Crystallographic Structure Determination of Neocarzinostatin, An Antitumor Protein - Chromophore Complex

Thesis by
Kyoung-Hee Kim

Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1995
(submitted May 25 , 1995)

To my mother and father

Acknowledgments

I would like to thank my research advisor Professor Douglas Rees for his support, encouragement, guidance and enthusiasm. He not only has been of great help in many difficulties in my research projects but also understands my crazy work hour. I would like to thank all the members of the Rees laboratory for their support and friendship. I am especially grateful to Jongsun Kim for introducing me the protein crystallography when I first came to Caltech and for teaching me the basic crystallographic techniques; Barbara Hsu for helping me data collection and processing; Dr. Leemor Joshua-Tor for showing me the cryocrystallography; Xiaotian Zhu, Mike Day, Jamie Schlessman and Michael Stowell for their assistance.

I owe thanks to Professor Andrew Myers and Dr. Byoung-Mog Kwon for providing neocarzinostatin for the structural study and for helpful discussions; Matthews Taylor and Professor Peter Dervan for providing many triplex samples for crystallizations; Noam Adir for the cocrystals of the reaction center and cytochrome c_2: Professor Pamela Bjorkman for discussions and for allowing me to use an R-AXIS imaging plate system for data collections.

Finally I would like to thank my parents, Cheon-Soo and Soon-Duk, and my sisters and brothers for their love, support and understanding. I would like to thank my friend Anna Rhee for reading my manuscripts. I would also like to thank the members of the Caltech Korean community for their friendship.

Abstract

Crystal structures of the protein-chromophore complex and the apoprotein form of the natural antitumor antibiotic neocarzinostatin (NCS) have been determined at 1.8 Å resolution. NCS is composed of a labile chromophore component with DNA-cleaving activity and a protein component that binds and stabilizes the chromophore. The NCS protein consists of the large domain of a seven-stranded β barrel and the small domain. The chromophore is bound noncovalently in a pocket between the two domains, primarily through van der Waals contacts with many nonpolar residues. Phe^{78} and Phe^{52} contact the two π-faces of the nine-membered enediyne ring through an edge of each benzene ring. Phe^{78} additionally contacts the aminosugar group through its π-face. The positioning of Phe^{78} represents the most significant difference between the holo- and apo-protein structures; in apo-NCS, Phe^{78} rotates to a more solvent-exposed position. The epoxide and C12 (the site of nucleophilic thiol addition during activation of the chromophore) are sequestered from the solvent, which likely contribute to the stability of the chromophore in holo-NCS. The amino group of the chromophore is oriented above C12 at a distance approximately the van der Waals diameter of a sulfur atom, supporting the idea that this group plays a role in the thiol activation mechanism. While the basic protein structure is conserved among at least five chromoprotein antibiotics, Phe^{78} is unique to NCS, perhaps accounting for the binding specificity of neocarzinostatin for its chromophore.

We have systematically investigated the effects of parameters in the molecular replacement methods on NCS using three different programs. The model with side chains of conserved residues and the maximum vector length or the integration radius of about half the diameter of the molecule, gave the best result. The low resolution limit of 15 Å was needed for the rotation function calculation in both X-PLOR and AMoRe. The program MERLOT failed to give molecular replacement solutions in this case.

Table of Contents

List of Tables

Appendix A.

Appendix B.

List of Figures

Chapter 1.

Chapter 2.

Chapter 3.

Appendix A.

Appendix B.

Chapter 1. Introduction to Neocarzinostatin

Enediyne drugs are a novel group of naturally occurring DNA-damaging antibiotics that have generated much interest in recent years because of their pronounced cytotoxicity and potential antitumor activity *(1,2)*. They include the calichemicins *(3,4)*, the esperamicins *(5,6)*, and the dynemicins *(7)*, the neocarzinostatin chromophore and the kedarcidin chromophore. In general, the enediyne drugs consist of three components; a system to deliver the drug to its DNA target, a DNA-damaging component (i.e., an enediyne moiety within a strained nine- or ten-membered ring) and a trigger to initiate the series of reactions resulting in DNA damage. Typically, the enediyne binds noncovalently to the minor groove and is triggered to undergo the Masamune-Bergman reaction to generate a benzenoid diradical under the appropriate conditions. In addition to the DNA damage, the cytotoxicity of certain enediyne structures has been attributed to drug-induced cell death *(8)*.

Neocarzinostatin (NCS) is isolated from the culture filtrate of *Streptomyces carzinostaticus* var. F-41. First reported in 1965 by Ishida *et al. (9)*, this antibiotic complex was later resolved into its components, a binding protein (apo-NCS) and a labile chromophore (NCS-chrom). The chromophore is extractable from NCS by organic solvents such as methanol *(10-12)*. All the biological activities of the chromoprotein reside exclusively within the labile chromophore and the protein component plays an important role in stabilizing and transporting the chromophore *(13,14)*. The biological activities of NCS include potent antitumor and antibacterial actions. In the presence of a thiol cofactor, NCS-chrom induces cleavage of single- and double-stranded DNA both *in vivo* and *in vitro*.

The novel structure of NCS-chrom **1** was proposed by Edo *et al. (15)* and has been determined through a series of spectroscopic, degradative, and synthetic studies to possess

the unique, highly strained epoxybicyclo[7.3.0]dodecadienediyne system having 2-hydroxy-7-methoxy-5-methyl-1-naphtholenecarboxylic acid, α-D-methylfucosamine and an ethylene cyclic carbonate. The proposal for the mechanism of actions of NCS-chrom (Fig. 1-1) was put forward by Myers *(16)*. The cascade of reactions leading to DNA damage is initiated by stereospecific nucleophilic attack at C12 by thiols. This triggering event is accompanied by rearrangement of the ring skeleton with epoxide opening and formation of cumulene intermediate **2**. The cumulene intermediate then undergoes rapid cycloaromatization to form diradical **3** that abstracts hydrogen atoms from deoxyribose of DNA to produce single- and double-stranded breaks. This proposal has been demonstrated in several laboratories using methyl thioglycolate *(16-18)* and sodium borohydrate *(19)* as nucleophiles. The cumulene intermediate formed from NCS-chrom and methyl thioglycolate in tetrahydrofuran : acetic acid solvent has been observed at low temperature by ^{1}H NMR spectroscopy *(18)*. The absolute configuration of the chromophore was determined from the structure of the methyl thioglycolate adducts of NCS-chrom *(17)*. Although thiol stimulates DNA damage by greater than 1000-fold *(20,21)*, NCS-chrom does cleave DNA in the absence of thiol, and this reaction is favored at more acidic pH's *(22)*, possibly acid-induced opening of the epoxide is the initiating event.

It has been demonstrated that at least 80% of DNA cleavage by NCS-chrom leads to the 5'-aldehyde of the adenine and thymine residues selectively *(23)*. The chemistry leading to these breaks involves hydrogen abstraction from C5' by the diradical species and oxygen atom transfer from molecular oxygen to the resulting DNA radical and eventually, the aldehyde residues are formed. Less than 20% of the strand breaks result from pathways initiated by hydrogen atom abstraction from C4' *(24-27)* and C1' *(24)*. The double-stranded breaks are sequence-specific and most of the breaks occur at the thymine residue of a guanine-thymine step *(27)*. They are more important than the single-stranded breaks in terms of mutagenicity and cytotoxicity. The single-stranded breaks are rapidly

repaired, whereas persistent double-stranded breaks result in cell-killing *(28-30)*. The nature of the thiol used is critical in determining the extent of formation of the double-stranded breaks; glutathione is superior to 2-mercaptoethanol or dithiothreitol in the double-stranded breaks *(31,32)*.

The structure of a complex formed by duplex DNA and the postactivated glutathione adduct of NCS-chrom have recently been determined by NMR spectroscopy *(33)*. NCS-chrom binds in the minor groove of duplex DNA by the intercalation of its naphthoate group and specific base pair recognition by its aminosugar group. The methylamino group of aminosugar moiety probably functions as an internal base in the addition reaction of thiols to C12 *(34)*. The cyclic carbonate group appears to take part in the passage of NCS-chrom through cellular and nuclear membrane, judging from the activities of the two derivatives of cyclic carbonate, the vicinal diol and the acyclic hydroxy carbonate *(10,35)*. The epoxide functionality is important in the activity of NCS-chrom, since the modification of the epoxide group to give either chlorohydrin or diol monomethyl ether did not seem to alter the binding affinities of NCS-chrom in spite of the differences in their biological activities *(36)*.

Recently, it has been found that NCS-chrom generates a single, site-specific break at a bulge in DNA in a thiol-independent reaction, involving intramolecular drug activation under general base catalysis *(37,38)*. Determination of the chemical structures of the drug products generated in the presence and the absence of bulged DNA has led to the proposal shown in Fig. 1-2 for their formation *(39)*. In this mechanism, the cumulene is stereoselectively generated via an intramolecular addition at C12 by the enolate anion **1b**, which is a resonance form of the naphthoate anion of NCS-chrom, resulting in the formation of the diradical **6**. Drug product **7** is generated only in the presence of a substrate-bulged DNA, which implies that the DNA bulge plays a vital role in drug activation. In the absence of bulged DNA, the diradical is quenched by other proton

sources to form **8**. The thiol-independent cleavage reaction is very slow compared to the thiol-dependent reaction, however it is highly efficient and selective *(39)*.

Apo-NCS binds the chromophore with high affinity ($K_d \sim 10^{-10}$ M) *(40)* and greatly stabilizes the labile chromophore. In the absence of the apoprotein, the chromophore is exceedingly unstable, undergoing rapid decomposition at high pH ($t_{1/2}$ = 0.6 min, at 0 °C and pH 8) *(40)* and upon incubation with thiols *(41)*. It has been suggested that the protein may function in binding to a cell surface receptor, or to transport the chromophore across the cell membrane into the cytoplasm, possibly for release at the level of cell nucleus *(42-45)*.

NCS-chrom shows a quite different reactivity toward thiols when complexed with the apoprotein (Fig. 1-3); nucleophilic attack of 2-mercaptoethanol at C12 of **1** complexed with the apoprotein gives a different type of cumulene intermediate that undergoes cycloaromatization followed by hydration to provide **11**. The fact that the binding constant of **11** toward the apoprotein is one order of magnitude larger than that of the chromophore suggests that the stability of the apoprotein-**11** complex may be closely related to the preferential formation of **11** in an apoprotein complex *(46,47)*.

The binding between apo-NCS and NCS-chrom is specific and no other proteins are exchangeable for apo-NCS. It had been observed that the carboxyl-terminal-43-peptide residue, obtained by trypsin digestion of apo-NCS, can bind to NCS-chrom and retains its antibacterial activity *(48)*. There is a strong hydrophobic interaction between apo-NCS and NCS-chrom as well as an ionic interaction. A release of the chromophore from NCS has been observed in a substrate-specific manner as well as in a pH-dependent manner. β-naphthol completely released the chromophore at 3 mM, while only 3% release was detected at the same concentration on α-naphthoic acid. There was also a large difference between D-galactosamine and D-galactose for the chromophore release at 10 - 500 mM, indicating the importance of the methylamino group in the chromophore for the chromophore-apoprotein interaction *(49)*. Selective oxidation of tryptophan by N-

bromosuccinimide revealed that Trp^{83} was exposed for the chemical modification but was not necessary for the chromophore binding whereas Trp^{39} was resistant to N-bromosuccinimide oxidation when the chromophore was bound *(50)*.

The amino acid sequence of apo-NCS was first investigated by conventional and gas-phase Edman degradation procedures *(51,52)* and was revised on the basis of fast atom bombardment and gas chromatographic-mass spectroscopy *(53)*. Recently, the sequence deduced from the base sequence of the gene identified inconsistent amino acid residues among different research groups *(54)*. Apo-NCS is an acidic single-chain polypeptide (pI = 3.2) with 113 amino acid residues. It possesses two disulfide bonds and a large portion of alanine, glycine, valine, serine and threonine (69 out of 113 residues). The conformation of the peptide backbone of apo-NCS was studied by infrared spectroscopy and circular dichroism measurement *(55)*. The preliminary x-ray study on NCS was reported by Sieker *et al. (56)*. These studies have suggested that the secondary structure of the apoprotein is rich in β structures. NMR solution structures *(57-59)* and a crystal structure *(60)* of apo-NCS at 1.5 Å resolution have been reported. A NMR solution structure of holo-NCS has also been reported *(61)*. The modeling study on holo-NCS has been described *(62)*.

A potent cytotoxic agent, NCS has undergone clinical evaluation in the treatment of various cancers, including cancers of the bladder, the stomach and leukemia *(63-66)*. NCS can inhibit tumor cell growth at the nanomolar range. The major inhibition for wide clinical use is its severe toxicity and the extremely short half-life both *in vivo* (in blood) and *in vitro*. SMANCS, a conjugate of poly(styrene-co-maleic acid/anhydride) with NCS, was developed to improve the stability of the NCS complex and to reduce its toxicity *(67,68)*. SMANCS was approved by the Japanese FDA and has been marketed in Japan.

Other chromoprotein antibiotics have been isolated from diverse microbial sources (Table 1-1). They include actinoxanthin (ACX) *(69)*, auromomycin which is also known as macromomycin (MCM) *(70)*, C-1027 *(71)* and kedarcidin *(72)*. They exhibit the

antibacterial activity as a result of DNA cleavage and each of them consists of an apoprotein and an active chromophore. The protein component of sixth antitumor chromoprotein, maduropeptin (M.W. ~ 31 kD) is not related to the NCS class proteins, though its chromophore has the nine-membered enediyne ring system *(73,74)*. Apoproteins in this family exhibit high sequence identities, higher than 36% for each pair (Table 1-2). ACX and C-1027 apoproteins are almost identical. The sequence alignments of five apoproteins are shown in Fig. 1-5. The first four apoproteins show higher sequence identities. Crystal structures of ACX *(75)* and MCM *(76)* were reported before we started this project. A preliminary x-ray study on the C-1027 apoprotein *(77)* and the NMR structure of the kedarcidin apoprotein *(78)* have also been published. Each apoprotein has a similar architecture that consists of two domains, separated by the cleft for the chromophore binding. Holo C-1027 and the kedarcidin apoprotein, as well as apo-NCS, are reported to possess specific proteolytic activity *(79-80)*.

The C-1027 chromophore *(81)* and the kedarcidin chromophore *(82)* have common nine-membered enediyne ring systems but different functional groups, which are bigger than those of NCS-chrom (Fig. 1-4). Similar ways of activation to that of NCS-chrom are proposed for both chromophores. Whereas NCS-chrom and 10-membered enediynes have been reported to have some trigger systems in their structures to start the reaction cascade leading to the diradical generation, there seems to be no trigger in the C-1027 chromophore which undergoes the spontaneous cycloaromatization without reducing agents such as thiols. Each chromophore is distinct from NCS-chrom, suggesting that each apoprotein maintains specificity for its respective chromophore.

The three-dimensional structure of NCS complex is required to observe the structure of NCS-chrom in the protein and to understand how the apoprotein recognizes and stabilizes the chromophore. We have determined the structures of apo-NCS and holo-NCS at 1.8 Å resolution. Chapter 2 describes the details of the structure determination including crystallization, data collection, heavy atom derivative screening, multiple

isomorphous replacement phase calculation, noncrystallographic symmetry determination, averaging, model building and refinement. Structures of the chromophore and the apoprotein and the comparisons with other chromoprotein structures are described in Chapter 3. In Chapter 4, molecular replacements on NCS with three different programs and the practical aspect of the molecular replacement are discussed.

Table 1-1. Comparison among antibiotic chromoproteins

name	source	amino acid residue	structure chromophore	protein
neocarzinostatin	*Streptomyces carzinostaticus*	113	yes	NMR, X-ray
actinoxanthin	*Actinomycete globisporus*	108	no	X-ray
macromomycin	*Streptomyces macromomycetius*	112	no	X-ray
C-1027	*Streptomyces globisporus*	110	yes	none
kedarcidin	*Actinomycete* strain L-586-6	114	yes	NMR

Table 1-2. Sequence identities and similarities among apoproteins

	MCM	ACX	C-1027	kedarcidin	
NCS	62.2	59.3	59.1	57.5	similarity
	46.8	48.1	48.2	39.6	identity
MCM		56.5	56.4	54.9	
		38.0	38.2	36.0	
ACX			94.4	55.1	
			93.5	41.1	
C-1027				57.8	
				44.0	

Figure 1-1. Proposed mechanism of thiol-dependent reaction of NCS-chrom

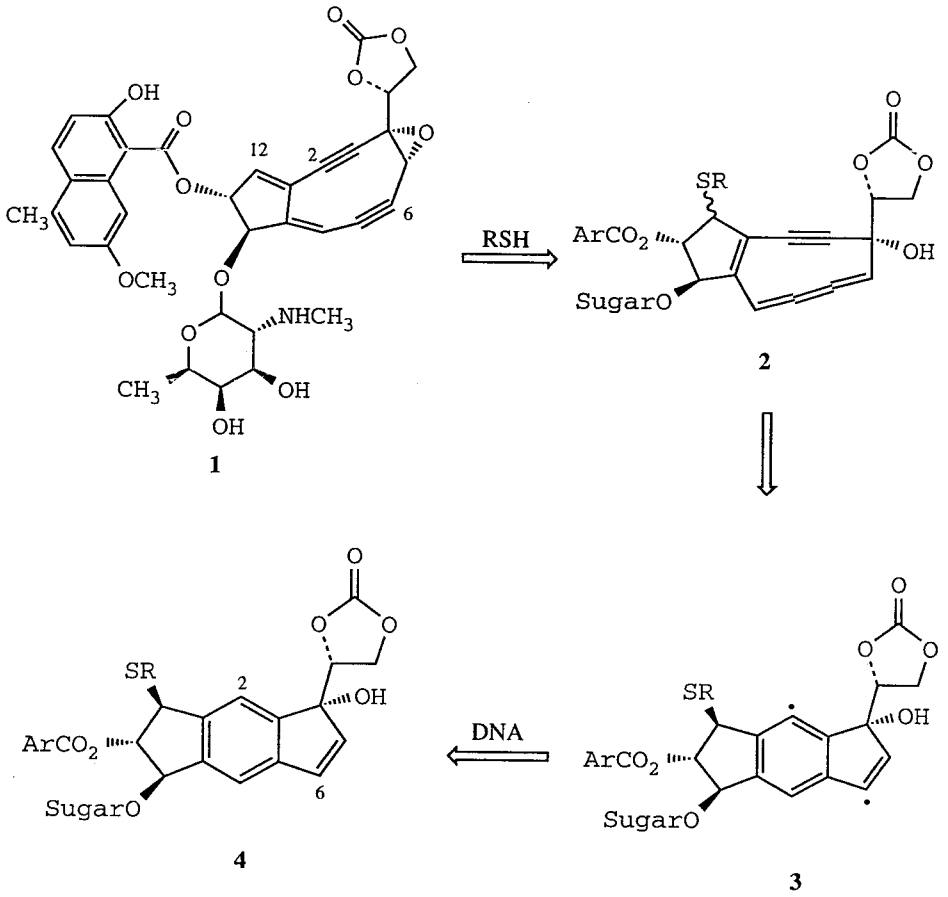


Figure 1-2. Proposed mechanism of thiol-independent reaction of NCS-chrom with bulged DNA

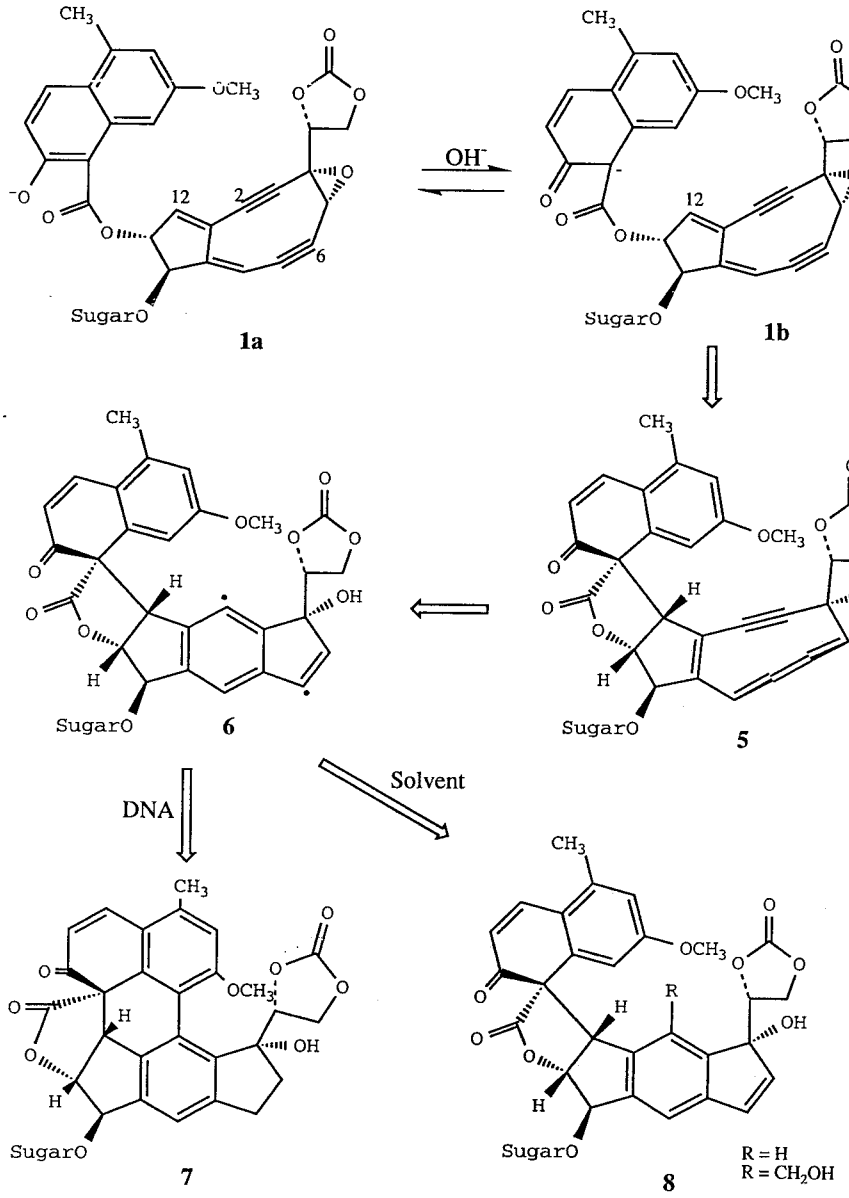


Figure 1-3. Proposed mechanism of apoprotein-assisted reaction of NCS-chrom

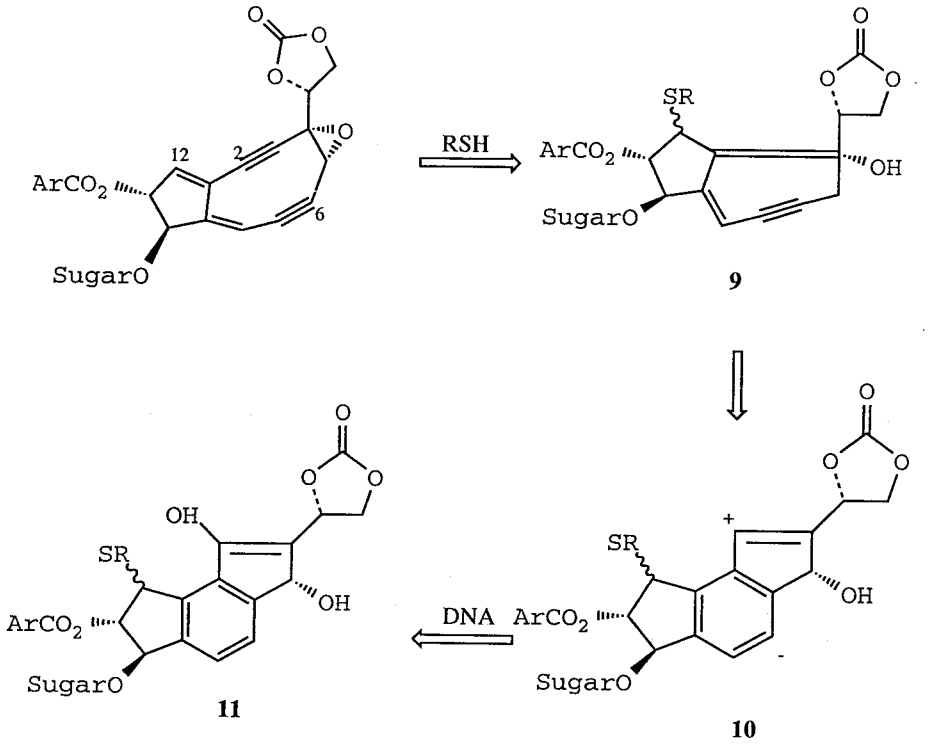
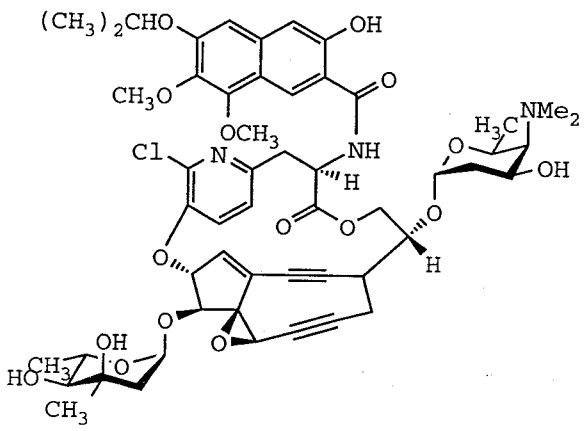


Figure 1-4. The structures of the chromophores.

(a) the kedarcidin chromophore



(b) the C-1027 chromophore

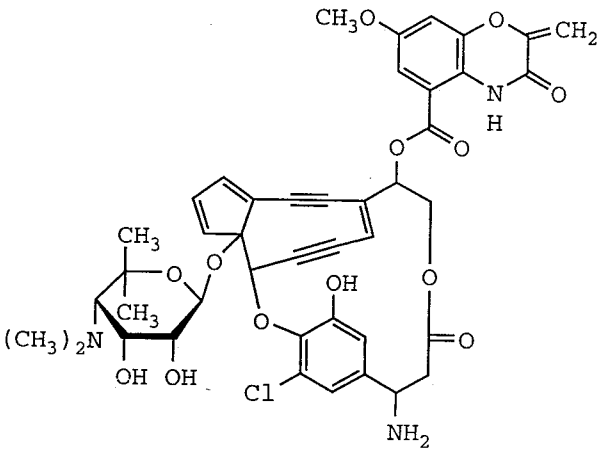


Figure 1-5. Amino acid sequence alignment of NCS *(54)*, MCM *(83)*, ACX *(53)*, C-1027 *(84)* and kedarcidin (KED) *(85)*. Residues that are absolutely conserved in all five proteins are shown in bold type.

```
                                .         .         .         .
NCS       1 AAPTATVTPSSGLSDGTVVKVAGAGLQAGTAYDVGQCAWV 40
MCM       1 .APGVTVTPATGLSNGQTVTVSATGLTPGTVYHVGQCAVV 39
ACX       1 .APAFSVSPASGLSDGQSVSVSV..AAAGETYYIAQCAPV 37
C1027     1 .APAFSVSPASGLSDGQSVSVSVSGAAAGETYYIAQCAPV 39
KED       1 ASAAVSVSPATGLADGATVTVSASGFATSTSATALQCAIL 40

                                .         .         .         .
NCS      41 DTGVLACNPADFSSVTADANGSASTSLTVRRSFEGFL.FDG 80
MCM      40 EPGVIGCDATTSTDVTADAAGKITAQLKVHSSFQAVVGADG 80
ACX      38 G.GQDACNPATATSFTTDASGAASFSFTVRKSYAGQT.PSG 76
C1027    40 G.GQDACNPATATSFTTDASGAASFSFVVRKSYTGST.PEG 78
KED      41 ADGRGACNVAEFHDFSLSG.GEGTTSVVVRRSFTGYVMPDG 80

                                .         .         .
NCS      81 TRWGTVDCTTA..ACQVGLSDAAGNGPEGVAISFN 113
MCM      81 TPWGTVNCKVV..SCSAGLGSDSGEGA.AQAITFA 112
ACX      77 TPVGSVDCATD..ACNLGAGNSGLNLGH.VALTFG 108
C1027    79 TPVGSVDCATA..ACNLGAGNSGLDLGH.VALTFG 110
KED      81 PEVGAVDCDTAPGGCQIVVGGNTGEYGNA.AISFG 114
```

References

(1) K. C. Nicolaou and W.-M. Dai, *Angew. Chem. Int. Ed. Engl.*, **30**, 1387 (1991)

(2) K. C. Nicolaou, A. L. Smith, E. W. Yue, *Proc. Natl. Acad. Sci. USA*, **90**, 5881 (1993)

(3) M. D. Lee, T. S. Dunne, M. M. Siegel, C. C. Chang, G. O. Morton, D. B. Borders, *J. Am. Chem. Soc.*, **109**, 3464 (1987)

(4) M. D. Lee, T. S. Dunne, C. C. Chang, G. A. Ellestad, M. M. Siegel, G. O. Morton, W. J. McGahren, D. B. Borders, *J. Am. Chem. Soc.*, **109**, 3466 (1987).

(5) J. Golik, J. Clardy, G. Dubay, G. Groenewold, H. Kawaguchi, M. Konishi, B. Krishnan, H. Ohkuma, K. Saito, T. W. Doyle, *J. Am. Chem. Soc.*, **109**, 3462 (1987)

(6) J. Golik, G. Dubay, G. Groenwold, H. Kawaguchi, M. Konishi, B. Krishnan, H. Ohkuma, K. Saito, T. W. Doyle, *J. Am. Chem. Soc.*, **109**, 3462 (1987).

(7) M. Konishi, H. Ohkuma, T. Tsuno, T. Oki, G. D. VanDuyne, J. Clardy, *J. Am. Chem. Soc.*, **112**, 3715 (1990).

(8) K. C. Nicolaou, P. Stabila, B. Esmaeli-Azad, W. Wrasidlo, A. Hiatt, *Proc. Natl. Acad. Sci. USA*, **90**, 3142 (1993)

(9) N. Ishida, K. Miyazaki, K. Kumagi, M. Rikimaru, *J. Antibiot.*, **A18**, 68 (1965)

(10) M. A. Napier, B. Holmquist, D. J. Strydom, I. H. Goldberg, *Biochem. Biophys. Res. Commun.*, **89**, 635 (1979)

(11) W. Koide, F. Ishii, K. Hasuda, Y. Koyama, K. Edo, S. Katamine, F. Kitame, N. Ishida, *J. Antibiot.*, **33**, 342 (1980)

(12) H. Suzuki, K. Miura, K. Kumada, T. Takeuchi, N. Tanaka, *Biochem. Biophys. Res. Commun.*, **94**, 255 (1980)

(13) M. A. Napier, L. S. Kappen, I. H. Goldberg, *Biochemistry*, **19**, 1767 (1980)

(14) L. S. Kappen, M. A. Napier, I. H. Goldberg, *Proc. Natl. Acad. Sci. USA*, **77**, 1970 (1980)

(15) K. Edo, M. Mizugaki, Y. Yoshida, H. Seto, K. Furihata, N. Otake, N. Ishida, *Tetrahedron Lett.*, **26**, 331 (1985)

(16) A. G. Myers, *Terahedron Lett.*, **28**, 4493 (1987)

(17) A. G. Myers, P. J. Proteau, T. M. Handel, *J. Am. Chem. Soc.*, **110**, 7212 (1988)

(18) A. G. Myers and P. J. Proteau, *J. Am. Chem. Soc.*, **111**, 1146 (1989)

(19) O. D. Hensens and I. H. Goldberg, *J. Antibiot.*, **42**, 761 (1989)

(20) T. A. Beerman, R. Poon, I. H. Goldberg, *Biochim. Biophys. Acta*, **475**, 294 (1977)

(21) L. S. Kappen and I. H. Goldberg, *Nucleic Acids Res.*, **5**, 2959 (1978)

(22) D.-H. Chin, C.-H. Zeng, C. E. Costello, I. H. Goldberg, *Biochemistry*, **27**, 8106 (1988)

(23) L. S. Kappen and I. H. Goldberg, *Biochemistry,* **22**, 4872 (1983)

(24) L. S. Kappen, I. H. Goldberg, S. H. Wu, J. Stubbe, L. Worth, Jr., J. W. Kozarich, *J. Am. Chem. Soc.*, **112**, 2797 (1990)

(25) L. S. Kappen, I. H. Goldberg, B. L. Frank, L. Worth, Jr., D. F. Christner, J. W. Kozarich, J. Stubbe, *Biochemistry*, **30**, 2034 (1991)

(26) I. Saito, H. Kawabata, T. Fujiwara, H. Sugiyama, T. Matsura, *J. Am. Chem. Soc.*, **111**, 8302 (1989)

(27) B. L. Frank, L. Worth, Jr., D. F. Christner, J. W. Kozarich, J. Stubbe, L. S. Kappen, I. H. Goldberg, *J. Am. Chem. Soc.*, **113**, 2271 (1991)

(28) T. Hatayama and I. H. Goldberg, *Biochim. Biophys. Acta*, **563**, 59 (1979)

(29) L. F. Povirk and I. H. Goldberg, *Proc. Natl. Acad. Sci. USA*, **82**, 3182 (1985)

(30) L. S. Kappen, C.-Q. Chen, I. H. Goldberg, *Biochemistry*, **27**, 4331 (1988)

(31) P. C. Dedon and I. H. Goldberg, *Biochemistry*, **31**, 1909 (1992)

(32) P. C. Dedon, Z.-W. Jiang, I. H. Goldberg, *Biochemistry*, **31**, 1917 (1992)

(33) X. Gao, A. Stassinopoulos, J. S. Rice, I. H. Goldberg, *Biochemistry*, **34**, 40 (1995)

(34) A. G. Myers, P. M. Harrington, B.-M. Kwon, *J. Am. Chem. Soc.*, **114**, 1086 (1992)

(35) M. A. Napier, I. H. Goldberg, O. D. Hensens, R. S. Dewey, J. M. Liesch, G. Albers-Schronberg, *Biochem. Biophys. Res. Commun.*, **100**, 1703 (1981)

(36) S. H. Lee and I. H. Goldberg, *Mol. Pharmacol.*, **33**, 396 (1988)

(37) L. S. Kappen and I. H. Goldberg, *Science*, **261**, 1319 (1993)

(38) L. S. Kappen and I. H. Goldberg, *Biochemistry*, **32**, 13138 (1993)

(39) O. D. Hensens, D.-H. Chin, A. Stassinopoulos, D. L. Zink, L. S. Kappen, I. H. Goldberg, *Proc. Natl. Acad. Sci. USA*, **91**, 4534 (1994)

(40) L. F. Povirk and I. H. Goldberg, *Biochemistry*, **19**, 4773 (1980)

(41) L. S. Kappen and I. H. Goldberg, *Biochemistry*, **19**, 4786 (1980)

(42) H. Maeda, S. Aikawa, A. Yamashita, *Cancer Res.*, **35**, 554 (1975)

(43) H. Lazarus, V. Razo, T. S. Anantha Samy, *Cancer Res.*, **37**, 3731 (1977)

(44) J. Takeshita, H. Maeda, K. Koike, *J. Biochem.*, **88**, 1071 (1980)

(45) S. Sakamoto, H. Maeda, J. Ogata, *Experiments* , **35**, 1233 (1979)

(46) H. Sugiyama, K. Yamashita, T. Fugiwara, I. Saito, *Tetrahedron*, **50**, 1311 (1994)

(47) D.-H. Chin and I. H. Goldberg, *J. Am. Chem. Soc.*, **115**, 9341 (1993)

(48) K. Edo, K. Saito, Y. Akiyama-Murai, M. Mizugaki, Y. Koide, N. Ishida, *Chem. Pharm. Bull.*, **34**, 5180 (1988)

(49) K. Edo, K. Saito, Y. Akiyama-Murai, M. Mizugaki, Y. Koide, N. Ishida, *J. Antibiotics*, **41**, 554 (1988)

(50) K. Edo, K. Saito, Y. Matsuda, Y. Akiyama-Murai, M. Mizugaki, Y. Koide, N. Ishida, *Chem. Pharm. Bull.*, **39**, 170 (1991)

(51) H. Maeda, C. B. Glaster, K. Kuromizu, J. Meienhofer, *Arch. Bioc. Biop.*, **164**, 379 (1974)

(52) K. Kuromizu, S. Tsunasawa, H. Maeda, O. Abe, F. Sakiyama, *Arch. Bioc. Biop.*, **246**, 199 (1986)

(53) B. W. Gibson, W. C. Herlihg, T. S. Anantha Samy, K.-S. Hahm, H. Maeda, J. Meienhofer, K. Biemann, *J. Biol. Chem.*, **259**, 10801 (1984)

(54) N. Sakata, S. Minamitani, T. Kanabe, M. Hori, M. Hamada, K. Edo, *Biol. Phar. Bull.*, **16**, 26 (1993)

(55) K. Saito, Y. Sato, K. Edo, Y. Akiyama-Murai, Y. Koide, N. Ishida, M. Mizugaki, *Chem. Pharm. Bull.*, **37**, 3078 (1989)

(56) L. C. Sieker, L. H. Jensen and T. S. Anantha Samy, *Biochem. Biophys. Res. Comm.*, **68**, 358 (1976)

(57) X. Gao, *J. Mol. Biol.*, **225**, 125 (1992)

(58) E. Adjadj, E. Quiniou, J. Mispelter, V. Favaudon, J.-M. Lhoste, *Eur. J. Biochem.*, **203**, 505 (1992)

(59) M. Lyndsay Remerowski, S. J. Glaser, L. C. Sieker, T. S. Anantha Samy, G. P. Drobny, *Biochemistry* , **29**, 8401 (1990)

(60) A. Teplyakov, G. Obmolova, K. Wilson, K. Kuromizu, *Eur. J. Biochem.*, **213**, 737 (1993).

(61) T. Tanaka, M. Hirama, K.-i. Fujita, S. Imajo, M. Ishiguro, *J. Chem. Soc. Chem. Commun.*, **15**, 1205 (1993)

(62) M. Ishiguro, S. Imajo, M. Hirama, *J. Med. Chem.*, **34**, 2366 (1991)

(63) H. Maeda, *Anticancer Res.*, **1**, 175 (1981)

(64) D. D. Von Hoff, D. A. Amato, J. H. Kaufman, G. Falkson, T. J. Cunningham, *Am. J. Clin. Oncol.*, **7**, 135 (1984)

(65) G. Falkson, J. M. MacIntyre, J. A. Schutt, B. Coetzer, L. A. Johnson, I. W. Simson, H. O. Douglas, *J. Clin. Oncol.*, **2**, 581 (1984)

(66) T. Konno, *Eur. J. Canc.*, **28**, 403 (1992) and references therein.

(67) A. Kimoto, T. Konno, T. Kawaguchi, Y. Miyauchi, H. Maeda, *Cancer Res.*, **52**, 1013 (1992)

(68) H. Maeda, *Advanced Drug Delivery Reviews*, **6**, 181 (1991)

(69) A. S. Khokholov, B. Z. Cherches, P. D. Reshetov, G. M. Smirnova, I. B. Sorokina, T. A Prokoptzeva, T. A. Koloditskaya, V. V. Smirnov, S. M. Navashin, I. P. Formina, *J. Antibiot.*, **22**, 541 (1969)

(70) H. Chimura, M. Ishizuka, M. Hamada,, S. Hori, K. Kimura, J. Iwanga, T. Takeuchi, H. Umezawa, *J. Antibiot.*, **21**, 44 (1968)

(71) J. Hu, Y.-C. Xue, M.-Y. Xie, R. Zhang, T. Otani, Y. Minami, Y. Yamada, T. Marunaka, *J. Antibiot.*, **41**, 1575 (1988)

(72) K. S. Lam, G. A. Hesler, D. R. Gustavson, A. R. Crosswell, J. M. Veitch, A. Forenza, K. Tomita, *J. Antibiot.*, **44**, 472 (1992).

(73) M. Hanada, H. Ohkuma, T. Yonemoto, K. Tomita, N. Ohbayashi, H. Kaemi, T. Miyaki, M. Konishi, H. Kawaguchi, *J. Antibiotics*, **44**, 403 (1991)

(74) D. R. Schroeder, K. L. Colson, S. E. Khlor, N. Zein, D. R. Langley, M. S. Lee, J. A. Matson, T. W. Doyle, *J. Am. Chem. Soc.*, **116**, 9351 (1994)

(75) V. Z. Pletnev, A. P. Kuzin, S. D. Trakhanov, P. V. Kostetsky, *Biopolymers,* **21**, 287 (1982)

(76) P. Van Roey and T. A. Beerman, *Proc. Natl. Acad. Sci. USA*, **86**, 6587 (1989)

(77) P. Briozzo, K. Inaka, Y. Minami, T. Otani, K. Miki, *Acta Cryst.*, **D49**, 372 (1993)

(78) K. L. Constantine, K. L. Colson, M. Wittekind, M. S. Friedrichs, N. Zein, J. Tuttle, D. R. Langley, J. E. Leet, D. R. Schroeder, K. S. Lam, B. T. Farmer II, W. J. Metzler, R. E. Bruccoleri, L. Mueller, *Biochemistry*, **33**, 11438 (1994)

(79) N. Sakata, K. S. Tsuchiya, Y. Moriya, H. Hayashi, M. Mori, T. Otani, M Nagai, T. Aoyagi, *J. Antibiotics*, **45**, 113 (1992)

(80) N. Zein, A. M. Casazza, T. W. Doyle, J. E. Leet, D. R. Schroeder, W. Solomon, S. G. Nadler, *Proc. Natl. Acad. Sci. USA*, **90**, 8009 (1993)

(81) K.-i. Yoshida, Y. Minami, R. Azuma, M. Saeki, T. Otani, *Tetrahedron Letters*, **34**, 2637 (1993)

(82) J. L. Leet, D. R. Schroeder, D. R. Langley, K. L. Colson, S. Haung, S. E. Khlor, M. S. Lee, J. Golik, S. J. Hofstead, T. W. Doyle, J. A. Matson, *J. Am. Chem. Soc.*, **115**, 8432 (1993)

(83) T. S. Anantha Samy, K.-S. Hahm, E. H. J. Modesr, G. W. Lampman, H. T. Keutmann, H. Umezawa, W. C. Herlihy, B. W. Gibson, S. A. Carr, K. Biemann, *J. Biol. Chem.*, **258**, 183 (1983)

(84) N. Sakata, S. Keno, M. Hori, M. Hamada, T. Otani, *Biosci. Biotech. Biochem.*, **56**, 1592 (1992)

(85) S. J. Hofstead, J. A. Matson, A. R. Malacko, H. Marquardt, *J. Antibiotics,* **45**, 1250 (1992)

Chapter 2. Structure Determination of Neocarzinostatin

2-1. Crystallization

Neocarzinostatin (NCS) powder was a generous gift of Kayaku Co., Ltd. to Dr. A. G. Myers. A 20 mg/ml NCS solution was prepared in 50 mM sodium acetate pH 5.0. All the crystallization experiments were done with the conventional hanging drop method and in the cold room, since no release of the chromophore from the protein was measured at 4 °C in 0.1 M sodium acetate pH 5.0 *(1)*. Initial screenings for the crystallization conditions were performed using the modified Jaru factorial solutions. No crystals were obtained from these conditions and most of the conditions failed to yield precipitates, which indicates the high solubility of NCS, due to the protein's acidic nature. Ammonium sulfate solutions at various pH's were tried, since NCS had been previously crystallized from it *(2)*. Precipitates, instead of crystals, were obtained at high concentrations of ammonium sulfate (>2.0 M) regardless of pH.

The NCS crystals were grown from 70% 2-methyl-2,4-pentanediol (MPD) and 50 mM sodium acetate pH 4.6. The hanging drops contained equal volumes of the protein solution and the reservoir solution. The rod-shaped crystals (Fig. 2-1) were obtained after 2 weeks and had dimensions of 0.3 x 0.1 x 0.1 mm^3.

2-2. Heavy atom derivative screening

Heavy atom derivatives are used to obtain the phase information by the isomorphous replacement method. For NCS, heavy atom derivatives were prepared by soaking the crystal in mother liquor containing the heavy atom compounds. Heavy atom solutions were added directly into the hanging drop with a pipetteman. Typically, 2 µl of heavy atom solution was added to a 4 µl crystallization drop. Table 2-1 lists names of compounds, soaking concentrations, time periods and *R* factors with the native data set I.

Twenty-four compounds were tried, only four of which were good enough to use in the phase calculations. They are trimethyllead acetate (Pb1), trimethyllead chloride (Pb2), samarium acetate (Sm) and iridous potassium chloride (Ir).

2-3. Data collection and processing

Because the crystals were very sensitive to x-ray radiation, all data sets were collected at around -160 °C *(3)*. In the cold room, the crystal was picked up with a small glass spatula, which consisting of a tiny glass plate glued to the end of a thin glass fiber, which is soldered into a copper pin (Fig. 2-1 (b)). Excess liquid was removed with a small piece of filter paper. The crystal was then transferred to the x-ray room in an ice box to minimize the damage due to the temperature change and put under the liquid nitrogen stream as quickly as possible. Since NCS crystals were grown from high percentages of MPD, a cryosolvent wasn't required for the freezing. At room temperature or 4 °C, crystals lasted about three hours at best. In contrast, one crystal lasted about two weeks without any sign of the radiation damage under the liquid nitrogen stream.

The native data set I and derivative data sets were collected with a Siemens multiwire area detector mounted on a Siemens rotating anode generator operating at 50 kV x 90 mA and were processed with the program package XENGEN *(6)* and ROCKS *(7)*. The native data set II was collected on an R-AXIS imaging plate system mounted on a Rigaku rotating anode generator and processed with the R-AXIS software. CuKα radiation (λ= 1.54 Å) was used for the diffraction studies. Table 2-2 lists data statistics for native and derivative data sets.

The native data set I was 92.7% complete to 2.4 Å resolution with an *R*merge of 7.11% and was used for calculation of MIR phases, averaging, solvent flattening, and initial stages of model building and refinement. The native data set II was 94.6% complete to 1.8 Å resolution and was used in later stages of model building and refinement. Data between 1.8 Å to 1.7 Å resolution shell was not used in refinement due to the low

completeness and <I/σ> value (51.3% and 1.97, respectively). The *R* factor between two native data sets was 14.25% for the resolution range of 50 to 2.4 Å. Data sets of heavy atom derivatives were more than 90% complete up to 2.7 Å resolution. *R* factors with the native data set I were about 12%. Ir derivative gave the best data set with an <I>/<σ> of 26.2 and an *R*merge of 4.7%.

NCS crystals are belong to the space group $P6_5$ with cell dimensions of $a= b= 90.4$ Å and $c = 52.5$ Å. This crystal form is distinct from the crystals reported by Sieker *et al. (2)* and Teplykov *et al. (4)*, which belong to the space group $P2_12_12_1$ with the cell constants of $a= 27$ Å, $b= 34$ Å and $c= 102$ Å. Two molecules per asymmetric unit and 54% of solvent content are expected based on the Matthews coefficient *(5)*.

2-4. Self-rotation function

Since two molecules of NCS are expected in the asymmetric unit, there should be an axis that relates them. In order to find the noncrystallographic symmetry axis, the self-rotation functions were calculated using the X-PLOR package *(8)* and the Crowther's fast rotation function *(9)*. The Crowther's self-rotation function showed the features of the sixfold axis along the *c* axis (Fig. 2-2). X-PLOR gave no significant peaks except the origin peak. The noncrystallographic axis might be on the same direction with the crystallographic axis or the programs were unable to detect the noncrystallographic symmetry with given data.

2-5. Molecular replacement

The molecular replacement method has been widely used to solve three-dimensional structures of molecules when the structures of similar or homologous molecules are available. The coordinates of actinoxanthin (ACX) *(10)* and macromomycin (MCM) *(11)* are available from the Brookhaven Data Bank as 1acx.pdb and 2mcm.pdb, respectively. ACX and MCM have two structural domains, which form a cleft for the chromophore

binding. The average distance between corresponding Cα atoms between ACX and MCM is 1.62 Å.

The rotation function was searched using the X-PLOR package and the Crowther's fast rotation function. The translation function was searched on the basis of the correlation coefficient between the observed and model structure factors. Possible solutions were checked by difference Fourier syntheses using the observed structure factors of heavy atom derivatives and the model phases. Several models were tried such as the whole molecule, the molecule with conserved side chains, the molecule with main chain atoms only and the molecule without the small domain. Various resolution ranges and parameters were tried with each program and model, no promising peaks were found. After the complete structure determination, the molecular replacement method was tried again and the results are described in Chapter 4.

2-6. MIR phasing and solvent flattening

In order to determine the heavy atom sites, isomorphous difference Patterson maps were calculated (Fig. 2-4). The binding site of Pb1 data set was determined from difference Patterson maps and the other sites were determined from difference Fourier syntheses (Table 2-3). Each derivative has one binding site, so that it was not possible to determine the noncrystallographic symmetry from the heavy atom sites. The z coordinates of all heavy atom sites were very close to z= 0 values.

Anomalous scattering data of Sm and Ir derivatives were used to distinguish the enantiomorphic space group. Sm derivative gave negative anomalous occupancy for the space group $P6_1$ and the space group of NCS crystals was determined to be $P6_5$.

The PHARE program *(12)* was used to refine positions and occupancies of the heavy atom sites and to calculate the multiple isomorphous replacement (MIR) phases. The phasing statistics are presented in Table 2-4. The mean figure of merit was 0.58 for the resolution range of 10 to 3 Å. The phases at high resolution were determined mainly

by Pb derivatives, since Ir derivative lost its contribution to the phase determination after 4.0 Å resolution and phasing powers of Sm derivative decreased dramatically after 5.0 Å resolution. Electron density maps with the MIR phases were too noisy to build the model (Fig. 2-5).

To improve the quality of MIR maps, the solvent flattening was performed with the program PHASES. The solvent content was set to 54% or less and the protein envelope was generated automatically by Wang's algorithm in the reciprocal space *(14)*. The MIR phases were iteratively refined by negative density truncation and solvent flattening. The *R* factor decreased to about 27% and the mean figure of merit increased to 0.67, but the map wasn't improved much. NCS molecules in the solvent-flattened map stacked on top of each other along the *c* axis and their boundary was not obvious.

2-7. Noncrystallographic symmetry determination and averaging

Since there are two molecules in the asymmetric unit, the noncrystallographic symmetry in the crystal was determined in order to improve the phase by averaging. The native Patterson map calculated at low resolution (~5 Å) revealed a large peak at u= 0, v= 0, w= 1/2 (Fig. 2-3). Its peak height was one-fourth of that of the origin peak, which indicated that two molecules were roughly related by a translation of 1/2 along the *c* axis. This relationship is not an exact one since the peak height decreased dramatically at the high resolution limits. For two equivalent positions (x, y, z) and (-x, -y, z+1/2) in the space group $P6_5$, the positions related by the translational symmetry are (x, y, z+1/2) and (-x, -y, z+1). The noncrystallographic twofold axis was determined from positions (x, y, z) and (-x, -y, z+1) to be parallel to the crystallographic *c* axis. Assuming the twofold axis goes through the heavy atom binding site of Sm derivative, it was refined with the program RHOPRP *(15)* that refines a special twofold axis based on the correlation of electron density maps. The twofold axis was about 2.3° tilted from the *c* axis and related two molecules with a rotation of 180° and a translation of 90.4 Å along the *a* axis. The

averaged map based on this twofold axis wasn't adequate for the model building, partly because of the poor MIR phases and partly because of the inaccurate twofold axis.

To accurately determine the noncrystallographic symmetry, the MCM structure was manually fitted to β sheet regions of the 5 Å averaged map using the graphics program TOM/FRODO *(16)* and the second molecule was generated by the known twofold axis. The model was refined with TNT rigid body refinement *(17)* and its correctness was checked with difference Fourier syntheses to locate the heavy atom binding sites. A model was found after several trials that gave an *R* factor of 54.5% and all heavy atom sites. The relationship between two molecules was extracted and the process of the averaging and rigid body refinement was repeated. The converged model gave a 45.1% *R* factor in the resolution range of 20.0 to 3.0 Å. The final relation between two molecules was described by the Euler angles θ_1= 165.15°, θ_2= 10.20°, θ_3= 22.37° and a translation of 91.61Å, -3.69Å and -1.75Å, respectively, in an orthogonal coordinate system. The noncrystallographic twofold axis was tilted about 10° from the *c* axis. Two molecules were related by a rotation of 187.57° and the twofold center was moved about 2.4 Å from the initially assumed position.

The model phases then were combined with the MIR phases and the iterative averaging was performed with local program. The envelope was generated based on the MCM structure. The *R* factor dropped from 35.5 % to 15.5 % and the mean figure of merit increased from 0.66 to 0.94 with the resolution range of 10.0 to 3.0 Å. The average phase change from the MIR phases was 68.06°. Fig. 2-6 shows the 5.0 Å averaged map in the same section as shown in Fig. 2-5. The averaged map shows solvent channel along the *c* axis and the part of β sheet in the protein.

2-8. Model building and refinement

The first model building was mainly to modify the MCM structure based on the averaged map with the program TOM/FRODO. Significant changes in the main chain

were found after residue 76, due to a deletion at residue 78 and an insertion at residue 106. The two molecules in the asymmetric unit were built separately. The 3.0 Å averaged maps are shown in Fig. 2-7 for the external β sheet region and the chromophore binding site in molecule B. The densities for the carbonate group and a part of naphthoate group were present, but the connection between two densities was not in the averaged map.

The first and second models were refined with TNT and later models with X-PLOR. X-PLOR refinements included slowcooling from 3000 °K to 300 °K in 25 °K steps, and with individual B factor refinement. After each refinement, the model was modified based on the (2|Fo|-|Fc|) and (|Fo|-|Fc|) maps. At the sixth refinement, the resolution limit was extended to 2.4 Å and the *R* factor became 22.2% with r.m.s. deviations from ideal bond distances and angles of 0.016 Å and 3.22°, respectively. The chromophore was found only in molecule B, and not in molecule A, in the (|Fo|-|Fc|) map, even though the initial crystallization solution contained a 1:1 complex of the protein and the chromophore. The apparent dissociation of the chromophore from about half of the NCS molecules in crystals grown from the low pH, aqueous MPD solution is consistent with the known dissociation behavior of NCS in an acidic, organic medium. That only molecule B has the chromophore explains the weak density for the chromophore in the averaged map. The dictionary for the chromophore was written based on the NMR structure of the chromophore, provided by Dr. A. G. Myers. At this point, the native data set II was collected and used in refinements. The resolution was further extended to 1.8 Å and one MPD molecule was found in the chromophore binding cleft of molecule A. Water molecules were added based on the (2|Fo|-|Fc|) map, the hydrogen-bonding network and the B factor. Water molecules with the B factors greater than 40 $Å^2$ were deleted. Changes in final *R* factors during refinements are shown in Fig. 2-8. The numbers of degrees of freedom and of reflections for the final model are 7323 and 20513, respectively.

The final model contains two protein molecules, one chromophore, one MPD and 161 water molecules (1773 atoms total) in the asymmetric unit. This model has an *R*

factor of 19.9% between 5.0 Å and 1.8 Å resolution with r.m.s. deviations from ideal bond distances and angles of 0.014 Å and 2.84°, respectively. The average B factor of the protein and the chromophore is about 22 $Å^2$ and that of solvents is about 27 $Å^2$. The occupancies of the chromophore and an MPD molecule are 0.86 and 0.88, respectively.

Fig. 2-9 shows the (2|Fo|-|Fc|) map for the external β sheet region of molecule B, in the same orientation with Fig. 2-7 (a), and around an MPD molecule in molecule A. The electron densities for the chromophore are shown in Fig. 2-10. The holes within the enediyne ring and the naphthoate ring are clearly visible in the 1.8 Å resolution maps with 1.3σ contour. The aminosugar group has the weakest density in the chromophore.

The Ramachandran plots (Fig. 2-11) for apo-NCS (molecule A) and holo-NCS (molecule B) were prepared with PROCHECK *(17)*. More than 90% of the residues are in the most favored regions and no residue is in unfavorable regions. Pro^9 adopts the *cis* peptide conformation and is conserved in the structures of MCM and ACX. The stereochemistry of the main chain and that of the side chain are all very good based on PROCHECK analysis. The Luzzatti plot *(18)* (Fig. 2-12) estimates the maximum coordinate error as about 0.22 Å. The structure coordinates have been deposited in the Protein Data Bank (1NCO.PDB).

Table 2-1. Heavy atom derivative screening

compound	concentration (mM)	soaking time (hours)	*R*iso (20.0-3.0 Å) (%)
$Hg(CN)_2$	32.5	3.5	12.66
NaEMTS*	10.0	13	10.97
$Hg(C_2H_5)PO_4$	10.0	24	11.78
K_2HgI_4	20.0	3	10.63
$Hg(CH_3)_2$	saturated	4	8.91
$K_2Pt(NO_2)_4$	15.0	48	11.95
$PtNH_2CH_2CH_2Cl_2$	5.0	24	10.03
K_2PtCl_4	saturated	24	10.14
Pb(acethylacetonate)	saturated	24	11.28
$(CH_3)_3Pb(OAc)$!	10.0	24	11.23
$(CH_3)_3PbCl$!	14.3	48	10.04
$PbMoO_4$	saturated	48	8.02
$Sm(OAc)_3$!	10	24	14.00
$[(NH_3)_5IrCl]Cl$	saturated	60	9.15
K_3IrCl_6!	saturated	24	13.19
$AuCl_3$	12.5	7	10.03
$AuNaS_2O_3$	15	18	14.69
$K_3UO_2F_5$	10.0	24	7.59
KI	10	84	6.68
$NaWO_4$	10	24	5.93
$KReO_4$	saturated	18	6.00
CdI_2	10	24	6.77
Rhodamine Perchloride	12	24	8.40
$Pr_2(C_2O_4)_3$	saturated	20	8.60

*EMPS = Ethylmercury Thiosalicylate

! derivatives used in the phase calculation

Table 2-2. Data statistics of NCS

(a) Statistics of native data sets

	Native I	Native II
cell constants (Å)	a= b= 90.37	90.35
	c= 52.38	52.49
collected refls (no.)	31730	146707
unique refls (no.)	9024	23556
resolution (Å)	2.40	1.70
completeness(%)	92.7	87.8
overall <I/σ>	17.5	8.9
*R*merge*	7.11	7.62

(b) Statistics of native data set I in each resolution shell

resolution (Å)	4.35	3.45	3.02	2.74	2.55	2.40
possible refls (no.)	1656	1635	1618	1694	1612	1611
collected refls (no.)	1650	1629	1606	1588	1584	967
*R*merge	5.82	6.20	7.93	9.42	12.54	14.31

(c) Statistics of native data set II in each resolution shell

resolution (Å)	10.0	7.0	4.0	3.0	2.5	2.2	2.0	1.9	1.8
possible refls	143	262	1703	2840	3559	3938	4086	2730	3369
collected refls	149	264	1729	2841	3551	3890	3936	2440	2607
<I/σ>	11.85	11.76	12.86	10.73	7.01	4.35	3.12	2.47	2.22

(d) Statistics of heavy atom derivatives

data set	Pb1	Pb2	Sm	Ir
compound	$(CH_3)_3Pb(OAc)$	$(CH_3)_3PbCl$	$Sm(OAc)_3$	K_3IrCl_6
collected refls (no.)	20021	18553	24770	19985
unique refls (no.)	5883	5678	5865	6280
resolution (Å)	2.77	2.79	2.78	2.68
completeness (%)	94.6	91.1	94.3	91.5
$\langle I/\sigma \rangle$	15.0	14.6	14.1	26.2
Rmerge	7.23	10.43	9.07	4.70
Riso† (50.0 to 3.0 Å)	11.9	11.1	12.6	13.0

*$Rmerge = \Sigma_h\Sigma_i \mid I(\mathbf{h})_i - \langle I(\mathbf{h}) \rangle \mid / \Sigma_h\Sigma_i \; I(\mathbf{h})_i$, where $I(\mathbf{h})$ is the intensity of reflection **h**, Σ_h is the sum over all reflections, and Σ_i is the sum over the i measurements of reflection **h**.

†$Riso = \Sigma_h \mid F_{PH} - F_P \mid / \Sigma_h \; F_P$, where Σ_h is the sum over all reflections and F_P and F_{PH} are the native and derivative structure factor amplitudes respectively.

Table 2-3. Heavy atom binding sites and occupancies

data set	sites	occupancy	location in the structure
Pb1	0.956,0.341,0.000	0.300	Asp^{A33}
Pb2	0.955,0.342,0.018	0.210	Asp^{A33}
Sm	0.013,0.496,0.140	1.305	Glu^{A106} and Glu^{B106}
Ir	0.081,0.197,0.004	0.345	Arg^{B71}

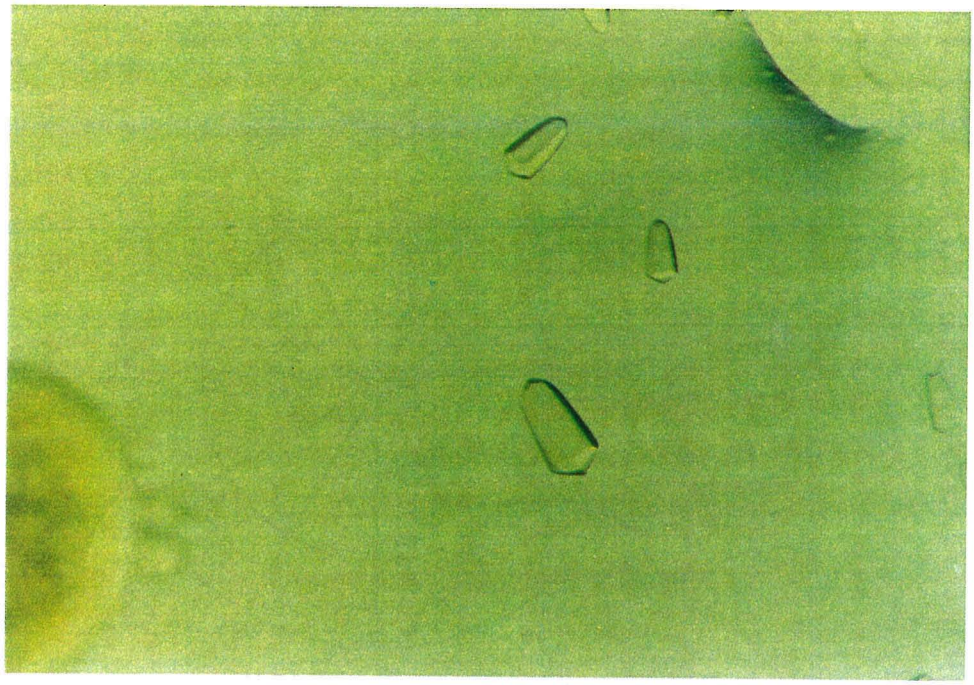
Table 2-4. Phasing statistics

res.	no. of rfls	mean F.O.M.	phasing power Pb1	Pb2	Ir	Irano*	Sm
7.74	164	0.76	1.87	1.32	1.39	2.28	1.48
6.32	253	0.73	2.08	1.37	1.68	1.25	1.35
5.33	348	0.71	1.90	1.35	1.25	1.33	1.00
4.62	489	0.63	1.55	1.26	0.91	1.11	0.52
4.07	630	0.56	1.34	1.06	0.79	0.84	0.28
3.64	788	0.50	1.37	1.00	0.61	0.57	0.15
3.29	978	0.54	1.46	1.37	0.58	0.62	0.10
3.00	1129	0.57	1.68	1.79	0.51	0.48	0.06
Total	4779	0.58	1.54	1.30	0.75	0.76	0.46

* Irano is the anomalous data of Ir derivative.

Figure 2-1. The NCS Crystal

(a) Crystals in the hanging drop



(b) Crystal on the spatula. This photo was taken after the data collection.

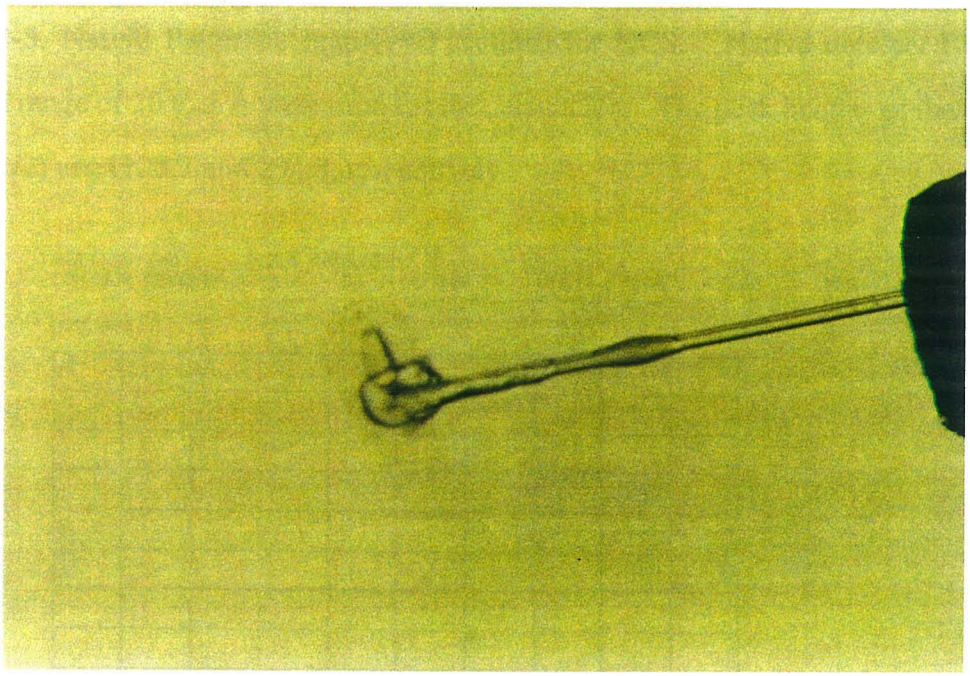


Figure 2-2. Crowther's self-rotation function of NCS. Native data set I and a resolution range of 10 to 5 Å were used for this calculation.

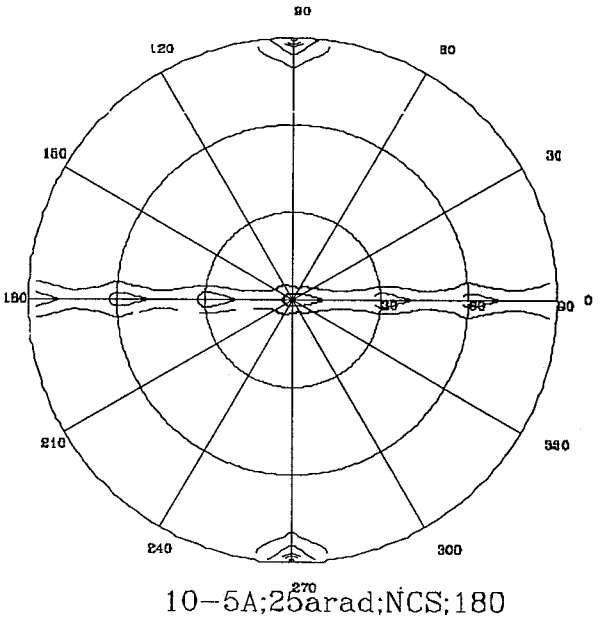


Figure 2-3. Native Patterson map (v=0 section) for NCS. Native data set I and a resolution range of 20 to 5 Å were used for the calculation. The peak heights of the origin and (0,0,1/2) are 1125.7 and 294.8, respectively

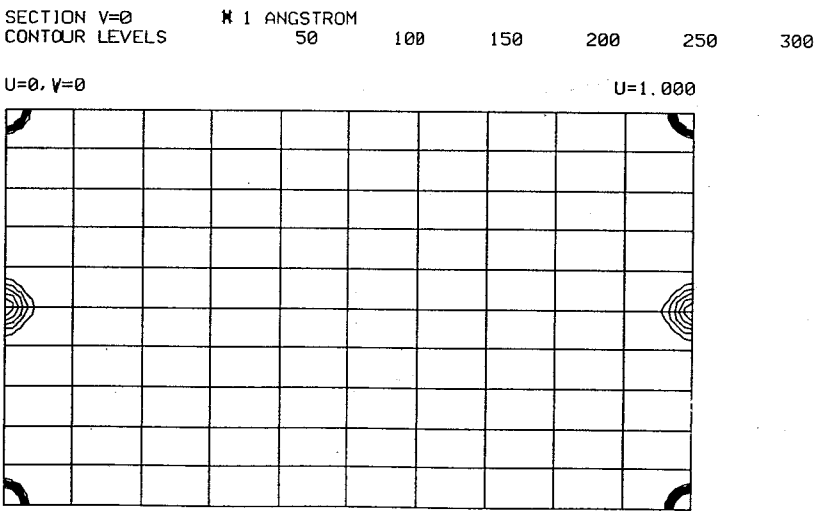
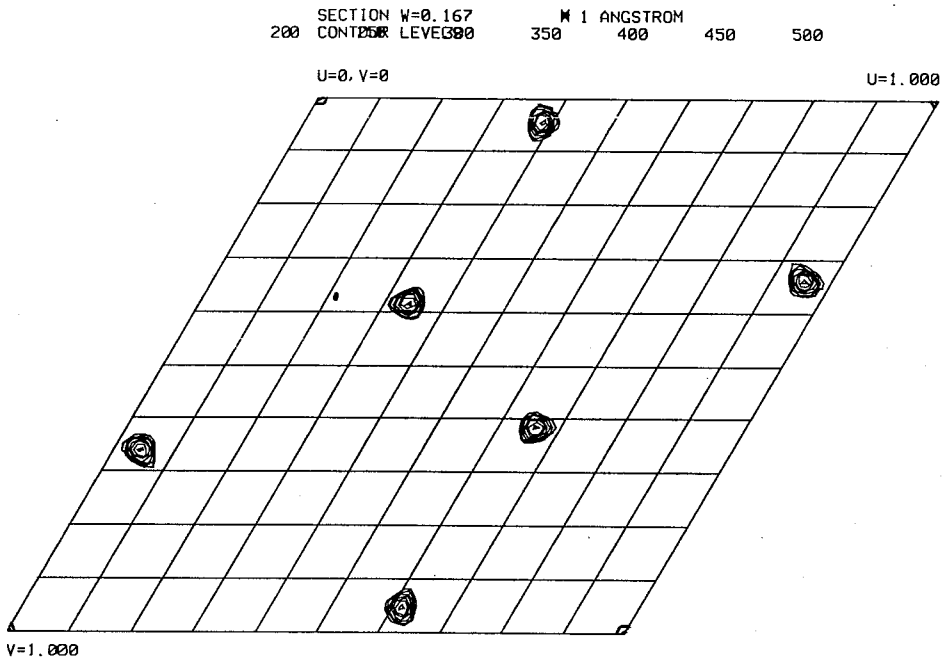
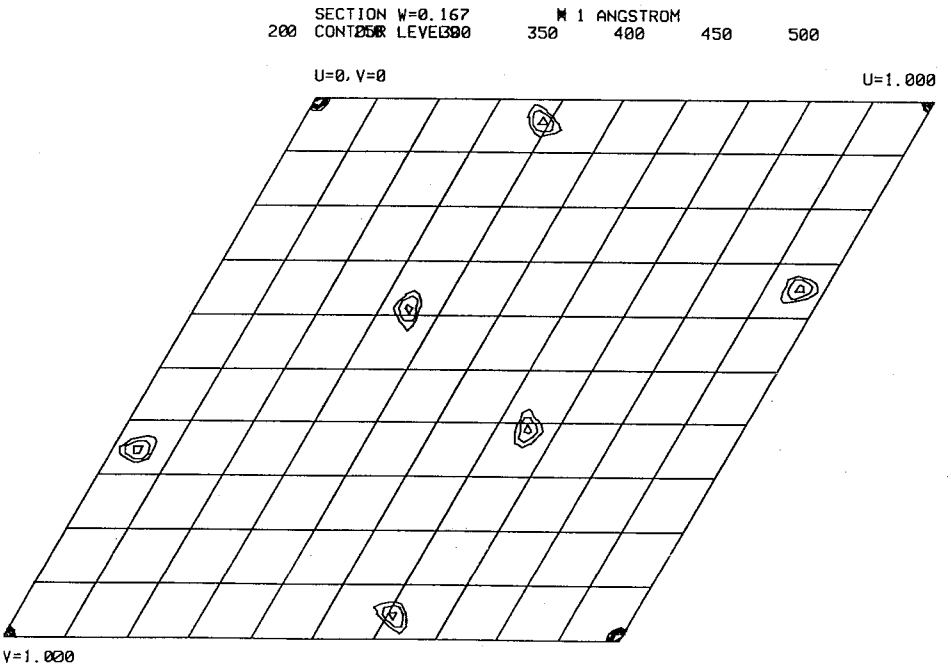


Figure 2-4. Isomorphous Difference Patterson maps. The Harker section of w=1/6 is shown for each derivative, calculated with a resolution range of 20 to 5 Å.

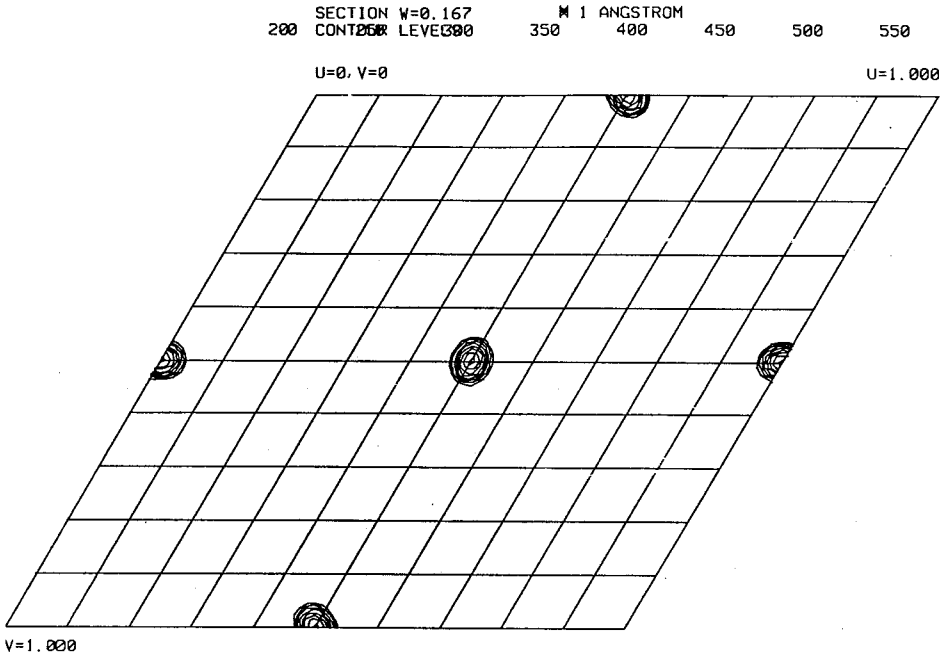
(a) $(CH_3)_3Pb(OAc)$



(b) $(CH_3)_3PbCl$



(c) $Sm(OAc)_3$,



(d) K_3IrCl_6

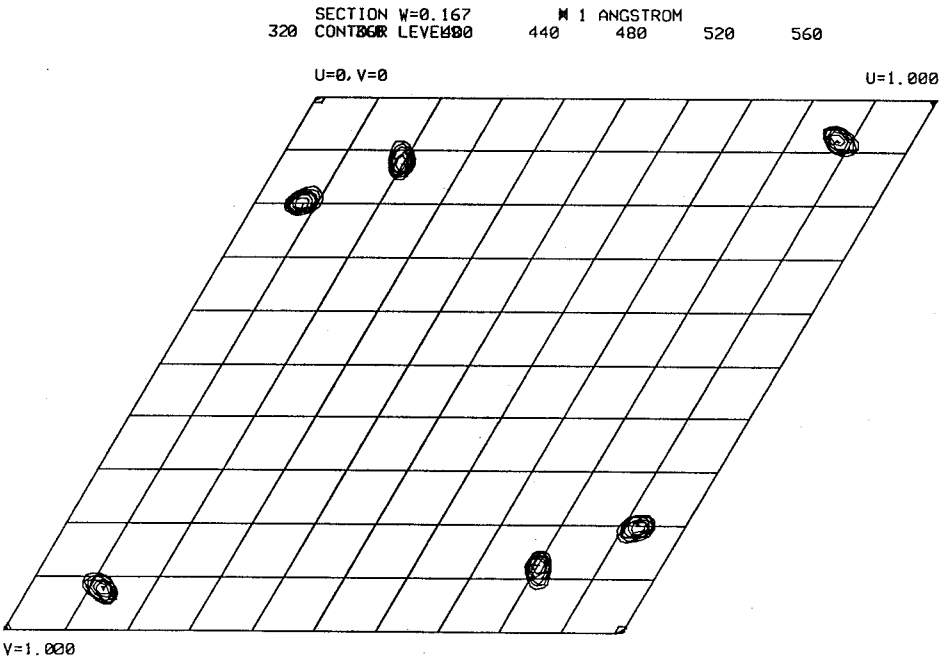


Figure 2-5. The 5.0 Å MIR map (z= 0 to 0.074)

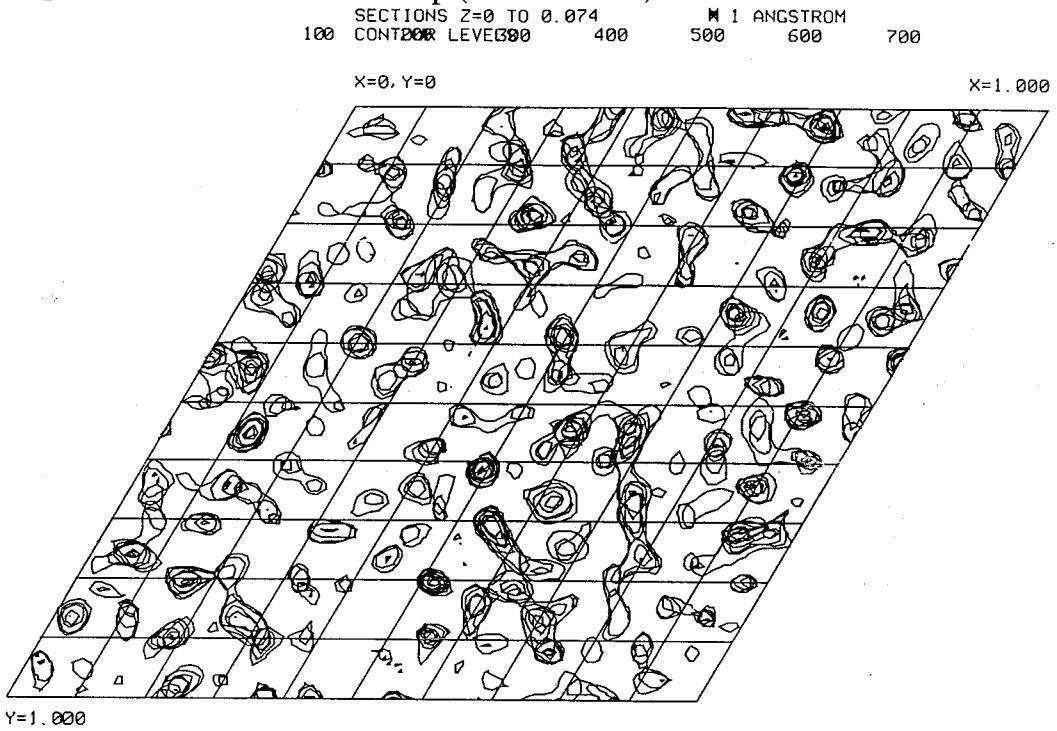


Figure 2-6. The 5.0 Å Averaged map (z= 0 to 0.074)

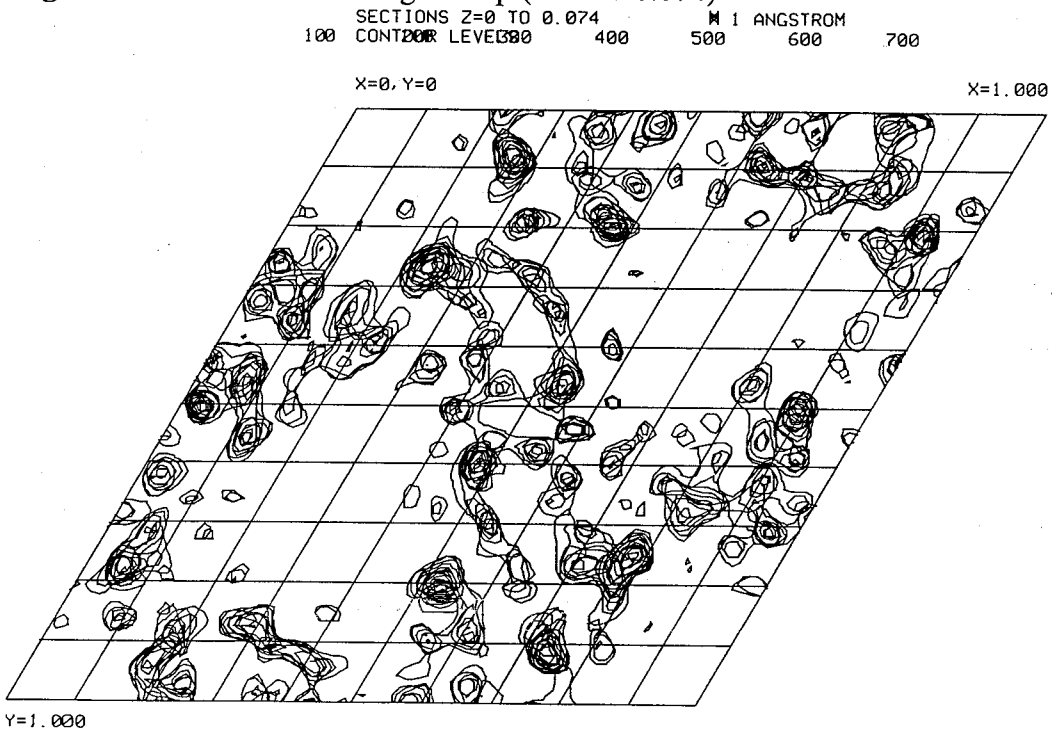
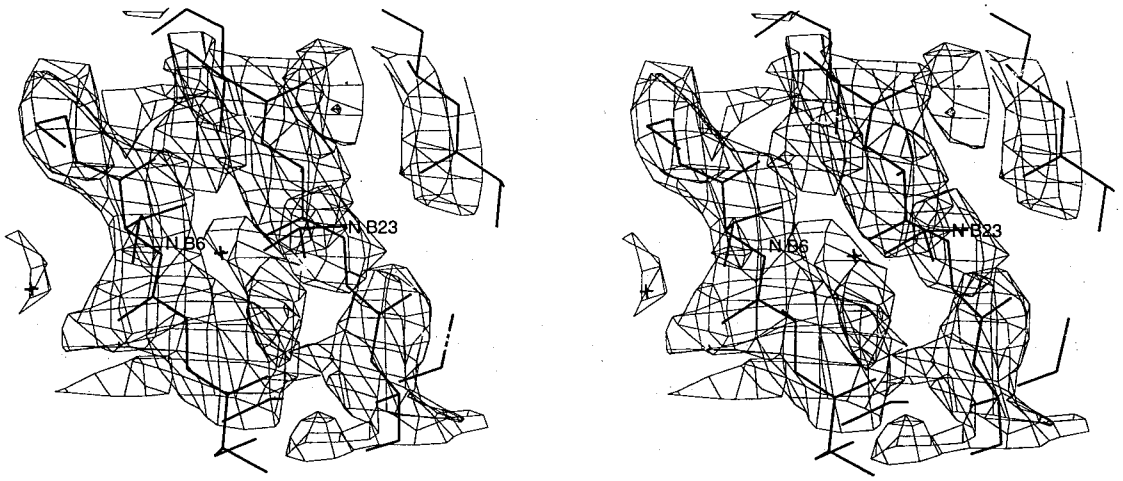


Figure 2-7. The 3.0 Å Averaged maps

(a) The external β sheet region of B molecule (1σ contour)



(b) The chromophore binding region of B molecule (1σ contour)

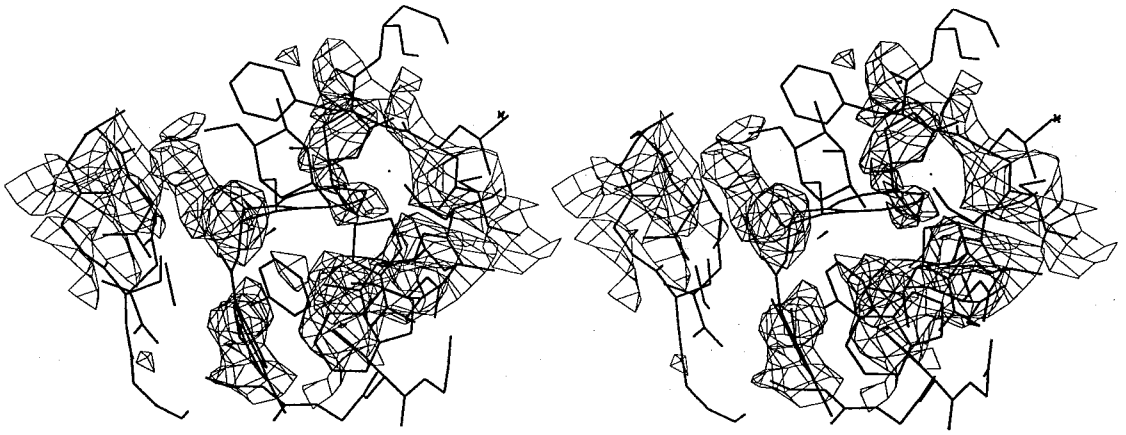


Figure 2-8. Refinements and *R* factors. First two refinements were performed with the program TNT and the rest were performed with the program X-PLOR. At the 7th refinement, the chromophore was added into the structure and at the 12th refinement, waters were added. The resolution was extended to 2.5 Å at the 3rd, to 2.4 Å at the 6th, to 2.0 Å at the 9th, to 1.9 Å at the 11th and to 1.8 Å at the 14th refinement.

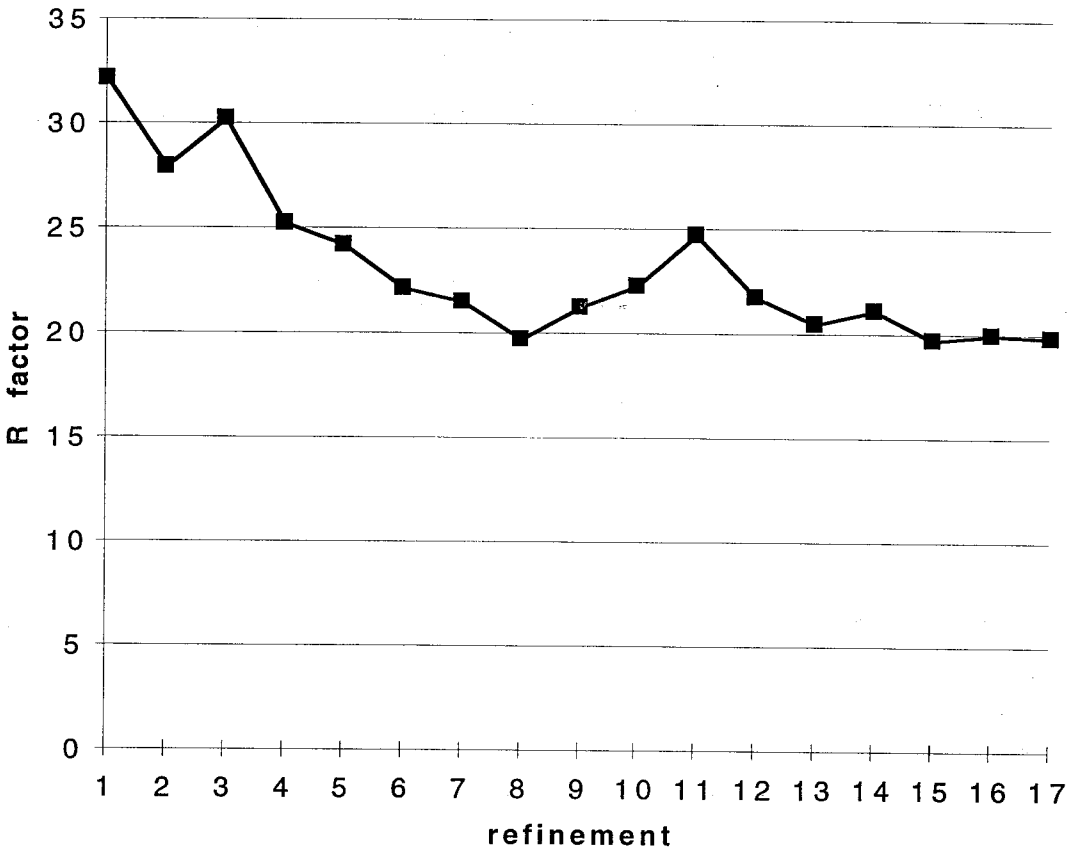
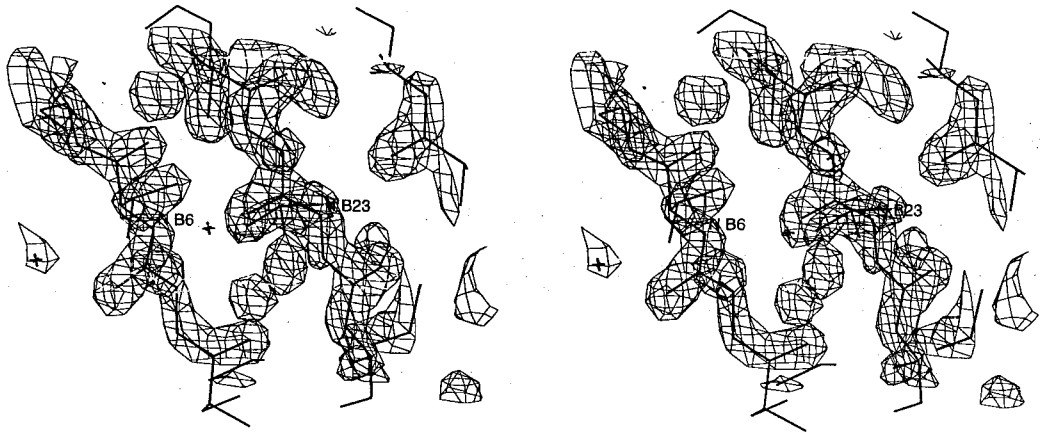


Figure 2-9. The (2|Fo|-|Fc|) electron density maps

(a) The external β sheet region of molecule B, shown in Figure 2-7 (a) (2σ contour)



(b) The density around the MPD molecule in molecule A (2σ contour)

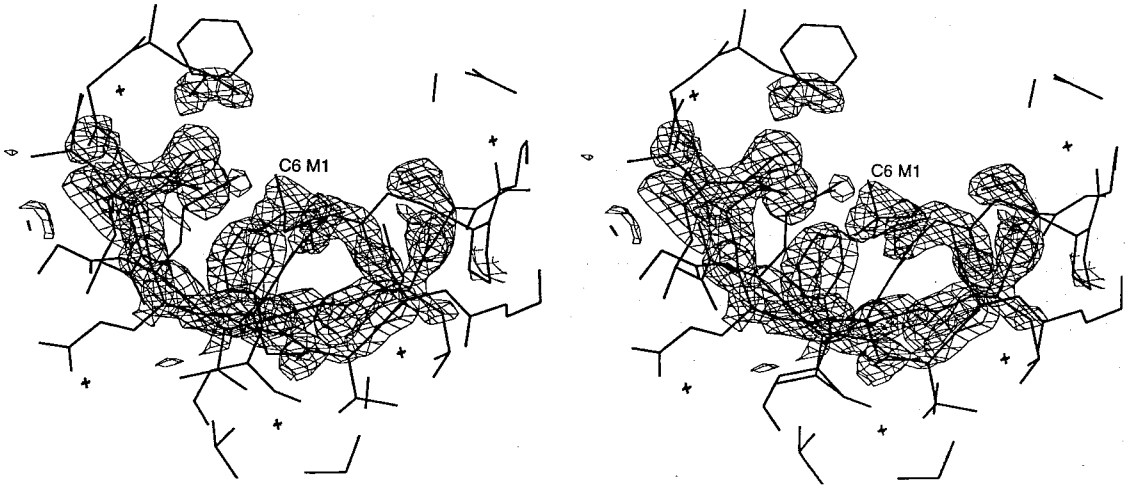
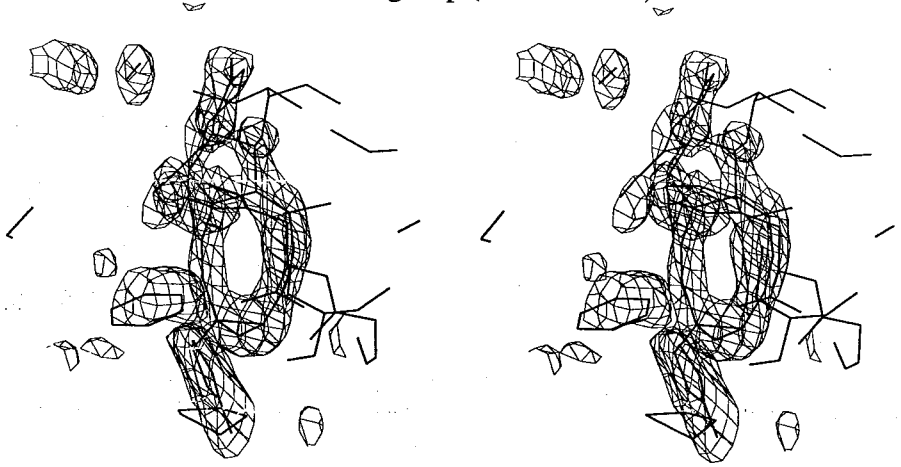
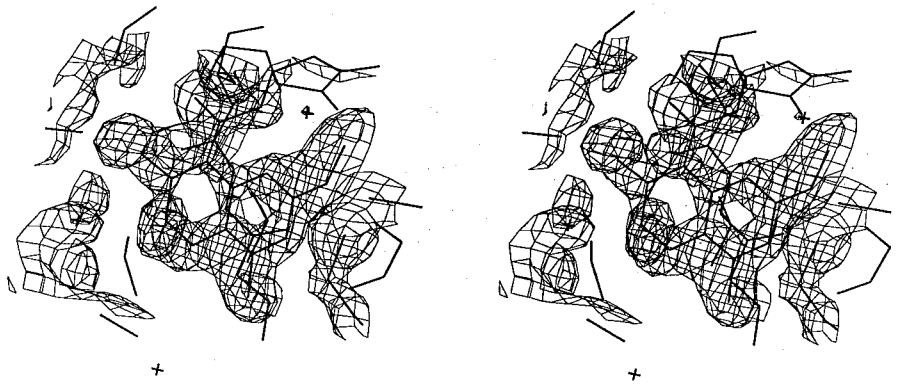


Figure 2-10. The (2|Fo|-|Fc|) maps of the chromophore

(a) The core ring and the cyclic carbonate group (1.3σ contour)



(b) The naphthoate group (1.3σ contour)



(c) The aminosugar group (1σ contour)

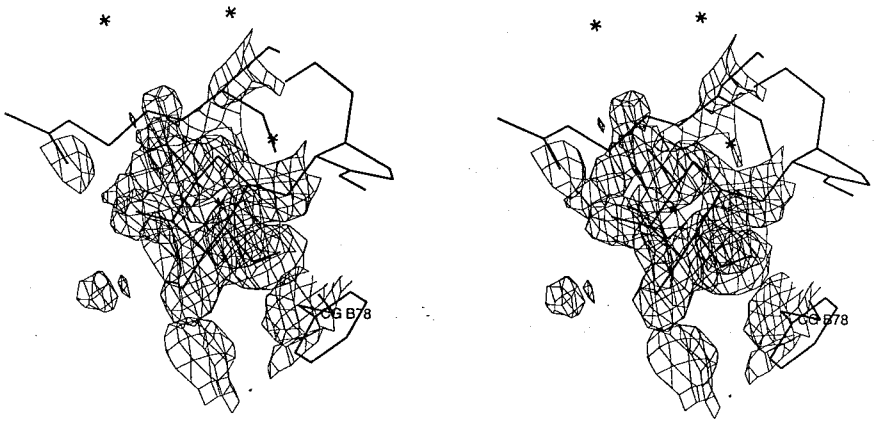
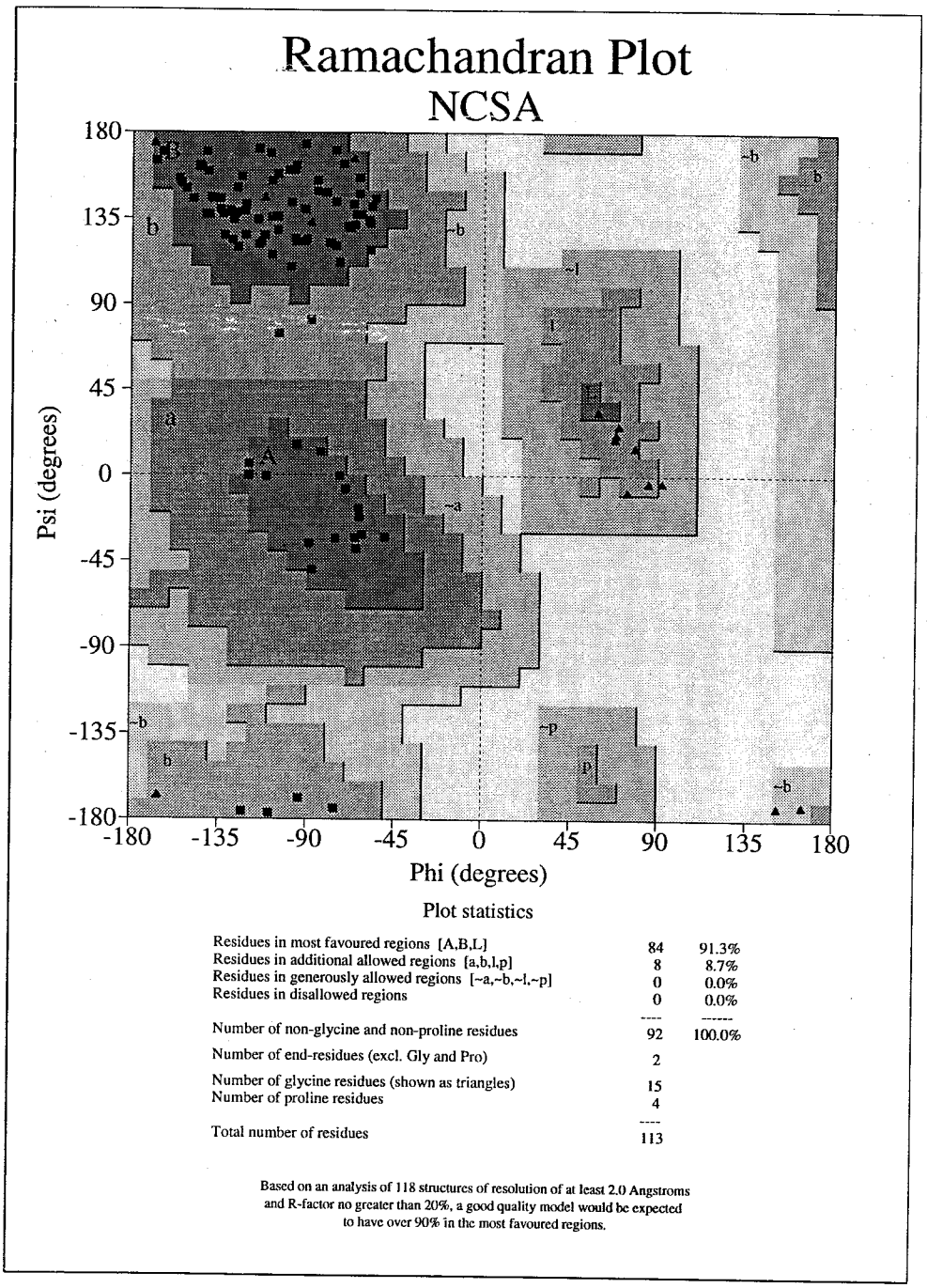
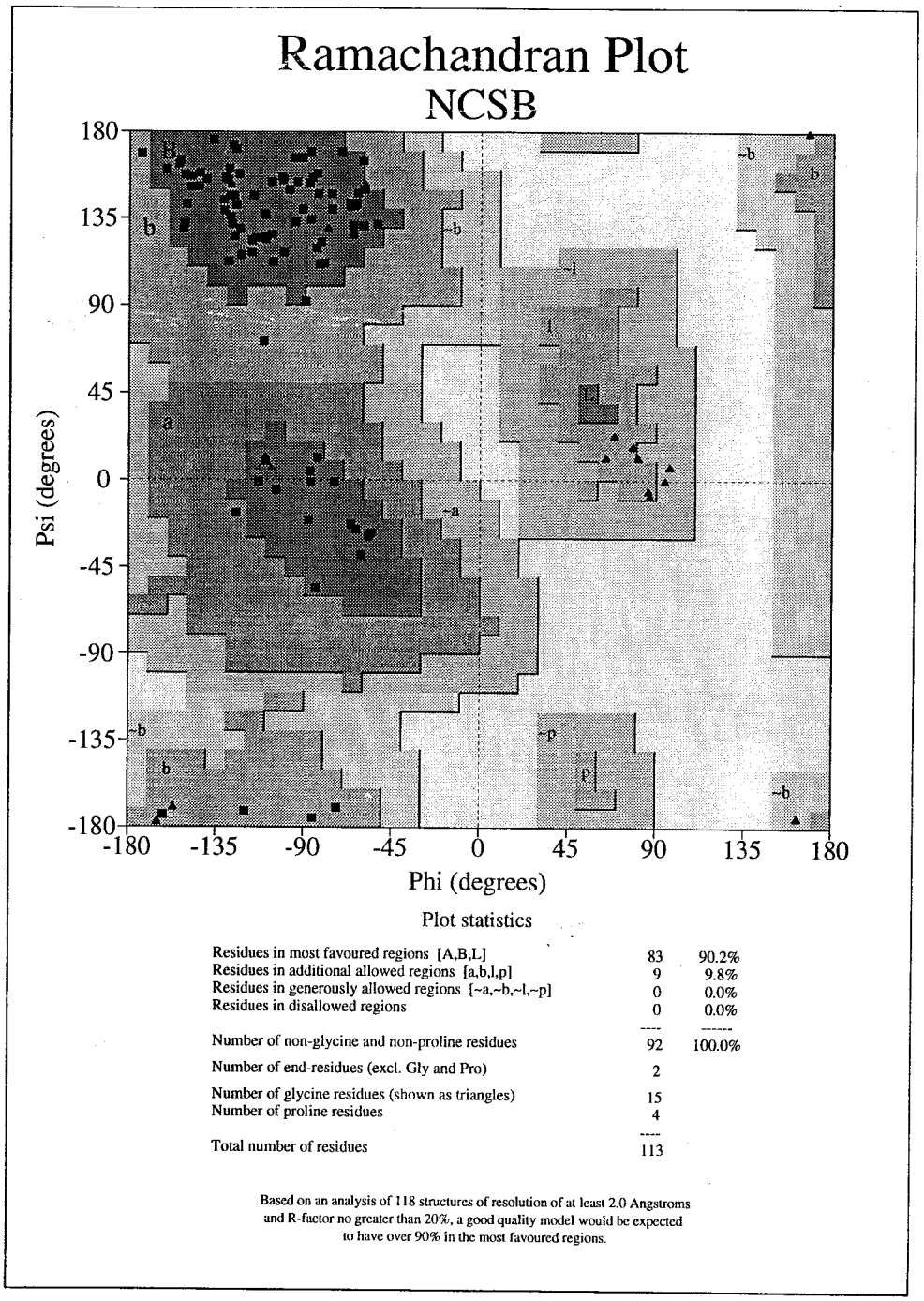


Figure 2-11. Ramachandran plots

(a) Apo-NCS



(b) Holo-NCS

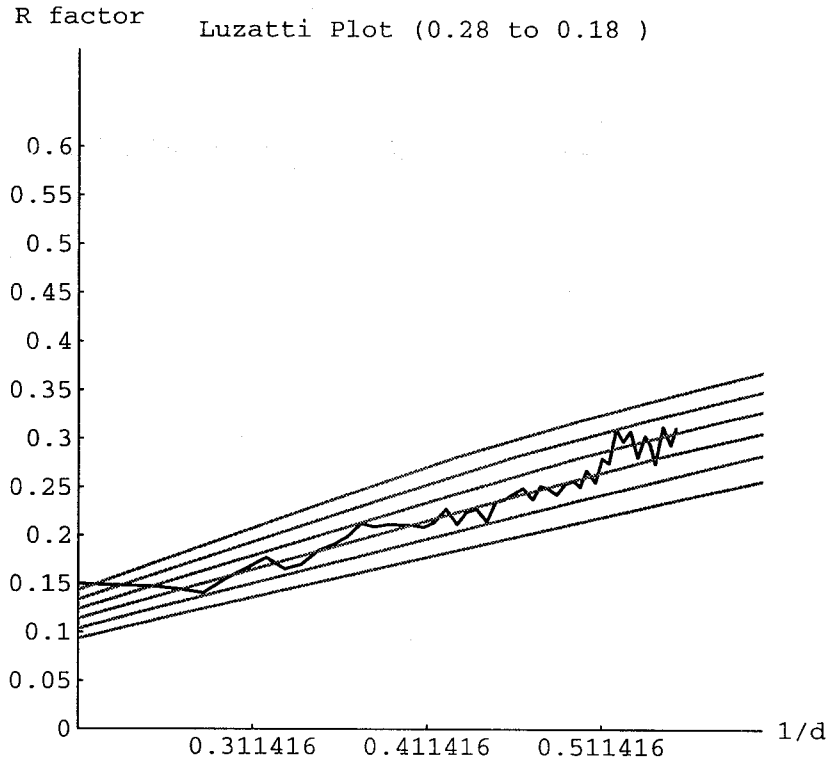


Plot statistics

Residues in most favoured regions [A,B,L]	83	90.2%
Residues in additional allowed regions [a,b,l,p]	9	9.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	92	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	15	
Number of proline residues	4	
Total number of residues	113	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 2-12. The Luzzatti plot. The *x* axis is the inverse of the resolution and three numbers correspond to 3.21 Å, 2.43 Å and 1.95 Å resolution. The dotted lines are errors of .18 Å, .20 Å, .22 Å, .24 Å, .26 Å and .28 Å from below and the estimated error of the structure is about 0.22 Å.



References

(1) K. Edo, K. Saito, Y. Akiyama-Murai, M. Mizugaki, Y. Koide, N. Ishida, *J. Antibiot.*, **41**, 554 (1988)

(2) L. C. Sieker, L. H. Jensen, T. S. Anantha Samy, *Biochem. Biophys. Res. Comm.*, **68**, 358 (1976)

(3) A. Teplyakov, G. Obmolova, K. Wilson, K. Kuromizu, *Eur. J. Biochem.*, **213**, 737 (1993)

(4) H. Hope, *Acta. Cryst.*, **B44**, 22 (1988)

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

(6) A. J. Howard *et al.*, *J. Appl. Cryst.*, **20**, 383 (1987)

(7) G. N. Reeke, *J. Appl. Cryst.*, **17**, 125 (1984)

(8) A. T. Brunger, *Acta Cryst.*, **A46**, 46-57 (1990)

(9) R. A. Crowther, in *Molecular Relacement Method*, edited by M. G. Rossmann, p173, New York: Gordon and Breach (1972)

(10) V. Z. Pletnev, A. P. Kuzin, S. D. Trakhanov, P. V. Kostetsky, *Biopolymers* , **21**, 287 (1982)

(11) P. Van Roey and T. A. Beerman, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 6587 (1989)

(12) PHARE was originally distributed by G. Bricogne (MRC, Cambridge, England) and subsequently modified for the VAX by J. L. Crawford and J. E. Ladner

(13) B. C. Wang, *Methods Enzymol.*, **115**, 41 (1985)

(14) J. M. Cox, *J. Mol. Biol.*, **28**, 151 (1967)

(15) T. A. Jones, *Methods Enzymol.*, **115**, 157 (1985)

(16) D. E. Tronrud, L. F. Ten Eyck, B. W. Matthews, *Acta Cryst.*, **A43**, 489 (1987)

(17) R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.*, **26**, 283 (1993)

(18) P. V. Luzzatti, *Acta Cryst.*, **5**, 802 (1952)

Chapter 3. Structure of Neocarzinostatin

3-1. Overview of the NCS structure

The structure of the NCS protein in both the holo- and apo-forms is L-shaped and has overall dimensions of 20 Å by 25 Å by 40 Å (Fig. 3-1 and 3-2). It consists primarily of β sheets organized into two domains, separated by a cleft for the chromophore binding. The large domain forms a seven-stranded antiparallel β barrel that contains two sheets: an external sheet consisting of strands 1, 2 and 5 and an internal sheet formed by strands 4, 3, 6, and 7 (Table 3-1). The small domain consists of two twisted, two-stranded antiparallel β sheets - strands 3a and 3b and strands 5a and 5b. These two β sheets are almost perpendicular to each other. In total, 53 residues are involved in β structures. The strand 3a is continuous from the β barrel to the small β sheet and is one of most conserved regions among different apoproteins. The disulfide bond of Cys^{37} - Cys^{47} closes a small β sheet with a β turn and that of Cys^{87} - Cys^{94} forms a ring with no regular secondary structure. There is a short segment of 3_{10} helix between residues Pro^{49} and Phe^{52}, which induces the stacking of the proline ring of residue 49 and the benzene ring of residue 52. In contrast, there are no helical segments in the structures of macromomycin (MCM) and actinoxanthin (ACX). With respect to the hydrogen bonding network in the β structure, the crystal structures are in agreement with NMR structures *(1-3)*.

The β barrel folding of NCS is similar to that of the immunoglobulin (Ig) variable domain (Fig. 3-3) *(4)*. Two differences are however noticeable. First, NCS does not have the Ig-like disulfide bond across the β barrel connecting strands 2 and 6. The second one comes from the manner in which the two sheets are packed at the level of the strand 4. In NCS, the strand 4 belongs to the second sheet; in contrast the strand 4 belongs to the first sheet in the Ig variable domain The topology with the strand 4 "sheet switching" is also observed in the structures of the bacterial chaperonin PapD *(5)*, the D2 domain of CD4

(6), a fibronectin type III domain *(7)*, and the extracellular domains of human growth hormone receptor *(8)*. Despite the topological similarity of the large domain of NCS with the immunoglobulin family, this similarity was not detected by a profile-based sequence analysis *(9)*.

NCS is a highly acidic protein, containing 8 aspartate and 2 glutamate residues but only 3 arginine and 1 lysine residues. Its hydrophilic property can be attributed to the fact that charged residues are fully exposed to the solvent (Fig. 3-4). Arg^{70} makes a salt bridge with the carboxylate group of Asp^{50} and has a close contact with the side chain of Asp^{15}. The conformations of side chains of these three residues are well conserved both in apo-NCS and in holo-NCS, in spite of their solvent accessibility. The acidic groups of the protein appear to play an important role in its biological activity *(10)*. The modification of the carboxyl groups into isopropylamide groups resulted in a broad shift of the protein pI toward higher values, accompanied by a marked reduction of toxicity *in vitro* and reduced binding of protein to cells. It has also shown that the modified protein has lower affinity for the chromophore. Lys^{20} is isolated from other charged residues in accord with the report that acylation of this residue results in no change in the protein conformation *(11)*. SMANCS, a conjugate of poly(styrene-co-maleic acid) with NCS, was made by attaching the polymer to the amino terminal and amino groups of Lys^{20} for the purpose of stabilizing the complex in the blood stream and reducing the toxicity *(12)*.

Fig. 3-5 shows the side chains of all Ile, Leu, Trp, Phe, Tyr and Val residues, with the same orientation as that used in Fig. 3-4. In contrast to the hydrophilic residues, many of small hydrophobic residues (nine valines and three leucines) are buried in the protein interior of the large domain. Six out of eight aromatic residues are present in the small domain. Trp^{39}, Phe^{52}, Phe^{76} and Phe^{78} are involved in the chromophore binding while Trp^{83} is exposed to the solvent, as predicted in the oxidation experiment with N-bromosuccinimide *(13)*. Tyr^{32} and Phe^{112} are buried in the protein and are well conserved among different apoproteins, except Ala^{32} in kedarcidin. Aromatic residues appear to be

important in the side chain packing. Stacking interactions are observed between the methyl groups and the aromatic rings for residues Val^{86}/Phe^{73}, Leu^{13}/Phe^{112}, Val^{45}/Trp^{39} and Val^{40}/Trp^{83}.

Two NCS protein molecules are present in the asymmetric unit (Fig. 3-6). One is apo-NCS (designated with residue numbers prefixed with "A") and the other is holo-NCS (designated by residue numbers prefixed with "B"). Apo-NCS contains an MPD molecule in the binding pocket. The dimer interactions between the two molecules occur between loop A of apo-NCS and loop C of holo-NCS and *vice versa*. Seven intermolecular hydrogen bonds are found between apo-NCS and holo-NCS in the asymmetric unit (Table 3-2). The solvent accessible surface areas of proteins are reduced about 9% by the dimer contacts. The amino terminal regions of two molecules in the asymmetric unit are parallel to each other and the noncrystallographic twofold axis goes through their center.

3-2. The NCS chromophore

The resolution of the refined holo-NCS structure is sufficiently high to establish the details of the chromophore structure (Fig. 3-7). The NCS-chrom structure is in complete agreement with the structure that was proposed on the basis of synthesis data and ^{1}H NMR spectroscopy *(14)*, including stereochemical assignments. The epoxy nonadiyne ring of the chromophore is essentially planar, with an r.m.s. deviation of the nine carbon atoms from the least squares plane of only 0.06 Å. The four C-C≡C bond angles of the two triple bonds in the enediyne ring, which were not restrained during refinement, display marked nonlinearity with an average value of 161.5° ± 1.2°. The aminosugar group of the chromophore adopts a chair conformation, with the amino group oriented above C12 at a distance of 4.97 Å that would accommodate the sulfur atom of a thiol between these two groups. This relationship supports the proposal that the amino group plays a catalytic role in the thiol activation mechanism *(15)*. Differences between NCS-chrom in the protein and the free form exist in the torsion angles of the functional groups (Fig. 3-7 (b)).

NCS-chrom is bound in the hydrophobic pocket formed by the internal β sheet of the large domain and the small domain. The binding pocket can be divided into three parts. Residues 34 to 52 and 94 to 98 form the bottom of the pocket for the naphthoate group. Loop C with residues 76 to 80 interacts with the carbonate group and loop A with residues 100 to 102 contacts the aminosugar group. Two loops are separated by the side chain of Phe^{B78} in the structure. The enediyne ring system is placed above the Cys^{B37} - Cys^{B47} disulfide bond and the naphthoate group is placed at the bottom of the pocket. The aminosugar and the carbonate groups are facing outwards and the epoxide is facing down towards the bottom of the pocket. The stabilization of NCS-chrom is achieved by forming van der Waals contacts with many apolar residues and several aromatic residues in the binding pocket. Table 3-3 lists residues and numbers of contacts with the chromophore. Gln^{B94} and Ser^{B98} are the only polar side chains around the chromophore. Phe^{B78} interacts with the chromophore most extensively.

The two π-faces of the nine-membered enediyne ring are sandwiched between the Phe^{B78} benzene ring on one side and the Phe^{B52} benzene ring and the Cys^{B37} - Cys^{B47} disulfide bond on the other. The side chains of Phe^{B52} and Phe^{B78} are perpendicular to the enediyne ring (Fig. 3-8 (a)), which is a favorable orientation often exhibited by aromatic systems *(16,17)*. The distance between C12 and the Cδ1 atom of Phe^{B78} is 3.85 Å and the solvent accessible area of C12 decreased from 24.2 $Å^2$ in the free form to 2.5 $Å^2$ in the protein-chromophore complex. The positioning of the Phe^{B78} side chain, in particular, should stabilize the chromophore by precluding attack of nucleophiles at C12. The C2 - C3 triple bond has close contacts with the Cδ1 atom of Phe^{B78} and the Cε2 atom of Phe^{B52}. The C6 - C7 triple bond is just above the sulfur atom of Cys^{B47}. The distances from C6 and C7 to the Sγ atom of Cys^{B47} are 3.57 Å and 3.30 Å, respectively. The other side of the C6 - C7 triple bond has contacts with the main chain atoms of Phe^{B76}, Leu^{B77} and Phe^{B78}. The epoxide is completely buried by the side chains of Cys^{B47}, Pro^{B49} and Phe^{B52}.

The naphthoate group is in proximity to the bottom of the pocket where strand 3 and 6 are exposed (Fig. 3-8 (b)). The base of the binding pocket is very flat due to the absence of side chains at positions 35 and 96, both of which are glycines. The naphthoate group forms two distorted hydrogen bonds to the protein - one between the carbonyl oxygen atom and the Oγ atom of Ser^{B98} (3.09 Å), and the other between the methoxy oxygen atom and the nitrogen atom of Gly^{B35} (3.14 Å). There is another hydrogen bond occurring at a lattice contact between the naphthoate hydroxyl oxygen atom and the nitrogen atom of Ala^{A1}, the amino terminus of the other protein molecule in the asymmetric unit (2.64 Å). The methyl group (C12") of the naphthoate has close contacts with Gln^{B94}, Val^{B95} and Gly^{B96} and the methoxy carbon (C11") is close to Gly^{B35} and Phe^{B52}. The sulfur atom of Cys^{B37} has contacts with the C6" - C9" site of the naphthoate and Leu^{B45} has contacts with the C1" - C4" site. The hydroxyl oxygen is close to the carboxyl oxygen of Gly^{B102} (3.71 Å). Two contacts of Trp^{B39} are made with C3" in the naphthoate; distances from C3" to Cη2 and Cζ3 atoms are 3.54 Å and 3.97 Å, respectively.

The sugar methylamino group stacks against the Phe^{B78} benzene ring in holo-NCS, which suggests that the amino group in the crystal is protonated. The distance between the nitrogen atom of methylamino group and the benzene ring of residue 78 is about 3.5 Å. The favorable interaction between a (presumed) positively charged group and the π-face of an aromatic system has been observed in a variety of protein and model systems *(18-20)*. The carbon atom of methylamino group is close to the loop region of Ala^{B100}, Ala^{B101} and Gly^{B102}. The carbonate group of the chromophore has contacts with the side chain of Pro^{B49} and the carboxyl oxygen atoms of Leu^{B77} and Phe^{B78}. The carbonyl oxygen of the carbonate has contacts with Gly^{B80}.

The solvent accessible areas for the chromophore and the protein are calculated as defined by Lee and Richards *(21)*. The bound chromophore exposes ~171 $Å^2$ of surface area to the solvent, or about 18% of the total accessible area of the free NCS-chrom. The solvent-exposed regions of NCS-chrom are provided primarily by the aminosugar, the

carbonate, and the naphthoate hydroxyl groups (Fig. 3-8 (d)). The complete burial of the active epoxide in the protein suggests that the protein stabilizes the chromophore by sequestration of the active group from the solvent. The changes in solvent accessible areas of the protein before and after the chromophore binding are compared in Table 3-4. The biggest change occurs in residue 78. Cys^{B37}, Cys^{B47}, Leu^{B45}, Gly^{B96} and Ser^{B98} are completely buried upon the chromophore binding. The surface areas of Gly^{B102}, Trp^{B39}, Phe^{B52} and Pro^{B49} are also decreased significantly.

The binding interactions of the chromophore to the protein observed crystallographically differ significantly from models based on NMR spectroscopy. In the modeling study *(18)*, the triple bonds of the chromophore were located around the disulfide bond Cys^{37}-Cys^{47} and the carbonate group interacted with Ser^{98} and Leu^{45}. The naphthoate group formed an aromatic stacking interaction with Phe^{52} and the aminosugar group interacted with Asp^{33}, Ser^{54} and Asp^{99} through the hydrogen bonds. The enediyne ring in this model positioned about the same location found in the crystal structure but its orientation was upside down. Positions of aminosugar and carbonate groups are switched with each other and the naphthoate position is completely wrong. No stacking interaction of the naphthoate group with the benzene ring of Phe^{52} or the hydrogen bonds with Asp residues were observed.

In the binding conformation of the chromophore determined by NMR *(23,24)*, the enediyne ring was located on the Cys^{37} - Cys^{47} disulfide bond and the aminosugar and the carbonate groups were facing outwards. The C2 - C3 triple bond was just above the sulfur atom of Cys^{47} and its side is covered by Phe^{52}. The benzene ring of Phe^{78} and the side chain of Leu^{45} were close to the enediyne ring. There was a salt bridge between the protonated aminosugar methylamino group and the Asp^{33} carboxylate. This chromophore binding position is very similar to that of the crystal structure, but detailed interactions are different, such as the hydrogen bonds and the salt bridge. The chromophore seemed to

move about 1 Å closer to loop A in the solution structure and differences in the torsion angles of the functional groups are also noticeable.

3-3. Comparison between apo-NCS and holo-NCS

The structures of apo-NCS and holo-NCS are similar, with r.m.s. deviations of 0.59 Å for 113 Cα atoms and 0.31 Å for 35 Cα atoms in the seven-stranded β barrel of the large domain. Differences between apo-NCS and holo-NCS are found in strands 5a and 5b, and in three loops around the binding pocket (Fig. 3-9). The most significant change is the position of the Phe^{78} benzene ring. This side chain rotates from a stacking position over the enediyne ring in holo-NCS to a solvent-exposed position in apo-NCS, where it abuts the side chain of Asp^{A79}. The distance between the Phe^{A78} benzene ring and the Asp^{A79} carboxylate is about 4.2 Å (Fig. 3-10). This amino acid interaction may underlie the pH-sensitive dissociation of chromophore from NCS. Differences between apo-NCS and holo-NCS increase the width of the binding pocket about 5 Å to accommodate the chromophore. The distance between the Gly^{B102} Cα atom and the Phe^{B78} Cζ atom is 4.54 Å and the distance between the Gly^{A102} Cα atom and the Phe^{A78} Cβ atom is 9.54 Å. The opening of the binding pocket is contributed by the repositioning of the Phe^{78} benzene ring. In the absence of NCS-chrom, a hydrogen bond in the crystal between MPD and the Oγ atom of Ser^{A98} of apo-NCS (2.57 Å) replaces the distorted hydrogen bond between the chromophore and the Oγ atom of Ser^{B98} in holo-NCS.

The mean B factors of apoproteins in apo-NCS and holo-NCS are 23.26 $Å^2$ and 21.90 $Å^2$, respectively. B factors averaged over the main chain atoms are plotted in Fig. 3-11. The loop between strands 3a and 3b has the highest B factors. The biggest differences between apo-NCS and holo-NCS are found in the loop between strands 6 and 7, which participates in the chromophore binding. Strands 4 and 5 and the loop between them have significantly higher B factors in apo-NCS than in holo-NCS.

Teplyakov *et al.* determined the crystal structure of apo-NCS at 1.5 Å resolution *(25)* whose coordinate is available from Protein Data Bank as 1noa.pdb. The crystals were obtained from ammonium sulfate and the structure was solved by the molecular replacement method using the ACX structure as a model. The r.m.s. deviations of the main chain atoms from apo-NCS and holo-NCS are 0.54 Å and 0.73 Å, respectively. Differences between 1noa and holo-NCS are found in residues 42 to 43, 78 to 84 and 100 to 102, which are the loops around the binding pocket (Fig. 3-12). The conformations of side chains are in agreement with those of apo-NCS and holo-NCS, except the Phe^{78} benzene ring and the Trp^{39} ring. The Phe^{78} benzene ring intrudes into the binding pocket and the distances from the Gly^{102} Cα atom to Cζ and Cβ atoms of Phe^{78} are 7 Å and 9.29 Å, respectively. The side chain conformations of Asp^{15}, Asp^{51} and Arg^{70} are also well conserved in the 1noa structure.

3-4. Comparison with other apoproteins

The NCS, MCM, and ACX proteins share the same basic structural architecture. They consist of a seven-stranded β barrel and a small domain of two antiparallel β sheets. The MCM structure *(26)* was determined to 1.6 Å resolution with the final *R* factor of 17.9%. It was well refined as shown in the Ramachandran plot (Fig. 3-13) and has higher content of β structures (66 residues) than NCS. The ACX structure *(27)* was determined to 2.5 Å resolution and was not refined well. Two residues of ACX lie in disallowed regions of Ramachandran plot and only 65% of residues are in most favorable regions. The structure has several breaks in β strands based on PROCHECK analysis *(28)*. The aromatic residues in ACX contribute to the hydrophobic interior of the large domain; in contrast, half of them participate in chromophore binding in NCS. The structure of the kedarcidin apoprotein has recently been determined by NMR spectroscopy *(29)* to have similar features.

Twenty-two residues are conserved among five apoproteins in the family of antitumor antibiotic chromoprotein, including two disulfide bonds (Fig. 1-5). The most conserved sequence region in the structure is the strand 3a that connects the large and small domains. It has three conserved residues (Gln^{36}, Cys^{37}, Ala^{37}) out of five resdiues total and must be important to the architecture of the protein. The Nε2 and Oε1 atoms of Gln^{36} form two hydrogen bonds with carboxyl oxygen and amino nitrogen atoms of Arg^{70}, which is also a conserved residue in every protein except in MCM. Residues 80 to 88 are also highly conserved, though they are exposed to the solvent. Strand 6 and the loop of residues 99 to 107 form the least conserved region. They are important in chromophore recognition as indicated in the solvent accessible area changes before and after chromophore binding in NCS.

The MCM structure can be superimposed to apo-NCS and holo-NCS with r.m.s deviations of 1.02 Å and 1.04 Å, respectively, for 111 common Cα atoms. The r.m.s. deviations of the ACX structure from apo-NCS and holo-NCS are 1.84 Å and 1.97 Å, respectively, for 107 common Cα atoms. As shown in Fig. 3-14, the three-stranded β sheet of NCS superimposes well on the other two structures, while the four-stranded β sheet is shifted significantly across the strands. The major differences occur in the loops around the binding cleft. Differences in Cα positions of equivalent residues of the loops go up to 6.9 Å for Gly^{102} of NCS and Leu^{98} of ACX.

Chothia and Lesk parameterized the relation between the divergence of sequence and structure in proteins *(30)*. Core residues are defined as, the main chain atoms of major elements of secondary structure, plus additional atoms at both ends of elements of secondary structure, as long as the deviation in the positions of the atoms are no greater than 3 Å. The r.m.s. differences in the positions of main chain atoms between two homologous proteins can be fit as an exponential function of the fraction of mutated residues. Table 3-5 shows the predicted and measured r.m.s. differences for the NCS, ACX and MCM structures. The common core of these proteins contains 90 residues (80%

of the total) and the predicted values are within 36% of observed values. The r.m.s. deviations for the common core are 0.76 Å between holo-NCS and MCM, 1.27 Å between holo-NCS and ACX, and 1.17 Å between MCM and ACX. The regions outside the common core (residues 26 to 30, 41 to 44, 78 to 79 and 99 to 107) are less conserved and except for Phe^{78} and Asp^{79}, correspond to the complementarity-determining regions in the Ig variable domains.

The binding pockets of NCS, ACX, and MCM have similar overall shapes, but the locations of side chains around the binding pockets are quite different. Among 20 residues close to NCS-chrom, only 3 - Cys^{37}, Cys^{47}, and Gly^{80} - are conserved. The chromophore binding pocket is biggest in kedarcidin, followed by ACX and smallest in NCS and MCM. The distance between the Cα atoms of Met^{77} and Gly^{104} is 18.86 Å in kedarcidin and that between the Cα atoms of Pro^{74} and Asn^{95} is 12.69 Å in ACX. The distance between the Cα atoms of Ala^{78} and Asp^{100} is 9.61 Å in MCM. These results are consistent with considerably large chromophore of kedarcidin.

In NCS and MCM, the bottom of the binding pocket is flat, due to residues Gly^{35} and Gly^{96}. Ala^{35} replaces Gly^{35} in ACX and Leu^{35}, while Val^{97} are the corresponding residues in kedarcidin. Kedarcidin also has the side chains of Val^{49} and Phe^{52}, as well as the disulfide bond, in the bottom of the pocket, which is crowded with side chains and retains the hydrophobic nature. The functional group at the bottom of the pocket in kedarcidin should be smaller than the naphthoate group of NCS-chrom. It is likely that the enediyne ring of the kedarcidin chromophore is located above the disulfide bond, with other functional groups facing outward. The wall of the binding pocket in NCS is formed by strands 3a and 3b on one side and by strands 3 and 6 on the other, and contains the disulfide bond and Leu^{45}, Gln^{94} and Ser^{98}. The two exits are closed by Trp^{39}, Pro^{49} and Phe^{52} to limit the solvent accessibility of the naphthoate group (Fig. 3-10). The corresponding residues are Val^{38}, Ile^{44}, Ala^{48}, Ser^{51}, Ser^{94}, and Gly^{98} in MCM and Pro^{36}, Gln^{40}, Pro^{45}, Ala^{48}, Asn^{90} and Gly^{94} in ACX. The hydrophobic nature of the

wall is conserved in both MCM and ACX, but the two exits are more open to solvent (Fig. 3-15). The functional groups at the bottom of the binding pocket in MCM and ACX are expected to be bigger than the naphthoate group in NCS-chrom.

Loop A in NCS contains small amino acids - Ala^{100}, Ala^{101} and Gly^{102} - while loop C has the residue Phe^{78} and the main chain atoms of Phe^{76} and Leu^{77}. In MCM, loop A is replaced with Asp^{100}, Ser^{101} and Gly^{102} and loop C with Val^{75}, Val^{76}, Gly^{77} and Ala^{78}. Asp^{100} might interact with the MCM chromophore through a hydrogen bond or a salt bridge. Loop C of MCM has small amino acids as loop A of NCS. Asp^{79} does not seem to interact with the chromophore, since its side chain pointed away from the binding pocket. His^{32} is likely to participate in the chromophore recognition, though Asp^{33} in NCS is positioned away from the chromophore. In ACX, loop A is replaced with Ser^{96}, Gly^{97} and Leu^{98} and loop C with Gln^{72}, Thr^{73} and Pro^{74}. The side chain of Leu^{98} is pointed away from the binding pocket and Asp^{99} and His^{102} as well as Tyr^{30} might be involved in the chromophore recognition. The kedarcidin apoprotein has Asn^{101}, Thr^{102} and Gly^{103} in loop A and Tyr^{76}, Val^{77}, Met^{78} and Pro^{79} in loop C. Loop A has similar side chains to loop A of MCM. Tyr^{106} and Met^{78} are likely to participate in the chromophore recognition, since they displayed high disorder in the NMR structure of the kedarcidin apoprotein.

The position and orientation of the enediyne ring system above the disulfide bond is expected to be different among the different apoproteins. NCS-chrom nicely fits into the binding pocket of MCM, but not to the binding pocket of ACX. The binding selectivity of a chromophore and its partner protein must be the result of interactions between the chromophore and specific side chains, such as Phe^{78} of NCS, that represent variable elements within the protein structures. The detailed comparisons on the chromophore recognition by different apoproteins will only be possible when the structures of other chromoproteins are determined.

Table 3-1. Elements of secondary structure in NCS apoprotein. The assignment is according to *(31)*.

types of structure	number	residues
external β sheet	1	Thr^{4} - Thr^{8}
	2	Val^{19} - Ala^{24}
	5	Ser^{62} - Ser^{66}
internal β sheet	4	Ser^{53} - Ala^{57}
	3	Thr^{30} - Gly^{35}
	6	Gln^{94} - Ser^{98}
	7	Val^{108} - Ile^{110}
two-stranded β sheet	3a	Gln^{36} - Val^{40}
	3b	Val^{44} - Asn^{48}
two-stranded β sheet	5a	Ser^{72} - Gly^{75}
	5b	Gly^{84} -Asp^{87}
3_{10} helix		Pro^{49} - Phe^{52}

Table 3-2. Intermolecular hydrogen bonds between apo-NCS (prefixed with "A") and holo-NCS (prefixed with "B") in the asymmetric unit

donor	acceptor	distance (Å)
Ala^{A1} N	Gly^{B102} O	2.54
Asn^{A103} Nδ2	Gly^{B25} O	3.02
Gly^{A104} N	Ala^{B1} O	2.88
Ala^{B1} N	Gly^{A102} O	3.07
Ala^{B2} N	Gly^{A104} O	3.08
Gln^{B27} Nε2	Gln^{A27} Oε1	2.78
Asn^{B103} Nδ2	Leu^{A26} O	2.86

Table 3-3. Van der Waals interactions between the chromophore and the protein. The numbers of contacts within 4.0 Å are listed Residues closer than 3.5 Å are shown in bold.

residue	#	residue	#	residue	#	residue	#
Val^{B34}	1	**Gly^{B35}**	8	**Cys^{B37}**	4	Trp^{B39}	2
Gly^{B43}	1	Leu^{B45}	9	**Cys^{B47}**	8	**Pro^{B49}**	7
Phe^{B52}	8	**Phe^{B76}**	5	**Leu^{B77}**	9	**Phe^{B78}**	30
Gly^{B80}	7	Gln^{B94}	3	Val^{B95}	2	**Gly^{B96}**	10
Ser^{B98}	10	Ala^{B100}	1	Ala^{B101}	2	**Gly^{B102}**	6

Table 3-4. Solvent accessible areas before and after the chromophore binding in holo-NCS. Residues are given in the order of the difference.

residue		solvent accessible area (Å^2)		
		before	after	difference
78	Phe	199.0367	88.5460	110.4907
45	Leu	50.5610	0.0000	50.5610
102	Gly	57.2188	12.6222	44.5966
98	Ser	31.5169	0.0203	31.4966
39	Trp	129.9102	99.1307	30.7795
52	Phe	67.9540	37.2060	20.7480
49	Pro	54.6405	29.3450	25.2955
43	Gly	74.2263	54.1005	20.1258
80	Gly	57.8212	38.5422	19.2790
37	Cys	12.6572	0.0000	12.6572
101	Ala	97.5843	85.2720	12.3123
96	Gly	10.3867	0.0000	10.3867
47	Cys	9.9735	0.0001	9.9734
107	Gly	18.8036	10.4519	8.3517
105	Pro	9.3135	2.8068	6.5067
33	Asp	70.0089	53.1564	6.1213

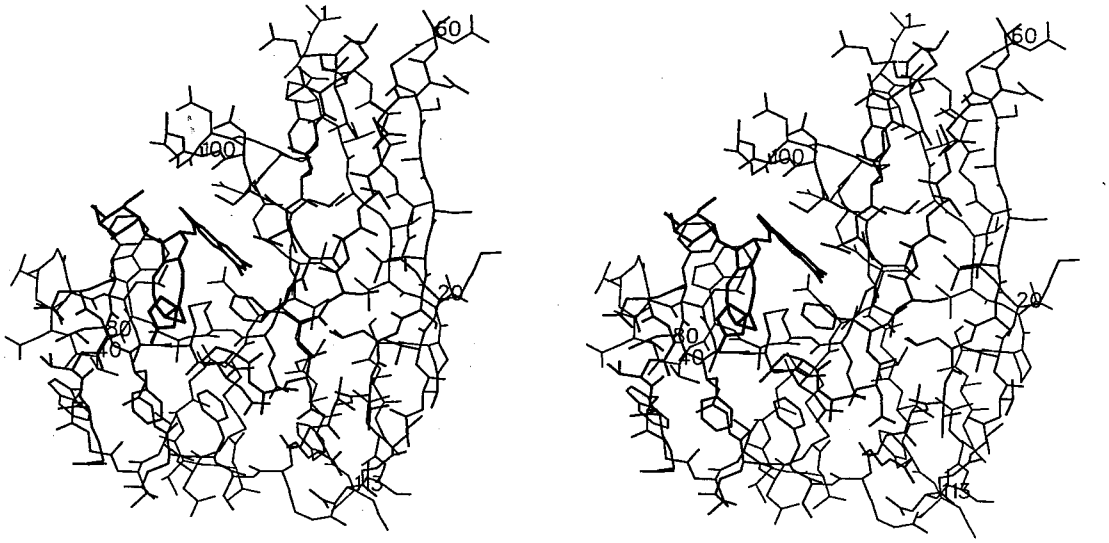
106	Glu	151.7626	147.8849	3.8777
97	Leu	3.6663	0.0045	3.6618
77	Leu	35.5267	31.9920	3.5347
76	Phe	55.0264	52.1564	2.8700
99	Asp	18.2965	16.4008	1.8957
35	Gly	1.7282	0.0000	1.7282
44	Val	45.9129	44.6101	1.3028
94	Gln	27.5571	26.1041	1.4530
100	Ala	102.2661	101.3400	0.9261
34	Val	0.4209	0.0000	0.4209
104	Gly	20.5672	20.2957	0.2715
95	Val	0.0182	0.0000	0.0182

Table 3-5. R.m.s. differences between the NCS, MCM and ACX apoproteins. The predicted r.m.s. difference of main-chain atoms, Δ_C, is given as 0.40·exp(1.87H), where H is the fraction of mutated residues *(30)*. Δ_O is the calculated r.m.s. difference.

	NCS MCM	NCS ACX	MCM ACX	NCS MCM	NCS ACX	MCM ACX
# of core residues	109	93	95	90	90	90
% of core residues(%)	96.4	82.3	84.8			
# of conserved residues	51	47	38	44	47	38
H	0.532	0.485	0.600	0.511	0.478	0.578
Δ_C	1.082	1.009	1.228	1.040	0.977	1.178
Δ_O	0.926	1.265	1.183	0.761	1.265	1.168
% error	16.8	20.2	3.8	36.8	22.7	0.9

Figure 3-1. The structure of holo-NCS

(a) All atoms are shown.



(b) A ribbons diagram, prepared with Molscript *(32)*

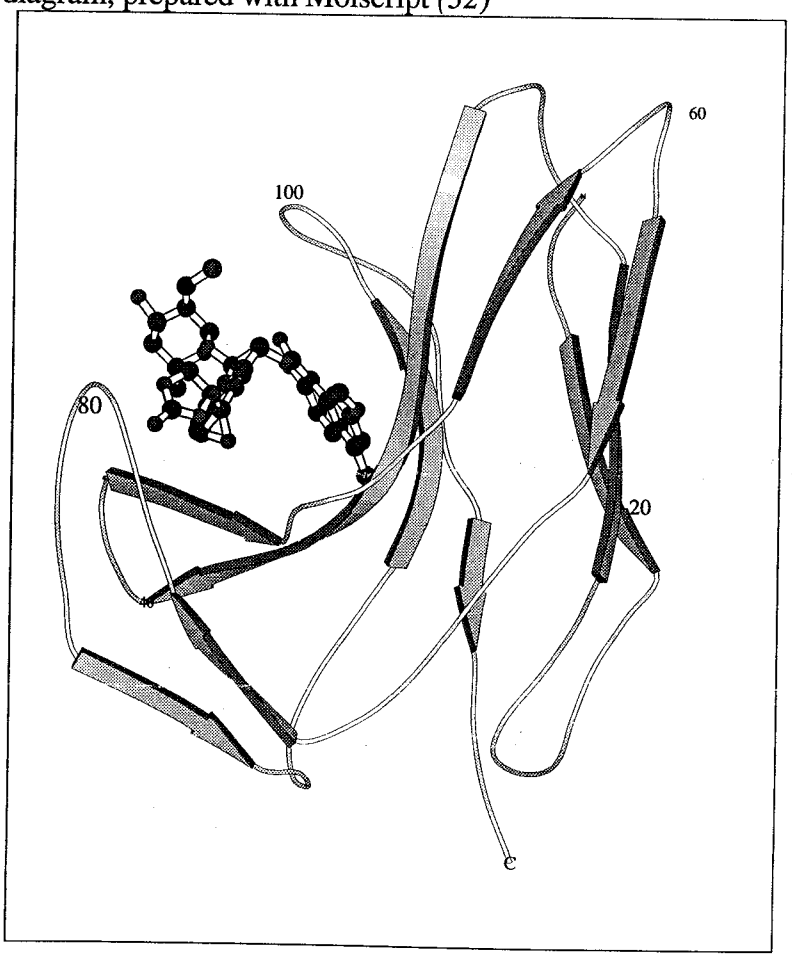
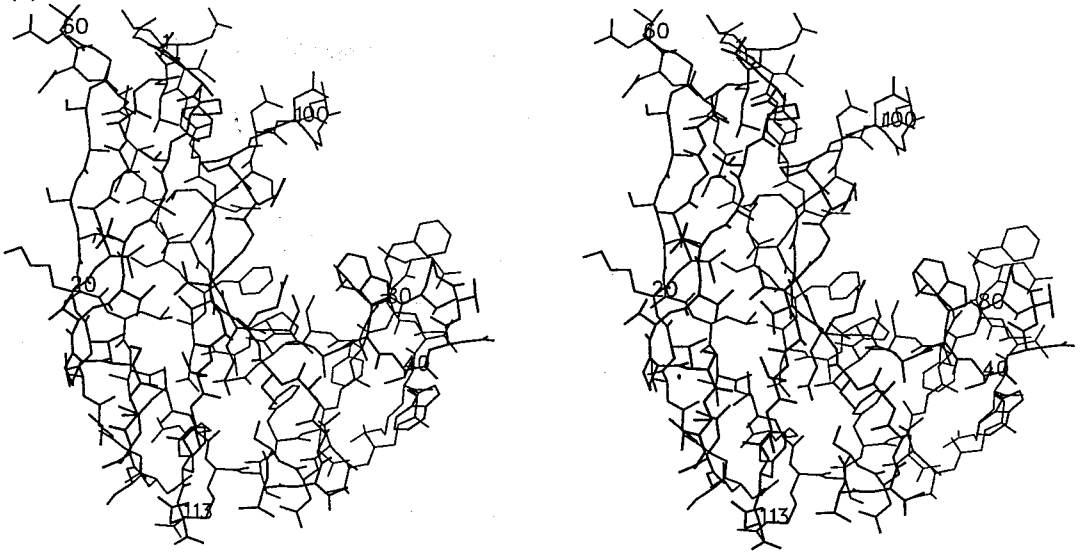


Figure 3-2. The structure of apo-NCS

(a) All atoms are shown.



(b) A ribbons diagram

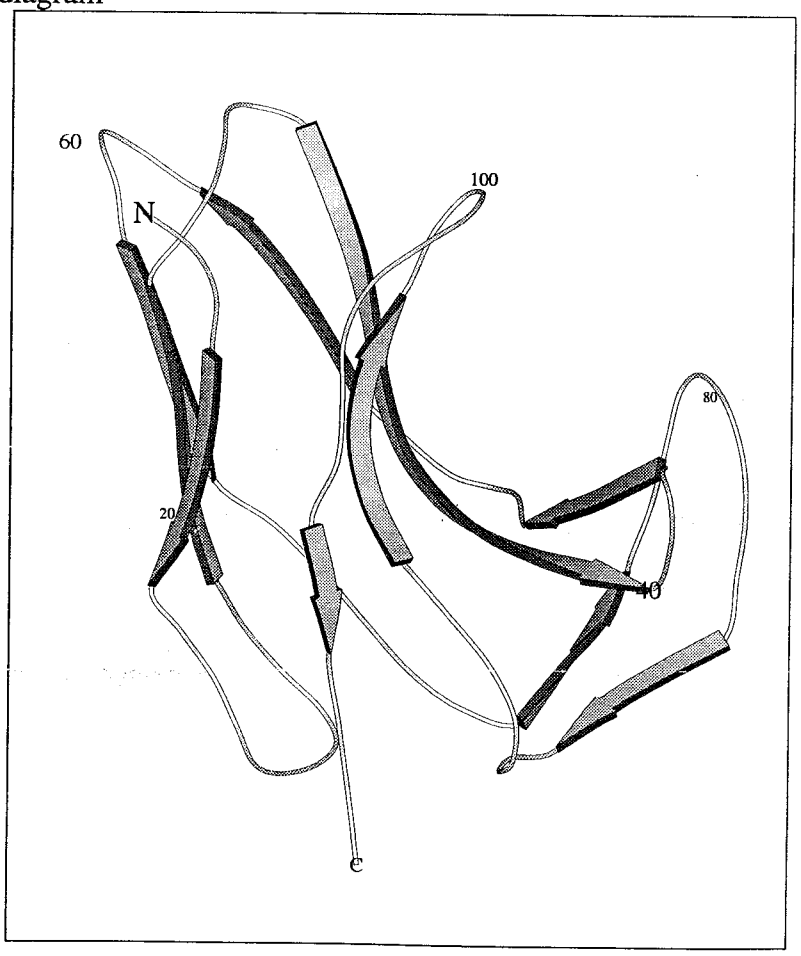
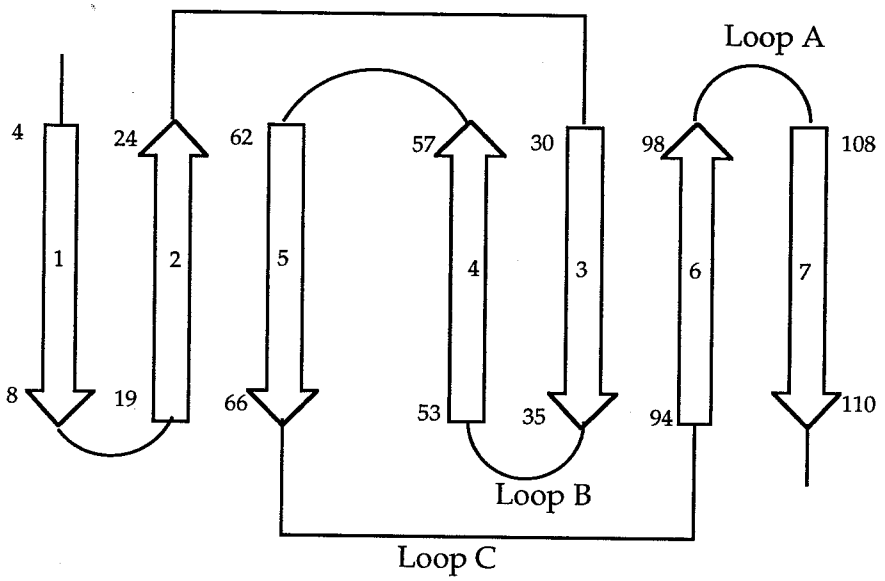


Figure 3-3. Topological diagrams of (a) NCS and (b) immunoglobulin variable domains. For NCS, the strands are defined by the indicated residue numbers. Loops A to C of NCS and CDR1 to CDR3 of the variable domain form the active site of these proteins.

(a)



(b)

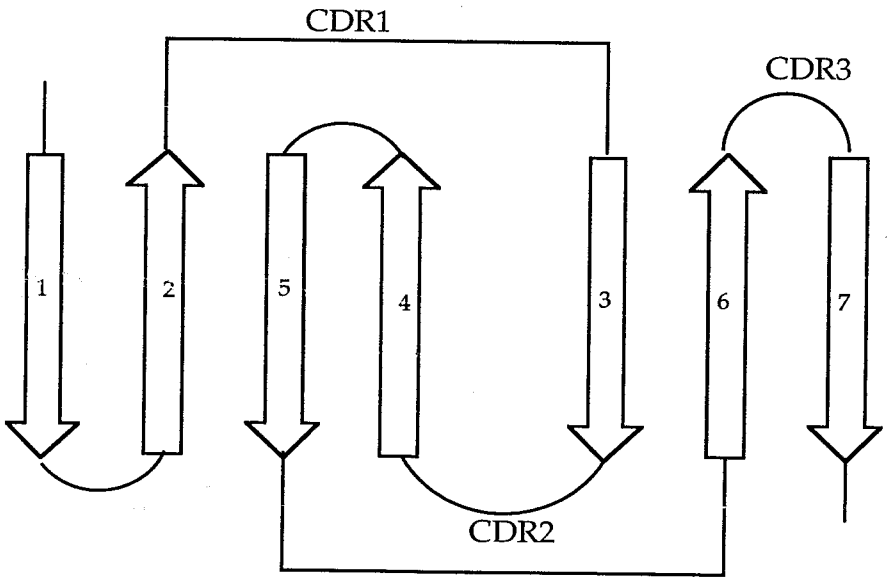


Figure 3-4. The side chains of charged residues. Asp, Glu, Lys and Arg residues of molecule B are shown.

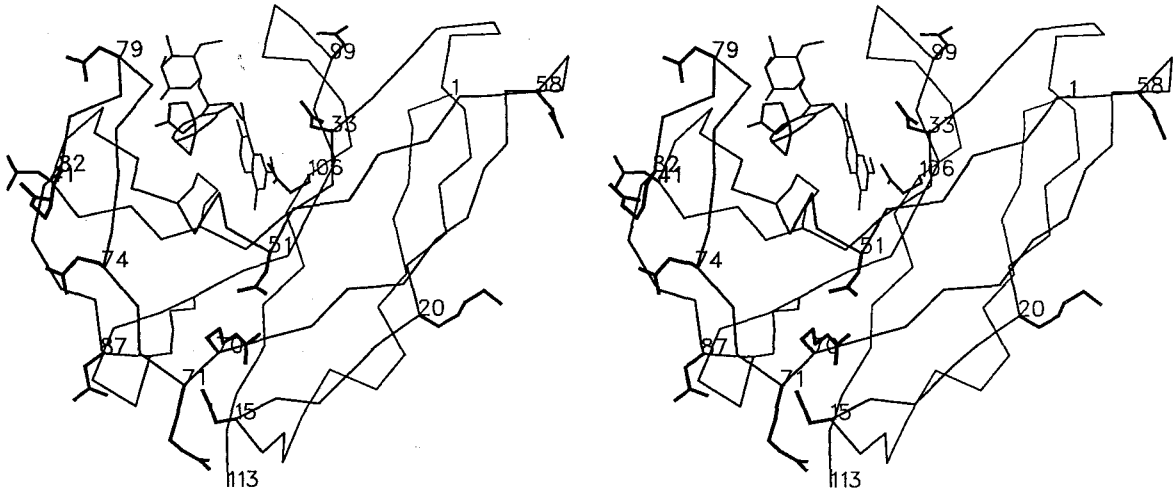


Figure 3-5. The side chains of hydrophobic residues. Phe, Tyr, Trp, Leu, Ile and Val residues of molecule B are shown.

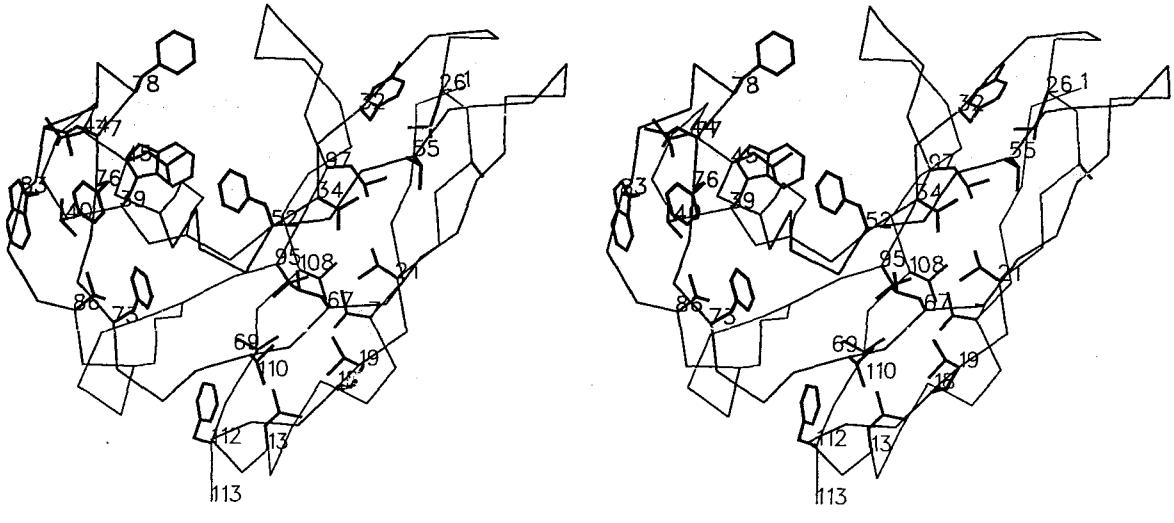


Figure 3-6. Two NCS protein molecules in the asymmetric unit. Holo-NCS contains the chromophore and apo-NCS contains an MPD in the binding pocket. The noncrystallographic twofold axis is perpendicular to the paper.

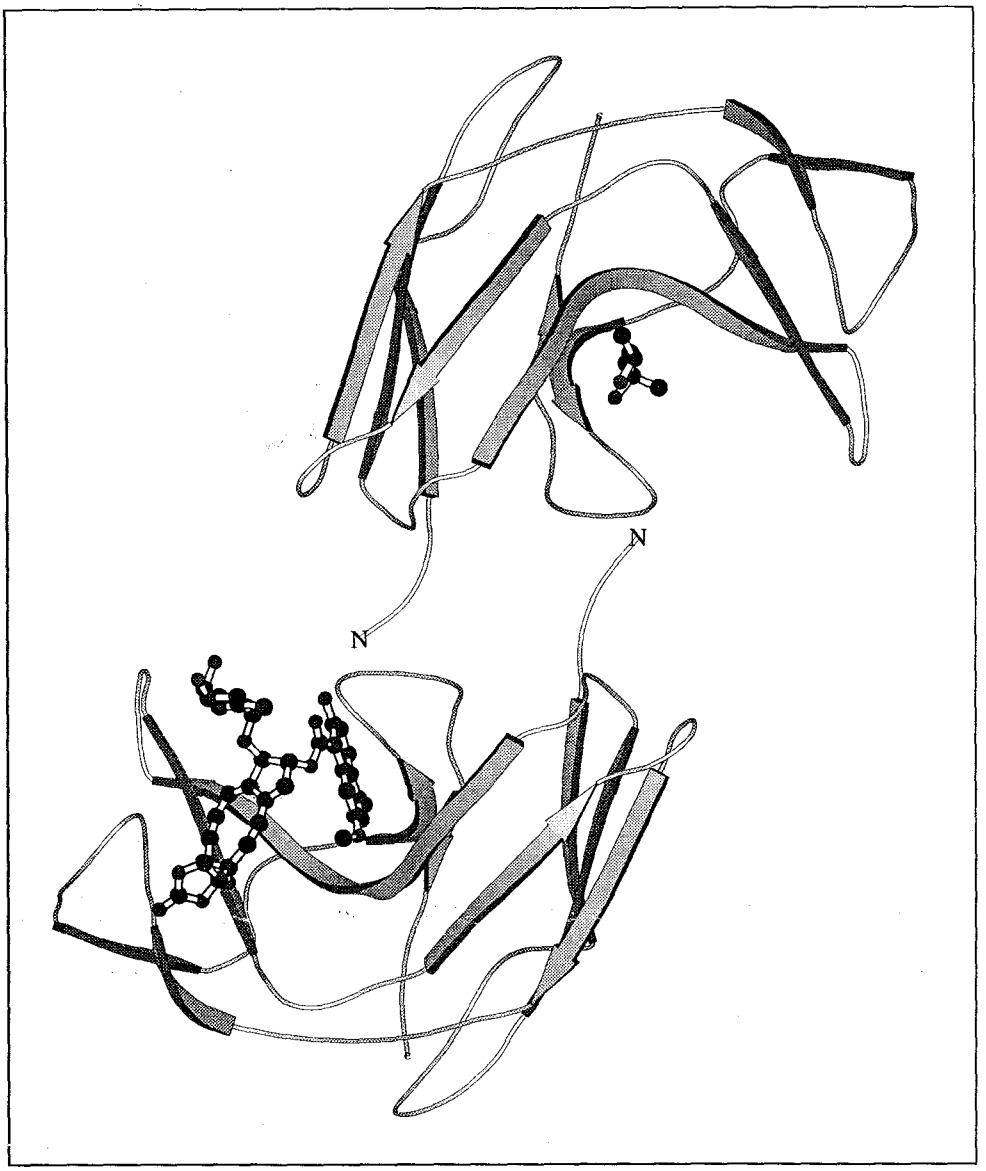
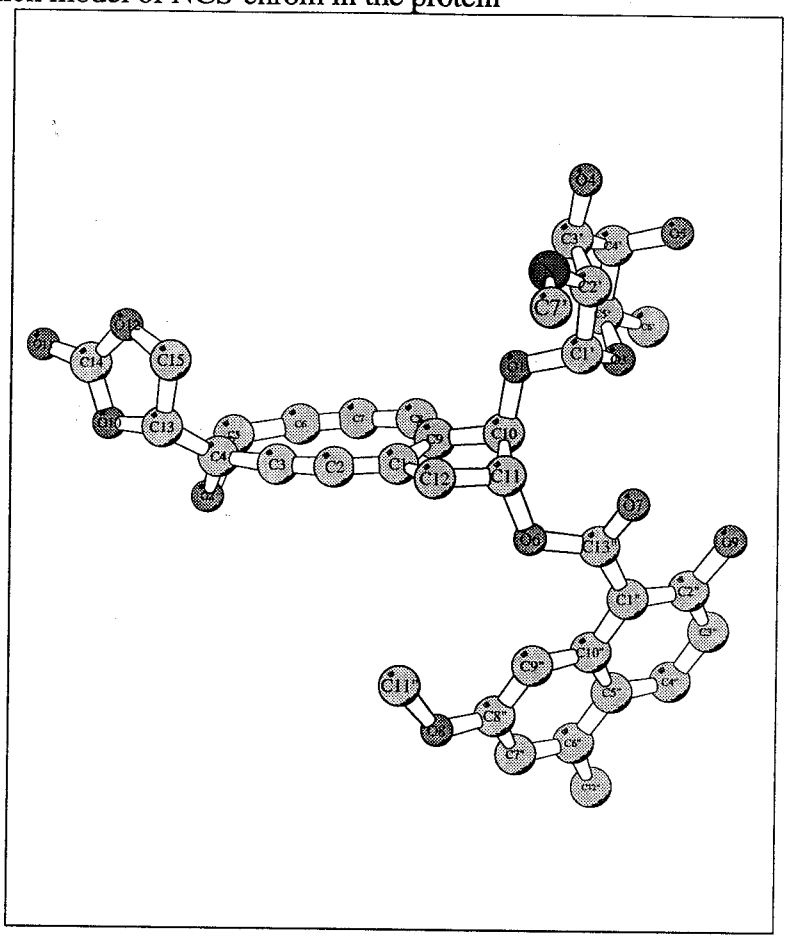


Figure 3-7. The structure of NCS-chrom

(a) Ball-and-stick model of NCS-chrom in the protein



(b) The chromophore in the protein (thick line) and in solution (thin line) *(33)*

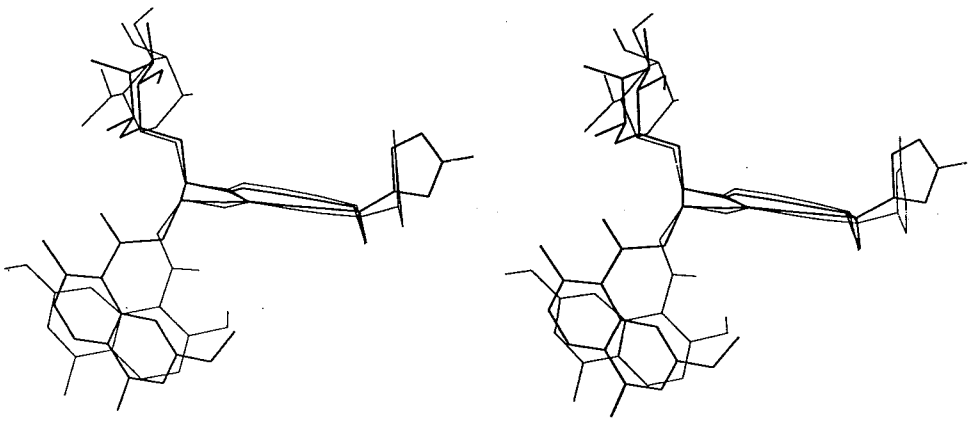
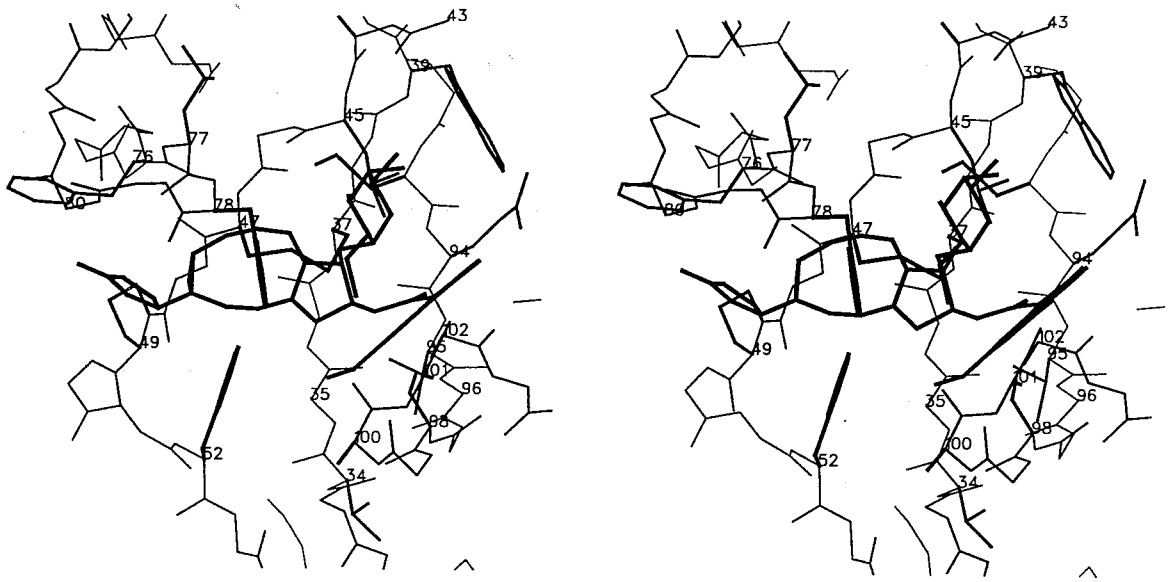
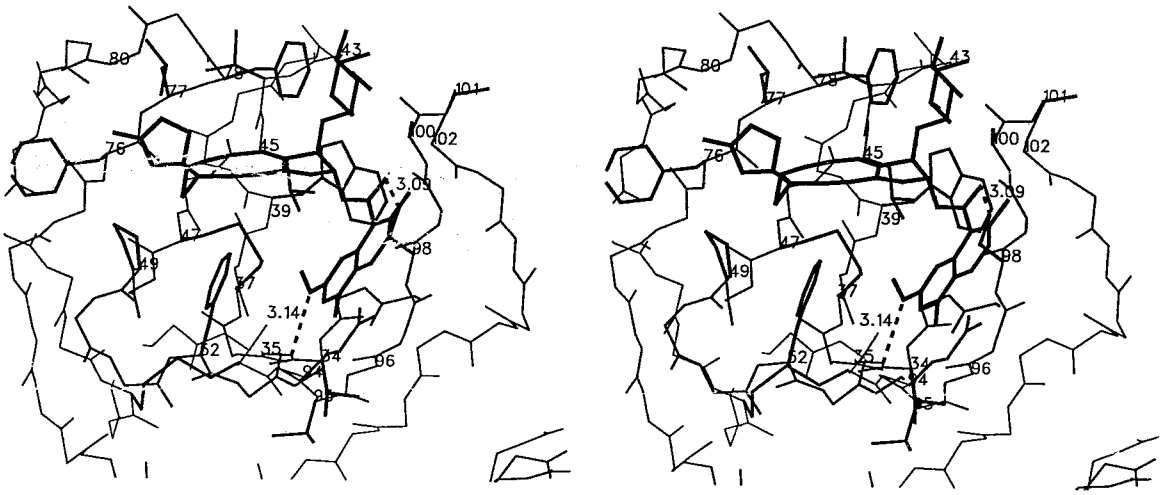


Figure 3-8. The binding of NCS-chrom in the protein. Residues close to the NCS chromophore are numbered.

(a) A top view



(b) A side view



(c) A view with the secondary structure of the protein

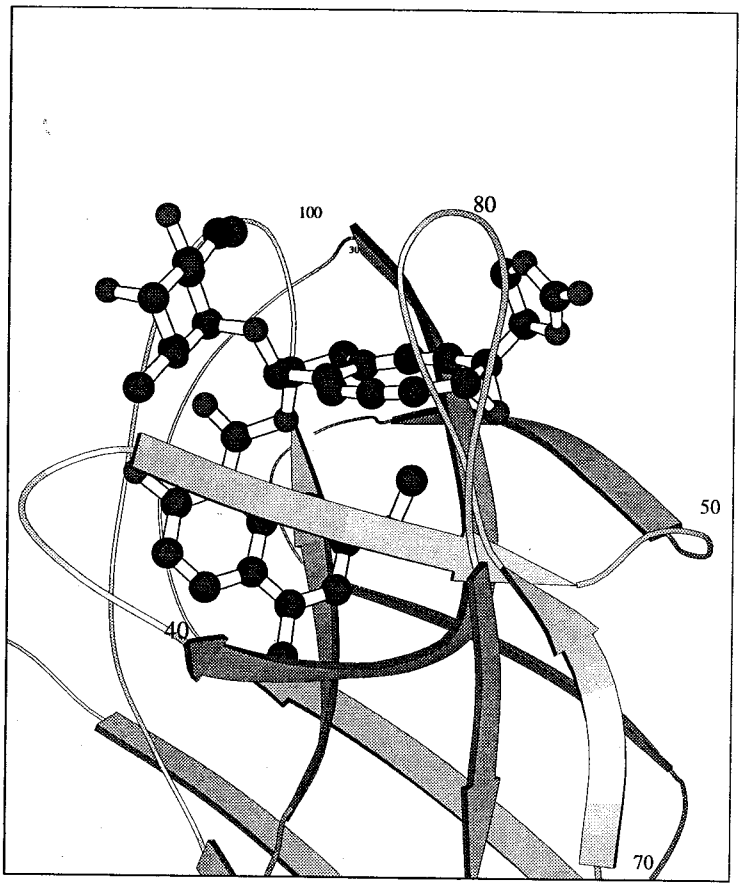


Figure 3-9. Comparison of apo-NCS (thin line) and holo-NCS (thick line). N, C and Cα atoms and the side chains of Phe^{78} and Asp^{A79} are shown.

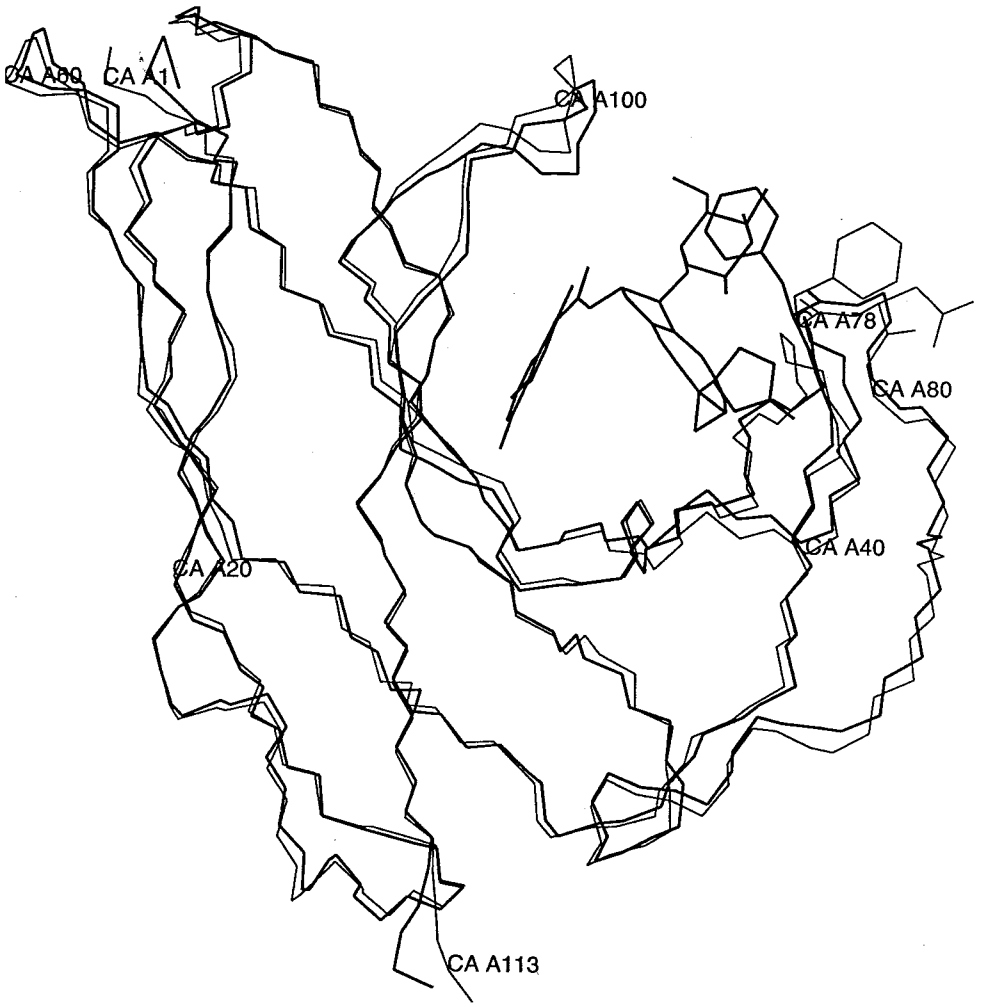


Figure 3-10. The binding pocket of apo-NCS

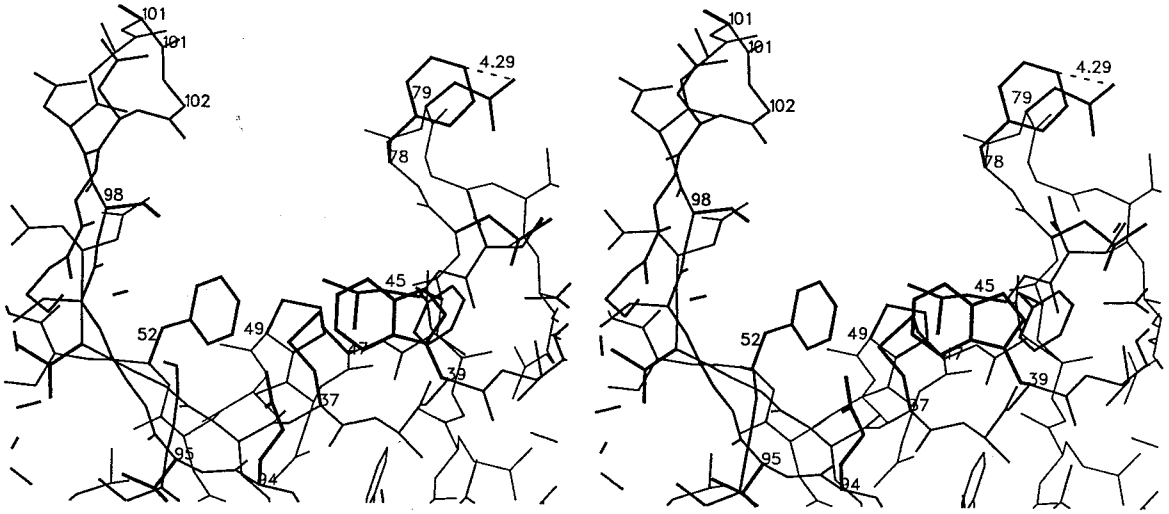


Figure 3-11. B factors averaged over the main chain atoms. Apo-NCS is in whole line and holo-NCS in dashed line.

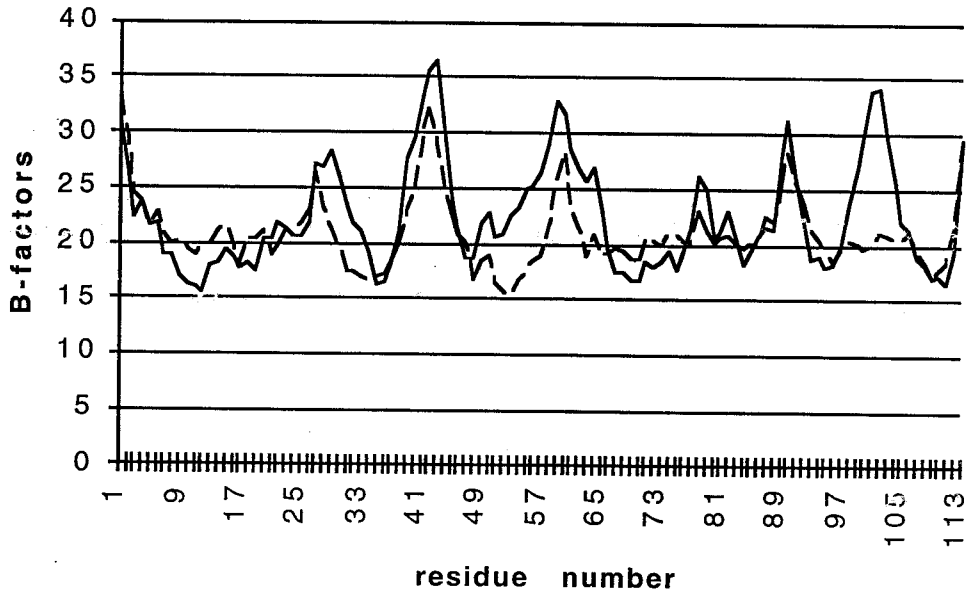


Figure 3-12. Comparison of holo-NCS (thick line) and apo-NCS from 1noa.pdb (thin line). Cα atoms and the side chains of Phe^{78} are shown.

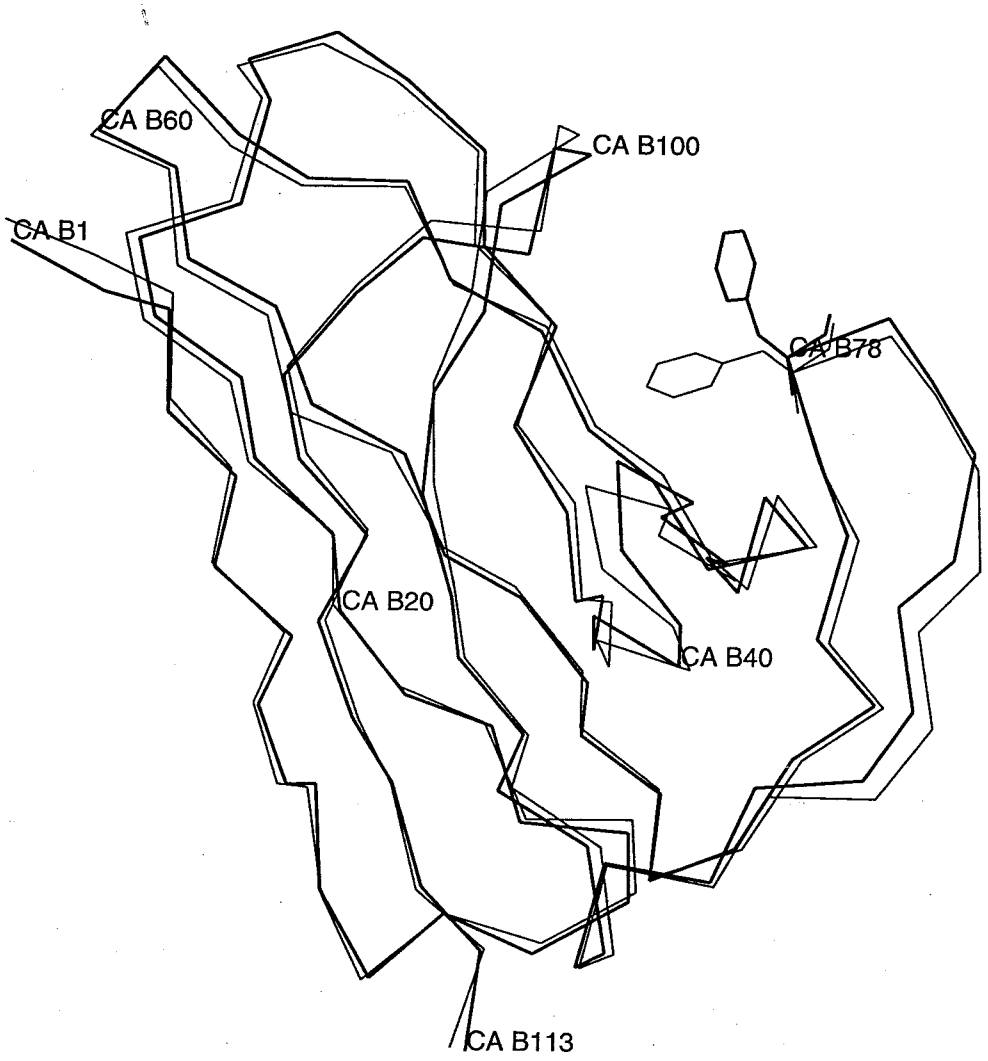
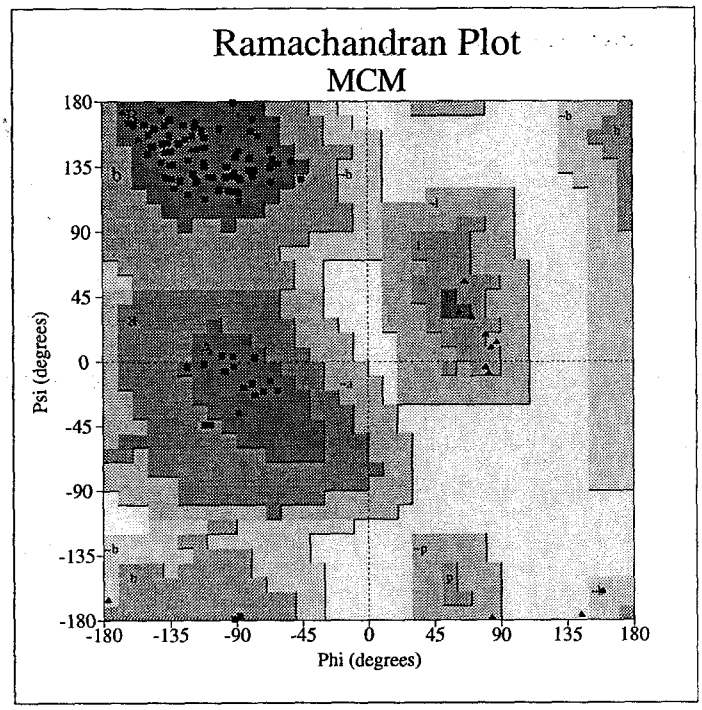


Figure 3-13. Ramachandran plots of MCM and ACX

(a) MCM



(b) ACX

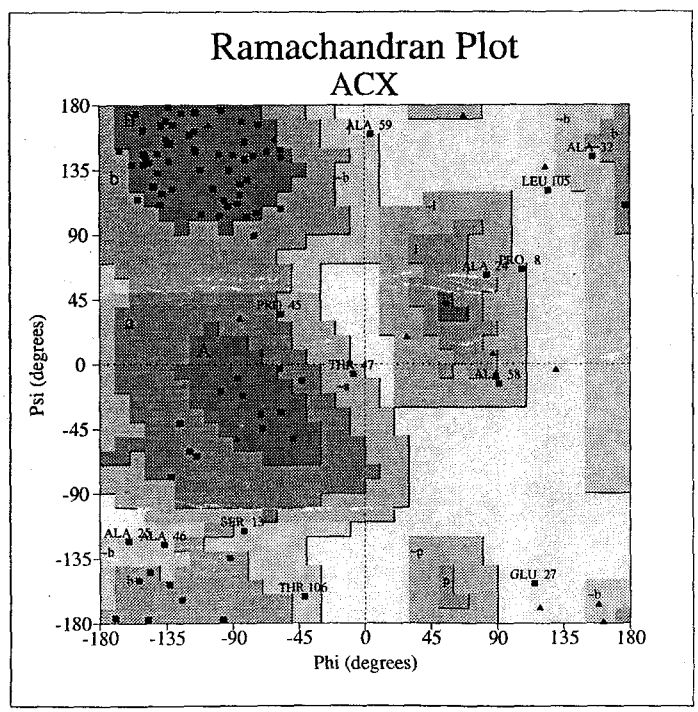


Figure 3-14. Comparison of apo-NCS (thick line), MCM (thin line), and ACX (dashed line) proteins. Cα atoms are shown.

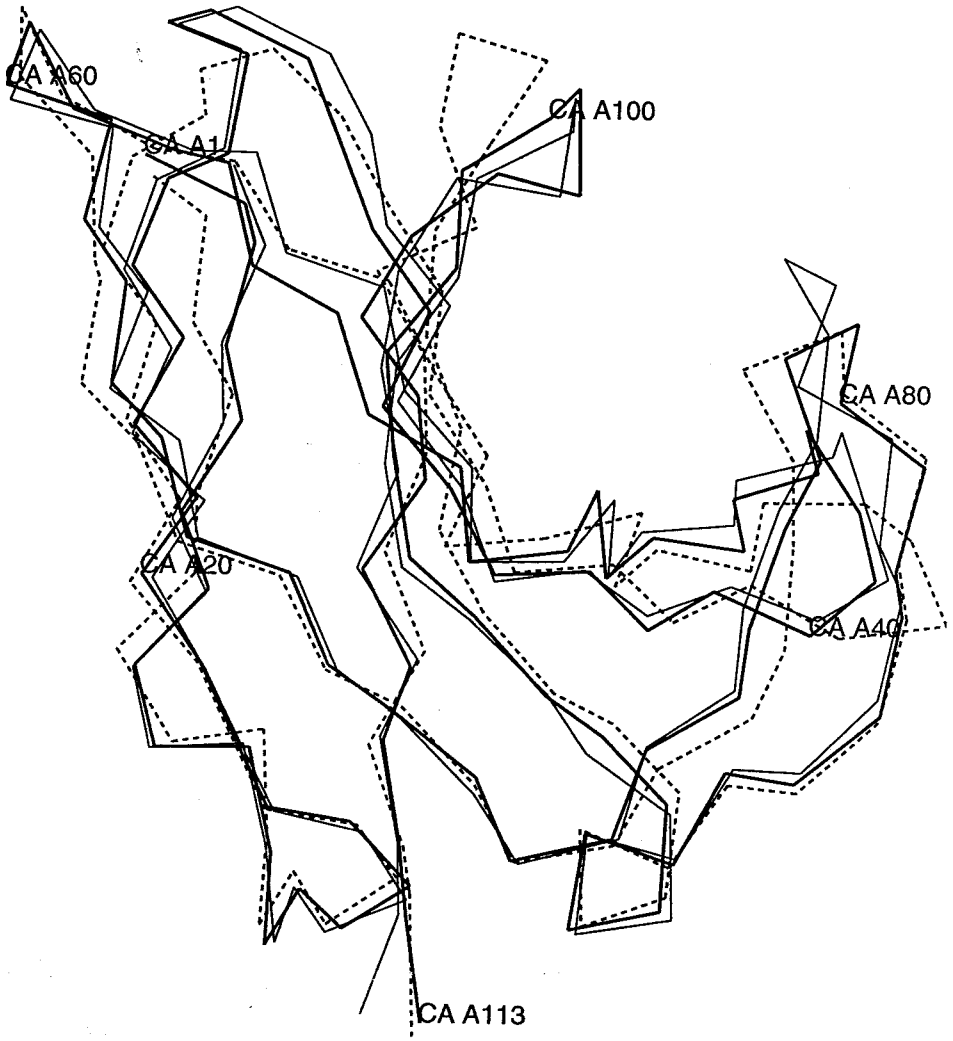
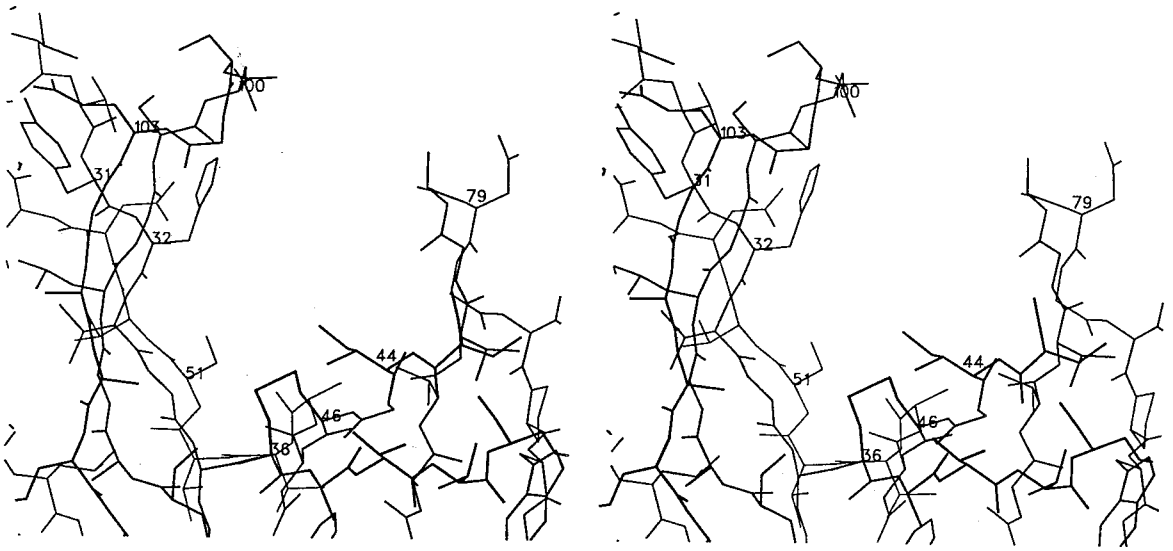
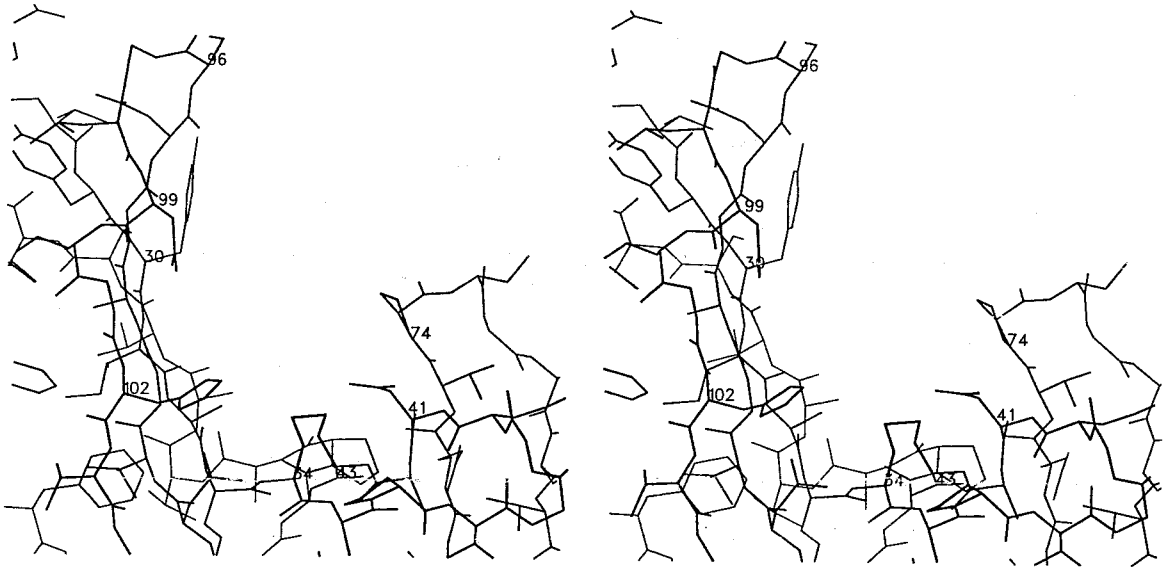


Figure 3-15. The binding pockets of MCM and ACX

(a) MCM



(b) ACX



References

(1) X. Gao, *J. Mol. Biol.*, **225**, 125 (1992)

(2) E. Adjadj, E. Quiniou, J. Mispelter, V. Favaudon, J.-M. Lhoste, *Eur. J. Biochem.*, **203**, 505 (1992)

(3) M. Lyndsay Remerowski, S. J. Glaser, L. C. Sieker, T. S. Anantha Samy, G. P. Drobny, *Biochemistry* , **29**, 8401 (1990)

(4) J. Deisenhofer, *Biochemistry*, **20**, 2361 (1981)

(5) S.-E. Ryu, P. D. Kwong, A. Truneh, T. G. Porter, J. Arthos, M. Rosenberg, X. Dai, N.-H. Xuong, R. Axel, R. W. Sweet, W. A. Hendricson, *Nature*, **348**, 419 (1990)

(6) A. Holmgren and C. I. Branden, *Nature*, **342**, 248 (1989)

(7) J. Leahy, W. A. Hendrickson, I. Aukhil, H. P. Erickson, *Science*, **258**, 987 (1992)

(8) A. M. deVos, M. Ultsch, A. A. Kossakoff, *Science*, **255**, 306 (1992)

(9) P. Bork and R.F. Doolittle, *Proc. Natl. Acad. Sci. USA* , **89**, 8990 (1992)

(10) F. Schonlau, W. Kohnlein, M. C. Garnett, *Mol. Pharmacology*, **45**, 1268 (1994)

(11) T. A. Anantha Samy, *Biochemistry*, **16**, 5573 (1977)

(12) H. Maeda, *Advanced Drug Delivery Reviews*, **6**, 181 (1991)

(13) K. Edo, K. Saito, Y. Matsuda, Y. Akiyama-Murai, M. Mizugaki, Y. Koide, N. Ishida, *Chem. Pharm. Bull.*, **39**, 170 (1991)

(14) A. G. Myers, P. J. Proteau, T. M. Handel, *J. Am. Chem. Soc.*, **110**, 7212 (1988)

(15) A. G. Myers, P. M. Harrington and B.-M. Kwon, *J. Am. Chem. Soc.*, **114**, 1086 (1992)

(16) S. K. Burley and G. A. Petsko, *Science*, **229**, 23 (1985)

(17) D. A. Dougherty and D. A. Stauffer, *Science*, **250**, 1558 (1990)

(18) J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, I. Silman, *Science*, **253**, 872 (1991)

(19) D. A. Rodham, S. Suzuki, R. D. Suenram, F. J. Lovas, S. Dasgupta, W. A. Goddard III, G. A. Blake, *Nature*, **362**, 735 (1993)

(20) M. Levitt and M. F. Perutz, *J. Mol. Biol.*, **201**, 751 (1988)

(21) B. Lee and F. M. Richards, *J. Mol. Biol.*, **33**, 515 (1992)

(22) M. Ishiguro, S. Imajo, M. Hirama, *J. Med. Chem.*, **34**, 2366 (1991)

(23) T. Tanaka, M. Hirama, M. Uceno, S. Imajo, M. Ishiguro, M. Mizugaki, K. Edo, H. Komatsu, *Tetrahedron Letters*, **32**, 3175 (1991)

(24) T. Tanaka, M. Hirama, K.-i. Fujita, S. Imajo, M. Ishiguro, J. *Chem. Soc. Chem. Commun.*, **15**, 1205 (1993)

(25) A. Teplyakov, G. Obmolova, K. Wilson, K. Kuromizu, *Eur. J. Biochem.*, **213**, 737 (1993).

(26) P. Van Roey and T. A. Beerman, *Proc. Natl. Acad. Sci. USA*, **86**, 6587 (1989)

(27) V. Z. Pletnev, A. P. Kuzin, S. D. Trakhanov, P. V. Kostetsky, *Biopolymers,* **21**, 287 (1982)

(28) R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.*, **26**, 283 (1993)

(29) K. L. Constantine, K. L. Colson, M. Wittekind, M. S. Friedrichs, N. Zein, J. Tuttle, D. R. Langley, J. E. Leet, D. R. Schroeder, K. S. Lam, B. T. Farmer II, W. J. Metzler, R. E. Bruccoleri, L. Mueller, *Biochemistry*, **33**, 11438 (1994)

(30) C. Chothia and A. M. Lesk, *EMBO J.*, **5**, 823 (1985)

(31) W. Kabsch and C. Sander, *Biopolymers*, **22**, 2577 (1983)

(32) P. J. Kraulis, *J. Appl. Cryst.*, **24**, 946 (1991)

(33) The coordinates of the NCS chromophore in solution was provided by A. G. Myers.

Chapter 4. Molecular Replacement on Neocarzinostatin

4-1. Introduction

The molecular replacement (MR) method has been widely used to solve three-dimensional structures of molecules when the structures of similar or homologous molecules are available. The objective of MR is to determine the coordinate transformation required to properly position the known molecule(s) in the unit cell of the unknown compound. In the general case, six parameters must be determined for each molecule, three rotation angles and three translations, to fully describe this transformation. As first described by Rossmann and Blow *(1)*, the MR problem is typically divided into two parts that formally involve comparison between the Patterson functions of the known and unknown structures as a function on the orientation of the known model. In the first step, the orientational (rotation) parameters are determined by calculation of the rotation function and in the second step, the translational (positional) parameters are determined by calculation of the translation function. In this way, the general six-dimensional problem is reduced to a series of two sequential three-dimensional problems.

In favorable situations, finding the correct solution to the MR problem is straightforward. In too many cases, however, the rotation function and/or the translation function is poorly contrasted and the correct solutions are not among the highest peaks. The critical dependence of the MR calculation on the parameters that define the rotation and translation functions implies that in these cases the peak height does not provide a reliable indication of the correct solution. MR trials on the structure of NCS were one of these situations.

While several studies have appeared systematically examining the dependence of the MR calculations on various search models and parameters *(2,3)*, few studies have been reported systematically analyzing cases why the MR method apparently failed. As

described in Chapter 2, we were unable to obtain MR solution from any model structures available at the time of the structural analysis. Consequently, we decided to reinvestigate the MR studies to assess the influence of various parameters on the rotation and translation function. Particular emphasis was placed on the effects of the integration radius or maximum vector length in the RF search; the high and low resolution limits of the diffraction data used; the type of model used for the known search structure (with or without conserved side chains); and the program package used to perform the MR calculations.

The crystal structures of macromomycin (MCM) *(4)* and actinoxanthin (ACX) *(5)*, whose coordinates are available from the Brookhaven Data Bank as 2mcm.pdb and 1acx.pdb, respectively, were used as search models. The sequence identity of NCS with MCM is 46.8% (52 out of 112 residues) and with ACX is 48.1% (52 out of 108 residues). The MCM structure was determined to 1.6 Å resolution with the *R* factor of 17.9%. The ACX structure was determined at 2.5 Å resolution (no final R factor for coordinate refinement was reported). The r.m.s. deviations of main chain atoms of MCM and ACX from holo-NCS are 1.03 Å and 1.93 Å, respectively. The r.m.s. deviation between apo-NCS and holo-NCS is 0.57 Å for the main chain atoms. The proteins have approximate dimensions of 20 Å x 25 Å x 40 Å and are composed mainly of β strands, that are folded into two domains with a cleft for the chromophore binding. NCS crystals grow in space group $P6_5$ with cell dimensions $a = b = 90.4$Å and $c = 52.5$Å. There were two protein molecules per asymmetric unit. The real rotation function (RF) solutions from the MCM coordinates are $\theta_1 = 169.74°$, $\theta_2 = 68.28°$, $\theta_3 = 39.90°$ for molecule A and $\theta_1 = 174.04°$, $\theta_2 = 77.53°$, $\theta_2 = 46.36°$ for molecule B in the Euler angles (Table 4-1). It is also interesting to see whether different rotation function search algorithms distinguish two close solutions as close as about 10°. The NCS crystals contained a translational symmetry determined by the native Patterson, which is a translation of 1/2 along the *z* axis.

The real translation function (TF) solutions are x= .65, y= .98, z= .0 for molecule A and x= .68, y= .0, z= .50 for molecule B, in the fractional coordinates.

Three available MR packages, X-PLOR *(6)*, MERLOT *(7)* and AMoRe *(8)*, which emplcy different algorithms for the evaluation of the rotation and translation functions, were used for the MR calculations. For comparison, two NCS models, apo-NCS and holo-NCS (with the chromophore deleted from the coordinate file) were also used in MR attempts. NCSA and NCSB represent whole protein molecules of apo-NCS and holo-NCS. Coordinate sets NCSA2 and NCSB2 are derived from these cordinates after truncation of side chains to the Cβ atom. MCM1 is MCM molecule with conserved side chains and MCM2 is the alanine chain of MCM molecule. All models were placed in a common starting orientation by superposition upon the MCM coordinates, using the algorithm of Kabsch *(9)*. For structure factor calculations, the models were placed in a P_1 cell with dimensions a= b= c= 80 Å and α= β= γ= 90°, which is about twice the longest dimension of the models *(10)*. The native data set II of the NCS structure determination was used for the observed data set. There were 4291 reflections in the resolution range of 15 - 3 Å (99.9% complete). 63% and 42% of these reflections are included in the resolution ranges of 15 - 3.5 Å and 15 - 4 Å, respectively.

4-2. The rotation function

The rotation function is defined by Rossmann and Blow *(1)* as the overlap between the Patterson function of the search model (P_{model}) and the observed Patterson function (P_{obs}) usually over a sphere centered on the Patterson origin. The rotation function will have the largest value when peaks in P_{model} are superimposed on peaks in P_{obs}. The rotation function is given as

$$R(\alpha,\beta,\gamma) = \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \sum_{\mathbf{p}} |F_{\mathbf{p}}|^2 G_{\mathbf{hp}}$$

where $F_{\mathbf{h}}$ and $F_{\mathbf{p}}$ are the observed and model structure factors and $G_{\mathbf{hp}}$ is the spherical interference function.

In the reciprocal space, Crowther expanded the Pattersons in terms of the normalized spherical harmonics and the normalized spherical Bessel functions *(11)*. The final expression for the fast rotation function R(α,β,γ) is

$$\mathrm{R}(\alpha,\beta,\gamma) = \sum_{lmm'n} a^{*}_{lmn}\, b_{lm'.n} D^{l}_{m'm}(\alpha,\beta,\gamma)$$

where a_{lmn} and $b_{lm'n}$ are expansion coefficients that depend on the intensities of the model and the crystal and $D^{l}_{m'm}$ are the matrices of the irreducible representation of the rotation group and are independent of the particular Patterson. In MERLOT, HARMCO program calculates a_{lmn} coefficients for each crystal and CROSUM evaluates R(α,β,γ) for each angle.

In AMoRe, the Patterson is expanded with spherical harmonics, but the radial variables are handled differently. The rotation function takes the form of

$$R(\alpha,\beta,\gamma) = \sum_{l=0}^{\infty} \sum_{m,m'=-l}^{l} C^{l}_{mm'}\, D^{l}_{mm'}\ (\alpha,\beta,\gamma)$$

The $C^{l}_{mm'}$ is the integration over the spherical domain of the product between the radial functions of the crystal and model. The detailed evaluation of the $C^{l}_{mm'}$ can be found in the reference *(12)*.

In X-PLOR, the real-space Patterson search method was employed for the orientation of the model. The rotation function $R(\theta_1,\theta_2,\theta_3)$ is evaluated as the sum of the product between the rotated values of P_{model} by the rotation of $(\theta_1,\theta_2,\theta_3)$ and the interpolated values of P_{obs} at the rotated position. Peaks of P_{model} are selected based on the length and the magnitude. A large number of the highest peaks after the RF search are subjected to Patterson correlation refinement. In PC refinement, the model is oriented according to the rotation angles and placed in a triclinic unit cell identical in geometry to that

of the crystal. Correlation coefficients between the squares of the normalized observed and model structure factors are calculated after rigid body refinement.

Different programs have different rotation angle conventions. In MERLOT and AMoRe, three rotation operations are defined with respect to the orthogonal system as the rotation α about the z axis, the rotation β about the new y axis and the rotation γ about the new z axis. In X-PLOR, the Euler angles are defined as the rotation θ_1 about the z axis, the rotation θ_2 about the new x axis and the rotation θ_3 about the new z axis. Though the definitions of the rotation matrix and the rotation angles are different, the matrix elements should be the same for each and all methods. The relationship between (α, β, γ) in MERLOT and AMoRe and (θ_1, θ_2, θ_3) in X-PLOR is $\theta_1 = 270^\circ - \gamma$: $\theta_2 = \beta$: $\theta_3 = 90^\circ - \alpha$. The angular distance between two rotation angles is evaluated as $\cos^{-1}\{0.5*(\mathrm{tr}(M)-1)\}$, where tr(M) is the trace of the rotation matrix, M, that relates them.

4-3. The translation function

The goal of the translation function is to find the positions of the correctly oriented models with respect to a crystallographic origin. Whereas the rotation functions in current use exploit properties of the Patterson function, the translation function may use the Patterson function (Crowther and Blow's T_0, T_1, T_2 functions *(13)*), the correlation between the observed and model structure factors, as in the *R* factor search *(14)*, or phased structure amplitudes, as in the phased translation function *(15)*. A packing function can also be used as a criterion in the TF search by evaluating the likelihood of packing arrangements of the search model and its symmetry in the crystal *(16)*. The correct solution is expected to have an optimally packed structure and subsequently a maximum of the packing function. The translation function is considered only after the determination of the RF solution and is sensitive to the accuracy of the RF solution.

In MERLOT, the TF search is done with the T_1 function and the *R* factor search. In AMoRe, the T_1 function is used as a mean of selecting reasonable peaks. After rigid

body refinement, the *R* factor and correlation coefficient between the observed and model structure factors are calculated. The TF search was carried out in X-PLOR by computing correlation coefficients between the squares of the normalized observed and model structure factors.

4-4. MERLOT

In MERLOT, the orientations of molecules are determined roughly with HARMCO and CROSUM and refined with LATSUM *(17)* and their positions are determined with TRNSUM, assisted with PAKFUN and RVAMAP. The RF search was performed over angular ranges of α= 0° - 60°, β= 0° - 180°, γ= 0° - 360° for the space group $P6_5$ *(18)*. A 2.5° step size in β was used and step sizes in α and γ were determined by the program. The result of RF searches with NCSA are shown in Table 4-4, which lists the resolution ranges, the outer integration radii, the RF solutions, the r.m.s. deviations and the distances from the real RF solutions. Two peaks are shown for each conditions; one is the correct peak and the other is the highest or the next highest peak. One consistent peak at around α= 17°, β= 70°, γ= 100° was found to be about 30° off from both of two real solutions. Options such a origin removal and the intensity cutoff did not improve the result. Similar result was obtained with NCSB. Since the rotation angles were too far off from the real ones, TF searches and trials with the MCM and ACX structure were abandoned.

The order of Bessel function in the RF search is determined based on the high resolution limit and the outer integration radius. The maximum Bessel order of the program is 30, which restricts the integration radius especially at high resolution (Table 4-2). This could be a reason why the RF routine in MERLOT couldn't find correct solutions even at high resolution. The same conditions using the same resolution ranges and the same orders of Bessel functions were tried with the program AMoRe and the RF solutions occurred among top ten peaks (data not shown). The limit on the maximum order of Bessel function in the program was not seemed to be a reason for the failure.

4-5. X-PLOR

In X-PLOR, 0.5 Å grid Patterson maps were calculated from the observed and model structure factors. Patterson peaks were selected whenever their values are greater than THREshold and their lengths fall within RANGe. THREshold used was 2.0 and RANGe was from 4 Å to the maximum vector length (r_{max}). The asymmetric unit of the rotation function is θ_1= 0° - 360°, θ_2= 0° - 180°, θ_3= 0° - 60° for the space group P65 *(19)*. Step sizes of 1.5° in θ_1, θ_2 and θ_3 were used. Five-thousand rotation angles are selected by RF values and rotation angles different by less than 6° are considered as the same (ε= 0.15). Table 4-3 lists numbers of reflections and selected Patterson peaks for given resolution range in case of MCM1. The numbers of selected Patterson peaks were doubled when r_{max} increased from 15 Å to 35 Å. When a lower value of THREshold was used, the number of selected Patterson peaks and the calculation time increased, but the result remained the same. This result implies that strong Patterson peaks contribute most in the rotation function. PC refinement was done with 60 steps of rigid body refinement. Several rotation angles converged into similar values during PC refinement. If the distance between two rotations after PC refinement is less than 5°, two rotations were considered as the same. There was no correlation between RF values and correlation coefficients after PC refinement. The TF search was performed for the top ten RF solutions within the *xy* plane, due to the origin ambiguity along *z* axis. For each rotation, the second molecule was generated by a translation of 1/2 along the *z* axis and the *R* factor was calculated after 40 steps of rigid body refinement. When there were two RF solutions, two translational parameters in the *xy* plane and their relative position along the *z* axis were determined.

NCS models were tried first (Table 4-5). Two RF solutions were found among the top ten peaks with the resolution ranges of 8.0 - 3.0 Å, 8.0 - 3.5 Å and 8.0 - 4.0 Å. The resolution range of 8.0 - 5.0 Å failed to give the correct RF solutions. Alanine chain

models gave similar results with lower correlation coefficients than whole molecules. The RF solutions were found within 2.5° from the real solutions in most cases. TF searches were followed and gave obvious peaks above one standard deviation (σ) from the next highest peak. *R* factors after rigid body refinement with two molecules were about 34% in case of whole molecules and about 44% in case of alanine chain models.

RF searches with MCM models were tried with various maximum vector lengths, resolution ranges and data sets. Table 4-6 lists angle distances from the real solutions and peak orders after PC refinement. The correct solutions occurred among the top ten peaks after PC refinement, and were within 7° of the real solutions. Correlation coefficients were independent of the model and the resolution range. However, high values of correlation coefficients didn't guarantee the correctness of a solution, which was only confirmed after the TF search. The best result was from MCM1 and 25 Å r_{max} (8 of two solutions and 1 of one solution) and the worst was from MCM2 and 15 Å r_{max} (no two solutions and 7 of one solution). MCM1 was the best in finding the RF solutions and MCM was better than MCM2. The r_{max} of 25 Å gave the best results (13 of two solutions and 11 of one solution), 35 Å was next (10 of two solutions and 13 of one solution) and 15 Å was the worst (6 of two solutions and 16 of one solution). Considering the diameter of the model is about 50 Å, the best vector length for the RF calculation was half the diameter of the model, as previously recommended in Joynson *et al. (11)*. The high resolution limit of 3.0 Å and the low resolution limit of 15.0 Å were needed for finding two RF solutions. Similar results were obtained with different data sets.

TF searches were followed (Table 4-7). In case where two RF solutions were found, two translational parameters were found without exception and relative position of two molecules was a translation of 1/2 along the *z* axis. In case where one RF solution was found, four cases out of thirteen successfully located the second molecule with the translational symmetry of the crystal. The correlation coefficient was more sensitive than

the *R* factor in the TF search. Though RF searches did not have any preference for the solutions, solution A gave translational parameters with higher correlation coefficients (about 10%) compared to those of solution B. Solution B was difficult to distinguish from wrong RF solutions with correlation coefficients in some cases (data not shown). The packing function was also given at the end of the TF search, but it was not as indicative as the *R* factor. *R* factors with one molecule were about 52% for MCM and MCM1 and about 55% for MCM2. *R* factors after rigid body refinement with two molecules were about 46% for MCM and MCM1 and about 51% for MCM2. MCM1 gave the correct solutions in most of the resolution ranges, with the RF search results of 25 Å r_{max}. The high resolution limit of 3 Å and the low resolution limit of 15 Å were best in finding MR solutions. The resolution range of 8.0 - 4.0 Å failed to give the correct MR solutions.

ACX models were also tried (data not shown). ACX molecules with conserved side chains and without side chains failed to give any correct RF solutions. ACX gave two RF solutions and translational parameters at the resolution ranges of 10.0 - 3.0 Å and 15.0 - 3.0 Å. The *R* factor after rigid body refinement with two molecules was 50%, about 4% higher than MCM. ACX also gave one RF solution and two translational parameters at the resolution ranges of 10.0 - 4.0 Å and 15.0 - 4.0 Å.

4-6. AMoRe

In AMoRe, ROTING is for the fast rotation function search and TRAING is for the translation function search. The maximum order of Bessel function is set to be 124 in the program and practically there is no limitation in the outer integration radius. The inner integration radius of 4 Å was used for the RF calculation. RF peaks that were 3σ above the average were selected. Two rotation angles different by less than 5° were considered as the same. One-molecule TF searches of the top ten RF peaks were performed to find the solution with the highest correlation coefficient. Two-molecule TF searches were performed with this MR solution and the top ten RF peaks. For each set of rotation

angles, three translation parameters were selected on the basis of the T_1 function value, followed by 40 cycles of rigid body refinement to calculate correlation coefficients and *R* factors.

Apo-NCS and holo-NCS gave two RF solutions as top peaks in the resolution ranges of 8.0 - 3.0 Å, 8.0 - 3.5 Å and 8.0 - 4.0 Å (Table 4-8). With 8.0 - 5.0 Å resolution range, two solutions are among the top five peaks. The RF solutions occurred within 5° from the real solutions. The correct RF solutions are more obvious with the higher resolution limit. Two correct peaks had r.m.s. values of 2σ above than that of the next highest peaks with 3.0 Å resolution limit, but r.m.s. values didn't mean much with 4.0 Å resolution limit. Alanine chain models gave correct solutions with the resolution ranges of 8.0 - 3.0 Å and 8.0 - 3.5 Å. TF searches were followed. The differences between correct and false TF solutions become clear after two-molecule TF searches in most cases. Correlation coefficients and *R* factors between the observed and model structure factors were given in Table 4-8 (b) and (c). Correlation coefficients were about 70% and *R* factors were about 30% with two molecules. Whole molecules gave better results than alanine chain models.

RF searches on MCM models were tried with various resolution ranges and outer integration radii. Table 4-9 lists angle distances from the real solutions and peak orders. The RF solutions were found among the top ten peaks and occurred within 8° from the real solutions in most cases. In five cases in Table 4-9 (a), the RF solutions occurred within 15° from the real solutions, which gave correct translational parameters. Solution A was found more frequently than solution B. Only solution A was found from all models with 15 Å radius. MCM1 is better in finding the RF solutions than MCM or MCM2, in the sense that MCM1 model gave 9 of two solutions and MCM2 model gave 3 of two solutions. The integration radius of 35 Å gave two solutions more often than that of 25 Å.

One-molecule and two-molecule TF searches were performed as in NCS models (Table 4-10). The correct TF solutions had about 10% higher correlation coefficients than

false solutions in one-molecule TF searches. Correlation coefficients were ranged from 30% to 40% and *R* factors were about 50%. In case of 35 Å radius, MCM1 gave the best result (7 of correct solutions) and MCM gave the worst result (3 of correct solutions) in two-molecule TF search. Five cases gave two translational parameters from one RF solution. In the case of a 25 Å radius, the correct solutions were found from one RF solution, except 6 cases and each model gave similar result after two-molecule TF searches. In the case of a 15 Å radius, the correct solutions were found from one RF solution. MCM1 and MCM2 were slightly better than MCM, in the sense that 19 of the correct MR solutions were found for MCM1 and MCM2 and 15 were found for MCM. The program was rather insensitive to the existence of side chains in models. The RF solutions of 25 Å radius (21 of correct solutions) were better than those of 35 Å radius or 15 Å radius (16 of correct solutions, each) in finding the translational parameters. It is interesting that the integration radius in the RF calculation affects the result of the TF calculation, since the accuracy of the RF solutions is relatively insensitive to different integration radii, and the integration radius is not a parameter in the TF calculation. The low resolution limit of 15 Å was needed for the calculation. The resolution limits of 4 Å or 3.5 Å found the correct solutions more often than a 3 Å resolution limit. The final correlation values ranged from 50% to 60% and the final *R* factors were about 40%. The correct MR solutions in AMoRe distribute uniformly among various resolution ranges and models.

ACX models were tried next. No correct RF solution was found from any models at various resolution ranges and the outer integration radius of 35 Å.

4-7. Conclusion

The MR approach on NCS was tried with three different programs. The interests of this study are that NCS is a small molecule, mainly composed of β strands and that the NCS crystal has two molecules in the asymmetric units that are related by a translation 1/2 along the *z* axis. The RF solutions of two molecules are different about 10°.

With X-PLOR, MCM1 was the best model, MCM was the next and MCM2 was the worst. The maximum vector length of 25 Å - about half the diameter of the molecule - was best, and 35 Å was next and 15 Å was worst in finding the correct RF solutions. The high resolution data (3 Å) and the low resolution data (~15 Å) were needed to find the correct MR solutions. The RF search did not show any preference for two solutions, but the TF search gave higher correlation values for solution A.

The RF search in AMoRe found solution A more frequently than solution B. The 25 Å radius was better than 35 Å or 15 Å radii in finding the correct MR solutions. The MR results of AMoRe were insensitive to the existence of the side chain in the model, though MCM1 worked slightly better than the other two models. The low resolution limit of 15 Å was needed for the calculation. The resolution limits of 4 Å or 3.5 Å found the correct MR solutions more often than a 3 Å resolution limit.

The correct RF solutions were found among top ten peaks with both X-PLOR and AMoRe. In both programs, the correlation coefficient or r.m.s. values after the RF search didn't guarantee the correctness of the solution. The correctness of the solution was determined after the TF search. Two RF solutions were found more frequently with X-PLOR than with AMoRe. AMoRe seems to have larger radius convergence in rigid body refinement than X-PLOR, since most of the correct MR solutions in AMoRe were found from one RF solution. If the rotation angles are about 10° off from the real solutions in X-PLOR, the TF search failed to find correct translational parameters. About a third of conditions gave the correct MR solutions in X-PLOR and about two-thirds gave the correct MR solutions in AMoRe. It was surprising AMoRe did so well in finding the MR solutions compared to MERLOT.

MERLOT failed to give the correct RF solutions with NCS models. A consistent solution was found 30° off from the real solutions. The limit on the maximum order of Bessel function in the program did not seem to be a reason for the failure. ACX was not a good model for NCS. X-PLOR found the correct solutions only in limited conditions and

AMoRe failed to find any solution with ACX models. Though ACX has similar sequence identity with NCS to that of MCM, the r.m.s. deviation from NCS is about twice greater than that of MCM from NCS and the refinement statistics of the ACX structure is bad.

MR approaches are computational experiments to determine best set of parameters that give the best results with the given data and model. Various resolution ranges and choice of parameters should be tried to determine the correctness of a solution. On the basis of on these studies, several general conclusions about the parameter values were obtained. The model with side chains of conserved residues and the maximum vector length or the integration radius of about half the diameter of the molecule gave the best result. The low resolution limit of 15 Å was needed for the RF calculation. The high resolution limit of 3.0 Å was needed in *X-PLOR*, but 4.0Å limit did better in *AMoRe*. The structure similarity and good refinement statistics of the model were important for the success of the MR method.

Table 4-1. The correct MR solutions. Solution A is for apo-NCS and solution B is for holo-NCS in the crystal. Solutions for six equivalent positions are listed.

	Solution A			Solution B		
	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
1	169.74	68.28	219.90	174.04	77.53	46.36
2	169.74	62.87	80.69	174.04	72.25	272.78
3	169.74	58.64	348.91	174.04	73.03	182.66
4	169.74	68.28	39.90	174.04	77.53	226.36
5	169.74	62.87	260.69	174.04	72.25	92.78
6	169.74	58.64	168.91	174.04	73.03	2.66

Table 4-2. The resolution limits and the orders of Bessel function.

(a) MERLOT. MERLOT has the maximum Bessel function order of 30, which limits in outer integration radius.

resolution limit	3.0	3.0	4.0	4.0	5.0	5.0
integration radius	14	17	20	23	25	29
Bessel function order	22	28	24	28	24	28

(b) AMoRe

resolution limits	3.0	3.0	4.0	4.0	5.0	5.0
integration radius	25	35	25	35	25	35
Bessel function order	44	64	32	46	26	38

Table 4-3. Numbers of selected Patterson peaks in X-PLOR. Numbers of reflections and selected Patterson peaks are given for the native data set II and the MCM1 model. A THREshold of 2.0 and RANGe of 4 Å to the given length were used as the selection criteria for the Patterson peaks.

resolution ranges	# of reflections	# of Patterson peaks, selected		
		15 Å	25 Å	35 Å
8.0-3.0	4685	520	1173	1323
10.0-3.0	4819	569	1204	1350
15.0-3.0	4921	538	1170	1364
8.0-3.5	2875	316	782	864
10.0-3.5	3009	352	770	854
15.0-3.5	3111	326	739	854
8.0-4.0	1853	252	499	555
10.0-4.0	1987	260	533	594
15.0-4.0	2087	246	501	584

Table 4-4. RF search of apo-NCS with MERLOT. Res. is the resolution range, arad. is the outer radius of integration and RMS is the r.m.s. deviation from the average values. A and B are the angle distances from solutions A and B and the small integer is the corresponding equivalent position (see Table 4-1).

res.	arad.	rf solutions	1-100	RMS	A	B
8-3	14	15.83 70.00 100.00	100.00	5.69	29.12 2	27.97 1
		39.17 97.50 145.00	82.51	4.71	137.76 3	141.42 6
8-3	17	17.50 70.00 100.00	100.00	6.16	30.79 2	26.38 1
		41.67 97.50 245.00	70.96	4.39	138.11 3	141.70 6
8-4	20	17.50 67.50 100.00	100.00	5.50	30.74 2	27.10 1
		17.50 50.00 100.00	67.83	3.72	34.56 2	36.74 1
8-4	23	17.50 67.50 100.00	100.00	6.04	30.74 2	27.10 1
		50.83 95.00 215.00	67.46	4.07	116.82 4	120.42 1
8-5	25	34.17 67.50 265.00	100.00	4.59	125.52 5	132.80 2
		15.00 72.50 100.00	89.10	4.09	28.50 2	28.30 1
8-5	29	36.67 67.50 265.00	100.00	4.47	126.15 5	133.42 2
		14.17 65.00 105.00	94.93	4.25	30.05 2	30.50 1
15-3	14	44.17 140.00 165.00	100.00	3.78	77.70 2	74.27 5
		20.83 65.00 95.00	98.88	3.73	31.88 4	26.27 1
15-3	17	44.17 140.00 165.00	100.00	3.78	77.70 2	74.27 5
		20.83 65.00 95.00	98.88	3.73	31.88 4	26.27 1
15-4	20	18.33 67.50 100.00	100.00	5.15	31.57 2	26.35 1
		51.67 147.50 175.00	79.95	4.11	83.22 2	78.29 5
15-4	23	17.50 67.50 100.00	100.00	5.27	30.74 2	27.10 1
		40.00 105.00 270.00	66.34	3.49	151.65 3	152.31 6
15-5	25	15.83 80.00 105.00	100.00	4.03	31.84 2	27.60 1
		24.17 135.00 140.00	82.24	3.31	62.99 2	59.49 5
15-5	29	35.83 135.00 160.00	100.00	4.10	72.58 2	70.52 5
		24.17 62.50 95.00	79.88	3.28	29.09 4	24.83 1

Table 4-5. X-PLOR Result of NCS models.

(a) RF search. RANGe is 4 Å to 35 Å. Angle distances from the real solutions and peak orders after PC refinement are given.

	NCSA		NCSA2		NCSB		NCSB2	
	A	B	A	B	A	B	A	B
8.0-3.0	1.12 1	0.00 2	0.00 1	0.00 2	0.00 2	1.13 1	0.00 2	0.00 1
8.0-3.5	1.89 1	1.14 2	1.04 1	0.00 2	1.64 2	1.16 1	1.06 2	1.18 1
8.0-4.0	2.18 1	X	1.43 1	2.08 10	2.44 2	2.44 1	0.93 1	2.09 4
8.0-5.0	X	2.47 4	X	X	X	6.20 1	X	X

(b) One-molecule TF search. Correlation coefficients are given.

	NCSA		NCSA2		NCSB		NCSB2	
	A	B	A	B	A	B	A	B
8.0-3.0	.3680	.3079	.2973	.2388	.2842	.3549	.2218	.2735
8.0-3.5	.3729	.3113	.3039	.2402	.3210	.3273	.2378	.2573
8.0-4.0	.3689	X	.2402	.3092	.3208	.3111	.2710	.2480
8.0-5.0	X	.3009	X	X	X	.2976	X	X

(c) Two-molecule TF search. *R* factors are given.

	NCSA	NCSA2	NCSB	NCSB2
8.0-3.0	.3460	.4453	.3459	.4454
8.0-3.5	.3365	.4425	.3431	.4451
8.0-4.0	X	.4270	.3386	.4354
8.0-5.0	X	X	X	X

Table 4-6. X-PLOR RF search of MCM models. Angle distances from the real solutions and peak orders after PC refinement are given.

(a) RANGe is 4 Å to 35 Å.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	0.00 1	4.32 5	1.75 2	4.06 1	2.88 1	4.23 2
10.0-3.0	0.00 4	4.32 1	0.00 2	4.40 1	X	X
15.0-3.0	X	4.32 1	4.27 2	4.42 1	2.43 1	3.56 2
8.0-3.5	X	2.77 6	X	1.72 2	X	X
10.0-3.5	3.41 5	X	X	1.77 2	X	1.82 4
15.0-3.5	3.39 3	2.44 4	2.50 1	1.96 2	X	1.60 3
8.0-4.0	X	X	5.28 5	X	X	X
10.0-4.0	6.86 3	X	5.10 1	X	5.57 2	X
15.0-4.0	6.93 3	X	5.72 1	1.96 6	5.15 1	X

(b) RANGe is 4 Å to 25 Å.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	0.00 4	X	1.67 2	4.07 1	2.88 1	X
10.0-3.0	1.28 2	4.38 1	0.00 1	4.22 2	2.75 1	2.83 2
15.0-3.0	X	6.30 1	0.00 2	4.16 1	2.63 1	3.16 2
8.0-3.5	X	2.77 4	1.64 1	1.72 2	X	X
10.0-3.5	X	X	1.73 1	1.77 3	X	1.79 4
15.0-3.5	3.49 3	2.42 4	2.15 1	1.79 2	2.54 1	1.63 3
8.0-4.0	X	X	5.14 5	X	X	X
10.0-4.0	6.81 3	X	5.04 1	3.70 6	5.00 2	X
15.0-4.0	7.07 4	X	5.73 1	3.47 6	5.37 2	X

(c) RANGe is 4 Å to 15 Å.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	0.00 3	4.33 1	1.81 1	4.63 3	2.88 1	X
10.0-3.0	X	4.38 1	X	4.21 1	X	X
15.0-3.0	1.62 2	4.34 1	X	4.18 1	X	4.23 1
8.0-3.5	X	2.71 4	X	1.71 1	X	2.04 8
10.0-3.5	X	X	X	1.74 2	X	1.81 2
15.0-3.5	1.58 3	2.50 4	2.12 1	1.78 2	X	1.60 1
8.0-4.0	X	X	X	X	X	X
10.0-4.0	6.39 3	X	X	1.37 2	4.16 1	X
15.0-4.0	7.15 4	X	5.79 1	2.29 4	4.07 1	X

(d) RANGe is 4 Å to 35 Å. Native data I was used for the calculation.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	1.24 7	4.00 2	1.18 4	3.82 1	3.04 2	X
10.0-3.0	1.92 7	4.12 4	1.87 2	3.93 1	3.16 1	X
15.0-3.0	X	4.05 1	2.20 3	3.948 1	3.21 1	X
8.0-3.5	X	X	X	1.16 5	X	X
10.0-3.5	3.07 10	X	2.74 2	1.31 1	X	0.91 8
15.0-3.5	11.93 2	2.31 4	3.06 2	1.31 1	3.57 1	0.85 2
8.0-4.0	X	X	5.82 10	X	X	X
10.0-4.0	7.72 5	X	6.00 4	X	3.94 10	X
15.0-4.0	7.65 4	X	6.62 1	X	4.01 5	X

Table 4-7. X-PLOR TF search of MCM models.

(a) One-molecule TF search. Correlation coefficients are given for the RF solutions of Table 4-6 (a).

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	.2256	.1253	.2454	.1310	.2004	.1084
10.0-3.0	.2374	.1360	.2680	.1460	X	X
15.0-3.0	X	.1402	.2561	.1454	.2218	.1424
8.0-3.5	X	.1556	X	.1710	X	X
10.0-3.5	.2438	X	X	.1962	X	.1784
15.0-3.5	.2407	.1687	.2922	.1932	X	.1810
8.0-4.0	X	X	.2088	X	X	X
10.0-4.0	.1749	X	.2551	X	.2036	X
15.0-4.0	.1776	X	.2441	.2158	.2192	X

(b) Two-molecule TF search. *R* factors are given for the RF solutions of Table 4-6 (a). The solution gave the second molecule by the translational symmetry of the crystal in (*).

	MCM	MCM1	MCM2
8.0-3.0	.4577	.4560	.4917
10.0-3.0	.4613	.4622	X
15.0-3.0	*.4771	.4687	.5108
8.0-3.5	X	X	X
10.0-3.5	*.4800	X	X
15.0-3.5	.4615	.4545	*.5160
8.0-4.0	X	X	X
10.0-4.0	X	X	X
15.0-4.0	X	.4346	*.4831

(c) Two-molecule TF search. *R* factors are given for the RF solutions of Table 4-6 (b). The solution gave the second molecule by the translational symmetry of the crystal in (*).

	MCM	MCM1	MCM2
8.0-3.0	.X	.4559	X
10.0-3.0	.4614	.4624	.5011
15.0-3.0	*.4686	.4681	.5108
8.0-3.5	X	.4419	X
10.0-3.5	X	.4488	X
15.0-3.5	.4547	.4544	.5039
8.0-4.0	X	X	X
10.0-4.0	X	.4289	X
15.0-4.0	X	.4347	*.4848

(d) Two-molecule TF search. *R* factors are given for the RF solutions of Table 4-6 (c). The solution gave the second molecule by the translational symmetry of the crystal in (*).

	MCM	MCM1	MCM2
8.0-3.0	.4575	.4563	*.4916
10.0-3.0	X	X	X
15.0-3.0	.4676	X	*.5124
8.0-3.5	X	X	X
10.0-3.5	X	X	X
15.0-3.5		.4545	*.5181
8.0-4.0	X	X	X
10.0-4.0	X	X	*.4831
15.0-4.0	X	.4344	*.4844

Table 4-8. AMoRe Result of NCS models.

(a) RF search. The integration radii of 4 Å to 35 Å were used. Angle distances from the real solutions and peak orders are given. Solution A is (50.3, 68.3, 100.1) and solution B is (43.1, 77.6, 96.3) in the Euler angles.

	NCSA				NCSA2				NCSB				NCSB2			
	A		B		A		B		A		B		A		B	
8.0-3.0	2.98	1	0.00	2	2.47	1	2.39	2	2.21	2	2.55	1	2.80	2	0.92	1
8.0-3.5	2.73	1	0.87	2	3.68	1	2.31	3	2.29	1	2.15	1	3.20	3	1.57	1
8.0-4.0	1.87	1	1.00	2	5.35	1		X	2.08	2	1.59	1	2.33	5		X
8.0-5.0	4.33	3	2.90	5	4.29	9		X	4.79	4	1.30	1		X		X

(b) One-molecule TF search. Correlation coefficients are given.

	NCSA		NCSA2		NCSB		NCSB2	
	A	B	A	B	A	B	A	B
8.0-3.0	51.28	45.89	46.62	42.37	44.25	49.97	41.07	44.79
8.0-3.5	44.05	38.43	42.75	38.83	38.01	41.80	37.03	39.84
8.0-4.0	39.01	35.03	36.06	X	33.40	37.34	31.49	X
8.0-5.0	33.88	33.52	30.79	X	31.16	31.83	X	X

(c) Two-molecule TF search. *R* factors and correlation coefficients are given.

	NCSA		NCSA2		NCSB		NCSB2	
8.0-3.0	31.65	77.45	40.02	65.75	32.06	76.09	40.45	64.57
8.0-3.5	30.48	75.76	37.64	66.02	31.36	73.90	38.57	64.20
8.0-4.0	30.58	73.64	X		31.35	71.16	*38.71	60.70
8.0-5.0	34.52	66.05	X		35.73	62.44	X	

Table 4-9. AMoRe RF search of MCM models. Angle distances from the real solutions and peak orders are given.

(a) The integration radii of 4 Å to 35 Å were used.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	3.96 4	2.43 5	3.44 2	1.00 5	1.40 1	X
10.0-3.0	4.60 1	X	1.94 1	3.34 3	0.00 1	3.26 5
15.0-3.0	6.79 1	X	10.22 1	2.06 4	11.59 1	2.42 8
8.0-3.5	X	X	2.79 3	1.12 8	2.20 3	X
10.0-3.5	6.40 3	X	4.62 1	2.32 9	1.79 1	3.13 6
15.0 3.5	8.67 1	X	12.34 1	2.71 8	10.23 1	X
8.0-4.0	X	X	5.44 5	X	3.64 3	X
10.0-4.0	5.10 1	X	6.89 1	X	4.94 1	X
15.0 4.0	7.30 1	X	10.35 1	X	8.75 2	X

(b) The integration radii of 4 Å to 25 Å were used.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	2.41 4	X	3.41 2	1.44 9	0.00 1	X
10.0-3.0	2.13 1	X	4.38 1	X	1.51 1	X
15.0-3.0	3.62 1	2.60 4	1.32 1	2.08 2	4.07 1	X
8.0-3.5	3.07 7	X	1.48 2	X	1.67 2	X
10.0-3.5	5.10 2	X	2.97 1	X	2.42 1	X
15.0-3.5	5.65 1	1.72 5	4.21 1	2.22 3	6.92 1	X
8.0-4.0	5.30 8	X	6.44 4	X	2.75 8	X
10.0-4.0	5.27 1	X	6.47 1	X	3.41 1	X
15.0-4.0	6.34 3	3.34 10	8.26 1	X	6.67 1	X

(c) The integration radii of 4 Å to 15 Å were used.

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	4.08 1	X	4.65 1	X	1.42 1	X
10.0-3.0	2.44 1	X	3.85 1	X	4.38 1	X
15.0-3.0	2.18 1	X	5.39 1	X	2.21 1	X
8.0-3.5	5.28 4	X	2.68 2	X	2.13 1	X
10.0-3.5	3.69 1	X	4.70 1	X	1.45 1	X
15.0-3.5	5.97 1	X	3.19 1	X	4.51 1	X
8.0-4.0	3.93 5	X	1.25 2	X	3.08 1	X
10.0-4.0	3.55 1	X	0.90 1	X	2.36 1	X
15.0-4.0	5.14 2	X	3.93 1	X	4.62 1	X

Table 4-10. AMoRe TF search of MCM models

(a) One-molecule TF search. Correlation coefficients are given for the RF solutions of Table 4-9 (a).

	MCM		MCM1		MCM2	
	A	B	A	B	A	B
8.0-3.0	33.98	28.82	37.41	31.07	35.33	X
10.0-3.0	36.37	X	40.07	34.60	39.46	36.08
15.0-3.0	39.30	X	42.86	38.31	43.92	41.09
8.0-3.5	X	X	33.63	26.55	33.37	X
10.0-3.5	32.22	X	37.25	31.45	39.36	35.60
15.0-3.5	36.61	X	41.17	36.49	45.29	X
8.0-3.0	X	X	29.19	X	27.95	X
10.0-4.0	29.15	X	34.10	X	35.99	X
15.0-4.0	34.32	X	38.85	X	34.51	43.55

(b) Two-molecule TF search. *R* factors and correlation coefficients are given for the RF solutions of Table 4-9 (a). Two molecules were found from one RF solution in (*).

	MCM		MCM1		MCM2	
8.0-3.0	44.59	48.78	43.86	51.98	X	
10.0-3.0	X		43.72	55.18	46.79	52.12
15.0-3.0	X		43.47	58.85	46.30	57.74
8.0-3.5	X		41.37	53.15	X	
10.0-3.5	X		41.17	57.36	43.58	55.68
15.0-3.5	*42.13	57.40	40.68	62.02	*42.73	62.71
8.0-4.0	X		X		X	
10.0-4.0	X		X		*42.15	55.39
15.0-4.0	*40.75	57.98	*39.27	62.55	41.03	63.80

(c) Two-molecule TF search. *R* factors and correlation coefficients are given for the RF solutions of Table 4-9 (b). Two molecules were found from one RF solution in (*).

	MCM		MCM1		MCM2	
8.0-3.0	X		43.87	51.99	X	
10.0-3.0	X		X		*47.74	50.58
15.0-3.0	44.41	55.15	43.52	58.88	*46.34	57.55
8.0-3.5	*42.33	49.57	X		*46.59	44.77
10.0-3.5	*42.04	52.80	*41.26	57.40	*43.57	55.69
15.0-3.5	41.85	57.55	40.70	62.05	*42.73	62.73
8.0-4.0	*41.42	48.32	*40.21	51.67	X	
10.0-4.0	*41.08	52.30	*40.08	56.12	*42.15	55.46
15.0-4.0	40.75	57.98	*39.35	62.51	*41.31	63.65

(d) Two-molecule TF search. *R* factors and correlation coefficients are given for the RF solutions of Table 4-9 (c). Two molecules were found from one RF solution in (*).

	MCM		MCM1		MCM2	
8.0-3.0	X		X		X	
10.0-3.0	X		X		*46.82	52.06
15.0-3.0	X		*43.55	58.85	*46.34	57.56
8.0-3.5	X		X		*44.21	49.05
10.0-3.5	*42.04	52.81	X		*43.58	55.72
15.0-3.5	*41.84	57.56	*40.70	62.05	*42.71	62.77
8.0-4.0	*41.42	48.43	*40.22	51.66	X	
10.0-4.0	*41.07	52.32	*39.98	56.99	X	
15.0-4.0	*40.76	57.98	*39.26	62.55	*41.02	63.82

References

(1) M. G. Rossmann and D. M. Blow, *Acta Cryst.*, **15**, 24 (1962)

(2) M. Cygler and W. F. Anderson, *Acta Cryst.*, **A44**, 38 (1988)

(3) C. F. Aguilar, M. P. Newman, J. Sanz Aparicio, J. B. Cooper, I. J. Tickle, T. L. Blundell, *Acta Cryst.*, **A49**, 306 (1993)

(4) P. Van Roey and T. A. Beerman, *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 6587 (1989)

(5) V. Z. Pletnev, A. P. Kuzin, S. D. Trakhanov, P. V. Kostetsky, *Biopolymers* , **21**, 287 (1982)

(6) A. T. Brunger, *Acta Cryst.*, **A46**, 46 (1990)

(7) P. M. D. Fitzgerald, *J. Appl. Cryst.*, **21**, 273 (1988)

(8) J. Navaza, *Acta Cryst.*, **A50**, 157 (1994)

(9) W. Kabsch, *Acta Cryst.*, **A32**, 922 (1976)

(10) M. A. Joynson, A. C. T. North, V. R. Sarma, R. E. Dickerson, L. K. Steinrauf, *J. Mol. Biol.*, **50**, 137 (1970)

(11) R. A. Crowther, in *The molecular replacement method*', edited by M. G. Rossmann, p173, New York: Gordon and Breach (1972)

(12) J. Navaza, *Acta Cryst.*, **A43**, 645 (1987)

(13) R. A. Crowther and B. M. Blow, *Acta Cryst.*, **A23**, 544 (1967)

(14) C. A. Taylor and K. A Morley, *Acta Cryst.*, **12**, 101 (1959)

(15) P. M. Colman, H. Fehlhammer, K. Bartels, In *Crystallographic Computing technique*, edited by F. R. Ahmed, K. Huml and B. Sedlace, p248, Copenhagen: Munksgaard

(16) W. A. Hendrickson and K. B. Ward, *Acta Cryst.*, **A32**, 778 (1976)

(17) E. Lattman and W. E. Love, *Acta. Cryst.*, **B26**, 1854 (1970)

(18) S. N. Rao, J.-H. Jih, J. A. Hartsuck, *Acta Cryst.*, **A36**, 878 (1980)

Appendix A. DNA Triple Helices : Crystallizations and itsImplications

A-1. Introduction to the DNA triple helix

The association between complementary deoxyribonucleic acid (DNA) strands in aqueous solution yields a double helix (duplex) with additional hydrogen-bonding donors and acceptors directed along the major groove. These donors and acceptors are exposed to the surrounding environment and are able to interact with specific binding molecules, such as proteins to form a specific complex or other DNA molecules to form a triple helix (triplex). Triplex structures were discovered more than three decades ago in RNA sequences *(1)* and later found to occur in DNA sequences as well *(2,3)*, but such structures were generally assumed to be biologically irrelevant. Recently, there has been a resurgence of interest in the triplex stimulated by their possible roles in biological processes and by the potential for recognition of specific sequences in duplex DNA or RNA (for reviews, *4)*.

Homopurine-homopyrimidine (pur-pyr) sequences have been mapped to several sites in the regulatory regions of eukaryotic genes. These sequences have been suspected to exhibit unusual DNA structures, since they are hypersensitive to single-strand-specific nucleases such as S1 and undergo a transition in plasmids to an underwound state under conditions of moderately acidic pH and negative supercoiling *(5)*. The most widely accepted model is that mirror-repeated pur-pyr sequences loop out to form a pyr-pur-pyr triplex and a single stranded homopurine stretch, termed H-DNA *(6,7)*. It has also been suggested that pur-pyr sequences may play a role in telomer formation *(8)* and chromosome folding *(9)*. Triplex formation occurs in the recombination with the help of recombinases, such as the RecA protein of *E. coli (10)*.

Triplex formation has also been exploited as a strategy for the recognition and site-specific cleavage of DNA duplex. Moser and Dervan found that synthesized homo-pyrimidine oligodeoxyribonucleotide with Fe-EDTA attached binds to corresponding tract in DNA to form the pyr-pur-pyr triplex and cleaves the duplex DNA at the recognition sites *(11)*. This work has generated great interest because of the potential applications in chromosome mapping and in chemotherapeutic applications. The latter potential has been explored in studies on the inhibition of c-*myc* expression *in vitro* by triplex formation *(12)*.

Sequence-selective recognition of pur-pyr sequences within long molecules of duplex DNA was first reported in plasmids of a few thousand base pairs, where a single cut was observed when the triplex-forming oligonucleotide was equipped with either Fe-EDTA *(11)* or Cu-phenanthroline *(13)*. With Fe-EDTA as a cleaving reagent, Strobel *et al.* showed that a triplex-forming oligonucleotide with 18 base pairs (18mer) was able to cut the DNA of bacteriophage λ at a single position within the 48502 base pair genome *(14)*, and that a 20mer recognized and cleaved a single site inserted in chromosome III of the yeast *Saccharomyces cerevisiae*, which contains 3.4 x 10^5 base pairs *(15)*. Dervan *et al.* then developed a strategy to cleave long DNA sequences at a single site *(16,17)*. An oligonucleotide recognizing a pur-pyr sequence which overlaps a methylase recognition site prevents methylation at this site through triplex formation, while all other sites are methylated. When the endonuclease is added, only the unmethylated site is cleaved and the others are protected by methylation. This technique is similar to that described by Szybalsky *et al. (18,19)* in which a DNA-binding protein is used to protect a DNA sequence from being methylated. The restriction endonuclease then cleaves selectively this protected site.

In the triplexes, the third strand is located in the major groove of the duplex and its bases recognize the purine bases in the target sequences through specific hydrogen bonds (Fig A-1). With all nucleotides in the anti conformation, a pyrimidine third strand binds parallel to the purine strand in pyr-pur-pyr triplex. Thymine and protonated cytosine

residues in third strand form sequence-specific hydrogen bonds with adenine and guanine residues, respectively, leading to isomorphous TxA·T and C^+xG·C triplets (x represents Hoogsteen or reverse Hoogsteen hydrogen bond and · represents Watson-Crick hydrogen bond). The specific hydrogen bonds are referred to as Hoogsteen hydrogen bonds since they are discovered by Hoogsteen in cocrystals of adenine and thymine *(20)*. The pK value of isolated deoxycytidine is 4.3, but the apparent pK is raised in oligonucleotides and multistranded nucleic acids due to their polyanionic character. Pyr-pur-pyr triplex is stable at acidic pH and its stability decreases as the pH decreases. The pH dependence of this triplex is also affected by the distribution of cytosines within the oligonucleotide sequence. Intermolecular and intramolecular pyr-pur-pyr triplexes have been observed by ^{1}H NMR spectroscopy *(21-23)*.

In pur-pur-pyr triplex, a purine third strand binds antiparallel to the purine strand in the duplex. Guanine and adenine in the third strand form reverse Hoogsteen hydrogen bonds with guanine and adenine residues in the duplex, respectively, yielding non-isomorphous GxG·C and AxA·T triplets. The overall stability of the pur-pur base pairs is similar to that of the Watson-crick base pairs *(24)*. With the lack of an acidic pH requirement, pur-pyr-pur triplexes are appealing choices for use in applications involving oligonucleotide targeting of duplex DNA *in vitro* and *in vivo*.

Thermodynamic analyses show that triplexes are less stable than duplexes. The stability is expected to depend on the base triplet sequence and length, the nature and concentration of cations, and, in the case of cytosine-containing third strands, on pH. The melting temperatures (T_ms) were measured for the combination of duplexes 34 base-pair long containing 15 central pur-pyr tract and oligopyrimidine 15mers *(25)*. The T_ms are 32 °C and 44 °C for T_{15} and C_{15} third strands, respectively. The T_m of a short triplex is lower than that of a long triplex. Na^+ ion increases the T_m of TxA·T containing triplexes, but decreases the T_m of triplexes containing more than 70% of C^+xG·C *(26)*. Mg^{2+} ions stabilize the triplex as well as the duplex. The presence of Mg^{2+} ions is not required for

intramolecular triplex formation *(27)*. Polyamines favor triplex formation at the neutral pH *(28)*. Thermal-denaturation analyses of $[d(TC)]_n \cdot [d(GA)]_n$ show that at low ionic strength and pH 7, 3 μM spermine was sufficient to cause triplex formation from all the duplex. In 0.2 M NaCl, the mid-point of the duplex-to-triplex transition occurred at a pH of 5.8, but this was increased by nearly one pH unit in the presence of 500 μM spermine. Modified bases or backbone such as 5-methylcytosine *(29)* or phosphorothioate *(30)* have been introduced into triplex-forming oligonucleotides to improve the stability under physiological conditions.

There is no crystal structure yet available for short triple-stranded oligonucleotides. Fiber diffraction data are available on several triplexes *(31-33)*. It has been proposed that the structure of $(dT)_n x(dA)_n \cdot (dT)_n$ had a large axial rise per base triplet (3.26 Å) with the three strands forming a 12-fold helix of A'-DNA with C3'-endo sugar puckers. This structure has a very deep major groove which should be able to accommodate a third strand. Recently single crystals of DNA triplexes have been reported to exhibit fiber-type x-ray diffraction patterns *(34)*. The $(dT)_{12} x(dA)_{12} \cdot (dT)_{12}$ triplex with two-residue overhangs has 13 residues per turn and a rise per residue of 3.26 Å and the triplex of hetero sequence has 12 residues per turn and a rise per residue of 3.20 Å. FTIR spectroscopy studies with $(dT)_n x(dA)_n \cdot (dT)_n$ and $(dC^+)_n x(dG)_n \cdot (dC)_n$ showed that the DNA triplex has B-form geometry with C2'-endo sugar pucker *(35,36)*. A molecular dynamics simulation of the DNA triplex, $d(TC)_5 x d(GA)_5 \cdot d(TC)_5$ has also revealed considerable deviation from the fiber diffraction model *(37)*.

Triplexes are the subject of intensive studies by one- and two-dimensional NMR spectroscopy *(21-24,38-43)*. The presence of protonated cytosine was first observed in the intermolecular pyr-pur-pyr triplex by Rajagopal and Feigon *(22)*. Studies of a 31-base DNA oligonucleotide, which folds to form a stable intramolecular pyr-pur-pyr triplex at acidic pH, have shown that most of the triplex sugars are close to the C2'-endo conformation *(40)*. However, stacking interactions between the purines in an

intermolecular triplex containing 11 base triplets have suggested that the oligopurine strand exhibits an A-helical base stacking conformation *(21)*. The structure of a GxT·A triplet in a pyr-pur-pyr triplex *(40,41)* and that of a mismatch of TxC·G in a pur-pur-pyr triplex *(43)* have been also reported.

The crystal structure of DNA triplex is necessary to obtain detailed structural information, such as the alignment of base triplets, the sugar geometry and the local conformation. Thirteen intermolecular pyr-pur-pyr triplexes were used for the crystallization and the preliminary x-ray diffraction studies on half of them were performed. The crystallization conditions and the implications of the x-ray diffraction experiments are described in this section.

A-2. Crystallization

Oligonucleotides were synthesized by Matthew Taylor in Peter Dervan's group using standard phosphoramidite chemistry *(44)*. The oligonucleotides were deprotected by an overnight incubation in aqueous ammonia at 55 °C and purified by reverse-phase HPLC in two stages. After salt exchange with Na^+ ions, the oligonucleotides were hybridized and concentrated to 24 mg/ml solutions. The sequences of thirteen triplexes that used for the crystallization trials are shown in Fig A-2. Eight of them (YM10 to YM17) have the consensus sequence of (5'-CTCTCTTCTT) x (5'-GAGAGAAGAA) · (5'-TTCTTCTCTC) and have different overhangs of thymine residue at the 5' end of each strand. YM18 is YM11 with an insertion of two thymine residues in the middle of sequence. YM20 to YM22 contain the overhang of thymine residue at the 3' end of the Watson-Crick pyrimidine strand.

Crystallization conditions were screened with the conventional hanging drop method at 4 °C, since the melting temperature of a 10mer triplex is slightly higher than room temperature. Triplex solutions were diluted to 3 mg/ml in 5 mM sodium cacodylate, pH 6.0. The buffers used were sodium citrate or sodium cacodylate at pH's of 5.0, 5.4,

5.6, 5.8 and 6.0. The precipitates used were 2-methyl-2,4-pentanediol (MPD), polyethylene glycols (PEGs) with different molecular weights, ammonium sulfate and isopropanol. Different divalent cations such as magnesium, barium and calcium were tried for the crystallizations. Polycations such as spermine tetrahydrochloride and cobaltamine trichloride were added to stabilize the triplexes. The hanging drop contained equal volumes of the triplex solution and the reservoir solution. The sitting drop method and crystallizations at room temperature were also tried to refine crystallization conditions.

About five thousand crystallization trials were set up. Crystallization conditions are listed in Table A-1 for each triplex. YM10 crystals were grown from 10% PEG 1K, 10 mM $MgCl_2$, 40 μM spermine and 10 mM MES pH 6.0. YM12 and YM13 gave rod- or plate-shaped crystals, depending on the conditions. The HPLC analysis indicated that the YM12 crystal contained two pyrimidine strands and one purine strand, which confirmed the crystallized material was a triplex, not a duplex. YM11 and YM18 gave rod-shaped crystals (Fig. A-3). The average size of rod-shaped crystals was about 0.1 mm x 30 μm x 30 μm. YM11c was the biggest crystal with the dimensions of 0.4 x 0.1 x 0.1 mm^3. The YM18 crystals had short life time; crystals were fully grown within one week and deteriorated after 10 days. YM14, YM15, YM19 and YM20 gave showers of needle-shaped crystals. YM21 and YM22 gave fragile oval-shaped crystals with the dimensions of 0.1 x 0.1 x 0.03 mm^3. YM16 and YM17 gave large cubic crystals from 18% PEG 1K, 10 mM EDTA, 20 mM $MgCl_2$, 0.1 M NaCl, 100 μM spermine and 10 mM citrate pH 5.5, but they turned out to be salt crystals.

A-3. X-ray diffraction studies

X-ray diffraction studies were performed with an R-AXIS imaging plate system and a Siemens area detector with the CuKα radiation (λ = 1.54 Å). Typical crystal-to-detector distance was 100 mm. Table A-1 lists the results of x-ray diffraction studies for each crystal. YM12 was the first crystal to be tested with x-rays. Two precession photos

taken at room temperature showed weak spots at 2.8 Å resolution. The crystal turned out to be very sensitive to x-ray radiation and the following diffraction experiments were done at -160 °C. Since crystallization conditions contained lower concentrations of MPD or PEG 1K, higher concentrations of MPD or PEG 1K with proper salt concentrations were used as cryosolvents. Crystals were first transferred to the cryosolvent. After 5 or 10 minutes, crystals were picked up and put under the cold nitrogen stream. Gradients of MPD and PEG concentrations were also tried to minimize the damage in the crystal. YM12 and YM13 crystals gave only couple of spots at 20 Å resolution on R-AXIS.

The YM11a crystal diffracted x-ray to 4.0 Å resolution. One of the frames contained a zone, indicating a hexagonal space group (Fig. A-4 (a)). With the YM11c crystal, 16 still and 24 oscillation frames were measured on R-AXIS (Fig. A-4 (b)). They showed strong streak at about 3.4 Å resolution and the background of fiber diffraction pattern of a B-type DNA. Single spots at 2.0 Å resolution were also found in the same frame, which indicated part of the crystal was well ordered. The indexing was unsuccessful with the program DENZO *(45)* or the R-AXIS software due to the lack of spots. The program XDS *(46)* indexed 48 spots among 57 measured spots with the cell constants of $a= b= 25.6$ Å, $c= 230.07$ Å, $\alpha= \beta= 90°$, $\gamma= 120°$. The YM11d crystal gave many spots up to 2.4 Å resolution with rather large mosaicity. One orientation data was collected but the program XDS was unable to index them.

The YM18b crystal showed the apparent fiber diffraction pattern and the rest of YM18 crystals gave only couple of spots at 20 Å resolution. The YM22 crystals gave the characteristic hexagonal spots at 20 Å resolution.

A-4. Implications

Triplexes with the consensus sequences but different overhangs were crytallized from similar conditions. The HPLC analysis confirmed that the crystallized material was in fact a triplex. As expected, the overhangs had large effects on the crystal growth.

Triplexes with the overhang at the 5' end of the Watson-Crick pyrimidine strand gave needle-shaped crystals but those with the overhang at the 5' end of the Hoogsteen pyrimidine strand gave relatively big crystals. The effect of the overhangs at the 5' end of the Watson-Crick purine strand was not obvious. YM11 and YM18 triplexes gave similar shaped and diffraction-quality crystals, indicating the end residues are more important than the length of the triplex in the crystal packing. The biggest problem in crystallization was the reproducibility of the crystal. Different ages of the same solution and different batches of the same triplex gave different crystallization results. There may be an equilibrium between duplex and triplex in solutions.

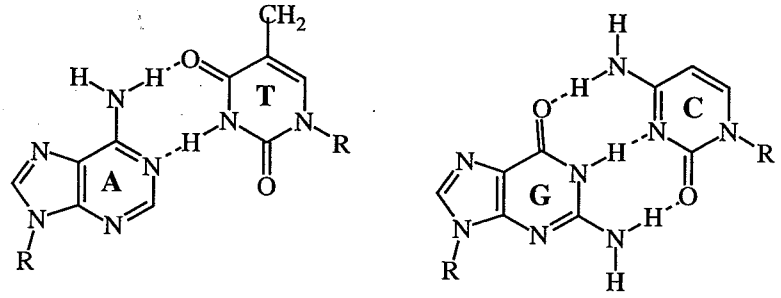
The preliminary diffraction studies provide some information about the triplex structure. YM11 and YM18 crystals gave single spots at the high resolution and the background of the fiber diffraction, which implied that part of the crystal was well ordered in three dimensions and the rest was ordered only in one dimension. Strong streak at around 3.4 Å resolution and the pattern of fiber diffraction suggested that a B-type DNA was present in the crystal. The approximate cell constants of the YM11c crystal indicated that there were 72 base pairs stacked along the *z* axis in one cell and 12 base pairs in one asymmetric unit, if the axial rise per base is 3.20 Å. The preliminary results are in agreement with other studies based on NMR spectroscopy and fiber diffraction studies. The details on the triplex structure will be obtained only with the well-diffracting crystals and the complete structure determination.

Table A-1. Crystallization conditions and results of x-ray diffraction

sample	crystallization conditions	x-ray diffraction
YM12	20% MPD, 20mM MgCl2, 10mM MES pH 6.0	Precession photos spots at 2.8 Å
YM13	15% PEG 1K, 20mM MgCl2, 50μM spermine, 10mM MES pH 6.0	R-AXIS, spots at 20 Å
YM11a	20% PEG 1K, 20mM $BaCl_2$, 0.1M NaCl, 0.1mM spermine, 10mM citrate pH 5.0	R-AXIS, spots at 4.0 Å hexagonal spots
YM11b	22% MPD, 0.1mM $Co(NH_3)_6Cl_3$, 0.1M NaCl, 0.5mM spermine, 10mM cacodylate pH 5.6	R-AXIS, spots at 2.7 Å
YM11c	20% MPD, 0.1mM $Co(NH_3)_6Cl_3$, 0.1M NaCl, 0.5mM spermine, 10mM cacodylate pH 5.6	R-AXIS, spots at 2.7 Å strong streak at 3.4 Å unit cell of 25x25x230?
YM11d	16% MPD, 10mM $MgCl_2$, 0.1M NaCl, 80μM spermine, 10mM citrate pH 5.0	Siemens many spots up to 2.4 Å
YM11e	26% MPD, 0.14mM $Co(NH_3)_6Cl_3$, 0.5mM spermine, 10mM citrate pH 5.0	Siemens, spots at 20 Å
YM18a	20% MPD, 10mM $MgCl_2$, 0.1M NaCl, 150μM spermine, 10mM citrate pH 5.6	R-AXIS, spots at ~3 Å ice formation
YM18b	14.5% MPD, 8mM $MgCl_2$, 0.1M NaCl, 150μM spermine, 10mM cacodylate pH 5.6	R-AXIS, background of fiber diffraction pattern
YM18c	20% MPD, 10mM $MgCl_2$, 0.1M NaCl, 150μM spermine, 10mM citrate pH 5.6	Siemens, spots at 20 Å
YM22a	18% PEG 1K, 24mM $MgCl_2$, 0.2mM spermine, 0.1mM $Co(NH_3)_6Cl_3$, 10mM citrate pH 5.0	Siemens, spots at 20 Å
YM22b	15% PEG 1K, 5% isopropanol, 0.1M NaCl, 20mM MgCl2, 0.1mM $Co(NH_3)_6Cl_3$, 0.2mM spermine, 10mM citrate pH 5.4	Siemens, spots at 20 Å

Figure A-1. (a) Watson-Crick base pair (b) base triplets formed by A·T and G·C base pairs with T, protonated C, G and A.

(a)



(b)

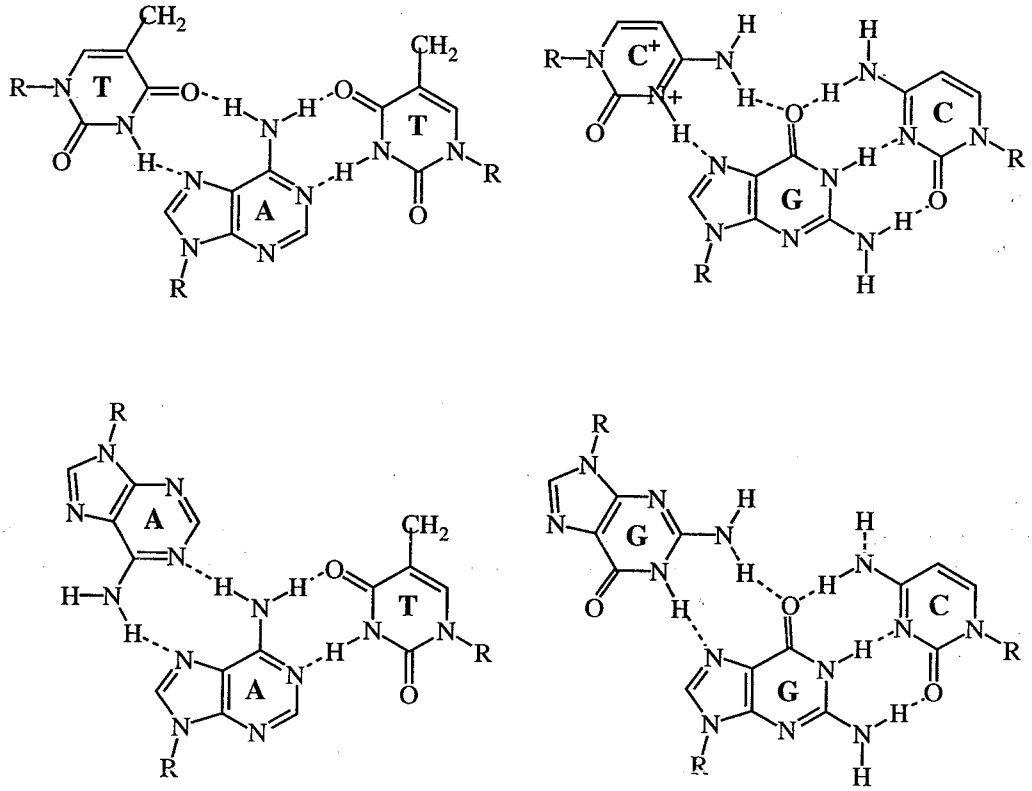


Figure A-2. DNA triplex sequences

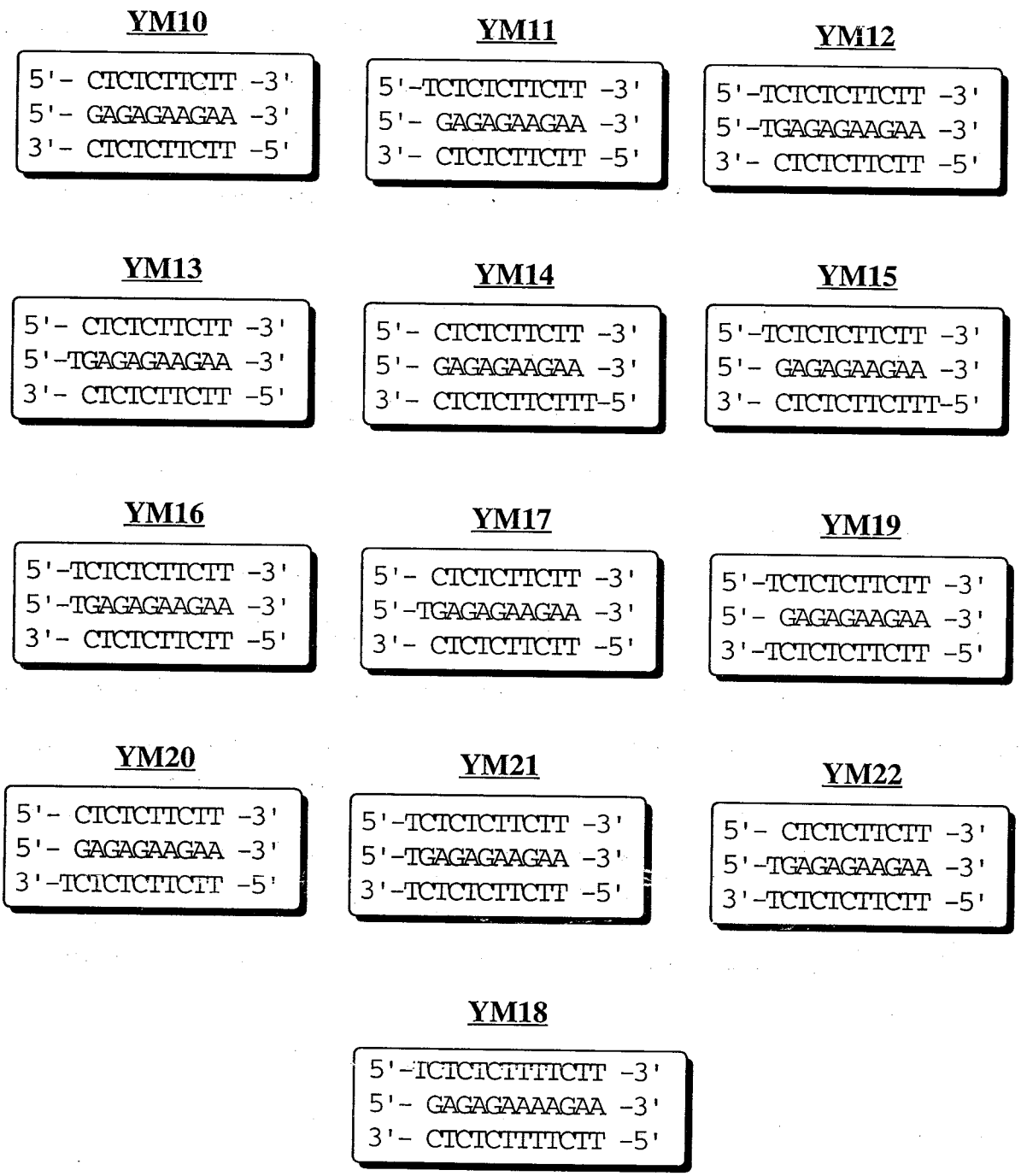
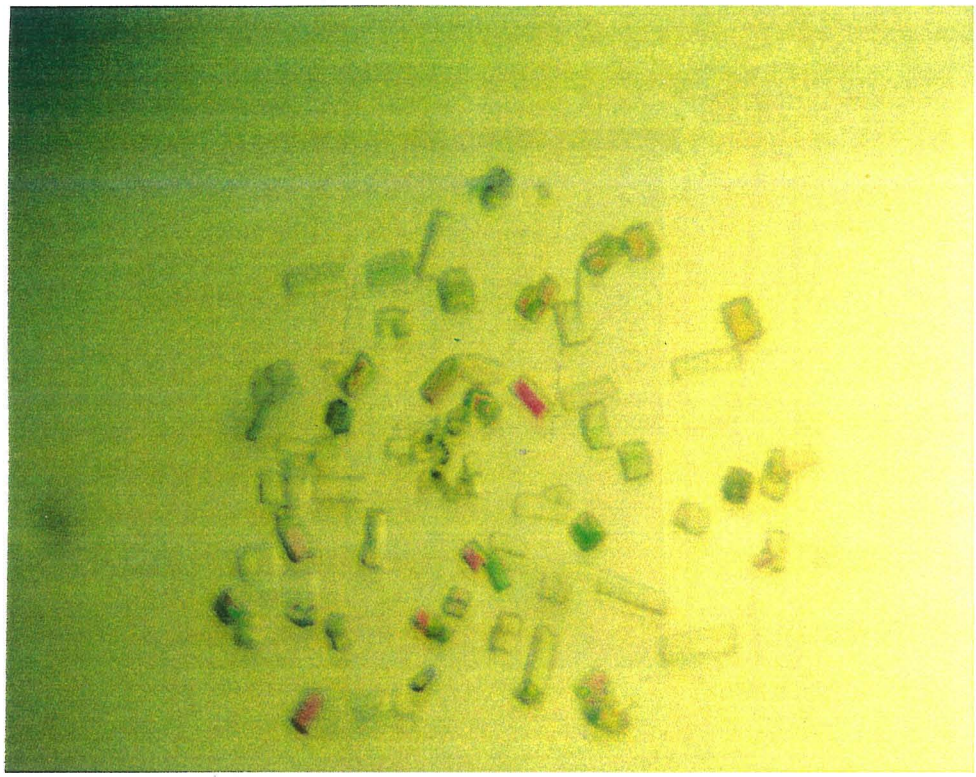


Figure A-3. Triplex Crystals (a) YM11 (b) YM18

(a)



(b)

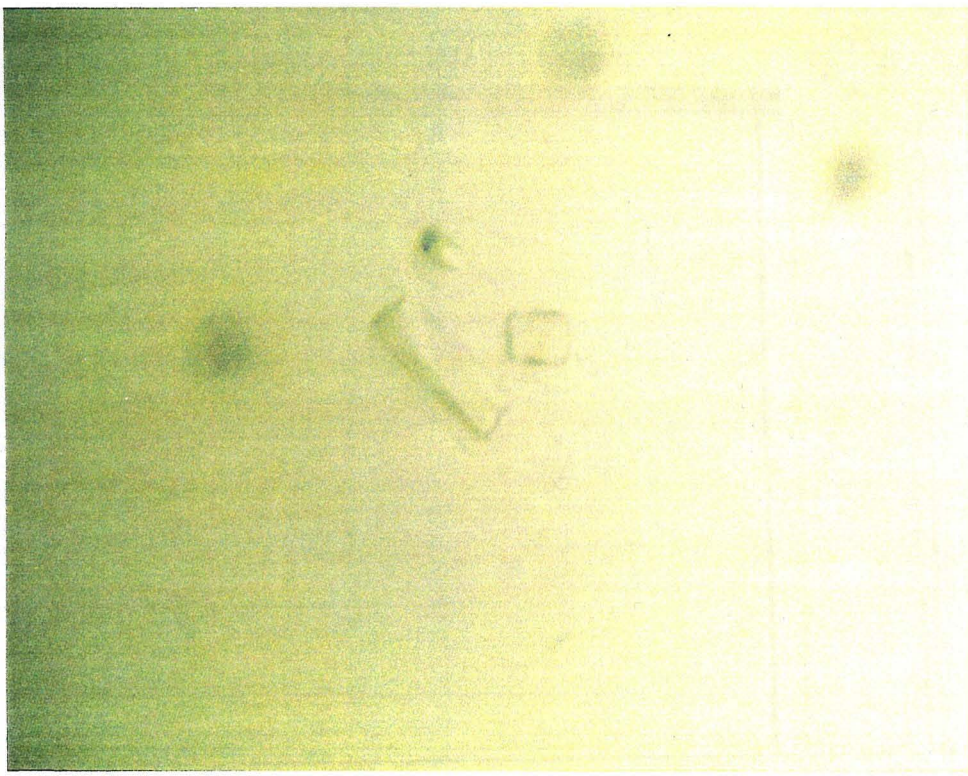
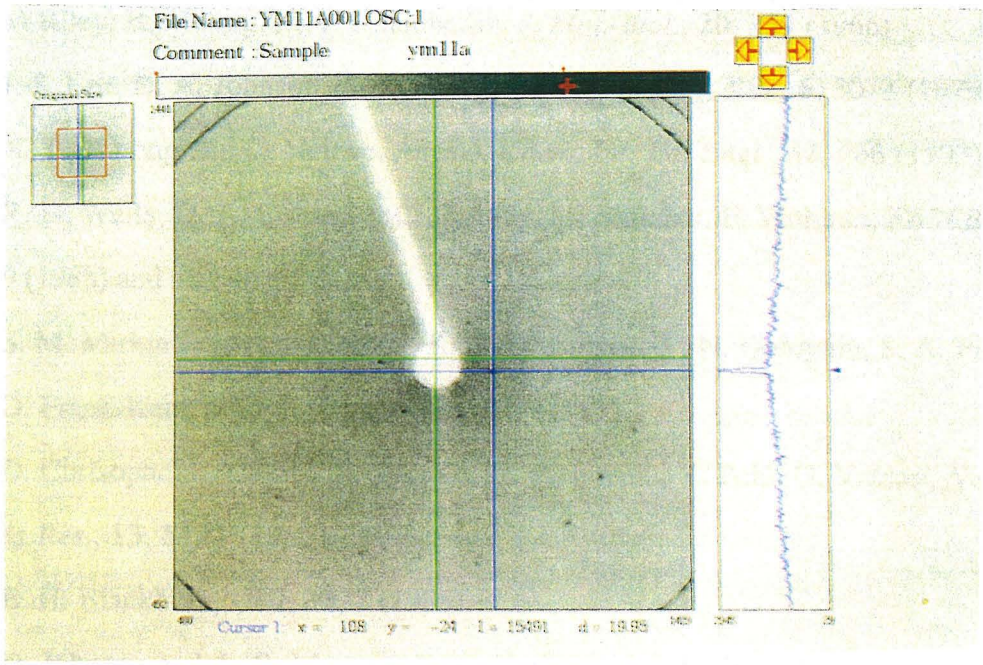
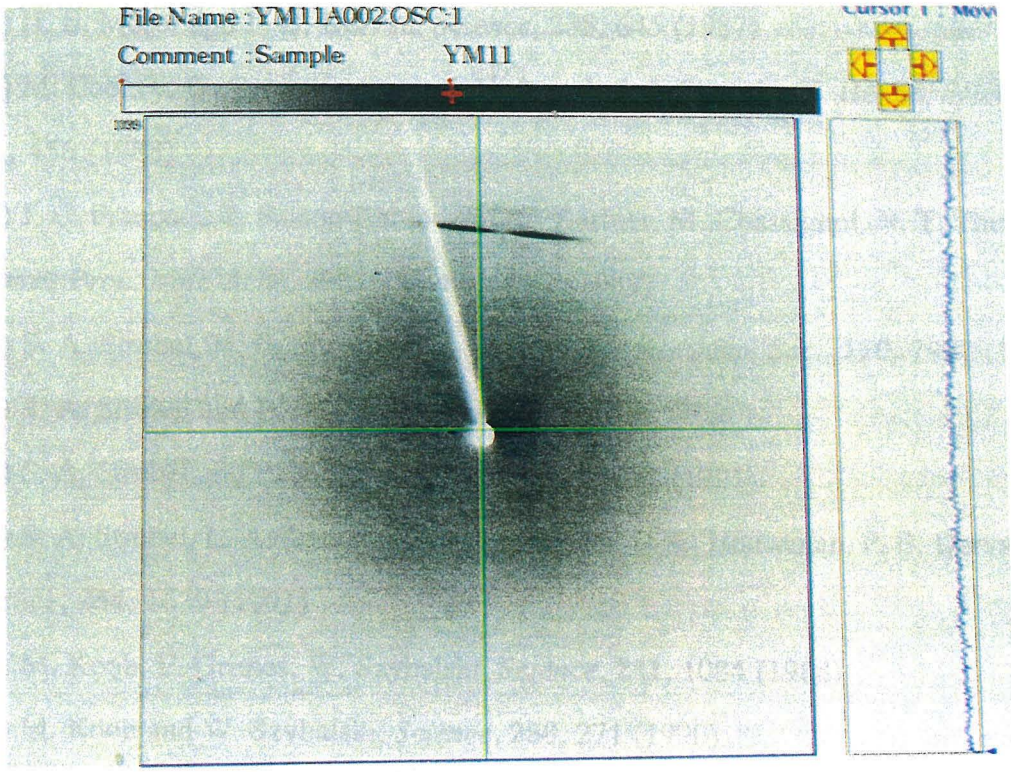


Figure A-4. Diffraction Patterns (a) YM11a (b) YM11c

(a)



(b)



References

(1) G. Felsenfeld, D. Davies, A. Rich, *J. Am. Chem. Soc.*, **79**, 2023 (1957)

(2) M Riley, B. Maling, M. J. Chamberlin, *J. Mol. Biol.*, **20**, 359 (1966)

(3) J. S. Lee, D. A. Johnson, A. R. Morgan, *Nucleic Acids Res.*, **6**, 3073 (1979)

(4) N. T. Thuong and C. Helene, *Angew. Chem. Int. Ed. Engl.*, **32**, 666 (1993)

(5) R. D. Wells, D. A. Collier, J. C. Hanvey, M. Shimizu, F. Wohlrab, *FASEB J.*, **2**, 2939 (1988) and references therein

(6) S. M. Mirkin, V. I. Lyamuchev, K. N. Drushlyak, V. N. Dobrynin, S. A. Filippov, M. D. Frank-Kamenetskii, *Nature*, **330**, 495 (1987)

(7) D. Chrisophe, B. Cabrer, A. Bacolla, H. Targovnik, V. Pohl, G. Vassart, *Nucleic Acids Res.*, **13**, 5127 (1985)

(8) E. H. Blackburn, *Cell*, **37**, 7 (1985)

(9) D. Johnson and A. R. Morgan, *Proc. Natl. Acad. Sci. USA*, **75**, 1637 (1978)

(10) P. Hsieh, C. S. Camerini-Otero, R. D. Camerini-Otero, *Genes Dev.*, **4**, 1951 (1990)

(11) H. E. Moser and P. B. Dervan, *Science*, **238**, 645 (1987)

(12) M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, *Science*, **241**, 456 (1988)

(13) J. C. Francois, T. Saison-Behmoaras, C. Barbier, M. Chassignol, N. T. Thuong, C. Helene, *Proc. Natl. Acad. Sci. USA*, **86**, 9702 (1989)

(14) S. A. Strobel, H. E. Moser, P. B. Dervan, *J. Am. Chem. Soc.*, **110**, 7927 (1989)

(15) S. A. Strobel and P. B. Dervan, *Science*, **249**, 73 (1990)

(16) S. A. Strobel and P. B. Dervan, *Nature*, **350**, 172 (1991)

(17) S. A. Strobel, L. A. Doucette-Stamm, L. Riba, D. E. Houseman, P. B. Dervan, *Science*, **254**, 1639 (1991)

(18) M. Koob, E. Grimes, W. Szybalski, *Science*, **241**, 1084 (1988)

(19) M. Koob and W. Szybalski, *Science*, **250**, 271 (1990)

(20) K. Hoogsteen, *Acta Cryst.*, **12**, 822 (1959)

(21) C. de los Santos, M. Rosen, D. Patel, *Biochemistry*, **28**, 7282 (1989)

(22) P. Pajagopal and J. Feigon, *Nature*, **339**, 637 (1989)

(23) V. Sklenar and J. Feigon, *Nature*, **345**, 836 (1990)

(24) D. S. Pilch, C. Levenson, R. H. Shafer, *Proc. Natl. Acad. Sci. USA*, **87**, 1942 (1990)

(25) R. G. Shea, P. Ng, N. Bischofberger, *Nucleic Acids Res.*, **18**, 4859 (1990)

(26) M. Rougee, B. Faucon, J. L. Mergny, F. Barcelo, C. Giovannangeli, T. Garestier, D. Helene, *Biochemistry*, **31**, 9269 (1992)

(27) F.-M. Chen, *Biochemistry*, **30**, 4472 (1991)

(28) K. J. Hampel, P. Crosson, J. S. Lee, *Biochemistry*, **30**, 4455 (1991)

(29) J. S. Lee, M. L. Woodsworth, L. J. P. Latimer, A. R. Morgan, *Nucleic Acids Res.*, **12**, 6603 (1984)

(30) L. J. P. Latimer, K. Hampel, J. S. Lee, *Nucleic Acids Res.*, **17**, 1549 (1989)

(31) A. Arnott and E. Selsing, *J. Mol. Biol.*, **88**, 509 (1974)

(32) S. Arnott, P. J. Bond, E. Selsing, P. J. C. Smith, *Nucleic Acids Res.*, 2459 (1976)

(33) J. Lourdes Campos and J. A. Subirana, *J. Biomol. Struct. Dyn.*, **8**, 793 (1991)

(34) K. Liu, H. Todd Miles, K. D. Parris, V. Sasisekharan, *Struct. Biol.*, **1**, 11 (1994)

(35) F. B. Howard, H. T. Miles, K. Liu, J. Frazier, G. Raghunathan, V. Sasisekharan, *Biochemistry*, **31**, 10671 (1992)

(36) A. Akhebat, C. Dagneaux, J. Liquier, E. Taillandier, *J. Biomol. Struct. dyn.*, **10**, 577 (1992)

(37) M. M. W. Mooren, D. E. Pulleyblank, S. S. Wijmenga, M. J. J. Blomers, C. W. Hilbers, *Nucleic Acids Res.*, **18**, 6523 (1990)

(38) R. Macaya, E. Wang, P. Schultze, V. Sklenar, J. Feigon, *J. Mol. Biol.*, **225**, 755 (1992)

(39) E. Wang, S. Malek, J. Feigon, *Biochemistry*, **31**, 4838 (1992)

(40) I. Radhakrishnan and D. J. Patel, *J. Am. Chem. Soc.*, **114**, 6913 (1992)

(41) I. Radhakrishnan and D. J. Patel, *J. Am. Chem. Soc.*, **115**, 1615 (1993)

(42) K. Ditrich, J. Gu, R. Tinder, M. Hogan, X. Gao, *Biochemistry*, **33**, 4111 (1994)

(43) C. A. Laughton and S. Neidle, *J. Mol. Biol.*, **223**, 519 (1992)

(44) L. J. McBride and M. H. Caruthers, *Tetrahedron Lett.*, **24**, 245 (1983)

(45) Z. Otwinowski, *Denzo: a film processing program for macromolecular crystallography*, Yale University, New Haven (1991)

(46) W. Kabsch, *Acta Cryst.*, **21**, 916 (1988)

Appendix B. Molecular Replacement of the Reaction Center in the Tetragonal Space Group

B-1. Introduction to the Reaction Center

Photosynthesis is the biological process by which the electromagnetic energy (light) is converted to chemical energy. In the photosynthetic bacteria, the photochemical reactions occur between pigments bound within membrane protein complexes called the reaction centers (RCs). The structure of the RC from *Rhodopseudomonas (Rps.) viridis* RC has been determined at 2.3 Å resolution *(1,2)*. Three structures of the RCs from *Rhodobacter (Rb.) sphaeroides* have been separately determined by the molecular replacement method using appropriately modified coordinates of the *Rps. viridis* RC as a search model *(3-9)*.

The RC from *Rb. sphaeroides* is built up from three subunits, which are called L (light), M (medium) and H (heavy) for apparent molecular weights determined with SDS gel electrophoresis. Subsequent amino acid sequence determination showed, however, that the H subunit is in fact the smallest with 260 amino acids, followed by the L subunit with 281 amino acids. The M subunit is the largest polypeptide with 307 amino acids *(10-12)*. The L and M subunits are homologous with the sequence identity of about 25%. The L and M subunits are firmly anchored in the membrane, each by five hydrophobic transmembrane α helices. The H subunit has one transmembrane α helix near the amino terminus and the carboxy terminal region is positioned on the cytoplasmic side and interacts with the other two subunits. The presence of eleven transmembrane α helices had been predicted form hydropathy analysis of the amino acid sequence *(12)*. In addition to these subunits, the RC contains ten cofactors: four bacteriochlorophylls (BChls, two of which form the strongly interacting dimer, called 'the special pair'), two bacteriopheophytins (BPhes), one non-heme iron atom, two quinones (Q_A and Q_B), and one carotenoid. The

cofactors are bound to the L and M subunits, most of them to the transmembrane helices. The structurally similar L and M subunits are related by a pseudo-twofold symmetry axis, which lies approximately through the special pair and non-heme iron cofactors.

In the RC, the absorption of a photon excites the primary electron donor (the special pair). The electron from the special pair is transferred to one of the BPhe molecules, to the primary acceptor, Q_A, and finally to the secondary acceptor, Q_B *(13)*. Following a second electron transfer reaction, the doubly reduced quinone picks up two protons to form dihydroquinone, which leaves the RC to be replaced by an exogenous quinone. The mobile dihydroquinone transfers electrons to a second membrane bound complex, the cytochrome bc_1, concomitantly releasing protons to create proton gradients across the membrane that provides the driving force for ATP synthesis *(14)*. There are two BPhe molecules, one on each side of the twofold axis, that in principle could accept the electrons, but only one pathway is used for electron transfer. This behavior must be due to different cofactor-protein interactions along the two branches *(1,15)*.

The amino acid sequences of subunits M, L, and H show 59, 49 and 39 % homology, respectively, between *Rps. viridis* and *Rb. sphaeroides*: their three dimensional structures are very similar. The differences are in the types of the cofactors and in the secondary electron donor which reduces the photooxidized primary donor. The RC from *Rb. sphaeroides* has a-type BChl and BPhe and two identical quinone (ubiquinone-10). The RC from *Rps. viridis* has b-type BChl and BPhe and a menaquinone-9 and a ubiquinone-9. In *Rps. viridis*, a tetra-heme c-type cytochrome is tightly bound to the transmembrane protein core of the RC. In *Rb. sphaeroides*, the secondary electron donor is an exogenous cytochrome c_2 (cyt c_2), that forms transient complex with the RC.

The cyt c_2 is a 14 kD water-soluble protein and contains a single covalently bound heme as its prosthetic group (Fig. B-1). The crystal structure of the cyt c_2 from *Rb. sphaeroides* has recently been determined to 1.6 Å resolution and contains five α helices, a short stretch of antiparallel β sheet and an antiparallel β loop *(16)*.

Different models for the interaction between the cyt c_2 and the RC have been proposed. Allen *et al.* proposed that the cyt c_2 binds in a cleft in the RC located directly above the special pair *(3)*. Tiede and Chang postulated that the binding site is on the M subunit of the RC *(17)*. In both models, a series of negatively charged amino acids on the RC are assumed to interact electrostatically with positively charged amino acids in the cyt c_2. To understand the details of the interactions between the cyt c_2 and the RC, co-crystallographic studies of the RC and cyt c_2 were performed. The data processing, the molecular replacement (MR) method and the solvent flattening on a tetragonal crystal containing the RC and cyt c_2 are described in the first part of this appendix.

To date, four reports of crystallization of orthorhombic crystal forms of the RC from *Rb. sphaeroides* have been published *(18-21)*. These crystals have similar cell constants and diffract x-rays from about 3.7 Å to 2.8 Å resolution. A trigonal crystal form has been obtained which diffracts x-rays to at least 2.65 Å resolution *(22)*. A tetragonal crystal form of the RC from *Rb. sphaeroides* diffracts x-rays up to 2.8 Å resolution. The data processing, the MR method and the averaging on this tetragonal RC crystal are described in the second part of this appendix.

B-2. The cocrystal of the RC and cyt c_2

B-2-1. Data processing

A possible cocrystal of the RC and cyt c_2 was obtained by Noam Adir of the University of California, San Diego. Three data sets were collected with different detectors and processed with different programs. Data statistics are listed in Table B-1. The 'coxtal1' data set was collected at room temperature, using CuKα radiation (λ= 1.54 Å) from an R-AXIS imaging plate system and processed with the R-AXIS software by Barbara Hsu. The data set was collected from a single crystal and contained 13503 unique reflections. The merging *R* factor was about 12 % and the data was 82.5% complete to 4.5 Å resolution.

The 'coxtal2' data set was collected with synchrotron radiation at the wavelength of 1.08 Å and the detector distance of 250 mm. Eighteen images with 2° oscillation were collected and processed with MOSFLM *(23)* and merged with Agravata/Rotavata by myself. *R* factors increased from 18.0% in the first frame to 32.5% in the last frame, due to the severe radiation damage. The data set contained 12683 independent reflections and had an *R*merge of 22.6% and the completeness of 56.2% to 4.0 Å resolution. The *R* factor with the 'coxtal1' data set is 22% for the resolution range of 50 to 5 Å.

The 'coxtal3' data set was collected with a Siemens area detector with CuKα radiation and processed with XDS *(24)* and XSCALE by Douglas Rees. The data were collected from two orientations and measured 5965 unique reflections. The data set was 86.3% complete to 6 Å resolution with an *R*merge of 9.21%. This data set agreed better with 'coxtal2' data than 'coxtal1' data; *R* factors with the 'coxtal1' and 'coxtal2' data sets were 17.75% and 13.51%, respectively, for the resolution range of 50 to 6 Å.

The crystals belong to the space group $P4_32_12$ or $P4_12_12$, based on the systematic absence, and have cell constants of $a= b= 142.7$ Å and $c= 254.8$ Å. The solvent content and the numbers of molecules in the asymmetric unit were analyzed with the Matthews equation *(25)*. With the Matthews coefficient for globular proteins ($V_M= 2.61$), the predicted number of molecules are 2.59 and 2.26 for the RC only and the RC plus cyt c_2, respectively. If the crystal has the Matthews coefficient of the orthorhombic RC ($V_M=$ 4.03) and the solvent content of 69.5%, the predicted number of molecules are 1.68 and 1.46 for the RC only and the RC plus cyt c_2, respectively. The number of molecules in the asymmetric unit could be one or two.

B-2-2. Self-rotation function

The self-rotation function was calculated using the X-PLOR package *(26)* and Crowther's fast rotation function *(27)*. Fig. B-2 (a) shows the Crowther's self-rotation function, calculated with the resolution range of 8 to 5 Å, the integration radius of 30 Å and

the 'coxtal1' data set. The peaks are the crystallographic fourfold axis along the *c* axis and the twofold axis along *a* and *b* axes. A result of X-PLOR self-rotation function search is listed in Fig. B-2 (b), calculated with the 'coxtal1' data set and the resolution range of 8 to 5 Å. Apparently there is no significant peak except the origin peak. Peaks for the noncrystallographic symmetry axis or the pseudo-twofold axis in the RC molecule were not present in the self-rotation function. The 'coxtal2' and 'coxtal3' data sets gave similar results in the self-rotation function calculations. The noncrystallographic symmetry might be on the same direction with the crystallographic axis or there might be only one molecule in the asymmetric unit.

B-2-3. Molecular replacement

The MR method was used to find the orientations and positions of the RC in the crystal. The calculations were performed with the programs MERLOT *(28)* and X-PLOR. The detailed descriptions about the MR method and the programs are given in Chapter 4. The structure of the RC from *Rb. sphaeroides* strain R-26 (Brookhaven Data Bank coordinate set 4rcr.pdb) was used as the model, whose is available from . It was refined to an *R* factor of 22.7% at 2.8 Å resolution. The average r.m.s. deviations of bond length and angle are 0.019 Å and 3.10°, respectively. The model was placed in an orthorhombic P_1 cell with the dimensions of 160 x 160 x 160 $Å^3$ and all protein atoms and cofactors were used for structure factor calculations.

In MERLOT, the rotation function (RF) search was performed over angular ranges of α= 0° - 90°, β= 0° - 90°, γ= 0° - 360° *(29)*. A 2.5° step size in β was used and step sizes in α and γ were determined by the program as 1.25° and 5°, respectively. The RF was evaluated at various resolution limits and integration radii. Table B-3 lists resolution ranges, integration radii, RF solutions and r.m.s. deviations of the highest peaks. The exact values of the RF solution were determined using the LATSUM program, which calculates the rotation function according to the method of Lattman and Love *(30)*.

LATSUM does not require the radius of integration and is more flexible in the sampling of rotation space. TF searches were done with the home program GENTF, a brute force translation function (TF) search program *(31)*. The properly oriented model is placed in the cell of the unknown crystal and the structure factors are calculated. The program calculates *R* factors and correlation coefficients at every positions by applying phase changes due to the translations.

In X-PLOR, the asymmetric unit of the rotation function is θ_1= 0° - 360°, θ_2= 0° - 90°, θ_3= 0° - 90°. Patterson peaks were selected whenever their values are greater than 0.0 and their lengths fall between 5 Å and 40 Å. Step sizes of 2.5° in θ_1, θ_2 and θ_3 were used. Five-thousand rotation angles were selected by RF values and rotation angles different by less than 10° are considered as the same (ε=0.25). Patterson correlation refinement was done with 40 steps of rigid body refinement. If the distance of two rotations after PC refinement is less than 10°, two rotations were considered as the same. Table B-4 lists results of RF searches for each data set and the resolution range. The highest peak and the second highest peaks were shown. TF searches were performed in the fractional coordinates within the translational asymmetry unit of *x*= 0 - 1, *y*= 0 - 1 and *z*= 0 - 1/2. Correlation coefficients between the squares of the model and observed structure factors and the packing functions were calculated. *R* factors for the highest peaks were also given.

With the 'coxtal1' data set, RF searches initially gave solutions around β= 0 section with both MERLOT and X-PLOR. In β= 0 section, the rotation with same (α+γ) values are the same, which makes the interpretation of the result difficult. The model was rotated by 45° along the β axis and RF searches were repeated. Several consistent RF solutions were found, but the TF solutions could not be determined unambiguously. At this point, the cell constants of the model were increased to 160 Å and the data at the low resolution were included for the calculation. MERLOT RF searches showed two consistent peaks around α= 85.0°, β= 87.5°, γ= 45.0° and α= 2.5°, β= 90.0°, γ= 225.0° with some

exceptions, such as the resolution range of 10 to 7 Å and 25 Å integration radius (Table B-2 (a)). The r.m.s. value of the highest peak becomes higher as the integration radius increases. Two peaks were determined after the LATSUM calculation as α= 90.0°, β= 90.0°, γ= 45.5° and α= 0.0°, β= 90.0°, γ= 225.5°, which were crystallographically related. X-PLOR RF searches gave the similar solutions in the resolution ranges of 10 - 8 Å, 15 - 7 Å and 15 - 8 Å (Table B-3 (a)). The correctness of the solution in these resolution ranges was obvious from the values of correlation coefficients (about 3% higher than the next highest peak). TF searches were done with both GENTF and X-PLOR (Table B-5 (a) and (b)). The solution in the space group $P4_32_12$ gave higher correlation coefficient and lower *R* factor than those of space group $P4_12_12$ and $P4_22_12$. The correlation coefficient was more sensitive indicator for the correctness of the solution than the *R* factor or the packing function. The final *R* factor after rigid body refinement was 40.98% for the resolution range of 15 to 5 Å.

The MR with the 'coxtal2' data set was done as the 'coxtal1' data set. MERLOT gave an incorrect solution as the highest peak in almost every condition (Table B-2 (b)) and the correct solution as the next highest peaks in some cases. The r.m.s. values for the highest peaks were about one standard deviation lower than those from the 'coxtal1' data set. X-PLOR gave the correct RF solution with high correlation coefficients (Table B-3 (b)). The RF search failed to provide possible correct solution in two cases where the correct solution was added to the list of the RF solutions before PC refinement. The TF solution agreed well with the result of the 'coxtal1' data set. The completeness of the 'coxtal2' data set is about 56% in each resolution shell. The low completeness of data affected the accuracy of the RF solution more in MERLOT than in X-PLOR. The final *R* factor after rigid body refinement was 36.61% for the resolution range of 15 to 5 Å.

The MR with the 'coxtal3' data set was straight forward. RF searches gave the correct solution both in MERLOT and in X-PLOR (Tables B-2 (c) and B-3 (c)). TF

searches also gave obvious results (Table B-5 (d)). The final *R* factor after rigid body refinement was 36.27% for the resolution range of 15 to 6 Å.

The cross-rotation function was also calculated with the observed structure factors. The structure factors of the orthorhombic RC ('rc19') and the tetragonal RC ('rctet') were available. The orthorhombic RC has cell constants of a= 138.0 Å, b= 77.5 Å, and c= 141.8 Å and one molecule in the asymmetric unit. The tetragonal RC has the cell constants of a= b= 142.3 Å and c= 276.6 Å and two molecules in the asymmetric unit. The MR on the tetragonal RC are described in section B-3. The cross-rotation functions were calculated with the program MERLOT (Table B-4). An advantage in using the observed structure factors is that the very low resolution data can be included for the calculation. The correct cross-rotation solutions were found between 'rc19' and 'rctet' sets in the resolution ranges of 50 - 6 Å and 50 - 7 Å. The 'coxtal1' and 'rc19' sets gave completely wrong RF solutions in each and every conditions. The 'coxtal1' and 'rctet' sets gave a consistent RF solution of θ_1= θ_2= 0.0°, θ_3= 90.0°, which is about 20° off from the correct RF solution. The 'coxtal3' and 'rc19' sets gave the correct solution, except in the resolution range of 50 to 7 Å and the integration radius of 40 Å. The 'coxtal3' and 'rctet' sets gave similar results to those of 'coxtal1' and 'rctet' sets.

The correct MR solution was determined unambiguously from three different data sets (Table B-6). There was only one RC molecule in the asymmetric unit. Three RF solutions were within 1° from each other and three TF solutions agreed within a distance of 1.5 Å. The low resolution data were needed for the RF calculation and the accurate determination of the RF solution was important for the TF search.

The (2|fol-|fcl) and (|fol-|fcl) electron density maps were calculated with the observed structure factors and the rigid body refined model. The molecular packing of the RC in the crystal provided a possible position on the periplasmic side of the M subunit, which was adjacent to part of the cytoplasmic domain of the H subunit of a symmetry-related RC molecule. The cyt c_2 has a heme group whose density should be significant compared to

the rest of the molecule. Fig. B-3 shows the electron density maps around a possible cytochrome binding position with 1.5σ contour. Unfortunately, densities for the heme or α helices in the cytochrome could not be found in the difference maps calculated with any data sets. With the 'coxtal1' data set, several residual densities were found in the expected binding position but the density for the heme group was ambiguous. With the 'coxtal2' data set, several stretches of densities are clearly visible. Interestingly, these densities only extended along the *z* direction, which implies they were originated from the lack of reflections along that direction. There is no apparent density for the cyt c_2 with the 'coxtal3' data set.

B-2-4. Solvent flattening

Since the cocrystal has high solvent content, the solvent flattening was performed with the program PHASES. PHASES is the program package for phasing reflection data for macromolecular crystals, which contains routines for the phase refinement and the solvent flattening with the negative density truncation. The protein-solvent boundary is automatically determined by Wang's algorithm in the reciprocal space *(32)*. The MR phases calculated with the program X-PLOR were imported to the program PHASES. The starting value for the figure of merit was arbitrary and those of the Hendrickson-Lattman coefficients were zeros. Three cycles of solvent flattening were performed. At the beginning of each cycle, a new molecular envelope was created. The program replaces densities in the solvent region with the average solvent density and negative densities in the protein region with zeros.

Different solvent contents were tried for the calculation. *R* factors and correlation coefficients are better with lower solvent content, but the average phase changes remained at about 26° regardless of the solvent content (Table B-7). The solvent-flattened maps are shown in Fig. B-4. With the 'coxtal1' data set, several continuous densities are present away from the expected binding position and the evidence for the heme group was

inconclusive. With the 'coxtal3' data set, the solvent-flattened maps contain an extra density, that is close to the M subunit but too small for the cyt c_2 molecule. The position of the extra density remained the same but its size changed with different solvent contents. The extra density might originate from the truncation effect in the Fourier synthesis due to the resolution limit of the data.

Unfortunately, the above analysis did not provide positive proof for the existence of the cyt c_2 in the crystal. The lack of the density for the cyt c_2 could be due to the low occupancy of the cyt c_2 in the binding position or to disorder of small cyt c_2 molecule within solvent region of the crystal.

B-3. The RC in the tetragonal crystal form

B-3-1. Data processing

The data set of the tetragonal crystal form of the RC from *Rb. sphaeroides* was collected with synchrotron radiation at the wavelength of 1.08 Å and the detector distance of 220 mm. The crystals diffracted to 2.8 Å resolution. One hundred frames with 1° oscillation were processed with the program DENZO *(33)* and scaled with SCALEPACK. The crystals belong to the space group $P4_32_12$ and have cell constants of a= b= 142.3 Å and c= 276.6 Å. Two molecules are present in the asymmetric unit. Data statistics are listed in Table B-8. The data set had an *R*merge of 15.7% and a completeness of 84.3%, for the resolution range of 50 - 3 Å. High values of merging *R* factors at high resolution shells were due to the short detector distance, which made spots overlap each other. With the 2σ cutoff on the observed structure factors, the completeness decreased to 66.9% and the completeness of the last shell was only 28.2%. This indicates that the data set contained a lot of weak reflections, especially at the high resolution. The data above 3.5 Å resolution seemed to be unreliable.

B-3-2. Molecular replacement

The orientations and positions of two molecules in the asymmetric unit were found with the X-PLOR package. The structure of the orthorhombic RC was used as the search model (rcw241.pdb). The structure was refined to an *R* factor of 22.3% to 3.0 Å resolution. The average r.m.s. deviations for the bond lengths and angles are 0.013 Å and 3.41°, respectively. The RF and TF searches were done with the same parameters described in section B-2-3. Table B-9 lists results of RF searches for various resolution ranges. The highest peaks and the second highest peaks were listed. Two distinct RF solutions and their TF solutions were determined unambiguously. The relative position of two TF solutions were determined with rigid body refinement (Table B-10). There were eight possible relative positions between the two TF solutions, due to the origin ambiguity. One position was found to have an *R* factor of 38%, which was about 8% lower than *R* factors of other positions.

The final *R* factor after rigid body refinement was 34.3% for the resolution range of 15 - 3 Å. The MR solutions are a rotation of θ_1= 113.97°, θ_2= 90.53°, θ_3= 89.21° and a translation of *x*= 64.05 Å, *y*= 55.91 Å, *z*= -12.35 Å for the first molecule and a rotation of θ_1= 236.23°, θ_2= 90.08°, θ_3= 90.56° and a translation of *x*= 37.76 Å, *y*= 55.26 Å, *z*= 41.80 Å for the second molecule. The rotations are given in the Euler angles and the translations are given in the orthogonal coordinates.

B-3-3. Averaging

Averaging *(34)* between one RC molecule in the orthorhombic crystal form and two RC molecules in the tetragonal crystal form was performed with a local program. Two nearest RC molecules in the tetragonal cell are selected for the averaging. One angstrom grid maps were calculated for the orthorhombic and tetragonal forms. Two molecules in the tetragonal form were averaged first and followed by the averaging between two crystal forms. New maps were calculated by combining the MR phases and the averaged phases

and the process was repeated. Transformations among molecules and the result of threefold averaging are given in Table B-12. The average phase changes of the orthorhombic form and tetragonal form were 40.52° and 50.55°, respectively.

X-PLOR refinement was performed with two independent RC molecules in the asymmetric unit.. Refinement included slowcooling from 3000 °K to 300 °K, with 25 °K steps and individual B-factor refinement. The *R* factor decreased from 33.9% to 19.7%. The average r.m.s difference between the Cα atoms of two molecules is 1.04 Å. The most variable regions in the structure were residues M99 to M108 and H40 to H62. With the 2σ cutoff on the observed structure factors, the number of reflections was 36518 and the number of degrees of freedom was 48186. Since the structure was seriously underdetermined with the present data set, more complete data set is needed for the further structure determination.

Table B-1. Data Statistics of the cocrystal

(a) overall statistics

Data set	coxtal1	coxtal2	coxtal3
cell constants (Å)	a=b=142.71	141.90	141.8
	c=254.77	252.67	252.7
detector	R-AXIS	Synchrotron	Siemens
data processing	R-AXIS	MOSFLM	XDS
number of reflections	13500	12683	5965
resolution limits (Å)	4.5	4.0	6.0
completeness (%)	82.5	56.2	86.3
R_{merge} (%)	~12	22.6	9.21

(b) $\langle I \rangle / \langle \sigma \rangle$ in each resolution shell for the 'coxtal1' data

Res.	15.0	10.0	9.0	8.0	7.0	6.0	5.0	4.5
$\langle I \rangle / \langle \sigma \rangle$	13.61	5.38	5.16	3.58	2.11	1.94	1.94	1.89

(c) $\langle I \rangle / \langle \sigma \rangle$ and R_{merge} in each resolution shell for the 'coxtal2' data

Res.	12.30	8.83	7.25	6.29	5.64	5.15	4.77	4.47	4.21	4.00
$\langle I \rangle / \langle \sigma \rangle$	11.76	9.69	6.10	4.92	3.48	2.39	2.05	2.05	1.81	1.65
R(%)*	5.1	6.4	10.7	14.6	21.1	28.1	29.5	3.2	37.2	40.7

(d) R_{merge} in each resolution shell for the 'coxtal3' data

Res.	20.41	14.43	11.79	10.21	9.13	8.33	7.72	7.22	6.80	6.45	6.15	5.89
R	4.42	5.13	6.45	6.65	7.88	9.64	12.59	16.24	18.88	21.41	24.20	25.58

Table B-2. MERLOT RF search of the cocrystal. Res. is the resolution range, arad. is the radius of integration and RMS is the r.m.s. deviation.

(a) 'Coxtal1' data set

res.	arad.	α	β	γ	RMS
10-7	25.	0.0	57.5	30.0	3.22
	30.	86.25	85.0	45.0	3.42
	35.	88.75	85.0	45.0	3.54
	40.	0.0	87.5	45.0	4.07
10-8	25.	7.5	85.0	220.0	3.54
	30.	78.75	82.5	45.0	3.77
	35.	77.5	82.5	45.0	4.03
	40.	78.75	82.5	45.0	4.28
	45.	78.75	82.5	45.0	4.33
15-6	25.	86.25	87.5	45.0	3.20
	30.	85.0	85.0	45.0	3.46
	35.	88.75	87.5	50.0	3.88
15-7	25.	85.0	87.5	45.0	4.56
	30.	85.0	85.0	45.0	4.72
	35.	87.5	87.5	50.0	4.95
	40.	88.75	87.5	45.0	5.56
15-8	25.	82.5	87.5	45.0	4.49
	30.	81.25	87.5	45.0	4.59
	35.	80.0	87.5	45.0	4.78
	40.	83.75	90.0	50.0	5.14
	45.	2.5	90.0	225.0	5.32

(b) 'Coxtal2' data set

res.	arad.	α	β	γ	RMS
10-5	25.	88.75	85.0	15.0	2.33
	29.	88.75	85.0	15.0	2.37
10-6	20.	3.75	77.5	215.0	2.01
	25.	1.25	77.5	215.0	2.18
	30.	0.0	77.5	215.0	2.22
	35.	0.0	77.5	215.0	2.49
10-7	25.	50.0	40.0	285.0	2.35
	30.	47.5	45.0	285.0	2.35
	35.	46.25	47.5	290.0	2.55
	40.	45.0	47.5	290.0	2.69
15-5	25.	33.75	55.0	305.0	1.98
	29.	37.5	55.0	310.0	2.05
15-6	25.	38.75	50.0	300.0	2.37
	30.	30.0	50.0	305.0	2.43
	35.	33.75	50.0	305.0	2.50
15-7	25.	35.0	50.0	300.0	2.58
	30.	41.25	47.5	300.0	2.59
	35.	35.0	50.0	305.0	2.70
	40.	37.5	50.0	300.0	2.83
15-8	25.	37.5	47.5	300.0	2.75
	30.	37.5	47.5	305.0	2.77
	35.	36.25	47.5	300.0	2.91
	40.	37.5	50.0	300.0	2.98
	45.	27.5	50.0	300.0	3.04

(c) 'Coxtal3' data set

res.	arad.	α	β	γ	RMS
10-6	30.	0.0	87.5	225.0	4.32
	35.	0.0	87.5	225.0	4.89
10-7	30.	2.5	80.0	210.0	3.96
	35.	0.0	87.5	225.0	4.38
	40.	88.75	87.5	225.0	5.01
10-8	30.	7.5	82.5	225.0	3.79
	35.	15.0	77.5	225.0	4.30
	40.	15.0	77.5	230.0	4.54
	45.	15.0	80.0	225.0	4.67
15-6	30.	1.25	90.0	230.0	4.73
	35.	0.0	90.0	225.0	5.28
15-7	30.	88.75	90.0	50.0	4.60
	35.	1.25	90.0	225.0	5.12
	40.	0.0	90.0	225.0	6.11
15-8	30.	80.0	82.5	45.0	4.30
	35.	86.25	87.5	45.0	4.66
	40.	1.25	90.0	225.0	5.43
	45.	1.25	90.0	225.0	5.59

Table B-3. X-PLOR RF search of the cocrystal. Corr. is the correlation coefficient between the squared model and observed structure factors.

(a) 'Coxtal1' data set

res. (Å)	θ_1	θ_2	θ_3	Corr.
10-7	177.30	46.84	92.00	0.05572
	177.08	59.62	1.41	0.05034
10-8	41.54	90.06	359.61	0.06506
	266.74	68.50	358.19	0.05572
15-6	178.99	26.46	10.22	0.05989
	213.20	48.42	12.82	0.05527
	42.67	90.27	359.23	0.05025
15-7	223.24	88.79	359.69	0.08467
	43.25	91.21	0.33	0.08466
	231.80	61.46	0.73	0.05875
15-8	44.73	89.73	0.86	0.07522
	176.71	55.34	0.71	0.03556

(b) 'Coxtal2' data set. The correct RF solution was added to the list of possible solutions before PC refinement in (*).

res.(Å)	θ_1	θ_2	θ_3	Corr.
10-5	44.26	87.82	0.40	0.10817
	224.93	89.98	359.80	0.10344
	238.45	59.20	86.52	0.05822
10-6	44.39	91.18	359.06	0.11754*
	316.21	35.71	137.55	0.05088
10-7	44.96	91.01	0.77	0.12208*
	7.92	47.83	76.99	0.07305
15-5	44.61	93.24	0.17	0.08383
	344.68	36.70	2.52	0.03256
15-6	44.37	91.07	359.06	0.10965

	59.96	89.83	1.18	0.05061
15-7	42.89	90.56	0.84	0.10210
	43.38	87.79	65.11	0.07012
15-8	68.32	87.15	0.39	0.06336
	57.92	90.33	0.03	0.05422

(c) 'Coxtal3' data set

res. (Å)	θ_1	θ_2	θ_3	Corr.
8-6	42.54	90.62	359.91	0.09433
	272.31	39.60	2.09	0.04088
10-6	43.44	88.92	90.10	0.07807
	272.79	39.35	1.65	0.03891
10-7	268.84	90.65	25.38	0.05057
	281.40	37.23	355.86	0.04806
15-6	223.10	89.62	0.05	0.08265
	43.48	89.67	0.04	0.08225
	38.23	88.34	91.03	0.05527
15-7	223.80	89.19	88.86	0.08784
	43.89	90.89	91.19	0.08782
	238.08	68.40	88.05	0.05042

Table B-4. The cross-rotation function between the observed structure factors.

(a) 'Rc19' and rctet' data sets

res.	arad.	α	β	γ	RMS
50-6	30.	0.0	90.0	152.5	3.70
	35.	0.0	90.0	155.0	4.03
50-7	30.	0.0	90.0	27.5	2.78
	35.	0.0	90.0	67.5	2.85
	40.	0.0	90.0	67.5	3.27

(b) 'Coxtal1' and 'rc19' data sets

res.	arad.	α	β	γ	RMS
50-6	30.	0.0	10.0	0.0	1.76
	35.	0.0	10.0	0.0	2.00
50-7	30.	0.0	40.0	90.0	1.99
	35.	52.5	12.5	35.0	2.30
	40.	0.0	90.0	132.5	2.55

(c) 'Coxtal1' and 'rctet' data sets

res.	arad.	α	β	γ	RMS
50-6	30.	57.5	0.0	32.5	2.54
	35.	52.5	0.0	37.5	3.06
50-7	30.	15.0	0.0	75.0	2.76
	35.	8.75	0.0	81.25	3.19
	40.	46.25	0.0	43.75	3.54

(d) 'Coxtal3' and 'rc19' data sets

res.	arad.	α	β	γ	RMS
50-6	30.	0.0	90.0	47.5	3.
	35.	0.0	90.	47.5	3.61
50-7	30.	0.0	90.0	45.0	2.64
	35.	0.0	90.0	47.5	2.80
	40.	0.0	90.0	132.5	3.44

(e) 'Coxtal1' and 'retet' data sets

res.	arad.	α	β	γ	RMS
50-6	30.	0.0	10.0	0.0	2.57
	35.	75.0	0.0	12.5	2.77
50-7	30.	0.0	12.5	00	2.56
	35.	88.75	0.0	7.5	2.74
	40.	0.0	7.5	0.0	3.04

Table B-5. TF search of the cocrystal

(a) GENTF with the 'coxtal1' data set. The RF solution of (43.25, 91.21, 0.33) and the resolution range of 15 to 7 Å were used. The TF is give in the orthogonal coordinates. Corr. is the correlation coefficient between the model and observed structure factors.

space group	x	y	z	Corr.	R
$P4_12_12$	54.96	21.52	101.97	0.3985	42.574
$P4_32_12$	54.77	21.63	102.03	0.3782	43.144

(b) XPLOR TF search with the 'coxtal1' data set. The RF solution of (43.25, 91.21, 0.33) and the resolution range of 15 to 7 Å were used. Corr. are the correlation coeffcient between the squared model and observed structure factors.

	x	y	z	corr.	R	packing
$P4_12_12$	54.84	21.61	38.11	0.2998	44.98	0.1456
$P4_22_12$	54.74	20.27	38.10	0.1776	47.64	0.1443
$P4_32_12$	54.84	21.68	101.85	0.4832	40.24	0.1459

(c) XPLOR TF search with the 'coxtal2' data set. The RF solution of (42.89, 90.56, 0.84) and the resolution range of 15 to 5 Å were used.

	x	y	z	corr	R	packing
$P4_32_12$	52.71	22.30	100.27	0.3799	46.49	0.1493

(d) XPLOR TF search with the 'coxtal3' data set. The RF solution of (43.47, 90.24, 0.26) and the resolution range of 15 to 6 Å were used.

	x	y	z	corr	R	Packing
$P4_12_12$	55.04	20.39	36.98	0.2344	47.20	0.1462
$P4_22_12$	54.74	20.27	36.10	0.2032	48.18	0.1450
$P4_32_12$	54.69	20.26	99.86	0.4612	41.68	0.1492

Table B-6. Summary of MR of the cocrystal. The rotation function is given in Euler angles and the translation function is given in the orthogonal coordinates. The *R* factor is calculated after rigid body refinement at the given resolution range.

data set	rotation function			translation function			Rfactor
	θ_1	θ_2	θ_3	x	y	z	
coxtal1	43.51	90.34	0.00	55.20	21.56	100.66	40.98 (15-5 Å)
coxtal2	43.32	89.87	0.20	54.44	21.20	99.45	36.61 (15-5 Å)
coxtal3	43.59	90.20	0.32	54.25	21.30	99.71	36.29 (15-6 Å)

Table B-7. Solvent flattening of the cocrystal

	initial	final	initial	final
solvent content	0.5		0.6	
R factor	.274	.180	.292	.218
correlation coefficient	.950	.978	.941	.969
figure of merit	.809	.869	.803	.849
mean phase change	26.31		26.36	

Table B-8. Data statistics of the tetragonal RC

res.	number of reflections	completeness (%)	<I>/<σ>	R factor
8.13	2873	91.3	16.17	0.073
6.46	2772	94.5	18.37	0.084
5.64	2756	94.9	14.22	0.108
5.13	2759	95.5	13.68	0.124
4.76	2741	95.7	14.12	0.127
4.48	2714	94.8	12.80	0.145
4.26	2694	95.4	10.51	0.170
4.07	2715	95.2	8.18	0.211
3.91	2678	95.0	6.30	0.276
3.78	2671	94.3	5.13	0.358
3.66	2643	93.7	3.81	0.482
3.56	2646	94.5	2.68	0.660
3.46	2630	93.5	2.27	0.784
3.38	2592	92.5	1.98	0.955
3.30	2616	93.2	1.54	0.000
3.23	2608	93.3	1.14	0.000
3.17	2589	92.2	0.88	0.000
3.11	2574	91.9	0.75	0.000
3.05	2593	92.3	0.62	0.000
3.00	2516	91.0	0.58	0.000
Total	53380	93.7	6.66	0.157

Table B-9. X-PLOR RF search of the tetragonal RC

res. (Å)	θ_1	θ_2	θ_3	Corr.
15-5	293.84	89.58	90.82	0.08140
	235.43	89.06	0.23	0.06404
	236.60	78.91	94.68	0.04039
15-6	113.14	90.42	359.62	0.06843
	234.64	88.79	0.36	0.05992
	236.78	76.78	5.16	0.04407
15-7	292.76	88.86	359.68	0.06408
	112.07	90.15	89.26	0.06304
	216.88	83.68	13.04	0.04399

Table B-10. X-PLOR TF search of the tetragonal RC

(a) One-molecule TF search. The resolution range of 15 to 6 Å was used. The RF solutions were (113.84, 90.43, 90.82) for the first molecule (mol1) and (235.43, 89.06, 90.23) for the second molecule (mol2), in the Euler angles. The TF solution was given in the orthogonal coordinates.

	x	y	z	R factor
mol1	8.08	58.56	126.07	53.15
mol2	38.36	54.57	40.67	51.98

(b) Two-molecule TF search. 40 steps of rigid body refinement were done with the resolution range of 15 to 6 Å.

relative translation of mol1	initial R factor	final R factor
(0.0,0.0,0.0)	52.9	51.6
(0.5,0.0,0.0)	50.2	47.6
(0.0,0.5,0.0)	50.2	47.7
(0.5,0.5,0.0)	52.5	51.4
(0.0,0.0,0.5)	52.5	50.8
(0.5,0.0,0.5)	43.6	**38.4**
(0.0,0.5,0.5)	50.0	46.9
(0.5,0.5,0.5)	52.8	51.3

Table B-11. Averaging between the orthorhombic RC and the tetragonal RC

(a) Transformations used in averaging

from	to	rotation			translation		
rcw241	mol1	113.967	90.467	89.265	-64.117	55.945	-12.216
rcw241	mol2	56.112	89.968	89.371	-32.915	156.685	27.540
mol1	mol2	180.290	57.877	180.158	30.888	116.231	-13.607

(b) Equivalent positions used in averaging

	orthogonal coordinates			fractional coordinates		
rcw241	37.004	29.601	68.094	0.268	0.382	0.480
mol1	4.638	44.868	32.851	0.033	0.317	0.119
mol2	35.664	112.235	41.792	0.252	0.794	0.151

(c) Results of the threefold averaging

	orthogonal		tetragonal	
	initial	final	initial	final
R factors	22.1	13.9	32.9	23.0
Correlation coefficients	.715	.956	.663	.905
mean phase change	40.52		50.55	

Figure B-1. Cα atoms of the RC and cyt c_2. The cofactors and the heme group of the cyt c_2 in the proposed position are also shown.

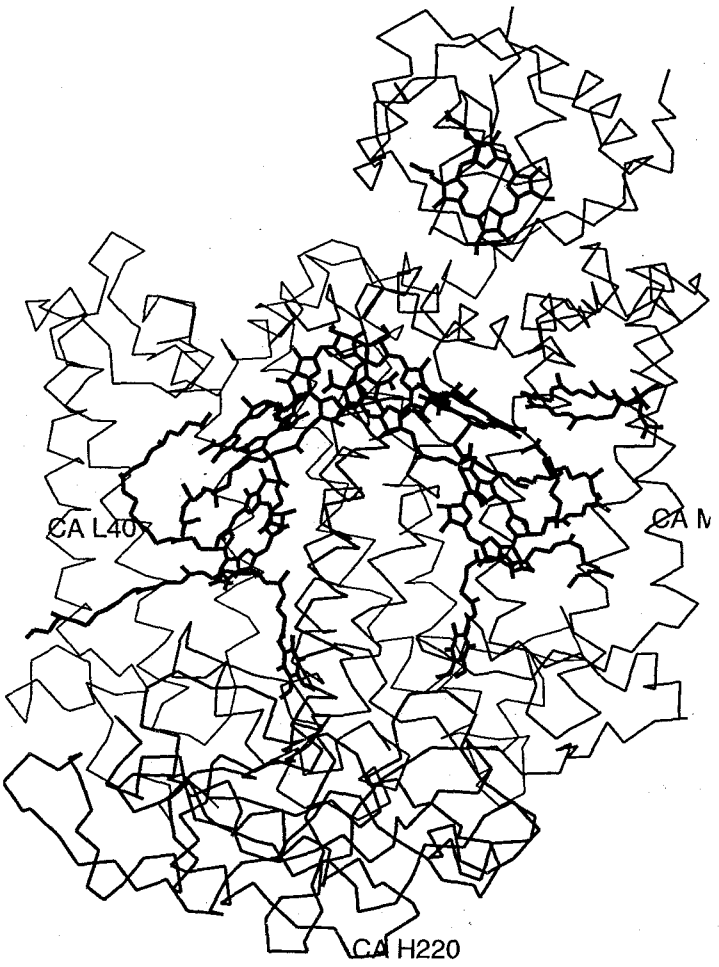
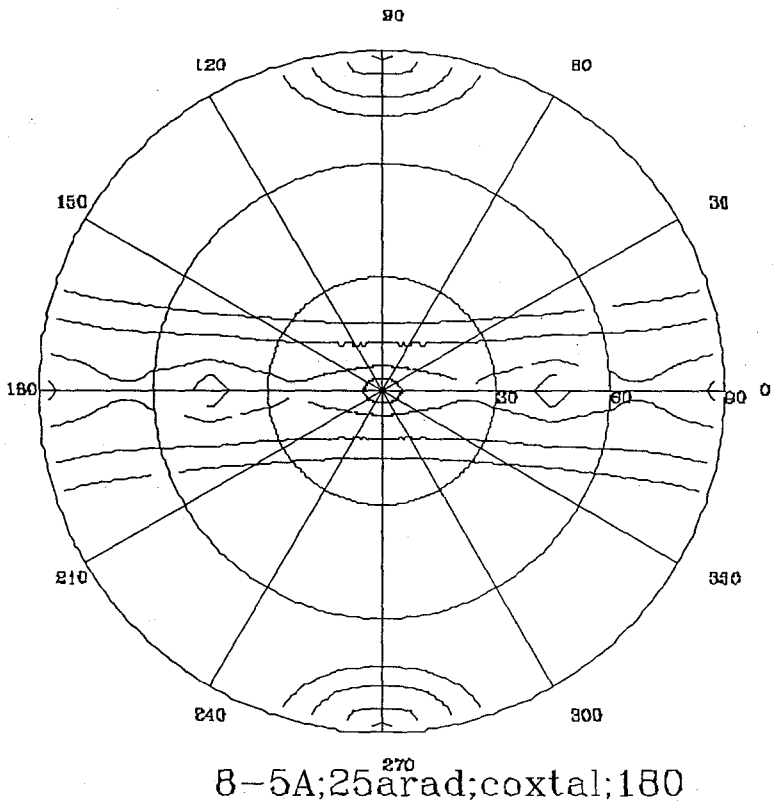


Figure B-2. Self-rotation function

(a) Crowther's self-rotation function. 'Coxtal1' data set and a resolution range of 8 to 5 Å were used for the calculation. A κ=180° section is shown.

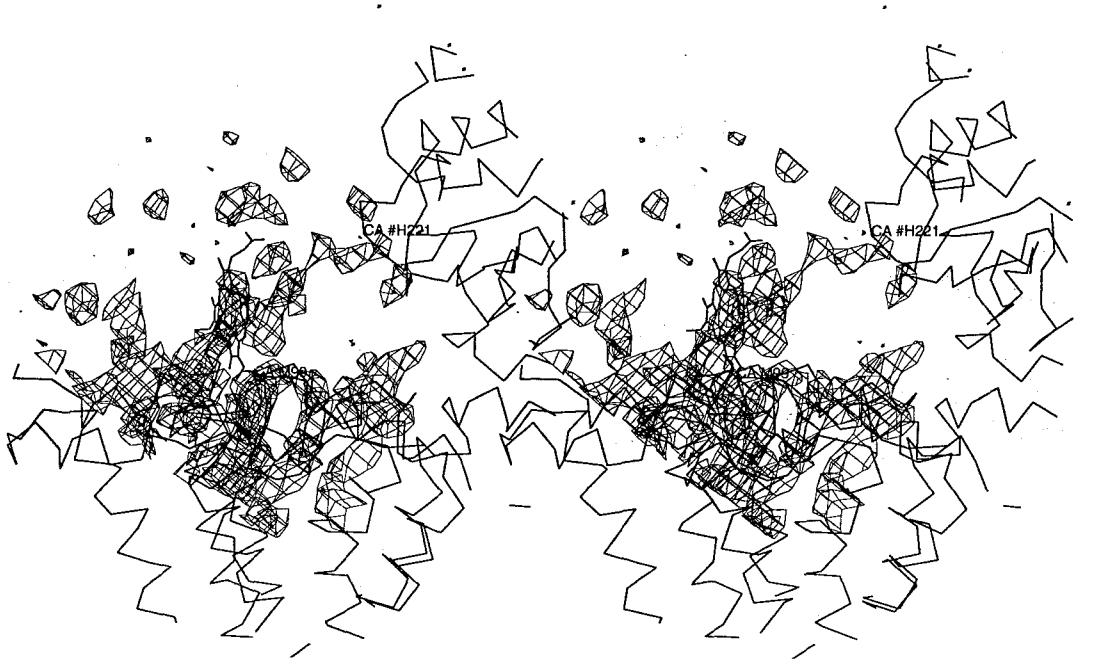


(b) X-PLOR self-rotation function. 'Coxtal1' data set and a resolution range of 8 Å to 5 Å were used for the calculation. RF is the value of the rotation function.

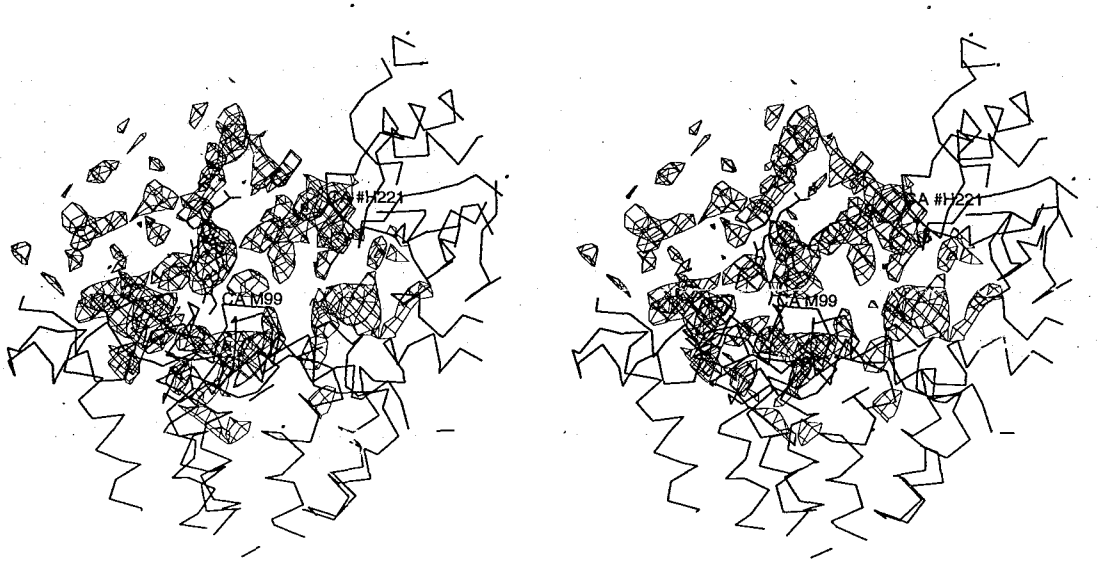
Peak Order	θ_1	θ_2	θ_3	RF
1	0.	0.	0.	11.2041
3475	12.0	0.	0.	6.299
4554	42.0	9.	42.0	5.839

Figure B-3. The electron density maps after rigid body refinement around the proposed cyt c_2 binding position. The maps are contoured at 1.5σ level.

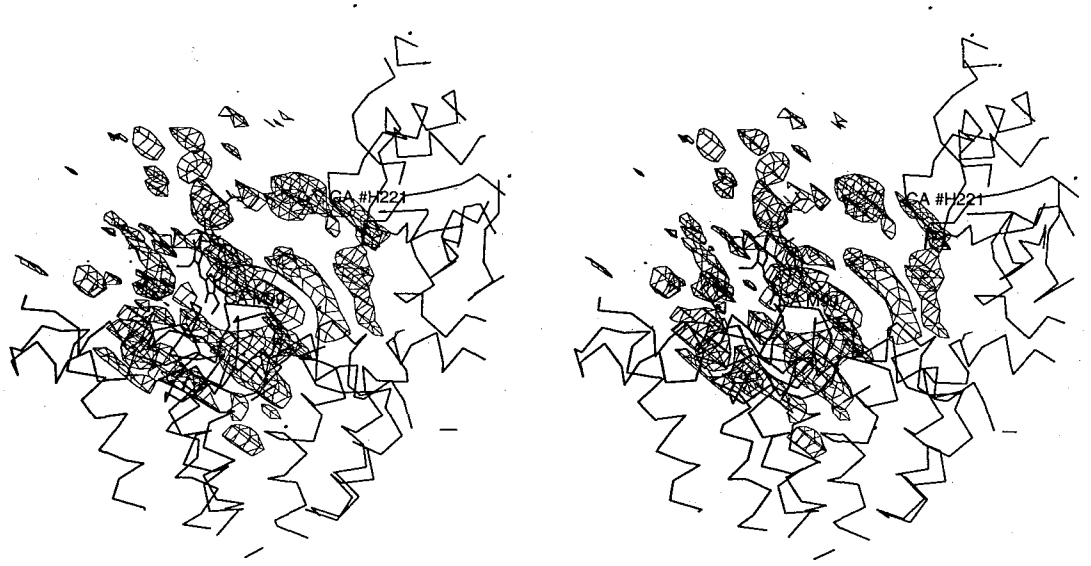
(a) The (2|Fo|-|Fc|) map of the 'coxtal1' data set



(b) The (|Fo|-|Fc|) map of the 'coxtal1' data set



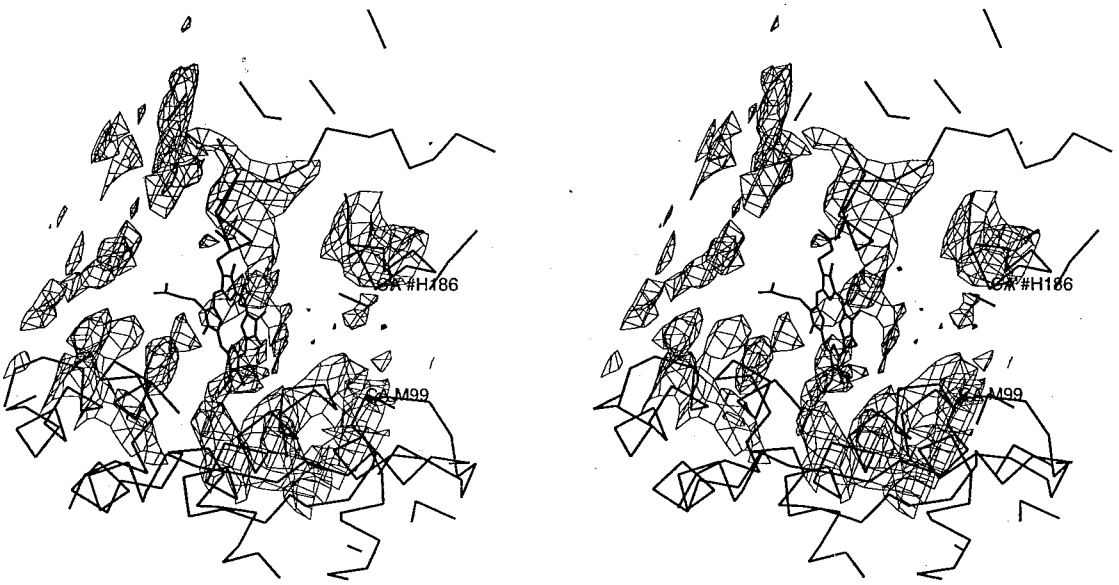
(c) The (|Fo|-|Fc|) map of the 'coxtal2' data set



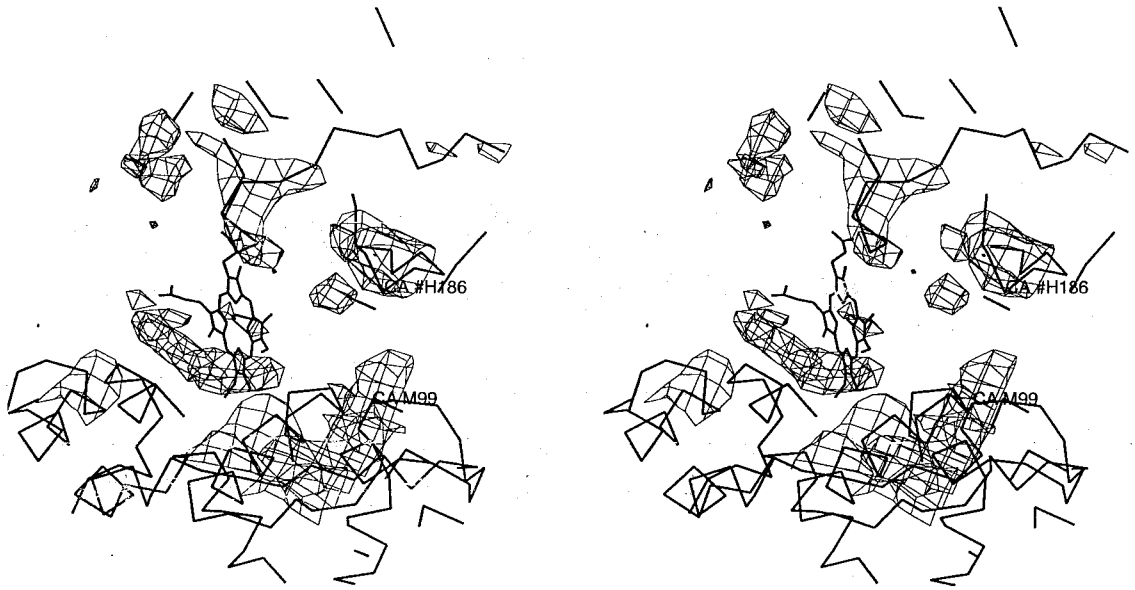
(d) The (|Fo|-|Fc|) map of the 'coxtal3' data set

Figure B-4. The solvent-flattened maps. The maps are contoured at 1σ level.

(a) The solvent content of 60% and the 'coxtal1' data set



(b) The solvent content of 60% and the 'coxtal3' data set



References

(1) J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, *Nature*, **318**, 618 (1985)

(2) J. Deisenhofer, H. Michel, *EMBO J.*, **8**, 2149 (1989)

(3) J. P. Allen, G. Feher, T. O. Yeates, H. Komiya, D. C. Rees, *Proc. Natl. Acad. Sci. USA*, **84**, 5730 (1987)

(4) J. P. Allen, G. Feher, T. O. Yeates, H. Komiya, D. C. Rees, *Proc. Natl. Acad. Sci. USA*, **84**, 6162 (1987)

(5) T. O. Yeates, H. Komiya, A. Chirino, D. C. Rees, J. P. Allen, G. Feher, *Proc. Natl. Acad. Sci. USA*, **85**, 7993 (1988)

(6) J. P. Allen, G. Feher, T. O. Yeates, H. Komiya, D. C. Rees, *Proc. Natl. Acad. Sci. USA*, **85**, 8487 (1988)

(7) C. H. Chang, D. Tiede, J. Tang, U. Smith, J. Norris, *FEBS Lett.*, **205**, 82 (1986)

(8) C.-H. Chang, O. El-Kabbani, D. Tiede, J. Norris, M. Schiffer, *Biochemistry*, **30**, 5352 (1991)

(9) U. Erlmer, G. Fritzsch, S. K. Buchanan, H. Michel, *Structure*, **2**, 925 (1994)

(10) J. C. Williams, L. A. Steiner, R. C. Ogden, M. I. Simon, G. Feher, *Proc. Natl. Acad. Sci. USA*, **80**, 6505 (1983)

(11) J. C. Williams, L. A. Steiner, G. Feher, M. I. Simon, *Proc. Natl. Acad. Sci. USA*, **81**, 7303 (1986)

(12) J. C. Williams, L. A. Steiner, G. Feher, *Proteins*, **1**, 312 (1986)

(13) G. Feher, J. P. Allen, M. Y. Okamura, D.C. Rees, *Nature*, **339**, 111 (1989)

(14) M. Y. Okamura and G, Feher, *Ann. Rev. Biochem.*, **61**, 861 (1992)

(15) H. Komiya, T. O. Yeates, D. C. Rees, J. P. Allen, G. Feher, *Proc. Natl. Acad. Sci. USA*, **85**, 9012 (1988)

(16) H. L. Axelrod, G. Feher, J. P. Allen, A. J. Chirino, M. W. Day, B. T. Hsu, D. C. Rees, *Acta Cryst. D*, **50**, 596 (1994)

(17) D. M. Tiede and C.-H. Chang, *Isr. J. Chem.*, **28**, 183 (1988)

(18) J. P. Allen and G. Feher, *Proc. Natl. Acad. Sci. USA*, **81**, 4795 (1984)

(19) C.-H. Chang, M. Schiffer, D. Tiede, U. Smith, J. Norris, *J. Mol. Biol.*, **186**, 201 (1985)

(20) A. Ducruix and F. Reiss-Husson, *J. Mol. Biol.*, **193**, 419 (1987)

(21) H. A. Frank, S. S. Taremi, J. R. Knox, *J. Mol. Biol.*, **198**, 139 (1987)

(22) S. K. Buchanan, G. Fritzschm U, Ermler, H. Michel, *J. Mol. Biol.*, **230**, 1311 (1993)

(23) A.G. W. Leslie, in *Joint CCP4 and ESF-EACMB Newletter on Protein Crystallography*, no. 26, (Daresbury Laboratory, Warrington, United Kingdom, 1992)

(24) W. Kabsch, *Acta Cryst.*, **21**, 916 (1988)

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

(26) A. T. Brunger, *Acta Cryst.*, **A46**, 46 (1990)

(27) R. A. Crowther, in *Molecular Relacement Method*, edited by M. G. Rossmann, p173, New York: Gordon and Breach (1972)

(28) P. M. D. Fitzgerald, *J. Appl. Cryst.*, **21**, 273 (1988)

(29) S. N. Rao, J.-H. Jih. J. A. Hartsuck, *Acta Cryst.*, **A36**, 878 (1980)

(30) E. Lattman and W. E. Love, *Acta. Cryst.*, **B26**, 1854 (1970)

(31) M. Fujinaga and R. J. Read, *J. Appl. Cryst.*, **20**, 517 (1987)

(32) B. C. Wang, *Methods Enzymol.*, **115**, 41 (1985)

(33) Z. Otwinowski, *Denzo: a film processing program for macromolecular crystallography*, Yale University, New Haven (1991)

(34) G. Bricogne, *Acta Cryst.*, **32**, 832 (1976)