POLYPEPTIDE SECONDARY STRUCTURE MAY DIRECT THE SPECIFICITY OF PROHORMONE CONVERSION

M. J. GEISOW

National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

Received 12 January 1978

1. Introduction

Hormones and other polypeptides secreted by cells are frequently biosynthesized as larger precursors. The necessary enzymic processing of these occurs prior to the secretion of hormones but after the secretion of blood factors and zymogens. These conversion processes are of particular importance in regulatory mechanisms [1] such as blood clotting. At present, the mechanisms of conversion from precursors to hormones are poorly understood in all but a few instances. Furthermore, hormones themselves are sometimes found as several homologous species, each with some activity, but differing in peptide length [2,3]. A clearer understanding of the intracellular processing events may help to distinguish between primary hormonal products and artefacts produced by non-specific enzyme attack [4].

A general mechanism for the intracellular processing of propolypeptides has been proposed. In this a trypsin-like enzyme cuts precursors specifically at adjacent basic amino acids. The resulting basic COOH-terminal residues are then removed by a carboxypeptidase B-like enzyme [5,6]. An example is the release of the C-peptide from proinsulin by cleavage at the carboxyl ends of –Arg–Arg– and –Lys–Arg– sequences. Both trypsin and carboxypeptidase B-like activities have been observed during the conversion of proparathyrin [16].

It is not known whether these enzymes are the same in all cells producing secretory material, or whether they are specific to particular polypeptides. For example albumin, glucagon and parathyrin all retain paired basic amino acid sequences, although their precursors are cleaved at identical or analogous residues. The present report indicates that the apparently high specificity of conversion may arise from readily predictable conformational properties of the precursors. The observations support a growing view that there are a limited number of converting enzymes with a wide cellular distribution.

2. Structure prediction methods

Regions of secondary structure were located in sequences published [19] by using the empirical rules in [9]. α-Helix and β-structure were deduced using recent tabulations of structure–formation parameters [20]. Beta-turns were located by computation of cumulative probabilities for tetrapeptides to form a reverse turn [21]. If extended regions of overlapping reverse turns were predicted by this method (due to a high frequency of serine, glycine and proline residues), the tetrapeptide with the highest probability of β-turn formation was chosen. Where appropriate, information from species variants was used in deciding structure. Usually sequence differences were not sufficient to alter the predicted secondary structure significantly.

3. Results and discussion

Amino acid sequences of precursors which contain pairs of basic residues are listed in table 1. An inspection of these did not reveal any additional features which could account for the restriction of proteolysis to the

Table 1
The sequences at double basic sites in propolypeptides

Propolypeptide	Processing sites	Sites not processed
Proalbumin (Human)	5 –Gly–Val–Phe–ARG–ARG▼–Asp–Ala–His–	–Glu–Ile–Ala–ARG–ARG–His–Pro–Tyr– –Glu–Tyr–Ala–ARG–ARG–His–Pro–Asp– –Arg–Tyr–Thr–LYS–LYS–Val–Pro–Gln–
Proparathyrin (Bovine)	5 –Ser–Val–Lys–LYS–ARG▼–Ala–Val–Ser–	–Trp–Leu–ARG–LYS–LYS–Leu–Gln–Asp– –Arg–Pro–ARG–LYS–LYS–Glu–Asp–Asn–
Proglucagon (Porcine)	30 –Met–Asn–Thr–LYS–ARG▼–Asn–Lys–Asn–	–Leu–Asp–Ser–ARG–ARG–Ala–Gln–Asp–
Progastrin (Porcine)	15 –Asp–Pro–Ser–LYS–LYS▼–Gln–Gly–Pro–	
Proinsulin (Porcine)	30 –Pro–Lys–Ala–ARG–ARG▼–Glu–Ala–Glu– 65 –Pro–Pro–Gln–LYS–ARG▼–Gly–Ile–Val–	
Lipotropin (Porcine)	40 –Ala–Ala–Glu–LYS–LYS▼–Asp–Glu–Gly– 60 –Pro–Lys–Asp–LYS–ARG▼–Tyr–Gly–Gly– 90 –Asn–Ala–His–LYS–LYS▼–Gly–Gln–OH	

indicated sites. The observed selectivity must therefore result either from the existence of very specific converting enzymes or from the inaccessibility of some double basic sequences to these enzymes. This latter alternative appears unlikely since lysine and arginine invariably appear on the surface of globular proteins, particularly in this instance where they occur in pairs. A more acceptable explanation would be that unprocessed sequences occur in a rigid conformation, such as an α-helix, and are thus poor substrates for enzymes. This idea is supported by the known ability of trypsin in small amounts to release active peptides from precursors. It should be noted that the pancreatic enzyme has the potential to split at any site containing a basic residue, so that the activation sites of precursors must be uniquely susceptible to tryptic attack.

Although small polypeptides do not generally possess secondary or tertiary structure in solution, their higher molecular weight precursors probably do have defined folding. Only proparathyrin, proinsulin and proalbumin represent the complete precursor structures in table 1, and the latter two are known to be structured in dilute solution. In order to examine the hypothesis that unprocessed paired basic sequences occur in elements of secondary structure, α-helix and β-strands were located exactly from known crystal structures or to a good approximation by the empirical rules in [9]. The results are recorded in fig.1.

All the known precursor processing sites were

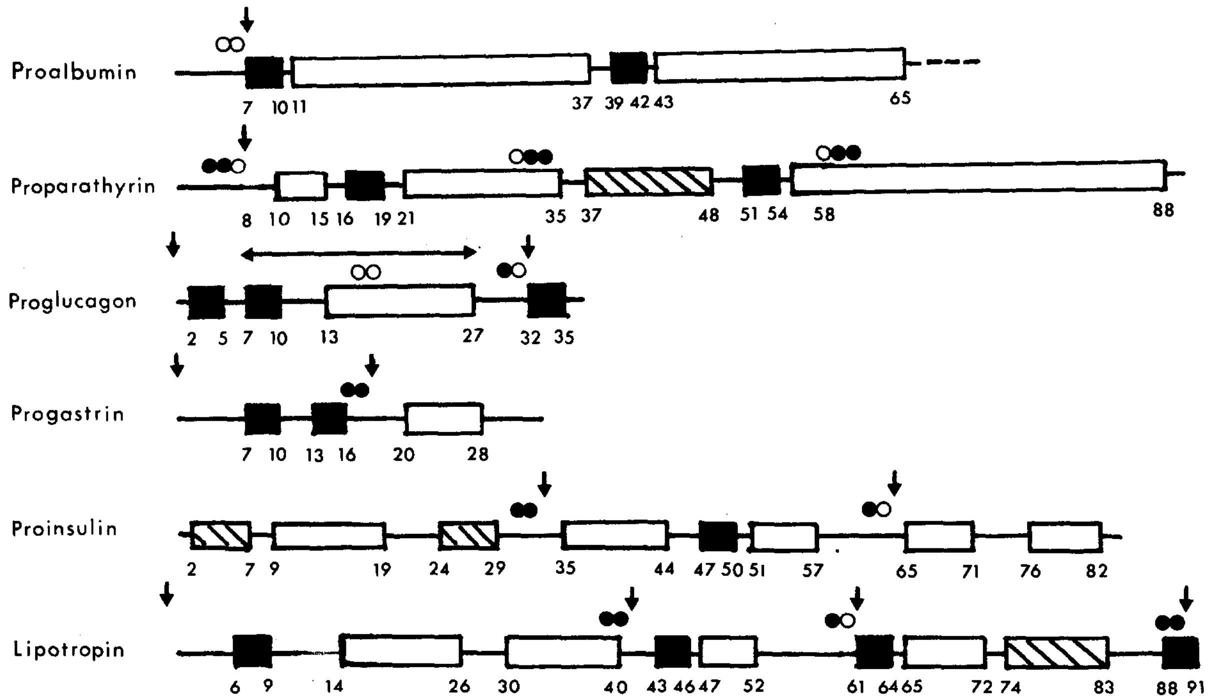

Fig.1. Secondary structure predicted in propolypeptides and hormones containing double basic residue sequences. Filled rectangles refer to β-turns, empty rectangles represent α-helix and hatched rectangles represent β-structure. The numbers indicate the boundaries of the secondary structural elements. Arrows mark known sites of precursor processing. These have been firmly established in the cases of lipotropin [8] and proinsulin [7]. Circles represent lysine (filled) and arginine (empty). The secondary structure shown for the A and B chains of insulin was taken from the published X-ray structure [17]. Predictions of the proinsulin structure are in [18]. The arrow above the glucagon structure indicates the extent of α-helix present in the crystal structure of glucagon [10]. The proalbumin NH_2 region only is shown because of the length of this molecule.

located in unstructured regions of the polypeptide chain or at an α-helix boundary. These sites were frequently found to be close to reverse turns in the polypeptide chain. On the other hand, paired basic sequences not split by processing were invariably found in predicted α-helical regions: –Arg–Lys–Lys– in proparathyrin, –Lys–Lys– and –Arg–Arg– sequences in proalbumin. The unprocessed –Arg–Arg– sequence in proglucagon is in the centre of a predicted α-helix. The same feature is present at this position in the glucagon crystal structure [10]. Of 9 unprocessed paired basic amino acid sequences in human albumin only 3 occur outside regions of predicted secondary structure [11]. However, these sites appear to be inaccessible to both trypsin [12] and pepsin [13,14].

On the basis of this model of precursor processing, one might predict that after formation, small hormones such as glucagon and lipotropin or its fragments would be further cleaved by the trypsin-like enzyme. These two hormones do not appear to have secondary structure in dilute solution. It is, however, a common feature of small polypeptide hormones that they become highly associated even at moderate concentrations. This association process is known to induce secondary structure in glucagon [15]. The rate of proteolysis of glucagon by trypsin was found to be markedly slowed as the concentration of the hormone was increased [22]. At the concentrations which occur in mature storage granules, insulin and glucagon actually crystallize [10]. Under such conditions it is evident that no further proteolysis can occur until the secretory event itself.

It is proposed that precursor processing is restricted

to relatively flexible regions of the folded molecules at double basic recognition sites. The polar nature of basic residues as well as the proximity of processing sites to reverse turns indicates that these sites have a superficial location, where they are accessible to enzymes. On the other hand, it is predicted that double basic sequences essential to the structure and function of the products of processing are restricted to regions of secondary structure. Although these potential sites may also be accessible to enzymes they are maintained in a rigid conformation by hydrogen bonding and thus represent a poor substrate for proteolytic enzymes.

The consistency of the above observations, as applied to the known precursors, demonstrates that empirical structural rules may be used to locate potential processing sites. This model of propolypeptide processing represents a considerable economy, since the specificity of conversion is achieved by precursor conformation and not by the elaboration of many individual converting enzymes. It will be interesting to see how this model can be applied as the sequences of more precursors become known.

Acknowledgements

I am grateful to Dr D. G. Smyth and Sir Arnold Burgen for valuable discussion.

References

[1] Neurath, H. (1975) in: Proteases and Biological Control (Reich, E., Rifkin, D. B. and Shaw, E. eds) pp. 51–64, Cold Spring Harbor Lab., Cold Spring Harbor, NY.
[2] Walsh, J. H. (1977) Fed. Proc. Fed. Am. Soc. Exp. Biol. 36, 1948–1951.
[3] Schafer, D. J. and Szelke, M. (1976) in: Amino-acids, peptides and proteins, vol. 8, (Sheppard, R. C. ed) The Chemical Society, Burlington House, London.
[4] Smyth, D. G. and Snell, C. R. (1977) FEBS Lett. 78, 225–228.
[5] Steiner, D. F., Kemmler, W., Tager, H. S., Rubenstein, A. H., Lernmark, A. and Zühlke, H. (1975) in: Proteases and Biological Control (Reich, E., Rifkin, D. B. and Shaw, E. eds) pp. 531–545, Cold Spring Harbor Lab., Cold Spring Harbor, NY.
[6] Bradbury, A. F., Smyth, D. G. and Snell, C. R. (1976) in: Polypeptide Hormones: Molecular and Cellular Aspects, CIBA Found. Symp. No. 41, pp. 61–75, Churchill, London.
[7] Kemmler, W., Peterson, J. D., Rubenstein, A. H. and Steiner, D. F. (1972) Diabetes 21, 572–583.
[8] Bradbury, A. F., Smyth, D. G. and Snell, C. R. (1976) Biochem. Biophys. Res. Commun. 69, 950–956.
[9] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–245.
[10] Blundell, T. L., Dockerill, S., Sasaki, K., Tickle, I. J. and Wood, S. P. (1976) Metabolism 25, 1331–1341.
[11] McLachlan, A. D. and Walker, J. E. (1977) J. Mol. Biol. 112, 543–558.
[12] Peters, T., Jr. and Feldhoff, R. C. (1975) Biochemistry 14, 3384–3391.
[13] Feldhoff, R. C. and Peters, T., Jr. (1975) Biochemistry 14, 4508–4514.
[14] Geisow, M. J. and Beaven, G. H. (1977) Biochem. J. 161, 619–625.
[15] Blanchard, M. H. and King, M. V. (1966) Biochem. Biophys. Res. Commun. 25, 298–303.
[16] Habener, J. F., Chang, H. T. and Potts, J. T., jr (1977) Biochemistry 16, 3910–3917.
[17] Blundell, T., Dodson, G., Hodgkin, D. and Mercola, D. (1972) Avan. Protein Chem. 26, 279–402.
[18] Snell, C. R. and Smyth, D. G. (1975) J. Biol. Chem. 250, 6291–6295.
[19] Dayhoff, M. O. ed, Atlas of Protein Sequence and Structure, vol. 1–5, Nat. Biomed. Res. Found. Washington DC.
[20] Fasman, G. D., Chou, P. Y. and Adler, A. J. (1976) Biophys. J. 16, 1201–1238.
[21] Chou, P. Y., Adler, A. J. and Fasman, G. D. (1975) J. Mol. Biol. 96, 29–45.
[22] Geisow, M. J. Unpublished results.