INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-1

SMR/96 - 8

SUMMER COLLEGE IN BIOPHYSICS

(2 - 27 August 1982)

LECTURE 3 : Refinement of Protein Crystal Structures.

LECTURE 4 : The Biosynthesis and Evolution of Insulin and Related Molecules.

G.G. DODSON

Department of Chemistry
University of York
Heslington
York YO1 5DD
U.K.

LECTURE 3

Refinement of Protein Crystal Structures

Introduction

In these notes I will outline the principles of the procedures used in the refinement of protein crystal structures and will discuss how refinement has contributed to our picture of protein structure and function.

The essential approach to refining a protein structure is the same as that for small molecules. Thus the atomic parameters are usually refined by least squares minimisation of the expression

$$\sum_{hkl} w_{hkl} \left(|F_{o,hkl}| - |F_{c,hkl}| \right)^2$$

where w_{hkl} is a weighting factor, |Fc| & |Fo| are the observed and calculated structure amplitudes. The corrections to the parameters are derived from the matrix equation

$$\Delta \underline{X} = H^{-1} \underline{V} \quad \text{where V is the gradient vector}$$

$$V_i = \sum_{hkl} w_{hkl} \Delta F_{hkl} \, \partial F_{c,hkl} / \partial p_i$$

& where H is the normal matrix

$$H_{ij} = \sum_{hkl} w_{hkl} \frac{\partial F_{c,hkl}}{\partial p_i} \cdot \frac{\partial F_{c,hkl}}{\partial p_j}$$

The main problems in refining protein crystals are

(1) the large number of parameters (these typically lie between 4 - 40,000) which presents formidable problems in computation.

(2) the limited resolution of the diffraction data (|Fo|). Only occasionally do the data extend to beyond 1.5Å spacing and this means the ratio of parameters to observations is unfavourable. Typically this ratio is about 2-4 observations per atomic parameter which limits the power of refinement methods.

All the procedures currently used in protein refinement are designed to alleviate these limitations

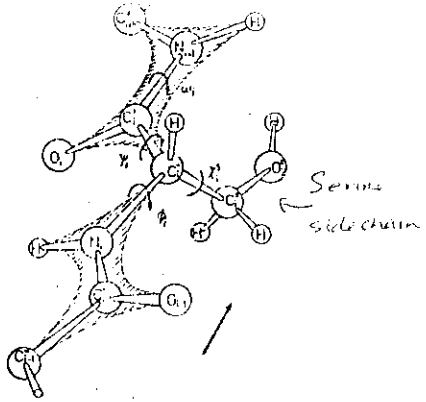
Calculation Methods

Treatment of the individual atoms separately by full matrix refinement is too large a problem even now. Other refinement procedures based on difference Fourier methods and real space calculations have been developed successfully for large molecules. In the last 5 years or so reciprocal space least squares minimisation techNiques have superceeded these largely owing to the rapid calculation of parameter shifts and methods for effectively improving the rates of observations to parameters.

Initial Stages

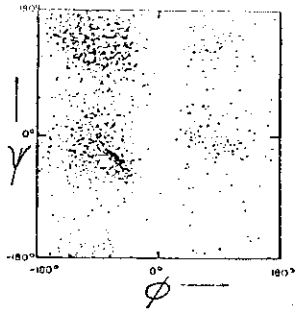
Before a protein can be refined it must be interpreted and a set of initial atomic positions exist. The coordinates can be surprisingly inaccurate - a typical rms shift in refinement for initial to final coordinates is .6Å. Thus automatic procedures are not applicable, and particularly at the early stages the structure needs to be reviewed carefully and systematically in Fourier maps during the refinement.

Secondly the atoms in proteins are relatively mobile with rms displacements from their mean position ranging from .1 to 1Å. The less ordered atoms scatter very weakly and in the Fourier maps appear as poorly resolved features. Positioning these atoms accurately is a major difficulty - and in a protein as much as a quarter of the atoms can have rms displacements of > .5Å. The solvent molecules, which usually constitute 30-50% of the cell are mostly even less well ordered than the protein and often extend into what is essentially a network of channels which contain many alternative solvent configurations.



'EPTIDE

Figure 1(a)



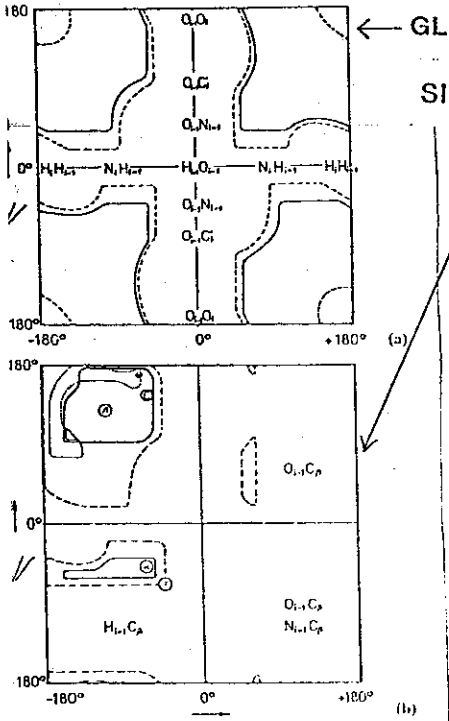
Ψ AND φ ANGLES

OF 13 PROTEINS

(UNREFINED)

Figure 1(b)

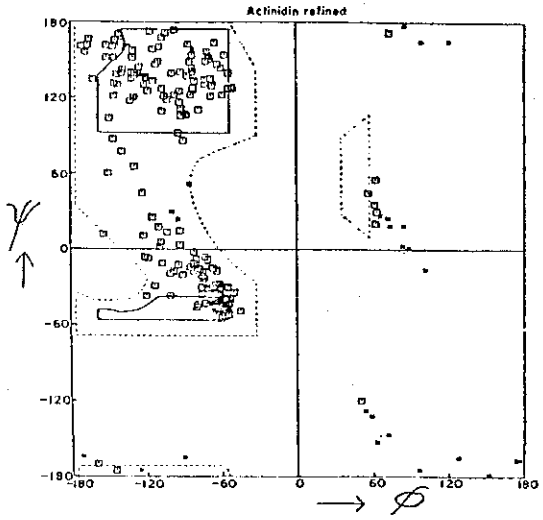
Figure 1(c)



← GLYCINE

SIDE CHAIN AMINO ACIDS

Figure 1(d)



Ψ AND φ ANGLES

OF ACTINIDIN (REFINED)

RAMACHANDRAN

Constrained Refinement

The principle in this method is to reduce the number of atomic parameters by constraining certain atomic groups to defined geometry. Thus benzene rings and peptide bonds for example are treated as rigid groups. Under these conditions the angular rotation about the bonds linking these groups are refined. This reduces the number of parameters by about an order of magnitude, as illustrated for the amino acid residue

serine.

(Figure 1(a))

a (Figure 1(c))

The size of the rigid group chosen can vary according to the data available and the character of the structure itself. It would be possible to refine the linking bonds between rigidly defined helical structures - a procedure followed in tRNA studies.

A limitation in rigid body refinement is that the geometry is not exactly known and that certain angles can vary by as much as 10°. Thus there is either considerable inaccuracy in the placing of the refined structure or special allowance for points of angular flexibility has to be made.

Restrained Refinement

An alternative philosophy is to increase the number of observations by including the chemically defined interatomic distances and angles in the normal equations. By this device the individual atomic parameters can be reliably refined

by least squares minimisation.

As with constrained refinement the restraints can be selected by the investigator in ways appropriate for the protein under study. It is for example possible to include not only the chemically invariant parameters such as the benzene and peptide bond ring distances and angles but also less exactly defined parameters produced by H bonding and non bonded contact. This flexibility is most useful and when used with judgement will hasten convergence.

The principal procedures for this method of refinement have been established by Konnert, Isaacs and Aganone (Agarwal), Moss and Levitt and Jack. There is little real (practical) difference in these approaches and while only few comparisons have been made they indicate that the controlling factor in the refinement is the investigators judgement!

The main uncertainty in restrained refinement is how to weight the two classes of observation (the X-ray structure amplitudes and the target bond lengths) together. Various empirical paths are followed. Commonly the weight of the X-ray terms are reduced at the early stages and steadily increased as convergence is approached. This makes allowance for the initial inaccuracy of the atomic positions and the necessarily inaccurate shift calculated from the electron density at this stage.

Another important factor is the tightness of the restraints. The looser these conditions are on the atomic coordinates the better the agreement with the X-ray observations. But of course the protein structure will conform less well to proper stereochemistry.

The progress of a typical constrained refinement (or tRNA) and a restrained refinement is followed by the reduction in the agreement factor $R = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$

Typically the initial value for R is 50% and this will fall to a value between 15 - 25%. The R value at convergence is a function of the data quality, its resolution and the tightness of the restraints.

Correcting and Extending the Protein Structure

The initial coordinates usually contain serious errors and will (might) always even include all the protein atoms and will invariably exclude most if not all the solvent atoms owing to the deficiencies in the original isomorphously phased map. Phases calculated from the initial coordinates are often an improvement on the isomorphous phases. With refinement the phases are further improved. This makes possible the calculation

$$\Delta\rho = \frac{1}{V} \sum_{hkl} ||F_o| - |F_c|| e^{i\alpha_c} e^{-2\pi i(hx + ky + lz)}$$

where $|F_o|$ = observed structure amplitude

$|F_c|$ = calculation " "

$\Delta\rho$ = difference fourier electron density

& α_c = the phase angle

From these calculations the atoms excluded from the phasing can be detected. Generally the best defined atoms with highest electron density are detected first, more disordered atoms only being identified as the refinement converges, and the phases surface and the errors in the electron density are reduced. Wrongly placed atoms can be detected by their behaviour in refinement, their apparent thermal motion increases continuously and there are indications too in the electron density in the difference Fourier as well.

The rate of convergence and the extent to which the poorly defined protein and solvent atoms can be described is largely determined by the quality and resolution of the data.

How refinement has contributed to our picture of proteins.

(1) Accurate coordinates have obviously extended the detail in the description of proteins. This is nicely illustrated in the Ramachandian plot of a refined protein's dihedral angles (ϕ & ψ) which fall precisely into the limits defined by internal geometry of the peptide. In contrast undefined proteins exhibit a very wide spread of the dihedral angles and these often populate energetically unfavourable regions. (Figures (a), (b))

The conformational behaviour of the polypeptide chain is now much better understood; the simple picture of a helix and β sheet for example is complicated by the finding that helical structures are often intermediate between α & 310, that β sheets have an inherent twist and so on.

Also the estimates of error that are obtained in refinement are a valuable [illegible] of the [illegible] variation in the accuracy of the atomic positions in different parts of the molecule

(2) Refinement of protein structures provides a sensible measure of apparent atomic motion through the magnitude of the thermal parameter B. The two components of this parameter are static disorder and real vibrations. It is not easy to separate these two effects.

The atomic thermal movement also contains two components, one from the motion of the molecule and the second from the motion of the atom itself. It is in principle possible to analyse the molecular vibrations and separate the two motions but there are considerable difficulties. Study of lysozyne has revealed evidence of overall molecular motion. More interesting perhaps however is the discovery that the molecule contains regions of unusually large atomic motion around the active site. This finding suggests that, at least in the case of lysozyme, the active site is more flexible and reflects its need to move rapidly through a series of conformational changes during catalysis. (Artymiuk et al., Nature, 280, 563, 1979)

(3) Accurate coordinates have obviously improved the description of protein interactions. One important example is the analysis and refinement of enzymes. The stereochemistry in the catalytic site, including the water molecules, is now known in some enzymes to an accuracy of .05 - .1Å in the atomic positions.

The symmetry in the 2 Zn insulin hexamer has also been clarified by the refinement of the crystal structure. The two independent subunits within the hexamer (which constitutes the unit cell) are almost exactly two fold related about a local axis : this symmetry breaks down further from the axis , largely owing to crystal contacts.

4. With high resolution data (d < 1.2Å) it is possible to locate someof the hydrogen atoms, bringing much more detail and accuracy into our picture of the hydrogen bonding in protein.

5. Refinement calculations have enabled us to exploit the structural similarities between related protein's very effectively. There are procedures, known as molecular replacement, by which an unknown crystal structure of a protein with known chain folding can be solved approximately. The crystal structure thus obtained is of course often very inaccurate and it is essential to correct it by refinement procedures.

LECTURE 4

The Biosynthesis and Evolution of Insulin and Related Molecules

Introduction

Insulin is an important metabolic hormone whose most well documented effects are the regulation of glucose levels in the blood and enhancement protein, nucleic acid and fatty acid synthesis. The hormone exerts its effects via interaction with a membrane bond receptor on the cell surface. In this lecture I will discuss the chemical and cellular events in insulin biosynthesis, the sequence and structural relationships between insulin and its relations and finally the fundamental insights given by the structure and organisation of the insulin gene.

Biosynthesis of Insulin

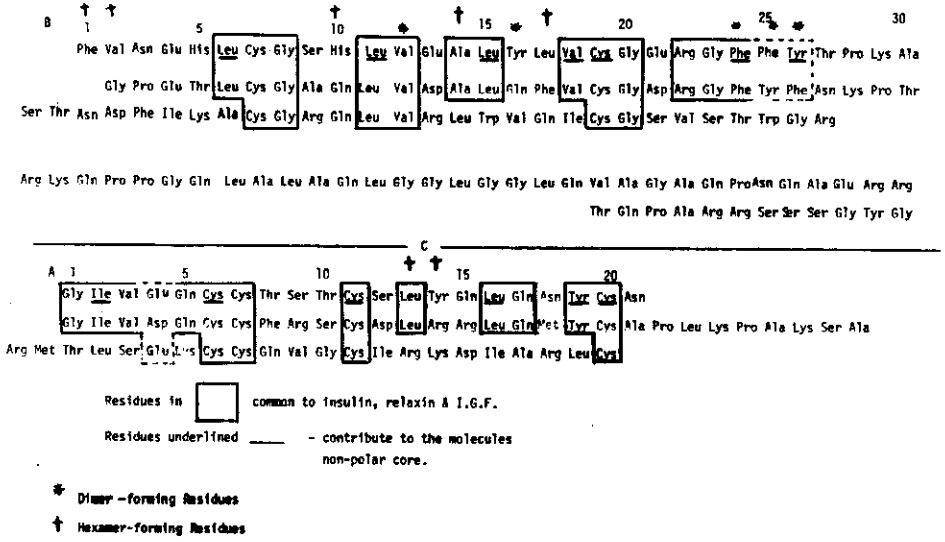
The accompanying paper fills out the details of the process which has evolved to release insulin at rates independent of its biosynthesis. The important features in insulins biosynthesis are:

(1) The correspondence between the chain structure and the chemical and physical demands in biosynthesis (i.e.) a long high chain is necessary for export from the ribosome; its cleavage to single chain proinsulin which can fold spontaneously.

(2) The incorporation of Zn^{2+} ions. These provide a means for the insulin dimer to aggregate to a hexamer and amplify the prohormone's solubility and the hormones insolubility. There are also possible roles for Zn^{2+} (and Ca^{2+}) after release of the insulin.

Table I

Sequences of Pig Proinsulin (Sanger et al., Steiner et. al.)
Human I.G.F. (Humbel et al.)
Pig Relaxin (James et. al., Schwabe et al.)

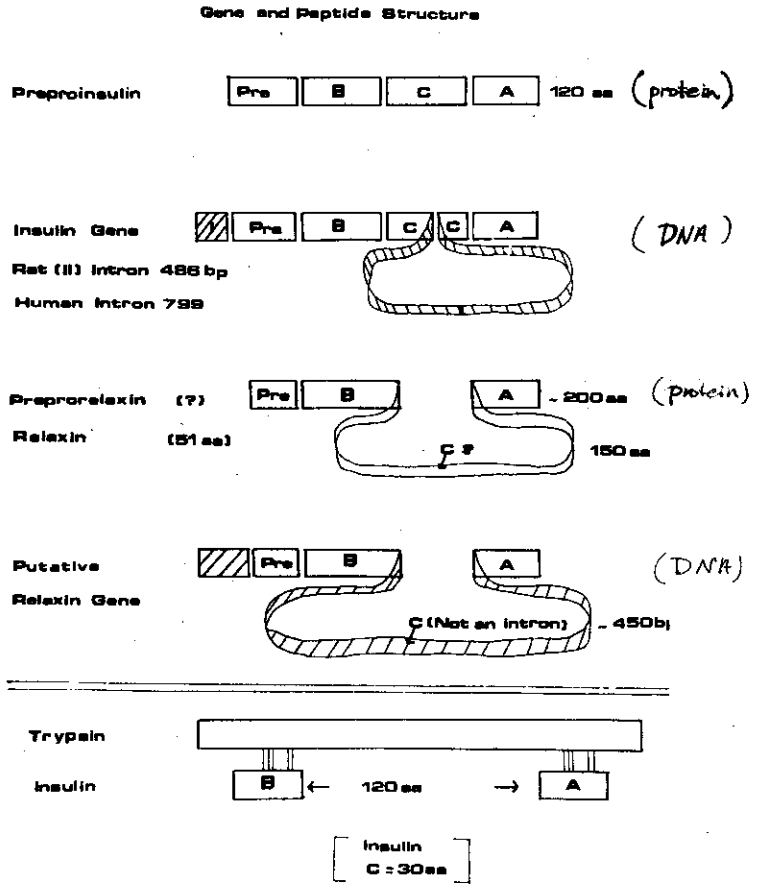


The Sequence and Structural Relationships

The sequences of the two molecules related to insulin. have been established. The first of these is relaxin, produced in the corpus luteum and which by relaxing the pelvic girdle facilitates birth. The second, insulin like growth factor, is synthesized probably in the liver and is a general growth factor.

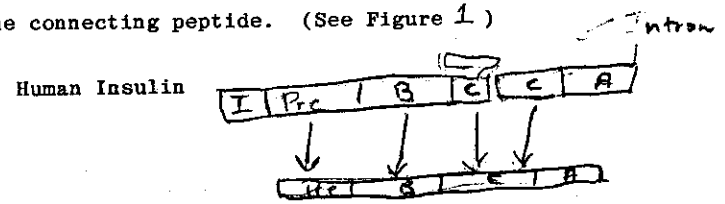
Table I shows the sequences for human pro insulin, the insulin like growth factors (IGF) and for the 3 known relaxins. The extent of the homologies between the insulin and the IGFs is striking - clearly these molecules are related and have similar folding.

The homology between insulin and the relaxins is much more limited - only the structurally important disulphide bonds and the glycines are preserved. The idea that these few residues do imply an "insulin" fold is confirmed by the observation that all the residues that in relaxin would be buried in the molecules non-polar core by insulin's folding are indeed now polar. The attached paper (Isaacs et al., 1977) gives a detailed account of the structural and sequence arguments.



Insulin's Gene Structure and its Evolutionary Significance

The DNA sequence of the insulins gene from the human, rat and other animals have been determined. (Bell et al., Nature, 284, 26, 1980). These studies show that the gene contains two segments which are not translated into proteins. These stretches of short DNA are called introns and interestingly in insulin one of these occurs in the middle of the segment corresponding to the connecting peptide. (See Figure 1)



The persistence of this large intron at precisely the same position in the C peptide region in all known insulin genes is at present a mystery. However it does give us an insight into one of the striking differences between proinsulin (with a C peptide of 30 amino acids) and prorelaxin (with a C peptide of 105 amino acids). The extra length in prorelaxins C peptide we can speculate arises from the expression of gene structure that in insulin is silent. The gene structure of relaxin is as yet unknown but when it is determined it will probably give us further insight into the relationships between insulin and relaxin.

Trypsin like semi proteases and ProInsulin.

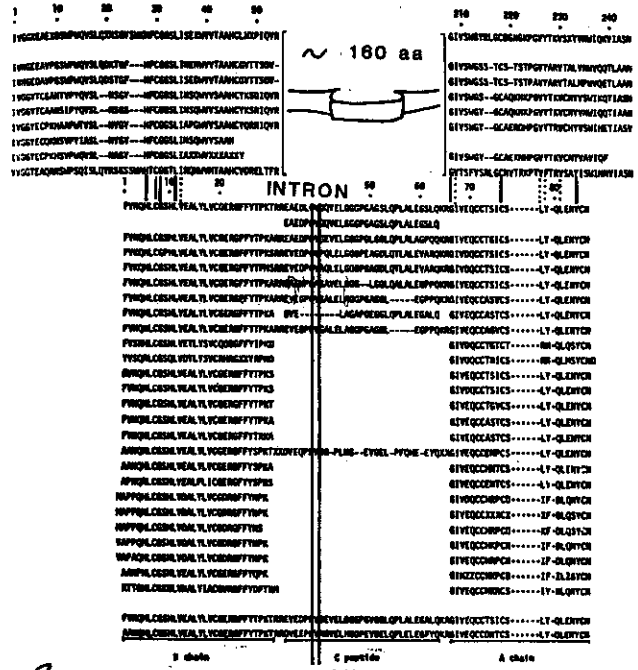
The amino acid homology between proinsulin and trypsin has been noted by several writers and has been exhaustively examined by de Haen, ~~(1976)~~ 1976 *. The main objection to the homology being real was that the homologous between the insulin B chain and A chain was separated in pro-insulin by some 40 residues ~~which~~ while in the

* de Haen. et al., J. Mol. Biol., 106, 639, 1976

enzyme it was separated by 200 residues. This discrepancy however

Figure 2

does not exist at the genetic level where the intron in the insulin gene corresponds to the structure between the distant homologous regions in trypsin.



Trypsin family.

Proinsulin family.

Figure 2

(Reprinted from Nature, Vol. 271, No. 5642, pp. 278–281, January 19, 1978)

Relaxin and its structural relationship to insulin

THE sequence of the ovarian peptide, relaxin, has recently been reported[1,2] and the observation[1] that it can be accommodated into the insulin fold has been discussed[3,4]. Only 11 residues, including the cystines, are common to insulin and relaxin (Fig. 1), but the probable identity of the cystine pairings and the preservation of the hydrophobic character of buried residues suggests some structural homology between the two hormones. The extent of sequence changes and the remote relationship between the corpus luteum and the pancreas, the respective sources of relaxin and insulin, makes homology of the two hormones most interesting. We have therefore rigorously examined the relaxin conformation by a computer graphics system and found it possible to accommodate the relaxin sequence within the insulin main-chain geometry.

We first considered the packing of residues whose equivalents in insulin are buried or partly buried. We chose molecule

a

A Chain

	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Relaxin	Arg	Met	Thr	Leu	Ser	Glu	Lys	Cys	Cys	Gln	Val	Gly	Cys	Ile	Arg	Lys	Asp	Ile	Ala	Arg	Leu	Cys	
Insulin			Gly	Ile	Val	Glu	Gln	Cys	Cys	Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu	Glu	Asn	Tyr	Cys	Asn

B Chain

	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Relaxin	PCA	Ser	Thr	Asn	Asp	Phe	Ile	Lys	Ala	Cys	Gly	Arg	Glu	Leu	Val	Arg	Leu	Trp	Val	Glu	Ile	Cys	Gly	Ser	Val	Ser	Thr
Insulin						Phe	Val	Asn	Gln	His	Leu	Cys	Gly	Ser	His	Leu	Val	Glu	Ala	Leu	Tyr	Leu	Val	Cys	Gly	Glu	Arg	Gly	Phe

Invariant in insulins with high potency

Common residue in insulin & relaxin → Relaxin: 30 29 28 27 26 25 = Arg Gly Trp

Buried or partly buried residues → Insulin: Ala Lys Pro Thr Tyr Phe

† Invariant as an aromatic residue

* Totally invariant residue

} in insulin

b

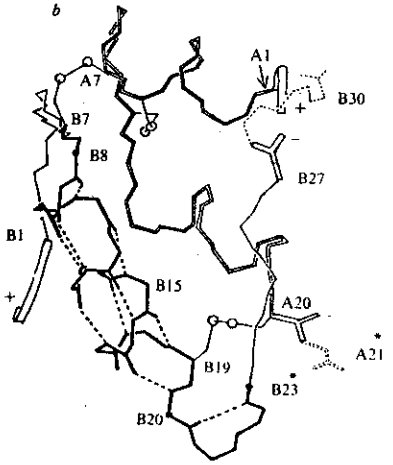


Fig. 1 *a*, The amino acid sequences of insulin and relaxin; *b*, the insulin main chains viewed from a direction that illustrates the relative positioning of the central B-chain α helix and the two roughly antiparallel A-chain helical stretches. The hydrogen bonds in the well defined B-chain α helix and subsequent turn are dotted; we expect this structural element to be least affected by the sequence changes in relaxin. *, Residues which seem to be crucial to the integrity of the insulin structure and which differ in relaxin. The extensions at the N termini in relaxin relative to insulin are represented as ribbons, the deletions at the C termini by dashed lines.

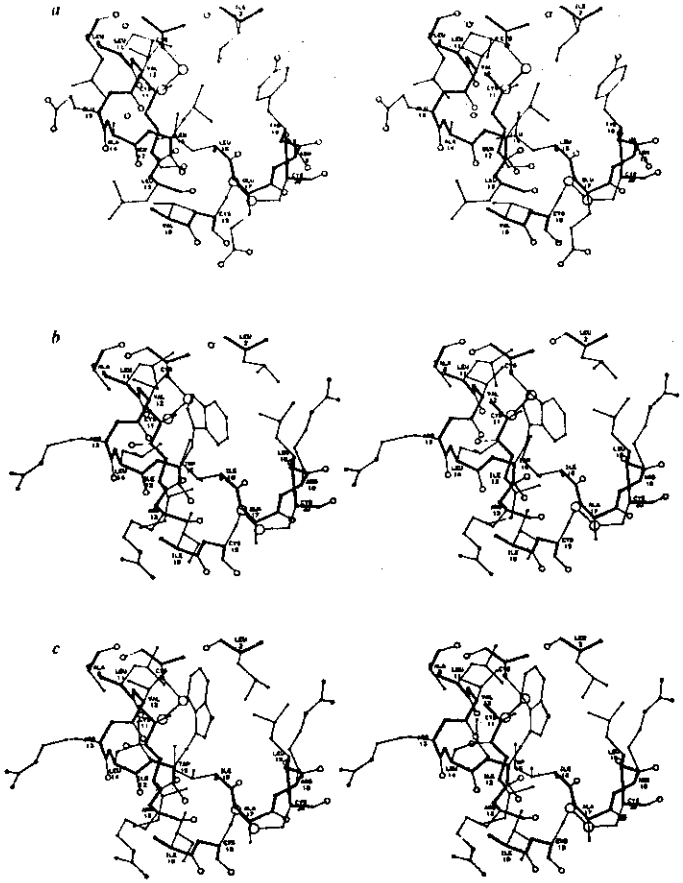


Fig. 2 Buried residues in the central core: *a*, 2-zinc insulin; *b* and *c*, alternative arrangements of Trp B_{15} in relaxin. The A chain is shown by open bonds and the B-chain by solid bonds.

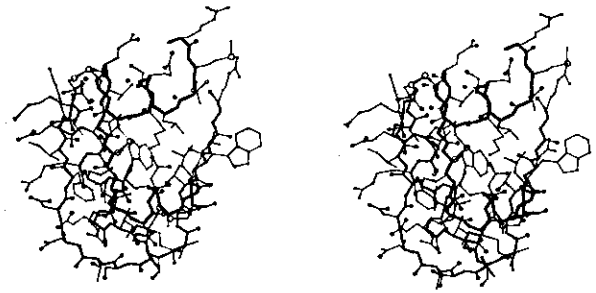


Fig. 3 A stereo view of the complete relaxin molecule with B_{15} as in Fig. 2*b*.

2 of the 2-zinc insulin[5] structure for the comparison since the conformations in this monomer are most closely related to those of 4-zinc pig insulin[6] and the dimeric hagfish insulin (Cutfield, J. F. *et al.*, in preparation). The atomic positions of this insulin molecule are now accurately known and for the buried residues the estimated errors are 0.1–0.2Å. Using Labquip models, we found that the internal hydrophobic residues of insulin could be replaced by the equivalent relaxin residues without the apparent need to adjust the main-chain conformation. In particular, there was an extensive cavity into which tryptophan B_{15} seemed to fit nicely. Because of the limitations and inaccuracies in model building which we consider might be misleading, we then used a computer graphics system[7,8] to superimpose the relaxin chains on to the insulin backbone. Any short contacts were indicated immediately by dotted lines between the offending atoms. Side-chain conformation swere adjusted by rotations about single bonds until acceptable stereochemistry and packing was achieved. We found that tryptophan B_{15} can indeed be directed into the relaxin core with no contacts unacceptable in relation to the precision of the insulin coordinates. This and the equivalent insulin structure are shown in Figs 2*b* and 2*a*.

Some alternative arrangements of buried side-chains are possible without adjustment of the main chain. An advantage of the computer graphics system is that such alternatives can be rapidly explored. In particular, we found that tryptophan B_{15} can be orientated in a second distinct way (Fig. 2*c*) related principally by 180° rotation about the C_β–C_γ bond, and corresponding essentially to that reported by Bedarker *et al.*[4]. It gives a rather more open hydrophobic core than our preferred alternative of Fig. 2*b*.

We built the terminal stretches of chain beyond the disulphide bridges as they occur in insulin. The additional residues at the N termini were built in arbitrary but sensible conformations giving reasonable contacts to the main body of the molecule. There were two difficulties at the B-chain C-terminus. The sequence reported by Schwabe *et al.*[9] differs from that of James *et al.*[1] from B_{25} onwards, possibly because of microheterogeneity. (This does not affect the central region of the molecule.) We have followed the sequence of James *et al.* and for simplicity identify relaxin residues by their position in insulin. In insulin, B_{23} is glycine and its substitution by an L-amino acid (serine in the sequences of James *et al.*, valine in that of Schwabe *et al.*) would give impossible contacts, so there must be a conformational difference here between relaxin and insulin. We continued beyond B_{22} in a way which left the residues near the positions of their equivalents in insulin. In contrast to B_{23}, the glycines at B_8 and B_{20} are preserved in relaxin.

The main features of our proposed model are: (1) of the two tryptophans, B_{13} is buried while B_{25} is exposed. This fits the chemical and fluorescence observations of Schwabe and Braddon[9], who found that one tryptophan reacted readily without altering the hormone's potency, while reaction of the second was very difficult and inactivated the hormone; (2) there are compensating changes in side chains that are spatially related, such as B_9 and B_{14} and A_2 and A_{16}, which are consistent with insulin and relaxin having the same internal organisation[11] (see Fig. 1*a*); (3) there is complementary distribution of the many charged side chains, particularly noticeable on the helical stretches. This not only supports the proposed folding but also suggests that the relaxin molecule achieves stability as a monomer through these favourable interactions[10]; (4) detailed changes near B_{23} modify the characteristics of the monomer surface involved in dimer formation in insulin; (5) residues B_9 to B_{19} fit well in an unbroken α helix, so that the improvement in sequence homology achieved by displacing the two sequences at B_{11} to B_{17} (ref. 1) seems to have no significance in structural terms.

We consider that relaxin has a three-dimensional organisation similar to that of insulin. We do not yet know the closeness of this relationship but note that insulin itself shows considerable structural variation[6]. Thus, while the B-chain α helix is closely preserved, we find widely different conformations of the terminal residues and some alterations of the A-chain helical structure. Therefore, we might expect greatest similarity between relaxin and insulin in the α-helical B-chain and increasing differences in the A chain and the chain termini.

Amongst the many sequence differences between insulin and relaxin, there are two of particular relevance to function. The first is the absence of residue A_{21} in relaxin. When A_{21} asparagine is removed from insulin so that, like relaxin, its A-chain terminates at A_{20} cystine, the hormone's solution and spectral properties are profoundly altered[12], its A-chain is apparently perturbed[13] and its biological potency is much reduced[10,11]. Second, the substitution of B_{23} glycine by an L-amino acid has been shown by the Chinese studies effectively to abolish insulin's biological activity[14]. In diverging from insulin at A_{21} and B_{23}, relaxin therefore will be without the capacity to behave like insulin. (We note that in the insulin-like peptide responsible for the non-suppressible insulin-like activity in the serum (insulin-like growth factor), B_{23} is glycine and A_{21} is alanine, though this is not the C-terminal amino acid[15]).

X-ray analysis of relaxin and other hormones with recognised homology to insulin[17] could clarify the relationships between their sequences, structures and biological behaviour.

We thank Professor T. L. Blundell for showing us the paper by Bedarkar *et al.* before publication. The coordinates of our model have been deposited with the Protein Data Bank.

NEIL ISAACS
ROBERT JAMES
HUGH NIALL
Department of Biochemistry,
Howard Florey Institute of Experimental Physiology and Medicine,
University of Melbourne,
Victoria, Australia 3052

GILLIAN BRYANT-GREENWOOD
Department of Anatomy and Reproductive Biology,
University of Hawaii,
Honolulu, Hawaii

GUY DODSON
Chemistry Department,
University of York,
York, UK

ALAN EVANS
ANTHONY C. T. NORTH
Astbury Department of Biophysics,
University of Leeds,
Leeds, UK

Received 19 September; accepted 23 November 1977.

1. James, R., Niall, H. D., Kwok, S. & Bryant-Greenwood, G. D. *Nature* 267, 544–546 (1977).
2. Schwabe, C. J., McDonald, K. & Steinetz, B. C. *Biochem. biophys. Res.Comm.* 70, 397–405 (1976); 75, 503–510 (1977).
3. Schwabe, C. J. & McDonald, K. J. *Science* 197, 914–915 (1977).
4. Bedarkar, S., Turnell, W. G., Blundell, T. L. & Schwabe, C. J. *Nature* (in the press).
5. Blundell, T. L. *et al. Nature* 231, 506–511 (1971).
6. Bentley, G., Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. *Nature* 261, 166–168 (1976).
7. Barry, C. D. & North, A. C. T. *Cold Spring Harb. Symp. quant. Biol.* 36, 577–584 (1971).
8. Denson, A. K., North, A. C. T. & Willoughby, T. V. *Acta crystallogr.* A31, 1279 (1975).
9. Schwabe, C. J. & Braddon, E. N. *Biochem. biophys. Res. Comm.* 68, 1126–1132 (1976).
10. Perutz, M. F. & Raidt, H. *Nature* 255, 256–259 (1975).
11. Carpenter, F. H. *Am. J. Med.* 40, 750–756 (1966).
12. Arquilla, L. K., Dorio, R. J. & Brogman, T. M. *Diabetes* 25, No. 5, 397–403 (1976).
13. Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. *Adv. Protein Chem.* 26, 280–402 (1972).
14. Zimmerman, A. L., Moule, M. L. & Yip, C. C. *J. biol. Chem.* 249, 4025–4029 (1974).
15. Insulin Group *Scientia sin.* XVII, No. 6, 779–792 (1974).
16. Rinderknecht, E. & Humbel, R. E., reported at the 11th FEBS Meeting, Copenhagen 1977.
17. Rinderknecht, E. & Humbel, R. E. *Proc. natn. Acad. Sci. U.S.A.* 73, 4379–4381 (1976).

Diabetologia 19, 174–182 (1980)

Diabetologia

Hypothesis

Role of Zinc in Insulin Biosynthesis

Some Possible Zinc-Insulin Interactions in the Pancreatic B-cell

S. O. Emdin[1], G. G. Dodson[2], J. M. Cutfield[3], and S. M. Cutfield[3]

[1]Department of Pathology, University of Umeå, Umeå, Sweden, [2]Department of Chemistry, University of York, York, England, and [3]Department of Biochemistry, University of Otago, Dunedin, New Zealand

Summary. The behaviour of proinsulin and insulin in the presence of zinc suggests it plays an important role in insulin's production in the B-cell for the vast majority of animal species. The postulate that proinsulin forms a zinc containing hexamer soon after its synthesis and that this organization of the molecule is maintained through all the subsequent processes is supported by our observation that the proinsulin hexamer is converted readily into the insulin hexamer. In addition the zinc ions enhance proinsulin's solubility and render insulin insoluble. Zinc ions also appear to play an important role in the microcrystalline character of the precipitated insulin granule. There may be advantages in condensing the stored material in this way; it will reduce contact with the surrounding membrane where the converting, and possibly other enzymes, are thought to be located, and it will tend to exclude incompletely converted hexamers.

Key words: Zinc, insulin, proinsulin, insulin biosynthesis.

It is now relatively clear that insulin is synthesized by means of a series of coordinated events, each of which imposes its own constraints on the hormone's structure [1]. As the molecule moves from one process to the next, different alterations in its structure allow it to satisfy its new environments. Any attempt to understand the molecule's structure and behaviour must take into account these different biosynthetic stages.

Insulin is initially manufactured at the ribosome as a single chain polypeptide of some 110 amino acid residues, pre-proinsulin [1]. It carries a hydrophobic N-terminal extension of about 25 residues which it has been suggested induces binding of the ribosome to the microsomal membrane [2]. The N-terminal extension is then rapidly removed, giving proinsulin which has the capacity to fold spontaneously into the correct three-dimensional structure [1].

In most species the pancreatic islet B-cell is known to contain high levels of zinc [3, 4]. Since zinc induces proinsulin hexamerization in vitro [5] it seems likely that proinsulin monomers will aggregate into hexamers in B-cells rich in zinc. There is evidence that the zinc-proinsulin hexamer is closely related to the insulin hexamer and that the insulin moiety in the proinsulin hexamer is organized as the insulin hexamer [5]. While insulin is known to precipitate rapidly at low zinc concentrations, proinsulin remains soluble in the presence of zinc. Hence, proinsulin presumably remains in solution inside the rough endoplasmic reticulum and during the subsequent transport to the Golgi apparatus, where conversion of the substrate to insulin begins [1]. The proteolytic removal of the connecting peptide continues within the storage vesicles [1], which themselves most likely develop from the Golgi apparatus. The newly formed insulin is rapidly precipitated and, in most animals, is frequently observed to assume a crystalline shape [1].

In response to metabolic requirements the vesicle contents are expelled into the circulation, where they bring about the diverse anabolic and regulatory effects of insulin.

The objects of this paper are two-fold: first to describe experiments designed to reveal more about the role of zinc in proinsulin conversion and subsequent granule formation; secondly to discuss how chemical and structural principles, operating in insulin production, explain aspects of insulin's structure.

0012-186X/80/0019/0174/$01.80

Calculations, Experiments and Results

Zinc and Insulin Levels in the B-cell

There are a number of reports on the amount of zinc in islet tissue from a variety of species [3, 4, 6]. The zinc levels are usually distributed in two ranges: 2–3 mg/100 g tissue ($0.3–0.5 \times 10^{-3}$ mol/l) and 10–12 mg/100 g tissue ($1.5–1.8 \times 10^{-3}$ mol/l). The lower levels are found in animals whose insulins are not crystallized or aggregated to hexamers in the presence of zinc [7, 8]. The higher concentrations are found in species where the insulins have a specific zinc binding and coordinating capacity [9]. The lower levels of zinc found in islet tissue from guinea pig and the Atlantic hagfish correspond to the levels of zinc found in many other tissues where no particular affinity for zinc has been observed [4]. A similarly low level of zinc (accompanied by a low insulin level) was reported by Berglund and Hellman in severely diabetic mice [10]. On the above basis it seems reasonable to assume that zinc levels of 10 mg/100 g islet tissue is associated with the insulin processing machinery. Given the continuity between the storage vesicle and the cisternal space we will assume a uniform zinc distribution within this compartment, comprising 30% of the B-cell [11]. Hence, the zinc concentration within the insulin processing compartment can be estimated as 5×10^{-3} mol/l.

Mammalian islets contain 1–3% insulin on a wet weight basis (assuming 80% water content and 75% B-cells per islet) [12]. About 60–80% of that insulin is associated with a sedimentable granule fraction [13, 14], comprising 12% of the cell volume [11]. This gives an insulin concentration of $1–3 \times 10^{-2}$ mol/l, were it in solution inside the storage vesicle. Similar values can be arrived at by using the number of insulin molecules per granule calculated in three different ways by Howell [15] and the volume of a granule estimated by Lange and Dean [3, 11]. Morphometric and chemical analyses show that hagfish insulin is present in its B-cell storage vesicle at about the same level as in other species studied (Emdin, unpublished). At this high concentration insulins are insoluble around neutral pH and must exist in a crystalline or precipitated state.

The approximately equimolar stoichiometry between zinc and insulin arrived at above, corresponds well with the observation made by Berglund and Hellman [10]. They observed a close correlation between zinc and insulin content in islets from ob/ob mice, and the amounts were roughly equimolar.

Since the biosynthetic process is not in chemical equilibrium we cannot estimate the distribution of zinc between the soluble proinsulin hexamer and the crystalline insulin which both interact with zinc, nor describe the distribution of zinc within the insulin processing machinery. One approach to understand the zinc interactions, which sharpen the discrimination between proinsulin and insulin was to determine the structure of 2 Zn pig insulin crystals soaked in excess zinc ions.

Preparation of the Zinc Soaked Crystals

Rhombohedral 2 zinc pig insulin crystals (1 mm in their longest dimension) were crystallized from 0.05 mol/l citrate buffer at pH 6.2 following the procedure of Schlichtkrull [19]. The crystals were washed in 2–5 ml 0.01 mol/l acetate (pH 6.2–7.0) until citrate free. They were then soaked in 1.5×10^{-3} mol/l zinc sulphate at pH 7.0 for six days. These conditions minimized crystal damage which reduced the quality of X-ray diffraction pattern.

Data Collection

For X-ray diffraction studies the zinc soaked 2-zinc insulin crystal was mounted in a thin-walled glass capillary with a small reservoir of solution and sealed.

The X-ray diffraction data were collected on an automatic computer controlled Hilger and Watt 4 circle diffractometer. Data were collected out to a spacing of 4.5 Å and corrected for Lorentz, polarization and absorption effects. The agreement factor between equivalents for the 506 independent X-ray reflections, Rm, was 0.047 where:

$$Rm = \frac{\sum_{hkl} \left(\sum_{i=1}^{n} I_i - \bar{I} \right)}{\sum_{hkl}^{n} n\bar{I}}$$

(I is the intensity of the X-ray reflection).

The data were scaled to the native series and correlation made for the larger temperature factor ($\Delta B = 14$ Å^2) of the zinc soaked crystals.

Determination of the Bound Zinc Positions

The difference Fourier electron density

$$\Delta\varrho_{xyz} = \frac{1}{V} \sum_h \sum_k \sum_l \Delta F \, e^{-2\pi i(hx + ky + lz)}$$

where $\Delta F = m(|F_{Zn}| - |F_{nat}|) \, e^{2\pi i \alpha_{nat}}$

and F_{Zn} = the structure factor for the zinc soaked crystal structure

F_{nat} = the structure factor for the native crystal

was calculated using phases α_{hkl} determined by isomorphous replacement [16]. The figure of merit, which is a measure of the phases' quality, is denoted by m.

There were two prominent peaks in the map. Their positions and occupancies were refined by least squares, minimizing:

$$E^2 = \Sigma w\,(K|F_{Zn}| - |F_{nat} + f_{Zn}|)$$

where K = the scale to native and zinc soaked series, W = the weighting term: in this case the figure of merit for the isomorphous phase.

The figure of merit was cut off in these calculations at 0.85 reducing the number of terms to 433.

A subsequent difference Fourier was calculated in which the electron density

$$\Delta\varrho = \frac{1}{V}\sum_{hkl} \Delta F\, e^{-2\pi i\,(hx+ky+lz)}$$

where

$$\Delta F = m\,(|F_{Zn}| - |F_{nat} + f_{Zn}|)\, e^{2\pi i\,\alpha_{nat}}$$

and f_{Zn} = the calculated zinc ion contribution to F_{Zn}. This showed further peaks of lower occupancy. Zinc ions were placed at the positions of the peaks and were subjected to least squares refinement. Those zinc positions which refined to an occupancy greater than 7ē are listed in Table 1. The nearest residue to each zinc position is also listed.

The agreement factor

$$R = \frac{\sum_{hkl}(|F_{Zn}| - |F_{nat} + f_{Zn}|)}{\sum_{hkl}|F_{Zn}|}$$

fell from 19.0%, assuming no zinc binding, to 14.4% when the bound zinc ions were included.

The Bound Zinc Positions

Distinction between the centrally coordinated zinc ions, which bind to histidyl side chains and organize the hexamer, and those which bind less specifically to carboxylate groups has to be made. We do this by referring to the former as coordinated and the latter as bound zinc. Depending on the insulin's conformation and the condition in the vesicle, there would be between 2 and 4 zinc ions coordinated per hexamer corresponding to the known 2 and 4 Zn insulin crystals. The two central zinc ions coordinated to the B_{10} histidines in 2 Zn insulin are not affected by the presence of the additional zinc ions and there is no evidence for the development of extra zinc coordination by B_5 histidine as seen for example in 4 Zn insulin

Table 1. Details of zinc coordination and binding in 2 Zn pig insulin crystals

Site number	Coordinates x	y	z	Occupancy	Nearest residue
1	0.0	0.0	0.23	1.00	B_{10} (I) histidine
2	0.0	0.0	−0.24	1.00	B_{10} (II) histidine
3	0.03	0.02	0.01	0.80	B_{13} (I + II) glutamic acid
4	0.13	0.30	0.57	0.67	B_{30} (II) carboxylic acid A_4 (II) glutamic acid
5	0.10	0.10	0.51	0.37	B_5 (I + II) histidine
6	0.33	0.14	0.98	0.38	A_{17} glutamic acid
7	0.29	0.17	0.09	0.38	A_{17} (I) glutamic acid B_1 (II) α-amino
8	0.11	0.07	0.37	0.33	B_5 (I) histidine
9	0.18	0.27	0.27	0.27	B_{21} (II) glutamic acid

The I and II refer to the two independent molecules in the 2 Zn pig insulin crystal asymmetric unit
The zinc positions are in fractional coordinates referred to hexagonal axes

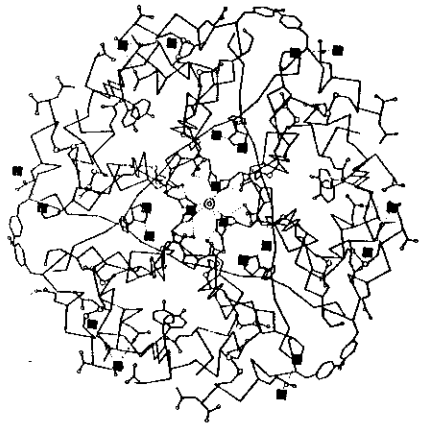


Fig. 1. The 2 Zn pig insulin hexamer viewed down the three fold crystal axis. The main chain structure is represented by the α-carbon positions only. The tyrosine sidechains, which make contact in dimer and hexamer formation are shown, as well as the histidine sidechains. Sidechains which are interacting with zinc ions are connected by dots to the bound zinc ions represented as squares

[17]. The most fully occupied new site is at the hexamer centre between the two B_{13} glutamic acid sidechains paired together by the association of insulin dimers in the hexamer (Fig. 1). At lower pH (6.0 – 6.5) divalent metal ion binding still occurs at this site but is reduced at the other sites (Dodson, unpublished). It seems likely that substitution at B_{13} sidechains can occur when the hexamers is in solution.

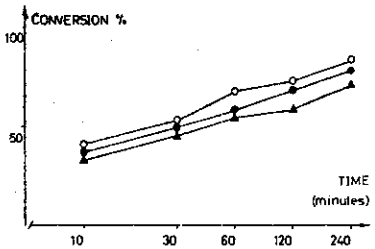


Fig. 2. Time course of proinsulin conversion with and without zinc. Pig proinsulin (0.8×10^{-3} mol/l) was incubated with trypsin (1:300) at pH 7.3, 37 °C for the times indicated. The rate of conversion was analyzed by gel electrophoresis and calculated as described in the text. Conditions were: (○) with 0.5×10^{-3} mol/l EDTA, (●) with two zinc ions per proinsulin hexamer and (▲) with six zinc ions per hexamer

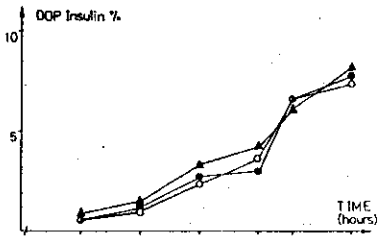


Fig. 3. Time course of desoctapeptide (DOP) – insulin formation with and without zinc. Pig insulin (0.5×10^{-3} mol/l) was incubated with trypsin (1:100) at pH 7.3, 37 °C. The rate of DOP-insulin formation was calculated after gel electrophoresis (see text). Conditions were: (○) with 0.5×10^{-3} mol/l EDTA, (●) with two zinc ions per insulin hexamer and (▲) with six zinc ions per insulin hexamer

Binding at the three symmetry related sites at B_{13}, together with the two centrally coordinated zincs, gives in total about five zincs per hexamer.

The additional zinc binding in the 2-zinc insulin crystal takes place on the hexamer's surface and involves interactions with mostly glutamic acid and histidine sidechains. The position of these sites and their adjacent side chains are illustrated in Figure 1; altogether there are seven zinc binding sites and two to four coordination sites per hexamer. Some of these zinc binding sites are produced by crystal packing and would probably not occur in solution. There is a total amount of 11.6 zinc ions in the hexamer; this corresponds to about 2.3%. This figure is sensibly near 2.5%, which is the value reported by Schlichtkrull for the zinc content in rhombohedral insulin crystal preparations in acetate buffer at neutral pH [18].

Effect of Zinc on Tryptic Proinsulin Conversion

Pig proinsulin, with a zinc content of 0.02% was dissolved in 0.05 mol/l Tris-HCl pH 7.3 to a concentration of 0.8×10^{-3} mol/l. The proinsulin was incubated at 37 °C with trypsin (diphenyl-carbamyl-chloride-treated, Sigma) 1:300 (w/w) and either EDTA 0.5×10^{-3} mol/l or zinc acetate yielding two or six zinc ions per proinsulin hexamer. At intervals samples were withdrawn and combined with an excess of soybean trypsin inhibitor (Sigma). The samples were analyzed by gel electrophoresis at pH 8.7 [19]. The protein bands were stained with Coomassie Brilliant Blue [20] and quantitated with a gel scanner. The amount of proinsulin converted was calculated as the sum of the optical densities of des B_{30}-insulin and the intermediate(s) moving to a position in between proinsulin and insulin [21], divided by the sum of proinsulin, des B_{30}-insulin and intermediate(s). During the course of the reaction a precipitate (diarginyl-insulin) formed after about four minutes both with and without zinc. The time-course of the reaction is shown in Figure 2. At the most, a slight reduction of the conversion-rate in the presence of zinc was noted. In similar experiments with potassium phosphate buffer (0.1 mol/l, pH 7.5) this difference of rates was even further reduced. Hence, the overall time course of tryptic proinsulin conversion was not significantly perturbed by the presence of zinc and the different aggregation state induced by the cation.

Effect of Zinc on Tryptic Insulin Degradation

Trypsin induced insulin degradation was studied by measuring the rate of desoctapeptide (DOP) insulin formation. Pig insulin with a zinc content of 0.02% and at a concentration of 0.5×10^{-3} mol/l was incubated with trypsin (1:100 w/w) under conditions otherwise similar to above. After gel electrophoresis and optical scanning, the amount of DOP-insulin formed was calculated as; DOP insulin: (DOP insulin + insulin). The over-all time course of this reaction is shown in Figure 3. When zinc was added to a level of six zinc ions per hexamer insulin precipitated instantaneously, but the rate of DOP-insulin formation was not affected. As with proinsulin, experiments were also performed in phosphate buffer at pH 7.5 but we found no difference with or without zinc. Hence, insulin degradation by trypsin was neither affected by the different states studied here, dimeric and hexameric, nor the precipitation of the substrate with equimolar amounts of zinc ions.

Calcium and Zinc as Insulin Precipitating Agents

Based on experiments with islets cultured in calcium-free medium, Howell et al. suggested that this cation may have an important function in insulin storage [22]. Calcium is known to precipitate insulin, and to bind to rhombohedral zinc insulin crystals (Blundell, Bedarkar, private communication; Dodson, unpublished). There is also the evidence that calcium is present in the secretion granules [15]. In an attempt to compare the abilities of zinc and calcium to precipitate insulin the simple experiment presented in Table 2 was done. It showed that zinc was far more able than calcium to precipitate insulin. This difference, more than one order of magnitude, was most pronounced at low insulin concentrations.

Table 2. Calcium and zinc as insulin precipitating agents

Pig insulin concentrations mol/l	Number of cations/insulin monomer giving precipitation	
	Calcium	Zinc
$5 \cdot 10^{-6}$	No ppt[a]	No ppt[a]
$25 \cdot 10^{-6}$	No ppt[a]	1.4
$50 \cdot 10^{-6}$	No ppt[a]	1
$250 \cdot 10^{-6}$	No ppt[a]	1
$330 \cdot 10^{-6}$	440	1
$500 \cdot 10^{-6}$	74	1

[a] ppt = precipitate

Small volumes of concentrated solutions of either $CaCl_2$ or $ZnCl_2$ were added to a series of insulin solutions (0.05 mol/l Tris-HCl, pH 7.3 at 37 °C) ranging in concentration from 5 to 500×10^{-6} mol/l. This was repeated every five minutes until precipitation or turbidity appeared. In cases where no precipitation was seen the $CaCl_2$ and $ZnCl_2$ solutions were added until they were at 1 mol/l concentration 1 mmol/l concentration respectively

Discussion

Generally accepted functions for zinc in the production of insulin by the zinc-rich B-cell are:

1. Assembly of the proinsulin and the derived insulin hexamers [5, 9].

2. Solubility of the zinc proinsulin hexamer [5].

3. Precipitation and crystallization of the nascent insulin [5, 9].

4. Crystal formation which presumably reduces the rate of proteolysis, especially by the converting enzyme(s) [9, 23].

It is clear from this list that zinc ions play a central role in insulin biosynthesis. Since there is now considerable experimental evidence that the hexameric structure has the right properties for the chemical and structural events that take place at each stage after the folding of the proinsulin monomer, we shall assume the hexameric organization is preserved intact throughout the processing after the initial biosynthetic steps.

Assembly, Solubility and Zinc Binding of Proinsulin

As soon as a sufficient number of proinsulin molecules inside the rough endoplasmic reticulum are released from their N-terminal extensions they will, after folding, presumably assemble into hexamers in which two or four zinc ions are centrally coordinated provided enough zinc is present [17]. Proinsulin is a remarkably soluble molecule. Grant et al. have shown that proinsulin in vitro is organized as a zinc hexamer with aggregation properties closely similar to those of insulin [5]. The pig proinsulin hexamer, which at neutral pH carries a net charge of −12, has a capacity to bind 30 zinc ions per hexamer, of which two to four are coordinated. Proinsulin's solubility as a hexamer is thus enhanced by the addition of 60 positive charges, present as a coat of zinc ions, roughly equivalent to the 54 negatively charged groups on the surface of the pig proinsulin hexamer [5].

The zinc binding sites in rhombohedral 2 Zn pig insulin are usually found adjacent to carboxylate groups. The dispersal of these sites on the molecule's surface is expected but their fractional occupancy suggests they do not and cannot assume the appropriate geometry for favourable coordination. There is, however, one highly occupied site at the hexamer's centre between the two B_{13} glutamic acid side chains brought together by the molecule's assembly. Here the double zinc charges are balanced completely by the two carboxylate groups (Fig. 1).

This pattern of high occupancy being associated with the pairing of the negatively charged carboxylate groups suggests an explanation for the zinc binding capacities of proinsulin. There are 54 carboxylate groups in the pig proinsulin hexamer and as the hexamer binds some 30 zinc ions in vitro then it is likely they are all involved in interaction with the cation. We propose this stoichiometry in the zinc binding might be achieved through pairing of the carboxylic acid side chains, brought about by the folding of the connecting peptide and the slight restructuring in the insulin moiety. Among the zinc-binding insulins the near invariance of four glutamic acid residues and of the aspartic acid residue in the connecting peptide sequence is consistent with them having such a specific structural role on the surface. Snell and Smyth have attempted to predict the folding of the proinsulin connecting peptide [24]. There is no pairing of carboxylate groups in their proposed structure.

High solubility in the cell will presumably be achieved by the process of zinc binding, facilitating the transport of the prohormone to its site of conversion. In this organized state the proinsulin has buried all the residues thought to be involved in the expression of the biological activity, either by aggregation or by folding of the connecting peptide [10]. Whether the burial of these residues is a necessary biological device we do not know.

Conversion of Proinsulin

Kemmler et al. have shown that under strictly controlled conditions a mixture of trypsin and carboxypeptidase B will convert zinc-free bovine proinsulin to native insulin [25]. We suggest that upon arrival at the Golgi apparatus some five to ten minutes later, the prohormone, still organized as a hexamer, is converted to an insulin hexamer by the enzyme(s) with tryptic and carboxypeptidase-B-like specificity [1]. We would expect this from proinsulin's behaviour in solution and from the observation that conversion with trypsin in vitro is virtually unaffected by the presence of zinc (i. e. hexameric conformation). We can understand the equal ability of the zinc proinsulin hexamer and zinc-free proinsulin to be converted by reference to the 2 Zn pig insulin hexamer structure. Here we find the exposure of the B-chain's C-terminus and the A-chain's N-terminus unaffected by the assembly to dimers and hexamers [16].

With the removal of the 6 connecting peptides, possibly completed in one exposure to the enzyme, the pig proinsulin hexamer loses 24 carboxylic acid side groups. We propose that the loss of these charges and the structural changes at the A-chain N-terminus and B-chain C-terminus drastically reduces the molecule's capacity to bind zinc. Grant et al. [5] report that insulin's solubility falls sharply as the bound zinc increases from 3–6 zinc ions per hexamer. We note that 6 zinc ions will exactly neutralize the net 12 negative charges present and favour precipitation; perhaps this favourable stoichiometry explains the role zinc plays in the insulin hexamer's lower solubility. Moreover, the loss of the 24 negative charges with the removal of the C-peptide and the 24 basic connecting residues (4 per monomer) and the exposure of 12 benzene rings (A_{19} tyrosines and B_{25} phenylalanines) may further reduce solubility by increasing the non-polar surface of the hexamer.

The insolubility of the converted product will favour complete hydrolysis of the soluble precursor. A certain fraction of proinsulin monomers (about 2–4%) remain unconverted in the hexamer and are incorporated into the crystal and expelled into the circulation [26], which may be seen as an unfavourable consequence of the insulin hexamer's insolubility.

Finally we have shown that calcium compared with zinc, is a poor insulin precipitating agent and we consider that the initial precipitation of the newly converted insulin is induced by zinc not by calcium and occurs at insulin concentrations several orders of magnitude below insulin's point of saturation.

Crystallization within the Storage Vesicle

In animals whose insulins are coordinated by zinc examination of electron micrographs reveals crystalline granules in the B-cells [27]. The abundance of such apparently crystalline structures may vary with fixation and the species under study and they may be scarce within the B-cell. Lange, who has studied this problem in detail, has interpreted crystallization in B-granules as being accidental rather than essential [29]. The constitution of the stored insulin crystals has only be examined directly in the grass snake and the salamander using optical transforms [29]. In these two animals the crystals are not rhombohedral. Comparable studies have not been reported on insulins whose crystals have been characterized in vitro, such as that of the rat and pig. Two crystalline forms of the rat insulins have been obtained in the laboratory, one rhombohedral and the other cubic. These have been related to the two characteristic crystalline forms observed in the granules in the rat pancreatic B-cell [20]. Some variation in the insulin storage crystal in the presence of the other numerous molecular species, such as calcium ions, arginine, lysine and the connecting peptide present in the storage granule, is perhaps to be expected. Supporting this is the observation that in the laboratory with pig and beef insulin we find the crystal form and the conditions of precipitation can easily be changed by adding various molecules to the solution (Cutfield, Dodson, Tolley, unpublished, [18]).

Schlichtkrull [18] has shown that insulin precipitates as microcrystals in the presence of zinc ions between pH 4.8–5.6. In the B-cell storage vesicle the picture is complicated. The presence of other molecules in the vesicle is also likely to alter the pH range of the phenomenon. Insulin's high concentration in the vesicle may itself buffer the pH to nearer 6. These factors may explain Lange's observation that no in vivo insulin crystal corresponds to any known crystal form grown in the laboratory [29].

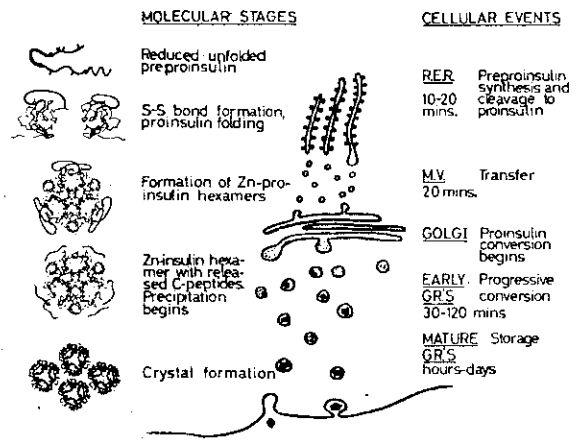


Fig. 4. Schematic drawing of the elements in the B-cell responsible for insulin biosynthesis. Alongside this scheme are the corresponding molecular species present at the different stages of insulin's biosynthesis. (Partly redrawn from Steiner et al. [34])

Crystallisation of the converted hexamer may help ensure high conversion in the insulin hexamer since the crystal packing requirements will effectively exclude proinsulin surfaces from being incorporated into the crystal. In support of this Steiner [31] has demonstrated that proinsulin inhibits the growth of insulin crystals, though it must be borne in mind this experiment was done at low zinc and insulin concentrations.

Discussion of what role insulin's crystallisation plays in preventing proteolysis during storage is limited by our ignorance of the converting enzyme's specificity. Experiments with trypsin, which is specific for basic residues, suggest there are advantages in dimerisation [32]. A pancreatic protease isolated by Yip was able to cleave the $B_{22} - B_{23}$ bond only in the absence of zinc; clearly the hexameric structure protected the hormone from attack by this enzyme [33]. We find that hexameric insulin in solution, or in the precipitated state, is as labile to trypsin as the dimer. This observation is consistent with the idea that crystallisation does provide protection to the stored insulin from an enzyme with trypsin like properties simply by condensing the molecules into crystals.

Zinc-free Insulins

In considering the role of zinc in the insulin processing machinery it is interesting to compare, where possible, the biosynthetic events in B-cells with high and low levels of zinc. An important difference is the appearance of the insulin storage vesicle. In zinc-rich B-cells the granules characteristically appear crystalline while in B-cells [27] with low levels of zinc the stored insulin is dispersed more uniformly throughout the vesicles. In the hagfish B-cell the granules are never condensed [28, 35], whereas in the guinea pig a minority of the granules are condensed, but they do not appear crystalline (Howell, personal communication). Loading of the proinsulin hexamer with zinc ions appears an important factor in its solubility. It is a question whether hagfish and guinea pig proinsulins can and need to bind zinc in their cellular environment where the zinc levels are significantly lower. The distribution of carboxylic acid chains in guinea pig proinsulin differs somewhat from the other proinsulins [24], and there is in addition evidence of structural alterations in the guinea pig insulin moiety [36]. These differences may be sufficient to abolish its zinc binding capacity but consideration of the sequence does suggest guinea pig proinsulin might bind zinc ions. Similar speculations about the hagfish proinsulin cannot be made since its structure is not known. If these two proinsulins cannot bind sufficient zinc then their solubilities must be determined solely by the character of their surface residues, as must be also the solubilities of the newly produced insulin, which will be precipitated only when the concentration builds up to high enough levels. This lag between the appearance of insulin and its precipitation is in sharp contrast to the rapid zinc-induced crystalline precipitation on the zinc-rich B-cell.

It is interesting to note that the amount of zinc required for insulin production is very small in relation to the total zinc content of the body. By processing its insulin at low levels of zinc, rather than at high levels, a guinea pig saves only some 0.1% of its total zinc content. One possible explanation for the lack of condensed crystalline granules in non zinc binding insulin storage vesicles is that these insulins are very resistant to further cleavage by their converting enzyme(s). The loss of zinc in the guinea pig and coypu insulin producing machinery is not then a consequence of a zinc shortage [23] but follows the ability of the insulin to dispense with aggregation in storage. This, in turn, removes the structural constraints on the residues responsible for assembly [9, 23], and the zinc ions, no longer necessary, are lost.

Effects during and after Secretion

There remain further possible roles for zinc ions in the events that take place during and after insulin's secretion. It has been established that insulin crystals and granules become more stable in the presence of surplus zinc ions [15, 18, 37]. It is hard to estimate the speed with which the zinc ions in the crystal will disperse as the vesicle contents are expelled via the basal membranes into the circulating medium; but while they are present the zinc ions will affect the crystal's rate of dissolution.

The balance of the evidence suggests that the insulin molecule moves into the blood as a hexamer and, we may speculate, with a local packet of liberated zinc ions. The observation that zinc ions affect insulin's binding and degradation in the liver [38] suggests a further possible role for zinc ions outside the B-cell. If this is a function for zinc then one might expect differences in the insulin binding and degradation properties of guinea pig and hagfish livers since in these animals the B cell does not have high levels of zinc ions.

Conclusion

The biological processes that occur during insulin's production are reflected in the sequences, structures and chemical behaviour of proinsulin and insulin. An interesting feature in the biosynthesis of insulin is the role of the zinc ion whose chemical properties are incorporated into the whole process. This is illustrated schematically in Figure 4. The zinc ions are not, however, essential – several animals produce insulin without using zinc ions. Their incorporation, however, amplifies with great advantage the chemistry available to the proinsulin and insulin molecules, particularly in their aggregation and solubility properties.

There are still gaps in our knowledge; the converting enzyme(s) has not yet been characterized and we do not know enough about what other proteins and enzymes are present in the vesicle membrane. As these details become known we will be in a better position to understand not only insulin's structure, but also the structural differences in insulins from different animal species.

Acknowledgements. Supported by grants from The British Medical Research Council and the Swedish Medical Research Council (12X-718), and the British and Swedish Diabetes Associations. Some of the zinc binding studies were carried out in the Laboratory of Molecular Biophysics at Oxford University. We are grateful for the help and interest of Professor Dorothy Hodgkin and David Phillips.

References

1. Steiner DF (1976) Insulin today. Diabetes 26: 322–340
2. Blobel G, Sabatine DD (1971) Ribosome-membrane interaction in eukaryotic cells. Biomembranes 2: 193–195
3. Lange RH (1973) Histochemistry of the islets of Langerhans. In: Graumann W, Neumann K (eds) Handbuch der Histochemie. Gustav Fischer Stuttgart, Bd VIII/I, pp 1–269
4. Havu N, Lundgren G, Falkmer S (1977) Zinc and manganese contents of micro-dissected pancreatic islets of some rodents. Acta Endocrinol (Kbh) 86: 570–577
5. Grant PT, Coombs TL, Frank BH (1972) Differences in the nature of the interactions of insulin and proinsulin with zinc. Biochem J 126: 433–440
6. Havu N, Lundgren G, Falkmer S (1977) Microchemical assays of glutathione, zinc, cobalt and manganese in micro-dissected areas of the endocrine pancreas in the hagfish, Myxine glutinosa. Acta Endocrinol (Kbh) 86: 561–569
7. Peterson JD, Coulter CL, Steiner DF, Emdin SO, Falkmer S (1974) Structural and crystallographic observations on hagfish insulin. Nature 251: 239–240
8. Zimmerman AE, Yip CC (1974) Guinea pig insulin. J Biol Chem 249: 4021–4025
9. Blundell T, Dodson G, Hodgkin D, Mercola D (1972) Insulin: The structure in the crystal and its reflection in chemistry and biology. Adv Protein Chem 26: 279–402
10. Berglund O, Hellman B (1976) Evidence for a role of zinc in the storage of insulin in the mouse pancreatic B-cells. Diabetologia 12: 380–381
11. Dean PM (1973) Ultrastructural morphometry of the pancreatic βcell. Diabetologia 9: 115–119
12. Hellman B, Täljedal I-B (1972) Histochemistry of the pancreatic islet cells. In: Steiner DF, Freinkel N (eds) Handbook of physiology, endocrinology I. Williams and Wilkins, Baltimore, p 91–110
13. Howell SL, Fink CJ, Lacy PE (1969) Isolation and properties of secretory granules from rat islets of Langerhans. J Cell Biol 41: 154–161
14. Noe BD, Baste CA, Bauer GE (1977) Studies on proinsulin and proglucagon biosynthesis and conversion at the subcellular level. J Cell Biol 74: 578–588
15. Howell SL (1974) The molecular organization of the βgranule of the islets of Langerhans Adv Cytopharmacol 2: 319–327
16. Adams MJ, Blundell TL, Dodson EJ, Dodson GG, Vijayan M, Baker EN, Harding MM, Hodgkin DC, Rimmer B, Sheat S (1969) Structure of rhombohedral 2 Zinc insulin crystals. Nature 224: 491–495

17. Bentley G, Dodson E, Dodson G, Hodgkin D, Mercola D (1976) Structure of insulin in 4-zinc insulin. Nature 261: 166–168
18. Schlichtkrull J (1961) Insulin crystals. Thesis Univ. of Copenhagen
19. Ornstein L, Davies JB (1962) Disc electrophoresis. Distillation Products Industries, Rochester, New York
20. Diezel W, Kopperschläger G, Hofmann E (1972) An improved procedure for protein staining in polyacrylamide gels with a new type of Coomassie Brilliant Blue. Anal Biochem 48: 617–620
21. Steiner DF, Clark JL, Nolan C, Rubinstein AH, Margoliash E, Aten B, Oyer PE (1969) Proinsulin and the biosynthesis of insulin. Recent Prog Horm Res 25: 207–282
22. Howell SL, Tyhurst M, Duvefelt H, Andersson A, Hellerström C (1978) Role of zinc and calcium in the formation and storage of insulin in the pancreatic β-cell. Cell Tissue Res 188: 107–118
23. Blundell TL, Wood SP (1975) Is the evolution of insulin Darwinian or due to selectively natural mutation? Nature 257: 197–203
24. Snell CR, Smyth DG (1975) Proinsulin: A proposed three-dimensional structure. J Biol Chem 250: 6291–6295
25. Kemmler W, Peterson JD, Steiner DF (1971) Studies on the conversion of proinsulin to insulin. J Biol Chem 246: 6786–6791
26. Steiner DF, Kemmler W, Clark JL, Oyer PE, Rubinstein AH (1972) The biosynthesis of insulin. In: Steiner DF, Freinkel N (eds) Handbook of physiology, endocrinology 1. Williams and Wilkins, Baltimore, p 175–198
27. Sato T, Herman L, Fitzgerald PJ (1966) The comparative ultrastructure of the pancreatic islet of Langerhans. Gen Comp Endocrinol 7: 132–157
28. Östberg Y (1976) The entero-insular endocrine organ in a cyclostome Myxine glutinosa. Thesis. Umeå Univ. Med. Diss. New Ser. No. 15
29. Lange RH (1976) Crystallography of islet secretory granules – a contribution to the problem of chemical composition of secretion granules. In: Fujita T (ed) Endocrine gut and pancreas. Elsevier Scientific Publishing Co, Amsterdam, p 167–178
30. Wood SD, Tickle IJ, Blundell TL, Wollmer A, Steiner DF (1978) Insulin polymorphism: Physical and biological properties of rat insulins. Arch Biochem Biophys 186: 175–183
31. Steiner DF (1973) Cocrystallisation of proinsulin and insulin. Nature 243: 528–530
32. Wang S-S Carpenter FH (1969) Kinetics of the tryptic hydrolysis of zinc-free bovine insulin. J Biol Chem 244: 5537–5543 (1969)
33. Yip CC (1971) A bovine pancreatic enzyme catalyzing the conversion of proinsulin to insulin. Proc Natl Acad Sci USA 68: 1312–315
34. Steiner DF, Kemmler W, Tager, HS, Rubinstein AH, Lernmark Å, Zühlke H (1975) Proteolytic mechanisms in the biosynthesis of polypeptide hormones. In: Reich E, Rifkin DB, Shaw E (eds) Proteases and biological control. Cold Spring Harbor Laboratory, p 531–549
35. Emdin SO, Falkmer S (1977) Phylogeny of insulin. Acta Paediatr Scand [Suppl] 270: 15–25
36. Massey DE, Smyth DG (1975) Guinea pig proinsulin. J Biol Chem 250: 6288–6290
37. Coore HG, Hellman B, Pihl E, Täljedahl I-B (1970) In: Falkmer S, Hellman B, Täljedahl I-B (eds): The structure and metabolism of the pancreatic islets. Pergamon Press, Oxford, p 385–395
38. Arquilla ER, Packer S, Tarmas W, Miaymoto S (1978) The effect of zinc on insulin metabolism. Endocrinology 103: 1440–1449

Received: September 18, 1979.
and in revised form: April 25, 1980

G. G. Dodson
Chemistry Department
York University
York, Y01 5DD
England

Reprint from: Insulin
Chemistry, Structure and Function of Insulin and Related Hormones
Editors: D. Brandenburg, A. Wollmer

POSSIBLE RELATIONSHIPS IN THE PROCESSING, STORAGE AND SECRETION OF SOME INSULIN RELATED PEPTIDES

G. G. Dodson
Department of Chemistry, University of York, York, U.K.

N. Isaacs
St. Vincent's School of Medical Research, Fitzroy, Victoria, Australia

R. A. Bradshaw
Department of Biological Chemistry, Washington University School of Medicine, St. Louis, U.S.A.

H. D. Niall
Howard Florey Institute, University of Melbourne, Parkville, Victoria, Australia.

Introduction

The amino acid sequences of relaxin (1),(2), insulin-like growth factor (IGF) (3) and nerve growth factor (NGF) (4) show a relationship that suggests they have evolved from a common precursor. Relaxin has two polypeptide chains similar in length to those in insulin; IGF and NGF are single chain molecules more analogous to proinsulin (Figure 1) where the A & B chains are joined by a connecting peptide. This feature appears to provide an insight into the different mechanisms that have evolved to release these hormones in response to the quite different requirements of their various physiological activities.

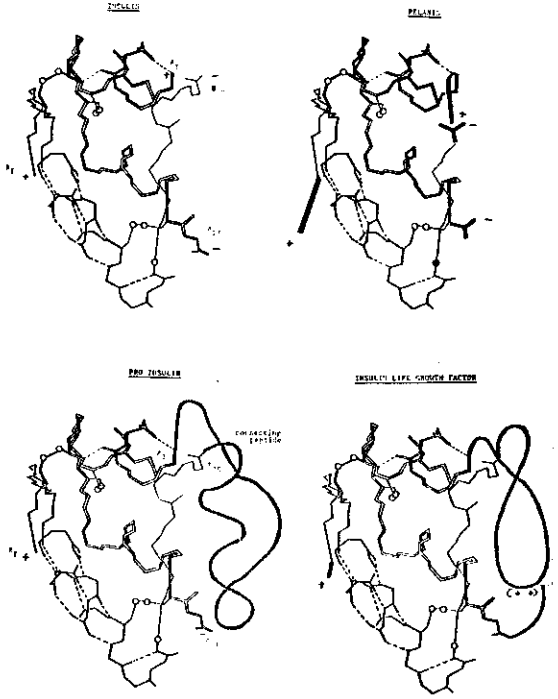
Insulin, Chemistry, Structure and Function of Insulin and Related Hormones

Figure 1

Converted
Peptides

Unconverted
Peptides

NGF
not shown



If the relationship with insulin is judged by the number of changes in the invariant insulin residues, then insulin is most closely related to IGF, less so to relaxin and much more distantly to NGF (5),(6). The conservation of the six half-cystines in the same relative sequence positions in insulin, IGF and relaxin suggests their spatial structure is related as well. Model building studies have demonstrated that the insulin folding can accommodate the IGF sequence as well as the very much more different relaxin sequence (1),(7),(8),(9). In NGF only three of the six half-cystines found in insulin are present, although two are paired analogous to B-19/A-20 disulphide. Certain structural features that occur in insulin, such as the α-helix between B9 - B19, may exist in NGF although the character of its sequence appears unfavourable for generally similar folding (10),(11).

The biosynthetic stages in the production of insulin by the β cell are now largely understood (12),(13). Insulin is synthesized as a single polypeptide chain with a signal (or "pre") segment preceding the B chain and a connecting peptide joining the carboxyl terminus of the B chain and the amino terminus of the A chain (12). The "pre" segment is removed from the nascent chain during or immediately following its penetration of the rough endoplasmic reticulum membrane. The prohormone is then transported from the rough endoplasmic reticulum (RER) to the Golgi apparatus where it is incorporated into vesicles (12). The enzymatic cleavages and the resulting removal of the connecting peptide change the physical chemistry of the molecule dramatically. Whereas proinsulin is very soluble and remains in solution even in the presence of high concentrations of zinc ions, insulin precipitates when the stoichiometry exceeds 2 gram atoms/ insulin hexamer, the amount required for the formation of the 2 zinc hexameric species (14). It has been proposed that it is the soluble proinsulin hexamer that is transported from the RER to the Golgi apparatus and that this is the species converted to the 2 Zn hexamer (14),(15). As the susceptible bonds at the carboxyl terminus of the B chain and amino terminus of the A chain are on the surface of the insulin hexamer, the removal of the connecting peptide from a hexameric proinsulin species is plausible. See Figure 2 and 3.

The high concentration of unbound zinc in the β cell ensures that the conversion of the proinsulin hexamer will result in the immediate precipitation of the insulin hexamer with concomitant granule formation. This removal of insulin from solution presumably will not only favour the continuing conversion reactions but also would protect the insulin

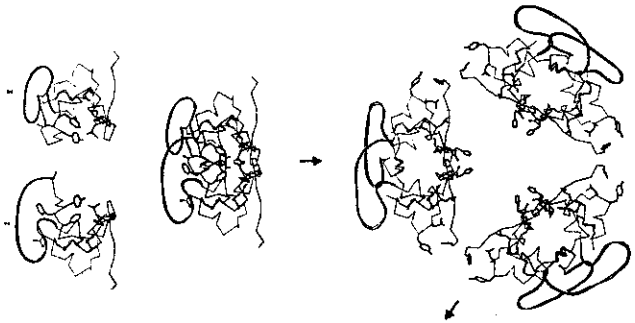


Figure 2

Proinsulin
Hexamer
Formation
(Soluble)

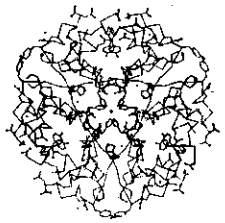


Figure 3

Insulin
Hexamer
(Insoluble)

hexamer from further proteolysis by reducing the contact with the vesicle membrane which contains the converting enzymes. Subsequently, in response to increases in the glucose concentration of the blood, the granules expel their content into the bloodstream where the insulin hexamers rapidly dissolve owing to the dilution and the higher pH of this milieu.

Although direct evidence is lacking, it seems highly probable that the synthesis and storage of relaxin is handled in much the same way. This is consistent with the increased synthesis and granule storage of relaxin by the granulosa cells of porcine corpora lutea during pregnancy (16). Just prior to parturition, the relaxin is released as a bolus into the bloodstream presumably influencing events concerned with the delivery of the foetus. We propose that the formation of relaxin granules within the storage vesicles in the corpora lutea is, as with insulin, associated with the excision of a connecting peptide leading to the formation of the two chain structure. Circumstantial evidence provided by amino acid sequence of relaxin itself and by the existence of putative intermediates in a proteolytic conversion mechanism lend further support to this hypothesis. (Niall, H. & Kwok, S. unpublished results.)

In contrast, IGF and NGF present a different picture. Their single chain structure shows there is no proteolytic removal of a connecting segment, suggesting that there is no precipitous change in the solution properties of these hormones during their synthesis. We propose therefore that neither IGF nor NGF form storage granules as insulin does. This could then be the reason why granule formation has not been detected for either growth factor. In fact, with one exception, neither IGF nor NGF is apparently "stored" in any tissue. The one instance where a high concentration is found is the adult male mouse submaxillary gland, which

contains substantial amounts of NGF (17). This situation however, is known to be unrelated to the action of NGF, representing rather a general derepression of the expressed genome of the convoluted tubule cells at puberty (18), (19). In addition a different solution for the storage of the large amounts in this one tissue has evolved, where instead of proteolytic processing, a high molecular weight complex (7S NGF), utilizing two other types of polypeptides, is used (20). As might be expected, this form of NGF is apparently unique to this one tissue.

The absence of NGF or IGF granules suggests a continuous diffusion from their cells of origin. This is entirely consistent with the continuous synthesis of NGF by end organs of the sympathetic nervous system followed by its local diffusion to the target neurons (21). Such a mechanism would also suit IGF, in its capacity as a putative somatomedin or growth hormone mediator (22). Upon induction by growth hormone, the IGF would diffuse to the skeletal tissues, where it would exert its growth stimulatory effects. Such a model for IGF action raises the possibility that the systemic levels of IGF represent only unused or "spent" hormone as opposed to material in active transport to a target tissue (5). Thus, the difficulties so far encountered in locating the physiologically relevant cells of origin of NGF and IGF can be seen, not as a puzzle, but as a consequence of the processes which produce them as a continuous secretion.

REFERENCES

1. James,R., Niall,H., Kwok,S. & Bryant-Greenwood,G., Nature, 267, 544-546 (1977).
2. Schwabe,C., McDonald,J.K. & Steinetz,B.G., Biochem. Biophys. Res. Commun. 75, 503-510 (1977).
3. Rinderknecht,E. & Humbel,R.E., J. Biol. Chem. 253, 2769-2776 (1978).
4. Angeletti,R.H. & Bradshaw,R.A., Proc. Natl. Acad. Sci. USA., 68, 2417-2420 (1971).
5. Bradshaw,R.A. & Niall,H.D. in Proceedings of the 6th Asia & Oceania Congress of Endocrinology, (C.J.Seng et al. eds.) Singapore, Vol. 1, 358-364 (1978).
6. Niall,H.D., Bradshaw,R.A. & Bryant-Greenwood,G. in Proceedings of the ICN-UCLA Symposium on Transmembrane Signalling (J.Collier, D.F.Steiner, M. Bitensky, & C.F.Fox, eds) Alan Liss, N.Y. (1978) in press.
7. Bedarkar,S., Turnell,W.G., Blundell,T.L. & Schwabe,C. Nature, 270, 449-451 (1977).
8. Isaacs,N., James,R., Niall,H., Bryant-Greenwood,G., Dodson,G., Evans,A. & North,A.C.T., Nature, 271, 278-281 (1978).
9. Blundell,T.L., Bedarkar,S., Rinderknecht,E. & Humbel,R.E. Proc. Natl. Acad. Sci. USA., 75, 180-184 (1978).
10. Frazier,W.A., Hogue-Angeletti,R.A., Sherman,R. & Bradshaw,R.A., Biochemistry, 12, 3281-3293 (1973).
11. Argos,P., Biochem. Biophys. Res. Commun., 70, 805-811 (1976).
12. Chan,S.J. & Steiner,D.F., Trends in Biochem. Sci., 2, 254-256 (1977).
13. Grant,P.T. & Coombs,T.L., Essays in Biochemistry, (Campbell,P.N. & Dickens,T., eds) Academic Press, New York, 6, 69-92 (1970).
14. Grant,P.T., Coombs,T.L. & Frank,B.H., Biochemical J., 126, 433 (1972).
15. Blundell,T., Dodson,G., Hodgkin, D. & Mercola, D. Adv. Protein Chem., 26, 279-402 (1972).
16. Belt,W.D., Anderson,L.L., Cavazos,L.F. & Melampy,R.M., Endocrinology, 89, 1-10 (1971).
17. Cohen,S., Proc. Natl. Acad. Sci. USA., 46, 302-311 (1960).
18. Junqueira,L.C.U., Fajer,A., Rabinovitch,M. & Frankenthal,L., J. Cellular Comp. Physiol., 34, 129-158 (1949).
19. Levi-MonFalcini,R. & Angelletti,P.U. in Regional Neurochemistry (S.S.Ketz and J.Elkes, eds.) Pergamon Press, N.Y., 362-376 (1961).
20. Varon,S., Nomura,J. & Shooter,E.M., Biochemistry, 6, 2202-2209 (1967).
21. Bradshaw,R.A., Ann. Rev. Biochem., 47, 191-216 (1978).
22. Froesch,E.R., Burgi,H., Muller,W.A., Humbel,R.E., Jakob,A. & Labhart,A., Recent Prog. Horm. Res., 23, 565-616 (1967).