PLATYPUS INSULIN:
STRUCTURAL ANALYSIS, CHEMICAL SYNTHESIS
AND
PHYSICO-CHEMICAL PROPERTIES

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

of the

AUSTRALIAN NATIONAL UNIVERSITY

by

AMANDA NOURSE

August 1996

The Platypus (*Ornithorhynchus anatinus*)

(Painting by Rod Scott that appeared on the cover of Australian Geographic, 1988, Number 12. Permission was obtained for reproduction by the editor.)

STATEMENT

The research work described in this thesis was undertaken in the Division of Biochemistry and Molecular Biology and the Biomolecular Resource Facility at the John Curtin School of Medical Research, Institute of Advanced Studies, Australian National University between March 1993 and July 1996. All the experimental results reported in the text were obtained by the author alone except for the purification (Dr G Barbara Treacy) and the amino-acid sequence determination (Dr Denis C Shaw) of native platypus insulin. The work presented in this thesis has not previously been submitted for a degree of this or any other University.

A Nourse

Amanda Nourse

ACKNOWLEDGMENTS

I would like to thank Dr Peter D Jeffrey and Dr Peter J Milburn for their most valuable assistance, advice and encouragement given to me throughout the course of this work. Both have my deepest gratitude and admiration.

Financial support from the Australian National University in the form of a PhD scholarship is gratefully acknowledged.

I also wish to thank the following persons:
All the members of the Biomolecular Resource Facility for their assistance and practical help;
Mr Kerry N McAndrew for the chemical synthesis of the peptide chains of insulin;
Dr Denis C Shaw for help with the sequence comparison studies;
Prof Frank Gibson for assistance with the molecular modelling studies;
Dr Rosemary L Martin who kindly provided me with rat epididymal tissue for the biological activity measurements of insulin;
Mr Kingsley Sutton for assistance with the initial biological activity measurements;
Dr Mervyn Griffiths for the donation of platypus pancreases;
The members of the Photography department for the production of slides as well as printing the coloured figures that appeared in thesis and
The members of the computer section of the John Curtin School of Medical Research for assistance with computers and programmes.

Finally I would also like to thank my husband, Leonard, for his invaluable support during the course of this work as well as during the compilation and writing of this thesis; my two children, Audrey and Henry, who have been part of this effort and my parents, Inus and Jackie Grobler, who have always encouraged and supported me in attaining my goals.

LIST OF ABBREVIATIONS

Å	angstrom
α	alpha
Acm	acetamidomethyl
β	beta
Boc	tert-butyloxycarbonyl
BSA	bovine serum albumin
C-terminal	carboxyl-terminal
cDNA	complementary deoxyribonucleic acid
CH_3CN	acetonitrile
CD	circular dichroism
DAS	diaminosuberic acid
DCC	*N,N*-dicyclohexylcarbodiimide
D glucose	dextro glucose
D.I. H_2O	distilled and deionised water
DKP	B10 Asp; B28 Lys; B29 Pro human insulin
DMF	*N,N*-dimethyformamide
DNA	deoxyribonucleic acid
2D NMR	two dimensional nuclear magnetic resonance
DTDP	2,2'-dithiodipyridine
EDT	ethanedithiol
EDTA	ethylenediaminetetra acetic acid
FB24S	B24 Phe replaced by a Ser
Fmoc	9-fluorenylmethoxycarbonyl
FW	formula weight
g	gram
GCG	Genetics Computer Group
Hepes	(*N*-[2-hydroxyethyl]piperazine-*N*'-[2-ethanesulfonic acid])

HOBt	1-hydroxybenzotriazole
HPCE	high performance capillary electrophoresis
HPLC	high performance liquid chromatography
HB10D	B10 His replaced by Asp
IDI	indefinite duoisodesmic
IRS-1	insulin receptor substrate-1
KRBH	Krebs Ringer Bicarbonate HEPES
L	litre
M	molar
MRW	mean residual weight in daltons
MALDI-TOF-MS	matrix-assisted laser desorption ionisation time of flight-mass spectrometry
MA13Nle	A13 Met replace by a nor leucine
MA13Mso	A 13 Met replaced by a Met-sulfoxide
min	minutes
mg	milli gram
mL	milli litre
mM	milli molar
ng	nano gram
nM	nano molar
fM	femto molar
MM	molecular mass
N-terminal	amino-terminal
nm	nano meter
NMP	*N*-methyl-2-pyrrolidone
OP	Two-fold axis (monomer-monomer interface)
OQ	Two-fold axis (dimer-dimer interface)
PB2Nma	B2 Pro replace by N-methyl alanine
PMC	2,2,5,7,8-pentamethylchroman-6-sulfonyl

RP-HPLC	reversed phase high performance liquid chromatography
SB9D	B9 Ser replaced by Asp
SH2	Src homology 2
S-Pyr	2-pyridinesulfenyl
TEAA	triethylethanolamine
TEA	triethylamine
TRIS	(Tris[hydroxymethyl]aminomethane)
TFA	Trifluoroacetic acid
Trt	trityl
Tris	tris(hydroxymethyl)aminutesutesomethane
tBu	tert-butyl
µg	micro gram
µL	micro litre
µM	micro molar
UV	ultra violet
VIS	visible
Zn	zinc

Amino-acids

Ala	A	alanine
Arg	R	arginine
Asp	D	aspartic acid
Asn	N	asparagine
Cys	C	cysteine
Glu	E	glutamic acid
Gly	G	glycine
Gln	Q	glutamine
His	H	histidine

Ile	I	isoleucine
Leu	L	leucine
Lys	K	lysine
Met	M	methionine
Phe	F	phenylalanine
Pro	P	proline
Ser	S	serine
Thr	T	threonine
Trp	W	tryptophan
Tyr	Y	tyrosine
Val	V	valine

GLOSSARY OF SYMBOLS

a_i	weight activity of species *i*.
c_i	weight concentration of species *i*.
°C	degrees Celsius
$\bar{c}$	total weight concentration
$\varepsilon_{\lambda}^{1cm}$	optical extinction coefficient at wavelength λ of a solution in a 1 cm cuvette
$\Delta\varepsilon$	differential dichroic absorption (M^{-1} cm^{-1})
ED_{50}	the log[Dose] when response=50%
λ	wavelength of light
h	thickness of ultracentrifuge cell
i	solute species
I	ionic strength of supporting electrolyte (buffer)
K	Kelvin
K_A	molar-based indefinite duoisodesmic self-association constant (monomer-monomer interface)
K_B	molar-based indefinite duoisodesmic self-association constant (dimer-dimer interface)
K_I	molar-based isodesmic self-association constant
$K_{I,6}$	molar-based equilibrium constant for isodesmic self-association of hexamer units
m_i	molarity of species *i*
M_i	molecular weight of species *i*
M_w	weight-average molecular weight
$M_{w,app}$	apparent weight-average molecular weight
OD_{λ}	optical density at wavelength λ
Ω(r)	function relating the weight activity fraction of monomer at some constant weight concentration to that at any other concentration

P	pressure
p	density of the solution
r	radius from axis of rotation
r_b	radial position of the base of the solution column
r_m	radial position of the meniscus of the solution column
r_F	reference radial position
(r)	shows functional dependence on radius e.g. $\bar{c}$ (r) the total weight concentration at radius r
R	gas constant
T	temperature
θ	sector angle of the ultracentrifuge cell (ultracentrifugation); observed rotation in millidegrees (circular dichroism)
$[\theta]$	mean residue ellipticity (deg cm^2/dmol)
$[\theta]_\lambda$	mean residue ellipticity at a particular wavelength (λ) (deg cm^2/dmol)
μ_i^0	standard chemical potential per gram of solute i
μ_i	chemical potential per gram of solute i
$\bar{\upsilon}$	partial specific volume
ω	angular velocity

NOTATION

The notation for each residue is to give the chain letter (A or B) first, followed by its number in the sequence, followed by the number 1 or 2 representing molecule 1 or 2, followed by the residue name; eg. B25.1 Phe represents B25 phenylalanine of molecule 1.

Mutant insulins are described in turn by the native residue name, the chain letter and number in the sequence, the name of the substituted residue and the species name and insulin. Residue names are given in one letter codes. For example HB10 D platypus insulin, where the His (native) is substituted by Asp (mutant).

ABSTRACT

Platypus Insulin: Indications from the Amino-acid Sequence of Significant Differences in Structure from Porcine Insulin.

Insulin from a monotreme, the platypus (*Ornithorhynchus anatinus*), was isolated and the amino-acid sequence determined. It differs from pig insulin at eleven amino-acid sites, mainly on the surface of the monomer. Substitutions relative to pig insulin occur in the monomer-monomer interface, the dimer-dimer interface and the receptor binding region. The residues A5 Glu, A8 Lys and A13 Met have not been reported before in any insulin. Multiple sequence comparison studies reveal a relatively close relationship with the nearest group of relatives to the platypus, the mammals. The relationship of the platypus sequence to reptilian insulin sequences (and amphibian and avian insulin sequences in this case) is sufficiently close to support the observation that platypus has retained some ancient reptilian characteristics over the course of evolution. Model building the platypus insulin sequence on the structure of porcine insulin indicates that there may be some interesting differences.

Total Chemical Synthesis of Native and Mutant HB10D Platypus Insulins

To investigate the biological and physico-chemical properties of platypus insulin, it was synthesised by totally chemical means. The A- and B-chains were synthesised separately by an automatic solid phase peptide synthesiser using Fmoc chemistry and acetamidomethyl (Acm) protection of Cys A6, Cys A11, Cys A20 and Cys B19. The method of synthesis used was largely based on the stepwise semi-regioselective formation of firstly the CysA7-CysB7 disulfide bond followed by the removal of *S*-Acm groups and concomitant formation of the other two disulfide bonds by means of iodine oxidation. The synthesis products were identified by mass spectrometric analysis. The method of total chemically synthesis employed was very successful. Also preliminary investigations into the self-association of synthetic platypus insulin by sedimentation analysis pointed to some reduction in the self-association at the dimer-dimer interface compared to bovine insulin. Furthermore, predicted differences

in the spatial arrangement of residues in this interface of the molecular model also suggested that the self-associating properties in this region may be affected, seemingly lowered.

This prompted an investigation of the properties of the dimer-dimer interface with platypus insulin analogues designed to probe that region. One of these, HB10D platypus insulin, was also chemically synthesised to investigate its self-association pattern. Both this analogue and the synthetic native platypus insulin were obtained in a pure form. A comparison of their circular dichroism spectra showed them to be similar to that of porcine insulin. *In vitro* measurement of the biological activity showed that of platypus insulin was identical to and HB10D platypus insulin double that of porcine insulin.

The Self-Association of Zinc-Free Native and Mutant HB10D Platypus Insulins

An extensive study of the self-association patterns of zinc-free synthetic native platypus and mutant (HB10D) platypus insulins in solution (pH = 7.0; I = 0.1 M; 25 °C) has been undertaken. Sedimentation equilibrium was chosen as the method of study as precise information could be obtained on the dependence of both weight-average molecular weight and of concentration of the monomer (m_1) on total protein concentration. Experimental data were fitted to a model, called indefinite duoisodesmic (IDI) of self-association, to obtain the relevant association constants.

The method of analysis of the sedimentation equilibrium data was as follows:

The sedimentation equilibrium data were analysed by the omega-method to obtain the thermodynamic parameter a_1 (the activity of the monomer in equilibrium with a series of polymeric species) as a function of total protein concentration. In the omega analysis, omega (Ω) function *versus* total protein concentration curves were extrapolated to yield the infinite dilution value, Ω^0 , whereby the thermodynamic activity of the self-associating monomer was determined at a reference point in the centrifuge cell. The activity of the monomer at this reference point was used to

calculate the activity of the monomer as a function of the total solute concentration. The concentration of the monomer m_1 was then calculated from the activity a_1 and known molecular weight of the monomer.

Once the sedimentation equilibrium data was analysed and transformed to concentration of the monomer m_1 as a function of total protein concentration it was fitted to a mathematical equation (Mark and Jeffrey, 1990) describing the Indefinite Duoisodesmic (IDI) model of self-association. From this the relevant association constants describing polymerising systems were calculated.

(IDI describes the indefinite self-association of a bivalent insulin monomer along the monomer-monomer (α-α) and dimer-dimer (β-β) interfaces governed by two association constants K_A and K_B respectively. The two initial pathways of dimerization are then followed by the indefinite addition of monomer under the control of the same two association constants. The above mathematical equation describes the composition of a solution consisting of an infinite array of odd and even numbered polymeric species).

The weight-fraction of each species was also calculated as a function of total insulin concentration using an expression in terms of the weight-fraction of species as a function of total concentration for the IDI model with known K_A and K_B values. The expression describes the complex distribution of species within the insulin system in solution. A distinction can be made between the weight-fraction of the two types of dimer α and β and tetramer α and β. It stresses the importance of the odd-numbered species, the dominant species by weight in bovine insulin under these particular conditions. The possible presence of odd-numbered species other than monomer have been taken into account.

In the studies on the self-association of the synthetic native and mutant platypus insulins each was compared with bovine insulin. It has been concluded that there is some reduction in the extent of the self-association of platypus insulin compared to bovine insulin. A reduction in specifically the dimer-dimer interaction is indicated by

the higher K_A and lower K_B values. HB10D platypus insulin shows a dramatic reduction in self-association compared to bovine insulin. (Similar results have been obtained with a HB10D human insulin). Analysis of the self-association pattern suggests that a K_B value of effectively zero virtually abolishes its dimer-dimer interaction.

Thus platypus insulin has essentially the same biological activity as porcine but somewhat lower self-association, while the introduction of one amino-acid in a critical region increases the activity two-fold while abolishing self-association beyond dimer.

TABLE OF CONTENTS

Chapter 2

Literature Review: The Structure-Function Relationship of Insulin

Chapter 3

Platypus Insulin: Structural Analysis

Chapter 4

Total Chemical Synthesis and Characterisation of Platypus and HB10D Platypus Insulins

Chapter 5

The Self-Association Pattern of Zinc-Free Bovine, Synthetic Platypus and HB10D Platypus Insulin: A Sedimentation Equilibrium Study

Chapter 1

Introduction

1.1 Aim of this project

Since the discovery of insulin by Banting and Best in the 1920's its use in the treatment of human insulin-dependent *diabetes mellitus* has been one of the spectacular successes of pharmaceutical science. The problems in conventional insulin therapy are that none of the presently available commercial preparations, in any combination, adequately reproduce the endogenous insulin-secretion-pattern resulting in poor blood-sugar control. Poor blood-sugar control can lead to microvascular complications; one such example is retinopathy in the eye suffered by many diabetics (Bristow, 1993). Many problems associated with its exogenous administration, namely nasal delivery, intravenous and in particular subcutaneous injection, are caused by the aggregation of insulin (Renard *et al.*, 1995; Olsen *et al.*, 1995; Lougheed *et al.*, 1980; Bristow, 1993) and this is an ongoing problem in insulin therapy (Galloway and Chance, 1994; Renard *et al.*, 1995; Olsen *et al.*, 1995).

The aggregation phenomenon may also lead to the formation of insoluble fibrils, amorphous precipitates and aggregates in therapeutic insulin formulations used in delivery systems. Aggregated forms of insulin also present numerous problems on application; for instance the generation of a significant percentage of immune reactions seen in diabetic patients (Robbins *et al.*, 1987), their implication in the formation of the amyloid growths at injection sites (Dische *et al.*, 1988), difficulties in controlling dosage and the likelihood that they may impair the activity of monomeric insulin by competing for receptor binding sites.

The problems of concern in this study are associated with the self-association of soluble insulin. The significance of these problems have been discussed by Galloway and Chance (Galloway and Chance, 1994). The approach used was to explore properties of an insulin with natural variations in its amino-acid sequence. Platypus is

such an exotic species with natural variation. The idea being that the amino-acid sequence of platypus insulin could hopefully show characteristics which would be of value in the design of insulin analogue(s) that were more user friendly in insulin therapy. By far the most advantageous property that can be built into a designed insulin molecule is that it does not self-associate while retaining full activity.

In order to investigate the structure and properties of platypus insulin it was necessary to have sufficient material but for ethical and moral reasons it was not contemplated that this could be obtained from platypus pancreases since it is an unique and protected animal of Australia. For the study, however, it was imperative that the amino-acid sequence of platypus insulin be known and fortunately sufficient material could be purified for its determination from platypus which had died of natural causes. To obtain sufficient material for further investigations other means of obtaining this was necessary and to do this chemical synthesis was thus chosen.

Also researched was the method for the total chemical synthesis of insulins by the strategy of the stepwise semi-regioselective formation of the disulfide bonds for the purpose of producing designed insulin analogues of a specific sequence for evaluation. The insulin molecule consists of two chains, namely the A- (21 amino-acids) and B- (30 amino-acids) chains, linked by two inter-chain disulfide bridges and one intra-chain disulfide bridge on the A-chain (Blundell *et al.*, 1972; Baker *et al.*, 1988). To synthesise this complex protein by total chemical means has to date been a difficult and laborious task (Brandenburg, 1990). Thus in this study considerable effort was put into optimising and modifying existing methods of chemical synthesis of insulin (and insulin-like molecules) to produce insulins in a pure form in a relatively short period of time. Although higher quantities of insulin can be produced by DNA recombinant technology it is not possible to produce insulins with some substitutions by this method, as for instance NA21P human insulin (Markussen *et al.*, 1988). An added advantage of chemical synthesis is that non-natural amino-acids as well as chemical side chains with known chemical properties can be built into the insulin molecule at specific sites for evaluation of their effect(s).

Since reducing or abolishing the self-association of therapeutic insulin is pivotal to improving insulin therapy, it is important to be able to analyse critically the self-association pattern of insulin analogues for potential use in solution. In this thesis sedimentation analysis has been used for evaluating chemically synthesised insulins in detail. Association constants, characteristic of the self-association pattern of a specific insulin, are derived using this analysis allowing quantitative assessment of the effects of sequence changes.

1.2 Problems of insulin therapy in the treatment of *diabetes mellitus*

1.2.1 The aggregation of insulin

Insulin self-associates reversibly in solution characteristically forming an array of polymeric species in equilibrium, of which the hexamer is most prominent. The zinc hexamer is the form in which insulin molecules are stored as inactive crystals in the pancreas. Hexamers in solution are evident at quite low concentrations even in the absence of zinc (Jeffrey, 1990). Many of the problems associated with the exogenous administration of insulin originate as a result of this aggregate formation at the high application concentrations that are used. These aggregates engage the potent monomers in interactions to form biologically inactive polymers.

Insulin can also form irreversible aggregates in solution resulting in precipitation (fibril or gel formation). This irreversible aggregation is a nucleation-driven process initiated by partial denaturated states of insulin (Sluzky *et al.*, 1991) and results in insulin fibrils that contains a perturbed native structure with increased β-structure (Yu *et al.*, 1974; Burke and Rougvie, 1972). This nucleation process is thought to take place in any of the following; the solution, the gas-liquid interface and the solid-liquid interface (Dathe *et al.*, 1990). Insulin adsorbs to various materials such as a variety of plastics and glass (Sefton and Antonacci, 1984; Cuatrecasas and Hollenberg, 1975). It is thought that the basic element for the nucleation process at solid-liquid interfaces is the monomer or dimer and are assumed to be adsorbed to the interface via their hydrophobic regions (Dathe *et al.*, 1990). Upon adsorption, insulin may alter its

conformation and subsequent molecules might bind to the adsorbed layer which progressively grows (Grau *et al.*, 1982).

The aggregation of insulin has limited development of insulin infusion devices for intravenous administration of insulin (Lougheed *et al.*, 1980). Changes of temperature, agitation and dissimilar metal pump components can also trigger precipitation in insulin solutions and limit the reproducibility of the dosages (Lougheed *et al.*, 1980).

1.2.2 Reproduction of the endogenous insulin-secretion-pattern

The endogenous insulin-secretion-pattern in humans consists of two components; namely the secretion of insulin at a continuous basal level and supplementation by rapid post prandial increases in response to elevated blood-sugar levels (Eaton *et al.*, 1980). Both components of insulin secretion are essential, the former to maintain normal blood-sugar levels (euglycaemia), particularly at night and in the morning, and the latter to regulate blood-sugar levels following meals.

Obviously the intended application of insulin therapy is to match or reproduce this secretion-pattern as closely as possible; namely the secretion at a continuous basal level in conjunction with a response to elevated blood glucose levels after a meal. Physiologically, insulin secretion peaks within one hour of an intake of an oral glucose load and returns to basal levels within approximately three hours (Brange *et al.*, 1990). The normal serum concentration of insulin is in the 1 nM range but to effect a lowering of post-prandial blood glucose levels in diabetics the insulin concentrations used in soluble formulations for injections are in the 4 mM range (Jeffrey, 1990; Bristow, 1993). This is about a million times more concentrated than that in circulation in normal individuals although incidentally the quantity applied per application is similar in quantity to the level in total blood volume. The exogenous administration of insulin results in the peaking of plasma levels only 1.5-2 hours after injection which remain elevated for 3-5 hours. As a consequence diabetics who

administer insulin at the time of a meal tend to experience hyperglycaemia within the first hour and hypoglycaemia somewhat later (Kang *et al.*, 1991b).

At the normal serum concentration of insulin (1 nM range) it exists for all practical purposes as a monomeric molecule, the form which is considered to be active. At the concentrations used in soluble formulations for injections (about 4 mM) it consists mostly as dimers and higher order polymers. Calculations from the studies of insulin self-association show that at 1 nM only one insulin molecule in ten thousand is not a monomer while at 4 mM only one insulin in about 50 is a monomer (Jeffrey, 1990). These are all problems which emphasise the need of research to produce more user friendly insulin preparations and in particular those associated with the aggregation of insulin.

1.3 Therapeutic insulins to counteract insulin therapy problems

In the treatment of diabetics attempts have been made to counteract some of these problems of aggregation of insulin and also of improving post-prandial control of peripheral glucose metabolism (Bristow, 1993) by producing insulin analogues with desirable properties such as:

- fully potent insulins that do not self-associate
- insulins with altered physico-chemical, pharmacokinetic and biological properties

  that can be used in formulations to mimic the normal insulin-secretion-pattern.

These insulin analogues can be produced by three different means:

- recombinant DNA technology,
- total chemical synthesis and
- enzymatic modification reactions.

1.3.1 Therapeutic insulins that are active and monomeric

It is generally thought that insulin aggregates have to dissociate to monomers in order to be absorbed from the subcutaneous injection site (Kang *et al.*, 1991b). The benefits of altering the insulin molecule to prevent self-association are an increase in the rate of absorption; quickening of the onset of effect and shortening the duration of effect (Kang *et al.*, 1991a; Kang *et al.*, 1991b). This would suggest that formulations which contain predominantly monomers would be advantageous. These would be considered quick-acting insulins and their use would also eliminate other problems arising from aggregation.

Since the indications are that monomeric insulin formulations would have the advantages of limited aggregation and rapid absorption, attempts have been made to keep native human insulin in this highly desirable state in solution (Brange *et al.*, 1988). Some of these were by addition (Bringer *et al.*, 1981) or removal of (Jeffrey, 1986) substances associated with therapeutic insulin solutions. Attempts have been made to produce insulin analogues with the desired properties. Molecules where the amino-acids in the insulins from human, pig and cow are strategically replaced or modified have been synthesised (Brange *et al.*, 1988; Weiss *et al.*, 1991; Brems *et al.*, 1992). The strategies that were employed to achieve this to date were to substitute or modify residues in the interfaces between molecules in dimers and higher polymers, the so-called monomer-monomer interface and the dimer-dimer interface, so as to reduce aggregation without adversely affecting the activity of the insulin.

At present several insulin analogues that are thought to be predominantly monomeric have and are being investigated for potential therapeutic use. With some of the analogues tested it has been claimed that they display clinically similar pharmacokinetics on injection (subcutaneously) at meal times to the endogenous insulin secretion seen in response to a meal (Kang *et al.*, 1991b). The methods used to assess the association states of these insulin analogues are not very sensitive and their true monomeric status in solution would still seem to be uncertain (Jeffrey, 1990). To date none of these "monomeric" insulins have been implemented in insulin therapy.

1.3.2 Therapeutic insulins with altered physico-chemical, pharmacokinetic and biological properties

Many insulin analogues with altered physico-chemical, pharmacokinetic and biological properties have been produced which show considerable potential as therapeutic agents in the management of diabetes (Bristow, 1993). These include production of

- superpotent analogues by increasing receptor affinity and alterations of the physiological actions of insulin
- insulin formulations which are quick-acting (in solution) and which contain monomeric insulin analogues or
- long-acting insulin (crystalline) which crystallises at the injection site or
- mixtures of both.

The production of superpotent analogues with a higher affinity for the insulin receptor have also received attention due to the efficacy of smaller therapeutic doses. B10 Asp (HB10D) human insulin is one such candidate that is superpotent *in vitro* but preliminary clinical evaluation, whilst clearly indicating the advantages due to the analogue's rapid absorption, failed to demonstrate an overall increased potency in man (Kang *et al.*, 1991a). It is possible that analogues with different receptor affinities might prove to be useful clinically on account of their differing specificities for the various insulin-responsive tissues. At present, however, clinical studies have not indicated any significant advantages in the use of the superpotent insulin analogues.

One example of a long-acting-insulin is an analogue that crystallises at physiological pH when injected thus delaying onset of action while increasing its duration. Such preparations have considerable clinical potential since a single daily injection could be prepared which closely matches the physiological basal rate of insulin secretion. Another that has been tried is proinsulin which displays about 10% of the activity of insulin and which is normally present in humans in small amounts and also has a prolonged plasma half-life. It also exhibits reduced self-association compared with native human insulin (Pekar and Frank, 1972). Evaluation of proinsulin as a

potentially clinically useful therapeutic agent in the treatment of diabetes led it to be discontinued due to an apparent increase in cardiovascular morbidity and mortality in patients when compared with those receiving insulin (Galloway *et al.*, 1992).

It is safe to say that despite many decades of experience and improvements in the quality of therapeutic insulin as well as the delivery devices and formulations available, serious problems to conventional insulin therapy remain (Galloway and Chance, 1994; Bristow, 1993).

1.4 Pursuing a fully active monomeric insulin

The production of a fully active monomeric insulin is still being pursued. Apart from the therapeutic advantages summarised above, monomeric insulins provide an important new tool for biochemical and ^{1}H-NMR (nuclear magnetic resonance) studies of insulin structure and dynamics in solution. Such studies are limited by aggregation, a phenomenon which leads to line-broadening in the physiological pH range. To analyse the solution structure of insulin NMR studies of insulin analogues that are mainly monomeric under physiological conditions have been undertaken (Shoelson *et al.*, 1992; Weiss *et al.*, 1991).

1.4.1 Design of an active monomeric insulin

For the design of an active monomeric insulin the following need to be considered:

- The three dimensional structure of insulin and its relationship to function
- The three interaction surfaces of insulin as independent targets for design
- The incorporation of reduced self-association properties of naturally occurring insulins, insulin-like molecules and insulin analogues
- The method of evaluation of the monomeric status of insulin analogues in solution

The three dimensional structure of insulin and its relationship to function

To study the structure-function relationship of insulin detailed knowledge of the three dimensional structure and an understanding of how it relates to function and molecular interactions is needed. This knowledge may also be used in the design of insulins with desired characteristics of which the active monomeric insulin would be most beneficial in the treatment of diabetes. A database of the three dimensional structures of many insulins and analogues have been compiled. Three dimensional structures can be obtained by two means, namely X-ray crystallography and in solution using multi and two dimensional nuclear magnetic resonance (2D NMR). The crystal structures of insulins are subjected to crystal packing forces which probably do not reflect the conformation that it acquires in solution as a monomer at physiological concentrations. On the other hand in NMR experiments, aggregation of insulin at concentrations where these studies need to be undertaken impairs the acquisition of detailed information. Improvements in the detail of the structures have been obtained with designed insulin analogues (Chapter 2) that are predominantly monomeric in the buffers used.

What is apparent from this large database of three dimensional structures is that insulin is conformationally flexible and exists in different conformations. The flexible segments of the molecule could easily be displaced upon insulin receptor binding (Badger, 1992) and it is believed that a conformational change takes place upon binding of insulin to its receptor and that the receptor induces or assists in this change. How insulin works in detail as a hormone at the level of molecular interactions is still not clearly understood. This may become clear when the co-crystal structure of insulin bound to its receptor or the ectodomain of the receptor is solved. In the meantime any information bearing on the conformation assumed by the monomer as it binds to its receptor would be helpful.

The three interaction surfaces of insulin as independent targets for design

It seems that the properties of biological competence and aggregation are so deeply embedded in the insulin structure that it is a formidable challenge to alter the one without the other. Despite this there have been analogues developed, based on the sequence of human insulin, where substitutions in discrete positions on the surface of insulin have seemingly independent effects on its three dimensional structure and function. The functions of concern are the two self-association surfaces and the receptor recognition surface (Shoelson *et al.*, 1992). An example is the DKP-human insulin (B10 Asp; B28 Lys; B29 Pro) with substitutions in both the interaction surfaces (B10 Asp in the dimer-dimer interface and B28 Lys and B29 Pro in the monomer-monomer interface). This insulin analogue has reduced tendencies to aggregate but has an enhanced affinity for the insulin receptor (Weiss *et al.*, 1991). A comparison of a family of DKP analogues having potency-altering substitutions at the B24 and B25 positions with similar analogues of human native insulin which exhibit parallel trends in receptor binding potency irrespective of which insulin template was studied (Shoelson *et al.*, 1992).

The incorporation of reduced self-association properties of naturally occurring insulins, insulin-like molecules and insulin analogues

Attempts have been made to prepare insulin analogues for potential therapeutic use with amino-acid substitutions at positions in the two self-association interfaces which exhibit reduced or lack of ability to form polymers. Examples of approaches that have been followed in these attempts are:

- Charge repulsion, for example by insertion of opposing negative charges at the interface, and steric hindrance, by insertion of bulky amino-acid side chains (Brange *et al.*, 1988; Brange *et al.*, 1990). In general, charge repulsion has proven to be the more successful of these two approaches.
- Deleting the hydrophobic B26-B30 region (Kubiak and Cowburn, 1986) (C-terminal of the B-chain). A des-(B26-B30)-human insulin shows reduced self-association properties (Jeffrey, 1986).

- Using the reduced self-association properties of insulins from animals like the hagfish (Schwartz *et al.*, 1987; Peterson *et al.*, 1975; Peterson *et al.*, 1974). Hagfish insulin has an Asp in the B10 position and does not form hexamers and only exists as dimers in solution and also crystallises as a dimer. (His in that position is essential for hexemerisation.) A synthesised HB10D human insulin with Asp instead of His in this position did in fact show these reduced self-association properties (Weiss *et al.*, 1991).
- Using the reduced self-association properties of an insulin-like molecule, namely the insulin-like growth factor-1 (IGF-1). IGF-1 (with about 50 % structural homology with insulin) has at the carboxy-terminus of its B-chain, lysine at position 28 and proline at position 29 (Rinderknecht and Humbel, 1978). It does not self-associate as strongly as native human insulin. Using computational chemistry it was possible to ascertain that this modification of insulin would self-associate poorly (Shields, 1993). The KP (B28 Lys; B29 Pro) human insulin that has been produced and shown to have reduced self-association properties (Weiss *et al.*, 1991) is fast acting (Galloway and Chance, 1994).

Evaluation of monomeric status of insulin analogues in solution

It is crucial that a reliable and reproducible method for measuring self-association behaviour be used to monitor molecular changes designed to modify this behaviour. It is argued here that sedimentation analysis in the analytical ultracentrifuge is still the best technique for characterising protein interactions (Jeffrey, 1990). The self-association pattern of insulin analogues can be quantitatively compared by using a model of self-association that best describes the system and by calculating association constants that describe the relative strengths of interaction at the two association interfaces. In this study the self-association pattern of synthesised platypus insulin and a mutant has been analysed by making use of a well established background of theory for analysing self-associating systems and of experimental studies of insulin with the analytical ultracentrifuge.

There are cases where methods used to assess the changes in self-association of insulin analogues have been relatively insensitive and this is illustrated where osmometry by the NOVO group was used (Brange *et al.*, 1988). Closer examination of this work led to the conclusion that although good progress has been made in increasing the amount of monomeric species at concentrations used for insulin formulations, the monomer was not the major molecular species at the chosen concentration (Mark and Jeffrey, 1990; Jeffrey, 1990). This emphasises the need for critical quantitative study of changes to self-association as well as to potency.

1.5 Exploring another avenue for identifying an insulin with desirable properties

The purpose of the present undertaking was to explore another avenue for identifying an insulin with desirable properties as a template, namely to exploit natural variation as exhibited in an exotic species like the platypus. Modifications of self-association behaviour without loss of activity may succeed in one of the more exotic insulins because its different composition and properties offer a different baseline for modification from the common mammalian insulins.

1.5.1 Using the platypus insulin sequence to explore and identify desirable properties

The platypus (*Ornithorhynchus anatinus*) is the only extant species of the family Ornithorhynchidae belonging to the order Monotremata. It is an ancient, highly specialised animal well adapted to its amphibious existence in the waterways of eastern Australia (Grant, 1989). It, like mammals, has mammary glands but lays eggs. The monotreme eggs are very similar in appearance to reptile eggs, with a soft and rubbery shell (Griffiths, 1988). Its biological features indicate that it is more closely related to marsupial and placental mammals than to any group of reptiles, despite retaining several ancestral reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988).

The availability of a small quantity of platypus pancreatic tissue and the subsequent purification of enough native insulin (ng) to determine its amino-acid sequence gave us the opportunity to investigate the composition, structure, biological activity and self-association of insulin from this unique animal. This study consists predominantly of three parts; namely:

- the structural analysis of platypus insulin,
- the total chemical synthesis of the active native platypus and a mutant platypus insulin and the determination of their biological activities and
- finally the sedimentation analysis of the self-association pattern of these synthesised insulins.

Structural analysis of platypus insulin

The primary amino-acid sequence of platypus insulin differs from that of human at eleven amino-acid sites; of which three residues have not been observed previously in other organisms. It also differs from pig and human insulin in two other significant aspects. The one being at its self-associating surfaces (monomer-monomer and dimer-dimer interfaces) where it differs by 4 residues; offering the possibility of a change in association properties. The other amino-acid substitutions were at its receptor binding surface where platypus insulin differs from human by 2 residues; here offering the possibility of a change in biological activity.

In chapter 3 the primary sequence of platypus insulin is compared with all insulin sequences in the database. Comparative studies include pairwise sequence analysis as well as multiple sequence analysis. A three dimensional model of platypus insulin based on the sequence was built with computer programmes. Predictions of the physico-chemical and biological properties of platypus insulin, namely its self-association , receptor binding and activity, were tentatively made on the basis of the sequence as well as the model of the three dimensional structure.

Total chemical synthesis of native platypus and a mutant platypus insulin

To test predictions about the structure-function relationship of platypus insulin, it was necessary to obtain adequate material by producing the insulin by either DNA technology or by total chemical synthesis. The latter was chosen because of the expertise available in the laboratory and because this route had the added advantage of being able to produce insulins with non-proteinogenic residues. Chapter 4 describes the strategy for the chemical synthesis of platypus insulin.

The total chemical synthesis was successful and preliminary investigations into the self-association of synthetic platypus insulin pointed to some reduction in the self-association at the dimer-dimer interface compared to bovine insulin (Chapter 5). Furthermore predicted differences in the spatial arrangement of residues in the dimer-dimer interface of the three-dimensional molecular model (Chapter 3) also suggested that the self-associating properties in this region may be affected, seemingly lowered. It was also decided to synthesise a platypus insulin mutant (B10 His replaced by an Asp) with a substitution in the hexamer forming region (dimer-dimer interface) and analyse its self-association pattern. The reason was to obtain an insulin with reduced self-association properties and increased biological activity as was found with a similar HB10D human insulin (Schwartz *et al.*, 1987). The effect of this mutation on the self-association pattern of platypus insulin was analysed and investigated in detail.

Both synthetic insulins were obtained in a pure form and biological activity and the circular dichroism spectra were determined.

Sedimentation analysis of the self-association pattern of chemically synthesised insulins

An extensive study of the self-association patterns of zinc-free synthetic native platypus and mutant (HB10D) platypus insulins in solution, at room temperature and neutral pH, has been undertaken and is described in Chapter 5. The sedimentation equilibrium method was chosen because of the accuracy of the data that could be

obtained. The omega-method of analysis was used to obtain the thermodynamic parameter a_1 (the activity of the insulin monomer in equilibrium with a series of polymeric species) as a function of total insulin concentration.

Once the sedimentation equilibrium data was analysed and transformed to concentration of the monomer m_1 as a function of total protein concentration it was fitted to a mathematical equation (Mark and Jeffrey, 1990) describing a model of self-association previously shown to be the best description of the complex insulin system. From this the relevant association constants describing the two pathways of formation of insulin dimers and higher polymeric species were derived. This detailed analysis of changes to self-association brought about by specific residue changes allows comparison of self-association in different insulins in quantitative terms and the quantitative linkage of self-association changes to changes in specific structural loci.

Chapter 2

Literature Review: The Structure-Function Relationship of Insulin

2.1 Introduction

The importance of the hormone, insulin, has made it the most intensively investigated protein in human biology and medicine. It was also the first protein to be sequenced in full (Sanger *et al.*, 1959). One aspect on which attention has been focused is its structure, with the ultimate goal being defining the role of its conformation in relation to its function. An important step toward this goal was the determination of the three-dimensional crystal structure of native 2 Zn porcine insulin (Blundell *et al.*, 1972; Baker *et al.*, 1988). Subsequently, several other crystal structures of insulin have been reported, as for example 4 Zn insulin, hagfish and des-(B26-B30) insulins (Bentley *et al.*, 1976; Cutfield *et al.*, 1979a; Cutfield *et al.*, 1979b; Baker *et al.*, 1988; Derewenda *et al.*, 1989) and conformation variations have been observed especially in the so-called flexible regions. The flexibility, combined with the rigid structure of the rest of the molecule, is most likely to be of importance for the expression of insulin activity (Baker *et al.*, 1988 and Blundell 1972).

A detailed understanding of the many facets of insulin structure and flexibility in solution is of fundamental importance to the understanding of its activity. Binding affinities and conformations in solution are among the most important properties of peptides that determine activity and ultimately, potency (Smith *et al.*, 1991). An understanding of these properties at a fundamental molecular level will ultimately enhance future ability to design therapeutic insulins with desirable properties.

The biological activity of insulin arises from the ligand-receptor complex and is in equilibrium with free insulin. This equilibrium occurs at low insulin concentrations

(Schäffer, 1994; Sonne, 1986). Consequently, an understanding of the biological activity of insulin may best be achieved by

- determining its physico-chemical properties in solution and
- knowledge of the three-dimensional conformation of insulin in solution as it binds to the receptor

Also knowledge of the multiple conformations in dynamic exchange in solution at higher concentrations (as needed in insulin therapy) may lead to an understanding of its self-association properties.

The crystal structure has been used as a basis to predict the tertiary fold of insulin in solution. This has been carried out in a number of ways namely, the analysis of the crystal packing of insulin (Chothia *et al.*, 1983), by energy minimisation calculation (Wodak *et al.*, 1985), and by molecular dynamic computer simulations (Kruger *et al.*, 1987; Mark *et al.*, 1991). An experimental determination of the three-dimensional solution structure of proteins became possible with the new developments in the field of nuclear magnetic resonance (NMR) (Wüthrich, 1986).

Since knowledge of the crystal structure of 2 Zn insulin has been of paramount importance in understanding the relationship between the structure and function of insulin as well as to the work to be described here, an outline of this structure is presented. This may be used as a basis for the study of the structure of insulin in solution and how that relates to its biological action as well as its self-association properties.

The properties of the residues that contribute to the physico-chemical characteristics and biological activity have the following indications. Firstly is the significance of the contribution of specific amino-acid residues to the properties of the three interfaces of the insulin molecule. These interfaces are the regions of self-interaction (monomer-monomer-interface and dimer-dimer interface) and the insulin region of receptor interaction. Some amino-acid residues are common to both the monomer-monomer interface and the receptor binding region (Shoelson *et al.*, 1992). Another indication is the self-association

behaviour of insulin at the concentrations used in both the therapeutic management of diabetes and in physico-chemical studies where in each case it is difficult to mimic the concentrations applicable in the physiological action of insulin (Mark *et al.*, 1990).

Also presented in this literature review is a description of the crystal and related solution structures of the mutant HB10D human insulin. This information is of importance since in the work presented here (Chapter 4) a similar mutant, HB10D platypus insulin, was chemically synthesised and the structure-function relationship investigated. Special attention has been given to its self-association properties.

Also knowledge of the, as yet unknown, three-dimensional structure of the insulin receptor and insulin-bound receptor are fundamental to an understanding of the biological action of insulin. The most recent information in respect to the structure of the insulin receptor as well as the most recent model of binding to and activation of receptor are discussed briefly.

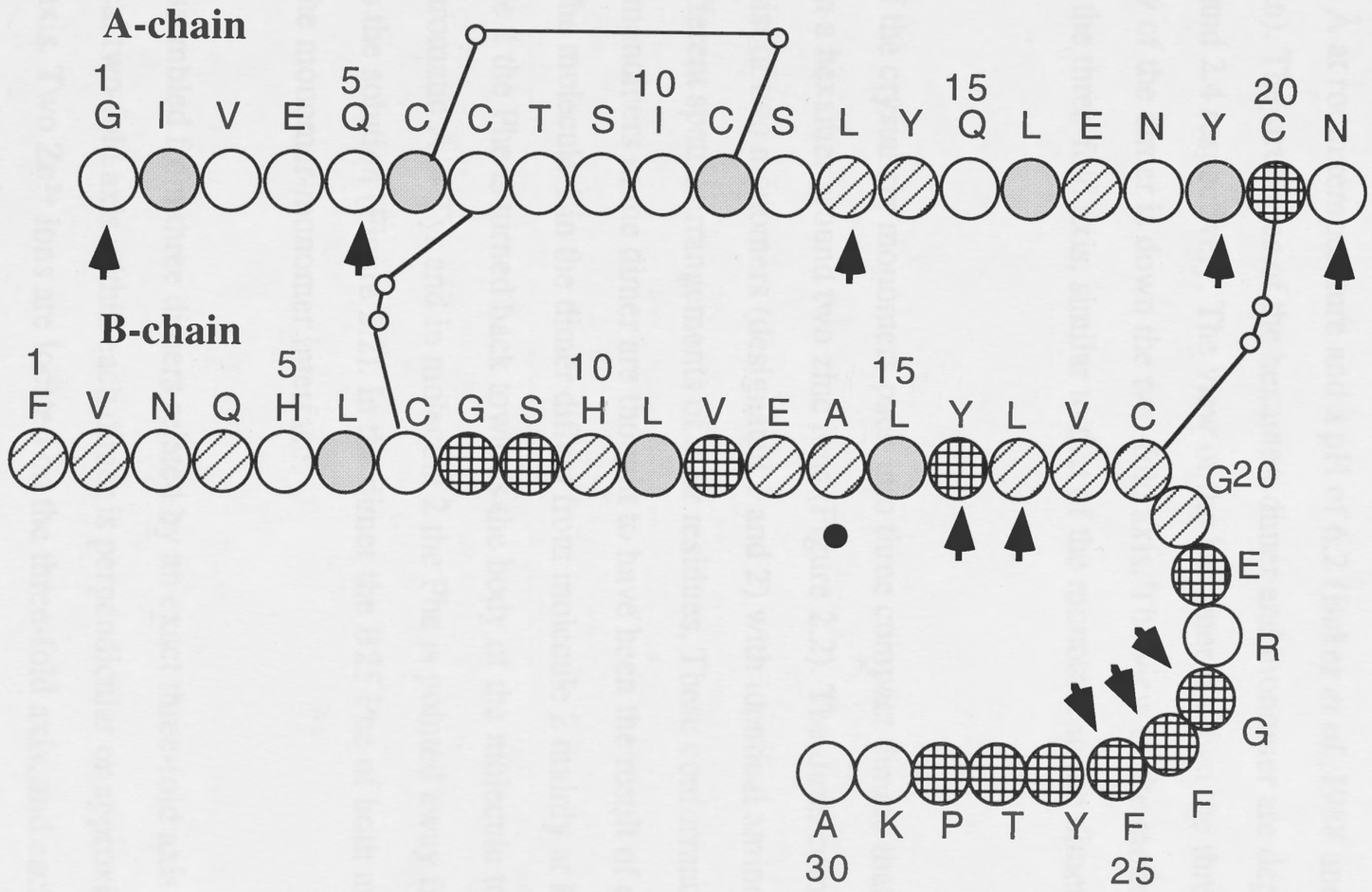
2.2 The structure of insulin

The primary amino-acid sequence of insulin (porcine), with the residues at specific positions responsible for the characteristics and properties of insulin, is presented in Figure 2.1. The insulin molecule consists of two chains (A and B) connected by two inter disulphide bridges (A7Cys-B7Cys and A20Cys-B19Cys). There is also an intra disulphide bridge (A6Cys-A11Cys). The current image of the three-dimensional structure of insulin is primarily based on X-ray crystallographic data. Flexible regions based on relatively high thermal parameters have been observed in the N-terminal of the A-chain (A1-A6), the N-terminal (B1-B4) and the C-terminal of the B-chain (B27-B30). (The thermal parameter is a quantitative estimate of the mean square displacement of each atom from its mean position (Baker *et al.*, 1988).) Less is known of the structure of insulin in solution. This is primarily because native insulin, close to its physiological pH and at

Figure 2.1

The amino-acid sequence of porcine insulin. (Human insulin differ at B30 Thr.) The insulin molecule consists of two chains, called A and B connected by two inter disulphide bridges (A7Cys-B7Cys and A20Cys-B19Cys). There is also an intra disulphide bridge (A6Cys-A11Cys). Residues (speckled) that are found in the inner core of the molecule; residues (cross-hatched) that contribute to the monomer-monomer interface; residues (diagonal line shading) that contribute to the dimer-dimer interface; the residues indicated by arrows contribute to the receptor binding region (many residues share properties with the monomer-monomer interface) and B14 Ala (black dot) that are found in the inner core and contribute to the dimer-dimer interface.

Figure 2.1


A-chain
1 5 10 15 20
G I V E Q C C T S I C S L Y Q L E N Y C N
B-chain
1 5 10 15
F V N Q H L C G S H L V E A L Y L V C
G20
E
R
G
F
A K P T Y F
30 25

concentrations normally required for physico-chemical studies, strongly self-associates in solution to form a complex mixture of polymeric species (Mark *et al.*, 1987).

2.2.1 The crystal structure of 2 Zn pig insulin

The crystal structure of insulin described in this section will be that of the rhombohedral crystals of 2 Zn pig insulin as seen in electron density maps calculated from X-ray data extending to 1.5 Å at room temperature and a pH of 6.2 (Baker *et al.*, 1988 and references therein). The structures of the hexamer, dimer and monomer are depicted in Figures 2.2, 2.3 and 2.4 respectively. The view of the hexamer is down the three-fold axis and the view of the dimer is down the two-fold axis. The view of the monomer is perpendicular to the three-fold axis, similar to that of the monomer in the dimer in Figure 2.3.

In the unit cell of the crystal six monomers pack into three compact dimers that further aggregate to form a hexamer around two zinc ions (Figure 2.2). The three compact dimers each consist of two monomers (designated 1 and 2) with identical amino-acid sequences but different spatial arrangements of their residues. These conformational variations of the monomers in the dimer are thought to have been the result of crystal packing forces. The molecule 1 in the dimer differs from molecule 2 mainly at B25 Phe where in molecule 1 the Phe is turned back towards the body of the molecule to make contact with the aromatic A19 Tyr and in molecule 2 the Phe is pointed away from the molecule towards the solution (Figure 2.2). In the dimer the B25 Phe of both monomers make contact in the monomer-monomer interface.

The hexamer is assembled from three dimers related by an exact three-fold axis of symmetry, the local two-fold axis within each dimer is perpendicular or approximately so to the three-fold axis. Two Zn^{2+} ions are located on the three-fold axis, and each is coordinated by three symmetry-related B10 His residues from separate monomeric units.

Figure 2.2

Hexamer: A view down the three-fold axis of the 2 Zn crystal structure of porcine insulin. The three dimers in yellow, orange and green are assembled in a way that brings the B10 His residues (pink) together to coordinate 2 Zn^{2+} ions (not shown) in the centre of this column. The model of the hexamer was based on the coordinates of X-ray crystal structure of rhombohedral 2 Zn porcine insulin, entry 3 ins (Wlodawer *et al.*, 1989) of the Brookhaven Protein Databank.

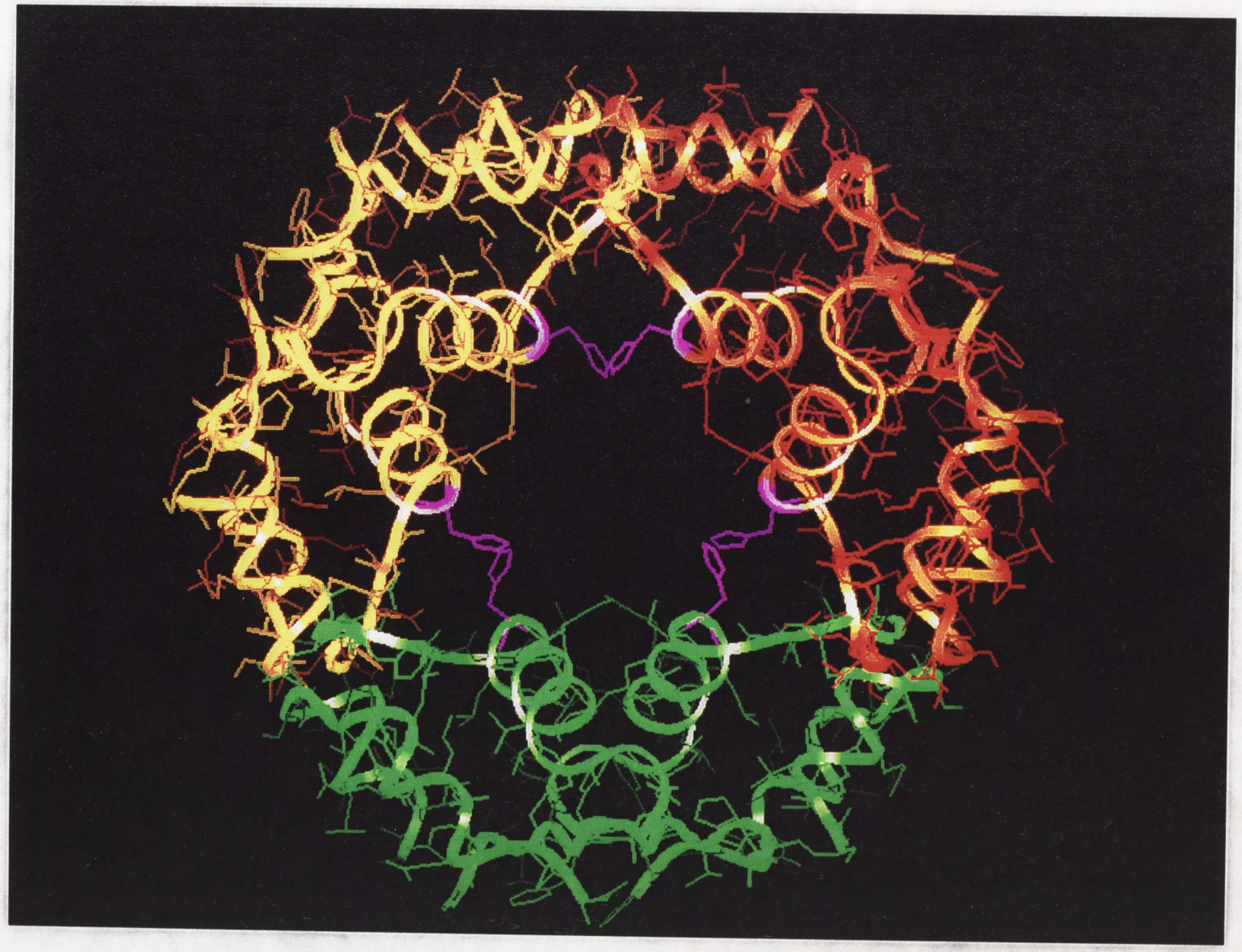


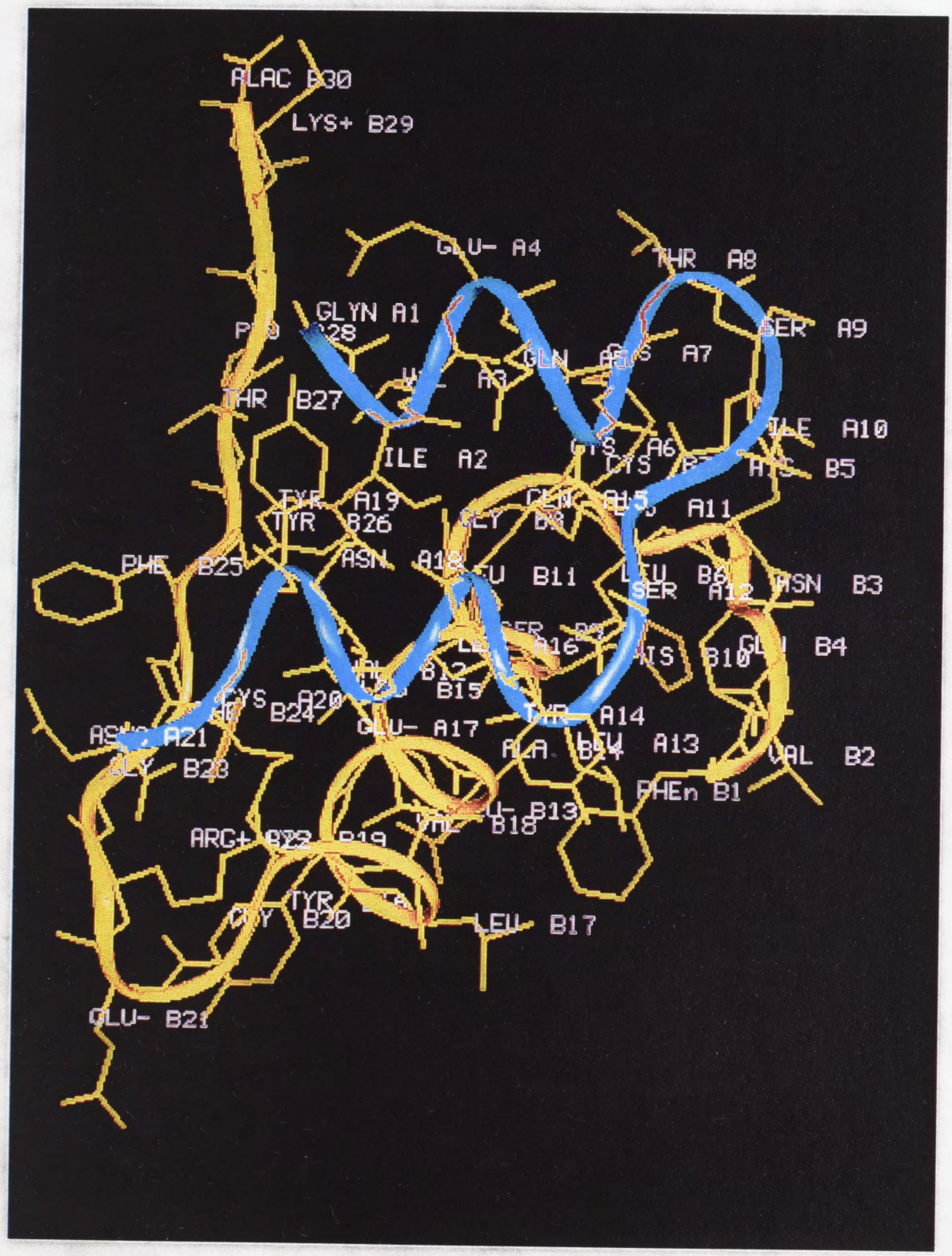
Figure 2.3

Dimer: A view down the two-fold axis of the 2 Zn crystal structure of porcine insulin depicting the monomer-monomer interface. Molecule 1 has an A-chain in pink and B-chain in white; molecule 2 has an A-chain in yellow and a B-chain in blue. The orientation of the B25.1 and B25.2 Phe are depicted in orange. The disulphide bonds are in red. The aromatic residues that line the monomer-monomer interface are illustrated. The model of the dimer was based on the coordinates of X-ray crystal structure of rhombohedral 2 Zn porcine insulin, entry 3 ins (Wlodawer *et al.*, 1989) of the Brookhaven Protein Databank.


B1
B25
B25
B1

Figure 2.4

Monomer: A view perpendicular to the three-fold axis of the 2 Zn crystal structure of porcine insulin. Depicted is molecule 2 with B25 Phe in an outward orientation. The A-chain is in yellow and the B-chain in blue. The model of the monomer was based on the coordinates of X-ray crystal structure of rhombohedral 2 Zn porcine insulin, entry 3 ins (Wlodawer *et al.*, 1989) of the Brookhaven Protein Databank.



ALAC B30
LYS+ B29
GLU- A4
THR A8
GLYN A1
SER A9
A7
GLN A5
A3
THR B27
ILE A10
ILE A2
CYS A6
B5
TYR A19
TYR B26
A11
PHE B25
ASN A18
B11
ASN B3
SER A12
B4
A16
B10
B15
CYS B24
A20
GLU- A17
TYR A14
A21
ALA B14
A13
GLY B23
VAL B2
PHEn B1
B13
B18
ARG+ B22
B19
TYR B20
LEU B17
GLU- B21

The three-fold axis of symmetry of the hexamer (Figure 2.2) is a view of the hexamer down the Zn coordinated cavity (column) lined by the B10 and B5 histidines.

Non-crystallographic lines of approximately two-fold axes of symmetry can be traced between the monomers. The two-fold axes are labelled OP (along the monomer-monomer interface) and OQ (along the dimer-dimer interface) (Blundell *et al.*, 1972). In the crystal each monomer is in direct contact with five monomers of the alternative variety, i.e. 1 is closely surrounded by molecules of type 2 and *vice versa.*

The conformation of A-chain in the crystal structure

The A-chain consists of two irregular α-helical regions, namely the A2-A8 and the A12-A19 regions that run very nearly antiparallel and are joined by a stretch of extended polypeptide chain from A9 to A12. The view in Figure 2.4 of the monomer shows that the A-chain is folded into a plane in which the N- and C-terminal residues are brought together, Thus the A19 Tyr side chain and the A2 Ile side chain are at van der Waal's contact. The A2-A8, α-helical region, contains the A7 Cys that forms a disulphide bridge with the B7 Cys contained in a α-helical region on the B-chain that runs antiparallel with the A2-A8 α-helix. There is a marked difference between the A-chain projections in molecules 1 and 2. This difference is largely because of a rotation of 32° about (ϕ) of residue A6 Cys in molecule 2 and this in turn forces the rotation of A19 Tyr and the displacement of B25 Phe relative to molecule 1. This rotation in molecule 1 is largely responsible for destroying the exact two-fold symmetry. In both molecules residues A2-A8 follow somewhat distorted helical conformations. Also the N-terminal (A1-A6) region is considered to be mobile based on its relatively high thermal parameters (Baker *et al.*, 1988).

The conformation of the B-chain in the crystal structure

The structure of the B-chain consists of three segments, an N-terminal extended β-chain (B1-B8), a central α-helix and a C-terminal extended β–chain (B21-B30) (Figure 2.4). The initial B-chain residues are free of interchain links and can move away from the body of the molecule. Different conformations are apparent in different insulin crystal structures (Cutfield *et al.*, 1981). The central α-helix (B9-B19) packs in a concave fashion towards the rest of the molecule. At each end of the helix (at B7 and B19) there is a cystine linking it to the A-chain. The C-terminal-chain extends back in the reverse direction at B20 Gly to B21 Glu . This sharp turn is made possible by the conformation of the B20 Gly which is that of the D-amino-acid (Baker et al, 1988). Multiple intramolecular bonds maintain the insulin fold in this region. Examples of these bonds are H-bonds between the peptide carbonyl and NH groups of B20 Gly and B23 Gly; B23 Gly and A21 Asn and A19 Tyr and B25 Phe as well as the ionic interaction between B22 Arg and the A21 C-terminal (Baker *et al.*, 1988). There are also intramolecular contacts (H-bonds) that allow insulin to dimerise and some are between B24 Tyr and B26 Tyr where the chain is involved in an antiparallel H-bond with its two-fold related equivalents in the dimer (Figure 2.3). The C-terminal-chain (residues B27-B30) is mobile in the monomer as can be seen by the latter's relatively high thermal parameters in the crystal. Also relatively high thermal parameters have been observed for the N-terminal-chain (residues B1-B4), indicating their mobility (Baker *et al.*, 1988).

The disulfide bonds

The insulin monomer is a hetero dimer consisting of an A-chain and a B-chain coupled covalently by two inter chain disulfides bridges (A20-B19 and A7-B7) and one intra disulfide bridge (A6-A11) (Figure 2.4). Of the three disulfide bonds in each molecule, A6-A11 is completely buried, A20-B19 is largely buried, whereas A7-B7 is exposed; both sulfur atoms are in contact, rather differently, with water molecules (Baker *et al.*, 1988).

At each end of the central α-helix (B9-B19) at B7 and B19 there is a cystine linking it to the A-chain. The positions of all the other cystine residues (A6, A7, A11 and A20) are also close to the ends of the α-helices (A2-A8 and A12-A19 in the A-chain). The disulfide bridges probably contribute to the rigidity and stability of the core of the molecule. The precise relative positions of the A-and B-chains are strengthened at a number of points by hydrogen bonds and ionised contacts (Baker *et al.*, 1988).

The conformation of the monomer in the crystal

The inner core of the insulin molecule is predominantly comprised of wholly buried non-polar residues that form a sphere with a radius of 6 Å. These residues include the invariant leucine residues A16, B6, B11 and B15, the invariant cystine A6-A11, the A19 Tyr ring, also A2 Ile and B14 Ala. The atoms of this inner core make contact with other non-polar residues, A3 Val, A13 Leu, B12 and B18 Val, B24 Phe and B19 Cys, which in turn are in contact with most of the remaining non-polar residues (See Figures 2.1 and 2.4). Since members of the last two groups are all partly on the outside of the molecule and make contact with other non-polar residues in the monomer-monomer and dimer-dimer interfaces, there is in fact an irregular belt of continuous non-polar character around the three-fold axis within the hexamer. It is notable that none of these buried residues are aromatic; in marked contrast with the situation in the aggregated surfaces (Baker *et al.*, 1988).

The formation of the dimer in the crystal

The C-terminal end of the B-chain is mainly involved in the dimer formation (monomer-monomer interface). The buried surface of the monomers in the dimer is flat, studded with aromatic and aliphatic residues. In the dimer, as found in the crystal, the dissimilar monomers 1 and 2, are rotated relative to the two-fold axis (monomer-monomer interface) (Figure 2.3). The respective extended β-chains (N-terminal B-chains) of the

monomers are packed face to face. The relative positions of the extended β-chains in the dimer surface are secured by an antiparallel β-sheet of hydrogen bonds involving the B24 Phe and B26 Tyr residues. The forces leading to aggregation are predominantly non-polar (Jeffrey and Coates, 1966a), reinforced and given direction by hydrogen bonds. All but one of the contacts involve B-chain residues. The packing of the aromatic rings in this interface is of interest. The six residues, two B16 Tyr's, two B24 Phe's and two B26 Tyr's are linked together in a network of loose contacts 4-4.5 Å long. A second smaller group of three aromatic residues is formed by the two B25 Phe of which B25.1 (see notation, p ix, this thesis) is in contact with A19.1 Tyr. The B25.2 Phe side chain lies across the two-fold axis in contact with B25.1 (Figure 2.2 and 2.3). The B12.2 Val residue fills in the cavity formed at the centre of the aromatic cluster.

The formation of the hexamer in the crystal

In the 2 Zn crystals three dimers are assembled around two central zinc ions on the three-fold-axis each octahedrally coordinated to three symmetry-related N3-atoms of B10 His's in molecules 1 and 2 (Figure 2.2). The two zincs are related by the same two-fold symmetry that relates the dimers. The centre of the hexamer is surrounded by six glutamic acid B13 residues brought together in a circle. They are arranged in pairs, the two within each dimer in direct contact. Both pairs of carboxylic oxygen ions are hydrogen bonded to well-defined water molecules in the central cavity, two of which are linked to B10 His. There is a polar channel at the centre of the hexamer made up of coordinating B-chain histidines, B9 Ser and B13 Glu. Associated with these groups is a well defined, but complex water structure which links the hydrogen bonding atoms in a elaborate chain of the structure. The structure of the six central glutamic acids is arranged as three hydrogen bonded pairs and suggests that the charge repulsion that would be normally operating is avoided by protonation (Baker *et al.*, 1988; Blundell *et al.*, 1972).

When the hexamer is assembled through the zinc coordination, both polar and non-polar residues are buried between the dimers. The packing in the dimer-dimer interface is correspondingly much looser than in the molecular interface within each dimer (monomer-monomer interface). The initial residue of each B-chain, B1 Phe, from each monomer is buried in a non-polar environment (A13 Leu, A14 Tyr) in the adjacent dimer. This crossover lends extra stability to the dimer. Between the B-chain N-terminal extended β-chains are buried non-polar residues whose contacts complete a hydrophobic circle which connects up the non-polar cores of the monomers already in touch within the dimer (Baker *et al.*, 1988). In addition, the association is also strengthened by ionic attraction between the N-terminal B1 amino group and the A17 Glu carboxyl ion (Baker *et al.*, 1988).

2.2.2 Conformational states of insulin hexamers in crystals

The zinc-insulin hexamer is the storage form of insulin in the pancreas. There are three distinct conformational states observed in crystals, designated T_6, $(TR)_3$ and R_6. These are based on their ligand binding properties, allosteric behaviour and pseudo point symmetries (Wollmer *et al.*, 1987; Kaarsholm *et al.*, 1989; Choi *et al.*, 1993). The T_6 state is associated with the 2 Zn insulins; the $(TR)_3$ with the 4 Zn insulins and R_6 with the phenol hexamers. In the T-state, residues 1-9 of the B-chain are in an extended β-conformation (as in the 2 Zn porcine insulin crystals) and in the R-state in an α-helix , with the B-chain α-helix now encompassing residues 1-19. The T- to R-state conformational transition comes about in the presence of phenol, phenol analogues and lyotropic anions. The $(TR)_3$ hexamer is composed of asymmetric dimers wherein each dimer contains one insulin molecule in the T-conformation with the other in the R-conformation (Brzovic *et al.*, 1994). It is also called the 4 Zn insulin, which is grown from high chloride ion concentrations (Smith *et al.*, 1984).

Clearly the contacts between the dimers in these three hexamer types are very different. The T_6 hexamer binds 2 zincs and is that of the 2 Zn insulin crystal structure where the

hexamer constitutes the rhombohedral unit cell and the dimer the asymmetric unit as discussed in the previous section (Derewenda *et al.*, 1990).

The $(TR)_3$ hexamer binds 4 zincs at high chloride ion concentrations. A rearrangement of contacts in the hexamer are generated to accommodate the binding of these ions and thus stabilising the structure. New compensating dimer-dimer contacts in the hexamer are generated by the extended β-conformation-α-helix transition that takes place only in monomer 2 at the N-terminal B1-B8 which now form a continued helix B1-B19. Extra zinc sites are created as a result in molecule 2 by the B5 His positioned on the helix and the B10 His in the adjacent dimer which swivel away from the central three-fold hexamer axis thus creating a tetrahedral binding site. The tetrahedral binding site consists of the two histidyl N atoms (B5 and B10), a chloride ion and a water molecule. Also the packing of the leucines in the R-molecule buries the zinc site, thus significantly increasing the stability of the hexamer. In this hexamer $(TR)_3$ there are three tetrahedral sites and one octahedral axial site that coordinates the 4 zincs compared to T_6 that has two octahedral axilla sites binding 2 zincs. The zinc binding site in monomer 1 is unchanged (Derewenda *et al.*, 1990).

The R_6 hexamer structure binds two zincs and in the presence of phenol to form monoclinic crystals. The presence of phenol causes an extension of the helical structure in the N-terminal B-chain to the B1-B19 α-helix in all the monomers. These phenol molecules bind in a non-polar cavity created by the packing of the B1-B8 helix against the A-chain of the three-folded related other dimer. The R_6 hexamer is also stabilised by these contacts that are made by these six phenol molecules that are specifically bound between the dimers (Derewenda *et al.*, 1990). The phenol makes two hydrogen bonds to the main-chain groups (one to the A6 carbonyl oxygen , the other to the peptide nitrogen of A11) and has displaced water molecules while some close non-polar contacts are enhanced (Derewenda *et al.*, 1989). The aromatic ring of the phenol packs at van der Waal's distance against the B5 His imidazole group. Since all the monomers are helical at B1-B8 the contacts between the dimers are equivalent. It has the same organisation of its

symmetry elements as the rhombohedral 2 Zn insulin hexamer. The contact between the B1-B19 helices that are made by B6 Leu along the axis are strengthened and both zinc sites are buried (Dodson *et al.*, 1993). All the monomers show increased separation at B25-A19 and show variations in the conformation of the B25 Phe (Derewenda *et al.*, 1990).

2.2.3 The structure of HB10D insulin in the crystal

This mutant crystallised into dimers in the absence of zinc. Analysis shows that the B10 Asp and B13 Glu carboxylate groups, both negatively charged, are directed away from each other (Dodson *et al.*, 1993). The mutation HB10D in insulin abolishes the capacity of the molecule to hexamerise in the presence of zinc. Zinc ions are important for hexamer formation by native insulin and in this mutant the zinc-binding residue His B10 is substituted for an Asp, therefore altering hexamer assembly. In the presence of zinc, however, this mutant forms a dodecamer in which two sets of three dimers are coordinated to two zinc ions on the central three-fold axis (Dodson *et al.*, 1993). The zinc coordination is through the B5 His imidazole ring and is reminiscent of the tetrahedral coordination that is seen in the R_6 insulin hexamer. The monomers in the structure have the T-conformation and the dimer-dimer contacts include many of the same residues that are used in the T_6 insulin hexamer. The contacts are, however, very different. In particular, a critical salt bridge is made by the mutated B10 Asp to the B-chain α-NH_2 group. Residues B10 Asp and B13 Glu, due to their negative charge, repel each other and do not form the polar channel at the centre of hexamer as does B10 His and B13 Glu in native insulin. The B13 Glu side chains are not packed together, in contrast with the native insulin hexamers, but are hydrogen bonded to the A9 Ser in one dimer interaction and to the A16 Tyr in the other. Another essential interaction is between the two B4 Gln residues, which form three pairs of the complementary hydrogen bonds about the three-fold axis (Dodson *et al.*, 1993).

2.3 The solution structure of insulin: Relationship to crystal structure

The current image of the structure of insulin is one primarily based on X-ray crystallographic data and little is known in regard to the solution structure of the most probable active form of the hormone. (Mark *et al.*, 1991). There are several well defined crystal structures of insulin in different conformations available, and some may be interconvertible (Hua *et al.*, 1991 and references therein). The information provided by these can give clues as to the structure of insulin in solution and from these a picture of the active conformation may emerge. [The structure of insulin in solution is that of multiple of structures in dynamic exchange. The structure that is determined in solution is that of a population of conformations averaged over a specific time period (Smith *et al.*, 1991)]. There is some indication that further changes from the crystalline structure may be necessary for insulin's biological action because it is thought that the biologically active structure(s) is not represented in the series of known structures (See later section) Also it is thought that the monomer is the biologically active species in solution (Baker *et al.*, 1988; Blundell *et al.*, 1972). Thus in trying to understand the biological action and self-association of insulin it is necessary to have knowledge of its conformational variations in solution. This may shed light in structural terms on the sequence of events that lead to optimal binding of insulin to its receptor and in its concomitant action, as well as the nature of its self-association behaviour in solution. This will enable manipulation and modification of the function and behaviour of insulin in a more effective way. Unfortunately no detailed structure of monomeric insulin under physiological solution conditions is as yet available. The reason for this is primarily because native insulin, close to physiological pH and at concentrations normally required for physico-chemical studies, strongly self-associates in solution to form a complex mixture of polymeric species (Mark *et al.*, 1987). In particular this has thus far prevented detailed structural analysis using high-resolution NMR, although changes in the NMR spectrum and circular dichroism spectrum of native insulin at high dilution indicates that there are conformational differences between the monomeric and aggregated forms (Pocker and Biswas 1980; Roy *et al.*, 1990). Some progress has, however, been made using native insulin under extreme

solvent conditions (Roy *et al.*, 1990) and using mutant insulins with altered association properties (Weiss *et al.*, 1991). This will be discussed below. The 2 Zn insulin crystal structure acts as a starting point to an understanding of the action of insulin in solution (see previous section)

Solution structure of insulin monomer

For the study of the structure of insulin in solution two dimensional nuclear magnetic resonance (2D NMR) methods have predominantly been employed. Unfortunately, it is only possible to study the insulin structure at concentrations where insulin aggregates (near its pharmacological concentrations). Also broad NMR signals are obtained with insulin aggregates in contrast to monomers (Weiss *et al.*, 1989). In order to overcome these problems several approaches have been followed:

- Mutant insulins have been designed that are thought to be predominantly monomeric at the desired concentrations (Weiss *et al.*, 1991).
- Solvents (among others 20% acetic acid in water) have been used that are known to render insulins predominantly monomeric at the desired concentrations (Weiss *et al.*, 1989; Hua and Weiss, 1991a; Hua and Weiss, 1991b)
- Simulation of the structure in solution by molecular dynamics (Mark *et al.*, 1991) using known crystal structure coordinates.

Available solution structures of native insulin are not well defined (Hua and Weiss, 1991b) but distinct features of the solution structure of native insulin has been identified with the following indications:

- it is globally flexible,
- the N-terminal B-chain residue B1 Phe has a flexible configuration (Kruger *et al.*, 1987; Hua *et al.*, 1991; Hua and Weiss, 1991b; Weiss *et al.*, 1989),
- the B25 Phe has an outward configuration (molecule 2) and

- A14 Tyr is exposed to solvent (Weiss *et al.*, 1991).

Structure of a mutant insulin that is predominantly monomeric in solution

To obtain a more detailed NMR structure of insulin in solution a mutant DKP (B10 His, B28 Lys and B29 Pro are replaced with Asp, Pro and Lys respectively) human insulin, that was shown to be predominantly monomeric in solution, was used (Weiss *et al.*, 1991). This reduced self-association was according to sedimentation equilibrium studies as well as from indications of 2D NMR spectrum line-widths that were appropriate for monomers (Weiss *et al.*, 1991). This mutant retains native receptor-binding potency. The spectra of DKP-insulin provides a model for the solution structure of insulin monomer and it most closely resembles the hexameric 2 Zn crystal form of porcine insulin with flexibility in the N- and C-terminus of the B-chain.

The 2D NMR study of DKP-insulin shows that the C-terminal region of the B-chain (residues B23-30) remains closely associated with the remainder of the protein (Weiss *et al.*, 1991) and is an indication that B24 and B26 Tyr each pack against the hydrophobic core, as observed in their crystal structures (Blundell *et al.*, 1972; Baker *et al.*, 1988). There is also an indication of flexibility in the B23-30 region. There is no evidence for the existence of the monomer 1. An outward orientation for B25 Phe (molecule 2 of the 2 Zn crystal structure) is suggested that would enable this side chain to interact directly with the insulin receptor. The absence of the B23 Gly-A21 Asn interaction (Baker *et al.*, 1988, see previous section) suggests that the proximity of these two conserved residues, as observed in the crystal structures, is not stably maintained in solution. Fluctuations in the B23-A21 distance have been modelled by molecular dynamics simulations (Kruger *et al.*, 1987; Mark *et al.*, 1991) and reflect both local motions of the A21 side chain and the global changes in the orientation of the C-terminal α-helix of the A-chain relative to the B-chain.

Despite the ingenious experimental approaches outlined above, the structure of the monomeric insulin molecule as it interacts with the receptor is still not defined.

Solution structure of dimers

Detailed knowledge of the contacts between the monomer is essential for the proper understanding of this interaction in solution. The solution structure of the dimer is thus of interest not only for its comparison with the monomer but also because of the continuing interest in the nature of the self-interaction itself. Based on the crystal structure of insulin it is highly likely that the monomer-monomer interaction region designated OP is the site of aggregation of insulin molecules into dimers in solution (Dodson *et al.*, 1983). Also this area of the monomer surface excluded from contact with water by dimer formation is sufficient to provide the favourable entropic increase necessary to drive the assembly process (Chothia and Janin, 1975; Baker *et al.*, 1988). To determine the solution structure of dimers in solution the mutant SB9D human insulin was used as a model. This mutant forms dimers in solution at low pH and its detailed structure was determined by 2D NMR (Jørgensen *et al.*, 1992) and is discussed below. (SB9D insulin had been synthesised by Brange *et al.* (Brange *et al.*, 1988) for the purpose of producing a monomeric insulin at pharmacological concentrations and neutral pH).

The SB9D human insulin dimer was well defined and no detectable differences between the monomers were found. The conformation of both the monomers was that of molecule 1. The association of the two monomers in the solution dimer is relatively loose as compared with that in the crystal dimer (Jørgensen *et al.*, 1992). This was also found in a molecular dynamics simulation study of the insulin dimer in water (Mark *et al.*, 1991), with the contacts in most cases identical to those found by Jørgensen *et al.* (Jørgensen *et al.*, 1992). The overall secondary and tertiary structures of the monomers were similar to the monomers in the 2 Zn porcine insulin crystal. Minor but significant deviations from this structure, as well as from the structure of monomeric insulin in solution, were shown

and ascribed to the absence of the hexamer and crystal packing forces as well as to the presence of monomer-monomer interactions.

Compared with the monomer in the 2 Zn crystal structure these deviations show the following:

- a destabilisation of the A-chain α-helix (A2-A8),
- an extension of the B-chain α-helix (B9-B19) to include B8 Gly,
- a less well defined β-turn in the B-chain (B21-B22) that might indicate a relatively high flexibility of this structural element,
- a bending of the N-terminal end of the B-chain (B1-B4) towards the A-chain loop (Millican and Brems, 1994).

All of these deviations are located at or close to the monomer-monomer interface and are consistent with structural constraints imposed by dimerisation.

Zn insulin hexamer and its structure in solution

The existence of the different conformations of hexamers in the T_6, $(TR)_3$ and R_6 conformation in crystals suggests that insulin is capable of undergoing transitions between the three conformational states. It has been suggested that the three conformational states observed in crystals are indicative of the sequential conformational transitions involved in the T to R conformational change in solution (Brzovic *et al.*, 1994) and that the inter-conversion of T_6 and R_6 occurs via a third species assigned as $(TR)_3$. The transition from the T-state to the R-state involves a coil-to-helix transition in residues B1-B8 producing a continuous α-helix from B1-B19. This conformational transition is also accompanied by small changes in the positions of A-chain residues and other B-chain residues. Anions act as heterotropic effectors that shift the distribution of hexamer conformations in favour of the R-state (Brzovic *et al.*, 1994). These three conformational states in solution have also been suggested by circular dichroism studies

and detailed analysis were performed of the changes in secondary structure associated with the hexamer-to-monomer dissociation (Melberg and Johnson, Jr. 1990).

2.4 Properties of the insulin molecule

There are three identifiable regions that are associated with the physico-chemical characteristics and biological activity on the surface of the insulin molecule (Baker *et al.*, 19888; Blundell *et al.*, 1972). These regions are the interfaces of self-association (monomer-monomer-interface and dimer-dimer interface) and the receptor binding region. The residues that are thought to contribute to these properties (Shoelson *et al.*, 1992) are shown in Table 2.1 and Figure 2.1. There are residues in the monomer-monomer interaction region (in bold) and also in the dimer-dimer interaction region (underlined) that are shared by the residues in the receptor binding region.

2.4.1 Residues involved in the self-association of insulin

The monomer-monomer interface

The interaction between the residues that are thought to contribute to the properties of the monomer-monomer interaction region (shown in Table 2.1 and Figure 2.1) (Shoelson *et al.*, 1992) on the surface of the monomers leads to the formation of the dimer. Many of the side-chain residues that participate in this interaction are situated on the central α-helix (B9-19) (B9 Ser, B12 Val, B16 Tyr) and the C-terminal end of the B-chain (B23-B28). Also the C-terminal residue A21 Asn is included in this group and lies on the monomer surface adjacent to the B23 Gly, the first residue in the extended β-chain of the C-terminal B-chain. (Baker *et al.*, 1988). Upon dimerisation the extended β-chains (B24 Phe, B25 Phe and B26 Tyr) of each monomer form an antiparallel β-sheet that is notable for an extensive set of van der Waal's interactions between adjoining hydrophobic surfaces (Baker *et al.*, 1988).

Table 2.1 Amino-acid residues that contribute to the three interaction regions in the insulin molecule.

Monomer-monomer interface

A21 Asn	B8 Gly	B9 Ser	**B12 Val**	**B16 Tyr**	B20 Gly	B21 Glu
B23 Gly	**B24 Phe**	**B25 Phe**	B26 Tyr	B27 Thr	B28 Pro	

Dimer-dimer interface

<u>A13 Leu</u>	A14 Tyr	A17 Glu	B1 Phe	B2 Val	B4 Gln	B10 His
B13 Glu	B14 Ala	<u>B17 Leu</u>	B18 Val	B19 Cys	B20 Gly	

Receptor binding region

A1 Gly	A5 Gln	<u>A13*Leu</u>	A19 Tyr	**A21 Asn**	**B12 Val**	**B16 Tyr**
<u>B17*Leu</u>	**B23 Gly**	**B24 Phe**	**B25 Phe**			

Residues (**bold**) in the monomer-monomer interface and the receptor binding region

Residues (<u>underlined</u>) in the dimer-dimer interface and the receptor binding region

* Residues that contribute to the second receptor binding site (Schäffer et al., 1994).

The removal of B26-B30 in the despentapeptide insulin abolishes self-association totally (Jeffrey, 1986) and by implication dimerisation. This is because B26 Tyr is essential for hydrogen bonding and the despentapeptide contains approximately a third of the dimerising surface (Baker *et al.*,1988; Spoden *et al.*,1995). Both hydrogen bonding and the non-polar contacts are obviously important in stabilising the dimer. The removal of B26-B30 does not significantly alter the folding of the rest of the molecule in neither the crystal (Bi *et al.*,1984) nor that in solution (Hua *et al.*,1992), nor does it affect the biological activity provided the new C-terminus is amidated (Fisher *et al.*,1985).

Reduced dimerisation was observed in IGF-1 with residues B28 Lys and B29 Pro. This led to the production of the DK human insulin where the residues B28 and B29 are reversed (i.e. B28 Pro to Lys; B29 Lys to Pro) which also displays marked reduction in dimerisation (Weiss *et al.*,1991). In the native insulin B28 Pro is on the monomer surface and is largely buried by dimer formation. The extended β-chain structure also packs the prolyl sidechain against the B26 tyrosil side chain. The Pro in the B29 position most likely introduces a steric property that interferes with this packing and thus reducing the interaction with the other monomer. The structural basis of this effect however is presently not understood (Weiss *et al.*,1991).

The dimer-dimer interface

The interaction between the residues that are thought to contribute to the properties of the dimer-dimer interaction region (shown in Table 2.1 and Figure 2.1) (Shoelson *et al.*,1992) on the surface of the dimers leads to the formation of mainly hexamers in the presence of zinc (Baker *et al.*,1988) at neutral pH and dimers at low pH (Badger, 1992) (Badger 1992). (In a zinc -free solution depending on the total protein concentration, insulin may form an indefinite array of polymeric species interacting along both the monomer-monomer interface and the dimer-dimer interface at neutral pH (Mark *et al.*,1987 and Chapter 5)). These residues are situated on the surface on the opposite side

of the monomer and are mainly found in the central B-chain α-helix (B9-B19) (namely B10 His, B13 Glu, B14 Ala, B17Leu, B18 Val and B19 Cys), the N-terminal B-chain (B1 Phe, B2 Val and B4 Gln) and the C terminal α-helix (A12-A19) in the A-chain (A13 Leu, A14 Tyr and A17 Glu). It is noted that the dimer-dimer interaction region shares no residues with the monomer-monomer interaction region. Also the B10 His is, in this group, essential for the coordination of Zn^{2+} in the hexamer (See previous sections).

Modification of B1 Phe with large groups may interfere with the formation of hexamers, whereas deletion of B1 Phe and B2 Val may weaken hexamer formation but would not prevent it (Brandenburg, 1969; Blundell *et al.*, 1972). Neither B1 and B2 is involved in stabilising interactions that might be important to the tertiary structure (Blundell *et al.*, 1972). Specific mutation of residues involved in the dimer-dimer interface can disrupt association properties without affecting the activity. This was shown with the mutant HB10D human insulin where self-association (especially hexamerisation) was reduced as shown by sedimentation equilibrium (Weiss *et al.*, 1991) without effecting the activity of the molecule *in vivo* (Vora *et al.*, 1988; Bristow, 1993). This is an aspect investigated in this work.

2.4.3 Residues involved in the biological activity of insulin

Receptor binding region

The residues that are thought to contribute to the properties of the receptor binding region (Shoelson *et al.*, 1992) and binding of insulin to its receptor are shown in Table 2.1 and Figure 2.1. Most of these residues are found on the surface of the monomer situated on the N-terminal A-chain (A1 Gly and A5 Gln), the central α-helix of the B-chain (B12 Val, B16 Tyr and B17 Leu), the C-terminal B-chain (B23-B25) and also the C-terminal A-chain (A19 Tyr and A21 Asn). There is a large overlap in the residues involved in the monomer-monomer interaction (previous section) as well as those

participating in receptor binding. These overlapping residues are mainly situated on the central α-helix B-chain (B12, B16, B17), the C-terminal B-chain (B23-B25) and also the C-terminal A-chain residue Asn. Residues A13 Leu and B17 Leu are also thought to be implicated in this binding and are also notably overlapping residues with the dimer-dimer interface that is thought to constitute the second receptor binding site (Schäffer, 1994). It is thus highly unlikely that this monomer-monomer interaction mode of dimerisation can be eliminated without adversely affecting the activity of the insulin. By implication it may be more beneficial to target the dimer-dimer interface.

Additional residues A2 Ile and A3 Val not found on the surface in the receptor binding region may also contribute to the binding of insulin to its receptor (Nakagawa and Tager, 1992). The reason for this is that it has been suggested that the C-terminal B-chain of insulin moves away from the rest from the N-terminal A-chain exposing the underlining hydrophobic core containing A2, A3, and possibly B15 (Baker *et al.*,1988) upon binding to the receptor. This separation includes B25 Phe and possible B24 Phe. This conformational change that has been suggested is based on observations that several insulins and insulin analogues with native-like structures , especially at the active surfaces, show reduced potency One of these is the (des B30) insulin in which A1 α-amino is peptide bonded to B29, now the COOH terminal residue (Markussen 1985). This cross-linked insulin is inactive but the residues in the identified active surfaces have an identical structure to that seen in active insulins. Although it crystallised in the 4 Zn structure, where B1-B8 changes into the helical conformation (see previous section) under the appropriate solution conditions, the C-terminal B-chain is unable to move away to expose the underlying A-chain residues. This suggests that this inability to move is the reason for loss of activity. Supporting the notion that a conformational change is necessary for activity is the insulin covalently crosslinked between B26 and B16 with a 80 % loss in activity (Derewenda *et al.*,1990). Another observation relevant to insulin's active surfaces is that despite the removal of B26-B30 this despenta insulin (DPI) is fully active provided the B25 carboxylate group is amidated. This indicates that the DPI

contains all the active surfaces necessary for the expression of biological activity (Derewenda *et al.*, 1990).

In the above mentioned cross-linked insulins the C-terminal B-chain is unable to move away from the A-chain to expose the underlining residues necessary for binding to the receptor explaining the loss of activity. In the DPI insulin those underlining residues are already exposed with the removal of the B26-B30 B-chain explaining the retention of full biological activity. Thus residues A2 Ile and A3 Val may also be added to the list of residues necessary for binding to the receptor (Derewenda *et al.*, 1990).

In a study where the steric requirements of the position B12 for high biological activity in insulin were investigated it was found that Val, a invariant residue (see following section) in the B12 position fulfills special side-chain packing requirements involved in the stability of the structure of insulin (Hu *et al.*, 1993).

Positions B24 Phe and B25 Phe are very important for biological activity and mutations at these positions are associated with diabetes in man (Tager *et al.*, 1979; Shoelson *et al.*, 1983a; Shoelson *et al.*, 1983b; Haneda *et al.*, 1984; Hua *et al.*, 1993). Various studies designed to elucidate the role of these aromatic groups (Mirmira and Tager, 1989; Mirmira and Tager, 1991; Mirmira *et al.*, 1991) have suggested that whereas none of these residues seem to directly influence the affinity in insulin-receptor interactions, they both play important roles in determining the three-dimensional structure of the insulin peptide backbone in the C-terminal region; a region which is critical for the receptor binding of the hormone (Mirmira *et al.*, 1991). This was indicated by replacement of B24 Phe by any variety of amino-acids with D-configuration or by Gly without major loss in receptor-binding potency (Mirmira and Tager, 1989). One of these, the mutant FB24G insulin where Phe is substituted by Gly in the same position, still displayed nearly full biological activity. Its solution structure indicates that the C-terminus of the B-chain is displaced from the native conformation exposing part of the underlying hydrophobic core (Hua *et al.*, 1991). Also B25 Phe can be replaced by other residues as long as the

C-terminal pentapeptide of the insulin B-chain is removed (Nakagawa and Tager, 1986; Nakagawa and Tager, 1987; Casaretto *et al.*, 1987; Schwartz *et al.*, 1989). Taking the above information into account and from a study where the peptide bonds in the C-terminal B-chain domain of insulin were replaced by a $\psi(CH_2\text{-}NH)$ linker, it seems though that sequential peptide bonds between insulin residues B23-B26 each play an important role in determining an acceptable insulin structure(s) as well as the potential for high-affinity insulin-receptor interactions (Nakagawa *et al.*, 1993), although the specific structural requirements of most of these positions are still unclear (Svoboda *et al.*, 1994). Currently suggestions regarding the structural requirements for position B25 have, however, been proposed . These are that the B25 imino group has a direct role in the formation of a hydrogen bond with the receptor (Wollmer *et al.*, 1994) and that the characteristics necessary for high affinity binding of this position are: the distance of side-chain branching and amide bond in B25, the planarity (sp^2-hybridization) at the $C\gamma$ atom in B25 and the potential for hydrogen bond formation, but not necessarily the aromatic structure (Spoden *et al.*, 1995).

The role of the B26 Tyr in receptor recognition is not clear because despite the fact that it is absent in des-(B26-B30) insulin amide the insulin still exhibits native bioactivity (Fischer *et al.*, 1985). Thus by implication the C-terminal pentapeptide of the B-chain is unnecessary for activity.

The insulin receptor binding region shares residues with the monomer-monomer interaction region (A21 Asn, B12 Val, B16 Tyr, B23 Gly, B24 Phe and B25 Phe) (Shoelson *et al.*, 1992). Schäffer (Schäffer, 1994) proposed a second binding domain on the opposite side of the molecule from the classical binding site. Residues A13 Leu and B17 Leu, located in the dimer-dimer interaction region (hexamer-forming surface), seemed to be involved in this second binding site. A model for receptor binding has been proposed where each of the two α-subunits of the insulin receptor (see later) contributes with a different binding region to the formation of the high-affinity binding site. Subsequently, a second molecule of insulin is able to bind to a low-affinity site involving

one of the α-subunits, thus accounting for the curvilinear Schatchard plot (Schäffer, 1994).

Growth promoting activity

An important point that should be raised in this chapter is that insulin has both metabolic and growth-promoting activities. One of these growth-promotion activities includes rapid effects on gene transcription (Messina, 1990). Which residues participate in growth promoting activities has been tentatively postulated; a postulation. The suggestions are based on a comparison of the residues in insulins as well as insulin-like growth factors (both of which show enhanced growth promoting effects) with mammalian insulins that show high potency for metabolic effects, but low potency for growth. Some of these insulins with growth promoting effects also show enhanced metabolic effects (turkey insulin) while others show low potency for metabolic effects (hystricomorph (guinea pig, porcupine, coypu and casiragua insulins) and the insulin-like growth factors (IGF)). The residues identified are A8 His in turkey insulin (Ala or Thr in mammals), B10 Asn, Gln or Glu in some hystricomorphs and IGF's (His in mammals); B13 Asp in some hystricomorphs and IGF's (Glu in mammals), A4 Asp in some hystricomorphs and IFG's (Glu in mammals) and B25 Tyr in some hystricomorphs and IGF's (Phe in mammals) (Horuk *et al.*, 1979; King and Kahn, 1981). Another feature of some of the hystricomorph insulins is that they have a substitute residue at B10 (Gln in coypu and Asn in guinea pig). Without a His in that position insulin is unable to coordinate around zinc ions to form a hexamer structure as found in pig insulin (Blundell, 1972).

2.5.1 Invariant amino-acids

The invariant residues and the residues in the receptor binding region that are important for biological activity are shown in Figure 2.5. The residues in insulin that are conserved may give insight into the structural properties required for the biological action of insulin. These invariant amino-acids can be divided into two groups

I A6 Cys, A7 Cys, A11 Cys, A20 Cys, B7 Cys and B19 Cys; A16 Leu, B6 Leu, B11 Leu, B15 Leu; B12 Val; B8 Gly and B23 Gly.

II A1 Gly, A19 Tyr and A21 Asn.

The first group (I), one which includes all the cysteines that form the disulphide bridges, are thought to be essential to the construction of the molecule. The second contains surface residues that may be extended to include certain others conserved in most insulins, A2 Ile, A3 Val, A5 Glu, B22 Arg, B24 Phe, B25 Phe and B26 Tyr. They lie on two continuous surfaces, the B-chain monomer-monomer interface (B24 -B26) and a polar surface containing the A-chain residues (A2, A3, A5) and B22. Some of these are in the flexible regions (N-terminal A-chain and C-terminal B-chain) where changes of the molecular arrangement in different crystals affect their conformation (Baker *et al.*, 1988).

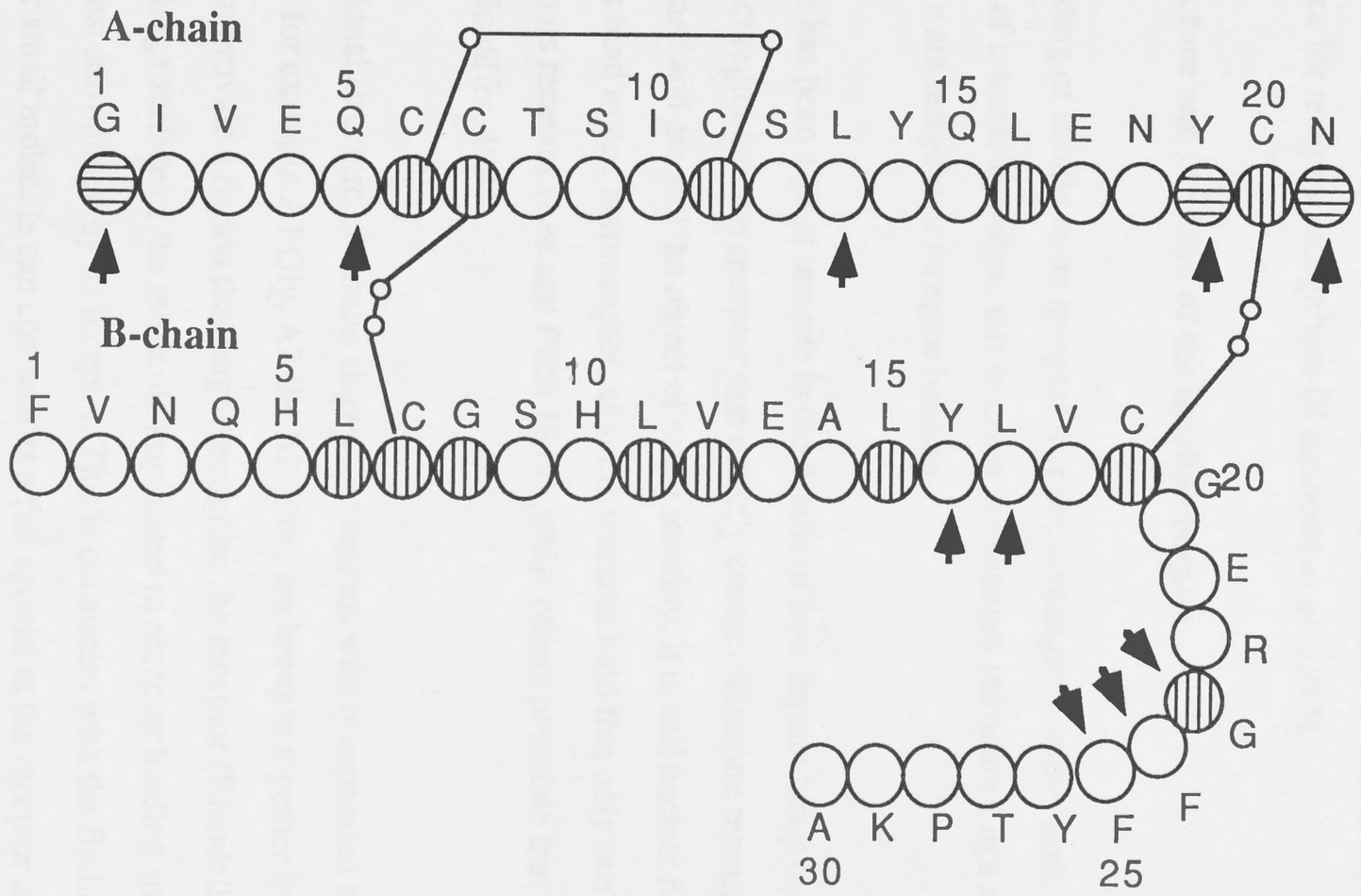
2.6 Stability of the insulin monomer fold with bearing to structure-activity relationship

For diabetic patients that use therapeutic insulin daily it is important that insulin should be stable. Considerable enhancement of folding stability is engineered by mutations in the N-terminal capping boxes of helices (Harper and Rose, 1993). Helices that may be considered for this stability engineering of the insulin molecule are the central α-helix of the B-chain (B9-B19) and the α-helix of the A-chain (A2-A8) (Kaarsholm *et al.*, 1993). In fact mutations that have been produced to confer increased stability in these regions

Figure 2.5

The amino-acid sequence of porcine insulin. Residues that are invariant in group I (text) are marked with horizontal line shading, residues that are invariant in group II are marked with vertical line shading and the residues indicated by arrows contribute to the receptor binding region.

Figure 2.5


A-chain
1 5 10 15 20
G I V E Q C C T S I C S L Y Q L E N Y C N
B-chain
1 5 10 15
F V N Q H L C G S H L V E A L Y L V C
G 20
E
R
G
F
A K P T Y F
30 25

are identical to those that lead to enhanced biological activity. Mutations in two regions were examined, namely A8 (the A2-A8 α-helix) where the threonine was substituted for a His or an Arg, and B10 (the B9-B19 α-helix) where the His was substituted for a Glu or an Asp. It was found that the substitutions affect stability in a nearly independent fashion. These stabilising mutations were generally found to enhance the cooperativity of the unfolding transition. It was concluded from this study that these highly potent insulin mutants elicit enhanced activity because these mutations stabilise structural motifs of critical importance for receptor recognition (Kaarsholm *et al.*, 1993).

2.7 The structure and function of the insulin receptor

Although the binding of insulin to its receptor was not investigated in the work presented in this thesis it is of interest to discuss this because the changes introduced in a desirable therapeutic insulin are relevant to receptor binding.

Historically, there has been a great interest in the details of how insulin binds to its receptor. The exact regions of the receptor that directly contact hormone remain incompletely defined and are still the object of much scrutiny. It is still unclear how many insulin monomers bind to one holoreceptor, as some workers hold that only one molecule of insulin binds to its receptor (Lee and Pilch 1994), while others postulate that two monomers bind (Schäffer, 1994).

The three-dimensional structure of insulin shows that regions widely separated in the primary structure, for example A1 Gly, A2 Ile, B25 Phe, are brought together in the tertiary structure to provide a region that might recognise the receptor (Blundell *et al.*, 1972). In fact the large surface of the molecule implicated in receptor binding suggests that insulin is almost surrounded by its receptor. This is consistent with the finding that no insulin mimetic small molecule can compete as a full agonist at the receptor and no insulin modified at a single amino-acid residue is a full agonist (Baker *et al.*, 1988). Also Nakagawa and Tager (Nakagawa and Tager, 1993) suggested that the insulin receptor

shows considerable capacity to participate in the conformational adjustment of the ligand molecules and that receptor mediated movements in native insulin lead to important structural adjustments without the need for major changes in the usual insulin fold.

The insulin receptor is a large glyco-protein molecule of molecular mass of about 300 000 daltons and a member of the tyrosine kinases family; a group of enzymes that play a critical role in cell division, metabolism and development (McDonald *et al.,* 1995). It consists of two alpha and two beta chains organised in a generally similar way to that in immunoglobulin molecules. A model of a ligand-free, inactive form of the structure of the receptor is shown in Figure 2.6.

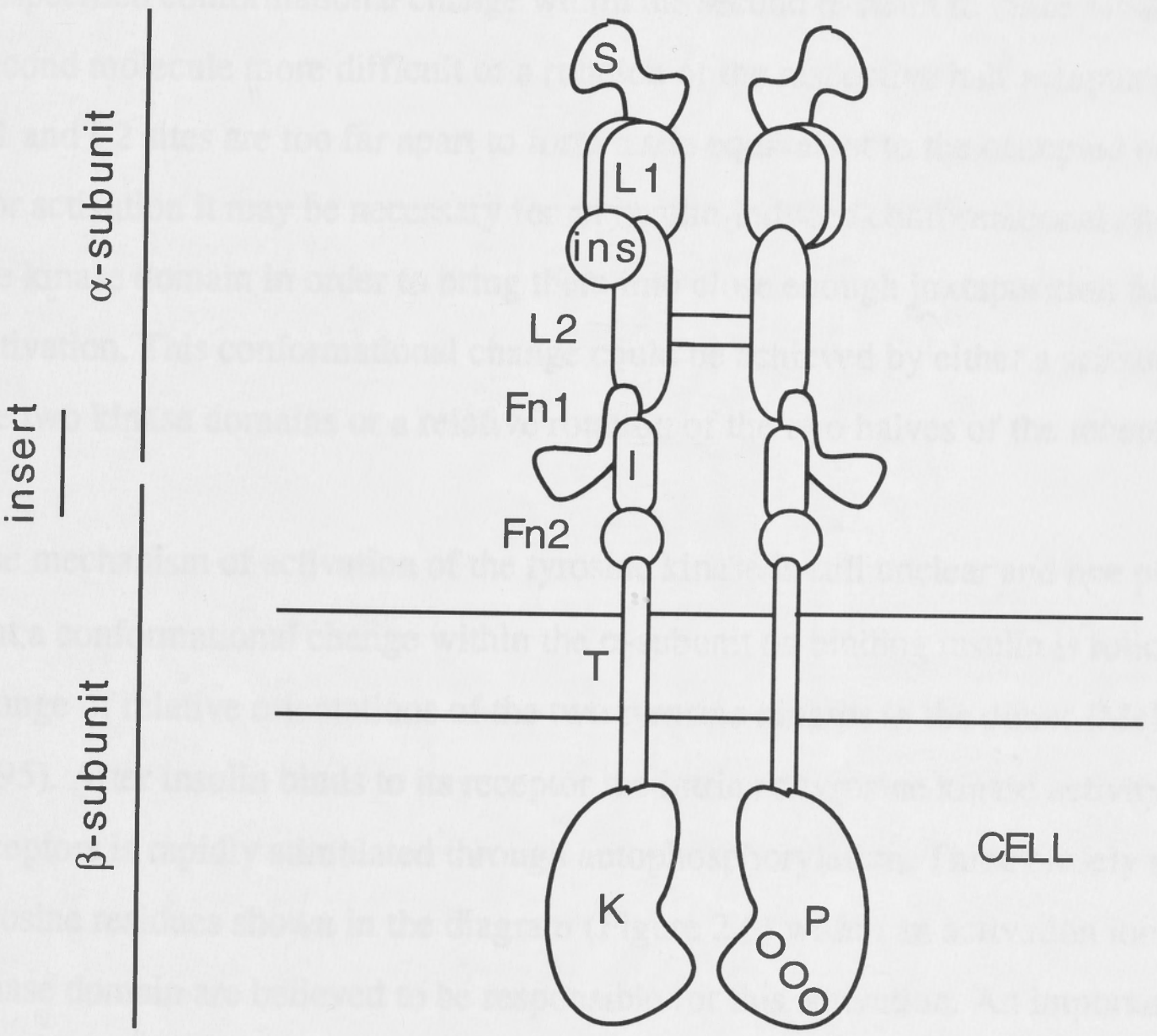
Cloning and sequencing the cDNA of the insulin receptor revealed similar features to those of other tyrosine receptor kinases (Ebina *et al.,* 1985; Ullrich *et al.,* 1985). These include an extracellular cystine-rich domain followed by a single transmembrane region which is probably helical, and then an intracellular tyrosine kinase. The insulin receptor is proteolytically processed into α and β subunits that remain covalently associated by disulphide bond(s). These $\alpha\beta$ units are further disulphide linked to give the mature $(\alpha\beta)2$ heterotetrameric receptor (McDonald *et al.,* 1995).

Some progress has been made in identifying receptor residues involved in binding of insulin and the residues implicated are in the L1, S and L2 domains (Figure 2.6) (Lee and Pilch, 1994). Studies with purified insulin receptor indicate that there are two classes of binding sites. One binding a single monomer of insulin which occurs at low concentration of the hormone. The other which occurs at higher concentrations of insulin where at least one further insulin monomer binds, but with a lower affinity. These bindings, together with the accelerated dissociation rate as measured using labelled insulin from the receptor in the presence of unlabelled insulin, is characteristic of the phenomenon of "negative cooperativity". Negative cooperativity has been interpreted either in terms of competition for binding sites on the receptor or in terms of insulin dimer formation competing with

Figure 2.6

A schematic representation of the model of the insulin receptor. The receptor comprises the α and β-units which are linked by disulphide bonds. There is a further disulphide bond which joins the two α-subunits of the tetramer. L1 and L2 are the duplicated large domains each comprising four or five copies of a helix-turn-strand motif, S is the cystine-rich domain, which contains three repeats. The first fibronectin domain F1 is interrupted by an insert which spans the α/β junction followed by the second fibronectin domain F2. T is the transmembrane helix and K the kinase domain. The circles on the kinase domain represent the three regulatory tyrosines of the kinase domain that are subjected to phosphorylation/dephosphorylation.

Figure 2.6



insulin receptor complexes. Most recently it has been explained in terms of a model in which bound insulin cross-links the subunits of the receptor (DeMeyts, 1994).

Two insulin molecules bind to the α-subunits of the (αβ)2 heterotetrameric receptor complex (Schäffer, 1994). The most current model for the binding of insulin to its receptor (to account for the negative cooperativity phenomenon) and activation of the receptor is that proposed by McDonald *et al.* (McDonald *et al.*, 1995) where they propose that: insulin may bind to the L1 and L2 domains in two possible ways. This is either to the L1 and L2 of the same α-chain or the two α-chains of the complex. Negative cooperativity exhibited upon binding insulin may be either through an unspecified conformational change within the second α-chain to make binding of the second molecule more difficult or a rotation of the respective half receptors so that the L1 and L2 sites are too far apart to form a site equivalent to the occupied binding site. For activation it may be necessary for an insulin-induced conformational change effecting the kinase domain in order to bring them into close enough juxtaposition for trans-activation. This conformational change could be achieved by either a scissors motion of the two kinase domains or a relative rotation of the two halves of the receptor.

The mechanism of activation of the tyrosine kinase is still unclear and one possibility is that a conformational change within the α-subunit on binding insulin is followed by a change of relative orientations of the two tyrosine kinases in the dimer (McDonald *et al.*, 1995). After insulin binds to its receptor the intrinsic tyrosine kinase activity of the receptors is rapidly stimulated through autophosphorylation. Three closely spaced tyrosine residues shown in the diagram (Figure 2.6) within an activation loop of the kinase domain are believed to be responsible for this activation. An important consequence of the activation is the recruitment of IRS-1 (insulin receptor substrate-1) and other effector and adaptor molecules. Each of these effector and adaptor molecules contain SH2 (Src homology 2) domains that bind phosphotyrosines with high affinity. IRS-1 is the main target for these SH2-containing proteins rather than the insulin receptor itself (White and Kahn, 1994), although recent evidence indicates parallel

pathways, independent of IRS-1 for activating kinases, also exist (Araki *et al.,* 1994). Eventually these transient complexes formed with IRS-1 activate a number of pathways involving kinases which can ultimately lead to the initiation of protein synthesis (Lin and Lawrence, 1994) and the translocation of an insulin-sensitive glucose transporter to the cell surface for glycolysis and lipogenesis.

Chapter 3

Platypus Insulin: Structural Analysis

3.1 Introduction

Investigations of the composition, structure, biological activity and self-association of insulin from more exotic species increases the possibility of identifying an insulin with desirable properties by exploiting natural variation. The platypus (*Ornithorhynchus anatinus*) is one such exotic animal. It is the only extant species of the family Ornithorhynchidae belonging to the order Monotremata, the egg laying mammals. Thus a detailed understanding of the many facets of the platypus insulin structure, primary, secondary and tertiary are of fundamental importance to the understanding of its activity and physico-chemical properties. These include properties of amino-acids in the sequence that contribute to:

- the structural integrity of the overall three-dimensional structure (invariant residues),
- the self-association interfaces (monomer-monomer and dimer-dimer)
- the receptor binding region
- the metabolic potency and
- the growth promotion effects of platypus insulin.

Also an understanding of these above facets at a fundamental molecular level will ultimately enhance future ability to design therapeutic insulins with desirable properties.

The physico-chemical properties, namely the self-association behaviour in solution at physiological pH, of platypus insulin are of particular interest. This self-association only becomes of significance in those instances where the insulin is present in high concentrations, as is the case in diabetes therapy and in the determination of the three-dimensional structure of insulin in solution by the method of 2D NMR where aggregation of insulin produces broad NMR signals which reduce the ability to determine structures

in detail. (See Chapter 2). There are two regions implicated in self-association, namely, the monomer-monomer interaction region (OP) and on the opposite side of the molecule, the weaker dimer-dimer interaction region (OQ) (Chapter 2). These two interaction regions as well as the receptor binding surface can be regarded as independent design features when engineering mutants with altered functional properties (Shoelson *et al.*, 1992). There is, however, a large overlap of residues involved in the monomer-monomer interface and those participating in receptor binding (Chapter 2). Residues targeted for mutation in the monomer-monomer interface to bring about the reduction in self-association must be chosen carefully so as not to disrupt the hormone's receptor binding capability and thus its activity. As the monomer-monomer interaction is the stronger interaction (compared with the dimer-dimer interaction) this area has been the main target for producing mutations that will reduce the strength of this interaction and thus the overall self-association of insulin. Mutations in this area are also used to probe the binding characteristics of the hormone to its receptor without taking into account its self-association properties. (Binding of insulin to its receptor occurs at concentrations (ng) at which insulin does not aggregate). It might, however, be of an advantage to produce mutations in the dimer-dimer interaction region to reduce self-association since this region is on the opposite side of the molecule and independent of the receptor binding region.

In the first part of this chapter the sequence of platypus insulin is compared by the method of pairwise sequence comparison as well as multiple sequence alignment to the sequence of all the insulins of different animals. This was done in order to identify the degree of similarity at the amino-acid sequence level of structural elements as well as the regions in the insulin molecule that contribute to its the physico-chemical and biological properties. In this sequence comparison analysis by computer the amino-acids are described with a set of physico-chemical properties. In the construction of a multiple sequence alignment qualitative descriptions of the residues found at each position are used (Gracy *et al.*, 1993). To do this the Genetics Computer Group sequence analysis software package was used (Devereux *et al.*, 1984). This sequence comparison analysis

may also give information in respect to the evolutionary relationship of platypus insulin with other insulins from different organisms. The method of multiple sequence alignment is well suited to the latter. It was also of interest to determine the relationship of the platypus insulin sequence with that of the hystricomorph insulins (guinea pig, porcupine, coypu and casiragua) as some of the latter insulins have distinct physico-chemical and biological properties (Chapter 2). For example some of these hystricomorph insulins do not self-associate to the extent that mammalian insulins do. This is because B10 His, essential for coordinating zinc in hexamer formation in some of these insulins is substituted by Asn or Gln (Blundell, et al., 1972). Also its growth promoting activity (relatively high) and metabolic potency (relatively low) are more comparable with insulin-like growth factors than with mammalian insulins (Chapter 2 and King and Kahn, 1981).

The full function of each amino-acid in a protein sequence is manifest only in its context in the tertiary structure. Therefore to make meaningful alterations to the sequence requires a three-dimensional model (Taylor, 1988). Thus in order to analyse the structure of platypus insulin in a meaningful way, information about its three-dimensional structure is necessary. Only the amino-acid sequence of platypus insulin was available so in order to obtain some information about its three-dimensional structure platypus insulin was modelled on the basis of its complete sequence using the coordinates of the X-ray crystal structure of porcine insulin. These studies form part of a program with the aim of evaluating insulins and insulin derivatives, both new and previously reported, with the desired characteristics.

3.2 Materials and Methods

3.2.1 Isolation and purification of native platypus insulin

Platypus insulin was isolated by G. Barbara Treacy (Nourse *et al.*, 1996). The procedure is outlined below in detail.

Platypus pancreatic tissue was obtained from 9 animals which had been deep frozen at death. For the preparation the bodies were kept in a refrigerator overnight and the tissue dissected out before defrosting was complete.

The extraction of the pancreases was carried out in the cold by a modification of the method of Treacy *et al.* (Treacy *et al.*, 1989). This consisted of homogenisation in cold acidic ethanol, centrifugation, adjustment of the pH of the supernatant to 8 and removal of the resulting precipitate by centrifugation. The supernatant was adjusted to pH 5.3, 2 M ammonium acetate solution at pH 5.3 added and the crude insulin and pancreatic polypeptide precipitated by the addition of cold absolute ethanol and cold diethyl ether. After being left overnight in the cold the precipitate was collected by decantation and centrifugation.

At this point it was observed that the precipitate did not separate cleanly from the liquid but rather as a dark coloured oily-looking liquid between the yellow precipitate and the other layer. This is not normally seen when preparing insulins from other organisms.

Platypus insulin was obtained from this "oily supernatant". Reversed phase (RP)-HPLC was employed to purify the insulin by applying successive 25µl aliquots, diluted to 250µl with starting solution, to a HPLC RP300C8 column, 4.6 mm i.d., and collecting the appropriate peak in each case. Elution was in 0.09% aqueous trifluoroacetic acid using a linear gradient from 8% to 50% acetonitrile in 20 min. The insulin was purified further by combining the material from several runs and reapplying them to the RP300C8 column.

3.2.2 Determination of the amino-acid sequence of platypus insulin

The amino-acid sequence of insulin from platypus *Ornithorhynchus anatinus* was determined using the Applied Biosystems Model 477, a pulsed liquid phase sequencer with on-line Model 120A PTH-Analyser (Applied Biosystems Inc., Foster City, CA, U.S.A.) modified for picomole sensitivity. Methods used were as described by Treacy *et*

al. (Treacy *et al.*, 1989). The molecular mass of platypus insulin, calculated as 5865 daltons, has been confirmed with a matrix-assisted laser desorption ionisation (MALDI) time-of-flight mass spectrometer by the author. For a more detailed description of the mass analysis see chapter 4 section 4.2.2.

3.2.3 Comparative phylogenetic analysis of the amino-acid sequence of platypus insulin

For the amino-acid sequence comparison studies the Genetics Computer Group (GCG) Sequence Analysis Software Package Version 7.1 (Devereux *et al.*, 1984) that include the FASTA, GAP, PILEUP and SEQED programs, was used. The computationally fast algorithm FASTA, (Pearson and Lipman, 1988) that allows a homology search of one sequence against the entire SWISS-PROT amino-acid sequence database was employed to search for insulin from different organisms with amino-acid sequences related to the amino-acid sequence of platypus insulin. FASTA uses a Ktup value of 2 (the default value) and a PAM250 scoring matrix. Scores between platypus insulin amino-acid sequence and each of the database entries were calculated and presented in order of decreasing best scores with the highest scores corresponding to the strongest relationship. The best 55 scores (all the insulin amino-acid sequences) were subjected to analysis by a more sensitive sequence comparison program, GAP (Needleman and Wunch, 1970). The GAP program that covers the complete sequence, maximises the number of matches and minimises the number of gaps. Data entries, where the amino-acid sequences included the signal and C chain, had been edited out with SEQED for the GAP alignments.

A dendogram displaying the clustering relationships of the primary sequence of platypus insulin with insulins from other organisms was compiled with PILEUP.
The PILEUP program (GCG) is an extension of the GAP program (GCG) where a multiple sequence alignment (algorithm of Feng and Doolittle) (Feng and Doolittle, 1987) is created from a group of related sequences using progressive pairwise alignment.

3.2.4 Molecular modelling of the three-dimensional structure of platypus insulin

Platypus monomer insulin structure was modelled using the coordinates of the X-ray crystal structure of rhombohedral 2 Zn porcine insulin, monomer 2, entry 3 ins (Wlodawer *et al.*, 1989) of the Brookhaven Protein Databank. Monomer 1 (Chinese nomenclature) differs from monomer 2 mainly at the B25 Phe in the 2 Zn dimeric crystal structure. In monomer 1 the B25 Phe side chain turns inward to contact A19 Tyr, whereas the side chain in monomer 2 points outward from the molecule (Baker *et al.*, 1988).

The porcine insulin monomer 2 structure was subjected to 250 iterations of steepest decent energy minimisation followed by 5000 iterations of conjugate gradient *in vacuo*. This energy minimised porcine insulin monomer 2 structure has been used for modelling the structure of platypus monomer insulin with the appropriate different residues A5, A8, A9, A10, A13, B2, B22, B25, B27, B29 and B30. The platypus monomer insulin has been subjected to 250 iterations of steepest decent energy minimisation followed by 5000 iterations of conjugate gradient energy minimisation *in vacuo*. The maximum derivative achieved was 0.0093 kcal/mol-Å.

Molecular dynamics for 100 ps of simulation at 200 K have subsequently been performed *in vacuo*. Analysis was performed using the transient structures written at intervals of 1 ps. Computations for molecular dynamics were carried out using the Fujitsu VP of the Australian National University Supercomputer Facility. During the simulations the charges were on and constant pressure off. The number of iterations were 100 000 with a time step of 1 fsec. No cross terms energies and no Morse potential for bond energies were included. The leapfrog integration algorithm was employed. A history file was written every 1000 steps.

Transient structures written at 15 ps; 55 ps; 81 ps and 97 ps were chosen and each were energy minimised by using the conjugate gradient algorithm for 5000 steps *in vacuo*. In

all cases the maximum derivative was below 0.001 kcal/mol-Å. Energy minimisation was performed with charges on. Cross term energies and Morse potentials for bond energies were also included.

Both the energy minimisations and simulations were carried out using the program Discover Version 94.0 (May 1994) on a Silicon Graphics workstation as well as the analysis using the program Analysis, all generated by the computer program Insight II version 2.3.0 (December 1993).

3.3 Results

3.3.1 Isolation and purification of native platypus insulin

Platypus insulin was obtained from the "oily supernatant" after the precipitation step in the isolation process (Section 3.2). The yield of the platypus insulin was just sufficient to meet the amount required (> μg) for the amino-acid sequence determination. No peak of platypus insulin was found from the yellow precipitate where it was expected but which instead contained a pancreatic polypeptide. This platypus pancreatic polypeptide was identified by amino-acid sequence and comparison with the SWISS-PROT database.

3.3.2 The amino-acid sequence of platypus insulin

The amino-acid sequence of platypus insulin shown in Figure 3.1, differs from that of human and porcine insulin at eleven residues. Most of these, summarised in Table 3.1, represent relatively conservative substitutions in regions that have been observed as varying from highly conserved to variable. Three, A5 Glu, A8 Lys and A13 Met have not been observed in other organisms (Table 3.2). Most substitutions occur on the surface, or at least partial surface, in regions important to receptor binding and aggregation of the molecule. Substitutions at all three interactive surfaces include B25 Phe→Tyr and

B27 Thr→Ile (monomer-monomer interface); A13 Leu→Met and B2 Val→Pro (dimer-dimer interface) and A5 Gln→Glu and B25 Phe→Tyr (receptor binding region).

Figure 3.1

The amino-acid sequence of insulin from platypus (*Ornithorhynchus anatinus).* The 11 residues that differ in porcine insulin are shown below the platypus sequence.

A-Chain

```
1         5         10        15        20
G I V E E C C K G V C S M Y Q L E N Y C N
      Q     T S I     L
```

B-Chain

```
1         5         10        15        20        25        30
F P N Q H L C G S H L V E A L Y L V C G E K G F Y Y I P R M
  V                                   R     F     T     K A
```

3.3.3 Pairwise amino-acid sequence comparisons of platypus insulin with that of all known insulin sequence

The pairwise comparisons of platypus insulin sequences with insulin sequences from other organisms using the GAP program is presented in Table 3.3 in order of decreasing percent identity with the platypus insulin sequence. The sequence comparison program GAP make comparisons between pairs of amino-acids by looking up a value in a symbol comparison table, in this case the Dayhoff PAM-250 matrix . The symbol comparison table has a real number for the match quality of every possible pair of symbols and the

Table 3.1: A phylogenetic comparison of the amino-acid sequence of platypus and porcine insulin. Only platypus amino-acid substitutions relative to porcine insulin are presented.

Residue	Amino-acid		Side Chain Properties		Structural domain [a,b]	Type [a,b]
	Porcine	Platypus	Porcine	Platypus		
A5	Gln	Glu*	polar	acidic	α-helix, receptor binding	conserved
A8	Thr	Lys*	polar	basic	α-helix	variable
A9	Ser	Gly	polar		surface residue	variable
A10	Ile	Val	non polar	non polar	extended β-conformation	variable
A13	Leu	Met*	non polar	non polar, S	α-helix, dimer-dimer interface	variable
B2	Val	Pro	non polar	non polar	dimer-dimer interface	variable
B22	Arg	Lys	basic	basic	β-bend, ionic pair	conserved
B25	Phe	Tyr	aromatic	aromatic	monomer-monomer interface, receptor binding region	conserved
B27	Thr	Ile	polar	non polar	monomer-monomer interface	variable
B29	Lys	Arg	basic	basic	extended β-chain	variable
B30	Ala	Met	polar	non polar, S	extended β-chain	variable

[a] Baker *et al.*, 1988; [b] Shoelson *et al.*, 1992; S: sulfur containing residue; *Previously not observed in insulin

Table 3.2: Comparison of platypus insulin with insulin sequences from other organisms. Only platypus insulin residues different from porcine insulin are shown. Sequence library used was the SWISS-PROT database.

Organisms that share residues with platypus insulin.	Platypus insulin residues differing from porcine	
None	A5 [a)]	Glu
None	A8 [a)]	Lys
None	A13 [a)]	Met
Elephant, goat, porcupine, sheep, horse, guinea pig	A9 [c)]	Gly
Elephant, goat, porcupine, sheep, cat, camel, bovine	A10 [c)]	Val
Rattle snake, cantor, electric ray, dogfish, ratfish, elephantfish	B2 [c)]	Pro
Electric ray, dogfish, salmon, alligator	B22 [c)]	Lys
Rattle snake, cantor, electric ray, dogfish, degu, copyu, casiragua	B25 [b)]	Tyr
Guinea pig	B27 [c)]	Ile
Rattle snake, cantor	B29 [c)]	Arg
Sponge	B30 [c)]	Met

[a)]Residues not observed in other organisms; [b)]Residues conserved, but observed in other organisms; [c)]Residues variable and shared with other organisms.

number is assigned based on chemical similarity or evolutionary distance. Percent identity is the percent of the symbols that actually match (Devereux *et al.*, 1984).

The percent identity of platypus insulin with other insulins is not particularly high, the highest being only about 82% (Table 3.3). This is the group of mammalian insulins that include elephant, goat and sheep. In the percent identity range of 72-82 % are represented insulins of mammals, reptiles, birds and amphibians. Mammalian insulins are predominantly at the higher end and amphibian insulins at the lower end of this range. Outside this range, (54-72%), are found the insulins of fish and the hystricomorphs (guinea pig, porcupine, coypu and casiragua), the latter at the lower end.

3.3.4 Multiple sequence alignment of platypus insulin with that of all known insulin sequences

The insulin sequences from different organisms were also subjected to multiple sequence alignment on the basis of pairwise similarity (PILEUP) and presented as a dendogram (Figure 3.2). The vertical branch distance of the dendograms are proportional to the similarity between sequences (Devereux *et al.*, 1984).

In the dendogram the insulin sequences cluster into 4 main branches. The first two branches consist of the amphibian insulin sequence (cluster 1) and mammalian insulin sequence (cluster 2). These two together form cluster 3. The mammalian insulin sequences (cluster 2) can be roughly divided into subcluster 2.1, the mouse and rat insulin sequences and subcluster 2.2, which contain the very similar sequences of porcine, bovine, human, elephant etc. It can be pointed out that the marsupial insulin sequence (*Didelphis marsupials virginiana* identical to kangaroo *Macropus giganteus*) (Treacy *et al.*, 1989) and the low potency insulins of *Aotus trivirgatus* (Seino *et al.*, 1987), *Hystrix cristata* (King and Kahn, 1981) and *Chinchilla brevicaudata* (Horuk *et al.*, 1979) are outside subclusters 2.1 and 2.2 which indicate that they are more distantly related. The third branch is represented by the single platypus insulin sequence which participates in a

Table 3.3: Pairwise comparison of platypus insulin with insulin sequences from other organisms in order of decreasing percent identity. Amino-acid sequence comparisons were performed using the Genetics Computer Group programmes GAP which used the PAM 250 scoring matrix. The protein sequences and their respective references are stored in the SWISS-PROT database.

SWISS-PROT Nomenclature	Organisms	% Identity
Plin[a)]	*Ornithorhynchus anatinus* (platypus)	100
Ins_Elema	*Elephas maximus* (Indian elephant)	82.353
Ins_Caphi	*Capra hircus* (goat)	82.353
Ins_Sheep	*Ovis arie* (sheep)	82.353
Ins_Horse	*Equus caballus* (horse)	80.392
Ins_Camdr	*Camelus dromedarius* (Arabian camel)	80.392
Ins_Bovin	*Bos taurus* (bovine)	80.392
Ins_Balph	*Balaenoptera physalus* (finback whale)	78.431
Ins_Macfa	*Macaca fascicularis* (monkey)	78.431
Ins_Canfa	*Canis familiaris* (dog)	78.431
Ins_Pig	*Sus scrofa* (pig)	78.431
Ins_Human	*Homo sapiens* (man)	78.431
Ins_Rabit	*Oryctolagus cuniculus* (rabbit)	78.431
Ins_Felca	*Felis catus* (cat)	78.431
Ins_Balbo	*Balaenoptera borealis* (sei whale)	78.431
Ins_Anapl	*Anas plathyrhynchos* (duck)	76.471
Ins_Croat	*Crotalus atrox* (rattlesnake)	76.471
Ins_Ansan	*Anser anser anser* (goose)	76.471
Ins_Chick	*Gallus gallus* (chicken)	76.471
Ins_Crilo	*Cricetulus longicaudatus* (Chinese hamster)	76.471
Ins_Hyscr	*Hystrix cristata* (porcupine)	76.471
Ins_Lepsp	*Lepisosteus spatula* (alligator)	74.510
Ins_Zoadh	*Zaocys dhumnades* (cantor)	74.510
Ins_Acoca	*Acomys cahirinus* (Egyptian spiny mouse)	74.510
Ins_Didma	*Didelphis marsupials virginiana* (opossum)	74.510
Ins2_Rat	*Rattus norvegicus* 2 (rat)	74.510
Ins2_Mouse	*Mus musculus* 2 (mouse)	74.510
Ins_Allmi	*Alligator mississippiensis*	74.000

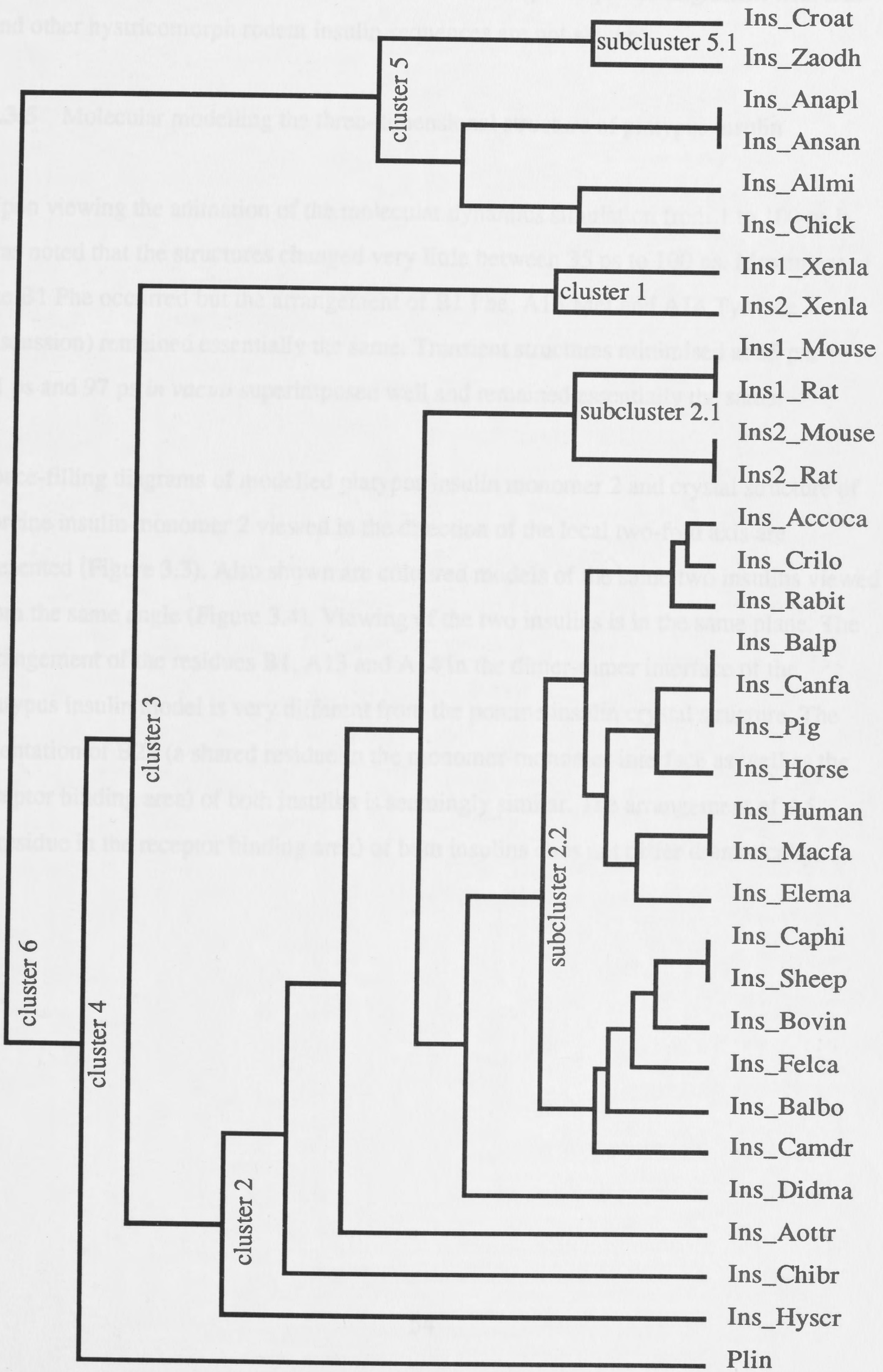
Ins_Geocy	*Geodia cydonium* (sponge)	74.000
Ins2_Xenla	*Xenopus laevis* 2 (African clawed frog)	72.549
Ins1_Mouse	*Mus musculus* 1 (mouse)	72.549
Ins1_Rat	*Rattus norvegicus* 1 (rat)	72.549
Ins_Squac	*Squalus acanthias* (spiny dogfish)	72.549
Ins1_Xenla	*Xenopus laevis* 1 (African clawed frog)	70.588
Ins_Aottr	*Aotus trivirgatus* (night monkey)	70.588
Ins_Torma	*Torpedo marmorata* (electric ray)	70.588
Ins_Chibr	*Chinchilla brevicaudata* (chinchilla)	68.627
Ins_Oncgo	*Onchorhynchus gorbuscha* (salmon)	68.000
Ins_Loppi	*Lophius piscatorius* (goosefish)	67.347
Ins_Cypca	*Cyprinus carpio* (carp)	67.347
Ins_Oncke	*Oncorhynchus keta* (salmon)	66.667
Ins_Myosc	*Myoxocephalus scorpius* (sculpin)	66.000
Ins_Thuth	*Thunnus thynnus* 2 (bluefin tuna)	66.000
Ins_Plafe	*Platichthys flesus* (flounder)	66.000
Ins_Petma	*Petromyzon marinus* (sea lamprey)	66.000
Ins_Calmi	*Callorhynchus milii* (elephantfish)	64.706
Ins_Hydco	*Hydrolagus colliei* (spotted ratfish)	64.706
Ins1_Batsp	*Batrachoididae sp* 1 (toadfish)	64.000
Ins_Gadca	*Gadus callarias* (Baltic cod)	64.000
Ins2_Batsp	*Batrachoididae sp* (toadfish)	63.265
Ins_Katpe	*Katsuwonus pelamis* (skipjack tuna)	62.000
Ins_Myxgl	*Myxine glutinosa* (Atlantic hagfish)	60.784
Ins_Octde	*Octodon degus* (degu)	60.000
Ins_Cavpo	*Cavia porcellus* (guinea pig)	58.824
Ins_Myoco	*Myocastor coypus* (coypu)	54.000
Ins_Progu	*Proechimys guairae* (casiragua)	54.000

[a)] Name assigned for platypus insulin.

Figure 3.2

Dendogram displaying the clustering relationships of the primary sequence of platypus insulin with insulins from other organisms using the PILEUP program. Platypus insulin was given the name plin. Names of insulins from different organisms are SWISS-PROT designations and their sequences and their respective references are stored in the SWISS-PROT database. Own clustering nomenclature was used.

Figure 3.2


Ins_Croat
subcluster 5.1
Ins_Zaodh
cluster 5
Ins_Anapl
Ins_Ansan
Ins_Allmi
Ins_Chick
Ins1_Xenla
cluster 1
Ins2_Xenla
Ins1_Mouse
Ins1_Rat
subcluster 2.1
Ins2_Mouse
Ins2_Rat
Ins_Accoca
Ins_Crilo
Ins_Rabit
Ins_Balp
Ins_Canfa
Ins_Pig
Ins_Horse
Ins_Human
Ins_Macfa
Ins_Elema
subcluster 2.2
Ins_Caphi
Ins_Sheep
Ins_Bovin
Ins_Felca
Ins_Balbo
Ins_Camdr
Ins_Didma
cluster 3
Ins_Aottr
Ins_Chibr
cluster 2
Ins_Hyscr
cluster 4
Plin
cluster 6

clustering relationship with the amphibian and mammalian insulin sequences (cluster 3) to form cluster 4. The final branch consists of reptilian and avian insulin sequences (cluster 5) that clusters with cluster 4 to form cluster 6. (Multiple sequence alignment with fish and other hystricomorph rodent insulin sequences are not shown).

3.3.5 Molecular modelling the three-dimensional structure of platypus insulin

Upon viewing the animation of the molecular dynamics simulation from 1 to 100 ps it was noted that the structures changed very little between 35 ps to 100 ps. Movement of the B1 Phe occurred but the arrangement of B1 Phe, A13 Met and A14 Tyr (see discussion) remained essentially the same. Transient structures minimised at 55 ps; 81 ps and 97 ps *in vacuo* superimposed well and remained essentially the same.

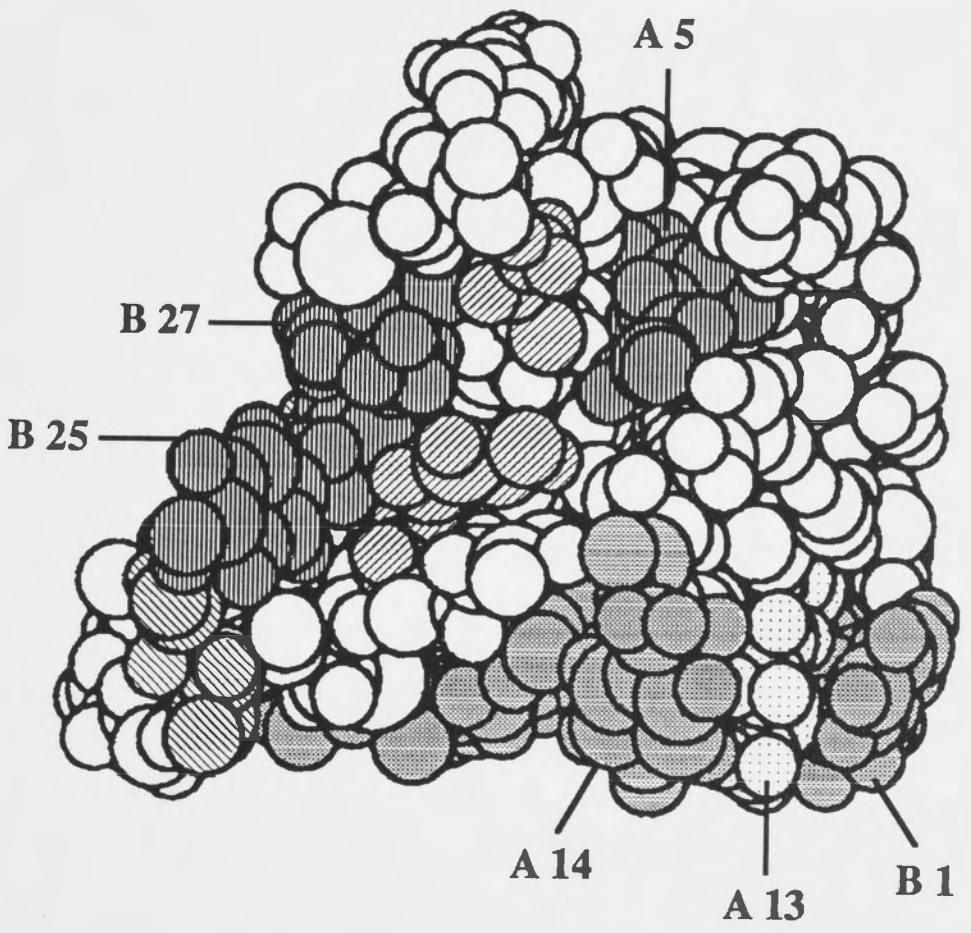
Space-filling diagrams of modelled platypus insulin monomer 2 and crystal structure of porcine insulin monomer 2 viewed in the direction of the local two-fold axis are presented (Figure 3.3). Also shown are coloured models of the same two insulins viewed from the same angle (Figure 3.4). Viewing of the two insulins is in the same plane. The arrangement of the residues B1, A13 and A14 in the dimer-dimer interface of the platypus insulin model is very different from the porcine insulin crystal structure. The orientation of B25 (a shared residue in the monomer-monomer interface as well as the receptor binding area) of both insulins is seemingly similar. The arrangement of A5 (a residue in the receptor binding area) of both insulins does not differ dramatically.

Figure 3.3

Space-filling diagrams of modelled platypus insulin monomer 2 and porcine insulin monomer 2 viewed in the direction of the local two-fold axis. Receptor binding region (diagonal right to left line shading,), monomer-monomer interface (diagonal left to right line shading), dimer-dimer interface (shaded), A13 (speckled) ; platypus insulin A5, B25, and B27 (vertical line shading).

Figure 3.3

Platypus Insulin



Porcine Insulin

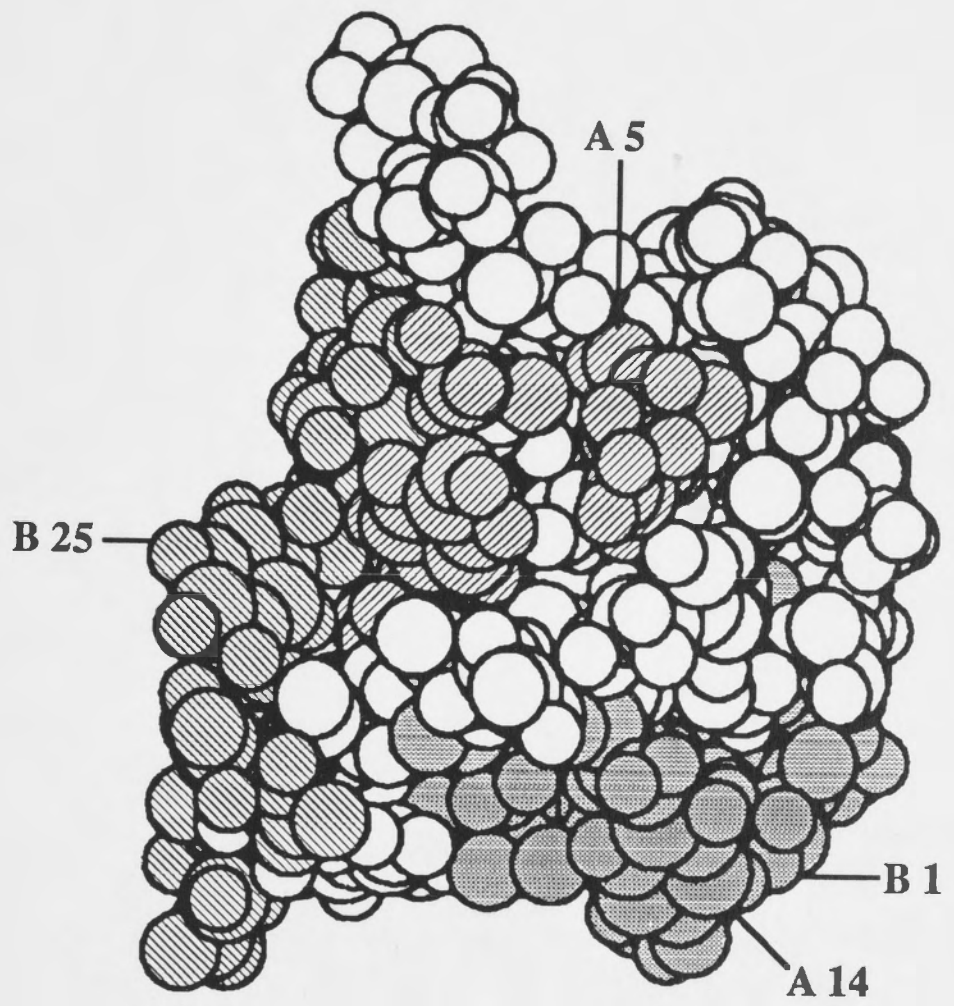
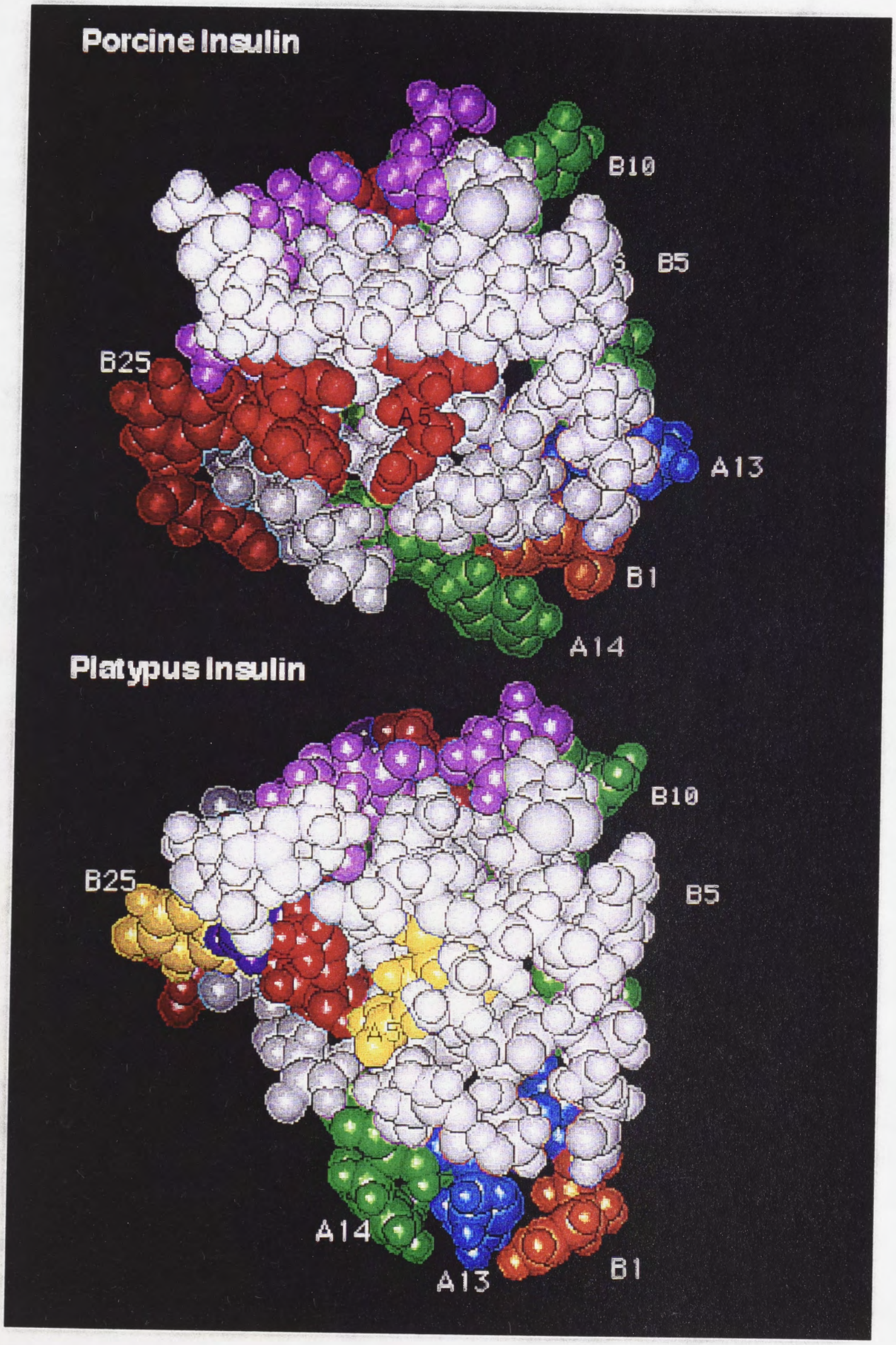


Figure 3.4

CPK coloured models of platypus and porcine insulin. The monomer-monomer interface is in pink, the dimer-dimer interface in green and the receptor binding area in red. There is considerable overlap of the monomer-monomer interface and the receptor binding area. Residue B1 is in orange and A13 in blue. Residues different in platypus insulin B25 and A5 are in yellow and B27 in purple. The view is along the two-fold axis.


Porcine Insulin
B10
B5
B25
A5
A13
B1
A14
Platypus Insulin
B10
B25
B5
A14
A13
B1

3.4 Discussion

3.4.1 Amino-acid sequence of platypus insulin

Structural integrity of platypus insulin structure (invariant residues)

Most of the eleven residue substitutions of platypus insulin (compared with porcine insulin) occur on the surface, or at least partial surface, in regions important to receptor binding and aggregation of the molecule. The amino-acid sequence reveals strict conservation of the residues known to be involved in maintaining the three-dimensional structure (Baker *et al.*, 1988). These include the thirteen invariant residues (the six half-cystine residues, A6, A7, A11, A20, B7, B19, glycine residues at B8 and B23, leucine residues at A16, B6, B11, and B15 and B12 Val) as well as the invariant group of surface residues believed to be involved in receptor binding and expression of biological activity (A1 Gly, A19 Tyr and A21 Asn) (Baker *et al.*, 1988). Some of the highly conservative sites A2 Ile, A3 Val, B24 Phe and B26 Phe to be found on the surface of the molecule are present in the platypus insulin structure. A5, B22 and B25, that are grouped with the latter, are conservative changes of

A5 Gln→Glu, B22 Arg→Lys and B25 Phe→Tyr in platypus insulin.

Self-association interfaces (monomer-monomer and dimer-dimer)

There are three distinct surfaces on the insulin molecule, namely, the monomer-monomer interface (A21, B8, B9, B12, B16, B20, B21 and B23-B28), the dimer-dimer interface (A13, A14, A17, B1, B2, B4, B10, B13, B14, and B17-20) that constitute the protein-protein interactive surfaces and the receptor binding surface (A1, A5, A19, A21, B12, B16, B23, B24 and B25) (Baker *et al.*, 1988; Shoelson *et al.*, 1992). Platypus insulin has substitutions at all three interactive surfaces namely B25 Phe→Tyr and B27 Thr→Ile (monomer-monomer interface); A13 Leu→Met and B2 Val→Pro (dimer-dimer interface) and A5 Gln→Glu and B25 Phe→Tyr (receptor binding region).

Self-associating forces that lead to dimerisation of insulin (Baker *et al.*, 1988; Shoelson *et al.*, 1992) are predominantly non-polar, reinforced by hydrogen bonding (Baker *et al.*, 1988; Pocker and Biswas, 1981). Thus B25 and B27 Ile, the residues different in platypus and part of monomer-monomer interface, considered conservative changes, fulfil these criteria. All other residues that form part of this surface, including B28 Pro, the latter especially important for high affinity self association (Brems *et al.*, 1992), are unchanged in platypus insulin. It can be concluded then that it is highly likely that the platypus insulin will form dimers and since B5 His and B 10 His are conserved, hexamerization can be expected in the presence of zinc (Bentley *et al.*, 1976; Blundell *et al.*, 1972).

Receptor binding region and biological activity (metabolic and growth)

A model for insulin binding to the receptor was recently proposed (Schäffer, 1994) where the involvement of A13 and B17 in a putative second receptor binding site were discussed. Platypus insulin has a Met in position A13 compared to Leu in porcine, which has not been seen in other organisms. B25 Tyr, also part of the receptor binding region and different from porcine insulin, is Phe in most mammals. The aromatic ring seems to be important for biological activity (Nakagawa and Tager, 1986). It has been shown that substitution at A8 Thr for Arg or His (positively charged residues) provide strong stabilisation of the A-chain NH_2-terminal helix which leads to enhanced biological activity (Kaarsholm *et al.*, 1993). Substitutions at A5 Gln→Glu (receptor binding region and charged) and A8 Thr→Lys (positively charged residue) of platypus insulin introduce possible charge interactions in this region. The A5 Glu and B 25 Tyr substitutions are known from the IGF-1 molecule (Rinderknecht and Humbel, 1978) and in combination with substitutions in A8, A13 and B22, indicate that binding properties to the insulin and IGF-1 receptors would be interesting to investigate. Receptor interaction and biological activity may be affected by the above substitutions and the structural consequences need direct investigation. Also some members of the insulins of the hystricomorphs which

show enhanced *in vitro* growth promoting activity, reduced insulin receptor binding affinity and low metabolic potency but enhanced *in vitro* growth promoting activity, share residues with the sequence of platypus insulin (Table 3.2). These residues are A9 Gly, B27 Ile and B25 Tyr while only the latter B25 Tyr is postulated to participate in this growth-promoting activity (King and Kahn, 1981).

The consequences of these residue changes of platypus insulin were subsequently investigated on a chemically synthesised platypus insulin. The *in vitro* biological activity (metabolic) (Chapter 4 section) and the self-association properties in a Zn free neutral pH solution (Chapter 5, section) were investigated.

3.4.2 Pairwise amino-acid sequence comparisons of platypus insulin with that of all known insulin sequences

The relatively low percentage identity (82%) of platypus insulin with its nearest relatives (Table 3.3), a group of mammalian insulins from elephant, goat and sheep, emphasises the unique biological features of this monotreme. This low percentage identity is reflected in the platypus insulin sequence differing at a great number of residues from its nearest relatives (8) and possessing amino-acids (3) that have not been reported before.

In the relatively small percent identity range of 72-82%, it is expected to find mammalian insulin sequences. The fact that reptilian, avian and amphibian insulins are also represented in this range, albeit at the lower end, shows their relationship to platypus insulin. This relationship on a molecular level fits in with anatomical and biological studies of this remarkable egg-laying mammal that has retained several ancestral reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988). It is interesting to note that the sponge insulin is also to be found in this range and has a percentage identity of 74% with platypus insulin. It also shares a B30 Met with platypus insulin (Table 3.2). Also the relative percent identity of platypus insulin compared with that of the hystricomorph insulins was the lowest, namely 54-60% (Table 3.3). The

conclusion that could be drawn from this pairwise sequence comparison study is that the relationship of the platypus insulin with that of the hystricomorph insulins is relatively low.

3.4.3 Multiple sequence alignment of platypus insulin with that of all known insulin sequences

On taxonomic grounds it is believed that the monotremes are the last survivors of a group of early mammals that evolved independently of the creatures that gave rise to today's marsupials and other mammals (Grant, 1989; Griffiths, 1988). The PILEUP dendogram, displaying interrelationships of insulin sequences from different animals (Figure 3.2), reflects this theory. Although the platypus insulin sequence is not part of subcluster 2.2 that contains the similar sequences of human, porcine, and elephant etc. and from which some rodent and marsupial insulin sequences are excluded, its closest relationship appears to be with the mammalian insulin sequence as a single member of a more distantly related branch. (cluster 2 of Figure 3.2 and highest percent identity, Table 3.3). Its relationship with reptilian insulin sequences (and amphibian and avian insulin sequences in this case) is sufficiently close to support the observation that platypus has retained some ancient reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988). The insulin sequence of the other living member of the monotremes the echidna, two species *Tachyglossus aculeatus* and *Zaglossus bruijnii* (Grant, 1989; Griffiths, 1988) are still unknown. It would be expected that the sequence is similar (or identical) to that of platypus and thus should subcluster with it. In support of this proposal is the similar subclustering of the mouse and rat in subcluster 2.1 as well as the snakes in subcluster 5.1 (own nomenclature). The results of the sequence comparison studies fits in with known taxonomic data used to classify and compile a hypothesis for the evolutionary relationship of platypus with these animals (Grant, 1989; Griffiths, 1988).

3.4.4 A comparison of the computer molecular model of the three-dimensional structure platypus insulin with the crystal structure of porcine insulin

Comparison of the structure of the molecular model (monomer 2) with that of porcine insulin reveals it to be different. The A2-8 α-helix seems more distorted and the B21-23 β-turn more irregular. The C-terminal B28-B30 end of the B-chain, which is known to be flexible, folds slightly back towards the main body of the monomer. A two-fold axis view reveals residues A13, A14, A17, B1 and B2 (part of the dimer-dimer interface) to be involved in a different spatial arrangement from porcine insulin (Figure 3.3). The non polar nature has been preserved in the two amino-acids (A13 Met and B2 Pro) of platypus insulin that differ from porcine insulin in this interface. In the porcine insulin crystal there exists an A14 Tyr-B1 Phe interaction in the dimer-dimer interface (Baker *et al.*, 1988; Blundell *et al.*, 1972), although in 2D NMR studies of insulin monomers in solution these residues seem to be flexible and the interaction unlikely to be maintained in solution (Weiss *et al.*, 1989). In the platypus insulin model A13 Met is sandwiched between A14 and B1 and the B1 Phe is pushed away from the body of the monomer (Figure 3.3). Such structural differences should be interpreted with caution and thus a postulated altered dimer-dimer interaction as well as altered self-association properties of platypus insulin needs to be verified experimentally under solution conditions when sufficient material is available.

Viewing the space filling models of both platypus and porcine insulin, by looking directly at the receptor binding region, revealed no marked difference. As viewed from the two-fold axis porcine insulin is mainly heart shaped while platypus insulin has a squarer structure. The invariant amino-acids responsible for the stable core structure of insulin (Baker *et al.*, 1988) are identical in both species.

3.4.5 Conclusion

The biological features of platypus insulin indicate that it is more closely related to marsupial and placental mammals than to any group of reptiles, despite retaining several ancestral reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988). There is also only a relatively distant relationship with the hystricomorph insulins. This study of the amino-acid sequence of platypus insulin showed it to be quite varied, one that might possibly lend itself to the study of identifying the desirable properties needed to produce an insulin that is monomeric in solution (Chapters 1 and 2). Platypus insulin differs from that of the human at eleven amino-acid sites; of which three residues have not been observed in other organisms. It also differs from porcine and human insulin at its self-associating surfaces (monomer-monomer and dimer-dimer interfaces) by 4 residues, offering the possibility of a change in association properties, and at its receptor binding surface by 2 residues, offering the possibility of a change in biological activity. It might be of interest to note that the B25 Tyr has been postulated as participating in growth promoting activity (Chapter 2) and this residue is present in platypus insulin. In designing an insulin for therapeutic use the possibility of growth promoting activity must be taken into account.

In conclusion, platypus insulin's self-association behaviour under physiological conditions needs to be evaluated. The influence of the changes in amino-acid composition on biological activity and receptor binding characteristics of platypus insulin also requires an experimental study. Platypus insulin possesses a unique evolutionary-designed primary structure. The study of the ramifications of this, especially its molecular interactions and physico-chemical behaviour in its physiological environment, namely in solution, blood, and in contacts with residues of its receptor upon binding to elicit a biological response, should contribute to our understanding of the structure-function relationship of insulin in general. The combination of knowledge gained from the resources of evolutionary designed and protein-engineered insulins will in turn contribute to a more informed

insulin design and evaluation strategy in the quest for a monomeric, active, stable and immunologically safe insulin for diabetes therapy.

Platypus insulin and the HB10D mutant have been chemically synthesised (Chapter 4) based on its known sequence. Sufficient material has been obtained for *in vitro* biological activity tests (Chapter 4) and self-association studies in solution (Chapter 5).

Chapter 4

Total Chemical Synthesis and Characterisation of Platypus and HB10D Platypus Insulins

4.1 Introduction

Platypus insulin has been isolated to purity by others but the amount obtained was very limited; being sufficient only to determine the amino-acid sequence and mass (Nourse *et al.*, 1996).

To investigate the biological and physico-chemical properties of platypus insulin relatively large amounts (mg quantities) of insulin were needed. For ethical and moral reasons the purification of insulin from its natural source, the pancreases of platypus, was not contemplated on this scale since it is a unique and protected animal of Australia and pancreas are available only rarely from animals that have died of natural causes. Alternative means of obtaining the required amounts of this insulin were contemplated, namely by chemical synthesis and DNA recombinant technology. For these studies its synthesis by totally chemical means was chosen. This had the potential advantage of being able to make mutants of the insulin using non-natural amino-acids.

The A- and B-chains were synthesised separately by an automatic solid phase peptide synthesiser using standard Fmoc chemistry and acetamidomethyl (Acm) protection of cysteines Cys A6, Cys A11, Cys A20 and Cys B19. The sulfhydryl groups of the other cysteines Cys A7 and Cys B7 were not protected. In the synthesis a strategy of the stepwise semi-regioselective formation of the disulfide bonds was employed. This strategy was based on the chemical synthesis of bombyxin II (an insulin-like molecule) (Nagata *et al.*, 1992). The CysA7-CysB7 disulfide bond was formed first, followed by

the removal of Acm groups on the other cysteines and concomitant formation of the two disulfide bonds by means of iodine oxidation.

Although the published methods work for synthesising bombyxin it was found that they were not suitable for synthesising platypus insulin. With platypus it was necessary to introduce numerous critical modifications and optimise other reaction steps. These for the AB chain combination reaction were:

- Addition of acetonitrile to the prescribed buffers to aid solubilization of the A- and B-chain derivatives
- An accurate determination of the A- and B-chain molar ratios.

Other modifications were:

- More dilute reaction mixtures for all reaction steps than prescribed by Nagata *et al.* (Nagata *et al.*, 1992)
- Optimising the reaction time of the oxidation reaction responsible for the formation of the other two disulfide bonds. This was found to be longer with platypus and even longer for the mutant platypus insulin
- Modifying the RP-HPLC separation conditions and gradient for purification of the final reaction product, namely insulin.

The method of total chemical synthesis employed yielded mg quantities of purified platypus insulin. Preliminary investigations into the self-association of synthetic platypus insulin by sedimentation analysis pointed to some reduction in the self-association at the dimer-dimer interface compared to bovine insulin (Chapter 5). Furthermore, observed differences in the spatial arrangement of residues in this interface of the molecular model (Chapter 3) also suggested that the contribution to the self-associating properties of this region may be lowered. This prompted an investigation of the properties of the dimer-dimer interface with platypus insulin analogues designed to probe that region. It was anticipated that the introduction of different residues, including non-proteinogenic amino-acids, into this interface would modulate self-association without having an adverse

effect on the biological activity. Consequently the total chemical synthesis of HB10D platypus insulin (B10 His is substituted for an Asp in the dimer-dimer interface) was also undertaken to analyse the self-association pattern of this mutant. A similar HB10D human insulin (Schwartz *et al.*, 1987) displayed a reduction in self-association properties (Weiss *et al.*, 1991) but at the same time had increased biological activity *in vitro*.

Both synthetic insulins were obtained in a pure form and biological activity and the circular dichroism spectra were determined.

4.2 Materials and methods

4.2.1 The total chemical syntheses of platypus and HB10D platypus insulin.

4.2.1.1 Materials

Protected amino-acids, Fmoc-Asn-OCH_2 phenoxymethyl and Fmoc-Met-OCH_2 phenoxymethyl resin were purchased from Applied Biosystems (Applied Biosystems Inc., Foster City, CA, U.S.A.). 2,2'-dithiodipyridine(DTDP) was from Sigma (Sigma Chemical Company, Sigma-Aldrich Pty Ltd., Castle Hill, NSW, Australia) and α-cyano-4-hydroxycinnamic acid from Aldrich (Aldrich Chemical Company Inc., Castle Hill, NSW, Australia). Collagenase A (from *C. histolyticum*) was from Boehringer Mannheim (Boehringer Mannheim Australia Pty Ltd., Castle Hill, NSW, Australia). Other reagents and solvents were of the highest quality available and were used without further purification.

4.2.1.2 Analytical procedures

Preparative HPLC

All the preparative reversed phase (RP) HPLC experiments were performed with a BIO-RAD Model 700 Chromatography Work Station (Bio-Rad Laboratories, Richmond, CA, U.S.A) with 1355T Soft-Start pumps and a Bio-Rad Model 1706 UV/VIS Monitor. Samples (peptides) were applied to a 5x300 mm C8 83-323-CHO 10087 Dynamax-300Å pore column (25x300 mm) from Rainin (Rainin Instrument Co. Inc., Emerville, CA, U.S.A), eluted at room temperature at a flow rate of 24 ml/min by gradients of acetonitrile (CH_3CN) and monitored at 215 nm. Mass spectra of respective peaks were obtained and analysed by matrix-assisted laser desorption ionisation time of flight-mass spectrometry (MALDI-TOF-MS) using a VG Tofspec Fisons Analytical Instrument (Fisons Instruments, Manchester, England).

The separation conditions (buffers and gradients) of the peptides on the preparative RP-HPLC column are shown in Tables 4.1 and 2. The various peaks were collected, molecular mass analysis performed to confirm the identities of the appropriate peptides, usually in the major peaks, (Table 4.4) and lyophilised. The peaks containing [Cys(Acm)$^{A6, A11, A20}$] A-chain, [Cys(Acm)^{B19}] and [Cys(Acm)^{B19}][HB10D] B-chains were lyophilised in the presence of 0.1% (v/v) β-mercaptoethanol to prevent oxidation.

Analytical HPLC

Analytical RP-HPLC was performed with a BIO-RAD Model 700 Chromatography Work Station with 1350T Soft-Start pumps and a Bio-Rad Model 1706 UV/VIS Monitor. Both platypus and HB10D platypus insulin were individually applied to a Waters RCM column (8x10 mm, Millipore), eluted at room temperature at a flow rate of 1 ml/min in 0.09% TFA with a linear 10-50% CH_3CN gradient within 30 minutes

and monitored at 214 nm. Ultimate product purity was assessed by analytical RP-HPLC, mass analysis and high performance capillary electrophoresis (HPCE).

4.2.1.3 Total chemical synthesis of platypus and HB10D platypus insulin

4.2.1.3.1 Strategy of semi-regioselective disulfide bond formation

The strategy of the semi-regioselective disulfide bond formation of insulin is shown in a diagram (Figure 4.1). This strategy was followed for both the syntheses of the native and mutant platypus insulin. The [Cys(Acm)$^{A6, A11, A20}$] A-chain (*S*-Acm A-chain) and the [Cys(Acm)^{B19}] B-chains (*S*-Acm B-chain) were synthesised separately by an automatic solid phase peptide synthesiser. During the synthesis the trityl-function was used as a protecting group for the sulfhydryl groups of cysteines Cys A7 and Cys B7 and the acetamidomethyl for Cys A6, A11, A20 and Cys B19. Trityl is TFA labile while acetamidomethyl is I_2 labile (Nagata *et al.*, 1992). Cleavage of the peptides from the resin support by standard TFA methods (King *et al.*, 1990; Applied Biosystems, 1988) removed all protecting groups except the *S*-Acm groups and left free the thiol functions of the Cys A7 *S*-Acm A-chain and the Cys B7 of the *S*-Acm B-chain. The free thiols of Cys B7 of the *S*-Acm B-chain peptides were first subjected to *S*-2-pyridinesulfenylation with 2,2'-dithiodipyridine (DTDP) (Nagata *et al.*, 1992; Maruyama *et al.*, 1990) to give [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain which were then coupled with equivalent amounts of [Cys(Acm)$^{A6, A11 A20}$] A-chain to yield the disulfide bond CysA7-CysB7 of [Cys(Acm)$^{A6, A11, A20, B19}$] insulin (Ruiz-Gayo *et al.*, 1988). The *S*-Acm groups were then removed and the other disulfide bonds formed by iodine oxidation in glacial acetic acid and HCl (Nagata *et al.*, 1992; Kamber *et al.*, 1980).

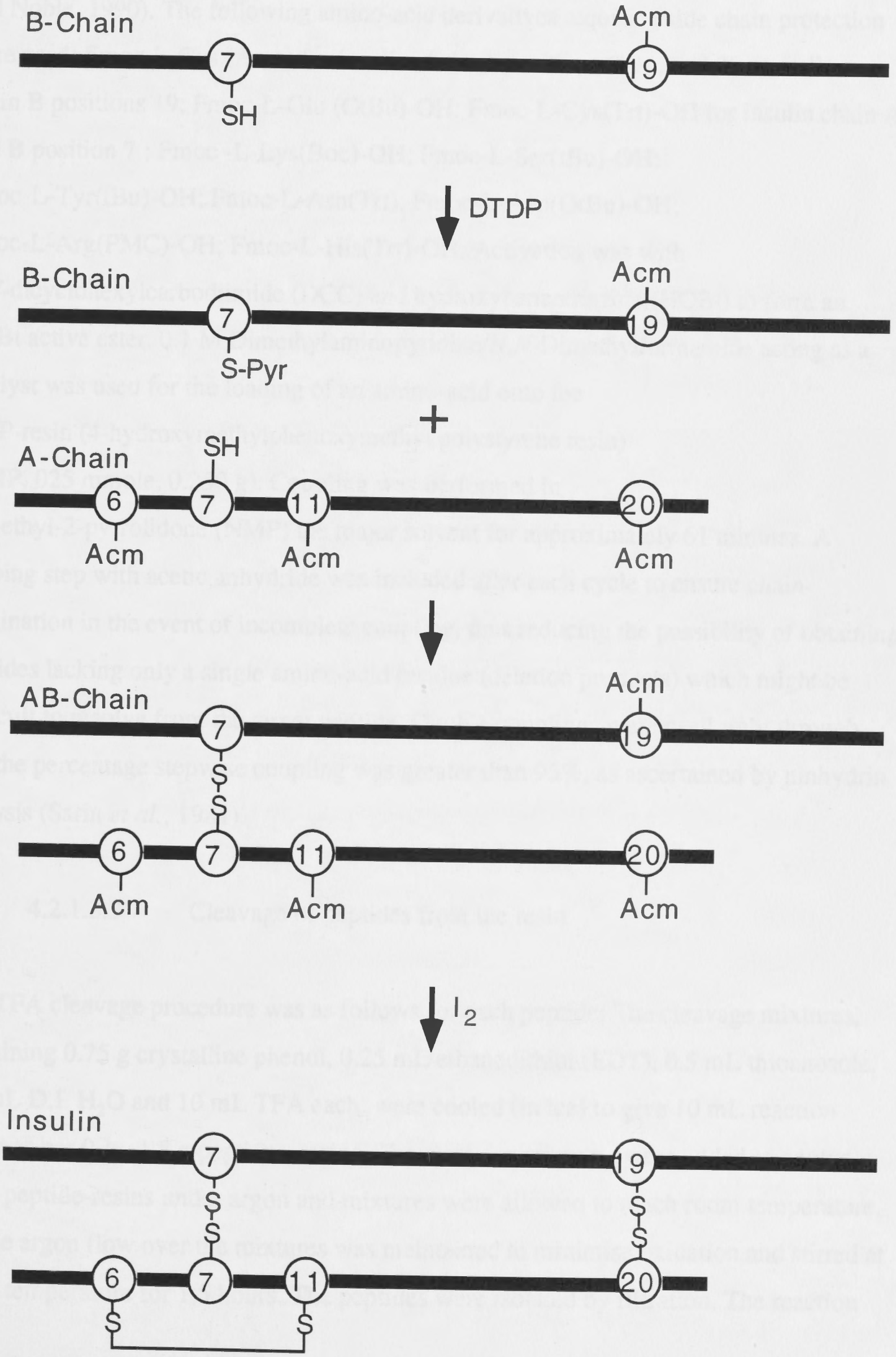
4.2.1.3.2 Solid-phase synthesis of the peptide chains

The solid phase peptide synthesis of [Cys(Acm)$^{A6, A11, A20}$] A-chain, [Cys(Acm)^{B19}] and [Cys(Acm)^{B19}][HB10D] B-chains of platypus and HB10D platypus insulin were carried

Figure 4.1

Strategy of the semi-regioselective disulfide bond formation.

Figure 4.1


B-Chain
Acm
7
19
SH
DTDP
B-Chain
Acm
7
19
S-Pyr
+
A-Chain
SH
6
7
11
20
Acm
Acm
Acm
AB-Chain
Acm
7
19
S
S
6
7
11
20
Acm
Acm
Acm
I2
Insulin
7
19
S
S
S
S
6
7
11
20
S
S

out on a Model 430A peptide synthesiser (Applied Biosystems, Foster City, CA, U.S.A) using the Fmoc α-amino group protecting strategy (Chang and Meienhofer, 1978; Fields and Noble, 1990). The following amino-acid derivatives requiring side chain protection were used; Fmoc-L-Cys (Acm) for insulin chain A positions 6, 11, and 20 as well as chain B positions 19; Fmoc-L-Glu (OtBu)-OH; Fmoc-L-Cys(Trt)-OH for insulin chain A and B position 7 ; Fmoc -L-Lys(Boc)-OH; Fmoc-L-Ser(tBu)-OH; Fmoc-L-Tyr(tBu)-OH; Fmoc-L-Asn(Trt); Fmoc-L-Asp(OtBu)-OH; Fmoc-L-Arg(PMC)-OH; Fmoc-L-His(Trt)-OH. Activation was with *N*,*N*-dicyclohexylcarbodiimide (DCC) and hydroxybenzotriazole (HOBt) to form an HOBt active ester. 0.1 M Dimethylaminopyridine/*N*,*N*-Dimethylformamide acting as a catalyst was used for the loading of an amino-acid onto the HMP-resin (4-hydroxymethylphenoxymethyl polystyrene resin) (HMP, 025 mmole, 0.250 g). Coupling was performed in *N*-methyl-2-pyrrolidone (NMP) the major solvent for approximately 61 minutes. A capping step with acetic anhydride was included after each cycle to ensure chain-termination in the event of incomplete coupling, thus reducing the possibility of obtaining peptides lacking only a single amino-acid residue (deletion products) which might be difficult to resolve from the target peptide. Double couplings were used right through and the percentage stepwise coupling was greater than 95%, as ascertained by ninhydrin analysis (Sarin *et al.*, 1981).

4.2.1.3.3 Cleavage of peptides from the resin

The TFA cleavage procedure was as follows for each peptide: The cleavage mixtures, containing 0.75 g crystalline phenol, 0.25 mL ethanedithiol (EDT), 0.5 mL thioanosole, 0.5 mL D.I. H_2O and 10 mL TFA each, were cooled (in ice) to give 10 mL reaction volumes per 0.1 - 1.5 g peptide - resins. The cleavage mixtures were added to cooled dried peptide-resins under argon and mixtures were allowed to reach room temperature. Gentle argon flow over the mixtures was maintained to minimise oxidation and stirred at room temperature for 1.5 hours. The peptides were isolated by filtration. The reaction

mixtures were then vacuum filtered through a medium-porosity, fritted glass funnel to separate them from the resin supports. The remaining resins were washed with 1 mL TFA. The filtrates were concentrated on a rotary evaporator under high vacuum to volumes of approximately 1 to 2 mL. The temperature of the water bath was maintained below 40 °C to prevent damage of the peptides. 50 mL of cold ethanol was added to concentrated reaction mixtures to precipitate the peptides.

Collection and purification of [Cys(Acm)$^{A6, A11, A20}$] A-chain

This A-chain derivative was collected by filtration, dried and redissolved in 10 mL 3 M Guanidinium/thiocyanate, 100 mM KH_2PO_4 pH 4.0, and the pH was adjusted to 7 to aid dissolution. This represented the aqueous phase. The phenol and other organic chemicals were extracted with chloroform and methanol followed by tert-butyl methyl ether. The A-chain derivative was dissolved in the aqueous phase and isolated from this phase by a preparative RP-HPLC column, purification method I (Table 4.1 and 2). β-mercaptoethanol was added to prevent oxidation before lyophilising. Analysis by MALDI mass spectrometry confirmed a molecular mass of 2598 daltons.

Collection and purification of [Cys(Acm)^{B19}] and [Cys(Acm)^{B19}][HB10D] B-chains

After filtering and drying, the native and mutant B-chain derivatives were further purified separately. The peptides were dissolved separately in about 15 mL glacial acetic acid, the solutions diluted to 40% acetic acid, centrifuged (the precipitates contained minimal amounts of peptide) and the supernatants applied to a 51x16 mm (i.d.) Sephadex G10 column. The [Cys(Acm)^{B19}][HB10D] B-chain was markedly less soluble under these conditions. The peptides were eluted in 40% acetic acid and collected in the void volume peak. β-mercaptoethanol was added to prevent oxidation and the peptides lyophilised. Analysis with the MALDI-TOF mass spectrometer confirmed molecular masses of 3557 and 3535 daltons for [Cys(Acm)^{B19}] and [Cys(Acm)^{B19}][HB10D] B-chain respectively.

The native and mutant B-chain derivatives were further purified on a preparative RP-HPLC column as follows. The lyophilisates were dissolved separately, firstly in CH_3CN plus 0.09% TFA, sonicated and this diluted to 20% CH_3CN plus 0.09% aqueous TFA, (total volume of 100 mL; approx. 0.75 mg/mL and 260 mL; approx. 0.27 mg/mL respectively). The solutions were applied separately to the column in several runs. The purification methods II and III (Table 4.1 and 2) for the native and mutant B-chain derivatives respectively were used. The identities of the peptides eluted in the major peaks of different runs were confirmed by molecular mass analysis (3557 and 3535 daltons respectively).

4.2.1.3.4 *S*-2-Pyridinesulfenylation of both native and mutant [Cys(Acm)^{B19}]B-chains at Cys B7 to form their [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain derivatives

The *S*-2-Pyridinesulfenylation of each of [Cys(Acm)^{B19}] native and [Cys(Acm)^{B19}][HB10D] B-chains was carried out using 50 mg (14 µmole) and 25 mg (7.1 µmole) respectively of each of the peptides by the following procedure: The process involved making two solutions. In one the particular peptide was first dissolved in 0.5 M ammonium acetate buffer (pH 4.0)/acetonitrile (65:35; v/v, 100 mL). In the other 2,2'-Dithiodipyridine (DTDP, 330 mg) was dissolved in acetonitrile (17.5 ml) to which was added 0.5 M ammonium acetate buffer (pH 4.0, 32.5 ml) to give a 30 mM DTDP solution. The two solutions were added together and the mixture stirred at room temperature for 20 minutes. The [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] [HB10D] B-chain was less soluble and stirring needed to be continued for another 40 minutes. The resulting peptide solutions were first diluted to 20% CH_3CN with 0.09% aqueous TFA and then subjected to preparative RP-HPLC using purification method II for the native B-chain derivative and III for the mutant B-chain derivative. The appropriate peak containing the [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain of both the native and the mutant samples was collected, lyophilised and their molecular masses confirmed by mass analysis as 3667 and

Table 4.1: Buffers with gradients used during the synthesis: Method I-IV; starting buffer A for 2 minutes
Method V-VI; starting buffer A for 8 minutes.

Method	Buffer A	Buffer B	Gradient
I	5 % CH_3CN in 30 % aqueous TEAA	CH_3CN	5-40 % CH_3CN (30 min)
II	0.09 % TFA in 20 % aqueous CH_3CN	0.09 % TFA in CH_3CN	20-60 % CH_3CN (30 min)
III	0.09 % TFA in 20 % aqueous CH_3CN	0.09 % TFA in CH_3CN	20-50 % CH_3CN (30 min)
IV	0.09 % TFA in 10 % aqueous CH_3CN	0.09 % TFA in CH_3CN	10-60 % CH_3CN (30 min)
V	20 % CH_3CN in 100 mM H_3PO_4/ 20 mM TEA/50 mM $NaClO_4$ pH 3.0	40 % CH_3CN in 100 mM H_3PO_4/ 20 mM TEA/50 mM $NaClO_4$ pH 3.0	20-28 % CH_3CN (2 min); 28-33 % CH_3CN (30 min); 33- 40 % CH_3CN (2 min); 40 % CH_3CN (8 min).
VI	20 % CH_3CN in 100 mM H_3PO_4/ 20 mM TEA/50 mM $NaClO_4$ pH 3.0	40 % CH_3CN in100 mM H_3PO_4/ 20 mM TEA/50 mM $NaClO_4$ pH 3.0	20-29 % CH_3CN (2 min); 29-34 % CH_3CN (30 min); 34- 40 % CH_3CN (2 min); 40 % CH_3CN (8 min).

Table 4.2: Methods used in the preparative RP-HPLC purification of peptides

Method	Peptide/Protein
I	[Cys(Acm)$^{A6, A11, A20}$] A-chain
II	[Cys(Acm)^{B19}] B chain [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B chain [Cys(Acm)$^{A6, A11, A20, B19}$] platypus inulin
III	[Cys(Acm)^{B19}][HB10D] B-chain [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}][HB10D] B-chain [Cys(Acm)$^{A6, A11, A20, B19}$][HB10D] platypus insulin
IV	Platypus and HB10D platypus insulin (desalting step)
V	Platypus insulin (I_2 oxidation step)
VI	HB10D platypus insulin (I_2 oxidation step)

3645 daltons respectively. The B-chain mutant derivative was less soluble in these solvents.

4.2.1.3.5 The combination of [Cys(Acm)$^{A6, A11, A20}$]A-chain with [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain of native or mutant to give their respective [Cys(Acm)$^{A6, A11, A20, B19}$] insulins

Molar ratios are critical to the outcome of this bond formation reaction. Amino-acid analysis (performed by the Biomolecular Resource Facility at the John Curtin School of Medical Research, Institute of Advanced Studies, Australian National University, Australia) was used to measure accurately the concentrations of the solutions containing the relevant peptides. From this the molar extinction coefficients of 1.0 M peptide solutions at 280 nm in the particular solvent had been calculated; ε_{280}=16500 M^{-1} cm^{-1} for [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain in 40% CH_3CN, 0.5 M Tris·HCl, pH 8.6 and ε_{280}=4300 M^{-1} cm^{-1} for [Cys(Acm)$^{A6, A11, A20}$] A-chain in 40% CH_3CN, 5 mM HCl. In the concentration measurements of the mutant B-chain derivative the ε_{280} of the native B-chain derivative was used.

The [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chains of each of the native and mutant preparations were dissolved in 40% CH_3CN plus 0.5 M Tris·HCl, pH 8.6. With the native sample 12 µmole (44 mg) was dissolved in 20 mL while with the mutant 2.7 µmole (10 mg) was dissolved in 32 mL. [Cys(Acm)$^{A6, A11, A20}$] A-chain was dissolved in 40% CH_3CN plus 5 mM HCl. For that to be used with the native B-chain derivative sample an equimolar amount (31 mg) was dissolved in 30 mL while with the mutant 14 mg was dissolved in 48 mL giving a B/A chain molar ratio of 1.4. To combine the chains the solutions of [Cys(Acm)$^{A6, A11, A20}$] A-chain were added dropwise to the solutions of [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chains of both the native and mutant samples separately with continuous mixing for 60 minutes at 45 °C. The resultant reaction mixtures (total volume 50 mL and 80 mL for the native and mutant respectively) were diluted with an equal volume of 0.09% aqueous TFA to give a 20% CH_3CN solution and subjected to

preparative RP-HPLC. The purification method II was used for the native sample and method III for the mutant sample (Table 4.1 and 2). The main peaks were collected and confirmed to be [Cys(Acm)$^{A6, A11, A20, B19}$] platypus and [Cys(Acm)$^{A6, A11, A20,B19}$][HB10D] platypus insulin by molecular mass analysis where the mass values were measured to be 6153 and 6131 daltons respectively.

4.2.1.3.6 Iodine oxidation of both native and mutant [Cys(Acm)$^{A6, A11, A20, B19}$] platypus insulin to give native and mutant platypus insulins

50 mg (8 μmole) of [Cys(Acm)$^{A6, A11, A20, B19}$] platypus and 15 mg (2.4 μmole) [Cys(Acm)$^{A6, A11, A20, B19}$][HB10D] platypus insulins respectively were dissolved in 1.5 mL and 3 mL H_2O, after which glacial acetic acid (10.5 mL and 21 mL respectively) and 2 M HCl (60 and 120 μL respectively) were added. A solution of iodine, 13 mM, in glacial acetic acid (18 mL and 36 mL respectively), was added dropwise to these solutions with continual shaking at 45 °C, and continued for 60 minutes with the native and 90 minutes with the mutant. The oxidation reaction was terminated by the addition of 0.25 M L-ascorbic acid (6 mL and 12 mL respectively). The acetic acid was evaporated off in a vacuum centrifuge, the Speed-Vac concentrator (Savant Instruments Inc., Hicksville, NY, U.S.A), until the volumes were 1 mL and 8 mL respectively. To these concentrated solutions a further addition of 0.25 M L-ascorbic acid (1 mL and 3.6 mL respectively) was made to ensure the reduction of all the iodine. These reaction mixtures were diluted with 18 mL and 15 mL , respectively, of 100 mM H_3PO_4 /20 mM Triethylamine/50 mM $NaClO_4$ at pH 3.0 buffer containing 20% CH_3CN. The diluted samples were in turn applied to a preparative RP-HPLC column and purified using method V for native platypus and VI for the mutant platypus insulin (Table 4.1 and 2). In both cases essentially two major peaks were eluted, the smaller, at 31% CH_3CN (33% in the case of the mutant) and the larger at 40% CH_3CN (Figure 4.9 and Figure 4.10). The 31% or 33% CH_3CN peaks containing respectively the native and HB10D mutant platypus insulins. These were lyophilised and desalted in turn on a preparative RP-HPLC

column using method IV for the native sample and III for the mutant sample (Table 4.1 and 2). They eluted as a single symmetrical Gaussian peak that contained the specific insulin. These were confirmed by molecular mass analysis. (Found 5865; expected 5865 daltons for native platypus insulin and found 5844; expected 5843 daltons for the mutant.)

4.2.2 Characterisation of the synthesis products

4.2.2.1 Mass analysis of peptides

Mass analysis of peptides was performed using a matrix-assisted laser desorption ionisation time of flight mass spectrometer (MALDI-TOF-MS) with a maximum power ≥ 180 μJ. (It is a linear time of flight spectrometer that uses an ultraviolet nitrogen laser. The laser beam is approximately 150-400 μm^2 in size and power is continually variable by means of a diaphragm.) Positive molecular ions were detected in linear mode, with excitation at 330 nm when the sample was mixed with α-cyano-4-hydroxycinnamic acid matrix in 40% CH_3CN and crystallised on a target.

A sufficient amount (a speck on the tip of a spatula) of each peptide (lyophilised) to give a definable signal in the MALDI-TOF-MS was dissolved in 40% CH_3CN/H_2O (concentration in the pico molar range). The following calibration compounds were used as either internal or external references: GramicidinS (Sigma), FW 1141.5; $(M + H)^+$ 1142.5 and bovine insulin (Sigma), FW 5733.5; $(M + H)^+$ 5734.5; $(M + 2H)^{2+}$ 2867.75 daltons. Calibration compounds or unknown peptides were prepared together or separately for internal or external calibration by mixing 1-2 μL peptide solution(s) with about 10 μL of saturated matrix solution. The mixtures (3.5 μL) were crystallised on a stainless steel mass spectrometer target before loading onto the magnetic probe of the mass spectrometer. Mass spectra were obtained, calibrated with either the internal or external calibration compounds and the mass of the unknown noted.

4.2.2.2 Biological characterisation of synthetic insulins

Insulin stimulated lipogenesis was measured by the method of Moody *et al.* (Moody *et al.*, 1974) as the incorporation of [3-^{3}H]D-glucose into toluene-extractable lipids of rat epididymal fat cells. The [3-^{3}H]D-glucose was obtained from Amersham (Amersham International plc, Buckinghamshire, England) and the specific activity of the batch was 73.5 mCi/mg.

Isolation of Adipocytes

Adipocytes were isolated from epididymal fat pads of 2 male Sprague-Dawley rats (weight approximately 150 g) fed *ad libitum.* Fat pads were digested by gentle agitation of sliced tissues for 45 minutes in a Krebs-Ringer-Bicarbonate buffer containing 10 mM HEPES, 1.5 mg/mL collagenase A, 0.55 mM glucose and 35 mg/mL bovine serum albumin (fatty acid free, fraction V, from Sigma) at pH 7.4. The cell suspension was then passed through a nylon mesh (200 micron) and washed 3 times with Krebs-Ringer-Bicarbonate-HEPES (KRBH) buffer pH 7.4 containing 1.5% bovine serum albumin (BSA) with the last wash containing 3% BSA at room temperature. The cell suspension was diluted for use in the assay with KRBH buffer pH7.4 and the concentration (mg dry-weight cells/ mL) determined by weighing a blotted aliquot of cell suspension on filter paper after drying in a desiccator.

[3-^{3}H]D-glucose incorporation into cellular lipids

Cell suspensions (0.5 mL, corresponding to 10-20 mg dry-weight cells) (Joshi *et al.*, 1990) and varying concentrations of insulin (10 µL) were pre incubated in 5 ml polyethylene vials for 15 minutes at 37 °C in KRBH pH 7.4 buffer containing 3% BSA. The gas phase was 95% O_2; 5% CO_2 . [3-^{3}H]D-glucose (10 µL) was then added (final concentration 1.0 µCi/mL) and incubation continued in the presence of 95% O_2; 5% CO_2

for 60 minutes. The reaction was stopped by the addition of 4 mL of toluene-based liquid scintillant Econofluor (Du Pont Australia Ltd, Sydney, Australia) suitable for radioactivity measurements in non-aqueous solutions, thus extracting the newly formed ^{3}H-lipid without interference from water-soluble [3-^{3}H]D-glucose. Counting was in a Packard Tri-Carb 460. A standardised dose response curve of porcine insulin (Sigma) was constructed with the same cells. Duplicate experiments were performed and the data points for each indicated concentration represent the mean of duplicate determinations (Figures 4.16 and 4.17). Stimulation, expressed as percent of maximum, was plotted as a function of insulin concentration and data points were fitted to the sigmoidal equation of the program GraphPadPRISM with a non-linear regression algorithm and ED_{50} (the log[Dose] when response =50%) values for all the insulins calculated with this program (GraphPad Software, San Diego, CA, U.S.A). The relative potency on a molar basis for the synthetic insulins were calculated as follows:

Relative % Activity = ED_{50} (Porcine Insulin)/ED_{50} (Synthetic Insulin) x 100

(Equation 4.1)

4.2.2.3 Biophysical characterisation of synthetic insulins

Circular dichroism spectra of synthetic insulins

The secondary structure of the synthetic insulins were investigated using circular dichroism (CD). CD spectra were recorded with a Jobin-Yvon Mark V dichrograph calibrated with (+)-10-camphorsulfonic acid. The buffer used was 1.0 mM HCl pH 3.2. All insulin concentrations were determined photometrically (Varian Techtron 634) using the molar extinction coefficient, ε_{276}=6200 M^{-1} cm^{-1}, at 276 nm (Frank and Veros, 1968). All spectra were recorded at room temperature. Far-UV CD spectra were scanned from 185 nm to 250 nm, using a cell with a pathlength of 0.05 cm and the concentration of porcine, synthetic platypus and HB10D platypus insulins were 25, 24 and 50 μM respectively. All spectra were smoothed before subtraction of the appropriate solvent

blanks. The CD data (in differential dichroic absorption, $\Delta\varepsilon$ (M^{-1} cm^{-1})) were converted to mean residue ellipticity ($[\theta]^{MRW}$) using a mean residue weight (M_R) of 112 for both native and mutant platypus insulin according to the following formula:

$$[\theta]^{MRW} = (\theta^{\circ})(M_R)/10[C](L) \qquad 4.2$$

where θ° is the observed rotation in millidegrees, M_R is the mean residual weight in daltons, C is the protein concentration in mg/mL, L is the cell path length in centimetres, and $[\theta]^{MRW}$ is mean residue ellipticity and has the units of deg cm^2/dmole (Adler *et al.*, 1973). The converted data was then plotted as $[\theta]^{MRW}$ *versus* wavelength (nm) and the $[\theta]_{208}/[\theta]_{223}$ ratios were calculated with the GraphPadPRISM program. (For more details about this program see Chapter 5).

4.3 Results

4.3.1 Total chemical synthesis of platypus and HB10D platypus insulin

4.3.1.1 Solid-phase synthesis of the peptide chains

The peptide chains, $[Cys(Acm)^{A6, A11, A20}]$ A-chain; $[Cys(Acm)^{B19}]$ native and $[Cys(Acm)^{B19}]$[HB10D] B-chains, were synthesised by the solid -phase method using the Fmoc α-amino group protecting strategy (Chang and Meienhofer, 1978; Fields and Noble, 1990). The final weights of the peptides (on the resin) for the A-chain, native B-chain and mutant B-chain were 69 mg, 92 mg and 83 mg respectively and the % yields calculated from the theoretical yields of resin-peptide 75.3, 72.6 and 68.8% respectively (Table 4.3). After cleavage of the peptides from the resin the A-chain was partially purified on a preparative RP-HPLC column (Figure 4.2) and the native and mutant B-chains in turn on a Sephadex G10 column followed by a preparative RP-HPLC column (Figures 4.3 and 4). The total weight yields for the A-chain and the native and mutant

Figure 4.2

Preparative RP-HPLC of the resin-cleaved [Cys(Acm)$^{A6, A11, A20}$] A-chain of platypus insulin (arrow) on a Dynamax-300Å pore column at 24 mL/min. Purification method I (Table 4.1 and 2) was followed. The A-chain eluted at approximately 28% CH_3CN.

Figure 4.2

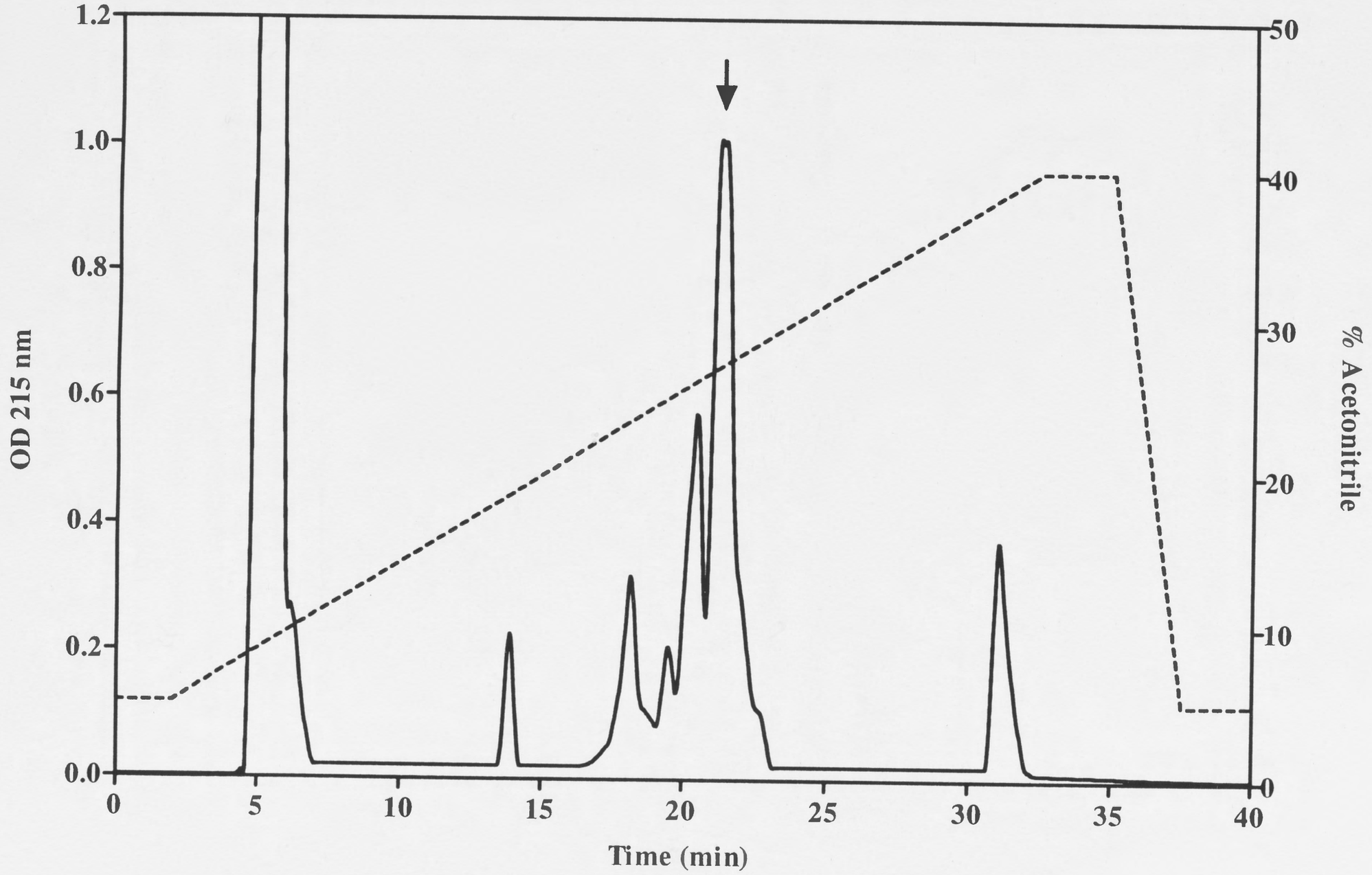

1.2
1.0
0.8
0.6
0.4
0.2
0.0
OD 215 nm
50
40
30
20
10
0
% Acetonitrile
0
5
10
15
20
25
30
35
40
Time (min)

Figure 4.3

Preparative RP-HPLC of [Cys(Acm)^{B19}] B-chain of platypus insulin (arrow). Purification method II (Table 4.1 and 2) was followed. The B-chain eluted at approximately 45% CH_3CN.

Figure 4.3

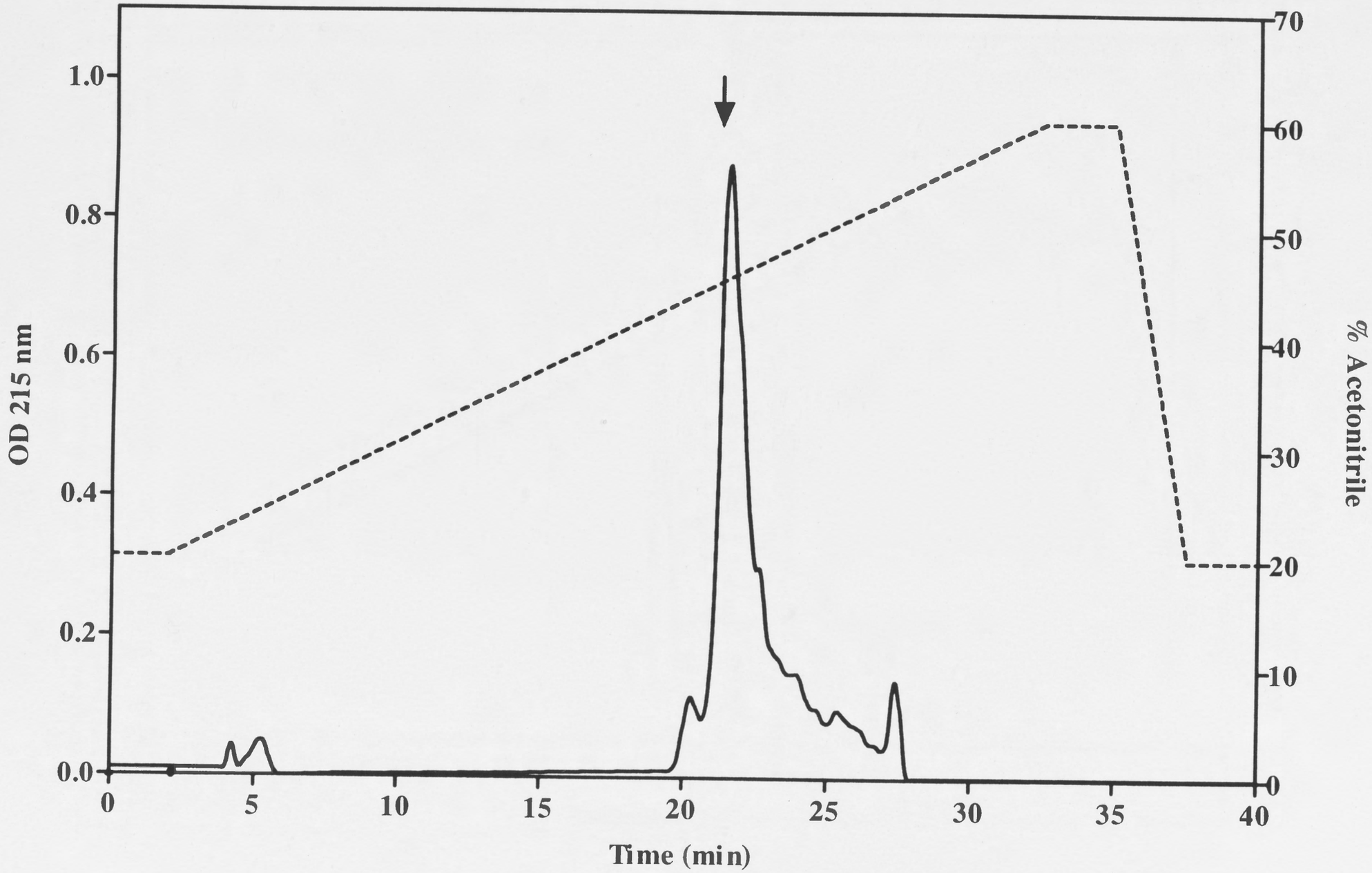
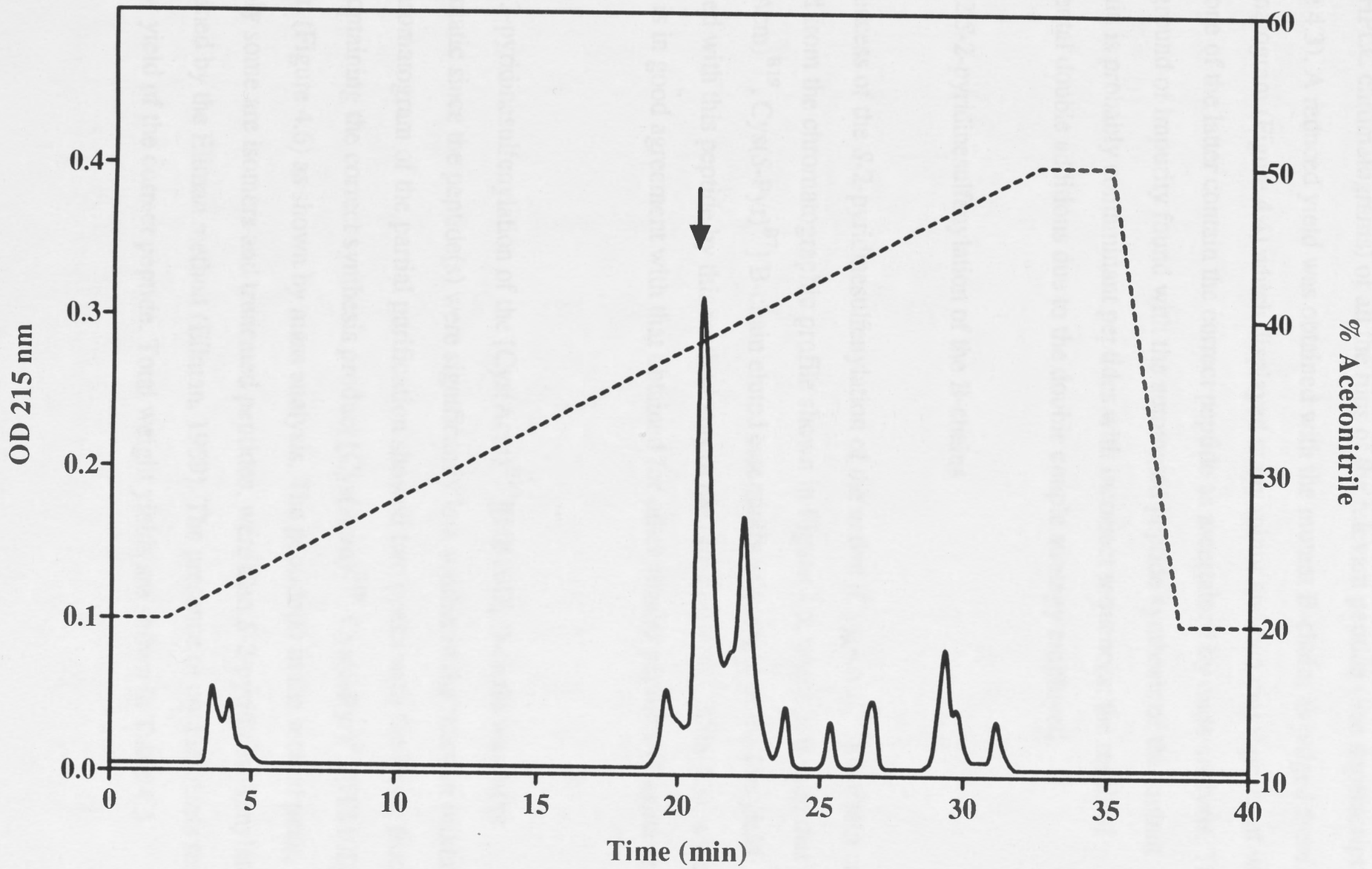

OD 215 nm
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
40
Time (min)
70
60
50
40
30
20
10
0
% Acetonitrile

Figure 4.4

Preparative RP-HPLC of [Cys(Acm)^{B19}][HB10D] B-chain of platypus insulin (arrow). Purification method III (Table 4.1 and 2) was followed. The B-chain eluted at approximately 38% CH_3CN.

Figure 4.4


OD 215 nm
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
Time (min)
60
50
40
30
20
10
% Acetonitrile

B-chains were approximately 50 mg, 65 mg and 45 mg respectively. Some weight determinations are only approximations based on the summation of the peak areas (RP-HPLC chromatograms) of all the runs of the relevant peptide (See superscript [d], Table 4.3). A reduced yield was obtained with the mutant B-chain, as judged from the chromatogram (Figure 4.4) which displayed many minor and two major peaks of which only one of the latter contain the correct peptide as ascertained by mass analysis. This background of impurity found with the automated peptide synthesis of the mutant B-chain is probably contaminant peptides with incorrect sequences; the result of accidental double additions due to the double couple strategy employed.

4.3.1.2 *S*-2-pyridinesulfenylation of the B-chains

The success of the *S*-2-pyridinesulfenylation of the native [Cys(Acm)^{B19}] B-chain can be judged from the chromatographic profile shown in Figure 4.5, where it is seen that the [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain eluted essentially as a single peak. The yield obtained with this peptide by this method was in the order of 90% (Table 4.3), a yield which is in good agreement with that obtained for other similar peptides (Nagata *et al.*, 1992).

The *S*-2-pyridinesulfenylation of the [Cys(Acm)^{B19}][HB10D] B-chain was more problematic since the peptide(s) were significantly less soluble in the reaction mixture. The chromatogram of the partial purification showed two peaks with the lesser front peak containing the correct synthesis product [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}][HB10D] B-chain (Figure 4.6) as shown by mass analysis. The peptide(s) in the second peak, probably some are isomers and truncated peptides, were also *S*-2-pyridinesulfenylated as ascertained by the Ellman method (Ellman, 1959). The presence of contaminants resulted in a low yield of the correct peptide. Total weight yields are shown in Table 4.3.

Table 4.3a: Approximate total weight and percentage yields of each step of the synthetic process.

Synthetic Platypus Insulin

	Wt (mg) Start	Wt (mg) Final	% Yield
[Cys(Acm)$^{A6, A11, A20}$] A-chain (on resin) (Resin-peptide → peptide)	69 [a)]	ND	
[Cys(Acm) $^{A6, A11, A20}$] A-chain (Cleavage from resin → RP HPLC)	ND	50 [d)]	72
[Cys(Acm)^{B19}] B-chain (on resin) (Resin-peptide → peptide)	92 [a)]	ND	
[Cys(Acm)^{B19}] B-chain (G10 → RP HPLC)	75 [b)]	65 [d)]	87 (70) [d)]
[Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain (*S*-2-pyridinesulfenylation → RP HPLC)	65 [d)]	60 [d)]	92
[Cys(Acm)$^{A6, A11, A20, B19}$] Insulin (CysA7-CysB7 → RP HPLC)	44 (B-chain) 31 (A-chain) [c)]	50 [d)]	67
Insulin (Iodine oxidation → RP HPLC)	50 [d)]	3.5 [d)]	7.0 [d)] (8)[e)]

Table 4.3b: Approximate total weight and percentage yields of each step of the synthetic process.

Synthetic HB10D Platypus Insulin

	Wt (mg) Start	Wt (mg) Final	% Yield
[Cys(Acm)$^{A6, A11, A20}$] A-chain (Cleavage resin → RP HPLC)	ND	50	
[Cys(Acm)^{B19}] B-chain (on resin) (Resin-peptide → peptide)	83 [a)]	ND	
[Cys(Acm)^{B19}] B-chain (G10 → RP HPLC)	70 [b)]	45 [d)]	60
[Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain (2-*S*-pyridinesulfenylation → RP HPLC)	45 [d)]	20 [d)]	44
[Cys(Acm)$^{A6, A11, A20, B19}$] Ins (Cys A7-Cys B7 → RP HPLC)	10 (B-chain) 14 (A-chain) [c)]	15 [d)]	63
Insulin (Iodine oxidation → RP HPLC)	15 [d)]	1.0 [d)]	6.7 [d)] (10) [e)]

Figure 4.5

Preparative RP-HPLC of [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain of platypus insulin (arrow). Purification method II (Table 4.1 and 2) was followed. The B-chain eluted at approximately 40% CH_3CN.

Figure 4.5

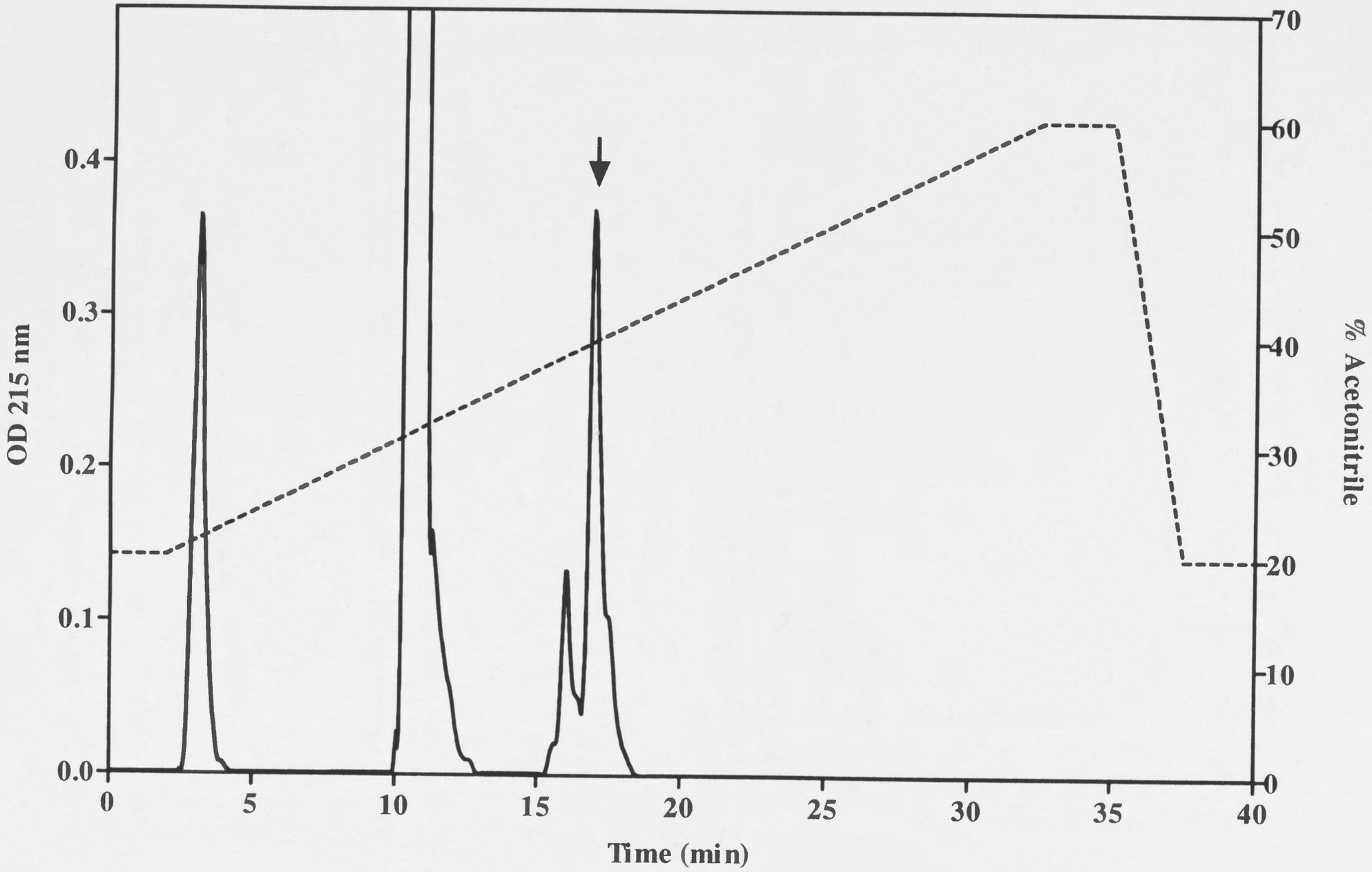
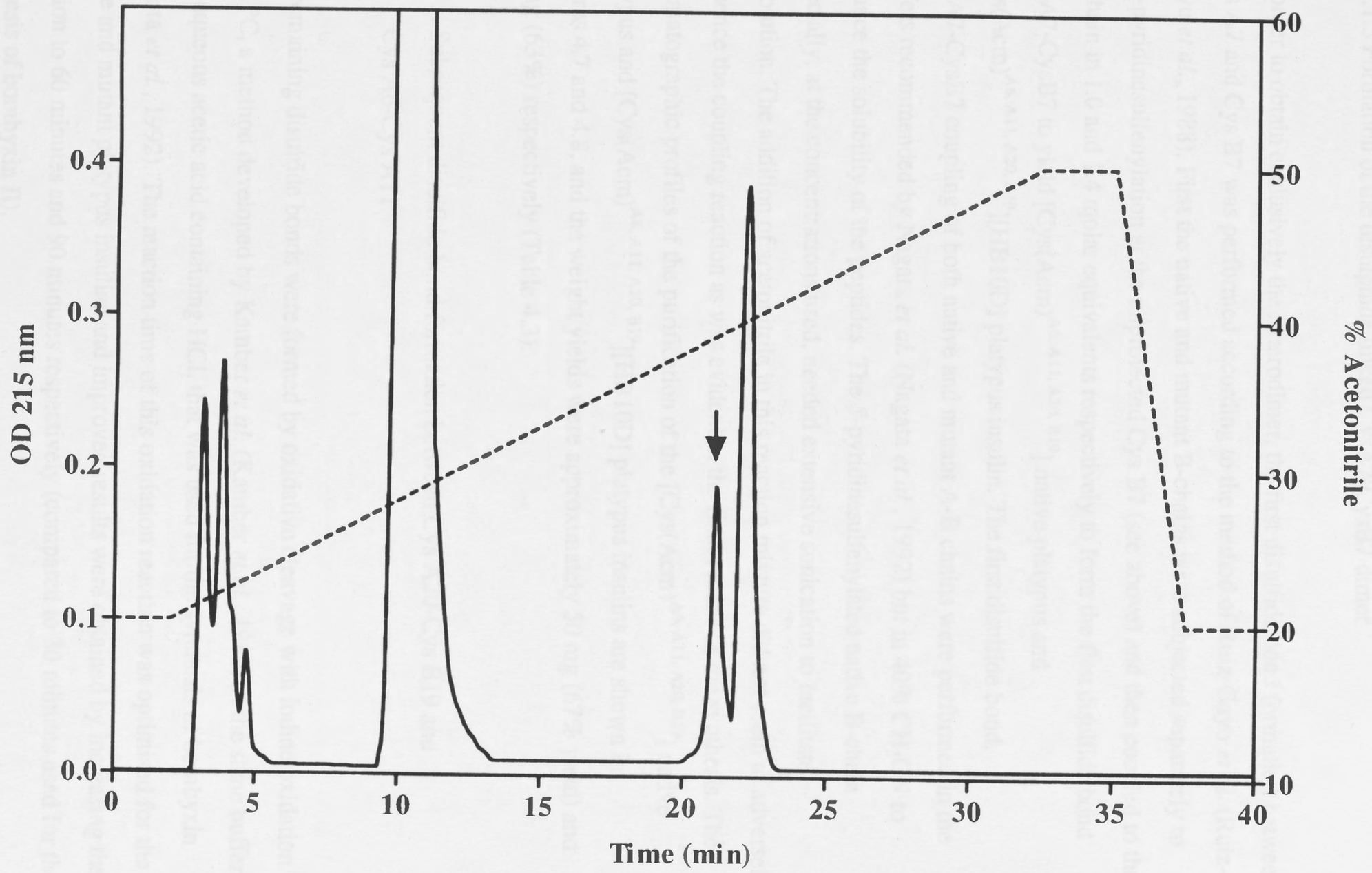

OD 215 nm
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
Time (min)
70
60
50
40
30
20
10
0
% Acetonitrile

Figure 4.6

Preparative RP-HPLC of [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}][HB10D] B-chain of platypus insulin (arrow). Purification method III (Table 4.1 and 2) was followed. The B-chain eluted at approximately 39% CH_3CN.

Figure 4.6


OD 215 nm
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
Time (min)
60
50
40
30
20
10
% Acetonitrile

4.3.1.3 Formation of the disulphide linked CysA7-CysB7 dimer

In order to obtain exclusively the heterodimer, the first disulfide bond formation between Cys A7 and Cys B7 was performed according to the method of Ruiz-Gayo *et al.* (Ruiz-Gayo *et al.*, 1988). First the native and mutant B-chains were subjected separately to *S*-2-pyridinesulfenylation at the unprotected Cys B7 (see above) and then coupled to the A-chain in 1.0 and 1.4 molar equivalents respectively to form the first disulfide bond CysA7-CysB7 to yield [Cys(Acm)$^{A6, A11, A20, B19}$] native platypus and [Cys(Acm)$^{A6, A11, A20, B19}$][HB10D] platypus insulin. The first disulfide bond, CysA7-CysB7 coupling of both native and mutant A-B chains were performed in the buffers recommended by Nagata *et al.* (Nagata *et al.*, 1992) but in 40% CH_3CN to enhance the solubility of the peptides. The *S*-pyridinesulfenylated native B-chain especially, at the concentrations used, needed extensive sonication to facilitate dissolution. The addition of acetonitrile in this reaction mixture did not seem to adversely influence the coupling reaction as was evident in the final results of the synthesis. The chromatographic profiles of the purification of the [Cys(Acm)$^{A6, A11, A20, B19}$] native platypus and [Cys(Acm)$^{A6, A11, A20, B19}$][HB10D] platypus insulins are shown in Figures 4.7 and 4.8, and the weight yields were approximately 50 mg (67% yield) and 15 mg (63%) respectively (Table 4.3).

4.3.1.4 Subsequent disulfide bond formation between Cys A20-Cys B19 and Cys A6-Cys A11

The remaining disulfide bonds were formed by oxidative cleavage with iodine oxidation at 45 °C, a method developed by Kamber *et al.* (Kamber *et al.*, 1980) in the same buffer, 95% aqueous acetic acid containing HCl, that was used for the synthesis of bombyxin (Nagata *et al.*, 1992). The reaction time of this oxidation reaction was optimised for the native and mutant platypus insulins and improved results were obtained by increasing the duration to 60 minutes and 90 minutes respectively (compared to 30 minutes used for the synthesis of bombyxin II).

Figure 4.7

Preparative RP-HPLC of

[Cys(Acm)$^{A6, A11, A20, B19}$] platypus insulin (arrow). Purification method II (Table 4.1 and 2) was followed. The AB-chain eluted at approximately 39% CH_3CN.

Figure 4.7

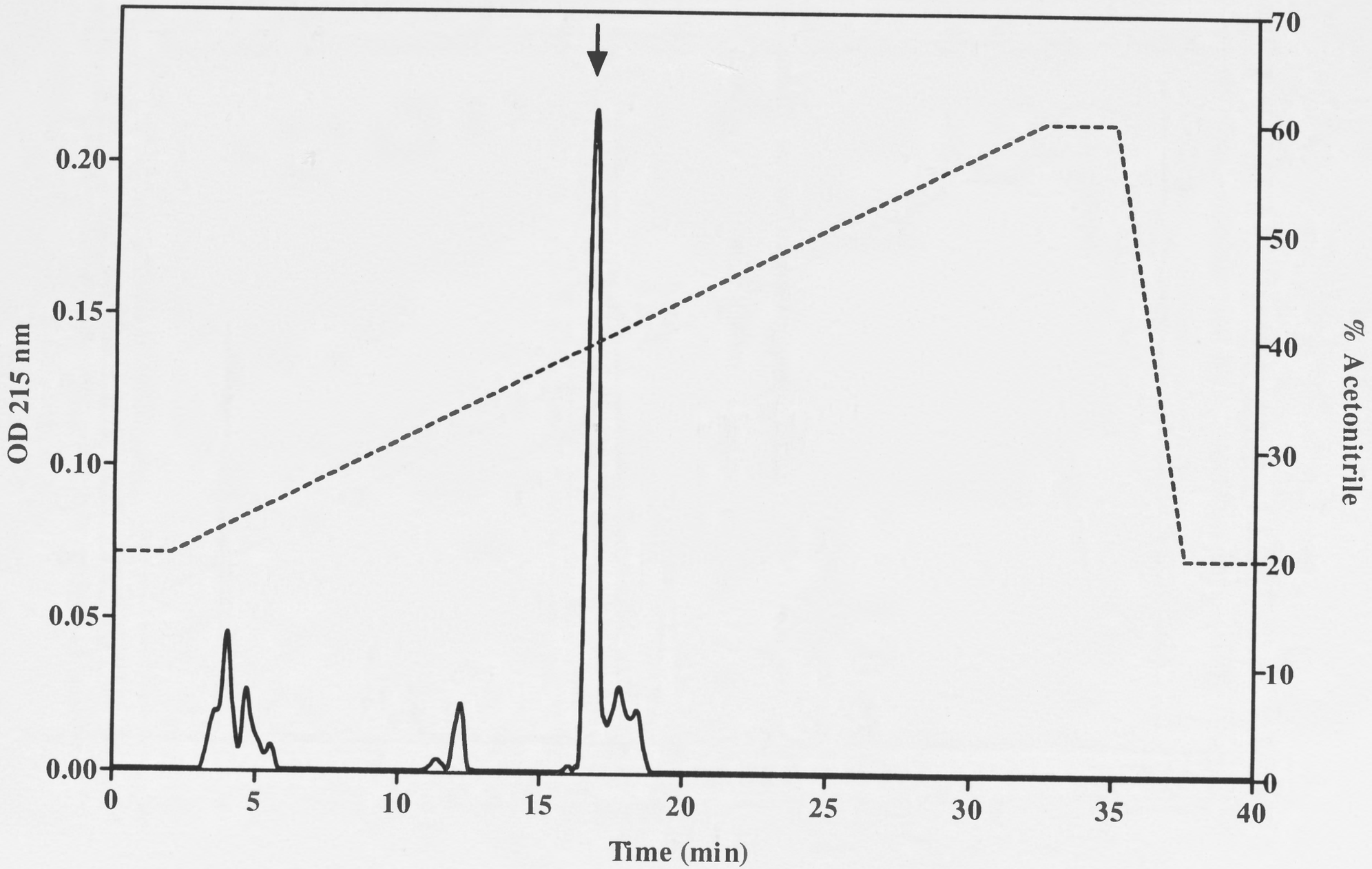
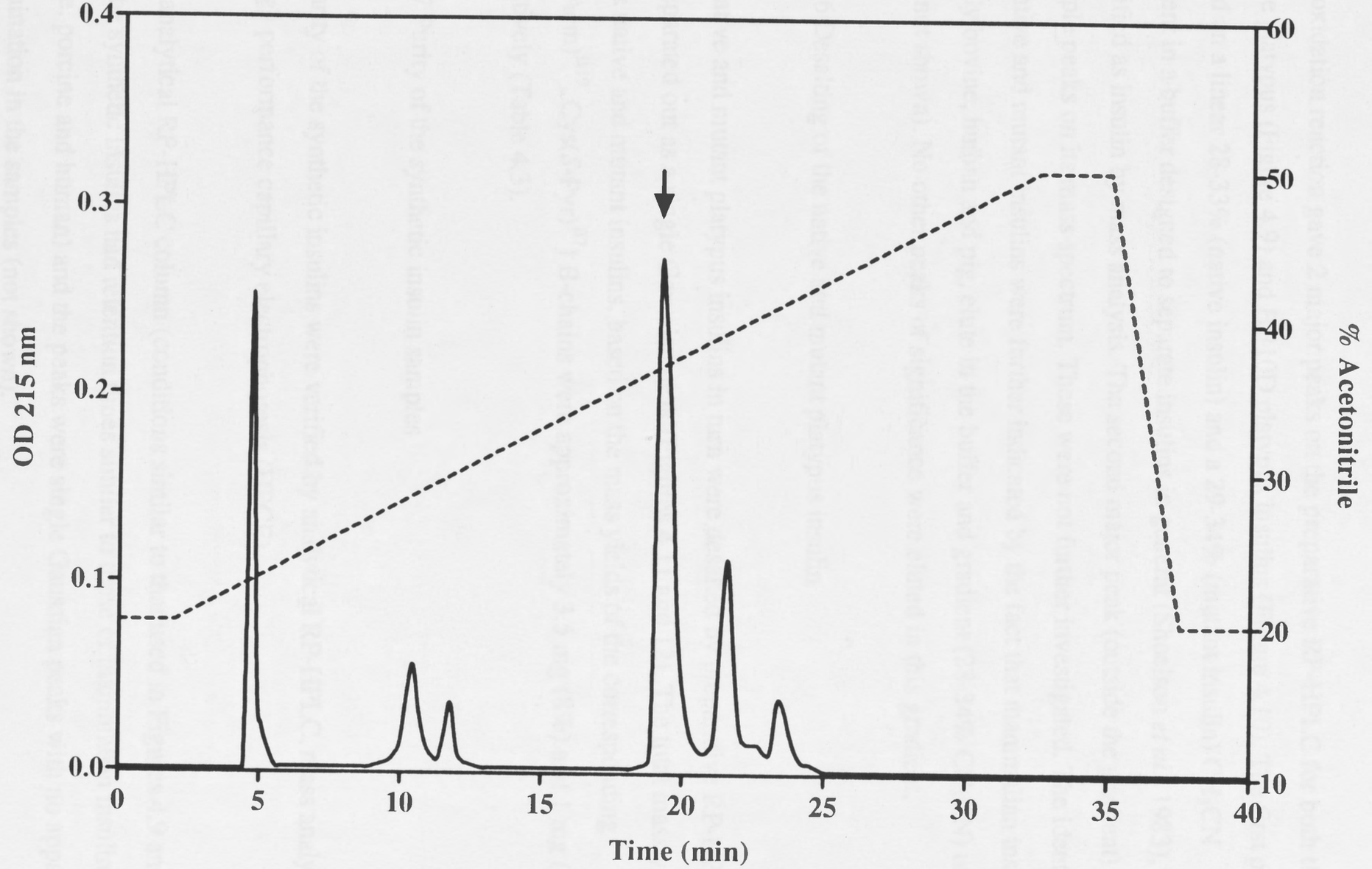

OD 215 nm
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
30
35
40
Time (min)
70
60
50
40
30
20
10
0
% Acetonitrile

Figure 4.8

Preparative RP-HPLC of [Cys(Acm)$^{A6, A11, A20, B19}$][HB10D] platypus insulin (arrow). Purification method III (Table 4.1 and 2) was followed. The AB-chain eluted at approximately 37% CH_3CN.

Figure 4.8


0.4
0.3
0.2
0.1
0.0
OD 215 nm
60
50
40
30
20
10
% Acetonitrile
0
5
10
15
20
25
30
35
40
Time (min)

4.3.1.5 Purification of synthesised native and mutant insulins

The oxidation reaction gave 2 major peaks on the preparative RP-HPLC for both the native platypus (Figure 4.9) and HB10D platypus insulins (Figure 4.10). The first peak, eluted on a linear 28-33% (native insulin) and a 29-34% (mutant insulin) CH_3CN gradient in a buffer designed to separate insulins in general (Shoelson *et al.*, 1983), was identified as insulin by mass analysis. The second major peak (outside the gradient) gave multiple peaks on its mass spectrum. These were not further investigated. The identity of the native and mutant insulins were further indicated by the fact that mammalian insulins, namely bovine, human and pig, elute in the buffer and gradient (28-34% CH_3CN) used (data not shown). No other peaks of significance were eluted in this gradient.

4.3.1.6 Desalting of the native and mutant platypus insulin

The native and mutant platypus insulins in turn were desalted by preparative RP-HPLC and separated out as a single Gaussian peak (Figures 4.11 and 12). The total mass yields for the native and mutant insulins, based on the mass yields of the corresponding [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chains were approximately 3.5 mg (8%) and 1 mg (10%) respectively (Table 4.3).

4.3.1.7 Purity of the synthetic insulin samples

The purity of the synthetic insulins were verified by analytical RP-HPLC, mass analysis and high performance capillary electrophoresis (HPCE).

On an analytical RP-HPLC column (conditions similar to that used in Figures 4.9 and 4.10) the synthetic insulins had retention times similar to those of mammalian insulins, (bovine, porcine and human) and the peaks were single Gaussian peaks with no apparent contamination in the samples (not shown).

Figure 4.9

Preparative RP-HPLC of platypus insulin (arrow). Purification method V (Table 4.1 and 2) was followed. The insulin eluted at approximately 31% CH_3CN.

Figure 4.9

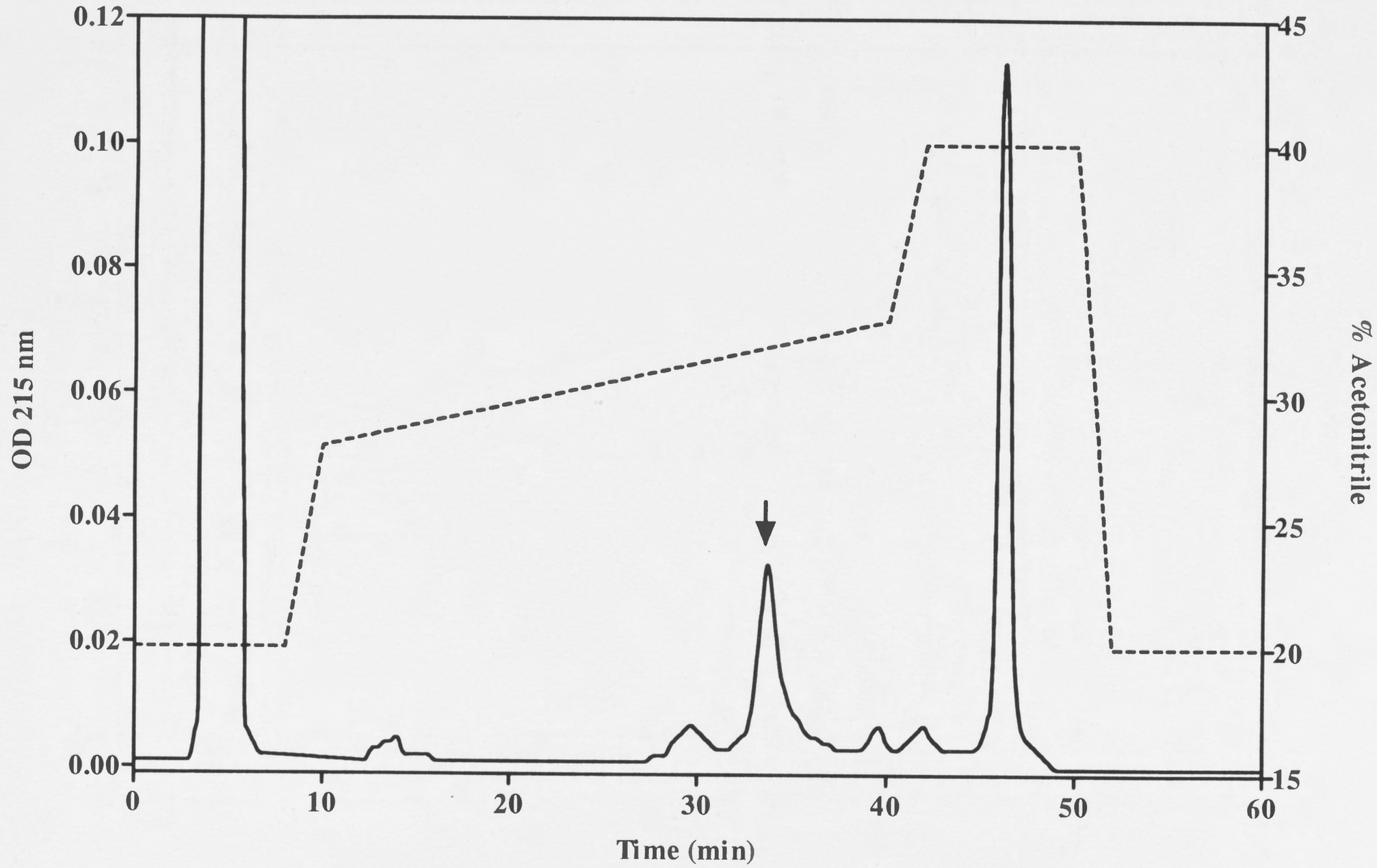

0.12
0.10
0.08
0.06
0.04
0.02
0.00
OD 215 nm
45
40
35
30
25
20
15
% Acetonitrile
0
10
20
30
40
50
60
Time (min)

Figure 4.10

Preparative RP-HPLC of [HB10D] platypus insulin (arrow). Purification method V (Table 4.1 and 2) was followed. The insulin eluted at approximately 33% CH_3CN.

Figure 4.10

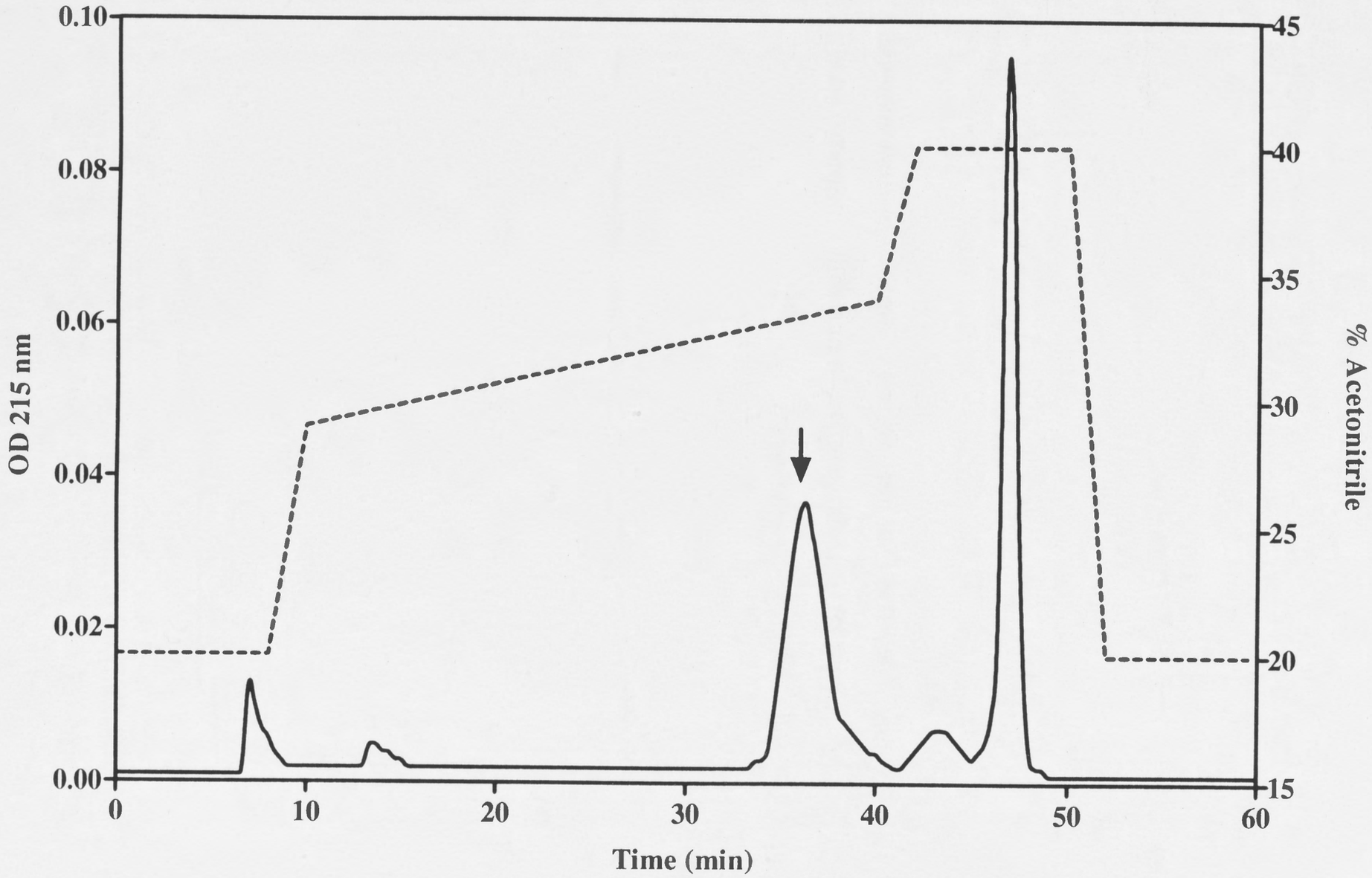

0.10
0.08
0.06
0.04
0.02
0.00
OD 215 nm
45
40
35
30
25
20
15
% Acetonitrile
0
10
20
30
40
50
60
Time (min)

Figure 4.11

Preparative RP-HPLC of platypus insulin (arrow). Purification method IV (Table 4.1 and 2) was followed. The insulin eluted at approximately 35% CH_3CN. This is the desalting step.

Figure 4.11

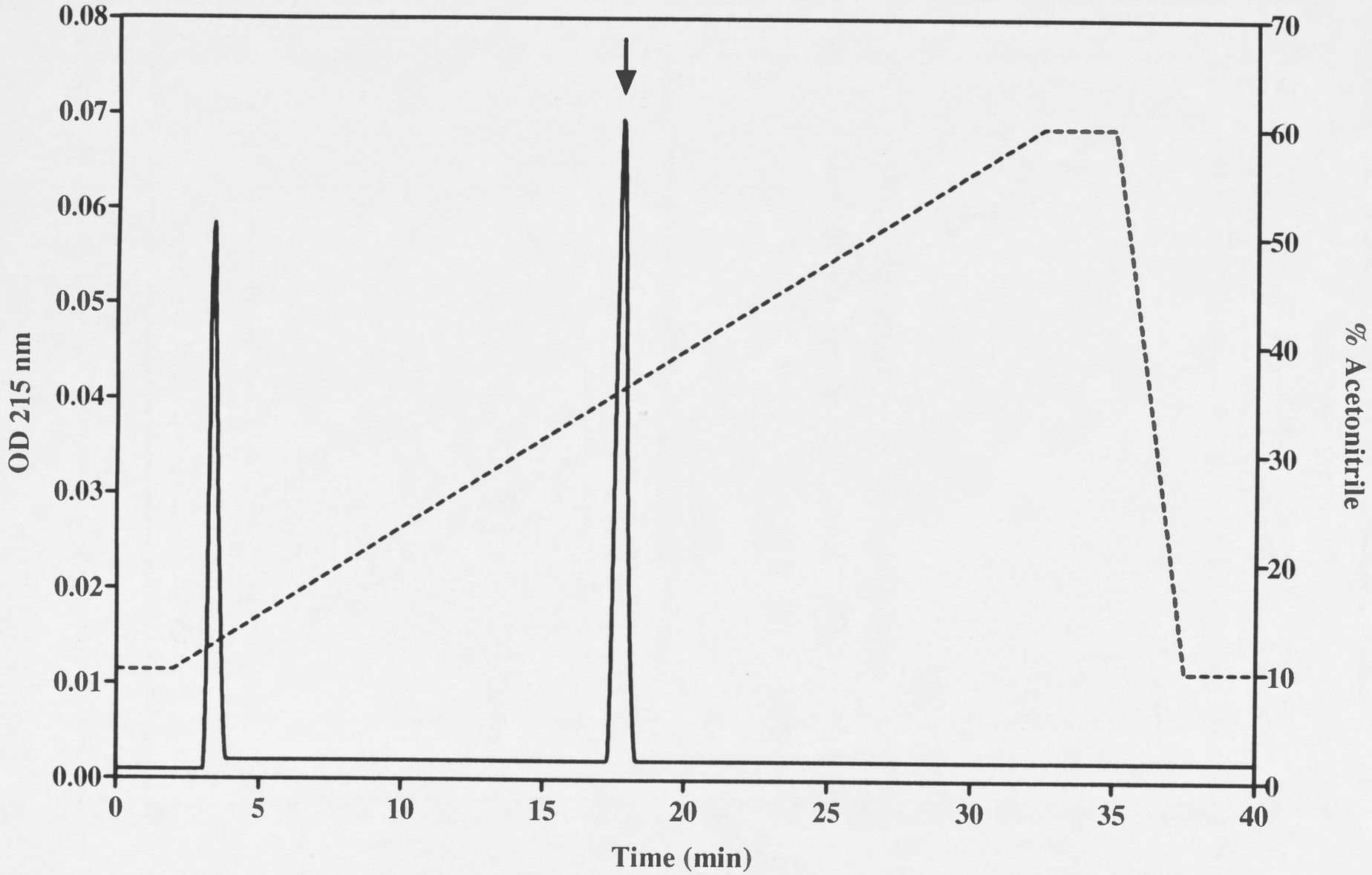
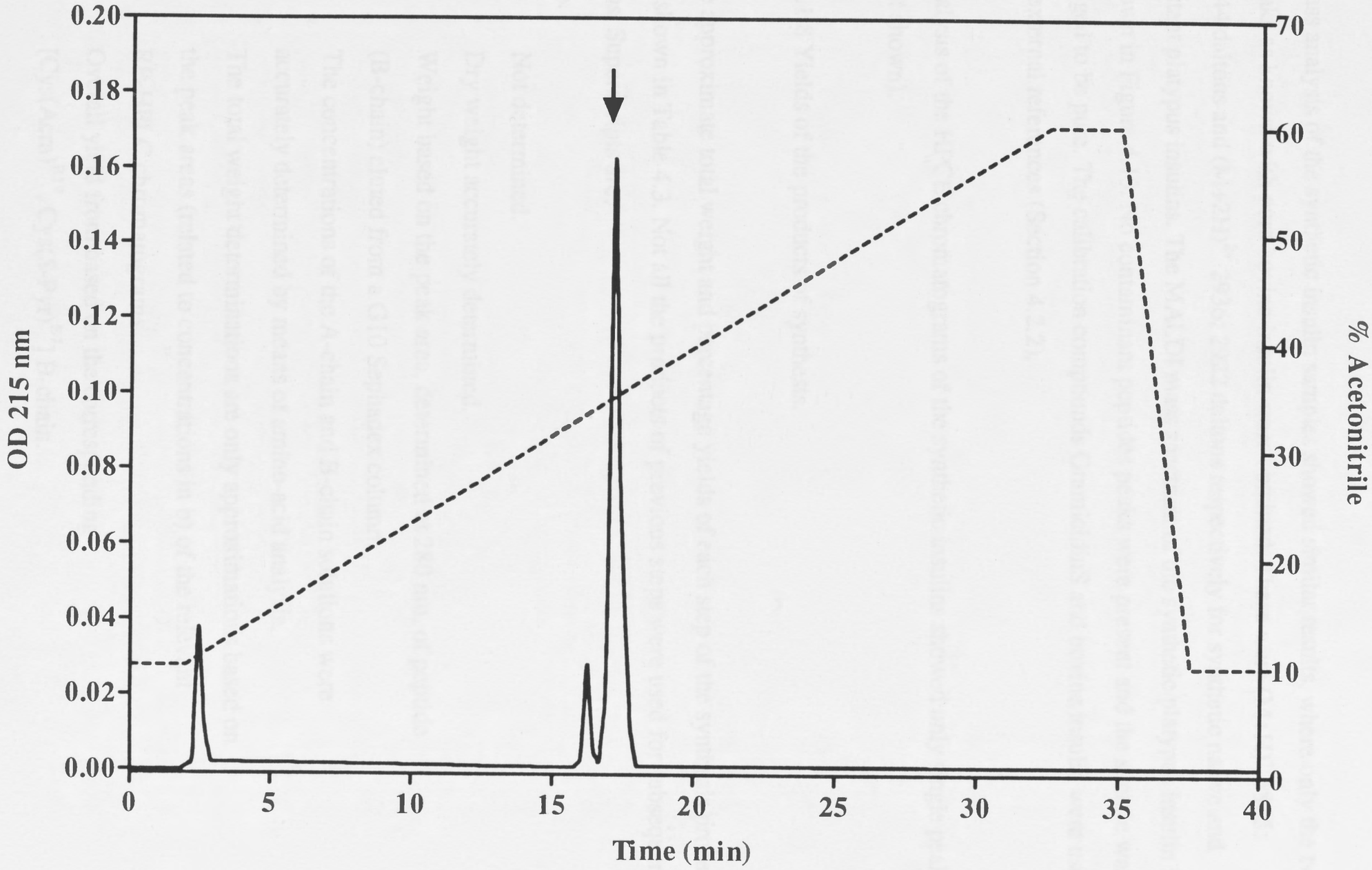

0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
OD 215 nm
70
60
50
40
30
20
10
0
% Acetonitrile
0
5
10
15
20
25
30
35
40
Time (min)

Figure 4.12

Preparative RP-HPLC of [HB10D] platypus insulin (arrow). Purification method IV (Table 4.1 and 2) was followed. The insulin eluted at approximately 35% CH_3CN. This is the desalting step.

Figure 4.12


0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
OD 215 nm
70
60
50
40
30
20
10
0
% Acetonitrile
0
5
10
15
20
25
30
35
40
Time (min)

Mass analysis of the synthetic insulin samples showed similar results, where only the two peaks as known with pure bovine insulin were obtained. These were $(M+H)^{+}$ 5865; 5844 daltons and $(M+2H)^{2+}$ 2936; 2922 daltons respectively for synthetic native and mutant platypus insulins. The MALDI mass spectrum of the synthetic platypus insulin is shown in Figure 4.13. No contaminant peptides peaks were present and the sample was judged to be pure. The calibration compounds GramicidinS and bovine insulin were used as external references (Section 4.2.2).

Analysis of the HPCE chromatograms of the synthetic insulins showed only single peaks (not shown).

4.3.1.8 Yields of the products of synthesis.

The approximate total weight and percentage yields of each step of the synthesis process are shown in Table 4.3. Not all the products of previous steps were used for subsequent steps. Superscripts are;

ND Not determined.

a) Dry weight accurately determined.

b) Weight based on the peak area, determined at 280 nm, of peptide (B-chain) eluted from a G10 Sephadex column.

c) The concentrations of the A-chain and B-chain solutions were accurately determined by means of amino-acid analysis.

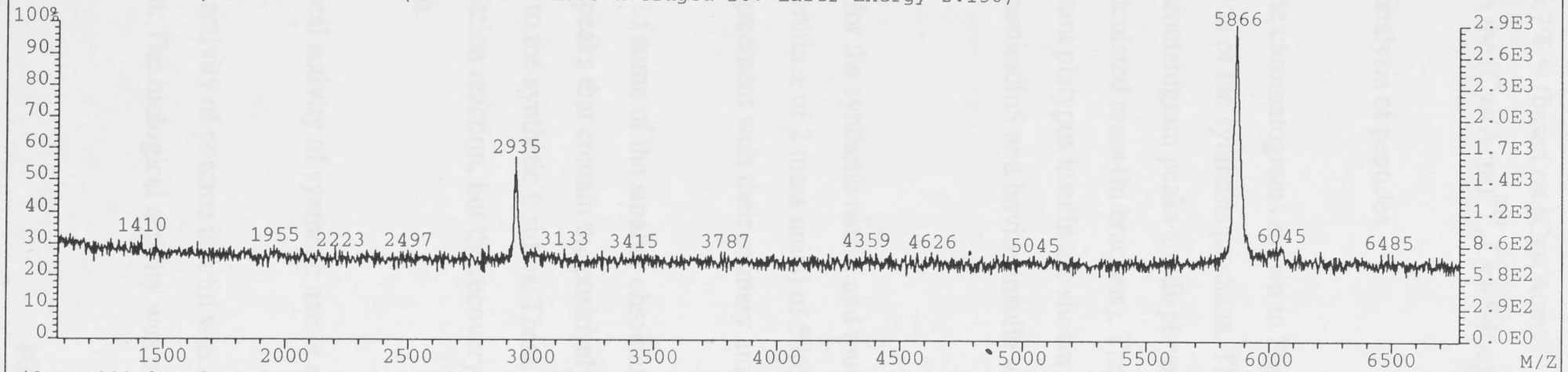
d) The total weight determinations are only approximations based on the peak areas (related to concentrations in b) of the relevant RP-HPLC chromatograms.

e) Overall yield from based on the corresponding [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain.

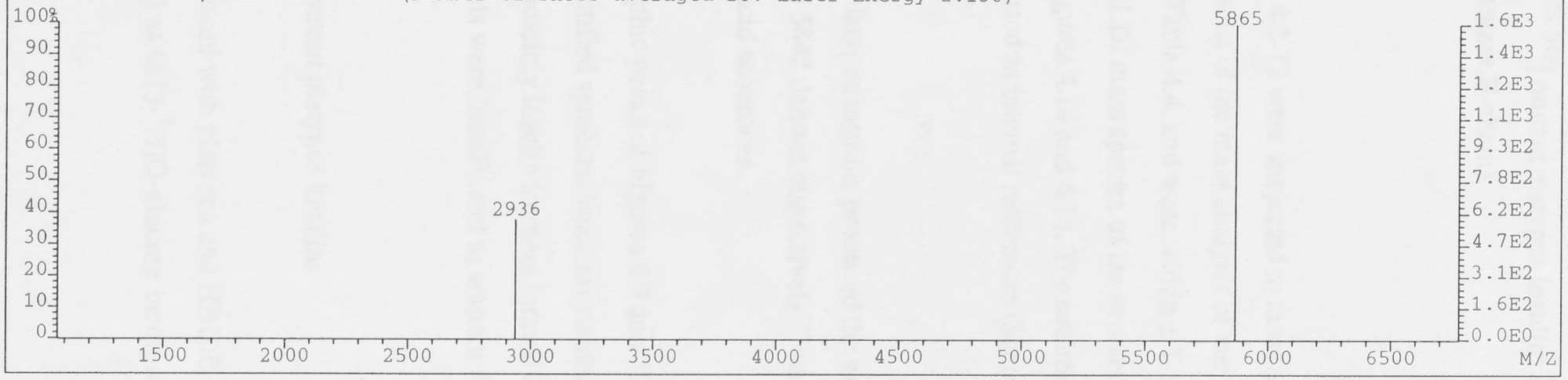
Figure 4.13

MALDI mass spectrum of the synthetic native platypus insulin. The calibration compounds GramicidinS and bovine insulin were used as external references. The peaks are platypus insulin $(M + H)^+$ 5865 and $(M + 2H)^{2+}$ 2936 daltons

File:AN109_2A Ident:1 Acq:22-NOV-1995 12:35:15 Cal:AN109_15A
TOFSPEC LDI+ TOF BpI:21504 TIC:35110716
File Text:508.2; PLINC (Number of shots averaged 28. Laser Energy 2:156)



File:AN109_2A Ident:1 SMO(2,9) PKD(7,3,7,0.50%,0.0,80.00%,F,F) SPEC(Heights,Centroid) Acq:22-NOV-1995 12:35:15 Cal:AN109_1»
TOFSPEC LDI+ TOF BpI:21504 TIC:35110716 Flags:NORM
File Text:508.2; PLINC (Number of shots averaged 28. Laser Energy 2:156)



The relative yields of the native platypus insulin were 7% [d] (based on [Cys(Acm)$^{A6, A11, A20, B19}$] native platypus insulin) and 8% [e] (based on the [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain) while the relative yields of the mutant platypus insulin were 6.7% [d] (based on [Cys(Acm)$^{A6, A11, A20, B19}$] mutant platypus insulin) and 10 % [e] (based on the [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] mutant B-chain).

4.3.1.9 Mass analysis of peptides

All peaks of the chromatograms shown in Figures 4.2-12 were subjected to mass analysis for identification of the synthesis products. The results of the mass analysis of the appropriate chromatogram peaks are displayed in Table 4.4. and were within ± 1-5 mass units of the calculated mass (in brackets). The MALDI mass spectra of the synthetic native and mutant platypus insulin are shown in Figures 4.14 and 4.15. The calibration compounds GramicidinS and bovine insulin were used as internal references (Section 4.2.2).

Mass analysis for the synthetic native and mutant platypus insulins produced the expected mass values (within ± of 2 mass units) of 5865 and 5843 daltons respectively. These values are in agreement with their primary amino-acid sequences.

Mass analyses of some of the smaller chromatographic peaks of Figures 4.9 and 10 (excluding the peaks that contain the positively identified synthetic insulins) yielded masses similar to the synthetic insulins. These are possibly insulin isomers formed during the iodine oxidation reaction, but the recovery yields were insufficient to warrant further characterisation.

4.3.2 Biological activity of synthetic native and mutant platypus insulins

The biological activity of porcine insulin was compared with platypus and HB10D platypus insulin. The biological activity was plotted as % [3-^{3}H]D-glucose incorporation

Figure 4.14

MALDI mass spectrum of the synthetic native platypus insulin. The calibration compounds GramicidinS and bovine insulin were used as internal references. The peaks are: GramicidinS $(M + H)^+$ 1139 daltons; bovine insulin $(M + H)^+$ 5737 and $(M + 2H)^{2+}$ 2868 daltons; platypus insulin $(M + H)^+$ 5864 and $(M + 2H)^{2+}$ daltons. (See Table 4.4 for calculated mass values.)

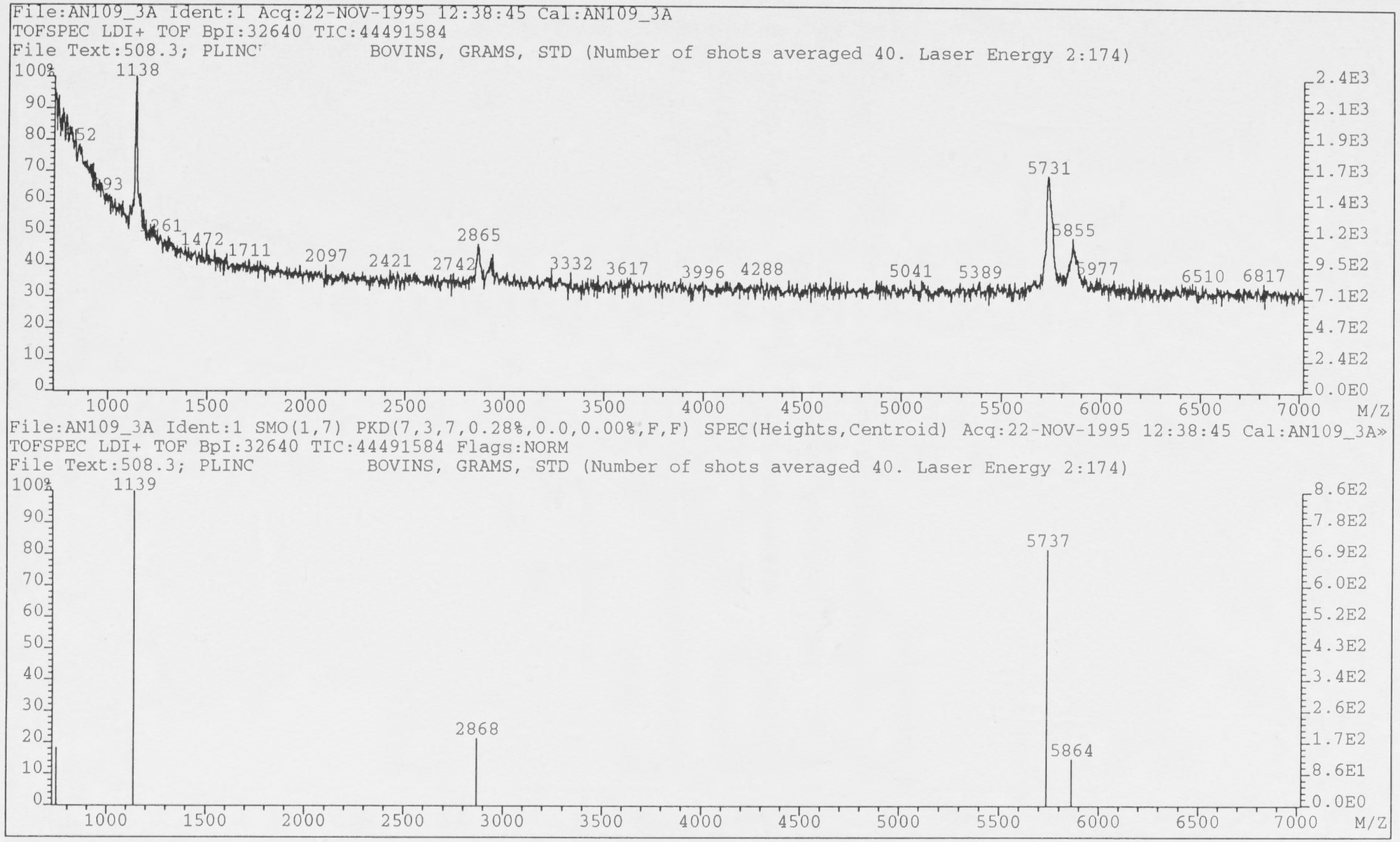

File:AN109_3A Ident:1 Acq:22-NOV-1995 12:38:45 Cal:AN109_3A
TOFSPEC LDI+ TOF BpI:32640 TIC:44491584
File Text:508.3; PLINC BOVINS, GRAMS, STD (Number of shots averaged 40. Laser Energy 2:174)
1138
1052
993
1261
1472
1711
2097
2421
2742
2865
3332
3617
3996
4288
5041
5389
5731
5855
5977
6510
6817
M/Z
File:AN109_3A Ident:1 SMO(1,7) PKD(7,3,7,0.28%,0.0,0.00%,F,F) SPEC(Heights,Centroid) Acq:22-NOV-1995 12:38:45 Cal:AN109_3A»
TOFSPEC LDI+ TOF BpI:32640 TIC:44491584 Flags:NORM
File Text:508.3; PLINC BOVINS, GRAMS, STD (Number of shots averaged 40. Laser Energy 2:174)
1139
2868
5737
5864
M/Z

Figure 4.15

MALDI mass spectrum of the synthetic mutant HB10D platypus insulin. The calibration compounds GramicidinS and bovine insulin were used as internal references. The peaks are: GramicidinS $(M + H)^+$ 1143 daltons; bovine insulin $(M + H)^+$ 5734 and $(M + 2H)^{2+}$ 2868 daltons; HB10D platypus insulin $(M + H)^+$ 5847 and $(M + 2H)^{2+}$ 2926 daltons. (See Table 4.4 for calculated mass values.)

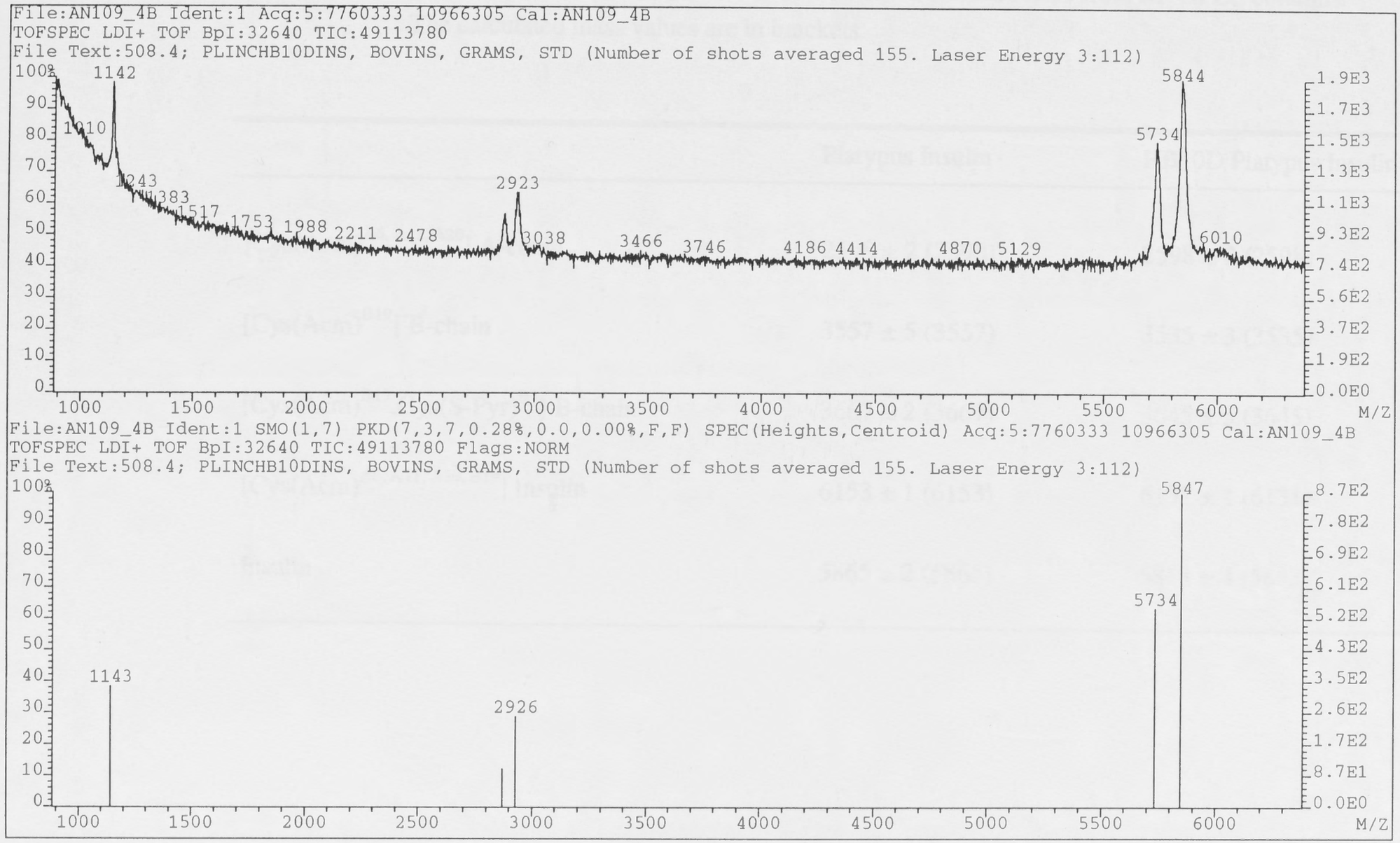

File:AN109_4B Ident:1 Acq:5:7760333 10966305 Cal:AN109_4B
TOFSPEC LDI+ TOF BpI:32640 TIC:49113780
File Text:508.4; PLINCHB10DINS, BOVINS, GRAMS, STD (Number of shots averaged 155. Laser Energy 3:112)
100%
90
80
70
60
50
40
30
20
10
0
1142
1010
1243
1383
1517
1753
1988
2211
2478
2923
3038
3466
3746
4186
4414
4870
5129
5734
5844
6010
1.9E3
1.7E3
1.5E3
1.3E3
1.1E3
9.3E2
7.4E2
5.6E2
3.7E2
1.9E2
0.0E0
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
M/Z
File:AN109_4B Ident:1 SMO(1,7) PKD(7,3,7,0.28%,0.0,0.00%,F,F) SPEC(Heights,Centroid) Acq:5:7760333 10966305 Cal:AN109_4B
TOFSPEC LDI+ TOF BpI:32640 TIC:49113780 Flags:NORM
File Text:508.4; PLINCHB10DINS, BOVINS, GRAMS, STD (Number of shots averaged 155. Laser Energy 3:112)
100%
90
80
70
60
50
40
30
20
10
0
1143
2926
5734
5847
8.7E2
7.8E2
6.9E2
6.1E2
5.2E2
4.3E2
3.5E2
2.6E2
1.7E2
8.7E1
0.0E0
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
M/Z

Table 4.4: MALDI-TOF mass analysis (m/z) of peptides in daltons eluted from RP-HPLC columns. The calculated mass values are in brackets.

	Platypus Insulin	HB10D Platypus Insulin
[Cys(Acm)$^{A6, A11, A20}$] A-chain	2598 ± 2 (2598)	2598 ± 2 (2598)
[Cys(Acm)^{B19}] B-chain	3557 ± 5 (3557)	3535 ± 3 (3535)
[Cys(Acm)^{B19},Cys(S-Pyr)^{B7}] B-chain	3667 ± 2 (3667)	3645 ± 5 (3645)
[Cys(Acm)$^{A6, A11, A20, B19}$] Insulin	6153 ± 1 (6153)	6131 ± 1 (6131)
Insulin	5865 ± 2 (5865)	5843 ± 4 (5843)

into rat adipocytes *versus* log insulin concentration (Figures 4.16 and 17). The GraphPadPRISM program was used to calculate ED_{50}. These values for porcine, platypus and HB10D platypus insulin were 5.23×10^{-10}; 5.47×10^{-10} and 2.56×10^{-10} M respectively. The relative potency on a molar basis for native and HB10D platypus insulin were 96 and 200% respectively.

4.3.3 Biophysical characterisation of synthetic native and mutant platypus insulins

Circular dichroism spectra of synthetic native and mutant platypus insulin

A comparison of the CD spectra (far UV) of porcine, platypus and HB10D platypus insulin at pH 3.2 are shown in Figure 4.18. It is clear that the features of the CD spectra of the two synthetic insulins are similar to that of porcine insulin in the same solution. It may thus be concluded that the final synthetic products are both correctly folded insulins. Enhanced concentration and association of insulin produce CD-spectral effects that have been observed in the CD spectra of the mutant platypus insulin, notably a decrease in the $[\theta]_{208}/[\theta]_{223}$ ratio (Pocker and Biswas, 1980); in fact the features are very similar to that of human insulin at pH 7.8 in the presence of zinc (Wollmer *et al.*, 1994). The $[\theta]_{208}/[\theta]_{223}$ ratios for porcine, platypus and HB10D platypus insulin are shown in (Table 4.5).

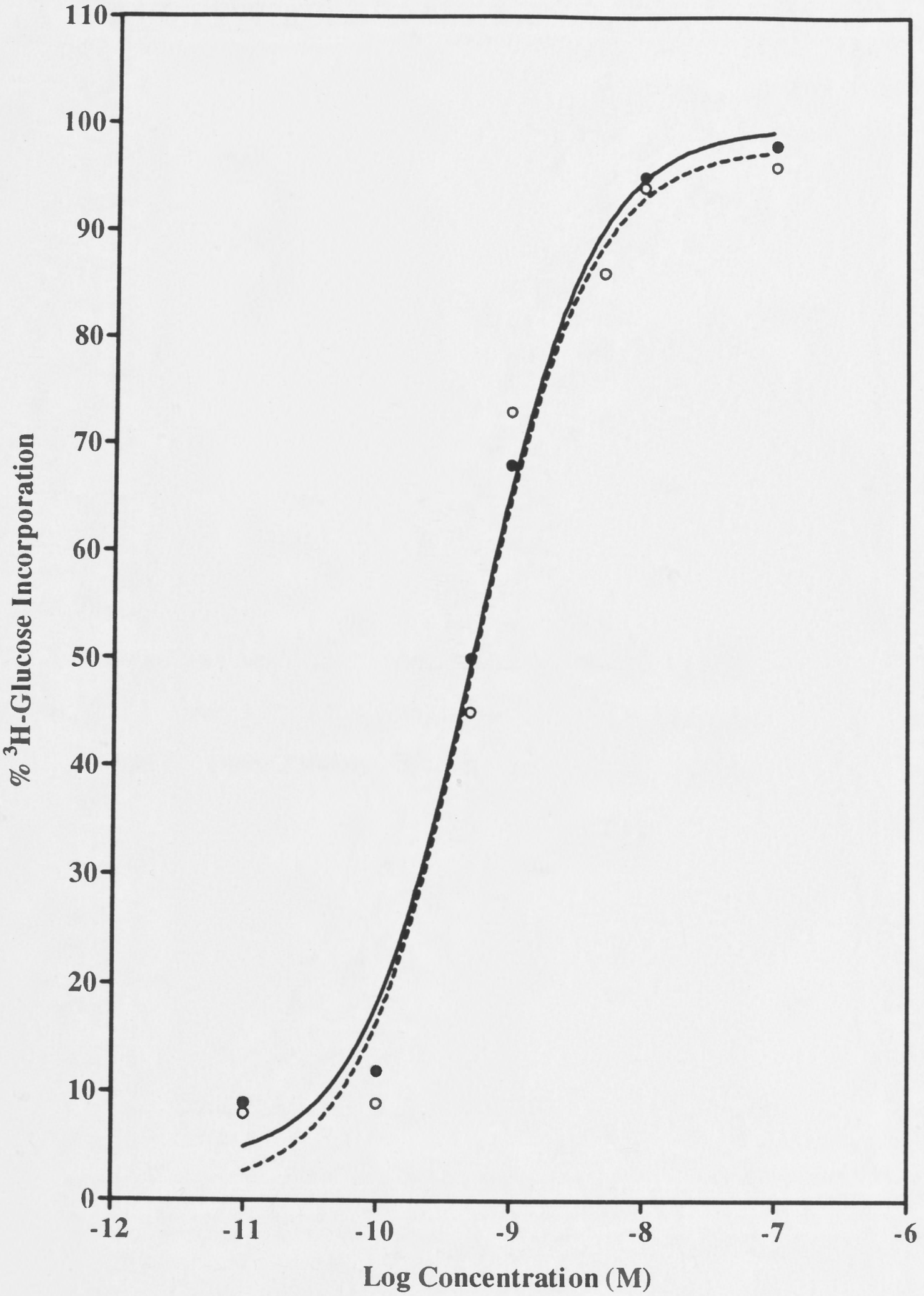
Table 4.5: CD data of porcine and synthetic insulins in 1.0 mM HCl, pH 3.2

	Concentration (μM)	$[\theta]_{208}/[\theta]_{223}$
Platypus insulin	24	1.53
HB10D platypus insulin	50	1.21
Porcine insulin	25	1.48

Figure 4.16

A comparison of the stimulation of lipogenesis (expressed as percent of the maximum) in isolated rat adipocytes by platypus and porcine insulin. Broken line (o) is porcine, solid line (•) platypus insulin.

Figure 4.16



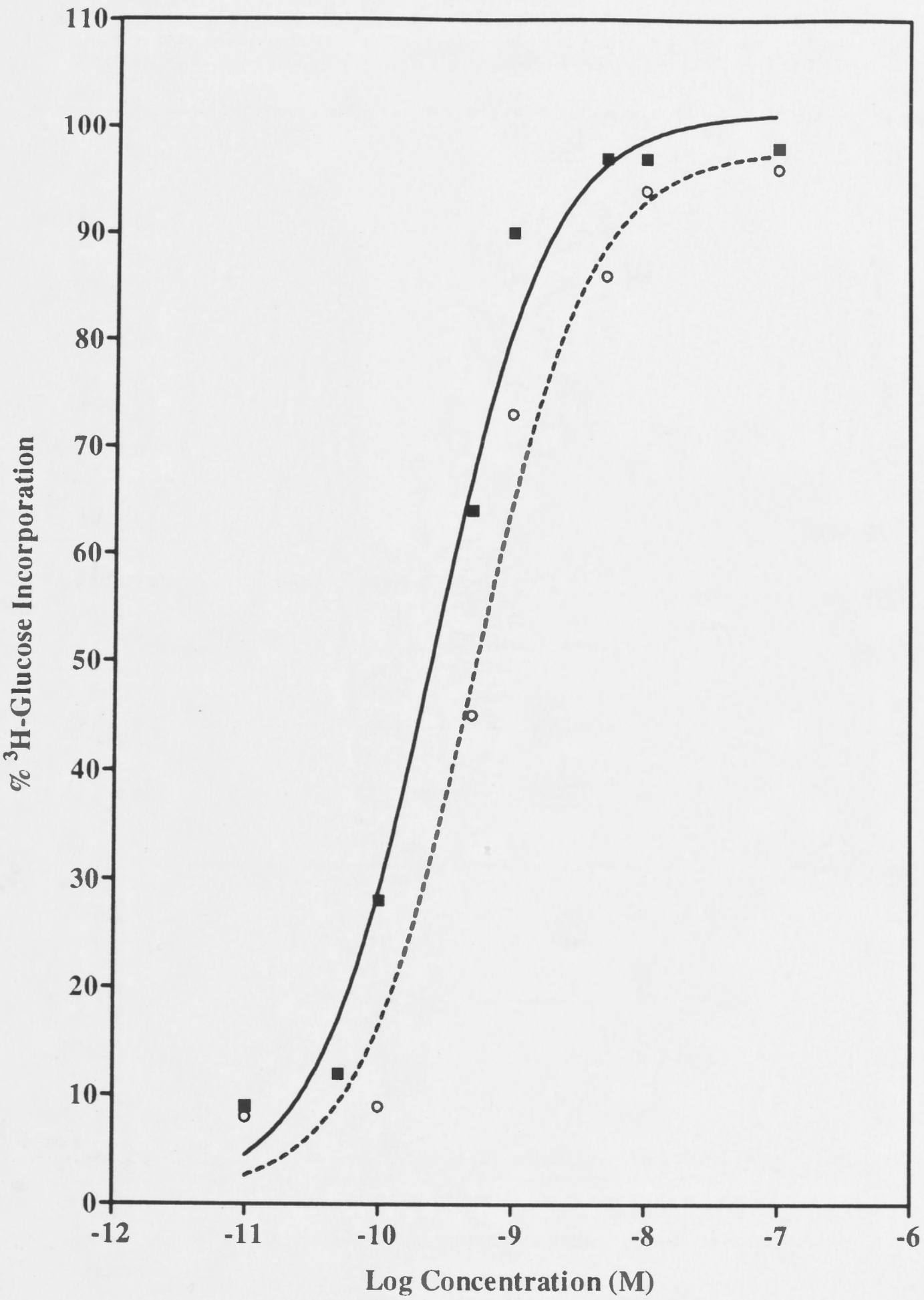
● Platypus Insulin

○ Porcine Insulin

Figure 4.17

A comparison of the stimulation of lipogenesis (expressed as percent of the maximum) in isolated rat adipocytes by HB10D platypus and porcine insulin. Broken line (o) is porcine, solid line (■) HB10D platypus insulin.

Figure 4.17



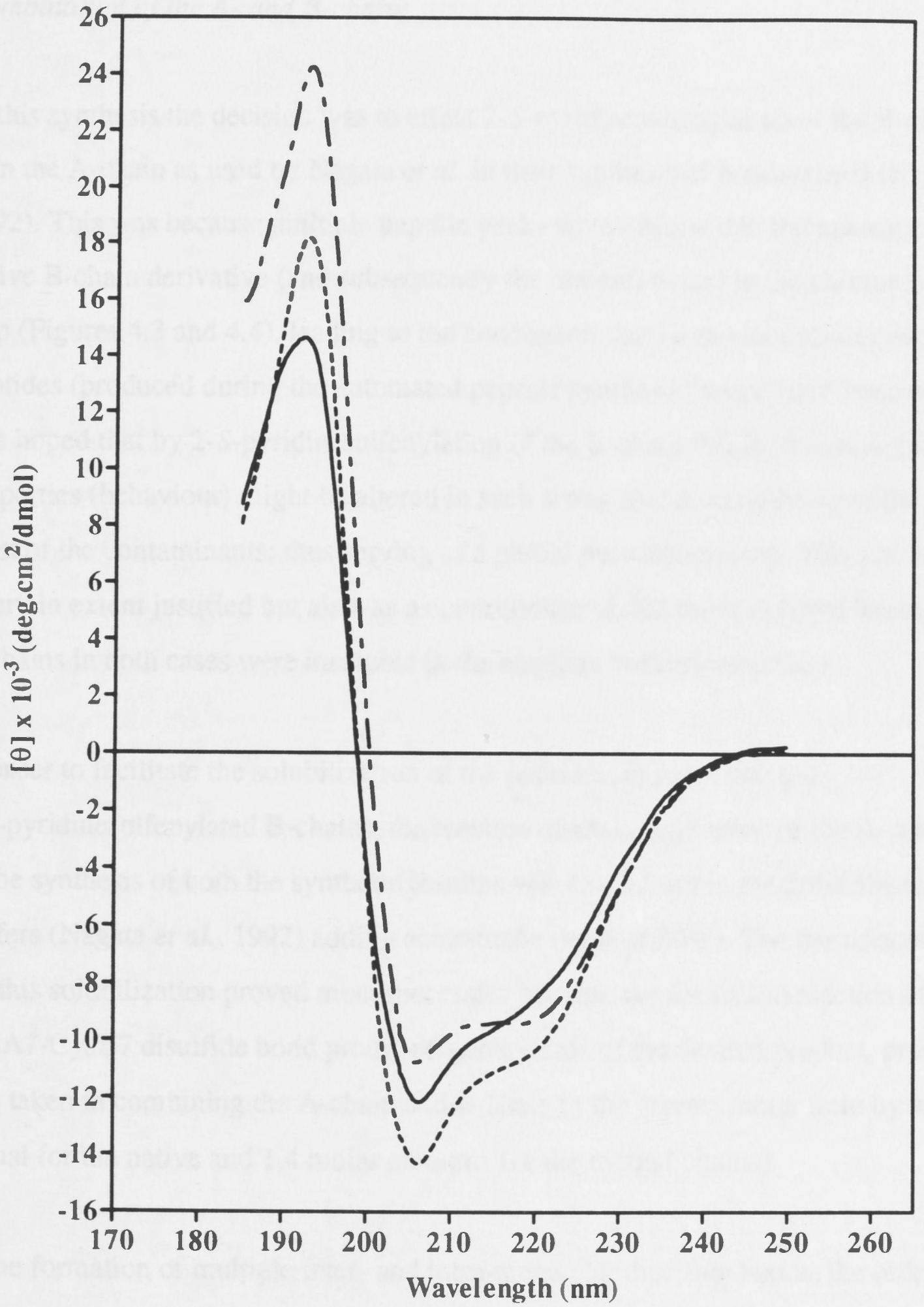
■ HB10D Platypus Insulin

○ Porcine Insulin

Figure 4.18

A comparison of the far UV CD spectra of porcine (-----), platypus (——) and HB10D platypus insulins (— - —) at 25, 24 and 50 μM respectively in 0.001 M HCl, pH 3.2.

Figure 4.18


26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
[θ] x 10-3 (deg.cm2/dmol)
170
180
190
200
210
220
230
240
250
260
Wavelength (nm)
HB10D Platypus Insulin
Porcine Insulin
Platypus Insulin

4.4 Discussion

Combination of the A- and B-chains

In this synthesis the decision was to effect 2-*S*-pyridinesulfenylation of the B-chain rather than the A-chain as used by Nagata *et al.* in their synthesis of bombyxin-II (Nagata *et al.*, 1992). This was because multiple peptide peaks were observed in the mass spectra of the native B-chain derivative (and subsequently the mutant) eluted in the chromatographic step (Figures 4.3 and 4.4), leading to the conclusion that large amounts of contaminant peptides (produced during the automated peptide synthesis) might have been present. It was hoped that by 2-*S*-pyridinesulfenylation of the B-chain that its chromatographic properties (behaviour) might be altered in such a way that it could be separated from most of the contaminants; thus serving as a partial purification step. This strategy was to a certain extent justified but also, as a consequence of this these 2-*S*-pyridinesulfenylated B-chains in both cases were insoluble in the aqueous buffers prescribed.

In order to facilitate the solubilization of the peptides, in particular the 2-*S*-pyridinesulfenylated B-chains, the reaction of the combination of the A- and B-chains in the synthesis of both the synthetic insulins was carried out in the prescribed aqueous buffers (Nagata *et al.*, 1992) adding acetonitrile (total of 40%). The use of acetonitrile for this solubilization proved most successful because the formation reaction of the CysA7-CysB7 disulfide bond produced good yields of the desired product, provided care was taken in combining the A-chain and B-chain in the correct molar ratio by weight (equal for the native and 1.4 molar amounts for the mutant chains).

In the formation of multiple inter- and intra-molecular disulfide bonds, the order of formation of disulfide bonds is an important factor. The strategy where the disulfide bond CysA7-CysB7 was formed first, instead of the CysB19-CysA20, was chosen because of the higher yield of the correct product over the incorrect disulfide linkages that were

obtained by Nagata *et al.* (Nagata *et al.*, 1992) in their synthesis of bombyxin-II. It has been assumed that no recognisable exchange of the disulfide linkages occurred during the iodine oxidation and that all the synthesis products should retain the disulfide bond CysA7-CysB7. The correct mode of disulfide linkages of natural insulins is as follows: CysA7-CysB7; CysA6-CysA11 and CysA20-CysB19 (Figure 4.1). Other products that could have different modes of disulfide linkages from the natural insulin will be CysA7-CysB7; CysA6-CysA20 and CysA11-CysB19 (isomer 1) as well as CysA7-CysB7; CysA11-CysA20 and CysA6-CysB19 (isomer 2).

Iodine oxidation

In the work here iodine oxidation was carried out in the presence of HCl despite there being no Trp in the peptides studied. This step was however, included for two reasons. In the first instance it was used because it was known to yield successful results and also it was reported by Nagata *et al.* (Nagata *et al.*, 1992) that during iodine oxidation considerable side reactions occurred that lowered the yield of this reaction. If some of these side reactions were unrelated to the presence of Trp then the use of HCl would prevent these happening so its use was adopted as a precautionary measure since it was not the intention here to study side reactions caused by iodine oxidation. From the yields obtained here, however, which are in agreement to those obtained by Maruyama *et al.* (Maruyama *et al.*, 1992) in their synthesis of bombyxin-IV without the addition of HCl, the use of HCl might not have been necessary (see below).

Relative yields of native and mutant platypus insulin

The percentage yields of native and mutant platypus insulin, based on the quantities of the corresponding [Cys(Acm)^{B19}, Cys(*S*-Pyr)^{B7}] B-chain, were 8% and 10% respectively. These do not compare very favourably with the 32% yield obtained by Nagata *et al.* (Nagata *et al.*, 1992) for the synthesis of the bombyxin-II, a brain secretory peptide of the silkworm *Bombyx mori* whose amino-acid sequence shows considerable homology

with the vertebrate insulin family peptides (Nagata *et al.*, 1992 and references therein). The high loss or low percentage yield was mainly caused by the I_2 oxidation step (Section 4.2.1.3.6 and Table 4.3) but was balanced by the high purity of both products, which were homogenous according to several high performance analytical methods (mass spectrometry, high performance capillary electrophoresis, analytical RP-HPLC and sedimentation equilibrium analysis).

The yields with the insulin products here, however, are in general agreement with those of others using chemical synthesis (Brandenburg, 1990). Taking into account the fact that in chemical synthesis chemical bonding is not directed by a template, the yields obtained would seem to compare favourably with those of insulin produced by DNA recombinant technology; which were in the order of 11-52% for a series of insulin analogues (Markussen *et al.*, 1988). Purity of the final product is, however, of paramount importance and in this work the purity was greater than 99% and this compares well with that obtained by the mentioned workers. An important advantage of this chemical method over DNA recombinant technology is that insulin analogues containing non-proteinogenic amino-acids can be synthesised (Brandenburg, 1990).

Characterisation of synthetic insulins: CD spectra and biological activity

The CD spectra of the final synthetic products (native and mutant platypus insulins) show typical features of a correctly folded insulin, and it can be reasonable assumed that they are indeed insulins. The CD spectrum of the synthetic mutant insulin at pH 3.2 shows spectral effects that are typical of insulin in a conformational state that it adopts in the presence of Zn^{2+} or at high concentrations where it is in a state of enhanced self-association (Pocker and Biswas, 1980; Wollmer *et al.*, 1994; Goldman and Carpenter, 1974). This observation has not been investigated further but it is tempting to suggest that the reason for the possible enhanced self-association in an acid environment is due to the B10 Asp. The latter resulting in a reduction of net charge when compared to the protonated His that it replaced, and thus may contribute to an increase in hydrophobic

self-interaction. (Hydrophobic self-interaction being the main force driving the association process (Jeffrey and Coates, 1966; Baker *et al.*, 1988).

The success of the chemical synthesis of both the native and mutant platypus insulins are measured by the *in vitro* biological activity test. Synthesised native platypus insulin has virtually the same potency as human (porcine) insulin. In agreement with HB10D human insulin, which in activity tests *in vitro* has a potency of 200% (Schwartz *et al.*, 1987), that of HB10D platypus insulin is of the same order.

Conclusion: Verification, yields and homogeneity

The conclusion that the final synthetic products are indeed the native and mutant platypus insulins is supported by the results of the biological activity measurements, the features of the circular dichroism spectra, very similar retention times on a RP-HPLC gradient developed to separate insulins (Shoelson *et al.*, 1983), mass analysis and sedimentation equilibrium analysis (Chapter 5).

Possible disulfide bond isomers (incorrect disulfide linkage products) of the synthetic insulins which could be expected to be obtained due to the strategies of semi-regioselective disulfide bond formation were not identified and characterised.

Chapter 5

The Self-association Pattern of Zinc-free Bovine, Synthetic Platypus and HB10D Platypus Insulin: A Sedimentation Equilibrium study.

5.1 Sedimentation analysis

In an ultracentrifuge parameters of the molecules under investigation can be related to their behaviour in a gravitational field. These parameters include molecular weight, shape and density of the molecule. Solvents influence this behaviour and their effects must be accounted for as the molecules (solute) are in solution.

The general ultracentrifugation equation or Lamm equation (Bowen, 1970) describes the effects of sedimentation, diffusion and angular velocity on the concentration changes and thus behaviour of solute with time. In this equation the centrifugal cell dimensions and the fact that the centrifugal field is not homogeneous but increases linearly with r, the distance from the axis of rotation, are taken into account. The solute movement can be divided into two processes namely sedimentation towards the bottom of the cell and diffusion in the opposite direction.

The flow equation describes the movement of material across a plane at r as follows;

$$\frac{dm}{dt} = \theta rh\left(cs\omega^2 r - D\frac{dc}{dr}\right) \qquad 5.1$$

Where dm is the net movement of the material, r the distance from the axis of rotation, c the concentration of the solute, s the sedimentation coefficient, ω the angular velocity of the rotor and D the diffusion coefficient. θrh relates to the dimensions of the centrifugal cell where θ is the sector angle, and h the solution thickness.

If sedimentation and diffusion are assumed to be independent of concentration the Lamm equation can be derived from the flow equation;

$$\frac{dc}{dt} = D\left\{\frac{d^2c}{dr^2} + \frac{1}{r}\frac{dc}{dr}\right\} - s\omega^2\left\{r\frac{dc}{dr} + 2c\right\} \qquad 5.2$$

Symbols as before (Bowen, 1970).

Expressions involving the molecular weight of the solute can be formulated from equations 5.1 and 5.2. In the Beckman Optima XL-A analytical ultracentrifuge the total concentration at any point in the cell can be measured by absorption optics. From these data molecular weights of solute in the native state as they exist in solution can be calculated directly. There are two main approaches for the determination of molecular weight by sedimentation techniques namely, the sedimentation velocity method where a sufficiently high gravitational field is applied to cause the sedimentation of a molecule that can be directly measured, and the sedimentation equilibrium method. The latter is the method of choice, especially for self-associating proteins, and was used exclusively in the work to be described.

5.1.1 Sedimentation equilibrium

In sedimentation equilibrium a protein solution in a cell is inserted into a rotor and spun at a constant angular velocity, ω (radian/second), and constant temperature, T (absolute), until the resulting distribution of total concentration of solute, $\bar{c}$ (g/L), *versus* radial distance from the axis of rotation, r, becomes invariant with time. In other words centrifugation is carried on for an appropriate period of time at low speed until a steady state is reached where the processes of sedimentation and diffusion are in equilibrium. No net movement of the material occurs, $dm/dt = 0$ and the flow equation, 5.1, becomes

$$\frac{s}{D} = \frac{dc/dr}{c\omega^2 r} \qquad 5.3$$

Equation 5.3 applies at the equilibrium state no matter how s and D may depend on concentration (Bowen, 1970; Archibald, 1963) and can be substituted into the very well known Svedberg equation (Bowen, 1970; Svedberg and Pedersen, 1940)

$$M = \frac{RTs}{D(1-\bar{\upsilon}\rho)} \qquad 5.4$$

to give the equation for the determination of molecular weights by the method of sedimentation equilibrium

$$\frac{\ln(dc_r)}{dr^2} = \frac{M(1-\bar{\upsilon}\rho)\omega^2}{2RT} \qquad 5.5$$

where M is the solute molecular weight (g/mol), c_r is the concentration of the solute (g/mL) at the radial distance r from the axis of rotation, ω the angular velocity of the rotor (radians/second), ρ the density of the solution, $\bar{\upsilon}$ the partial specific volume of the solute, T the absolute temperature, and R the gas constant.

The sedimentation flow is proportional to $\omega^2 r$, and r increases towards the cell bottom. There is thus a greater sedimenting tendency at the bottom of the cell, and in consequence a greater balancing tendency for diffusion in the opposite direction. Since diffusion is driven by the gradient of chemical potential (which depends on the concentration gradient) it follows that the concentration gradient increases towards the cell bottom. In other words the centrifugal field increases towards the bottom of the cell (Williams *et al.*, 1958).

A plot of ln c_r versus r^2 will yield a straight line with a slope proportional to M for a single ideal solute. When multiple species are present, or for an associating system, the measured molecular weight shows an increase from the meniscus towards the cell bottom and also upward curvature. The slope of the ln c_r versus r^2 plot is then proportional to the average molecular weight; an average based on the proportion by weight of the various species present known as the weight-average molecular weight

$\overline{M}_w$. Thus for a non linear ln c_r versus r^2 plot, tangents to the curve yield a weight-average molecular weight for the mixture of species present at each radial position (Aune, 1978; Kim *et al.*, 1977; Teller, 1973). The weight-average molecular weight is defined as

$$\overline{M}_w = \frac{\Sigma M_i c_i}{\Sigma c_i} \qquad 5.6$$

Non-ideality arising from the finite size of macromolecules and the charge they carry tends to reduce apparent molecular weight. In the presence of non-ideality, the concentration distribution at equilibrium leads to the apparent weight-average molecular weight, $M_{w,app}$ .The plot must be assessed very carefully as upward curvature can be partially obscured by non-ideality, as the latter tends to cause downward curvature (Aune, 1978).

Overlaying plots of data from samples of different starting concentrations and angular speeds will provide information about the presence of more than one species in the cell. Results from sedimentation equilibrium experiments have thus the ability to detect polydisperse and self-associating systems and determine the monomer molecular weight and stoichiometry. The effects of non-ideality can be taken into account (Albright and Williams, 1967; Harding, 1985; Munk and Halbrook, 1976; Soucek and Adams, 1976).

Sedimentation equilibrium of a single solute species

It must be pointed out that equation 5.5, although simple, does not satisfy detailed analysis nor have thermodynamic theory and consideration of non-ideality been taken into account. In fact, many assumptions have been made to derive equation 5.5. In the following section some general principles of thermodynamics will be employed to deduce equations which are very similar in form to the above but are modified to contain various terms such as activity coefficients. In a later section (5.1.4) the

consideration of non-ideality effects with the increase in total concentration of the solute in the treatment of experimental results will be discussed.

At equilibrium the total potential is constant at each point in the cell, in other words the variation with radial distance of the centrifugal potential ψ_i, tending to cause sedimentation, is exactly balanced by the variation with radial distance of the chemical potential, μ_i tending to cause back-diffusion.

The mathematical equation of sedimentation equilibrium at constant temperature is as follows

$$\frac{d\psi_i}{dr} - \frac{d\mu_i}{dr} = 0 \qquad \text{(Fujita, 1962)} \qquad 5.7$$

The term $d\psi_i / dr$ may be formulated as $M_i\, r\, \omega^2\, dr$ where M_i is the molecular weight of the species i and the second term $d\mu_i/dr$ is obtained from the partial differentiation of $\mu_i = \mu_i^0 + RT\, ln\, a_i$ where μ_i^0 is the standard chemical potential per mole and a_i the thermodynamic activity of the species i. Thus the basic sedimentation equilibrium equations can be formulated assuming constant pressure applies as follows

$$\frac{d \ln a_i(r)}{dr^2} = \frac{M_i \omega^2 (1 - \bar{\upsilon}_i \rho)}{2RT} \qquad 5.8$$

which may be integrated utilising the reasonable assumption that the buoyancy term $(1 - \bar{\upsilon}_i \rho)$ is constant with respect to radial distance, and hence to pressure variation to give

$$a_i(r) = a_i(r_F) \exp\left\{\phi_i M_i (r^2 - r_F^2)\right\} \qquad 5.9$$

$$\phi_i = \frac{\omega^2 (1 - \bar{\upsilon}_i \rho)}{RT} \qquad 5.9a$$

where r and r_F are two radial distances between the meniscus, r_m and the base of the solution column r_b. It has been assumed that the species i is a non-electrolyte or at least a electroneutral component as defined by Casassa and Eisenberg (Casassa and Eisenberg, 1964). For a single non-associating solute

$$a_i = y_i c_i \text{ and } \bar{c} = \sum_i c_i \qquad 5.9b$$

Where c_i is the weight concentration of species i and y_i is the activity coefficient on the same concentration scale.

Sedimentation equilibrium of a single solute species which self-associates to form an equilibrium mixture of states of different molecular weights

For each state equation 5.8 satisfies the conditions of sedimentation equilibrium as well as the condition that chemical equilibrium is maintained at each point in the cell (Nichol and Ogston, 1965). This can be illustrated by formulating the following ratio for a monomer-dimer system based on equation 5.9.

$$\frac{a_2(r)}{a_1^2(r)} = \frac{a_2(r_F)\exp\left\{\phi_2 2M_1(r^2 - r_F^2)\right\}}{a_1^2(r_F)\exp\left\{2\phi_1 M_1(r^2 - r_F^2)\right\}} \qquad 5.10$$

when $\phi_2 = \phi_1$, the partial specific volume of monomer and dimer, are identical and there is no volume change upon reaction. The exponential term in equation 5.10 cancels and the dimerization constant is identical at each point in the cell. Even when a volume change accompanies dimerization equation 5.9 is still valid as the correct dependence of the equilibrium constant on pressure can be predicted, consequent on changing radial distance (Howlett *et al.*, 1970).

5.1.2 The omega analysis

Milthorpe *et al.* (Milthorpe *et al.*, 1975) developed the theoretical background for the omega analysis where sedimentation equilibrium results obtained with polymerising systems can be analysed directly and therefore the errors inherent in differentiation and integration steps avoided. This analysis permits the evaluation of the thermodynamic parameter a_1, the activity of the monomer, in equilibrium with a series of polymeric species, as a function of total weight concentration $\bar{c}$. The omega function (Ω) provides a means of fitting reaction models to sedimentation equilibrium data obtained with polymerising systems. The method is capable of detecting possible volume changes inherent on polymer formation, of treating systems where activity coefficients of solute species are functions of total concentration and of describing the system in terms of relevant equilibrium constants (Milthorpe *et al.*, 1975).

Omega function *versus* protein concentration curves can be extrapolated to yield the indefinite dilution value, Ω^0, whereby the thermodynamic activity of the self-associating monomer can be determined at a reference point in the centrifuge cell. The activity of the monomer at the reference point can subsequently be used to calculate the activity of the monomer as a function of the total solute concentration (Milthorpe *et al.*, 1975). The activity of the monomer for a particular set of reaction conditions is a function only of the total solute concentration and can be used for the estimation of relevant equilibrium constants describing polymerising systems belonging to a particular model. For a thermodynamically ideal system, the equations required to fit the experimental data are those expressing the relations between m_1, the molar concentration of monomer (M_1 must be known), and $\bar{c}$, the total concentration of the protein. The activity of the monomer, a_1, a thermodynamic parameter, as a function of total concentration of the solute, $\bar{c}$, can also be used to evaluate non-ideality effects. (see later)

The experimentally obtained plot of $\bar{c}$ *versus* r may be analysed to yield the activity of the monomer a_1 as a function of total weight concentration $\bar{c}$ (Milthorpe *et al.*,

1975). This is achieved by defining an experimentally determinable dimensionless function $\Omega\ (r)$ as :

$$\Omega(r) = \frac{\bar{c}(r)\exp\left[\phi_1 M_1 (r_F^2 - r^2)\right]}{\bar{c}(r_F)} \qquad 5.11$$

$$\phi_1 = \frac{(1-\bar{\upsilon}_1\rho)\omega^2}{RT}$$

$\bar{\upsilon}_1$ is the partial specific volume of the effective monomer, ρ the solution density, ω^2 the angular velocity, R, the universal gas constant and T the absolute temperature of the sedimentation equilibrium experiment. M_1 is the effective-monomer molecular weight, $\bar{c}(r)$ and $\bar{c}(r_F)$ the total protein concentrations at radial positions r and r_F between or at r_m and r_b, the meniscus and base of the cell, respectively.

Provided M_1 and $\bar{\upsilon}_1$ are known and any reference radial distance, r_F, is chosen to lie between or at r_m and r_b, a plot construct of $\Omega\ (\ r\ )$ *versus* $\bar{c}(r)$ from experimental results can be obtained (in the case of the Beckmann Optima XL-A from absorption optical measurements) across the entire distribution. Raw data sets are measurements of $c\ (r\ )$ *versus* radial distance r.

By definition $\Omega(r) = 1$ at $\bar{c}(r_F)$ 5.11a

Rewriting equation 5.10

$$\Omega(r) = \frac{a_1(r_F)\bar{c}(r)}{a_1(r)\bar{c}(r_F)} \qquad 5.12$$

Equation 5.12 makes it possible to evaluate $a_1(r)$ as a function of $\bar{c}(r)$, as described below, without differentiation of experimental results that can introduce error. The value of $\bar{c}(r_F)$ used to compute $\Omega(\ r\ )$ may be selected as any value in the observed total concentration range.

As $\bar{c}(r) \rightarrow 0$, $a_1(r) \rightarrow \bar{c}(r)$ since the activity coefficient tends to unity and in addition $\bar{c}(r) \rightarrow c_1(r)$ since dilution favours dissociation. A plot of Ω (r) *versus* $\bar{c}(r)$ can be extrapolated to infinite dilution to yield a value for Ω^0.

$$\lim_{c(r)\rightarrow 0} \Omega(r) = \Omega^0 = \frac{a_1(r_F)}{\bar{c}(r_F)} \qquad 5.13$$

Since the selected value of $\bar{c}(r_F)$ is known, the monomer activity, $a_1(r_F)$ can be calculated. The value of $a_1(r_F)$ is then used to determine the monomer activity, $a_1(r)$, at any radial position and hence at any measured concentration. Equation 5.9 can be written in terms of $a_1(r)$ as follows

$$a_1(r) = a_1(r_F)\exp\left[\phi_1 M_1\left(r_F^2 - r^2\right)\right] \qquad 5.14$$

allowing a_1 to be calculated as a function of r .

A plot of $a_1(r)$ *versus* $\bar{c}(r)$ can now be constructed and fitted with various reaction models. The values of the reaction model parameters can be determined by simultaneous solution of the reaction model equations using the data from a number of arbitrarily chosen points (Milthorpe *et al.*, 1975). Alternatively the entire data set could be used for the fitting procedure using non-linear regression.

Combination of equation 5.12 and 5.13 simplifies the analysis and the significance of this extrapolated value can thus be appreciated

$$a_1(r) = \frac{\Omega^0 \bar{c}(r)}{\Omega(r)} \qquad 5.15$$

A plot of $\Omega(r)$ *versus* $\bar{c}(r)$ provides the value of Ω^0 which can then be used to construct a plot of $a_1(r)$ *versus* $\bar{c}(r)$ using equation 5.14.

By conducting several experiments with varying initial concentrations and/or angular velocities several $\Omega(r)$ *versus* $\bar{c}(r)$ plots can be constructed, extrapolated to indefinite dilution and the value of Ω^0 determined. To assist in the required extrapolation it is desirable to include an experiment which is of the meniscus depletion design so that the lowest measurable $\bar{c}(r)$ can be included in the analysis (Milthorpe *et al.*, 1975). The method permits correlation of results obtained in different experiments conducted over a reasonable range of concentrations. Provided chemical equilibrium has been reached and samples are homogeneous, coincident plots of $\Omega(r)$ *versus* $\bar{c}(r)$ can then be obtained in their common total concentration range for different experiments by selecting a $c(r_F)$ common to all (Milthorpe *et al.*, 1975). This is a very sensitive test for determining the presence of polymerising systems. The extrapolated value of Ω^0 obtained for all experiments are the same since the weight-fraction of monomer is a function of the variable $\bar{c}(r)$ and the equilibrium constants are pressure independent. It follows that Ω^0 of this coincident plot can be used to calculate the activity of the monomer at the reference concentration and hence values of $a_1(r)$ are available for the entire concentration range encompassed by all experiments using equation 5.15. Different reference points within the concentration distribution will provide the same result.

The advantage of this method is that the thermodynamic activity of the monomer a_1, as a function of total weight concentration $\bar{c}$, is obtained rather than an apparent quantity which defines non-ideality effects in a less explicit way. For a single non-associating solute the omega analysis directly yields the concentration dependence of the activity coefficient, y_i, since $a_i = y_i c_i$ (Jeffrey *et al.*, 1977). (See section on non-ideality effects).

Extrapolation of the omega function

The accuracy of the analysis described above is highly dependent on the accuracy with which the $\Omega(r)$ *versus* $\bar{c}(r)$ curve can be extrapolated to zero concentration.

The extrapolation of an omega function curve to infinite dilution for a nearly ideal, weakly self-associating solute is relatively straightforward because there is little or no curvature in the omega function near zero concentration. This is demonstrated by the fact that when a small reference concentration is chosen, the intercept has a value close to its maximum possible value of 1. In this way if an error is made in the extrapolation, the percentage error is likely to be small and therefore the error in $a_i\,(r_F)$ and other monomer activities will be small (Morris and Ralston, 1985).

As the strength of the self-association increases extrapolation becomes more difficult. For mild-associations the omega function curve becomes steeper at low concentrations and the curvature in this region becomes more difficult to determine. Even if a small reference concentration is chosen, the intercept value will be substantially smaller than 1.0. In these cases, a small error in the extrapolation will result in a large percentage error in the intercept value and hence the monomer activity values. This percentage error will tend to increase as the strength of the self-association increases. Therefore for mild and strong self-associating systems there is a possibility of rejecting the correct self-associating model because of an error made in extrapolating the omega function curve to zero concentration.

In principle the extrapolation and its associated problems can be avoided entirely by fitting the omega function directly to a reaction model equation, since the omega function, like the activity function, is dependent only on the total solute concentration and can be written in terms of the self-association parameters (Morris and Ralston, 1985).

5.1.3 Associating systems at low total solute concentrations

Over a range of low $\bar{c}$ the system may be considered, to a good approximation, as thermodynamically ideal (all $y_i \cong 1$). In this region therefore the omega analysis yields directly corresponding values of m_1 ($m_1 = a_1 / \overline{M}_w$) and $\bar{c}$, appropriate for use in self-associating model equations (see later). The results obtained in this way can be

used to explore the nature of the association pattern, using the available theory, and to obtain first estimates of relevant equilibrium constants.

The sedimentation equilibrium equation

$$\frac{d\ln a_i(r)}{dr^2} = \frac{M_i\omega^2(1-\bar{\upsilon}_i\rho)}{2RT} \qquad 5.8$$

can be rewritten assuming that all $y_i \approx 1$; $\phi_i = \phi$ and using equation 5.9

$$\phi_i = \phi = \frac{(1-\bar{\upsilon}\rho)\omega^2}{RT}$$

as

$$\frac{d\ln c_i(r)}{d(r^2)} = \phi M_i \qquad 5.16$$

It follows that

$$\frac{d\ln \bar{c}(r)}{d(r^2)} = \phi \overline{M}_w \qquad 5.17$$

The slope of the tangent to a plot of $\ln \bar{c}$ *versus* r^2 obtained from the sedimentation equilibrium experiments yield, with the knowledge of ϕ, the weight-average molecular weight of the system at the corresponding value of $\bar{c}$. The monomer molecular weight, M_1, can be obtained as it is equal to $\overline{M}_w$ at $\bar{c} \rightarrow 0$. An increase in $\overline{M}_w$ with increasing $\bar{c}$ can give the first insight into the nature of the association.

5.1.4 Associating systems at high values of total solute concentrations $\bar{c}$: non-ideality effects

Over a range of high $\bar{c}$ solute concentrations thermodynamic non-ideality must be included in all calculations where $\bar{c}$ is present in the equation. This thermodynamic non-ideality correction is demonstrated in the following equation

$$\ln y_i = \sum_j \alpha_{ij} m_i \qquad 5.18$$

where each of the subscripts *i, j* is allowed to span the set of monomeric and polymeric species independently. Only the first term of the expansion has been retained because a relatively low total concentration have been considered. This approximation is applicable to experiments with insulin due to its low solubility in aqueous solutions (up to approximately 4 mg/mL).

The a_{ij} may be calculated by using the expression (Wills *et al.*, 1980)

$$\alpha_{ij} = \frac{4\pi N\left(r_i + r_j\right)^3}{3} + \frac{Z_i Z_j\left(1 + xr_i + xr_j\right)}{2I\left(1 + xr_i\right)\left(1 + xr_j\right)} - M_j \bar{\upsilon}_j \qquad 5.19$$

where the first term represents the co-volume contribution, based on spherical geometry; r_i and r_j being radii of the impenetrable spheres used to approximate protein molecules; the second term gives the charge-charge interactions, born by the spheres in terms of their net charge Z_i and Z_j, the ionic strength, I and the Debye inverse-screening length x; and the third term expresses the molar volume of species j. The required Stokes radii and electrostatic charges for the different insulin species were described in detail previously (Mark *et al.*, 1987).

5.2 Review of the self-association of zinc-free insulin

5.2.1 The indefinite duoisodesmic (IDI) model of self-association

No marked volume change accompanies the self-association of insulin especially in the low concentration range (Howlett *et al.*, 1970). Thus equation 5.5 and its differentiated form are applicable to the sedimentation equilibrium distribution of all insulin species which self-associate to form an equilibrium mixture of different molecular weights.

The pattern of self-association of zinc-free bovine insulin has been extensively studied by a wide variety of techniques including nuclear magnetic resonance (NMR) (Bradbury *et al.*, 1981; Kadima *et al.*, 1992), circular dichroism (CD) (Goldman and Carpenter, 1974), sedimentation velocity (Jeffrey *et al.*, 1976), sedimentation equilibrium (Jeffrey *et al.*, 1976) and light scattering (Kadima *et al.*, 1993).

Mark *et al.* (Mark *et al.*, 1987) have undertaken an extensive study of the self-association pattern of zinc-free bovine insulin in solution over a wide range of experimental conditions of pH, ionic strength and temperature. Sedimentation equilibrium was chosen as the method of study as precise information can be obtained on the dependence of both weight-average molecular weight and of concentration of the monomer m_1 on total protein concentration. Mark *et al.* (Mark *et al.*, 1987) based their model on interactions observed in the crystal between subunits where insulin acts as a bi-functional monomer having two distinct non-identical interfaces both capable of self-interaction. Experimental data from sedimentation equilibrium experiments were fitted to the proposed model of self-association described by a mathematical equation in terms of the dependence of weight-average molecular weight and monomer concentration on total protein concentration (Nichol *et al.*, 1984). By relating pH, temperature and ionic strength dependence of the association constants with properties of various amino-acid residues on the surface of the insulin monomer, each constant has been tentatively assigned to a particular reaction domain.

The inclusion of terms allowing for thermodynamic non-ideality in the analysis made it the most comprehensive treatment of the insulin system available.

The proposed model of Mark *et al.* (Mark *et al.*, 1987), called indefinite duoisodesmic (IDI), can be described as the indefinite self-association of a bivalent insulin monomer along the monomer-monomer and dimer-dimer interfaces governed by two binding constants K_A and K_B respectively. The two initial pathways of dimerization are followed by the indefinite addition of monomer under the control of the same two equilibrium constants.

The monomer possesses two independent non-identical self-association sites designated α (monomer-monomer interface) and β (dimer-dimer interface), both capable of self-interaction. This account for two different sets of interactions between monomers in the crystallographically determined zinc hexamer structure. These are defined as interactions about two 2-fold axis, OP and OQ, respectively (Baker *et al.*, 1988; Bundell *et al.*, 1972).

Two types of dimers are formed, one involved in an α-α interaction, leaving the two β sites exposed and governed by an association constant K_A; the other involving β-β interaction, leaving two α sites exposed and governed by an association constant K_B. Linear chain growth proceeds by successive addition of monomer so that all polymers, both odd and even numbered, coexist in equilibrium. Each of these polymers possesses alternating α-α and β-β bonds with even numbered polymers having either two α-sites or two β-sites exposed and odd numbered polymers having an α-site at one end and β-site at the other. The authors postulated it is likely that given the relative magnitudes of K_A and K_B that they referred to the interactions about OP and OQ respectively. In other words the interactions about the OP interface (represented by K_A and also referred to as the monomer-monomer interface) appears to be stronger than the interactions about the OQ interface (represented by K_B and also referred to as the dimer-dimer interface).

The conclusion made by the authors (Mark *et al.*, 1987) was that the α-α interactions may be identified with the OP interactions and the β-β interactions most certainly exist in solution. Thus it is possible to describe the solution behaviour of zinc-free insulin in terms of a single association pattern at neutral pH and 25 °C. In general the solution will contain an equilibrium mixture of all polymeric forms of insulin of which the composition depends on the total concentration and the values of K_A and K_B for a specific environment. The hexamer is merely one of these species and has no particular significance.

The mathematical equation developed by Nichol *et al.* (Nichol *et al.*, 1984) describing this model taking both K_A and K_B into consideration and distinguishing between the non-identical self-association domains is as follows

$$\bar{c} = M_1 m_1 \left[\frac{(1+2K_A m_1)(1+2K_B m_1)}{1-4K_A K_B m_1^2} \right] \qquad 5.20$$

where M_1 is the molecular weight of the monomer, m_1 its concentration on the molar scale and $\bar{c}$ is the total protein concentration in g/L. Thus the above mathematical equation describes the composition of a solution consisting of an infinite array of odd- and even numbered polymeric species. The weight-fraction of each species can be calculated as a function of total insulin concentration (see later, equation 5.17).

Mark *et al.* (Mark *et al.*, 1987) developed an expression describing the dependence of "reduced" weight-average molecular weight ($\overline{M}_w / M_1$) on total concentration $\bar{c}$.

$$\frac{\overline{M}_w}{M_1} = \frac{P}{Q} \qquad 5.21$$

where

$$P = 16K_A^2K_B^2m_1^4 + 16K_AK_B(K_A + K_B)m_1^3 + 24K_AK_Bm_1^2 + 4(K_A + K_B)m_1 + 1$$

$$Q = (1 - 4K_AK_Bm_1^2)(1 + 2K_Am_1)(1 + 2K_Bm_1)$$

When $K_A = K_B$ i.e. the association sites are equivalent, the above equation reduces to the same form as the equation for an indefinitely self-associating system governed by a single equilibrium constant, K_I (isodesmic) (Tang *et al.*, 1977).

$$\frac{\overline{M}_w}{M_1} = \frac{1 + K_Im_1}{1 - K_Im_1}; K_Im_1 < 1 \qquad 5.22$$

with

$$2K_A + 2K_B = K_I$$

Equations 5.14 and 5.15 describe a smooth monotonically increasing dependence of $\overline{M}_w / M_1$ on $\overline{c}$.

5.2.2 Weight-fractions of species

Mark *et al.* (Mark *et al.*, 1987) formulated an expression in terms of the weight-fraction of species as a function of total concentration for the IDI model. The weight-fraction of the insulin species present in solution under the relevant conditions as a function of total insulin concentration can be calculated provided K_A and K_B are known.

The weight-fraction, ϕ_i, of each species, *i* being given by:

$$\phi_1 = \frac{c_i}{\bar{c}} = \frac{iX_i m_1^{i-1}(1-4K_A K_B m_1^2)^2}{(1+2K_A m_1)(1+2K_B m_1)} \qquad 5.23$$

where

$$X_i = (4K_A K_B)^{(i-1)/2} \qquad i \text{ odd} \qquad 5.23a$$

$$X_i = (K_A + K_B)(4K_A K_B)^{(i-2)/2} \qquad i \text{ even} \qquad 5.23b$$

The complex distribution of species within the insulin system in solution can be represented graphically. A distinction can be made between the weight-fraction of the two types of dimer α and β and tetramer α and β. It stresses the importance of the odd-numbered species, the dominant species by weight in bovine insulin under these particular conditions. The possible presence of odd-numbered species other than monomer have been taken into account.

Thermodynamic non-ideality effects are assessed basically in terms of co-volume and charge-charge interaction effects. Non-ideality effects in the present system are not large due to the relatively small size of insulin, the small net charge borne by it and the low concentration range (0- 4 mg/mL) that was examined. Insulin is insoluble at higher concentrations in these solution conditions.

5.2.3 Comparison of IDI with other proposed models

Mark *et al.* (Mark *et al.*, 1987) formulated a model based on interactions observed in the crystal between subunits and showed that it could be satisfactorily applied to describe insulin's self-association over all sets of experimental conditions.

The question that inevitably arises is whether more than one model can be found that fits all conditions and, if so, what might be the consequences for instance for the

pharmacological applications of insulin in a concentration range where the self-association is likely to be significant. Mark and Jeffrey (Mark and Jeffrey, 1990) addressed this problem in a subsequent paper where they have made similar systematic attempts to fit all of the available data with other reasonable models.

Mark and Jeffrey (Mark and Jeffrey, 1990) applied data fitting procedures both with and without correction for non-ideality to experimentally measured concentration distributions of zinc-free insulin obtained over a wide range of experimental conditions (pH 2, 7, and 10, ionic strengths 0.05 and 0.1, T = 25 °C and 37 °C) using four different models of self-association patterns based on known physico-chemical properties of insulin in solution and patterns already in the literature. All the models have in common the inclusion of the monomer, known to be present in solution from earlier studies (Jeffrey, 1974; Jeffrey and Coates, 1966b), and the other species , the dimer and the hexamer, known to be stable (Blundell *et al.*, 1972).

The four models considered are the following:

(i) Indefinite isodesmic self-association of monomers (IIM).

(ii) An initial dimerisation, followed by indefinite isodesmic self-association of dimers (IID).

(iii) Two initial pathways of dimerisation followed by the indefinite addition of monomer under the control of the same two equilibrium constants (IDI).

(iv) Dimerisation, followed by hexamerization, and then indefinite isodesmic self-association of hexamers (IIH).

The IID model (ii) (Jeffrey *et al.*, 1976) describes the self-association of insulin at pH 7; The IDI model (iii), described in equation 5.20, is the pattern based on interaction in the crystal that has already been used (Mark *et al.*, 1987) to fit results over a wide range of experimental conditions, mentioned above, and discussed extensively in a paper by Mark *et al.* (Mark *et al.*, 1987), and the IIH model (iv) by Pekar and Frank (Pekar and Frank, 1972) to fit the results obtained with zinc-free pig insulin at pH 7. The IIM model (i) (not published before) was included in this analysis as being the simplest one that could be formulated if account is taken of measured

molecular masses ranging from less than that of the insulin dimer to greater than that of the hexamer.

Values for the association constants and their uncertainties for all 4 models, fitted by both the ideal and non-ideal procedures, were calculated. In almost all instances the constants evaluated, when ideality was assumed, were well within the limits of uncertainty characteristic of the non-ideal computations. This suggests that there was little to gain in the relatively low concentration range considered here by the non-ideality correction (Mark and Jeffrey, 1990).

Comparing the standard deviations for the models and conditions it was found that IID, IDI and IIH were all acceptable models for all the experimental conditions if the fits including non-ideality were discounted. The model IDI that was originally proposed for these conditions (Mark *et al.*, 1987) was the only one that was satisfactory when non-ideality corrections were included in the analysis. Model IIM failed badly at pH 2 and was therefore excluded by the authors. They could, however, not justifiably exclude either of the other models.

The 4 models have been compared at three total insulin concentrations of practical importance, namely "physiological", 3ng/mL (0.5 nM); "pharmacological" , 3 mg/mL (0.5 mM); and "physico-chemical", 30 mg/mL (5mM). The calculated values of the association constants were used to compute the concentrations and mass fractions of any desired insulin species as a function of total insulin concentration for all of the models and all of the conditions. The calculations were performed using the constants obtained at pH 7, $I = 0.1$ and 25 °C, the conditions closest to those appropriate to the three total concentrations chosen.

The conclusions arrived at were as follows:

(i) At serum concentrations (" physiological ") the distribution is independent of the model, essentially all of the protein is monomeric.

(ii) At any of the three concentrations the fraction of monomer is almost independent of the model as indeed is the fraction of the dimer.

(iii) At the pharmacological concentration regardless of the model, the fraction of monomer, the active species, is very low, being less than 10%.

(iv) At the concentration required for the studies by NMR spectroscopy of the solution structure of insulin (" physico-chemical "), the fraction of monomer, the molecule of interest here, is extremely low whatever model is considered.

(v) All models show that at concentrations well above the serum level, insulin solutions contain a very complex set of polymers of different sizes, dominated by species hexameric or bigger.

The above findings were independent of the model chosen to represent the self-association, and may be taken to reflect real properties of insulin at different concentrations.

In an attempt to make monomeric insulins at pharmacological concentrations (0.6 mM) which have largely preserved their biological activity, single amino-acid substitutions have been introduced in mainly the monomer-monomer interface of the hormone (Brange *et al.*, 1988). The term, "association state", has been used to monitor changes in self-association characteristic of these modified insulins. Association state is defined as a number which for a thermodynamically ideal solution, at a given concentration, is equal to the ratio of the number-average molecular mass, $\overline{M}_n$, to the molecular mass of the monomer. (The number-average was chosen because only osmotic pressure measurements were available to these workers to estimate molecular weight changes).

$$\overline{M}_n = \frac{\Sigma M_i n_i}{\Sigma n_i} \qquad 5.24$$

where $\overline{M}_n$ is the number -average molecular mass and n_i the number of moles. For a homogeneous solute, the ratio of $\overline{M}_w$ to $\overline{M}$ is 1.0. The greater the degree of heterogeneity the greater the ratio of $\overline{M}_w / \overline{M}_n$.

An expression for the number-average molecular mass in terms of the IDI model of the self-association pattern of insulin can be derived as follows

$$\overline{M}_n = \frac{\bar{c}\left[1-(4K_A K_B)m_1^2\right]}{m_1\left[1+(K_A+K_B)m_1\right]} \qquad 5.25$$

The fractions of any desired polymeric species at any specified pair of values of $\overline{M}_n$. $\overline{M}_n$ and $\bar{c}$ can thus be calculated. The process can be applied to especially modified insulins and for comparison to give insight into their particular association state.

In the IDI model equation the interaction at the monomer-monomer interface is represented by K_A and the interaction at the dimer-dimer interface by K_B (Equation 5.23). Mark and Jeffrey (Mark and Jeffrey, 1990) determined by calculation the K_A values for these modified insulins by determining what changed value of K_A corresponds to the paired values of $\overline{M}_n$ and $\bar{c}$ measured by osmometry by using the IDI model equation and leaving K_B values unaltered. This assumption was reasonable because the amino-acid residues in the dimer-dimer interface, represented by K_B in the IDI model, were unmodified in zinc-free insulin at pH 7. The fractions of any polymers at any total insulin concentration desired were then calculated using equation 5.23. The "association state", calculated when K_A is given the value that it has in unmodified zinc-free insulin at $\bar{c}$ = 0.2 mM, is 3.1, in exact agreement with that measured by osmometry (Brange *et al.*, 1988). This agreement confirms that model IDI provides a valid description of insulin self-association

The conclusion that emerges from studying these results refered to above (Brange *et al.*, 1988), is that the modified insulins of about normal activity are not likely to be more than 35 % monomeric under the reported conditions. In fact there are more dimers than monomer at each concentration for all the values of K_A. The apparently paradoxical increase in the fraction of dimer as K_A is weakened reflects the formation of dimer at the expense of higher species via the β-β interaction in this model. The calculated fractions show that the amino-acid substitutions have resulted in a doubling or trebling of the amount of monomeric insulin at each concentration, consistent with

the reported two to three times better absorption and with changes detected quantitatively in circular dichroism spectra.

It emerges from the calculations that the theoretical lower limit (when $K_A = 0$) for the association state for the concentration range studied here is 1.5. If one could confirm that model IDI is indeed the correct description of self-association, these results would imply that the only way an insulin with a higher proportion of monomer could be produced is to modify appropriate residues in the β-β interaction region.

Thus the IDI model is the favoured model, both on the grounds that it is most consistent with the available information from structural studies and that it is the best fit to the experimental data.

It must be noted that some of the values of the association state estimated by osmometry with these modified insulins are slightly lower than the theoretical lower limit of 1.5. The operation of thermodynamic non-ideality depresses these values, especially due to the introduction of extra charged side chains into the insulin molecule as was the case (Brange *et al.*, 1988). In addition, conformational changes induced by modification of the native molecule could also affect the β-β interaction, thereby lowering the K_B . Even without using a specific model as a basis for calculation, the values of the association state imply a relatively low level of monomer at the concentrations used.

The analysis emphasises that it is important to employ a model for the self-association process that defines the composition of the solution under a given set of conditions in a way that reflects as closely as possible the real state of affairs. Rationally designed attempts to stop the self-association would be more productive if it were known that a second route of dimerization, as in model IDI, is actually utilised.

In conclusion it is highly desirable to define as precisely as possible the types and proportions of insulin species present in solutions of the protein. To be able to define uniquely the self-association pattern, experimental approaches that distinguish

quantitatively between different pathways of polymer formation must be devised. It is also desirable to concomitantly, also experimentally, remove as many as possible of the assumptions that make thermodynamic non-ideality corrections ambiguous. This information is important in the design of strategies to suppress insulin polymerisation completely while retaining full activity in the modified monomer.

5.3 Materials and methods

Glassware that had been in contact with the insulin samples was treated with 0.1 mM EDTA for 24 hours to remove contaminating divalent cations followed by thorough rinsing in HPLC grade distilled water. Protein and peptide concentrations were determined spectrophotometrically using a Varian series 634 spectrophotometer. All dialysis was performed at 4 °C using Spectrapor No 3 membrane tubing (M_w cutoff 3,500 daltons). The tubing was thoroughly rinsed with HPLC grade distilled water. Solutions were dialysed against a minimum of 100 volumes of buffer for at least 24 hours with the buffer being changed at least twice. Experimental solutions were allowed to equilibrate for at least 10 hours after the final buffer change before ultracentrifugation.

Three insulin preparations were used in these self-association studies. These were bovine insulin, totally chemically synthesised platypus insulin and totally chemically synthesised HB10D platypus insulin.

5.3.1 Bovine insulin preparation

Bovine insulin (Sigma) was purified before use in a PrePak Waters RCM Base preparative Reverse Phase (RP) HPLC column as follows: Crystalline bovine insulin (approximately 10 mg) was dissolved in 100 mM H_3PO_4/20 mM Triethylamine/50 mM $NaClO_4$ pH 3.0 buffer containing 20% CH_3CN. This solution was injected on to a preparative RP-HPLC column and purification method V (Table 4.1, Chapter 4) was followed. Bovine insulin eluted in a major peak at 31% CH_3CN on the linear gradient and was lyophilised. The lyophilisate was dissolved in 10% CH_3CN plus 0.09% aqueous TFA , total volume 10 mL, and desalted on the preparative RP-HPLC column using method IV (Table 4.1, Chapter 4). Bovine insulin as a single symmetrical Gaussian peak. The identity of the insulin was confirmed by molecular mass analysis (5734 daltons) using a matrix-assisted laser desorption ionisation (MALDI) time-of-flight mass spectrometery. This lyophilised product was used in all the experiments.

5.3.2 Sedimentation equilibrium studies

Only samples judged to be homogeneous by HPLC, mass analysis and high performance capillary electrophoresis were used in the experiments. The composition of the buffer employed in all the sedimentation equilibrium studies was as follows: 20 mM Tris, 18 mM HCl, 79 mM NaCl, 1 mM EDTA, pH 7.0, ionic strength, *I*, 0.1 M. The concentrations of the insulins in solution were determined spectrophotometrically at 276 nm employing an extinction coefficient of 10.5 for a 1% solution, 1 cm light path length (Frank and Veros, 1968). These were substantiated again when measured in a Beckman Optima XL-A using its optical system and a light path length of 1.2 cm during sedimentation equilibrium studies. Dilutions where necessary were performed with dialysate.

Sedimentation equilibrium experiments were carried out in a Beckman Optima XL-A Analytical Ultracentrifuge (Beckman Instruments, Inc., Fullerton, Ca, U.S.A.) and final sedimentation equilibrium concentration distributions were obtained from scans at 280 nm using its absorbance optical system. Dialysed solution (100 μL) and dialysate (110 μL) were placed in the sample and solvent sector respectively of a double sector carbon filled epon centrepiece. Experiments were conducted at different speeds and initial concentrations as detailed later (Table 5.1). Baseline scans at 360 nm were subtracted to remove the effects of oil droplets on lenses and windows, as well as of optical imperfections in the windows and lenses.

5.3.3 Analysis of sedimentation equilibrium data

The computer program XLACALC/RALSTON/XLAMW , using the slope of the tangent to the plot of ln $\bar{c}$ *versus* r^2 (equation 5.17), was employed to calculate the weight-average molecular weight, $\overline{M}_w$, as a function of total insulin concentration $\bar{c}$ after the baseline has been subtracted with the computer program XLACALC/RALSTON/XLABASE. The solution density ρ has been taken as

1.00 g/mL and it has been assumed that $\bar{\upsilon}$, the partial specific volume of insulin, is the same for all insulin species namely 0.725 mL/g (Jeffrey, 1974; Jeffrey, 1986). Non-ideality effects were not taken into account in the calculations used in this thesis but $\overline{M}_w$ was used rather than $\overline{M}_{w,app}$ for convenience in nomenclature.

The program XLACALC/RALSTON/XLAMW was also employed to calculate omega values as a function of total concentration using equation 5.11. A reference concentration of $\bar{c}(r_F) = 0.50$ g/L was chosen for all omega analyses. The value of the reference concentration used to compute the omega plots may be selected as any value in the observed total concentration range (Milthorpe *et al.*, 1975).

For the "reduced" weight-average molecular weight ($\overline{M}_w / M_1$), versus $\bar{c}$ plots the monomer molecular weight M_1 values for bovine insulin, 5734 daltons; platypus insulin, 5865 and HB10D platypus insulin, 5843 daltons, previously calculated with the PEPTIDESORT program from GCG (Chapter 3), were used in these calculations.

All other equations referred to in this section and used for the calculations were installed in the application program GraphPadPRISM™ Version 1.0 from Graphpad Software Inc., San Diego, CA, U.S.A. The non linear regression algorithm of this application program was employed in the curve fitting procedures.

5.4 Results

5.4.1 The dependence of "reduced" weight-average molecular weight on total bovine, platypus and HB10D platypus insulin concentration

In the sedimentation equilibrium experiments conducted with zinc-free insulin, the ionic strength was fixed at 0.1 M and the temperature 25 °C to permit the correlation of results over a range of total insulin concentration at 0-1.5 g/L. This correlation was achieved in several experiments (Table 5.1) by employing different loading concentrations and angular velocities. In this range of protein concentrations it is

Table 5.1: A summary of the experimental conditions applicable to sedimentation equilibrium experiments conducted with zinc-free insulin. Optical path length 1.2 cm and $\bar{c}_0$ starting concentration.

Insulin	Experimental run	$\bar{c}_0$ (g/L)	Speed (rev/min)	Time (hour)
Bovine				
	bin3a1	1.38	30 K	6
	bin3a2	0.50	30 K	6
	bin3a3	0.25	30 K	6
	bin4a2	0.50	50 K	12
	bin4a3	0.25	50 K	12
Platypus	plin3a1	0.37	30 K	6
	plin5a1	0.85	30 K	6
	plin5a3	0.86	30 K	6
	plin6a3	0.86	42 K	12
HB10D Platypus				
	doa1a2	0.34	30 K	6
	doa1a3	0.66	30 K	6
	doa2a2	0.34	42 K	12
	doa2a3	0.66	42 K	12

assumed that the samples obey the Beer-Lambert Law where absorbance and concentration are proportional. The zinc-free buffer employed in all cases was 20 mM Tris, 18 mM HCl, 79 mM NaCl, 1 mM EDTA, pH 7.0, ionic strength, I, 0.1 M.

The weight-average molecular weight $\overline{M}_w$ and the omega values as a function of total concentration were calculated for the three insulins using the RALSTON computer program. An increase in $\overline{M}_w$ with increasing total concentration can give the first insight into the nature of the association.

Plots of $\overline{M}_w / M_1$ (the "reduced" weight-average molecular weight) as a function of total protein concentration $\overline{c}$ were constructed for all the insulins for all experiments listed in Table 5.1. All the results from all experiments for bovine insulin (Figure 5.1), regardless of the loading concentrations and angular velocity, were coincident. The same was found for the experiments with platypus (Figure 5.2) and HB10D platypus insulins (Figure 5.3). This supports the observation that all the insulin samples were homogeneous. It is clear from the above plots for bovine and platypus insulin that the $\overline{M}_w / M_1$ increases with increasing total protein concentration, an indication of self-association. The plots for HB10D platypus insulin (Figure 5.3) also indicate self-association but to a lesser extent. In fact, this value does not increase above two at a total concentration of 0.5 g/L and higher. This is an indication that the major associated species of this analogue is dimer over the concentration range and conditions used.

5.4.2 Omega Analysis.

Omega values $\Omega(r)$, as a function of total concentration $\overline{c}(r)$, using equation 5.11 were calculated for the three insulins using the RALSTON computer program. All data sets listed in Table 5.1, used for the calculation of $\overline{M}_w$ as a function of total concentration using equation 5.17, were used in these calculations. Different

Figure 5.1

The overall extent of association of zinc-free bovine insulin (pH 7.0; I = 0.10 M; T = 25 °C) reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets mbin3a1 (o) and mbin3a2 (•)(Table 5.1) are displayed. The K_A and K_B values, shown in Table 5.2, were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\overline{c}$ using equation 5.21. This calculated solid curve is displayed superimposed on the experimental points.

Figure 5.1

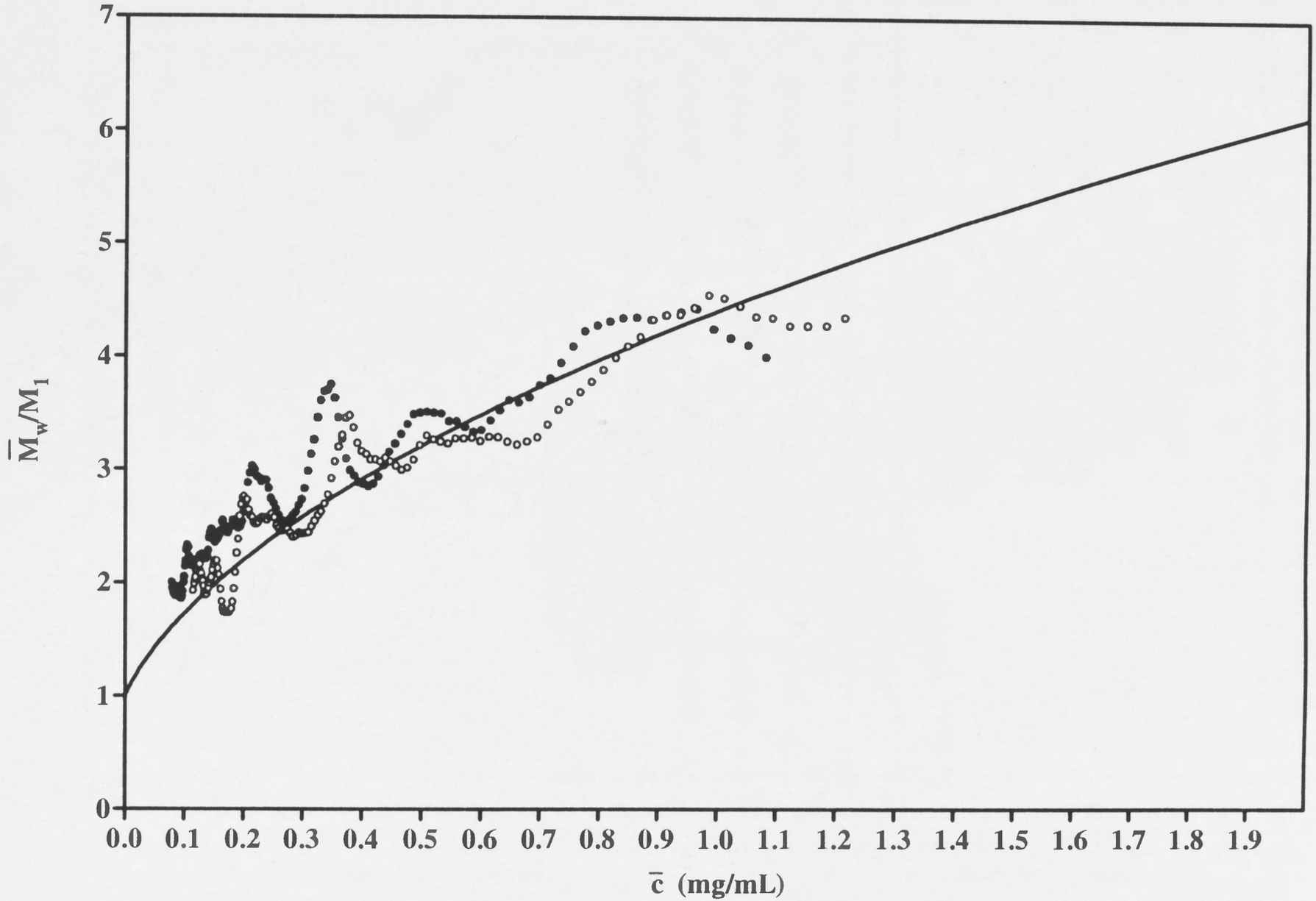

7
6
5
4
3
2
1
0
$\bar{M}_w/M_1$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9
$\bar{c}$ (mg/mL)

Figure 5.2

The overall extent of association of zinc-free chemically synthesised platypus insulin reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets mplin5a1 (•) and mplin3a1 (o) (Table 5.1) are displayed. The K_A and K_B values, shown in Table 5.2, were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\bar{c}$ using equation 5.21. This calculated solid curve is displayed superimposed on the experimental points.

Figure 5.2

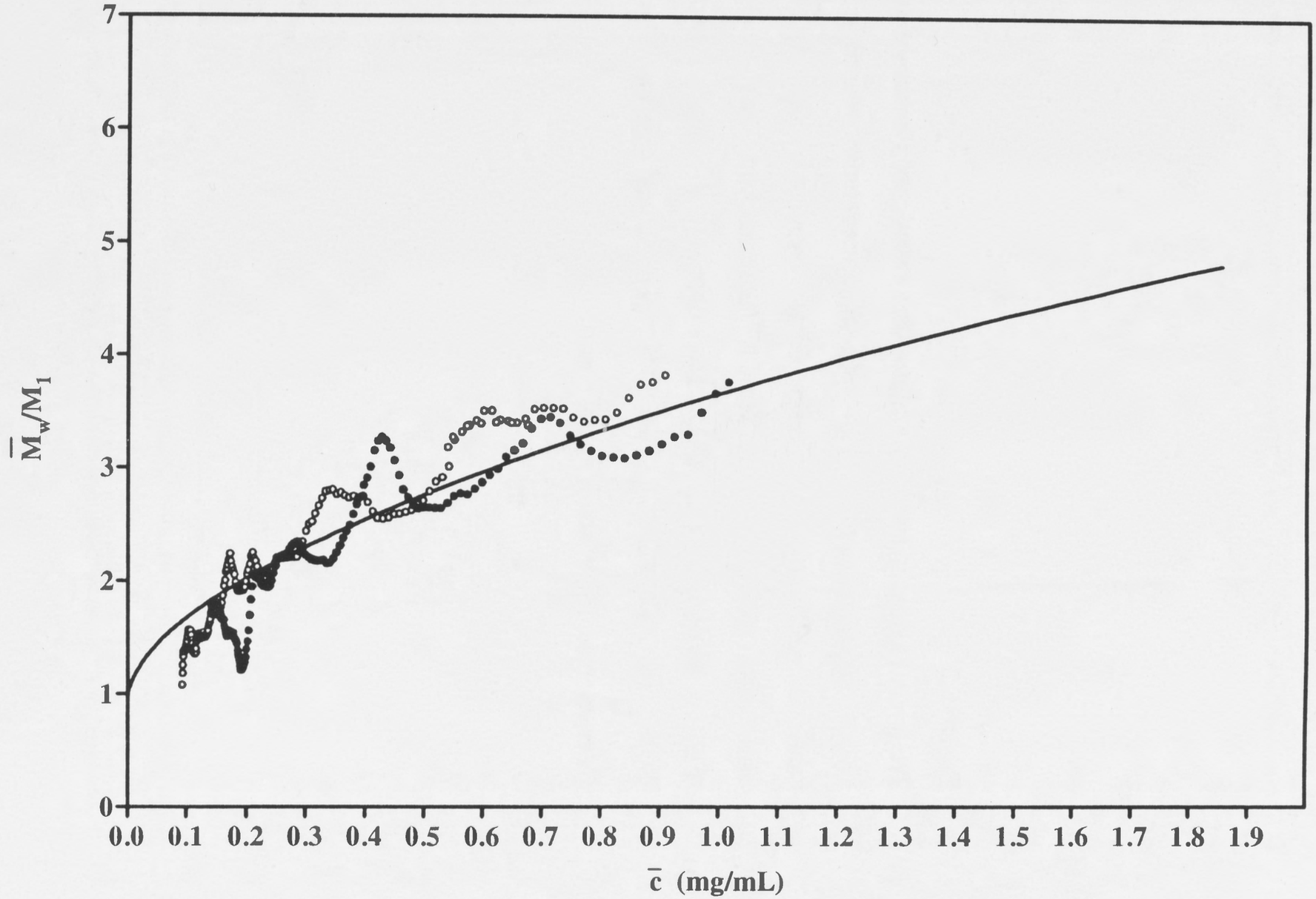
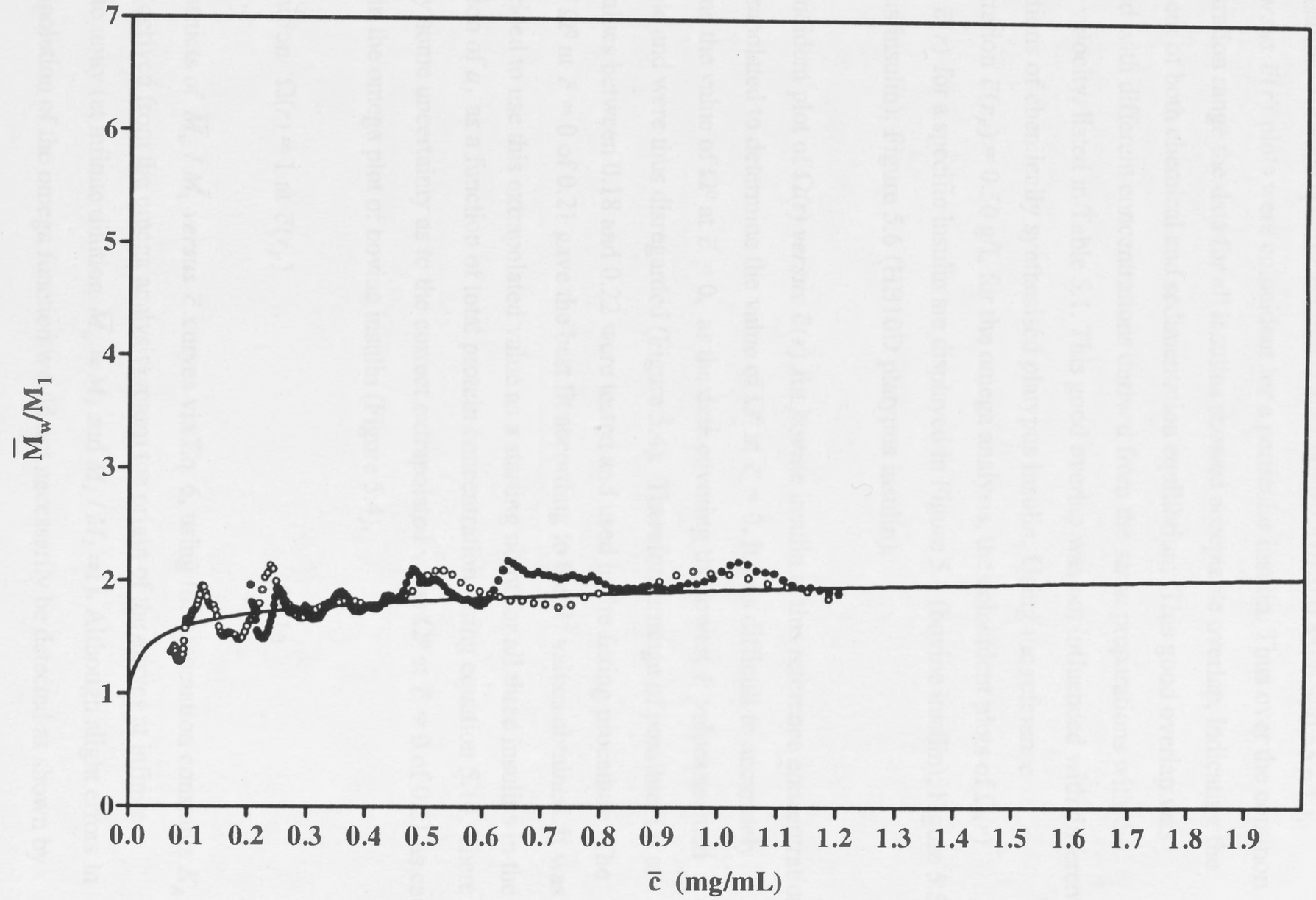

7
6
5
4
3
2
1
0
$\bar{M}_w/M_1$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9
$\bar{c}$ (mg/mL)

Figure 5.3

The overall extent of association of zinc-free chemically synthesised HB10D platypus insulin reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets mdoa1a3 (•) and mdoa2a3 (o) (Table 5.1) are displayed. The K_A and K_B values, shown in Table 5.2, were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\bar{c}$ using equation 5.21. This calculated solid curve is displayed superimposed on the experimental points.

Figure 5.3


7
6
5
4
3
2
1
0
$\bar{M}_w/M_1$
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9
$\bar{c}$ (mg/mL)

reference concentrations within the concentration distribution of each insulin, namely 0.4; 0.45; 0.5; 0.55 g/L respectively, were used in the omega analysis.

Plots of $\Omega(r)$ *versus* $\bar{c}(r)$, for each reference concentration, were constructed for the three insulins. For each specific reference concentration (0.4; 0.45; 0.5; 0.55 g/L), $\Omega(r)$ *versus* $\bar{c}(r)$ plots were coincident for a particular insulin. Thus over the common concentration range the data for all insulins showed acceptable overlap, indicating the attainment of both chemical and sedimentation equilibrium. This good overlap was achieved with different concentrations derived from the same preparations with the angular velocity, listed in Table 5.1. This good overlap was not influenced with different preparations of chemically synthesised platypus insulin. Using the reference concentration $\bar{c}(r_F) = 0.50$ g/L for the omega analysis, the coincident plots of $\Omega(r)$ *versus* $\bar{c}(r)$ for a specific insulin are displayed in Figure 5.4 (bovine insulin); Figure 5.5 (platypus insulin); Figure 5.6 (HB10D platypus insulin).

The coincident plot of Ω(r) *versus* $\bar{c}(r)$ for bovine insulin at this reference concentration was extrapolated to determine the value of Ω^0 at $\bar{c} = 0$. It was difficult to accurately determine the value of Ω^0 at $\bar{c} = 0$, as the data covering the lowest $\bar{c}$ values seemed unreliable and were thus disregarded (Figure 5.4). Therefore a range of possible Ω^0 at $\bar{c} = 0$ values between 0.18 and 0.22 were tested and used in the fitting procedure. The value of Ω^0 at $\bar{c} = 0$ of 0.21 gave the best fit according to the R^2 values obtained. It was thus decided to use this extrapolated value as a starting point for all three insulins in the calculation of a_1 as a function of total protein concentration, using equation 5.14. There is clearly some uncertainty as to the correct extrapolated value Ω^0 at $\bar{c} = 0$ of 0.21 as can be seen in the omega plot of bovine insulin (Figure 5.4).

By definition $\Omega(r) = 1$ at $\bar{c}(r_F)$ 5.11a

and constructs of $\overline{M}_w / M_1$ *versus* $\bar{c}$ curves via Eq. 6, using the association constants K_A and K_B (derived from the omega analysis) return the origin of the curves at infinite dilution to unity (at infinite dilution $\overline{M}_w = M_1$ and $M_1 / M_1 = 1$). Although slight errors in the extrapolation of the omega function would not necessarily be detected as shown by

Figure 5.4

An illustrative plot of the omega-method used to analyse the sedimentation equilibrium results obtained with zinc-free bovine insulin. All the experimental data sets listed for bovine insulin in Table 5.2 were used in the analysis. The ordinate values were obtained using equation 5.11 utilising the common reference concentration of 0.5 g/L, while the abscissa values refer to the corresponding total concentrations for which omega values were determined. The extrapolation to give $\Omega^0 = 0.21$ is shown.

Figure 5.4

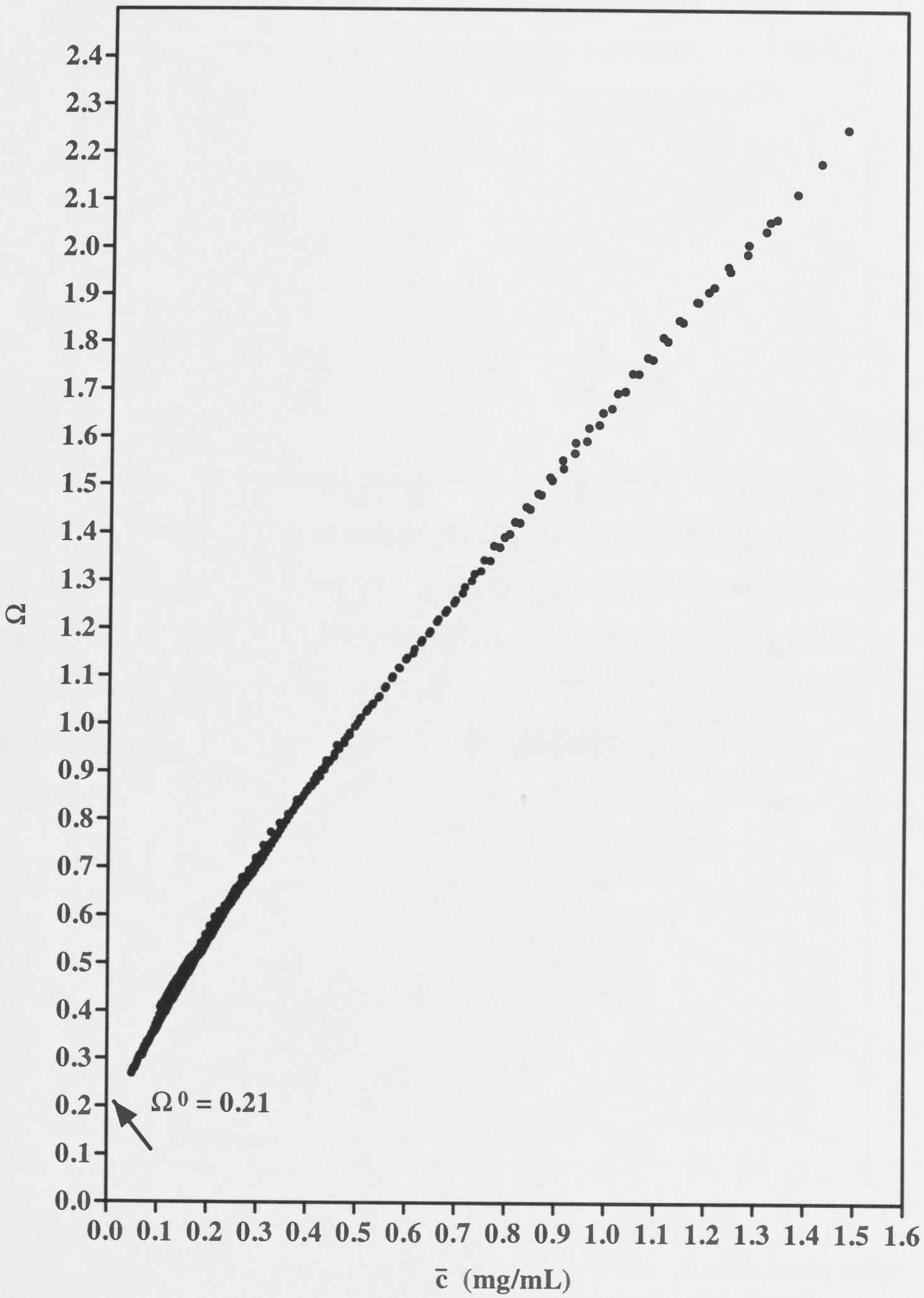

Ω
2.4
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Ω 0 = 0.21
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6
c̄ (mg/mL)

Figure 5.5

An illustrative plot of the omega-method used to analyse the sedimentation equilibrium results obtained with zinc-free chemically synthesised platypus insulin. All the experimental data sets listed for platypus insulin in Table 5.2 were used in the analysis. The ordinate values were obtained using equation 5.11 utilising the common reference concentration of 0.5 g/L, while the abscissa values refer to the corresponding total concentrations for which omega values were determined.

Figure 5.5

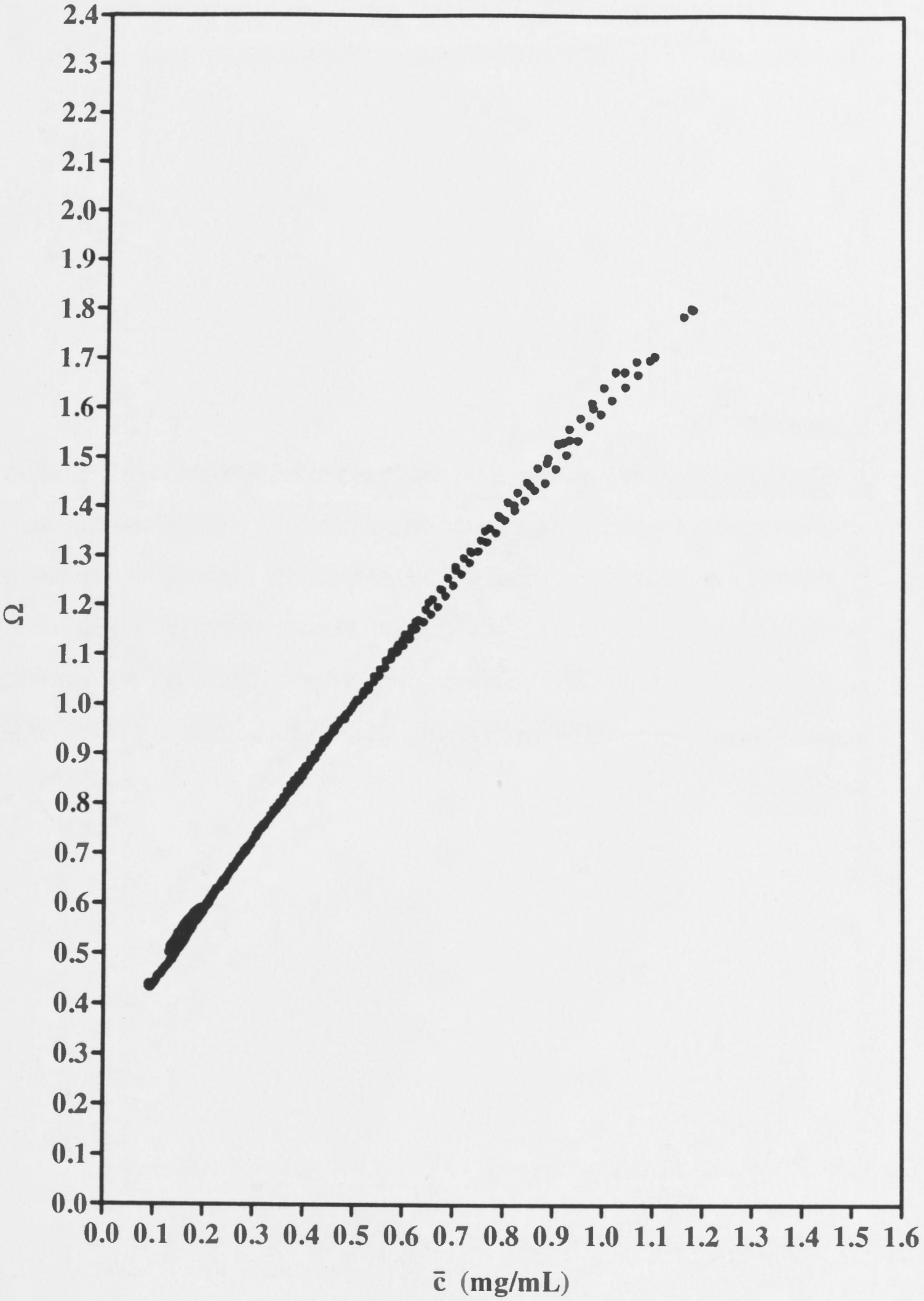
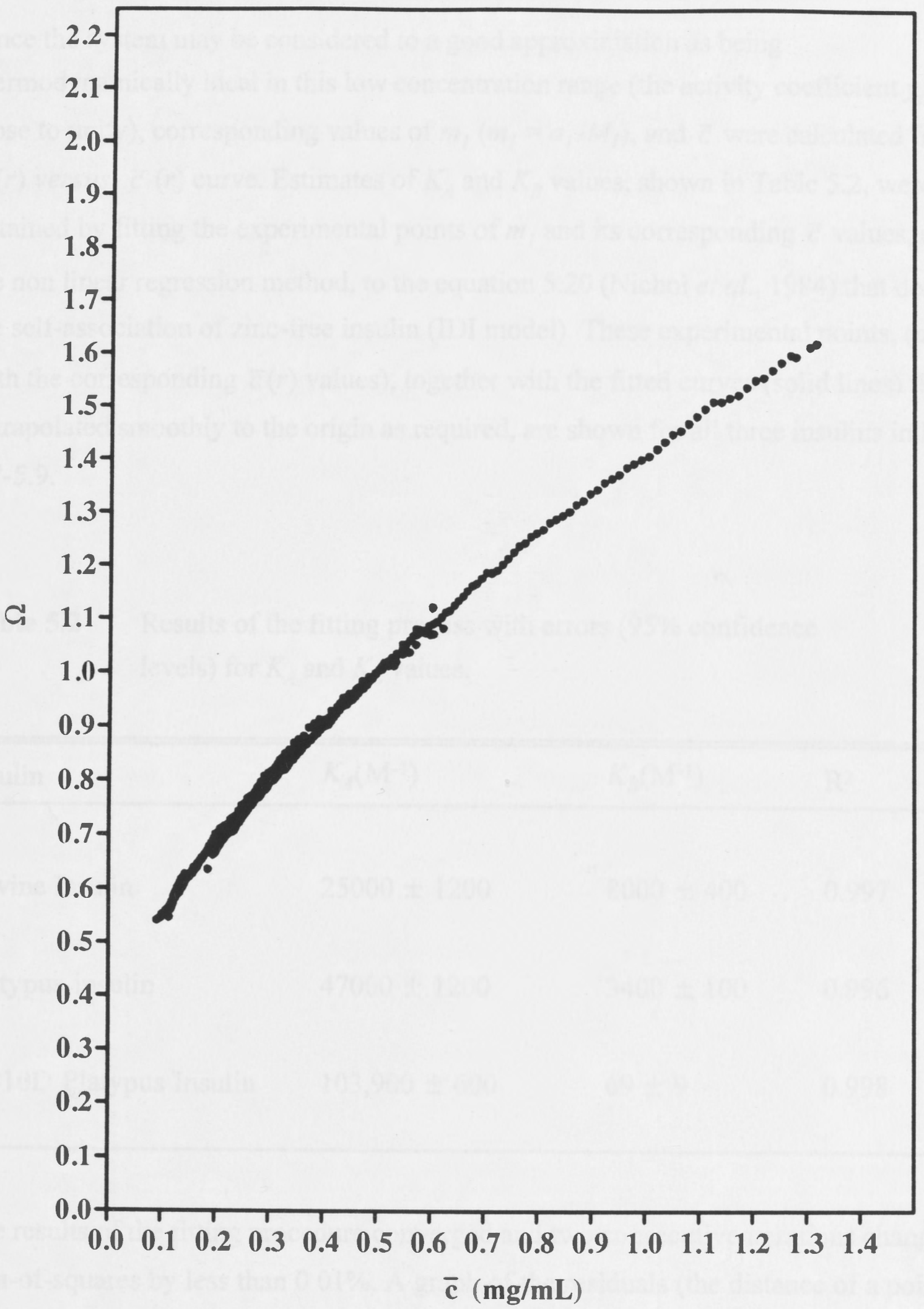

2.4
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
c
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6
c̄ (mg/mL)

Figure 5.6

An illustrative plot of the omega-method used to analyse the sedimentation equilibrium results obtained with zinc-free chemically synthesised HB10D platypus insulin. All the experimental data sets listed for HB10D platypus insulin in Table 5.2 were used in the analysis. The ordinate values were obtained using equation 5.11 utilising the common reference concentration of 0.5 g/L, while the abscissa values refer to the corresponding total concentrations for which omega values were determined.

Figure 5.6


2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
C
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4
c̄ (mg/mL)

Morris and Ralston (1985), it was found that the chosen Ω^0 value of 0.21 (reference concentration 0.5 g/L) gave the best fit of the experimental data to the IDI model equation for all three insulins.

Since the system may be considered to a good approximation as being thermodynamically ideal in this low concentration range (the activity coefficient y_i being close to unity), corresponding values of m_1 ($m_1 = a_1/M_1$), and $\bar{c}$ were calculated from the $a_1(r)$ *versus* $\bar{c}$ (r) curve. Estimates of K_A and K_B values, shown in Table 5.2, were obtained by fitting the experimental points of m_1 and its corresponding $\bar{c}$ values, using the non linear regression method, to the equation 5.20 (Nichol *et al.*, 1984) that describes the self-association of zinc-free insulin (IDI model). These experimental points, ($a_1(r)$ with the corresponding $\bar{c}(r)$ values), together with the fitted curves (solid lines) that extrapolated smoothly to the origin as required, are shown for all three insulins in Figures 5.7-5.9.

Table 5.2 Results of the fitting process with errors (95% confidence levels) for K_A and K_B values.

Insulin	$K_A(M^{-1})$	$K_B(M^{-1})$	R^2
Bovine Insulin	25000 ± 1200	8000 ± 400	0.997
Platypus insulin	47000 ± 1200	3400 ± 100	0.996
HB10D Platypus Insulin	103,900 ± 600	69 ± 9	0.998

The results of the fitting procedure converged and two consecutive iterations changed the sum-of-squares by less than 0.01%. A graph of the residuals (the distance of a point from the curve) in all cases showed that the deviation of the experimental data points

Figure 5.7

A plot of the thermodynamic activity of monomer, a_1 *versus* the total weight concentration from the sedimentation equilibrium experiments conducted with zinc-free bovine insulin. Equation 5.15 was used to calculate the experimental points from the plot shown in Figure 5.4. The solid line shows the final result of the fitting procedure of experimental m_1 values with non-linear regression to equation 5.20 (describing the IDI model of self-association) and a_1 values were back calculated with $m_1 = a_1/M_1$. The solid curve was calculated using equation 5.20 with estimates values $K_A = 25000\ M^{-1}$ and $K_B = 8000\ M^{-1}$ and shows the final result of the curve fitting procedure used to obtain these estimates.

Figure 5.7

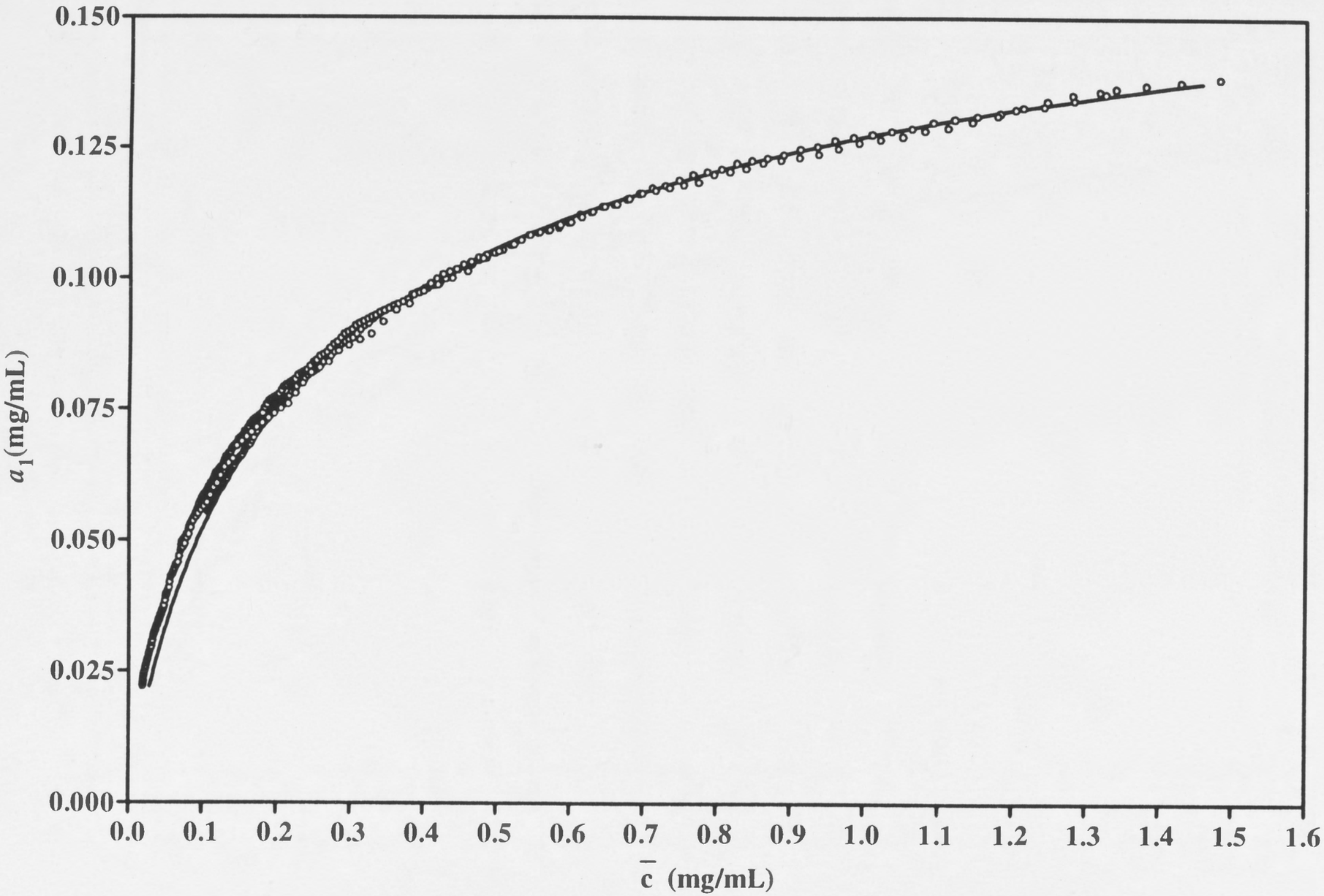

0.150
0.125
0.100
0.075
0.050
0.025
0.000
a_1(mg/mL)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
$\bar{c}$ (mg/mL)

Figure 5.8

A plot of the thermodynamic activity of monomer, a_1 *versus* the total weight concentration from the sedimentation equilibrium experiments conducted with zinc-free chemically synthesised platypus. Equation 5.15 was used to calculate the experimental points from the plot shown in Figure 5.4. The solid line shows the final result of the fitting procedure of experimental m_1 values with non-linear regression to equation 5.20 (describing the IDI model of self-association) and a_1 values were back calculated with $m_1 = a_1/M_1$. The solid curve was calculated using equation 5.20 with estimates values $K_A = 47000\ M^{-1}$ and $K_B = 3400\ M^{-1}$ and shows the final result of the curve fitting procedure used to obtain these estimates.

Figure 5.8

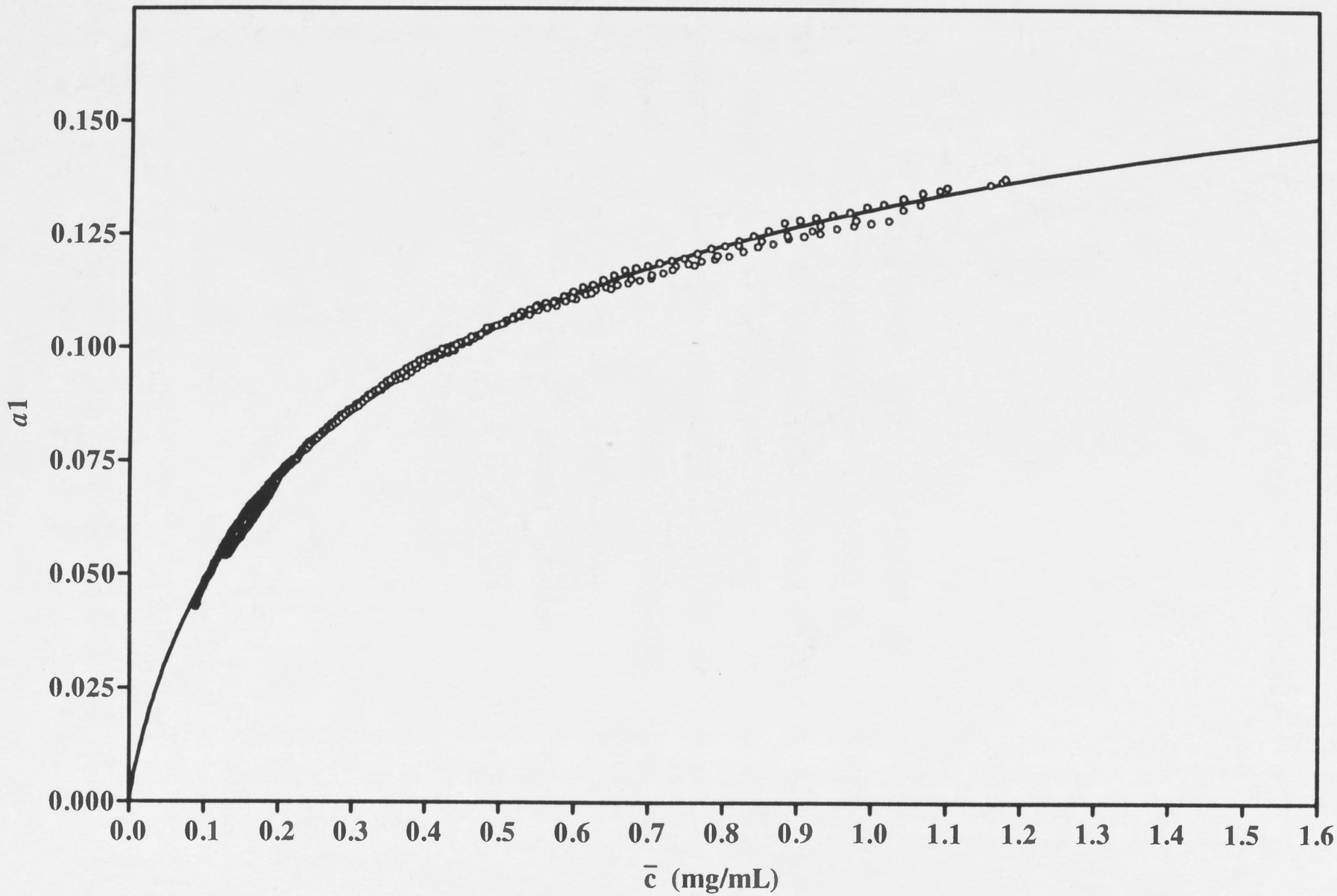
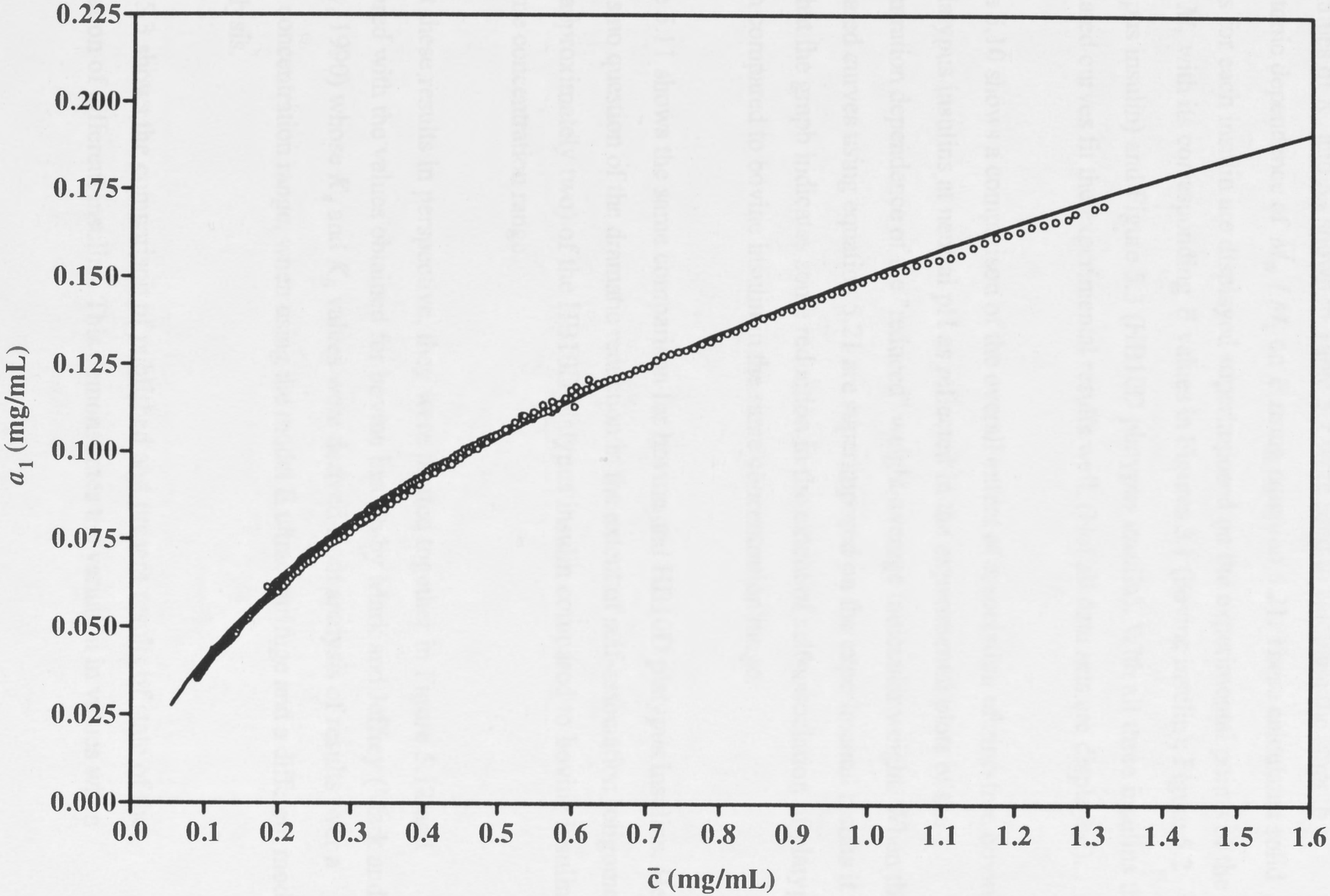

0.150
0.125
0.100
0.075
0.050
0.025
0.000
a1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
c̄ (mg/mL)

Figure 5.9

A plot of the thermodynamic activity of monomer, a_1 *versus* the total weight concentration from the sedimentation equilibrium experiments conducted with zinc-free chemically synthesised HB10D platypus. Equation 5.15 was used to calculate the experimental points from the plot shown in Figure 5.4. The solid line shows the final result of the fitting procedure of experimental m_1 values with non-linear regression to equation 5.20 (describing the IDI model of self-association) and a_1 values were back calculated with $m_1 = a_1/M_1$. The solid curve was calculated using equation 5.20 with estimates values $K_A = 103{,}900\ M^{-1}$ and $K_B = 69\ M^{-1}$ and shows the final result of the curve fitting procedure used to obtain these estimates.

Figure 5.9


0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
a_1 (mg/mL)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
1.6
$\bar{c}$ (mg/mL)

from the the best fit curve were within experimental error. The R^2 values, a measurement of goodness of fit, were all greater than 0.99 in all cases (Table 5.2); indicating excellent fits within experimental precision.

The values of K_A and K_B shown in Table 5.2 were used to calculate the smooth monotonic dependence of $\overline{M}_W / M_1$ on $\bar{c}$ using equation 5.21. These calculated solid curves for each insulin are displayed superimposed on the experimental points of the $\overline{M}_W / M_1$ with its corresponding $\bar{c}$ values in Figures 5.1 (bovine insulin); Figure 5.2 (platypus insulin) and Figure 5.3 (HB10D platypus insulin). With all three insulins the calculated curves fit the experimental results well. (Not all data sets are displayed).

Figure 5.10 shows a comparison of the overall extent of association of zinc-free bovine and platypus insulins at neutral pH as reflected in the experimental plots of the concentration dependence of the "reduced" weight-average molecular weight. When the calculated curves using equation 5.21 are superimposed on the experimental results it is seen that the graph indicates some reduction in the extent of self-association of platypus insulin compared to bovine insulin in the same concentration range.

Figure 5.11 shows the same comparison for bovine and HB10D platypus insulins. Here, there is no question of the dramatic reduction in the extent of self-association (oligomeric state, approximately two) of the HB10D platypus insulin compared to bovine insulin in the same concentration range.

To put these results in perspective, they were plotted together in Figure 5.12 and compared with the values obtained for bovine insulin by Mark and Jeffrey (Mark and Jeffrey, 1990) whose K_A and K_B values were derived from analysis of results over a bigger concentration range, when using the model E ultracentrifuge and a different mode of analysis.

Table 5.3 shows the comparison of published and present results of state of the association of different insulins. This demonstrates the variation in values when

Figure 5.10

A comparison of the overall extent of association of zinc-free bovine insulin and chemically synthesised platypus insulin reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets for bovine insulin mbin3a1 (o) and mbin3a2 (•) and from platypus insulin mplin3a1 (Δ) and mplin5a1($\blacktriangle$) (Table 5.1) are displayed. The K_A and K_B values, shown in Table 5.2, were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\bar{c}$ using equation 5.21. The calculated solid curves of bovine and platypus insulin, indicated on the graph, are displayed superimposed on the experimental points.

Figure 5.10

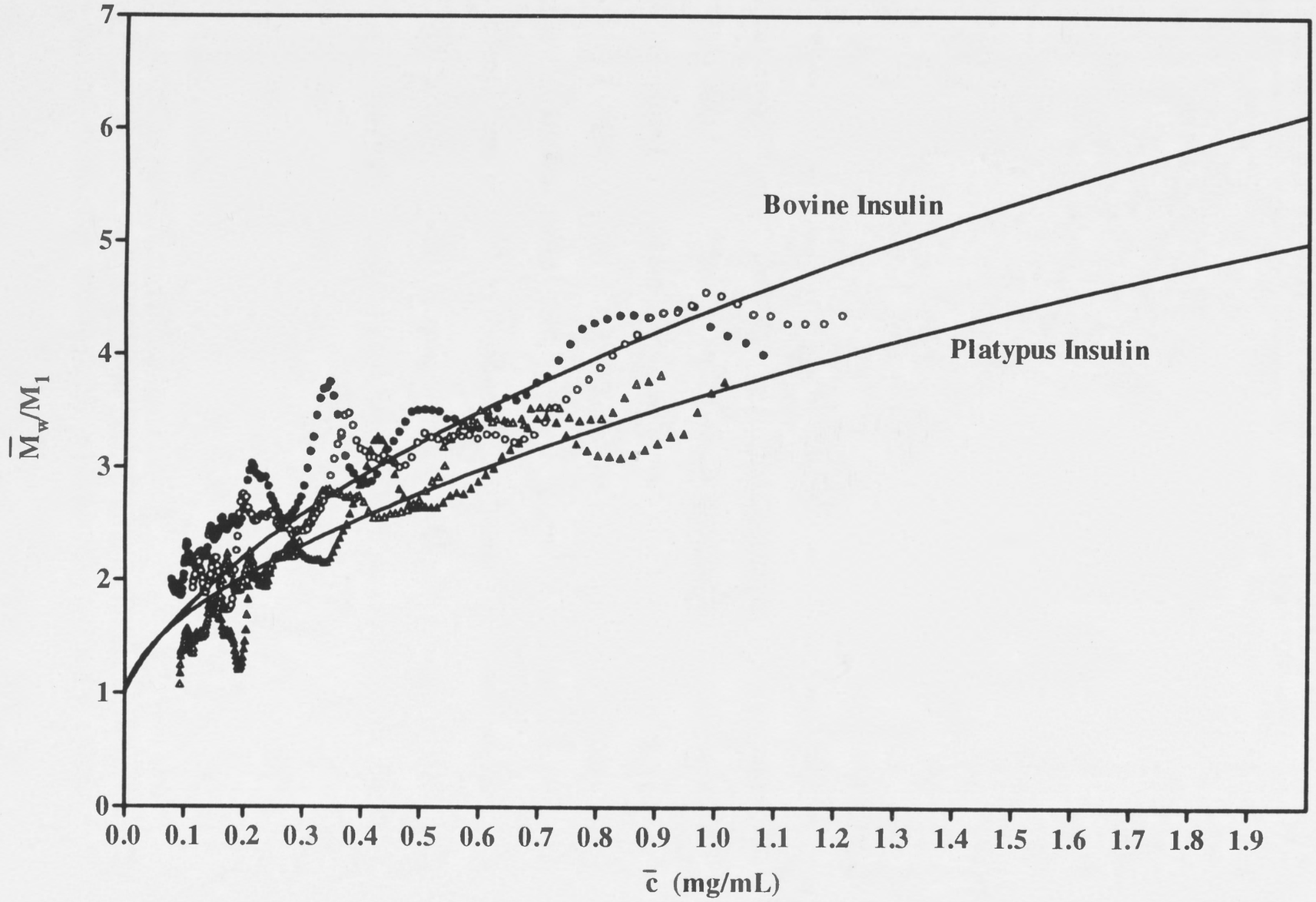

Bovine Insulin
Platypus Insulin
$\bar{M}_w/M_1$
$\bar{c}$ (mg/mL)
0
1
2
3
4
5
6
7
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9

Figure 5.11

A comparison of the overall extent of association of zinc-free bovine insulin and chemically synthesised HB10D platypus insulin reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets for bovine insulin mbin3a1 (o) and mbin3a2 (•) and for HB10D platypus insulin mdoa1a3 (■) and mdoa2a3 (❑) (Table 5.1) are displayed respectively. The K_A and K_B values, shown in Table 5.2, were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\overline{c}$ using equation 5.21. The calculated solid curves of bovine and platypus insulin, indicated on the graph, are displayed superimposed on the experimental points.

Figure 5.11

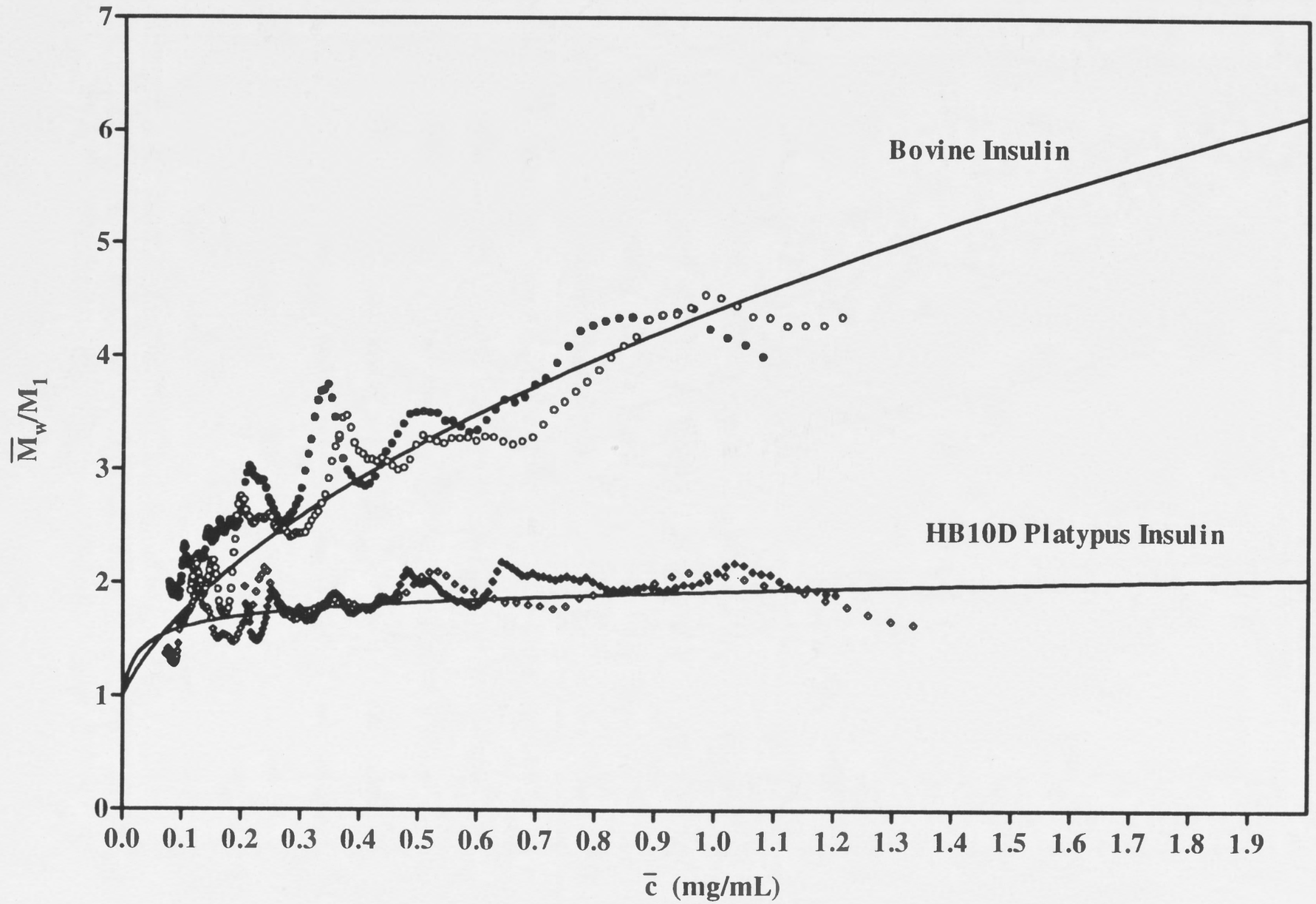

Bovine Insulin
HB10D Platypus Insulin
$\overline{M}_w/M_1$
$\overline{c}$ (mg/mL)
0
1
2
3
4
5
6
7
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9

Figure 5.12

A comparison of the overall extent of association of zinc-free bovine, chemically synthesised platypus and HB10D platypus insulin and bovine insulin from work by Mark *et al.* (Mark *et al.*, 1987) as reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight. The experimental data sets for platypus insulin mplin5a1 (o) and 3a1 (❑) (Table 5.1) are displayed. The K_A and K_B values, shown in Table 5.2 for the three insulins in this study, and values obtained by Mark *et al.* (Mark *et al.*, 1987) for bovine insulin were used to calculate the smooth monotonic dependence of $\overline{M}_w / M_1$ on $\overline{c}$ using equation 5.21. The calculated solid curves of the three insulins and of the bovine insulin in this study (AEM Bovine Insulin) (Mark *et al.*, 1987) shown on the graph, are displayed superimposed on the experimental points of platypus insulin.

Figure 5.12

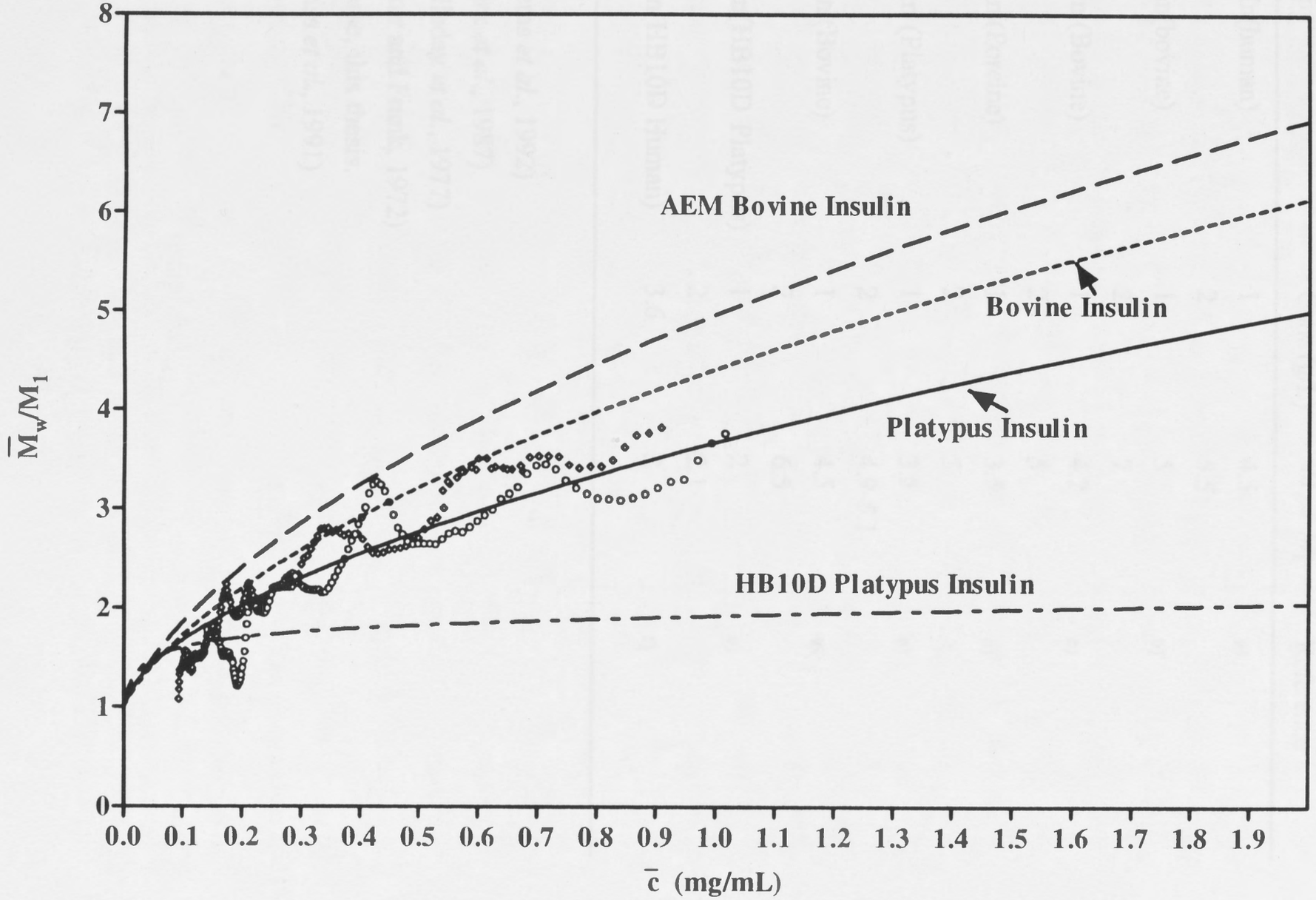

AEM Bovine Insulin
Bovine Insulin
Platypus Insulin
HB10D Platypus Insulin
$\bar{M}_w/M_1$
$\bar{c}$ (mg/mL)
0
1
2
3
4
5
6
7
8
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9

Table 5.3: The comparison of the state of association $\overline{M}_w / M_1$ at three different total concentrations (1 g/L, 2 g/L and 3.6 g/L) of the different insulins in solution, neutral pH and zinc-free. Published and own values were used.

Insulin	Conc (g/L)	$\overline{M}_w / M_1$	Reference
Insulin(human)	1	4.5	a)
	2	5.5	
Insulin(bovine)	1	5	b)
	2	7	
Insulin (Bovine)	1	4.2	c)
	2	5	
Insulin (Porcine)	1	3.5	d)
	2	5	
Insulin (Platypus)	1	3.5	e)
	2	4.9-5.1	
Insulin (Bovine)	1	4.5	e)
	2	6.5	
Insulin(HB10D Platypus)	1	2	e)
	2	2.1	
Insulin(HB10D Human)	3.6	2	f)

a) (Brems *et al.*, 1992)

b) (Mark *et al.*, 1987)

c) (Holladay *et al.*, 1977)

d) (Pekar and Frank, 1972)

e) Nourse, this thesis.

f) (Weiss *et al.*, 1991)

different methods of sedimentation equilibrium data collection are used. The solution conditions also vary slightly. The significance of these results is discussed later.

5.4.3 Evaluation of the weight-fractions of all species up to and including hexamer of bovine, platypus, and HB10D platypus insulin as a function of total concentration.

According to the IDI model of self-association of insulin in the concentration range 0-2g/L (0 - 0.33 mM) in a zinc-free neutral pH solution, bovine insulin (and human insulin) exists as a complex distribution of monomers, dimers, trimers, tetramers, pentamers, hexamers and higher order polymers. This relative distribution is dependent on the total concentration of insulin and can be analysed in terms of the weight-fraction of a particular species as a function of total concentration.

The values for weight-fractions of all species up to and including hexamer ($i = 6$) present in solutions of zinc-free insulin (pH 7.0; $I = 0.1$ M; $T = 25$ °C) as a function of total concentration were calculated, using equation 5.23, and are displayed graphically in Figures 5.13 - 5.15. Values for K_A and K_B used in the calculations are from Table 5.2.

The graphs of both bovine insulin (Figure 5.13) and platypus insulin (Figure 5.14) illustrate the presence of monomers, α- and β-dimers, trimers, α- and β-tetramers, pentamers and hexamers in the specified concentration range, albeit in different amounts as related by the difference in K_A and K_B values. From the weight distribution of all species of HB10D platypus insulin as a function of total concentration (Figure 5.15), it can be noted that under the conditions examined and in the concentration range of 0 - 2 g/L, only monomers, β-dimers and β-tetramers (in very small amounts) exists in solution.

Figure 5.13

The weight-fractions (ϕ_i) of all species up to and including ($i = 6$) present in a solution of zinc-free bovine insulin as a function of total concentration is shown. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled α-dimer and α-tetramer refer to the species of dimer and tetramer with two exposed α-faces whereas the lines labelled β-dimer and β-tetramer refer to the species of dimer and tetramer with two exposed β-faces.

Figure 5.13

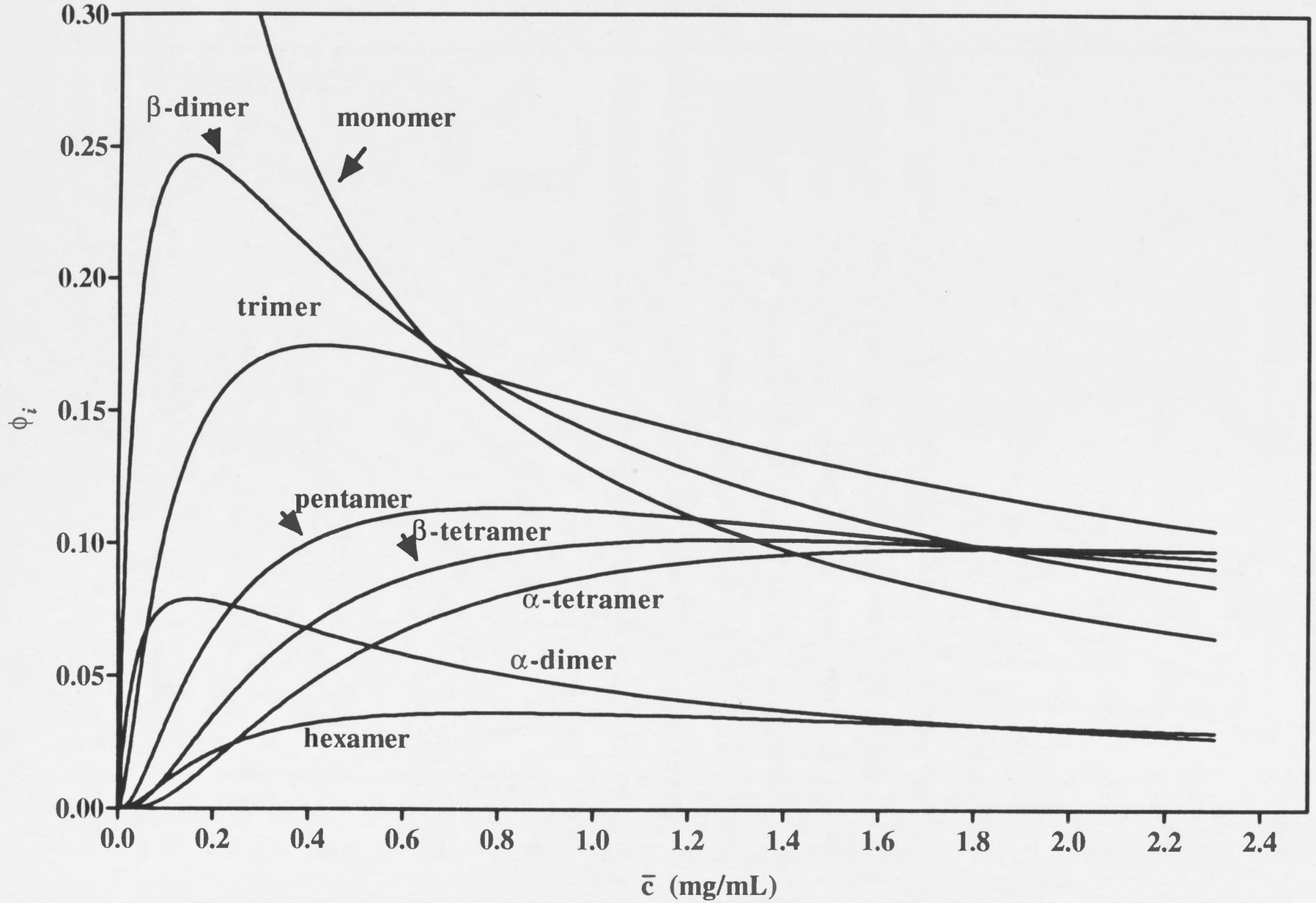

0.30
0.25
0.20
0.15
0.10
0.05
0.00
ϕ_i
β-dimer
monomer
trimer
pentamer
β-tetramer
α-tetramer
α-dimer
hexamer
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
$\bar{c}$ (mg/mL)

Figure 5.14

The weight-fractions (ϕ_i) of all species up to and including ($i = 6$) present in a solution of zinc-free chemically synthesised platypus insulin as a function of total concentration is shown. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled α-dimer and α-tetramer refer to the species of dimer and tetramer with two exposed α-faces whereas the lines labelled β-dimer and β-tetramer refer to the species of dimer and tetramer with two exposed β-faces.

Figure 5.14

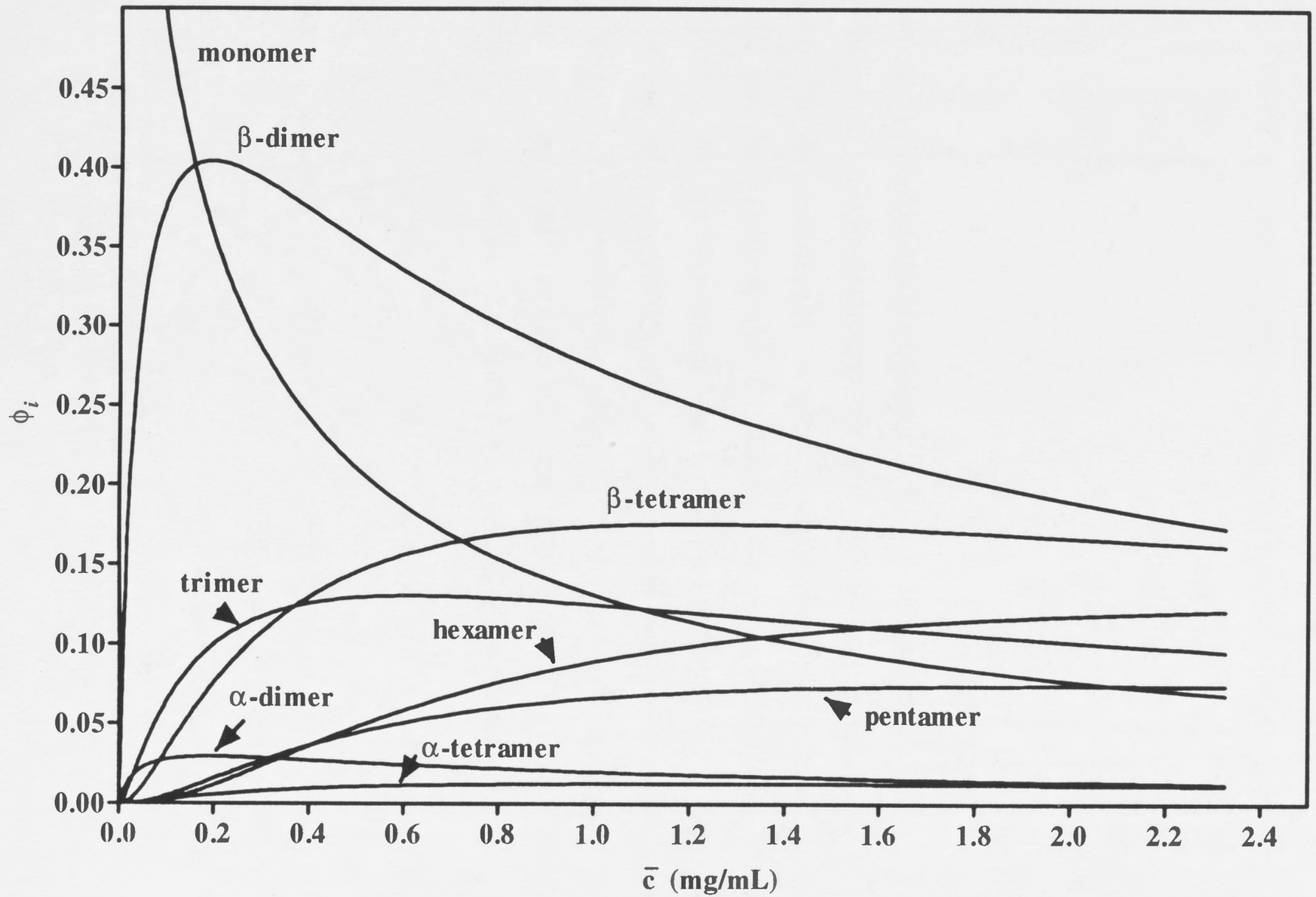
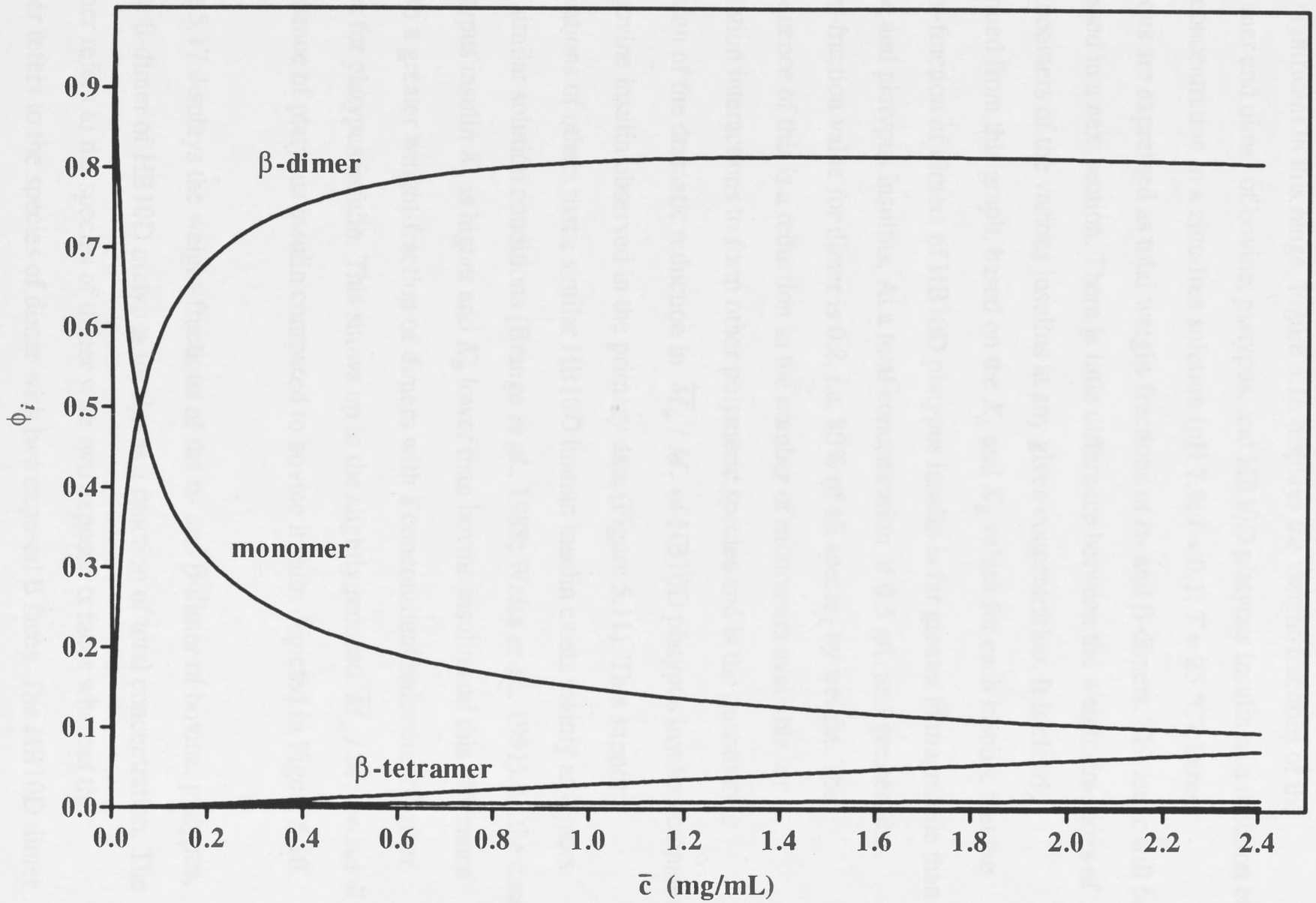

monomer
β-dimer
β-tetramer
trimer
hexamer
α-dimer
pentamer
α-tetramer
ϕ_i
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
$\bar{c}$ (mg/mL)

Figure 5.15

The weight-fractions (ϕ_i) of all species up to and including ($i = 6$) present in a solution of zinc-free chemically synthesised HB10D platypus insulin as a function of total concentration is shown. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled β-dimer and β-tetramer refer to the species of dimer and tetramer with two exposed β-faces. (No other oligomers than those depicted were present at significant weight-fractions in the concentration range shown). No α-dimer, trimers, α-tetramers, pentamers and hexamers than those depicted were present at significant weight-fractions in the concentration range shown.

Figure 5.15


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
ϕ_i
β-dimer
monomer
β-tetramer
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
$\bar{c}$ (mg/mL)

5.4.4 Comparison of the weight-fractions of monomers and oligomers of the three insulins

The concentration range 0-2 g/L of insulin samples was used in the analysis because the changes in weight-fractions of the different polymeric (oligomeric) species are most significant in this range. Figure 5.16 displays the weight-fractions of the monomer and dimer of bovine, platypus, and HB10D platypus insulin as a function of total concentration in a zinc-free solution (pH 7.0; $I = 0.1$; $T = 25$ °C). Dimer fractions are expressed as total weight-fractions of α- and β-dimers. The latter will be discussed in a next section. There is little difference between the weight-fractions of the monomers of the various insulins at any given concentration. It is clearly illustrated from this graph, based on the K_A and K_B values for each insulin, that the weight-fraction of dimers of HB10D platypus insulin is far greater in magnitude than bovine and platypus insulins. At a total concentration of 0.5 g/L and greater, the weight-fraction value for dimer is 0.8, i.e. 80% of all species by weight. The consequence of this is a reduction in the number of monomers available for association interactions to form other polymeric species and is the quantitative reflection of the dramatic reduction in $\overline{M}_w / M_1$ of HB10D platypus insulin compared with bovine insulin observed in the primary data (Figure 5.11). This supports observations of others that a similar HB10D human insulin exists mainly as dimers under similar solution conditions (Brange *et al.*, 1988; Weiss *et al.*, 1991). In the case of platypus insulin K_A is higher and K_B lower than bovine insulin and this difference leads to a greater weight-fraction of dimers with a concomitant reduction of other species for platypus insulin. This shows up in the slightly reduced $\overline{M}_w / M_1$ *versus* $\overline{c}$ dependence of platypus insulin compared to bovine insulin depicted in Figure 5.10.

Figure 5.17 displays the weight-fractions of the α- and β-dimer of bovine, platypus, and the β-dimer of HB10D platypus insulin as a function of total concentration. The α–dimer refers to the species of dimer with two exposed α faces whereas the β-dimer refers to the species of dimer with two exposed β faces. The HB10D dimer exists exclusively as a β-dimer in this concentration range. Compared to other

Figure 5.16

A comparison of the weight-fractions (ϕ_i) of monomers and dimers in solutions of zinc-free bovine, chemically synthesised platypus and HB10D platypus insulins as a function of total concentration. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled dimer include both α- and β-dimers. Plin = platypus and Bin = bovine.

Figure 5.16

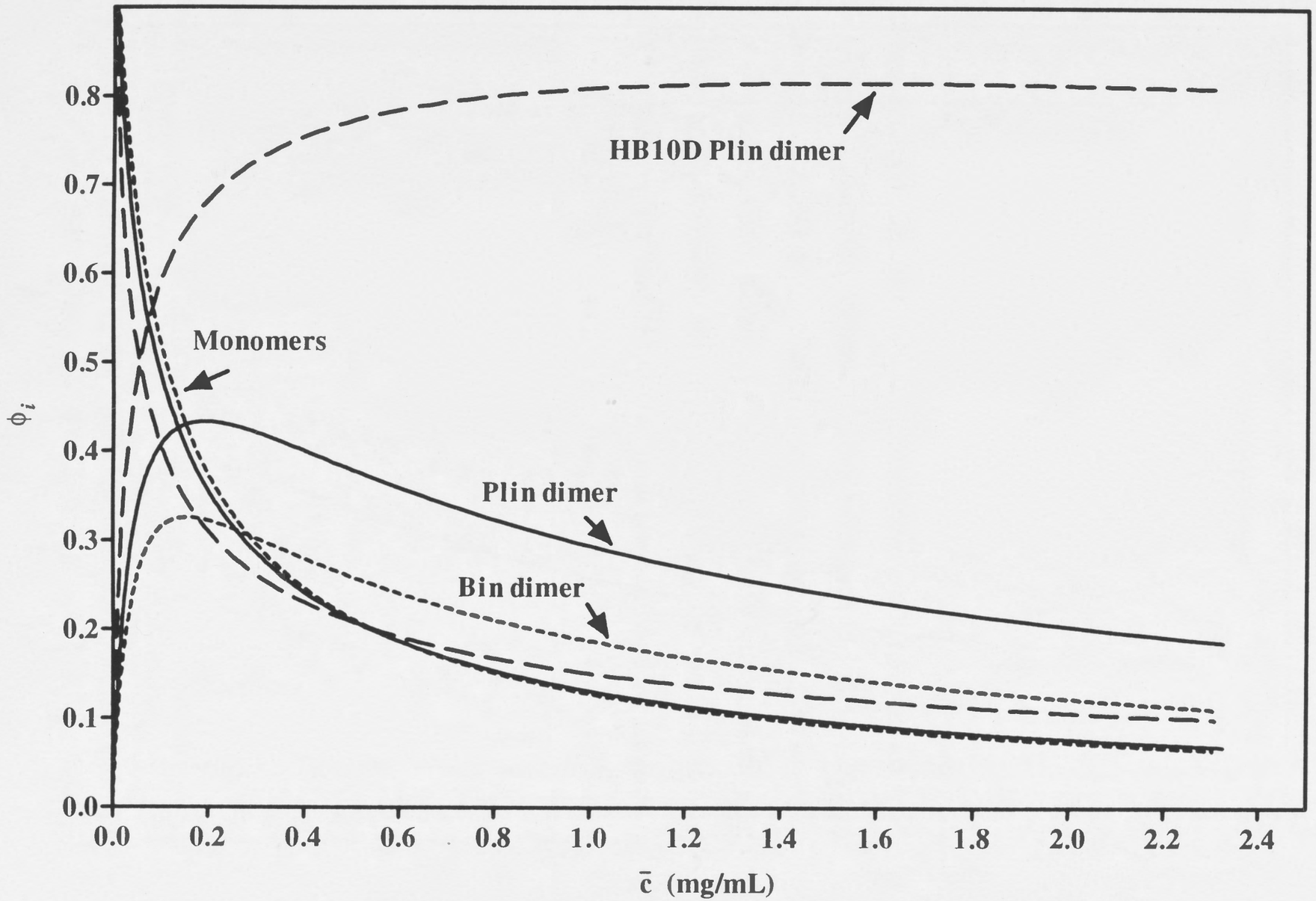
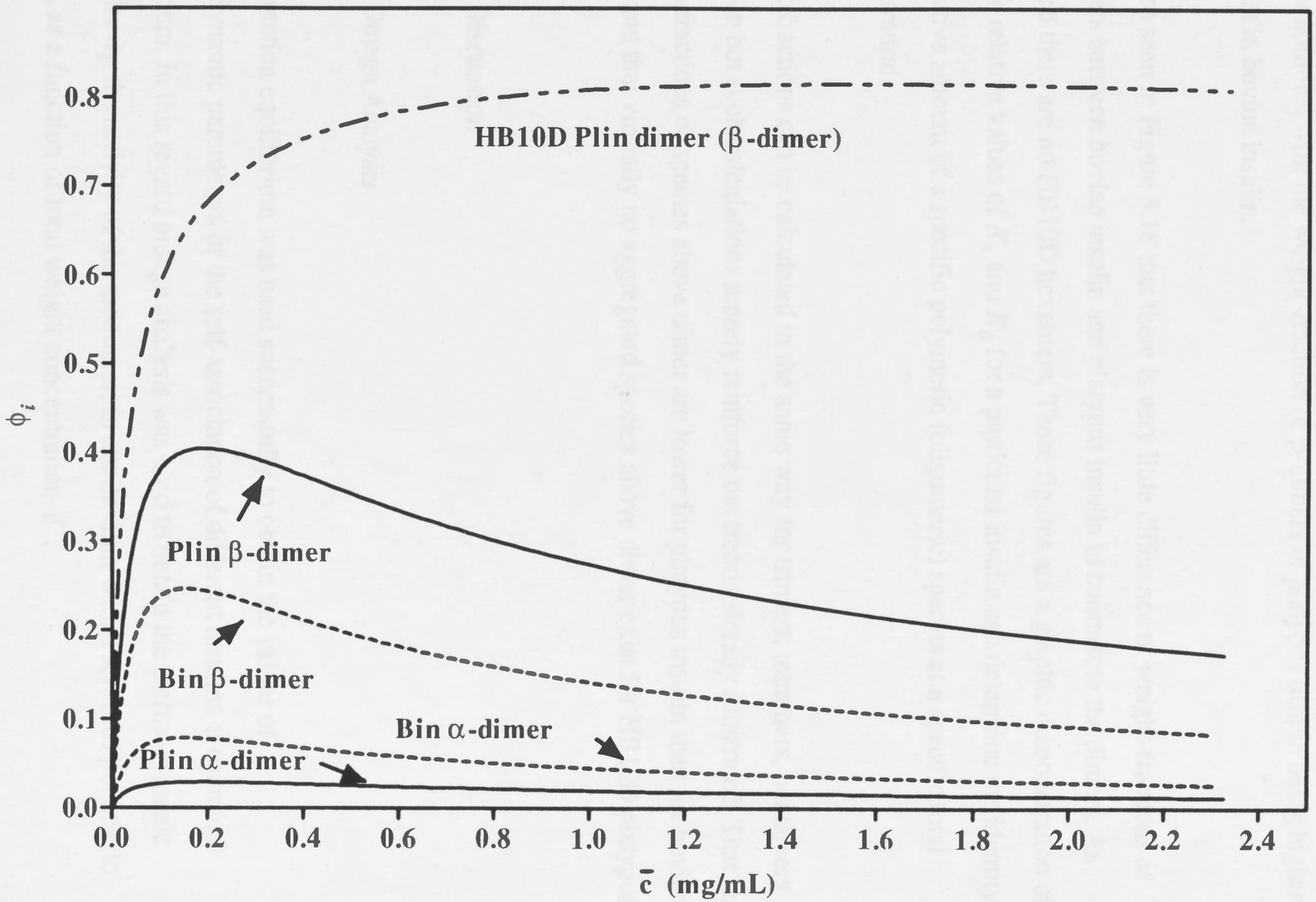

HB10D Plin dimer
Monomers
Plin dimer
Bin dimer
ϕ_i
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
$\bar{c}$ (mg/mL)

Figure 5.17

A comparison of the weight-fractions (ϕ_i) of α- and β-dimers in solutions of zinc-free free bovine, chemically synthesised platypus and HB10D platypus insulins as a function of total concentration. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled α-dimer refer to the species of dimer with two exposed α-faces whereas the lines labelled β-dimer refer to the species of dimer with two exposed β-faces. Plin = platypus and Bin = bovine.

Figure 5.17


HB10D Plin dimer (β-dimer)
Plin β-dimer
Bin β-dimer
Bin α-dimer
Plin α-dimer
ϕ_i
$\bar{c}$ (mg/mL)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4

insulins, the K_A value is huge (indicating a strong α-α interaction) and the K_B value very low, an indication that the β-β interaction has been virtually destroyed.

Comparing the weight-fractions of the α- and β-dimers of bovine and platypus insulin it can be seen as expected that the β-dimers are the dominant dimer in solution in these conditions, with the weight-fraction of β-dimers of platypus insulin being higher than that in bovine insulin.

It can be seen in Figure 5.18 that there is very little difference of weight-fractions of hexamers between bovine insulin and platypus insulin in contrast to the dimers. As expected there are no HB10D hexamers. These figures are a graphic demonstration of how the relative values of K_A and K_B for a particular insulin can determine the identity and relative amount of a specific polymeric (oligomeric) species at a specific total concentration.

Weight-fractions can be calculated in the same way for trimers, tetramers, pentamers and so on but such calculations merely reinforce the trends already illustrated. That is, that the fractions of species above dimer are lower for platypus insulin than for bovine insulin and that virtually no aggregated species above dimer exist for HB10D platypus insulin.

5.5 Discussion

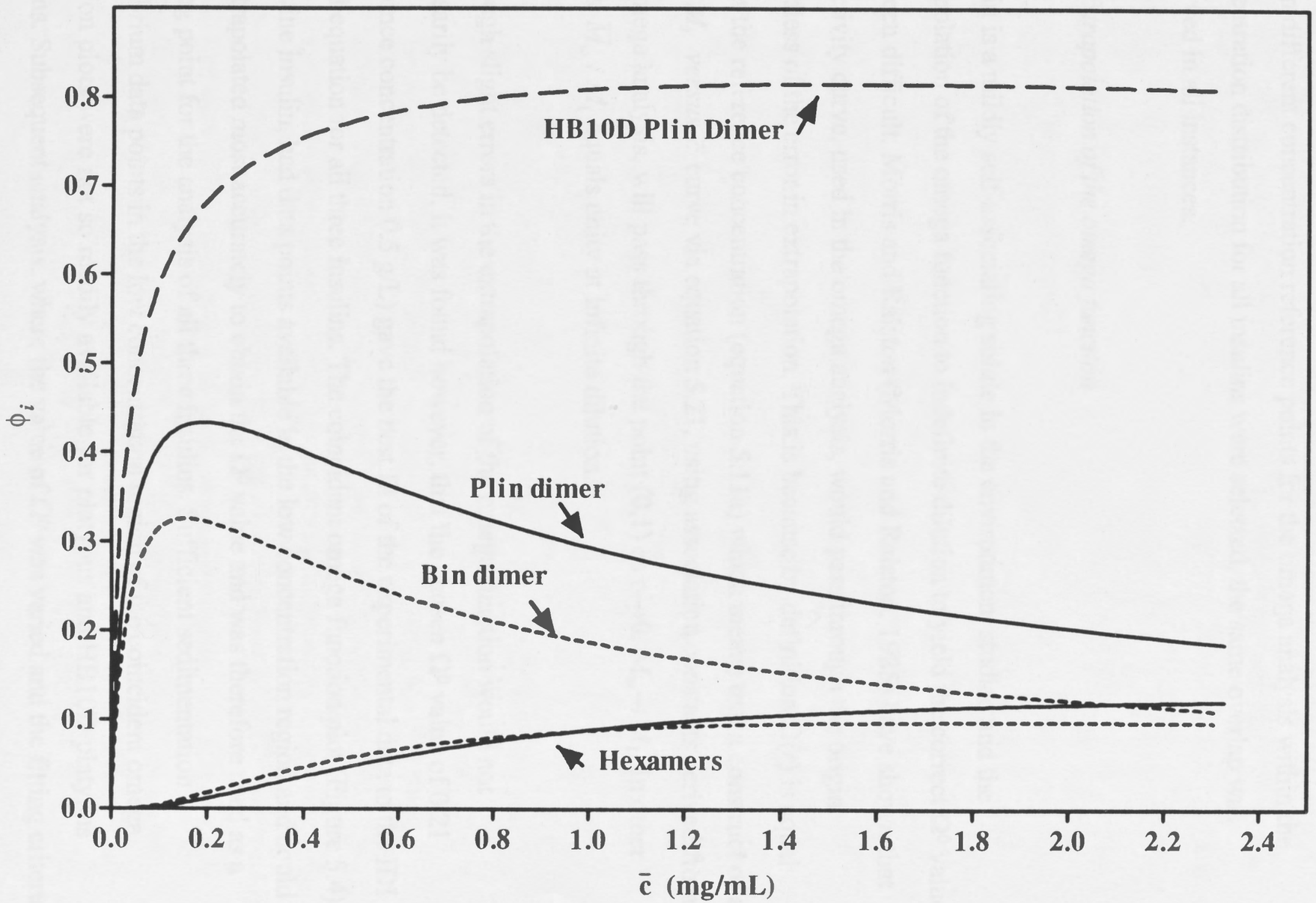
5.5.1 Omega Analysis

Sedimentation equilibrium was used successfully to obtain the values of thermodynamic parameters of the self-association of different insulins at chemical equilibrium. In this regard omega analysis was used to obtain the thermodynamic parameter a_1, the activity of the monomer, in equilibrium with a series of polymeric species, as a function of total weight concentration $\bar{c}$.

Figure 5.18

A comparison of the weight-fractions (ϕ_i) of dimers solutions of zinc-free bovine, chemically synthesised platypus and HB10D platypus insulins as a function of total concentration. The values for the weight-fractions were calculated using equation 5.23 assuming values for K_A and K_B from Table 5.2. The lines labelled dimer include both α- and β-dimers. Plin = platypus and Bin = bovine.

Figure 5.18


HB10D Plin Dimer
Plin dimer
Bin dimer
Hexamers
ϕ_i
$\bar{c}$ (mg/mL)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4

The coincidence of the omega plots

The overlap of omega function curves (Figure 5.1-3) for the different insulins shows that chemical equilibrium had been achieved and in all cases no significant concentration of solute was present that did not participate in the self-association. When different concentration reference points for the omega analysis within the concentration distribution for all insulins were selected, the same overlap was observed in all instances.

The extrapolation of the omega function

Insulin is a mildly self-associating solute in the environment studied and the extrapolation of the omega function to indefinite dilution to yield the correct Ω^0 value has been difficult. Morris and Ralston (Morris and Ralston, 1985) have shown that the activity curve, used in the omega analysis, would pass through the origin regardless of the error in extrapolation. This is because by definition $\Omega(r)$ is equal to 1 at the reference concentration (equation 5.11a) which means that a construct of a $\overline{M}_w / M_1$ *versus* $\bar{c}$ curve via equation 5.21, using association constants derived from the omega analysis, will pass through the point (0,1) as $c \rightarrow 0$, $M_w \rightarrow M_1$. In other words $\overline{M}_w / M_1$ equals unity at infinite dilution.

Although slight errors in the extrapolation of the omega function would not necessarily be detected, it was found however, that the chosen Ω^0 value of 0.21 (reference concentration 0.5 g/L) gave the best fit of the experimental data to the IDI model equation for all three insulins. The coincident omega function plot (Figure 5.4) of bovine insulin, had data points available in the low concentration region and could be extrapolated most accurately to obtain the Ω^0 value and was therefore used as a starting point for the analysis of all three insulins. Sufficient sedimentation equilibrium data points in the low concentration region of the coincident omega function plots were not so readily available for platypus and HB10D platypus insulins. Subsequent analysis, where the value of Ω^0 was varied and the fitting criteria applied, also showed that this was the best value that could be found for these two insulins.

The choice of the reference concentration

The value of $\bar{c}(r_F)$, the reference concentration value, chosen to compute $\Omega(r)$ values as a function of $\bar{c}(r)$ may be selected as any value in the observed total concentration range. Irrespective of a possible error in the choice of the reference concentration, a good extrapolation to indefinite dilution of the omega function curve (omega $\Omega(r)$ *versus* $\bar{c}(r)$) will yield the correct value of the reference monomer activity at the appropriate radial position r (Morris and Ralston, 1985; Milthorpe *et al.*, 1975). In this study different reference concentrations, within the concentration distribution of each insulin, were used in the omega analysis and subsequent fitting procedure of the experimental data as described in the preceding sections. The results of this analysis for each reference concentration were compared for the different insulins respectively and found to give similar results in terms of calculated association constants (not shown).

Thermodynamic non-ideality effects

The effects of non-ideality in the present system are not large due to the relatively small size of, and net charge borne by, the insulin protein and because a restricted range of total concentration (0-1.5 g/L) was examined. At higher concentrations, however, the system can exhibit non-ideality (from excessive crowding or charge effects) and can obscure the existence of higher association states. It should be noted that the determination of the thermodynamic activity of monomer is independent of any assumption concerning the nature of the self-association pattern. There is, however, nothing to gain in the low concentration range considered here by the non-ideality correction; a decision based on the results of Mark *et al.* (Mark *et al.*, 1987) when taking non-ideality effects into account over a wide concentration range 0.01 - 3.3 g/L for the calculation of the K_A and K_B found that the values obtained differ little from those obtained neglecting non-ideality.

The allowance for non-ideality assumes greater significance in experimental environments where charge-charge interactions are increased in magnitude. The results (Mark *et al.*, 1987) have shown that in the environment under discussion the values of the activity coefficients y_i differ very little from unity, even for higher polymeric species.

5.5.2 The fitting of the m_1 *versus* $\bar{c}$, data to the IDI model

Fitting the experimental data to a mathematical equation describing the IDI model of self-association of insulin was excellent as was interpreted from the results of the non linear regression analysis, namely R^2 values, standard errors and 95% confidence intervals. A visual inspection of the fitted curve and experimental data (Figures 5.7-9) for all the different insulins support the goodness of fit parameters in all cases.

It was found that HB10D platypus insulin is dimeric in a zinc-free neutral pH environment and is similar to HB10D human insulin (Weiss *et al.*, 1991). The successful fitting of the sedimentation equilibrium data of this dimeric insulin analogue, HB10D platypus insulin, to the IDI model of insulin self-association demonstrates the extensive applicability of this model. In addition, the association constants derived for this dimeric insulin analogue make biological sense in terms of what is known about the structure of insulin (Baker *et al.*, 1988).

It can be concluded that the IDI model for the self-association of insulin is supported by data in the present study for a specified solution environment (zinc-free, neutral pH, $I = 0.1$, T = 25 °C) and that previously obtained by others.

The association constants K_A and K_B

For all insulins the association constants K_A and K_B calculated from the omega analysis and returned in the $(\overline{M}_w / M_1)$ *versus* $\bar{c}$ solid line curve, using equation 5.21, are in good agreement with the experimental data points (Figure 5.1-3) by visual inspection.

The K_A derived for bovine insulin in this study, 25000 M^{-1} , compares well with the value, 27400 M^{-1} , obtained by Mark and Jeffrey (Mark and Jeffrey, 1990) in their sedimentation equilibrium study (ideal treatment) under identical solution conditions. The value for K_B , 8000 M^{-1} , in this study is, however, slightly lower than the value of 11800 M^{-1} , derived by Mark and Jeffrey (Mark and Jeffrey, 1990). This may be due to thermodynamic non-ideality effects as they used a much wider concentration range (0.1 - 3.3 g/L), using the Rayleigh interference optical system than was used in the present study (0-1.5 g/L). This overall good agreement in two completely independent studies adds confidence in the results of the analysis of the platypus insulin samples and is also a good indication of the global error in the association constants resulting from this kind of analysis. It should be noted that neutral pH insulin does not dissolve beyond 4 g/L (Mark *et al.*, 1987).

5.5.3 The dependence of "reduced" weight-average molecular weight on total insulin concentration : The overall extent of the self-association of zinc-free insulin

The overall extent of the self-association of bovine, platypus and HB10D platypus insulin differ markedly as illustrated in Figures 5.7-5.9 respectively and is governed by the relative values of association K_A and K_B . In all instances self-association is present as is illustrated by the dependence of "reduced" weight-average molecular weight $(\overline{M}_w / M_1)$ of insulin on the total concentration. Even HB10D platypus insulin self-associates, although its highest association state is two in these solution conditions. In the following section the extent and the role the association constants play in the self-association process of all the insulins will be discussed.

Comparison of the self-association pattern of bovine and platypus insulin

From the comparison of the overall extent of association of zinc-free bovine and platypus insulins at neutral pH as reflected in plots of the concentration dependence of the "reduced" weight-average molecular weight ($\overline{M}_w / M_1$) *versus* $\overline{c}$ (Figure 5.10), it can be reasonably concluded that the extent of the self-association of platypus insulin is clearly less than that for bovine insulin. This reduction in self-association is not very large but it is considered to be just significant in the light of the estimation of the magnitude of the errors as assessed from the comparison of the two studies of bovine insulin referred to above. Assuming experimental conditions are identical and sample preparations homogeneous the difference in the extent of self-association can be attributed to the different amino-acid residues in the interaction regions (monomer-monomer interface and dimer-dimer interface) of monomeric insulin. The biological activity of the two insulins has been found to be identical (Chapter 4). The amino-acid residue changes, A5 Gln→Glu and B25 Phe→Tyr, in the receptor interaction region are not expected to change the self-association behaviour of insulin (Baker *et al.*, 1988).

The K_A value of platypus insulin (47000 M^{-1}) is nearly two fold higher in magnitude than that of bovine insulin (25000 M^{-1}). The changes in amino-acid residues from bovine insulin to platypus insulin in the monomer-monomer interface (α-α interaction region or OP interaction region in the crystal) are B25 Phe→Tyr and B27 Thr→Ile respectively. Comparing the properties of the residues (Table 3.1, Chapter 3) and according to previous studies the B25 Phe→Tyr is not expected to have an effect on the extent of monomer-monomer interaction in this region or its receptor binding capabilities. The B27 Thr→Ile substitution, however, introduces a more hydrophobic amino-acid residue in the monomer-monomer interface (α-α interaction region) and has the potential to increase the extent of this self-association as hydrophobic interactions are considered to be the dominant force that govern this interaction as shown by (Jeffrey and Coates, 1966a; Baker *et al.*, 1988). Mark *et al.* (Mark *et al.*,

1987) in their sedimentation equilibrium studies of the self-association of bovine insulin over a wide pH range and at different ionic strengths and temperatures at pH 7.0, have found that the value of K_A (which in all experimental environments is greater than K_B) increases with increased temperature and increasing ionic strength; both characteristics of interactions which are predominantly hydrophobic in nature (Kauzmann, 1959). The relative small dependence of K_A on pH (in comparison to K_B) at fixed ionic strength and temperature is consistent with the above finding for the nature of the α-α interaction region (Mark *et al.*, 1987).

The K_B value, representing the dimer-dimer interface (β-β interaction region or OQ interaction region in the crystal) according to the IDI model, of platypus insulin (3400 M^{-1}) is more than two-fold smaller in magnitude than that of bovine insulin (8000 M^{-1}). The changes in amino-acid residues from bovine to platypus insulin in the dimer-dimer interface are B2 Val→Pro and A13 Leu→Met . When comparing the properties of the residues (Table 3.1, Chapter 3) it is seen that neither substitution would be expected to alter the overall hydrophobic nature of this region. The substitution of B2 Val→Pro, however, has the potential of having an effect on the spatial arrangement of residues in its region as Pro is known to be a breaker of secondary structure (Fasman, 1989) (Chapter 3). These properties of Pro in the B2 position in platypus insulin could have an influence on the overall local arrangement of the amino-acid side chains in the dimer-dimer interface and thus translate into a reduced β-β interaction that is reflected in the sedimentation equilibrium experiments on the self-association of platypus insulin.

In addition, the hydrophobicity (the value of free energies of transfer) calculated in kcal $mole^{-1}$ for Pro is slightly lower than that for Val (Fasman, 1989). The A13 Leu→Met substitution introduces a sulfur containing amino-acid, slightly less hydrophobic than Leu (Fasman, 1989), into this region which in turn could effect the spatial arrangement of its neighbouring residues. Although the sulphur may have some H-bonding capability, the major role of methionine is to provide a very flexible hydrophobic side chain (Fasman, 1989). It was indeed found, when using molecular modelling (Chapter 3), that the spatial arrangement of residues A17, A13 and A14 in

the dimer-dimer interface is very different in platypus insulin. It is possible that this observation could translate into a change (reduction) in the extent of the self-interaction in the dimer-dimer interaction region and which indeed was found.

Comparison of the self-association pattern of bovine and HB10D platypus insulin

X-ray analysis of mammalian insulins suggests that the B10 His residue is important for the formation of zinc hexamers (Baker *et al.*, 1988). This residue also participates in the hexamer contact (so called dimer-dimer interface) of non-zinc hexamers and thus forms part of the residues involved in the β-β interaction of insulin (OQ interaction in the crystal structure). It must be pointed out however, that it is by no means certain that the groups involved in the OQ interaction in the crystal (Baker *et al.*, 1988) are necessarily the same ones involved in the associated states of zinc-free insulin in free solution (Mark *et al.*, 1987). Weiss *et al.* (Weiss *et al.*, 1991) with their 2D NMR studies on a monomeric DKP human insulin in solution, also list B10 His as involved in the dimer-dimer interface.

Studies of HB10D human insulin were originally motivated by the identification of Asp B10 proinsulin in association with *diabetes mellitus* and hyperproinsulinemia in man (Gruppuso *et al.*, 1984; Chan *et al.*, 1987). The mutation apparently impairs the regulated pathway of proinsulin processing, storage and release (Carroll *et al.*, 1988; Gross *et al.*, 1989). The corresponding HB10D human insulin analogue was synthesised and shown to exhibit enhanced bioactivity and receptor binding affinity (Schwartz *et al.*, 1987). Its self-association was explored by the methods of osmometry (Brange *et al.*, 1988) and sedimentation equilibrium (Weiss *et al.*, 1991) and found to be dimeric in the solution conditions and in the concentration range studied. The latter concluded from their 2D NMR studies on this analogue that it is structurally conservative, in other words no significant structural changes occur upon the substitution at the B10 position.

Hagfish insulin (HB10D) together with insulins from guinea pig (HB10N); coypu (HB10Q), degu (HB10N) and casiragua (HB10Q) that have a substitution in the B10

position do not form stable zinc hexamers (Burke *et al.*, 1984). In addition, these insulins are found to be much less active than the B10 His mammalian insulins which do form stable zinc hexamers (Emdin *et al.*, 1980; Peterson *et al.*, 1975). In contrast HB10D human insulin which also do not form stable zinc hexamers has a high biological activity. The difference in activity is evidently due to the difference in the background sequence. The nature of the B10 residue position has been explored with the chemical synthesis of B10 Asn, B10 Leu and B10 Lys human insulins (Burke *et al.*, 1984). The data obtained in this particular study indicated that the steric character rather than the polarity of the amino-acid in this position is important for high biological activity. This view tends to agree with the finding that the B10 Asp human insulin analogue exhibits a potency of 2 to 4 times that of the natural hormone in an *in vitro* lipogenesis assay with rat fat cells (Schwartz *et al.*, 1987). Unfortunately it is however a characteristic of insulin that changes in the activity measured *in vitro* are often not reflected in *in vivo* assays (Bristow, 1993). B10 Asp human insulin, despite having a two to three fold increase in both receptor-binding and *in vitro* bioactivity, is equipotent with human insulin *in vivo* (Bristow, 1993).

In this study it was undertaken to synthesise a HB10D platypus insulin analogue by a totally chemical means. The analogue displayed increased biological activity *in vivo* and the expected reduction in the extent of self-association in a zinc-free neutral pH solution environment. This is in contrast to a similar HB10D human insulin analogue which does not show this increase in biological activity *in vivo* (Bristow, 1993).

The HB10D platypus insulin analogue also shows a dramatic reduction in the self-association as demonstrated in Figure 5.9; a self-association state which does not increase beyond two. A similar finding for the state of association of HB10D human insulin (a similar analogue) was found under similar solution conditions (Weiss *et al.*, 1991) and the conclusion was made that this analogue is dimeric at 0.6 mM. No detailed analyses were however undertaken. The same result was also evident from osmometry studies with this analogue in similar conditions (Brange *et al.*, 1988). To confuse matters it has even been suggested that it is monomeric at pharmacological concentrations (Bristow, 1993). The approach in the

work presented here was to analyse the self-association pattern in a zinc-free neutral pH solution, fitting sedimentation equilibrium data to a model of self-association and calculating the association constants. This has not been attempted before for an insulin with an Aspartic acid in the B10 position.

Although the background sequences between human and platypus insulins are different, it can reasonably be assumed that this does not contribute in a major way to the reduction of self-association as the difference in the extent of the self-association of bovine and platypus is not very marked. In the HB10D platypus insulin one amino-acid residue change in the dimer-dimer interface has been introduced that presumably destroys the β-β interaction under these conditions. This hypothesis can be supported by analysing the self-association pattern of this analogue.

The K_A value (103,900 M^{-1}) that represents the α-α interaction region (monomer-monomer interface) according to the IDI model of HB10D platypus insulin, is just over four times higher in magnitude than that of bovine insulin (25000 M^{-1}). In contrast, the K_B value (69 M^{-1}), that represents the β-β interaction region (dimer-dimer interface) according to the IDI model, of HB10D platypus insulin, is effectively zero (bovine insulin 8000 M^{-1}). It can concluded that this huge reduction in K_B value virtually abolishes the β-β interaction involving the dimer-dimer interface. The increase in the K_A value for HB10D platypus insulin could be explained as follows. It is possible that with the abolishment of the β-β interaction, and thus removing of this previously (in the native sequence insulins) competing self-association event , that the availability for association via the α α interaction region increases. In other words the α - α interaction is stronger in the absence of a competing β-β interaction. These quantitative findings demonstrate nicely that analysis of the self-association pattern employing sedimentation equilibrium and the omega method allows meaningful correlations to be made between changes at the molecular level and their subsequent expression at the structural and biological activity level.

Comparison with published results of state of association

Table 5.3 demonstrates the variation in the state of association ($\overline{M}_w / M_1$) that can be found when different methods, sometimes over various concentration ranges, for collection of sedimentation equilibrium data in similar solution environments. The difference in the association state between platypus and bovine insulin only came to light when the latter insulin is used as a standard against which the state of association of other insulins are measured under the exact conditions. This difference, in the light of these data, must however be interpreted with caution. It can be concluded that the state of association of platypus insulin does not deviate significantly from that found for the other mammalian insulins. Since the state of association of platypus insulin is not dramatically different from other mammalian insulins, it validates the success of the use of chemical synthesis (this thesis). The same conclusion regarding the success of chemical synthesis can be made for HB10D platypus insulin. The association state of this analogue in this study was found to be similar to that for HB10D human insulin.

5.5.4 Weight-fractions as a function of total concentration

In the following discussion it will be assumed that the IDI model for the self-association of insulin applies. The rationale for utilising this model has been given in the theory section of this chapter. As noted already, the sedimentation data acquired and analysed in this study for all insulins also supports the validity of this model of self-association. Thus, variations of the association constants can be explained in terms of the known properties of amino-acid residues in the postulated interaction regions (self-association and receptor binding regions).

Figures 5.13-5.15 clearly demonstrate, in terms of weight-fractions, the complex distribution of species within each of the three insulins under study in a zinc-free solution (pH 7.0; $I = 0.1$ M ; $T = 25$ °C) as well as the effect of the relative difference in K_A and K_B values for the different insulins.

Despite the significant difference between the K_A and K_B values of the different insulins, the weight-fractions of the monomers including HB10D platypus insulin, do not differ significantly, even in the higher concentration range (Figure 5.16). This is reflected in the experimental finding that the extrapolated Ω^0 values of all three insulins at a reference concentration of 0.5 g/L are identical within experimental error.

Bovine, platypus and HB10D platypus insulins are overwhelmingly monomeric at these low concentrations as is shown graphically in Figure 5.16. This can be expected since in humans the plasma insulin concentration range is 0.7 - 3 ng/mL (Fajans *et al.*, 1972) and *in vitro* binding and biological activity studies in cells maximum stimulation occurs at 10 ng/mL (Marsh *et al.*, 1984). Insulin solutions containing 10 ng/mL of insulin would comprise 99.987% monomer by weight using equation 5.23. Apart from the existence of zinc hexamer crystals in the β granules of the pancreas (Baker *et al.*, 1988), in a biological context no function for the self-association of insulin has been found *in vivo* (Jeffrey *et al.*, 1976; DeMeyts *et al.*, 1976).

The magnitude of the K_A value describes the strength of the α-α interaction along the monomer-monomer interface of both monomers in the β- dimer. The constant K_B defines the strength of the β-β interaction to produce α-dimers. Figure 5.18 illustrates the comparison of the weight-fractions of α- and β dimers of bovine and platypus insulin and β-dimers of HB10D platypus insulin. The magnitude of the K_A value of these insulins decrease as follows; HB10D platypus insulin > platypus insulin > and bovine insulin (Table 5.2) and is reflected in the concurrent decrease in the weight-fraction of their β-dimers in the same order with the highest weight- fraction being that of HB10D platypus insulin. The weight-fraction of β-dimer of HB10D platypus insulin is approximately 0.8 (80% by weight) at concentrations 0.4 g/L and higher. In contrast, bovine- and platypus insulin under the above solution conditions consist of an array of different oligomers and polymers.

A similar trend is noticed with the weight-fractions of the α-dimers of these insulins. With a decrease in the magnitude of the K_B value there is a decrease in the weight-

fraction of the α–dimer with the existence of α-dimers in a solution of HB10D platypus insulin negligible, showing that the β-β interaction is virtually wiped out. Likewise the lower K_B value for platypus than bovine insulin is reflected by the lower weight-fraction of the α-dimer of platypus insulin. Comparing the different insulins, the difference in strength of the α-α interaction of the β-dimer and β-β interaction of the α-dimer manifest itself also in the difference of the weight-fractions of the other species, trimers, α- and β-tetramers, pentamers and hexamers. These are all slightly lower for platypus than for bovine insulin in the direction predicted by the modelling exercises (Chapter 3). The really significant result is of course found in HB10D platypus insulin. There are effectively no oligomeric species present in this synthesised, fully active mutant platypus insulin except β-dimers.

5.6 Future work

The production of a therapeutic insulin which is more user friendly is still being pursued. The reason for this is that there are still problems associated with the exogenous administration of insulin. Specific physico-chemical and biological properties have been identified as desirable, for example a fully active insulin that exhibits reduced self-association in solution at the concentration administered and at physiological pH (Chapter 1). Directions to take in the design of these therapeutic insulins may be indicated and become apparent from ongoing research. Some of these indications which have become apparent to this author are outlined below.

The success of the total chemical synthesis of the native and mutant platypus insulins (this thesis) sets the stage for the synthesis of mutant insulins containing non-proteinogenic amino-acids. It is hoped that by doing this it might be possible to abolish or reduce self-association by substituting residues involved in this self-association. An insulin with desirable properties could thus be obtained that would improve insulin therapy (Chapter 1). A13 Met is unique in platypus and B2 Pro is only found in insulins from reptiles and fishes (Chapter 3). The importance of these positions in the insulin molecule is that they are involved in dimer-dimer interaction (Chapter 3) and are thought to be responsible for the reduced self-association that was observed for platypus insulin (this thesis). Thus to investigate the effect of substituting non-proteinogenic amino-acid residues with specific properties in positions A13 and B2 on self-association, mutant A- and B-chains MA13Nle, MA13Mso and PB2Nma were chemically synthesised by the author. These will be combined with their respective native A- and B-chains to form the mutants insulins based on the platypus insulin sequence.

The residue properties of the following non-proteinogenic amino-acids were chosen:

- Nle (Mutant MA13Nle; A13 Met replaced by a nor-leucine).
  Compared to Met (the residue in platypus insulin in this position) Nle is less hydrophobic. Both have similar side-chain lengths. Also Met is involved in a specific interaction with the aromatic rings of the tyrosine on either side of the Met side-chain as observed in the molecular model of platypus insulin (Chapter 3).The effect of disrupting that specific interaction on receptor binding and self-association can be investigated. Also compared to Val (the residue in human insulin) Nle has a longer side-chain length by one carbon. Steric properties on self-association and receptor binding can be investigated.
- Mso (Mutant MA13Mso; A 13 Met replaced by a Met-sulfoxide).
  Compared to Met (the residue in platypus insulin in this position), Mso is more polar and the effect of a more polar residue (similar properties and size) on receptor binding and self-association can be investigated.
- Nma (Mutant PB2Nma; B2 Pro replace by *N*-methyl alanine).
  Compared to Pro (the residue in platypus insulin in this position), Nma is similar in size and polarity but does not have the steric restrictive characteristic of Pro that is known to be a breaker of regular secondary structure (Fasman, 1989). Also Pro is occasionally found to be the second residue in N-terminal capping boxes of helices (Harper and Rose, 1993). B1-8 in insulin is normally in the extended β-conformation in insulin and a Pro in the B2 position may give this a more helical character. (This was observed in the molecular model of platypus insulin, not shown). The self-association properties of this mutant can be investigated.

The residue at position A13 in insulin is implicated in modulating the binding to its receptor (Schäffer, 1994) and platypus insulin has an unique Met in that position.

Therefore receptor binding studies need to be performed on the synthetic native and mutant platypus insulins presented in this study as well as those proposed above since the different amino-acids in the receptor binding region might be significant in influencing a change in binding with its receptor.

Insulin has both metabolic and growth-promoting activities. These growth-promotion activities have been shown to include rapid effects on gene transcription (Messina, 1990). The hystricomorph (guinea pig, porcupine, coypu and casiragua) insulins and insulin-like growth factors display relatively higher potencies for growth promotion effects than most mammalian insulins. Platypus insulin shares residues in the receptor binding region with some of these hystricomorph insulins and insulin-like growth factors (Rinderknecht and Humbel, 1978) that display high potencies for growth promotion effects (King and Kahn, 1981; Blundell, 1972). These residues are A5 Glu and B25 Tyr. The residue B10 Asp in the mutant platypus insulin sequence with similar properties than the corresponding residue in some of the hystricomorph insulins and insulin-like growth factors (B10 Glu) may also contribute to growth. Another is A8 Lys in the mutant and native platypus insulin sequence. His in this A8 position in chicken insulin is thought to contribute to its high potency for growth (King and Kahn, 1981). Lys and His are both basic amino acids in contrast to most mammalian insulins with Thr in that position. (All the amino-acid sequences were found in the SWISSPROT amino-acid sequence database (Chapter 3). It is of interest that the marsupial kangaroo insulin also has a basic amino-acid Asn in that position (Treacy *et al.*, 1989). Thus in the light of some residues of the mutant and native platypus sequence that are implicated in higher potencies for growth than most mammalian insulins this activity should be tested.

Structural information of these insulin mutants from nuclear magnetic resonance (NMR) experiments as well as from X-ray crystallography may also contribute to a better understanding of the self-association properties and biological activity of insulin.

It is noted that an increase in temperature resulted in an increase of the K_A value and a decrease of the K_B value when analysing the self-association pattern of bovine insulin as obtained using sedimentation equilibrium and the IDI model of self-association (Mark *et al.*, 1987). This is a demonstration that temperature has an influence on the self-association pattern of bovine insulin. It would be interesting to investigate the influence of temperature on the self-association behaviour of platypus insulin since platypus has a body temperature of only 32 °C. Since its insulin functions at lower temperatures, the self-association pattern of platypus insulin might be significantly different from the patterns of other mammalian insulins. From such a study some desirable properties may come to light that may be used in the design of therapeutic insulins.

The ideal therapeutic insulin would be one that can be taken orally avoiding the discomfort of application by injection and the problems associated with the latter (Chapter 1). (Also nasal application of insulin as an alternative to injection has been associated with problems.) Such an insulin must not be degraded by the proteolytic enzymes in the digestive system, be absorbed into the circulation and must not be immunogenic. If this is possible it is envisaged that it can best be accomplished by the total chemical synthesis of insulin and substituting natural amino-acids in the sequence with non-proteinogenic amino-acids which have all the specific desirable properties and which also block proteolytic degradation.

It is a hope of this author to be able to extend the work already started and even work in a project with the goal of producing the ideal therapeutic insulin.

Bibliography

Adler, A.J., Greenfield, N.J., and Fasman, G.D. (1973). Circular dichroism and optical rotary dispersion of proteins and polypeptides. Methods Enzymol. *27*, 675-735.

Albright, D.A. and Williams, J.W. (1967). Sedimentation equilibria in polydisperse nonideal solutions. J. Phys. Chem. *71*, 2780-2786.

Applied Biosystems (1988a). Chemistry. In Model 430A Peptide Synthesiser. User's Manual. Version 1.3B. (Foster City, California, U.S.A). pp. 6-1-6-158.

Applied Biosystems (1988b). Introduction to cleavage techniques. In Strategies in peptide synthesis. pp. 4-60.

Araki, E., Lipes, M.A., Patti, M.E., Bruning, J.C., Haag, B., Johnson, R.S., and Kahn, C.R. (1994). Alternative pathway of insulin signalling in mice with targeted disruption of the IRS-1 gene. Nature *372*, 186-190.

Archibald, W.J. (1963). Ultracentrifugal analysis in theory and experiment (New York-London: Academic Press).

Aune, K.C. (1978). Molecular weight measurements by sedimentation equilibrium: some common pitfalls and how to avoid them. (New York: Academic Press).

Badger, J. (1992). Flexibility in crystalline insulins. Biophys. J. *61*, 816-819.

Baker, E.N., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.M., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D., Sakabe, K., Sakabe, N., and Vijayan, N.M. (1988). The structure of 2Zn pig insulin crystals at 1.5 Å resolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *319*, 369-456.

Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., and Mercola, D. (1976). Structure of insulin in 4-zinc insulin. Nature *261*, 166-168.

Bi, R.C., Dauter, Z., Dodson, E.J., Dodson, G., Giordano, F., and Reynolds, C. (1984). Insulin's structure as a modified and monomeric molecule. Biopolymers *23*, 391-395.

Blundell, T.L., Cutfield, J.F., Dodson, E.J., Dodson, G.G., Hodgkin, D.C., and Mercola, D.A. (1972). The crystal structure of rhombohedral 2 zinc insulin. Cold Spring Harb. Symp. Quant. Biol. *36*, 233-241.

Bowen, T.J. (1970). An introduction to ultracentrifugation (London: Wiley-Interscience).

Bradbury, J.H., Ramesh, V., and Dodson, G. (1981). ^{1}H nuclear magnetic resonance study of the histidine residues of insulin. J. Mol. Biol. *150*, 609-613.

Brandenburg, D. (1969). [Des-PheB1-insulin, a crystalline analogue of bovine insulin]. Hoppe Seylers. Z. Physiol. Chem. *350*, 741-750.

Brandenburg, D. (1990). Insulin Chemistry. In Insulin. P. Cuatrecasas and S. Jacobs, eds. (Berlin, Heidelberg, New York: Springer-Verlag), pp. 3-22.

Brange, J., Ribel, U., Hansen, J.F., Dodson, G., Hansen, M.T., Havelund, S., Melberg, S.G., Norris, F., Norris, K., Snel, L., Sorensen, A.R., and Voigt, H.O. (1988).

Monomeric insulins obtained by protein engineering and their medical implications. Nature *333*, 679-682.

Brange, J., Owens, D.R., Kang, S., and Volund, A. (1990). Monomeric insulins and their experimental and clinical implications. Diabetes Care *13*, 923-954.

Brems, D.N., Alter, L.A., Beckage, M.J., Chance, R.E., DiMarchi, R.D., Green, L.K., Long, H.B., Pekar, A.H., Shields, J.E., and Frank, B.H. (1992). Altering the association properties of insulin by amino-acid replacement. Protein Eng. *5*, 527-533.

Bringer, J., Heldt, A., and Grodsky, G.M. (1981). Prevention of insulin aggregation by dicarboxylic amino-acids during prolonged infusion. Diabetes *30*, 83-85.

Bristow, A.F. (1993). Recombinant-DNA-derived insulin analogues as potentially useful therapeutic agents. Trends. Biotechnol. *11*, 301-305.

Brzovic, P.S., Choi, W.E., Borchardt, D., Kaarsholm, N.C., and Dunn, M.F. (1994). Structural asymmetry and half-site reactivity in the T to R allosteric transition of the insulin hexamer. Biochemistry *33*, 13057-13069.

Burke, M.J. and Rougvie, M.A. (1972). Cross- protein structures. I. Insulin fibrils. Biochemistry *11*, 2435-2439.

Carroll, R.J., Hammer, R.E., Chan, S.J., Swift, H.H., Rubenstein, A.H., and Steiner, D.F. (1988). A mutant human proinsulin is secreted from islets of Langerhans in increased amounts via an unregulated pathway. Proc. Natl. Acad. Sci. U. S. A. *85*, 8943-8947.

Casaretto, M., Spoden, M., Diaconescu, C., Gattner, H.G., Zahn, H., Brandenburg, D., and Wollmer, A. (1987). Shortened insulin with enhanced *in vitro* potency. Biol. Chem. Hoppe Seyler *368*, 709-716.

Casassa, E.F. and Eisenberg, H. (1964). Thermodynamic analysis of multicomponent systems. Adv. Protein Chem. *19*, 287-395.

Chan, S.J., Seino, S., Gruppuso, P.A., Schwartz, R., and Steiner, D.F. (1987). A mutation in the B chain coding region is associated with impaired proinsulin conversion in a family with hyperproinsulinemia. Proc. Natl. Acad. Sci. U. S. A. *84*, 2194-2197.

Chang, C. and Meienhofer, J. (1978). Solid-Phase peptide synthesis using mild base cleavage of N^{α}-Fluorenylmethyloxycarbonylamino-acids, exemplified by a synthesis of dihydrosomatostatin. Int. J. Pept. Protein Res. *11*, 246-249.

Choi, W.E., Brader, M.L., Aguilar, V., Kaarsholm, N.C., and Dunn, M.F. (1993). The allosteric transition of the insulin hexamer is modulated by homotropic and heterotropic interactions. Biochemistry *32*, 11638-11645.

Chothia, C., Lesk, A.M., Dodson, G.G., and Hodgkin, D.C. (1983). Transmission of conformational change in insulin. Nature *302*, 500-505.

Chothia, C. and Janin, J. (1975). Principles of protein-protein recognition. Nature *256*, 705-708.

Cuatrecasas, P. and Hollenberg, M.D. (1975). Binding of insulin and other hormones to non-receptor materials: saturability, specificity and apparent "negative cooperativity". Biochem. Biophys. Res. Commun. *62*, 31-41.

Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Emdin, S.F., and Reynolds, C.D. (1979). Structure and biological activity of hagfish insulin. J. Mol. Biol. *132*, 85-100.

Cutfield, S.M., Dodson, G.G., Schwertner, E., and Zahn, H. (1979). X-ray diffraction of crystals of half-synthetic sheep insulin. Hoppe Seylers. Z. Physiol. Chem. *360*, 783-785.

Dathe, M., Gast, K., Zirwer, D., Welfle, H., and Mehlis, B. (1990). Insulin aggregation in solution. Int. J. Pept. Protein Res. *36*, 344-349.

DeMeyts, P., Bainco, A.R., and Roth, J. (1976). Site-site interactions among insulin receptors. Characterization of the negative cooperativity. J. Biol. Chem. *251*, 1877-1888.

DeMeyts, P. (1994). The structural basis of insulin and insulin-like growth factor I receptor binding and its relevance to mitogenic *versus* metabolic signalling. Diabetologia *37*, S135-S148.

Derewenda, U., Derewenda, Z., Dodson, E.J., Dodson, G.G., Reynolds, C.D., Smith, G.D., Sparks, C., and Swenson, D. (1989). Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. Nature *338*, 594-596.

Derewenda, U., Derewenda, Z., Dodson, E.J., Dodson, G.G., Bing, X., and Markussen, J. (1991). X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. J. Mol. Biol. *220*, 425-433.

Derewenda, U., Derewenda, Z.S., Dodson, G.G., and Hubbard, R.E. (1990). Insulin structure. In Insulin. P. Cuatrecasas and S. Jacobs, eds. (Berlin, Heidelberg, New York: Springer-Verlag), pp. 23-39.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic. Acids. Res. *12*, 387-395.

Dische, F.E., Wernstedt, C., Westermark, G.T., Westermark, P., Pepys, M.B., Rennie, J.A., Gilbey, S.G., and Watkins, P.J. (1988). Insulin as an amyloid-fibril protein at sites of repeated insulin injections in a diabetic patient. Diabetologia *31*, 158-161.

Dodson, E.J., Dodson, G.G., Hubbard, R.E., and Reynolds, C.D. (1983). Insulin's structural behavior and its relation to activity. Biopolymers *22*, 281-291.

Dodson, E.J., Dodson, G.G., Lewitova, A., and Sabesan, M. (1978). Zinc-free cubic pig insulin: crystallization and structure determination. J. Mol. Biol. *125*, 387-396.

Dodson, G.G., Hubbard, R.E., and Reynolds, C.D. (1983). Insulin's structural variations and their relation to activity. Biopolymers *22*, 281-292.

Dodson, G.G., Dodson, E.J., Turkenburg, J.P., and Bing, X. (1993). Molecular recognition in insulin assembly. Biochem. Soc. Trans. *21*, 609-614.

Eaton, R.P., Allen, R.C., Schade, D.S., and Standefer, J.C. (1980). "Normal" insulin secretion: the goal of artificial insulin delivery systems? Diabetes Care *3*, 270-273.

Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J.H., Masiarz, F., Kan, Y.W., Goldfine, I.D., and et al (1985). The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling. Cell *40*, 747-758.

Ellman, G.L. (1959). Arch. Biochem. Biophys. *82*, 70-77. (Abstract)

Emdin, S.O., Sonne, O., and Gliemann, J. (1980). Hagfish insulin: the discrepancy between binding affinity and biologic activity. Diabetes *29*, 301-303.

Emdin, S.O. and Falkmer, S. (1977). Phylogeny of insulin. Some evolutionary aspects of insulin production with particular regard to the biosynthesis of insulin in *Myxine glutinosa*. Acta Paediatr. Scand. Suppl. 15-25.

Fajans, S.S., Floyd, J.C., Jr., Knopf, R.F., Pek, S., Weissman, P., and Conn, J.W. (1972). Amino-acids and insulin release in vivo. Isr. J. Med. Sci. *8*, 233-243.

Fasman, G.D. (1989). Prediction of protein structure and the principles of protein conformation (New York: Plenum Press).

Feng, D.F. and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. *25*, 351-360.

Fields, G.B. and Noble, R.L. (1990). Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino-acids. Int. J. Pept. Protein Res. *35*, 161-214.

Fischer, W.H., Saunders, D., Brandenburg, D., Wollmer, A., and Zahn, H. (1985). A shortened insulin with full *in vitro* potency. Biol. Chem. Hoppe Seyler *366*, 521-525.

Fisher, W.H., Saunders, D., Brandenburg, D., Wollmer, A., and Zahn, H. (1985). The shortened insulin with full *in vitro* potency. Biol. Chem. Hoppe Seyler *366*, 521-525.

Frank, B.H. and Veros, A.J. (1968). Physical studies of proinsulin - association behaviour and conformation in solution. Biochem. Biophys. Res. Commun. *32*, 155-160.

Fujita, H. (1962). Mathematical theory of sedimentation analysis. (New York: Academic Press).

Galloway, J.A., Hooper, S.A., Spradlin, C.T., Howey, D.C., Frank, B.H., Bowsher, R.R., and Anderson, J.H. (1992). Biosynthetic human proinsulin. Review of chemistry, *in vitro* and *in vivo* receptor binding, animal and human pharmacology studies, and clinical trial experience. Diabetes Care *15*, 666-692.

Galloway, J.A. and Chance, R.E. (1994). Improving insulin therapy: Achievements and challenges. Horm. Metab. Res. *26*, 591-598.

Goldman, J. and Carpenter, F.H. (1974). Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. Biochemistry *13*, 4566-4574.

Gracy, J., Chiche, L. and Sallantin, J. (1993). Improved alignment of weakly homologous protein sequences using structural information. Prot. Engin. *6*, 821-829.

Grant, T. (1989). The Platypus. A unique mammal. (Kensington, Australia. New South Wales University Press), pp. 1-14.

Grau, U., Seipke, G., Obermeier, R., and Thurow, H. (1982). Neue Insuline. (K.G. Petersen, K.G. Schluter, and L. Kerp, eds.) (Freiburg: Hoechst), pp. 411-418.

Griffiths, M. (1988). The platypus. Scientific American *258*, 60-67.

Gross, D.J., Halban, P.A., Kahn, C.R., Weir, G.C., and Villa Komaroff, L. (1989). Partial diversion of a mutant proinsulin (B10 aspartic acid) from the regulated to the constitutive secretory pathway in transfected AtT-20 cells. Proc. Natl. Acad. Sci. U. S. A. *86*, 4107-4111.

Gruppuso, P.A., Gorden, P., Kahn, C.R., Cornblath, M., Zeller, W.P., and Schwartz, R. (1984). Familial hyperproinsulinemia due to a proposed defect in conversion of proinsulin to insulin. N. Engl. J. Med. *311*, 629-634.

Haneda, M., Polonsky, K.S., Bergenstal, R.M., Jaspan, J.B., Shoelson, S.E., Blix, P.M., Chan, S.J., Kwok, S.C., Wishner, W.B., Zeidler, A., and et al (1984). Familial hyperinsulinemia due to a structurally abnormal insulin. Definition of an emerging new clinical syndrome. N. Engl. J. Med. *310*, 1288-1294.

Harding, S.E. (1985). The representation of equilibrium solute distributions for nonideal polydisperse systems in the analytical ultracentrifuge. Biophys. J. *47*, 247-250.

Harper, E.T. and Rose, G.D. (1993). Helix stop signals in proteins and peptides: the capping box. Biochemistry *32*, 7605-9.

Holladay, L.A., Ascoli, M., and Puett, D. (1977). Conformational stability and self-association of zinc-free bovine insulin at neutral pH. Biochim. Biophys. Acta *494*, 245-254.

Horuk, R., Goodwin, P., O'Connor, K., Neville, R.W.J., and Lazarus, N.R. (1979). Evolutionary change in the insulin receptors of hystricomorph rodents. Nature *279*, 439-440.

Howlett, G.J., Jeffrey, P.D., and Nichol, L.W. (1970). The effects of pressure on the sedimentation equilibrium of chemically reacting systems. J. Phys. Chem. *74*, 3607-3610.

Hu, S., Burke, G.T., Schwartz, G.P., Ferderigos, N., Ross, J.B.A. and Katsoyannis, P.G. (1993). Steric requirements at position B12 for high biological activity in insulin. Biochemistry *32*, 2631-2635.

Hua, Q.X., Shoelson, S.E., Kochoyan, M., and Weiss, M.A. (1991). Receptor binding redefined by a structural switch in a mutant human insulin. Nature *354*, 238-241.

Hua, Q.X., Kochoyan, M., and Weiss, M.A. (1992). Structure and dynamics of des-pentapeptide-insulin in solution: the molten-globule hypothesis. Proc. Natl. Acad. Sci. U. S. A. *89*, 2379-2383.

Hua, Q.X., Shoelson, S.E., Inouye, K., and Weiss, M.A. (1993). Paradoxical structure and function in a mutant human insulin associated with diabetes mellitus. Proc. Natl. Acad. Sci. U. S. A. *90*, 582-586.

Hua, Q.X. and Weiss, M.A. (1991a). Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: sequential resonance assignment and implications for protein dynamics and receptor recognition. Biochemistry *30*, 5505-5515.

Hua, Q.X. and Weiss, M.A. (1991b). Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-specific resonance assignments and effects of solvent composition. Biochim. Biophys. Acta *1078*, 101-110.

Jeffrey, P.D. (1974). Polymerization behavior of bovine zinc-insulin at neutral pH. Molecular weight of the subunit and the effect of glucose. Biochemistry *13*, 4441-4447.

Jeffrey, P.D., Milthorpe, B.K., and Nichol, L.W. (1976). Polymerization pattern of insulin at pH 7.0. Biochemistry *15*, 4660-4665.

Jeffrey, P.D., Nichol, L.W., Turner, D.R., and Winzor, D.J. (1977). The combination of molecular covolume and frictional coefficient to determine the shape and axial ratio of a rigid macromolecule. Studies on ovalbumin. J. Phys. Chem. *81*, 776-781.

Jeffrey, P.D. (1986). Self-association of des-(B26-B30)-insulin. The effect of Ca^{2+} and some other divalent cations. Biol. Chem. Hoppe Seyler *367*, 363-369.

Jeffrey, P.D. (1990). Insulin: Good things come in small packages. Today's Life Science *2*, 45-48.

Jeffrey, P.D. and Coates, J.H. (1966a). An equilibrium ultracentrifuge study of the effect of ionic strenght on the self-association of bovine insulin. Biochemistry *12*, 3820-3824.

Jeffrey, P.D. and Coates, J.H. (1966b). An equilibrium ultracentrifuge study of the self-association of bovine insulin. Biochemistry *5*, 489-498.

Jørgensen, A.M., Kristensen, S.M., Led, J.J., and Balschmidt, P. (1992). Three-dimensional solution structure of an insulin dimer. A study of the B9(Asp) mutant of human insulin using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. J. Mol. Biol. *227*, 1146-1163.

Joshi, S., Burke, G.T., and Katsoyannis, P.G. (1990). An insulin-like compound consisting of the B-chain of bovine insulin and an A-chain corresponding to a modified A- and the D-domains of human Insulin-like Growth Factor I. J. Protein Chem. *9*, 235-246.

Kaarsholm, N.C., Ko, H.C., and Dunn, M.F. (1989). Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin. Biochemistry *28*, 4427-4435.

Kaarsholm, N.C., Norris, K., Jørgensen, R.J., Mikkelsen, J., Ludvigsen, S., Olsen, O.H., Sorensen, A.R., and Havelund, S. (1993). Engineering stability of the insulin monomer fold with application to structure-activity relationships. Biochemistry *32*, 10773-10778.

Kadima, W., Roy, M., Lee, R.W., Kaarsholm, N.C., and Dunn, M.F. (1992). Studies of the association and conformational properties of metal-free insulin in alkaline sodium chloride solutions by one-and two-dimensional ^{1}H NMR. J. Biol. Chem. *267*, 8963-8970.

Kadima, W., Ogendal, L., Bauer, R., Kaarsholm, N., Brodersen, K., Hansen, J.F., and Porting, P. (1993). The influence of ionic strength and pH on the aggregation properties of zinc-free insulin studied by static laser light scattering. Biopolymers *33*, 1643-1657.

Kamber, B., Hartmann, A., Eisler, K., Riniker, B., Rink, H., and Sieber, P. (1980). The synthesis of cystine peptides by iodine oxidation of *S*-trityl-cysteine and *S*-acetamidomethyl-cysteine peptides. Helv. Chim. Acta *63*, 899-915.

Kang, S., Brange, J., Burch, A., Volund, A., and Owens, D.R. (1991a). Absorption kinetics and action profiles of subcutaneously administered insulin analogues (AspB9GluB27, AspB10, AspB28) in healthy subjects. Diabetes Care *14*, 1057-1065.

Kang, S., Creagh, F.M., Peters, J.R., Brange, J., Volund, A., and Owens, D.R. (1991b). Comparison of subcutaneous soluble human insulin and insulin analogues (AspB9, GluB27; AspB10; AspB28) on meal-related plasma glucose excursions in type I diabetic subjects. Diabetes Care *14*, 571-577.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. Adv. Protein Chem. *14*, 1-63.

Kim, H., Deonier, R.C., and Williams, J.W. (1977). The investigation of self-association reactions by equilibrium ultracentrifugation. Chem. Rev. *77*, 659-690.

King, D.S., Fields, C.G., and Fields, G.B. (1990). A cleavage method which minimizes side reactions following Fmoc solid phase peptide synthesis. Int. J. Pept. Protein Res. *36*, 255-266.

King, G.L. and Kahn, C.R. (1981). Non-parallel evolution of metabolic and growth-promoting functions of insulin. Nature *292*, 644-646.

Kruger, P., Strassburger, W., Wollmer, A., van Gunsteren, W.F., and Dodson, G.G. (1987). The simulated dynamics of the insulin monomer and their relationship to the molecule's structure. Eur. Biophys. J. *14*, 449-459.

Kubiak, T. and Cowburn, D. (1986). Enzymatic semisynthesis of porcine despentapeptide (B26-30) insulin using unprotected desoctapeptide (B23-30) insulin as a substrate. Model studies. Int. J. Pept. Protein Res. *27*, 514-521.

Lee, J. and Pilch, P.F. (1994). The insulin receptor: structure, function, and signaling. Am. J. Physiol. *266*, C319-34.

Lenz, V., Gattner, H.G., Sievert, D., Wollmer, A., Engels, M., and Hocker, H. (1991). Semisynthetic des-(B27-B30)-insulins with modified B26-tyrosine. Biol. Chem. Hoppe Seyler *372*, 495-504.

Lin, T.A. and Lawrence, J.C. (1994). Phas-1 as a link between mitogen-activated protein-kinase and translation initiation. Science *266*, 653-656.

Lougheed, W.D., Woulfe Flanagan, H., Clement, J.R., and Albisser, A.M. (1980). Insulin aggregation in artificial delivery systems. Diabetologia *19*, 1-9.

Mark, A.E., Nichol, L.W., and Jeffrey, P.D. (1987). The self-association of zinc-free bovine insulin. A single model based on interactions in the crystal that describes the association pattern in solution at pH 2, 7 and 10. Biophys. Chem. *27*, 103-117.

Mark, A.E., Berendsen, H.J., and van Gunsteren, W.F. (1991). Conformational flexibility of aqueous monomeric and dimeric insulin: a molecular dynamics study. Biochemistry *30*, 10866-10872.

Mark, A.E. and Jeffrey, P.D. (1990). The self-association of zinc-free bovine insulin. Four model patterns and their significance. Biol. Chem. Hoppe Seyler *371*, 1165-1174.

Markussen, J. (1985). Comparative reduction/oxidation studies with single chain des-(B30) insulin and porcine proinsulin. Int. J. Protein Res. *25*, 421-434.

Markussen, J., Diers, I., Engesgaard, A., Hansen, M.T., Hougaard, P., Langkjaer, L., Norris, K., Ribel, U., Sorensen, A.R., and Voigt, H.O. (1987). Soluble, prolonged-acting insulin derivatives. II. Degree of protraction and crystallizability of insulins substituted in positions A17, B8, B13, B27 and B30. Protein Eng. *1*, 215-223.

Markussen, J., Diers, I., Hougaard, P., Langkjaer, L., Norris, K., Snel, L., Sorensen, A.R., Sorensen, E., and Voigt, H.O. (1988). Soluble, prolonged-acting insulin derivatives. III. Degree of protraction, crystallizability and chemical stability of insulins substituted in positions A21, B13, B23, B27 and B30. Protein Eng. *2*, 157-166.

Marsh, J.W., Westley, J., and Steiner, D.F. (1984). Insulin-receptor interactions. Presence of a positive cooperative effect. J. Biol. Chem. *259*, 6641-6649.

Maruyama, K., Nagasawa, H., Isogai, A., Tamura, S., Ishizaki, H., and Suzuki, A. (1990). Synthesis of bombyxin-IV, an insulin-like heterodimeric peptide from the silkworm, *Bombyx mori*. Peptides *11*, 169-171.

Maruyama, K., Nagata, K., Tanaka, M., Nagasawa, H., Isogai, A., Ishizaki, H., and Suzuki, A. (1992). Synthesis of bombyxin-IV, an insulin superfamily peptide from the silkworm, Bombyx mori, by stepwise and selective formation of three disulfide bridges. J. Protein Chem. *11*, 1-12.

McDonald, N.Q., Murray Rust, J., and Blundell, T.L. (1995). The first structure of a receptor tyrosine kinase domain: a further step in understanding the molecular basis of insulin action. Structure. *3*, 1-6.

Melberg, S.G. and Johnson, W.C., Jr. (1990). Changes in secondary structure follow the dissociation of human insulin hexamers: a circular dichroism study. Proteins *8*, 280-286.

Messina, J.L. (1990). Regulation of gene expression by insulin. In Insulin. P. Cuatrecasas and S. Jacobs, eds. (Berlin, Heidelberg, New York: Springer-Verlag), pp. 398-419.

Millican, R.L. and Brems, D.N. (1994). Equilibrium intermediates in the denaturation of human insulin and two monomeric insulin analogs. Biochemistry *33*, 1116-1124.

Milthorpe, B.K., Jeffrey, P.D., and Nichol, L.W. (1975). The direct analysis of sedimentation equilibrium results obtained with polymerizing systems. Biophys. Chem. *3*, 169-176.

Mirmira, R.G., Nakagawa, S.H., and Tager, H.S. (1991). Importance of the character and configuration of residues B24, B25, and B26 in insulin-receptor interactions. J. Biol. Chem. *266*, 1428-1436.

Mirmira, R.G. and Tager, H.S. (1989). Role of the phenylalanine B24 side chain in directing insulin interaction with its receptor. Importance of main chain conformation. J. Biol. Chem. *264*, 6349-6354.

Mirmira, R.G. and Tager, H.S. (1991). Disposition of the phenylalanine B25 side chain during insulin-receptor and insulin-insulin interactions. Biochemistry *30*, 8222-8229.

Moody, A.J., Stan, M.A., Stan, M., and Gliemann, J. (1974). A simple free fat cell bioassay for insulin. Horm. Metab. Res. *6*, 12-16.

Morris, M. and Ralston, G.B. (1985). Determination of the parameters of self-association by direct fitting of the omega function. Biophys. Chem. *23*, 49-61.

Munk, P. and Halbrook, M.E. (1976). Sedimentation equilibrium of polymers in good solvents. Macromolecules *9*, 568-574.

Nakagawa, S.H., Johansen, N.L., Madsen, K., Schwartz, T.W., and Tager, H.S. (1993). Implications of replacing peptide bonds in the COOH-terminal B chain domain of insulin by the (CH_2-NH) linker. Int. J. Peptide Protein Res. *42*, 578-584.

Nagata, K., Maruyama, K., Nagasawa, H., Urushibata, I., Isogai, A., Ishizaki, H., and Suzuki, A. (1992). Bombyxin-II and its disulfide bond isomers: synthesis and activity. Peptides *13*, 653-662.

Nakagawa, S.H. and Tager, H.S. (1986). Role of the phenylalanine B25 side chain in directing insulin interaction with its receptor. Steric and conformational effects. J. Biol. Chem. *261*, 7332-7341.

Nakagawa, S.H. and Tager, H.S. (1987). Role of the COOH-terminal B-chain domain in insulin-receptor interactions. Identification of perturbations involving the insulin mainchain. J. Biol. Chem. *262*, 12054-12058.

Nakagawa, S.H. and Tager, H.S. (1992). Importance of aliphatic side-chain structure at positions 2 and 3 of the insulin A chain in insulin-receptor interactions. Biochemistry *31*, 3204-3214.

Nakagawa, S.H. and Tager, H.S. (1993). Importance of main-chain flexibility and the insulin fold in insulin-receptor interactions. Biochemistry *32*, 7237-7243.

Needleman, S.B. and Wunch, C.D. (1970). A general method applicable to the search for similarities in the amino-acid sequence of two proteins. J. Mol. Biol. *48*, 443-453.

Nichol, L.W., Sculley, M.J., Jeffrey, P.D., and Winzor, D.J. (1984). Indefinite self-association of a solute in linear and branched arrays. J. Theor. Biol. *109*, 285-298.

Nichol, L.W. and Ogston, A.G. (1965). Sedimentation equilibrium in reacting systems of the type $mA + nB \rightarrow$ C. J. Phys. Chem. *69*, 4365-4367.

Nourse, A., Treacy, G.B., Shaw, D.C., and Jeffrey, P.D. (1996). Platypus insulin: Indications from the amino-acid sequence of significant differences in structure from porcine insulin. Biol. Chem. Hoppe-Seyler *377*, 147-153.

Olsen, C.L., Turner, D.S., Iravani, M., Waxman, K., Selam, J.L., and Charles, M.A. (1995). Diagnostic procedures for catheter malfunction in programmable implantable intraperitoneal insulin infusion devices. Diabetes Care *18*, 70-76.

Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U. S. A. *85*, 2444-2448.

Pekar, A.H. and Frank, B.H. (1972). Conformation of proinsulin. A comparison of insulin and proinsulin self-association at neutral pH. Biochemistry *11*, 4013-4016.

Peterson, J.D., Coulter, C.L., Steiner, D.F., Emdin, S.O., and Falkmer, S. (1974). Structural and crystallographic observations on hagfish insulin. Nature *251*, 239-240.

Peterson, J.D., Steiner, D.F., Emdin, S.O., and Falkmer, S. (1975). The amino-acid sequence of the insulin from a primitive vertebrate, the Atlantic hagfish (*Myxine glutinosa*). J. Biol. Chem. *250*, 5183-5191.

Pocker, Y. and Biswas, S.B. (1980). Conformational dynamics of insulin in solution. Circular dichroic studies. Biochemistry *19*, 5043-5049.

Pocker, Y. and Biswas, S.B. (1981). Self-association of insulin and the role of hydrophobic bonding: a thermodynamic model of insulin dimerization. Biochemistry *20*, 4354-4361.

Pullen, R.A., Lindsay, D.G., Wood, S.P., Tickle, I.J., Blundell, T.L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., and Gammeltoft, S. (1976). Receptor-binding region of insulin. Nature *259*, 369-373.

Renard, E., Baldet, P., Picot, M.C., Jacques Apostol, D., Lauton, D., Costalat, G., Bringer, J., and Jaffiol, C. (1995). Catheter complications associated with implantable systems for peritoneal insulin delivery. An analysis of frequency, predisposing factors, and obstructing materials. Diabetes Care *18*, 300-306.

Rinderknecht, E. and Humbel, R.E. (1978). The amino-acid sequence of human insulin like growth factor I and its structural homology with proinsulin. J. Biol. Chem. *253*, 2769-2776.

Robbins, D.C., Cooper, S.M., Fineberg, S.E., and Mead, P.M. (1987). Antibodies to covalent aggregates of insulin in blood of insulin-using diabetic patients. Diabetes *36*, 838-841.

Roy, M., Lee, R.W., Brange, J., and Dunn, M.F. (1990). ^{1}H NMR spectrum of the native human insulin monomer. Evidence for conformational differences between the monomer and aggregated forms. J. Biol. Chem. *265*, 5448-5452.

Ruiz-Gayo, M., Albericio, F., Pons, M., Royo, M., Pedroso, E., and Giralt, E. (1988). Uteroglobin-like peptide cavities. I. Synthesis of antiparallel and parallel dimers of bis-cysteine peptides. Tetrahedron Lett. *29*, 3845-3848.

Sanger, F. (1959). Science *129*, 1340-1344.

Sarin, V.K., Kent, S.B., Tam, J.P., and Merrifield, R.B. (1981). Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction. Anal. Biochem. *117*, 147-157.

Schäffer, L. (1994). A model for insulin binding to the insulin receptor. Eur. J. Biochem. *221*, 1127-1132.

Schwartz, G.P., Burke, G.P., and Katsoyannis, P.G. (1987). A superactive insulin: [B10-Aspartic acid]insulin(human). Proc. Natl. Acad. Sci. U.S.A *84*, 6408-6411.

Schwartz, G.P., Burke, G.T., and Katsoyannis, P.G. (1989). A highly potent insulin: des-(B26-B30)-[AspB10,TyrB25-NH_2]insulin(human). Proc. Natl. Acad. Sci. U. S. A. *86*, 458-461.

Sefton, M.V. and Antonacci, G.M. (1984). Adsorption isotherms of insulin onto various materials. Diabetes *33*, 674-680.

Seino, S., Steiner, D.F., and Bell, G.I. (1987). Sequence of New World primate insulin having low biological potency and immunoreactivity. Proc. Natl. Acad. Sci. U. S. A. *84*, 7423-7427.

Shields, J.E. (1993). Solution conformation model of LysPro insulin analog. Diabetes *42*, 223A (Abstract)

Shoelson, S., Fickova, M., Haneda, M., Nahum, A., Musso, G., Kaiser, E.T., Rubenstein, A.H., and Tager, H. (1983a). Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution. Proc. Natl. Acad. Sci. U. S. A. *80*, 7390-7394.

Shoelson, S., Haneda, M., Blix, P., Nanjo, A., Sanke, T., Inouye, K., Steiner, D., Rubenstein, A., and Tager, H. (1983b). Three mutant insulins in man. Nature *302*, 540-543.

Shoelson, S.E., Lu, Z.X., Parlautan, L., Lynch, C.S., and Weiss, M.A. (1992). Mutations at the dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. Biochemistry *31*, 1757-1767.

Sluzky, V., Tamada, J.A., Klibanov, A.M., and Langer, R. (1991). Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces. Proc. Natl. Acad. Sci. U. S. A. *88*, 9377-9381.

Smith, P.E., Al-Obeidi, F., and Pettitt, B.M. (1991). Aspects of the design of conformationally constrained peptides. Methods Enzymol. *202*, 411-436.

Smith, G.D., Swenson, D.C., Dodson, E.J., Dodson, G.G., and Reynolds, C.D. (1984). Structural stability in the 4-zinc human insulin hexamer. Proc. Natl. Acad. Sci. U. S. A. *81*, 7093-7097.

Sonne, O. (1986). The reversible receptor binding of insulin in isolated rat adipocytes measured at 37 °C. The binding is not rate limiting for cellular uptake. Biochim. et Biophys. *886*, 302-309.

Soucek, D.A. and Adams, E.T. (1976). Molecular weight distributions from sedimentation equilibrium of nonideal solutions. J. Colloid. Interface Sci. *55*, 571-582.

Spoden, M., Gattner, H.G., Zahn, H., and Brandenburg, D. (1995). Structure-function relationships of des-(B26-B30)-insulin. Int. J. Pept. Protein Res. *46*, 221-227.

Svedberg, T. and Pedersen, K.O. (1940). The ultracentrifuge (Oxford: Oxford University Press).

Svoboda, I., Brandenburg, D., Barth, T., Gattner, H.G., Jiracek, J., Velek, J., Blaha, I., Ubik, K., Kasicka, V., Pospisek, J., and et al (1994). Semisynthetic insulin analogues modified in positions B24, B25 and B29. Biol. Chem. Hoppe Seyler *375*, 373-378.

Tager, H., Given, B., Baldwin, D., Mako, M., Markese, J., Rubenstein, A., Olefsky, J., Kobayashi, M., Kolterman, O., and Poucher, R. (1979). A structurally abnormal insulin causing human diabetes. Nature *281*, 122-125.

Tang, L.H., Powell, D.R., Escott, B.M., and Adams, E.T., Jr. (1977). Analysis of various indefinite self-associations. Biophys. Chem. *7*, 121-139.

Taylor, W.R. (1988). Pattern matching methods in protein sequence comparison and structure prediction. Protein Eng. *2*, 77-86.

Teller, D.C. (1973). Characterization of proteins by sedimentation equilibrium in the analytical ultracentrifuge (New York: Academic Press).

Treacy, G.B., Shaw, D.C., Griffiths, M.E., and Jeffrey, P.D. (1989). Purification of a marsupial insulin: amino-acid sequence of insulin from the eastern grey kangaroo *Macropus giganteus*. Biochim. Biophys. Acta *990*, 263-268.

Ullrich, A., Bell, J.R., Chen, E.Y., Herrera, R., Petruzzelli, L.M., Dull, T.J., Gray, A., Coussens, L., Liao, Y.C., Tsubokawa, M., and et al (1985). Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature *313*, 756-761.

Vora, J.P., Owens, D.R., Dolben, J., Atiea, J.A., Dean, J.D., Kang, S., Burch, A., and Brange, J. (1988). Recombinant DNA derived monomeric insulin analogue: comparison with soluble human insulin in normal subjects. BMJ. *297*, 1236-1239.

Weiss, M.A., Nguyen, D.T., Khait, I., Inouye, K., Frank, B.H., Beckage, M., O'Shea, E., Shoelson, S.E., Karplus, M., and Neuringer, L.J. (1989). Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. Biochemistry *28*, 9855-9873.

Weiss, M.A., Hua, Q.X., Lynch, C.S., Frank, B.H., and Shoelson, S.E. (1991). Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective ^{2}H and ^{13}C labeling with application to protein design. Biochemistry *30*, 7373-7389.

White, M.F. and Kahn, C.R. (1994). The insulin signaling system. J. Biol. Chem. *269*, 1-4.

Williams, J.W., Van Holde, K.E., Baldwin, R.L., and Futjita, H. (1958). The theory of sedimentation analysis. Chem. Rev. *58*, 715-806.

Wills, P.R., Nichol, L.W., and Siezen, R.J. (1980). The indefinite self-association of lysozyme: consideration of composition-dependent activity coefficients. Biophys. Chem. *11*, 71-82.

Wlodawer, A., Savage, H., and Dodson, G. (1989). Structure of insulin: results of joint neutron and X-ray refinement. Acta Crystallogr. B. *45*, 99-107.

Wodak, S.J., Alard, P., Delhaise, P., and Renneboog Squilbin, C. (1985). Simulation of conformational changes in 2 Zn insulin. J. Mol. Biol. *181*, 317-322.

Wollmer, A., Rannefeld, B., Johansen, B.R., Hejnaes, K.R., Balschmidt, P., and Hansen, F.B. (1987). Phenol-promoted structural transformation of insulin in solution. Biol. Chem. Hoppe Seyler *368*, 903-911.

Wollmer, A., Gilge, G., Brandenburg, D., and Gattner, H.G. (1994). An insulin with the native sequence but virtually no activity. Biol. Chem. Hoppe Seyler *375*, 219-222.

Wüthrich, K. (1986). NMR of proteins and nucleic acids (New York: Wiley and Sons Inc.).

Yu, N.T., Jo, B.H., Chang, R.C., and Huber, J.D. (1974). Single-crystal Raman spectra of native insulin. Structures of insulin fibrils, glucagon fibrils, and intact calf lens. Arch. Biochem. Biophys. *160*, 614-622.

Biol. Chem. Hoppe-Seyler, Vol. 377, pp. 147–153, February 1996 ·

Short Communication

Platypus Insulin: Indications from the Amino Acid Sequence of Significant Differences in Structure from Porcine Insulin

Amanda Nourse, G. Barbara Treacy, Denis C. Shaw and Peter D. Jeffrey*

Protein Biochemistry Group, Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, Canberra, 2601, Australia

* Corresponding author

Insulin from a monotreme, the platypus *(Ornithorhynchus anatinus)*, was isolated and the amino acid sequence determined. It differs from pig insulin at eleven amino acid sites, mainly on the surface of the monomer. Substitutions relative to pig insulin occur in the monomer-monomer interface, the dimer-dimer interface and the receptor binding region. The residues A5 Glu, A8 Lys and A13 Met have not been reported before in any insulin. Multiple sequence comparison studies reveal a relatively close relationship with the nearest group of relatives to the platypus, the mammals. The relationship of the platypus sequence to reptilian insulin sequences (and amphibian and avian insulin sequences in this case) is sufficiently close to support the observation that platypus has retained some ancient reptilian characteristics over the course of evolution. Model building the platypus insulin sequence on the structure of porcine insulin indicates that there may be some interesting differences.

Key words: Amino acid sequence / Insulin structure / Molecular modelling / Platypus / Sequence comparison.

The platypus *(Ornithorhynchus anatinus)* is the only extant species of the family Ornithorhynchidae belonging to the order Monotremata, the egg laying mammals. It is an ancient highly specialised animal well adapted to its amphibious existence in the waterways of eastern Australia (Grant, 1989). Its biological features indicate that it is more closely related to marsupial and placental mammals than to any group of reptiles, despite retaining several ancestral reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988).

Mammalian insulin, under physiological conditions of pH and at concentrations required for exogenous administration, strongly self-associates in solution to form a complex mixture of polymeric species (Blundell *et al.*, 1972; Jeffrey *et al.*, 1976). As insulin is biologically active in its monomeric form, attempts have been made to keep insulin in this highly desirable state (Brange *et al.*, 1988). Numerous strategies have been employed in efforts to try and achieve this, mainly on insulins from human, pig and cow, without much success. These are addition (Bringer *et al.*, 1981) and removal (Jeffrey, 1986) of substances to insulin solutions and replacement or modification of strategically chosen amino-acids (Brange *et al.*, 1988; Weiss *et al.*, 1991). Investigation of the composition, structure, biological activity and self-association of insulin from more exotic species, like the platypus, increases the possibility of identifying an insulin with desirable properties by exploiting natural variation.

The uniqueness of platypus insulin presents this opportunity to study the structure-function relationship because it differs from that of human at eleven amino-acid sites, three of which exhibit residues that have not been observed in other organisms. It also differs from pig and human insulin at its self-associating surfaces (monomer-monomer and dimer-dimer interfaces) by 4 residues, offering the possibility of a change in association properties, and at its receptor binding surface by 2 residues, offering the possibility of a change in biological activity.

On the basis of the complete platypus insulin amino-acid sequence, a three dimensional structure was modelled on the X-ray crystal structure of porcine insulin. These studies form part of a program with the aim of evaluating insulins and insulin derivatives, both new and previously reported, with the desired characteristics.

A Chain

1				5				10				15				20
G	I	V	E	E	C	C	K	G	V	C	S	M	Y	Q	L	E	N	Y	C	N
				Q			T	S	I			L								

B Chain

1				5				10				15				20				25				30
F	P	N	Q	H	L	C	G	S	H	L	V	E	A	L	Y	L	V	C	G	E	K	G	F	Y	Y	I	P	R	M
	V																				R			F	T		K	A	

Fig. 1 Amino Acid Sequence of Insulin from Platypus *Ornithorhynchus anatinus*.
The residues of porcine insulin at sites where they differ are shown below. Peptide sequences were determined using the Applied Biosystems Model 477, a pulsed liquid phase sequencer with on-line Model 120A PTH-Analyser modified for nanomole sensitivity. Methods used were as described previously (Treacy *et al.*, 1989). The molecular mass of platypus insulin, calculated as 5865 daltons, has been confirmed with the matrix-assisted laser desorption ionisation (MALDI) time-of-flight mass spectrometer.

Table 1 A Comparison of the Amino Acid Sequence of Platypus and Porcine Insulin.
Only platypus amino acid substitutions relative to porcine insulin are presented.

Residue	Amino acid		Side chain properties		Structural domain[a,b]	Type[a,b]
	Porcine	Platypus	Porcine	Platypus		
A5	Gln	Glu	polar	acidic	α-helix, receptor binding	conserved
A8	Thr	Lys	polar	basic	α-helix	variable
A9	Ser	Gly	polar		surface residue	variable
A10	Ile	Val	non polar	non polar	extended β conformation	variable
A13	Leu	Met	non polar	non polar, S	α-helix, dimer-dimer interface	variable
B2	Val	Pro	non polar	non polar	dimer-dimer interface	variable
B22	Arg	Lys	basic	basic	β bend	conserved
B25	Phe	Tyr	aromatic	aromatic	monomer-monomer interface, receptor binding region	conserved
B27	Thr	Ile	polar	non polar	monomer-monomer interface	variable
B29	Lys	Arg	basic	basic	extended β chain	variable
B30	Ala	Met	polar	non polar, S	extended β chain	variable

[a] Baker *et al.*, 1988.
[b] Shoelson *et al.*, 1992.

Table 2 Comparison of Platypus Insulin with Insulin Sequences from other Organisms.
Only platypus insulin residues different from porcine insulin are shown. Sequence library used was the SWISS-PROT database.

Organisms that share residues with platypus insulin	Platypus insulin residues	
	A5[a]	E
	A8[a]	K
	A13[a]	M
Rattle snake, cantor, electric ray, dogfish, degu, copyu, casiragua	B25[b]	Y
Elephant, goat, porcupine, sheep, horse, guinea pig	A9[c]	G
Elephant, goat, porcupine, sheep, cat, camel, bovine	A10[c]	V
Rattle snake, cantor, electric ray, dogfish, ratfish, elephantfish	B2[c]	P
Electric ray, dogfish, salmon, alligator	B22[c]	K
Guinea pig	B27[c]	I
Rattle snake, cantor	B29[c]	R
Sponge	B30[c]	M

[a] Residues not observed in other organisms.
[b] Residues conserved, but observed in other organisms.
[c] Residues variable and shared with other organisms.

Platypus pancreatic tissue was obtained from 9 animals which had been deep frozen at death. The extraction of insulin from the pancreases was carried out in the cold according to the method of Treacy *et al.* (1989). The precipitate did not separate cleanly from the liquid but rather as a dark coloured oily-looking liquid between the yellow precipitate and the other layer. Platypus insulin was obtained from this 'oily supernatant'. Reverse phase HPLC was employed to purify the insulin. Elution was in 0.09% aqueous trifluoroacetic acid using a linear gradient from 8% to 50% acetonitrile in 20 minutes.

The amino-acid sequence of platypus insulin, shown in Figure 1, differs from that of human and pig insulin at eleven residues. Most of these, summarised in Table 1, represent relatively conservative substitutions in regions that have been observed as varying from highly conserved to variable. Three, A5 Glu, A8 Lys and A13 Met have not been observed in other organisms (Table 2). Most substitutions occur on the surface, or at least partial surface, in regions of important receptor binding and aggregation of the molecule. The amino acid sequence reveals strict conservation of the residues known to be involved in maintaining the three dimensional structure (Baker *et al.*, 1988). These include the thirteen invariant residues (the six half-cystine residues, A6, A7, A11, A20, B7, B19, glycine residues at B8 and B23, leucine residues at A16, B6, B11, and B15 and B12 Val) as well as the invariant group of surface residues believed to be involved in receptor binding and expression of biological activity (A1 Gly, A19 Tyr and A21 Asn) (Baker *et al.*, 1988). Some of the conservative sites A2 Ile, A3 Val, B24 Phe and B26 Phe to be found on the surface of the molecule are present in the platypus insulin structure. A5, B22 and B25, that are grouped with the latter, are conservative changes of A5 Gln→Glu, B22 Arg→Lys and B25 Phe→Tyr in platypus insulin.

There are three distinct surfaces on the insulin molecule, namely the monomer-monomer interface (A21, B8, B9, B12, B16, B20, B21 and B23–B28), the dimer-dimer interface (A13, A14, A17, B1, B2, B4, B10, B13, B14, and B17–20) that constitute the protein-protein interactive surfaces and the receptor binding surface (A1, A5, A19, A21, B12, B16, B23, B24 and B25) (Baker *et al.*, 1988; Shoelson *et al.*, 1992). Platypus insulin has substitutions at all three interactive surfaces, namely B25 Phe→Tyr and B27 Thr→Ile (monomer-monomer interface); A13 Leu→Met and B2 Val→Pro (dimer-dimer interface) and A5 Gln→Glu and B25 Phe→Tyr (receptor binding region).

Self-associating forces that lead to dimerisation of insulin (Baker *et al.*, 1988; Shoelson *et al.*, 1992) are predominantly non-polar, reinforced by hydrogen bonding (Baker *et al.*, 1988; Pocker and Biswas, 1981). Thus B25 and B27 Ile,

the residues different in platypus and part of monomer-monomer interface, considered conservative changes, fulfil these criteria. All other residues that form part of this surface, including B28 Pro, the latter especially important for high affinity self association (Brems *et al.*, 1992), are unchanged in platypus insulin. It can be concluded then that it is highly likely that the platypus insulin will form dimers and as B5 His and B 10 His are conserved, hexamerization can be expected in the presence of zinc (Bentley *et al.*, 1976; Blundell *et al.*, 1972).

A model for insulin binding to the receptor was recently proposed (Schäffer, 1994) where the involvement of A13 and B17 in a putative second receptor binding site were discussed. Platypus insulin has a Met in position A13, which has not been seen in other organisms. B25 Tyr, also part of the receptor binding region and different from pig insulin, is Phe in most organisms. The aromatic ring seems to be important for biological activity (Nakagawa and Tager, 1986). It has been shown that substitution at A8 Thr for Arg or His (positively charged residues) provides strong stabilisation of the A-chain NH_2-terminal α-helix and lead to enhanced biological activity (Kaarsholm *et al.*, 1993). Substitutions at A5 Gln→Glu (receptor binding region and charged) and A8 Thr→Lys (positively charged residue) of platypus insulin introduce possible charge interactions in this region. The A5 Glu and B25 Tyr substitutions are known from the IGF-1 molecule (Rinderknecht and Humbel, 1978) and in combination with substitutions in A8, A13 and B22, indicate that binding properties to the insulin and IGF-1 receptors would be interesting to investigate. Receptor interaction and biological activity may be affected by the above substitutions and the structural consequences need direct investigation.

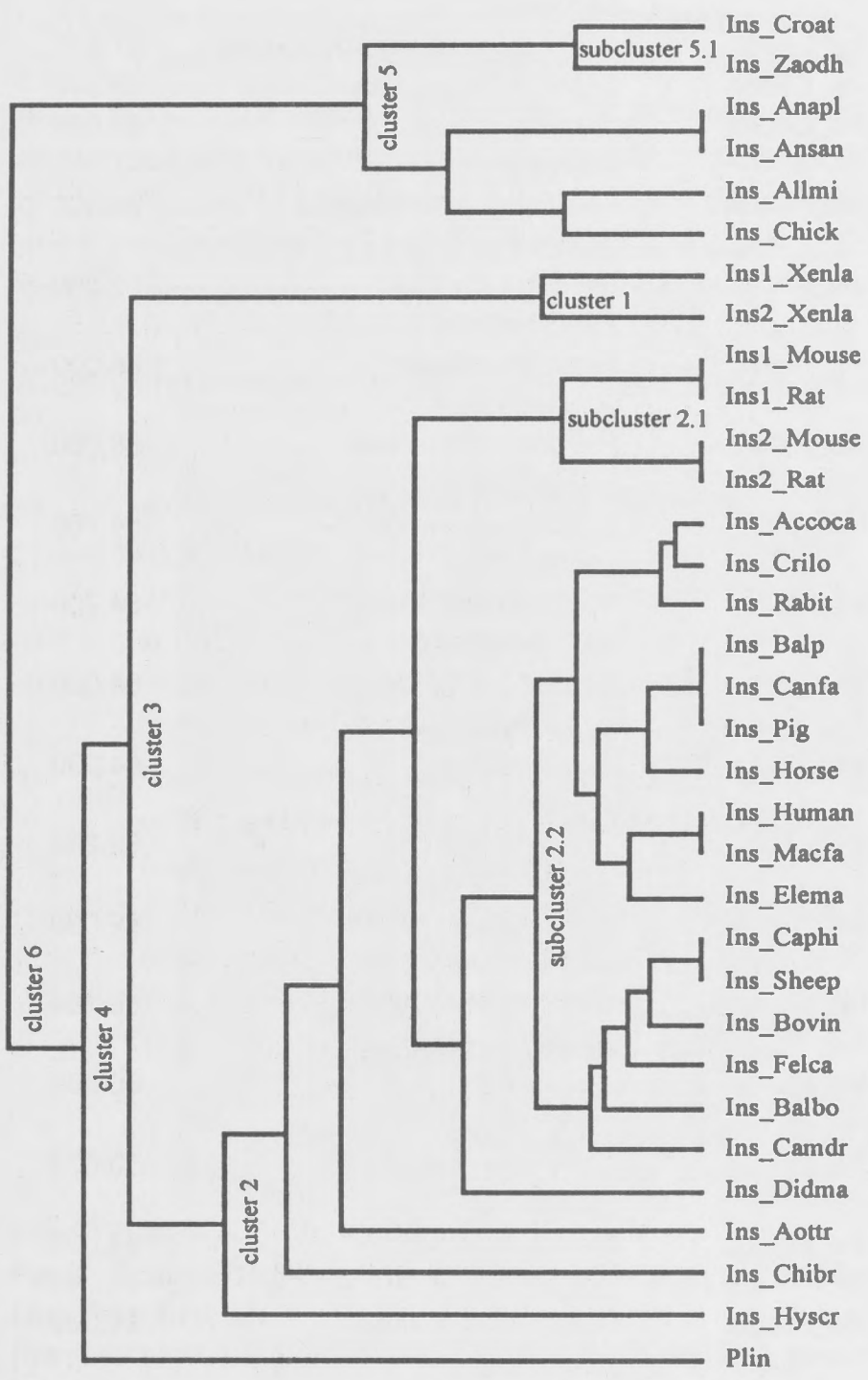


Fig. 2 Dendogram Displaying the Clustering Relationships of the Primary Sequence of Platypus Insulin with Insulins from Other Organisms Using the PILEUP Program, Genetics Computer Group (GCG) Sequence Analysis Software Package Version 7.1 (Devereux *et al.*, 1984).
PILEUP is an extension of GAP (GCG) where a multiple sequence alignment (Feng and Doolittle, 1987), is created from a group of related sequences using progressive pairwise alignment. Platypus insulin was given the name plin. Names of insulins from different organisms are SWISS-PROT designations and their sequences and their respective references are stored in the SWISS-PROT database. Own clustering nomenclature was used.

Pairwise comparisons of platypus insulin sequences with insulin sequences from other organisms were performed using the GAP sequence comparison algorithm of Needleman and Wunch (Needleman and Wunch, 1970). This is presented in Table 3 in order of decreasing percent identity with the platypus insulin sequence.

The insulin sequences from different organisms were subjected to multiple sequence alignment on the basis of pairwise similarity. The vertical branch distance of the dendogram are proportional to the similarity between sequences (Devereux *et al.*, 1984). In the resultant dendogram the insulin sequences cluster into 4 main branches (Figure 2).

The first two branches consist of the amphibian insulin sequence (cluster 1) and mammalian insulin sequence (cluster 2). These two together form cluster 3. The mammalian insulin sequences (cluster 2) can be roughly divided into subcluster 2.1, the mouse and rat insulin sequences and subcluster 2.2, which contain the very similar sequences of porcine, bovine, human, elephant etc. It can be pointed out that the marsupial insulin sequence (*Didelphis marsupials virginiana* identical to kangaroo *Macropus giganteus*) (Treacy *et al.*, 1989) and the low potency insulins of *Aotus trivirgatus* (Seino *et al.*, 1987), *Hystrix cristata* (King and Kahn, 1981) and *Chinchilla brevicaudata* (Horuk *et al.*, 1979) are outside subclusters 2.1 and 2.2 which indicate that they are more distantly related. The third branch is represented by the single platypus insulin sequence which participates in a clustering relationship with the amphibian and mammalian insulin sequences (cluster 3) to form cluster 4. The final branch consist of reptilian and avian insulin sequences (cluster 5) that clusters with cluster 4 to form cluster 6. (Multiple sequence alignment with fish and other hystricomorph rodent insulin sequences are not shown.)

On taxonomic grounds it is believed that the monotremes are the last survivors of a group of early mammals that evolved independently of the creatures that gave rise to today's marsupials and other mammals (Grant, 1989;

Table 3 Pairwise Comparison of Platypus Insulin with Insulin Sequences from Other Organisms in Order of Decreasing Percent Identity.

SWISS-PROT Nomenclature	Organisms	% Identity	SWISS-PROT Nomenclature	Organisms	% Identity
Plin[a]	*Ornithorhynchus anatinus* (platypus)	100	Ins_Geocy	*Geodia cydonium* (sponge)	74.000
			Ins2_Xenla	*Xenopus laevis* 2 (African clawed frog)	72.549
Ins_Elema	*Elephas maximus* (Indian elephant)	82.353	Ins1_Mouse	*Mus musculus* 1 (mouse)	72.549
Ins_Caphi	*Capra hircus* (goat)	82.353	Ins1_Rat	*Rattus norvegicus* 1 (rat)	72.549
			Ins_Squac	*Squalus acanthias* (spiny dogfish)	72.549
Ins_Sheep	*Ovis arie* (sheep)	82.353	Ins1_Xenla	*Xenopus laevis* 1 (African clawed frog)	70.588
Ins_Horse	*Equus caballus* (horse)	80.392	Ins_Aottr	*Aotus trivirgatus* (night monkey)	70.588
			Ins_Torma	*Torpedo marmorata* (electric ray)	70.588
Ins_Camdr	*Camelus dromedarius* (Arabian camel)	80.392	Ins_Chibr	*Chinchilla brevicaudata* (chinchilla)	68.627
Ins_Bovin	*Bos taurus* (bovine)	80.392	Ins_Oncgo	*Onchorhynchus gorbuscha* (salmon)	68.000
Ins_Balph	*Balaenoptera physalus* (finback whale)	78.431	Ins_Loppi	*Lophius piscatorius* (goosefish)	67.347
			Ins_Cypca	*Cyprinus carpio* (carp)	67.347
Ins_Macfa	*Macaca fascicularis* (monkey)	78.431	Ins_Oncke	*Oncorhynchus keta* (salmon)	66.667
Ins_Canfa	*Canis familiaris* (dog)	78.431	Ins_Myosc	*Myoxocephalus scorpius* (sculpin)	66.000
Ins_Pig	*Sus scrofa* (pig)	78.431	Ins_Thuth	*Thunnus thynnus* 2 (bluefin tuna)	66.000
Ins_Human	*Homo sapiens* (man)	78.431	Ins_Plafe	*Platichthys flesus* (flounderer)	66.000
Ins_Rabit	*Oryctolagus cuniculus* (rabbit)	78.431	Ins_Petma	*Petromyzon marinus* (sea lamprey)	66.000
Ins_Felca	*Felis catus* (cat)	78.431	Ins_Calmi	*Callorhynchus milii* (elephantfish)	64.706
Ins_Balbo	*Balaenoptera borealis* (sei whale)	78.431	Ins_Hydco	*Hydrolagus colliei* (spotted ratfish)	64.706
Ins_Anapl	*Anas plathyrhynchos* (duck)	76.471	Ins1_Batsp	*Batrachoididae sp* 1 (toadfish)	64.000
Ins_Croat	*Crotalus atrox* (rattlesnake)	76.471	Ins_Gadca	*Gadus callarias* (Baltic cod)	64.000
Ins_Ansan	*Anser anser anser* (goose)	76.471	Ins2_Batsp	*Batrachoididae sp* (toadfish)	63.265
Ins_Chick	*Gallus gallus* (chicken)	76.471			
Ins_Crilo	*Cricetulus longicaudatus* (Chinese hamster)	76.471	Ins_Katpe	*Katsuwonus pelamis* (skipjack tuna)	62.000
Ins_Hyscr	*Hystrix cristata* (porcupine)	76.471	Ins_Myxgl	*Myxine glutinosa* (Atlantic hagfish)	60.784
Ins_Lepsp	*Lepisosteus spatula* (alligator)	74.510			
Ins_Zoadh	*Zaocys dhumnades* (cantor)	74.510	Ins_Octde	*Octodon degus* (degu)	60.000
Ins_Acoca	*Acomys cahirinus* (Egyptian spiny mouse)	74.510	Ins_Cavpo	*Cavia porcellus* (guinea pig)	58.824
Ins_Didma	*Didelphis marsupials virginiana* (opossum)	74.510	Ins_Myoco	*Myocastor coypus* (coypu)	54.000
Ins2_Rat	*Rattus norvegicus* 2 (rat)	74.510			
Ins2_Mouse	*Mus musculus* 2 (mouse)	74.510	Ins_Progu	*Proechimys guairae* (casiragua)	54.000
Ins_Allmi	*Alligator mississippiensis*	74.000			

[a] Name assigned for platypus insulin.

Amino-acid sequence comparisons were performed using the Genetics Computer Group (GCG) Sequence Analysis Software Package Version 7.1 (Devereux *et al.*, 1984) that include the FASTA, GAP and SEQED programmes. The computationally fast algorithm FASTA, (Pearson and Lipman, 1988) that allows a homology search of one sequence against the entire SWISS-PROT amino acid sequence database was employed to search for insulin from different organisms with amino acid sequences related to the amino acid sequence of platypus insulin. Scores between platypus insulin amino acid sequence and each of the database entries were calculated and presented in order of decreasing best scores with the highest scores corresponding to the strongest relationship. The best 55 scores (all the insulin amino acid sequences) were subjected to a more sensitive sequence comparison program, GAP (Needleman and Wunch, 1970). The GAP program that covers the complete sequence, maximises the number of matches and minimises the number of gaps. Data entries, where the amino acid sequences included the signal and C chain, had been edited out with SEQED for the GAP alignments.

Griffiths, 1988). The PILEUP dendogram, displaying interrelationships of insulin sequences from different animals (Figure 2) reflects this theory. Although the platypus insulin sequence is not part of subcluster 2.2 that contains the similar sequences of human, porcine, and elephant etc. and from which some rodent and marsupial insulin sequences are excluded, its closest relationship appears to be with the mammalian insulin sequence as a single member of a more distantly related branch (cluster 2 of Figure 2 and highest percent identity, Table 3). Its relationship with reptilian insulin sequences (and amphibian and avian insulin sequences in this case) is sufficiently close to support the observation that platypus has retained some ancient reptilian characteristics over the course of evolution (Grant, 1989; Griffiths, 1988). The insulin sequence of the other living member of the monotremes the echidna, two species *Tachyglossus aculeatus* and *Zaglossus bruijnii* (Grant, 1989; Griffiths, 1988) is still unknown. It would be expected that the sequence is similar (or identical) to that of platypus and thus should subcluster with it. In support of this proposal is the similar subclustering of the mouse and rat in subcluster 2.1 as well as the snakes in subcluster 5.1 (own nomenclature). The results of the sequence comparison studies fits in with known taxonomic data used to classify and compile a hypothesis for the evolutionary relationship of platypus with these animals (Grant, 1989; Griffiths, 1988).

In the computer modelling exercise of the platypus insulin, monomer 2 (Chinese convention) was energy minimised and molecular dynamics for 100 ps of simulation at 200 K have been performed *in vacuo*. Upon viewing the animation of the molecular dynamics simulation from 1 to 100 ps it was noted that the structures changed very little between 35 ps to 100 ps. Movement of the B1 Phe occured but the arrangement of B1 Phe, A13 Met and A14 Tyr (see later) remained essentially the same. Transient structures written at 15 ps; 55 ps; 81 ps and 97 ps were chosen and each were energy minimised *in vacuo*. Structures minimised at 55 ps; 81 ps and 97 ps superimposed well and remained essentially the same.

Comparison of the structure of the molecular model with that of pig insulin reveals it to be different. The A2–8 α-helix seems more distorted and the B21–23 turn more irregular. The C-terminal B28–B30 end of the B-chain, which is known to be flexible, folds slightly back towards the main body of the monomer. A two-fold axis view reveals residues A13, A14, A17, B1 and B2 (part of the dimer-dimer interface) to be involved in a different spatial arrangement from pig insulin (Figure 3). The non polar nature has been preserved in the two amino acids (A13 Met and

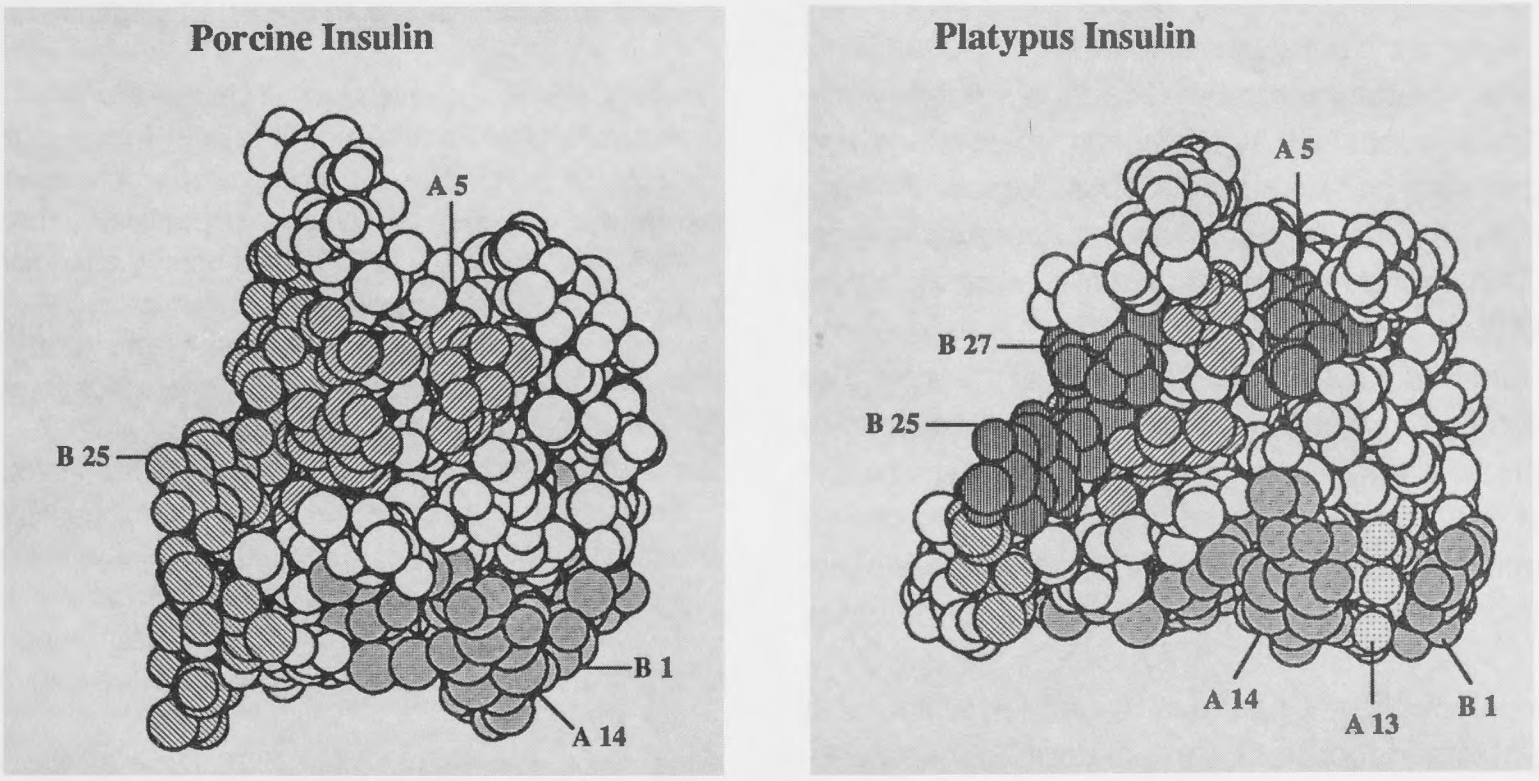


Fig. 3 Space-Filling Diagrams of Modelled Platypus Insulin Monomer 2 and Porcine Insulin Monomer 2 Viewed in the Direction of the Local Two-Fold Axis.
Receptor binding region (diagonal right to left line shading), monomer-monomer interface (diagonal left to right line shading), dimer-dimer interface (shaded), A13 (speckled); platypus insulin A5, B25, and B27 (vertical line shading). Platypus monomer insulin structure was modelled on the X-ray crystal structure of rhombohedral 2 Zn porcine insulin, monomer 2, entry 3 ins (Wlodawer *et al.*, 1989) of the Brookhaven Protein Databank. The pig insulin monomer 2 structure was subjected to 250 iterations of steepest decent energy minimisation followed by 5000 iterations of conjugate gradient *in vacuo* and was used for modelling the structure of platypus monomer 2 insulin with the appropriate different residues. The platypus monomer 2 insulin has been subjected to 250 iterations of steepest descent energy minimisation followed by 5000 iterations of conjugate gradient energy minimisation *in vacuo*. Molecular dynamics of 100 ps of simulation at 200 K have subsequently been performed *in vacuo*. Analysis was performed using the transient structures written at intervals of 1 ps. Computrations for molecular dynamics were carried out using the Fujitsu VP of the Australian National University Supercomputer Facility. The number of iterations were 100 000 with a time step of 1 fsec. A history file was written every 1000 steps. Transient structures written at 15 ps; 55 ps; 81 ps and 97 ps were chosen and each was energy minimised by using the conjugate gradient algorithm for 5000 steps *in vacuo*. In all cases the maximum derivative was below 0.001 kcal/mol-Å. Both the energy minimisations and simulations were carried out using the program Discover Version 94.0 (May 1994) on a Silicon Graphics workstation as well as the analysis using the program Analysis, all generated by the computer program Insight II version 2.3.0 (December 1993).

B2 Pro) of platypus insulin that differ from pig insulin in this interface. In the porcine insulin crystal there exists an A14 Tyr-B1 Phe interaction in the dimer-dimer interface (Baker *et al.*, 1988; Blundell *et al.*, 1972), although in 2D-NMR studies of insulin monomers in solution these residues seem to be flexible and the interaction unlikely to be maintained in solution (Weiss *et al.*, 1989). In the platypus insulin model A13 Met is sandwiched between A14 and B1 and the B1 Phe is pushed away from the body of the monomer (Figure 3). Such structural differences should be interpreted with caution, thus a possible altered dimer-dimer interaction and thus altered self-association properties of platypus insulin needs to be evaluated experimentally under solution conditions when sufficient material is available.

Viewing the space filling models of both platypus and pig insulin, by looking directly at the receptor binding region, revealed no marked difference. As viewed from the two-fold axis pig insulin is mainly heart shaped while platypus insulin has a squarer structure. The invariant amino acids responsible for the stable core structure of insulin (Baker *et al.*, 1988) are identical in both species.

In conclusion, platypus insulin's self-association behaviour under physiological conditions needs to be evaluated. The influence of the changes in amino acid composition on biological activity and receptor binding characteristics of platypus insulin also require experimental study. Platypus insulin possesses a unique evolutionary-designed primary structure. The study of the ramifications of this, especially its molecular interactions and physico-chemical behaviour in its physiological environment, namely in solution, blood, and in contacts with residues of its receptor upon binding to elicit a biological response, should contribute to our understanding of the structure-function relationship of insulin in general. The combination of knowledge gained from the resources of evolutionary designed and protein-engineered insulins will in turn contribute to a more informed insulin design and evaluation strategy in the quest for a monomeric, active, stable and immunologically safe insulin for diabetes therapy.

To date only sufficient platypus insulin for amino acid sequence determination has been isolated. Chemical peptide synthesis, based on the known sequence, is currently under way to obtain sufficient material to conduct the studies referred to.

Acknowledgements

We thank Prof. Frank Gibson for help with the molecular modelling studies, Dr. Mervyn Griffiths for the donation of the platypus pancreases and Dr. Peter J. Milburn for helpful discussions.

References

Baker, E.N., Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.M., Hubbard, R.E., Isaacs, N.W., Reynolds, C.D., Sakabe, K., Sakabe, N., and Vijayan, N.M. (1988). The structure of 2 Zn pig insulin crystals at 1.5 Å resolution. Philos. Trans. R. Soc. Lond. Biol. Sci. *B319*, 369–456.

Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., and Mercola, D. (1976). Structure of insulin in 4-zinc insulin. Nature *261*, 166–168.

Brange, J., Ribel, U., Hansen, J.F., Dodson, G.,Hansen, M.T., Havelund, S., Melberg, S.G., Norris, F., Norris, K., Snel, L., Sørenson, A.R., and Voigt, H.O. (1988). Monomeric insulins obtained by protein engineering and their medical implications. Nature *333*, 679–682.

Brems, D.N., Alter, L.A., Beckage, M.J., Chance, R.E., DiMarchi, R.D., Green, L.K., Long, H.B., Pekar, A.H., Shields, J.E., and Frank, B.H. (1992). Altering the association properties of insulin by amino acid replacement. Protein. Eng. *5*, 527–533.

Bringer, J., Heldt, A., and Grodsky, G.M. (1981). Prevention of insulin aggregation by dicarboxylic amino acids during prolonged infusion. Diabetes *30*, 83–85.

Blundell, T., Dodson, G., Hodgkin, D., and Mercola, D. (1972). Insulin: The structure in the crystal and its reflection in chemistry and biology. Adv. in Prot. Chem. *26*, 279–402.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic. Acids. Res. *12*, 387–395.

Dische, F.E., Wernstedt, C., Westermark, G.T., Westermark, P., Pepys, M.B., Rennie, J.A., Gilbey, S.G., and Watkins, P.J. (1988). Insulin as an amyloid-fibril protein at sites of repeated insulin injections in a diabetic patient. Diabetologia *31*, 158–161.

Feng, D.F., and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. *25*, 351–360.

Grant, T. (1989). The Platypus. A Unique Mammal. (Kensington Australia: New South Wales University Press), pp. 1–14.

Griffiths, M. (1988). The platypus. Scientific American *258*, 60–67.

Horuk, R., Goodwn, P., O'Connor, K., Neville, R.W.J., Lazarus, N.R., and Stone, D. (1979). Evolutionary change in the insulin receptors of hystricomorph rodents. Nature *279*, 439–440.

Jeffrey, P.D. (1986). Self-association of des-(B26–B30)-insulin. The effect of Ca^{2+} and some other divalent cations. Biol. Chem. Hoppe-Seyler *367*, 363–369.

Jeffrey, P.D., Milthorpe, B.K., and Nichol, L.W. (1976). Polymerization pattern of insulin at pH 7.0. Biochemistry *15*, 4660–4665.

Kaarsholm, N.C., Norris, K., Jørgensen, R.J., Mikkelsen, J., Ludvigsen, S., Olsen, O.H., Sorensen, A.R., and Havelund, S. (1993). Engineering stability of the insulin monomer fold with application to structure-activity relationships. Biochemistry *32*, 10773–10778.

King, G.L., and Kahn, C.R. (1981). Non-parallel evolution of metabolic and growth-promoting functions of insulin. Nature *292*, 644–646.

Nakagawa, S.H., and Tager, H.S. (1986). Role of the phenylalanine B25 side chain in directing insulin interaction with its receptor. Steric and conformational effects. J. Biol. Chem. *261*, 7332–7341.

Needleman, S.B., and Wunch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453.

Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA *85*, 2444–2448.

Pocker, Y., and Biswas, S.B. (1981). Self-association of insulin and the role of hydrophobic bonding: a thermodynamic model of insulin dimerization. Biochemistry *20*, 4354–4361.

Rinderknecht, E., and Humbel, R.E. (1978). The amino acid sequence of human insulin like growth factor I and its structural

homology with proinsulin. J. Biol. Chem. *253*, 2769–2776.

Schäffer, L. (1994). A model for insulin binding to the insulin receptor. Eur. J. Biochem. *221*, 1127–1132.

Seino, S., Steiner, D.F., and Bell, G.I. (1987). Sequence of a New World primate insulin having low biological potency and immunoreactivity. Proc. Natl. Acad. Sci. USA *84*, 7423–7427.

Shoelson, S.E., Lu, Z.X., Parlautan, L., Lynch, C.S., and Weiss, M.A. (1992). Mutations at the dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. Biochemistry *31*, 1757–1767.

Treacy, G.B., Shaw, D.C., Griffiths, M.E., and Jeffrey, P.D. Purification of a marsupial insulin: amino-acid sequence of insulin from the eastern grey kangaroo *Macropus giganteus.* (1989). Biochim. Biophys. Acta. *990*, 263–268.

Wlodawer, A., Savage, H., and Dodson, G. (1989). Structure of insulin: results of joint neutron and X-ray refinement. Acta. Crystallogr. *B45*, 99–107.

Weiss, M.A., Hua, Q.X., Lynch, C.S., Frank, B.H., and Shoelson, S.E. (1991). Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective ^{2}H and ^{13}C labeling with application to protein design. Biochemistry *30*, 7373–7389.

Weiss, M.A., Nguyen, D.T., Khait, I., Inouye, K., Frank, B.H., Beckage, M., O'Shea, E., Shoelson, S.E., Karplus, M., and Neuringer, L.J. (1989). Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. Biochemistry *28*, 9855–9873.

Received August 17, 1995; accepted October 19, 1995