Precursors for peptide hormones share common secondary structures forming features at the proteolytic processing sites

Mohamed Rholam, Pierre Nicolas and Paul Cohen

Groupe de Neurobiochimie Cellulaire et Moléculaire, Université Pierre et Marie Curie, Unité Associée 554 au CNRS, 96 boulevard Raspail, 75006 Paris, France

Received 5 August 1986

We have analyzed the amino acid sequences situated around the putative proteolytic cleavage sites in twenty different biosynthetic precursors of peptide hormones by processing enzymes. The prediction of the probability for forming secondary structures around the basic amino acids, constituting the cleavage sites, was made using the modified method of Chou and Fasman. The results indicate that the processing sequences which are cleaved in vivo, are in all cases located inside regions with high β-turn formation probability or else immediately adjacent to these structures. The β-turn forming region at the cleavage locus, is flanked on both sides by amino acid sequences with a high probability for forming highly ordered structures, either β-sheet or α-helix. These conformational features are not found in precursors around dibasic pairs, i.e. putative cleavage loci, but which are not cleaved in vivo and appear to be conserved. We hypothesize that β-turns including the basic amino acids doublets, flanked by highly ordered secondary structures (either β-sheet or α-helix) may constitute a minimal requirement for the recognition by the endoproteases involved in the processing of these precursors.

Proteolytic processing *Protease* *β-turn*

1. INTRODUCTION

Activation of precursor molecules by proteolytic processing is a quite largely expanded mechanism both in lower and higher organisms. This appears to be essential in the replication of picornaviruses [1], the activation of either plasma proteins or zymogens [2]. In the case of secretory proteins, conversion of the biosynthetic precursor into the biologically active form(s) is dependent upon post-translational modifications including proteolytic cleavage performed by presumed selective proteases. Examination of the amino acid sequences of prepro-peptide hormones as derived from the corresponding cloned cDNAs nucleotide sequences indicated that the biologically active hormone sequence is always flanked, or preceded, by basic amino acids, Lys and/or Arg, often arranged as pairs or even quadruplets and sometime as a singlet or a triplet (see references in table 1). Since during processing these basic amino acids are completely removed from the precursor molecules and are generally no longer found in the 'mature' peptide hormones, this implies that an adequate and selective enzyme equipment can excise these amino acids and that they probably participate in recognition signals for these endoproteases. Examination of the amino acid sequences around these cleavage signals was unsuccessful in revealing the conservation of any 'consensus' primary sequence and so raised several questions with respect to the origin of selectivity for the relevant endoproteases.

Abbreviations: H, E and T, α-helix, β-sheet and β-turn structures; K and R, Lys and Arg amino acid residues, respectively

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

Recently a few of these enzymes possibly involved in the processing of pro-enkephalins [3–5], pro-opiomelanocortin [6], pro-ocytocin-neurophysin I [7] and pro-somatostatin [8] were characterized. Several types of cleavage occurring either before, after or else inside the pair of basic amino acids were observed by different authors. In some cases the use of peptide analogs [4,9] or the destruction of secondary structure by denaturation [10] strongly suggested that other structural parameters, beyond the lone basic amino acid arrangements, might play a role in the 'specificity' of the recognition by the relevant endoproteases. Among the above cited enzyme systems, while basic amino acid doublets are hydrolyzed, lone Arg or Lys residues in the precursors or hormone sequences are not. Substitution of one out of the two basic amino acids of a doublet by a neutral hydrophobic residue (Nle) or else its stereochemical modification (by D-Arg or D-Lys) were efficient in abolishing the protease action (this laboratory, unpublished).

To test for the possibility that the secondary structure around the cleavage signals may adopt given conformations to provide 'selective domains' to be recognized by the processing proteases we have evaluated the secondary structure forming potentialities of 94 sequences around excision sites enclosed in 20 polypeptide hormone precursors. The results allow us to formulate an hypothesis with respect to the existence of highly ordered domains at the processing sites.

2. MATERIALS AND METHODS

The secondary structure of different precursors was analyzed by a modification of Chou and Fasman's original method which uses the multiplication approach instead of the arithmetic [11]. α-helices and β-structures were deduced using tabulations of structure-formation parameters already determined [12]. Chou and Fasman's numerical method for β-turn was adopted [13].

3. RESULTS AND DISCUSSION

An inspection of the primary sequence of 20 pro-polypeptides did not reveal any stricking features in the sequences around the cleavage sites except that among the various types of basic doublets, the Lys-Arg pair (table 1) was found as the most abundant (54%). Therefore, it can be inferred that this observed selectivity of the proteolytic processing must result from the existence of specific enzymes (not as yet clearly defined) or else from the steric hindrance of some basic doublets, or both. To search for common structural features that could be related to this enzymatic processing, we have investigated the secondary structure of 20 precursors based upon the available primary sequence data.

The general analysis of 94 sets of basic amino acids (Lys or/and Arg), whether they were presumed to be cleaved, in vivo, or not, indicates that 82% of them are localized in reverse turns (table 1). According to their position in the precursors, with respect to the known hormonal sequences, these sets could be distributed among 3 sub-groups (either flanking, outside or inside the hormone sequence). The results indicate that among 66 sets of

Table 1

Secondary structure analysis of sequences around basic amino acids in proproteins

Precursor sequences analyzed	Predicted structures	Localization of basic residues
PRO-ENKEPHALIN A (26)		
98 - 103	T	K(98)-K(99)
105 - 110	T	K(105)-R(106)
111 - 118	H	K(112)-K(113)
134 - 139	T	K(134)-R(135)
140 - 144	H	K(141)-K(142)
175 - 191	T	K(186)-R(187)
196 - 201	T	K(197)-R(198)
210 - 216	T	K(210)-R(211)
		R(217)-R(218)
230 - 235	T	K(230)-R(231)
236 - 242	H	K(237)-R(238)
261 - 266	T	K(261)-R(262)
PRO-OPIOMELANOCORTIN (27)		
(-89)-(-55)	T	R(-57)-R(-56)
(-47)-(-32)	T	R(-43)-R(-42)
(-31)-(-22)	H	K(-28)-R(-27)
(-21)- 3	T	K(- 2)-R(- 1)
10 - 28	T	K(15)-K(16)-R(17)-R(18)
29 - 43	H	K(40)-R(41)
83 - 89	T	K(82)-K(83)
95 - 107	T	K(102)-R(103)
117 - 206	H	K(131)-K(132)

Table 1 (contd)

Precursor sequences analyzed	Predicted structures	Localization of basic residues
PRO-THYREOLIBERIN (28)		
48 - 52	T	R(51)-R(52)
68 - 77	H	K(75)-R(76)
78 - 83	T	K(81)-R(82)
104 - 109	H	K(107)-R(108)
110 - 114	T	R(113)-R(114)
151 - 154	H	K(152)-R(153)
155 - 161	T	R(158)-R(159)
162 - 173	H	K(170)-R(171)
173 - 177	T	R(176)-R(177)
198 - 202	T	K(200)-R(201)
205 - 210	T	K(206)-R(207)
PRO-TACHYKININ (29)		
58 - 62	T	R(57)
69 - 77	T	K(70)-R(71)
99 - 102	T	K(96)-R(97)
108 - 111	T	K(109)-R(110)
123 - 127	T	R(127)-R(128)-R(129)-K(130)
PRO-RELAXIN (30)		
25 - 31	T	K(29)-R(30)
94 - 108	T	K(105)-K(106)
126 - 137	T	K(134)-K(135)-R(136)-R(137)
151 - 154	T	K(153)-R(154)
PRO-VIP/PHM (31)		
71 - 91	T	R(80)
92 - 110	H	K(100)-K(101) K(109)-R(110)
111 - 121	T	K(123)-R(124)
136 - 152	H	R(138)-K(139) K(144)-K(145)
153 - 160	T	K(154)-R(155)
PRO-INSULIN-LIKE GROWTH FACTOR II (32)		
21 - 42	T	R(37)-R(38)
60 - 66	T	R(68)
95 - 136	H	R(103)-R(104) R(113)-R(114) K(129)-R(130)
137 - 180	T	R(179)-K(180)
PRO-RENIN (33)		
27 - 59	H	K(34)-K(35)
60 - 64	T	K(62)-R(63)
298 - 318	H	K(310)-R(311)
332 - 354	T	R(352)-R(353)
373 - 401	H	R(382)-K(383)
PRO-GLUCAGON (34)		
53 - 65	T	K(51)-R(52)
66 - 78	H	R(69)-K(70)
79 - 84	T	K(82)-R(83)
90 - 101	T	K(89)-R(90)
112 - 123	H	R(122)-R(123)
PRO-CORTICOTROPIN-RELEASING FACTOR (35)		
121 - 131	T	R(123)-R(124)
154 - 160	T	R(152)-R(153)
186 - 189	T	R(188)-K(189)
190 - 196	H	R(196)
PRO-GROWTH HORMONE-RELEASING FACTOR (36)		
27 - 38	H	R(30)-R(31)
39 - 43	T	R(42)-K(43)
44 - 62	H	R(51)-K(52)
72 - 77	T	R(77)
PRO-INSULIN (37)		
19 - 31	T	R(31)-R(32)
53 - 66	T	K(64)-R(65)
PRO-PARATHYROID HORMONE (38)		
(-10)-(- 5)	T	K(-3)-K(-2)-R(-1)
14 - 41	H	R(25)-K(26)-K(27)
42 - 54	T	R(52)-K(53)-K(54)
PRO-UROTENSIN (39)		
100 - 109	T	K(101)-R(102)
136 - 141	T	R(137)-K(138)
142 - 145	T	K(145)
PRO-SOMATOSTATIN (22)		
(-16)- 2	T	R(-15) R(-2)-K(- 1)
PRO-VASOPRESSIN-NEUROPHYSIN II (40)		
6 - 13	T	K(11)-R(12)
101 - 107	E	R(105)-R(106)
108 - 123	T	R(108)
PRO-OCYTOCIN-NEUROPHYSIN I (40)		
6 - 13	T	K(11)-R(12)
PRO-CALCITONIN (23)		
72 - 87	T	K(83)-R(84)
115 - 120	T	K(118)-K(119)-R(120)
PRO-CALCITONIN GENE RELATED PEPTIDE (23)		
82 - 88	T	K(81)-R(82)
110 - 123	T	R(121)-R(122)-R(123)
PRO-GONADOTROPIN-RELEASING HORMONE (41)		
(-11)-(- 1)	T	K(-2)-R(-1)
49 - 53	T	K(54)-K(55)

basic amino acids, flanking the hormone sequences, all of them are directly included in reverse turns or immediately adjacent to such structures (at -1 or -2 amino acid). Therefore it can be inferred that the situation of cleavage loci signals in β-turns is a minimal requirement for the recognition by convertases or maturases.

In contrast, among 17 sets of basic residues situated inside the hormone sequences, 10 were found either in an α-helix or a β-sheet. They could be expected to be preserved for cleavage if these structures are conserved after release of the known hormone sequence. The remaining 7 sets were found in the region with β-turn forming capacity. In vivo data about the possible role of these signals for proteolytic cleavage are still lacking. On the basis of these results, 92% out of the 83 above described sets are to be predicted in suitable structures by this method. Finally, 11 sets of basic amino acids were localized outside the hormone sequence in the precursors. If we assume the above hypothesis, 54% of them could be predicted as cleaved because situated in turn structures while the remaining 46% should not be.

We can predict that the presence of basic pairs into high β-turn forming capacity sequences is compatible with recognition by selective endoproteases and that they are consequently expected to be cleaved under adequate conditions. This is supported by the observations, from literature data, that this prerequisite is fulfilled and that such loci are indeed cleaved.

This general tendency of cleavage sites, to be located in a β-turn, is in agreement with the functions attributed generally to these structures [14]: i.e. (i) they provide a flexible region to the cleavage sites and therefore allow accessibility for enzymatic processing and (ii) in cases where the β-turn is adjacent to the cleavage sites, these regions may provide a structure in the nascent polypeptide chain to trigger the processing. On another hand, the unprocessed sequences appear to be included in a rigid conformation (α-helix or β-sheet structures) and thus represent poor substrates for proteolytic enzymes. However, after release from precursors, small hormones, such as VIP, glucagon or lipotropin, do not appear to possess secondary structure in dilute solutions [15]. Under such conditions, they could be cleaved at their internal basic residues. On the basis of these data, one might suggest that: (i) to preserve their full sequences from cleavage, the hormones self-associate to induce their secondary structure [16] or (ii) these active peptides are indeed cleaved in vivo under certain conditions but their fragments still remain to be characterized.

Considering 57 analyzed domains (table 2), corresponding to loci cleaved in vivo, it can be calculated that the probability for predicting the structure H/T/H is the highest (48%) while the others (H/T/E; E/T/H; E/T/E) rank respectively for 18%, 24% and 10%. These values were calculated assuming that certain domains may be predicted to adopt either an α-helix or a β-sheet conformation. Since α-helices can be expected to be generally more exposed than β-sheets in the tertiary structure of proteins [17,18], this implies that the sequential order in cleavage may be determined by the above defined type of interactions.

Out of the 48 analyzed doublets (table 2), known to be cleaved in vivo, 73% were of the type Lys-Arg. They themselves are distributed among H/T/H (56%), H/T/E (18%), E/T/H (22%) and E/T/E (4%). In a similar way, if we compare the localization of Lys-Arg pairs, in a given structure, to other doublets (Arg-Lys, Lys-Lys and Arg-Arg), roughly 70–75% of the doublets belong to the Lys-Arg type. This indicates that no mutual relation-

Table 2

Relative percentage of various basic amino acids combination in different secondary structures

Type of secondary structures	% of basic residues cleaved in vivo	% of K-R related to other doublets	% of K-R found in each structure
H/T/H	48%	75%	56%
H/T/E	18%	73%	18%
E/T/H	24%	72%	22%
E/T/E	10%	70%	4%

ships exist between Lys-Arg doublets and any secondary structure, except that the predominance of Lys-Arg doublets in H/T/H structures may explain the preferential cleavage of Lys-Arg doublets in multifunctional precursors.

As a corollary from such observations, it can be proposed that (i) the basic doublet is necessary but it does not appear to play a critical role in specifying the type of secondary structure flanking the β-turn, (ii) the degree of exposure of a processing domain depends upon its secondary structure. Therefore, the unfolding of the precursor in a sequentially ordered process may regulate the cleavage at the basic pairs. (iii) Only a few proteolytic enzymes may participate in the processing of these peptide hormone precursors. For example, the somatostatin-28 convertase of the rat brain cortex [7,19] while active in converting somatostatin-28 (S-28) into its S-28 (1–12) and S-14 NH_2- and COOH-terminal domains could not release the hormones from their 12 kDa common precursor [19]. Similarly, in the case of pro-opiomelanocortin, processing products appear in vivo in a given order [20]. This implies that the basic amino acids at the cleavage loci are necessary for the hydrolysis to occur but they can hardly explain both the enzyme specificity and the sequence of events.

Comparative examination of the primary structures of the cleavage domains in the series of both pro-ocytocin-neurophysin I and pro-vasopressin-neurophysin II among 6 animal species [21] reveals that some variations of the amino acid sequences occur at position 8, 14, 15, 16, 18, 19 and 21 of either precursors (i.e. amino acid 8 of vasopressin and residues 2–4, 6, 7 and 9 of neurophysins). Nevertheless prediction for secondary structure formation indicates the conservation of a β-sheet/β-turn/α-helix domain (this laboratory, unpublished). Similarly, comparison of the amino acid sequences of several pro-somatostatin-28 molecules indicates high variation in the S-28 (1->12) NH_2-terminal fragment preceding the Arg-Lys pair. But these variances in amino acid sequences have no significant effect on the conservation of a highly ordered secondary structure [22]. The same conclusion applied (this laboratory, unpublished) when comparing pro-calcitonin and pro-CGRP [23] through various species.

Finally, in the case of the pro-vasopressin-neurophysin II and pro-ocytocin-neurophysin I genes [24] for instance, it can be observed that a single exon (exon-I) encodes for the whole NH_2-terminal domain of the precursor (1->21) including the nonapeptide hormone sequence, the processing sequence and the first nine amino acids of neurophysins [24,6]. Preliminary examination of the gene structure for a few prohormones [25] indicates that in all cases the hormone(s) sequence(s) is/are coded by a single exon which also codes for the mRNA specifying the processing domains either preceding, following or flanking the latter. Further elucidation of other prohormones gene structures will indicate if this is, as can be expected, a general rule.

ACKNOWLEDGEMENTS

This work was supported by funds from the Université Pierre et Marie Curie, the Centre National de la Recherche Scientifique (UA 554 and ARI Chimie-Biologie), the Fondation pour la Recherche Médicale Française, the Mutuelle Générale de l'Education Nationale and the Institut National de la Santé et de la Recherche Médicale (CRE 834006).

REFERENCES

[1] Nicklin, M.J.H., Toyoda, H., Murray, M.G. and Wimmer, E. (1986) Biotechnology 4, 33–36.
[2] Barrett, A.J. (1979) Proteinases in Mammalian Cells and Tissues, North Holland, Amsterdam.
[3] Lindberg, I., Yang, H.Y.T. and Costa, E. (1982) Biochem. Biophys. Res. Commun. 106, 186–193.
[4] Mizuno, K., Kojima, M. and Matsuo, H. (1985) Biochem. Biophys. Res. Commun. 128, 884–891.
[5] Nyberg, F., Nordstrom, K. and Terenius, L. (1985) Biochem. Biophys. Res. Commun. 131, 1069–1074.
[6] Loh, Y.P., Parish, D.C. and Tujeta, R. (1985) J. Biol. Chem. 260, 7194–7205.
[7] Clamagirand, C., Camier, M., Boussetta, H., Fahy, C., Morel, A., Nicolas, P. and Cohen, P. (1986) Biochem. Biophys. Res. Commun. 134, 263–318.
[8] Gluschankof, P., Morel, A., Benoit, R. and Cohen, P. (1985) Biochem. Biophys. Res. Commun. 128, 1051–1057.
[9] Gomez, S., Gluschankof, P., Morel, A. and Cohen, P. (1985) J. Biol. Chem. 260, 10541–10545.
[10] Miura, S., Amaya, Y. and Mori, M. (1986) Biochem. Biophys. Res. Commun. 134, 1151–1159.

[11] Dufton, M.J. and Hider, R.C. (1977) J. Mol. Biol. 115, 177–193.
[12] Levitt, M. (1979) Biochemistry 17, 4277–4285.
[13] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[14] Rose, G.D. and Roy, S. (1980) Proc. Natl. Acad. Sci. USA 77, 4643–4647.
[15] Maher, M.H., Robinson, R.M. and Mattice, W.L. (1983) Proc. Natl. Acad. Sci. USA 79, 345–349.
[16] Blanchard, M.H. and King, M.V. (1966) Biochem. Biophys. Res. Commun. 25, 298–303.
[17] Kanehisha, M.I. and Tsong, T.Y. (1980) Biopolymers 19, 1617–1628.
[18] Ghelis, K. and Yon, J. (1982) in: Protein Folding, pp. 503–569. Academic Press, New York.
[19] Gluschankof, P., Morel, A., Gomez, S., Nicolas, P., Fahy, C. and Cohen, P. (1984) Proc. Natl. Acad. Sci. USA 81, 6662–6666.
[20] Mains, R.E. and Eipper, B.A. (1981) J. Biol. Chem. 256, 5683–5688.
[21] Marrakchi, N. (1985) Thèse de 3ème Cycle, Faculté des Sciences de Tunis, Tunisia.
[22] Argos, P., Taylor, W.L., Minth, C.D. and Dixon, J.E. (1983) J. Biol. Chem. 258, 8788–8793.
[23] Lasmoles, F., Julienne, A., Day, F., Minvielle, S., Milhaud, M. and Mokhtar, M.S. (1985) EMBO J. 4, 2603–2607.
[24] Schmale, H., Heinsohn, S. and Richter, D. (1983) EMBO J. 2, 763–767.
[25] Douglas, J., Civelli, O. and Herbert, E. (1984) Annu. Res. Biochem. 53, 665–715.
[26] Howells, R.D., Kilpatrick, D.L., Bhatt, R.S., Monahan, J.J., Poonian, M.S. and Udenfriend, S. (1984) Proc. Natl. Acad. Sci. USA 81, 7651–7655.
[27] Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S.N. (1979) Nature 278, 423–427.
[28] Lechan, R.M., Wu, P., Jackson, I.M.D., Wolf, H., Cooperman, S., Mandel, G. and Goodman, R.H. (1986) Sciences 231, 159–161.
[29] Nawa, H., Hirose, T., Takashima, H., Inayama, S. and Nakanishi, S. (1983) Nature 306, 32–36.
[30] Hudson, P., Haley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J. and Niall, H. (1983) Nature 301, 628–631.
[31] Itoh, N., Obata, K., Yanaihara, N. and Okamoto, H. (1983) Nature 304, 547–549.
[32] Bell, G.I., Merryweather, J.P., Sanchez-Pescador, R., Stempien, M.M., Priestley, L., Scott, J. and Rall, L.B. (1984) Nature 310, 775–777.
[33] Panthier, J.J., Foote, S., Chambraud, B., Strosberg, A.D., Corvol, P. and Rougeon, F. (1982) Nature 298, 91–92.
[34] Lund, P.K., Goodman, R.H., Dee, P.C. and Habener, J.H. (1982) Proc. Natl. Acad. Sci. USA 79, 345–349.
[35] Shibahara, S., Morimoto, Y., Furutani, Y., Notake, M., Takahashi, H., Shimizu, S., Horikawa, S. and Numa, S. (1983) EMBO J. 2, 775–779.
[36] Gubler, U., Monahan, J., Lomedico, P.T., Bhatt, R.S., Collier, K.J., Hoffman, J., Bohlen, P., Esch, F., Ling, N., Zeytin, F., Brazeau, P., Poonian, M.S. and Gage, L.P. (1983) Proc. Natl. Acad. Sci. USA 80, 4311–4314.
[37] Bell, G.I., Swain, W.F., Pictet, R., Cordell, B., Goodman, H.M. and Rutter, W.J. (1979) Nature 282, 525–527.
[38] Kronenberg, H.M., McDevitt, B.E., Majzoub, J.A., Nathans, J., Sharp, P.A., Potts, J.T., jr and Rich, A. (1979) Proc. Natl. Acad. Sci. USA 76, 4981–4985.
[39] Ishida, I., Ichikawa, T. and Deguchi, T. (1986) Proc. Natl. Acad. Sci. USA 83, 308–312.
[40] Land, H., Grez, M., Ruppert, S., Schmale, H., Rehbein, M., Richter, D. and Schutz, G. (1983) Nature 302, 342–344.
[41] Nickolics, K., Mason, A.J., Szonyi, E., Ramachandran, J. and Seeburg, P.H. (1985) Nature 316, 511–517.