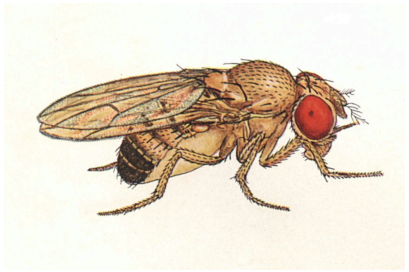
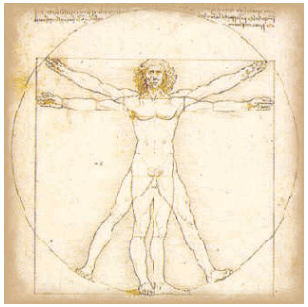
CARNEGIE INSTITUTION FOR SCIENCE | EMBRYOLOGY

Exploring Similarities between Man and Beast Using Publicly Available Genomic Resources



Frederick J Tan
Bioinformatics Research Faculty
23 November 2013

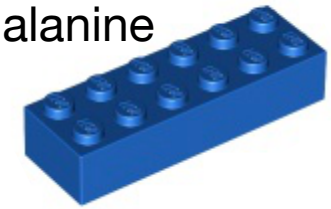


http://emb.carnegiescience.edu/cty

Special thanks to Valeriya Gaysinskaya, Pavol Genzor, Safia Malki, Zehra Nizami, and Gaëlle Talhouarne for devising these examples


CAGTCAGTGACCTTAACTGCT
TACTGCACAGACCCAACTG
CCGAGCACTCAGGAGAG
CCAGGCTAGTTTTGGAT
TGGGGGTTGGGGGGAG
AGGTTTGACCCAGC
GTAGAAGGTTCAG
CACACAAATACATAAAGTGTCTT

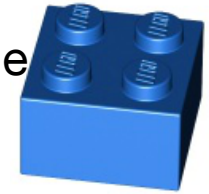
alanine



tryptophan

lysine

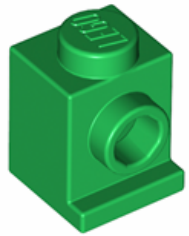
arginine



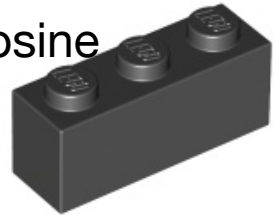
glycine

glutamate

aspartate

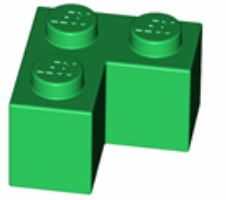


tyrosine



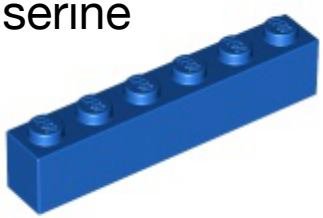
threonine

proline



phenylalanine

serine



UniProt

Downloads · Contact · Documentation/H

Search | Blast | Align | Retrieve | ID Mapping

Search in Protein Knowledgebase (UniProtKB)

Query Search | Advanced Search » | Clear

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is **not** reviewed. Includes complete and reference proteome sets.
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations, taxonomy, keywords, subcellular locations, cross-referenced databases and more.

NEWS

UniProt release 2013_11 - Nov 13, 2013

Forever young and cancer-free, in a black hole

› Statistics for UniProtKB:
Swiss-Prot · TrEMBL

› Forthcoming changes

› News archives

Follow @uniprot 729 followers

SITE TOUR

unitprot.org

Sequences

Sequence		Length	Mass (Da)	Tools
P02417 [UniParc].	FASTA	149	16,311	Blast go

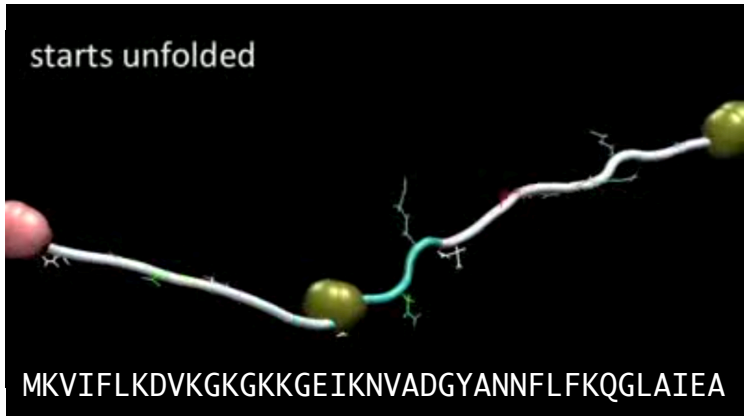
Last modified May 1, 1991. Version 2.
Checksum: 8252B98C1677C130

```
        10         20         30         40         50         60
MKVIFLKDVK GKGKKGEIKN VADGYANNFL FKQGLAIEAT PANLKALEAQ KQKEQRQAAE

        70         80         90        100        110        120
ELANAKKLKE QLEKLTVTIP AKAGEGGRLF GSITSKQIAE SLQAQHGLKL DKRKIELADA

       130        140
IRALGYTNVP VKLHPEVTAT LKVHVTEQK
```

« Hide

Simulation of a millisecond folder: NTL9

youtube.com/watch?v=gFcp2Xpd29I



Home

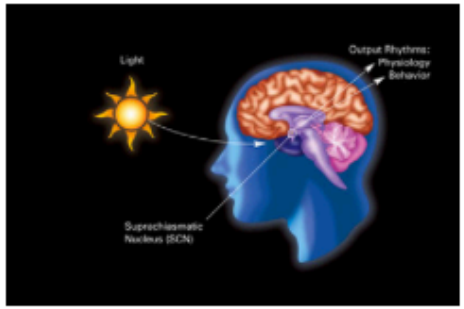
CTY Family Day 2013

Carnegie Institution for Science
Department of Embryology

Recent technological advancements have led to an explosion in the amount of information available about the genes and proteins present in organisms ranging from viruses to bacteria to plants to humans. In this workshop, you will be introduced to some of the publicly accessible tools and genomic databases that allow anyone with an internet connection to explore how genes in fundamental biological processes are conserved between model organisms and man.

Night and Day: Regulating circadian rhythm with cryptochromes

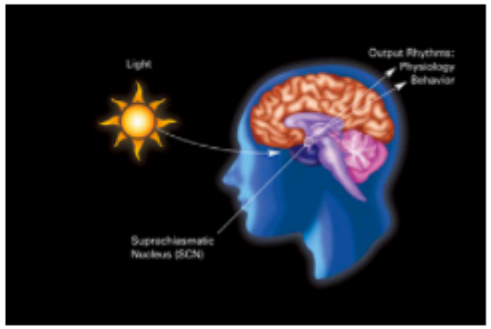
Many organisms vary their activities according to the time of day. These circadian (circa- "approximately" -dian "a day") rhythms help find prey and avoid predators. A protein called cryptochrome acts to coordinate circadian rhythms with changes in the daily sunrise and sunset.



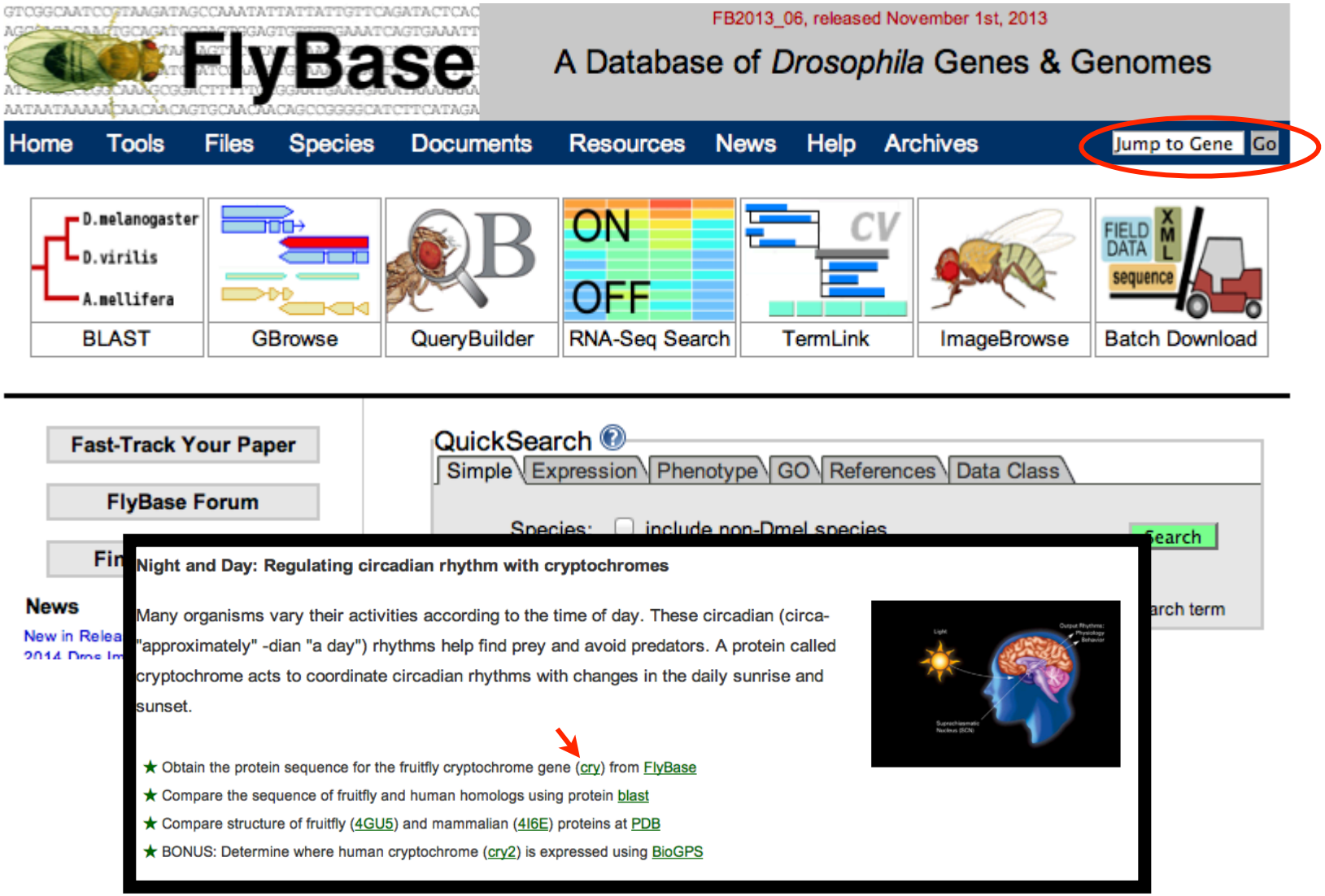
- ★ Obtain the protein sequence for the fruitfly cryptochrome gene (cry) from FlyBase
- ★ Compare the sequence of fruitfly and human homologs using protein blast
- ★ Compare structure of fruitfly (4GU5) and mammalian (4I6E) proteins at PDB
- ★ BONUS: Determine where human cryptochrome (cry2) is expressed using BioGPS

Night and Day: Regulating circadian rhythm with cryptochromes

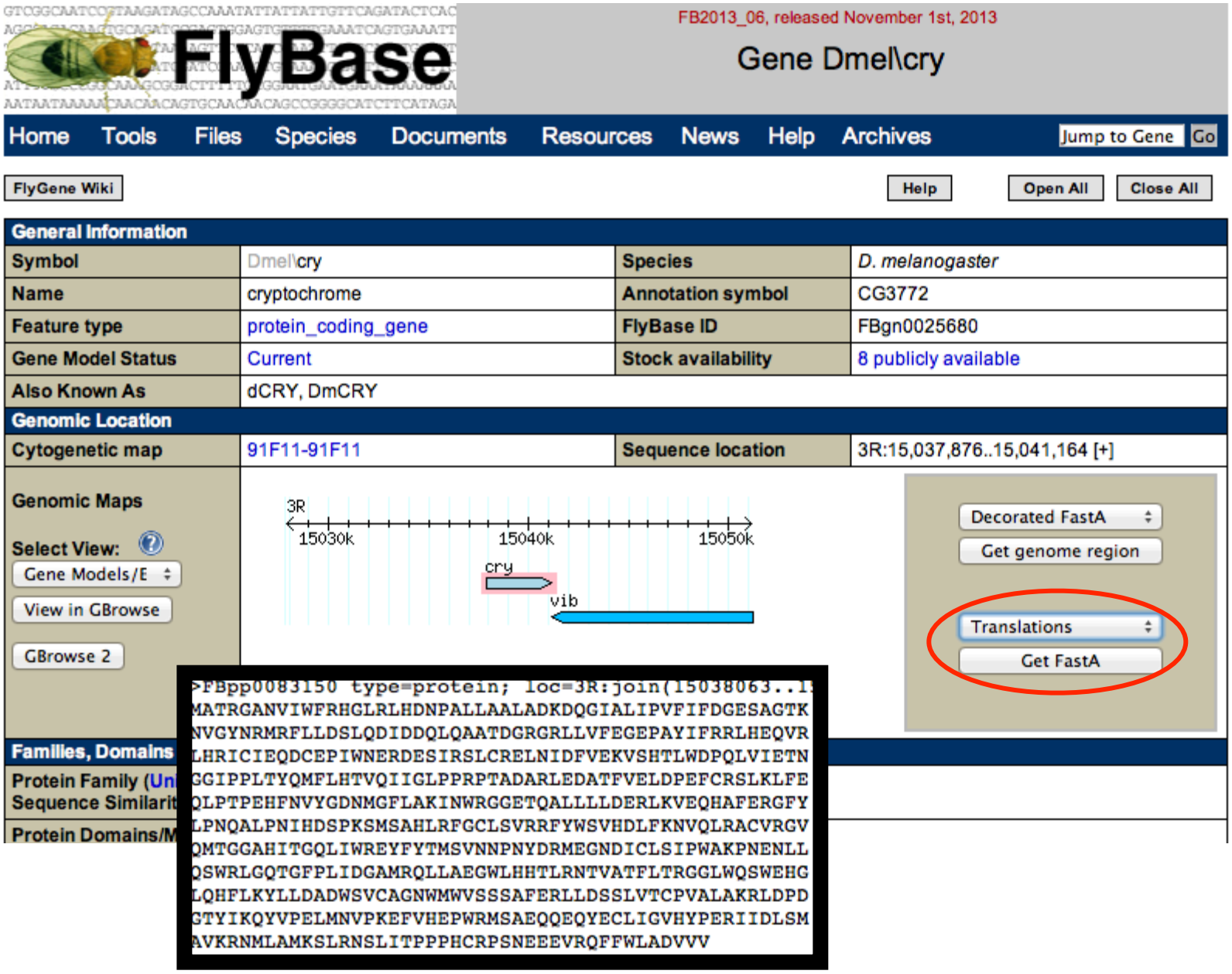
Many organisms vary their activities according to the time of day. These circadian (circa- "approximately" -dian "a day") rhythms help find prey and avoid predators. A protein called cryptochrome acts to coordinate circadian rhythms with changes in the daily sunrise and sunset.

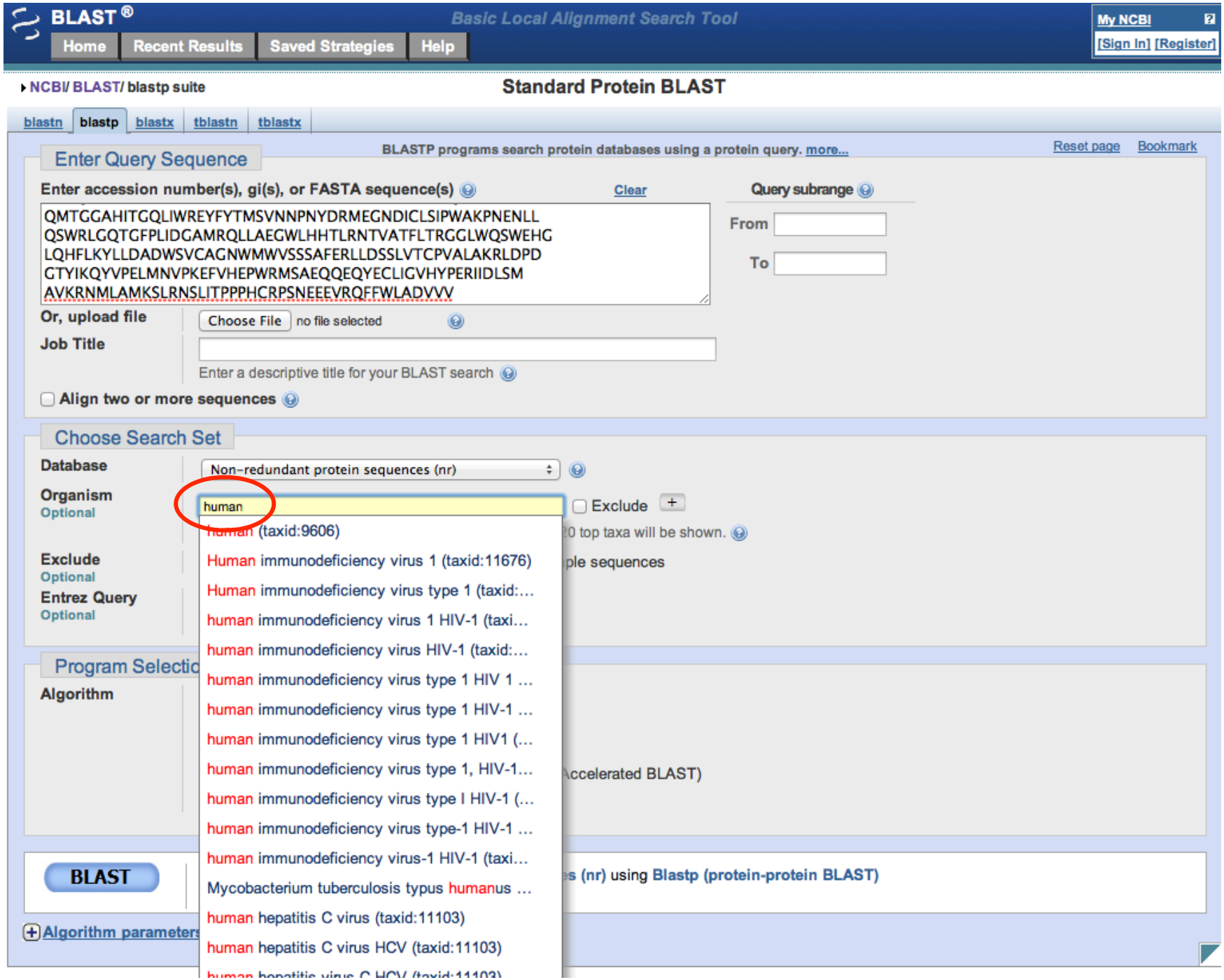


- ★ Obtain the protein sequence for the fruitfly cryptochrome gene (cry) from FlyBase
- ★ Compare the sequence of fruitfly and human homologs using protein blast
- ★ Compare structure of fruitfly (4GU5) and mammalian (4I6E) proteins at PDB
- ★ BONUS: Determine where human cryptochrome (cry2) is expressed using BioGPS


FlyBase
FB2013_06, released November 1st, 2013
A Database of Drosophila Genes & Genomes
Home
Tools
Files
Species
Documents
Resources
News
Help
Archives
Jump to Gene
Go
D.melanogaster
D.virilis
A.mellifera
BLAST
GBrowse
QueryBuilder
ON
OFF
RNA-Seq Search
CV
TermLink
ImageBrowse
FIELD DATA
XML
sequence
Batch Download
Fast-Track Your Paper
FlyBase Forum
QuickSearch
Simple
Expression
Phenotype
GO
References
Data Class
Species:
include non-Dmel species
Search
News
Night and Day: Regulating circadian rhythm with cryptochromes
Many organisms vary their activities according to the time of day. These circadian (circa- "approximately" -dian "a day") rhythms help find prey and avoid predators. A protein called cryptochrome acts to coordinate circadian rhythms with changes in the daily sunrise and sunset.
★ Obtain the protein sequence for the fruitfly cryptochrome gene (cry) from FlyBase
★ Compare the sequence of fruitfly and human homologs using protein blast
★ Compare structure of fruitfly (4GU5) and mammalian (4I6E) proteins at PDB
★ BONUS: Determine where human cryptochrome (cry2) is expressed using BioGPS

flybase.org


FlyBase
FB2013_06, released November 1st, 2013
Gene Dmel\cry
Home Tools Files Species Documents Resources News Help Archives
Jump to Gene Go
FlyGene Wiki
Help
Open All
Close All
General Information
Symbol
Dmel\cry
Species
D. melanogaster
Name
cryptochrome
Annotation symbol
CG3772
Feature type
protein_coding_gene
FlyBase ID
FBgn0025680
Gene Model Status
Current
Stock availability
8 publicly available
Also Known As
dCRY, DmCRY
Genomic Location
Cytogenetic map
91F11-91F11
Sequence location
3R:15,037,876..15,041,164 [+]
Genomic Maps
Select View:
Gene Models/E
View in GBrowse
GBrowse 2
3R
15030k
15040k
15050k
cry
vib
Decorated FastA
Get genome region
Translations
Get FastA
>FBpp0083150 type=protein; loc=3R:join(15038063..1
MATRGANVIWFRHGLRLHDNPALLAALADKDQGIALIPVFIFDGESAGTK
NVGYNRMRFLLDSLQDIDDQLQAATDGRGRLLVFEGEPAYIFRRLHEQVR
LHRICIEQDCEPIWNERDESIRSLCRELNIDFVEKVSHTLWDPQLVIETN
GGIPPLTYQMFLHTVQIIGLPPRPTADARLEDATFVELDPEFCRSLKLFE
QLPTPEHFNVYGDNMGFLAKINWRGGETQALLLLDERLKVEQHAFERGFY
LPNQALPNIHDSPKSMSAHLRFGCLSVRRFYWSVHDLFKNVQLRACVRGV
QMTGGAHITGQLIWREYFYTMSVNNPNYDRMEGNDICLSIPWAKPNENLL
QSWRLGQTGFPLIDGAMRQLLAEGWLHHTLRNTVATFLTRGGLWQSWEHG
LQHFLKYLLDADWSVCAGNWMWVSSSAFERLLDSSLVTCPVALAKRLDPD
GTYIKQYVPELMNVPKEFVHEPWRMSAEQQEQYECLIGVHYPERIIDLSM
AVKRNMLAMKSLRNSLITPPPHCRPSNEEEVRQFFWLADVVV
Families, Domains
Protein Family (Un
Sequence Similarit
Protein Domains/M

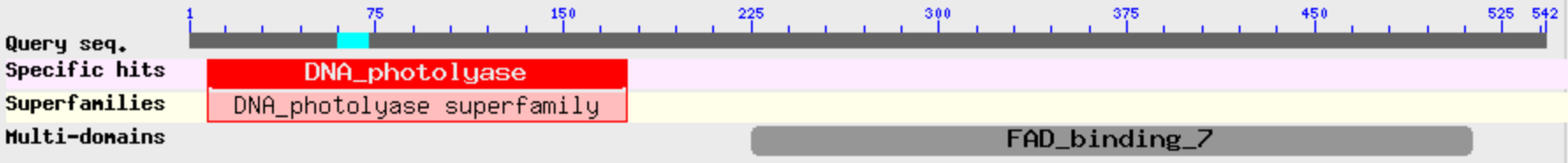


blast.ncbi.nlm.nih.gov

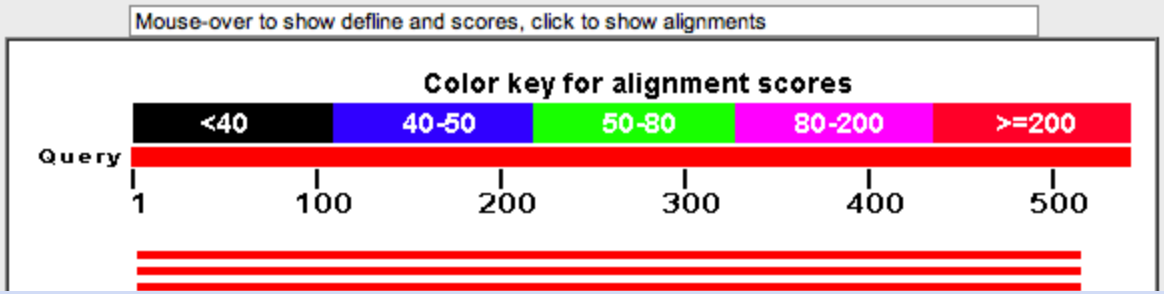
Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of 13 Blast Hits on the Query Sequence



Alignments

Download GenPept Graphics ▼ Next ▲ Previous Descriptions

KIAA0658 protein [Homo sapiens]

Sequence ID: dbj|BAA31633.1| Length: 589 Number of Matches: 1

Range 1: 18 to 506 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
375 bits(962)	4e-121	Compositional matrix adjust.	214/521(41%)	298/521(57%)	41/521(7%)

```
Query  5    GANVIWFRHGLRLHDNPALLAALADKDQGIALIP-VFIFDGESAGTKNVGYNRMRFLLDS  63
            ++V WFR GLRLHDNPALLAA+    +G   +  V+I D   A + +VG NR RFLL S
Sbjct  18   ASSVHWFRKGLRLHDNPALLAAV----RGARCVRCVYILDPWFAASSSVGINRWRFLLQS  73

Query  64   LQDIDDQLQAATDGRGRLLVFEGEPAYIFRRLHEQVRLHRICIEQDCEPIWNERDESIRS  123
            L+D+D  L+       RL V  G+PA +F RL ++  + R+  E D EP   ERD +I
Sbjct  74   LEDLDTSLRKLNS---RLFVVRGQPADVFPRLFKEWGVTRLTFEYDSEPFGKERDAAIMK  130

Query  124  LCRELNIDFVEKVSHTLWDPQLVIETNGGIPPLTYQMFLHTVQIIGLPPRPTADARLEDA  183
            + +E  ++ V + SHTL+D   +IE NG  PPLTY+ F   +  + LP +P       +
Sbjct  131  MAKEAGVEVVTENSHTLYDLDRIIELNGQKPPLTYKRFQAIISRMELPKKPVGLVTSQQM  190

Query  184  TFVELDPEFCRSLKLFEQLPTPEHFNVYG----DNMGF----LAKINWRGGETQALLLLD  235
                  E CR+ ++ E      H   YG    + +GF    L    W+GGET+AL  LD
Sbjct  191  -------ESCRA-EIQEN-----HDETYGVPSLEELGFPTEGLGPAVWQGGETEALARLD  237
```

Related Information

Gene - associated gene details

Map Viewer - aligned genomic context

RCSB PDB PROTEIN DATA BANK → PDB-101

A MEMBER OF THE PDB | EMDataBank

An Information Portal to Biological Macromolecular Structures

As of **Tuesday Nov 19, 2013 at 4 PM PST** there are **95644** Structures | **PDB Statistics**

Search
Advanced
Browse

Everything | Author | Macromolecule | Sequence | Ligand

e.g., PDB ID, molecule name, author

Search History , Previous Results

Customize This Page

Available on the App Store

PDB-101 Hide

Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

MyPDB Hide

Login to your Account
Register a New Account
MyPDB Help Page

Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

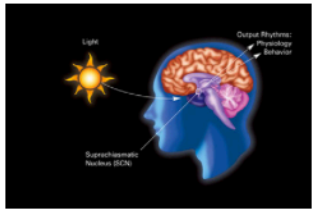
Biological Macromolecular Resource

Full Description

New Features Hide

Night and Day: Regulating circadian rhythm with cryptochromes

Many organisms vary their activities according to the time of day. These circadian (circa- "approximately" -dian "a day") rhythms help find prey and avoid predators. A protein called cryptochrome acts to coordinate circadian rhythms with changes in the daily sunrise and sunset.

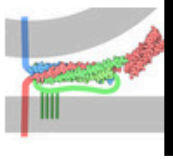


★ Obtain the protein sequence for the fruitfly cryptochrome gene (cry) from FlyBase
★ Compare the sequence of fruitfly and human homologs using protein blast
★ Compare structure of fruitfly (4GU5) and mammalian (4I6E) proteins at PDB
★ BONUS: Determine where human cryptochrome (cry2) is expressed using BioGPS

Learn: Featu

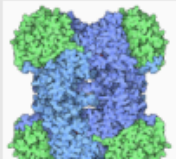
Structural View

PDB-101



are used inside cells to transport digestive enzymes from the Golgi to their final location in lysosomes. They are also used to deliver molecules out of the cell: for example, neurotransmitters are released from vesicles that fuse with the cell membrane at nerve synapses. The 2013 Nobel Prize was awarded to three researchers who have revealed the central molecular machinery for this process of vesicle fusion.

Full Article



Protein Structure Initiative Featured System

Methylation of Arginine

The function of many proteins is tuned and regulated after they are synthesized by modification of key amino acids. These modifications change the chemistry of the amino acid, creating a distinctive mark that can be used as a signal or to customize interactions with other molecules. Phosphate groups, for

Website Release Archive:

RCSB PDB News Hide

Weekly | Quarterly | Yearly

2013-11-19

Help Desk Questions and Answers

PDB-101 Hide
Structural View of Biology
Understanding PDB Data
Molecule of the Month
Educational Resources
Author Profiles

MyPDB Hide
Login to your Account
Register a New Account
MyPDB Help Page

Home Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools

Tools Hide
Download Files
Compare Structures

Summary | 3D View | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Structure of Full-length Drosophila Cryptochrome

4GU5

Display Files
Download Files
Share this Page

DOI:10.2210/pdb4gu5/pdb

ENTRY 4GU5 SUPERSEDES 3TVS

Primary Citation

Updated structure of Drosophila cryptochrome.

Levy, C., **Zoltowski, B.D.**, **Jones, A.R.**, **Vaidya, A.T.**, **Top, D.**, **Widom, J.**, **Young, M.W.**, **Scrutton, N.S.**, **Crane, B.R.**, **Leys, D.**

Journal: (2013) Nature **495:** E3-E4

PubMed: 23518567
PubMedCentral: PMC3694752
DOI: 10.1038/nature11995
Search Related Articles in PubMed

PubMed Abstract:

No abstract available... [Read More & Search PubMed Abstracts]

Related Citations in PDB Entry (REMARK 1) Hide

Structure of full-length Drosophila cryptochrome.
Zoltowski, B.D., Vaidya, A.T., Top, D., Widom, J., Young, M.W., Crane, B.R.
(2011) *Nature* ***480:*** *396* PubMed

Molecular Description Hide

Classification: Signaling Protein

Biological Assembly 1

3D View | More Images...

No symmetry
Stoichiometry: **Monomer**
Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:
Simple Viewer | Protein Workshop
Kiosk Viewer

RCSB PDB PROTEIN DATA BANK → RCSB PDB-101

A MEMBER OF THE PDB | EMDataBank

An Information Portal to Biological Macromolecular Structure

As of **Tuesday Nov 19, 2013 at 4 PM PST** there are **95644** Structures | **PDB Statistics**

Search
Advanced
Browse

Everything | Author | Macromolecule | Sequence | Ligand

e.g., PDB ID, molecule name, author

Search History , Previous Results

PDB-101 Hide
- Structural View of Biology
- Understanding PDB Data
- Molecule of the Month
- Educational Resources
- Author Profiles

MyPDB Hide
- Login to your Account
- Register a New Account
- MyPDB Help Page

Home Hide
- News & Publications
- Usage/Reference Policies
- Deposition Policies
- Website FAQ
- Deposition FAQ
- Contact Us
- About Us
- Careers
- External Links
- Sitemap
- New Website Features

Deposition Hide
- All Deposit Services
- Electron Microscopy
- X-ray | NMR
- Validation Server
- BioSync Beamlines/Facilities
- Related Tools

Tools Hide
- Download Files
- Compare Structures
- Drug & Drug Target Mapping
- File Formats

Summary | 3D View | Sequence | Annotations | Seq. Similarity | 3D Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Structure of Full-length Drosophila Cryptochrome

4GU5

Display Files
Download Files
Share this Page

NOTE: Use your mouse to drag, rotate, and zoom in and out of the structure.

Structure Details

Structure	Biological Assembly 1
Symmetry Type	Global Symmetry
Symmetry	C1
Stoichiometry	A

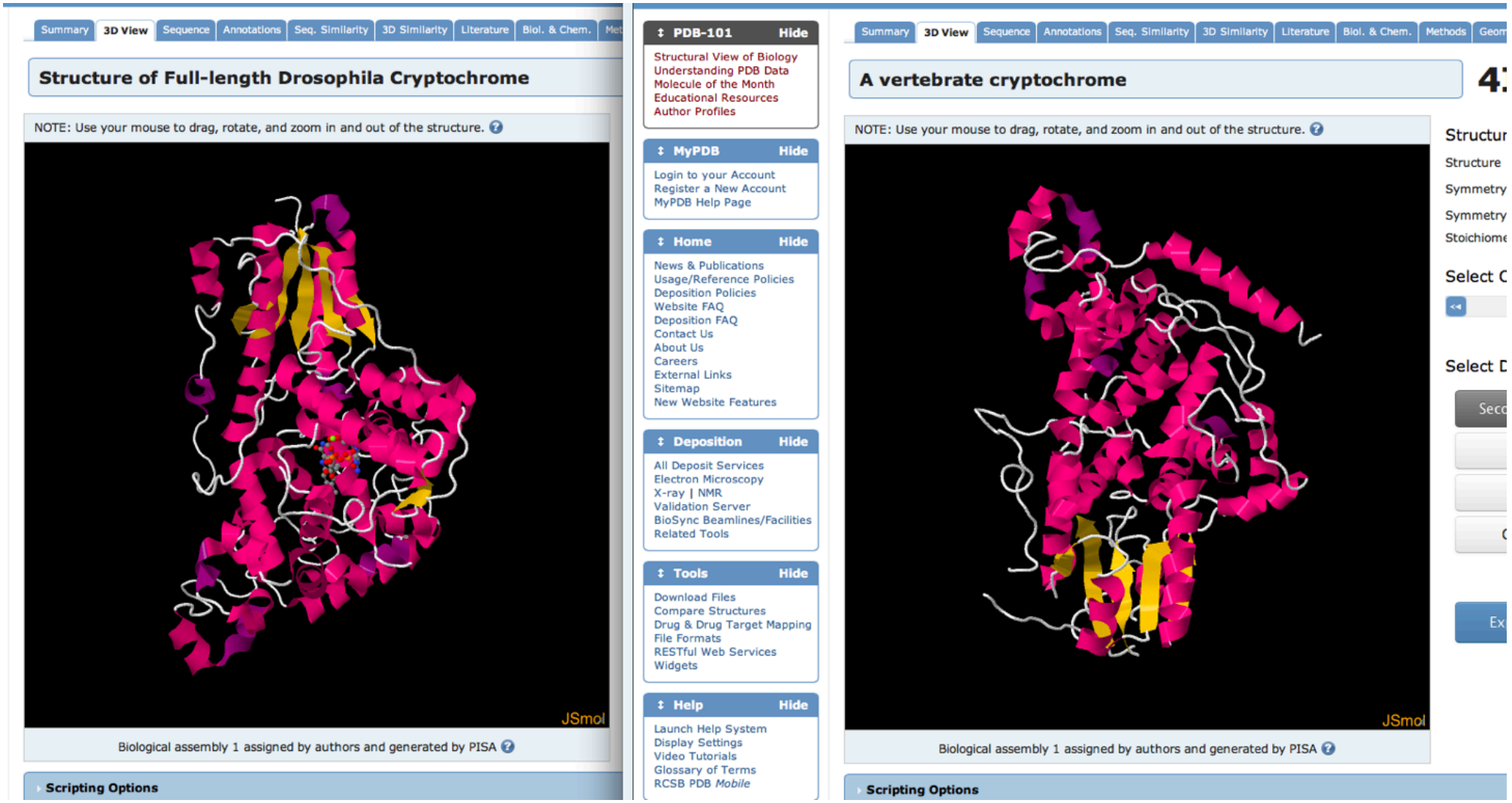
Select Orientation

◂◂ *Front* ▸▸

Select Display Mode

Secondary Structure
Subunit
Symmetry
Custom View

Export 3D Image



What is the same? What is different?
What parts do you think are important?