The transthyretin-related protein family

Therese Eneqvist[1,2], Erik Lundberg[1], Lars Nilsson[1], Ruben Abagyan[2] and A. Elisabeth Sauer-Eriksson[1]

[1]*Umeå Centre for Molecular Pathogenesis, Umeå University, Sweden;* [2]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, California, USA*

A number of proteins related to the homotetrameric transport protein transthyretin (TTR) forms a highly conserved protein family, which we present in an integrated analysis of data from different sources combined with an initial biochemical characterization. Homologues of the transthyretin-related protein (TRP) can be found in a wide range of species including bacteria, plants and animals, whereas transthyretins have so far only been identified in vertebrates. A multiple sequence alignment of 49 TRP sequences from 47 species to TTR suggests that the tertiary and quaternary features of the three-dimensional structure are most likely preserved. Interestingly, while some of the TRP orthologues show as little as 30% identity, the residues at the putative ligand-binding site are almost entirely conserved. RT/PCR analysis in *Caenorhabditis elegans* confirms that one TRP gene is transcribed, spliced and predominantly expressed in the worm, which suggests that at least one of the two *C. elegans* TRP genes encodes a functional protein. We used double-stranded RNA-mediated interference techniques in order to determine the loss-of-function phenotype for the two TRP genes in *C. elegans* but detected no apparent phenotype. The cloning and initial characterization of purified TRP from *Escherichia coli* reveals that, while still forming a homo-tetramer, this protein does not recognize thyroid hormones that are the natural ligands of TTR. The ligand for TRP is not known; however, genomic data support a functional role involving purine catabolism especially linked to urate oxidase (uricase) activity.

Keywords: *Escherichia coli*; homology model; purine catabolism; sequence analysis; transthyretin-related protein.

Transthyretin (TTR) is a transport protein in extracellular fluids of vertebrates, where it distributes the two thyroid hormones 3,5,3′-triiodo-L-thyronine (T_3) and 3,5,3′,5′-tetra-iodo-L-thyronine (thyroxine, T_4), as well as vitamin A in complex with retinol-binding protein [1]. TTR has so far been identified in piscine, amphibian, reptilian, avian, marsupial, and eutherian vertebrates [2,3]. The three-dimensional structure of TTR is a homotetramer of 55 kDa. Each monomer of 125–130 amino acids comprises eight β-strands denoted A–H organized into two four-stranded β-sheets and one short α-helix [4,5]. The dimer–dimer association creates a central hydrophobic channel where the two hormone-binding sites are situated [6], while the two retinol-binding protein binding sites are positioned on the surface of the molecule [7,8]. Human TTR is associated with two clinical forms of amyloidosis; senile systemic amyloidosis involves the native protein [9], whereas familial amyloidotic polyneuropathy is caused by single point mutations [10,11]. More than 70 mutations distributed over the entire sequence are associated with the disease [3,12]. So far, it is not known if TTR can cause amyloidosis in other species. We are studying a novel family of TTR-related proteins (TRPs) and have identified 49 sequences from 47 different species (Table 1). The predicted protein sequences from *Escherichia coli*, *Bacillus subtilis*, *Schizosaccharomyces pombe* and *Caenorhabditis elegans* are listed as TTR-like in SwissProt and trEMBL [13–15]. In this study we show that the extent of organisms carrying this gene is large and comprises bacteria, plants, and animals including vertebrate species. The four amino acid sequence motif Y-R-G-S at the C-terminal end of the protein unambiguously separate members of the TRP family, not only from TTR but also from other sequences listed as TTR-like in databases (with a particularly large number of representatives found in *C. elegans*). By analysing data from existing gene expression profile analysis based on DNA micro arrays in *C. elegans* [16,17], we find that the TRP genes are transcriptionally regulated during development and thus most likely encode functional proteins. We have performed an RT/PCR analysis that confirms the expression of one of these genes in *C. elegans*, and used double-stranded (ds) RNA-mediated interference in order to determine the loss-of-function phenotype for the TRP genes. We have also cloned and expressed TRP from *E. coli* and performed a characterization of the protein with size exclusion chromatography, thyroid hormone-binding studies, and amyloid formation by partial acid denaturation in comparison to human and fish TTR. A recent study by Shultz and colleagues showed that the gene *yunM* encoding TRP in *Bacillus subtilis* is essential for urate oxidase (uricase)

Table 1. Sequences from TTR related proteins. The sequences comprise putative protein sequences from SwissProt/TrEBML[SP] and GenPept[GP], and translated genome sequences[GB] and expressed sequence tags[EST] from GenBank, genome sequence projects[G], in some cases from unpublished, incomplete sequencing data[(G)].

Accession code	Species	Species description
NT_010542 [GB]	*Homo sapiens*	Human
AV594153 [EST], AV594154 [EST]	*Bos taurus*	Cattle or cow
BAB23318 [GP], BAB28659 [GP]	*Mus musculus*	Common house mouse
BF282409 [EST]	*Rattus norvegicus*	Norway rat or brown rat
AW637709 [EST]	*Xenopus laevis*	African clawed frog
BI888796 [EST]	*Danio rerio*	Zebra fish
BG935139 [EST]	*Salmo salar*	Atlantic salmon
BE468943 [EST]	*Ictalurus punctatus*	Channel catfish
AE003828 [GB]	*Drosophila melanogaster*	Fruit fly
AC084648 [GP]	*Caenorhabditis briggsae*	Roundworm, a free-living nematode
(1) Q21882 [SP]	*Caenorhabditis elegans*	Roundworm, a free-living nematode
(2) O44578 [SP]		
BG734131 [EST]	*Ostertagia ostertagi*	Parasitic stomach worm
BE919514 [EST]	*Solanum tuberosum*	Potato
BF114284 [EST]	*Lycopersicon esculentum*	Tomato
BE824466 [EST]	*Glycine max*	Soybean
BAA96913 [GP]	*Arabidopsis thaliana*	Thale cress or common wall cress
AW680248 [EST]	*Sorghum bicolor*	Broomcorn
AI783313 [EST]	*Zea mays*	Corn, Indian corn or maize
AL503126 [EST]	*Hordeum vulgare*	Barley
BE586587 [EST]	*Secale cereale*	Rye
BE444381 [EST]	*Triticum aestivum*	Wheat
BE582998 [EST]	*Phytophthora sojae*	Parasitic fungi that causes brown rot
AL435917 [EST]	*Pichia angusta*	Yeast-like fungi
CAA20843 [GP]	*Schizosaccharomyces pombe*	Fission yeast growing on decaying matter
AAF10734 [GP]	*Deinococcus radiodurans*	Gram-positive radiation-resistant coccus
NC_002974 [(G)]	*Mycobacterium smegmatis*	Non-pathogenic mycobacterium
T34864 [SP]	*Streptomyces coelicolor*	Gram-positive spore forming bacterium
BAB04479 [GP]	*Bacillus halodurans*	Alkaliphilic spore-forming bacterium
O32142 [SP]	*Bacillus subtilis*	Gram-positive nonpathogenic soil bacterium
AAK24588 [GP]	*Caulobacter crescentus*	Gram-negative bacterium in water and soil
NC_002969 [(G)]	*Brucella suis*	Gram-negative bacterium that causes brucellosis
NP_105848 [GP]	*Mesorhizobium loti*	Gram-negative nitrogen-fixing bacterium
AAK88066 [GP]	*Agrobacterium tumefaciens*	Gram-negative nitrogen-fixing bacterium
(1) CAC49566 [GP]	*Sinorhizobium meliloti*	Gram-negative nitrogen-fixing bacterium
(2) CAC49189 [GP]		
NC_002718 [(G)]	*Rhodobacter sphaeroides*	Phototrophic gram-negative bacterium
NC_002710 [(G)]	*Burkholderia cepacia*	Gram-negative opportunistic bacterium
NC_002930 [(G)]	*Burkholderia pseudomallei*	Gram-negative bacterium that causes melioidosis
P76341 [SP]	*Escherichia coli*	Gram-negative bacterium, abundant in colon
NC_002961 [(G)]	*Salmonella dublin*	Gram-negative bacterium that causes enteritis
CAD08223 [GP]	*Salmonella typhi*	Gram-negative bacterium that causes typhoid
AAL20029 [GP]	*Salmonella typhimurium*	Gram-negative bacterium that causes paratyphoid
NC_002924 [(G)]	*Actinobacillus actinomycetemcomitans*	Gram-negative bacterium found in lesions
AAG04907 [GP]	*Pseudomonas aeruginosa*	Gram-negative opportunistic bacterium
NC_002716 [(G)]	*Pseudomonas fluorescens*	Gram-negative fluorescent bacterium
NC_002947 [(G)]	*Pseudomonas putida*	Gram-negative opportunistic bacterium
NC_002949 [(G)]	*Pseudomonas syringae*	Gram-negative bacterium pathogenic to plants
CAB72989 [GP]	*Campylobacter jejuni*	Gram-negative bacterium that causes enteritis

activity and is coregulated with three other genes (*yunJ*, *yunK*, and *yunL*) encoding two permease homologues presumed to be responsible for uric acid transport and a putative urate oxidase [18]. In this report we attempt to summarize the data relating to this protein available from public databases and discuss this information with regard to its putative role in purine catabolism.

Experimental procedures

Multiple sequence alignment and analysis

Sequences were derived from GenBank, GenPept, Swiss-Prot and TrEMBL using BLAST [19] and FASTA [20], conveniently managed with the Biology Workbench

available from San Diego Supercomputer Center at http://workbench.sdsc.edu. In most cases default parameters were applied. For example, in the BLAST searches an amino acid sequence was used to scan either protein sequences (blastp) or translated nucleotide sequences (tblastn), using the BLOSUM62 substitution matrix with a gap-opening penalty of 11 and a gap extension cost of 1. The only significant homologues to TTR and TRP (according to alignment scores and *E*-values) were members of these two families. The TRPs were easily separated from TTRs by their characteristic C-terminal consensus sequence Y-R-G-S. Preliminary sequence data was obtained from The DOE Joint Genome Institute at http://www.jgi.doe.gov, The Institute for Genomic Research website at http://www.tigr.org, the Advanced Center for Genome Technology at the University of Oklahoma at http://www.genome.ou.edu, the Department of Microbiology at the University of Illinois at http://www.salmonella.org, and the Sanger Centre at http://www.sanger.ac.uk. The bovine sequence is the result of combined EST sequences, none of which contains the whole protein coding sequence. The human TTR-like protein was derived from translated chromosomal DNA and is a sum of partly overlapping nucleotide stretches from different reading frames and no evidence of expression has yet been observed. Similarly, the sequence from fruit fly comprises a section of translated chromosomal DNA. The putative signal peptides were predicted using the SignalP WWW server at the Center for Biological Sequence Analysis [21], and predictions of cellular localization were performed with PSORT [22]. The multiple sequence alignment was constructed with CLUSTALW [23], using the Gonnet weight matrix with gap opening and gap extension penalties of 10.0 and 0.20, respectively. The phylogenetic tree was created from the prealigned sequences using the Neighbour Joining method [24] and plotted with DRAWGRAM, which is part of the program package PHYLIP [25].

Three-dimensional model of the *E. coli* protein by homology modelling

The homology model of *E. coli* TRP was based on the 1.5 Å crystal structure of human TTR [Protein Data Base (PDB) code 1F41] and refined using the Internal Coordinate Mechanics (ICM) energy optimization method [26,27]. Briefly, the starting model that displays idealized geometry and comprises all atoms including hydrogens was created from a structure-sequence alignment generated by the zero end gap dynamic programming algorithm where the backbone and conserved side chains adopt the same conformation as the template. Loop regions defined by the sequence–structure alignment are subject to search against a database of loop structures from the PDB and loops with the closest matching sequences and loop end positions are inserted into the homology model. The structure was relaxed to relieve the steric strain by a regularization procedure [26], before prediction of the side chain conformations effected with the Biased Probability Monte Carlo method [28], followed by a second regularization procedure. Coordinates for this model can be requested from T.E.

Detection and characterization of the *C. elegans* TRP transcript

Total RNA was isolated from a mixed-stage population of *C. elegans* nematodes by a guanidine thiocyanate procedure [29]. The detection and characterization of the TRP transcripts by RT/PCR was performed using the Superscript One-step RT/PCR system (Invitrogen) and 2 μg total *C. elegans* RNA as a template. PCR amplification was performed with the 5′-primer 5′-GGTTTAATTACCCAA GTTTGAG-3′ that corresponds to the *C. elegans* spliced-leader 1 (SL1) sequence [30]. The 3′-primers correspond to distal portions of the cDNA sequence specific for either R09H10.3 (5′-TTTGGTACCTTATGATCCACGGTAT GTAGAGTATC-3′) or the ZK697.8 gene (5′-TTTGGTA CCAGTTGCTAAAAATCTTCTAATTTG-3′). The splicing pattern as well as the extreme 5′ end of the R09H10.3 transcript were determined by sequence analysis of the DNA fragment amplified by the R09H10.3 gene specific primer.

RNAi in *C. elegans*

Standard methods were used for culturing *C. elegans* on nematode growth medium [31]. A segment of the R09H10.3 and ZK697.8 genes, designated for RNA-mediated interference (RNAi) were amplified from genomic DNA prepared from the wild type N2/Bristol *C. elegans* strain. The primer pair 5′-TTTTTCATGATTCACGCAAGACAATGGG-3′ and 5′-TTTGGTACCTTATGATCCACGGTATGTAG-3′ amplified a 225-bp segment spanning exon 3 of R09H10.3, while the primers 5′-TTTTTCATGAGTACAAATTAGA AGATTTTTAGC-3′ and 5′-TTTGGTACCTGTGATCC AATATTAGTCCAT-3′ amplified a 170-bp segment spanning one of the predicted exons of ZK697.8. The fragments were subcloned into the vector L4440 [32], between two T7 promoters in inverted orientation. The cloned plasmids were individually transformed into the *E. coli* strain HT115(DE3). This strain is RNAseIII-deficient and carries isopropyl thio-β-D-galactoside (IPTG) inducible expression of T7 polymerase, which has been shown to be beneficial for RNAi by feeding [33]. The optimized feeding conditions reported by Kamath *et al.* were used to maximize observable phenotypes [34]. Briefly, transformed HT115 were grown overnight, mixed and seeded onto nematode growth medium plates containing 1 mM IPTG and 50 μg·mL^{-1} ampicillin followed by induction at room temperature overnight. L4 stage hermaphrodite worms were placed onto nematode growth medium plates containing seeded bacteria expressing dsRNA for either R09H10.3 or ZK697.8 and incubated for 24 h at 20 °C. Subsequently, three worms were replica plated onto plates seeded with the same bacteria and allowed to lay eggs for an additional 24 h before being removed. Progeny were scored for embryonic lethality after a further 24 h at 20 °C (presence of unhatched eggs) and for postembryonic phenotypes (such as sterility, aberrant morphology, uncoordinated movements, egg-laying defects, or slow growth) after several successive 12–24 h intervals.

Cloning of *E. coli* TRP

The construct corresponding to the complete amino acid sequence was amplified from chromosomal DNA of *E. coli*

strain K12-MG1655 using the primers 5′-CATGCC ATGGTAAAGCGTTATTTAGTACTC-3′ tagged with a 5′-*Nco*I cleavage site and 5′-TTTCGAGCTCTTAACTG CCACGATAGGTTG-3′ tagged with a 3′-*Sac*I site (Interactiva Virtual Laboratory). A construct corresponding to the mature protein without the predicted signal sequence [21] was amplified in a similar manner using the same 3′-*Sac*I primer and the primer 5′-CATGCCATGGCA CAACAAAACATTCTTAG-3′, introducing a N-terminal methionine and a *Nco*I cleavage site. After digestion with *Nco*I and *Sac*I (New England Biolabs/Amersham Pharmacia Biotech), the fragment was introduced into a pET24d vector (kindly provided by Gunter Stier, EMBL-Heidelberg, Germany) also cleaved with *Nco*I and *Sac*I, using the T4 DNA ligase Ready-To-Go kit (Amersham Pharmacia Biotech). The ligated vector was used to transform [35] *E. coli* DH5α, which were plated onto Luria–Bertani (LB) agar plates containing 30 μg·mL^{-1} kanamycin (Km). The subsequent transformants were collected for plasmid preparation using Wizard Plus SV Minipreps (Promega). The plasmids were digested with *Bam*HI (New England Biolabs), whose cleavage site is situated within the region of the pET24d cloning cassette supposedly replaced by the *E. coli* TRP gene, and used for a second transformation of DH5α plated on 30 μg·mL^{-1} Km LB agar plates. The constructs were sequenced using the DYEnamic ET terminator kit (Amersham Pharmacia Biotech) and an ABI 377 sequencer.

Protein expression and purification

Competent *E. coli* BL21 cells were transformed [35] and plated onto LB agar plates containing 30 μg·mL^{-1} Km. One colony was picked and grown in LB with 30 μg·mL^{-1} Km at 37 °C to optical density (OD)$_{600\ nm}$ = 0.9, induced with 0.2 mM IPTG for 2 h, harvested by centrifugation and stored at −20 °C. Frozen cells were thawed and lysed in 10 mL water including ≈ 1 mg lysozyme and 1 mM MnCl$_2$ for 10 min. DNase I was added followed by incubation for another 10 min and centrifugation at 25 000 g for 15 min. The construct of the immature protein generates two products that were analysed by N-terminal sequencing. One product corresponds to the intact sequence and the other represents the processed mature protein, which proves that the signal sequence was cleaved after Ala23 as predicted. In all subsequent experiments the construct of the mature protein was used after purification by ion exchange batch chromatography using SP-sepharose (Amersham Pharmacia Biotech) using 20 mM Hepes and 50 mM NaCl, pH 7.0 as wash and loading buffers, respectively. The elution buffer included also 1 M (NH$_4$)$_2$SO$_4$. Protein fractions were analysed by SDS/PAGE on 20% polyacrylamide gels using the Phast system (Amersham Pharmacia Biotech). Fractions containing pure *E. coli* TRP were pooled, dialysed against 50 mM Tris pH 7.5 with 200 mM NaCl, concentrated to 5 mg·mL^{-1} (Centriprep, Amicon) and stored at −20 °C. The molecular weight of the purified protein was determined by mass spectrometry to 13 013 Da for the monomer, which was 130 Da lower than expected from the sequence. Most likely this reduction corresponds to incomplete incorporation of the initial methionine residue. Recombinant fish TTR cloned from

Sparus aurata cDNA (T. Eneqvist & A.E. Sauer-Eriksson, unpublished data), human TTR and the ATTR V30M mutant were expressed in a similar fashion as the *E. coli* protein then purified by preparative native PAGE on a 10% gel (Model 491 Prep Cell, Biorad) equilibrated with 0.025 M Tris pH 8.5/1.9 M glycine [5]. Fractions containing pure TTR were pooled, dialysed against 50 mM Tris pH 7.5, and concentrated to 5 mg·mL^{-1} (Centriprep, Amicon), then stored at −20 °C.

Size exclusion chromatography

A Superdex 75 column (Amersham Pharmacia Biotech) was pre-equilibrated with 50 mM Tris pH 7.0 containing 200 mM NaCl. Approximately 2 mL purified *E. coli* TRP at 1 mg·mL^{-1} in the same buffer was injected, and the eluted protein was detected by measuring the absorbance at 280 nm. As molecular weight standard, a set of low molecular mass gel filtration standards (Amersham Pharmacia Biotech) containing ribonuclease A (13.7 kDa), chymotrypsinogen A (25.0 kDa), ovalbumin (43.0 kDa), and BSA (67.0 kDa) was analysed under similar conditions.

Partial acid denaturation

Using the protocol described for human TTR [36], purified proteins of *E. coli* TRP, human TTR, and the human amyloidogenic variant ATTR V30M were diluted to final concentrations of 0.2 mg·mL^{-1} in buffers appropriate for the desired pH (e.g. 50 mM NaOAc/NaPO4 and 100 mM KCl). After 72 h of incubation at 37 °C, all samples were thoroughly vortexed to distribute equally all potential amyloid fibrils, and analysed by optical density (OD) measurements at 330 nm in a standard UV cell.

Thyroid hormone binding

Two poly(vinylidene difluoride) membranes were washed in methanol followed by TBS buffer (20 mM Tris pH 8.2, 1 M NaCl). The membranes were allowed to semidry before circles were marked with a pencil. Three μL human TTR, fish TTR, *E. coli* TRP, and BSA at four different concentrations (2, 1, 0.5, and 0.1 mg·mL^{-1}) were applied in their allocated rings. The drops (≈ 110, 55, 28, and 5.5 pmol) were allowed to dry before the membranes were placed in a 5% skim-milk/TBS, and gently shaken for 1–2 h at 4 °C. The membranes were subsequently placed into two separate solutions: one including ≈ 6 μCi (≈ 7.7 pmol) T$_3$ and the second ≈6 μCi (≈ 6.2 pmol) T$_4$, both in 30 mL TBS. The membranes were further incubated for 1–2 h. The filters were washed in Tween/TBS for 10 min, and the level of T$_4$ and T$_3$ binding was evaluated using a phosphoimager. T$_4$ and T$_3$ were purchased from New Life Science Products, Inc.

ANS-binding studies

Fluorescence measurements were made with a Fluoro-Max-2 spectrofluorometer (Jobin Yvon) scanning emission fluorescence from 440 to 550 nm. Emission spectra of 8-anilinonaphthalene-1-sulphonate (ANS) were recorded in a solution of phosphate-buffered saline (NaCl/Pi, 137 mM NaCl, 3 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$

pH 7.4) in the absence or presence of either human TTR or *E. coli* TRP (14 ng·μL^{-1}). Experiments during which ANS concentrations varied from 1 to 40 μM yielded essentially identical results (data not shown).

Results

Sequence analysis of the TRP family

A multiple sequence alignment with 49 TTR-related protein sequences from 47 species was compared to human TTR (Fig. 1). The TRP sequences are ≈ 35% identical to the TTR family, while the sequence identity within the TRP family is 30–95%. However, the TRPs have a very distinguished consensus sequence that clearly identifies them as belonging to a separate protein family. The consensus is particularly evident in the C-terminal end, where the TRP sequences differ remarkably from those of transthyretin. The sequence identity between TRP and TTR from mouse is 32%. Interestingly, the sequence identity between mouse TRP and fish TTR is higher (37%). The TTR-related proteins from rat and mouse are ≈ 95% identical, analogous to the identity between the TTR sequences from those species.

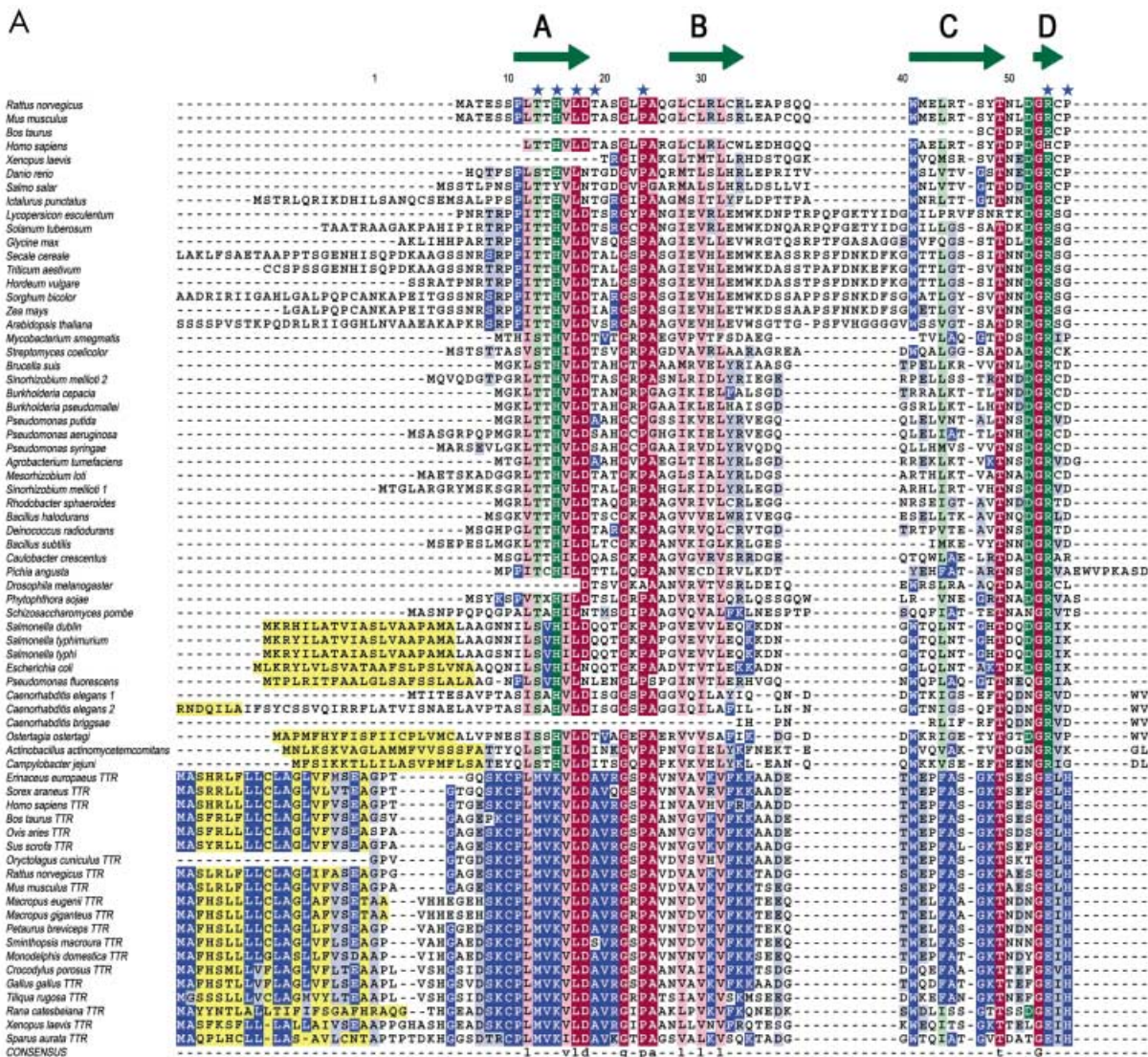From human chromosomal DNA a 111-amino acid sequence with 76% similarity to the mouse TRP could be

Fig. 1. Multiple sequence alignment. Amino acid sequences of TTR-related proteins from 47 species aligned and compared with TTR sequences from 20 species (reviewed by Eneqvist *et al.* [3]). Similarity was defined as amino acid substitutions within one of the following groups: FYW, IVLM, RK, DE, GA, TS, and NQ. Positions that are more than 80% identical are red, and those more than 80% similar are pink. Residues displaying an identity of 80% or higher within the TRP family are shown in dark green, while those more than 80% similar are light green. Similarly, positions displaying above 80% identity and 80% similarity in the TTR family are shown in dark and light blue, respectively. Confirmed or predicted signal peptides are indicated with yellow background colouring. Numbering and secondary structure elements are based on human TTR and are shown as green arrows (β-strands) and a red box (α-helix). Residues lining the hormone-binding channel in TTR are marked with blue stars. The N-terminal sequences of TRPs (residues preceding 10 according to human TTR numbering) were not aligned, whereas these residues in TTR were aligned manually.

Fig. 1. (Continued).

assembled from partly overlapping nucleotide stretches in the long arm of chromosome 16. This region is part of a working draft sequence segment and contains several repetitive elements, which makes it unreliable. Similarly, the incomplete TRP sequence from *Drosophila melanogaster* was also derived from genomic data from a translated region. Therefore, it is still unclear if these species have a functional TRP gene.

According to predictions by the SignalP WWW server at the Center for Biological Sequence Analysis [21] the majority of TTR-related proteins are cytoplasmic. In the Gram-negative enterobacteria *E. coli*, *Salmonella* and *Campylobacter jejuni*, the fluorescent bacterium *Pseudomonas fluorescens* (but not in the remaining *Pseudomonas* species), and *Actinobacillus actinomycetemcomitans*, putative signal sequences suggest that these proteins are localized at the periplasm. Indeed, a large-scale N-terminal sequencing project has verified the expression and the predicted cleavage site of TRP in *E. coli* [37]. Signal sequences were also predicted in sequences from the nematodes *C. elegans*

and *Ostertagia ostertagi*, which implies that those proteins are secreted.

As the majority of the sequences identified in higher organisms are derived from expressed sequence tags (ESTs), their N-terminal ends are incomplete. However, using the cellular localization server PSORT [22] the amino acid composition of TRP from plants and animals was considered to be peroxisomal. The proteins do not contain any of the two identified peroxisome signal signatures [38], but several experimentally verified peroxisomal proteins have yet unidentified means of targeting.

The putative protein sequence AAC33718 from *Salmonella dublin* is the only TRP that contains a longer C-terminal end, with 36 additional residues following the TTR-related part. However, the preliminary *Salmonella dublin* genome sequence data from the Department of Microbiology at the University of Illinois (accession code NC_002961) contradicts that fact and instead indicates a protein with the familiar C-terminal end (Fig. 1).

Two nonidentical sequences were found in the nematode *C. elegans* and the nitrogen-fixing bacterium *Sinorhizobium meliloti*. In *C. elegans* the two sequences, O44578 located on chromosome V (gene product of ZK697.8) and Q21882 on chromosome IV (gene product of R09H10.3) are 91% identical over the TRP domain, but very different in length. The first protein contains 70 amino acids preceding the TTR-homology element predicted to represent an unusually long signal peptide with a cleavage site after position 52, whilst the second gene product consists of the TTR-related component alone. The two *Sinorhizobium meliloti* sequences, none of which contain a predicted signal sequence, are of more similar length with 129 vs. 123 residues and are 64% identical to each other.

The TRP gene derived from chromosomal DNA of *Arabidopsis thaliana* encodes a protein of 324 residues. Preceding the TTR-related domain is an N-terminal domain of ≈ 190 amino acids, which is 27% identical to the N-terminal half of a putative uricase from *Bacillus subtilis* (NP_391125).

Detection and characterization of the *C. elegans* TRP transcripts

Gene expression profile studies using DNA microarray technology suggest that both R09H10.3 and ZK697.8 are expressed in the worm [16,17,39]. However, whereas two EST clones are available for R09H10.3 (yk1092605 and yk869d09), no ESTs are available for ZK697.8. To determine if any of these two TRP genes are in fact expressed in *C. elegans* we performed RT/PCR analysis, using-3′ primers specific for either R09H10.3 or ZK697.8 (Fig. 2). We were able to demonstrate that the shorter gene R09H10.3 is expressed in *C. elegans* (Fig. 2A, lane 1), but were unable to amplify any cDNA derived from the ZK697.8 gene, which suggests that this gene is not expressed under normal growth conditions (Fig. 2A, lane 2). By comparing the sequence of the amplified cDNA of R09H10.3 to that of the available genomic sequence, we were able to deduce the organization of exons and introns. The splicing between the two protein coding exons predicted in the databases was confirmed, and we also identified an additional small exon located ≈ 1 kb upstream (Fig 2B and C). This exon contains an additional ATG in frame with the TRP reading frame of the following exons, representing a possible alternative translational start site for R09H10.3.

To characterize the 5′ end of the TRP cDNA we used a 5′ primer that corresponds to the *C. elegans* spliced-leader 1 (SL1) sequence [30]. The 5′ termini of most *C. elegans* mRNAs are modified by incorporation of a 22-nucleotide, nontranslated 'leader' sequence that is donated by a distinct 100-nucleotide SL1 RNA transcript. This *trans*-splicing event generates a short 5′ untranslated region and introduces an essential tri-methylguanosine cap at the 5′ end of the mRNA. The organization of the R09H10.3 cDNA with an SL1 DNA appended to the R09H10.3 transcripts through a *trans*-splicing reaction, suggests that nucleotides −46 to −1 constitute the true 5′ terminus of R09H10.3.

The gene expression profile for R09H10.3 suggests that it is regulated during development, being more abundant in the larval stage L4 and adults [16], and with a higher expression in adult males compared with adult
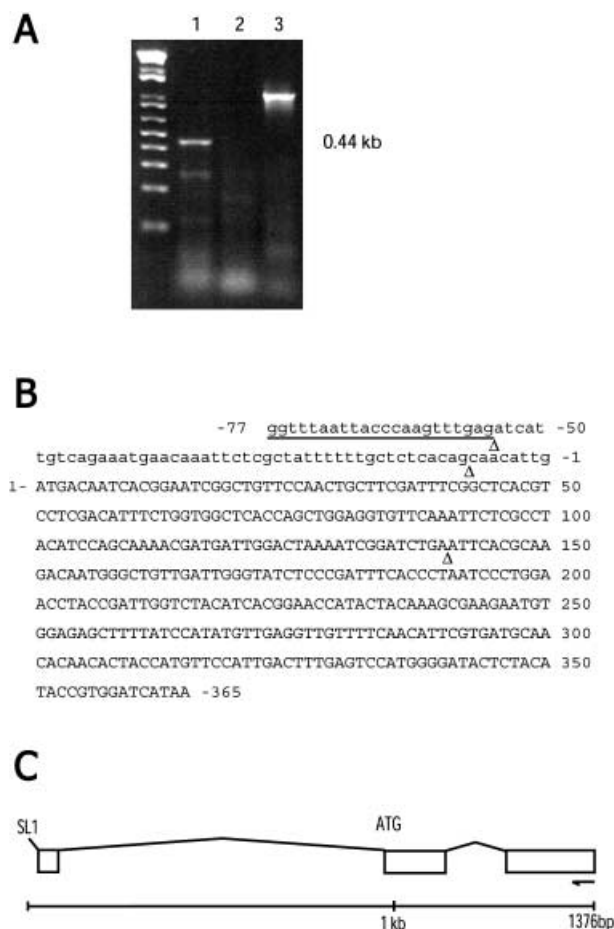

Fig. 2. Detection and characterization of the *C. elegans* TRP transcript. (A) *C. elegans* TRP cDNA was synthesized using RT/PCR and analysed by electrophoresis in a 1.5% agarose gel stained with ethidium bromide. The 440-bp fragment corresponding to R09H10.3 cDNA was consistently amplified (lane 1), whereas no cDNA amplification was observed for the second TRP gene ZK697.8 (lane 2). The robust amplification of cDNA from gene T03D8.1 served as a positive control (lane 3). (B) Sequence of the 440-bp R09H10.3 cDNA fragment with the positions of the intron/exon boundaries indicated (Δ). Capital letters represent the predicted TRP ORF and the SL1 sequence is underlined. (C) The arrangement of exons in the *C. elegans* TRP R09H10.3 gene. Exons are shown as boxes with connecting lines displaying splicing patterns, and transcription proceeds from left to right. The 5′ splice site used for splicing of the SL1 *trans*-spliced leader sequence (0) and the position of the 3′ primer used in RT/PCR and sequencing are indicated. The structure and sequence of the extreme 3′ end of R09H10.3 was not determined.

hermaphrodites [17]. Furthermore, assembled data from several independent DNA micro array experiments have shown that R09H10.3 is coregulated with a group of 803 genes, many of which are known or believed to be expressed specifically in the intestine [39], suggesting that R09H10.3 might be expressed in this tissue.

RNA interference in *C. elegans*

In *C. elegans* injection of dsRNA results in the specific inactivation of genes containing homologous sequences, a

technique termed RNA-mediated interference (RNAi) [40]. RNAi can also be achieved by feeding worms *E. coli* expressing dsRNA corresponding to a specific gene [32]. We have used RNAi through feeding in order to determine the loss-of-function phenotype for R09H10.3 and ZK697.8. Feeding normal wild-type *C. elegans* with bacteria producing dsRNA homologous to R09H10.3 and ZK697.8 resulted in no obvious phenotype when looking for gross phenotypes using a dissecting microscope. However, it is possible that the loss-of-function phenotype is more subtle than could be detected in this study.

Predicted three-dimensional structure

Comparison of the amino acid sequences of aligned TTR-related proteins with the three-dimensional structure of TTR shows that insertions and deletions are situated exclusively at the N- and C-terminal ends, the surface exposed BC-, CD-, DE-, and FG-loops, and the α-helix, while the AB- and GH-loops comprising the dimer–dimer interface in TTR are well conserved both in sequence and in length (Fig. 1). Thus, it is very likely that TRP and TTR share a similar structure.

A homology model of the *E. coli* protein (Fig. 3) based on the X-ray crystallographic structure of human TTR was created using the program ICM [26,27]. The crystal structures from human, rat and chicken TTR have been solved [4,41,42]. Chicken and rat TTR display somewhat higher sequence identity to *E. coli* TRP than the human protein (36.5% and 33.9% compared to 30.4% of the structurally ordered residues), but their structures are very similar to that of human TTR [3]. We chose the human protein as template because it represents the best-characterized TTR structure available and is determined to the highest resolution. The resulting model looks reasonable in that the hydrophobic core is well preserved and the side chains could be fitted without large structural adjustments. The differences at the hormone-binding site are clearly visible, and suggest that the members of the TRP family are designed for ligands different from thyroid hormones. The residues lining the hormone-binding channel in TTR include Met13, Lys15, Leu17, Pro24, Glu54, Thr106, Ala108, Leu/Gln110, Ser/Thr112, Ser115, Ser/Thr117, Thr119, and Val/Ile/Leu121 [3,6]. The corresponding residues are highly conserved within the TRP family, though some are different from TTR; Thr/Ser7, His9, Leu11, Pro18, Arg47, His98, Pro100, Leu/Thr102, Ser104, Ser/Gly107, Ser/Thr109, Tyr111, and Gly113 (numbering according to the mature *E. coli* TRP). The majority of these amino acids are situated at the highly conserved C-terminal end of the TTR-related proteins (residues His98–Ser114). This region shows very low sequence homology with TTR; in particular, the four residue stretch Y-R-G-S at the C
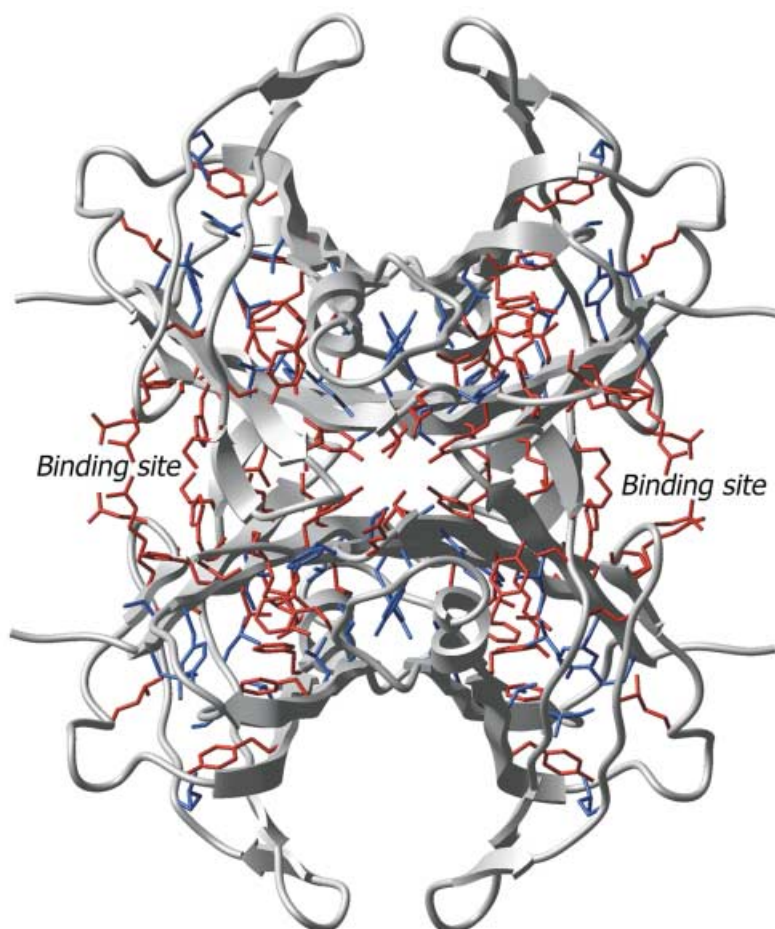

Fig. 3. Visualizing the conservation of the three-dimensional structure. The *E. coli* TRP model based on human TTR (PDB accession code 1F41), with residues displaying more than 80% identity (red) or 80% similarity (blue) within the TRP family drawn as sticks.

terminus that is distinctive to members of the TRP family (Fig. 1).

The hormone-binding channel of TTR provides room for two extended thyroid hormone molecules (Fig. 4). According to the ICM model, the TRP binding pocket is not as deep because it is closed off by the large tyrosine residue at position 111 and a different side chain conformation of Leu110 due to the larger side chains at positions 13 (Gln instead of Ala) and 100 (Pro instead of Ala). The electrostatic surface potential of *E. coli* TRP at the putative binding site is predominantly positive, while the same region in human TTR is distinctly negative. Therefore, it does not seem likely that this protein would bind the same ligand.

Putative function involving uric acid catabolism

In several of the bacterial species the gene encoding TRP is situated in the same region as genes encoding proteins involved in purine catabolism, for example xanthine dehydrogenase, uricase, allantoicase, and ureidoglycolate

hydrolase. However, no such correlation could be found in *E. coli*, *Salmonella* and *Campylobacter jejuni*, which appear to have a periplasmic form of TRP. The TRP in *Bacillus subtilis* (YunM) is expressed as part of an operon including two alleged permeases and a putative uricase, and inactivation of the *yunM* gene results in a uricase-defective phenotype [18]. The putative uricase (YunL) consists of a C-terminal domain homologous to other uricases and a 170-residue N-terminal domain reported to show similarity to alkyl hydroperoxide reductase C (accession code S70169) although the identity is 33% it covers only a 63 amino acid overlap. Interestingly, this domain is 22% identical to the N-terminal domain in TRP from *Arabidopsis thaliana* and these domains seem to belong to a unique protein family showing a range of 20–60% identity, encoded by individual genes in *Streptomyces coelicolor* (T34863), *Bacillus halodurans* (NP_241624), *Pseudomonas aeruginosa* (AAG04905), *Caulobacter crescentus* (NP_421407), *Agrobacterium tumefaciens* (NP_355285), *Sinorhizobium meliloti* (from which two sequences were found, NP_437708 and NP_437328),
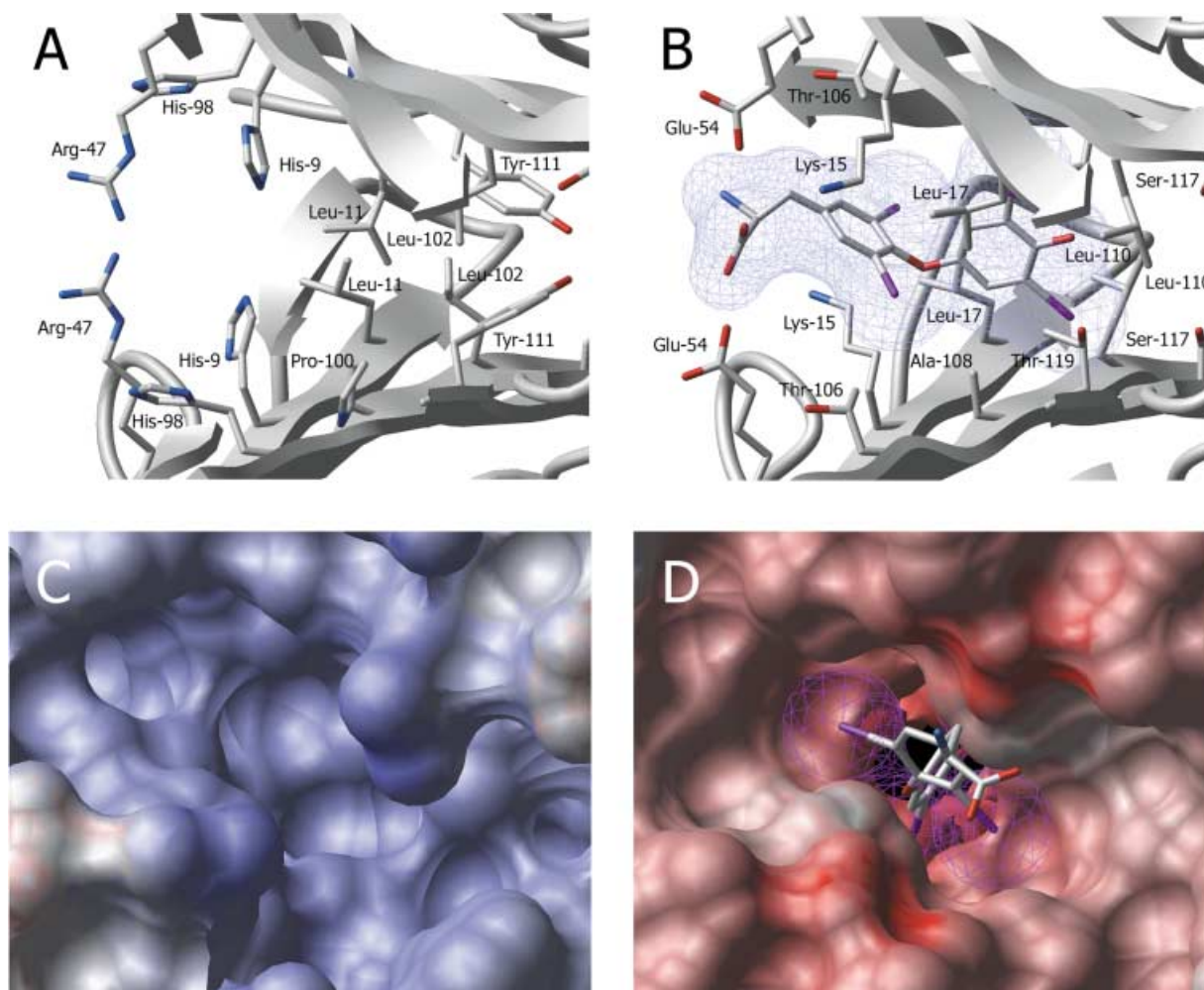

Fig. 4. Homology model of the *E. coli* TRP protein. (A) The ligand-binding site of *E. coli* TRP. (B) The ligand-binding site of human TTR in complex with thyroxine (PDB accession code 2ROX). Noticeable differences in side chains include His9 for Lys15, Arg47 for Glu54, His98 for Thr106, and Tyr111 for Thr119. (C) and (D) show the same as (A) and (B), looking straight through the binding channel with the electrostatic surface potential displayed in blue (positive) and red (negative). The van der Waals' radii of the iodine atoms are outlined in magenta.

Fig. 5. Purification of *E. coli* TRP. (A) Size exclusion chromatography. The purified protein migrates as a single peak showing a tetrameric protein of ≈ 50 kDa. The migration of four proteins in the gel filtration calibration kit (Amersham Pharmacia Biotech) is indicated as diamonds (ribonuclease A, 13.7 kDa; chymotrypsinogen A, 25.0 kDa; ovalbumin, 43.0 kDa; and BSA, 67.0 kDa). (B) SDS/PAGE (20% gel) analysis showing the purity of the protein. Lane 1, molecular mass standards (kDa); lane 2, after SP-sepharose; lane 3, after gel filtration.
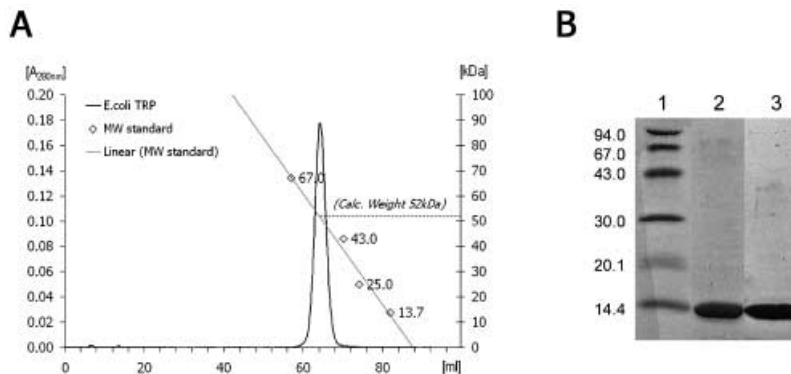
Mesorhizobium loti (NP_105847), and *Mus musculus* (EST sequence BI328404 derived from liver). These sequences show very weak similarities to a number of different proteins hence no relevant relationship to a known protein or function could be determined for this group. In the rhizobia *Sinorhizobium meliloti* (NP_437328) and *Mesorhizobium loti* (NP_105847) this molecule is merged with yet another conserved unknown protein that contains a polysaccharide deacetylase domain (Pfam 01522 [43]). This putative deacetylase is also found in many species, for example *Pseudomonas aeruginosa* (AAG04906), *Caulobacter crescentus* (NP_421406), *Salmonella typhi* (CAD02964), *Salmonella typhimurium* (AAL22006), *Agrobacterium tumefaciens* (NP_355286), *Sinorhizobium meliloti* (NP_437709), and *Schizosaccharomyces pombe* (CAB10114), but unfortunately it is not clear if it exists in *B. subtilis*. None of these uncharacterized proteins are predicted to contain a signal sequence and based on the positioning of these genes in the genomes both appear to be associated with TRP and uricase.

Available ESTs suggest that in the fungi *Phytophora sojae* and *Pichia angusta* TRP mRNA is transcribed in the mycelium. In plants evidence of expression comes mainly from roots, but also from above-ground organs like leaves and flowers. In fish there is proof of expression in the head kidney of *Ictalurus punctatus*, in the embryo of *Danio rerio*, and in the adult liver of *Salmo salar*. Other sources of ESTs include unfertilized eggs from the frog *Xenopus laevis*, foetus cartilage of *Bos taurus*, ovary, spleen and eye of *Rattus norvegicus*, and embryo, liver, pancreas, brain, mammary glands and mandible of *M. musculus*.

Characterization of TRP from *E. coli*

We have cloned, expressed and purified the TTR-related protein from the Gram-negative bacterium *E. coli*. The construct including the signal sequence generates two protein products and MS confirms that one corresponds to the mature protein starting with residue Ala24. The optimized purification scheme is simple and based solely on the high pI of the protein (≈ 8.4) that allows strong binding to SP-sepharose under conditions where most *E. coli* proteins display low affinity for the same gel material. Size exclusion chromatography on a gel filtration column confirms that the *E. coli* TRP forms a tetramer of a similar size to TTR (Fig. 5A). Expression of the protein is high with

typically ≈ 50–60 mg of pure protein per litre of *E. coli* culture (Fig. 5B). We have investigated the eventual amyloidogenic properties of *E. coli* TRP using a protocol based on partial acid denaturation routinely used to induce human TTR amyloid *in vitro* [44]. The *E. coli* TRP does not show any propensity for pH-induced amyloid formation (Fig. 6). It migrates as a monomer on SDS/PAGE at pH intervals ranging from 3.5 to 7.5 (data not shown), while human TTR migrates as a dimer at pH levels above 5.0 if not extensively boiled prior to loading onto the gel. This suggests that the dimer and tetramer assembly is less stable in *E. coli* TRP than in human TTR.

In order to investigate the thyroid hormone binding properties of *E. coli* TRP we performed a dot-blot analysis using radioactively labelled hormones (Fig. 7). Human TTR has 4–10 times higher binding affinity for thyroxine (T_4) than triiodo-thyronine (T_3) (the dissociation constant K_d for thyroxine lies between 3.1×10^{-10} and 1.3×10^{-7}) [1,45]. Fish TTR on the other hand has higher binding affinity for T_3 compared to T_4 [14,46]. As controls we used human and sea bream (*Sparus aurata*) TTR as well as BSA, another thyroid-hormone carrier in plasma [47]. We could confirm the differences in affinity for human and fish transthyretin, but did not observe any binding of T_3 to human TTR (Fig. 7). This was a surprise, and therefore we tested T_4 and T_3 binding to human TTR using the standard
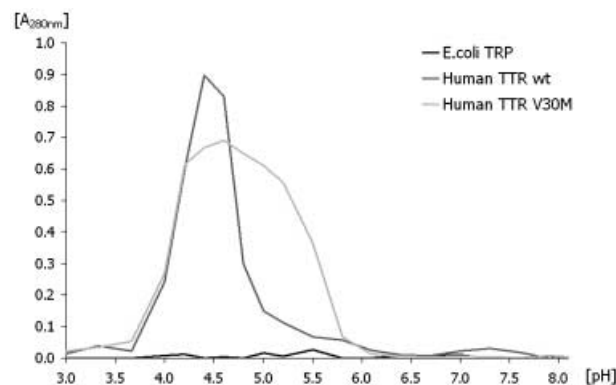

Fig. 6. Aggregation of human TTR and *E. coli* TRP. The level of aggregation was measured at 330 nm after incubation for 72 h at acid denaturing conditions.
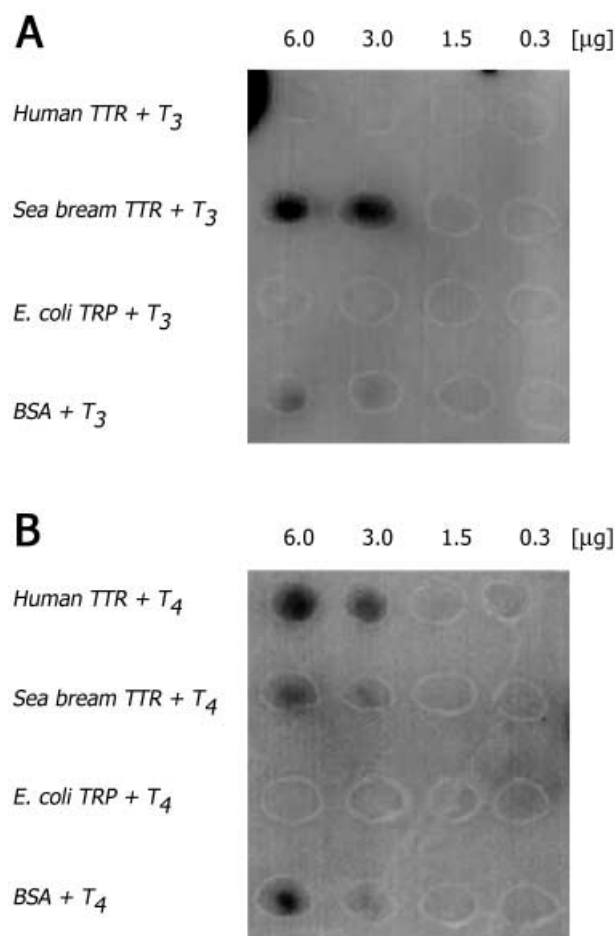
Fig. 7. Dot-blot analysis of T_3 and T_4 binding to human TTR, fish TTR, *E. coli* TRP, and BSA. (A) T_3. (B) T_4. The final amount of protein in each dot is indicated.

Fig. 8. ANS binding to *E. coli* TRP. (Top) Displacement of ANS bound to human TTR by T_4. The addition of TTR causes a shift in the fluorescence emission maximum and an increase in emission intensity. This shift is quenched by the addition of T_4. These results are in agreement with those reported previously [53]. (Bottom) The same experiment shows no apparent binding of ANS to *E. coli* TRP.

method of electrophoresis followed by autoradiography [48] (this method does not work for *E. coli* TRP due to its high pI of 8.4). The analysis confirmed that T_3 does not bind to human recombinant TTR (data not shown). Previous studies describing hormone binding to human TTR were performed on serum protein [45,48]. Our results indicate that human TTR expressed in bacteria displays subtle conformational changes in its hormone-binding channel compared to protein purified from serum. The binding to BSA appears reasonable, it has K_d values of 1.89×10^{-6} and 4.59×10^{-7} for T_3 and T_4, respectively [49,50]. We did not observe any hormone binding to *E. coli* TRP, which confirms our model-based hypothesis that it has no or only very low binding affinity for T_3 and T_4.

ANS is a useful reagent for exploring hydrophobic surfaces on proteins and studying protein interactions with small molecules [51,52]. Quenching of ANS fluorescence by competitive displacement has been used to analyse the binding of TTR to T_3 and T_4 [53]. We performed a similar study on *E. coli* TRP using human TTR as a reference and did not detect any binding of ANS to *E. coli* TRP (Fig. 8). This shows that even though the central binding channels of TTR and TRP might be structurally similar, their shape or
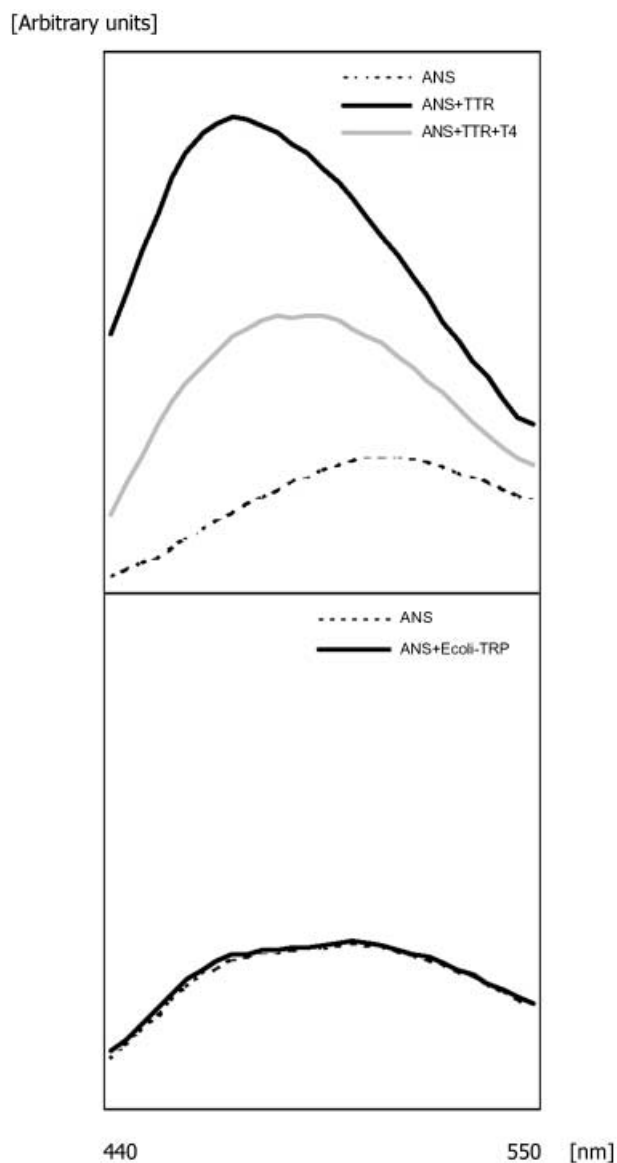
hydrophobic properties are not the same suggestive of different ligand-binding specificities.

Discussion

The TRPs represent a protein family related to TTR but present in a broader range of species, including bacteria, plants and animals. A phylogenetic tree based on both TRP and TTR sequences shows that the TTRs form a separate branch, which most likely originated from a gene duplication event in a prevertebrate species (Fig. 9). It appears that TRP is not only the ancestor of TTR, but has also remained conserved as the TTRs evolved alongside TRP in
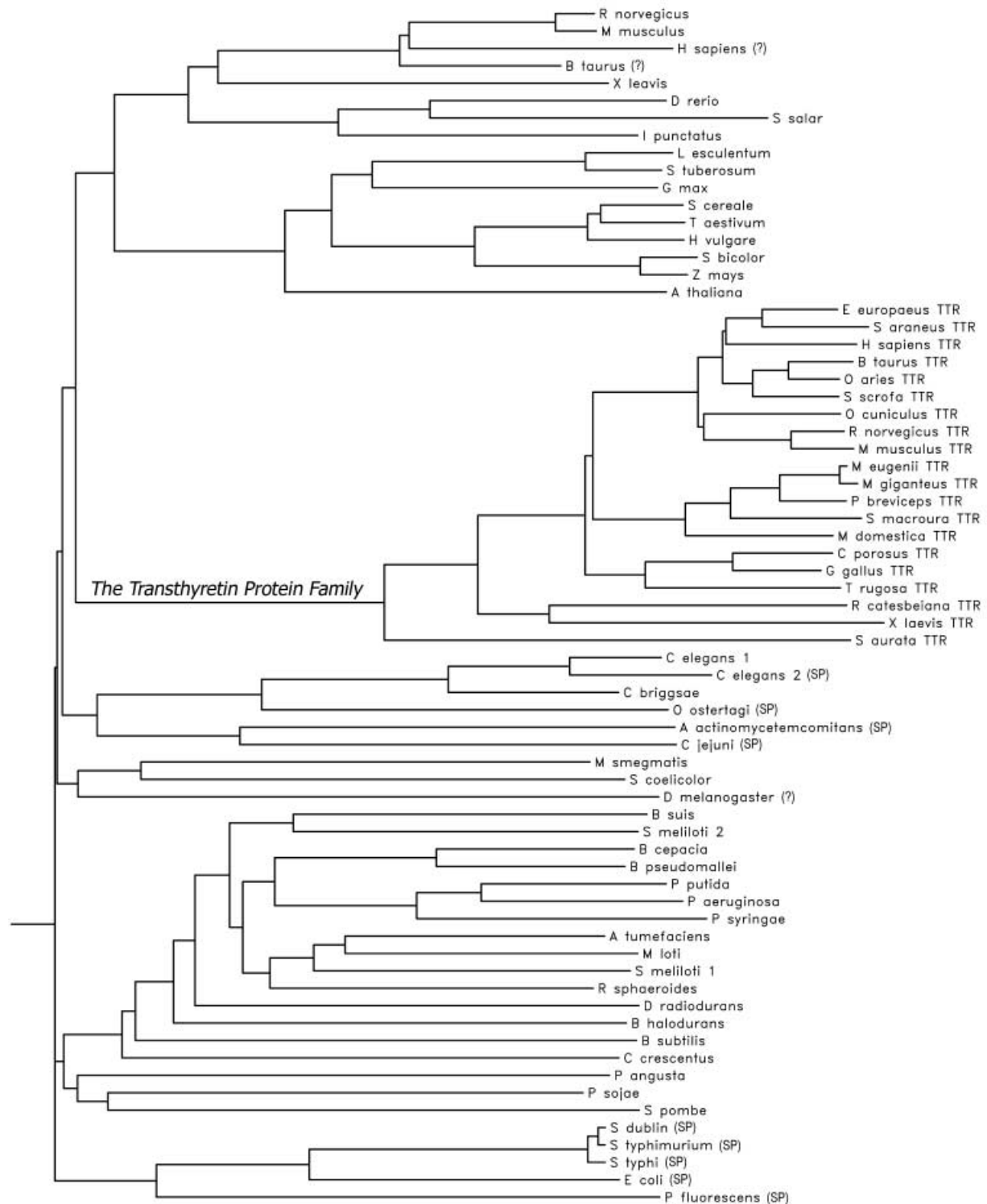
Fig. 9. Phylogenetic tree of TRP and TTR. The tree was based on the multiple sequence alignment comprising 49 TRP sequences and 20 TTR sequences presented in Fig. 1. TRP sequences from species where it is unclear if a functional TRP gene exists and those with predicted signal peptides are marked with (?) and (SP), respectively. The TTR family branch represented by vertebrates is also indicated.

vertebrates. The considerable sequence similarity at the binding site within the TRP family indicates that they perform a unique and important function, which is separate from that of TTRs.

The preliminary characterization of the TRP isolated from *E. coli* agrees well with predictions derived from multiple sequence analysis and the computer-generated homology model. Like TTR the protein forms a stable

homotetramer, while no binding to the thyroid hormones T_3 or T_4 could be detected. Of the 15 residues lining the binding channel as many as 12 are conserved within the TRP family, but only five are similar to TTR. The shape and the hydrophobic properties of the binding channel are clearly different between the two proteins, and the electrostatic potential is predominantly positive in TRP while that of TTR is negative (Fig. 4). Therefore, it is not surprising that *E. coli* TRP does not bind thyroid hormones. It is quite clear that TRP binds a different, and as yet unknown ligand, and considering the outline of TTR as a tetrameric hormone-binding protein it is likely that TTR evolved from a similar transport protein, or possibly an enzyme designed for another small molecule.

The most important information concerning the function of TRP comes from a study of genes involved in purine catabolism in *B. subtilis*, which showed that TRP is essential for uricase activity [18]. Urate oxidase or uricase (E.C. 1.7.3.3) is an enzyme that catalyses the oxidation of uric acid to allantoin by reduction of O_2 to H_2O_2. Uricase homologues are found in a wide range of species [54], however, its metabolic role varies. Interestingly, several of the bacterial TRP genes are situated close to those of uricase homologues in the genome. Expressed TRP sequence tags from plants were found predominantly in roots where urate oxidation is known to occur, and four TRP sequences were identified in symbiotic rhizobia. Despite the fact that ESTs were identified from the liver of both mouse and fish and that the proteins were indicative to be peroxisomal, we did not detect an obvious correlation to urate oxidase activity in vertebrates.

Humans and other primates lack uricase, which remains in the genome as a nonfunctional pseudogene [55]. The human TRP sequence was derived from partly overlapping translated chromosomal DNA from the strand opposite that coding for the growth arrest-specific gene 11 situated at 16q24.3, a region commonly deleted in breast and prostate carcinomas [56]. Since no EST sequences from humans were found, it is not clear if this region contains a functional TTR-related gene. It is noticeable that mouse ESTs available from National Institutes of Health, Mammalian Gene Collection [57] and RIKEN Mouse ESTs [58] show that this transcript is expressed in tumours derived from liver and mammary glands, and that urate has been implied to protect against cancer caused by oxygen radicals [59,60].

In conclusion, available data suggests that the TRP family is related but separate from TTR. We have demonstrated that at least one of the two TRP genes in *C. elegans* is expressed, and that the transcript generated is properly post-transcriptionally modified (SL1 *trans*-spliced). No obvious phenotype was detected when removing TRP gene activity from *C. elegans* by means of RNAi. Our current working hypothesis is that TRP function is associated with urate oxidase activity as shown for *B. subtilis* [18]. However, the ability to bind urate or any other metabolite in the purine catabolism pathway, including any prospective enzymatic properties remains to be experimentally verified. Further investigations of *E. coli* TRP are in progress to determine the functional and biological role of TRP and its structural relationship to TTR.

References

1. Schreiber, G. & Richardson, S.J. (1997) The evolution of gene expression, structure and function of transthyretin. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **116**, 137–160.
2. Power, D.M., Elias, N.P., Richardson, S.J., Mendes, J., Soares, C.M. & Santos, C.R. (2000) Evolution of the thyroid hormone-binding protein, transthyretin. *Gen. Comp. Endocrinol.* **119**, 241–255.
3. Eneqvist, T. & Sauer-Eriksson, A.E. (2001) Structural distribution of mutations associated with familial amyloidotic polyneuropathy in human transthyretin. *Amyloid: Int. J. Exp Clin. Invest.* **8**, 149–168.
4. Blake, C.C., Geisow, M.J., Oatley, S.J., Rérat, B. & Rérat, C. (1978) Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* **121**, 339–356.
5. Hörnberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A.E. (2000) A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J. Mol. Biol.* **302**, 649–669.
6. Wojtczak, A., Cody, V., Luft, J.R. & Pangborn, W. (1996) Structures of human transthyretin complexed with thyroxine at 2.0 Å resolution and 3′,5′-dinitro-*N*-acetyl-L-thyronine at 2.2 Å resolution. *Acta Crystallogr.* **D52**, 758–765.
7. Monaco, H.L., Rizzi, M. & Coda, A. (1995) Structure of a complex of two plasma proteins: transthyretin and retinol-binding protein. *Science* **268**, 1039–1041.
8. Naylor, H.M. & Newcomer, M.E. (1999) The structure of human retinol-binding protein (RBP) with its carrier protein transthyretin reveals an interaction with the carboxy terminus of RBP. *Biochemistry* **38**, 2647–2653.
9. Westermark, P., Sletten, K., Johansson, B. & Cornwell, G.G.D. (1990) Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proc. Natl Acad. Sci. USA* **87**, 2843–2845.
10. Saraiva, M.J. (1995) Transthyretin mutations in health and disease. *Hum. Mutat.* **5**, 191–196.
11. Benson, M.D. & Uemichi, T. (1996) Transthyretin amyloidosis. *Amyloid: Int. J. Exp Clin. Invest.* **3**, 44–56.
12. Connors, L.H., Richardson, A.M., Théberge, R. & Costello, C.E. (2000) Tabulation of transthyretin (TTR) variants as of 1/1/2000. *Amyloid: Int. J. Exp Clin. Invest.* **7**, 54–69.
13. Monaco, H.L. (2000) The transthyretin-retinol-binding protein complex. *Biochim. Biophys. Acta* **1482**, 65–72.
14. Yamauchi, K., Nakajima, J., Hayashi, H. & Hara, A. (1999) Purification and characterization of thyroid-hormone-binding protein from masu salmon serum. A homolog of higher-vertebrate transthyretin. *Eur. J. Biochem.* **265**, 944–949.
15. Prapunpoj, P., Yamauchi, K., Nishiyama, N., Richardson, S.J. & Schreiber, G. (2000) Evolution of structure, ontogeny of gene expression, and function of xenopus laevis transthyretin. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **279**, R2026–R2041.
16. Hill, A.A., Hunter, C.P., Tsung, B.T., Tucker-Kellogg, G. & Brown, E.L. (2000) Genomic analysis of gene expression in *C. elegans*. *Science.* **290**, 809–812.

17. Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. & Kim, S.K. (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **98**, 218–223.

18. Schultz, A.C., Nygaard, P. & Saxild, H.H. (2001) Functional analysis of 14 genes that constitute the purine catabolic pathway in *Bacillus subtilis* and evidence for a novel regulon controlled by the PucR transcription activator. *J. Bacteriol.* **183**, 3293–3302.

19. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.

20. Pearson, W.R. & Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448.

21. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.

22. Nakai, K. & Kanehisa, M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**, 897–911.

23. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

24. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

25. Felsenstein, J. (1989) PHYLIP – Phylogeny Inference Package Version 3.2. Cladistics. 5, 164–166.

26. Abagyan, R.A., Totrov, M.M. & Kuznetsov, D.N. (1994) ICM – a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506.

27. Cardozo, T., Totrov, M. & Abagyan, R. (1995) Homology modeling by the ICM method. *Proteins* **23**, 403–414.

28. Abagyan, R. & Totrov, M. (1994) Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235**, 983–1002.

29. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.

30. Krause, M. & Hirsh, D. (1987) A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* **49**, 753–761.

31. Brenner, S. (1974) The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94.

32. Timmons, L. & Fire, A. (1998) Specific interference by ingested dsRNA. *Nature* **395**, 854.

33. Timmons, L., Court, D.L. & Fire, A. (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112.

34. Kamath, R.S., Martinez-Campos, M., Zipperlen, P., Fraser, A.G. & Ahringer, J. (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**. (RESEARCH), 0002.1–0002.10.

35. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580.

36. Lai, Z., Colón, W. & Kelly, J.W. (1996) The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **35**, 6470–6482.

37. Rudd, K.E. (2000) EcoGene: a genome sequence database for Escherichia coli K-12. *Nucleic Acids Res.* **28**, 60–64.

38. Subramani, S., Koller, A. & Snyder, W.B. (2000) Import of peroxisomal matrix and membrane proteins. *Annu. Rev. Biochem.* **69**, 399–418.

39. Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N. & Davidson, G.S. (2001) A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092.

40. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. & Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811.

41. Sunde, M., Richardson, S.J., Chang, L., Pettersson, T.M., Schreiber, G. & Blake, C.C. (1996) The crystal structure of transthyretin from chicken. *Eur. J. Biochem.* **236**, 491–499.

42. Wojtczak, A. (1997) Crystal structure of rat transthyretin at 2.5 Å resolution: first report on a unique tetrameric structure. *Acta Biochim. Pol.* **44**, 505–517.

43. Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. & Sonnhammer, E.L. (2000) The Pfam protein families database. *Nucl. Acids Res.* **28**, 263–266.

44. Colon, W. & Kelly, J.W. (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31**, 8654–8660.

45. Chang, L., Munro, S.L., Richardson, S.J. & Schreiber, G. (1999) Evolution of thyroid hormone binding by transthyretins in birds and mammals. *Eur. J. Biochem.* **259**, 534–542.

46. Santos, C.R. & Power, D.M. (1999) Identification of transthyretin in fish (*Sparus aurata*): cDNA cloning and characterisation. *Endocrinology* **140**, 2430–2433.

47. Robbins, J. (1976) Thyroxine-binding proteins. *Prog. Clin. Biol. Res.* **5**, 331–355.

48. Richardson, S.J., Bradley, A.J., Duan, W., Wettenhall, R.E., Harms, P.J., Babon, J.J., Southwell, B.R., Nicol, S., Donnellan, S.C. & Schreiber, G. (1994) Evolution of marsupial and other vertebrate thyroxine-binding plasma proteins. *Am. J. Physiol.* **266**, R1359–R1370.

49. Okabe, N., Mano, N. & Tahira, S. (1989) Binding characteristics of a major thyroid hormone metabolite, 3,3′,5′- triiodo-L-thyronine, to bovine serum albumin as measured by fluorescence. *Biochim. Biophys. Acta* **990**, 303–305.

50. Okabe, N. & Hokaze, M. (1992) Thyroxine binding properties of glycosylated bovine serum albumin. *Chem. Pharm. Bull. (Tokyo)* **40**, 3314–3315.

51. Stryer, L. (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.* **13**, 482–495.

52. Edelman, G.M. & McClure, W.O. (1968) Fluorescent probes and the conformation of proteins. *Acta Chem. Res.* **1**, 65–70.

53. Cheng, S.Y., Pages, R.A., Saroff, H.A., Edelhoch, H. & Robbins, J. (1977) Analysis of thyroid hormone binding to human serum prealbumin by 8-anilinonaphthalene-1-sulfonate fluorescence. *Biochemistry* **16**, 3707–3713.

54. Wu, X.W., Lee, C.C., Muzny, D.M. & Caskey, C.T. (1989) Urate oxidase: primary structure and evolutionary implications. *Proc. Natl. Acad. Sci. USA* **86**, 9412–9416.

55. Wu, X.W., Muzny, D.M., Lee, C.C. & Caskey, C.T. (1992) Two independent mutational events in the loss of urate oxidase during hominoid evolution. *J. Mol. Evol.* **34**, 78–84.

56. Whitmore, S.A., Settasatian, C., Crawford, J., Lower, K.M., McCallum, B., Seshadri, R., Cornelisse, C.J., Moerland, E.W., Cleton-Jansen, A.M., Tipping, A.J., Mathew, C.G., Savnio, M., Savoia, A., Verlander, P., Auerbach, A.D., Van Berkel, C., Pronk, J.C., Doggett, N.A. & Callen, D.F. (1998) Characterization and screening for mutations of the growth arrest- specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer. *Genomics* **52**, 325–331.

57. Strausberg, R.L., Feingold, E.A., Klausner, R.D. & Collins, F.S. (1999) The mammalian gene collection. *Science* **286**, 455–457.

58. Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K. & Togawa, Y. (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* **10**, 1757–1771.

59. Ames, B.N., Cathcart, R., Schwiers, E. & Hochstein, P. (1981) Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. *Proc. Natl Acad. Sci. USA* **78**, 6858–6862.

60. Sevanian, A., Davies, K.J. & Hochstein, P. (1991) Serum urate as an antioxidant for ascorbic acid. *Am. J. Clin. Nutr.* **54**, 1129S–1134S.