Announcements

- Videos: SciVe / www.rostlab.org
- THANKS:
  - Tim Karl + Manfred Roos
- Special lectures:
  - Andrea Schafferhans: May 8, 10 & 22
- NO lectures (not final):
  - holidays: May 1, 17, 29, June 07
  - Student Assembly: May 15
  - June 14
- LAST lecture: Jul 5
- Examen: Jul 12, 10:30 (likely this room)
  - Makeup: likely: Oct 18 - afternoon
Lecture plan (PP1: Structure)

01: 2012/04/17: skipped
02: 2012/04/19: welcome: who we are
03: 2012/04/24: intro I - acids/structure
04: 2012/04/26: INSERT - Machine learning in biology (cross-validation asf) - TMH1
05: 2012/05/01: holiday (May 1)
06: 2012/05/03: INSERT 2 - TMH2
07: 2012/05/08: Andrea Schafferhans: intro II - 3D comparisons: Andrea Schafferhans
08: 2012/05/10: Andrea Schafferhans: Comparative Modeling 1
09: 2012/05/15: no lecture: student assembly
10: 2012/05/17: holiday (Ascension Day)
11: 2012/05/22: Andrea Schafferhans: Comparative Modeling 2
12: 2012/05/24: Alignment 1
13: 2012/05/29: holiday (Pentecost/Whitsun)
14: 2012/05/31: Alignment 2
15: 2012/06/05: intro III - 3D->1D: sec str
16: 2012/06/07: holiday (Corpus Christi)
17: 2012/06/12: sec str pred 1 (white board)
18: 2012/06/14: no lecture
19: 2012/06/19: sec str pred 2
20: 2012/06/21: sec str pred 3
21: 2012/06/26: transmembrane helix prediction
22: 2012/06/28: transmembrane strand prediction, solvent accessibility
23: 2012/07/03: 3D prediction
24: 2012/07/05: summary: what we do in our group
25: 2012/07/10: no lecture
26: 2012/07/12: no lecture
27: 2012/07/17: no lecture
28: 2012/07/24: no lecture
29: 2012/07/26: no lecture
Protein Prediction - Part 1: Structure

1 Introduction
Marco Punta contributed the slides

- PhD in Trieste (MD for membrane proteins)
- Postdoc @ Columbia Univ in the City of New York (contact predictions)
- Senior scientist in NYCOMPS (Target selection for membrane proteins)
- IAS Fellow @ TUM
- Project manager @ Pfam @ Sanger Inst. Hinxton (Cambridgeshire)
TOC today

- Protein introduction
  - amino acids
  - protein structure
  - bonds & energies

- NEXT WEEK (tuesday/thursday)
  - structure comparison

- 2 WEEKs (tuesday)
  - structure comparison
Life: The Players
Life is diverse

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

Tuesday April 24, 2012
How does life work?

Leonardo Da Vinci (1452-1519)
Vitruve Luc Viatour (~1492)
Central dogma of molecular biology

DNA → RNA → Protein

information, code, library, manual

intermediate step

machinery of life

© Laszlo Kajan, TUM
Central dogma of molecular biology

DNA  RNA  Protein
information, code, library, manual
intermediate step
machinery of life

© Laszlo Kajan, TUM
Haemophilus Influenzae
(Pfeiffer’s bacillus)
Haemophilus Influenzae  
(Pfeiffer’s bacillus)  

Mycoplasma Genitalium  
Haemophilus Influenzae
(Pfeiffer’s bacillus)

Mycoplasma Genitalium

Sacharomyces cerevisiae
Haemophilus Influenzae
(Pfeiffer’s bacillus)

Mycoplasma Genitalium

Sacharomyces cerevisiae

Caenorhabditis elegans, nematode - worm
Entire Manual / Library

- **Haemophilus Influenzae** (Pfeiffer’s bacillus)
  - 40 authors

- **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 - yeast</em></td>
<td>6,600</td>
</tr>
<tr>
<td><em>Drosophila melanogaster - fruit fly</em></td>
<td>13,600</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans - worm</em></td>
<td>19,000</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana - mustard</em></td>
<td>26,735</td>
</tr>
<tr>
<td><em>Oryza sativa - rice</em></td>
<td>50,000</td>
</tr>
</tbody>
</table>

* Estimate from 1999
## 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>1740</td>
</tr>
<tr>
<td>Methanococcus jannaschi</td>
<td>1,738.00</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>4,288.00</td>
</tr>
<tr>
<td>Sacharomyces cerevisiae - yeast</td>
<td>6,600.00</td>
</tr>
<tr>
<td>Drosophila melanogaster - fruit fly</td>
<td>13,600.00</td>
</tr>
<tr>
<td>Caenorhabditis elegans - worm</td>
<td>19,000.00</td>
</tr>
<tr>
<td>Arabidopsis thaliana - mustard</td>
<td>26,735.00</td>
</tr>
<tr>
<td>Oryza sativa - rice</td>
<td>50,000.00</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>*100,000</td>
</tr>
</tbody>
</table>

* Estimate from 1999
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."
not Jan 1, 2000
Manual for human

☐ not Jan 1, 2000

Science

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."

Top, Dr. Francis Collins and J. Craig Venter joined President Bill Clinton at the White House on Monday to announce the completion of the first draft of the human genome.
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

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"
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
## 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>Saccharomyces cerevisiae - yeast</em></td>
<td>6,600</td>
</tr>
<tr>
<td><em>Drosophila melanogaster - fruit fly</em></td>
<td>13,600</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans - worm</em></td>
<td>*21,000</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana - mustard</em></td>
<td>*26,735</td>
</tr>
<tr>
<td><em>Oryza sativa - rice</em></td>
<td>*50,000</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>*25,000</td>
</tr>
</tbody>
</table>

* Estimates from 2010
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.

http://static.open.salon.com

http://i42.tinypic.com
Central dogma of molecular biology

DNA → RNA → Protein

information, code, library, manual

intermediate step

machinery of life
FANTOM 3: Functional annotation of mouse

Science 2005 309:1559-63
The Transcriptional Landscape of the Mammalian Genome

Numbers:
- Total transcripts: 181,047
- New protein-coding transcripts: 16,247
- 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

- DNA: information, code, library, manual
- RNA: intermediate step
- Protein: machinery of life

Still the central dogma, but we know that reality is more complicated
Can we compute life?

Tough challenge, we address it from all possible angles

But
- We still do not understand the “book”/genome
- We don’t even understand the first principles, completely

Nevertheless
New life in the test tube!

Craig Venter Institute

first synthesis of an entire organism

Mycoplasma genitalium

Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Gibson et al. Science 319(5867): 1215-1220. 2008

© Burkhard Rost (TU Munich)
We can dream but we cannot compute life, yet
Reality and images
Georges Braque - Houses at L'Estaque
Illustration by David S. Goodsell, the Scripps Research Institute, UCSD, USA
Mycoplasma genitalium

Protein synthesis (labels in black)
1. DNA
2. DNA polymerase
3. single-stranded-DNA binding protein (protects single-stranded portions during replication)
4. RNA polymerase
5. messenger RNA
6. ribosome
7. transfer RNA (in pink) and elongation factor Tu and Ts
8. elongation factor G
9. aminoacyl-tRNA synthetases
10. topoisomerases
12. Rec system for DNA repair: a) RecA, b) RecBC
13. chaperonin GroEL (helps folding of new proteins)
14. proteasome ClpA (destroys old proteins)

Enzymes for energy production (labels in red)
15. glycolytic enzymes
16. pyruvate dehydrogenase complex

Membrane proteins (labels in blue)
17. ATP synthase
18. secretory proteins
19. sodium pump
20. zinc transporter
21. magnesium transporter
22. ABC transporter (different ABC transporters transport different types of molecules-ABC is short for "ATP-binding cassette"
23. magnesium transporter
24. lypoglycan (long carbohydrate chains connected to lipid in the membrane)

Illustration by David S. Goodsell, the Scripps Research Institute, UCSD, USA

© Burkhard Rost (TU Munich)
Eukaryotic cell

Illustration by David S. Goodsell, the Scripps Research Institute, UCSD, USA
Doyle et al. (1998) Science **280**:69-77 - The structure of the potassium channel: molecular basis of K+ conduction and selectivity
Umberto Boccioni - Dynamism of a soccer player
Two views of a protein structure

(a) and (b) show different perspectives of a protein structure.

Wu et al. unpublished
Different levels of abstraction

(a) [Diagram of a molecular structure with labels Ser132, Phe 34, and others]

(b) [Diagram of a molecular structure with a network of lines]

Wu et al. unpublished

Photograph: Filippo Monteforte/AFP/Getty Images

Marco Punta, TU München 2010
© Burkhard Rost (TU Munich)
What are proteins
Protein sequence

>gi|16128674|ref|NP_415226.1| potassium translocating ATPase, subunit A [Escherichia coli K12]
MAAQGFLLIATFLLVLMVLRPLGSGLARLINDIPLPGTTGVERVLFRALGVSDREMNWK
QYLCAILGLNMLGLAVLFFMLLGQHYLPLNPQQLPGGLSWDLALNTAVSFVTNTNWQSYSG
ETTLSYFSQMAGLTQVNFLSAASGIAVIFALIRAFTQRSMTSLGNAWVDLLRITLWLVP
VALLIALFIIQQGALQNLPYQAQNTVEGAAQQLLPMPGVASQEAIKMLGTMGGGFFNANS
SHPFENPTALTNFVQLAIFLIPTALCFAFGEVMGDRQRQRMLLWAMSVIFVICVGVMW
AVEVGNPHLLALGTDSINMEGKESRFGVLVSSLFAVVTAAEASCGAVIAMHDSFTALGGM
VPMWLMQIGEVVGGVGSLGMYGMMLFVLLAVFIALGLMGRTPFELGKGKIDVREMKLTALA
ILVTPTLVLMGAALAMMTDAGRSAMLNPQPHGFSEVLAYVSSAAANNGGSAFAGLSANSPF
WNCLLAFCMFVGRFGVIIPVMAIGSLVSKKSQAASSGTLPHTGPFVGLLLIGTVVLLVGA
LTFIPALALGPVAEYLS
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
Proteins = gene products = machinery of life

From the book: “DNA: The Secret of Life” by James Watson and Andrew Berry
Constituents of proteins: amino acids
Amino acid

side-chain

backbone
Joining amino acids into proteins

isolated amino acid

side-chain

backbone
Joining amino acids into proteins

a dipeptide

From Wikipedia
Joining amino acids into proteins

A dipeptide

From Wikipedia

http://www.webchem.net/notes/chemical_bonding/covalent_bonding.htm
Joining amino acids into proteins

polypeptide chain
Joining amino acids into proteins

side-chain

backbone
Joining amino acids into proteins
Joining amino acids into proteins

Joining amino acids into proteins
Joining amino acids into proteins
Joining amino acids into proteins
Joining amino acids into proteins
Joining amino acids into proteins
Rationalizing biophysical features of constituents
Side chain properties

- Alanine
- Arginine
- Aspartic acid
- Asparagine
- Glutamic acid
- Glutamine
- Glycine
- Cysteine
- Histidine
- Isoleucine
- Leucine
- Lysine
- Methionine
- Phenylalanine
- Proline
- Serine
- Threonine
- Trpophan
- Tyrosine
- Valine
Positively charged amino acids

- Alanine
- Aspartic Acid
- Asparagine
- Glutamic Acid
- Glutamine
- Histidine
- Isoleucine
- Leucine
- Lysine
- Methionine
- Phenylalanine
- Proline
- Serine
- Threonine
- Tryptophan
- Tyrosine
- Valine

© Burkhard Rost (TU Munich)
Negatively charged amino acids

- Aspartic Acid
- Glutamic Acid

Ala
Arg
Asn
Glu
Cys
Gly
His
Ile
Leu
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val
Polar amino acids

N  Q  S  T  Y

Alanine  Arginine  Asparagine  Aspartic Acid  Glutamine  Glutamic Acid  Histidine  Isoleucine  Leucine  Lysine  Methionine  Phenylalanine  Proline  Serine  Threonine  Trypsphan  Tyrosine  Valine
amino acids
Some components of the energetics of proteins
Non-covalent interactions in proteins

<table>
<thead>
<tr>
<th>Covalent (electron sharing)</th>
<th>Hydrogen bond</th>
<th>Van der Waals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/10</td>
<td>1/100-1/1000</td>
</tr>
</tbody>
</table>
Non-covalent interactions in proteins

Covalent (electron sharing)  <  Hydrogen bond  <  Van der Waals

1  <  1/10  <  1/100-1/1000

From web.njit.edu
Non-covalent interactions in proteins

Covalent (electron sharing) \( < \) Hydrogen bond \( < \) Van der Waals

1 \( \frac{1}{10} \) 1/100-1/1000

From Wikipedia
Salt bridges

Arg, Lys, (His) -> Glu, Asp
Salt bridges

Arg, Lys, (His) -> Glu, Asp

π-π interactions


Trp, Tyr, Phe -> Trp, Tyr, Phe
π-π interactions

PDBid: 2eve

Trp, Tyr, Phe -> Trp, Tyr, Phe
cation-π interactions

Arg, Lys -> Trp, Tyr, Phe

Disulfide bridges

Cys -> Cys

R
SH

oxidation

R
SH

H
H

R

+ 2H⁺ + 2e⁻

From Wikipedia, author: Benjah-bmm27

From Wikipedia, author: Marco Punta, TU München 2010

© Burkhard Rost (TU Munich)
“components” of protein structure: secondary structure
Protein structure
Protein structure
Protein secondary structure
Protein secondary structure: helix
Protein secondary structure: helix
Protein secondary structure: helix

from 3goe, structure at 0.97 Å resolution
Protein secondary structure: helix

from 3goe, structure at 0.97 Å resolution
Protein secondary structure: sheets

from 3goe, structure at 0.97 Å resolution
Protein secondary structure: sheets

from 3goe, structure at 0.97 Å resolution
Secondary structure: Ramachandran

from: http://boscoh.com/protein/the-mysterious-regions-of-the-ramachandran-plot

Ramachandran plot (Ramachandran, 1963, JMB 7:99)
GN Ramachandran (Gopalasamudram Narayana Iyer Ramachandran)

- 8 October 1922 - 7 April 2001
- 1942: Electrical engineering -> physics (Master with Raman) / optics
- 1949: PhD: Cavendish Lab in Cambridge: crystallography
- 1952: Madras: shift to biological molecules: triple helical structure of collagen
- 1963 (JMB): analysis of backbone confirmations: Ramachandran plots
Secondary structure: Ramachandran

Ramachandran plot (Ramachandran, 1963, JMB 7:99)

from: http://boscoh.com/protein/the-mysterious-regions-of-the-ramachandran-plot
Secondary structure: Ramachandran

Ramachandran plot (Ramachandran, 1963, JMB 7:99)

from: http://boscoh.com/protein/the-mysterious-regions-of-the-ramachandran-plot
Secondary structure: Ramachandran

From Wikipedia
Author: Dcrjsr

from: http://boscoh.com/protein/the-mysterious-regions-of-the-ramachandran-plot

Ramachandran plot (Ramachandran, 1963, JMB 7:99)
Secondary structure: Ramachandran

from: http://boscoh.com/protein/the-mysterious-regions-of-the-ramachandran-plot

Ramachandran plot (Ramachandran, 1963, JMB 7:99)
Secondary structure: Ramachandran

Ramachandran plot (Ramachandran, 1963, JMB 7:99)

Ho and Brasseur BMC Structural Biology 2005 5:14

Marco Punta, TU München 2010
© Burkhard Rost (TU Munich)
Protein structure
Protein secondary structure: sheets

from 3goe, structure at 0.97 Å resolution
Protein structure

A

B

Ho and Brasseur BMC Structural Biology 2005 5:14
Protein structure
Protein structure: cartoon
Protein structure cartoons
Protein structure cartoons
“components” of protein structure: domains
Domains
Domains