## A single amino acid change in the acetylcholinesterase-like domain of thyroglobulin causes congenital goiter with hypothyroidism in the cog/cog mouse: A model of human endoplasmic reticulum storage diseases

PAUL S. KIM\*<sup>†</sup>, Shaikh Abu Hossain<sup>\*</sup>, Young-Nam Park<sup>‡</sup>, Ike Lee<sup>§</sup>, Sung-Eun Yoo<sup>\*</sup>, and Peter Arvan<sup>‡</sup>

\*Division of Endocrinology and Metabolism, University of Cincinnati School of Medicine, Cincinnati, OH 45267; <sup>‡</sup>Division of Endocrinology and Department of Developmental and Molecular Biology, Albert Einstein College of Medicine, Bronx, NY 10461; and <sup>§</sup>Division of Cardiology, Beth Israel Hospital, Boston, MA 02215

Edited by Donald F. Steiner, University of Chicago, Chicago, IL, and approved July 2, 1998 (received for review May 6, 1998)

ABSTRACT Newly synthesized thyroglobulin (Tg), the major secretory glycoprotein of the thyroid gland, folds and homodimerizes in the endoplasmic reticulum (ER) before its export to the site of iodination, where it serves as the precursor for thyroid hormone synthesis. In families with defective Tg export, affected individuals suffer from a thyroidal ER storage disease characterized by a distended thyrocyte ER containing misfolded Tg, along with induced ER molecular chaperones. Inherited as an autosomal recessive trait, deficient Tg causes congenital hypothyroidism in newborns that, if untreated, results in goiter along with serious cognitive and growth defects. Recently, a similar phenotype has been observed in inbred *cog/cog* mice, although the precise molecular defect has remained undefined. Here, we have isolated and cloned a full-length 8.5-kb Tg cDNA from cog/cog mice and unaffected isogenic AKR/J mice. Comparison of the complete sequences reveals that cog/cog mice express a Leu-2263  $\rightarrow$  Pro missense mutation in the acetylcholinesterase-homology domain of Tg. Heterologous expression studies in COS cells indicate that cog Tg exhibits a severe defect in exit from the ER. Site-directed mutagenesis of cog Tg to convert the single amino acid back to Leu-2263 restores normal Tg secretion. We conclude that the cog mutation in Tg is responsible for this ER storage disease that causes thyroid dyshormonogenesis.

For trafficking through the eukaryotic secretory pathway, exportable proteins must achieve a certain level of conformational maturity to be competent for intracellular transport as well as for their ultimate biological function. An increasing number of genetically transmitted metabolic diseases involving intracellular retention of exportable proteins have been recognized, caused by mutations that lead to tertiary and quaternary structural defects (1). In a family of disorders of protein trafficking known as endoplasmic reticulum storage diseases (ERSDs; ref. 2), mutations as small as a single amino acid change can lead to failure of protein export from the ER (an essential step in overall protein export), such as in cystic fibrosis [mutant CFTR (3)], juvenile pulmonary emphysema [mutant  $\alpha$ -L-antitrypsin (4)], osteogenesis imperfecta [mutant type I procollagen (5)], juvenile diabetes insipidus [mutant vasopressin precursor (6)], and hypercholesterolemia [mutant low density lipoprotein receptor (7)]. Because export of the mutant protein from the ER is prevented, it never reaches the destination at which its physiologic function normally takes place.

Mutations in thyroglobulin (Tg), the major thyroid secretory glycoprotein, can also cause inherited disease (2, 8, 9). Normally, Tg is exported through the secretory pathway to the lumen of thyroid follicles to undergo iodination, which leads to thyroid hormone synthesis. Congenital hypothyroidism due to deficient Tg is inherited as an autosomal recessive trait in  $\approx$ 1:40,000 human newborns (10). In several human kindreds that have been studied with this disorder, defective trafficking of Tg from the ER to Golgi has been suggested to be responsible for the congenital thyroid hormone deficiency (11). In at least one of these kindreds, the underlying molecular defect has been proved to be a mutation resulting in an altered Tg coding sequence (12).

A similar form of severe congenital hypothyroidism with colloid-deficient goiter along with abnormal growth and central nervous system development is observed in cog/cog mice (13). Indeed, electron microscopic studies of the thyroid glands of cog/cog mice indicate an abnormally distended ER (14)-a phenotype comparable to that observed from thyroid biopsies of children suffering from congenital goiter with defective Tg (11, 15-17). In the cog/cog mouse, Tg mRNA is abundant and normal in size, but purified Tg protein exhibits abnormal biophysical properties including enhanced susceptibility to proteolysis (18-20). We have recently established that abnormal folding, dimerization, and export of Tg, in association with markedly elevated levels of five ER molecular chaperones, characterizes the thyroid defect, indicating an ERSD (21). Moreover, in thyroid tissue from homozygous *cog/cog* mice, Tg exhibits temperature-sensitive export from the ER, suggesting that the cog defect might be caused by a point mutation in the Tg coding sequence (21).

The *cog* trait originally appeared as a spontaneous autosomal recessive phenotype in the inbred AKR/J strain of mice (22). The *cog* gene has been mapped near the Tg gene locus at the central region of mouse chromosome 15 (23), but the precise genetic defect has remained elusive. Here, we have prepared thyroid cDNA libraries from *cog/cog* and isogenic AKR/J mice, which, by cDNA sequencing, has allowed us to deduce the complete Tg primary structure in both cases. These data demonstrate that the hormonogenic domains of *cog* Tg are identical to those of the normal Tg. However, we have found a single nucleotide substitution in *cog* Tg, changing Leu-2263 to Pro-2263 in a region strictly conserved in the Tgs from all known species, which is evolutionarily derived from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

<sup>© 1998</sup> by The National Academy of Sciences 0027-8424/98/959909-5\$2.00/0 PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviations: Tg, thyroglobulin; ERSD, endoplasmic reticulum storage disease; AChE, acetylcholinesterase.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF076186).

<sup>&</sup>lt;sup>†</sup>To whom reprint requests should be addressed at: 231 Bethesda Ave, Medical Science Building 5564, Cincinnati, OH 45267. e-mail: Paul.Kim@uc.edu.

	1 ATG ACG GCC CTG GTC TTG TGG GTC TCT ACT TTG TGG AGC TCA GTC TGC CTG GTA GCA GCC	
61	20 M T A L V L W V S T L L S S V C L V A A	
181	N I F E Y Q V D A Q P L R P C E L Q R E K A P L K Q A E Y V P Q C S B D	GSFQ
41	T V Q C Q N D G Q S C W C V D S D G K K V P G S R Q L G R P T I C L S P	C Q L H
81	UNICANA COSIANC ENSIGNOS ANCINAS ANTINAS ASCIACCIGATIGOTICITE CITECITE CONSIGNATIVA GOSIANCITATIGOS COSIANTIVAS AS ANTINAS ASCINTATIGOS COSIANTIVAS AS ANTINAS AS ANTINAS ANTINAS AS ANTINAS ANTINAS ANTINAS AS ANTINAS ANTINAS AS ANTINAS ANTINAS ANTINAS AS ANTINAS ANTINAS ANTINAS ANTINAS AS ANTINAS ANTINAS AS ANTINAS ANTINAS AS ANTINAS ANTINA	TIG CAA TGC TGG
421	ίατ στη όλο κάν αφο όσο λης όλα ότο την όσο κός όσο όλο όλο όρο όσο του τότο τότο τότο τότο όλο λόο και τότο τότο τότο τότο τότο τότο τότο τότ	GA GAC AGG TCA
541	ών τος	SAT GCA TIT GTG
161 661	PPQCTADGEFMPVQCKFVNTDTDMMIFDLIHNYNRFP VCTCTCAGCGGGCAGGGGCAGTTTCCCTGAAGTGGTGGGGGGGG	DAFV SAAATT TAT GAC
201	TFSSFRGRFPEVSGYCYCADSCALLABTGLEELL	EIYD
241	T I F A G L D Q A S T F T Q S T M Y R I L O R F L A I O L V I S G R A CALTA STARTS AND A CONSTRAINT A CONST	R C P T
901 281	AG TOT GAA GTA GAA CAG TITI GOT GCC ACC AGA TITI GOT CAT TOT TAC APIC COG AGA TOGC CAC AGA GAT GOT CAC TAC CAA ACA GTA CAG TGC CAG AGA GOG ATG ' K C E V E O F A T F F G H S V T P F C H S V T O F C H S V O T O C H V O C O C F C A A CAG A GOG ATG '	NGC TGG TGT GTG
1021	μπ σες κλα το το το το το το το τός τος τός τος τός τος τος τος τος τος τος τος τος τος το	TAC TIT GAA ACC
1141	LAQ GREVPETTRUQUES CALESENT CALESENT CALESENTE OF CONTENTS AND A CALESENT TO C	Y F E T TT GAC TCT GGA
361	PDYFSPQDLLSSEDRLAPVSGVRSDTSCPPRZXES	V D S G
401	L L R S I A V B H Y Q R L S E S R S L L R E A Y F P S R E L A G L	A L Q F
1381 441	CC ACC AAC CCC AAG CGA CTC CAG CAG AAC CTG TTT GGA GGG ACA TTT TTG GGG AAT GCT GCC CAG TTT AAC TTG TCA GGA GCC CTT GGT ACA AGA AGC ACA TTT AAC ' $^{\mathrm{TT}}$ TO N P K R L O O N L F G G T F L A N A A O P N L S G A L G T R S T F N	TC AGT CAG TTT
1501	το όλο όλο τμα όδι όμο τος τος νός της ότα και κάν από κάι κάι τος τάν ότα ότα ότα της ότα της όλο τος και τος τός όλα της ότο ότα τος	TG TCT GAG AAG
1621	CACT ACT ANG AAA CGA GTT GTG GGT ACT TIG GGT TC AAA GTC AAC CTA CAA GAA GAA CAA GAA GCC TTA AAA TTT CTC GTT TCT CTC CTG GAA CTT CCA GAA	V S E K FIC CTT GTC TTC
521 1741	TVAMNKRVVGNPGFORTOTTOTTOTTOTTOTAL AND GATTER GATTER GATTER GALANT GALANT GATTEN GATTEN GATTEN A GATTER GATTEN GALANT GALANT GATTEN GALANT GATTEN GATTAN GATTEN GATT	F L V F
561	L Q R A V S V S V F E D I A R D L G D V M E M V F S A Q A C K Q M P G K F F	V P S C
601	The G S Y E D I Q C Y A G E C W C V D S R G K E L D G S R V R G G R PAGE C W C V D S R G K E L D G S R V R G G R P	R C P T
1981	LG TOT GAA AAG CAG AGA GAT CAG ATG CAA AGC CTC GGA AGC GCT CAG CCC GCC GGC TYC AGC TITT TITT GTT CCT AGT AGT AGT AGA GAG GAA TAT TITT CTC CCT GTC (	AG TGC TTC AAC
2101	ών σμα τών τῶν τῶν τῶν τῶν τῶν τῶν τῶν τῶν τῶν τῶ	ET TTC CTT COG
2221	SECYCVDTEGQV1PGTQSTCVQCDTEGQV1PGTQSTVGEAKCCCPSVCQLQAEQ TN GTAGGGTTCTCTGAACTCCAGCTGGTGCCCTCCATCCCAATGTTACATCCTCAGTGCAGTGCCAGTGGCGCTGGAGGCATTGGCGGTGGAGGGATGTGGCAGTGGCG	A F L G
721	V V G V L L S N S S M V P S I S N V Y I P Q C S A S G Q W R H V Q C D Q C D G I TO THE TAK AND AN AN AND AN AN AND AN AN AND AN	PHEQ
761	V P E W Y E R W K T Q N G D G Q E L T P A L L M K I V S Y R E V A S R	N F S L
2461 801	TT CTT CAA AGT CTT TAT GAT GOT GOC CAG CAA AGA ATC TTT CCA GTG CTA GOA CAG TAC CCA TCT CTT CAA GAT GTT CCA CAG GTA GTG TTG GAG GGA GCC ACA ACT ( $F$ L O S L Y D A G O O R I F P V L A O Y P S L O D V P O V V L E G A T	CG CCT GGG GAG
2581	ΔC ΑΤΤ ΤΤΟ CTIG GAT CCG ΤΑC ΑΤC ΤΤΤΤ TOG CAG ATC CTL ΑΑΤ GGT CAG COC AGC CAÂ ΤΑC CCA GGT COC ΤΑ̈́C ΤCA GAC TTT AĂC ATG COT TTG GAA CAC TTC ΑΑC TTC ΑΑC TC ( Ν Τ Ρ Τ Γ Γ Ρ Υ Τ Γ Ρ Υ Γ Γ Γ Ρ Υ Γ Γ Γ Γ Γ Γ Γ Γ Γ Γ Γ Γ	GG AGC TGT TGG
2701	סָד סָדָר ﻣَאָר ﻣَאָר ﻣָבָר ﻣָבָר ﻣָבָר ﻣָשָׁ ﻣָדָר ﻣָבָד אָרֵה רָהָר אָבָר רָאָר כָשָׁ ﻣָבָר אָב רָאָר מָגָר באָר מָגר מאָר מאָר מאָר מאָר מאָר מאָר מאָר מאָ	TT ANG GAA ACA
881 2821	C V D E A G Q K L D G T Q T K P G E I P A C P G P C E E V K L R V L K P NA GAJATI GTC TCA GCC TCC AN GCT TCT TCA TTE GCA GTG GGA GTG TTC TTA GTG GCC ANG GGA GCA GTG GGA GTT GCC TCA '	I K E T
921	EEIVSASNASSPPLGGESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	F P S R
961	D A F S E K F L R G G E Y A I R L A A Q S T L T F Y Q S L R A S L G K S	D G A A
3061 1001	GT CTT CTG TGG TCA GGG CCC TAC ATG CCA CAA TGC AAT ATG ATT GGA GGC TGG GAA CCT GTC CAG TGC TAG GGA CAG TGT TGG TGT GTG GAT GGA AGG ' S L L W S G P Y M P O C N M T G G W B P V O C H A G T G O C W C V D G P '	GA GAG TIC ATT
3181	CA GOC TOG CHE AND TOA COC TOA CAG AND COC CAG TOT COC AGA AAC TOT GAG CHA TOC AGA GOA ADT GOG THE ANT TOT GOA TOG ADA CAG GOT GOA COC CAA	GA AAT CCA GGC
3301	ός σα από τος της της από τος της τος της τος τος από τας της από της από της από τος τος τος τος τος τος τος τ	CA GTG AAC ACA
1081 3421	PGDLFIPVCLQTAGTCTCGGCTCTGTGTGTGTGTGTGTGGTGTGTGGGGGTGTGTGGGGGTGT	PVN T
1121	N G S A Q C P G L C D V L K S R A L S R K V G L G Y S P V C E A L D G A	FSPV
1161	AN TOFUSE THE GEC CALING THE THE THE THE OT GOLA AND CONSIGNAL GALING AND AND THE THE OTHER OF A CHARACT CONCAC THE THE AND A CONCAC THE AND A	P L P F
3661 1201	"CT GGG TCA GAT GTG GCT GAT GGA GTG ATC TTC TSC GAA ACA GCC TCT AGC TCA GGA GTG AGG ACT GTT CAG CAG TGC CAG CTG TGC CSC CAG GGC CTC CGG AGT I S G S D V A D G V I P C E T A S S S G V T T V O C C O L L C P O G L P S	A F S P
3781	ώς σόν σύς νών μαι νός σύς ανό νόν σύς τος σύς από τος σόν σόν σόν σόν σός σύν κός του σύν νός σών σύς σύς σύς σύς της (	AG CTC CTG CTC
3901	CA CÉA GĂA AĂG ATG TỐC AĞT GTT GĂC TĂC TĆC GỐC TTG CĂG GĆC TTC CĂO GTC TTC ATA CTG GĂT GĂO CTG ATA GOC CTG TG CTT TC TCC CAG ATC CĂG GTA AĂG J	CA TTT GGG ACC
1281	PPGKMCSVDYSGLLQAFQVFLDAGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TFGT
1321	L = V = S = T + C = D + S = I = Q + G = C + L = A = R + L = Q + N + T = R + L = Q + N + L = Q + S = I = S + Q + G = L = C + Z + L = Q + S = C + Z + S + Z + Z + Z + Z + Z + Z + Z + Z	V G S L
4141	CA GAPITIG TAC AGC ATT GAA AGG GGG GTA AGG GGC CAG GAT CTC CTT GGG GGC TTT GGC GGT CTG GATC GGC AGG GTTC CAG CTT CAT CTG GAC TCC AAG AGA P D L Y S I E R A V T G Q D L L G R F A D L I O S G R F O L H L D S K T	F S A D
4261	CC ACC TTG TAT TTT CTC ALT GOT GAC ACC TTT GTC ACC TCT COC ACC ACT CAG CTT GOA TGT ATG GAA GGG TTC TAC AGA GTC CCA ACC ACA CAG GAT GCT CTG G	SC TGT GTG AAA
4381	CO COT GAA GGA AGC TTO TTO CAG GAT GGG CGA TGC ACT COC TOT COT GOT GGC AGC TAC CAA GAA CAG GCA GGC AGC TCA GCC TGC ATC CCA TGC CCA AGG GGC AGA	CA ACC ATC ACC
1441 4501	CPEGSFXQAACCEPEGSVQDGRCTPCPAGAACACGTVQGCTVAGACACACCTVAGACAACGAACGAACGAACGACGACGACGACGACGACGAC	TTIT
1481	T G A F S K T H C V T D C Q K N B A G L Q C D Q N G Q Y Q A S Q K N R D T	S G E V
1521	TC TO GING GALAGE GANG GANG GANG GANG CHALGAR THALGAR GANG GOT GANG GOT CHALGARA TOT CAGE TOT CHEGATER GANG AND TOT GANG AND GOT CCALGARA TOT $F$ C V D S E G R K L Q W L Q T E À G L S E S Q C L M I R K F D K À P E S	AGGTT ATC TTT K V I F
4741 1561	AT GOT ANT TOT COT GTG ATA GTC AAG TCC AOT GTC COC AGT GOT GAC TCC CGG CTG CAG CAC TCC TTG ACA GAT TGT GCA AAT GAT GAG GCC TGC AGC TTC CTT ACA ( D A N S P V T V K S S V P S A D S P I V O C I T D C A N P F A C S P I V	TG TCC ACA ATG
4861	μα τύν αύα από τόν τόν ανα την τών τον του και το του του του του του του του του του	GG AGC CTC AGG
4981	ESEVSCEAGTIG AAA ATG AGA AAT AG CAAT ATT TRUD N FACVT SDQEQ DA MGSLKA ATG AGA AGG AGT GAA ATG AGT CAGA AGG AGC TTG GAAGA AGG AGG CATG GAA ATG AGT AGG AGG GAGT GGA AGT AGG AGG	G S L R CAA AAT GTG CTC
1641 5101	C Q V K V R N S G K D S L A V Y V K K G Y E S T A A G Q K S F E P T G P T GG TH TAN NG C TOT GTG GTG TTO TTOT GTG GTG AG GGC AN GTG A AG AN AG AG AG THE GTG GTG GTG GTG GA AN GTG TTOT GTG GTG GTG GTG GTG GTG GA AG	Q N V L
1681	G L Y S P V V F S A S G A N L T D T H T Y C L L A C D N D S C C D G F	IITQ
1721	TI NAM GOAG GEE ACE AFE THE THE THE AGE THE CHE GALE AT CHE CHE ARE ARE ARE GALE GALE ACE AFE GEE AGE CAST GET ACE THE AFE AFE AFE AFE AFE AFE AFE AFE AFE AF	A G V T
5341 1761	AT GAC CAG GGG AGC CGC CAG ATG ACT TG AGT CTG GGA GGC CCA AGAG TTC TG CAG GGT CTG GCA CTC TTG GAA GGA ACT CAA GAC AGC TTTT ACC AGC TTC CAG CGG Y D O G S R O M T L S L G O C F F L O G L A L L F G T O D S P T S R O O	TC TAC CTC TOG
5461	ο ότο τός της της της ότο ότα το τός το το το της του το της του ότα το	AC ATT ACC CAG
5581	TE ATE OTC AAC AGE CAT TO ATE CE AS CAS CAS TAC TOSE TO TTC AGE CAC CHE THIT TO AGE AGE GAS GAS AC THE THE TO TOSE TOST OCC CAS	D I T Q BAG CCC ATT TTC
1841 5701	VIVNTSHSLPSLPSQQQYWLFTTHLFSQQQXWLFTALFSAEQANLWCLSRCAQ GTCAACTIGGCGGCATAACAAAAGTTCATCTITTGTATTTACACTGCTTGCGAGTGCGGGTGGGGGGGGGG	E P I F
1881	C Q L A D I T K S S S L Y F T C F L Y F E A Q V C D N V M E S N A K N C	SQIL
1921	PHQPTAL FROM A LFR RR K V V L N D K V N N FY T R L P F Q K L T G I S I R	D K V P
5941 1961	11G YOT GGA AAA CYC ATT YOC AAT GGG TYC TTT GAG TOT GAG COG CTC TOT GGAC AGG GAC CCT MOC TGCA CAT GGC TTT GGT TTT CTA AAT GTC TCC CAG CTG CAA GGA ( M S G K L I S N G F F E C E R L C D R D P C C T G F G F L N V S O L O G	GAGAGGTGACC GEVT
6061	ώτ το να το γία να να το και το γία το και το και το και το και το	TT GGG TGG TAC
6181	AG AÁG CT GČT GTA TSU AET GAC AČT CČT AGC TTC TGC CČT TČT GČT GCC CTU CAA TCC CTC AGT GAG GAG GTT ACT TCG GAC TT TGG GAG AG CTT ACT GAG GAG GTT ACT GAG GAG AGT ACT TGG GAG TT ACT TGG TGG TGG TT ACT TGG GAG TT ACT TGG TGG TGG TGG TGG TGG TGG TGG TGG T	P G W Y
2041	Q K P A V W S D T P S F C P S A A L Q S L T E E K V T S D <u>s</u> W Q T L A L	S S V I
2081	TANGAN CAN CAN CARC AND AND THE GAT GAT CARCAR AND	D C L V
6421 2121	CT ACC CTO CAG ATC CAG COT GOG GFT GTG AGA TOTT GTG TTC TAC COT GAT ATA CAG AAC TGC ATA CAG AGC TCG GG CCC AGG TGC TGG CTC CAT GAA ( T T L C I O P G V V P V P V P D T O N C T H S L B S H T C W L L L P C	AG GCT ACC TAC
6541	το της τος λής του τος της τος της τος τος τος τος τος τος λος τος λος τος τος τος τος της τος της τος της τος της τος της λα	TG GGC ACT GCC
6661	1 T K K S G I F L V Q S D V T S T F S V K I D S F G Q L Q G S Q V I K G AG AG CAS CTA GCA CCA GAG GTC TTG GAG GGT CAT GAG GT CAT GG AG GAG CAS CAS CTA GTC CAG GAG GTC TTG GAG GGT CAT GGAG GAT GAT TG GAG GGT CAT GGAG GAT GAT TG GAG GAT CAT GGAG GAT GAT GT D G AG GAT GAT GAT GAG TG TG GAG GAT GAT	CC AAG CCA AGG
2201	W K Q V Y R P L G V P Y A A P P L A D N R P R A P E V L N W T G S W D A	TKPR
2241	A S C W Q P G T R T P T P P Q I N E D C L Y 🛃 N V P V P E N L V S N A S	VLVF
6901	το όκο μαι ένα ένα ένα ένα ένα ένα ότι όσι όσι όσι όχο της μαι όσο τος έτο ένα ότι όσι όσι όσι ένα της ένα ότι ό	GG GTC TTC GGC
7021	TO THE AT THA GOL THE AT GAL GAL GAL GAL THE GOL GAL THE GAL GAL CAN ATT GAL GAL CAN ATT GAL GAL CAL THE GAL GAL THE THE GAL THE GAL THE GAL THE GAL THE GAL THE GAL THE G	AG COT GTG ACA
2321 7141	FLSSGSSDEVAGNNGLLLDQVAALTWVQSHIGGSCAGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	Q R V T
2361	L A A D R S G A D V A S I H L I I S R T R L Q L F K K A L L H G S A	L S P A
2401	A I I S P E K A Q Q Q A A T L A K E V G C P N F I H P G S G I M F R Q K	P À N I
7381 2441	TE ANT GAT GCT GAG ACC AAG CTC GTG GCT GTG AGT GGG CCT TTC CAC TAG GGG CCT GTG GTG GAT GGC CAG TAC CGT GAA, CTT CCA TCC AGA GGG TTG AAG L L N D À Q T K L L À V S G P P H Y W G P V V D G O V T. R R T. D S P D V D D C.	GG CCT CTG CCA
7501	ມວ່ານ້ອ ດມິດສາ ແມ່ນ ແລະ ປະສາຊັດ ແລະ ແລະ ເພື່ອມ ເພື່ອ ແລະ ແມ່ນ ແມ່ນ ທີ່ມີ ແລະ ແມ່ນ ທີ່ ແມ່ນ ແມ່ນ ແມ່ນ ແມ່ນ ແມ່ນ ແ	CC TTT TAC CAG
7621	ε τη στη στη στη τάν τλα τα τα τα σλα στη στη στη τάν ταν ταν ταν του του του της στη στη του της του του του του της του της του της του της του της της του της	co cro cro cas
2521 7741	A L V R S L G G E D S D A R I L A A A V W Y Y S L E R S T D D Y A S F S A GCC ACC GG GAC TAC TIT AIC AFC TOT COC ATG GCT AGG CATA GCG CAG GAG AACC COT GGG AAT FOR TAC CAT THE CAC GAR AGC TAT (	R A L E AGC CAT GGC AGC
2561 7861	NATRDYFIIC PNVNMASLWARD AR TROUT PNYNWSSEN NATRDYFI	G H G S
2601	The L L A D V Q Y A P G L P P Y S A Y Q G Q P S T E E Q S L S L K V M Q	Y F S N
7981 2641	TL ATLA AWA TLA WAA AAT CCT HAC TAC CCA CAT GAG GAT GTA TAG AAG GAG GCA GCT GGA GTAG TTT GCC ACA CCT TGG CCA GAC TTT ATT CCT GGA GCA GGA GGA GGA GGA GGA GGA GGA GGA	AG GAG CTC AGC K E L S
8101	ας αλα στη τος λλη σας αλα σας στο λλλα του σας τος τος της τος τος λλα της λής της λλα στη το λλα σλη σαλ σας λήα σλη σαλ τος λάτος της τος τος τος της τος τος τος της τος της τος της τος τος τος της της τος της της τος της της τος της της της της της της της της της τη	AG AGT GAA GAG
2681 8221	ми и гикии и и ки а и с S F W S K Y I Q T L K D A D G A K D A Q L T A A GA TH GAA GT GAA GT GAA GT GAA TH GA A GA A	K S E E
2721	EDLEVGPGLEEDLSGSLEPVPKSYSK * 2746	

FIG. 1. Tg cDNA sequence of cog/cog mice with congenital hypothyroid goiter differs by one nucleotide from that of euthyroid AKR/J mice. The complete cog Tg coding sequence and the primary structure of the deduced cog Tg protein are shown. The preparation of thyroid cDNA libraries from the two isogenic strains of mice, library screening, and DNA sequencing of selected clones, is described in Materials and Methods. The deduced cog Tg protein sequence differs in six locations from that recently reported in GenBank (accession no. U76389) for normal outbred mice (the differing codons are shown as underlined areas). Of these differences, only the nucleotide substitution (CTC to CCC) at position 6848, resulting in a Leu-2263  $\rightarrow$  Pro mutation which resides in the acetylcholinesterase-homology domain of Tg, differs between the Tg sequences of cog/cog and AKR/J mice.



FIG. 2. The *cog* mutation is located within a conserved structural domain. (A) The amino acids surrounding Leu-2263 (denoted with an arrowhead, at bottom) are highly conserved in Tg from all four species that have thus far been examined, as well as in AChE and other members of the " $\alpha/\beta$  hydrolase fold" family (a partial list is shown; h, human; b, bovine; r, rat; m, mouse; t, torpedo; c, candida; s, salmon; a, avian; d, *Drosophila*). The amino acid numbering for rat Tg is not established because the cDNA sequence has not yet been determined in its entirety. (B) Given the close proximity of Leu-2263 to Cys-2260, which forms an intrachain disulfide bond (solid bracket) that provides important structural stability to AChE (see text) and is likely to play a similar role in Tg (dashed bracket), we hypothesize that the cog mutation (L2263P) may destabilize this region, potentially by preventing proper formation of this disulfide bridge (signified by question mark).

acetylcholinesterase (AChE) and is known to be essential for the structural stability of the AChE molecule. Heterologous expression studies confirm that *cog* Tg exhibits deficient intracellular transport, whereas PCR-mediated reversion of Pro-2263 to Leu-2263 corrects the defect. Thus, the Leu-2263  $\rightarrow$  Pro mutation in *cog* Tg is responsible for the ERSD that causes thyroid dyshormonogenesis.

## MATERIALS AND METHODS

Animals. Homozygous *cog/cog* and AKR/J mice were purchased from The Jackson Laboratories. At 6 months of age, thyroid goiters were removed from four mutant mice for generation of a thyroid cDNA library. For the control cDNA library, freshly isolated thyroids from 12 AKR/J mice were employed.

Isolation and Sequencing of Tg cDNAs. Using the Gubler-Hoffman technique (24), thyroid cDNA libraries from cog/cog and AKR/J mice were generated. Bidirectional DNA strands were synthesized in a single tube without intervening extractions. Because Tg is encoded by a long message, the initial reaction using oligo(dT) (containing a terminal XhoI site) employed Superscript II reverse transcriptase (Moloney murine leukemia virus-reverse transcriptase devoid of RNase H from GIBCO/BRL). The first cDNA strand was protected by methylation. Fragments of mRNA that had been nicked by RNase H served as primers for DNA polymerase I in making the second cDNA strand. Following ligation of an EcoRI linker and digestion with XhoI, double-stranded cDNAs were then cloned into predigested  $\lambda$  arms of the ZAP express vector (Stratagene). The bacteriophage libraries contained a primary titer of >1 million plaque-forming units/ml.

Seven clones from the cog/cog library, presumably representing the full-length ( $\geq$ 8.2 kb) Tg cDNA, were initially selected by probing with a cDNA fragment containing 1.7 kb of the extreme 5' coding sequence of rat Tg (25). Five positive clones gave identical restriction patterns when digested with EcoRI and XhoI. By using the enzymatic dideoxy method (26), the terminal 100 bp at the 5' and 3' ends of the coding region



FIG. 3. Defective intracellular trafficking of *cog* Tg in COS cells. COS-7 cells were either untransfected or transiently tranfected with Tg cDNAs as described in *Materials and Methods*. By Western blot analysis, unlike the control Tg, a protein encoded by the *cog* Tg cDNA could be found intracellularly but was secreted into the overnight collection medium at levels below the limits of detection.

were found to be identical for all five clones. Three of these full-length clones that contained the most 5'-untranslated sequence were then subcloned into pBS+ (Stratagene). One such clone was then selected for sequencing (see Fig. 1), beginning with standard T3 and T7 oligonucleotide primers. The Tg cDNA was also digested with BamHI and PstI and appropriate fragments ligated into pBS+, so that additional internal sequence could be obtained from reactions by using these same standard primers. The remaining internal cDNA sequence was obtained from reactions by using primers designed on the basis of cog Tg sequences that were available from earlier sequencing runs. Using this approach, the cog Tg cDNA was completely sequenced in both directions, which helped to eliminate sequencing errors. Then, by using specific cog Tg cDNA fragments as probes, several large and overlapping fragments from the AKR/J library were isolated and sequenced (no single full-length Tg cDNA clone was isolated from this library). From this approach, the entire Tg coding sequence from the control mice was obtained. All sequencing was performed in the DNA sequencing facility at Beth Israel Hospital (Boston).

**Transient Expression of Mutant and Normal Tg.** COS-7 cells were grown in DMEM containing 10% FBS. Using lipofectamine (GIBCO/BRL), subconfluent cells were then tranfected either with the full-length *cog* Tg or normal Tg cDNA, both subcloned into the pBK expression vector (Stratagene); a nonmutant Tg cDNA in the pCB6 expression vector (27) was used in Fig. 2c. These vectors drive Tg expression from the immediate early cytomegalovirus promoter. At 48 h after transfection, the medium was changed to include 10 mM Na butyrate and the cells further incubated for 1 day. The bathing medium was then collected and precipitated with trichloroacetic acid, while the cells were lysed in 50 mM Tris (pH 7.4) plus boiling 2% SDS and 25 mM DTT. Both medium



FIG. 4. Cog Tg is retained in the ER and degraded intracellularly. COS-7 cells, transfected with nonmutant or cog Tg cDNAs as in Fig. 3, were pulse-labeled with [ $^{35}$ S]methionine/cysteine, chased for various times, and the medium collected and cells lysed as described. (A) Equal aliquots of cell lysates and chase medium were combined and immunoprecipitated with polyclonal antiserum against mouse Tg, prior to digestion with endo H (+) or mock digestion (-). As shown at 6 h of chase,  $\geq 60\%$  of normal Tg had passed the medial Golgi as measured by acquisition of resistance to endo H digestion (R) with secretion into the medium (not shown), whereas cog Tg remained entirely sensitive (S) to endo H digestion (open arrow), indicating that the mutant Tg protein never reached the Golgi for complex carbohydrate addition. (B) Equal aliquots of lysates of cells transfected with the cog Tg cDNA, and the corresponding chase medium samples, were immunoprecipitated with antisera against Tg and analyzed by SDS/PAGE and fluorography. Although  $\approx 40\%$  of cog Tg is slowly degraded over the first 12 h, during the second 12-h chase, the remaining newly synthesized cog Tg was more rapidly degraded with an undetectable fraction appearing in the secreted media, consistent with biphasic degradation kinetics (39).

and cell lysate samples were then diluted into gel sample buffer, resolved by 5% SDS/PAGE, and electrophoretically transferred to nitrocellulose. Immunoblotting employed rabbit polyclonal anti-Tg sera, with a goat anti-rabbit secondary antibody coupled to peroxidase. Blots were developed by using enhanced chemiluminescence with luminol. For immunoprecipitation analysis, transfected COS-7 cells were pulse-labeled with [<sup>35</sup>S]methionine/cysteine and chased for various times prior to lysis in a nondenaturing buffer containing a standard mixture of protease inhibitors (28). Equal aliquots of cell lysates or chase medium were precleared with Zysorbin prior to immunoprecipitation with polyclonal anti-Tg, followed by reducing SDS/PAGE. Endoglycosidase H digests were performed as described (21). Tg bands were detected by phosphorimaging (Molecular Dynamics).

## **RESULTS AND DISCUSSION**

By detecting for both the 5' and 3' ends of the Tg cDNA, five independent clones that contained full-length inserts ( $\geq$ 8.2 kb)



FIG. 5. Pro-2263  $\rightarrow$  Leu-corrected Tg is secreted like the wild-type protein. Using PCR-mutagenesis, the *cog* Tg cDNA was converted back to wild-type at residue Leu-2263. COS-7 cells, transfected with either the mutant or Pro-2263  $\rightarrow$  Leu-corrected Tg cDNA, were examined by immunoprecipitation, and SDS/PAGE as in Fig. 4. Radiolabeled Tg bands were quantitated by phosphorimaging.

were isolated from a thyroid cDNA library prepared from cog/cog mice (see Materials and Methods). Complete, bidirectional DNA sequence obtained for one clone (referred to as cog Tg) included an 8301-nt insert containing an ORF region encoding an intact pre-prohormone that contained a 20residue signal peptide and a 2746-residue mature protein, consistent with a full-length Tg translation product (Fig. 1). When compared with the Tg sequence recently reported in GenBank for normal outbred mice, the deduced cog Tg translation product differed in six locations (Fig. 1, underlines), reflecting heterogeneity among murine Tgs, similar to that reported to occur between normal humans (29). To determine which if any of these six sequence differences might be responsible for the cog phenotype, we also prepared a thyroid cDNA library and obtained the complete Tg cDNA sequence from AKR/J mice, the original parental inbred strain that does not exhibit any thyroid abnormality. Because the cog/cog and AKR/J stains are isogenic, if the cog defect is found in the Tg coding sequence, then only a single region of sequence difference is to be expected, whereas if the cog gene is not the Tg gene, the two Tg sequences should be identical.

When the two Tg cDNAs were compared, the sole observed difference was a CTC  $\rightarrow$  CCC substitution at position 6848 of the ORF, creating the mutation Leu-2263  $\rightarrow$  Pro in the mature *cog* Tg protein (Fig. 1, bold type). This was confirmed by selective sequencing of two independently isolated, full-length *cog* Tg cDNA clones. The proline substitution is not one of the sequence variations reported for normal outbred mice (Fig. 1). Importantly, the wild-type leucine residue in this position is conserved in all species for which suitable Tg sequence has been reported (Fig. 2*a*).

The C-terminal 544 amino acids of Tg exhibit significant homology to the entirety of acetylcholinesterase (AChE; refs. 30–32), which shares with Tg the features of dimer formation (33, 34) and transport through the secretory pathway (34). The highest degree of homology is contained in a region conserved between virtually all members of the large " $\alpha/\beta$  hydrolase fold" family, including hydrolytic enzymes and cell adhesion molecules that are thought to have diverged from a common ancestral gene (35). This conserved region in the crystal structure of AChE (Fig. 2*a*), and presumably in other family members, is contained within a peptide loop connecting antiparallel  $\beta$ -strands (33) that are thought to play a critical structural role.

An expanded view (Fig. 2b) demonstrates that an intrachain disulfide bond, which is adversely affected by changes in the neighboring amino acids (36), is in close proximity to the conserved peptide loop in AChE. Notably, a Asp  $\rightarrow$  Asn substitution in this conserved sequence has been shown to result in deficient protein export (36), a consequence of structural instability that probably represents either failure to form the intrachain disulfide bond (Fig. 2b) or the proper antiparallel alignment of  $\beta$ -strands. Based on such observations, we hypothesize that a structurally critical disulfide bond also exists in this region of Tg, and that the neighboring cog mutation may cause similar structural instability leading to deficient Tg export. Given that homodimerization of Tg occurs the ER (37), the structural instability provided by mutations in this region can account for the established deficiency of cog Tg in homodimerization and intracellular transport (21). This in turn can explain the abnormally distended thyrocyte ER with a vast induction of molecular chaperones (21).

By using the *cog* Tg cDNA, we transiently expressed the protein in COS cells. As shown by immunoblotting, untransfected COS cells contained only background bands in cell lysates (Fig. 3 *Left*) and no bands in the medium (Fig. 3 *Right*). In addition, COS cells transfected with a control cDNA encoding nonmutant Tg exhibited a specific Tg band intracellularly and secreted into the medium. By contrast, the expressed *cog* Tg protein was detectable only intracellularly (Fig. 3). These data confirm that the cloned *cog* cDNA encodes a Tg protein defective for secretion.

To determine if the recombinantly expressed *cog* Tg ever reached the Golgi complex, we examined its sensitivity to digestion with endoglycosidase H (21). As shown in Fig. 4*a*, *cog* Tg remained completely sensitive to digestion, indicating that the mutant protein was retained in the ER. Moreover, we found that the fate of transiently expressed *cog* Tg involves intracellular degradation as a function of time (Fig. 4*b*). Although the kinetics of degradation appeared relatively slow, it is consistent with the recently described ER-associated degradation process (38) that occurs in many ERSDs (2). Studies are currently ongoing to characterize this process for Tg in greater detail.

Finally, to prove that the transport defect in *cog* Tg can be specifically attributed to Pro-2263, we used PCR-based mutagenesis to specifically convert the *cog* Tg cDNA back to wild-type only at residue Leu-2263. As shown in Fig. 5, correction of this single residue restored normal Tg secretion, proving that the conserved leucine in this region is essential for the structural stability (21) and intracellular transport (this report) of the Tg protein.

In conclusion, we have identified a point mutation contained within the AChE-homology domain of the Tg coding sequence as the molecular basis for congenital hypothyroid goiter in *cog/cog* mice, a prototypic ERSD. Further, though formally possible, it is unlikely that the *cog/cog* mouse has a second genetic lesion. Future work will be aimed at precisely defining how the AChE-like domain may regulate the stability and assembly of Tg homodimers. Finally, although the molecular lesion is known in relatively few human cases, the fact that a similar thyroid ERSD is observed in numerous patients suffering from congenital hypothyroid goiter suggests that subtle defects in the Tg coding sequence in these kindreds may be expected.

This work was supported by National Institutes of Health Grants DK-40344 to P.A. and DK-02113 and DK-52076 to P.S.K., as well as support from Knoll Pharmaceuticals.

- 1. Hammond, C. & Helenius, A. (1995) Curr. Opin. Cell Biol. 7, 523–529.
- 2. Kim, P. S. & Arvan, P. (1998) Endocr. Rev. 19, 173-202.

- Ward, C. L., Omura, S. & Kopito, R. R. (1995) *Cell* 83, 121–127.
  Sifers, R. N., Finegold, M. J. & Woo, S. L. (1992) *Semin. Liver Dis.* 12, 301–310.
- 5. Prockop, D. J. & Kivirikko, K. I. (1995) Annu. Rev. Biochem. 64, 403–434.
- Ito, M., Jameson, J. L. & Ito, M. (1997) J. Clin. Invest. 99, 1897–1905.
- Pathak, R. K., Merkle, R. K., Cummings, R. D., Goldstein, J. L., Brown, M. S. & Anderson, R. G. (1988) *J. Cell Biol.* 106, 1831–1841.
- Medeiros-Neto, G. & Stanbury, J. B. (1994) Inherited Disorders of the Thyroid System (CRC, Boca Raton, FL), pp. 107–138, 221.
- Ricketts, M. H., Simons, M. J., Parma, J., Mercken, L., Dong, Q. & Vassart, G. (1987) Proc. Natl. Acad. Sci. USA 84, 3181–3184.
- Ieiri, T., Cochaux, P., Targovnik, H. M., Suzuki, M., Shimoda, S., Perret, J. & Vassart, G. (1991) J. Clin. Invest. 88, 1901–1905.
- Medeiros-Neto, G., Kim, P. S., Yoo, S. E., Vono, J., Targovnik, H., Camargo, R., Hossain, S. A. & Arvan, P. (1996) *J. Clin. Invest.* 98, 2838–2844.
- Targovnick, H. M., Vono, J., Billerbeck, E. C., Cerrone, G. E., Varela, V., Mendive, F., Wajchenberg, B. L. & Medeiros-Neto, G. (1995) *JCEM* 80, 3356–3360.
- 13. Sugisaki, T., Beamer, W. G. & Noguchi, T. (1992) Neurochem. Res. 17, 1037–1040.
- Mayerhofer, A., Amador, A. G., Beamer, W. G. & Bartke, A. (1988) J. Hered. 79, 200–203.
- Michel-Bechet, M., Cotte, G., Codaccioni, J.-L. & Athouel-Haon, A.-M. (1969) Acta Anat. 73, 389–409.
- Lissitzky, S., Torresani, J., Burrow, G. N., Bouchilloux, S. & Chabaud, O. (1975) *Clin. Endocrinol.* 4, 363–392.
- Ohyama, Y., Hosoya, T., Kameya, T., Suzuki, N., Nakamura, S., Kazahari, K., Shibayama, K., Yokota, Y. & Matsura, N. (1994) *Clin. Endocrinol.* 41, 129–135.
- Basche, M., Beamer, W. G. & Schneider, A. B. (1989) *Endocrinology* **124**, 1822–1829.
- Adkison, L. R., Taylor, S. & Beamer, W. G. (1990) J. Endocrinol. 126, 51–58.
- Fogelfeld, L., Harel, G., Beamer, W. G. & Schneider, A. B. (1992) *Thyroid* 2, 329–335.
- Kim, P. S., Kwon, O.-Y. & Arvan, P. (1996) J. Cell Biol. 133, 517–527.
- Beamer, W. G., Maltais, L. J., DeBaets, M. H. & Eicher, E. M. (1987) *Endocrinology* 120, 838–840.
- Taylor, B. A. & Rowe, L. (1987) Proc. Natl. Acad. Sci. USA 84, 1986–1990.
- 24. Gubler, U. & Hoffman, B. J. (1983) Gene 25, 263-269.
- 25. Graves, P. N. & Davies, T. F. (1990) Mol. Endocrinol. 4, 155-161.
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. & Smith, M. (1977) *Nature (London)* 265, 687–695.
- 27. Muresan, Z. & Arvan, P. (1997) J. Biol. Chem. 272, 26095–26102.
- 28. Kim, P., Bole, D. & Arvan, P. (1992) J. Cell Biol. 118, 541-549.
- van de Graaf, S. A., Pauws, E., de Vijlder, J. J. & Ris-Stalpers, C. R. (1997) Eur. J. Endocrinol. 136, 508–515.
- Schumacher, M., Camp, S., Maulet, Y., Newton, M., MacPhee-Quigley, K., Taylor, S. S., Friedmann, T. & Taylor, P. (1986) *Nature (London)* 319, 407–409.
- Swillens, S., Ludgate, M., Mercken, L., Dumont, J. E. & Vassart, G. (1986) Biochem. Biophys. Res. Commun. 137, 142–148.
- Mori, N., Itoh, N. & Salvaterra, P. M. (1987) Proc. Natl. Acad. Sci. USA 84, 2813–2817.
- Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. & Silman, I. (1991) *Science* 253, 872–879.
- 34. Bourne, Y., Taylor, P. & Marchot, P. (1995) Cell 83, 503-512.
- Cousin, X., Hotelier, T., Giles, K., Toutant, J. P. & Chatonnet, A. (1998) *Nucleic Acids Res.* 26, 228–230.
- Shafferman, A., Kronman, C., Flashner, Y., Leitner, M., Grosfeld, H., Ordentlich, A., Gozes, Y., Cohen, S., Ariel, N., Barak, D., Harel, M., Silman, I., Sussman, J. L. & Velan, B. (1992) *J. Biol. Chem.* 267, 17640–17648.
- 37. Kim, P. S. & Arvan, P. (1991) J. Biol. Chem. 266, 12412-12418.
- 38. Brodsky, J. L. & McCraken, A. A. (1997) Trends Cell Biol. 7, 151–156.
- Tsao, Y. S., Ivessa, N. E., Adesnik, M., Sabatini, D. D. & Kreibich, G. (1992) J. Cell Biol. 116, 57–67.