00: Protein Prediction 1 - structure

pp1_hello

Protein Prediction 1 (for Computational Biology) - Protein structure
TUM summer semester
Computational biologists only
Videos: YouTube / www.rostlab.org

THANKS:

+ Dmitrij Nechaev

Special lectures:
- 07/xx Predrag Radivojac - Indiana Univ.
- 06/xx Yana Bromberg - Rutgers Univ.

No lecture:
- 05/15 Ascension day
- 05/23 Student assembly (SVV)
- 06/06 Whitsun holiday
- 06/15 Corpus Christi

LAST lecture: bef: Jul 11
after: Jul 28

Examen:
- Makeup:

WEDNESDAY(!!) July 12: 18:00-19:30 TBA

Exercises:

TBA

CONTACT: Lothar Richter richter@rostlab.org

Announcements
Protein Prediction
- Part 1: Structure
Definitions

- Computational Biology

- Bioinformatics

- http://www.rostlab.org/talks/
Definitions

- **Computational Biology**
  - biology replacing experiments by computers (include neurobiology, image processing)

- **Bioinformatics**
  - anything that has to do with storing and using the information about bio-sequences

- [http://www.rostlab.org/talks/](http://www.rostlab.org/talks/)
Who are we?
Burkhard Rost
Some Professional Positions
Study complex systems
Inga Weise assistant@rostlab.org
Lothar Richter teaching@rostlab.org
Nobel Prizes in Informatics?
2013 Nobel Prize in Chemistry

for what?

who are they?
2013 Nobel Prize in Chemistry

“for the development of multiscale models for complex chemical systems” (nobelprize.org)

Martin Karplus
Harvard Univ, USA & Univ Strasbourg, France

Michael Levitt
Stanford Univ, USA

Arieh Warshel
USC, Los Angeles, USA
Protein structure comparisons

All-alpha
3sdh

All-beta
1bww

AlphaBeta
1xne
Predict protein function
Classical physics vs quantum physics

Michael Levitt (AP): It’s sort of nice ... to see that computational science, computational biology is being recognized. ... It’s become a very large field and it’s always in some ways been the poor sister, or the ugly sister, to experimental biology.

Arieh Warshel (NobelPrize) [experimental results are] like seeing a watch and wondering how actually it works. ... what we developed is a way which requires a computer to look ... and then to eventually understand how exactly [a protein] does what it does.
TOC today

- **Biology introduction**
  - Organisms
  - Genes
  - Central dogma

- **Protein introduction**
  - numbers asf

- **NEXT lectures**
  - amino acids & energy
  - structure & domains
  - 3D comparisons
Background
Life is diverse

Trilobite Bergeroniellus spinosus
Lena River Gorge, Siberia
http://www.emory.edu/COLLEGE/ENVS/research/ichnology
How does life work?

Leonardo Da Vinci (1452-1519)
Vitruve Luc Viatour (~1492)
How does life work?

Leonardo Da Vinci (1452-1519)
Vitruve Luc Viatour (~1492)
Common 2 Life?

Trilobite Bergeroniellus spinosus
Lena River Gorge, Siberia
http://www.emory.edu/COLLEGE/ENVS/research/ichnology
Central dogma of molecular biology

DNA → RNA → Protein

information, code, library, manual

intermediate step

machinery of life

4 nucleotides
GATC

4 nucleotides
GAUC

20 amino acids
ACDEFGHIKLMNPRSTVWY

transcription

translation

© Wikipedia

© Burkhard Rost
ROSTLAB.
TUM
Images
What is similar?
What is similar?
What is similar?

- Silicone
- Glass

© Wikipedia
© http://www.lionsons.com
© Burkhard Rost
© TUM
What is this?

Diamond

Graphite/coal

© Wikipedia

© Burkhard Rost

ROSTLAB.

TUM

© 38/128
What is this?

Diamond

Graphite/coal

carbon

diamond lattice

© Wikipedia

© Burkhard Rost
Which molecule dominates our body weight?
Which molecule makes most of our body?

H₂O - water

how much?
What elements make most of our body?

H2O - water

about 50-65% in adults
What elements make most of our body?
Which ones make up life?

- Oxygen - O: 65.00%
- Hydrogen - H: 9.70%
- Body weight: 96.5%

The periodic table shows the elements that make up life, with oxygen and hydrogen being the most abundant.

© Wikipedia
Which ones make up life?

<table>
<thead>
<tr>
<th>Element</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxygen - O</td>
<td>65.00</td>
</tr>
<tr>
<td>carbon - C</td>
<td>18.60</td>
</tr>
<tr>
<td>hydrogen - H</td>
<td>9.70</td>
</tr>
<tr>
<td>nitrogen - N</td>
<td>3.20</td>
</tr>
<tr>
<td>calcium - Ca</td>
<td>1.80</td>
</tr>
<tr>
<td>phosphorus - P</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Body weight: 96.5
Which ones make up life?

<table>
<thead>
<tr>
<th>Element</th>
<th>Atom</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxygen</td>
<td>O</td>
<td>65.90</td>
</tr>
<tr>
<td>carbon</td>
<td>C</td>
<td>18.60</td>
</tr>
<tr>
<td>hydrogen</td>
<td>H</td>
<td>9.70</td>
</tr>
<tr>
<td>nitrogen</td>
<td>N</td>
<td>3.20</td>
</tr>
<tr>
<td>calcium</td>
<td>Ca</td>
<td>1.80</td>
</tr>
<tr>
<td>phosphorus</td>
<td>P</td>
<td>1.00</td>
</tr>
<tr>
<td>potassium</td>
<td>K</td>
<td>0.40</td>
</tr>
<tr>
<td>sodium</td>
<td>Na</td>
<td>0.20</td>
</tr>
<tr>
<td>chlorine</td>
<td>Cl</td>
<td>0.20</td>
</tr>
<tr>
<td>magnesium</td>
<td>Mg</td>
<td>0.10</td>
</tr>
<tr>
<td>sulfur</td>
<td>S</td>
<td>0.05</td>
</tr>
<tr>
<td>iron</td>
<td>Fe</td>
<td>0.03</td>
</tr>
<tr>
<td>iodine</td>
<td>I</td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Body weight: 96.5%**

---

© Burkhard Rost

© Wikipedia
What is this?
What is this?
What is this?

What is this?

Escherichia coli /E coli

What is this?
What is this?

Streptococcus
What is this?
What is this?

Salmonella bongori

© Wikipedia
What is this?

What is this?

Avian or bird flu (virus)

What is this?

What is this?

HIV virus

What is this?
Velcro - invented by George de Mestral

George de Mestral (1907-1990)

- Swiss electrical engineer
- Invented velcro 1949

http://www.engr.sjsu.edu/WofMatE/polymers.htm

© Wikipedia
© Burkhard Rost
© ROSTLAB
Velcro copied from plants

George de Mestral (1907-1990)

- Swiss electrical engineer
- Invented velcro 1949

Velcro copied from plants

http://www.engr.sjsu.edu/WofMatE/polymers.htm

© Wikipedia

burr (seed of fruit)

© Wikipedia
Bacteria differ in shape

© Wikipedia
What do they have in common?
Common to bacteria: single cells

single cell

© Wikipedia
Cells

Eukaryote

- Membrane-enclosed nucleus
- Nucleolus
- Mitochondrion
- Ribosomes
- Cell Membrane

Prokaryote

- Nucleoid
- Capsule (some prokaryotes)
- Flagellum
- Cell Wall (in some eukaryotes)

© Wikipedia
Do we carry bacteria?

*Leonardo Da Vinci (1452-1519)*
*Vitruve Luc Viatour (~1492)*
we carry bacteria when we are sick
Our bodies are teeming with bacteria


© NATURE REVIEWS MICROBIOLOGY 9:244-53, COPYRIGHT 2011
Common?

Trilobite Bergeroniellus spinosus
Lena River Gorge, Siberia
http://www.emory.edu/COLLEGE/ENVS/research/ichnology
DNA/Gene

DNA
DEOXYRIBONUCLEIC ACID

Cytosine
Guanine
Adenine
Thymine

© Wikipedia

www.chemistryexplained.com
Where do we have genes?
Genes in nucleus and mitochondria

green: cell (actin)
red: mitochondria
blue: nucleus

© Wikipedia
Genes anywhere else in eukaryotic cell?
Genes in nucleus and mitochondria

green: cell (actin)
red: mitochondria
blue: nucleus
Central dogma of molecular biology

DNA → RNA → Protein

information, code, library, manual

intermediate step

machinery of life

© Laszlo Kajan, TUM
Protein synthesis

© Wikipedia

© Laszlo Kajan, TUM
1. what is the smallest building block of life that can replicate?
   - protein - quarks - cells - organelles

2. different cells in typical human?
   - 20 - 200 - 400 - 1000

3. what are the parts of cells called?
   - cellular parts - organelles - organs - celluoplasts

4. which part of cells is called the “powerhouse”?
   - nucleus - Golgi apparatus - mitochondria - Endoplasmic reticulum

5. what part of a plant cell is involved with photosynthesis?
   - mitochondria - nucleus - smooth reticulum - chloroplast

6. what is mitosis?
   - cell death - cellular respiration - cell division - cellular communication

7. who first used the term cell?
   - Aristotle - Captain Hooke - Robert Hooke - James Watson

8. how many elements are found in amounts larger than trace amounts (0.01%) in our bodies?
   - 92 - 48 - 13 - 11

9. when communities of living things interact with non-living things they are called ...?
   - population - community - biosphere - ecosystem

10. the most common molecule in the human body is?
Q & A

1. what is the smallest building block of life that can replicate?
   - protein - quarks - cells - organelles

2. different cells in typical human?
   - 20 - 200 - 400 - 1000

3. what are the parts of cells called?
   - cellular parts - organelles - organs - celluoplasts

4. which part of cells is called the “powerhouse”?
   - nucleus - Golgi apparatus - mitochondria - Endoplasmic reticulum

5. what part of a plant cell is involved with photosynthesis?
   - mitochondria - nucleus - smooth reticulum - chloroplast

6. what is mitosis?
   - cell death - cellular respiration - cell division - cellular communication

7. who first used the term cell?
   - Aristotle - Captain Hooke - Robert Hooke - James Watson

8. how many elements are found in amounts larger than trace amounts (0.01%) in our bodies?
   - 92 - 48 - 13 - 11

9. when communities of living things interact with non-living things they are called ...?
   - population - community - biosphere - ecosystem

10. the most common molecule in the human body is?
    - H2O
Life: The Players
What is life?
Life is diverse

Trilobite Bergeroniellus spinosus
Lena River Gorge, Siberia
http://www.emory.edu/-College/Envs/research/ichnology
Can you define life?
Descriptive definition of life

- **Homeostasis**
  (regulation of internal environment to maintain constant state)

- **Organization - unit: cells**

- **Metabolism**
  (transfer of energy)

- **Growth**

- **Adaptation**

- **Response to stimuli**

- **Reproduction**
Descriptive definition of life

- Homeostasis
  (regulation of internal environment to maintain constant state)
- Organization - unit: cells
- Metabolism
  (transfer of energy)
- Growth
- Adaptation
- Response to stimuli
- Reproduction

Viruses = life?
Central dogma of molecular biology
modern DNA/RNA sequencing
Next generation sequencing

# Illumina HiSeq 2016

## Table 1: HiSeq X System Sequencing Capacity

<table>
<thead>
<tr>
<th></th>
<th>HiSeq X Ten System</th>
<th>HiSeq X Five System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Number of Instruments</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Annual Genome Capacity</td>
<td>&gt; 18,000</td>
<td>&gt; 9000</td>
</tr>
<tr>
<td>Price per 30x Genome</td>
<td>&lt; $1000</td>
<td>&lt; $1500</td>
</tr>
</tbody>
</table>

~2TB/run (<3 days)
Haemophilus Influenzae
*(Pfeiffer’s bacillus)*

Mycoplasma Genitalium

Sacharomyces cerevisiae

Caenorhabditis elegans, nematode - worm

Drosophila melanogaster fruit fly 195 authors
## Genome sizes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mycoplasma genitalium</em></td>
<td>470</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>1,740</td>
</tr>
<tr>
<td><em>Methanococcus jannaschi</em></td>
<td>1,738</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>4,288</td>
</tr>
<tr>
<td><em>Sacharomyces cerevisiae</em> - yeast</td>
<td>6,600</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em> - fruit fly</td>
<td>13,600</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em> - worm</td>
<td>19,000</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em> - mustard</td>
<td>26,735</td>
</tr>
<tr>
<td><em>Oryza sativa</em> - rice</td>
<td>50,000</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>*</td>
</tr>
</tbody>
</table>

* Estimate from 1999
## Genome sizes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycoplasma genitalium</td>
<td>470</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>1740</td>
</tr>
<tr>
<td>Methanococcus jannaschi</td>
<td>1,738</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>4,288</td>
</tr>
<tr>
<td>Sacharomyces cerevisiae - yeast</td>
<td>6,600</td>
</tr>
<tr>
<td>Drosophila melanogaster - fruit fly</td>
<td>13,600</td>
</tr>
<tr>
<td>Caenorhabditis elegans - worm</td>
<td>19,000</td>
</tr>
<tr>
<td>Arabidopsis thaliana - mustard</td>
<td>26,735</td>
</tr>
<tr>
<td>Oryza sativa - rice</td>
<td>50,000</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>*100,000</td>
</tr>
</tbody>
</table>

* Estimate from 1999
Manual for human

not Jan 1, 2000

June 27, 2000

Genetic Code of Human Life Is Cracked by Scientists

By NICHOLAS WADE

WASHINGTON, June 26 -- In an achievement that represents a pinnacle of human self-knowledge, two rival groups of scientists said today that they had deciphered the hereditary script, the set of instructions that defines the human organism.

"Today we are learning the language in which God created life.

© Burkhard Rost

© ROSTLAB.
Manual for human

☐ not Jan 1, 2000

☐ number of genes/proteins:
   Oct 1999
   (after >5 years):
   100,000
Manual for human

☐ not Jan 1, 2000

☐ number of genes/proteins:
  Oct 1999
  (after >5 years):
  100,000

☐ Nov 1999:
  oops there are only 30,000
Degree of novelty: biology vs. physics

**Human proteins**

<table>
<thead>
<tr>
<th>Year</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>100,000</td>
</tr>
<tr>
<td>2003</td>
<td>30,000</td>
</tr>
<tr>
<td>2015</td>
<td>20,000</td>
</tr>
</tbody>
</table>

**Speed of light**

<table>
<thead>
<tr>
<th>Who</th>
<th>Year</th>
<th>Value [km/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huygen</td>
<td>1690</td>
<td>220,000</td>
</tr>
<tr>
<td>Foucaul</td>
<td>1862</td>
<td>298,000</td>
</tr>
<tr>
<td>CGPM</td>
<td>1972</td>
<td>299,792</td>
</tr>
</tbody>
</table>

---

**novelty in biology**

**novelty in physics**

© Burkhard Rost

ROSTLAB.

TUM
## Genome sizes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycoplasma genitalium</td>
<td>470</td>
</tr>
<tr>
<td>Haemophilus influenzae</td>
<td>1,740</td>
</tr>
<tr>
<td>Methanococcus jannaschi</td>
<td>1,738</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>4,288</td>
</tr>
<tr>
<td>Sacharomyces cerevisiae - yeast</td>
<td>6,600</td>
</tr>
<tr>
<td>Drosophila melanogaster - fruit fly</td>
<td>13,600</td>
</tr>
<tr>
<td>Caenorhabditis elegans - worm</td>
<td>*21,000</td>
</tr>
<tr>
<td>Arabidopsis thaliana - mustard</td>
<td>*26,735</td>
</tr>
<tr>
<td>Oryza sativa - rice</td>
<td>*50,000</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>*25,000</td>
</tr>
</tbody>
</table>

* Estimates from 2010
Descriptive definition of life

- **Homeostasis** (regulation of internal environment to maintain constant state)
- **Organization** - unit: cells
- **Metabolism** (transfer of energy)
- **Growth**
- **Adaptation**
- **Response to stimuli**
- **Reproduction**

Viruses = life?
How many genes do viruses have?
Mimivirus

Mimivirus

Hitlist in 2013 | Nprotein
---|---
Megavirus chilensis | 1100
Mamavirus | 1023
Mimivirus | 979
Mimivirus M4 | 620
Cafeteria roenbergenensis virus | 544


© Wikipedia
Given manual: now we understand human?

Genetic Code of Human Life Is Cracked by Scientists

By NICHOLAS WADE

WASHINGTON, June 26 -- In an achievement that represents a pinnacle of human self-knowledge, two rival groups of scientists said today that they had deciphered the hereditary script, the set of instructions that defines the human organism.

"Today we are learning the language in which God created life."
Now we know it all?

Like for every good manual: you hardly ever find what you look for when you find it, it is difficult to understand!

Anna Tramontano, La Sapienza Rome, Italy

http://static.open.salon.com

http://i42.tinypic.com
Central dogma

DNA Polymerase

replication (DNA -> DNA)

DNA

transcription (DNA -> RNA)

RNA Polymerase

RNA

translation (RNA -> Protein)

Ribosome

Protein

Function

Structure

slide: Andrea Schafferhans
© Burkhard Rost
ROSTLAB.
Central dogma of molecular biology

DNA → RNA → Protein

information, code, library, manual

intermediate step

machinery of life
The Transcriptional Landscape of the Mammalian Genome

Science 2005 309:1559-63
The Transcriptional Landscape

Numbers:
- Total transcripts: 181,047
- New protein-coding transcripts
- New proteins: 5,154
- Multiple splice variants: 65%
- 1.35 5’ start sites for each 3’ end
- 1.83 3’ ends for each 5’ end
Central dogma of molecular biology

DNA → RNA → Protein

Still the central dogma, but we know that reality is more complicated
What are proteins?
Protein functions

- Defense (e.g. antibodies)
- Structure (e.g. collagen)
- Enzymes – metabolism, catabolism
- Communication / Signaling (e.g. insulin)
- Ligand binding / Transport (e.g. hemoglobin)
- Storage (e.g. ferritin)
Gallery of proteins 1

slide: Andrea Schafferhans

© Burkhard Rost

ROSTLAB.

111/128
Gallery of proteins 2
Protein sequence

>gi|16128674|ref|NP_415226.1| potassium translocating ATPase, subunit A [Escherichia coli K12]
MAAQGFLLIATFLLVLMVLARPLGSGLARLINDIPLPGTTGVERVLFRALGVSDREMNWKQYLCAILGLNMLGLAVLFFMLLGQHYLPLNPQQPLPGLSWDLALNTAVSFVTNTNWQSYSGETTLSYFSQMAGLTQVNFLSAASGIAVIFALIRAFTQRSMSTLGNAWVDLLRITLWVLVPVALLIALFFIQQQGALQNFLPYQAVNTVEGAQQLLPMGPVASQEAIKMLGTNGGFFNANSHPFENPTALTNFQMLAIFlIPTALCFAFGEVMGDRRQGRMLIWAMSVIFVICVGVVMWAEVQGNPHLLALGTDSSINMEGKESRFGVLVSSLFAVTTAASCAGAVIAMHDSFTALGGMVPMWLMQIGEVVFGGVGSGLYGMMLFVLLAVFIAGLMIGRTPEYLGKKIDVREMKTALALILVTPTLVLMAALAMMTDAHSMLNPQPHGFSEVLYAVSSAANNNGSAFAGLSANSFPWNCLLAFCMFVGRFGVIIIPVMAIAAGSLVSKKSQAASSGTLPTHGPLFVGLLIGTVLLL VGA LTFIPALALGPVAEYLS
Some facts about proteins

- how many in human?
  - 20-25K in human

- how long are they?
  - ~35-30k, median around 400

- do they consist of units?
  - most proteins have more than 2 domains

- how many proteins known?
Bio-sequence data explodes

Number of entries in UniProtKB/TrEMBL

© Uniprot EBI Cambridgeshire
**Big data - big CPU**

<table>
<thead>
<tr>
<th></th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/09</td>
<td>$8 \times 10^6$</td>
</tr>
<tr>
<td>2014/04</td>
<td>$55 \times 10^6$</td>
</tr>
</tbody>
</table>

- Double every 2 years
- 6.9 fold in 55 months
# Big data - big CPU

<table>
<thead>
<tr>
<th></th>
<th>Proteins</th>
<th>comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/09</td>
<td>$8 \cdot 10^6$</td>
<td></td>
</tr>
<tr>
<td>2014/04</td>
<td>$55 \cdot 10^6$</td>
<td></td>
</tr>
</tbody>
</table>

- **Cytochrome P450 - Mentha piperita - peppermint**
  - MELQLWSALILWVTISSLINOWRKPKPOGKFPFLGFHLHLWGGKLPOHALASVAKEYGPVAHVQVLGEVFSVVLSSREATKEAMKLVDPCANRFESIGTRIMWYDNEDII

- **Cytochrome P450 - Mentha spicata - spearmint**
  - MELDLLSAIILVATYIVSLLINQWRKSKSQNLPSSPPKLVPVIGHLHLWGGLPQHVFRSIAQKYGPAHVQVLGEVYSVVLSSAEAAKQAMKVLDFADRDGIQSRIMWYDKDDII

---

*double every 2 years*
# Big data - big CPU

## Proteins

<table>
<thead>
<tr>
<th>Year</th>
<th>Proteins</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/09</td>
<td>8 \times 10^6</td>
<td>64 \times 10^{12}</td>
</tr>
<tr>
<td>2014/04</td>
<td>55 \times 10^6</td>
<td>3025 \times 10^{12}</td>
</tr>
</tbody>
</table>

- **Double every 2 years**
- **>20-fold every 2 years**

- **47 fold**
- **55 months**
### Big data - big CPU

<table>
<thead>
<tr>
<th>Year</th>
<th>Proteins</th>
<th>Comparisons</th>
<th>We Need</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/09</td>
<td>$8 \cdot 10^6$</td>
<td>$64 \cdot 10^{12}$</td>
<td>5 TB</td>
</tr>
<tr>
<td>2014/04</td>
<td>$55 \cdot 10^6$</td>
<td>$3025 \cdot 10^{12}$</td>
<td>300 TB</td>
</tr>
</tbody>
</table>

- Proteins: 8 million proteins in 2009/09, 55 million proteins in 2014/04, an increase of double every 2 years.
- Comparisons: 64 trillion comparisons in 2009/09, 3025 trillion comparisons in 2014/04, an increase of 20-fold every 2 years.
- We need: 5 TB in 2009/09, 300 TB in 2014/04, an increase of >5-fold every 2 years.
Protein WITH annotations grow less

Number of entries in UniProtKB/Swiss-Prot

© Swiss-Prot Geneva/Lausanne
Raw protein sequence data vs. annotations

Number of entries in UniProtKB/TrEMBL

© Uniprot EBI Cambridgeshire
PDB: Proteins with experimentally known 3D structure

© PDB Rutgers Univ

© Burkhard Rost

ROSTLAB.
PDB: known structure ~ 0.2% of sequences

UniProt=all (0.2%)

Swiss-Prot=Annotation (20%)

PDB=3D
Proteins = gene products
= machinery of life

From the book: “DNA: The Secret of Life” by James Watson and Andrew Berry
Proteins - genetic code

From the book: “DNA: The Secret of Life” by James Watson and Andrew Berry

http://img.sparknotes.com/figures/1/132e5720f3b37836571a1aeb9d163ac3/genetic_code.gif
Kingdoms similar in amino acids usage

organisms DO differ in the codon usage

\begin{align*}
\text{nucleotide} & \quad \text{amino acid} \\
\text{GCT} & \quad \text{A} \\
\text{GCA} & \quad \rightarrow \\
\text{GCC} & \\
\text{GCG} &
\end{align*}
Lecture plan (PP1 structure)

01: 04/25 Tue: no lecture
02: 04/27 Thu: no lecture
03: 05/02 Tue: Organization of lecture: intro into cells & biology
04: 05/04 Thu: Intro I - acids/structure - domains
05: 05/09 Tue: Possibly no lecture - Alignment 1
06: 05/11 Thu: Alignment 2
07: 05/16 Tue: Alignment 3
08: 05/18 Thu: Comparative modeling & experimental structure determination & secondary structure assignment
09: 05/23 Tue: SKIP: student assembly (SVV)
10: 05/25 Thu: SKIP: Ascension Day
11: 05/30 Tue: SKIP: Whitsun holiday (05/15-17)
12: 06/01 Thu: 1D: Secondary structure prediction 1
13: 06/06 Tue: SKIP: Whitsun holiday (06/03-06)
14: 06/08 Thu: 1D: Secondary structure prediction 2
15: 06/13 Tue: 1D: Secondary structure prediction 3 / Transmembrane structure prediction 1
16: 06/15 Thu: SKIP: Corpus Christi
17: 06/20 Tue: 1D: Transmembrane structure prediction 2 / Solvent accessibility prediction
18: 06/22 Thu: 1D: Disorder prediction
19: 06/27 Tue: 2D prediction
20: 06/29 Thu: 3D prediction / Nobel prize symposium
21: 07/04 Tue: TBA
22: 07/06 Thu: recap 1
23: 07/11 Tue: recap 2
24: 07/12 Thu: examen
25: 07/13 Tue: TBA
26: 07/18 Thu: TBA
27: 07/20 Tue: TBA
28: 07/22 Thu: TBA
29: 07/25 Thu: TBA