Structural shifts of aldehyde dehydrogenase enzymes were instrumental for the early evolution of retinoid-dependent axial patterning in metazoans

Tiago J. P. Sobreira[a,1], Ferdinand Marlétaz[b,1], Marcos Simões-Costa[c,d,1], Deborah Schechtman[e], Alexandre C. Pereira[a], Frédéric Brunet[b], Sarah Sweeney[f], Ariel Pani[g,h], Jochanan Aronowicz[h], Christopher J. Lowe[h], Bradley Davidson[f], Vincent Laudet[b], Marianne Bronner[i], Paulo S. L. de Oliveira[a], Michael Schubert[b,2,3], and José Xavier-Neto[c,2,3]

[a]Laboratório de Genética e Cardiologia Molecular, Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, 05403-000, São Paulo-SP, Brazil; [b]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 07, France; [c]Laboratório Nacional de Biociências, Campus do Laboratório Nacional de Luz Síncrotron, 13083-970, Campinas-SP, Brazil; [d]Departamento de Biologia Celular e do Desenvolvimento, Instituto de Ciências Biomédicas, University of São Paulo, 05508-900, São Paulo-SP, Brazil; [e]Departamento de Bioquímica, Instituto de Química da Universidade de São Paulo, 05508-900, São Paulo-SP, Brazil; [f]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85724; [g]Committee on Evolutionary Biology, University of Chicago, Chicago, IL 60637; [h]Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA 93950; and [i]Division of Biology 139-74, California Institute of Technology, Pasadena, CA 91125

Edited* by John Gerhart, University of California, Berkeley, CA, and approved November 10, 2010 (received for review August 17, 2010)

Aldehyde dehydrogenases (ALDHs) catabolize toxic aldehydes and process the vitamin A-derived retinaldehyde into retinoic acid (RA), a small diffusible molecule and a pivotal chordate morphogen. In this study, we combine phylogenetic, structural, genomic, and developmental gene expression analyses to examine the evolutionary origins of ALDH substrate preference. Structural modeling reveals that processing of small aldehydes, such as acetaldehyde, by ALDH2, versus large aldehydes, including retinaldehyde, by ALDH1A is associated with small versus large substrate entry channels (SECs), respectively. Moreover, we show that metazoan ALDH1s and ALDH2s are members of a single ALDH1/2 clade and that during evolution, eukaryote ALDH1/2s often switched between large and small SECs after gene duplication, transforming constricted channels into wide opened ones and vice versa. Ancestral sequence reconstructions suggest that during the evolutionary emergence of RA signaling, the ancestral, narrow-channeled metazoan ALDH1/2 gave rise to large ALDH1 channels capable of accommodating bulky aldehydes, such as retinaldehyde, supporting the view that retinoid-dependent signaling arose from ancestral cellular detoxification mechanisms. Our analyses also indicate that, on a more restricted evolutionary scale, ALDH1 duplicates from invertebrate chordates (amphioxus and ascidian tunicates) underwent switches to smaller and narrower SECs. When combined with alterations in gene expression, these switches led to neofunctionalization from ALDH1-like roles in embryonic patterning to systemic, ALDH2-like roles, suggesting functional shifts from signaling to detoxification.

Aldehyde dehydrogenase phylogeny | *Branchiostoma floridae* | *Ciona intestinalis* versus *Ciona savignyi* | evolution of retinoic acid signaling | origins of morphogen-dependent signaling

In animal development, major signaling pathways are controlled by morphogens, diffusible molecules whose evolutionary origins are difficult to assess. Aldehyde dehydrogenase (ALDH) enzymes are attractive subjects to study the evolution of morphogen signaling for two main reasons. First, in addition to their acknowledged role in protecting animals by catabolizing reactive biogenic and xenobiotic aldehydes, some ALDHs also synthesize signaling molecules (1–3). Prime examples for these two ALDH enzyme roles are the ALDH2s, which degrade small toxic aldehydes, such as the acetaldehyde derived from ethanol metabolism (1, 2), and the ALDH1s, which process larger aldehydes, including retinaldehyde, a vitamin A-derived precursor of the morphogen retinoic acid (RA). RA plays a critical role during embryonic development of chordates (i.e., amphioxus, tunicates, and vertebrates) and has been suggested to have already been involved in patterning the last common ancestor of bilaterian animals (4–8). Second, ALDHs are among the best-characterized proteins, and their structure and substrate profiles have been determined with exquisite precision (9–15). Thus, structural modeling of these proteins can be used to study the evolution of substrate specificity without extensive biochemical analyses (16–20).

ALDH1 and ALDH2 enzymes share a high degree of sequence identity, indicating a very close phylogenetic relationship (3). Pioneer observations by Moore et al. on human ALDH2 and sheep ALDH1 (17) suggested that their respective abilities to detoxify small aldehydes and to process large aldehydes are correlated with the size and shape of their substrate entry channels (SECs), the intramolecular cavities that direct aldehydes to the catalytic sites of ALDH enzymes. Human ALDH2 displays a narrow SEC with a constricted entrance, whereas sheep ALDH1A1 exhibits a large SEC with a broad opening (17, 18). Thus, SEC topology influences ALDH1/2 substrate preference. For example, although retinaldehyde is a good substrate for vertebrate ALDH1s and acetaldehyde is a natural substrate of ALDH2s, ALDH2s cannot process retinaldehyde and ALDH1s process acetaldehyde only extremely inefficiently (16, 17–22).

To understand the evolutionary origins of the substrate preferences of ALDH1 and ALDH2 enzymes, as well as to illuminate how signaling and protective functions are connected to these different enzyme activities, we used an integrated approach that combined genomic, phylogenetic, and structural analyses. The resulting comprehensive data set was complemented with information on developmental gene expression of ALDH1/2s in the cephalochordate amphioxus (*Branchiostoma floridae*) and the ascidian tunicate *Ciona intestinalis*. These two invertebrate chordate models possess functional RA signaling cascades and are pivotal models for understanding vertebrate origins from both a genomic and a developmental perspective (4, 23–26). Together, this work provides support for the hypothesis that some intercellular signaling mechanisms evolved from cellular detoxification pathways.

Author contributions: M.S. and J.X.-N. designed research; T.J.P.S., F.M., M.S.-C., D.S., F.B., S.S., A.P., J.A., C.J.L., B.D., P.S.L.d.O., M.S., and J.X.-N. performed research; D.S., C.J.L., B.D., M.B., and P.S.L.d.O. contributed new reagents/analytic tools; T.J.P.S., F.M., M.S.-C., D.S., A.C.P., F.B., S.S., A.P., J.A., C.J.L., B.D., V.L., P.S.L.d.O., M.S., and J.X.-N. analyzed data; and T.J.P.S., F.M., D.S., B.D., V.L., M.B., P.S.L.d.O., M.S., and J.X.-N. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]T.J.P.S., F.M., and M.S.-C. contributed equally to this work.

[2]M.S. and J.X.-N. contributed equally to this work.

[3]To whom correspondence may be addressed. E-mail: Michael.Schubert@ens-lyon.fr or xavier.neto@lnbio.org.br.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1011223108/-/DCSupplemental.

Results

SEC Volumes Distinguish Vertebrate ALDH1s from ALDH2s. To test whether SEC differences between sheep ALDH1 and human ALDH2 reflect fundamental evolutionary differences between these enzymes, we used the crystal structures of these proteins to create 3D structural models of ALDH1/2s. These models were then analyzed to determine molecular parameters, such as SEC volume, which are implicated in substrate preference (17). Our dataset shows that ALDH1s generally display larger channel volumes than ALDH2s (589 ± 59 Å^3 for ALDH1s versus 403 ± 53 Å^3 for ALDH2s, mean ± SD, $P < 0.001$) (Dataset S1 and Dataset S2). Therefore, channel volume represents a fundamental difference between ALDH1 and ALDH2, reflecting conserved structural requirements associated with processing of large and small aldehydes, respectively.

Because it is likely that the overall geometry of the SEC, rather than its volume alone, determines ALDH1/2 specificities, we looked for further mechanistic clues to the evolution of substrate preference in ALDH1/2 sequences that diverge between the biochemically well-characterized vertebrate ALDH1s and ALDH2s. We hence compiled a list of 34 amino acid sequence signatures distinguishing the large-channeled ALDH1As (known to synthesize RA) from the narrow-channeled ALDH2s (with well-characterized roles in detoxification of small toxic aldehydes). Six signatures mapped to a subset of the 27 amino acids that form the SEC (Table S1). Of those signatures, only three signatures are positioned at critical locations at the ALDH1/2 channel, suggesting that they are implicated in determining ALDH1 and ALDH2 channel volumes/functions. The remaining signatures marked residues at oligomerization domains, surface loops or inside the molecule, which, after careful examination, did not suggest immediate functional correlates (Table S1).

The first of the three SEC signatures includes amino acid 124 at the entrance ("mouth") of the ALDH1/2 channel (17). In human ALDH2, a bulky Met124 (124 Å^3 van der Waals volume) protrudes into the channel (Table S1), allowing access of only small aldehydes, whereas in sheep ALDH1A1, a small, unobtrusive Gly124 (48 Å^3) allows entry of large aldehydes (17, 18). Thus, the first amino acid signature performs a size selection function. The second channel signature includes amino acid 459 at the proximal third ("neck") of the ALDH1/2 SEC (Table S1) (17). In vertebrate ALDH2s, this amino acid is a large Phe459 (135 Å^3). In contrast, vertebrate ALDH1s typically display the smaller Val (93 Å^3) or Leu (124 Å^3) (Table S1). The third signature corresponds to aa303, close to the catalytic Cys302 (86 Å^3) at the end of the channel ("bottom") (17). In vertebrate ALDH2s, this amino acid is Cys303, whereas vertebrate ALDH1s typically display Thr303 (93 Å^3), Ile303 (124 Å^3), or Val303 (Table S1).

To understand the roles of the amino acid signatures in substrate interaction, we performed docking studies on human ALDH2 and sheep ALDH1 with small aldehydes (Movie S1, Movie S2, and Movie S3). In the ALDH2 channel, acetaldehyde is kept close to the γ-sulfur of the catalytic Cys302 (Movie S1), whereas the ALDH1 channel does not favor this close association between the aldehyde substrate and Cys302, neither for acetaldehyde (Movie S2), nor for formaldehyde (Movie S3). Analysis of the structural motifs involved in substrate retention close to the ALDH catalytic site indicates that, in ALDH2, Cys302 and Phe465 keep acetaldehyde close to the catalytic γ-sulfur and that Phe465 is latched in position by Phe459, which, in turn, is fixed by Cys303. In ALDH1, this mechanism is not present, because the interaction surfaces between Cys303 and Phe459 are missing, due to their substitution by Ile303 and Val459, respectively (Movie S2). Thus, neck and bottom signatures keep the aldehyde moiety of small substrates close to the catalytic Cys302 in ALDH2s, consistent with a structural specialization of narrow-channeled ALDHs to process small aldehydes.

Switches from Small to Large ALDH1/2 Substrate Entry Channels Operated at the Origin of RA Signaling. To understand the evolution of affinities for small and large aldehydes in ALDH2 and ALDH1, we performed large-scale phylogenetic analyses of the ALDH superfamily (Fig. 1, Figs. S1 and S2, Dataset S1, and Dataset S2). Contrary to traditional views, we found that metazoan ALDH1s and ALDH2s do not form independent families but are members of a single, well-supported ALDH1/2 clade, with ALDH2s forming a distinct group nested within this ALDH1/2 clade (Fig. 1 and Figs. S1 and S2). In contrast to ALDH2s, ALDH1s underwent multiple lineage-specific duplications. For example, the genomes of amphioxus (*B. floridae*) and the ascidian tunicates *C. intestinalis* and *C. savignyi* contain, respectively, six, four, and two ALDH1 duplicates, whereas in the hemichordate *Saccoglossus kowalevskii*, there are five ALDH1 genes (Fig. 1, Figs. S1 and S2, Dataset S1, and Dataset S2) (27).

Analyses of channel size distribution in eukaryote ALDH1s and ALDH2s indicate that SEC variation can be subdivided into small (<420 Å^3), medium (420–508 Å^3), and large channels (>508 Å^3) (Fig. S3, Dataset S1, and Dataset S2). In metazoans, small channels dominate in ALDH2s, whereas large channels preponderate in ALDH1s (Fig. 1). In plants, there are also two major ALDH1/2 groups, one with small and medium channels and another one with predominantly large channels (345 ± 47 Å^3 versus 561 ± 78 Å^3, $P < 0.05$), suggesting that a single ALDH1/2 ancestor duplicated early in plant evolution, giving rise to two distinct functional classes (28). Fungi experienced a distinct diversification pattern with various fungal lineages independently duplicating a single ALDH1/2 ancestor (29). These duplicates subsequently underwent SEC alterations, leading to fungal ALDH1/2 enzymes with small, medium or large SECs (Fig. 1).

To understand how small, medium, and large SEC volumes arose during evolution of the ALDH1/2 clade, we reconstructed ancestral sequences at selected nodes of the ALDH1/2 phylogeny by using the maximum likelihood method (Table S2). For example, the reconstructed ancestral eukaryote ALDH1/2 enzyme displays a small SEC, in which a bulky Met124 blocks the channel mouth, Phe459 constricts the channel neck, and Cys303 occupies the channel bottom, similar to modern metazoan ALDH2 SECs (Table S2). This reconstruction indicates that the vertebrate ALDH2 SEC signatures that distinguish them from vertebrate ALDH1s are ancestral, rather than derived, reflecting ancient structural adaptation to process small aldehydes. Table S2 also shows that the emergence of large-channeled ALDH1 enzymes was accompanied by substitution of an ancestral bulky Met124 at the SEC mouth by small amino acids, such as Gly and Ala. This observation indicates that the wide open channels of ALDH1 enzymes, which process large aldehydes, such as retinaldehyde, evolved from a background of small, constricted channels similar to those displayed by vertebrate ALDH2s, which process small, toxic aldehydes.

ALDH1 Duplication and Divergence Switched Large into Narrow Substrate Entry Channels in Invertebrate Chordates. Curiously, the three SEC signatures that efficiently discriminate large-channeled vertebrate ALDH1s from narrow-channeled vertebrate ALDH2s do not distinguish their invertebrate chordate counterparts, because some invertebrate chordate ALDH1s display channel signatures typical for vertebrate ALDH2s (Table S1). To determine how these SEC variations evolved in the framework of the multiple, lineage-specific ALDH1 duplications of invertebrate chordates, we focused on the cephalochordate amphioxus, whose genome encodes six ALDH1s and a single ALDH2 (27), and on two closely related ascidian tunicates: *C. intestinalis* with four ALDH1s and a single ALDH2, and *C. savignyi* with two ALDH1s and a single ALDH2.

Because the first ALDH2 signature at the channel mouth probably selects for smaller substrates, we hypothesized that bulky and small residues may be similarly present in invertebrate chordate ALDH2s and ALDH1s, respectively. Accordingly, a bulky Leu124 protrudes into the channel mouth of amphioxus, *C. intestinalis*, and *C. savignyi* ALDH2s (Fig. 2, and Figs. S4 and S5). Amphioxus ALDH1s are heterogeneous in that a small Gly124 is embedded into the channel border without constricting the

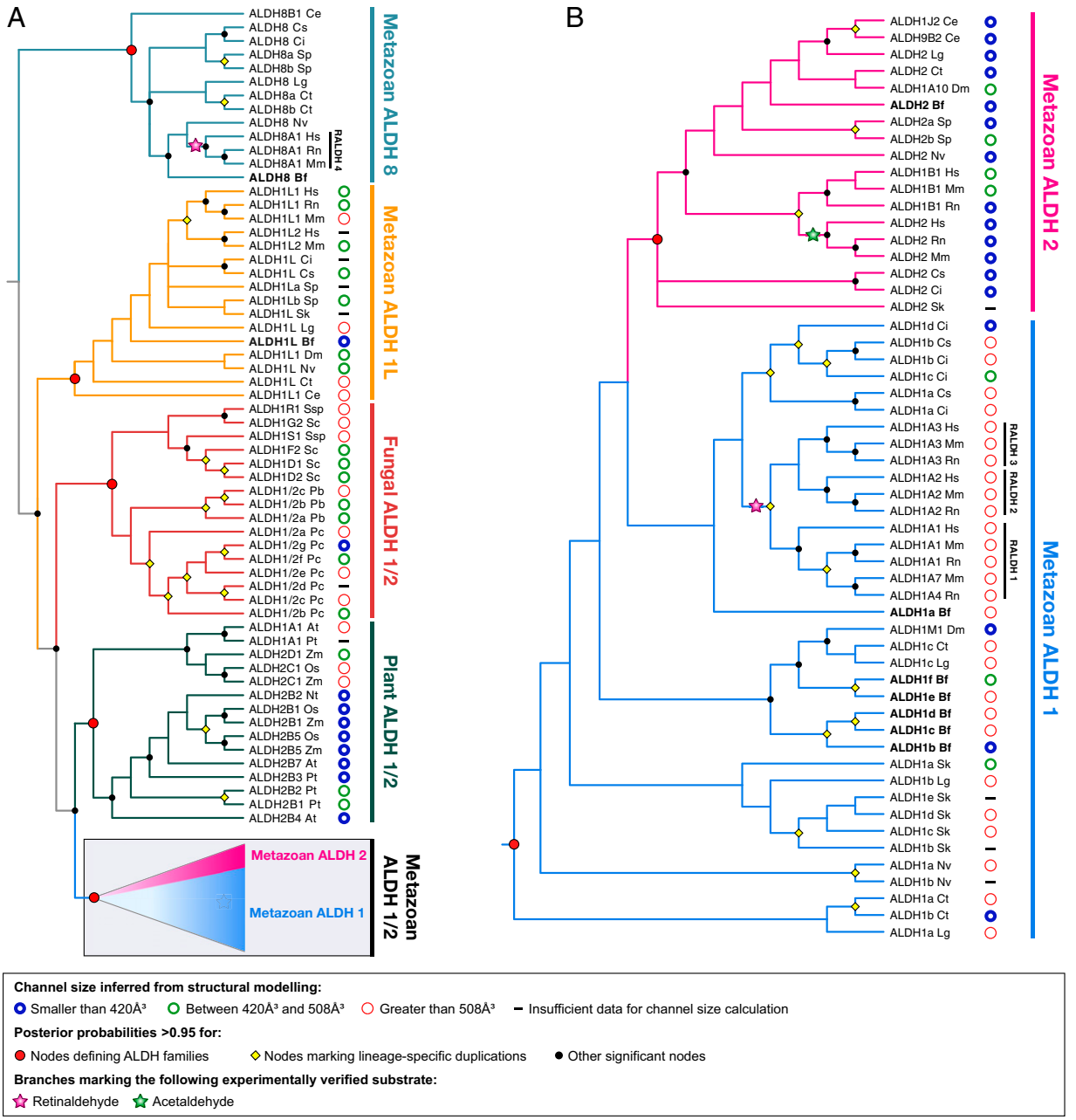

Fig. 1. Phylogeny of ALDH1/2s, ALDH1Ls, and ALDH8s. The overall topology with the ALDH8s used as outgroup is shown in *A*. The boxed area in *A* highlights the metazoan ALDH1/2 clade, which is depicted in *B*. Nodes with significant posterior probabilities (>0.95) are highlighted with icons. The actual values are given in Fig. S2. Channel size categories are shown diagrammatically for ALDH1/2s and ALDH1Ls with sufficient sequence information. ALDH8s are too divergent for these calculations. At, *Arabidopsis thaliana*; Bf, *Branchiostoma floridae*; Ct, *Capitella teleta*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Cs, *Ciona savignyi*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Lg, *Lottia gigantea*; Mm, *Mus musculus*; Nv, *Nematostella vectensis*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; Pc, *Phanerochaete chrysosporium*; Pb, *Phycomyces blakesleeanus*; Pt, *Populus trichocarpa*; Rn, *Rattus norvegicus*; Sc, *Saccharomyces cerevisiae*; Sk, *Saccoglossus kowalevskii*; Sp, *Strongylocentrotus purpuratus*; Ssp, *Schizosaccharomyces pombe*; Zm, *Zea mays*.

channel in amphioxus ALDH1a and ALDH1d, whereas the larger Glu124 (109 Å^3) or Ser124 (73 Å^3) partially obstruct the channel in the other four amphioxus duplicates, leaving insufficient space to accommodate the β-ionone moiety of retinaldehyde (Fig. 2). The ALDH1s from *C. intestinalis* and *C. savignyi* are also heterogeneous: *C. intestinalis* ALDH1a and ALDH1d, as well as *C. savignyi* ALDH1a, display the small Gly and *C. savignyi* ALDH1b, the small Ala (67 Å^3) at position 124, which do not obstruct the channel entrance, whereas *C. intestinalis* ALDH1b and ALDH1c, respectively, display the larger Ser124 and Ile124, which interfere with retinaldehyde accommodation in the ALDH1 channel (Fig. 2, and Figs. S4 and S5).

In the ALDH2s from amphioxus, *C. intestinalis*, and *C. savignyi*, the amino acid at the second channel signature is Phe459, which constricts the channel neck with its large aromatic ring. As in vertebrate ALDH1s, in some amphioxus ALDH1s, the bulky Phe459 is substituted by smaller amino acids, such as Ile459 in ALDH1a, ALDH1b, and ALDH1d, Thr459 (93 Å^3) in ALDH1c or Gly459 in ALDH1e and ALDH1f. In ascidian tunicates, only *C. intestinalis* ALDH1d displays a Leu459, whereas all other

ALDH1s display bulky amino acids, such as Phe and Met at position 459 (Figs. S4 and S5), similar to vertebrate ALDH2s.

As in vertebrates, in amphioxus, *C. intestinalis*, and *C. savignyi*, the amino acids of the third ALDH2 signature at the channel bottom is Cys303. Amphioxus ALDH1s are heterogeneous in that ALDH1a and ALDH1d display the vertebrate ALDH1 pattern (Thr303 and Ile303, respectively), whereas all other amphioxus ALDH1s display the vertebrate ALDH2 pattern (Fig. 2). In ascidian tunicates, only *C. intestinalis* and *C. savignyi* ALDH1a display the vertebrate ALDH1 pattern with Thr303, whereas other ALDH1s show the vertebrate ALDH2 pattern (Figs. S4 and S5). Thus, after lineage-specific duplication, some ALDH1 enzymes of invertebrate chordates incorporated amino acids and/or structural motifs similar to those of the vertebrate ALDH2 channel, shifting from a large, wide open configuration to constricted SEC topologies.

Synteny analyses in amphioxus and *C. intestinalis* suggest that the ALDH1a genes from both species are the sister groups to the other cephalochordate and ascidian tunicate ALDH1s and that the amphioxus duplicates ALDH1b, ALDH1c, ALDH1d, ALDH1e, and ALDH1f and *C. intestinalis* ALDH1b, ALDH1c, and ALDH1d evolved by duplication from an ALDH1a-like ancestor (Fig. S6). Moreover, reconstructions of ancestral SEC signatures and channel structures indicate that amphioxus and *C. intestinalis* ALDH1 ancestors displayed large SECs, structurally consistent with the capacity to process retinaldehyde into RA (Figs. 1 and 2, and Figs. S2 and S4). The reconstructed amphioxus ALDH1 ancestor already displayed a typical ALDH1 SEC with a 512 $Å^3$–wide opening lined by Gly124, an unobstructed neck flanked by Val459, and a bottom Cys303. The reconstructed *C. intestinalis* ALDH1 ancestor exhibits a large, 540 $Å^3$ SEC, but, curiously, with ancestral Met124, Phe459, and Cys303 SEC signatures, suggesting that large ALDH1 channels emerged independently and by different mechanisms in cephalochordates and ascidian tunicates.

Invertebrate Chordate ALDH1 Channel Switch Is Associated with Transitions Between Restricted and Pleiotropic Expression. Our next goal was to understand the specific developmental contexts in which the invertebrate chordate ALDH1/2s are deployed. We hence assessed developmental expression of the large-channeled ALDH1s structurally capable of accommodating retinaldehyde for RA synthesis, the narrow-channeled ALDH2 genes adapted for small aldehyde detoxification, and the divergent, narrow-channeled ALDH1s.

The ALDH1a and ALDH1d genes of amphioxus and *C. intestinalis* encode enzymes with large and unobstructed SECs. Amphioxus ALDH1a is expressed caudally close to the developing tail bud with a sharp anterior boundary in the neurula (at 12 h) (Fig. 2). In *C. intestinalis*, ALDH1a is expressed in a sharp, posterior mesodermal domain in the early embryo (at 5–8 h) (Fig. S4) (30). At later embryonic stages, ALDH1a is detectable in a distinct domain in the posterior gut endoderm of the amphioxus late embryo (at 24 h) and in the posterior trunk of *C. intestinalis* (at 10–12 h). In amphioxus, expression of ALDH1d overlaps that of ALDH1a at the neurula stage, but diverges from that of ALDH1a in the late embryo. At this stage, amphioxus ALDH1d expression is broad and inconspicuous with a moderate accentuation of the signal in the posterior gut endoderm. In *C. intestinalis*, ALDH1d expression is diffuse and weak in the gastrula and becomes diffusely transcribed in the trunk at 10–12 h of development. Thus, in both amphioxus and *C. intestinalis*, ALDH1a is consistently expressed in patterns that are entirely consistent with a role in anteroposterior patterning and that are reminiscent of expression of vertebrate ALDH1A2 (RALDH2), which defines posterior identity in the vertebrate embryo (31).

In both amphioxus and *C. intestinalis*, there is only a single gene encoding a narrow-channeled ALDH2 enzyme. In amphioxus, ALDH2 expression is restricted to posterior, mesendodermal tissues at 12 h of development and, subsequently, spreads throughout the embryo at 24 h (Fig. 2). In *C. intestinalis*, ALDH2 expression is strong and diffuse at early and late developmental stages (Fig. S4). Thus, ALDH2 genes are expressed in widespread patterns during development of invertebrate chordates.

There are a total of six genes encoding ALDH1s with narrow channels in amphioxus and *C. intestinalis*: amphioxus ALDH1b, ALDH1c, ALDH1e, and ALDH1f and *C. intestinalis* ALDH1b and ALDH1c. In amphioxus, these duplicates are weakly expressed in posterior domains overlapping that of ALDH1a in the neurula (at 12 h) (Fig. 2). However, by 24 h, they are expressed diffusely and weakly throughout the amphioxus embryo with a weak to moderate concentration of the signal for ALDH1b, ALDH1c, and ALDH1e in the posterior gut endoderm. In *C. intestinalis*, ALDH1b is diffusely transcribed in trunk and tail, whereas ALDH1c is not detectable by in situ hybridization in developing embryos (Fig. S4). Thus, the genes encoding narrow-channeled ALDH1s in amphioxus and *C. intestinalis* generally display either widespread or inconspicuous expression patterns during development, suggesting that these enzymes are not playing major roles in anteroposterior patterning of the embryo.

Discussion

Structural Insights into ALDH1 and ALDH2 Function. Our modeling studies confirm the notion, determined by Moore et al. (17) with only two enzymes, that substrate access channel size is a crucial determinant of ALDH1 and ALDH2 function. Here, we extend this concept to the whole metazoan ALDH1/2 clade and provide insights into the structural adaptations underlying the ability of the narrow-channeled ALDH2s to process small aldehydes and of the large-channeled ALDH1s to process large, bulky aldehydes.

We determined that position 124 at the channel entrance is a selective gate for aldehyde size. Ancestral reconstructions show that, throughout metazoan ALDH1/2 evolution, increases in channel size are associated with substitution of the bulky, ancestral, Met124 by small Ala124 or Gly124 (Table S2). The selective abilities of ALDH1/2 channels are thus regulated by the presence or absence of a steric hindrance to the entry of large aldehydes into the ALDH1/2 channel. Thus, we show that the ALDH2 channel cannot accommodate retinaldehyde, which is consistent with earlier studies (16, 22) showing that retinaldehyde is not an ALDH2 substrate, but a competitive inhibitor of acetaldehyde degradation. This behavior contrasts with the ease with which retinaldehyde is admitted into the ALDH1 channel. Thus, size selection is a fundamental feature of substrate discrimination by the ALDH1/2 channel. We have also shown that, although ALDH2 can keep small aldehydes close to the catalytic Cys302 long enough for catalysis, ALDH1s cannot. Therefore, small aldehyde processing by ALDH2s requires specific structural adaptations to reduce substrate mobility inside their channels, an ability that is lacking in large-channeled ALDH1s (16).

Switches Between Small and Large Substrate Entry Channels Are Common in ALDH1/2 Evolution. The changes in the ALDH1 and ALDH2 SEC that we describe reflect a tendency of eukaryote ALDH1/2 genes to alter the structures of their encoded enzymes after gene duplication (Fig. 2 and Figs. S4, S5, and S7). By accumulating mutations at the mouth, neck, and/or SEC bottom, ALDH1/2s underwent changes in SEC geometry and/or overall volume, which affected their structural abilities to accommodate their original substrates. These changes led to switches from small, constricted channels adapted to the handling of small aldehydes to large, broadly opened channels adjusted to receive large aldehyde molecules or vice versa (Fig. S8). Our results thus provide an important contrast to studies proposing a general nonreversibility of amino acid changes involved in the functional adaptation of proteins (32).

ALDH1/2 Switches and the Origins of RA Signaling. Although ALDH enzymes can catalyze a range of different substrates, the molecular switches between ALDH1 and ALDH2 SECs reported here very likely represent functional transitions between the ability of ALDH1/2s to process small, toxic aldehydes for defense against endogenous and xenobiotic aldehyde aggression and its capacity

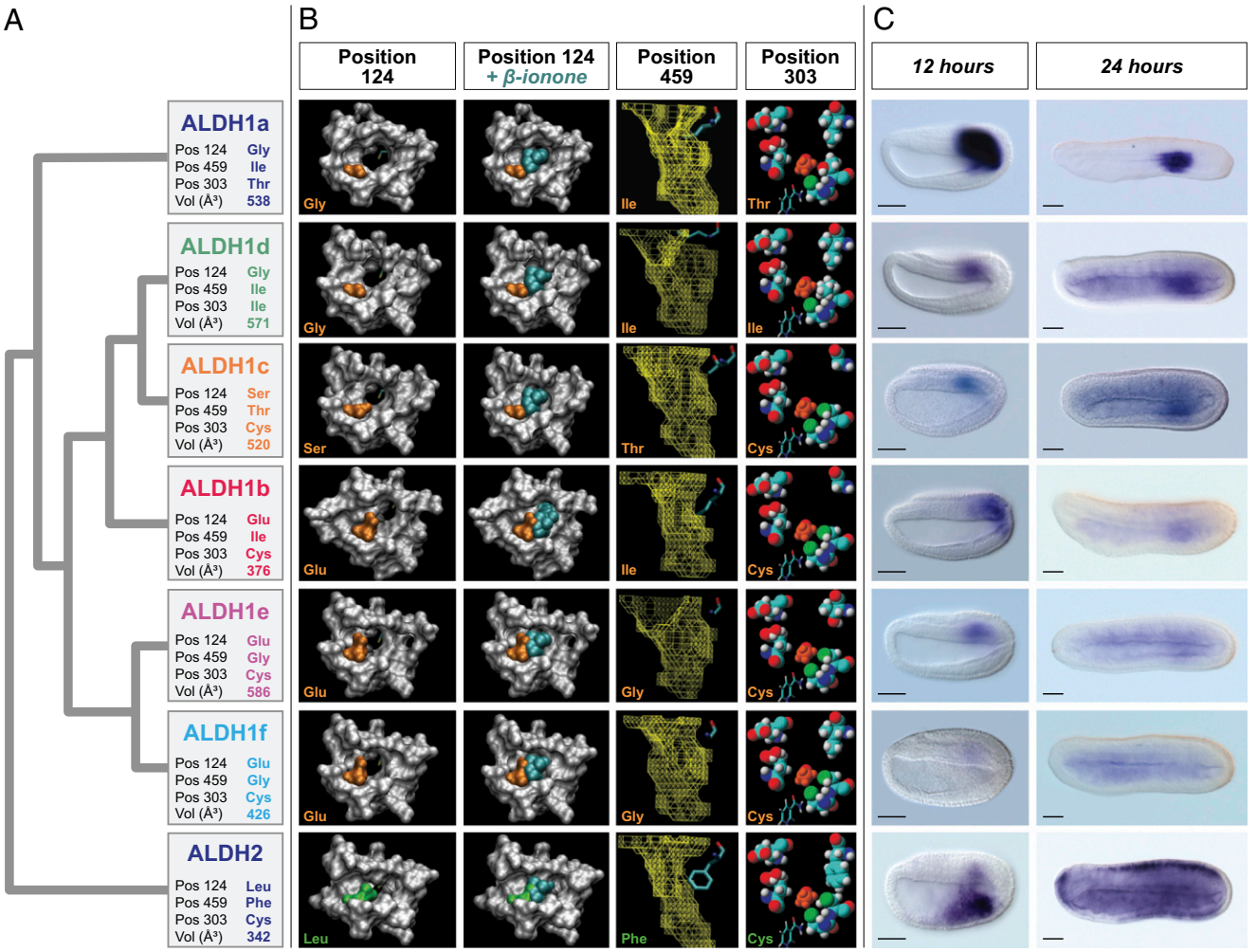

Fig. 2. Amphioxus ALDH1/2 duplicates. Phylogeny (*A*), channel structure (*B*), and developmental expression (*C*). Amino acid signatures of the substrate entry channel at positions 124 (the mouth), 459 (the neck), and 303 (the bottom) are indicated. For the expression analyses, neurulae (12 h) and late embryos (24 h) are shown. (Scale bars: 50 μm.)

to generate signaling molecules from larger aldehyde precursors. Using ancestral sequence reconstruction, we provide evidence that ALDH1/2 switches were important for the emergence of ALDH1 retinaldehyde dehydrogenases, which probably originated after gene duplication early in metazoan evolution, when a small, narrow-channeled ALDH1/2 ancestor, structurally related to modern ALDH2s, gave rise to a gene encoding a larger SEC capable of accommodating bulkier molecules, including retinaldehyde. This evolutionary scenario supports the view that RA signaling evolved from enzymes implicated in detoxification (3) and, combined with the description of a retinoic acid receptor (RAR) and of other RA signaling cascade members in both protostomes and deuterostomes, pushes the origins of RA signaling to much earlier times than traditionally assumed (4, 7).

ALDH1s Underwent Independent Duplication and Extensive Diversification in Metazoans. Our data substantiate the notion that the metazoan ALDH1 ancestor originated from a eukaryote ALDH1/2 ancestor and underwent duplication before the origin of bilaterian animals. It is also evident that ALDH1s duplicated independently in various animal groups and underwent extensive diversification, which is supported by two ALDH1 genes (one with a large SEC) in the cnidarian *Nematostella vectensis* and, in amphioxus and *C. intestinalis*, by structurally dissimilar ALDH1 ancestors and by the distinct arrangement of ALDH1 SEC signatures. Although duplication and diversification are common in ALDH1s, the metazoan ALDH2s are typically preserved as single copies, and their SECs have kept the same constricted features of the eukaryote ALDH1/2 ancestor.

Invertebrate ALDH1 Switches Suggest Shifts from Patterning to Detoxification. The presence of ALDH1 duplicates in a given animal raises questions about the roles of each duplicate within the RA signaling cascade (4, 7). ALDH1 duplicates in amphioxus and *C. intestinalis* originated by duplication from an ALDH1 ancestor with a large SEC structurally compatible with RA synthesis (Fig. S6). This structure is present in their ALDH1a paralogs, which display sharp posterior domains, consistent with early embryonic anteroposterior patterning. In contrast, genes encoding ALDH1b, ALDH1c, ALDH1e, and ALDH1f in amphioxus and ALDH1b and ALDH1c in *C. intestinalis* accumulated mutations resulting in constricted ALDH SECs poorly suited to accommodate large molecules, but still capable of admitting small aldehydes. These genes display broad expression patterns, suggesting that they have evolved novel functions probably associated with the processing of small, toxic aldehydes for protection against endogenous or xenobiotic aldehydes (1–3).

A plausible scenario leading from patterning to protective ALDH roles can be derived from the expression patterns of the ALDH1d genes of amphioxus and *C. intestinalis*. The molecular structures of the SECs of these two ALDH1ds are consistent with retinaldehyde processing. However, ALDH1d expression is rather

broad throughout the amphioxus embryo and the *C. intestinalis* embryonic trunk, suggesting that changes in gene regulation of these two genes occurred after duplication and that these changes were not accompanied by structural remodeling of the SEC.

The transition from restricted signaling functions to generalized roles has not been completed to equivalent degrees in each of the divergent amphioxus duplicates. ALDH1b, ALDH1c, ALDH1e, and ALDH1f developed diffuse patterns in the late embryo, while curiously maintaining weak, but restricted, posterior domains during neurulation and, except for Aldh1f, a relative concentration of expression in the posterior gut endoderm, which are likely to represent the ancestral, ALDH1a-like pattern. In *C. intestinalis*, ALDH1b is diffusely and inconspicuously expressed in the trunk, whereas ALDH1c expression is not detectable during embryogenesis, but seems to be restricted to adult tissues, as indicated by EST database searches.

The fate of these ALDH1 duplicates in amphioxus and *C. intestinalis* is consistent with an evolutionary scenario involving neofunctionalization after duplication, with gene duplicates acquiring more generalized functions during embryogenesis and, possibly, in the adult. Therefore, in amphioxus and *C. intestinalis*, some duplicated ALDH1s experienced modifications of gene regulation and protein structure that resulted in neofunctionalization of the duplicates, possibly away from roles in axial patterning, toward generalized, pleiotropic functions similar to those of ALDH2, an enzyme important for protection against small aldehyde toxicity in chordates (33).

The ALDH1/2 Case and Its Implications for Anatomic and Physiological Evolution. Mutations of regulatory regions have been regarded as the major driving force of morphological evolution in development (34), whereas mutations in coding regions are viewed as major determinants of physiological evolution (35). Here, we describe regulatory alterations affecting duplicated ALDH1/2 genes of amphioxus and ascidian tunicates that are accompanied by fundamental structural shifts of the proteins they encode. Therefore, our data suggest that distinctions between anatomical and physiological evolution may not always be so clear cut and that rapid evolution of novel functions can be achieved when regulatory and protein structure mutations are superimposed after gene duplication, a hypothesis that provides a common ground for these two evolutionary mechanisms that have traditionally been thought to depend on distinct mechanisms. In sum, the ALDH1/2 case probably represents one of many examples that are likely to emerge with the incorporation of protein structure analyses into the collection of approaches used to study the evolution of body plans.

Materials and Methods

Whole genomes, EST databases, and trace repositories were mined for ALDH sequences using both signature (InterPro IPR002086) and global similarity searches. Amphioxus and *C. intestinalis* ALDH1/2 clones were obtained, respectively, from cDNA libraries (36) and from the Gene Collection Release 1 (37). ALDH amino acid residue numbers are based on the classical numbering of the mature human ALDH2 enzyme with the catalytic Cys at position 302 (17).

For additional details, see *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Gérard Benoit, Tiago Pereira, Linda Z. Holland, and Nicholas D. Holland for critical reading of the manuscript. We are indebted to the Faculty of Medicine of the University of São Paulo for access to its high-performance computing cluster. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo Grant 06/50843-0 (to J.X.-N.), by funds from Agence Nationale de Recherche (ANR-07-BLAN-0038 and ANR-09-BLAN-0262-02), Centre National de la Recherche Scientifique, and Ministere de l'Education Nationale de la Recherche et de Technologie (to M.S.), and by the Consortium for Research into Nuclear Receptors in Development and Aging (CRESCENDO), a European Union Integrated Project of FP6. M.S.-C. was supported by a travel fellowship from the Company of Biologists.

1. Seiler N, Eichentopf B (1975) 4-aminobutyrate in mammalian putrescine catabolism. *Biochem J* 152:201–210.
2. Kikonyogo A, Pietruszko R (1996) Aldehyde dehydrogenase from adult human brain that dehydrogenates gamma-aminobutyraldehyde: Purification, characterization, cloning and distribution. *Biochem J* 316:317–324.
3. Yoshida A, Rzhetsky A, Hsu LC, Chang C (1998) Human aldehyde dehydrogenase gene family. *Eur J Biochem* 251:549–557.
4. Campo-Paysaa F, Marlétaz F, Laudet V, Schubert M (2008) Retinoic acid signaling in development: Tissue-specific functions and evolutionary origins. *Genesis* 46:640–656.
5. Marlétaz F, Holland LZ, Laudet V, Schubert M (2006) Retinoic acid signaling and the evolution of chordates. *Int J Biol Sci* 2:38–47.
6. Duester G (2008) Retinoic acid synthesis and signaling during early organogenesis. *Cell* 134:921–931.
7. Simões-Costa MS, Azambuja AP, Xavier-Neto J (2008) The search for non-chordate retinoic acid signaling: Lessons from chordates. *J Exp Zoolog B Mol Dev Evol* 310: 54–72.
8. Theodosiou M, Laudet V, Schubert M (2010) From carrot to clinic: An overview of the retinoic acid signaling pathway. *Cell Mol Life Sci* 67:1423–1445.
9. Lamb AL, Newcomer ME (1999) The structure of retinal dehydrogenase type II at 2.7 Å resolution: Implications for retinal specificity. *Biochemistry* 38:6003–6011.
10. Dräger UC, Wagner E, McCaffery P (1998) Aldehyde dehydrogenases in the generation of retinoic acid in the developing vertebrate: A central role of the eye. *J Nutr* 128(Suppl):4635–4665.
11. Graham CE, Brocklehurst K, Pickersgill RW, Warren MJ (2006) Characterization of retinaldehyde dehydrogenase 3. *Biochem J* 394:67–75.
12. Lin M, Zhang M, Abraham M, Smith SM, Napoli JL (2003) Mouse retinal dehydrogenase 4 (RALDH4), molecular cloning, cellular expression, and activity in 9-*cis*-retinoic acid biosynthesis in intact cells. *J Biol Chem* 278:9856–9861.
13. Niederreither K, Subbarayan V, Dollé P, Chambon P (1999) Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* 21:444–448.
14. Wang X, Penzes P, Napoli JL (1996) Cloning of a cDNA encoding an aldehyde dehydrogenase and its expression in *Escherichia coli*. Recognition of retinal as substrate. *J Biol Chem* 271:16288–16293.
15. Zhao D, et al. (1996) Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase. *Eur J Biochem* 240:15–22.
16. Klyosov AA (1996) Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes. *Biochemistry* 35:4457–4467.
17. Moore SA, et al. (1998) Sheep liver cytosolic aldehyde dehydrogenase: The structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* 6:1541–1551.
18. Steinmetz CG, Xie P, Weiner H, Hurley TD (1997) Structure of mitochondrial aldehyde dehydrogenase: The genetic component of ethanol aversion. *Structure* 5:701–711.
19. Yoshida A, Hsu LC, Davé V (1992) Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. *Enzyme* 46:239–244.
20. Yoshida A, Hsu LC, Yanagawa Y (1993) Biological role of human cytosolic aldehyde dehydrogenase 1: Hormonal response, retinal oxidation and implication in testicular feminization. *Adv Exp Med Biol* 328:37–44.
21. Duester G (2001) Genetic dissection of retinoid dehydrogenases. *Chem Biol Interact* 130-132:469–480.
22. Klyosov AA, Rashkovetsky LG, Tahir MK, Keung WM (1996) Possible role of liver cytosolic and mitochondrial aldehyde dehydrogenases in acetaldehyde metabolism. *Biochemistry* 35:4445–4456.
23. Holland LZ, et al. (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res* 18:1100–1111.
24. Cañestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932–942.
25. Schubert M, Escriva H, Xavier-Neto J, Laudet V (2006) Amphioxus and tunicates as evolutionary model systems. *Trends Ecol Evol* 21:269–277.
26. Putnam NH, et al. (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
27. Cañestro C, Postlethwait JH, Gonzàlez-Duarte R, Albalat R (2006) Is retinoic acid genetic machinery a chordate innovation? *Evol Dev* 8:394–406.
28. Skibbe DS, et al. (2002) Characterization of the aldehyde dehydrogenase gene families of Zea mays and Arabidopsis. *Plant Mol Biol* 48:751–764.
29. Perozich J, Nicholas H, Wang BC, Lindahl R, Hempel J (1999) Relationships within the aldehyde dehydrogenase extended family. *Protein Sci* 8:137–146.
30. Fujiwara S, Kawamura K (2003) Acquisition of retinoic acid signaling pathway and innovation of the chordate body plan. *Zoolog Sci* 20:809–818.
31. Niederreither K, McCaffery P, Dräger UC, Chambon P, Dollé P (1997) Restricted expression and retinoic acid-induced downregulation of the retinaldehyde dehydrogenase type 2 (RALDH-2) gene during mouse development. *Mech Dev* 62:67–78.
32. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461:515–519.
33. Chen C-H, et al. (2008) Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* 321:1493–1495.
34. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134:25–36.
35. Hoekstra HE, Coyne JA (2007) The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* 61:995–1016.
36. Yu J-K, et al. (2007) Axial patterning in cephalochordates and the evolution of the organizer. *Nature* 445:613–617.
37. Satou Y, et al. (2002) A cDNA resource from the basal chordate *Ciona intestinalis*. *Genesis* 33:153–154.