**EXPERIMENT -3**

**Aim-** To explore KEGG database through the following exercise.

**Theory-** KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

**How to find a standard metabolic pathway in KEGG ?** 

* From the KEGG [table of contents](http://www.genome.ad.jp/kegg/kegg2.html) click on the link '[Metabolic pathways](http://www.genome.ad.jp/kegg/metabolism.html)' under pathway category.
* A list with all pathways will be available.
* To find the pathway link for **'Lysine biosynthesis'**, scroll down on the browser to the group of pathways called 'amino acid metabolism' and click on the link.

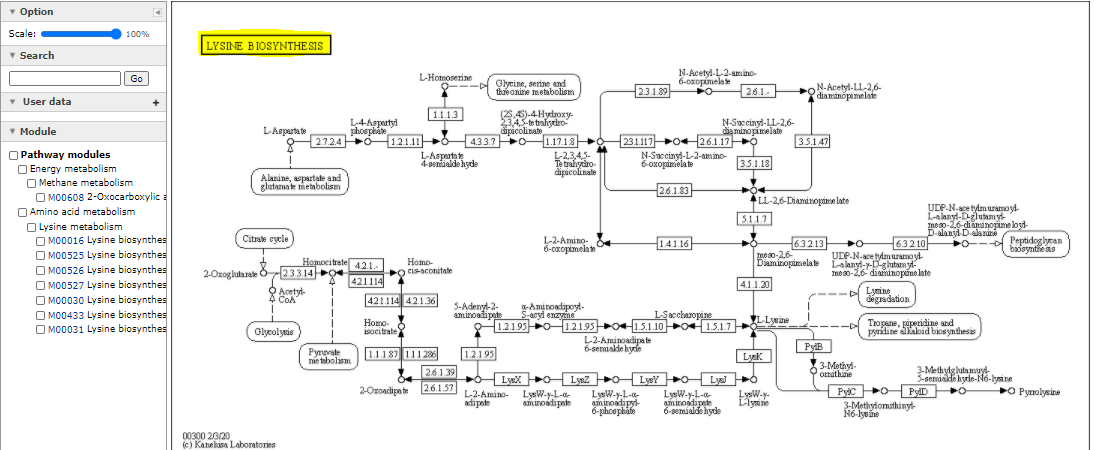


**Fig 1.0**

**Scroll down to the portion where heading 1.5 is given or press ctrl+f and then type in LYSINE BIOSYNTHESIS**

**Visualizing the standard map after clicking on the link**

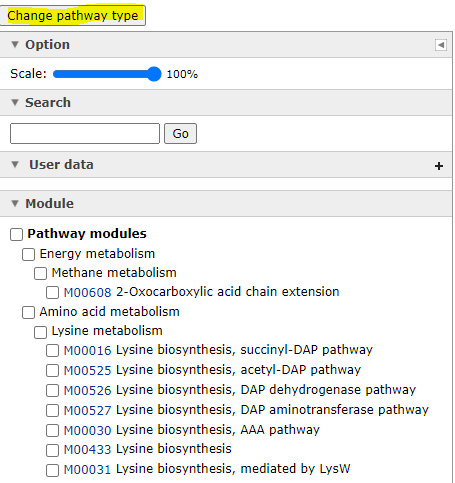
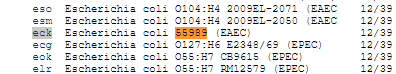
You should now see the standard pathway [MAP00300](http://www.genome.ad.jp/kegg/pathway/map/map00300.html) for lysine biosynthesis.

****

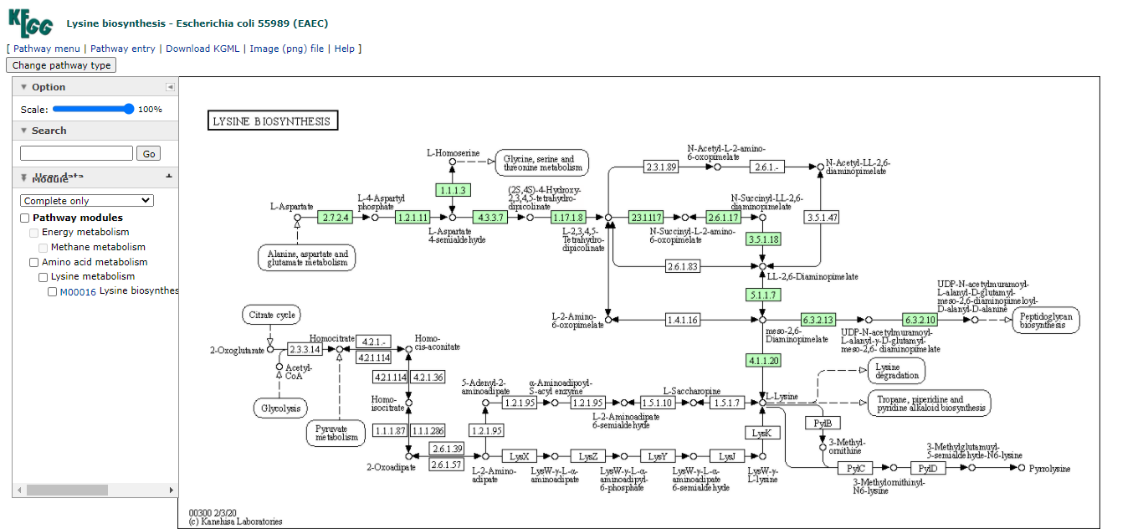
**Fig 2.0**

**1.How to find a species specific pathway map?**

**Go to change pathway type search the organism or the species you want to select (ctrl+f ) and then click the link left hand side of it .**

 ****

**Fig 3.0 Fig 4.0**

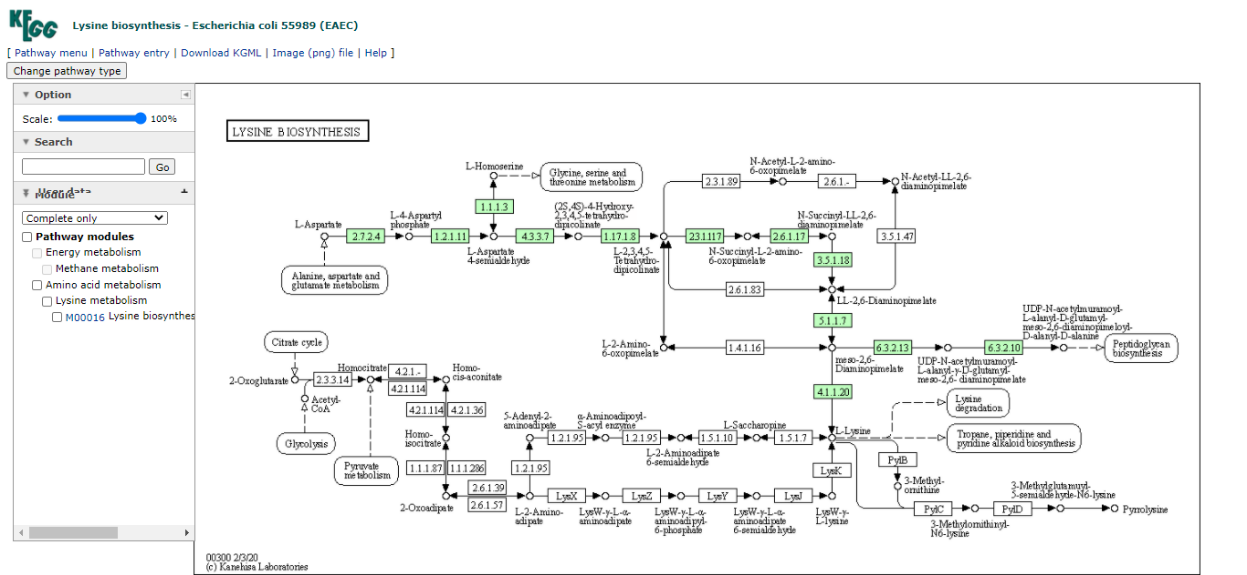


**Fig 5.0**

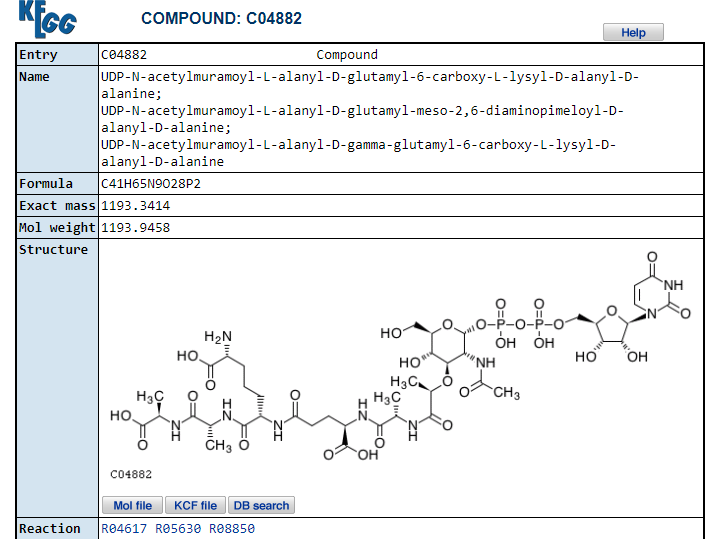
* All known E.coli enzymes for which a database entry exists are labelled in green. Not all enzymes are coloured which indicates that the corresponding enzyme (pathway) does not exist in this organism or that no information is found in any of the public databases (Swiss Prot, GenBank, Protein Data Base etc.).
* Since the [complete genome](http://www.genome.ad.jp/dbget-bin/get_htext?E.coli.kegg+B) for E.coli has been sequenced this very likely indicates that complete information is available about this Gram-negative bacteria. However, about 39% of all predicted genes (called ORFs or open reading frames) are not associated with any known protein, biochemical or physiological activity and therefore not all pathways are known and not described in KEGG.
* Some organisms such as mouse and human are not listed for lysine biosynthesis but show a map for [glutamate](http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00251), for example. **L-lysine, unlike glutamate, is an essential amino acid for both human and mouse. L-lysine has to be part of our diet, because we lack the necessary enzymes for its biosynthesis!!!**

1. **How to find a chemical structure and metabolite information from a pathway map?**

* Note that the immediate precursor of L-lysine also serves as a substrate for peptidoglycan synthesis to form the activated polypeptide **UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelotl-D-alanyl-D-alanine** the structure of which can be found by clicking on the small circle next to the compound name.

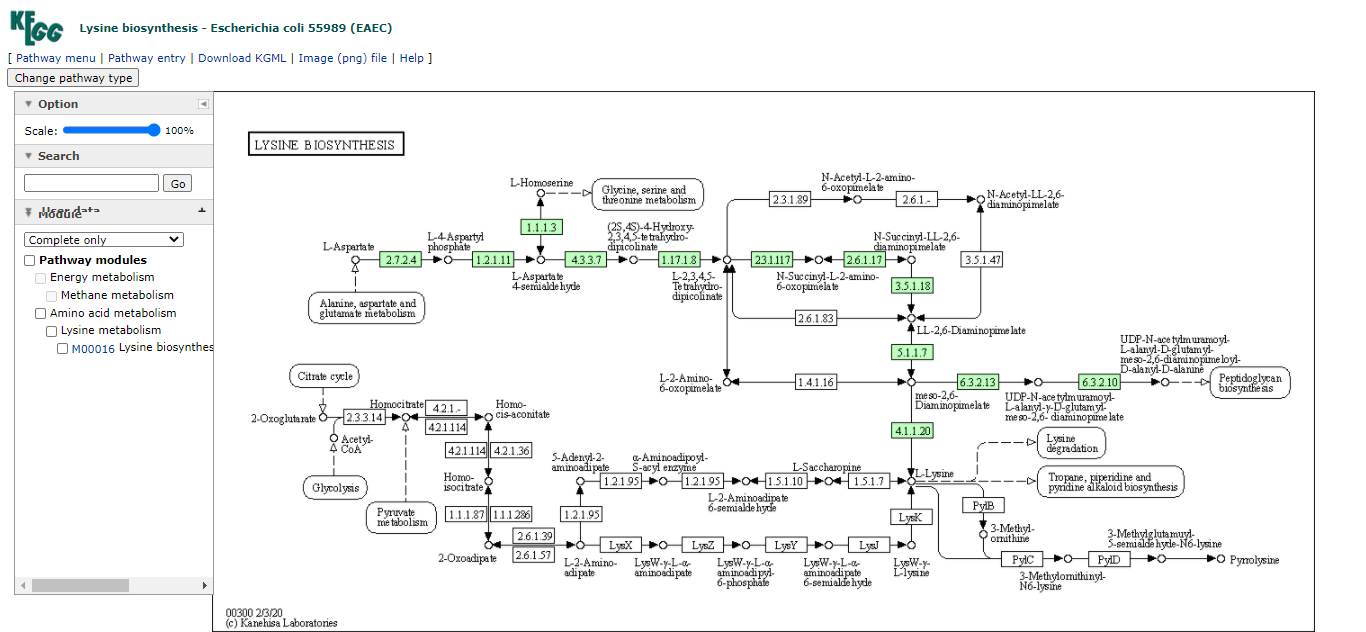
****

**Fig 6.0**

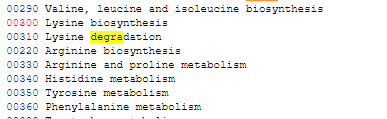


**Fig 7.0**

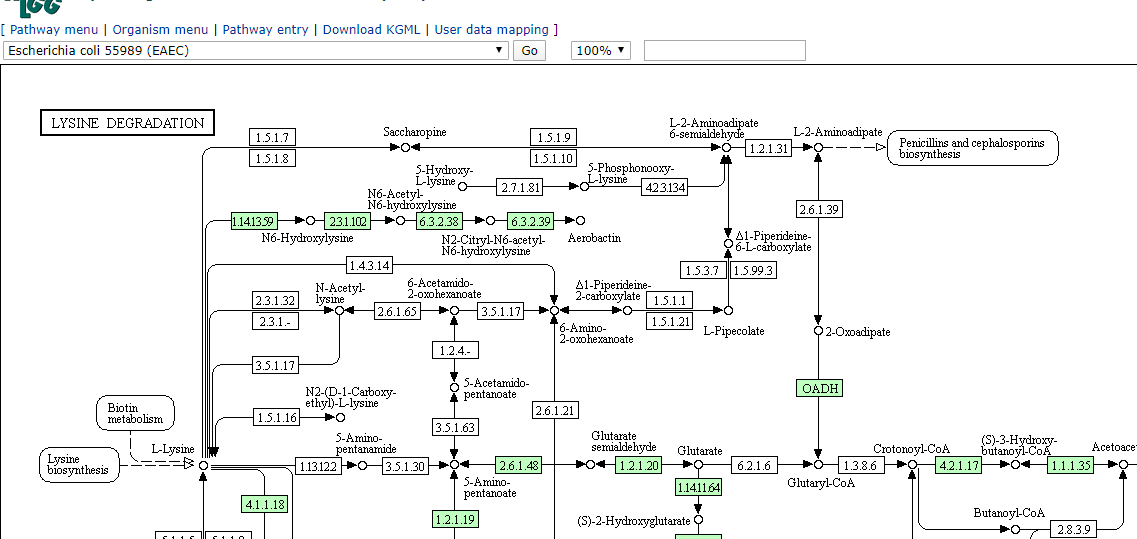
* You will see an information page for KEGG entry [C04882](http://www.genome.ad.jp/dbget-bin/www_bget?cpd:C04882).
* The preceding problem gives an example of how to find information about a metabolite, namely its chemical structure, formula, KEGG entry number, and pathway map numbers for which it is an intermediate.
* Clicking on a substrate name (or its circle) on a pathway map page is the easiest way to find relevant chemical information about a substrate and its shared pathways.
* The same can be done for an enzyme by clicking on the E.C. number box or for an intersecting pathway indicated in a roundish shaped box.
* For example, a link exists from the lysine biosynthesis map to the 'lysine degradation' pathway map. Clicking on the box marked '[Lysine degradation](http://www.genome.ad.jp/kegg/pathway/map/map00310.html)' will bring us to the corresponding catabolic processes. Note that the species selection will not change (we last selected *E.coli* pathways above).



**Fig 8.0**

****

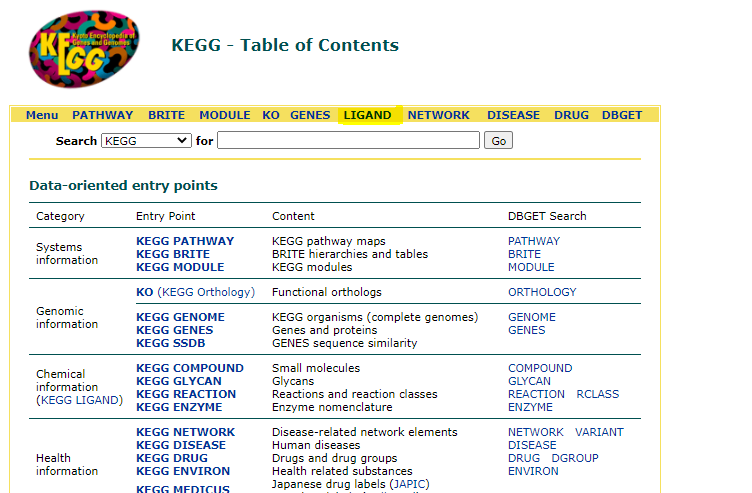
**Fig 9.0**



**Fig 10.0**

**1: How do I find a chemical structure, metabolite information by keyword search**?

* To find a pathway metabolite or enzyme, the table of contents offers a direct link to the DBGET Ligand database at KEGG.
* This search mode can be found at the 'table of contents' page under the 'enzyme' category, DBGET search.
* Click on the link called '[Ligand](http://www.genome.ad.jp/dbget-bin/www_bfind?ligand)' to access a generic search mode that allows a keyword entry.

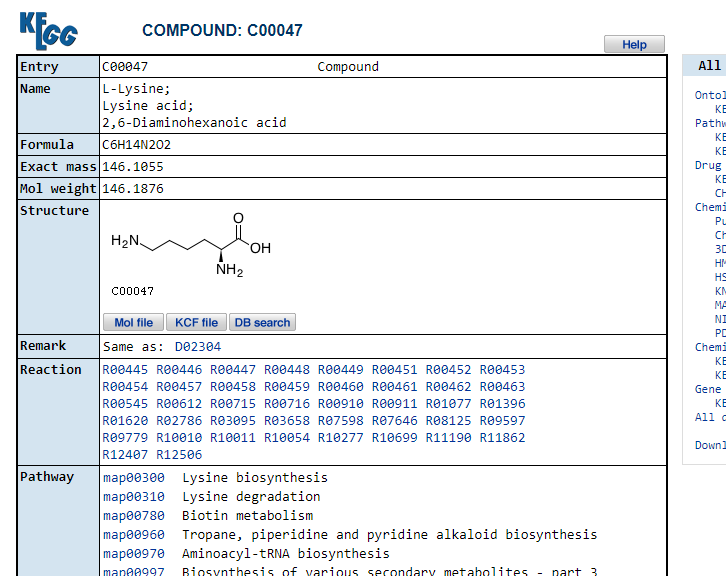
****

**Fig 11.0**



**Fig 12.0**

* Note that an exact enzyme number or compound number is not necessary in the DBGET database.
* To find information about lysine or L-lysine, type in 'lysine' and hit the return (enter) key. You will receive a return list with [159 hits](http://www.genome.ad.jp/dbget-bin/www_bfind_sub?dbkey=ligand&keywords=lysine&mode=bfind&max_hit=1000). The search returned all KEGG entries that contained the word 'lysine' anywhere in the enzyme or compound name.
* The list contains 45 enzyme links (ec: x.x.x.xx) and 51 compound links (cpd: Cxxxxx), one being L-lysine ([cpd:C00047](http://www.genome.ad.jp/dbget-bin/www_bget?cpd:C00047)) and all others derivatives thereof.



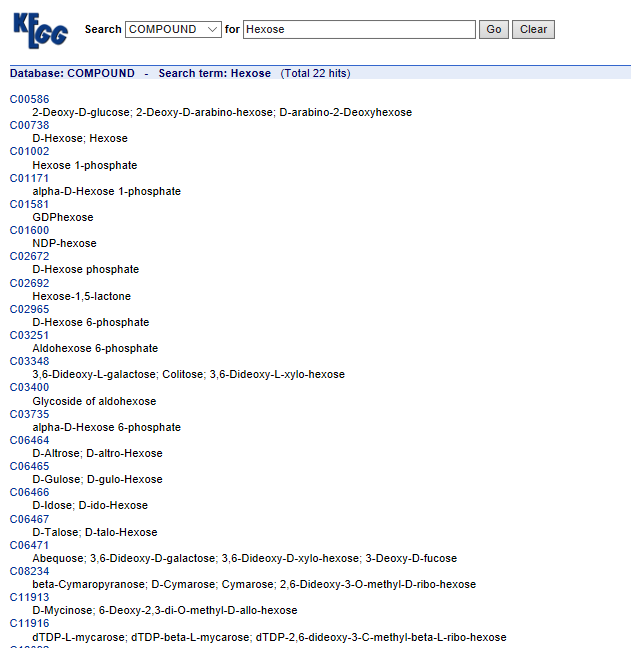
**Fig 13.0**

* Clicking on the cpd number will bring you to the chemical structure information sheet. This sheet lists compound entry number for L-lysine (note that D-lysine has a different entry, of course), common name(s), formula, structure, all pathway maps that contain L-lysine as metabolite (5 maps for L-lysine including synthesis and degradation, biotin metabolism, alkaloid biosynthesis II, and Aminoacyl-tRNA biosynthesis), and finally a list of all known enzymes that use L-lysine as a substrate.

**General information on biological molecules**

* One additional feature that is very helpful to use are the molecular catalogue entries, specifically the '[compound classification](http://www.genome.ad.jp/kegg/catalog/compounds.html)'.
* This link leads to a catalogue of metabolites classified according to their functional class, e.g., carbohydrates, fatty acids, phospholipids, neurotransmitters etc..

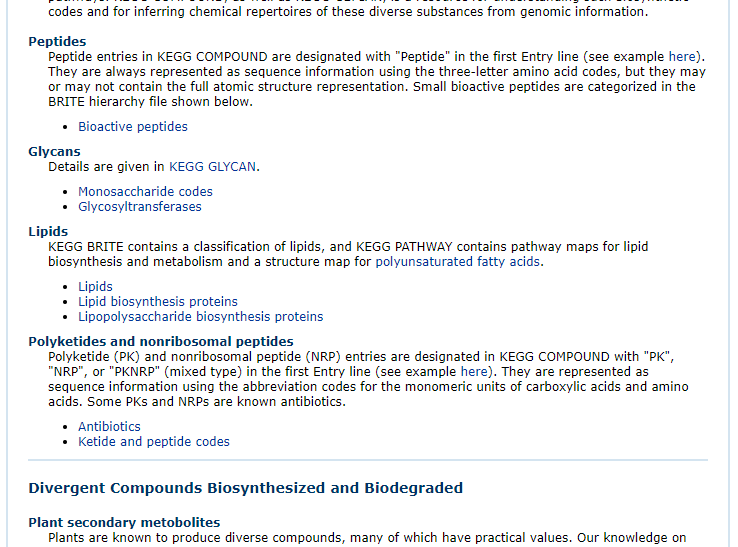
If you want to look up the structures of a class of molecules like the amino acids or various hexoses, this link will give you the best and broadest result to quickly find what you need.



**Fig 14.0**

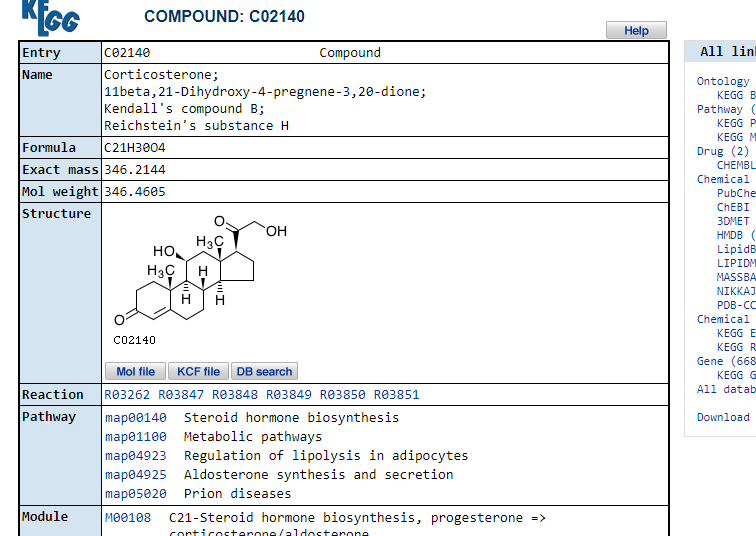
**Can be accessed from Table Of Contents**

* Use this link as a reference for structure information. As an example we are interested in the general structure of [steroid hormones](http://www.genome.ad.jp/kegg/catalog/cpd_steroid_hormone.html).
* A link provided in the category 'Lipids' provides a page containing the names and chemical structures of seven cholesterol derived steroid hormones.



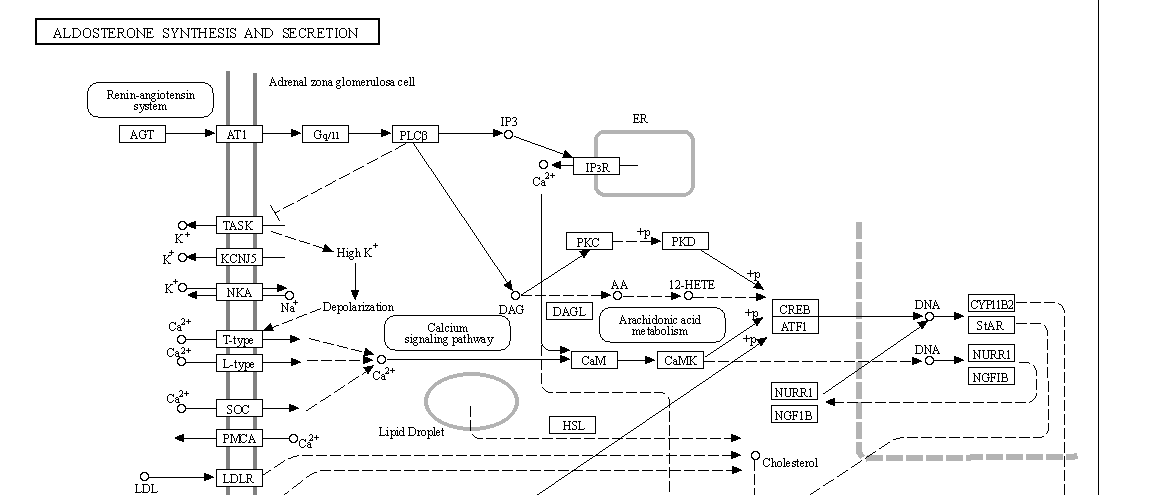
**Fig 15.0**

* Clicking on the name link '[aldosterone](http://www.genome.ad.jp/dbget-bin/www_bget?cpd:C01780)' connects to a structure information page providing a link to the pathway map for C21 steroid hormone metabolism ([MAP00140](http://www.genome.ad.jp/dbget-bin/show_pathway?MAP00140+C01780)).



**Fig 16.0**

* Following the pathway map link results in the standard pathway for steroid hormone metabolism with aldosterone position marked as a red circle (because we started our search from aldosterone).

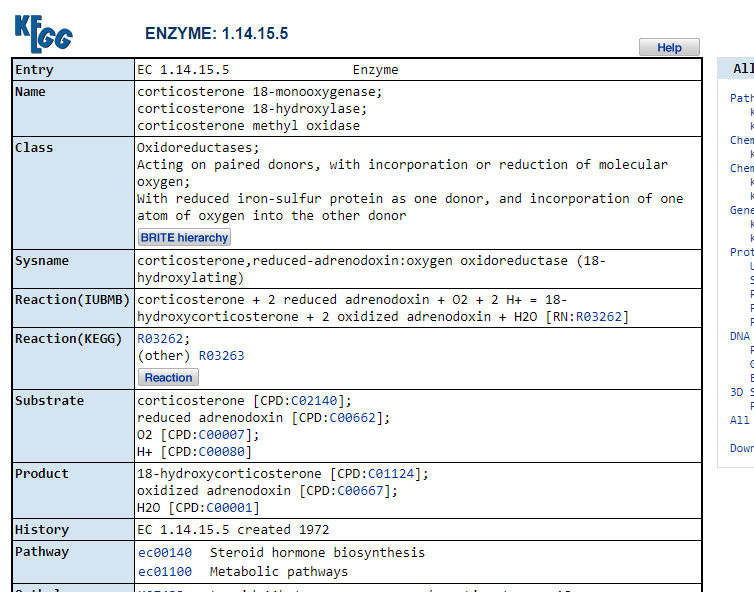


**Fig 17.0**

* Selecting the [Homo sapiens](http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00140) version of the map shows as a variety of pathways whereas no corresponding [bacterial map for E.coli](http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=eco&mapno=00140) exists. Not surprisingly, because microorganisms lack the capability of synthesizing steroid hormones. The database offers only links to eukaryotic organisms.

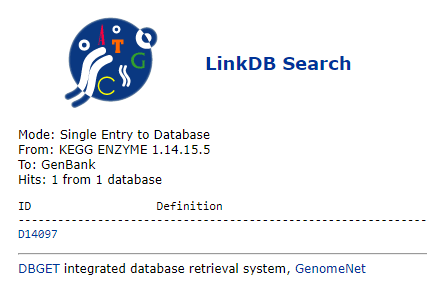
**Why are some enzymes not colored even though they are part of a pathway of enzymes that are colored?**

* Sometimes an enzyme is not marked where you would expect it like in the aldosterone pathway above.
* This pathway map shows all known reactions summarized in a standard pathway map. Species specific enzymes are marked green.
* Missing enzymes that appear to interrupt a pathway occur when no entry for this enzyme (gene, amino acid sequence, protein structure) exists in *any* database, not only KEGG.
* The enzyme with the entry [EC 1.14.15.5](http://www.genome.ad.jp/dbget-bin/www_bget?enzyme+1.14.15.5) is *Corticosterone 18-monooxygenase* and converts corticosterone into aldosterone.



**Fig 18.0**

* Following the E.C. link for this enzyme to the entry in [GenBank](http://www.genome.ad.jp/dbget-bin/www_bget?genbank-today+D14097) (mirrored from NCBI) shows one nucleic acid sequence report for rat (exon 9 of Rat CYP11B2 gene for aldosterone synthase).



**Fig 19.0**