

Fig. 3 Schematic enzyme mechanisms for intracellular motility cycles⁷⁻⁹ comparing the actomyosin and dynein ATPase cycles (a) with the vesicle transport cycle (b). a, 'ATPase' refers to either myosin or dynein and T refers to the translocatable partners (that is, filamentous actin or the B subfibril). NTP refers to nucleotide (for example, ATP or AMP-PNP) and Pr refers to the products of ATP hydrolysis. The drive stroke for myosin and dynein occurs after step 3. b, 'ATPase' refers to the putative ATPase in the crossbridge that mediates the attachment of the translocatable vesicles (V) to the microtubules (MT)¹⁵. We propose that ATP modulates the affinity of the ATPase-vesicle complex for the substrate microtubule. This hypothesis is supported by the observation that vesicles prepared from squid axoplasm retain their motile ATPase and can translocate along flagellar microtubules¹⁷. For simplicity the scheme is drawn as if the vesicle transport ATPase has a stable association with the vesicle. However, the ATPase can apparently dissociate from the vesicle. The step at which products (Pr) are released remains to be determined in both (a) and (b). Note that the effect of the binding of nucleotide (NTP) in the actomyosin and dynein cycles (a) is the opposite of its effect in the vesicle-transport cycle (b).

1-2 min of ATP addition, and vesicles began to move in the centre of the axoplasm 10 min after ATP addition.

Based on these results, we propose that AMP-PNP operates by competitively attaching to a critical ATP binding site in the motility machinery of vesicle transport. Miller and Lasek¹⁵ have recently proposed that the critical ATP binding site is an ATPase that is located on the 16-18-nm-long crossbridges that attach the transported vesicle to its substrate microtubule. Thus, like the actomyosin system and the dynein-microtubule system, motility in the vesicle transport system may be based on an ATP-dependent crossbridge cycle.

ATP has two important functions in the crossbridge cycle of muscles and axonemes⁷⁻⁹; it provides energy for movement and reduces the affinity of the crossbridge (that is, myosin or dynein) for its motile partner (that is, actin or tubulin) (Fig. 3a). In muscle, the biochemical cycle producing movement begins with the binding of ATP to the stable actin-myosin (rigor) complex⁹ which induces a rapid release of the myosin crossbridge from actin. Myosin hydrolyses the ATP, producing an activated ADP-myosin complex that has a much higher affinity for actin than the ATP-myosin complex. When the activated ADP-myosin complex rebinds to actin, it releases its stored energy in the motile power stroke. A similar scheme has been proposed for axonemes^{7,8}.

The effects of AMP-PNP on vesicle transport suggest that the enzymatic cycle of vesicle transport differs in at least two important ways from the cycle proposed for muscle and axonemes. First, dynein and myosin ATPases are weakly inhibited by AMP-PNP, that is, AMP-PNP will not produce complete inhibition of the biochemical cycle at concentrations of ATP that support motility^{6,14}. By contrast, the vesicle transport ATPase is completely inhibited by AMP-PNP even in the presence of concentrations of ATP that readily support vesicle transport. Second, in both muscles and axonemes, nucleotide binding promotes the dissociation of the motility complex (Fig. 3a, step 1)^{6,10,13}. In contrast, AMP-PNP facilitates a stable association of vesicles with microtubules. This suggests that nucleotide binding (for example, ATP or AMP-PNP) to the vesicle-transport ATPase promotes the formation of the basic motility complex (the vesicle-crossbridge-microtubule complex; Fig. 3b, step 1), which is the opposite of the effect of nucleotide on myosin and dynein (Fig. 3a).

The scheme proposed for the vesicle-transport cycle in Fig. 3b is further supported by observations of vesicle movement in ATP-depleted axoplasm. When ATP is eliminated from squid axoplasm by conversion to AMP and pyrophosphate with the

enzyme apyrase, vesicle transport is inhibited and most vesicles are subject to a constrained brownian motion². Apparently, rigor complexes are not favoured in the vesicle-transport system in low ATP concentrations.

The structural organization of the vesicle transport system may be analogous in some respects to that of the actomyosin and the dynein-microtubule systems¹⁵⁻¹⁷. Significantly, in these three systems a crossbridge is the structural link between the stationary substrate and the translocated partner^{5,8,15}. However, although there may be structural analogies between these motile systems, the results presented here and others^{18,19} indicate that the enzymatic properties of the motile protein of vesicle transport differ significantly from the crossbridge proteins in both muscle and axonemes (that is, myosin and dynein, respectively).

Received 14 January; accepted 7 June 1985.

1. Brady, S. T., Lasek, R. J. & Allen, R. D. *Science* **218**, 1129-1131 (1982).
2. Brady, S. T., Lasek, R. J. & Allen, R. D. *Cell Motil.* **5**, 81-101 (1985).
3. Allen, R. D. *et al. J. Cell Biol.* **100**, 1736-1752 (1985).
4. Schnapp, B., Vale, R. D., Sheetz, M. P. & Reese, T. S. *Cell* **40**, 449-454 (1985).
5. Huxley, H. E. *Cold Spring Harb. Conf. Cell Prolif.* **3**, 115-128 (1976).
6. Eisenberg, E. & Greene, L. A. *Rev. Physiol.* **42**, 293-309 (1980).
7. Gibbons, I. R. *J. Cell Biol.* **91**, 107s-124s (1981).
8. Satir, P., Wais-Steider, J., Lebduska, S., Nasr, A. & Avolio, J. *Cell Motil.* **1**, 303-327 (1981).
9. Lynn, R. W. & Taylor, E. W. *Biochemistry* **10**, 4617-4624 (1971).
10. Mitchell, D. R. & Warner, F. D. *J. Biol. Chem.* **256**, 12535-12544 (1981).
11. Yount, R. G., Babcock, D., Ballantyne, W. & Ojala, D. *Biochemistry* **10**, 2484-2489 (1971).
12. Greene, L. E. & Eisenberg, E. *J. Biol. Chem.* **255**, 543-548 (1980).
13. Reedy, M. C., Reedy, M. K. & Goody, R. S. *J. Muscle Res. Cell Motil.* **4**, 55-81 (1983).
14. Okuno, M. & Brokaw, C. J. *J. Muscle Res. Cell Motil.* **2**, 131-140 (1981).
15. Miller, R. H. & Lasek, R. J. *Biol. Bull.* **167**, 503-504 (1984).
16. Smith, D. S. *Phil. Trans. R. Soc. B261*, 395-405 (1971).
17. Gilbert, S. P., Allen, R. D. & Sloboda, R. D. *Nature* **315**, 245-248 (1985).
18. Forman, D., Brown, K. J., Livengood, D. R. *J. Neurosci.* **3**, 1279-1288 (1983).
19. Forman, D., Brown, K. J. & Promersberger, M. E. *Brain Res.* **272**, 194-197 (1983).
20. Lasek, R. J. *Curr. Topics Membranes Transport* **22**, 39-53 (1984).

Primary structure of bovine thyroglobulin deduced from the sequence of its 8,431-base complementary DNA

Luc Mercken, Marie-Jeanne Simons, Stéphane Swillens, Marc Massaer & Gilbert Vassart

Institut de Recherche Interdisciplinaire et Service de Chimie Clinique, Faculté de Médecine, Université Libre de Bruxelles, Campus Erasme, 808 route de Lennik, 1070 Bruxelles, Belgium

In mammals, an adequate supply of thyroid hormones is essential for normal growth and neurological development. The biosynthesis of thyroid hormones involves an iodinated precursor protein, thyroglobulin, which may be considered an extreme example of a pro-hormone¹. Thyroglobulin is a dimeric glycoprotein of relative molecular mass (M_r) 660,000 (660K), which is secreted by the thyrocyte and stored in the lumen of the thyroid follicle. The hormonogenic reaction is extracellular, and involves iodination of tyrosyl residues of thyroglobulin and the intramolecular coupling of a subset of these into thyroxine (T_4) and triiodothyronine (T_3), which remain part of the polypeptide chain. Secretion of hormones results from the endocytosis of thyroglobulin followed by its complete hydrolysis in lysosomes. Considering that the maximum yield of hormones is ~6-8 per 660K protein², the whole process is apparently wasteful. However, the efficiency of thyroglobulin as a thyroid hormone precursor is extremely high when the supply of iodine is short; in such conditions, almost all the iodine incorporated is found in iodothyronines³. Hence it is suggested that the thyroglobulin structure has evolved to allow for the preferential and efficient iodination and coupling of the hormonogenic tyrosines. Here we report the complete primary structure of bovine thyroglobulin, derived from the sequence of its 8,431-base-pair complementary DNA. The 2,769-amino-acid sequence is characterized by a pattern of imperfect repeats derived from three cysteine-rich motifs. Four hormonogenic tyrosines have been precisely localized near the amino and carboxyl ends of the protein.

The complete nucleotide sequence of bovine thyroglobulin messenger RNA (Fig. 1) was obtained by sequencing six recombinant plasmids containing 99% of its cDNA⁴ and one genomic

GGCGAGCGCTTCTACCGCTTCTCTCCGAGCGCTCCGAA																																			
1	ATG	GCC	CTG	GCC	CTA	TGG	GTG	TTG	CTT	CTG	CTG	GAC	TTA	ATC	TTC	TTC	GTA	Leu	Ala	Leu	Asp	Leu	Ile	Cys	Leu	Ala	Ser	Gly	Ala	Val	Ala	Ala	Gln	Pro	
	Met	Ala	Leu	Pro	Cys	Glu	Glu	Leu	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Val	Ala	Gln	Pro
																																			10
91	CTC	GCG	CCA	TGT	GAG	CTG	CAG	AUG	GAG	AGG	GCT	TTT	CTG	AAG	GGA	GAA	TAC	TAC	CTC	GCC	CAG	TCC	GCC	GAG	GAT	GGC	AGC	TTT	CAG	ACT					
	Leu	Arg	Pro	Cys	Glu	Glu	Leu	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	
																																			40
181	GTG	CAG	TGC	GCG	AGG	GCC	TCC	TGC	TGG	GCT	GTG	GAT	GAT	GCG	AGA	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	GAT	
	Val	Gln	Cys	Glu	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	
																																			30
271	GCC	TGT	CTG	TGC	TTC	CAG	CTG	CAG	AGG	CAG	ATT	TTG	CTC	AGC	AGC	TAC	ATC	AAC	ACA	GCC	AGC	TCC	TAC	CTA	CCC	CAG	TGT								
	Ala	Cys	Leu	Ser	Phe	Phe	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	
																																			80
361	CAG	GAT	TCA	GCG	GAT	TAC	TGC	GCC	GTG	CAG	TGC	GAG	TGC	GAG	CGG	CGG	GAG	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	TGC	
	Ala	Cys	Leu	Ser	Phe	Phe	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	
																																			90
451	ACC	GCG	CAG	CAG	AGG	CGG	CGG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	
	Thr	Arg	Gln	Gln	Gln	Arg	Pro	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	
																																			150
541	CCC	CAG	TGT	TCC	CAG	GAT	GCG	GCG	TTC	AGG	GCG	GAT	Gln	Cys	TGC	AGG	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	
	Pro	Gln	Cys	Pro	Ser	Pro	Ala	Ala	Ala	Ala</																									

0691	ACT	TUG	TUG	AGA	CCG	GAC	GAC	TTC	GGA	GTT	CCG	TAT	GCC	GCA	CCC	CTG	GCG	GAG	AAA	CGC	TTC	CGG	GAC	GAC	CAC	TTC	Thr	Ser	Tyr	Lys	Pro	Val	Asp	Gln	Phe	Leu	Gly	Ala	Pro	Pro	Leu	Gly	Ala	Pro	Arg	Ala	Pro	Glu	His	Leu	1620	
0751	AGC	TUG	ACT	GCG	ACC	ACC	AGA	CCG	ASG	TGC	GAG	CCA	GCG	ATC	AUG	ACG	GCC	AGC	AGC	CTC	CCG	GCG	GTC	ACT	GAT	Thr	Asn	Tyr	Gly	Ser	Pro	Gln	Ala	Thr	Lys	Pro	Ala	Arg	Ala	Gly	Tyr	Pro	Pro	Pro	Gly	Ala	Gly	Ala	1620			
0891	GAT	TGC	CTG	TAT	CTC	AAT	GTG	TTT	CTG	CGG	AGC	ATG	GCC	AGC	TGC	CTG	GTG	TTC	CAC	AAT	GCC	GCG	GAG	GCG	AGC	Thr	Asp	Cys	Leu	Tyr	Leu	Asn	Val	Phe	Val	Phe	His	Met	Ala	Pro	Asn	Ser	Val	Leu	Leu	Ala	Gly	Lys	1620			
0931	GUC	AAT	GCG	CGG	CGG	GCA	GTC	GAT	GCG	TTC	CTG	GCT	GCT	GTC	GGC	ACC	ATC	ATT	GTG	GTG	AGC	AGC	ASC	TAT	CGC	Ala	Gly	Ser	Asp	Arg	Pro	Gln	Ala	Val	Asn	Leu	Val	Val	Val	Thr	Ala	Ser	Tyr	Arg	Thr	Ala	1620					
7021	TTC	GCG	CTC	CTG	AAT	TTC	GCG	TCC	GCG	AGC	AGC	AGC	TGG	GCG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	Thr	Phe	Gly	Phe	Leu	Ser	Gly	Ser	Glu	Leu	Ser	Gly	Asn	Tyr	Gly	Leu	Leu	Thr	Tyr	Val	Gln	1620					
7111	ATC	CAG	GGT	TTC	GCG	GAC	CTC	GCG	GGC	GTC	ACC	CTG	GCA	GAC	GAC	GCG	ATC	GAT	ATA	GCG	GCG	ATC	CAC	CTT	GTA	ACC	Thr	Leu	Gln	Ala	Phe	Gly	Ala	Pro	Arg	Val	Thr	Leu	Ala	Ala	Ala	Ser	His	Leu	Val	1620						
7201	AGC	GCC	AGC	ATC	Ser	AGA	CTC	TTC	GCG	AGA	GCT	GTC	ATG	GCT	GGC	CTC	CTC	CTC	GGC	CTC	ATC	AGA	GAG	AGC	GCT	Thr	Arg	Ala	Asn	Ser	Arg	Pro	Ala	Val	Leu	Met	Gly	Ser	Ala	Val	Ala	Arg	Glu	Ala	1620							
7291	GCG	CAG	GCA	GCT	GCT	TTC	GCA	AGC	GAG	GTT	GCT	GTC	CCC	AGC	TGC	CTC	GCA	AAA	ATG	GTG	TGC	CTC	CGC	CAG	GCG	CCC	Arg	Gln	Gln	Ala	Ala	Leu	Ala	Gly	Val	Gly	Cys	Pro	Ser	Ser	Val	Gln	Gly	Met	Val	Ser	Cys	Leu	Arg	Gln	Ala	1620
7381	AGA	ATC	CTC	AAT	GAT	CCC	CAG	ACC	AAI	CTT	CTG	GCT	AGT	GCT	CTC	CAC	TAC	TGG	GCT	CGG	GTG	CTC	GAG	CAG	CTC	GTA	CCC	Arg	Ala	Leu	Asp	Ala	Gln	Thr	Ala	Val	Ser	Gly	Pro	His	Tyr	Gly	Pro	Val	Val	Ala	Asp	Gly	Thr	Leu	1620	
7471	GAC	AGC	CCA	GCC	GCA	GTG	CTG	CAG	AGC	GCC	CCA	CGG	CTC	AAG	GTG	GAT	CTC	ATC	GGG	AGT	TCT	CAG	GAC	GAC	CTC	ATC	AGA	Glu	Thr	Pro	Ala	Arg	Val	Leu	Gln	Arg	Ala	Pro	Arg	Val	Lys	Val	Asp	Leu	Leu	Gly	Ser	Gln	Asp	Gly	Leu	1620
7561	GCC	AGC	GCT	GAG	CAG	TAT	GAG	GAA	CAA	TAT	GAG	GAA	AGT	CAA	GCG	AGC	AGT	ACC	AAA	ACA	GCG	ATC	AGC	CTC	ATC	AGA	AGA	Thr	Pro	Ala	Arg	Val	Leu	Gln	Arg	Ala	Pro	Arg	Val	Lys	Val	Asp	Leu	Leu	Gly	Ser	Gln	Asp	Gly	Leu	1620	
7651	GAC	GCA	GCA	GAC	GCC	GCT	GTC	GAG	GCT	GGG	GCG	AGC	TGG	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	TAC	Thr	Glu	Ala	Ala	Ala	Val	Gln	Ala	Ala	Ala	Thr	Tyr	Tyr	Ser	Ser	Leu	Glu	His	Asp	Ser	Ser	Phe	Ser	Arg	Ala	1620
7741	CTC	GAG	CAG	GCG	AGC	GGA	GAC	TTC	ATC	ATC	TTC	ATC	TTC	CCC	GTG	ATC	AGC	ATC	AGC	ATC	AGC																															

Fig. 1 Sequence of thyroglobulin cDNA and the deduced primary structure of the protein. Nucleotides are numbered (on the left of the figure) from the ATG initiation codon. The deduced amino acids are numbered from the amino-terminal asparagine of the mature protein. The polyadenylation signal⁸ is underlined. The putative *N*-glycosylation sites (Asn-X-Ser/Thr) and the tyrosine residues involved in hormonogenesis are underlined and overlined, respectively.

Methods. Six recombinant plasmids⁴ and one genomic DNA fragment (G. de Martynoff *et al.*, in preparation) were used as starting material. Subcloning in phage M13 vectors²⁴ and generation of overlapping fragments with DNase were performed as described in ref. 25. The dideoxy chain termination method²⁶ and buffer gradient gels²⁷ were used for DNA sequence analysis. Confidence was obtained by reading each nucleotide on both DNA strands. The overlapping recombinant plasmids⁴ differ at two positions: nucleotide 4,585 is A in pBTg 1.9 and G in pBTg 4.7 (changing amino-acid 1,510 from Ser to Gly), and nucleotide 7,141 is G in pBTg 4.7 and A in pBTg 2.5 (changing amino-acid 2,362 from Val to Met). These differences may reflect sequence polymorphism. However, an error in the reverse transcriptase reaction cannot be excluded.

However, an error in the reverse transcriptase reaction cannot be excluded.

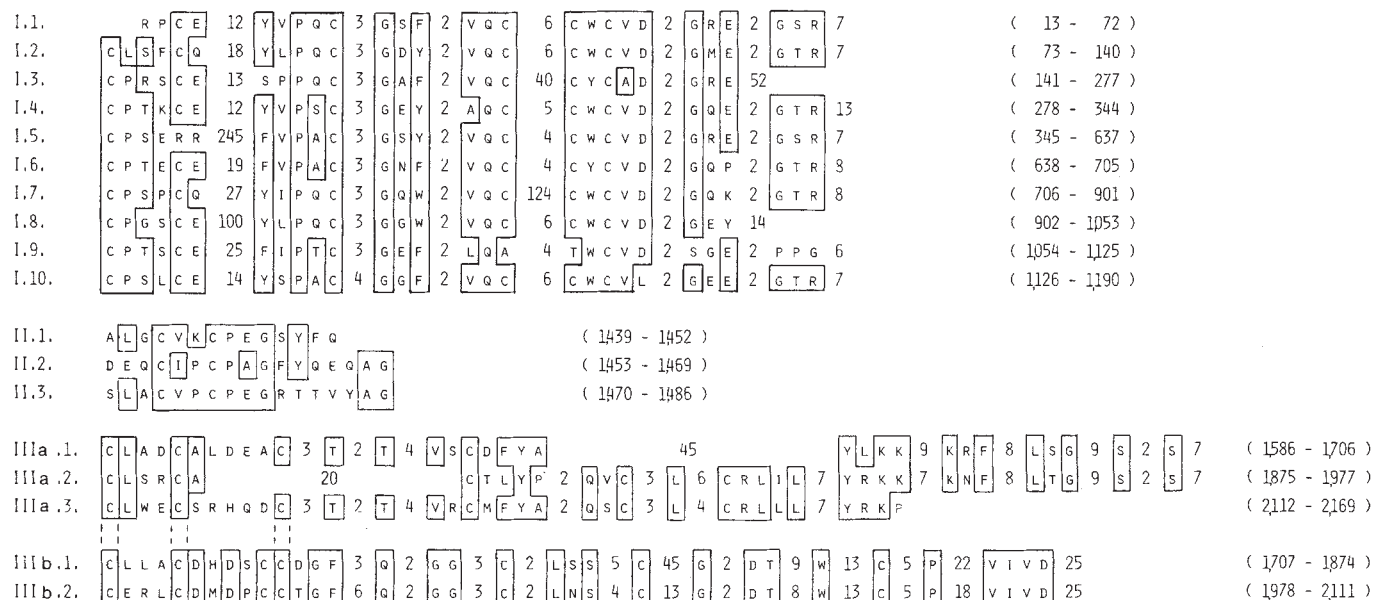


Fig. 2 Internal homology in thyroglobulin primary structure. The three types of homologous domains (given in the one-letter code) are aligned. Insertions or deletions have been introduced to maximize homology. Conserved amino acids are boxed according to the following rule: for type I repeats, the amino acids should be present at least five times; for type II and III repeats, at least twice. Residues showing no conservation between the domains are represented by numbers. Homologies in Cys positions between domains IIIa and IIIb are indicated by broken lines. Localization of the domains is shown on the right; the numbering is as in Fig. 1.

DNA fragment containing the 5'-terminal portion of the message (G. de Martynoff *et al.*, in preparation). Partial nucleotide sequences have been reported for three species⁵⁻⁷ but they do not exceed 3,000 bases long. The general organization of the mRNA sequence is as follows: a 41-nucleotide 5'-untranslated segment precedes an open reading frame of 8,307 bases. To our knowledge, this represents the longest coding sequence determined to date. The 3'-untranslated segment is 83 bases long and contains the canonical AATAAA polyadenylation signal⁸.

As judged from the nucleotide sequence, a 19-amino-acid signal peptide is followed by 2,750 residues, constituting a polypeptide of expected M_r 302,253; this value is in perfect agreement with the M_r obtained for the translation product of thyroglobulin mRNA *in vitro*⁹ or in *Xenopus* oocyte¹⁰ and with the M_r of 330,000 for the glycosylated protomer¹¹. The amino-acid composition of the protein agrees with published results (not shown). In particular, despite their role in hormonogenesis, the number of tyrosine residues (72, that is, 2.5%) does not deviate significantly from the average observed in a protein data bank (3.3%; ref. 12). The sequence contains 15 potential acceptor sites for *N*-glycosylation (Fig. 4), which represents a slight excess compared with the 20 chains characterized in the mature dimeric protein¹³.

When analysed for the presence of internal homology, the protein sequence shows a striking repetitive pattern involving three types of cysteine-containing motifs (Fig. 2). The type I motif is repeated 10 times between positions 13 and 1,190; it extends over ~60 amino acids and contains 24 highly conserved residues, including six cysteines. Sequences of variable length, unrelated to the motif, interrupt some of the repeats at fixed positions (Fig. 2). Preliminary data indicate that the positions of these insertions correlate with the borders of exons (J. Parma, in preparation). The type II motif is much shorter (17 amino acids) and is only repeated three times between positions 1,439 and 1,486. The type III motif may be subdivided into two subtypes (IIIa, IIIb; Fig. 2), which display a similar pattern of cysteine residues. Both subtypes are repeated twice in tandem, between positions 1,586 and 2,111. A third copy of motif IIIa is found further downstream (Fig. 4). Over 75% of the thyroglobulin sequence is thus involved in some kind of repetitive structure, clearly indicating that this exceptionally large protein evolved from the serial duplication of a limited number of building blocks. Interestingly, the last 600 residues at the carboxyl end show no internal homology, are poor in cysteine and contain a cluster of tyrosines of which three are hormonogenic (see below and Fig. 4).

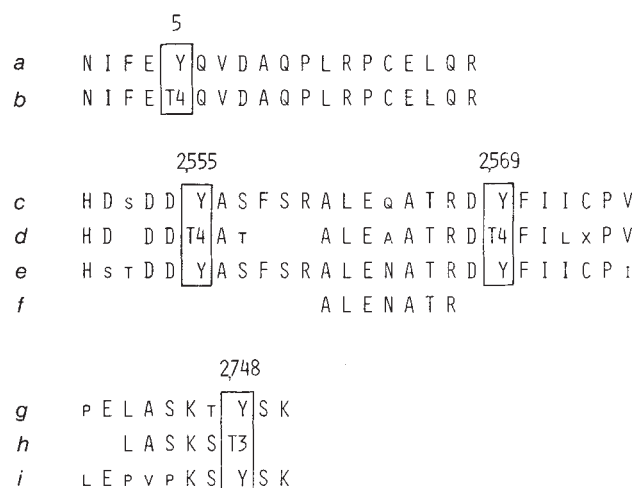


Fig. 3 Hormonogenic domains in thyroglobulin. The hormone-containing domains identified to date in various species are represented by the one-letter code and are aligned for maximum homology. The capital letters correspond to amino acids occurring at least twice in different species. The four hormonogenic tyrosines are boxed and their positions in the protein are indicated. *a*, *c*, *g*, the peptide sequences derived from the present study; *b*, the four identical hormone-containing peptides determined from bovine, human, ovine and porcine thyroglobulin^{17,28,29}; *d*, the sequence of three peptides obtained from porcine thyroglobulin, named b₁Ad₃, T6 and b₁B in ref. 18; *e*, *i*, protein sequence derived from a partial rat cDNA sequence⁷; *f*, a tryptic peptide, isolated from human thyroglobulin, containing a glycosidic chain attached to the asparagine residue³⁰; *h*, sequence of a hormone-containing peptide isolated from porcine thyroglobulin¹⁹.

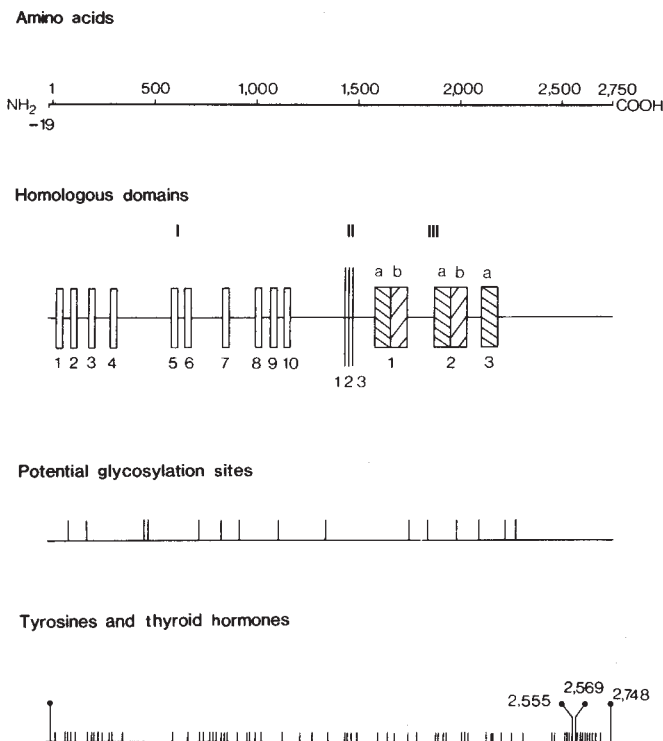


Fig. 4 Schematic representation of the thyroglobulin precursor. For clarity, a space has been introduced between the homologous domains even when they are contiguous in the protein (see Fig. 2). The tyrosine and thyroid hormone residues are represented by short and longer bars, respectively.

The sequence organization of thyroglobulin (Fig. 4) is reminiscent of the structure of some other large polypeptides, including the epidermal growth factor (EGF) precursor¹⁴, EGF receptor¹⁵ and low-density lipoprotein receptor¹⁶. Although there exists no direct sequence homology between thyroglobulin and these proteins, they probably share an analogous evolutionary history and may have similar three-dimensional structures.

Four tryptic peptides containing thyroid hormones have been isolated from thyroglobulin of various species and sequenced¹⁷⁻¹⁹. A systematic search for homologous peptides in our sequence allowed the identification and precise localization of all four hormonogenic tyrosines (Figs 3, 4). Two of these map at subterminal positions (residues 5 and 2,748); they correspond to sites involved preferentially in the synthesis of T₄ and T₃, respectively^{17,19}. The other two sites are closely linked around position 2,560. The amino-terminal domain is adjacent to the first, type I, repeat and might even share sequence characteristics with it⁵. In contrast, the three other hormonogenic tyrosines map in the non-repeated carboxy-terminal segment (Fig. 4). Comparison of the sequences surrounding the hormonogenic residues reveals no clear homology except for the presence of at least one charged amino acid (Fig. 3) and some clustering of tyrosines (Fig. 4)—the significance of this observation awaits the identification of the donor tyrosine involved in the coupling reaction.

Compilation of the sequences of the hormonogenic segments presently available shows a remarkable conservation of the amino-terminal domain (Fig. 3). It has been reported that as much as 50% of the thyroxine in mature thyroglobulin may originate from this domain²⁰. Although less information is available for the other hormonogenic domains, they do not display the same degree of sequence conservation (Fig. 3). Given these considerations, it is possible that the amino-terminal region of thyroglobulin has an important role in hormonogenesis. It is noteworthy, in this context, that a 10K amino-terminal peptide is cleaved from the protein during the hormonogenic process²¹.

The chemical or enzymatic reactions responsible for this phenomenon are unknown; it must be quite specific, however, as the carboxyl terminus of the 10K peptide could be unambiguously determined and was readily localized (Gln 80) in a partial protein sequence²².

Knowledge of the primary structure of thyroglobulin should provide the basis for studying the structure-function relationships of the hormonogenic process *in vitro*, in reconstructed synthetic systems, and should help to elucidate the molecular basis for genetic defects of thyroglobulin production²³.

We thank Dr J. E. Dumont for his continuous support and for critical reading of the manuscript, and Mrs D. Leemans for preparation of the manuscript. This study was supported by grants from the Ministère de la Politique Scientifique, NIH (21732), FRSM, IRSIA and ARBD.

Received 7 May; accepted 18 June 1985.

1. Van Herle, A. J., Vassart, G. & Dumont, J. E. *New Engl. J. Med.* **301**, 239-249, 307-314 (1979).
2. Marriq, C., Arnaud, C., Rolland, M. & Lissitzky, S. *Eur. J. Biochem.* **111**, 33-47 (1980).
3. Rolland, M., Montfort, M. F. & Lissitzky, S. *Biochim. biophys. Acta* **303**, 338-347 (1973).
4. Christophe, D., Mercken, L., Brocas, H., Pohl, V. & Vassart, G. *Eur. J. Biochem.* **122**, 461-469 (1982).
5. Mercken, L., Simons, M. J., de Martynoff, G., Swillens, S. & Vassart, G. *Eur. J. Biochem.* **147**, 59-64 (1985).
6. Malthiery, Y. & Lissitzky, S. *Eur. J. Biochem.* **147**, 53-58 (1985).
7. Di Lauro, R. *et al. Eur. J. Biochem.* **148**, 7-11 (1985).
8. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211-214 (1976).
9. Chebath, J., Chabaud, D., Becarevic, A., Cartouzou, G. & Lissitzky, S. *Eur. J. Biochem.* **77**, 243-252 (1977).
10. Vassart, G., Refetoff, S., Brocas, H., Dinsart, C. & Dumont, J. E. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3839-3843 (1975).
11. Lissitzky, S., Mauchamp, J., Reynaud, J. & Rolland, M. *FEBS Lett.* **60**, 359-363 (1975).
12. Barker, W. C. *et al. Protein Sequence Data Base: Atlas of Protein Sequence and Structure Version 7* (Natl. Biomed. Res. Fdn, Washington, D. C., 1983).
13. Arima, T., Spiro, M. J. & Spiro, R. J. *biol. Chem.* **247**, 1825-1835 (1972).
14. Gray, A., Dull, T. J. & Ullrich, A. *Nature* **303**, 722-725 (1983).
15. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).
16. Yamamoto, T. *et al. Cell* **39**, 27-38 (1984).
17. Rawitch, A. B., Chernoff, S. B., Litwer, M. R., Rouse, J. B. & Hamilton, J. W. *J. biol. Chem.* **258**, 2079-2082 (1983).
18. Marriq, C., Rolland, M. & Lissitzky, S. *EMBO J.* **1**, 397-401 (1982).
19. Marriq, C., Rolland, M. & Lissitzky, S. *Biochem. biophys. Res. Commun.* **112**, 206-213 (1983).
20. Marriq, C., Lejeune, P. J., Malthiery, Y., Rolland, M. & Lissitzky, S. *FEBS Lett.* **175**, 140 (1984).
21. Chernoff, S. & Rawitch, A. J. *biol. Chem.* **256**, 9425-9430 (1981).
22. Rawitch, A. B., Mercken, L., Hamilton, J. W. & Vassart, G. *Biochem. biophys. Res. Commun.* **119**, 335-342 (1984).
23. Ricketts, M. H. *et al. EMBO J.* **4**, 731-737 (1985).
24. Vieira, J. & Messing, J. *Gene* **19**, 259-268 (1982).
25. Hong, G. F. *J. molec. Biol.* **158**, 539-549 (1982).
26. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161-178 (1980).
27. Biggin, M. D., Gibson, T. J. & Hong, G. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3963-3965 (1983).
28. Lejeune, P. J., Marriq, C., Rolland, M. & Lissitzky, S. *Biochem. biophys. Res. Commun.* **114**, 73-80 (1983).
29. Rawitch, A. B. *et al. Biochem. biophys. Res. Commun.* **118**, 423-429 (1984).
30. Rawitch, A. B., Liao, T. H. & Pierce, J. G. *Biochim. biophys. Acta* **160**, 360-367 (1968).

Ethylation interference and X-ray crystallography identify similar interactions between 434 repressor and operator

Frederic D. Bushman, John E. Anderson, Stephen C. Harrison & Mark Ptashne

Fairchild Biochemistry Building, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

In the crystal structure of a repressor-operator complex (the 434 repressor DNA-binding domain and its 14-base pair (bp) operator), Anderson *et al.*¹ elsewhere in this issue identify six positions of likely contact between repressor protein and phosphates of the DNA backbone. At each of these positions, electron densities of protein and DNA merge. Experiments presented here indicate that intact 434 repressor approaches these phosphates very closely when it is bound to DNA in solution. Specifically, when any one of these phosphates is ethylated², repressor cannot bind to the modified operator. We also identify another position where ethylation has