Chapter 4

The evolution of the retinol metabolism

Abstract

Introduction

The retinol metabolism comprises a series of enzymatic reactions that convert dietary vitamin A (retinol) into various bioactive compounds, primarily retinal for vision (REF) and retinoic acid for gene regulation (REF), ensuring the proper functioning of visual processes and other physiological roles in the body (REFS).

Retinol (Vitamin A1) is an essential micronutrient derived primarily from diet. It can be obtained directly from animal sources as retinyl esters or indirectly from plant sources as pro-vitamin A carotenoids, which are then converted into retinol in the body (REFS). Once in the cell, retinol is esterified to retinyl ester by the enzyme lecithin retinol acyltransferase (LRAT) (REF). When needed, retinyl ester is hydrolysed back to retinol (REF). Retinol is oxidized to retinal by retinol dehydrogenases (RDHs). Several other enzymes are involved in various steps of the retinol metabolism pathway as schematically shown in Figure 4.1 that summarizes what is known about the pathway according to the KEGG Pathway Database (Kanehisa et al. 2021). In addition, Table 4.1 provides a comprehensive list of these enzymes ranked by the number of pathways they participate in according to KEGG. Involvement in one or few pathways serves as an indicator of enzyme specificity to the retinol metabolism, as opposed to broad spectrum enzymes.

Retinal, particularly 11-cis-retinal, plays a crucial role in vision (REFS). 11-cis-retinal binds to the protein opsin in photoreceptor cells forming rhodopsin. Upon absorbing a photon, 11-cis-retinal is isomerized to all-trans-retinal, leading to a conformational change in opsin, and initiating a cascade of events called phototransduction (REFS) (see Chapter 3). After light exposure, all-trans-retinal is reduced to all-trans-retinol and then converted back to 11-cis-retinal through a series of enzymatic reactions. This part of the visual cycle is essential as it ensures the retina’s responsiveness to light (REFS). The regulation of the metabolic steps ensures sufficient 11-cis-retinal availability and prevents toxic build-up of intermediates. Additionally, retinal can be further oxidized to retinoic acid by retinaldehyde dehydrogenases (RALDHs). Retinoic acid serves as a signalling molecule that regulates gene expression and is critical for numerous developmental processes (REFS).

Retinol metabolism has been described mainly in human and ?. While little is known about its presence and composition in other organisms ?..

Given the importance of the retinol metabolism, it is compelling to delve into its evolutionary history, especially when considering the broader evolution of vision. Hence, the work presented in this chapter aimed to unravel this intricate history. The initial step was to identify the genetic components involved and determine their evolutionary relationships to answer questions such as: Do the gene families belong to overarching orthogroups? How closely related are they? The subsequent objective was to uncover the distribution of these components across the animal kingdom and, more broadly, within eukaryotes, to pinpoint the specific point in time when all the components came into place. The final endeavour was to delineate the main evolutionary events characterizing each orthogroup, to discern, for instance, if certain gene families have undergone a greater number of evolutionary events and contextualizing them within the evolutionary tree of life.

Results and Discussion

Enzymes involved in retinol metabolism belong to 12 major orthogroups.

To understand the evolution of the retinol metabolism, I decided to reconstruct the evolution of all the enzymes involved in the pathway, as described by KEGG (Kanehisa et al. 2021) (Figure 4.1 and Table 4.1). To do this, I explored the genes encoding these enzymes in 101 species spanning all of Eukarya (Table 4.2 and its supp version with more info).

The first step to study the evolution of these genes was to first determine to which gene families or orthogroups they belonged to. In fact, while the number of enzymes participating in the pathway is relatively high, some of them might belong to a broader gene family, or to put it more precisely, orthogroup, i.e., a group of orthologs and paralogs deriving from the same original gene duplication.

Therefore, the first part of the analyses aimed to identify the orthogroups that the enzymes belong to. For this, a preliminary BLASTP (REF) was performed using sequences from the KEGG Orthology lists (Kanehisa 2019) as queries versus our database of 101 eukaryotes. A mention of the fact that the RPH component doesn’t have any seqs on kegg and therefore was discarded from further search until future work can provide at least some preliminary seqs to work with. The results of this blast were used as input for orthogroup identification pipeline. The details can be found in the Methods section, but briefly, two alternative software (Broccoli (REF) and Orthofinder (REF)) were used to independently assess orthogroups, then the results were compared, and consensus groups were defined. Briefly mention key differences of methods?

The results of the orthogroup identifications and the comparison between the two methods is shown in Figure 4.2. First of all, we can see how the two methods are largely consistent, with many cases of one-to-one correspondence of orthogroups. However, it is also immediately noticeable that Orthofinder tended to provide fewer and larger orthogroups, while Broccoli provided more and in some cases smaller orthogroups. As a consequence, some gene families appeared fragmented into multiple smaller orthogroups according to Broccoli only. Quickly mention that we found a GK group (or subgroup) but this gene has nothing to do with retinol metabolism so excluded from further research. Table 4.3 summarizes the final 12 orthogroups identified and shows the comparison of Orthofinder and Broccoli with each other and the original KEGG groups.

Overall we identified some interesting and unexpected findings: such as that DGAT and DGAT2L4 are not to be considered the same gene family/orthogroup (according to both Orthofinder and Broccoli) and that the SDR and RDH families are intermingled, possibly indicating that they belong to a broader orthogroup. For the latter, to discriminate more rigorously the relationship between SDR and RDH, all orthougroups were collected as one big orthogroup for phylogenetic analysis.

Possvm and Generax describe substructure of orthogroups.

Aim of this section.

Differences between two methods. What does possvm do, what does generx do. What are pros and cons of each. (Does possvm output a rooted or unrooted tree). Possvm is fast, easy to run (from laptop?) and easy to annotate.. generax is slow but very precise (?)

“As the species overlap algorithm relies on the implicit taxonomic information contained in the gene tree’s topology, this approach is suitable for cases where the species tree is unknown or unavailable.” – Possvm paper. 😊

**RETSAT**

The Retinol Saturase (RETSAT) enzyme catalyses the reaction that saturates the 13-14 double bond of all-trans-retinol to produce all-trans-13,14-dihdriretinol (Moise et al. 2004). This enzyme appears to be involved only in retinol metabolism (Table 4.1) according to the KEGG Database (REF), meaning it is very specific to this pathway.

The orthogroups identified for RETSAT by OrthoFinder and Broccoli present a clear one-to-one relationship with high degree of identity (Figure 4.2), indicating no mixture with any other orthogroup examined. The consensus RETSAT orthogroup contained 338 sequences distributed throughout all major eukaryotic clades (Figure 4.3A).

Ortholog sorting with Possvm identified 7 orthogroups within the RETSAT family, with one orthogroup containing RETSAT together with the related enzyme PYRD2 as well as CTR enzymes (Figure 4.3B). Gene tree to species tree reconciliation with GeneRax revealed a high number of events (especially losses) in proportion to the size of the orthogroup (Figure 4.3C). More details about tree topologies?

Check out Weber et al 2020 for further discussion?

**PNPLA4**

The Patatin Like Phospholipase Domain Containing 4 (PNPLA4) enzyme plays a role in the hydrolysis of retinyl esters to retinol (Schreiber et al. 2012). It is involved in one other pathway according to KEGG (Table 4.1).

Both OrthoFinder and Broccoli identify one distinct orthogroup for PNPLA4 independent from all other orthogroups (Figure 4.2). The consensus orthogroup contains 215 sequences. While being present in both major eukaryotic clades, this orthogroup appears to be missing in basal groups of Amorphea, such as the Holomycota branch that includes Fungi (Figure 4.4A).

Possvm identified 9 orthogroups within this family. PNPLA1-5 belonged to the same orthogroup, with PNPLA4 being sister group to the other genes (Figure 4.4B). The GeneRax reconciled tree recovered the same topology and identified a moderate number of events (Figure 4.4C). The tight relationship between PNPLA4 and other PNPLA genes is in accordance with evidence suggesting that some of them are also involved in retinol metabolism (REF).

**ALDH1**

**BCMO1/RPE65**

BCMO1 (also known as BCO1)...

**LRAT**

**RDH/DHRS**

Even when this was one of the largest orthogroups (~5000 sequences) with by far the most complex substructure, it may very well be that the mega orthogroup could be expanded even more to include other enzymes. The possible incompleteness may stem from the fact that our original seeds for the blast came only from enzymes used by the retinol metabolism (focus of this chapter) however other enzymes are involved in other pathways. Future research that aims to understand the details of the evolution of RDHs and SDRs should broaden even more the scope of investigation, even if this will be extremely challenging due to the difficulties in performing phylogenetic analysis with such large datasets.

Although the genes involved in retinol metabolism may be considered specific according to the number of kegg pathways it is involved in, the broad orthogroup is very broad definetly involved in numerous physiological processes throughout eukarya.

Fig is collapsed in such a way to keep one clade per possvm orthogroup.

**DGAT1**

**DGAT2LA4**

Also known as AWAT2...

...Belongs to a broader OG that includes DGAT2. But is clearly distinct from DGAT1 (see above).

**CYP**

Largest, but more compact groups compared to RDH/DHRS.

**AOX**

**ADH**

**UGT**

Conclusions

Summary of results. Answer questions posed in intro! Example of relationships amongs groups: DGAT1 vs DGAT2L4; closeness between RDH+DHRS and some components of ADH group (see fig 4.2)...

Ctenophores are the group with least OGs; could either reflect weird biology or missing data due to incomplete genomes...

Evolution:

Origins: The visual pigments, opsins, and the associated visual cycle have ancient origins. It's believed that the basic phototransduction machinery was present in the common ancestor of most extant animals.

Opsin Diversification: Over evolutionary time, different opsin classes evolved, allowing animals to detect light across different wavelengths. This diversification is tightly linked with changes in retinoid usage and the visual cycle.

Vertebrate Adaptations: The transition from aquatic to terrestrial environments necessitated changes in the visual system, including modifications in the retinoid pathway.

Retinoid Binding Proteins: Evolutionary changes in retinoid-binding proteins and enzymes have fine-tuned retinoid transport, storage, and metabolism, allowing diverse organisms to adapt to their specific visual environments.

Methods

Identification of orthogroups for retinol metabolism enzymes.

**Species list (and species tree)**

**Data mining**

- Collection of kegg “orthologs” for each components as queries

- preliminary loose blastp

**Orthogroup definition**

- with orthofinder

- with broccoli

- comparison of the two, visualisation with Cytoscape

- final definition of consensus orthogroups

Reconstruction of evolutionary history of orthogroups.

**Phylogenetic trees**

- alignments, trimmings

- trees

- ETE (only for generax = necessary for some of the subsequent steps)

**Reconstruction substructure/orthologs with Possvm**

**Reconstruction of evolutionary events with Generax**

Data Availability

Acknowledgements

For this chapter I would like to thank Riccardo Kyriacou who during his time as a summer intern assisted me in optimising Broccoli parameters for orthogroup detection and then comparing Broccoli orthogroups with those from Orthofinder using Cytoscape. My thanks also go to Julien Devilliers for his invaluable coding assistance, which facilitated the automation of various steps within this chapter.

References

Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Science* 28:1947–1951.

Kanehisa M, Sato Y, Kawashima M. 2021. KEGG mapping tools for uncovering hidden features in biological data. *Protein Science* [Internet] n/a. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4172

Moise AR, Kuksa V, Imanishi Y, Palczewski K. 2004. Identification of All-trans-Retinol:All-trans-13,14-dihydroretinol Saturase \*. *Journal of Biological Chemistry* 279:50230–50242.

Schreiber R, Taschler U, Preiss-Landl K, Wongsiriroj N, Zimmermann R, Lass A. 2012. Retinyl ester hydrolases and their roles in vitamin A homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1821:113–123.