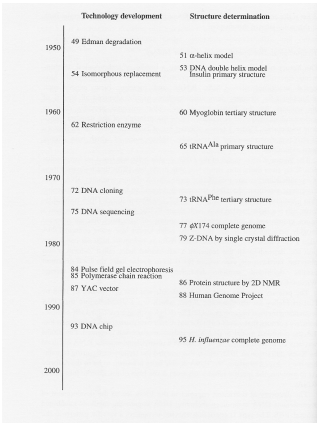
Computational Biology

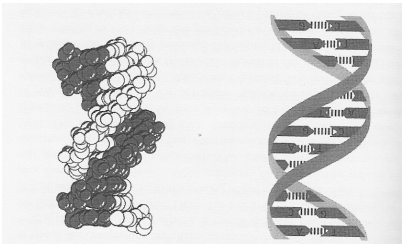
part II technological developments in molecular biology & databases

Jaap Kaandorp

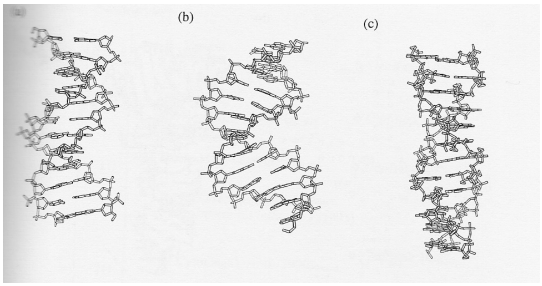
History of structure determination for nucleic acids and proteins I: DNA,RNA, proteins



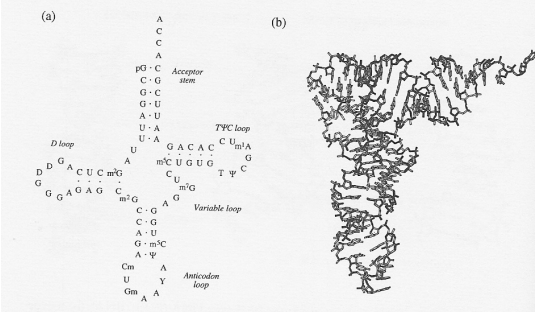
DNA double helix



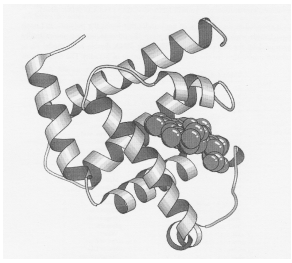
Polymorphic DNA tertiary structures



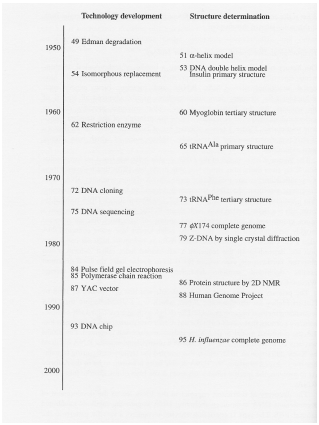
Transfer RNA a) primary and secondary structure b) tertiary structure



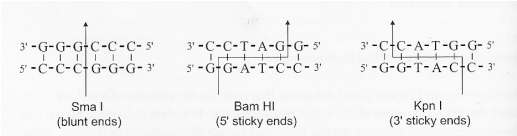
3D protein structure protein: example myoglobin



History of structure determination for nucleic acids and proteins II: restriction enzymes



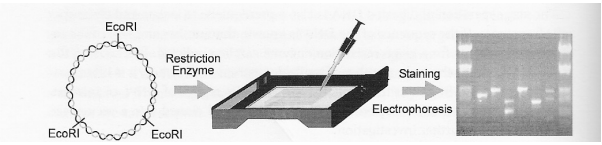
Restriction enzymes



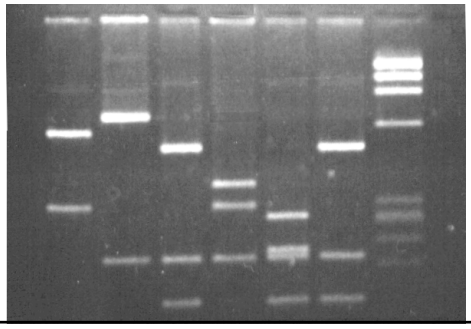
Restriction enzymes, different recognition sequences

Name	Organism	Recognition sequence
Alu I	*Arthrobacter luteus*	5′ A G ↑ C T 3′
Bam HI	*Bacillus amyloliquefaciens*	5′ G ↑ G A T C C 3′
Bgl II	*Bacillus globigii*	5′ A ↑ G A T C T 3′
Eco RI	*Escherichia coli*	5′ G ↑ A A T T C 3′
Eco RV	*Escherichia coli*	5′ G A T ↑ A T C 3′
Hind III	*Haemophilus influenzae*	5′ A ↑ A G C T T 3′
Kpn I	*Klebsiella pneumonia*	5′ G G T A C ↑ C 3′
Pme I	*Pseudomonas mendocina*	5′ G T T T ↑ A A A C
Sau96 I	*Staphylococcus aureus*	5′ G ↑ G N C C 3′
Sma I	*Serratia marcescens*	5′ C C C ↑ G G G 3′
Xma I	*Xanthomonas malvacearum*	5′ C ↑ C C G G G 3′

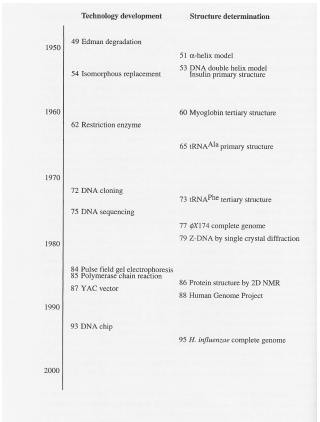
Gel electrophoresis of DNA restriction fragments



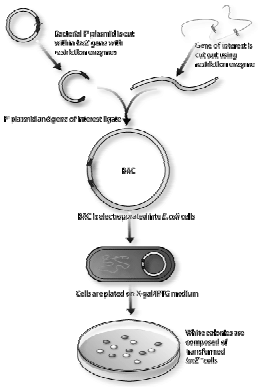
Gel electrophoresis of DNA restriction fragments



History of structure determination for nucleic acids and proteins III: DNA cloning



Cloning using bacteria (Shizuya et al, PNAS, 1992)



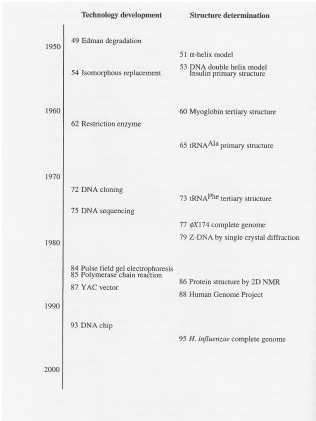
DNA libraries

- Genomic DNA libraries, large number of fragments is cloned in bacteria. Used in for example DNA chips
- Complementarty DNA (cDNA) libraries, use mRNA pool of cell or tissue of interest as starting point.. Used to study proteins which are expressed in small amounts

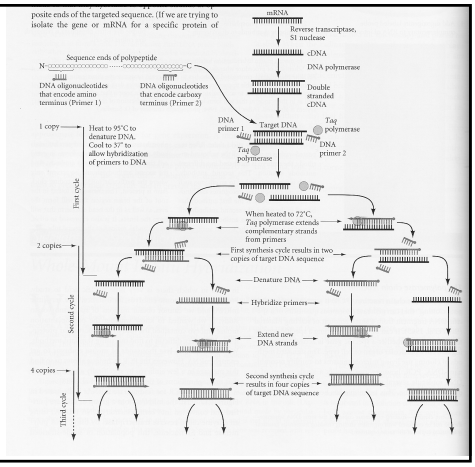
cDNA libraries differ from genomic libraries:

- Complementary DNA
- Only contain coding regions
- Tissue specific, snapshot of current gene expression
- Frequency of clones in library indicates the level of gene expression

History of structure determination for nucleic acids and proteins IV:PCR



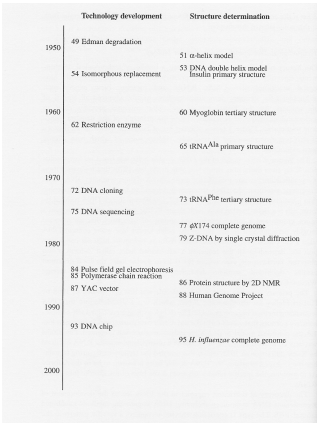
Polymerase Chain Reaction (Gilbert, 2001)



PCR

- Has revolutionized molecular genetics
- You can clone genes directly from genomic DNA (without DNA libraries)
- You can clone mRNA with this technique
- Very sensitive technique (method of choice in forensic studies)
- Real time PCR can provide quantitative information

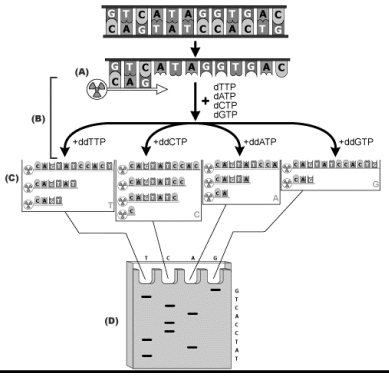
History of structure determination for nucleic acids and proteins V: DNA sequencing



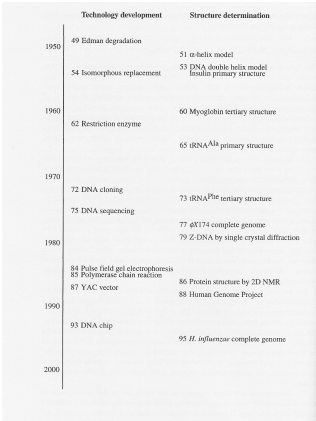
DNA-sequencing

- Use same procedure like in PCR to produce copies of target sequences
- Instead of synthesizing copies of the complete sequence, copies are forced to terminate before reaching the end
- Chain termination method by adding dideoxyncluotides (ddNTPs), chemical modified nucleotid nothing can be added after adding a ddNTP
- Resulting copies are separated by gel electrophoresis
- Four different reactions are visualized on 4 different lanes on the gel

DNA-sequencing



History of structure determination for nucleic acids and proteins V: DNA chips (microarrays)



DNA chips (microarrays) I

- Method for high-throughput gene expression analysis
- Monitor the expression of several thousand of genes in one experiment, not a single gene
- You need a DNA library

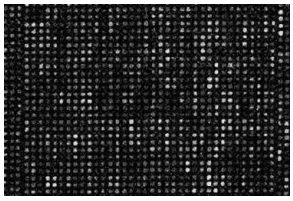
DNA chips (microarrays) II

- Construction of a chip from a DNA library
- Amplify individual genes by PCR
- individual genes are spotted onto glass slides

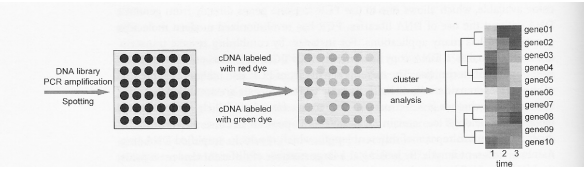
DNA chips (microarrays) III

- Extract mRNA from 2 samples you want to compare
- Use reverse transcriptase mRNA->cDNA
- Label cDNA with fluorescent colours
- cDNA is hybridized with DNA chip

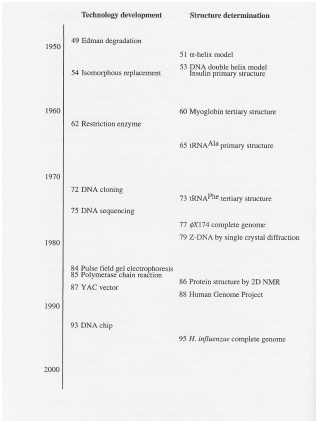
Microarrays III



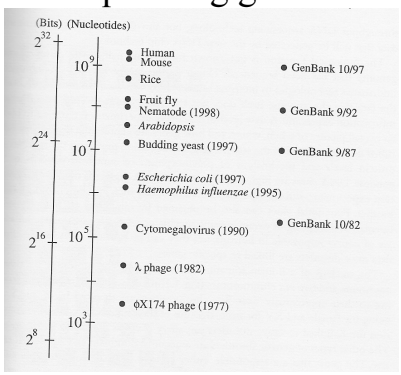
Microarrays IV



History of structure determination for nucleic acids and proteins V: sequenced genomes (human genome 2001)



Sequencing genomes



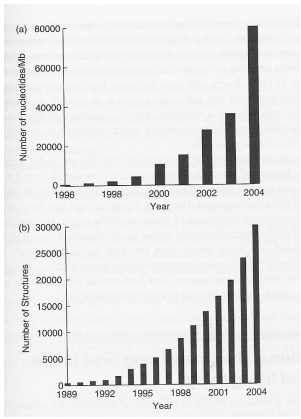
Genome sequences

- Viruses genome size 10^3-10^5nucleotids
- bacteria 10^6 nucleotids
- Saccharomyces cerevisae (yeast) 10^7 nucleotids
- Caenorhabdites elegans 10^8 nucleotids
- free-living organisms 10^6-10^9 nucleotids

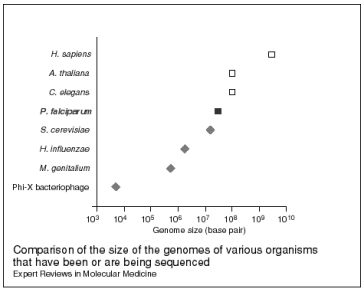
Genome sequences: increasing genome size -> increasing fraction of non-coding regions

- Bacteria and archeabacteria genes every 1000 nucleotids
- Saccharomyces cerevisae (yeast) genes every 2000 nucleotids
- Caenorhabdites elegans genes every 5000 nucleotids

Number of nucleotids / proteins in databases



Genomes different organisms I



Comparison of the size of the genomes of various organisms that have been or are being sequenced
Expert Reviews in Molecular Medicine

Arabidopsis thaliana

Genomes in different species II

TABLE I

Nuclear genome size in different species

Common name	Scientific name	Nuclear genome size (1)
Wheat	*Triticum aestivum*	15,966
Onion	*Allium cepa*	15,290
Garden pea	*Pisum sativum*	3,947
Corn	*Zea mays*	2,292
Asparagus	*Asparagus officinalis*	1,308
Tomato	*Lycopersicum esculentum*	907
Sugarbeet	*Beta vulgaris*	758
Apple	*Malus X domestica*	743
Common bean	*Phaseolus vulgaris*	637
Cantaloupe	*Cucumis melo*	454
Grape	*Vitis vinifera*	483
Man	*Homo sapiens*	2,910

1: Expressed in Megabases (1Mb:1,000,000 bases)

Genomes different organisms III

Organism	Number of chromosomes (haploid genome)	Genome size (base pairs; genes)
Mycoplasma genitalium (prokaryote)	1 circular chromosome	$580 \cdot 10^3$ bp; 480 genes
Escherichia coli (prokaryote)	1 circular chromosome	$4.6 \cdot 10^6$ bp; 4,290 genes
Saccharomyces cerevisiae (budding yeast; eukaryote)	16 chromosomes	$12.5 \cdot 10^6$ bp; 6,186 genes
Arabidopsis thaliana (flowering plant; eukaryote)	5 chromosomes	$100 \cdot 10^6$ bp; ~25,000 genes
Drosophila melanogaster (fruit fly, eukaryote)	4 chromosomes	$180 \cdot 10^6$ bp; ~14,000 genes
Mus musculus (mouse, eukaryote)	20 chromosomes	$2.5 \cdot 10^9$ bp; ~30,000 genes
Homo sapiens (human, eukaryote)	23 chromosomes	$2.9 \cdot 10^9$; ~30,000 genes

Rice genome

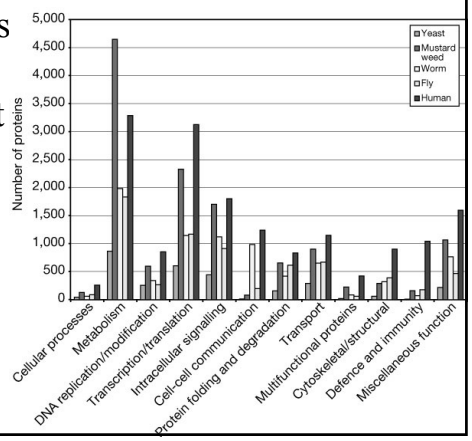
- Goff, S. A. et al. A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science 296, 92 - 100 (2002).

  That means rice probably has more genes than we do, despite having only one-seventh as much DNA.One reason for this difference is that plants' genes seem to be on average much shorter than mammals'. A rice gene is usually about 4,500 DNA letters long. The average human gene probably stretches to over 30,000 letters.

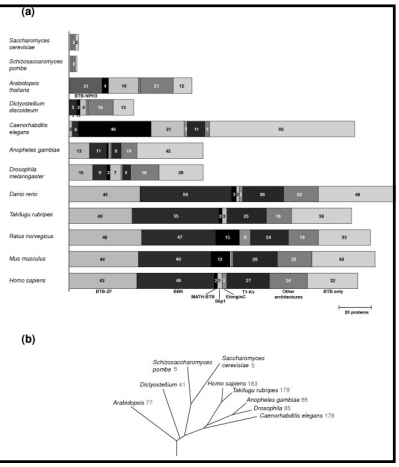
Rice genome II

- Plants are also prone to copying genes, chromosomes, and sometimes their entire genome. This might be a way for them to generate genetic diversity. Mammals can rearrange individual genes to make several different sorts of protein. Plants show much less of this, seeming to rely on a larger number of shorter genes.

Genomes in different species IV



Genomes in different species V



Sequenced genomes

- The genomes of more than 180 organisms have been sequenced since 1995. The Quick Guide includes descriptions of these organisms and has links to sequencing centers and scientific abstracts

  http://www.genomenewsnetwork.org/resources/sequenced_genomes/genome_guide_p1.shtml

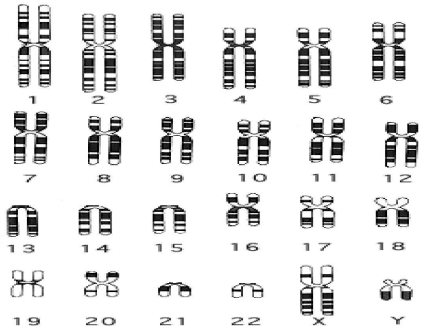
Sequence genomes II

- Aeropyrum pernix
  Agrobacterium tumefaciens
  Anabaena
  Anopheles gambiae
  Apis mellifera
  Aquifex aeolicus
  Arabidopsis thaliana
  Archaeoglobus fulgidus
  Ashbya gossypii
  Bacillus anthracis
  Bacillus cereus
  Bacillus halodurans
  Bacillus licheniformis
  Bacillus subtilis
  Bacteroides fragilis
  Bacteroides thetaiotaomicron
  Bartonella henselae
  Bartonella quintana
  Bdellovibrio bacteriovorus
  Bifidobacterium longum
  Blochmannia floridanus
  Bordetella bronchiseptica
  Bordetella parapertussis
  Bordetella pertussis
  Borrelia burgdorferi
  Bradyrhizobium japonicum
  Brucella melitensis
  Brucella suis
  Buchnera aphidicola

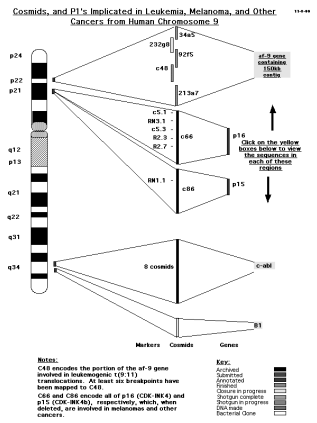
Human Genome project

- Since initiation Human Genome project (1988, finalized in 2001) computational biology (bioinformatics) has become an integral part of molecular biology
- Make biological sense out of sequence data, data produced by DNA chips etc.

Human chromosomes



Human chromosomes



Genetic disorders

- X-linked recessive disorders
- Autosomal dominant disorders
- Autosomal recessive disorders

Computational Biology

end part II technological developments in molecular biology & databases

Jaap Kaandorp