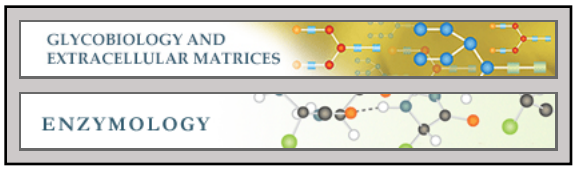
Glycobiology and Extracellular Matrices:

Identification and Biochemical Characterization of Two Functional CMP-Sialic Acid Synthetases in *Danio rerio*



Wiebke Schaper, Joachim Bentrop, Jana Ustinova, Linda Blume, Elina Kats, Joe Tiralongo, Birgit Weinhold, Martin Bastmeyer and Anja-K. Münster-Kühnel

J. Biol. Chem. 2012, 287:13239-13248.
doi: 10.1074/jbc.M111.327544 originally published online February 20, 2012

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 16, pp. 13239–13248, April 13, 2012

Identification and Biochemical Characterization of Two Functional CMP-Sialic Acid Synthetases in *Danio rerio*[*] [S]

Received for publication, December 5, 2011, and in revised form, February 1, 2012 Published, JBC Papers in Press, February 20, 2012, DOI 10.1074/jbc.M111.327544

Wiebke Schaper[‡], Joachim Bentrop[§], Jana Ustinova[§], Linda Blume[‡], Elina Kats[‡], Joe Tiralongo[¶], Birgit Weinhold[‡], Martin Bastmeyer[§], and Anja-K. Münster-Kühnel[‡1]

From the [‡]Institute of Cellular Chemistry, Hannover Medical School (MHH), Carl-Neuberg-Strasse1, D-30625 Hannover, Germany, the [§]Zoologisches Institut I, Zell- und Neurobiologie, Karlsruhe Institute of Technology (KIT), Haid-und-Neu-Strasse 9, D-76131 Karlsruhe, Germany, and the [¶]Institute for Glycomics, Griffith University, Gold Coast Campus, Southport, Queensland 4222, Australia

Background: Addition of sialic acid to the nonreducing end of glycoconjugates requires activation by CMP-sialic acid synthetase (CMAS).
Results: In zebrafish, we identified two CMAS enzymes that differ in expression pattern, activities, and intracellular localization.
Conclusion: Maintenance of two CMAS paralogues is attributed to subfunctionalization.
Significance: Unraveling the individual functions of CMAS paralogues helps to elucidate the impact of sialylation in vertebrate development.

Sialic acids (Sia) form the nonreducing end of the bulk of cell surface-expressed glycoconjugates. They are, therefore, major elements in intercellular communication processes. The addition of Sia to glycoconjugates requires metabolic activation to CMP-Sia, catalyzed by CMP-Sia synthetase (CMAS). This highly conserved enzyme is located in the cell nucleus in all vertebrates investigated to date, but its nuclear function remains elusive. Here, we describe the identification and characterization of two Cmas enzymes in *Danio rerio* (dreCmas), one of which is exclusively localized in the cytosol. We show that the two *cmas* genes most likely originated from the third whole genome duplication, which occurred at the base of teleost radiation. *cmas* paralogues were maintained in fishes of the Otocephala clade, whereas one copy got subsequently lost in Euteleostei (*e.g.* rainbow trout). In zebrafish, the two genes exhibited a distinct spatial expression pattern. The products of these genes (dreCmas1 and dreCmas2) diverged not only with respect to subcellular localization but also in substrate specificity. Nuclear dreCmas1 favored *N*-acetylneuraminic acid, whereas the cytosolic dreCmas2 showed highest affinity for 5-deaminoneuraminic acid. The subcellular localization was confirmed for the endogenous enzymes in fractionated zebrafish lysates. Nuclear entry of dreCmas1 was mediated by a bipartite nuclear localization signal, which seemed irrelevant for other enzymatic functions. With the current demonstration that in zebrafish two subfunctionalized *cmas* paralogues co-exist, we introduce a novel and unique model to detail the roles that CMAS has in the nucleus and in the sialylation pathways of animal cells.

Sialic acids (Sia), a family of nine carbon α-keto acids, are mostly found as terminal sugars on glycoproteins and glycolipids. Due to their exposed position and negative charge, Sia influence numerous cell interaction and cell recognition processes by charge repulsion or by acting as part of recognition structures (1). In mice, interference with Sia biosynthesis has been shown to be lethal before embryonic day 10 (2).

In addition to the mouse model, the zebrafish (*Danio rerio*) has been introduced as a valuable system to study the biological significance of sialylation in vertebrates, particularly of linkage-specific sialyltransferases. Mass spectrometry studies revealed that *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the major Sia derivatives in zebrafish (3, 4). Sia has been found in mono- and oligosialylated structures with up to seven residues bound to glycoproteins and glycolipids (3, 4). In addition, long homopolymeric chains of α-2,8-linked Neu5Ac (polySia)[2] were detected in zebrafish on neural cell adhesion molecules Ncam1a and Ncam1b (5, 6). Removal of polySia results in deficits in posterior commissure formation (5). The formation of polySia during embryonic development is catalyzed by polysialyltransferase ST8Sia2 (7). ST8Sia4, the second polysialyltransferase identified in zebrafish, is generally capable of adding polySia to Ncam1a and Ncam1b but is not involved in polysialylation in the embryo (6, 7). ST8Sia3, the third zebrafish sialyltransferase analyzed to date, catalyzes the formation of oligosialic acid chains. Down-regulation of ST8Sia3 entails anomalous somite morphology, which shows that Sia also play a role in non-neuronal development (8).

A prerequisite for the formation of sialoglycoconjugates is the activation of Sia to its cytidine monophosphate diester

* This work was supported by the German Research Foundation (DFG) Grant MU1849/1 (to A. K. M. K.) and BA 1034/15-19 (to M. B. and J. B.) and by funding from the DFG for the Cluster of Excellence REBIRTH (From Regenerative Biology to Reconstructive Therapy).
[S] This article contains supplemental Tables 1–3 and Fig. 1.
The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) JQ015186, JQ015187, JQ015188, and JQ015189.
[1] To whom correspondence should be addressed. Tel.: 49-511-532-8245; Fax: 49-511-532-8801; E-mail: muenster.anja@mh-hannover.de.

[2] The abbreviations used are: polySia, polysialic acid; CMAS, CMP-Sia synthetase; dreCmas, *D. rerio* Cmas; mmuCMAS, *M. musculus* CMAS; omyCmas, *O. mykiss* Cmas; KDN, 5-deamino-neuraminic acid; NCAM, neural cell adhesion molecule; EST, expressed sequence tag; NLS, nuclear localization signal; EPC, epithelioma papulosum cyprini; contig, group of overlapping clones; WGD, whole genome duplication; AP, alkaline phosphate; hpf, hours post-fertilization; BC, basic cluster.

(CMP-Sia), which is catalyzed by one of the key enzymes of the sialylation pathway, CMP-Sia synthetase (CMAS or CSS, EC 2.7.7.43). Only activated sugars are transported into the Golgi apparatus, where they serve as a substrate for sialyltransferases. Interference with the activation reaction abolishes sialylation on the cell surface (9). The CMAS enzyme is conserved from bacteria to human with five conserved primary sequence motifs forming the active site pocket (10). In contrast to all other sugar-activating enzymes analyzed to date, the vertebrate CMAS is localized in the cell nucleus (11). This unusual intracellular localization was first recognized in the lens epithelial layer by E. L. Kean in 1969 (12) and later confirmed in a variety of biochemical studies using different tissues and species (11, 13). Although the nuclear localization signals have been identified in recombinant mouse and rainbow trout CMAS (14, 15), the biological relevance of this unusual intracellular localization still remains an enigma.

In this study, we identified and cloned two distinct *cmas* genes in zebrafish. We purified the recombinant proteins and demonstrated that both dreCmas enzymes assemble into tetramers and are enzymatically active *in vitro* as well as in cellular systems. Remarkably, however, the dreCmas enzymes showed significant differences not only in terms of substrate specificity but also with respect to their subcellular localization and spatial expression. Although dreCmas1 was transported to the nuclear compartment by a bipartite nuclear localization signal, dreCmas2, in contrast to all other vertebrate CMAS analyzed thus far, remained in the cytoplasm.

EXPERIMENTAL PROCEDURES

DNA and Protein Sequence Analysis—Two *cmas* homologues were identified in the zebrafish genome using known vertebrate *Cmas* sequences in BLAST searches. Sequences of *cmas* homologues of other species (supplemental Table 1) were obtained from ENSEMBL and GenBank™ databases or by using BLASTP or TBLASTN algorithms applied to protein, mRNA, and EST databases. In the latter case, overlapping ESTs were downloaded, aligned, and sorted for each species according to sequence identities. Regions of unsure sequencing were deleted, and consensus sequences were inferred manually.

For phylogenetic analyses, nucleotide sequences were aligned using CLUSTALW implemented in MEGA (16). The maximum likelihood tree was constructed using Mega5 and tested by bootstrap analysis with 1000 replications. The ambiguously aligned N-terminal and C-terminal codons were excluded from analysis, resulting in a 1221-bp-long (407 amino acids) alignment. All sites in triplets were used, and missing data and alignment gaps were deleted in pairwise comparisons. We used the general time-reversible model of substitutions and uniform rates of substitutions among sites. The tree was inferred using the nearest neighbor interchange maximum likelihood heuristic method.

The programs EBI-ClustalW (17) and Bio Edit 7.0.5 (Tom Hall, Ibis Therapeutics, Carlsbad, CA) were used for multiple sequence alignments for visualization of conserved domains and nuclear localization signals. Prediction of putative nuclear localization signal (NLS) sequences was performed by eye and by use of PsortII and PredictNLS. Structure prediction was done using Phyre (18).

Blocks of synteny were determined with the help of the Synteny database (19) or on sight using the latest versions of genome projects provided by ENSEMBL database. In the latter case, chromosomal location and orientation of orthologues of up to 10 genes upstream and downstream of the *cmas* genes of zebrafish were searched for in other species. Only the relative chromosomal position was taken into consideration.

We included species abbreviations in the gene names (*e.g.* dreCmas). For hemichordates and fish, we used the nomenclature recommended for zebrafish (gene, *cmas*; protein, Cmas); for vertebrates in general, we used the mouse nomenclature (gene, *Cmas*; protein, CMAS).

Isolation of cDNA—Total RNA was isolated from pooled 24 and 35 h post-fertilization (hpf) zebrafish embryos using TRIzol (Invitrogen). cDNA was generated using the SuperScript™ II cDNA synthesis kit (Invitrogen). Full open reading frames of dre*cmas1* and dre*cmas2* were amplified using specific primers (supplemental Table 2) and Phusion DNA Polymerase (Finnzymes). PCR products were subcloned into the pCR-Blunt II-TOPO vector (Invitrogen). Their identities were confirmed by sequencing.

Cmas Purification and Size Exclusion Chromatography—We subcloned dre*cmas1* and dr*ecmas2* cDNAs into a modified pET22b-Strep vector allowing the expression of N-terminally Strep II-tagged (IBA) proteins. Primer sequences are provided in supplemental Table 2. Recombinant dreCmas1 and dreCmas2 were expressed in *Escherichia coli* BL21(DE3) (Novagen) at 15 °C in Power Broth (AthenaES) and purified by Strep-Tactin affinity chromatography (IBA). Peak fractions were desalted (HiPrep 26/10, GE Healthcare) and concentrated to 1–2 mg ml^{-1} in buffer containing 50 mM Tris-HCl (pH 8), 20 mM $MgCl_2$, 150 mM NaCl, and 1 mM DTT. Purified protein samples were flash-frozen in liquid nitrogen and stored at −80 °C until required. Protein concentrations were determined using the absorption at 280 nm and the specific extinction coefficient calculated using the ProtParam tool. Size exclusion chromatography was performed as described previously using the above-mentioned buffer (20). Following purification and size exclusion chromatography, Western blots were stained with Strep-Tactin-alkaline phosphate (AP) conjugate (IBA).

In Vitro Activity Assay—Sia activation was analyzed using the EnzChek pyrophosphate assay kit (Invitrogen) and half-area 96-well microplates (Greiner). The reaction was performed essentially as described (20) in 50 mM Tris-HCl, pH 7.5, 25 mM $MgCl_2$ and started by the addition of 0.5 ng/μl dreCmas1 or dreCmas2. CTP, UTP, ATP, or GTP was used at final concentration 1000 μM, and Neu5Ac, Neu5Gc, or KDN was used at 4000 μM.

Construction of Plasmids for Cell Culture Experiments—We subcloned dre*cmas1* and dre*cmas2* cDNAs into two different modified pcDNA3 vectors allowing the expression of N-terminally FLAG-tagged or C-terminally Myc-V5-tagged proteins (supplemental Table 2). Deletion mutants were generated using overlap extension PCR (21) and Phusion Polymerase (Finnzymes) and subsequently cloned in the pcDNA3 vector with C-terminal Myc and V5 tag. Deletion primers were

designed to delete nucleotide triplets encoding selected amino acids (supplemental Table 3).

In Vivo Activity Assay—The functionality of wild-type and mutant dreCmas was analyzed in complementation studies using CHO LEC29.Lec32 cells as described previously (14). Transfected cells were harvested and subdivided into two aliquots, which were incubated at 37 °C for 30 min in either the absence or the presence of 100 ng of endosialidase E to remove polySia. Equal protein amounts were separated by 7 and 10% SDS-PAGE, respectively. PolySia and dreCmas expression was analyzed by Western blotting using anti-polySia mAb 735 (5 μg/ml) (22), anti-V5 mAb (Sigma), and goat anti-mouse 680 IRDye secondary antibody (LI-COR). Scanning was performed with the LI-COR Odyssey infrared imaging system.

Transfection of NIH-3T3 and EPC Cells and Indirect Immunofluorescence Analysis—Epithelioma papulosum cyprini (EPC) cells (ATCC CRL-2872; a *Pimephales promelas*-derived cell line) were cultured in minimal essential medium with Earle's salts and L-Gln (PAA) and 10% fetal calf serum at 5% CO_2 and 26 °C. 24 h before transfection, 3.5×10^5 EPC cells were cultured on glass coverslips in a 12-well plate. Cells were transfected with 0.5 μg of DNA using 3 μl of Fugene 6 (Roche Diagnostics). The next day, cells were fixed in 80% acetone at −20 °C for 10 min (23). Maintenance and transfection of NIH-3T3 cells and immunofluorescence analysis were performed essentially as described (14), except anti-FLAG M5 (4 μg/ml, Sigma), anti-Myc 9E10 (1.3 μg/ml), and sheep anti-mouse IgG Cy3 (3 μg/ml, Sigma) were used for indirect immunofluorescence staining.

Animal Care—Wild-type and golden zebrafish strains were maintained and crossed according to standard procedures. Developmental stages are indicated in hpf according to Kimmel *et al.* (24).

Nuclear and Cytoplasmic Extracts—Zebrafish embryos were dechorionated and deyolked according to Link *et al.* (25) and flash-frozen in N_2. 100 embryos were used to prepare nuclear and cytosolic extracts according to Dignam *et al.* (26). The extracts were analyzed by Western blotting using antibodies that were originally derived against murine CMAS and that specifically recognize the distinct dreCmas enzymes.[3]

Whole Mount in Situ Hybridization—Whole mount *in situ* hybridization of digoxigenin-labeled RNA probes was carried out according to standard protocols (27, 28). To generate specific probes, the 3′ regions of dre*cmas1* and dre*cmas2* were amplified by RT-PCR with specific primers (supplemental Table 2) and cloned into pCR-Blunt II-TOPO (Invitrogen). Antisense and sense (control) riboprobes were synthesized from linearized plasmids using the digoxigenin RNA labeling kit (Roche Applied Science). Pictures were taken using a Zeiss SteREO Lumar.V12 equipped with an AxioCam HRc camera. Images were edited with Photoshop 6.0 (Adobe).

RESULTS

Identification of Two cmas Genes in the D. rerio Genome—Analyses of the completely sequenced genome of *D. rerio* revealed two homologues of the *cmas* gene. The first is located on chromosome 4 (dre*cmas1*). According to the current version of the *D. rerio* genome assemblage, the second gene (dre*cmas2*) is split into two fragments situated on different strands but in direct proximity to each other on chromosome 25. These fragments covered a large portion of the *cmas* gene including the N-terminal and C-terminal portions, but with a gap of about 135 bp. As both fragments were found on different contigs, we hypothesized that they may represent an intact but incorrectly assembled gene. Indeed, we were able to amplify cDNAs of both dre*cmas* genes by RT-PCR using gene-specific primers. The amplified dre*cmas1* cDNA encoded a protein of 430 amino acids with a calculated molecular mass of 48.2 kDa and was identical to that previously reported (accession number CAK18993). The amplified dre*cmas2* contained an open reading frame (ORF) that encoded a protein of 423 amino acids (calculated molecular mass of 47.7 kDa). At the primary sequence level, both proteins shared 56% identity. An alignment of *D. rerio* Cmas amino acid sequences with those of *Mus musculus* (mmuCMAS) and *Oncorhynchus mykiss* (omyCmas, rainbow trout) is presented in Fig. 1*A* and highlights identical residues in *black* and conserved residues in *gray*. Like all other known CMAS enzymes, both dreCmas sequences harbored five conserved primary sequence motifs (motifs I to V) in the N-terminal domain (Fig. 1*A*). In line with other vertebrate CMAS enzymes, dreCmas1 and dreCmas2 possessed an additional C-terminal domain (Fig. 1*A*) composed of 173 and 175 amino acids, respectively. When compared with the N-terminal domains, the C-terminal domains showed less homology to the primary sequences of mmuCMAS and omyCmas. Protein structure prediction revealed homology to phosphatases of the haloacid dehalogenase family. However, phosphatase activity was not observed for either of the dreCmas enzymes *in vitro* (data not shown).

Next, we addressed the question whether other fish species also possess two paralogues of the *cmas* gene. Analyses of the latest fish genome assemblages as well as of the National Center for Biotechnology Information (NCBI) fish EST database did not give evidence for two different intact *cmas* genes in any of the species belonging to Euteleostei (as defined by Li *et al.* (29)). In contrast, consistent with zebrafish, two sequence variants were found among ESTs of species belonging to the Otocephala clade, namely catfish (*Ictalurus furcatus*, *Ictalurus punctatus*) and the fathead minnow (*P. promelas).* The maximum likelihood phylogenetic tree demonstrates that fish *cmas* genes segregated into two distinct clusters comprising *cmas1* and *cmas2*, respectively (Fig. 1*B*). Both clades have a monophyletic origin and putatively originate from the ancient whole genome duplication (WGD) that occurred in Teleostei.

To further specify the phylogenetic relationship between *Cmas* genes, we analyzed the syntenic organization of surrounding genomic regions. Within vertebrates, we found traces of conserved synteny surrounding the *Cmas1* genes between zebrafish chromosome 4 and chicken chromosome 1 despite numerous translocations and inversions that presumably occurred in both lineages (Fig. 2*A*, *left*). A significant intragenomic synteny between zebrafish chromosomes 4 and 25 corroborated that both *cmas* genes were paralogues resulting from

[3] M. Sellmeier, W. Schaper, B. Weinhold, R. Gerardy-Schahn, and A.-K. Münster-Kühnel, manuscript in preparation.

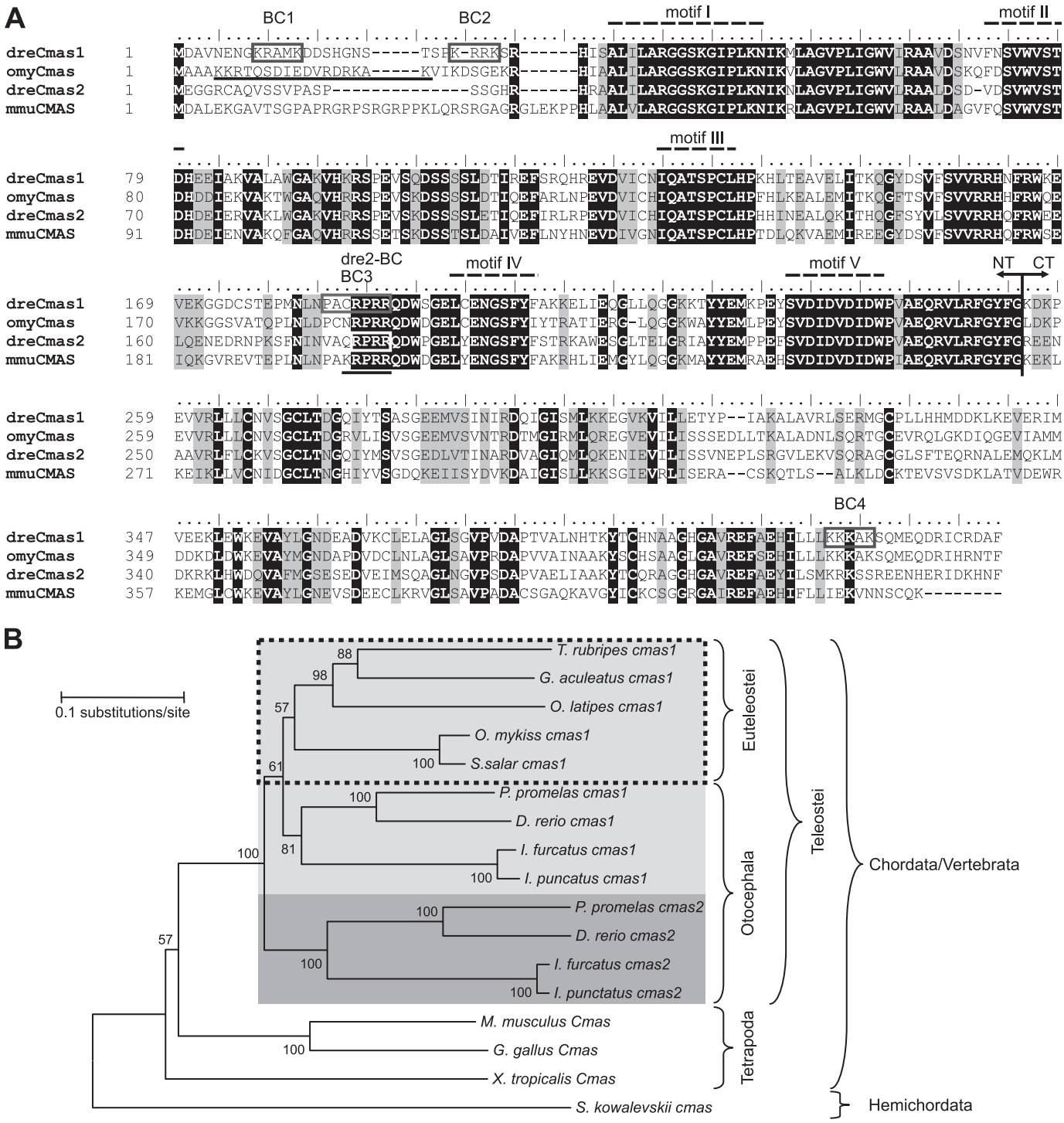


FIGURE 1. **Relations between selected vertebrate CMAS sequences.** *A*, multisequence alignment. Amino acid sequences of the two *D. rerio* Cmas proteins (dreCmas1 and dreCmas2) have been aligned with *O. mykiss* (omyCmas) and *M. musculus* CMAS (mmuCMAS) using ClustalW to display maximum homology. Strictly and highly conserved residues are shaded in *black* and *gray*, respectively. The five conserved CMAS motifs essential for enzymatic activity are marked with *dashed lines*. The nuclear localization signals of mmuCMAS and omyCmas are *underlined* (14, 15). Basic clusters identified in dreCmas1 (BC1 to BC4) and dreCmas2 (dre2-BC) are boxed in *gray* and *white*, respectively. The boundary between the N- and C-terminal domains (*NT* and *CT*) is marked by *arrows*. *B*, maximum likelihood phylogenetic tree showing evolutionary relationships among *Cmas* genes. *cmas* homologues in other fish species were identified either in the latest fish genome assemblages with dre*cmas2* as a query or in the NCBI fish EST database. In the latter, each species with multiple *cmas*-like ESTs (10 in *Salmo salar*, 16 in *Gasterosteus aculeatus*, 21 in *Oryzias latipes*) was checked for the presence of more than one sequence variant that could not be explained by allelic polymorphism and alternative splicing. Fish *cmas2* homologues are shaded in *dark gray*, and fish *cmas1* genes are in *light gray*. Euteleost species that have a single *cmas* gene are framed (*dashed rectangle*). The branch lengths correspond to evolutionary distance indicated by the scale. *Numbers* at the bifurcations are bootstrap values (1000 replicates). The tree is based on 1221 bp long nucleotide alignment. Ambiguously aligned C- and N-terminal codons were excluded from the analyses. We used the general time-reversible model of substitutions, uniform rates of substitutions among sites, and all positions in codons. Genera are abbreviated by their initials. *T. rubripes*, *Takifugu rubripes*; *G. gallus*, *Gallus gallus*; *S. kowalevskii*, *Saccoglossus kowalevskii*; *X. tropicalis*, *Xenopus tropicalis*.

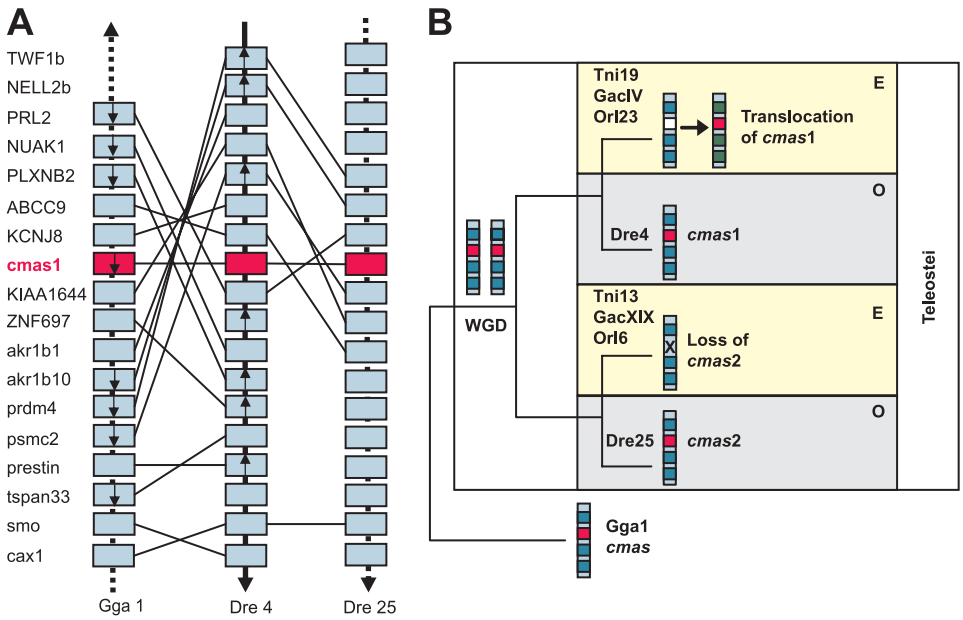


FIGURE 2. **Conserved synteny elucidates evolution of *cmas* genes in fishes.** *A*, synteny among chicken (*G. gallus, Gga*) chromosome 1, zebrafish (*D. rerio, Dre*) chromosome 4, and zebrafish chromosome 25. Each box represents a gene. *Red boxes* show *cmas* homologues. Gene names on the *left side* correspond to the *center column* showing zebrafish chromosome 4 (*Dre 4*). The order of genes on the Dre4 is real and is centered at *cmas1*. The order of genes on chicken chromosome 1 (*Gga 1*) and zebrafish chromosome 25 (*Dre 25*) is relative (other genes, not shown here, can reside between two boxes). Direction of arrows connecting genes is 5′ to 3′. *Boxes with arrows* indicate genes on the reverse strand of the chromosome. *Thin solid lines* connect homologous genes on different chromosomes. *B*, the proposed scenario of evolution of *cmas* genes in Teleostei. The ancestral chromosome harboring *cmas* gene (which was similar to the chicken Gga1 and zebrafish Dre4) was duplicated during the WGD in ray-finned fish. Paralogues of *cmas* reside on chromosomes Dre4 and Dre25 in zebrafish. Apparently, *cmas1* was translocated to another site from the euteleost ancestral chromosome, homologous to Dre4, and *cmas2* was lost from the euteleost ancestral chromosome, homologous to the Dre25 (compare supplemental Fig. 1). *Colored strips* represent chromosomes, and similar colors indicate homology. *O*, Otocephala; *E*, Euteleostei.

the WGD (Fig. 2*A*, *right*). In Euteleostei, *cmas1* was excised from a homologue of the *D. rerio* chromosome 4 and transferred to a foreign location, whereas *cmas2* was excised from the homologue of *D. rerio* chromosome 25 and lost (probably during translocation) (Fig. 2*B* and supplemental Fig. 1). In summary, the two *cmas* genes most likely originated from the third WGD at the base of teleost radiation. During evolution, one *cmas* copy got lost in Euteleostei (*e.g.* rainbow trout), whereas both paralogues were maintained in fishes of the Otocephala clade (*e.g.* zebrafish).

drecmas1 and drecmas2 mRNAs Show Different Spatial Expression in Zebrafish Embryos—The expression pattern of dre*cmas1* and dre*cmas2* was analyzed in the developing zebrafish by *in situ* hybridizations using specific probes derived from the 3′-UTR of both mRNAs. In general, dre*cmas1* showed a stronger and more distinct expression than dre*cmas2*. At 90% epiboly (9 hpf), besides a basal, more or less ubiquitous expression, dre*cmas1* was detected in the axial mesoderm, especially in the notochord primordium (Fig. 3*A*). At 18 hpf, dre*cmas1* showed a robust expression in the entire central nervous system, the somites, the notochord, and the developing pronephric duct (Fig. 3*C*). With progressing development, expression of dre*cmas1* was down-regulated in the trunk. It persisted in the central nervous system and was up-regulated in the kidney and the liver primordium (Fig. 3, *H* and *I*). dre*cmas2* was expressed at lower levels and in less sharply defined regions. It was detected more or less ubiquitously from the end of gastrula through segmentation (Fig. 3, *B* and *D*). During the pharyngula stage, dre*cmas2* expression was restricted to the brain (Fig. 3*G*), and it was down-regulated around hatching (Fig. 3*K*). Although dre*cmas2* showed expression in the heart at 48 hpf (Fig. 3*I*), we could not detect it in skeletal muscle, liver, or kidney.

dreCmas Proteins Exhibit Different Substrate Specificities and Assemble into Tetramers—To analyze whether both dre*cmas* genes encode active enzymes, the proteins were expressed with an N-terminal Strep II tag in *E. coli* BL21(DE3) and purified to homogeneity via Strep-Tactin affinity and subsequent anion-exchange chromatography. Lysate, flow-through, and the purified proteins were analyzed by Coomassie Brilliant Blue Staining and Western blot analysis (Fig. 4*A*). Enzymatic activity was investigated *in vitro* by using the EnzChek pyrophosphate assay kit. Equal protein concentrations were used. In addition to CTP and Neu5Ac, the nucleotide donors UTP, ATP, and GTP, as well as the Sia derivatives Neu5Gc and KDN, were tested as alternative substrates. All substrates were used in non-limiting concentrations. The specific enzymatic activities clearly demonstrated that both enzymes were strictly dependent on CTP (data not shown) but differed in terms of the preferred Sia derivative (Table 1). dreCmas1 had highest activity toward Neu5Ac, lower activity toward Neu5Gc, and just basal activity toward KDN. In contrast, dreCmas2 preferentially activated KDN and showed only basal activity toward Neu5Ac and Neu5Gc.

To determine the oligomeric state of dreCmas1 and dreCmas2, size exclusion chromatography was performed with the purified recombinant enzymes (Fig. 4, *B* and *D*). dreCmas1 eluted as a single peak at 11.87 ml corresponding to an apparent molecular mass of 236.1 kDa as calculated from the log molecular mass *versus* retention volume plot (Fig. 4*B*, *inset*). The apparent molecular mass to theoretical mass (50.2 kDa) ratio of 4.7 indi-

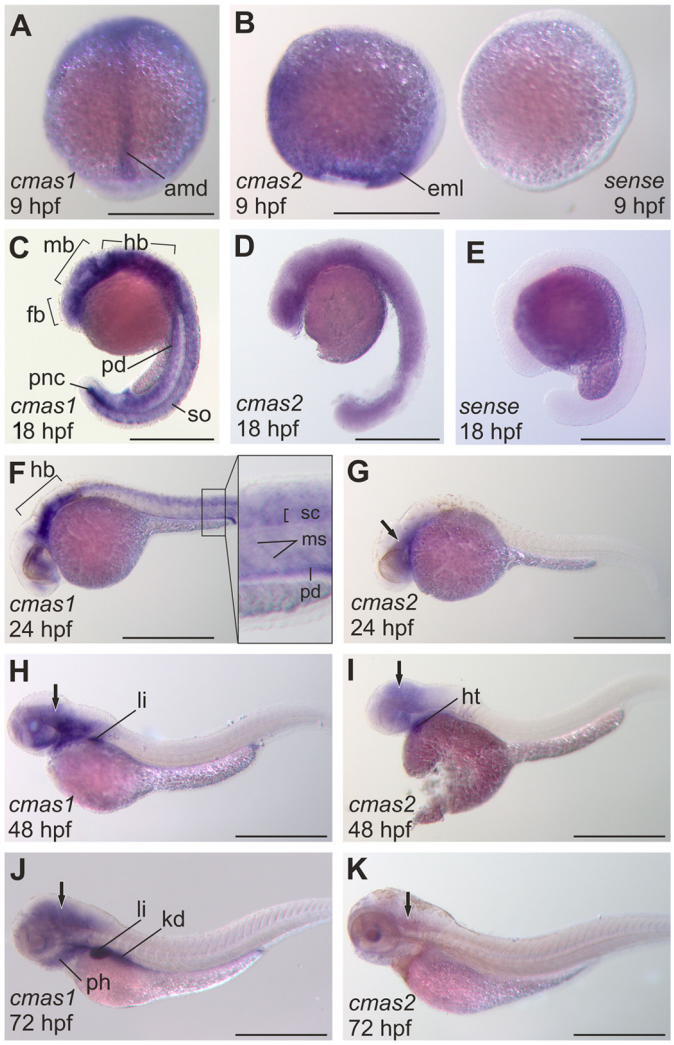


FIGURE 3. **Expression of dre*cmas1* and dre*cmas2* mRNAs in zebrafish embryos.** Embryos were hybridized *in situ* with digoxigenin-labeled riboprobes. *A*, *C*, *F*, *H*, and *J*, dre*cmas1*; *B*, *D*, *G*, *I*, and *K*, dre*cmas2*; *B* and *E*, sense controls. The *inset* in *F* shows a higher magnification of the indicated region. Ages are indicated in hpf. *A* and *B*, dorsal views. *D–K*, lateral views. dre*cmas1* is expressed in the posterior axial mesoderm at 9 hpf (*A*), throughout the central nervous system at all analyzed time points (*C*, *F*, *H*, and *J*), in the somites during segmentation (*C*), in the pronephric kidney starting around 18 hpf (*C*, *F*, *H*, and *J*), and in the liver (*H* and *J*). dre*cmas2* is weakly but ubiquitously expressed in early developmental stages (up to 18 hpf), and it is restricted to the brain regions from 24 hpf onwards (*G*, *I*, *J*, and *K*). *amd*, axial mesoderm; *eml*, endomesodermal layer; *fb*, forebrain; *hb*, hindbrain; *ht*, heart; *kd*, kidney; *li*, liver; *mb*, midbrain; *ms*, myoseptum; *pd*, pronephric duct; *ph*, pharynx; *pnc*, posterior notochord; *sc*, spinal cord; *so*, somites. The *arrows* indicate a more or less uniform expression in the brain. *Scale bars* represent 500 μm.

cated that the recombinant Strep II-tagged dreCmas1 formed a tetramer or a pentamer. The same result was obtained in repeated experiments with protein lacking the epitope tag. Because it is known from crystal structure analyses of *Neisseria meningitidis* CMAS (30) and mmuCMAS-NT (31) that the active unit (N-terminal domain) is formed by a dimer, it is reasonable to conclude that dreCmas1 assembled into a tetramer. The recombinant dreCmas2 eluted as a single peak at 12.29 ml corresponding to an apparent molecular mass of 193.9 kDa (Fig. 4*D*). A ratio of 3.9 from the apparent molecular mass to the theoretical mass (49.6 kDa) clearly indicated that the recombinant Strep II-tagged dreCmas2 also formed a tetramer. The integrity of the proteins was confirmed by SDS-PAGE followed by Coomassie Blue staining as well as Western blot analysis of the fractions (Fig. 4, *C* and *E*).

The Two Zebrafish Cmas Enzymes Localize to Different Cellular Compartments—To investigate the intracellular localization of the two zebrafish Cmas enzymes, the cDNAs were expressed in EPC cells, derived from the cyprinid *P. promelas*. To minimize the influence of the epitope tag, N-terminally FLAG-tagged as well as C-terminally Myc-V5-tagged proteins were analyzed by indirect immunofluorescence analysis. Regardless of the position of the epitope tag, dreCmas1 was localized in the nuclear compartment (Fig. 5*A*). Intriguingly, and in contrast to all other vertebrate CMAS proteins, dreCmas2 was found in the cytoplasm of transfected EPC cells. These results were confirmed in a mouse fibroblast cell line (NIH-3T3 cells; data not shown). To investigate the intracellular destination of the endogenous enzymes, nuclear and cytoplasmic extracts were prepared from 48 hpf zebrafish embryos and analyzed by Western blotting (Fig. 5*B*). The specificity of the antibodies for either of the dreCmas enzymes was confirmed with purified recombinant proteins (Fig. 5*C*). In agreement with the results obtained in EPC cells, endogenous dreCmas proteins were found either in the nuclear (dreCmas1) or in the cytosolic fraction (dreCmas2) (Fig. 5*B*).

A Bipartite Nuclear Localization Signal Targets dreCmas1 to Nuclear Compartment—To identify the NLS essential for nuclear targeting of dreCmas1, primary sequence analysis was performed by eye and by the use of the programs PSORT and PredictNLS. Four basic clusters (BC1 to BC4) were identified in the dreCmas1 sequence (Fig. 1*A*). BC1 ($K^{9}RAMK^{13}$) and BC2 ($K^{24}RRK^{27}$) were located at the N terminus, BC3 ($P^{184}ACRPRR^{190}$) was located at the center, and BC4 ($K^{412}KKAK^{416}$) was located at the C terminus. BC4 and the surrounding 12 amino acid residues were strictly conserved in rainbow trout Cmas (Fig. 1*A*). Because they do not serve as NLS (15), we concentrated on the analysis of BC1 to BC3. All BCs as well as BC1 and BC2 in combination were deleted by site-directed mutagenesis in C-terminally V5-Myc-tagged dreCmas1. Deletion mutants were expressed in EPC cells (Fig. 6) and NIH-3T3 cells (data not shown). The subcellular localization was analyzed by indirect immunofluorescence microscopy. Although the deletion of BC3 did not impair nuclear import of dreCmas1, deletion of BC1 or BC2, individually or in combination, entailed retention in the cytoplasm. Thus, both BC1 and BC2 were essential for nuclear import of dreCmas1 and formed a bipartite NLS ($K^{9}RAMK^{13}(X)_{11}K^{24}RRK^{27}$) according to the consensus sequence $(K/R)_2X_{10-12}(KR)_3$ (32). dreCmas2 in contrast contained a single short BC ($R^{179}PRR^{182}$) (dre2-BC) at the center of the protein (Fig. 1*A*), which did not fit to the consensus sequence of a monopartite NLS $(K/R)_{4-6}$ (32). The absence of an NLS was in perfect agreement with the cytosolic localization of dreCmas2 (Fig. 5).

Nuclear Sequestration Is Not Required for Enzymatic Activity of dreCmas1—To analyze the enzymatic activity of dreCmas in a cellular system and to determine the importance of the BCs for activity, full-length dreCmas as well as deletion mutants were analyzed in a complementation approach using the

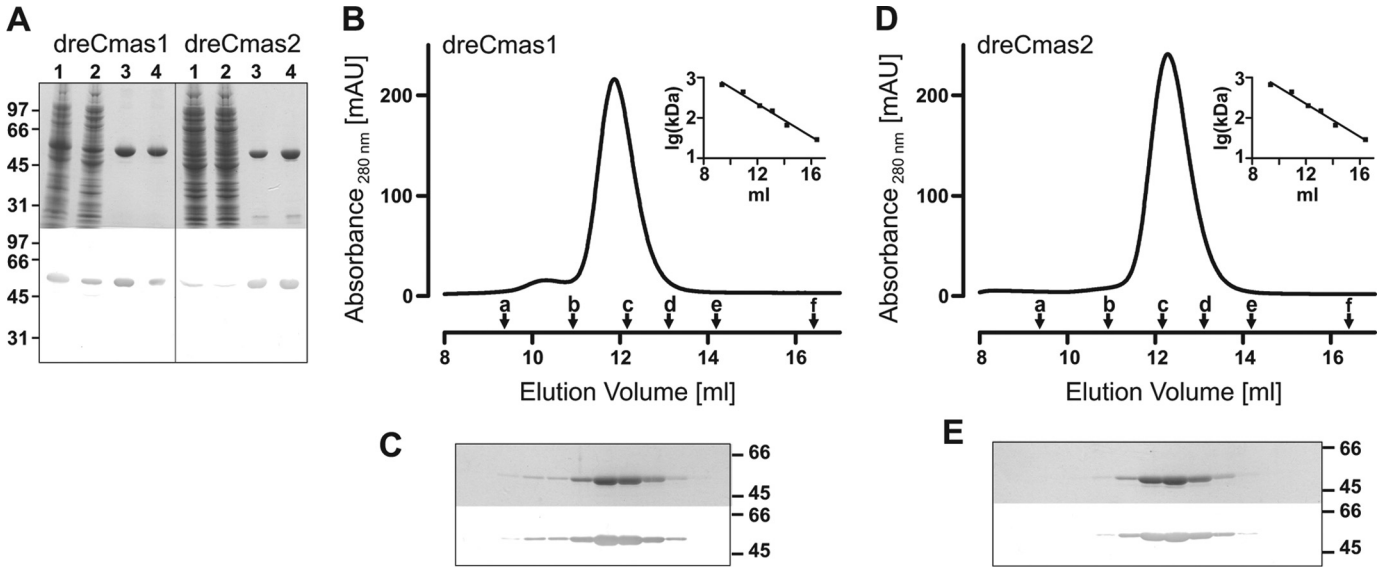


FIGURE 4. **Purification and size exclusion chromatography of recombinant dreCmas1 and dreCmas2.** *A*, N-terminally Strep II-tagged dreCmas1 and dreCmas2 were expressed in *E. coli* and purified by Strep-Tactin affinity chromatography. Purification was followed by SDS-PAGE with Coomassie Brilliant Blue (*upper panel*) and Strep-Tactin-AP Western blot staining (*lower panel*). *Lane 1*, bacterial lysate; *lane 2*, flow-through of the Strep-Tactin column; *lane 3*, eluate of the Strep-Tactin column; *lane 4*, flow-through of anion-exchange column. *B–E*, determination of quaternary organization was performed by size exclusion chromatography of Strep II-tagged dreCmas1 (*B*) and dreCmas2 (*D*). Protein standards are indicated by *arrows* (*a–f*: 669, 443, 200, 150, 66, and 29 kDa). The apparent molecular masses of the two enzymes were determined by standard curves (*inset*). Elution was traced by 12% SDS-PAGE followed by Coomassie Brilliant Blue Staining and Strep-Tactin-AP Western blotting (*C* and *E*) of fractions. *mAU*, milliabsorbance units.

TABLE 1

Substrate specificity of zebrafish Cmas enzymes

Specific enzymatic activities (μmol(substrate) mg^{-1}(enzyme) min^{-1}) of purified recombinant StrepII-dreCmas1 and StrepII-dreCmas2 were determined in the presence of different sugar substrates (4 mM). CTP was used in a non-limiting concentration (1 mM). Values are given as means ± S.D. from four independent experiments.

	dreCmas1	dreCmas2
Neu5Ac	4.5 ± 0.3	0.4 ± 0.1
Neu5Gc	2.7 ± 0.1	0.2 ± 0.1
KDN	0.3 ± 0.1	3.3 ± 0.4

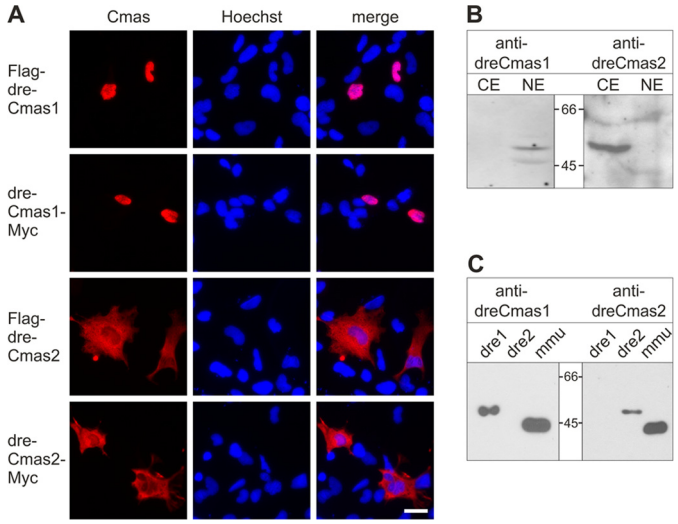


FIGURE 5. **Intracellular localization of dreCmas1 and dreCmas2.** *A*, indirect immunofluorescence analysis of dreCmas1 and dreCmas2 transiently expressed in EPC cells. N-terminally FLAG- or C-terminally Myc-V5 tagged proteins (as indicated) were detected with anti-FLAG or anti-Myc mAb and visualized with a Cy3-conjugated secondary antibody (*Cmas panels*). Nuclei were stained with Hoechst 33258. Merged pictures are shown in the *right panels*. *Scale bar*: 20 μm. *B*, Western blot analysis of nuclear (*NE*) and cytosolic (*CE*) extracts generated from zebrafish. Proteins were separated by 12% SDS-PAGE and analyzed by Western blotting with CMAS specific antibodies. Although dreCmas1 was visualized in the nuclear fraction of zebrafish extracts, dreCmas2 was detected in the cytosolic fraction. *C*, specificity of the antibodies was controlled with purified epitope-tagged recombinant proteins. The antibodies specifically detect either dreCmas1 (*dre1*) or dreCmas2 (*dre2*). Both antibodies recognize the murine CMAS (*mmu*).

CMAS-negative Chinese hamster ovary (CHO) cell line LEC29.Lec32 (9). The *lec32* mutation in LEC29.Lec32 causes the expression of asialoglycoconjugates at the cell surface, a defect that can be complemented by recombinant expression of an active Cmas. As shown in Fig. 7 (*upper panel*), reconstitution of the defect by both dreCmas proteins led to reappearance of polySia on the cell surface, which is visible as a smear due to microheterogeneity of the polySia chain length. Moreover, specificity of polySia expression was controlled by use of endosialidase E leading to the disappearance of the polySia signal (Fig. 7). In all experiments, dreCmas expression was controlled by Western blot staining (Fig. 7, *Cmas lane*). dreCmas1 was able to complement the defect in LEC29.Lec32 cells, and neither the deletion of BC1 or BC2 nor deletion of the bipartite NLS (ΔBC1+2) in dreCmas1 altered enzymatic activity in the cellular system. Also, expression of dreCmas2, which preferentially activates KDN and showed only residual activity with Neu5Ac *in vitro*, produced sufficient amounts of CMP-Neu5Ac to form polySia, a homopolymeric Neu5Ac chain, in the cellular system. In contrast, deletion of the central basic cluster in dreCmas1 (ΔBC3) as well as deletion of the corresponding residues in dreCmas2 (ΔBC) completely abolished enzymatic activity. These data show that dreCmas1 and dreCmas2 were not only enzymatically active *in vitro* but also in a cellular system. Enzymatic activity was associated with conserved basic residues in the center of both dreCmas proteins, but not with the bipartite NLS in dreCmas1. Thus, nuclear import is not a prerequisite for enzymatic activity of dreCmas1.

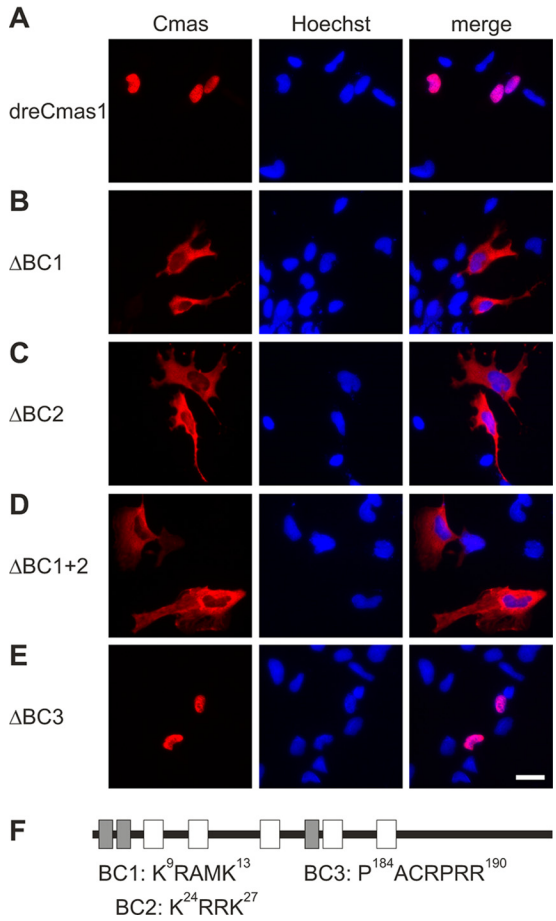


FIGURE 6. **Intracellular localization of dreCmas1 mutants in EPC cells.** *A–E*, indirect immunofluorescence analysis of C-terminally Myc-V5 tagged dreCmas1 (*A*) and deletion mutants (*B–E*) lacking individual basic clusters (see *panel F* and Fig. 1*A*) after transient expression in EPC cells. Staining was performed with anti-Myc mAb and a Cy3-conjugated secondary antibody (*Cmas panels*). Nuclei were stained with Hoechst 33258. Pictures are merged in the *right panels*. *Scale bar*: 20 μm. *F*, schematic representation of zebrafish dreCmas1. *Gray boxes* indicate identified BCs, and *open boxes* the conserved CMAS motifs.

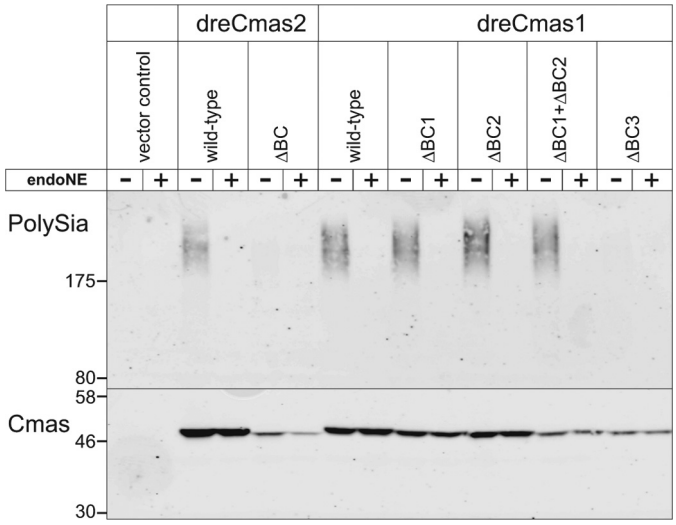


FIGURE 7. ***In vivo* activity analysis of wild-type and mutant dreCmas in LEC29.Lec32 cells.** C-terminally Myc-V5 tagged dreCmas1 and dreCmas2 and deletion mutants (ΔBC) were transiently expressed in LEC29.Lec32 cells. Empty pcDNA3-MycV5 vector was used as control. Whole cell lysates were separated by 7% SDS-PAGE, and the expression of polySia was monitored by Western blot analysis using mAb 735 (*upper panel*). Specificity of the mAb 735 was controlled in a parallel aliquot of the cell lysates by degradation of polySia with endosialidase E treatment (*endoNE*+). Expression levels of the recombinant proteins were analyzed by 12% SDS-PAGE followed by Western blot analysis using anti-V5 mAb (*lower panel*). Reduced protein expression levels were observed for the inactive dreCmas1ΔBC3 and dreCmas2ΔBC and the active dreCmas1ΔBC1+2.

DISCUSSION

In the present study, we report the identification and characterization of two *cmas* genes in zebrafish. We showed that both paralogues encoded active enzymes that differ with regard to their spatial expression pattern, intracellular localization, and substrate specificity. In contrast to mammalian genomes, which contain a single *Cmas* gene, we identified two *cmas* genes in *D. rerio* and other fish belonging to the Otocephala clade, such as *P. promelas*, *I. furcatus,* and *I. punctatus*. Representatives of Euteleostei lacked the second gene. We found a syntenic correspondence between chromosomes carrying zebrafish dre*cmas*1 and the chicken *Cmas* gene, suggesting that the order of genes surrounding dre*cmas1* corresponds to the gene arrangement in the common ancestor of the teleost fish. The gene duplication observed in Otocephala most likely originated from the third WGD, which occurred some 305–450 million years ago in the ray-finned fish linage at the base of teleost radiation (33). This WGD has often been regarded as a driving force for the diversification of teleost fish, which constitute the most specious vertebrate lineage (34). Because Euteleostei lacked not only the second copy of *cmas2*, but also the small circumjacent genomic region, it can be concluded that *cmas2* was lost in this lineage. Instability of this genomic region is also emphasized by translocation of *cmas1* in Euteleostei from an ancestral chromosome to a new site. The reason for this instability remains to be elucidated.

At the primary sequence level, the two zebrafish Cmas enzymes shared 56% identity. Divergent residues were found throughout the molecules with increased frequency in the C-terminal domain. Accordingly, sequence similarity to vertebrate CMAS enzymes was concentrated in the enzymatically active N-terminal domain, but was still significant in the C-terminal domain. In the murine enzyme, the C-terminal domain mediates tetramerization and thereby modulates the kinetic properties (20). Based on the observed sequence similarity, it is reasonable to conclude that the C-terminal domain is also responsible for the quaternary organization of the dreCmas paralogues. Tetramerization is consistent with results obtained for purified endogenous CMAS enzymes from different vertebrate tissues and species revealing trimeric to pentameric forms (for review, see Refs. 11 and 13). The similarity of vertebrate CMAS enzymes is also reflected in the organization of the active site pocket, which is built by the five conserved primary sequence motifs. Deletion of the central basic cluster (Figs. 1 and 7) abolished enzymatic activity as reported for mouse (14), rainbow trout (15), and zebrafish CMAS enzymes (this study). X-ray analysis of the enzymatically active domain of murine CMAS confirmed that this amino acid stretch is part of the active site (31).

A comparative analysis of fish genomes by Kassahn *et al.* (35) revealed that a minimum of 4% of protein-encoding genes have

been retained in duplicate in the teleost lineage from the last WGD. To avoid genetic redundancy, permanent preservation of two paralogues is assumed to be due either to subdivision of the ancestral function (subfunctionalization) or to the acquirement of a new function (neofunctionalization). Regarding the two dreCmas paralogues, subfunctionalization was manifested, for example, in substrate specificity. dreCmas1 showed the highest activity toward Neu5Ac and was only poorly active using the deaminated sugar KDN *in vitro*. KDN, however, was the preferred substrate for dreCmas2, which showed poor *in vitro* activity with Neu5Ac. This observation was unexpected because dreCmas2 resembles omyCmas in substrate specificity (KDN≫Neu5Ac>Neu5Gc), whereas on the amino acid level, omyCmas and dreCmas1 are more closely related (36). With further identification and characterization of *cmas* genes from other fish and higher vertebrate species, the apparent discrepancies between sequence homologies and similarities in enzymatic properties may be resolved, and the importance of single amino acids in functional domains may be elucidated. The functional relevance of dreCmas2 expression with high *in vitro* preference for KDN remains elusive because thus far, no KDN has been detected in zebrafish embryos. Only Neu5Ac and Neu5Gc are found during 0.5–48 hpf (3, 4). However, KDN was shown to be the major Sia derivative in the skin mucus of the closely related carp (*Cyprinus carpio*) (37). These findings may either reflect evolutionary differences within the order Cypriniformes or point to the possibility that KDN is expressed in zebrafish tissues in developmental stages not analyzed to date. Furthermore, the results obtained in the cellular system indicate that dreCmas2 can participate in Neu5Ac activation *in vivo*.

In addition to the substrate specificity, differences in the spatial expression pattern underline the functional divergence of the two dreCmas paralogues. dre*cmas*1 was prominently expressed in regions of active neurogenesis, presumably postmitotic neurons in forebrain, midbrain, and hindbrain, as well as in the somites. In contrast, dre*cmas*2 was weakly but ubiquitously detected from 9 hpf throughout somitogenesis and quickly down-regulated thereafter. Both dre*cmas* paralogues were expressed before and maintained during the expression of sialyltransferases acting downstream of CMAS in Sia biosynthesis. The expression of the sialyltransferases *st8sia1* and *st8sia5* (38) as well as of polysialyltransferase *st8sia2* starts around 10 hpf in the developing nervous system (7, 39). The oligosialyltransferase *st8sia3* shows a highly dynamic expression in somites and somite-derived structures (8). It is also detected in the developing nervous system (38) corresponding to *drecmas1* expression in early embryonic stages. Additionally, ubiquitous expression of the monosialyltransferase *st8sia6* and the GM3 synthase *st3gal4* is observed during *D. rerio* development (38, 40). Thus, dre*cmas* spatial expression patterns coincide with those of sialyltransferases. Differences in the spatial and/or temporal expression profile during embryogenesis have been reported for nearly all *D. rerio* duplicate gene pairs, suggesting that their sub- and neofunctionalization enables a more specialized control of development (35).

Corroborating our results, functional divergence of zebrafish paralogues has been reported in terms of subcellular localization (35). So far, vertebrate CMAS enzymes, independent of tissue or species, have been predominantly found in nuclear fractions with just minor amounts in other compartments (11, 13). Nuclear sequestration has been confirmed with recombinant proteins in different cellular systems (15, 41, 42). In this study, we identified and characterized the first cytosolic vertebrate CMAS. Only dreCmas1 was targeted to the nuclear compartment, whereas dreCmas2 was exclusively retained in the cytoplasm. Entry of dreCmas1 to the cell nucleus was mediated by a bipartite NLS ($K^9R(X)_{13}$KRRK), which is related to the NLS in omyCmas ($K^5KR(X)_{10}$RKAK) in terms of intramolecular localization (15). Whether the observed differences in subcellular localization of the dreCmas proteins affect their enzymatic properties or indicate different functions remains to be elucidated.

In summary, we identified the first vertebrate CMAS exclusively found in the cytoplasm *in vivo* and demonstrated the existence of a second Cmas in zebrafish that was directed to the nuclear compartment by a bipartite NLS. Both enzymes assembled into tetramers and showed enzymatic activity *in vitro* and in a cellular system. Like other duplicated genes that have arisen in WGD, zebrafish Cmas paralogues diverged in function, as is obvious by differences in their expression patterns, subcellular distributions, and substrate specificities. In parallel to unraveling the biological consequences of expressing duplicated paralogues of *cmas* and their individual function in vertebrate development, these differences provide a base for resolving the molecular and cellular requirements of CMAS enzymatic activity.

Acknowledgments—EPC cells (ATCC CRL-2872) were kindly provided by Prof. Dr. Dieter Steinhagen (Stiftung Tierärztliche Hochschule Hannover, Germany). We thank Dr. Steffen Scholpp (Institute of Toxicology and Genetics (ITG), Karlsruhe Institute of Technology, Germany) for discussing the drecmas expression patterns and Kerstin Weber for expert technical assistance.

REFERENCES

1. Cohen, M., and Varki, A. (2010) The sialome: far more than the sum of its parts. *OMICS.* **14,** 455–464
2. Schwarzkopf, M., Knobeloch, K. P., Rohde, E., Hinderlich, S., Wiechens, N., Lucka, L., Horak, I., Reutter, W., and Horstkorte, R. (2002) Sialylation is essential for early development in mice. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 5267–5270
3. Guérardel, Y., Chang, L. Y., Maes, E., Huang, C. J., and Khoo, K. H. (2006) Glycomic survey mapping of zebrafish identifies unique sialylation pattern. *Glycobiology* **16,** 244–257
4. Chang, L. Y., Harduin-Lepers, A., Kitajima, K., Sato, C., Huang, C. J., Khoo, K. H., and Guérardel, Y. (2009) Developmental regulation of oligosialylation in zebrafish. *Glycoconj. J.* **26,** 247–261
5. Marx, M., Rutishauser, U., and Bastmeyer, M. (2001) Dual function of polysialic acid during zebrafish central nervous system development. *Development* **128,** 4949–4958
6. Langhauser, M., Ustinova, J., Rivera-Milla, E., Ivannikov, D., Seidl, C., Slomka, C., Finne, J., Yoshihara, Y., Bastmeyer, M., and Bentrop, J. (2012) Ncam1a and Ncam1b: two carriers of polysialic acid with different functions in the developing zebrafish nervous system. *Glycobiology* **22,** 196–209
7. Marx, M., Rivera-Milla, E., Stummeyer, K., Gerardy-Schahn, R., and Bastmeyer, M. (2007) Divergent evolution of the vertebrate polysialyltransferase Stx and Pst genes revealed by fish-to-mammal comparison. *Dev.*

Biol. **306,** 560–571
8. Bentrop, J., Marx, M., Schattschneider, S., Rivera-Milla, E., and Bastmeyer, M. (2008) Molecular evolution and expression of zebrafish St8SiaIII, an α-2,8-sialyltransferase involved in myotome development. *Dev. Dyn.* **237,** 808–818
9. Potvin, B., Raju, T. S., and Stanley, P. (1995) Lec32 is a new mutation in Chinese hamster ovary cells that essentially abrogates CMP-*N*-acetylneuraminic acid synthetase activity. *J. Biol. Chem.* **270,** 30415–30421
10. Münster-Kühnel, A. K., Tiralongo, J., Krapp, S., Weinhold, B., Ritz-Sedlacek, V., Jacob, U., and Gerardy-Schahn, R. (2004) Structure and function of vertebrate CMP-sialic acid synthetases. *Glycobiology* **14,** 43R–51R
11. Kean, E. L., Münster-Kühnel, A. K., and Gerardy-Schahn, R. (2004) CMP-sialic acid synthetase of the nucleus. *Biochim. Biophys. Acta* **1673,** 56–65
12. Kean, E. L. (1969) Sialic acid activating enzyme in ocular tissue. *Exp. Eye Res.* **8,** 44–54
13. Kean, E. L. (1991) Sialic acid activation. *Glycobiology* **1,** 441–447
14. Munster, A. K., Weinhold, B., Gotza, B., Muhlenhoff, M., Frosch, M., and Gerardy-Schahn, R. (2002) Nuclear localization signal of murine CMP-Neu5Ac synthetase includes residues required for both nuclear targeting and enzymatic activity. *J. Biol. Chem.* **277,** 19688–19696
15. Tiralongo, J., Fujita, A., Sato, C., Kitajima, K., Lehmann, F., Oschlies, M., Gerardy-Schahn, R., and Münster-Kühnel, A. K. (2007) The rainbow trout CMP-sialic acid synthetase utilizes a nuclear localization signal different from that identified in the mouse enzyme. *Glycobiology* **17,** 945–954
16. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol. Evol.* **28,** 2731–2739
17. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948
18. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4,** 363–371
19. Catchen, J. M., Conery, J. S., and Postlethwait, J. H. (2009) Automated identification of conserved synteny after whole genome duplication. *Genome Res.* **19,** 1497–1505
20. Oschlies, M., Dickmanns, A., Haselhorst, T., Schaper, W., Stummeyer, K., Tiralongo, J., Weinhold, B., Gerardy-Schahn, R., von Itzstein, M., Ficner, R., and Münster-Kühnel, A. K. (2009) A C-terminal phosphatase module conserved in vertebrate CMP-sialic acid synthetases provides a tetramerization interface for the physiologically active enzyme. *J. Mol. Biol.* **393,** 83–97
21. Heckman, K. L., and Pease, L. R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–932
22. Frosch, M., Görgen, I., Boulnois, G. J., Timmis, K. N., and Bitter-Suermann, D. (1985) NZB mouse system for production of monoclonal antibodies to weak bacterial antigens: isolation of an IgG antibody to the polysaccharide capsules of *Escherichia coli* K1 and group B meningococci. *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1194–1198
23. Heppell, J., Lorenzen, N., Armstrong, N. K., Wu, T., Lorenzen, E., Einer-Jensen, K., Schorr, J., and Davis, H. L. (1998) Development of DNA vaccines for fish: vector design, intramuscular injection and antigen expression using viral haemorrhagic septicaemia virus genes as model. *Fish Shellfish Immunol.* **8,** 271–286
24. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995) Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203,** 253–310
25. Link, V., Shevchenko, A., and Heisenberg, C. P. (2006) Proteomics of early zebrafish embryos. *BMC. Dev. Biol.* **6,** 1
26. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11,** 1475–1489
27. Schulte-Merker, S., Ho, R. K., Herrmann, B. G., and Nüsslein-Volhard, C. (1992) The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development* **116,** 1021–1032
28. Thisse, C., and Thisse, B. (2008) High-resolution *in situ* hybridization to whole mount zebrafish embryos. *Nat. Protoc.* **3,** 59–69
29. Li, C., Lu, G., and Ortí, G. (2008) Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol.* **57,** 519–539
30. Mosimann, S. C., Gilbert, M., Dombroswki, D., To, R., Wakarchuk, W., and Strynadka, N. C. (2001) Structure of a sialic acid-activating synthetase, CMP-acylneuraminate synthetase in the presence and absence of CDP. *J. Biol. Chem.* **276,** 8190–8196
31. Krapp, S., Münster-Kühnel, A. K., Kaiser, J. T., Huber, R., Tiralongo, J., Gerardy-Schahn, R., and Jacob, U. (2003) The crystal structure of murine CMP-5-*N*-acetylneuraminic acid synthetase. *J. Mol. Biol.* **334,** 625–637
32. Christophe, D., Christophe-Hobertus, C., and Pichon, B. (2000) Nuclear targeting of proteins: how many different signals? *Cell Signal.* **12,** 337–341
33. Ravi, V., and Venkatesh, B. (2008) Rapidly evolving fish genomes and teleost diversity. *Curr. Opin. Genet. Dev.* **18,** 544–550
34. Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res.* **13,** 382–390
35. Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C., and Ragan, M. A. (2009) Evolution of gene function and regulatory control after whole genome duplication: comparative analyses in vertebrates. *Genome Res.* **19,** 1404–1418
36. Nakata, D., Münster, A. K., Gerardy-Schahn, R., Aoki, N., Matsuda, T., and Kitajima, K. (2001) Molecular cloning of a unique CMP-sialic acid synthetase that effectively utilizes both deaminoneuraminic acid (KDN) and *N*-acetylneuraminic acid (Neu5Ac) as substrates. *Glycobiology* **11,** 685–692
37. Inoue, S., and Kitajima, K. (2006) KDN (deaminated neuraminic acid): dreamful past and exciting future of the newest member of the sialic acid family. *Glycoconj. J.* **23,** 277–290
38. Chang, L. Y., Mir, A. M., Thisse, C., Guérardel, Y., Delannoy, P., Thisse, B., and Harduin-Lepers, A. (2009) Molecular cloning and characterization of the expression pattern of the zebrafish α2, 8-sialyltransferases (ST8Sia) in the developing nervous system. *Glycoconj. J.* **26,** 263–275
39. Thisse, B., Pfumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X. Q., and Thisse, C. (2001) Expression of the zebrafish genome during embryogenesis. ZFIN on-line publication (NIH R01 RR15402) http://zfin.org/cgi-bin/webdriver?MIval=aa-pubview2.apg&OID=ZDB-PUB-010810-1
40. Sohn, H., Kim, Y. S., Kim, H. T., Kim, C. H., Cho, E. W., Kang, H. Y., Kim, N. S., Kim, C. H., Ryu, S. E., Lee, J. H., and Ko, J. H. (2006) Ganglioside GM3 is involved in neuronal cell death. *FASEB J.* **20,** 1248–1250
41. Münster, A. K., Eckhardt, M., Potvin, B., Mühlenhoff, M., Stanley, P., and Gerardy-Schahn, R. (1998) Mammalian cytidine 5′-monophosphate *N*-acetylneuraminic acid synthetase: a nuclear protein with evolutionarily conserved structural motifs. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 9140–9145
42. Lawrence, S. M., Huddleston, K. A., Tomiya, N., Nguyen, N., Lee, Y. C., Vann, W. F., Coleman, T. A., and Betenbaugh, M. J. (2001) Cloning and expression of human sialic acid pathway genes to generate CMP-sialic acids in insect cells. *Glycoconj. J.* **18,** 205–213