Bioinformatics 1
Biology, Sequences, Phylogenetics
Part 1
Sepp Hochreiter
Master Bioinformatics

→ 3 credits (plus 3 credits for exercises)

→ first basic course: module M6 (required for other modules)
Biological Chemistry

2.6 credits (plus 3.2 credits for exercises)

Module “informatics”
Examination at the end of the class

Course manuscript:
http://www.bioinf.jku.at/teaching/ws2012/bin1/

Also videos
Blocked class (see KUSSS):

Course 347391: “Introduction into Instrumental Analytics for Life-Sciences”

First meeting Molecular Biology at 2.10.2012:
320007, VO Molekulare Biologie der Zelle I
320061, UE Molekulare Biologie der Zelle I
<table>
<thead>
<tr>
<th>Time</th>
<th>Montag/Monday</th>
<th>Dienstag/Tuesday</th>
<th>Mittwoch/Wednesday</th>
<th>Donnerstag/Thursday</th>
<th>Freitag/Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30-9:15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:15-10:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:15-11:00</td>
<td>366554</td>
<td></td>
<td>365062</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Statistik 2</td>
<td></td>
<td>Sequence Analysis and Phylogenetics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2VO</td>
<td></td>
<td>2UE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3 047</td>
<td></td>
<td>S3 047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00-11:45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00-12:45</td>
<td>344014</td>
<td></td>
<td>340023</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Artificial Intelligence</td>
<td></td>
<td>Algorithms and Data Structures 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2VO</td>
<td></td>
<td>2VO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS 5</td>
<td></td>
<td>HS 19 / HS 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Alternative: 326011, Thursday)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:45-14:30</td>
<td>340023</td>
<td></td>
<td>351001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Algorithms and Data Structures 2</td>
<td></td>
<td>Information Systems 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2VO</td>
<td></td>
<td>2VO</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HS 19 / HS 7</td>
<td></td>
<td>HS 1</td>
<td></td>
<td>347334</td>
</tr>
<tr>
<td></td>
<td>(Alternative: 326011, Thursday)</td>
<td></td>
<td>Information Systems 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2VO</td>
<td></td>
<td>2UE</td>
</tr>
<tr>
<td>14:30-15:15</td>
<td>351001</td>
<td></td>
<td>347334</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information Systems 1</td>
<td></td>
<td>Chemistry for Physist II</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2VO</td>
<td></td>
<td>2VO</td>
<td></td>
<td>HF 9904 / S3 055</td>
</tr>
<tr>
<td></td>
<td>HS 1</td>
<td></td>
<td>HS 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30-16:15</td>
<td>365060</td>
<td></td>
<td>351003/351004</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequence Analysis and Phylogenetics</td>
<td></td>
<td>Information Systems 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2VO</td>
<td></td>
<td>2UE</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2 059</td>
<td></td>
<td>T 911 / S3 055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:15-17:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bioinformatics 1: Biology, Sequences, Phylogenetics
What is Bioinformatics?

- Interface of biology and computers
- Analysis of proteins, genes and genomes using computer algorithms and computer databases
- Analysis and storage of the billions of DNA base pairs that are sequenced by genomics projects
What is Bioinformatics?

Computer Science

Bioinformatics

Biology & Life Sciences

Mathematics & Statistics
What is Bioinformatics?

“Bioinformatics is a new subject of genetic data collection, analysis and dissemination to the research community.”

_Hwa A. Lim (1987)_

“Bioinformatics: Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.”

_NIH working definition (2000)_
What is Bioinformatics?

http://en.wikipedia.org/wiki/Bioinformatics

- creation and advancement of *databases*, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the *management* and analysis of biological data.  
  \( \rightarrow \) BI IV

- mapping and analyzing DNA and protein sequences, *aligning* different DNA and protein sequences to compare them and *creating and viewing 3-D models of protein structures*.  
  \( \rightarrow \) BI III

- focus on developing and applying computationally intensive techniques (e.g., *pattern recognition*, *data mining*, *machine learning algorithms*, and *visualization*)  
  \( \rightarrow \) BI II

- Major research efforts in the field include *sequence alignment*, *gene finding*, *genome assembly*, *protein structure alignment*, *protein structure prediction*, *prediction of gene expression and protein-protein interactions*, *genome-wide association studies* and *the modeling of evolution*.  
  \( \rightarrow \) BI III
Questions Answered by Bioinformatics

- From where came the first human?
- Is Anna Anderson the tsar’s daughter Anastasia?
- Are the neanderthals the ancestors of the humans?
- What are the evolutionary relationships between species?
A) Not from Africa: Homo erectus is human ancestor

B) From Africa: Competition with Homo erectus
Is Anna Anderson the tsar’s daughter Anastasia Romanov?

Anastasia (1909) and Anna Anderson (1959)

The kids of the tsar
Are the neanderthals the human ancestors or a different species?
Phylogeny: history of species

Phylogenetic knowledge: evolutionary trees

mammals
snakes
saurians
crocodiles
birds
turtles

Bioinformatics 1: Biology, Sequences, Phylogenetics
Three Answers

1. From where came the first human?
   - Africa!

2. Is Anna Anderson the tsar’s daughter Anastasia?
   - No!

<table>
<thead>
<tr>
<th>Anna Anderson</th>
<th>Carl Maucher (Grand nephew F. Schanzkowska)</th>
<th>Prince Philip (Grand nephew zar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCACCATGAATATTGC</td>
<td>CCACCATGAATATTGC</td>
<td>TCACCATGAATATTGT</td>
</tr>
<tr>
<td>TAGTCAAATCCCTT</td>
<td>TAGTCAAATCCCTT</td>
<td>CAGTCAAATCCCTC</td>
</tr>
<tr>
<td>91</td>
<td>106</td>
<td>324</td>
</tr>
<tr>
<td>337</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Are the neanderthals the ancestors of the humans?
   - No! Separate Species

<table>
<thead>
<tr>
<th>Modern Humans</th>
<th>Neanderthals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bioinformatics 1: Biology, Sequences, Phylogenetics
Literature

D. W. Mount, Bioinformatics: Sequences and Genome analysis, CSHL Press, 2001
D. Gusfield, Algorithms on strings, trees and sequences: computer science and computational biology, Cambridge Univ. Press, 1999
M. Waterman, Introduction to Computational Biology, Chapmann & Hall, 1995
Setubal and Meidanis, Introduction to Computational Molecular Biology, PWS Publishing, 1997
Pevzner, Computational Molecular Biology, MIT Press, 2000
J. Felsenstein: Inferring phylogenies, Sinauer, 2004
Contents

1 Biological Basics
1.1 The Cell
1.2 Central Dogma of Molecular Biology
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Introns, Exons, and Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
Contents

2 Bioinformatics Resources
  2.1 Data Bases
  2.2 Software
  2.3 Articles
3 Pairwise Alignment
3.1 Motivation
3.2 Sequence Similarities and Scoring
3.2.1 Identity Matrix
3.2.2 PAM Matrices
3.2.3 BLOSUM Matrices
3.2.4 Gap Penalties
3.3 Alignment Algorithms
3.3.1 Global Alignment - Needleman-Wunsch
3.3.2 Local Alignment - Smith-Waterman
3.3.3 Fast Approximations: FASTA, BLAST and BLAT
3.4 Alignment Significance
3.4.1 Significance of HSPs
3.4.2 Significance of Perfect Matches
Contents

4 Multiple Alignment
4.1 Motivation
4.2 Multiple Sequence Similarities and Scoring
  4.2.1 Consensus and Entropy Score
  4.2.2 Tree and Star Score
  4.2.3 Weighted Sum of Pairs Score
4.3 Multiple Alignment Algorithms
  4.3.1 Exact Methods
  4.3.2 Progressive Algorithms
  4.3.3 Other Multiple Alignment Algorithms
4.4 Profiles and Position Specific Scoring Matrices
5 Phylogenetics
5.1 Motivation
5.1.1 Tree of Life
5.1.2 Molecular Phylogenies
5.1.3 Methods
5.2 Maximum Parsimony Methods
5.2.1 Tree Length
5.2.2 Tree Search
5.2.3 Weighted Parsimony and Bootstrapping
5.2.4 Inconsistency of Maximum Parsimony
5.3 Distance-based Methods
5.3.1 UPGMA
5.3.2 Least Squares
5.3.3 Minimum Evolution
5.3.4 Neighbor Joining
5.3.5 Distance Measures
5.4 Maximum Likelihood Methods
5.5 Examples
Bioinformatics processes data from molecular biology

Molecular biology attempts at discovering the principles of the cell which is the largest unit all lifeforms have in common
The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
• **nucleus**: eukaryotes - a cell's information center → chromosomes, DNA

• **nucleolus**: ribosome subunits assembly

• **mitochondria** and **chloroplasts**: the power generators (oxidative phosphorylation, photosynthesis); own genome from mother

• **endoplasmic reticulum** (ER): eukaryotes - transport network for certain modifications and specific destinations; rough ER has ribosomes on its surface and secretes proteins into the cytoplasm

• **golgi apparatus**: eukaryotes - process & pack macromolecules; vesicles

• **ribosomes**: RNA/protein complex that synthesizes proteins from mRNA

• **lysosomes**: enzymes (acid hydrolases) that digest old organelles, food particles, and engulfed viruses or bacteria

• **peroxisomes**: rid the cell of toxic peroxides

• **centrosome**: cytoskeleton, microtubules for directing the transport through ER and golgi apparatus

• **vacuoles**: store food and waste
Eukaryotic cells possess a nucleus (plants, vertebrates)

Prokaryotic cells do not possess a nucleus (bacteria, archaea)
The Cell

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
The Cell

Pathway glycogen
Pathway hormone
Central Dogma

How are the nano-machines in the cell constructed?

- These machines are proteins or protein-RNA complexes

Where is the information about these machines stored?

- Everything is stored in the DNA

How is the information in the DNA used to build proteins?

- Central dogma: DNA → RNA → Proteins
1. **transcription** (mRNA)
2. transport
3. **translation** (ribosom, tRNA)
4. folding (protein)

codons/basetriplets

Amino acid
Central Dogma

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
Deoxyribonucleic acid (DNA) codes all information of life

> double helix as sequence of nucleotides with a deoxyribose

> ends are called 5' and 3'; DNA is written from 5' to 3'

> upstream is towards the 5' end downstream towards the 3'

> 5 nucleotides (nucleobases, bases): adenine (A), thymine (T), cytosine (C), guanine (G), and uracil (U)

> first 4 in DNA whereas uracil in RNA instead of thymine

> two classes: purines (A, G) / pyrimidines (C, U, T)
DNA

- hydrogen bonds between purines and pyrimidines
- base pairs: A—T and C—G
- each helix of the DNA is complementary to the other
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
1.3 DNA

- Nucleotides are the building blocks of DNA.
- There are four types of nucleotides: Adenine (A), Thymine (T), Guanine (G), and Cytosine (C).
- DNA consists of two polynucleotide chains that are antiparallel.
- The chains are held together by hydrogen bonds between the bases.

1.5 Transcription

- Transcription is the process by which DNA is transcribed into RNA.
- It occurs in the nucleus.
- RNA polymerase enzyme is involved in transcription.

1.9 Translation

- Translation is the process by which RNA is translated into proteins.
- It occurs in the cytoplasm.
- Ribosomes are the sites of translation.

Bioinformatics 1: Biology, Sequences, Phylogenetics
Bioinformatics 1: Biology, Sequences, Phylogenetics
The DNA is condensed in the nucleus in the chromosomes.

DNA wraps around histones resulting in chromatin.

Two chromatins linked at the centromere are a chromosome.
Single DNA nucleotides differ at each human.

Small differences are inherited from both parents (except maternal mitochondrial DNA).

Variation in the DNA at the same position in at least 1% of the population: single nucleotide polymorphism (SNP -- pronounced snip).

SNPs occur all 100 to 300 base pairs.

Current research relate diseases to SNPs (schizophrenia or alcohol dependence).
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
RNA

- Ribonucleic acid (RNA): sequence of nucleotides

- Contrast to DNA: ribose rings instead of deoxyribose; uracil instead of thymine

- Transcribed from DNA through RNA polymerases

- Kind of RNA:
  - mRNA (messenger)
  - tRNA (codon coding)
  - dsRNA (double stranded)
  - miRNA (micro)
  - RNAi (interference)
  - siRNA (small interfering)
  - ncRNA (non-coding) like rRNA (ribosomal)
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
RNA
RNA

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA

1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
Transcription is the process of reading out a RNA (mRNA) from the DNA.
Transcription Initiation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
Transcription Initiation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
   1.5.1 Initiation
   1.5.2 Elongation
   1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
   1.9.1 Initiation
   1.9.2 Elongation
   1.9.3 Termination
1.10 Folding

Transcription

Bioinformatics 1: Biology, Sequences, Phylogenetics
Transcription Inhibition

Bioinformatics 1: Biology, Sequences, Phylogenetics
After 8 nucleotides the sigma-subunit is dissociated from polymerase.

For elongation there exist promoters.
Transcription Termination

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
Splicing, Exons and Introns

- Splicing modifies pre-mRNA released after transcription
- Non-coding sequences: introns (intragenic regions)
- Coding sequences: exons are glued together
- A snRNA complex, the spliceosome, performs the splicing but some RNA sequences can perform autonomous splicing
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
  1.5.1 Initiation
  1.5.2 Elongation
  1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
  1.9.1 Initiation
  1.9.2 Elongation
  1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
pre-mRNA can be spliced in different ways: alternative splicing, therefore a gene can code different proteins

Alternative splicing is controlled by signalling molecules
Amino Acids

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
Amino Acids

ALIPHATIC AMINO ACIDS

Glycine (Gly) G
Alanine (Ala) A
Valine (Val) V
Leucine (Leu) L
Isoleucine (Ile) I

AMINO ACIDS WITH HYDROXYL- OR SULFUR-CONTAINING SIDE CHAINS

Serine (Ser) S
Cysteine (Cys) C
Threonine (Thr) T
Methionine (Met) M
Proline (Pro) P

CYCLIC AMINO ACID

Bioinformatics 1: Biology, Sequences, Phylogenetics
Amino Acids

AROMATIC AMINO ACIDS

- Phenylalanine (Phe) F
- Tyrosine (Tyr) Y
- Tryptophan (Trp) W

BASIC AMINO ACIDS

- Histidine (His) H
- Lysine (Lys) K
- Arginine (Arg) R

ACIDIC AMINO ACIDS AND THEIR AMIDES

- Aspartic acid (Asp) D
- Glutamic acid (Glu) E
- Asparagine (Asn) N
- Glutamine (Gln) Q
### Amino Acids

#### Hydrophobic (nonpolar):
- glycine (Gly, G)
- alanine (Ala, A)
- valine (Val, V)
- leucine (Leu, L)
- isoleucine (Ile, I)
- methionine (Met, M)
- phenylalanine (Phe, F)
- tryptophan (Trp, W)
- proline (Pro, P)

#### Hydrophilic (polar):
- serine (Ser, S)
- threonine (Thr, T)
- cysteine (Cys, C)
- tyrosine (Tyr, Y)
- asparagine (Asn, N)
- glutamine (Gln, Q)

#### Acidic (-, hydrophilic):
- aspartic acid (Asp, D)
- glutamic acid (Glu, E)

#### Basic (+, hydrophilic):
- lysine (Lys, K)
- arginine (Arg, R)
- histidine (His, H)

Cysteine and methionine: disulfide bonds
Amino Acids

Glycine + Alanine → Glycyllalanine

Bioinformatics 1: Biology, Sequences, Phylogenetics
all proteins consist of these 20 amino acids

3D interactions of the amino acids results in nano-machines

Genetic code: instructions for producing proteins from DNA

Protein in coded through a gene which is transcribed into mRNA and then translated into an amino acid sequence which automatically configures into a protein

Genetic code gives the rules for translation

Rules are simple: 3 nucleotides (codon) = one amino acid

AUG and CUG: start codon
## Genetic Code

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU</td>
<td>phe</td>
<td>UAU</td>
<td>UGU</td>
</tr>
<tr>
<td></td>
<td>UUC</td>
<td></td>
<td>UAC</td>
<td>UGC</td>
</tr>
<tr>
<td></td>
<td>UUA</td>
<td>leu</td>
<td>UAA</td>
<td>UGA</td>
</tr>
<tr>
<td></td>
<td>UUG</td>
<td></td>
<td>UAG</td>
<td>UGG</td>
</tr>
<tr>
<td>C</td>
<td>CUU</td>
<td>leu</td>
<td>CAU</td>
<td>CGU</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td></td>
<td>CAC</td>
<td>CGC</td>
</tr>
<tr>
<td></td>
<td>CUA</td>
<td></td>
<td>CAA</td>
<td>CGA</td>
</tr>
<tr>
<td></td>
<td>CUG</td>
<td></td>
<td>CAG</td>
<td>CGG</td>
</tr>
<tr>
<td>A</td>
<td>AAU</td>
<td>ile</td>
<td>AAA</td>
<td>AGU</td>
</tr>
<tr>
<td></td>
<td>AUC</td>
<td></td>
<td>AAG</td>
<td>AGC</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td></td>
<td></td>
<td>AGA</td>
</tr>
<tr>
<td></td>
<td>AUG</td>
<td></td>
<td></td>
<td>AGG</td>
</tr>
<tr>
<td>G</td>
<td>GCU</td>
<td></td>
<td>GAU</td>
<td>GGU</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td></td>
<td>GAC</td>
<td>GGC</td>
</tr>
<tr>
<td></td>
<td>GCA</td>
<td></td>
<td>GAA</td>
<td>GGA</td>
</tr>
<tr>
<td></td>
<td>GCG</td>
<td></td>
<td>GAG</td>
<td>GGG</td>
</tr>
</tbody>
</table>

C = Cytosin, U = Uracil, A = Adenin, G = Guanin

Base pairs DNA: A-T and C-G (T = Thymin)

Bioinformatics 1: Biology, Sequences, Phylogenetics
Translation

After transcription the pre-mRNA is spliced, edited, transported out of the nucleus into the cytosol (eukaryotes)

The ribosome (protein production machinery) assembles the amino acid sequences out of the mRNA

Ribosome consists of two subunits 60S and 40S in eukaryotes and 50S and 30S in bacteria
Translation Initiation

- Inactive ribosomes have dissociated subunits
- Ribosome binds to site at mRNA marked by AGGAGGU (Shine-Dalgarno)
- At this site the initiation factors IF1, IF2, IF3 and 30S ribosomal subunit bind
- The initiator tRNA binds to the start codon
- Then the 50S subunit binds to the complex and translation can start
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Translation Elongation

1. Biological Basics
2. The Cell
3. Central Dogma
4. DNA
5. RNA
6. Transcription
7. Initiation
8. Elongation
9. Termination
10. Splicing
11. Amino Acids
12. Genetic Code
13. Translation
14. Initiation
15. Elongation
16. Termination
17. Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
Translation Elongation

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Ribosome
Termination by a stop codon (UAA, UAG, UGA) which enters the A-site

tRNAs cannot bind, however release factors bind at or near the ribosome

Amino acid chain is released and the 70S ribosome dissociates

30S subunit remains attached to the mRNA and searching for the next Shine-Dalgarno pattern
Translation Termination

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding

Bioinformatics 1: Biology, Sequences, Phylogenetics
Folding of the Protein

- Only the correct folded protein functions correctly (cf Creutzfeld-Jacob, Alzheimer, BSE, Parkinson)
- Proteins always fold into their specific 3D structure
- Complicated procedure with lots of interactions
- Folding pathways are not unique and have intermediate states
- Folding is assisted by special chaperones (hide the hydrophobic regions or act as containers)
- Folding time: milliseconds up to minutes or hours
- Major tasks in bioinformatics is the prediction of the 3D structure to guess the function or to design new proteins
Folding of the Protein
Folding of the Protein

1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding
1 Biological Basics
1.1 The Cell
1.2 Central Dogma
1.3 DNA
1.4 RNA
1.5 Transcription
1.5.1 Initiation
1.5.2 Elongation
1.5.3 Termination
1.6 Splicing
1.7 Amino Acids
1.8 Genetic Code
1.9 Translation
1.9.1 Initiation
1.9.2 Elongation
1.9.3 Termination
1.10 Folding