ON THE PRIMARY AND TERTIARY STRUCTURE OF RELAXIN FROM THE SAND TIGER SHARK (*Odontaspis taurus*)

L. K. GOWAN, J. W. REINIG, C. SCHWABE, S. BEDARKAR[+] and T. L. BLUNDELL[+]

Department of Biochemistry, Medical University of South Carolina, Charleston, SC 29425, USA and [+]*Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England*

Received 5 May 1981

1. Introduction

The structural variability between the two known relaxin sequences from the pig [1–3] and rat [4] far exceed that observed for any insulin pair, yet their primary sequences can readily be folded into an insulin conformation [5,6]. This raises the question of how much more variability in the primary sequence may be tolerated without preventing an insulin-like folding to occur and how much of the surface may be varied without loss of receptor interaction. The evolutionary diversion of relaxins from each other and the possible existence of a common ancestral gene for relaxins and insulins led us to examine the structure of shark relaxin. Based on the assumption that sharks have existed in their present form for at least 500 million years, i.e., have lived close to the point at which fishes and the ancestors of mammals are assumed to have branched from each other and where the insulin gene might have undergone duplication to give rise to the relaxin gene, we expected to observe a closer relation between the relaxin of this quasi-prehistoric species and insulins.

2. Materials and methods

Ovaries from the sand tiger shark (*Odontaspis taurus*) were obtained with the help of the Charleston Shark Fishing Club. The ovaries were removed from pregnant animals (pup size from 8–10 in.) and immediately placed on ice. The extraction of relaxin was performed as in [7] and relaxin fractions were identified by the bioassay in guinea pigs [8]. The primary structure analysis was performed on a Beckmann 809TC automatic sequencer using a 0.1 M quadrol program no. 122974 and polybrene to aid retention of peptides in the sequencer cup. PTH residues were identified by high-performance liquid chromatography (HPLC) using a Waters liquid chromatograph system by the procedure in [9].

Shark relaxin was reduced and alkylated with iodo-[^{3}H]acetic acid in 6 M guanidine–HCl. The chains were subsequently separated by HPLC as in [7].

For real-time display of relaxin models the Evans and Sutherland picture system II was employed. We have used the computer program FRODO [10] that was extensively modified and recoded for the Evans and Sutherland picture system by T. A. Jones and I. J. Tickle. The main-chain and disulphide conformations were assumed to be identical with those of insulin [12].

3. Results and discussion

The complete sequence of the A chain of the shark relaxin was obtained by automatic sequencing of ~20 nmol material. Sequential yields remained as high as 85–90% up to the last 3 residues when the yield dropped as expected. However, no ambiguities arose and the sequence was confirmed by amino acid analysis and automatic sequence analysis of tryptic peptides obtained from a second sample of the A chain.

The B chain was much more difficult to obtain in pure form at reasonable yields than that of porcine relaxin. The N-terminus was blocked by a PCA residue that could not be removed from the whole chain because of solubility problems. The PCA peptide was

Elsevier/North-Holland Biomedical Press

RELAXIN SEQUENCES

A-CHAIN	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Porcine					Arg	Met	Thr	Leu	Ser	Glu	Lys	Cys	Cys	Glu	Val	Gly	Cys	Ile	Arg	Lys	Asp	Ile	Ala	Arg	Leu	Cys
Rat			PCA	Ser	Gly	Ala	Leu	Leu	Ser	Glu	Gln	Cys	Cys	His	Ile	Gly	Cys	Thr	Arg	Arg	Ser	Ile	Ala	Lys	Leu	Cys
Shark			Ala	Thr	Ser	Pro	Ala	Met	Ser	Ile	Lys	Cys	Cys	Ile	Tyr	Gly	Cys	Thr	Lys	Lys	Asp	Ile	Ser	Val	Leu	Cys

B-CHAIN	-7	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Porcine								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	Trp			
Rat	Arg	Val	Ser	Glu	Glu	Trp	Met	Asp	Gln	Val	Ile	Gln	Val	Cys	Gly	Arg	Gly	Tyr	Ala	Arg	Ala	Trp	Ile	Glu	Val	Cys	Gly	Ala	Ser	Val	Gly	Arg	Leu	Ala	Leu	
Shark								PCA	-	-	-	-	-	-	-	Leu	Cys	Gly	Arg	Gly	Phe	Ile	Arg	Ala	Ile	Ile	Phe	Ala	Cys	Gly	Gly	Ser	Arg			

Numbering as for insulin

Residues underlined are identical to insulin

Fig.1. Amino acid sequences of relaxin from hog, rat and shark. Blocks of identity are marked and positions of identity of hog relaxin and hog insulin are underlined.

later lost after trypsin digestion as was a possible extension of the B chain beyond Arg 23. The remaining peptides Leu–Cys–Gly–Arg, Gly–Phe–Ileu–Arg and Ala–Ileu–Ileu–Phe–Ala–Cys–Gly–Gly–Ser–Arg were isolated by HPLC and the primary sequence determined by automatic sequence analysis. The peptides were ordered by analogy to obtained sequences. The primary structure of the core of shark relaxin is displayed in fig.1 together with the sequences of rat and pig relaxin. Because of the seasonal migration of *Odontaspis taurus* the additional ovarian tissue required to complete the analysis of the shark relaxin will not be available for several months. The portion of the structure presented is that which is primarily responsible for the folding of the molecule and thus of greatest interest. Blocks of homology extending over all 3 chains are indicated in fig.1. Invariant positions are by and large occupied by amino acids that are important for chain folding and crosslinking. All residues forming the hydrophobic core in insulin are also hydrophobic in the relaxins and all cysteines are invariant as are the glycines at B8 and B20 which allow a main chain conformation similar to that of insulin as shown in stereo view in fig.2. One of the more surprising aspects is the existence of B11 Phe and B12 Ileu which are Leu and Val, respectively, in porcine relaxin as well as in all insulins. This part of the structure is shown in fig.3. Similar changes occur in rat relaxins where B11 and B12 are tyrosine and alanine, respectively. In fact the B11 can be accommodated as an aromatic residue by allowing the side-chain to point into the region occupied by tryptophan

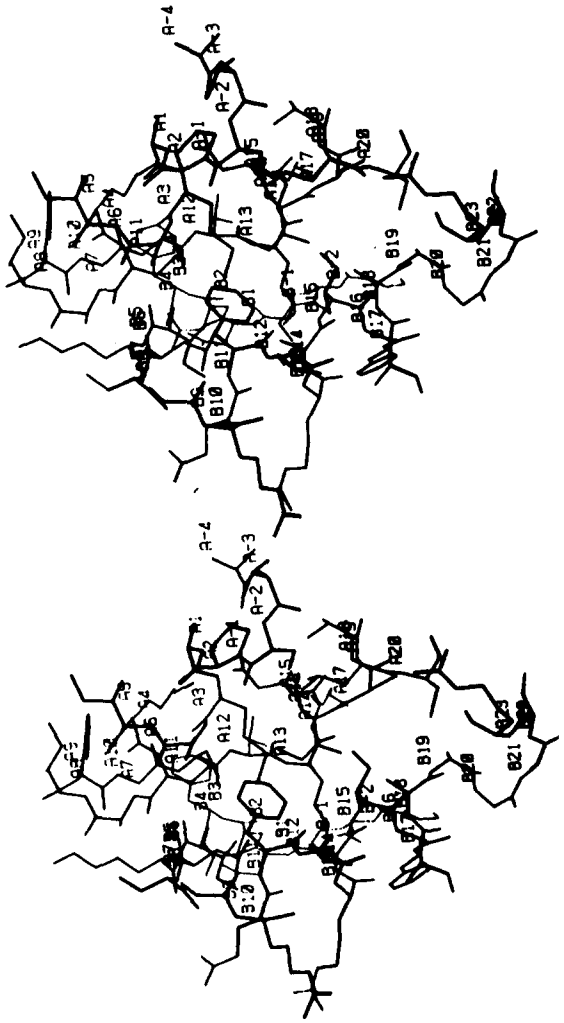

Fig.2. Stereo view of shark relaxin.

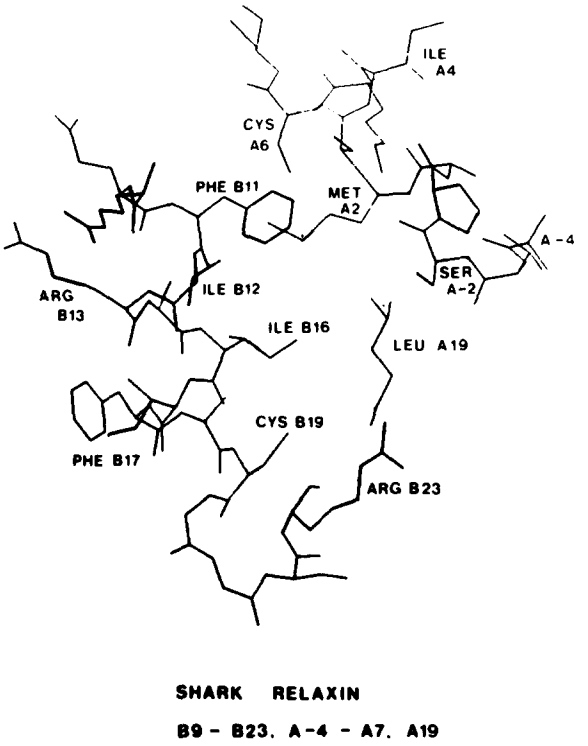

Fig.3. Surface residues of shark relaxin viewed at right angles to the axis of the A chain N-terminal helix.

in porcine relaxin. Perhaps the rat relaxin sequence is most surprising in this respect in having aromatic residues at both positions, B11 and B15. It is interesting that in shark relaxin B6 Leu and B14 Ala are identical to most insulin sequences. The ion pairs observed in porcine relaxin [5] are conserved in shark relaxin at A5 Lys and A15 Asp only and in rat relaxin at B13 Arg and B17 Glu only.

Our studies of relaxin structures suggest that there exists a stringent requirement to maintain the tertiary structure in spite of the total absence of common immunological determinants and active sites. The maintenance of the three-dimensional structure provides an interesting contrast to the maintenance of active site configuration as exemplified by the differences between subtilisin and mammalian serine proteases of the trypsin family [11].

Because of the variability of the primary structure of relaxin it is difficult if not impossible to predict which of the relaxin surfaces are important for bioactivity. However, we note with interest the conservation of positively-charged residues, arginines or lysines, at positions B9, B13, A13 and A14. While it was suggested by us that the Trp B15 was important for relaxin activity we note that shark relaxin does not possess this residue. Shark relaxin is not active in mice but fully active in guinea pigs suggesting that variability of the relaxin surfaces (Trp B15, for example), might be tolerable because of receptor variability far beyond that observed for insulin receptors. The selective force for the evolution of biomolecules is a function which appears to depend, in the case of relaxin, on the three-dimensional projection of the various protein folds. If the prediction of highly variable receptors for relaxin is incorrect, then the next 2 or 3 sequences of relaxins from different species should allow us to outline the various sites of interaction with antibodies and receptors as it has been done successfully for insulin.

Acknowledgements

The expert technical assistance of Mr Hal Crow, Ms Nancy McKnight and Ms Mary Berry are gratefully acknowledged. This work was supported by grant HD10540 from the National Institute of Child Health and Human Development, National Institutes of Health, and the Kroc Foundation.

References

[1] Schwabe, C., McDonald, J. K. and Steinetz, B. G. (1976) Biochem. Biophys. Res. Commun. 70, 397–405.
[2] Schwabe, C., McDonald, J. K. and Steinetz, B. G. (1977) Biochem. Biophys. Res. Commun. 75, 503–510.
[3] James, R., Niall, H. D., Kwok, S. and Bryant-Greenwood, G. (1977) Nature 267, 544–546.
[4] John, M. J., Borjesson, B. W., Walsh, J. R. and Niall, H. D. (1981) Endocrinology 108, 726–729.
[5] Bedarkar, S., Turnell, W. G., Blundell, T. L. and Schwabe, C. (1977) Nature 270, 449–451.
[6] Isaacs, N., James, R., Niall, H. D., Bryant-Greenwood, G., Dodson, G., Evans, A. and North, A. C. T. (1978) Nature 271, 278–281.
[7] Reinig, J. W., Daniel, L. N., Schwabe, C., Gowan, L. K., Steinetz, B. G. and O'Byrne, E. M. (1981) Endocrinology, in press.
[8] Kroc, R. L., Steinetz, B. G. and Beach, V. L. (1959) Ann. NY Acad. Sci. 75, 942–980.
[9] Bhown, A. S., Mole, J. E., Weissinger, A. and Bennett, J. C. (1978) J. Chromatog. 148, 532–535.
[10] Jones, T. A. (1978) J. of Applied Crystallography 11, 268–272.
[11] Hartley, B. S. (1970) Phil. Trans. Roy. Soc. London B257, 77–87.
[12] Blundell, T. L. and Humbel, R. E. (1980) Nature 287, 781–787.