Bioinformatik

Mathias Weyland

27. November 2008

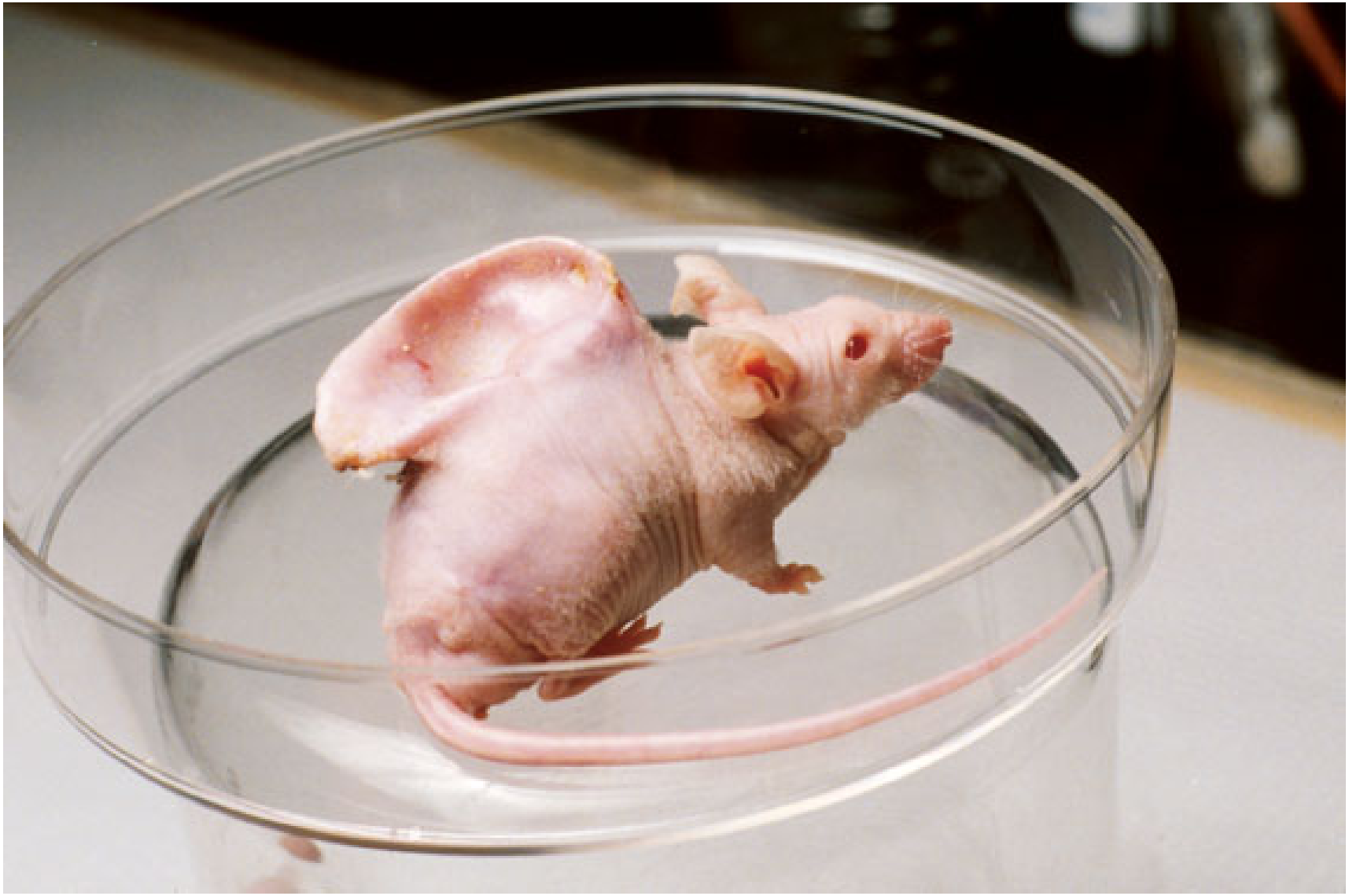
Bioinformatik

Einleitung

Agenda

- Einleitung
- DNA
- Proteine
- Simulation
- Fragen

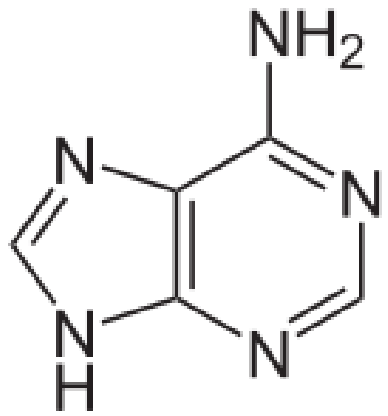
Biologie verstehen



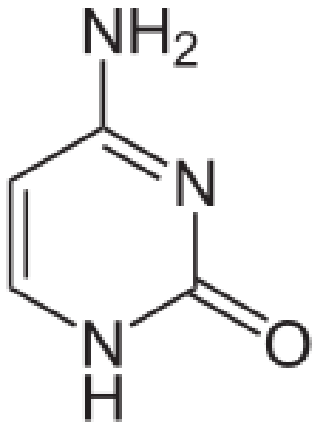
Bioinformatik

DNA

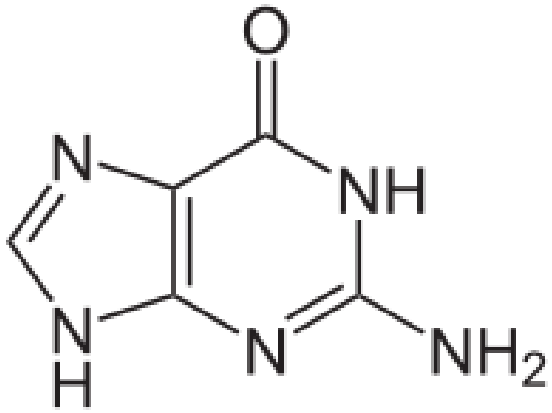
Nukleobasen



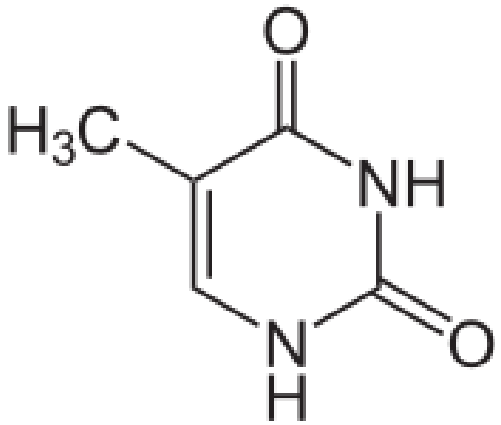
A



C

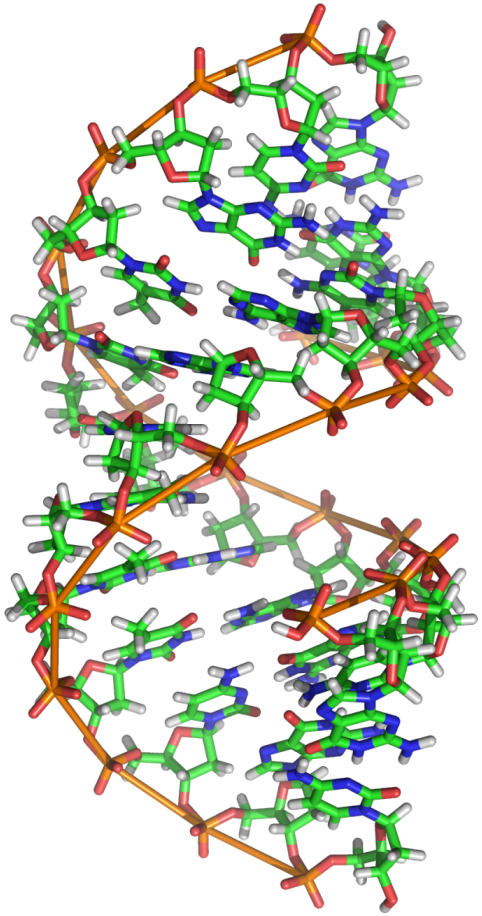


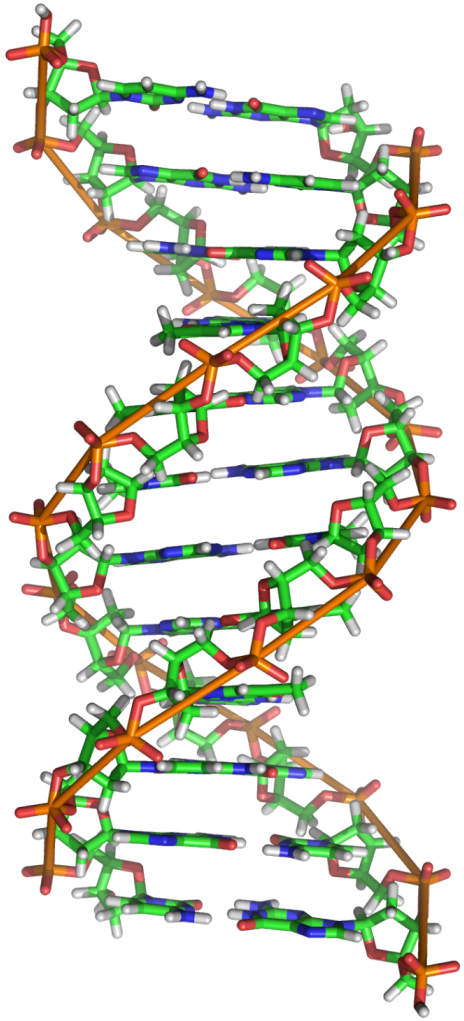
G

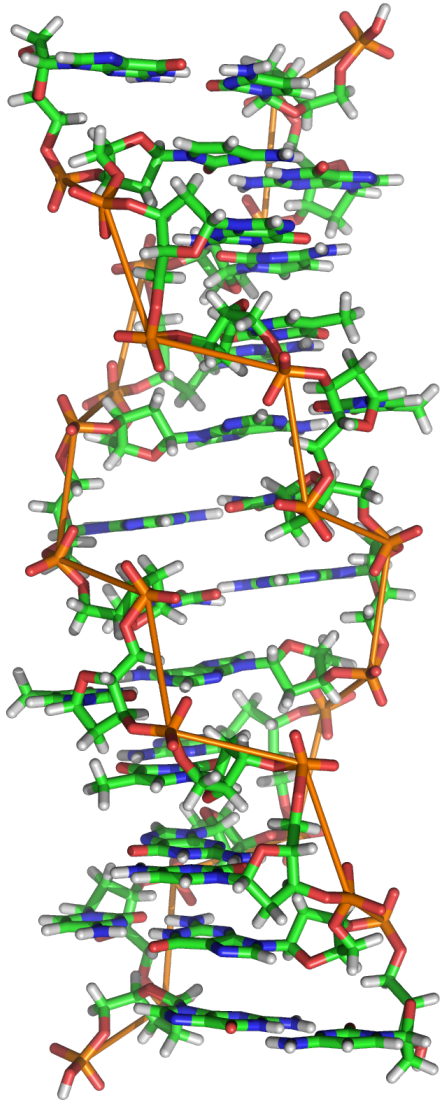


T

DNA (De(s)oxyribonukleinsäure)







TATA-Box

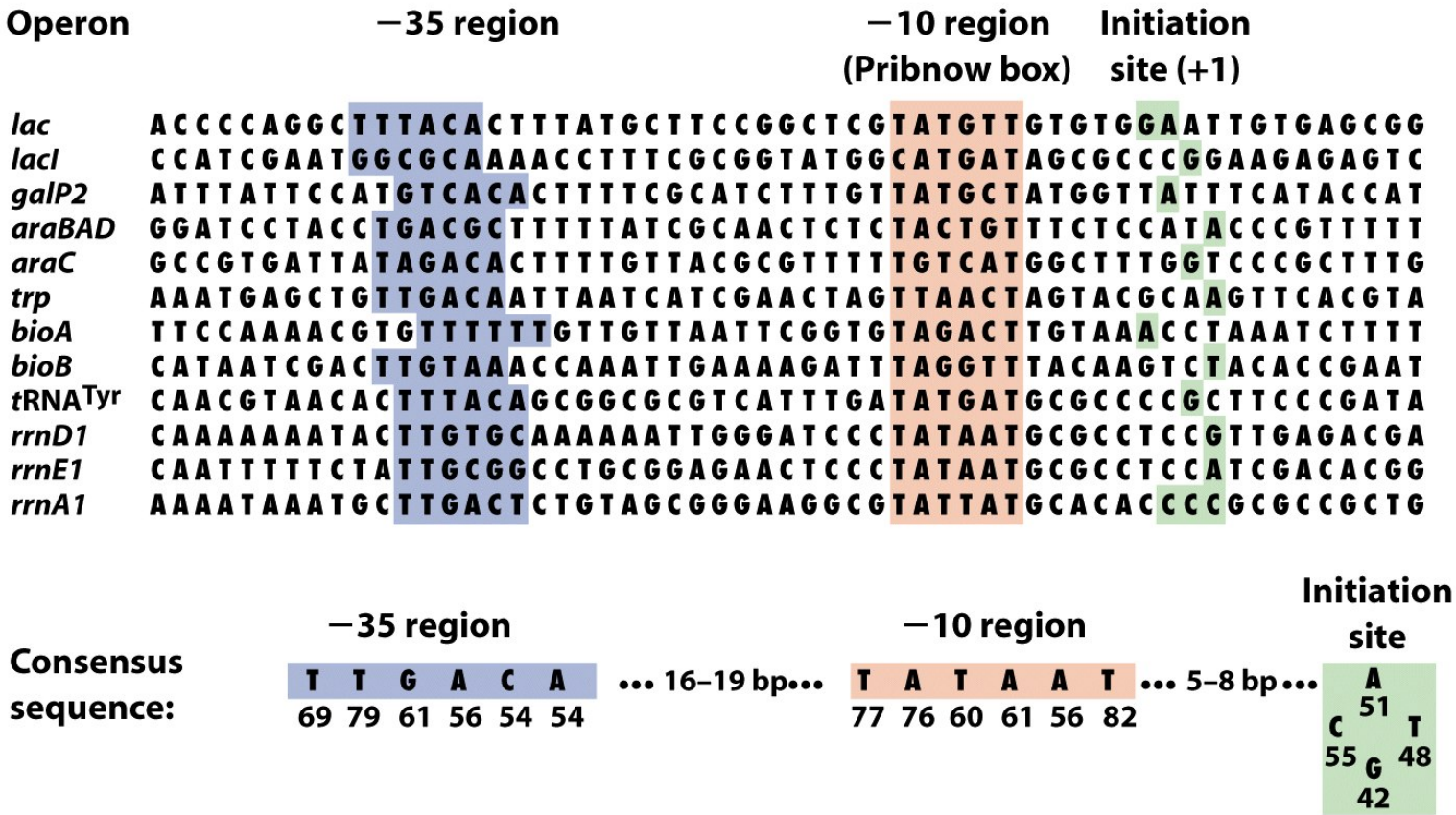


Figure 25-5 Fundamentals of Biochemistry, 2/e

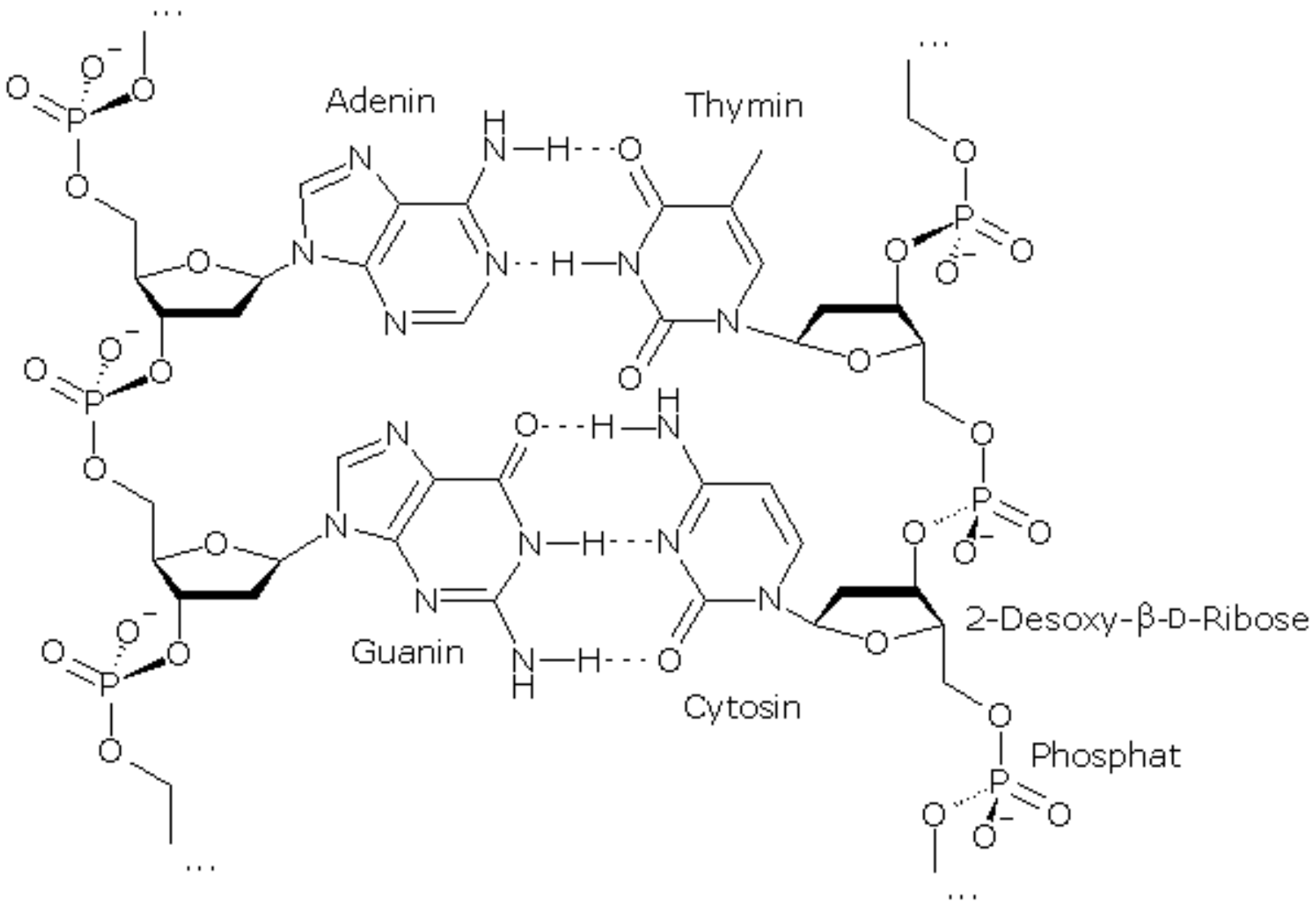
Anzahl Gene

Organismus	Anzahl Gene
E. coli	4'400
S. cerevisiae	5'800
Drosophila melanogaster	13'700
Arabidopsis thaliana	25'500
Homo Sapiens	27'000
Mus musculus	29'000
Oryza satiav	50'000

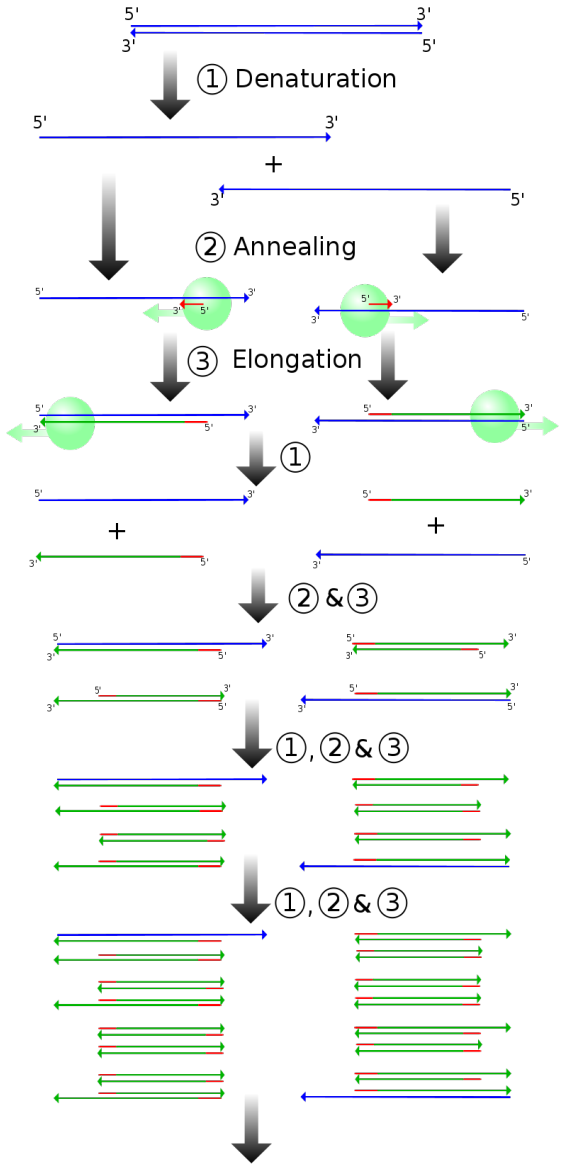
Elektronegativität

Element	Elektronegativität
Wasserstoff (H)	2.1
Kohlenstoff (C)	2.5
Stickstoff (N)	3.0
Sauerstoff (O)	3.5

Wasserstoffbrücken

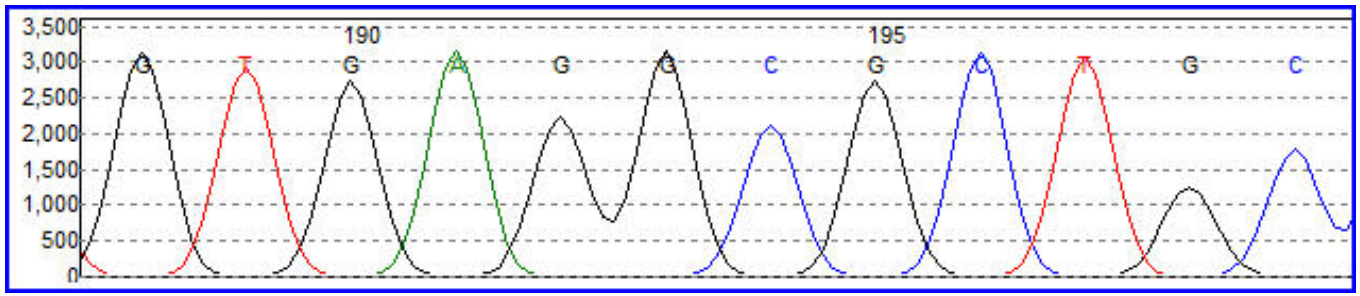


PCR (Methode)

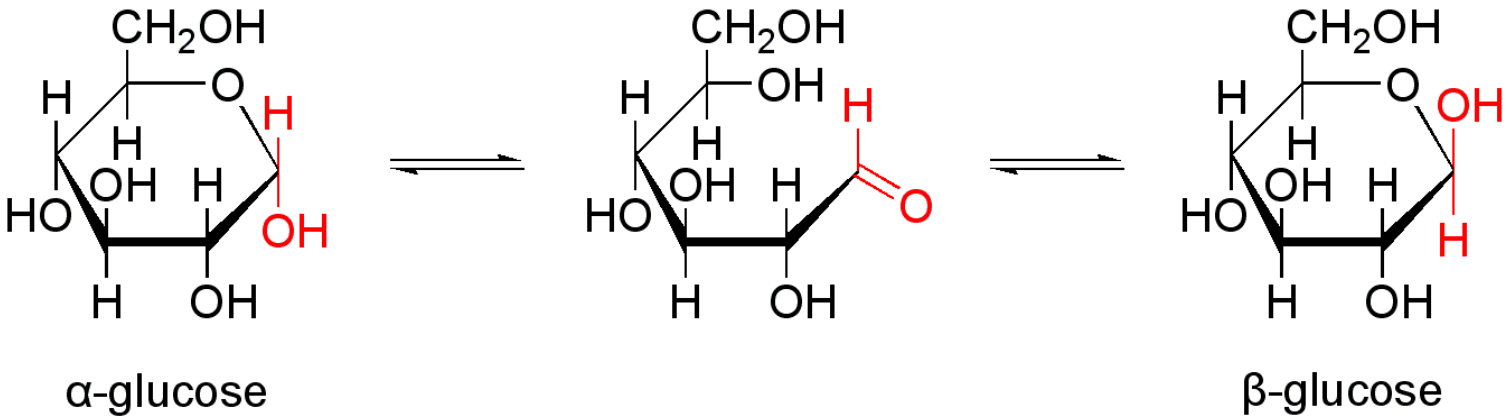


Exponential growth of short product

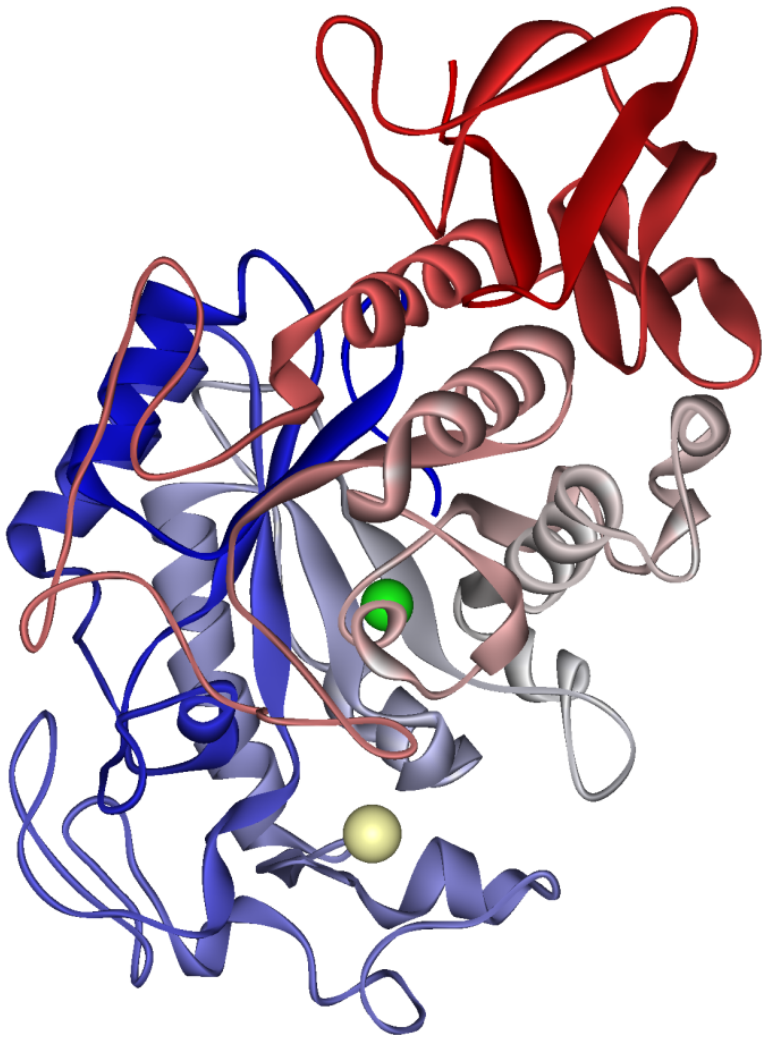
Moderner Output



α- und β-Glukose



3D-Struktur der Hydrolase



α-Glukosidase Protein

MASSHITLVS HDDGFTLSYQ QRTILHHSRQ TPCLWAGAGT ADIEMFRGNF
SIKDRLDEKI ALTQAAVSER ASGWVIRFTR GETIGATLEV STDDHDRLVL
RLKNDAVAHN RLWLRLAAKP ADHIYGCGEQ FSYFDLRGKP FPLWTSEQGV
GRNKQSHVTW LADCKENAGG DYYWTFFPQP TFVSTQKYYC HVDNSGYMNF
DFSAPHYHEL AFWENHATLR FECAPTYIEL LEKLTALLGR QPELPDWVYD
GVTLGIQGGT EVCQQKLDVM RRGGVKVNGI WAQDWSGIRM TSFGKRVMWN
WKWNRDLYPQ LDTRIAQWKR EGVQFLAYIN PYVASDRDLC EEAASRGYLT
QDAAGKDYHV EFGEFYAGVV DLTNPEAYAW YKEVIKKQLI ELGCGGWMAD
FGEYLPTDTH LYNEVSAEIM HNAWPALWAK CNYEALEETG KLGEVLFFMR
AGYTGSQKYS TMMWAGDQNV DWSLDDGLAS VIPAALSLAM SGHGLHHSDI
GGYTTLFEMK RSKELLLRWC DFSAFTPMMR THEGNRPGDN WQFDGDAQTI
AHFARMTTIF STLKPYIKQA VAQNARTGLP VMRPLFLHYE DEPRAYTLKY
QYLFGRDLLV APVYEEGRQD WSLWLPDDRW VNVWTGETHG GGDITVDAPL
GKPPVFYREG

α-Glukosidase Gensequenz (I)

ATGGCTTCGT CACACATTAC ACTCGTTTCC CATGACGACG GATTTACGTT
GTCGTACCAG CAAAGAACGA TCCTGCATCA CTCAAGGCAA ACACCCTGCC
TGTGGGCGGG CGCAGGCACA GCGGATATCG AGATGTTCCG CGGCAACTTC
AGCATTAAAG ACCGGCTCGA TGAGAAAATC GCCCTGACCC AGGCGGCCGT
TAGTGAGCGC GCTTCTGGCT GGGTGATTCG CTTTACGCGC GGCGAGACGA
TCGGTGCCAC GCTTGAGGTG AGCACCGATG ACCATGATCG TCTGGTGCTG
AGGCTCAAAA ACGACGCCGT CGCGCACAAT CGTCTCTGGT TGCGGCTGGC
GGCGAAGCCA GCCGATCATA TTTACGGCTG CGGCGAGCAG TTCTCTTATT
TCGATCTGCG CGGTAAACCG TTCCCGCTGT GGACCAGCGA ACAGGGCGTG
GGCCGCAATA AACAGAGCCA CGTTACCTGG CTTGCCGACT GTAAAGAGAA
CGCGGGCGGC GACTACTACT GGACCTTCTT TCCGCAGCCA ACGTTCGTCA
GCACCCAAAA GTACTACTGC CACGTCGATA ACAGCGGCTA TATGAACTTC
GATTTCAGCG CGCCGCACTA TCATGAGCTG GCGTTCTGGG AAAACCATGC
CACGCTGCGC TTCGAATGCG CGCCGACCTA CATCGAACTG CTGGAAAAAC
TCACCGCCCT GCTGGGCCGC CAGCCGGAGC TGCCGGACTG GGTTTACGAT
GGCGTAACGC TCGGCATTCA GGGCGGTACC GAAGTCTGCC AGCAGAAACT

α-Glukosidase Gensequenz (II)

CGACGTGATG CGCCGCGGCG GCGTGAAGGT AAACGGCATC TGGGCGCAGG
ACTGGTCCGG CATCCGCATG ACCTCCTTCG GCAAGCGCGT GATGTGGAAC
TGGAAATGGA ACCGCGATCT TTACCCGCAG CTTGATACGC GCATTGCACA
GTGGAAACGC GAAGGCGTGC AGTTCCTCGC TTATATCAAC CCGTACGTGG
CAAGCGATCG CGATTTGTGC GAAGAAGCGG CGTCGCGTGG TTATCTGACA
CAAGATGCCG CAGGCAAGGA TTATCACGTC GAGTTTGGCG AGTTCTACGC
AGGCGTCGTC GACCTGACCA ACCCGGAAGC GTACGCCTGG TATAAAGAGG
TCATCAAAAA ACAGCTGATT GAACTGGGTT GCGGCGGCTG GATGGCTGAT
TTTGGCGAAT ATCTGCCGAC CGACACGCAC CTGTATAACG AGGTCAGCGC
GGAGATTATG CACAACGCCT GGCCTGCGCT GTGGGCGAAG TGCAACTACG
AGGCGCTTGA AGAGACCGGA AAACTCGGCG AGGTGCTGTT CTTTATGCGC
GCCGGTTATA CCGGCAGCCA GAAATATTCC ACGATGATGT GGGCAGGCGA
TCAGAACGTC GACTGGAGTC TCGACGACGG GCTCGCCTCG GTGATCCCCG
CCGCGCTGTC TTTGGCGATG TCCGGGCACG GCCTGCACCA CAGCGATATC
GGCGGATACA CCACGCTGTT TGAAATGAAA CGCAGCAAAG AGCTACTGCT
ACGCTGGTGC GATTTCAGCG CGTTTACGCC GATGATGCGC ACGCATGAGG

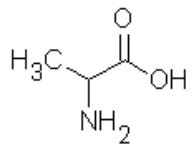
α-Glukosidase Gensequenz (III)

GCAACCGCCC TGGCGATAAC TGGCAGTTCG ACGGCGACGC GCAAACCATC
GCGCATTTTG CCCGCATGAC GACAATTTTC AGCACGCTCA AACCCTATAT
CAAACAGGCG GTGGCGCAGA ATGCCCGCAC CGGCCTGCCG GTGATGCGTC
CGCTGTTTTT GCACTATGAG GATGAACCCC GCGCCTACAC GCTGAAATAT
CAGTATCTGT TTGGCCGCGA TCTGCTGGTG GCACCGGTGT ATGAAGAAGG
ACGCCAGGAC TGGTCGCTGT GGTTGCCTGA CGATCGCTGG GTAAACGTCT
GGACCGGTGA AACCCACGGC GGCGGCGACA TTACGGTGGA TGCGCCGCTC
GGCAAACCGC CGGTGTTCTA CCGCGAGGGC AGCGAATGGC AGTCCCTTTT
CGCCACGCTG CGTCAGATAA ACCCCATGTA A

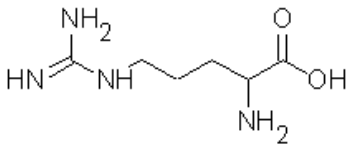
Bioinformatik

Proteine

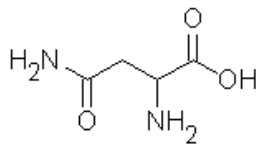
Aminosäuren (I)



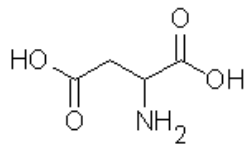
Alanin **(Ala)**



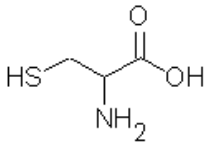
Arginin **(Arg)**



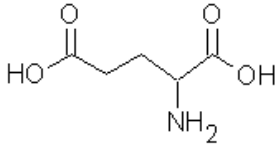
Asparagin **(Asn)**



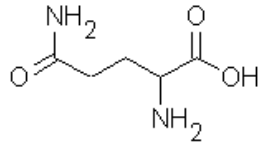
Asparaginsäure **(Asp)**



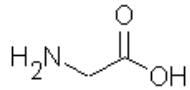
Cystein **(Cys)**



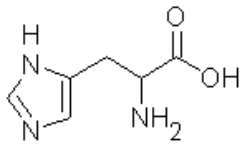
Glutaminsäure **(Glu)**



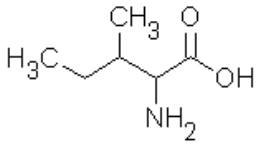
Glutamin **(Gln)**



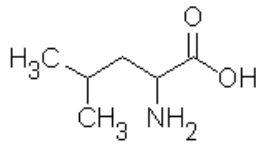
Glycin **(Gly)**



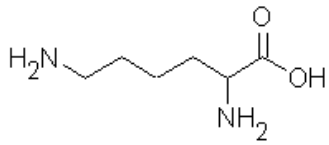
Histidin **(His)**



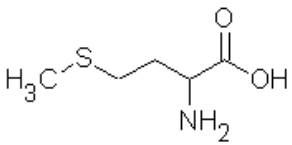
Isoleucin **(Ile)**



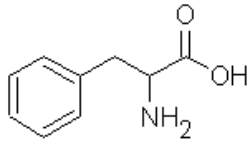
Leucin **(Leu)**



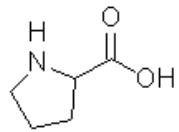
Lysin **(Lys)**



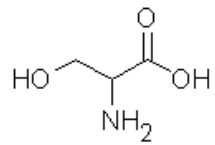
Methionin **(Met)**



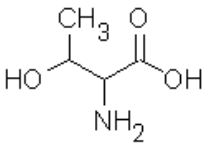
Phenylalanin **(Phe)**



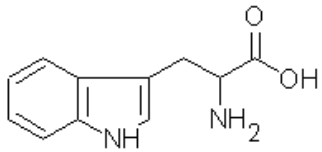
Prolin **(Pro)**



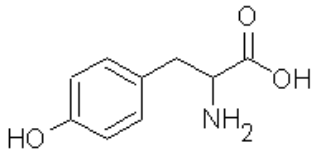
Serin **(Ser)**



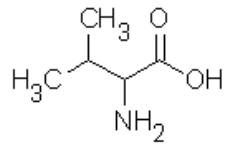
Threonin **(Thr)**



Tryptophan **(Trp)**

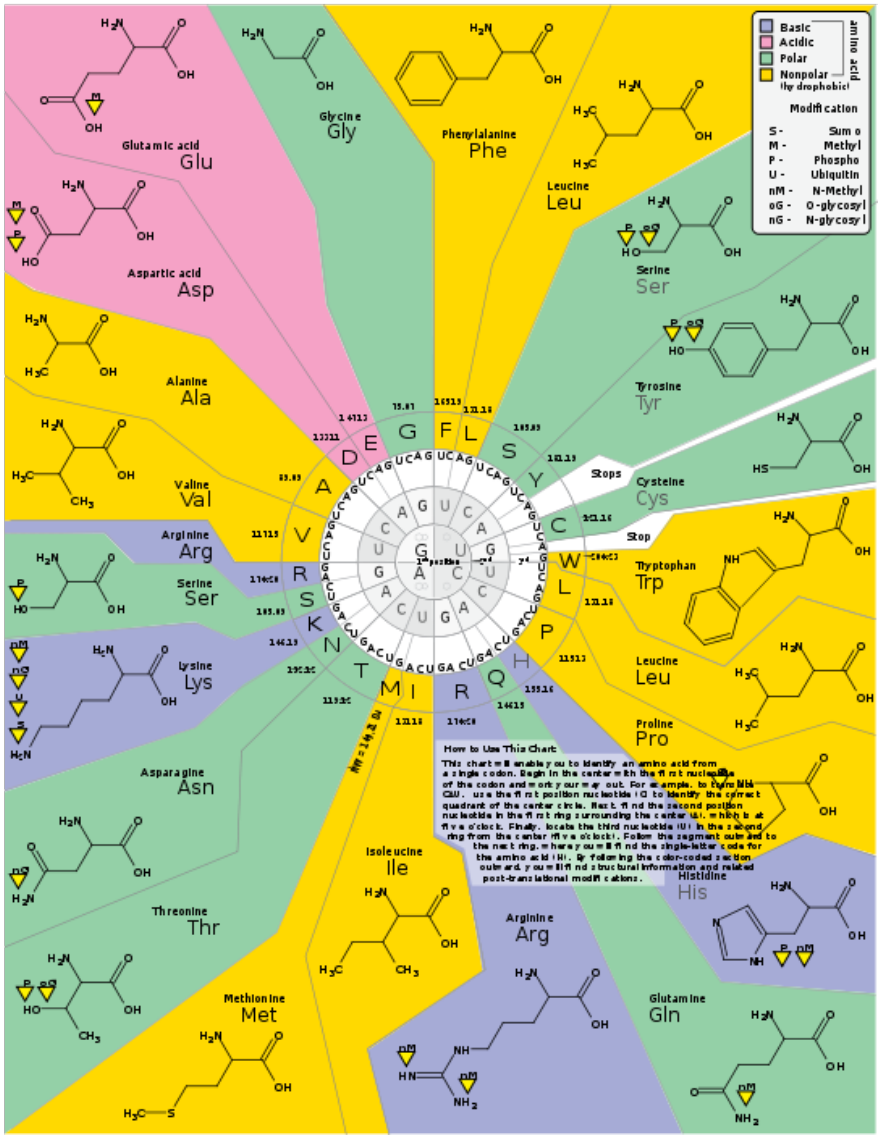


Tyrosin **(Tyr)**



Valin **(Val)**

Aminosäuren (II)



RNA-Transkription

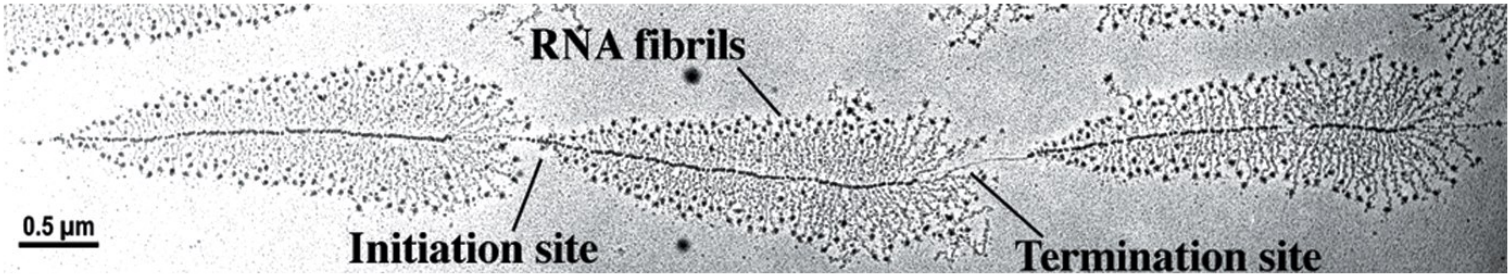
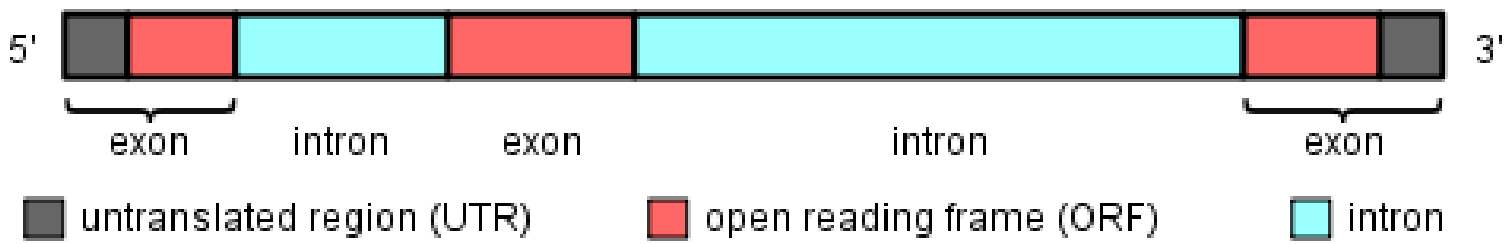
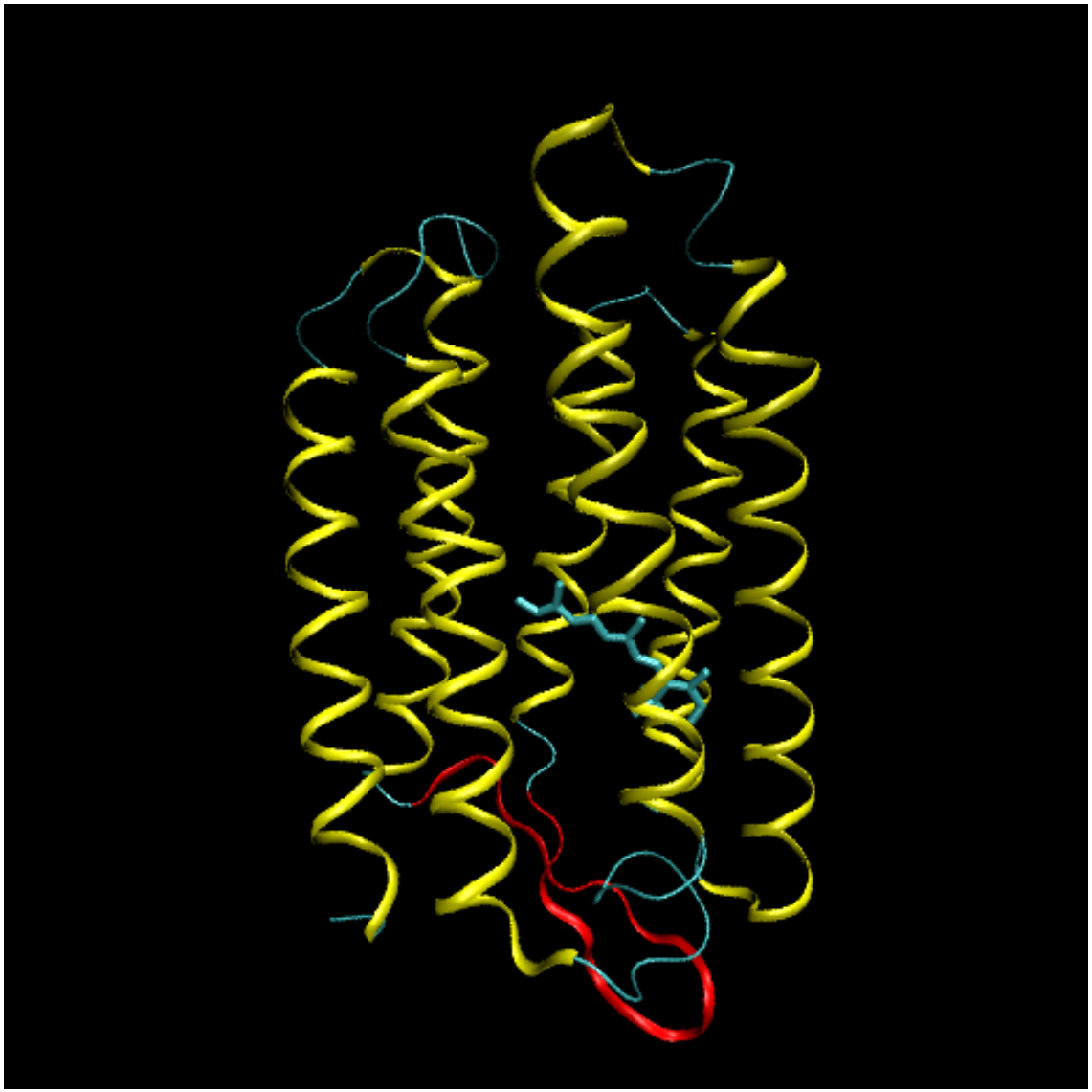


Figure 25-8 Fundamentals of Biochemistry, 2/e

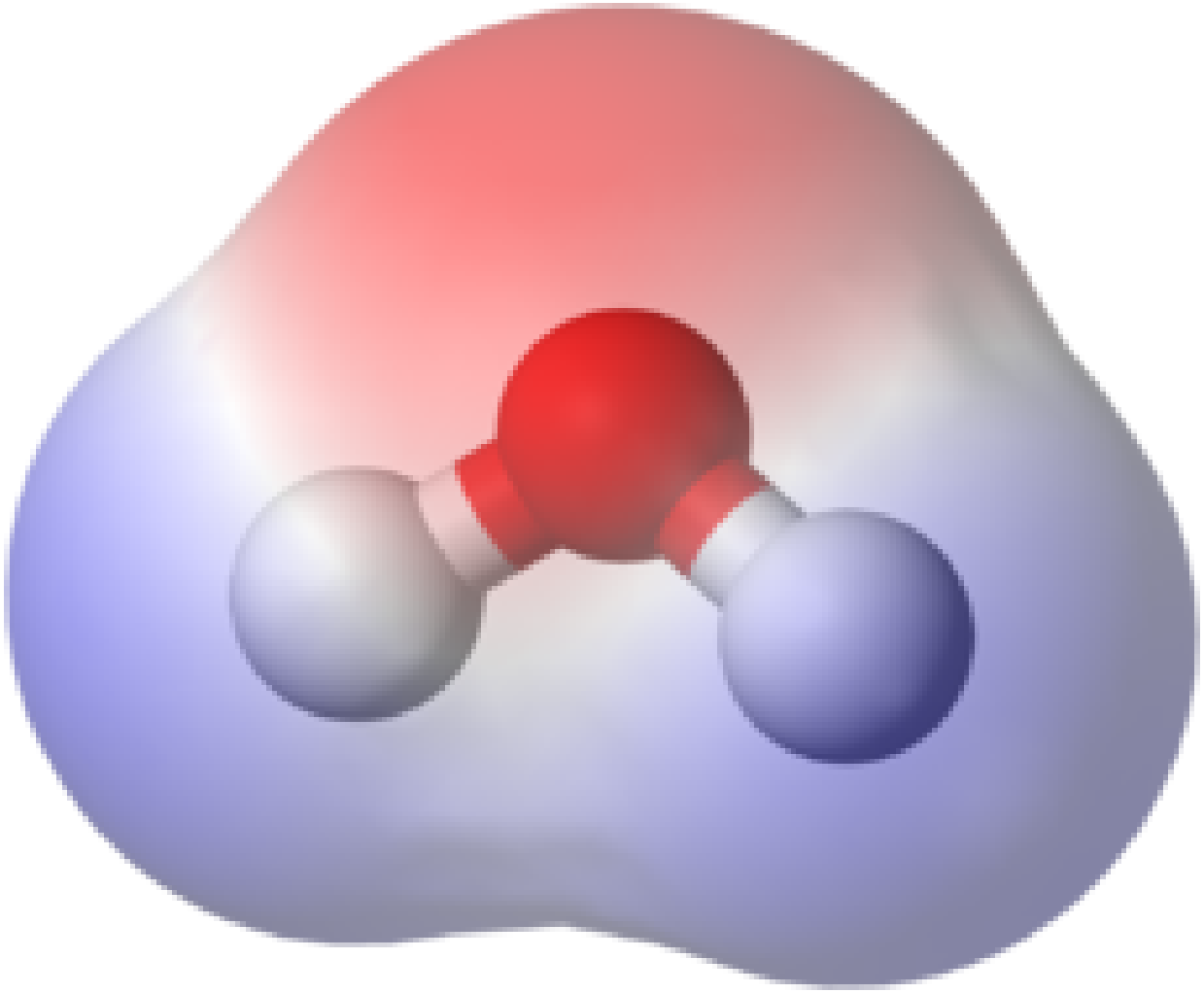
Introns und Exons



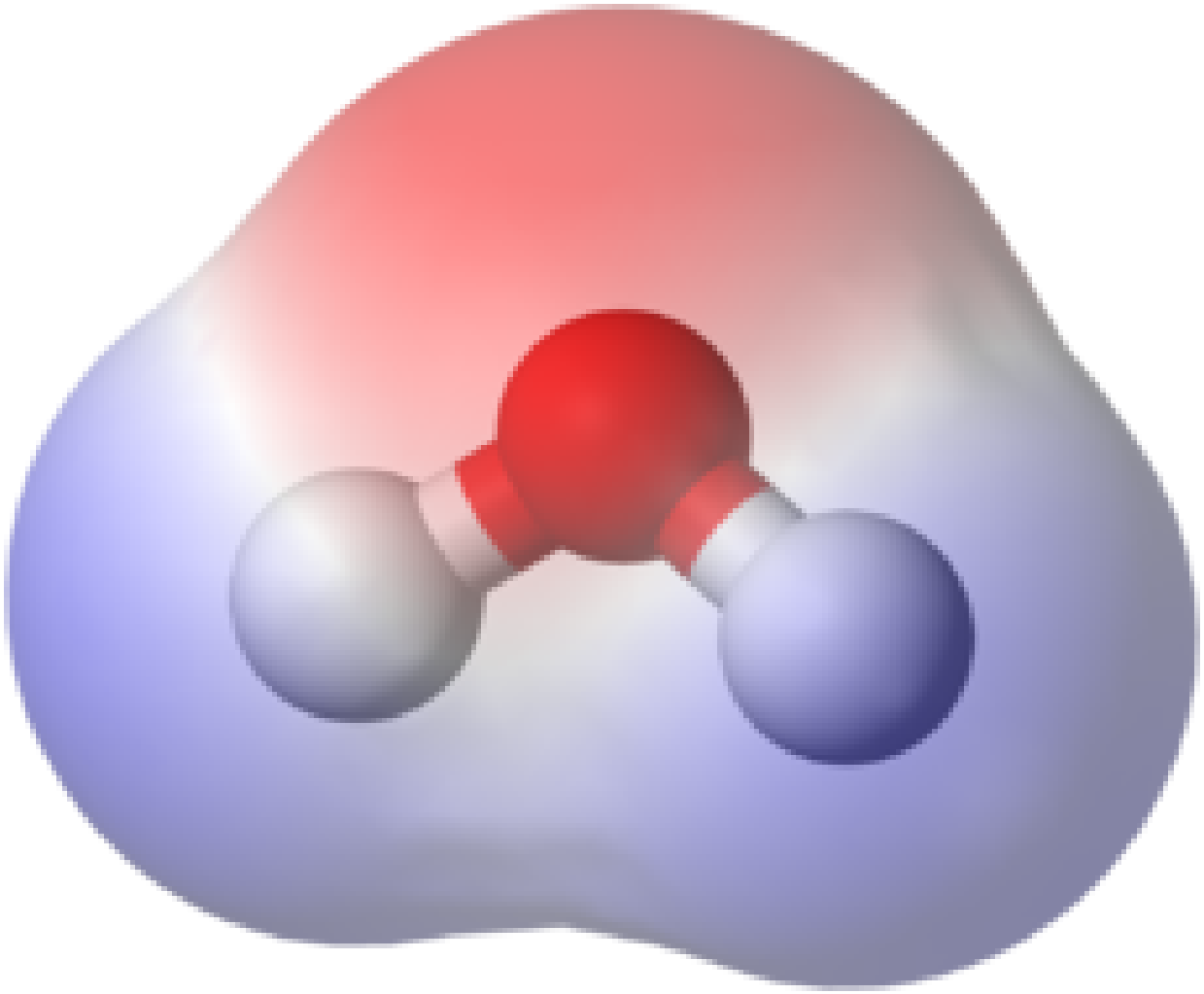
α-Helix



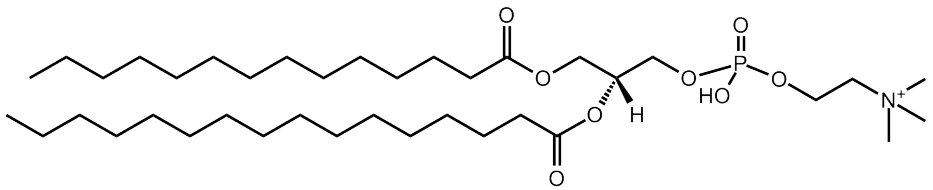
β-Sheet



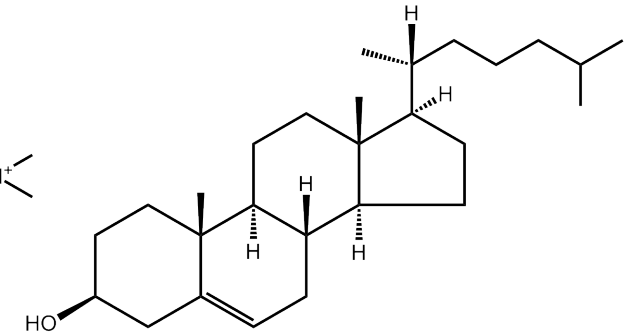
Wasser



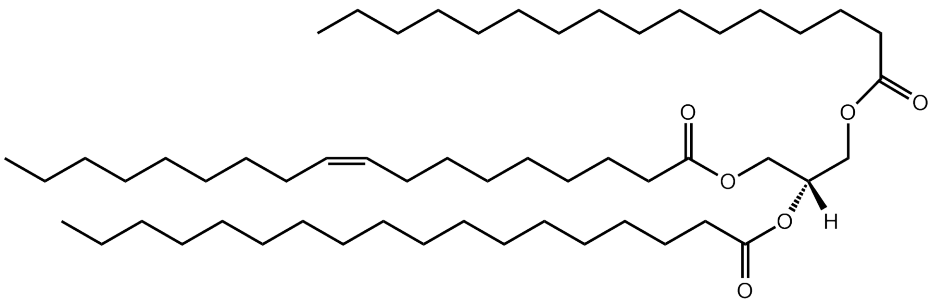
Lipide



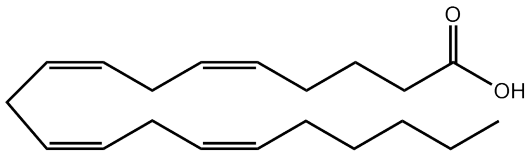
1-myristoyl-2-palmitoyl-*sn*-glycerophosphocholine
(Glycerophospholipids)



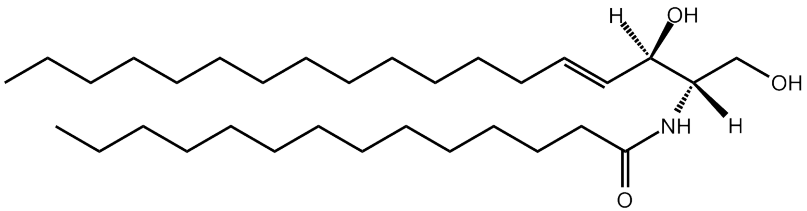
Cholesterol (Sterol lipids)



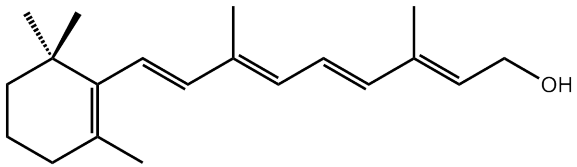
1-oleoyl-2-stearoyl-3-palmitoyl-*sn*-glycerol
(Glycerolipids)



Arachidonic acid (Fatty acyls)

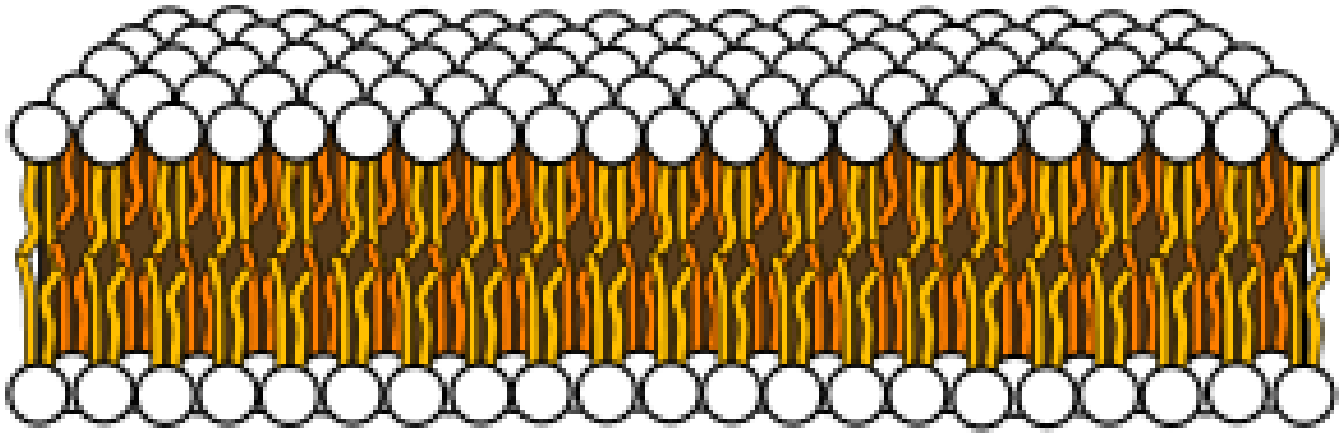


N-myristoyl-sphing-4-enine
(Sphingolipids)

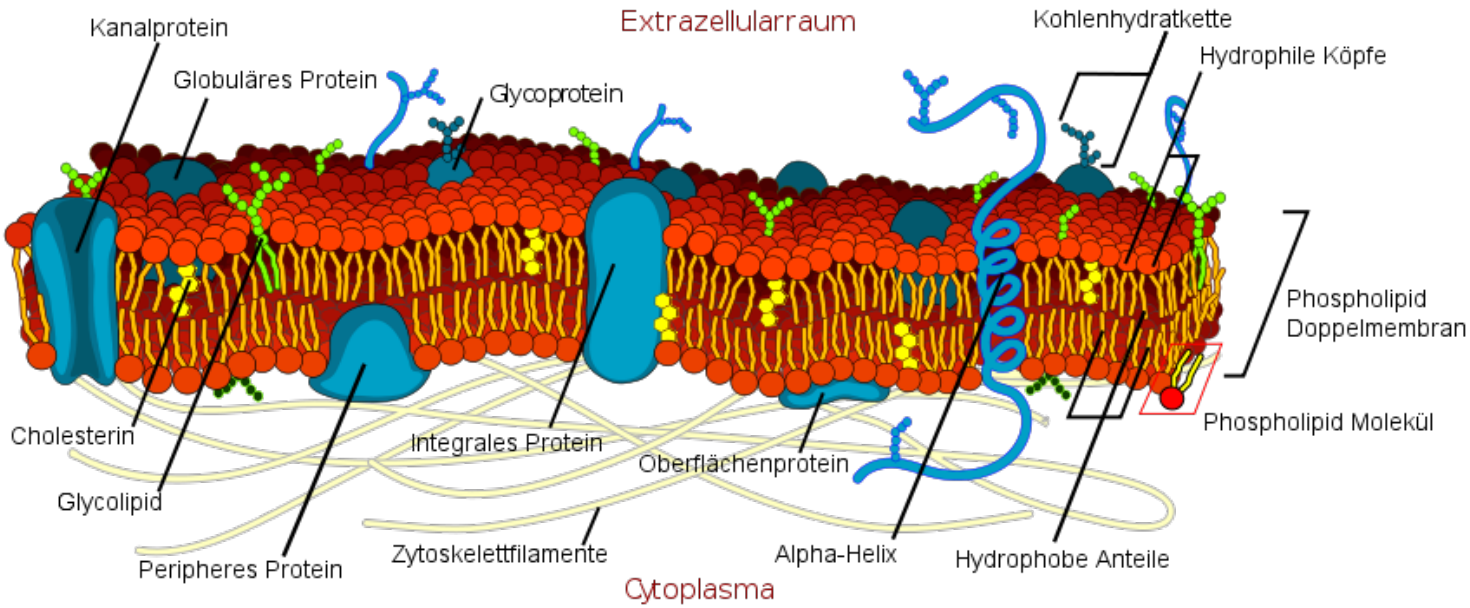


Retinol (Prenol lipids)

Lipid-Doppelmembran (Bilayer)



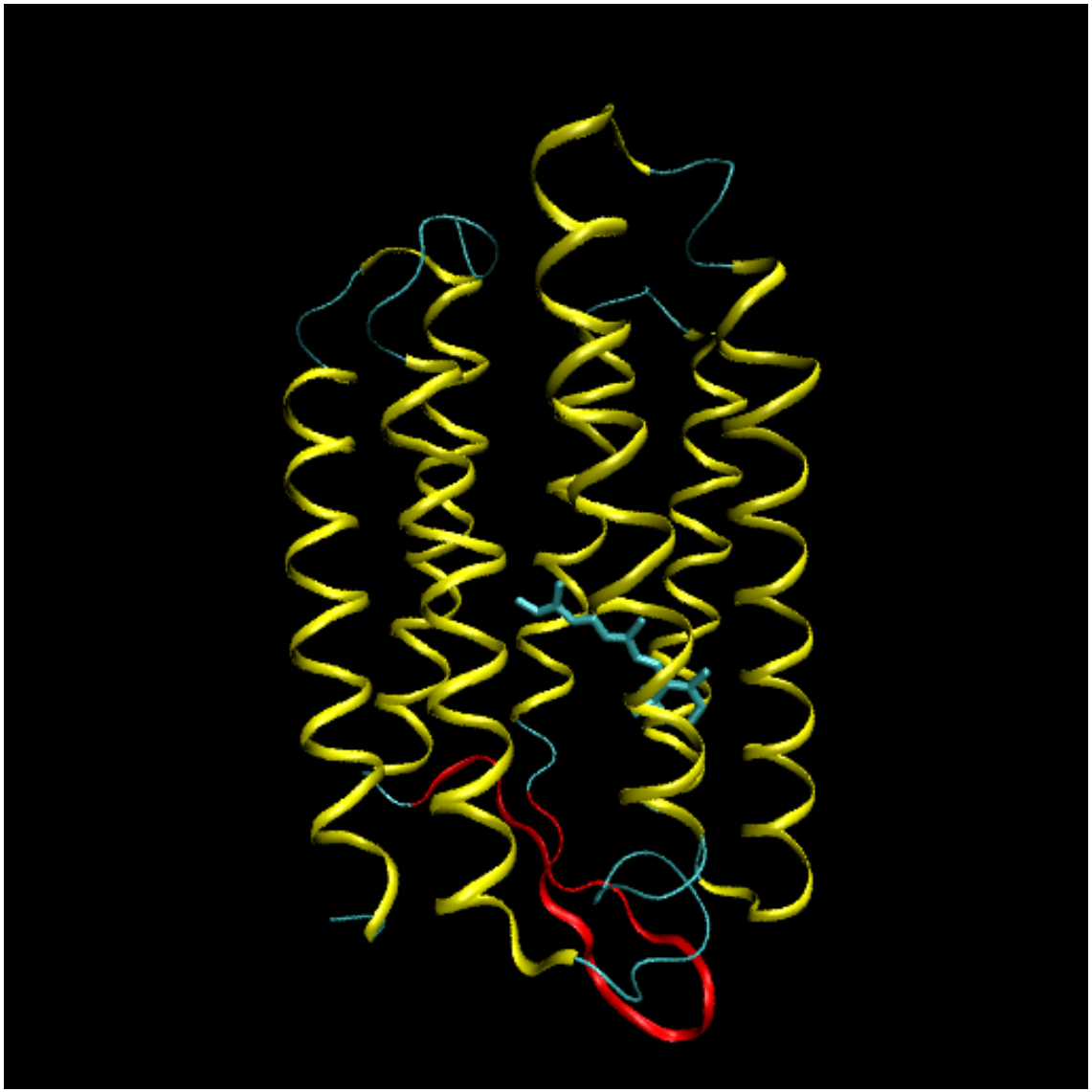
Lipid-Doppelmembran (Bilayer) mit Membranproteinen



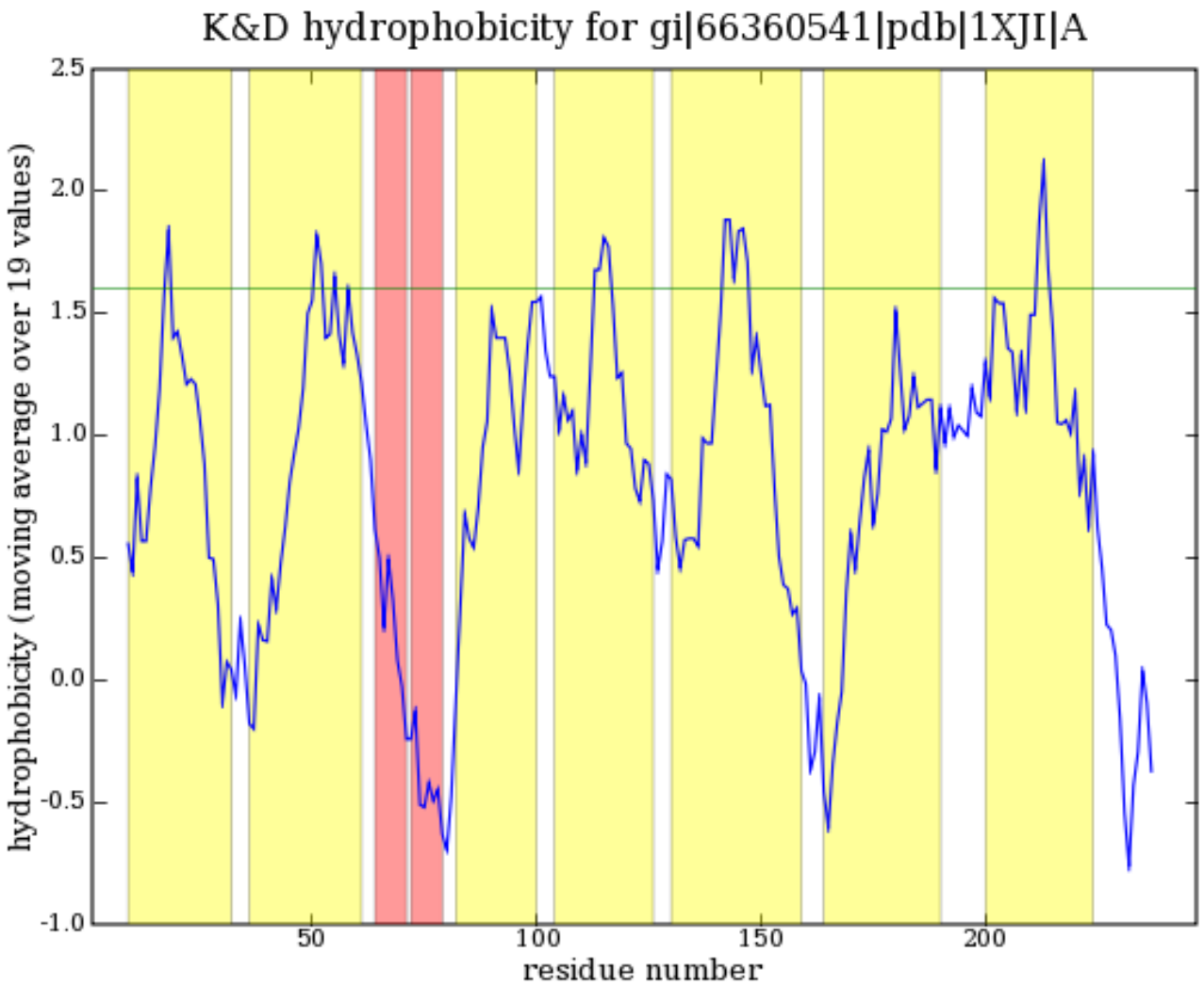
Membranprotein gi:66360541 (Protein)

```
AQITGRPEWI WLALGTALMG LGTLYFLVKG MGVSDPDAKK FYAITTLVPA
IAFTMYLSML LGYGLTMVPF GGEQNPIYWA RYADWLFTTP LLLLDLALLV
DADQGTILAL VGADGIMIGT GLVGALTKVY SYRFVWWAIS TAAMLYILYV
LFFGFTSKAE SMRPEVASTF KVLRNVTVVL WSAYPVVWLI GSEGAGIVPL
NIETLLFMVL DVSAKVGFGL ILLRSRAIFG EAEAPEPSAG DGAAATS
```

Membranprotein gi:66360541 (Protein)



Hydrophobizitätsplot



Bioinformatik

Simulation

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

Simulation von Drift

- 500 Flaschen mit je 32 Fruchtfliegen
- Pro Flasche 16 Weibchen und 16 Männchen
- Pro Geschlecht 8 Wt und 8 bn-Mutanten
- Paarung
- Zufällige Auswahl von 16 Fliegen pro Geschlecht und Flasche
- 40 Zyklen

drift.py (I)

```
bottles=500
individuals=32
wild=16
generations=40

flies=[[[() for x in range(individuals/2)]
    for y in range(2)] for z in range(bottles)]
matrix=[[0 for x in range(individuals+1)]
    for y in range(generations)]
```

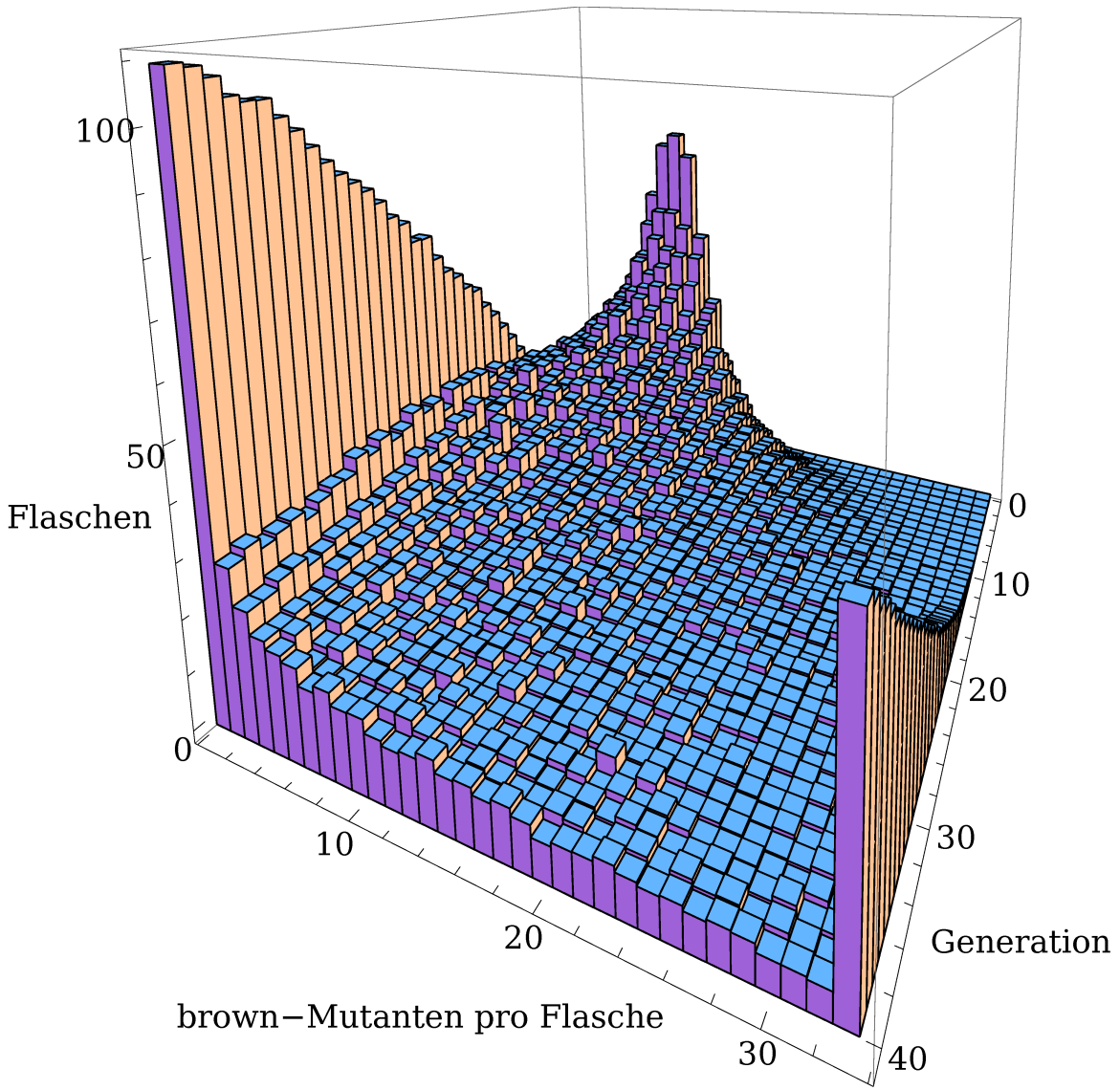
drift.py (II)

```
for bottle in range(bottles):
  for gender in range(2):
    for fly in range(individuals/2):
      if fly<wild/2:
        flies[bottle][gender][fly]='B','B'
      else:
        flies[bottle][gender][fly]='b','b'
```

drift.py (III)

```
for generation in range(generations):
  for bottle in range(bottles):
    tmp=[[() for x in range(individuals/2)] for y in range(2)]
    for gender in range(2):
      for fly in range(individuals/2):
        tmp[gender][fly]= (
          flies[bottle][0]
             [random.randrange(individuals/2)][random.randint(0,1)],
          flies[bottle][1]
             [random.randrange(individuals/2)][random.randint(0,1)])
    flies[bottle]=tmp
    matrix[generation][
        flies[bottle][0].count(('b','b'))+flies[bottle][1].count(('b','b'))] +=1
  print matrix[generation]
```

Genetische Drift



Bioinformatik

Fragen