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

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

Published
2012

Journal Title
Journal of Biological Chemistry

DOI
https://doi.org/10.1074/jbc.M111.327544

Copyright Statement
This research was originally published in Journal of Biological Chemistry (JBC). Joe Tiralongo, Wiebke Schaper, Joachim Bentrop, Jana Ustinova, Linda Blume, Elina Kats, Birgit Weinhold, Martin Bastmeyer, Anja-K. Münster-Kühnel, Identification and Biochemical Characterization of Two Functional CMP-Sialic Acid Synthetases in Danio rerio, Journal of Biological Chemistry (JBC), 2012; Vol. 287(16), pp. 13239-13248. Copyright the American Society for Biochemistry and Molecular Biology. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. Please refer to the journal's website for access to the definitive version.

Downloaded from
http://hdl.handle.net/10072/48768
Characterization of two D. rerio CMP-sialic acid synthetases

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

Wiebke Schaper¹, Joachim Bentrop², Jana Ustinova², Linda Blume¹, Elina Kats³, Joe Tiralongo³, Birgit Weinhold¹, Martin Bastmeyer² and Anja Münster-Kühnel¹

¹Institute of Cellular Chemistry, Hannover Medical School (MHH), Carl-Neuberg-Straße1, D-30625 Hannover, Germany
²Zoologisches Institut I, Zell- und Neurobiologie, Karlsruhe Institute of Technology (KIT), Haid-und-Neu-Straße 9, D-76131 Karlsruhe, Germany
³Institute for Glycomics, Griffith University, Gold Coast Campus, Queensland 4222, Australia

Running title: Characterization of two D. rerio CMP-sialic acid synthetases

Address correspondence to: Dr. Anja-K. Münster-Kühnel, Institute of Cellular Chemistry, Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany; Tel. +49 511 532 8245; Fax: +49 511 532 8801; E-mail: muenster.anja@mh-hannover.de

Keywords: Sialic acid, CMP-sialic acid synthetase, zebrafish, glycobiology

Background: Addition of sialic acid to the non-reducing end of glycoconjugates requires activation by CMP-sialic acid synthetase (CMAS).

Results: In zebrafish, we identified two CMAS enzymes which 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.

ABSTRACT
Sialic acids (Sia) form the non-reducing end of the bulk of cell surface expressed glycoconjugates. They are, therefore, major elements in intercellular communication processes. 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 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, while 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 (Neu5Ac), while the cytosolic dreCmas2 showed highest affinity for 5-deamino-neuraminic acid (KDN). 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.

INTRODUCTION
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 influences 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-glycolyneuraminic 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) were detected in zebrafish on neural cell adhesion molecules Ncam1a (NCAM) and Ncam1b (PCAM) (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 oligoSia 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 (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. While 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 (Supplementary 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 (ML) 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 (407 aa) long alignment. All sites in triplets were used, 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 ML heuristic method.

The programs EBI-ClustalW (17) and Bio Edit 7.0.5 (Tom Hall, Ibis Therapeutics, Carlsbad, USA) were used for multiple sequence alignments for visualization of conserved domains and nuclear localization signals. Prediction of putative NLS sequences was performed by eye and by use of PsortII (http://psort.ims.u-tokyo.ac.jp) and Cokol (http://cubic.bioc.columbia.edu/predictNLS). Structure prediction was done using Phyre (18).
Characterization of two *D. rerio* CMP-sialic acid synthetases

Blocks of synteny were determined with the help of 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 ten 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 heterochordates 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 hpf zebrafish embryos using Trizol (Invitrogen). cDNA was generated using the SuperScriptII™cDNA synthesis kit (Invitrogen). Full open reading frames of dre*cmas1* and dre*cmas2* were amplified using specific primers (Supplementary Table 2) and Phusion DNA-Polymerase (Finnzymes). PCR-products were sub-cloned into pCR-BluntII-Topo vector (Invitrogen). Their identities were confirmed by sequencing.

**Cmas purification and size exclusion chromatography** - We sub-cloned dre*cmas1* and dre*cmas2* cDNAs into a modified pET22b-Strep vector allowing the expression of N-terminally Strep-II-tagged (IBA) proteins. Primer sequences for the Topo vector (Invitrogen). Full open reading frames of dre*cmas1* and dre*cmas2* were amplified using specific primers (Supplementary Table 2) and Phusion DNA-Polymerase (Finnzymes). PCR-products were sub-cloned into pCR-BluntII-Topo vector (Invitrogen). Their identities were confirmed by sequencing.

**Peak fractions were desalted (HiPrep 26/10, GE Healthcare)**. 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 MgCl₂, 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 http://expasy.org/tools/protparam.html. Size exclusion chromatography was performed as previously described using the abovementioned buffer (20). Western blots following purification and size exclusion chromatography 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 TrisHCl pH 7.5, 25 mM MgCl₂ and started by the addition of 0.5 ng/µl dreCmas1 or dreCmas2. CTP, UTP, ATP, or GTP were used at f.c. 1000 µM and Neu5Ac, Neu5Gc or Kdn at 4000 µM.

**Construction of plasmids for cell culture experiments** - We sub-cloned 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 (Supplementary 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 (Supplementary 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 either in the absence or presence of 100 ng endoNE to remove polySia. Equal protein amounts were separated by 7% and 10% SDS-PAGE, respectively. PolySia and dreCmas expression were analyzed by Western blotting using anti-polySia mAb 735 (5 µg/ml), and goat anti-mouse 680 IRDye secondary antibody (LICOR). Scanning was performed with the LICOR 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 (MEM) with Earles salts and L-Gln (PAA, #E15-825) and 10% fetal calf serum at 5% CO₂ and 26°C. 24 h before transfection 3.5×10⁵ EPC cells were cultured on glass coverslips in a 12-well plate. Cells were transfected with 0.5 µg DNA using 3 µl Fugene6 (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 hours post-fertilization (hpf) according to Kimmel et al. (24).

Nuclear and cytoplasmatic extracts - Zebrafish embryos were dechorionated and deyolked according to Link et al. (25) and flash frozen in N₂. 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 which were originally derived against murine CMAS and which specifically recognize the distinct dreCmas enzymes (Sellmeier et al., manuscript in preparation).

Whole mount in situ hybridization - Whole mount in situ hybridization of Digoxigenin (DIG) labeled RNA probes was carried out according to standard protocols (27;28). To generate specific probes the 3’regions of drcmas1 and drcmas2 were amplified by RT-PCR with specific primers (Supplementary Table 2) and cloned into pCR-Blunt II-TOPO (Invitrogen). Antisense and sense (control) riboprobes were synthesized from linearized plasmids using the DIG RNA labeling kit (Roche). 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 (drcmas1). According to the current version of the D. rerio genome assembly, the second gene (drcmas2) 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 drcmas genes by RT-PCR using gene specific primers. The amplified drcmas1 cDNA encoded a protein of 430 amino acids with a calculated molecular weight of 48.2 kDa and was identical to that previously reported (acc. no. C8K18993). The amplified drcmas2 contained an open reading frame (ORF) that encoded a protein of 423 amino acids (calculated molecular weight 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 Figure 1A and highlights identical residues in black and conserved residues in grey. Like all other known CMAS enzymes, both dreCmas sequences harbored five conserved primary sequence motifs (motifs I to V) in the N-terminal domain (NT, Figure 1A). In line with other vertebrate CMAS enzymes, dreCmas1 and dreCmas2 possessed an additional C-terminal domain (CT, Figure 1A) composed of 173 and 175 amino acids, respectively. Compared to the N-terminal domains, the CTs 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 NCBI fish EST database did not give evidence for two different intact cmas genes in any of the species belonging to Euteleostei (sensu 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 punctatus) and the fathead minnow (Pimephales promelas). The Maximum Likelihood phylogenetic tree demonstrates that fish cmas genes segregated into two distinct clusters comprising cmas1 and cmas2, respectively (Figure 1B). 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 trace 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 (Figure 2A, left). A significant intra-genomic synteny between zebrafish chromosomes 4 and 25 corroborated that both cmas genes were paralogues resulting from the WGD (Figure 2A, right). In Euteleostei,
Characterization of two *D. rerio* CMP-sialic acid synthetases

*cmasl* was excised from a homologue of the Dre chromosome 4 and transferred to a foreign location, whereas *cmas2* was excised from the homologue of Dre chromosome 25 and lost (probably during translocation) (Figure 2B and Supplementary Figure 1). Summarized, 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), while 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 **drecmas1** and **drecmas2** was analyzed in the developing zebrafish by *in situ* hybridizations using specific probes derived from the 3’UTR of both mRNAs. In general, **drecmas1** showed a stronger and more distinct expression than **drecmas2**. At 90% epiboly (9 hpf) - beside a basal, more or less ubiquitous expression - **drecmas1** was detected in the axial mesoderm, especially in the notochord primordium (Figure 3A). At 18 hpf, **drecmas1** showed a robust expression in the entire central nervous system, the somites, the notochord, and the developing pronephric duct (Figure 3C). With progressing development, expression of **drecmas1** was down-regulated in the trunk. It persisted in the central nervous system and was up-regulated in the kidney and the liver primordium (Figure 3H, J). **drecmas2** 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 (Figure 3B, D). During the pharyngula stage, **drecmas2** expression was restricted to the brain (Figure 3G), and it was down-regulated around hatching (Figure 3K). While **drecmas2** showed expression in the heart at 48 hpf (Figure 3I), 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 **drecmas** genes encode active enzymes, the proteins were expressed with an N-terminal StrepII-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 Blue staining as well as by Western blot analysis of the fractions (Figure 4C and E). **drecmas1** showed expression in the heart at 48 hpf (Figure 3I), we could not detect it in skeletal muscle, liver, or kidney. While **drecmas2** showed expression in the heart at 48 hpf (Figure 3I), we could not detect it in skeletal muscle, liver, or kidney. While **drecmas2** showed expression in the heart at 48 hpf (Figure 3I), we could not detect it in skeletal muscle, liver, or kidney.

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 *Pimephales 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 (Figure 5A). 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 distribution of the endogenous enzymes, nuclear and cytoplasmatic extracts were prepared from 48 hpf zebrafish embryos and analyzed by Western blotting (Figure 5B). The specificity of the antibodies for either of the dreCmas enzymes was confirmed with purified recombinant proteins (Figure 5C).

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) (Figure 5B).

A bipartite nuclear localization signal targets dreCmas1 to the nuclear compartment - To identify the nuclear localization signal (NLS) essential for nuclear targeting of dreCmas1, primary sequence analysis was performed by eye and by use of the programs PSORT and COKOL. Four basic clusters (BC1 to BC4) were identified in the dreCmas1 sequence (Figure 1A). BC1 (K\textsubscript{179}RAMK\textsuperscript{181}) and BC2 (K\textsubscript{27}RRK\textsubscript{37}) were located at the N-terminus, BC3 (P\textsuperscript{184}ACPRR\textsuperscript{196}) at the center, and BC4 (K\textsubscript{412}KKAK\textsuperscript{416}) at C-terminus. BC4 and the surrounding twelve amino acid residues were strictly conserved in rainbow trout Cmas (Figure 1A). Since 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 (Figure 6) and NIH-3T3 cells (data not shown). The subcellular localization was analyzed by indirect immunofluorescence microscopy. While 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\textsubscript{179}RAMK\textsubscript{181}(X)\textsubscript{18}K\textsubscript{27}RRK\textsubscript{37}) according to the consensus sequence (K/R)\textsubscript{X}10-12(K/R)\textsubscript{3} (32). dreCmas2 in contrast contained a single short BC (R\textsubscript{170}PRR) (dre2-BC) at the center of the protein (Figure 1A) which did not fit to the consensus sequence of a monopartite NLS (K/R)\textsubscript{4-6} (32).

The absence of an NLS was in perfect agreement with the cytosolic localization of dreCmas2 (Figure 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 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 Figure 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 (+endoNE) leading to the disappearance of the polySia signal (see Figure 7). In all experiments dreCmas expression was controlled by Western blot staining (Figure 7, Cmas). 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 \textit{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 \textit{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.

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 \textit{Cmas} gene, we identified two \textit{cmas} genes in \textit{D. rerio} and other fish belonging to the Otocephala clade, such as \textit{P. promelas}, \textit{I. furcatus}, and \textit{I. punctatus}. Representatives of Euteleostei lacked the second gene. We found a syntenic correspondence
between chromosomes carrying zebrafish dreCmas1 and the chicken Cmas gene, suggesting that the order of genes surrounding dreCmas1 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 lineage 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 driving force for the diversification of Teleost (34). Since Euteleostei lacked fish which constitute the most specious WGD has often been regarded as a driving force for the diversification of Teleost fish which constitute the most specious vetebrate linage (34). Since Euteleostei lacked 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 (11;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 and coworkers (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 acquisition of a new function (neofunctionalization). Regarding the two dreCmas paralogues, subfunctionalization was manifested, for example in substrate specificity. dreCmas1 showed the highest activity towards 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) while, 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 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 to 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. dreCmas1 was prominently expressed in regions of active neurogenesis, presumably postmitotic neurons in forebrain, midbrain, and hindbrain, as well as in the somites. In contrast dreCmas2 was weakly but ubiquitously detected from 9 hpf throughout somitogenesis and quickly down-regulated thereafter. Both dreCmas 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 oligo-sialyltransferase 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 mono-sialyltransferase st8sia6 and the GM3 synthase st3gal4 is observed during D. rerio development (38;40). Thus, dreCmas 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, while dreCmas2 was exclusively retained in the cytoplasm. Entry of dreCmas1 to the cell nucleus was mediated by a bipartite NLS (K^S(R(X)_{10})KRRK), which is related to the NLS in omyCmas (K^SKR(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 indicates 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.

REFERENCES

Characterization of two D. rerio CMP-sialic acid synthetases


ACKNOWLEDGEMENTS

EPC cells (ATCC CRL-2872) were kindly provided by Prof. Dr. Dieter Steinheigen (Stiftung Tierärztliche Hochschule Hannover, Germany). We thank Dr. Steffen Scholpp (ITG, Karlsruhe Institute of Technology, Germany) for discussing the dreCmas expression patterns and Kerstin Weber for expert technical assistance. This work received financial support from the German Research Foundation (DFG) to AKMK (MU1849/1), and to MB and JB (BA 1034/15-19) and is supported by funding from the DFG for the Cluster of Excellence REBIRTH (From Regenerative Biology to Reconstructive Therapy).

FIGURE LEGENDS

**Figure 1. Relations between selected vertebrate CMAS sequences.** (A) Multi-sequence alignment. Amino acid sequences of the two *D. rerio* Cmas proteins (dreCmas1 and dreCmas2) have been aligned with *Oncorhynchus mykiss* (omyCmas) and *Mus musculus* CMAS (mmuCMAS) using ClustalW to display maximum homology. Strictly and highly conserved residues are shaded in black and grey, 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 (BC) identified in dreCmas1 (BC1 to BC4) and dreCmas2 (dre2-BC) are boxed in grey 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 either identified in the latest fish genome assemblages with dreCmas2 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 grey, fish cmas1 genes in light grey. Euteleost species which 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: Danio (rerio), Gasterosteus (aculeatus), Gallus (gallus), Ictalurus (furcatus and punctatus), Mus (musculus), Oncorhynchus (mykiss), Oryzias (latipes), Pimephales (promelas), Saccoglossus (kowalevskii), Salmo (salar), Takifugu (rubripes), Xenopus (tropicalis).

**Figure 2. Conserved synteny elucidates evolution of cmas genes in fishes.** (A) Synteny among chicken (Gallus gallus, Gga) chromosome 1, zebrafish (Danio 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 Chr4). The order of genes on the Dre4 is real and is centered at cmas1. The order of genes on chicken chromosome 1 (Gga1) and zebrafish chromosome 25 (Dre Chr25) 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 whole genome duplication (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 Supplementary Figure 1). Colored balks represent chromosomes, similar colors indicate homology. O – Otocephala, E – Euteleostei.
Characterization of two *D. rerio* CMP-sialic acid synthetases

**Figure 3. Expression of dreccmas1 and dreccmas2 mRNAs in zebrafish embryos.** Embryos were hybridized *in situ* with digoxigenin-labeled riboprobes. (A,C,F,H,J) dreccmas1; (B,D,G,I,K) dreccmas2; (B,E) sense controls. Insert in (F) shows a higher magnification of the indicated region. Ages are indicated in hpf (hours post fertilization). (A-B) dorsal views; (D-L) lateral views. dreccmas1 is expressed in the posterior axial mesoderm at 9 hpf (A), throughout the central nervous system at all analyzed timepoints (C,F,H,J), in the somites during segmentation (C), in the pronephric kidney starting around 18 hpf (C,F,H,J), and in the liver (H,J). dreccmas2 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,L). 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.

**Figure 4. Purification and size-exclusion chromatography of recombinant dreCmas1 and dreCmas2.** N-terminally StrepII-tagged dreCmas1 and -2 were expressed in *E. coli* and purified by Strep-Tactin affinity chromatography. Purification was followed by SDS-PAGE with Coomassie (A upper panel) and Strep-Tactin-AP Western blot staining (A 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 StrepII-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 staining and Strep-Tactin-AP Western blotting (C and E) of fractions.

**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). 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. While 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).

**Figure 6. Intracellular localization of dreCmas1 mutants in EPC cells.** Indirect immunofluorescence analysis of C-terminally Myc-V5 tagged dreCmas1 (A) and deletion mutants (B–E) lacking individual basic clusters (see F and Figure 1A) after transient expression in EPC cells. Staining was performed with anti-Myc mAb and a Cy3-conjugated secondary antibody (Cmas). Nuclei were stained with Hoechst 33258. Merged pictures are shown in the right panels. Scale bar: 20 µm. (F) Schematic representation of zebrafish dreCmas1. Grey 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 -2 and deletion mutants (∆BC) 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 dreCmas1∆BC3 and dreCmas2∆BC and the active dreCmas1∆BC1+2.
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.

<table>
<thead>
<tr>
<th></th>
<th>dreCmas1</th>
<th>dreCmas2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neu5Ac</td>
<td>4.5 +/-0.3</td>
<td>0.4 +/-0.1</td>
</tr>
<tr>
<td>Neu5Gc</td>
<td>2.7 +/-0.1</td>
<td>0.2 +/-0.1</td>
</tr>
<tr>
<td>KDN</td>
<td>0.3 +/-0.1</td>
<td>3.3 +/-0.4</td>
</tr>
</tbody>
</table>
Characterization of two *D. rerio* CMP-sialic acid synthetases

**Figure 1**

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>Motif I</th>
<th>Motif II</th>
<th>Motif III</th>
<th>Motif IV</th>
<th>Motif V</th>
</tr>
</thead>
<tbody>
<tr>
<td>dreCms1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>cmyCms1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>dreCms2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>mmuCms5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

B

Phylogenetic tree showing the evolutionary relationships among different species.
Figure 2
Characterization of two *D. rerio* CMP-sialic acid synthetases

Figure 3
Characterization of two *D. rerio* CMP-sialic acid synthetases

**Figure 4**

A: gel images showing bands for dreCmas1 and dreCmas2

B: chromatogram for dreCmas1 with absorbance at 280 nm plotted against elution volume (ml)

C: Western blot for dreCmas1 showing bands at 66 and 45 kDa

D: chromatogram for dreCmas2 with absorbance at 280 nm plotted against elution volume (ml)

E: Western blot for dreCmas2 showing bands at 66 and 45 kDa
Figure 5

A

Flag-dre-Cmas1

dre-Cmas1-Myc

Flag-dre-Cmas2

dre-Cmas2-Myc

B

<table>
<thead>
<tr>
<th>anti-dreCmas1</th>
<th>anti-dreCmas2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>NE</td>
</tr>
<tr>
<td>-65</td>
<td>-45</td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>anti-dreCmas1</th>
<th>anti-dreCmas2</th>
</tr>
</thead>
<tbody>
<tr>
<td>dre1</td>
<td>dre2</td>
</tr>
<tr>
<td>-65</td>
<td>-45</td>
</tr>
</tbody>
</table>
Characterization of two *D. rerio* CMP-sialic acid synthetases

Figure 6

A

dreCmas1

B

ΔBC1

C

ΔBC2

D

ΔBC1+2

E

ΔBC3

F

BC1: K^{RAMK}_{13}
BC2: K^{RRK}_{27}
BC3: P^{ACPRR}_{190}
Characterization of two *D. rerio* CMP-sialic acid synthetases

**Figure 7**

![Image](image_url)