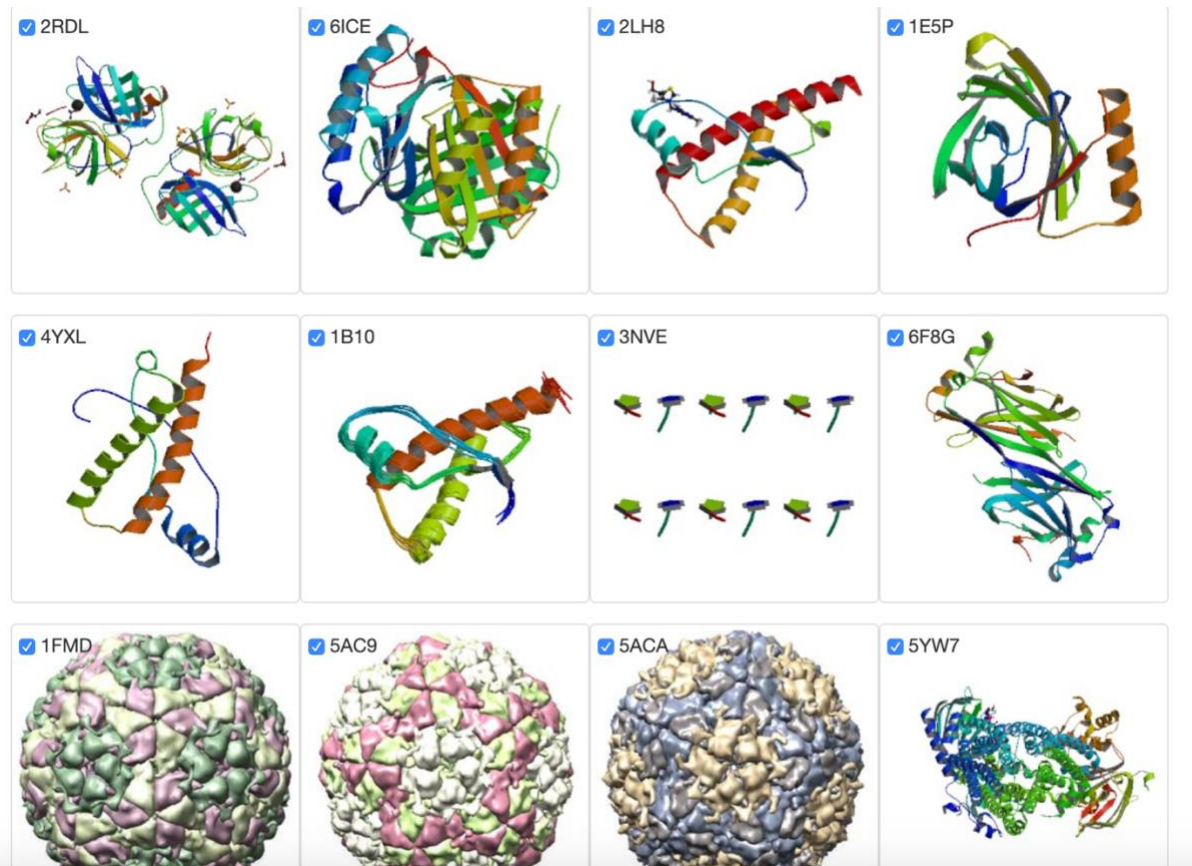


## Question1. 2

Mnemonic	i	MESAU
Taxon identifier	i	10036
Scientific name	i	Mesocricetus auratus
Taxonomy navigation		<a href="#">↑</a> > <a href="#">Mesocricetus</a> <a href="#">↓</a> Terminal (leaf) node.
Common name	i	Golden hamster
Synonym	i	-
Other names	i	>Golden hamsters >Mesocricetus auratus Waterhouse, 1839 >Syrian golden hamster >Syrian golden hamsters >Syrian hamster <a href="#">More »</a>
Rank	i	SPECIES
Lineage	i	> cellular organisms > <a href="#">Eukaryota</a> > Opisthokonta > <a href="#">Metazoa</a> > Eumetazoa > Bilateria > Deuterostomia > <a href="#">Chordata</a> > <a href="#">Craniata</a> > <a href="#">Vertebrata</a> > Gnathostomata > Teleostomi > <a href="#">Euteleostomi</a> > Sarcopterygii > Dipnotetrapodomorpha > Tetrapoda > Amniota > <a href="#">Mammalia</a> > Theria > <a href="#">Eutheria</a> > Boreoeutheria > <a href="#">Euarchontoglires</a> > <a href="#">Glires</a> > <a href="#">Rodentia</a> > <a href="#">Myomorpha</a> > <a href="#">Muroidea</a> > <a href="#">Cricetidae</a> > <a href="#">Cricetinae</a> > <a href="#">Mesocricetus</a>
Strains	i	> Syrian ,Golden Syrian
See also		> <a href="#">animaldiversity.org</a> > <a href="#">www.arkive.org</a> > <a href="#">www.iucnredlist.org</a> > <a href="#">www.maniform.com</a> > <a href="#">www.planet-mammiferes.org</a> > <a href="#">en.wikipedia.org</a> > <a href="#">www.biolib.cz</a> > <a href="#">www.departments.bucknell.edu</a> > <a href="#">NCBI</a>

## [Uniprot.org](#)

Uniprot uses TrEMBL Uniport is a good tool because it is easy to use and easy to find as well because it has a Powerful searching tool

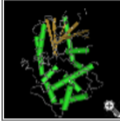


[rcsb.org](http://rcsb.org)

RCSB uses PDB. RCSB is also an amazing tool because it is convenient and visual in contrast to uniprot. The view of the results could be set differently if you want details images or etc. It also had a 3d view of the structures. Overall, this website was the best one.

cd00066: **G-alpha**

Download alignment ?



**Alpha subunit of G proteins (guanine nucleotide binding)**  
The alpha subunit of G proteins contains the guanine nucleotide binding site. The heterotrimeric GNP-binding proteins are signal transducers that communicate signals from many hormones, neurotransmitters, chemokines, and autocrine and paracrine factors. Extracellular signals are received by receptors, which activate the G proteins, which in turn route the signals to several distinct intracellular signaling pathways. The alpha subunit of G proteins is a weak GTPase. In the resting state, heterotrimeric G proteins are associated at the cytosolic face of the plasma membrane and the alpha subunit binds to GDP. Upon activation by a receptor GDP is replaced with GTP, and the G-alpha/GTP complex dissociates from the beta and gamma subunits. This results in activation of downstream signaling pathways, such as cAMP synthesis by adenylyl cyclase, which is terminated when GTP is hydrolyzed and the heterotrimers reconstitute.

Links ?

Source: [cd00882](#)  
Taxonomy: [Eukaryota](#)  
PubMed: [12 links](#)  
Book: [4 links](#)  
Protein: [Representatives](#)  
[Specific Protein](#)  
[Related Protein](#)  
[Related Structure](#)  
[Architectures](#)  
Superfamily: [cd21455](#)

Conserved Features/Sites ?

PubMed References ?

GTP/Mg2+

adenylyl

beta - gamma

GoLoco binding

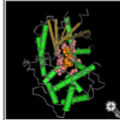
putative

**Feature 1:** GTP/Mg2+ binding site [chemical binding site]

**Evidence:**

- Structure:** 1AZT: Bos taurus Gs-alpha binds GTP-gamma-S, a GTP analog and Mg2+, defined using 3.5 Å contacts  
- [View structure](#) with Cn3D
- Comment:** binding site for GTP and Mg2+ (bound by conserved Thr)
- Citation:** [PMID 9242920](#)

[Download Cn3D for Viewing 3D Structure](#)
[Scroll to Sequence Alignment Display](#)



Statistics ?

PSSM-Id: 206639  
View PSSM: [cd00066](#)  
Aligned: 142 rows  
ThresholdBitScore: 227.409  
ThresholdSettingGi: [1169858](#)  
Created: 1-Nov-2000  
Updated: 17-Jan-2013

Structure ?

Structure View  
Program: [Cn3D](#)  
Drawing: [All Atoms](#)

cd00066 is part of a hierarchy of related CD models.  
Use the graphical representation to navigate this hierarchy.  
cd00066 is a member of the superfamily [cd21455](#)

cd00066 Sequence Cluster

Zoom In



Sub-family Hierarchy

Interactive Display with CDTree ?



- cd00882 Ras\_like\_GTPase
- cd00066 G-alpha**
- cd00154 Rab
- cd01860 Rab5\_related
- cd01861 Rab6

[ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)

NCBI uses SWISS-PROT. NCBI was really hard to use but this difficulty came with a prize, this site was the most detail of them all, but the amount of detail was overwhelming, and the structure of the webpage wasn't pleasant either.

### Question 1. 3

I cannot understand much because I have no prior knowledge of biology, it is also very unreadable, but I would assume the computers can understand the codes like 1B10 which is completely unintelligible to me

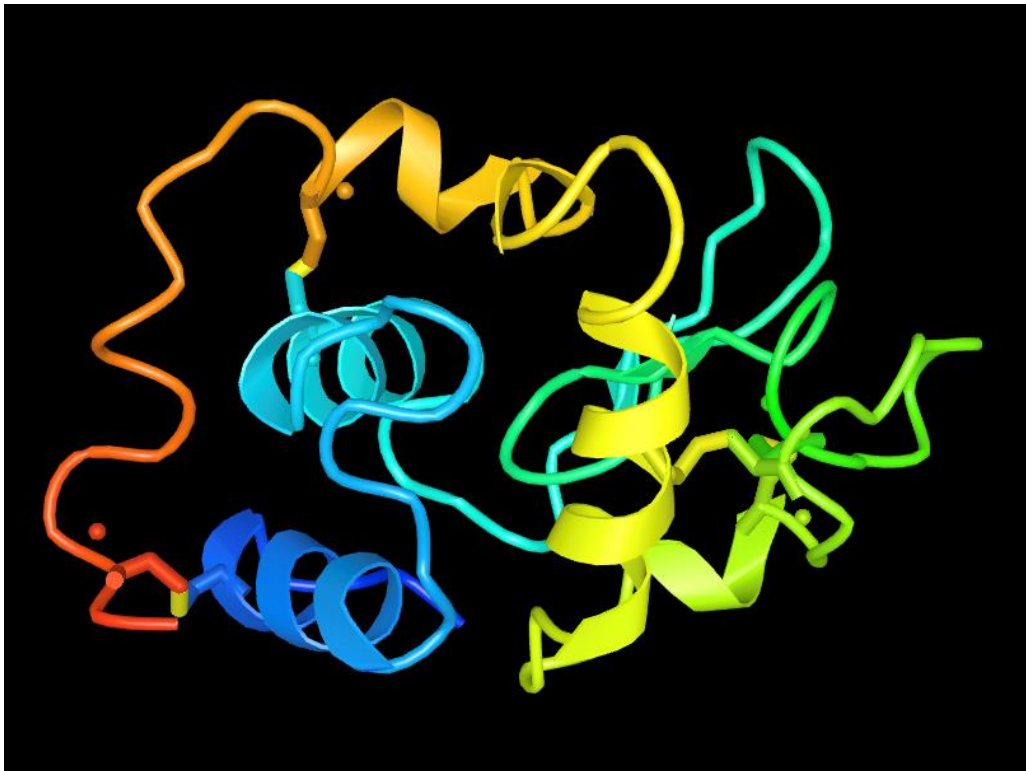
### Question1. 4

I have no prior knowledge of biology and such a complex topic like genes does not make any sense to me, but I can understand a few things, what I suspect is all these data is made of type string, the computer. I also had an observation in PDB database, all the biological structures that start with 1F in their name are classified as "*structural proteins*" and all the sphere shapes are viruses ,I also observed that viruses that start with 6M in their codes are "*Eastern equine encephalitis virus (strain Florida 91-469)*" organisms so I assume these biological structures are named according to their classifications and organisms.

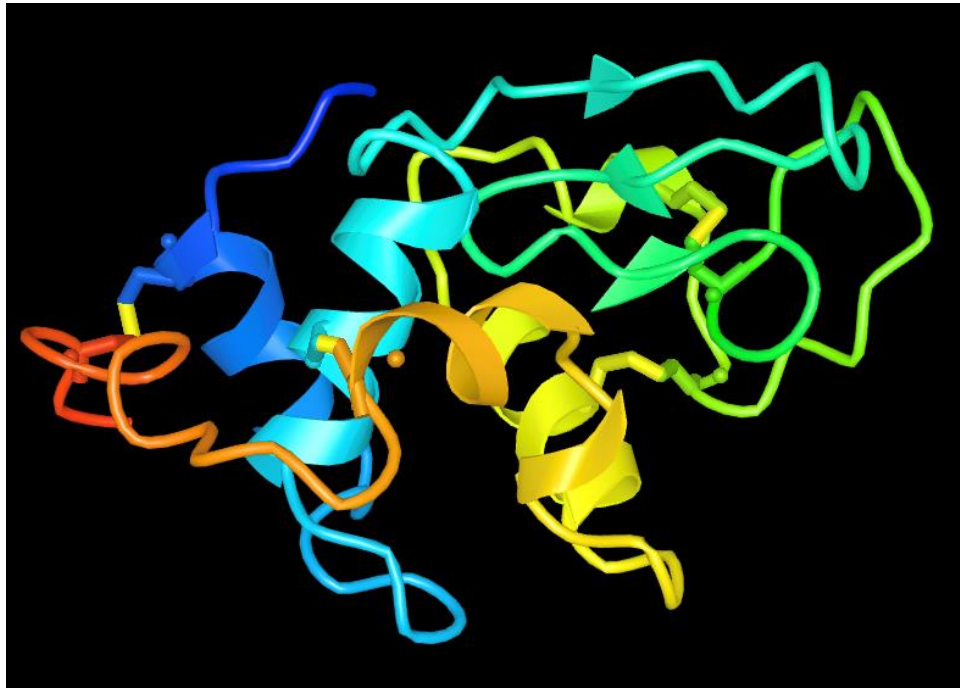
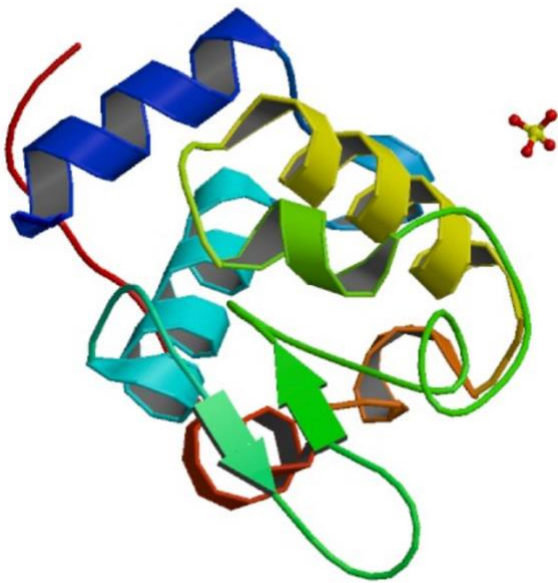
### Question 2.1

```
>sp|P81708|LYSC1_CANLF Lysozyme C, milk isozyme OS=Canis lupus  
familiaris OX=9615 PE=1 SV=1  
KIFSKCELARKLKSMGMDGFHGYSLANWVCMAEYESNFNTQAFNGRNSNGSSDYGIFQLN  
SKWWCKSNSSHSSANACNIMCSKFLDDNIDDDIACAKRVVKDPNGMSAWVAWVKHCKGKDL  
SKYLASCNL
```

### Question 2.2



### Question 2.3



Most of the ribbon-like parts of the shape are similar. They are similarly drawn as well. The Distill version has random dots (spheres) that are not in the PDB version that I got from RCSB. The tube-like parts of the structure are also different. Overall the shapes are similar but not identical.

### Question 3.3

From what I understood the computers that can predict the shape of proteins' structure are essential to development of science and especially biology, biotechnology and medicine. The lecturer used a great analogy which I'm going to use as well. He said molecular biology is like having a lot of text in an alien language.

Today, computers can almost understand human speech but for them these speeches are nothing more than 0s and 1s. So how do computers understand our sentences? Well the first step is to understand the meaning of the words. Computers use different techniques to understand words. One of the common ways is called distributional semantics. In this technique the computer understands the meaning of the word by the context of the sentence.

But How would we turn this into mathematics that computer can understand?

We use a simple technique called count vector this means we check how many times a word appears in the same article as the other common word excluding the stop words (i.e. and, be, to, the, as, etc.). this way the computer can make patterns of the word that are commonly used together.

If we go back to the lecturer's analogy again, we see that we can apply the same logic for molecular biology as well, if we consider it an alien language. To the computer it doesn't make a difference whether the word has a meaning, it is just series of codes and symbols given to the different molecules or proteins. The computer puts the similar structures together to predict a new structure.

## Bibliography

Flender, S. ., Machines that learn by doing. *Towards Data Science* .

Garbade, D. M. , A Simple Introduction to Natural Language Processing.

Pierre Baldi and Gianluca Pollastri. , A Machine Learning Strategy for Protein Analysis.

CrashCourse Youtube channel, Natural Language Processing [episode 7 of crashcourse AI],  
[Recorded by Jabrel in PBS Digital Studios].