41. Ogata, R. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 5851-5854.

42. Ogata, R. and Gilbert, W. (1979) J. Mol. Biol. 132, 709-728.

43. Ohlendorf, D. H., Anderson, W. F., Fisher, R. F., Takeda, Y. and Matthews, B. W. (1982) Nature 298, 718-723.

44. Ohlendorf, D. H., Anderson, W. F., Lewis, M., Pabo, C. O. and Matthews, B. W. (1983) J. Mol. Biol. 169, 757-769.

45. Ohshima, Y., Horiuchi, T. and Yanagida, M. (1975) J. Mol. Biol. 91, 515-519.

46. Otsuka, A. and Price, P. (1974) Anal. Biochem. 62, 180-187.

47. Pabo, C. O. and Lewis, M. (1981) Nature 298, 443-447.

48. Platt, T., Files, J. and Weber, K. (1973) J. Biol. Chem. 248, 110-121.

49. Ribiero, A. A., Wemmer, D., Bray, R. P , Wade-Jardetzky, N. G. and Jardetzky, O. (1981) Biochem. 20, 823-829.

50. Richmond, T. and Steitz, T. (1976) J. Mol. Biol. 103, 25-38.

51. Riggs, A. D. and Bourgeois, S. (1968) J. Mol. Biol. 34, 361-364.

52. Rosenberg, J. M., Kallai, O. B., Kopka, M. L., Dickerson, R. E. and Riggs, A. D. (1977) Nucl. Acids Res. 4, 567-572.

53. Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom J. H., Kaptein, R., Rueterjans, H. and Beyreuther, K. (1983) Biochem. 22, 228-235.

54. Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. and Sigler, P. B. (1985) Nature 317, 782-786.

55. Schlotmann, M., Beyreuther, K., Geisler, N. and Mueller-Hill, B. (1975) Biochem. Soc. Trans. 3, 1123-1124.

56. Schmitz, A. and Galas, D. J. (1979) Nucl. Acids Res. 6, 111-137.

57. Schnarr, M., Durand, M. and Maurizot, J-C. (1983) Biochem. 22, 3563-3570.

58. Schnarr, M. and Maurizot, J-C. (1982) Biochim. Biophys. A. 702, 155-162.

59. Steitz, T. A., Richmond, T. J., Wise, D. and Engelman, D. (1974) Proc. Natl. Acad. Sci. USA 71, 593-597.

60. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. and Matthews, B. W. (1982) Proc. Natl. Acad. Sci. USA 79, 3097-3100.

61. Takeda, Y., Ohlendorf, D. H., Anderson, W. F. and Matthews, B. W. (1983) Science 221, 1020-1026.

62. Tanaka, I., Appelt, K., Dijk, J., White, S. W. and Wilson, K. S. (1984) Nature 310, 376-381.

63. Vucelic, V., Vucelic, R., Bray, P. and Jardetzky, O. (1981) FEBS Letters124, 204-206.

64. Wade-Jardetzky, N., Bray, R. P., Conover, W. W., Jardetzky, O., Geisler, N. and Weber, K. (1979) J. Mol. Biol. 128, 259-264.

65. Wang, J., Barkley, M. and Bourgeois, S. (1974) Nature 251, 247-249.

66. Weber, K., Files, J. G., Platt, T., Ganem, D. and Miller, J. H. in Protein-Ligand Interactions, Sund H and Blauer G, eds. (Walter de Gruyter, New York, 1975).

67. Wemmer, D., Ribiero, A. A., Bray, R. P., Wade-Jardetzky, N. G. and Jardetzky, O. (1981) Biochem. 20, 829-833.

68. Zingsheim, H. P., Geisler, N., Weber, K. and Mayer, F. (1977) J. Mol. Biol. 115, 565-570.

69. Zuiderweg, E. R. P., Billeter, M., Boelens, R., Scheek, R. M., Wuethrich, K. and Kaptein, R. (1984) FEBS Letters 174, 243-247.

70. Zuiderweg, E. R. P., Kaptein, R. and Wuethrich, K. (1983) Proc. Natl. Acad. Sci. USA 80, 5837-5841.

71. Zuiderweg, E. R. P., Scheek, R. M. and Kaptein, R. (1985) Biopolymers 24, 2257-2277.

Table One: Headpiece Precipitant Trials

Trials were performed by placing 10 microliters of headpiece protein at a concentration of 6-7 mg/ml in a buffer containing 100 mM KCl, 10 mM Na cacodylate pH 7.5, 0.1 mM dithiothreitol, 0.1 mM phenylmethane sulfonyl chloride, 0.02% Na azide in a depression slide, and adding the precipitant at the given starting concentration while watching the spot under a dissecting microscope until the solution turned cloudy.

1) 2-methyl-2,4-pentane diol (MPD, 100%): clear at 40% MPD, cloudy at 50%; oily precipitate.

2) Polyethylene glycol 600 (PEG 600, 40%): slightly cloudy at 8%, cloudy at 12%; precipitate lights and extinguishes.

3) PEG 1000 (40%): clear at 4%, cloudy at 8%; oil or precipitate?

4) PEG 1500 (50%): clear at 5%, precipitate at 10%; precipitate does not light.

5) PEG 6000 (50%): clear at 5%, cloudy at 10%; brownish precipitate — oil?

6) Ethanol (100%): hazy at 20%, cloudy at 30%; brownish precipitate.

7) Ammonium acetate (5 M): clear through 3 M.

8) Ammonium citrate pH 9.0 (3 M): clear at 0.9 M, cloudy at 1.2M.

9) Ammonium formate (5 M): clear through 3 M.

10) Ammonium phosphate pH 8.5 (4 M): clear at 0.4 M, cloudy at 0.8 M; oily precipitate.

11) Ammonium sulfate pH 8.5 (3.7 M): clear at 0.6 M, cloudy at 1 M; oily precipitate.

12) Magnesium chloride (5 M): clear through 4 M.

13) Magnesium sulfate (2 M): clear at 1.2 M, cloudy 1.4 M; oily precipitate.

14) Sodium chloride (5 M): clear through 4 M.

Table Two: Headpiece Crystallization Trials

The following is a list of all crystallization trials with the lac headpiece protein and their results. Trials are grouped according to the precipitant used, followed by crystallization technique, and ending with the specific conditions of each trial. All trials were set up using a headpiece protein stock at an initial concentration of 6-7 mg/ml, containing 100 mM KCl, 10 mM Na cacodylate pH 7.5, 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM phenylmethane sulfonyl fluoride, and 0.02 % Na azide (HP buffer). Crystallizations in spot plates were prepared by adding concentrated solutions of precipitants and counterions to 10 microliter samples of the headpiece protein in HP buffer to give the initial starting levels of each component in each well. Those in hanging drops were prepared by adding 5 microliters of the reservoir solution to 5 microliters of the protein stock in HP buffer and suspending this over the reservior on a siliconized coverslip. Those in microdialysis tubes were prepared simply by adding 10 microliters of the protein stock to the tube and allowing the reservoir solution to diffuse in through the membrane, except for setups with PEG, where a low level of the precipitant was added directly to the tube as well.

I. 2-methyl-2,4-pentanediol (MPD)

A. Microdialysis; 20 % to 70 % in steps of 10 % every three days.

1) 100 mM ammonium phosphate pH 5.1
   Oil droplets on membrane at 20 %; precipitate at 60 % that may light.
2) 100 mM ammonium phosphate pH 7.2
   Amorphous precipitate on membrane at 60 % that may be protein.
3) 100 mM ammonium phosphate pH 9.1
   Amorphous precipitate on membrane at 70 % that may be protein.

B. Vapor diffusion (spot plates); 20 % to 100 % in steps of 10 % every seven days.

1) 0.85 HP buffer plus 50 mM Na cacodylate pH 7.5
   Dozens of specs by 40 % that light and extinguish. Precipitate ring around spot at 50 %. Spot started to precipitate at 70 %, was completely precipitated by 90 % with a brownish precipitate that did not light (oil?).
2) 0.85 HP buffer plus 0.4 M NaCl, 50 mM Na cacodylate

pH 7.5
Some specs floating on surface at 30 % that light and extinguish, that increased in number slightly up to 90 %, when protein oiled out.

3) 0.85 HP buffer plus 80 mM $MgCl_2$, 50 mM Na cacodylate pH 7.5
Specs seen on setup. Some small, cube-like crystals seen at 30 % that light and extinguish poorly but have good shape; these were gone at 40 %. Oiled out at 90 %. Thousands of microcrystals in spot at 100 %; probably salt crystals.

4) 0.85 HP buffer plus 0.4 M $MgCl_2$, 50 mM Na cacodylate pH 7.5
Specs seen on setup. Some small, cube-like crystals seen at 30 % that light and extinguish poorly and have poor shape; these were gone at 40 %. Salt crystals at 90 % (crystal violet test).

5) 0.85 HP buffer plus 125 mM Na_2SO_4, 50 mM Na cacodylate pH 7.5
Specs seen on setup. Salt crystal and haze in spot at 50 %. Spot increasingly precipitated up to 100 % with precipitate lighting faintly.

6) 0.85 HP buffer plus 100 mM KPO_4 pH 7.5
Some floating cubic forms at 30 % that light and extinguish poorly; gone at 40 %. Spot oiled out at 60 %. Precipitate in spot at 100 %.

II. Polyethylene glycol 600 (PEG 600)

A. Microdialysis; 8 % to 36 % in steps of 4 % every three days.

1) 100 mM ammonium phosphate pH 5.25
Membrane appeared grainy immediately (oil droplets?); remained so up to 36 %, when two multinucleated crystals appeared on the glass wall.

2) 100 mM ammonium phosphate pH 7.15
Nothing at all seen up through 36 %.

3) 100 mM ammonium phosphate pH 9.0
Nothing at all seen up through 36 %.

III. Ammonium sulfate

A. Microdialysis; 1 M to 4 M in steps of 0.2 M every three days.

1) 10 mM Na acetate pH 5.1
"Grainy" membrane at 1 M (oil drops?); resolved to hundreds of large and small oil drops on membrane and on meniscus by 1.2 M; a triangular chunk, possible crystal, in one drop at 2.2 M; some more chunks on membrane at 3.4 M; lots of these in oil drops at 3.6 M.
2) 10 mM Na cacodylate pH 7.1
"Grainy" membrane at 1 M; hundreds of large and small oil drops on membrane and meniscus at 1.2 M; a few multinucleated crystals on membrane at 2.4 M; some fibers and possible crystals at 3.2 M that seem to light and extinguish; crystals or precipitate at 3.6 M; hundreds of microcrystals at 4 M that light and extinguish.
3) 10 mM Na cacodylate pH 8.95
A number of large (up to 0.5 mm across), birefringent clumps at 1 M that light and extinguish weakly; some small possible crystals and oil drops on meniscus at 1.6 M; oil drops with small crystals in them, and a huge, amorphous clump in tube at 1.8 M; more extensive oiling by 2.2 M; microcrystalline-like debris at 2.4 M; several floating clumps at 3 M; several patches of material that light and extinguish radially on walls at 3.6 M.

B. Microdialysis; 0.6 M to 1.8 M in steps of 0.1 M every three days; then 1.8 to 3.4 M in 0.4 M steps every three days; then 3.6 to 4 M in 0.2 M steps every three days.

1) 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM Na azide pH 8.0
A few fibrous, sheet-like crystals on the membrane that may light and extinguish at 0.6 M; a clump at 0.8 M; a few small, multinucleated crystals that light and extinguish weakly at 1.0 M; oil on the walls and meniscus at 1.3 M; a few dozen small, tightly-formed oil drops on the membrane at 1.4 M; several multiple layered crystals on the walls that light at 1.6 M; patches that light and extinguish radially on walls at 3.4 M; several uniformly extinguishing, 0.2 mm crystals on walls at 3.8 M.
2) 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM Na azide pH 8.5
A few microcrystals that light at 0.6 M; one or two clumps that light on glass at 0.8 M; a few multinucleated crystals on the glass at 1 M; some tightly formed oil drops on the membrane and several poorly formed crystals on the walls at 1.2 M; light oiling on the walls and meniscus at 1.3 M; six to twelve small crystals that light and extinguish weakly, have poor shape at 1.4 M; several multinucleated crystals on the walls at 1.7 M; some

small material on the meniscus that lights at 2.2 M; the same but larger at 2.6 M; radially lighting patches on the walls and some microcrystals on the membrane at 3.4 M; large radially lighting and extinguishing patches on the walls at 3.8 M.

3) 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM Na azide pH 9.0

Numerous poorly formed crystals on walls and possibly on membrane that light and may extinguish at 0.6 M; three or four poorly formed clumps at 0.8 M; small, radially lighting and extinguishing patches on the walls at 0.9 M; numerous microcrystals on glass and some floating clumps at 1 M; increasing number of these at 1.2 M; oil on walls, meniscus and membrane, along with some poorly formed crystals on walls at 1.3 M; some floating clumps that light and extinguish at 1.4 M; numerous multinucleated crystals at 1.6 M; more patches on the walls, less well ordered than at lower pH, at 2.6 M.

4) 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM Na azide pH 9.5

Some clumps on the glass that do not light at 0.8 M; some sheet-like material on the membrane at 1.1 M; small oil drops on membrane at 1.2 M; extensive oiling at 1.3 M; heavy by 1.7 M; small patches on the walls at 3.8 M.

C. Microdialysis; 0.6 M to 3.95 M in steps of 0.2 M every three days

1) 50 mM ammonium phosphate pH 9.0

Hundreds of tiny specs that may light at 0.6 M; oil on membrane and some microcrystals on glass that light and extinguish multiply at 1 M; oil on walls and membrane at 1.2 M; a large fiber that lights and extinguishes, and several disordered microcrystals on wall, along with oiling throughout tube, at 1.6 M; radially lighting and extinguishing patches on walls at 3.4 M; hundreds of "starburst" crystals at 3.6 M.

2) 250 mM ammonium phosphate pH 9.0

Fibers and amorphous material on walls and membrane at 0.6 M; oil drops on membranes at 1 M; heavy oil drops on walls and in liquid at 1.2 M; tight oil drops and some microcrystals at 1.4 M; fibers and oil drops at 2.6 M; radially lighting and extinguishing patches at 3.4 M.

3) 50 mM $MgCl_2$ pH 9.0

Oil at membrane at 1 M; oil throughout liquid at 1.2 M, heavy by 1.4 M; poorly lighting "starburst" crystals at 3.6 M.

4) 250 mM $MgCl_2$ pH 9.0

Rectangular, multinucleated crystals on membrane that

might light and extinguish at 0.6 M; oil throughout solution at 1.2 M; heavy at 1.6 M; some poor crystals on walls and membrane at 1.8 M.

5) 50 mM KCl pH 9.0
   Several radially lighting and extinguishing patches on walls at 0.8 M; light oiling at 1.2 M; large, heavy oil drops on meniscus at 2.2 M.
6) 250 mM KCl pH 9.0
   Microcrystals on membrane at 0.6 M; oil droplets on membrane at 1 M; oil throughout solution at 1.2 M; heavy at 1.4 M; disordered blobs at 3.8 M.

D. Vapor diffusion (spot plates); 1.2 M to 2.5 M at 0.25 M every seven days, then 2.5 M to 4 M at 0.5 M every seven days.

1) 50 mM Na cacodylate pH 7.5
   A few particles that light and extinguish in sections at 1.25 M; some debris that does not light at 1.75 M; precipitate ring around spot at 2 M; lightly oiling in solution at 2.5 M; microcrystals at 3 M; some radially lighting and extinguishing patches at edge at 3.5 M; many at 4 M.
2) 0.4 M NaCl and 50 mM Na cacodylate pH 7.5
   A few floating microcrystals that light and extinguish at 1.5 M; a small, single crystal (?) at 2.5 M; precipitate ring around spot and oiling on meniscus at 3.5 M; extensive oiling at 4 M.
3) 1 M NaCl and 50 mM Na cacodylate pH 7.5
   A few floating microcrystals at 1.5 M; some blobs that light with a cross pattern at 2 M; precipitate ring around spot and light oiling in solution at 3.5 M.
4) 80 mM $MgCl_2$ and 50 mM Na cacodylate pH 7.5
   A few microcrystals that light and extinguish at 1.25 M; a small (0.05 by 0.15 mm) rectangle that lights and extinguishes in a wave at 1.75 M; precipitate ring around spot at 2.25 M; oiling out at 3 M.
5) 0.4 M $MgCl_2$ and 50 mM Na cacodylate pH 7.5
   A large fiber that lights and extinguishes at 1.25 M; microcrystals at 1.75 M; some blobs that light with a cross pattern at 2 M; light oiling, and a small single crystal growing off of the fiber at 3 M; precipitate ring around spot at 3.5 M.
6) 100 mM KPO_4 pH 7.5
   Microcrystals at 1.25 M; precipitate ring around spot at 2 M; oiling in solution at 2.25 M; heavily twinned microcrystals around edge and in some of the oil drops at 3.5 M.

IV. Ammonium phosphate

A. Vapor diffusion (hanging drop); 4 M ammonium phosphate in reservoir.

1) Ammonium phosphate pH 5.0
   Fiber with several small crystals growing from it, oiling out at one week; radially lighting and extinguishing patches at three weeks.
2) Ammonium phosphate pH 5.5
   Oil drops at one week; small crystal at two weeks.
3) Ammonium phosphate pH 6.0
   Oiling at one week; fibers at three weeks.
4) Ammonium phosphate pH 6.5
   Oil drops at one week; microcrystals at two weeks.
5) Ammonium phosphate pH 7.0
   Oil drops and a few small microcrystals which light and extinguish at one week; fibers at two weeks.
6) Ammonium phosphate pH 7.5
   Oiled out immediately on mixing.
7) Ammonium phosphate pH 8.0
   Oiling immediately on mixing; fibers at three weeks.
8) Ammonium phosphate pH 8.5
   Oiling immediately on addition; possible small crystal at two weeks.
9) Ammonium phosphate pH 9.0
   A couple of small crystals, one single, one multiply nucleated at one week.

B. Vapor diffusion (spot plates); 1.25 M to 2.5 M at 0.25 M every week; 2.5 M to 4.0 M at 0.5 M every week.

1) 1 M Ammonium phosphate pH 7.5
   A floating clump that lights and extinguishes in sections at 1.25 M; a "snowball" at 1.75 M; precipitate ring around spot at 2.25 M; scum (oiling out or precipitate?) at 3 M.
2) 0.92 M Ammonium phosphate and 0.4 M NaCl pH 7.5
   Several clumps, some of which light, at 1.25 M; fibers and a multinucleated crystal at 2.5 M.
3) 1 M NaCl and 0.9 M ammonium phosphate pH 7.5
   A salt crystal at 1.25 M.
4) 0.94 M ammonium phosphate and 0.125 M Na_2SO_4 pH 7.5
   Spot clouded on setup; clear after one week with some debris; oiling at 2.25 M; precipitate ring around spot at 2.5 M; completely oiled out at 4 M.

V. Ammonium citrate

A. Vapor diffusion (spot plates); 1 M to 3 M at 0.25 M every seven days.

1) 50 mM Na cacodylate pH 7.5
   Some fibers at 1.25 M; precipitate ring around spot and some microcrystals that light and extinguish at 1.75 M; radially lighting and extinguishing patches at 2.25 M; larger microcrystals and a faintly birefringent skin over bottom and surface of spot at 2.5 M; fibers along surface of spot at 2.75 M.
2) 0.43 M NaCl and 50 mM Na cacodylate pH 7.5
   Some fibers at 1 M; microcrystals at 1.75 M.
3) 1.1 M NaCl and 40 mM Na cacodylate pH 7.5
   Some fibers at 1 M; more at 1.75 M; bottom oiled out at 2.5 M.
4) 87 mM $MgCl_2$ and 47 mM Na cacodylate pH 7.5
   A few fibers at 1.25 M; some microcrystals at 1.5 M; precipitate ring around spot at 1.75 M; oil at 2.25 M; around six large rosettes (approximately 2 mm across) on bottom at 2.5 M; birefringent patches on surface and some 0.1 mm needle crystals on bottom at 2.75 M.
5) 0.43 M $MgCl_2$ and 47 mM Na cacodylate pH 7.5
   Fibers at 1.25 M; some microcrystals at 1.75 M; lots of fibers plus some small specs that light and extinguish at 3 M.
6) 140 mM Na_2SO_4 and 48 mM Na cacodylate pH 7.5
   Fibers at 1.25 M; precipitate ring around spot at 2 M; hundreds of microcrystals plus needle crystals running along surface that light and extinguish at 2.25 M.
7) 100 mM KPO_4 pH 7.5
   Fibers at 1.25 M; precipitate ring around spot at 1.75 M; microcrystals on bottom that light and extinguish at 2.25 M; birefringent skin on surface at 2.5 M; a small fiber crystal (0.05 by 0.05 mm) that lights and extinguishes evenly at 3 M.

B. Microdialysis; 1.5 M to 3 M in 0.25 M steps every three days

1) pH 5.0
   Some material on membrane at 1.5 M; oiling and multiple sheet-like crystals at 2 M.
2) pH 6.0
   Hundreds of possible twinned fiber crystals on membrane at 1.5 M; some oiling and small crystals on surface of membrane (tube had dried out) at 2 M.
3) pH 7.0
   Hundreds of oil drops on membrane at 1.5 M; oiling on surface at 2 M; precipitate ring around tube and crystalline precipitate in tube at 2.5 M.
4) pH 8.0
   Oil drops at 1.5 M; precipitate around meniscus,

radially lighting and extinguishing patches on walls and scummy precipitate on membrane at 2.25 M.

5) pH 9.0
Oiling at 1.5 M; crystalline precipitate around meniscus at 2.25 M; precipitate on membrane at 2.5 M.

C. Vapor diffusion (spot plates); 2 M to 3 M ammonium citrate pH 7.5 in steps of 0.25 M every seven days.

1) pH 7.5
Half-dozen blobs that light with a cross at 2 M; precipitate ring around spot, skin on meniscus and numerous small (up to 0.2 mm) rosettes in oily ring around spot that light and extinguish radially at 2.25 M.
2) 37 mM $MgCl_2$ pH 7.5
Some blobs that light with a cross at 2 M; precipitate ring around spot, skin on meniscus and some small rosettes at 2.25 M.
3) 209 mM $MgCl_2$ pH 7.5
Some microcrystals, some oblong crystals that light but do not extinguish and a thread at 2 M; a crystalline "half-ring" at 2.5 M; salt crystals at 2.75 M.
4) 43 mM Na_2SO_4 pH 7.5
Blobs that light with a cross at 2 M; skin on meniscus, precipitate ring around spot and rosettes around edge at 2.25 M; several large patches that barely light but extinguish at 2.75 M.
5) 209 mM Na_2SO_4 pH 7.5
Microcrystals, a film on the surface of the spot with thread-like fibers running along it and a slight precipitate ring around spot at 2 M; rosettes at 2.25 M.
6) 43 mM KPO_4 pH 7.5
Some thread-like crystals and microcrystals at 2 M; a skin with a thready birefringence on the surface of and a precipitate ring around the spot at 2.25 M; "snowballs" at 2.5 M.
7) 209 mM KPO_4 pH 7.5
Some microcrystals and crystalline aggregates that light at 2 M; skin on surface with "snowballs" around edge at 2.5 M.

D. Vapor diffusion (spot plates); 2 M to 3 M ammonium citrate pH 8.5 in steps of 0.25 M every seven days.

1) pH 8.5
Crystalline ring around spot and birefringent skin on spot at 2 M; "snowballs" at 3 M.
2) 50 mM $MgCl_2$ pH 8.5

Some blobs that light with a cross at 2 M; skin on surface and a crystalline ring of 0.15 mm long fibers around spot at 2.25 M.

3) 250 mM $MgCl_2$ pH 8.5
   A few blobs that light with a cross at 2 M; a ring of microcrystals around edge of spot at 2.25 M; skin over surface at 2.5 M; crystalline patches on skin at 2.75 M; salt crystals at 3 M.
4) 50 mM Na_2SO_4 pH 8.5
   Skin on surface and crystalline ring around spot at 2 M.
5) 250 mM Na_2SO_4 pH 8.5
   Ring of fiber crystals around spot and nonbirefringent skin on surface at 2 M; crystals growing larger with increasing salt concentration; a chunk that lights and extinguishes weakly at 2.75 M.
6) 50 mM KPO_4 pH 8.5
   Nonbirefringent skin on surface and crystalline fiber ring around spot at 2 M; a chunk that lights weakly at 2.5 M.
7) 250 mM KPO_4 pH 8.5
   Precipitate ring around spot and nonbirefringent skin on surface at 2 M; did not change.

Table Three: Headpiece/Operator Crystallization Trials

The following is a list of all cocrystallization trials with the lac headpiece protein and the 21 base-pair lac operator DNA oligomer and their results. Trials are grouped according to precipitant used, followed by crystallization technique, and ending with the specific conditions of each trial. All trials were set up using a headpiece protein stock at an initial concentration of 6-7 mg/ml, containing 100 mM KCl, 10 mM Na cacodylate pH 7.5, 0.1 mM ethylenediamine tetraacetic acid, 0.1 mM dithiothreitol, 0.1 mM phenylmethane sulfonyl fluoride, and 0.02 % Na azide (HP buffer); operator stock was at a concentration of 12 mg/ml in water. Crystallizations in spot plates were prepared by adding concentrated solutions of precipitants and counterions to 5 microliter samples of the operator DNA with various amounts of the protein to give the initial starting levels of each component in each well. NOTE: In this chart a 1:1 ratio of headpiece protein to operator means two protein molecules per DNA strand, since the operator has two headpiece binding sites in it.

I. Polyethylene glycol (PEG) 6000

A. Vapor diffusion (spot plates): 10% to 30%, at 5% per week.

1) 1:1 headpiece:operator
   A large (0.1 x 0.25 mm) twinned crystal that lights and extinguishes multiply, plus a slight ring and specs at 10%; the ring gone but the rest the same at 15%.
2) 2:1 headpiece:operator
   Some specs and a slight ring at 10%; oiling out on surface, plus thousands of small (<0.05 mm) shapes that light with a cross pattern at 25%.
3) 1:1 headpiece:operator plus 10 mM $MgCl_2$
   Some specs and a ring at 10%; hundreds of oil drops of all sizes (up to 0.1 mm) that light with a cross on surface and bottom of spot at 25%.
4) 2:1 headpiece:operator plus 10 mM $MgCl_2$
   Some specs and a ring at 10%; oil drops and numerous shapes that light with a cross at 15%; larger (up to 0.15 mm across) at 25%.
5) 1:1 headpiece:operator plus 100 mM $MgCl_2$
   Some specs and a ring at 10%; numerous shapes on surface and bottom that light with a cross at 15%.
6) 2:1 headpiece:operator plus 100 mM $MgCl_2$
   Some specs and a slight ring at 10%; shapes that light with a cross and a few floating crystals that

light and extinguish at 15%.

II. 2-methyl-2,4-pentanediol (MPD)

A. Vapor diffusion (spot plates): 20% to 80%, at 10% per week.

1) 1:1 headpiece:operator
Shapes that light with a cross at 20%; some multiple crystals at 40%; hundreds of long, thin crystals that light and extinguish singly, none greater than 0.05 mm long, on surface and bottom at 60%; some square crystals that appear deep red under polarizer at 70%.
2) 2:1 headpiece:operator
Shapes that light with a cross at 20%; slight ring around spot at 50%; hundreds of tiny rectangular crystals that light and extinguish singly at 60%; salt crystal at 70%.
3) 1:1 headpiece:operator plus 10 mM $MgCl_2$
Some fibers and specs at 20%; some shapes and a few small, ill-formed crystals (the largest 0.025 x 0.1 mm) that light and extinguish singly at 40%; hundreds of twinned crystals, each two rectangular crystals joined end-to-end at a slight angle, most having poor form, but some being well-formed and lighting and extinguishing singly, at 50%; spot slightly precipitated at 60%; less precipitated at 70%.
4) 2:1 headpiece:operator plus 10 mM $MgCl_2$
Some specs at 20%; hundreds of well-ordered but smaller twinned crystals at 50%; larger, some up to 0.1 mm long, also spot slightly precipitated at 60%; fewer crystals at 70%.
5) 1:1 headpiece:operator plus 100 mM $MgCl_2$
Some specs at 20%; thousands of shapes that light with a cross at 50%; hundreds of small (<0.05 mm on edge), cube-like crystals that are well-formed, not twinned, light and extinguish singly and appear deep red under the polarizer, plus some precipitate at 60%.
6) 2:1 headpiece:operator plus 100 mM $MgCl_2$
Some specs at 20%; thousands of shapes that light with a cross at 50%; thousands of specs at 60%; smaller crystals plus amorphous blobs and twinned clusters of crystals at 70%.

B. Vapor diffusion (spot plates): 20% to 60%, at 10% per week.

1) 1:1 headpiece operator plus 10 mM $CaCl_2$
Some specs at 20%; a few shapes that light with a

cross at 30%; hundreds of multinucleated crystals, plus a few twinned crystals, measuring up to 0.05 x 0.2 mm that light and extinguish beautifully as two crystals joined end-on-end, at 40%; clusters of multinucleated crystals and only a few tiny single crystals at 50%.

2) 2:1 headpiece:operator plus 10 mM $CaCl_2$
   Some specs at 20%; some fibers and specs and a slight ring at 40%; thousands of powderlike crystals on the surface and bottom that light and extinguish at 50%.
3) 1:1 headpiece:operator plus 10 mM $Sr(NO_3)_2$
   Some specs at 20%; faint ring around spot at 30%; thousands of end-on-end twinned crystals, none larger than 0.05-0.1 mm at best and not well-formed but which light and extinguish at 40%; hundreds of small rectangular crystals that light and extinguish on surface and on bottom of spot at 50%.
4) 2:1 headpiece:operator plus 10 mM $Sr(NO_3)_2$
   Some specs at 20%; faint ring around spot at 30%; numerous end-on-end twinned crystals that are smaller than in 3) but better formed and have a reddish streak in them under polarizer at 40%.
5) 1:1 headpiece:operator plus 10 mM $BaCl_2$
   Some specs at 20%; ring around spot at 30%; thousands of very small end-on-end twinned crystals that light and extinguish at 40%; thousands of needles on surface and bottom at 50%.
6) 2:1 headpiece:operator plus 10 mM $BaCl_2$
   Some spheres that light with a cross at 20%; a ring around spot at 30%; thousands of fibrous, needle crystals that light and extinguish at 40%; slightly larger needles at 50%

C. Vapor diffusion (spot plates): 20% to 60%, at 10% per week

1) 1:1 headpiece:operator plus 7.55 mM $MgCl_2$
   Some spheres that light with a cross at 20%; a ring around spot at 40%; tiny needle crystals, single, multiple and in clusters at 50%.
2) 0.5:1 headpiece:operator plus 12.78 mM $MgCl_2$
   Some specs at 20%; hundreds of "snowball" crystals plus some end-on-end twinned crystals and cubic crystals that light and extinguish, all <0.05 mm, at 40%; hundreds of cubic crystals on surface that light and extinguish as multiple crystals at 50%.
3) 1.5:1 headpiece:operator plus 10 mM $MgCl_2$
   Some specs at 20%; a ring around spot at 40%; hundreds of tiny needles at 50%.
4) 2:1 headpiece:operator plus 10 mM $MgCl_2$
   Some specs at 20%; dozens of end-on-end twinned crystals, not very good shape but up to 0.05 x 0.3 mm

that light and extinguish at 40%.

5) 2:1 headpiece:operator plus 20 mM $MgCl_2$
Some specs at 20%; "snowball" crystals, plus some small end-on-end twinned crystals at 40%; the latter larger (up to 0.2 mm) but not well-ordered at 50%.

D. Vapor diffusion (spot plates): 20% to 60%, at 10% per week

1) 2:1 headpiece:operator plus 20 mM $MgCl_2$
Some specs at 20%; numerous specs, plus some small (0.5 mm on edge) cubes that light and extinguish at 30%; scores of end-on-end twinned crystals (up to 0.05 x 0.15 mm), with reasonable shape and that light and extinguish, at 40%.

2) 2:1 headpiece:operator plus 1 mM spermine·HCl pH 7.0, 20 mM $MgCl_2$
Some specs at 20%; thousands of spheres that light with a cross around edge of spot at 30%; thousands of tiny microcrystals on surface and on bottom of spots at 40%.

3) 2:1 headpiece:operator plus 3 mM spermine·HCl pH 7.0, 66 mM $MgCl_2$
Spot precipitated on addition of spermine and did not resuspend until 66 mM Mg added; some specs at 20%; hundreds of tiny crystals on surface plus dozens of large spiral crystals on bottom that light and extinguish at 40%.

4) 2:1 headpiece:operator plus 40 mM $MgCl_2$
Some specs at 20%; hundreds of cubic crystals and end-on-end twinned crystals on surface and bottom that have good shape and light and extinguish, but are too small, at 40%; small crystals and rosettes at 50%.

5) 1:1 headpiece:CGCGAATTCGCG (plus 1 spermine, 25 $MgCl_2$)
Some specs at 20%; half a dozen large (up to 0.15 x 0.8 mm) native DNA crystals (no changes in 0KL from native pattern) at 40%.

6) 2:1 headpiece:CGCGAATTCGCG (plus 1 spermine, 25 $MgCl_2$)
Some specs at 20%; some poorly formed crystals at 30%.

Figures

Figure 1. Sephadex G-50 Fine column. Fractions 41-52 represent excluded volume; fractions 71-86 the headpiece fraction; fractions 105-115 the fully included volume.

Figure 2. SDS polyacrylamide gel showing the course of the headpiece purification. Gel is a 5 to 20% linear gradient gel containing 7M urea, according to the protocol of Horigome et al. (Horigome, T., Yoshida, K., Kanai, Y. and Sugano, H (1980) Seibutsu Butsuri Kagaku 24, 187-190).

Lane 1: repressor protein before proteolysis.

Lane 2: repressor after 3 hours of proteolytic digestion at 20°C with alpha-chymotrypsin (1 mg enzyme per 100 mg repressor).

Lane 3: digest after first lima bean trypsin inhibitor column.

Lane 4: load-wash of first P-11 column.

Lane 5: high-salt elution peak of first P-11 column.

Lane 6: First peak (fractions 41-52) of Sephadex G-50 Fine column.

Lane 7: Second peak (fractions 71-86) of G-50 column.

Lane 8: second P-11 column high-salt elution peak.

Lane 9: final, concentrated headpiece (approx. 6 mg/ml).

Lane 10: headpiece final product after a 14 day incubation at 4°C in HP storage buffer.

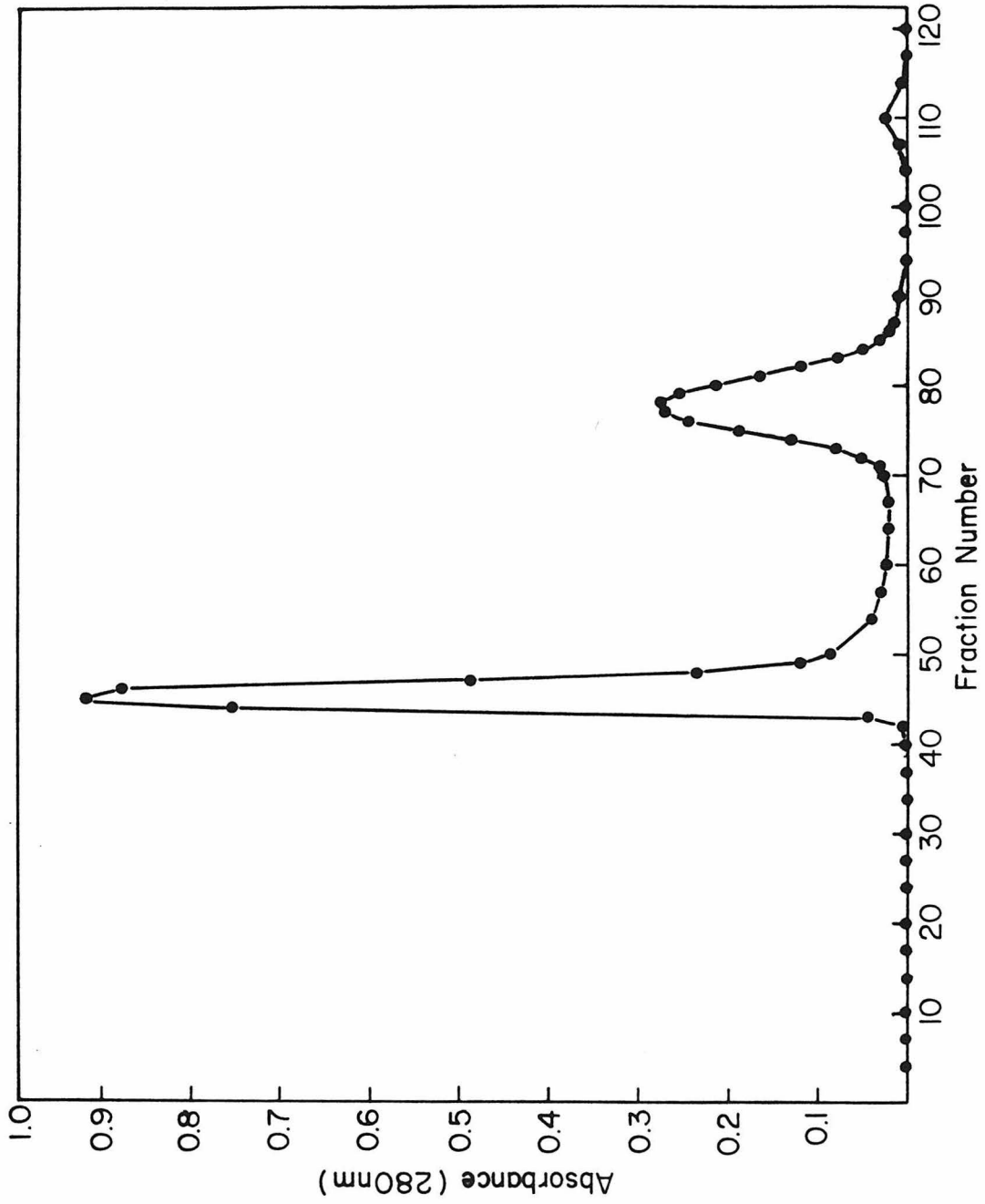

Absorbance (280nm)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
10
20
30
40
50
60
70
80
90
100
110
120
Fraction Number

Figure 1

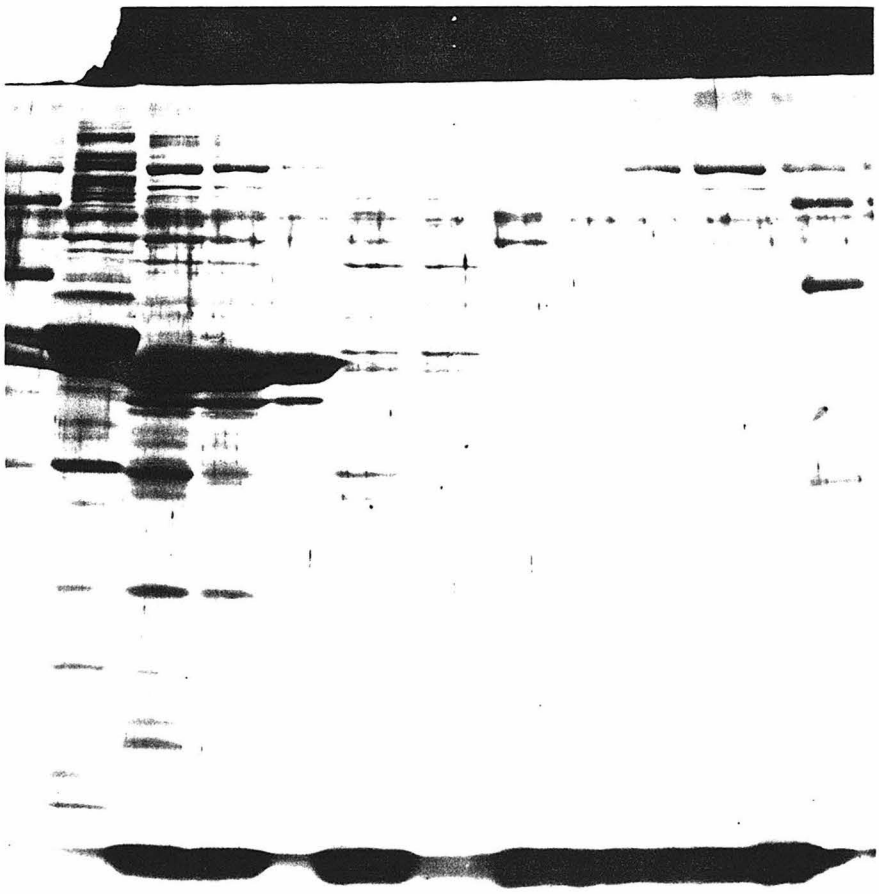


Figure 2

CHAPTER 3

DRUGS THAT BIND THE DOUBLE HELIX: HOECHST 33258

Introduction

The final chapter describes our structural study of the B-DNA 12mer C-G-C-G-A-A-T-T-C-G-C-G complexed with the DNA fluorochrome Hoechst 33258. Hoechst is a member of a large and very important class of DNA binding molecules that are characterized by their ability to bind to B-form DNA in the minor groove. As reviewed recently by Zimmer and Waehnert (5), this group includes a number of important antibiotic and antitumor agents, such as distamycin A, netropsin and a host of others, which all share a common set of characteristics. First of all, each one interacts with DNA in a nonintercalative manner by binding in the minor groove. Characteristically, rather than disrupt the structure of the DNA, they instead strengthen the double helix, often leading to such effects as an increase in melting temperature and greater stability of the B conformation. Also, they tend to be sequence-preferring rather than sequence-specific, for example, having a high affinity for AT-rich over GC-rich regions or vice versa. Another characteristic is a dual mode of binding, with many of the compounds exhibiting a more specific, tight-binding mode to certain sequences along with a less specific, weaker binding to DNA in general. Work by Denny et al. (1) has shown that the nature of this tight-binding interaction may be correlated to antitumor activity in at least some of these compounds, suggesting that proper design of such molecules may lead to drugs with high activity against a number of diseases.

Our interest in these minor groove binders began with our analysis of a complex of the antitumor compound netropsin with C-G-C-G-A-A-T-T-^{Br}C-G-C-G (3, 4). The structure revealed how this AT-specific binder fit into the minor groove of the AT center of the 12mer, steered into place by a series of bifurcated hydrogen bonds between the drug and groups on the base-pair edges of the groove. In the process, it replaces a series of ordered waters in the groove of the DNA thought to help stabilize it, which we call the "spine of hydration," forming an even stiffer spine and thereby accounting for the dramatic rise in melting temperature seen on netropsin binding to double-helical B DNA (2). In addition, the reason for the molecule's strong preference for AT-rich regions was shown to be due to a series of close steric contacts between the drug and the bottom of the groove, contacts that would not be allowed in GC-rich regions due to the presence of the guanine N2 amine group. This latter observation led to the suggestion of a new series of molecules that could exploit this van der Waals contact to preferentially steer them to either GC- or AT-rich regions, which we refer to as "lexitropsins."

We became interested in Hoechst because both its structural features and its known high affinity for AT-rich regions of DNA suggested to us that it might be another minor groove-binding compound like netropsin. Possessing anthelmintic properties as well as being a widely used DNA fluorochrome, it was intriguing because it not only would bind to AT-rich regions but also had a tolerance for GC-rich regions as well, making it an example of both types of binding in one molecule. The structure confirmed our predictions for minor groove

binding, reinforcing our model for the way in which AT-specific binding compounds work. It also provided an example of a GC-specific binding interaction with the groove and unexpectedly revealed how conformational variation in the molecule can influence sequence preference. The detailed structure analysis of Hoechst 33258 is given in the following paper.

References

1. Denny, W. A., Atwell, G. J., Baguley, B. C. and Cain, B. F. (1979) J. Medicinal Chem. 22, 134-150.

2. Kopka, M. L., Fratini, A. V., Drew, H. R. and Dickerson, R. E. (1983) J. Mol. Biol. 163, 129-146.

3. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) Proc. Natl. Acad. Sci. USA 82, 1376-1380.

4. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) J. Mol. Biol. 183, 553-563.

5. Zimmer, C. and Waehnert, U. (1986) Prog. Biophys. Molec. Biol. 47, 31-112.

DRUGS THAT READ THE DOUBLE HELIX: HOECHST 33258

Philip Pjura, Kazimierz Grzeskowiak, and Richard E. Dickerson

Molecular Biology Institute, University of California at Los Angeles

Los Angeles, CA 90024, U.S.A.

Summary

An x-ray crystal structure analysis has been carried out of the complex between the antibiotic and DNA fluorochrome Hoechst 33258 and a synthetic B-DNA dodecamer of sequence: C-G-C-G-A-A-T-T-C-G-C-G.

The drug molecule, which can be schematized as: phenol-benzimidazole-benzimidazole- piperazine, sits within the minor groove in the A-A-T-T region of the DNA double helix, displacing the spine of hydration that is found in drug-free DNA. The N-H of benzimidazoles form bridging three-center hydrogen bonds between adenine N3 and thymine O2 atoms on edges of base pairs, in a manner both mimicking the spine of hydration and calling to mind the binding of the antitumor drug netropsin. Two conformers of Hoechst are seen in roughly equal population, related by 180^{o} rotation about the central benzimidazole-benzimidazole bond: one form in which the piperazine ring extends out from the surface of the double helix, and another in which it is buried deep within the minor groove. Steric clash between drug and DNA dictates that the phenol-benzimidazole-benzimidazole portion of Hoechst 33258 binds only to AT regions of DNA, whereas the piperazine ring demands the wider groove characteristic of GC regions.

1. Introduction

Hoechst 33258 (Figure 1a) is a member of a large class of compounds whose useful and interesting properties derive from their ability to bind to nucleic acids. A synthetic preparation of the Hoechst Pharmecutical company, it was originally developed as one of a series of N-methylpiperazine derivatives of potential clinical use (36). It showed no antineoplastic activity, but was found to be fairly effective as an anthelmintic (30, 31). Its major usefulness, however, comes from its ability to fluoresce specifically when bound to DNA. It has been used widely as a cytological stain, both in vitro and in vivo, forming a brightly fluorescent complex with chromosomes, with a marked specificity for AT-rich regions (12, 13, 28). In DNA in which thymidine has been substituted by bromodeoxyuridine (BrdU), the fluorescence of Hoechst 33258 is quenched, and this effect has been used as a sensitive probe for DNA replication and strand exchange in dividing cells (17, 33, 46). Growth of cells on low levels of Hoechst results in a number of effects on replicating chromosomes, including decondensation in AT-rich regions (14, 40), the induction of breaks in "fragile" sites (44), stage-specific inhibition of cell growth (15), the induction of polyploidy and endoreplication in cell lines (29), and, in conjunction with BrdU, an increased sensitivity to UV killing in growing cells (49, 51, 52). Other uses have included fluorescence-activated sorting of chromosomes (27); a method of separating DNA based on AT content due to induced changes in electrophoretic mobility (38, 44); and as a diagnostic test for

mycoplasma contamination in mammalian cell lines (2).

How Hoechst works is determined by the nature of its interaction with DNA, and physiochemical studies of this interaction both with DNA in solution and in the cellular milieu have revealed the following picture. Hoechst exhibits two modes of binding to nucleic acids: a low-level, tight-binding mode and a high-level, low-affinity mode, which have quite different characteristics (1, 53). The former mode occurs at a drug-to-phosphate ratio of less than 0.05–0.20 (42), with a tight-binding constant of approximately 10^6–10^7 M^{-1} (43), is unaffected by 0.01 to 1.0 M salt, is proportional to the percent AT base-pairs in the duplex (35), is noncooperative, and leads to induction of fluorescence which is also proportional to percent AT (34). The latter mode begins to predominate at a drug-to-phosphate ratio of greater than 0.1, has a weaker binding constant of 10^4–10^5 M^{-1}, is abolished at high salt, is insensitive to the base composition of the polymer, is highly cooperative, and causes quenching of the fluorescence generated by specific binding. Because its physical characteristics are those that desribe the behavior of Hoechst as a fluorochrome and a drug, the tight-binding mode is the one of biochemical and pharmaceutical interest and the one that we have studied.

This tight-binding interaction seems to represent binding to double-stranded DNA in the B form, since binding to single-stranded forms (48), to RNA-RNA or RNA-DNA duplexes (43), or to DNA under conditions in which it assumes the A form (42), is seen to be of the low-affinity type. Binding causes characteristic shifts in the drug's

absorption maximum from 338 to 345-356 nm, the induction of circular dichroism at 355 nm, and of fluorescence emission at 505 nm, all of which are observed to be proportional to the percent AT of the nucleic acid under high salt conditions (3, 35, 56). The binding does not appreciably disrupt the DNA structure, as seen by the small changes in its electric dichroism and birefringence properties (1). The main effect is a dramatic rise in T_m, increasing by 4° C in poly (CG) and 40° C in poly (AT), with this increase in mixed polymers being proportional to percent AT (3). Competition with other compounds known to bind to DNA by various intercalative and nonintercalative modes of bonding show effects on Hoechst binding only in the case of minor groove, AT-specific binders (32, 50), although GC bases lying within AT stretches are tolerated by the dye (16). The minor groove is also implicated as the specific site of binding because poly (IC) is more similar to poly (AT) than to poly (CG) in binding properties, and because Hoechst binds well to phage T6 DNA, in which the major groove is blocked by glycosylation (42). When bound, the dye is shown to lie at a 45 to 49° angle to the helix axis, consistent with minor groove binding (20, 39). Sensitivity of binding to the presence of guanidine·HCl but not to NaCl or urea suggests a combination of hydrogen bonding and electrostatic components to binding, and titration curves of the dye with DNA of various base compositions indicate that three hydrogen bonding sites are involved (42, 53). Several footprinting studies indicate that the Hoechst binding site spans three to five bases (11, 41). Based on this evidence, a model has been proposed in which the dye binds nonintercalatively in the

minor groove, with the N groups of the two benzimidazole rings and the N-methyl group of the piperazine ring making three hydrogen bonding interactions with the N3 and O2 groups of AT base pairs in the bottom of the groove (11, 42).

We became interested in Hoechst because these properties made it a member of an important class of DNA-binding drugs, such as distamycin and netropsin, which bind to DNA in the minor groove. We had previously studied a complex of the AT-specific binding netropsin (Figure 1b) with a twelve base-pair DNA oligomer of sequence C-G-C-G-A-A-T-T-^{Br}C-G-C-G by single crystal x-ray diffraction analysis (25, 26). Our results showed that this drug indeed binds to the B-form dodecamer by fitting into the minor groove of the central A-A-T-T region. In doing so, it displaces a series of ordered water molecules seen in the native DNA structure, termed the "spine of hydration," which help to stabilize the B conformation. Netropsin mimics this spine by forming a series of bifurcated hydrogen bonds with the base edges in the groove, explaining both its binding preference for AT sequences and its stabilizing effect on B-DNA. Hoechst's physical and physiochemical similarities to netropsin prompted us to carry out a similar analysis to a resolution of 2.2 Angstrom. Our results show that the same mode of minor groove binding occurs in Hoechst as in netropsin, confirming and expanding the conclusions of previous studies. In addition, Hoechst exhibits an unexpected conformational duality in binding, which explains its ability to tolerate GC base-pairs. These results have important implications for understanding the compound's properties as a DNA

fluorochrome and drug, and valuable lessons for the understanding and design of molecules that bind to DNA.

2. Results: Drug Binding Within the Minor Groove

The crystal structure of the complex of Hoechst 33258 with a double-helical B-DNA oligomer of sequence: C-G-C-G-A-A-T-T-C-G-C-G was solved as outlined under Experimental Procedures. Plate IA and Figure 2 show how the Hoechst melecule sits within the minor groove, and how it displaces an ordered zig-zag string of water molecules, or spine of hydration, that is present in the free DNA (7, 23). This spine of hydration has two layers: A first hydration shell of water molecules forms hydrogen-bonded ridges between adenine N3 and/or thymine O2 atoms on opposite helix strands and on adjacent base pairs — between the bases in these pairs that are brought closer together by helix rotation. A second shell of waters then bridges the first-shell molecules, giving them an approximate local tetrahedral coordination. The spine is well-developed in the A-A-T-T center of the DNA helix, but breaks up as it enters the C-G-C-G regions above and below, mainly because of steric hindrance from guanine C2 amine groups. This spine of hydration is believed to make an important contribution to the stabilization of B-DNA, relative both to interconversion to another helix type, and to unwinding into singe strands.

The Hoechst 33258 molecule displaces the spine of hydration in the region where the drug binds (Figure 2b). But its nitrogen atoms on the concave side of the molecule form bifurcated or two-center hydrogen bonds (19, 54) that bridge exactly the same pairs of DNA atoms as did the spine of hydration. The hydrogen-bonded backbone of

the spine is replaced by a covalently bonded drug backbone, conferring added stability on the B-DNA double helix. This behavior of Hoechst is very much like that of netropsin (25, 26).

Actual distances between potentially hydrogen-bonding N and O atoms of DNA and drug or water are given in Figure 3. In the center of the spine of hydration in free DNA, the N-O and N-N distances are reasonable for conventional hydrogen bonds, ca. 2.80 Angstrom. As the spine enters the C·G base pair regions above and below, it pulls away from the floor of the minor groove, making the N-O and N-N distances considerably longer. Ultimately, the spine of hydration in the dodecamer crystals ends when its way is blocked by 3'-OH groups from the ends of overlapping neighboring molecules.

In the Hoechst complex, the two benzimidazole nitrogens, N1 and N3, sit at somewhat long but reasonable distances from DNA O2 and N3 atoms, but the piperazine ring is less comfortably fitted into the groove. Because of the geometry of its attachment to the first benzimidazole, the phenol group projects up at an angle to the bottom of the minor groove, and its -OH is not in a favorable position for direct hydrogen bonding to base edges. In fact, as Figures 2b and 3b indicate, the phenol -OH functions as a second-shell hydrating atom, with a bond to water no. 71 and from there to the DNA O2 and N3 atoms.

A somewhat disordered half of the original spine of hydration continues up the minor groove in the region not occupied by the drug.

The two aromatic benzimidazole rings of Hoechst, like the pyrrole rings of netropsin, are slotted tightly into the narrow A·T region of a minor groove that is barely wide enough to admit them. The

tightness of fit can be appreciated from the cross section through the first benzimidazole (Bz1) in Plate ID, showing van der Waals' nonbonded packing surfaces. As will be seen in the next section, this fit is so constricted that the DNA must adapt slightly in order to admit the drug molecule. In order that each benzimidazole group should fit parallel to the walls of the groove in its own local region, Bz1 and Bz2 must be twisted by 36^{o} around the bond that joins them. The phenol ring, in contrast, remains coplanar with Bz1 and hence is not parallel to the walls of the groove in its immediate neighborhood (Plate IC). This probably represents a tradeoff in energy between efficient packing within the groove, vs. the resonance stabilization that occurs if Phe and Bz1 constitute one large, planar conjugated bond system. Such a tradeoff is permissible because Phe does not sit as deep down in the minor groove as do Bz1 and Bz2. The bond between Phe and Bz1, being symmetrical about the midline of Bz1, does not allow the drug molecule to follow the curvature of the groove itself.

The piperazine ring is involved in yet another type of conformational energy tradeoff. Steric clash between piperazine $-CH_2-$ and benzimidazole -CH- makes the most stable situation for the free drug in solution one in which the best mean planes through Pip and Bz2 are roughly at right angles to one another. Yet a snug fit within the minor groove would favor these two rings' being nearly coplanar. The compromise between these mutually conflicting demands is seen in cross section in Plate IE. The best plane through the puckered Pip ring lies approximately 60^{o} away from the midplane of the minor groove.

The narrowness of the groove apparently forbids its turning any farther about the bond connecting it with Bz2. Even this compromise is possible only because Pip sits in a G·C region of the groove, a region that is intrinsically wider than the A·T region (see Figure 8 of Reference 9). Pip thus actually favors a G·C region over an A·T, in contrast to the behavior of Bz1 and Bz2. Moreover, the guanine $-NH_2$ that would prevent Bz1 and Bz2 from approaching the bottom of the groove closely enough to act as donors in hydrogen bonds actually aids in holding down Pip; the rotated, puckered ring does not sit deeply enough in the groove for the -NH2 to be a steric hindrance, and the $-NH_2$ actually donates a long hydrogen bond to the terminal N6 atom of piperazine.

There is yet one further twist to the piperazine story. The observed electron density during x-ray structure analysis suggested a statistical distribution of the piperazine ring between two alternate positions, shown in Plate IB and Figure 4. In the position that has been discussed so far, Pip nests down in the minor groove (Figure 4a), whereas in the alternate position it extends out of the groove (Figure 4b). Least-squares refinement suggested a roughly equal population between these two states. Note that interconversion between the two Pip conformations is not something that can occur when Hoechst 33258 is bound to DNA, since it requires an 180^{o} rotation about the bond connecting Bz1 and Bz2. Binding to DNA can only "freeze out" a conformation equilibrium that pre-existed in the free drug molecule in solution. But as will be mentioned later in the Discussion, these two conformers of Hoechst enable it to tolerate both A·T and B·T base

pairs at one end of its binding site. Hence, in this respect, the base specificity of Hoechst 33258 differs from that of netropsin.

3. Results: Effects of Drug Binding on DNA Structure

As with netropsin, so the binding of Hoechst 33258 has very little effect on the local helix parameters of DNA. Table I compares mean helix parameters for these two drugs, with the respective variants of free DNA to which they were complexed. (Hoechst should be compared with native DNA, and netropsin with MPD7). Helical twist and rise per base pair are virtually unchanged by the binding of either drug, as are the inclination of base pairs to the helix axis and their displacement from that axis. The main alterations in helix parameters produced by binding of Hoechst are a 5^{o} mean increase in propeller twist and a shift of main chain torsion angle delta (C5'-C4'-C3'-O3') at purines by an average of 10^{o} toward a more C2'-endo-like conformation. Pyrimidines remain clustered in the C1'-exo region. The structural significance of these small changes is not clear, since equivalent small changes were observed between the two variants of the drug-free DNA helix: native and MPD7.

Both Hoechst and netropsin do exert two overall effects on the DNA: they widen the minor groove and bend the helix axis backward by $7\text{-}8^{o}$ at the binding site. Both of these effects can be seen in Figure 5, which is a superposition of the DNA helix alone with that when the DNA is complexed with Hoechst. The most marked widening of the groove occurs in the region of Bz2 and Pip and is produced primarily by movement of phosphates P10 and P11. These shifts are shown quantitatively in Figure 6. The widening, as measured by nearest P-P distances across the groove, is +0.76 Angstrom (A) at step P9-P20 near

Bz1, +2.07 A at P10-P19 near Bz2, +1.70 A at P11-P18 near Pip, and +1.40 A at P12-P17 beyond the end of the drug molecule. Corresponding groove widenings at these same four positions in the netropsin complex are: +0.16 A, +1.94 A, +1.54 A, and +1.16 A, respectively.

The bending back of the helix axis produced by binding of Hoechst can be shown quantitatively in the base pair normal vector plot, Figure 7. The view in this plot is down the helix axis, and points 1-12 are the ends of vectors that describe the change in normal vector orientation produced by drug binding. Points for the upper half and the lower half of the helix cluster around two different regions, whose separation on this plot indicates a bend of ca. 7^{o} in helix axis. Moreover, the bend is concave away from the Bz2 binding site, suggesting that it is a consequence of drug binding.

This is precisely the region of greatest minor groove-widening as well. Netropsin shows similar behavior: greatest widening near the propylamidinium end of the molecule (lower left in Figure 1b), and a bending of 8^{o} in the helix axis. In Hoechst, this drug-induced bending is superimposed on a prior natural bend of 19^{o} in the axis of the drug-free DNA, a bend that is roughly at right angles to that produced by the drug (7).

4. Discussion

The observed DNA binding behavior of Hoechst 33258 is explained very cleanly by the structure of the complex. Hoechst, like netropsin, binds within the minor groove, displacing the spine of hydration that is present in the free DNA (Figure 8). The covalent backbone of the drug molecules stabilizes the B helix conformation even more effectively than did the hydrogen-bonded spine. The DNA-drug complex is held together by a combination of specific bridging hydrogen bond to base edges, electrostatic attraction between the cationic drug and the anionic DNA, and hydrophobic interactions resulting from the burying of aromatic rings deep within a narrow cleft. Part of this hydrophobic stabilization undoubtedly is entropic in origin — created by the dispersal of water molecules that formerly sat in regular array down the minor groove.

Each drug makes the greatest perturbation in DNA structure at its lower end as drawn in Figure 8 — the piperazine ring for Hoechst and the propylamidinium tail for netropsin. The groove is opened up more at that end of the drug than at the other. Both the phenol ring of Hoechst and the guanidinium of netropsin are lifted up somewhat from the floor of the groove and are not fitted in as tightly. But for both drugs, the snug fit within the groove forces the helix axis to bend backwards by 7^{o} or 8^{o} in the region of binding.

Since the Hoechst and netropsin molecules are of similar size, it is not surprising that footprinting experiments show them to protect the same size binding site. Hoechst makes one fewer bridging hydrogen

bonds than does netropsin, which probably explains why its binding constant to DNA is lower by two orders of magnitude.

The observed requirement of A•T regions of DNA for binding of both Hoechst and netropsin (as well as a great number of other groove-binding drugs) arises not from anything as sophisticated as a hydrogen bond pattern, but from simple steric hindrance. The C2 amine of a guanine would not permit the drug to fit deeply enough into the groove to make a stable complex. Hence, the aromatic benzimidazole rings of Hoechst and pyrrole rings of netropsin are A•T-reading structural elements. In contrast, the piperazine ring of Hoechst can be regarded as a weak G•C-reading element. Steric hindrance about the bond to benzimidazole forces the piperazine ring to lie roughly crossways across the goove. This means that the ring is sufficiently far off the floor of the groove that a guanine C2 amine group no longer is an impediment; quite the contrary, it contributes an additonal hydrogen bond to the terminal nitrogen of the piperazine ring. Moreover, the crossways positioning of piperazine means that it cannot fit within the intrinsically narrower A•T regions of a B-DNA minor groove. It must have the added width that comes only with G•C regions. If such a wider groove is unavailable, then Hoechst can reorient about its central benzimidazole-benzimidazole bond, to yield an alternate conformation in which the piperazine ring extends out of the groove (Plate IB, Figure 4b). Piperazine evidently makes a relatively unimportant contribution to the binding of Hoechst, since even in the most favorable case as observed in this crystal structure analysis, the external-piperazine conformation is adopted by roughly

half the molecules in the crystal.

Hoechst 33258 is an interesting mixture of adaptation and maladaption to the geometry of B-DNA. The fluorescence behavior of the fused five- and six-ring benzimidazole molecule is what makes Hoechst useful as a chromosomal stain. The fused rings also create an asymmetric bond geometry that induces a 30^{o} bend from one unit to the next in a chain built by linked benzimidazoles. This is well-suited for following the minor groove as it winds its way up the DNA helix; the analogous bend for the [pyrrole-amide] unit in netropsin is 36^{o}. But this curvature in Hoechst is interrupted at the phenol; the -OH in the <u>para</u> position fails to curve and is held too far away from the floor of the groove for hydrogen bonding. In fact, as Figures 2b and 8c show, the phenol -OH takes the place of a <u>second</u> shell water in the spine of hydration, binding instead to another water molecule that itself makes a normal O2-to-N3 bridge across two bases. If the phenol -OH were in the <u>meta</u> position, then the 60^{o} bend introduced in the drug molecule would bring the -OH too close to the floor of the groove. At the other end of the Hoechst molecule, the piperazine ring is functionally similar to a benzimidazole, with nonpolar sides and polar, hydrogen-bonding keel. But its cross section is utterly different: broad and shallow rather than tall and thin. Hence, it requires a different kind of minor groove region for binding: the broader groove that is associated with G·C base pairs.

The preference of a puckered six-membered ring for G·C regions of DNA is also seen in other drug molecules, including some that intercalate between base pairs. In the complex of the antitumor drug

daunomycin with the B-DNA hexamer C-G-T-A-C-G (47), the aglycone ring of the drug intercalates between a C·G and a G·C base pair at each end of the helix, and the amino sugar lies flat in the bottom of the minor groove in a manner comparable to that of the piperazine in Hoechst, with even a hydrogen bond from an -OH of the aglycone ring A to guanine N2 and N3 atoms. Nonintercalative minor groove-binding drugs that prefer G·C-containing binding sites are rare, but chromomycin, mithramycin, and olivomycin are three related examples. Van Dyke and Dervan (55) have shown by footprinting studies that these drugs prefer at least two adjacent G·C base pairs. In the light of our Hoechst structure it is interesting to note that all three of these antibiotics contain multiple hexopyranose rings; one might expect to find them nested within the wider G·C regions of the minor groove, anchored down to guanine amines via hydrogen bonds, like the piperazine of Hoechst.

One might think that a long-chain polymer of benzimidazoles would be an efficient and tight binder to DNA. But another problem arises when one follows this line of reasoning: the repeating benzimidazole unit, like the repeating pyrrole-amide unit of netropsin, is 25% too long for a match with the repetition of base pairs along the floor of the minor groove (10). In a long polymer of benzimidazoles, the hydrogen-bonding N-H should quickly go out of registration with the base edge N3 and O2 atoms. This out-of-phasedness is clearly visible in Figure 5. If one considers the <u>first</u> atom of the Phe-Bz1, Bz1-Bz2, and Bz2-Pip bonds as marker atoms, then in Figure 5 the first marker atom (on Phe) is approximately on the same level with base pair

A6·T19, the second marker atom (on Bz1) is significantly below the plane of T7·A18, and the third marker atom (on Bz2) is halfway between T8·A17 and C9·G16. Clearly, the benzimidazole backbone of Hoechst 33258 is a poor model for longer DNA-bonding molecules.

When this helix repeat mismatch problem was considered with netropsin, a solution was found that involved replacing each -CO-NH- amide of a poly(amide-pyrrole) chain with a shorter -CO- or a -NH- (10). This made the repeat of the drug polymer match that of the DNA and still preserved the hydrogen-bonding properties necessary for strong DNA binding. A possible way of shortening a poly(benzimidazole) chain in similar fashion would be to alternate benzimidazole and simple imidazole rings, or to synthesize the repeating polymer: $(\text{-benzimidazole-benzimidazole-imidazole-})_n$. This in turn suggests an interesting analogue of Hoechst 33258 that would bind very tightly and specifically to AT regions of DNA and would exhibit the fluorescent behavior of Hoechst (Figure 9). Loewe and Urbanietz (37) observed that phenol could be replaced by another group as long as it remained aromatic (and, hence, presumably capable of delocalizing electrons into the Bz1 system). Replacement of phenol by imidazole would accomplish this and would also furnish one more positive charge and one more hydrogen bond donor, oriented (unlike the -OH of phenol) toward the floor of the groove. At the other end, Loewe and Urbanietz also found that piperazine could be replaced by another basic cation. Imidazole for piperazine would preserve both the charge and the hydrogen bond donor and would now be sufficiently narrow that it could fit down into the narrow groove of an AT region.

The compound in Figure 9, which might be called IBBI for imidazole-benzimidazole-benzimidazole-imidazole, should require a binding site of five successive A•T base pairs and should be a fluorescent DNA stain of considerable affinity for DNA.

5. Experimental Procedures

Synthesis of Oligonucleotide and Crystallization of Complex

The dodecadeoxynucleotide d(C-G-C-G-A-A-T-T-C-G-C-G) was synthesized by the liquid-phase phosphotriester method (4, 45). The protected dodecamer was deblocked, and an analytical HPLC run was used to verify removal of the blocking groups. Two passes of the deblocked dodecamer through an ion-exchange DEAE Sephadex A-25 column were sufficient to produce material that was pure enough to crystallize readily.

Crystals of the complex of DNA dodecamer plus Hoechst 33258 were grown by vapor diffusion from a solution containing 0.21 mM DNA, 0.21 mM Hoechst 33258, 5.38 mM $Mg(OAc)_2$, 0.21 mM spermine acetate (pH 7.0) and 20% 2-methyl-2,4-pentanediol (MPD). The DNA was observed to precipitate out of solution upon addition of the dye, forming a yellowish, oily mass that slowly went back into solution during the course of the crystallization. Crystals grew over a period of eight weeks, appearing after equilibration against a reservoir containing 35% MPD. The largest crystals were approximately 0.15 x 0.20 x 0.80 mm. Precession survey photography showed them to adopt orthorhombic space group $P2_12_12_1$, with cell dimensions: a = 25.04 A, b = 40.33 A, c = 65.85 A.

Data Collection and Structure Analysis

Three-dimensional x-ray diffraction data from the DNA/drug complex crystals were collected on a Syntex P1-Bar diffractometer

using omega scan geometry, with graphite monochrometered Cu-Ka radiation. Data were collected to a resolution of 2.2 A, with approximately 45% of the 3487 unique reflections being observed above the two-sigma level.

The structure of the DNA/drug complex was solved by molecular replacement, using the final coordinates of the native d(C-G-C-G-A-A-T-T-C-G-C-G) as a starting model (8). The residual error or crystallographic R factor between the observed and calculated reflections was 46.6% (zero-sigma, 8.0-2.2 A resolution) prior to refinement.

These initial DNA coordinates then were adjusted, without any assumption as to the locaton or even the presence of a drug molecule, during four successive stages of refinement at increasing resolution, using the Jack-Levitt (18) restrained refinement method. The DNA was refined to convergence at each stage, adjusting atomic coordinates but holding all temperature factors uniform at 20.0. Only when coordinate refinement was complete at 2.2 A resolution were temperature factors also set free to refine to convergence. The R factor or residual error at this stage was 31.7%.

At this point, solvent peaks were clearly visible in the Fourier or electron density map, along with a putative Hoechst 33258 molecular image centered within the minor groove. In order to improve the phasing as much as possible, and hence to produce the best possible image of the drug molecule for examination, 40 solvent peaks were added to the structure during several further rounds of combined positional and temperature factor refinement, during which the R

factor fell to 28.0%. No solvent peaks were added, however, in the region where the drug moleucle was thought to lie, in order not to confuse the interpretation of the drug image.

This portion of the refined electron density map then was examined on an Evans and Sutherland Picture System II graphics station, using the molecular model-building program FRODO (21). The otherwise unexplained electron density down the minor groove did indeed have the shape expected for an image of the Hoechst 33258 molecule at 2.2 A resolution, and the framework of the drug molecule was built into the density. Refinement, now of the complete DNA/drug complex, continued for several more rounds of refinement and examination of difference maps to pick out more solvent peaks, until the point was reached where it was judged that the remaining peaks visible in the difference map fell below the noise level of the data. The final structure contained one DNA double-helical dodecamer, one Hoechst 33258 drug molecule, and 175 solvent peaks. The final residual error was 0.140 for two sigma data (0.191 for all data), for reflections between 8.0 A and 2.2 A resolution.

As with all structures from our laboratory, both the original x-ray diffraction data and the final refined coordinates have been put in the public domain by depositing them with the Brookhaven Protein Data Bank.

Acknowledgements

We would like to thank David Goodsell for advice and help with the color graphics system, and Mary L. Kopka for extensive discussions

about the Hoechst vs. netropsin comparisons. This work was supported by the American Cancer Society Grant No. NP-504, and NIH Grant GM-31299.

References

1. Bontemps, J., Houssier, C. and Fredericq, E. (1975) Nucl. Acids Res. 2, 971-984.
2. Chen, T. R. (1977) Exptl. Cell Res. 104, 255-262.
3. Comings, D. E. (1975) Chromosoma (Berl.) 52, 229-243.
4. Denney, W. A., Leupin, W. and Kearns, D. R. (1982) Helvetica Chim. Acta 65, 2372-2393.
5. Dickerson, R. E. and Drew, H. R. (1981) J. Mol. Biol. 149, 761-786.
6. Dickerson, R. E., Kopka, M. L. and Pjura, P. (1983) Proc. Natl. Acad. Sci. USA 80, 7099-7103.
7. Drew, H. R. and Dickerson, R. E. (1981) J. Mol. Biol. 151, 535-556.
8. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. (1981) Proc. Natl. Acad. Sci. USA 78, 2179-2183.
9. Fratini, A. V., Kopka, M. L., Drew, H. R. and Dickerson, R. E. (1982) J. Biol. Chem. 257, 14686-14707.
10. Goodsell, D. and Dickerson, R. E. (1986) J. Med. Chem. 29, 727-733.
11. Harshman, K. D. and Dervan, P. B. (1985) Nucl. Acids Res. 13, 4825-4835.
12. Herzog, A. and Schutze, H. R. (1968) Deutsche Tierarztliche Wochenschrift 19, 476-478.
13. Hilwig, I. and Gropp, A. (1972) Exptl. Cell Res. 75,

122-126.

14. Hilwig, I. and Gropp, A. (1973) Exptl. Cell Res. 81, 474-477.

15. Hirschberg, J., Lavi, U., Goitein, R. and Marcus, M. (1980) Exptl. Cell Res. 130, 63-72.

16. Holmquist, G. (1975) Chromosoma (Berl.) 49, 333-356.

17. Holmquist, G. P. and Comings, D. E. (1975) Chromosoma (Berl.) 52, 245-259.

18. Jack, A. and Levitt, M. (1978) Acta Cryst. A34, 931-935.

19. Jeffrey, G. A. and Maluszynska, H. (1982) Int. J. Biol. Macromol. 4, 173-185.

20. Jennings, B. R. and Ridler, P. J. (1983) Biophys. Struct. Mech. 10, 71-79.

21. Jones, T. A. (1978) J. Appl. Crystallog. 11, 268-272.

22. Jurnak, F. A. and McPherson, A., eds. Biological Macromolecules and Assemblies. Vol. 2: Nucleic Acids and Interactive Proteins (Wiley-Interscience, New York, 1985) Appendix, pp. 471-494.

23. Kopka, M. L., Fratini, A. V., Drew, H. R. and Dickerson, R. E. (1983) J. Mol. Biol. 163, 129-146.

24. Kopka, M. L., Pjura, P. E., Goodsell, D. S. and Dickerson, R. E. In Nucleic Acids and Molecular Biology (D. Lilley and F. Eckstein, eds.), Vol. 1 (Springer-Verlag, Berlin, 1986).

25. Kopka ,M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) Proc. Natl. Acad. Sci. USA 82, 1376-1380.

26. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson,

R. E. (1985) J. Mol. Biol. 183, 553-563.

27. Kroll, W. (1984) Histochem. 80, 493-496.

28. Kucherlapati, R. S., Hilwig, I., Gropp, A. and Ruddle, F. H. (1975) Humangenetik 27, 9-14.

29. Kusyk, C. J. and Hsu, T. C. (1979) Cytogenet. Cell Genet. 23, 39-43.

30. Laemmler, G., Herzog, H., Saupe, E. and Schutze, H. R. (1971) Bull. WHO 44, 751-756.

31. Laemmler, G. and Raether, W. (1968) Eighth Int. Cong. on Trop. Med. Malaria, Teheran.

32. Langlois, R. G. and Jensen, R. H. (1979) J. Histochem. Cytochem. 27, 72-79.

33. Latt, S. A. (1973) Proc. Natl. Acad. Sci. USA 70, 3395-3399.

34. Latt, S. A. and Stetten, G. (1976) J. Histochem. Cytochem. 24, 24-33.

35. Latt, S. A. and Wohlleb, J. C. (1975) Chromosoma (Berl.) 52, 297-316.

36. Loewe, H. and Urbanietz, J. (1968) Eighth Int. Cong. on Trop. Med. Malaria, Teheran.

37. Loewe, H. and Urbanietz, J. (1974) Arzneimittel-Forschung 24, 24-33.

38. Loucks, E., Chaconas, G., Blakesley, R. W., Wells, R. D. and van de Sande, J. H. (1979) Nucl. Acids Res. 6, 1869-1879.

39. Makarov, V. L., Poletaev, A. I., Sveshnikov, P. G., Knodrat'eva, N. O., Pis'menskii, V. F., Doskocil, J., Kondelka, J. and Vol'kenshtein, M. V. (1979) Mol. Biol. (USSR) 13, 339-352.

40. Markus, M., Nielsen, K., Goitein, R. and Gropp, A. (1979) Exptl. Cell Res. 122, 191-201.

41. Martin, R. F. and Holmes, N. (1983) Nature 302, 452-454.

42. Mikhailov, M. V., Zasedatelev, A. S., Krylov, A. S. and Gurskii, G. V. (1981) Mol. Biol. (USSR) 15, 541-554.

43. Mueller, W. and Gautier, F. (1975) Eur. J. Biochem. 54, 385-394.

44. Mueller, W., Hattesohl, I., Schuetz, H. J. and Meyer, G. (1981) Nucl. Acids Res. 9, 95-119.

45. Narang, S. A. (1983) Tetrahedron 39, 3-22.

46. Perry, P. and Wolff, S. (1974) Nature 251, 156-158.

47. Quigley, G. J., Wang, A. H-J., Ughetto, G., van der Marel, G., van Boom, J. H. and Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 7204-7208.

48. Raposa, T. and Natarajan, A. T. (1974) Humangenetik 21, 221-226.

49. Rosenstein, B. S. (1982) Photochem. Photobiol. 35, 163-166.

50. Sahar, E. and Latt, S. A. (1980) Chromosoma (Berl.) 79, 1-28.

51. Stetten, G., Davidson, R. L. and Latt, S. A. (1977) Exptl. Cell Res. 108, 447-452.

52. Stetten, G., Latt, S. A. and Davidson, R. L. (1976) Somatic Cell Genet. 2, 285-290.

53. Stokke, T. and Steen, H. B. (1985) J. Histochem. Cytochem. 33, 333-338.

54. Taylor, R., Kennard, O. and Versichel, W. (1984) J. Am. Chem.

Soc. 106, 244-248.

55. Van Dyke, M. W. and Dervan, P. B. (1983) Biochem. 22, 2373-2377.

56. Weisblum, B. and Haenssler, E. (1974) Chromosoma (Berl.) 46, 255-260.

Table 1. Mean Helical Parameters and Changes Produced by Drug Binding

Parameter	DNA Helix Alone		DNA-drug complex	
	Native	MPD7	Netropsin	Hoechst 33258
t_g = helical twist	$35.8^o(4.2^o)$	$36.0^o(4.7^o)$	$36.0^o(3.5^o)$	$35.8^o(3.8^o)$
n = base pairs /turn	10.0	10.0	10.0	10.0
hg = rise per base pair	3.34(0.44)A	3.37(0.52)A	3.33(0.54)A	3.39(0.44)A
Base pair inclination to helix axis	$-1.5^o(7.3^o)$	$+2.0^o(4.6^o)$	$-2.9^o(5.9^o)$	$-2.6^o(4.5^o)$
Base pair displacement from helix axis	-0.17(0.95)A	-0.24(0.41)A	-0.46(0.94)A	-0.10(1.07)A
Propeller twist	$13.0^o(5.1^o)$	$17.7^o(6.4^o)$	$15.2^o(7.1^o)$	$18.2^o(5.7^o)$
Torsion angle delta:				
Purines	$128.4^o(20.8^o)$	$131.2^o(25.0^o)$	$123.6^o(19.6^o)$	$136.1^o(18.7^o)$
Pyrimidines	$117.0^o(19.8^o)$	$107.5^o(24.7^o)$	$105.8^o(20.8^o)$	$115.5^o(27.9^o)$

Notes: The Native DNA structure has the sequence C-G-C-G-A-A-T-T-C-G-C-G, and has a 19^o sideways bend in the overall helix axis (5). The MPD7 structure has 5-bromocytosine in place of cytosine at the 9th position and has an unbent helix axis (9). Netropsin was cocrystallized with the bromocytosine derivative, whose helix axis remained straight, whereas Hoechst 33258 was cocrystallized with the native DNA sequence, and the helix retained the sideways bend that is visible in Figure 2. In addition, the binding of each drug caused the helix axis to bend backward at the binding site by 7^o or 8^o. For more precise definitions of helix parameters, see Reference 9 and the Appendix to Reference 22.

Figure Captions

Figure 1. (a) Hoechst 33258 and (b) netropsin. Note the similarity between the two DNA-binding drugs: both are crescent-shaped molecules with flat aromatic rings and consequently would fit well into the minor groove of DNA. Both have hydrogen bond-forming N-H groups along their concave edge and an overall positive charge (Subscripts on nitrogens are identifying labels, not chemical symbols).

Figure 2. Hydration and drug binding within the minor groove of the B-DNA dodecamer of sequence: C-G-C-G-A-A-T-T-C-G-C-G. View is directly into the minor groove, with base pair C1·G24 at top and G12·C13 at bottom. (a) DNA alone, with the oxygen atoms of the string of water molecules or spine of hydration down the minor groove shown as crossed spheres. (b) Complex of DNA with Hoechst 33258, with the drug molecule in the lower half of the minor groove, and the undisplaced spine of hydration in the upper half.

Figure 3. Schematics of (a) the spine of hydration, and (b) Hoechst 33258 with the residual spine, within the minor groove of C-G-C-G-A-A-T-T-C-G-C-G. Distances are from water oxygens or drug nitrogens, to adenine N3 or thymine O2 of DNA base pairs. Distances less than 4 A are dashed, and distances greater than 4 A are dotted. The C12 3'-OH group at the top of each diagram is from a neighboring DNA molecule.

Figure 4. Stereo views depicting the alternate position of the piperazine ring in the DNA-Hoechst complex. (a) Piperazine buried

within the minor groove. (b) Piperazine extended out, away from the double helix. Transitions between these two states must involve a 180^{o} rotation about the bond connecting the two benzimidazoles, and hence must take place when the drug is not bound to DNA.

Figure 5. Least-squares superposition of the DNA-Hoechst complex (dark bonds) on the structure of the DNA alone (open bonds). Minor groove opens toward right. Six key phosphorus atoms are numbered: 9 - 10 - 11 and 18 - 19 - 20, in the region of maximum widening of minor groove upon binding of Hoechst molecule. For an equivalent drawing looking into the minor groove, see Figure 4b of Reference 24.

Figure 6. Width of the minor groove opening, defined as shortest phosphorus-phosphorus distances, less 5.8 A representing two phosphate group radii. Phosphate numbering at bottom corresponds to that of Figure 5. The widening of the groove produced by Hoechst binding at P10-P19 and P11-P18 occurs in the vicinity of the second benzimidazole and piperazine rings and is easily seen in Figure 5.

Figure 7. Base plane normal vector plot, showing the bend in helix axis produced by the binding of Hoechst 33258. To understand this plot, imagine that the DNA dodecamer helix sits with its axis normal to the plane of the page, and with its minor groove (toward you in Figure 2) facing downward, along the positive y axis. The viewing direction of Figure 5 then is in from the left in this plot. Pip-Bz2-Bz1-Phe indicate locations of the four components of Hoechst as they wind around the helix axis. The x and y axes are components of the changes in direction cosines of the normal vectors to base pairs 1 through 12, changes that result from the binding of Hoechst.

The observation that the normal vector points for base pairs 1 through 6 cluster around one position marked by +, and that those for base pairs 7 through 12 cluster around a different position +, means that the helix axis is bent backward at its center. The distance between the two points +, roughly 0.12, indicates a bending of $\sin^{-1}(0.12) = 7^{o}$. The bending is in a direction away from Bz2, the second benzimidazole. For normal vector plots of other variants of the helix C-G-C-G-A-A-T-T-C-G-C-G, see References 6 and 26.

Figure 8. Summary schematics of the minor groove in free B-DNA of sequence: C-G-C-G-A-A-T-T-C-G-C-G, and in its complexes with two drug molecules. Sugar-phosphate backbones are the curved sides of a ladder, with rungs for base pairs. Pyrimidine O2 and purine N3 (N2 for guanines) are circled on the rungs. Letters W indicate water oxygens. Dotted lines mark hydrogen bonds, although some of these are quite long by conventional standards. (a) Spine of hydration in the drug-free B-DNA helix. (b) Complex with netropsin, in which amide N-H groups replace the first shell waters of the spine. (c) Complex with Hoechst 33258, with a residual continuation of the spine of hydration beyond the drug-binding site.

Figure 9. Synthetic analogue of Hoechst 33258 that should show very strong DNA binding and AT-specificity, while retaining the fluorescent properties of Hoechst. The rightmost imidazole retains the aromatic character of phenol, but provides another positive charge and another potential hydrogen-bonding group. The leftmost imidazole retains the H-bonding character and positive charge of N-methyl piperazine, but is narrow enough to slip into an AT region of the

minor groove.

Plate I. Computer graphics depiction of space-filling models of Hoechst 33258 bound to C-G-C-G-A-A-T-T-C-G-C-G. Colored dots define the van der Waals nonbonded surfaces. Stick bonds delineate the skeleton within. In most cases only a relatively narrow slab of structure is shown, in order to focus on one feature and keep the drawing comprehensible.

A. Skeletal drawing of the DNA double helix, with space-filling dot representation of the Hoechst molecule within the minor groove. Red atom at upper end of Hoechst is the oxygen of the phenol -OH. Blue bands along the drug are the exposed benzimidazole nitrogens.

B. Oblique cross section through the Hoechst-DNA complex, in the plane of the Hoechst molecule. The Hoechst phenol and two benzimidazoles are in light violet; the piperazine ring is shown in orange in its in-groove position, and in blue in its out-of-groove position. The DNA base pairs beneath the drug are shown in oblique cross section. The upper edge of the DNA image follows the floor of the minor groove; the cavity at the bottom defines a cross section across the major groove, with phosphate groups pendant to either side.

C. Cross section of the minor groove, through the phenol ring. The phenol ring has a purple skeleton with white dots outlining the van der Waals nonbonded contact surface. The DNA has a white skeleton with blue surface dots. The minor groove opens up and slightly to the right. Its opening is flanked by phosphate groups, with five-membered sugar rings below and base pairs in oblique cross section beneath.

Note that the phenol ring does not lie parallel to the walls of the groove, and is not fitted snugly into the bottom of the groove.

D. Cross section of the minor groove, through the first benzimidazole ring (purple, with white surface-defining dots). The benzimidazole ring sits snugly within a narrow groove that is barely wide enough to accommodate it.

E. Cross section of the minor groove, through the piperazine ring. The best mean plane through the puckered piperazine ring (purple) would make an angle of roughly 60^{o} with a plane bisecting the minor groove. The piperazine can be accommodated in this twisted orientation only because G·C regions of the minor groove are intrinsically wider in B-DNA than are A·T regions. Phosphates and sugars are at the top of the two walls of the groove; base pairs are below.

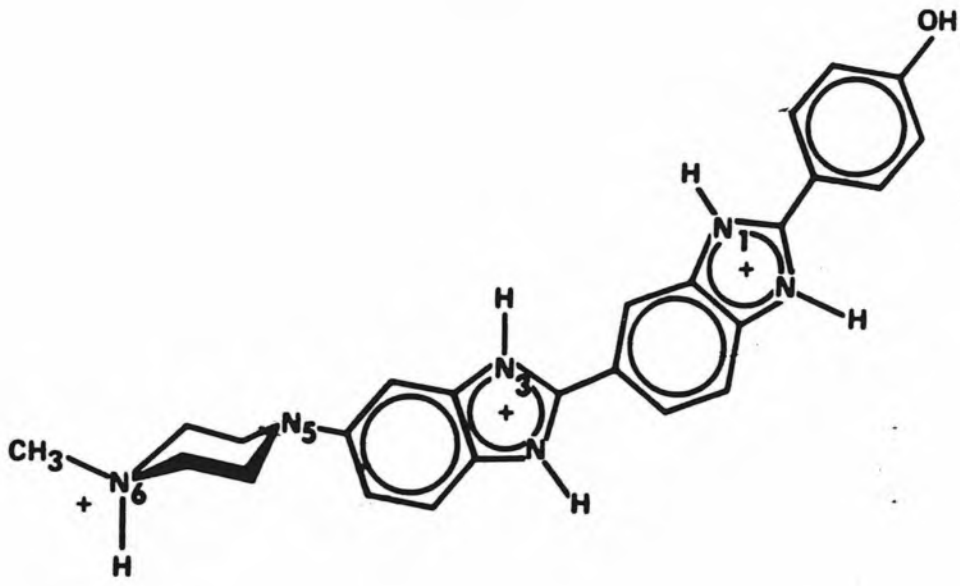


Figure 1a

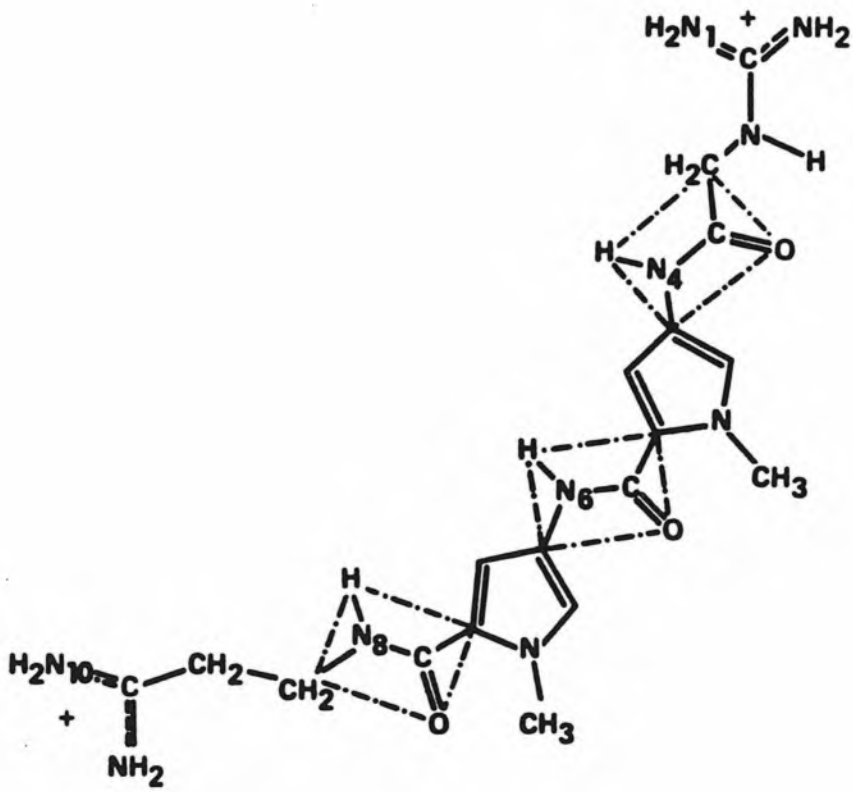


Figure 1b

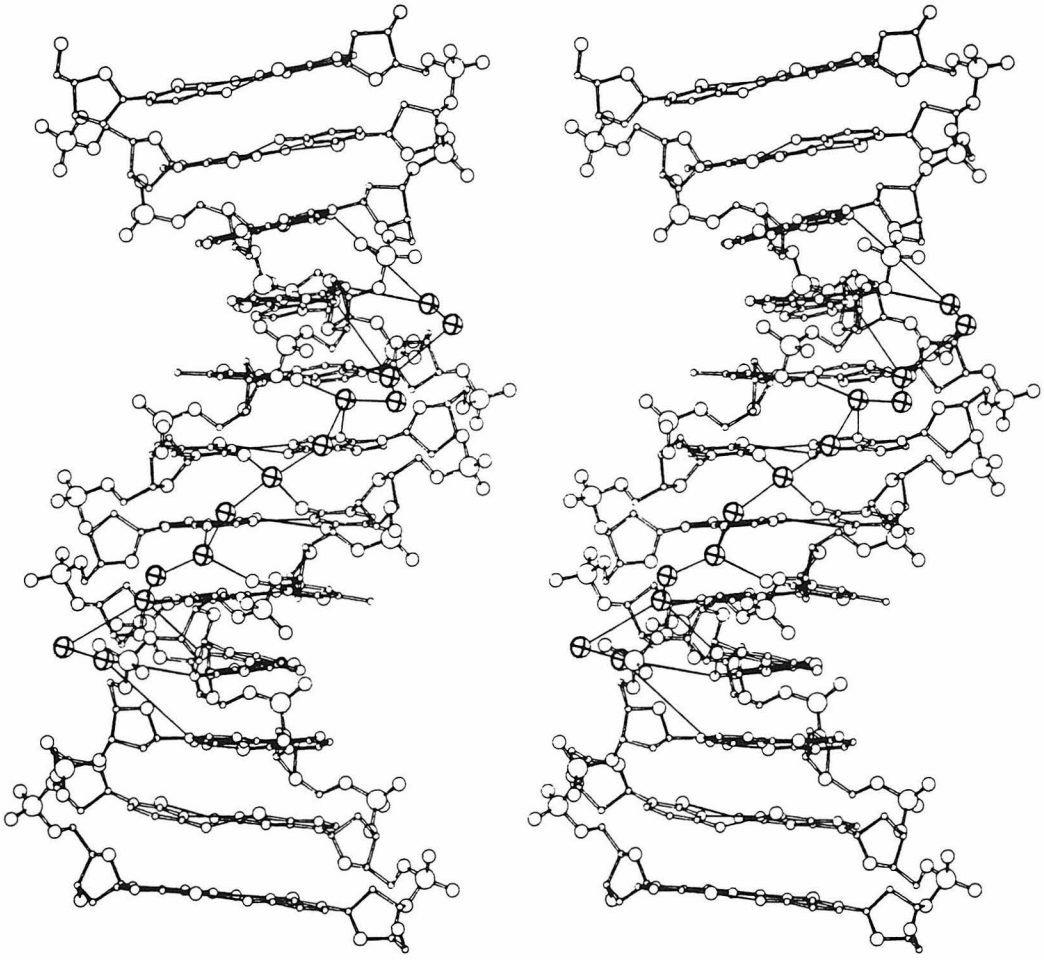


Figure 2a

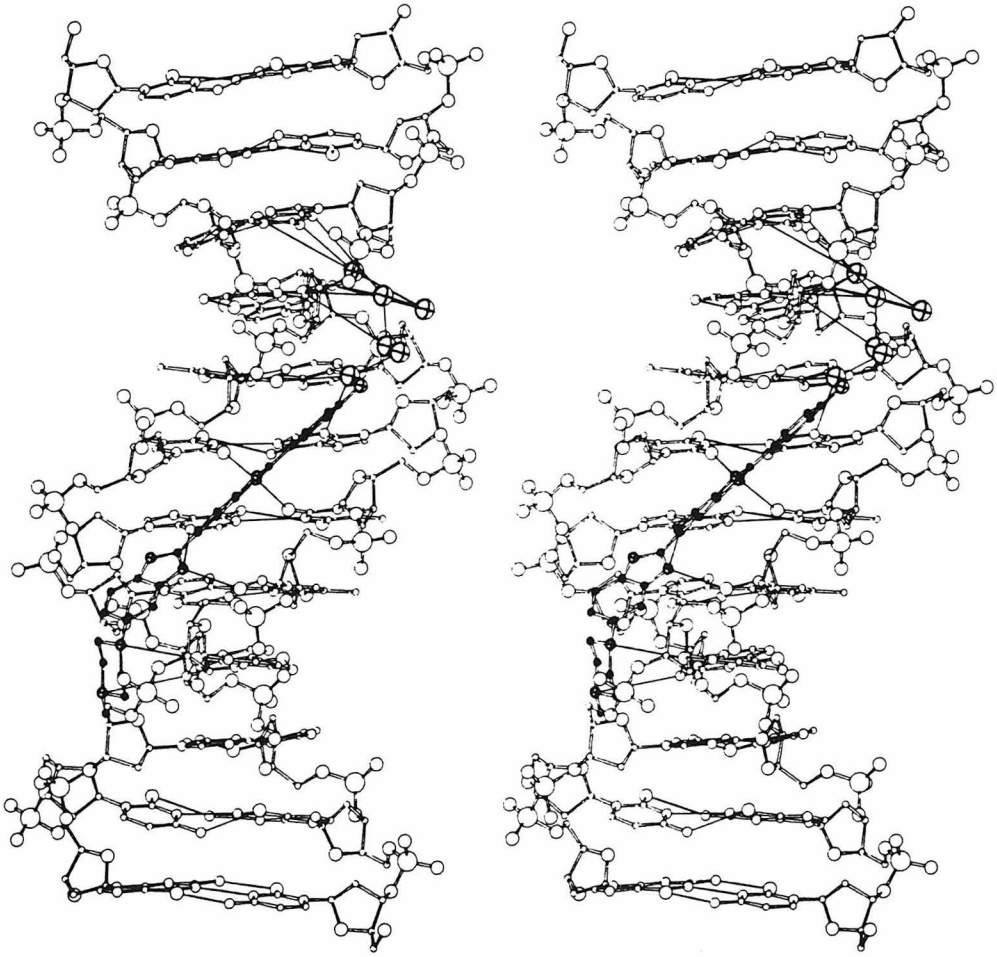


Figure 2b

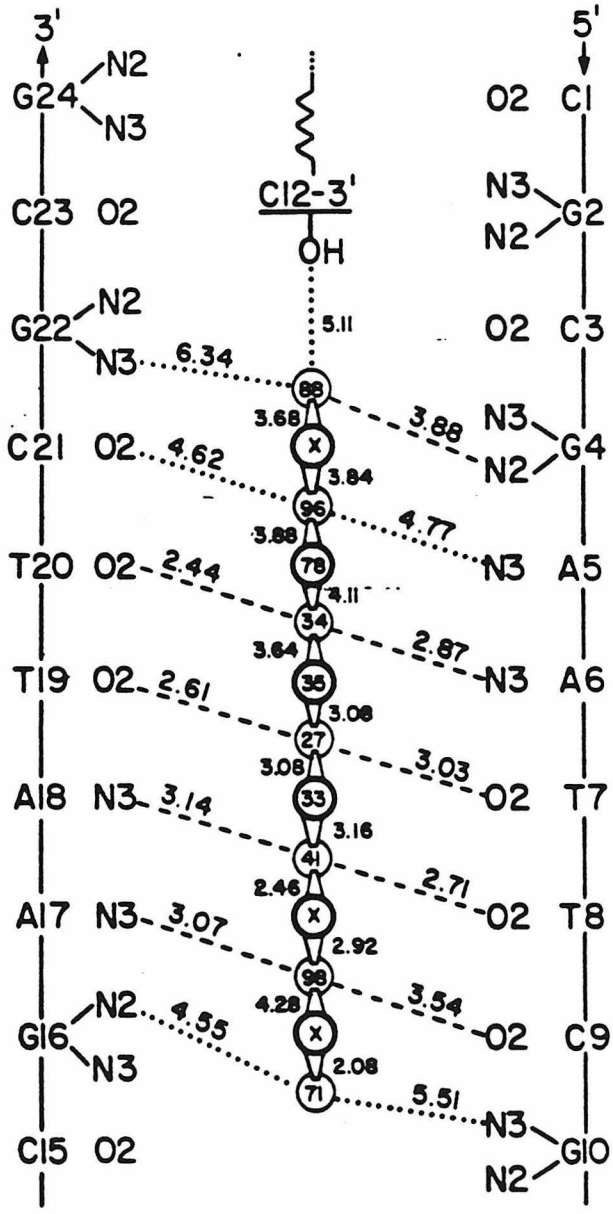

3'
5'
G24 N2 N3
C23 O2
G22 N2 N3
C21 O2
T20 O2
T19 O2
A18 N3
A17 N3
G16 N2 N3
C15 O2
C12-3'
OH
5.11
6.34
3.68
4.62
3.84
3.88
4.77
3.88
2.44
4.11
3.64
2.87
2.61
3.08
3.08
3.03
3.14
3.16
2.46
2.71
3.07
2.92
4.28
3.54
4.55
2.08
5.51
O2 C1
N3 G2 N2
O2 C3
N3 G4 N2
N3 A5
N3 A6
O2 T7
O2 T8
O2 C9
N3 G10 N2

Figure 3a

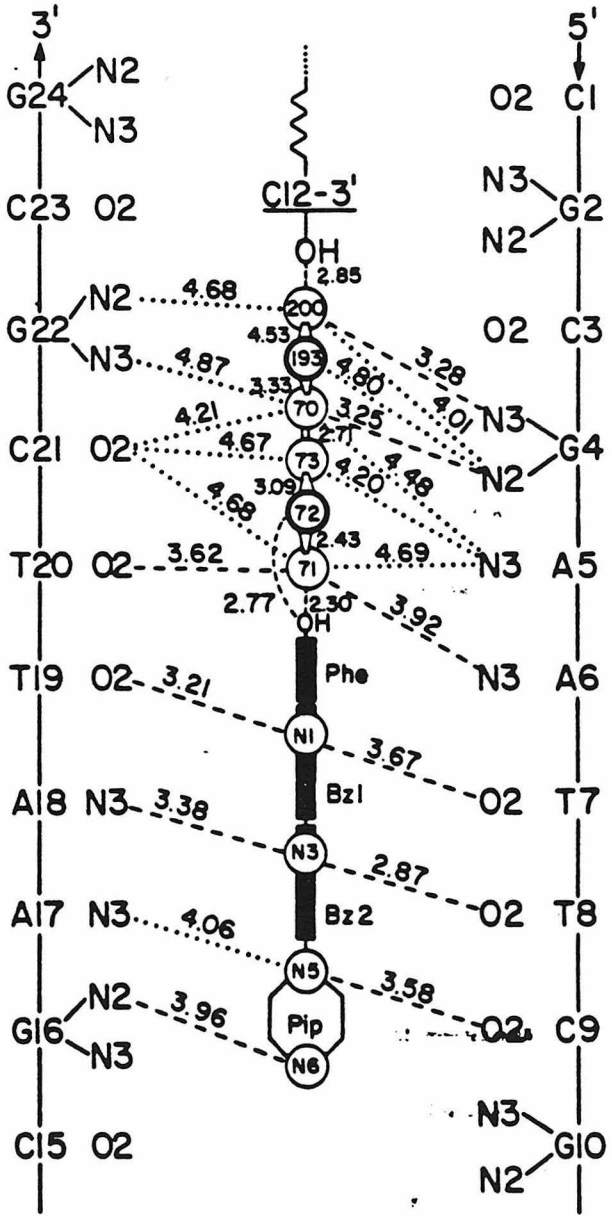

3'
5'
G24 N2
N3
C23 O2
G22 N2
N3
C21 O2
T20 O2
T19 O2
A18 N3
A17 N3
G16 N2
N3
C15 O2
C12-3'
OH
2.85
200
193
70
73
72
71
OH
Phe
N1
Bz1
N3
Bz2
N5
Pip
N6
4.68
4.53
4.87
3.33
4.21
4.67
3.09
3.62
2.77
2.30
3.21
3.38
4.06
3.96
3.28
4.60
3.25
2.71
4.20
4.69
3.92
3.67
2.87
3.58
O2 C1
N3 G2
N2
O2 C3
N3 G4
N2
N3 A5
N3 A6
O2 T7
O2 T8
O2 C9
N3 G10
N2

Figure 3b

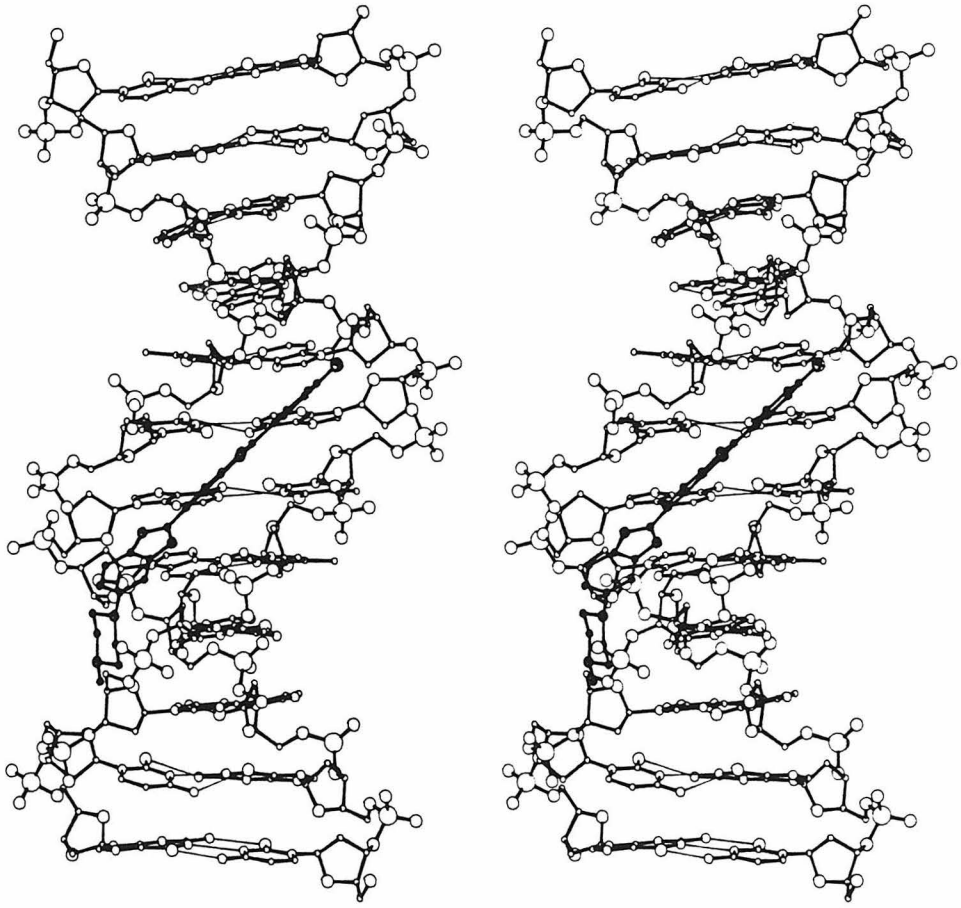


Figure 4a

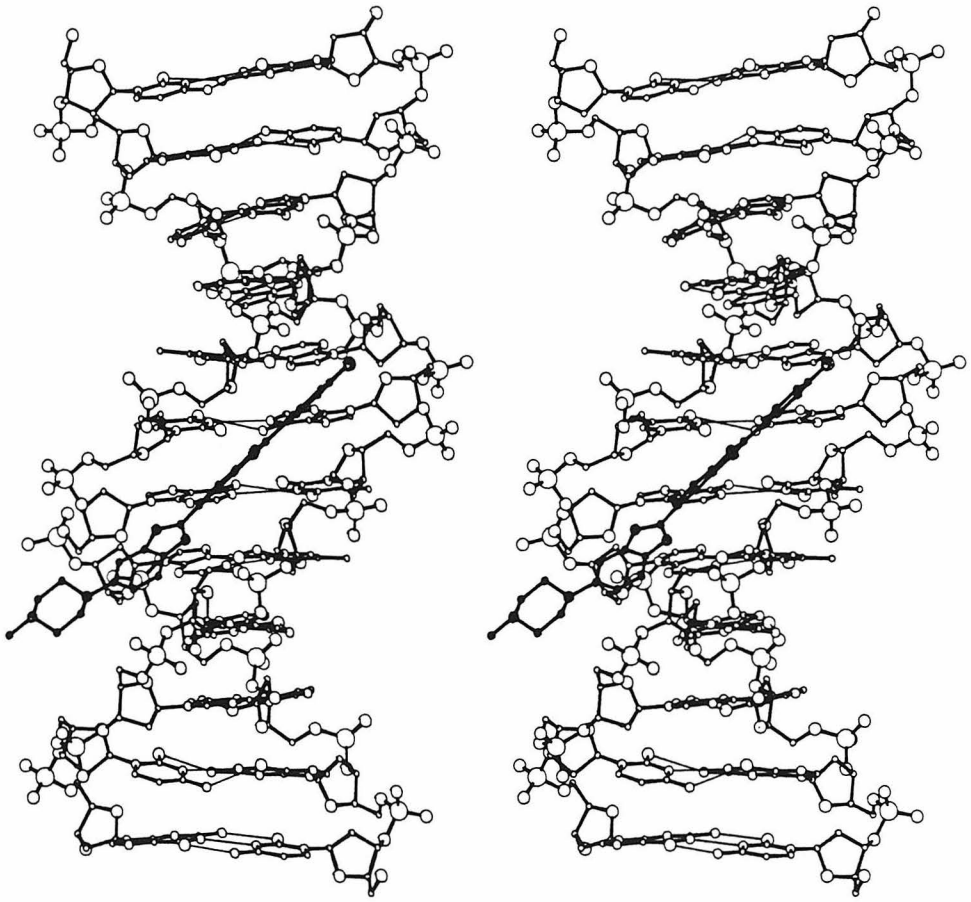


Figure 4b

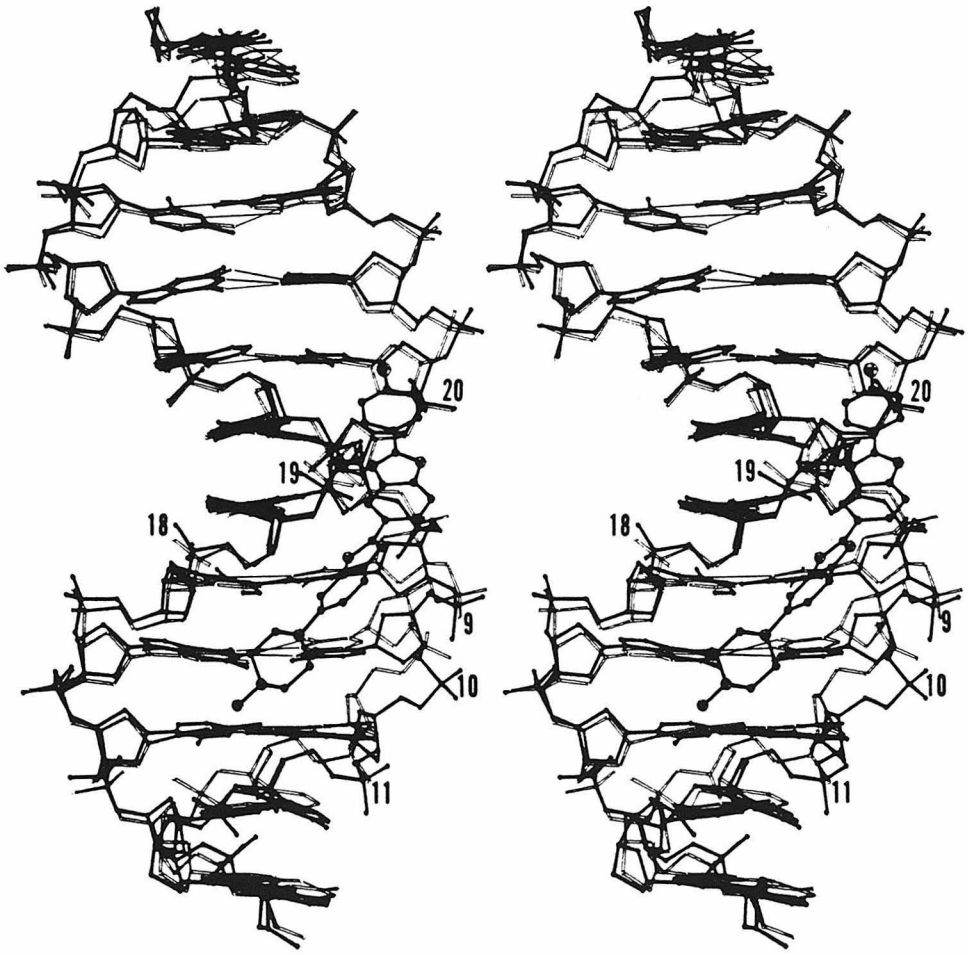


Figure 5

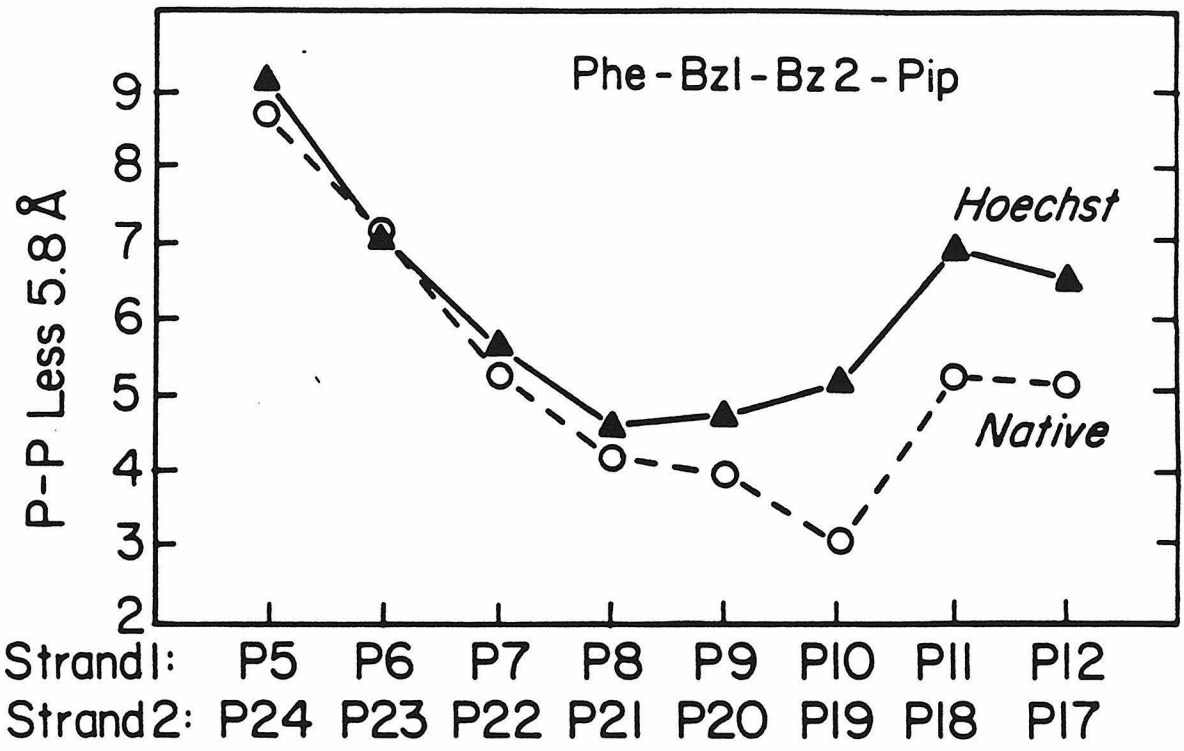

Phe - Bzl - Bz 2 - Pip
Hoechst
Native
P-P Less 5.8 Å
9
8
7
6
5
4
3
2
Strand1: P5 P6 P7 P8 P9 P10 P11 P12
Strand2: P24 P23 P22 P21 P20 P19 P18 P17

Figure 6

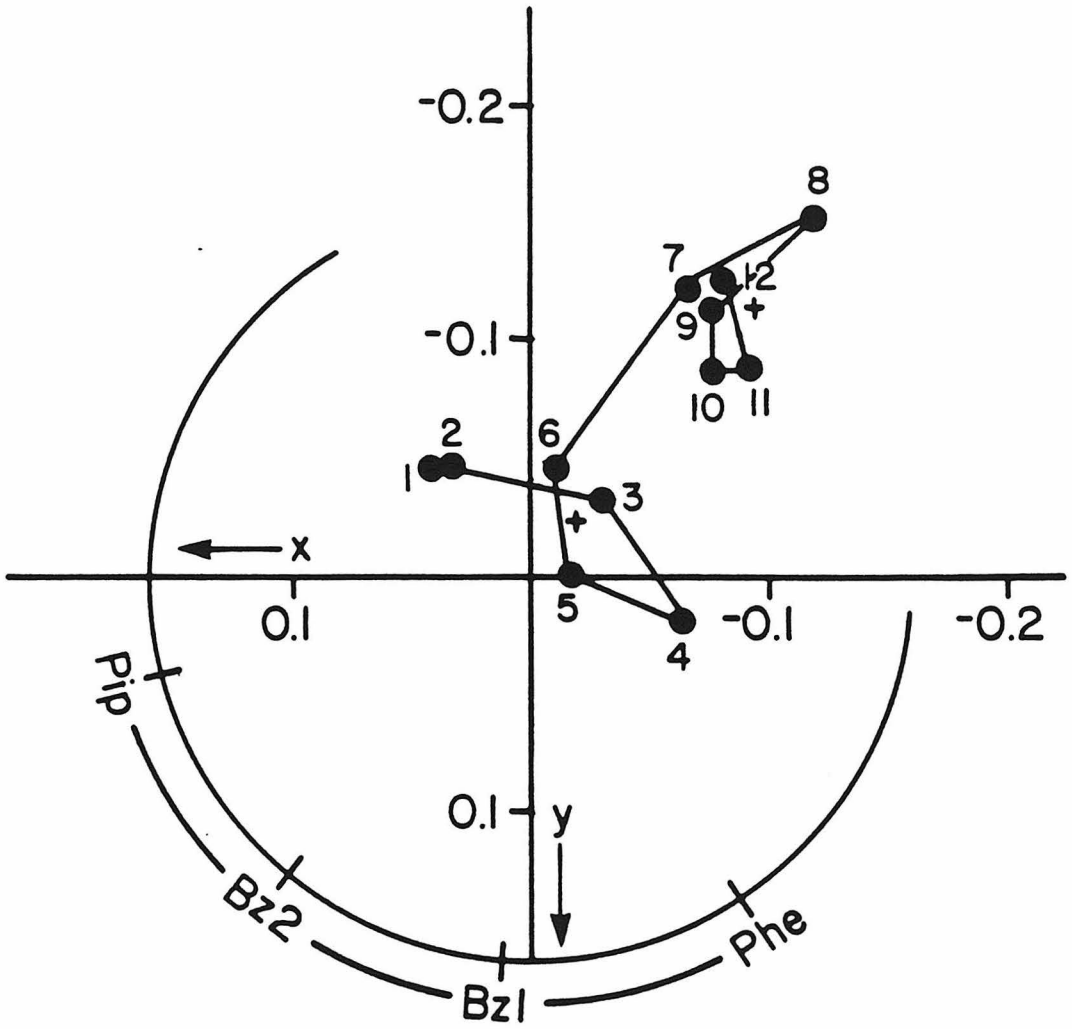


Figure 7

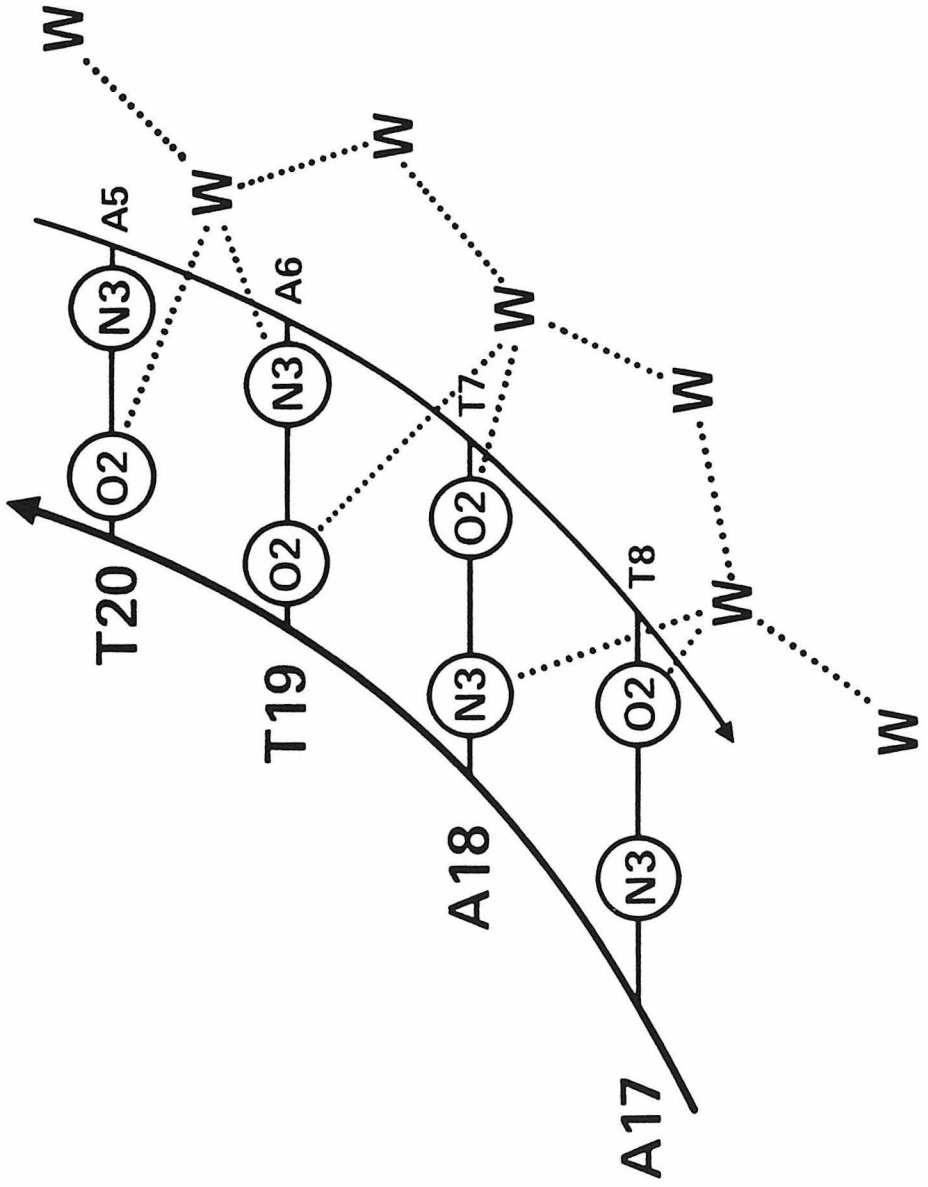


Figure 8a

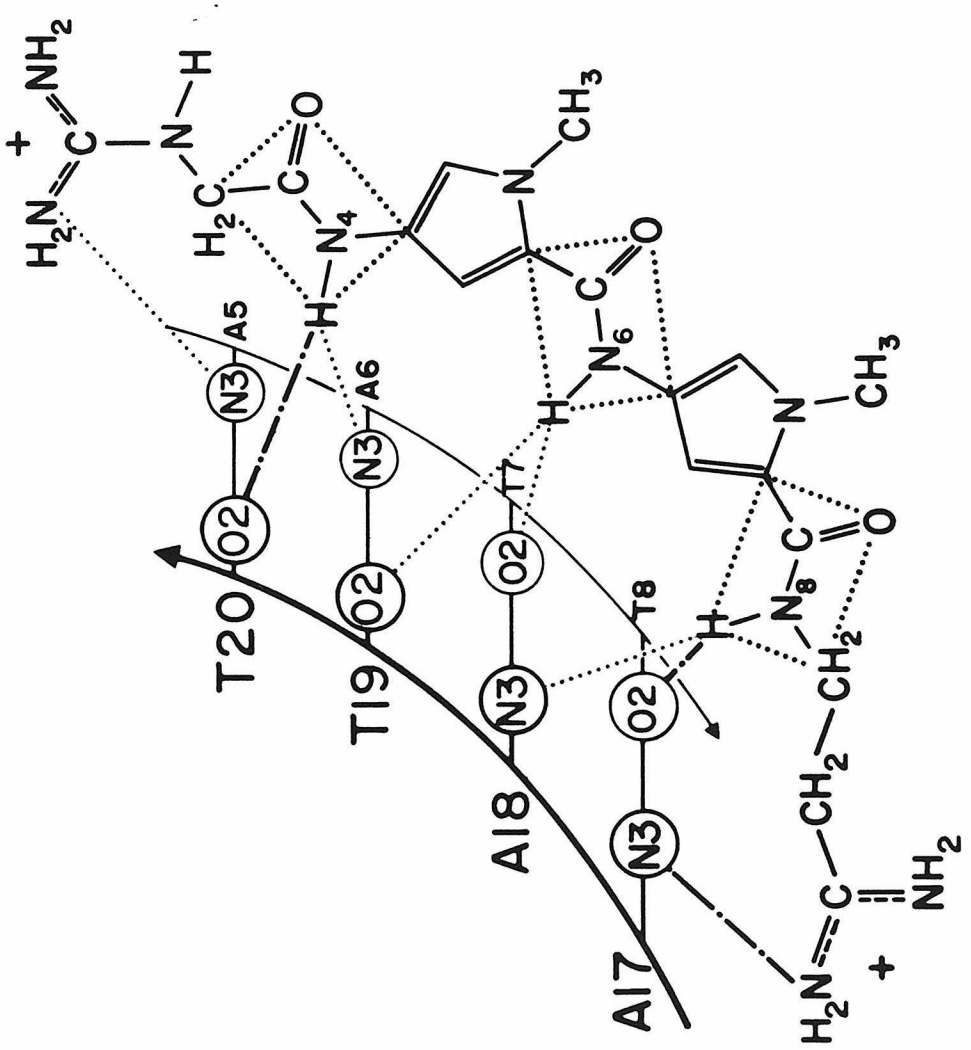


Figure 8b

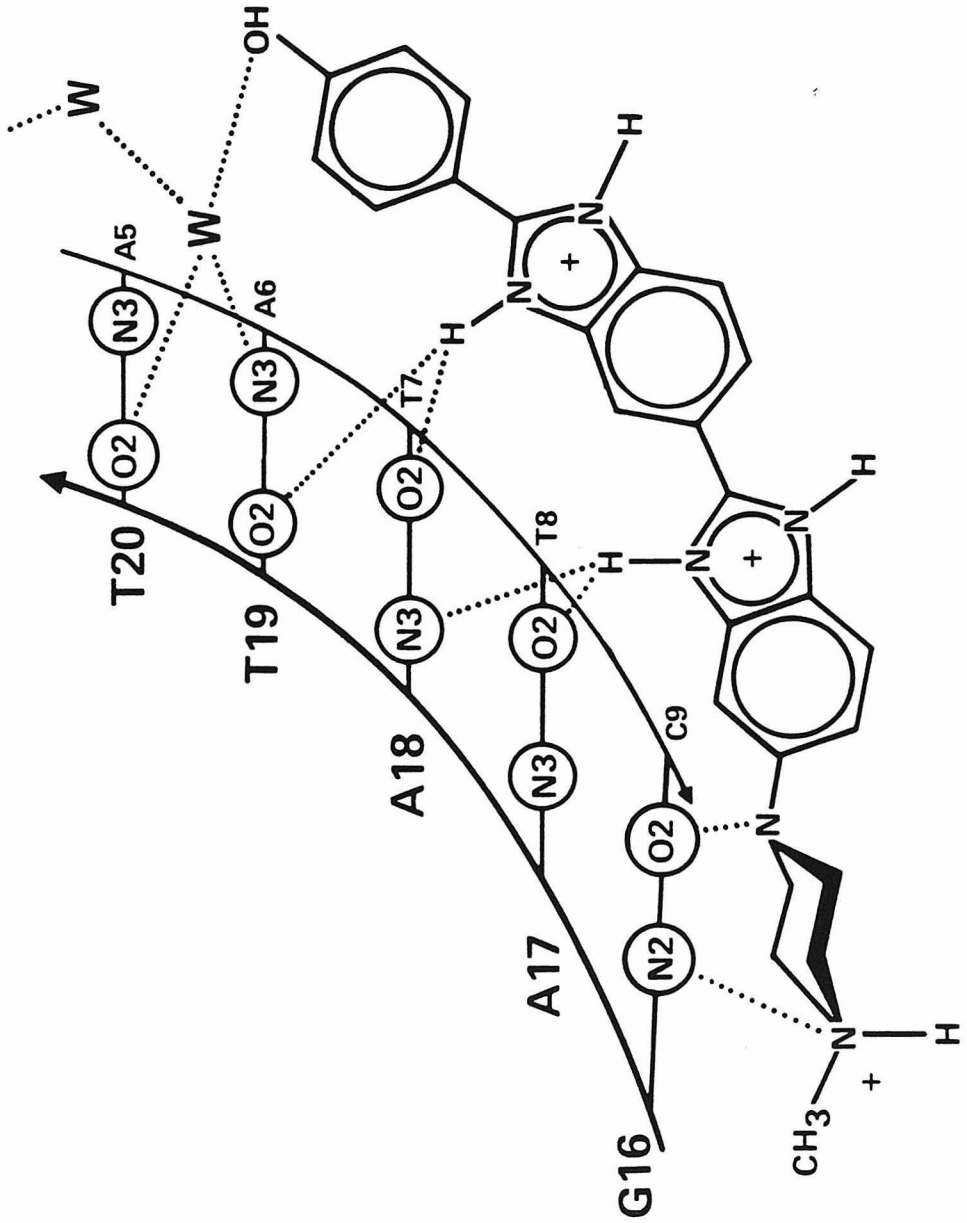

W
W
OH
A5
A6
N3
N3
O2
O2
T20
T19
T7
O2
N3
A18
T8
O2
N3
A17
G16
C9
O2
N2
H
N
+
CH3

Figure 8c

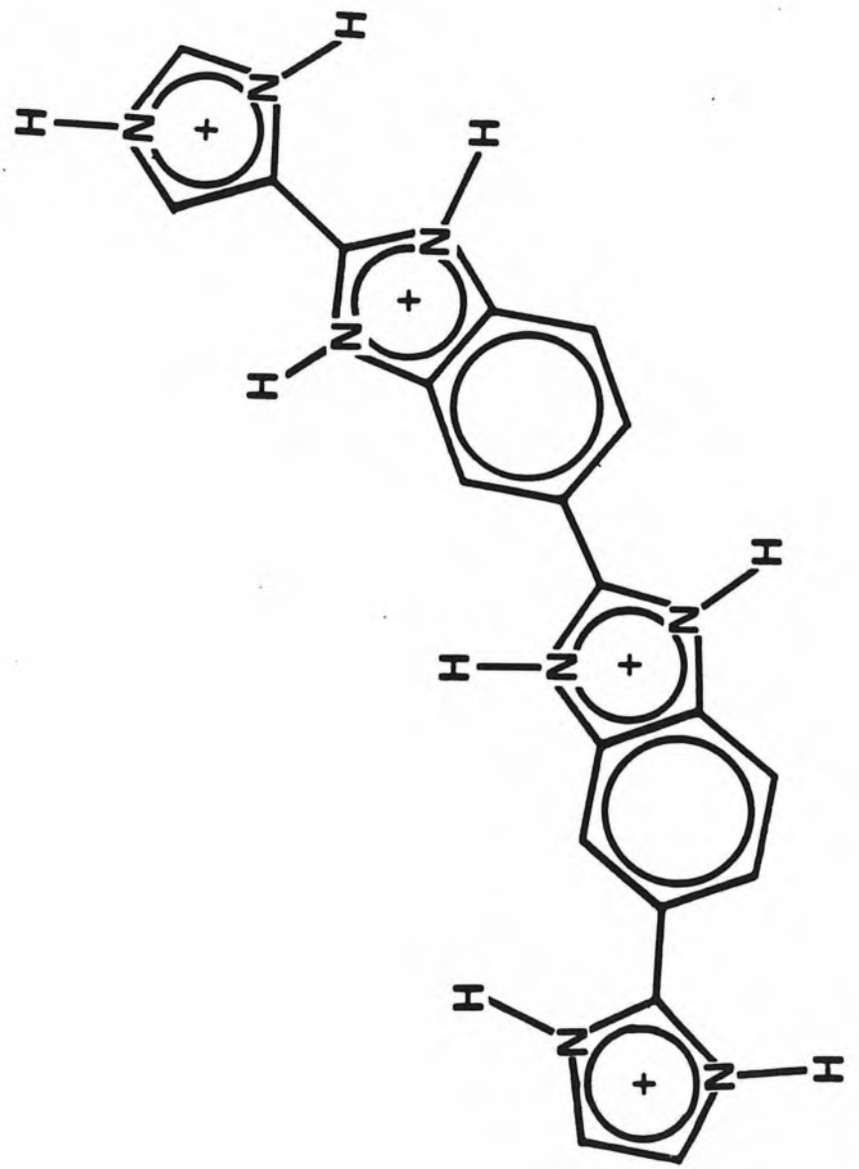


Figure 9

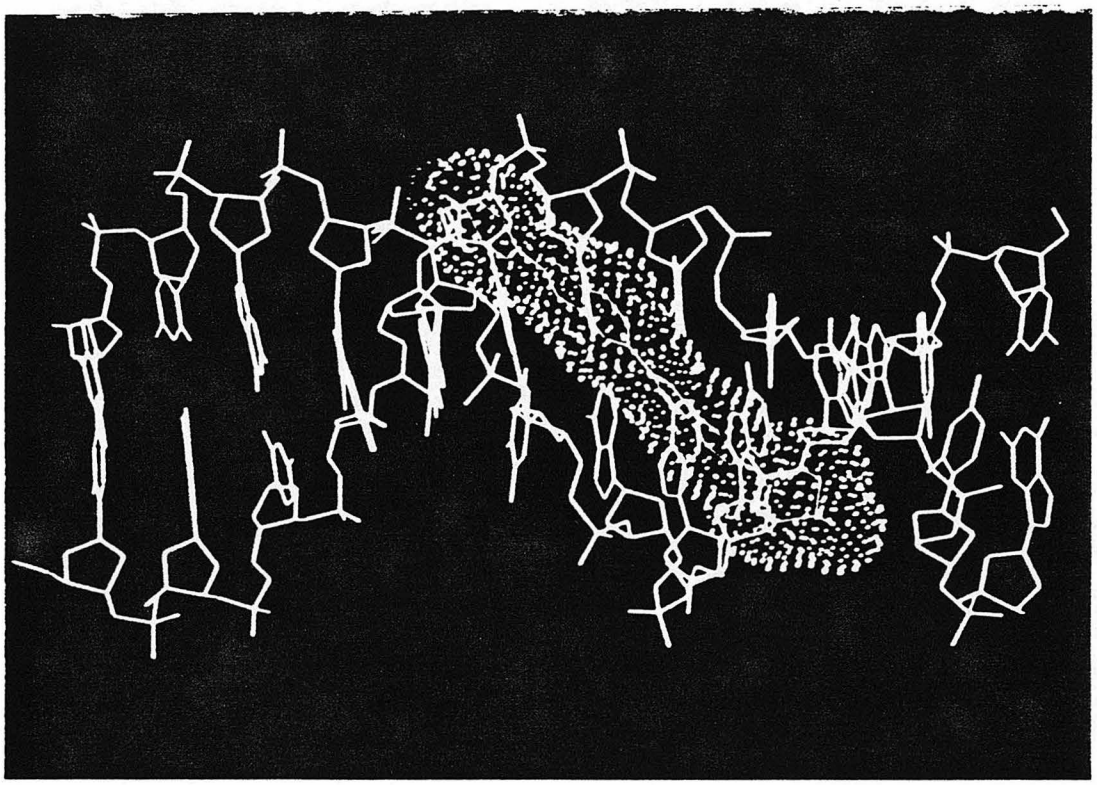


Plate IA -- DNA/Hoechst complex

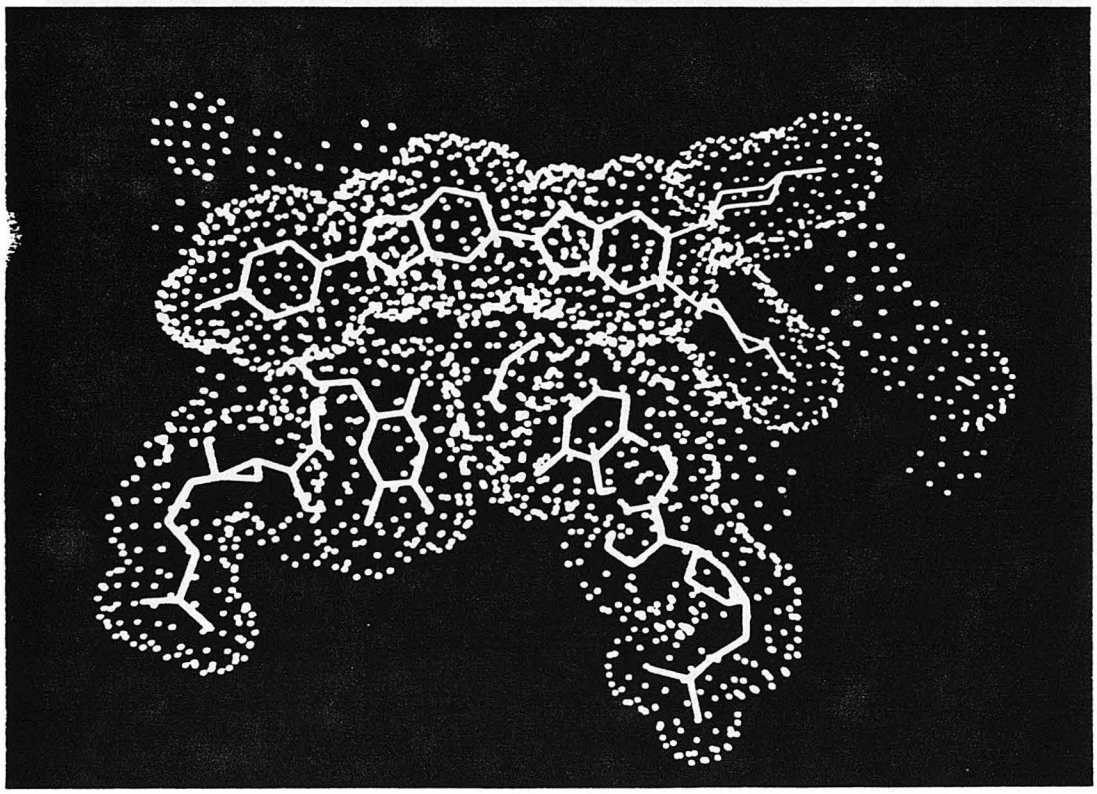


Plate IB -- DNA/Hoechst cross section with piperazine in two positions

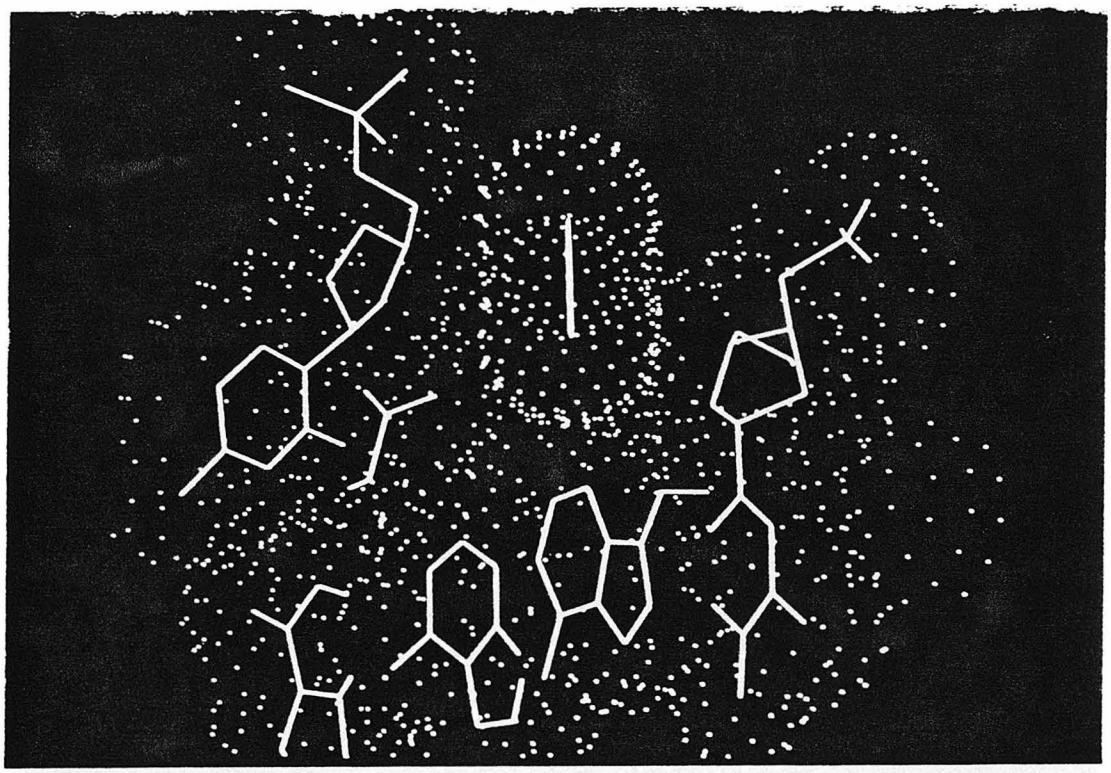


Plate IC -- Cross section through phenol

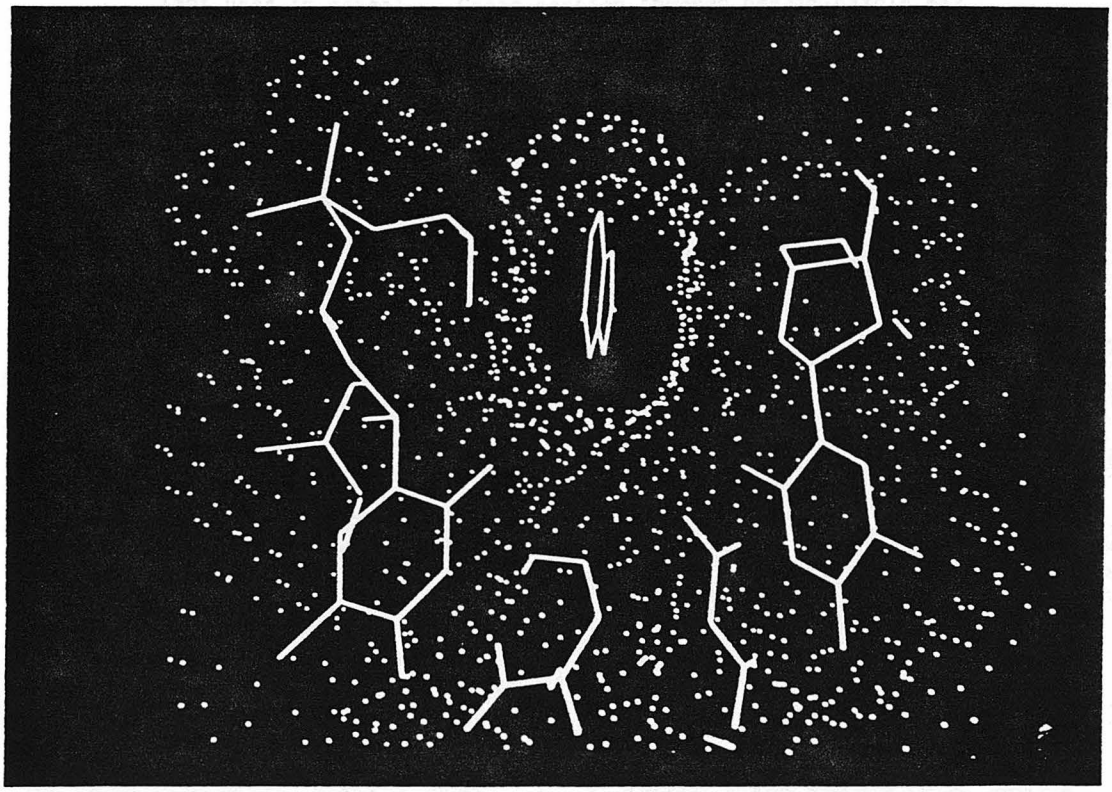
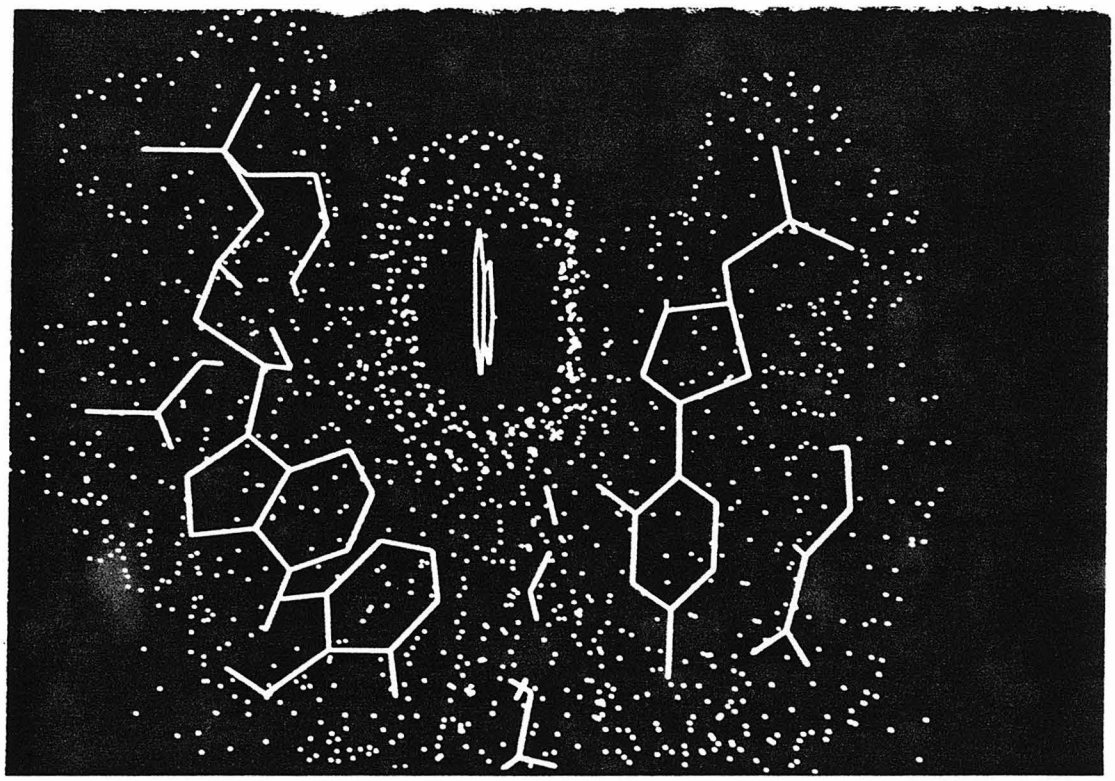


Plate ID -- Cross section through benzimidazole Bzl



(Not used in paper) -- Cross section through benzimidazole Bz2

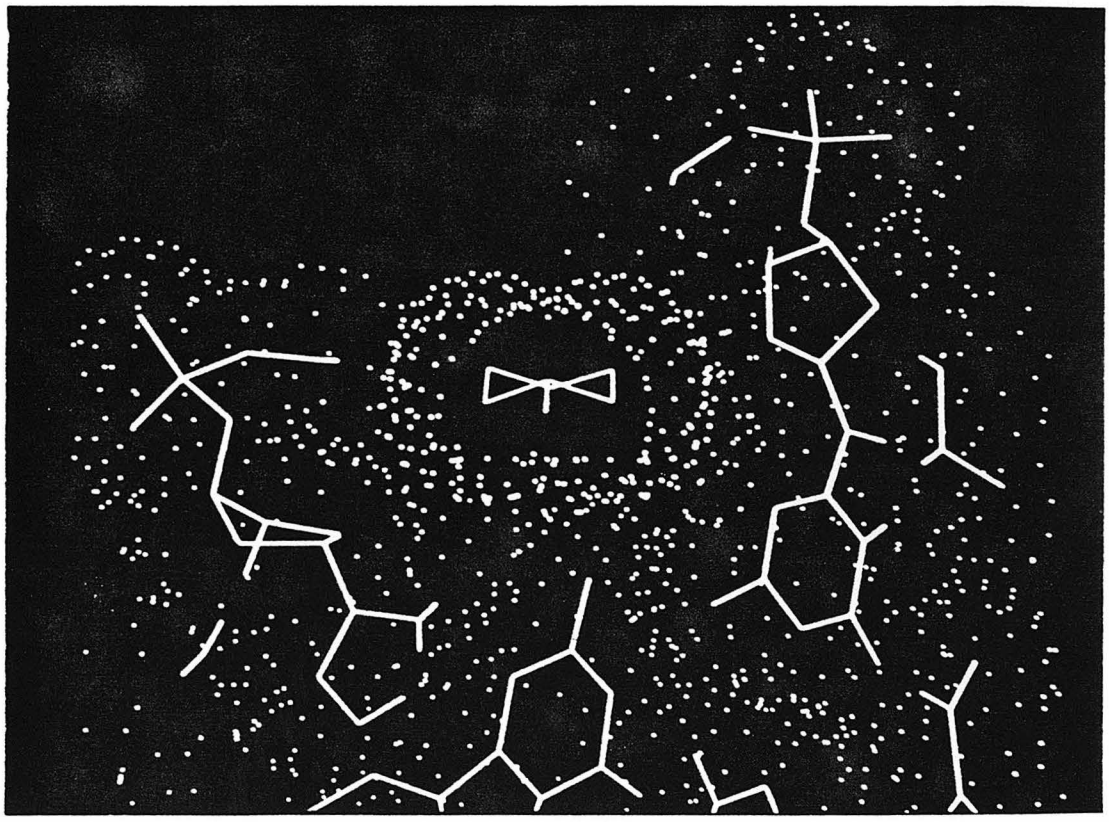


Plate IE -- Cross section through piperazine buried inside groove