STRUCTURE-BASED MUTATIONAL ANALYSIS OF S. POMBE $p13^{suc1}$

OKI KWOSHI DZIVENU

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of philosophy in the University of Oxford.

Laboratory of Molecular Biophysics
and Wolfson College, Oxford

Trinity Term, 1999.

SCIENCE LIBRARY OXFORD
RADCLIFFE
30 SEP 1999
THESIS

Structure-based Mutational Analysis of *S. pombe* p13^{suc1}

Oki Kwoshi Dzivenu
Laboratory of Molecular Biophysics and Wolfson College, Oxford

Submitted for the degree of Doctor of Philosophy
Trinity Term, 1999

Abstract

p13^{suc1} from *schizosaccharomyces pombe* is a member of a family of non-enzymatic cell cycle regulatory proteins called CKS for Cyclin-Dependent kinases Subunit. Other members of this family include CKS1 (*S. cerevisiae*), CksHs1 and CksHs2 (Homo sapiens). The CDKs (CDK1-CDK8) for Cyclin-Dependent Kinases are a class of Ser/Thr kinases that regulate the cell cycle. The *suc1*$^{+}$ gene was initially identified as a seppresor of certain CDK1 temperature sensitive mutations. Despite the wealth of crystallographic models available plus ample – albeit, sometimes conflicting - evidence from genetics and biochemical studies, an account of the exact physiological role of the CKS proteins remains an elusive goal. In a quest to identify the residues and hence the particular surface region involved in mediating protein-protein interactions with CDK2, I employed the X-ray structure of Suc1 at 2.7A resolution as guide for a site-specific mutagenesis study. Comparative biochemical and biophysical characterisation of Suc1 and the mutants led to the conclusion that isoleucine-94 and Leucine-96 (located in the hydrophobic patch) are involved in mediating protein-protein interactions with GST-CDK2. This conclusion has since been confirmed by the publication of the X-ray structure of monomeric CksHs1 in a complex with CDK2 by Bourne et al., 1996 (Cell 84: 863-874). An extension of the mutational study to test the hypothesis that Suc1 may utilise conserved residues at the anion-binding site to mediate protein-protein interactions with the Anaphase Promoting Complex (APC) is described. X-ray data has been collected on wild type Suc1 crystals at 100K to 2.3Å resolution. The structure has been resolved and refined to a crystallographic R-factor of 22.6%. *S. pombe* Suc1 exists as a zinc-stabilised, non strand-exchanged dimer in both the 2.7Å and 2.3Å models. Structural analyses of two Suc1 folding mutants are also presented. The cyclins (A - H) are positive regulatory subunits of CDKs. They share a high degree of homology over a region of about 100 amino acid residues called the "cyclin box". The determination of the crystal structure of cyclin A3 (an N-terminal truncated version of bovine cyclin A) and a CDK2-cyclin A3 complex by other workers has revealed the mechanism of activation of CDKs by cyclins. In *S. pombe,* the CDK1-cyclin B heterodimeric complex constitutes the mitotic kinase. In order to understand the specificity underlying the CDK-cyclin interaction, I embarked on a structural study of *S. pombe* cyclin B and CDK1. Both full-length proteins have proven intractable to attempts to overproduce them in a soluble form in *E. coli* for crystallisation studies. A truncated version of cyclin B (similar to cyclin A3) was designed, cloned and overproduced in *E. coli*. The cyclin B3 protein was directed into inclusion bodies as insoluble aggregates. Extensive attempts – both *in vivo* and *in vitro* - to produce a soluble cyclin B3 proved unsuccessful. Finally, an *E. coli* co-expression system designed to overproduce CDK1-cyclin A3, CDK1-cyclin B3, CDK1-cyclin B and CDK1-Suc1 complexes is described.

"The man dies in all who keep silent in the face of tyranny"

Wole Soyinka, 1971.

Acknowledgements

On numerous occasions, I've come out of a discussion meeting with my supervisor, Professor Louise N. Johnson with the firm conviction that I was a fellow of the royal society (F.R.S). This state of delusion lingers till I complete the journey from her office to S1 (the lab) to don a lab coat and pick up a pipette. Her unceasing ability to point out the positive in an avalanche of negative results is enough to banish the blues and shore up one's confidence to the extent that, I always feel like going back to work hard till I drop dead! I am very grateful to her for the rare opportunity to study in LMB, years of coaching and lots of intellectual encouragement. I will also, like to thank Dr. Jane Endicott for equipping me with invaluable techniques in gene manipulation and various aspects of wet work. A positive attribute I yearn for, is Dr. Nick Brown's thorough knowledge of protein biochemistry and his knack for getting all the details from a publication after just a single reading! I owe a huge debt of gratitude to Dr. Elspeth Garman for revealing to me (with oceans of patience) that the computer sitting on my desk could do much more than send and receive electronic mail. Apart from Elspeth, I thank Dr. Martin Noble for taking on the onerous task of attempting to demystify protein X-ray crystallography to me. Dr. David Barford has been my advisor for the past five years and I am very grateful for his assistance throughout. I am indebted to Stephen Lee for assistance with the figures, Irene Taylor and Mike Pickford for technical assistance, Richard Bryan and Yuguang for computing. I thank other members (Simon, Paul and Alison) of the cell cycle group for their support.

Long, Long before Oxford, there was Tema, Ghana where a bouncing baby boy was born to Janet and Vincent Dzivenu. The immense love that these two beings showered on Robert/Bob is the reason why the author of this thesis has come this far in life. My brothers (Dela & Kwaku Edem) and my only sister (Ama) have consistently made my world a highly enriched and blessed one; their love and support is immeasurable. Many thanks to the people of Sasieme, Nima and Nungua.

Acknowledgements

Then came Mawuli School, Ho, where I spent my formative years. The memories are as vivid as ever! Frequently, I've been forcibly woken up from sleep by my wife to explain why I was laughing aloud. Invariably, this rude awakening for an explanation always tended to be a disruption of a most beautiful dream in which Arnold Degboe, Edem Ameko, Jeremiah Nyaho, Mickylow and myself were either being chased away from the girls dormitory (after the lights-out call) by the caretaker or some other pleasant mischief. Those days will never be forgotten! After Mawuli, London: I owe a debt of gratitude to Prof. Shoolingen-Jordan, Dr. Martin Warren and Dr. Sarah-Peters. Also, I thank Nehar G, Rozzy and Moses Mushipe (Vuka!) for all the good times at Queen Mary & Westfield College. Other notable contributors to my success and welfare are Demia (Mr. E. Kaikai, F.R.S), Charley, Akofa fofo (Mr. Wilson Hormenu), Akofa Dada (Auntie Charity), John Townsend and Mr. Moffat at St. Thomas More RC School.

Back to Oxford: this village has so far, enriched my life with two of the most beautiful beings on the planet. A wife who is the living embodiment of beauty –both internal and external- and a son who is a daily reminder of God's miraculous presence in one's life. I shall be eternally grateful to Oxford for Leonora and Tornyeli. I must not forget my in-laws – Auntie & Papa Leigh – for all those days of prayers, laughter and much "provoking"! Some of the numerous people who have made the lab and Oxford in general, a joyful place to live and work in are Mary, John, Andrew, David, Eilis, Peter, Neil and the Mayosi family. Thanks to Julie Tucker for allowing me to use her Macintosh computer and also, for all those stimulating, nocturnal discussions on any topic under the sun! MaxRes to the VIP crowd of Mulenga Pamu, Alfred, Justin, Thomas, Judas. Kilavuka, C-D'dee, Einstein, Ben Accam, Lady-M, Auntie, Martha, Romeo, Joseph Somevi, Titanic and Aligator. Susan Hall has been a most generous and kind neighbour.

The monotony of a prolonged campaign of writing was frequently broken by interesting discussions/fights on Okyeame, Afrika Think Tank and Ghanaforum. I'm grateful to the members of these Ghanaian internet discussion fora.

Apart from thanking my sponsors (M.R.C), I shall end this verbose piece by registering my profound gratitude to Elspeth for attempting to, and succeeding in easing the overwhelming grief I felt at the loss of my dearest – Sista! Elspeth, your ability to empathise is unparalleled; thank you very much!

"When you do your best and it is not good enough, it is time to realise that, it is enough to do your best"

Kwamena Kwansah-Aidoo, May 1998

This is a special dedication:

To Vincent Oki Kwasi Dzivenu
&
Janet Yaowovi Dzidzornu
for their unceasing and unconditional love

To Leonora Leigh & Tornyeli Dzivenu
for continuing to share with me, a love that is very unique
and sweet, even at the most galling of times

Sista!

Angels smiled,
On this
Destiny's child
A
Woman personified.

In the past,
When I was last,
You
Made me first.

Head decorated with sores,
Social outcast,
But
You
Remained
My Sista.

When others denied me,
Crucified me,
You
Were there
In
My corner.

When the world was crowded,
And
I was lonley,
You were
All of that
And more.

And so
This day
I've but one
Thing to say
Thanks Sista
For
Being
My mother.

(Adapted from an original by Rodney Coates at www.ulbobo.com/umoja).

Abbreviation		Meaning
1.	ATP	Adenosine 5'-triphosphate
2.	B.M.	Boeringher Mannheim
3.	DNA	Deoxyribonucleic acid
4.	DTT	dithiothreoitol
5.	DTE	dithioethylene
6.	EDTA	ethylenediamine tetra acetic acid
7.	EGTA	Ethyleneglycol-bis-(β-amino ethyl ether) N,N,N',N'-tetraacetic acid
8.	EtBr	Ethidium bromide
9.	HEPES	(N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
10.	HBS	HEPES-buffered saline
11.	IPTG	isopropylthiol-β-D-galactoside
12.	LB	Luria broth
13.	MOPS	(3-[N-morpholino]propanesulfonic acid) buffer
14.	N.E.B	New England Biolabs
15.	O.D.	Optical density
16.	PAGE	Polyacrylamide gel electrophoresis
17.	PEG-6000	Polyethylene glycol-6000
18.	PMSF	Phenylmethanesulfonyl flouride
19.	r.t.	Room temperature
20.	SDS	Sodium dodecyl sulphate
21.	TAE	Tris-acetate EDTA buffer
22.	TBE	Tris-borate EDTA buffer
23.	Wt	wild type

CONTENTS

Chapter 1 Introduction 1

Chapter 2 The crystal structure of *S. pombe* p13^{suc1} at 2.3 Angstroms (Å) resolution 60

Chapter 3 Structural analysis of *S. pombe* p13^{suc1} folding mutants 84

Chapter 4 Mutational analysis of *S. pombe* p13^{suc1} 108

Chapter 5 Co-expression of heterodimeric cell cycle regulatory proteins in *E. coli* 187

Chapter 6 The targeting model of Cks function 260

Appendix I A brief account of protein crystallography 271

"Nothing is permanent but change"

- Heraclitus.

Chapter 1: Introduction

CONTENTS

LIST OF FIGURES **2**

LIST OF TABLES **3**

1.1 THE CELL CYCLE **3**

1.2 THE CYCLIN-DEPENDENT KINASES (CDKS) AND THEIR CYCLIN SUBUNITS **5**

1.2.1 Primary CDKs and their cognate cyclins *7*
1.2.2 Secondary CDKs and their cognate cyclins *8*
1.2.3 Single cyclins *12*

1.3 MULTIFACETED REGULATION OF CDKS **13**

1.3.1 Activation by cyclin binding *14*
1.3.2 Activation by phosphorylation and dephosphorylation *18*
1.3.3 Inhibition by both phosphorylation and dephosphorylation *19*
1.3.4 Inhibition by the binding of inhibitory protein subunits *20*
1.3.5 Structural modulation by Cks binding *25*
1.3.6 A summary (structural) of CDK regulation *26*

1.4 CHECKPOINT REGULATION 28

1.4.1 G1 to S-phase 28
1.4.2 G2 to M-phase 31

1.5 REGULATED PROTEOLYSIS IN CELL CYCLE CONTROL 33

1.5.1 The Anaphase Promoting Complex (APC) pathway 34
1.5.2 The Skp1-Cullin-F-box (SCF) pathway 36

1.6 THE CYCLIN-DEPENDENT KINASE SUBUNITS (CKS) PROTEINS 39

1.6.1 Biological importance of the Cks proteins 39
1.6.2 A wealth of X-ray structures 40
1.6.3 Modulation of CDK2 structure by the binding of Cks1Hs2 43
1.6.4 Functional models of Cks function 43
1.6.5 A summary of cell cycle control 44
1.6.6 A preview of the thesis 45

1.7 LITERATURE CITED 46

List of figures

Figure 1.1 Stages of the cell cycle ---------- 3
Figure 1.2 Cellular events that impinge on the core cell cycle apparatus ---------- 9
Figure 1.3 Structure of ATP-bound human HsCDK2 ---------- 13
Figure 1.4 Regulation of CDK activity ---------- 14
Figure 1.5 The X-ray structures of cyclins A and H ---------- 15
Figure 1.6 HsCDK2-HsCyclin A3 complex structure ---------- 16
Figure 1.7 Conformational changes induced by T160 phosphorylation ---------- 18
Figure 1.8 Phosphorylated HsCDK2-HsCyclin A3- $p27^{Cip2}$ ternary complex ---------- 21
Figure 1.9 Structural rearrangements induced by the binding of p27 to phosphorylated HsCDK2-HsCyclin A3 ---------- 22
Figure 1.10 Reshaping of the HsCDK2 N-lobe by p27 ---------- 23
Figure 1.11 CDK6-$p16^{INK4a}$ complex structure ---------- 24
Figure 1.12 The structure of HsCDK2-CksHs1 complex ---------- 26
Figure 1.13 Mechanistic insights gained from structural analysis of monomeric CDK2 and various complexes ---------- 27
Figure 1.14 Schematic representation of human pRB, p107 and p130 proteins ---------- 29
Figure 1.15 The G1→S-phase molecular switch ---------- 30
Figure 1.16 Control of G2→M-phase ---------- 32
Figure 1.17 The ubiquitination process ---------- 34
Figure 1.18 Substrate-specific activators of the APC ---------- 35
Figure 1.19 Ubiquitination of specific substrates by SCF complexes ---------- 37
Figure 1.20 Regulated proteolysis and cell control in S. cerevisiae ---------- 38

Figure 1.21 The dual conformational state of SpSuc1 ---------- *41*
Figure 1.22 Summary of cell cycle regulation ---------- *45*

List of Tables

Table 1.1 Effect of the overproduction or deletion of Cks proteins in different organisms _____ *40*

1.1 The cell cycle

After a proliferating cell has attained a certain critical mass, it duplicates its genome by replication (S-phase) and then divides mitotically (M-phase) into two daughter cells with identical genomes. These two key events are punctuated by growth periods or gaps called G1 and G2. Thus, a period of growth (G1) followed by DNA replication (S-phase) and a second period of growth (G2) followed by cell division (M-phase) constitute a cyclical series of events, of uneven duration, referred to as the cell cycle (Norbury and Nurse, 1992; Nurse *et al.*, 1998).



Figure 1.1 Stages of the cell cycle *The goals of this process are to faithfully replicate DNA and segregate the two copies (nuclear) and also, duplicate and segregate organelles (cytoplasmic). Reproduced from (Weinberg, 1996).*

Regulation of the cell cycle is maintained at two main levels. The first is a multi-faceted regulation of a family of kinases called cyclin-dependent kinases (CDKs). The second is through the erection of a few barriers along the route (in G1and G2 phases) manned by a group of proteins called checkpoint regulators (Collins *et al.*, 1997). Regulation of the CDKs is responsible for ensuring a tightly orchestrated transition from

one phase of the cell cycle to the other i.e. control of cell proliferation, while checkpoint regulators are required to detect aberrations such as improperly replicated DNA and respond by either slowing down the pace of the cell cycle so that DNA repair enzymes can rectify that damage or shepherd that particular cell to a state of quiescence known as G_0. The imperative to understand the regulation of the cell cycle in as much detail as possible becomes obvious when we examine the debilitating consequences on a host organism when the regulatory mechanism governing cellular division goes awry. Cancer can be defined simply as a disease of uncontrolled cell proliferation brought about by the deregulation of the cell cycle machinery (Collins *et al.*, 1997; Kamb, 1995; Stinchcomb, 1995). The affluent parts of the world have largely overcome poverty-related diseases such as tuberculosis, cholera etc... but the scourge of cancer – in all its myriad forms – continues to exact a high toll on human life in these regions. The forecast is that about 40% of North Americans may suffer a form of cancer and about 20% of these may be fatal (Ames *et al.*, 1995; Rennie and Rusting, 1996). The global trend is not too dissimilar (Gorman, 1998). So, apart from its intrinsic scholarly value, a detailed understanding of the cell cycle would hopefully assist pharmaceutical research scientists in their quest to find means of preventing, eliminating or managing diseases that arise from anomalous or perturbed cell cycles.

Like most subfields in biomedicine, our current knowledge about the cell cycle has been produced along the following lines of enquiry; the geneticist identifies and isolates the gene(s) whose products are implicated in the regulation of the cell cycle, the biochemist then characterises this gene product(s) in terms of physico-chemical and sometimes functional properties. The structural biologist closes this circle by presenting us with an atomic model of the protein. It is obvious from this simplistic account of knowledge accumulation that in most cases, structural knowledge would tend to lag behind those from genetic and biochemical investigations. This means that atomic models are published long after some genetic and biochemical models have become the accepted wisdom in the field. This places the structural biologist in the enviable position of either confirming and subsequently elaborating on accepted biochemical models or completely overturning and replacing them with new ones gleaned from structural data. Sometimes,

the paucity of genetic and biochemical data coupled with the peculiar arrangement of a particular protein molecule in a crystal structure has led many a structural biologist to propose wrong (with the benefit of hindsight) mechanistic models for proteins whose three-dimensional structures have been elucidated. This buttresses the point that an interdisciplinary approach to cell cycle research is the most effective way of understanding this complex phenomenon. All the available textbooks in this field (Hutchison and Glover, 1995; Meijer *et al.*, 1995; Mitchison, 1971; Murray and Hunt, 1993; Ruffolo *et al.*, 1997) provide – to varying extents - a rich account of the genetics and biochemistry of the cell cycle but lack a cogent account of the contribution of atomic structures to the understanding of this complex and rapidly evolving field. Instead, one has to consult primary journals and reviews for this sort of information. In this introductory chapter to the thesis, the reader will be presented with the basics of cell cycle regulation with the emphasis on the mechanistic insights gleaned from the atomic structures of some cell cycle regulatory proteins.

1.2 The Cyclin-Dependent Kinases (CDKs) and their cyclin subunits

A Cyclin-Dependent Kinases (CDK) can be defined as a serine/threonine kinase that requires a cyclin subunit for activation. All CDKs possess a signatory motif at the N-terminus which is as close as possible to the consensus sequence PSTAIRE i.e. a proline (P) residue followed by a serine (S), threonine (T), alanine (A), isoleucine (I), arginine (R) and ending with a glutamate (E). According to this definition, there are so far, eight known CDKs in higher eukaryotes termed CDK1 - CDK8. In common with nearly all kinases, CDKs possess an ATP binding motif, which is signified by the Gly-X-Gly-X-X-Gly (where X is any amino acid). In general, CDKs consist of about 300 amino acid residues and they share a high degree of homology (35-65%). With regards, to the protein kinase superfamily, CDKs are no larger than minimal protein kinase catalytic units (Hardie and Hanks, 1995; Johnson *et al.*, 1998; Johnson *et al.*, 1996; Morgan, 1997).

The cyclins are a family of structurally related proteins that are synthesised and degraded in a cell cycle-dependent manner. They range in size from about 35 to 90 Kilodaltons (KDa). Each cyclin is given an alphabetical suffix as it is discovered. To date, nine cyclins (A-I) have been discovered in vertebrates (Edwards *et al.*, 1998; Nakamura *et al.*, 1995; Pines, 1995; Wei *et al.*, 1998). The cyclins can be divided into three broad groups (depending on when they are synthesised and where they act in the course of the cell cycle); the mitotic cyclins (A and B), the G1 cyclins (D and E) and the transcription cyclins (C, F, G, H, I, K and T). The exact physiological role of members of the latter group is yet to be characterised in detail. Multiple sequence alignment, site-specific mutagenesis studies and structural elucidation of some cyclins has led to the identification of conserved regions that represent functional domains. All cyclins possess a region of about 100-150 amino acids called the cyclin box. These residues determine the specificity of the cyclin-CDK interaction (Kobayashi *et al.*, 1992; Lees and Harlow, 1993; Noble *et al.*, 1997). A consensus sequence at the N-terminus called the destruction (D) box is present in the cyclins that are degraded at the end of mitosis (Glotzer *et al.*, 1991). A requirement for exit from mitosis is the degradation of mitotic cyclins via the poly-ubiquitination of lysine residues near the destruction box (section 1.5). The consensus D-box motif in B-type cyclins is RXALGXIXN where R is an arginine residue, A is alanine, L is a leucine, I is an isoleucine, N is an asparagine and X represents any amino acid (Glotzer *et al.*, 1991). Apart from cyclin B3, the arginine at position 1 and the leucine at position 4 are highly conserved in all B-type cyclins. Another signatory motif present in these proteins is the PEST sequence i.e. proline (P), glutamate (E), serine (S) and threonine (T). PEST sequences signify an unstable protein that is rapidly turned over (Rechsteiner and Rogers, 1996). Cln2 and Cln3 of *S. cerevisiae* contain PEST sequences. That in itself is not a targeting or degradation motif. For degradation to occur, residues in the PEST motif have to be phosphorylated (Rechsteiner and Rogers, 1996).

Due to the rapidity with which new CDKs and cyclins are being discovered and the fact that at present, we know little of the substrates of CDK-cyclin complexes, the field of cell cycle research has gained a reputation for complexity (Holloway, 1996).

However, a critical examination of the literature has led to the classification of the CDKs into two subfamilies of core and auxiliary cell cycle regulators (Morgan, 1997). In the next section, the CDKs and cyclins that are required for core cell cycle control (designated primary CDKs and cyclins) would be distinguished from those which play other cellular functions in addition to cell cycle regulation (designated secondary CDKs and cyclins).

1.2.1 Primary CDKs and their cognate cyclins

Cdc28 from *Saccharomyces cerevisiae* and cdc2 from *Schizosaccharomyces pombe* are the founding members of the CDK family of serine/threonine protein kinases. Genetic screens in fission yeast resulted in the identification of a gene called *cdc2* - short for cell-division cycle 2 (Nurse and Bissett, 1981). The product of this gene was designated $p34^{cdc2}$ because it encoded a protein of 34KDa. Although several putative CDKs were subsequently isolated from vertebrates (Draetta *et al.*, 1988; Meyerson *et al.*, 1992; Nigg, 1995), cdc2 and CDC28 remained the only CDKs in yeast till the recent discovery of mcs6 from *S. pombe* and kin28 from *S. cerevisiae* (Molz and Beach, 1993; Valay *et al.*, 1993). Cdc2 was renamed CDK1 and the other members of this family of enzymes were given numerical suffixes as they were discovered. The basic cell cycle regulatory activity of CDK1 is well conserved from yeast to humans. In *S. pombe*, CDK1 in a complex with cyclin B (also called cdc13) is responsible for mitotic regulation. Two other cyclins called cig1 and cig2 form a complex with CDK1 to initiate and regulate DNA synthesis in *S. pombe* (Stern and Nurse, 1996). The number of known CDKs and cyclins increases as one moves from *S. pombe* through *S. cerevisiae* to humans. Three different cyclins called Cln1-3 are required to regulate the G1 function of CDK1 in *S. cerevisiae*. Another set of cyclins called Clb1-4 and Clbs5 and 6 are required for the initiation and control of mitosis and DNA synthesis, respectively (Nasmyth, 1996). In vertebrates, the CDK2-cyclin E complex is required to initiate DNA synthesis in S-phase (Donnellan and Chetty, 1999). After this event, the cyclin E is degraded and cyclin A synthesis begins. The latter replaces cyclin E in the CDK2-cyclin complex to oversee progress through S-phase (Donnellan and Chetty, 1999). Degradation of cyclin A begins

at the end of S-phase: the remnants form a complex with CDK1 and this is thought to prepare the cell in G2 for M-phase. Cyclin A degradation is completed shortly after entry into mitosis. CDK1-cyclin A kinase complexes are situated in the nucleus where they are active (Cardoso *et al.*, 1993; Pines and Hunter, 1991). The CDK1-cyclin B complex is responsible for initiating and regulating mitosis. The CDK1-cyclin B complex accumulates in the cytoplasm as an inactive kinase due to inhibitory phosphorylation (section 1.3.4). Just before nuclear envelope breakdown at prophase, the CDK1-cyclin B complexes are activated and they enter the nucleus (Pines and Hunter, 1991). It appears that core cell cycle regulatory functions can be performed by a single CDK (yeasts) and two CDKs (vertebrates) in conjunction with their cognate cyclin partners. What then is the role of the other CDKs discovered in yeasts and vertebrates?

1.2.2 Secondary CDKs and their cognate cyclins

The plethora of CDKs and cyclins we see in vertebrates does not correspond to a radically different or more complex cell cycle machinery in these organisms. It appears that other CDKs and cyclins have evolved to play diverse cellular roles like transcriptional regulation (Dynlacht, 1997; Sanchez and Dynlacht, 1996) that impinge on cell cycle regulation (fig. 1.2).



Figure 1.2 Cellular events that impinge on the core cell cycle apparatus

In lower eukaryotes, the core cell cycle apparatus (consisting of a single CDK and a few cyclins) regulates normal cellular proliferation. This feature is conserved in higher eukaryotes where the core cell cycle apparatus plays an additional role as a choreographer of cellular function (Lees and Weinberg, 1999). The latter role is achieved by using secondary CDKs and their cyclin partners to acquire growth-regulatory signals from diverse sources (red and green arrows), integrate them into the basic cell cycle apparatus, process the new information, make some majors decisions before relaying signals to various cellular targets situated in different parts of the cell (Jacks and Weinberg, 1998; Lees and Weinberg, 1999). This additional role of the core cell cycle apparatus in vertebrates explains the evolution of extra CDKs and cyclins whose broad functional role(s) include the relaying of signals to and from the core cell cycle apparatus and other diverse cellular units. Reproduced from (Weinberg, 1996).

In the absence of the core cell cycle regulators, these secondary CDK-cyclin complexes can sometimes perform the function of the core cell cycle CDKs (Morgan, 1997). For example, in *S. cerevisiae*, the Pho85 CDK and cyclins Pcl1 and Pcl2 have

been implicated in cell cycle control based on the observation that a Pho85-Pcl complex can maintain the cell through G1 in the absence of Cdc28-Cln complexes (Lenburg and O'Shea, 1996), but the defined role for this CDK lies in the regulation of gene expression (Santos *et al.*, 1995). Similarly, a poorly defined cell cycle regulatory role has been ascribed to CDK3 based on primary amino acid similarity to CDK1 and CDK2 and its ability to rescue *S. cerevisiae* cells deficient in CDK1 but a cyclin partner is yet to be isolated for this protein (Hofmann and Livingston, 1996).

One of the well-established examples of a CDK coupling other cellular signals to the cell cycle can be seen in the action of CDK4/6 and the D-type cyclins. As described in section 1.4, the decision to commit to a round of cell division depends on factors such as the availability of nutrients. CDK4/6-cyclin D complexes are formed in vertebrate cells in response to growth factors (Sherr, 1996). These complexes phosphorylate the retinoblastoma protein pRB (described in section 1.4.1). Two closely related proteins to pRB (p107 and p130) are phosphorylated by CDK2-cyclin A/E complexes. These phosphorylated pocket proteins respond by releasing bound transcription factors (E2Fs) which in turn trigger the expression of genes that allow the cell cycle to traverse late G1 and into S-phase (Sherr, 1996). It has been shown that pRB deficient cells can progress through G1 in the absence of CDK4/6-Cyclin D complexes (Sherr, 1996). Thus, the CDK4/6-Cyclin complex can be seen as the link between the core cell cycle machinery, extracellular signals and the transcriptional apparatus.

In common with most CDKs, CDK5 was cloned via homology to yeast CDK1 (Lew and Wang, 1995). CDK5 and its cyclin partner p35 are found mainly in terminally differentiated neuronal cells (Tang *et al.*, 1997). This represents one of the few examples where a CDK-cyclin complex is found in a non-dividing cell, performing a role that is not connected to cell cycle regulation. CDK5-p35 complexes are known to phosphorylate tau (a neuron-specific microtubule associated protein) *in vivo* (Lew and Wang, 1995). Improper phosphorylation of these proteins is associated with neurodegenerative diseases such as Alzheimers.

In vertebrates, CDK7 in a complex with cyclin H can phosphorylate CDK1 *in vitro* (on T161), CDK2 (on T160), CDK3 (on T167) and CDK4 (on T172) (Kato *et al.*,

1994; Larochelle *et al.*, 1998; Morgan, 1995). The CDK7-cyclin H complex is further stabilised by a third protein called MAT1 (Devault *et al.*, 1995; Tassan *et al.*, 1995a). The ability of this complex to phosphorylate some CDKs on the activatory residues is known as CAK (for CDK-activating kinase) activity (Fisher *et al.*, 1995). CDK7-cyclin H complexes have been isolated in association with the general transcription factor THIIH where the CDK-cyclin complexes act as the CDKs that phosphorylate the C-terminal domain (CTD) of RNA polymerase II (RNAPII) during transcription (Svejstrup *et al.*, 1996; Yankulov and Bentley, 1997). This is another example where a CDK couples a different cellular function to core cell cycle regulation.

A cDNA encoding a protein called Cak1/Civ1 has been isolated from *S. cerevisiae* (Thuret *et al.*, 1996). Although Civ1 is a kinase, it lacks the signatory PSTAIRE motif that is present in all CDKs and at a molecular weight of 45KDa, it is about 10KDa bigger than the average CDK protein. Civ1 also lacks the canonical consensus Gly-X-Gly-X-X-Gly (where X is any amino acid) that is present in the ATP-binding loop of most protein kinases (Draetta, 1997; Thuret *et al.*, 1996). In *S. cerevisiae*, Civ1 is required for CDK1 activation *in vivo*. Civ1 also differs from CDK7 in that it does not require additional protein subunits or modifications to become active. This begs the question "which is the real CAK *in vivo*?" The CDK7-cyclin H homologue in *S. pombe* is the Mcs6-Mcs2 complex which possesses CAK activity but a MAT1 homologue is yet to be isolated and Mcs6-Mcs2 kinase has not been shown to be indispensable to CDK1 phosphorylation *in vivo*. In contrast, the *S. cerevisiae* homologue of CDK7 called kin28 does not possess CAK activity but like CDK7, it associates with yeast TFIIH where it phosphorylates the CTD of RNAPII. There are two plausible answers to the above question. *S. cerevisiae* may have evolved two separate molecules (Civ1 and Kin28) to perform functions that can be executed by a single molecule (CDK7) in higher eukaryotes or CDK7 may not be the real CAK *in vivo* in the latter group of organisms. As yet, a Civ1 homologue is yet to be isolated from vertebrates.

Vertebrate cyclin C was originally isolated via its ability to rescue *S. cerevisiae* cells deficient in G1 cyclin function. The kinase partner CDK8 was isolated later and it was shown to form a direct complex with cyclin C (Leclerc *et al.*, 1996; Tassan *et al.*,

1995b). The CDK8-cyclin C kinase associates with RNAPII and it also phosphorylates the CTD of the polymerase (Hengartner *et al.*, 1998; Maldonado *et al.*, 1996; Rickert *et al.*, 1996).

CDK9 belongs to the PITALRE family of CDK-like protein kinases (Marshall and Price, 1995; Yang *et al.*, 1997). It is a 42KDa molecule whose cyclin partner is a C-type, 87KDa protein called cyclin T (Wei *et al.*, 1998). CDK9 is an RNAPII transcription elongation factor and the kinase hyperphosphorylates RNAPII CTD *in vitro* (Yang *et al.*, 1996).

Human Cyclin G has a cyclin box domain which is most similar to those of human cyclin A and *S. pombe* Cig1 (Bueno *et al.*, 1991; Tamura *et al.*, 1993). Since then it has been demonstrated that cyclin G is a transcriptional target of the p53 tumour supressor protein (Okamoto and Beach, 1994; Zauberman *et al.*, 1995) suggesting that it may play a role in the connection between cell cycle control and apoptosis. The biochemical regulatory mechanism and physiological function of cyclin G is unknown. A possible kinase partner for cyclin G, called GAK for cyclin G-associated kinase has been cloned (Kanaoka *et al.*, 1997). GAK migrates as a 160KDa molecule and the N-terminal region of this protein encodes a Ser/Thr kinase. It has a long C-terminal extension which shares homology with tensin and auxilin. GAK-cyclin G association has been demonstrated by immunoprecipitation, western blot and BIAcore analyses but as yet, the role of this complex is unknown. Since GAK lacks a PSTAIRE motif, it is evidently not a core cell cycle regulatory kinase but it may have a putative role in connecting other cellular processes such as apoptosis to cell cycle regulation (Kanaoka *et al.*, 1997).

1.2.3 Single cyclins

Human cyclin I was cloned from a cDNA library of forebrain cortex (Nakamura *et al.*, 1995). It contains a cyclin box at the N-terminus that is most similar to those of cyclins G and E and the C-terminus is rich in PEST sequences. Cyclin I is expressed at constant levels throughout the cell cycle (Nakamura *et al.*, 1995).

Human cyclin K was isolated through its ability to rescue *S. cerevisiae* cells lacking Cln1/2/3 (Edwards *et al.*, 1998). It consists of 357 amino acids and it has a cyclin

box which is most related to those of cyclins C and H. Homologues have been isolated from mouse and *Drosophila melanogaster*. Cyclin K coimmunoprecipitates with the large subunit of RNAPII and in a complex with it as yet unidentified kinase partner, it can phosphorylate the CTD *in vitro* (Edwards *et al.*, 1998).

Human cyclin F was isolated as a suppressor of a G1→S deficiency of an *S. cerevisiae* Cdc4 mutant (Bai *et al.*, 1994). It is an 87KDa molecule that contains extensive PEST sequences at the C-terminus and a cyclin box domain that bears similarity to those of A and B type cyclins. The identification of a conserved stretch of 43 amino acids (originally observed in cyclin F) in a host of other proteins including Cdc4 has given birth to a new family of molecules called the F-box proteins. They are important for regulated proteolysis of specific substrates (section 1.5.2).

1.3 Multifaceted Regulation of CDKs

The X-ray structure of monomeric, unphosphorylated human CDK2 (HsCDK2) was determined in 1993 (De Bondt *et al.*, 1993). This is an inactive molecule with a global topology that is similar to the prototypical kinase fold as first observed in the structure of the cAMP-dependent protein kinase (PKA) (Johnson *et al.*, 1998; Johnson *et al.*, 1996; Knighton *et al.*, 1991; Pavletich, 1999).

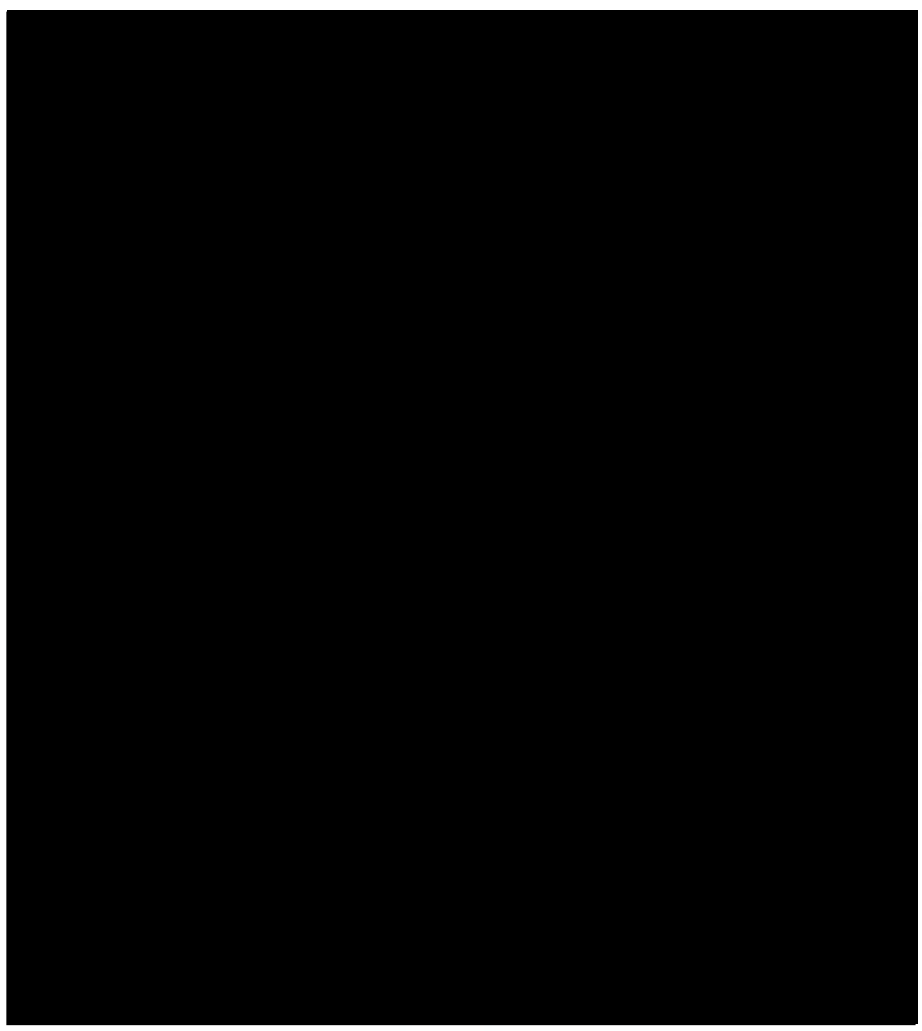


Figure 1.3 Structure of ATP-bound human HsCDK2

The overall structure consists of an active site sandwiched between two structural domains: a smaller N-terminal lobe, comprising mainly beta sheets and a relatively larger C-terminal lobe, which is mainly α-helical. The PSTAIRE

helix is located at the entrance of the active site cleft. The ATP substrate sits in a hydrophobic pocket in the active site cleft. It is depicted in a ball-and-stick representation. The T-loop (residues 146–170) is highlighted in orange. This figure was produced with Aesop (Dr. Noble, unpublished work).

Before the X-ray structure was published, biochemical evidence showed that CDKs were regulated by at least five post-translational mechanisms (Martin-Castellanos and Moreno, 1997; Morgan, 1995). They are:

(i). Activation by cyclin binding

(ii). Activation by phosphorylation and dephosphorylation

(iii). Inhibition by dephosphorylation

(iv). Inhibition by the binding of inhibitory (CKI) proteins

(iv). Modulation by the binding of the Cks proteins



Figure 1.4 Regulation of CDK activity
Cyclin binding can render an inactive CDK partially active. Phosphorylation on T160 in CDK2 by CAK makes the CDK-cyclin complex fully active. The binding of members of the CKI family of proteins can abolish the activity of the kinase complex. Reproduced from (Pavletich, 1999).

1.3.1 Activation by cyclin binding

Whereas CDK levels remain constant throughout the cell cycle, cyclins are produced and degraded as and when they are required. The binding of a cyclin to a CDK serves to activate the latter and also target the heterodimeric complex to particular substrate targets (Schulman *et al.*, 1998). The crystal structure of a fragment (residues 172 – 432) of bovine cyclin A3 (BvCyclin A3) was reported in 1995 (Brown *et al.*,

1995). That of human cyclin H (HsCyclin H) has also been determined (Anderson *et al.*, 1997; Kim *et al.*, 1996).



Figure 1.5 The X-ray structures of cyclins A and H

(a) Both BvCyclin A3 and (b) HsCyclin H consist of α-helical secondary structure elements. The conserved core of each molecule contains 10 helices which are numbered α1-α5 for the first domain and α1'-α5' for the second domain. The main difference between the two proteins lies outside of the core where the amino- and carboxy-terminal (N and C, respectively) helices differ in their relative positions. This suggests that cyclin H cannot interact with CDK2 in the manner observed for cyclin A in the X-ray structure of the complex (see below). Reproduced with modification from (Morgan, 1996).

The details of the mechanism by which a cyclin molecule activates a CDK was revealed by the publication of the HsCDK2-HsCyclin A3 complex structure (Jeffrey *et al.*, 1995). Structural comparison of the inactive HsCDK2 and partially active heterodimeric complex explains why the former cannot phosphorylate target substrates.

There are two main reasons why an unphosphorylated, monomeric CDK molecule is inactive: first, the large flexible T-loop (so-called because it contains the phosphorylatable threonine 160) blocks substrate binding at the entrance of the active site cleft. Second, helix L12 that is situated near the active site cleft disrupts the position of the PSTAIRE helix of the N-lobe. This results in the mispositioning of catalytic residues involved in ATP binding. The orientation of the beta (β) and gamma (γ) phosphates of the bound ATP substrate is perpendicular to the optimal alignment necessary for phosphotransfer (Jeffrey *et al.*, 1995; Morgan, 1997).





Figure 1.6 HsCDK2-HsCyclin A3 complex structure

(a) The HsCDK2-HsCyclin A3 heterodimer complex has a large interface that consists predominantly of hydrophobic contacts. This extensive interface reflects the relatively low (hence high affinity) dissociation constant (K_D) of about 50nm measured for this complex (Heitz et al., 1997). The major intersubunit contacts involve the PSTAIRE helix of HsCDK2 and α3

and α5 of the conserved cyclin box domain of HsCyclin A3. Another important binding interaction involves the C-terminal lobe of HsCDK2 and the N-terminal helix of HsCyclin A3. Since this helix is not conserved among cyclins, it may confer a degree of specificity on the HsCDK2-HsCyclin A3 complex. Structural comparison with free BvCyclin A3 (Brown et al., 1995) showed that the cyclin is unchanged upon complex formation suggesting that the cyclin is a rigid template on which the conformation of a CDK is modified to optimise the catalytic efficiency of the complex. This figure was produced with Aesop (Dr. Noble, unpublished work).

(b) Free HsCDK2 (grey) is superimposed on HsCyclin A3-bound HsCDK2 (cyan) to highlight the differences in the two structures. Both the PSTAIRE helix and the T-loop of free HsCDK2 are coloured green. That of the complexed HsCDK2 is shown in red (i.e. PSTAIRE) and yellow (T-loop), respectively. The realignment of Glu51, Lys33 and Asp145 are depicted in the close-up view of the PSTAIRE helix. Close up views of the T-loop in the absence and presence of HsCyclin A3 are depicted in space-filling representations. The residue phosphorylated by CAK (T160) is shown in red. Figure 1.6b was reproduced from (Pavletich, 1999).

The complex structure showed that cyclin binding induced conformational changes in HsCDK2. In particular, the PSTAIRE helix of HsCDK2 was moved into the catalytic cleft and rotated by about 90° (Jeffrey *et al.*, 1995; Pavletich, 1999). A glutamate residue (Glu 51 in HsCDK2) that is conserved in all eukaryotic protein kinases is situated on the PSTAIRE helix. In the absence of cyclin, Glu51 resides outside the catalytic cleft. Cyclin binding and the attendant changes in the PSTAIRE helix brings the side chain of Glu51 into the active site of HsCDK2 where in conjunction with a lysine residue (Lys33), an aspartate residue (Asp145) and a magnesium ion, it co-ordinates the phosphate atoms of the ATP and places them in the proper orientation for phosphotransfer. Another conformational change in HsCDK2 brought about by HsCyclin A3 binding occurs in the T-loop. This structure is modified and parts of it are moved by over 20Å to relieve the blockade to the catalytic cleft of HsCDK2 (Jeffrey *et al.*, 1995; Pavletich, 1999). Another effect of cyclin-induced changes in the T-loop is the exposure

to solvent of the phosphorylation site (T160). This sets the stage for upgrade of the catalytic status of the HsCDK2-HsCyclin A3 complex from partially to fully active after phosphorylation by CAK.

1.3.2 Activation by phosphorylation and dephosphorylation

Phosphorylation of T160 induces an additional 80-300 fold increase in the catalytic activity of the HsCDK2-HsCyclin A3 complex (Morgan, 1997). The T-loop undergoes additional conformational changes when a phosphate is transferred to the threonine residue (Pavletich, 1999; Russo *et al.*, 1996a).



Figure 1.7 Conformational changes induced by T160 phosphorylation *The unphosphorylated HsCDK2-HsCyclin A3 complex (grey) has been superimposed on the phosphorylated complex (HsCDK2 is coloured cyan and HsCyclin A3 is magenta). The T-loop of the unphosphorylated complex is in red whilst that of the phosphorylated complex is highlighted in yellow. The ATP is depicted in ball-and-stick representation and the phosphate group on T160 is shown as a yellow sphere. Reproduced from (Pavletich, 1999).*

The phosphate group on T160 is bound by three arginines (one each from the N-lobe, C-lobe and the T-loop). These arginines form hydrogen bonds to HsCDK2 and HsCyclin A3 residues resulting in the organisation of this region (which is part of the catalytic cleft) in preparation for the binding of the putative polypeptide substrate of a target protein. Thus, it seems that phosphorylation of T160 is an extension of the re-organisation of the active cleft of the HsCDK2 that was initiated by HsCyclin A3 binding (Pavletich, 1999; Russo *et al.*, 1996a).

The above conformational changes result in a fully activated kinase, which is now ready to phosphorylate substrates of which we know very little. There are yet other forms of regulation (of which our understanding is relatively clear) overlaid onto the CDK molecule.

1.3.3 Inhibition by both phosphorylation and dephosphorylation

In the yeasts, CDKs involved in cell cycle control can be inhibited by phosphorylation on Tyr15. In vertebrates, additional inhibitory phosphorylation occurs on Thr14. The X-ray structures described above provide an explanation as to how this adverse effect on CDK catalytic activity can be achieved. The side chains of these residues are close to the glycine-rich loop that forms the roof of the ATP phosphate-binding site (De Bondt *et al.*, 1993; Jeffrey *et al.*, 1995; Morgan, 1997). In the HsCDK2 structure, Thr14 and Tyr15 are buried beneath the T-loop and are not accessible to solvent but cyclin binding (as observed in the HsCDK2-HsCyclin A3 complex structure) exposes these residues and makes them solvent accessible (Morgan, 1997). This explains the biochemical observation that phosphorylation of Thr14 and Tyr15 is cyclin-dependent (Morgan, 1997). Perhaps, the inhibitory effect of phosphate groups on Thr14 and Tyr15 stems from the perturbation of the optimal alignment of the ATP substrate and the side chains of HsCDK2 catalytic residues (Morgan, 1997). The functional significance of inhibitory phosphorylation is to restrain the cell from a premature entry into mitosis (described under section 1.4.2). At the end of M-phase, the CDK in the mitotic kinase (initially by CDK1-cyclin A and then CDK1-cyclin B) is returned to the inactive state by dephosphorylation of T160 (T161 for SpCDK1) and degradation of the cyclin A/B

subunit. The removal of the phosphate group is catalysed by an enzyme called KAP for CDK-associated phosphatase (Hannon *et al.*, 1994; Poon and Hunter, 1995). Another form of inhibiting the kinase activity of partially or fully active CDK-cyclin complexes involves the binding of structurally disruptive protein molecules belonging to the cyclin-dependent kinase inhibitor (CKI) family.

1.3.4 Inhibition by the binding of inhibitory protein subunits

There are two groups of CKIs in vertebrates: the Cip and INK4 families (Nakayama, 1998). Together, they provide a tissue-specific mechanism by which cell cycle progression can be stopped in response to both extra- and intracellular signals (Harper *et al.*, 1993; Morgan, 1997). Members of the Cip family include p21, p27, and p57. They inhibit CDK2- and CDK4/6-cyclin complexes involved in G1 and G1/S control. The Ink4 family contains four members (p15, p16, p18, and p19) which act on only CDK4/6-cyclin D complexes (Morgan, 1997). So far, three CKIs have been identified in *S. cerevisiae:* Far1, acts on Cdc28-Cln complexes leading to cell cycle arrest in G1-phase in the presence of mating pheromone (Morgan, 1997; Peter and Herskowitz, 1994). Sic1 ensures that the cells enter S phase appropriately by inhibiting Cdc28-Clb5,6 complexes until they are required at the end of G1 (Mendenhall, 1993; Morgan, 1997). Pho81 inhibits Pho85-Pho80 kinase complexes in response to phosphate starvation (Lenburg and O'Shea, 1996; Morgan, 1997). In *S. pombe*, the Rum1 protein contributes to the G2→M-phase barrier by inhibiting the mitotic kinase (SpCDK1-SpCyclin B) complexes during G1 (Correa-Bordes and Nurse, 1995; Morgan, 1997).

Mechanistic insights into how the CKIs affect CDKs have come from structural studies on vertebrate CKIs (Russo *et al.*, 1996b). The activity of a fully active HsCDK2-HsCyclin A3 can be completely abolished by the binding of $p27^{Cip2}$.

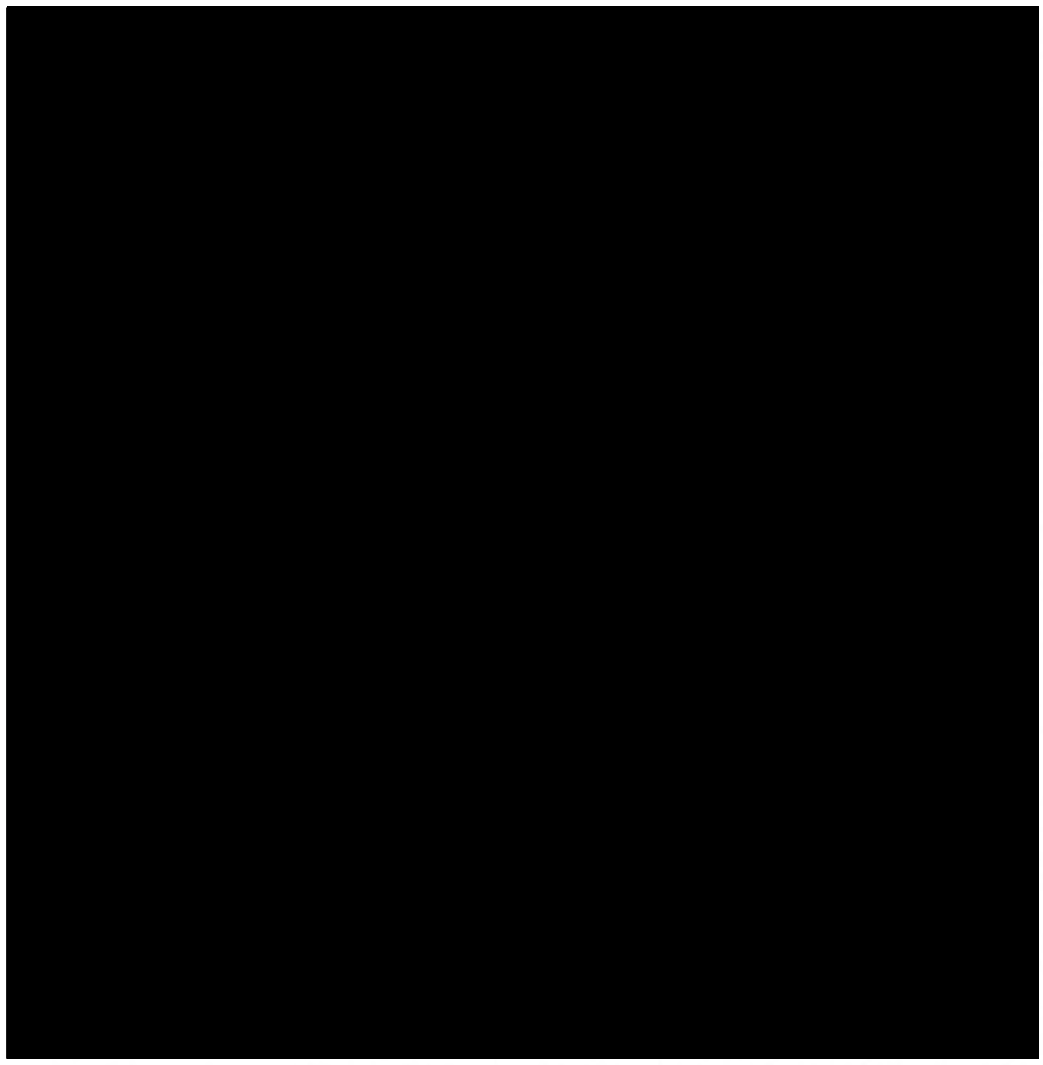


Figure 1.8 Phosphorylated HsCDK2-HsCyclin A3-p27^{Cip2} ternary complex *The p27 inhibitor protein (highlighted in red) embraces both HsCDK2 and HsCyclin A3 from the top of the kinase complex. The C-terminal region of p27 interacts with a binding pocket on HsCyclin A3 whereas the N-terminal region binds to the N-lobe of HsCDK2. The cyclin structure is unaffected but the CDK molecule undergoes drastic structural changes described below. The T-loop is highlighted in yellow and the side chains of Tyr15 and Thr160 are shown. Drawn by Julie Tucker with Aesop (Dr. Noble, unpublished work).*

The inhibition is achieved via two mechanisms: first, a small 3_{10}-helix of p27 is inserted into the catalytic cleft of HsCDK2. Structural comparison of HsCDK2-HsCyclin A3-p27 and the ATP-bound HsCDK2-HsCyclin A3 complexes shows that the inserted 3_{10}-helix is a positional mimic of the ATP substrate in this part of the HsCDK2 molecule. A highly conserved tyrosine residue of p27 mimics the van der Waals contacts made by the purine of ATP. This "foreign" tyrosine residue also retains the hydrogen bonds that are made by N1 and N6 groups of ATP. The backbone carbonyl groups of the C-terminal end of the p27 3_{10}-helix forms hydrogen bonds with the lysine residues at the active site of HsCDK2. Normally, these hydrogen bonds are made to the nucleotide phosphates of the ATP. Thus, the net effect of the insertion of 3_{10}-helix deep within the active site cleft of HsCDK2 is a direct occlusion of the ATP substrate (Morgan, 1996; Morgan, 1997; Pavletich, 1999; Russo *et al.*, 1996b).



Figure 1.9 Structural rearrangements induced by the binding of p27 to phosphorylated HsCDK2-HsCyclin A3

Superimposition of ATP and p27 bound HsCDK2-HsCyclin complexes. This illustration shows that with respect to its position and the contacts that are made, the conserved Tyr88 residue of p27 and the ATP purine are similar. Hydrogen bonds are denoted by dotted lines. Reproduced from (Pavletich, 1999).

It has been demonstrated that even in the absence of the disruptive 3_{10}-helix, p27 retains potent inhibitory ability (Pavletich, 1999; Polyak *et al.*, 1994). This occurs via the conformational changes p27 makes on binding to the N-terminal lobe of HsCDK2 (Morgan, 1996; Morgan, 1997; Pavletich, 1999; Russo *et al.*, 1996b). A p27 β-hairpin (fig. 1.10) forms hydrophobic contacts with the N-lobe of HsCDK2 that sits on top of the catalytic cleft. The N-lobe of an unbound HsCDK2 in the HsCDK2-HsCyclin A3

heterodimeric complex is normally curved (fig. 1.3) and it possesses a hydrophobic core. Upon binding, p27 removes β1 of HsCDK2 in order to gain entry into the hydrophobic core of the N-lobe. Next, a p27 β-strand is inserted to take the place of the displaced HsCDK2 β-strand. This results in the flattening of the N-lobe and the formation of an intermolecular β sandwich comprising of a new hydrophobic core (fig. 1.10). This dramatic structural re-arrangement has been described as a co-refolding of the HsCDK2 N-lobe and p27 into a new structure (Pavletich, 1999).



Figure 1.10 Reshaping of the HsCDK2 N-lobe by p27

The ATP-bound N-lobe (right) of HsCDK2 is compared with the p27-bound (left) HsCDK2 is coloured blue, HsCyclin A3 magenta and p27 yellow. Residues involved in both the intra- and intermolecular hydrophobic contacts are shown. The overall effect of p27 binding is the disruption of critical ATP binding elements such as the glycine rich loop. Reproduced from (Pavletich, 1999).

Members of the INK4 class of CDK inhibitors can associate with monomeric CDK4 *in vivo* (Serrano *et al.*, 1993). Also, *in vitro*, they can bind to active CDK4-cyclin D complexes and inhibit their kinase activity without dissociating the cyclin subunit (Guan *et al.*, 1996; Hirai *et al.*, 1995). This ambiguity made the understanding of their exact mechanism unclear until the recent determination of the HsCDK6-p19^{INK4a} and

HsCDK6-p16^{INK4a} complex structures (Brotherton *et al.*, 1998; Russo *et al.*, 1998). The structures provided an explanation for the above *in vitro* observations as follows: p16 binds to the opposite (with respect to cyclin binding) face of the CDK and interacts with the catalytic cleft of the CDK (fig. 1.11). Since the binding site of both the inhibitor and activator do not overlap, p16 can adversely affect the kinase activity of the CDK4-cyclin D complex without displacing the cyclin.

(a)



(b)

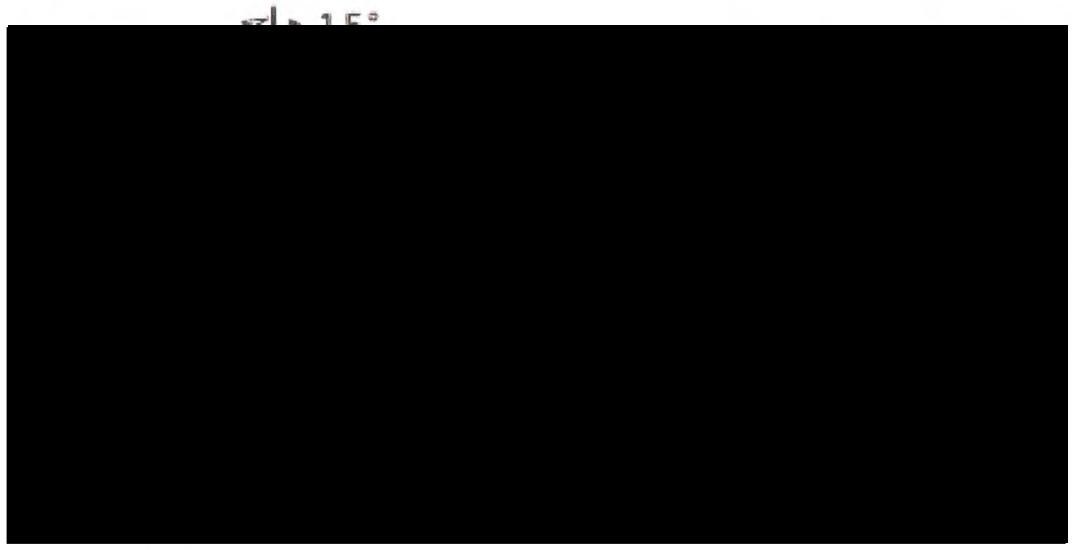


Figure 1.11 CDK6-p16^{INK4a} complex structure

(a) The p16-bound CDK6 and the cyclin A-bound CDK2 structures have been superimposed via the alignment of their respective C-lobes. The binding of the p16 molecule to the CDK leads to a twist in the N- and C-lobes that disrupts critical CDK-cyclin interactions. This is depicted schematically in (b). The two lobes of the CDK are rotated through a vertical axis by about 15° upon p16 binding. Reproduced from (Pavletich, 1999).

The p16 inhibitor achieves its inhibitory effect by two main mechanisms. The first method is similar to the action of p27 i.e. the perturbation of ATP binding site in order to disrupt the optimal alignment of ATP and active site residues necessary for phosphotransfer. The second method involves the initiation of allosteric changes in CDK4 that travels to and alters the CDK4-cyclin D interface so that key activating interactions between CDK4 and cyclin D are no longer possible. So, although the remaining interactions are enough to keep the cyclin attached to p16-CDK4, they are ineffective as far as CDK4 activation is concerned (fig. 1.11).

*1.3.5 **Structural modulation by Cks binding***

Potentially, the functional surface of a CDK can be extended by the binding of a protein subunit belonging to the cyclin-dependent kinase subunit (Cks) family of proteins. Genetic analysis in fission and budding yeasts indicates that Cks proteins are required for CDK function *in vivo*, but their precise role remains elusive (see section 1.6). In *S. pombe*, the Suc1 protein binds to CDK1-cyclin B complexes and it is necessary for governing mitosis, although the exact mechanism is unknown. Analysis of the crystal structure of HsCDK2-CksHs1 (fig. 1.12) and other biochemical evidence (Patra and Dunphy, 1998; Sudakin *et al.*, 1997) have revealed that these proteins may serve as adapters linking CDK-cyclin complexes with specific, cellular regulatory components like the anaphase promoting complex (APC) or proteosome.

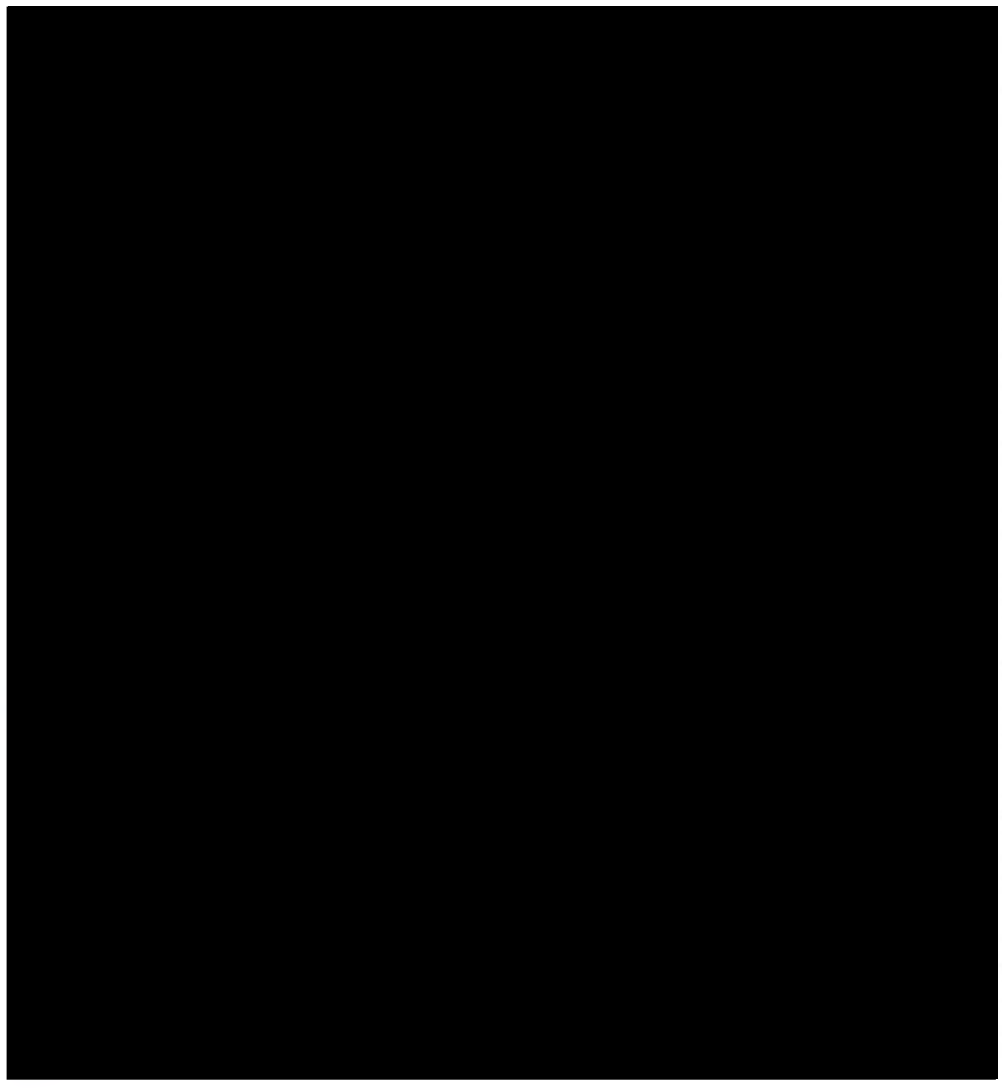


Figure 1.12 The structure of HsCDK2-CksHs1 complex

The side chain residues in the conserved phosphate-binding site (Lys11, Arg 20, Trp54 and Arg71) are highlighted. The conserved anion-binding site can be used to target the complex to phosphoproteins. The binding of CksHs1 does not affect the structure of the CDK. The figure was generated with Aesop (Dr. Noble, unpublished work).

The biochemical analysis of CDK-Cks interactions is described in chapter 4.

1.3.6 A summary (structural) of CDK regulation

The foregoing description has shown that CDK molecules possess an inherent structural malleability that allows their structural and hence functional properties to be modified by a diverse repertoire of regulatory molecules. The central importance of the cell cycle and the diverse pathways that impinge on it, may have directed the evolution of a class of molecules that possess intrinsic structural flexibility to oversee cell cycle progression in conjunction with other important cellular functions.



Figure 1.13 Mechanistic insights gained from structural analysis of monomeric CDK2 and various complexes

This array of CDK structures include inactive monomeric CDK2, inactive HsCDK2-CksHs1, partially active HsCDK2-HsCyclin A3, a fully active, phosphorylated HsCDK2-HsCyclin A3, an inhibited HsCDK2-HsCyclin A3-p27 ternary and non-activatable CDK6-p16INK4a complexes. The PSTAIRE helix is highlighted in red in both the monomeric and cyclin A bound HsCDK2 structures whereas their T loops are highlighted in yellow. In the HsCDK2-HsCyclin A3 complex, the phosphate group is represented by a yellow sphere and the T loop is coloured yellow. In the relevant structures, ATP is depicted in a ball-and-stick representation. Reproduced with modification from (Pavletich, 1999).

1.4 CHECKPOINT REGULATION

START is a point in G1 in which yeast cells decide whether to embark on a journey that begins with DNA replication and ends in a mitotic division or choose a route involving sexual conjugation and meiotic division (Cross, 1995; Woollard and Nurse, 1995). Whatever route is chosen is in response to environmental factors such as the availability of nutrients. The equivalent of START in mammalian cell cycles is called the restriction (R) point (Zetterberg *et al.*, 1995). It was observed that at a certain point in its life cycle, the growth of a normal mammalian cell depended on the presence of growth factors in the media (Pardee, 1974). This dependence occurred at a discrete time point in mid-to-late G1 phase. After the R-point is traversed, the cell commits to entering S-phase and the completion of a round of the cell cycle (Bartek *et al.*, 1996; Planas-Silva and Weinberg, 1997). Whereas cells in the pre-R phase are strictly dependent on the presence of serum in the media, the crossing of the R-point into post-R phase seemed to act as a molecular switch leading to the abolition of serum-dependency in the post-R period of G1, S, G2 and M-phases. The molecular basis of this mechanism is beginning to be understood.

1.4.1 G1 to S-phase

The retinoblastoma protein (pRb) functions as a negative regulator of cell growth and it is also a tumour suppressor. pRB, p107 and p130 belong to the pocket protein family. They share a highly conserved region – the pocket domain – through which they mediate protein-protein interactions with other molecules.



Figure 1.14 Schematic representation of human pRB, p107 and p130 proteins

*Conserved regions among all three proteins are highlighted in grey whilst those between only p107 and p130 are highlighted in black. The percentage degree of similarities between each pair and all there proteins are indicated. The highly conserved **A** and **B** pocket regions are used for mediating protein-protein interactions with other proteins. It has a high affinity for cellular targets containing the LXCXE motif where **L** is a leucine, **C** is a cysteine, E is glutamate and **X** represents any amino acid. Reproduced from (Mulligan and Jacks, 1998).*

The state of phosphorylation of pRB is critical to the regulation of the G1→S-phase transition (Mittnacht, 1998). When pRB is hypophosphorylated, it inhibits cell cycle progression by binding to and inhibiting transcription factors such as E2F, which are necessary for S-phase entry. When it is hyperphosphorylated, pRB indirectly permits cell cycle progression by releasing the bound transcription factors, including the E2F family. These transcription factors activate expression of genes necessary for the cell to traverse from G1 to S phase.

Phosphorylation of pRB is a multi-step process that is initiated by the CDK4/6-cyclin D kinase complexes (Planas-Silva and Weinberg, 1997; Sherr, 1996). This step can be blocked by the action of the INK4 class of inhibitors as described mechanistically under section 1.3.4). Phosphorylation is completed by the CDK2-cyclin E kinase complex, which is normally activated at about the time the cell crosses the R-point (Planas-Silva and Weinberg, 1997). A form of regulation whose exact mechanism is poorly understood is that the ability of the CDK2-cyclin E kinase to complete the phosphorylation of pRB depends on the prior partial phosphorylation of the common substrate by CDK4/6-cyclin D complexes (Hatakeyama *et al.*, 1994; Resnitzky and Reed, 1995) 16).



Figure 1.15 The G1→S-phase molecular switch *Important cellular decisions about whether the cell should proliferate or enter a state of quiescence are carried out late in G1 phase of the cell cycle (Lees and Weinberg, 1999). For most of this phase up to the pre-R point, the cell is responsive to extracellular signals from mitogenic growth factors and also by anti-proliferation signallers like transforming growth factor (TGF-β). These extracellular signals which are channelled through the cytoplasm (see fig. 1.2) into the nucleus where they are integrated into the core cell cycle apparatus results in the activation (by cyclin D) or inhibition (by $p16^{INK4a}$) of CDK4/6 molecules. This determines the state of phosphorylation of pRB which in turn determines whether the molecular switch is "on" (i.e. proliferate) or "off" (i.e. arrest). In effect, pRb is acting as a transcriptional repressor that controls gene expression by physically associating with transcription factors such as those of the E2F family. Reproduced from (Weinberg, 1996).*

The E2F family consists of five transcription factors. pRB preferentially binds to E2F-1, -2, and -3 whereas E2F-4 and -5 predominantly associate with p107 and p130. The E2Fs bind to the DP family of transcription factors and heterodimers of E2F and DP regulate the expression of genes necessary for G1/S-phase transition. E2F-1 plays an important role in the G1→S-phase transition: when it is bound to pRb, cell cycle progression is repressed. In the absence of pRb, E2F-1 stimulates cellular proliferation. The functional significance of the regulation of this checkpoint is reflected in the observation that the inactivation or deletion of pRB is found in many cancers, including retinoblastomas and carcinomas of the lung, breast, bladder, and prostate (Herwig and Strauss, 1997).

1.4.2 G2 to M-phase

The molecular details of this checkpoint are well documented in the literature (Dunphy, 1994; Lock and Wickramasinghe, 1994; Moreno *et al.*, 1989; Murray, 1993). In G2-phase just prior to the onset of mitosis, the mitotic kinase (CDK1-cyclin B complex) is held in an inactive state by phosphorylation at Tyr15 (in *S. pombe*) and also at Thr14 (in higher eukaryotes). The Wee1 enzyme is responsible for Tyr15 phosphorylation in most organisms but in vertebrates, the Myt1 kinase can phosphorylate both Thr14 and Tyr15 (Liu *et al.*, 1997; Lundgren *et al.*, 1991; Mueller *et al.*, 1995b). Entry into mitosis is triggered by dephosphorylation of both residues by Cdc25 phosphatase and subsequent activation of CDK1 (Murray, 1993).



Figure 1.16 Control of G2→M-phase

This scheme illustrates some of the positive feedback mechanisms employed to ensure that the cell enters and completes a round of mitosis only when it is desirable. Both Wee1 and Myt1 kinases and the Cdc25 phosphatase are all phosphorylated during mitosis. Phosphorylation of the kinases is inhibitory whilst that of the phosphatase is activatory. The potential for positive feedback has been shown by the observation that part of the mitotic phosphorylation on Wee1 and Cdc25 is catalysed by CDK1-cyclin B (Hoffmann et al., 1993; Mueller et al., 1995a). Other kinases like Plx, (a frog Polo-like kinase) also phosphorylates Cdc25 in the course of mitosis (Kumagai and Dunphy, 1996). An active Cdc25 can be inactivated by the action of protein phosphatase 2A (PP2A) (Millward et al., 1999). In S. pombe, the Nim1 protein kinase negatively regulates Wee1 by phosphorylating its C-terminal catalytic domain (Dunphy, 1994).

Execution of mitosis is dependent on an accurate completion of S-phase (Fisher and Nurse, 1996). In *S. pombe,* incomplete DNA replication prevents mitosis by retaining a leash on CDK1-cyclin B kinase activity: this is achieved by maintaining Tyr15 of CDK1 in a state of phosphorylation (Elledge, 1996; Lew and Kornbluth, 1996; Morgan,

1995; Morgan, 1997). It has also been observed that in several organisms, when the DNA material is damaged in either S phase or G2, initiation of mitosis is blocked by the stabilisation of the inhibitory phosphates on CDK1 (Jin *et al.*, 1996; Morgan, 1997; Rhind *et al.*, 1997; Ye *et al.*, 1997). This dependency of one important phase (mitosis) on another equally vital phase (DNA replication) acts in a concerted manner with other essential processes like DNA repair to ensure that the nuclear material bequeathed to each daughter cell is as error free as possible.

Exit from mitosis requires CDK1 inactivation by dephosphorylation of T161 and degradation of the cyclin subunit (Luca, 1993) by the ubiquitin-mediated anaphase promoting complex (APC) pathway (see section 1.5). Towards the end of mitosis, the mitotic kinase which started off as a CDK1-cyclin A complex to initiate mitosis and then a CDK1-cyclin B complex to execute mitosis exists as a ternary complex consisting of CDK1-cyclin B-Suc1 (in *S. pombe*) to end mitosis. The third component - SpSuc1 - is the subject of this thesis and it is described in more detail under section 1.6.

1.5 Regulated proteolysis in cell cycle control

The destruction of proteins that are no longer required for certain phases of the cell cycle is an integral part of cell cycle regulation (Hershko, 1997; Hoyt, 1997; Udvardy, 1996). It is well established that the covalent attachment of ubiquitin to certain target proteins acts as a flag for their degradation by the proteosome. For example, mitotic cyclins which are longer required at the end of M-phase are destroyed by the ubiquitin-mediated degradation pathway (Holloway *et al.*, 1993; Murray, 1995; Sudakin *et al.*, 1995). To date, the most detailed understanding of this process has come from studies on *S. cerevisiae.* Anaphase in *S. cerevisiae* is initiated by the destruction of Pds1 which is an inhibitor of sister chromosome separation (Cohen-Fix *et al.*, 1996; Michaelis *et al.*, 1997). Degradation of the CKI, Sic1 activates CDK1–Clb kinase which in turn initiates S-phase (Schwob *et al.*, 1994).Two distinct ubiquitination systems trigger these events: Pds1 and mitotic cyclins are degraded by the anaphase-promoting complex (APC) pathway (King *et al.*, 1996) whereas Sic1 is ubiquitinated by the Skp1–cullin–F-box (SCF) protein complex pathway (Deshaies, 1997; Feldman *et al.*, 1997; Krek, 1998; Patton *et al.*, 1998; Skowyra

et al., 1997). Homologues of several APC and SCF components have been identified in several organisms where they have been shown to function in ubiquitin-dependent degradation of proteins involved in cell cycle control and other biological processes (Krek, 1998).

1.5.1 The Anaphase Promoting Complex (APC) pathway

The APC or the cyclosome is a 1500KDa multiprotein complex consisting of about 12-13 polypeptide subunits. It is required for the metaphase-to-anaphase transition *in vivo* and also, high-rate cyclin-ubiquitination *in vitro* (Hershko and Ciechanover, 1998; Irniger *et al.*, 1995; King *et al.*, 1995; Sudakin *et al.*, 1995).

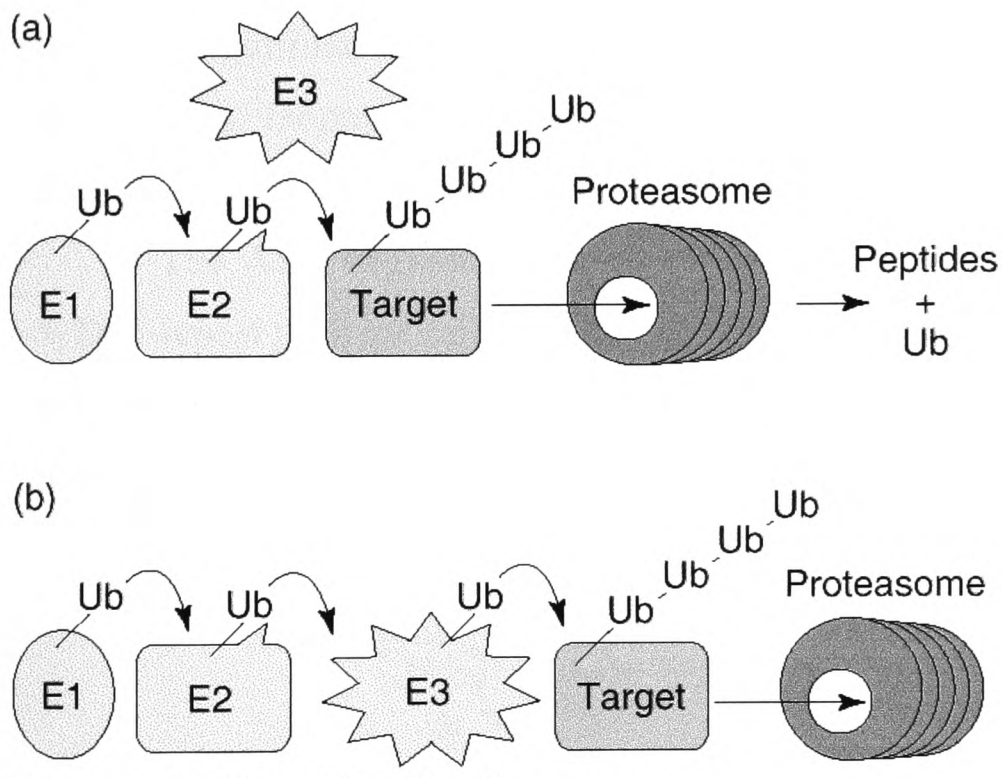


Figure 1.17 The ubiquitination process

The family of enzymes responsible for this pathway are the E1s, E2s, E3s and the isopeptidases (Hershko and Ciechanover, 1998). E1 activates ubiquitin by forming an E1-ubiquitin thiolester bond in the presence of ATP (Hershko and Ciechanover, 1998). The activated ubiquitin is then transferred to E2. The latter are a family of ubiquitin-conjugating enzymes. At this stage, some members of the E2 family can directly transfer the conjugated ubiquitin to a specific protein substrate (a) by catalysing the formation of an isopeptide bond between the carboxyl terminus of ubiquitin and the ε-amino group of

an internal lysine in the substrate. (b) other members of the E2 family depend on E3 (ubiquitin-ligases) to transfer the activated ubiquitin to a protein substrate. Substrate specificity of the ubiquitination reaction is conferred by E2s (Hershko and Ciechanover, 1998). Reproduced from (Townsley and Ruderman, 1998).

It appears that mono-ubiquitination is inadequate for degradation by the proteosome. Only poly-ubiquitinated substrates are degraded. In *S. pombe*, the E2-Ub conjugate is recognised by the anaphase-promoting complex (APC) which is an E3. Cyclin B is degraded by the proteosome after the APC has been activated by phosphorylation (Hershko and Ciechanover, 1998). The isopeptidases are a family of de-ubiquinating enzymes that can reverse the ubiquitination process (Hershko and Ciechanover, 1998). It is has been suggested that this provides fidelity of co-operativity in the ubiquitination process so that only selected protein targets are chosen for destruction. Isopeptidases may also serve to recycle ubiquitin at the end of the degradation process (Hershko and Ciechanover, 1998).

The APC degrades both anaphase inhibitors and mitotic cyclins. Two related WD40-repeat proteins called Cdc20 and Hct1/Cdh1 in *S. cerevisiae* and Fizzy and Fizzy-related in *D. melanogaster* are required for APC dependent proteolysis and they are important for substrate specificity (Kramer *et al.*, 1998).

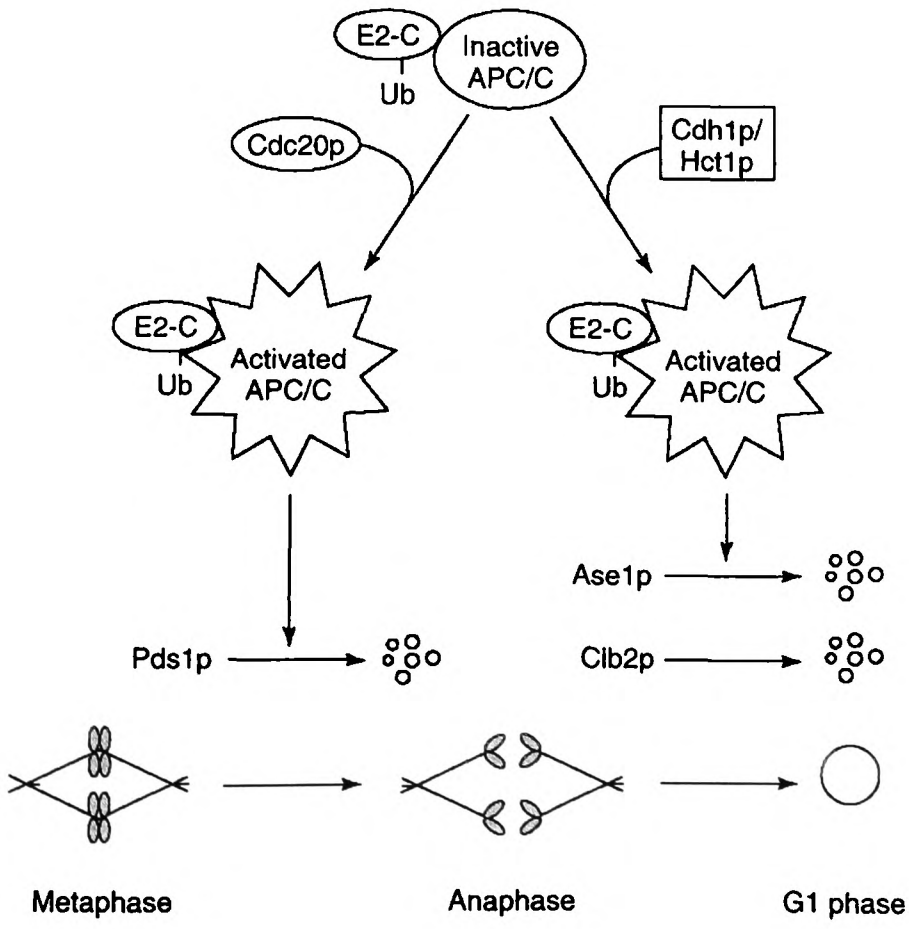


Figure 1.18 Substrate-specific activators of the APC

The Cdc20 and Cdh1/Hct1 proteins target the APC to different substrates during the cell cycle. Evidence from S. cerevisiae shows that Cdc20 targets substrates that are degraded at the metaphase-to-anaphase transition to the APC

whereas the Cdh1/Hct1 proteins target proteins that are degraded late in mitosis. Reproduced from (Townsley and Ruderman, 1998).

1.5.2 The Skp1-Cullin-F-box (SCF) pathway

In *S. cerevisiae*, SCF consists of Cdc53, Cdc34, Skp1 and Cdc4. Cdc53 is a cullin and Cdc34 is a ubiquitin-conjugating enzyme i.e. an E2 (Krek, 1998; Patton *et al.*, 1998). The SCF is an E3. The cullins can be defined as a family of proteins that are homologous to *S. cerevisiae* Cdc53 (Jackson, 1996). Cdc4 is a member of the growing family of proteins containing an F-box. The F-box is a conserved stretch of amino acids originally identified in cyclin F (Patton *et al.*, 1998). It has subsequently been demonstrated that other proteins like Cdc4 mediate protein-protein interactions with Skp1 via the F-box motif (Hershko and Ciechanover, 1998; Lisztwan *et al.*, 1998). Recent admissions into the family of cell cycle regulators are two S-phase kinase-associated proteins called $p19^{SKP1}$ and $p45^{SKP2}$ (Bai *et al.*, 1996; Connelly and Hieter, 1996; Zhang *et al.*, 1995). The exact function of the Skp proteins are not known but several lines of evidence point to a role in the recruitment of other proteins involved in the ubiquitination pathway. Skp2 is an F-box containing protein and it binds to Skp1 via this motif. The N- and C-terminal regions of Skp2 bind to CDK2-cyclin A complexes in the course of S-phase but the exact function of this quaternary complex is poorly understood (Yam *et al.*, 1999). In addition to Cdc4, yeast cells express several other F-box proteins including Grr1 (Hershko and Ciechanover, 1998; Krek, 1998; Patton *et al.*, 1998). SCF is necessary for the ubiquitination and subsequent degradation of Sic1, Far1 and Cdc6 (Drury *et al.*, 1997; Henchoz *et al.*, 1997).



Figure 1.19 Ubiquitination of specific substrates by SCF complexes

SCF is a ubiquitin ligase (an E3) which interacts with both a ubiquitin-conjugating enzyme and a substrate. SCF substrates need to be phosphorylated for subsequent ubiquitination and it is thought that phosphorylation of the substrate is part of the recognition signal which mediates binding to the SCF complex. Skp1-Cdc34-Cdc53 can form complexes with both (a) Cdc4 and (b) Grr1. It seems that the nature of the F-box employed defines substrate specificity for the SCF complex. Cdc4 is required for recruiting phosphorylated Sic1 for ubiquitination by the E2 enzyme Cdc34, whereas Grr1 is required for the degradation of the G1 cyclins Cln1 and Cln2. Grr1 is dispensable for the degradation of Sic1 and Far1 (Krek, 1998; Patton et al., 1998). Reproduced from (Townsley and Ruderman, 1998).

The interplay between cell cycle progression and regulated proteolysis by the two different ubiquitin pathways described above is summarised in figure 1.20.



Figure 1.20 Regulated proteolysis and cell control in S. cerevisiae

To cross START in G1 requires the activation of CDK1 by Clns 1-3 (G1 cyclins) and Clbs 1-6 (mitotic cyclins). In order to relieve the restraint on the S-phase kinase at START, CDK1-Clns1/2 phosphorylates Sic1 protein, which is subsequently ubiquinated and degraded by the SCF^{Cdc4} complex. Autophosphorylation targets CDK1-Clns1,2 complexes for the same fate via the SCF^{Grr1} complex. The elimination of Sic1 and Clns1,2 frees the CDK1 protein to bind to Clbs5/6 for the initiation and completion of DNA replication. The CDK1-Clbs2/3/4 kinase complexes initiate entry into mitosis. The APC degrades the Pds1 protein, the Clbs and other proteins to bring about a metaphase-to-anaphase transition. The SCF pathways (purple) are constitutively active during cell cycle progression but APC activity (green) is limited to anaphase till the end of G1 phase. Red and black arrows indicate phosphorylation and activation events, respectively and lines ending with bars indicate inhibition. Cdc28 is the S. cerevisiae CDK1

homologue and the letter ***p*** *at the end of the name of a gene refers to the gene product. Reproduced from (Patton et al., 1998).*

1.6 The Cyclin-dependent kinase subunits (Cks) proteins

1.6.1 Biological importance of the Cks proteins

The *SpSuc1*$^+$ gene was initially identified in *S. pombe* as a seppresor of certain CDK1 temperature-sensitive mutations (Hayles *et al.*, 1986). Homologues have been identified from several organisms. They share a high degree (~60-80%) of primary amino acid sequence conservation (Pines, 1996; Vogel and Baratte, 1996). The gene product of *SpSuc1*$^+$ binds to CDK1: in fact, SpSuc1 covalently coupled to sepharose beads is sold commercially as a reagent for fishing out CDK1 from complex mixtures (Pines, 1996; Vogel and Baratte, 1996). Also, biochemical studies have detected ternary complexes containing SpSuc1, a CDK1 and a cyclin B (Booher *et al.*, 1989). It has been more than a decade since the initial discovery of the *SpSuc1*$^+$ gene, yet an exact account of the physiological role remains an elusive goal. This is because of the lack of a single functional model that can accommodate all the conflicting evidence from genetic and biochemical studies.

SpSuc1 can rescue some temperature-sensitive mutants of *S. pombe* CDK1 *in vitro* (Moreno *et al.*, 1989) This is the basis of a widely used functional test in cell cycle research known as the histone kinase assay (see chapter 4, section 4.2.18c). Deletion of the SpSuc1/CKS1 gene from yeast leads to a loss of cell viability (Dunphy and Newport, 1989). It has been observed that the cells arrest in late mitosis showing mitotic spindles of varying lengths and condensed chromosomes. This state of arrested development is also characterised by the presence of high levels of mitotic cyclins and concomitant high CDK1 activity suggesting that the inability to destroy these mitotic cyclins (a pre-requisite for mitotic exit) is the cause of the cessation of the cell cycle in late mitosis. Overexpression of SpSuc1/CKS1 in yeast results in a delayed entry into mitosis (Hayles *et al.*, 1986; Hindley *et al.*, 1987; Richardson *et al.*, 1990). Paradoxically, this observation (i.e. delayed entry into mitosis) is seen in *Xenopus* interphase egg extracts not because of overexpression but as a

result of the immunodepletion of Xe-p9 (Patra and Dunphy, 1996). *In vitro*, the addition of excess SpSuc1 or Xe-p9 prevents CDC25 phosphatase from removing the inhibitory phosphate on Tyr-15 of CDK1 (Dunphy and Newport, 1989; Gautier *et al.*, 1991; Patra and Dunphy, 1996). A summary of the effect of CKS proteins on different organisms is presented below. In the next section, we shall look at the contribution of structural biology to our understanding of the structure and function of the CKS family of proteins.

[redacted]

Table 1.1 Effect of the overproduction or deletion of Cks proteins in different organisms

Reproduced from (Pines, 1996).

1.6.2 A wealth of X-ray structures

The Cks family of proteins are ideal targets for structural elucidation by X-ray crystallography because they are readily overexpressed in recombinant systems in a soluble form and also, they are relatively easy to purify and crystallise. To date, six different structures of Cks proteins have been determined by this method (Arvai *et al.*, 1995; Bourne *et al.*, 1995; Bourne *et al.*, 1996; Endicott *et al.*, 1995; Khazanovich *et al.*, 1996; Parge *et al.*, 1993). A common feature of these atomic structures is that Cks structures exist in two oligomeric states: a single, monomeric fold and a ß-strand exchanged dimer (fig. 1.21).

(a)



(b)



Figure 1.21 The dual conformational state of SpSuc1

(a) SpSuc1 β-hairpin single domain fold (b) β-interchanged SpSuc1 dimer fold. The two subunits are highlighted in red and yellow. The side chains forming the β-hinge region (His88, Val89, Pro90, Glu91, Pro92 and His93) are depicted in a ball-and-stick representation. Figure generated with Aesop (Dr. Noble, unpublished work).

The atomic structure of CksHs2 was the first to be resolved (Parge *et al.*, 1993). This showed that it formed a hexamer composed of three ß-strand-exchanged dimers. It was suggested that CksHs2 might act as a hub for the multimerisation of Cyclin-CDK1 complexes *in vivo* (Parge *et al.*, 1993) So far, only the monomeric and dimeric (but not the hexameric) forms have been observed in solution (Birck *et al.*, 1995), suggesting that the hexamer may have arisen as a result of the conditions (50mM sodium citrate at a non-physiological pH of 5.5) used to crystallise CksHs2 protein. In contrast, the structure of human CksHs1 showed that it was a monomer consisting of four central β-sheets bordered by two α-helices (Arvai *et al.*, 1995). A comparison of the CksHs1 and CksHs2 structures reveals a stretch of conserved residues (Glu 61-His65) that forms a novel β-hinge region. How this particular region folds determines the global fold of the polypeptide. If the β-hinge region folds back on itself, we get a single monomeric structure as seen in the CksHs1 but if this region unfolds to form an extended β-strand that can be exchanged between two CksHs2 molecules, we see a strand-exchanged molecule as in the CksHs2 structure. This structural theme is repeated in the atomic structures of SpSuc1. It has been shown that SpSuc1 is capable of adopting two distinct conformations: a zinc-mediated dimer and a β-interchanged dimer. Each domain in the β-interchanged dimer consists of three α-helices and a four-stranded β-sheet with the last β-strand being exchanged between two subunits to form an intimate intersubunit β-strand-exchanged dimer. The zinc-mediated dimer consists of four α-helices and a four-stranded β-sheet. The zinc-mediated dimer does not oligomerise to form a hexamer as seen in the CksHs2 structure. As yet, there is no evidence that zinc ions are required for the *in vivo* function of SpSuc1. A comparison of the two structures reveals a novel β-hinge motif consisting of the conserved HVPEPH (His-88 to His 93) sequence. It is this β-hinge that determines the conformation and assembly of SpSuc1. The opening of the β-hinge produces the intersubunit β-strand interchanged dimer while its closure produces an intrasubunit β-hairpin and single-domain fold.

Attempts to reconcile the above structural data with biochemical evidence in a quest to arrive at a better understanding of the structure-functional relationships of CKS proteins is yet to produce a cogent functional model for these class of CDK regulators.

One proposal is that CKS proteins may belong to a class of proteins known as the domain-swappers (Bennett *et al.*, 1994) These are a family of functionally variant proteins that form dimers by either exchanging secondary structure elements; for example, bovine seminal ribonuclease dimerises by exchanging N-terminal α-helices (Piccoli *et al.*, 1992) or sometimes whole domains as in the case of diptheria toxin (Bennett *et al.*, 1994). It is yet to be demonstrated whether the monomeric or dimeric forms of a domain-swapper perform different biological functions or not. For the Cks proteins, it has been gleaned from the available atomic structures that the single monomeric fold possesses unique surface features (like an extensive hydrophobic patch and an anion-binding site) that are unavailable in the β-exchanged dimer, (Morgan, 1996). This has led to the speculation that the ability to exchange β-strands and thereby form two distinct molecules may have evolved to extend the functional utility of the Cks and other domain-swapping proteins (Schlunegger *et al.*, 1997).

1.6.3 Modulation of CDK2 structure by the binding of Cks1Hs2

The structure of the CDK2-Cks1 complex revealed that a single Cks protein binds the lower part of the kinase C-terminal lobe, without affecting the active site cleft or cyclin-binding site (Bourne *et al.*, 1996). Thus, as suggested by previous work, it is unlikely that the CksHs1 protein directly affects kinase activation or substrate binding. On the other hand, the Cks protein is in a position where it could serve as an extension of the substrate-binding surface or could offer a recognition surface for regulatory proteins controlling CDK function.

1.6.4 Functional models of Cks function

The earliest model for Cks function suggested that these small proteins inhibited CDK activity by blocking the Cdc25 phosphatase dephosphorylation of Tyr15 (Vogel and Baratte, 1996). The elucidation of the structure of the first Cks protein led to the suggestion that they may act as a hub for the multimerisation of CDKs (Parge *et al.*, 1993). However, the structure of the CDK2-CksHs1 complex showed that those models

were implausible because the Cks binding site on CDKs was at the C-terminal lobe which is far away from Tyr15 (situated near the active site). The complex structure also showed that only monomeric not dimeric or hexameric Cks molecules interact with a CDK. The functional model that is gaining currency among cell cycle researchers is the so-called targeting model. This model suggests that members of the Cks family may be adapter proteins targeting CDK complexes to specific cellular substrates. The evidence for this model comes from the demonstration that the active, phosphorylated form of the APC from clam oocytes binds to SpSuc1-sepharose beads whilst the interphase form (unphosphorylated) does not. Since the bound APC was specifically eluted by phosphate containing compounds, such as p-nitrophenol phosphate, it was suggested that the APC bound the phosphate-binding site of SpSuc1 (Sudakin *et al.*, 1997). The ability of SpSuc1 to bind both SpCDK1 and phosphorylated APC was used to explain the lag kinetics of APC activation by the SpCDK1-SpCyclin B kinase as follows: activation of the APC requires multiple phosphorylations which can be divided into an initial, slow phosphorylation event and a later, fast phosphorylation event to complete the process. The initial phosphorylation event is probably carried out by the CDK1-cyclin B kinase and this results in a tighter association of CDK1-cyclin B-Suc1 to the partially activated APC to accelerate the activation process. The lag kinetics can be reconstituted *in vitro* and it is probably important to prevent premature self-inactivation of the mitotic kinase towards the end of mitosis (Hershko, 1997; Patra and Dunphy, 1998; Sudakin *et al.*, 1997).

1.6.5 A summary of cell cycle control

The foregoing account suggests that cell cycle transitions are in essence, the regulated activation and deactivation of CDKs. The inherent structural flexibility of a CDK allows it to be modulated (structurally and functionally) by several distinct families of cell cycle proteins. A universal feature of this important cellular function is that the molecular mechanisms that trigger the onset of cell cycle events possess all-or-nothing switch characteristics and progress through each phase is irreversible (Morgan, 1997).

The molecular basis for cell cycle control and its interrelationship with other cellular function like apoptosis is summarized in the figure below.



Figure 1.22 Summary of cell cycle regulation

A = growth-promoting signals issued by neighbouring cells

B = growth-inhibitory signals issued by neighbouring cells

C = Transforming growth factor beta (an inhibitor)

D = Active pRB protein complexed with inactive transcription factors (master brake)

E = DNA damage or oxygen deprivation. Reproduced from (Weinberg, 1996).

1.6.6 A preview of the thesis

The main thrust of the work presented here has been concerned with investigation into the function of SpSuc1 using the X-ray structure as a basis for probing the biochemical properties of this molecule. Chapter 2 is a description of the structure of SpSuc1 using a relatively higher resolution (compared to the published model) X-ray data set for the structural analysis. Another source of complexity in the quest to offer an accurate description of the physiological function of Cks proteins has been the

observation that these molecules exhibit a duality of conformational state (i.e. monomers and dimers). The folding pathway of SpSuc1 is currently under investigation by Dr. Itzhaki of the University of Cambridge. To this end, Dr. Itzhaki's group have produced several folding mutants of SpSuc1 for biophysical analysis. Chapter 3 contains an account of the elucidation of the structure of two such mutants. After these detailed structural examinations in (Endicott *et al.*, 1995) and chapters 2 and 3, certain notable features were noted and several point mutants were generated to find the functional significance of these molecular properties of SpSuc1. An account of the generation of the mutants and their characterisation (biochemical and biophysical) is presented in chapter 4. Chapter 5 describes an attempt at using protein engineering techniques to produce a fragment of SpCyclin B for structural studies, an account of the construction of an *E. coli* expression vector for dual plasmid co-expression studies, *in vivo* strategies for heterologous protein overexpression in *E. coli*, *in vitro* strategies for refolding a protein encapsulated in inclusion bodies and finally, a computer-assisted generation of a homology model of SpCyclin B in an attempt to deduce a rationale for the persistent aggregation property of this recombinant protein. Chapter 6 contains a brief synthesis of the work described in this thesis plus several recent publications in an attempt to provide a plausible functional model for SpSuc1. A simple account of the theory of X-ray crystallography is provided in the appendix.

1.7 LITERATURE CITED

Ames, B.N., Gold, L.S. and Willett, W.C. (1995) The causes and prevention of cancer. *Proc Natl Acad Sci U S A*, **92**, 5258-5265.

Anderson, G., Busso, D., Poterszman, A., Hwang, J.R., Wurtz, J.-M., Ripp, R., Thierry, J.-C., Egly, J.-M. and Moras, D. (1997) The structure of cyclin H: common mode of kinase activation and specific features. *The EMBO Journal*, **16**, 958-967.

Arvai, A.S., Bourne, Y., Hickey, M.J. and Tainer, J.A. (1995) Crystal structure of the human cell cycle protein CksHs1: single domain fold with similarity to kinase N-lobe domain. *J Mol Biol*, **249**, 835-842.

Bai, C., Richman, R. and Elledge, S.J. (1994) Human cyclin F. *EMBO J.*, **13**, 6087-6098.

Bai, C., Sen, P., Hofman, K., Ma, L., Goebl, M., Harper, J.W. and Elledge, S.J. (1996) SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, F-box. *Cell*, **86**, 263-274.

Bartek, J., Bartkova, J. and Lukas, J. (1996) The retinoblastoma protein pathway and the restriction point. *Curr Opin Cell Biol*, **8**, 805-814.

Bennett, M.J., Choe, S. and Eisenberg, D. (1994) Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA*, **91**, 3127-3131.

Birck, C., Raynaud-Messina, B. and Samama, J.P. (1995) Oligomerization state in solution of the cell cycle regulators p13suc1 from the fission yeast and p9cksphy from the myxomycete Physarum, two members of the cks family. *FEBS Lett*, **363**, 145-150.

Booher, R.N., Alfa, C.E., Hyams, J.S. and Beach, D.H. (1989) The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localization. *Cell*, **58**, 485-497.

Bourne, Y., Arvai, A.S., Bernstein, S.L., Watson, M.H., Reed, S.I., Endicott, J.E., Noble, M.E., Johnson, L.N. and Tainer, J.A. (1995) Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci U S A*, **92**, 10232-10236.

Bourne, Y., Watson, M.H., Hickey, M.J., Holmes, W., Rocque, W., Reed, S.I. and Tainer, J.A. (1996) Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle regulatory protein CksHs1. *Cell*, **84**, 863-874.

Brotherton, D.H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P.J., Volyanik, E., Xu, X., Parisini, E., Smith, B.O., Archer, S.J., Serrano, M., Brenner, S.L., Blundell, T.L. and Laue, E.D. (1998) Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d [published erratum appears in Nature 1998 Nov 26;396(6709):390]. *Nature*, **395**, 244-250.

Brown, N.R., Noble, M.E.M., Endicott, J.A., Garman, E.F., Wakatsuki, S., Mitchell, E.P., Rasmussen, B., Hunt, T. and Johnson, L.N. (1995) The crystal structure of cyclin A. *Structure*, **3**, 1235-1247.

Bueno, A., Richardson, H., Reed, S.I. and Russell, P. (1991) A fission yeast B-type cyclin functioning early in the cell cycle [published erratum appears in Cell 1993 Jun 4;73(5):following 1049]. *Cell*, **66**, 149-159.

Cardoso, M.C., Leonhardt, H. and Nadal-Ginard, B. (1993) Reversal of terminal differentiation and control of DNA replication: cyclin A and Cdk2 specifically localize at subnuclear sites of DNA replication. *Cell*, **74**, 979-992.

Cohen-Fix, O., Peters, J.M., Kirschner, M.W. and Koshland, D. (1996) Anaphase initiation in Saccharomyces cerevisiae is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p. *Genes Dev*, **10**, 3081-3093.

Collins, K., Jacks, T. and Pavletich, N.P. (1997) The cell cycle and cancer. *Proc Natl Acad Sci U S A*, **94**, 2776-2778.

Connelly, C. and Hieter, P. (1996) Budding yeast SKP1 encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell*, **86**, 275-285.

Correa-Bordes, J. and Nurse, P. (1995) p25rum1 orders S phase and mitosis by acting as an inhibitor of the p34cdc2 mitotic kinase. *Cell*, **83**, 1001-1009.

Cross, F.R. (1995) Starting the cell cycle: what's the point? *Curr Opin Cell Biol*, **7**, 790-797.

De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.-H. (1993) Crystal structure of cyclin-dependent kinase 2. *Nature*, **363**, 595-602.

Deshaies, R.J. (1997) Phosphorylation and proteolysis: partners in the regulation of cell division in budding yeast [published erratum appears in Curr Opin Genet Dev 1997 Jun;7(3):424]. *Curr Opin Genet Dev*, **7**, 7-16.

Devault, A., Martinez, A.-M., Fesquet, D., Labbe, J.-C., Morin, N., Tassan, J.-P., Nigg, E.A., Cavadore, J.-C. and Doree, M. (1995) MAT1 (menage a trois) a new RING finger protein subunit stabilising cyclin H-CDK7 complexes in starfish and xenopus CAK. *EMBO J.*, **14**, 5027-5036.

Donnellan, R. and Chetty, R. (1999) Cyclin E in human cancers. *Faseb J*, **13**, 773-780.

Draetta, G., Piwnica-Worms, H., Morrison, D., Druker, B., Roberts, T. and Beach, D. (1988) Human cdc2 protein kinase is a major cell-cycle regulated tyrosine kinase substrate. *Nature*, **336**, 738-744.

Draetta, G.F. (1997) Cell cycle: will the real Cdk-activating kinase please stand up. *Curr Biol*, **7**, R50-52.

Drury, L.S., Perkins, G. and Diffley, J.F. (1997) The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *Embo J*, **16**, 5966-5976.

Dunphy, W.G. (1994) The decision to enter mitosis. *Trends in cell biology*, **4**, 202-207.

Dunphy, W.G. and Newport, J.W. (1989) Fission yeast p13 blocks mitotic activation and tyrosine dephosphorylation of the Xenopus cdc2 protein kinase. *Cell*, **58**, 181-191.

Dynlacht, B.D. (1997) Regulation of transcription by proteins that control the cell cycle. *Nature*, **389**, 149-152.

Edwards, M.C., Wong, C. and Elledge, S.J. (1998) Human cyclin K, a novel RNA polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity. *Mol Cell Biol*, **18**, 4291-4300.

Elledge, S.J. (1996) Cell cycle checkpoints: Preventing an identity crisis. *Science*, **274**, 1664-1672.

Endicott, J.A., Noble, M.E., Garman, E.F., Brown, N., Rasmussen, B., Nurse, P. and Johnson, L.N. (1995) The crystal structure of $p13^{suc1}$, a p34cdc2-interacting cell cycle control protein. *EMBO J.*, **14**, 1004-1014.

Feldman, R.M., Correll, C.C., Kaplan, K.B. and Deshaies, R.J. (1997) A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p [see comments]. *Cell*, **91**, 221-230.

Fisher, D.L. and Nurse, P. (1996) A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S- phase and mitosis in the absence of G1 cyclins. *Embo J*, **15**, 850-860.

Fisher, R.P., Jin, P., Chamberlin, H.M. and Morgan, D.O. (1995) Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. *Cell*, **83**, 47-57.

Gautier, J., Solomon, M.J., Booher, R.N., Bazan, J.F. and Kirschner, M.W. (1991) cdc25 is a Specific Tyrosine Phosphatase That Directly Activates $p34^{cdc2}$. *Cell*, **67**, 197-211.

Glotzer, M., Murray, A.W. and Kirschner, M.W. (1991) Cyclin is degraded by the ubiquitin pathway. *Nature*, **349**, 132-138.

Gorman, C. (1998) Curing cancer: the hope and the hype. *TIME*, Vol. 151, pp. 44-54.

Guan, K.L., Jenkins, C.W., Li, Y., O'Keefe, C.L., Noh, S., Wu, X., Zariwala, M., Matera, A.G. and Xiong, Y. (1996) Isolation and characterization of p19INK4d, a p16-related inhibitor specific to CDK6 and CDK4. *Mol Biol Cell*, **7**, 57-70.

Hannon, G.J., Casso, D. and Beach, D. (1994) KAP: A dual specificity phosphatase that interacts with cyclin-dependent kinases. *Proc. Natl. Acad. Sci. (U.S.A.)*, **91**, 1731-1735.

Hardie, G. and Hanks, S. (eds.) (1995) *The protein kinase facts book.* Academic Press, London.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K. and Elledge, S.J. (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin- dependent kinases. *Cell*, **75**, 805-816.

Hatakeyama, M., Brill, J.A., Fink, G.R. and Weinberg, R.A. (1994) Collaboration of G1 cyclins in the functional inactivation of the retinoblastoma protein. *Genes Dev*, **8**, 1759-1771.

Hayles, J., Beach, D., Durkaz, B. and Nurse, P. (1986) The fission yeast cell cycle control gene cdc2: isolation of a sequence suc1 that suppresses cdc2 mutant function. *Mol. Gen. Genet.*, **202**, 291-293.

Heitz, F., Morris, M.C., Fesquet, D., Cavadore, J.C., Doree, M. and Divita, G. (1997) Interactions of cyclins with cyclin dependent kinases: a common interactive mechanism. *Biochemistry*, **36**, 4995-5003.

Henchoz, S., Chi, Y., Catarin, B., Herskowitz, I., Deshaies, R.J. and Peter, M. (1997) Phosphorylation- and ubiquitin-dependent degradation of the cyclin- dependent kinase inhibitor Far1p in budding yeast. *Genes Dev*, **11**, 3046-3060.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S. and Young, R.A. (1998) Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell*, **2**, 43-53.

Hershko, A. (1997) Roles of ubiquitin-mediated proteolysis in cell cycle control. *Curr Opin Cell Biol*, **9**, 788-799.

Hershko, A. and Ciechanover, A. (1998) The ubiquitin system. *Annu Rev Biochem*, **67**, 425-479.

Herwig, S. and Strauss, M. (1997) The retinoblastoma protein: a master regulator of cell cycle, differentiation and apoptosis. *Eur J Biochem*, **246**, 581-601.

Hindley, J., Phear, G., Stein, M. and Beach, D. (1987) Suc1+ encodes a predicted 13 kilodalton protein that is essential for cell viability and directly involved in the division cycle of Schizosaccharomyces pombe. *Mol. Cell Biol.*, **7**, 504-511.

Hirai, H., Roussel, M.F., Kato, J.Y., Ashmun, R.A. and Sherr, C.J. (1995) Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6. *Mol Cell Biol*, **15**, 2672-2681.

Hoffmann, I., Clarke, P.R., Marcote, M.J., Karsenti, E. and Draetta, G. (1993) Phosphorylation and activation of human cdc25-c by cdc2 cyclin-b and its involvement in the self-amplification of mpf at mitosis. *Embo J*, **12**, 53-63.

Hofmann, F. and Livingston, D.M. (1996) Differential effects of cdk2 and cdk3 on the control of pRb and E2F function during G1 exit. *Genes Dev*, **10**, 851-861.

Holloway, S.L. (1996) Getting control of the cell cycle literature. *Cell*, **87**, 597-600.

Holloway, S.L., Glotzer, M., King, R.W. and Murray, A.W. (1993) Anaphase is initiated by proteolysis rather than by inactivation of maturation promoting factor. *Cell*, **73**, 1393-1402.

Hoyt, M.A. (1997) Eliminating all obstacles: regulated proteolysis in the eukaryotic cell cycle [comment]. *Cell*, **91**, 149-151.

Hutchison, C. and Glover, D.M. (eds.) (1995) *Cell Cycle Control*. IRL Press, Oxford.

Irniger, S., Piatti, S., Michaelis, C. and Nasmyth, K. (1995) Genes involved in sister chromatid separation are needed for B-type cyclin proteolysis in budding yeast [published erratum appears in Cell 1998 May 1;93(3):487]. *Cell*, **81**, 269-278.

Jacks, T. and Weinberg, R.A. (1998) The expanding role of cell cycle regulators [comment]. *Science*, **280**, 1035-1036.

Jackson, P.K. (1996) Cull and destroy. *Current biology*, **6**, 1209-1212.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J. and Pavletich, N.P. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, **376**, 313-320.

Jin, P., Gu, Y. and Morgan, D.O. (1996) Role of inhibitory CDC2 phosphorylation in radiation-induced G2 arrest in human cells. *J Cell Biol*, **134**, 963-970.

Johnson, L.N., Lowe, E.D., Noble, M.E. and Owen, D.J. (1998) The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett*, **430**, 1-11.

Johnson, L.N., Noble, M.E.M. and Owen, D.J. (1996) Active and inactive protein kinases. *Cell*, **85**, 149-158.

Kamb, A. (1995) Cell-cycle regulators and cancer. *Trends Genets*, **11**, 136-140.

Kanaoka, Y., Kimura, S.H., Okazaki, I., Ikeda, M. and Nojima, H. (1997) GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. *FEBS Lett*, **402**, 73-80.

Kato, J.Y., Matsuoka, M., Strom, D.K. and Sherr, C.J. (1994) Regulation of cyclin D-dependent kinase 4 (cdk4) by cdk4-activating kinase. *Mol Cell Biol*, **14**, 2713-2721.

Khazanovich, N., Bateman, K., Chernaia, M., Michalak, M. and James, M. (1996) Crystal structure of the yeast cell-cycle control protein, p13suc1, in a strand-exchanged dimer. *Structure*, **4**, 299-309.

Kim, K.K., Chamberlin, H.M., Morgan, D.O. and Kim, S. (1996) Three dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase. *Nat. Struct. Biol.*, **3**, 849-855.

King, R.W., Deshaies, R.J., Peters, J.-M. and Kirschner, M.W. (1996) How proteolysis drives the cell cycle. *Science*, **274**, 1652-1659.

King, R.W., Peters, J.M., Tugendreich, S., Rolfe, M., Hieter, P. and Kirschner, M.W. (1995) A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. *Cell*, **81**, 279-288.

Knighton, D.R., Zheng, J., Ten Eyck, L.F., Ashford, V.A., Xuong, N.-H., Taylor, S.S. and Sowadski, J.M. (1991) Crystal structure of the catalytic subunit of cyclic adenosinemonophosphate-dependent protein kinase. *Science*, **253**, 407-413.

Kobayashi, H., Stewart, E., Poon, R., Adamczewski, J.P., Gannon, J. and Hunt, T. (1992) Identification of the domains in cyclin A required for binding to , and activation of, p34cdc2 and p32cdk2 protein kinase subunits. *Mol. Biol. Cell.*, **3**, 1279-1294.

Kramer, E.R., Gieffers, C., Holzl, G., Hengstschlager, M. and Peters, J.M. (1998) Activation of the human anaphase-promoting complex by proteins of the CDC20/Fizzy family. *Curr Biol*, **8**, 1207-1210.

Krek, W. (1998) Proteolysis and the G_1-S transition: the SCF connection. *Curr. Opin. Genets Develop.*, **8**, 36-42.

Kumagai, A. and Dunphy, W.G. (1996) Purification and molecular cloning of Plx1, a Cdc25-regulatory kinase from Xenopus egg extracts. *Science*, **273**, 1377-1380.

Larochelle, S., Pandur, J., Fisher, R.P., Salz, H.K. and Suter, B. (1998) Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. *Genes Dev*, **12**, 370-381.

Leclerc, V., Tassan, J.P., O'Farrell, P.H., Nigg, E.A. and Leopold, P. (1996) Drosophila Cdk8, a kinase partner of cyclin C that interacts with the large subunit of RNA polymerase II. *Mol Biol Cell*, **7**, 505-513.

Lees, E.M. and Harlow, E. (1993) Sequences within the conserved cyclin box of human cyclin A are sufficient for binding to and activation of cdc2 kinase. *Mol Cell Biol*, **13**, 1194-1201.

Lees, J.A. and Weinberg, R.A. (1999) Tossing monkey wrenches into the clock: new ways of treating cancer [In Process Citation]. *Proc Natl Acad Sci U S A*, **96**, 4221-4223.

Lenburg, M.E. and O'Shea, E.K. (1996) Signaling phosphate starvation. *Trends Biochem Sci*, **21**, 383-387.

Lew, D.J. and Kornbluth, S. (1996) Regulatory roles of cyclin dependent kinase phosphorylation in cell cycle control. *Curr Opin Cell Biol*, **8**, 795-804.

Lew, J. and Wang, J.H. (1995) Neuronal cdc2-like kinase. *Trends in biochemistry*, **20**, 33-87.

Lisztwan, J., Marti, A., Sutterluty, H., Gstaiger, M., Wirbelauer, C. and Krek, W. (1998) Association of human CUL-1 and ubiquitin-conjugating enzyme CDC34 with the F-box protein p45(SKP2): evidence for evolutionary conservation in the subunit composition of the CDC34-SCF pathway. *Embo J*, **17**, 368-383.

Liu, F., Stanton, J.J., Wu, Z. and Piwnica-Worms, H. (1997) The human Myt1 kinase preferentially phosphorylates Cdc2 on threonine 14 and localizes to the endoplasmic reticulum and Golgi complex. *Mol Cell Biol*, **17**, 571-583.

Lock, L.F. and Wickramasinghe, D. (1994) Cyclin with CDKs. *Trends in cell biology*, **4**, 404-405.

Luca, F.C. (1993) Does degradation lead to segregation? *Current biology*, **3**, 716-718.

Lundgren, K., Walworth, N., Booher, R., Dembski, M., Kirschner, M. and Beach, D. (1991) mik1 and wee1 cooperate in the inhibitory tyrosine phosphorylation of cdc2. *Cell*, **64**, 1111-1122.

Maldonado, E., Shiekhattar, R., Sheldon, M., Cho, H., Drapkin, R., Rickert, P., Lees, E., Anderson, C.W., Linn, S. and Reinberg, D. (1996) A human RNA polymerase II complex associated with SRB and DNA-repair proteins [published erratum appears in Nature 1996 Nov 28;384(6607):384]. *Nature*, **381**, 86-89.

Marshall, N.F. and Price, D.H. (1995) Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem*, **270**, 12335-12338.

Martin-Castellanos, C. and Moreno, S. (1997) Recent advances on cyclins, CDKs and CDK inhibitors. *Trends in cell biology*, **7**, 95-98.

Meijer, L., Guidet, S. and Tung, H.Y.L. (eds.) (1995) *Progress in cell cycle research*. Plenum press, New York.

Mendenhall, M.D. (1993) An inhibitor of p34CDC28 protein kinase activity from Saccharomyces cerevisiae. *Science*, **259**, 216-219.

Meyerson, M., Enders, G.H., Wu, C.-L., Su, L.-K., Gorka, C., Nelson, C., Harlow, E. and Tsai, L.-H. (1992) A family of human cdc2-related protein kinases. *EMBO J.*, **11**, 2909-2917.

Michaelis, C., Ciosk, R. and Nasmyth, K. (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, **91**, 35-45.

Millward, T.A., Zolnierowicz, S. and Hemmings, B.A. (1999) Regulation of protein kinase cascades by protein phosphatase 2A. *Trends Biochem Sci*, **24**, 186-191.

Mitchison, J.M. (1971) *The biology of the cell cycle*. Cambridge university press, Cambridge.

Mittnacht, S. (1998) Control of pRB phosphorylation. *Curr Opin Genet Dev*, **8**, 21-27.

Molz, L. and Beach, D. (1993) Characterization of the fission yeast mcs2 cyclin and its associated protein kinase activity. *Embo J*, **12**, 1723-1732.

Moreno, S., Hayles, J. and Nurse, P. (1989) Regulation of p34cdc2 protein kinase during mitosis. *Cell*, **58**, 361-372.

Morgan, D. (1996) The dynamics of cyclin-dependant kinase structure. *Curr. Opinion in Cell Biology*, **8**, 767-772.

Morgan, D.O. (1995) Principles of CDK regulation. *Nature*, **374**, 131-134.

Morgan, D.O. (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol*, **13**, 261-291.

Mueller, P.R., Coleman, T.R. and Dunphy, W.G. (1995a) Cell cycle regulation of a Xenopus Wee1-like kinase. *Mol Biol Cell*, **6**, 119-134.

Mueller, P.R., Coleman, T.R., Kumagai, A. and Dunphy, W.G. (1995b) Myt1: a membrane-associated inhibitory kinase that phosphorylates Cdc2 on both threonine-14 and tyrosine-15. *Science*, **270**, 86-90.

Mulligan, G. and Jacks, T. (1998) The retinoblastoma gene family: cousins with overlapping interests. *Trends Genet*, **14**, 223-229.

Murray, A. (1995) Cyclin ubiquitination: The destructive end of mitosis. *Cell*, **81**, 149-152.

Murray, A. and Hunt, T. (1993) *The cell cycle*. W H Freeman, New York.

Murray, A.W. (1993) Turning on mitosis. *Current biology*, **3**, 291-293.

Nakamura, T., Sanokawa, R., Sasaki, Y.F., Ayusawa, D., Oishi, M. and Mori, N. (1995) Cyclin I: a new cyclin encoded by a gene isolated from human brain. *Exp Cell Res*, **221**, 534-542.

Nakayama, K. (1998) Cip/Kip cyclin-dependent kinase inhibitors: brakes of the cell cycle engine during development. *Bioessays*, **20**, 1020-1029.

Nasmyth, K. (1996) At the heart of the budding yeast cell cycle. *Trends Genet*, **12**, 405-412.

Nigg, E.A. (1995) Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays*, **17**, 471-480.

Noble, M.E.M., Endicott, J.A., Brown, N.R. and Johnson, L.N. (1997) The cyclin box fold: protein recognition in cell-cycle and trasncription control. *Trends in Biochemical Sciences*, **22**, 482-487.

Norbury, C. and Nurse, P. (1992) Animal cell cycles and their control. *Ann. Rev. Biochem.*, **61**, 441-470.

Nurse, P. and Bissett, Y. (1981) Gene required in Gl for commitment to cell cycle and in G2 for control of mitosis in fission yeast. *Nature*, **292**, 558-560.

Nurse, P., Masui, Y. and Hartwell, L. (1998) Understanding the cell cycle. *Nat Med*, **4**, 1103-1106.

Okamoto, K. and Beach, D. (1994) Cyclin G is a transcriptional target of the p53 tumor suppressor protein. *EMBO J.*, **13**, 4816-4822.

Pardee, A. (1974) A restriction point for the control of normal animal cell proliferation. *PNAS*, **71**, 1286-1290.

Parge, H.E., Arvai, A.S., Murtari, D.J., Reed, S.I. and Tainer, J.A. (1993) Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science*, **262**, 387-394.

Patra, D. and Dunphy, W.G. (1996) Xe-p9, a Xenopus Suc1/Cks homolog, has multiple essential roles in cell cycle control. *Genes Dev*, **10**, 1503-1515.

Patra, D. and Dunphy, W.G. (1998) Xe-p9, a Xenopus Suc1/Cks protein, is essential for the Cdc2-dependent phosphorylation of the anaphase- promoting complex at mitosis. *Genes Dev*, **12**, 2549-2559.

Patton, E.E., Willems, A.R. and Tyers, M. (1998) Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. *Trends Genet*, **14**, 236-243.

Pavletich, N.P. (1999) Mechanisms of cyclin-dependent kinase regulation: structures of cdks, their cyclin activators, and cip and INK4 inhibitors [In Process Citation]. *J Mol Biol*, **287**, 821-828.

Peter, M. and Herskowitz, I. (1994) Direct inhibition of the yeast cyclin-dependent kinase Cdc28-Cln by Far1. *Science*, **265**, 1228-1231.

Piccoli, R., Tamburrini, M., Piccialli, G., Di Donato, A., Parente, A. and D'Alessio, G. (1992) The dual-mode quaternary structure of seminal RNase. *Proc Natl Acad Sci U S A*, **89**, 1870-1874.

Pines, J. (1995) Cyclins and cyclin-dependant kinases: a biochemical view. *Biochem J.*, **308**, 697-711.

Pines, J. (1996) Cell cycle: reaching for a role for the Cks proteins. *Curr Biol*, **6**, 1399-1402.

Pines, J. and Hunter, T. (1991) Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. *J. Cell Biol.*, **115**, 1-17.

Planas-Silva, M.D. and Weinberg, R.A. (1997) The restriction point and control of cell proliferation. *Curr Opin Cell Biol*, **9**, 768-772.

Polyak, K., Kato, J.-y., Solomon, M.J., Sherr, C.J., Massague, J., Roberts, J.M. and Koff, A. (1994) $p27^{Kip1}$, a cyclin-Cdk inhibitor, links transforming growth factor-β and contact inhibition to cell cycle arrest. *Genes Develop.*, **8**, 9-22.

Poon, R.Y.C. and Hunter, T. (1995) Dephosphorylation of Cdk2 Thr160 by the cyclin dependent kinase-interacting phosphatase KAP in the absence of cyclin. *Science*, **270**, 90-93.

Rechsteiner, M. and Rogers, S.W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci*, **21**, 267-271.

Rennie, J. and Rusting, R. (1996) Making headway against cancer. *Sci Am*, **275**, 56-59.

Resnitzky, D. and Reed, S.I. (1995) Different roles for cyclins D1 and E in regulation of the G1-to-S transition. *Mol Cell Biol*, **15**, 3463-3469.

Rhind, N., Furnari, B. and Russell, P. (1997) Cdc2 tyrosine phosphorylation is required for the DNA damage checkpoint in fission yeast. *Genes Dev*, **11**, 504-511.

Richardson, H.E., Stueland, C.S., Thomas, J., Russell, P. and Reed, S.I. (1990) Human cDNAs encoding homologs of the small p34Cdc28/Cdc2-associated protein of Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Genes Dev*, **4**, 1332-1344.

Rickert, P., Seghezzi, W., Shanahan, F., Cho, H. and Lees, E. (1996) Cyclin C/CDK8 is a novel CTD kinase associated with RNA polymerase II. *Oncogene*, **12**, 2631-2640.

Ruffolo, R.R., Poste, G. and Metcalf, B.W. (eds.) (1997) *Cell cycle regulation*. Harwood academic publishers.

Russo, A., Jeffrey, P.D. and Pavletich, N.P. (1996a) Structural basis of cyclin dependent kinase activation by phosphorylation. *Nat. Struct. Biol.*, **3**, 696-700.

Russo, A.A., Jeffrey, P.D., Patten, A.K., Massague, J. and Pavletich, N.P. (1996b) Crystal structure of the $p27^{Kip1}$ cyclin-dependent-kinase inhibiotr bound to the cyclinA-CDK2 complex. *Nature*, **382**, 325-331.

Russo, A.A., Tong, L., Lee, J.O., Jeffrey, P.D. and Pavletich, N.P. (1998) Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature*, **395**, 237-243.

Sanchez, I. and Dynlacht, B.D. (1996) Transcriptional control of the cell cycle. *Curr Opin Cell Biol*, **8**, 318-324.

Santos, R.C., Waters, N.C., Creasy, C.L. and Bergman, L.W. (1995) Structure-function relationships of the yeast cyclin-dependent kinase Pho85. *Mol Cell Biol*, **15**, 5482-5491.

Schlunegger, M.P., Bennett, M.J. and Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem*, **50**, 61-122.

Schulman, B.A., Lindstrom, D.L. and Harlow, E. (1998) Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. *Proc Natl Acad Sci U S A*, **95**, 10453-10458.

Schwob, E., Bohm, T., Mendenhall, M.D. and Nasmyth, K. (1994) The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae [see comments] [published erratum appears in Cell 1996 Jan 12;84(1):following 174]. *Cell*, **79**, 233-244.

Serrano, M., Hannon, G.J. and Beach, D. (1993) A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4 [see comments]. *Nature*, **366**, 704-707.

Sherr, C.J. (1996) Cancer cell cycles. *Science*, **274**, 1672-1677.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J. and Harper, J.W. (1997) F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex [see comments]. *Cell*, **91**, 209-219.

Stern, B. and Nurse, P. (1996) A quantitative model for the cdc2 control of S phase and mitosis in fission yeast. *Trends Genet*, **12**, 345-350.

Stinchcomb, D.T. (1995) Constraining the cell cycle: regulating cell division and differentiation by gene therapy [comment]. *Nat Med*, **1**, 1004-1006.

Sudakin, V., Ganoth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V. and Hershko, A. (1995) The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol Biol Cell*, **6**, 185-197.

Sudakin, V., Shteinberg, M., Ganoth, D., Hershko, J. and Hershko, A. (1997) Binding of activated cyclosome to p13suc1: Use for affinity purification. *J Biol Chem*, **272**, 18051-18059.

Svejstrup, J.Q., Vichi, P. and Egly, J. (1996) The multiple roles of transcription/repair factor TFIIH. *Trends in Biochem. Sci.*, **21**, 346-350.

Tamura, K., Kanaoka, Y., Jinno, S., Nagata, A., Ogiso, Y., Shimizu, K., Hayakawa, T., Nojima, H. and Okayama, H. (1993) Cyclin G: a new mammalian cyclin with homlogy to fission yeast Cig1. *Oncogene*, **8**, 2113-2118.

Tang, D., Chun, A.C.S., Zhang, M. and Wang, J.H. (1997) Cyclin-dependent kinase 5 (Cdk5) activation domain of neuronal Cdk5 activator. Evidence of the existence of cyclin fold in neuronal Cdk5a activator. *J Biol Chem*, **272**, 12318-12327.

Tassan, J.-P., Jaquenoud, M., fry, A.M., Frutger, S., Hughes, G.J. and Nigg, E.A. (1995a) In vitro assembley of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. *EMBO J.*, **14**, 5608-5617.

Tassan, J.P., Jaquenoud, M., Leopold, P., Schultz, S.J. and Nigg, E.A. (1995b) Identification of human cyclin-dependent kinase 8, a putative protein kinase partner for cyclin C. *Proc Natl Acad Sci U S A*, **92**, 8871-8875.

Thuret, J.-Y., Valay, J.-G., Faye, G. and Mann, M. (1996) Civ1 (CAK in vivo), a novel Cdk-Activating kinase. *Cell*, **86**, 565-576.

Townsley, F.M. and Ruderman, J.V. (1998) Proteolytic ratchets that control progression through mitosis. *Trends Cell Biol*, **8**, 238-244.

Udvardy, A. (1996) The role of controlled proteolysis in cell-cycle regulation. *Eur J Biochem*, **240**, 307-313.

Valay, J.G., Simon, M. and Faye, G. (1993) The kin28 protein kinase is associated with a cyclin in Saccharomyces cerevisiae. *J Mol Biol*, **234**, 307-310.

Vogel, L. and Baratte, B. (1996) Suc1: cdc2 affinity reagent or essential cdk adaptor protein? *Prog Cell Cycle Res*, **2**, 129-135.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H. and Jones, K.A. (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*, **92**, 451-462.

Weinberg, R.A. (1996) How cancer arises. *Sci Am*, **275**, 62-70.

Woollard, A. and Nurse, P. (1995) G1 regulation and checkpoints operating around START in fission yeast. *Bioessays*, **17**, 481-490.

Yam, C.H., Ng, R.W., Siu, W.Y., Lau, A.W. and Poon, R.Y. (1999) Regulation of cyclin A-Cdk2 by SCF component Skp1 and F-box protein Skp2. *Mol Cell Biol*, **19**, 635-645.

Yang, X., Gold, M.O., Tang, D.N., Lewis, D.E., Aguilar-Cordova, E., Rice, A.P. and Herrmann, C.H. (1997) TAK, an HIV Tat-associated kinase, is a member of the cyclin-dependent family of protein kinases and is induced by activation of peripheral blood lymphocytes and differentiation of promonocytic cell lines. *Proc Natl Acad Sci U S A*, **94**, 12331-12336.

Yang, X., Herrmann, C.H. and Rice, A.P. (1996) The human immunodeficiency virus Tat proteins specifically associate with TAK in vivo and require the carboxyl-terminal domain of RNA polymerase II for function. *J Virol*, **70**, 4576-4584.

Yankulov, K.Y. and Bentley, D.L. (1997) Regulation of CDK7 substrate specificity by MAT1 and TFIIH. *Embo J*, **16**, 1638-1646.

Ye, X.S., Fincher, R.R., Tang, A. and Osmani, S.A. (1997) The G2/M DNA damage checkpoint inhibits mitosis through Tyr15 phosphorylation of p34cdc2 in Aspergillus nidulans. *Embo J*, **16**, 182-192.

Zauberman, A., Lupo, A. and Oren, M. (1995) Identification of p53 target genes through immune selection of genomic DNA: the cyclin G gene contains two distinct p53 binding sites. *Oncogene*, **10**, 2361-2366.

Zetterberg, A., Larsson, O. and Wiman, K.G. (1995) What is the restriction point? *Curr Opin Cell Biol*, **7**, 835-842.

Zhang, H., Kobayashi, R., Galktionov, K. and Beach, D. (1995) p19Skp1 and p45Skp2 are essential elemets of the cyclin A-CDK2 S phase kinase. *Cell*, **82**, 915-925.

Chapter 2: The crystal structure of *S. pombe* p13^{suc1} at 2.3 Angstroms (Å) resolution.

CONTENTS

LIST OF FIGURES **61**

LIST OF TABLES **61**

2.1 INTRODUCTION **61**

2.2 MATERIALS, METHODS & RESULTS **62**

2.2.1 Crystallisation of S. pombe Suc1 ..... *62*
2.2.2 Data collection and processing *63*
2.2.3 Structure refinement *67*
2.2.4 The structure of S. pombe Suc1 ..... *68*
2.2.5 Comparison of 100K and room temperature structures of Suc1 ..... *72*

2.3 DISCUSSION **80**

2.4 LITERATURE CITED **81**

List of figures

Figure 2.1 Diffraction image of a Suc1 crystal ---------- 65
Figure 2.2 The structure of Suc1 viewed approximately along the 2-fold axis ---------- 70
Figure 2.3 Illustration of the co-ordination of Zn1 and Zn2 in dimer and tetrameric association ---------- 71
Figure 2.4 Comparative temperature factor plots for Suc1 structures ---------- 73
Figure 2.6 Bond lengths between the zinc atoms and their respective ligands ---------- 77
Figure 2.7 Stereo diagram showing part of the electron density at the hinge region ---------- 78
Figure 2.8 Schematic representation of the water molecules in (a). 293K and (b). 100K Suc1 structures --- 79

List of Tables

Table 2.1 Summary of data collection strategy for 100K Suc1 crystals _____ 64
Table 2.2 Summary of reflections, intensities and R-factors by shells as output by SCALEPACK _____ 66
Table 2.3 Refinement statistics for wild type Suc1 protein _____ 68

2.1 INTRODUCTION

During the course of collecting X-ray data, protein crystals tend to suffer from radiation induced decay when they are exposed to X-ray beams at room temperature (Blundell and Johnson, 1976). It is thought that this radiation damage is the result of the generation and diffusion of free radicals throughout the crystal lattice. The adverse effect of this phenomenon is that there is rapid loss of intensity especially for the highest resolution reflections during the course of data collection. Protein crystals also suffer from thermal damage when irradiated with high intensity X-ray beams (Chayen *et al.*, 1996). The combined effect of radiation and thermal damage can limit both the quality and quantity of an X-ray data set especially with small protein crystals (Rodgers, 1994). Performing the data collection experiment on crystals flash-cooled at low temperatures (around 100K) minimises the effects of radiation and thermal damage. Apart from a few exceptions, cryo-crystallography (Appendix I) has become the method of choice for macromolecular crystallography data collection experiments (Garman and Schneider, 1997). Compared with room temperature data sets, a cryo-data set of the same protein crystal often leads to an extension of the resolution limit because of the reduced atomic

thermal vibrations. Another advantage of cryo-crystallography is that a whole data set can be collected from a single crystal. In general, protein structures from cryo-data sets have low atomic temperature factors (B-factors) and more than twice as many ordered water molecules as equivalent room temperature data (Chayen *et al.*, 1996).

The structure of *S. pombe* Suc1 was first determined in this laboratory with data collected to 2.7Å at room temperature (Endicott *et al.*, 1995). The data set was collected during the commissioning phase of the European Synchrotron Radiation Facility (ESRF) in 1994. The high brightness and high intensity of the BL3/ID19 beamline (Hellemans, 1994) and the relatively small crystal size (20 x 20 x 100μm^3) were such that significant radiation damage was inflicted on the Suc1 crystal. A cryo-cooling facility was yet to be installed at the beamline, so the data set was collected at room temperature using several crystals for the native and derivative data sets. The structure was solved by Multiple isomorphous replacement (MIR) and it revealed that *S. pombe* Suc1 was a zinc-mediated dimer, not a strand-exchanged dimer as is CksHs2 (Parge *et al.*, 1993), although the hinge region (His88 - Val92) that determines whether Cks proteins might fold either as a monomer or a strand-exchanged dimer (see chapter 1) was supported by relatively poor electron density. Also, the first five amino acids of the N-terminus and the last twelve residues of the C-terminus of the Suc1 structure could not be modelled because they had no supporting electron density. In order to extend the quality of the Suc1 protein structure and also confirm the conclusion(s) inferred from the 2.7Å model, a new set of X-ray data to 2.3Å resolution was collected at the ESRF at a temperature of 100K. A comparative analysis of the two protein structures is presented below.

2.2 MATERIALS, METHODS & RESULTS

2.2.1 Crystallisation of S. pombe Suc1

Suc1 was crystallised (Endicott *et al.*, 1995) using the hanging-drop vapour diffusion method (McPherson, 1985). Characterisation of Suc1 by these workers showed that it was monomeric in a solution of 10mM Tris, 0.2M NaCl pH7.4. Also, it was observed that appropriate concentrations of zinc ions could induce dimerisation of Suc1. This led to the inclusion of zinc sulphate in the crystallisation solutions. Briefly, an

aliquot (1µl) of the protein stock (~18.5-32.5mg/ml in 10mM Tris pH 7.5) was mixed with 1µl of mother liquor from the Crystal screen I kit (Hampton Research) on Thermanox (Nunc Inc.,) coverslips. Each mother liquor was supplemented with zinc sulphate such that the final molar ratio of Suc1 protein to zinc ions in the crystallisation drop was 2:1 i.e. Suc1: zinc ion molar concentration. Poor quality crystals grew overnight at 20°C from a solution consisting of 9% PEG (8K), 100mM zinc acetate and half-equimolar zinc sulphate (i.e. at a ratio of 2 moles of Suc1 protein to a mole of zinc ions), 0.05M sodium cacodylate pH 6.5. Substitution of zinc acetate by ammonium acetate led to a significant improvement in the quality of the crystals (Endicott *et al.*, 1995). The final crystallisation conditions are 5% PEG (8K), 100mM ammonium acetate and 1.5mM zinc sulphate, 0.1% dioxane and 0.1M MES pH 6.5.

2.2.2 Data collection and processing

Data were collected on a 30cm diameter MAR Research image plate at the X-ray diffraction beamline (BL3/ID19) at the ESRF in Grenoble, France in September 1995 by Drs Noble and Endicott. The crystals used in these experiments were grown by Dr. Endicott. A single crystal was soaked for 2 minutes in a cryo-solution (30% glycerol in mother liquor i.e. 5% PEG 8K, 0.1M ammonium acetate and 1.5mM zinc sulphate and 0.1M MES pH 6.5) and flash-cooled to 100K in a stream of gaseous nitrogen. The data collection strategy was as follows: a single diffraction image was collected and auto-indexed with the program DENZO (Otwinowski and Minor, 1997) in order to obtain the unit cell dimensions and orientation of the crystal with respect to the beam. After the refinement of the experimental parameters and subsequent integration of the reflections on the first image, the results were fed into the program PREDICT (M.E.M Noble, unpublished work) in order to obtain guidelines such as the starting phi (ϕ) angle and the total ϕ range required to collect a complete data set. In order to overcome the problem of overloaded reflections, at low resolution when collecting high resolution data sets, the data collection experiment was divided into two parts: first, a high resolution data set was collected with an oscillation angle ($\Delta\phi$) of 1.0° per image before the detector was moved back (from 340mm to 420mm) for the collection of a low resolution data set with a $\Delta\phi$ of 2.5° per image. A smaller oscillation angle was used for the high resolution pass to avoid

the reflections overlapping each other. A summary of the data collection strategy is presented in table 2.1.

	High resolution data set	Low resolution data set
Maximum resolution	2.18Å	2.64Å
Crystal-to-detector distance (d)	340mm	420mm
Number of diffraction images	42	11
Oscillation range ($\Delta\phi$)/image	1°	2.5°
Total ϕ range	42°	27.5°
Exposure time	120s	120s

Table 2.1 Summary of data collection strategy for 100K Suc1 crystals

The X-ray wavelength was 0.9Å. The crystal belonged to spacegroup $P6_522$ with unit cell dimensions a=b=60.1Å and c=264.5Å and there was one dimer of Suc1 molecule per asymmetric unit. The data were integrated with DENZO (Otwinowski and Minor, 1997) and scaled with Scalepack (Otwinowski and Minor, 1997). The average mosaicity was 0.29° as compared to 0.1° for some of the room temperature Suc1 crystals. The measured intensities were converted to structure factor amplitudes with the program TRUNCATE (CCP4, 1994). About 70 reflections from the scaled data were too weak to be converted to structure factor amplitudes by TRUNCATE.

(a).

(b).

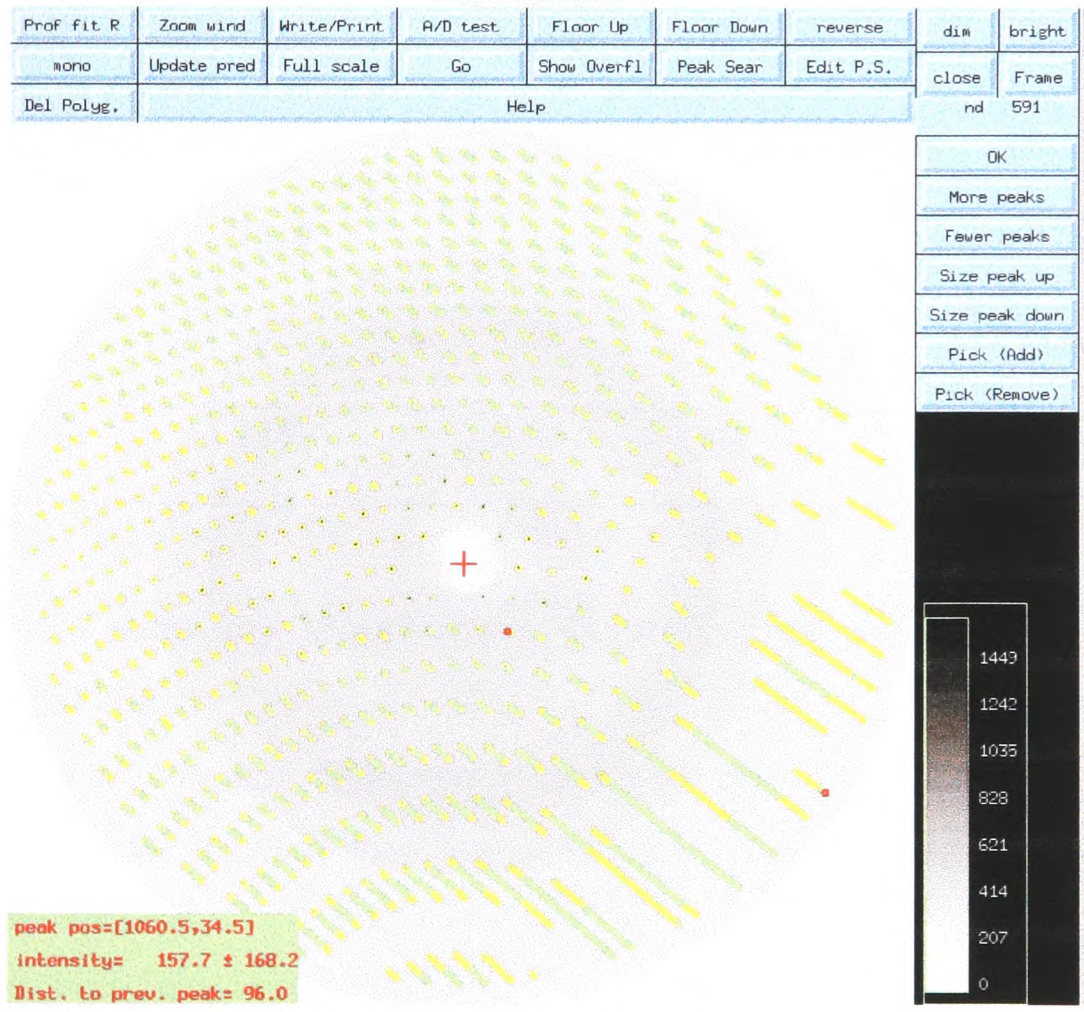


Figure 2.1 Diffraction image of a Suc1 crystal

(a) Raw data. (b) The predicted spots have been superimposed on the image with the auto-indexing program DENZO. Yellow circles represent partially recorded reflections,

green ones represent fully recorded reflections and red spots are indicative of anomalies such as overlapping spots, the shadow of the backstop, edge of the image plate or overloading of that particular spot.

Resolution shell (Å)	Reflections[¶]	<I/σ(I)>	%Completeness[¶]	R_{merge}*	Multiplicity
20.00-4.93	1531	22.49	99.3	0.056	4.06
4.93-3.93	1398	26.76	99.1	0.055	4.43
3.93-3.43	1359	24.61	99.7	0.063	4.46
3.43-3.12	1332	19.47	99.6	0.077	4.52
3.12-2.90	1330	13.09	99.6	0.109	4.60
2.90-2.73	1304	10.00	99.8	0.145	4.67
2.73-2.59	1320	7.61	99.9	0.183	4.69
2.59-2.48	1308	5.70	100.0	0.229	4.68
2.48-2.38	1311	4.18	100.0	0.293	4.62
2.38-2.30	1280	3.58	100.0	0.35	4.67
20.00-2.30	13473	19.56	99.7	0.074	4.43

Table 2.2 Summary of reflections, intensities and R-factors by shells as output by SCALEPACK

Although the crystals diffracted to as high a resolution as 2.18Å, data beyond 2.30Å, were excluded because they merged poorly (an R_{merge} of 56.1% for the 2.28-2.20Å resolution shell).

[¶]*Number of unique reflections.*

¶*Percentage of theoretically possible data*

$$*R_{merge} = \frac{\sum_i \sum_h (I_i(h) - \langle I(h) \rangle)}{\sum_i \sum_h I(h)}$$

*where $I_i(h)$ and $\langle I(h) \rangle$ are the **ith** measurement and the average measured intensity of reflection **h**.*

2.2.3 Structure refinement

The previously determined protein structure of Suc1 at 2.7Å was used to calculate the initial phases for the 2.3Å structure. Comparison of the unit cell dimensions of the room temperature data set with that of the 100K data set showed that the two unit cells were not completely isomorphous. Freezing the Suc1 crystal had shrunk the unit cell dimensions by 0.77Å in the ***a*** and ***b*** axes and 1.34Å in the ***c*** axis. This accounts for the relatively high initial R values: the initial R_{work} was 55% and R_{free} was 52%. In order to determine the correct position of the new model in the unit cell, the crude model was subjected to a round of rigid body refinement with the program XPLOR (Brunger, 1992). This reduced the R_{work} value to 49.4% and the R_{free} to 48.7%. At this stage, an electron density map (2Fo-Fc contoured at the nominal electron density of 0.3 e/Å^3 which corresponds to the 1σ level calculated for the map) was calculated and the position of each amino acid in the electron density was carefully examined with the graphics program O (Jones, 1991). This was followed by a round of positional refinement in XPLOR in which the ***A*** and ***B*** subunits (which are related by a non-crystallographic 2-fold axis) of the Suc1 molecule were treated as two separate objects instead of one rigid entity. This resulted in a further decrease in the R_{work} to 34.2% and R_{free} was 38.3%. A new 2Fo-Fc electron density map was calculated and the ***A*** subunit was superimposed on the ***B*** subunit in order to look for regions where the two subunits were identical. This was done to exploit the presence of the 2-fold non-crystallographic symmetry for the refinement process (Vellieux and Read, 1997). Non-crystallographic symmetry (ncs) restraints were imposed on these identical regions and the model was subjected to a round of simulated annealing refinement with a starting temperature of 2000K using XPLOR. This gave an R_{work} value of 28.3% and R_{free} of 36.6%. A new 2Fo-Fc electron density map was calculated and the regions of similarity between the two subunits were inspected

and modified before further positional refinement was carried out with the program REFMAC (CCP4, 1994). This gave an R_{work} of 30.6% and R_{free} of 33.9%. After anisotropic scaling and temperature factor refinement in REFMAC, the R_{work} decreased to 25.5% and the R_{free} value was 30.1%. At this stage, water molecules were added using the program ARP (CCP4, 1994) with subsequent refinement of the water positions with REFMAC. The final protein model contains 175 well-defined water molecules, three zinc atoms and the final R_{work} is 22.6% and R_{free} is 28.98%.

REFINEMENT	100K Suc1 structure	293K Suc1 structure (Endicott *et al.*, 1995)
Resolution Range (Å)	20 – 2.3Å	20 – 2.7Å
Total number of recorded reflections	13403	Not available
Number of reflections in working set	12900	Not available
Number of reflections in test set	493	Not available
Final $^{\delta}R_{work}$ (R_{free})	22.8% (28.98%)	20.0% (27.0%)
	NUMBER OF ATOMS	
Protein	1628	1628
Zinc	3	3
Waters	175	23
Average B-factor of protein atoms, zinc ions and water molecules, respectively	22.9Å^2, 16.5Å^2, 32.3Å^2	45.1Å^2, 37.5Å^2, 42.8Å^2
RMS deviation from ideal geometry:		
Bond lengths→	0.0231Å	0.012Å
Bond angles →	2.5°	1.9°

Table 2.3 Refinement statistics for wild type Suc1 protein

*$^{\delta}$Crystallographic $R_{work}= \sum_h |F_o-F_c| / \sum_h F$ where F_o and F_c are the observed and calculated structure factor amplitudes of reflection **h**, respectively. R_{free} is calculated as for R_{work} but using about 5% of the data set that had been excluded from the refinement.*

2.2.4 The structure of S. pombe Suc1

Apart from the regions with high B-factors in the room temperature Suc1 structure, the 100K Suc1 structure at 2.3Å is nearly identical to the 2.7Å (Endicott *et al.*,

1995) room temperature structure. The overall structure is a zinc-mediated dimer (fig. 2.2) with two zinc atoms at the dimer interface. The two monomers (denoted ***A*** and ***B*** subunits) are related by 180° rotation. The third zinc atom is sandwiched between two dimers that are related by a crystallographic 2-fold axis. These zinc atoms also make contact to symmetry related molecules in the unit cell so that the packing in the unit cell can be considered as stemming from a repeating tetrameric unit (fig. 2.3). Suc1 is a small (13KDa) and stable protein consisting of four β-sheets joined by loop regions to four α-helices (fig. 2.2). The secondary structure elements begin at the N-terminus with an α-helix (α1), a pair of β-strands (β1a and β1b) followed by a pair of α-helices (α2 and α3) then comes α4 and a pair of β-strands (β3 and β4) at the C-terminus. The following salt-bridge interactions are present in each of the monomers that constitute the zinc-stabilised dimer: Arg8-Asp54, Arg8-Glu14, Glu37-Arg99, Arg64-Glu70, Glu68-Arg72 and His88-Glu91 (subunit ***A*** only) and Glu86-His88 (subunit ***B*** only). Two notable surface features of the Suc1 molecule include a region predominated by aromatic and hydrophobic residues (Tyr31, Tyr38, His40, Tyr85, His88, Pro90, Glu91, His93, Ile94 and Leu96): all these residues are conserved and have exposed side-chains. The second feature of interest is a region consisting of a cluster of positively charged side-chains contributed by Arg30, Arg39 and Arg99.



Figure 2.2 The structure of Suc1 viewed approximately along the 2-fold axis

*In this cartoon representation, subunit **A** is coloured green, subunit **B** is coloured blue, and the zinc atoms are coloured magenta. The chain begins at Val6 of the N-terminus and ends at Lys101 of the C-terminus because of the absence of electron density for Ser2-Gly5 and Asp102-Gly113. There are three zinc atoms in the Suc1 dimer. Zn1 and Zn2 are located at the dimer interface while Zn3 forms a crystal lattice contact site with the side-chains of Glu86 and His88, His88' and a water molecule. The dimer interface is further characterised by several charged-residue to charged-residue interactions: GluA16-ArgB30, GluA16-ArgB99, GluA19-ArgB30, GluB16-ArgA30, GluB16-ArgA99 and*

AspA23-HisB26. The hinge region consists of residues His88, Val89, Pro90, His91 and Pro92. This figure was produced with Aesop (M.E.M Noble, unpublished work).



Figure 2.3 Illustration of the co-ordination of Zn1 and Zn2 in dimer and tetrameric association

The subunits of the non-crystallographic dimer are shown as one light grey and one dark grey ellipse labelled A and B, respectively. Residues from both subunits A and B are involved in the co-ordination of Zn1 and Zn2. The interaction between Zn2 and Zn2' is mediated by two water molecules. Reproduced from (Endicott et al., 1995).

2.2.5 Comparison of 100K and room temperature structures of Suc1

Globally, the two protein structures are very similar. Superimposition of the room temperature and 100K structures (using C_α atoms only), gave a root mean square deviation (RMSD) of 0.36Å. Superimposition of subunit ***A*** of the room temperature Suc1 structure on subunit ***A*** of the 100K structure gave an RMSD of 0.34Å and that for the ***B*** subunit was 0.33Å. There was a relatively poorer fit between the ***A*** and ***B*** subunits of the same protein structure e.g. the ***A*** and ***B*** subunits of the room temperature Suc1 structure (RMSD) was 0.42Å and that for the 100K structure was 0.60Å, indicating that inter-subunit differences are more pronounced in the 100K Suc1 structure. A detailed examination of each equivalent pair of residues (after superimposition of the ***A*** subunit on the ***B*** subunit of the 100K Suc1 structure) revealed that most of the main chain atoms bar two regions were identical. The first discrepant region consists of residues 31-38. Examination of the electron density around this loop shows that it is more ordered in subunit ***B*** than in ***A***. The significance of this difference is unclear but this is a solvent accessible loop with relatively high temperature factors (an average of 33.6Å^2 compared with an average of 22.9Å^2 for all the protein atoms). The second region consists of residues 88-92. This is part of the hinge region. As described in the preceding chapter, this region ought to be inherently flexible if it is to play the structural role ascribed to it (Bourne *et al.*, 1995). The average B-factor for the hinge region in the 100K and 293K Suc1 structures are 18.7Å^2 and 36.4Å^2 , respectively. Hence, the B-factors for this region in the 100K structure are lower than the average for the entire protein suggesting that the different conformations of the hinge represent different static structures and do not arise from thermal mobility. Apart from the above two regions, most of the side chains of equivalent residues in the ***A*** and ***B*** subunits of the 100K Suc1 model occupied identical positions in the electron density. Side chains of residues Val6, Arg17, Gln24, His26, Met42, Glu59, Thr60, Leu80 differed in conformation. Val6 and Lys101 are the last observed residues at the N- and C-terminal residues, respectively. Both residues are exposed to solvent and also they possess higher than average temperature factors. The rest of the side-chain differences do not appear to be significant and they seem to arise as a result of a more ordered structure at 100K. In both the room temperature and 100K Suc1 structures, the first five residues of the N-terminus and the last twelve residues of

the C-terminus were not modelled because of the absence of supporting electron density. Extrapolation of the B-factor plots at both termini suggests that these missing regions are likely to have high B-factors. Variation of temperature factors follows the same trend in both subunits of the 100K and the 293K Suc1 structures. A summary of the mean B factors for both the room temperature and 100K Suc1 structure is shown in fig. 2.4 where corrected values (i.e. $B_{corrected} = B_{c\alpha} - B_{mean}$) are shown for main chain atoms only. After correction for average main chain B factors (45.1Å^2 for 293K structure; 22.9Å^2 for 100K structure) it can be seen that the variation in B-factor with residue number are very similar for both structures.

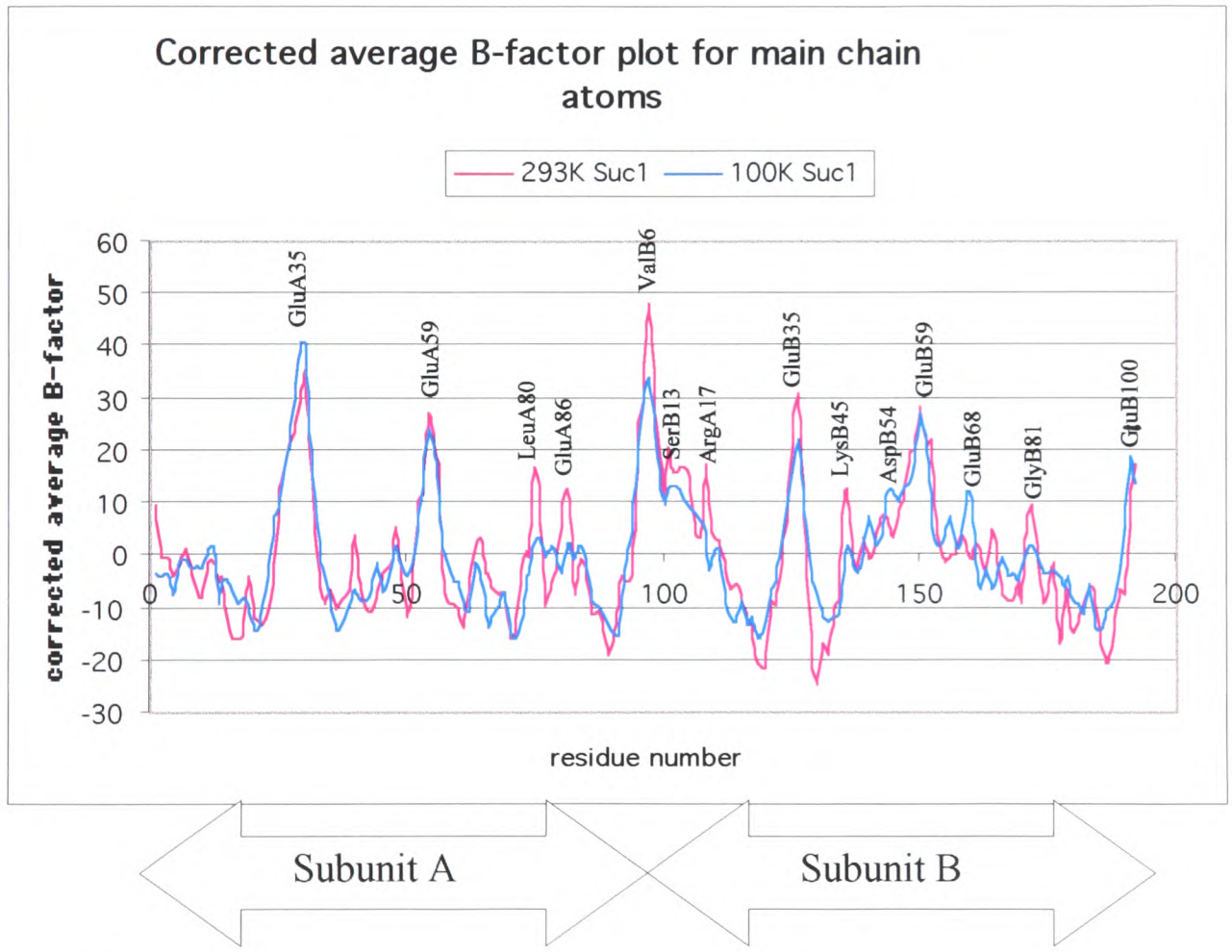


Figure 2.4 Comparative temperature factor plots for Suc1 structures

The blue line represents the plot for the 293K Suc1 structure and that of the100K structure is magenta. Refer to the text for a discussion of the temperature variation in the two structures.

For the ***A*** subunit, the N-terminal residues up to Glu A35 have average B factors. In contrast, the ***B***-subunit N-terminal residues from Val 6 to Arg 17 show higher than average B factors (e.g. Val B6 B = 66.7Å^2; Arg 17 B = 34.1Å^2). This difference may be attributed to the manner in which the two subunits are packed in the unit cell of the crystal. For the rest of the chain, the variation in B factors shows similar trends. There is a high B-factor associated with Glu35 (67.1Å^2 in the *A* subunit). Tyr27-41 forms a solvent exposed loop whose average temperature factor (33.5Å^2) is 10.1Å^2 higher than the average temperature factor of 22.9Å^2 for the 100K Suc1 structure. The intervening residues Met42-Asp54 between this peak and the next one have a relatively low average B-factor (16.4Å^2). The region from Tyr55-Leu66 forms a second region of relatively high B-factors. It has an average B-factor of 26.7Å^2 and the peak (Glu59) has a B-factor of 52.1Å^2.

The 100K Suc1 structure shows an improvement in the stereochemistry of the main chain atoms: in the room temperature structure, 89.6% of residues are in the most favoured region and all the residues are in allowed/generously allowed regions of the Ramachandran plot. This compares with 92.1% of residues in the 100K model being in the most favoured region and 100% of the residues being in the allowed/generously allowed regions (fig. 2.5).

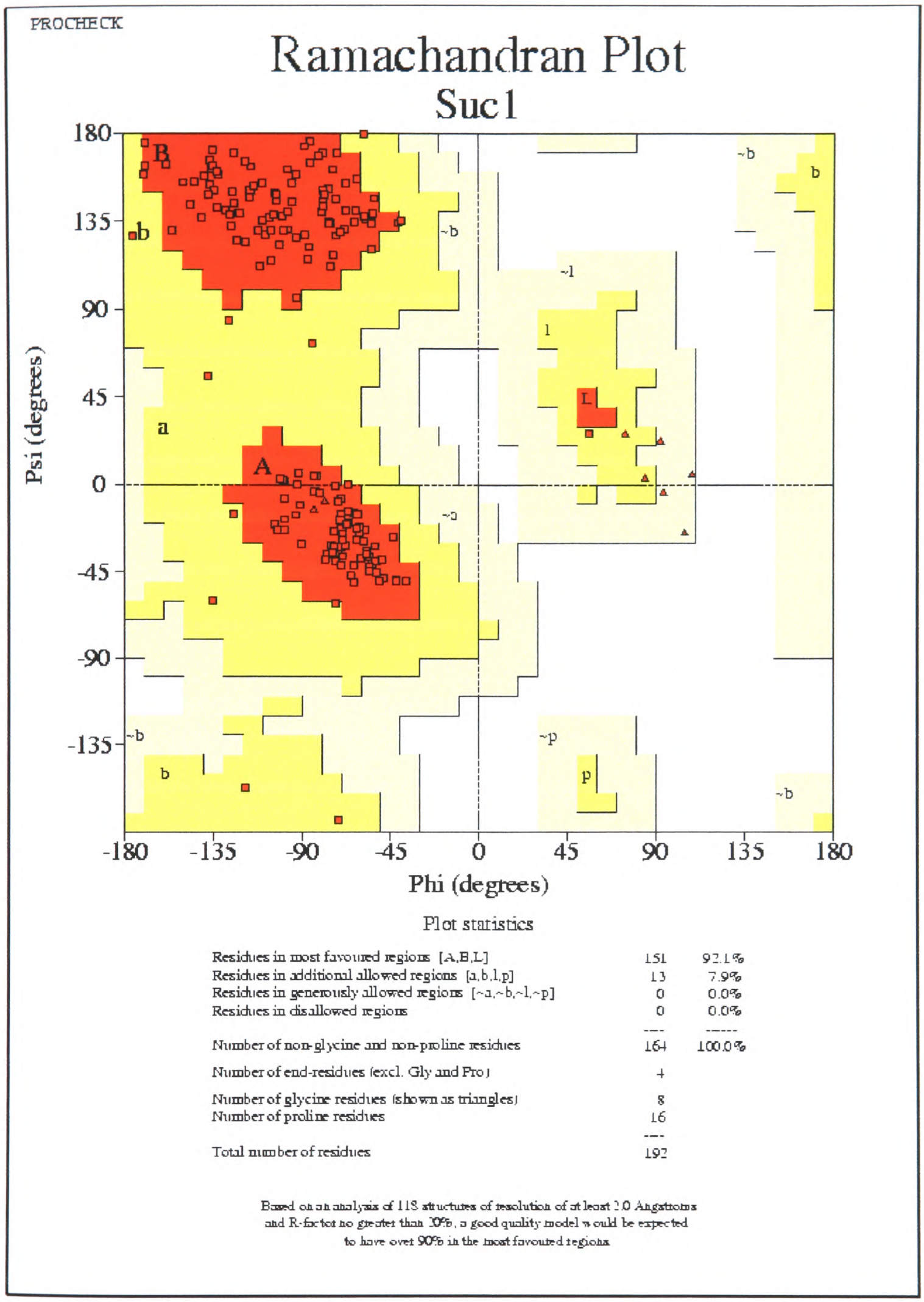


Figure 2.5 Ramachandran plot for 100K Suc1 main chain (N, C_α, C, O) atoms

The following residues; AspB23, AspA33, SerB79 and GluA86, have moved from allowed/generously allowed regions in the room temperature (293K) Suc1 structure to either the most favoured or to additional, allowed regions in the 100K structure.

Two of the three zinc ions in Suc1 are sandwiched at the dimer interface between subunits ***A*** and ***B***. The tetrahedral co-ordination of the zinc atoms observed in the 293K Suc1 structure (Endicott *et al.*, 1995) has been retained in the high resolution 100K structure. The ligands that interact with zinc1 are His26 from the ***A*** subunit, Asp23 of subunit ***B***, Glu91 from a crystallographically related molecule and a water molecule. Zinc2 is co-ordinated by AspA23, HisB26, HisB40 from a crystallographically related molecule and a tightly bound water molecule. The third zinc atom, zinc3, has the following ligands: His88 from subunit ***B***, Glu86 from subunit ***B***, His88 from the ***A*** subunit of a crystallographically related molecule and a water molecule. X-ray crystallographic analysis of zinc ions in tetrahedral complexes of relatively small organic molecules like di(histidino) pentahydrate and di (L-histidino) zinc(II) dihydrate (Harding and Cole, 1963) and (Kretsinger and Cotton, 1963) suggests that in tetrahedral complexes, the Zn—N bond length varies from 2.00 to 2.05Å whilst that for the Zn—O was 2.8-2.9Å. In Suc1, the average Zn—N bond length varies from 2.0-2.4Å for the 100K and 1.9-2.9Å for the 293K structure. Similarly, the Zn—O bond length ranges from 1.8-2.6Å and 2.2-2.8Å for the 100K and 293K structures, respectively (fig. 2.6).

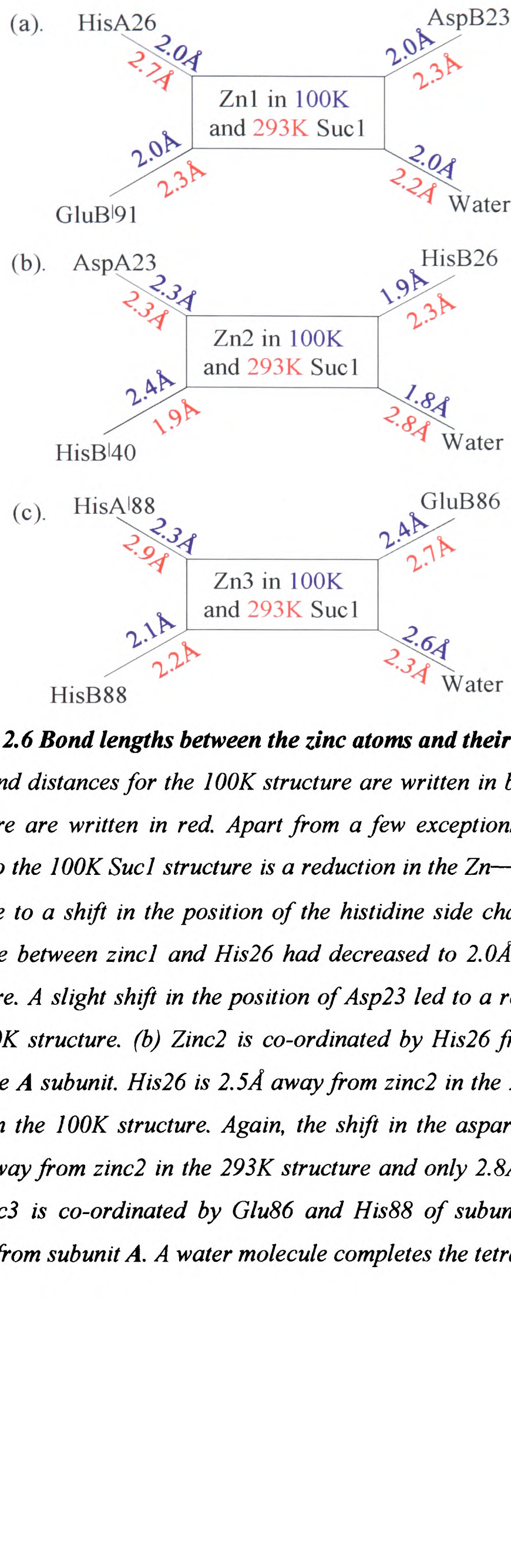


Figure 2.6 Bond lengths between the zinc atoms and their respective ligands

The bond distances for the 100K structure are written in blue ink and that for the 293K structure are written in red. Apart from a few exceptions, the general trend from the 293K to the 100K Suc1 structure is a reduction in the Zn—O and Zn—N bond distances. (a) Due to a shift in the position of the histidine side chain in the 100K structure, the distance between zinc1 and His26 had decreased to 2.0Å from 2.7Å in the 293K Suc1 structure. A slight shift in the position of Asp23 led to a reduction from 3.0Å to 2.8Å in the 100K structure. (b) Zinc2 is co-ordinated by His26 from the ***B*** *subunit and Asp23 from the* ***A*** *subunit. His26 is 2.5Å away from zinc2 in the 293K Suc1 structure and 1.9Å away in the 100K structure. Again, the shift in the aspartate residue is far less, being 2.9Å away from zinc2 in the 293K structure and only 2.8Å away in the 100K structure. (c) Zinc3 is co-ordinated by Glu86 and His88 of subunit* ***B*** *and a symmetry-related His88' from subunit* ***A****. A water molecule completes the tetrahedral geometry.*

Although the part of the model that forms the hinge-region (His88-Pro92) residues was unambiguously assigned in the electron density by Endicott et. al. 1995, the equivalent region in the 100K Suc1 structure provides a slightly clearer picture of the position of each atom in the corresponding electron density (fig. 2.7). Superimposition of these residues from subunit *A* of the 100K structure on that from the 293K structure gave an RMSD of 0.354Å.



Figure 2.7 Stereo diagram showing part of the electron density at the hinge region

The final electron density from REFMAC for the Suc1 structure was used for (a) 293K Suc1 and (b) 100K Suc1. The electron density was contoured at the nominal electron density of 0.75 e/Å^3, which corresponds to the 2.5σ level calculated for the 2Fo-Fc map. Carbon atoms are coloured green, nitrogen blue and oxygen is red. The residues displayed are Val87, His88, Val89, Pro90, Glu91 and Pro92. This figure was produced with Aesop (M.E.M Noble, unpublished work).

Perhaps, the most dramatic difference in the higher resolution 100K Suc1 structure is the increase in the number of well-ordered water molecules from 20 in the 293K structure to 175 (shown in fig 2.8).The waters in the 100K structure have a mean B-factor of 32.3Å whereas those of the 293K Suc1 structure have a mean B-factor of 42.8Å.

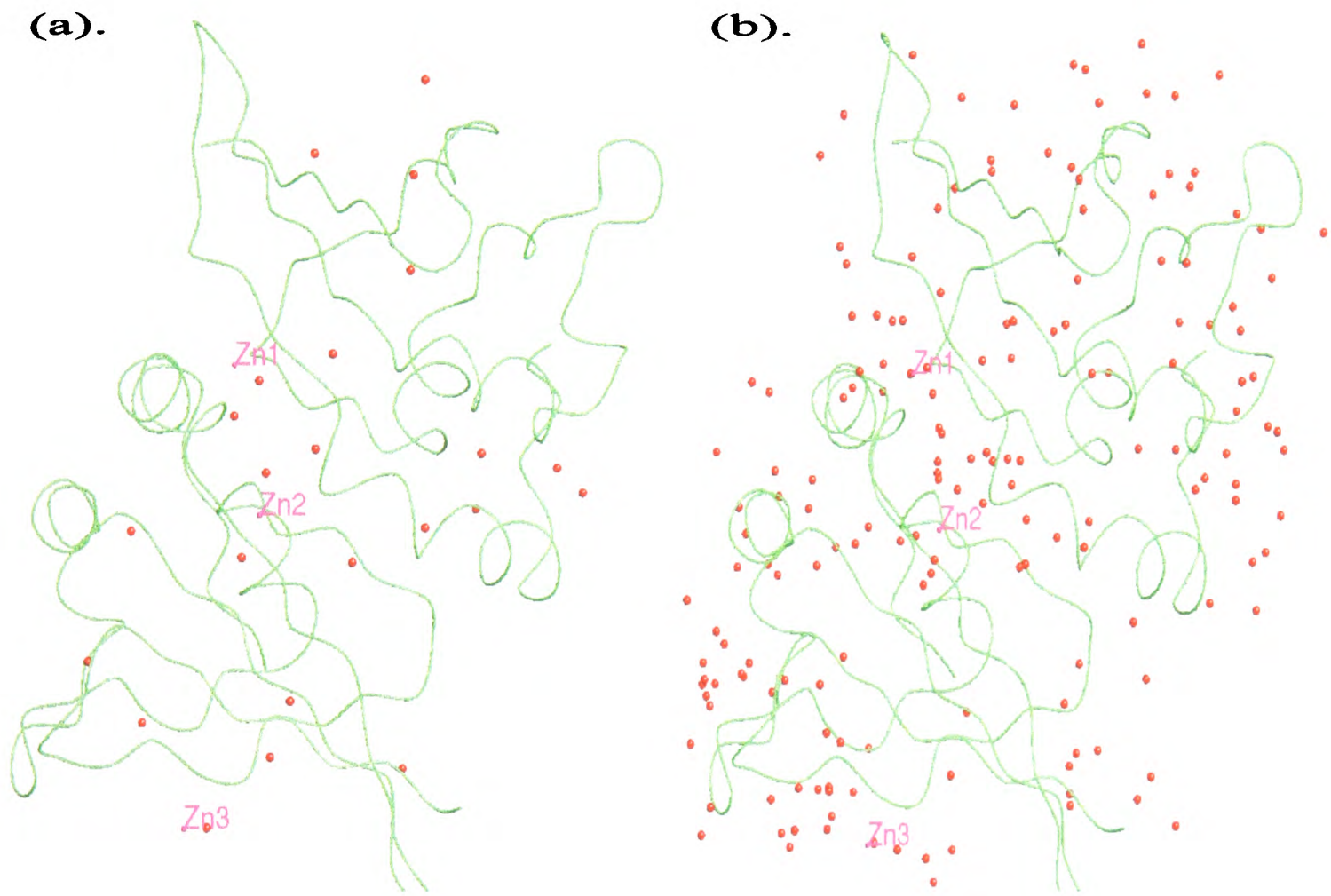


Figure 2.8 Schematic representation of the water molecules in (a). 293K and (b). 100K Suc1 structures

*The orientation of each molecule is the same as in Fig. 2.2. In this worm representation, the **A** and **B** subunits of each Suc1 molecule are coloured green, zinc atoms are coloured magenta and water molecules are coloured red. Most of the additional water molecules in the 100K Suc1 structure constitute a second hydration shell surrounding the protein molecule. In the 293K Suc1 structure, there are about 5 water molecules (compared to at least 20 for the 100K structure) at the interface of the two subunits. A cluster of water molecules surround the zinc3 ion in the 100K Suc1 structure; the significance of this observation is not clear.*

2.3 DISCUSSION

The 100K Suc1 protein structure presented in this chapter provides a more precise description of the structure of the protein compared to the room temperature model. Comparison of the room temperature (293K) and 100K Suc1 structures show a decrease in the overall temperature factor by 22.2Å^2 from 45.1Å^2 to 22.9Å^2. This reduction is consistent with values reported for similar comparative studies (Gregoriou *et al.*, 1998; Tilton *et al.*, 1992). However, a strict comparison of B-factors between the room temperature Suc1 structure at 2.7Å and the 100K Suc1 structure at 2.3Å resolution is not possible because of the different resolutions of the data sets and because of the different refinement programs used in the latter stages of refinement. In particular, B-factors at 2.7Å resolution are difficult to define precisely (Dr. Eleanor Dodson, University of York, personal communication). As expected, variation of temperature factors followed the same trend in both subunits of the 100K and 293K Suc1 structures (fig. 2.4). Despite the extension of the resolution limit from 2.7Å to 2.3Å, the first 5 residues from the N-terminus and the last 12 residues at the C-terminus could not be modelled in the 100K Suc1 structure because of the absence of interpretable electron density for these regions. This was the case for the 293K structure as well and suggests that these terminal residues are either extremely mobile and/or may have static disorder.

The role of zinc ions in proteins can be divided into two broad categories i.e. enzymatic and structural (Alberts *et al.*, 1998; Vallee and Auld, 1990). The zinc ions in carbonic anhydrase are involved in the catalytic mechanism of this enzyme (Coleman, 1992) whilst those in aspartate transcarbamylase are required solely for stabilising this multi-domain protein (Honzatko *et al.*, 1982). The special properties of zinc that confers this degree of versatility are as follows: as a divalent cation, Zn^{2+} has a full complement of 10 electrons in the ***d*** shell. The consequences of this electronic configuration means that zinc lacks a ligand field stabilisation energy when co-ordinated by ligands in any geometry (Berg and Shi, 1996). In contrast, ions with partially filled ***d*** shells have a ligand field stabilisation energy, which selects against certain arrangements of ligands. The second reason for the versatility of zinc is due to fact that it does not belong to any of the extremes, with respect to hard-soft acid-base theory. This feature enables zinc to form

tight interactions with a variety of ligands such as the sulphur from a cysteine residue, the nitrogen from a histidine and the oxygen from glutamate, aspartate and a water molecule. The third reason is due to the lack of multiple redox states. Zinc cannot be oxidised to Zn^{3+} or reduced to Zn^{-} under physiological conditions. This ensures that unwanted chemical reactions would not occur in a protein in which zinc is employed in either a catalytic or structural role. Finally, Zn^{2+} is widespread in protein molecules because it is kinetically labile i.e. it can diffuse in and out of proteins at a relatively rapid rate (Berg and Shi, 1996). All of the three zinc atoms in the 293K Suc1 structure were tetrahedrally co-ordinated (Endicott *et al.*, 1995). This co-ordination geometry is retained by all three ions in the high resolution, 100K Suc1 structure. The main difference between the two structures as far as zinc co-ordination is concerned is the decrease in the bond lengths. All the shifts that occur in the residues that co-ordinate the zinc atoms are more pronounced in the histidine ligands. This is a consequence of a better-ordered structure at 100K i.e. very little dynamic disorder.

Analysis of a relatively high resolution (2.3Å) cryo-cooled, X-ray structure of *S. pombe* Suc1 confirms the conclusion deduced from the previous model (2.7Å) that the protein structure was a zinc-mediated dimer not a strand-exchanged dimer as seen for the human homologue, CksHs2 (Parge *et al.*, 1993). Extension of the resolution limit has provided a more precise description of the Suc1 structure than had been available previously.

2.4 LITERATURE CITED

Alberts, I.L., Nadassy, K. and Wodak, S.J. (1998) Analysis of zinc binding sites in protein crystal structures. *Protein Sci*, **7**, 1700-1716.

Berg, J.M. and Shi, Y. (1996) The galvanization of biology: a growing appreciation for the roles of zinc. *Science*, **271**, 1081-1085.

Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press, London.

Bourne, Y., Arvai, A.S., Bernstein, S.L., Watson, M.H., Reed, S.I., Endicott, J.E., Noble, M.E., Johnson, L.N. and Tainer, J.A. (1995) Crystal structure of the cell cycle-regulatory

protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci U S A*, **92**, 10232-10236.

Brunger, A.T. (1992) *X-PLOR: Version 3.1; a system for protein crystallography and NMR.* Yale University Press, New Haven.

CCP4. (1994) The CCP4 (Collaborative Computational Project Number 4) suite: programmes for protein crystallography. *Acta Cryst. D*, **D50**, 760-763.

Chayen, N.E., Boggon, T.J., Cassetta, A., Deacon, A., Gleichmann, T., Habash, J., Harrop, S.J., Helliwell, J.R., Nieh, Y.P., Peterson, M.R., Raftery, J., Snell, E.H., Hadener, A., Niemann, A.C., Siddons, D.P., Stojanoff, V., Thompson, A.W., Ursby, T. and Wulff, M. (1996) Trends and challenges in experimental macromolecular crystallography. *Q Rev Biophys*, **29**, 227-278.

Coleman, J.E. (1992) Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu Rev Biochem*, **61**, 897-946.

Endicott, J.A., Noble, M.E., Garman, E.F., Brown, N., Rasmussen, B., Nurse, P. and Johnson, L.N. (1995) The crystal structure of p13^{suc1}, a p34cdc2-interacting cell cycle control protein. *EMBO J.*, **14**, 1004-1014.

Garman, E.F. and Schneider, T.R. (1997) Macromolecular cryocrystallography. *J. Appl. Cryst.*, **30**, 211-237.

Gregoriou, M., Noble, M.E., Watson, K.A., Garman, E.F., Krulle, T.M., de la Fuente, C., Fleet, G.W., Oikonomakos, N.G. and Johnson, L.N. (1998) The structure of a glycogen phosphorylase glucopyranose spirohydantoin complex at 1.8 A resolution and 100 K: the role of the water structure and its contribution to binding. *Protein Sci*, **7**, 915-927.

Harding, M.M. and Cole, S.J. (1963) The crystal structure of di (histidino) zinc pentahydrate. *Acta Crystallographica*, **16**, 643-650.

Hellemans, A. (1994) Europe's New X-ray Source Beams With Confidence. *Science*, Vol. 266, pp. 26 - 27.

Honzatko, R.B., Crawford, J.L., Monaco, H.L., Ladner, J.E., Ewards, B.F., Evans, D.R., Warren, S.G., Wiley, D.C., Ladner, R.C. and Lipscomb, W.N. (1982) Crystal and molecular structures of native and CTP-liganded aspartate carbamoyltransferase from Escherichia coli. *J Mol Biol*, **160**, 219-263.

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

Kretsinger, R.H. and Cotton, F.A. (1963) The crystal and molecular stucture of di-(L-histidino)-zinc (II) dihydrate. *Acta Crystallographica*, **263**, 651-657.

McPherson, A. (1985) Crystallization of macromolecules: general principles. *Methods Enzymol*, **114**, 112-120.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collection Collected in Oscillation Mode. In Carter, C.W. and Sweet, R.M. (eds.), *Methods in Enzymology*. Academic Press, London, Vol. 276, pp. 307 - 325.

Parge, H.E., Arvai, A.S., Murtari, D.J., Reed, S.I. and Tainer, J.A. (1993) Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science*, **262**, 387-394.

Rodgers, D.W. (1994) Cryocrystallography. *Structure*, **2**, 1135-1140.

Tilton, R.F., Dewan, J.C. and Petsko, G.A. (1992) Effects of temperature on protein structure and dynamics: X-ray crystallographic studies on the protein ribonucleas-A at nine different temperature from 98 to 320 K. *Biochemistry*, **31**, 2469-2481.

Vallee, B.L. and Auld, D.S. (1990) Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry*, **29**, 5647-5659.

Vellieux, F.M.D. and Read, R.J. (1997) Noncrystallographic Symmetry Averaging in Phase Refinement and Extension. In Carter, C.W. and Sweet, R.M. (eds.), *Methods in Enzymology*. Academic Press, London, Vol. 276, pp. 18 - 52.

Chapter 3: Structural analysis of *S. pombe* p13^{suc1} folding mutants

CONTENTS

LIST OF FIGURES **85**

LIST OF TABLES **85**

3.1 INTRODUCTION **85**

3.2 MATERIALS, METHODS & RESULTS **88**

3.2.1 Crystallisation of Ala32Gly and Ser13Gly suc1 mutants *88*

3.2.2 Data collection and processing of the Ala32Gly mutant. *89*

3.2.3 Structure determination and refinement of Ala32Gly protein model *91*

3.2.4 Data collection and processing of the Ser13Gly mutant. *93*

3.2.5 Structure determination and refinement of Ser13Gly protein model *95*

3.2.6 Comparative structural analysis of wild-type SpSuc1, Ala32Gly and Ser13Gly mutants *100*

3.3 DISCUSSION **105**

3.4 LITERATURE CITED **105**

List of figures

Figure 3.1 Putative folding pathway of SpSuc1 ---- 87
Figure 3.2 Ramachandran plot for Ala32Gly main chain (N, C_α, C, O) atoms ---- 98
Figure 3.3 Ramachandran plot for Ser13Gly main chain (N, C_α, C, O) atoms ---- 99
Figure 3.4 SpSuc1 structure depicting the two residues that were mutated in Ser13Gly and Ala32Gly, respectively ---- 100
Figure 3.5 Comparative temperature factor plot for SpSuc1 ---- 102
Figure 3.6 Superposition of wild type and mutant SpSuc1 ---- 103
Figure 3.7 Stereo representation of the electron density around Loop31-38 in subunit A *of (a) wild-type SpSuc1 (b) Ala32Gly and (c) Ser13Gly* ---- 104

List of tables

Table 3.1 Summary of data collection strategy for Ala32Gly crystals ____ 90
Table 3.2 Summary of reflections, intensities and R-factors by resolution shells as output by SCALEPACK for the Ala32Gly data set ____ 91
Table 3.4 Summary of data collection strategy for Ser32Gly crystals ____ 94
Table 3.4 Summary of reflections, intensities and R-factors by resolution shells as output by SCALEPACK for the Ser13Gly data set ____ 94
Table 3.5 Refinement statistics for Ala32Gly and Ser13Gly crystals ____ 96

3.1 INTRODUCTION

SpSuc1 exists in two distinct conformational states: a stable, monomeric form and a strand-exchanged dimer (Bourne *et al.*, 1995). The strand-exchanged dimer is formed by the exchange of a terminal β-strand between two monomeric molecules. Due to this feature, SpSuc1 has been added to the family of proteins called domain-swapping proteins (Bennett *et al.*, 1994; Bennett *et al.*, 1995; Schlunegger *et al.*, 1997). Although the functional significance of this conformational duality is as yet unknown, this intriguing feature has piqued the interest of researchers working in the field of protein folding. They have employed the powerful technique of protein engineering to make SpSuc1 mutants that could be used to probe the folding pathway of the two

conformational species of SpSuc1. The current state of knowledge regarding the folding of SpSuc1 is presented below.

Under native conditions, about 95% of SpSuc1 molecules exist in the monomeric form (Birck *et al.*, 1995; Birck *et al.*, 1996). The equilibration rate between the monomer and strand-exchanged dimer is relatively slow and it can take a few hours to several days depending on the concentration of the protein (Rousseau *et al.*, 1998). It has been shown that the rate of the monomer-dimer equilibration can be increased if SpSuc1 is equilibrated in low concentrations of a chemical denaturant. Dimerisation of SpSuc1 is enhanced when a high concentration (>1mM) of the protein is denatured in a solution of 8M Urea and then refolded by dialysis (Dr. Endicott, University of Oxford, personal communication) Under such conditions, about 30% of the SpSuc1 molecules end up in the strand-exchanged dimeric conformation. Dimerisation of SpSuc1 can also be achieved by heating the protein to a temperature above its melting point before slowly cooling it to room temperature (Dr. Itzhaki, University of Cambridge, personal communication). The folding pathway of SpSuc1 is via a partly folded intermediate state. Analysis of this intermediate state has revealed that it is an expanded form of the native, monomeric SpSuc1. It possesses the majority of the interactions present in the native, monomeric SpSuc1 albeit, in a relatively weakened state (Dr. Itzhaki, University of Cambridge, personal communication). The nucleation-condensation model of protein folding suggests that the folding nucleus (around which the rest of the structure condenses) corresponds to the region of the protein that is most highly structured in the transition state (Itzhaki *et al.*, 1995). In the case of SpSuc1, the folding nucleus consists of the central β-strands i.e. β2 and β4. Ongoing studies in the laboratory of Dr. Itzaki at the University of Cambridge, are beginning to suggest that the folding pathway of SpSuc1 begins to diverge into two independent (monomeric and dimeric) pathways after the formation of the expanded folding intermediate (fig. 3.1).

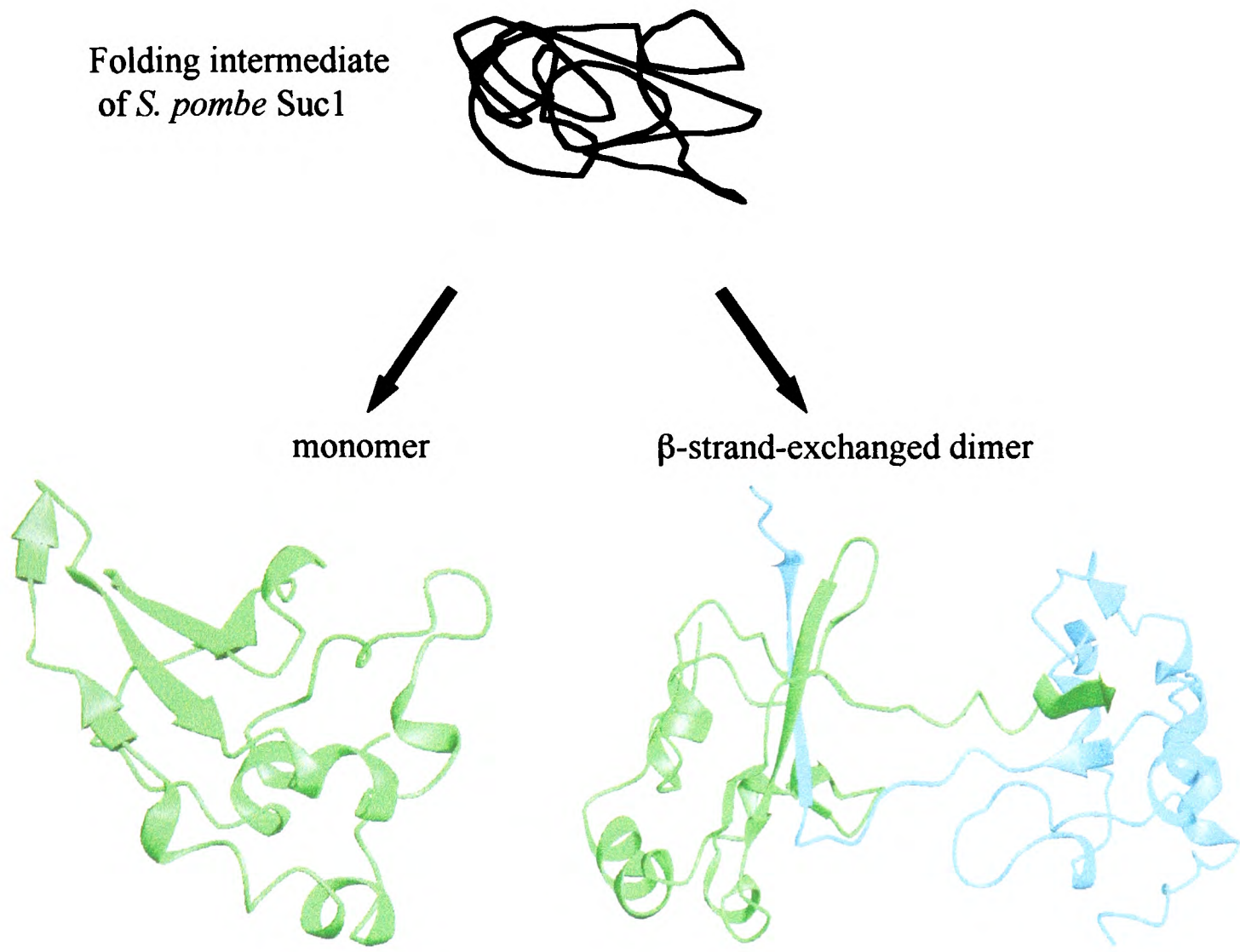


Figure 3.1 Putative folding pathway of SpSuc1

Under native conditions, it is unlikely that a properly folded SpSuc1 monomer would spontaneously convert to a β-strand-exchanged dimer because of the high energetic barrier that must be traversed. This model explains the experimental observations mentioned above i.e. the observation that the formation of the dimeric species of SpSuc1 in vitro, is enhanced when a correctly folded monomer is (a) treated with a mild chemical denaturant (e.g. 1-2M Urea) and left to stand for several hours or days at 4°C (b) denatured in 8M Urea and renatured by dialysis (c) heated above the melting temperature (T_m) and allowed to cool slowly to room temperature and (d) maintaining the protein at the non-physiological pH of 5.5 for a few hours at 4°C. It appears that under these experimental conditions, the SpSuc1 monomer reverts to a partially destabilised conformation that is similar to the expanded folding intermediate. It is plausible that in vitro, the half-life of the intermediate is greater than that in vivo, hence explaining the higher percentage of dimers produced under the former conditions.

It has been observed that mutation of Pro90 to Ala or Gly completely abolished the ability of SpSuc1 to dimerise (Rousseau *et al.*, 1998). This proline residue is part of the hinge region (HVPEP). Other domain-swapping proteins in which the swapping of secondary structure elements occur via similar arm-opening mechanisms possess single or multiple proline residues in the critical hinge region (Bennett *et al.*, 1994; Bennett *et al.*, 1995; Schlunegger *et al.*, 1997). It has been speculated that the presence of proline residues may confer rigidity on the backbone in the hinge region, thus allowing the extension of this region during the process of β-strand-exchanged dimer formation (Dr. Itzhaki, University of Cambridge, personal communication). In the case of SpSuc1, it has been observed that mutation of Pro92 - the second proline in the repeat – does not impair dimerisation (Dr. Itzhaki, University of Cambridge, personal communication). This suggests that the particular location of the proline is a critical factor. A common assumption underpinning protein-engineering analysis of folding pathways is that the point mutations have no adverse effect(s) on either the global structure of the protein or the folding pathway. Wherever possible, it is necessary to test the validity of this assumption by determining the atomic structures of these folding mutants, especially those that appear to have different biophysical properties from the wild-type protein. To this end, Dr. Itzhaki prepared and sent several milligrams of purified suc1 mutants to Dr. Endicott for structural elucidation by X-ray crystallography. Dr. Endicott crystallised Ala32Gly and Ser13Gly mutants. The folding pathway of these two mutants is identical to that of the wild-type protein. An X-ray data set for Ala32Gly was collected by Drs Noble, Endicott and Garman. The author of the thesis took part in the collection of Ser13Gly data set. An account of the processing and a comparative structural analysis is presented below.

3.2 MATERIALS, METHODS & RESULTS

3.2.1 Crystallisation of Ala32Gly and Ser13Gly suc1 mutants

Both mutants were crystallised under conditions similar to that of the wild-type SpSuc1 crystals, by the hanging-drop vapour diffusion method (McPherson, 1985). In the case of Ala32Gly, an initial screen for novel crystallisation conditions using the

reagents in crystal screen I did not prove successful. However, when each of the solutions in the kit was supplemented with 1.5mM of zinc sulphate, a cluster of single crystals grew out of solution number 25 (0.1M imidazole pH 6.5 and 1.0M sodium acetate) of crystal screen I (Hampton Research). Briefly, 1µl of the protein stock of 20mg/ml was mixed with 1µl of the mother liquor (0.1M imidazole pH 6.5, 1.0M Sodium acetate trihydrate and 1.5mM Zinc sulphate). A cluster of single Ala32Gly crystals grew out of this solution after about 16 hours incubation at 21°C.

For the Ser13Gly mutant, 1µl of the protein stock (20mg/ml) was mixed with 1µl of the mother liquor (0.1M MES pH 6.4, 2-2.5% PEG 8000, 200-250mM ammonium acetate and 1.5mM Zinc sulphate). This also yielded a cluster of single crystals after an overnight incubation at 21°C.

3.2.2 Data collection and processing of the Ala32Gly mutant.

Data were collected on a 34.5cm diameter MAR Research image plate at the X-ray diffraction beamline at the ELETTRA Synchrotron Radiation Source (Trieste, Italy) in May 1998. A single crystal was soaked for 2 minutes in a cryo-solution (30-35% glycerol in 0.1M imidazole pH 6.5, 1.0M Sodium acetate trihydrate and 1.5mM Zinc sulphate) and flash-cooled to 100K in a stream of gaseous nitrogen. The data collection strategy was similar to that described for the wild type SpSuc1 crystals in the preceding chapter. Briefly, a single diffraction image was collected and auto-indexed with DENZO (Otwinowski and Minor, 1997). The unit cell dimensions and orientation of the crystal with respect to the beam were used as the input parameters for PREDICT (M.E.M Noble, unpublished work) in order to obtain the necessary data collection parameters such as the starting phi (ϕ) angle and the total ϕ range necessary for the collection of a complete data set. A summary of the data collection strategy is presented in table 3.1.

	High resolution data set	Low resolution data set
Maximum resolution	2.0Å	2.18Å
Crystal-to-detector distance	370mm	400mm
Number of diffraction images	18	49
Oscillation range (Δϕ)/image	0.5°	0.5°
Total ϕ range	9°	24.5°
Exposure time	120s	120s

Table 3.1 Summary of data collection strategy for Ala32Gly crystals

The X-ray wavelength was 0.88Å. The crystal belonged to spacegroup $P6_522$ with unit cell dimensions a=b=59.3Å and c=264.4Å and there was one dimer molecule of Ala32Gly molecule per asymmetric unit.

The data were integrated with DENZO (Otwinowski and Minor, 1997) and scaled with SCALEPACK (Otwinowski and Minor, 1997). The mosaicity was 0.28 as compared to 0.29 for the wild type, 100K SpSuc1 crystals. The measured intensities were converted to structure factors with the program TRUNCATE (CCP4, 1994). About 45 reflections from the scaled data were too weak to be converted to structure factor amplitudes by TRUNCATE.

Resolution shell (Å)	Reflections[¶]	<I/σ(I)>	%Completeness[¶]	R_{merge}*	Multiplicity
50.00-4.60	1773	27.56	93.0	0.035	3.44
4.60-3.65	1703	19.94	98.3	0.037	3.65
3.65-3.19	1648	14.91	98.3	0.045	3.81
3.19-2.90	1631	10.75	98.3	0.064	3.80
2.90-2.69	1596	7.08	98.3	0.101	3.78
2.69-2.53	1606	6.34	98.2	0.132	3.80
2.53-2.40	1586	4.61	98.0	0.197	3.78
2.40-2.30	1411	3.58	87.9	0.231	3.69
50.00-2.30	12954	16.84	96.3	0.048	3.72

Table 3.2 Summary of reflections, intensities and R-factors by resolution shells as output by SCALEPACK for the Ala32Gly data set

The high resolution cut off for the Ala32Gly data set was 2.3Å, although the crystals diffracted to as high a resolution as 2.0Å. Data beyond 2.30Å were excluded because they merged poorly with the rest of the data.

[¶]*Number of unique reflections.*

¶*Percentage of theoretically possible data*

$$*R_{merge} = \frac{\sum_i \sum_h (I_i(h) - \langle I(h) \rangle)}{\sum_i \sum_h I(h)}$$

where $I_i(h)$ and $\langle I(h) \rangle$ are the ***ith*** *measurement and the average intensity of reflection* ***h****.*

3.2.3 Structure determination and refinement of Ala32Gly protein model

The phasing model for the Ala32Gly data set was the 2.3Å wild-type SpSuc1 structure. The water molecules were excluded from the co-ordinate file of the latter. The unit cell dimensions of the 100K SpSuc1 data (a=b=60.1Å and c=264.5Å) were nearly

isomorphous with that of Ala32Gly a=b=59.3Å and c=264.4Å. There was a reduction in the unit cell dimensions by 0.8Å in the ***a*** and ***b*** axes and 0.1Å in the ***c*** axis. The initial R-values were 47.2% and 46.6% for the R_{work} and R_{free}, respectively. A round of rigid body refinement with a subsequent refinement of the temperature factors using REFMAC (Murshudov *et al.*, 1997) reduced the R_{work} value to 29.5% and the R_{free} to 33.8%. An electron density map (2Fo-Fc contoured at the nominal electron density of 0.11 e/Å^3 which corresponded to the 1σ level calculated for the map) was calculated and the position of each amino acid in the electron density was carefully examined with the graphics program O (Jones, 1991). It was noticed that the position of the loop region (consisting of residues 31-38) in the electron density, differed in the two subunits of this nascent model. The fit of the loop to the electron density was good in subunit ***B*** but out of place in subunit ***A***. A round of simulated annealing refinement (with an initial temperature of 2000K) was carried out on the Ala32Gly model using a developmental version of the Crystallography & NMR (CNS v0.4) software program (Brunger *et al.*, 1998). In order to find the appropriate value for the weights to be applied to the crystallographic terms, the CNS program scanned several different values and then recommended the value that gave the greatest reduction in the R_{free} value. A 2Fo-Fc map was calculated and the model was carefully examined again. Apart from Loop31-38, the rest of the model coincided well with the corresponding electron density. The R_{work} was 32.1% and the R_{free} was 36.0%. There was a lack of a supportable electron density for the first five N-terminal and the last 12 C-terminal residues. The part of the structure corresponding to Loop31-38 was omitted from the model and the new co-ordinates of the Ala32Gly model were put through an energy minimisation refinement step in CNS. The R-values at the end of this process were 31.2% for R_{work} and 35.1% for the R_{free} value. A new 2Fo-Fc 'omit' map was calculated and the deleted loop region was rebuilt into an improved electron density for that region with the aid of the O graphics program. The alanine to glycine mutation was effected in both subunits and the three zinc atoms were placed in their corresponding electron densities. The new co-ordinates were written out and the weights were updated in CNS. Another energy minimisation step was carried out using the updated weight value. This gave an R_{work} of 26.8% and an R_{free} of 32.7%. An

inspection of the Ala32Gly model revealed that Loop31-38 had as good a fit in the corresponding electron density in subunit ***A*** as it did in subunit ***B***. A bulk solvent correction was applied and water molecules were added using the facility in CNS to search through an Fo-Fc map for peaks in the electron density. The following rejection criteria were applied to the water peaks; the minimum and maximum distances between water and any atom was set 2.6Å and 4.0Å, respectively. Also, the minimum hydrogen bonding distance between water and an oxygen or nitrogen atom was set to 2.0Å. This was followed by a restrained, B-factor refinement of all the atoms in the Ala32Gly model. About 175 well-defined water molecules were added. The final R_{work} was 23.1% and the R_{free} was 29.0%.

3.2.4 Data collection and processing of the Ser13Gly mutant.

Data were collected on a 30cm diameter MAR Research image plate at beamline 9.6 of the Synchrotron Radiation Source (Daresbury, U.K.) in May 1998. A single crystal was soaked for 2 minutes in a cryo-solution (30% glycerol in 0.1M MES pH 6.4, 2-2.5% PEG 8000, 200-250mM ammonium acetate and 1.5mM Zinc sulphate). The crystal was then scooped with the aid of a cryo-loop (Hampton Research) and flash-cooled to 100K in a stream of gaseous nitrogen. The data collection strategy differed from that described for both the wild type SpSuc1 and the Ala32Gly crystals: severe time constraints as a result of the instability of the X-ray beam meant that only one data set instead of two (a high and low resolution) were collected. A summary of the data collection strategy is presented in table 3.4

	High resolution data set
Maximum resolution	2.24Å
Crystal-to-detector distance	350mm
Number of diffraction images	16
Oscillation range ($\Delta\phi$)	1.5°
Total ϕ range	13.5°
Exposure time	600s
Wavelength	0.9Å
Spacegroup	$P6_522$
Unit cell dimensions	a=b=60.0Å and c=261.1Å

Table 3.4 Summary of data collection strategy for Ser32Gly crystals

Resolution shell (Å)	Reflections[¶]	<I/σ(I)>	%Completeness[¶]	R_{merge}*	Multiplicity
50.00-4.50	1871	22.83	92.6	0.043	2.55
4.50-3.57	1725	21.79	94.7	0.055	2.82
3.57-3.12	1683	10.28	94.8	0.114	2.87
3.12-2.83	1656	4.93	94.2	0.231	2.84
2.83-2.63	1621	2.65	93.2	0.434	2.96
2.63-2.48	1612	1.78	93.7	0.627	2.89
2.48-2.35	1633	1.09	94.8	0.986	2.70
50.00-2.35	13274	10.56	92.9	0.099	2.75

Table 3.4 Summary of reflections, intensities and R-factors by resolution shells as output by SCALEPACK for the Ser13Gly data set

Although the crystals diffracted to 2.24Å and the 2.35-2.25Å resolution shell contained 1473 reflections, the high resolution cut off for the Ser13Gly data set was 2.35Å because the error in the measurement of the Intensity (I) was too large for the 2.35-2.25Å resolution shell and the error in the measured intensity (I/σ(I)) was less than 1 (data not shown).

¶Number of unique reflections.

¶Percentage of theoretically possible data

$$*R_{merge} = \frac{\sum_i \sum_h (I_i(h) - \langle I(h) \rangle)}{\sum_i \sum_h I(h)}$$

*where $I_i(h)$ and $\langle I(h) \rangle$ are the **ith** measurement and the average intensity of reflection **h**.*

3.2.5 Structure determination and refinement of Ser13Gly protein model

The co-ordinates of the 2.3Å SpSuc1 wild-type structure minus the water molecules were used to calculate the initial phases for the Ser13Gly data set. The unit cell dimensions of the Ser13Gly crystals were a=b=60.1Å and c=264.1Å which is nearly isomorphous to the native 100K SpSuc1 crystals (chapter 2). The initial R-values were 45.3% for R_{work} and 44.8% for the R_{free}. Rigid body refinement followed by temperature factor refinement in REFMAC (Murshudov *et al.*, 1997) reduced the R_{work} value to 30.1% and the R_{free} to 35.1%. Examination of the nascent model using a 2Fo-Fc electron density map contoured at the nominal electron density of 0.12 e/Å^3, which corresponded to the 1σ level calculated for the map revealed that most regions of the protein model (including Loop31-38) were supported by relatively good electron density. Weight optimisation was followed by a round of simulated annealing with a starting temperature of 2000K. An improved 2Fo-Fc map was calculated and Ser13 was changed to a glycine residue using the facility in the graphics program O (Jones, 1991). The three zinc atoms were also positioned in their respective electron densities. The new model was put through a round of energy minimisation before water molecules were added. Application of a bulk solvent correction was followed by a round of restrained B-factor refinement.

The final R_{work} was 22.9% and R_{free} was 30.1%. A summary of the refinement statistics for both Ala32Gly and Ser13Gly is presented in Table 3.5.

REFINEMENT	Ala32Gly	Ser13Gly
Resolution Range (Å)	20 – 2.3Å	20 – 2.35Å
Total number of reflections used	12909	12492
Number of reflections in working set	11757	11390
Number of reflections in test set for R_{free}	1152	1102
Final $^{\delta}R_{work}$ (R_{free})	23.1% (29.0%)	22.9% (30.1%)
Unit cell volume occupied by solvent	48.1%	47.9%
	NUMBER OF ATOMS	
Protein	1628	1628
Zinc	3	3
Waters	175	176
Average B-factors of protein atoms, zinc ions and water molecules, respectively	48.8Å^2, 57.6Å^2, 62.4Å^2	46.9Å^2, 41.0Å^2, 46.8Å^2
RMS deviation from ideal geometry		
Bond lengths →	0.0205Å	0.010Å
Bond angles →	2.2°	1.7°
Improper angles →	1.4°	1.21°

Table 3.5 Refinement statistics for Ala32Gly and Ser13Gly crystals

$^{\delta}$Crystallographic $R=\sum_h|F_o-F_c|/\sum_h F_c$ where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively. R_{free} is calculated as for R_{work} but using about 10% of the data set that had been excluded from the refinement. The average B-factors for the two mutant SpSuc1 structures (48.8Å^2 for Ala32Gly and 46.9Å^2 Ser13Gly) are the same as the temperature factors estimated from a Wilson plot of the intensities (48.8Å^2 for Ala32Gly and 46.9Å^2 Ser13Gly). These average B-factors (obtained from refinement) are higher than those observed for the 100K wild-type structure where the average B-factor was 22.9Å^2. The Wilson plot for the 100K wild-type structure gave a temperature of 48.4Å^2 for the intensities. The possibility that the differences in refined B factors between the wild-type and the mutants had arisen from

the use of different refinement programs (REFMAC versus CNS) was investigated by refining the wild-type structure with CNS with a protocol that included a simulated annealing, energy minimisation and B-factor refinement and the overall B-factor was 23.4Å^2, which suggests that this is the value acceptable by the refinement process. In an attempt to see whether these values had become trapped in a false minimum, all the B-factors were reset to 50Å^3 and the structure was again put through a round of simulated annealing, minimisation and B-factor refinement in CNS. The resulting average B factor was 21.3. Hence the values obtained by different refinement protocols (REFMAC and CNS) are consistent to within the errors in the refinement. The reason for the large discrepancy between the values from the Wilson plot and the refinement average B-factor for the wild-type 100K was not resolved. The differences between the wild-type and the mutant structures average B-factor values could have arisen from different crystal quality or differences arising from cryo-protection and cooling.

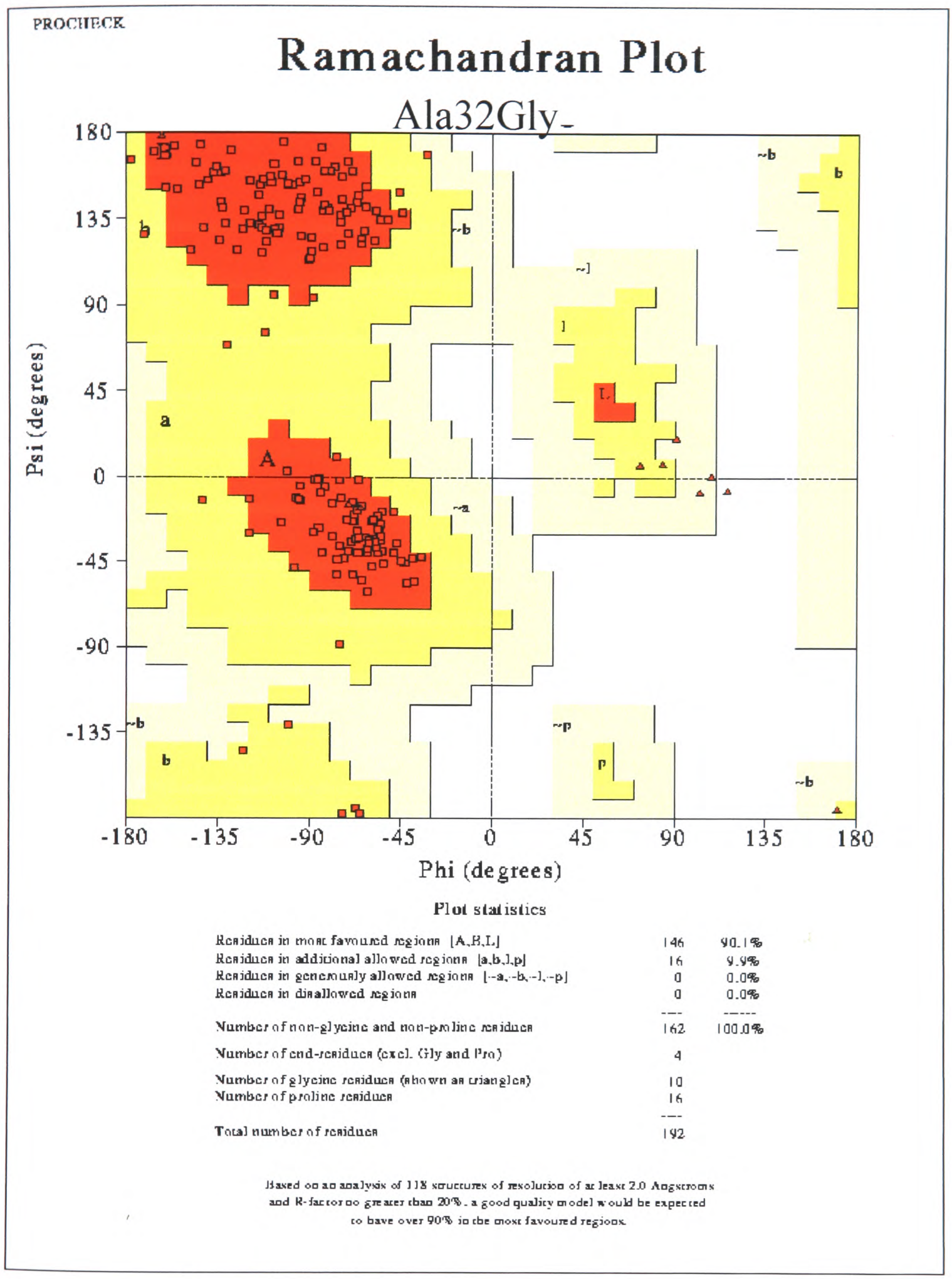


Figure 3.2 Ramachandran plot for Ala32Gly main chain (N, C_α, C, O) atoms

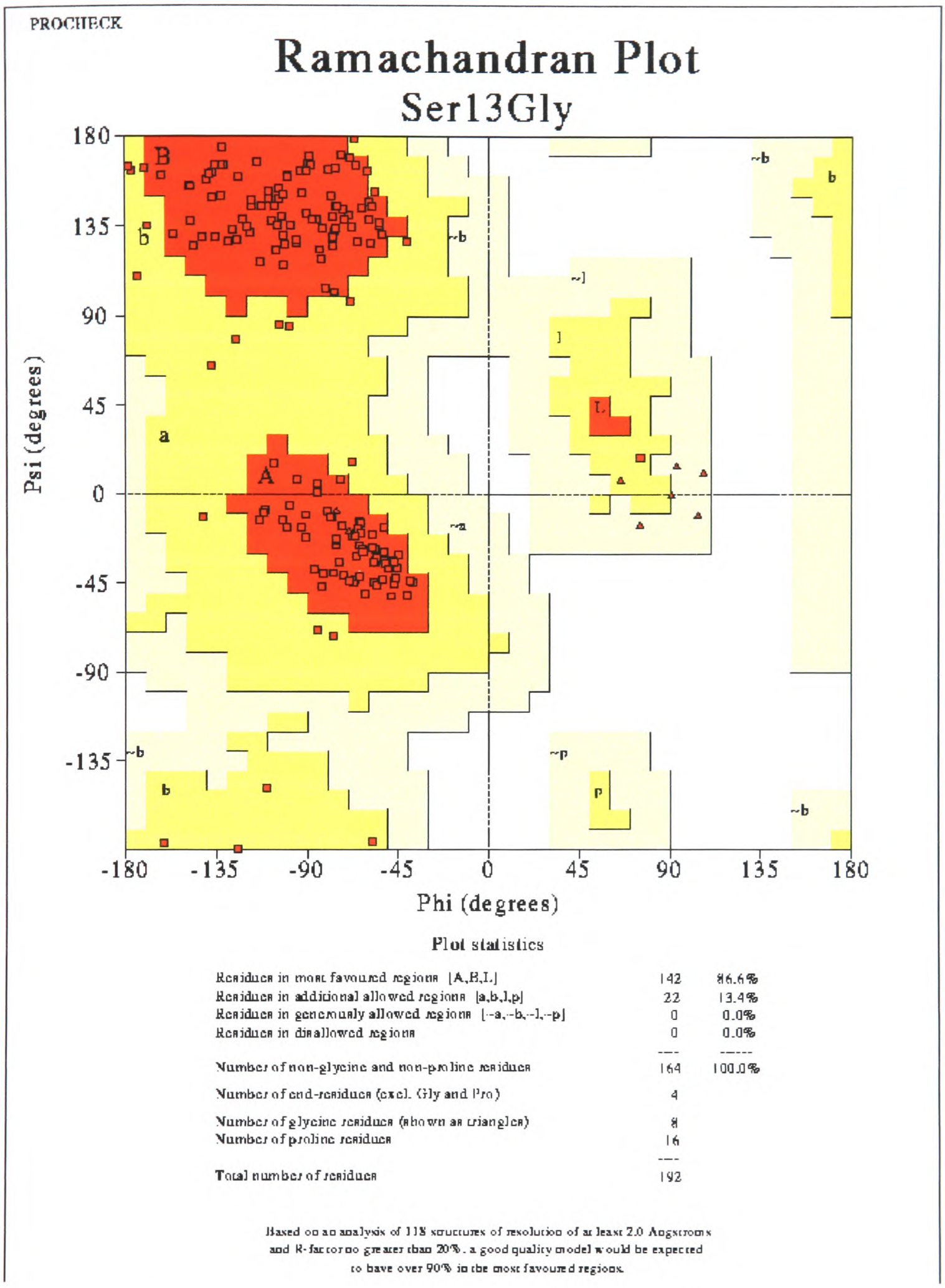


Figure 3.3 Ramachandran plot for Ser13Gly main chain (N, C_α, C, O) atoms

3.2.6 Comparative structural analysis of wild-type SpSuc1, Ala32Gly and Ser13Gly mutants

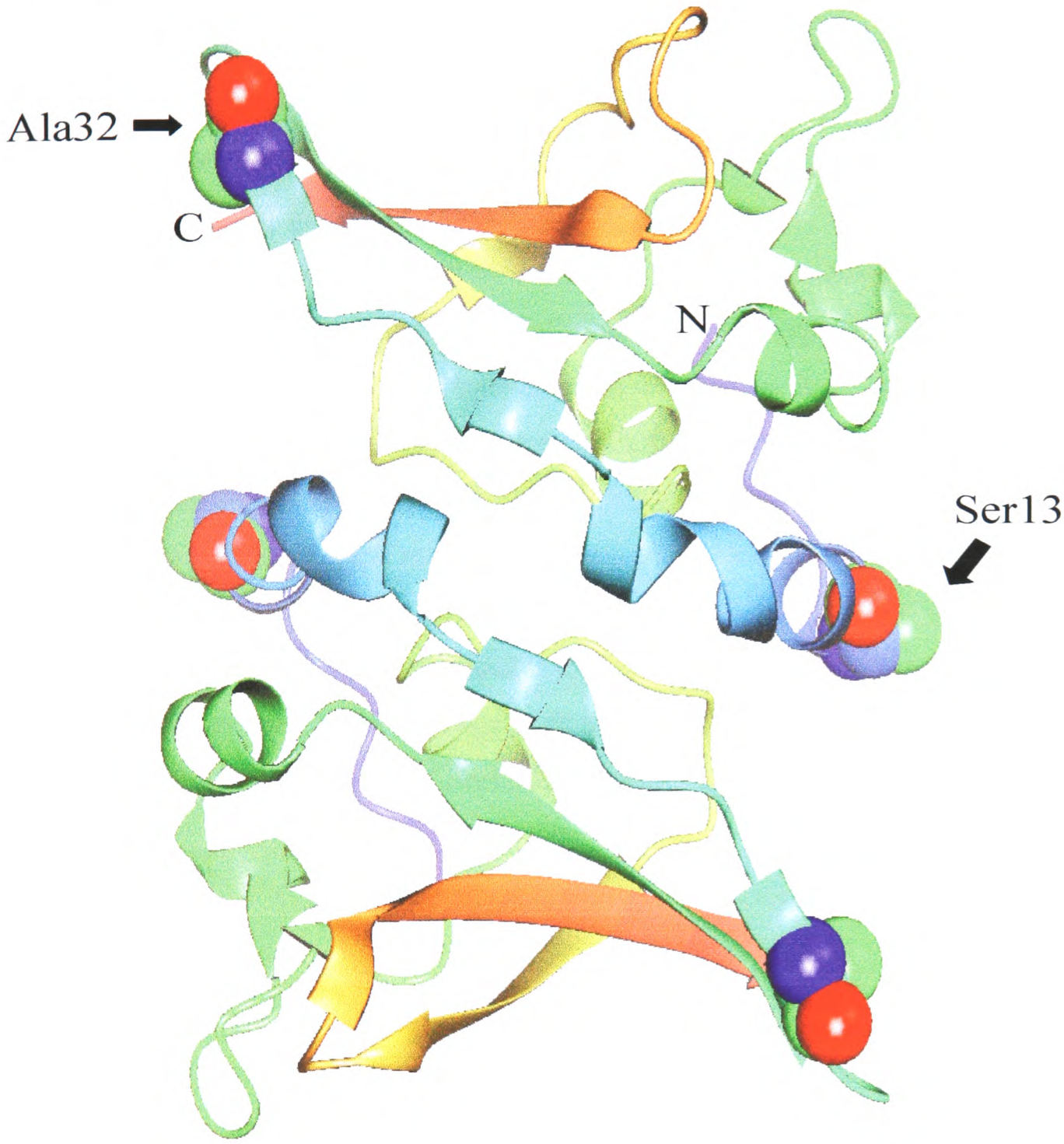


Figure 3.4 SpSuc1 structure depicting the two residues that were mutated in Ser13Gly and Ala32Gly, respectively

The colouring is based on the position of each residue in the sequence; starting with blue for N-terminal residues and ending with orange for the C-terminal residues. Ser13 and Gly32 are drawn as space-filling models with carbon atom coloured green, oxygen red

and nitrogen blue. Figures 3.4 and 3.5 were produced with Aesop (Dr. Noble, unpublished work).

In general, the three protein structures i.e. 100K wild-type SpSuc1, Ala32Gly and Ser13Gly have identical global structures. There was no electron density for the first 5 N-terminal and the last 12 C-terminal residues. Superposition of the Cα atoms of Ala32Gly on wild-type SpSuc1 gave an RMSD of 0.37Å, and Ser13Gly on wild-type SpSuc1 gave an RMSD of 0.31Å, whilst Ala32Gly on Ser13Gly gave an RMSD of 0.42Å suggesting that all three structures had similar overall folds. A similar operation on the two subunits of each molecule suggested that inter-subunit differences were more pronounced. The RMSD for superimposing subunit ***A*** on subunit ***B*** for the Ala32Gly structure was 0.53Å and that for the Ser13Gly was 0.59Å. The corresponding value for wild-type SpSuc1 was 0.60Å. Examination of each structure revealed that the inter-subunit differences were more noticeable around Loop31-38. Superimposition of loopA31-38 on LoopB31-38 gave an RMSD of 0.53Å, 0.51Å and 0.43Å for the wild-type SpSuc1, Ser13Gly and Ala32Gly, respectively. The average B-factor was 22.9Å^2, 48.8Å^2 and 46.9Å^2 for the wild-type SpSuc1, Ala32Gly and Ser13Gly structures, respectively.

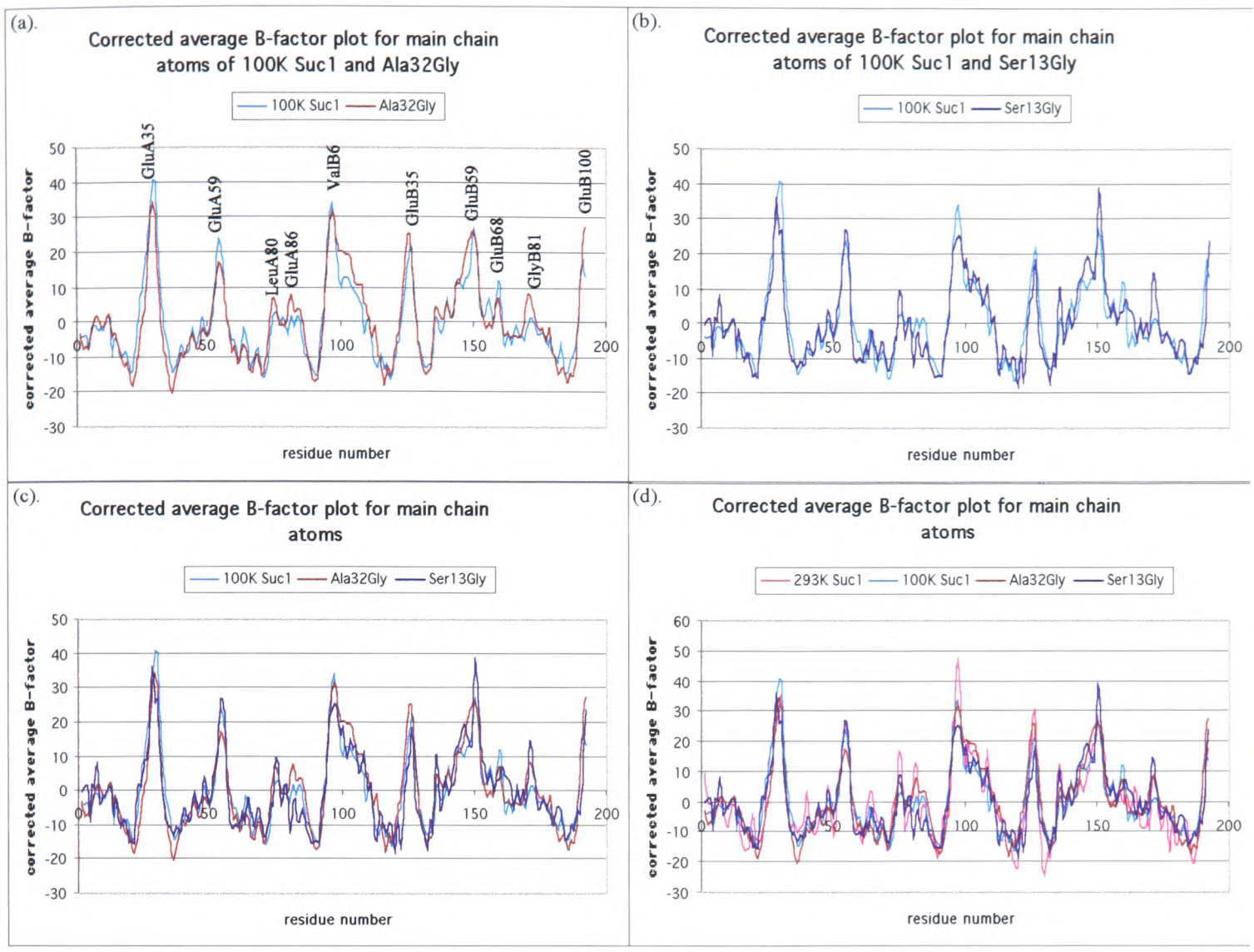


Figure 3.5 Comparative temperature factor plot for SpSuc1

Stereochemically (as judged from the Ramachandran plots), Ala32Gly and wild-type SpSuc1 were relatively better models than Ser13Gly. In the latter two, more than 90% of the residues were in the most favoured regions of the Ramachandran plot. The equivalent figure for Ser13Gly was 87%. This is a reflection of the quality of the respective X-ray data. The co-ordination geometry of the zinc atoms in all three structures was tetrahedral. The number of water molecules in the mutant structures (176 for Ala32Gly and 175 for Ser13Gly) is consistent with the value obtained for the wild-type SpSuc1 structure of comparable atomic resolution.

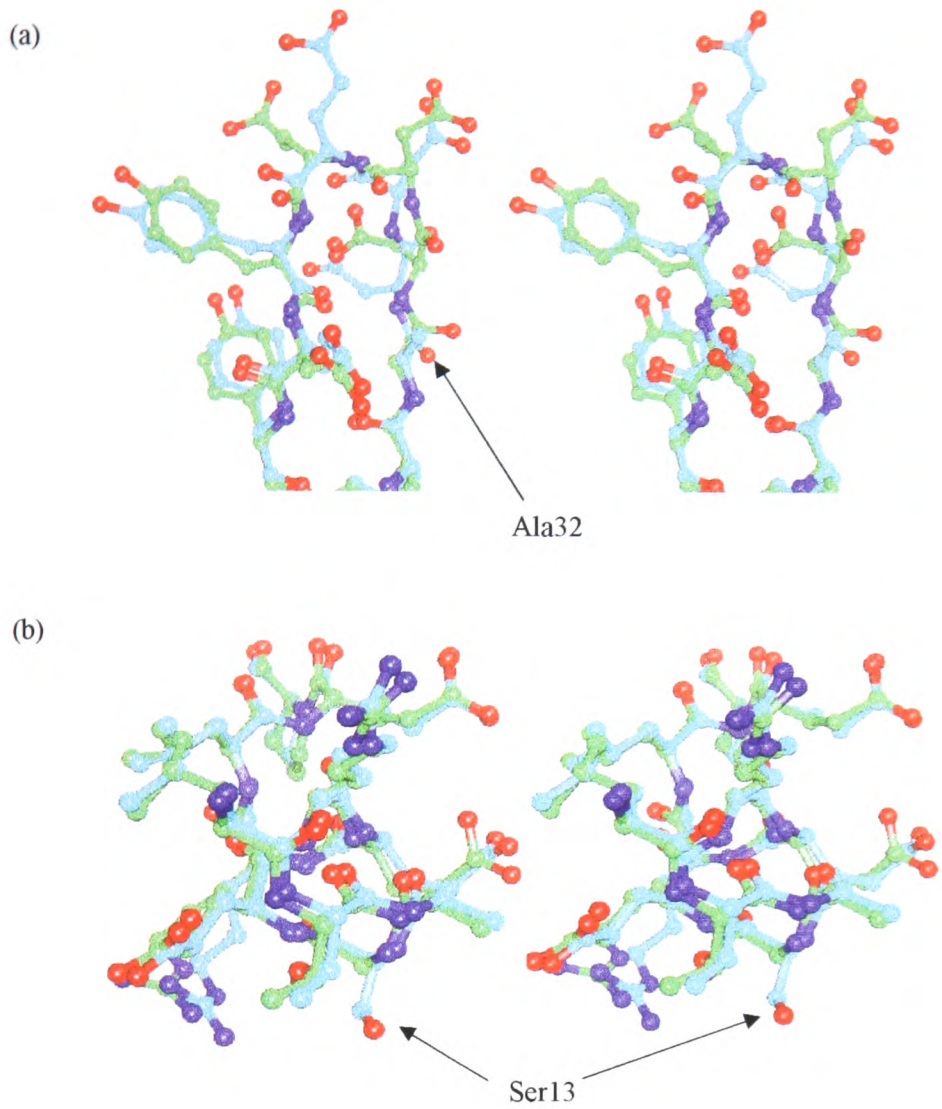


Figure 3.6 Superimposition of wild type and mutant SpSuc1

*(a) Wild type SpSuc1 is coloured cyan and the Ala36Gly mutant is green. Superimposition of the two loop regions (residues 31 to 38) from subunit **A** gave an r.m.s deviation of 0.71Å.*

*(b) Wild type SpSuc1 is coloured cyan and the Ser13Gly mutant is green. Superimposition of the two helices (residues 10 to 22) from subunit **A** gave an r.m.s deviation of 0.15Å.*

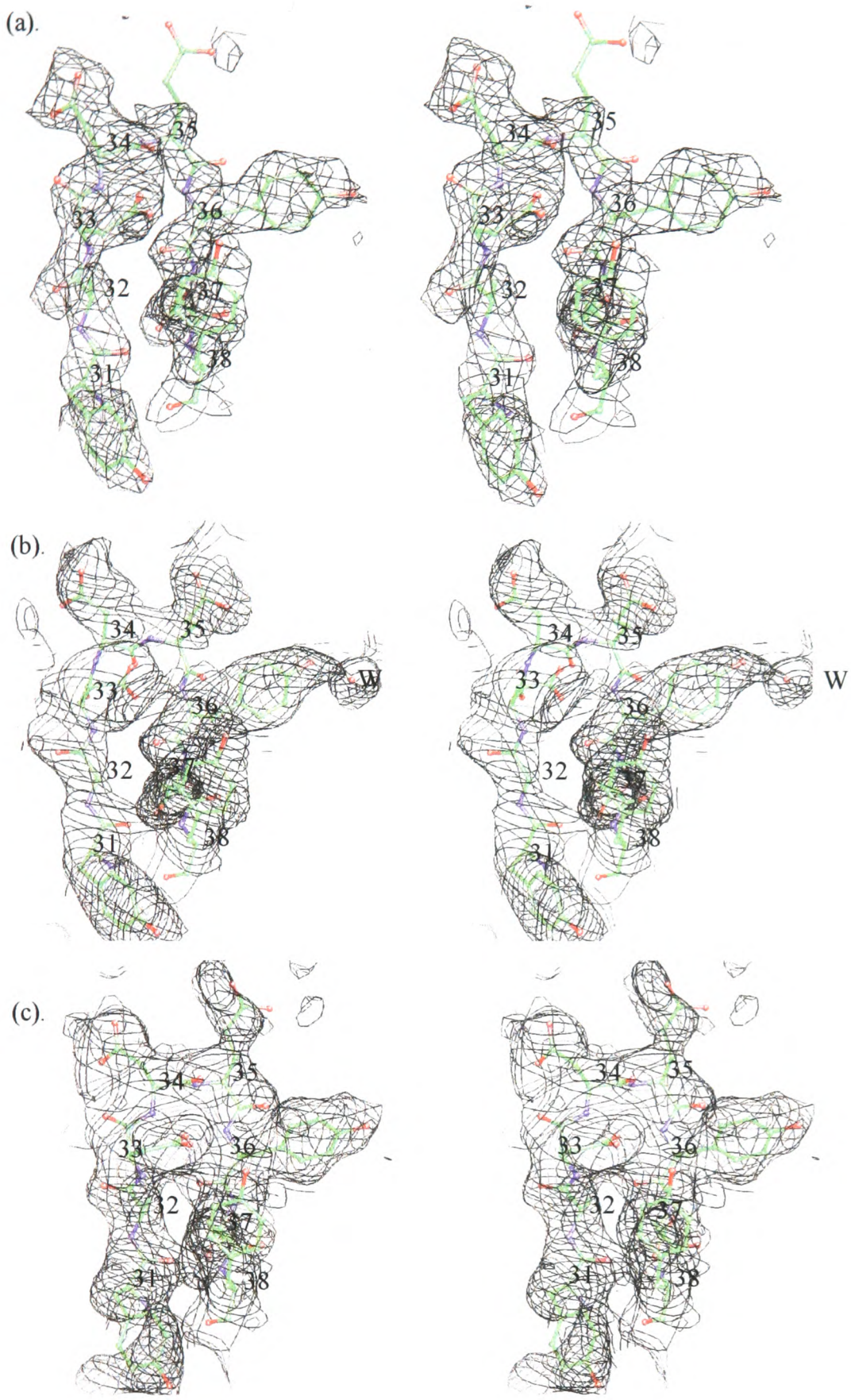


Figure 3.7 Stereo representation of the electron density around Loop31-38 in subunit A of (a) wild-type SpSuc1 (b) Ala32Gly and (c) Ser13Gly

All the 2Fo-Fc maps were contoured at a nominal electron density of 0.81-0.83 e/Å^3, which corresponds to the 2.8σ level calculated for each map.

3.3 DISCUSSION

In many respects, the mutation of Ala32 and Ser13 to glycine has not altered the tertiary structure of SpSuc1. This is not a surprising finding when we consider the location and apparent role of these residues in the wild-type SpSuc1 structure. Ser13 is part of a three-turn α-helix (α1a) consisting of residues 12-22. Ala32 is located on a loop region between β1a and β1b. Both residues are located on the surface of the molecule. The folding pathways of both mutants are indistinguishable from that of the wild type. Both α1a and Loop31-38 are formed late in the folding pathway. The earliest secondary structural elements to be formed constitute the core of the expanded, folding intermediate. These are β2 and β4. Although, the three structures have identical conformations, it appears that mutating Ala32 to glycine has reduced the mobility of this solvent-accessible, loop region in subunit *A*. Detailed examination of the electron density map around this loop region shows that the Ala32Gly mutation has resulted in a better ordering of this loop region. This is also reflected in the RMSD and B-factor values. Superimposition of Cα atoms of LoopA31-38 on LoopB31-38 (RMSD of 0.53Å and 0.43Å for wild-type SpSuc1 and Alal32Gly, respectively) reveals that there was a 19% reduction in the inter-subunit differences pertaining to this region, as a result of the single point mutation. The results presented in this chapter confer a degree of validity on the common assumption in protein-engineering analysis of folding pathways that, in general, point mutations of surface residues do not have adverse effect(s) on either the global structure of the protein or the folding pathway. However, point mutations may cause some disordering or ordering in a particular local region of the protein molecule.

3.4 LITERATURE CITED

Bennett, M.J., Choe, S. and Eisenberg, D. (1994) Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA*, **91**, 3127-3131.

Bennett, M.J., Schlunegger, M.P. and Eisenberg, D. (1995) 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci*, **4**, 2455-2468.

Birck, C., Raynaud-Messina, B. and Samama, J.P. (1995) Oligomerization state in solution of the cell cycle regulators p13suc1 from the fission yeast and p9cksphy from the myxomycete Physarum, two members of the cks family. *FEBS Lett*, **363**, 145-150.

Birck, C., Vachette, P., Welch, M., Swaren, P. and Samama, J.P. (1996) Is the function of the cdc2 kinase subunit proteins tuned by their propensities to oligomerize? Conformational states in solution of the cdc2 kinase partners p13suc1 and p9cksphy. *Biochemistry*, **35**, 5577-5585.

Bourne, Y., Arvai, A.S., Bernstein, S.L., Watson, M.H., Reed, S.I., Endicott, J.E., Noble, M.E., Johnson, L.N. and Tainer, J.A. (1995) Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci U S A*, **92**, 10232-10236.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

CCP4. (1994) The CCP4 (Collaborative Computational Project Number 4) suite: programmes for protein crystallography. *Acta Cryst. D*, **D50**, 760-763.

Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol*, **254**, 260-288.

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

McPherson, A. (1985) Crystallization of macromolecules: general principles. *Methods Enzymol*, **114**, 112-120.

Murshudov, G.N., Vagen, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum -likelihood method. *Acta Cryst.*, **D53**, 240-255.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collection Collected in Oscillation Mode. In Carter, C.W. and Sweet, R.M. (eds.), *Methods in Enzymology*. Academic Press, London, Vol. 276, pp. 307 - 325.

Rousseau, F., Schymkowitz, J.W., Sanchez del Pino, M. and Itzhaki, L.S. (1998) Stability and folding of the cell cycle regulatory protein, p13(suc1). *J Mol Biol*, **284**, 503-519.

Schlunegger, M.P., Bennett, M.J. and Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem*, **50**, 61-122.

Chapter 4: Mutational analysis of *S. pombe* p13^{suc1}

CONTENTS

LIST OF FIGURES **110**

LIST OF TABLES **111**

4.1 INTRODUCTION **111**

4.2 MATERIALS & METHODS **115**

4.2.1 Escherichia coli host strains and plasmid vectors *115*
4.2.2 Schizosaccharomyces pombe host strains and plasmid vectors *115*
4.2.3 Site-specific mutagenesis *115*
4.2.4 Growth of uracil-containing phagemid *116*
4.2.4 Titration of phagemid stock *117*
4.2.5 Extraction of phagemid DNA *117*
4.2.6 Synthesis of mutagenic strand *118*
4.2.7 Purification of oligonucleotide primers *119*
4.2.8 Annealing of mutagenic primer to template *119*
4.2.9 Synthesis of complementary strand *119*
4.2.10 Gel analysis of reaction products *120*
4.2.11 Preparation of competent E. coli cells by the Calcium Chloride method *120*

4.2.12 Transformation of reaction products *120*
4.2.13(a) Screening for mutants by dideoxy DNA sequencing *120*
4.2.13(b) Annealing reaction *121*
4.2.13(c) Labelling reaction *121*
4.2.13(d) Termination reactions *121*
4.2.14 Subcloning mutant SpSuc1 genes into pRK172 vector *122*
4.2.15 Ligation reaction *122*
4.2.16 Expression and purification of mutant SpSuc1 ..... *123*
4.2.17(a) Characterisation of mutant SpSuc1 protein *124*
4.2.17(b) Electrospray mass spectrometry *124*
4.2.17(c) Circular dichroism spectroscopy *124*
4.2.17(d) Fast protein liquid chromatography (FPLC) *124*
4.2.17(e) Real time biomolecular interaction analysis (BIAcore™ 2000) *125*
4.2.17(f) Expression and purification of GST-3C-CDK2 ..... *128*
4.2.17(g) Isothermal calorimetry (ITC) *129*
4.2.18(a) In vitro functional assay *129*
4.2.18(b) Preparation of yeast extracts *129*
4.2.18(c) Histone H1 Kinase assays *130*
4.2.19(a) In vivo functional assay *130*
4.2.19(b) Subcloning SpSuc1 and Ile94Ala:Leu96Ala mutant into yeast expression vectors *132*
4.2.19(c) Preparation of "ultra" competent E. coli cells *132*
4.2.19(d) Preparation of pREP DNA stocks *132*
4.2.19(e) Restriction digest of the pREP vectors *133*
4.2.19(f) Extraction and purification of vector and insert fragments for the ligation reaction .. *133*
4.2.19(g) Analysis of transformants *134*
4.2.19(h) Preparation and transformation of competent S. pombeΔSuc1 cells by: *134*
4.2.19(i) Isolation and analysis of haploid spores *135*
4.2.20(a) Activation of the Anaphase Promoting Complex (APC) *136*
4.2.20(b) Generation of Arg30Ala:Arg39Ala double and Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutants *136*
4.2.20(c) Assay of Cyclin-Ubiquitin Ligase Activity *138*
4.2.20(d) Western blot analysis of SpSuc1 mutants *139*

4.3 RESULTS **141**

4.3.1 Identification of mutant clones *141*
4.3.2 Purification of mutant SpSuc1 proteins *142*
4.3.3 Electrospray mass spectrophotometry (ES MS) *144*
4.3.4 Circular dichroism spectroscopy *146*
4.3.5 Zinc-induced dimerisation of SpSuc1 ..... *148*
4.3.6 Binding of GST-CDK2 to SpSuc1 mutants *149*
4.3.7 Histone kinase(HK) assays *161*
4.3.8 Yeast in vivo rescue assay *162*
4.3.8(a) Subcloning SpSuc1 gene fragment into yeast expression vectors *162*
4.3.8(b) Transformation of recombinant plasmid DNA into S. pombeΔSuc1 ..... *164*
4.3.9 APC activation assay *165*

4.4 DISCUSSION 169

4.4.1 Retention of structural integrity 169
4.4.2 The Oligomerisation State of SpSuc1 169
4.4.3 Comparative binding studies 170
4.4.4 The hydrophobic patch 172
4.4.5 The anion-binding site 178
4.4.6 Summary 182

4.5 LITERATURE CITED 183

List of figures

Figure 4.1 Illustration of the hydrophobic/aromatic and charged cluster sites on the surface of SpSuc1 --- 112
Figure 4.2 The principle of Surface plasmon resonance (SPR) ---------- 125
Figure 4.3 The sensogram ---------- 126
Figure 4.4 The chemistry of immobilisation to the sensor chip ---------- 127
Figure 4.5 Coupling procedure ---------- 127
Figure 4.6 The CDK1-cyclin B kinase assay ---------- 129
Figure 4.7 Illustration of pREP yeast expression vectors ---------- 131
Figure 4.8 Illustration of the PCR-based site-directed mutagenesis protocol ---------- 137
Figure 4.9 Manual DNA sequencing of (a) Asp54Asn mutant and (b) wild-type SpSuc1 ---------- 141
Figure 4.10a A typical SpSuc1 protein elution profile from the S100 gel filtration column ---------- 142
Figure 4.10b Expression and purification of Asp54Asn mutant ---------- 143
Figure 4.11 Eluants of Asp54Asn from the H.I.C. column ---------- 143
Figure 4.12 Electro-spray mass spectrum of SpSuc1 mutants ---------- 145
Figure 4.13 Comparative CD spectra of wild-type SpSuc1 and the mutants ---------- 146
Figure 4.14 Zinc-induced dimerisation of SpSuc1 ---------- 148
Figure 4.15 Overlay plot of SPR profiles obtained from binding interactions between the ligand (wild-type SpSuc1 and mutants) and the analyte (GST-CDK2) ---------- 150
Figure 4.16 CDK2 binding to immobilised wild-type SpSuc1 ---------- 153
Figure 4.17 Binding of GST-3C-CDK2 to immobilised wild-type SpSuc1 (1.9μM) ---------- 155
Figure 4.18 Binding of GST-3C-CDK2 to immobilised Ile94Ala:Leu96Ala double mutant ---------- 157
Figure 4.19 Comparative sensograms depicting the binding of GST-3C-CDK2 to (a) wild-type SpSuc1 and (b) Ile94Ala:Leu96Ala double mutant ---------- 159
Figure 4.20 Histone H1 kinase activity of wild-type SpSuc1 and Ile94Ala:Leu96Ala mutant ---------- 161
Figure 4.21 Restriction digest of pREP and pTZ18R/SpSuc1 plasmid vectors ---------- 162
Figure 4.23 The product of the PCR-based mutagenesis reaction ---------- 165
Figure 4.24 Elution profile of purified Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant on an S200 (Pharmacia) gel filtration column ---------- 166
Figure 4.25 Western blot analysis of a dimeric SpSuc1 species ---------- 167
Figure 4.26 The X-ray structure of CksHs1-HsCDK2 ---------- 173
Figure 4.27 Amino acid sequence alignment of Cks proteins ---------- 174
Figure 4.28 The conserved hydrophobic patch residues of SpSuc1 and CksHs1 ---------- 177

Figure 4.29 CDK sequence alignment ---------- *178*
Figure 4.30 The conserved phosphate-binding site of SpSuc1 and CksHs1 ---------- *179*
Figure 4.31 Surface representation of Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant --------- *181*

List of Tables

Table 4.1 Summary of SpSuc1 mutants _____ *114*
Table 4.2 Mutagenic primers _____ *118*
Table 4.3 Nucleotide sequence of the primers _____ *137*
Table 4.4 Competency test on E. coli SURE cells _____ *163*
Table 4.5 Transformation of recombinant pREP plasmid DNA into competent E. coli SURE cells _____ *163*
Table 4.6 APC activation assay with SpSuc1 _____ *168*
Table 4.7 CDK-Cks intersubunit contact residues _____ *176*

4.1 INTRODUCTION

The detailed examination of the atomic structures of *S. pombe* (SpSuc1) in (Endicott *et al.*, 1995) and in the preceding chapters has set the stage for the structure-based, site-specific mutagenesis experiments described in this chapter. The 100K SpSuc1 structure was unavailable at the time of these experiments so, the 293K SpSuc1 model was used as the basis for these studies. This work was initiated in October 1994 when the X-ray structure of SpSuc1 had just been determined but the function of the protein still remained an enigma. However, the atomic structure provided a few clues as to how SpSuc1 could mediate protein-protein interactions with other cell cycle regulatory units. An examination of this model revealed certain notable features that were considered worthy of further investigations to ascertain their physiological role(s): the first was a region predominated by aromatic and hydrophobic residues (Tyr31, Tyr38, His40, Tyr85, His88, Pro90, Glu91, His93, Ile94 and Leu96). All these residues are conserved with exposed side-chains. The second feature of interest was a region consisting of a cluster of positively charged side-chains contributed by Arg30, Arg39 and Arg99.

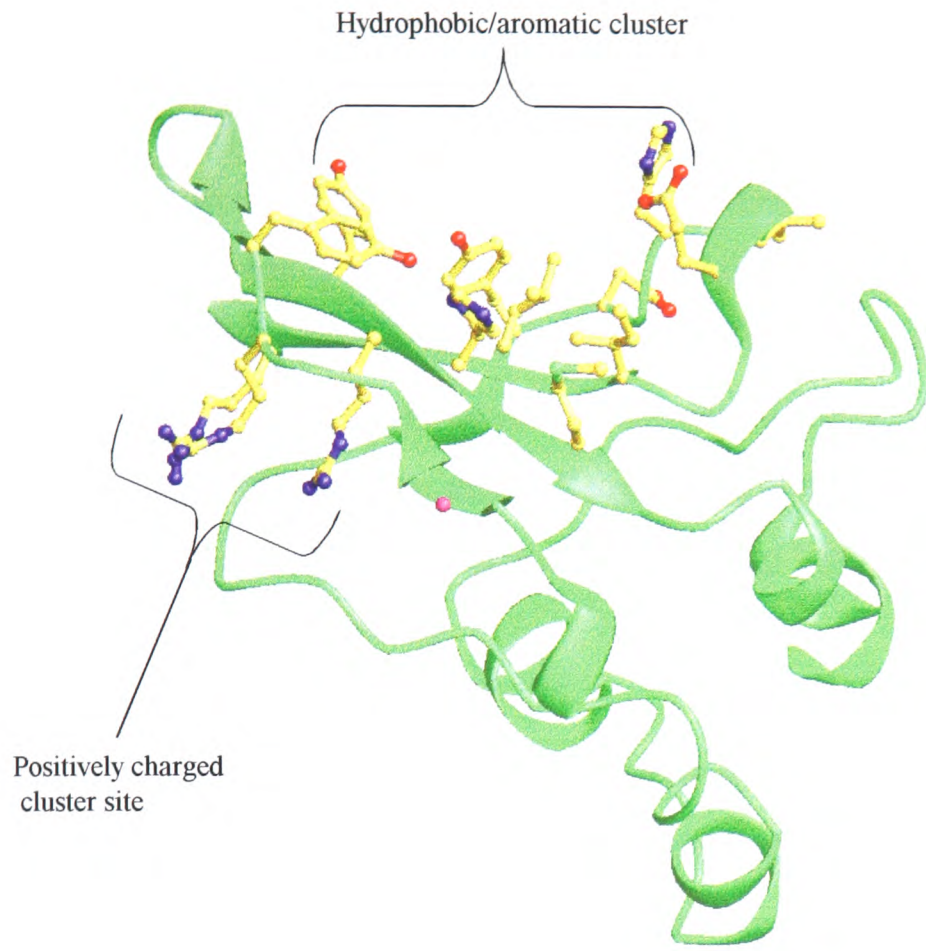


Figure 4.1 Illustration of the hydrophobic/aromatic and charged cluster sites on the surface of SpSuc1 *This shows a single subunit of the zinc-mediated dimer in the same orientation as that in figure 2.1 (chapter 2). The highlighted residues are mentioned in the test.*

Initially, the aim was to identify the residues on SpSuc1 that were responsible for mediating SpSuc1-CDK2 interactions, using among other techniques, site-specific mutagenesis and surface plasmon resonance (SPR). Due to the dificulties encountered during extensive attempts (Chapter 5 and Dr. Jane Endicott, University of Oxford, personal communication) to obtain appropriate quantities of soluble *S. pombe* CDK1 (the natural partner of SpSuc1), the CDK used in this study was the human homologue of CDK2 which was expressed and purified as a glutathione-S-transferase (GST) fusion protein. Consideration was given to the production of several mutants; those described in this thesis include:

1. An Arg30Ala mutant was designed to investigate the importance of the charged cluster sites.
2. Ile94Ala:Leu96Ala double mutant to investigate the role of residues in the hydrophobic region in protein-protein interactions.
3. The following salt-bridge interactions are present in each of the monomers that constitute the zinc-stabilised dimer: Arg8-Asp54, Arg8-Glu14, Glu37-Arg99, Arg64-Glu70, Glu68-Arg72 and His88-Glu91 for subunit ***A*** and Glu86-His88 in subunit ***B***.

An Asp54Asn mutant was made in order to sabotage one of the salt bridges (Arg8-Asp54) in an attempt to create a temperature-sensitive mutant of SpSuc1.

4. Zn3 forms a crystal lattice contact site with the side-chains of GluB86, HisB88, His88' and a water molecule. The importance of this zinc atom was to be investigated by a Glu86Ala mutant. An oligonucleotide primer to effect a Glu86Cys change was synthesised when efforts were being made to solve the crystal structure of SpSuc1. As it turned out (Endicott *et al.*, 1995) this proved unnecessary but I went ahead and made this mutant in an attempt to test the efficacy of this mutagenesis method. Apart from the binding studies presented on page 150, a discussion of this mutant has been omitted from the thesis.

After preliminary results from this study had been obtained, the crystal structure of CDK2-CksHs1 complex was reported (Bourne *et al.*, 1996) and this lent some credence to the initial findings reported in this chapter that only monomeric SpSuc1 binds to GST-CDK2 and this interaction is via the hydrophobic patch of SpSuc1. Although, this important finding dispelled some of the mystery surrounding the function of this small, cell cycle regulatory protein, an account of the exact physiological role still remained elusive. In a quest to answer this question, the most promising (from a functional point of view) mutant from the initial study was subcloned into yeast expression vectors and transformed into a special strain of *S. pombe* lacking one chromosomal copy of the *SpSuc1*$^{+}$ gene in order to examine the phenotypic effects (if any) on a yeast cell of harbouring a SpSuc1 molecule that was deficient in its binding to *S. pombe* CDK1 (SpCDK1).

The notable surface features of SpSuc1 (i.e. the hydrophobic patch and the charged cluster site) are structurally conserved among the atomic structures determined so far. The crystal structure of CksHs2 (Parge *et al.*, 1993) showed that the positively charged cluster site had a sulphate ion bound to it. That of CksHs1 had phosphate and the phosphate analogue, vanadate bound to it at this site (Arvai 1995) hence the name anion-binding site. Two groups have proposed that this anion-binding site could bind to a phosphorylated threonine e.g. Thr161 (Parge *et al.*, 1993) of human CDK1 (Thr167 in *S. pombe* CDK1) or the phosphorylated Tyrosine 15 of CDK1 (Endicott and Nurse, 1995).

Interestingly, the binding of CksHs1 to human CDK2 via the hydrophobic patch leaves the charged cluster site exposed to bind to an appropriate target molecule (Bourne *et al.*, 1996). In a recent study (Sudakin *et al.*, 1997), it has been demonstrated that the active, phosphorylated form of the anaphase promoting complex (APC) binds to SpSuc1 and the bound SpSuc1 is eluted most efficiently by phosphate-containing compounds such as p-nitrophenyl phosphate (pNPP). In a collaborative effort with the laboratory of Prof. Avram Hershko (Technion-Israel Institute of Technology), two new SpSuc1 mutants were made in an effort to test the hypotheses that SpSuc1 binds to the phosphorylated APC via the positively charged cluster site. A summary of the mutants described in this chapter is provided in the table below.

NAME	REASON
Arg30Ala	To probe the functional significance of the charged cluster site
Ile94Ala:Leu96Ala double mutant	To investigate the role of residues in the hydrophobic patch
Asp54Asn	An attempt to make a temperature-sensitive mutant
Glu86Ala	To test the importance of zinc to the structure and function of SpSuc1.
Glu86Cys	This was originally designed to aid the search for a derivative for crystal structure determination
Pro90Ala	Made by (Rousseau *et al.*, 1998) to investigate the importance Pro90 in the folding pathway of SpSuc1
Δ87-89	Also made by (Rousseau *et al.*, 1998) to investigate the role of the so-called hinge region in SpSuc1 folding.

Table 4.1 Summary of SpSuc1 mutants

4.2 MATERIALS & METHODS

4.2.1 Escherichia coli host strains and plasmid vectors

E. coli strains CJ236 and MV1190 were used for growing and amplifying plasmid DNA for the mutagenesis reaction. JM101, SURE™ and DH5α strains were used for amplifying and harvesting plasmid DNA stocks. BL21(DE3)pLysS was used for protein expression. The vector pTZ18R was used for cloning and pRK172 was used for expression. Standard microbiological techniques were adhered to (Sambrook *et al.*, 1989).

4.2.2 Schizosaccharomyces pombe host strains and plasmid vectors

CDK1-33^{ts} and 972h^{-} were the temperature-sensitive and wild-type strain of *S. pombe*, respectively. Crude extracts from these strains were used for the histone kinase assay. The SpSuc1 deletion strain *S. pombeΔSuc1* was used for the *in vivo* study. This strain lacks one chromosomal copy of *SpSuc1*$^{+}$ hence, cell viability depends on maintaining this strain in the diploid state. The extrachromosomally replicating pREP (pREP1, pREP41 and pREP81) and the integrative pRIP (pRIP5 and pRIP45) expression plasmids were constructed (Basi *et al.*, 1993; Maundrell, 1993) by placing the promoter and polyadenylation signal of the thiamine-repressible gene, *nmt1* of *S. pombe* upstream of the multiple cloning site (MCS) of pUC19-derived plasmid. Placing a gene under the control of the *nmt1* promoter permits thiamine-mediated regulation of the expression of the target gene. All the yeast strains and vectors were gifts from Dr. Jacky Hayles (Imperial Cancer Research Fund, London) to Dr. Jane Endicott (University of Oxford).

4.2.3 Site-specific mutagenesis

The protocol for this experiment was taken from the Muta-Gene phagemid *in vitro* mutagenesis kit (Bio-Rad). The principle (Kunkel *et al.*, 1987) is to allow selective degradation of the non-mutagenized strand and preservation of the mutant strand in a double-stranded DNA (dsDNA). The target DNA is synthesised in a *dut ung* double mutant of a strain of *Esherichia coli* (CJ236). The *dut* gene encodes for a protein called dUTPase: an enzyme responsible for degrading dUTP. Thus, the absence of dUTPase in

CJ236 allows for the accumulation of dUTP in the cell. The *ung gene* product is called uracil N-glycosylase; an enzyme that degrades uracil-containing DNA. The absence of this enzyme in CJ236 prevents the degradation of uracil-containing DNA. Infection of a culture of CJ236 with helper phage produces uracil-containing single-stranded DNA (ssDNA) that serves as a template for the *in vitro* synthesis of a complementary strand, which is primed by the mutagenic oligonucleotide. The product is a dsDNA which is then transformed into an *E. coli* cell with an active uracil N-glycosylase so that the uracil-containing strand can be degraded, leaving the non-uracil containing strand, which also has the desired mutation, to replicate. The frequency of mutation normally exceeds 50%, making the identification of mutants by sequence analysis less arduous.

4.2.4 Growth of uracil-containing phagemid

The SpSuc1 gene present on an Nde I-Eco RI fragment in pTZ18R was transformed into competent CJ236 cells. A single colony was picked and grown overnight with aeration at 37°C in 20ml of LB media containing 50µg/ml of ampicillin and 35µg/ml of chloramphenicol. An inoculum (1.0ml) of the overnight culture (ONC) was placed in 50ml of 2X yeast tryptone (YT) media containing ampicillin and chloramphenicol at the above concentrations. It was grown with shaking to an OD_{600} of 0.4. Helper phage M13KO7 was added to a multiplicity of infection (M.O.I) of about 20 phage/cell. The infected culture was incubated with shaking at 37°C for 60 minutes before kanamycin was added to a final concentration of 70µg/ml. Incubation was continued overnight. Thirty millilitres (30ml) of the ONC was placed in a centrifuge tube and spun at 12000 rpm on the sorvall SS-34 rotor for 15 minutes at 4°C. The supernatant was transferred to a fresh centrifuge tube and spun at 17000xg for 15 minutes at 4°C. The supernatant was then decanted into a fresh centrifuge tube and 150µg of RNase A was added followed by incubation at room temperature for 30 minutes. A mixture of 4.5M ammonium acetate plus 20% PEG-6000 was added to the phagemid-containing supernatant in a ratio of 1:4. This was properly mixed before incubation on ice for 30 minutes to precipitate the phagemid. The phagemid were collected by centrifugation at 17000xg for 15 minutes. The supernatant was discarded and the pellet was properly drained and resuspended in 200µl of high salt buffer (300mM NaCl, 100mM Tris, pH

8.0, 1mM EDTA). The resuspended phagemids were transferred to a 1.5ml eppendorf tube, chilled on ice for 30 minutes and centrifuged for 2 minutes to get rid of insoluble material. The supernatant was transferred to a fresh eppendorf tube and stored at 4°C.

4.2.4 Titration of phagemid stock

MV1190 has an active uracil N-glycosylase whereas CJ236 lacks this enzyme; if the phagemids have incorporated uracil to a significant extent, we would expect them to survive much more readily in CJ236 than in MV1190. This experiment was done to gauge the extent of uracil incorporation. Overnight cultures (ONCs) of MV1190 and CJ236 were grown and 0.5ml of the former was used to inoculate 50ml of 2X YT; 1ml of the latter was used to inoculate a separate 50ml 2X YT media. The two cultures were incubated at 37°C with shaking until the OD_{600} was between 0.3 - 0.35. The CJ236 culture attained the required OD_{600} before the MV1190 culture so its growth was attenuated by placing it on ice while incubation of MV1190 continued at 37°C. Ten microlitres (10µl) of the uracil-containing phagemid stock was added to each culture and incubation at 37°C was continued for 2 hours. The MV1190 culture was diluted 10-fold, 100-fold and 50µl of each dilution and undiluted culture was spread on ampicillin-containing LB agar plates. The CJ236 culture was diluted 10^3-fold, 10^4-fold and 10^5-fold and 50µl of each dilution was spread onto separate ampicillin-containing plates. The plates were incubated at 37°C overnight. The presence of at least a 10^4-fold fewer colonies from the MV1190 culture relative to that of CJ236 is indicative of satisfactory uracil incorporation into the phagemid.

4.2.5 Extraction of phagemid DNA

It is necessary, after obtaining a good titre, to purify the viral DNA by extraction before using it in an *in vitro* mutagenesis reaction. The phagemid stock (190µl) was extracted twice with an equal volume of neutralised phenol and once with phenol/chloroform (1:1:1/48 phenol: chloroform: isoamyl alcohol). Extraction was continued until there was no visible interface. Each extraction mixture was vortexed for 1 minute. Back extraction was performed as follows: 100µl of TE (10mM Tris, pH 8.0, 1mM EDTA) was added to the first phenol extraction tube, vortexed and spun briefly.

The resultant aqueous phase was added to the next phenol extraction tube, vortexed and so on. The aqueous phases were pooled and ammonium acetate was added (7.8M; one-tenth of the volume of the pooled phases). Ethanol (100%; 2.5X the volume of the mixture) was also added. The mixture was incubated at -70°C for 30 minutes to precipitate the DNA. The DNA was pelleted by centrifugation at 4°C for 15 minutes; the supernatant was carefully removed and discarded. The pellet was gently washed with 90% ethanol and resuspended in 20µl of TE (10mM Tris-HCl pH 7.5, 1mM EDTA) buffer. The dissolved ssDNA was transferred to a new eppendorf tube and quantitated by reading the absorbance of a diluted stock at 260nm*.

*1 OD_{260} unit of a ssDNA = 33µg/ml (Sambrook *et al.*, 1989).

4.2.6 Synthesis of mutagenic strand

The stage was now set for performing the *in vitro* mutagenesis reaction using the purified ssDNA as the template. A synthetic oligonucleotide containing the desired mutation plus any other novel sequence (e.g. a new restriction site) are required. The Oxford Centre for Molecular Sciences (OCMS) nucleotide synthesis facility prepared eight oligonucleotide primers. The sequences are as follows:

Mutant	**Primer sequence (5'- 3')**	**New restriction site**
Ile94Ala:Leu96Ala	ACGCTTAAAT**GC**CAGG**GC**GT GGCTCTGGTAC	Rsa I
Glu86Ala	GACATGGAC**AG**C**G**TACATTT CCCATCC	Rsa I
Asp54Asn	ATTGAAATAAT**T**GGTGGGTA TCGCTTTAGC	None
Arg30Ala	ATCATCAGC**G**TAC**GC**AAGGA GAATAATGG	Rsa I

Table 4.2 Mutagenic primers

Bold and enlarged letters represent mismatch bases and underlined letters represent silent restriction sites. The oligonucleotide primers were all phosphorylated at the 5' position by the supplier.

4.2.7 Purification of oligonucleotide primers

The mixture containing the oligonucleotides was heated at 55°C for 16 hours to expel ammonia in solution. Aliquots (360µl) were placed in sterile eppendorf tubes and 40µl of 3M sodium acetate was added to each aliquot, followed by 1.0ml of ethanol. The oligonucleotides were precipitated at -20°C for 30 minutes and then spun at maximum speed (13000 rpm) for 10 minutes in a microcentrifuge. The pellet was washed twice with 500µl of 70% ethanol. After drying the ethanol off, the pellet was resuspended in 100µl of sterile distilled water. The process - from precipitation to resuspension - was repeated and the absorbance at 260nm (required for quantitation) was taken.

4.2.8 Annealing of mutagenic primer to template

A reaction mixture containing 0.3pmol (200ng) of uracil-containing DNA, 6/9pmol of mutagenic primer and 1µl of 10X annealing buffer (20mM Tris-HCl pH7.4, 2mM $MgCl_2$ and 50mM NaCl) was made to a final volume of 10µl using sterile distilled water. A control reaction containing all the above ingredients less the mutagenic primer was also prepared. The reaction mixtures were placed in a 70°C heating block and allowed to cool slowly to 30°C over a 40-minute period. The reactions were placed in an ice-water bath to stabilise primer-template hybridisation.

4.2.9 Synthesis of complementary strand

The following components were added to the annealing reaction mixtures in an ice-water bath: 1µl of 10X synthesis buffer (4mM each dNTP, 7.5mM ATP, 175mM Tris-HCl pH7.4, 37.5mM $MgCl_2$ and 15mM DTT), 3 units of T4 DNA ligase (Boehringer Mannheim) and 1 unit of T4 DNA polymerase Boehringer Mannheim (B.M.). The reactions were incubated on ice for 5 minutes, then at 25°C for a further 5 minutes and finally at 37°C for 90 minutes. The reaction was stopped by the addition of 90µl of stop buffer (10mm Tris-HCl pH 8.0, 10mM EDTA).

4.2.10 Gel analysis of reaction products

DNA loading buffer (1µl) was added to 9µl of the *in vitro* mutagenesis reaction. This was properly mixed and electrophoresced at 150V on a 1% agarose gel in Tris-acetate EDTA (TAE) buffer.

4.2.11 Preparation of competent E. coli cells by the Calcium Chloride method

A single colony of bacteria was picked and inoculated into 5.0ml of LB media overnight at 37°C. The overnight culture ONC (50µl) was transferred to 20ml of LB media in a 50ml-falcon tube. This was grown with aeration at 37°C till the OD_{600} was between 0.2 - 0.5. The cells were centrifuged at 5000xg for 5 minutes. The pellet was resuspended in 5.0ml of ice-cold 0.1M $CaCl_2$ and incubated on ice for 45 minutes followed by centrifugation. The pellet was resuspended in 2.0ml of resuspension solution (7% glycerol and 50mM $CaCl_2$). The competent cells were stored on an ice-water bath at 4°C and used for transformation within 8 days.

4.2.12 Transformation of reaction products

The reaction mixture was placed in an eppendorf tube containing 200µl of competent *E. coli* cells. The tube was inverted twice and placed on an ice-water bath for 10 minutes. The mixture was heat-shocked for 2 minutes at 42°C on a heating block to allow the cells to take up the plasmid DNA. LB media (600µl) was added and mixed by inversion of the tube twice. The cells were incubated at 37°C for 20 minutes in order to allow for the expression of the antibiotic resistance gene before spreading onto LB agar plates containing the appropriate antibiotic(s).

4.2.13(a) Screening for mutants by dideoxy DNA sequencing

The Sequenase™ kit (Version 2.0 DNA polymerase) supplied by United States Biochemical (USB) was used for this experiment. The template DNA was obtained by miniprepping (using alkaline lysis/boiling/Wizard™ minipreps DNA purification system/QIAGEN spin plasmid kit) methods of a 3ml ONC of transformed cells.

4.2.13(b) Annealing reaction

In order to produce single-stranded template DNA (ssDNA) the double-stranded plasmid DNA was subjected to alkaline denaturation before it was incorporated into the annealing mixture [2.6µg of template DNA, 2µl of 5X reaction buffer (final concentration 40mM Tris-HCl pH 7.5, 20mM $MgCl_2$ and 50mM NaCl), 6.6ng of primer]. The volume was made up to a total of 10µl with sterile distilled water. The primer was annealed to the template by heating at 65°C for 2 minutes and then cooled slowly to less than 35°C over a 30-minute period. This was chilled on ice for later use. While the annealing mixture was cooling, 2.5µl of each termination mixture (G, A, T and C) was placed in 4 sterile eppendorf tubes and kept covered at room temperature (25°C). The labelling mix was diluted 1:10 with sterile distilled water (final concentration 0.75µM dGTP, 0.75µM dCTP and 0.75µM dTTP). The Sequenase polymerase was diluted 8-fold prior to use. The 4 termination tubes were pre-warmed at 37°C for 5 minutes.

4.2.13(c) Labelling reaction

The following were added to the ice-cold annealed DNA mixture: 1µl of 0.1M DTT, 2µl of the diluted labelling mix, 0.5µl of [^{35}S]dATP (0.5µCu) and 2µl (0.8units) of diluted sequenase polymerase. These were mixed and incubated at room temperature for 2-5 minutes.

4.2.13(d) Termination reactions

Aliquots of the labelling reaction (4.5µl) were transferred to each termination tube and mixed properly. Incubation of the termination reactions was continued at room temperature for 5 minutes. The reactions were stopped by the addition of 4µl of stop buffer. The samples were electrophoresed at 1500V on a 7% Tris-borate EDTA (TBE) gel till the first dye had reached the bottom of the gel. The gel was soaked in 10% acetic acid followed by 12% methanol for 15 minutes to rid it of urea. It was then transferred to a piece of 3MM paper and dried at 80°C for 60 minutes. The dried gel was exposed to an X-ray film for 24hr. An X-OMAT machine (Kodak) was used to develop the autoradiograph.

4.2.14 Subcloning mutant SpSuc1 genes into pRK172 vector

Mutant SpSuc1 genes were subcloned from pTZ18R on a Nde I-Eco RI fragment into an expression vector called pRK172. The T7 promoter directs transcription of the target gene in pRK172. The use of pRK172 in conjunction with cell lines like BL21(DE3)pLysS provides two main advantages; the first is a highly regulated expression system resulting in little or no basal expression of target genes. This enables the expression system to handle genes whose protein products could be lethal to the cell. Secondly, the existence of a plasmid-encoded lysozyme gene allows for a relatively easy lysis of the cells to release expressed proteins (Studier, 1991). Plasmid-bound mutant SpSuc1 genes were purified and transformed into BL21(DE3)pLysS for induction and expression. Two units of the restriction enzyme Nde I (New England Biolabs) and 2µl of 10X reaction buffer (50mM potassium acetate, 20mM Tris-acetate, 10mM magnesium acetate, 1mM dithiothreitol pH 7.9) were added to 5µg of the recombinant plasmid in an eppendorf tube. The total volume was made up to 20µl with sterile distilled water and the reaction mixture was properly mixed before incubation at 37°C for 60 minutes. A further 2 units of Nde I was added and incubation was continued at that temperature for another hour. The eppendorf was spun briefly before the addition of 10X Eco RI reaction buffer (50mM NaCl, 100mM Tris-HCl, 10mM $MgCl_2$, 0.025% TritonX-100 pH 7.5) to a final concentration of 1X. Two units of Eco RI were added and incubation was continued at 37°C for 60 minutes. An equivalent amount of the pRK172 vector was also cut with the two (Nde I and Eco RI) restriction enzymes. DNA loading dye was added and the sample was electrophoresced on a 1% agarose gel containing ethidium bromide (EtBr) at 80V for 60 minutes. The insert (mutant SpSuc1) and the linearised vector were excised from the gel and purified - using QIAEX II gel extraction kit from QIAGEN Ltd., Germany - for the subsequent ligation reaction.

4.2.15 Ligation reaction

Two microlitres (2µl) of 10X ligase buffer (660mM Tris-HCl, pH 7.5, 50mM $MgCl_2$, 10mM DTE and 10mM ATP) and 1unit of T4 DNA ligase (New England Biolabs) were added to 17µl of the purified mixture from above. The reaction was either incubated at room temperature for 60 minutes or at 20°C for 16 hours. The DNA was

transformed (see section 4.2.12) into competent *E. coli* JM101 cells. Stocks of pRK172/mutant were harvested from the transformed cells and 1µg of each pRK172/mutant was transformed into competent BL21(DE3)pLysS cells; a single colony of the transformant was picked and checked for overexpression of the cloned gene.

4.2.16 Expression and purification of mutant SpSuc1

The established protocol for producing wild-type SpSuc1 (Brizuela *et al.*, 1987) was used to express and purify the mutant proteins. A single colony of transformed BL21(DE3)pLysS was picked and grown overnight in 2X YT media containing ampicillin (50µg/ml) and chloroamphenicol (35µg/ml). The ONC was inoculated into 500ml of LB media in a 2l flask. Ampicillin was added to a final concentration of 10µg/ml. The cell culture was grown to mid-log phase (i.e. OD_{600} of 0.4 - 0.5) at 37°C before protein expression was induced by the addition of isopropyl thiol-β-D-galactoside (IPTG) to a final concentration of 0.4mM. Growth was continued for 2-4 hours. The cells were harvested by centrifugation at 5000 rpm in a Sorval RC-5B plus at 4°C. The cell pellet was resuspended in GF buffer (50mM NaCl, 1mM EDTA, 1mM DTT and 50mM Tris-HCl pH 7.5). A cocktail of protease inhibitors containing aprotinin (2µg/ml), leupeptin (2µg/ml) and pepstatin A (4µg/ml) was added. The cells were lysed by freeze-thaw and sonication. The broken cells were spun at 100000g for 60 minutes at 4°C. Like the wild-type protein, the mutant was present in the supernatant. Polyethyleneimine [0.1% (v/v)] was added to the supernatant to precipitate anionic matter. This was followed by further precipitations with ammonium sulphate; first with 30% (w/v) and then 60% (w/v). The precipitated mutant protein was collected by centrifugation and then resuspended in GF buffer. Two chromatographic columns were used to purify the mutant protein at 4°C; gel filtration column (S100, Pharmacia 2.6 x 100cm) which had undergone prior equilibration in GF buffer at a flow rate of 1.75ml/min followed by hydrophobic interaction chromatography (H.I.C) (phenyl-sepharose CL-4B, Pharmacia 2.6 x 8.5cm). The H.I.C column was equilibrated in a mixture of 0.85M ammonium sulphate plus 50mM Tris-HCl (pH 7.5) at a flow rate of 4.5ml/min. Fractions containing mutant SpSuc1 proteins were identified by SDS-PAGE. The fractions from the H.I.C column were pooled and prepared for dialysis using centriprep-3 (Amicon, Beverly,

M.A., U.S.A). Dialysis of the samples into 10mM Tris-HCl (pH 7.5) and subsequent concentration using centricon-10 from the above suppliers yielded a pure protein (≈95% homogeneity) of about 18 - 30mg/ml.

4.2.17(a) Characterisation of mutant SpSuc1 protein

The following techniques/experiments were employed to characterise the mutant proteins:

i). Electrospray mass spectrometry (ES MS).

ii). Circular dichroism (CD) spectroscopy.

iii). Fast protein liquid chromatography (FPLC).

iv). Real-time biomolecular interaction analysis (BIAcore™).

v). Isothermal calorimetry (ITC).

vi). Histone kinase (HK) assays.

vii). Yeast *In vivo* rescue assay.

viii). Activation of the Anaphase promoting complex (APC)

4.2.17(b) Electrospray mass spectrometry

Samples (1mg/ml in 10mM Tris-HCl pH 7.5) were analysed on a BIOQUE instrument (VG Biotech) by Dr. Robin Aplin.

4.2.17(c) Circular dichroism spectroscopy

Samples of the wild-type and mutant proteins (0.2mg/ml) were dialysed into 10mM $NaPO_4$ (pH 8.0) before analysis on a Jasco J720 under the following parameters; (a) Cell length: 0.5cm, (b) Temperature: room temperature, (c) Start: 250nm, (d) End: 190nm, (e) Resolution: 1nm, (f) Band width: 1.0nm, (g) Sensitivity: 20 mdeg, (h) Response: 1sec, (i) Speed: 10nm/min, (j) Accumulation: 4.

4.2.17(d) Fast protein liquid chromatography (FPLC)

The oligomerisation state of the mutants was checked in the presence and absence of Zn^{2+}. Each sample (1mg/ml) was prepared in a buffer containing 10mM Tris-HCl pH 7.4, 200mM NaCl with or without 1mM $ZnSO_4$. The samples were ran at a flow rate of

1ml/min on either a superose-12 (HR 10/30) gel filtration column (Pharmacia) or a Superdex-75 (HR 10/30) gel filtration column (Pharmacia).

4.2.17(e) Real time biomolecular interaction analysis (BIAcore™ 2000)

The interaction of SpSuc1 mutant proteins with GST-CDK2 was investigated using BIAcore™ 2000 (Pharmacia Biosensor, Uppsala, Sweden). The essential parts of the BIA (for Biomolecular Interaction Analysis) system are a sensor chip, an optical system and an automated liquid handling system. The sensor chip is a biospecific surface on which binding interactions occur. BIAcore utilises a biosensor (defined as "an instrument that combines a biological recognition mechanism with a sensing device or transducer") to monitor biomolecular interactions in real time.

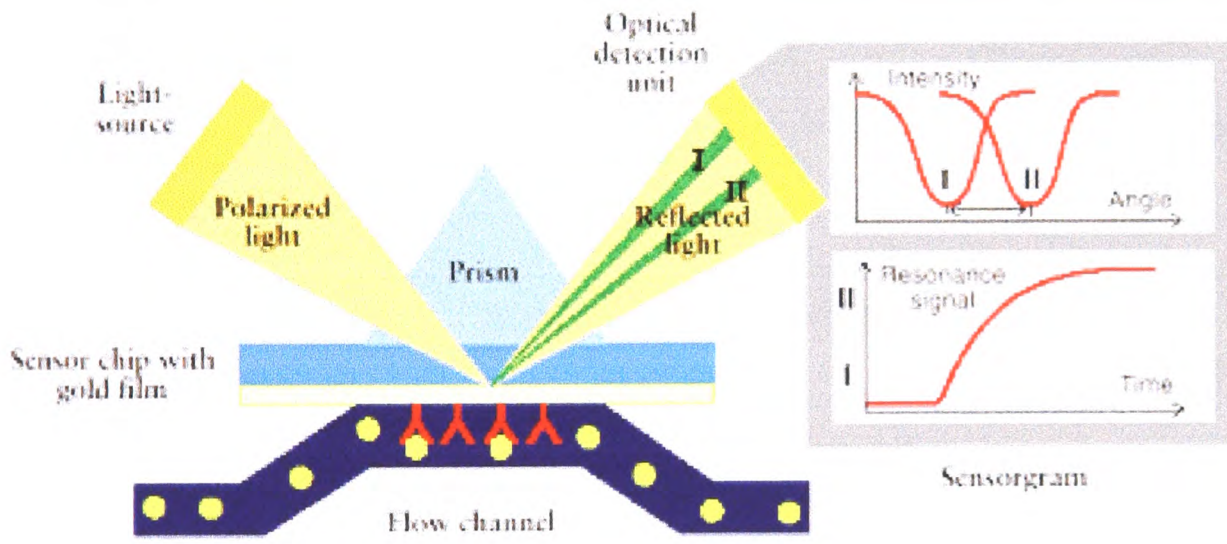


Figure 4.2 The principle of Surface plasmon resonance (SPR)

This is an optical phenomenon that is employed to monitor biomolecular interactions in the BIAcore system. Conditions of total internal reflection (TIR) in thin metal films give rise to SPR. This fact is employed in the BIAcore instrument as follows; plane-polarised light is incident on a sensor chip in a manner that leads to TIR. This can be achieved by ensuring that the angle of incidence is greater than a certain critical angle. A monitor detects the intensity of the reflected light, which is sensitive to its location in the detection chamber. SPR produces a sharp decrease in the intensity of the reflected light at a particular angle called the SPR angle or the resonance angle. The position of the resonance angle corresponds to the refractive index on the surface of the metal film (non-

illuminated surface) and this in turn corresponds to the concentration of solubilised material. Changes in SPR are reported in response units (RU).

The SPR profile i.e a plot of RU against time is called a sensogram .

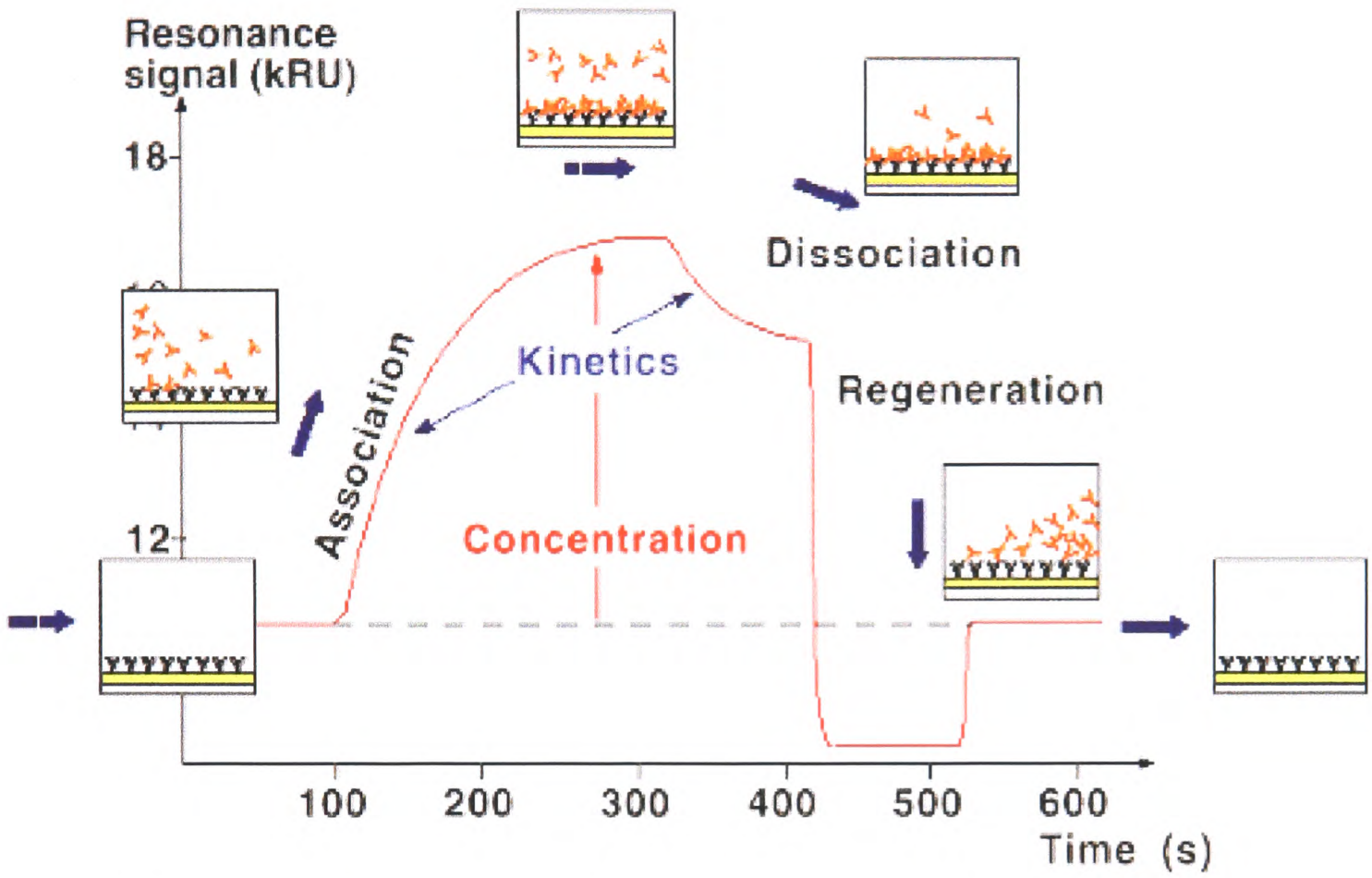


Figure 4.3 The sensogram

This shows the profile of the interaction event in real time.

Immobilisation procedure: One of the reactants (the ligand) in a BIAcore experiment is always immobilised on a sensor chip and the other reactant (analyte) is passed over the ligand (Pharmacia, 1994; Raghavan and Bjorkman, 1995; Szabo *et al.*, 1995). Currently, one can immobilise a ligand on a sensor chip in six different ways i.e via surface thiol, ligand thiol, aldehyde groups, amine groups (fig 4.4), avidin-biotin and nitrilotriacetic acid (NTA) (Pharmacia, 1994; Szabo *et al.*, 1995). The ligands in this study (SpSuc1 and the various mutants) were immobilised via amine coupling to the carboxymethylated dextran surface of the sensor chip.

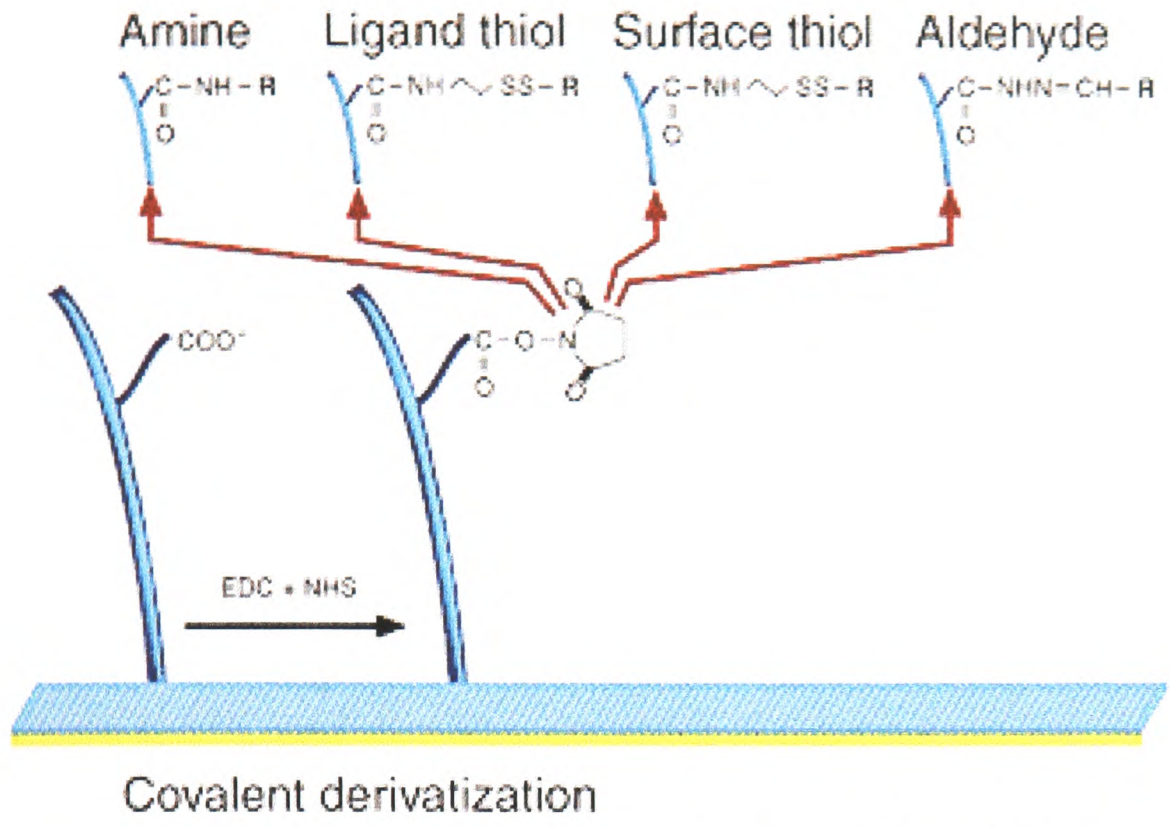


Figure 4.4 The chemistry of immobilisation to the sensor chip

A 1:1 mixture of N-hydroxysuccinimide (NHS) and N-ethyl-N-(dimethyl-aminopropyl)-carbodiimide (EDC) was used to introduce N-hydrosuccinimide esters into the hydrogel. This leads to the esterification of the negatively charged carboxymethyl groups on the dextran surface. The esters react with free amine groups on the ligand (SpSuc1 and mutants) in a spontaneous reaction to form covalent links. In the preliminary study, 1.9μM of each ligand was immobilised.

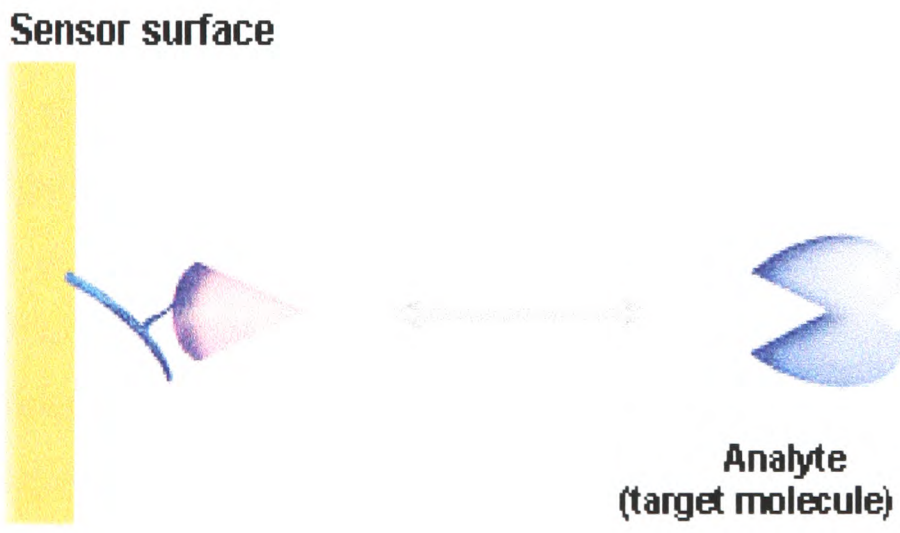


Figure 4.5 Coupling procedure

The analyte (GST-CDK2) was injected onto the SpSuc1-bound sensor chip at a flow rate of 10μl/min in HBS buffer (10mM HEPES pH 7.4, 150mM NaCl, 4.4mM EDTA, 0.05% Surfactant P20) at several

different protein concentrations and at a temperature of 25°C. The interaction was then followed in real time in the resulting sensogram. (figs 4.1-4.4 were reproduced from BIAweb at http://www.biaweb.com).

4.2.17(f) Expression and purification of GST-3C-CDK2

A glycerol stock of transformed B834(DE3)pLys-S cells (a generous gift from Dr. Hanlon) were streaked onto an LB plate containing 100µg/ml of ampicillin and incubated overnight at 37°C. A single colony was picked and grown in a 10ml culture of LB media containing ampicillin (50µg/ml) and this was grown with shaking at 37°C. This inoculum was then transferred to 500ml of LB media containing ampicillin and grown with aeration at 37°C until the OD_{600} reached to 0.6 (≈4 hours). The flask was cooled on ice for 5 min by gentle swirling to lower the temperature of the medium to ≈20°C. Protein expression was induced by the addition of IPTG (1M stock in water) to a final concentration of 80µM and the cell culture was incubated with aeration at 18°C for 12 hours. The flask was cooled on ice and the cell culture was harvested by centrifugation at 4000 rpm for 15 min at 4°C. The cell pellet was resuspended in 15ml of ice-cold equilibration buffer [30mM Tris pH 8.2, 150mM NaCl, 3mM DTT plus a cocktail of complete™ protease inhibitors (Boehringer Mannheim)]. The suspension was sonicated for 1min on ice 3 times. The mixture was spun at 10000xg for 5 min at 4°C. The supernatant was transferred into a capped 50ml culture tube and filtered through a disposable 0.22µm filter unit (Whatman). A 3ml slurry of Glutathione Sepharose 4B beads (Pharmacia) was added and the mixture was rotated for 1 hr at 4°C to allow binding to the fusion protein. The mixture was poured into a column and washed with equilibration buffer till the absorbance reading at 280nm returned to as near the baseline as possible. The fusion protein was eluted in a buffer containing 10mM Glutathione (Sigma), 50mM Tris-HCl (pH7.5) and 150 mM NaCl at 4°C. Addition of a powder of glutathione to the buffer lowered the pH from 7.5 to ≈4.8 so it was necessary to re-pH the elution buffer before applying it the column. The fusion protein was dialysed into HBS buffer and concentrated. The yield was 17.5mg of GST-3C-CDK2 from a litre of bacterial culture.

4.2.17(g) Isothermal calorimetry (ITC)

For this experiment, 10μM of GST-3C-CDK2 and 50μM of wild-type SpSuc1 were dialysed overnight into HBS buffer (10mM HEPES pH 7.4, 250mM NaCl, 10% glycerol, 1mM EDTA and 0.5mM DTT) and sent to Dr. John Ladbury (University of London) for analysis.

4.2.18(a) In vitro functional assay

This *in vitro* functional assay was conducted to determine the ability of the mutant proteins to rescue a CDK1 temperature-sensitive mutant.

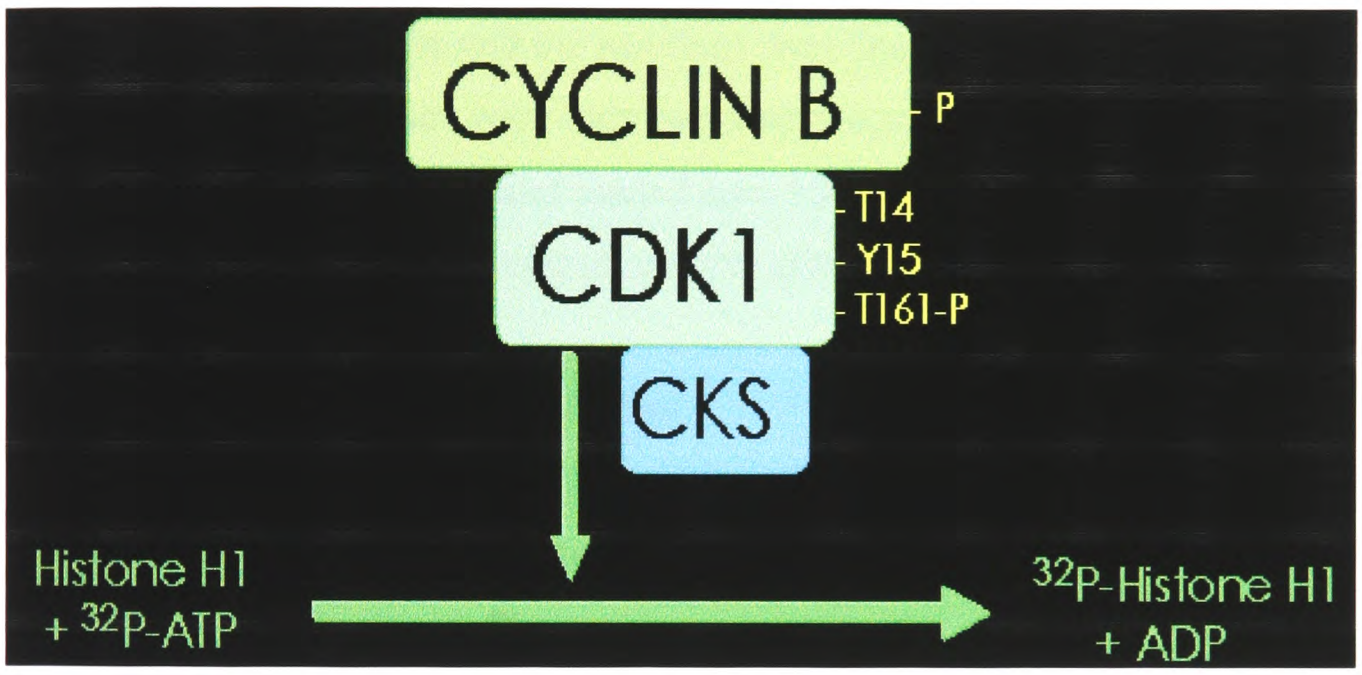


Figure 4.6 The CDK1-cyclin B kinase assay

Figures 1.16 and 4.6 were reproduced with modification from Dr. Meijer's slide collection at www.sb-roscoff.fr/CyCell.

4.2.18(b) Preparation of yeast extracts

Fission yeast strains (CDK1-33 and 972h$^-$) were grown in separate minimal media at 25°C to an OD_{595} of 0.2-0.5; the former is the temperature-sensitive mutant and the latter is the wild-type strain. The cells were harvested by centrifugation at 3000 rpm at 4°C and resuspended in acetate buffer (100mM potassium acetate, 2.5mM magnesium acetate and 1mM DTT). The cells were lysed by french pressing at 20000 psi. The extract was centrifuged at 10000g for 10 minutes to separate the lipid phase from the rest of the

cell debris. The supernatant was removed and frozen by shooting 50µl aliquots over liquid nitrogen. The yeast cell extracts were stored at -70°C.

4.2.18(c) Histone H1 Kinase assays

Four microlitres (4µl) of yeast extract supernatant was added to 6µl of HB-5 buffer (60mM β-glycerophosphate, 15mM p-nitrophenyl phosphate, 25mM MOPS pH 7.2, 15mM $MgCl_2$, 5mM EGTA, 1mM DTT, 0.1mM Na orthovanadate, 0.2mM phenylmethanesulfonyl flouride (PMSF), 20µg/ml leupeptin and 40µg/ml aprotinin) before preincubation at 25°C or 40°C for 15 minutes. Ten microlitres of KIN buffer (final concentration: 1mg/ml histone H1, 40µCi/ml γ^{32}P-ATP in HB-5 buffer, 100µM ATP) was added to the mixture and the reactions were incubated at 25°C or 40°C for a further 20 minutes. The reactions were terminated by the addition of 20µl of 2X PAGE sample buffer, boiled for 3 minutes and loaded onto 15% SDS-PAGE minigels. Phosphorylated histone H1 was detected by autoradiography after a 24-hour exposure at 80°C.

4.2.19(a) In vivo functional assay

In order to examine the phenotypic effects on an *S. pombe* cell of harbouring a copy of SpSuc1 that was defective in binding to CDK1, the wild-type and the Ile94Ala:Leu96Ala double mutant were subcloned into yeast expression vectors and transformed into *S. pombeΔSuc1*.

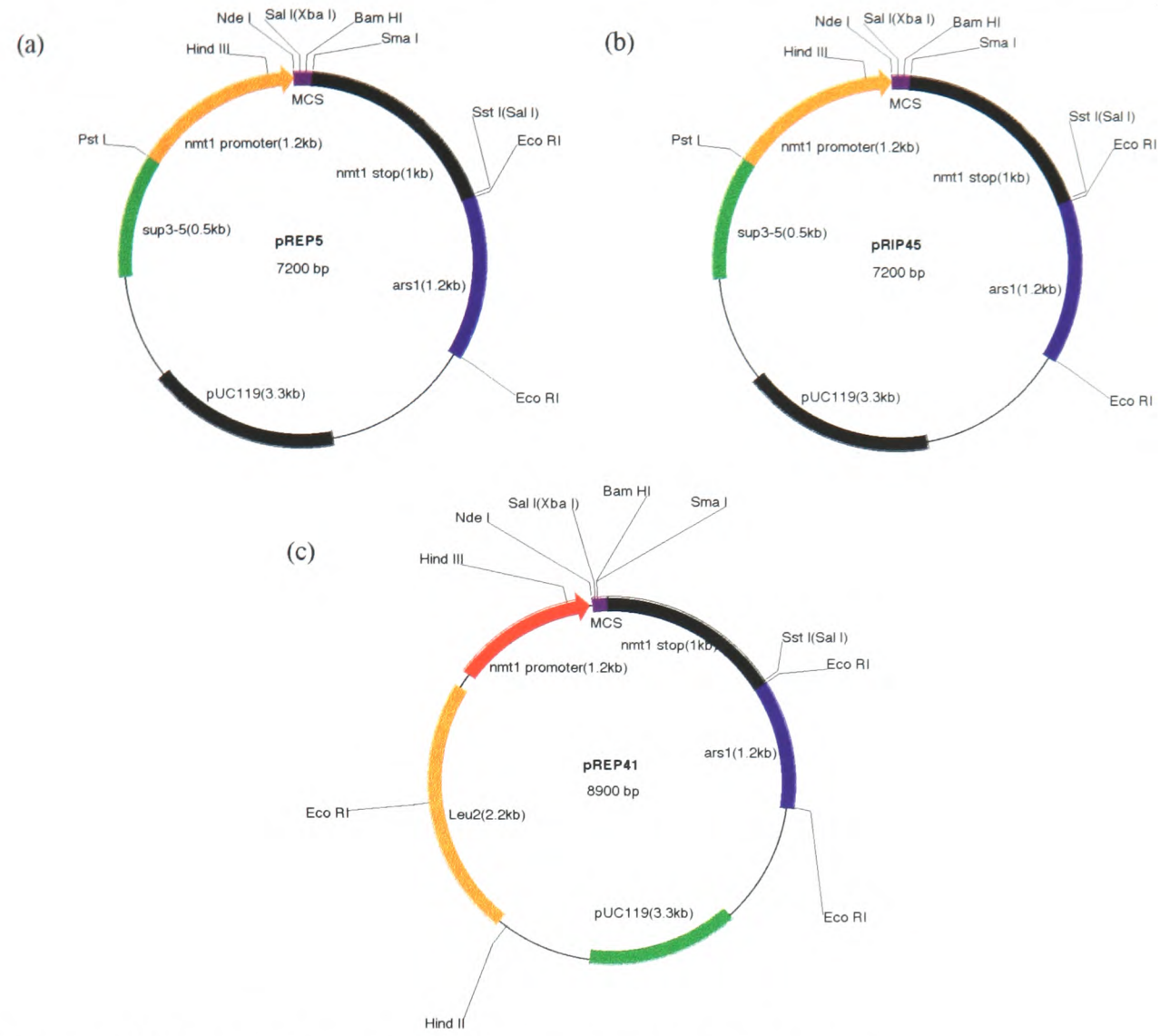


Figure 4.7 Illustration of pREP yeast expression vectors

All the vectors were pUC119-based and the presence of both the bacterial origin of replication and the yeast autonomous replication sequences (ars1) allowed for maintenance and replication in bacterial cells as well as yeast cells hence the designation of these type of plasmids as shuttle vectors. In bacteria, the selectable marker used was ampicillin whilst in yeast, it was leucine. Sup3-5 is a nonosense suppressor which is useful for selecting single copy intergrants i.e. in the case of pRIP5 and pRIP45 vectors. Expression of the target gene was under the control of the thiamine-repressible nmt1 promoter. Three versions of the pREP vectors with variable promoter strengths (pREP1>pREP41>pREP81) were used in this study. It has been shown that the nmt1 promoter does not switch off completely in that many genes expressed under nmt1 control are able to complement deficiencies in S. pombe strains even in the weakest promoter (pREP81) but others are not (Forsburg, 1993).

4.2.19(b) Subcloning SpSuc1 and Ile94Ala:Leu96Ala mutant into yeast expression vectors

For the bacterial expression, SpSuc1 was subcloned on an Nde I-Eco RI fragment into the pRK172 expression vector. In this experiment, the SpSuc1 genes were subcloned on an Nde I-Sal I fragment into the pREP/pRIP vectors for the yeast expression because Eco RI is not a unique site in the yeast expression vectors (fig. 4.5). Sal I is present at the 3' untranslated region (UTR) of the SpSuc1 gene fragment.

4.2.19(c) Preparation of "ultra" competent E. coli cells

The competency of the *E.coli* cells produced by the calcium chloride method was about 10^3-10^4 per microgram (μg) of DNA. This method was discarded for a more efficient process which regularly yielded 10^7-10^8 colonies per microgram (μg) of plasmid DNA. They were prepared according to an original method (Inoue *et al.*, 1990) modified by Dr. Yoshi Watanabe (Imperial Cancer Research Fund, London). A single colony of *E.coli* DH5α cells were inoculated into LB media (2 ml) and grown overnight with shaking at 37°C. One millilitre (1ml) of the overnight culture (ONC) was transferred to 500ml of LB media supplemented with 5ml of 1M $MgCl_2$. This was grown with vigorous shaking at 18°C until the OD_{595} was between 0.25-0.7. The cells were then cooled on ice for 10 minutes followed by centrifugation at 2500xg for 10 minutes. The pellet was resuspended in 80mls of ice cold TB buffer (10mM Pipes, 55mM $MnCl_2$, 15mM $CaCl_2$ and 250mM KCL) and incubated in an ice bath for 10 minutes. The cells were harvested again by centrifugation and the pellet was resuspended in 20ml of TB buffer to which DMSO had been added to a final concentration of 7%. The cells were mixed by gentle swirling on ice. After a 10-minute incubation on ice, the cells were aliquoted into sterile eppendorfs. The competent cells were immediately chilled in an ethanol-dry ice bath and stored at -80°C.

4.2.19(d) Preparation of pREP DNA stocks

Two microlitres (2μl) of plasmid DNA was transformed into competent *E. coli* DH5α cells as described in section 4.2.12. A single colony of the transformant was picked and grown in 500ml of LB media for 16 hours (overnight). The cells were

harvested by centrifugation and the plasmid DNA was isolated using the QIAGEN midi-prep kit. For DNA quantitation, 10µl of the stock sample was made up to 100µl with Tris-EDTA pH 7.4 (TE) buffer and the OD_{260} value was taken against a TE buffer blank. Typical yields were: pREP1=480µg/ml, pREP5=460µg/ml, pREP41=320µg/ml, pRIP45=360µg/ml and pREP81=520µg/ml.

4.2.19(e) Restriction digest of the pREP vectors

Each yeast vector was subjected to a double restriction digest. The buffer composition was 150mM NaCl, 10mM Tris-HCl, 10mM $MgCl_2$, 1mM dithiothreitol (pH 7.9 at 25°C) supplemented with 100µg/ml BSA. All the buffers, enzymes and BSA were supplied by the New England Biolabs (NEB). All the digests were incubated at 37°C for 90 min. The digests were stopped by the addition of DNA loading dye and each reaction mixture was loaded onto a 1% agarose gel and ran at 200V for about 2 hours. About 3-6µg of each pREP vector was subjected to a double restriction digest using 2µl each of Nde I (40U) and Sac I (40U). A similar digest was performed on pTZ18R/SpSuc1 recombinant plasmids to release wild-type SpSuc1 and Ile94Ala:Leu96Ala DNA inserts for the subsequent ligation reaction.

4.2.19(f) Extraction and purification of vector and insert fragments for the ligation reaction

The QIAGEN gel extraction kit was used to purify the linearised pREP vectors and the SpSuc1 inserts, separately. For quantification, 10µl of each DNA sample was made up to 100µl with sterile distilled water and the *OD_{260} was read against a water blank. Ligation reactions were set as described under section 4.2.15 for the ligation of the SpSuc1 and Ile94Ala:Leu96Ala fragments into pRIP5, pREP41, pRIP45 and pREP81 vectors using 100ng each of vector and insert in a reaction volume of 15µl. Each reaction mixture was supplemented with ATP to a final concentration of 2mM (supplied 10X buffer contains 10mM ATP). The ATP supplement was added to offset the fact that repeated freeze-thawing of the supplied buffer tended to hydrolyse the ATP component. All the ligation reactions were incubated at 16°C for 16 hours before they were

transformed into competent *E.coli* DH5α cells as described in section 4.2.12 of this chapter.

* 1 OD_{260} unit of double-stranded DNA = 50μg/ml (Sambrook *et al.*, 1989).

4.2.19(g) Analysis of transformants

A single representative colony of each transformant from the LB agar plate from was picked and inoculated into 10ml of LB media and grown overnight (≈16 hours) at 37°C. The QIAGEN miniprep DNA kit was used to isolate DNA from the overnight culture according to the manufacturer's instructions. Each DNA sample was subjected to a double digest using Nde I and Sac I restriction enzymes to ascertain the presence or otherwise of the DNA insert.

4.2.19(h) Preparation and transformation of competent S. pombeΔSuc1 cells by:

1. The Lithium acetate method

A single colony of yeast cells *S. pombeΔSuc1* was grown in a 150ml culture of MB medium to a density of 8 x 10^6 cells/ml (OD_{595nm} = 0.42*). The cell culture was harvested by centrifugation at 3000rpm for 5 minutes at room temperature. The supernatant was decanted and the cells were washed with 40ml of sterile distilled water and spun as before. In order to perforate the relatively tough yeast cell wall, the cell pellet was resuspended at a density of 1 x 10^9 cells/ml in 0.1M lithium acetate (adjusted to pH 4.9 with acetic) and dispensed as 100μl aliquots into eppendorf tubes. The cells were then incubated at 30°C for 60 minutes. A microgram (1μg) of the purified plasmid DNA to be transformed in (15μl of TE buffer pH 7.5) was added to the sedimented cells in each tube and this was mixed by gentle vortexing. A 290μl aliquot of 50% (w/v) PEG 4000 pre-warmed at 30°C was added and the entire mixture was gently mixed again before incubation at 30°C for another 60 minutes. After this incubation period, the cells were heat-shocked at 43°C for 15 minutes and then cooled to room temperature for 10 minutes. The mixture was centrifuged at 5000 rpm for 2 minutes in a microcentrifuge and the supernatant was carefully removed with the aid of an aspirator. The cells were then resuspended in 1ml of 1/2 YE5S with the aid of a P1000 pipette. The cell suspension was

transferred to a 50ml falcon tube and then diluted with 9ml of 1/2 YE5S. This was incubated with shaking at 32°C for 60 minutes. The cells were spun again and then resuspended in 1ml of 1/2 YE5S. Aliquots of 200µl were spread onto minimal media plates and incubated at 32°C for 4-5 days. Transformants began to appear on the 4th day.

*An OD_{595nm} of 0.5 = a cell density of 1x 10^7 cells/ml.

2. Electroporation

The cells were grown to a density of 7.4 x 10^5 cells/ml (OD_{595nm} = 0.37) in minimal medium and then harvested by centrifugation at 3000rpm for 5 minutes at 20°C. They were washed once by re-suspending in ice-cold water and then harvested again before re-suspending in ice-cold 1M sorbitol. The mixture was then re-harvested and re-suspended in ice-cold 1M sorbitol at a density of 1-5 x 10^9 cells/ml. Aliquots (100µl) of the cell suspension were added to pre-chilled cuvettes containing the DNA for transformation (100ng) and the mixture was incubated on ice for 5 minutes. The electroporator (Jensen/Flowgen) was set to 1.5kV, 132Ω, 40µF and the cells were pulsed with a time constant of about 4 - 4.5. The transformed cells were kept on ice till the rest of the cells had been electroporated. The cells were plated onto minimal media and incubated at 32°C for 4-6 days. Transformants began to appear after 5 days.

4.2.19(i) Isolation and analysis of haploid spores

In order to sporulate the cells, a single colony of the transformants was picked and streaked onto a plate of malt extract (ME) media. This was incubated at 30°C for 3 days. After this period, a loopful of the patch of cells was taken and suspended in 1ml of sterile distilled water containing 2% (v/v) helicase enzyme to strip away the ascus wall. This was then incubated at room temperature for 16 hours to release the spores into solution. The mixture was resuspended and an aliquot (100µl) of the resuspended spores was spread onto minimal media plates supplemented with adenine (0.25µg/ml). The plates were divided into two sets: those containing thiamine (5µg/ml) and those without. The rest of the spores were stored at 4°C. The plated spores were incubated at 32°C for 4 days and the haploids were distinguished from the diploids by patching the colonies onto

minimal plates containing a relatively low amount (10mg/litre) of adenine. The haploid (adenine$^-$) grew as pink patches while the diploids (adenine $^+$) grew as white patches. The cells were fixed on a microscope slide and the nucleus was stained with 4,6-diamidino-2-phenyllindole 2HCl (DAPI) fluorescent stain. The images were captured with a light microscope set at a magnificent of 10 x 100 (1000).

4.2.20(a) Activation of the Anaphase Promoting Complex (APC)

Active, phosphorylated APC binds to SpSuc1-sepharose beads but the inactive, unphosphorylated form does not (Sudakin *et al.*, 1997). It has been suggested from these findings that SpSuc1 may be targeting CDK1-Cyclin B kinase complex to the APC for the subsequent degradation of the cyclin B subunit. *In vivo*, the activity of the APC is enhanced by the addition of recombinant SpSuc1 to a mixture of CDK1-Cyclin B and the APC. In order to test the hypothesis that SpSuc1 binds to the APC via the charged cluster site, a new SpSuc1 mutant (Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala) was prepared, expressed, purified and sent together with the Ile94Ala:Leu96Ala double mutant to Dr. Michal Shteinberg in the laboratory of Prof. Avram Hershko (Technion-Israel Institute of Technology) for the APC activation studies. The Arg30Ala:Arg39Ala double mutant was also prepared but it did not express because of the introduction of spurious mutations.

4.2.20(b) Generation of Arg30Ala:Arg39Ala double and Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutants

In this case, a mutagenesis method (Fisher and Pei, 1997) based on the polymerase chain reaction (PCR) was employed to make the second batch of SpSuc1 mutants. This PCR-based method coupled with relatively fast automated DNA sequencing made the generation of mutants less laborious, time-consuming and more efficient than the Kunkel method (see section 4.2.3).

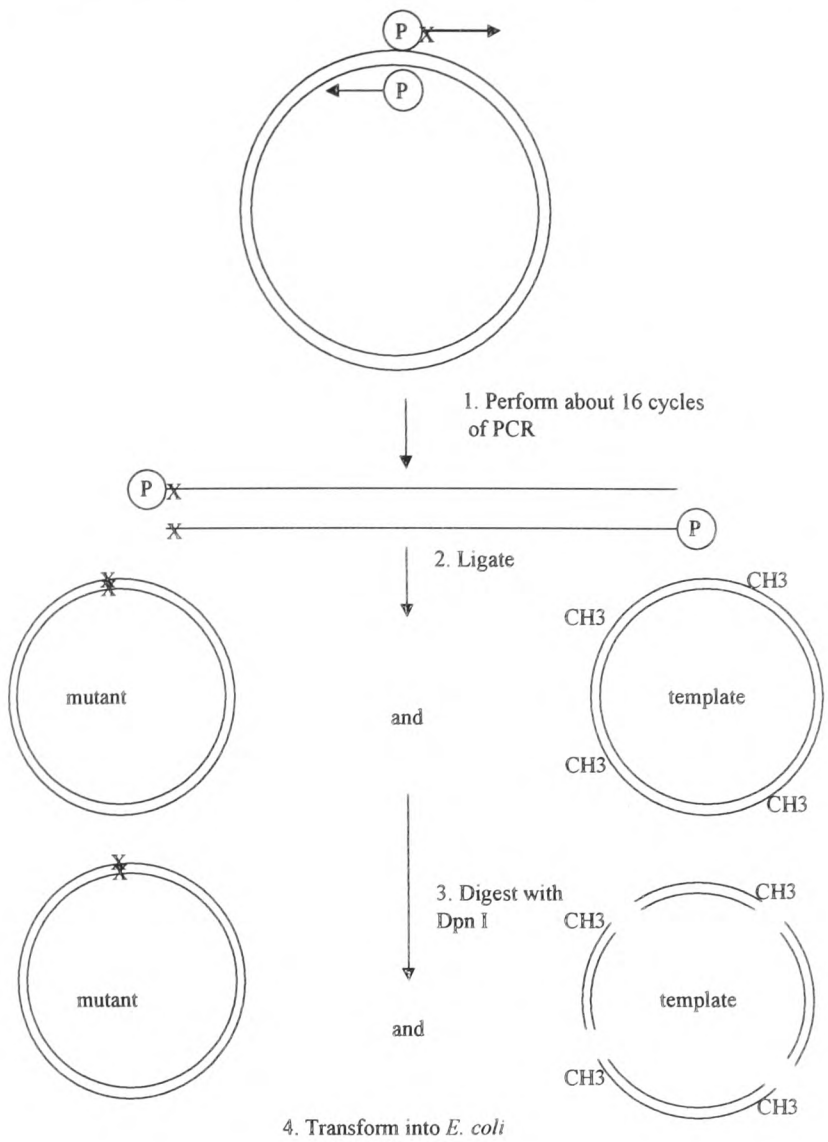


Figure 4.8 Illustration of the PCR-based site-directed mutagenesis protocol

Briefly, adjacent primers were designed on opposite strands so that the intended mutation was encoded at the 5' end of one primer. Both primers were 5' phosphorylated by the manufacturer (OCMS nucleotide sequencing facility). The PCR was carried out in a 100µl mixture using 1U Pfu DNA polymerase (Stratagene) with 25ng of template DNA and 50pmol of each primer, 200µM of each dNTP in the supplied Pfu reaction buffer [20mM Tris-HCl, pH 8.8, 10mM KCL, 10mM $(NH4)_2SO_4$, 2mM $MgSO_4$, 0.1% Triton X-100, 100µg/ml bovine serum album (BSA)].

The amplification parameters were an initial 4min at 94°C, followed by 16 cycles of 1 min at 94°C, 1 min at the annealing temperature of 58°C and 10 min (2.5min/1Kb) plus 5 seconds auto-extension at 72°C.

Primer	**Sequence (5' - 3')**
Forward	CCTGCGTATGCTGATGATGAATATG AATATGCTCATGTT
Reverse	TAGATAATGGATTTGATCAATGAAT GGCTCTAGACG

Table 4.3 Nucleotide sequence of the primers

The forward primer was designed to effect an Arg30Ala:Arg39Ala double mutation. Those codons have been underlined.

At the end of the PCR reaction, 17µl of 6x DNA loading dye was added to the reaction mixture (100µl) and it was ran on a 1% agarose gel to confirm the presence of the reaction product. The band was extracted with the QIAGEN gel extraction kit and a ligation reaction was set up as described in section 4.2.15 of this chapter. At the end of the ligation period, the samples were heat-inactivated by incubation at 65°C for 20 minutes. The samples were diluted to a final volume of 50µl in NEB buffer 4 (50mM potassium acetate, 20mM Tris-acetate, 10mM Magnesium acetate, 1mM DTT pH 7.9 with 60 U of DpnI (New England Biolabs) and incubated at 37°C for 90 minutes. The reaction mixture was then transformed into "ultra" competent DH5α *E. coli* cells and plated onto LB agar plate supplemented with ampicillin (100µg/ml). About 250-300 colonies were generated per sample using this protocol. Two colonies from each plate were picked and grown for DNA isolation. The purified plasmid DNA was sequenced by the automated dye-terminator method at the DNA sequencing facility (Sir William Dunn School of Pathology, University of Oxford).

Another advantage of this PCR-based method of introducing point mutations over the Kunkel method is that it is independent of the type of (cloning or expression) vector used thus removing the need to sub-clone from a cloning vector to an expression vector as was done with the Kunkel method. In this case, the mutaganesis was carried out on a SpSuc1 gene fragment in the pRK172 expression vector. Confirmed mutants were transformed into BL21(DE3)pLys-S and the protein was expressed and purified as described in section 4.2.16. The purified samples were packaged on dry ice and sent to the collaborators in Israel for the APC activation assay analysis.

4.2.20(c) Assay of Cyclin-Ubiquitin Ligase Activity

Dr. Michal Shteinberg (Technion-Israel Institute of Technology) conducted these experiments. This assay was based on the observation that SpSuc1 has the ability to enhance the activation of the APC by CDK1 i.e. the conversion of the APC from its inactive, interphase form to its active, mitotic form. The activity of the APC was assayed as its ability to ligate ubiquitin to Cyclin B in the presence of E1 and E2C. Briefly, reaction mixtures containing the following in a volume of 10µl: 40mM Tris-HCl (pH 7.6), 1 mg/ml rcm-BSA, 1mM DTT, 5mM $MgCl_2$, 10mM phosphocreatine, 50mg/ml

creatine phosphokinase, 50μM ubiquitin, 1 μM ubiquitin aldehyde, 1pmol of E1, 5pmol of E2-C, 1μM okadaic acid, enzyme (e.g. GST-88-cyclin B-CDK1, SpSuc1, APC), and 1-2pmol (~105 cpm) of ^{125}I-labeled cyclin B-(13-91)/protein A were prepared. It has been demonstrated that the N-terminal fragment of cyclin B (i.e. cyclin B-(13-91) is a good substrate for destruction box-specific, cell cycle-regulated ubiquitinylation and degradation of mitotic cyclins (Sudakin *et al.*, 1997). The samples were incubated at 18°C for 90 minutes. The samples were agitated during incubation when enzyme activity associated with Sepharose beads was determined. At the end of the incubation period, the samples were subjected to electrophoresis on a 12.5% polyacrylamide-SDS gel. The reaction product was quantified with a PhosphorImager and the amount of radioactivity in all cyclin-ubiquitin conjugates was expressed as the percentage of total radioactivity in each lane. The reactions were conducted in the range that was linear with enzyme concentration; this was 5-40% ^{125}I-cyclin ligated to ubiquitin. A unit of ligase activity was defined as that converting 1% ^{125}I-cyclin to ubiquitin conjugates under the reaction conditions described above.

4.2.20(d) Western blot analysis of SpSuc1 mutants

The identities of the first batch (Glu86Ala, Asp54Asn, Ile94Ala:Leu96Ala) of SpSuc1 mutants were confirmed by N-terminal protein sequencing using the service at the MRC immunochemistry unit, University of Oxford. The second batch (Pro90Ala, SpSuc1Δ87-89 and Ile94Ala:Leu96Ala:Arg30Ala:Arg39Ala) were confirmed as the correct mutants by full-length DNA sequencing. The Pro90Ala (Rousseau *et al.*, 1998) and SpSuc1Δ87-89 mutants were constructed in the laboratory of Dr. Laura Itzhaki (University of Cambridge) and sent to Dr. Jane Endicott (University of Oxford) for structural studies. I expressed, purified and characterised their binding interaction with GST-CDK2 via BIAcore analysis. The introduction of a third column at the end of the purification revealed the presence of a potential dimeric species for the Ile94Ala:Leu96Ala:Arg30Ala:Arg39Ala quadruple mutant. A comparative western blot analysis was carried out on the second batch of SpSuc1 mutants in order to confirm the identity of this putative dimeric species. Protein samples were resuspended in sample buffer [2X stock consisted of 0.5M Tris-HCl pH 6.8, 4.4% (w/v) SDS, 20% (v/v)

glycerol, 2% (v/v) mercaptoethanol and 0.01mg/ml Bromophenol blue) and ran on an SDS-PAGE gel at a constant current of 20mA. At the end of the electrophoresis, the gel was carefully removed and equilibrated in blot buffer (25mM Tris-HCl pH 8.3, 150mM glycine, 0.05% SDS, 5% methanol) for 5 minutes. The gel was placed onto a wet immobilion membrane and both the gel and membrane were sandwiched on each side by 3 filter papers that had undergone prior immersion in blot buffer. All the air bubbles were gently squeezed out of the blot sandwich. The blot was ran at a constant current of 100mA for 60 minutes at room temperature in order to transfer the protein from the gel to the membrane. To block non-specific binding sites, the membrane was incubated in 15-20ml of 3% BSA in PBS for 45 minutes. This was followed by another incubation of the membrane in 10ml of 3% BSA/PBS to which a primary antibody, anti-SpSuc1 had been added at a 1:500 dilution. The membrane was washed thrice in phosphate-buffered saline (PBS) solution. and the bound primary antibody was detected by a 1:10000 dilution of secondary antibody i.e. goat anti-rabbit IgG alkaline phosphatase. After washing thrice with PBS, the membrane was developed with 10ml of NBT/BCIP. When the bands were visible enough, the reaction was terminated by washing the blot in PBS containing 20mM EDTA to chelate the Mg^{2+} ions.

4.3 RESULTS

4.3.1 Identification of mutant clones

With the Kunkel method of mutagenesis, the introduction of silent restriction sites in some of the mutagenic primers facilitated the distinction of mutant clones from non-mutants by comparative restriction analysis (data not shown). Dideoxy DNA sequencing (Sanger *et al.*, 1977) was used to confirm that they were the expected mutants and also, to check for any aberrant mutations that may have occurred.

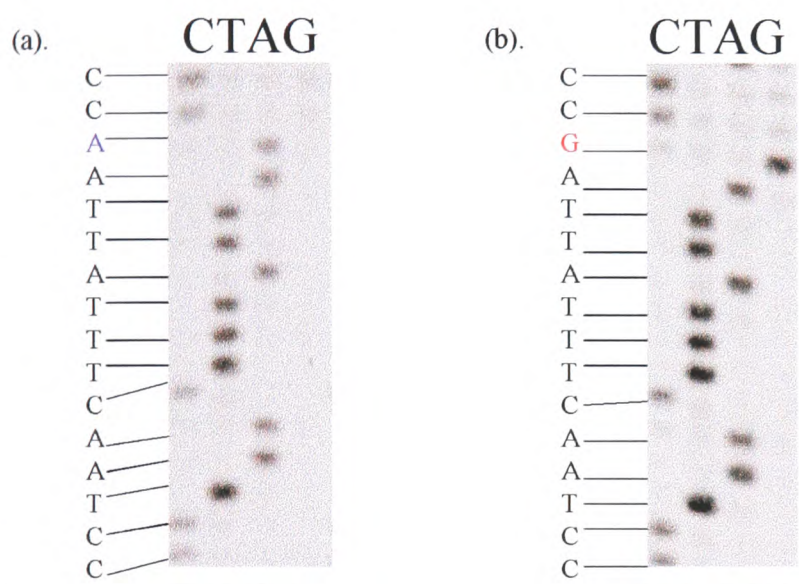


Figure 4.9 Manual DNA sequencing of (a) Asp54Asn mutant and (b) wild-type SpSuc1

The reactions were prepared as described in the preceding section. Two and a half microlitres (2.5µl) of each reaction mixture was loaded onto a pre-heated (20 minutes) Tris-borate EDTA (TBE) gel (7%) and ran at 1800V. The codon for Asp (TAG) has been altered to TAA for Asn. No new restriction site was introduced in this case. Four other different mutants were confirmed at the DNA level by manual DNA sequencing. They are Glu86Ala, Arg30Ala, Ile94Ala:Leu96Ala double mutant and Asp33Ala:Asp34Ala:Glu35Ala triple mutant. The introduction of silent restriction sites allowed for the identification of putative mutant clones by comparative restriction digests.

The expression and purification protocol of the mutants was essentially unchanged from that of the wild-type SpSuc1. Two mutants (Arg30Ala and Asp33Ala:Asp34Ala:Glu35Ala triple mutant) did not express at all. It is possible that the mutagenesis reaction introduced spurious mutations (such as stop codons) into the DNA fragment of these mutants.

4.3.2 Purification of mutant SpSuc1 proteins

The first separation column in the purification protocol for SpSuc1 protein was an S100 gel filtration column. A clarified supernatant from a 3-litre culture was applied to the pre-equilibrated column and the various fractions were collected.

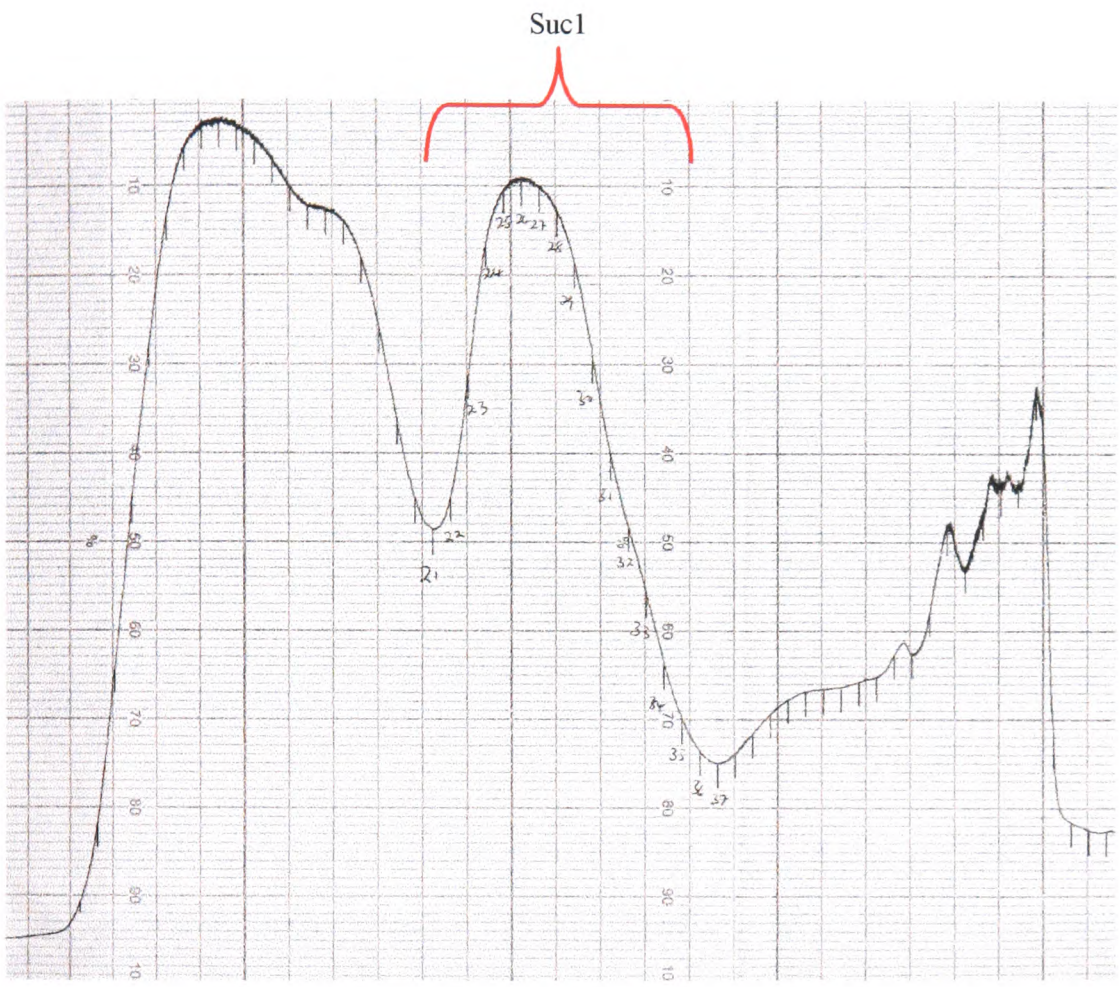


Figure 4.10a A typical SpSuc1 protein elution profile from the S100 gel filtration column

The chart-recorder was set to full-scale deflection of 1.0 and the flow-rate was 1ml/min. The middle peak represents the SpSuc1-containing peak.

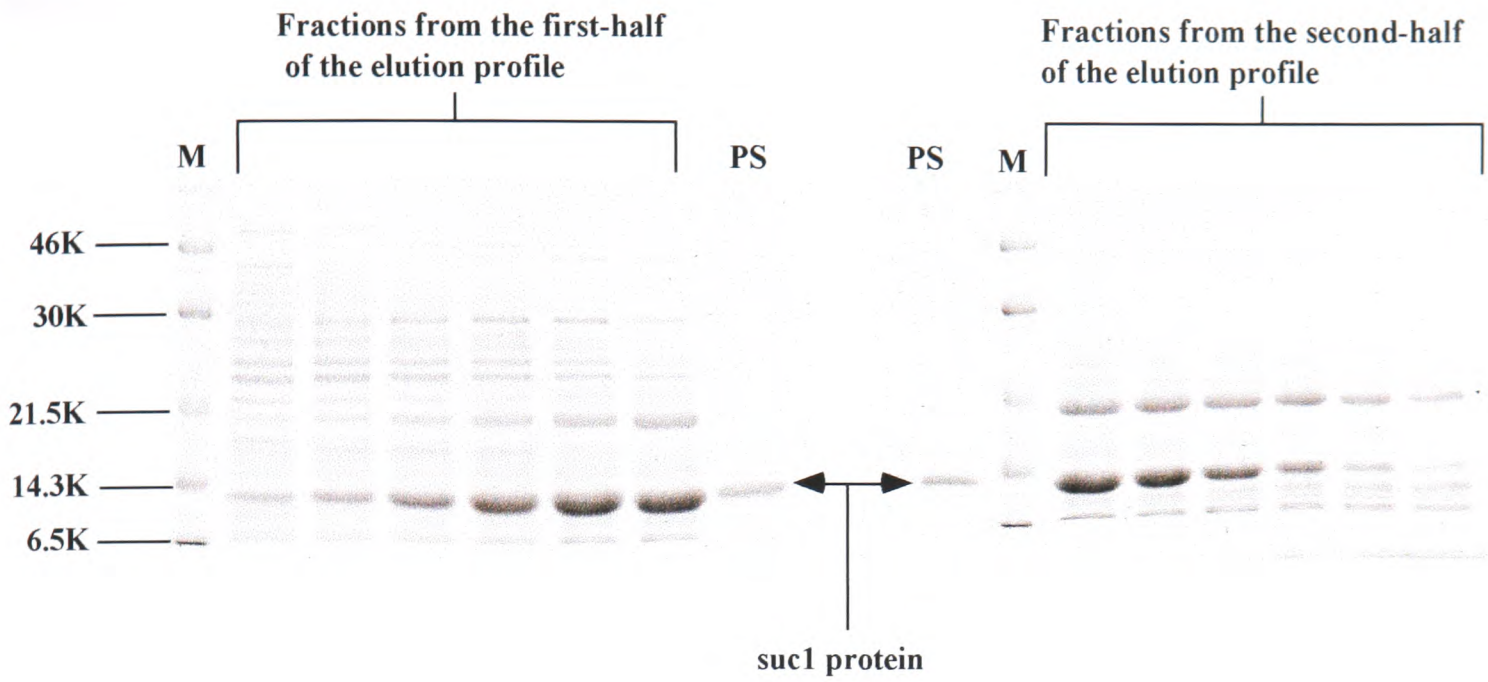


Figure 4.10b Expression and purification of Asp54Asn mutant

Fractions from S100 (Pharmacia) gel filtration column ran on a 10% Tris-tricine protein gel and stained with coomassie blue. M = molecular weight markers, PS = purified wild-type SpSuc1 protein. The double arrow indicates the expected position of a 13KDa protein.

The SpSuc1-containing fractions were pooled and loaded onto a hydrophobic interaction chromatography (H.I.C.) column and the bound protein was eluted with a decreasing gradient of ammonium sulphate (0.85 – 0M).



Figure 4.11 Eluants of Asp54Asn from the H.I.C. column

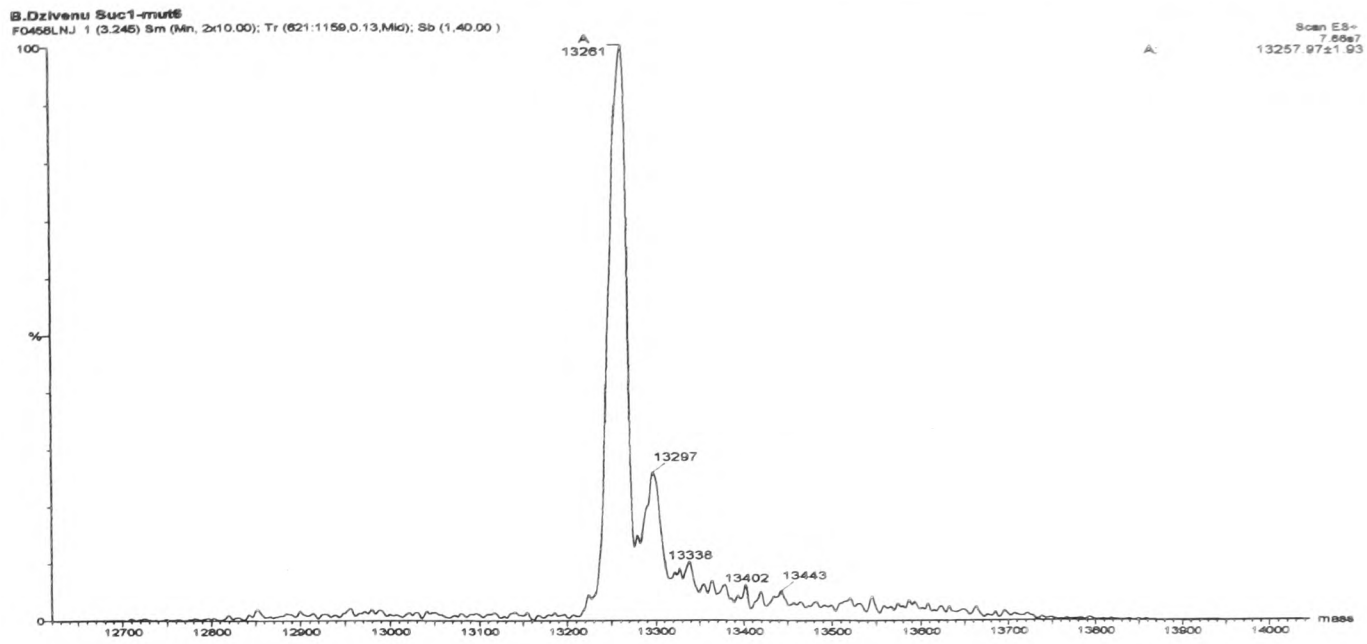
Bound Asp54Asn was eluted at 0.30-0.40M $(NH_4)_2SO_4$. The pooled fractions were dialysed into 10mM Tris-HCl pH 7.4 before concentrating to ≈500µl. The yields as judged by a Pierce assay were about 18-30mg from a 6L-culture for all the mutants.

The level of expression of the mutants was relatively poor with respect to the wild-type protein. The yield for the wild-type protein was about 20mg from a 3-liter culture. Perhaps the conditions for expressing the wild-type protein were not optimal for the mutant proteins.

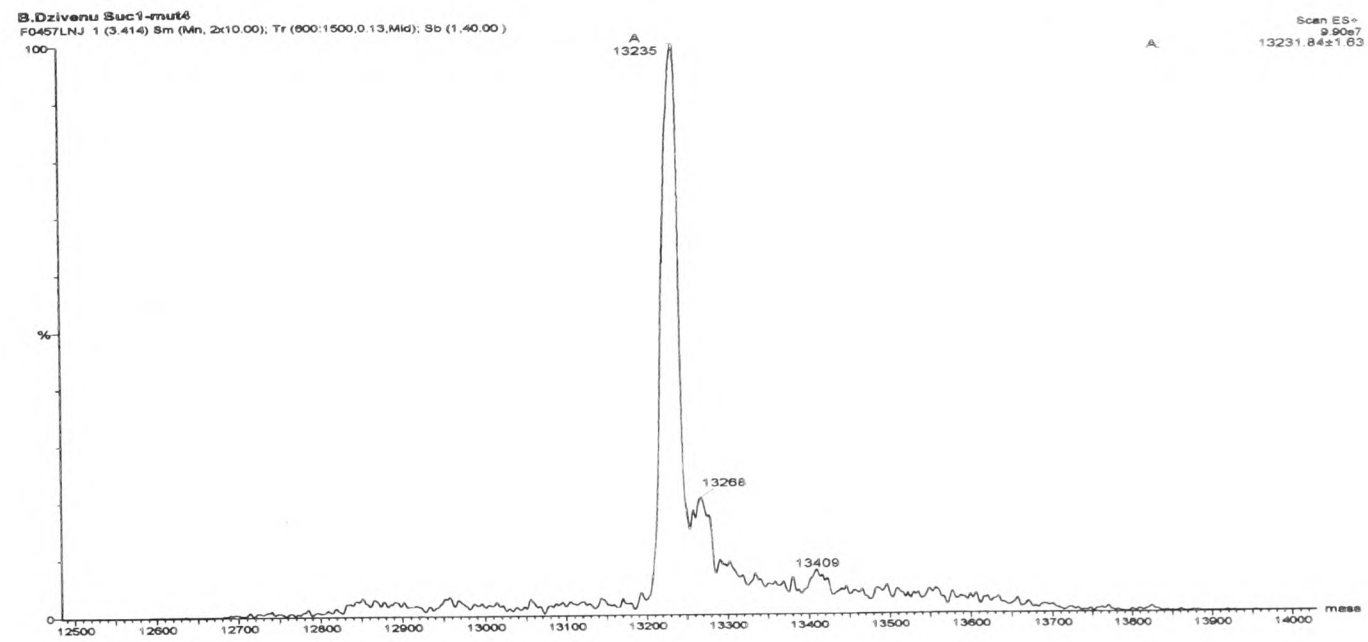
4.3.3 Electrospray mass spectrophotometry (ES MS)

ES MS has become a useful technique for the analysis of recombinant proteins. The protein can be analysed either as a purified intact molecule or as a mixture of polypeptides e.g. after a limited protease digestion (Mann and Wilm, 1995; Robinson and Radford, 1995). Due to the accuracy to which the mass of the protein can be determined (to 0.01% of the mass of a protein), ES MS was particularly useful in this case for confirming that the intended mutations had occurred at the DNA level. This was particularly important in the case of the first round of mutants where a manual DNA sequencing method was used to sequence only a limited region of the gene (i.e. region where the mutation was effected). Another advantage of ES MS is that it gives a good indication of the state of homogeneity or heterogeneity of the sample.

(a).



(b).



(c).

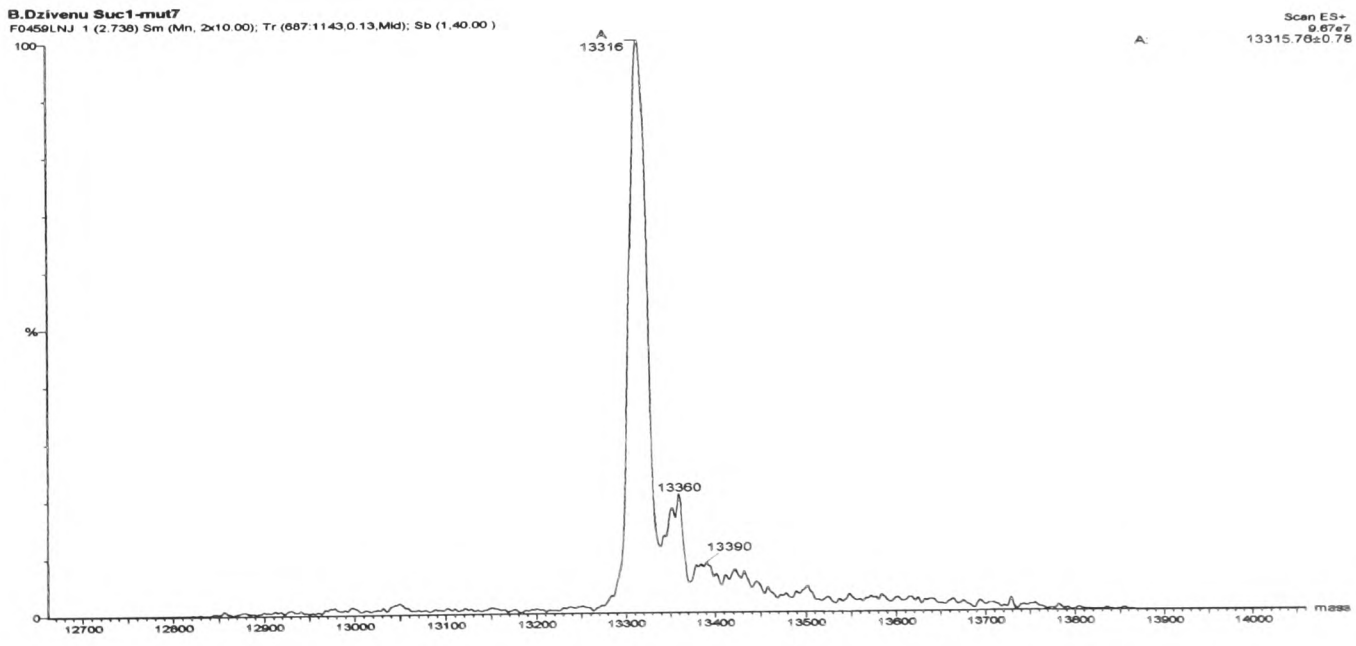


Figure 4.12 Electro-spray mass spectrum of SpSuc1 mutants

(a) Glu86Ala mutant. Observed mass = 13258, expected mass = 13257 daltons. (b) Ile94Ala:Leu96Ala double mutant, observed mass = 13232 daltons, expected mass =

13231daltons (c). Asp54Asn mutant. The theoretical mass of Asp54Asn is 13314 daltons which is almost identical to the observed mass of 13315 daltons. The mass of the wild-type protein was 13316 daltons. In all cases, it is assumed that the first residue (Met) has been lost just as it was with the wild-type protein (Endicott *et al.*, 1995).

4.3.4 Circular dichroism spectroscopy

To check whether the structural integrity of each mutant protein had been retained or not, the wild-type protein and the various mutants were compared by circular dichroism (CD) spectroscopy. Samples (1mg/ml) were dialysed into 10mM $NaPO_4$ pH8.0 and 0.2mg/ml of each sample was analysed on a Jasco J720 machine (Fig. 4.8).

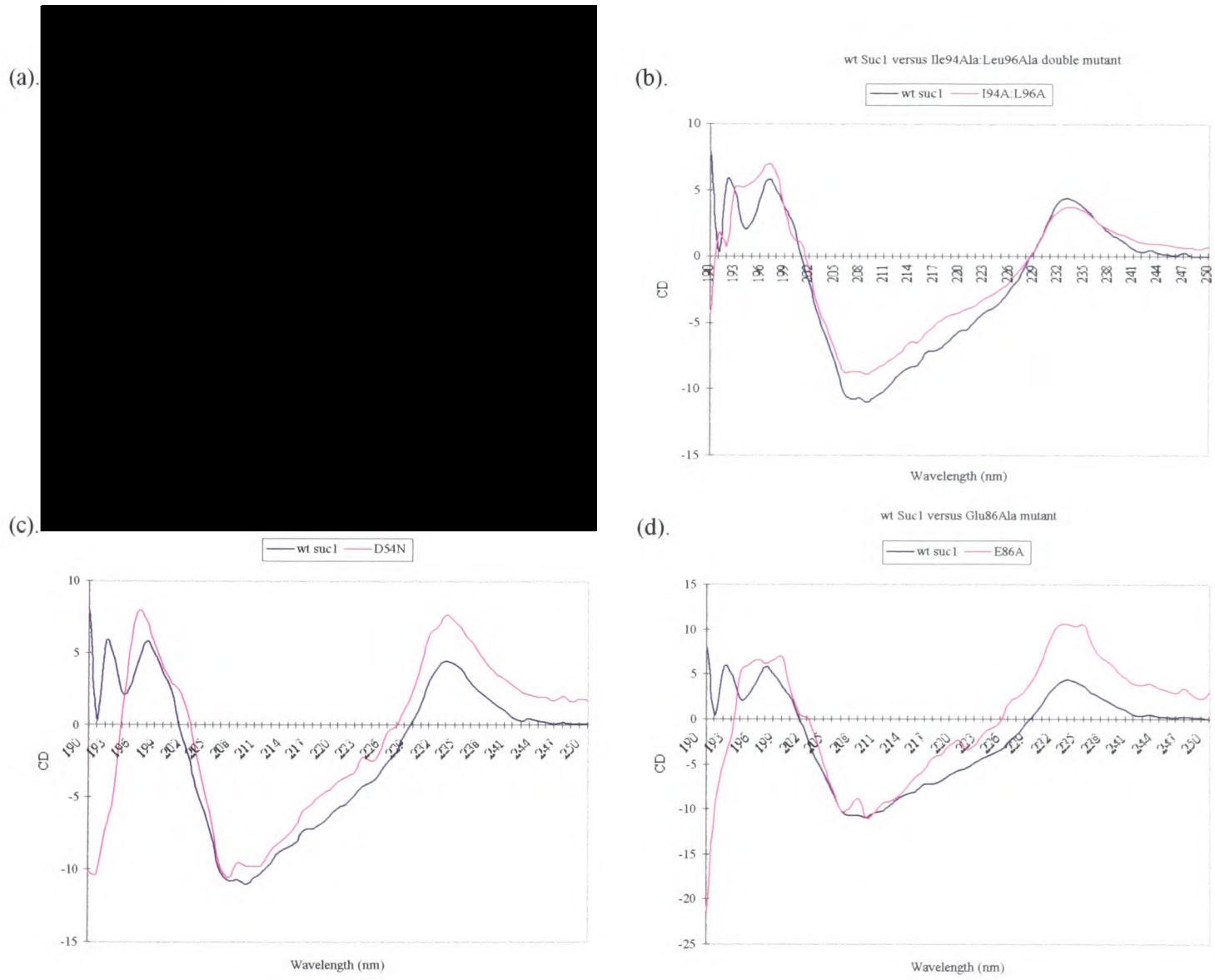


Figure 4.13 Comparative CD spectra of wild-type SpSuc1 and the mutants

(a) CD spectra of the α-helix, β-sheet, and random-coil forms of a model polypeptide poly-L-lysine. It has been reproduced from (Campbell and Dwek, 1984) *to assist in*

gauging the contributions to the spectra by α-helices, β-sheets and random-coils. (b) Ile94Ala:Leu96Ala double mutant wild-type SpSuc1. (c) Asp54Asn versus wild-type SpSuc1. (d) Glu86Ala mutant versus wild-type SpSuc1. Apart from the Ile94Ala:Leu96Ala spectrum, adjustments were made to each of the mutant spectra to account for slight differences in protein concentration.

As seen in figure 4.13, the CD spectrum of each mutant is similar to that of the wild-type protein, indicating that the intended site-specific alterations have not led to a gross conformational distortion. This result is not unexpected considering the fact that all the altered amino acid residues are exposed, surface residues, which are not crucial to the maintenance of the protein structure. A structural basis for this assertion was provided in Chapter 3 where the mutation of two surface residues (Ser13 and Ala32) did not perturb the global structure of the protein.

The crystal structure of SpSuc1 shows that the protein consists of four α-helices and a four-stranded β-sheet (Endicott *et al.*, 1995 and chapter 2). The CD spectra of the wild-type protein and the mutants reflect this dominance of α-helical structures (Fig. 4.13). In general, a prominent positive absorption at ≈198nm and an intense negative absorption at ≈208nm and ≈222nm, respectively, are indicative of absorption by α-helical structures (van Stokkum *et al.*, 1990). A β-sheet also displays a fairly intense positive absorption at ≈198nm (Johnson, 1990). Inspection of the spectra does not reveal a negative peak at ≈222nm; since these spectra are a composite of the absorptions by all the secondary structures present in the protein, it is possible that this peak has been masked by a positive absorption by another structure e.g. a random coil or aromatic residues. Preponderance of aromatic residues in SpSuc1 is a feature that has already been noted elsewhere in the thesis: the side-chains of these residues are normally optically inactive but they can be induced to absorb at ≈250-310nm (Campbell and Dwek, 1984). The intense positive peak at ≈235nm reflects the absorption by the cluster of 7 tyrosine residues in SpSuc1. In general, the CD result is identical to a latter, independent publication containing the CD spectrum of wild-type SpSuc1 (Rousseau *et al.*, 1998).

4.3.5 Zinc-induced dimerisation of SpSuc1

To determine the oligomerisation state of the mutants and compare it to the wild-type protein, each sample was dialysed into two different buffers and analysed by FPLC on either a Superose-12 or a Superdex-75 column (Pharmacia).

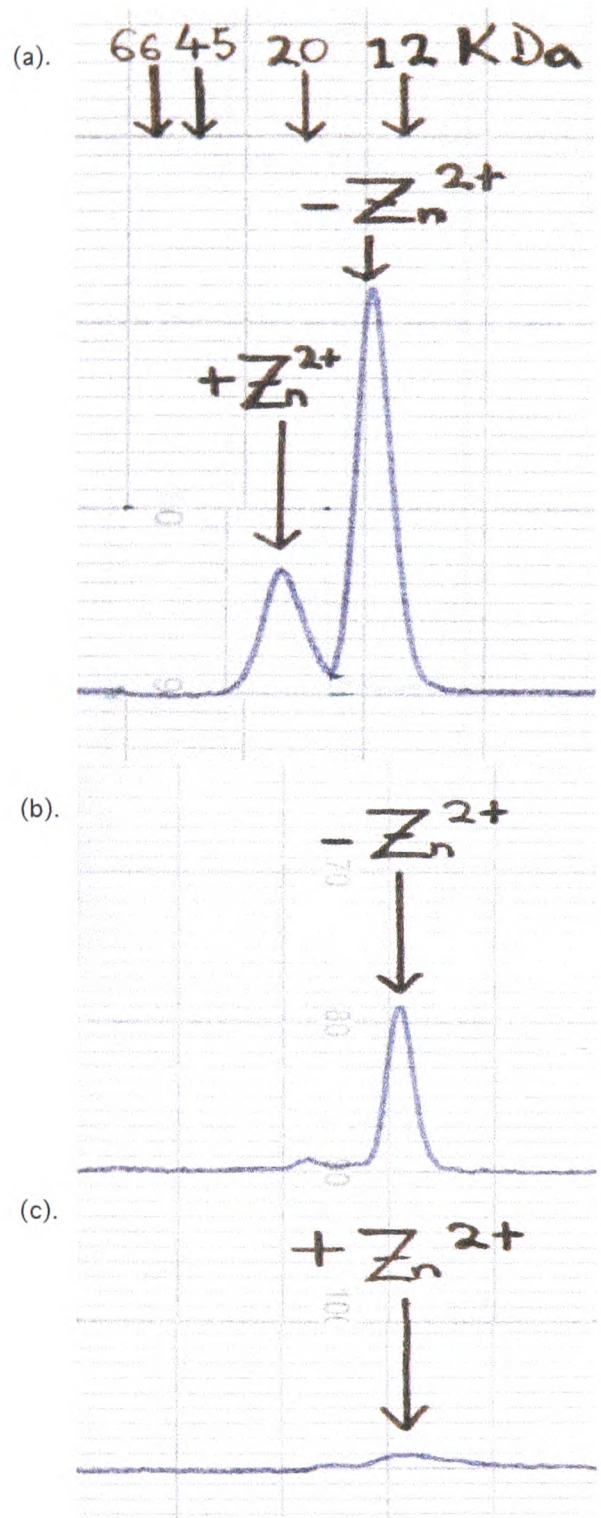


Figure 4.14 Zinc-induced dimerisation of SpSuc1 *(a) The oligomerisation state of wild-type SpSuc1 in the presence and absence of Zn^{2+} has been established* (Endicott *et al.*, 1995). *The elution profile of the wild-type SpSuc1 protein in the absence of zinc ions is monomeric but in the presence of half-equimolar zinc ion concentration, it forms a dimer. The elution profile of the wild-type protein in the presence and absence of Zn^{2+} have been superimposed to depict the shift from monomer to dimer and (b) The profile of the Ile94Ala:Leu96Ala double mutant in the absence and (c) presence of half-equimolar Zn^{2+}. In this case, the double mutant precipitated in the presence of half-equimolar zinc ions. Like the wild-type protein, Glu86Ala, Asp54Asn dimerised in the presence of half-equimolar zinc (chromatogram not shown). The mass of SpSuc1 as judged by SDS-PAGE and electrospray MS was 13KDa but according to the calibration markers the size of monomeric SpSuc1 was ≈17KDa and the corresponding dimer was ≈34KDa. The reason for this anomaly is related to the shape of the protein molecule.*

The preceding results enabled one to proceed with the preliminary binding (BIA) and functional (HK) assays. The former allowed for the determination of the existence -

if any - and the strength of binding interactions between the mutant proteins and GST-CDK2 while the latter could provide an answer to the question, "Does a reduced or non-existent binding affinity between a particular mutant and GST-CDK2 correlate with the inability of that particular mutant to stabilise CDK1-33^{ts} + cyclin B interactions?"

4.3.6 Binding of GST-CDK2 to SpSuc1 mutants

Dr. Nick Brown (University of Oxford) kindly donated the GST-CDK2 protein used in the preliminary biacore binding studies. In subsequent studies, Dr. Alison Lawrie kindly donated CDK2 prepared from insect cells and GST-3C-CDK2 was prepared as described in section 4.2.17(f). For the preliminary binding experiments, SpSuc1 and the various mutants were covalently immobilised in different flowcells via free amine groups to the carboxymethylated hydrogel. The immobilisation process produced a response of about ≈250RU in each case (data not shown) indicating* that about 0.3μM of the ligand had been linked to the sensor chip. GST-CDK2 was passed over the ligand at a flow rate of 10μl/min and the resulting sensogram was evaluated using BIAevaluation software v2.1 (Pharmacia Biosensor, Sweden). A control reaction in which the analyte was passed over a naked sensor chip was also performed. Initially, each binding experiment was performed at three different concentrations of analyte (7.6μM, 3.8μM and 1.9μM) but only data for 7.6μM of analyte has been reported (Fig. 4.15). Apart from an expected decrease in the SPR profiles, the response at the lower concentrations of analyte yielded identical affinity values (data not shown).

*A RU of 60000 corresponds to 1g/l of immobilised protein (Kalinin *et al.*, 1995).

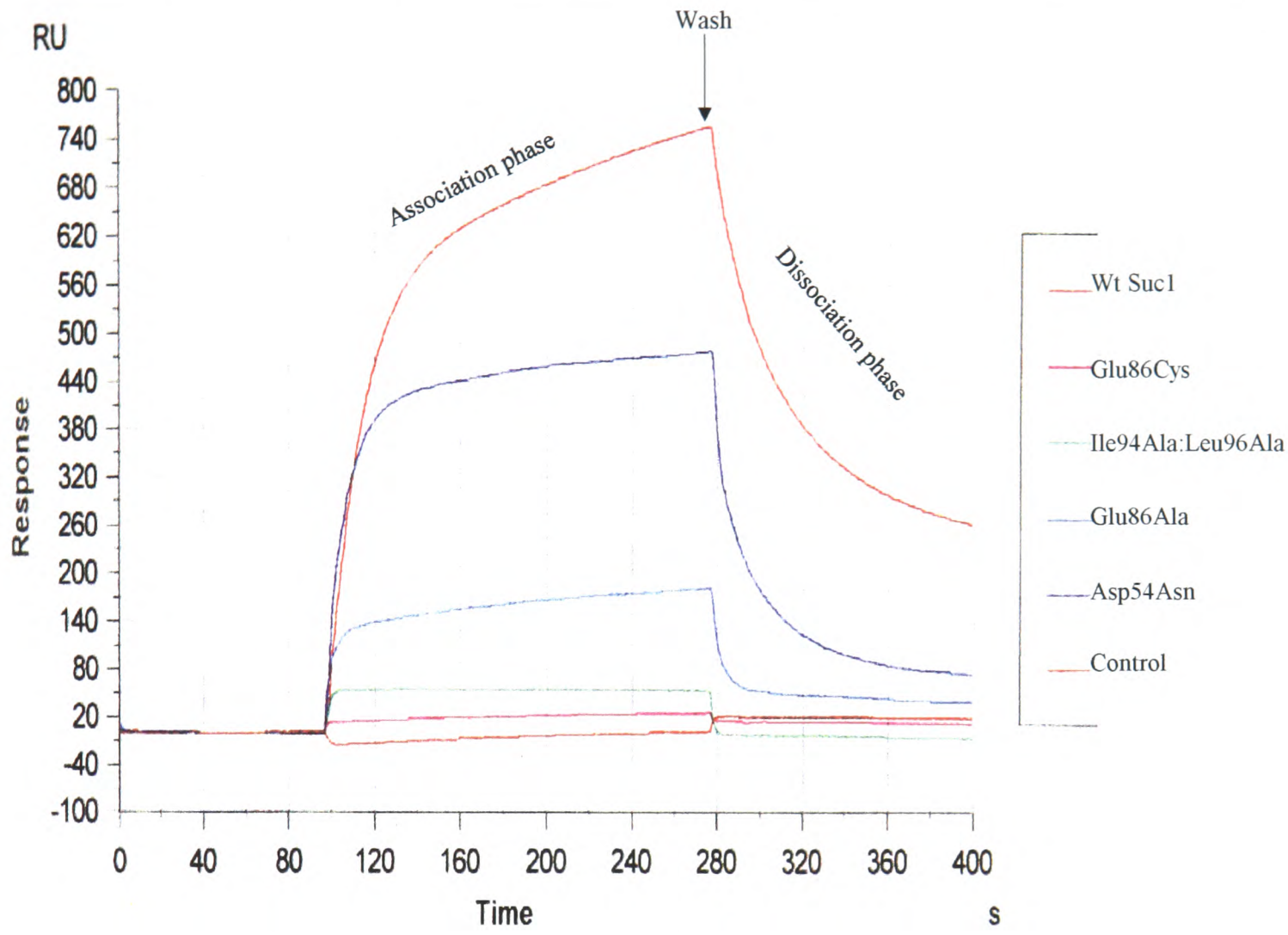


Figure 4.15 Overlay plot of SPR profiles obtained from binding interactions between the ligand (wild-type SpSuc1 and mutants) and the analyte (GST-CDK2)

GST-CDK2 + wild-type SpSuc1; Apparent K_d = 67nM.

GST-CDK2 + E86C; Apparent K_d was negligible.

GST-CDK2 + Ile94Ala:Leu96Ala; Apparent K_d was negligible.

GST-CDK2 + Glu86Ala; Apparent K_d = 84nM;

GST-CDK2 + Asp54Asn; Apparent K_d = 60nM.

A control reaction with GST-CDK2 only over a naked sensor chip was performed. K_d = k_{OFF}/k_{ON} and $K_d = 1/K_a$ (see text for definitions). k_{ON} and k_{OFF} rates were calculated from (140-240)s and (300-380)s, respectively.

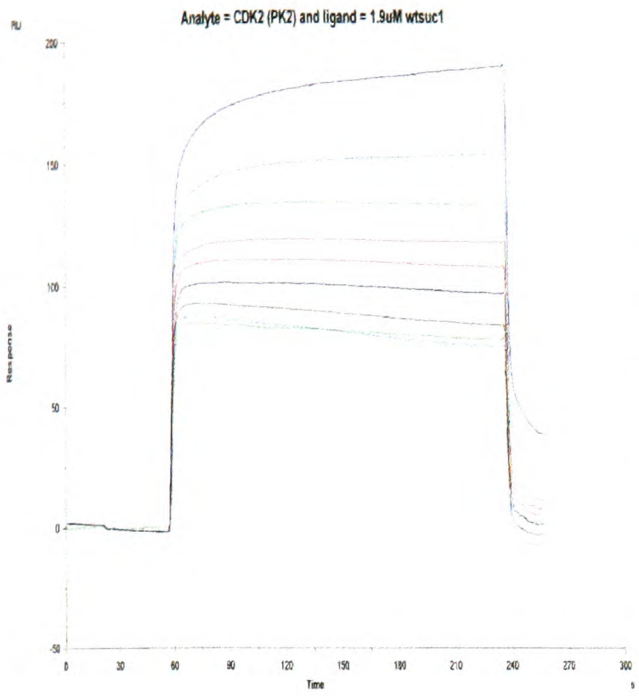
There are two ways of obtaining kinetic constants from BIAcore experiments; the first is to use the relationship $K_d = k_{OFF}/k_{ON}$ where K_d is the equilibrium dissociation constant and k_{ON} is the rate at which analyte binds to the ligand on the chip and k_{OFF} is the rate at which analyte can be removed from the chip with a buffer solution. The inverse of K_d gives K_a - the association constant. The underlying assumption in this

method is that the k_{ON} and k_{OFF} rates can be accurately determined from the sensogram. Factors such as the flow rate, concentration of analyte (GST-CDK2) and the amount of ligand immobilised on the hydrogel may have adverse effects on the accuracy of k_{ON} and k_{OFF} rates. The boundary between the arrival of bulk buffer and real time biospecific events is poorly defined in sensograms and this gives rise to bulk refractive index effects. In cases where the association event is extremely fast, the K_a value is limited by the arrival of free analyte on the chip. The second and more reliable method is to measure the RU at equilibrium at several analyte concentrations and perform a Scatchard analysis on the data. This has the added advantage of ascertaining the appropriateness of the model one has chosen to fit the data (Ladbury *et al.*, 1995; Lee *et al.*, 1995)

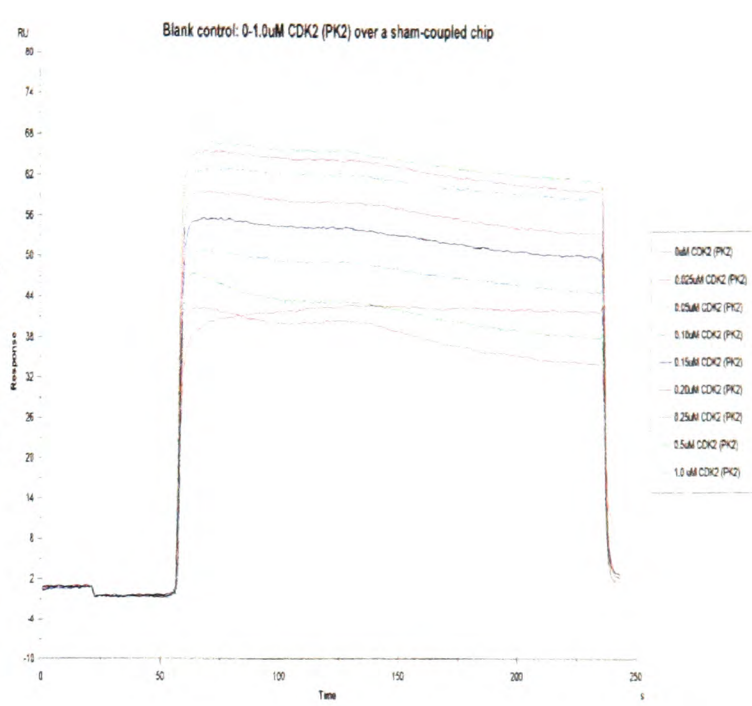
The preliminary experiments were aimed at obtaining comparative measurements of the binding of the different mutants. The preliminary curves showed that during the time of the experiment, the binding of GST-CDK2 to wild type SpSuc1 had not reached saturation. Kinetic analysis suggested an approximate K_D value of 67nM. The binding of GST-CDK2 to Asp54Asn showed rather different kinetics; it yielded an approximate K_D value of 60nM, similar to that for the wild-type SpSuc1. Asp54 is removed from the hydrophobic CDK2 binding site and evidently, the mutation has had only a minimal effect on binding. Asp54 is involved in an external salt-bridge linkage with Arg8 but the interaction as judged by the crystallographic model is not so strong as to be indispensable to the protein. Arg8 is sandwiched between Glu14 and Asp54 forming hydrogen bond interactions with both residues. Arg8 and Asp54 are about 3Å apart and there are two hydrogen bond interactions between these two residues (Endicott *et al.*, 1995 and chapter 2). One of these hydrogen bonds is expected to be retained by Asn54 in the Asp54Asn mutant. So, instead of four hydrogen bond interactions between Asp54-Arg8-Glu14 in wild-type SpSuc1, there would be three between Asn54-Arg8-Glu14 in the Asp54Asn mutant in an exposed part of the protein. The aim of this alteration was to create a temperature-sensitive mutant of SpSuc1. In the histone kinase assay, Asp54Asn stabilised CDK1-33^{ts} protein at the elevated temperature of 40°C just as well as the wild-type protein (data not shown). So, like other groups, (Basi and Draetta, 1995) an attempt to create a temperature-sensitive mutant of SpSuc1 was unsuccessful - perhaps a testament to the stability of this fission yeast cell cycle regulatory protein.

The Glu86Ala mutant showed less binding than wild type or Asp54Asn. The kinetic analysis yielded an apparent K_D of 87nM. In the crystal structure of SpSuc1, Glu86 binds to Zn3 at a crystal lattice contact site. Evidently, a mutation at this site had some effect on the ability to bind to GST-CDK2. This residue is situated at the same side of the molecule as the hydrophobic patch but it is not in the putative CDK binding region. In contrast, the Glu86Cys mutant showed no binding at all. In comparison with Glu86Ala mutant, the characteristics of the Glu86Cys mutant could be attributed to the effect of the cysteine residue rather than the absence of a crucial Glu86 residue. The Ile94Ala:Leu96Ala double mutant showed just a small amount of binding but the kinetic analysis was unable to produce a K_D value. Both residues are situated in the putative CDK binding site and hence, a more extensive BIAcore analysis was performed with this mutant. In the first of these latter BIAcore experiments, the aim was to overcome the disadvantages of using a GST-fusion protein. These disadvantages are discussed in section 4.4.3. Native CDK2 produced by the baculovirus expression system was used as the analyte. Wild-type SpSuc1 protein was immobilised as before and a range of concentrations of the analyte (CDK2) was passed over the immobilised ligand (Fig. 4.11). Even at the highest concentrations of analyte (5μM) and ligand (7.6μM), there was no detectable binding. Repetition of this experiment using fresh protein samples yielded the same result. In a different experimental set up, a GST-antibody was immobilised on a BIAcore sensor chip and the GST moiety of GST-CDK2 was captured by the antibody. A range of several concentrations of up to 10μM SpSuc1 protein was passed over the ligand and once again, there was no detectable binding. A control reaction in which cyclin A was passed over the same chip produced an expected binding response (data not shown). In figures 4.17-4.19, the experimental set-up was reverted to that used for the preliminary BIAcore experiments (fig. 4.15) and this resulted in the expected response units. Evidently, the GST tag helps promote association although GST by itself does not bind to immobilised SpSuc1 (data not shown).

(a)



(b)



(c)

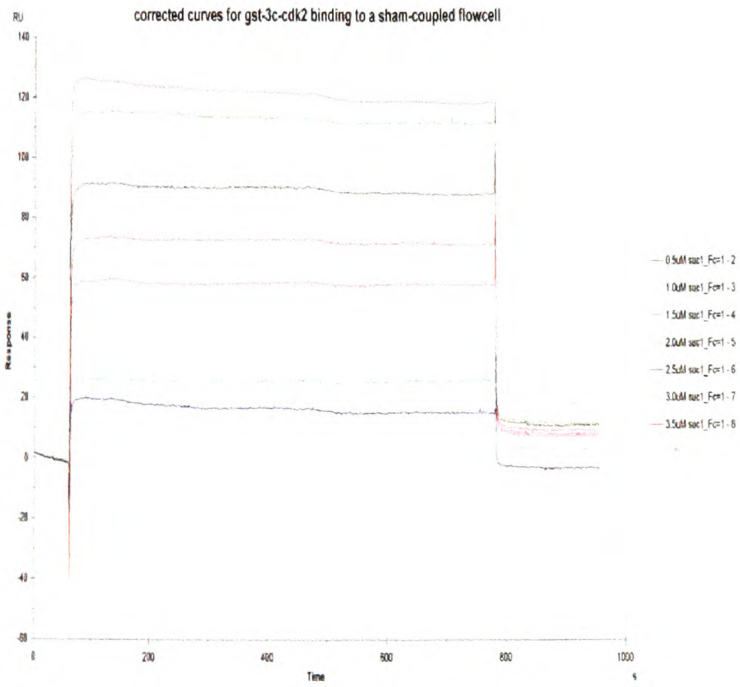


Figure 4.16 CDK2 binding to immobilised wild-type SpSuc1

(a) A range (0-1μM) of CDK2 was passed over immobilised SpSuc1 (1.9μM). A comparison with the blank control reaction results (b) showed just a marginal increase in the response units (RUs). The responses (identical to bulk refractive index), suggests that

there was no binding interaction between these two proteins. (c) shows the corrected curves for GST-3C-CDK2 for the range of 0.5 – 3.5μM. Even increasing the highest concentration for the CDK2 to 5μM did not show any binding interaction between the two proteins (data not shown).

In the subsequent set of BIAcore experiments, systematic attempts were made to obtain kinetic values for the interaction between GST-3C-CDK2 and wild-type SpSuc1 and the Ile94Ala:Leu96Ala double mutant using Scatchard analysis. A range of concentrations (1.9, 3.8 and 7.6μM) of the ligand wild-type or mutant were immobilised in the four different flowcells of a sensor chip and a range (0 - 3.5μM) of analyte concentrations were passed through each flowcell. To correct for the contribution of bulk refractive index differences to each sensogram, the blank control for each concentration of analyte was subtracted from each reaction curve.

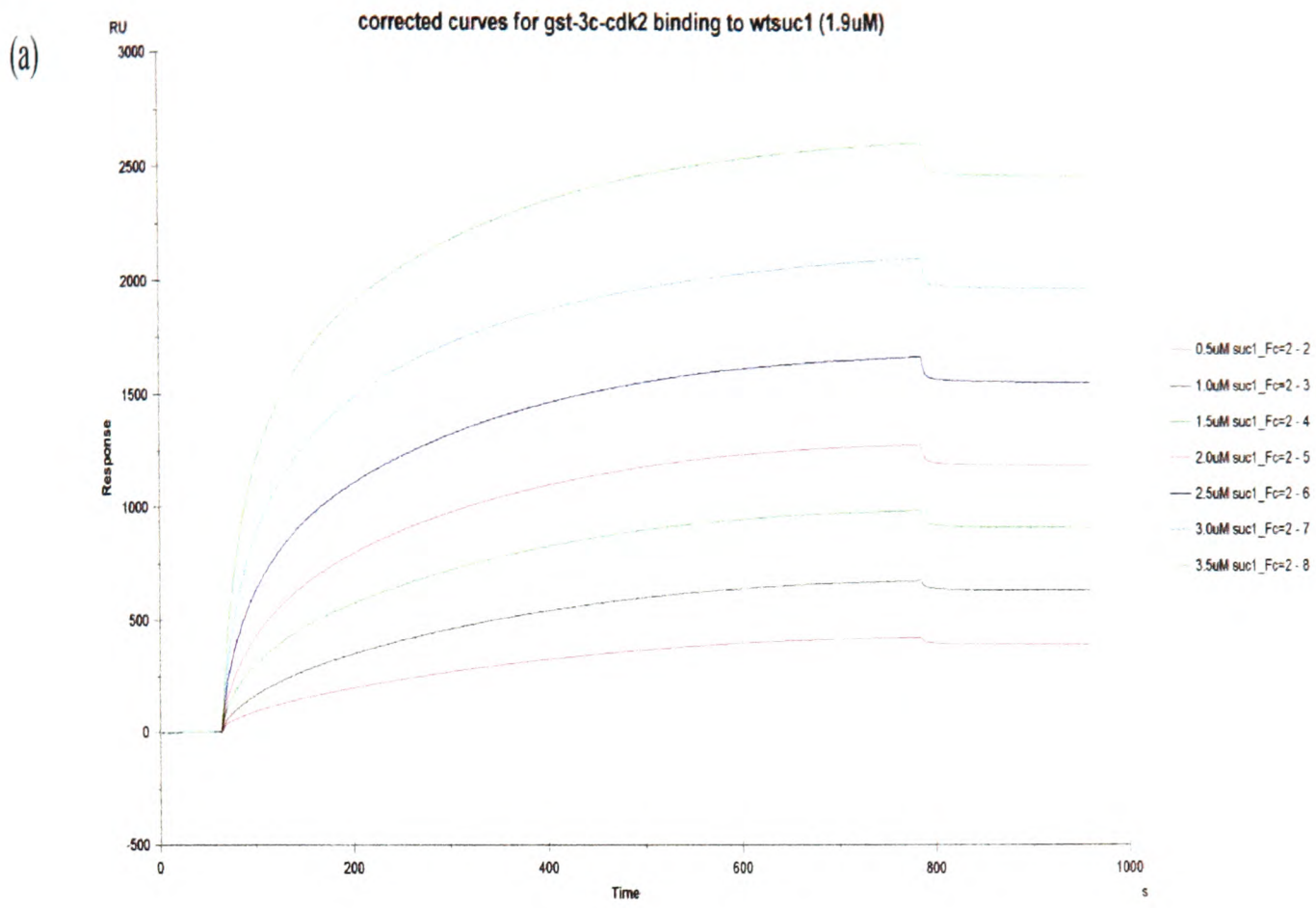


Figure 4.17 Binding of GST-3C-CDK2 to immobilised wild-type SpSuc1 (1.9μM)

(a) corrected sensograms showing that there was a binding interaction between SpSuc1 and the analyte at every concentration that was tested. Scatchard plots for (a) 1:1 (CDK2:SpSuc1) binding and (c) 2:1 (CDK2:SpSuc1) binding models. In both cases the

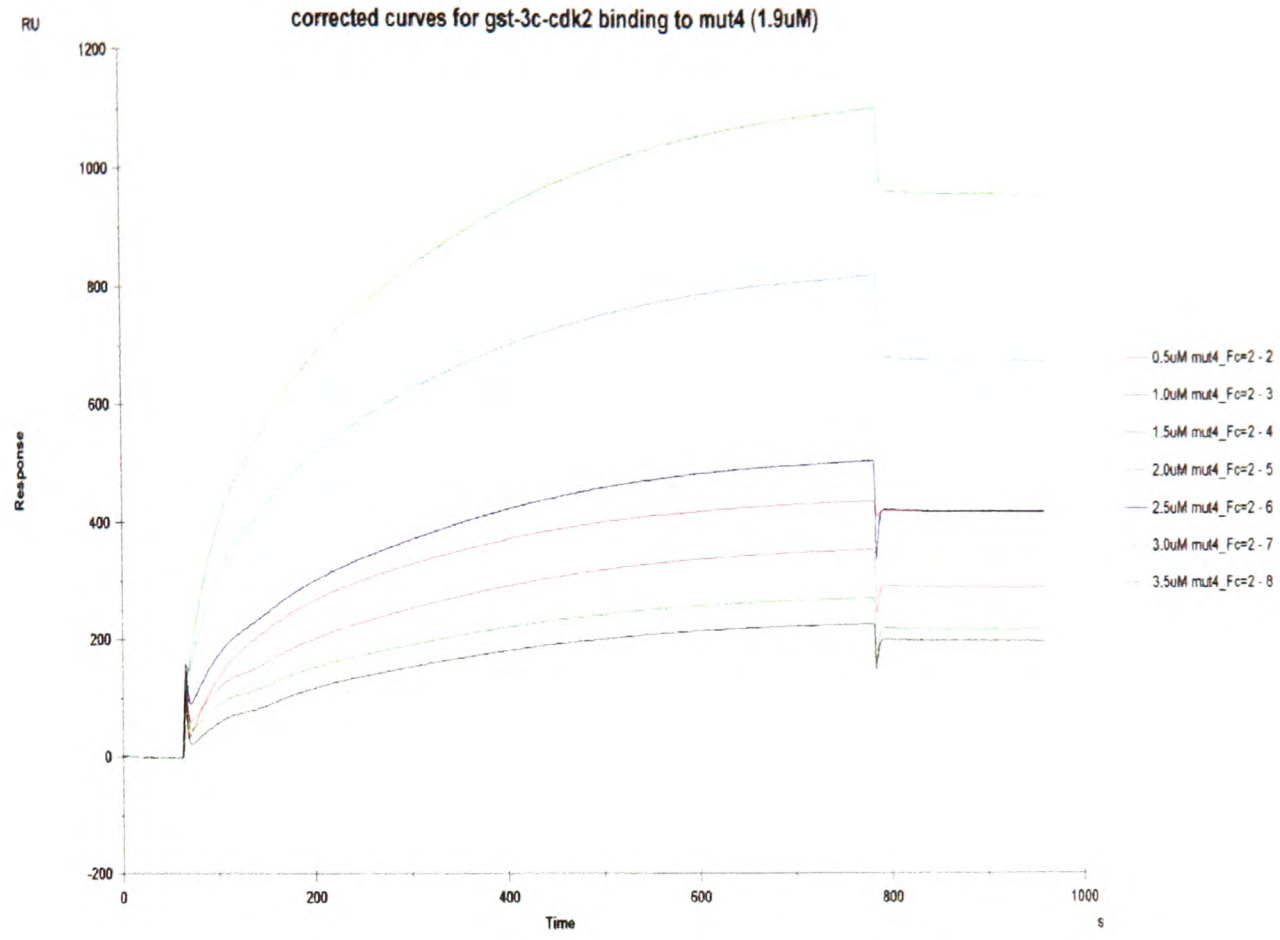
plots were non-linear thus, precluding the deduction of the K_D *value from these plots. For the Scatchard plot,*

$$R = R_{max} - K_D \frac{R}{L} \qquad (1)$$

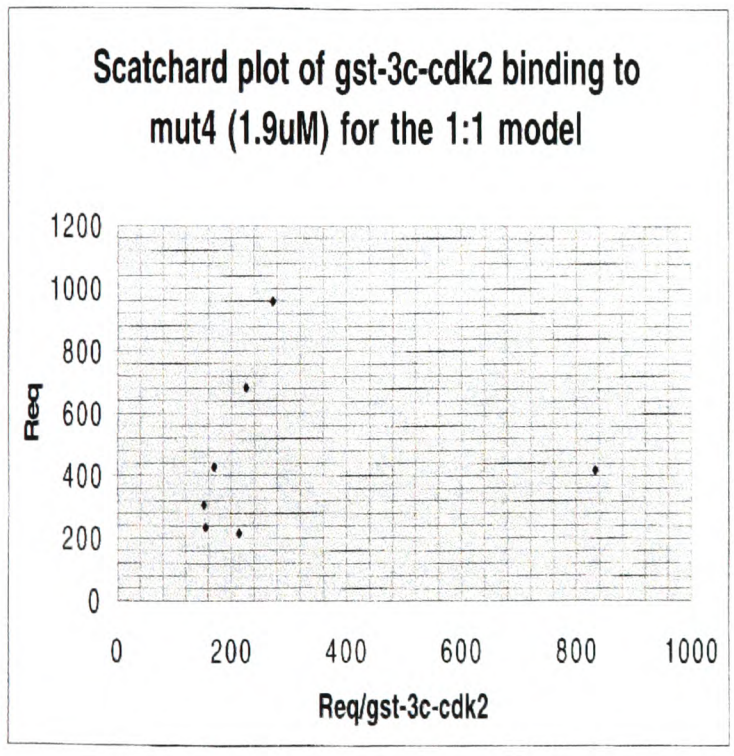
Where R = response units of the binding interaction, R_{max} *= the response unit at equilibrium and L = ligand concentration. A plot of R versus R/L gives slope =* $-K_D$.

A similar experiment was performed on Ile94Ala:Leu96Ala double mutant and the results are shown in the next figure (Fig. 4.18).

(a)



(b)



(c)

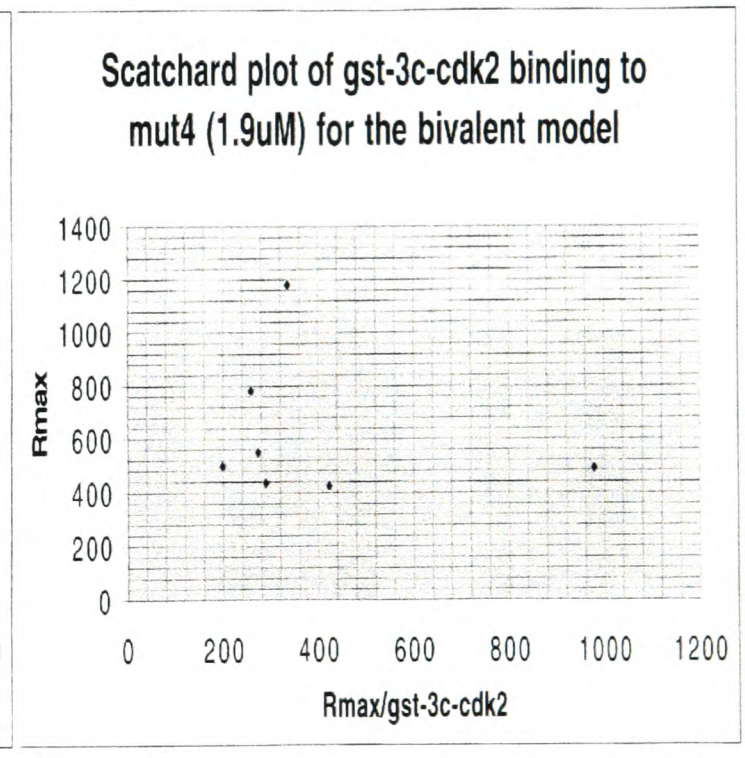


Figure 4.18 Binding of GST-3C-CDK2 to immobilised Ile94Ala:Leu96Ala double mutant

Although there was a binding interaction at all the concentrations tested, a comparative analysis of the RUs showed that in general, the double mutant resulted in a decreased RU

for the same analyte concentration. The Scatchard plots (b) and (c) showed that neither a 1:1 nor a 2:1 model was appropriate for the binding interactions observed.

The experiments conducted at higher concentrations of immobilised ligand i.e. 3.8μM (see fig. 4.19) and 7.6μM showed the same trend as that described in figure 4.18 except for the expected increases in the responses units due to the increases in concentration of proteins being used.

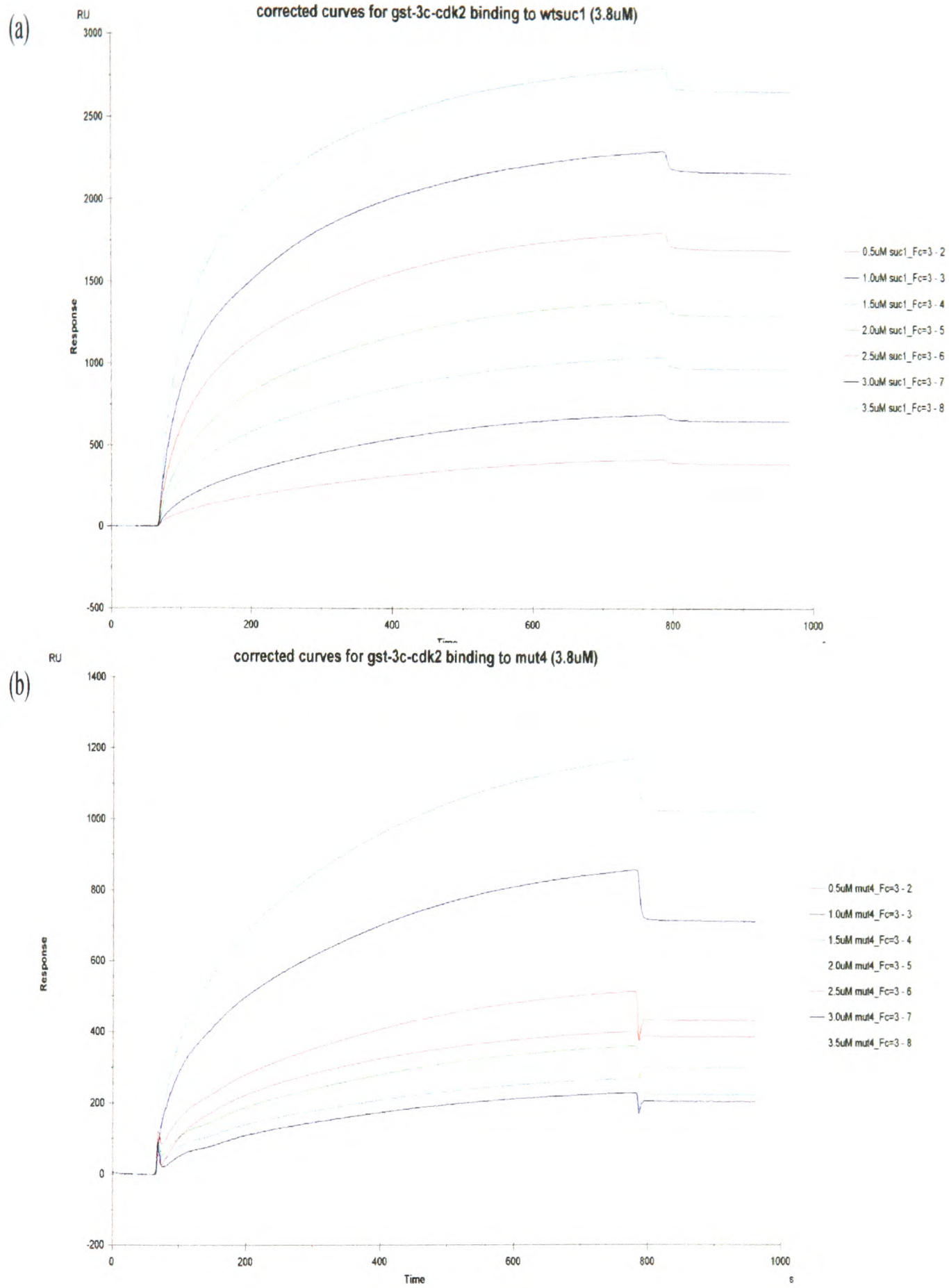


Figure 4.19 Comparative sensograms depicting the binding of GST-3C-CDK2 to (a) wild-type SpSuc1 and (b) Ile94Ala:Leu96Ala double mutant

The response unit for each concentration of analyte were as follows: for 0.5μM GST-3C-CDK2, SpSuc1 = 300RU and Ile94Ala:Leu96Ala = 220RU, for 1.0μM GST-3C-CDK2,

SpSuc1 = 700RU and Ile94Ala:Leu96Ala = 250RU, for 1.5μM GST-3C-CDK2, SpSuc1 = 950RU and Ile94Ala:Leu96Ala = 340RU, for 2.0μM GST-3C-CDK2, SpSuc1 = 1300RU and Ile94Ala:Leu96Ala = 390RU, for 2.5μM GST-3C-CDK2, SpSuc1 = 1750RU and Ile94Ala:Leu96Ala = 500RU, for 3.0μM GST-3C-CDK2, SpSuc1 = 2250RU and Ile94Ala:Leu96Ala = 840RU and for 3.5μM GST-3C-CDK2, SpSuc1 = 2750RU and Ile94Ala:Leu96Ala = 1800RU. This showed that ability of the Ile94Ala:Leu96Ala double mutant to bind to GST-3C-CDK2 had been attenuated not abolished as was suggested by the preliminary results (see fig. 4.10).

Attempts to obtain quantitative results using the BIAevaluation software (v3.0) showed that the 1:1 binding model gave very large chi^2 values (much greater than anticipated) and impossibly low k_d values. For the 2:1 model, the chi^2 values were relatively lower and there was some consistency between k_{a1}, k_{d1}, k_{a2} and k_{d2} for the different concentrations of analyte. A Scatchard plot gave anomalous behaviour in all the plots. The K_D for the linear part of the Scatchard (when the concentration of GST-3C-CDK2 was low) was 1.1μM. The one consistent but odd feature of the sensograms was that (fig. 4.19) as the concentration of GST-3C-CDK2 increased, so did the RU_{max} and there was no evidence of the saturation that one would expect with either a 1:1 or a 2:1 binding model. Normally, this anomalous behaviour could be attributed to the immobilisation of too much ligand on the chip but varying the amount immobilised made no difference in this particular case. Also, increasing the flow rate to offset possible re-association events in the resulting sensograms failed to rectify this anomaly. Also, an attempt to obtain thermodynamic data by the isothermal calorimetry (ITC) method (performed at University College London by Dr. J. Ladbury) showed that there was no measurable binding interaction between GST-3C-CDK2 and SpSuc1 in solution. Those latter attempts with the BIAcore to obtain more quantitative data in the binding of SpSuc1 and Ile94Ala:Leu96Ala mutant to GST-3C-CDK2 were disappointing. Nevertheless, both the early and later experiments showed that the binding of the mutant was severely attenuated but not completely abolished.

4.3.7 Histone kinase(HK) assays

As described in chapter 1, SpSuc1 was discovered via its ability to rescue a strain of *S. pombe* carrying a temperature-sensitive (ts) mutant copy of SpCDK1 (Hayles *et al.*, 1986b). It has been assumed that this effect is the result of SpSuc1 binding to SpCDK1 and providing additional stability. Therefore, if the CDK1 binding site on SpSuc1 has been disrupted, it might be anticipated that there would be diminished rescue of the ts-mutant at non-permissive temperatures. Histone Kinase assays were performed with Asp54Asn and Ile94Ala:Leu96Ala mutants. The autoradiograph of the assay involving the Ile94Ala:Leu96Ala mutant is shown in the figure below.

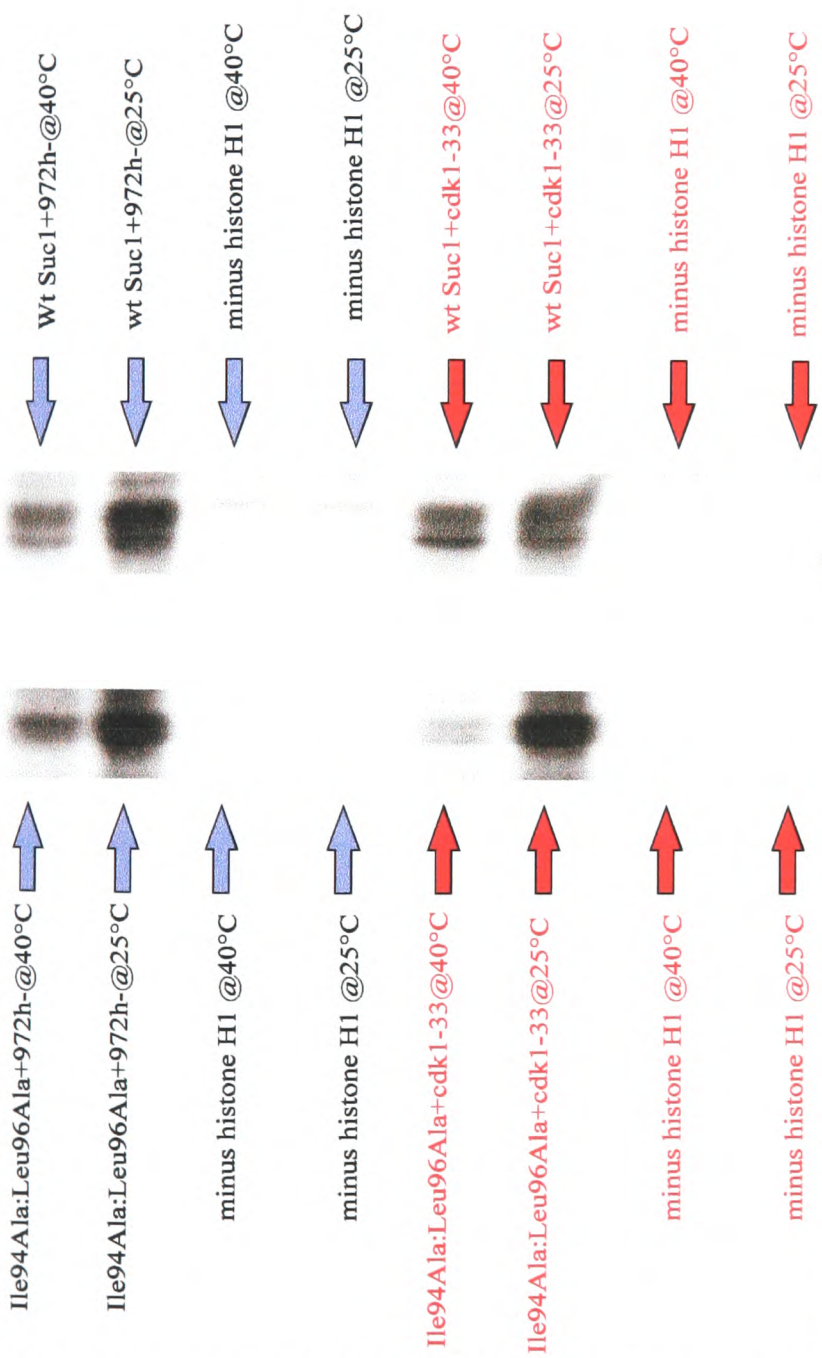


Figure 4.20 Histone H1 kinase activity of wild-type SpSuc1 and Ile94Ala:Leu96Ala mutant

Each assay was performed with 5μM of recombinant SpSuc1 protein on crude yeast extracts with relatively impure histone H1 from Sigma. The autoradiograph was scanned and the background exposure was subtracted from the entire film till the contrast became clearer. The arrows point to the phosphorylated histone H1. Blue and red arrows point to reaction lanes of the wild type yeast strain (972h-) and CDK1-33^{ts} mutant, respectively. The mutation of the side chains of residues of Ile94 and Leu96 led to a reduction in the ability of the mutant SpSuc1 to rescue CDK1-33^{ts}. In contrast, the Asp54Asn mutant was able to stabilise CDK1-33^{ts} + cyclin B complex at the elevated temperature (data not shown).

4.3.8 Yeast in vivo rescue assay

I am grateful to Dr. J. Hayles of the ICRF for the supply of materials (vectors, yeast media and reagents) and also, lab space and expert advice in the course of this part of the work. Mr. S. Montgomery in the Kearsey group (Zoology Department, University of Oxford) generously provided similar assistance (i.e. the provision of lab space, yeast media, reagents and expertise).

4.3.8(a) Subcloning SpSuc1 gene fragment into yeast expression vectors

S. pombe cells lacking SpSuc1 are not viable (Hayles *et al.*, 1986a). In these experiments, the ability of wild type SpSuc1 and Ile94Ala:Leu96Ala mutant to rescue ΔSpSuc1 haploid yeast cells were investigated. The initial step involved the subcloning of the gene fragments of SpSuc1 (fig. 4.21) and the mutant into the yeast expression vectors.

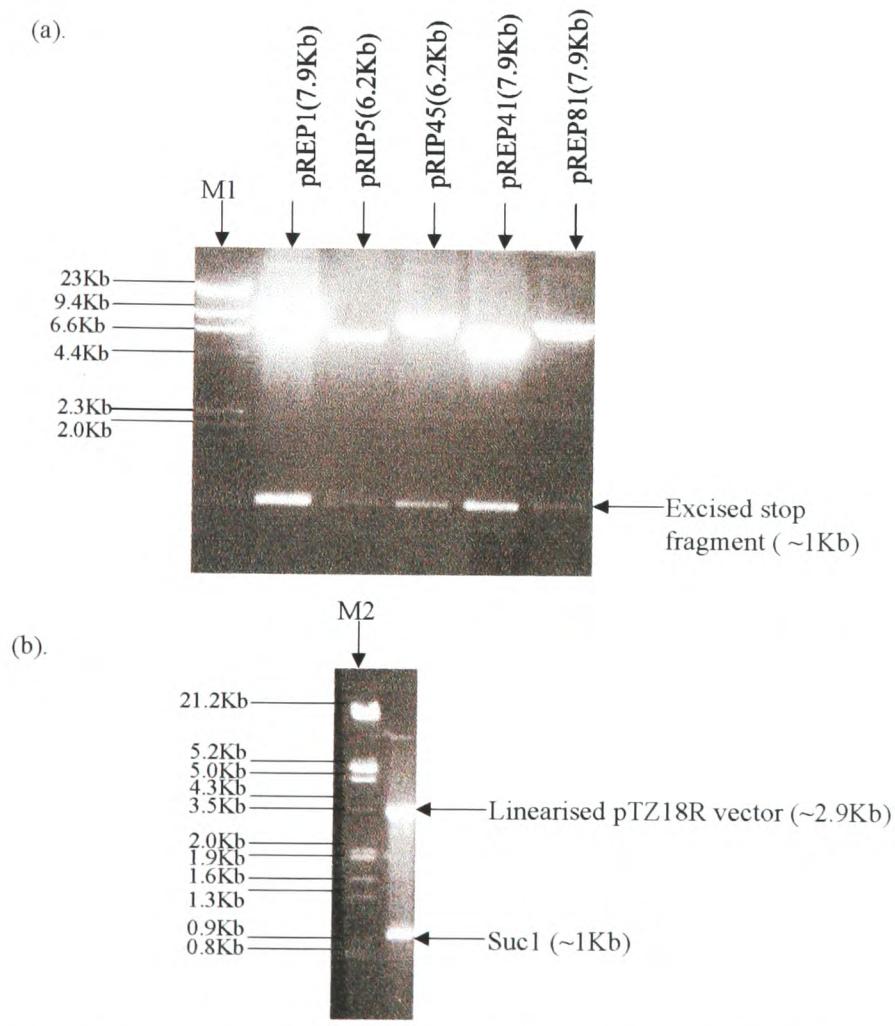


Figure 4.21 Restriction digest of pREP and pTZ18R/SpSuc1 plasmid vectors

(a) Each sample was double-digested with Nde I and Sal I restriction endonucleases and ran on a 1% agarose gel in TAE buffer. The reaction linearised each vector and released the stop fragment of about 1Kb (see fig. 4.5). M1 = lambda-hind III digested DNA size-markers. (b) The SpSuc1 gene fragment was also excised from the pTZ18R cloning vector using the restriction enzymes. M2 = lambda-hind III and Eco RI digested DNA size-markers

The results showed that the restriction enzymes successfully cut the DNA at the appropriate sites to release DNA fragments of the expected sizes. Repeated attempts at transforming the ligation mixture into *E. coli* cells made competent by the calcium chloride method failed so a different method of making competent cells was used and the competency of the cells were tested. The results are presented in the table below.

Amount of DNA (pUC18) transformed	Number of colonies on each plate	Number of colonies per microgram (µg) of DNA
10pg	136	1.4×10^7
50pg	671	1.3×10^7
100pg	A lawn of colonies (>1000)	
10000pg	A lawn of colonies (>1000)	

Table 4.4 Competency test on E. coli SURE cells

The results showed that the competency of the SURE cells was comparable to that of commercially available super or "ultra" competent (1 x 10^8/µg of DNA) cells.

Recombinant plasmid	Number of colonies per plate
pREP1/SpSuc1	192
pRIP5/SpSuc1	352
pREP41/SpSuc1	179
pRIP45/SpSuc1	306
pREP81/SpSuc1	131
pREP1/Ile94Ala:Leu96Ala	353
pRIP5/Ile94Ala:Leu96Ala	353
pREP41/Ile94Ala:Leu96Ala	74
pRIP45/Ile94Ala:Leu96Ala	233
pREP81/Ile94Ala:Leu96Ala	115

Table 4.5 Transformation of recombinant pREP plasmid DNA into competent E. coli SURE cells

Representative colonies were picked from each plate and minipreped to isolate the recombinant plasmid for restriction analysis. After confirmation that the insert had been incorporated into the yeast vectors, stocks of each recombinant plasmid were made and stored in Tris-EDTA (TE) buffer pH 7.4.

4.3.8(b) Transformation of recombinant plasmid DNA into S. pombeΔSuc1

The chemical transformation method (Lithium acetate) was more efficient ($\approx 10^5$ colonies/µg of DNA than the electroporation method ($\approx 10^3$ colonies per µg of DNA). For the chemical method, the colonies appeared on the fourth day of incubation at 32°C whereas colonies appeared on the fifth day for the electroporated cells suggesting that the relatively harsher treatment of mild electrocution may have had an adverse effect on the growth rate of the transformed yeast cells. The colonies from the Lithium acetate method were used for subsequent analysis. After sporulation, approximately equal numbers (≈60) of colonies grew on the MM+adenine plates in the presence or absence of thiamine for wild-type SpSuc1, Ile94Ala:Leu96Ala mutant and the vector only (i.e. pREP1, pREP41 and pREP81) plates. This suggested that the colonies were a mixture of haploids and diploids but since there was no significant numerical difference in the number of colonies on the reaction and the control plates, one could conclude that the Ile94Ala:leu96Ala mutant was able to rescue a yeast strain lacking a copy of the *SpSuc1*$^+$ gene.

To overcome the difficulty encountered in distinguishing haploid from diploid cells, the colonies (after sporulation) were maintained on minimal media with low adenine and this gave a colour contrast between the two types of spores (pink = haploid and white = diploid). In all the cases in which the promoter was repressed (presence of thiamine in the medium) pink patches appeared on the plate (data not shown) suggesting that either these colonies were false positives or that there was significant read through despite the presence of thiamine in the media. The pREP vectors have been shown to exhibit this laxity in promoter regulation (Forsburg, 1993). The presence of pink (i.e. haploid) colonies on the selection plate also suggested that the Ile94Ala:Leu96Ala double mutant had retained the ability to rescue *S. pombeΔSuc1*.

4.3.9 APC activation assay

As stated in the chapter 1, SpSuc1 appears to be an important mediator of the activation of the APC. In these experiments both the hydrophobic mutant (Ile94Ala:Leu96Ala) and the quadruple mutant (Arg30Ala:Arg39Ala: Ile94Ala:Leu96Ala) were examined for their ability to activate the APC. Production of the quadruple mutant was performed by PCR (see fig. 4.8). The PCR results are shown in the figure below.

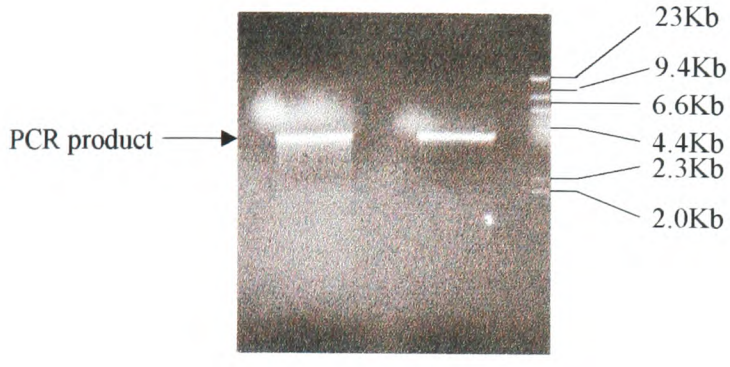


Figure 4.23 The product of the PCR-based mutagenesis reaction

Because the mutagenesis reaction was carried out on a pRK172-borne SpSuc1 gene fragment, the expected size of the product would correspond to a linearised pRK172/SpSuc1 vector which equals ≈3.4kb.

The results of the full-length DNA sequencing showed that only the intended mutation had occurred in the Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant but that for the Arg30Ala:Arg39Ala double mutant showed that a number of unintended mutations had occurred. Not surprisingly, in the subsequent experiment to check for protein expression, only the quadruple mutant produced an identifiable SpSuc1 protein product. Protein expression was scaled up and purified as before except that this time, a third column (S200 gel filtration column was introduced) prior to the setting up of crystallisation trials (Appendix I). This last step was introduced in order to isolate a homogenous species of purified protein for the growth of protein crystals. The chromatogram from this experiment showed an unexpected second peak (≈30Kda) in addition to the expected monomeric SpSuc1 peak (≈13Kda).

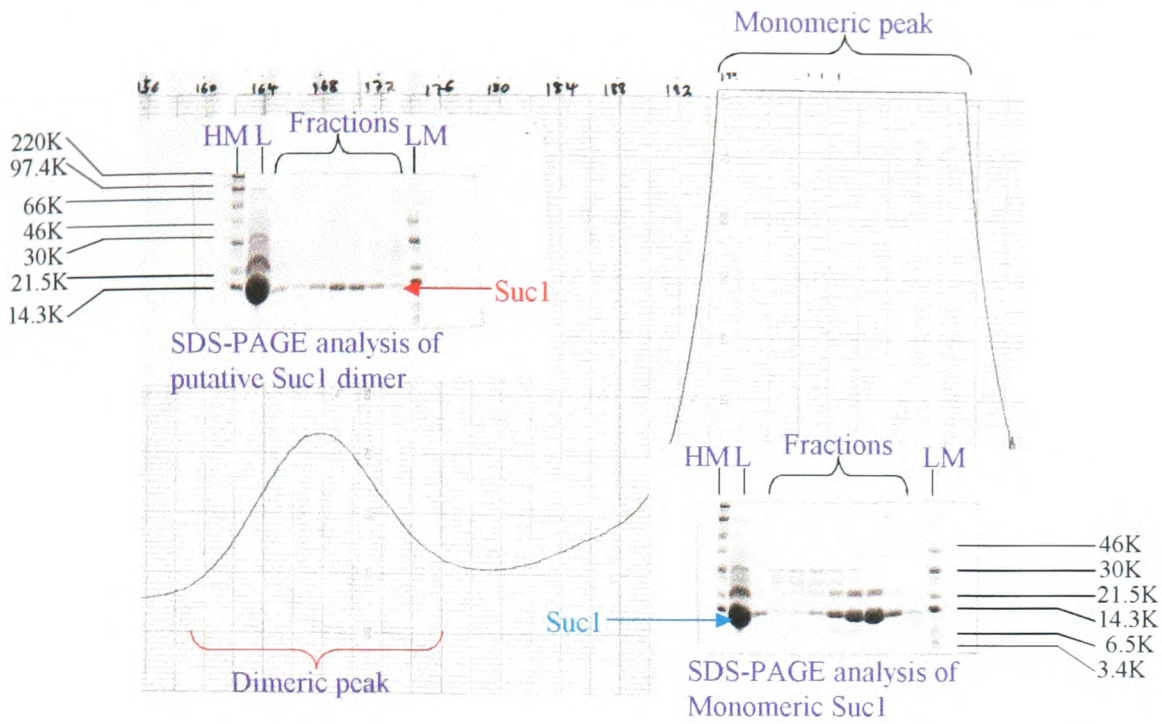


Figure 4.24 Elution profile of purified Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant on an S200 (Pharmacia) gel filtration column

A sample (5ml) of the concentrated eluate from the HIC column was applied to the column at a flow rate of 5ml/min in GF buffer and a full scale deflection (f.s.d.) of 1 absorbance unit at 280nm. A 100µl aliquot from each fraction (4ml) in each chromatographic peak was analysed on a 10% tris-tricine gel.

In order to confirm the identity of the putative dimeric species, the sample was run on an SDS-PAGE gel and probed with an anti-SpSuc1 antibody (fig. 4.25). The result suggested that the putative dimeric peak contained SpSuc1 protein.

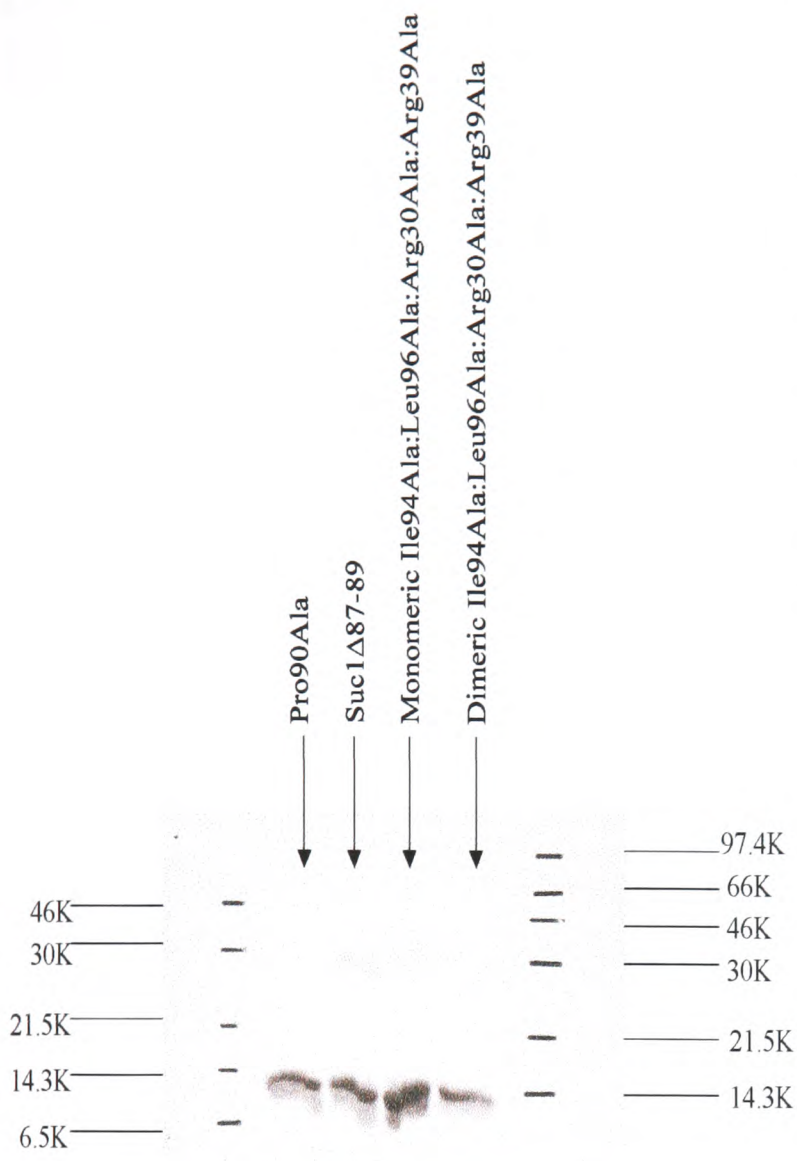


Figure 4.25 Western blot analysis of a dimeric SpSuc1 species

Pro90Ala and SpSuc1Δ87-89 proteins were also probed alongside the monomeric and putative dimericArg30Ala:Arg39Ala:Ile94Ala: Leu96Ala quadruple mutant. The result suggested that the putative dimeric peak consisted predominantly of SpSuc1.

The purified wild-type and mutant proteins were sent to Prof. Hershko. Dr. M. Shteinberg performed the APC activation assay as described in section 4.2.20(c). The results of the APC activation assay are presented in the table below:

Reaction mixture	**Unit of APC activity**
Blank control	0
1µl of APC	0.14
1µl of APC + 125 units SpCDK1-Cyclin B	0.92
1µl of APC + 125 units SpCDK1-Cyclin B + "Haifa" SpSuc1	6.5
1µl of APC + 125 units SpCDK1-Cyclin	6.8

B + "Oxford" SpSuc1	
1µl of APC + 125 units SpCDK1-Cyclin B + Ile94Ala:Leu96Ala double hydrophobic mutant	0.68
1µl of APC + 125 units SpCDK1-Cyclin B + Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple anionic and hydrophobic mutant	0.99

Table 4.6 APC activation assay with SpSuc1

The results show that both the "Oxford" and "Haifa" SpSuc1 were able to activate the APC resulting in a seven-fold increase in APC activity (i.e. iodination of the substrate). In contrast both the double and quadruple mutants gave levels corresponding to basal APC activity. Only the peak containing monomeric quadruple mutant was used in this assay.

4.4 DISCUSSION

The preliminary results of this study showed that monomeric SpSuc1 from *S. pombe* mediates protein-protein interactions with GST-CDK2 via certain residues in the hydrophobic region; the evidence pinpointed Ile94 and Leu96 as two residues that were important for the binding of SpSuc1 to GST-CDK2. The experimental design could not discriminate between Ile94 and Leu96 in terms of contribution to binding energy.

4.4.1 Retention of structural integrity

The DNA sequencing and ES MS results showed that the intended mutations had been effected and that no aberrant mutations had occurred in the *SpSuc1*$^{+}$ gene (Figs. 4.9 and 4.12) for the mutants that were successfully expressed. Since these results were obtained, the complete DNA fragment of Ile94Ala:Leu96Ala and Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala SpSuc1 mutants have been sequenced and they were shown to be free of spurious mutations. Circular dichroism spectra of SpSuc1 and the mutants (fig. 4.13) showed that there had been a retention of structural integrity among the mutants. Apart from Asp54, all the residues chosen for mutational analysis were surface residues; in the light of the structural analysis of point mutants described in chapter 3, it is plausible that the mutation of the isoleucine and leucine side-chains of residues 94 and 96 to alanine had minimal or no effect on the global architecture of the protein although local conformational changes cannot be ruled out. Since the global structure of the protein has not been perturbed by these point mutations (i.e. the mutants folded correctly), any biochemical or functional differences observed between wild-type SpSuc1 and the mutants can be attributed to the specific residues that were altered.

4.4.2 The Oligomerisation State of SpSuc1

Duality of conformational states (monomeric versus strand-exchanged dimer) among the Cks family of proteins has been a source of intrigue for workers in the field since it was observed (Schlunegger *et al.*, 1997). The predominant molecular species of SpSuc1 in solution are mono-dispersed monomeric proteins. Although small amounts of dimeric SpSuc1 have been detected by other workers (Birck *et al.*, 1995) they did not

find an equilibrium between the two forms. Also, the hexameric assemblies observed (Parge *et al.*, 1993) in the crystal structure of the human CksHs2 homologue has not been detected in solution by other workers (Birck *et al.*, 1995). The reported species so far, have been monomers and dimers in solution. This phenomenon has been observed in this study (fig. 4.24). The zinc-mediated dimer of SpSuc1 is a special case and it is not yet proven that *in vivo*, SpSuc1 function requires zinc ions. The presence of zinc ions in the crystal structure of SpSuc1 is probably due to the requirement of the crystallisation conditions. In this study, the application of a relatively pure protein sample to a gel filtration column (in a buffer lacking zinc ions) prior to the setting up of crystallisation trials revealed the presence of a putative dimeric species for the Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant. The ratio of this dimer (≈5%) to the monomer was consistent with previous observations (Rousseau *et al.*, 1998). Indeed, in some laboratories (Neyroz *et al.*, 1996), this is the established method for isolating strand-exchanged dimeric species of SpSuc1 for biophysical studies. The emerging picture is that the strand-exchanged SpSuc1 dimer is an *in vitro* artefact arising when monomeric SpSuc1 is maintained at a relatively high concentration (>20mg/ml). A putative functional role has been assigned to the strand-exchanged dimer as a negative regulator of CDK1 (Parge *et al.*, 1993) but in the light of the paucity of this dimeric species *in vivo*, and the absence of direct empirical evidence, it is the author's view that this putative functional role for the Cks strand-exchanged dimer is implausible.

4.4.3 Comparative binding studies

The use of GST fusion proteins in SPR studies has been criticised for overestimating the quantitative data obtained because of the divalency of GST and the subsequent avidity effect associated with this phenomenon (Ladbury *et al.*, 1995). Others (Panayotou *et al.*, 1993) have compared sensograms obtained from free and GST fusion SH2 domains and they concluded that there was no significant difference in the affinity values. The preliminary K_d values reported in this study (fig. 4.15) were reported only as guides to show how the mutants differed from the wild-type protein with respect to the binding to GST-CDK2. Subsequently, more detailed studies were carried out to check the appropriateness of the model chosen and also, to eliminate experimental variables such as

flowrate, concentration of analyte and immobilised ligand, and the elimination of re-association events. In the preliminary analysis, the binding interaction between wild-type SpSuc1 and GST-CDK2 was estimated as 67nM. The value for CksHs1 and CDK2 obtained by BIAcore analysis was 77nM (Bourne *et al.*, 1996). In the binding experiment, the interaction between SpSuc1 and human CDK2 (HsCDK2) was examined. The natural partner for the former is SpCDK1 and that for the latter is CksHs1. The sequence similarity between HsCDK2 and SpCDK1 is about 65% and SpSuc1 and CksHs1 share a 53% sequence identity. In this study, SpSuc1 did not bind to CDK2 with a high enough affinity to be detectable by the BIAcore method. In solution phase, SpSuc1 did not bind to CDK2 or GST-CDK2 but when SpSuc1 was immobilised, there was a binding interaction with GST-CDK2 but not with free CDK2. This could be attributed to the avidity effect conferred on GST-CDK2 by the inherent tendency for two GST molecules to associate to form a functional homodimer in solution and the lack of binding between immobilised SpSuc1 and free CDK2 could be attributed to species difference i.e. SpSuc1 being an *S. pombe* protein and CDK2 being a human protein. The evidence suggests that when SpSuc1 is immobilised, it adopts a conformation that is able to bind to GST-CDK2 but SpSuc1 in solution lacks this putative CDK2-competent conformation hence the absence of a binding interaction between these two molecules in solution. Dr. Nick Brown (University of Oxford, personal communication) noticed similar observations when he attempted to form an *in vitro* complex of HsCDK2-SpSuc1 protein complex for crystallisation studies by mixing purified samples of each individual protein before applying the mixture to a gel filtration column in an attempt to isolate a protein-protein complex. Despite extensive attempts using several different buffer conditions, the separation column always resolved the two components. In contrast, the binding of GST-CDK2 to SpSuc1-sepharose was with such a high affinity that the complex could not be eluted with non-denaturing buffer (Dr. Nick Brown, University of Oxford, personal communication). This is consistent with the observations made in this chapter and leads to the conclusion that the conformation of immobilised SpSuc1 may differ from free SpSuc1 in solution and that this conformational difference allowed immobilised SpSuc1 to bind to GST-CDK2 whilst free SpSuc1 in solution could not bind to GST-CDK2. Another plausible reason is that immobilisation of SpSuc1 gives rise to high local

concentrations that would artificially enhance the binding interaction of an otherwise weak binder (Egan and Solomon, 1998). For example, the concentration of SpSuc1/Cks protein coupled to sepharose beads has been estimated to be about 100μM. This underlies the effectiveness of using Suc1-sepharose beads to isolate various forms (monomeric, cyclin-bound, active and inactive) of CDK1 from eukaryotic cells although experiments conducted with (sub)physiological amounts of human Cks proteins showed that these proteins preferred to bind to phosphorylated, cyclin-bound CDK1 and CDK2 (Egan and Solomon, 1998).

4.4.4 The hydrophobic patch

The most interesting mutant in the preliminary study was the Ile94Ala:Leu96Ala double mutant. Both residues are completely conserved across species. In the X-ray crystal structure of SpSuc1, the two residues are located in the hydrophobic region near the β-hinge-forming-conserved HVPEPH sequence. To the left of this hydrophobic region is the aromatic cluster formed by Tyr31, Tyr38 and Tyr85. The evidence indicates that Ile94 and Leu96 of p13^{suc1} are necessary for binding and stabilisation. At first, the BIAcore studies suggested that this mutant could not bind to GST-CDK2 but further experiments showed that the binding interaction had been markedly reduced but not completely abolished. This latter assessment fitted with the results of the histone kinase assay which showed that although this mutant had retained the ability to stabilise a temperature-sensitive CDK1 mutant, the extent had been severely reduced by the mutation of the side chains of Ile94 and leu96. This ameliorated binding explains the ability of this mutant to rescue an *S. pombeΔSuc1* strain in the *in vivo* assay: the deficit accruing from the reduction in binding ability could be compensated for by the overproduction of the protein to several orders of magnitude above the required physiological levels.

The nature of the hydrophobic region in monomeric SpSuc1 and the strand-exchanged dimer are markedly different (Bourne *et al.*, 1995; Endicott *et al.*, 1995). In particular, Ile94 and leu96 cannot mediate protein-protein interactions in the manner described for β-strand-exchanged SpSuc1. The crystal structure of CksHs1-CDK2 (fig.

4.26) showed that the β-strand-exchanged dimer could not mediate interactions with CDK2 via the hydrophobic patch.



Figure 4.26 The X-ray structure of CksHs1-HsCDK2

CksHs1 binds to the helix 5 and loop14 (L14) which follows this helix. The anion-binding site is exposed and this has the potential to interact with phosphorylated substrates or subunits. The structure shows that 15 out of the 17 residues responsible for mediating protein-protein interactions are conserved among the Cks family. Reproduced with modification from (Morgan, 1996).



Figure 4.27 Amino acid sequence alignment of Cks proteins

The aligned sequences are the two human homologues (CksHs1 and CksHs2) and the two yeast homologues (S. pombe and S. cerevisiae Cks1). The X-ray structures of all four proteins have now been determined (Bourne et al., 1995; Khazanovich et al., 1996; Parge et al., 1993). Residues that have been shown by the CDK2-CksHs1 complex crystal structure to be involved in complex formation are in red and those that form the anion-binding site are in blue. The β-hinge region (red bar) is conserved in all four proteins. For CksHs1, this region consists of His60 to His65 and this forms a β-bend between β3 and β4 resulting in a monomeric unit. But in CksHs2, this region is conformationally different in that it forms an extended conformation promoting a β-strand exchange that interlocks two subunits into a dimer. This feature is also to be found in the yeast proteins where in SpSuc1, the β-bend region His85-His92 bends over to form a monomeric unit but in Cks1, it extends as in CksHs2 to form a strand-exchanged dimer. Using CksHs1 as an example, the position and identity of the four β-strands and helices are indicated beneath the sequence alignment. Reproduced with modification from (Bourne et al., 1996).

The alignment reveals two differences in the CDK-binding region of the Cks proteins: first, in CksHs1, residue 58 is a methionine but in both SpSuc1 and *S. cerevisiae*, it is a glutamic acid. Second, in CksHs1, residue 61 is a glutamic acid but the equivalent residue in SpSuc1 is a valine and that in *S. cerevisiae* Cks1 is an alanine. In *S. pombe* CDK1, there is a Glutamate at position 224 that may form a hydrogen bond with the glutamate in the Met58→Glu variation from CksHs1 to SpSuc1. The analogous situation in *S. cerevisiae* is a possible salt-bridge between Arg226 of *S. cerevisiae* CDK1 and the glutamate residue in Cks1. The change from Met58 to Glu in SpSuc1 may explain why SpSuc1 binds with lower affinity than human CksHs1 to human CDK2 (HsCDK2) since Met58 docks to hydrophobic patch. The Glu61→Val(SpSuc1) or Ala(Cks1) variation would lead to the abolition of a unique salt-bridge linkage between a conserved Lys227 in CDK2 and the glutamate conserved in human Cks proteins (Bourne *et al.*, 1996). The effect of this alteration on the overall binding interaction may not be dramatic but it could also contribute to the reduced affinity observed between human CDK2 and SpSuc1.

Apart from the case of the two hydrophobic residues discussed in this chapter, there are precedents (Lee *et al.*, 1995) for the important role a single amino acid may play in protein-protein interactions spanning an area of several angstroms and which at first glance may seem to involve several residue-to-residue contacts. This observation has led some workers (Cunningham and Wells, 1993) to classify residues involved in protein-protein interactions into two categories: contact and stabilisation residues. The former are residues that have been shown by atomic structures to make intersubunit contacts and the latter are those that apart from making intersubunit contacts, also make substantial contributions to binding and stabilisation energies between the two proteins. Thus, it is possible to delete the side chains of several contact residues without impeding complex formation but the deletion of one or two important binding residues may result in a dramatic reduction in the binding affinity between the two proteins in question. In the CDK2-CksHs1 complex, the contact residues from the X-ray structure are tabulated below:

HsCDK2 contact residue (from X-ray structure)	**CksHs1 contact residue (from X-ray structure)**		**SpCDK1 contact residue (from theoretical model)**	**SpSuc1 contact residue (from X-ray structure)**
Asp206	Met23		Leu181	His88
Ser207	Glu63		Ser214	Ile94 His40
Glu208	Glu63 His60		Glu215	Glu91 Pro90
Ile209	Glu63 Met58 His60		Ile216	Glu91 His88 Leu96
Asp210	His21		Phe220	Tyr31
Phe213	Tyr12 Leu68		Gln224	Ala32
Asp235	Pro62		Asp242	His88 Pro90
Lys237	Ile59 Glu61		Lys244	Glu86
Ser239	Arg70 Tyr57 Met58		Thr246	Tyr85
Phe240	Met58		Pro248	Tyr38 Asp33
Pro241	Tyr19		Arg249	Asp33 Asp34 Glu35 Tyr38
Lys242	Asp14		Trp250	Asp33 Asp34 Tyr31 Ala32
Trp243	Asp13 Asp14 Tyr12			

Table 4.7 CDK-Cks intersubunit contact residues

Predictably, this shows several residues in addition to Ile94 and leu96. But the relative importance of each contact residue with respect to the contribution to the binding and stabilisation energy can only be judged by a careful residue by residue point mutation study.

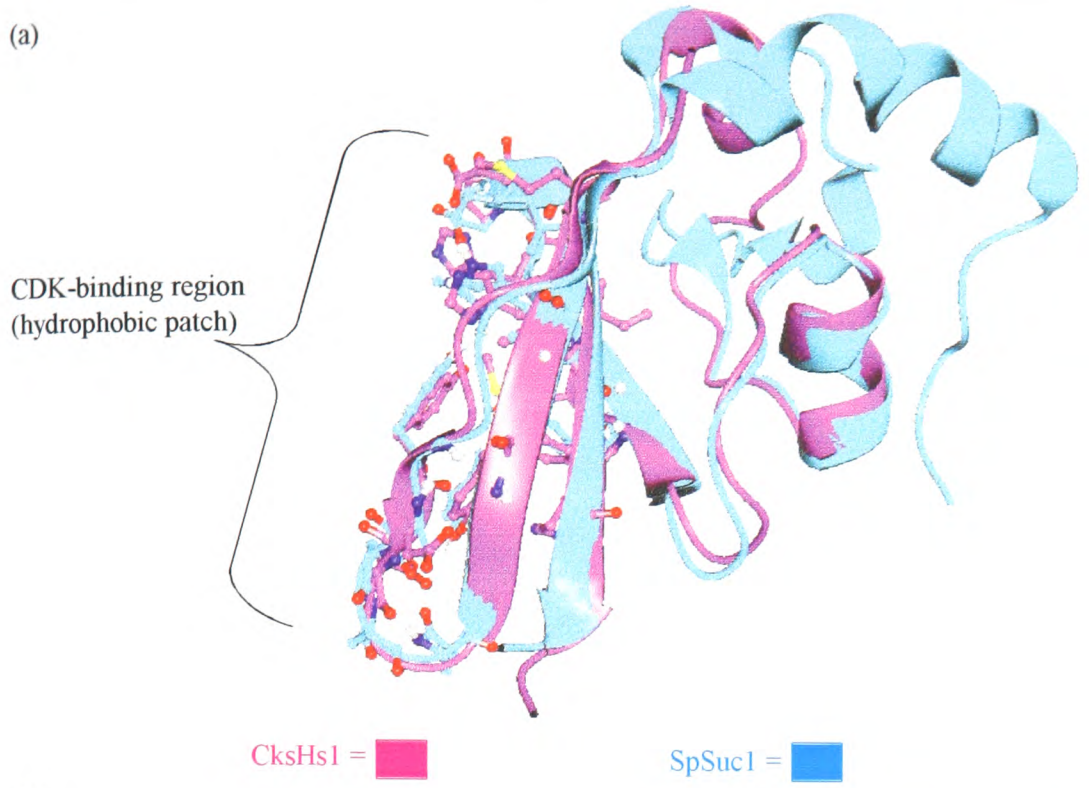


Figure 4.28 The conserved hydrophobic patch residues of SpSuc1 and CksHs1

(a) Superimposition of the two structures gave an r.m.s. deviation of 1.71Å for 12 C_α atom. (b) Stereo view of the hydrophobic patch in the two structures. The side chains of the residues listed in Table 4.7 are highlighted. This shows that the CDK interacting residues are highly conserved both in the primary amino acid sequence (fig. 4. 27) and in the tertiary structures.

```
HsCDK2  205 GDSEIDQLFRIFRTL  232 SMPDYKPSFPKW-AR
HsCDC2  206 GDSEIDQLFRIFRAL  233 SLQDYKNTFPKW-KP
Spcdc2  212 GDSEIDEIFKIFQVL  239 LLQDYKSTFPRW-KR
ScCdc28 214 GDSEIDQIFKIFRVL  241 YLPDFKPSFPQW-RR
HsCDK3  205 GDSEIDQLFRIFRML  232 QLPDYKGSFPKW-TR

HsCDK4  215 GNSEADQLGKIFDLI  243 SLP------------
HsCDK5  203 GNDVDDQLKRIFRLL  230 KLPDYKPTYPMYPAT
                <---α5--->          L14
```

Figure 4.29 CDK sequence alignment

The primary amino acid sequence for human CDK2 (HsCDK2), human CDK1 (denoted above as HsCDC2), S. pombe CDK1 (denoted above as Spcdc2), S. cerevisiae CDK1 (denoted above as ScCdc28), human CDK3 (HsCDK3), CDK4 (hsCDK4) and CDK5 (HsCDk5). The last two are not known to bind to bind to Cks proteins. The residues in the two regions of CDK2 (α5 and L14) that bind to CksHs1 are in green. What seems to distinguish a Cks binding from a non-Cks binding CDK is the conservation of the GDSEID and (S,T)FPXW motifs and the presence of an insert region between helices α5 and α6 of CDK2. Analysis of the binding interactions between SpSuc1, CDK1 and Crp from Dictyostelium (Sharma et al., 1999) has shown that the conservation of the GDSEID motif is the most important of the two motifs for SpSuc1 binding. Reproduced with modification from (Bourne et al., 1996).

4.4.5 The anion-binding site

In the CDK2-CksHs1 structure, it was observed that apart from the cluster of exposed and conserved hydrophobic residues involved in making contacts with residues in HsCDK2, there was on the other side of CksHs1, a cluster of conserved and positively charged pocket of residues (Lys11, Arg20, Tyr54 and Arg71) that formed a phosphate anion-binding site based upon structures of phosphate and vanadate bound to CksHs1 (Arvai *et al.*, 1995) at this region. The equivalent residues in SpSuc1 are Arg30, Arg39, Trp82 and Arg99.

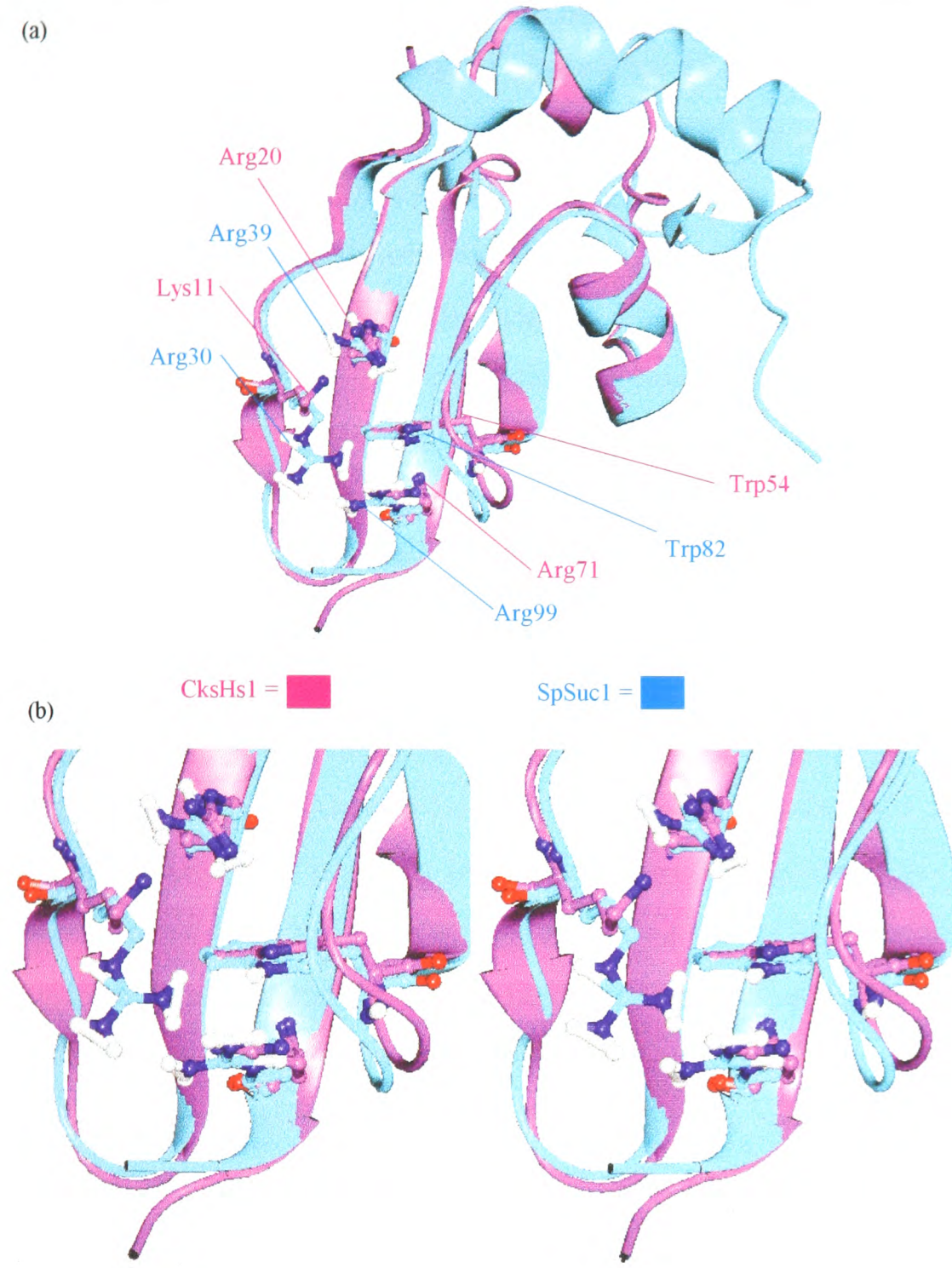


Figure 4.30 The conserved phosphate-binding site of SpSuc1 and CksHs1

The two structures have been superposed as in Figure 4.28. Although the global superposition is relatively poor, there is good superposition of the amino acids that form the conserved anion-binding site.

Buoyed by the demonstration that SpSuc1 bound to phosphorylated APC (Sudakin *et al.*, 1997) two new anion-binding mutants (Arg30Ala:Arg39Ala and Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala) were designed in an attempt to obtain direct empirical evidence that SpSuc1 binds both SpCDK1 and the APC at the same time using two independent (hydrophobic and positively charged cluster site) binding surfaces.

Because the Arg30Ala:Arg39Ala and Arg30Ala (one of the preliminary mutants) did not express and also the inability of the author to go back and prepare these particular mutants due to the pressure of time constraints, the results obtained from this line of enquiry using only the hydrophobic and quadruple mutants are preliminary. Both the double hydrophobic and quadruple mutants failed to activate the APC beyond the basal level (Table 4.6). The ambiguity stemmed from the fact that the failure of both mutants to activate the APC could have the following interpretation: the activation of the APC is a phosphorylation event carried out by the CDK1-Cyclin B-SpSuc1 ternary complex. This targeting model requires that SpSuc1 bring the ternary complex in close proximity to the partially phosphorylated APC so that the kinase (CDK1-Cyclin B) can activate the APC by fully phosphorylating the APC. The fully phosphorylated APC can then bring mitosis to an end by degrading cyclin B. But the ability of both Ile94Ala:Leu96Ala and Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala mutants to bind to CDK1 had been markedly reduced by the Ile94Ala and Leu96Ala mutations. So, the lack of activation of the APC could be due to the inability of both mutants to interact with CDK1-Cyclin B. In the absence of an Arg30Ala:Arg39Ala double mutant, the lack of activation observed for the quadruple mutant cannot be attributed to the deletion of the arginine (Arg30 and Arg39) side chains. That is, the question "was the lack of activation of the APC observed for the Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala quadruple mutant due to its inability to bind to the putative phosphorylated unit of the APC?" could not be answered. Further studies are currently underway (Simon Holton, University of Oxford) to resolve this ambiguity.

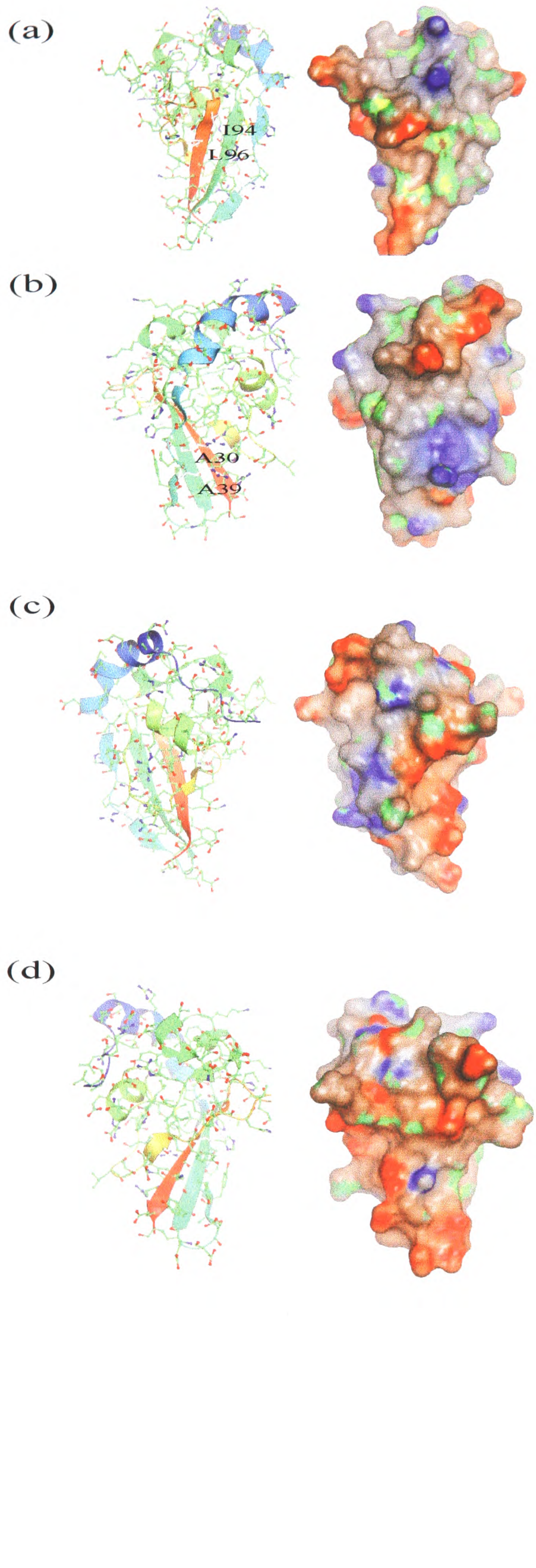


Figure 4.31 Surface representation of Arg30Ala:Arg39Ala:Ile94Ala :Leu96Ala quadruple mutant

The left panel of each diagram shows a panoramic view of the secondary structure elements and side-chains of the molecule in the same orientation as that in the right panel which is an AESOP representation of the surface properties of the molecule in the left panel. Hydrophobic patches are green, positively charged patches are blue and negatively charged patches are red. All other surface features are coloured grey. In (a) to (d), the molecule is progressively rotated by 90° about a vertical axis. (a) monomeric SpSuc1 molecule with the viewer looking directly (perpendicular to the page) at the hydrophobic patch. This shows that the two mutated residues are located in the central part of the green (hydrophobic) patch. (b) This view brings the anion-binding

site (blue) into view and the two arginine residues are located. (c) and (d) show that the two notable surface areas are unique to only the hydrophobic and anion-binding sites. This figure was created with the assistance of Dr. Noble (University of Oxford) using AESOP (M.E.M. Noble, unpublished work).

4.4.6 Summary

At the start of the experiments described in this chapter, the structure of SpSuc1 had been determined but its mode of interaction with CDK1 was unknown. The preliminary results presented in this chapter suggested that Ile94 and Leu96 (located at the hydrophobia patch) of SpSuc1 were involved in mediating CDK-SpSuc1 protein-protein interactions. The publication of the HsCDK2-CksHs1 complex structure (Bourne *et al.*, 1996) confirmed these preliminary conclusions. However, the function of SpSuc1 remained an elusive quest so further functional studies were carried out with the most interesting (from a functional point of view) mutant from the preliminary studies. These further studies suggested that the mutation of Ile94 and Leu96 to alanine led to a reduction but not abolition of binding interaction with CDK1. This attenuation in binding was probably offset by the *in vivo* overexpression of Ile94Ala:Leu96Ala double mutant in the *S. pombeΔSuc1* yeast strain. The timely publication (Sudakin *et al.*, 1997) demonstrating for the first time that SpSuc1 may bind to a phosphorylated APC subunit via the anion-binding site moved the investigations described here to a different level in which the goal was to use mutational analysis to identify these residues at the anion-binding site of SpSuc1. Although the results were not definitive, it was encouraging to note that wild-type SpSuc1 (prepared by two slightly different methods) activated the APC to identical levels but the SpSuc1 mutants whose binding abilities at both the hydrophobic and anionic-binding sites had been sabotaged was unable to elicit an APC activatory response in the assay. The body of evidence presented here in conjunction with recent papers from the literature (Bourne *et al.*, 1996; Egan and Solomon, 1998; Patra and Dunphy, 1998; Sudakin *et al.*, 1997) support "The Targeting Model" of Cks function (chapter 6).

4.5 LITERATURE CITED

Arvai, A.S., Bourne, Y., Hickey, M.J. and Tainer, J.A. (1995) Crystal structure of the human cell cycle protein CksHs1: single domain fold with similarity to kinase N-lobe domain. *J Mol Biol*, **249**, 835-842.

Basi, G. and Draetta, G. (1995) p13suc1 of Schizosaccharomyces pombe regulates two distinct forms of the mitotic cdc2 kinase. *Mol Cell Biol*, **15**, 2028-2036.

Basi, G., Schmid, E. and Maundrell, K. (1993) TATA box mutations in the Schizosaccharomyces pombe nmt1 promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. *Gene*, **123**, 131-136.

Birck, C., Raynaud-Messina, B. and Samama, J.P. (1995) Oligomerization state in solution of the cell cycle regulators p13suc1 from the fission yeast and p9cksphy from the myxomycete Physarum, two members of the cks family. *FEBS Lett*, **363**, 145-150.

Bourne, Y., Arvai, A.S., Bernstein, S.L., Watson, M.H., Reed, S.I., Endicott, J.E., Noble, M.E., Johnson, L.N. and Tainer, J.A. (1995) Crystal structure of the cell cycle-regulatory protein suc1 reveals a beta-hinge conformational switch. *Proc Natl Acad Sci U S A*, **92**, 10232-10236.

Bourne, Y., Watson, M.H., Hickey, M.J., Holmes, W., Rocque, W., Reed, S.I. and Tainer, J.A. (1996) Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle regulatory protein CksHs1. *Cell*, **84**, 863-874.

Brizuela, L., Draetta, G. and Beach, D. (1987) p13suc1 acts in the fission yeast cell cycle as a component of the p34cdc2 protein kinase. *EMBO J.*, **6**, 3507-3514.

Campbell, I.D. and Dwek, R.A. (1984) *Biological Spectroscopy*, California.
Cunningham, B.C. and Wells, J.A. (1993) Comparison of a structural and a functional epitope. *J. Mol. Biol.*, **234**, 554-563.

Egan, E.A. and Solomon, M.J. (1998) Cyclin-stimulated binding of Cks proteins to cyclin-dependent kinases. *Mol Cell Biol*, **18**, 3659-3667.

Endicott, J.A., Noble, M.E., Garman, E.F., Brown, N., Rasmussen, B., Nurse, P. and Johnson, L.N. (1995) The crystal structure of p13^{suc1}, a p34cdc2-interacting cell cycle control protein. *EMBO J.*, **14**, 1004-1014.

Endicott, J.A. and Nurse, P. (1995) The cell cycle and suc1: from structure to function? *Structure*, **3**, 321-325.

Fisher, C.L. and Pei, G.K. (1997) Modification of a PCR-based site-directed mutagenesis method. *Biotechniques*, **23**, 570-571, 574.

Forsburg, S.L. (1993) Comparison of Schizosaccharomyces pombe expression systems. *Nucleic Acids Res*, **21**, 2955-2956.

Hayles, J., Aves, S. and Nurse, P. (1986a) suc1 is an essential gene involved in both the cell cycle and growth in fission yeast. *EMBO J.*, **5**, 3373-3379.

Hayles, J., Beach, D., Durkaz, B. and Nurse, P. (1986b) The fission yeast cell cycle control gene cdc2: isolation of a sequence suc1 that suppresses cdc2 mutant function. *Mol. Gen. Genet.*, **202**, 291-293.

Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*, **96**, 23-28.

Johnson, W.C., Jr. (1990) Protein secondary structure and circular dichroism: a practical guide. *Proteins*, **7**, 205-214.

Kalinin, N.L., Ward, L.D. and Winzor, D.J. (1995) Effects of solute multivalence on the evaluation of binding constants by biosensor technology: studies with concanavalin A and interleukin-6 as partitioning proteins. *Anal Biochem*, **228**, 238-244.

Khazanovich, N., Bateman, K., Chernaia, M., Michalak, M. and James, M. (1996) Crystal structure of the yeast cell-cycle control protein, p13suc1, in a strand-exchanged dimer. *Structure*, **4**, 299-309.

Kunkel, T.A., Roberts, J.D. and Zakour, R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol*, **154**, 367-382.

Ladbury, J.E., Lemmon, M.A., Zhou, M., Green, J., Botfield, M.C. and Schlessinger, J. (1995) Measurement of the binding of tyrosyl phophopeptides to SH2 domains: a reappraisal. *Proc. Natl. Acad. Sci. (U.S.A.)*, **92**, 3199-3203.

Lee, C.H., Leung, B., Lemmon, M.A., Zheng, J., Cowburn, D., Kuriyan, J. and Saksela, K. (1995) A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *Embo J*, **14**, 5006-5015.

Mann, M. and Wilm, M. (1995) Electrospray mass spectrometry for protein characterisation. *TIBS*, **20**, 219-224.

Maundrell, K. (1993) Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene*, **123**, 127-130.

Morgan, D. (1996) The dynamics of cyclin-dependant kinase structure. *Curr. Opinion in Cell Biology*, **8**, 767-772.

Neyroz, P., Menna, C., Polverini, E. and Masotti, L. (1996) Intrinsic fluorescence properties and structural analysis of p13(suc1) from Schizosaccharomyces pombe. *J Biol Chem*, **271**, 27249-27258.

Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M.J., Hiles, I., Pawson, T. and Waterfield, M.D. (1993) Interactions between SH2 domains and tyrosine-phosphorylated platelet- derived growth factor beta-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. *Mol Cell Biol*, **13**, 3567-3576.

Parge, H.E., Arvai, A.S., Murtari, D.J., Reed, S.I. and Tainer, J.A. (1993) Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science*, **262**, 387-394.

Patra, D. and Dunphy, W.G. (1998) Xe-p9, a Xenopus Suc1/Cks protein, is essential for the Cdc2-dependent phosphorylation of the anaphase- promoting complex at mitosis. *Genes Dev*, **12**, 2549-2559.

Pharmacia. (1994) *BIAtechnology handbook*. Merk AB, Uppsala.

Raghavan, M. and Bjorkman, P.J. (1995) BIAcore: a microchip-based system for analyzing the formation of macromolecular complexes. *Structure*, **3**, 331-333.

Robinson, C.V. and Radford, S.E. (1995) Weighing the evidence for structure: electrospray ionization mass spectrometry of proteins. *Structure*, **3**, 861-865.

Rousseau, F., Schymkowitz, J.W., Sanchez del Pino, M. and Itzhaki, L.S. (1998) Stability and folding of the cell cycle regulatory protein, p13(suc1). *J Mol Biol*, **284**, 503-519.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a loboratory manual*. Cold spring harbour laboratory press, New York.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-5467.

Schlunegger, M.P., Bennett, M.J. and Eisenberg, D. (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem*, **50**, 61-122.

Sharma, S.K., Michaelis, C., Lee, K.Y., Wang, J.H. and Weeks, G. (1999) Binding and catalytic properties of the Cdc2 and Crp proteins of Dictyostelium. *Eur J Biochem*, **260**, 603-608.

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol*, **219**, 37-44.

Sudakin, V., Shteinberg, M., Ganoth, D., Hershko, J. and Hershko, A. (1997) Binding of activated cyclosome to p13suc1: Use for affinity purification. *J Biol Chem*, **272**, 18051-18059.

Szabo, A., Stolz, L. and Granzow, R. (1995) Surface plasmon resonance and its use in biomolecular interaction analysis (BIA). *Curr Opin Struct Biol*, **5**, 699-705.

van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R. and Groen, F.C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem*, **191**, 110-118.

Chapter 5: Co-expression of heterodimeric cell cycle regulatory proteins in *E. coli*

CONTENTS

LIST OF FIGURES **189**

LIST OF TABLES **189**

5.1 INTRODUCTION **189**

5.2 MATERIALS & METHODS **194**

5.2.1 Escherichia coli host strains and plasmid vectors *194*
5.2.2 Production of cyclinBΔ195 DNA fragment *194*
5.2.3 Subcloning of cyclinBΔ195 DNA fragment *195*
5.2.4 Production of recombinant protein *197*
5.2.5 In vivo strategies for producing soluble SpCyclin B3 ..... *198*
5.2.5(a) Effect of growth media *198*
5.2.5(b) Effect of temperature *198*
5.2.5(c) Effect of IPTG titration *199*
5.2.5(d) Co-expression of SpCyclin B3 with E. coli chaperonins *199*

5.2.5(e) Addition of 3% ethanol or 10% glycerol to the cell culture *200*
5.2.5(f) Expression of SpCyclin B3 as a thioredoxin-fusion protein *201*
5.2.6 In vitro strategies for producing soluble SpCyclin B3 ..... *201*
5.2.7 Preparation of inclusion bodies (IBs) *202*
5.2.8 Refolding recombinant SpCyclin B3 with 8M Urea buffer *202*
5.2.9 Refolding of SpCyclin B3 by immobilised metal affinity chromatography (IMAC) *203*
5.2.10 Iterative refolding of SpCyclin B3 ..... *204*
5.2.11 Homology modelling of SpCyclin B3 ..... *205*
5.2.12 Homology modelling of proteins by email *205*
5.2.13 Construction of pLNJ1 expression vector *207*
5.2.14 Restriction digests, ligation reactions and the harvesting of pLNJ1 expression vector *208*
5.2.15 Construction of pLNJ/SpCDK1 and pLNJ1/SpCyclin B plasmids *211*
5.2.16 Co-expression strategies *213*
5.2.17 Non-denaturing affinity purification of his-tagged proteins *214*
5.2.18 Simultaneous refolding of SpCDK1-SpCyclin B3 complex *214*

5.3 RESULTS **215**

5.3.1 Amplification and cloning of SpCyclin B3 ..... *215*
5.3.2 In vivo attempts at producing soluble SpCyclin B3 ..... *216*
5.3.3 In vitro attempts at producing soluble SpCyclin B3 ..... *219*
5.3.4 A homology model of SpCyclin B3 ..... *222*
5.3.5 DNA manipulations and construction of pLNJ1 vectors *224*
5.3.6 Recombinant protein expression and SDS-PAGE analyses *226*
5.3.7 Expression of pLNJ1/Cyclin B plus pRK172/SpCDK1 ..... *227*
5.3.8 Expression of pLNJ1/SpCDK1 plus pSJW1/SpSuc1 ..... *229*
5.3.9 Expression of pLNJ1/SpCDK1 plus pET21d/Cyclin A3. ..... *229*
5.3.10 Expression of pLNJ1/SpCDK1 plus pET21d/Cyclin B3. ..... *231*
5.3.11 Affinity Purification of soluble SpCDK1-Cyclin complexes *232*
5.3.12 Co-denaturation and refolding of SpCDK1 and Cyclin B3 ..... *234*
5.3.13(a) Theoretical models of SpCDK1 and SpCyclin B *235*
5.3.13(b) Interactions at the CDK-cyclin dimer interface *236*
5.3.14(a) The PSTAIRE region *244*
5.3.14(b) The activation segment *245*
5.3.14(c) The N-terminal lobe of HsCDK2 ..... *245*
5.3.14(d) The C-terminal lobe of HsCDK2 ..... *246*

5.4 DISCUSSION **246**

5.4.1 Protein folding in vivo and in vitro *246*

5.5 LITERATURE CITED **251**

List of Figures

Figure 5.1 Schematic representation of the sub-cloning procedure ---- 196
Figure 5.2 Construction of pLNJ1 expression vector ---- 208
Figure 5.3 Schematic representation of pLNJ1 expression vector ---- 210
Figure 5.4 Overview of pLNJ1/CDK1 and pLNJ1/Cyclin B expression constructs ---- 212
Figure 5.6 Determination of optimal expression conditions for SpCyclin B3 protein ---- 216
Figure 5.7 Expression of SpCyclin B3 as a thioredoxin fusion protein ---- 218
Figure 5.8 Refolding of SpCyclin B3 protein using the iterative method ---- 221
Figure 5.9 Aligned amino acid sequence of BvCyclin A3 and SpCyclin B3 ---- 222
Figure 5.10 Surface properties of BvCyclin A3 X-ray structure and SpCyclin B3 model ---- 223
Figure 5.11 Restriction digest of pACYC177 and pET28a vectors ---- 224
Figure 5.12 Restriction analysis of pLNJ1/CDK1, pLNJ1/CYCLIN B and pLNJ1 vectors ---- 225
Figure 5.13 Expression of SpCDK1 ---- 226
Figure 5.14 Co-expression of pRK172/CDK1 & pLNJ1/CYCLIN B (1) and pSJW1/Suc1 & pLNJ1/CDK1 (2) in B834(DE3) cells ---- 228
Figure 5.15 Co-expression of pET21d/Cyclin A3 & pLNJ1/CDK1 (3) and pET21d/Cyclin B3 & pLNJ1/CDK1 (4) in B834(DE3) cells ---- 230
Figure 5.16 Co-expression of cell cycle regulatory proteins at 25°C ---- 231
Figure 5.17 Chromatogram depicting the purification of his-tagged CDK1-Cyclin A3 ---- 233
Figure 5.18 Affinity purification of CDK-Cyclin complexes and SDS-PAGE analyses ---- 234
Figure 5.19 Refolding of CDK1 and Cyclin B3 from inclusion bodies ---- 235
Figure 5.20 Primary amino acid sequence alignment of SpCDK1 and HsCDK2 ---- 236
Figure 5.20 Formation of CDK-cyclin complexes ---- 237

List of Tables

Table 5.1 Pairs of expression constructs used for co-expression ____ 213
Table 5.2 Summary of co-expression results ____ 232
Table 5.3 Residues involved inter-subunit contacts at the CDK-cyclin dimer interface ____ 244

5.1 INTRODUCTION

The overproduction of both *S. pombe* $p34^{cdc2}$ (SpCDK1), and *S. pombe* $p56^{cdc13}$ (SpCyclin B) in an *E. coli* host in quantities suitable for crystallographic studies has proven difficult (Dr. Endicott, personal communication). The initial part of the experiments described in this chapter was concerned with producing a truncated version of SpCyclin B for structural studies. This project was initiated after the successful protein engineering and subsequent crystallisation and structure determination of a truncated version of bovine cyclin A3 (Brown *et al.*, 1995). The protein engineering technique

utilised in this case was applied to SpCyclin B with a view to making this protein amenable to overproduction in a soluble form for structural studies. The aim was to address the molecular basis of CDK-cyclin specificity. The rest of the chapter is an account of experiments designed to co-express both SpCDK1 and SpCyclin B in an attempt to circumvent the difficulties encountered in the individual expression of these protein molecules. The crystal structure of the CDK2-Cyclin A complex (Jeffrey *et al.*, 1995) showed that the interactions at the protein-protein interface were mainly hydrophobic. Since CDKs and Cyclins are highly conserved families of proteins, the assumption underpinning the work described in this part of the chapter was that SpCDK1 and SpCyclin B probably possess hydrophobic patches which, if left unprotected by a protein subunit or an appropriate chaperone, might predispose these individual proteins to form higher order aggregates. One of the conclusions from Chapter 4 of this thesis and (Bourne *et al.*, 1996) was that the hydrophobic surface of SpSuc1 was used in mediating protein-protein interactions with SpCDK1. Since SpSuc1 is a highly soluble protein, it was anticipated that co-expressing SpSuc1 with SpCDK1 would enable SpSuc1 to bind to and rescue the nascent SpCDK1 polypeptides before they are packaged into inclusion bodies.

The co-expression of two cistrons in *E. coli* may be achieved via either the use of a bicistronic construct or two plasmids with compatible replicons. A bicistronic construct can be defined as a single plasmid bearing two over-expressible cistrons. The cistrons are commonly but not always arranged in tandem. Replicon-compatible vectors can be defined as two vectors that can co-exist in the same *E. coli* host cell by virtue of the mutual compatibility of their origins of replication. For example, the vector pFL260 carries the ColE1 replicon, which permits it to co-exist with the vector pFL261 (which carries the p15A replicon) in the same *E. coli* host cell (Larimer *et al.*, 1990). A bicistronic construct has the following advantages over the use of compatible plasmids; first, it is the method of choice in cases where strict control of stoichiometry is necessary to form a functional protein unit. Second, it requires the use of only one antibiotic resistance marker for plasmid selection whereas, an ideal replicon-compatible overexpression system requires that each plasmid carries the gene encoding a different antibiotic resistance e.g. β-lactamase (*bla*) to confer ampicillin resistance and

chloroamphenicol acetyltransferase (*cat*) to confer chloroamphenicol resistance. Third, a bicistronic construct allows us to reap the benefits (tight regulation of basal transcription) of using expression host cells like the pLysS and pLysE strains which already harbour a plasmid bearing the gene for lysozyme. One advantage of the replicon-compatible method over a bicistronic construct is the relative ease with which the former can be assembled. It is possible to maintain two incompatible plasmids with different antibiotic markers in *E. coli* by growing the transformed cells in the presence of high concentrations of antibiotics (Elce *et al.*, 1995; Graham-Siegenthaler *et al.*, 1994) but this approach has severe drawbacks such as rendering the growth media unconducive for viable cell growth, significant differences in plasmid copy number and poor yields of recombinant heterodimeric protein (Elce *et al.*, 1995; Larimer *et al.*, 1990).

"In the case of simple dimeric proteins, a binomial distribution of homo- and heterodimer forms is expected, with hybrid dimers equalling the 2pq term of the equation

$$\mathbf{1 = p^2 + 2pq + q^2} \qquad (1)$$

where **p** and **q** are the proportion of each type of subunit, **A** and **B**. This distribution is dependent upon the level of expression of each monomer, which is in turn dependent upon the level of copy number of each plasmid given identical expression cassettes" (Larimer *et al.*, 1990). This distribution also depends on the affinity of molecule **A** for **B.**

There are several accounts (Better *et al.*, 1988; Elce *et al.*, 1995; Graham-Siegenthaler *et al.*, 1994; Larimer *et al.*, 1990; Li *et al.*, 1997; Skerra and Pluckthun, 1988) of successful co-expression experiments in *E. coli* showing that this organism can be manipulated to assemble a functional heterodimeric protein unit in sufficient amounts for structural and biochemical studies. Attempts to express either whole or functional antigen-binding fragments of members of the immunoglobulin (Ig) family of proteins in *E. coli* has met with limited success (Boss *et al.*, 1984; Cabilly *et al.*, 1984) because the antibody was regularly directed into inclusion bodies (IBs) as insoluble aggregates. The yield of functional entities from refolding experiments tended to be too poor for structural studies. The V_L and V_H domains of a Fv fragment of the myeloma protein McPC603 have been successfully co-expressed in *E. coli* (Skerra and Pluckthun, 1988). The two cistrons, each equipped with a Shine-Dalgarno sequence, were placed in tandem to form a bicistronic construct under the control of the *lac* promoter. This resulted in the production

of a functional heterodimeric Fv fragment in the periplasm of the bacteria cells. With the expression conditions yet to be optimised, the authors reported a yield of 0.2mg/ml of purified Fv protein from a litre of cell culture (Skerra and Pluckthun, 1988). This investigation was carried out before the advent of the widely used and much more powerful T7 promoter system (Studier *et al.*, 1990) so there is scope for an increased yield of recombinant Fv. Similarly, the Fd fragment (24KDa) of a human antibody and the Kappa fragment (25.3KDa) from mouse were engineered into a bicistronic construct and placed under the regulation of the *arb* promoter from *S. cerevisiae* (Better *et al.*, 1988). Induction of protein expression in an *E. coli* host resulted in a yield of 2mg of an active heterodimeric unit (48Kda) from a litre of cell culture. This example shows that bacteria cells are capable of assembling a correctly folded mouse-human chimeric construct that displays a complex pattern of both inter- and intra-subunit disulphide linkages. The assembly process took place in the periplasm of the bacteria cells; this is a relatively less reducing environment, compared to the cytoplasm. Other reports of successful co-expression experiments in *E. coli* host cells include the production of functional human haemoglobin using a bicistronic cassette comprising α- and β-globin genes under the control of the *tac* promoter. The tetrameric protein product ($\alpha_2\beta_2$) constituted 5-10% of total cellular protein. Individual expression of these two genes produced soluble α-globin and completely insoluble β-globin protein (Looker *et al.*, 1994). Another tetrameric holoenzyme ($\alpha_2\beta_2$) successfully co-expressed in an *E. coli* host is human casein kinase II [(CKII), (Bodenbach *et al.*, 1994). The α- and β-subunit genes were placed in tandem to form a bicistronic construct in a pET (Studier *et al.*, 1990) expression vector. More recently (Li *et al.*, 1997), the retinoic acid receptor (RAR) and the retinoid-X receptor (RXR) have joined the list of success stories with respect to bicistronic co-expression in *E. coli.* Individually, RXR can be over-expressed to produce soluble protein but the majority of RAR was expressed as insoluble aggregates in inclusion bodies. The small amount of RAR that was soluble could not mediate protein-protein interactions with RXR. Co-expression resulted in a dramatic increase in the yield of soluble, active and stable RAR-RXR heterodimeric protein.

In the field of cell cycle research in this laboratory, Dr. Lawrie found that expressing CDK7 and Cyclin H as individual polypeptides in insect cells resulted in both

subunits being directed into inclusion bodies as insoluble aggregates. After changing the expression strategy to co-infection by two viruses so that both subunits were co-produced in the same cell, Dr. Lawrie obtained a yield of 5 milligrams of soluble, active CDK7-Cyclin H complex per litre of cell culture (Dr. Lawrie, personal communication). However, a drawback of co-infection of baculoviruses with two or more viruses is the lack of stoichiometry in the amount of proteins synthesised. This is analogous to the use of two compatible plasmids in *E. coli*. Multi-gene expression vectors that direct the simultaneous synthesis of two or more proteins in the same infected cell provide a method of producing stoichiometric amounts of multi-protein complexes. For example, insertion of the L3 and M7 genes of bluetongue virus (BTV) into a baculovirus dual expression transfer vector led to the overproduction of stoichiometric amounts of the gene products i.e. VP3 and VP7, respectively (French and Roy, 1990). This technique has been successfully extended to overproduce triple (Belyaev and Roy, 1993), quadruple (Belyaev and Roy, 1993; French *et al.*, 1990) and quintuple (Belyaev *et al.*, 1995) BTV structural proteins in a baculovirus expression system.

The study that provided the greatest impetus to pursue the course of experiments described in the latter part of this chapter was the investigation into the co-expression of calpains on two replicon-compatible plasmids and also, via the use of a bicistronic construct in *E. coli* (Elce *et al.*, 1995). In this paper, the authors described how one can transform the freely available cloning vector pACYC177 into an expression vector equipped with the powerful T7 promoter expression systems. Since the resultant vector carried the p15A replicon from pACYC177, it could co-exist in the same *E. coli* host cell with most *E. coli* expression vectors, including the popular pET expression vectors. This approach was extended to co-express SpCDK1 with SpCyclins B (full-length), SpCyclin B3, BvCyclin A3 and SpSuc1. An *E. coli* expression system was chosen over other systems for the following reasons; ease of genetic manipulation, efficient transformation and fast rate of growth, ease of scaling up e.g. for fermentation and comparatively low cost.

5.2 MATERIALS & METHODS

5.2.1 Escherichia coli host strains and plasmid vectors

E. coli strains HB101, DH5α and B834(DE3) were used for plasmid amplification (the first two) and expression, respectively. The pACYC177 vector was obtained from New England Biolabs (UK) Ltd. It possesses the p15A origin of replication and also, it carries the kanamycin-resistance gene and the *bla* gene for ampicillin resistance. It is a low copy number (~15/cell) plasmid. The pET28a vector was a generous gift of Dr. Leigh (MRC Immunochemistry unit, University of Oxford). It was originally obtained from Novagen. It is a medium copy number plasmid that carries the ColE1 origin of replication. It is kanamycin resistant. The pREP4-groESL plasmid (Dale *et al.*, 1994) was a gift from Dr. Kishore (MRC Immunochemistry unit, University of Oxford). The pET21d expression vector was a gift from the laboratory of Dr. Wigley (University of Oxford). The pET32c thioredoxin fusion vector and the corresponding bacterial host strains [AH494(DE3) and AH494(DE3)pLys-S] were gifts from Dr. Iorin (Department of Biochemistry, University of Oxford). The gene for SpSuc1 was subcloned into the pSJW1 expression vector (Winder and Kendrick-Jones, 1995) on an Nde I-Eco RI fragment as described in chapter 4, that for BvCyclin A3 was present in pET21d expression vector and the gene for full-length Cyclin B (*cdc13*$^{+}$) was borne by the pTZ19R cloning vector. All the restriction endonucleases and DNA modifying enzymes used in this chapter were obtained from New England Biolabs (UK) Ltd.

5.2.2 Production of cyclinBΔ195 DNA fragment

The following oligonucleotide primers were designed and purified according to the method described in chapter 4 section 4.2.7.

Forward primer = 5'-AAGAATCCATGGACCCTCTTATGGTTTCTG-3'

Reverse primer = 5'-TACACTCTCGAGCCATTCTTCATGTTTCATCTC-3'

Nco I and Xho I restriction sites (underlined) were incorporated into the forward and reverse primers, respectively. For the initial reaction, a cocktail containing 200μM of each dNTP (dCTP, dTTP, dATP and dGTP), 1unit of Deep Vent^{TM} DNA polymerase and $MgCl_2$ at three different concentrations (2, 4 and 6μM) was prepared. Eight hundred

nanograms (800ng) of each primer (forward and reverse) and 0.001 pmoles of the template was added to each reaction mixture on an ice-water bath. The total volume was made up to 50µl with sterile distilled water. Each reaction mixture was properly mixed, centrifuged briefly, and then topped with 50µl of mineral oil. Amplification parameters were an initial 4 minutes at 94°C, followed by 40 cycles of 1 minutes at 94°C, 1 minute at the annealing temperature of 54°C and 2 minutes at 72°C for the extension phase. DNA loading dye was added to each reaction and the mixture was analysed on a 1% agarose gel in TAE buffer.

5.2.3 Subcloning of cyclinBΔ195 DNA fragment

Several attempts at digesting the PCR product with Nco I and Xho I for subsequent subcloning directly into pET21d expression vector proved unsuccessful. Therefore, the cloning strategy was altered to one in which an intermediate cloning vector (pUC18) was used to subclone the flushed PCR product via a blunt-end ligation reaction (fig. 5.1).

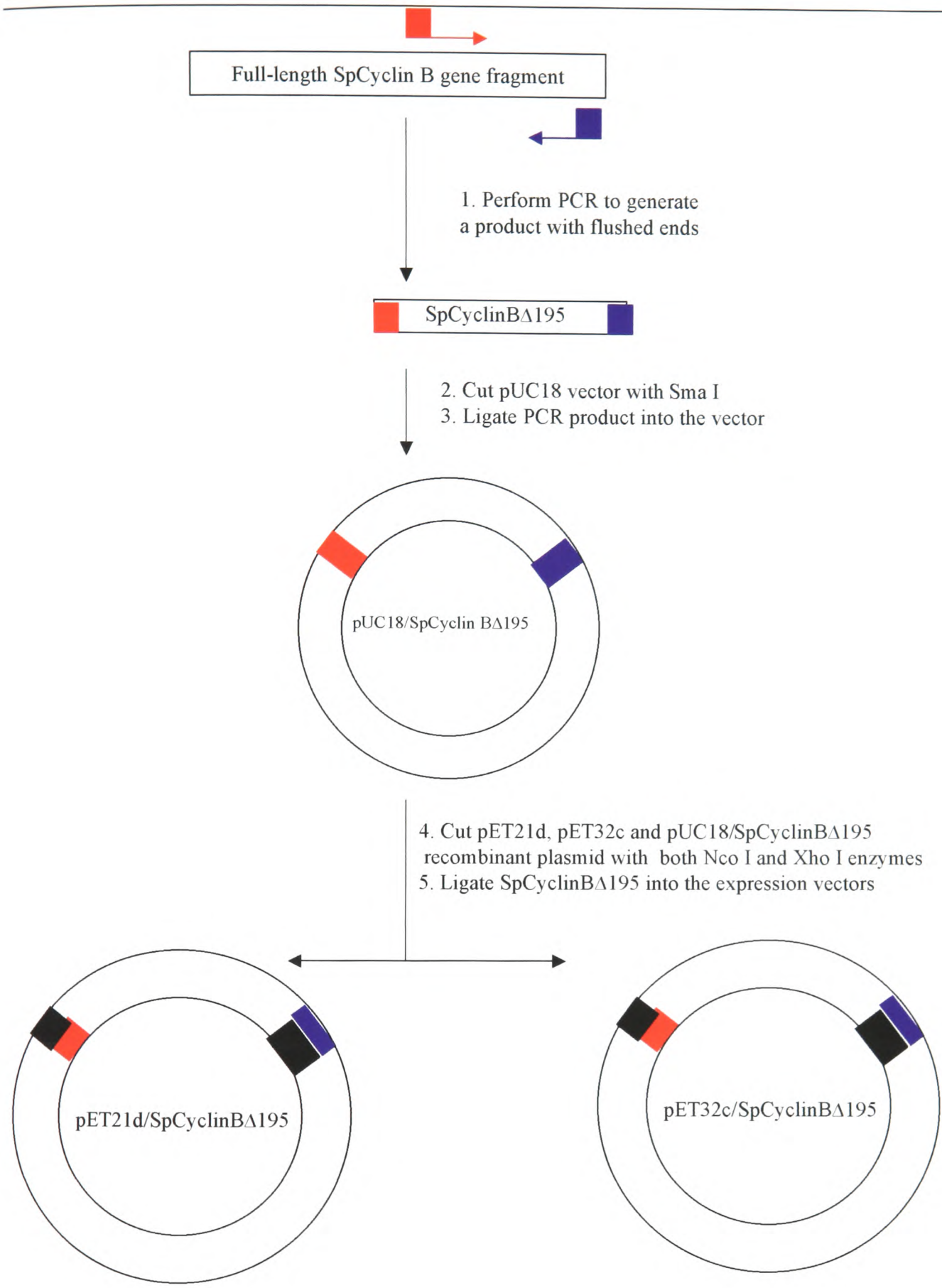


Figure 5.1 Schematic representation of the sub-cloning procedure

See text for details.

Briefly, about 3µg of pUC18 plasmid DNA was digested with 3 units of Sma I enzyme in the presence of Sma I buffer at 25°C in a total reaction volume of 10µl for 60

minutes. The Sma I enzyme was inactivated by heating the reaction mixture at 65°C for 20 minutes. In order to prevent the re-ligation of the vector in the subsequent ligation reaction, the 5'-phosphate group was removed with Calf Intestinal alkaline Phosphatase (CIP). Five units of CIP and 2µl of reaction buffer were added to the restriction digest in a total reaction volume of 20µl. This was incubated at 37°C for 60 minutes. The reaction was stopped by the addition of DNA loading dye. The mixture was analysed on a 1% agarose gel ran in TAE buffer at 140V for 60 minutes. The linearised pUC18 vector (2.7kb) was identified, extracted and purified using the DNA CLEAN-UP^{TM} kit (Promega, Southampton, U.K.) according to the manufacturer's instructions. Since the oligonucleotide primers used for the PCR reaction lacked a 5'-phosphate group, it was necessary (for the ligation reaction), to perform a kinase reaction on the PCR product to attach a 5'-phosphate group. To this end, 5ul of T4 DNA ligase buffer was added to 200pmoles of cdc13Δ195. The mixture was heated to 70°C for 5 minutes and then chilled on ice. Twenty (20) units of T4 polynucleotide kinase was added and the reaction mixture was incubated at 37°C for 30 minutes. The insert was purified and ligated into pUC18. The ligation mixture was then transformed into competent HB101 cells using standard procedures.

5.2.4 Production of recombinant protein

The pET21d/cdc13Δ195 expression construct was transformed into competent BL21(DE3)pLYS-S and B834(DE3)pLYS-S host strains. The latter has been shown to produce higher yields of recombinant protein (Doherty *et al.*, 1995). The transformed cells were spread on LB agar plates supplemented with ampicillin (100µg/ml) and chloroamphenicol (35µg/ml) and incubated at 37°C for 16 hours. Colonies were picked and tested for recombinant protein expression. For the rest of the thesis, the protein product of the construct shall be referred to as SpCyclin B3. A sterile 10µl pipette tip was used to transfer a single colony of transformed cells into LB media (10ml in five different 50ml-falcon tubes) containing ampicillin (50µg/ml) and chloroamphenicol (35µg/ml). The cell culture was grown with aeration at 37°C to mid-log phase before IPTG was added at four different concentrations (0.1, 0.2, 0.4 and 1.0µM) to induce protein expression. Cell growth was continued for four hours post-induction. The cells were

harvested and analysed on a coomassie-stained SDS-PAGE gel. A putative SpCyclin B3 band running at about 31KDa was identified and submitted for N-terminal protein sequencing. Examination of the soluble and insoluble fractions showed that the putative SpCyclin B3 protein was present exclusively in the pellet. Research efforts were directed towards ways of producing milligram quantities of soluble SpCyclin B3.

5.2.5 In vivo strategies for producing soluble SpCyclin B3

The initial focus was on the exploration of experimental conditions that would enable the transformed cells to overexpress SpCyclin B3 in a soluble form. Although the direction of the target polypeptide into inclusion bodies was a nuisance, fortunately one can exploit the inherent physiological plasticity of bacteria cells to improve the folding of target proteins via strategies such as the use of different growth media, alteration of growth temperature, titration of the chemical inducer of protein expression, and co-expression of homologous or heterologous folding accessory proteins and the use of fusion proteins (Georgiou and Valax, 1996; Hockney, 1994; Makrides, 1996; Weickert *et al.*, 1996).

5.2.5(a) Effect of growth media

There is at least one example in the literature (Moore *et al.*, 1993) demonstrating the preferential production of higher amounts of active protein in cells grown in rich media. This observation was also incorporated into the optimisation strategy for SpCyclin B3 production *in vivo*. Overnight cultures of the transformed cells were grown in LB media and 5ml of the cell culture was transferred to a 2litre flask containing either 500ml of LB (5g bacto tryptone, 2.5g bacto yeast extract and 5g NaCl), TB (6g bacto tryptone, 12 bacto yeast extract and 2ml of 100% glycerol) or SOB [2% (w/v) bacto tryptone, 0.5% (w/v/) yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$ pH 6.7) media and protein expression was conducted as described above under section 5.2.4.

5.2.5(b) Effect of temperature

Growing the bacterial culture at a temperature lower than 37°C either before or post-induction is a well-known technique for facilitating the correct folding of target

proteins (Cabilly, 1989; Makrides, 1996; Shirano and Shibata, 1990; Vasina and Baneyx, 1996). Some of the reasons why a lower temperature may favour the native state of the protein include a decrease in the driving force for protein self-association, a slower rate of protein synthesis and changes in the folding kinetics of the polypeptide chain (Georgiou and Valax, 1996). This variable was applied to the expression of SpCyclin B3 as follows: the overnight culture and pre-incubation cultures were grown at 37°C in the three different growth media mentioned in the previous section 5.2.5(a). When each starting culture had grown to mid-log phase, it was incubated for 30 minutes at 18°C and 25°C. IPTG (0.6mM) was then added and protein expression was continued for 4 hours in the case of the 25°C induction and 16 hours for the 18°C-degree induction. The 37°C induction was conducted as described under section 5.2.4. At the end of the induction period, the cell culture was harvested and aliquots (100µl) were processed as described in chapter 4 and analysed on SDS-PAGE gels.

5.2.5(c) Effect of IPTG titration

The experience of some members of the Laboratory of Molecular Biophysics had shown that screening a range of IPTG concentrations can result in the empirical determination of a particular concentration of IPTG that increases the yield and solubility of the target protein. The reasons for the success of this technique are similar to that given above for the effect of temperature. This technique was applied to the production of SpCyclin B3 in conjunction with different media types (LB, TB and SOB) and variable temperature 18, 25 and 37°C. The transformed cells were grown to mid-log phase and then induced with 50µM, 500µM and 1000µM of IPTG. The induced cells were harvested at the end of the various induction periods and analysed on SDS-PAGE gels.

5.2.5(d) Co-expression of SpCyclin B3 with E. coli chaperonins

Chaperone proteins were initially characterised as heat shock proteins (hsps) that were produced by bacteria during stress insult. Several families of chaperonins have now been identified including the hsp90, hsp70, hsp60, proline isomerase and protein disulphide isomerase families. The utility of chaperones as protein-folding reagents both *in vitro* and *in vivo* is well established (Cole, 1996). The general technique for carrying

out chaperonin-assisted protein expression involves transforming *E.coli* with separate chaperone-expressing and target-protein expressing plasmids bearing distinct antibiotic resistance functions and compatible replicating origins. In studies with protein tyrosine kinase C (Csk) (Amrein *et al.*, 1995) the expression of full-length Csk under standard conditions led to very low yields of soluble protein. In contrast, when the GroESL plasmid (pREP 4-groESL) was co-introduced with the Csk expressing plasmid (pD556/RBSII-Csk) into *E.coli,* the majority (≈70-80%) of the expressed Csk protein was found to be soluble under standard growth conditions. In the case of SpCyclin B3, competent *E. coli* B834(DE3) cells containing pET21/B3 expression plasmid were transformed with pREP 4-groESL and selected on LB agar plates containing ampicillin (100μg/ml) and kanamycin (35μg/ml). A single colony of the doubly transformed cells was picked and grown overnight in LB media supplemented with ampicillin (50μg/ml) and kanamycin (35μg/ml). This inoculum was used to check for protein expression in the three different media (LB, TB and SOB) and at three different temperatures (18, 25 and 37°C) using 0.6mM IPTG to induce protein expression in all cases.

5.2.5(e) Addition of 3% ethanol or 10% glycerol to the cell culture

Addition of ethanol (3% v/v) to the growth medium of the transformed cells post-induction could increase the yield of soluble protein (Thomas and Baneyx, 1996). The addition of ethanol is part of a variety of external stimuli (others include high temperature, bacteriophage infection, UV irradiation, presence of amino acid analogues, heavy metals or organic solvents in the growth medium) that induces the production of heat shock proteins (Thomas and Baneyx, 1996) which as described above, play an essential role in sherperding the nascent polypeptide to the properly folded, native state. It has been shown by the *in vitro* studies of β-lactamase folding that an intermediate with the characteristics of a molten globule is populated at intermediate denaturant concentration. This entity is prone to aggregation. Conditions such as the addition of sucrose, lower temperatures and the absence of reducing agents have been shown to prevent protein aggregation *in vitro* and correlate well with observations *in vivo* on inclusion body formation (Georgiou and Valax, 1996). Although glycerol, polyols and sucrose do not permeate the cell membrane and consequently cannot exert a direct effect

on the folding of cytoplasmic proteins, their addition to a growing cell culture results in an increase in the osmotic pressure which leads to the accumulation of osmoprotectants, such as glycine betaine. The latter has an effect similar to sucrose in stabilising the native protein structure. For example addition of sorbitol to cells grown at 25°C have been shown to produce >400-fold higher levels of recombinant protein (Blackwell and Horgan, 1991). *E. coli* cells transformed with pET21d/B3 were grown to mid-log phase and protein expression was induced as described. As in the other cases, this new factor was introduced both under standard conditions established for SpCyclin B3 (i.e. growth in TB media and induction at mid-log with 0.6mM IPTG and in conjunction with previous variable factors (i.e. different media and temperatures).

5.2.5(f) Expression of SpCyclin B3 as a thioredoxin-fusion protein

E. coli thioredoxin (product of *trxA* gene) has been shown to increase the yield and solubility of heterologous proteins when either co-expressed or fused to the target protein (LaVallie *et al.*, 1993; LaVallie and McCoy, 1995; Stoll *et al.*, 1998; Yasukawa *et al.*, 1995). The success of this method and its wide application has led to the development of several commercially available kits for producing thioredoxin-fusion proteins (Novagen, 1999). The SpCyclin B3 fragment was excised from pET21d by a double restriction digest with Nco I–Xho I and subcloned into pET32c using recombinant techniques described in the preceding chapter. The pET32c/SpCyclin B3 recombinant plasmid was then transformed into competent AD494(DE3) and AD494(DE3)pLys-S. The AD494 strain is a thoredoxin reductase mutant that allows disulfide bond formation in the *E. coli* cytoplasm (Novagen, 1999). The latter was plated onto LB agar plates containing ampicillin, kanamycin and chloroamphenicol and the former, onto plates supplemented with the first two antibiotics only.

5.2.6 In vitro strategies for producing soluble SpCyclin B3

The yield of recombinant protein from *in vitro* refolding experiments are about 30% for the best and more frequently below 10% for most recombinant proteins. Increasing the starting material can offset the losses (Cole, 1996). Protein expression under unfavourable thermodynamic and kinetic factors would enhance protein mis-

folding and lead to the packaging of more of the recombinant protein into inclusion bodies. So, to optimise the production of SpCyclin B3 inclusion bodies, the cell culture was grown to an OD_{600} of 1.0 at 37°C with shaking at 280rpm. Protein expression was induced by the addition of 1mM IPTG. The cells were harvested by centrifugation and either stored overnight at -20°C or used immediately for the subsequent inclusion body preparation.

5.2.7 Preparation of inclusion bodies (IBs)

The cell pellet from a 500-ml cell culture was thawed in 200ml of lysis buffer [100mM Tris-HCl (pH 7.5), 5mM EDTA and 270µg/ml of lysozyme plus 4 tablets of complete™ protease inhibitor cocktail (Boehringer Mannheim)]. Sodium deoxycholate [final of 0.1% (w/v)] was added to dissolve membranous material and the mixture was stirred at room temperature. Prior to the addition of DNase (10µg/ml) and RNase (5µg/ml) to degrade nuclear material, 8mM of $MgCl_2$ (pH 6.5) was added to the mixture. These enzymes require Mg^{2+} ions for efficient catalysis. The mixture was aliquoted (40ml) into 50-ml falcon tubes and each was sonicated on ice (3 x 20s) and centrifuged for 45 minutes at 8000rpm to collect the IBs. The IBs were washed with 100ml of 0.1% Triton X-100 containing protease inhibitors. After that, the IBs were spun, washed again as above and centrifuged to collect the inclusion bodies. The final washing step was conducted once with 200ml of distilled water containing protease inhibitors. These were also spun as above to collect the IBs.

5.2.8 Refolding recombinant SpCyclin B3 with 8M Urea buffer

The IBs were solubilised in 200ml of Urea buffer (8M urea, 50mM Tris-HCl pH 8.5, 20mM DTT) and stirred overnight at 4°C. The mixture was clarified by centrifugation at 8000rpm at 4°C. The supernatant was filtered through a 0.45µm filter unit. The volume was made up to 200ml (containing about 0.3-0.6µg/ml of protein) with urea buffer and the protein mixture was dialysed in a step-wise manner against 2l of buffer containing decreasing (6, 5, 4, 3 and 2.5 and 2.0M) of urea. At between 2.5-2.0M, the protein solution got noticeably turgid so the refolding process was stopped and the

mixture was clarified by centrifugation at 8000rpm at 4°C for 30 minutes. The supernatant was concentrated and analysed by SDS-PAGE.

5.2.9 Refolding of SpCyclin B3 by immobilised metal affinity chromatography (IMAC)

The ability of the Ni-NTA column to bind selectively to polyhistidine-contatining proteins under both denaturing and non-denaturing conditions was exploited in an attempt to refold SpCyclin B3 using the IMAC method (Crowe *et al.*, 1994; Hochuli, 1990; Hochuli *et al.*, 1987). The immobilised protein was refolded either in a stepwise (i.e. from 6 to OM GdnHCl) (Holzinger *et al.*, 1996; Oswald and Rinas, 1996; Rogl *et al.*, 1998) or gradual manner (gradient from 6-0.5M GdnHCl) (Holtet *et al.*, 1994; Shi *et al.*, 1997; Stempfer *et al.*, 1996). An overnight culture of B834(DE3)pLys-S cells containing pET21d/B3/ was grown in 10ml of TB media supplemented with ampicillin (50μg/ml) and chloramphenicol (35μg/ml). Five millilitres (5ml) of the ONC was transferred into fresh 500ml of TB media supplemented with ampicillin (10μg/ml) and grown at 37°C with vigorous shaking at 300rpm. The cell culture was allowed to grow to an OD_{600} of 0.8 before protein expression was induced by the addition of IPTG at a final concentration of 0.6mM. The bacterial cells were harvested after 4hrs and the pellet was re-suspended in 25ml of buffer I (6M GdnHCl, 20mM Tris-HCl pH 7.9, 500mM NaCl and 4mM octylglucoside). The cells were lysed by sonication and the soluble and insoluble fractions separated by centrifugation at 40000rpm for 60 minutes. The increased speed of centrifugation was necessary to remove existing aggregates that can act as nuclei to trigger aggregation during refolding. The supernatant was clarified by passing it through a 0.45μm filter paper (Whatman) and then applied to an Ni-NTA agarose column which had previously been equilibrated with buffer I. The NTA column was washed extensively with buffer II (20mM Tris-HCl pH7.9, 150mM NaCl) and the immobilised protein eluted with the same buffer containing 50mM EDTA. The eluted fractions were assessed on an 18% SDS-PAGE. A single band running at about 31KDa was observed and submitted for N-terminal protein sequencing. The pooled fractions were concentrated to an appropriate volume before application to a Superdex-75 (Pharmacia) gel filtration column.

5.2.10 Iterative refolding of SpCyclin B3

This is a variation on the method described above in which the target protein was immobilised on the column matrix in denaturing buffer and then washed with a buffer devoid of denaturant. In this method, the target protein is immobilised as described above but whilst it is still immobilised on the column, it is subjected to more than 20 cycles of renaturation and denaturation (Holtet *et al.*, 1994). Each cycle consisted of a renaturation step where the column was exposed to a non-denaturing buffer followed by a pulse of denaturing buffer. In the overall protocol, the concentration of denaturant was reduced from cycle to cycle. At the end of the iterative refolding process, the His_6 fusion protein was eluted from the column and analysed by SDS-PAGE. According to the authors, this method was successfully used to refold about 20 different disulphide-containing, eukaryotic proteins expressed in *E. coli* (Holtet *et al.*, 1994). The iterative refolding protocol is based on the notion that the folded state of a protein is thermodynamically or kinetically more stable than the unfolded and intermediate forms. Also, completely folded proteins tend to unfold more slowly than intermediately or misfolded proteins when exposed to denaturants. The overall process can be separated into three phases; phase I in which all the misfolded and newly refolded protein is unfolded during the denaturation pulse. Phase II (productive phase) where the misfolded protein is unfolded during the denaturation pulse but the folded protein is not; this leads to an incremental accumulation of folded protein during these cycles. Phase III (final phase) where the denaturation pulse affects the structure of neither misfolded nor folded protein. SpCyclin B3 was expressed as described in the preceding section and the total protein was extracted into a lysis buffer [6M GdnHCl, 150mM NaCl, 50mM Tris-HCl pH 8.0 and 0.1% (v/v) β-mercaptoethanol]. The mixture was clarified by centrifugation at 40000rpm at 25°C for 60 minutes. The supernatant was filtered through a 0.44µm filter unit and then applied to a pre-equilibrated, Ni-NTA agarose column at room temperature. The His_6-SpCyclinB3 was purified under denaturing (8M urea, 150mM NaCl, 50mM Tris-HCl pH 8.0 and 0.1% (v/v) β-mercaptoethanol) and non-denaturing (150mM NaCl, 50mM Tris-HCl pH 8.0 and 0.1% (v/v) β-mercaptoethanol) buffers with an FPLC (Pharmacia) programme that successively exposed the column to less and less amount (by volume) of the denaturing buffer but more of the renaturing buffer from the first

cycle to the last (≈26 hours) so that at the end, the column was washed exclusively with the renaturing buffer before the protein was eluted with an increasing gradient of imidazole (0-0.5M) in the renaturing buffer.

5.2.11 Homology modelling of SpCyclin B3

In certain cases, the presence of exposed, hydrophobic patches on the surface of a protein molecule can predispose the nascent protein to aggregation with similar molecules *in vivo* (Lijnzaad *et al.*, 1996). The expression of HIV 1 integrase for crystallographic studies was an intractable problem because of its low solubility and propensity to aggregate (Jenkins *et al.*, 1995). However, a systematic point mutagenesis approach in which each hydrophobic residue in the core domain was mutated to either a lysine or alanine led to the production of soluble protein which was crystallised and the X-ray structure was successfully determined (Dyda *et al.*, 1994). The point mutation of lysine for phenylalanine at position 185 resulted in a core domain of HIV 1 integrase that was highly soluble, monodispersed in solution and retained catalytic activity (Jenkins *et al.*, 1995). This formed the impetus to build a homology model of SpCyclin B3 in an attempt to identify any extensive hydrophobic patches that may exist on the surface of the protein molecule. In this analysis, a crude homology model of SpCyclin B3 was built using the crystal structure of BvCyclin A3 as a template. The primary amino acid sequence of BvCyclin A3 and SpCyclin B3 were aligned with the GAP programme in the GCG software suite (Senger *et al.*, 1998) and displayed with ALSCRIPT (Barton, 1993). Non-conserved residues in the BvCyclin A3 crystal structure were then mutated to the equivalent in SpCyclin B3 using the graphics program O (Jones, 1991). The co-ordinates of the SpCyclin B3 model were written out and the model were subjected a brief round of energy minimisation in XPLOR (Brunger, 1992). Surface hydrophobicity was calculated with GRASP (Nicholls *et al.*, 1993) and the image was displayed in XOBJECTS (Dr. Noble, unpublished work).

5.2.12 Homology modelling of proteins by email

The SpCDK1 and SpCyclin B 3-D models were generated by submitting the SWISSPROT accession codes (P04551 for SpCDK1 and P10815 for SpCyclin B) to a

web-based modelling server at www.expasy.ch/swissmod/SWISS-MODEL.html. This service utilises an automated, knowledge-based, protein-modelling program called ProMod (Peitsch, 1995; Peitsch, 1996; Peitsch *et al.*, 1995) to predict the structure of a protein from primary sequence information. Swiss-model predicts protein structure by the extrapolation of a model for a new (target) sequence from the known atomic structures of related family members (templates). Although the number of protein sequences in the swiss-prot database is about 80000 (swiss-prot release 37.0), there are only about 8000 sets of atomic co-ordinates in the protein data bank (PDB). Since a high sequence similarity (>50%) often leads to distinct structural similarities between 2 proteins, knowledge-based modelling can be used to glean structural information in the absence of an experimentally determined atomic model. It has been estimated that for protein cores sharing 50% residue identity, every pair of corresponding atoms in the 2 structures are, on average, only 1Å apart (Chothia and Lesk, 1986; Peitsch, 1995). So, a knowledge-based model can (with a degree of confidence) be utilised in the rational design of site-directed mutagenesis experiments and also act as a basis for designing protein constructs for overexpression *in vivo*. The limitations of knowledge-based modelling are first, the resolution of the template can sometimes be low thus affecting the accuracy of the target model, and second, the method would fail in the absence of "knowledge" i.e. an experimentally determined template structure that shares both sequence and 3-D structural similarities with the target protein (Peitsch, 1995).

Depending on the state of traffic on the world-wide web (www), the results of a modelling request can be returned within an hour; it took about 5-10 minutes for the models presented in this chapter. Operationally, when a sequence is submitted to the swiss-model server, the first thing it does is to search through the PDB using BLAST (Altschul *et al.*, 1990), FastA (Pearson and Lipman, 1988) and SIM (Huang and Miller, 1991) looking for template proteins (experimentally determined structures) that show primary sequence similarities with the target protein. The modelling task is aborted if no protein of known structure shares at least 35% sequence identity with the target protein (Peitsch, 1995). If appropriate templates are found, swiss-model then builds a framework structure (skeleton) for the target model using sequence alignment tools and 3-D superimposition tools. At this stage, swiss-model achieves optimal alignment by ignoring

insertions and deletions (which are generally restricted to loop regions) that interrupt the core secondary structure elements (α-helices and β-sheets). After the construction of the framework structure, ProMod generates the atomic co-ordinates of the nascent target model. The quality (i.e. correctness) of the final model is crucially dependent on these two initial steps i.e. the correct identification of the conserved core or target protein and its alignment with one or more similar template sequences. Completion of the model involves the following steps: first, the rebuilding of the non-conserved loops from their stems by structural similarity searches through the PDB, second, completion of the main chain using a library of backbone elements (pentapeptides) derived from high resolution (<2Å) X-ray structures, third, the addition of side chains that are absent from the template proteins and fourth, the correction of existing side chains using a library of permitted rotamers. After the addition of all the biochemical moieties, the target model is optimised by a few cycles of minimisation using CHARMM (Brooks *et al.*, 1983). This process is aimed at producing the best possible overall stereochemistry of the target model by improving intramolecular contacts and relieving strain. Finally, assessment of the new model (akin to the validation stage of an X-ray structural determination experiment, see Appendix I) is carried out by the calculation of 3-D profiles (Luthy *et al.*, 1992) and a reliability factor (called the C-factor) is derived for each residue in the new model based on the degree of structural similarity between template and target model.

5.2.13 Construction of pLNJ1 expression vector

The 1094bp Bam HI-Sma I fragment containing the C-terminal half of the kanamycin-resistance gene was excised from the pACYC177 vector and substituted with the 1470bp Bgl II–SmaI fragment from pET28a vector to give a circular DNA plasmid vector of 4316bp. Since the latter fragment also contains the C-terminal half of the kanamycin-resistance gene, this operation re-created a full-length kanamycin-resistance gene in the resulting plasmid expression vector called pLNJ1. The presence of the β-lactamase gene in the pACYC177 template conferred ampicillin resistance on pLNJ1.

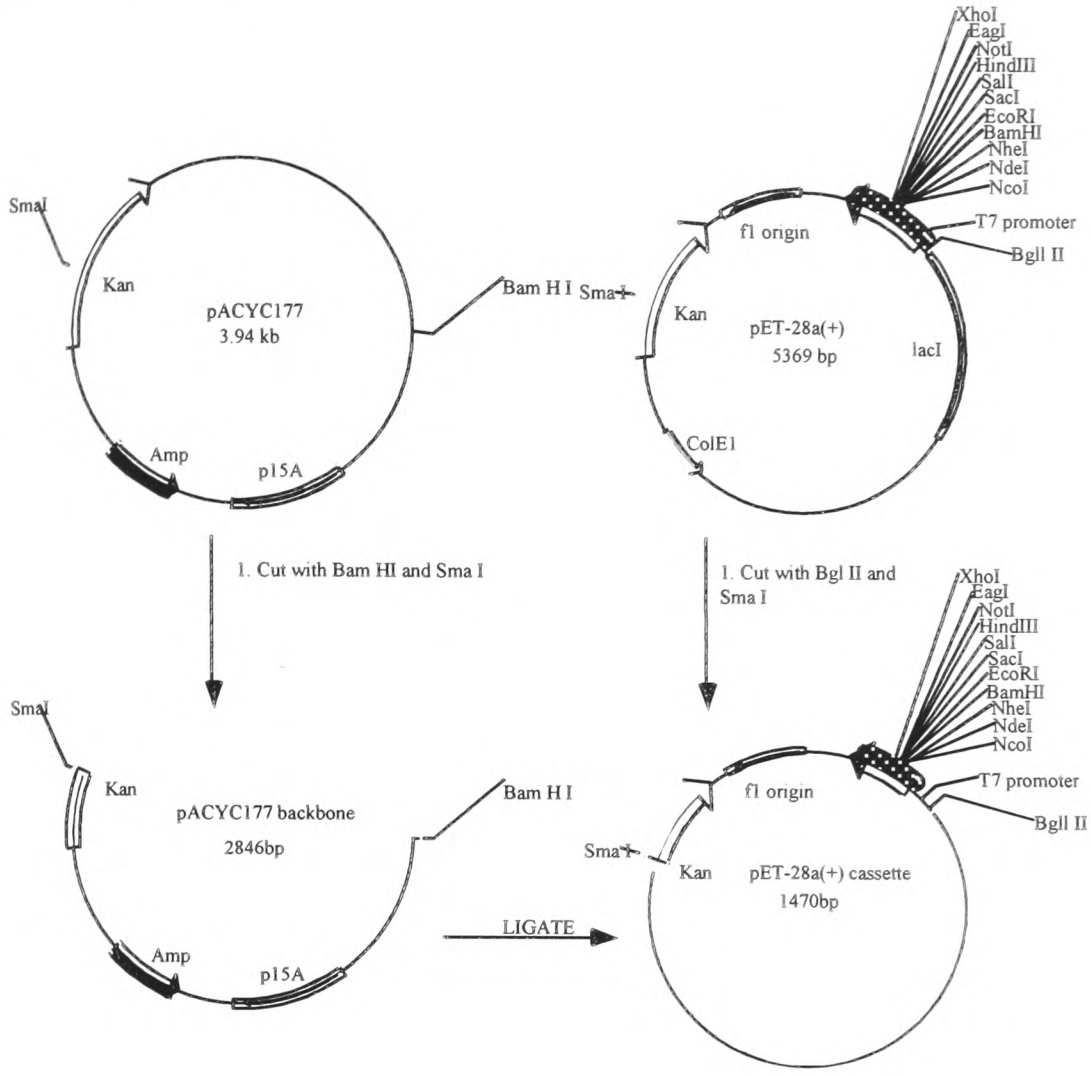


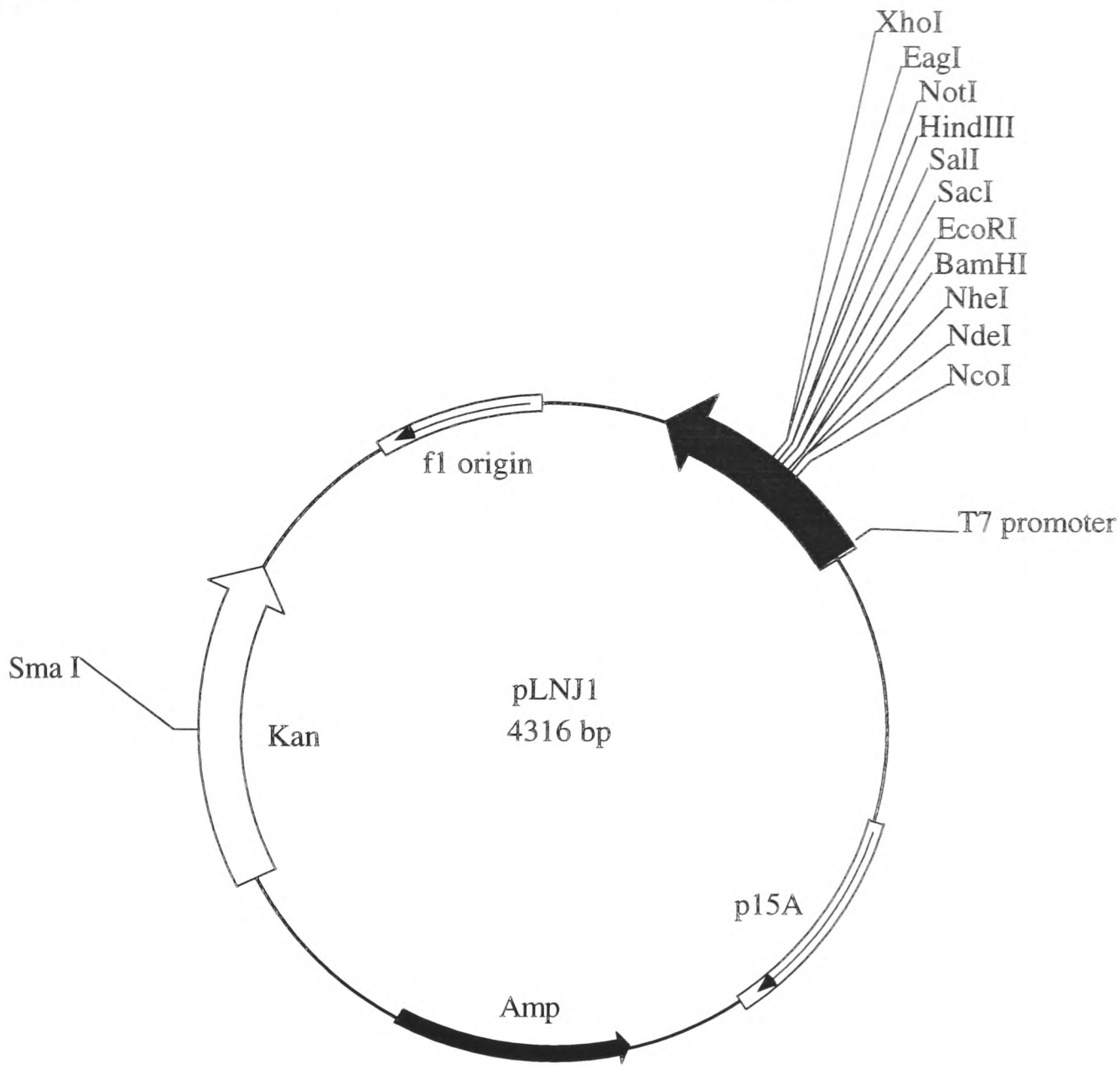
Figure 5.2 Construction of pLNJ1 expression vector

See text for details.

5.2.14 Restriction digests, ligation reactions and the harvesting of pLNJ1 expression vector

Three units of the restriction enzyme Bam HI and 2µl of 10X reaction buffer were added to 3-5µg of the plasmid DNA in a sterile eppendorf tube. The total volume was made up to 20µl with sterile distilled water. The reaction mixture was incubated at 37°C for 1 hour followed by heat inactivation at 65°C for 20 minutes. Three units of Sma I enzyme and 3µl of 10X-reaction buffer were added and the total volume was made up to 30µl with sterile distilled water. This was incubated at 25°C for an hour at the end of which DNA loading dye was added and the digest was loaded onto a 1% DNA agarose gel and electrophoresced at 140V for 60 minutes in 1X-TAE (0.04M Tris-Acetate, 0.001M EDTA pH 8.0) buffer. The relevant DNA fragments were identified under a UV

lamp and excised with clean blades into separate sterile eppendorfs for the subsequent purification step using Wizard™ DNA clean-up kit (Promega). One hundred nanograms (100ng) of the 2846bp Bam HI-Sma I fragment of the pACYC177 backbone was added to a 100ng of the 1470bp Bgl II-Sma I insert carrying the T7 promoter, Shine-Dalgarno sequence, an S-tag, multiple cloning sites (MCS), N- and C-terminal hexahistidine tags and a termination codon. The reaction mixture was completed by the addition of ATP (1mM), 1µl of 10X-ligation buffer and 1µl of T4 DNA ligase. The total volume was made up to 15µl with sterile distilled water and the reaction was incubated and transformed into competent DH5α cells as described in chapter 4. Typical yields were about 100-300 colonies per plate. A single colony was picked and transferred into 10ml of LB media supplemented with ampicillin (50µg/ml) in a 50ml falcon tube. This was grown with aeration at 37°C for 16 hours. This initial inoculum was used to inoculate a litre of fresh LB media containing ampicillin (50µg/ml). Cell growth was continued at 37°C with aeration for a further 8 hours before the addition of chloroamphenicol ((35µg/ml).The culture was left to grow overnight with aeration at 37°C. The cells were harvested by centrifugation and the QIAGEN midi-prep kit was used according to the manufacturer's instructions to isolate pure pLNJ1 expression plasmid. Amplification with chloroamphenicol increased the yield of this low copy number plasmid from 80µg/ml to about 500µg/ml from a litre of cell culture.



Plasmid name: pLNJ1
Plasmid size: 4316 bp
Constructed by: Oki K. Dzivenu
Construction date: 18/6/1997.
Comments: Low copy number plasmid. Both ampicillin and kanamycin resistant. Requires amplification with choroamphenicol in order to obtain sufficient yields from DNA preparations.

Figure 5.3 Schematic representation of pLNJ1 expression vector

The following features from pET28a have been incorporated into the pLNJ1 expression vector: N- and C-terminal hexahistidine tags, a thrombin cleavage site and a T7 tag. The others are T7 promoter, transcription start and terminator sequences. Also, pLNJ1 carries an f1 origin so that infection of a transformed bacterial culture with helper phage (e.g. M13KO7) would produce ssDNA for other applications such as sequencing or mutagenesis.

5.2.15 Construction of pLNJ/SpCDK1 and pLNJ1/SpCyclin B plasmids

The 882bp SpCDK1 gene was excised from the pRK172 expression vector on an Nde I-Bam HI fragment and subcloned into pET28a expression plasmid. The resulting recombinant plasmid, pET28a/SpCDK1 was cut with Bgl II and Sma I restriction enzymes to release the 2352bp expression cartridge containing the SpCDK1 gene under the control of the T7 promoter. This fragment was then ligated into the pACYC177 vector (which had undergone prior digestion with Bam HI and Sma I enzymes) to form a new recombinant expression plasmid called pLNJ1/SpCDK1.

Similarly, in the construction of pLNJ1/SpCDK1 the 1446bp *S. pombe cdc13+* gene was excised from pTZ19R cloning vector on an Nde I-Bam HI fragment and subcloned into pET28a to give pET28a/Cyclin B recombinant plasmid. The 2916bp Bgl II-Sma I expression cassette containing the *cdc13+* gene was excised from pET28a/Cyclin B and slotted into the linearised pACYC177 vector to give a new plasmid expression construct called pLNJ11/Cyclin B.

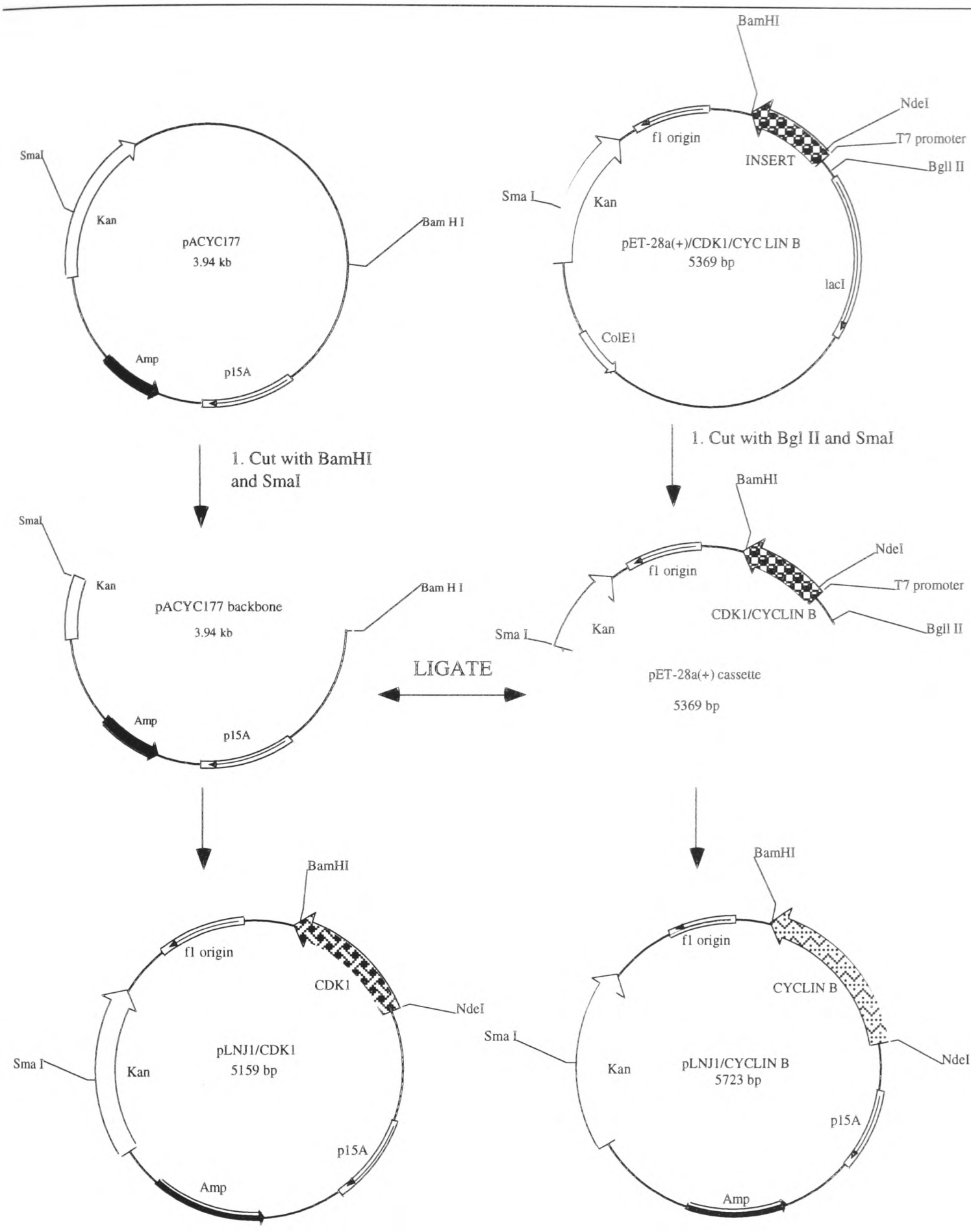


Figure 5.4 Overview of pLNJ1/CDK1 and pLNJ1/Cyclin B expression constructs

See text for details.

5.2.16 Co-expression strategies

The following plasmids were transformed into competent B834(DE3) cells; pET21d/A3, pET21d/B3, pSJW1/SpSuc1 and pRK172/SpCDK1. Each transformed cell was picked and made competent using the Transformation and Storage Solution (TSS) method (Chung *et al.*, 1989). Briefly, a fresh overnight culture of bacteria was diluted 1:100 with pre-warmed LB broth containing ampicillin (50μg/ml). This was incubated at 37°C with aeration (shaking at 250rpm) to an OD_{600} of 0.3-0.5. An equal volume of ice-cold 2X TSS (20% PEG 3350, 10% DMSO, 50mM $MgCl_2$ at a final pH of 6.5) was added and the cell suspension was mixed gently. For transformation, a 200μl aliquot of cells was pipetted into a cold polypropylene tube containing 100pg of pLNJ1/Ω (where Ω is either SpCDK1 or SpCyclin B) plasmid DNA and the mixture was gently mixed. The cell/DNA mixture was incubated for 30 minutes at 4°C (ice-water bath). A 600μl aliquot of LB broth was added and the cells were incubated at 37°C with shaking at 250rpm for 60 minutes to allow for the expression of the antibiotic-resistance genes (ampicillin and kanamycin). The cells were spread on LB agar plates containing ampicillin (100μg/ml), Kanamycin (35μg/ml). After an overnight incubation at 37°C, transformants were picked and tested for protein expression as described in the Materials & Methods section of chapter 4.

Plasmid DNA	**Co-expressed with**
pLNJ1	-
pLNJ1/SpCDK1	-
pRK172/SpCDK1	-
pET21d/Cyclin A3	pLNJ1/SpCDK1
pET21d/SpCyclin B3	pLNJ1/SpCDK1
pSJW1/SpSuc1	pLNJ1/SpCDK1
pRK172/SpCDK1	pLNJ1/full-length Cyclin B

Table 5.1 Pairs of expression constructs used for co-expression

The ideal co-expression vector should carry only one antibiotic resistance gene, which should be different from the marker in the second plasmid to allow for dual selection of both plasmids. pLNJ1 is both ampicillin and kanamycin resistant. All the other vectors are ampicillin resistant. This meant that pLNJ1 expression constructs could only be transformed into the expression host strain after the other plasmid partner has been transformed.

5.2.17 Non-denaturing affinity purification of his-tagged proteins

Although SDS-PAGE analysis showed that SpCyclin B3 was exclusively directed into inclusion bodies, an attempt was made to exploit the high affinity binding that the hexahistidine tag has for certain divalent metal ions to capture any soluble SpCyclin B3 molecules that may be present in the soluble cytoplasmic extract but were quantitatively, too little to be observed on SDS-PAGE. Transformed cells were grown, induced and harvested as described. The cell pellet from a 500ml bacterial culture was extracted into 50ml of start buffer (0.02M Na_2HPO_4, 0.5M NaCl pH7.2) supplemented with 10mM $MgCl_2$, RNase (2µg/ml), DNase (4µg/ml) plus a tablet of complete™ protease inhibitor cocktail (Boehringer Mannheim). The mixture was sonicated and clarified by centrifugation as described and the supernatant was filtered through a 0.22µm filter paper (Whatman) before it was loaded onto a pre-equilibrated 1ml HiTrap™ (Pharmacia) column charged with Co^{2+} ions according to the manufacture's instructions using a Gradifrac™ (Pharmacia) protein purification system. The column was washed with start buffer till the chromatographic profile returned to near-baseline. Bound protein was eluted with an increasing gradient of imidazole in elution buffer (0.02M Na_2HPO_4, 0.5M $NaCl_2$ and 0.5M imidazole). Aliquots of the eluates were analysed on SDS-PAGE gels.

5.2.18 Simultaneous refolding of SpCDK1-SpCyclin B3 complex

The IMAC method (gradient refolding) was used to co-denature and refold co-expressed SpCDK1-SpCyclin B3 molecules from IBs.

5.3 RESULTS

5.3.1 Amplification and cloning of SpCyclin B3

The preliminary amplification reactions were performed with a titration of $MgCl_2$ in the reaction buffer (2mM, 4mM and 6mM) in order to determine the optimal reaction conditions. Both the 2mM and the 6mM reaction mixture yielded a secondary reaction product of about 500bp (data not shown). The 4mM $MgCl_2$ reaction mixture produced a single product of the expected size (861bp). A subsequent reaction was performed using these conditions (fig. 5.5).

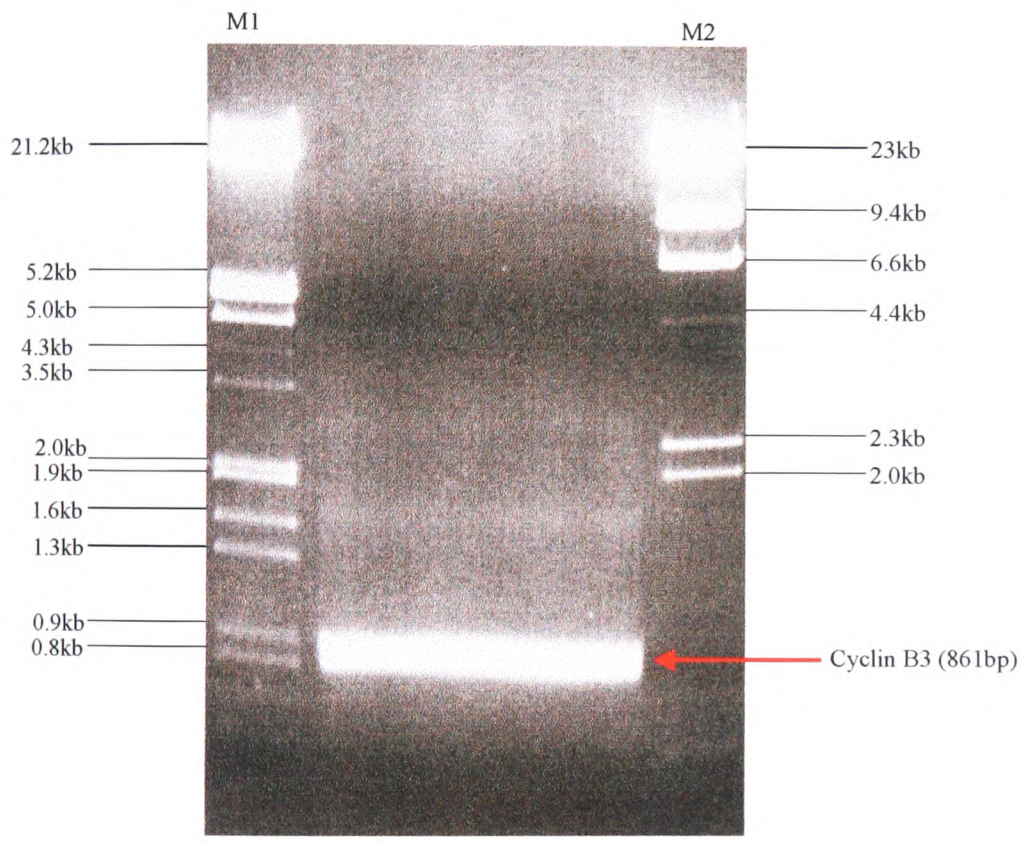


Figure 5.5 Production of SpCyclin B3Δ195 DNA fragment

DNA loading dye was added to the total reaction mixture (100µl) and this was ran on 1% preparative agarose DNA gel in TAE buffer.

Several attempts at subcloning the PCR product directly into pET21d (after digesting both the vector and the insert with Nco I and Xho I restriction enzymes were not successful). The inability of one or both enzymes to digest SpCyclin B3Δ195 could be attributed to insufficient length of overhangs at the 5' and 3' ends of the insert. The SpCyclin B3Δ195 fragment was equipped with the recommended minimum number of overhangs for Nco I (6 nucleotide bases) and Xho I (6 nucleotide bases) (NEB, 1999). To overcome this obstacle, the uncut PCR product was cloned in a blunt-ended ligation reaction into pUC18 cloning vector and then subcloned successfully on an Nco I-Xho I

DNA fragment into pET21d and pET32c expression vectors. Full-length DNA sequencing of SpCyclin B3Δ195 showed that the intended mutations to incorporate the 5' and 3' restriction site had occurred and the gene fragment was devoid of spurious mutations.

5.3.2 In vivo attempts at producing soluble SpCyclin B3

The exploratory experiments were geared towards the empirical determination of optimal expression conditions for SpCyclin B3. Competent BL21(DE3), BL21(DE3)pLys-S, B834(DE3), and B834(DE3)pLys-S were transformed with pET21d/B3 and tested for protein expression. Examination of the uninduced and induced lanes for each expression host strain showed that the protein was being expressed in all the different bacteria strains to comparable levels (data not shown). Subsequent optimisation experiments were carried with transformed B834(DE3)pLys-S because it has been demonstrated that this host strain produces relatively higher yields of recombinant protein (Doherty *et al.*, 1995).

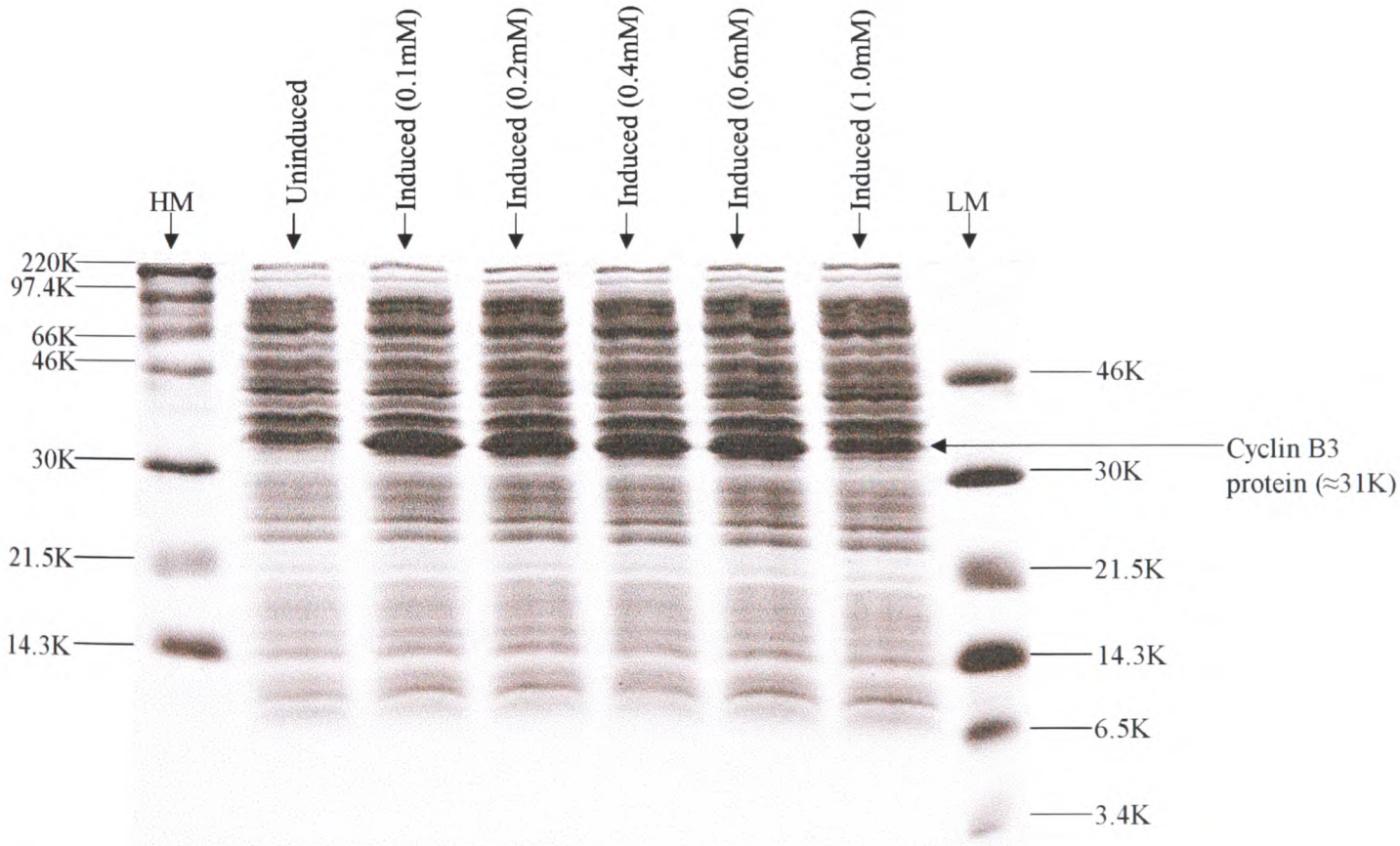


Figure 5.6 Determination of optimal expression conditions for SpCyclin B3 protein

Preliminary expression results had shown that SpCyclin B3 was best expressed in TB medium at 37°C with 0.4mM IPTG. In a subsequent experiment to further optimise the expression conditions, all other factors were maintained but the IPTG concentration was varied: examination of the SDS-PAGE gels showed that induction of protein expression with 0.6mM IPTG produced a marginally higher yield of SpCyclin B3. Theoretical calculation of the expected size was 29KDa with a pI of 5.4 (www.expasy.ch/tools/pi_tool.html), but the band identified by N-terminal protein sequencing as SpCyclin B3 molecule migrated at bout 31KDa. The increased size was due to the extraneous vector-encoded amino acid residues attached to the C-terminus of SpCyclin B3.

Various attempts at exploiting the physiological versatility of *E. coli* to determine empirical conditions that would lead to the overexpression of soluble and active SpCyclin B3 *in vivo* were not successful. The optimal conditions for the production of SpCyclin B3 were determined as induction with 0.6mM IPTG at mid-log phase (OD_{595} of 0.4-0.6) at 37°C for 4 hours in TB media. Expression at temperatures below 37°C (18 and 25°C) led to a reduction in the amount of SpCyclin B3 expressed (data not shown) and almost all the expressed protein was directed into IBs. Attempts at slowing down the rate of protein synthesis by a combination of lower temperatures and relatively lower concentrations of IPTG yielded similar results (data not shown). Addition of either 3% ethanol or 10% glycerol post-induction did not result in the overexpression of SpCyclin B3 in a soluble and active form (data not shown). Co-expression of SpCyclin B3 with plasmid-encoded *E. coli* chaperones (GroESL) gave an intriguing result (data not shown). Both chaperones were highly overexpressed but the supernatant contained only about 50% of the GroEL (≈56KDa) molecule and a similar amount of GroES (≈14KDa) as judged by SDS-PAGE. The other half of the chaperones were encapsulated in IBS together with almost all of the overexpressed SpCyclin B3 molecules. Finally, an attempt was made to render SpCyclin B3 soluble *in vivo* by expressing it as a fusion protein to *E. coli* thioredoxin. The latter is a small (109 amino acids, 11.6KDa) protein which is highly soluble in the *E. coli* cytoplasm, thermally stable and localises to the cytoplasmic face of the adhesion zones between the inner and outer cell membranes. If these features of thioredoxin are

conferred on the resulting fusion protein, it was expected to produce a highly soluble molecule, which would be correctly folded and relatively easy to purify. The thioredoxin moiety can be cleaved with enterokinase.

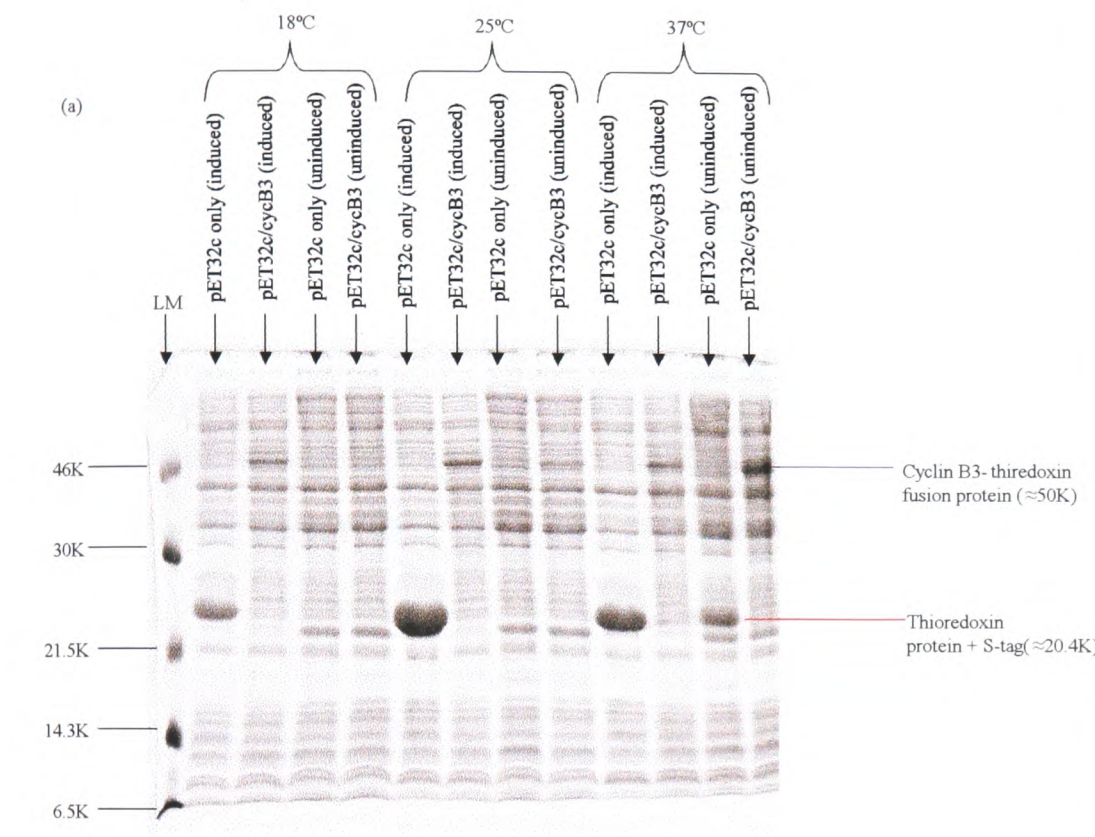


Figure 5.7 Expression of SpCyclin B3 as a thioredoxin fusion protein *B834(DE3)pLys-S cells transformed with pET32c/B3 were grown in TB media and tested for protein expression at (a) various temperatures using 0.6mM IPTG and (b) 37°C using three different concentrations of IPTG.*

Although the size of native *E. coli* thioredoxin is 11.6KDa, the pET23c vector was constructed so that the *trx*-tag was upstream of the multiple cloning site (MCS) and since the trx gene had to be devoid of a 3' stop codon so that there could be uninterrupted read through from the 5' end of the *trx* gene all the way to the 3' end of the SpCyclin B3 gene fragment. This arrangement of the expression cassette meant that in the absence of any insert, there would be uninterrupted read through from the 5' end of the fused *trx* gene to the T7 terminator sequence (Novagen, 1999) thus adding 98 amino acids to the C-terminal end of thioredoxin. This was expected to add at least an extra 10KDa to the 11.6KDa native protein hence the band migrating at about 21KDa.

5.3.3 In vitro attempts at producing soluble SpCyclin B3

Heterologous expression of eukaryotic proteins in *E. coli* often leads to the production of insoluble aggregates of misfolded protein called inclusion bodies. The sole advantage of having a target protein directed into inclusion bodies is that it provides a relatively easy means of isolating the target protein. However, the solubilisation of the target protein can only be achieved in the presence of strong denaturants such as 8M urea or 6M GdnHCl. The main problem with this approach is the inability to achieve efficient refolding *in vitro*. The difficulty lies in how to lower the denaturant concentration to allow folding while at the same time preventing aggregation. Published methods for eliminating the chemical denaturant have involved dialysis and dilution (London *et al.*, 1974). The advent of the metal affinity purification protocols have resulted in the evolution of variants of column refolding protocols where the target protein is immobilised on the column via a hexahistidine tag and the denaturant is either eliminated in a stepwise manner (Holzinger *et al.*, 1996; Oswald and Rinas, 1996; Rogl *et al.*, 1998) or gradually (Holtet *et al.*, 1994; Shi *et al.*, 1997; Stempfer *et al.*, 1996). Initially, the dialysis method was used to refold SpCyclin B3 protein. The denatured protein was dialysed from 8 to 2M urea concentration. The protein solution got noticeably milky in appearance between 2.5-2M urea concentration so the refolding process was stopped and the mixture was clarified by centrifugation. Analysis of the supernatant and the resulting pellet showed that most of the SpCyclin B3 molecules had precipitated (data not shown). The process of dialysis exposed the SpCyclin B3 protein to a decreasing gradient of

denaturant over several hours at 4°C. This meant that the mixture of protein molecules remained exposed for an extended period of time to an intermediate denaturant concentration (2 to 4M urea) where the SpCyclin B3 molecules were not yet folded but no longer denatured (i.e. folding intermediates) and thus extremely prone to aggregation (London *et al.*, 1974). In a subsequent refolding experiment, the aim was to determine a denaturant concentration at which the majority of the SpCyclin B3 molecules were correctly folded but less prone to aggregation because of the presence of mild denaturant. Experiments with lysozyme had shown that the atomic structure of a protein crystallised at several concentrations of urea-containing buffer ((0.7, 2, 3, 4, 5 and 9M) were identical to the native structure (Pike and Acharya, 1994). IBs containing SpCyclin B3 molecules were prepared and denatured as before. The sample was split into two halves and sample I was dialysed till it got noticeably milky at bout 2.5M urea buffer. The dialysis of sample II was halted at 3M urea concentration and the protein mixture was spun at 40000rpm for 20 minutes to remove any aggregation nuclei. The supernatant was concentrated and an aliquot was analysed by SDS-PAGE. This revealed an absence of a band corresponding to the SpCyclin B3 molecule. A thin spatula was used to scrap some material from the centriprep-10 membrane and this sample was analysed on SDS-PAGE. It showed that most of the SpCyclin B3 protein had precipitated during the process of concentration at 4°C. The dialysis method was abandoned for the immobilised metal affinity chromatography (IMAC) method.

In two separate experiments (see section 5.2.9), SpCyclin B3 was refolded on the Ni-NTA column using a stepwise elution protocol (Holzinger *et al.*, 1996) and gradual elution (Rogl *et al.*, 1998; Shi *et al.*, 1997; Stempfer *et al.*, 1996). The latter method resulted in the precipitation of the protein on the column but the former yielded a soluble protein that migrated on SDS-PAGE at about 30KDa and it eluted from a superdex-75 (Pharmacia) gel filtration column at about 30KDa (data not shown). This band was submitted for N-terminal sequencing and it was shown to be a host-encoded protein called WHP (Wulfing *et al.*, 1994). WHP is a two-domain protein that is extremely rich in metal-binding amino acids such as histidines, cysteines and acidic amino acids. It was noted as a persistent contaminant during IMAC purification of target proteins from *E.*

coli. It migrates on SDS-PAGE as a 27KDa protein and it bound to Ni-NTA in the presence of 6M GdnHCl (Wulfing *et al.*, 1994).

Finally, the iterative method was employed to refold SpCyclin B3. The result was similar to the gradient IMAC method in that at the end of the refolding process, the eluate did not contain any SpCyclin B3; almost all of the latter had precipitated on the column (fig. 5.8).

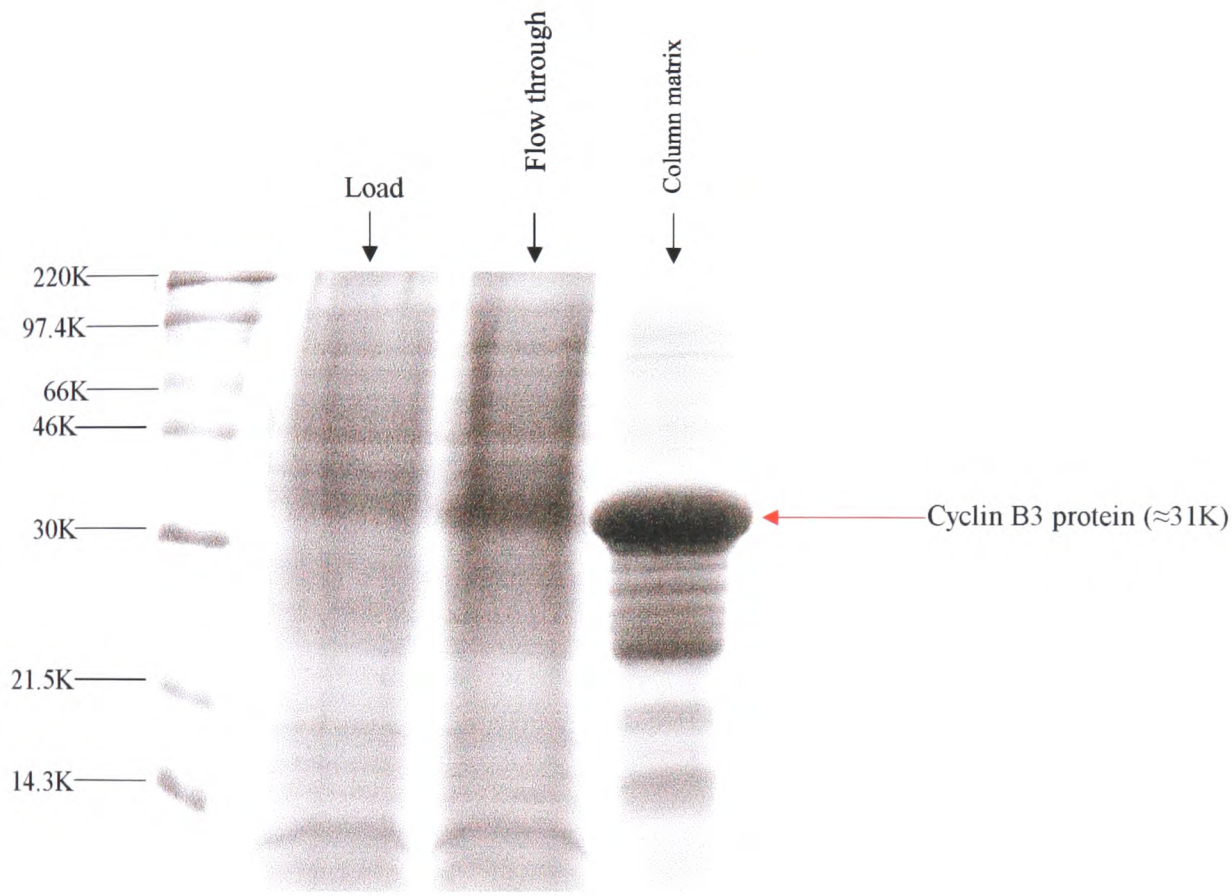


Figure 5.8 Refolding of SpCyclin B3 protein using the iterative method

At the end of the refolding process, the eluate as judged by the chromatographic profile (data not shown) contained a negligible amount of protein. The eluted fractions were concentrated using strataclean beads (Stratagene) and analysed on SDS-PAGE (data not shown) and this confirmed the absence of SpCyclin B3 molecules. The Ni-NTA agarose column was dismantled and a spatula-tip amount of the column matrix was fetched and analysed on SDS-PAGE. This showed that almost all the SpCyclin B3 protein had precipitated on the column either during or at then end of the refolding process.

5.3.4 A homology model of SpCyclin B3

Through primary amino acid sequence, mutational analysis and X-ray structural studies, a conserved region of about 100 amino acid residues called the cyclin box has been shown to be responsible for the binding of cyclins to CDKs (Noble *et al.*, 1997). Since both BvCyclin A3 and SpCyclin B3 were designed to retain the cyclin box fold (where all cyclins share considerable sequence similarity), it was deemed appropriate to use the crystal structure of BvCyclin A3 as a platform to generate a crude homology model of SpCyclin B3 with a view to the identification of any extensive hydrophobic patches on the surface of SpCyclin B3 that may be pre-disposing it to aggregation.

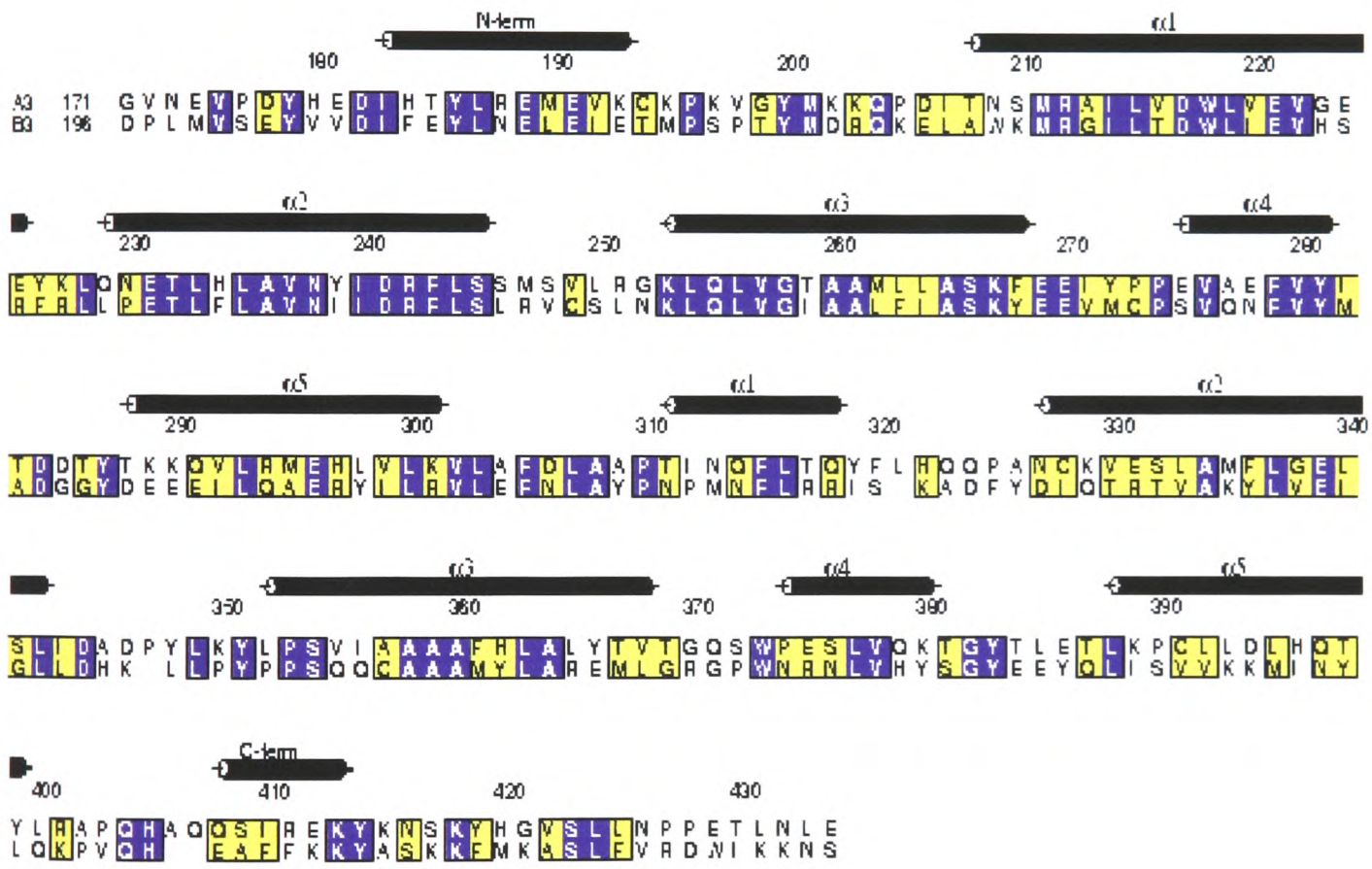


Figure 5.9 Aligned amino acid sequence of BvCyclin A3 and SpCyclin B3

The primary sequences were aligned with the program GAP in the GCG software suite (Senger et al., 1998) and this gave a percentage similarity and identity of 54% and 35%, respectively. The aligned sequences were displayed with ALSCRIPT (Barton, 1993). Conserved residues are boxed blue and non-conserved residues are in yellow. The secondary structure elements experimentally determined for BvCyclin A3 (Brown et al., 1995) are represented above the alignment as helices.

Examination of the aligned sequences showed that there was a preponderance of non-polar residues in the non-conserved regions of SpCyclin B3. To generate a 3-D model of SpCyclin B3, each non-conserved residue in BvCyclin A3 was mutated to the aligned equivalent in SpCyclin B3 using the "mutate_replace" command in the graphics program O (Jones, 1991). The model was subjected to minimisation in XPLOR and subsequent examination did not show any bad contacts in the interior of the protein.

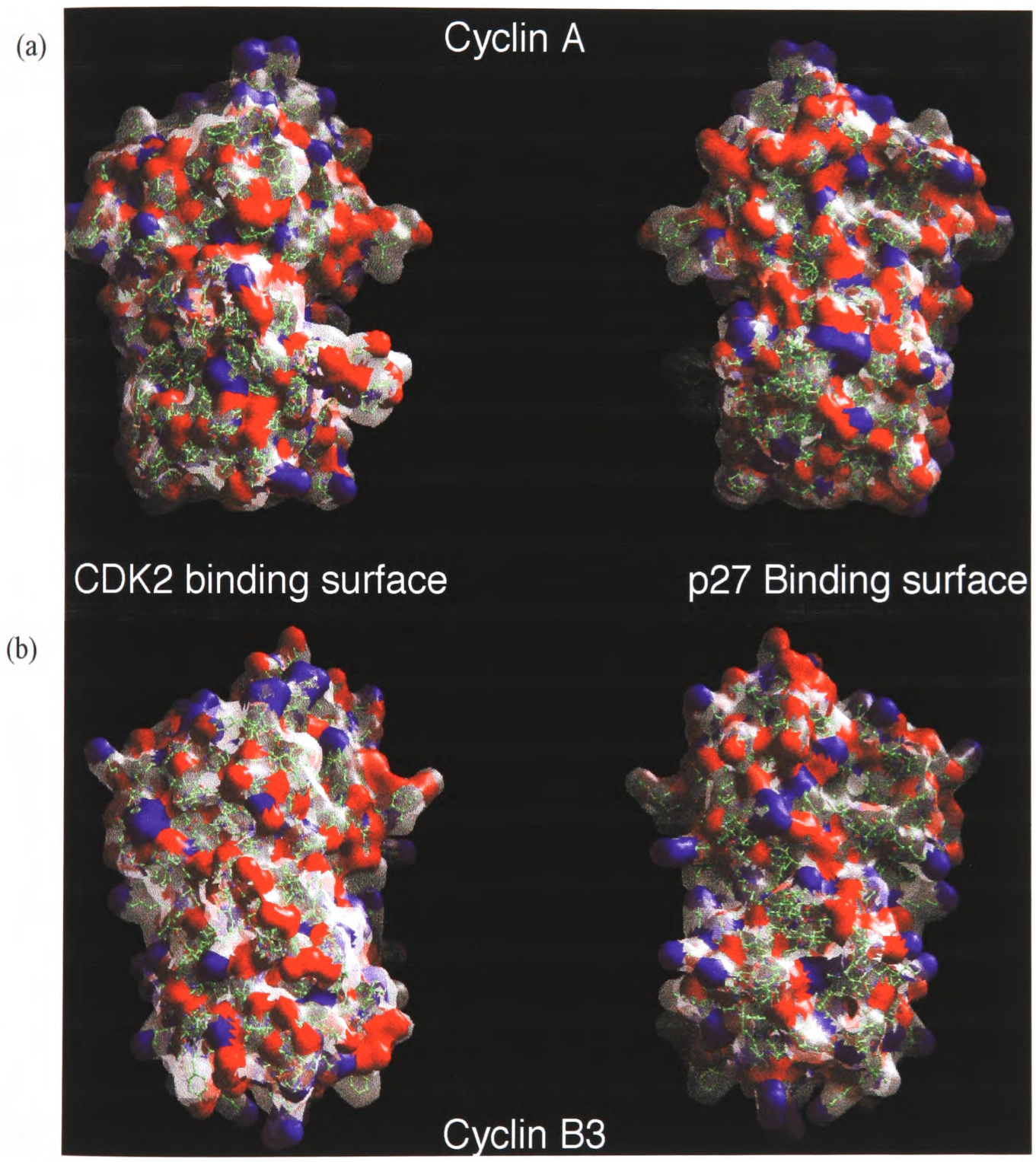


Figure 5.10 Surface properties of BvCyclin A3 X-ray structure and SpCyclin B3 model

Surface hydrophobicity was calculated with the program GRASP (Nicholls *et al.*, 1993) *and each molecule was displayed with the program XOBJECTS (Dr. Noble, unpublished work). In this representation, oxygen atoms are red, nitrogen atoms are blue and all others are grey. A semi-transparent surface envelope has been built onto the secondary structure elements of each molecule.*

A comparative examination of the surface properties of the BvCyclin A3 X-ray model and SpCyclin B3 homology model showed that the surface properties were similar overall and also at both the CDK2 and Cip/Kip binding surfaces. In the absence of an extensive region of hydrophobicity on SpCyclin B3 that could account for the propensity of this molecule to aggregate, the research focus was redirected to finding an *in vivo* mechanism to produce soluble active SpCyclin B3 with bacterial cells. The final strategy in this quest involved the co-expression of SpCyclin B3 with its cognate CDK partners in *E. coli*.

5.3.5 DNA manipulations and construction of pLNJ1 vectors

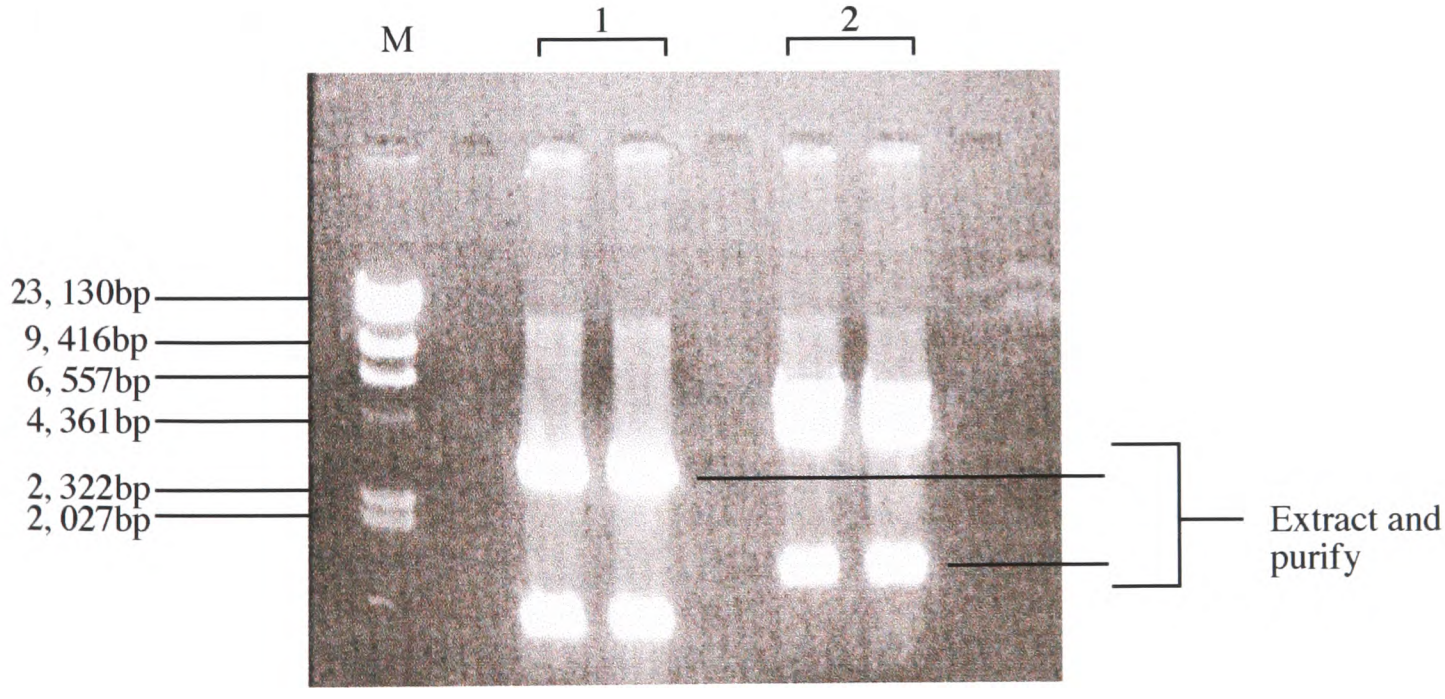


Figure 5.11 Restriction digest of pACYC177 and pET28a vectors

Each sample of pure plasmid DNA was double-digested with restriction endonucleases and ran on a 1% agarose gel in TAE buffer. M = Lambda-Hind III digested size standards. 1 = a duplicate ran of pACYC177 vector cut with Bam HI-Sma I. The sizes of

the two fragments are 1094bp and 2846bp. 2 = a duplicate ran of pET28a vector cut with Bgl II-Sma I. The sizes of the three fragments are 1470bp, 3899bp and 5369bp. The latter fragment is due to incomplete digestion by one of the restriction enzymes.

The purified vector (2846bp) and insert (1470bp) were quantitated by reading the OD_{260} value of a 1:10 diluted stock in sterile distilled water. The purity was assessed by calculating the OD_{260}/OD_{280} ratio. The yield of the vector was 40.35µg/ml in a volume of 200µl with an OD_{260}/OD_{280} ratio of 1.71. That of the insert was 33.71µg/ml in a volume of 200µl with an OD_{260}/OD_{280} ratio of 2.02, suggesting that both the vector and insert were pure enough for the subsequent ligation reaction.

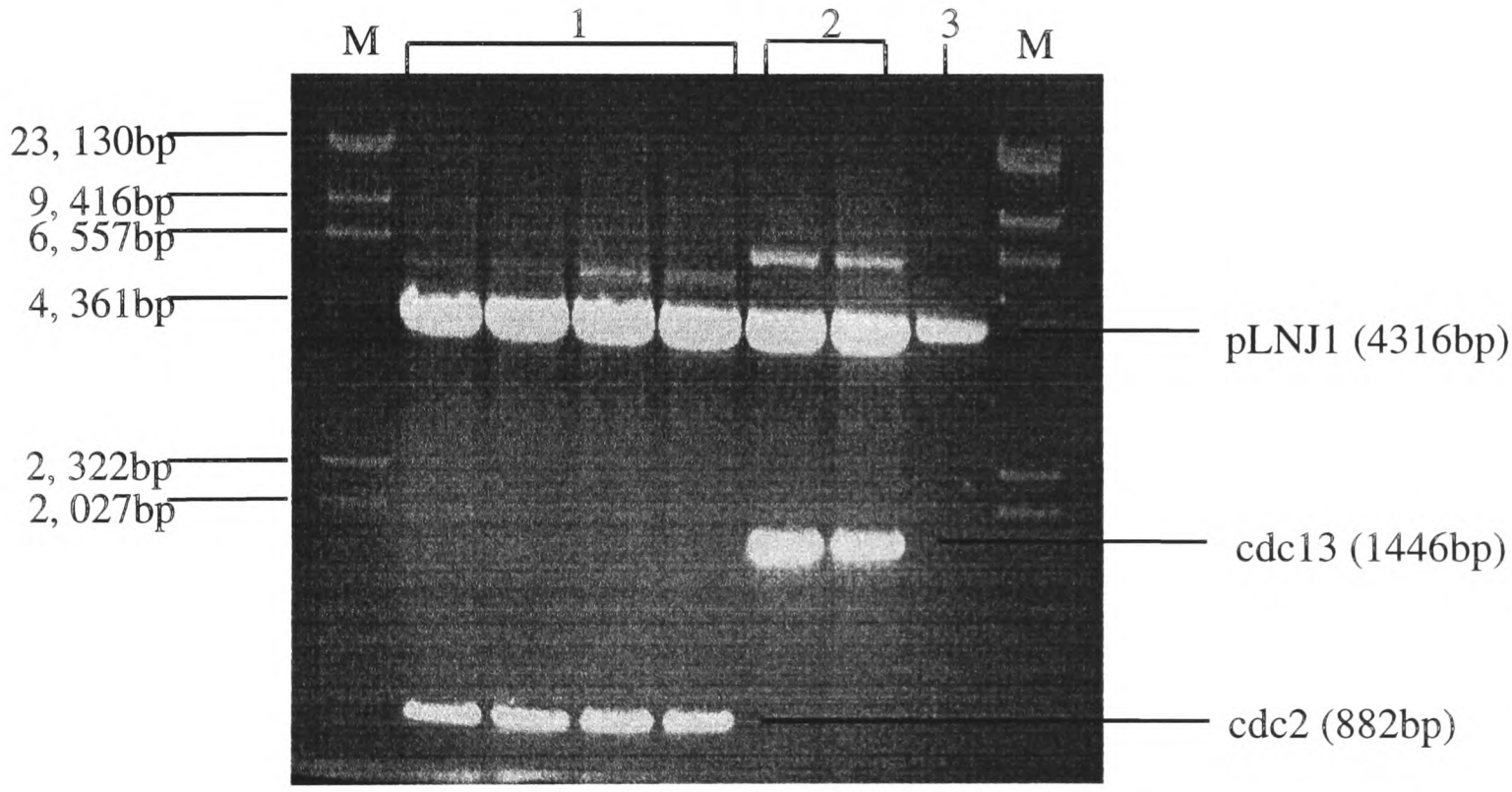


Figure 5.12 Restriction analysis of pLNJ1/CDK1, pLNJ1/CYCLIN B and pLNJ1 vectors

M = Lambda-Hind III digested size standards. 1 = pLNJ1/CDK1 plasmid DNA (isolated from four separate colonies) double-digested with Nde I and Bam HI. 2 = pLNJ1/Cyclin B plasmid DNA (isolated from two separate colonies) double-digested with Nde I and Bam HI. 3 = pLNJ1 vector double-digested with Nde I and Bam HI. The third fragment in lanes 1 and 2 is the result of incomplete digestion by one of the restriction enzymes.

5.3.6 Recombinant protein expression and SDS-PAGE analyses

After demonstrating that the recombinant clones contained the appropriate inserts, it was necessary to confirm that cells harbouring pLNJ1/SpCDK1 plasmid were expressing SpCDK1 protein under conditions established for pRK172/SpCDK1. To this end, transformants containing pRK172/SpCDK1 only and pLNJ1/SpCDK1 only were grown and checked for protein expression. The results (fig. 5.13) showed that the pattern of expression of SpCDK1 from both pRK172 and pLNJ1 expression vectors were similar. However, an examination of the fractions in fig 5.13 show that cells containing pRK172/SpCDK1 produced slightly more SpCDK1 protein than those containing pLNJ1/SpCDK1.

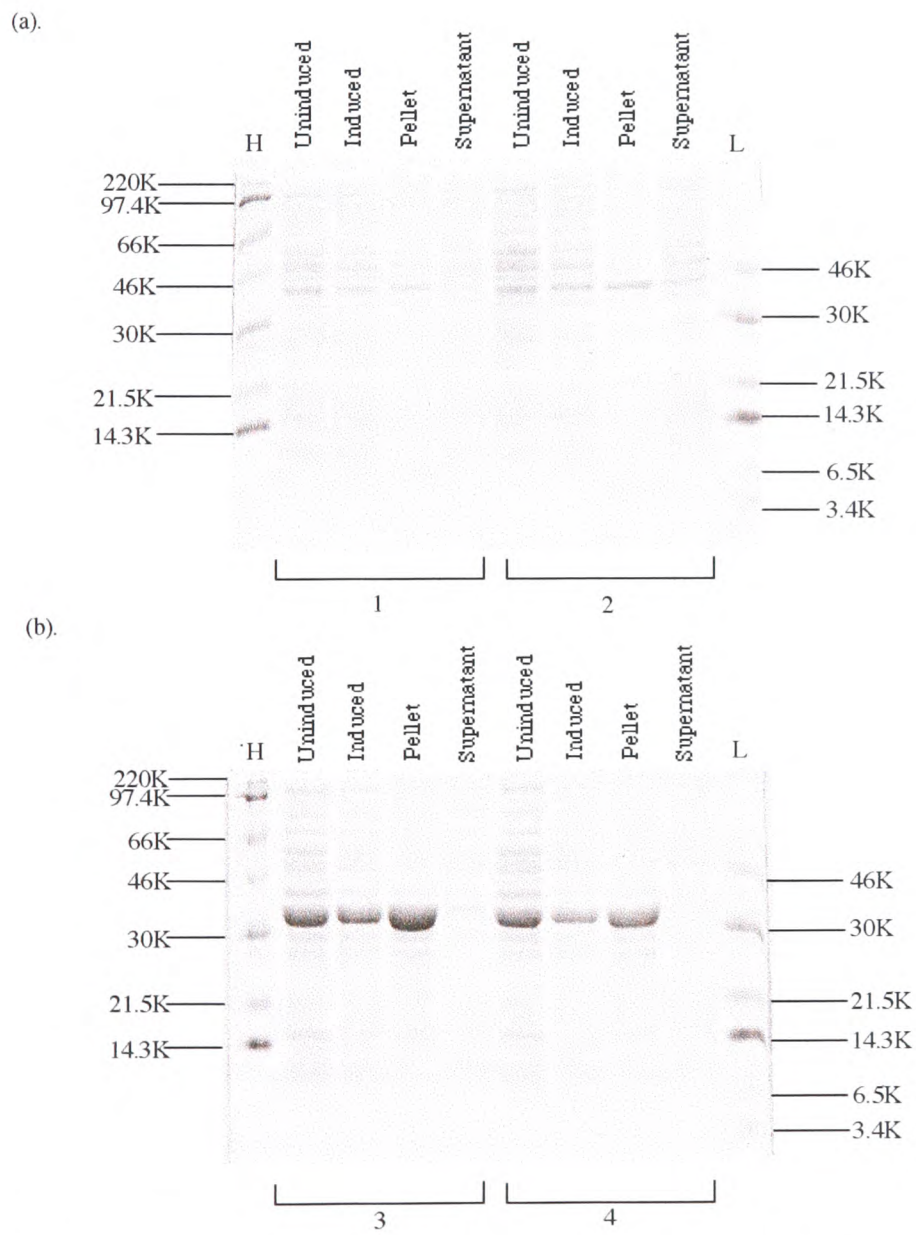


Figure 5.13 Expression of SpCDK1 *(a) SDS-PAGE analysis of untransformed B834(DE3) (1) and cells transformed with pLNJ1 (2). (b) SDS-PAGE analysis of B834(DE3) cells transformed with pRK172/CDK1 (3) and pLNJ1/CDK1 (4). H and L refer to high and low molecular weight standards, respectively. The cells were grown in LB medium and protein expression was induced by the addition of IPTG (0.4mM). Protein*

expression was continued for 4 hours before the samples were harvested and run on 10% Tris-tricine gels

In both cases, the SpCDK1 protein was present exclusively in the pellet. A common feature of all the protein expression experiments was the level of recombinant protein in the uninduced fraction (figs 5.13-5.16). This is due to laxity in the regulation of basal transcription. The use of the T7 promoter system in conjunction with expression host strains such as B834(DE3), B834(DE3)pLysS and B834(DE3)pLysE provide varying degrees of regulation of basal transcription (described in chapter 4 section 4.2.14). B834(DE3)pLysS and B834(DE3)pLysE cells harbour a chloroamphenicol-resistant plasmid carrying the T7 lysozyme gene. T7 lysozyme binds to and inhibits T7 RNA polymerase. pLysE cells provide the tightest control of basal transcription because they produce the highest amount of T7 lysozyme (Studier *et al.*, 1990). pLysS cells provide a less tight control and non-pLysS cells provide minimal control hence the level of basal transcription observed in almost all the uninduced cultures. The need to maintain two plasmids with compatible origins precluded the use of either pLysS or pLysS cells for the co-expression experiments.

A standard recommended concentration of 0.4mM IPTG (Studier *et al.*, 1990) was used in all the protein expression experiments. Sometimes, it was necessary to do an IPTG titration to determine an optimal (with respect to solubility and yield) concentration for the particular target gene.

5.3.7 Expression of pLNJ1/Cyclin B plus pRK172/SpCDK1

Co-expression of full-length Cyclin B and SpCDK1 was conducted at three different temperatures i.e. overnight at 18°C, four hours at 25°C and 37°C. For the experiments conducted at 18°C and 37°C, the effect of two different IPTG concentrations (0.4mM and 1.0mM) were also examined. There was minimal basal transcription of target genes when the two proteins were co-expressed at 18°C (fig. 5.14a).

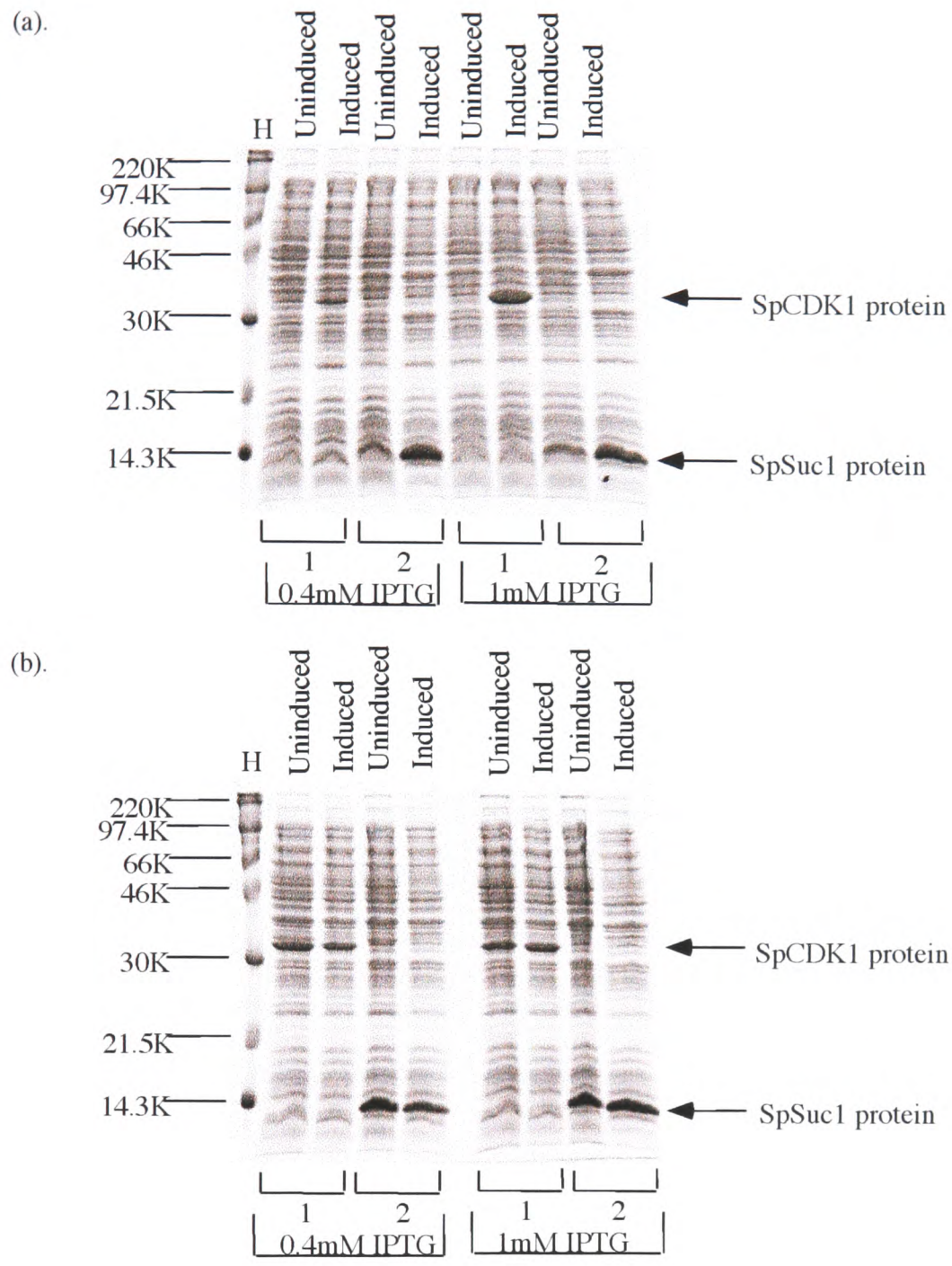


Figure 5.14 Co-expression of pRK172/CDK1 & pLNJ1/CYCLIN B (1) and pSJW1/Suc1 & pLNJ1/CDK1 (2) in B834(DE3) cells *(a) The cells were grown at 37˚C to mid-log phase and protein expression was induced at 18˚C for 16 hours. (b) Cell growth and protein expression was conducted at 37˚C throughout. Protein expression was continued for fours. All the samples were ran on 10% Tris-tricine gels.*

Induction of protein expression at 18°C with 1mM IPTG produced relatively more SpCDK1 protein than those induced with 0.4mM IPTG (fig. 5.14a). In general, it was observed that the greater the temperature the higher the level of basal transcription. A protein band corresponding to full-length Cyclin B (56KDa) was not detected on all the coomassie-stained SDS-PAGE gels. This was not pursued further. Analysis of the soluble and insoluble fractions showed that in all cases, the SpCDK1 protein was present predominantly in the pellet (data not shown).

5.3.8 Expression of pLNJ1/SpCDK1 plus pSJW1/SpSuc1

Induction of protein expression with 0.4mM and 1mM IPTG at 18°C produced comparable amounts of SpSuc1 protein with minimal basal transcription (fig. 5.14a). Again, the level of basal transcription seems to increase with increasing temperature. In this experiment, SpSuc1 was preferentially overexpressed over SpCDK1, which could not be seen at all on the coomassie-stained SDS-PAGE gel. The SpSuc1 protein was, as expected, almost entirely soluble (data not shown).

5.3.9 Expression of pLNJ1/SpCDK1 plus pET21d/Cyclin A3.

As mentioned in Chapter 4, bovine Cyclin A3 is readily overexpressed in *E. coli* in a mainly soluble and active form (Brown *et al.*, 1995). It is well-established that CDK1 forms an active complex with Cyclin A, but whether phosphorylation of CDK1 on T167 is an absolute requirement to form a high affinity complex with Cyclin A or not, remains unresolved (Desai *et al.*, 1995; Heitz *et al.*, 1997). Nevertheless, an attempt was made to produce SpCDK1 and BvCyclin A3 with the notion that a pool of free, soluble BvCyclin A3 in the cytoplasm would be able to bind to the nascent SpCDK1 polypeptides and form a soluble complex *in vivo*. Induction of protein expression at 18°C with 0.4mM IPTG led to the overproduction of bovine Cyclin A3 and SpCDK1 (fig. 5.15a). The yield of the latter was about a third of that of the former.

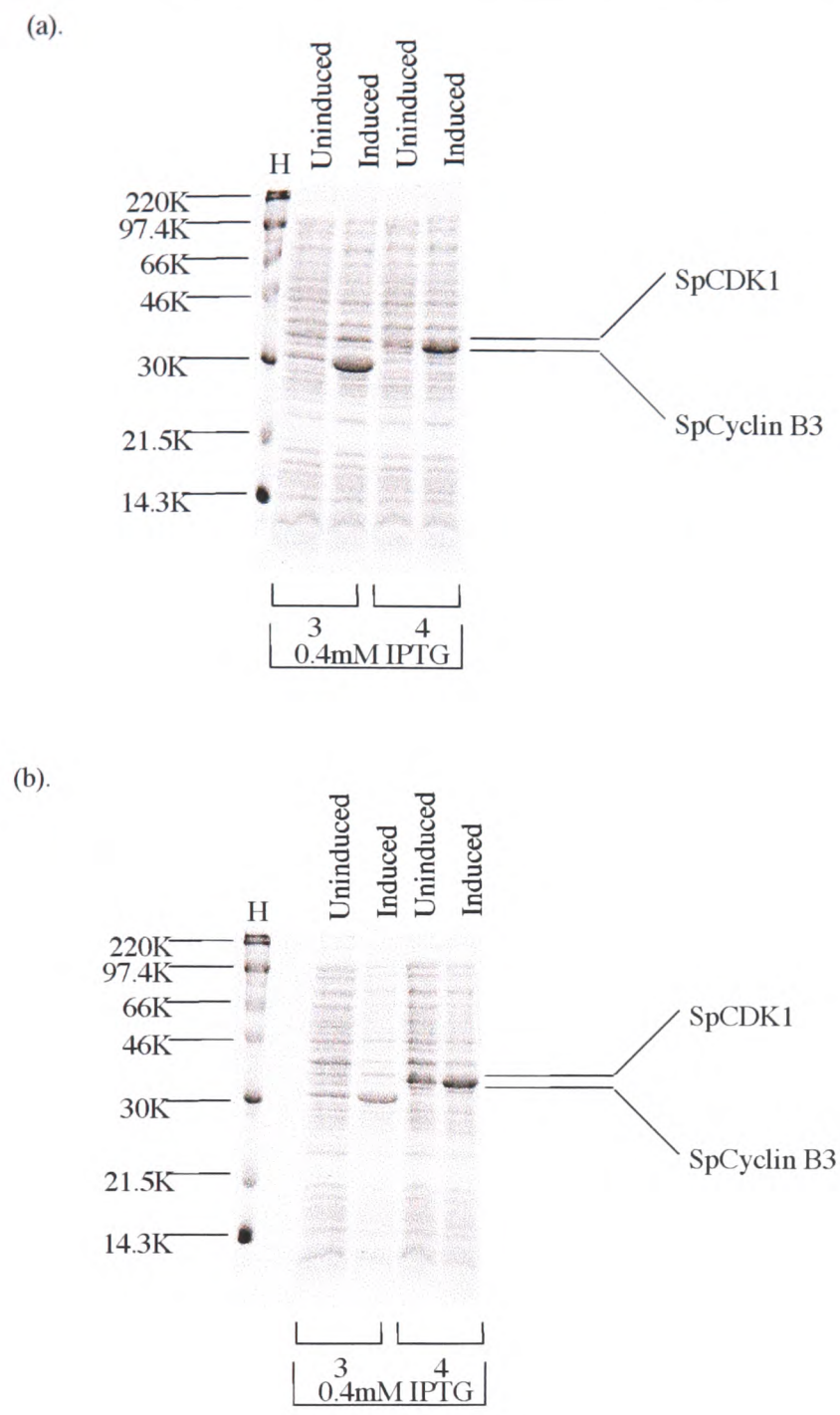


Figure 5.15 Co-expression of pET21d/Cyclin A3 & pLNJ1/CDK1 (3) and pET21d/Cyclin B3 & pLNJ1/CDK1 (4) in B834(DE3) cells

(a) The cells were grown at 37˚C to mid-log phase but protein expression was induced at 18˚C for 16 hours. (b) Cell growth and protein expression was conducted at 37˚C. Protein expression was continued for hours. All the samples were ran on 10% Tris-tricine gels.

Also, the amount of both proteins decreased when the experiment was conducted at higher (25˚C and 37˚C) temperatures [figs 5.15(b) and 5.16]. In all cases, accumulation of BvCyclin A3 exceeded that of SpCDK1. About 60% of the BvCyclin A3 protein was found in the soluble fraction. The rest, with almost all of the overexpressed SpCDK1 protein was present in the pellet.

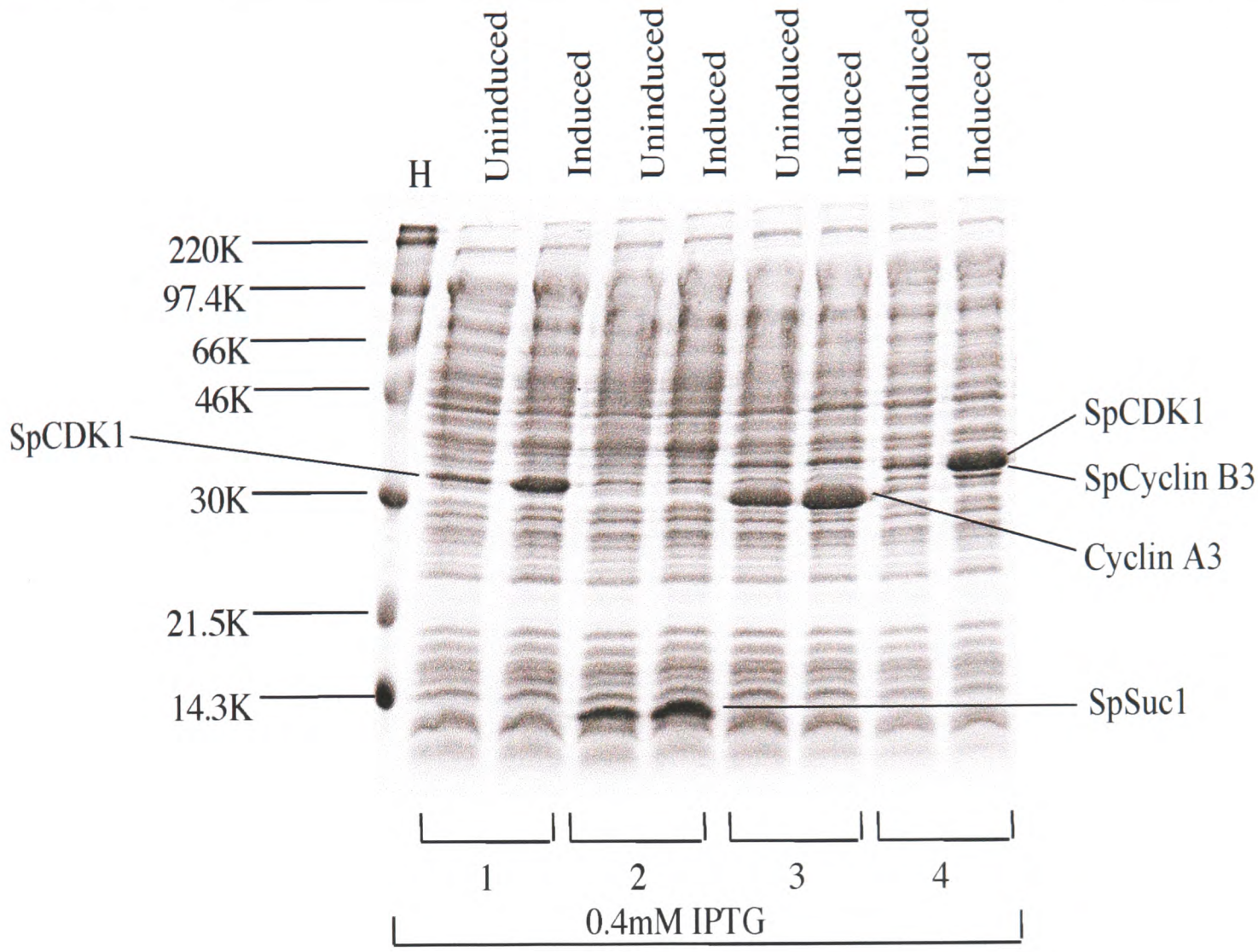


Figure 5.16 Co-expression of cell cycle regulatory proteins at 25°C

1 = pRK172/CDK1 + pLNJ1/Cyclin B

2 = pSJW1/Suc1 + pLNJ1/CDK1

3 = pET21d/Cyclin A3 + pLNJ1/CDK1

4 = pET21d/Cyclin B3+ pLNJ1/CDK1

The cells were grown with shaking at 250rpm to mid-log phase at 37°C before incubation at 25°C for 60 minutes. IPTG (0.4mM) was added and protein expression was allowed to proceed for 4 hours.

5.3.10 Expression of pLNJ1/SpCDK1 plus pET21d/Cyclin B3.

In contrast to the above, individual expression of either SpCDK1 or Cyclin B3 resulted in high-level production of insoluble aggregates. It was envisaged that simultaneous translation of the two mRNA transcripts in the same bacteria cell, would

enable the two newly formed polypeptides to bind and form a soluble and active protein-protein complex. Although both proteins were overexpressed, they were both predominantly found in the pellet.

Co-expressed constructs	Overexpressed?	Soluble or insoluble
pLNJ1/Cyclin B + pRK172/ SpCDK1	SpCDK1 was overexpressed. Cyclin B was not.	SpCDK1 was ~100% insoluble
pSJW1/SpSuc1 + pLNJ1/SpCDK1	SpSuc1 was overexpressed. SpCDK1was not.	SpSuc1 was ~100% soluble
pET21d/Cyclin A3 + pLNJ1/SpCDK1	Yes, both.	Cyclin A3 was ~50% soluble but SpCDK1 was ~100% insoluble
pET21d/Cyclin B3 + pLNJ1/SpCDK1	Yes, both	Both were ~100% insoluble.

Table 5.2 Summary of co-expression results

5.3.11 Affinity Purification of soluble SpCDK1-Cyclin complexes

An examination of the soluble and insoluble fractions on the SDS-PAGE gels showed that most of the SpCDK1 protein was present in inclusion bodies. Nevertheless, an attempt was made to purify soluble SpCDK1-Cyclin A3 complexes because "even when inclusion bodies are formed, some proportion of the target protein is usually soluble within the cell. With the high expression levels of the pET system and the power of the his-tag purification method, even when most of the target protein mass is in aggregates there may be a significant amount of soluble material as well" (Novagen, 1999). BvCyclin A3 and SpCyclin B3 recombinant proteins had C-terminal hexahistidine tags whilst the SpCDK1 protein had an N-terminal tag. The resulting CDK-Cyclin complex would possess two hexahistidine tags which if available for binding, would result in an enhanced affinity for the cobalt-chelated matrix. This has two main advantages: first, even if the level of the soluble fraction is minimal, the induced culture can be scaled up in volume and the affinity purification step can be used to purify and concentrate the soluble fraction. Second, an enhanced binding would result in a delayed elution of the bound protein which would, in turn produce relatively purer eluted fractions. The starting

bacterial culture was scaled to 6 litres and the supernatant was applied to a 1ml HiTrap™ (Pharmacia) metal chelating column charged with Co^{2+} ions.

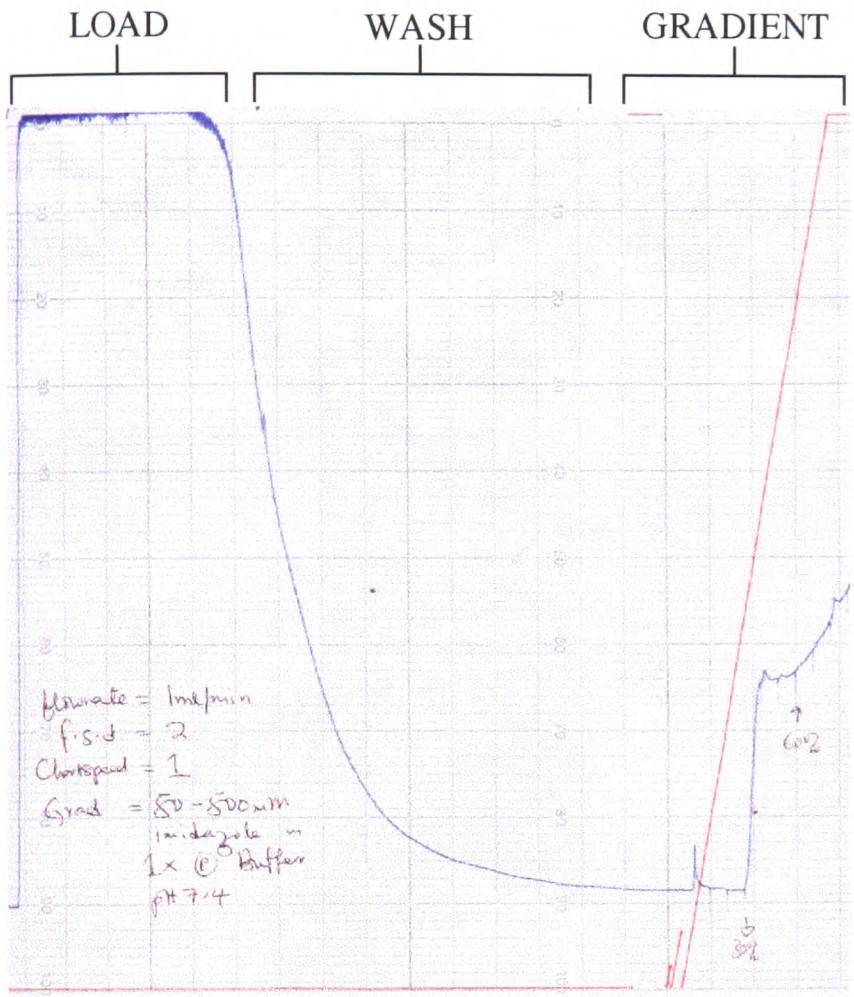
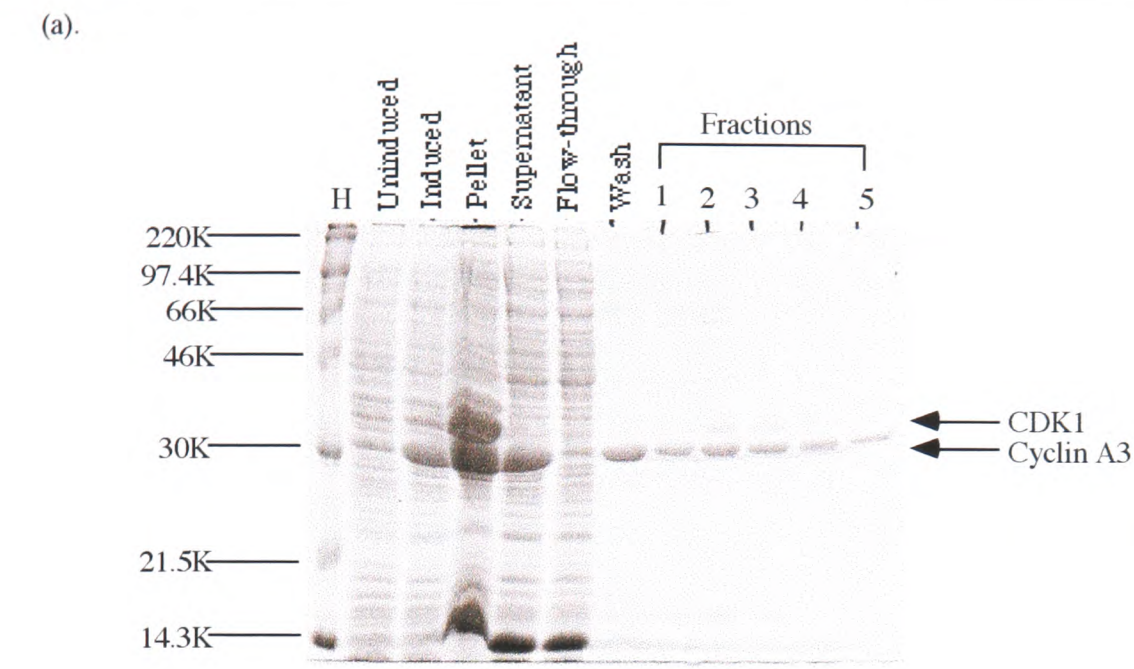


Figure 5.17 Chromatogram depicting the purification of his-tagged CDK1-Cyclin A3

The clarified protein mixture containing SpCDK1-BvCyclin A3 was applied to a Hitrap metal chelating column charged with a 0.1M solution of Cobalt chloride. The charged column had a binding capacity of 5mg of a hexahistidine-tagged protein per ml of resin. The column was washed to near-baseline before the bound proteins were eluted with a linear gradient of 50 - 500mM imidazole in 1X-phosphate (0.02M Na_2HPO_4, 0.5M NaCl pH 7.2) buffer.

Aliquots of the eluates were analysed on SDS-PAGE gels (fig. 5.18).



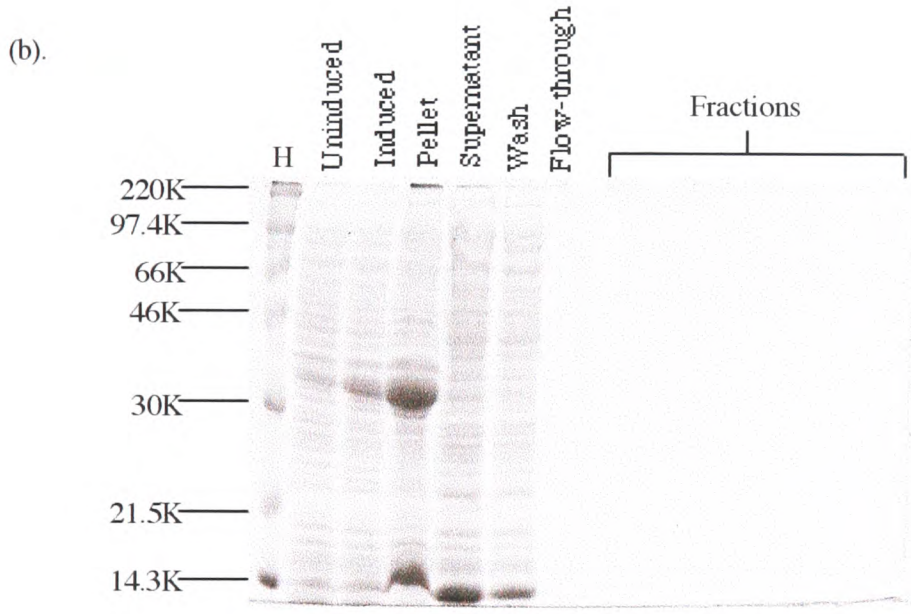


Figure 5.18 Affinity purification of CDK-Cyclin complexes and SDS-PAGE analyses

(a) Most of the CDK1-Cyclin A3 complex was present in the insoluble fraction. Fractions 2-4 contain two bands corresponding to CDK1 (~34KDa) and Cyclin A3 (~30KDa). These fractions were pooled, concentrated and passed onto a Superdex-75 gel filtration column (Pharmacia). The single peak on the resulting chromatogram contained only Cyclin A3 (data not shown). (b) Although both proteins were overexpressed, there was no trace of either SpCDK1 or SpCyclin B3 in the collected fractions ran on a coomassie-stained, 10% Tris-tricine gel.

5.3.12 Co-denaturation and refolding of SpCDK1 and Cyclin B3

Since both SpCDK1 and SpCyclin B3 were being directed into IBs even in the co-expression system, an attempt was made to isolate, denature and reconstitute an active complex *in vitro*.

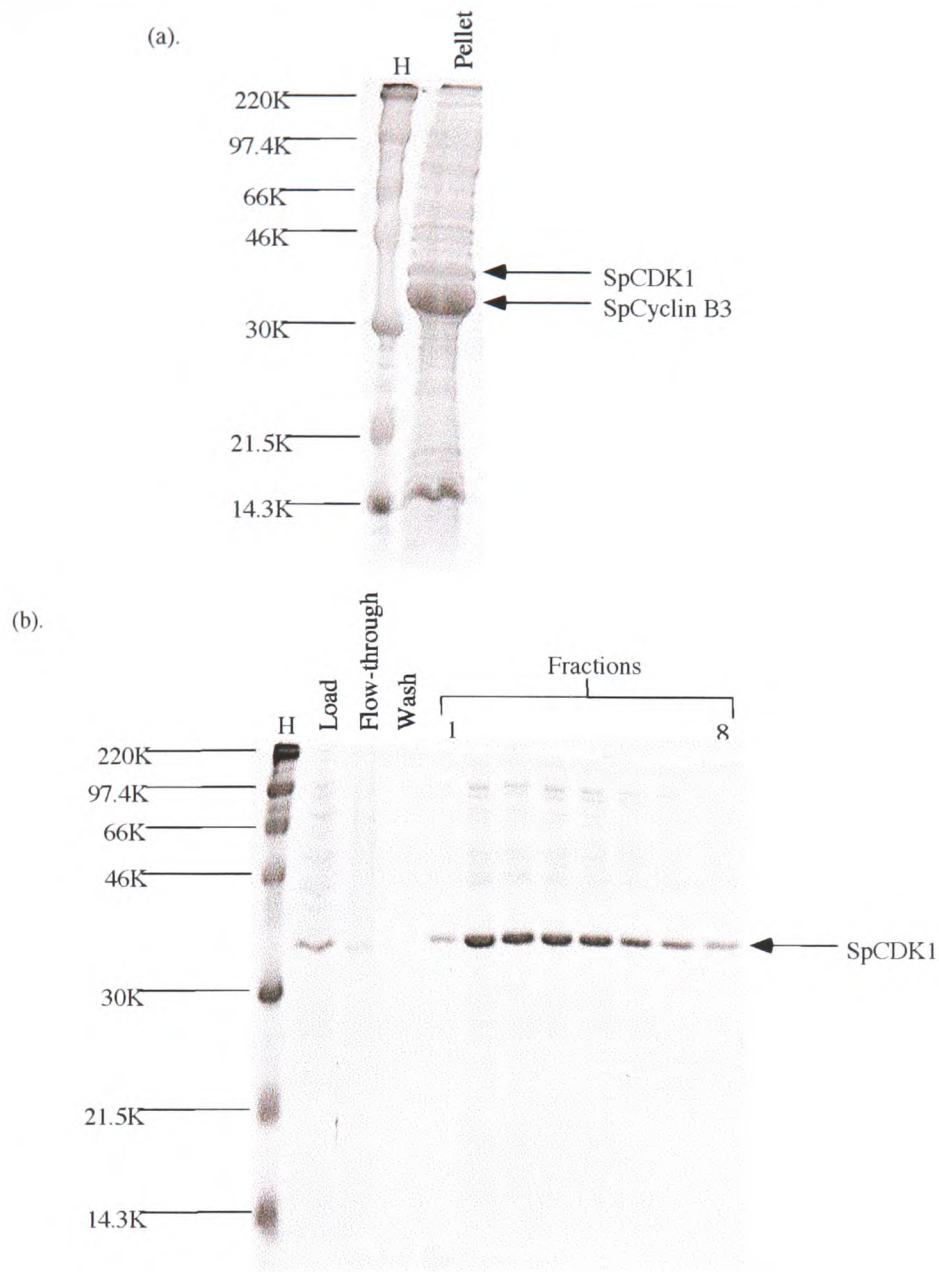


Figure 5.19 Refolding of CDK1 and Cyclin B3 from inclusion bodies

(a) The insoluble fraction from which the inclusion bodies were isolated and denatured in 8M urea. (b) SDS-PAGE analysis of the purification procedure. The band corresponding to SpCyclin B3 is not present in the Load fraction suggesting that it may have precipitated in the course of the refolding process. The amount of protein in each eluted fraction (4ml) was so low (< 1ug/ml) that each fraction had to be concentrated using strataclean resin (Stratagene) before the electrophoresis step. The identity of the putative SpCDK1 band was confirmed by N-terminal protein sequencing (data not shown).

5.3.13(a) Theoretical models of SpCDK1 and SpCyclin B

In the absence of experimentally determined structural (NMR or X-ray) models for SpCDK1 and SpCyclin B, the full-length sequences of these proteins were submitted to the Glaxo-Wellcome web-based modelling server for the generation of structural models (Peitsch, 1996; Peitsch *et al.*, 1995). The SpCDK1 co-ordinates were generated using the following co-ordinates from the protein database (PDB): 1fin.pdb (HsCDK2-HsCyclin A3 complex, 2.3Å), 1jst.pdb (phosphorylated HsCDK2-HsCyclin A3, 2.6Å), 1hck.pdb (HsCDK2 1.9Å), 1b38.pdb (HsCDK2 2.0Å) and 1buk.pdb (HsCDK2-CksHs1

complex, 2.6Å). The SpCDK1 theoretical model contained residues 1-297 whilst that of SpCyclin B consisted only residues 198-450. The latter model was based on the following co-ordinates: 1fin.pdb, 1jst.pdb and 1vin.pdb (BvCyclin A3, 2.0Å).

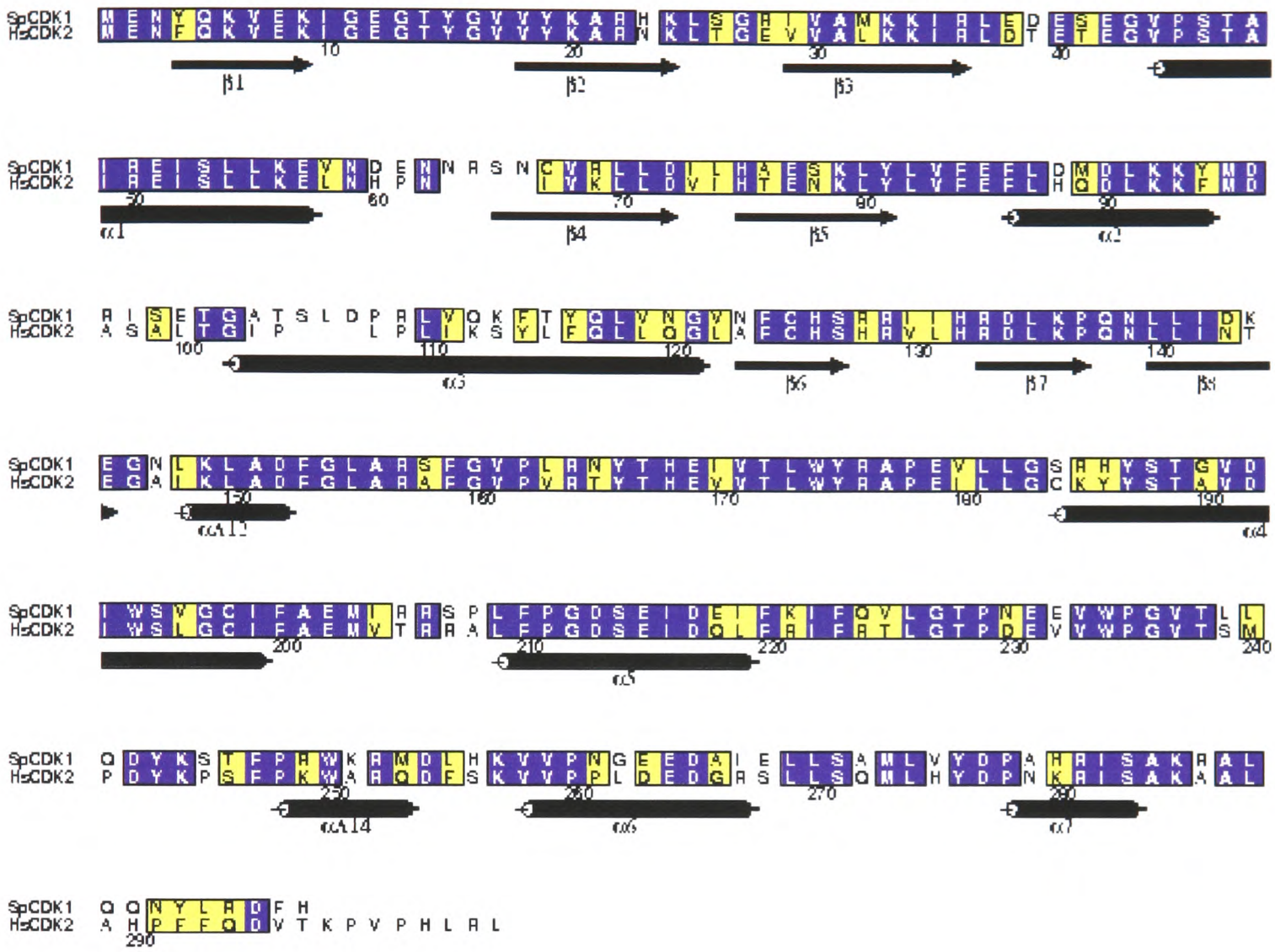


Figure 5.20 Primary amino acid sequence alignment of SpCDK1 and HsCDK2

Identical and similar residues are boxed in blue and yellow, respectively. The sequences were aligned with Clustal W (Higgins et al., 1992) and displayed with ALSCRIPT (Barton, 1993). Experimentally determined secondary structural elements (arrows for β-sheet and barrels for α-helices) for HsCDK2 are depicted below the aligned sequences. The two proteins share about 65% sequence similarity.

5.3.13(b) Interactions at the CDK-cyclin dimer interface

To generate an SpCDK1-SpCyclin B complex model, the SpCDK1 model was superimposed onto the experimentally determined (X-ray) HsCDK2 model and the

SpCyclin B model was similarly superimposed onto the X-ray structure of HsCyclin A3. The co-ordinates were written out, edited and fed into a program called CONTACTS (CCP4, 1994) in order to determine the residues involved in mediating protein-protein interactions.

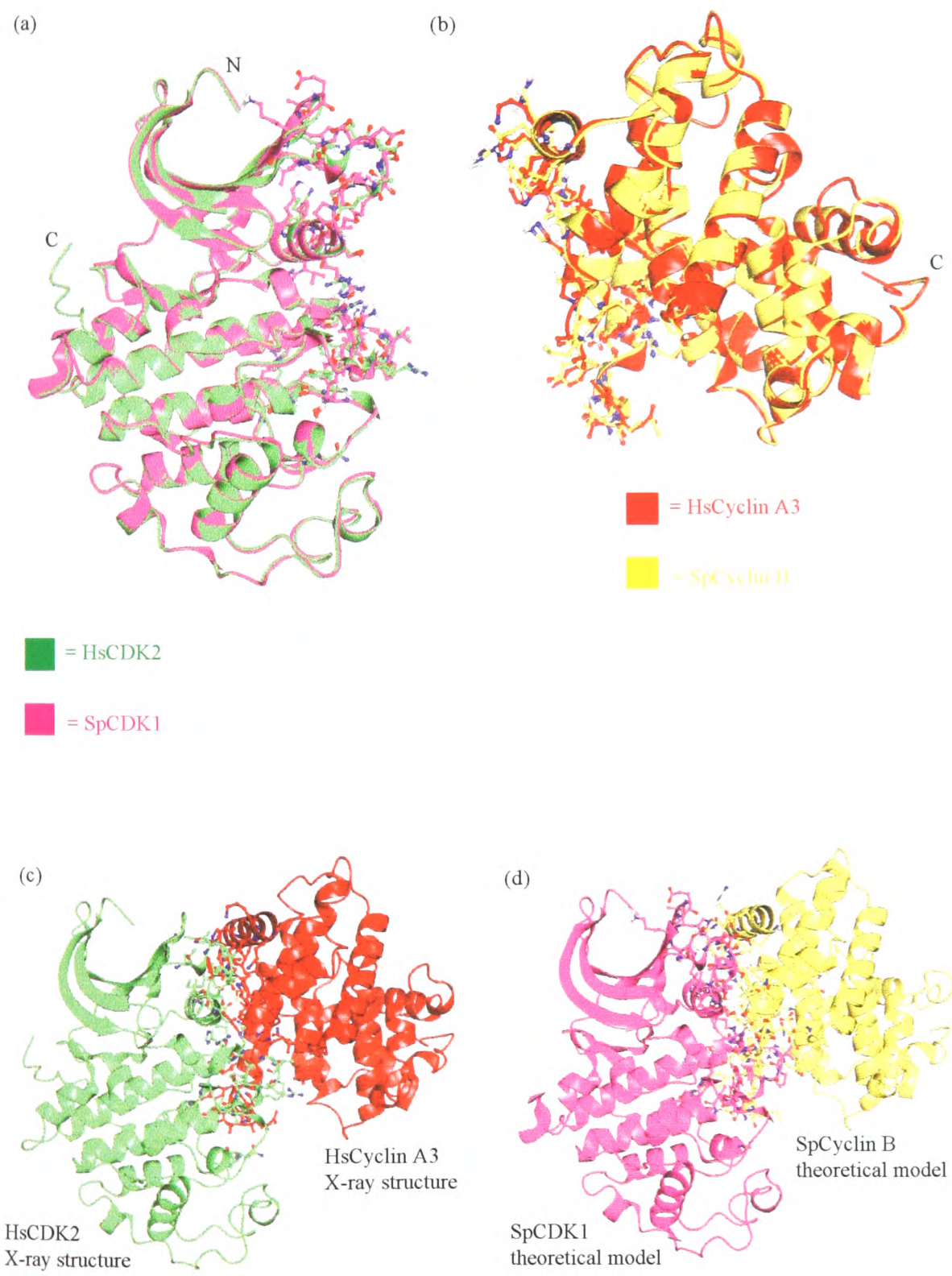


Figure 5.20 Formation of CDK-cyclin complexes

HsCDK2 contact residue (from X-ray structure)	**HsCyclin A3 contact residue (from X-ray structure)**		**SpCDK1 contact residue (from theoretical model)**	**SpCyclin B contact residue (from theoretical model)**
Leu37→ (conserved)	←His296 (non-conserved, Arg321 in SpCyclin B)		Leu37→ (conserved)	←Arg321 (non-conserved, His296 in HsCyclin A3)
Thr39→ (non-conserved, Asp in SpCDK1)	←Lys289 (non-conserved, Glu314 in SpCyclin B)		Asp39→ (non-conserved, Thr in HsCDK2)	←Glu314(non-conserved, Lys289 in HsCyclin A3) ←Leu317 (conserved)
Glu40→ (conserved)	←Leu292 (conserved) ←Lys288(non-conserved, Glu3143 in SpCyclin B)		Glu40→ (conserved)	←Glu314(non-conserved, Lys289 in HsCyclin A3) ←Glu313(non-conserved, Lys288 in HsCyclin A3)
Thr41→ (non-conserved, Ser in SpCDK1)	←Lys288(non-conserved, Glu3143 in SpCyclin B)		Ser41→ (non-conserved, Thr in SpCDK1)	←Glu313(non-conserved, Lys288 in HsCyclin A3) ←Val300 (conserved)
Glu42→ (conserved)	←Lys266 (conserved)		Glu42→ (conserved)	←Val300 (conserved) ←Lys291 (conserved) ←Phe287(non-conserved, Leu262 in HsCyclin A3)
Gly43→ (conserved)	←Glu295 (conserved)		Gly43→ (conserved)	←Phe287(non-conserved, Leu262 in HsCyclin A3) ←Lys291 (conserved) ←Leu317 (conserved)
Val44→ (conserved)	←Lys266 (conserved) ←Glu295 (conserved) ←Leu299 (conserved)		Val44→ (conserved)	←Lys291 (conserved) ←Phe287(non-conserved, Leu262 in HsCyclin A3) ←Leu317

	←His296(non-conserved, Arg321 in SpCyclin B)			(conserved) ←Glu320 (conserved)
Ser46→ (conserved)	←Lys266 (conserved) ←Pro272(non-conserved, Cys297 in SpCyclin B)		Pro45→ (conserved)	←Lys291 (conserved)
Ile49→ (conserved)	←Leu263(non-conserved, Ile288 in SpCyclin B) ←Lys266 (conserved)		Ser46→ (conserved)	←Lys291 (conserved) ←Cys297(non-conserved, Pro272 in HsCyclin A3)
Arg50→ (conserved)	←Phe267(non-conserved, Tyr292 in SpCyclin B) ←Lys266 (conserved)		Ile49→ (conserved)	←Tyr292(non-conserved, Phe267 in HsCyclin A3) ←Leu331 (conserved) ←Lys291 (conserved)
Ile52→ (conserved)	←Phe304 (conserved)		Arg50→ (conserved)	←Lys291 (conserved) ←Tyr292(non-conserved, Phe267 in HsCyclin A3) ←Glu294 (conserved) ←Val295(non-conserved, Ile270 in HsCyclin A3)
Ser53→ (conserved)	←Phe304 (conserved) ←Leu306 (conserved) ←Phe267(non-conserved, Tyr292 in SpCyclin B)		Ile52→ (conserved)	←Phe329 (conserved)
Lys56→ (conserved)	←Phe304 (conserved) ←Asp305(non-conserved, Asn330 in		Ser53→ (conserved)	←Tyr292(non-conserved, Phe267 in HsCyclin A3) ←Ala332 (conserved)

	SpCyclin B) ←Thr303(non-conserved, Glu228 in SpCyclin B)			←Phe329 (conserved) ←Leu331 (conserved)
Glu57→ (conserved)	←Tyr185 (conserved) ←Met189(non-conserved, Leu214 in SpCyclin B)		Leu54→ (conserved)	←Tyr292(non-conserved, Phe267 in HsCyclin A3)
His71→ (conserved)	←His296(non-conserved, Arg321 in SpCyclin B)		Lys56→ (conserved)	←Glu328(non-conserved, Trp303 in HsCyclin A3) ←Phe329 (conserved)
Glu73→ (conserved)	←His296(non-conserved, Arg321 in SpCyclin B)		Glu57→ (conserved)	←Tyr210 (conserved) ←Asn330(non-conserved, Asp305 in HsCyclin A3)
His119→ (conserved)	←Ile182 (conserved) ←Tyr178 (conserved)		Ile73→ (non-conserved, Val in HsCDK2)	←Phe329 (conserved)
Ser120→ (conserved)	←Asp181 (conserved) ←Ile182 (conserved) ←Tyr178 (conserved)		His75→ (conserved)	←Phe329 (conserved) ←Arg321(non-conserved, His296 in HsCyclin A3) ←Arg325(non-conserved, Lys300 in HsCyclin A3)
His121→ (non-conserved, Arg in SpCDK1	←Tyr185 (conserved)		Ala76→ (non-conserved, Thr in HsCDK2)	←Arg321(non-conserved, His296 in HsCyclin A3)
Arg122→ (conserved)	←Ile182 (conserved) ←Ala307 (conserved)		Glu77→ (conserved)	←Arg321(non-conserved, His296 in HsCyclin A3)
Arg150→ (conserved)	←Glu269 (conserved) ←Glu268 (conserved)		Lys79→ (conserved)	←Arg321(non-conserved, His296 in HsCyclin A3)

	←Phe267(non-conserved, Tyr292 in SpCyclin B) ←Ile270(non-conserved, Val295 in SpCyclin B)			
Phe152→ (conserved)	←Asn173(non-conserved, Leu198 in SpCyclin B) ←Ile182 (conserved)		Leu80→ (conserved)	←Phe329 (conserved)
Gly153→ (conserved)	←Gln313(non-conserved, Asn338 in SpCyclin B)		Asn123→ (non-conserved, Ala in HsCDK2)	←Tyr203 (conserved)
Val154→ (conserved)	←Glu268 (conserved) ←Thr316(non-conserved, Arg341 in SpCyclin B) ←Asn312(non-conserved, Met337 in SpCyclin B)		His126→ (conserved)	←Tyr203 (conserved) ←Ile207 (conserved)
Pro155→ (conserved)	←Asn173(non-conserved, Leu198 in SpCyclin B)		Ser127→ (conserved)	←Tyr203 (conserved) ←Asp206 (conserved) ←Ile207 (conserved)
Val156→ (non-conserved, Ile in SpCDK1)	←An173(non-conserved, Leu198 in SpCyclin B)		Arg128→ (non-conserved, His in HsCDK2)	←Tyr210 (conserved) ←Leu214 (non-conserved, Met190 in HsCyclin A3) ←Ile207 (conserved)
Arg157→ (conserved)	←Ile270(non-conserved, Val295 in SpCyclin B)		Arg129→ (conserved)	←Ile207 (conserved) ←Arg342(non-conserved, Gln317 in HsCyclin A3) ←Arg341(non-conserved, Trp316 in

				HsCyclin A3) ←Leu211 (conserved) ←Phe208 (non-conserved, His183 in HsCyclin A3) ←Tyr210 (conserved)
Tyr159→ (conserved)	←Ile270(non-conserved, Val295 in SpCyclin B)		Arg157→ (conserved)	←Tyr292 (non-conserved, Phe267 in HsCyclin A3) ←Val295 (non-conserved, Ile270 in HsCyclin A3) ←Glu293 (conserved) ←Glu294 (conserved)
Tyr180→ (conserved)	←Asn173(non-conserved, Leu198 in SpCyclin B)		Ser158→ (non-conserved, Ala in HsCDK2)	←Tyr292 (non-conserved, Phe267 in HsCyclin A3)
Ser181→ (conserved)	←Asn173(non-conserved, Leu198 in SpCyclin B)		Phe159→ (conserved)	←Leu198 (non-conserved, Asn173 in HsCyclin A3) ←Ile207 (conserved) ←Arg342 (non-conserved, Gln317 in HsCyclin A3) ←Met199 (non-conserved, Glu174 in HsCyclin A3)
Thr182→ (conserved)	←Asn173(non-conserved, Leu198 in SpCyclin B) ←Val175 (conserved)		Gly160→ (conserved)	←Arg341 (non-conserved, Trp316 in HsCyclin A3) ←Asn338 (non-conserved, Gln313 in HsCyclin A3)
Pro271→ (conserved)	←Val175 (conserved)		Val161→ (conserved)	←Arg341 (non-conserved, Trp316 in HsCyclin A3) ←Asn335 (non-conserved, Trp310 in HsCyclin A3) ←Asn338 (non-

				conserved, Gln313 in HsCyclin A3) ←Glu255 (conserved)
Asn272→ (non-conserved)	←Glu174 (non-conserved, Met199 in SpCyclin B)		Pro162→ (conserved)	←Arg341 (non-conserved, Trp316 in HsCyclin A3) ←Leu198 (non-conserved, Asn173 in HsCyclin A3)
Ser276→ (conserved)	←Val175 (conserved) ←Asp177(non-conserved, Asp202 in SpCyclin B) ←Tyr178 (conserved)		Leu163→ (non-conserved, Val in HsCDK2)	←Leu198 (non-conserved, Asn173 in HsCyclin A3)
Ala277→ (conserved)	←Tyr178 (conserved)		Tyr166→ (conserved)	←Val295 (non-conserved, Ile270 in HsCyclin A3) ←Met296 (non-conserved, Tyr271 in HsCyclin A3)
Lys278→ (conserved)	←Tyr178 (conserved) ←Asp177(non-conserved, Asp202 in SpCyclin B) ←Asp181 (conserved)		Tyr187→ (conserved)	←Leu198 (non-conserved, Asn173 in HsCyclin A3)
			Ser188→ (conserved)	←Leu198 (non-conserved, Asn173 in HsCyclin A3) ←Val200 (conserved) ←Met199 (non-conserved, Glu174 in HsCyclin A3)
			Thr189→ (conserved)	←Leu198 (non-conserved, Asn173 in HsCyclin A3) ←Val200 (conserved)

				←Tyr203 (conserved)
			Gly190→ (non-conserved, Ala in HsCDK2)	←Leu198 (non-conserved, Asn173 in HsCyclin A3) ←Val200 (conserved)
			Ala279→ (non-conserved, Asn in HsCDK2)	←Val200 (conserved)
			Ser283→ (conserved)	←Glu202 (non-conserved, Asp176 in HsCyclin A3) ←Tyr203 (conserved)
			Ala284→ (conserved)	←Tyr203 (conserved)
			Lys285→ (conserved)	←Tyr203 (conserved) ←Asp206 (conserved)

Table 5.3 Residues involved inter-subunit contacts at the CDK-cyclin dimer interface

The co-ordinate file for the SpCyclin B theoretical model was re-numbered (198-450) and given a co-ordinate chain identifier "B" using the program PDBSET (CCP4, 1994) before superimposition onto BvCyclin A3 in the graphics program O (Jones, 1991). The two co-ordinate files (i.e. SpCDK1 and SpCyclin B) were merged, edited and the inter-subunit contact residues within a range of 0-3.9Å were determined.

5.3.14(a) The PSTAIRE region

The crystal structure of HsCDK2-HsCyclin A3 showed that the cyclin box clamped onto and surrounded the middle part of the PSTAIRE helix (residues 46 – 56) of HsCDK2. This interaction was mediated by extensive hydrophobic interactions between Ile49, Ile52 and Leu54 of HsCDK2 and the following HsCyclin A3 residues; Leu263 (non-conserved), Phe267 (non-conserved), Leu299 (conserved), Phe304 (conserved), Leu306 (conserved) and Ala307 (conserved). These interactions are conserved in the SpCDK1-SpCyclin B theoretical model. The differences between the two complex models include a Leu263→Ile(258) change and a Phe267→Tyr292. Also, there was a

network of side chain-backbone hydrogen bond interactions in the loop preceding the PSTAIRE helix i.e. Glu42, Gly43 and Val44 of HsCDK2 and Lys266 and Glu295 of HsCyclin A3. These residues and the interactions are also conserved in the SpCDK1-SpCyclin B theoretical model. In this part of the structural models, the main difference lies in the C-terminal portion of the PSTAIRE helix where Lys56 (conserved) of HsCDK2 donates two hydrogen bonds to the Asp305 (Asn330 in SpCDK1) side chain and the Thr303 (Glu 228 in SpCDK1) backbone carbonyl of HsCyclin A3. Glu57 (conserved) of HsCDK2 accepts a hydrogen bond from the Tyr185 (conserved) of HsCyclin A3.

5.3.14(b) The activation segment

In the free, uncomplexed HsCDK2 structure (De Bondt *et al.*, 1993), the activation segment blocked the entrance to the catalytic cleft. In the HSCDK2-HsCyclin A3 complex, this blockade was relieved and the N-terminal portion of the activation segment interacted with helices α1, α2, α3 and the N-terminal helix of HsCyclin A3. There were van der Waals contacts between Ala151, Phe152 (conserved) and Tyr159 (conserved) of HsCDK2 and Phe267 (Tyr292 in SpCyclin B), Ile182 (conserved) and Ile270 (Val295 in SpCyclin B) of HsCyclin A3, respectively. Also, the guanidinium group of Arg150 (conserved) of HsCDK2 formed hydrogen bonds with the backbone carbonyl groups of Glu269 (conserved) and Ile270 (Val 295 in SpCyclin B) of HsCyclin A3. The side chains of Arg157 (conserved) of HsCDK2 interacted with that of Gln228 of HsCyclin A3.

5.3.14(c) The N-terminal lobe of HsCDK2

The interactions between the N-terminal β-sheet of HsCDk2 and helix α5 of HsCyclin A3 were mainly hydrophobic. The side chains of residues Val69, His71 (conserved), Leu76 and Leu37 (conserved) of HsCDK2 packed against those of His296 (Arg321 in SpCDK1) and Phe304 (conserved) of HsCyclin A3.

5.3.14(d) The C-terminal lobe of HsCDK2

The interactions between the C-terminal lobe of HsCDK2 and the N-terminal helix of HsCyclin A3 were mainly polar in nature. This feature is also conserved in the theoretical model of the SpCDK1-SpCyclin B complex model.

5.4 DISCUSSION

The overall aim of the experiments described in this chapter was to understand the molecular basis of CDK-cyclin specificity by determining the atomic structure of SpCDK1-SpCyclin B3 and then embarking on a comparative analysis with other known CDK-cyclin atomic structures. The *cdc2*$^{+}$ and *cdc13*$^{+}$ genes from *S. pombe* encode for SpCDK1 and SpCyclin B proteins, respectively. Although, these two genes were identified and cloned long before homologues from higher animals were discovered, the complete absence in the literature of successful overproduction strategies despite sustained investigations belies the difficulty of embarking on X-ray crystallographic analysis of these two important cell cycle regulatory proteins. Studies with CDK2 of the CDK family of proteins have enjoyed the greatest success with respect to structural studies (see chapter 1). Because the level of both recombinant SpCDK1 and SpCyclin B proteins (as judged by SDS-PAGE) produced by the expression system were relatively high, the author was convinced that a sustained campaign aimed at producing soluble and active forms of each protein in quantities suitable for crystallisation trials would eventually bear fruit. The perseverance did not pay off as I expected but it led me into unplanned but fruitful alleyways of exploring strategies for improving heterologous protein expression in *E. coli*, protein folding (*in vivo* and *in vitro*) and comparative homology modelling.

5.4.1 Protein folding in vivo and in vitro

Protein folding is perhaps the single most important process in biology because it converts linear polypeptide chains into 3-D structures that endows proteins with all their vital structural and functional properties (Ellis and Hartl, 1996). It is an active area of research whose importance is becoming increasingly more and more relevant in these

days of genome sequencing (Dobson and Ellis, 1998). It is well established that the information for the folding process is encoded within the aminoacyl sequence of the polypeptide chain (Anfinsen, 1973). The *in vivo* environmental milieu in which a nascent polypeptide chain folds into an active 3-D entity is vastly more complex than that faced by an unfolded polypeptide chain *in vitro*. It has been estimated that each cell of *E. coli* growing at 37°C with a doubling time of 40 minutes makes about 1000 polypeptide chains of average mass 40KDa every second (Ellis and Hartl, 1996; Neidhardt, 1987). These polypeptide chains occupy a cytoplasmic volume of about $0.6\mu m^3$ and they are synthesised by about 15×10^3 ribosomes; these organelles would almost be touching one another when arranged in polysomes. A related contribution to the complexity of the *in vivo* milieu is the phenomenon of macromolecular crowding (Dobson and Ellis, 1998; Ellis and Hartl, 1996). This refers to the fact that the concentration of macromolecules inside living cells tends to be so high that a significant proportion of the cellular volume is physically occupied and therefore unavailable to nascent macromolecules. There are regulatory mechanisms in place e.g. programmed proteolysis, efficient packing and compartmentalisation that ensures that the cell retains only the amount of macromolecules that it can physically accommodate. Nevertheless, the effective concentration of nascent polypeptides is greatly increased as a result of this phenomenon. For instance, it has been estimated that the total concentration of protein and RNA inside an *E. coli* cell is about 340mg/ml (Dobson and Ellis, 1998). This amount would be much greater in most of the cases described in chapters 4 and 5 of this thesis and other recombinant expression systems where the target protein(s) is being overexpressed to levels that may be several orders of magnitude greater than the normal physiological concentrations. The adverse effect of macromolecular crowding is the increased opportunity for nascent polypeptides and folding intermediates to interact *in vivo*. Macromolecular association constants for proteins in *E. coli* are predicted to be several orders of magnitude greater than that predicted for the dilute protein solutions required for efficient *in vitro* refolding experiments. Thus increased macromolecular association constants leads to an increased chance of protein self-association leading to aggregation. The folding intermediate via which a denatured protein folds to the native state as described in chapter 3, is extremely prone to aggregation because it displays, albeit

transiently, surface hydrophobic patches which can bind to similar molecules to form insoluble aggregates. This is why in practice, successful refolding of a pure denatured polypeptide *in vitro* is favoured by the use of low concentrations of protein and low temperatures; both of which do not favour hydrophobic interactions (Jaenicke, 1998; Jaenicke and Seckler, 1997). Relatively, SpCyclin B3 has more hydrophobic residues than BvCyclin A3 and this could render the folding the folding process of the former a more fragile process.

The complexity of the environment in which proteins fold *in vivo* compared to refolding *in vitro* requires some form of regulatory mechanisms to prevent aggregation. One of the major functions of molecular chaperones is to combat aggregation *in vivo* by binding to nascent polypeptide chains, folding intermediates and misfolded proteins to oversee their proper folding to the native state (Cole, 1996). Other enzymes that are important for folding *in vivo* are protein disulphide isomerase (PDI) and peptidyl prolyl cis-trans isomerase. Despite the evolution of elaborate *in vivo* mechanisms for ensuring the successful folding of proteins, the advent of recombinant DNA technology and the subsequent widespread application of heterologous protein expression systems has shown that the destination of a great number of foreign genes when expressed in *E. coli*, is not the desired state of soluble and active entities but as insoluble, inactive aggregates in inclusions bodies. Often, the IBs contain the target protein in a highly enriched but aggregated form (Rudolph and Lilie, 1996). Protein misfolding and aggregation occurs naturally and it is the basis of several pathological conditions (Dobson and Ellis, 1998; Kaufman, 1998). Due to the recognition that inclusion body formation is the rule rather than the exception during heterologous expression in *E. coli* (Marston, 1986), refolding science has mushroomed a billion dollar industry in which important therapeutic targets (like cyctokines) are produced from the refolding of IBs (Anon., 1992). Although they were originally defined as aggregates of misfolded protein (Kane and Hartley, 1988), there is ample evidence that some proteins encapsulated in IBs retain significant levels of native structure and activity (Chaffotte *et al.*, 1992; Hoess *et al.*, 1988; Tokatlidis *et al.*, 1991). It is now clear that inclusion body formation is a manifestation of a stress response. Among the several reasons why the expression of a target protein in *E. coli* may invoke such a response include the fact the cellular milieu in which a foreign gene is

expressed in *E. coli* may be markedly different with regards to chemical properties such as redox potential, intracellular pH, concentration of unbound water etc...etc. It may also differ in the type and specificity of molecular chaperones present and the levels which catalyse post-translational modifications. In most cases, the target protein in its natural environment may be expressed at very low levels in highly differentiated cell types but in a recombinant system, the foreign gene is almost always being induced to accumulate the target protein to unphysiological levels. The survival of the *E. coli* cell may then depend on the efficient packaging of the target protein away as relatively harmless IBs. Although all the above factors may have contributed to the persistent direction of SpCyclin B3 and SpCDK1 into IBs, the inability to refold these two molecules *in vitro* despite extensive attempts suggests another plausible reason.

The nucleotide codons AGA and AGG (both encode for the amino acid arginine) occur at a much higher frequency in eukaryotic and viral genes than those from *E. coli* (Brinkmann *et al.*, 1989; Schenk *et al.*, 1995). The AGA codon is translated in *E. coli* by the rare $tRNA_{arg4}$, which is encoded by the *argU* gene. Also, the *E. coli dnaY* gene codes for the minor $tRNA_{AGA/AGG}$ for arginine (Garcia *et al.*, 1986). An increased level of tRNAarg4 *in vivo* has led to additional translation of the AGG codon (Spanjaard *et al.*, 1990). This tRNA is also indirectly involved in the process of DNA replication possibly by translating the rare codons of proteins essential to DNA replication (Garcia *et al.*, 1986). The frequency and arrangement of these rare codons in a target gene coupled with the paucity of the corresponding tRNA in *E. coli* has had adverse effects on several heterologous bacteria expression systems (Bonekamp and Jensen, 1988; Robinson *et al.*, 1984; Spanjaard *et al.*, 1990). The meagre pool of $tRNA_{arg4}$ frequently causes ribosomal frameshifts of +1 at tandem AGA and AGG codons (Spanjaard *et al.*, 1990). The expression of genes containing two or more consecutive AGA/AGG triplets often leads to the formation of truncated protein products as a result of an out-of-frame termination signal (Schenk *et al.*, 1995). In the production of IGF-1 in *E. coli*, mistranslation of AGA codons into lysine instead of arginine led to the formation of a heterogeneous product (Setharam *et al.*, 1988). Co-expression of a plasmid-encoded argU or dnaY gene with the target eukaryotic gene in *E. coli* has led to the efficient production of the recombinant protein in several cases (Brinkmann *et al.*, 1989). SpCDK1 and SpCyclin B3 contain rare

AGA/AGG codons. The former has 3 AGA and 2 AGG whilst the latter has 4 AGA and 6 AGG codons but both gene fragments lack a consecutive repeat of these rare codons. Consequently, the overexpression of SpCyclin B3 and SpCDK1 did not result in the formation of truncated products (as judged by the gels and N-terminal sequencing) of bands corresponding to the expected sizes of the products of both *cdc2+* and SpCyclinBΔ195 genes. The overexpression of these two genes at high levels, the lack of success in both the *in vitro* refolding and *in vivo* co-expression experiments (using cell cycle subunits and GroESL) suggests that the presence of these rare codons in SpCDK1 and SpCyclin B3 may have led to the mistranslation of one or more crucial structural amino acids into new residues that could not hold the tertiary structures of these proteins together as the natural amino acids do in (say) *S. pombe* cells. Full length amino acid sequencing of the protein product could have shown whether this was plausible or not. The sample was submitted for N-terminal sequencing and only the first 12 residues were determined.

The result of the co-expression of GroESL with SpCyclin B3 supports this hypothesis. One of the functions of GroESL is to hold on to a mis-folded polypeptide and re-direct it back to the correct folding pathway. If SpCyclin B3 were mis-folded due to any other reason other than mis-translation, one would have expected a certain amount of the SpCyclin B3 molecules to be successfully chaperoned onto the correct folding pathway *in vivo*. The observation that all the GroESL molecules that recognised SpCyclin B3 and bound to it were directed into IBs suggests that the basic problem (misfolding as a result of mistranslation) could not, as expected, be rectified by these *E. coli* chaperones. It is plausible that the recombinant SpCyclin B3 and SpCDK1 molecules were being produced in *E. coli* as inherently misfolded proteins due to the mistranslation of the one or more rare arginine codons present in these two genes. If that were case, it would explain the extensive but unsuccessful attempts at refolding SpCyclin B3 (individually) and with SPCDK1 because the native state of these recombinant molecules had been altered to inherently misfolded proteins as result of mistranslation. This hypothesis can be tested by co-expressing SpCyclin B3 and SpCDK1 with overexpressible constructs of the argU/dnaY genes.

On the basis of kinetic studies and the crystal structure of HsCDK2-HsCyclin A3, it was proposed that the interaction of any cyclin with any CDK involved the same hydrophobic contacts (Heitz *et al.*, 1997) and this was predicted to induce a marked conformational change in the catalytic cleft of the CDKs as was achieved by HsCyclin A3 on binding to HsCDK2 (Jeffrey *et al.*, 1995). Natural and unnatural CDK-cyclin complexes have been created *in vitro* and the strength of the binding interactions have been measured by fluorescence spectroscopy (Heitz *et al.*, 1997). CDK1, CDK2 and CDK7 bound to their respective cyclin partners with a dissociation constant of about 50nM (Heitz *et al.*, 1997). Unnatural complexes like CDK1-cyclin H, CDK2-cyclin H and CDK7-cyclin A were isolated *in vitro* by size-exclusion HPLC and identified by SDS-PAGE (Heitz *et al.*, 1997). These complexes were stable at salt concentrations of up to 400mM, suggesting that the interactions were mainly hydrophobic in nature (Heitz *et al.*, 1997). The affinity of a CDK for an unusual cyclin partner was about 25-50 fold lower than that for a natural partner (Heitz *et al.*, 1997). Based on these results and the X-ray structure of HsCDK2-HsCyclin A3, the authors proposed a common interactive mechanism for CDK-cyclin interactions. Comparative homology modelling (data not shown) of human CDK1, CDK2, CDK7, cyclins A, B, E and H has provided a molecular basis for the *in vitro* observations of unnatural CDK-cyclin complexes. The analysis showed that in general, any of the above CDKs mediate protein-protein interactions with any of the cyclins tested with a conserved core residues and this may account for the relatively high affinity observed for unnatural CDK-cyclin complexes.

5.5 LITERATURE CITED

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.

Amrein, K.E., Takacs, B., Stieger, M., Molnos, J., Flint, N.A. and Burn, P. (1995) Purification and characterization of recombinant human p50csk protein- tyrosine kinase from an Escherichia coli expression system overproducing the bacterial chaperones GroES and GroEL. *Proc Natl Acad Sci U S A*, **92**, 1048-1052.

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223-230.

Anon. (1992) Cytokine pipeline coming under steam. *European biotechnology newsletter*, **139**.

Barton, G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Engineering*, **6**, 37-40.

Belyaev, A.S., Hails, R.S. and Roy, P. (1995) High-level expression of five foreign genes by a single recombinant baculovirus. *Gene*, **156**, 229-233.

Belyaev, A.S. and Roy, P. (1993) Development of baculovirus triple and quadruple expression vectors: co- expression of three or four bluetongue virus proteins and the synthesis of bluetongue virus-like particles in insect cells. *Nucleic Acids Res*, **21**, 1219-1223.

Better, M., Chang, C.P., Robinson, R.R. and Horwitz, A.H. (1988) Escherichia coli secretion of an active chimeric antibody fragment. *Science*, **240**, 1041-1043.

Blackwell, J.R. and Horgan, R. (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett*, **295**, 10-12.

Bodenbach, L., Fauss, J., Robitzki, A., Krehan, A., Lorenz, P., Lozeman, F.J. and Pyerin, W. (1994) Recombinant human casein kinase II. A study with the complete set of subunits (alpha, alpha' and beta), site-directed autophosphorylation mutants and a bicistronically expressed holoenzyme. *Eur J Biochem*, **220**, 263-273.

Bonekamp, F. and Jensen, K.F. (1988) The AGG codon is translated slowly in E. coli even at very low expression levels. *Nucleic Acids Res*, **16**, 3013-3024.

Boss, M.A., Kenten, J.H., Wood, C.R. and Emtage, J.S. (1984) Assembly of functional antibodies from immunoglobulin heavy and light chains synthesised in E. coli. *Nucleic Acids Res*, **12**, 3791-3806.

Bourne, Y., Watson, M.H., Hickey, M.J., Holmes, W., Rocque, W., Reed, S.I. and Tainer, J.A. (1996) Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle regulatory protein CksHs1. *Cell*, **84**, 863-874.

Brinkmann, U., Mattes, R.E. and Buckel, P. (1989) High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene*, **85**, 109-114.

Brooks, B.R., Bruccoleri, E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *Journal of computational chemistry*, **4**, 187-217.

Brown, N.R., Noble, M.E.M., Endicott, J.A., Garman, E.F., Wakatsuki, S., Mitchell, E.P., Rasmussen, B., Hunt, T. and Johnson, L.N. (1995) The crystal structure of cyclin A. *Structure*, **3**, 1235-1247.

Brunger, A.T. (1992) *X-PLOR: Version 3.1; a system for protein crystallography and NMR.* Yale University Press, New Haven.

Cabilly, S. (1989) Growth at sub-optimal temperatures allows the production of functional, antigen-binding Fab fragments in Escherichia coli. *Gene*, **85**, 553-557.

Cabilly, S., Riggs, A.D., Pande, H., Shively, J.E., Holmes, W.E., Rey, M., Perry, L.J., Wetzel, R. and Heyneker, H.L. (1984) Generation of antibody activity from immunoglobulin polypeptide chains produced in Escherichia coli. *Proc Natl Acad Sci U S A*, **81**, 3273-3277.

CCP4. (1994) The CCP4 (Collaborative Computational Project Number 4) suite: programmes for protein crystallography. *Acta Cryst. D*, **D50**, 760-763.

Chaffotte, A.F., Guillou, Y. and Goldberg, M.E. (1992) Inclusion bodies of the thermophilic endoglucanase D from Clostridium thermocellum are made of native enzyme that resists 8 M urea. *Eur J Biochem*, **205**, 369-373.

Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *Embo J*, **5**, 823-826.

Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A*, **86**, 2172-2175.

Cole, P.A. (1996) Chaperone-assisted protein expression. *Structure*, **4**, 239-242.

Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Stuber, D. and Henco, K. (1994) 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol Biol*, **31**, 371-387.

Dale, G.E., Schonfeld, H.J., Langen, H. and Stieger, M. (1994) Increased solubility of trimethoprim-resistant type S1 DHFR from Staphylococcus aureus in Escherichia coli cells overproducing the chaperonins GroEL and GroES. *Protein Eng*, **7**, 925-931.

De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. and Kim, S.-H. (1993) Crystal structure of cyclin-dependent kinase 2. *Nature*, **363**, 595-602.

Desai, D., Wessling, H.C., Fisher, R.P. and Morgan, D.O. (1995) Effects of phosphorylation by CAK on cyclin binding by CDC2 and CDK2. *Mol. Cell. Biol.*, **15**, 345-350.

Dobson, C.M. and Ellis, R.J. (1998) Protein folding and misfolding inside and outside the cell. *Embo J*, **17**, 5251-5254.

Doherty, A.J., Ashford, S.R., Brannigan, J.A. and Wigley, D.B. (1995) A superior host strain for the over-expression of cloned genes using the T7 promoter based vectors. *Nucleic Acids Res*, **23**, 2074-2075.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, R. and Davies, D.R. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases [see comments]. *Science*, **266**, 1981-1986.

Elce, J.S., Hegadorn, C., Gauthier, S., Vince, J.W. and Davies, P.L. (1995) Recombinant calpain II: improved expression systems and production of a C105A active-site mutant for crystallography. *Protein Eng*, **8**, 843-848.

Ellis, R.J. and Hartl, F.U. (1996) Protein folding in the cell: competing models of chaperonin function. *Faseb J*, **10**, 20-26.

French, T.J., Marshall, J.J. and Roy, P. (1990) Assembly of double-shelled, viruslike particles of bluetongue virus by the simultaneous expression of four structural proteins. *J Virol*, **64**, 5695-5700.

French, T.J. and Roy, P. (1990) Synthesis of bluetongue virus (BTV) corelike particles by a recombinant baculovirus expressing the two major structural core proteins of BTV. *J Virol*, **64**, 1530-1536.

Garcia, G.M., Mar, P.K., Mullin, D.A., Walker, J.R. and Prather, N.E. (1986) The E. coli dnaY gene encodes an arginine transfer RNA. *Cell*, **45**, 453-459.

Georgiou, G. and Valax, P. (1996) Expression of correctly folded proteins in Escherichia coli. *Curr Opin Biotechnol*, **7**, 190-197.

Graham-Siegenthaler, K., Gauthier, S., Davies, P.L. and Elce, J.S. (1994) Active recombinant rat calpain II. Bacterially produced large and small subunits associate both in vivo and in vitro. *J Biol Chem*, **269**, 30457-30460.

Heitz, F., Morris, M.C., Fesquet, D., Cavadore, J.C., Doree, M. and Divita, G. (1997) Interactions of cyclins with cyclin dependent kinases: a common interactive mechanism. *Biochemistry*, **36**, 4995-5003.

Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *Comput Appl Biosci*, **8**, 189-191.

Hochuli, E. (1990) Purification of recombinant proteins with metal chelate adsorbent. *Genet Eng*, **12**, 87-98.

Hochuli, E., Dobeli, H. and Schacher, A. (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J Chromatogr*, **411**, 177-184.

Hockney, R.C. (1994) Recent developments in heterologous protein production in Escherichia coli. *Trends Biotechnol*, **12**, 456-463.

Hoess, A., Arthur, A.K., Wanner, G. and Fanning, E. (1988) Recovery of soluble biologically active recombinant proteins from total bacterial lysates using ion exchange resin. *Bio/Technology*, **9**, 1214.

Holtet, T.L., Nielsen, K.L., Etzerodt, M., Moestrup, S.K., Gliemann, J., Sottrup-Jensen, L. and Thogersen, H.C. (1994) Receptor-binding domain of human alpha 2-macroglobulin. Expression, folding and biochemical characterization of a high-affinity recombinant derivative. *FEBS Lett*, **344**, 242-246.

Holzinger, A., Phillips, K.S. and Weaver, T.E. (1996) Single-step purification/solubilization of recombinant proteins: application to surfactant protein B. *Biotechniques*, **20**, 804-806, 808.

Huang, J. and Raff, J.W. (1999) The disappearance of cyclin B at the end of mitosis is regulated spatially in Drosophila cells. *Embo J*, **18**, 2184-2195.

Huang, X. and Miller, M.A. (1991) A time-efficeint, linaer-space local similarity algorithm. *Advances in Applied Mathematics*, **12**, 337-367.

Jaenicke, R. (1998) Protein self-organization in vitro and in vivo: partitioning between physical biochemistry and cell biology. *Biol Chem*, **379**, 237-243.

Jaenicke, R. and Seckler, R. (1997) Protein misassembly in vitro. *Adv Protein Chem*, **50**, 1-59.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J. and Pavletich, N.P. (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, **376**, 313-320.

Jenkins, T.M., Hickman, A.B., Dyda, F., Ghirlando, R., Davies, D.R. and Craigie, R. (1995) Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc. Natl. Acad. Sci. USA*, **92**, 6057-6061.

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

Kane, J.F. and Hartley, D.L. (1988) Formation of recombinant protein inclusions in *E. coli*. *Trends in biotechnology*, **6**, 95.

Kaufman, R.J. (1998) Meeting report: molecular chaperones and the heat shock response. *Biochimica et Biophysica Acta*, **1423**, R13-R27.

Larimer, F.W., Mural, R.J. and Soper, T.S. (1990) Versatile protein engineering vectors for mutagenesis, expression and hybrid enzyme formation. *Protein Eng*, **3**, 227-231.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Biotechnology (N Y)*, **11**, 187-193.

LaVallie, E.R. and McCoy, J.M. (1995) Gene fusion expression systems in Escherichia coli. *Curr Opin Biotechnol*, **6**, 501-506.

Li, C., Schwabe, J.W., Banayo, E. and Evans, R.M. (1997) Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc Natl Acad Sci U S A*, **94**, 2278-2283.

Lijnzaad, P., Berendsen, H.J. and Argos, P. (1996) Hydrophobic patches on the surfaces of protein structures. *Proteins*, **25**, 389-397.

London, J., Skrzynia, C. and Goldberg, M.E. (1974) Renaturation of Escherichia coli tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. *Eur J Biochem*, **47**, 409-415.

Looker, D., Mathews, A.J., Neway, J.O. and Stetler, G.L. (1994) Expression of recombinant human hemoglobin in Escherichia coli. *Methods Enzymol*, **231**, 364-374.

Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83-85.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol Rev*, **60**, 512-538.

Marston, F.A. (1986) The purification of eukaryotic polypeptides synthesized in Escherichia coli. *Biochem J*, **240**, 1-12.

Moore, J.T., Uppal, A., Maley, F. and Maley, G.F. (1993) Overcoming inclusion body formation in a high-level expression system. *Protein Expr Purif*, **4**, 160-163.

NEB. (1999) *New England Biolabs catalogue*. New England Biolabs, Inc.

Neidhardt, F.C. (ed.) (1987) *Chemical composistion of Escherichia coli*. American society for microbiology, Washington D.C.

Nicholls, A., Bharadwai, R. and Honig, B. (1993) GRASP - graphical representation and analysis of surface properties. *Biophysical journal*, **64**.

Noble, M.E.M., Endicott, J.A., Brown, N.R. and Johnson, L.N. (1997) The cyclin box fold: protein recognition in cell-cycle and trasncription control. *Trends in Biochemical Sciences*, **22**, 482-487.

Novagen. (1999) pET System Manual. .

Oswald, T. and Rinas, U. (1996) Chloramphenicol resistance interferes with purification of histidine- tagged fusion proteins from recombinant Escherichia coli. *Anal Biochem*, **236**, 357-358.

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, **85**, 2444-2448.

Peeper, D.S., Parker, L.L., Ewen, M.E., Toebes, M., Hall, F.L., Xu, M., Zantema, A., VanderEb, A.J. and Piwnica-Worms, H. (1993) A- and B-type cyclins differentially modulate substrate specificity of cyclin-cdk complexes. *EMBO J.*, **12**, 1947-1954.

Peitsch, M.C. (1995) Protein modelling by email. *Biotechnology*, **13**, 658-660.

Peitsch, M.C. (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem Soc Trans*, **24**, 274-279.

Peitsch, M.C., Wells, T.N., Stampf, D.R. and Sussman, J.L. (1995) The Swiss-3DImage collection and PDB-Browser on the World-Wide Web. *Trends Biochem Sci*, **20**, 82-84.

Pike, A.C. and Acharya, K.R. (1994) A structural basis for the interaction of urea with lysozyme. *Protein Sci*, **3**, 706-710.

Pines, J. and Hunter, T. (1991) Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. *J. Cell Biol.*, **115**, 1-17.

Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) Codon usage can affect efficiency of translation of genes in Escherichia coli. *Nucleic Acids Res*, **12**, 6663-6671.

Rogl, H., Kosemund, K., Kuhlbrandt, W. and Collinson, I. (1998) Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett*, **432**, 21-26.

Rudolph, R. and Lilie, H. (1996) In vitro folding of inclusion body proteins. *Faseb J*, **10**, 49-56.

Schenk, P.M., Baumann, S., Mattes, R. and Steinbiss, H.H. (1995) Improved high-level expression system for eukaryotic genes in Escherichia coli using T7 RNA polymerase and rare ArgtRNAs. *Biotechniques*, **19**, 196-198, 200.

Senger, M., Flores, T., Glatting, K., Ernst, P., Hotz-Wagenblatt, A. and Suhai, S. (1998) W2H: WWW interface to the GCG sequence analysis package. *Bioinformatics*, **14**, 452-457.

Setharam, R., Heeren, R., Wong, E.Y., Braford, S.R., Klein, B.K., Aykent, S., Kotts, C.E., Mathis, K.J., Bishop, B.F., Jennings, M.J., Smith, C.E. and Siegel, N.R. (1988) Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. *Biochemistry and biophysics research communications*, **155**, 518-523.

Shi, P.Y., Maizels, N. and Weiner, A.M. (1997) Recovery of soluble, active recombinant protein from inclusion bodies. *Biotechniques*, **23**, 1036-1038.

Shirano, Y. and Shibata, D. (1990) Low temperature cultivation of Escherichia coli carrying a rice lipoxygenase L-2 cDNA produces a soluble and active enzyme at a high level. *FEBS Lett*, **271**, 128-130.

Skerra, A. and Pluckthun, A. (1988) Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science*, **240**, 1038-1041.

Spanjaard, R.A., Chen, K., Walker, J.R. and van Duin, J. (1990) Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg). *Nucleic Acids Res*, **18**, 5031-5036.

Stempfer, G., Holl-Neugebauer, B. and Rudolph, R. (1996) Improved refolding of an immobilized fusion protein. *Nat Biotechnol*, **14**, 329-334.

Stoll, V.S., Manohar, A.V., Gillon, W., MacFarlane, E.L., Hynes, R.C. and Pai, E.F. (1998) A thioredoxin fusion protein of VanH, a D-lactate dehydrogenase from Enterococcus faecium: cloning, expression, purification, kinetic analysis, and crystallization. *Protein Sci*, **7**, 1147-1155.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*, **185**, 60-89.

Thomas, J.G. and Baneyx, F. (1996) Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing Heat-shock proteins. *J Biol Chem*, **271**, 11141-11147.

Tokatlidis, K., Dhurjati, P., Millet, J., Beguin, P. and Aubert, J.P. (1991) High activity of inclusion bodies formed in Escherichia coli overproducing Clostridium thermocellum endoglucanase D. *FEBS Lett*, **282**, 205-208.

Vasina, J.A. and Baneyx, F. (1996) Recombinant protein expression at low temperatures under the transcriptional control of the major Escherichia coli cold shock promoter cspA. *Appl Environ Microbiol*, **62**, 1444-1447.

Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. (1996) Optimization of heterologous protein production in Escherichia coli. *Curr Opin Biotechnol*, **7**, 494-499.

Winder, S.J. and Kendrick-Jones, J. (1995) Protein production in three different expression vectors from a single polymerase chain reaction product. *Anal Biochem*, **231**, 271-273.

Wulfing, C., Lombardero, J. and Pluckthun, A. (1994) An Escherichia coli protein consisting of a domain homologous to FK506- binding proteins (FKBP) and a new metal binding motif. *J Biol Chem*, **269**, 2895-2901.

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., Fujimoto, J., Yamamoto, T. and Ishii, S. (1995) Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. *J Biol Chem*, **270**, 25328-25331.

Chapter 6: The targeting model of Cks function

CONTENTS

LIST OF FIGURES **260**

6.1 GENERAL DISCUSSION **261**

6.1.1 Interplay of spatial, temporal and proteolytic regulatory mechanisms in mitosis *264*

6.1.2 Future work *268*

6.2 LITERATURE CITED **268**

List of figures

Figure 6.1 Ternary complex of HsCDK2-HsCyclin A3-CksHs1 ---------- *262*

Figure 6.2 Regulation of mitotic exit in S. cerevisiae after (Bachant and Elledge, 1999; Visintin et al., 1999) ---------- *266*

6.1 General discussion

The work presented in this thesis has been concerned with elucidating the function of the cell regulatory protein called $p13^{suc1}$. The approach was to exploit the wealth of X-ray structures available for this small protein as a basis for biochemical analysis. As a prelude to the biochemical results obtained, the atomic structure of Suc1 was presented in chapters 2 and 3. Two notable molecular features of the SpSuc1 molecule were observed and investigations into their potential functional properties were carried out as described in chapter 4. In the final chapter of the thesis, I shall put the findings of the investigations described in the thesis and recent evidence (from diverse organisms) culled from several recent publications into a plausible functional model for Cks proteins.

As described in chapter 4, good structural evidence for the targeting model was first made available by the determination of the crystal structure of HsCDK2-CksHs1 complex structure (Bourne *et al.*, 1996). This showed that Cks proteins mediate protein-protein interactions with their CDK partners via their hydrophobic face leaving the putative phosphate binding site exposed to solvent. The structure-based mutational investigations reported in Chapter 4 support this conclusion. Then came evidence that SpSuc1 binds to the active, phosphorylated APC from clam oocytes (Sudakin *et al.*, 1997). During interphase, the APC is inactive and it is converted to the active form in late mitosis by a phosphorylation event initiated by CDK1-cyclin B. The demonstration that the interaction with the APC was mediated via the mitotic kinase and also that the bound APC was most efficiently eluted from SpSuc1-sepharose beads by p-nitrophenyl phosphate strengthened the suggestion that the conserved, exposed phosphate-binding site of Cks proteins may be involved in mediating protein-protein interactions. In the crystal structure of monomeric CksHs2, the conserved anion-binding site had a phosphate analogue bound at this site (Arvai *et al.*, 1995). Taken together, the evidence suggests that at some point in late mitosis, SpSuc1 could be sandwiched between CDK1-cyclin B and the APC. Structurally, the assembly of a CDK-cyclin-Cks ternary complex is feasible and this has been isolated *in vivo* (Booher *et al.*, 1989).

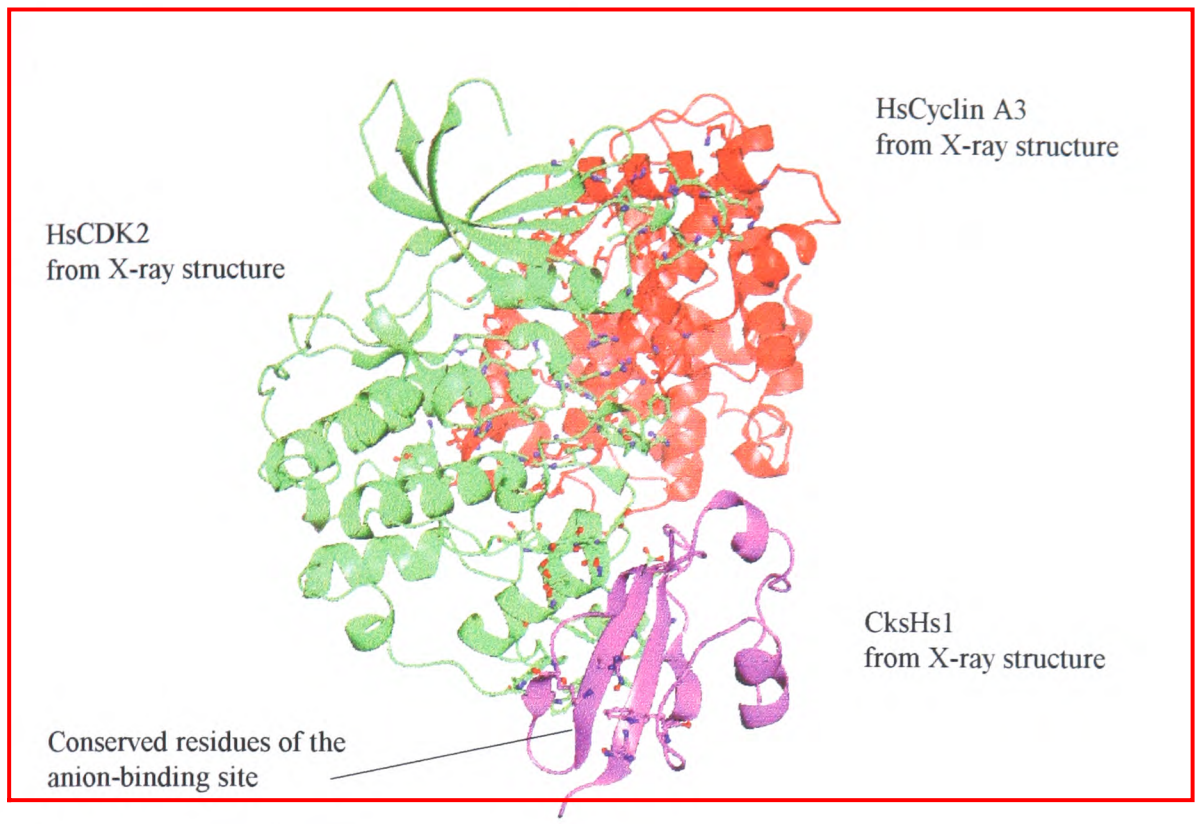


Figure 6.1 Ternary complex of HsCDK2-HsCyclin A3-CksHs1

This structural model was generated from the X-ray structures of HsCDK2-HsCyclin A3 and HsCDK2-CksHs1 complexes. Both the cyclin and the Cks protein can bind to the CDK at the same time at two different but specific sites without any stearic hindrance, as observed in the graphics program O (Jones, 1991). Also, in this assembly, the conserved phosphate-binding site is tantalisingly poised to interact with a phosphorylated target. Figure generated with Aesop (Dr. Noble, unpublished work).

The results of the APC activation assay presented in chapter 4 suggest that the inability of SpSuc1 (Ile94Ala:Leu96Ala) to bind to CDK1 and possibly the APC (Arg30Ala:Arg39Ala:Ile94Ala:Leu96Ala) leads to a loss in APC activation. The question as to which form of the mitotic kinase (i.e. active or inactive) was targeted to the APC by the Cks protein was answered (Egan and Solomon, 1998) by evidence showing that at physiological concentrations, CksHs2 preferentially bound to T161-phosphorylated HsCDK1-Cyclin B. Also, CksHs2 preferentially interacted with active, phosphorylated HsCDK2-cyclin A (Egan and Solomon, 1998) overturning the commonly held view that Cks proteins interacted with all forms of CDK1 and CDK2. As mentioned in chapter 4, the latter view was born out of the widespread use of SpSuc1-sepharose beads to isolate eukaryotic CDKs and it may not represent the situation *in vivo*. The selective binding of

Cks proteins to only the active, cyclin-bound kinase points to a role in recruiting a phosphorylating agent to the APC, (perhaps). Selective recruitment of the active kinase would serve two purposes: first, the elevation of the APC catalytic activity from "primed" to fully active mode by CDK1-cyclin B phosphorylation of a primed (i.e. partially phosphorylated APC). Second, the initiation of the CDK1-cyclin B kinase inactivation process (absolutely required for mitotic exit) by the degradation of cyclin B. In this sense the Cks protein would be playing a role in stopping a cell cycle as suggested for the *Arabidopsis thaliana* homologue (Suc1-At) of SpSuc1 (Dr. P. Fobert, personal communication, Plant Biotechnology Institute, Canada and at genome-www.stanford.edu/Arabidopsis/madison97/at97abs/).

The absolute requirement of Cks proteins for cyclin B degradation was provided by investigations carried out with the Cks homologue from frog (Patra and Dunphy, 1998). It was shown that interphase egg extracts depleted of Xe-p9 lost the ability to carry out cyclin B proteolysis (Patra and Dunphy, 1998) and this results in M-phase arrest. Extracts devoid of Xe-p9 after APC activation were able to degrade cyclin B and exit from mitosis showing that the Cks proteins are not directly involved in the protein degradation process (Patra and Dunphy, 1998). Another important observation made from Xe-p9 depleted egg extracts was that they lost the ability to hyperphosphorylate the Cdc27 subunit of the APC at the beginning of mitosis. Hyperphosphorylation of Cdc27 forms part of the APC activation initiation process. Also, small amounts of Xe-p9-Cdc27 complex could be isolated from M-phase but not interphase egg extracts, suggesting a direct physical association with Cks-APC subunit in the course of mitosis. The final evidence from this work was the observation that CDK1-cyclin B phosphorylation of Cdc27-APC and purified, bacterially produced, his-tagged Cdc27 was strongly enhanced by the presence of Xe-p9 suggesting that Cks proteins, modulate substrate specificity of CDK1-cyclin B in order to achieve, first, activation of the APC and second, the destruction of cyclin B. Both events culminate in the end and exit from M-phase. In summary, empirical evidence from yeast, vertebrates and plants show that:

1. Suc1-At may be in involved in stopping cell division.
2. CksHs2 preferentially binds to the fully active CDK1-cyclin B kinase.

3. SpSuc1 mediates protein-protein interactions with HsCDK2 via its (SpSuc1) hydrophobic patch.
4. SpSuc1 is necessary for APC activation.
5. Xe-p9 binds to phosphorylated Cdc27 (a subunit of the APC).
6. The presence of Xe-p9 *in vitro* enhances Cdc27 phosphorylation by CDK1-cyclin B.
7. Recent evidence from a yeast system supports the observations seen in *Xenopus*. This suggests that not only does *S. cerevisiae* CKS1 associate with the 19S subunit of the 26S proteosome, but the Cks protein may also regulate the activity of the proteosome (Kaiser *et al.*, 1999; Koepp *et al.*, 1999).
8. In *S. cerevisiae*, elimination of CKS1 resulted in the stabilisation of proteins that are normally ubiquinated and degraded to allow mitotic exit (Kaiser *et al.*, 1999).

In the next section, the latest evidence from investigations into mitotic regulation in *S. cerevisiae* and *D. melanogaster* would be woven together with the evidence above to present a plausible model of Cks function.

6.1.1 Interplay of spatial, temporal and proteolytic regulatory mechanisms in mitosis

As described in chapter 1, cyclin synthesis and degradation, specific phosphorylation and dephosphorylation, and association of CKI proteins all play an important role in ensuring that individual CDK-cyclin complexes act at specific points in the course of the cell cycle in order to maintain the correct temporal ordering of cell cycle events. The role of subcellular localisation of CDK-cyclin complexes and their regulatory subunits as a form of cell cycle and in particular mitotic control is beginning to be established (Pines, 1999; Scheffner, 1999; Yang and Kornbluth, 1999).

The events of mitosis include nuclear envelope breakdown, condensation of chromosomes, the assembly of mitotic spindles and in anaphase, the segregation of chromosomes to opposite poles of the cell. This is followed by cytokinesis in which the cell is cleaved in two and the two daughter cells enter interphase where they prepare for the next round of DNA replication and cell division. The main regulator of this process is

the CDK1-cyclin B kinase. Whereas the S-phase cyclins A and E are localised in the nucleus, cyclin B (whose synthesis begins in S-phase), accumulates in the cytoplasm in an inactive complex with CDK1. At prophase, cyclin B is phosphorylated and this seems to promote the rapid translocation of the inactive kinase complex into the nucleus where the kinase becomes activated to phosphorylate mitotic substrates. Important CDK regulators like Wee1, CDC25 and $p27^{Cip2/kip}$ are also spatially regulated (Lopez-Girona *et al.*, 1999; Pines, 1999; Scheffner, 1999; Tomoda *et al.*, 1999). For instance during interphase, inactive CDC25 is found in a complex with 14-3-3 proteins in the cytoplasm but just before mitosis, CDC25 (like cyclin B), is rapidly translocated into the nucleus where it (presumably) removes the inhibitory phosphate on CDK1 to promote mitosis. In dividing Drosophila cells, exit from mitosis also involves spatially regulated activation of the APC followed by the proteolysis of cyclin B (Huang and Raff, 1999). Is there an agent in the nucleus whose functional role is to bring, at the appropriate time (i.e. late mitosis) the active CDK1-cyclin B kinase and the partially phosphorylated APC together? The functional role of the Cks proteins could be the answer to this question. Although the most recent and elaborate account of the regulation of mitotic exit does not mention the plausible role of the Cks proteins, I think the targeting function can be accommodated into this model (figure 6.2).

Exit from mitosis in all dividing cells involves the inactivation of the mitotic kinase (CDK1-cyclin B). In *S. cerevisiae*, this is achieved by two mechanisms: the degradation of the Clb (cyclin B) protein subunit via the APC pathway and the inhibition of the remaining CDK1-cyclin B kinase by Sic1. Association with Cdh1/Hct1 activates the APC. In the course of mitosis, Cdh1/Hct1 is phosphorylated and rendered inactive by CDK1-cyclin B kinase. When the cell is ready to get out of mitosis Cdh1 is dephosphorylated (activated) by the newly discovered Cdc14 phosphatase (Shou *et al.*, 1999).

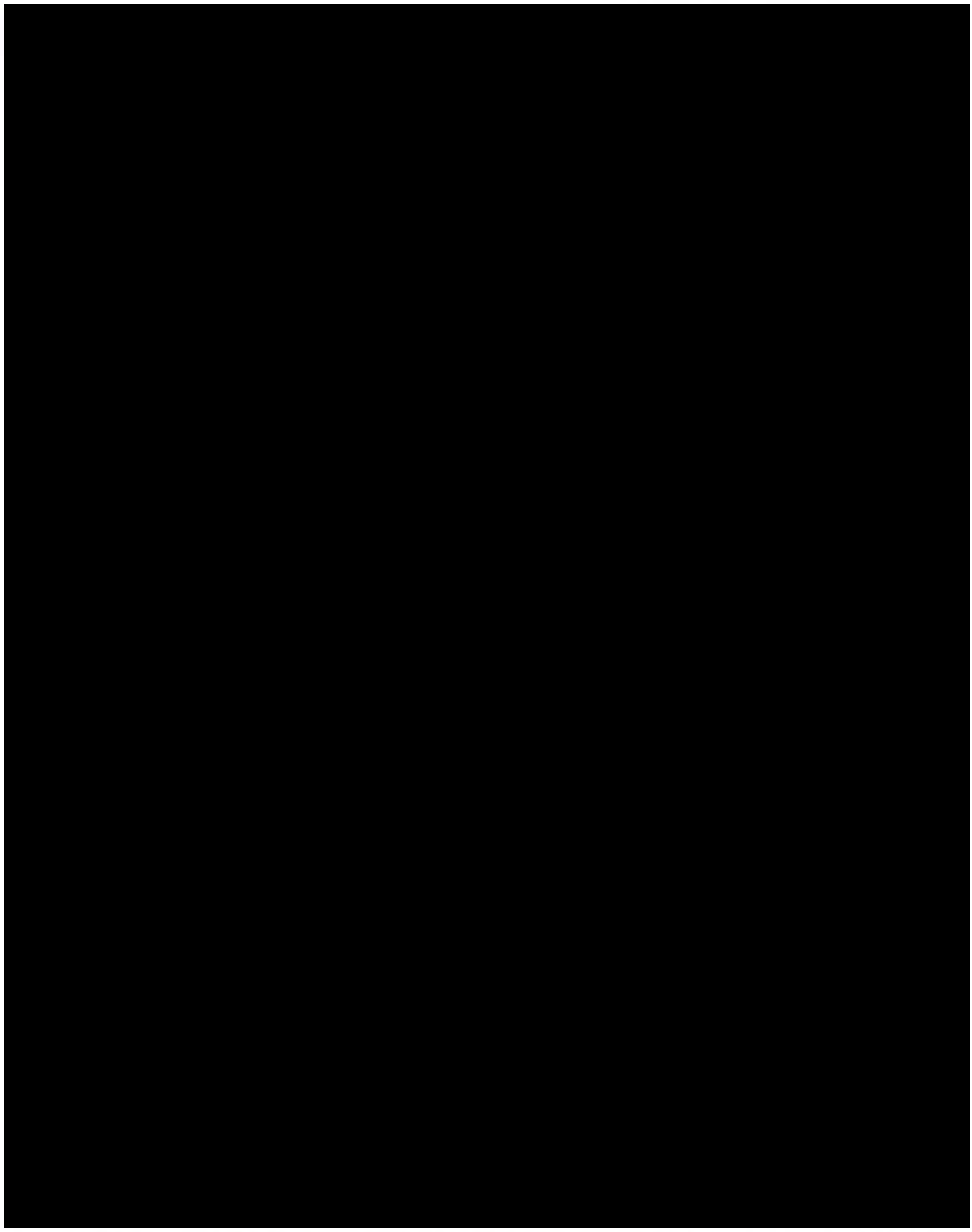


Figure 6.2 Regulation of mitotic exit in S. cerevisiae after (Bachant and Elledge, 1999; Visintin et al., 1999)

A key player in this process is a newly discovered protein phosphatase called Cdc14. In G1, the binding to an inhibitory protein called Cfi1 keeps Cdc14 in the nucleolus as an inactive enzyme. At the G1→S-phase boundary, the Sic1 protein (a CKI) is degraded thus releasing CDK1 to bind to the newly synthesised cyclin B. Accumulation of inactive CDK1-cyclin B complexes proceeds in the cytoplasm as the cell nears the end of DNA synthesis (S-phase). Cdc14 remains localised in the nucleolus when the cell enters mitosis but at the metaphase-to-anaphase transition point, Cdh14 is released into the entire nucleus and parts of the cytoplasm. This initiates a series of events whose ultimate goal is to inactivate CDK1 and hence, exit from M-phase: first, Cdc14 dephosphorylates and thus activates the Swi5 transcription factor allowing fresh synthesis and accumulation of Sic1 protein. The newly synthesised Sic1 protein is phosphorylated and hence targeted for degradation but this fate is avoided by Cdc14-catalysed dephosphorylation of the phosphorylated Sic1. The liberated Sic1 begins to reduce CDK1-cyclin kinase activity by direct inhibition. Next, Cdc14 dephosphorylates Cdh1/Hct1 and this initiates APC activation. CKS1 (S. cerevisiae homologue of SpSuc1) then binds to the remaining CDK1-cyclin B complexes using its (CKS1) hydrophobic patch. Cdh1/Hct1 then targets the partially activated APC to the CDK1-cyclin B-CKS1 ternary complex and this results in further phosphorylations of the APC (perhaps on the yeast equivalent of Cdc27 seen in the Xenopus system) by the CDK1-cyclin B-CKS1 kinase. The fully activated APC then initiates polyubiqutination and subsequent proteolysis of the cyclin Clb subunit thus returning CDK1 to the inactive state and allowing the cell to exit mitosis.

The evolutionary conservation of the core cell cycle apparatus and the regulatory molecules - from yeast to humans - validates the presumption implicit in the model presented that these features are essentially similar in all organisms.

6.1.2 Future work

The immediate aim could be the unambiguous establishment of the phosphate-binding site of SpSuc1 as the site used in mediating protein-protein interactions with the APC. This could be done by repeating the APC activation assay using a mutant of SpSuc1 in which only the residues in the putative phosphate-binding site has been altered. Although theoretical models of a CDK-cyclin-Cks ternary complex are informative, elucidation of the structure of a HsCDK2-HsCyclin A3-CksHs1/2 ternary complex would increase the supporting experimental data for the targeting model. The putative physiological function of the Cks proteins lies in the process of regulated proteolysis in cell cycle control. Our current state of knowledge with respect to the biochemical and structural mechanisms involved in this process is relatively incomplete. A lot of the current understanding has come from the field of genetics. In the future, this knowledge deficit can be plugged by the provision of biochemical and structural data on the regulatory molecules involved in this process. Apart from the fact that it consists of about 12-13 subunits and it is about 1500KDa big, there is as yet, no experimental evidence relating the structure of the APC to its function. In the future, elucidation of the structure of all or some of the individual subunits of the APC should be a challenging but hopefully rewarding biological project. Since $Cdc27_{his6}$ from Xenopus is bacterially expressed and it can be phosphorylated *in vitro* by CDK1-cyclin B, it would be feasible to attempt to gain structural insights into their mode of interaction by crystallising a phosphorylated Cdc27-Xe-p9 complex for an atomic structure determination project.

6.2 Literature cited

Arvai, A.S., Bourne, Y., Hickey, M.J. and Tainer, J.A. (1995) Crystal structure of the human cell cycle protein CksHs1: single domain fold with similarity to kinase N-lobe domain. *J Mol Biol*, **249**, 835-842.

Bachant, J.B. and Elledge, S.J. (1999) Mitotic treasures in the nucleolus [news]. *Nature*, **398**, 757-758.

Booher, R.N., Alfa, C.E., Hyams, J.S. and Beach, D.H. (1989) The fission yeast cdc2/cdc13/suc1 protein kinase: regulation of catalytic activity and nuclear localization. *Cell*, **58**, 485-497.

Bourne, Y., Watson, M.H., Hickey, M.J., Holmes, W., Rocque, W., Reed, S.I. and Tainer, J.A. (1996) Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle regulatory protein CksHs1. *Cell*, **84**, 863-874.

Egan, E.A. and Solomon, M.J. (1998) Cyclin-stimulated binding of Cks proteins to cyclin-dependent kinases. *Mol Cell Biol*, **18**, 3659-3667.

Huang, J. and Raff, J.W. (1999) The disappearance of cyclin B at the end of mitosis is regulated spatially in Drosophila cells. *Embo J*, **18**, 2184-2195.

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

Kaiser, P., Moncollin, V., Clarke, D.J., Watson, M.H., Bertolaet, B.L., Reed, S.I. and Bailly, E. (1999) Cyclin-dependent kinase and Cks/Suc1 interact with the proteasome in yeast to control proteolysis of M-phase targets. *Genes Dev*, **13**, 1190-1202.

Koepp, D.M., Harper, J.W. and Elledge, S.J. (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle [In Process Citation]. *Cell*, **97**, 431-434.

Lopez-Girona, A., Furnari, B., Mondesert, O. and Russell, P. (1999) Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein [see comments]. *Nature*, **397**, 172-175.

Patra, D. and Dunphy, W.G. (1998) Xe-p9, a Xenopus Suc1/Cks protein, is essential for the Cdc2-dependent phosphorylation of the anaphase- promoting complex at mitosis. *Genes Dev*, **12**, 2549-2559.

Pines, J. (1999) Cell cycle. Checkpoint on the nuclear frontier [news; comment]. *Nature*, **397**, 104-105.

Scheffner, M. (1999) Moving protein heads for breakdown [news; comment]. *Nature*, **398**, 103-104.

Shou, W., Seol, J.H., Shevchenko, A., Baskerville, C., Moazed, D., Chen, Z.W., Jang, J., Charbonneau, H. and Deshaies, R.J. (1999) Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. *Cell*, **97**, 233-244.

Sudakin, V., Shteinberg, M., Ganoth, D., Hershko, J. and Hershko, A. (1997) Binding of activated cyclosome to p13suc1: Use for affinity purification. *J Biol Chem*, **272**, 18051-18059.

Tomoda, K., Kubota, Y. and Kato, J. (1999) Degradation of the cyclin-dependent-kinase inhibitor p27Kip1 is instigated by Jab1 [see comments]. *Nature*, **398**, 160-165.

Visintin, R., Hwang, E.S. and Amon, A. (1999) Cfi1 prevents premature exit from mitosis by anchoring Cdc14 phosphatase in the nucleolus . *Nature*, **398**, 818-823.

Yang, J. and Kornbluth, S. (1999) All aboard the cyclin train: subcellular trafficking of cyclins and their CDK partners. *Trends in cell biology*, **9**, 207-210.

Appendix I: A brief account of Protein crystallography

CONTENTS

LIST OF FIGURES **272**

7.1 INTRODUCTION **272**

7.1.1 Basic terminology *273*
7.1.2 The art of protein crystallisation *274*
7.1.3 Diffraction of X-rays by crystals *276*
7.1.4 X-ray sources *279*
7.1.5 X-ray detectors *281*
7.1.6 Cryocrystallography *283*
7.1.7 X-ray data collection *285*
7.1.8 The concept of the reciprocal space and lattice *289*
7.1.9 The phase problem *290*
7.1.10 The isomorphous replacement solution(single or multiple) *290*
7.1.11 The molecular replacement solution *291*
7.1.12 The trivial case of phase transfer *291*

7.2 REFINEMENT OF MACROMOLECULAR STRUCTURES **292**

7.2.1 The need to refine the nascent model *292*
7.2.2 Common pitfalls of macromolecular refinement *294*
7.2.3 Some of the current solutions for overcoming the pitfalls in protein refinement *295*
7.2.4 Rigid body refinement *297*
7.2.5 Positional refinement *298*
7.2.6 Torsion angle molecular dynamics refinement *298*
7.2.7 Simulated annealing *298*
7.2.8 Temperature factor refinement *300*

7.3 CALCULATION AND INTERPRETATION OF THE ELECTRON DENSITY 300

7.3.1 Qualitative analysis of macromolecular structures *301*

7.4 CRYSTALLOGRAPHIC PROGRAMS USED 303

7.5 LITERATURE CITED 304

List of figures

Figure 7.1 The unit cell ---------- *273*
Figure 7.2 A typical phase diagram based on variable protein and precipitant concentration ---------- *274*
Figure 7.3 Schematic representation of two commonly used crystallisation methods ---------- *276*
Figure 7.4 Geometrical construction of Bragg's law ---------- *278*
Figure 7.5 Schematic representation of the atomic energy levels for copper ---------- *280*
Figure 7.6 Typical experimental set-up for a cryocrystallographic data collection ---------- *284*
Figure 7.7 The effects of crystal alignment on the required oscillation angle to avoid overlapping reflections ---------- *285*
Figure 7.8 X-ray data collection ---------- *288*
Figure 7.9 Schematic diagram of the simulated annealing refinement process ---------- *299*

7.1 Introduction

An X-ray protein crystallography experiment can be divided simply into a "wet" and "dry" work phase. The former is called "wet" because it encompasses procedures that are performed on the laboratory bench i.e. gene manipulations, protein expression, purification and crystallisation. After growing suitable crystals, the experiment progresses into the "dry" phase, which is mostly computational. In this phase, a single crystal is exposed to an X-ray beam and the pattern of diffraction is recorded on an X-ray detector. Coaxing a meaningful atomic model from a collection of diffraction spots involves the use of a plethora of computer programs plus a lot of skill from a protein

crystallographer. In addition to the various statistical checks that are used to monitor the correct processing of the X-ray data, it is important that the final atomic model agrees with biochemical data unless there are good scientific reasons (for example, incorrect biochemical models) to do otherwise. In this appendix, I present a simplified account of protein crystallography that I, as a non-mathematical biochemist, have found useful to the understanding of the theory behind some of the techniques employed in Chapters 2 and 3 of this thesis. This section has been adapted from an introductory text (Rhodes, 1993) to the subject for biochemists and molecular biologists. Specialist and rigorous accounts of protein crystallography can be found in any of the following texts (Blundell and Johnson, 1976; Carter and Sweet, 1997; Drenth, 1994).

7.1.1 Basic terminology

A crystal is a solid consisting of a highly ordered arrangement of many millions of copies of a basic unit (e.g. a protein molecule). It is usually bounded by well-formed planes, which meet at edges and corners. The faces and edges of a crystal are normally arranged with some degree of symmetry. By definition, the unit cell is the basic building block upon which the entire volume of the crystal is built. In later sections of this appendix, it would become apparent that these features of crystals i.e. repetition, periodicity and symmetrical arrangements underpin the technique of crystallography.

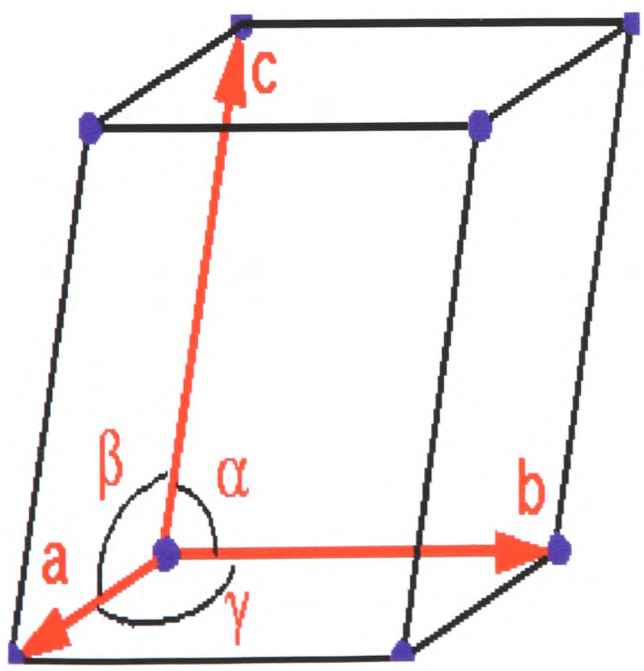


Figure 7.1 The unit cell

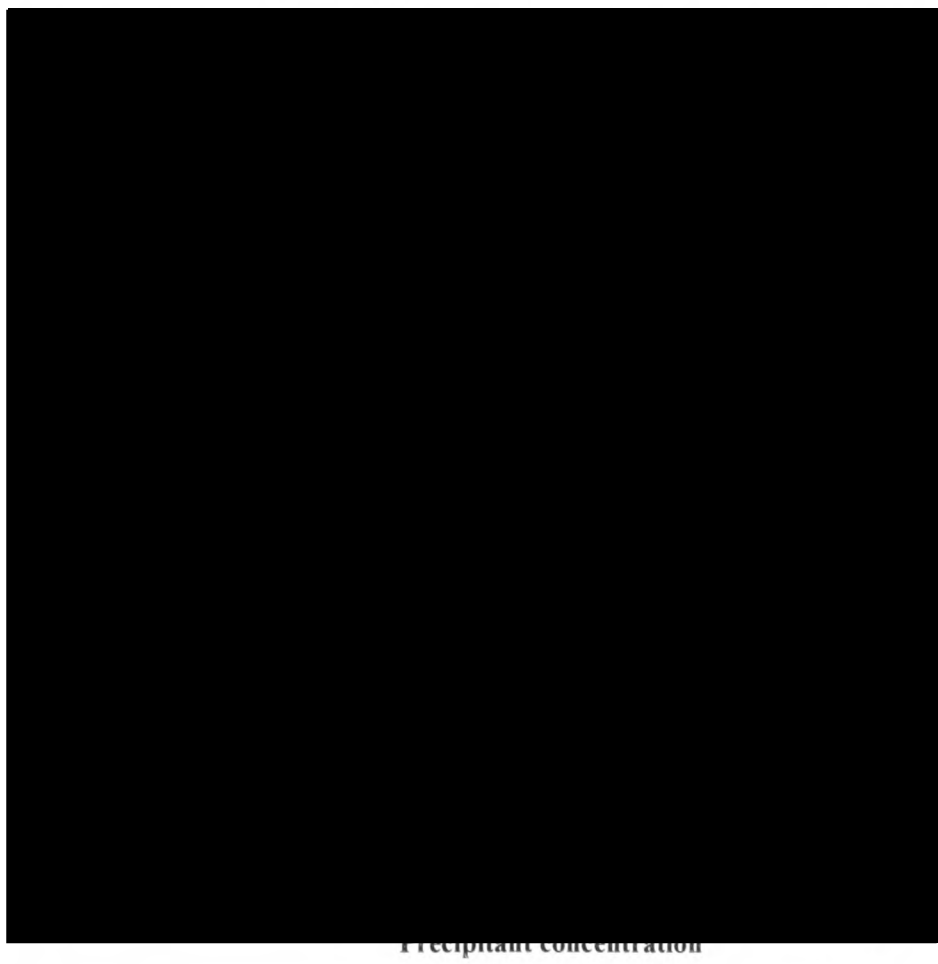
The unit cell has a shape which is described by the cell dimensions a, b, c. The latter are the lengths of the unit cell axes. The angle between the ***a*** *and* ***b*** *axes is given the Greek letter γ, that between* ***a*** *and* ***c*** *is called β and α denotes the interaxial angle between the* ***b*** *and* ***c*** *axes.*

The term crystal lattice refers to a regular three-dimensional array of points upon which the contents of the unit cell can be arranged by infinite repetition. Crystals exhibit symmetry properties that conform to one of 230 possible space groups. Protein crystals are composed of only L-amino acids; this feature means that they cannot exhibit mirror planes or centres of symmetry. This limits the number of space groups of protein crystals to 65. For example, Suc1 crystals belong to the hexagonal space group $P6_522$.

7.1.2 The art of protein crystallisation

Growing suitable protein crystals is arguably the most important step in any protein crystallography experiment, but it remains the least understood with regards to the factors that govern why certain proteins will crystallise but others do not. The process of crystallisation is a phase transition event (Chayen *et al.*, 1996). Protein crystals tend to grow from an aqueous solution that has been brought to supersaturation. Due to this feature, phase diagrams can be used to assist in the design of crystal growth conditions. Such a diagram would show which state (i.e. liquid, crystalline or amorphous, solid precipitate), would be stable under different crystallisation parameters. The protein crystallographer can use such a diagram to assess the effect(s) of factors such as the concentrations of protein, precipitant(s), additive(s), pH and temperature on the production of crystals (Chayen *et al.*, 1996).

Figure 7.2 A typical phase diagram based on variable protein and precipitant concentration



The figure shows four possible outcomes:

(a) a zone where the protein will just precipitate,

(b) a region where spontaneous nucleation will take place,

(c) an area of lower supersaturation just below the nucleation zone where crystals are stable and may grow but no further nucleation will take place. This zone is called the metastable zone and it is thought to contain the best conditions for growth of large, well-ordered crystals), (d) an unsaturated area where the protein is fully dissolved in the solution and therefore will never crystallise. Reproduced from (Chayen et al., 1996).

The target protein sample must satisfy the following criteria:

(i) Purity: It is vital that the target protein is very pure (>90%). Experience shows that apart from being chemically pure, the protein in solution must be a homogeneous species i.e. a mixture of unfolded and correctly folded protein would fail to crystallise. A mixture of different oligomerisation states (e.g. monomers and dimers) will also fail to crystallise. In general, the purification protocol normally ends with a gel-filtration chromatography step so that the oligomerisation state of the protein sample can be ascertained prior to setting up crystallisation trials. Some of the methods for assessing chemical and conformational purity are: SDS gel electrophoresis (presence of a single band on the gel), isoelectric focusing gel electrophoresis (for detecting charge heterogeneity), electrospray mass spectrometry (for an accurate assessment of the mass and purity of the sample) and dynamic light-scattering (to check the state of aggregation of the sample).

(ii) Concentration: Initially, the crystallographer requires about a milligram of purified protein at a concentration of 1-20mg/ml to begin the search for conditions under which the protein will crystallise. Once these conditions have been determined, it is necessary to hone in on those conditions in order to grow bigger and better (sharp edges) crystals.

The crystallisation process can be divided into two phases called nucleation and growth. The former is a pre-requisite for the latter. Crystals grow around pre-formed nuclei. The nucleation phase demands a higher degree of supersaturation than the growth phase. The parameters that can influence the crystallisation process are the concentrations of precipitant, protein, additives, pH and temperature. The crystallisation solution containing the target protein can be brought to a state of supersaturation by varying these

parameters. Currently, the standard method for bringing a crystallisation solution to a state of supersaturation is either by hanging- or sitting-drop vapour diffusion (Fig. 7.3).

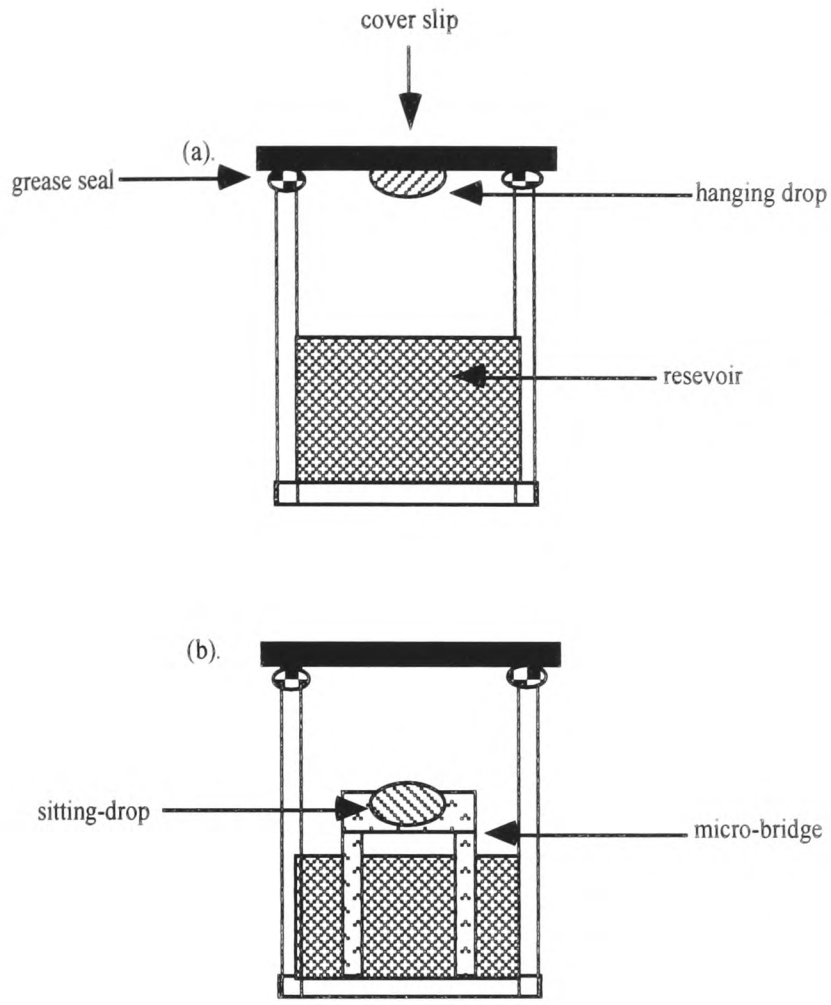


Figure 7.3 Schematic representation of two commonly used crystallisation methods

(a) The hanging-drop and

(b) Sitting-drop vapour diffusion methods.

At supersaturation, a nucleus may form around which the crystal would grow. Instead of the desired single crystal, one often gets thousands of small crystals after the nucleation and growth phases. In theory, because these two phases can be separated for some target proteins, the crystallographer has a degree of control and can direct the crystallisation trial towards the desired outcome, but in practice each step of the crystallisation experiment can be fraught with insurmountable and inexplicable obstacles.

7.1.3 Diffraction of X-rays by crystals

After growing protein crystals of appropriate quality and size, diffraction data can be collected by mounting a single crystal in a beam of X-ray radiation with an X-ray detector positioned behind the crystal to record the diffraction pattern. The goal of the crystallographer is to obtain an image of the contents of the unit cell in atomic detail. The

aim is to take a photograph of the protein molecules in the crystal akin to how one would capture an image of a yeast cell with a light microscope. The latter captures the image of the yeast cell when light reflected from each of the millions of molecules that constitute the cell is focused by a lens. We do not see the individual molecules of the yeast cell because the reflection from a molecule is too weak, and also the dimensions of a molecule are too small - compared to the wavelength of light (400-700nm) - to allow for the resolution of two individual atoms. But a composite of the reflections from all the molecules is strong enough to be focused to recreate an image of the yeast cell. In protein X-ray crystallography, the electron cloud of each atom (carbon, nitrogen, oxygen etc) in the protein is responsible for scattering the incident X-ray beam. So, if we could place an appropriate lens behind the diffracting crystal (X-ray microscope), we could focus the reflected rays onto a detector and re-create an image of the electron cloud of the entire protein. Alas, such a lens is yet to be discovered, so the protein crystallographer has to use indirect methods to obtain the information required to re-create the overall electron cloud of a protein molecule in a single protein crystal. In this case too, the reflection by a single protein molecule is too weak to be detected, but the orderly and repetitive nature of the contents of the crystal means that each reflection is, in effect amplified millions of times to produce a strong, measurable spot on the X-ray detector. How the crystallographer achieves this feat can be explained by considering how a crystal diffracts X-rays. In 1913, Sir W.H. Bragg and his son Sir W.L. Bragg derived an equation to explain why crystals appear to reflect X-ray beams at certain angles of incidence to the crystal surface called theta (θ).

$$n\lambda = 2d \sin\theta \qquad (1)$$

where the variable ***d*** is the distance between atomic layers in the unit cell of the crystal, the variable lambda (λ) is the wavelength of the X-ray beam and ***n*** is an integer.

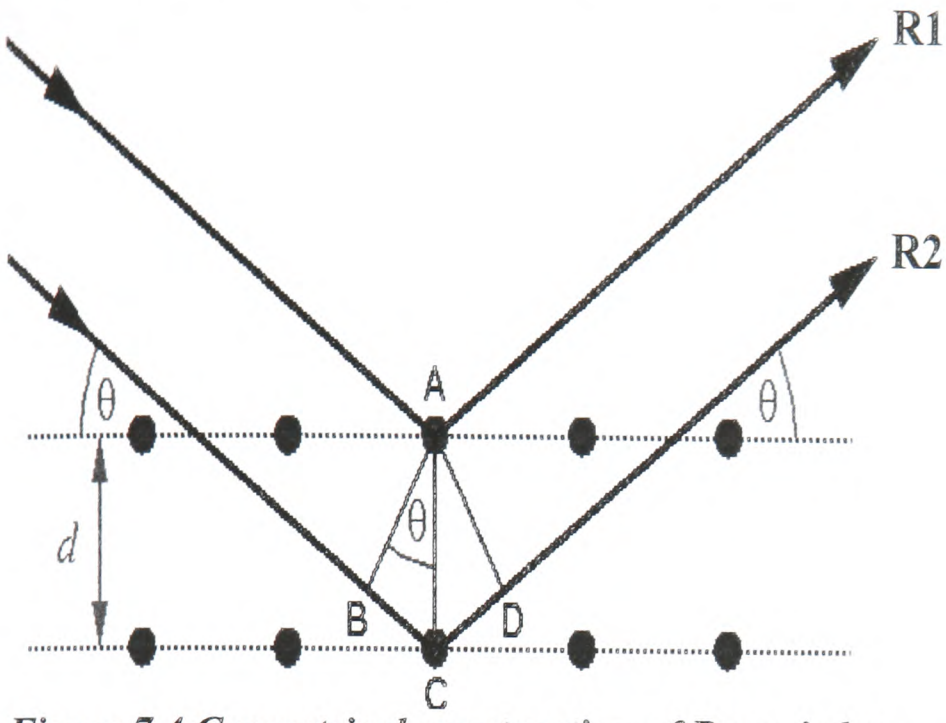


Figure 7.4 Geometrical construction of Bragg's law

The black dots represent a set of parallel planes (perpendicular to the page) of the unit cell. Two rays, R1 and R2 impinge and emerge from two sets of adjacent parallel planes at an angle θ. If we represent the interplanar distance by the letter ***d****, then the extra distance that R2 has to travel to reach the X-ray detector compared to R1 is BC + CD = 2d sin θ. BC + CD is the path difference between R1 and R2. If this path difference is an integral multiple of the wavelength then we get positive (constructive) interference between the two waves and hence a strong reflected beam impinges on the detector. In the latter case, the Bragg condition has been satisfied. If the path difference is not an integral multiple of the wavelength, R1 and R2 are out of phase and therefore interfere destructively i.e. they cancel each other out and this leads to an absence of a spot on the detector. So, a diffraction pattern is the consequence of the superposition of waves when radiation (X-rays, visible light, etc.) are scattered by objects that are the same order of size as the wavelength of the radiation. The superposition results in constructive and destructive interference and this leads to regions of brightness and darkness on an appropriate detector. The angle (θ) at which one observes such a pattern is called the Bragg angle and it is inversely related to the interplanar distance* ***d****. This means that crystals with large unit cells would give small angles of diffraction so that many reflections will fall within a range where they can be conveniently measured. Conversely, crystals with small unit cells would give large angles of diffraction and therefore if the size of the detector remains the same, fewer reflections will be measured.*

7.1.4 X-ray sources

After growing suitable crystals, the two main requirements for collecting X-ray data are an X-ray source and a detector. X-rays are electromagnetic radiation with wavelengths of 1000-0.1Å i.e. 10^{-7} - 10^{-11}m. The two main sources of X-rays for protein crystallography are a laboratory source (e.g. a rotating anode generator) and a synchrotron source. In a rotating anode generator, the generation of X-rays involves the acceleration of electrons in a vacuum by an electric field. These high-speed electrons are directed against a metal target (usually copper i.e. the anode) and the subsequent multiple collisions rapidly slow down the electrons. After each impact with the target metal, the energy of motion of the electron is converted mainly into heat. For this reason, the target tends to be a high-melting point metal and it is cooled continually. A fraction of the fast-moving electrons lose most of their energy in single collisions to atoms in the target metal. The target atoms emit this energy in two different ways: as a continuum of X-rays and as sharp peaks at specific wavelength. The sharp intensity peaks are characteristic of the particular target metal used. If the anode is a metal with an atomic number of about 30, the electrons impinging on the focal point of the metal target eject electrons out of the inner K-shell (fig. 7.5). Electrons coming down from higher energy levels (e.g. L and M-shells) fill the vacancy. The energy emitted by the descending electrons gives two nearly monochromatic lines, K_α and K_β, respectively.

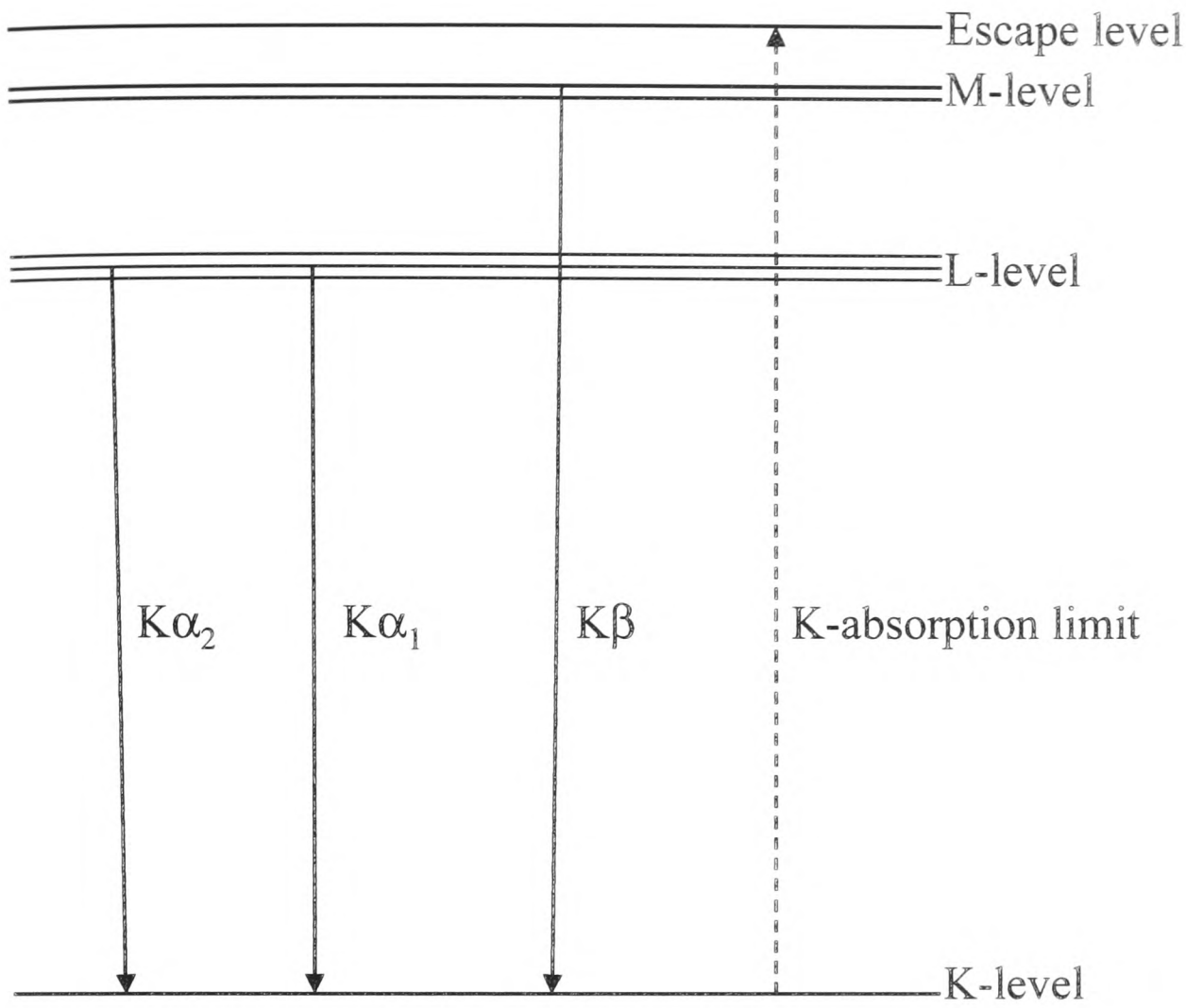


Figure 7.5 Schematic representation of the atomic energy levels for copper
The transitions that cause characteristic X-ray emissions are shown.

In a rotating anode generator, the anode is a rotating water-cooled copper cylinder.

In the past two decades, a hitherto unused by-product of particle accelerators called Synchrotron radiation has become a widely used tool in macromolecular crystallography (Chayen *et al.*, 1996; Helliwell, 1998; Mitchell *et al.*, 1999). When electrons are accelerated to velocities close to the speed of light along orbits varying in radii from a few to several hundred metres, they emit an intense beam of X-rays. This is exploited by macromolecular crystallographers to collect diffraction data on crystals which, for various other reasons, would have been difficult using conventional laboratory X-ray sources. The applications of it include the measurement of the highest possible resolution data, the study of large unit cells, collection of data from small crystals (e.g. Suc1 crystals), the exploitation of anomalous dispersion effects for phase determination and time-resolved studies. All the X-ray data used to solve the Suc1 structures in chapters

2 and 3 were collected at Synchrotron radiation sources. The main advantages of Synchrotron radiation over conventional laboratory sources of X-rays include high flux i.e. large number of photons, high brightness i.e. a well collimated beam, high brilliance i.e. small source size and good collimation, tunability i.e. the ability to select a particular wavelength, and the polarised nature of the beam. Since the Suc1 crystals were of relatively small size (20 x 20 x 100μm^3), some of the properties of Synchrotron radiation (e.g. high intensity) were particularly suited to the collection of X-ray diffraction data from these crystals. For the rotating anode generator, only the characteristic lines of copper are sufficiently intense enough to be used to collect protein X-ray data. Synchrotron radiation lacks this limitation since it produces an intense white beam from which any wavelength can be chosen with the help of a monochromator. This feature of Synchrotron radiation underpins structure determination techniques such as Multiwavelength Anomalous s Dispersion (MAD).

7.1.5 X-ray detectors

The simplest X-ray detector is the X-ray film. After an appropriate chemical development of an X-ray film, positions that have been impinged by incident X-rays are marked by dark spots. An optical scanner can be used to measure the precise position and intensity of each spot and this information can be relayed in a digital form to a computer for subsequent analysis. The advantage of an X-ray film is that it has a very high spatial resolution but this outweighed by its limitations which include the inability to automate the data collection process and a small dynamic range necessitating packs of film to provide an adequate range for the weak and high intensity spots. This leads to an increase in the amount of processing and the risk of systematic errors due to film scaling within the packs. Also, the background noise level caused by chemical fog can be relatively high and this results in a poor signal-to-noise ratio thereby rendering the weak high-resolution reflections difficult to utilise. For these reasons, X-ray film has been almost completely superseded by other detectors. Another device for recording X-ray data is the Scintillation counter. It counts X-ray photons so it can give an accurate measure of the intensities over a wide range. But the Scintillation counter is limited to the counting of one reflection at a time. This makes them impractical to protein

crystallography where hundreds to thousands of reflections are produced at a particular time. These days modern detectors provide an immediate output of the two-dimensional diffraction pattern containing hundreds of spots in a digital format, thus eliminating the wet developing and digital scanning that come with films (Gruner, 1994). Also, modern detectors combine the accuracy and wide range of Scintillation counters with the simultaneous measurement of many reflections by X-ray film, hence the name area detectors. A good area detector must have the following attributes: uniform sensitivity to incident X-rays, a well-defined spatial distortion to ensure reliable calibration, the spatial resolution, active area and dynamic range (i.e. the range of intensities that can be recorded) must be adequate and it must have a high detective quantum efficiency (DQE) (Garman, 1996). The latter refers to the number of incident photons that must be recorded by the detector. All the X-ray data analysed in Chapters 2 and 3 of this thesis were collected on image plate area detectors. These are the commercially available MAR Research (Norderstedt, Germany) image plates called Mar30 (100K Suc1 and Ser13Gly) and Mar34.5 (Ala32Gly). The numbers refer to the diameter of the detector in centimetres. Image plates consist of a veneer of an inorganic storage phosphor on a flat, plastic base. Incident X-ray photons excite electrons in the material to higher energy levels. The half-life of this metastable state is about 10 hours. A portion of this energy is emitted as normal fluorescent light in the visible wavelength region. The rest can be released as violet luminescence ($\lambda \approx 390$nm) on illumination by a red He-Ne laser ($\lambda \approx 633$nm). Since the emitted light is proportional to the number of photons absorbed by that particular area of the image plate, the diffraction image can be read out by scanning the image plate with a laser beam coupled to a photomultiplier system. The plate can then be erased by exposure to intense white light and reused to collect the next diffraction image. The advantages of image plates over X-ray film are: firstly that image plates are about 10 times more sensitive to X-rays and secondly that they have a much wider dynamic range (1: 100000). This means that the entire range of reflections - from strong to weak - can be collected with a single exposure on a single image plate. This feature of image plates, coupled with the ability to erase and re-use the same plate repeatedly, allows for the automation of the data collection process, thus reducing both the time the protein spends in the X-ray beam and also the time the experimenter spends collecting and digitising the

data. Other area detectors currently used in macromolecular crystallography are multi-wire gas proportional chambers and charge coupled devices (CCDs).

7.1.6 Cryocrystallography

As mentioned in the introduction to Chapter 2 of this thesis, when crystals of biological macromolecules are subjected to irradiation from an X-ray beam, they tend to deteriorate due to the effect of radiation induced decay. The damage is largely inflicted by the generation and diffusion of free radicals within the crystal (Blundell and Johnson, 1976). Crystals of biological macromolecules are especially prone to this sort of damage because of the high (30-95%) solvent content (Matthews, 1968) and the potential for diffusion within the crystal (Garman and Schneider, 1997). Some of the advantages of collecting X-ray data at cryogenic temperatures have already been noted (see chapter 2). First, the dramatic decrease in radiation damage allows for the collection of complete data sets from a single crystal. Second, depending on the extent of dynamic and static disorder within a particular protein crystal, the reduction in atomic thermal vibration at cryogenic temperatures often results in the extension of the resolution limit of diffraction. Third, employing cryogenic techniques removes the need to mount the crystals in a capillary tube, thus precluding background X-ray scatter from excess mother liquor and the glass/quartz capillary tube. The resultant improvement in the signal-to-noise is critical for experiments on weakly diffracting crystals that hitherto, would have been unsuitable for detailed structural studies. Fourth, it is possible to collect X-ray data on fragile crystals using cryocrystallographic techniques as opposed to room temperature techniques because the former requires less manipulation of the crystals. Finally, cryocrystallography affords the experimenter the freedom to screen for suitable crystals at the home laboratory and harvest them when they at their peak condition by freezing them in liquid nitrogen whilst waiting for a data collection time at a Synchrotron source (Chayen *et al.*, 1996; Garman and Schneider, 1997). The arrangement for collecting a cryo-data set is depicted in the diagram below.

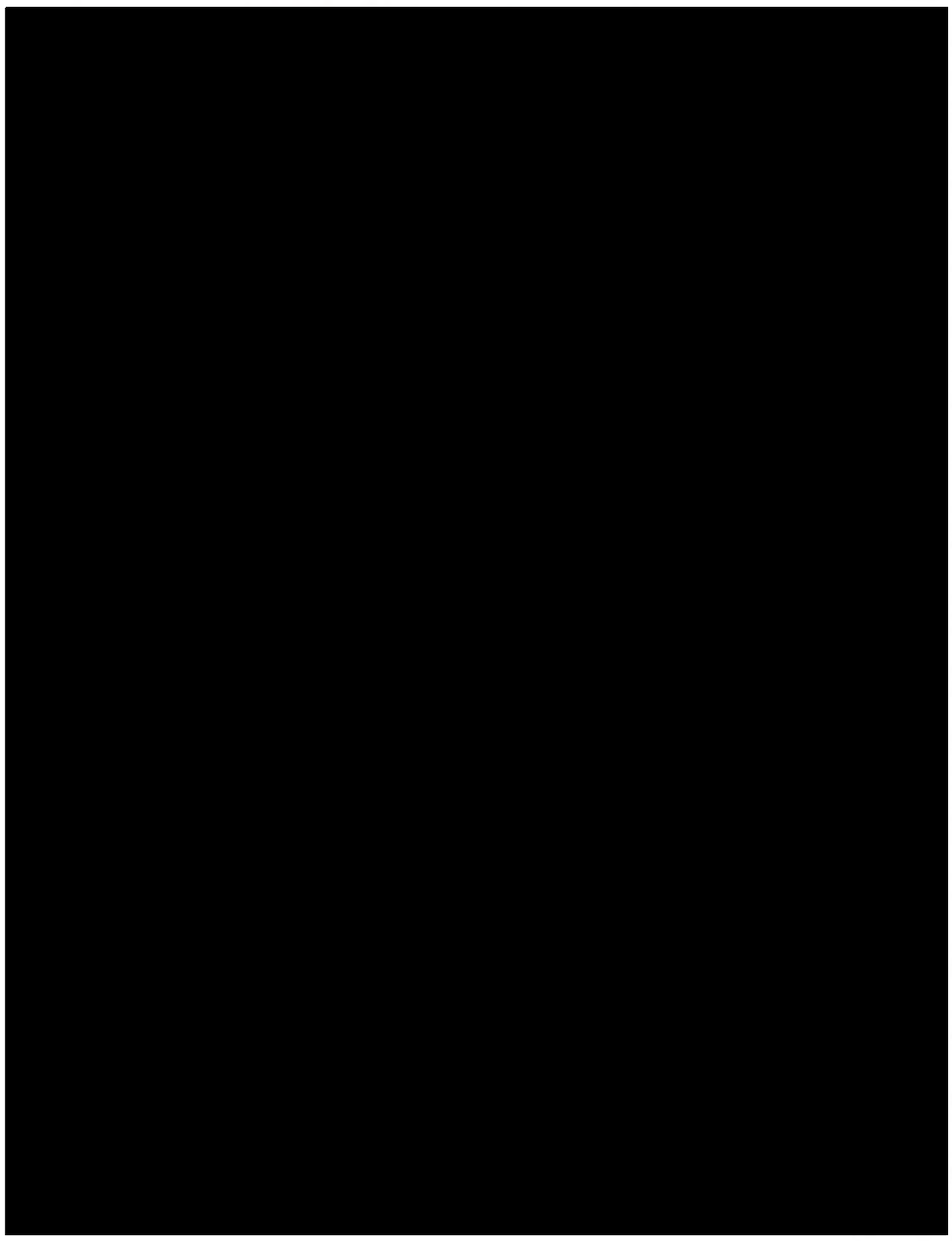


Figure 7.6 Typical experimental set-up for a cryocrystallographic data collection

In this arrangement, a magnetic rubber disc is pierced with a stainless steel pin; the top of the disc is rounded and it has the same diameter as a hole in the base of the pin holder or "top hat". The bottom of the disc fits into the hole in the goniometer head. The magnet must be strong enough to make a firm connection but weak enough to allow the experimenter to manoeuvre the top hat as easily as possible. Reproduced with permission from Dr. Elspeth Garman.

In practice, the macromolecular crystal is first prepared for flash freezing by screening for appropriate solutions (so-called cryo-protectants) that would prevent the damage caused by the formation of crystalline ice if an untreated protein crystal is frozen at cryogenic temperatures. The cryo-protectant forms vitreous amorphous ice instead and thus avoids the mechanical stress which would be suffered by the crystal due to the 9% increase in specific volume that accompanies the phase transition from water to hexagonal ice (Garman and Schneider, 1997). The role of the cryo-protectant is to reduce the crystallisation rates of both the water within the crystal and that within the surrounding solvent. The effect is to minimise any disruption of the crystal by ice-formation during the cooling procedure.

7.1.7 X-ray data collection

After a successful mounting of a single protein crystal in the X-ray beam, the diffraction data are commonly collected in a so-called oscillation mode. With this method, the crystal is rotated over a small angle (0.1 - 1.5°) about an axis perpendicular to the X-ray beam for a period of time from as short as 1 second to as long as 40 minutes. The first image or frame is recorded and the contiguous same small angle is collected as the next frame, till the range of angle through which the crystal must traverse in order to collect as complete a data set as possible has been reached. The angular range required for the collection of a complete data set is determined by the crystal symmetry or space group, unit cell dimensions and the geometry of the experiment i.e. crystal orientation relative to the beam and detector (Garman, 1996). For the relatively high symmetry space group $P6_522$ hexagonal Ala32Gly crystals, with unit cell dimensions of a=b=59.3Å and c= 264.4Å, using an oscillation angle of 0.5° can give (depending on the oreintation of the crystal with respect to the beam) a 96.3% complete data set after an angular range of only 33.5°.

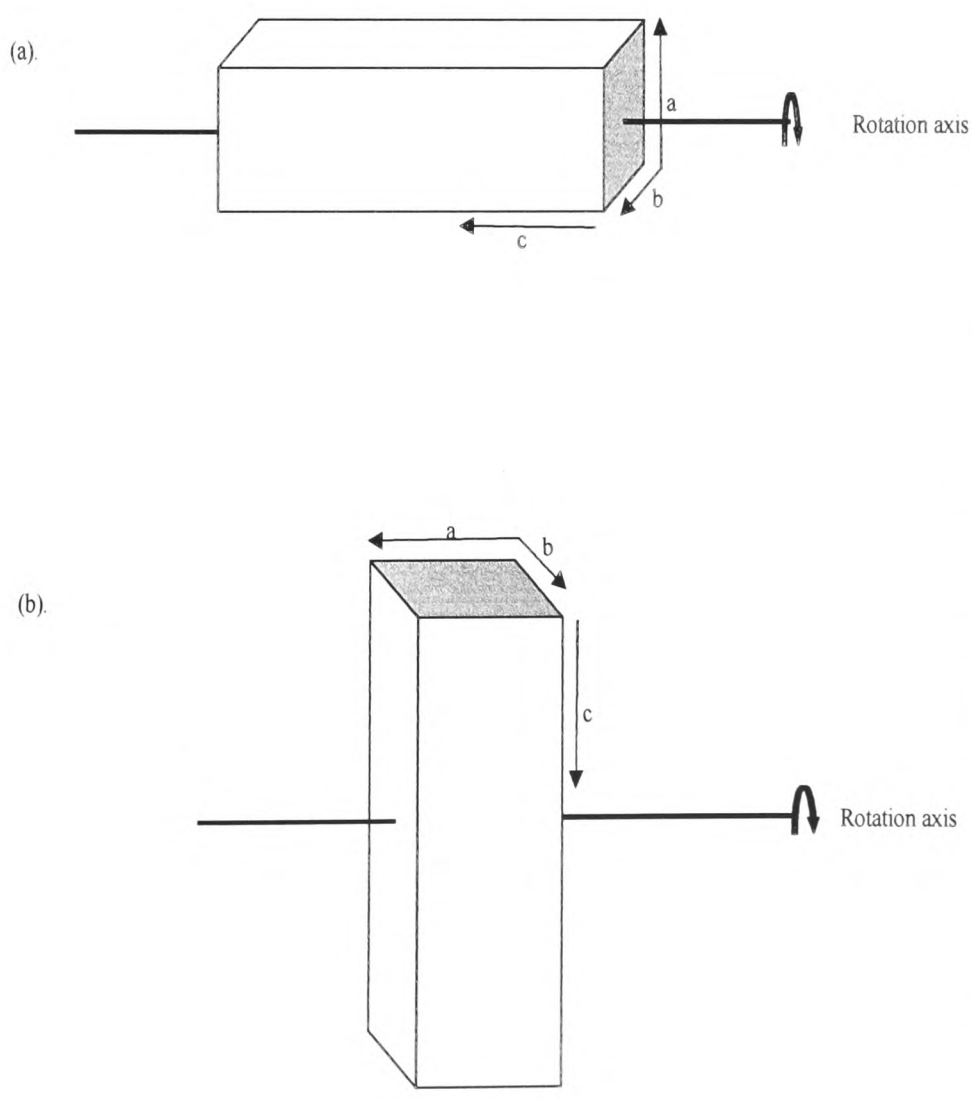


Figure 7.7 The effects of crystal alignment on the required oscillation angle to avoid overlapping reflections

Consider the case of Suc1 crystals with a=b=60Å and c=265Å. In case (a) where the crystal is aligned with the longest cell axis, ***c****, along the rotation axis, if the experimenter wants to collect a 2Å data, the formula* $\Delta\phi_{max} = d_{max} / C_{max}$ *(where* ϕ_{max} *is the maximum*

oscillation angle, d_{max} is the resolution of the data and C_{max} is the largest unit cell edge in the direction normal to the beam) gives an oscillation angle $\Delta\phi_{max}$ of 2/65 radians (≈ 1.8°). In case (b) where the crystal is mounted so that the cell axis becomes perpendicular to the rotation axis and the longest cell length becomes the determining factor for the maximum oscillation angle, $\Delta\phi_{max}$ = 2/265 radians (≈ 0.4°). In this illustration, the effect of mosaicity has been ignored.

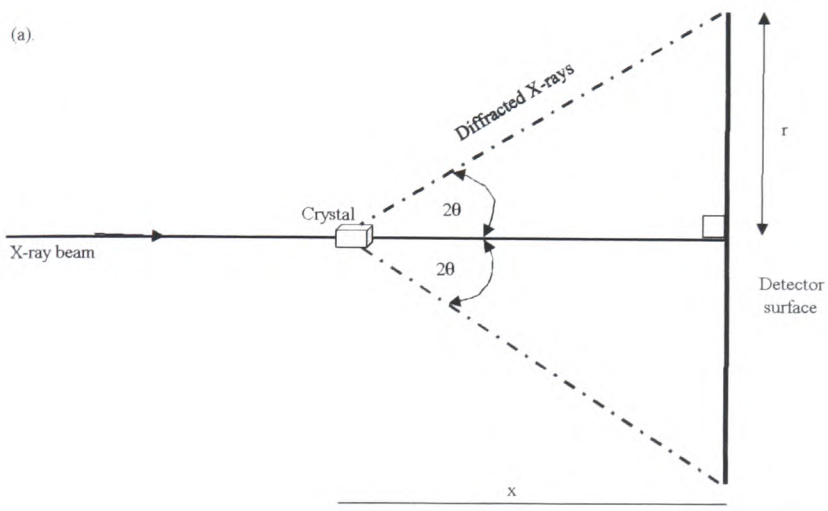


Figure 7.8 Data collection geometry *(a) Illustration of the geometry that determines the resolution of the data. (b) Construction of the Ewald sphere. This is the geometric expression of Bragg's law. In this construction, the protein crystal is at the centre* ***C*** *of a sphere of radius* ***1/λ***. *The vector* $\mathbf{k_o}$ *is the direction of the incident beam.* ***O*** *is the origin of the reciprocal lattice of the sample. Diffraction will occur in directions for which a point of the reciprocal lattice intercepts the Ewald sphere, such as* ***P***. *The vector* ***CP*** *and the scattering vector* ***OP*** *give the direction of the diffracted beam. When this condition is satisfied, a diffraction image like those depicted in the next figure (fig. 7.8) is recorded and the crystal is then rotated for the next reflecting sphere of lattice planes to intercept the detector and hence be recorded. The sphere of reflecting lattice planes is a property of the particular crystal whilst the diameter of the sphere is a property of the wavelength of the X-ray beam.*

In practice, the crystallographer collects the first diffraction image and from that, using auto-indexing software, one can obtain the parameters such as the space group and unit cells dimensions. These and other parameters such as the wavelength and crystal-to-detector distance are then fed into a computer program such as PREDICT (M.E.M Noble, unpublished work) from which one can obtain an output giving a recommended oscillation angle and the total angular range required to collect a complete data set.

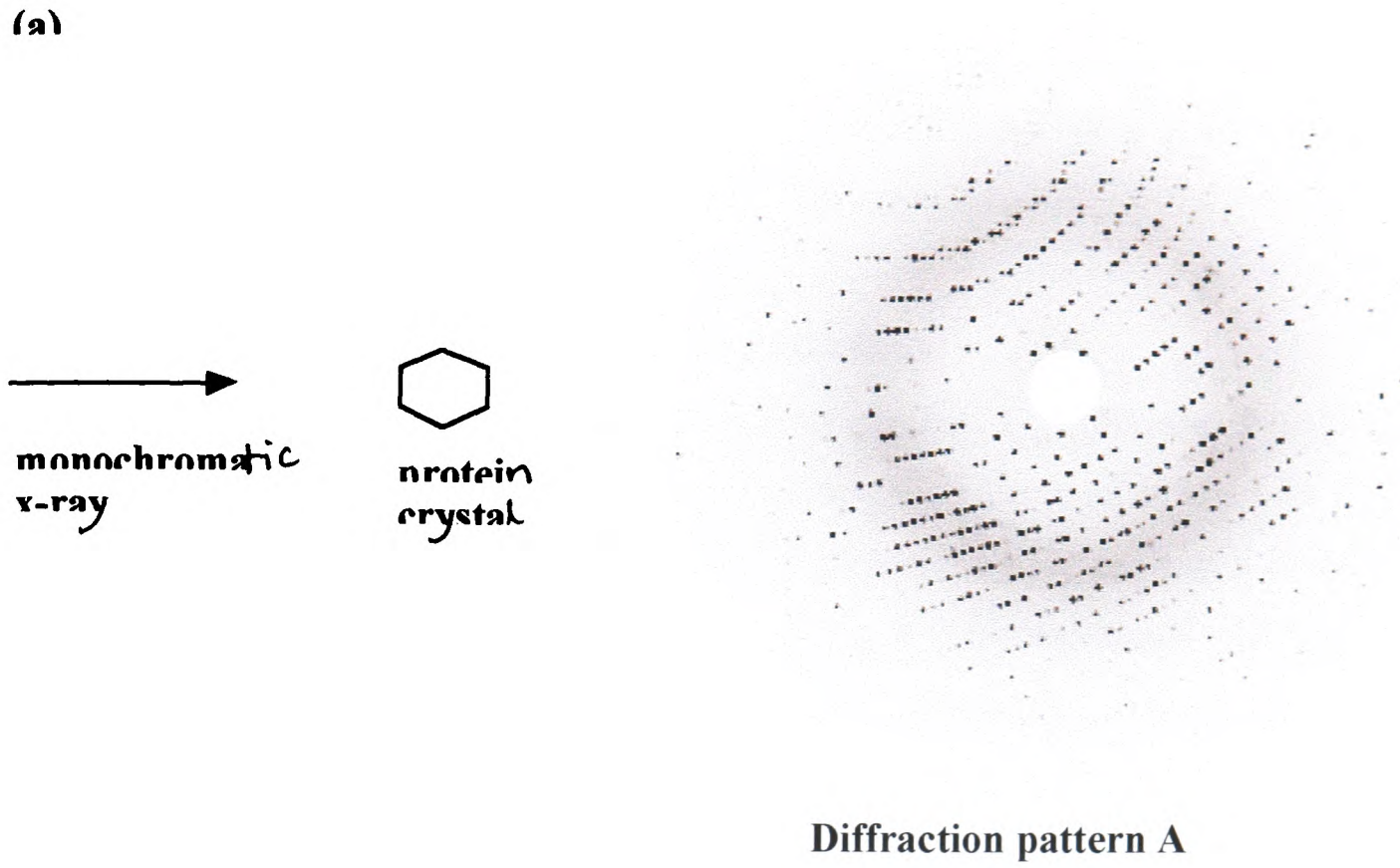


Figure 7.8 X-ray data collection

In practice, the mounted protein crystal is rotated in the beam so that the Bragg condition is satisfied for most of the sets of planes in the unit cell. Bragg's law allows us to consider the diffraction pattern as a reflection of an incident X-ray from a set of parallel planes that make up the unit cell of a crystal, hence the use of the term

reflections for the spots that one sees on the X-ray detector. The usefulness of Bragg's law is that it tells the crystallographer the precise conditions under which a strong diffraction spot will impinge on the X-ray detector. The purpose of data collection is to record as many of these spots as possible for subsequent analysis.

7.1.8 The concept of the reciprocal space and lattice

In the course of determining the crystal structure of a protein, the crystallographer shuffles between two co-ordinate systems as a result of the need to work in the three-dimensional environment of the unit cell (direct space and lattice) and a second three-dimensional environment of the diffraction pattern (reciprocal space and lattice). The position of an atom of a protein molecule in the unit cell is assigned the co-ordinates ***x, y, z*** with respect to an origin (000 for x=0, y=0 and z=0). This origin can be a vertex of the unit cell. The position of the reflection in reciprocal space is assigned co-ordinates ***h, k, l*** with respect to an origin, which is the centre of the diffraction pattern. According to Bragg's law, the distance between two reflections (***t***) is reciprocally related to the distance (***d***) between the unit cells. In order to compute the electron density, the amplitude and phase of each of the diffracting rays have to be measured The amplitude $\mathbf{F_{hkl}}$ can be derived for the intensity $\mathbf{I_{hkl}}$ of the diffracted rays $\mathbf{I_{hkl}} \alpha \mathbf{F^2_{hkl}}$ but the phase cannot be measured directly. The scattering from a single atom (j) in the unit cell can be represented as:

$$f_{hkl} = f_j e^{-Bj\sin 2\theta/\lambda 2} e^{2\pi i\,(hx_j + ky_j + lz_j)} \qquad (2)$$

where $\mathbf{f_j}$ is a form factor describing the diffracting power of each atom as a function of the Bragg angle and **B** is a temperature factor describing the motion of the atom about its mean position, ***x, y, z***.

This can be written for all atoms in the unit cell as;

$$F(hkl) = \sum_{j=1}^{N} f_j e^{-Bj\sin 2\theta/\lambda 2} e^{2\pi i(hxj+kyj+lzj)} \qquad (3)$$

Written in terms of an amplitude and a phase,

$$F(hkl) = |F(hkl)|e^{i\phi(hkl)} \qquad (4)$$

The measured entity (intensity) is related to the structure factor by

$$I(hkl) = |F(hkl)|^2 \qquad (5)$$

So, the crystallographer is burdened with the task of recovering the missing entity (ϕ) in order to obtain the structure factor for all the atoms in the unit cell.

7.1.9 The phase problem

As seen from the preceding section, the structure factor can be described in terms of an amplitude and a phase. If both the amplitude and phase are known, the cloud of electrons surrounding each atom within the unit cell can be reconstituted from the diffraction pattern using a mathematical technique called a Fourier synthesis. The immediate goal of a protein crystallography experiment is to calculate the electron density surrounding each atom of the protein molecule. The mathematical formula for calculating the electron density is:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l F(hkl) e^{[-2\pi i(hx + ky + lz)]} \qquad (6)$$

where **V** is the volume of the unit cell.

This equation needs the phase information for each reflection. Its determination constitutes the so-called phase problem of protein crystallography. The widely used techniques for recovering the lost phase are the isomorphous replacement, multiwavelength anomalous dispersion (MAD) and the molecular replacement methods.

7.1.10 The isomorphous replacement solution (single or multiple)

Although crystals are by definition solids, protein crystals are distinguished by the fact that they contain a high proportion (30-95%) of solvent (Matthews, 1968). One can envision a picture of millions of protein molecules immobilised in a pool of solvent. This feature allows the crystallographer to react the protein molecules in the unit cell with appropriate reactants or substrates. The crystal is normally soaked in a solution of a heavy metal and the product of this reaction is called a derivative. In reacting the protein

crystal with a heavy metal, a few atoms of the heavy metal might bind to certain specific sites of the protein molecule within the unit cell without altering the dimensions of the unit cell. When this occurs, the derivatised crystal is said to be isomorphous with the native (underivatisd) crystal. A heavy metal such as mercury is chosen because it has more electrons than the atoms that constitute amino acids and hence it is a strong scatterer of incident X-rays. If we represent the structure factor of the heavy atom as **FH** and that of the protein in the native and derivative crystals as **FP** and **FPH** respectively, then it can be shown mathematically that **FH** = **FPH** - **FP**. The crystallographer can measure the amplitudes, |**FPH**| and |**FP**| and hence obtain an estimate of the isomorphous difference |**FPH**| - |**FP**|. The co-ordinates of the heavy metal can be determined by a mathematical technique called difference Patterson maps. This provides **FH**, thus making it possible to calculate **FP**. In practise, this method can be beset with teething problems such as non-isomorphism, weak substitution and hence negligible scattering by the heavy atom, variations in native and derivative crystals and lack of appropriate bindings sites within the protein molecule for the heavy atom.

7.1.11 The molecular replacement solution

As more and more atomic structures are deposited in the protein data bank, there is an increasing likelihood that an unknown structure may have a similar structure to one that has already been solved. This is especially true of homologous proteins from different species e.g. human Ckshs1 and *S. cerevisiae* CKS1 have similar tertiary structures (Khazanovich *et al.*, 1996). When two proteins have similar or identical tertiary structures, it is possible to use the known atomic structure of one molecule to calculate the initial phases of the unknown structure. This is the essence of the molecular replacement technique.

7.1.12 The trivial case of phase transfer

A variation on the above technique is the case where the phases of an existing protein model are assumed to be identical to the phases of an unknown protein molecule. This can occur in three instances: first, when a higher resolution data set is collected for a previously determined structure as is the case of the subject of chapter 2; second, when a

data set is collected for a point mutant of a previously determined wild-type protein as is the case of the subject of chapter 3 of this thesis and finally, when a data set is collected for a protein-small ligand complex and there is already in existence the atomic model of the native protein. In all of these cases, the underlying assumption is that the diffraction patterns of each pair would be so similar that one can use the calculated phases of the known structure as a starting point for obtaining relatively crude phases for the unknown structure. In a sense, this technique and the molecular replacement method are the reverse of the journey from diffraction pattern to protein model via an electron density map. The mathematical facility that allows one to shuffle between these two destinations is called the Fourier transform. Once the initial phases have been determined for the unknown structure, a process of model adjustment (using a computer graphics program) and refinement of the crude phases tends to gradually move the model away from that of the starting model until the new structure is revealed. This involves alternate cycles of manual rebuilding and refinement of the model.

7.2 Refinement of macromolecular structures

7.2.1 The need to refine the nascent model

The X-ray structure of a macromolecule depends on the accuracy of the phases. Usually, the experimentally determined values are inaccurate. Subsequently, electron density maps phased with such poor initial phases tend to be of a poor quality. The atomic positions derived from such maps are frequently inaccurate. It has therefore become necessary to refine the atomic parameters as best as possible so that many important features of the macromolecular structure can be revealed. The aim of this process is to re-construct a model of the macromolecule which adequately explains the diffraction data whilst at the same time making physical, chemical and biological sense. During the process of refinement, the parameters of the model are altered to improve the agreement between amplitudes calculated from the model and the amplitudes obtained by measurement. Macromolecular refinement can be defined as "a chemically-restrained non-linear optimisation problem" (Adams *et al.*, 1997).

$$E = E_{chem} + w_{xray}E_{xray} \qquad (7)$$

where E_{chem} consists of empirical information about chemical interactions between atoms in the model. E_{chem} is a function of all atomic positions and includes information about both covalent (bond lengths, bond angles, torsion angles, chiral centres and planarity of aromatic rings and non-covalent interactions like van der waals, hydrogen-bonding and electrostatic bonds. E_{xray} describes the difference between observed and calculated diffraction data. W_{xray} is a weight that is chosen to balance the forces arising from the E_{chem} and E_{xray} terms (Adams *et al.*, 1997).

Traditionally, the agreement between the $F_{observed}$ and $F_{calculated}$ has been used to monitor the progress of the refinement procedure. This value is called the R value. More recently, a new value called the Free R value has been added to the criteria (Brunger, 1992a).

$$R_{work} = \frac{\sum_{hkl} \| F_{obs} - F_{calc} \|}{\sum_{hkl} | F_{obs} |}$$

R_{free} is equivalent to R_{work} for a disjoint 5-10% of the total reflections not used in the refinement.

The R_{free} value provides an independent assessment of the refinement. At the beginning of the refinement procedure, the diffraction data are split into two sets: a large working set of about 90% of the total available data, and a test set of about 10%. The test set is then used to calculate the R_{free} value (Brunger, 1992a) to give an unbiased indication of the model quality. The working set is used in the refinement procedure whilst the test set is not. At the end of the refinement procedure, it is expected that the R_{work} and R_{free} values would be about 20 and 26, respectively. Values greater than 25% for the former and 30% for the latter ought to be justified.

Finally, the model must also meet certain objective criteria. The stereochemistry of the protein structure should be consistent with stereochemical parameters for known small molecule structure. The peptide dihedral angles are normally depicted pictorially in a Ramachandran plot (figs 2.5, 3.2, 3.3 of chapters 2 and 3). Deviations from ideal values require justification. Also, the agreement between the expected and refined geometries is

normally reported. The ideal root mean square (r.m.s) deviations for bond lengths should be about 0.05Å and that for bond angles should be about 2°. The final protein structure would have well-defined water molecules.

7.2.2 Common pitfalls of macromolecular refinement

In practice, macromolecular refinement can be beset with difficulties arising from the low data-to-parameter ratio frequently encountered in the refinement of protein molecules. For instance, for Suc1 crystals with approximately 2000 atoms in the asymmetric unit, the number of parameters (x, y, z coordinates, occupancy and B-factors) is about 10000 compared to the number of observations of about 12000 reflections, giving a data-to-parameter ratio of 1.2:1. This compares unfavourably with small molecules where the data-to-parameter ratio is typically 10:1 or better. This poses the danger of overfitting the diffraction data during macromolecular refinement. Currently, the solution to this problem is to increase the apparent ratio of data to parameters by incorporating prior chemical information such as bond lengths and bond angle restraints obtained from ideal values observed in high resolution, small, organic molecule structures. The assumption is that bond lengths and angles do not change appreciably from small molecules to macromolecules. The second pitfall in macromolecular refinement is that the nascent model is often a crude model because of the limited quality of the experimental phases and this introduces a discrepancy between the crude model and the actual structure of the crystallized molecule. The overall goal of the refinement process is to attain a global energy minimum for the model. This target function depends on several atomic parameters such as atomic coordinates, temperature factors and occupancies. This large number of adjustable parameters (e.g. ≈10000 for Suc1 crystals) results in a very complicated target function. This results in the so-called multiple minima problem (Adams *et al.*, 1997); the refinement process can fail by getting trapped in any of the many local minima that exist (in addition to the global minimum) in the target function. The fourth impediment to successful refinement is the phenomenon of model bias in the electron density maps (Hodel *et al.*, 1992). This refers to the situation where the inclusion of calculated phase information from a model (e.g. 293K Suc1) carries the risk of biasing the refinement process towards the current atomic model instead of

moving away from it to reveal the true identity of the contents of the unit cell (e.g.100K Suc1). Model bias can obscure the detection of errors in atomic models if the phase information is either insufficient or of a poor quality. Several specialized optimization protocols for overcoming the above problems during macromolecular refinement have been developed. Some of the methods used in the refinement of Suc1 are described below.

7.2.3 Some of the current solutions for overcoming the pitfalls in protein refinement

Some of the specialized optimization methods employed in chapters 2 and 3 of this thesis include rigid-body refinement, simulated annealing, conjugate gradient energy minimization or positional refinement, torsion angle molecular dynamics and temperature factor refinement. Until recently, the method of least squares was the common minimization algorithm for most of the macromolecular refinement programs such XPLOR version 3.1. Least-squares minimisation is a technique for finding the best fit of a particular assumed model to a set of experimental data when there are more experimental observations than parameters to be determined (Glusker and Trueblood, 1985). With this method, the best parameters that correspond to the assumed model of the structure are found by reducing or minimizing the sum of the squares of the discrepancies between the observed values of $|F_{observed}|^2$ and those calculated $|F_{calculated}|^2$ for an appropriate trial structure or a partially refined version of it. This method of refinement as implemented in the XPLOR version 3.1 program was used in the initial stages of the refinement of the 100K Suc1 model. A limitation of the least squares minimization protocol as applied to macromolecular refinement is the assumption that the errors in the data and model follow a Gaussian distribution. But in protein crystallography, there is rarely a Gaussian distribution between the predicted and the observed data, hence least squares minimization is inadequate for treating the errors in the data and model. These errors are better accounted for by a recently implemented refinement algorithm called the maximum likelihood method. This method has been implemented in the refinement programs REFMAC (Murshudov *et al.*, 1997) which was employed in the latter stages of the refinement of 100K Suc1 in chapter 2 and in the early stages of the refinement of both Ala32Gly and Ser13Gly in chapter 3. Another refinement program called CNS (Brunger

et al., 1998a) which implements this algorithm was employed in the latter stages of the refinement of the two Suc1 mutants. The underlying statistical law of the Maximum likelihood theory is called Bayes' theorem (Matthews, 1998). Bayes discovered a formula for assessing the impact of new evidence (e.g. a model) on an existing theory or observation (e.g. X-ray data). According to Bayes, in order to obtain a better assessment of the new model being obtained given the invariant X-ray data, one has to combine the probability of the data i.e. prior probability with the weight of evidence provided by the new model i.e. the probability of the new model. This concept is encapsulated in the so-called likelihood ratio, which captures the likelihood of the new model being the basis of the X-ray data. The outcome of this formulation is an updated probability (Likelihood = L) of the X-ray data giving rise to the new model. In macromolecular refinement, the prior or *a priori* probability is the restraints to ideal geometry (covalent bonds, bond angles, chirality, planarity and van der Waals forces) imposed on the model. According to the maximum likelihood theory, the best atomic model must be the one which is most consistent with the observations i.e. the X-ray data. In this case, consistency is a statistical measurement of the probability, ***p***, that the observations would be made, given the current model.

$$p(\text{model; data}) = \frac{p(\text{model})p(\text{data;model})}{p(\text{data})} \qquad (8)$$

Since the data are invariant and the only parameters being altered in the course of refinement and manual model rebuilding are in the model, the denominator of equation (8) becomes a normalization constant. In so far as the goal i.e. the maximization of the likelihood is concerned, the presence of a constant factor does not alter the position of the maximum so equation (8) becomes

$$\mathbf{L}(\text{model; data}) = p(\text{model})\, p(\text{data;model}) \qquad (9)$$

In the course of refinement and manual model-building, if the current model is altered to make the observations more probable, (e.g. effecting the Ala32Gly or the Ser13Gly

mutation in the current model to reflect the electron density of a glycine residue instead of an alanine or serine residue), the model gets better and there is a subsequent improvement in the likelihood. In this formulation, the probabilities have to include all the effects of all sources of error, including errors in the model. If the refinement is going well, as the model improves and becomes a truer reflection of the contents of the unit cell, the errors get smaller and the probabilities become more accurate - this leads to an increase in the likelihood. (Read, 1997). It has been shown that in general, the maximum likelihood-based refinement protocols produce higher R_{work} values, but more accurate models and lower or equal R_{free} values compared with least-squares refinement protocols. This is because the maximum likelihood method significantly reduces the amount of overfitting of the data. Another advantage of this method is that the reduced overfitting leads to a reduction in model-bias in electron density maps. This is important during the manual rebuilding stages of refinement and also it reduces the time it takes to move from an initial crude model to a model that is a better reflection of the contents of the unit cell (Adams *et al.*, 1997; Brunger *et al.*, 1998b).

7.2.4 Rigid body refinement

After using the phases of a known structure (e.g. 293K Suc1 as in chapter 2 or 100K Suc1 as in chapter 3) to calculate the initial phases of an unknown structure, the R_{work} value, which is a measure of the difference between the $F_{observed}$ and $F_{calculated}$ was between 45-55% for the three X-ray structures reported in chapters 2 and 3. These relatively high initial R values can be attributed to the slight changes in unit cell dimensions between crystals of the known structure and those of the unknown. A consequence of these changes in cell dimension means that the model is not correctly positioned in the unit cell hence the high R values. In order to rectify this anomaly, one can move (rotation and translation) the contents of the unit cell (subunits ***A***, ***B*** and zinc atoms) as one rigid body till the correct (indicated by relatively low R-values) orientation in the new unit cell is determined. The process can be defined as a refinement procedure during which a group of residues, domains or subunits are restricted to move as one rigid entity.

7.2.5 Positional refinement

Conceptually, this is the opposite of the above where all or groups of atoms are moved together as a rigid entity. In positional refinement, all the atoms of the model have the freedom -subject to chemical restraints/constraints- to move into any position in conformational space. In positional refinement, only the route of steepest descent is permitted in the quest to attain the global minimum. Uphill motion against the energy gradient is not allowed. This limitation means that the refinement process can easily fail by getting trapped in any of the local energy minima along the route to the global minimum. The probability of this happening depends on how many and how deep the local minima are.

7.2.6 Torsion angle molecular dynamics refinement

Comparative analysis of bond lengths and angles from both small molecule and protein crystallography has shown that there is negligible deviation from ideal bond lengths and angles in macromolecular crystal structures. This has afforded crystallographers the opportunity to increase the observation-to-parameter ratio by incorporating prior chemical information into the formulation for X-ray refinement. Traditionally, this empirical information has been used to impose some restraints on the model during X-ray structure refinement so that the bond lengths and angle do not deviate too far from ideal values. In torsion angle molecular dynamics, this flexibility is removed by imposing fixed -length constraints on the model. This reduces the number of variables to search for in conformational space, and also it increases the radius of convergence in macromolecular refinement. Empirical evidence indicates that employing torsion angle molecular dynamics in conjunction with maximum likelihood targets has a synergistic effect on the radius of convergence (Adams *et al.*, 1997 and Brunger, 1998 #1103; Brunger *et al.*, 1997; Rice and Brunger, 1994).

7.2.7 Simulated annealing

Annealing is a physical process whereby a solid is heated until all the constituent particles arrange themselves randomly in a liquid phase: the latter is then slowly cooled so that all particles arrange themselves in the lowest energy state (Brunger *et al.*, 1997).

In macromolecular refinement, the protein model is subjected to the computational simulation of the annealing process. Compared to the above methods, simulated annealing increases the probability of the model attaining a global optimum because motion against the gradient is permitted. The tendency for this uphill motion is determined by the specified temperature. By specifying a fixed temperature, the model initially gains a certain inertia which allows it to cross energy barriers en route to the global energy minimum. It is essential that the target temperature is large enough to overcome small barriers but low enough to ensure that the system does not escape from the global minimum if it manages to arrive there.



Figure 7.9 Schematic diagram of the simulated annealing refinement process *Although, temperature itself is a global parameter of the system, temperature fluctuations arise owing to local conformational transitions - for example from an amino acid side chain falling into the correct orientation. These local changes tend to lower the value of the potential energy of the system, thus increasing the kinetic energy and hence the temperature, of the system. Reproduced from* (Brunger *et al.*, 1997).

Normally, a heat bath or sink is coupled to the system to remove this excess kinetic energy through heat dissipation. This renders the reverse transition an unlikely event, as it would require a localized increase in kinetic energy where the conformational change occurred in the first place. Thus, the effect of temperature coupling is to maintain

a sufficient amount of kinetic energy to rectify local conformational anomalies, but it does not supply enough energy to allow escape from the global minimum. This explains the general observation that the agreement with the experimental data will improve rather than deteriorate with simulated annealing. The advantages of simulated annealing are that it can overcome local minima in the target function and this leads to a reduction in the manual rebuilding required. The effectiveness of simulated annealing can be improved by the use of torsion angle dynamics and maximum likelihood targets. The new refinement package called CNS combines simulated annealing, cross-validation, torsion angle dynamics with maximum likelihood targets. REFMAC does not employ simulated annealing and torsion angle dynamics.

7.2.8 Temperature factor refinement

The B-factor incorporates the fact that atoms are not at rest but have thermal motion. This results in the modification of the structure factor amplitudes by $\exp -Bj\sin^2\theta/\lambda^2$ where $Bj = 8\pi^2\hat{u}_x^2$ and $\hat{u}_x^2$ is the mean square displacement in the x direction. For example, a B = 20Å^2 corresponds to a amplitude of 0.5Å and a B of 80Å^2 gives an amplitude of 1.01Å suggesting that the position of such an atom in the its electron density cannot be accurately assigned (Cruickshank, 1999).

In refinement, B values can take up other errors in scaling, lack of absorption correction, incorrect atomic positions etc...etc...Because of this it is usual to employ an overall B factor for all atoms in the early stages of refinement and then to employ individual B factor towards the latter stages of structure determination when the structure is almost correct.

7.3 Calculation and interpretation of the electron density

When the quality of the initial phases have been improved, an interpretable electron density map can be calculated. In the case of isomorphous replacement, a partial model of the protein can be traced into the electron density using the primary sequence as a guide. In the case of molecular replacement and phase transfer solutions, the initial

protein model would consist of both main chain and side-chain atoms. If the quality of the electron density is good enough, the new and correct sequence can be built in to replace that of the phasing model. Normally, at this stage of a structure determination experiment, the nascent model is relatively crude and full of errors. The main and side chain atoms of many residues may not be correctly positioned in the electron density. Loop regions of the protein which are inherently mobile are particularly prone to being out of density or not possessing a discontinuous density, making tracing relatively difficult. In order to rid the nascent model of these errors, it is necessary to improve the quality of the model to the highest resolution available. Computer graphics have had a major impact on the display and interpretation of electron density maps. In this thesis, the program O was used (Jones, 1991). This allows the model and map to be displayed, and also the model adjusted to optimise the fit to the electron density. When a trial set of calculated phases are available, as in the case of Suc1 2.7Å structure, conventional 2Fo-Fc maps (where Fo and Fc are the observed and calculated structure factors, respectively) allow the problem of model bias to be partly overcome. Model bias is the phenomenon where the inclusion of calculated phase information from a model (e.g. 293K Suc1) carries the risk of biasing the refinement process towards the current atomic model instead of moving away from it to reveal the true identity of the contents of the unit cell (e.g.100K Suc1). It can obscure the detection of errors in atomic models if the phase information is either insufficient or of a poor quality (Hodel *et al.*, 1992).

7.3.1 Qualitative analysis of macromolecular structures

There are several statistical criteria that are used to indicate the quality of the X-ray data, monitor the progress of the data processing and judge the correctness of the final model. In a typical protein crystallography experiment, several different diffraction images would be obtained for a single crystal as it is rotated in the X-ray beam. This feature of data collection means that a reflection (hkl) or its symmetry related counterpart would be recorded more than once. It takes longer to collect a dataset for a protein crystal belonging to a low symmetry space group than for a crystal belong to a high symmetry space group. A degree of redundancy is an inescapable but desirable feature of this mode of data harvesting. The various diffraction images can be reduced to a unique collection

using a program such as SCALEPACK (Otwinowski and Minor, 1997). The merging R value on intensities indicates the quality of the merged data. It is a measure of the accuracy of the scaled data.

$$R_{merge}(I) = \frac{\sum_h \sum_i |I_h - I_{hi}|}{\sum_h \sum_i I_{hi}}$$

where I_{hi} is the intensity of a reflection and I_h is the average intensity of that reflection.

In general, an R-merge of less than 10% is indicative of a satisfactory fit but ideally high precision (eg. $R_{merge} < 5\%$) should be aimed at. A criterion which is deemed to be as important - if not more so, as the R_{merge} is the average signal-to-noise ratio $I/\sigma(I)$. This is very important for the highest resolution range of the X-ray data set where the intensities are weak. A value of more than 2σ for most of the reflections in the highest resolution shell is deemed statistically significant. This value tells the user of the protein model whether the structure is actually at the quoted resolution. Another gauge for the latter attribute is the percentage completeness of the X-ray data. Given the spacegroup and cell dimensions of a protein crystal, a program such as SCALEPACK can predict the expected number of unique reflections and comparison of this with the number actually collected gives the percentage completeness. It is desirable for the completeness to be as high as possible because systematic gaps in the data can lead to misleading electron density maps and errors in the protein structure. Again, the value for the highest resolution shell is an indicator of the quoted resolution of the protein structure. The total number of reflections divided by the number of unique reflections, after outliers have been rejected is called the Average Redundancy of the scaled data set. A high value together with a low Rmerge, is an indication of the likely precision with which the data have been collected. When all the above criteria have been satisfied, the next phase of statistical monitoring occurs in the refinement stage of structure determination.

7.4 CRYSTALLOGRAPHIC PROGRAMS USED

1. PREDICT: A program for obtaining the oscillation angle and total angular range required to collect a complete data set given parameters such as space group and unit cell dimensions. Dr. Noble (unpublished work).
2. DENZO: A data reduction program (Otwinowski and Minor, 1997).
3. SCALEPACK: A program for scaling reduced data (Otwinowski and Minor, 1997).
4. TRUNCATE: A CCP4 program for calculating structure factor amplitudes for native and derivative data sets. It accepts the output from SCALEPACK and corrects for negative intensities (CCP4, 1994).
5. MAPMAN: A program for map conversions
6. XPLOR: A program for structure determination and refinement (Brunger, 1992b).
7. O : A graphics program for displaying and building models (Jones, 1991).
8. REFMAC: a CCP4 program for refining macromolecular structures (Murshudov *et al.*, 1997).
9. CNS: A program for structure determination and refinement (Brunger *et al.*, 1998a).
10. PROCHECK: A program for assessing the stereochemistry of a model (Laskowski *et al.*, 1993).
11. AESOP: A program for drawing macromolecules. Dr. Noble (unpublished work).
12. ALSCRIPT: A program for displaying and manipulating alligned amino acid sequences (Barton, 1993).
13. GRASP: A program for protein surface properties (Nicholls and Honig, 1991).

7.5 LITERATURE CITED

Adams, P.D., Pannu, N.S., Read, R.J. and Brunger, A.T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc Natl Acad Sci U S A*, **94**, 5018-5023.

Barton, G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Engineering*, **6**, 37-40.

Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press, London.

Brunger, A.T. (1992a) Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

Brunger, A.T. (1992b) *X-PLOR: Version 3.1; a system for protein crystallography and NMR*. Yale University Press, New Haven.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998a) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54**, 905-921.

Brunger, A.T., Adams, P.D. and Rice, L.M. (1997) New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*, **5**, 325-336.

Brunger, A.T., Adams, P.D. and Rice, L.M. (1998b) Recent developments for the efficient crystallographic refinement of macromolecular structures. *Curr Opin Struct Biol*, **8**, 606-611.

Carter, C.W. and Sweet, R.M. (1997) *Macromolecular Crystallography*. Academic Press, London.

CCP4. (1994) The CCP4 (Collaborative Computational Project Number 4) suite: programmes for protein crystallography. *Acta Cryst. D*, **D50**, 760-763.

Chayen, N.E., Boggon, T.J., Cassetta, A., Deacon, A., Gleichmann, T., Habash, J., Harrop, S.J., Helliwell, J.R., Nieh, Y.P., Peterson, M.R., Raftery, J., Snell, E.H., Hadener, A., Niemann, A.C., Siddons, D.P., Stojanoff, V., Thompson, A.W., Ursby, T. and Wulff, M. (1996) Trends and challenges in experimental macromolecular crystallography. *Q Rev Biophys*, **29**, 227-278.

Cruickshank, D.W. (1999) Remarks about protein structure precision. *Acta Crystallogr D Biol Crystallogr*, **55**, 583-601.

Drenth, J. (1994) *Principles of Protein Crystallography*. Springer-Verlag, New York.
Garman, E.F. (1996) Modern methods for rapid x-ray diffraction data collection from crystals of macromolecules. *Methods Mol Biol*, **56**, 87-126.

Garman, E.F. and Schneider, T.R. (1997) Macromolecular cryocrystallography. *J. Appl. Cryst.*, **30**, 211-237.

Glusker, J.P. and Trueblood, K.N. (1985) *Crystal structure analysis: a primer*. Oxford University Press, Oxford.

Gruner, S.M. (1994) X-ray detectors for macromolecular crystallography. *Current opinion in structural biology*, **4**, 765-769.

Helliwell, J.R. (1998) Synchrotron radiation facilities. *Nat Struct Biol*, **5 Suppl**, 614-617.
Hodel, A., Kim, S.-H. and Brunger, A.T. (1992) Model bias in macromolecular crytal structures. *Acta crystallographica*, **A48**, 851-858.

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

Khazanovich, N., Bateman, K., Chernaia, M., Michalak, M. and James, M. (1996) Crystal structure of the yeast cell-cycle control protein, p13suc1, in a strand-exchanged dimer. *Structure*, **4**, 299-309.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thorton, J.M. (1993) PROCHECK: a programme to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283-291.

Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.

Matthews, R. (1998) Flukes and flaws. *Prospect*, Vol. Issue 35: November 1998, pp. 20-24.

Mitchell, E., Kuhn, P. and Garman, E. (1999) Demystifying the synchrotron trip: a first time user's guide. *Structure Fold Des*, **7**, R111-121.

Murshudov, G.N., Vagen, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum -likelihood method. *Acta Cryst.*, **D53**, 240-255.

Nicholls, A. and Honig, B. (1991) A rapid finite difference algorithm, utilising successive over relaxation to solve the Poisson Boltzmann equation. *J. Comp. Chem.*, **12**, 435-445.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collection Collected in Oscillation Mode. In Carter, C.W. and Sweet, R.M. (eds.), *Methods in Enzymology*. Academic Press, London, Vol. 276, pp. 307 - 325.

Read, R.J. (1997) Maximum likelihood in protein crystallography. *CCP4 refinement workshop*, University of York.

Rhodes, G. (1993) *Crystallography Made Crystal Clear. A Guide for Users of Macromolecular Models*. Academic Press, San Diego.

Rice, L.M. and Brunger, A.T. (1994) Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins*, **19**, 277-290.