INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

X-RAY CRYSTALLOGRAPHIC ANALYSIS OF A KANAMYCIN KINASE

CRYSTALLOGRAPHIC STUDIES ON TWO STRUCTURES OF A KANAMYCIN KINASE: A Mg-AMPPNP AND A Mg-ADP COMPLEX

By

WAI-CHING HON, B. Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Doctor of Philosophy (1998)	McMaster University
(Biochemistry)	Hamilton, Ontario

TITLE: Crystallographic studies on two structures of a kanamycin kinase: a Mg-AMPPNP and a Mg-ADP complex.

AUTHOR: Wai-Ching Hon, B. Sc. (University of Waterloo)

SUPERVISOR: Dr. Daniel S. C. Yang

NUMBER OF PAGES: *xv*, 139

Abstract

3',5"-aminoglycoside phosphotransferase type IIIa (APH(3')-IIIa) belongs to a family of bacterial enzymes that phosphorylate the aminoglycoside antibiotics. The modified antibiotics are rendered ineffective due to a lowered affinity for their targets in the bacterial cells. APH(3')-IIIa phosphorylates regiospecifically the 3'- and/or the 5"-hydroxyl groups of nine aminoglycosides. To gain insight into the chemical mechanisms involved in phosphate transfer, the crystal structures of APH(3')-IIIa complexed with a Mg-ATP analog and Mg-ADP have been determined to 2.4 Å and 2.2 Å resolution, respectively. These complexes represent the first and the last step in the enzyme kinetic reaction pathway and their structures are almost identical. Half of the protein molecule shows striking structural resemblance to the catalytic cores of protein kinases, which share less than 6% amino acid sequence identity with APH(3')-IIIa. The folding of this structurally conserved region forms the nucleotide-binding pocket in which five of the seven invariant residues form an elaborate phosphate recognition network via two Mg^{2+} ions. In the Mg-AMPPNP complex, the coordination spheres surrounding the two Mg^{2+} ions are far from the preferred octahedral geometry, which is observed in the Mg-ADP complex. The geometry of the Mg^{2+} ion that binds to the carboxylate oxygens of an invariant Asp is particularly distorted. This Mg^{2+} ion may play a role in triggering or controlling the phosphoryl transfer reaction. This hypothesis would likely have extended implications in understanding the details of the catalytic mechanisms of protein kinases. A potential antibiotic-binding site on APH(3')-IIIa is also immediately obvious. A cluster of residues conserved among the APH(3') isozymes line a partially solvent exposed pocket

that displays a net electronegative surface potential. This binding pocket satisfies the requirements for binding to a broad spectrum of cationic aminoglycoside substrates.

Acknowledgements

Many sincere thanks go to my supervisor, Dr. S. C. Yang, for his scientific guidance and his generous spiritual support. I would especially like to express my appreciation for his provision of an environment for growth and exploration. During my graduate studies, I participated in two distinct collaborative research projects. The work in the first project, the isolation and characterization of a group of antifreeze proteins from cold-acclimated winter rye, is not described in this thesis. I would like to thank Dr. Marilyn Griffith (University of Waterloo, Canada) for providing me the opportunity to first test my potential as a scientist. The second project is the core of this thesis and originated from Dr. Gerard D. Wright's laboratory. For this, I would like to thank Gerry for the opportunity to track a different scientific terrain. Thanks are also extended to Dr. Albert M. Berghuis who co-supervised part of the x-ray crystallographic work, and other members of my supervisory committees, including Dr. Richard Rachubinski, Dr. David Andrews and Dr. Aled Edwards.

My work would not have been as enjoyable without the friendship and help of my labmates – past and present, immediate and extended. In an approximate chronological order of appearance, they are Ponita, Paul Ala, Frank Sicheri; Ernest, Susan Xue, Andrzej, Steve Bubanko, Jing, Roy Satmaka, Adrian, Laura, Adelaine and Quyen Hoang. Special thanks go to Mike, Jason and Mervi for sharing their enthusiasm about science.

Last but not least, my work would not have been possible without my parents, who not only gave me life, but also have been keeping me alive. Dad's soups and early chauffeuring service, and mom's junk food supplies and pocket money deserve mentioning.

Table of contents

List of figures

List of tables

List of abbreviations

AcSCoA	acetyl coenzyme A
ADP	adenosine 5'-diphosphate
AMPPNP	β,γ-imidoadenosine 5'-trophosphate
APH	aminoglycoside phosphotransferase
APH(3')	3'-aminoglycoside phosphotransferase
APH(3')-IIIa	3',5"-aminoglycoside phosphotransferase type IIIa
ATP	adenosine 5'-triphosphate
ATPase	adenosine 5'-triphosphate hydrolase
B-factor	temperature factor
cAPK	cAMP-dependent protein kinase
EDTA	ethylenediaminetetraacetic acid
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
IPTG	isopropyl β-D-thiogalactopyranoside
ISAS	iterative single-wavelength anomalous scattering
MAD	multi-wavelength anomalous diffraction
MIR	multiple isomorphous replacement
MIRAS	multiple isomorphous replacement complemented with anomalous scattering
NCS	noncrystallographic symmetry
NMR	nuclear magnetic resonance

OD	optical density
PDB	Brookhaven Protein Data Bank
PEG	polyethylene glycol
Phkγ	γ-subunit of glycogen phosphorylase kinase
PKI	peptide inhibitor (residues 5-22, Ser21→Ala) for cAPK
PMSF	phenylmethylsulfonylfluoride
r.m.s.d.	root mean squared deviation
rRNA	ribosomal ribonucleic acid
SAS	single-wavelength anomalous scattering
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
Se-Met	selenomethionine
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement complemented with anomalous scattering

Chapter 1

General Introduction

The antibiotic era

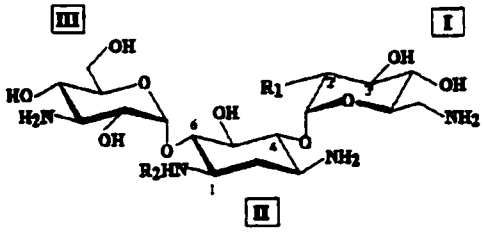
Antibiotics are antimicrobial agents secreted by certain microorganisms, presumably to gain ecological advantages in the natural environment. The era of antibiotics as "wonder drugs" began when Alexander Fleming noted that his culture of *Staphylococcus aureus* was killed by a contaminating mold, *Penicillium notatum* in 1929 (Atlas, 1988). The therapeutic value of penicillin was established after its isolation a decade later. Soon afterwards, in the early 1940s, Selman Waksman and co-workers found that several soil bacteria belonging to the *Actinomycetes* group produce antibiotics that have a broader spectrum of action. The best-known among this new generation of drugs is streptomycin, which is produced by *Streptomyces griseus* and is a member of the aminocyclitol-aminoglycoside group of antibiotics (hereafter referred to as aminoglycosides).

Aminoglycosides

Aminoglycosides are mostly used in the treatment of bacterial infections caused by susceptible aerobic Gram-negative and Gram-positive bacteria, including the opportunistic pathogens *Staphylococci* and *Enterococci*. They constitute a large group of compounds, both natural and semi-synthetic, that have aminosugars that are linked to one or two cyclic moieties through glycosidic bonds. They can be divided into three structural types, the 4,5- and the 4,6-disubstituted 2-deoxystreptamines containing aminoglycosides, and others (Umezawa & Kondo, 1982). A subset of the former two types of aminoglycosides is

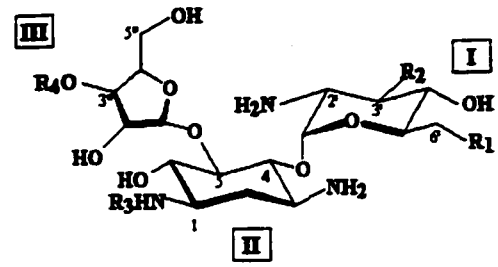
relevant to this study and their chemical structures are illustrated in figure 1-1, which also shows the numbering scheme of the cyclic moieties and the substituents on these rings. These compounds all contain a central 2-deoxystreptamine to which is attached two aminosugars and the total number of rings vary from three (as in kanamycin) to five (as in lividomycin A). The rings are highly substituted with polar groups, the majority of which are amino- and hydroxyl groups. The variation of the substituents mainly occurs on rings I and III. The structure of the central 2-deoxystreptamine ring stays constant except for amikacin (semi-synthetic) and butirosin (natural) which have the bulky substituent, [NH-CO-CH(OH)-CH_2-CH_2-NH_2], on C1.

Aminoglycosides interfere with protein synthesis by binding tightly to the 30S bacterial ribosomal subunits, and some have been shown to bind specifically to the acceptor-site (A-site) of the 16S rRNA (Fourmy, Recht, Blanchard & Puglisi, 1996). The accumulation of defectively translated proteins is generally believed to be the primary cause of cell death (Davis, 1987). The bactericidal action of aminoglycosides makes them unique among antibiotics that inhibit protein synthesis, which are mostly bacteriostatic (Levison, 1995). As drugs, they also produce dose-dependent toxicity in human, most notably causing proximal tubular necrosis in the kidney and damage of the sensory cells of the chochlea and vestibular apparatus (Beaucaire, 1995). Despite their adverse side effects, they remain as a clinically important group of drugs because of their effectiveness against a broad spectrum of pathogens, and particularly their synergistic effect in treating certain infections caused by Enterococci when used in conjunction with penicillins.



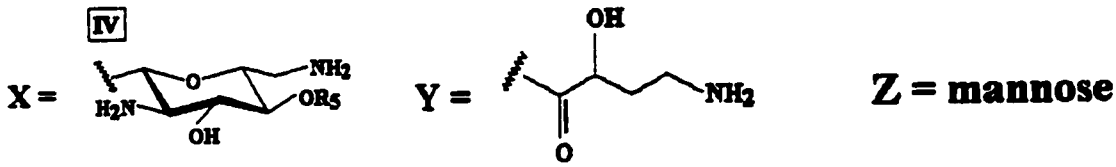
4,6-disubstituted 2-deoxystreptamine

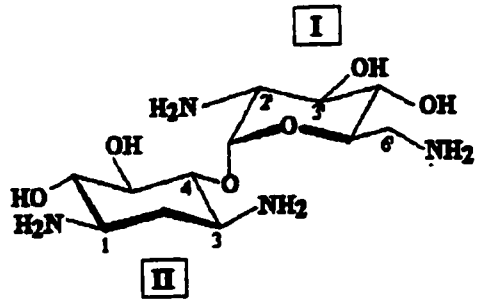
	R_1	R_2
kanamycin A	OH	H
kanamycin B	NH_2	H
amikacin	NH_2	Y



4,5-disubstituted 2-deoxystreptamine

	R_1	R_2	R_3	R_4	R_5
neomycin B	NH_2	OH	H	X	H
paromomycin	OH	OH	H	X	H
lividomycin A	OH	H	H	X	Z
ribostamycin	NH_2	OH	H	H	
butirosin	NH_2	OH	Y	H	





neamine

Figure 1-1. Chemical structures of the aminoglycoside substrates of APH(3')-IIIa. The ring numbers are indicated as Roman numerals. Neamine is a semi-synthetic compound representing the minimal substrate for APH(3')-IIIa.

Bacterial resistance to aminoglycosides

Like all other types of antibiotics that have been introduced into the clinical settings, aminoglycosides have met resistant strains of bacterial pathogens. The resistance is not only restricted to the natural forms of aminoglycosides, but also extends to the semi-synthetic forms like amikacin. The resistance mechanisms include mutation of the targets, impaired uptake of the antibiotics and the presence of plasmid-borne aminoglycoside-modifying enzymes (Courvalin & Carlier, 1981). A single base pair mutation in the 16S rRNA (Beauclerk & Cundliffe, 1987) or a single amino acid mutation in the ribosomal protein (Chang & Flaks, 1972) can result in a ribosome with lowered affinity for the antibiotics. Since aminoglycosides are highly polar molecules, uptake of these antibiotics by the bacteria requires energy. Mutations that affect the active transport system across the cell membrane decrease the accumulation of antibiotics inside the cell. The most prevalent resistance mechanism found in clinical isolates, however, is the production of inactivating enzymes. These enzymes fall into three classes, *O*-phosphotransferases (APH), *O*-nucleotidyltransferases (ANT) and *N*-acetyltransferases (AAC), which presumably inactivate the antibiotics by lowering their binding affinities to the targets (Shaw, Rather, Hare & Miller, 1993). Each class consists of families of isozymes that differ in their positions of modification on the antibiotics and in their substrate specificities. The reactions catalysed by these three classes of enzymes are illustrated below, with kanamycin as a representative aminoglycoside.

***O*-phosphotransferases**

$$kanamycin + ATP \rightleftharpoons kanamycin - P + ADP$$

***O*-nucleotidyltransferases**

$$kanamycin + ATP \rightleftharpoons kanamycin - AMP + PP_i$$

N-acetyltransferases

$$kanamycin + AcSCoA \rightleftharpoons kanamycin - Ac + HSCoA$$

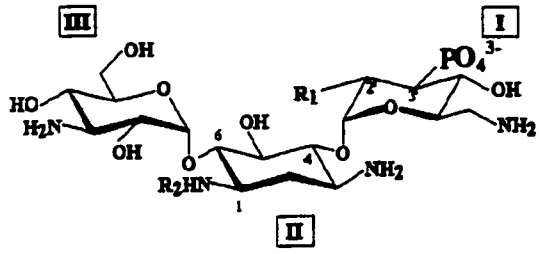
3',5"-aminoglycoside phosphotransferase type IIIa

Within the class of APH, twenty distinct types of proteins have been identified (Shaw, Rather, Hare & Miller, 1993). The APH(3') isozymes, kinases that phosphorylate the 3'-hydroxyl group of the 4,5- and 4,6-disubstituted 2-deoxystreptamines, constitute the largest family. They have different substrate spectra, which are generally fairly broad, and all of them can phosphorylate kanamycin. Hence, they are often referred to as kanamycin kinases. These isozymes share 24% of amino acid sequence identity (refer to sequence alignment in chapter 4, figure 4-1). There are other APH's that phosphorylate the 4-, 6-, or 3"-hydroxyl groups of the aminoglycosides (Shaw, Rather, Hare & Miller, 1993). The substrates of these enzymes, streptomycin (APH(3") and APH(6)) and hygromycin (APH(4)), fall into the "others" structural type of aminoglycosides. These enzymes share approximately 47% amino acid sequence similarity with the APH(3') isozymes, suggesting that they may have a similar three-dimensional structure. Based on the sequence alignment performed by Shaw *et al.* (Shaw, Rather, Hare & Miller, 1993), there are three invariant residues among all APH's, including His188, Asp190 and Asn195 (amino acid numbering according to 3',5"-aminoglycoside phosphotransferase type IIIa). As will be discussed in chapter 4, Asp190 and Asn195 form part of an evolutionary conserved Mg-ATP binding site, which is present in all APH(3') isozymes and contains Lys44, Glu60, Asp190, Asn195 and Asp208. Hence, APH's that do not belong to the APH(3') family may share

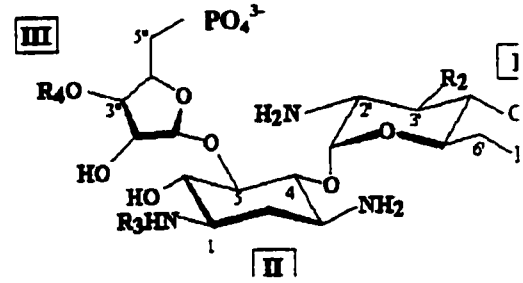
similar characteristics in the binding of Mg-ATP but may not have the same phosphoryl transfer mechanism.

One member in the APH(3') family, 3',5"-aminoglycoside phosphotransferase type IIIa (APH(3')-IIIa, a.k.a. kanamycin kinase), is the subject of this thesis work. It represents one of the best characterized member of the APH class, both in terms of enzyme kinetic mechanism (McKay, Thompson & Wright, 1994; McKay & Wright, 1995; McKay & Wright, 1996; Thompson, Hughes & Wright, 1996a) and substrate specificity and recognition (McKay, Robinson, Lane & Wright, 1994; McKay, Roestamadji, Mobashery & Wright, 1996; Thompson, Hughes & Wright, 1996b; Cox & Serpersu, 1997). Its nucleotide sequence encodes a 264-residue protein and was originally cloned from a *Streptococcus faecalis* plasmid, pJH1 (Trieu-Cuot & Courvalin, 1983). It has the broadest substrate spectrum among all the APH(3') isozymes (Shaw, Rather, Hare & Miller, 1993; McKay, Thompson & Wright, 1994). It catalyses regiospecific ATP-dependent phosphoryl transfer to at least nine different aminoglycosides (Shaw, Rather, Hare & Miller, 1993), the chemical structures of eight of these are given in figure 1-1. For the 4,5-disubstituted 2-deoxystreptamines, the site of reaction is doubly specific in that both the 3'- and 5"-hydroxyl groups can be phosphorylated (Trieu-Cuot & Courvalin, 1983; Thompson, Hughes & Wright, 1996b), and substrates that contain both reactive groups, doubly-phosphorylated products have been found *in vitro* (Thompson, Hughes & Wright, 1996b). The double specificity can be partly explained by the close spatial arrangement of the 3'- and the 5"-hydroxyl groups as revealed by the NMR structure of paromomycin complexed with the A site of 16S rRNA (Fourmy, Recht, Blanchard & Puglisi, 1996) and that of butirosin A bound to APH(3')-IIIa (Cox & Serpersu, 1997).

The chemical structures of a representative phosphorylated product from the 4,5- and the 4,6-disubstituted 2-deoxystreptamines are shown below.



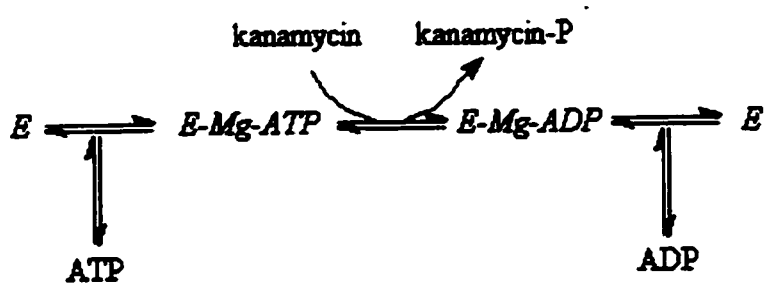
3'-phosphorylated kanamycin



5''-phosphorylated neomycin

Kinetic analyses performed with modified kanamycin A and neamine (a minimal substrate, see figure 1-1) have shown that an amino group on C-1 and either an amino or a hydroxyl group at the 2' and 6' positions are important for substrate recognition (McKay, Roestamadji, Mobashery & Wright, 1996).

Detailed kinetic studies have shown that APH(3')-IIIa follows a Theorell-Chance sequential-ordered reaction pathway (McKay & Wright, 1995; McKay & Wright, 1996). Binding of Mg-ATP to the enzyme is followed by that of the aminoglycoside, and phosphorylation is a fast step with the release of ADP constituting the rate-limiting step. The kinetic reaction pathway is illustrated below with kanamycin as an example.



The chemical mechanism of this enzyme, however, has been elusive. Since all APH's contain an invariant His (His188 in APH(3')-IIIa) in the C-terminal half of the molecule, a double-displacement mechanism was once proposed (Martin, Jullien & Courvalin, 1988), with His188 being phosphorylated first by Mg-ATP followed by transferring of the phosphate group to the aminoglycoside. However, positional isotope exchange experiments using γ-^{18}O-ATP did not support the presence of a phospho-histidine intermediate and a direct-transfer mechanism was in turn favoured (Thompson, Hughes & Wright, 1996). Using the FSBA (5'-[*p*-(fluorosulfonyl)-benzoyl]adenosine, an ATP analog) labeling technique, Lys33 and Lys44 have been implied to involve in the binding of ATP (McKay, Robinson, Lane & Wright, 1994). Lys44 is invariant among all APH(3') isozymes and Lys33 is unique to APH(3')-IIIa and –VIIa. Only the Ala site mutant of Lys44, but not that of Lys33, showed significantly reduced steady state kinetic reaction and the effect was almost exclusively on the Km (Hon, *et al.*, 1997).

Structural studies on aminoglycoside-modifying enzymes

To date, among the three classes of aminoglycoside-modifying enzymes, only the structures of 4',4"-aminoglycoside nucleotidyltransferase, a member of the ANT class, both as apoenzyme and as a ternary complex with adenosine 5'-α,β-methylenetriphosphate (AMPCPP, a non-hydrolyzable ATP analog) and kanamycin, have been determined (Sakon, *et al.*, 1993; Pedersen, Benning & Holden, 1995). The general fold of the enzyme is that the N-terminal is characterized by a five-stranded β-sheet whereas the C-terminal is formed by five α-helices. The enzyme crystallized as a dimer with the two subunits wrapping about one another such that both subunits contribute to the binding of the

antibiotic. The resolution of the ternary structure was 2.5 Å and the density for the kanamycin moiety was symmetrical such that both the 4'- and the 4"-hydroxyl groups could be placed close to the α-phosphate group of the nucleotide. In the published structure, the assigned orientation of the kanamycin molecule was such that the 4'-hydroxyl group was approximately 5 Å from the alpha-phosphorus of the nucleotide, poised for an in-line attack phosphoryl transfer mechanism.

Objectives

One approach to circumvent the current problem of bacterial resistance to antibiotics is to design inhibitors to known proteins that confer resistance. It is hoped that eventually, through correct identification of the resistant genotypes, a double drug therapy can be administered. A first step to such an approach requires both detailed biochemical and structural knowledge about the target proteins. The objective of this study is to employ x-ray crystallographic techniques to complement ongoing enzymatic analyses in elucidating the mechanisms of inactivation by APH(3')-IIIa. The structural insights gained from this enzyme may be extended to other homologous APH(3') family members.

Overview of this thesis

The work presented in this thesis was initiated when a diffraction data set of 2.7 Å resolution was collected on a crystal grown by Susan Xue using protein sample provided by Dr. Gerard Wright and Geoffrey McKay. Using the same crystallization condition, however, crystals of similar quality could not be reproduced. Subsequently, several rounds of extensive crystallization screening were performed not only to obtain diffraction

quality crystals, but also to produce crystals of protein complexed with substrates or products or their analogs. This effort was partly aided by the technical help of Laura Rossi and Adrian Cristalaw. Chapter two presents the methods and summarizes the results of this early work. Since a suitable model for molecular replacement was not available, the technique of isomorphous replacement was used to derive the structure of this enzyme. Attempts to generate heavy atom derivatives using the soaking technique only produced one platinum derivative that was of insufficient phasing power. The protein model was eventually determined using a multi-wavelength anomalous diffraction (MAD) data set collected on a selenomethionyl protein crystal. The phasing procedure was relatively straightforward despite the fact that the crystal contained ten selenium atoms per asymmetric unit, a substructure once considered large and presumed difficult to be determined. Moreover, post-structure determination data analysis showed that phases of comparable quality could also be obtained using single-wavelength anomalous diffraction data. The experimental work and results of data analyses are described in chapter three. Two structures have been determined for APH(3')-IIIa in this study, a Mg-AMPPNP (adenylyl imidodiphosphate, a non-hydrolysable ATP analog) complex and a Mg-ADP complex. They represent the kinetic stages before and after the phosphorylation. These structures will be described in chapter four, mostly with reference to the catalytic cores of protein kinases which share the basic folding pattern and phosphate recognition scheme. Implications for both phosphoryl transfer mechanism and antibiotic recognition will also be discussed. Chapters three and four are self-contained in that relevant background materials not introduced in this section and methods pertinent to the results are described.

References

Atlas, R. M. (1988). *Microbiology: fundamentals and applications.* New York: MacMillan Publishing Co.

Beaucaire, G. (1995). *J Chemother* **2**, 111-23.

Beauclerk, A. A., & Cundliffe, E. (1987). *J Mol Biol* **193**, 661-71.

Chang, F. N., & Flaks, J. G. (1972). *Antimicrob Agents Chemother* **2**, 294-307.

Cox, J. R., & Serpersu, E. H. (1997). *Biochemistry* **36**, 2353-9.

Courvalin, P., & Carlier, C. (1981). *J Antimicrob Chemother* **8 Suppl A**, 57-69.

Davis, B. D. (1987). *Microbiol Rev* **51**, 341-50.

Fourmy, D., Recht, M. I., Blanchard, S. C., & Puglisi, J. D. (1996). *Science* **274**, 1367-1371.

Hon, W.-C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D., & Berghuis, A. M. (1997). *Cell* **89**, 1-20.

Levison, M. E. (1995). *Infect Dis Clin North Am* **9**, 483-95.

Martin, P., Jullien, E., & Courvalin, P. (1988). *Mol Microbiol* **2**, 615-25.

McKay, G. A., Robinson, R. A., Lane, W. S., & Wright, G. D. (1994). *Biochemistry* **33**, 14115-20.

McKay, G. A., Thompson, P. R., & Wright, G. D. (1994). *Biochemistry* **33**, 6936-44.

McKay, G. A., & Wright, G. D. (1995). *J Biol Chem* **270**, 24686-92.

McKay, G. A., Roestamadji, J., Mobashery, S., & Wright, G. D. (1996). *Antimicrob Agents Chemother* **40**, 2648-50.

McKay, G. A., & Wright, G. D. (1996). *Biochemistry* **35**, 8680-5.

Pedersen, L. C., Benning, M. M., & Holden, H. M. (1995). *Biochemistry* **34**, 13305-11.

Sakon, J., Liao, H. H., Kanikula, A. M., Benning, M. M., Rayment, I., & Holden, H. M. (1993). *Biochemistry* **32**, 11977-84.

Shaw, K. J., Rather, P. N., Hare, R. S., & Miller, G. H. (1993). *Microbiol. Rev.* **57**, 138-163.

Thompson, P. R., Hughes, D. W., & Wright, G. D. (1996a). *Biochemistry* **35**, 8686-95.

Thompson, P. R., Hughes, D. W., & Wright, G. D. (1996b). *Chem Biol* **3**, 747-55.

Trieu-Cuot, P., & Courvalin, P. (1983). *Gene* **23**, 331-41.

Umezawa, H., & Kondo, S. (1982). In *Aminoglycoside antibiotics*, edited by Umezawa, H., & Hooper, I. R. pp. 267-292. Berlin, Heidelberg, New York: Springer-Verlag.

Chapter 2

Protein preparation and crystallization

Protein preparation

Materials

The DNA clone that encodes 3'5"-aminoglycoside phosphotransferase type IIIa (APH(3')-IIIa) was originally isolated from a *Streptococcus faecalis* plasmid (Trieu-Cuot & Courvalin, 1983). It was subcloned into an overexpression vector (McKay, Thompson & Wright, 1994) and the protein produced from the resulting plasmid, pETSacG1, lacks the N-terminal Met residue and hence consists of 263 residues. The *Escherichia coli* strain BL21(DE3) was used as the over-expression vehicle.

Methods

Protein production

The production of APH(3')-IIIa was carried out based largely on the previously described protocol (McKay, Thompson & Wright, 1994) with slight modification for the scale-up process. A 100 ml overnight saturated culture of *E. coli* BL21(DE3)/ pETSacG1 was used to inoculate a 20 L Luria-Bertani broth at 37 °C in a 40 L fermenter (New Brunswick Scientific). The culture medium contained 50 μg of ampicillin per ml of medium. The bacterial culture was allowed to grow to an OD_{595} between 0.6 and 0.7 before induction with 1 mM (final concentration) of IPTG. After a further 3 h growth at 37 °C, the cells were harvested by continuous flow centrifugation at room temperature. A 20 L fermenter preparation typically yielded about 55 g of cell paste. The cell paste was washed once with 200 ml lysis buffer containing 50 mM Tris-HCl (pH 8.0), 200 mM

NaCl and 1 mM EDTA. After centrifugation at 3000*g* at 4 °C, the pellet was resuspended in the same buffer, divided into roughly 10g aliquots, flash frozen in liquid nitrogen and stored at -20 °C until use.

Protein purification

Frozen cell solution was quickly thawed under running tap water and PMSF (100 mM stock solution in ethanol) was added to a final concentration of 1 mM. Cells were lysed in 50 ml aliquots by sonication at 4 °C (ten 30 sec bursts at full power with cooling on ice in between) and then centrifuged for 30 min at 10000 *g* at 4 °C. The contaminants in the lysate were partially removed by precipitation with 30% saturated (at 0 °C) ammonium sulfate. The resulting supernatant was dialysed against two changes of buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl and 1 mM EDTA at 4 °C. The target protein was isolated through two rounds of anion-exchange chromatography at 4 °C. The first round was through a 50 ml Macro-Prep Q (BioRad) that had been pre-equilibrated with the same buffer as that used for the starting protein solution. A linear gradient was run from 150-500 mM NaCl at a rate of 5 ml min^{-1} and 10 ml fractions were collected. APH(3')-IIIa typically eluted as a broad peak around 350 mM NaCl. Proteins from fractions within this peak were separated by SDS-PAGE and visualized by coomassie blue stain. Fractions that were more than 80 % enriched in the expected 31 kD protein were pooled, dialysed against the starting buffer and applied on a Resource-Q column (5 ml, Pharmacia). The elution buffers and the gradient were the same as those used for the Macro-Prep Q column. The elution rate was 1 ml min^{-1} and 1 ml fractions were collected. Fractions that contained near homogeneous APH(3')-IIIa protein were

pooled, dialysed against two changes of 5 mM Hepes (pH 7.5) and 10 mM $MgCl_2$ at 4 °C, and concentrated to 10-15 mg ml^{-1} using the Centriprep10 filtration units (Amicon). Particulate matters in the final solution were removed using the 0.2 μm spin filtration centrifuge tubes (DiaMed). The resulting protein stock solution was stored at 4 °C. Freshly purified protein samples contain both the monomeric and the dimeric forms. The proportion of the two species was not characterized nor separated by size-exclusion chromatography.

Crystallization

Materials

HEPES, sodium cacodylate, magnesium chloride, calcium acetate, magnesium salts of ADP, ATP and AMPPNP (β,γ-imidoadenosine 5'-triphosphate), and the aminoglycosides tobramycin, kanamycin, paromomycin, neomycin, amikacin, lividomycin, ribostamycin and butirosin were purchased from Sigma. PEG 4000 and PEG 8000 were purchased from Fluka. Reagent grade isopropanol and glycerol were purchased from Fisher Scientific. All solutions used for crystallization were pre-filtered through 0.2 μm membrane filtration units (Millipore).

Methods

Crystallization was performed by the vapour diffusion method (McPherson, 1989) mostly in the hanging drop mode, and sometimes in the sitting drop mode. With the hanging drop mode, the cover slips were siliconized by brief immersion in dimethyldichlorosilane, then rinsed with copious amount of water and sterilized by baking

at 190 °C for an hour. The equilibration setup routinely consisted of 500 µl of mother liquor in the wells and 2-3 µl each of protein solution, with or without ligands, and mother liquor in the drops. In order to simulate the different stages of the kinetic pathway of APH(3')-IIIa (see chapter 3), several rounds of screening experiments were performed to search for optimal crystal growth conditions of the protein alone, protein complexed with nucleotide, or with both nucleotide and antibiotic. The protein was typically pre-incubated with 1 mM nucleotide (ADP, ATP, AMPPNP), or 1 mM each of ATP and tobramycin (an inhibitor, equivalent to kanamycin lacking the 3'-OH group, see figure 1-1), or 1 mM each of ADP and one of the aminoglycosides listed above. For each of these mixtures, two sets of screening experiments were performed, one at room temperature and the other at 4 °C, using the sparse matrix method (Jancarik & Kim, 1991) as provided by the Crystal Screen kits I and II (Hampton Research). The starting protein concentration was 8 mg ml^{-1}. Conditions that yielded crystals were further refined by a fine matrix search by varying the concentrations of the salt or precipitant (in steps of ~10 % of the original values) and the protein stock concentration (5-9 mg ml^{-1}).

Results

Out of the 98 mixtures of crystallization reagents, 16 yielded crystals of various sizes and qualities. Six of these conditions were fine-tuned and reproducibly yielded crystals that were of diffraction quality, i.e. at least to a maximum resolution of 3.2 Å. Of all the complete diffraction data sets collected, three different structures have thus far been determined for APH(3')-IIIa. Two of these, a Mg-AMPPNP complex and a Mg-ADP complex, were determined by the author and are presented in this thesis. The third is the

apoenzyme that crystallized in the spacegroup $P4_32_12$ with cell dimensions a=55.5 Å, b=55.5 Å, c=185.1 Å. The structure has been determined to 3.2 Å resolution by Adelaine Leung and will not be discussed here. Protein crystals grown in the presence of aminoglycoside antibiotics are of the spacegroup $P2_12_12_1$ and approximately the same cell dimensions as those of the Mg-ADP and Mg-AMPPNP complexes (statistics of these crystals are given in chapters three and four). The presence of the antibiotics in these crystals have not been evaluated either chemically by techniques such as HPLC or crystallographically using the technique of difference Fourier.

Table 2-1. Crystallization conditions from the sparse matrix screen search that produced diffraction quality protein crystals either directly or after slight fine-tuning.

Salt	Buffer (0.1 M)	Precipitant [a]	Crystal type
2.0 M $(NH_4)_2SO_4$	none	5% isopropanol	apoenzyme
none	Hepes pH 7.5	10% isopropanol 20% PEG 4000	Mg-ADP, Mg-ATP, Mg-AMPPNP
0.2 M Mg acetate	Na cacodylate pH 6.5	20% PEG 8000	Mg-ADP + paromomycin or tobramycin or kanamycin or neomycin or lividomycin
none	Tris-HCl pH 8.5	8% PEG 8000	Mg-ADP + tobramycin

[a] Percentages of isopropanol and PEG solutions are given on (v/v) and (w/v) basis, respectively.

References

Jancarik, J., & Kim, S. H. (1991). *J. Appl. Cryst.* **24**, 409-411.

McKay, G. A., Thompson, P. R., & Wright, G. D. (1994). *Biochemistry* **33**, 6936-44.

McPherson, A. (1989). *Preparation and analysis of protein crystals.* Malabar, Fl.: Krieger Publishing Co.

Trieu-Cuot, P., & Courvalin, P. (1983). *Gene* **23**, 331-41.

Chapter 3

Phasing of selenomethionyl kanamycin kinase : evaluation of single and multiple-wavelength anomalous diffraction experiments

Abstract

The structure of an enterococcal kanamycin kinase has been determined using multi-wavelength anomalous diffraction (MAD) data collected on a single, selenomethionyl (Se-Met) protein crystal. The protein crystallizes as a 62-kD dimer with 12 Se-Met residues per asymmetric unit. The structure was phased by ten of the twelve Se atoms, using anomalous data measured at three wavelengths. In this report, the experimental details of the MAD experiment are reported and the feasibility of using data from fewer wavelengths are also investigated. Our analysis shows that it is possible to use single-wavelength anomalous scattering (SAS) data both to determine the Se substructure and to derive phases of sufficient quality to initiate model building.

Introduction

The techniques of isomorphous replacement and anomalous scattering were first proposed by Bijvoet about half a century ago (Bijvoet, 1954), and through the pioneering work of Perutz and coworkers (Green, Ingram & Perutz, 1954), have remained as the central phasing methods for the determination of unknown protein crystal structures. Both techniques require the presence of heavy atoms in the protein molecules, but each exploits different physical phenomena of x-ray scattering. Isomorphous replacement makes use of normal (Thomson) scattering which involves interaction of x-ray radiation with total

atomic electrons. The scattering is elastic and the intensity, which falls off with increasing diffraction angle, is proportional to the atomic number. On the other hand, anomalous scattering involves electronic transitions of inner shell electrons when the frequency of the incident x-ray approaches the natural absorption frequency of these electrons (James, 1982). These transitions can be observed as absorption edges in an x-ray absorption spectrum of a heavy atom and each edge represents electronic transition of a particular atomic orbital. In contrast to normal scattering, anomalous scattering results in the breakdown of Friedel's law, i.e. reflections that are related by inversion (Friedel pairs) or mirror (Bijvoet pairs) symmetry do not have equal intensities. The magnitude of the differences depends on the type of orbital electrons involved and the frequency of the incident x-ray, but not on the diffraction angle.

For both techniques, location of the heavy atoms is the first step in the derivation of protein phases. This commonly involves the differences in structure factor magnitudes that are directly related by the presence of heavy atoms. Since normal scattering involves all types of atoms, isomorphous replacement requires the subtraction of structure factor magnitudes between *two* diffraction data sets, one from a "native" crystal, the other from a "derivative" which contains additional heavy atoms. On the contrary, since light atoms of biological macromolecules produce negligible anomalous scattering at wavelengths suitable for carrying out x-ray diffraction experiments, subtraction can be carried out between the structure factor magnitudes of Bijvoet (or Friedel) pairs measured from *one* diffraction experiment. For single isomorphous replacement (SIR) and single-wavelength anomalous scattering (SAS) experiments, having positioned and derived the phases of the heavy atoms, calculation of the protein phases remains as a two-solution problem (Wang,

1985). This is the well-known problem of "phase ambiguity" first described by Bijvoet and coworkers (Bijvoet, 1949). A common approach in macromolecular crystallography is to use both phases (double-phase Fourier synthesis) and to employ multiple data sets to break the phase ambiguity, essentially building up the signals and averaging out the noise. In the case of isomorphous replacement, multiple derivatives are often prepared and the technique is called Multiple Isomorphous Replacement (MIR) (Blundell & Johnson, 1976; Ke, 1997). In the case of anomalous scattering, data can be measured at multiple wavelengths from a single crystal and the technique is known as multi-wavelength anomalous diffraction (MAD) (Smith, Andrews, Guest & Harrison, 1990; Hendrickson, 1991; Hendrickson & Ogata, 1997).

Until the recent successful application of MAD experiments in protein structure determination, anomalous scattering has mainly played a supplementary role in isomorphous replacement experiments. For most heavy atoms, the signals produced by anomalous scattering from the K and the L shell electrons are relatively weak as compared to those by normal scattering from the total atomic electrons. This poses stringent requirement on data collection practice. Bijvoet or Friedel pairs are preferably collected on the same frame or close in time, respectively. With accurately measured x-ray data, it is possible to determine protein structures based solely on anomalous diffraction experiments with Bijvoet diffraction ratios ($\langle|F_h|-|F_{-h}|\rangle / \langle|F_p|\rangle$) as low as 1.4% (table 3-4). A classical example is the structure determination of crambin based on the anomalous scattering of sulphur atoms at Cu $K\alpha$ wavelength (1.54 Å) (Hendrickson & Teeter, 1981). The idea of using tunable x-ray radiation to optimize the anomalous signal of heavy atoms and hence the possibility of determining protein structure with MAD data collected from a

single crystal was first described more than 20 years ago (Phillips, Wlodawer, Yevitz & Hodgson, 1976). The applicability of MAD experiments in phasing unknown protein structures was firmly established by the pioneering work of Hendrickson and coworkers (Hendrickson, *et al.*, 1989) using the technique of algebraic formalism (Karle, 1980). The MIR techniques can alternatively be used to phase MAD data. Such application was first reported by Kahn and coworkers (Kahn, *et al.*, 1985) and has been documented in detail by Ramakrishnan and colleagues (Ramakrishnan, Finch, Graziano, Lee & Sweet, 1993; Ramakrishnan & Biou, 1997).

The advent of MAD experiments has called attention to the beauty of anomalous scattering as a single-crystal phasing method for novel macromolecular structures. However, the execution of such experiments depends heavily on the availability of synchrotron beamtime. There are likely occasions, such as the occurrence of interruption due to equipment failure, when the allotted beamtime is insufficient for a full-scale MAD experiment. In this study, we demonstrate that a structure solution can be obtained with single-crystal single-wavelength anomalous diffraction data. The analysis was performed with part of a three-wavelength MAD data set that led to the structure solution of kanamycin kinase (Hon, *et al.*, 1997), formally known as 3',5"-aminoglycoside phosphotransferase type IIIa (APH(3')-IIIa). The 263-residue protein is within the typical size range of most protein structures solved these days. Our results simply restate that with accurately measured data, the error inherent in a SAS experiment can be effectively filtered out by standard crystallographic procedures such as solvent flattening (Wang, 1985) and noncrystallographic symmetry (NCS) averaging (Bricogne, 1976). A structure solution could thus be obtained without auxiliary phasing information.

Materials and methods

Protein preparation, crystallization and crystal handling

The selenomethionyl (Se-Met) protein sample of APH(3')-IIIa was prepared by Dr. G. D. Wright (McMaster University) (Hon, *et al.*, 1997) using the protocol as recommended by Hendrickson *et al.* (Hendrickson, Horton & LeMaster, 1990). The protein was purified by G. A. McKay (McMaster University) in a manner similar to that previously described (McKay, Thompson & Wright, 1994). Briefly, the procedure consisted of differential ammonium sulfate precipitation and large-scale anion-exchange chromatography using a 50-ml Macro-Prep Q gel column (BioRad). An extra step of anion-exchange chromatography was carried out using the Mono-Q anion-exchange column (5 ml, Pharmacia) with the same buffer system and elution gradient. This last step improved the purity of the protein sample. The purified protein was dialysed into a buffer of 5 mM Hepes, pH 7.5 and 10 mM $MgCl_2$.

The molecular mass of the Se-Met-derivatized protein was confirmed by electrospray mass spectrometric analysis (Department of Chemistry, University of Waterloo, Canada) to be consistent with a fully-substituted protein containing six Se-Met residues per monomer. The calculated molecular mass is 31124 per monomer and the measured values were 31125 and 62254 for the monomer and the dimer, respectively. The Se-Met protein crystallized in the same conditions as those used for the native protein. Crystals were grown using the hanging drop vapor diffusion method. The hanging drop typically consisted of 2 µl of protein (8 ml/mg) and Mg-ATP (1 mM) mixed with 2 µl of mother liquor (0.1 M Na cacodylate, pH 6.5, 20% (w/v) of PEG 4000, and 6% (v/v) of

isopropanol). The crystal growth period for the Se-Met protein was longer than that for the native protein (more than two weeks *vs* a few days). The crystals belong to the spacegroup $P2_12_12_1$, with unit cell dimensions of a = 50.7 Å, b = 93.0 Å and c = 134.1 Å. The crystals were stored at liquid nitrogen temperature and were transported to the synchrotron site in a gas-phase nitrogen transporter dewar (Nitrogen). The freezing procedure consisted of rinsing protein crystals for a few seconds in the crystallization mother liquor that was augmented to 30 % (w/v) PEG 4000, flash freezing in liquid nitrogen and storage in homemade cryo-vials of similar design to the commercial ones (Hampton Research).

Data collection, scaling and reduction

MAD data were collected from a single Se-Met protein crystal at the X12C beamline at the National Synchrotron Light Source, Brookhaven. During the time of the experiment, the synchrotron ring was operating at 2.5 GeV. The crystal was kept frozen at -160 °C using an Oxford cryostream cooling device. The Se absorption edge of the protein was scanned using a fluorescence detector. Diffraction data to 2.5 Å were recorded on a MAR Research imaging plate at the inflection (λ1) and peak (λ2) of the absorption edge, and at a remote wavelength (λ3) on the higher energy end from the edge (table 3-1). Collection of Bijvoet reflections was optimized by aligning the a-axis of the crystal parallel to the spindle axis. The data collection strategy was such that 1° oscillation frames of 180-210 s exposure time were collected for the three wavelengths, in the order of λ1, λ2 and λ3, in three consecutive 30° sectors. Data in the blind zone surrounding the a-axis were not collected. Intensities of the diffraction data were integrated and indexed using

DENZO and reduced using SCALEPACK of the HKL package (Otwinowski, 1993; Otwinowski & Minor, 1997). The Bijvoet reflections were scaled separately. The statistics of data collection and scaling are presented in table 3-1.

Determination of the Se positions

The Se sites were initially located using Bijvoet difference Patterson functions and the automated Patterson interpretation routine HASSP in the HEAVYv.4 package (Terwilliger, Kim & Eisenberg, 1987). For the MAD data analysis, Bijvoet reflections from the three wavelengths were merged into one SIRAS-type (single isomorphous replacement complemented with anomalous scattering) data file using the MADMRG routine in HEAVYv.4 (Terwilliger, 1997). The Bijvoet differences averaged from the three wavelengths, $\frac{1}{3}\sum_{\lambda 1}^{\lambda 3}\left(|F_h| - |F_{-h}|\right)$, were used as coefficients for the Patterson synthesis. For the SAS data analysis, Bijvoet differences based on λ2 data were used as coefficients for the Patterson synthesis. In all cases, data from 40.0 to 3.0 Å, with $F/\sigma_F \geq 4$ were used. The Patterson peak searches only produced a subset of the Se sites. The remainder Se sites were located by Bijvoet difference Fourier analyses using phases calculated after the first round of phase estimation, refinement and solvent flattening procedure based on the initial sites.

Phase determination, extension and modification

All procedures of heavy atom parameters refinement, phase determination and extension, solvent flattening and NCS averaging were performed by the PHASES package

(Furey & Swaminathan, 1990). In the case of MAD data analysis, phases were determined in a similar manner to the technique of Multiple Isomorphous Replacement complemented by Anomalous Scattering (MIRAS) (Kahn, *et al.*, 1985; Ramakrishnan & Biou, 1997). Figure 3-1 illustrates the principles of such treatment. Data from the three wavelengths were combined as five separate data sets where the λ3 data were chosen as the native data (table 3-1). The dispersive differences between λ1 and λ3, and between λ2 and λ3 constituted two isomorphous data sets, and Bijvoet differences from λ1 and λ2 two derivative anomalous data sets, and those from λ3 one native anomalous data set. For phasing by the isomorphous replacement technique, only centric reflections and acentric reflections with both Bijvoet pairs measured were used. The protein structure factor amplitudes, $|F_p|$, were approximated by the mean of the Bijvoet pairs, i.e. $(|F_h| + |F_{-h}|)/2$. "Derivative" data from λ1 and λ2 were scaled to the "native" data from λ3 using the Wilson's method. Initial least-square phase refinement and iterative cycles of solvent flattening (solvent content was estimated to be 44% using the formula as recommended by PHASES) and phase modification were carried out using data from 40.0 - 3.0 Å. Values of the anomalous scattering factors used in the phase refinement were copied from a previously published structure whose data were collected at the same synchrotron beamline (Ramakrishnan, Finch, Graziano, Lee & Sweet, 1993). Subsequently, phases were slowly extended to 2.7 Å by coupling to NCS averaging.

In the case of SAS data analysis, native Bijvoet differences from λ2 data were used for phasing by the same procedures as mentioned above. All steps involved in heavy atom parameter refinement and phase calculations were carried out independently of phasing information obtained from the MAD data analysis, except that molecular masks

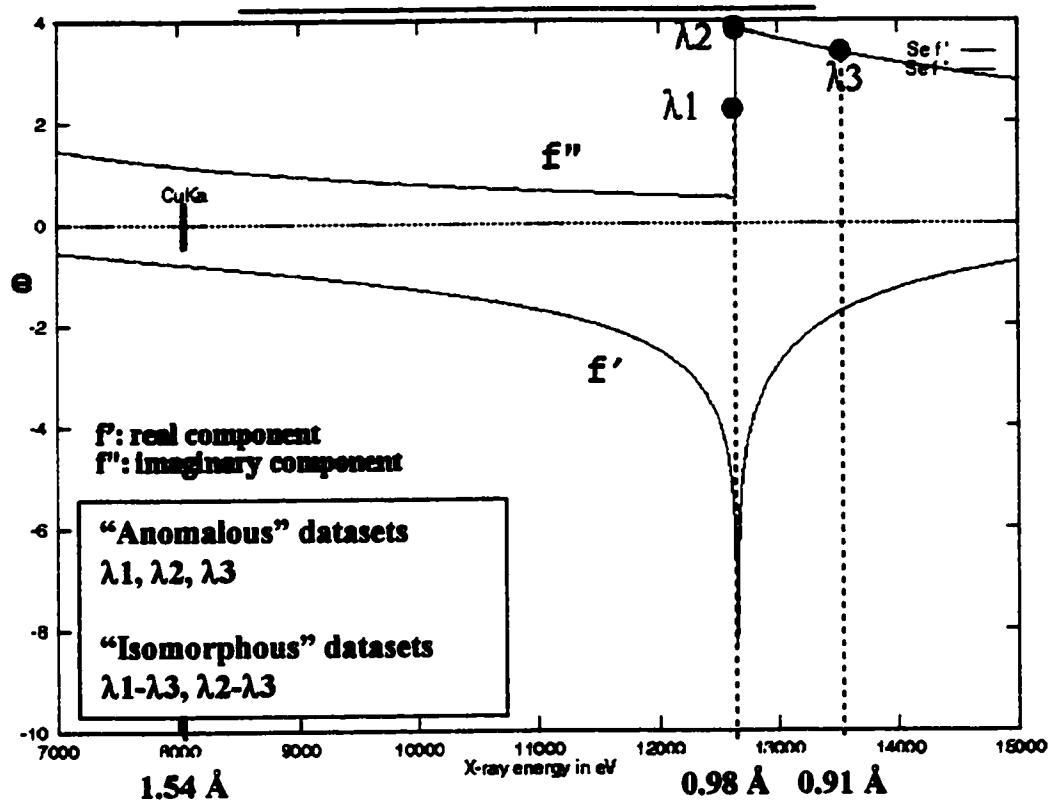


Figure 3-1. X-ray anomalous scattering spectrum of Se illustrating the principle of a MAD data set analysed in a MIRAS manner. The curves represent the change of anomalous scattering coefficients f' and f" (unit: number of electrons) as a function of incident x-ray energy (or wavelength). Data collected at the three indicated wavelengths in the vicinity of the absorption edge of Se can be combined together to generate five independent data sets as shown in the insert. The imaginary component (f") of the anomalous scattering factor gives rise to Bijvoet differences within each wavelength, resulting in three "anomalous" data sets. Significant differences in the real component (f') give rise to the dispersive differences between λ1 and λ3, and also between λ2 and λ3, resulting in two "isomorphous" data sets. The core of this diagram was obtained from Ethan A. Merritt's website "http://www.bmsc.washington.edu/scatter/AS_index.html".

Table 3-1. Statistics of data collection, phasing and model refinement

Data Collection

Wavelength (Å)	Resolution limit (Å)	Reflections Measured / Unique [a]	Completeness (%) All data / last shell	Overall $\langle I/\sigma(I)\rangle$	R_{sym} [b] (%)
0.9802 (inflection, λ1)	2.7	148910 / 26288	67.1 / 31.3	13.0	4.6
0.9795 (peak, λ2)	2.7	130214 / 22978	62.7 / 14.3	12.3	4.5
0.9100 (remote, λ3)	2.7	124782 / 25754	69.9 / 42.2	12.3	4.6

MIRAS phasing (40.0 -3.0 Å)

Data set	Scattering factors (no. of electrons) Δf'	f''	No. of phased reflections	Mean FOM[c]	Overall FOM	Phasing power [d]
λ1-λ3 (SIR)	7.63	-	9001	0.375	0.704	2.88
λ2-λ3 (SIR)	5.46	-	8725	0.325		2.32
λ1 (SAS)	-	3.15	8205	0.411		2.11
λ2 (SAS)	-	5.92	8009	0.434		2.49
λ3 (SAS)	-	3.37	8705	0.315		1.43

Model refinement (40.0 -3.0 Å, based on λ3 data)

Number of reflections		11136
R_{conv} [b] (%)		22.9
R_{free} [c] (%)		29.5
Number of non-hydrogen atoms	protein	4340
	ADP	54
	Mg^{2+}	4
Residues in Ramachandran plot [e] (%)	most favourable region	87.3
	additionally allowed region	11.8
	generously allowed region	0.2
	disallowed region	0.6

[a] The unique reflections only include those that have both Bijvoet mates recorded. The percentage of data completeness is also calculated based on these reflections.

[b] $R_{sym} = \Sigma(|I_i - \langle I\rangle|) / \Sigma\langle I\rangle$, where I_i is the intensity of an individual reflection and $\langle I\rangle$ is the mean intensity of that reflection.

[c] FOM = figure of merit.

[d] Phasing power = $\langle F_h\rangle/\langle E\rangle$, where $\langle F_h\rangle$ is the mean calculated heavy-atom structure factor amplitude, and $\langle E\rangle$ is the mean estimated lack of closure.

[e] As calculated by the program PROCHECK (Laskowski, MacArthur, moss & Thornton, 1993). The percentage represents non-glycine and non-proline residues.

generated interactively for the MAD data were used for NCS averaging. Similarly, phasing using the SIRAS technique was also performed with data from λ2 (derivative) and λ3 (native).

Model building and refinement

The program O (Jones, Zou, Cowan & Kjeldgaard, 1991) was used for model building and refitting. Model building was initiated using the 2.7 Å electron density map calculated with MAD phases that were improved by two-fold NCS averaging. Skeletonization of this experimental map guided the majority of the tracing unambiguously. A model that was more than 90% complete could be built unambiguously based on this experimental map. There were two regions, residues 99-106 and 157-163, that displayed discontinuous electron density. These regions were built using σ_A-weighted (Read, 1994) $2F_o$-F_c electron density map during refinement. Refinement was carried out using the positional and simulated annealing protocols in X-PLOR (version 3.843) (Brunger, 1993) and noncrystallographic symmetry restraint was applied at all stages. λ3 data from 40-2.7 Å were used with bulk-solvent correction (Jiang & Brunger, 1994) (F_p $=(|F_h| + |F_{-h}|)/2$, $F_p/\sigma_F \geq 2$). Cycles of model refitting and refinement were carried out. The final cycle of refinement included grouped B-factor refinement. The statistics of refinement are provided in table 3-1.

Results

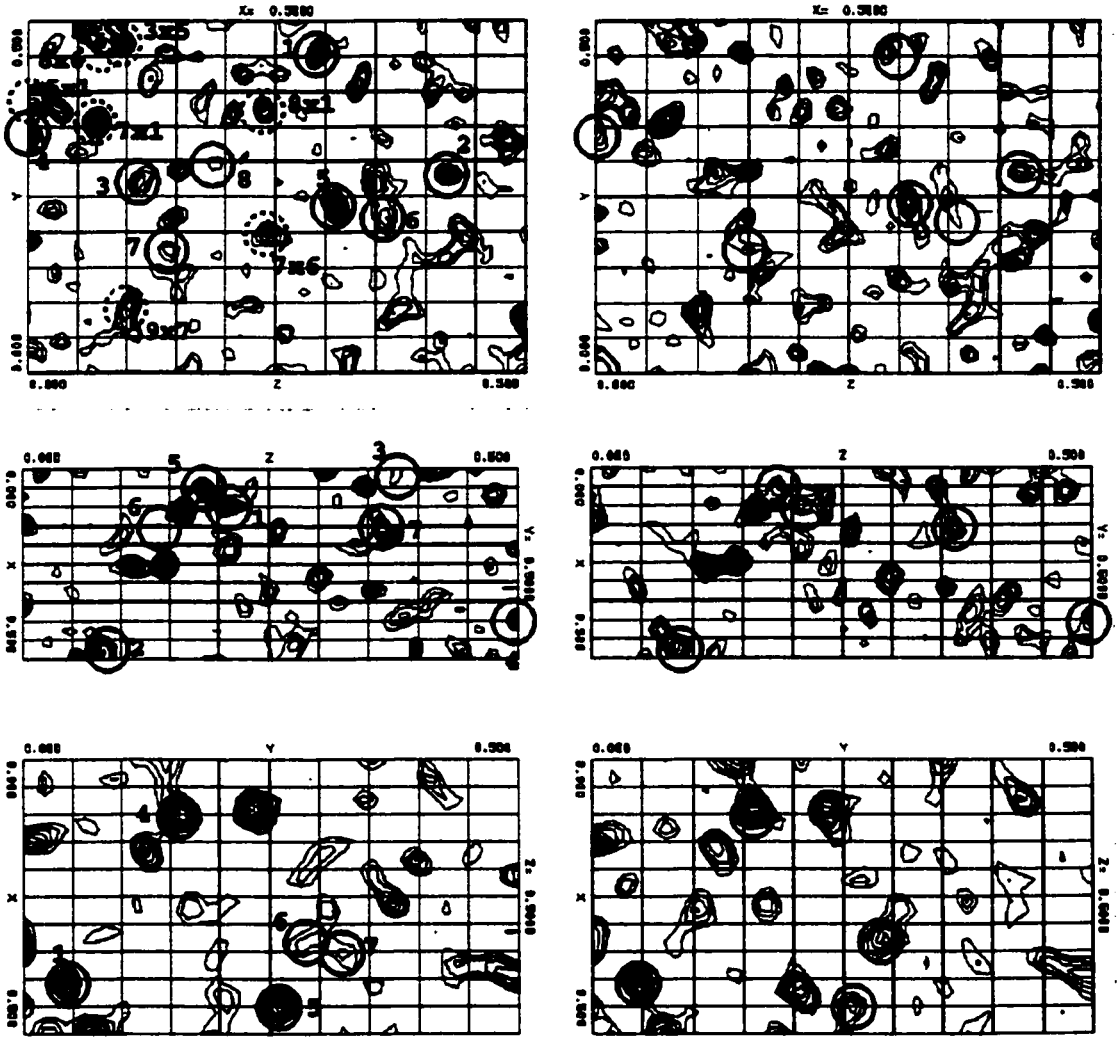
MAD data analysis: determination of the Se positions and phasing

Various means of calculating the Patterson function were attempted; the squares

of dispersive differences $(F_{\lambda 1} - F_{\lambda 3})$ and $[(F_{\lambda 1} - F_{\lambda 3}) + (F_{\lambda 2} - F_{\lambda 3})]/2$, and Bijvoet differences $(|F_h| - |F_{-h}|)_{\lambda 2}$ and $\frac{1}{3}\sum_{\lambda 1}^{\lambda 3}(|F_h| - |F_{-h}|)$ were used as Fourier coefficients. Harker sections with the most distinctive features were generated when the Bijvoet differences from the three data sets were averaged (figure 3-2). Eight Se sites were readily identified from this Patterson synthesis. After solvent flattening, the phases determined from these eight sites were used to calculate the Bijvoet difference Fourier map from which two more sites could be identified. These ten sites belong to a dimer and the non-crystallographic two-fold symmetry as revealed by the two sets of five sites (r.m.s.d. = 0.16 Å) validated their identities (figure 3-3). MIRAS phases derived from the ten Se sites were of high quality as indicated by the map correlation coefficient between the regions that are related by NCS (table 3-3). The electron density map computed with these phases was already clearly interpretable (figure 3-4). The phases were further improved by two-fold NCS averaging and the resulting electron density map was used for subsequent model building. More than 90% of the protein sequence could be built unambiguously. The missing two Se sites were found to belong to Met26 which is situated on a mobile loop on each of the two monomers. The model was later refined against a 2.2 Å native data set and the structure will be described in chapter 4.

SAS data analysis

Using the Bijvoet difference Patterson map calculated with data from λ2, six Se sites could be located (figure 3-2). The remaining four sites could be determined from the Bijvoet difference Fourier map calculated with preliminary phases determined from these



MAD SAS

Figure 3-2. Harker sections (x, y, z = 0.5, from top to bottom) of Bijvoet difference Patterson maps. Patterson maps were calculated with averaged (MAD) and λ2 (SAS) $(\Delta F_{\pm h})^2$ as Fourier coefficients. Self- (solid-line circle) and cross-vectors (broken-line circle) are numbered according to the peak list in table 3-2. All maps are contoured at 0.5σ interval, starting from 1σ level. Individual plots were generated with the program PHASES (Furey & Swaminathan, 1990).

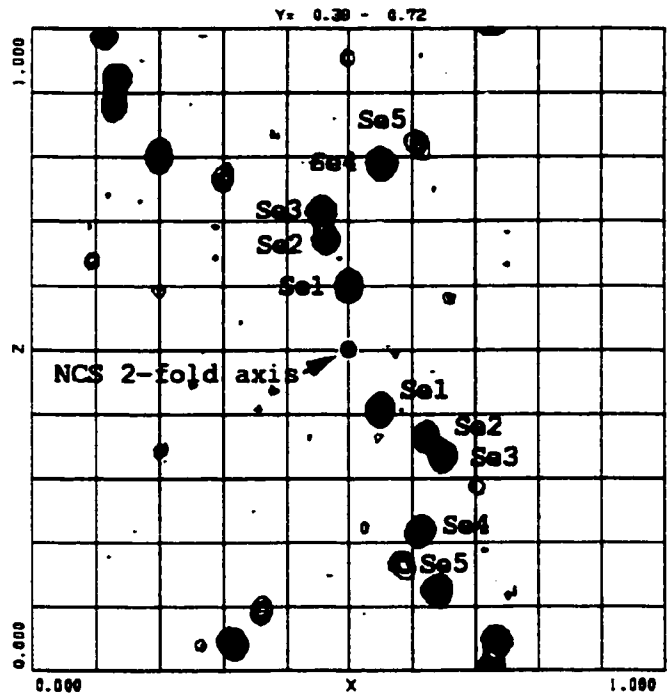


Figure 3-3. Projection of the Bijvoet difference Fourier map demonstrating the two-fold NCS of the ten Se sites. The map, calculated with MIRAS phases to 3.0 Å, was skewed such that the NCS axis is normal to the page. The Se sites are numbered arbitrarily to highlight the symmetry and do not correspond to the peak list in table 3-2. This figure was prepared by modifying the subroutines SKEW and FSFOUR and plotted with CTOUR in the program PHASES (Furey & Swaminathan, 1990).

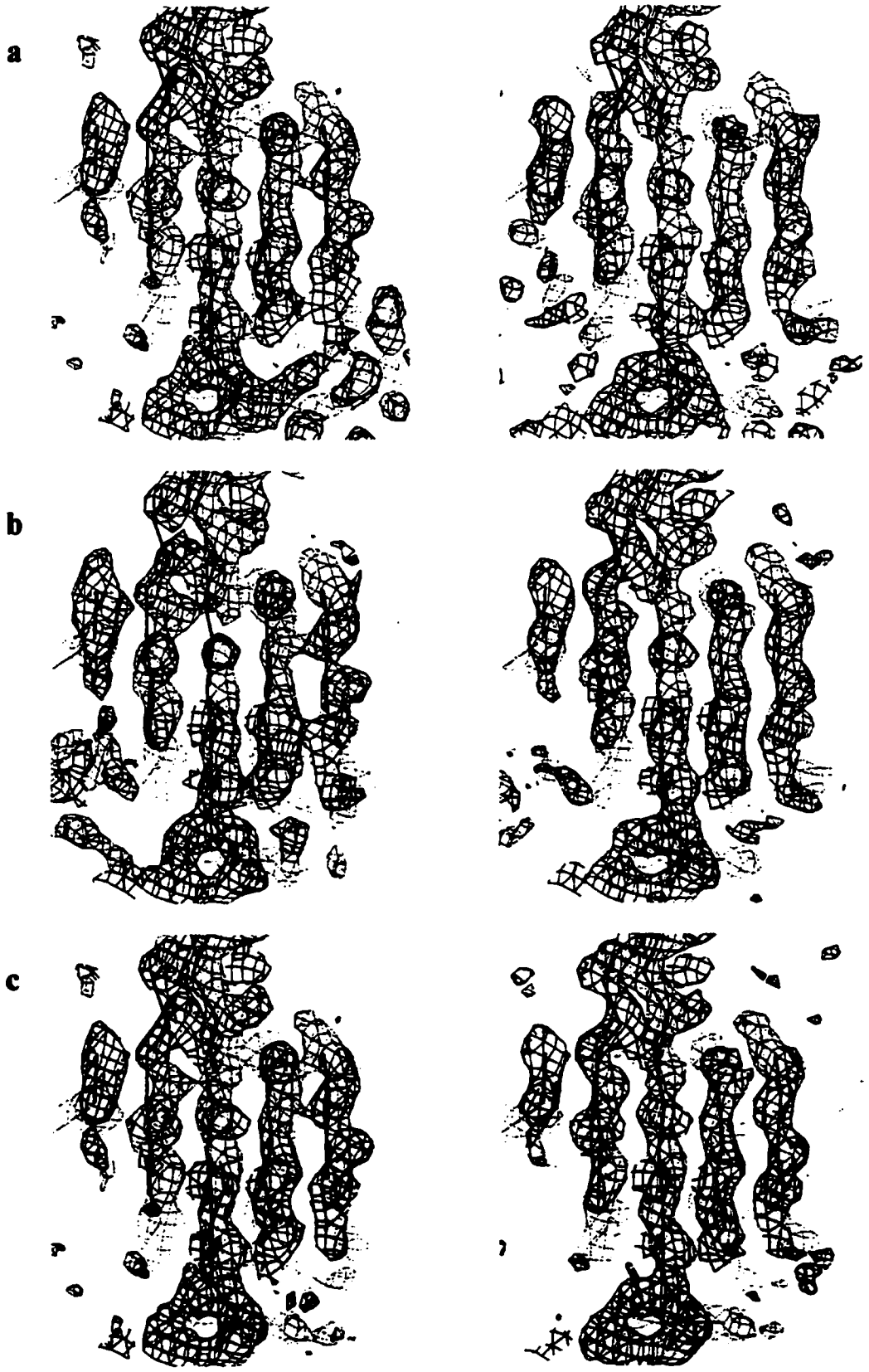

a
b
c

Figure 3-4. Sections of experimental electron density maps showing the five-stranded anti-parallel β-sheet of the N-terminal domain of APH(3')-IIIa. (**a**) λ2 data (SAS), (**b**) λ2 and λ3 data (SIRAS) and (**c**) λ1, λ2 and λ3 data (MIRAS) were used in phase calculations. The maps were calculated with data from 40 - 3.0 Å and contoured from 1σ level. The Cα trace of the refined model is superimposed on the maps. The maps were generated with phases calculated after solvent flattening (left panel) and additionally after NCS averaging (right panel). This figure was prepared with the program O (Jones, Zou, Cowan & Kjeldgaard, 1991).

six sites. The coordinates of these 10 sites are almost identical to those determined using the MAD data (table 3-2). The occupancies between the two Se substructures differ slightly. Solvent flattening alone was sufficient to resolve the phase ambiguity judging by the resulting electron density map that clearly shows the main course of the secondary structures (figure 3-4).

Evaluation of phases obtained with one-, two- and three-wavelength data

Parallel graphical and numerical comparisons of phases determined at different stages, using one- (SAS), two- (SIRAS) and three-wavelength (MIRAS) anomalous data are presented in figures 3-4 and 3-5, and table 3-3. The gross features of the experimental electron density maps are largely similar (figure 3-4). It is interesting to note that the use of NCS averaging has the same effect as adding a data set from another wavelength, judging by the phase errors as compared to the refined, 2.7 Å resolution protein model (figure 3-5, table 3-3). Although it is apparent that each additional set of phasing information improves the quality of the phases progressively, the minimal requirement for a structure solution is SAS data and solvent flattening.

Discussion

There has been increasing application of MAD experiments in the determination of novel protein structures in recent years (Hendrickson & Ogata, 1997). The development of techniques to produce Se-Met derivatives via biological vehicles, including both prokaryotic and eukaryotic overexpression systems, has been particularly important (Hendrickson, Horton & LeMaster, 1990; Yang, Hendrickson, Kalman & Crouch, 1990;

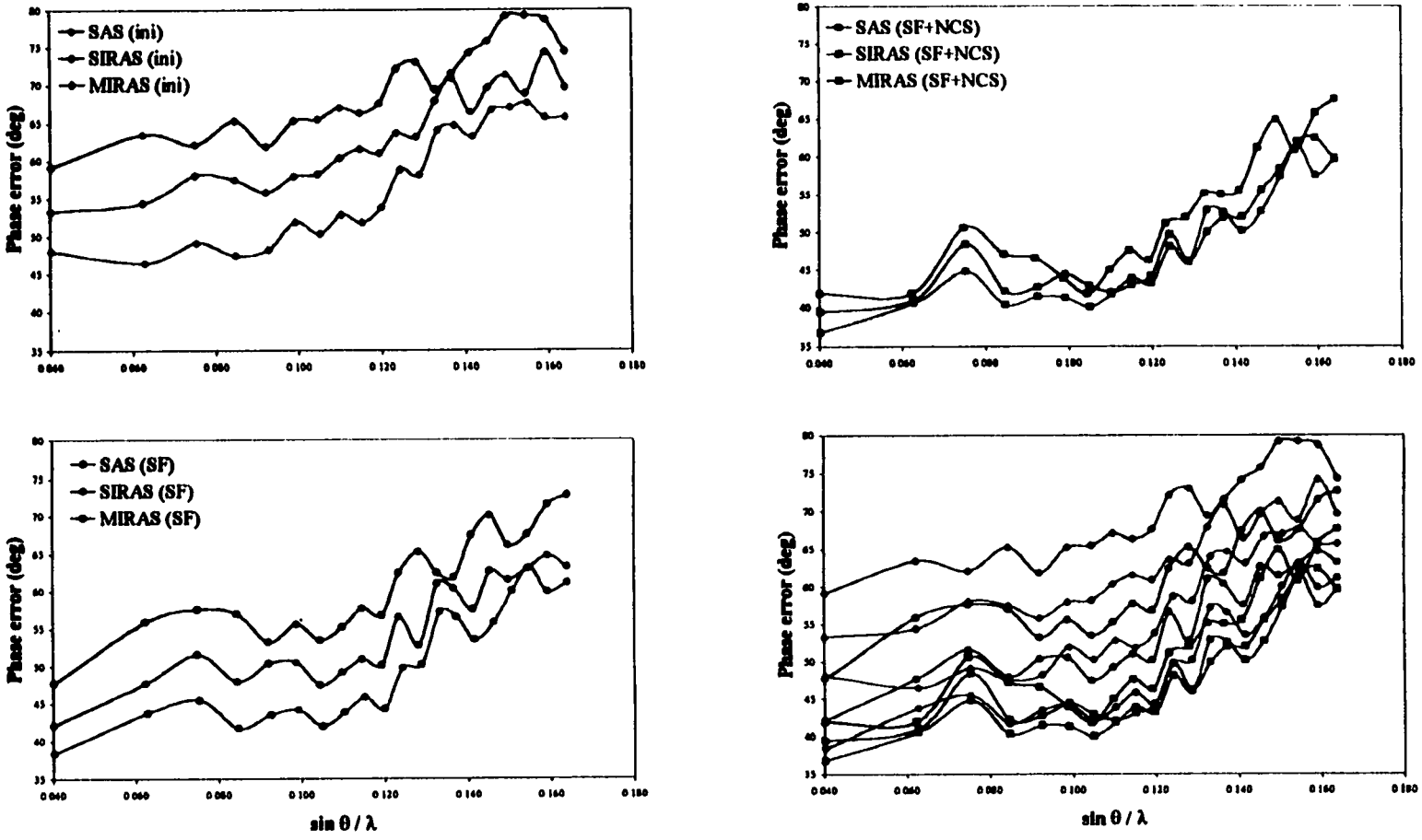


Figure 3-5a. Comparisons of phase errors with respect to phases of the refined 2.7 Å protein model.

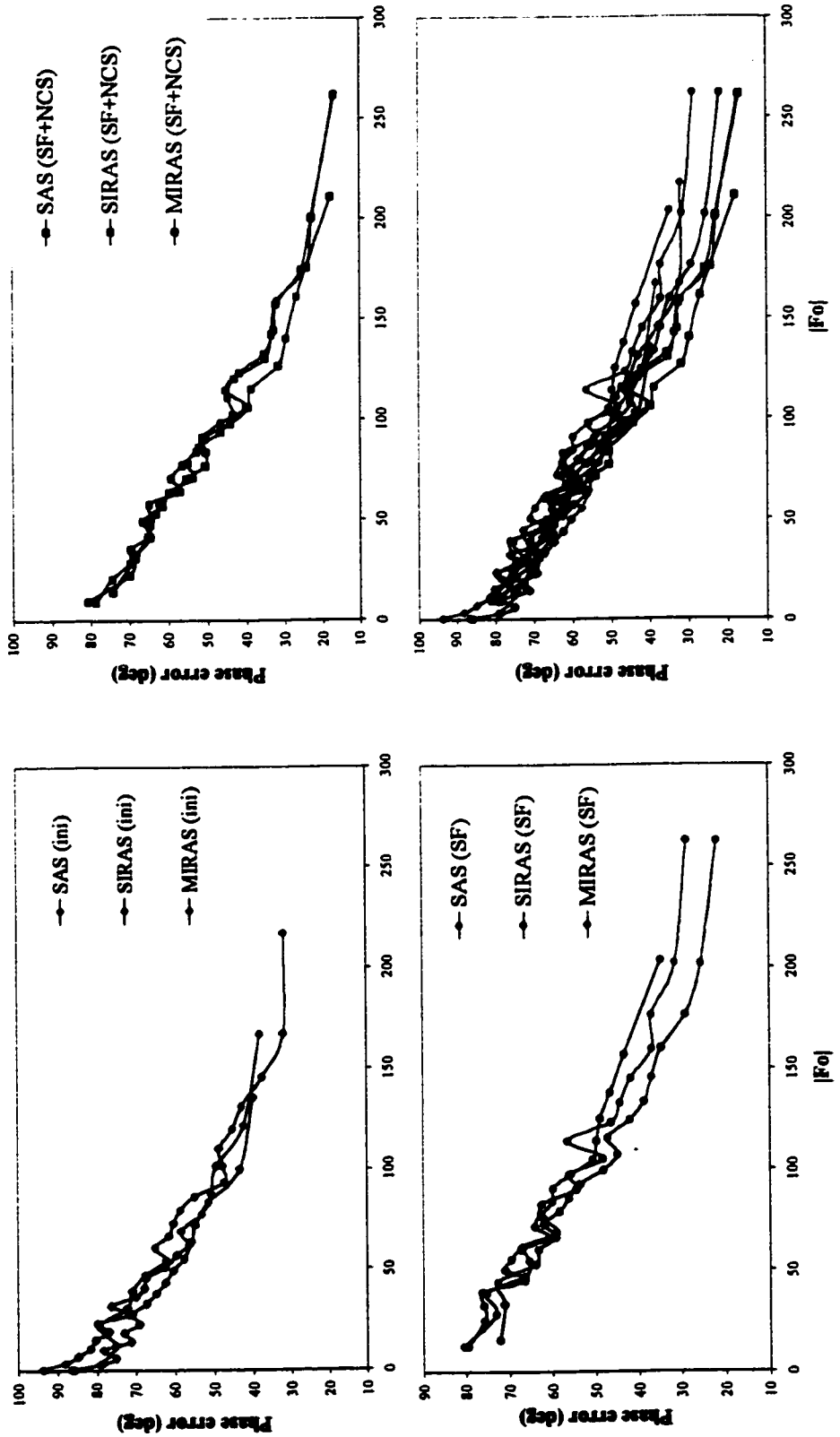


Figure 3-5b. Comparisons of phase errors with respect to phases of the refined 2.7 Å protein model.

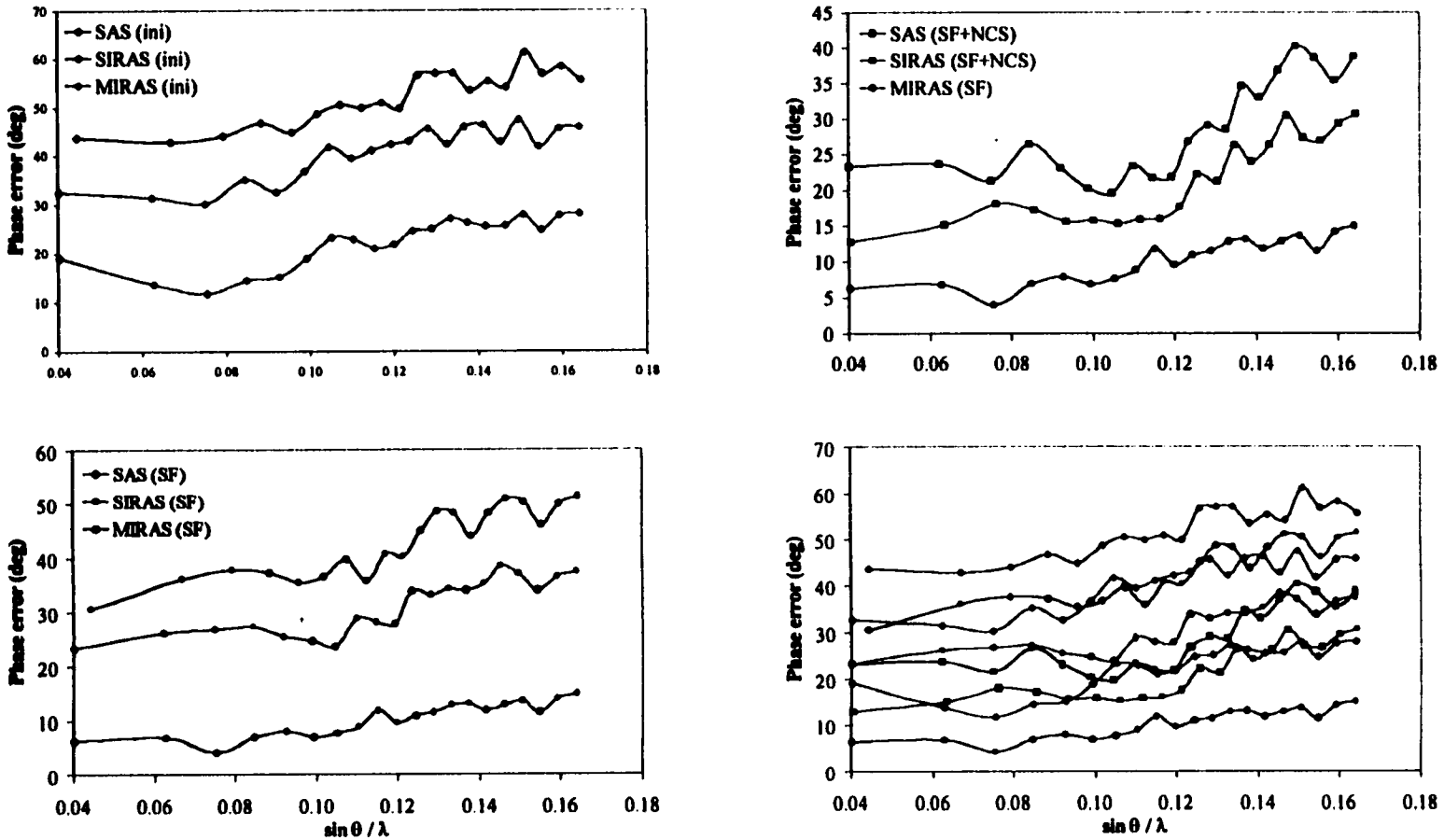


Figure 3-5c. Comparisons of phase errors with respect to MIRAS phases modified with solvent flattening and NCS averaging.

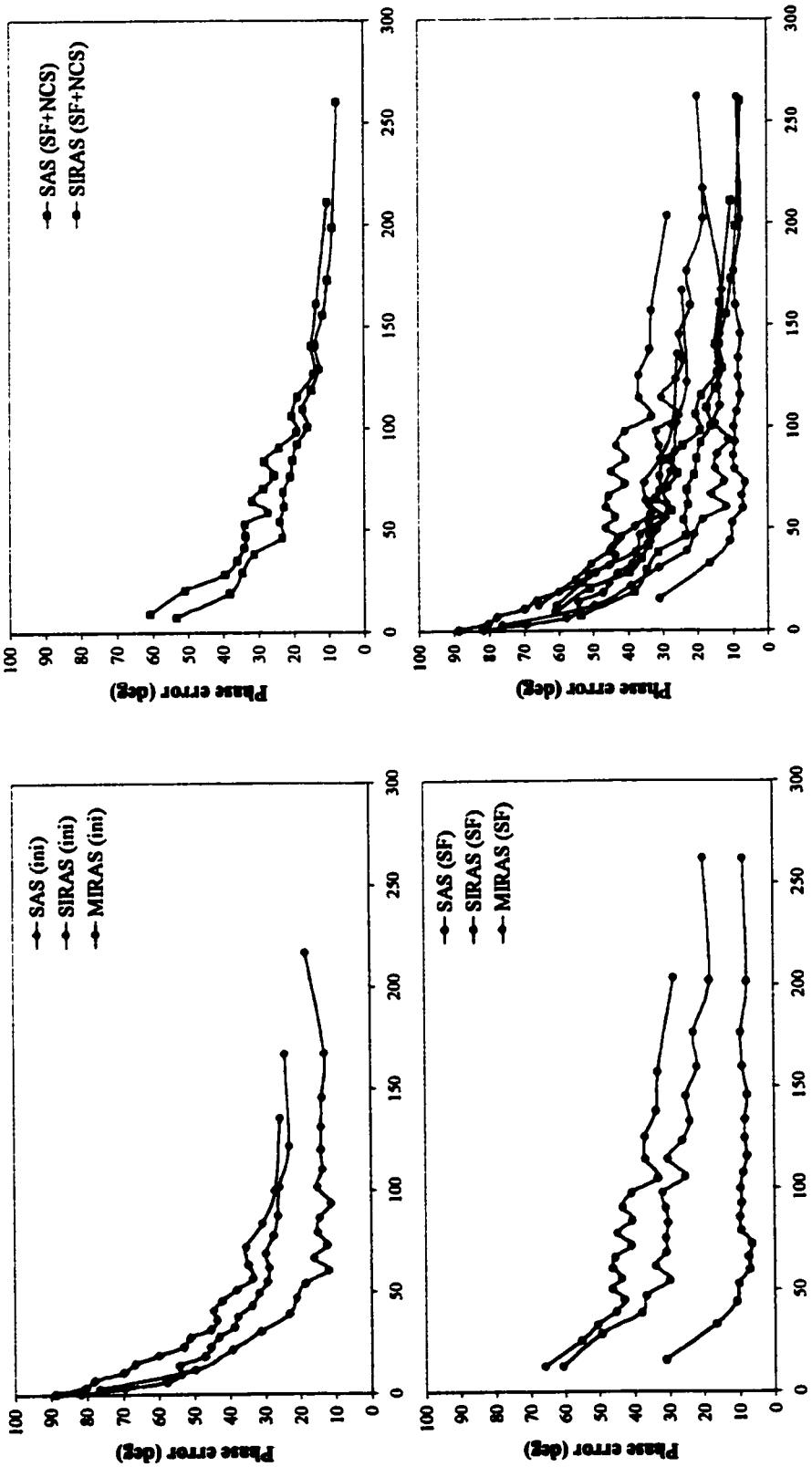


Figure 3-5d. Comparisons of phase errors with respect to MIRAS phases modified with solvent flattening and NCS averaging.

Table 3-2. Coordinates, relative occupancies and B-factors of the ten selenium atoms and their relative peak heights on Bijvoet difference Fourier maps calculated with phases after solvent flattening to 3.0 Å resolution.

Site		Coordinates (MAD/SAS) x	y	z	Occupancy	Group B factor	Relative peak height
1	*	0.45064	0.02166	0.64676	1.00000	22.2300	100.0
	*	0.44974	0.02120	0.64723	1.00000	9.6550	100.0
2	*	0.26487	0.10558	0.78996	0.92552		96.5
	*	0.26488	0.10662	0.79046	0.91725		97.2
3	*	0.48246	0.10662	0.56093	0.84588		88.4
		0.48271	0.10680	0.56128	0.82451		22.4
4	*	0.08057	0.14392	0.82101	0.83825		87.2
	*	0.08065	0.14443	0.82109	0.79094		81.8
5	*	0.02775	0.13108	0.84212	0.93374		83.1
	*	0.02737	0.13188	0.84317	0.81504		75.7
6	*	0.30180	0.07952	0.50184	0.75969		78.1
	*	0.29974	0.07985	0.50214	0.75493		81.8
7	*	0.46750	0.10357	0.59148	0.74959		77.9
		0.46824	0.10432	0.59175	0.72629		17.1
8	*	0.07747	0.16071	0.93108	0.70548		77.9
	*	0.07752	0.16006	0.93086	0.72133		75.7
9		0.55545	0.24427	0.03004	0.44988		29.0
		0.56184	0.24155	0.02890	0.39363		11.7
10		0.13511	0.14455	0.47626	0.44432		23.1
		0.13232	0.14463	0.47800	0.46162		12.3

* Denotes sites that could be identified from Bijvoet difference Patterson maps initially.

Table 3-3. Comparisons of map correlation coefficients and phase errors at 40-3 Å resolution.

	NCS map cor-coef [a]	Map cor-coef *wrt* F_o map of MIRAS (SF+NCS) [b] phases	Map cor-coef *wrt* F_o map of refined model phases	Map cor-coef *wrt* $2F_o$-F_c map of refined model phases	$\Delta\phi$ [c] (°) *wrt* MIRAS (SF+NCS) phases	$\Delta\phi$ (°) *wrt* refined model phases
SAS (λ2)						
Initial phase determination	0.343	0.561	0.390	0.404	52	70
Solvent flattening	0.466	0.464	0.519	0.538	43	60
NCS averaging coupled with phase extension to 2.7 Å	-	0.849	0.703	0.695	28	52
SIRAS (λ2, λ3)						
Initial phase determination	0.440	0.661	0.460	0.471	40	63
Solvent flattening	0.695	0.827	0.661	0.657	31	55
NCS averaging coupled with phase extension to 2.7 Å	-	0.935	0.769	0.760	21	49
MIRAS (λ1, λ2, λ3)						
Initial phase determination	0.588	0.830	0.583	0.587	22	57
Solvent flattening	0.732	0.951	0.719	0.712	10	50
NCS averaging coupled with phase extension to 2.7 Å	-	-	0.769	0.759	0	48

[a] Map correlation coefficient (map cor-coef) = $\dfrac{\Sigma(\rho_1-\Sigma\rho_1/\Sigma N)(\rho_2-\Sigma\rho_2/\Sigma N)}{\sqrt{\Sigma(\rho_1-\Sigma\rho_1/\Sigma N)^2\Sigma(\rho_2-\Sigma\rho_2/\Sigma N)^2}}$, where ρ_1 and ρ_2 represents electron density values at each grid point N of region1 and region2 as in the case of NCS map correlation coefficients, or map1 and map2 as in the case of conventional map correlation coefficients. Both NCS map correlation coefficients and phase errors were calculated using the program PHASES (Furey & Swaminathan, 1990). Conventional map correlation coefficients were calculated using the program MAPMAN (Kleywegt & Jones, 1996). [b] SF=solvent flattening, NCS=NCS averaging. [c] $\Delta\phi$=phase error.

Table 3-4. Protein structures determined based solely on single-wavelength anomalous scattering experiments performed at the Cu *K*α wavelength, 1.54 Å.

Protein name	N_A / N_{res} [a]	Absorption edge of anomalous scatterers (Å)	f" [b]	Estimated Bijvoet diffraction ratio [c] (%)	No. of crystals	Phasing techniques
Crambin [d]	6S / 46	5.02	0.6	1.4	1	resolved anomalous phasing
Rat Cd,Zn metallothionein [e,f]	5Cd / 61	3.50	5.0	12.9	1	ISAS
Bovine neurophysin II [g]	5I / 4x95	2.72	6.8	5.9	1	ISAS
Ocean pout antifreeze protein [h]	6I / 4x62	2.72	6.8	6.6	1	ISAS, 4-fold NCS averaging
(Kanamycin kinase) [i]	(10Se / 2x263)	(0.980)	(5.9)	(6.2)	(1)	(ISAS, 2-fold NCS averaging)

[a] N_A and N_{res} are the number of detectable anomalous scatterers, and number of residues per asymmetric unit, respectively. [b] Values of f" at Cu *K*α wavelength, estimated according to the theory of quantum mechanics, are taken from the *International Tables for X-ray Crystallography* (Macgillavry & Rieck, 1985). [c] Bijvoet diffraction ratio, $\frac{\langle |F_h| - |F_{-h}| \rangle}{\langle F_P \rangle}$, is given by $\sqrt{\frac{N_A}{2}} \frac{2f''}{\langle |F_h| \rangle}$ (Hendrickson & Teeta, 1981). $\langle |F_P| \rangle$ is estimated as $\sqrt{346 \times N_{res}}$, based on the approximation that amino acids have on average molecular weight of 110 and 7.7 non-hydrogen atoms, each with 6.7 electrons (Smith, Andrews, Guest & Harrison, 1990). [d] (Hendrickson & Teeta, 1981), [e] (Furey, *et al.*, 1986), [f] (Robbins, *et al.*, 1991), [g] (Chen, *et al.*, 1991), [h] (Yang, *et al.*, 1998), [i] this study. These examples were retrieved using "single" and "anomalous" as keywords in the search for MEDLINE records and the list may not be exhaustive. Only structures determined without auxiliary phasing information are included.

Doublié, 1997). This is because most proteins contain Met residues and can be produced by recombinant DNA technology. With the use of cryo-cooling techniques to reduce radiation damage of crystals, MAD experiments often allow structure determination using data collected from a single protein crystal. This method will no doubt be the primary crystallographic experiment for determining novel protein structures for years to come. In this study, we have presented our experience concerning two issues that would be of general interest to practitioners of MAD experiments.

First, employing standard techniques, we have located the positions of ten Se atoms in the asymmetric unit with relative ease. The size of the substructure is generally considered large and its determination was presumed to be difficult. This is partly due to experiences with conventional MIR experiments. The presence of such a large number of heavy atoms would likely results in severe non-isomorphism. On the other hand, anomalous diffraction data measured from a single crystal is free from such error. Our results indicate that using an automated Patterson interpretation algorithm such as HASSP (Terwilliger, Kim & Eisenberg, 1987), the technique of Bijvoet difference Patterson synthesis is very amenable to the determination of substructure of this size. Although a few of the Harker peaks appear on all three Harker sections, interpretation by visual inspection is difficult due to the presence of strong features that belong to cross-vectors or that are simply spurious (figure 3-2). We have found that averaged Bijvoet differences enhance the signal to noise ratio in the Patterson synthesis, allowing larger number of sites to be determined (figure 3-2). Averaging of Bijvoet differences is part of the routine MADMRG in the program HEAVYv.4 (Terwilliger, 1997). MADMRG merges MAD data into a single SIRAS type data set by first averaging the dispersive and the Bijvoet

differences separately, which are then given weights estimated by the Bayesian method (Terwilliger, 1994). It is the weighted, averaged dispersive or Bijvoet differences that were meant to be used as Patterson function coefficients by the author of the program. However, such maps generated with our data were of lesser quality (data not shown).

Due to its general applicability to proteins that can be produced by recombinant DNA techniques, Se-Met has become a valuable "designer" derivative. However, the concentration of Se atoms in the asymmetric unit remains as a concern of practicality. Assuming the Met residues are not situated on mobile regions, how many are enough for single-crystal structure determination? In a compilation of structures solved with MAD data till 1995 (Hendrickson & Ogata, 1997), streptavidin had the lowest concentration of Se-Met, with one Se-Met per 126 residues and a Bijvoet diffraction ratio of ~4% (Hendrickson, *et al.*, 1989). On the basis of the lowest Bijvoet diffraction ratio ever applied experimentally, i.e. that of crambin (1.4%) (Hendrickson & Teeter, 1981), it might be feasible to obtain a structure solution with a ratio of one Se-Met per 500 residues (using the equation given in table 3-4 and assuming f'=5.9 and Bijvoet diffraction ratio=2%). One has to bear in mind that the structure of crambin was determined to 1.5 Å and that extremely high resolution diffraction data may be necessary to phase the structure with such weak anomalous signals. However, the diffraction experiment was performed at room temperature for 150 h and only a subset of the reflections was used to phase the structure using the technique of resolved anomalous scattering (Hendrickson & Teeter, 1981). Employing more contemporary methods of data collection and phasing, it may be possible for the average protein crystals to approach the practical limit achieved for crambin, which is generally considered as non-protein-like. These methods include data collection at close

to liquid nitrogen temperature, a practice that can minimize error due to x-ray absorption arising from crystals mounted in glass capillaries. The accuracy of the data measured may also be enhanced by paying attention to the alignment of x-ray optics and the use of optimal methods of data collection and integration.

On the other hand, how many Se atoms would make determination of the substructure too difficult? Given accurately measured anomalous data, there seems to be no theoretical limit. Since anomalous data lack systematic errors due to non-isomorphism, determination of the substructure remains a matter of methodology. The currently available methods for determining heavy atom substructure can be broadly divided into two categories - Patterson synthesis and direct methods. Interpretation of Patterson maps that contain a large number of heavy-atom cross-vectors would necessitate the use of automated algorithms, such as HASSP as implemented in HEAVY (Terwilliger, Kim & Eisenberg, 1987) and SHELX (Sheldrick, 1990; Sheldrick, 1992). Direct methods include the traditional form, as implemented in the programs MULTAN (Germain & Woolfson, 1968) and SHELX, and the iterative form, Shake-and-Bake (Miller, *et al.*, 1993; Weeks, DeTitta, Hauptman, Thuman & Miller, 1994), as implemented in the program *SnB* (Miller, Gallo, Khalar & Weeks, 1994). Atomic-resolution data are not necessary since the data are overdetermined due to the large ratio of reflections to number of Se atoms. Both Patterson synthesis and direct methods can operate on $|^{\circ}F_A|$, as determined using the technique of algebraic formalism, dispersive or Bijvoet differences. Only Bijvoet differences allow the determination of substructure from single-wavelength anomalous data. The applicability and limitation of individual methods will have to await comparisons performed with more complex substructures.

To our knowledge, all phasing of protein structures based solely on Se-Met derivatives has been carried out with data measured at three to four wavelengths. Although three wavelengths are preferred to optimize both the Bijvoet and dispersive differences, two wavelengths have been considered as the minimal requirement for the performance of MAD experiments (Hendrickson, 1991; Ramakrishnan & Biou, 1997). This is because data from two wavelengths are necessary to solve for the unknown parameters in the wavelength-dependent linear equations that are the core of the algebraic formalism (Karle, 1980; Hendrickson, Smith & Sheriff, 1985). Moreover, two-wavelength anomalous data resemble two derivatives in the MIR treatment, facilitating the resolution of phase ambiguity. We have shown here that with single-wavelength anomalous data, it is possible to determine both the Se substructure and the protein structure. The technique of single-wavelength anomalous scattering (SAS), albeit rarely practiced, has been used successfully in phasing of protein structures with similar or less anomalous signals (table 3-4). These experiments were all performed at wavelengths remote from the absorption edges of the anomalous scatterers.

Although solvent flattening is a routine technique in macromolecular x-ray crystallography, the essence of the widely cited article which first described its full automated implementation (Wang, 1985) seems to have been largely ignored (table 3-4). Wang demonstrated that, with accurately measured diffraction data, iterative cycles of error filtering in both direct and reciprocal space were sufficient to resolve the phase ambiguity in a double-phased Fourier map. Hence, the structure solution could be revealed when only one derivative was available. For anomalous diffraction experiment, this technique was referred to as iterative single anomalous scattering (ISAS), and is

different from the resolved anomalous phasing technique (Hendrickson & Teeter, 1981) whose principle was derived from the quasi-anomalous dispersion method (Ramachandran & Srinivasan, 1970). The technique of resolved anomalous phasing operates entirely in the reciprocal space and relies critically on a subset of the reflections that have protein phase approximating the phase of the substructure. The method of ISAS was deemed to have broader applicability since no such approximation is required (Wang, 1985). Both methods have had success with data collected using Cu $K\alpha$ radiation, which is remote from the absorption edges of the anomalous scatterers (table 3-4). By inference, the anomalous signals of Se can also be exploited using home source x-ray radiation given large enough Bijvoet diffraction ratio. The use of Cr $K\alpha$ radiation (2.29 Å) was suggested to enhance the anomalous signals of sulphur atoms since the value of f'' doubles (Wang, 1985). Using Se-Met APH(3')-IIIa as an example, the Bijvoet diffraction ratios at Cu and Cr $K\alpha$ wavelengths are 1.3% (f''=1.3) and 2.5% (f''=2.4), respectively. These values are close to and higher than that for native crambin (table 3-4).

In summary, our data analysis suggests that single-crystal, single-wavelength anomalous diffraction experiments performed with Se-Met protein could be of general applicability. Our study has implication in the design of data collection strategy. Since a SAS data set can produce phases of sufficient quality to initiate model building, one might want to set the priority in collecting a complete data set at the peak of the absorption edge. Data from additional wavelengths can be collected afterwards to increase the interpretability of the experimental electron density map. The practice of cryo-crystallography should undermine the necessity of collecting reflections at different wavelengths close in time.

References

Bijvoet, J. M. (1949). *Proc. Koninklijke Ned. Akad. Wetenschappen* **52**, 313-314.

Bijvoet, J. M. (1954). *Nature* **173**, 888-891.

Blundell, T. L., & Johnson, L. N. (1976). *Protein Crystallography*. New York, London, San Francisco: Academic Press.

Bricogne, G. (1976). *Acta Cryst.* **A32**, 832-847.

Brunger, A. T. (1993). *X-PLOR, Verson 3.1: A System for X-ray Crystallography and NMR*. New Haven, CT, USA: Yale University Press.

Chen, L. Q., Rose, J. P., Breslow, E., Yang, D., Chang, W. R., Furey, W. F., Jr., Sax, M., & Wang, B. C. (1991). *Proc Natl Acad Sci U S A* **88**, 4240-4244.

Doublié, S. (1997). *Methods Enzymol.* **276**, 523-530.

Furey, W., & Swaminathan, S. (1990). *Am. Crystallogr. Assoc. Meet. Program Abstract* **18**, 73.

Furey, W. F., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C., & Stout, C. D. (1986). *Science* **231**, 704-710.

Germain, G., & Woolfson, M. M. (1968). *Acta Cryst.* **B24**, 91-96.

Green, D. W., Ingram, V. M., & Perutz, M. F. (1954). *Proc. R. Soc. Ser. A* **225**, 287-307.

Hendrickson, W. A. (1991). *Science* **254**, 51-58.

Hendrickson, W. A., Horton, J. R., & LeMaster, D. M. (1990). *Embo J* **9**, 1665-1672.

Hendrickson, W. A., & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494-523.

Hendrickson, W. A., Pahler, A., Smith, J. L., Satow, Y., Merritt, E. A., & Phizackerley, R. P. (1989). *Proc Natl Acad Sci U S A* **86**, 2190-2194.

Hendrickson, W. A., Smith, J. L., & Sheriff, S. (1985). *Methods Enzymol* **115**, 41-55.

Hendrickson, W. A., & Teeter, M. M. (1981). *Nature* **290**, 107-113.

Hon, W.-C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D., & Berghuis, A. M. (1997). *Cell* **89**, 1-20.

James, R. W. (1982). *The Optical Principles of the Diffraction of X-rays.* Woodbridge, Connecticut: Ox Bow Press.

Jiang, J.-S., & Brunger, A. T. (1994). *J. Mol. Biol.* **243**, 100-115.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991). *Acta Cryst.* **47**, 110-119.

Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O., & Wery, J. P. (1985). *FEBS Lett* **179**, 133-137.

Karle, J. (1980). *Int. J. Quantum. Chem. Symp.* **7**, 357-367.

Ke, H. (1997). *Methods Enzymol.* **276**, 448-461.

Kleywegt, G. J., & Jones, T. A. (1996). *Acta Cryst* **D52**, 826-828.

Macgillavry, C. H., & Rieck, G. D. (1985). *Physical and chemical tables.* Dordrecht, Holland: D. Reidel Publishing Co.

McKay, G. A., Thompson, P. R., & Wright, G. D. (1994). *Biochemistry* **33**, 6936-6944.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M., & Hauptman, H. A. (1993). *Science* **259**, 1430-1433.

Miller, R., Gallo, S. M., Khalar, H. G., & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613-621.

Otwinowski, Z. (1993). In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, edited by Sawyer, L., Issacs, N., & Bailey, S. pp. 56-62. Warrington, UK: SERC Daresbury Laboratory.

Otwinowski, Z., & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.

Phillips, J. C., Wlodawer, A., Yevitz, M. M., & Hodgson, K. O. (1976). *Proc Natl Acad Sci U S A* **73**, 128-132.

Ramachandran, G. N., & Srinivasan, R. (1970). *Fourier Methods in Crystallography.* New York: Wiley (Interscience).

Ramakrishnan, V., & Biou, V. (1997). *Methods Enzymol* **276**, 538-557.

Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., & Sweet, R. M. (1993). *Nature* **362**, 219-223.

Read, R. (1994). In *Proceedings of the CCP4 Study Weekend*, edited by Bailey, S., Hubbard, R., & Waller, D. pp. 31-40. Warrington, U.K.: Daresbury Laboratory.

Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N. H., Furey, W. F., Wang, B. C., & Stout, C. D. (1991). *J Mol Biol* **221**, 1269-1293.

Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467-473.

Sheldrick, G. M. (1992). In *Crystallographic Computing 5*, edited by Moras, D., Podjarny, A. D., & Thierry, J. C. pp. 145-157. : Oxford Univ. Press.

Smith, J. M., Andrews, S. C., Guest, J. R., & Harrison, P. M. (1990). *Biochem Soc Trans* **18**, 925-926.

Terwilliger, T. C. (1994). *Acta Cryst.* **D50**, 11-16.

Terwilliger, T. C. (1997). *Methods Enzymol* **276**, 530-537.

Terwilliger, T. C., Kim, S. H., & Eisenberg, D. (1987). *Acta. Cryst.* **A43**, 1-5.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90-117.

Weeks, C. M., DeTitta, G. T., Hauptman, H. A., Thuman, P., & Miller, R. (1994). *Acta Cryst.* **A50**, 210-220.

Yang, D. S.-C., Hon, W.-C., Bubanko, S., Xue, Y.-Q., Seetharaman, J., Hew, C. L., & Sicheri, F. (1998). *Biophys. J.* , in press.

Yang, W., Hendrickson, W. A., Kalman, E. T., & Crouch, R. J. (1990). *J Biol Chem* **265**, 13553-13559.

Chapter 4

Crystal structures of 3',5"-aminoglycoside phosphotransferase type IIIa complexed with Mg-AMPPNP and Mg-ADP

Abstract

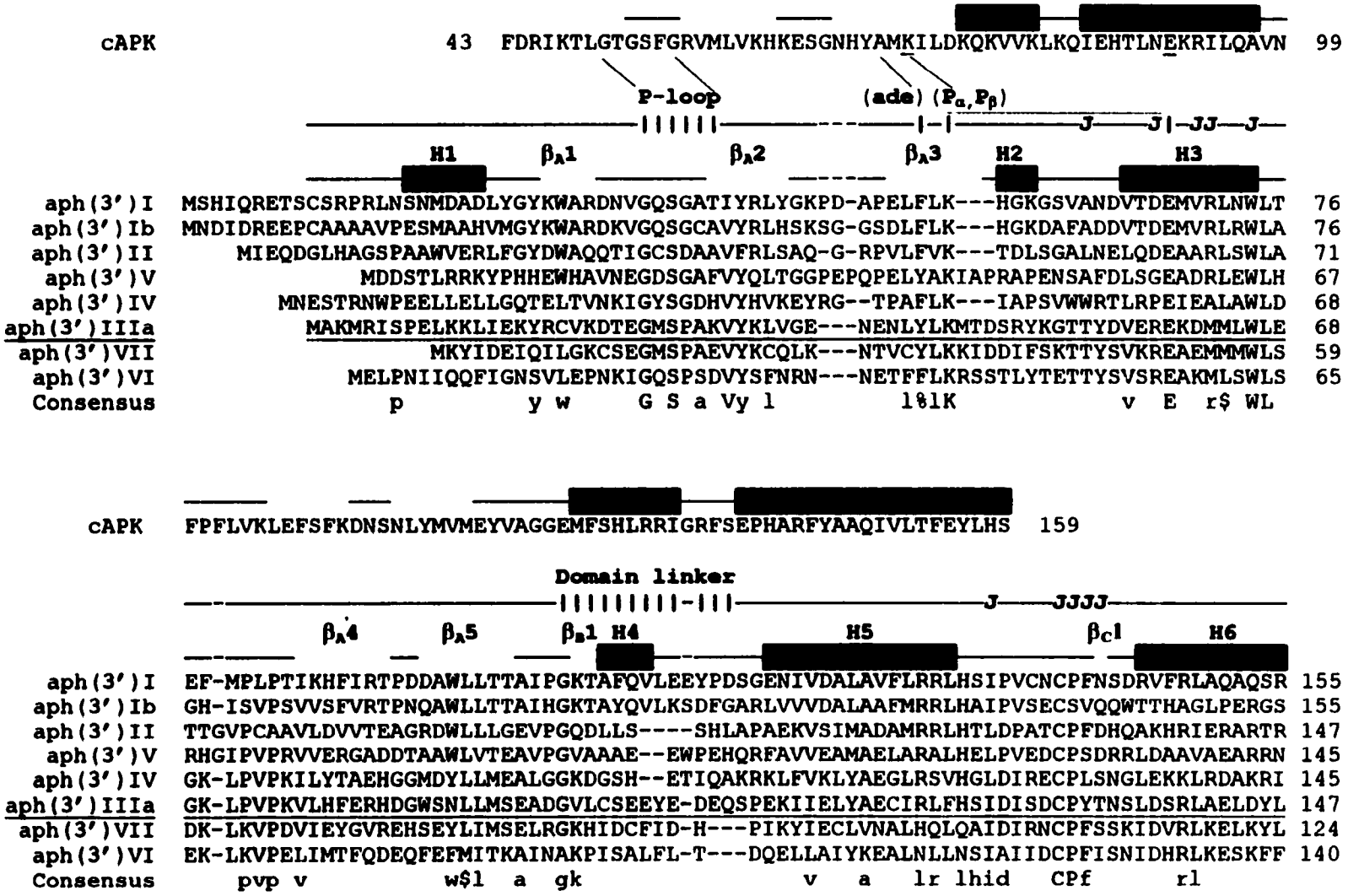
3',5"-aminoglycoside phosphotransferase type IIIa (APH(3')-IIIa) phosphorylates a group of aminoglycoside antibiotics and thereby confers drug resistance to bacteria that express the enzyme. The reaction begins with binding of ATP and ends with releasing of ADP. The crystal structures of APH(3')-IIIa complexed with Mg-AMPPNP (a non-hydrolysable Mg-ATP analog) and Mg-ADP have been refined to 2.4 and 2.2 Å, respectively. The two structures are almost identical, having backbone r.m.s.d. of 0.45 Å. The overall fold is bi-lobal, with the N-lobe dominated by a five-stranded anti-parallel β-sheet and the C-lobe composed mainly of six α-helices. The nucleotide binds in a deep cleft formed between the two lobes. Despite the presence of only 6% sequence identity, half of the molecule has topological equivalence in the catalytic cores of Ser/Thr or Tyr protein kinases. The structural resemblance is especially striking in the nucleotide-binding site where five of the seven invariant residues form an intricate phosphate recognition scheme. In particular, the γ-phosphate of AMPPNP is well positioned for nucleophilic attack via two Mg^{2+} ions in a manner similar to those of the catalytic subunit of cAMP-dependent protein kinase (cAPK) and the γ-subunit of glycogen phosphorylase kinase (Phkγ). This might imply a common phosphoryl transfer mechanism, which has not yet been clearly defined in either type of kinases. Comparison of the Mg-AMPPNP complex of APH(3')-IIIa, the Mn-ATP/peptide inhibitor (PKI(5-24)) ternary complex of cAPK and

the Mn-AMPPNP complex of Phkγ reveals that three of the five invariant residues form the D-N-D triad in the C-lobe, which is also structurally conserved. Moreover, the difference in the position of one of the two metal ions likely holds the key to the elucidation of the mechanistic details of phosphoryl transfer. Two structural features are unique to APH(3')-IIIa. First, a 17-residue flexible loop partially encloses an area on the C-lobe, forming the putative antibiotic-binding site. Second, the two lobes form extensive interactions on the side opposite to the entrance of the active site, hence, likely preventing inter-domain rotation as seen in cAPK.

Introduction

In this section, the crystal structures of the first and the last enzyme complexes in the kinetic pathway of APH(3')-IIIa are presented. These are the first 3D structures determined for the aminoglycoside phosphotransferases, over half of which are APH(3') isozymes (Shaw, Rather, Hare & Miller, 1993). The eight types of APH(3') isozymes shown in figure 4-1 share between 23 to 47 % amino acid sequence identity.

Despite the lack of overall sequence similarity, the structural scaffold of APH(3')-IIIa is similar to those of the catalytic cores of the Ser/Thr or Tyr protein kinases, which constitute a superfamily of enzymes that have been most extensively characterized in the eukaryotic signal transduction pathways (reviewed in (Hanks & Hunter, 1995)). The sizes of these kinases vary and they are often composed of multiple subunits. However, they all share a common catalytic core that is homologous in sequence (Bairoch & Claverie, 1988; Hanks & Hunter, 1995) (figure 4-2) and similar in 3D structure (Bossemeyer, 1995). Proteins that carry out similar reactions and have sequences homologous to this catalytic


cAPK 43 FDRIKTLGTGSFGRVMLVKHKESGNHYAMKILDKQKVVKLKQIEHTLNEKRILQAVN 99
P-loop
(ade) (Pα, Pβ)
H1
βA1
βA2
βA3
H2
H3
aph(3')I MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATIYRLYGKPD-APELFLK---HGKGSVANDVTDEMVRLNWLT 76
aph(3')Ib MNDIDREEPCAAAAVPESMAAHVMGYKWARDKVGQSGCAVYRLHSKSG-GSDLFLK---HGKDAFADDVTDEMVRLRWLA 76
aph(3')II MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQ-G-RPVLFVK---TDLSGALNELQDEAARLSWLA 71
aph(3')V MDDSTLRRKYPHHEWHAVNEGDSGAFVYQLTGGPEPQPELYAKIAPRAPENSAFDLSGEADRLEWLH 67
aph(3')IV MNESTRNWPEELLELLGQTELTVNKIGYSGDHVYHVKEYRG--TPAFLK---IAPSVWWRTLRPEIEALAWLD 68
aph(3')IIIa MAKMRISPELKKLIEKYRCVKDTEGMSPAKVYKLVGE---NENLYLKMTDSRYKGTTYDVEREKDMMLWLE 68
aph(3')VII MKYIDEIQILGKCSEGMSPAEVYKCQLK---NTVCYLKKIDDIFSKTTYSVKREAEMMMWLS 59
aph(3')VI MELPNIIQQFIGNSVLEPNKIGQSPSDVYSFNRN---NETFFLKRSSTLYTETTYSVSREAKMLSWLS 65
Consensus p y w G S a Vy l l%lK v E r$ WL
cAPK FPFLVKLEFSFKDNSNLYMVMEYVAGGEMFSHLRRIGRFSEPHARFYAAQIVLTFEYLHS 159
Domain linker
βA4
βA5
βB1
H4
H5
βC1
H6
aph(3')I EF-MPLPTIKHFIRTPDDAWLLTTAIPGKTAFQVLEEYPDSGENIVDALAVFLRRLHSIPVCNCPFNSDRVFRLAQAQSR 155
aph(3')Ib GH-ISVPSVVSFVRTPNQAWLLTTAIHGKTAYQVLKSDFGARLVVVDALAAFMRRLHAIPVSECSVQQWTTHAGLPERGS 155
aph(3')II TTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS----SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTR 147
aph(3')V RHGIPVPRVVERGADDTAAWLVTEAVPGVAAAE--EWPEHQRFAVVEAMAELARALHELPVEDCPSDRRLDAAVAEARRN 145
aph(3')IV GK-LPVPKILYTAEHGGMDYLLMEALGGKDGSH--ETIQAKRKLFVKLYAEGLRSVHGLDIRECPLSNGLEKKLRDAKRI 145
aph(3')IIIa GK-LPVPKVLHFERHDGWSNLLMSEADGVLCSEEYE-DEQSPEKIIELYAECIRLFHSIDISDCPYTNSLDSRLAELDYL 147
aph(3')VII DK-LKVPDVIEYGVREHSEYLIMSELRGKHIDCFID-H---PIKYIECLVNALHQLQAIDIRNCPFSSKIDVRLKELKYL 124
aph(3')VI EK-LKVPELIMTFQDEQFEFMITKAINAKPISALFL-T---DQELLAIYKEALNLLNSIAIIDCPFISNIDHRLKESKFF 140
Consensus pvp v w$l a gk v a lr lhid CPf rl

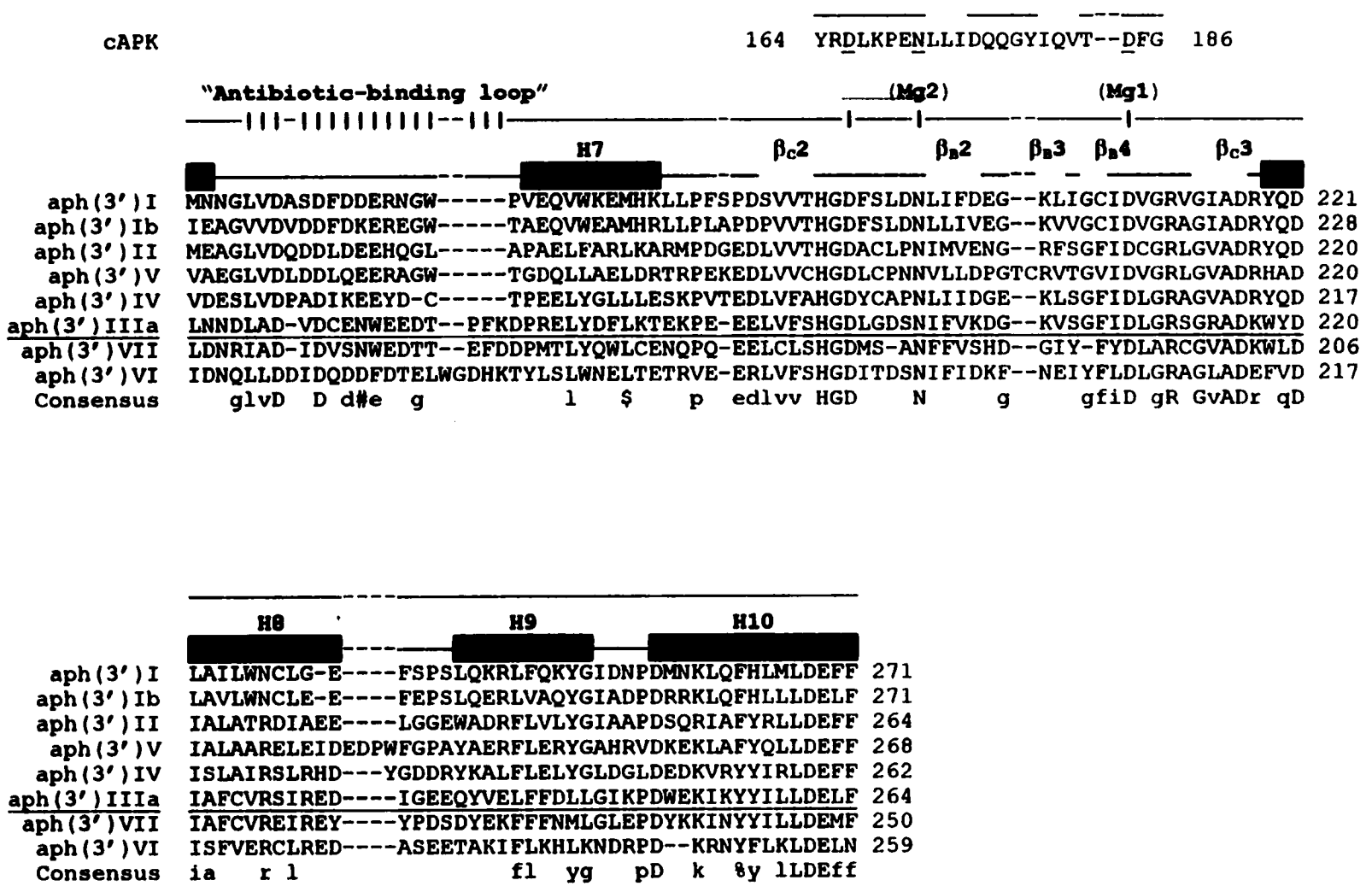
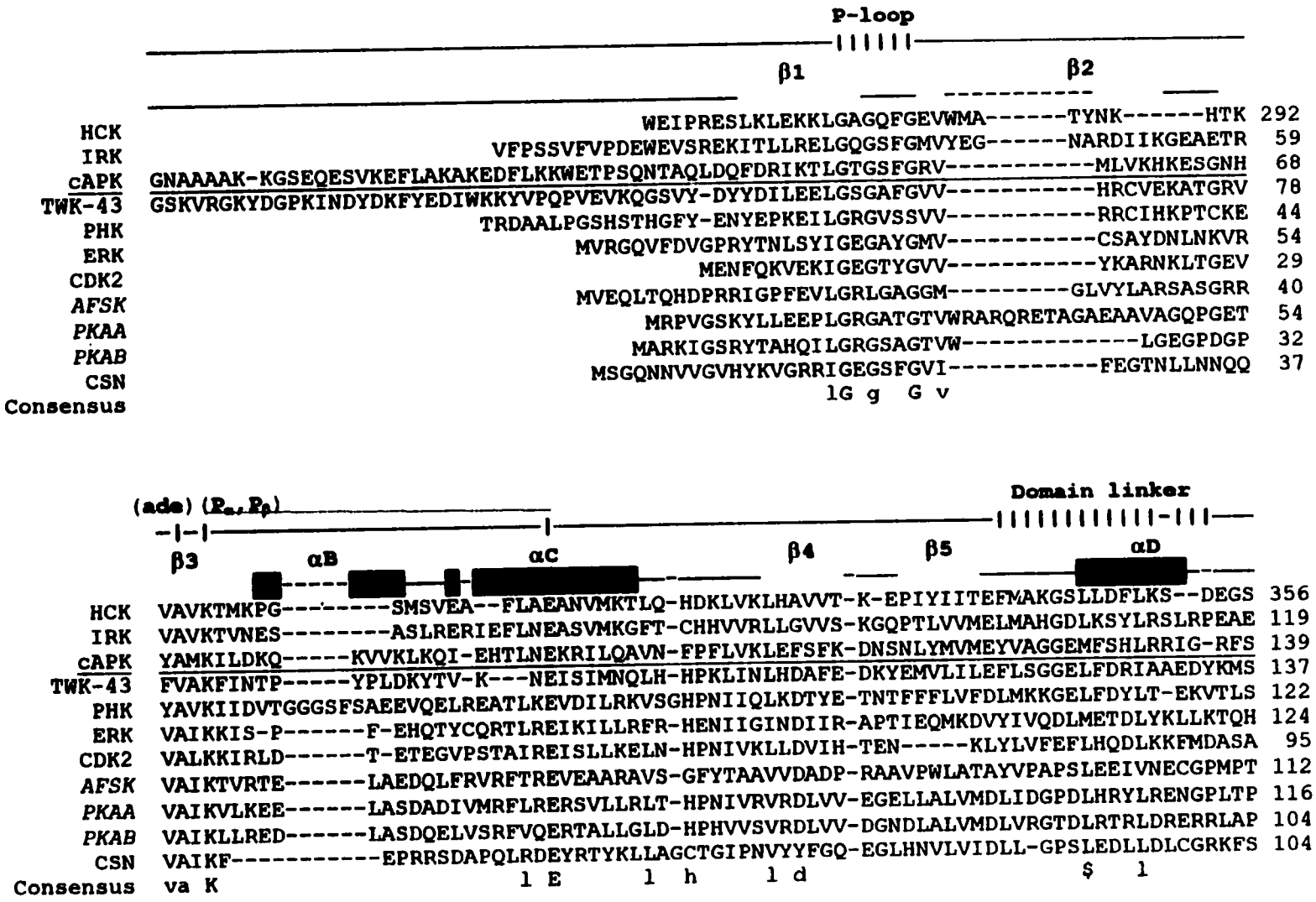

cAPK 164 YRDLKPENLLIDQQGYIQVT--DFG 186
"Antibiotic-binding loop"
(Mg2)
(Mg1)
H7
βc2
βB2
βB3
βB4
βc3
aph(3')I MNNGLVDASDFDDERNGW-----PVEQVWKEMHKLLPFSPDSVVTHGDFSLDNLIFDEG--KLIGCIDVGRVGIADRYQD 221
aph(3')Ib IEAGVVDVDDFDKEREGW-----TAEQVWEAMHRLLPLAPDPVVTHGDFSLDNLLIVEG--KVVGCIDVGRAGIADRYQD 228
aph(3')II MEAGLVDQDDLDEEHQGL-----APAELFARLKARMPDGEDLVVTHGDACLPNIMVENG--RFSGFIDCGRLGVADRYQD 220
aph(3')V VAEGLVDLDDLQEERAGW-----TGDQLLAELDRTRPEKEDLVVCHGDLCPNNVLLDPGTCRVTGVIDVGRLGVADRHAD 220
aph(3')IV VDESLVDPADIKEEYD-C-----TPEELYGLLLESKPVTEDLVFAHGDYCAPNLIIDGE--KLSGFIDLGRAGVADRYQD 217
aph(3')IIIa LNNDLAD-VDCENWEEDT--PFKDPRELYDFLKTEKPE-EELVFSHGDLGDSNIFVKDG--KVSGFIDLGRSGRADKWYD 220
aph(3')VII LDNRIAD-IDVSNWEDTT--EFDDPMTLYQWLCENQPQ-EELCLSHGDMS-ANFFVSHD--GIY-FYDLARCGVADKWLD 206
aph(3')VI IDNQLLDDIDQDDFDTELWGDHKTYLSLWNELTETRVE-ERLVFSHGDITDSNIFIDKF--NEIYFLDLGRAGLADEFVD 217
Consensus glvD D d#e g l $ p edlvv HGD N g gfiD gR GvADr qD
H8
H9
H10
aph(3')I LAILWNCLG-E----FSPSLQKRLFQKYGIDNPDMNKLQFHLMLDEFF 271
aph(3')Ib LAVLWNCLE-E----FEPSLQERLVAQYGIADPDRRKLQFHLLLDELF 271
aph(3')II IALATRDIAEE----LGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF 264
aph(3')V IALAARELEIDEDPWFGPAYAERFLERYGAHRVDKEKLAFYQLLDEFF 268
aph(3')IV ISLAIRSLRHD---YGDDRYKALFLELYGLDGLDEDKVRYYIRLDEFF 262
aph(3')IIIa IAFCVRSIRED----IGEEQYVELFFDLLGIKPDWEKIKYYILLDELF 264
aph(3')VII IAFCVREIREY----YPDSDYEKFFFNMLGLEPDYKKINYYILLDEMF 250
aph(3')VI ISFVERCLRED----ASEETAKIFLKHLKNDRPD--KRNYFLKLDELN 259
Consensus ia r l fl yg pD k %y lLDEff

Figure 4-1. Amino acid sequence alignment of eight types of APH(3') isozymes and secondary structure assignment of the Mg-AMPPNP complex of APH(3')-IIIa. Mutliple sequence alignment was performed by MULTIALIGN, an adapted version of CLUSTAL W (Corpet, 1988). Residues that are >90% and >50% conserved are highlighted in red and blue, respectively. Residues are defined to have similar properties as follows: ! is any one of IV, $ is any one of LM, % is any one of FY and # is any one of NDQE. Secondary structure assignment was performed by DSSP (Kabsch & Sander, 1983) as implemented in the graphic program RASMOL (Sayle & Millner-White, 1995). Magenta blocks represent α- and 3_{10}-helices (H); and yellow blocks represent beta-strand (β, where β_A1 = strand 1 of β-sheet A). Residues with known or predicted functions are highlighted. I = nucleotide-binding residue; ade = adenine base; P_α, P_β = α and β phosphate groups of ADP/ATP; --- = hydrogen bond; J = domain joint. Topologically equivalent sections of the sequence of cAPK (PDB ID 1ATP (Knighton, *et al.*, 1991)) are also shown for comparison. Residues of cAPK that are conserved within the protein kinases superfamily and have counterparts in the APH(3') isozymes are also coloured in red and blue (see also figure 4-2). Only those residues that share identity between the two types of kinases are underlined. GENBANK accession numbers are 66878 (aph(3')-I), 125463 (aph(3')-II), 125464 (aph(3')-IIIa), 125465 (aph(3')-IV), 125466 (aph(3')-V), 125467 (aph(3')-VI), 125468 (aph(3')-VII), 125469 (aph(3')-Ib (misnamed in the literature as aph(3')-VIII, Gerard D. Wright, personal communication)) and 349839 (cAPK).


P-loop
β1
β2
(ade) (Pα, Pβ)
Domain linker
β3
αB
αC
β4
β5
αD
HCK WEIPRESLKLEKKLGAGQFGEVWMA------TYNK------HTK 292
IRK VFPSSVFVPDEWEVSREKITLLRELGQGSFGMVYEG------NARDIIKGEAETR 59
cAPK GNAAAAK-KGSEQESVKEFLAKAKEDFLKKWETPSQNTAQLDQFDRIKTLGTGSFGRV----------MLVKHKESGNH 68
TWK-43 GSKVRGKYDGPKINDYDKFYEDIWKKYVPQPVEVKQGSVY-DYYDILEELGSGAFGVV----------HRCVEKATGRV 78
PHK TRDAALPGSHSTHGFY-ENYEPKEILGRGVSSVV----------RRCIHKPTCKE 44
ERK MVRGQVFDVGPRYTNLSYIGEGAYGMV----------CSAYDNLNKVR 54
CDK2 MENFQKVEKIGEGTYGVV----------YKARNKLTGEV 29
AFSK MVEQLTQHDPRRIGPFEVLGRLGAGGM---------GLVYLARSASGRR 40
PKAA MRPVGSKYLLEEPLGRGATGTVWRARQRETAGAEAAVAGQPGET 54
PKAB MARKIGSRYTAHQILGRGSAGTVW----------LGEGPDGP 32
CSN MSGQNNVVGVHYKVGRRIGEGSFGVI----------FEGTNLLNNQQ 37
Consensus lG g G v
HCK VAVKTMKPG----------SMSVEA--FLAEANVMKTLQ-HDKLVKLHAVVT-K-EPIYIITEFMAKGSLLDFLKS--DEGS 356
IRK VAVKTVNES--------ASLRERIEFLNEASVMKGFT-CHHVVRLLGVVS-KGQPTLVVMELMAHGDLKSYLRSLRPEAE 119
cAPK YAMKILDKQ-----KVVKLKQI-EHTLNEKRILQAVN-FPFLVKLEFSFK-DNSNLYMVMEYVAGGEMFSHLRRIG-RFS 139
TWK-43 FVAKFINTP-----YPLDKYTV-K---NEISIMNQLH-HPKLINLHDAFE-DKYEMVLILEFLSGGELFDRIAAEDYKMS 137
PHK YAVKIIDVTGGGSFSAEEVQELREATLKEVDILRKVSGHPNIIQLKDTYE-TNTFFFLVFDLMKKGELFDYLT-EKVTLS 122
ERK VAIKKIS-P------F-EHQTYCQRTLREIKILLRFR-HENIIGINDIIR-APTIEQMKDVYIVQDLMETDLYKLLKTQH 124
CDK2 VALKKIRLD------T-ETEGVPSTAIREISLLKELN-HPNIVKLLDVIH-TEN-----KLYLVFEFLHQDLKKFMDASA 95
AFSK VAIKTVRTE------LAEDQLFRVRFTREVEAARAVS-GFYTAAVVDADP-RAAVPWLATAYVPAPSLEEIVNECGPMPT 112
PKAA VAIKVLKEE------LASDADIVMRFLRERSVLLRLT-HPNIVRVRDLVV-EGELLALVMDLIDGPDLHRYLRENGPLTP 116
PKAB VAIKLLRED------LASDQELVSRFVQERTALLGLD-HPHVVSVRDLVV-DGNDLALVMDLVRGTDLRTRLDRERRLAP 104
CSN VAIKF----------EPRRSDAPQLRDEYRTYKLLAGCTGIPNVYYFGQ-EGLHNVLVIDLL-GPSLEDLLDLCGRKFS 104
Consensus va K l E l h l d $ l

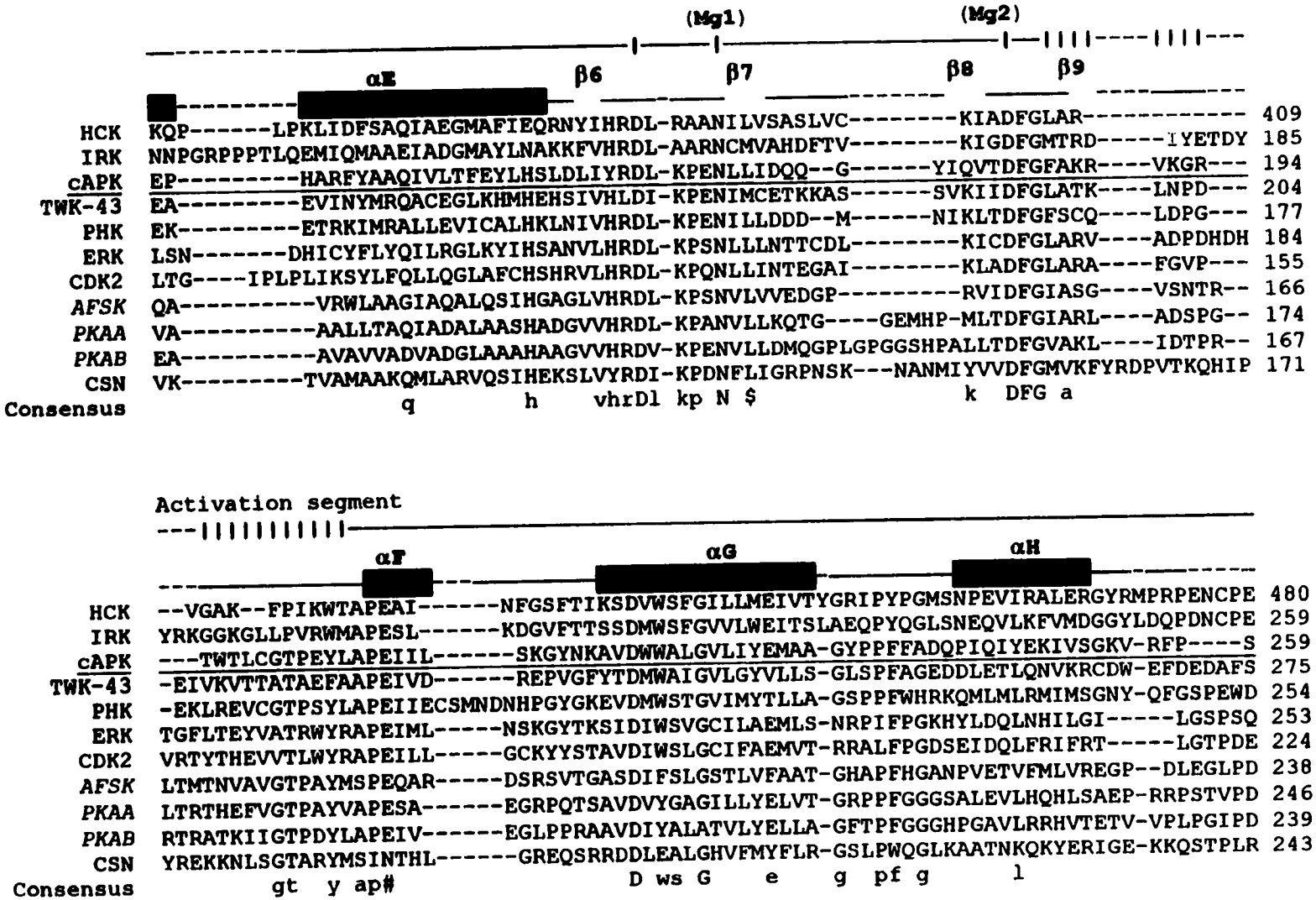

(Mg1)
(Mg2)
αE
β6
β7
β8
β9
HCK KQP------LPKLIDFSAQIAEGMAFIEQRNYIHRDL-RAANILVSASLVC--------KIADFGLAR---------- 409
IRK NNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDL-AARNCMVAHDFTV--------KIGDFGMTRD-----IYETDY 185
cAPK EP---------HARFYAAQIVLTFEYLHSLDLIYRDL-KPENLLIDQQ--G------YIQVTDFGFAKR----VKGR--- 194
TWK-43 EA---------EVINYMRQACEGLKHMHEHSIVHLDI-KPENIMCETKKAS------SVKIIDFGLATK----LNPD--- 204
PHK EK---------ETRKIMRALLEVICALHKLNIVHRDL-KPENILLDDD--M------NIKLTDFGFSCQ----LDPG--- 177
ERK LSN-------DHICYFLYQILRGLKYIHSANVLHRDL-KPSNLLLNTTCDL--------KICDFGLARV----ADPDHDH 184
CDK2 LTG----IPLPLIKSYLFQLLQGLAFCHSHRVLHRDL-KPQNLLINTEGAI--------KLADFGLARA----FGVP--- 155
AFSK QA----------VRWLAAGIAQALQSIHGAGLVHRDL-KPSNVLVVEDGP---------RVIDFGIASG----VSNTR-- 166
PKAA VA----------AALLTAQIADALAASHADGVVHRDL-KPANVLLKQTG----GEMHP-MLTDFGIARL----ADSPG-- 174
PKAB EA----------AVAVVADVADGLAAAHAAGVVHRDV-KPENVLLDMQGPLGPGGSHPALLTDFGVAKL----IDTPR-- 167
CSN VK---------TVAMAAKQMLARVQSIHEKSLVYRDI-KPDNFLIGRPNSK---NANMIYVVDFGMVKFYRDPVTKQHIP 171
Consensus q h vhrDl kp N $ k DFG a
Activation segment
αF
αG
αH
HCK --VGAK--FPIKWTAPEAI------NFGSFTIKSDVWSFGILLMEIVTYGRIPYPGMSNPEVIRALERGYRMPRPENCPE 480
IRK YRKGGKGLLPVRWMAPESL------KDGVFTTSSDMWSFGVVLWEITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPE 259
cAPK ---TWTLCGTPEYLAPEIIL------SKGYNKAVDWWALGVLIYEMAA-GYPPFFADQPIQIYEKIVSGKV-RFP----S 259
TWK-43 -EIVKVTTATAEFAAPEIVD------REPVGFYTDMWAIGVLGYVLLS-GLSPFAGEDDLETLQNVKRCDW-EFDEDAFS 275
PHK -EKLREVCGTPSYLAPEIIECSMNDNHPGYGKEVDMWSTGVIMYTLLA-GSPPFWHRKQMLMLRMIMSGNY-QFGSPEWD 254
ERK TGFLTEYVATRWYRAPEIML-----NSKGYTKSIDIWSVGCILAEMLS-NRPIFPGKHYLDQLNHILGI-----LGSPSQ 253
CDK2 VRTYTHEVVTLWYRAPEILL-----GCKYYSTAVDIWSLGCIFAEMVT-RRALFPGDSEIDQLFRIFRT-----LGTPDE 224
AFSK LTMTNVAVGTPAYMSPEQAR-----DSRSVTGASDIFSLGSTLVFAAT-GHAPFHGANPVETVFMLVREGP--DLEGLPD 238
PKAA LTRTHEFVGTPAYVAPESA------EGRPQTSAVDVYGAGILLYELVT-GRPPFGGGSALEVLHQHLSAEP-RRPSTVPD 246
PKAB RTRATKIIGTPDYLAPEIV------EGLPPRAAVDIYALATVLYELLA-GFTPFGGGHPGAVLRRHVTETV-VPLPGIPD 239
CSN YREKKNLSGTARYMSINTHL------GREQSRRDDLEALGHVFMYFLR-GSLPWQGLKAATNKQKYERIGE-KKQSTPLR 243
Consensus gt y ap# D ws G e g pf g l

Figure 4-2. Sequence alignment, common architecture and functional motifs of representative catalytic cores of protein kinases. Eukaryotic protein kinases are human hematopoietic cell kinase (HCK, tyrosine kinase domain), human insulin-receptor kinase (IRK, tyrosine kinase domain), yeast cAMP-dependent protein kinase (cAPK, catalytic subunit), mollusc twitchin kinase (TWK-43), rabbit phosphorylase kinase (PHK, γ-subunit), rat MAP kinase (ERK), human cyclin-dependent protein kinase 2 (CKD2) and yeast casein kinase-1 (CSN). Prokaryotic protein kinases, AFSK is from *Streptomyces griseus,* and PkaA and PkaB are from *Streptomyces coelicolor.* Mutliple sequence alignment was performed by MULTIALIGN, an adapted version of CLUSTAL W (Corpet, 1988). Residues that are >90% and >50% conserved are highlighted in red and blue, respectively. Residues are defined to have similar properties as follows: ! is any one of IV, $ is any one of LM, % is any one of FY and # is any one of NDQE. Highlighted, conserved secondary structure elements are based on the structure of cAPK (PDB ID 1ATP), and the original nomenclature are used (Knighton, *et al.*, 1991). Sequences following helix H (αH for cAPK) are not conserved and are therefore not shown in here. Colour schemes are the same as those described in figure 4-1. GENBANK accession numbers are 2194104 (HCK), 999510 (IRK), 349839 (cAPK), 1942209 (TWK-43), 1827741 (PHK), 2194078 (ERK), 1942964 (CDK2), 1127277 (CSN), 1703208 (AFSK), 666024 (PKAA) and 666025 (PKAB).

core have lately been found in the prokaryotes (reviewed in (Zhang, 1996)). Unless mentioned otherwise, the term "protein kinases" will hereafter be referred strictly to the catalytic cores of the Ser/Thr or Tyr protein kinases and not other types of protein kinases. The structures of the two nucleotide complexes of APH(3')-IIIa reveal the atomic details involved in phosphate recognition. The nucleotide-binding sites of these two structures will be compared against those of the Mn-ATP/PKI(5-24) complex of cAPK and the Mn-AMPPNP complex of Phky. The observed similarities and differences will form the basis of hypotheses about events that might happen during phosphoryl transfer. Moreover, the putative antibiotic-binding site can be delineated on the basis of the current structures and analysis of amino acid sequence homology among the APH(3') isozymes.

Materials and methods

Crystallization, data collection and processing

Protein crystals of both Mg-ADP and Mg-AMPPNP complexes could be obtained using the same crystallization condition as previously described (see chapter 3). These crystals were produced by co-crystallizing the protein with Mg-ATP and Mg-AMPPNP, respectively. The conversion of Mg-ATP to Mg-ADP could be due to hydrolysis of the compound in solution during crystallization. Diffraction data for the high resolution Mg-ADP complex was collected at the X12C beamline of the National Synchrotron Light Source. During data collection, the crystal was kept frozen at -160 °C using an Oxford cryostream cooling device. Diffraction data to 2.0 Å were recorded on a MAR Research imaging plate. For the Mg-AMPPNP complex, the diffraction experiment was performed on a RAXIS IIC area detector using Cu $K\alpha$ radiation generated by a RIGAKU RU200 rotating

anode generator operating at 3 kW (60 kV x 50 mA), and focused to a 0.2 mm focal spot by SUPPER double focusing mirrors. The data were collected at -167°C using the Molecular Structure Corporation cryosystem. The crystal diffracted to 2.2 Å. The cryo-cooling condition for both crystals was the same as that previously described (see chapter 3). The diffraction data of both data sets were indexed, integrated and scaled using the HKL package (Otwinowski & Minor, 1997). Both crystals contain one dimer per asymmetric unit. The statistics of data are shown in table 4-1. See also appendix 1.

Model building and refinement

All the model refitting and refinement were performed with the programs O (Jones, Zou, Cowan & Kjeldgaard, 1991) and X-PLOR (version 3.843) (Brunger, 1993), respectively. The 2.7 Å resolution protein model obtained with the Se-Met derivative (see chapter 3) was used to initiate refinement for both the Mg-AMPPNP and the high-resolution Mg-ADP complexes. Data for the two data sets were corrected for bulk solvent (Jiang & Brunger, 1994) and filtered by $F/\sigma_F \geq 2$ cutoff and noncrystallographic symmetry restraint was applied at all stages. Since neither the data for the Mg-AMPPNP complex nor the high resolution Mg-ADP complex were isomorphous with those of the 2.7 Å data, the procedure began with rigid-body refinement using 40.0-3.5 Å data. This was followed by positional powell minimisation and simulated annealing protocols as implemented in X-PLOR using data extending to 2.4 Å (Mg-AMPPNP complex) and 2.2 Å (Mg-ADP complex) resolution. There were four loop regions that required extensive rebuilding, including residues 23-27, 99-107, 152-163 and 231-237 which previously displayed poor main-chain and/or side-chain electron density. These regions were rebuilt for the two

Table 4-1. Statistics of data and refined structures of APH(3')-IIIa.

	Mg^{2+}-ADP complex	Mg^{2+}-AMPPNP complex
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (Å)	a=49.71, b=91.25, c=131.30	a=50.17, b=91.32, c=132.65
Maximum resolution (Å)	2.2	2.4
Observed reflections	210381	82458
Unique reflections	30934	24554
R_{sym} [a] (%)	5.5	5.5
Completeness (%)	96.4	99.2
Model refinement ($F/\sigma_F \geq 2$)		
Reflections used in R-factor calculations		
Number	28047	22401
Resolution range (Å)	40.0-2.2	40.0-2.4
Completeness of highest resolution shell (%)	82.5 (2.2-2.3 Å)	87.1 (2.4-2.5 Å)
R_{conv} [b] (%)	22.3	22.3
R_{free} [c] (%)	30.0	30.1
Number of non-hydrogen atoms		
protein	4340	4340
ADP / AMPPNP	54	62
Mg^{2+}	4	4
water	289	169
isopropanol	8	4
Rmsd in bond lengths (Å)[d]	0.009	0.009
Rmsd in bond angles (°)[d]	1.4	1.5
Average B factors (Å^2)		
Main-chain atoms	18.3	23.41
Side-chain atoms	21.2	26.88
Residues in Ramachandran plot [e] (%)		
most favourable	92.2	88.8
additionally allowed	7.6	10.8
generously allowed	0.2	0.4
disallowed	0.0	0.0
Luzzati estimated coordinate error (Å) [f,g] test set / work set	0.4 / 0.3	0.5 / 0.3

[a] $R_{sym} = \Sigma(|I_i - \langle I \rangle|) / \Sigma\langle I \rangle$ where I_i is the intensity of an individual reflection and $\langle I \rangle$ is the mean intensity of that reflection. [b] $R_{conv} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively. [c] R_{free} is equivalent to R_{conv} for 5 % of randomly selected F_o that were not included in the refinement (Brunger, 1992). [d] As calculated by X-PLOR 3.8 (Brunger, 1993). [e] As calculated by the program PROCHECK (Laskowski, MacArthur, Moss & Thornton, 1993). The percentage represents non-glycine

and non-proline residues. [f] (Luzzati, 1952). [g] These values were calculated using all data. See also appendix 1.

monomers in the asymmetric unit individually using σ_A-weighted (Read, 1994) $2F_o$-F_c electron density maps calculated with the corresponding region omitted. Cycles of model refitting and refinement were carried out. The fit to electron density was evaluated for the entire protein molecules at every cycle and model rebuilding was carried out where necessary. For the 2.4 Å resolution data, the σ_A-weighted $2F_o$-F_c electron density map showed clear electron density for the γ-phosphate in both monomers and the Mg-AMPPNP moiety was built in using a refined model from the database as a template (PDB ID 1CDK). When the conventional and free R-factors were refined to 26.3% and 31.4% for the Mg-AMPPNP complex, and 28.7% and 32.7% for the Mg-ADP complex, the solvent molecules were located and refined. Residual peak heights, hydrogen-bonding distances, steric contacts, individual refined B-factors and overall R-factors were taken into account during the selection process. The automatically selected water molecules were inspected on a 2Fo-Fc σ_A-weighted map calculated with the solvent molecule omitted to ensure good electron density peaks. Water molecules that are situated on density regions belonging to side-chains or symmetry related molecules were rejected. Inclusion of the water molecules improved the subsequent electron density maps and part of aforementioned loop regions was again refitted. The final cycle of refinement included restrained B-factor refinement. Statistics of the refinement are given in table 4-1.

Results and Discussion

Quality of the models

Statistics of the two refined models are shown in table 4-1. The main-chains of both monomers in the two crystal complexes have satisfactory geometry, with 88.8% (Mg-

AMPPNP complex) and 92.2% (Mg-ADP complex) of the non-glycine and non-proline residues placed in the most favoured regions and none in the disallowed regions in the Ramachandran plots (Laskowski, MacArthur, Moss & Thornton, 1993) (figure 4-3a). Glu105 and Gln106 of monomer A in the Mg-AMPPNP complex and Glu105 of monomer A in the Mg-ADP complex are placed in the generously allowed regions. The electron density of these regions was still ambiguous at the final stage of refinement. The backbone r.m.s.d. between monomer A and B is 0.32 Å for the Mg-AMPPNP complex, and 0.37 for the Mg-ADP complex. The regions that differ include residues 23-27, 103-107, 152-159 and 231-234, which are all situated on loop regions that display the higher than average main-chain B-factors (figure 4-3b).

Overall architecture of APH(3')-IIIa

The structures of the Mg-ADP and Mg-AMPPNP complexes are very similar, displaying an overall backbone r.m.s.d. of 0.44 Å for monomer A and 0.34 Å for monomer B. Thus far, the γ-subunit of the glycogen phosphorylase kinase (Phkγ) is the only protein kinase structure with both Mg-ADP and Mg-AMPPNP (or Mg-ATP) complexes reported (Owen, Noble, Garman, Papageorgiou & Johnson, 1995). Major structural change between the two complexes was also not observed. The only difference was a minute movement of the phosphate-binding loop (P-loop, see below) towards the phosphates in the ADP-complex. In the following, description of the APH(3')-IIIa structure will be based on the Mg-AMPPNP complex, and differences will be described when they arise. The structures of cAPK, a prototype of the catalytic core of Ser/Thr or Tyr protein kinases, will be used as the primary reference of comparisons. Residues corresponding to those of

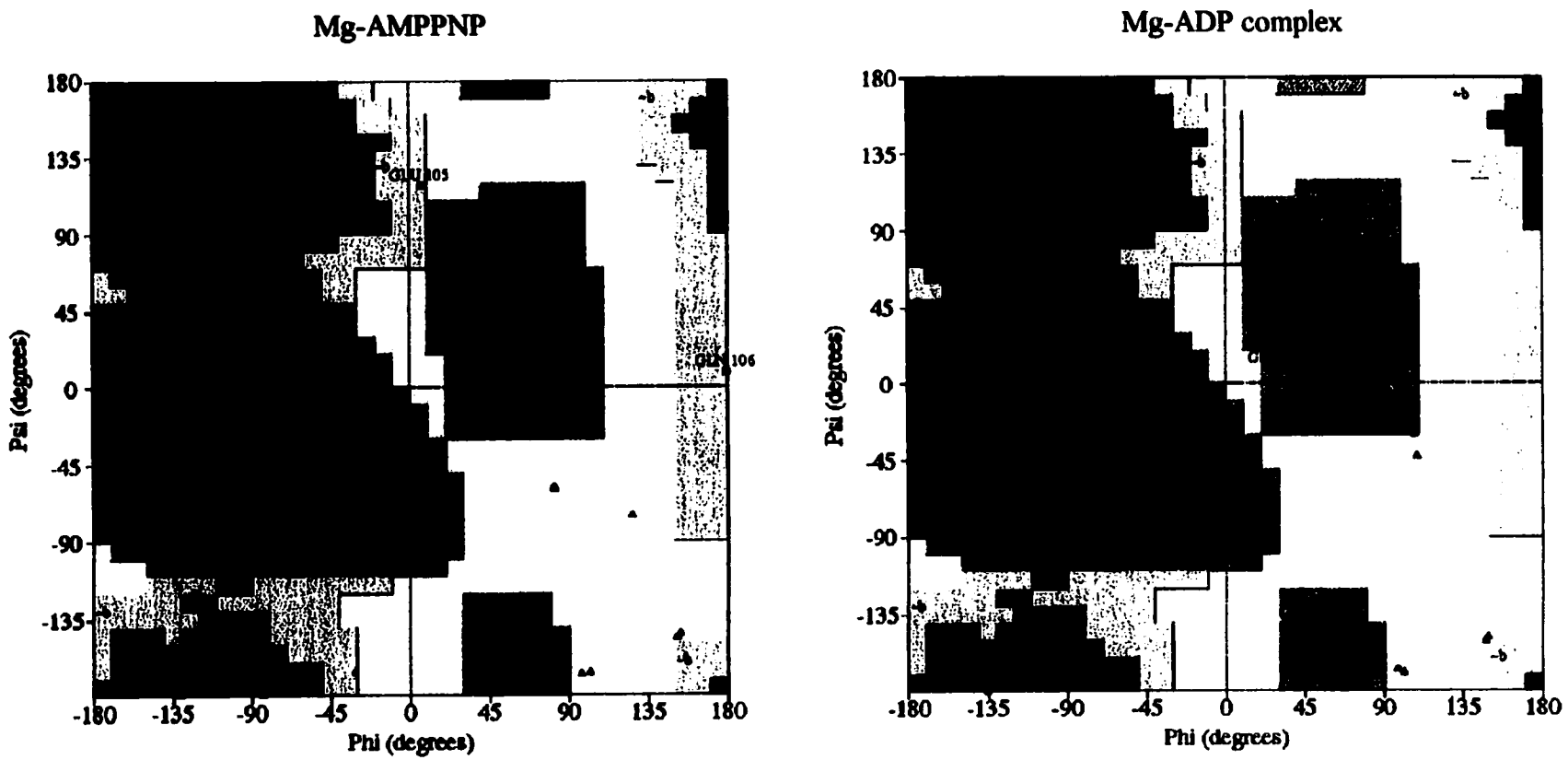
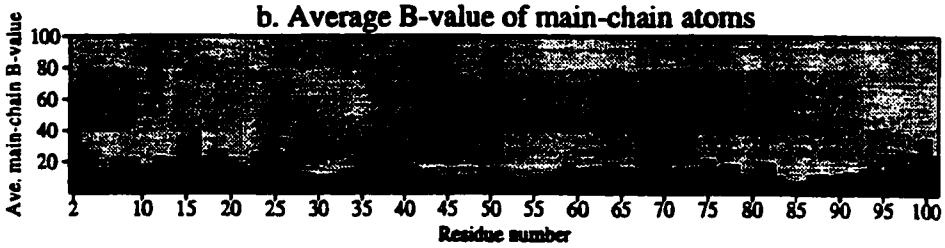
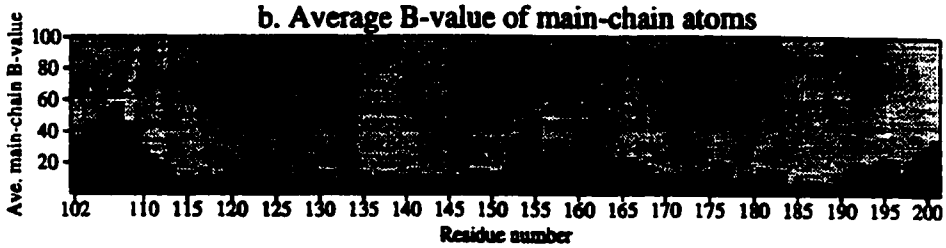


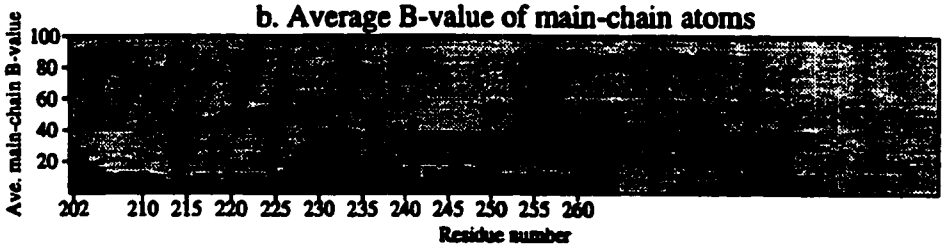
Figure 4-3a. Ramachandran plots of the refined models of APH(3')-IIIa complexed with Mg-AMPPNP and Mg–ADP. The plots were generated by the program PROCHECK (Laskowski, MacArthur, Moss & Thornton, 1993). The glycine residues are represented by solid triangles.

A

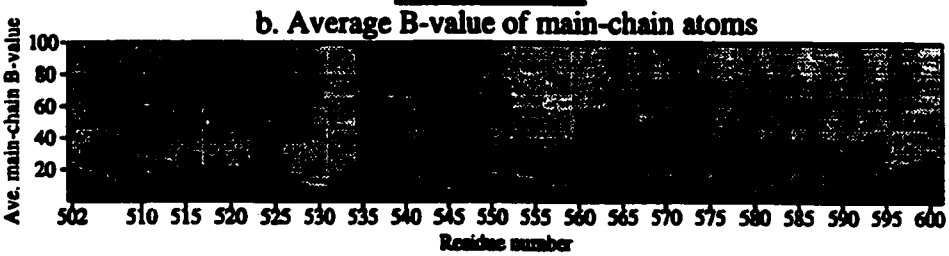
Monomer A

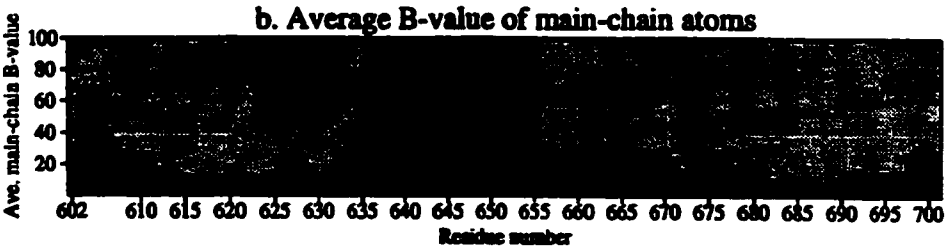


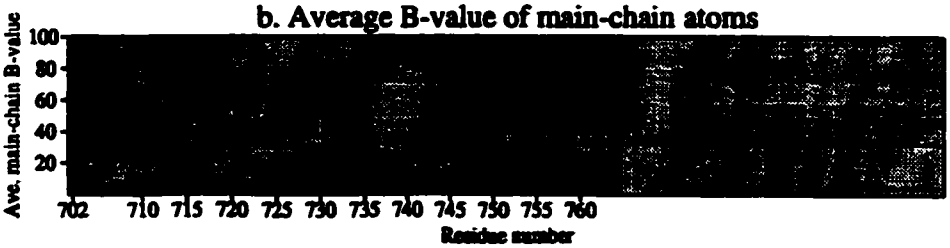




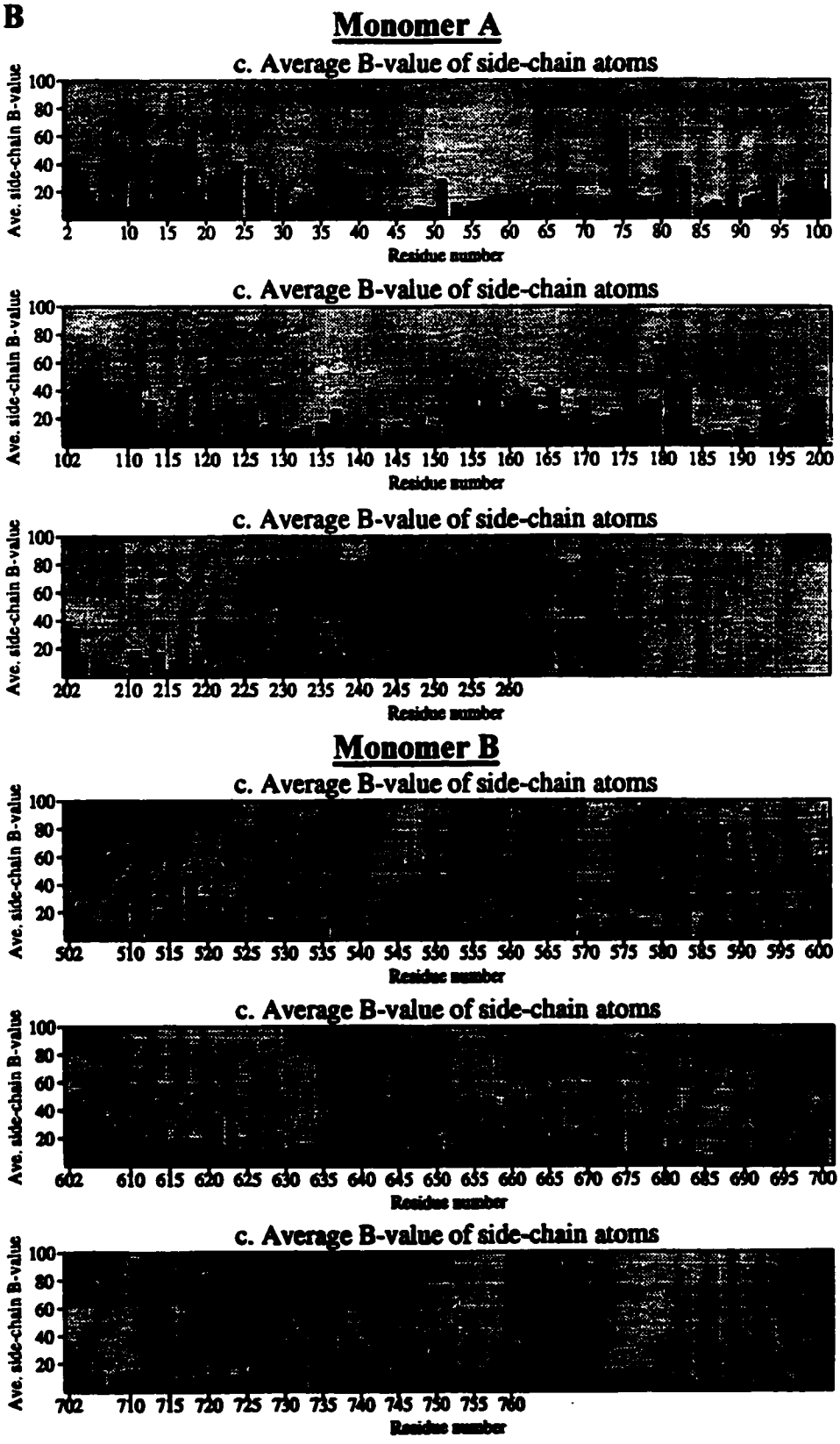
Monomer B



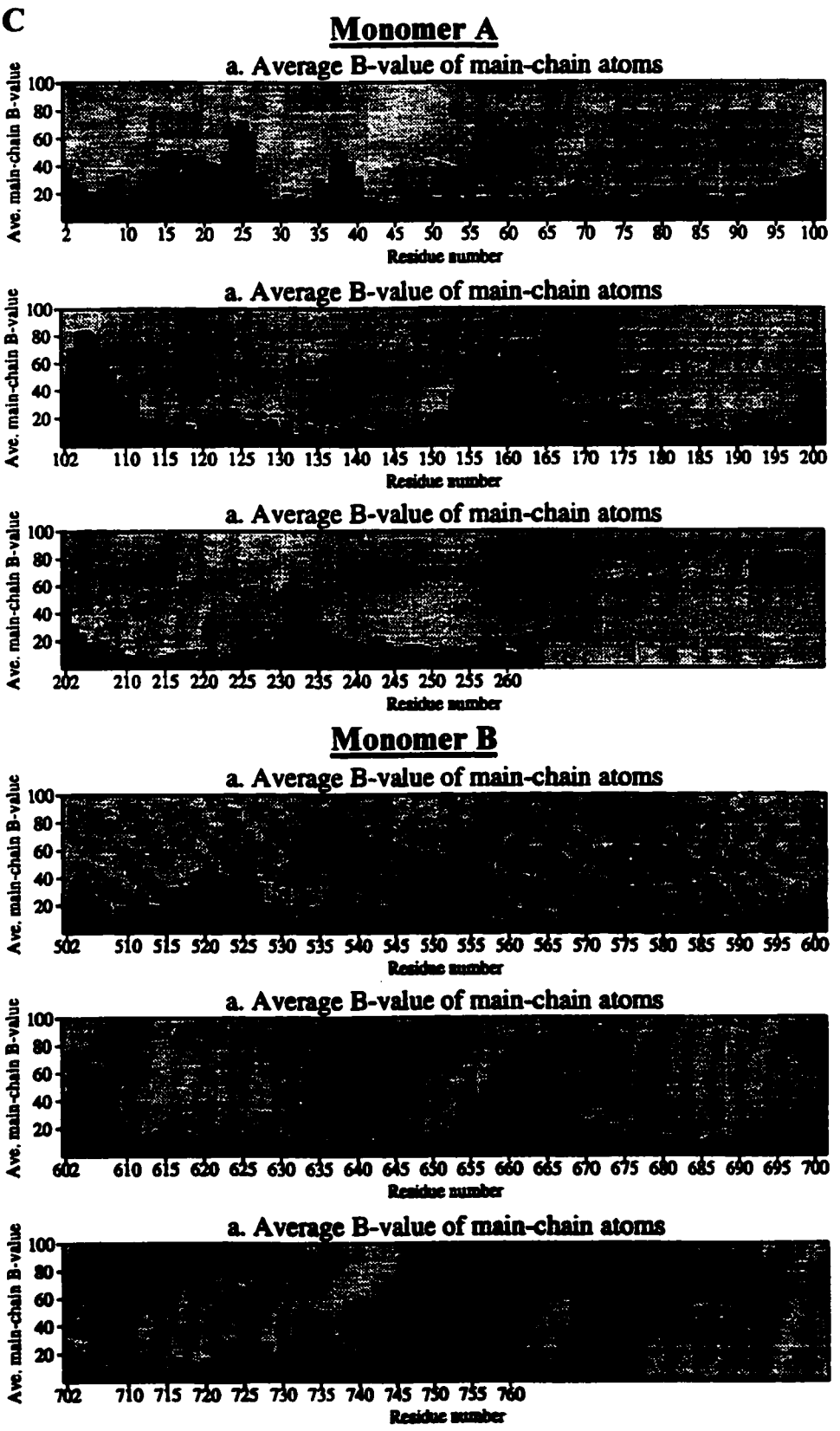




B


Monomer A
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
Monomer B
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
c. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number

C


Monomer A
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
2 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100
Residue number
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
102 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200
Residue number
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
202 210 215 220 225 230 235 240 245 250 255 260
Residue number
Monomer B
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
502 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600
Residue number
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
602 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700
Residue number
a. Average B-value of main-chain atoms
Ave. main-chain B-value
100
80
60
40
20
702 710 715 720 725 730 735 740 745 750 755 760
Residue number

D

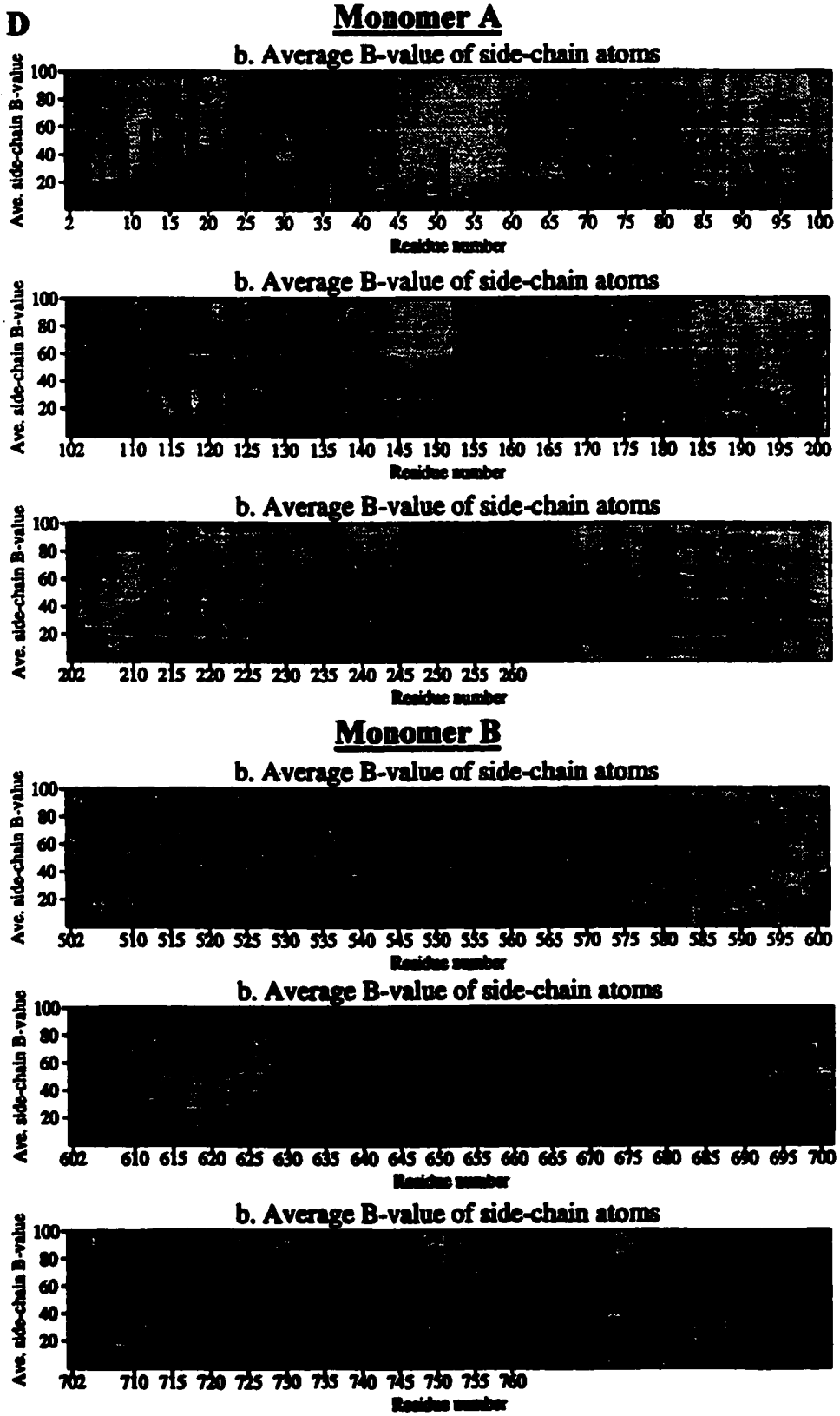

Monomer A
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
Monomer B
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number
b. Average B-value of side-chain atoms
Ave. side-chain B-value
Residue number

Figure 4-3b. Plots of average B-factors of main-chain and side-chain atoms of APH(3')-IIIa complexed with Mg-ADP (**A** and **B**) or complexed with Mg-AMPPNP (**C** and **D**). Residue numbers of monomer B are augmented by 500 with respect to those of monomer A. This figure was prepared with the program PROCHECK (Laskowski, MacArthur, Moss & Thornton, 1993).

APH(3')-IIIa will be given in curly brackets.

Figure 4-4 shows the overall architecture of the APH(3')-IIIa monomer. The protein has a bi-lobal shape where the two lobes are joined by a 12-residue linker (residues 95-106). The N-lobe (residues 2-94) mainly consists of a five-stranded anti-parallel β-sheet whereas the C-lobe (residues 107-264) is dominated by six α-helices. The deep cleft between the two lobes forms the nucleotide-binding site. This folding pattern is typical of the catalytic cores of Ser/Thr or Tyr protein kinases. Crystal structures of at least ten members have been determined to date (Knighton, *et al.*, 1991; De Bondt, *et al.*, 1993; Hu, *et al.*, 1994; Hubbard, Wei, Ellis & Hendrickson, 1994; Zhang, Strand, Robbins, Cobb & Goldsmith, 1994; Owen, Noble, Garman, Papageorgiou & Johnson, 1995; Goldberg, Nairn & Kuriyan, 1996; Longenecker, Roach & Hurley, 1996; Mohammadi, Schlessinger & Hubbard, 1996; Sicheri, Moarefi & Kuriyan, 1997; Xu, Harrison & Eck, 1997). Figure 4-5 shows the folding pattern of a representative subset of these structures. The similarity is remarkable considering that APH(3')-IIIa shares only less than 6% of sequence identity with these proteins in general (figure 4-1). The seven invariant residues are all located in the nucleotide-binding site (see below). There are two major segments of the protein molecule, comprising 130 residues in total, that have topological equivalence in protein kinases (figures 4-1 & 4-4). The term "topological equivalence" refers to structural elements that have the same secondary structure composition, overall architecture and sequence connectivity in the two types of kinases (Orengo, *et al.*, 1997). The first segment is composed of the majority of the N-lobe, including the β-strands β_A1-3, the helix H3 and β_A4-5, the domain linker and the first helix of the C-lobe, H5. The second segment consists of the C-lobe half of the highly conserved nucleotide-binding site.

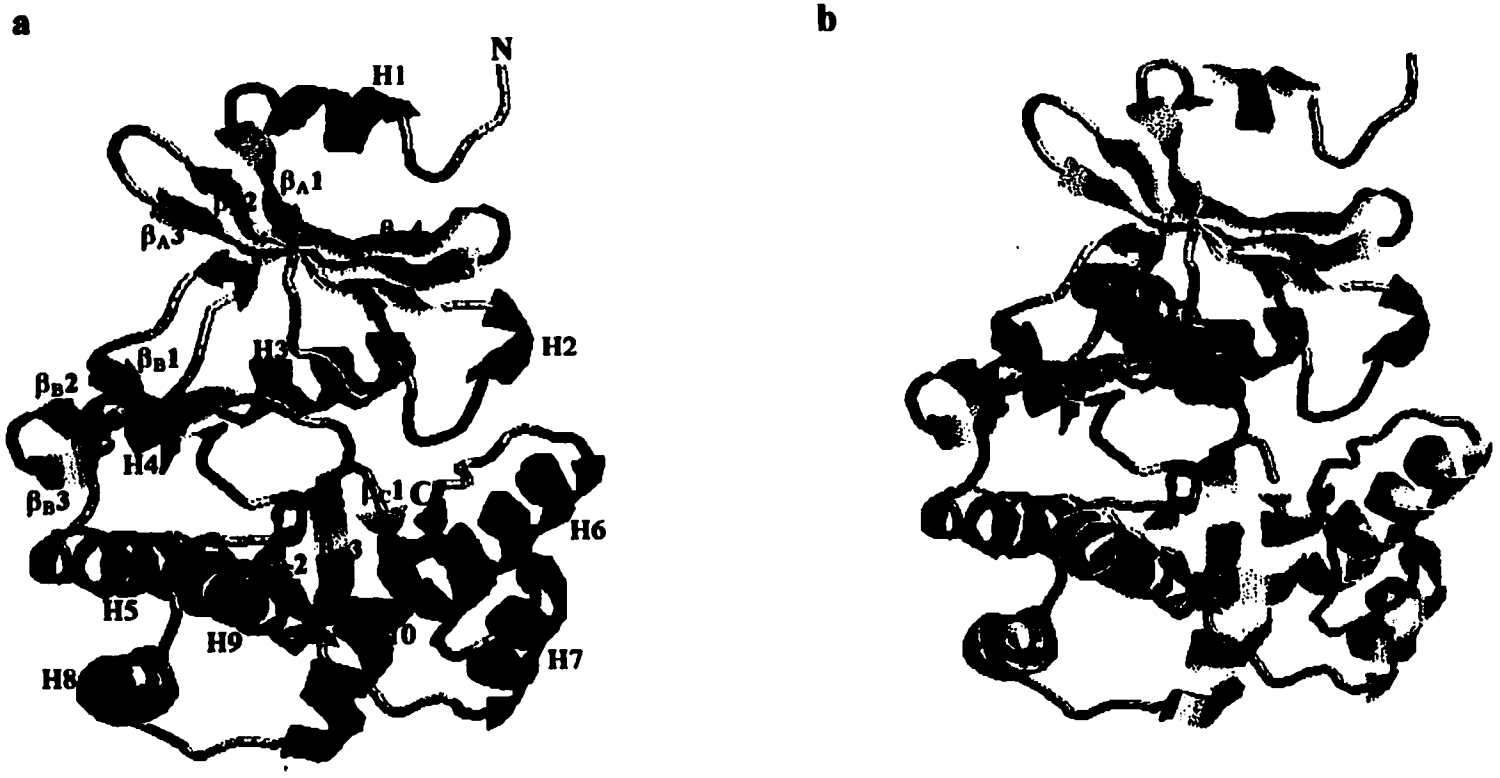
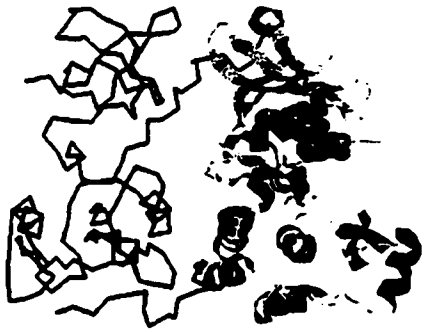


Figure 4-4. The overall architecture of APH(3')-IIIa complexed with Mg-AMPPNP. (a) The secondary structure elements are coloured in the following scheme: magenta for α- and 3_{10}- helices, yellow for β-strands, blue for turns and white for others. (b) The subdomains P-loop, domain linker and the putative antibiotic-binding loop are highlighted in green, greenblue and cyan, respectively. The Mg-AMPPNP molecule is shown as space-filling model. The carbon, nitrogen, oxygen, phosphorus and Mg atoms are coloured in grey, blue, red, orange and greenblue, respectively. Except for the putative antibiotic-binding loop, only portions of the structure that have topological equivalents in the catalytic cores of Ser/Thr or Tyr protein kinases are highlighted in colour. This view is hereafter adopted as the front view of the protein molecule.

HCK / Ca-AMPPNP



TWK-43

IRK

PHK / Mn-AMPPNP

cAPK / Mn-ATP

ERK / Mg-ATP

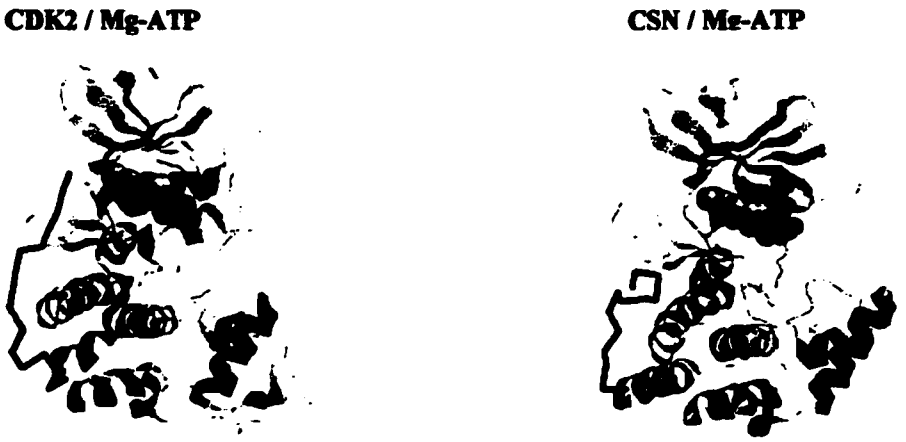


Figure 4-5. Ribbon diagrams illustrating the crystal structures of representative Ser/Thr or Tyr protein kinases. The sequences of these protein kinases are shown in figure 4-2. The conserved subdomains are colour-coded as follows: green (P-loop), greenblue (domain linker), cyan (activation segment). Domains or subdomains that are not part of the conserved catalytic core are shown as Cα trace coloured in grey. The N- and C-termini are highlighted in blue and red, respectively. The nucleotide substrates or analogs are presented as space-filling models. PDB ID's are 1AD5 (HCK), 1IRK (IRK), 1ATP (cAPK), 1KOB (TWK-43), 1PHK (PHKγ), 1GOL (ERK), 1HCK (CDK2) and 1CSN (CSN). This figure and all subsequent figures of protein structures were prepared with the graphic program RASMOL (Sayle & Millner-White, 1995). Common helices in the C-lobes are labelled in the diagram of cAPK following the nomenclature adopted in the original report (Knighton, *et al.*, 1991).

It is flanked by the conserved sequence motifs HGD and DXG (residues 188-190 and 208-210), which are equivalent to the HRD and DFG motifs {164-166, 184-186} in the catalytic cores of protein kinases. Part of this segment acts as the base of the nucleotide-binding site and part of it interacts with the domain linker through β_B1 and β_B2. In both types of kinases, the HXD and DXG motifs are located on loop regions. The loop on which the HRD motif is situated has been designated as the "catalytic loop" (Knighton, *et al.*, 1991). The Asp residue is invariant and has been proposed to serve as a catalytic base in the phosphoryl transfer mechanism (Knighton, *et al.*, 1991; Bossemeyer, Engh, Kinzel, Ponstingl & Huber, 1993; Zheng, *et al.*, 1993) (see discussion later).

The insertion between the two segments are unique to the APH(3') isozymes. This part includes the lower half of the "domain joint" and the putative "antibiotic-binding loop", both of which will be described in detail later. Beyond the nucleotide-binding segment in the C-lobe, topological similarities between the two types of kinases end. The two-stranded anti-parallel β-sheet (β_C1 and β_C2), which is connected to both ends of the C-lobe nucleotide-binding segment (figure 4-4), is unique to the structure of APH(3')-IIIa and its sequence is conserved among all APH(3') isozymes (figure 4-1). APH(3') isozymes lack the activation segment (also called activation-loop, T-loop or phosphorylation lip) (figure 4-2) that is believed to play an important role in regulating the activity of protein kinases through intramolecular steric hindrance (Johnson, Noble & Owen, 1996). This segment is flanked by the consensus motifs DFG and APE {184-186, 206-208} and contains the Thr and/or Tyr residues that are phosphorylated upon activation in certain protein kinases.

Overall, the N-lobe β-sheet in APH(3')-IIIa exhibits a much lesser degree of twist

than those observed in the catalytic cores of protein kinases (figure 4-6). The packing of the α-helices in the C-lobe also differs. In APH(3')-IIIa, the six α-helices in the C-lobe are arranged in two bundles of three (H5,H8,H9 and H6,H7,H10) that are oriented approximately orthogonal to each other (figure 4-4). The two helix-bundles are positioned on either side of β_C1 and β_C2, giving the C-lobe a semi-symmetrical appearance. In the catalytic core of protein kinases, the number of α-helices in the C-lobe varies but the common core is represented by five helices. Four of these {αE, αF, αH and αI} pack as a four-helix bundle situated on the side of the domain linker, approximately orthogonal to the fifth helix {αG} which is on the side of the activation segment (figure 4-2).

The APH(3')-IIIa-nucleotide complex crystallized as a dimer that is linked by two disulphide bridges (Cys19 on monomer A to Cys156 on monomer B, and *vice versa*) in a head-to-tail fashion (figure 4-7). The intermolecular contacts involve regions flanking the first helix (residues 3-19) and the putative "antibiotic-binding loop" (residues 151-166). A total of 14 residues take part in forming one disulphide bond, eight hydrogen bonds and nine van der Waals contacts (table 4-2), resulting in the burial of 2900 $Å^2$ solvent accessible surface area, which is roughly 10% of the added solvent accessible surface area of the two monomers. Freshly purified APH(3')-IIIa exists both as monomers and dimers (McKay, Thompson & Wright, 1994). Dimer formation seems to be favoured by the high protein concentration required for crystallization. Results from steady state kinetic analysis showed that the kinetic parameters, K_m and k_{cat}, as exhibited by the monomeric and dimeric forms were undifferentiable within experimental error (McKay, Thompson & Wright, 1994). Moreover, the disulphide Cys residues are not conserved among other APH(3') isozymes (figure 4-1). In the crystal structure of the apoenzyme, the protein

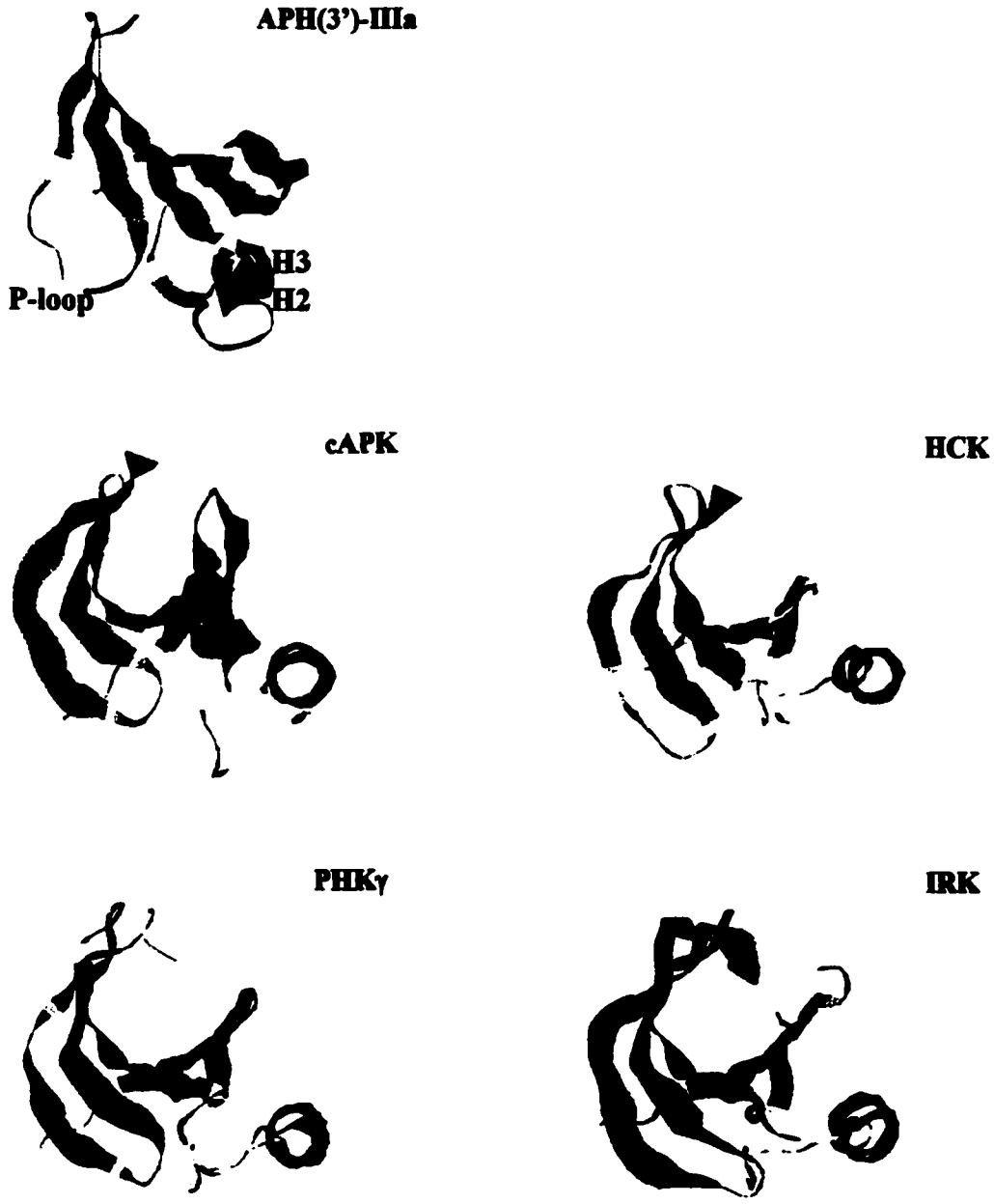


Figure 4-6. Comparison of the degree of twist of the N-lobe β-sheets of APH(3')-IIIa and representative catalytic cores of protein kinases. cAPK and Phkγ belong to the Ser/Thr protein kinase subfamily and HCK and IRK belong to the Tyr protein kinase subfamily. The β-sheets are viewed normal to the helix axis of H3 in APH(3')-IIIa and its equivalents.

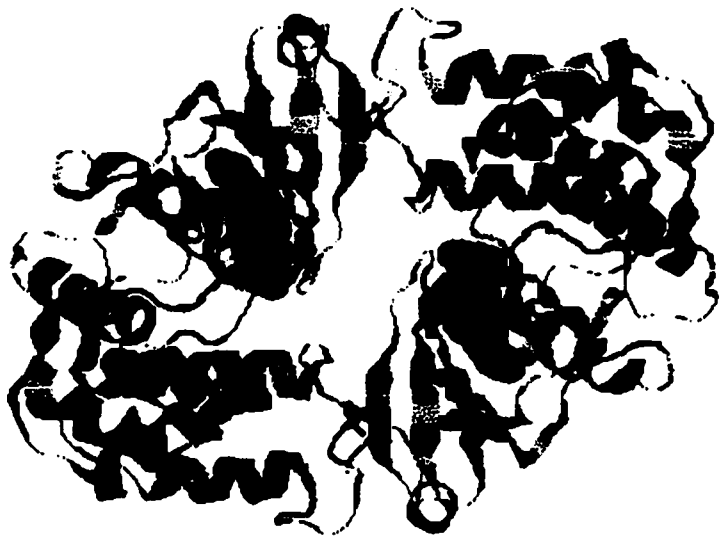


Figure 4-7. APH(3')-IIIa dimer viewing parallel to the noncrystallographic two-fold axis. The disulphide bridges are shown in stick form coloured in deep blue. The N-lobe, domain linker and C-lobe are coloured in orange, greenblue and violet, respectively.

Table 4-2. Dimeric interactions in APH(3')-IIIa.

Hydrogen bonding				
Monomer 1		Monomer 2		Distance (Å)
Lys3	NZ	Asn148	O	2.8
Lys3	NZ	Leu147	O	2.9
Arg5	NH2	Asp150	O	3.0
Arg5	NH2	Asp150	OD1	3.0
Arg5	NH2	Asp150	O	3.0
Arg5	NH2	Asp150	OD1	3.0
Arg5	NE	Asp150	OD1	2.9
Lys11	NZ	Lys166	O	2.7
van der Waals interactions				
Monomer 1		Monomer 2		Distance (Å)
Arg5	CZ	Val154	CG2	3.6
Arg5	CD	Val154	CG2	3.8
Cys19	SG	Cys156	CB	2.8
Cys19	CB	Cys156	SG	3.0
Cys19	CB	Cys156	CB	3.3
Ile6	CD1	Trp159	CD2	3.8
Ile6	CD1	Trp159	CG	3.9
Met4	CE	Trp159	CD1	3.9
Met4	CE	Pro168	CD	3.9

Analysis of these interactions was performed by the program LIGPLOT (Wallace, Laskowski & Thornton, 1995).

molecules are linked by disulphide bonds in a continuous fashion such that each subunit is linked to two other subunits (Adelaine Leung, personal communication). Therefore, the formation of dimeric APH(3')-IIIa is likely to be fortuitous and may not have any functional significance.

The P-loop

The consensus sequence GXSXXXV/I of the APH(3') isozymes (figure 4-1) forms a flap that covers the phosphate-binding subsite in the structures of APH(3')-IIIa (figure 4-4). This structural motif, flanked between β_A1 and β_A2, is equivalent to the G-rich P-loops (phosphate-recognition loop) in the catalytic cores of protein kinases, which contain the consensus sequence GXGXXGXV (figures 4-3). In the structures of the catalytic cores of protein kinases, this loop sometimes adopts the conformation of a loop (e.g. ERK2 (Zhang, Strand, Robbins, Cobb & Goldsmith, 1994)), and sometimes as β-turn-β (e.g. cAPK (Zheng, *et al.*, 1993) and Phkγ (Owen, Noble, Garman, Papageorgiou & Johnson, 1995)) (figure 4-5). Nonetheless, like the P-loop in APH(3')-IIIa, this region is characteristically flexible in almost all of the reported structures of the protein kinase catalytic cores. Its conformation can vary according to the molecule bound in the nucleotide-binding site. A vivid example is demonstrated by the crystal structures of the Tyr protein kinase domain of the fibroblast growth factor complexed to a family of indolinone inhibitors (Mohammadi, *et al.*, 1997). Its main role seems to be less in anchoring the nucleotide than in shielding the γ-phosphate from the bulk solvent.

The G-rich P-loop is not unique to protein kinases, but is common among nucleotide-binding proteins, including nucleotidyltransferases and ATPases (Walker,

Saraste, Runswick & Gay, 1982; Smith & Rayment, 1996). It has been nicknamed "Walker's sequence 1" and is generally used as a landmark for the identification of ATP-binding domains in proteins. Such prevalent use led to an earlier erroneous prediction of the G-rich region, GXXDXGRXG (residues 205-213), in the C-lobe of the APH(3') isozymes as the P-loop (Shaw, Rather, Hare & Miller, 1993). In certain types of kinases, such as adenylate kinase, the consensus sequence contains a basic residue, Lys or Arg, for interacting with the phosphates of ATP. The equivalent of this basic residue, Lys44, is situated on β_A3 of APH(3')-IIIa.

Nucleotide-binding site

This is the region that displays the highest degree of structural resemblance to the catalytic cores of protein kinases. The fold is unique among nucleotide-binding proteins. It is assembled as two parts, the N- and the C-lobes each contribute half, forming a large, relatively hydrophobic cavity in which the nucleotide buries. For both the Mg-APMPNP and Mg-ADP complexes, the σ_A-weighted $2F_o$-F_c electron density maps clearly reveal density for the entire molecule of the nucleotides and two Mg^{2+} ions (figure 4-8). Out of the eight sets of coordinates of protein kinase-nucleotide complexes that are currently available in the Brookhaven Protein Data Bank (PDB), only cAPK (PDB ID 1ATP, as ternary complex of Mn-ATP/peptide inhibitor (PKI(5-24)) (Zheng, *et al.*, 1993) and 1CDK (Bossemeyer, Engh, Kinzel, Ponstingl & Huber, 1993), as ternary complex of Mn-AMPPMP/PKI(5-24)) and Phkγ (PDB ID 1PHK (Owen, Noble, Garman, Papageorgiou & Johnson, 1995), as binary complex of Mn-ATP) have two divalent cations coordinated to the nucleotide phosphate groups. Mn1 bridges the α- and the γ-phosphates to the invariant

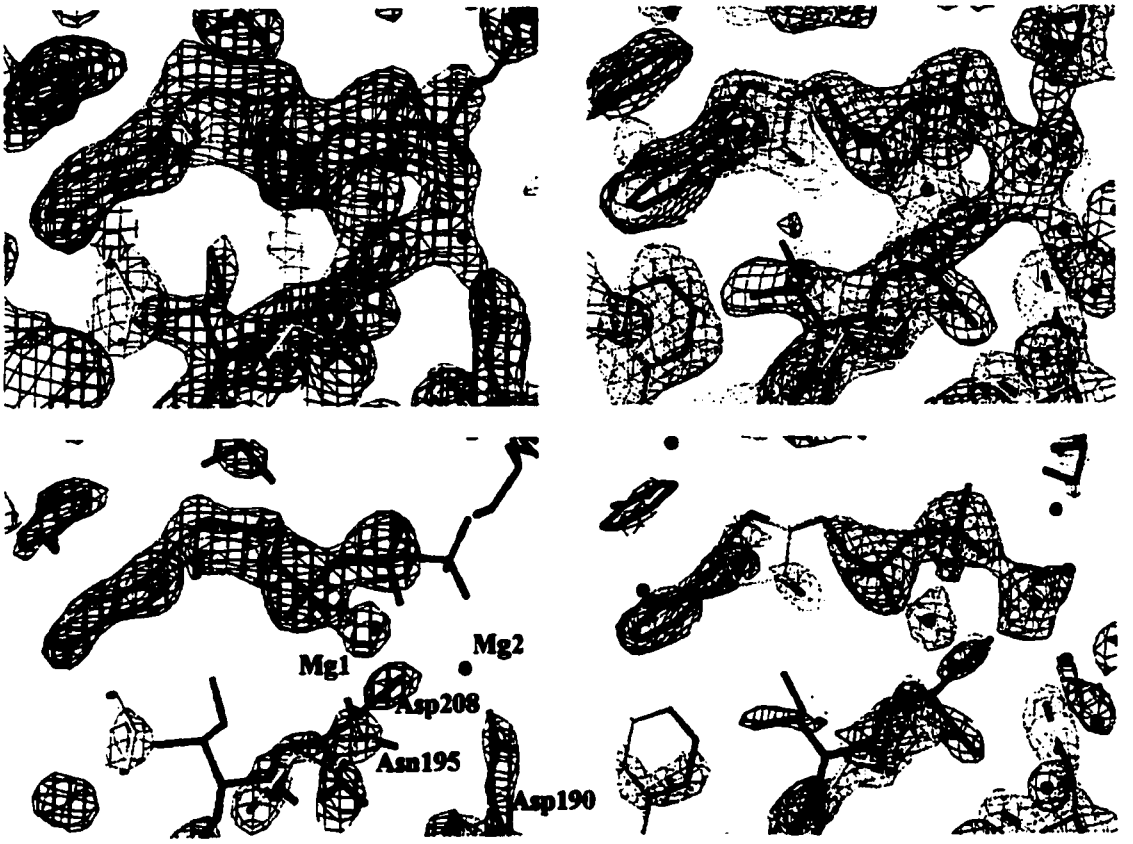


Figure 4-8. Sections of σ_A-weighted, simulated annealing $2F_o$-F_c omit maps representing the nucleotide-binding site in APH(3')-IIIa. The maps were calculated with the Mg-ADP or Mg-AMPPNP and the surrounding 5.0 Å sphere of atoms omitted, and contoured starting at 1 sigma level (top panel) and 2 sigma level (bottom panel). The Mg^{2+} ions and the water oxygens are shown as green and red spheres, respectively. This figure was prepared with the program O (Jones, Zou, Cowan & Kjeldgaard, 1991).

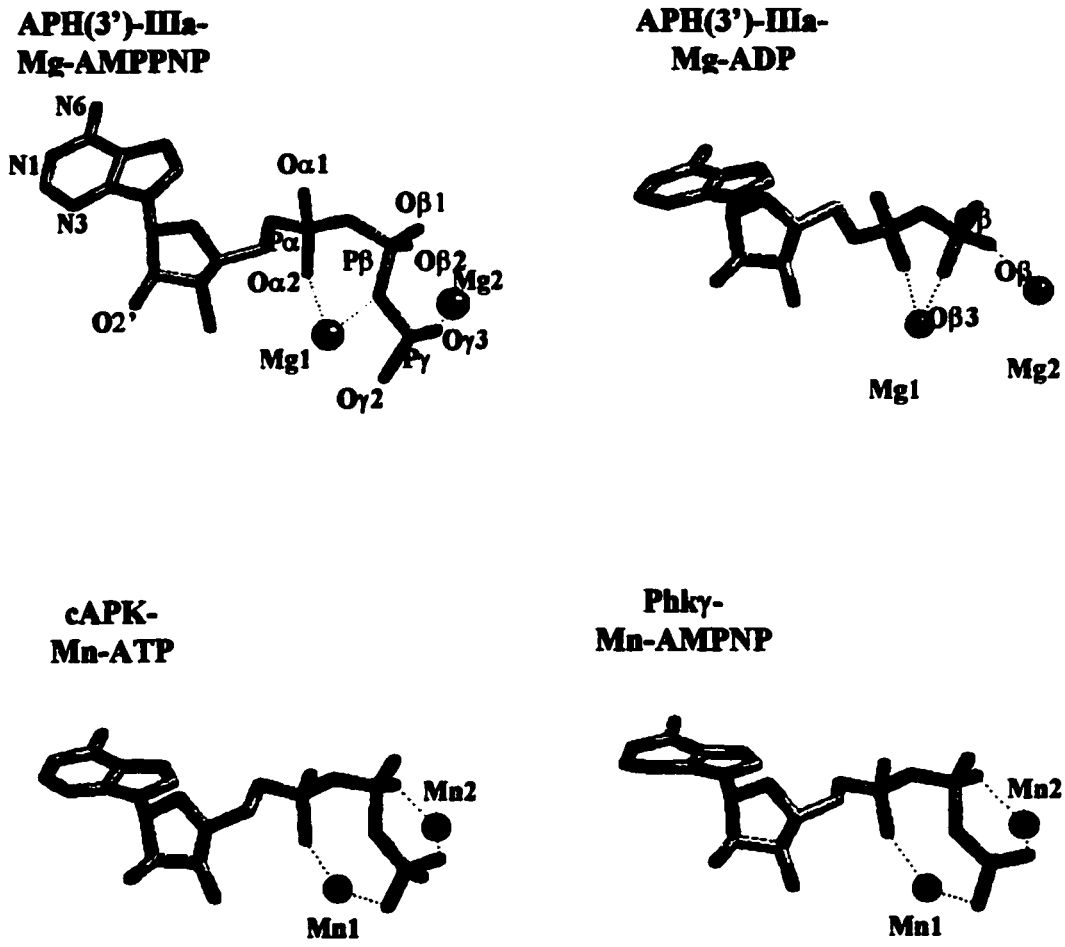


Figure 4-9. Conformation of the bound nucleotides and nomenclature of the phosphate and magnesium atoms. The carbon, nitrogen, oxygen, phosphorus and Mg (or Mn) atoms are coloured in grey, blue, red, orange and greenblue, respectively. Note that the nucleoside-triphosphate reported in the crystal structure of Phkγ was AMPPNP (Owen, Noble, Garman, Papageorgiou & Johnson, 1995) instead of ATP as disclosed in the available coordinate file (PDB ID 1PHK).

{Asn171} whereas Mn2 bridges the β- and the γ-phosphates to the invariant {Asp184}. The other five structures have one coordinated divalent cation and their modes of phosphate recognition differ. All three phosphate groups of AMPPNP in the structure of the human hematopoietic cell kinase (HCK, PDB ID 1AD5, (Sicheri, Moarefi & Kuriyan, 1997)) were coordinated to one Ca^{2+} ion, which is unlikely to be the physiological metal ion. In the structures mentioned below, only the metal ion centre equivalent to Mn1 in cAPK and Phkγ is occupied. Structure of the K52R mutant of mitogen-activated protein (MAP) kinase (ERK2, PDB ID 1GOL) forms a nonproductive complex with ATP (Robinson, *et al.*, 1996), whereas the γ-phosphate of Mg-ATP in the structure of the inactivated form of the human cyclin-dependent kinase 2 (CDK2, PDB ID 1HCK) was positioned inward in the nucleotide-binding site (Schulze-Gahmen, De Bondt & Kim, 1996). The Mg-ATP complex of yeast casein kinase I (CSN, PDB ID 1CSN) and the Mn-ATP complex of the activated form of CDK2 (PDB ID 1JST) has Mg1 (see figure 4-9 for nomenclature). The β- and γ-phosphate groups in these two structures have extremely high thermal factors, ranging from 80-100 $Å^2$. The presence of only one metal ion is likely a result of sub-optimal crystallization conditions. Hence, these five structures do not provide obvious structural evidence for the reasoning of possible phosphoryl transfer mechanisms. Comparisons regarding the mode of phosphate-binding will primarily be made with the structures of cAPK and Phkγ nucleotide complexes, using the PDB coordinates 1ATP and 1PHK, respectively. The nucleotide-binding site is customarily divided into the purine-, ribose- and phosphate-binding subsites. The molecular interactions observed in these subsites will be described separately in the following. The implications for the phosphoryl transfer mechanism as can be projected from the atomic

details in the phosphate-binding subsite will be discussed afterwards.

The bound nucleotide

The conformation of the bound AMPPNP in APH(3')-IIIa is similar to those of cAPK and Phkγ nucleotide complexes (figure 4-9). The conformation about the glycosyl C1'-N9 linkage is *anti* with χ=-166° and -162° for Mg-AMPPNP and Mg-ADP, respectively. The ribose adopts a C2'-exo, C3'-endo puckering conformation and the orientation about the C4'-C5' is +sc (γ=-58°) for the Mg-AMPPNP complex and –sc (γ=-135°) for the Mg-ADP complex. The nomenclature described here follows that of Saenger (Saenger, 1984).

The purine-binding subsite

The adenine base is deeply buried and forms multiple van der Waals interactions with Val31 (P-loop), Tyr42, Glu92 (N-lobe), and Phe197 (C-lobe). The N6 atom forms a hydrogen bond with the backbone carbonyl oxygen of Ser 91. Stacking of the base with an aromatic side-chain is unique to the structure of APH(3')-IIIa and would likely occur in all other APH(3') isozymes since Tyr42 is conserved as either Tyr or Phe (figures 4-1 & 4-10). In the structures of the catalytic cores of protein kinases, Tyr42 corresponds to the conserved Ala {Ala70}. N1 of the adenine base is also involved in hydrogen bonding with the backbone amide nitrogen of Ala93 in the domain linker region. This is similar to most of the protein kinase-nucleotide complex structures (Zheng, *et al.*, 1993; Owen, Noble, Garman, Papageorgiou & Johnson, 1995; Robinson, *et al.*, 1996; Schulze-Gahmen, De Bondt & Kim, 1996).

The ribose-binding subsite

The ribose ring forms van der Waals interactions with Asp22, Val31 and Ile207. A water molecule forms hydrogen bonds to both the O2' hydroxyl group of the ribose and the N3 atom of the adenine base. The O3' hydroxyl group forms a hydrogen bond with the backbone carbonyl O of Ser194. This is unlike the aforementioned nucleotide complexes of protein kinases in which the ribose rings are anchored to the protein molecule mainly through hydrogen bonding via the O2' and O3' hydroxyl groups. The hydrogen-bonding partners are mostly provided by the side-chain of an Asp or a Glu on the domain linker (in cAPK, Phkγ, ERK2 and CDK2), and a backbone carbonyl oxygen situated on the catalytic loop (in cAPK, Phkγ and CDK2).

The phosphate-binding subsite

The phosphate groups of Mg-AMPPNP and Mg-ADP are anchored in the nucleotide-binding site through an elaborate network of hydrogen and ionic bonds so that the γ-phosphate is positioned towards the entrance of the active site. The entire recognition pattern for the nucleoside triphosphate is very similar to those observed in the structures of cAPK (Zheng, *et al.*, 1993) and Phkγ (Owen, Noble, Garman, Papageorgiou & Johnson, 1995). It involves five of the seven invariant residues that are conserved between the APH(3') isozymes and the catalytic cores of protein kinases, including Lys44 and Glu60 {72 and 91} from the N-lobe, and Asp190, Asn195 and Asp208 {166, 171 and 184} (hereafter referred to as the D-N-D triad) from the C-lobe (figure 4-1). Figure 4-10 illustrates the interactions involved in phosphate-binding in the context of the surrounding

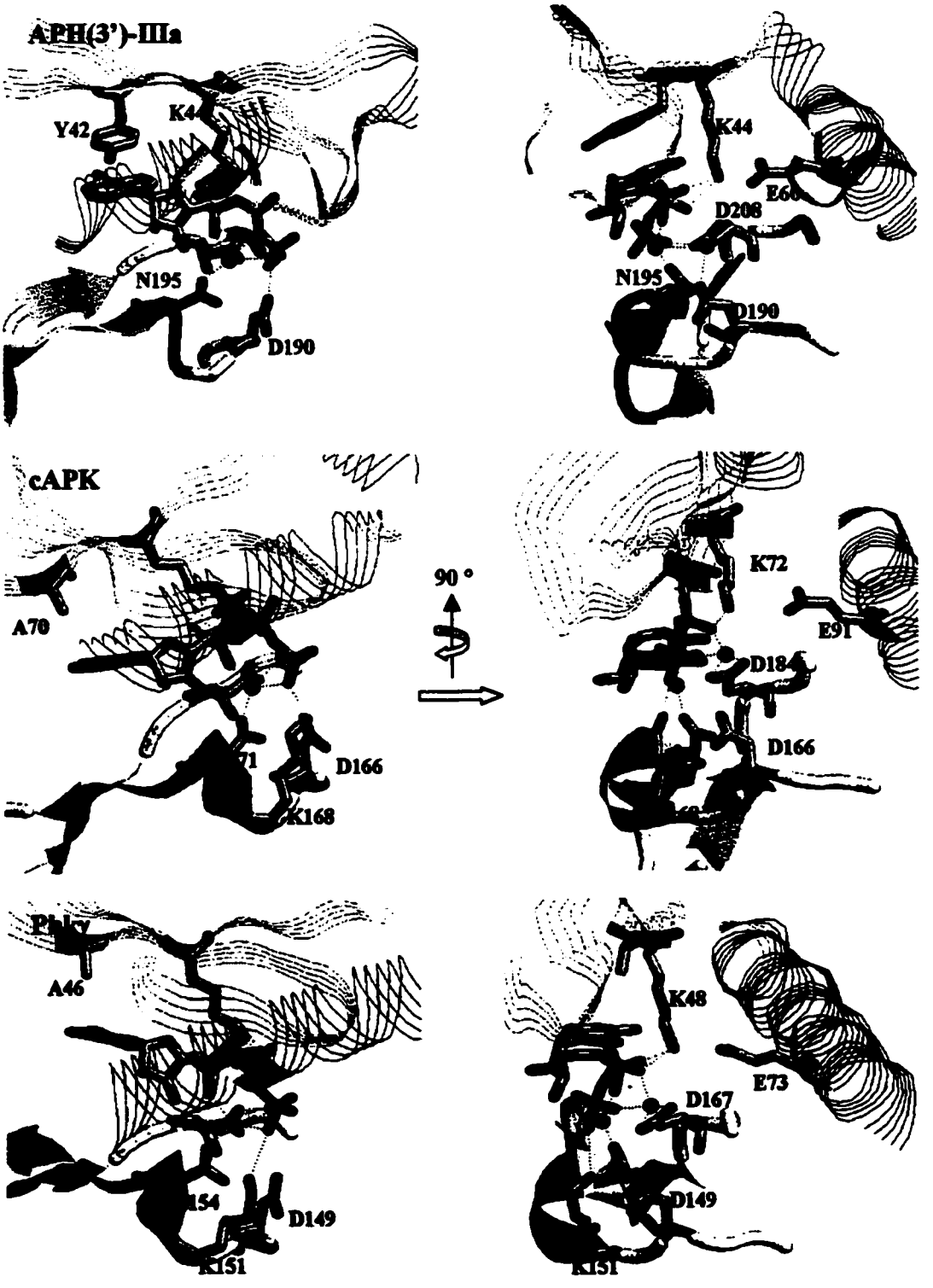

APH(3')-IIIa
Y42
K4
N195
D190
K44
E6
D208
N195
D190
cAPK
A70
D166
K168
90°
K72
E91
D184
D166
A46
D149
K48
E73
D167
D149

Figure 4-10. Phosphate recognition in the nucleotide-binding sites of APH(3')-IIIa, cAPK and Phkγ. The left and right panels display the front view and an orthogonal side view, respectively. Secondary structure elements in the N-lobe (including β_A1, β_A2, β_A3 and H3 or their equivalents in cAPK and Phkγ) are shown as strands whereas those in the C-lobe (regions spanning the HXD and DXG motifs) are shown as solid ribbons.

secondary structures and table 4-3 summarizes the atomic details of the interactions in APH(3')-IIIa, cAPK and Phkγ. The nomenclature of the atoms of both Mg-AMPPNP and Mg-ADP are provided in figure 4-9.

The Lys-Glu pair

Lys44 {72} in the N-lobe forms the only direct interactions between the protein molecule and the phosphate groups. Its side-chain amino group is hydrogen-bonded to the α- and β-phosphate oxygens (Oα1 and Oβ2) in both the Mg-AMPPNP and the Mg-ADP complexes of APH(3')-IIIa. These interactions are secured by the side-chain of the invariant Glu60 {91}, which positions the side-chain of Lys44 {72} in place via a three-centered hydrogen bond (Jeffrey & Saenger, 1991) (table 4-3). Similar interactions are also observed in the Mn-ATP/PKI complex of cAPK and the Mn-AMPPNP complex of Phkγ.

The metal ions

The C-lobe interacts with the phosphate groups indirectly via the D-N-D triad (Asp190-Asn195-Asp208) which forms a template for two metal ion centres that can bridge all three phosphate groups to the protein molecule. In accordance with the convention adopted in the structures of cAPK and Phkγ, the two Mg^{2+} ions are designated as Mg1 and Mg2. In both the Mg-AMPPMP and the Mg-ADP complexes, Mg1 is positioned between the amide oxygen (Oδ1) of Asn195 {171} and the carboxyl oxygen (Oδ2) of Asp208 {184}. In the Mg-AMPPNP complex, Mg1 also links with the α-phosphate oxygen (Oα2) and the bridging N atom between the β- and γ-phosphate groups. The resulting coordination geometry is octahedral although the Mg^{2+}···O distances are longer than the usual 2.1 Å and electron density for two water molecules required to

Table 4-3. Summary of interactions in the conserved phosphate-binding subsites of APH(3')-IIIa, and the catalytic cores of cAMP-dependent protein kinase (cAPK) and glycogen phosphorylase kinase γ-subunit (Phkγ)[a].

Residue/atom[b]	Metal ion[c]	Nucleotide[d] / H_2O	Interatomic distance (Å)			
			APH(3')-IIIa-Mg-ADP[e]	APH(3')-IIIa-Mg-AMPPNP[e]	cAPK-Mn-ATP	Phkγ-Mn-AMPPNP
E60/Oε1,Oε2 {91, 73}			3.0, 3.0	2.8, 2.9	[4.3][f], 3.6	3.6, 2.8
vs						
K44/Nζ {72, 48}						
N195/Oδ1 {171, 154}			2.7	2.8	2.6	2.7
vs						
D190/O {166, 149}						
K44/Nζ {72, 48}		Oα1, Oβ2	3.2, 3.1	2.8, 3.5	2.5, 2.6	2.9, 2.9
	Mg1	Oα2, Oβ3[f], Oγ2	2.2, 2.2, -	2.1, [2.8], [3.0]	2.0, -, 2.2	2.4, -, 2.3
N195/Oδ1 {171, 154}	Mg1		2.3	2.4	2.1	2.3
D208/Oδ1 {184, 167}	Mg1		2.4	2.6	2.1	2.4
	Mg1	H_2O	2.2, 2.3	-	2.1	2.5
	Mg2	Oβ2, Oγ3	2.2, -	[2.9], 2.6	2.0, 2.2	2.4, 2.5
D208/Oδ1,Oδ2 {184, 167}	Mg2		2.4, 2.4	[2.7], 2.3	2.2, 2.4	2.5, 2.4
D190/Oδ2 {166, 149}	Mg2		-	2.6	[3.8]	[4.2]
	Mg2	H_2O	2.1, 2.1, 2.1	2.3	2.2, [1.8]	2.2

[a] PDB ID for the nucleotide complexes of cAPK and Phkγ are 1ATP and 1PHK, respectively. [b] Residue numbers equivalent to those of APH(3')-IIIa are given in curly brackets in the order of cAPK and Phkγ. [c] The metal ions in both the cAPK and Phkγ nucleotide complex structures are Mn^{2+}. [d] Nomenclature of nucleotide atoms are given in figure 4-10. [e] Distances in square brackets are considered longer than reasonable hydrogen or ionic bonding distances but are included for comparison. [f] Oβ3 in ADP is equivalent to the β-γ bridging oxygen in ATP.

complete the hexa-coordination sphere could not be located (figure 4-11). This observation is slightly different from those in the Mn-ATP/PKI(5-24) ternary complex of cAPK and the Mn-AMPPNP complex of Phkγ. Mn1 in these two structures bridges both the α- and γ-phosphate oxygens (Oα2, Oγ2) and has a trigonal bipyramidal coordination geometry (figure 4-12, table 4-3).

In the Mg-ADP complex of APH(3')-IIIa, Mg1 bridges with both the α- and β-phosphate oxygens (Oα2, Oβ3) where the latter is equivalent to the β,γ-bridging oxygen in ATP. In this case, the coordination geometry of Mg1 is a hexa-coordinated octahedron with three water molecules to complete the coordination sphere (figure 4-11). The displacement of Mg1 upon superposition of the Mg-AMPPNP and the Mg-ADP complexes is 0.5 Å.

In the Mg-AMPPNP complex of APH(3')-IIIa, Mg2 is coordinated to the γ-phosphate oxygen (Oγ3), the carboxyl oxygen (Oδ2) of Asp208 {184} that is closest to the γ-phosphate, the carboxyl oxygens (Oδ2) of Asp190 {166} that is also closest to the γ-phosphate, and one water molecule. The coordination geometry centering on Mg2 is octahedral although two of the three water molecules required to complete the coordination sphere could not be located (figure 4-11). The electron density surrounding Mg2 in this complex is also not as discreet as that surrounding Mg1 or Mg2 in the Mg-ADP complex (figure 4-8). This implies that Mg2 may not be stationary and its refined position might be a time-average of more than one position. An interesting observation is that the position of Mg2 is such that it can potentially form a "direct bonding" with both carboxyl oxygens of Asp208 {184}, i.e. it is equidistant from the two carboxyl oxygens (figure 4-13).

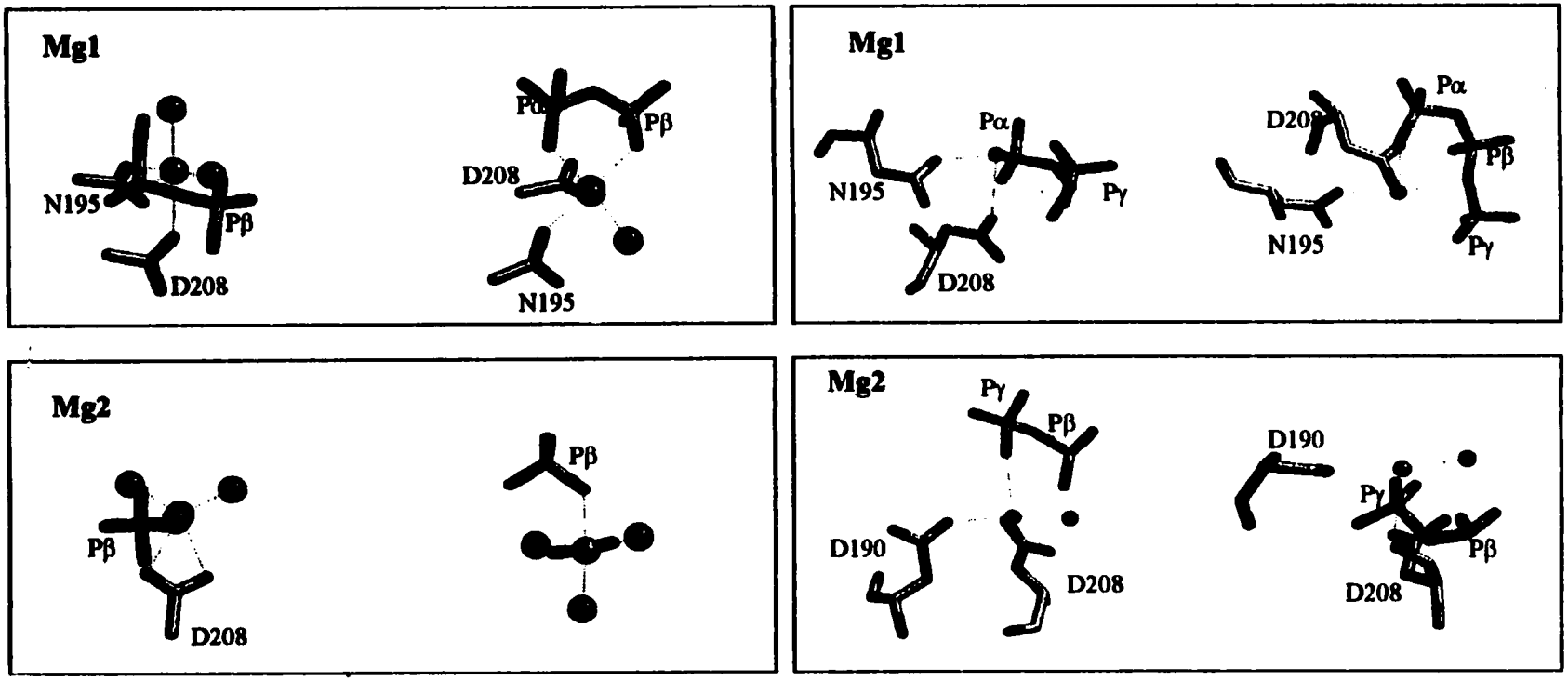
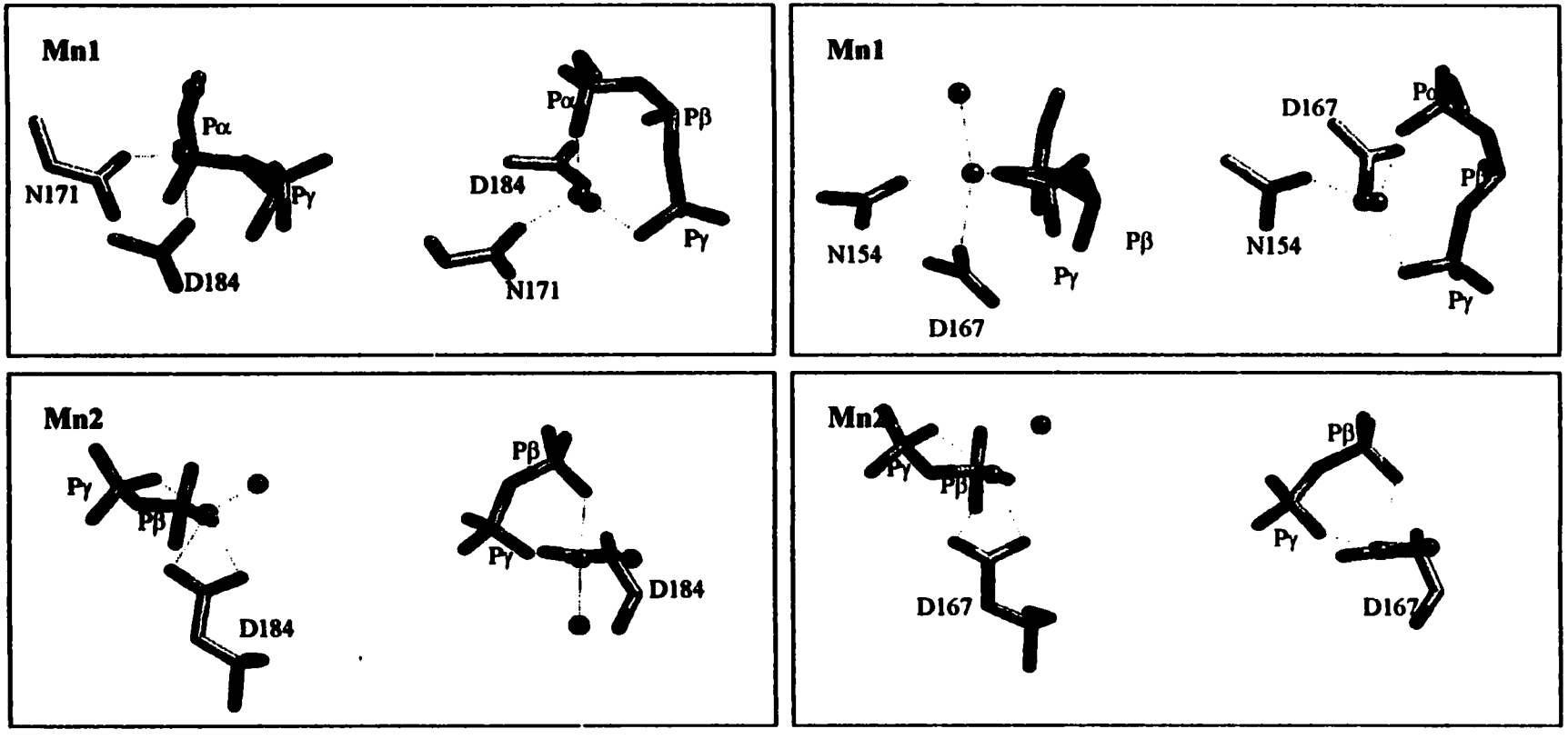


Figure 4-11. Coordination geometry of the two Mg^{2+} ions in the nucleotide-binding sites of APH(3')-IIIa complexed with Mg-ADP and Mg-AMPPNP. Two orthogonal views are presented for each ion. The Mg^{2+} ions and the oxygen atoms of water molecules are displayed as greenblue and red spheres, respectively. See also table 4-3 for a summary of the distances involved.



Mn-ATP complex of cAPK **Mn-ATP complex of Phkγ**

Figure 4-12. Coordination geometry of Mn^{2+} ions in the nucleotide-binding sites of cAPK and Phkγ. Two orthogonal views are presented for each ion. The Mn^{2+} ions and the oxygen atoms of water molecules are displayed as greenblue and red spheres, respectively. See also table 4-4 for a summary of the distances involved.

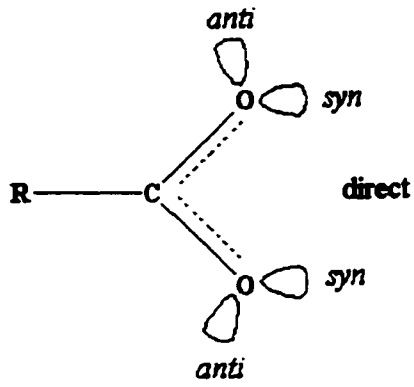


Figure 4-13. Bonding directions involving the *anti* and *syn* lone pairs of the carboxyl oxygens (Glusker, 1991).

However, it is located off the plane of the carboxyl group and its distance from the Oδ1 atom, 2.7 Å, is 0.4 Å longer than from the Oδ2 atom of Asp208. Oδ1 of Asp208 also does not fit into the octahedral coordination geometry of Mg2. This coordination scheme of Mg2 is different from those of Mn2 in the Mn-ATP/PKI(5-24) ternary complex of cAPK and the Mn-AMPPNP complex of Phkγ. Mn2 in these two structures is coordinated to the β- and γ-phosphate oxygens (Oβ2, Oγ3), both of the Asp208 {184} carboxyl oxygens via "direct bonding" and one water molecule(s) (figure 4-12, table 4-3). The coordination geometry centering on these two Mn^{2+} ions is a distorted octahedron.

In the Mg-ADP complex of APH(3')-IIIa, Mg2 is coordinated to the β-phosphate oxygen (Oβ2), both of the carboxyl oxygens of Asp208 {184} via "direct bonding" and three water molecules. The resulting coordination geometry is a distorted octahedron. The displacment of Mg2 upon superimposition of the Mg-AMPPNP and the Mg-ADP complexes is 1.1 Å.

The D-N-D triad

Asp190 {166}, Asn195 {171} and Asp208 {184} form the invariant D-N-D triad. There appears to be a need for stabilizing the spatial geometry of this triad, which is located on two loops formed partly by the conserved HXD {HGD} and DXG {DFG} motifs. In the structures of APH(3')-IIIa, the scaffold of these loops are supported on both ends by two sets of anti-parallel β-strands, $\beta_B2/\beta_B3/\beta_B4$ and β_C1/β_C2 (figure 4-14). As mentioned before, the sequence of the second set of anti-parallel β-strands are highly conserved among APH(3') isozymes. These β-strands, however, are absent in the catalytic cores of protein kinases where the two loops juxtapose through electrostatic interactions between the side-chains of His and Phe of the HRD and the DFG motifs, respectively. The nitrogen on the imidazole ring of the His residue is poised to interact with the aromatic π molecular orbital system of the Phe residue (figure 4-14). The His residue is sometimes replaced by a Tyr residue, whose aromatic ring interacts with that of Phe in an edge-to-surface manner (figure 4-14). Hence, the two loops are structurally secured either by the "button-like" His-Phe (or Tyr-Phe) side-chain interaction in the catalytic cores of protein kinases, or by the "zipper-like" anti-parallel β-strands in APH(3')-IIIa, and likely its isoforms.

Stabilization of these loop structures helps to constrain the side-chain rotamer conformation of the D-N-D triad, which attains a geometry close to an equilateral triangle in the binary complexes of APH(3')-IIIa, the Mn-ATP/PKI(5-24) complex of cAPK and the Mn-AMPPNP complex of Phky (figure 4-14). In these structures, the side-chain of Asn195 {171} is held fixed by hydrogen bonding between its amide nitrogen and the backbone carbonyl oxygen of the invariant Asp190 {166}. On the other hand, both the

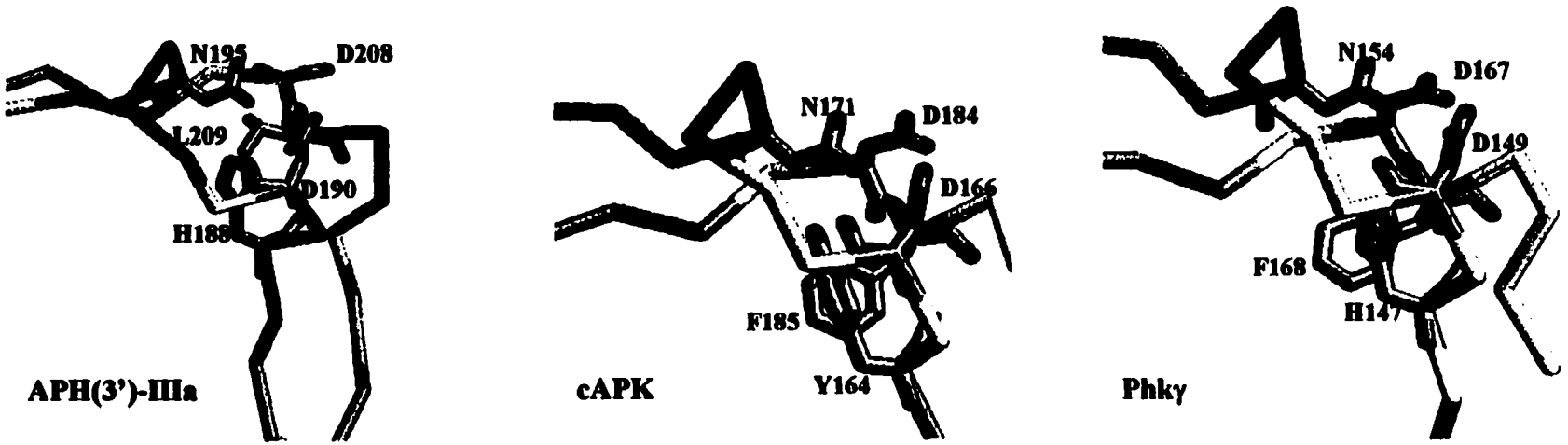


Figure 4-14. The HXD/DXG loop-loop stabilization and the spatial geometry of the conserved D-N-D triad.

side-chains of Asp190 {166} and Asp208 {184} are free-standing, in that they do not form intra-molecular interaction. The side-chain of Asp208 {184} appears to have a greater degree of freedom than that of Asp190 {166} and is held fixed through coordination with Mg1 {Mn1} in the nucleotide complex. This is based on the observation that the carboxyl group of {Asp184} rotates by 97°, while that of Asp{166} retains roughly the same orientation, upon superimposition of the structures of cAPK complexed with and without a nucleotide (result not shown).

Thus far, the significance of the D-N-D structural motif has not been emphasized.

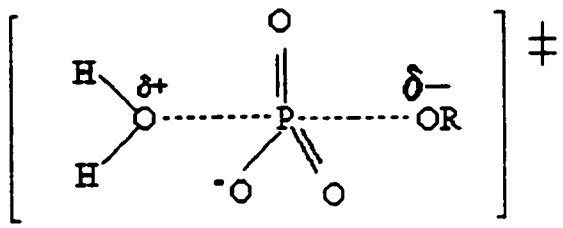
Implications for the mechanism of phosphoryl transfer

The presence of the invariant His188 among all the APH's once prompted the conjecture that these enzymes undergo a phospho-histidine enzyme intermediate during catalysis (Shaw, Rather, Hare & Miller, 1993). However, mutagenesis and positional isotope exchange experiments using γ-^{18}O-ATP did not support such a mechanism (Thompson, Hughes & Wright, 1996). The position of His188 in the structure is also inconsistent with a catalytic role. The side-chain of His188 is buried in a hydrophobic core and participates in the HXD/DXG loop-loop stabilization as discussed previously (figure 4-14).

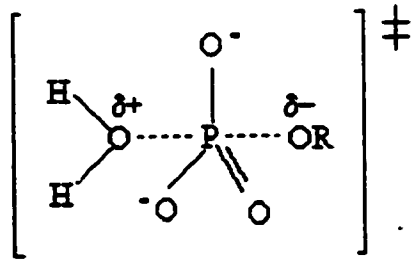
The high degree of structural similarity between the nucleotide-binding sites of APH(3')-IIIa and the catalytic cores of protein kinases indicates that there is likely a common phosphoryl transfer mechanism. This prediction is partly supported by the results of steady-state kinetic analyses on APH(3')-IIIa and cAPK (Cook, Neville, Vrana, Hartl & Roskoski, 1982; Kong & Cook, 1988; McKay & Wright, 1995; McKay & Wright, 1996).

Phosphoryl transfer in both enzymes appear to be a fast step, following binding of both ATP and substrate, and that ADP release is partially rate-limiting. Moreover, stereochemical studies suggest that transfer of γ-phosphate from ATP to the substrate hydroxyl group occurs via direct in-line geometry (Ho, Bramson, Hansen, Knowles & Kaiser, 1988; Thompson, Hughes & Wright, 1996). There are two extreme possibilities of a direct attack mechanism, one involving a meta-phosphate transition state (dissociative pathway) and the other a penta-coordinated phosphate transition state (associative pathway). These pathways are illustrated below.

Dissociative



Associative



In the dissociative pathway, the bond to the leaving group is almost broken before the bond is barely formed with the incoming nucleophile. This results in the formation of an unstable meta-phosphate that carries a full negative charge. In the associative pathway, bond formation with the incoming nucleophile dominates over bond cleavage, resulting in a

penta-coordinated phosphate group that carries two negative charges. In the structures of nucleotide complexes of APH(3')-IIIa, the negative charges developed on both the meta-phosphate and the penta-phosphate transition states can be adequately neutralized by the two Mg^{2+} ions. Hence, these two pathways cannot be differentiated based on the criterion of charge neutralization. However, if the ternary structure of cAPK complexed with ADP and substrate peptide is representative of this family of kinases (Madhusudan, *et al.*, 1994) (see figure 4-16), the plausible nature of the reaction mechanism may be deduced from the relative positioning of the leaving group and the nucleophile. The spacing between the γ-phosphorus and the incoming hydroxyl was 2.7 Å, which is 1.5 Å longer than a P-O covalent bond. Thus, there is enough room for a dissociative pathway to occur. This does not preclude the possibility of an associative pathway in which case a conformational change would be required to close the gap between the two substrates. Such conformational change would likely be transient and has thus far not been observed in crystallographic studies.

The details of the chemical mechanisms involved in phosphoryl transfer, on the other hand, are still unclear. Thus far, kinetic studies on the protein kinases have mainly been concentrated on the plausible role of the Asp190 {166} equivalents as a catalytic base and the effects of divalent metal ions. Structural comparisons performed herein suggest that elucidation of the phosphoryl transfer mechanism requires a clear understanding of the subtle interactions among the two Mg^{2+} ions, the D-N-D triad, the phosphate groups and the incoming substrate hydroxyl group.

The Asp catalytic base model

Asp166 of cAPK was assigned the role of a catalytic base that presumably deprotonates the substrate hydroxyl group, thereby activating nucleophilic attack on the γ-phosphorus of ATP (Knighton, *et al.*, 1991). This hypothesis was based on the observation that Asp166 is in close proximity to what would be the natural phosphate acceptor Ser {Ala21} of PKI(5-24) in the ternary complex structure. The observation that the side-chain of Asp166 is only 2.7 Å away from the hydroxyl oxygen of Ser of a peptide substrate lent further support to this hypothesis (Madhusudan, *et al.*, 1994). Ala mutants of this Asp residue showed dramatic decrease in k_{cat} in both cAPK and APH(3')-IIIa (Gibbs & Zoller, 1991; Hon, *et al.*, 1997). However, current kinetic analyses do not support the presence of a general-base catalyst in the reaction mechanism for cAPK (Zhou & Adams, 1997), nor in the Tyr protein kinase Csk (C-terminal src kinase) (Cole, Grace, Philips, Bum & Walsh, 1995). Moreover, there is a large difference between the pK_a values of the carboxyl group of an Asp and the hydroxyl group of a Ser or a Thr (estimated pK_a's ~4, ~14 and ~15 for the side-chains of Asp, Ser and Thr, respectively (Kyte, 1995)). The gap seems to be too wide to be closed by the protein environment which can influence the apparent pK_a values of amino acid side-chains. This consideration alone indicates that the Asp carboxyl group is unlikely to be able to deprotonate the incoming hydroxyl group.

The divalent metal ion effects

Kinetic studies performed on at least two protein kinases, including cAPK (Adams & Taylor, 1993) and Csk (Grace, Walsh & Cole, 1997), showed that Mg^{2+} is the preferred

divalent cation for phosphoryl transfer and that transition metal divalent cations, such as Mn^{2+} and Co^{2+}, decrease the k_{cat} by 2.5-4 fold. In cAPK, the apparent rate constants of the phosphoryl transfer decreased markedly but that of ADP dissociation only decreased slightly in the presence of Mn^{2+} and Co^{2+} (Adams & Taylor, 1993). For APH(3')-IIIa, the use of Mn^{2+} resulted in almost total inhibition (G. D. Wright, personal communication). A satisfactory explanation for these effects is still lacking.

Coincidentally, the coordination schemes of the two Mg^{2+} ions in the Mg AMPPNP complex of APH(3')-IIIa differ from those of the Mn^{2+} ions in the Mn-ATP/PKI complex of cAPK and Mn-AMPPNP complex of Phkγ. The difference between Mg2 and Mn2 is particularly pronounced. To investigate whether this is a result of the spatial differences in the phosphate-binding subsites, the three structures were superimposed based solely on the least-squares alignment of the main-chain coordinates of the D-N-D triad, which acts as a template for the binding of metal ions. Figure 4-15 illustrates the result of superimposition of the three phosphate-binding subsites containing the adenosine triphosphates.

The r.m.s.d. of the D-N-D main-chain coordinates are 0.19 Å and 0.21 Å between APH(3')-IIIa and cAPK, and between APH(3')-IIIa and Phkγ, respectively. Moreover, the positions of their side-chains are almost identical. This indicates that the spatial relationship of the D-N-D triad is strictly conserved throughout evolution. Despite the large positional deviations of the Lys-Glu pair, the coordinates of the phosphate groups superimpose relatively well and those of Mg1/Mn1 are almost identical (figure 4-15)

On the other hand, due to their different coordination schemes, Mg2 is 1.5 Å and 1.6 Å from Mn2 of cAPK and Phkγ, respectively. When the Mg-ADP complex of

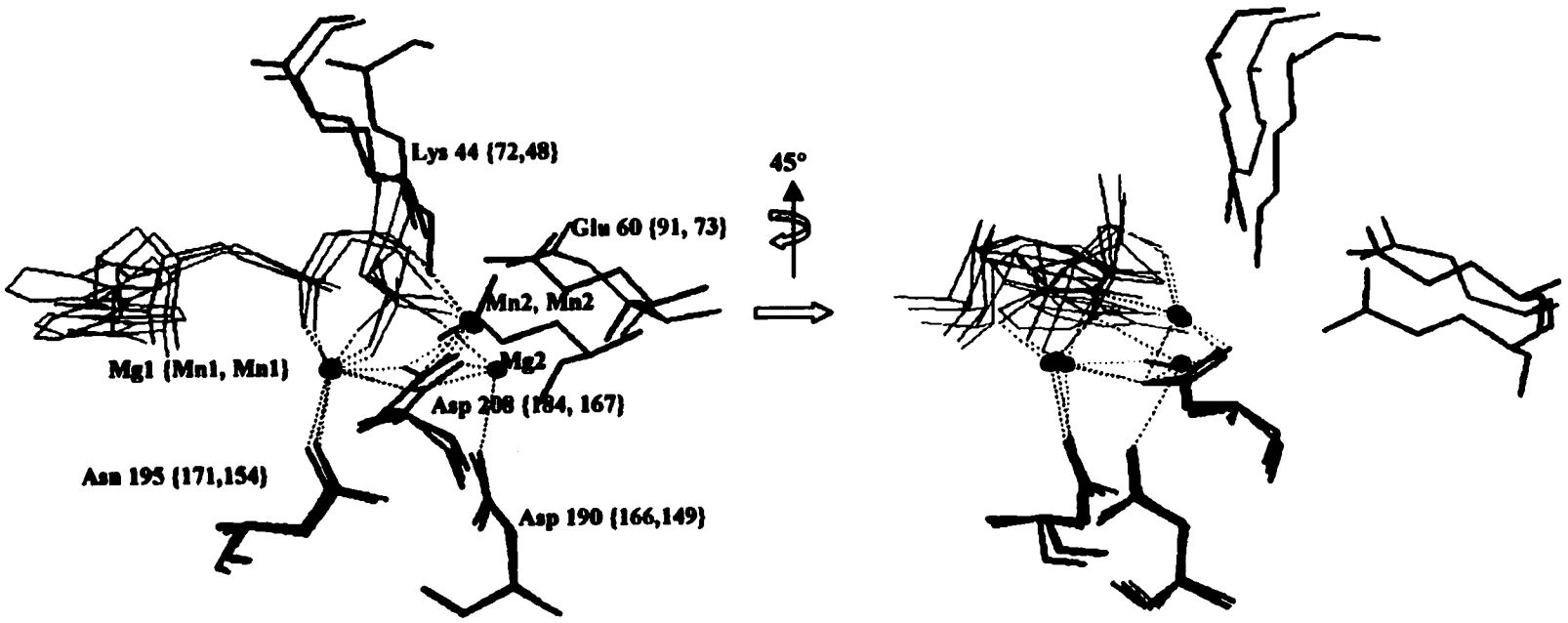


Figure 4-15. Superposition of the conserved phosphate-binding subsites of APH(3')-IIIa (red), cAPK (green) and Phkγ (blue). The alignment was performed using the coordinates of the D-N-D triads of the three proteins. The nucleoside triphosphates are illustrated as thin stick models and the dotted lines indicate the coordination schemes of the metal ions.

APH(3')-IIIa was superimposed in the same manner, the position of Mg2 is close to those of Mn2 of cAPK and Phkγ (differences in distance are 0.3 Å and 0.4 Å, respectively). There is also ample space in the phosphate-binding subsite of APH(3')-IIIa such that the phosphate groups can adjust their torsion angles and Mg2 can adopt a position close to those of Mn2 without encountering steric hindrances. If these crystallographic observations reflect the conditions of kinetic analyses, the significant positional difference between Mg2 and Mn2 is likely due to the difference in their chemical properties, and not due to subtle differences in the structures. A plausible explanation for the differences in coordination schemes could be Mg^{2+}'s preference for unidentate, instead of bidentate coordination. "Direct bonding" of Mg^{2+} ions to carboxyl groups, i.e. bidentate coordination, is rarely observed in small molecules (Glusker, 1991). The observed difference between the position of Mg2 and Mn2 may provide new insights into the roles of Mg^{2+} ions in catalysis.

Plausible roles of the Mg^{2+} ions

The two Mg^{2+} ions appear to serve multiple roles in the phosphoryl transfer reaction. Besides neutralizing the negative charges of the phosphates, both metal ions assist in proper positioning of the γ-phosphate for transfer and confer tight binding of ATP and ADP. Mg1 is proximal to both the β,γ-bridging nitrogen in the AMPPNP complex and the equivalent of the β,γ-bridging oxygen in the ADP complex. This implies that Mg1 plays a role in stabilizing the developing negative charge in the transition state (or intermediate) during phosphate transfer. This role cannot be substituted by positively charged protein side-chains since there is no Lys or Arg residues in the vicinity that are

poised to make contact with the bridging nirogen/oxygen. Mg1 also seems to assist in binding of Mg2 by fixing the rotamer conformation of Asp208 {184}.

The role of Mg2 is more difficult to discern. If the difference in the reaction rates is partly determined by the difference in the coordination scheme between Mg2 and that of Mn2, it might be informative to piece together a picture of Mg2 and Mn2 in the ternary complex. With the currently available coordinates, a hypothetical ground state complex in the active site of cAPK was composed (figure 4-16). The side-chain of the target Ser hydroxyl group for cAPK was modeled based on the coordinates of Ala21 of PKI in the Mn-ATP/PKI complex (see legends of figure 4-16 for details). The Ser hydroxyl oxygen is positioned approximately in-line with the P-O bond between the γ-phosphorus and the β,γ-bridging oxygen. The hydrogen bonding distance between this hydroxyl group and the carboxyl group of Asp{166} is 2.9 Å, which is close to the reported distance of 2.7 Å (Madhusudan, *et al.*, 1994). If Mg2 in the active ternary complex of cAPK takes up a similar position as that of Mg2 in the Mg-AMPPNP complex of APH(3')-IIIa, the same carboxyl oxygen of Asp{166} would simultaneously bond to both the incoming hydroxyl group and Mg2, with the former through the *syn* direction and the latter between the *anti* and the *syn* directions (figure 4-13). The angle between the Mg^{2+}···O bond and the hydrogen bond is 68°, which is much less than the 120° between the two lone pairs of electrons of a carboxyl oxygen. Even when the position of Mg2 is allowed an uncertainty of 0.3 Å (the distance between Mg2 in the Mg-ADP complex of APH(3')-IIIa and Mn2 in the Mn-ATP/PKI complex of cAPK upon superimposition of the two structures), this angle would still be within 80°. This might imply instability in such a bonding scheme. Since the incoming hydroxyl group is anchored indirectly through the extensive

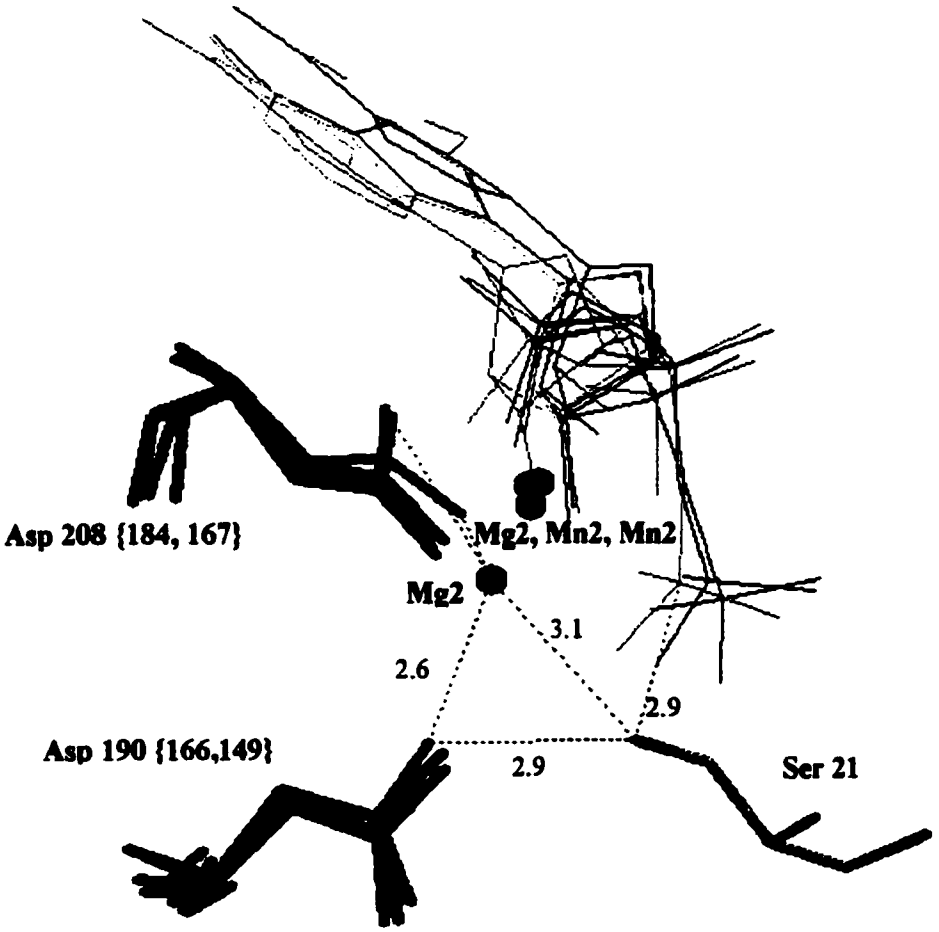


Figure 4-16. A hypothetical active ternary complex. This diagram is similar to figure 4-15, except that the conserved Lys, Glu, Asn and Mg1/Mn1 are omitted for clarity. cAPK, Phky and APH(3')-IIIa complexed with Mg-AMPPNP are coloured in red, blue and green, respectively. The Mg-ADP complex of APH(3')-IIIa (coloured in pink) is superimposed to illustrate the shift in position of Mg2 before and after the phosphoryl transfer reaction. The side-chain of the substrate Ser for cAPK is also superimposed to illustrate the approximate disposition of the ground state ternary complex. The side-chain of Ser21 was created based on the main-chain coordinates of Ala21 in the peptide inhibitor PKI (PDB ID 1ATP). The best rotamer in the library of the program O (Jones, Zou, Cowan & Kjeldgaard, 1991) that can form a hydrogen bond with Asp166 was selected. The reported hydrogen bonding distance in the literature was 2.7 Å (Madhusudan, *et al.*, 1994). Mg1 {Mn1} and Mg2 {Mn2} are shown as small and large spacefilling models, respectively. The numbers indicate the interatomic distances in Å.

interactions between the substrate and the protein molecule, and that the *syn* lone pairs of the carboxyl group is more basic than the *anti* lone pairs (Glusker, 1991), substrate binding may trigger displacement of Mg2. Migration of Mg2 would likely be guided by its bonding to the carboxyl group of Asp208 {184} and facilitated by the water molecules in the solvent pocket where it is situated. It is apparent that Mn2 is unlikely to be able to respond to such a triggering mechanism.

At this point, the generality of the role of Asp190 {166} in stabilizing the formation of a ground state complex through hydrogen bonding with the incoming hydroxyl group awaits more crystallographic studies of ternary complexes of this superfamily of kinases. The generality of the divalent ion effects among these kinases has thus far not been thoroughly evaluated. The location of Mg2 in the Mg-AMPPNP complex of APH(3')-IIIa is based on the currently refined model (see appendix 1) and its accuracy requires confirmation with higher resolution diffraction data. Moreover, it would be desirable to obtain an experimentally-derived model of Mn-AMPPNP complex of APH(3')-IIIa so that meaningful comparison can be performed. Although a lot of ground work needs to be established before the significance of the observed structural difference can be confirmed, it is clear that the conserved coordination environment of Mg2 is unusual. Mg^{2+} ions have a strong preference for perfect octahedral coordination geometry and rarely form "direct bonding" (figure 4-13) with a carboxyl group (Glusker, 1991). Unlike the Mg^{2+} ion that form "direct bonding" with an Asp carboxyl group in the crystal structures of D-xylose isomerase (Farber, Glasfeld, Tiraby, Ringe & Petsko, 1989), Mg2 in the Mg-AMPPNP complex of APH(3')-IIIa is not confined in space by protein or

substrate ligands. It seems that there is a built-in mechanism for its movement during catalysis.

The analysis presented here simply suggests that there could be a genuine structural difference between Mg2 and Mn2 and that this difference may hold the key to the elucidation of the chemical mechanism of these kinases.

The putative antibiotic-binding loop and the C-terminal helix

At the entrance of the nucleotide-binding site in the C-lobe, there are two structural elements that do not have counterparts in the catalytic core of protein kinases but share a high degree of sequence homology with other APH(3') isozymes. These are the extended flexible loop that spans residues 151-166 and the C-terminal helix (residues 250-264, H10) (figures 4-1 & 4-5). Together, they form the better part of a pocket that is likely to be the antibiotic-binding site. Because of its structural disposition and for reasons discussed below, this loop is tentatively called the "antibiotic-binding loop".

The electron density in the region of the "antibiotic-binding loop" is mostly weak and the average B-factor is 52 $Å^2$ for main-chain atoms (figure 4-3b). In the crystal structure, part of this loop is stabilized through disulphide-linkage and van der Waals contact with the dimer partner. In its monomeric form, this loop forms few interactions with the rest of the molecule, and hence, would likely to be even more flexible in its conformation under physiological conditions. It forms a flap that shelters the last half of the C-terminal helix from the bulk solvent and appears as a gateway to a large pocket formed by the C-lobe, underneath the nucleotide-binding site (figure 4-17). This loop is not analogous to the activation segments that span between the consensus motifs DFG and

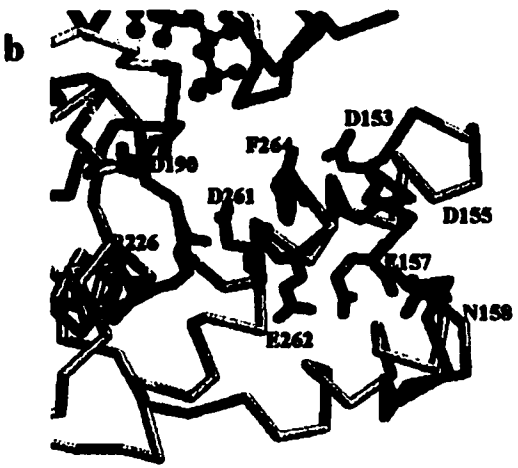


Figure 4-17. The putative antibiotic-binding site of APH(3')-IIIa. (**a**) The side-, front- and top-views of the protein molecule are shown from left to right. The P-loop, domain linker and the "antibiotic-binding loop" are coloured in green, greenblue and cyan, respectively. Side-chains of the conserved residues likely to be involved in antibiotic-binding are displayed as stick models. (**b**) A close-up view. Mg-AMPPNP is shown as ball-and-stick model and the side-chain of the conserved Asp190 is also displayed.

APE in the catalytic core of protein kinases (Johnson, Noble & Owen, 1996). The lengths of the activation segments vary (figure 4-2) and are wedged between two conserved helices {H6 and H7} in the C-lobe (figure 4-5).

The C-terminal helix is positioned beneath the P-loop with the entire surface of one face lining the "antibiotic-binding site". It has the highest degree of local sequence homology among all of the α-helices in APH(3')-IIIa (figure 4-1). The consensus sequence is DXXKXX(F/Y)YXXLDEXF, where X denotes any residue and the underlined residues have more than 70% occurrence in the type I to type VII APH(3') isozymes. The C-terminus, Phe264, is proximal to the γ-phosphate of AMPPNP. This residue not only is highly conserved, but also adopts an unusual side-chain rotamer conformation (figure 4-17). The aromatic ring tilts inward and is buried. Based on the rotamer library in the program O, this side-chain scores a r.m.s.d. of 49.1 Å from the best rotamer (Kleywegt & Jones, 1995). Both of the carboxyl oxygens, one of which forms a hydrogen bond (2.9 Å) with the hydroxyl oxygen of Tyr55, are solvent-accessible (figure 4-17). This arrangement is unlike those observed in most structures of the catalytic cores of protein kinases whose C-termini are on the periphery of the molecules (figure 4-5).

Implications for the binding of aminoglycosides

Although realization of the detailed molecular interactions between protein and antibiotic awaits the determination of the ternary structures, some structural features important for antibiotic-binding are immediately obvious. Considering the chemical properties of the substrates of APH(3')-IIIa (figure 1-1) and the results of kinetic analyses accumulated thus far, several characteristics of the antibiotic-binding site are expected.

First, the antibiotic-binding site should be flexible enough to accommodate the different sizes of aminoglycoside substrates (figure 1-1), and at the same time be able to stabilize their binding. Second, the aminoglycosides have multiple amino groups and electrostatic interactions have been predicted to play an important role in antibiotic recognition (McKay, Roestamadji, Mobashery & Wright, 1996; Cox & Serpersu, 1997). Third, it is likely to have a docking patch composed of conserved residues that can interact specifically with the common structural denominator of the various aminoglycosides, including rings I and II (figure 1-1). Since the structure of ring III is different in the 4,5- and 4,6-disubstituted 2-deoxystreptamines, there would likely be different binding subsites for this ring.

The "antibiotic-binding site" is lined by the solvent accessible surfaces of H8, H10 and the inner side of the "antibiotic-binding loop". The aminoglycoside substrates presumably dock in between the surface provided by H8 and H10, and the "antibiotic-binding loop" which would likely undergo local conformational change to secure binding of the antibiotic. Involvement of local conformational change, instead of global conformational change such as domain reorientation, in substrate-binding has been observed in the structure of triosephosphate isomerase (Knowles, 1991). Due to the size of the "antibiotic-binding loop", the "antibiotic-binding site" would most likely be only partially enclosed. This would allow substrates with bulky substituents to be accommodated. The equivalent of the "antibiotic-binding loop" is not observed in the structures of the catalytic core of protein kinases and instead, by inference from the structures of cAPK, the C-lobes act simply as rigid docking patches for the protein substrates (Narayana, Cox, Xuong, Eyck & Taylor, 1997). For APH(3')-IIIa, the

"antibiotic-binding loop" is likely to be crucial in stabilizing binding of the aminoglycoside substrates which, due to their smaller sizes, likely do not form interactions with the enzyme as extensive as those that occur between protein kinases and their substrates.

The entire molecular surface lining the "antibiotic-binding site" displays negative electrostatic potential (figure 4-18), suggesting that the enzyme is selective for positively charged molecules. The aminoglycoside substrates of APH(3')-IIIa consist of four to six amino groups (figure 1-1). The pK_a values of the amino groups of two of the substrates, amikacin and butirosin A, have been determined by ^{15}N-NMR spectroscopy (Cox & Serpersu, 1997). Other than N3 (amino group attached to C3), which has slightly lower pK_a values (7.6 and 7.4 for amikacin and butirosin A, respectively), the amino groups have pK_a values that vary from 8.1 to 9.7. Thus, these antibiotics are polycationic at physiological pH. The contributions of five amino groups to enzyme binding and catalytic efficiency have been analysed using modified substrates (McKay, Roestamadji, Mobashery & Wright, 1996). Upon removal of individual amino groups, the k_{cat}/K_m ratios varied from 160% for N3 as compared to neamine, a minimal substrate compound, to <1% for N6' as compared to kanamycin A. Both K_m and k_{cat} were affected, indicating that individual amino groups contribute both to binding affinity and proper orientation of the substrate in the formation of productive enzyme-antibiotic complexes. Since the amino groups were only substituted to hydrogen and not to hydroxyl groups, the effect of hydrogen-bonding through ion-pairing with protein side-chains could not be definitively evaluated.

Conceptually, electrostatic attraction may be the first event of substrate recognition in the bulk solvent and stabilization of enzyme-antibiotic complex formation could be mainly through hydrogen-bonding. The molecular surface of the putative

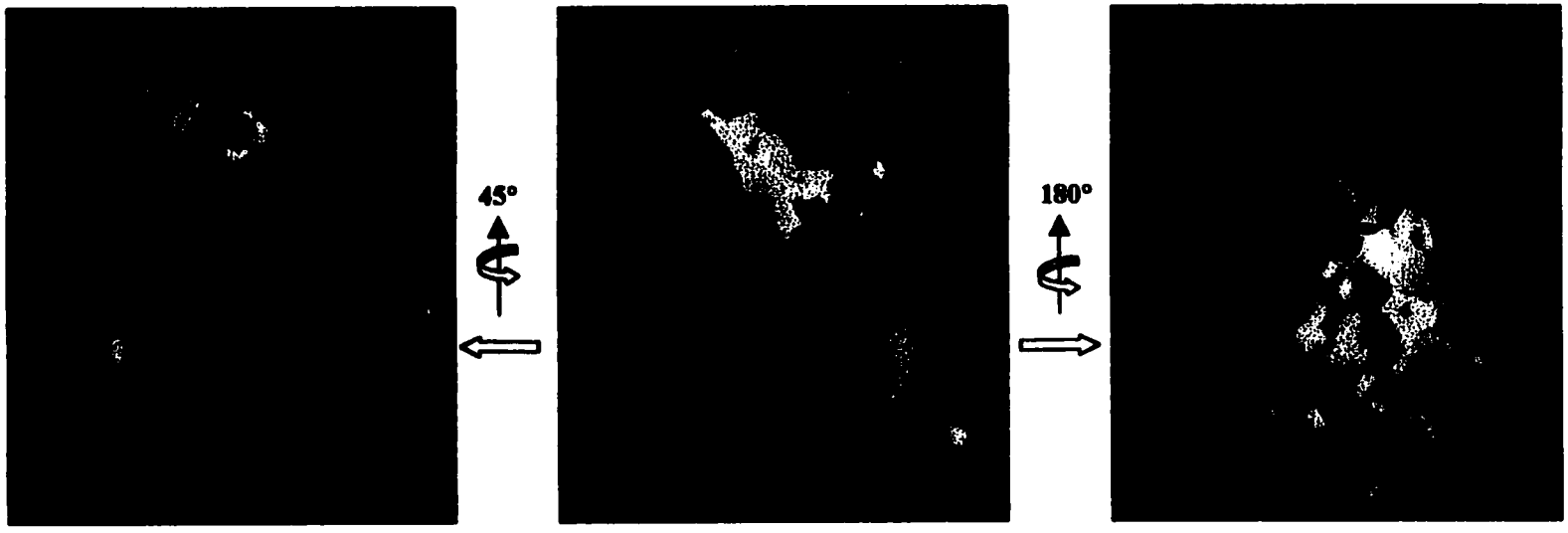


Figure 4-18. Surface electrostatic potential of APH(3')-IIIa. The potentials are coloured in linear gradients of red from –35 *k*T to neutral, and blue from neutral to 35 *k*T. Neutral potential (0 *k*T) is coloured in white. The AMPPNP molecule is shown as stick model coloured in light blue. The front view of the protein is shown in the centre. This figure is prepared with the program GRASP (Nicholls, Sharp & Honig, 1991).

antibiotic-binding site not only has a net electronegative potential, but also has a high density of polar atoms. This surface has the potential to form extensive hydrogen-bonding networks with the highly polar aminoglycoside substrates, which contain of 10 to 14 polar groups per molecule (NH_2 and OH groups inclusive, figure 1-1). Part of the specificity of antibiotic-binding is expected to be conferred by the conserved residues of the enzyme. In the vicinity of the γ-phosphate of AMPPNP is a cluster of highly conserved residues that have solvent-accessible polar groups (figure 4-17). These include Asp153, Asp155 Glu157 and Asn158 on the "antibiotic-binding loop", Arg226 on H8 and Asp261, Glu262 and Phe264 on H10. Part, or all of this cluster of residues may form the primary binding site for rings I and II, which are common to both 4,5- and 4,6-disubstituted 2-deoxystreptamines.

Implications for inhibitor design

The revelation of the striking similarity in the three-dimensional fold, and particularly in the geometry of the ATP-binding sites, between APH(3')-IIIa and the Ser/Thr or Tyr protein kinases has already led to a series of inhibitor studies on this enzyme (Daigle, McKay & Wright, 1997). Representatives of three classes of known protein kinase inhibitors were examined. These included staurosporine, which belongs to the indole carbazole class, genistein and quercetin, belonging to the flavanoid class, and four variants of the isoquinoline sulfonamide class (refer to the aforementioned citation for the chemical structures). Although all three classes of compounds are known to be competitive inhibitors of ATP for the protein kinases (Hidaka, Inagaki, Kawamoto & Sasaki, 1984; Srivastava, 1985; Srivastava & Chiasson, 1986; Akiyama, *et al.*, 1987;

Herbert, Seban & Maffrand, 1990; Akiyama & Ogawara, 1991; Casnellie, 1991), only the isoquinoline sulfonamide class of compounds are good inhibitors of APH(3')-IIIa and the bifunctional AAC(6')-APH(2"), which possesses both acetyltransferase and kinase activities. This latter class of compounds demonstrated competitive inhibition in the micromolar range with respect to ATP. The possibility of forming a hydrogen bond between the isoquinoline ring nitrogen and the backbone amide of Ala93, analogous to that formed between N1 of the adenine of ATP and Ala93, was attributed as a structural basis of inhibition (Daigle, McKay & Wright, 1997). This hydrogen bond seems to be conserved in the nucleotide-complexes of most protein kinases (Zheng, *et al.*, 1993; Owen, Noble, Garman, Papageorgiou & Johnson, 1995; Robinson, *et al.*, 1996; Schulze-Gahmen, De Bondt & Kim, 1996; Sicheri, Moarefi & Kuriyan, 1997). Two of these protein kinases, namely, phosphorylase kinase and hematopoietic cell kinase, are susceptible to quercetin inhibition (Srivastava, 1985). The differentiation of inhibition among the different classes of inhibitors may likely be due to steric hindrance at the backside of the ATP-binding site as most of these inhibitors have bulky and elaborate ring systems. The adenine ring in the Mg-AMPPNP and the Mg-ADP complexes of APH(3')-IIIa is situated in a relatively confined hydrophobic pocket. As will be discussed later, the extensive domain-domain interactions along the backside of the protein molecule would likely resist domain movement to accommodate inhibitors with bulky hydrophobic groups.

These inhibition studies provide an encouraging starting point for the contemplation of inhibitor design. Inhibitors that are specific to the APH isozymes would preferably encompass both the ATP- and the antibiotic-binding sites. The molecular surface lining the former is likely to be more constant than the latter, hence providing a

more general structural base for an inhibitor that would ideally be capable of inhibiting a spectrum of APH isozymes. Interactions with the antibiotic-binding sites would provide specificity towards the APH isozymes.

Interdomain connection

The N- and the C-lobes are connected to each other mainly through two regions. One is covalent in nature, the domain linker, and the other is non-covalent, which herein is referred to as the "domain joint". The term "domain linker" is deemed more appropriate than "domain hinge", which is frequently used for the equivalent region in the structures of the catalytic cores of protein kinases. Two conformations have been observed in the structures of cAPK, whose N-lobe rotates by 15° with respect to the C-lobe from an "open" to a "closed" conformation upon nucleotide-binding (reviewed in (Cox, Radzio-Andzelm & Taylor, 1994)). The "hinge" region was suggested to act as a pivot upon which the two domains could move, to a first approximation, as rigid bodies. However, such nucleotide-induced interdomain movement is not observed in the apoenzyme and ATP complex structures of CDK2 (Schulze-Gahmen, De Bondt & Kim, 1996). Large scale interdomain movement as observed in cAPK seems to be unlikely in APH(3')-IIIa. This is because the two domains interface via contacts formed by H3 from the N-lobe and the loop connecting H5 and H6 from the C-lobe (figure 4-19). A total of 11 residues are involved in forming four hydrogen bonds and 19 van der Waals contacts (table 4-4). Some of the residues involved are invariant or semi-invariant among APH(3') isozymes, including Trp66, Ile125, Cys130, Pro131 and Tyr132. The interactions extend over the entire backside (opposite to the entrance of the nucleotide-binding site) of the protein molecule.

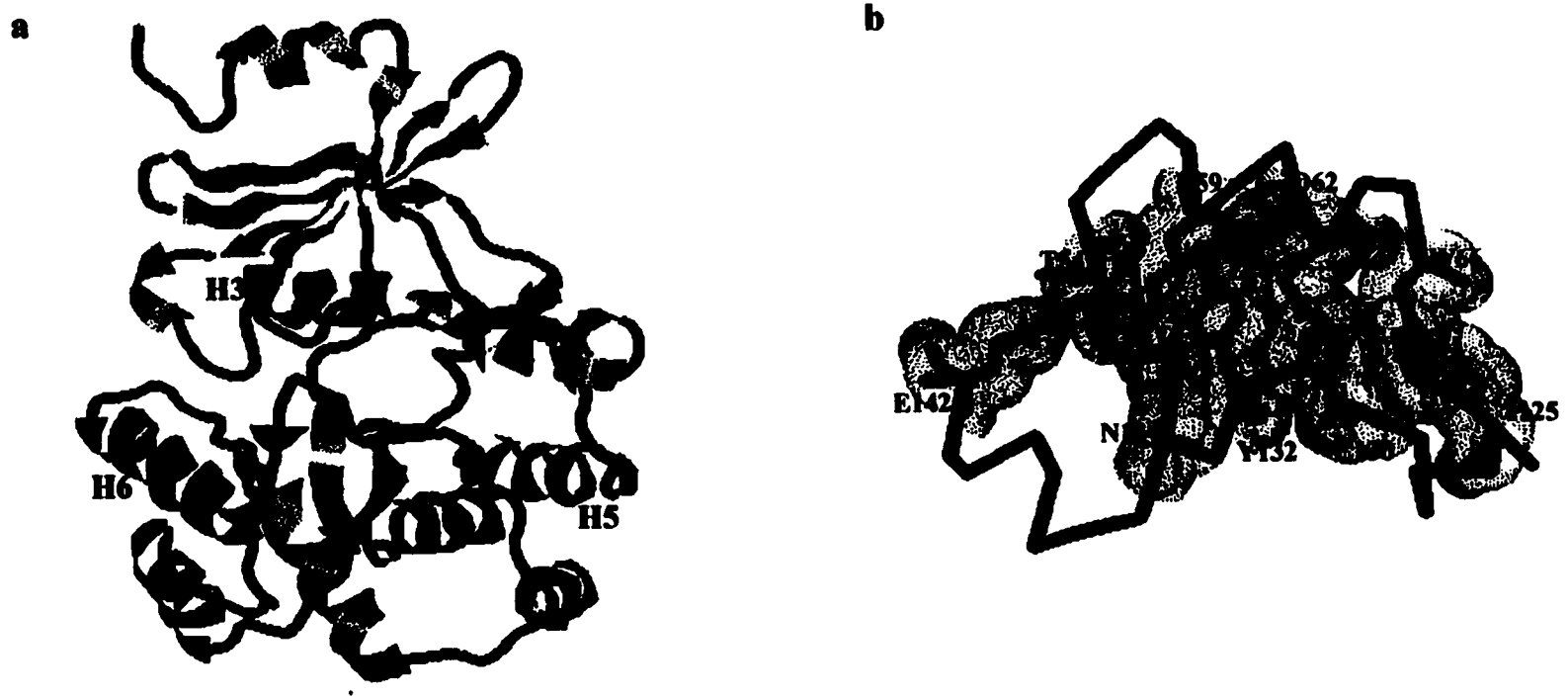


Figure 4-19. Domain joint on the back side of APH(3')-IIIa. (**a**) Global view. The P-loop, domain linker and "antibiotic-binding loop" are coloured in green, greenblue and cyan. The upper and lower parts of the domain joint are coloured in orange and redorange, respectively. (**b**) Close-up view. Side-chains of residues involved in domain contact are shown as stick models. The van der Waals surface of these residues is displayed in Connolly dot representation. Oxygen, nitrogen and sulphur atoms are coloured in red, blue and yellow, respectively.

Table 4-4. Interactions between the two lobes of APH(3')-IIIa along the domain joint.

Hydrogen bonding				
N-lobe		C-lobe		Distance (Å)
Thr54	OG1	Glu142	OE2	2.9
Thr54	N	Glu142	OE1	3.0
Arg59	NH2	Asn134	OD1	2.6
Asp62	OD2	Tyr132	OH	2.8
van der Waals interactions				
N-lobe		C-lobe		Distance (Å)
Arg59	CZ	Tyr132	CD1	3.6
Arg59	CZ	Tyr132	CG	3.9
Arg59	CG	Tyr132	CZ	3.6
Arg59	CG	Tyr132	CE2	3.5
Arg59	CG	Tyr132	CD2	3.8
Asp62	CB	Tyr132	CE2	3.8
Met63	CE	Cys130	SG	3.8
Met63	SD	Tyr132	CD2	3.8
Met63	SD	Pro131	CD	3.8
Met63	SD	Tyr132	CD2	3.8
Met63	SD	Pro131	CD	3.8
Trp66	CH2	Cys130	SG	3.9
Trp66	CH2	Ile125	CG2	3.8
Trp66	CZ3	Cys130	SG	3.8
Trp66	CZ2	Cys130	CA	3.8
Trp66	CE2	Cys130	CA	3.9
Trp66	CD2	Pro131	CD	3.7
Trp66	CG	Pro131	CG	3.9
Trp66	CG	Pro131	CD	3.7

Analysis of these interactions was performed by the program LIGPLOT (Wallace, Laskowski & Thornton, 1995).

A rigid-body type of inter-domain movement would likely be energetically costly.

Substrate-binding induced domain closure is frequently observed in the structures of kinases and well-known examples are represented by the yeast hexokinase (Steitz, Shoham & Bennett, 1981) and the *E. coli* adenylate kinase (Schulz, Müller & Diederichs, 1990). These enzymes carry out uni-substrate phosphorylation and specificity is ensured by bringing the two substrate-binding sites into an enclosed environment. This property may not be favourable to APH(3')-IIIa which seems to be optimized for inactivating a broad spectrum of aminoglycosides.

References

Adams, J. A., & Taylor, S. S. (1993). *Protein Sci.* **2**, 2177-2186.

Akiyama, T., Ishida, J., Nakagawa, S., Ogawara, H., Watanabe, S., Itoh, N., Shibuya, M., & Fukami, Y. (1987). *J Biol Chem* **262**, 5592-5.

Akiyama, T., & Ogawara, H. (1991). *Methods Enzymol* **201**, 362-70.

Bairoch, A., & Claverie, J. M. (1988). *Nature* **331**, 22.

Bossemeyer, D. (1995). *FEBS Lett* **369**, 57-61.

Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., & Huber, R. (1993). *Embo J* **12**, 849-59.

Brunger, A. T. (1993). *X-PLOR, Verson 3.1: A System for X-ray Crystallography and NMR*. New Haven, CT, USA: Yale University Press.

Casnellie, J. E. (1991). *Adv Pharmacol* **22**, 167-205.

Cole, P. A., Grace, M. R., Philips, R. S., Burn, P., & Walsh, C. T. (1995). *J. Biol. Chem.* **270**, 22105-22108.

Cook, P. F., Neville, M. E., Jr., Vrana, K. E., Hartl, F. T., & Roskoski, R., Jr. (1982). *Biochemistry* **21**, 5794-5799.

Corpet, F. (1988). *Nucl. Acids Res.* **16**, 10881-10890.

Cox, J. R., & Serpersu, E. H. (1997). *Biochemistry* **36**, 2353-9.

Cox, S., Radzio-Andzelm, E., & Taylor, S. S. (1994). *Curr Opin Struct Biol* **4**, 893-901.

Daigle, D. M., McKay, G. A., & Wright, G. D. (1997). *J Biol Chem* **272**, 24755-8.

De Bondt, H. L., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O., & Kim, S. H. (1993). *Nature* **363**, 595-602.

Farber, G. K., Glasfeld, A., Tiraby, G., Ringe, D., & Petsko, G. A. (1989). *Biochemistry*

28, 7289-97.Gibbs, C. S., & Zoller, M. J. (1991). *J. Biol. Chem.* **266**, 8923-8931.

Glusker, J. P. (1991). *Advances in Protein Chemistry* **42**, 1-76.

Goldberg, J., Nairn, A. C., & Kuriyan, J. (1996). *Cell* **84**, 875-87.

Grace, M. R., Walsh, C. T., & Cole, P. A. (1997). *Biochemistry* **36**, 1874-1881.

Hanks, S. K., & Hunter, T. (1995). *Faseb J* **9**, 576-96.

Herbert, J. M., Seban, E., & Maffrand, J. P. (1990). *Biochem Biophys Res Commun* **171**, 189-95.

Hidaka, H., Inagaki, M., Kawamoto, S., & Sasaki, Y. (1984). *Biochemistry* **23**, 5036-41.

Ho, M.-F., Bramson, H. N., Hansen, D. E., Knowles, J. R., & Kaiser, E. T. (1988). *J. Am. Chem. Soc.* **110**, 2680-2681.

Hon, W. C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D., & Berghuis, A. M. (1997). *Cell* **89**, 887-95.

Hu, S. H., Parker, M. W., Lei, J. Y., Wilce, M. C., Benian, G. M., & Kemp, B. E. (1994). *Nature* **369**, 581-4.

Hubbard, S. R., Wei, L., Ellis, L., & Hendrickson, W. A. (1994). *Nature* **372**, 746-54.

Jeffrey, G. A., & Saenger, W. (1991). *Hydrogen bonding in biological structures.* Berlin: Springer-Verlag.

Jiang, J.-S., & Brunger, A. T. (1994). *J. Mol. Biol.* **243**, 100-115.

Johnson, L. N., Noble, M. E. M., & Owen, D. J. (1996). *Cell* **85**, 149-156.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110-119.

Kabsch, W., & Sander, C. (1983). *Biopolymers* **22**, 2577-637.

Kleywegt, G. J., & Jones, T. A. (1995). *Acta Cryst.* **D52**, 829-832.

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., & Sowadski, J. M. (1991). *Science* **253**, 407-14.

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Xuong, N. H., Taylor, S. S., & Sowadski, J. M. (1991). *Science* **253**, 414-20.

Knowles, J. R. (1991). *Nature* **350**, 121-4.

Kong, C.-T., & Cook, P. F. (1988). *Biochemistry* **27**, 4795-4799.

Kyte, J. (1995). *Structure in Protein Chemistry.* New York & London: Garland Publishing, Inc. *pp.59.*

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283-291.

Longenecker, K. L., Roach, P. J., & Hurley, T. D. (1996). *J Mol Biol* **257**, 618-31.

Luzzáti, V. (1952). *Acta Cryst.* **5**, 802-810.

Madhusudan, Trafny, E. A., Xuong, N. H., Adams, J. A., Ten Eyck, L. F., Taylor, S. S., & Sowadski, J. M. (1994). *Protein Sci* **3**, 176-87.

McKay, G. A., Roestamadji, J., Mobashery, S., & Wright, G. D. (1996). *Antimicrob Agents Chemother* **40**, 2648-50.

McKay, G. A., Thompson, P. R., & Wright, G. D. (1994). *Biochemistry* **33**, 6936-44.

McKay, G. A., & Wright, G. D. (1995). *J Biol Chem* **270**, 24686-92.

McKay, G. A., & Wright, G. D. (1996). *Biochemistry* **35**, 8680-5.

Mohammadi, M., McMahon, G., Sun, L., Tang, C., Hirth, P., Yeh, B. K., Hubbard, S. R., & Schlessinger, J. (1997). *Science* **276**, 955-60.

Mohammadi, M., Schlessinger, J., & Hubbard, S. R. (1996). *Cell* **86**, 577-87.

Narayana, N., Cox, S., Xuong, N., Eyck, L. F. T., & Taylor, S. S. (1997). *Structure* **5**,

921-35.

Nicholls, A., Sharp, K. A., & Honig, B. (1991). *Proteins* **11**, 281-96.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., & Thornton, J. M. (1997). *Structure* **5**, 1093-108.

Otwinowski, Z., & Minor, W. (1997). *Methods Enzymol.* **276**, 307-326.

Owen, D. J., Noble, M. E., Garman, E. F., Papageorgiou, A. C., & Johnson, L. N. (1995). *Structure* **3**, 467-82.

Read, R. (1994). In *Proceedings of the CCP4 Study Weekend*, edited by Bailey, S., Hubbard, R., & Waller, D. pp. 31-40. Warrington, U.K.: Daresbury Laboratory.

Robinson, M. J., Harkins, P. C., Zhang, J., Baer, R., Haycock, J. W., Cobb, M. H., & Goldsmith, E. J. (1996). *Biochemistry* **35**, 5641-6.

Saenger, W. (1984). *Principles of nucleic acid structure.* New York, Berlin, Heidelberg, Tokyo: Springer-Verlag.

Sayle, R. A., & Millner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374-376.

Schulz, G. E., Müller, C. W., & Diederichs, K. (1990). *J Mol Biol* **213**, 627-30.

Schulze-Gahmen, U., De Bondt, H. L., & Kim, S. H. (1996). *J Med Chem* **39**, 4540-6.

Shaw, K. J., Rather, P. N., Hare, R. S., & Miller, G. H. (1993). *Microbiol. Rev.* **57**, 138-163.

Sicheri, F., Moarefi, I., & Kuriyan, J. (1997). *Nature* **385**, 602-9.

Smith, C. A., & Rayment, I. (1996). *Biophys J* **70**, 1590-602.

Srivastava, A. K. (1985). *Biochem Biophys Res Commun* **131**, 1-5.

Srivastava, A. K., & Chiasson, J. L. (1986). *Prog Clin Biol Res* **213**, 315-8.

Steitz, T. A., Shoham, M., & Bennett, W. S., Jr. (1981). *Philos Trans R Soc Lond B Biol*

Sci **293**, 43-52.

Thompson, P. R., Hughes, D. W., & Wright, G. D. (1996). *Chem Biol* **3**, 747-55.

Thompson, P. R., Hughes, D. W., & Wright, G. D. (1996). *Biochemistry* **35**, 8686-95.

Trieu-Cuot, P., & Courvalin, P. (1983). *Gene* **23**, 331-41.

Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982). *Embo J* **1**, 945-51.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). *Protein Eng* **8**, 127-34.

Xu, W., Harrison, S. C., & Eck, M. J. (1997). *Nature* **385**, 595-602.

Zhang, F., Strand, A., Robbins, D., Cobb, M. H., & Goldsmith, E. J. (1994). *Nature* **367**, 704-11.

Zheng, J., Knighton, D. R., ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., & Sowadski, J. M. (1993). *Biochemistry* **32**, 2154-61.

Zheng, J., Trafny, E. A., Knighton, D. R., Xuong, N.-H., Taylor, S. S., Ten Eyck, L. F., & Sowadski, J. M. (1993). *Acta Crystallogr. D* **49**, 362-5.

Zhou, J., & Adams, J. A. (1997). *Biochemistry* **36**, 2977-2984.

Chapter 5

Summary and General Discussion

Summary of experimental accomplishments

The structural studies of APH(3')-IIIa presented in this thesis have implications for both the crystallographic techniques used to determine novel protein structures and for understanding the chemical mechanisms of a unique group of kinases.

The unanticipated success of phasing the protein structure using multi-wavelength anomalous diffraction (MAD) data collected on a crystal containing a relatively large substructure of Se atoms provided an opportunity to re-evaluate the applicability of anomalous x-ray diffraction. In particular, single-wavelength anomalous diffraction data was demonstrated to be sufficient to produce a phase set of comparable quality to that generated with MAD data. The technique of single crystal, single-wavelength anomalous diffraction has rarely been practiced for the determination of macromolecular structures. The results presented in this study attest to the general applicability of this technique with Se-Met derivatives which are fast becoming an important tool in macromolecular crystallography. The computational techniques are well established and this precedent may help to overcome the conceptual hurdle of exploiting the relatively weak but useful signals from anomalous scattering.

The revelation of the structural similarity between APH(3')-IIIa and the catalytic cores of protein kinases aided in distilling the structural features important for a phosphoryl transfer mechanism that is likely to be common among kinases that adopt a similar nucleotide-binding site. A structurally conserved protein scaffold is employed to secure the relative spatial geometry of five invariant residues and two Mg^{2+} ions for

binding of the phosphate groups of ATP and ADP. The arrangement of these structural elements serve to correctly position the γ-phosphate for nucleophilic attack. In particular, the most evolutionary conserved structural motif – the D-N-D triad in the C-terminal domain, appears to provide an intrinsic mechanism that selects Mg^{2+} ions for optimal catalysis. The position of one of the two Mg^{2+} ions is proximal to the invariant Asp residue that is responsible for positioning the incoming hydroxyl group. This observation is novel since similar complex structures of two protein kinases contain Mn^{2+} ions which take up a different position. Elucidation of the chemical implications underlying these positional differences may provide an explanation for the difference in reaction rates and reveal the phosphoryl transfer mechanism of this superfamily of kinases.

Evolution of a "promiscuous kinase" structural template

It was Brenner who first called attention to the sequence similarity between the aminoglycoside phosphotransferases and protein kinases (Brenner, 1987). The template of sequence comparisons spanned the C-lobe half of the nucleotide-binding site. Predating the structure determination of any of the proteins, the two conserved Asp residues (190 and 208) were predicted to participate in binding the phosphate groups of ATP via Mg^{2+} and in their activation. A plausible clue to the evolutionary relationship between the aminoglycoside phosphotransferases and the eukaryotic type of protein kinases may come from a common genus of organisms - *Streptomyces*. These organisms not only produce various types of antibiotics, including aminoglycosides, but they are also one of a few kinds of bacteria in which the Ser/Thr or Tyr type of protein kinases have been identified (reviewed in (Zhang, 1996)). A number of genes have been identified in *Streptomyces*

coelicolor, including *afsK* (Matsumoto, Hong, Ishizuka, Horinouchi & Beppu, 1994), *pkaA* and *pkaB* (Urabe & Ogawara, 1995), which encode proteins that reveal high degrees of sequence similarity to the eukaryotic-type of protein kinases. The sequence homology encompasses both the N- and the C-lobes of the catalytic cores (figure 4-2). When over-expressed in *E. coli*, these proteins were found to be (trans-)autophosphorylated, singly or doubly, on Ser/Thr/Tyr residues (Matsumoto, Hong, Ishizuka, Horinouchi & Beppu, 1994; Urabe & Ogawara, 1995). Moreover, their occurrence seems to be widespread among other *Streptomyces* species on the basis of Southern hybridization experiments (Matsumoto, Hong, Ishizuka, Horinouchi & Beppu, 1994; Urabe & Ogawara, 1995). *Streptomyces* are Gram-positive, filamentous soil bacteria. Thus far, other prokaryotic genes showing homology to eukaryotic-type protein kinases have only been identified in species that undergo multi-cellular stages in their life cycles. These include the Gram-negative soil bacterium *Myxococcus xanthus* (Udo, Munoz-Dorado, Inouye & Inouye, 1995) and the multi-cellular filamentous cyanobacterium *Anabaena sp.*(Zhang, 1993). Hence, these protein kinases were suggested to play roles in intercellular interactions via signaling pathways that are distinct from the better-characterized bacterial signal-transduction pathways involving the His kinase superfamily (Alex & Simon, 1994; Zhang, 1996).

Aminoglycoside phosphotransferases are homologous to viomycin phosphotransferase (Brenner, 1987) (figure 5-1), an enzyme found in the viomycin-producing *Streptomyces vinaceus* (Skinner & Cundliffe, 1980; Thompson, *et al.*, 1982). Viomycin is a basic, 16-membered ring cyclic peptide and belongs to the tuberactinomycin type of antibiotics. Like aminoglycosides, viomycin also forms complexes with RNA and

```
                1                                                  50
   aph(3')I     MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATIYRLYGKP-DA
   aph(3')V                  MDDSTLRRKYPHHEWHAVNEGDSGAFVYQLTGGPEPQ
aph(3')IIIa               MAKMRISPELKKLIEKYRCVKDTEGMSPAKVYKLVGE---N
        vph     MRIIETHRDLLSRLLPGDTVGGLAVHEGQFHHVVIGSHRV----------
  Consensus                    p                  g s a Vy

               51                                                 100
   aph(3')I     PELFLK---HGKGSVANDVTDEMVRLNWLTEF-MPLPTIKHFIRTPDDAW
   aph(3')V     PELYAKIAPRAPENSAFDLSGEADRLEWLHRHGIPVPRVVERGADDTAAW
aph(3')IIIa     ENLYLKMTDSRYKGTTYDVEREKDMMLWLEGK-LPVPKVLHFERHDGWSN
        vph     -VCFARTRAAADRLPGRADVLRALAGIDLGFRTPQPLSEGGAQGTDEPPY
  Consensus        %lk                ea  l wL     pvp

               101                                                150
   aph(3')I     LLTTAIPGKTAFQVLEEYPDSGENIVDALAVFLRRLHSIPVCNCPFNSDR
   aph(3')V     LVTEAVPGVAAAEEWPEH--QRFAVVEAMAELARALHELPVEDCPSDRRL
aph(3')IIIa     LLMSEADGVL-CSEEYEDEQSPEKIIELYAECIRLFHSIDISDCPYTNSL
        vph     LVLSRIPGAPLEDDVLTSPEVAEAVARQYATLLSGLAAAGDEEKVRAALP
  Consensus     $   a  g                      ya  lr lh      cp

               151                                                200
   aph(3')I     VFRLAQAQSRMNNGLVDASDFDDERNGW-----PVEQVWKEMHKLLPFSP
   aph(3')V     DAAVAEARRNVAEGLVDLDDLQEERAGW-----TGDQLLAELDRTRPEKE
aph(3')IIIa     DSRLAELDYLLNNDLAD-VDCENWEEDT--PFKDPRELYDFLKTEKPE-E
        vph     EAPANEWQEFATGVRTELFPLMSDGGR------ERAERELAALDALPHLT
  Consensus       rl e        l #  d                el   l    p  e

               201                                                250
   aph(3')I     DSVVTHGDFSLDNLIFD--EG--KLIGCIDVGRVGIADRYQDLAILWNCL
   aph(3')V     DLVVCHGDLCPNNVLLD--PGTCRVTGVIDVGRLGVADRHADIALAAREL
aph(3')IIIa     ELVFSHGDLGDSNIFVK--DG--KVSGFIDLGRSGRADKWYDIAFCVRSI
        vph     SAVV-HGDLGGENVLWETVDGVPRMSGVVDWDEVGIGDPAEDLAAIGASY
  Consensus      lv  HGD    N       g    sgfiD  grG aD   Dia   r l

               251                                                300
   aph(3')I     -GE----FSPSLQKRLFQKYGIDNPDMNKLQFHLMLDEFF
   aph(3')V     EIDEDPWFGPAYAERFLERYGAHRVDKEKLAFYQLLDEFF
aph(3')IIIa     RED----IGEEQYVELFFDLLGIKPDWEKIKYYILLDELF
        vph     GEE---LLGRVLALGGWADNGTAERISAIRGTFALQQALYAQRDGDEEEL
  Consensus      e      g   a  f    g   pd  k   y 11#e f

               301
   aph(3')I
   aph(3')V
aph(3')IIIa
        vph     ADGLSGYR
  Consensus
```

Figure 5-1. Sequence alignment of representative APH(3') isozymes and viomycin phosphotransferase from *Streptomyces vinaceus* (vph, GENBANK accession number 139341). Convention of the symbols is the same as in figures 4-1 and 4-2.

is a potent inhibitor of bacterial protein synthesis (Skinner & Cundliffe, 1980). On the basis of amino acid sequence alignment (figure 5-1), the 3D-structure of viomycin phosphotransferase would likely have a topology closer to that of APH(3')-IIIa than to the eukaryotic-type protein kinases. Sequence homology between viomycin phosphotransferase and APH(3') isozymes resides in the last third of the molecules, starting a little upstream of the HGD motif of the nucleotide-binding site. The D-N-D triad and the equivalence of the central two-stranded β-sheet, β_C1 and β_C2, are also conserved. Sequence homology among the different types of APH enzymes is also largely confined to the last third of the sequences (Shaw, Rather, Hare & Miller, 1993). This suggests that viomycin phosphotransferase and the APH's are close cousins during evolution. This finding indicates that *Streptomyces* have evolved similar mechanisms to inactivate two different types of antibiotics that they produce using the same structural template, which plausibly shares its ancestry with that of the protein kinases. It is interesting that both aminoglycoside and viomycin kinases, and at least some of the protein kinases select for basic substrates.

Thoughts on antibiotic resistance

Antibiotic-producing strains of bacteria have long been suspected as the origin of bacterial antibiotic-resistance as they display resistance mechanisms similar to those found in clinical isolates (Walker & Walker, 1970; Benveniste & Davies, 1973). This hypothesis was supported by the discovery that resistance genes originally identified in the pathogenic bacteria were also found in the antibiotic-producing strains of bacteria (Skinner & Cundliffe, 1980; Thompson, *et al.*, 1982; Herbert, Sarwar, Ner, Giles & Akhtar, 1986;

Trieu-Cuot & Courvalin, 1986). The products of these genes, which presumably evolved for biosynthesis of secondary metabolites and for self-defense, have been widely disseminated to other microorganisms whose growth the antibiotics are meant to inhibit. In fact, most of the antibiotic-resistance genes are located on conjugative plasmids and/or transposons (Trieu-Cuot & Courvalin, 1986; Trieu-Cuot, Arthur & Courvalin, 1987). Horizontal transfer of genetic materials among bacteria through conjugation is a common phenomenon and has been found to occur even between distantly related bacterial genera (Frost, 1992; Clewell, 1993; Davies, 1994). This mode of gene shuttling is the most likely cause of multi-antibiotic resistance in bacteria. Plasmids that amass a collection of antibiotic resistant genes are not only widespread in clinical isolates (Davies, 1994; Leclercq, 1997), but have also been found in food borne bacteria (Perreten, *et al.*, 1997). The ability to accumulate multiple genes that confer resistance to different types of antibiotics is equally, if not more, threatening to the emergence of resistance to newly distributed antibiotics. The mechanisms of bacterial conjugation, however, are still poorly understood.

The widespread use of antibiotics in the treatment of bacterial infections has occurred for roughly half a century, which is a short period as compared to the history of chemical warfare waged within the microbial world. It seems that we are only beginning to understand our role as a third party in this tug-of-war.

References

Alex, L. A., & Simon, M. I. (1994). *Trends in Genet.* **10**, 133-138.

Benveniste, R., & Davies, R. (1973). *Proc. Natl. Acad. Sci. USA* **72**, 3628-3632.

Brenner, S. (1987). *Nature* **329**, 21.

Clewell, D. B. 1993. Bacterial conjugation. Plenum Press, New York.

Davies, J. (1994). *Science* **264**, 375-82.

Frost, L. S. (1992). *Can. J. Microbiol.* **38**, 1091-1096.

Herbert, C. J., Sarwar, M., Ner, S. S., Giles, I. G., & Akhtar, M. (1986). *Biochem J* **233**, 383-93.

Leclercq, R. (1997). *Clin Infect Dis* **24**, S80-4.

Matsumoto, A., Hong, S. K., Ishizuka, H., Horinouchi, S., & Beppu, T. (1994). *Gene* **146**, 47-56.

Perreten, V., Schwarz, W., Cresta, L., Boeglin, M., Dasen, G., & Teuber, M. (1997). *Nature* **389**, 801-2.

Shaw, K. J., Rather, P. N., Hare, R. S., & Miller, G. H. (1993). *Microbiol. Rev.* **57**, 138-163.

Skinner, R. H., & Cundliffe, E. (1980). *J Gen Microbiol* **120**, 95-104.

Thompson, C. J., Skinner, R. H., Thompson, J., Ward, J. M., Hopwood, D. A., & Cundliffe, E. (1982). *J Bacteriol* **151**, 678-85.

Trieu-Cuot, P., & Courvalin, P. (1986). *J Antimicrob Chemother* **18 Suppl C**, 93-102.

Trieu-Cuot, P., Arthur, M., & Courvalin, P. (1987). *Microbiol. Sci.* **4**, 263-266.

Udo, H., Munoz-Dorado, J., Inouye, M., & Inouye, S. (1995). *Genes Dev.* **9**, 972-983.

Urabe, H., & Ogawara, H. (1995). *Gene* **153**, 99-104.

Walker, M. S., & Walker, J. B. (1970). *J. Biol. Chem.* **245**, 6683-6689.

Zhang, C.-C. (1993). *Proc. Natl. Acad. Sci. USA* **90**, 11840-11844.

Zhang, C. C. (1996). *Mol Microbiol* **20**, 9-15.

Appendix 1

Note on model refinement

The plausible enzyme mechanism of APH(3')-IIIa implied in chapter 4 is based on the models built and refined independently by the author. The refinement statistics are different from those of the published model of the Mg-ADP complex (Hon, W. C., McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S., Wright, G. D., & Berghuis, A. M. (1997). *Cell* **89**, 887-95) and those of the Mg-AMPPNP complex which is currently being refined by Dr. Albert M. Berghuis (personal communication). At present, the two models of the Mg-AMPPNP complex differ mainly in the domain linker and the putative antibiotic-binding loop regions (results not shown). Pertinent statistics of the models refined by A.M.B. are presented in the following table (see also table 4-1).

Statistics of model refinement carried out by author and A.M. B.

	Mg^{2+}-ADP complex	Mg^{2+}-AMPPNP complex
Model refinement ($F/\sigma_F \geq 0$)		
Reflections used in R-factor calculations		
Number	29129 (28047)	23215 (22401)
Resolution range (Å)	40.0-2.2	40.0-2.4
R_{conv} [a] (%)	22.0 (22.3)	21.0 (22.3)
R_{free} [b] (%)	29.2 (30.0)	27.7 (30.1)
Number of non-hydrogen atoms		
protein	4340	4340
ADP / AMPPNP	54	62
Mg^{2+}	4	2 (4)
water	128 (289)	65 (169)
isopropanol	0 (8)	0 (4)
Residues in Ramachandran plot [c] (%)		
most favourable	89.0 (92.2)	89.2 (88.8)
additionally allowed	10.3 (10.3)	10.3 (10.8)
generously allowed	0.6 (0.2)	0.4 (0.4)
disallowed	0.0 (0.0)	0.0 (0.0)
Luzzati estimated coordinate error (Å)	0.4 / 0.3	0.5 / 0.4
test set / work set	(0.4 / 0.3)	(0.5 / 0.3)

[a] $R_{conv} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively. [b] R_{free} is equivalent to R_{conv} for 5% (author's) or 10 % (A.M.B.'s) of randomly selected F_o that were not included in the refinement (Brunger, 1992). [c] As calculated by the program PROCHECK (Laskowski, MacArthur, Moss & Thornton, 1993). The percentage represents non-glycine and non-proline residues. Statistics generated from the models described in this thesis are shown in brackets.

Appendix 2

List of refereed publication

1. Griffith, M., Ala, P., Yang, D. S.-C., **Hon, W.-C.**, and Moffat, B. A. (1992). Antifreeze protein produced endogenously in winter rye leaves. *Plant Physiol.* 100:593-596.

2. **Hon, W.-C.**, Griffith, M., Chong, P., and Yang, D. S.-C. (1994). Extraction and isolation of antifreeze proteins from winter rye (*Secale cereale* L.) leaves. *Plant Physiol.* 104:971-980.

3. **Hon, W.-C.**, Griffith, M., Mlynarz, A., Kwok, Y. C., and Yang, D. S.-C. (1995). Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiol.* 109:879-889.

4. Antikainen, M., Griffith, M., Zhang, J., **Hon, W.-C.**, Yang, D. S.-C., and Pihakaski-Maunsbach, K. (1996). Immunolocalization of antifreeze proteins in winter rye leaves, crowns, and roots by tissue printing. *Plant Physiol.* 110:845-857.

5. Griffith, M., Antikainen, M., **Hon, W.-C.**, Pihakaski-Maunsbach, K., X.-M. Yu, Chun, J. U., and Yang, D. S.-C. (1997). Antifreeze proteins in winter rye. *Physiologica Plantarum.* 100:327-332.

6. **Hon, W.-C.**, McKay, G. A., Thompson, P. R., Sweet, R. M., Yang, D. S. C., Wright, G. D., and Berghuis, A. M. (1997). Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell.* 89:1-20.

7. Yang, D. S.-C., **Hon, W.-C.**, Bubanko, S., Xue, Y.-Q., Seetharaman, J., Hew, C. L., and Sicheri, F. (1998). Identification of the ice-binding surface on a type III AFP

using a "flatness function" algorithm. *Biophys. J.* in press.

This list represents my contribution to two distinct projects during my graduate studies. The first, which I have not written up as a thesis, involved the isolation and characterization of a novel group of antifreeze proteins from the leaves of cold-acclimated winter rye. The work was, at that time, a collaborative effort between Dr. Marilyn Griffith's group (University of Waterloo) and Dr. Daniel Yang's group (McMaster University). Dr. Mervi Antikainen (Univerisity of Turku, Finland) later joined Griffith's group as a visiting Ph.D. student and further characterized these antifreeze proteins.