PROTEIN MODELLING AND ASSOCIATED DRUG DESIGN

A Thesis presented by

Richard K. Scott

in partial fulfilment of the requirements for the
Degree of Doctor of Philosophy at the
University of Newcastle upon Tyne

Department of Chemistry
University of Newcastle upon Tyne
September 1993

ACKNOWLEDGEMENTS

I would like to extend my thanks to my supervisor Prof. W. Clegg for his advice and guidance during my course of research. I would especially like to thank Dr. Nick Tomkinson for his enthusiasm and helpful comments. I would also like to thank Prof. B.T. Golding and Dr. R. Griffin from the Chemistry Department and Prof. A.H. Calvert and Dr. D.R. Newell from the Cancer Research Unit. I would also like to thank Prof. Harry Gilbert for help in protein purification and extraction and Prof. Roger Pain for his technical advice in secondary structure determination.
Finally, I would also like to thank Dr. R.M. Stroud for his kind gift of the crystal coordinates of Thymidylate Synthase, SERC for financial support throughout this work, Molecular Simulations for the use of their facilities and Nyree for the tremendous help and support she has given me in the preparation of this Thesis.

ABSTRACT

This Thesis is presented in two parts -

Part one covers an investigation into the secondary and tertiary structure of the protein *Xylanase* found in ***Pseudomonas fluorescens subspecies cellulosa.***

Part two documents the Computer Aided Design of Novel Quinazoline Antifolates for the enzyme **Thymidylate Synthase .**

Part one

Mature Xylanase protein consists of two distinct regions - a cellulose binding domain and a catalytic region, A and B respectively. Computer modelling of tertiary structure from primary sequence and secondary turn information proved difficult in the absence of experimental X-ray crystal data. Consequently, a series of modified proteins based on the Xylanase were prepared by Recombinant DNA technology for extraction and purification. The modified proteins were to be used as a bench mark for quantitative and definitive calculation and determination of the secondary structure of the xylanase. This was to provide an excellent reference point for theoretical modelling of tertiary structure. Part one of the Thesis documents Computer Modelling work and protein purification and extraction of the xylanase.

Part two

Thymidylate Synthase (TS) exists as dimer with a single active site in each subunit. It has been crystallised in two forms; a "reduced" (major), and "oxidised" (minor) form. The major form of TS contains dUMP covalently bound to cysteine in both active sites and in the presence of CB3717. One active site of the minor form contains dUMP non-covalently bound and in the presence of CB3717 while the other active site contains only inorganic phosphate and CB3717. The active site of TS is a large cavity that binds CB3717 into two possible confirmations. One is seen in the major form and one in the minor form.

Part two of my research documents an investigation into enzyme/inhibitor interaction in TS and covers the Computer Aided Design of a series of inhibitors based on the knowledge of the TS active site. Several of these compounds have been put forward as target compounds for synthesis.

CONTENTS

(xiv) Cellulose column extraction of xylanase
(xv) Protein assay of cellulose columns

1.3 RESULTS

(i) Computer modelling
(ii) Cloning
(iii) Periplasmic extraction
(iv) Xylanase assays
(v) Avicel binding
(vi) Guanidine extraction and protein dialysis
(vii) Cloning without pNM52 regulation
(viii) Further Xylanase assays
(ix) Cellulose column extraction
(x) Protein assays

1.4 DISCUSSION

(i) Cloning of Xylanase
(ii) Xylanase assays
(iii) Extraction Techniques
  a) Avicel binding
  b) Guanidine extraction
  c) Cellulose column extraction
(iv) Xylanase protein recovery
(v) Modelling of protein structure
(vi) Importance of crystal data
(vii) Xylanase as a target for protein modelling
(viii) Future work
(ix) Summary

PART 2

2.1 INTRODUCTION

(i) Thymidylate Synthase
(ii) Thymidylate Synthase Crystal structure
(iii) Computer modelling using QUANTA/CHARMm
    a) Overview
    b) Architecture of Modelling
    c) Atomic coordinate manipulation
(iv) Computational Algorithms
    a) Overview
    b) Energetics and force
    c) Internal energy terms
    d) External energy terms
    e) Electrostatic Energy
    f) van der Waals Energy
    g) Constraint Energy
(v) Minimisation
    a) Steepest Descents
    b) Powell
    c) Adopted basis-set Newton Raphson
(vii) Electrostatic Potential Surface

2.2 MATERIALS AND METHODS

(i) Software
(ii) Hardware
(iii) Data
(iv) Methods
    a) Comparison of the two TS forms
    b) Analysis of crystal structure
    c) Active site comparison
    d) Conformational preferences of CB3717
    e) THF and DHF conformations
    f) Application of patches
    g) Minimisation of TS containing CB3717, THF and DHF
    h) N^{10} substitution - N^{9} ligand design and cbnew series
    i) Minimisation of one subunit of TS

j) Bifurcated ligand design - the BIF series
k) External energy interactions for each atom in the BIF and CB3717 series at 4,8 and 12Å range
l) Comparison of internal and external energy contributions
m) van der Waals and electrostatic summation - bumps and gaps
n) Solvation energies - comparison of interaction
o) Most promising Ligands

2.3 RESULTS

(i) TS forms - major and minor conformations
(ii) Crystal Structure
(iii) Active site comparison
(iv) Conformational preferences of CB3717
(v) Conformational preferences of DHF and THF
(vi) Conformational preferences of DHF and THF in TS
(vii) The cbnew series of ligands
(viii) Monomer minimisation
(ix) The BIF series of ligands
(x) Interaction at 4, 8 and 12Å
(xi) Solvation energies and minimisation
(xii) Connolly surface calculation

2.4 DISCUSSION

(i) Molecular mechanics and drug design
(ii) History of computer aided drug design
(iii) Use and limitations of Modelling
(iv) Assessment of methodology
(v) Physiology of TS
(vi) Active site of TS
(vii) Effect of new ligands
(viii) Applications of new ligands
(ix) Further development
(x) Future work
(xi) Summary

REFERENCES

APPENDICES

Appendix 1.1	Transformation of Religated Plasmids
Appendix 1.2	L-Broth preparation and Agar plate preparation
Appendix 1.3	Preparation of periplasmic fraction from cultures
Appendix 1.4	Xylanase assay stock preparation and standard curves
Appendix 2.1	Custom parameter set
Appendix 2.2	Residue Topology Files
Appendix 2.3	PATCH data and awk files
Appendix 2.4	Solvation energies and input files

PART ONE
PROTEIN MODELLING AND XYLANASE

1.1 INTRODUCTION

(i) Xylan

Xylan is one of the major components of hemicellulose, the compound that cements together the cellulose fibres which form the structure of plant cell walls. Xylan is found predominantly in grasses where it forms chains made up of 1-4 linked anhydro-D-xylose units variously substituted by L-arabinose and 4-O-methyl D-glucuronic acid (see Figure 1.1).

Xylan can be hydrolysed by 3 classes of xylanase enzyme: the ß-xylanases (endo-1, 4-ß-D-xylanohydrolase) ß-xylosidases and a third enzyme xylobiase.[1] Endoxylanases hydrolyse the internal glycosidic links of xylan, while exoxylanases attack the external ends of xylan. These two enzymes break down xylan to xylobiose and xylose. The third enzyme xylobiase attacks the disaccharide xylobiose and breaks it down to the monosaccharide xylose.

Xylose can be metabolised to fructose-6-phosphate within the rumen and to lactic acid within the silage clamp. Xylanase enzymes have been identified from several organisms:

Ruminococcus flavefaciens has been found to contain a multifunctional enzyme which possesses xylanase and endocellulase activity.[2] Similarly *Cellulomonas uda* has been found to exhibit xylanase activity as well as glucanase activity. The xylanase enzyme was found to be cell bound and it degraded ß-xylobiose.[1] *Pseudomonas fluorescens* subsp. *cellulosa* contains 3 xylanase genes and 3 cellulase genes.[3] *Bacillus circulans* possesses 2 endo ß-D-xylanases and 1 ß-D-xylobiase.[4]

Such micro organisms are thought to utilise these multifunctional enzyme complexes to degrade both cellulose and hemicellulose and ultimately to break down plant cell walls into water soluble carbohydrate.

(ii) Cloning Methodology

Recombinant DNA technology has expanded as an industry in its own right over the last 20 years. Its applications are varied and widespread. Genes encoding for polypeptide sequences of interest i.e. proteins such as enzymes, are removed from the native chromosomal DNA and inserted into a suitable vehicle (Vector), so that the

Figure 1.1

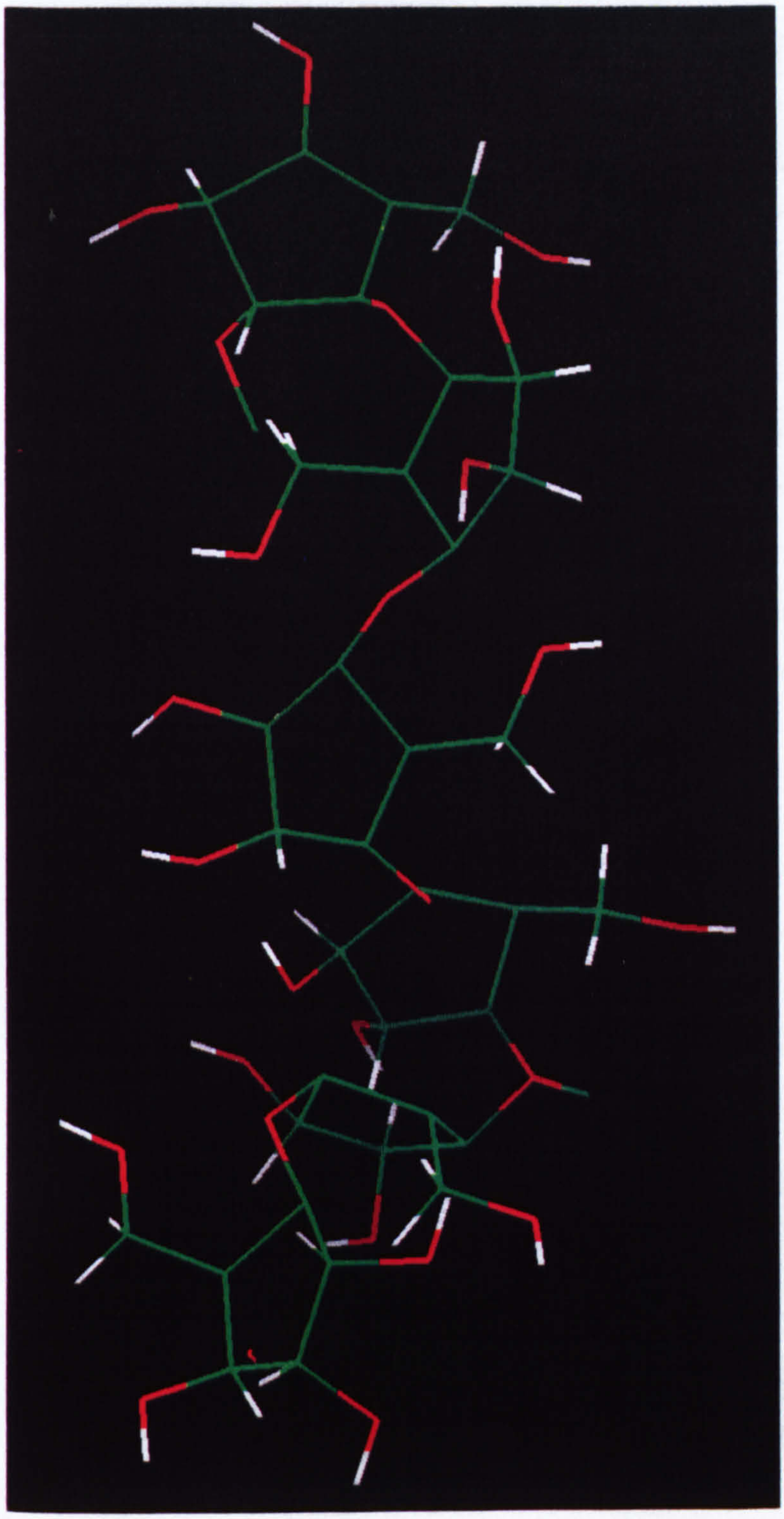


Figure 1.2

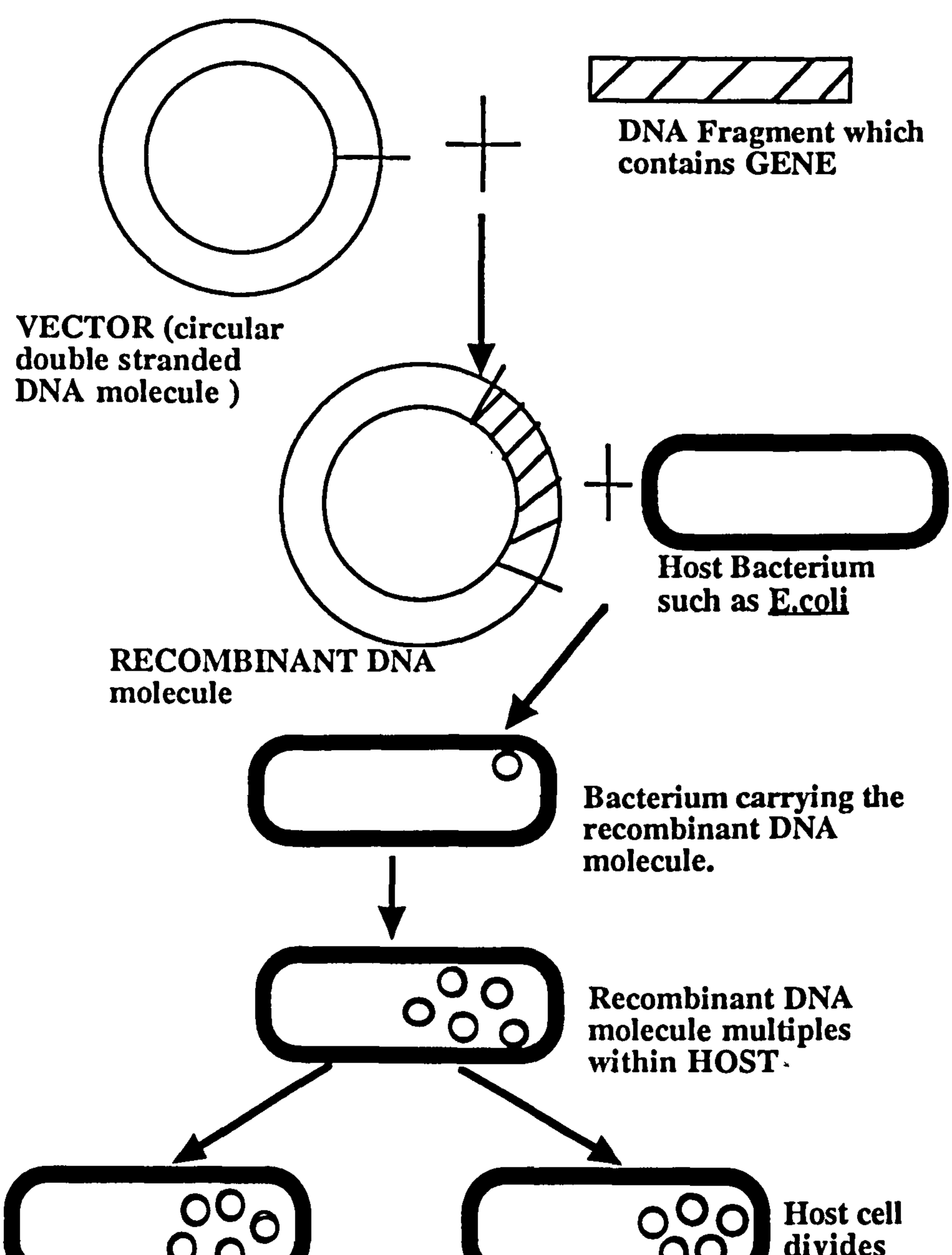


Diagram of Gene Replication
DNA Fragment which contains GENE
VECTOR (circular double stranded DNA molecule)
Host Bacterium such as E.coli
RECOMBINANT DNA molecule
Bacterium carrying the recombinant DNA molecule.
Recombinant DNA molecule multiples within HOST.
Host cell divides

gene can be replicated in a host cell - usually a bacterium. In this study, the gene of interest (XYLA) was taken from *Pseudomonas fluorescens* subsp. *cellulosa* and cloned into a strain of *E.coli* (see Figure 1.2).

The genes are manipulated using restriction endonucleases (RE's) which recognise specific nucleotide sequences in double-stranded DNA and cleave the double strand at this site.

For example, the RE BamH1* recognises the nucleotide sequence GATCC and cleaves the DNA to produce cohesive ends (see Figure 1.3)*.

The gene of interest can be inserted into double-stranded DNA vectors by digestion of both vector and gene with suitable RE's so that cohesive ends formed can interact to form hydrogen bonds. The hybrid molecule is joined up by another enzyme - DNA ligase.

DNA ligase seals the vector molecule and the gene together by replacing the phosphodiester bond between adjacent nucleotides (see Figure 1.3). Suitable DNA vectors include plasmids and bacteriophage which can be maintained autonomously in the host cell. The recombinant DNA molecule is introduced into the host cell by a process known as transformation whereby the living host cell takes up the foreign DNA molecule under suitable experimental conditions.

After multiple cell divisions a clone of identical host cells is produced , every cell in this clone containing one or more copies of the gene. Cells harbouring the recombinant DNA molecule can be selected for by growth, complementation or antibody screening.

(iii) Structure of Xylanase

The catalytic mechanisms of Xylanase have been thoroughly elucidated.[5-8] Limited information on the domains of the Xylanase has also been obtained:

Xylanase protein consists of two distinct regions A and B.

Region A is serine-rich (>16% SER) and contains a cellulose binding domain.

Region B contains the active region of the protein- Region A can be excised without the protein losing its activity (see Figure 1.4 and 1.5).

The cellulose binding domain shares considerable homology with the *P.fluorescens* endoglucanase. A hypothetical model has been proposed for the mode of enzyme action and its possible tertiary structure (see Figure 1.6).

The most likely mode of enzyme action involves region A attaching to the substrate,

---

* Bam H1 is the name of a commercially available restriction endonuclease

Figure 1.3

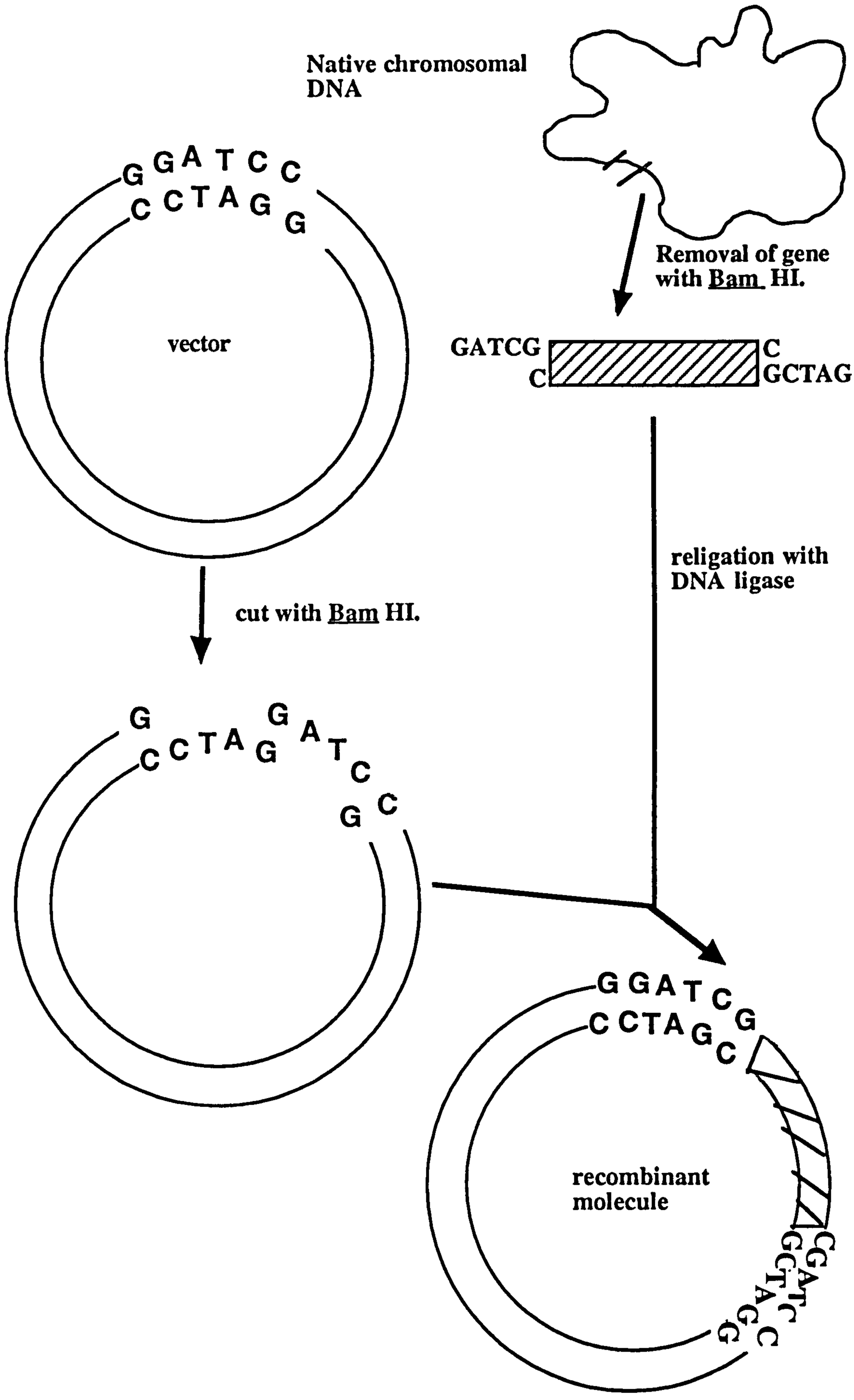
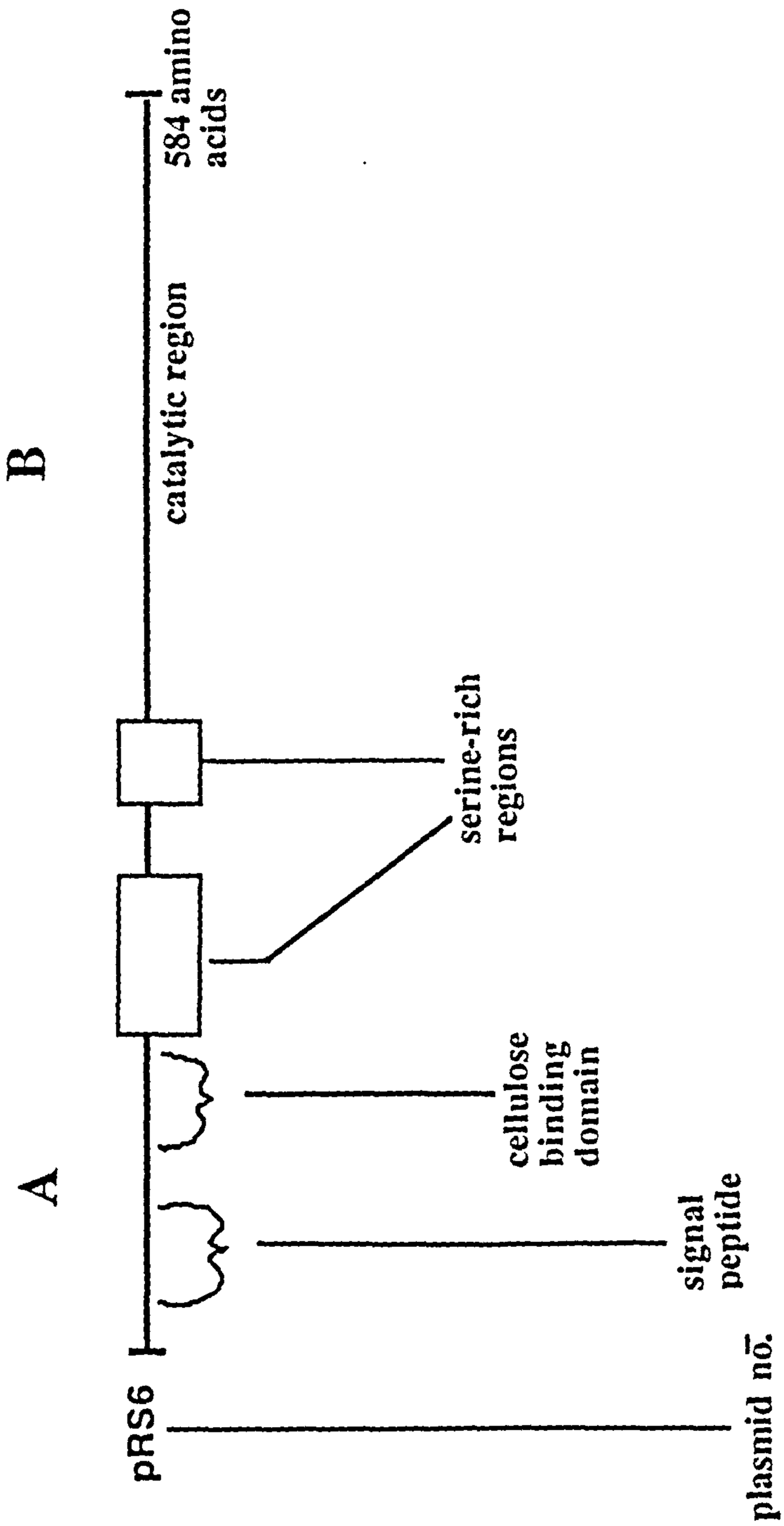

Native chromosomal DNA
Removal of gene with Bam HI.
GATCG
C
C
GCTAG
vector
cut with Bam HI.
religation with DNA ligase
recombinant molecule

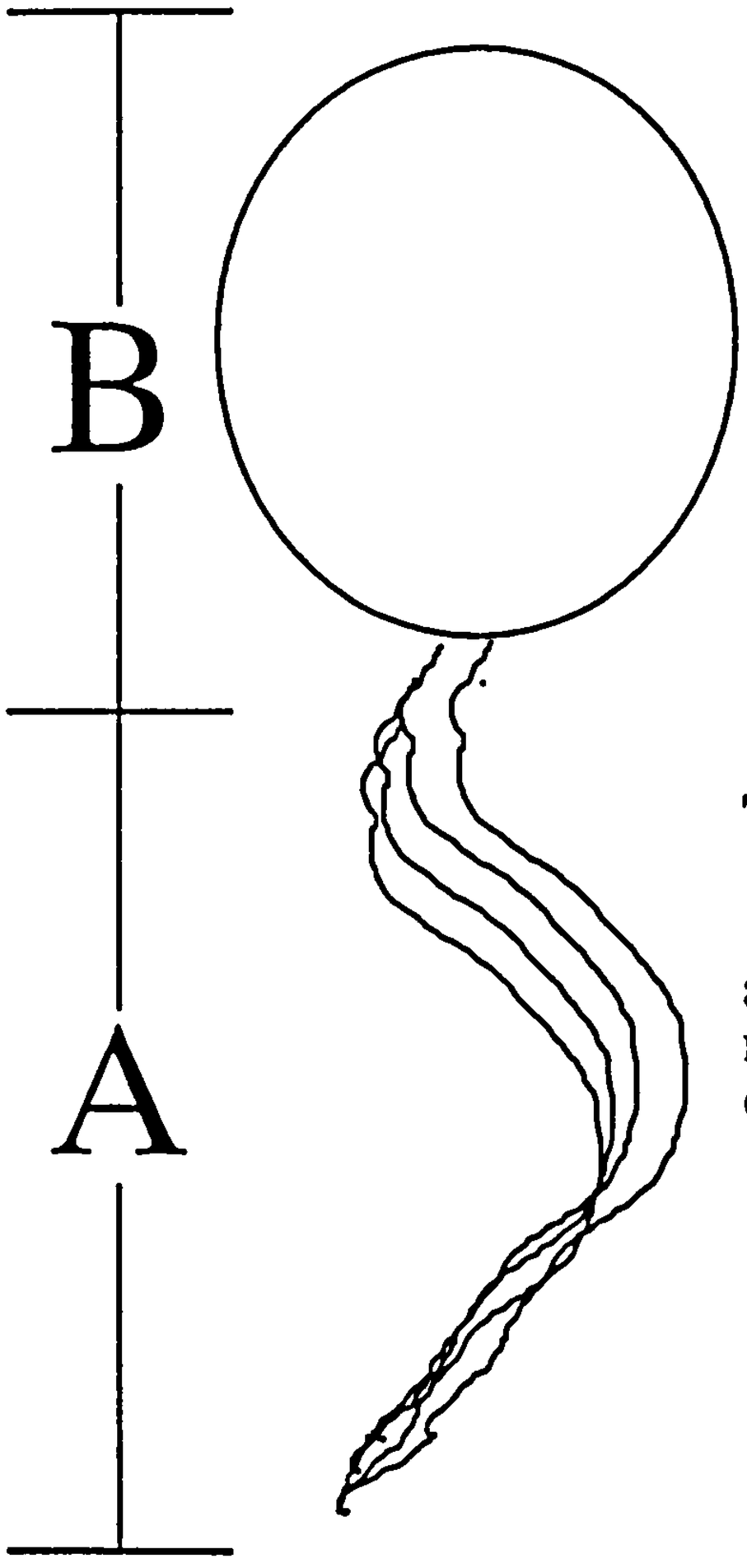
Figure 1.4


pRS6
plasmid no.
A
signal peptide
cellulose binding domain
serine-rich regions
B
catalytic region
584 amino acids

```
REGION_A      *1   *  AQTADCSYNI*11 *  TNEWNTGYIG*21 *  DITITNRGSS*31 *  AINGWSVNWQ*41 *  YATNRLSSSW*51 *  NANVSGSNPY
REGION_B      *1   *  IFTSSARQNI*11 *  VRAEFNQITA*21 *  ENIMKMSYMY*31 *  SGSNFSFTNS*41 *  DRLVSWAAQN*51 *  GQTVHGHALV

REGION_A      *61  *  SASNLSWNGN*71 *  IQPGQSVSFG*81 *  FQVNKNGGSA*91 *  ERPSVGGSIC*101*  SGSVASSSAP*111*  ASSVPSSIAS
REGION_B      *61  *  WHPSYQLPNW*71 *  ASDSNANFRQ*81 *  DFARHIDTVA*91 *  AHFAGQVKSW*101*  DVVNEALFDS*111*  ADDPDGRGSA

REGION_A      *121*  SSPSSVASSV*131*  ISSMASSSPV*141*  SSSSVASSTP*151*  GSSSGNQQCN*161*  WYGTLYPLCV*171*  TTTNGWGWED
REGION_B      *121*  NGYRQSVFYR*131*  QFGGPEYIDE*141*  AFRRAPRADP*151*  TAELYYNDFN*161*  TEENGAKTTA*171*  LVNLVQRLLN

REGION_A      *181*  QRSCIARSTC*191*  AAQPAPFGIV*201*  GSGSSTPVSS*211*  SSSSLSSSSV*221*  VSSIRSSSSS*231*  SSSSVA
REGION_B      *181*  NGVPIDGVGF*191*  QMHVMNDYPS*201*  IANIRQAMQK*211*  IVALSPTLKI*221*  KITELDVRLN*231*  NPYDGNSSND

REGION_A
REGION_B      *241*  YTNRNDCAVS*251*  CAGLDRQKAR*261*  YKEIVQAYLE*271*  VVPPGRRGGI*281*  TVWGIADPDS*291*  WLYTHQNLPD

REGION_A
REGION_B      *301*  WPLLFNDNLQ*311*  PKPAYQGVVE*321*  ALSGR

                                                       Hydrophobic
                                                       Neutral hydrophilic
                                                       Basic hydrophilic
                                                       Acidic hydrophilic
```

Figure 1.5

Figure 1.6



HEAD REGION

globular - containing active site. Hydrolyses xylan to water soluble carbohydrates.

TAIL SECTION

attatches to substrate molecule to overcome entropy.

bringing region B into contact with the substrate, where the active sites cleaves the ß, 1-4 bonds of the Xylan chain.

(iv) Applications of Xylanase

This enzyme found in *Pseudomonas* and other micro organisms such as *Cellumonas* is often part of a multifunctional enzyme suite.[9] These enzyme complexes degrade both cellulose and hemicellulose, ultimately breaking down plant cell walls to water soluble carbohydrate. Xylanase could, in theory, be used commercially as either a silage inoculant (by improving silage quality through lowering clamp pH quickly and liberating water soluble carbohydrates), or as a novel enzyme secreted by transgenic livestock which aids forage digestion and breakdown.

Recombinant DNA technology has advanced sufficiently so, that the gene encoding xylanase could be inserted into suitable plasmid DNA hosts. The gene could then be incorporated into animal cell lines and ultimately, the animal itself. With suitable expression sequences and a stable environment for the protein (e.g. liver tissue harbouring the Xylanase gene), the Xylanase could be expressed and have a positive *in vivo* effect on nutrient uptake in the duodenum. The host animal would be able to derive greater nutritional value from forage - as the plant cell walls were hydrolysed to carbohydrate by the novel enzyme.

(v) Advantages of Computer Modelling

Computer modelling of crystal structures gives a rapid indication of the effects of specific active site mutations.[10-11] It can give an insight into the exact nature of the way the enzyme binds to the substrate. Modelling can calculate the energetics of interactions such as substrate binding - in the case of Xylanase, modelling can help determine how the enzyme attaches to the substrate polymer chain, either by sequential breakage of ß, 1-4 bonds or random attach/reattachment along the chain.

(vi) Limitations of computer modelling

With the X-ray crystal data of the Xylanase catalytic region from Reading Institute not forthcoming, detailed docking studies using modelling packages could not be performed. Crystallisation of such a large protein proved difficult
(see Chapter1.4(vi)).
Any tertiary structure predictions are limited by both the software and relatively small database size of elucidated structures[12] - to date, there are homologous proteins

available from over 400 structures in the Brookhaven database. Consequently, efforts were moved towards production of purified xylanase protein. This would allow further biological experimentation (such as gel filtration and circular dichroism) in order to determine the tertiary protein structure of the xylanase. The number of chemical experiments that could be performed was dependent on the amount of pure Xylanase protein harvested. If the protein was difficult to obtain and purify (15 milligrammes of 90% pure protein needed to perform a 2D NMR run - this was considered a large amount in synthesis terms), then some experiments became unfeasible.

1.2 MATERIALS AND METHODS

(i) Computer Modelling

Using the molecular modelling package QUANTA v(3.0) and CHARMm[13] version (21.2) as supplied by Molecular Simulations on Silicon Graphics 4D20 personal Iris workstations, the raw amino acid sequences of both the A and B region of the Xylanase were written into the sequence builder module of QUANTA's Protein Package subprogram.

Using the standard amino acid residue topology file (AMINO.RTF), the sequence was read into QUANTA. The secondary conformation of Xylanase was then specified - residues taking up either the right-handed alpha helix or parallel ß sheet conformation. The secondary structure assignment routine then was implemented to add any available turn information (algorithm used according to Garnier *et al.*[14-15], default of 180^{o} turn). Hydrogen bonding was then calculated and the secondary structure of region A or B displayed on screen.

The CHARMm program was then used to minimise the two regions (see Chapter 2.1 for details).
A specified 500 steps of the "steepest descents" algorithm was employed to perform a first minimisation on each region, the coordinates for each step being updated every minimisation cycle. This process was repeated until the r.m.s (root mean square) force was less than 0.100 kcal/mol (r.m.s force gives an overall indication of energetic stability between residues).
No distance constraints were set out at this stage. The "Adopted-Basis Newton Raphson" algorithm[16-17] was then applied to the minimised structures until the r.m.s. force was below 0.100 kcal/mol. The resultant energy of each region was displayed in Kcals/mol.

(ii) Xylanase harvesting and purification

The table below indicates the plasmids and microbial strain used in this work:

E.coli strain	Genotype of Phenotype	Source
JM83	ara^{Δ} (lac-pro)Sm^{R}	Norrander
	thi ø 89 dlac ZMIS	et al (1983)[18]
Plasmids		
pNM52	lac Iq Tc^{R}	Gilbert et al [3]
pN23++	lac Iq Tc^{R} Amp^{R}	Gilbert et al [3]
B41	lac Iq Tc^{R} Amp^{R}	Gilbert et al [3]

All these materials were stored in cryo tubes at -70°C. The *E.Coli* strain JM83 is a commerically available host bacterium and the plasmids have inherent antibiotic resistance to Tetracycline (Tc^{R}) and Ampicillin (Amp^{R}).
Gilson pipettes were used throughout in all transformations and religations.
Sterilised pipette tips were used once, discarded to prevent reagent contamination and surgical gloves were worn when using enzyme to avoid contamination by exogenous DNA.
Thorough mixing was performed by vortex and cryo samples were spun by microcentrifuge.
All enzymes, enzyme buffers and cell-free extracts were stored at -20°C.

(iii) Transformation and religation of JM83 containing pNM52 with the plasmids B41 and pN23++

JM83 was grown up to competent status and transformed with either B41 or pN23++ plasmids (see Appendix 1.1).
Transformations were plated out, incubated and observed.

(iv) Determination of recombinants of pN23++ and B41

Plating ensured only recombinants of pN23++ and B41 grew on media containing ampicillin and tetracycline (pN23++ and B41 are resistant to these two antibiotics). Four colonies from E.coli harbouring pN23++ and four from E.coli harbouring B41 were picked and inoculated into sterile universal bottles containing 10ml L-Broth, 100µl Amp stock and 10ml TET stock. and 10 µl 0.1M IPTG.
These stock cultures were grown overnight at 37°C on an orbiting starter (100 revs).

(v) Extraction of periplasmic fraction

Cultures of JM83 harbouring pNM52 and transformed with recombinant pN23++ or B41 were grown to mid log phase (optical density of 0.4 at 550 nm) with final antibiotic concentrations of 100 µg/ml ampicillin and 15 µg/ml tetracycline.
These cultures were then prepared for Xylanase assay by extraction of the cell periplasmic fraction (see Appendix 1.3).

(vi) Xylanase assays

a) Non IPTG induced cultures

50 µl of each periplasmic fraction was incubated with 1 ml 1.3%w/w xylan in PC buffer and 95 µl PC buffer for 20 minutes at 37°C. Samples were removed and 2ml of DNSA reagent (see Appendix 1.4) was added to terminate the reaction . Samples were then vortexed, placed on ice for 5 minutes and then incubated in a boiling water bath for 20 minutes.
Samples were then microfuged for 2 minutes and the extinction value of the supernatants were measured at 575 nm.

b) IPTG induced cultures

Periplasmic fractions were obtained from cultures incubated with an additional 1µl IPTG stock per ml of culture and assayed as described previously.

c) Time course assays

This was performed to determine the ideal time to harvest the Xylanase protein.
Cultures were grown to mid log phase and then induced with IPTG stock (23 mg/ml) - 1µl per 1 ml of culture.
Samples were removed from the cultures at regular intervals and periplasmic fractions extracted and assayed for Xylanase content.

(vii) Avicel binding to Xylanase

To sterile universal bottles containing periplasmic fractions was added an equal volume of 5%w/v avicel pH 10.5 in 100 mM tris HCL pH 7.5 . The universal bottles were placed horizontally in ice on a shaker for 1 hour to let the avicel bind to the xylanase.

Samples were then shaken vigorously to resuspend the avicel and then washed with 3 x 10ml 100mM tris HCL pH 7.5 over a micro glass vacuum filter.
The washings were collected and assayed for xylanase activity to ensure that all the xylanase protein was bound to the avicel. The avicel cake was retained and resuspended in 1 ml 10% SDS (Sodium dodecyl sulphate) and boiled for 10 minutes in sterile universal bottles. Samples were decanted into eppendorfs and microfuged for 5 minutes and the supernatant retained for SDS-page mini gel electrophoresis.

(viii) SDS - page mini gel electrophoresis

SDS-Polyacrylamide gels are used to analyse translational products of genes i.e. proteins. Running a gel on periplasmic fraction samples gave an indication of the amount of protein present (full form xylanase or truncated inactive form).
SDS denatures the tertiary and secondary structure of the xylanase protein, leaving a polypeptide chain with a negative charge due to the SDS molecule binding to the chain. Denatured proteins can be separated by running them on a polyacrylamide gel with a voltage applied across the gel. The protein chains will migrate to the positive terminal through the gel matrix.
The gel was poured between two glass plates held apart by 1.5 cm width spacers and sealed with a rubber tube gasket. The gel was run vertically with reservoir buffer tanks situated at both ends of plates.

Preparation of stock solutions

i)	Acrylamide	30 grammes
	N,N^1-methylene bisacrylamide	0.6 grammes
	H_20 (double distilled) to	10 ml
ii)	Ammonium persulphate 10%	2.5 grammes
	H_20 (double distilled) to	25 ml

N.B. Due to the fact that Acrylamide is a powerful neurotoxin in monomer form, mask and gloves were used in preparation of stock solutions

Preparation of Running gel (for 2 gels)

Acrylamide stock 30%	8.35 ml
0.75m tris/SDS pH 8.8	12.50 ml
Ammonium persulphate	90 µl
Double distilled water to	25 ml

tris/SDS preparation:

Trizma base	71g
Tris/HCL	47.2g
SDS	2g

adjusted to pH 8.8 with 1M NaOH.
This was made up to 1 litre with double distilled water.

Plate preparation for pouring

The glass plates were washed in warm detergent solution and rinsed with single then double distilled water. The plates were then rinsed with ethanol and allowed to air-dry (note - care was taken to avoid contamination of the working surfaces of the plates). The rubber tube gasket was gently placed at the base of one plate and the other plate laid flush on top to form a sandwich. The plate and gasket assembly was clamped with bulldog clips to form a tight seal around the gasket. 30 µl of TEMED (1M) (N,N,N',N' - tetramethylene diamine) was added to the running gel, mixed and poured into the space between the plates. The plates were filled up to three quarters depth and isobutanol was layered on the gel to prevent oxidation of the air-exposed surface as the gel set. TEMED catalyses the reaction between acrylamide and bis-acrylamide to form the gel polymer that sets in about 15 minutes.

Preparation of Stacking gel (per 2 gels)

30% acrylamide stock	4 mls
0.25M Tris/SDS pH 6.8	7 mls
TEMED	18 µl
10 % ammonium persulphate	54 µl

Double distilled water to 15 mls total volume (approximately 3.5 ml) was added.
N.B. The stacking gel had to be used within 5 minutes after preparation before it set.
tris/SDS preparation

Tris/HCL	39.4g
SDS	2g
adjust to pH	6.8

Double distilled water was added to make the volume to 1 litre.

When the running gel set, the isobutanol was poured off and the gel rinsed with double distilled water repeatedly. The stacking gel was poured between the glass plates over the running gel and a spacer comb was inserted flush with the top of the assembly to

form wells in the gel (see Figure 1.7). Once set, the comb was carefully removed, the wells rinsed with double distilled water and the rubber gasket and clamps removed with care. The gel sandwiched between the glass plates was then slotted into an electrophoresis tank. Upper and lower reservoir tanks were filled with reservoir buffer.

Reservoir electrode buffer

Tris	15.15g
Glycine	72g
SDS	5g
adjust to pH 8.3	

This was made up to 5 litres volume with double distilled water.

Sample preparation for the gel

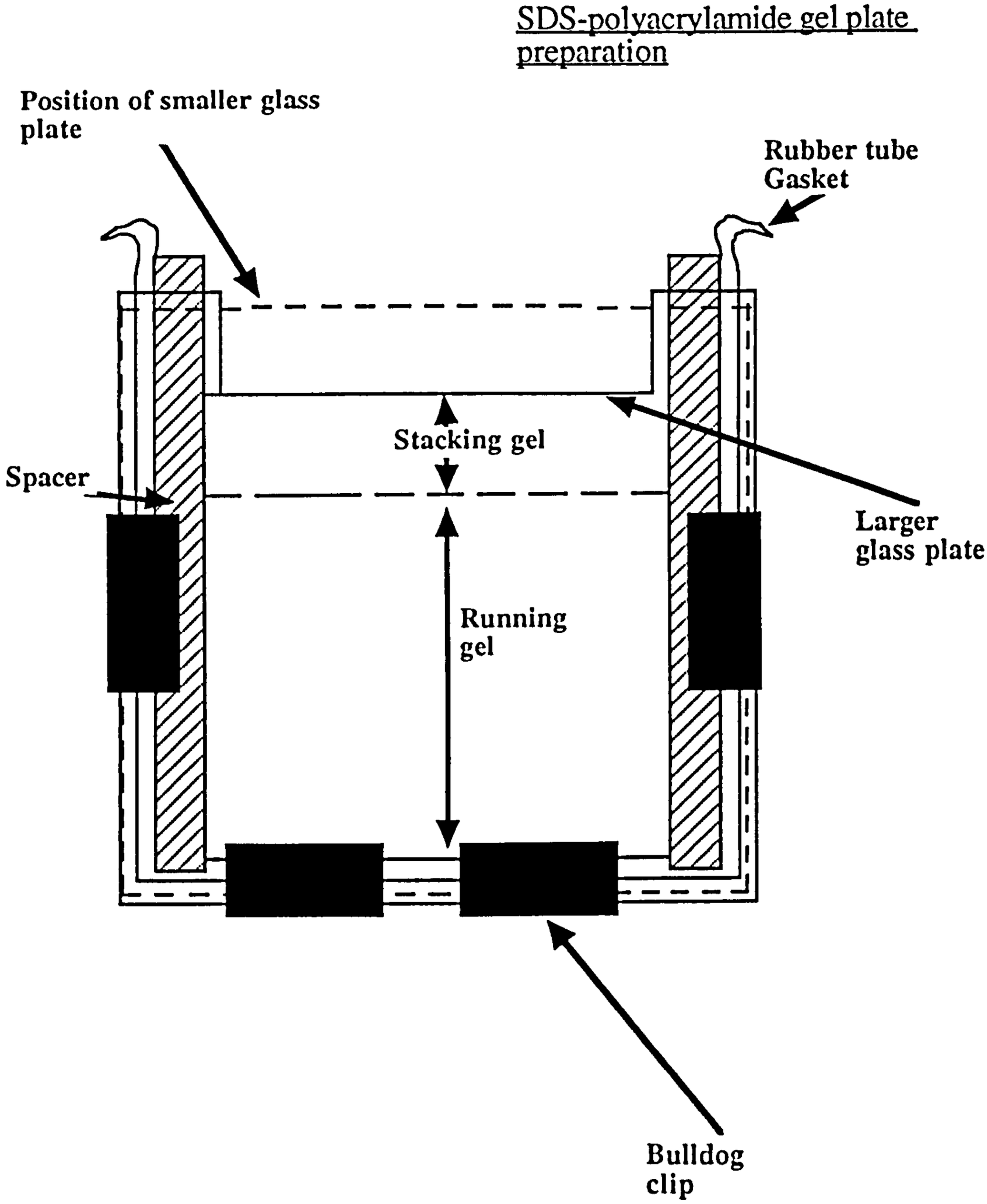
Sample (Periplasmic fraction or SDS fraction)	20 µl
Loading buffer	10 µl
Bromophenol blue	2 µl

Samples were mixed and boiled for 2 minutes then loaded into the wells (approximately 20 µl per well).

Protein Standard	Molecular weight(Dalton)
alpha-lactalbumin	14200
trypsin inhibitor	20100
trypsinogen	24000
carbonic anhydrase	29000
glyceraldehyde-3-phosphate dehydrogenase	36000
egg albumin	45000
bovine albumin	66000

Protein standards were mixed and boiled as before and loaded into wells at either end of the gel (20 µl total volume per well). The gel was run at 25 mA for 30 minutes until the samples ran through the stacking gel. The current was then increased to 40 mA for a further 120 minutes approximately. The gel was then removed from the plates and stained with Coomassie blue for 30 minutes. The gel was then de-stained with three washings of a 10% acetic/40% methanol mixture. The gel was then fixed in cellophane and heat-dried under vacuum. The presence of the xylanase was

Figure 1.7


SDS-polyacrylamide gel plate preparation
Position of smaller glass plate
Rubber tube Gasket
Stacking gel
Spacer
Larger glass plate
Running gel
Bulldog clip

elucidated by comparison of banding patterns from pN23 or B41 cultures with standards of known molecular weight.

Periplasmic Fraction

The Periplasmic fraction (total protein fraction) was analysed by SDS-page mini gel.

SDS fraction

20 µl of each SDS fraction (see Chapter 1.2 (vii)) was prepared for the SDS -page mini gel. Running of a mini gel using SDS fractions from avicel samples indicated how much (if any) of the xylanase bound directly to the avicel from 1.2(vi).

(ix) Large scale harvesting of xylanase protein.

4 x 0.5 litres of B41 and pN23 cultures were grown overnight for harvesting the following day. Periplasmic samples were prepared from these cultures for determination of the optimum amount of avicel necessary to bind B41 or pN23 xylanase forms. 5 ml samples of periplasm were mixed with varying amounts of avicel (see Chapter 1.2.(vi)).

avicel stock/ml	B41 periplasm/ml	pN23 periplasm/ml
1	5	5
3	5	5
5	5	5
7	5	5
10	5	5
15	5	5
20	5	5

SDS-page mini gel electrophoresis was performed on the SDS extracts. The intensity of the banding on the gel was directly proportional to the amount of protein that had bound to the avicel (see Figures 1.10 to 1.17).

Time course preparation of periplasmic fractions from cultures was also performed to determine the optimum time to harvest the cells.

(x) Guanidine extraction

Once the optimum amount of avicel necessary for xylanase binding had been determined, the periplasmic fractions could then be treated with guanidine hydrochloride. Guanidine does not denature protein - unlike SDS, so the xylanase enzyme retains its structure and activity. Periplasmic fractions from pN23 and B41 cultures were prepared as described previously and bound to the optimum amount of avicel.

Each avicel/periplasmic mixture was then resuspended in 1 ml 6M guanidine hydrochloride after washing with three times 10 ml 100mM tris pH 7.5. After one hour the mixture was spun down by microcentrifuge and the supernatant retained. Guanidine had stripped the xylanase protein from the avicel over this time. 5 litres of 100 mM tris pH 7.5 (60.55g of trizma base in 5 litres of double distilled water) was prepared, placed in a coldroom at 4^{o}C and mixed continuously using a magnetic stirrer.

6 inch portions of narrow dialysis tubing were boiled in double distilled water for 5 minutes. The dialysis tubes were then clamped at one end and each 1 ml sample of guanidine hydrochloride (containing the xylanase protein) was placed in separate dialysis tubes by Gilson pipette. The open end of each dialysis tube was then clamped, ensuring that any air bubbles were expelled from the tube. Each clamped dialysis tube was placed in the 5 litres of tris buffer and left to dialyse overnight. The 6M guanidine inside the dialysis tubing passed into the 5 litres of 100mM tris by osmosis, leaving the protein behind inside the tubing.

(xi) SDS-page mini gel analysis of dialysed protein extract

Protein extracts from dialysis (see Chapter 1.2 (x)) were examined as before using SDS-page mini gels.

(xii) Xylanase assays on dialysed protein extracts

Xylanase assays were performed on the dialysed protein extracts as before (see Chapter 1.2 (vi)).

(xiii) Transformation of B41 and pN23 into JM83 without the regulation of pNM52

As the level of protein extraction was found to be quantitatively low, the plasmids were transformed into JM83 without the regulatory multicopy pNM52 plasmids (see Appendix 1.1).

Starter Culture

5 ml L-Broth
50 µl Ampicillin stock
a wire loop of JM83 E.coli

Starter cultures were grown up overnight at 37°C and 100 r.p.m. on an orbiting shaker.

Religation

20 µl competent JM83 cells
10 µl plasmid (either B41 or pN23)
(see Appendix 1.1)

Transformants were plated out onto blue xylan agar (blue xylan stock 60 mg per 100 ml L-Broth) and grown overnight at 37°C. White colonies or "spots" indicated recombinants containing functional xylanase genes (the XYLA gene degrades xylan and the blue dye to a colourless product).
Putative colonies were picked and inoculated into 500 ml L-Broth under previous standard conditions. Cells were harvested as before and the periplasms were bound to avicel. The avicel was then resuspended in 10 % SDS as before. Banding patterns and corresponding banding intensities were examined.

Religation
200 µl competent JM83 cells
10 µl plasmids (either B41 or pN23++) (see Appendix 1.1)

Transformants were plated out onto blue xylan agar (xylan stock 60 mg/100 ml L-Broth/blue) guanidine hydrochloride 6M. The avicel and guanidine mixtures were spun down , the supernatant retained (first wash) and the avicel resuspended in 1 ml guanidine. This was repeated a further 2 times to obtain 4 washes of guanidine hydrochloride.

Each avicel pellet was then washed in 3 x 10 ml tris pH 7.5 100mM and resuspended in 1 ml 10% SDS - no protein should be left bound to the avicel. Samples were dialysed in 5 ml 100mM tris pH 7.5 as before. The protein samples were recovered from the dialysis tubing and prepared for SDS- page mini gel analysis.

(xiv) Cellulose column extraction of xylanase

The Guanidine extraction of xylanase proved difficult. Cellulose columns were prepared to allow the xylanase contained in the periplasmic fractions to bind to the cellulose.
5% w/v cellulose HL and 5% w/v cellulose HBS were prepared in 100mM tris pH 7.5 and autoclaved.
Mini columns of cellulose were prepared.Each column consisted of a plastic cylinder 4 inches long with an open neck and glass filter at a tapered base.

5 mls of each periplasm was added slowly to each column. The cellulose binding domain of B41 and pN23 xylanase proteins bound to the granular cellulose in the column while the filtrate passed through unbound. The filtrate was assayed for xylanase activity - it should be negligible, if the xylanase bound successfully to the column.
Once the periplasm had run through each column, the cellulose was washed with 4 x 5 ml tris pH 7.5 100mM. The columns were then removed and resuspended in 1 ml 10% SDS.
Samples were boiled for 10 minutes, spun down and the supernatant retained for SDS-page mini gel analysis. Protein assays were performed on the tris washes of the cellulose columns.

(xv) Protein assay of cellulose columns

The reagent for the protein assay detects the presence of polypeptide. The reaction mixture turns blue if protein is present or brown if there is no protein in the sample.
1 ml of Coomassie G-250 reagent was added to 0.1 ml of each tris wash (Bradford Method 1976). Further cellulose columns were prepared and samples fully hydrated by 2 x 500k centrifuge spins for 10 mins and resuspended in 100 mM tris pH 7.5.
Cellulose columns were finally washed and resuspended in 10ml 100mM NaCl rather than 10% SDS. The salt eluted protein samples from the cellulose columns were then dialysed in 4l 100mM tris pH 7.8 and an SDS page mini gel performed on the

recovered protein from dialysis.

1.3 RESULTS

(i) Computer Modelling

Initial computer modelling of secondary structure proved limited. Application of turn information and extensive minimisation of A and B regions of xylanase resulted in very linear three dimensional structures (see Figures 1.8 and 1.9). Although no homologous structures were available for comparison the two regions had begun to adopt some form of linear and spherical gross structure.

(ii) Cloning

The plasmids B41 and pN23 were successfully cloned into JM83 under the regulatory control of JM83. Bacteria that contained the B41 or pN23 plasmid were able to grow on media containing Ampicillin (Amp). Bacteria containing pNM52 as well as the xylanase containing plasmid were able to grow on media containing Tetracycline (TET^r)and Ampicillin. Hence bacterial colonies that were present on agar containing Amp and TET^r were those colonies that were picked for use in all xylanase cloning work

(iii) Periplasmic extraction

Periplasmic fractions of *E.coli* (JM83) containing the B41 or pN23++ plasmid under pNM52 regulatory control were prepared and subjected to SDS gel electrophoresis. A strong banding pattern was observed at 42 kD molecular weight for both b41 and pN23 samples. Banding was observed at 60 kD for pN23 only (see Figure 1.10).

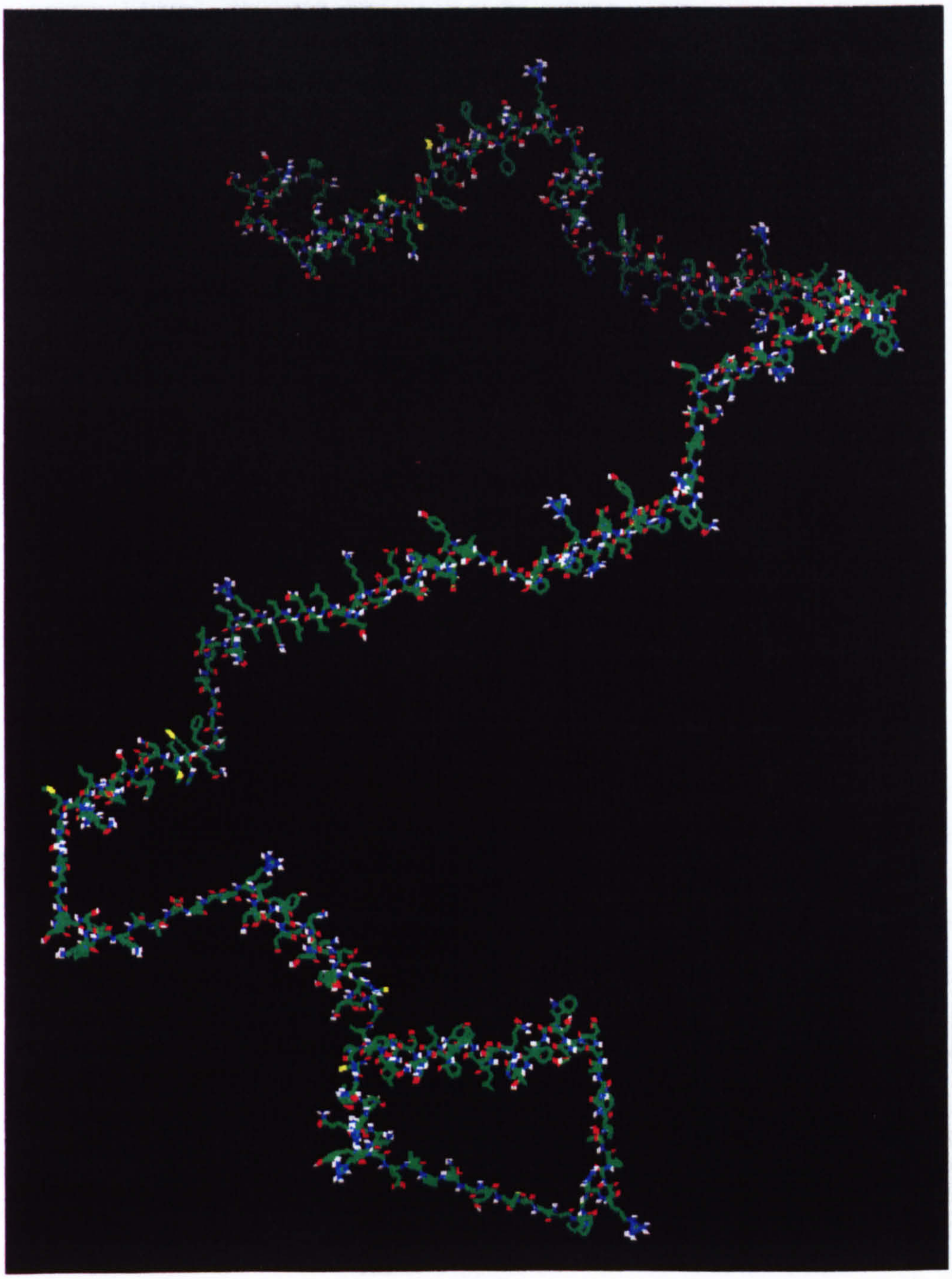
(iv) Xylanase assays

In order to make representative measurements of the xylanase activity of pN23++ and B41, xylanase assays were performed on either the periplasmic fractions or "sucrose wash" of *E.coli* in the mid log or stationary phase of their growth over various incubation periods. The basis of the assay is as follows:

The periplasmic fraction when incubated with blue xylan substrate (soluble in water), will cleave the blue xylan, releasing a blue dye, remazol brilliant blue, at

Figure 1.8

Figure 1.9



a rate proportional to the activity of the xylanase enzyme present. The reaction is stopped by addition of 2 volumes of DNSA reagent which terminate the reaction. Samples are assayed colorimetrically; the more dye liberated, the greater is the activity of xylanase in the periplasmic fraction (see Appendix 1.4).

(v) Avicel binding

SDS fractions from avicel pellets were prepared and subjected to SDS-page mini gel electrophoresis. This gel indicated qualitatively the amount of xylanase protein bound to the avicel from 1.2(vii). Again, banding was observed at the 42 kD for the B41 samples and 60 kD for the pN23 samples. A fainter band was observed at 42 kD in pN23 (see Figure 1.11).
Using varying amounts of avicel to bind to the xylanase (see 1(viii)), SDS fractions were prepared for SDS-page mini gel electrophoresis. From the banding patterns observed, the ideal amount of avicel needed to bind the maximal amount of xylanase was determined (see Figure 1.12 and 1.13).
From the gels, 5 ml of avicel was the optimum amount to bind 5 ml B41 periplasm and 3 ml avicel to bind pN23 periplasm.

(vi) Guanidine extraction and protein dialysis

SDS-page mini gels on dialysed guanidine extracts were prepared. From examination of the gels, almost negligible amounts of xylanase were successfully extracted (see Figure 1.14).
Xylanase assays on these extracts did indicate some xylanase activity.

(vii) Cloning without pNM52 regulation

As levels of protein extraction were quantitatively low, the B41 and pN23++ plasmids were transformed into *E.coli* (JM83) without pNM52 regulation. SDS-page mini gels were prepared as before. Much stronger banding patterns were observed at 60 kD and 42 kD, indicating quantitatively that more xylanase was present in this periplasmic fraction (see Figure 1.15).

(viii) Further Xylanase assays

Further assays were performed on the new periplasmic fractions of *E. coli* expressing the B41 or pN23++ plasmids without pNM52 regulation.

(ix) Cellulose column extraction

The SDS-page mini gels results gave very faint banding patterns for cellulose bound fractions (see Figure 1.16).

(x) Protein assays

Protein assays were performed on the cellulose washes to determine if all unbound protein had been flushed from the cellulose column.

	cellulose HL			cellulose HBS		
plasmids	B41	pN23	control	B41	pN23	control
1st wash	blue	brown	brown	blue	brown	brown
2nd to 5th wash	all solutions were brown after subsequent washes					

Repeated SDS-page mini gels on cellulose extracts revealed no bands whatsoever (See figure 1.17).

Figure 1.10

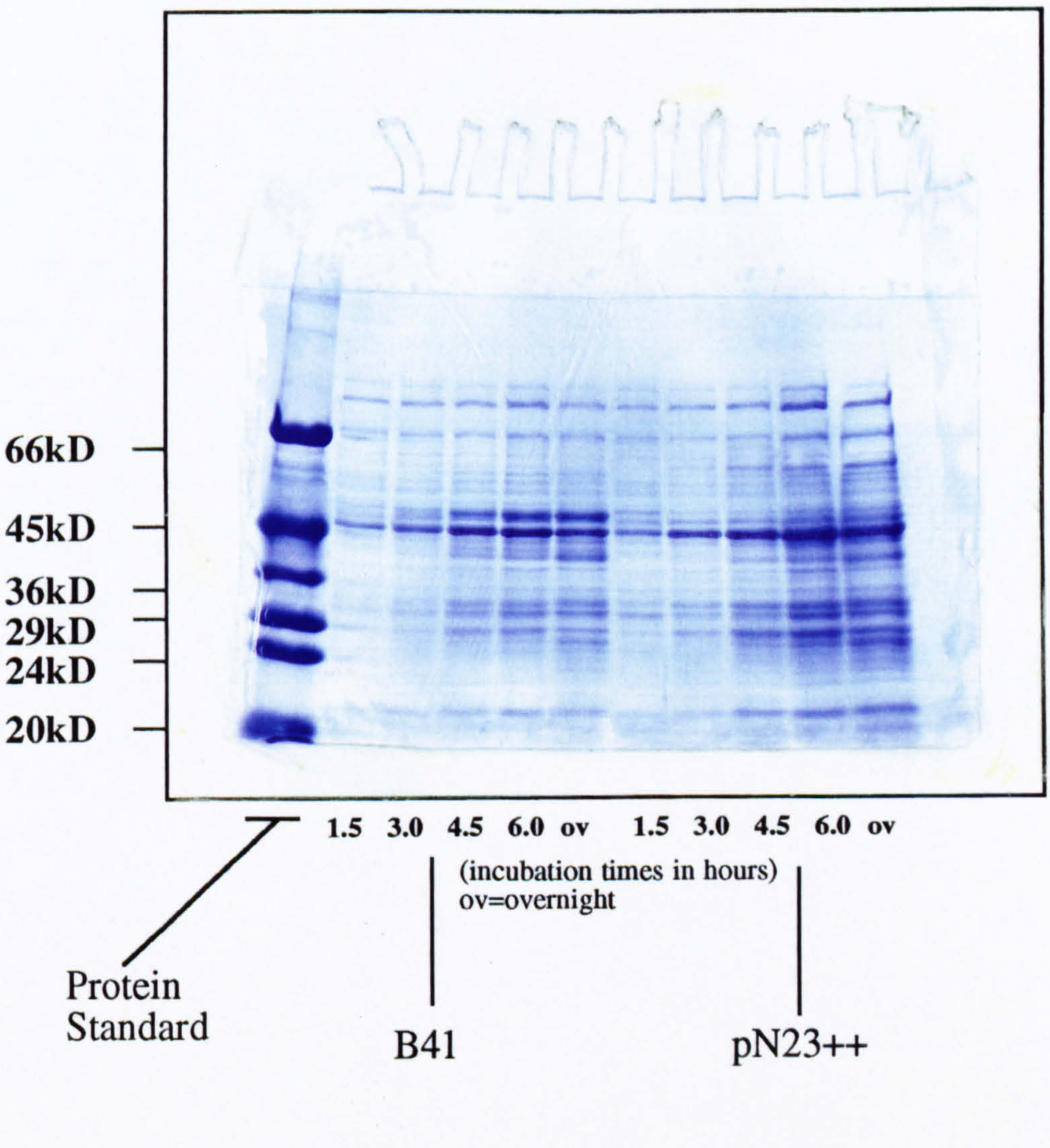

66kD
45kD
36kD
29kD
24kD
20kD
1.5 3.0 4.5 6.0 ov 1.5 3.0 4.5 6.0 ov
(incubation times in hours)
ov=overnight
Protein Standard
B41
pN23++

Figure 1.11

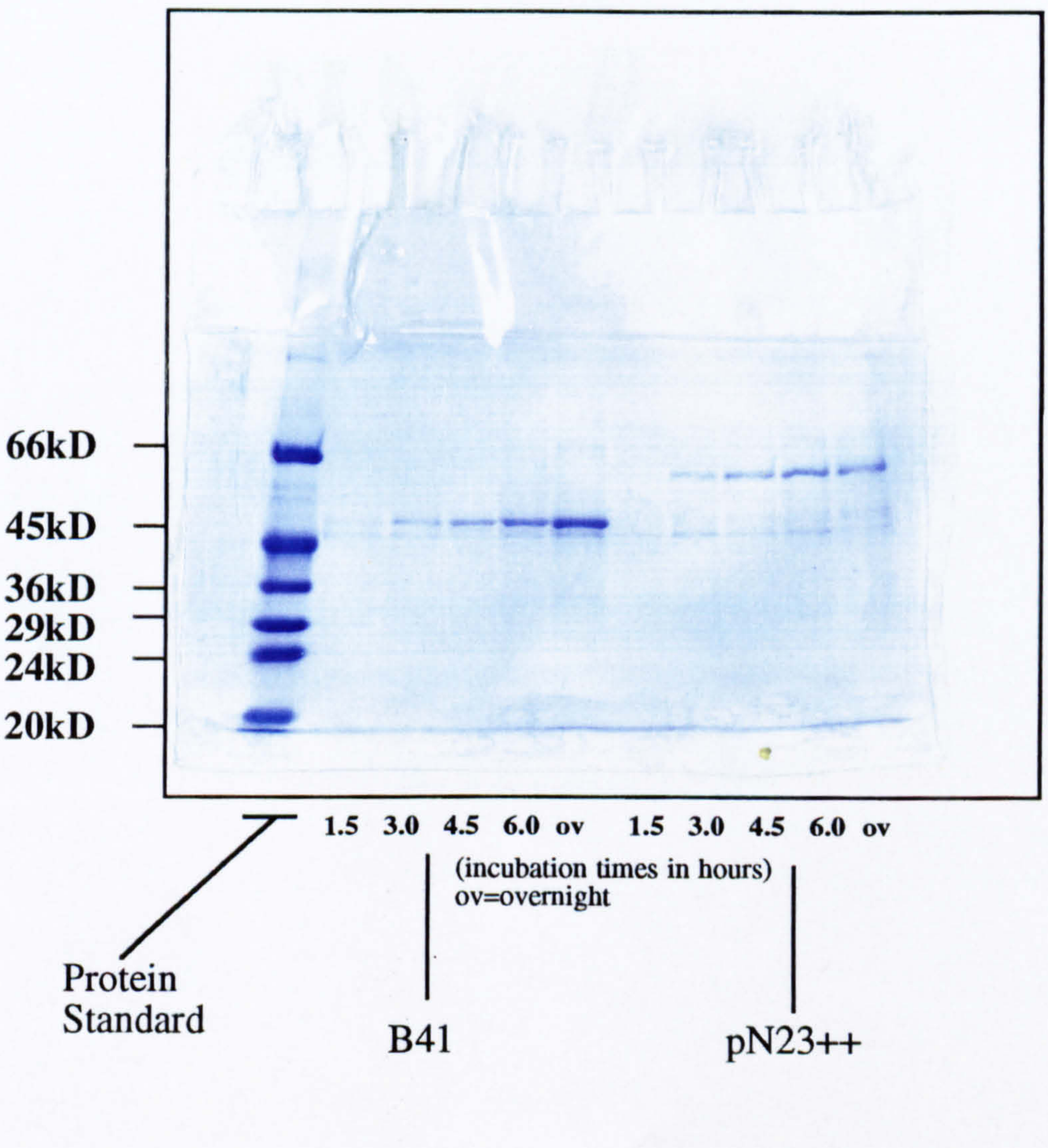

66kD
45kD
36kD
29kD
24kD
20kD
1.5 3.0 4.5 6.0 ov 1.5 3.0 4.5 6.0 ov
(incubation times in hours)
ov=overnight
Protein Standard
B41
pN23++

Figure 1.12

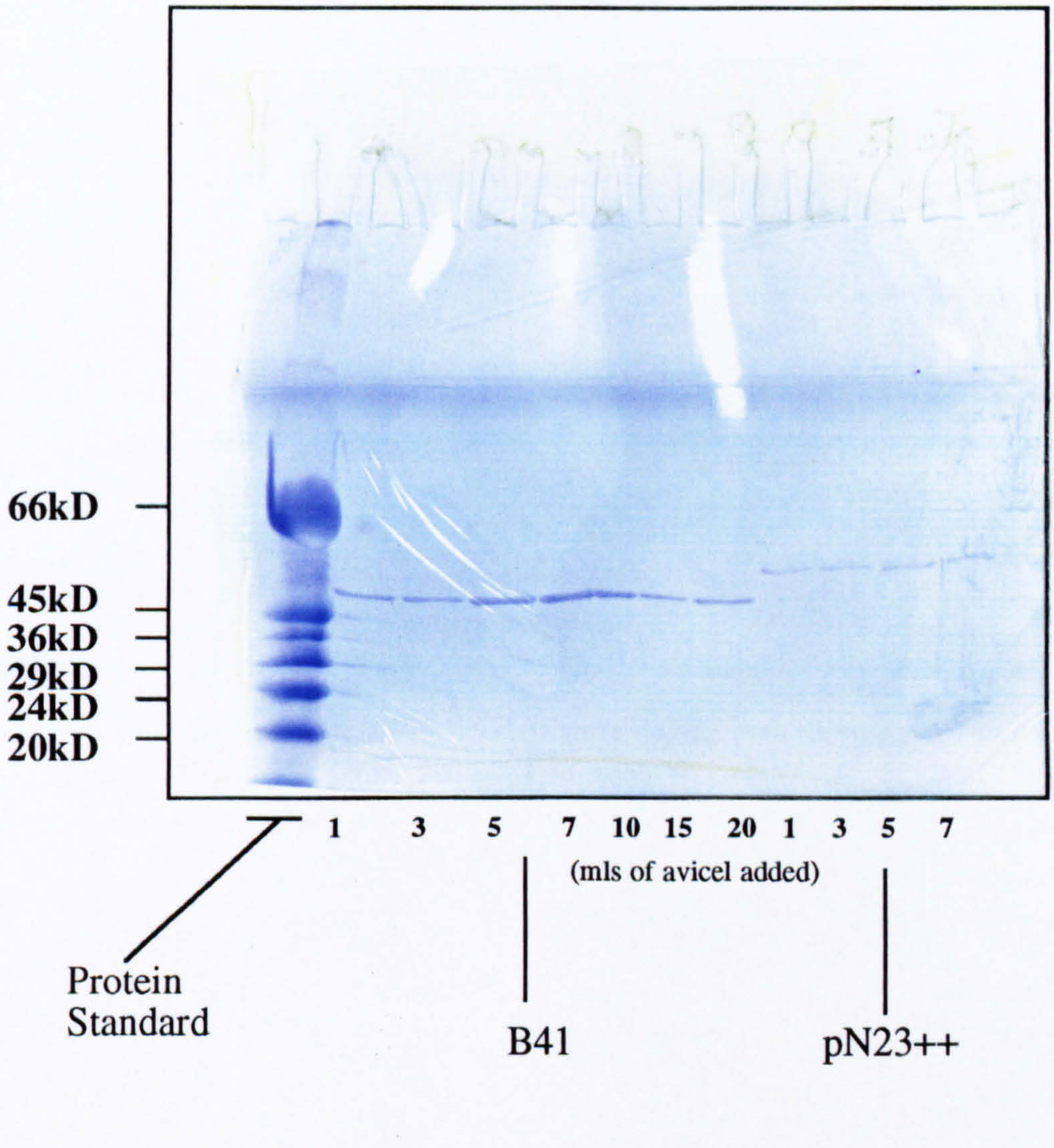

66kD
45kD
36kD
29kD
24kD
20kD
1
3
5
7
10
15
20
1
3
5
7
(mls of avicel added)
Protein Standard
B41
pN23++

Figure 1.13

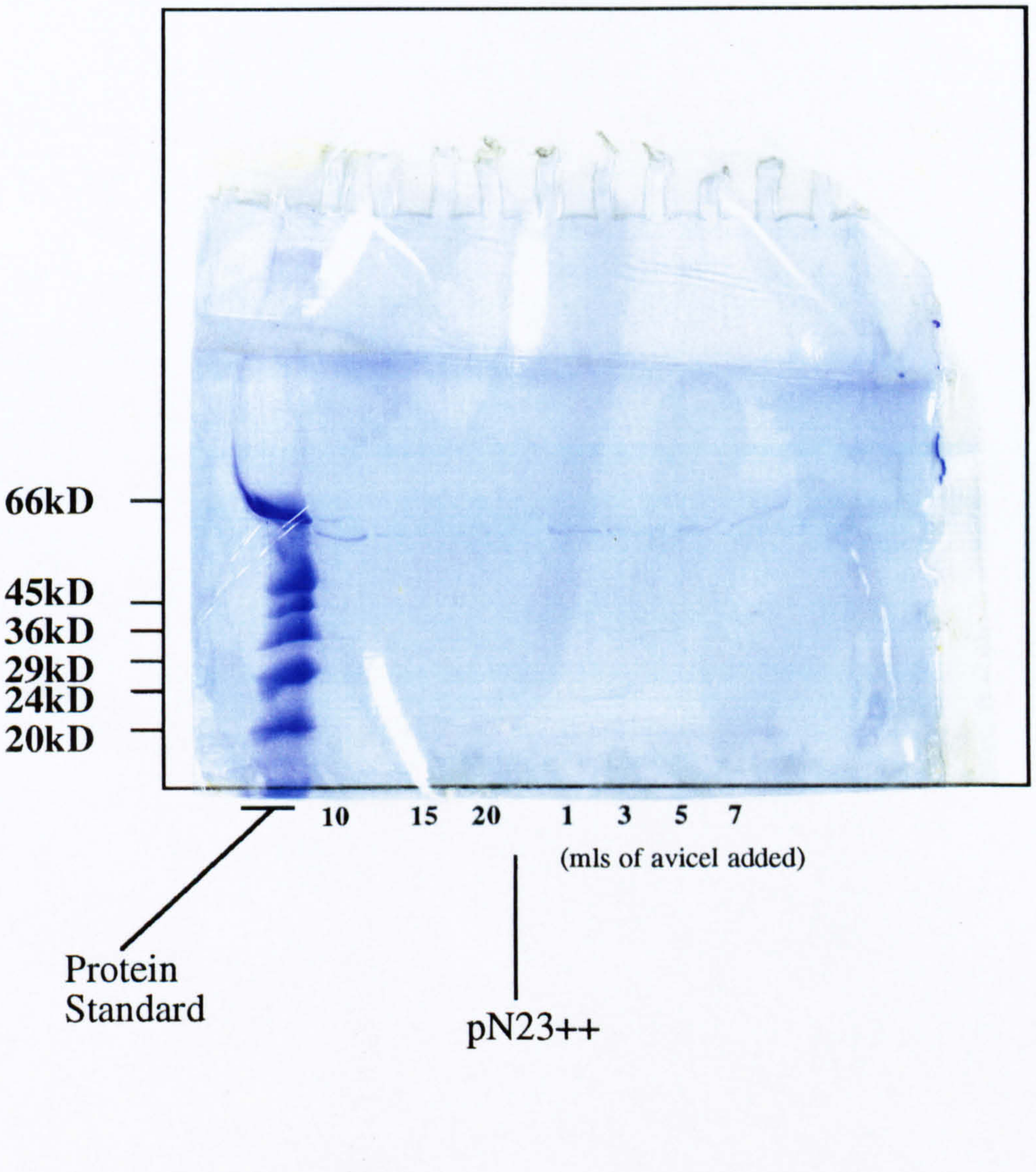

66kD
45kD
36kD
29kD
24kD
20kD
10
15
20
1
3
5
7
(mls of avicel added)
Protein Standard
pN23++

Figure 1.14

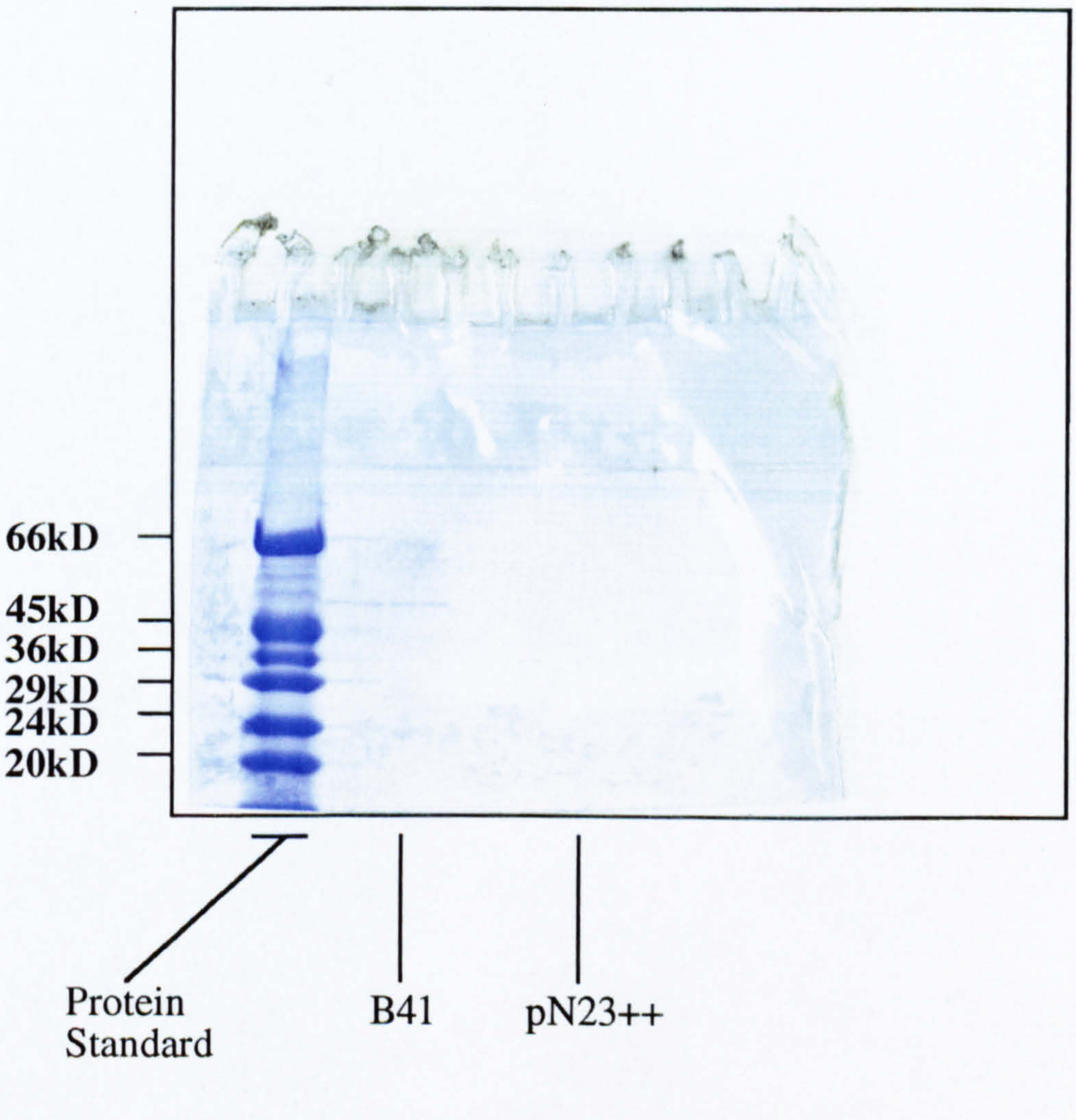


Figure 1.15

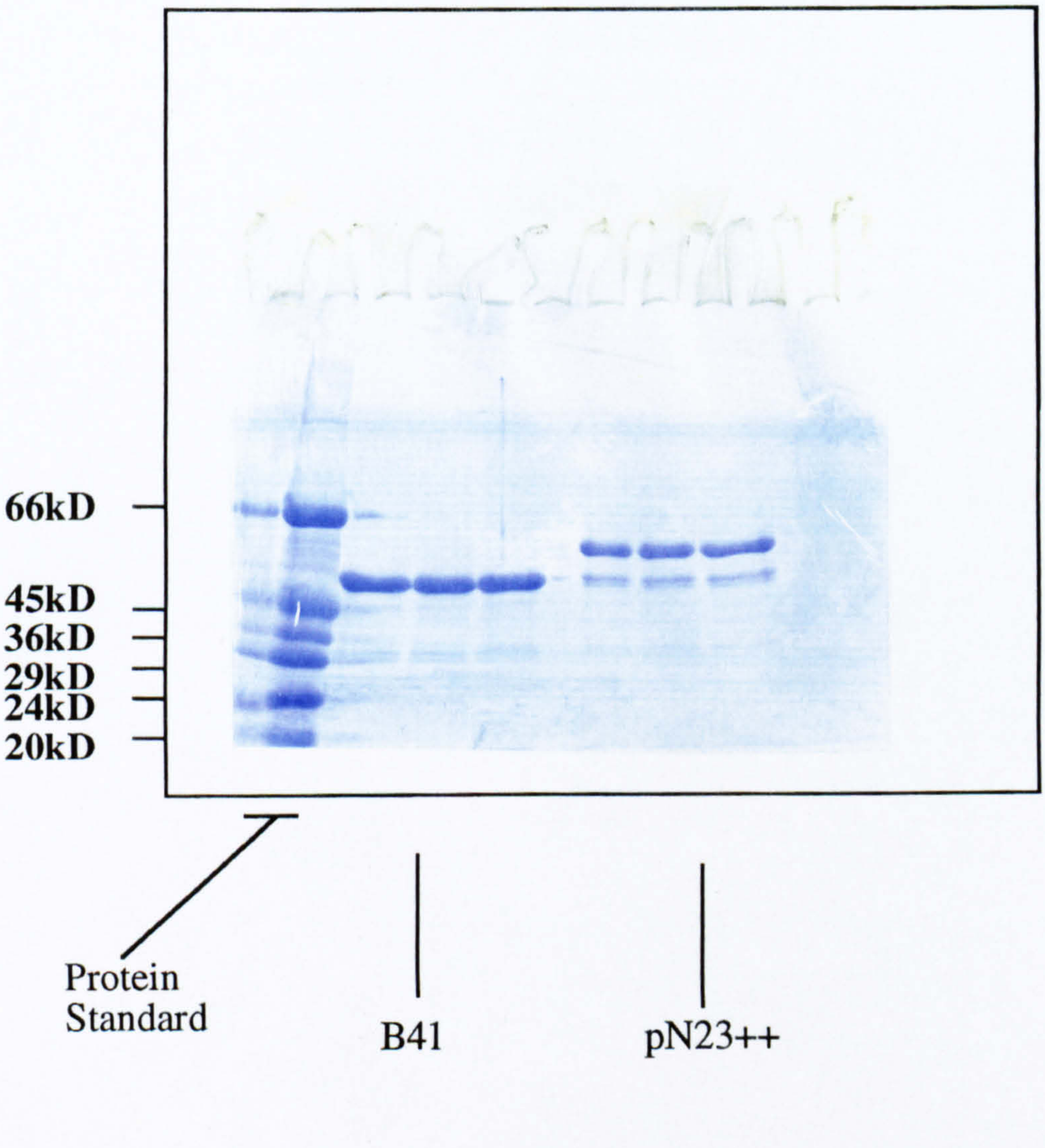


Figure 1.16

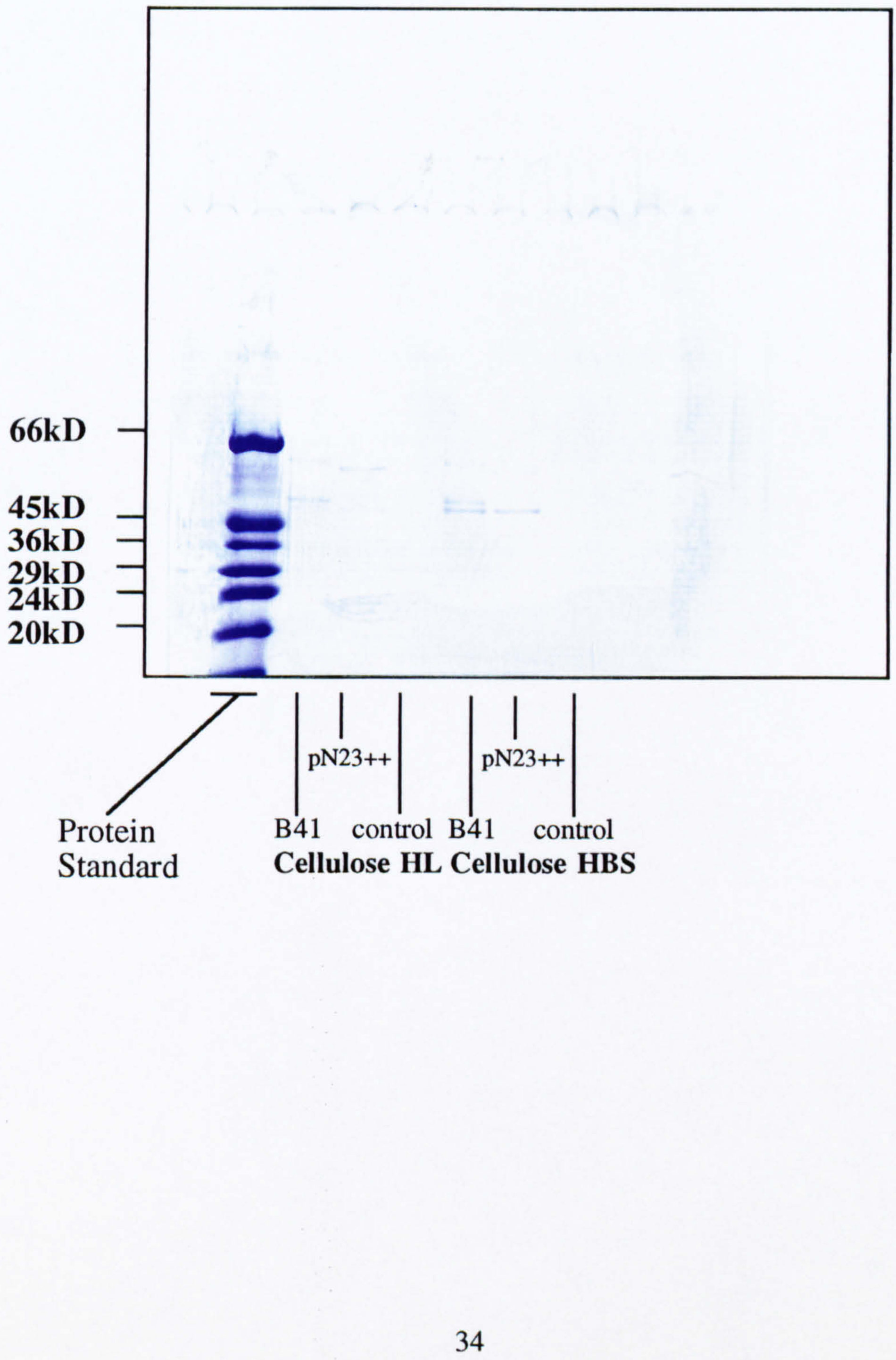
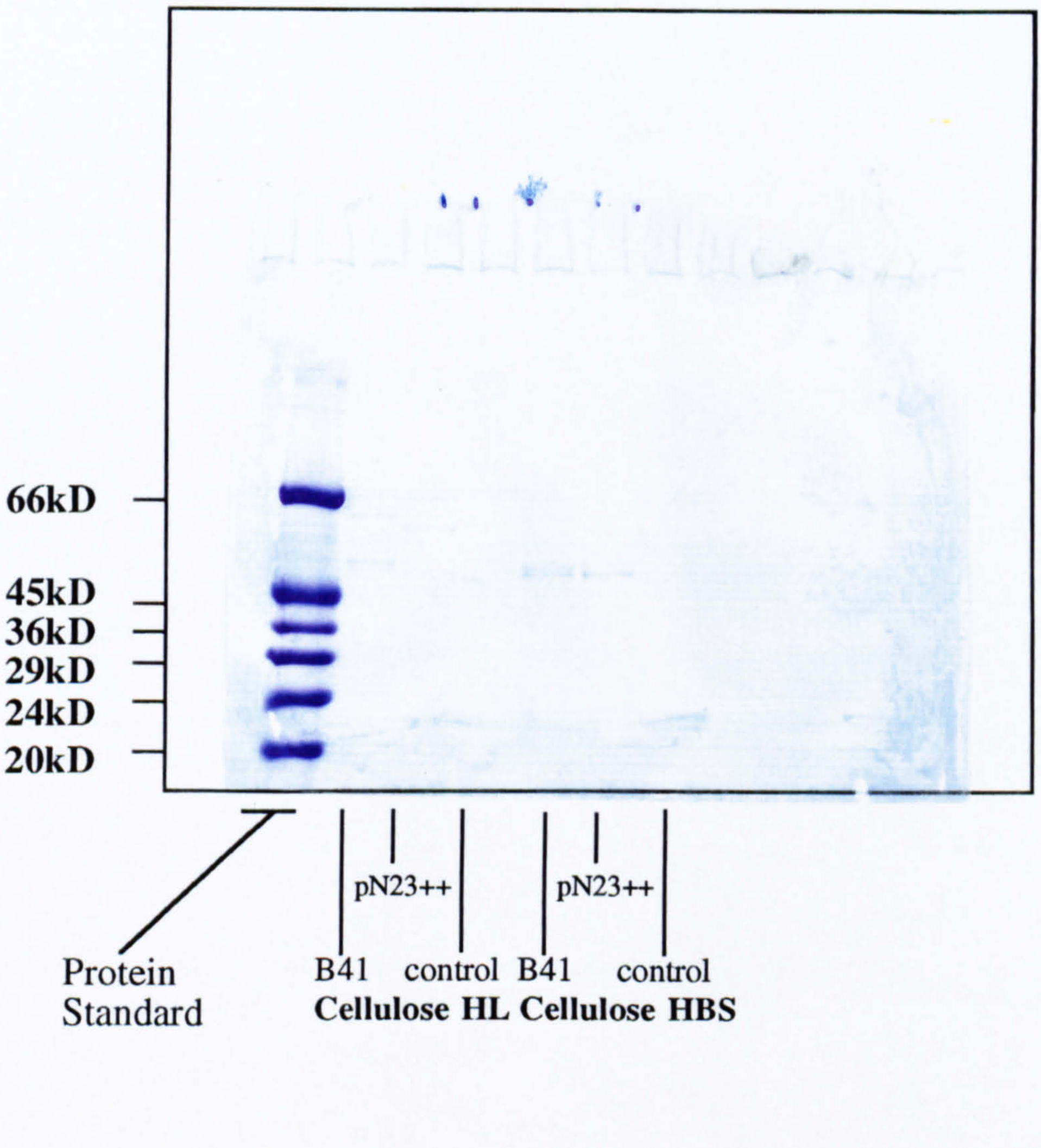

66kD
45kD
36kD
29kD
24kD
20kD
pN23++
pN23++
Protein Standard
B41
control
B41
control
Cellulose HL
Cellulose HBS

Figure 1.17


66kD
45kD
36kD
29kD
24kD
20kD
pN23++
pN23++
Protein Standard
B41
control
B41
control
Cellulose HL
Cellulose HBS

1.4 DISCUSSION

(i) Cloning of Xylanase

Once the XYLA gene had been successfully cloned into the JM83 *E.coli* (see Chapter 1(viii)a) there are many factors which affect the level of transcription and translation of the gene (see Figure 1.20).

The rate limiting step of gene translation depends on the binding affinity of RNA polymerase to the promoter sequence.[19] The DNA fragment encoding the XYLANASE protein is expressed in very low levels in native Pseudomonas. This is due to a weak promoter sequence in the native chromosomal DNA. To overcome low expression, XYLA was inserted into vectors under the control of an adapted Lac operon. The Lac operon has one of the strongest naturally occurring promoter sequences, the Lac p sequence.[20] This strong promoter increased the level of expression of the XYLA gene and ultimately protein synthesis. [3] However the lacP gene must be controlled in its function. Unregulated gene expression places too great an energy demand on host *E.coli* and cell death results. To overcome this lethal hyper-expression, the multicopy plasmid pNM52 was initially transformed into *E.coli* to regulate the XYLA expression. pNM52 has a plasmid copy number of about 30 in *E. coli* and contains the Lac I_q gene which synthesises repressor protein.[21]

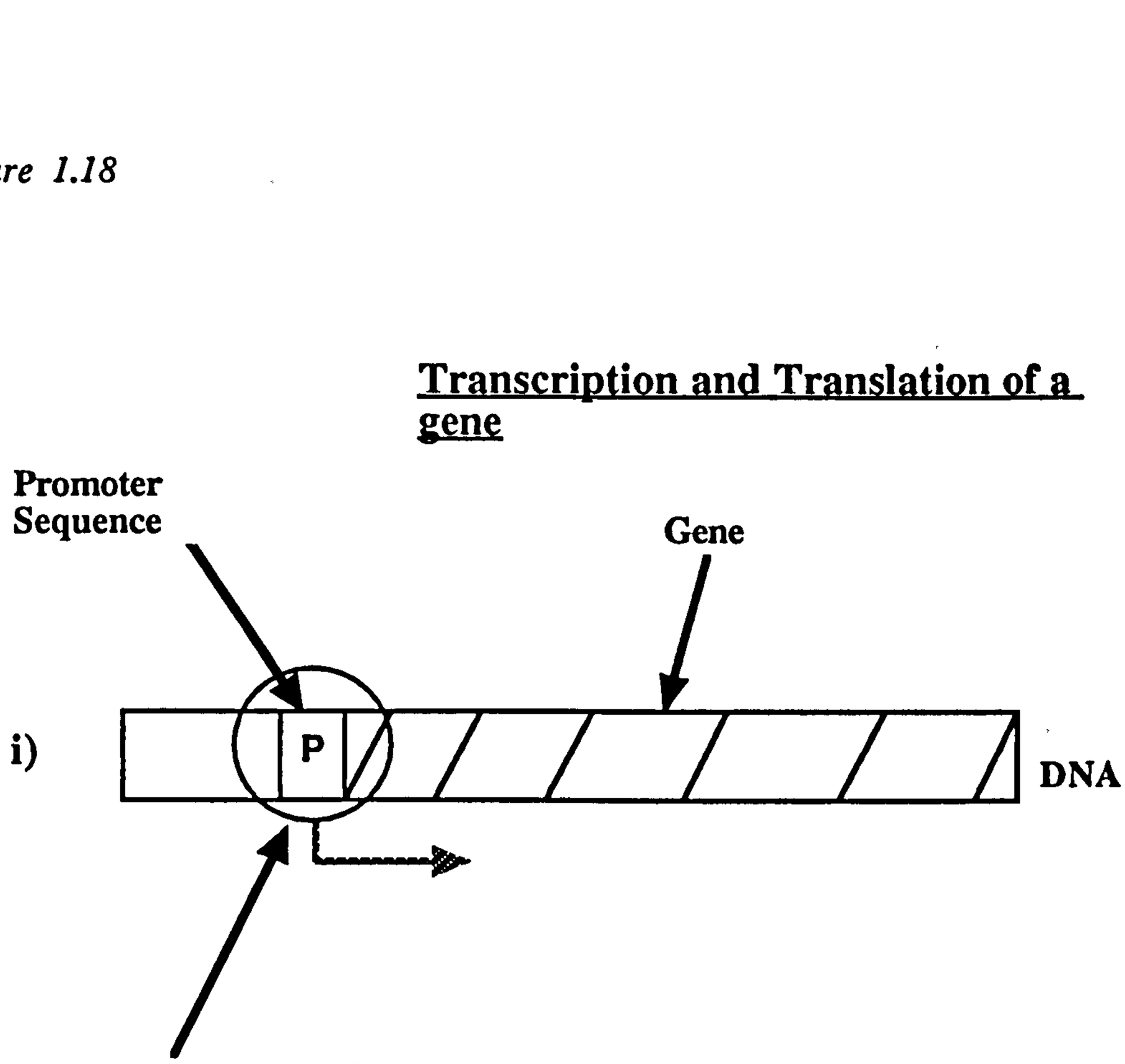
Normal *E.coli* strains contain only 10 molecules of repressor protein per cell, but *E.coli* transformed with pNM52 have 10x30 molecules of repressor i.e. 300. *E.coli* JM83 can harbour plasmid vectors up to a maximum copy number of 300. Therefore, in this work transformation of recombinant plasmids into JM83 *E.coli* with pNM52 gave 100% suppression of XYLA gene expression.
To activate host cell production of Xylanase, IPTG was used to induce gene expression through inactivation of the repressor protein . This IPTG activation was used to allow host bacterial cell numbers to reach maximal level before protein expression was initiated.

(ii) Xylanase Assays

This assay method gave a qualitative indication of the amount of enzyme activity present in the samples analysed. In cases of raw periplasmic fraction, considerable activity was indicated by the deep blue colouration of samples.[22.]

Figure 1.18

Transcription and Translation of a gene



Assays on samples that had undergone various methods of protein extraction did not display such marked relative activity. Several methods of protein extraction involved temporary denaturation of protein (such as the guanidine method) and the subsequent refolding of the secondary structure to form an active protein. It was clearly the case that much of the original active enzyme was unable to reform into the correct tertiary structure and hence maintain activity.

(iii) Extraction Techniques

Periplasmic extraction breaks open *E.coli* cells in two stages. Initial resuspension of a bacterial pellet in sucrose and EDTA partially breaks down the bacterial cell wall - in effect, the cell "leaks" small fragments of protein through holes in the bacterial cell wall. Vigorous resuspension and centrifugation completely breaks down the bacterial cells, releasing cell proteins (the periplasmic fraction). The periplasmic fraction contains the raw protein contents of the bacterium. In this work, the periplasmic fraction contained the xylanase protein expressed by B41 and pN23.
This extraction technique for obtaining "raw" protein was successful in that activity of samples were maintained. The xylanase enzyme was not subjected to any chemical attack in this technique and as such, maintained activity.

a) Avicel binding

Periplasmic fractions obtained from B41 or pN23 cultures bound successfully to avicel. However, to recover the protein bound to the avicel further treatment was necessary. Sodium dodecyl sulphate (SDS) used to strip the xylanase from the avicel completely denatured the protein, so the SDS procedure only indicated the amount of protein present in the periplasmic fraction. Binding studies using varying amounts of avicel to bind to periplasmic fraction indicated a 1:1 ratio of avicel to bind B41 periplasm and 3:5 ratio for pN23. This was a quantitative result taken from analysis of the intensity of the banding patterns present in the SDS-page mini gels. Once the ideal amounts of avicel required for periplasm binding had been elucidated, a separate method for extracting the xylanase without denaturation was employed.
As an extraction technique, this was unsatisfactory as little active xylanase was recovered from the avicel. SDS is an extremely powerful detergent and although it can strip avicel completely of bound protein, the protein structure is not maintained.

b) Guanidine extraction

It was reasoned that the high concentration of guanidine hydrochloride (6M) used in this method of extraction would strip the xylanase protein from avicel without significant denaturation of the protein. The enzyme would therefore retain its catalytic activity and tertiary structure. Analysis of the SDS-page mini gels showed only faint banding patterns corresponding to any xylanase protein washed from the avicel by guanidine. As very little xylanase was effectively removed from avicel by this method, the cellulose columns extraction technique was used.

c) Cellulose column extraction

From analysis of the SDS-page mini gels, this method of xylanase extraction from periplasm proved least successful. Again, only faint banding patterns were observed on the gels, indicating that negligible amounts of protein were binding to the cellulose column. Further repetition of cellulose columns using NaCl 100 mM to elute the protein also proved disappointing.

(iv) Xylanase protein recovery

It is clear from this work that the methods employed in the recovery of active xylanase protein were unable to provide sufficient quantities for further analysis.

Xylanase is a large and consequently a structurally complex enzyme. Most laboratory techniques for protein extraction involve some degree of tertiary structure degradation and subsequent protein refolding after recovery. It would therefore be logical to assume that xylanase protein is unstable under these extraction conditions and cannot be recovered by these methods. It should also be noted that the *E. coli* used to produce the full form protein were subjected to accelerated synthetic rates. This may have lead to incomplete or incorrect protein transcription/translation within the host cell.

The most likely avenue of further work in xylanase protein recovery is to concentrate on crystallisation of the protein and hence determine its crystal structure.

(v) Modelling of protein structure

There are many methods available for evaluation of protein conformation. Solvent accessibility can be used to predict the probability of an amino acid being buried or exposed in a protein.[23] The degree of ideal crystal packing indicates the arrangement of residues in a protein.[24]
Distance geometry methods allow constraints to be placed on atoms and residues thus defining relative positions within protein structure. Knowledge-based prediction of structure is most commonly used. These methods involve correlation of amino acid properties and amino acid sequence. There are three algorithms that are most widely used. The first two rely on parameters obtained from analysis of fully elucidated sequences and corresponding structures.[14-15,25-26] The third method is based on stereochemical criteria.[27]
Minimum energy calculations (as in this work) allow the protein sequence to fold and adopt a minimum energy conformation. There are two problems with this approach. Firstly appropriate parameters defining energy contributions of individual atoms and residues must be unified. There are many software packages available, each with a separate forcefield and parameter set. The second problem involves negotiating multiple minima (by crossing saddle points) in energy minimisation. Several algorithms cannot push the energy minimisation through a local energy well and consequently the "global" minimum is never reached. What is needed is a unified forcefield and parameter set to define all possible atomic interactions. Energy calculations are being constantly refined, but minimisation schemes have so far failed to predict chain folding accurately.[28-29] This has been seen in this work using the limited turn information.

Systematic searching of all possible conformations of sequential residues is another method of structure prediction.[30] Sections of protein (up to 6 residues) are subjected to conformational search and analysis. Each conformation is screened using the electrostatic energy, the exposed hydrophobic area and a solvent screening term. This is a computationally intensive approach with hundreds of possible structures generated from only a few selected torsions.

Class prediction of protein structure is based on division of a known set of crystal structures into four categories, α, β, $\alpha+\beta$ and α/β. These correspond to amino acid sequences of high α content, high β content, mixed α and β content and disordered structures.[31]

A more recent approach involves the use of neural networks which learn from existing structures how to predict the secondary structure of small sections of sequences.[32] The neural network approach only achieves predictive success rates of around 65%.[33] Introduction of tertiary structural information (such as NMR distance constraints) is necessary to improve prediction rate.
Other recent predictive methods have been developed to predict structure based on hydrophobicity of individual amino acids.[34-35] These "antigenicity values" are currently amongst the most popular methods of prediction.[36]
Finally, there are expert systems. These are customised computer programs specific to individual molecular systems. Each structure is minimised to a global minimum as quickly as possible, then database searching is used to compare homologous structures. Computer modelling is then used in conjunction with external data for each molecular system to provide a model for the protein structure.[35]
Current methodologies for tertiary structure prediction from primary and secondary data do not enable us to predict a protein's conformation.[36]
The simple algorithms available for prediction of secondary structure from raw sequence data have degrees of accuracy varying from 40% to 70% successful prediction (note that 50% predictive success constitutes a random guess!).[37]
Prospects for tertiary prediction are encouraging. There are three approaches in use for prediction of tertiary protein structure.

The first involves the use of sequence homology where peptides of known three dimensional structure are compared with homologous sequences. The homologous sequence is then mapped onto the tertiary structure of the known structure and refined using minimisation. The second method uses secondary structure prediction of short sequences of amino acids. These short units are then assembled into a compact structure.
These first two methods rely heavily on the data base of known crystal structures.[38-39] As the Brookhaven database contains relatively few well-resolved structures at present (around 200 solved to a resolution better than 2Å) this method is dependant on the improvement of X-ray crystallography techniques and NMR spectroscopy and the increase in published structures. Once the database is sufficiently large and varied, predictive schemes will improve as a consequence.
The third method is not reliant on this data. Empirical *ab initio* energy functions can be used to derive the tertiary structure at minimum energy potential.

Trial structures are evaluated and either rejected or accepted for further refinement.

Research is continuing into the modelling of tertiary protein structure. Ultimately, the identification of any predictive motifs is limited by the protein structure database size.[12] Many more data will be needed before protein structure can be precisely predicted. However, with novel X-ray crystal structures constantly being resolved, it may not be long before the database size will be large enough to perfect predictive schemes.

(vi) Importance of crystal data

The methodology of X-ray crystallography of proteins is a complex one and is not covered in this work. The reader should refer to a specific review available on this subject.[40-41] It should be stressed, however, that perhaps the greatest hurdle in protein crystallography is the growth of sufficiently large, high quality crystals.

This process can be the rate-limiting step for many research groups. Primarily, large amounts of pure protein are required for crystallisation - typically gram amounts. This is where recombinant DNA technology can provide the raw protein (both full-form and mutant enzymes).

Harvesting and purification of the raw protein is crucial to the success of such projects. Finally, even if all the previous criteria are fulfilled, the crystals must be grown. This can take anywhere from a month, one to two years and sometimes not at all.[12] In difficult cases, it might be possible to grow protein crystals complexed with a substrate or suitable inhibitor (see Part 2). Once suitable crystals have been obtained, sophisticated hardware and computer software must be employed to collect and process the crystal data.

(vii) Xylanase as a target for protein modelling

In terms of both scientific interest and commercial application, this enzyme should still be actively considered for further work. Xylanase is thought to be in the structural class of proteins known as 8-fold α/β–barrels.[36] Many proteins with little sequence homology adopt this conformation, indicating that amino acid composition and sequence does not necessarily dictate tertiary structure. Suggestions have been made that many α/β–barrel proteins evolved from a common ancestor.[42] Xylanase has no structurally-resolved related proteins, thus making it an ideal candidate for further study of the α/β class of proteins.

Xylanase is also unique amongst the endoxylanases in that is has a markedly high amount of SER residues in its sequence. Little is yet known about the function of these serine-rich areas of sequence which are highly conserved.[5] Further work must be pursued to determine the role these SER residues play in xylanase biological function.
On a commercial level, xylanase may well have a future in transgenic livestock or as a silage innoculant.[43] Elucidation of structure and modelling of specific active site mutations will lead to enhancement of specific activity and increased knowledge of reaction mechanism.

(viii) Future work

Perhaps the best approach is to combine all available data from as many sources of information as possible. Protein class (either globular or linear) can be established using circular dichroism. [44] Flexible pattern matching and sequence motifs can be used to predict secondary loops and turns. Since this work was completed, the alpha carbon backbone of region B of xylanase has been reportedly elucidated by X-ray crystallography at the Reading Institute (unpublished). QUANTA/CHARMm has automatic sidechain placement functionality, whereby sidechain residues are fitted to protein backbone. Any NMR distance constraints are applied and the residues are subjected to simulated annealing. This process minimises each sidechain group and performs local dynamics simulation. This allows the residues to occupy all possible areas of conformational space and adopt the lowest energy structure.

(ix) Summary

Current methodologies for modelling an unknown protein structure are limited by the amount of existing data from elucidated structures. With time, these limits will be surpassed as the crystal databases expand to encompass all the different classes of protein. This part of the research work was highly ambitious and tackled many of the issues that are faced by Protein Modelling groups. At this point in time, we are still dependent on the valuable data from the x-ray crystallographer in order to understand more about protein structure and function. It is my firm belief that in the next 20 years we will be able to have a significantly greater level of automation in the elucidation of novel protein structures. Computational instrumentation is still improving, and the computer-modeller has increasingly more and more powerful hardware and software at his/her disposal. When one considers that the computer equipment used in this

work has an equivalent today eight times more powerful, then one is confident that more elegant and sophisticated algorithms will be developed. These predictive tools will greatly aid the scientist in the pursuit of Nirvana - tertiary structure from primary amino acid sequence.

PART TWO
THYMIDYLATE SYNTHASE AND NOVEL DRUG DESIGN

2.1 INTRODUCTION

(i) Thymidylate Synthase

Thymidylate Synthase (TS) has been regarded as a key enzyme in cancer chemotherapy for well over a decade.[45-46] TS catalyses the reductive methylation of 2'-deoxyuridylate monophosphate (dUMP) to 2'-deoxythymidylate monophosphate (dTMP) with the accompanying conversion of 5,10-methylene tetrahydrofolate (5,10-met THF) to 7,8-dihydrofolate (7,8 DHF).[47]

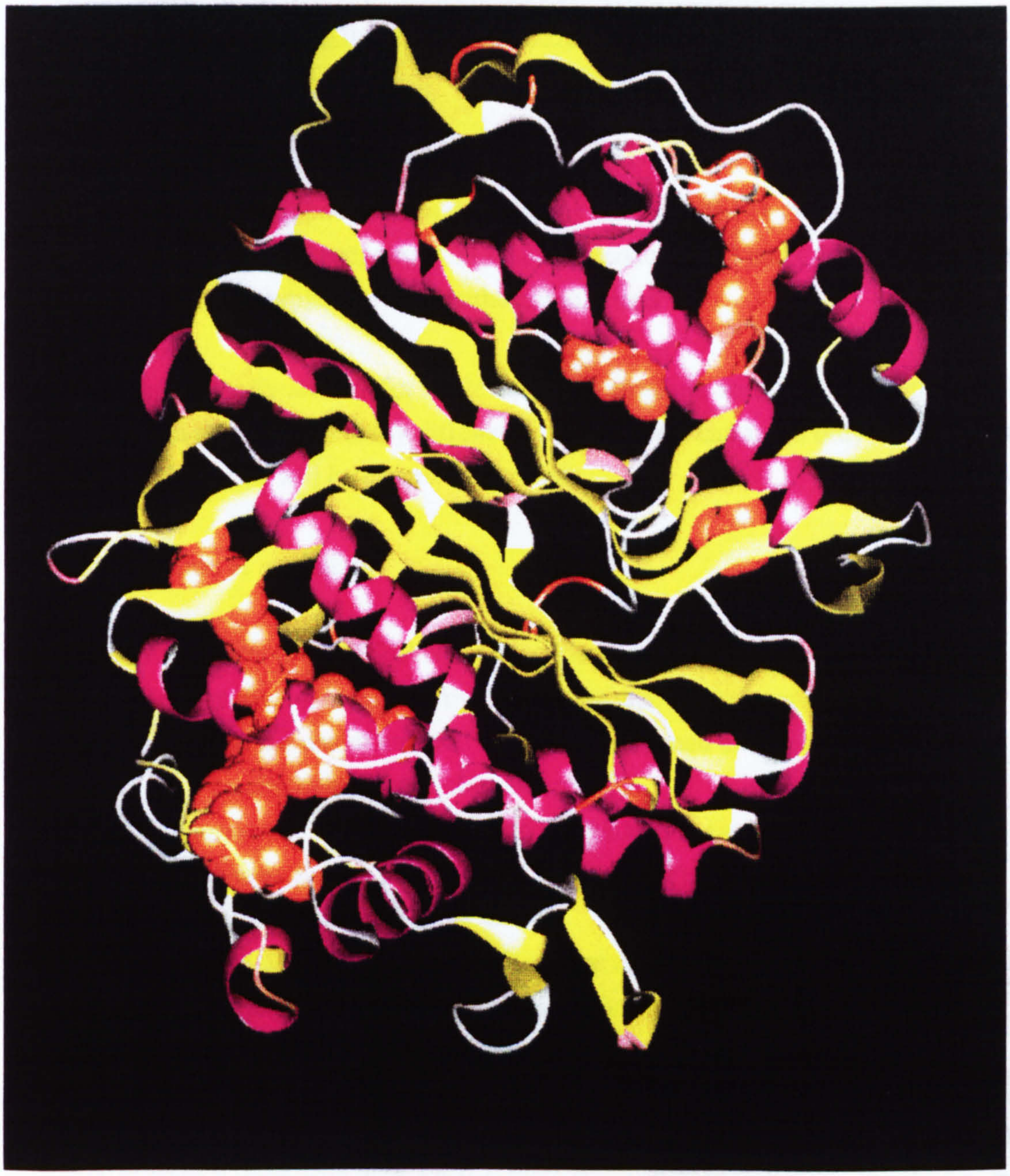
Blocking this dTMP synthesis results in "Thymineless death",[48-49] where rapidly proliferating cells become starved of the DNA necessary for cellular division. Protein and RNA biosynthesis is unaffected by TS inhibition thus making this enzyme a seemingly attractive target for cancer treatment.

Throughout the 1980's considerable effort was put into development of a potent antifolate, one of the most notable being N^{10}-Propargyl-5,8-dideazafolic acid (CB3717).[50-55] Phase I and phase II clinical studies were carried out using CB3717 with varying degrees of success.[56-59]
Although CB3717 proved a very tight binding inhibitor of TS and overcame methotrexate resistance, the drug was found to display renal toxicity. This renal toxicity was due to poor aqueous solubility and human studies were discontinued. Recent studies have involved modification of the CB3717 to improve solubility while maintaining cytotoxity to cancer cells.[60-64]

(ii) Thymidylate Synthase Crystal Structure

The structure of Escherichia coli thymidylate synthase complexed with the substrate dUMP and an analogue of THF (CB3717) was solved and refined at 1.97 angstroms by Robert M. Stroud and his group at the University of California in 1989.[65-66] This crystal structure was kindly made available for analysis.
TS exists as a dimer with a single active site in each subunit. Stroud crystallised TS in two forms; a "reduced" (major) form and "oxidised" (minor) form. The major form of TS contains dUMP covalently bound to cysteine in both dUMP sites and in the presence of CB3717 (see Figure 2.1).

Figure 2.1



One active site of the minor form contains dUMP noncovalently bound and in the presence of CB3717 while the other active site contains only inorganic phosphate and CB3717.
The active site of TS is a large cavity which binds CB3717 in two possible conformations. One is seen in the major form and one in the minor form (see Figure 2.2).

(iii) Computer modelling using QUANTA/CHARMm

a) Overview

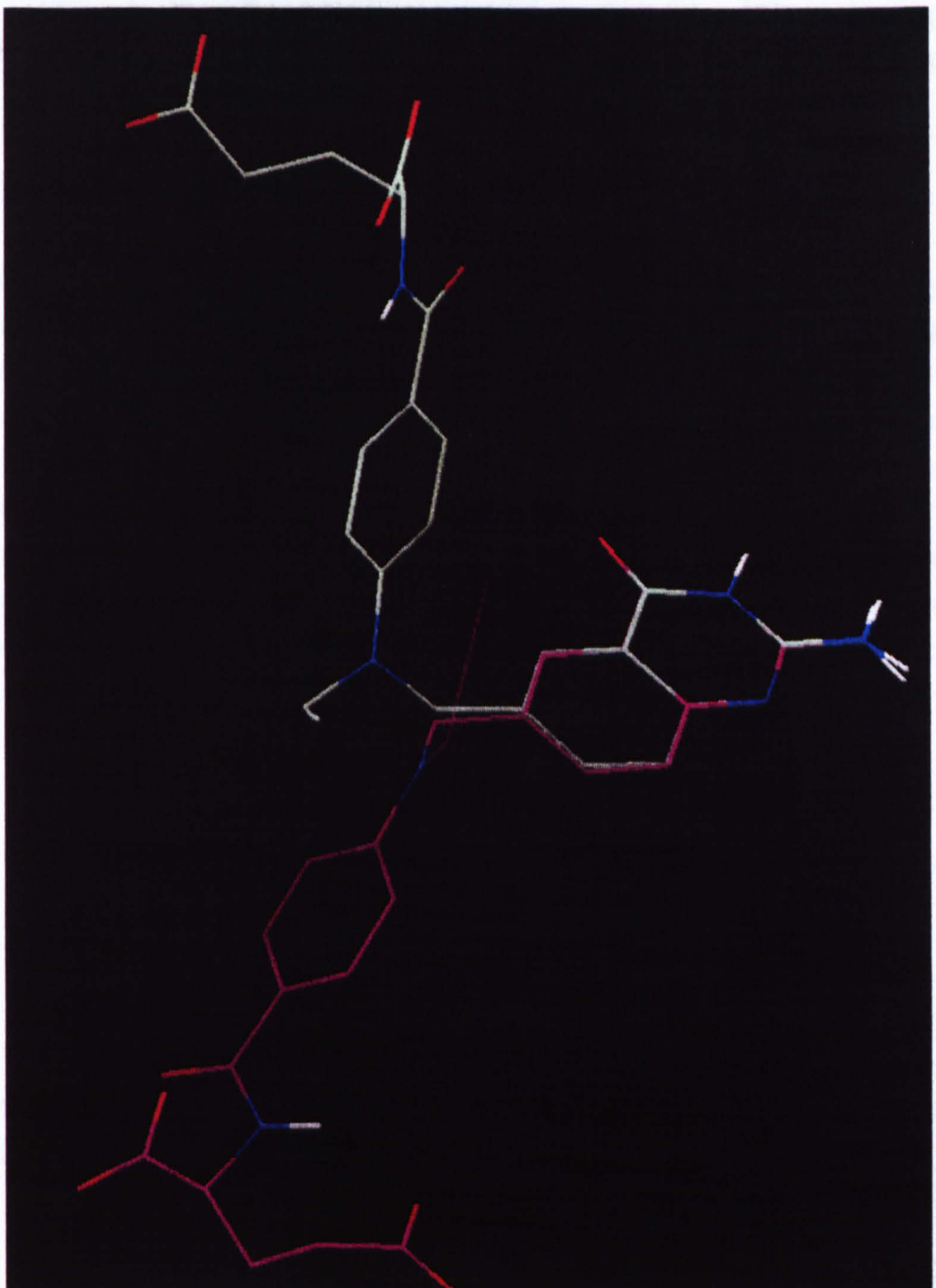
The three-dimensional coordinates provided by Stroud acted as the starting point for all modelling work.[66] The programs CHARMm (version 21.2) and QUANTA (version 3.0) were used as supplied by Molecular Simulations.[68]

CHARMm is a software tool for modelling both structure and function of molecules. The program can perform many different computational tasks, but only the functions utilised in this work will be discussed in depth:

b) Architecture of modelling

CHARMm consists of FORTRAN source code which can manipulate data in two ways: either directly, input created by the user fed into the program via a command stream, or through the use of external files that are opened and closed with computational results written to another external file.
Crystal Structures can be read directly into CHARMm for analysis but where definitive Cartesian coordinates are not available, the information must be generated within CHARMm using the standard parameter set .
This parameter set defines the atomic masses, partial charges, bonds, angles, dihedral angles and improper torsional angles between atoms. From this parameter set, a residue topology file (RTF) is created with the correct molecular topology for each individual residue in the molecule. In the case of large molecules, in order to reduce computational time, an extended atom representation is used in the RTF. An extended representation specifies methyl, methylene and methine groups as single 'atoms'. The remaining aliphatic hydrogens are not simulated.

Figure 2.2

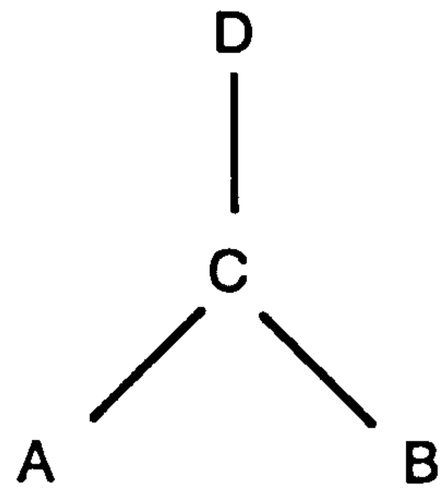


Only polar hydrogens are defined i.e. those hydrogen atoms bound to oxygen, nitrogen, a carbonyl carbon or an olefin carbon.[68]

Once all the RTF's have been created, the primary structure file (PSF) is generated and read into CHARMm.
The PSF only contains molecular topology information and partial atomic charges. All other data such as bond angle or force constraints are read in from the binary version of parameter file.

c) Atomic coordinate manipulation

A complete set of Cartesian coordinates is always necessary if molecular energy computations are to be performed. Crystal structure analysis yields these coordinates but in cases where all the coordinates are not available, it is possible to generate the entire structure from internal coordinates (I.C.'s). I.C.'s are generated from the internal coordinate table which contains the reference geometries about which the bond, angle or dihedral angles are determined. If the positions of any three atoms are known, the position of a fourth atom can be defined from the IC's. The IC's consist of three values - an angle, a distance and a dihedral angle. The IC table contains five values corresponding to angle, distance and dihedral values. Given four atoms, A,B,C and D arranged as below:-



The IC table contains the A-C bond distance, the A-B-C bond angle, the A-B-C-D dihedral angle, the B-C-D bond angle and the C-D bond distance.

(iv) Computational Algorithms

a) Overview

In addition to building molecular structures CHARMm provides algorithms for calculation of empirical energy in energy minimisation of structures, molecular dynamics and vibrational analysis. This work is concerned with the energy minimisation routines associated with TS and novel antifolate ligands.

b) Energetics and force

Provided that Cartesian coordinates, or parameter set and a PSF are available, the potential energy of any molecular structure can be calculated. The total energy of a molecule is described by the following equation where the total energy is the sum of the following terms:-

$$ENERGY(E) = E_b + E_\phi + E_\theta + E_\iota + \qquad \textit{internal}$$

$$E_{el} + E_{vdw} + E_{hb} + \qquad \textit{external}$$

$$E_{cons} + E_{user} \qquad \textit{special}$$

This comprehensive empirical energy function is based on the Born-Oppenheimer approximation.[69]

c) Internal Energy Terms

CHARMm describes energy empirically,[70] in that the internal and external energy terms are a function only of the nuclear coordinates and not of the electronic positions. This assumption
states that electronic and nuclear components of nuclei motion are separable as electrons can adjust their positions instantaneously in response to movement of the nuclei. The internal energy terms are described as follows:-

$$\text{Bond Energy:} \quad E_b = \sum k_b (r - r_O)^2$$

where k_b = force constant and r_o = equilibrium bond length.

Bond Angle Energy: $E_\theta = \sum k_\theta (\theta - \theta_o)^2$

where k_θ = force constant and θ_o = equilibrium bond angle.

Dihedral Energy: $E_\phi = \sum k_\phi [1 + \cos(n\phi - \delta)]$

where k_ϕ = force constant, ϕ = dihedral angle and n = periodicity

Improper Energy : $E_\iota = \sum k_\iota (\iota - \iota_o)^2$

where k_ι = force constant, i_o = equilibrium improper angle

d) External Energy Terms

Nonbonded interactions make up most of the computational time when potential energy is calculated. Any two atoms in a molecular structure have an interaction between them. For large molecules such as TS (over 5000 atoms), each atom will be affected by all the other atoms in the structure. The computational time needed to evaluate all these interactions is extremely long - fortunately CHARMm allows the user to specify a distance beyond which no interactions are calculated. This is called the switching function and is important in saving computational time when calculating potential energies. For example, a cut off distance of 11Å (i.e. for each atom, only atoms within an 11Å radius are involved in calculation of nonbonded interaction for that atom) ensures that 95% of the associated nonbonded energy is computed for that atom. The Electrostatic and Van der Waals interactions contribute nearly all the external (nonbonded) energy of a molecular system.

e) Electrostatic Energy

Electrostatic energy is due primarily to the interactions of polar and charged atoms. Polar atoms arise from non-uniform charge distribution of covalently bound atoms of different electronegativities. The total electrostatic energy results from the summation of attractive and repulsive partial or net charges and is greater than van der Waals energy. Electrostatic energy also acts at a longer range than van der Waals forces.

Electrostatic Energy: $$E_{el} = \sum_{i,j>i} \left(\frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \right)$$

where $q_i\ q_j$ = atomic charges, ε_o = dielectric constant and r_{ij} = interatomic distance

CHARMm approximates the positive and negative partial or net charge as point charges, centered at each individual atom. This is the monopole approximation, the energy between two point charges being dependent on the distance between them.

f) van der Waals Energy

van der Waals energy principally is due to the forces of electron shell repulsion and an attractive dispersion force between a pair of atoms. The attractive forces or London forces are the universal forces between atoms that hold non-polar molecules together in the liquid phase. This force is due to the fluctuating dipolar moment of molecules, resulting in a net attractive force.

van der Waals Energy: $$E_{vdw} = \sum_{i,j>j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

where r_{ij} = interatomic distance

Balancing the attractive forces are short range repulsive forces which result from electron cloud overlap between molecules. The repulsive force is more sensitive to distance between molecules than the attractive force. The resulting force is referred to as a 6-12 potential or the Lennard-Jones potential.

To maximise the efficiency of nonbonded calculations (van der Waals and electrostatic interactions), CHARMm creates a list of all the pairs of interactions that are to be calculated. Careful modification of the nonbonded list can dramatically reduce the computational time required to perform minimisation routines.
Atom pairs that are too far apart or directly connected (via bond or bond angle) are excluded from the nonbonded list. The nonbonded cutoff distance (CUTNB) used throughout this work is 12Å, atom pairs that are further than 12Å are not included in the nonbonded list.
The nonbonded list is constructed in two steps. Firstly, all residues that can interact are found and a minimum rectangular solid box about each separate residue is created. All

potentially interacting pairs are then computed. The second stage in the nonbonded list generation searches all atoms in the residue pairs that interact to create the final list.

For most calculations, a cutoff point of 12Å makes very little difference to the energy of a large molecule, however there is always the possibility that overall long range interactions outside the 12Å cutoff distance may not be negligible. Individual long range interactions are negligible as the van der Waals and electrostatic potentials beyond 11Å are very small.
In order to smooth out the change in energy beyond 12Å - automatically zero, as no atoms outside this range are considered, CHARMm includes two functions to smooth out the energy around the cutoff distance.

The switching function (VSWITCH) is applied to the van der Waals interaction and the shifting function (SHIFT) applied to the electrostatic function - the default applies the SWITCH function to both nonbonded terms.
The SHIFT function is more simple than the VSWITCH function - SHIFT modifies the potential energy and forces so that they disappear at a given cutoff distance (CUTOFNB). The CUTOFNB value in this work is 11Å. The function takes the following form:

$$\left(1-\frac{2r_{ij}^2}{r_{cut}^2}+\frac{r_{ij}^4}{r_{cut}^4}\right)$$

where r_{ij} = interatomic distance and r_{cut} = cutoff distance

The VSWITCH function requires two variables CTONNB and CTOFNB to define the switching function. CTONNB defines the distance at which the VSWITCH function is "on" and CTOFNB, the distance when the function is "off".
In this work CTONNB = 9Å and CTOFNB = 11Å (as above).

$$r<r_{on}:1 \qquad r_{on}<r<r_{off}:\frac{\left(r_{off}-r\right)^2\left(r_{off}+2r-3r_{on}\right)}{\left(r_{off}-r_{on}\right)^3} \qquad r_{off}<r:0$$

The update frequency of the nonbonded list is specified by INBFRQ - in this work the

nonbonded list is updated every 10 steps unless otherwise specified.

g) Constraint Energy

Atomic constraints fix atoms in space so that they do not move during minimisation of the molecular structure. The simplest form of constraint is to allow atoms no movement and all energy terms involving fixed atoms remain fixed. However, for most applications, it is desirable to allow atoms some freedom of movement whilst minimisation is performed.
In cases where the starting structure is well ordered and near to a conformational minimum, such as an X-ray crystal structure, a harmonic constraint is applied to the atoms.

$$E_{cons} = \sum_i k_i mass_i \left(r_i - r_{oi}\right)^{\exp}$$

where E_{cons} = the constraint energy, k_i and $mass_i$, the force constant and the mass of the atom i, r_i the position of the atom and r_{oi}, its reference position about which the atom is to be centred.
Harmonic constraints on atoms allow a molecular structure to relax during minimisation while avoiding large displacement of atoms.

(v) Minimisation

When potential energies and forces acting on the atoms in a molecular structure have been defined, it is possible to calculate the total potential of the structure. However, this is a calculation of a single fixed conformation and usually it is preferable to determine a "low energy" structure or a number of low energy structures of the molecule under investigation. Minimum energy conformations are obtained by minimisation. Minimisation is performed to relieve strain in conformations obtained experimentally and determine the conformational local energy minimum. In this work, local minima around the starting x-ray crystal structure are examined. Many minimisation techniques exist for finding low energy conformations. CHARMm supports five different, iterative minimisation algorithms. Of these, three algorithms were used to minimise the TS crystal structure.

a) Steepest Descents (SD)

This is the simplest non convergent first derivative method of minimisation. SD moves atoms directly down the energy gradient. A displacement opposite to the potential energy gradient is added to the Cartesian coordinates at each step. If a lower energy results, the step size is increased, otherwise it is decreased. This algorithm is robust and quickly improves poor starting geometries (such as close contacts between atoms). However, SD suffers from poor convergence near the minimum and can often overshoot a local minimum.
Generally, SD is useful for initial minimisation of large, well ordered structures (such as X-ray crystal structures) where small changes in atomic positions can relieve close contacts. SD can also minimise and relieve steric strain in small molecules with poorly defined structure.. SD stores no information on the previous directional movement of atoms, so cannot predict better movement directions.

b) Powell (POWE)

The second minimisation algorithm used in this work is the conjugate gradient Powell method minimiser. This is an iterative method of minimisation and makes use of the previous change in energy (as well as the current gradient) after each minimisation step. The POWE method converges to a minimum energy faster than SD, requiring fewer evaluations of the energy and energy gradient to achieve the same reduction in energy .
However, POWE does not update the non-bonded and hydrogen bonded interaction lists. Very poor starting conformations can cause numerical data overflows. POWE is a useful algorithm minimising low energy conformation structures, before employing the sophisticated Adopted Basis Set Newton-Raphson method

c) Adopted Basis Set Newton Raphson (ABNR)

ABNR is based on the second derivative minimisation method Newton Raphson (NR)[71]. NR is much more efficient in locating a minimum, although it has poor tolerance for bad starting structures. NR requires excessive computational storage - a 3N x 3N second derivative matrix (where 3N dimensions correspond to the 3N degrees of freedom of the structure) must be stored and inverted.

Computational time is extremely long using NR and is therefore is restricted to 200 or fewer atom systems in CHARMm. For most large molecular systems (200 atoms or

more) ABNR is normally the preferred algorithm. ABNR performs energy minimisation using the NR algorithm. The positions and forces from each of the previous steps are stored and the second derivative matrix is constructed by finite differences from the displacement and the first derivative vectors. SD steps are applied along the directions of zero or negative eigenvalues
(to avoid saddle points). Any new directions are incorporated into the displacements subspace (the basis set) by using a SD step along the residual gradient vector calculated at each step. ABNR is two to three times faster than NR as the second derivative matrix is much smaller.

(vi) Electrostatic Potential Surface

Effective representation of important atomic and molecular properties (such as accessibility and electrostatic potential) is provided within QUANTA. An adaptation of the Connolly Molecular surface program[72] can be applied from the **Calculate** menu under **Solvent Surfaces.** A probe sphere of radius 1.4Å (equivalent to a charged proton) is rolled over the surface of the molecule or atoms under consideration, resulting in a smooth surface of dots which represents the Electrostatic potential surface accessible to a water molecule, including internal cavities. Connolly surfaces are used to study the shape, depth and chemical characteristics of receptor sites of a molecular structure.

2.2 MATERIALS AND METHODS

(i) Software

The programs CHARMm (Version 21.2) and QUANTA (Version 3.2.3) were used as supplied by Molecular Simulations without modification of source code.

(ii) Hardware

Computational drug design was performed using Silicon Graphics 4D/20 Personal Iris workstations with 12 MB main memory and 380 MB disk storage capacity.

(iii) Data

The source of the three-dimensional coordinates of Thymidylate Synthase was in the form of the crystal structure of *Escherichia coli* Thymidylate Synthase (TS) complexed with the substrate dUMP and CB3717 (an analogue of the cofactor methylene tetrahydrofolate), solved by multiple isomorphous replacement and refined at 1.97Å resolution.[65,67]

The parameter set PARM30.PRM that defined atomic masses, partial charges etc. was slightly modified to include several customised parameters (see Appendix 2.1).
Novel compounds designed as TS inhibitors were derived from the crystal structure of CB3717 in TS.

(iv) Methods

The original crystal data for each form of TS was obtained from 8 separate files. All headers were removed from each structure file and catenated together.
The molecular structure files (MSF's) were generated using QUANTA's amino acid dictionary for **proteins with polar hydrogens only** (v 20.3 parameters), renamed, reordered and assigned to new segments within QUANTA as follows:

SEGMENT NAME	RESIDUE		NUMBER
1S	1	TO	264
2S	301	TO	564
1H	565	TO	566
2H	567	TO	568

This order is common to all MSF's of TS in this work (see Appendix 2.2 for full structure information).

a) Comparison of the two TS forms

TS was crystallised in two forms - a "reduced" (so called as it was recrystallised in a greater amount of mercaptoethanol) and an "oxidised" form (see Figure 2.3).
The monomers of each form were compared using the **Protein Package** within the **Applications** menu in QUANTA. The main chains of the monomer subunits were compared using **Comparison** and **Superposition.**

b) Analysis of crystal structure

Secondary structure was assigned to each TS form by using **Calculate H bonds** and **Calculate secondary structure** within the **Protein package.**
The *b* -value parameter in the raw crystal data was assigned by colour in the TS structure - red corresponding to the high *b* -value and blue to a low *b* -value. The *b* - value is an indicator of the thermal mobility, a high value indicates large fluctuations of movement while a low value indicates small fluctuations.
Polar hydrogens were added to the starting structure using the **Add polar hydrogens** within **Molecular Editor**

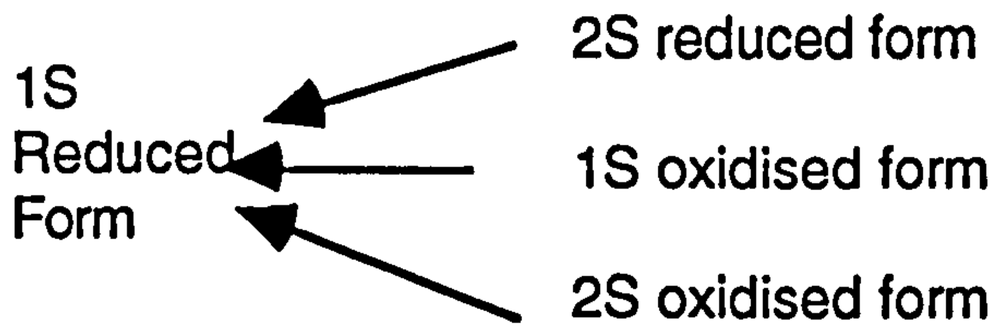
c) Active site comparison

From initial visual studies of the TS crystal structure, there appeared to be two distinct pockets that CB3717 could occupy.
A common RTF for CB3717 was generated in the **Molecular Editor** of QUANTA. Each monomer of TS contained one active site, so there are four active sites in all, when both TS forms are considered.

```
MAJOR_FORM    *1  *  MKQYLELMQK*11 *  VLDEGTQKND*21 *  RTGTGTLSIF*31 *  GHQMRFNLQD*41 *  GFPLVTTKRC*51 *  HLRSIIHELL
MINOR_FORM    *1  *  MKQYLELMQK*11 *  VLDEGTQKND*21 *  RTGTGTLSIF*31 *  GHQMRFNLQD*41 *  GFPLVTTKRC*51 *  HLRSIIHELL

MAJOR_FORM    *61 *  WFLQGDTNIA*71 *  YLHENNVTIW*81 *  DEWADENGDL*91 *  GPVYGKQWRA*101*  WPTPDGRHID*111*  QITTVLNQLK
MINOR_FORM    *61 *  WFLQGDTNIA*71 *  YLHENNVTIW*81 *  DEWADENGDL*91 *  GPVYGKQWRA*101*  WPTPDGRHID*111*  QITTVLNQLK

MAJOR_FORM    *121*  NDPDSRRIIV*131*  SAWNVGELDK*141*  MALAPCHAFF*151*  QFYVADGKLS*161*  CQLYQRSCDV*171*  FLGLPFNIAS
MINOR_FORM    *121*  NDPDSRRIIV*131*  SAWNVGELDK*141*  MALAPCHAFF*151*  QFYVADGKLS*161*  CQLYQRSCDV*171*  FLGLPFNIAS

MAJOR_FORM    *181*  YALLVHMMAQ*191*  QCDLEVGDFV*201*  WTGGDTHLYS*211*  NHMDQTHLQL*221*  SREPRPLPKL*231*  IIKRKPESIF
MINOR_FORM    *181*  YALLVHMMAQ*191*  QCDLEVGDFV*201*  WTGGDTHLYS*211*  NHMDQTHLQL*221*  SREPRPLPKL*231*  IIKRAPESIF

MAJOR_FORM    *241*  DYRFEDFEIE*251*  GYDPHPGIKA*261*  PVAIMKQYLE*307*  LMQKVLDEGT*317*  QKNDRTGTGT*327*  LSIFGHQMRF
MINOR_FORM    *241*  DYRFEDFEIE*251*  GYDPHPGIKA*261*  PVAIMKQYLE*307*  LMQKVLDEGT*317*  QKNDRTGTGT*327*  LSIFGHQMRF

MAJOR_FORM    *337*  NLQDGFPLVT*347*  TKRCHLRSII*357*  HELLWFLQGD*367*  TNIAYLHENN*377*  VTIWDEWADE*387*  NGDLGPVYGK
MINOR_FORM    *337*  NLQDGFPLVT*347*  TKRCHLRSII*357*  HELLWFLQGD*367*  TNIAYLHENN*377*  VTIWDEWADE*387*  NGDLGPVYGK

MAJOR_FORM    *397*  QWRAWPTPDG*407*  RHIDQITTVL*417*  NQLKNDPDSR*427*  RIIVSAWNVG*437*  ELDKMALAPC*447*  HAFFQFYVAD
MINOR_FORM    *397*  QWRAWPTPDG*407*  RHIDQITTVL*417*  NQLKNDPDSR*427*  RIIVSAWNVG*437*  ELDKMALAPC*447*  HAFFQFYVAD

MAJOR_FORM    *457*  GKLSCQLYQR*467*  SCDVFLGLPF*477*  NIASYALLVH*487*  MMAQQCDLEV*497*  GDFVWTGGDT*507*  HLYSNHMDQT
MINOR_FORM    *457*  GKLSCQLYQR*467*  SCDVFLGLPF*477*  NIASYALLVH*487*  MMAQQCDLEV*497*  GDFVWTGGDT*507*  HLYSNHMDQT

MAJOR_FORM    *517*  HLQLSREPRP*527*  LPKLIIKRKP*537*  ESIFDYRFED*547*  FEIEGYDPHP*557*  GIKAPVAI
MINOR_FORM    *517*  HLQLSREPRP*527*  LPKLIIKRAP*537*  ESIFDYRFED*547*  FEIEGYDPHP*557*  GIKAPVAI

                                                                                  Hydrophobic
                                                                                  Neutral hydrophilic
                                                                                  Basic hydrophilic
                                                                                  Acidic hydrophilic
```

Figure 2.3

Both forms of TS were read into QUANTA and mapped onto the reduced form of TS active site that contained CB3717 and dUMP covalently bound - within **Protein Package,,** the oxidised form of TS was **superimposed** in two segments onto segment 1S of the reduced form of TS



Then a **Comparison** using **LSQ keyboard** (the least squares algorithm) mapped all the monomers of the TS protein onto each other. Note that the CB3717 and dUMP molecules were not directly superimposed into the "averaged" active site, but were allowed to move in each active site.

d) Conformational preferences of CB3717

In order to determine any conformational preferences, CB3717 (the antifolate) was removed from both the oxidised and reduced form of TS , then minimised *in vacuo* using the steepest descents algorithm (SD) . Once the r.m.s. deviation was consistently below 0.3kcal between two coordinate sets, this provided a good indication of the overall change in conformation of a molecule between minimisation steps.

Then Adopted-basis set Newton Raphson (ABNR) was applied until r.m.s. deviation was below 0.1. Typically, an *in vacuo* minimisation of CB3717 underwent 150 steps SD and 250 steps ABNR with a step size of 0.1. Any change in CB3717 conformation after minimisation was noted.

e) THF and DHF conformations

Methylene tetrahydrofolate (THF) and dihydrofolate (DHF) are the normal cofactors of TS. CB3717 was selected from both the major and minor forms of TS and converted into DHF and THF using the **molecular editor.** Atom types and charges were assigned using charge templates within QUANTA - (charge smoothing over all atoms) and the coordinates saved with all hydrogen atoms included in the structure.

f) Application of Patches

In order to allow CHARMm to recognise the covalent bond between the two dUMP molecules and the cysteine residues in the major ("reduced") form of TS, a "patch" file was created. This was called MAJO.RTF and this patch specified the covalent S-H bond for residues 1S:146 to 1H:565 and 2S:446 to 2H:567 (CYS to dUMP).
(see Appendix 2.3)

g) Minimisation of TS containing CB3717, THF and DHF

For docking of DHF and THF into TS, the RTF's corresponding to these structures were changed to include polar hydrogens within QUANTA's Molecular Editor. The newly created RTF's with polar hydrogen representation were called THF.RTF and DHF.RTF

This was to determine preferred conformation of DHF and THF either within TS or *in vacuo* . Once minimisation of the substrate/enzyme complex was completed, DHF and THF were removed from the structure and reminimised as before to check if a reasonable fit was attained.

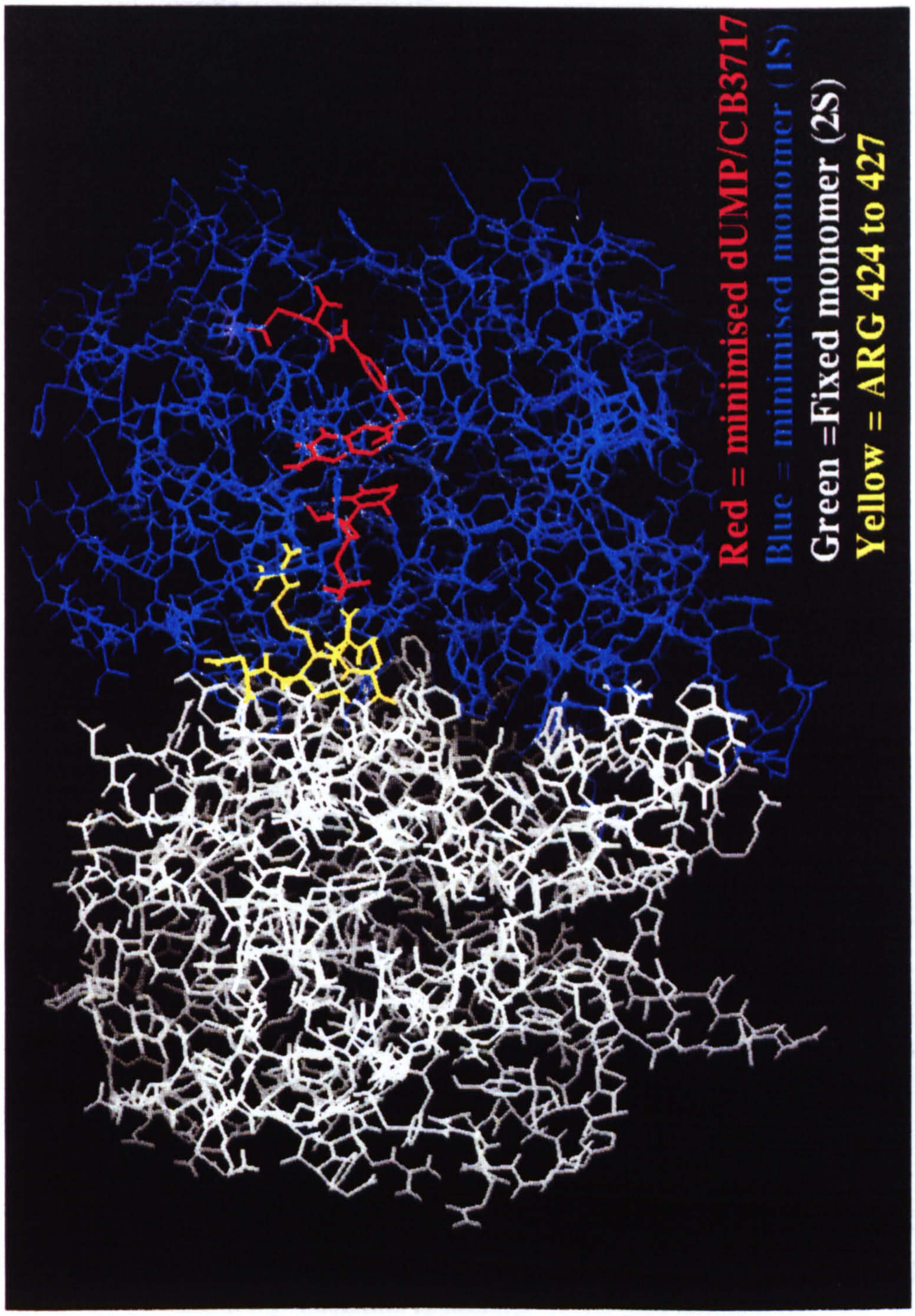
h) N^{10} substitution and N^{9} ligand design

Now that the behaviour of THF, DHF and CB3717 within TS and *in vacuo* had been determined, an investigation of the effect of the propargyl group (CP1, CP2, CP3 and H46) of CB3717 was undertaken.
The propargyl group is responsible for the highest specificity of the N^{10}-substituted quinazoline based series of ligands for TS.[73-83]

i) Minimisation of one subunit of TS

In order to significantly reduce the computational time required to minimise TS and associated bound novel ligands, half of the dimer crystal structure was minimised.
The minor form of TS had one monomer fixed so that the constrained atoms did not move during minimisation. Segments 2S and 2H were subjected to a CONS FIX command in CHARMm. This fixed all atoms in residues 301 to 564 and 567 to 568. Four ARG residues (424 to 427) were not fixed as these residues formed part of the active site of the second subunit (see Figure 2.4).

Figure 2.4


Red = minimised dUMP/CB3717
Blue = minimised monomer (1S)
Green =Fixed monomer (2S)
Yellow = ARG 424 to 427

Similarly the major form of TS had segments 1S and 1H fixed (residues 1 to 264 and 565 to 566) apart from the four ARG residues 124 to 127).
Each TS form was then subjected to 25 steps SD and 250 steps ABNR.
Minimised coordinate were read out to "TS_HALF_MIN.CRD" and "TS_RED_HALF_MIN.CRD" respectively.

j) Bifurcated ligand design - the BIF series

Bifurcated ligands based on the crystal structure of CB3717 were built as follows:
The 1S segment of the minor from of TS was matched with the 2S segment of the major form using **comparison** in QUANTA's **protein package.**
Residues 1H:566 from the TS minor form and 2H:568 from the major form were selected and saved to separate MSF's. In Molecular editor, the propargyl groups of each CB3717 were removed and the quinazoline ring system from 2H:568(major) added to N^{10} of the 1H:566 (minor) CB3717. This was called BIF1, the bifurcated ligand based on the CB3717 found in the minor form of TS in the 1S segment(see Figure 2.5).

Similarly, BIF2 was created by removing the propargyl groups of both template CB3717 molecules and the quinazoline ring system 1H:566(minor) added to N^{10} of 2H:566(major)

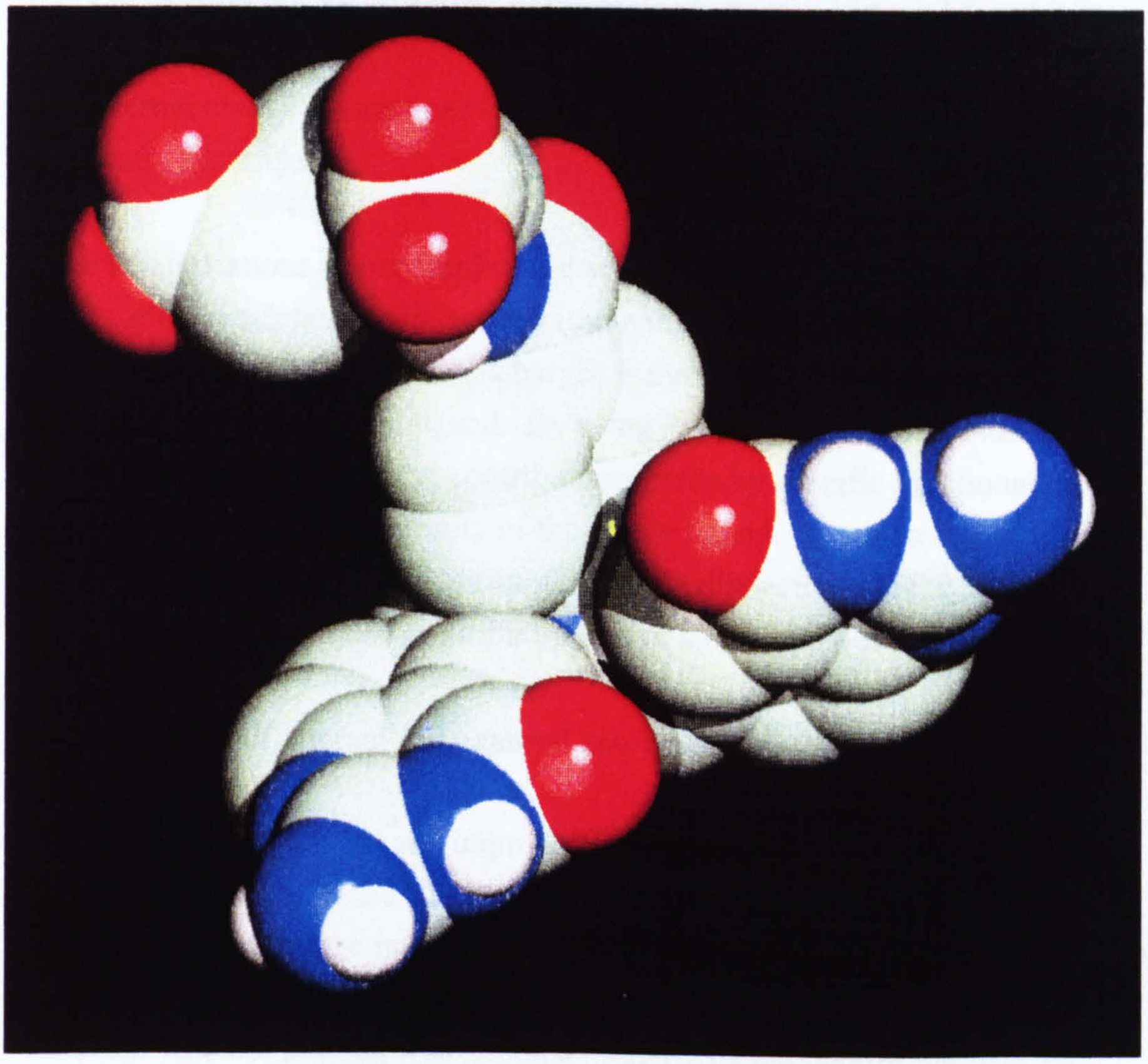
Atom numbers for both new bifurcated molecules were changed accordingly.

All atoms were retyped and bond types correctly assigned. Atoms were charged using **charge templates** option in molecular editor.

To relieve any steric strain or incorrect dihedral angle conformation around the N^{10} the two new ligands were separately minimised with a harmonic constraint force of 50 kcal/mol placed on all atoms in each new structure apart from C11, C9, C6, C26 and N10 which were allowed to minimise freely.

With these constraints in place the BIF ligands underwent 25 steps SD and 250 x 4 steps ABNR, while the harmonic constraint force was allowed to relax. Once harmonic constraints were removed, a further 25 steps SD and 250 x 4 steps ABNR were performed to give final minimised structures of bifurcated ligand. The new BIF molecules were introduced into their respective active sites: BIF1

Figure 2.5



into the locus previously occupied by 1H:566 CB3717 of the minor TS form and BIF2 into 1H:568 of the major form.

The BIF ligands were then fixed using the CONS FIX command and the enzyme allowed to minimise around them with 25 steps SD and 250 x 4 steps ABNR.

Finally, the BIF ligands were allowed to minimise freely within TS with 25 steps SD and 500 x 2 ABNR. This gave a final, minimised TS structure containing BIF1 in the minor form and BIF2 in the major form (see Figure 2.6 and 2.7).

k) External energy interactions for each atom in the BIF and CB3717 series at 4, 8 and 12Å range

Once minimisations were completed, a script file was set up to calculate van der Waals and electrostatic contributions (the nonbonded interactions) at a 4, 8 and 12Å sphere around each ligand. Charges were reported back from CHARMm for each atom in the bound ligand. By using "awk" files (see Appendix 2.3) it was possible to determine the contributions made by specific functional groups and hence assess binding affinity of the whole ligand. Summated contributions of the quinazoline ring systems and glutamate tails sections of CB3717 were noted and compared to the contributions from atoms in the BIF molecules.

l) Comparison of internal and external energy contributions

The dihedral, angle, bond and improper angle values were calculated using CHARMm for each ligand. (BIF1, BIF2 and the major and minor CB3717 ligands). These were the internal energy terms (as described in Chapter 2.1) and were summed and compared with the external energy contributions. This indicated whether binding was affected more by ligand conformation or by atomic structure.

m) Van der Waals and electrostatic summation - bumps and gaps

Once individual atomic contributions to nonbonded interaction were collated, those atoms that had poor van der Waals and electrostatic contributions were substituted with other atoms to create new ligands. This was performed in the molecular editor by changing one atom type at a time in either BIF1 or BIF2. Each structure was saved to a new filename and atoms retyped and recharged on

Figure 2.6

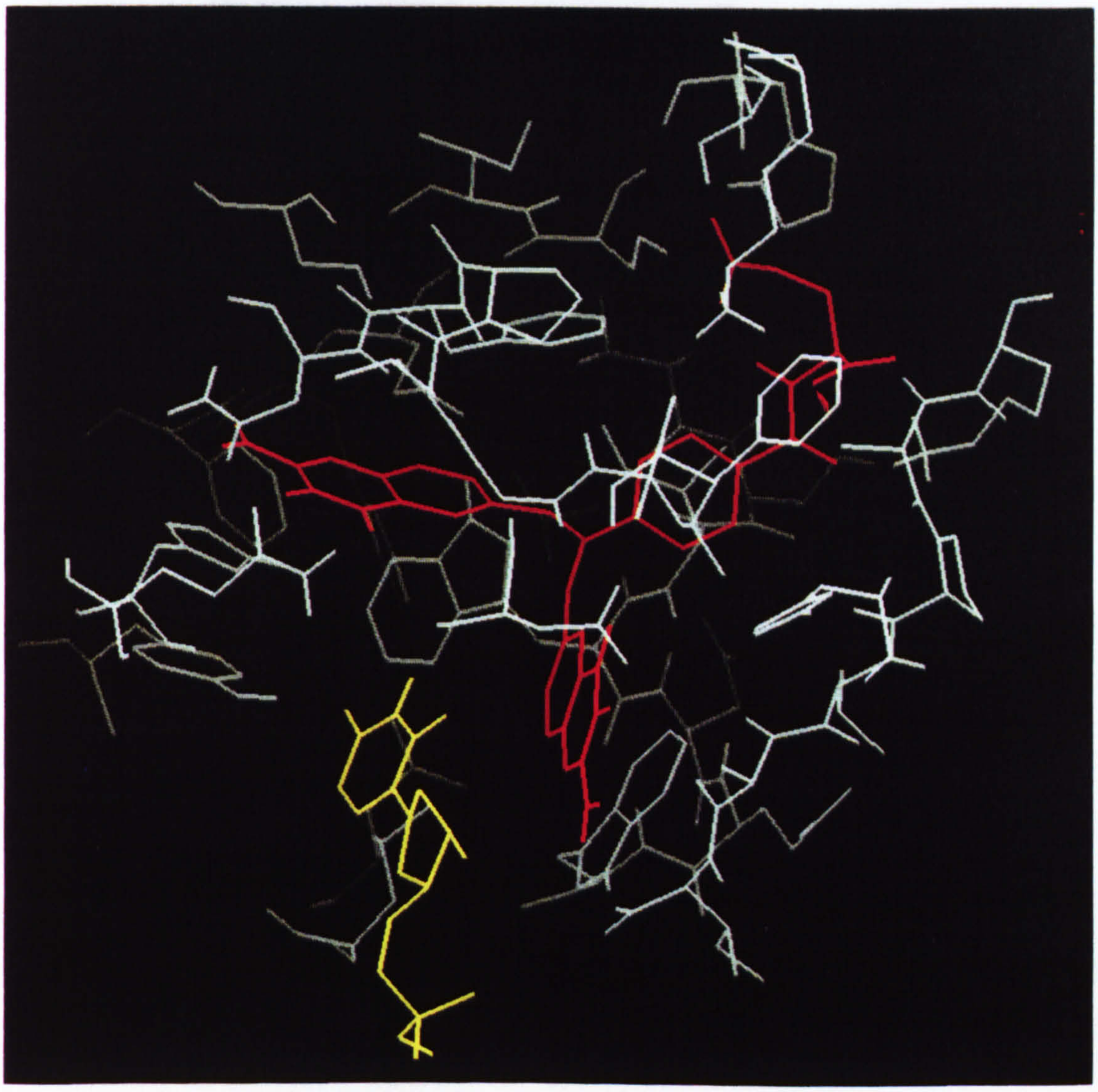


Figure 2.7

exit of the Molecular editor. In total there were 18 ligands generated from BIF1 and BIF2 making a total of 20 novel structures that could bind into the active site of TS in both major and minor forms of the enzyme.

The changes to BIF1 and BIF2 were also influenced by the assessment of close contacts within the active site. The close contacts were assessed using the neighbouring atoms tool in the main viewport of QUANTA. The dot surface generation program in QUANTA was also used to define the 3 dimensional space of the active site pocket and shape of the ligand. A dot surface (20 point density) was generated based on the van der Waals calculation of atomic size. The two dot surfaces (for the TS pocket and ligand) were attached to the structure using the **manage graphical objects** tool in QUANTA . This allowed visual inspection of the goodness
of fit of each structure generated.

Atoms in the ligand that visibly overlapped the active site pocket were selected and changed in the molecular editor. Areas that revealed "gaps" in the pocket were noted and in the cases of obvious gaps, an atom substitution was made in the ligand structure to fill the pocket more snugly.

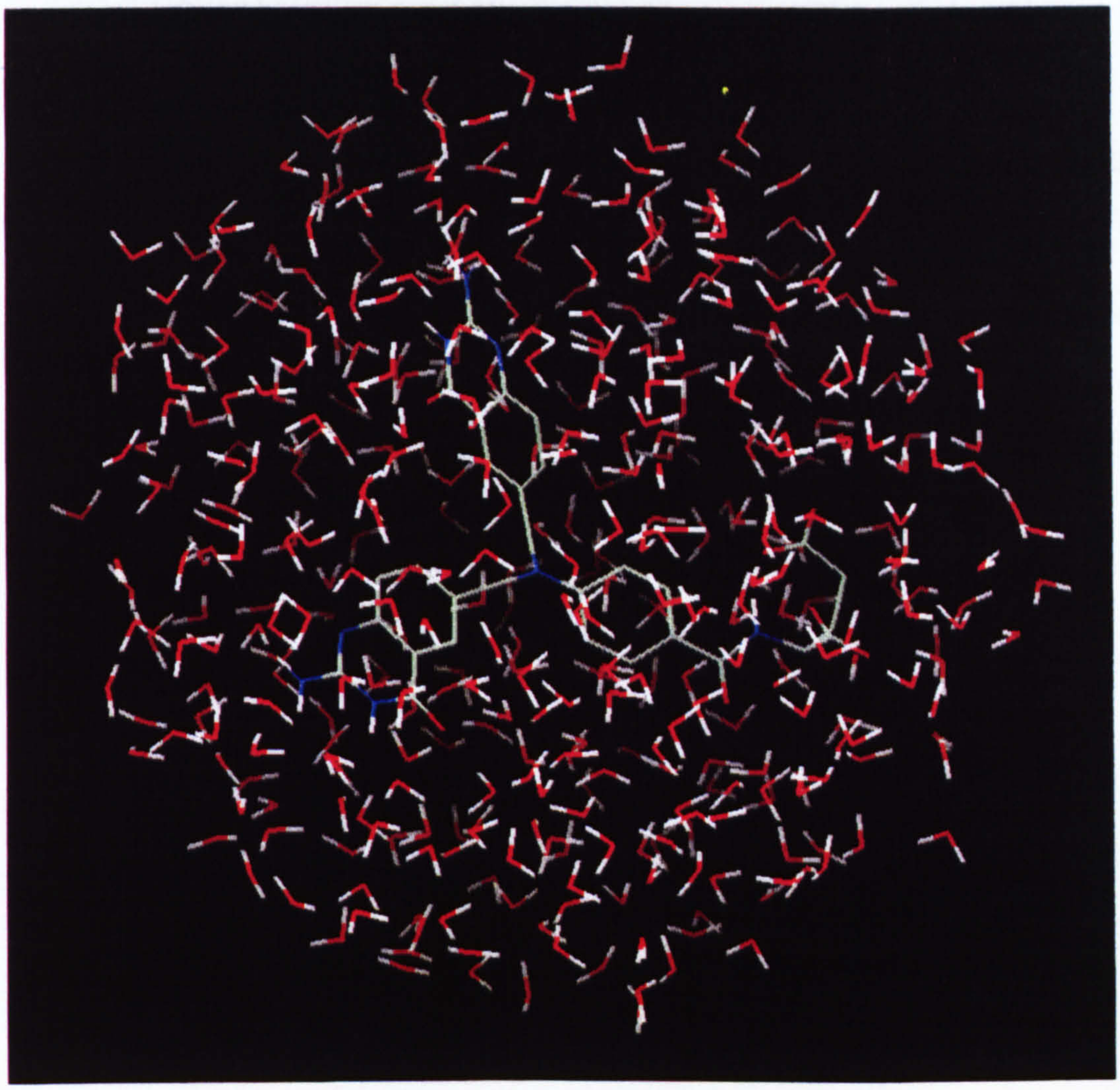
n) Solvation energies - comparison of interaction

To approximate the behaviour of each of the ligands (20 BIF and CB3717 parent structures) either complexed to TS or in solution the following approach was used.

A 15Å sphere of 420 water molecules was placed around each individual BIF ligand and minimised (see Figure 2.8) The final energy associated with the solvated ligand was noted. This value was compared to the equivalent solvation and subsequent minimisation of the two CB3717 ligands - again in 420 water molecules. This solvation was carried out using the **solvate** option under the CHARMm submenu of QUANTA.

To approximate each ligand's preference for either remaining in solution or bound to TS, the solvated CB3717 energy values were subtracted from the energy values from the final minimisation of TS containing each CB3717 ligand. This energy difference corresponded to a positive value relating to the parent ligand's preference for remaining in solution.

Figure 2.8



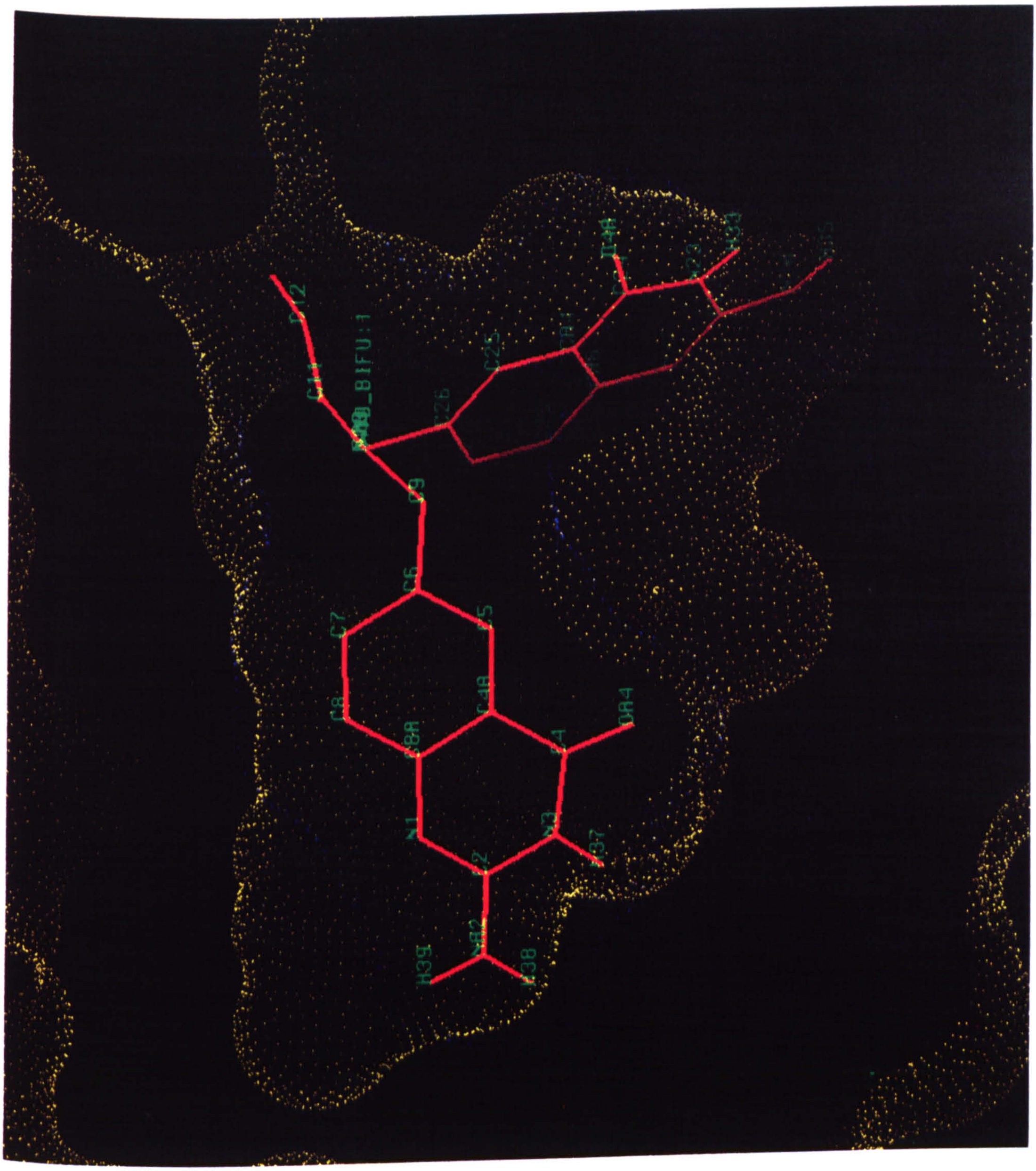
Each BIF ligand was similarly assessed, with solvated energies compared to final energies of bound ligand in TS. If the positive value obtained from this calculation was lower than the corresponding CB3717 value, the new bifurcated ligand was deemed to have a reduced preference for remaining in solution i.e. a greater preference for binding to the TS molecule.

m) Most promising ligands

Those new bifurcated ligands which showed greatest affinity for the TS enzyme were subjected to further scrutiny. A Connolly electrostatic surface was generated for each ligand and corresponding active site pocket. Visual assessments of bumps and gaps were made (see Figure 2.9 and 2.10), with emphasis on visualisation of electrostatically favourable interactions (e.g. a blue or negatively charged ligand surface close to a red or positively charged pocket surface). The Connolly surface was generated from the Calculate option in QUANTA, using a probe radius of 1.4Å and formal charge of +1. Electrostatic surfaces were rendered using a 20 point dot density with colouring of each surface ramped from red to blue (positive to negative).

Figure 2.9

Figure 2.10


C12
C11
C_BIFU:1
C9
C26
C25
C6
C7
C5
C8
C8A
C4A
C4
OA4
N1
N3
C2
H37
NA2
H39
H38

2.3 RESULTS

(i) TS Forms - major and minor conformations

The main chains (a carbon backbone and side chain residues) of the TS monomers are almost identical in structure. Mainchain comparison of each monomer gave an r.m.s. displacement of 0.535Å and a maximum displacement of 4.121Å.
The entire structure of each monomer gave an r.m.s. displacement of 1.123Å and a maximum displacement of 10.533Å. The global structure of the monomer is highly conserved in both forms of TS .[65]

The coordinate data obtained from Stroud is well-resolved and the residues in both TS forms are identical with the exception of one residue in each monomer - LYS 235 and LYS 535 in the major form dimer are replaced by ALA 235 and ALA 535 in the minor form dimer.

(ii) Crystal Structure

Assuming that TS crystal structure is an accurate representation of *in vivo* structure, it necessarily follows that computer modelling and drug design based on this structure is valid.

B-values from raw data indicate that the crystal lattice is highly stable around the core of the dimer molecule. Typical core protein b-values are of the order of 1 to 2. The surface of the crystal structure has a greater b-value range from 1 to 7. This is to be expected, as this is at the solvent interface in which the TS crystals were solved.

The overall structure of the TS molecule bound to CB3717 is similar to a previously solved TS unbound structure of *L.Casei* [84] or *E.Coli* .[85] The major and minor forms consist of 528 residues but the contents of each active site differs markedly.

(iii) Active site comparison

The residues that define the boundaries of the TS binding sites are highly conserved.[65-66]

TS exists as a dimer with a single active site in each subunit. The major from of TS contains dUMP covalently bound to cysteines 1S:146 and 2S:446 in both active sites in the presence of CB3717 (CB3:566 and CB3:568). One active site of the minor form contains dUMP (residue 1H:565) noncovalently bound in the presence of CB3717 (CB3:566) whilst the other contains only phosphate (2S:567) and CB3717.

The active site is a large cavity that allows CB3717 to bind in one conformation in the major form and a different one in the minor form. This second binding site[65] has CB3717 with the quinazoline ring rotated 147° away from the position occupied in the major form binding site.

The phosphate ion in the major form is lodged between four arginine residue side-chains, two ARG from each monomer and a serine (SER).

Analysis of CB3717 in the TS crystal structure reveals that in the major form (reduced) the glutamate tail is oriented in two different positions, approximately 180° about the C14 carbon. The benzyl group of CB3717 shifts only slightly along the plane in the major and minor forms. The major difference is the 147° shift of the quinazoline ring.
The "minor" pocket containing the alternate CB3717 conformation undergoes little sidechain disturbance, which may indicate a stable second pocket in TS.

The quinazoline rings of CB3717 in the minor form push dUMP out of the "major" pocket, while in the minor TS form, dUMP remains more planar, lining the pocket boundary. The pocket dimensions are approximately 7Å x 10Å x 10Å. This would indicate that the preferred conformation of CB3717 is the "minor" conformation when no dUMP is present.

In the minor form where dUMP is present, the pocket in the major form is blocked by the dUMP, so CB3717 adopts the 147° shift, occupying the "minor" pocket.

In the major form, covalently bound dUMP is drawn away from the pocket, giving CB3717 enough space to take up the "major" conformation.

(iv) Conformational preferences of CB3717

From initial *in vacuo* minimisation of CB3717 there appears to be no conformational preference for either "major" or "minor" positions. However, within TS, the preferred conformation of CB3717 is the "minor" structure. Even when noncovalently bound dUMP is present in the pocket, CB3717 still prefers the "minor" conformation. In the major form of TS, when the CYS:146 residue covalently binds to dUMP, conditions in the pocket that the "minor" CB3717 occupy become unfavourable. The dUMP moves across to sterically block the "minor" pocket. The "major" pocket is preferred. So, when no dUMP is present, CB3717 prefers the "minor" conformation. When dUMP is present but is noncovalently bound, the "minor" conformation is still preferred. Finally, when dUMP is covalently bound , the preferred conformation is the "major" conformation.

(v) Conformational preferences of DHF and THF

Minimisation *in vacuo* of DHF and 5,10-THF gave conflicting results. DHF appears to have no preferential conformation while 5,10-THF prefers the "reduced" (major) conformation. This basic minimisation procedure gave only a rough estimate of conformational preference - the final conformation of 5,10-THF may well have been governed by the atom constraints placed on 5,10-THF at C4A, C8A and C11 around the 5-membered ring.

(vi) Conformational preferences of DHF and THF in TS

From comparison of minimised structures of TS (containing either DHF or 5,10-THF substrates) compared to starting crystal structures, DHF and 5,10-THF seem to have no preference for which pocket they occupy.

(vii) The cbnew series of ligands

The N9 substituted series of ligands were found to be energetically unfavourable within the major and minor TS active sites. the N10 substituted CB3717[86] ligands were pushed out of the pocket as the propargyl group on C10 was not in the most favourable orientation. Close contacts (bumps) were found between the HG of SER 2S:444 and H38 of the cb3new ligand and HG of ASN 2S:441 and H39 of cb3new. Reorientation of the propargyl group to point into the active site

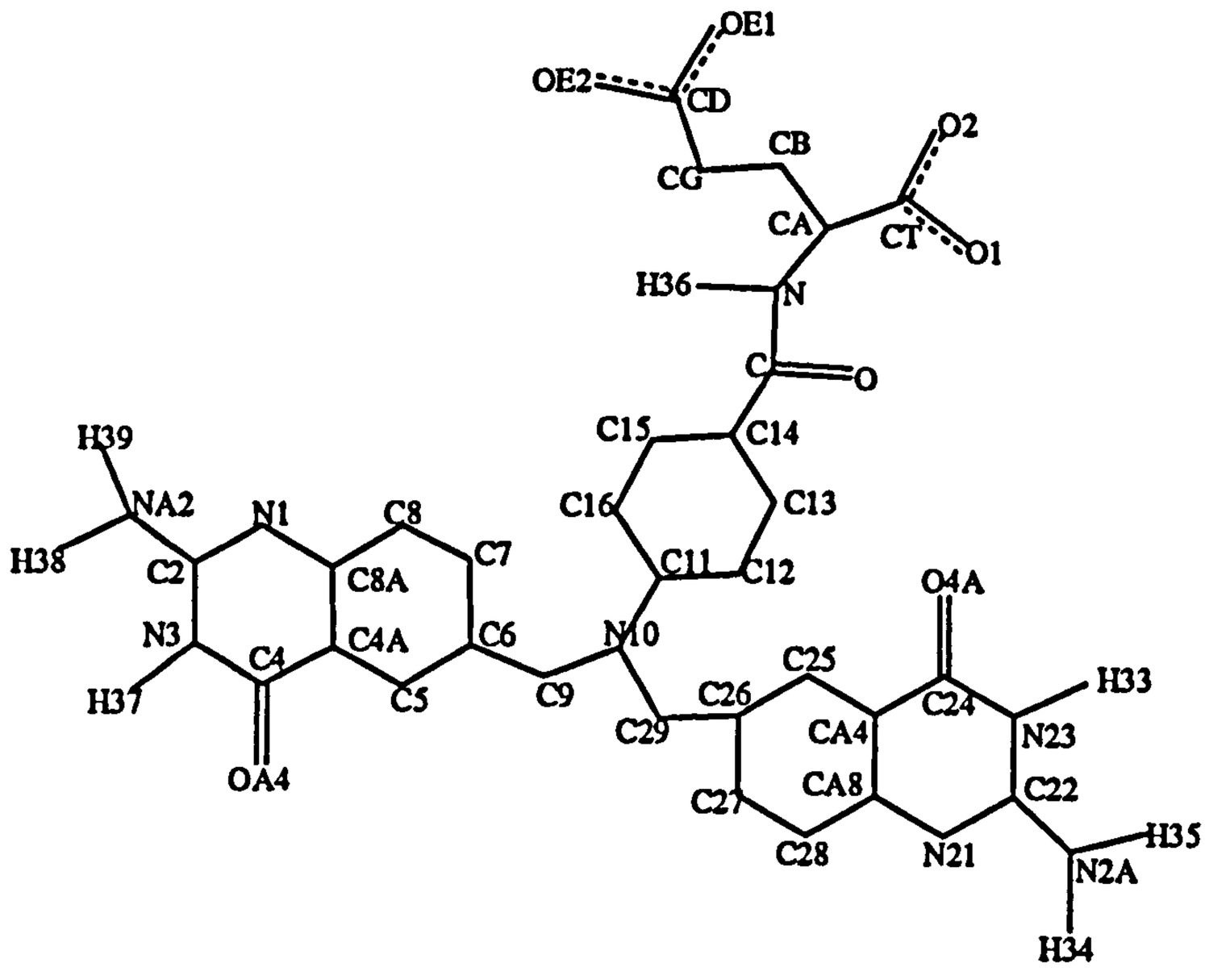
cavity did not stop the whole ligand from being forced out of the active site (cb3new) in both major and minor TS forms.
The ammonium group that then replaced the propargyl (cb4new) resulted in the ligand remaining within the active site. Extension of this ammonium group (by additions of CH2 units to the C10 group) allowed the NH3+ to hydrogen bond to the inorganic phosphate (in the minor form of TS). This extended carbon chain from the C10 group proved to be highly flexible. Consequently many energetically unfavourable conformations were generated. These cbnew ligands were deemed unsuitable for further investigation.

(viii) Monomer minimisations

By placing a CONS FIX constraint on a monomeric subunit of each form of TS, the run-time of each minimisation was reduced by up to 30% . Typically a series of half enzyme minimisations now took 5 hours of C.P.U. time rather than in excess of 8. No gross change in structure was observed at the interface between monomeric subunits. The total energy of major and minor forms of TS was -6151kcal mol^{-1} and -6113 kcal mol^{-1} respectively.

(ix) The BIF series of ligands

A comparison of the relative energy contributions of the new quinazoline ring system in the BIF ligands compared with the propargyl group in CB3717, revealed that the new quinazoline bifurcated portions contributed to a net decrease in energy of each enzyme system. Atom naming covention for BIF1 is shown below:



Comparison of Energy	bif1_quinazoline	-33.39 kcal /mol
	cb3717_minor_propargyl	-0.97 kcal/mol
	bif2_quinazoline	-19.76 kcal/mol
	cb3_major_propargyl	-4.85 kcal/mol
IC Energy Contributions	BIF1	7.74 kcal/mol
	CB3717 minor	4.11 kcal/mol
	BIF2	9.91 kcal/mol
	CB3717 major	2.78 kcal/mol
TOTAL ENERGY of TS	MINOR FORM	
	containing BIF1	-6144.59 kcal /mol
	containing CB3717	-6117.14 kcal /mol
	MAJOR FORM	
	containing BIF2	-6214.23 kcal /mol
	containing CB3717	-6153.99 kcal /mol

(x) Interactions at 4, 8 and 11Å.

Individual Van der Waals' and Electrostatic[87] energies were calculated for each atom in BIF1 and BIF2 ligands over a sphere radius of 4, 8 and 11Å. These energy contributions were compared to the parent CB3717 compound (see Appendix 2.4 for details). The atoms that gave poor contribution to overall energy were changed in subsequent BIF ligands. From the BIF1 and BIF2 ligands, these were the atoms that were changed: N21, N1, N3, N23, C8, C23 and C28.

LIGAND NUMBER	PARENT ATOM	NEW ATOM TYPE
MINOR FORM TS		
BIF3	N21	C21
BIF5	N1	C1
BIF7	N21 & N1	C21 & C1
BIF9	C28	N28
BIF11	C8	N8
BIF13	C28 & C8	N28 & N8
BIF15	N23	C23
BIF17	N3	C3
BIF19	N23 & N3	C23 & C3
MAJOR FORM TS		
BIF4	N21	C21
BIF6	N1	C1
BIF8	N21 & N1	C21 & C1
BIF10	C8	N8
BIF12	C28	N28
BIF14	C8 & C28	N8 & N28
BIF16	N3	C3
BIF18	N23	C23
BIF20	N23 & N3	C23 & C3

(xi) Solvation energies and minimisation

The BIF series ligands were minimised within TS and also within a 15Å water sphere containing 420 TIP3 water molecules. From comparison of CB3717 minimised within TS or in a sphere of 420 water molecules a comparison could be made between BIF ligands and CB3717. This comparison indicated the energy associated with each ligand for remaining in solution or bound to TS. From 20 novel compounds designed, studies indicated all but BIF2 had greater preference to be bound to TS rather than in solution when compared to their parent CB3717 ligand (see Figure 2.11 and 2.12). The most notable of these were BIF5, BIF7, BIF9, BIF10, BIF12 and BIF18.(see Appendix 2.4 and Figures 2.13 to 2.16)

(xii) Connolly surface calculations

Examination of electrostatic potential surface[88] for the BIF series of ligands showed that these new ligands fit snugly into the bifurcated pocket of TS with no overlapping electrostatic contacts. The Connolly surfaces also showed that there were no poor steric interactions within the TS pocket of both major and minor forms.

Figure 2.11

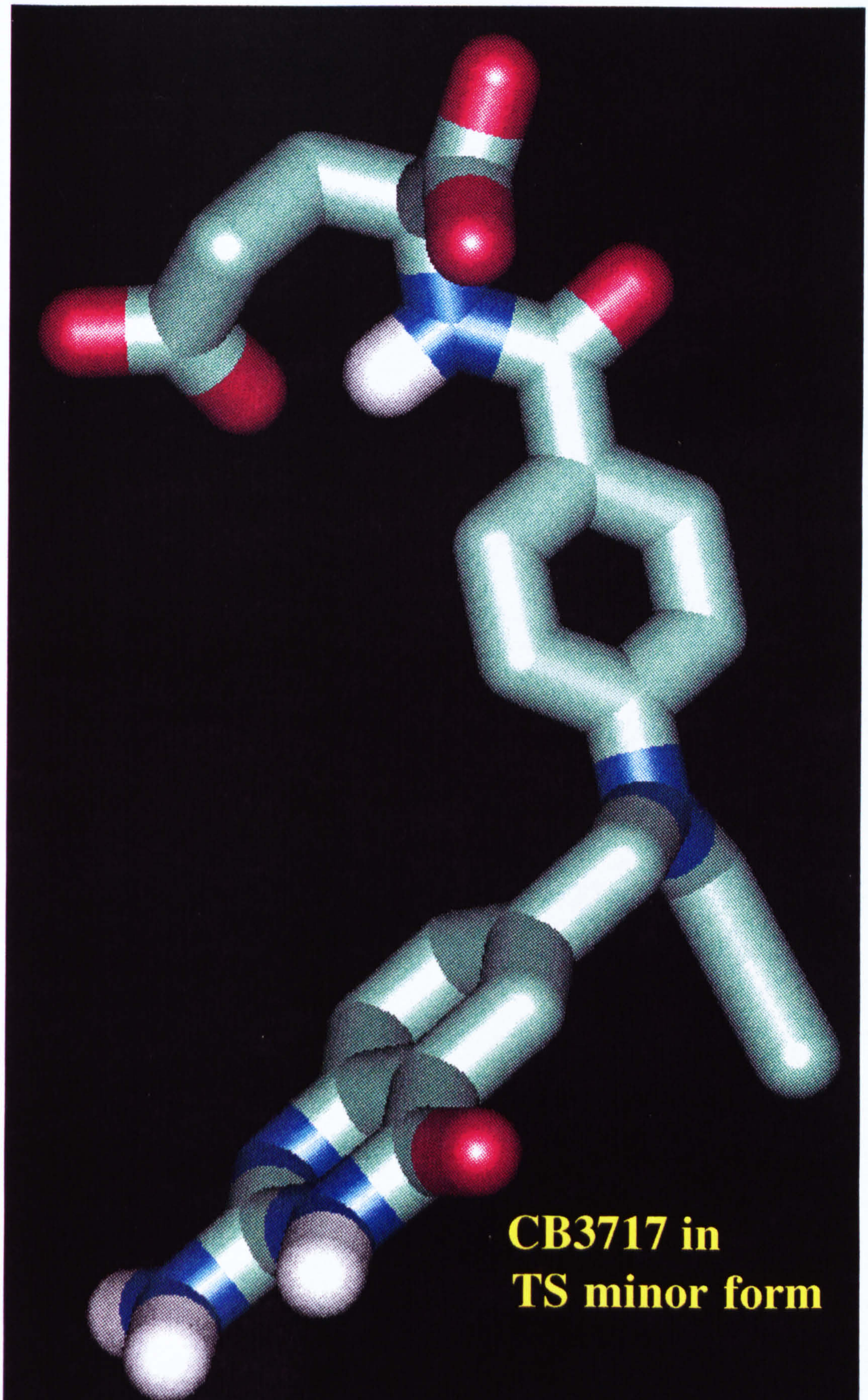


Figure 2.12

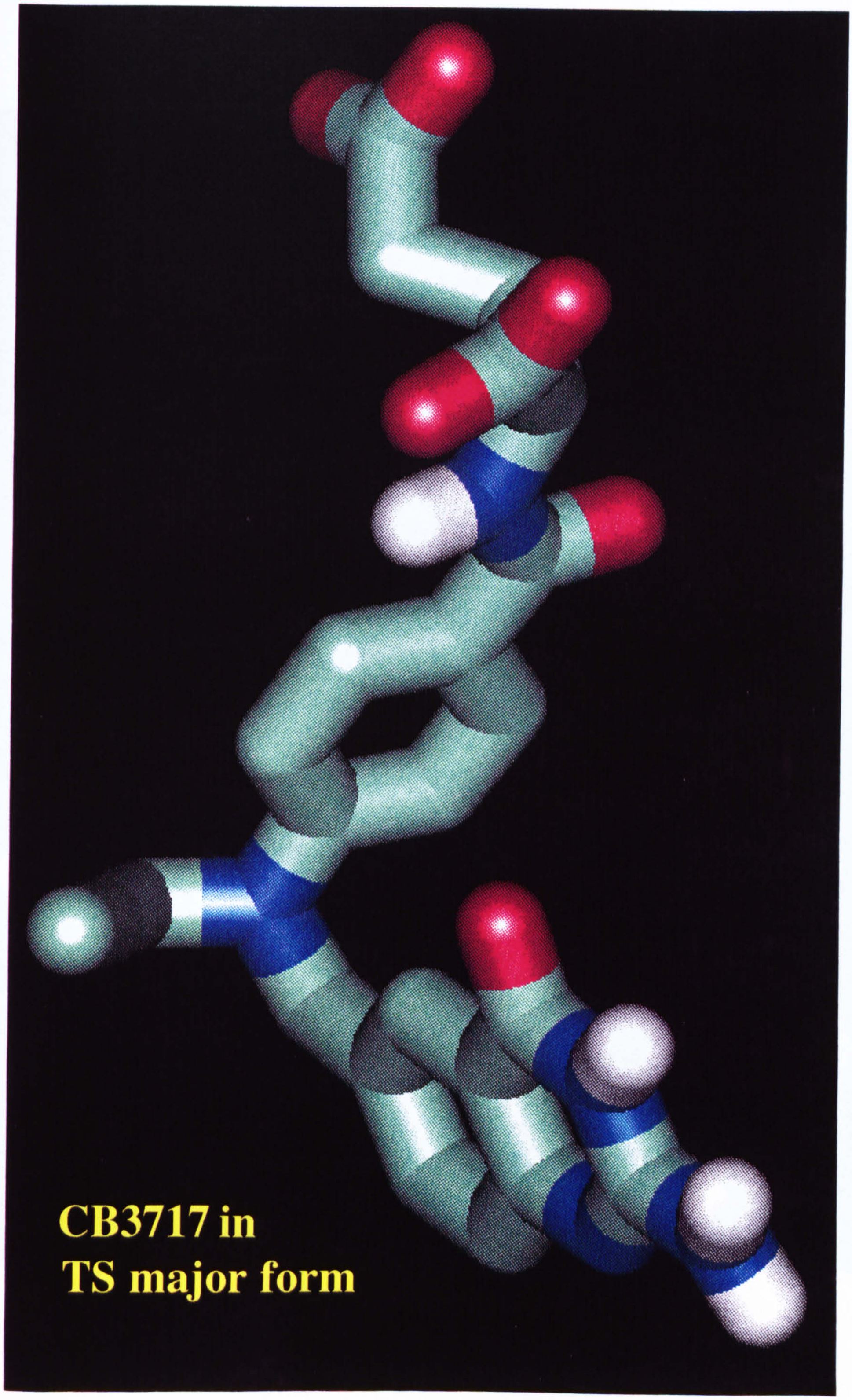


Figure 2.15

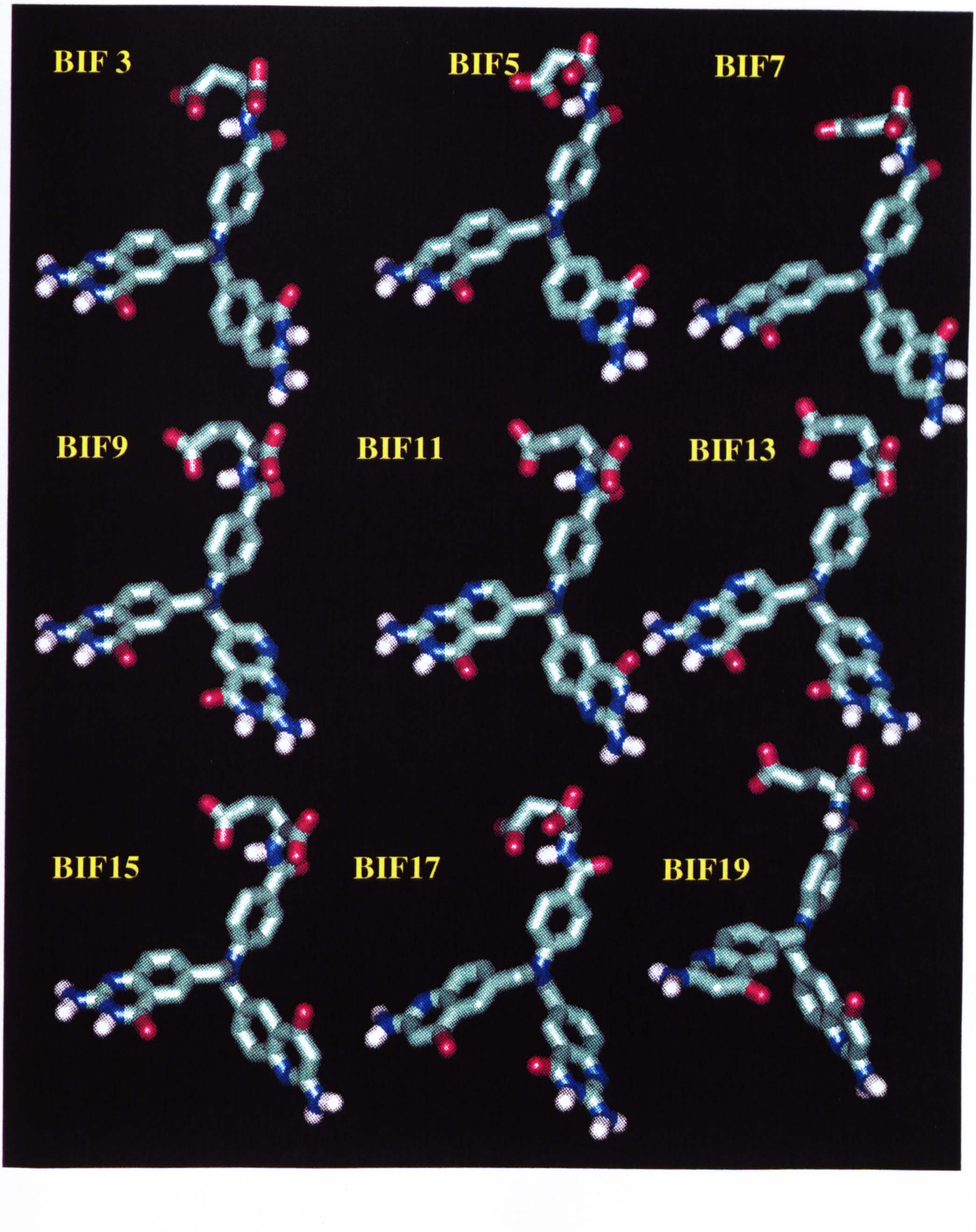
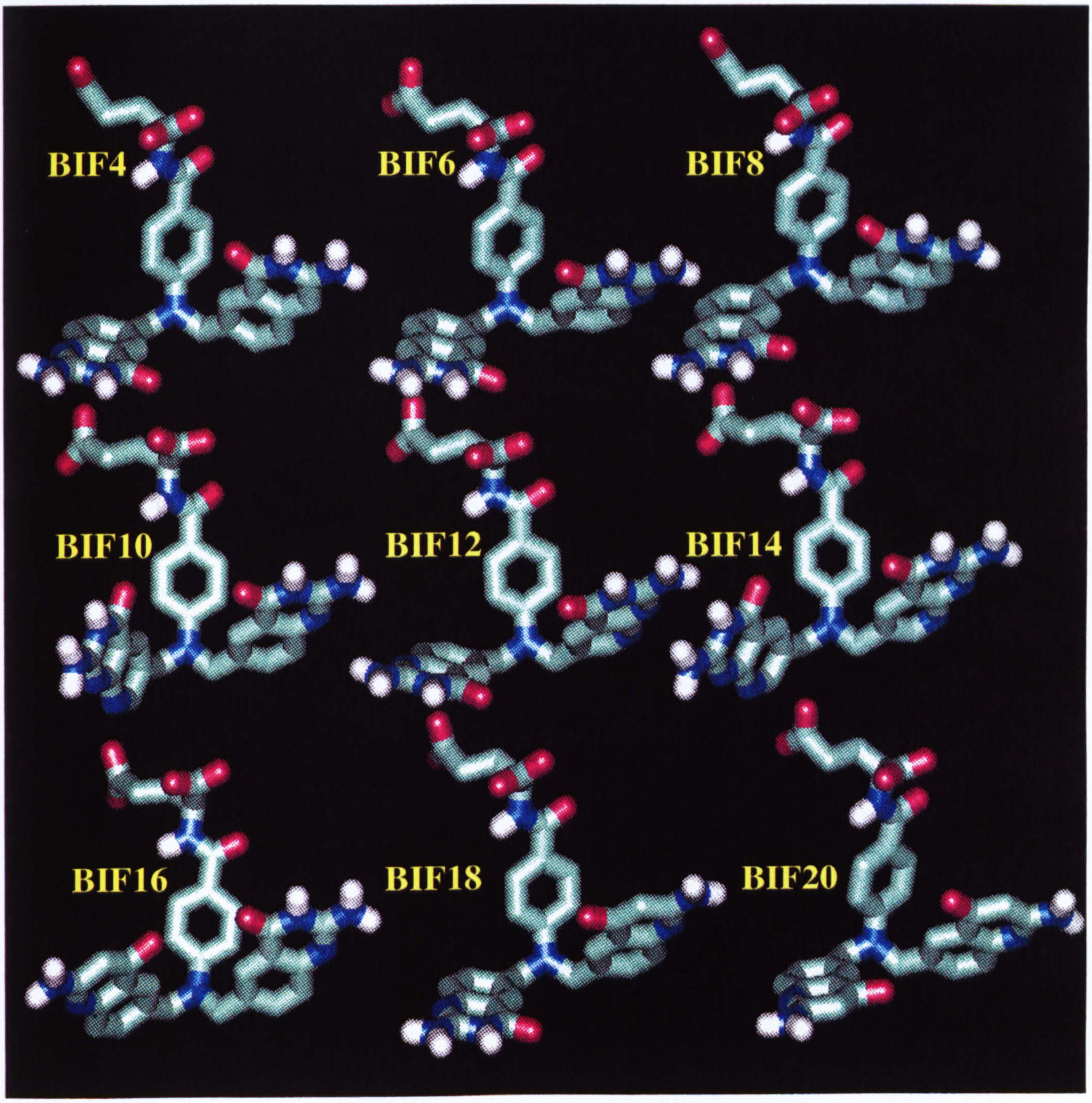


Figure 2.16



2.4 DISCUSSION

(i) Molecular Mechanics and Drug Design

Molecular mechanics is a classical method of expressing a molecular structure as a set of balls and springs, with series of potential energy functions expressing the molecular force field as a sum of these functions.[89] More elegant models describing these molecules include cross valence, polarization and nonbonded interactions.

Each individual energy term has a preferential equilibrium position (e.g. bond length, bond angle, van der Waals etc.) and force constants are used to associate energetic penalties with each individual deviation from the equilibrium. A "Force Field" is simply the complete set of analytical energy functions and their associated sets of numerical parameters (based on experimentally or theoretically calculated force constants).

CHARMm is one of the most extensively tested force fields[90] - others include MM2[89] and AMBER.[91-92] CHARMm force fields are considered the standard for macromolecules[90] and use the "united-atom" representation - all non-hydrogen bonding hydrogens are implicitly included by slightly increasing the effective size of the parent atom to which the hydrogens are bonded.[93]
In cases where the force field does not describe the parameters of certain functional groups, an estimation of bond, bond angle or dihedral angle is required (see Appendix 2.1).
Once all the necessary parameters have been obtained, molecular mechanics energy minimisation can be performed.

(ii) History of Computer-aided drug design

Man has been fighting disease for centuries. Many varieties of plants and herbs have found their way (often by chance or observation) into the front line in combating disease and illness. More recently, observations of effective compounds has been backed up with scientific reasoning - active substances in plants and herbs (which have been traditionally considered to be of medical importance) have undergone in-depth investigation.

At the start of this century, scientists began to follow up unexpected results of chemical investigations and biochemical experiments. Chemists began massive

screening programs of hundreds of derivatives of active compounds - empiricism ruled supreme!

In the 1950's drugs were designed, based only on vague notions of physical properties and functional groups. Structure-activity relationships were based more on chance and guesswork than rational drug design. By the mid 1960's, more key or "lead" drugs were reaching hospitals - good drugs were being replaced by better ones, often from just a small structural change to the original compound.
Advances in Physical chemistry and Biochemistry provided the foundation of knowledge on which a new quantitative approach to drug design was built.

In 1962, the Organic chemist Corwin Hansch brought together the quantitative terms that mathematics was now able to describe - structure and activity combined with the computational power of the emerging (and constantly improving) computers. This new approach to drug design opened the doors to a new area of scientific research.[94]

The first X-ray diffraction photograph was displayed in 1912 - from this simple X-ray picture of an inorganic structure developed the protein crystallography and protein modelling that is today working hand in hand with computational chemistry and computer-aided drug design.
From 1930 to 1980, equipment and techniques in X-ray crystallography steadily improved, especially with the introduction of automatic diffractometers and rapid data-processing with computers. In the 1980s, synchrotron radiation greatly enhanced protein crystal structure analysis. The high resolution and intensity available from synchrotron radiation allowed many more protein/inhibitor complexes to be resolved - even some protein/substrate complexes were elucidated.
1965 saw the introduction of a novel tool for the x-ray crystallographer. A line -drawing cathode ray tube controlled by a small computer was used to display small proteins and protein folding.
Within only 5 years, over 20 other graphical systems were in use, and as technology improved through the 1970's, interactive systems were developed.

By the 1980's not only could structures be represented but properties: van der Waals surfaces could be represented by contour lines, as could electrostatic potentials. Colour and dot surface generation brought structures out from the 2

dimensional screen. It became possible to *see* active sites of proteins and to *see* interactions between residues.

In 1982, Martin Karplus and his colleagues put forward a computer program to model complete macromolecular systems. This program could build structures (either from existing data or from scratch), energy minimise them and perform a range of tasks. These included molecular dynamics simulation and analysis of structural equilibrium and dynamic properties of molecules.

Today, state-of-the-art molecular modelling systems provide extensive computational and graphics facilities for analysing known structures and interactions.
It is now possible to gain accurate three-dimensional structural information on proteins and complexes formed between these proteins and inhibitors.

(iii) Use and Limitations of Modelling

Models are essential components of experimental and theoretical chemistry. Computer modelling draws together mathematical, conceptual, iconic and enactive modelling methods employed by the chemist. The mathematical model is described in the forcefield of the modelling program, the iconic model in the three-dimensional graphics displayed on the computer monitor, the enactive by the manipulations of structure made possible by the modelling program and the conceptual model is realised by the modeller as he/she *sees* the compound described by the previous models. Computer modelling acts as a stimulus to creative thinking. An experienced modeller can distinguish between "possible" and "improbable" compounds. This speeds up the selection process in the development of novel compounds and is a very important feature of computer modelling. Traditional benchmark chemistry is a slow and thorough process - computer modelling can help a chemist design novel compounds faster.

One must take care not to fall into the trap of relying on computer modelling to provide the definitive answer. The data generated by a modelling program acts as a pointer to chemical interaction and does not always generate the same results that benchwork provides. Modelling currently facilitates the work of a scientist, whether that scientist is building a structural model based on NMR distance constraints, X-ray crystal data refinement or a *de novo* model using theory alone.

(v) Physiology of TS

TS is not similar to other folate-based enzymes in its binding to substrate.[99] TS is an intracellular enzyme that is responsible for the synthesis of dTMP, a key building block in the production of DNA. Dihydrofolate reductase (DHFR) was a pervious target for drug design, but unlike TS, blocking of this enzyme by methotrexate resulted in purine synthesis and blocking of RNA production. Cells also became resistant to these drugs by simply raising the level of intracellular DHFR.[100-101] Many different classes of compound have an inhibitory effect on TS[102-109] and polyglutamation of these drugs is known to increase inhibitory effect up to one hundred times.[110] If the two pockets in TS can be utilised simultaneously, then a suitable drug analogue would have tetrahydrofolate and dihydrofolate chemical characteristics. The amount of intracellular dUMP present would obviously affect the relative binding of these compounds

(vi) Active site of TS

It was reasoned that if inorganic phosphate was needed in 1 pocket of TS for catalytic activity, then a NH3+ moiety interacting with the inorganic phosphate would induce allosteric blocking of enzyme activity. if the phosphate was an artefact (or not present under physiological conditions) then a carboxyl (COO-) group could be introduced to mimic the ARG residue binding to phosphate and resulting in competitive inhibition of THF.

(vii) Novel Ligands

From the computer modelling in this study, the following ligands are put forward for future work: BIF5, BIF7, BIF9, BIF10, BIF12 and BIF18. The structural data obtained from this study have been maintained on archived data tape and are available for analysis should provision be made for suitable experimental synthesis. Whether these compounds are indeed better inhibitors of TS *in vivo* will only be revealed after an adventurous organic chemist has produced them! To this end, I propose that BIF5 and BIF7 be put forward for synthesis as these are the best relative inhibitors of TS and are reasonably easy to synthesise.

(viii) Future Work

Should the crystal coordinates of human TS become available on the Brookhaven database, a full screen of existing compounds should be made.

Ligands with low binding affinity for TS should be compared to more active compounds. The use of computer modelling will greatly facilitate this process and lead to a greater understanding of biological activity of these antifolates. Site directed mutagenesis of TS will give an insight into the catalytic mechanisms of TS and the mode of inhibitor action. One of the BIF ligands at least should be put forward to enzyme assay and cell cytotoxicity studies, even if it proves inactive. The knowledge gained from this exercise will certainly benefit future work in the development of novel TS inhibitors.

(ix) Summary

TS is still a protein very much under the attention of the scientific community. This work has shown that there is a large pocket within TS into which a bifurcated molecule can fit. Initial studies indicate that such a compound may have increased binding affinity for TS and hence be a more potent inhibitor of the enzyme. This work also indicates that TS may exist in two distinct forms, dependent on the levels of substrate present. It is not known if both active sites are simultaneously involved in synthesis of dTMP. I sincerely hope that this work will help another to develop a potent antifolate that will go at least part of the way to combating cancer -an illness for which there must be a solution.

REFERENCES

1. Rapp, P. and Wagner, F. (1986). *Appli. Environ. Microbiol.*, **51**, 746-764.

2. Pettipher, G.L. and Latham, M.J. (1979). *J.Gen.Microbiol.*, **110**, 21-27.

3. Gilbert, H.J., Jenkins, G., Sullivan, D.A. and Hall, J. (1987). *Mol. Gen. Genet.*, **210**, 551-556.

4. Esterban, R., Villanueva, J.R. and Villa, T.G. (1982). *Can. J. Microbiol.*, **28**, 733-739.

5. Hall, J., Hazelwood, G.P., Huskinsson, N.S., Durrant, A.J. and Gilbert, H.J. (1989). *Molecular Microbiol.*, **3**, 1211-1219.

6. Hall, J. and Gilbert, H.J. (1988). *Mol. Gen. Genet.*. **213**, 112-117.

7. Gilbert, H.J., Hall, J., Hazelwood, G.P. and Ferreira, L.M.A. (1990). *Mol. Microbiol.*, **4**, 759-767.

8. Ferreira, L.M.A., Durrant, A.J., Hall, J., Hazelwood, G.P. and Gilbert, H.J. (1990). *Biochem. J.*, **269**, 261-264.

9. Wolff, B.R., Mudry, T.A., Glick, B.R. and Pasternak, J.J. (1986). *Appl. Environ. Microbiol.*, **51**, 1367-1369.

10. Appelt, K., Bacquet, R.J., Bertlett, C.A., Booth, C.L.J., Freer, S.T., Fuhry, M.A.M., Gehring, M.R., Herrmann, S.M., Howland, E.F., Janson, C.A., Jones, T.R., Kan, C., Kathardekar, V., Lewis, K.K., Marzoni, G.P., Matthews, D.A., Mohr, C., Moomaw, E.W., Morse, C.A., Oatley, S.J., Ogden, R.C, Reddy, M.R., Reich, S.H., Schoettlin, W.S., Smith, W.W., Varney, M.D.,Villafranca, J.E., Ward, R.W., Webber, S., Webber, S.E., Welsh, K.M. and White, J (1991). *J. Med. Chem.*. **34**, 1925-1934.

11. Hol, W.G.J. (1986). *Angewandte Chemie.*, **25**, 767-852.

12. Rooman, M.J., and Wodak, S.J. (1988). *Nature*, **355**, 45-50.

13. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983). *J.Comp.Chem.*, **4**, 187-217.

14. Garnier, J., Osguthorpe, D.J. and Robson, B.(1978). *J.Mol.Biol.*, **120**, 97-120

15. Gilbrat, J.F., Garnier, J. and Robson, B. (1987). *J.Mol.Biol.*, **198**, 425-443.

16. Dembo, R.S. (1982). *Nonlinear Optimization. 1981* , **361** ed. Powell, M.J.D. Academic Press, New York.

17. Ponder, J.W. and Richards, F.M. (1987). *J.Comp.Chem.*, **8**, 1016-1024.

18. Norrander, J., Kemp, T. and Messing, J. (1983). *Gene*, **26**, 101-116.

19. Hawley, D.K. and McClure, W.R. (1983). *Nucleic Acid Res.*, **II**, 2237-2255.

20. Viera, J. and Messing, J. (1982). *Gene,* **19,** 259-268.

21. Gilbert, H.J., Blazek, R., Bullman, H.M.S. and Minton, N.P. (1986). *J.Gen. Microbiol.,* **132,** 151-160.

22. Biely, P., Mislovicoan, D. and Toman, R. (1985). *Anal. Biochem.,* **144,** 142-146.

23. Lee, B. and Richards, F.M. (1971). *J.Mol.Biol.,* **55,** 379-400.

24. Crippen, G.M. and Kuntz, I.D. (1978). *Int.J.Pept.Protein Res.,* **12,** 47-56.

25. Chou, P.Y. and Fasman, G.D.(1974). *Biochemisty,* 211-222, 222-245.

26. Biou, V., Gilbrat, J., F., Levin, J.M., Robson, B., and Garnier, J. (1988). *Protein. Engineering,* **2,** 185-191.

27. Lim, V.I. (1974). *J.Mol.Biol.,* **88,** 857-872, 873-894.

28. Hagler, A.T and Honig, B. (1978). *Proc. Natl. Acad. Sci. USA,* **75,** 554-558.

29. Cohen, F.E. and Sternberg, M.J.E. (1980). *J.Mol.Biol.,* **138,** 321-333.

30. Moult, J. and James, M.N.G. (1987). *Proteins,* **1,** 146-163.

31. Klein, P. and DeLisi, C. (1986). *Biopolymers,* **25,** 1659-1672.

32. Holley, L.H. and Karplus, M. (1989). *Proc.Natl. Acad. Sci. USA,* **86,** 152-156.

33. Bohr, H., Bohr, J., Brunak, S. and Cotterill, R.M.J. (1988). *FEBS Lett.,* **241,** 223-228.

34. Welling, G.W., Weiser, W.J., van der Zee, R. and Welling-Webster, S. (1985). *FEBS Lett..* **188,** 215-218.

35. Robson, B., Platt, E., Fishleigh, R. V., Marsden, A. and Millard, P. (1987). *J.Mol. Graphics,* **5,** 8-17.

36. Goraj, K., Renard, A. and Martial, J.A. (1990). *Protein Engineering,* **3,** 259-266.

37. Kabsch, W. and Sander, C. (1984). *Proc. Nattl. Acad. Sci. USA,* **81,** 1075-1078.

38. Levitt, M. and Chothia, C. (1976). *Nature,* **261,** 552-558.

39. Taylor, W.R. and Thornton, J.M.(1984). *J.Mol.Biol.,* **173,** 487-514.

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

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

42. Lebioda, L. and Stec, B. (1988). *Nature,* **333,** 683-686.

43. McDonald, P., Edwards, R.A. and Greenhalgh, J.F.D. (1981). *Silage,* 367-376. Animal Nutrition (3rd Edition), Longmans.

44. Hurle, M.R., Mattews, C.R., Cohen, F.E., Kuntz, I.D., Toumadje, A and Johnson, Jr, W.C. (1987). *Proteins,* **2,** 210-224.

45. Freidkin, M. (1973). *Adv. Enzymol.,* **38,** 235-335.

46. Danenberg, P.V. (1977). *Biochhimica et Biophysica Acta.,* **473,** 73-85.

47. Santi, D.V. and Brewer, C.F. (1973). *Biochemistry,* **12,** 2416.

48. Curtin, N.J., Harris, A.L., James, O.F.W. and Bassendine, M.F. (1986). *British Journal of Cancer,* **53,** 361-368.

49. Curtin, N.J., Harris, A.L. and Aherne, G.W. (1991). *Cancer Research,* **51,** 2346-2352.

50. Jones, T.R., Calvert, A.H., Jackman, A.L., Brown, S.J., Jones, M. and Harrap, K.R. (1981). *Europ.J.Cancer,* **17,** 11-19.

51. Jackson, R.C., Jackman, A.L. and Calvert, A.H. (1983). *Biochem. Pharmacol..* **24,** 3783-3790.

52. Jackman, A.L., Taylor, G.A., Calvert, A.H. and Harrap, K.R. (1984). *Biochem. Pharmacol.,* **33,** 3269-3275.

53. Newell, D.R., Alison, D.L., Calvert, A.H., Harrap, K.R., Jarman, M., Jones, T.R., Manteuffel-Cymborowska M. and O'Connor P. (1986). *Cancer Treatment Reports,* **70,** 971-978.

54. Sikora, E.W., Jackman, A.L., Newell, D.R. and Calvert, A.H. (1988). *Biochem. Pharmacol.*, **37,** 4047-4054.

55. Jackman, A.L., Taylor, G.A., O'Connor, B.M., Bishop, J.A., Moran, R.G. and Calvert, A.H. *Cancer Research,* **50,** 5212-5218.

56. Cantwell, B.M.J., Earnshaw, M. and Harris, A.L.. (1986). *Cancer Treatment Reports,* **70,** 1335-1336.

57. Sessa, C., Zucchetti, M., Ginier, M., Williams, Y., 'Incalci, M.D and Cavalli, F. (1988). *Eur. J. Clinical. Oncology,* **24,** 769-775.

58. Vest, S., Bork, E. and Hansen, H.H. (1988). *Eur. J. Clinical. Oncology,* **21,** 201-204.

59. Cantwell, B.M.J., Macaulay, V., Harris, A.L., Kaye, S.B., Smith, I.E., Milsted, R.A.V. and Calvert, A.H. (1988). *Eur. J. Clinical. Oncology,* **24,** 733-736.

60. Patil, S.D., Jones, C., Nair, M.G., Galivan, J., Maley, F., Kisliuk, R.L., Gaumont, Y., Duck, D. and Ferone, R.. (1989). *J.Med.Chem.*, **32,** 1284-1289.

61. Jones, T.R., Thornton, T.J., Flinn, A., Jackman, A.L., Newell, D.R. and Calvert ,A.H. (1989). *J.Med.Chem.*, **32,** 847-852.

62. Jackman, A.L., Newell, D.R., Gibson, W., Jodrell, D.I., Taylor, G.A., Bishop, J.A., Huges, L.R. and Calvert, A.H.. (1991). *Biochem. Pharmacol.*, **42**, 1885-1895.

63. Jackman, A.L., Taylor, G.A., Gibson, W., Kimbell, R., Brown, M., Clavert, A.H., Hudson, I.R. and Huges, L.R. (1991). *Cancer Research*, **51**, 5579-5586.

64. Ward, W.H.J., Kimbell, R. and Jackman, A.L. Short Communications. (1992). *Biochem. Pharmacol.*, **43**, 2029-2031.

65. Montfort, W.R., Perry, K.M., Fauman ,E.B., Finer-Moore, J.S., Maley, G.F., Hardy, L., Maley, F. and Stroud, R.M. (1990) *Bicohemistry*, **29**, 6964-6977.

66. Finer-Moore, J.S., Montfort, W.R. and Stroud, R.M. (1990) *Biochemistry*, **29**, 6978-6986.

67. Matthews, D.A., Appelt, K., Oatley, S.J. and Xuong, N.H. (1990) *J.Mol.Biol.*, **214**, 923-936.

68. Brooks, B.R., Bruccoleri, R.E. , Olafson, B.D., States, D.J., Swaminatham, S. and Karplus, M. (1983). *J.Comp.Chem.*, **4**, 187-217.

69. Born, M. and Oppenheimer, R. (1927). *Annalen der Physik.*, **74**. 457-484.

70. Nilsson, L. and Karplus, M. (1986). *J. Comp. Chem.*, **7**, 591-616.

71. Ponder, J.W. and Richards, F.M. (1987). *J.Comp.Chem.*, **8**, 1016-1024.

72. Connolly, M.L. (1983). *Science*, **221**, 709

73. Jones, T.R., Calvert, A.H., Jackman, A.L., Eakin, M.A., Smithers, M.J., Betteridge, R.F., Newell, D.R., Hayter, A.J., Stocker, A., Harland, S.J., Davies, L.C. and Harrap, K.R. (1985). *J. Med. Chem.*, **28**, 1468-1476.

74. Marsham, P.R., Chamber, P., Hayter, A.J., Huges, L.R., Jackman, A.L., O'Connor, B.M., Bishop, J.A.M. and Calvert, A.H. (1989). *J. Med. Chem.*, **32**, 569-575.

75. Jones, T.R., Betteridge, R.F., Newell, D.R. and Jackman, A.L. (1989). *Heterocyclic Chem.*, **26**, 1501-1507.

76. Pawelczak, K., Jones, T.R., Kempny, M., Jackman, A.L., Newell, D.R., Krzyzanowski, L. and Rzeszotarska, B. (1989). *J. Med. Chem.*, **32**, 160-165.

77. Marsham, P.R., Jackman, A.L., Oldfield, J., Huges, L.R., Thornton, T.J., Bisset, G.M.F., O'Connor, B.M., Bishop, J.A.M and Calvert, A.H. (1990). *J. Med. Chem.*, **33**, 3072-3078.

78. Jackman, A.L., Marsham, P.R., Thornton, T.J., Hughes, L.R., Calvert, A.H. and Jones, T.R. (1990). *J. Med. Chem.*, **33**, 3067-3071.

79. Seither, R.L., Trent, D.F., Mikulecky, D.C., Rape, T.J. and Goldman, I.D. (1991). *J. Biological Chem.*, **266**, 4112-4118.

80. Marsham, P.R., Huges, L.R., Jackman, A.L., Hayter, A.J., Oldfield, J., Wardelworth, J.M., Bishop, J.A.M., O'Connor, B.M. and Calvert, A.H. (1991). *J. Med. Chem.*, **34**, 1594-1605.

81. Thornton, T.J., Jones, T.R., Jackman, A.L., Flinn, A., O'Connor, B.M., Warner, P. and Calvert, A.H. (1991). *J. Med. Chem.*, **34**, 978-984.

82. Rosowsky, A., Forsch, R.A. and Moran, R.G. (1992). *J. Med. Chem.*, **35**, 2626-2630.

83. Rosowsky, A., Galivan, J., Beardsley, G.P., Bader, H., O'Connor, B.M., Russello, O., Moroson, B.A., DeYarman, M.T., Kerwar, S.S. and Freisheim, J.H. (1992). *Cancer Research*, **52**, 2148-2155.

84. Hardy, L.M., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V. and Stroud, R.M. (1987). *Science*, **235**, 448-455.

85. Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Montfort, W.R., Maley, C.F. and Stroud, R.M. (1990). *Proteins*, **8**, 315-333.

86. Cichowicz, D.J., Hynes, J.B. and Shane, B. (1988). *Biochimica et Biophysica Acta.*, **957**, 363-369.

87. Head-Gordon, T. and Brooks III, C.L. (1987). *J. Phys. Chem.*, **91**, 3342-3349.

88. Weiner, P.K., Langridge, R., Blaney, J.M., Schaeffer, R. and Kollman, P.A. (1982). *Proc. Natl. Acad. Sci. USA.* **79**, 3754-3758.

89. Buckert, U. and Allinger, N.L. (1982). *Molecular Mechnics*, American Chemical Societ; Washington D.C.

90. Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P. and Barry, D.C. (1990). *J. Med. Chem.*, **33**, 883-895.

91. Weiner, P.K. and Kollman, P.A. (1981), *J. Comp. Chem.*, **2**, 287.

92. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Jnr. and Weiner, P.K. (1984). *J. Am. Chem. Soc.*, **106**, 765.

93. Ripka, W.C. and Blaney, J.M. (1991). *Topics in Stereochemistry*, **20**, 1-83.

94. Hansch, C. (1969). *Acc. Chem. Res.*, **2**, 232.

95. Momany, F.A. and Rone, R. (1992). *J. Comp. Chem.*, **13**, 888-900.

96. Clark, M., Cramer III, R.D. and Van Opdenbosch, N. (1989). *J. Comp. Chem.*, **10**, 982.

97. Mayo, S.L., Olafson, B.D. and Goddard III, (1990). *J. Phys. Chem.*, **94**, 8897.

98. Momany, F.A. and Klimkowski, V.J. (1990). *J. Comp. Chem.*, **11**, 654-662.

99. Fleischman, S.H. and Brooks III, C.L. (1990). *Proteins,* **7,** 52-61.

100. Tattersall, M.H.N., Jackson, R.C., Jackson, S.T.M. and Harrap, K.R. (1974). *Europ. J. Cancer,* **10,** 819-826.

101. Calvert, A.H., Jones, T.R., Dady, P.J., Grzelakowska-Szyabert,B., Paine, R.M., Taylor, G.A. and Harrap, K.R. (1980). *Europ. J. Cancer,* **16,** 713-722.

102. Spears, C.P., Hayes, A.A., Shahinian, A.H. and Danenburg, P.V. (1989). *Biochemical Pharmacology,* **38,** 2985-2993.

103. Chen, S.C., Daron, H.H. and Aull, J.L. (1989). *Int. J. Biochem.,* **21,** 1217-1221.

104. Huges, L.R., Jackman, A.L., Oldfield, J., Smith, R.C., Burrows, K.D., Marsham, P.R., Bishop, J.A.M., Jones, T.R., O'Connor, B.M. and Calvert, A.H. (1990). *J. Med. Chem.,* **33,** 3060-3066.

105. Thorndike, J., Kisliuk, R.L., Gaumont, Y., Piper, J.R. and Nair, M.G. (1990)*Arch. Biochem. and Biophysics,* **2,** 334-341.

106. Bures, A.K., Daron, H.H. and Aull, J.L. (1991). *Int. J. Biochem.,* **23,** 733-736.

107. Kalman, T.I., Marinelli, E.R. Xu, B., Vengopala-Reddy, A.R., Johnson, F. and Grollman, A.P. (1991). *Biochem. Pharmacol.,* **42,** 431-437.

108. Peters, G.J., Van Groeningen, C.J., Laurensse, E.J. and Pinedo, H.M. (1991).*Eur. J. Cancer,* **27,** 263-267.

109. Hynes, J.B., Tomazic, A., Parrish, C.A. and Fetzer, O.S. (1991). *J. Heterocyclic Chem.*, **28,** 1357-1363.

110. Rosowsky, A., Galivan, J., Beardsley, G.P., Bader, H., O'Connor, B.M., Moroson, B.A., DeYarman, M.T., Kerwar, S.S. and Freisheim, J.H. (1992).*Cancer Research,* **52,** 2148-2155.

111. Webber, S.E., Bleckman, T.M., Attard, J., Deal, J.G., Kathardekar, V., Welsh, K.M., Webber, S., Janson, C.A., Matthews, D.A., Smith, W.W., Freer, S.T., Jordan, S.R., Bacquet, R.J., Howland, E.F., Booth, C.L.J., Ward, R.W., Hermann, S.M., White, J., Morse, C.A., Hilliard, J.A. and Bartlett, C.A. (1993). *J.Med.Chem.*, **36,** 733-746.

Appendix 1.1

Transformation of Religated Plasmids

Cultures of JM83 containing pNM52 were grown overnight at 37°C in an orbiting shaker by taking 1 ml of stock culture of the strain and inoculating aseptically into 10 ml of prewarmed L-broth (see Appendix 1.2) containing 10 µl TET^R .

The following day 1 ml of the JM83 culture was pipetted into a sterilised 2 litre flask containing 200 ml of pre-warmed L-broth and 200 µl TET^R .

The Cells were grown at 200 rpm until the optical density of the culture was 0.4 at 550 nm. The Culture was cooled on ice for 10 minutes to prevent further growth. Approximately 2 x 50 ml of the culture was removed and centrifuged at 4000 rpm for 4 mins at 4°C.
After the supernatant was discarded, the cells were resuspended in half the volume (25 ml) of ice cold sterile 0.1m $MgCl_2$ and recentrifuged at 4000 rpm for 4 mins at 4°C (Approximately 5000 x g).
After the supernatant was decanted off, each cell culture pellet was resuspended in 4 ml ice cold sterile 0.1m $CaCl_2$.The Cells were maintained on ice for 2 hours, after which time they were competent and took up the B41 and pN23++ plasmids.

10 µl of either pN23++ or B41 plasmid was added to 200 ml competent JM83 (containing pNM52) cells and placed on ice for 30 minutes in eppendorfs.
Transformants were placed in a 42°C water bath for exactly 2 minutes and transferred to ice for a further 30 minutes to allow the cells to recover and express resistance.

Appendix 1.2

1. L-Broth preparation

Compound	g/litre
Bactotryptone	10
Bactoyeast	5
NaCl	10

For a working solution of 4 litres L-broth, slightly less than 2 litres of single distilled water were poured into a 5 litre beaker. The above ingredients (x4) were added, and the pH raised to 7.5 (the optimum pH for E.coli growth) using 1M NaOH.
The solution was made up to 4 litres with single distilled water. The L-Broth was decanted into 500 ml duran bottles (up to the 400 ml mark). The bottles were autoclaved at 120^{o}C for 15 minutes.

2. Agar Plate preparation

A 2% bacto agar medium was prepared by adding 8g of bacto agar to 400 ml of L-Broth duran bottles. The bottles were autoclaved at 120^{o}C for 15 minutes. Following sterilisation, plates were poured immediately, allowed to set for 30 minutes and dried inverted in a 37^{o}C incubator for 1 hour.

3. L-Agar Plates containing antibiotics

The same procedure was followed as with L-agar plates, except that after sterilisation, the agar was cooled to 55^{o}C (as antibiotics are thermolabile). The following antibiotic concentrations were added depending on the media required:

Media	Concentration add per 400 ml L-Broth	Final antibiotic concentration
L-Agar and Ampicillin	4 ml of 10 mg/ml stock ampicllin solution added	100 µg/ml
L-Agar and Tetracycline	700 µl of a $^{1}/_{1000}$ stock tetracycline solution added	15 µg/ml

Plates were poured as described previously.

Appendix 1.3

Preparation of periplasmic fraction from Cultures

Cultures were kept on ice whenever possible. For 100 mls volume of culture: the culture was cooled on ice for 10 minutes to prevent further growth.
The cells were spun down (2 x 50 ml) at 800 rpm at 4°C for 10 minutes.
After supernatant was discarded, the cells were resuspended gently in 5 ml 20% sucrose in 10 mM tris buffer pH 8.0 (ice cold). To this mixture was then added 0.25 ml EDTA (0.5m) and the cells were maintained on ice for 30 minutes.

The cells were then recentrifuged at 8000 rpm at 4°C for 10 minutes. The supernatant was decanted (this was the "sucrose wash" - if needed, then this was transferred to eppendorfs).
The cells were resuspended vigorously in 5 ml ice cold sterile water and maintained on ice for a further 30 minutes.

Finally, the cells were recentrifuged at 8000 rpm at 4°C for 10 minutes and supernatant decanted to eppendorfs (total volume approximately 5 ml). This was the periplasmic protein fraction of the cell culture.

Appendix 1.4

Xylanase assay stock preparation

For 1.3%w/v xylan stock:
1.3g xylan made up to 100 ml

DNSA reagent:
1% w/v DNSA
0.2% w/v phenol,
1% w/v NaOH

(to activate DNSA just before use, DNSA reagent made up fresh for all assays.To 100 ml of DNSA reagent was added Xylose stock 1 mg/ml in PC buffer 1 ml 5% w/v Sodium sulphate, and 10µl 20% w/v glucose)

Xylanase assay

Standard curve *Figure A*

Xylose Stock µg/ml	$E_{575\ nm}$ Extinction	
0	0.000	
50	0.090	
100	0.170	
200	0.360	
300	0.550	
400	0.800	
500	0.900	
600	1.220	
700	1.350	
B41 plasmid	Sucrose wash	Periplasm
	0.150	0.975
	0.195	0.975
	0.210	1.245
	0.250	1.155
Average extinction value:	**0.201**	**1.088**
pN23 plasmid	Sucrose wash	Periplasm
	0.233	0.425
	0.275	0.805
	0.285	0.695
	0.225	0.470
Average extinction value:	**0.255**	**0.599**

Standard Curve *Figure B*

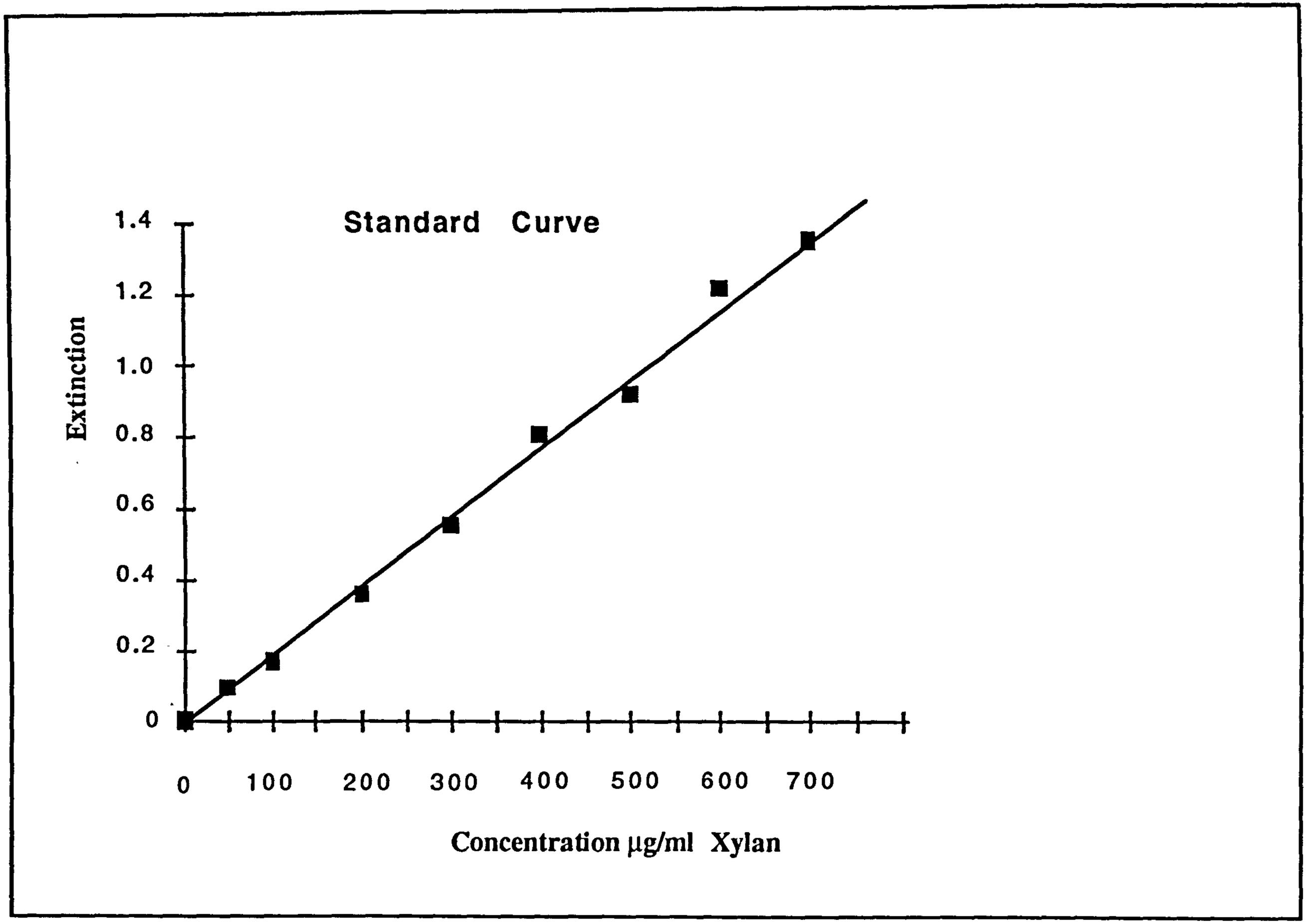
Xylose Stock µg/ml	E_{575nm} Extinction	
0	0.000	
200	0.230	
300	0.380	
400	0.500	
500	0.600	
600	0.720	
700	0.860	
1000	1.240	
B41 plasmid	Sucrose wash	Periplasm
	0.140	0.810
	0.190	1.060
	0.190	1.040
	0.080	1.050
Average extinction value:	**0.150**	**0.990**
pN23 plasmid	Sucrose wash	Periplasm
	0.150	0.580
	0.180	0.710
	0.100	0.790
	0.110	0.490
Average extinction value:	**0.135**	**0.643**

Standard curve ***Figure C***

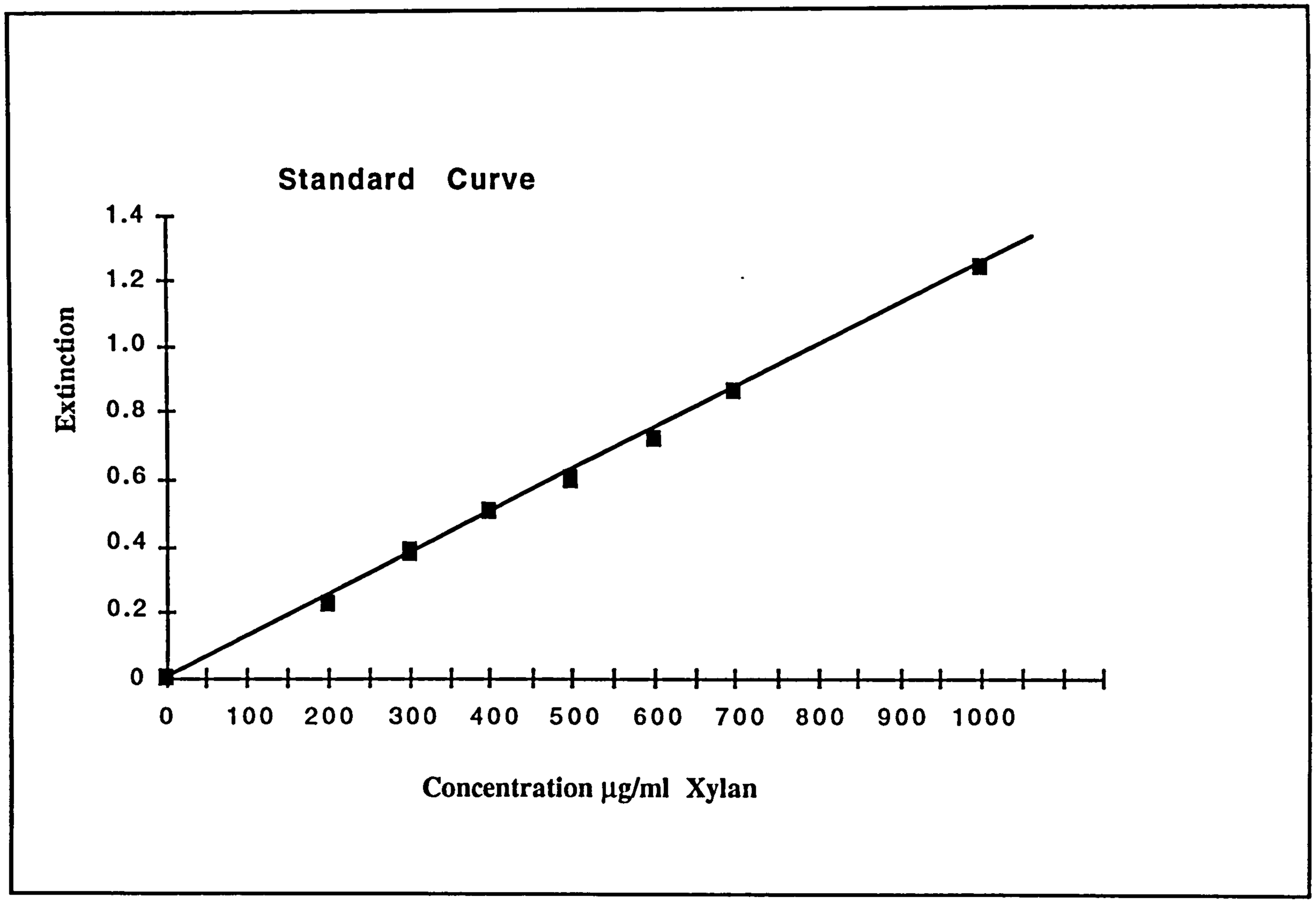
Xylose Stock µg/ml	E_{575nm} Extinction
0	0.000
50	0.100
100	0.170
200	0.320
300	0.500
400	0.610
500	0.810
600	0.910
700	1.080
800	1.140
B41 plasmid	0.000 (-0.130)
B41 plasmid + IPTG	1.675
pN23 plasmid	0.000 (-0.123)
pN23 plasmid + IPTG	1.005

Standard curve ***Figure D***

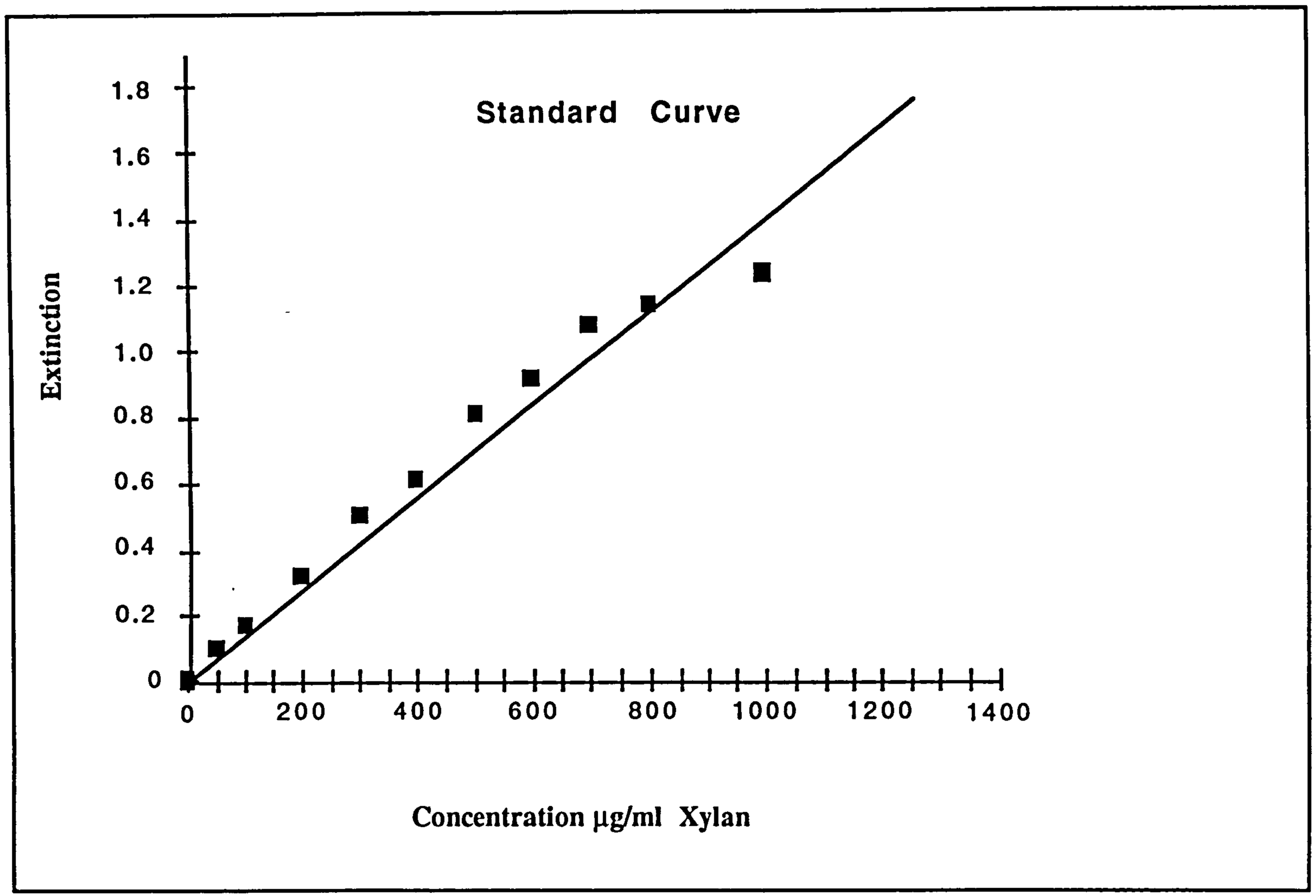
Xylose Stock µg/ml				E_{575nm} extinction
0				0.000
100				0.220
200				0.420
500				1.080
800				1.650
Time after IPTG induction (mins)	B41 neat periplasm	B41 (10^{-1})	pN23 neat periplasm	pN23 (10^{-1})
	E_{575nm}	E_{575nm}	E_{575nm}	E_{575nm}
90	0.245	0.167	0.178	0.250
120	0.832	0.329	0.829	0.176
270	1.006	0.666	1.931	0.553
360	2.018	0.781	1.262	0.527
960	1.758	0.938	1.987	0.859



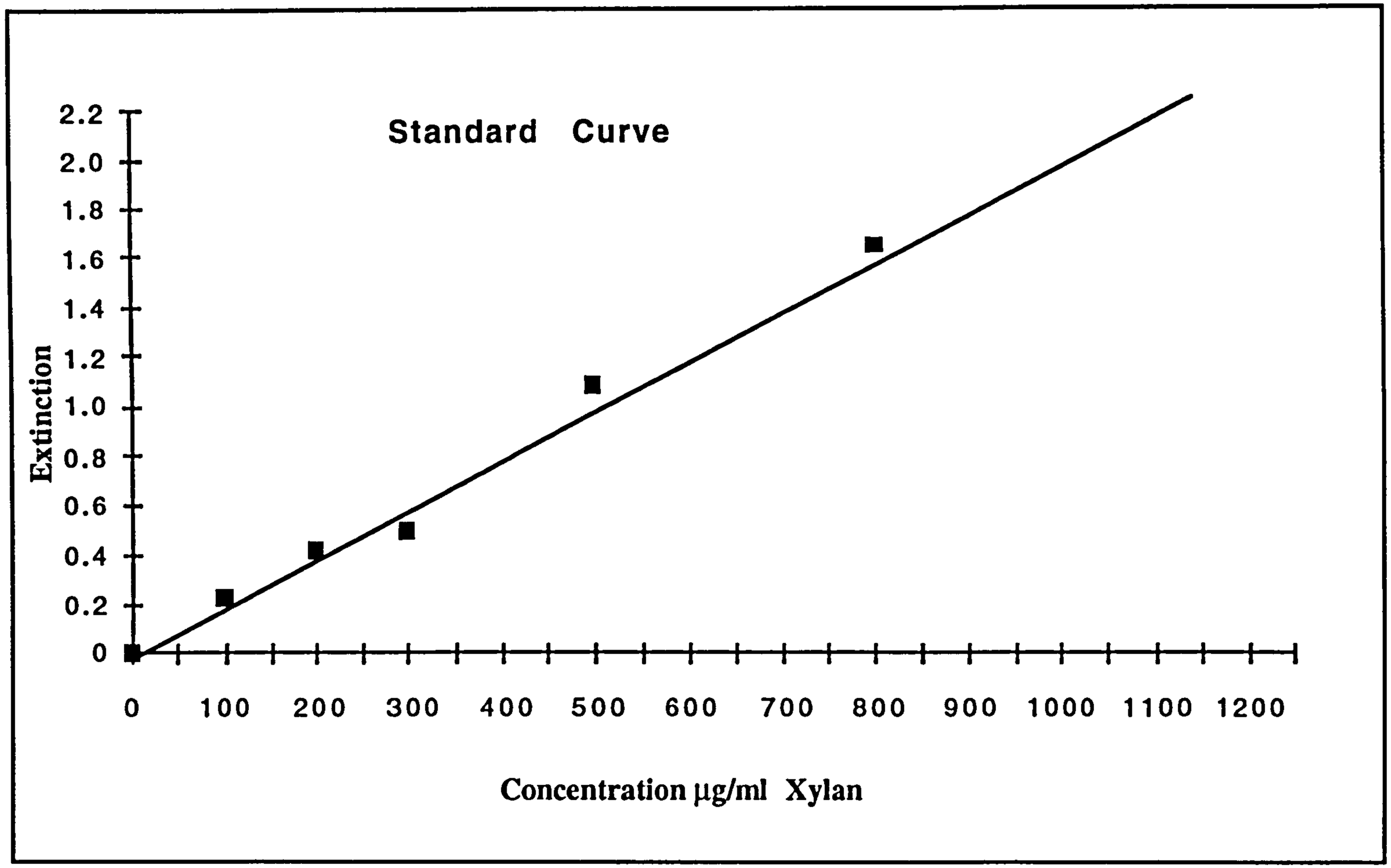
Xylanase Assay - Standard Curve ***Figure A***



Xylanase Assay - Standard Curve *Figure B*



Xylanase Assay - Standard Curve ***Figure C***



Xylanase Assay - Standard Curve ***Figure D***

Appendix 2.1

Custom parameter set

```
BOND

C     CR66    386.0          1.419   !* from C CR56
CH2E  CUY1    350.0          1.46    !* from CT CUY1


THETAS
C     C6R   NT        70.0      120.0   !* from C C6R NP
N6R   C6R   NT        70.0      120.0   !* from N6R C6R N5R
H     NT    C6R       35.0      109.47  !* from H NT C
CR66  C     N6R       62.0      111.3   !* from CR56 C N6R
CR66  C     O         86.0      128.8   !* from CR56 C O
C     CR66  CR66     116.0      119.2   !* from C CR56 CR56
C     CR66  C6R       70.0      120.0   !* from C6R CR66 C6R
C     CR66  C6RE      70.0      120.0   !* from C6R CR66 C6R
CUY1  CH2E  NT        50.0      109.47  !* from CUY1 CT NT
CH2E  CUY1  CUY1      25.0      180.0   !* from CT CUY1 CUY1
CH2E  C     N         58.0      120.0   !* from CT C N
C     C6R   N         70.0      120.0   !* from C C6R NP
C6RE  C     NP        70.0      117.5   !* from C6R C NP
CUY1  CH2E  NP        50.0      109.47  !* from CUY1 CT NT
C6R   C6RE  C         70.0      120.0   !* from C6R C6R C
C6RE  C     CR66      65.0      120.0   !* from C6RE C6R CR66


PHI
X     C6RE  C     X          3.1   2   180.0   !* from X C6RE C6RE X
X     CH2E  CUY1  X          0.4   3     0.0   !* from X CT CUY1 X
X     C6RE  CR66  X          3.1   2   180.0   !* from X C6R CR66 X
X     C     CR66  X          3.1   2   180.0   !* from X C6R CR66 X
```

Appendix 2.2

Residue Topology Files

PHOSPHATE and dUMP

CB3717

```
* ...
* Proprietory and Confidential Property of Polygen Corporation
* ....
* TEMPLATE FOR CREATING CHARMm RTF file (from ChemNote)
* ....
*      (%_m) replaced by first 4 characters of molecule name
*      (%_r) replaced by RTF format of residue definition
* ...
* Build RTF for phs
* ...
*  residue definition in RTF format built inline
*
   21     3
MASS     1 H       1.00800 ! Hydrogen bonding hydrogen (neutral group)
MASS     2 HC      1.00800 ! Hydrogen bonding hydrogen (charged group)
MASS     3 HA      1.00800 ! Aliphatic or aromatic hydrogen
MASS     4 HT      1.00800 ! TIPS3P water model hydrogen
MASS     5 LP      0.0     ! ST2 lone pair
MASS     6 MBE     9.01218 ! Beryllium
MASS     7 B      10.81    ! Boron
MASS     8 HO     1.00800  ! Hydrogen on an alcohol oxygen
MASS     9 HMU    1.00800  ! Mu-bonded hydrogen for metals and boron-hydride
MASS    10 CT    12.01100 ! Aliphatic carbon (tetrahedral)
MASS    11 CH1E  13.01900 ! Extended atom carbon with one hydrogen
MASS    12 CH2E  14.02700 ! Extended atom carbon with two hydrogens
MASS    13 CH3E  15.03500 ! Extended atom carbon with three hydrogens
MASS    14 C     12.01100 ! Carbonyl or Guanidinium carbon
MASS    15 CM    12.01100 ! Carbonmonoxide, or other triply bonded, carbon
MASS    16 CUA1  12.01100 ! Carbon in double bond, first pair
MASS    17 CUA2  12.01100 ! Carbon in double bond, second conjugated pair
MASS    18 CUY1  12.01100 ! Carbon in triple bond, first pair
MASS    19 CUY2  12.01100 ! Carbon in triple bond, second conjugated pair
MASS    20 CUA3  12.01100 ! Carbon in double bond, third conjugated pair
MASS    21 C5R   12.01100 ! Aromatic carbon in a five membered ring
MASS    22 C6R   12.01100 ! Aromatic carbon in a six  membered ring
MASS    23 C5RE  13.01900 ! Extended aromatic carbon in five membered ring
MASS    24 C6RE  13.01900 ! Extended aromatic carbon in six membered ring
MASS    25 CR55  12.01100 ! Aromatic carbon-merged five membered rings
MASS    26 CR56  12.01100 ! Aromatic carbon-merged five/six membered rings
MASS    27 CR66  12.01100 ! Aromatic carbon-merged six membered rings
MASS    28 C5RP  12.01100 ! for Aryl-Aryl bond between C5R rings
MASS    29 C6RP  12.01100 ! for Aryl-Aryl bond between C6R rings
MASS    30 N5RP  14.00670 ! for Ar-Ar bond between five membered rings
MASS    31 N     14.00670 ! Nitrogen; planar-valence of 3, i.e. nitrile, et
MASS    32 NP    14.00670 ! Nitrogen in peptide, amide, or related, group
MASS    33 NX    14.00670 ! Proline nitrogen, or similar
MASS    34 N5R   14.00670 ! Nitrogen in a five membered aromatic ring
MASS    35 N6R   14.00670 ! Nitrogen in a six membered aromatic ring
MASS    36 NT    14.00670 ! Nitrogen (tetrahedral), i.e. Amine, etc.
MASS    37 NC    14.00670 ! Charged guanidinuim-type nitrogen
MASS    38 NO2   14.00670 ! Nitrogen in nitro, or related, group
MASS    39 N6RP  14.00670 ! for Aryl-Aryl bond between six membered rings
MASS    40 O     15.99940 ! Carbonyl oxygen for amides, or related structur
MASS    41 OA    15.99940 ! Carbonyl oxygen for aldehydes, or related struc
MASS    42 OK    15.99940 ! Carbonyl oxygen for ketones, or related structu
MASS    43 OC    15.99940 ! Charged oxygen
MASS    45 OT    15.99940 ! Hydroxyl oxygen (tetrahedral)/Ionizable acid ox
MASS    46 OW    15.99940 ! TIP3P water model oxygen
```

```
MASS     47 OH2    15.99940 ! ST2  water model oxygen
MASS     48 OM     15.99940 ! Carbonmonoxide, or other triply bonded, oxygen
MASS     49 OS     15.99940 ! Ester oxygen
MASS     50 OE     15.99940 ! Ether oxygen / Acetal oxygen
MASS     51 OAC    15.99940 ! Carbonyl oxygen for acids
MASS     52 O5R    15.99940 ! Oxygen in five membered aromatic ring-radicals,
MASS     53 O6R    15.99940 ! Oxygen in aromatic ring - radicals etc.
MASS     55 OSI    15.99940 ! Oxygen in Si-O-Si bond
MASS     56 O2M    15.99940 ! Oxygen in Si-O-Al or Al-O-Al bond
MASS     57 OSH     0.00000 ! Massless O for zeolites, or related cage compou
MASS     60 PT     30.9738  ! Phosphorous, general; usually tetrahedral
MASS     61 PO3    30.9738  ! Phosphorous bonded to three oxygens
MASS     62 PO4    30.9738  ! Phosphorous bonded to four oxygens
MASS     63 PUA1   30.9738  ! Phosphorous double bond
MASS     64 P6R    30.9738  ! Phosphorous in aromatic 6-membered ring
MASS     70 ST     32.06000 ! Sulphur, general; usually tetrahedral
MASS     71 SH1E   33.06800 ! Extended atom sulphur with one hydrogen
MASS     72 S5R    32.0600  ! Sulphur in a five membered aromatic ring
MASS     73 S6R    32.0600  ! Sulphur in a six membered aromatic ring
MASS     74 SE     32.060   ! Thioether sulphur
MASS     75 SK     32.060   ! Thioketone sulphur
MASS     76 SO1    32.0600  ! Sulphur bonded to one oxygen
MASS     77 SO2    32.0600  ! Sulphur bonded to two oxygens
MASS     78 SO3    32.0600  ! Sulphur bonded to three oxygens
MASS     79 SO4    32.0600  ! Sulphur bonded to four oxygens
MASS     80 MLI     6.941   ! Lithium
MASS     81 MNA    22.9898  ! Sodium
MASS     82 MMG    24.305   ! Magnesiun
MASS     83 MK     39.098   ! Potassium
MASS     84 MCA    40.080   ! Calcium
MASS     85 MMN    54.938   ! Manganese
MASS     86 MFE    55.847   ! Iron
MASS     87 MZN    65.38    ! Zinc
MASS     88 MRB    85.4678  ! Rubidium
MASS     89 MCS   132.9054  ! Cesium
MASS     90 MSI    28.0855  ! Silicon
MASS     91 MAL    26.9815  ! ALuminum
MASS     92 XF     18.99840 ! Fluorine
MASS     93 XCL    35.45300 ! Chlorine
MASS     94 XBR    79.904   ! Bromine
MASS     95 XI    126.9045  ! Iodine
MASS     96 MCU    63.546   ! Copper
MASS     97 MV     50.9415  ! Vanadium
MASS     98 MCR    51.996   ! Chromium
MASS     99 MCO    58.9332  ! Cobalt
MASS    100 MNI    58.69    ! Nickel
MASS    101 MAS    74.9216  ! Arsenic
MASS    102 MSE    78.96    ! Selenium
MASS    103 MSR    87.62    ! Strontium
MASS    104 MY     88.9059  ! Yttrium
MASS    105 MZR    91.22    ! Zirconium
MASS    106 MNB    92.9064  ! Niobium
MASS    107 MMO    95.94    ! Molybdenum
MASS    108 MRU   101.07    ! Ruthenium
MASS    109 MRH   102.9055  ! Rhodium
MASS    110 MPD   106.42    ! Palladium
MASS    111 MAG   107.868   ! Silver
MASS    112 MCD   112.41    ! Cadmium
```

```
MASS    113 MSN   118.69     ! Tin
MASS    114 MSB   121.75     ! Antimony
MASS    115 MBA   137.33     ! Barium
MASS    116 MW    183.85     ! Tungsten
MASS    117 MOS   190.2      ! Osmium
MASS    118 MPT   195.08     ! Platinum
MASS    119 MAU   196.08     ! Gold
MASS    120 MHG   200.59     ! Mercury
MASS    121 MPB   207.2      ! Lead
MASS    122 MBI   208.9804   ! Bismuth
MASS    123 MLA   138.9055   ! Lanthanum
MASS    124 MCE   140.12     ! Cerium
MASS    125 MPR   140.9077   ! Praseodymium
MASS    126 MAC   227.0278   ! Actinium
MASS    127 MTH   232.0381   ! Thorium
MASS    128 MU    238.0289   ! Uranium
MASS    129 MTE   127.60     ! Tellurium
MASS    130 MPO   209.0      ! Polonium
MASS    131 XAT   210.0      ! Astatine
MASS    132 MTC    98.0      ! Technetium
MASS    133 MSC    44.9559   ! Scandium
MASS    134 MTI    47.88     ! Titanium
MASS    135 MGA    69.72     ! Gallium
MASS    136 MGE    72.59     ! Germanium
MASS    137 MIN   114.82     ! Indium
MASS    138 MHF   178.49     ! Hafnium
MASS    139 MTA   180.9479   ! Tantalum
MASS    140 MIR   192.22     ! Iridium
MASS    141 MTL   204.383    ! Thallium
MASS    142 MFR   223.0      ! Francium
MASS    143 MRA   226.0254   ! Radium
MASS    144 MND   144.24     ! Neodymium
MASS    145 MPM   145.0      ! Promethium
MASS    146 MSM   150.36     ! Samarium
MASS    147 MEU   151.96     ! Europium
MASS    148 MGD   157.25     ! Gadolinium
MASS    149 MTB   158.9254   ! Terbium
MASS    150 MDY   162.50     ! Dysprosium
MASS    151 MHO   164.9304   ! Holmium
MASS    152 MER   167.26     ! Erbium
MASS    153 MTM   168.9342   ! Thulium
MASS    154 MYB   173.04     ! Ytterbium
MASS    155 MLU   174.967    ! Lutetium
MASS    156 MPA   231.0359   ! Protactinium
MASS    157 MNP   237.0482   ! Neptunium
MASS    158 MPU   244.0      ! Plutonium
MASS    159 MAM   243.0      ! Americium
MASS    160 MCM   247.0      ! Curium
MASS    161 MBK   247.0      ! Berkelium
MASS    162 MCF   251.0      ! Californium
MASS    163 MES   252.0      ! Einsteinium
MASS    164 MFM   257.0      ! Fermium
MASS    165 MMD   258.0      ! Medelevium
MASS    166 MNO   259.0      ! Nobelium
MASS    167 MLR   260.0      ! Lawrencium
MASS    168 HE      4.00260  ! Helium
MASS    169 NE     20.179    ! Neon
MASS    170 AR     39.948    ! Argon
```

```
MASS   171 KR     83.80    ! Krypton
MASS   172 XE    131.29    ! Xenon
MASS   173 RN    222.0     ! Radon
MASS   174 MRF   261.0     ! Rutherfordium/Kurchatovium
MASS   175 MHA   262.0     ! Hahnium
MASS   176 MRE   186.31    ! Rhenium
MASS   177 MSIU  28.08550  ! Silicon when as double bond
MASS   180 NR56  14.00670  ! N at fused bond between 5 and 6-membered aromat
MASS   181 NR66  14.00670  ! N at fused bond between two 6-membered aromatic
MASS   182 NR55  14.00670  ! N at fused bond between two 5-membered aromatic
MASS   183 NR1   14.00670  ! Protonated nitrogen in neutral histidine ring
MASS   184 NR2   14.00670  ! Unprotonated nitrogen in neutral histidine ring
MASS   185 NR3   14.00670  ! Nitrogens in charged histidine ring
MASS   186 NC2   14.00670  ! for neutral guanidinium group - Arg sidechain
MASS   189 C5RQ  12.01100  ! for second Ar-Ar bond between C5RP rings (ortho
MASS   190 C3    12.01100  ! Carbonyl carbon in 3-membered aliphatic ring
MASS   191 CT3   12.01100  ! C in 3-membered aliphatic ring, usually tetrahe
MASS   192 C4    12.01100  ! Carbonyl carbons in 4-membered aliphatic ring
MASS   193 CT4   12.01100  ! C in 4-membered aliphatic ring, usually tetrahe
MASS   194 C6RQ  12.01100  ! Carbon of C6RP type ortho to C6RP pair
MASS   195 CQ66  12.01100  ! Third adjacent pair of CR66 types in fused ring
MASS   196 CS66  12.01100  ! Second adjacent pair of CR66 types in fused rin
MASS   197 CPH1  12.01100  ! CG and CD2 carbons in histidine ring
MASS   198 CPH2  12.01100  ! CE1 carbon in histidine ring
MASS   199 CP3   12.01100  ! Carbon on nitrogen in proline ring

AUTOGENERATE ANGLES DIHE
DEFA FIRS NONE LAST NONE

RESI phs -1.520
GROUP
ATOM P  PO4   1.36        ! n.b. these taken from DNA.RTF's for Thimine
ATOM OP1 OC  -0.72        !   taken on 3.10.91
ATOM OP2 OC  -0.72        ! note much higher charge than old RTF's
ATOM OP3 OC  -0.72
ATOM OP4 OC  -0.72
BOND P OP1
BOND P OP2
BOND P OP3
BOND P OP4
ACCE OP1 P
ACCE OP2 P
ACCE OP3 P
ACCE OP4 P

END
```

```
* ...
* Proprietory and Confidential Property of Polygen Corporation
* ....
* TEMPLATE FOR CREATING CHARMm RTF file (from ChemNote)
* ....
*      (%_m) replaced by first 4 characters of molecule name
*      (%_r) replaced by RTF format of residue definition
* ...
* Build RTF for dum
* ...
*  residue definition in RTF format built inline
*
   21    3
MASS     1 H       1.00800 ! Hydrogen bonding hydrogen (neutral group)
MASS     2 HC      1.00800 ! Hydrogen bonding hydrogen (charged group)
MASS     3 HA      1.00800 ! Aliphatic or aromatic hydrogen
MASS     4 HT      1.00800 ! TIPS3P water model hydrogen
MASS     5 LP      0.0     ! ST2 lone pair
MASS     6 MBE     9.01218 ! Beryllium
MASS     7 B      10.81    ! Boron
MASS     8 HO      1.00800 ! Hydrogen on an alcohol oxygen
MASS     9 HMU     1.00800 ! Mu-bonded hydrogen for metals and boron-hydride
MASS    10 CT     12.01100 ! Aliphatic carbon (tetrahedral)
MASS    11 CH1E   13.01900 ! Extended atom carbon with one hydrogen
MASS    12 CH2E   14.02700 ! Extended atom carbon with two hydrogens
MASS    13 CH3E   15.03500 ! Extended atom carbon with three hydrogens
MASS    14 C      12.01100 ! Carbonyl or Guanidinium carbon
MASS    15 CM     12.01100 ! Carbonmonoxide, or other triply bonded, carbon
MASS    16 CUA1   12.01100 ! Carbon in double bond, first pair
MASS    17 CUA2   12.01100 ! Carbon in double bond, second conjugated pair
MASS    18 CUY1   12.01100 ! Carbon in triple bond, first pair
MASS    19 CUY2   12.01100 ! Carbon in triple bond, second conjugated pair
MASS    20 CUA3   12.01100 ! Carbon in double bond, third conjugated pair
MASS    21 C5R    12.01100 ! Aromatic carbon in a five membered ring
MASS    22 C6R    12.01100 ! Aromatic carbon in a six  membered ring
MASS    23 C5RE   13.01900 ! Extended aromatic carbon in five membered ring
MASS    24 C6RE   13.01900 ! Extended aromatic carbon in six membered ring
MASS    25 CR55   12.01100 ! Aromatic carbon-merged five membered rings
MASS    26 CR56   12.01100 ! Aromatic carbon-merged five/six membered rings
MASS    27 CR66   12.01100 ! Aromatic carbon-merged six membered rings
MASS    28 C5RP   12.01100 ! for Aryl-Aryl bond between C5R rings
MASS    29 C6RP   12.01100 ! for Aryl-Aryl bond between C6R rings
MASS    30 N5RP   14.00670 ! for Ar-Ar bond between five membered rings
MASS    31 N      14.00670 ! Nitrogen; planar-valence of 3, i.e. nitrile, et
MASS    32 NP     14.00670 ! Nitrogen in peptide, amide, or related, group
MASS    33 NX     14.00670 ! Proline nitrogen, or similar
MASS    34 N5R    14.00670 ! Nitrogen in a five membered aromatic ring
MASS    35 N6R    14.00670 ! Nitrogen in a six membered aromatic ring
MASS    36 NT     14.00670 ! Nitrogen (tetrahedral), i.e. Amine, etc.
MASS    37 NC     14.00670 ! Charged guanidinuim-type nitrogen
MASS    38 NO2    14.00670 ! Nitrogen in nitro, or related, group
MASS    39 N6RP   14.00670 ! for Aryl-Aryl bond between six membered rings
MASS    40 O      15.99940 ! Carbonyl oxygen for amides, or related structur
MASS    41 OA     15.99940 ! Carbonyl oxygen for aldehydes, or related struc
MASS    42 OK     15.99940 ! Carbonyl oxygen for ketones, or related structu
MASS    43 OC     15.99940 ! Charged oxygen
MASS    45 OT     15.99940 ! Hydroxyl oxygen (tetrahedral)/Ionizable acid ox
MASS    46 OW     15.99940 ! TIP3P water model oxygen
```

```
MASS     47 OH2    15.99940 ! ST2  water model oxygen
MASS     48 OM     15.99940 ! Carbonmonoxide, or other triply bonded, oxygen
MASS     49 OS     15.99940 ! Ester oxygen
MASS     50 OE     15.99940 ! Ether oxygen / Acetal oxygen
MASS     51 OAC    15.99940 ! Carbonyl oxygen for acids
MASS     52 O5R    15.99940 ! Oxygen in five membered aromatic ring-radicals,
MASS     53 O6R    15.99940 ! Oxygen in aromatic ring - radicals etc.
MASS     55 OSI    15.99940 ! Oxygen in Si-O-Si bond
MASS     56 O2M    15.99940 ! Oxygen in Si-O-Al or Al-O-Al bond
MASS     57 OSH     0.00000 ! Massless O for zeolites, or related cage compou
MASS     60 PT     30.9738  ! Phosphorous, general; usually tetrahedral
MASS     61 PO3    30.9738  ! Phosphorous bonded to three oxygens
MASS     62 PO4    30.9738  ! Phosphorous bonded to four oxygens
MASS     63 PUA1   30.9738  ! Phosphorous double bond
MASS     64 P6R    30.9738  ! Phosphorous in aromatic 6-membered ring
MASS     70 ST     32.06000 ! Sulphur, general; usually tetrahedral
MASS     71 SH1E   33.06800 ! Extended atom sulphur with one hydrogen
MASS     72 S5R    32.0600  ! Sulphur in a five membered aromatic ring
MASS     73 S6R    32.0600  ! Sulphur in a six membered aromatic ring
MASS     74 SE     32.060   ! Thioether sulphur
MASS     75 SK     32.060   ! Thioketone sulphur
MASS     76 SO1    32.0600  ! Sulphur bonded to one oxygen
MASS     77 SO2    32.0600  ! Sulphur bonded to two oxygens
MASS     78 SO3    32.0600  ! Sulphur bonded to three oxygens
MASS     79 SO4    32.0600  ! Sulphur bonded to four oxygens
MASS     80 MLI     6.941   ! Lithium
MASS     81 MNA    22.9898  ! Sodium
MASS     82 MMG    24.305   ! Magnesiun
MASS     83 MK     39.098   ! Potassium
MASS     84 MCA    40.080   ! Calcium
MASS     85 MMN    54.938   ! Manganese
MASS     86 MFE    55.847   ! Iron
MASS     87 MZN    65.38    ! Zinc
MASS     88 MRB    85.4678  ! Rubidium
MASS     89 MCS   132.9054  ! Cesium
MASS     90 MSI    28.0855  ! Silicon
MASS     91 MAL    26.9815  ! ALuminum
MASS     92 XF     18.99840 ! Fluorine
MASS     93 XCL    35.45300 ! Chlorine
MASS     94 XBR    79.904   ! Bromine
MASS     95 XI    126.9045  ! Iodine
MASS     96 MCU    63.546   ! Copper
MASS     97 MV     50.9415  ! Vanadium
MASS     98 MCR    51.996   ! Chromium
MASS     99 MCO    58.9332  ! Cobalt
MASS    100 MNI    58.69    ! Nickel
MASS    101 MAS    74.9216  ! Arsenic
MASS    102 MSE    78.96    ! Selenium
MASS    103 MSR    87.62    ! Strontium
MASS    104 MY     88.9059  ! Yttrium
MASS    105 MZR    91.22    ! Zirconium
MASS    106 MNB    92.9064  ! Niobium
MASS    107 MMO    95.94    ! Molybdenum
MASS    108 MRU   101.07    ! Ruthenium
MASS    109 MRH   102.9055  ! Rhodium
MASS    110 MPD   106.42    ! Palladium
MASS    111 MAG   107.868   ! Silver
MASS    112 MCD   112.41    ! Cadmium
```

```
MASS    113 MSN   118.69     ! Tin
MASS    114 MSB   121.75     ! Antimony
MASS    115 MBA   137.33     ! Barium
MASS    116 MW    183.85     ! Tungsten
MASS    117 MOS   190.2      ! Osmium
MASS    118 MPT   195.08     ! Platinum
MASS    119 MAU   196.08     ! Gold
MASS    120 MHG   200.59     ! Mercury
MASS    121 MPB   207.2      ! Lead
MASS    122 MBI   208.9804   ! Bismuth
MASS    123 MLA   138.9055   ! Lanthanum
MASS    124 MCE   140.12     ! Cerium
MASS    125 MPR   140.9077   ! Praseodymium
MASS    126 MAC   227.0278   ! Actinium
MASS    127 MTH   232.0381   ! Thorium
MASS    128 MU    238.0289   ! Uranium
MASS    129 MTE   127.60     ! Tellurium
MASS    130 MPO   209.0      ! Polonium
MASS    131 XAT   210.0      ! Astatine
MASS    132 MTC    98.0      ! Technetium
MASS    133 MSC    44.9559   ! Scandium
MASS    134 MTI    47.88     ! Titanium
MASS    135 MGA    69.72     ! Gallium
MASS    136 MGE    72.59     ! Germanium
MASS    137 MIN   114.82     ! Indium
MASS    138 MHF   178.49     ! Hafnium
MASS    139 MTA   180.9479   ! Tantalum
MASS    140 MIR   192.22     ! Iridium
MASS    141 MTL   204.383    ! Thallium
MASS    142 MFR   223.0      ! Francium
MASS    143 MRA   226.0254   ! Radium
MASS    144 MND   144.24     ! Neodymium
MASS    145 MPM   145.0      ! Promethium
MASS    146 MSM   150.36     ! Samarium
MASS    147 MEU   151.96     ! Europium
MASS    148 MGD   157.25     ! Gadolinium
MASS    149 MTB   158.9254   ! Terbium
MASS    150 MDY   162.50     ! Dysprosium
MASS    151 MHO   164.9304   ! Holmium
MASS    152 MER   167.26     ! Erbium
MASS    153 MTM   168.9342   ! Thulium
MASS    154 MYB   173.04     ! Ytterbium
MASS    155 MLU   174.967    ! Lutetium
MASS    156 MPA   231.0359   ! Protactinium
MASS    157 MNP   237.0482   ! Neptunium
MASS    158 MPU   244.0      ! Plutonium
MASS    159 MAM   243.0      ! Americium
MASS    160 MCM   247.0      ! Curium
MASS    161 MBK   247.0      ! Berkelium
MASS    162 MCF   251.0      ! Californium
MASS    163 MES   252.0      ! Einsteinium
MASS    164 MFM   257.0      ! Fermium
MASS    165 MMD   258.0      ! Medelevium
MASS    166 MNO   259.0      ! Nobelium
MASS    167 MLR   260.0      ! Lawrencium
MASS    168 HE      4.00260  ! Helium
MASS    169 NE     20.179    ! Neon
MASS    170 AR     39.948    ! Argon
```

```
MASS   171 KR     83.80    ! Krypton
MASS   172 XE    131.29    ! Xenon
MASS   173 RN    222.0     ! Radon
MASS   174 MRF   261.0     ! Rutherfordium/Kurchatovium
MASS   175 MHA   262.0     ! Hahnium
MASS   176 MRE   186.31    ! Rhenium
MASS   177 MSIU  28.08550  ! Silicon when as double bond
MASS   180 NR56  14.00670  ! N at fused bond between 5 and 6-membered aromat
MASS   181 NR66  14.00670  ! N at fused bond between two 6-membered aromatic
MASS   182 NR55  14.00670  ! N at fused bond between two 5-membered aromatic
MASS   183 NR1   14.00670  ! Protonated nitrogen in neutral histidine ring
MASS   184 NR2   14.00670  ! Unprotonated nitrogen in neutral histidine ring
MASS   185 NR3   14.00670  ! Nitrogens in charged histidine ring
MASS   186 NC2   14.00670  ! for neutral guanidinium group - Arg sidechain
MASS   189 C5RQ   12.01100 ! for second Ar-Ar bond between C5RP rings (ortho
MASS   190 C3    12.01100  ! Carbonyl carbon in 3-membered aliphatic ring
MASS   191 CT3   12.01100  ! C in 3-membered aliphatic ring, usually tetrahe
MASS   192 C4    12.01100  ! Carbonyl carbons in 4-membered aliphatic ring
MASS   193 CT4   12.01100  ! C in 4-membered aliphatic ring, usually tetrahe
MASS   194 C6RQ  12.01100  ! Carbon of C6RP type ortho to C6RP pair
MASS   195 CQ66  12.01100  ! Third adjacent pair of CR66 types in fused ring
MASS   196 CS66  12.01100  ! Second adjacent pair of CR66 types in fused rin
MASS   197 CPH1  12.01100  ! CG and CD2 carbons in histidine ring
MASS   198 CPH2  12.01100  ! CE1 carbon in histidine ring
MASS   199 CP3   12.01100  ! Carbon on nitrogen in proline ring

AUTOGENERATE ANGLES DIHE
DEFA FIRS NONE LAST NONE

RESI DUM -1.000
GROUP
ATOM P PO4   1.356
ATOM OP1 OT -0.603
ATOM OP2 OC -0.623
ATOM OP3 OC -0.623
ATOM OR5 OS -0.463
ATOM CR5 CH2E  0.067
ATOM CR4 CH1E  0.097
ATOM CR3 CH1E -0.003
ATOM OR3 OT -0.653
ATOM CR2 CH2E  0.067
ATOM CR1 CH1E  0.197
ATOM OR1 OE -0.373
ATOM N1 N6R -0.183
ATOM C2 C   0.797
ATOM O2 O -0.553
ATOM N3 NP -0.813
ATOM C4 C   0.867
ATOM O4 O -0.553
ATOM C5 C6RE -0.293
ATOM C6 C6RE  0.257
ATOM H21 H   0.397
ATOM H24 HO  0.187
ATOM H25 H   0.447
BOND P OP1
BOND OP1 H21
BOND P OP2
```

```
BOND P OP3
BOND P OR5
BOND OR5 CR5
BOND CR5 CR4
BOND CR4 CR3
BOND CR3 OR3
BOND OR3 H24
BOND CR3 CR2
BOND CR2 CR1
BOND CR1 OR1
BOND OR1 CR4
BOND CR1 N1
BOND N1 C2
BOND C2 O2
BOND C2 N3
BOND N3 C4
BOND C4 O4
BOND C4 C5
BOND C5 C6
BOND C6 N1
BOND N3 H25
IMPH C6 N1 C2 N3
IMPH N1 C2 N3 C4
IMPH C2 N3 C4 C5
IMPH N3 C4 C5 C6
IMPH C4 C5 C6 N1
IMPH C5 C6 N1 C2
IMPH CR4 CR5 CR3 OR1
IMPH CR3 CR4 OR3 CR2
IMPH CR1 CR2 OR1 N1
IMPH N1 C2 C6 CR1
IMPH C2 N3 N1 O2
IMPH N3 C2 C4 H25
IMPH C4 C5 N3 O4
DONO H21 OP1
DONO H24 OR3
DONO H25 N3
ACCE OP1
ACCE OP2 P
ACCE OP3 P
ACCE OR5
ACCE OR1
ACCE OR3
ACCE N1
ACCE O2 C2
ACCE N3
ACCE O4 C4
IC OP2  P   OP1 H21    1.69 108.84 -84.98 109.47   1.09
IC OP3  P   OP1 H21    1.65 107.56 154.72 109.47   1.09
IC OR5  P   OP1 H21    1.61 106.58  25.39 109.47   1.09
IC OP1  P   OR5 CR5    1.51 106.58  78.38 125.71   1.41
IC OP2  P   OR5 CR5    1.69 102.92 -167.17 125.71   1.41
IC OP3  P   OR5 CR5    1.65 119.55 -43.67 125.71   1.41
IC P OR5   CR5 CR4    1.61 125.71 102.01 134.13   1.30
IC OR5 CR5   CR4 CR3    1.41 134.13 148.15 112.42   1.50
IC OR5 CR5   CR4 OR1    1.41 134.13   2.99 131.90   1.52
IC CR5 CR3  *CR4 OR1    1.30 112.42 153.43 108.06   1.52
IC CR5 CR4   CR3 OR3    1.30 112.42 109.52 110.86   0.86
```

```
IC OR1 CR4  CR3 OR3    1.52 108.06 -97.06 110.86   0.86
IC CR5 CR4  CR3 CR2    1.30 112.42 -132.26 105.06   1.56
IC CR4 OR3 *CR3 CR2    1.50 110.86 -115.48 109.53   1.56
IC CR4 CR3  OR3 H24    1.50 110.86 157.43 109.47   1.09
IC CR2 CR3  OR3 H24    1.56 109.53  41.95 109.47   1.09
IC CR4 CR3  CR2 CR1    1.50 105.06 -29.99 103.88   1.45
IC OR3 CR3  CR2 CR1    0.86 109.53  89.13 103.88   1.45
IC CR3 CR2  CR1 OR1    1.56 103.88  27.85 109.82   1.54
IC CR3 CR2  CR1 N1    1.56 103.88 177.44 114.65   1.55
IC CR2 OR1 *CR1 N1    1.45 109.82 -145.04 126.58   1.55
IC CR2 CR1  OR1 CR4    1.45 109.82 -15.08 104.13   1.52
IC N1 CR1  OR1 CR4    1.55 126.58 -160.12 104.13   1.52
IC CR1 OR1  CR4 CR5    1.54 104.13 141.46 131.90   1.30
IC CR1 OR1  CR4 CR3    1.54 104.13  -4.79 108.06   1.50
IC CR2 CR1  N1 C2    1.45 114.65  68.26 121.20   1.53
IC OR1 CR1  N1 C2    1.54 126.58 -148.12 121.20   1.53
IC CR2 CR1  N1 C6    1.45 114.65 -111.69 120.72   1.39
IC CR1 C2 *N1 C6    1.55 121.20 179.96 118.08   1.39
IC CR1 N1  C2 O2    1.55 121.20  -3.46 132.99   1.11
IC C6 N1  C2 O2    1.39 118.08 176.50 132.99   1.11
IC CR1 N1  C2 N3    1.55 121.20 175.44 109.39   1.44
IC N1 O2 *C2 N3    1.53 132.99 -178.83 117.61   1.44
IC N1 C2  N3 C4    1.53 109.39   4.51 130.33   1.45
IC O2 C2  N3 C4    1.11 117.61 -176.39 130.33   1.45
IC N1 C2  N3 H25    1.53 109.39 -175.47 114.84   1.09
IC C2 C4 *N3 H25    1.44 130.33 179.98 114.83   1.09
IC C2 N3  C4 O4    1.44 130.33 177.71 125.69   1.27
IC H25 N3  C4 O4    1.09 114.83  -2.31 125.69   1.27
IC C2 N3  C4 C5    1.44 130.33  -2.01 117.07   1.47
IC N3 O4 *C4 C5    1.45 125.69 179.72 117.23   1.47
IC N3 C4  C5 C6    1.45 117.07  -0.57 112.55   1.38
IC O4 C4  C5 C6    1.27 117.23 179.69 112.55   1.38
IC C4 C5  C6 N1    1.47 112.55  -0.19 132.40   1.39
IC C5 C6  N1 CR1    1.38 132.40 -176.95 120.72   1.55
IC C5 C6  N1 C2    1.38 132.40   3.09 118.08   1.53

END
```

```
* ...
* Proprietory and Confidential Property of Polygen Corporation
* ....
* TEMPLATE FOR CREATING CHARMm RTF file (from ChemNote)
* ....
*      (%_m) replaced by first 4 characters of molecule name
*      (%_r) replaced by RTF format of residue definition
* ...
* Build RTF for CB3
* ...
*  residue definition in RTF format built inline
*
   21    3
MASS     1 H       1.00800 ! Hydrogen bonding hydrogen (neutral group)
MASS     2 HC      1.00800 ! Hydrogen bonding hydrogen (charged group)
MASS     3 HA      1.00800 ! Aliphatic or aromatic hydrogen
MASS     4 HT      1.00800 ! TIPS3P water model hydrogen
MASS     5 LP      0.0     ! ST2 lone pair
MASS     6 MBE     9.01218 ! Beryllium
MASS     7 B      10.81    ! Boron
MASS     8 HO      1.00800 ! Hydrogen on an alcohol oxygen
MASS     9 HMU     1.00800 ! Mu-bonded hydrogen for metals and boron-hydride
MASS    10 CT     12.01100 ! Aliphatic carbon (tetrahedral)
MASS    11 CH1E   13.01900 ! Extended atom carbon with one hydrogen
MASS    12 CH2E   14.02700 ! Extended atom carbon with two hydrogens
MASS    13 CH3E   15.03500 ! Extended atom carbon with three hydrogens
MASS    14 C      12.01100 ! Carbonyl or Guanidinium carbon
MASS    15 CM     12.01100 ! Carbonmonoxide, or other triply bonded, carbon
MASS    16 CUA1   12.01100 ! Carbon in double bond, first pair
MASS    17 CUA2   12.01100 ! Carbon in double bond, second conjugated pair
MASS    18 CUY1   12.01100 ! Carbon in triple bond, first pair
MASS    19 CUY2   12.01100 ! Carbon in triple bond, second conjugated pair
MASS    20 CUA3   12.01100 ! Carbon in double bond, third conjugated pair
MASS    21 C5R    12.01100 ! Aromatic carbon in a five membered ring
MASS    22 C6R    12.01100 ! Aromatic carbon in a six  membered ring
MASS    23 C5RE   13.01900 ! Extended aromatic carbon in five membered ring
MASS    24 C6RE   13.01900 ! Extended aromatic carbon in six membered ring
MASS    25 CR55   12.01100 ! Aromatic carbon-merged five membered rings
MASS    26 CR56   12.01100 ! Aromatic carbon-merged five/six membered rings
MASS    27 CR66   12.01100 ! Aromatic carbon-merged six membered rings
MASS    28 C5RP   12.01100 ! for Aryl-Aryl bond between C5R rings
MASS    29 C6RP   12.01100 ! for Aryl-Aryl bond between C6R rings
MASS    30 N5RP   14.00670 ! for Ar-Ar bond between five membered rings
MASS    31 N      14.00670 ! Nitrogen; planar-valence of 3, i.e. nitrile, et
MASS    32 NP     14.00670 ! Nitrogen in peptide, amide, or related, group
MASS    33 NX     14.00670 ! Proline nitrogen, or similar
MASS    34 N5R    14.00670 ! Nitrogen in a five membered aromatic ring
MASS    35 N6R    14.00670 ! Nitrogen in a six membered aromatic ring
MASS    36 NT     14.00670 ! Nitrogen (tetrahedral), i.e. Amine, etc.
MASS    37 NC     14.00670 ! Charged guanidinuim-type nitrogen
MASS    38 NO2    14.00670 ! Nitrogen in nitro, or related, group
MASS    39 N6RP   14.00670 ! for Aryl-Aryl bond between six membered rings
MASS    40 O      15.99940 ! Carbonyl oxygen for amides, or related structur
MASS    41 OA     15.99940 ! Carbonyl oxygen for aldehydes, or related struc
MASS    42 OK     15.99940 ! Carbonyl oxygen for ketones, or related structu
MASS    43 OC     15.99940 ! Charged oxygen
MASS    45 OT     15.99940 ! Hydroxyl oxygen (tetrahedral)/Ionizable acid ox
MASS    46 OW     15.99940 ! TIP3P water model oxygen
```

```
MASS     47 OH2    15.99940 ! ST2  water model oxygen
MASS     48 OM     15.99940 ! Carbonmonoxide, or other triply bonded, oxygen
MASS     49 OS     15.99940 ! Ester oxygen
MASS     50 OE     15.99940 ! Ether oxygen / Acetal oxygen
MASS     51 OAC    15.99940 ! Carbonyl oxygen for acids
MASS     52 O5R    15.99940 ! Oxygen in five membered aromatic ring-radicals,
MASS     53 O6R    15.99940 ! Oxygen in aromatic ring - radicals etc.
MASS     55 OSI    15.99940 ! Oxygen in Si-O-Si bond
MASS     56 O2M    15.99940 ! Oxygen in Si-O-Al or Al-O-Al bond
MASS     57 OSH     0.00000 ! Massless O for zeolites, or related cage compou
MASS     60 PT     30.9738  ! Phosphorous, general; usually tetrahedral
MASS     61 PO3    30.9738  ! Phosphorous bonded to three oxygens
MASS     62 PO4    30.9738  ! Phosphorous bonded to four oxygens
MASS     63 PUA1   30.9738  ! Phosphorous double bond
MASS     64 P6R    30.9738  ! Phosphorous in aromatic 6-membered ring
MASS     70 ST     32.06000 ! Sulphur, general; usually tetrahedral
MASS     71 SH1E   33.06800 ! Extended atom sulphur with one hydrogen
MASS     72 S5R    32.0600  ! Sulphur in a five membered aromatic ring
MASS     73 S6R    32.0600  ! Sulphur in a six membered aromatic ring
MASS     74 SE     32.060   ! Thioether sulphur
MASS     75 SK     32.060   ! Thioketone sulphur
MASS     76 SO1    32.0600  ! Sulphur bonded to one oxygen
MASS     77 SO2    32.0600  ! Sulphur bonded to two oxygens
MASS     78 SO3    32.0600  ! Sulphur bonded to three oxygens
MASS     79 SO4    32.0600  ! Sulphur bonded to four oxygens
MASS     80 MLI     6.941   ! Lithium
MASS     81 MNA    22.9898  ! Sodium
MASS     82 MMG    24.305   ! Magnesiun
MASS     83 MK     39.098   ! Potassium
MASS     84 MCA    40.080   ! Calcium
MASS     85 MMN    54.938   ! Manganese
MASS     86 MFE    55.847   ! Iron
MASS     87 MZN    65.38    ! Zinc
MASS     88 MRB    85.4678  ! Rubidium
MASS     89 MCS   132.9054  ! Cesium
MASS     90 MSI    28.0855  ! Silicon
MASS     91 MAL    26.9815  ! ALuminum
MASS     92 XF     18.99840 ! Fluorine
MASS     93 XCL    35.45300 ! Chlorine
MASS     94 XBR    79.904   ! Bromine
MASS     95 XI    126.9045  ! Iodine
MASS     96 MCU    63.546   ! Copper
MASS     97 MV     50.9415  ! Vanadium
MASS     98 MCR    51.996   ! Chromium
MASS     99 MCO    58.9332  ! Cobalt
MASS    100 MNI    58.69    ! Nickel
MASS    101 MAS    74.9216  ! Arsenic
MASS    102 MSE    78.96    ! Selenium
MASS    103 MSR    87.62    ! Strontium
MASS    104 MY     88.9059  ! Yttrium
MASS    105 MZR    91.22    ! Zirconium
MASS    106 MNB    92.9064  ! Niobium
MASS    107 MMO    95.94    ! Molybdenum
MASS    108 MRU   101.07    ! Ruthenium
MASS    109 MRH   102.9055  ! Rhodium
MASS    110 MPD   106.42    ! Palladium
MASS    111 MAG   107.868   ! Silver
MASS    112 MCD   112.41    ! Cadmium
```

```
MASS     113 MSN   118.69     ! Tin
MASS     114 MSB   121.75     ! Antimony
MASS     115 MBA   137.33     ! Barium
MASS     116 MW    183.85     ! Tungsten
MASS     117 MOS   190.2      ! Osmium
MASS     118 MPT   195.08     ! Platinum
MASS     119 MAU   196.08     ! Gold
MASS     120 MHG   200.59     ! Mercury
MASS     121 MPB   207.2      ! Lead
MASS     122 MBI   208.9804   ! Bismuth
MASS     123 MLA   138.9055   ! Lanthanum
MASS     124 MCE   140.12     ! Cerium
MASS     125 MPR   140.9077   ! Praseodymium
MASS     126 MAC   227.0278   ! Actinium
MASS     127 MTH   232.0381   ! Thorium
MASS     128 MU    238.0289   ! Uranium
MASS     129 MTE   127.60     ! Tellurium
MASS     130 MPO   209.0      ! Polonium
MASS     131 XAT   210.0      ! Astatine
MASS     132 MTC    98.0      ! Technetium
MASS     133 MSC    44.9559   ! Scandium
MASS     134 MTI    47.88     ! Titanium
MASS     135 MGA    69.72     ! Gallium
MASS     136 MGE    72.59     ! Germanium
MASS     137 MIN   114.82     ! Indium
MASS     138 MHF   178.49     ! Hafnium
MASS     139 MTA   180.9479   ! Tantalum
MASS     140 MIR   192.22     ! Iridium
MASS     141 MTL   204.383    ! Thallium
MASS     142 MFR   223.0      ! Francium
MASS     143 MRA   226.0254   ! Radium
MASS     144 MND   144.24     ! Neodymium
MASS     145 MPM   145.0      ! Promethium
MASS     146 MSM   150.36     ! Samarium
MASS     147 MEU   151.96     ! Europium
MASS     148 MGD   157.25     ! Gadolinium
MASS     149 MTB   158.9254   ! Terbium
MASS     150 MDY   162.50     ! Dysprosium
MASS     151 MHO   164.9304   ! Holmium
MASS     152 MER   167.26     ! Erbium
MASS     153 MTM   168.9342   ! Thulium
MASS     154 MYB   173.04     ! Ytterbium
MASS     155 MLU   174.967    ! Lutetium
MASS     156 MPA   231.0359   ! Protactinium
MASS     157 MNP   237.0482   ! Neptunium
MASS     158 MPU   244.0      ! Plutonium
MASS     159 MAM   243.0      ! Americium
MASS     160 MCM   247.0      ! Curium
MASS     161 MBK   247.0      ! Berkelium
MASS     162 MCF   251.0      ! Californium
MASS     163 MES   252.0      ! Einsteinium
MASS     164 MFM   257.0      ! Fermium
MASS     165 MMD   258.0      ! Medelevium
MASS     166 MNO   259.0      ! Nobelium
MASS     167 MLR   260.0      ! Lawrencium
MASS     168 HE      4.00260  ! Helium
MASS     169 NE     20.179    ! Neon
MASS     170 AR     39.948    ! Argon
```

```
MASS    171 KR     83.80    ! Krypton
MASS    172 XE    131.29    ! Xenon
MASS    173 RN    222.0     ! Radon
MASS    174 MRF   261.0     ! Rutherfordium/Kurchatovium
MASS    175 MHA   262.0     ! Hahnium
MASS    176 MRE   186.31    ! Rhenium
MASS    177 MSIU  28.08550  ! Silicon when as double bond
MASS    180 NR56  14.00670  ! N at fused bond between 5 and 6-membered aromat
MASS    181 NR66  14.00670  ! N at fused bond between two 6-membered aromatic
MASS    182 NR55  14.00670  ! N at fused bond between two 5-membered aromatic
MASS    183 NR1   14.00670  ! Protonated nitrogen in neutral histidine ring
MASS    184 NR2   14.00670  ! Unprotonated nitrogen in neutral histidine ring
MASS    185 NR3   14.00670  ! Nitrogens in charged histidine ring
MASS    186 NC2   14.00670  ! for neutral guanidinium group - Arg sidechain
MASS    189 C5RQ  12.01100  ! for second Ar-Ar bond between C5RP rings (ortho
MASS    190 C3    12.01100  ! Carbonyl carbon in 3-membered aliphatic ring
MASS    191 CT3   12.01100  ! C in 3-membered aliphatic ring, usually tetrahe
MASS    192 C4    12.01100  ! Carbonyl carbons in 4-membered aliphatic ring
MASS    193 CT4   12.01100  ! C in 4-membered aliphatic ring, usually tetrahe
MASS    194 C6RQ  12.01100  ! Carbon of C6RP type ortho to C6RP pair
MASS    195 CQ66  12.01100  ! Third adjacent pair of CR66 types in fused ring
MASS    196 CS66  12.01100  ! Second adjacent pair of CR66 types in fused rin
MASS    197 CPH1  12.01100  ! CG and CD2 carbons in histidine ring
MASS    198 CPH2  12.01100  ! CE1 carbon in histidine ring
MASS    199 CP3   12.01100  ! Carbon on nitrogen in proline ring

AUTOGENERATE ANGLES DIHE
DEFA FIRS NONE LAST NONE

RESI CB3 -2.000
GROUP
ATOM N1 N6R -0.740
ATOM C2 C6R  0.900
ATOM NA2 NP -0.750
ATOM N3 N6R -0.560
ATOM C4 C  0.580
ATOM OA4 O -0.519
ATOM C4A CR66 -0.069
ATOM C5 C6RE  0.031
ATOM C6 C6R  0.031
ATOM C7 C6RE  0.031
ATOM C8 C6RE  0.031
ATOM C8A CR66 -0.069
ATOM C9 CH2E  0.031
ATOM N10 NP -0.369
ATOM CP1 CH2E  0.031
ATOM CP2 CUY1 -0.269
ATOM CP3 CUY1 -0.259
ATOM C11 C6R  0.231
ATOM C12 C6RE  0.031
ATOM C13 C6RE  0.031
ATOM C14 C6R  0.131
ATOM C15 C6RE  0.031
ATOM C16 C6RE  0.031
ATOM C C  0.580
ATOM O O -0.519
ATOM N NP -0.319
```

```
ATOM CA CH1E  0.131
ATOM CT C  0.171
ATOM O1 OC -0.540
ATOM O2 OC -0.540
ATOM CB CH2E  0.031
ATOM CG CH2E  0.031
ATOM CD C  0.171
ATOM OE1 OC -0.539
ATOM OE2 OC -0.540
ATOM H36 H  0.281
ATOM H37 H  0.331
ATOM H38 H  0.361
ATOM H39 H  0.361
BOND N1 C2
BOND C2 NA2
BOND NA2 H38
BOND NA2 H39
BOND C2 N3
BOND N3 C4
BOND C4 OA4
BOND C4 C4A    ! parameter for bond distance not found
BOND C4A C5
BOND C5 C6
BOND C6 C7
BOND C7 C8
BOND C8 C8A
BOND C8A N1
BOND C8A C4A
BOND C6 C9
BOND C9 N10
BOND N10 CP1
BOND CP1 CP2    ! parameter for bond distance not found
BOND CP2 CP3
BOND N10 C11
BOND C11 C12
BOND C12 C13
BOND C13 C14
BOND C14 C15
BOND C15 C16
BOND C16 C11
BOND C14 C
BOND C O
BOND C N
BOND N CA
BOND CA CT
BOND CT O1
BOND CT O2
BOND CA CB
BOND CB CG
BOND CG CD
BOND CD OE1
BOND CD OE2
BOND N H36
BOND N3 H37
IMPH C8A N1 C2 N3
IMPH N1 C2 N3 C4
IMPH C2 N3 C4 C4A
IMPH N3 C4 C4A C8A
```

```
IMPH C8A C4A C5 C6
IMPH C4A C5 C6 C7
IMPH C5 C6 C7 C8
IMPH C6 C7 C8 C8A
IMPH C7 C8 C8A C4A
IMPH C4A C8A N1 C2
IMPH N1 C8A C4A C4
IMPH C16 C11 C12 C13
IMPH C11 C12 C13 C14
IMPH C12 C13 C14 C15
IMPH C13 C14 C15 C16
IMPH C14 C15 C16 C11
IMPH C15 C16 C11 C12
IMPH C2 N3 N1 NA2
IMPH NA2 H39 C2 H38
IMPH N3 C2 C4 H37
IMPH C4 C4A N3 OA4
IMPH C4A C5 C8A C4
IMPH C6 C5 C7 C9
IMPH C8A C4A C8 N1
IMPH N10 CP1 C11 C9
IMPH C11 C12 C16 N10
IMPH C14 C13 C15 C
IMPH C N C14 O
IMPH N C CA H36
IMPH CA N CT CB
IMPH CT O1 O2 CA
IMPH CD OE1 OE2 CG
DONO H38 NA2
DONO H39 NA2
DONO H37 N3
DONO H36 N
ACCE N1
ACCE NA2
ACCE N3
ACCE OA4 C4
ACCE N10
ACCE O C
ACCE N
ACCE O1 CT
ACCE O2 CT
ACCE OE1 CD
ACCE OE2 CD
IC C8A N1  C2 NA2    1.44 123.91 -179.54 114.88   1.47
IC C8A N1  C2 N3    1.44 123.91  -1.60 119.99   1.47
IC N1 NA2 *C2 N3    1.15 114.88 -177.82 125.09   1.47
IC N1 C2  NA2 H39    1.15 114.88 -77.32 119.96   1.09
IC N3 C2  NA2 H39    1.47 125.09 104.85 119.96   1.09
IC N1 C2  NA2 H38    1.15 114.88 102.61 120.00   1.09
IC C2 H38 *NA2 H39    1.47 120.00 179.93 120.04   1.09
IC N1 C2  N3 C4    1.15 119.99   1.57 117.56   1.23
IC NA2 C2  N3 C4    1.47 125.09 179.28 117.56   1.23
IC N1 C2  N3 H37    1.15 119.99 -178.40 121.21   1.09
IC C2 C4 *N3 H37    1.47 117.56 179.97 121.23   1.09
IC C2 N3  C4 OA4    1.47 117.56 -178.93 117.03   1.22
IC H37 N3  C4 OA4    1.09 121.23   1.03 117.03   1.22
IC C2 N3  C4 C4A    1.47 117.56   0.95 125.09   1.31
IC N3 OA4 *C4 C4A    1.23 117.03 -179.89 117.89   1.31
```

```
IC N3 C4  C4A C5   1.23 125.09 177.92 120.53   1.35
IC OA4 C4  C4A C5   1.22 117.89  -2.19 120.53   1.35
IC N3 C4  C4A C8A   1.23 125.09  -3.20 118.76   1.38
IC C4 C5 *C4A C8A   1.31 120.53 -178.85 120.70   1.38
IC C4 C4A  C5 C6   1.31 120.53 -174.12 116.95   1.36
IC C8A C4A  C5 C6   1.38 120.70   7.02 116.95   1.36
IC C4A C5  C6 C7   1.35 116.95  -8.52 121.61   1.42
IC C4A C5  C6 C9   1.35 116.95 -178.42 126.74   1.48
IC C5 C7 *C6 C9   1.36 121.61 171.35 111.00   1.48
IC C5 C6  C7 C8   1.36 121.61   4.91 119.39   1.30
IC C9 C6  C7 C8   1.48 111.00 176.26 119.39   1.30
IC C6 C7  C8 C8A   1.42 119.39   0.47 119.97   1.36
IC C7 C8  C8A N1   1.30 119.97 173.89 124.40   1.44
IC C7 C8  C8A C4A   1.30 119.97  -1.90 120.88   1.38
IC N1 C4A *C8A C8   1.44 114.60 176.18 120.88   1.36
IC C4A C8A  N1 C2   1.38 114.60  -0.60 123.91   1.15
IC C8 C8A  N1 C2   1.36 124.40 -176.63 123.91   1.15
IC N1 C8A  C4A C4   1.44 114.60   2.94 118.76   1.31
IC C8 C8A  C4A C4   1.36 120.88 179.12 118.76   1.31
IC N1 C8A  C4A C5   1.44 114.60 -178.18 120.70   1.35
IC C5 C6  C9 N10   1.36 126.74 135.23 105.09   1.58
IC C7 C6  C9 N10   1.42 111.00 -35.56 105.09   1.58
IC C6 C9  N10 CP1   1.48 105.09  -8.64 116.04   1.50
IC C6 C9  N10 C11   1.48 105.09 132.99 109.46   1.46
IC C9 CP1 *N10 C11   1.58 116.04 -136.83 121.20   1.46
IC C9 N10  CP1 CP2   1.58 116.04 -74.58 114.97   1.55
IC C11 N10  CP1 CP2   1.46 121.20 148.59 114.97   1.55
IC N10 CP1  CP2 CP3   1.50 114.97  -5.28 176.11   1.30
IC C9 N10  C11 C12   1.58 109.46 132.79 124.69   1.37
IC CP1 N10  C11 C12   1.50 121.20 -87.90 124.69   1.37
IC C9 N10  C11 C16   1.58 109.46 -44.23 113.13   1.39
IC N10 C12 *C11 C16   1.46 124.69 176.77 122.11   1.39
IC N10 C11  C12 C13   1.46 124.69 -178.01 121.08   1.42
IC C16 C11  C12 C13   1.39 122.11  -1.24 121.08   1.42
IC C11 C12  C13 C14   1.37 121.08   0.62 116.84   1.34
IC C12 C13  C14 C15   1.42 116.84   0.19 123.59   1.40
IC C12 C13  C14 C   1.42 116.84 179.52 122.88   1.56
IC C13 C15 *C14 C   1.34 123.59 -179.39 113.53   1.56
IC C13 C14  C15 C16   1.34 123.59  -0.40 119.43   1.41
IC C C14  C15 C16   1.56 113.53 -179.78 119.43   1.41
IC C14 C15  C16 C11   1.40 119.43  -0.19 116.95   1.39
IC C15 C16  C11 N10   1.41 116.95 178.10 113.13   1.46
IC C15 C16  C11 C12   1.41 116.95   0.99 122.11   1.37
IC C13 C14  C O   1.34 122.88  48.86 120.34   1.22
IC C15 C14  C O   1.40 113.53 -131.75 120.34   1.22
IC C13 C14  C N   1.34 122.88 -168.32 111.81   1.34
IC C14 O *C N   1.56 120.34 -141.06 116.79   1.34
IC C14 C  N CA   1.56 111.81 -144.28 138.04   1.45
IC O C  N CA   1.22 116.79  -0.04 138.04   1.45
IC C14 C  N H36   1.56 111.81  35.83 110.97   1.09
IC C CA *N H36   1.34 138.04 179.88 110.98   1.09
IC C N  CA CT   1.34 138.04 166.65 108.30   1.57
IC H36 N  CA CT   1.09 110.98 -13.46 108.30   1.57
IC C N  CA CB   1.34 138.04  42.95 113.02   1.56
IC N CT *CA CB   1.45 108.30 124.77 111.23   1.56
IC N CA  CT O1   1.45 108.30 -11.13 128.27   1.32
IC CB CA  CT O1   1.56 111.23 113.64 128.27   1.32
IC N CA  CT O2   1.45 108.30 115.83 118.88   1.26
```

```
IC CA O1 *CT O2    1.57 128.27 -135.23  96.56    1.26
IC N CA  CB CG    1.45 113.02  15.12 111.03    1.52
IC CT CA  CB CG    1.57 111.23 -106.95 111.03    1.52
IC CA CB  CG CD    1.56 111.03 153.31 105.72    1.57
IC CB CG  CD OE1    1.52 105.72 -79.91 116.38    1.41
IC CB CG  CD OE2    1.52 105.72  53.53 118.09    1.27
IC CG OE1 *CD OE2    1.57 116.38 -137.19 109.52    1.27

END
```

Appendix 2.3

patch data and awk files

```
* ...
* Proprietory and Confidential Property of Polygen Corporation
* ....
* TEMPLATE FOR CREATING CHARMm RTF file (from ChemNote)
* ....
*      (%_m) replaced by first 4 characters of molecule name
*      (%_r) replaced by RTF format of residue definition
* ...
* Build RTF for CB3
* ...
*  residue definition in RTF format built inline
*
   21    3
MASS     1 H      1.00800 ! Hydrogen bonding hydrogen (neutral group)
MASS     2 HC     1.00800 ! Hydrogen bonding hydrogen (charged group)
MASS     3 HA     1.00800 ! Aliphatic or aromatic hydrogen
MASS     4 HT     1.00800 ! TIPS3P water model hydrogen
MASS     5 LP     0.0     ! ST2 lone pair
MASS     6 MBE    9.01218 ! Beryllium
MASS     7 B     10.81    ! Boron
MASS     8 HO    1.00800  ! Hydrogen on an alcohol oxygen
MASS     9 HMU   1.00800  ! Mu-bonded hydrogen for metals and boron-hydride
MASS    10 CT    12.01100 ! Aliphatic carbon (tetrahedral)
MASS    11 CH1E  13.01900 ! Extended atom carbon with one hydrogen
MASS    12 CH2E  14.02700 ! Extended atom carbon with two hydrogens
MASS    13 CH3E  15.03500 ! Extended atom carbon with three hydrogens
MASS    14 C     12.01100 ! Carbonyl or Guanidinium carbon
MASS    15 CM    12.01100 ! Carbonmonoxide, or other triply bonded, carbon
MASS    16 CUA1  12.01100 ! Carbon in double bond, first pair
MASS    17 CUA2  12.01100 ! Carbon in double bond, second conjugated pair
MASS    18 CUY1  12.01100 ! Carbon in triple bond, first pair
MASS    19 CUY2  12.01100 ! Carbon in triple bond, second conjugated pair
MASS    20 CUA3  12.01100 ! Carbon in double bond, third conjugated pair
MASS    21 C5R   12.01100 ! Aromatic carbon in a five membered ring
MASS    22 C6R   12.01100 ! Aromatic carbon in a six  membered ring
MASS    23 C5RE  13.01900 ! Extended aromatic carbon in five membered ring
MASS    24 C6RE  13.01900 ! Extended aromatic carbon in six membered ring
MASS    25 CR55  12.01100 ! Aromatic carbon-merged five membered rings
MASS    26 CR56  12.01100 ! Aromatic carbon-merged five/six membered rings
MASS    27 CR66  12.01100 ! Aromatic carbon-merged six membered rings
MASS    28 C5RP  12.01100 ! for Aryl-Aryl bond between C5R rings
MASS    29 C6RP  12.01100 ! for Aryl-Aryl bond between C6R rings
MASS    30 N5RP  14.00670 ! for Ar-Ar bond between five membered rings
MASS    31 N     14.00670 ! Nitrogen; planar-valence of 3, i.e. nitrile, et
MASS    32 NP    14.00670 ! Nitrogen in peptide, amide, or related, group
MASS    33 NX    14.00670 ! Proline nitrogen, or similar
MASS    34 N5R   14.00670 ! Nitrogen in a five membered aromatic ring
MASS    35 N6R   14.00670 ! Nitrogen in a six membered aromatic ring
MASS    36 NT    14.00670 ! Nitrogen (tetrahedral), i.e. Amine, etc.
MASS    37 NC    14.00670 ! Charged guanidinuim-type nitrogen
MASS    38 NO2   14.00670 ! Nitrogen in nitro, or related, group
MASS    39 N6RP  14.00670 ! for Aryl-Aryl bond between six membered rings
MASS    40 O     15.99940 ! Carbonyl oxygen for amides, or related structur
MASS    41 OA    15.99940 ! Carbonyl oxygen for aldehydes, or related struc
MASS    42 OK    15.99940 ! Carbonyl oxygen for ketones, or related structu
MASS    43 OC    15.99940 ! Charged oxygen
MASS    45 OT    15.99940 ! Hydroxyl oxygen (tetrahedral)/Ionizable acid ox
MASS    46 OW    15.99940 ! TIP3P water model oxygen
```

```
MASS     47 OH2    15.99940 ! ST2  water model oxygen
MASS     48 OM     15.99940 ! Carbonmonoxide, or other triply bonded, oxygen
MASS     49 OS     15.99940 ! Ester oxygen
MASS     50 OE     15.99940 ! Ether oxygen / Acetal oxygen
MASS     51 OAC    15.99940 ! Carbonyl oxygen for acids
MASS     52 O5R    15.99940 ! Oxygen in five membered aromatic ring-radicals,
MASS     53 O6R    15.99940 ! Oxygen in aromatic ring - radicals etc.
MASS     55 OSI    15.99940 ! Oxygen in Si-O-Si bond
MASS     56 O2M    15.99940 ! Oxygen in Si-O-Al or Al-O-Al bond
MASS     57 OSH     0.00000 ! Massless O for zeolites, or related cage compou
MASS     60 PT     30.9738  ! Phosphorous, general; usually tetrahedral
MASS     61 PO3    30.9738  ! Phosphorous bonded to three oxygens
MASS     62 PO4    30.9738  ! Phosphorous bonded to four oxygens
MASS     63 PUA1   30.9738  ! Phosphorous double bond
MASS     64 P6R    30.9738  ! Phosphorous in aromatic 6-membered ring
MASS     70 ST     32.06000 ! Sulphur, general; usually tetrahedral
MASS     71 SH1E   33.06800 ! Extended atom sulphur with one hydrogen
MASS     72 S5R    32.0600  ! Sulphur in a five membered aromatic ring
MASS     73 S6R    32.0600  ! Sulphur in a six membered aromatic ring
MASS     74 SE     32.060   ! Thioether sulphur
MASS     75 SK     32.060   ! Thioketone sulphur
MASS     76 SO1    32.0600  ! Sulphur bonded to one oxygen
MASS     77 SO2    32.0600  ! Sulphur bonded to two oxygens
MASS     78 SO3    32.0600  ! Sulphur bonded to three oxygens
MASS     79 SO4    32.0600  ! Sulphur bonded to four oxygens
MASS     80 MLI     6.941   ! Lithium
MASS     81 MNA    22.9898  ! Sodium
MASS     82 MMG    24.305   ! Magnesiun
MASS     83 MK     39.098   ! Potassium
MASS     84 MCA    40.080   ! Calcium
MASS     85 MMN    54.938   ! Manganese
MASS     86 MFE    55.847   ! Iron
MASS     87 MZN    65.38    ! Zinc
MASS     88 MRB    85.4678  ! Rubidium
MASS     89 MCS   132.9054  ! Cesium
MASS     90 MSI    28.0855  ! Silicon
MASS     91 MAL    26.9815  ! ALuminum
MASS     92 XF     18.99840 ! Fluorine
MASS     93 XCL    35.45300 ! Chlorine
MASS     94 XBR    79.904   ! Bromine
MASS     95 XI    126.9045  ! Iodine
MASS     96 MCU    63.546   ! Copper
MASS     97 MV     50.9415  ! Vanadium
MASS     98 MCR    51.996   ! Chromium
MASS     99 MCO    58.9332  ! Cobalt
MASS    100 MNI    58.69    ! Nickel
MASS    101 MAS    74.9216  ! Arsenic
MASS    102 MSE    78.96    ! Selenium
MASS    103 MSR    87.62    ! Strontium
MASS    104 MY     88.9059  ! Yttrium
MASS    105 MZR    91.22    ! Zirconium
MASS    106 MNB    92.9064  ! Niobium
MASS    107 MMO    95.94    ! Molybdenum
MASS    108 MRU   101.07    ! Ruthenium
MASS    109 MRH   102.9055  ! Rhodium
MASS    110 MPD   106.42    ! Palladium
MASS    111 MAG   107.868   ! Silver
MASS    112 MCD   112.41    ! Cadmium
```

```
MASS     113 MSN   118.69     ! Tin
MASS     114 MSB   121.75     ! Antimony
MASS     115 MBA   137.33     ! Barium
MASS     116 MW    183.85     ! Tungsten
MASS     117 MOS   190.2      ! Osmium
MASS     118 MPT   195.08     ! Platinum
MASS     119 MAU   196.08     ! Gold
MASS     120 MHG   200.59     ! Mercury
MASS     121 MPB   207.2      ! Lead
MASS     122 MBI   208.9804   ! Bismuth
MASS     123 MLA   138.9055   ! Lanthanum
MASS     124 MCE   140.12     ! Cerium
MASS     125 MPR   140.9077   ! Praseodymium
MASS     126 MAC   227.0278   ! Actinium
MASS     127 MTH   232.0381   ! Thorium
MASS     128 MU    238.0289   ! Uranium
MASS     129 MTE   127.60     ! Tellurium
MASS     130 MPO   209.0      ! Polonium
MASS     131 XAT   210.0      ! Astatine
MASS     132 MTC    98.0      ! Technetium
MASS     133 MSC    44.9559   ! Scandium
MASS     134 MTI    47.88     ! Titanium
MASS     135 MGA    69.72     ! Gallium
MASS     136 MGE    72.59     ! Germanium
MASS     137 MIN   114.82     ! Indium
MASS     138 MHF   178.49     ! Hafnium
MASS     139 MTA   180.9479   ! Tantalum
MASS     140 MIR   192.22     ! Iridium
MASS     141 MTL   204.383    ! Thallium
MASS     142 MFR   223.0      ! Francium
MASS     143 MRA   226.0254   ! Radium
MASS     144 MND   144.24     ! Neodymium
MASS     145 MPM   145.0      ! Promethium
MASS     146 MSM   150.36     ! Samarium
MASS     147 MEU   151.96     ! Europium
MASS     148 MGD   157.25     ! Gadolinium
MASS     149 MTB   158.9254   ! Terbium
MASS     150 MDY   162.50     ! Dysprosium
MASS     151 MHO   164.9304   ! Holmium
MASS     152 MER   167.26     ! Erbium
MASS     153 MTM   168.9342   ! Thulium
MASS     154 MYB   173.04     ! Ytterbium
MASS     155 MLU   174.967    ! Lutetium
MASS     156 MPA   231.0359   ! Protactinium
MASS     157 MNP   237.0482   ! Neptunium
MASS     158 MPU   244.0      ! Plutonium
MASS     159 MAM   243.0      ! Americium
MASS     160 MCM   247.0      ! Curium
MASS     161 MBK   247.0      ! Berkelium
MASS     162 MCF   251.0      ! Californium
MASS     163 MES   252.0      ! Einsteinium
MASS     164 MFM   257.0      ! Fermium
MASS     165 MMD   258.0      ! Medelevium
MASS     166 MNO   259.0      ! Nobelium
MASS     167 MLR   260.0      ! Lawrencium
MASS     168 HE      4.00260  ! Helium
MASS     169 NE     20.179    ! Neon
MASS     170 AR     39.948    ! Argon
```

```
MASS   171 KR     83.80    ! Krypton
MASS   172 XE    131.29    ! Xenon
MASS   173 RN    222.0     ! Radon
MASS   174 MRF   261.0     ! Rutherfordium/Kurchatovium
MASS   175 MHA   262.0     ! Hahnium
MASS   176 MRE   186.31    ! Rhenium
MASS   177 MSIU  28.08550  ! Silicon when as double bond
MASS   180 NR56  14.00670  ! N at fused bond between 5 and 6-membered aromat
MASS   181 NR66  14.00670  ! N at fused bond between two 6-membered aromatic
MASS   182 NR55  14.00670  ! N at fused bond between two 5-membered aromatic
MASS   183 NR1   14.00670  ! Protonated nitrogen in neutral histidine ring
MASS   184 NR2   14.00670  ! Unprotonated nitrogen in neutral histidine ring
MASS   185 NR3   14.00670  ! Nitrogens in charged histidine ring
MASS   186 NC2   14.00670  ! for neutral guanidinium group - Arg sidechain
MASS   189 C5RQ   12.01100 ! for second Ar-Ar bond between C5RP rings (ortho
MASS   190 C3    12.01100  ! Carbonyl carbon in 3-membered aliphatic ring
MASS   191 CT3   12.01100  ! C in 3-membered aliphatic ring, usually tetrahe
MASS   192 C4    12.01100  ! Carbonyl carbons in 4-membered aliphatic ring
MASS   193 CT4   12.01100  ! C in 4-membered aliphatic ring, usually tetrahe
MASS   194 C6RQ  12.01100  ! Carbon of C6RP type ortho to C6RP pair
MASS   195 CQ66  12.01100  ! Third adjacent pair of CR66 types in fused ring
MASS   196 CS66  12.01100  ! Second adjacent pair of CR66 types in fused rin
MASS   197 CPH1  12.01100  ! CG and CD2 carbons in histidine ring
MASS   198 CPH2  12.01100  ! CE1 carbon in histidine ring
MASS   199 CP3   12.01100  ! Carbon on nitrogen in proline ring

AUTOGENERATE ANGLES DIHE
DEFA FIRS NONE LAST NONE

PRES MAJO       -0.33
! Patch for dump to CYS This may not actually work but who knows?
GROUP
ATOM 1SG  SE    -0.12
GROUP
ATOM 2N1  NP    -0.20
ATOM 2C6  CH1E  -0.10
ATOM 2C5  CH2E  +0.12
BOND 1SG  2C6
ANGL  2N1 2C6 1SG       2C6 1SG 1CB       2C5 2C6 1SG
DIHE  2CR1 2N1 2C6 1SG
DIHE  2C2 2N1 2C6 1SG
DIHE  2C4 2C5 2C6 1SG
DIHE  2N1 2C6 1SG 1CB
DIHE  2C5 2C6 1SG 1CB
DIHE  2C6 1SG 1CB 1CA
IC     2CR1 2N1 2C6 1SG        0.0   0.0  180.0  0.0  0.0
IC     2C2 2N1 2C6 1SG         0.0   0.0  180.0  0.0  0.0
IC     2C4 2C5 2C6 1SG         0.0   0.0  180.0  0.0  0.0
IC     2N1 2C6 1SG 1CB         0.0   0.0  180.0  0.0  0.0
IC     2C5 2C6 1SG 1CB         0.0   0.0  180.0  0.0  0.0
IC     2C6 1SG 1CB 1CA         0.0   0.0  180.0  0.0  0.0

END
```

```
#! /bin/sh
nawk '
{if ($4 ~ /^(N2A|H33|H34|H35|C22|N23|N21|C24|O4A|CA4|CA8|C28|C25|C26|C27|C2
{
        {if (NF == 12)
        {sum += $11; sum2 += $12}
}
        {if (NF == 11)
        {sum += $10; sum2 += $11}
}
}
}
{if ($4 ~ /^(NA2|H39|H38|H37|C2|N3|N1|C4|OA4|C4A|C8A|C8|C5|C6|C7|C9)$/)
{
        {if (NF == 12)
        {sum5 += $11; sum6 += $12}
}
        {if (NF == 11)
        {sum5 += $10; sum6 += $11}
}
}
}
{if ($4 !~ /^(N2A|H33|H34|H35|C22|N23|N21|C24|O4A|CA4|CA8|C28|C25|C26|C27|C
{
        {if (NF == 12)
        {sum3 += $11; sum4 += $12}
}
        {if (NF == 11)
        {sum3 += $10; sum4 += $11}
}
}
{if ($4 ~ /^(O1|O2|CT)$/)
{
        {if (NF == 12)
        {sum7 += $11; sum8 += $12}
}
        {if (NF == 11)
        {sum7 += $10; sum8 += $11}
}
}
}
{if ($4 ~ /^(O1)$/)
{
        {if (NF == 12)
        {sum15 += $11; sum16 += $12}
}
        {if (NF == 11)
        {sum15 += $10; sum16 += $11}
}
}
}
{if ($4 ~ /^(O2)$/)
{
        {if (NF == 12)
        {sum17 += $11; sum18 += $12}
}
        {if (NF == 11)
```

```
        {sum17 += $10; sum18 += $11}
}
}
}
{if ($4 ~ /^(OE1)$/)
{
        {if (NF == 12)
        {sum19 += $11; sum20 += $12}
}
        {if (NF == 11)
        {sum19 += $10; sum20 += $11}
}
}
}
{if ($4 ~ /^(OE2)$/)
{
        {if (NF == 12)
        {sum21 += $11; sum22 += $12}
}
        {if (NF == 11)
        {sum21 += $10; sum22 += $11}
}
}
}
{if ($4 ~ /^(OE1|OE2|CD|CG|CB)$/)
{
        {if (NF == 12)
        {sum9 += $11; sum10 += $12}
}
        {if (NF == 11)
        {sum9 += $10; sum10 += $11}
}
}
}
{if ($4 ~ /^(C11|C12|C13|C14|C15|C16)$/)
{
        {if (NF == 12)
        {sum11 += $11; sum12 += $12}
}
        {if (NF == 11)
        {sum11 += $10; sum12 += $11}
}
}
}
{if ($4 ~ /^(C|O|N|H)$/)
{
        {if (NF == 12)
        {sum13 += $11; sum14 += $12}
}
        {if (NF == 11)
        {sum13 += $10; sum14 += $11}
}
}
}
END     {print "FILE = '$1'";
        print "TOTAL VDW of bif = ", sum;
        print "TOTAL ELEC of bif = ", sum2;
        print "TOTAL VDW of quin = ", sum5;
```

```
print "TOTAL ELEC of quin = ", sum6;
print "TOTAL nonbonded of both sides = ", sum5 + sum6 + sum + sum2;
print "TOTAL VDW of main = ", sum3;
print "TOTAL ELEC of main = ", sum4;
print "TOTAL nonbonded of main = ", sum3 + sum4;
print "TOTAL VDW = ", sum + sum3;
print "TOTAL ELEC = ", sum2 + sum4;
print "TOTAL VDW of short acid = ", sum7;
print "TOTAL ELEC of short acid = ", sum8;
print "TOTAL VDW of O1 = ", sum15;
print "TOTAL ELEC of O1 = ", sum16;
print "TOTAL VDW of O2 = ", sum17;
print "TOTAL ELEC of O2 = ", sum18;
        print "TOTAL VDW of long acid = ", sum9;
        print "TOTAL ELEC of long acid = ", sum10;
        print "TOTAL VDW of OE1 = ", sum19;
        print "TOTAL ELEC of OE1 = ", sum20;
        print "TOTAL VDW of OE2 = ", sum21;
        print "TOTAL ELEC of OE2 = ", sum22;
        print "TOTAL VDW of ring = ", sum11;
        print "TOTAL ELEC of ring = ", sum12;
        print "TOTAL VDW of amide = ", sum13;
        print "TOTAL ELEC of amide = ", sum14;
        print "TOTAL nonbonded = ", sum2 + sum4 + sum + sum3}' $1
nawk '
{if ($4 ~ /^(NA2|H39|H38|H37|C2|N3|N1|C4|OA4|C4A|C8A|C8|C5|C6|C7|C9)$/)
{
        {if (NF == 12)
        {sum += $11; sum2 += $12}
}
        {if (NF == 11)
        {sum += $10; sum2 += $11}
}
}
}
{if ($4 ~ /^(N2A|H33|H34|H35|C22|N23|N21|C24|O4A|CA4|CA8|C28|C25|C26|C27|C2
{
        {if (NF == 12)
        {sum5 += $11; sum6 += $12}
}
        {if (NF == 11)
        {sum5 += $10; sum6 += $11}
}
}
}
{if ($4 !~ /^(NA2|H39|H38|H37|C2|N3|N1|C4|OA4|C4A|C8A|C8|C5|C6|C7|C9)$/)
{
        {if (NF == 12)
        {sum3 += $11; sum4 += $12}
}
        {if (NF == 11)
        {sum3 += $10; sum4 += $11}
}
}
}
{if ($4 ~ /^(O1|O2|CT)$/)
{
        {if (NF == 12)
```

```
        {sum7 += $11; sum8 += $12}
}
        {if (NF == 11)
        {sum7 += $10; sum8 += $11}
}
}
}
{if ($4 ~ /^(O1)$/)
{
        {if (NF == 12)
        {sum15 += $11; sum16 += $12}
}
        {if (NF == 11)
        {sum15 += $10; sum16 += $11}
}
}
}
{if ($4 ~ /^(O2)$/)
{
        {if (NF == 12)
        {sum17 += $11; sum18 += $12}
}
        {if (NF == 11)
        {sum17 += $10; sum18 += $11}
}
}
}
{if ($4 ~ /^(OE1)$/)
{
        {if (NF == 12)
        {sum19 += $11; sum20 += $12}
}
        {if (NF == 11)
        {sum19 += $10; sum20 += $11}
}
}
}
{if ($4 ~ /^(OE2)$/)
{
        {if (NF == 12)
        {sum21 += $11; sum22 += $12}
}
        {if (NF == 11)
        {sum21 += $10; sum22 += $11}
}
}
}
{if ($4 ~ /^(OE1|OE2|CD|CG|CB)$/)
{
        {if (NF == 12)
        {sum9 += $11; sum10 += $12}
}
        {if (NF == 11)
        {sum9 += $10; sum10 += $11}
}
}
}
{if ($4 ~ /^(C11|C12|C13|C14|C15|C16)$/)
```

```
{
        {if (NF == 12)
        {sum11 += $11; sum12 += $12}
}
        {if (NF == 11)
        {sum11 += $10; sum12 += $11}
}
}
}
{if ($4 ~ /^(C|O|N|H)$/)
{
        {if (NF == 12)
        {sum13 += $11; sum14 += $12}
}
        {if (NF == 11)
        {sum13 += $10; sum14 += $11}
}
}
}
END     {print "FILE = '$2'";
        print "TOTAL VDW of bif = ", sum;
        print "TOTAL ELEC of bif = ", sum2;
        print "TOTAL VDW of quin = ", sum5;
        print "TOTAL ELEC of quin = ", sum6;
        print "TOTAL nonbonded of both sides = ", sum5 + sum6 + sum + sum2;
        print "TOTAL VDW of main = ", sum3;
        print "TOTAL ELEC of main = ", sum4;
        print "TOTAL nonbonded of main = ", sum3 + sum4;
        print "TOTAL VDW = ", sum + sum3;
        print "TOTAL ELEC = ", sum2 + sum4;
        print "TOTAL VDW of short acid = ", sum7;
        print "TOTAL ELEC of short acid = ", sum8;
        print "TOTAL VDW of O1 = ", sum15;
        print "TOTAL ELEC of O1 = ", sum16;
        print "TOTAL VDW of O2 = ", sum17;
        print "TOTAL ELEC of O2 = ", sum18;
        print "TOTAL VDW of long acid = ", sum9;
        print "TOTAL ELEC of long acid = ", sum10;
        print "TOTAL VDW of OE1 = ", sum19;
        print "TOTAL ELEC of OE1 = ", sum20;
        print "TOTAL VDW of OE2 = ", sum21;
        print "TOTAL ELEC of OE2 = ", sum22;
        print "TOTAL VDW of ring = ", sum11;
        print "TOTAL ELEC of ring = ", sum12;
        print "TOTAL VDW of amide = ", sum13;
        print "TOTAL ELEC of amide = ", sum14;
        print "TOTAL nonbonded = ", sum2 + sum4 + sum + sum3}' $2
nawk '
{if ($4 ~ /^(CP1|CP2|CP3)$/)
{
        {if (NF == 12)
        {sum += $11; sum2 += $12}
}
        {if (NF == 11)
        {sum += $10; sum2 += $11}
}
}
}
```

```
{if ($4 ~ /^(NA2|H39|H38|H37|C2|N3|N1|C4|OA4|C4A|C8A|C8|C5|C6|C7|C9)$/)
{
        {if (NF == 12)
        {sum5 += $11; sum6 += $12}
}
        {if (NF == 11)
        {sum5 += $10; sum6 += $11}
}
}
}
{if ($4 !~ /^(CP1|CP2|CP3)$/)
{
        {if (NF == 12)
        {sum3 += $11; sum4 += $12}
}
        {if (NF == 11)
        {sum3 += $10; sum4 += $11}
}
}
}
{if ($4 ~ /^(O1|O2|CT)$/)
{
        {if (NF == 12)
        {sum7 += $11; sum8 += $12}
}
        {if (NF == 11)
        {sum7 += $10; sum8 += $11}
}
}
}
{if ($4 ~ /^(O1)$/)
{
        {if (NF == 12)
        {sum15 += $11; sum16 += $12}
}
        {if (NF == 11)
        {sum15 += $10; sum16 += $11}
}
}
}
{if ($4 ~ /^(O2)$/)
{
        {if (NF == 12)
        {sum17 += $11; sum18 += $12}
}
        {if (NF == 11)
        {sum17 += $10; sum18 += $11}
}
}
}
{if ($4 ~ /^(OE1)$/)
{
        {if (NF == 12)
        {sum19 += $11; sum20 += $12}
}
        {if (NF == 11)
        {sum19 += $10; sum20 += $11}
}
```

```
}
}
{if ($4 ~ /^(OE2)$/)
{
        {if (NF == 12)
        {sum21 += $11; sum22 += $12}
}
        {if (NF == 11)
        {sum21 += $10; sum22 += $11}
}
}
}
{if ($4 ~ /^(OE1|OE2|CD|CG|CB)$/)
{
        {if (NF == 12)
        {sum9 += $11; sum10 += $12}
}
        {if (NF == 11)
        {sum9 += $10; sum10 += $11}
}
}
}
{if ($4 ~ /^(C11|C12|C13|C14|C15|C16)$/)
{
        {if (NF == 12)
        {sum11 += $11; sum12 += $12}
}
        {if (NF == 11)
        {sum11 += $10; sum12 += $11}
}
}
}
{if ($4 ~ /^(C|O|N|H)$/)
{
        {if (NF == 12)
        {sum13 += $11; sum14 += $12}
}
        {if (NF == 11)
        {sum13 += $10; sum14 += $11}
}
}
}
END     {print "FILE = '$3'";
        print "TOTAL VDW of prop = ", sum;
        print "TOTAL ELEC of prop = ", sum2;
        print "TOTAL VDW of quin = ", sum5;
        print "TOTAL ELEC of quin = ", sum6;
        print "TOTAL nonbonded of both sides = ", sum5 + sum6 + sum + sum2;
        print "TOTAL VDW of main = ", sum3;
        print "TOTAL ELEC of main = ", sum4;
        print "TOTAL nonbonded of main = ", sum3 + sum4;
        print "TOTAL VDW = ", sum + sum3;
        print "TOTAL ELEC = ", sum2 + sum4;
        print "TOTAL VDW of short acid = ", sum7;
        print "TOTAL ELEC of short acid = ", sum8;
        print "TOTAL VDW of O1 = ", sum15;
        print "TOTAL ELEC of O1 = ", sum16;
        print "TOTAL VDW of O2 = ", sum17;
```

```
        print "TOTAL ELEC of O2 = ", sum18;
        print "TOTAL VDW of long acid = ", sum9;
        print "TOTAL ELEC of long acid = ", sum10;
        print "TOTAL VDW of OE1 = ", sum19;
        print "TOTAL ELEC of OE1 = ", sum20;
        print "TOTAL VDW of OE2 = ", sum21;
        print "TOTAL ELEC of OE2 = ", sum22;
        print "TOTAL VDW of ring = ", sum11;
        print "TOTAL ELEC of ring = ", sum12;
        print "TOTAL VDW of amide = ", sum13;
        print "TOTAL ELEC of amide = ", sum14;
        print "TOTAL nonbonded = ", sum2 + sum4 + sum + sum3}' $3
nawk '
{if ($4 ~ /^(CP1|CP2|CP3)$/)
{
        {if (NF == 12)
        {sum += $11; sum2 += $12}
}
        {if (NF == 11)
        {sum += $10; sum2 += $11}
}
}
}
{if ($4 ~ /^(NA2|H39|H38|H37|C2|N3|N1|C4|OA4|C4A|C8A|C8|C5|C6|C7|C9)$/)
{
        {if (NF == 12)
        {sum5 += $11; sum6 += $12}
}
        {if (NF == 11)
        {sum5 += $10; sum6 += $11}
}
}
}
{if ($4 !~ /^(CP1|CP2|CP3)$/)
{
        {if (NF == 12)
        {sum3 += $11; sum4 += $12}
}
        {if (NF == 11)
        {sum3 += $10; sum4 += $11}
}
}
}
{if ($4 ~ /^(O1|O2|CT)$/)
{
        {if (NF == 12)
        {sum7 += $11; sum8 += $12}
}
        {if (NF == 11)
        {sum7 += $10; sum8 += $11}
}
}
}
{if ($4 ~ /^(O1)$/)
{
        {if (NF == 12)
        {sum15 += $11; sum16 += $12}
}
```

```
        {if (NF == 11)
        {sum15 += $10; sum16 += $11}
}
}
}
{if ($4 ~ /^(O2)$/)
{
        {if (NF == 12)
        {sum17 += $11; sum18 += $12}
}
        {if (NF == 11)
        {sum17 += $10; sum18 += $11}
}
}
}
{if ($4 ~ /^(OE1)$/)
{
        {if (NF == 12)
        {sum19 += $11; sum20 += $12}
}
        {if (NF == 11)
        {sum19 += $10; sum20 += $11}
}
}
}
{if ($4 ~ /^(OE2)$/)
{
        {if (NF == 12)
        {sum21 += $11; sum22 += $12}
}
        {if (NF == 11)
        {sum21 += $10; sum22 += $11}
}
}
}
{if ($4 ~ /^(OE1|OE2|CD|CG|CB)$/)
{
        {if (NF == 12)
        {sum9 += $11; sum10 += $12}
}
        {if (NF == 11)
        {sum9 += $10; sum10 += $11}
}
}
}
{if ($4 ~ /^(C11|C12|C13|C14|C15|C16)$/)
{
        {if (NF == 12)
        {sum11 += $11; sum12 += $12}
}
        {if (NF == 11)
        {sum11 += $10; sum12 += $11}
}
}
}
{if ($4 ~ /^(C|O|N|H)$/)
{
        {if (NF == 12)
```

```
        {sum13 += $11; sum14 += $12}
}
        {if (NF == 11)
        {sum13 += $10; sum14 += $11}
}
}
}
END     {print "FILE = '$4'";
        print "TOTAL VDW of prop = ", sum;
        print "TOTAL ELEC of prop = ", sum2;
        print "TOTAL VDW of quin = ", sum5;
        print "TOTAL ELEC of quin = ", sum6;
        print "TOTAL nonbonded of both sides = ", sum5 + sum6 + sum + sum2;
        print "TOTAL VDW of main = ", sum3;
        print "TOTAL ELEC of main = ", sum4;
        print "TOTAL nonbonded of main = ", sum3 + sum4;
        print "TOTAL VDW = ", sum + sum3;
        print "TOTAL ELEC = ", sum2 + sum4;
        print "TOTAL VDW of short acid = ", sum7;
        print "TOTAL ELEC of short acid = ", sum8;
        print "TOTAL VDW of O1 = ", sum15;
        print "TOTAL ELEC of O1 = ", sum16;
        print "TOTAL VDW of O2 = ", sum17;
        print "TOTAL ELEC of O2 = ", sum18;
        print "TOTAL VDW of long acid = ", sum9;
        print "TOTAL ELEC of long acid = ", sum10;
        print "TOTAL VDW of OE1 = ", sum19;
        print "TOTAL ELEC of OE1 = ", sum20;
        print "TOTAL VDW of OE2 = ", sum21;
        print "TOTAL ELEC of OE2 = ", sum22;
        print "TOTAL VDW of ring = ", sum11;
        print "TOTAL ELEC of ring = ", sum12;
        print "TOTAL VDW of amide = ", sum13;
        print "TOTAL ELEC of amide = ", sum14;
        print "TOTAL nonbonded = ", sum2 + sum4 + sum + sum3}' $4
```

Appendix 2.4

solvation energies and input files

MINOR TS		E kcal/mol	MAJOR TS		E kcal/mol
CB3717	TS + ligand	-6113.03	CB3717	TS + ligand	-6151.21
	H_2O+ ligand	-6628.62		H_2O+ ligand	-6618.23
	ΔEnergy	**-515.59**		ΔEnergy	**-467.02**
	H_2O only	-6352.93		H_2O only	-6394.04
BIF1	TS + ligand	-6136.85	BIF2	TS + ligand	-6127.05
	H_2O+ ligand	-6613.77		H_2O+ ligand	-6605.50
	ΔEnergy	**-476.92**		ΔEnergy	**-478.45**
	H_2O only	-6347.63		H_2O only	-6348.96
	ΔΔE	**-38.67**		ΔΔE	**11.43**
BIF3	TS + ligand	-6109.14	BIF4	TS + ligand	-6131.47
	H_2O+ ligand	-6573.33		H_2O+ ligand	-6577.30
	ΔEnergy	**-464.19**		ΔEnergy	**-445.83**
	H_2O only	-6346.02		H_2O only	-6329.25
	ΔΔE	**-51.40**		ΔΔE	**-21.19**
BIF5	TS + ligand	-6122.46	BIF6	TS + ligand	-6111.41
	H_2O+ ligand	-6535.69		H_2O+ ligand	-6560.46
	ΔEnergy	**-414.23**		ΔEnergy	**-449.05**
	H_2O only	-6342.39		H_2O only	-6333.36
	ΔΔE	**-101.36**		ΔΔE	**-17.97**
BIF7	TS + ligand	-6115.25	BIF8	TS + ligand	-6124.65
	H_2O+ ligand	-6520.20		H_2O+ ligand	-6575.71
	ΔEnergy	**-405.75**		ΔEnergy	**-451.06**
	H_2O only	-6342.94		H_2O only	-6333.25
	ΔΔE	**-109.84**		ΔΔE	**-15.96**
BIF9	TS + ligand	-6114.83	BIF10	TS + ligand	-6140.55
	H_2O+ ligand	-6550.45		H_2O+ ligand	-6564.85
	ΔEnergy	**-435.61**		ΔEnergy	**-424.30**
	H_2O only	-6341.59		H_2O only	-6250.86
	ΔΔE	**-79.98**		ΔΔE	**-42.72**
BIF11	TS + ligand	-6103.57	BIF12	TS + ligand	-6141.63
	H_2O+ ligand	-6560.13		H_2O+ ligand	-6505.89
	ΔEnergy	**-456.56**		ΔEnergy	**-364.26**
	H_2O only	-6354.63		H_2O only	-6556.65
	ΔΔE	**-59.03**		ΔΔE	**-102.76**
BIF13	TS + ligand	-6110.62	BIF14	TS + ligand	-6145.32
	H_2O+ ligand	-6562.25		H_2O+ ligand	-6600.84
	ΔEnergy	**-451.63**		ΔEnergy	**-455.52**
	H_2O only	-6331.56		H_2O only	-6338.98
	ΔΔE	**-63.96**		ΔΔE	**-11.50**

BIF15	TS + ligand	-6117.24	BIF16	TS + ligand	-6125.58
	H_2O+ ligand	-6556.09		H_2O+ ligand	-6560.82
	ΔEnergy	**-438.84**		ΔEnergy	**-435.50**
	H_2O only	-6340.65		H_2O only	-6350.72
	ΔΔE	**-76.75**		ΔΔE	**-31.52**
BIF17	TS + ligand	-6100.71	BIF18	TS + ligand	-6147.70
	H_2O+ ligand	-6558.39		H_2O+ ligand	-6572.22
	ΔEnergy	**-457.68**		ΔEnergy	**-424.52**
	H_2O only	-6351.04		H_2O only	-6345.47
	ΔΔE	**-57.91**		ΔΔE	**-42.50**
BIF19	TS + ligand	-6073.26	BIF20	TS + ligand	-6103.88
	H_2O+ ligand	-6528.56		H_2O+ ligand	-6545.30
	ΔEnergy	**-455.30**		ΔEnergy	**-441.42**
	H_2O only	-6356.12		H_2O only	-6351.34
	ΔΔE	**-60.29**		ΔΔE	**-25.60**

```
* Script file produced by QUANTA
*
!.....
! CHARMm startup file for QUANTA
!
! Redirect the CHARMM output
open write unit 7 Card name bif14_in_final.LOG line
outu 7
bomlev -2
bann
open read unit 11 file name $CHM_DATA/AMINO.BIN
read rtf unit 11 file
open read unit 12 card name $CHM_DATA/PATCHA.RTF
read rtf unit 12 card appe
open read unit 02 file name PARM.BIN
read para unit 02 file
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "DUM.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "CB3.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "MAJO.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "BI14.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
read sequ quanta
* Read sequence using special QUANTA option
*
    1
BI14  1
generate BI14 setup
read sequ quanta
* Read sequence using special QUANTA option
*
  264
MET   1    LYS   2    GLN   3    TYR   4    LEU   5    GLU   6
LEU   7    MET   8    GLN   9    LYS   10   VAL   11   LEU   12
ASP   13   GLU   14   GLY   15   THR   16   GLN   17   LYS   18
ASN   19   ASP   20   ARG   21   THR   22   GLY   23   THR   24
GLY   25   THR   26   LEU   27   SER   28   ILE   29   PHE   30
GLY   31   HSD   32   GLN   33   MET   34   ARG   35   PHE   36
ASN   37   LEU   38   GLN   39   ASP   40   GLY   41   PHE   42
PRO   43   LEU   44   VAL   45   THR   46   THR   47   LYS   48
```

```
ARG   49    CYS   50    HIS   51    LEU   52    ARG   53    SER   54
ILE   55    ILE   56    HSD   57    GLU   58    LEU   59    LEU   60
TRP   61    PHE   62    LEU   63    GLN   64    GLY   65    ASP   66
THR   67    ASN   68    ILE   69    ALA   70    TYR   71    LEU   72
HSD   73    GLU   74    ASN   75    ASN   76    VAL   77    THR   78
ILE   79    TRP   80    ASP   81    GLU   82    TRP   83    ALA   84
ASP   85    GLU   86    ASN   87    GLY   88    ASP   89    LEU   90
GLY   91    PRO   92    VAL   93    TYR   94    GLY   95    LYS   96
GLN   97    TRP   98    ARG   99    ALA   100   TRP   101   PRO   102
THR   103   PRO   104   ASP   105   GLY   106   ARG   107   HIS   108
ILE   109   ASP   110   GLN   111   ILE   112   THR   113   THR   114
VAL   115   LEU   116   ASN   117   GLN   118   LEU   119   LYS   120
ASN   121   ASP   122   PRO   123   ASP   124   SER   125   ARG   126
ARG   127   ILE   128   ILE   129   VAL   130   SER   131   ALA   132
TRP   133   ASN   134   VAL   135   GLY   136   GLU   137   LEU   138
ASP   139   LYS   140   MET   141   ALA   142   LEU   143   ALA   144
PRO   145   CYS   146   HSD   147   ALA   148   PHE   149   PHE   150
GLN   151   PHE   152   TYR   153   VAL   154   ALA   155   ASP   156
GLY   157   LYS   158   LEU   159   SER   160   CYS   161   GLN   162
LEU   163   TYR   164   GLN   165   ARG   166   SER   167   CYS   168
ASP   169   VAL   170   PHE   171   LEU   172   GLY   173   LEU   174
PRO   175   PHE   176   ASN   177   ILE   178   ALA   179   SER   180
TYR   181   ALA   182   LEU   183   LEU   184   VAL   185   HIS   186
MET   187   MET   188   ALA   189   GLN   190   GLN   191   CYS   192
ASP   193   LEU   194   GLU   195   VAL   196   GLY   197   ASP   198
PHE   199   VAL   200   TRP   201   THR   202   GLY   203   GLY   204
ASP   205   THR   206   HSD   207   LEU   208   TYR   209   SER   210
ASN   211   HSD   212   MET   213   ASP   214   GLN   215   THR   216
HSD   217   LEU   218   GLN   219   LEU   220   SER   221   ARG   222
GLU   223   PRO   224   ARG   225   PRO   226   LEU   227   PRO   228
LYS   229   LEU   230   ILE   231   ILE   232   LYS   233   ARG   234
LYS   235   PRO   236   GLU   237   SER   238   ILE   239   PHE   240
ASP   241   TYR   242   ARG   243   PHE   244   GLU   245   ASP   246
PHE   247   GLU   248   ILE   249   GLU   250   GLY   251   TYR   252
ASP   253   PRO   254   HSD   255   PRO   256   GLY   257   ILE   258
LYS   259   ALA   260   PRO   261   VAL   262   ALA   263   ILE   264
generate 1S   setup
read sequ quanta
* Read sequence using special QUANTA option
*
  264
MET   301   LYS   302   GLN   303   TYR   304   LEU   305   GLU   306
LEU   307   MET   308   GLN   309   LYS   310   VAL   311   LEU   312
ASP   313   GLU   314   GLY   315   THR   316   GLN   317   LYS   318
ASN   319   ASP   320   ARG   321   THR   322   GLY   323   THR   324
GLY   325   THR   326   LEU   327   SER   328   ILE   329   PHE   330
GLY   331   HSD   332   GLN   333   MET   334   ARG   335   PHE   336
ASN   337   LEU   338   GLN   339   ASP   340   GLY   341   PHE   342
PRO   343   LEU   344   VAL   345   THR   346   THR   347   LYS   348
ARG   349   CYS   350   HIS   351   LEU   352   ARG   353   SER   354
ILE   355   ILE   356   HSD   357   GLU   358   LEU   359   LEU   360
TRP   361   PHE   362   LEU   363   GLN   364   GLY   365   ASP   366
THR   367   ASN   368   ILE   369   ALA   370   TYR   371   LEU   372
HSD   373   GLU   374   ASN   375   ASN   376   VAL   377   THR   378
ILE   379   TRP   380   ASP   381   GLU   382   TRP   383   ALA   384
ASP   385   GLU   386   ASN   387   GLY   388   ASP   389   LEU   390
GLY   391   PRO   392   VAL   393   TYR   394   GLY   395   LYS   396
GLN   397   TRP   398   ARG   399   ALA   400   TRP   401   PRO   402
```

```
THR   403   PRO   404   ASP   405   GLY   406   ARG   407   HIS   408
ILE   409   ASP   410   GLN   411   ILE   412   THR   413   THR   414
VAL   415   LEU   416   ASN   417   GLN   418   LEU   419   LYS   420
ASN   421   ASP   422   PRO   423   ASP   424   SER   425   ARG   426
ARG   427   ILE   428   ILE   429   VAL   430   SER   431   ALA   432
TRP   433   ASN   434   VAL   435   GLY   436   GLU   437   LEU   438
ASP   439   LYS   440   MET   441   ALA   442   LEU   443   ALA   444
PRO   445   CYS   446   HSD   447   ALA   448   PHE   449   PHE   450
GLN   451   PHE   452   TYR   453   VAL   454   ALA   455   ASP   456
GLY   457   LYS   458   LEU   459   SER   460   CYS   461   GLN   462
LEU   463   TYR   464   GLN   465   ARG   466   SER   467   CYS   468
ASP   469   VAL   470   PHE   471   LEU   472   GLY   473   LEU   474
PRO   475   PHE   476   ASN   477   ILE   478   ALA   479   SER   480
TYR   481   ALA   482   LEU   483   LEU   484   VAL   485   HIS   486
MET   487   MET   488   ALA   489   GLN   490   GLN   491   CYS   492
ASP   493   LEU   494   GLU   495   VAL   496   GLY   497   ASP   498
PHE   499   VAL   500   TRP   501   THR   502   GLY   503   GLY   504
ASP   505   THR   506   HSD   507   LEU   508   TYR   509   SER   510
ASN   511   HSD   512   MET   513   ASP   514   GLN   515   THR   516
HSD   517   LEU   518   GLN   519   LEU   520   SER   521   ARG   522
GLU   523   PRO   524   ARG   525   PRO   526   LEU   527   PRO   528
LYS   529   LEU   530   ILE   531   ILE   532   LYS   533   ARG   534
LYS   535   PRO   536   GLU   537   SER   538   ILE   539   PHE   540
ASP   541   TYR   542   ARG   543   PHE   544   GLU   545   ASP   546
PHE   547   GLU   548   ILE   549   GLU   550   GLY   551   TYR   552
ASP   553   PRO   554   HSD   555   PRO   556   GLY   557   ILE   558
LYS   559   ALA   560   PRO   561   VAL   562   ALA   563   ILE   564
generate 2S   setup
read sequ quanta
* Read sequence using special QUANTA option
*
    2
DUM   565   CB3   566
generate 1H   setup
read sequ quanta
* Read sequence using special QUANTA option
*
    1
DUM   567
generate 2H   setup

! QUANTA coordinates included in script file

open read unit 10 card name bif14_in_ts.crd
read coor unit 10 card
close unit 10

cons fix sele (segid 1s .or. segid 1h) .and. .not. (resid 124 : 128) end
print coor sele (segid 1s .or. segid 1h) .and. .not. (resid 124 : 128) end

! Energy script for CHARMm
!

FAST 0

SHAKE OFF
```

```
SKIPE NONE


MINI SD -
 NSTE 25 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000 -
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000

MINI ABNR -
 NSTE 500 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000

MINI ABNR -
 NSTE 500 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000


open write unit 10 card name bif14_final_min.crd
write coor unit 10 card
close unit 10
```

```
* Script file produced by QUANTA
*
!.....
! CHARMm startup file for QUANTA
!
! Redirect the CHARMM output
open write unit 7 Card name bif15_in_final.LOG line
outu 7
bomlev -2
bann
open read unit 11 file name $CHM_DATA/AMINO.BIN
read rtf unit 11 file
open read unit 12 card name $CHM_DATA/PATCHA.RTF
read rtf unit 12 card appe
open read unit 02 file name PARM.BIN
read para unit 02 file
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "DUM.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "CB3.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "PHS.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
! Script for reading RTF
DELE ATOM SELECT ALL END
OPEN READ UNIT 77 CARD NAME -
 "BI15.RTF"
READ RTF CARD UNIT 77 APPEND
CLOSE UNIT 77
read sequ quanta
* Read sequence using special QUANTA option
*
    1
BI15  1
generate BI15 setup
read sequ quanta
* Read sequence using special QUANTA option
*
  264
MET   1     LYS   2     GLN   3     TYR   4     LEU   5     GLU   6
LEU   7     MET   8     GLN   9     LYS   10    VAL   11    LEU   12
ASP   13    GLU   14    GLY   15    THR   16    GLN   17    LYS   18
ASN   19    ASP   20    ARG   21    THR   22    GLY   23    THR   24
GLY   25    THR   26    LEU   27    SER   28    ILE   29    PHE   30
GLY   31    HSD   32    GLN   33    MET   34    ARG   35    PHE   36
ASN   37    LEU   38    GLN   39    ASP   40    GLY   41    PHE   42
PRO   43    LEU   44    VAL   45    THR   46    THR   47    LYS   48
```

```
ARG   49    CYS   50    HIS   51    LEU   52    ARG   53    SER   54
ILE   55    ILE   56    HSD   57    GLU   58    LEU   59    LEU   60
TRP   61    PHE   62    LEU   63    GLN   64    GLY   65    ASP   66
THR   67    ASN   68    ILE   69    ALA   70    TYR   71    LEU   72
HSD   73    GLU   74    ASN   75    ASN   76    VAL   77    THR   78
ILE   79    TRP   80    ASP   81    GLU   82    TRP   83    ALA   84
ASP   85    GLU   86    ASN   87    GLY   88    ASP   89    LEU   90
GLY   91    PRO   92    VAL   93    TYR   94    GLY   95    LYS   96
GLN   97    TRP   98    ARG   99    ALA   100   TRP   101   PRO   102
THR   103   PRO   104   ASP   105   GLY   106   ARG   107   HIS   108
ILE   109   ASP   110   GLN   111   ILE   112   THR   113   THR   114
VAL   115   LEU   116   ASN   117   GLN   118   LEU   119   LYS   120
ASN   121   ASP   122   PRO   123   ASP   124   SER   125   ARG   126
ARG   127   ILE   128   ILE   129   VAL   130   SER   131   ALA   132
TRP   133   ASN   134   VAL   135   GLY   136   GLU   137   LEU   138
ASP   139   LYS   140   MET   141   ALA   142   LEU   143   ALA   144
PRO   145   CYS   146   HSD   147   ALA   148   PHE   149   PHE   150
GLN   151   PHE   152   TYR   153   VAL   154   ALA   155   ASP   156
GLY   157   LYS   158   LEU   159   SER   160   CYS   161   GLN   162
LEU   163   TYR   164   GLN   165   ARG   166   SER   167   CYS   168
ASP   169   VAL   170   PHE   171   LEU   172   GLY   173   LEU   174
PRO   175   PHE   176   ASN   177   ILE   178   ALA   179   SER   180
TYR   181   ALA   182   LEU   183   LEU   184   VAL   185   HIS   186
MET   187   MET   188   ALA   189   GLN   190   GLN   191   CYS   192
ASP   193   LEU   194   GLU   195   VAL   196   GLY   197   ASP   198
PHE   199   VAL   200   TRP   201   THR   202   GLY   203   GLY   204
ASP   205   THR   206   HSD   207   LEU   208   TYR   209   SER   210
ASN   211   HSD   212   MET   213   ASP   214   GLN   215   THR   216
HSD   217   LEU   218   GLN   219   LEU   220   SER   221   ARG   222
GLU   223   PRO   224   ARG   225   PRO   226   LEU   227   PRO   228
LYS   229   LEU   230   ILE   231   ILE   232   LYS   233   ARG   234
ALA   235   PRO   236   GLU   237   SER   238   ILE   239   PHE   240
ASP   241   TYR   242   ARG   243   PHE   244   GLU   245   ASP   246
PHE   247   GLU   248   ILE   249   GLU   250   GLY   251   TYR   252
ASP   253   PRO   254   HSD   255   PRO   256   GLY   257   ILE   258
LYS   259   ALA   260   PRO   261   VAL   262   ALA   263   ILE   264
generate 1S   setup
read sequ quanta
* Read sequence using special QUANTA option
*
  264
MET   301   LYS   302   GLN   303   TYR   304   LEU   305   GLU   306
LEU   307   MET   308   GLN   309   LYS   310   VAL   311   LEU   312
ASP   313   GLU   314   GLY   315   THR   316   GLN   317   LYS   318
ASN   319   ASP   320   ARG   321   THR   322   GLY   323   THR   324
GLY   325   THR   326   LEU   327   SER   328   ILE   329   PHE   330
GLY   331   HSD   332   GLN   333   MET   334   ARG   335   PHE   336
ASN   337   LEU   338   GLN   339   ASP   340   GLY   341   PHE   342
PRO   343   LEU   344   VAL   345   THR   346   THR   347   LYS   348
ARG   349   CYS   350   HIS   351   LEU   352   ARG   353   SER   354
ILE   355   ILE   356   HSD   357   GLU   358   LEU   359   LEU   360
TRP   361   PHE   362   LEU   363   GLN   364   GLY   365   ASP   366
THR   367   ASN   368   ILE   369   ALA   370   TYR   371   LEU   372
HSD   373   GLU   374   ASN   375   ASN   376   VAL   377   THR   378
ILE   379   TRP   380   ASP   381   GLU   382   TRP   383   ALA   384
ASP   385   GLU   386   ASN   387   GLY   388   ASP   389   LEU   390
GLY   391   PRO   392   VAL   393   TYR   394   GLY   395   LYS   396
GLN   397   TRP   398   ARG   399   ALA   400   TRP   401   PRO   402
```

```
THR   403   PRO   404   ASP   405   GLY   406   ARG   407   HIS   408
ILE   409   ASP   410   GLN   411   ILE   412   THR   413   THR   414
VAL   415   LEU   416   ASN   417   GLN   418   LEU   419   LYS   420
ASN   421   ASP   422   PRO   423   ASP   424   SER   425   ARG   426
ARG   427   ILE   428   ILE   429   VAL   430   SER   431   ALA   432
TRP   433   ASN   434   VAL   435   GLY   436   GLU   437   LEU   438
ASP   439   LYS   440   MET   441   ALA   442   LEU   443   ALA   444
PRO   445   CYS   446   HSD   447   ALA   448   PHE   449   PHE   450
GLN   451   PHE   452   TYR   453   VAL   454   ALA   455   ASP   456
GLY   457   LYS   458   LEU   459   SER   460   CYS   461   GLN   462
LEU   463   TYR   464   GLN   465   ARG   466   SER   467   CYS   468
ASP   469   VAL   470   PHE   471   LEU   472   GLY   473   LEU   474
PRO   475   PHE   476   ASN   477   ILE   478   ALA   479   SER   480
TYR   481   ALA   482   LEU   483   LEU   484   VAL   485   HIS   486
MET   487   MET   488   ALA   489   GLN   490   GLN   491   CYS   492
ASP   493   LEU   494   GLU   495   VAL   496   GLY   497   ASP   498
PHE   499   VAL   500   TRP   501   THR   502   GLY   503   GLY   504
ASP   505   THR   506   HSD   507   LEU   508   TYR   509   SER   510
ASN   511   HSD   512   MET   513   ASP   514   GLN   515   THR   516
HSD   517   LEU   518   GLN   519   LEU   520   SER   521   ARG   522
GLU   523   PRO   524   ARG   525   PRO   526   LEU   527   PRO   528
LYS   529   LEU   530   ILE   531   ILE   532   LYS   533   ARG   534
ALA   535   PRO   536   GLU   537   SER   538   ILE   539   PHE   540
ASP   541   TYR   542   ARG   543   PHE   544   GLU   545   ASP   546
PHE   547   GLU   548   ILE   549   GLU   550   GLY   551   TYR   552
ASP   553   PRO   554   HSD   555   PRO   556   GLY   557   ILE   558
LYS   559   ALA   560   PRO   561   VAL   562   ALA   563   ILE   564
generate 2S   setup
read sequ quanta
* Read sequence using special QUANTA option
*
    1
DUM   565
generate 1H   setup
read sequ quanta
* Read sequence using special QUANTA option
*
    2
PHS   567   CB3   568
generate 2H   setup

! QUANTA coordinates included in script file

open read unit 10 card name bif15_in_ts.crd
read coor unit 10 card
close unit 10

cons fix sele (segid 2s .or. segid 2h) .and. .not. (resid 424 : 428) end
print coor sele (segid 2s .or. segid 2h) .and. .not. (resid 424 : 428) end
! Energy script for CHARMm
! Energy script for CHARMm
!


FAST 0

SHAKE OFF
```

```
SKIPE NONE


MINI SD -
 NSTE 25 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000 -
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000

MINI ABNR -
 NSTE 500 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000

MINI ABNR -
 NSTE 500 NPRI 5 TOLG 0.000000 STEP 0.020000 TOLS 0.000000 TOLENR 0.001000
 INBFRQ 10 CUTNB 12.000000 CTONNB 9.000000 CTOFNB 11.000000 -
 VSWITCH SHIFT RDIE EPS 1.000000 -
 IHBFRQ 0 -
 IMGFRQ 10 CUTIM 12.000000


open write unit 10 card name bif15_final_min.crd
write coor unit 10 card
close unit 10
```