

BIO-INFORMATICA

COURSE 3B VERGELIJKENDE GENOOMANALYSE

WEEK 4

A: EIWIT KARAKTERISERING

B: GENOOM EVOLUTIE EN HOMOLOGIE

OVERZICHT ONDERZOEKSPPLAN 3B

Week	Onderwerp	Activiteiten	M&M
1	Oriëntatie op onderwerp Sequenties verzamelen	Inlezen Stap 1	NCBI database
2	Genen en GC percentage	Stap 2	Python script
3	Codon gebruik	Stap 3	Python script
4	Eiwit karakterisering + genoom evolutie en homologie	Stap 4	Python script
5	Kenmerken oppervlakte proteïnen	Stap 5	Python script
6	Fylogenetisch onderzoek	Stap 6	Bioinf tools
7	Onderzoeksverslag		

DEZE WEEK

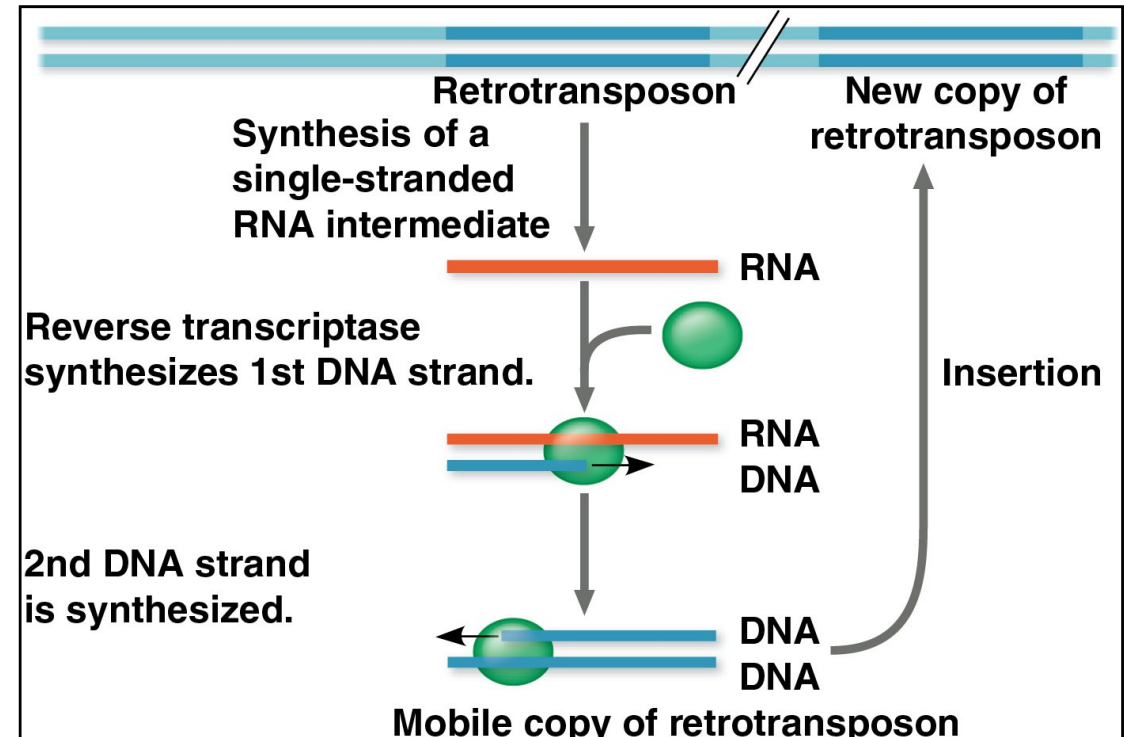
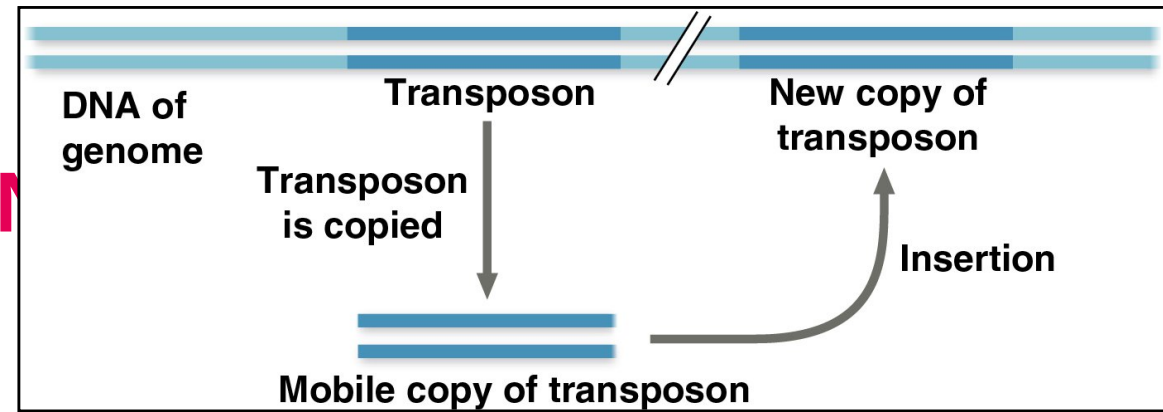
- Herhaling vorige weken
- Aminozuren, eiwitten en structuur niveau's.
- (Multi)gen families.
- Homologie: homologen, orthologen, paralogen, ...
- Campbell pagina 486-487 Genes and multigene families.
- Campbell 20.5.

HERHALING VORIGE WEKEN

- Het GC-percentage beïnvloedt de codon bias of de codon bias beïnvloedt het GC-percentage.
- Door de derde codon positie (wobble) kan de GC bias van een genoom of genoom regio heel erg correleren met de GC3 waarde (GC content op de derde codon).
 - Codons die eindigen met een G/C, zullen vaker gebruikt worden (bias) wanneer er een hogere GC percentage is. GC percentage veroorzaakt dus een codon bias.
 - De codon bias zal op hun beurt weer het gebruik van G/C in de codons beïnvloeden.
- Evolutionair is het voordelig om genen te hebben die passen bij de codon bias van het organisme (efficiëntere eiwit productie en mutatie bias (CG)).
- Belangrijke genen hebben een hogere GC percentage.
 - Stabiliteit of beter mutatie correctie?

HERHALING VORIGE WEKEN

- Transposon
 - Hebben een cut-and-paste of copy-and-paste mechanisme
 - Bewegen door een DNA intermediaire en hebben een enzym transposase nodig
- Retrotransposon (meest dominante vorm in eukaryoten)
 - Bewegen door een RNA intermediaire en gebruiken hierbij een reverse transcriptase.
 - Retrovirussen geëvolueerd uit dit systeem of andersom!

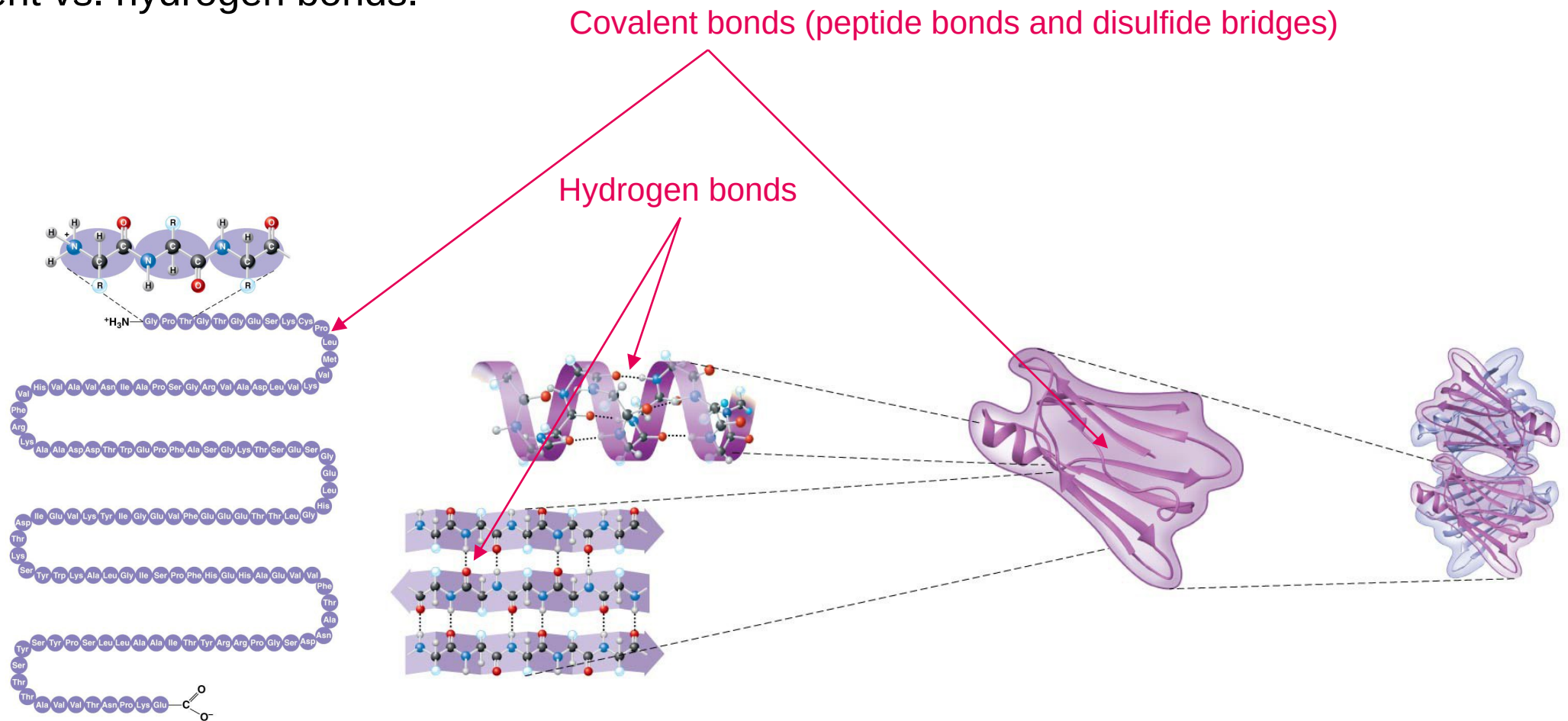


EIWIT STRUCTUUR NIVEAUS – WAT WETEN JULLIE NOG?

- Primaire structuur
- Secundaire structuur
- Tertiaire structuur
- Quaternaire structuur

EIWIT STRUCTUUR NIVEAUS – WAT WETEN JULLIE NOG?

Covalent vs. hydrogen bonds.



AMINOZUUR STRUCTUUR

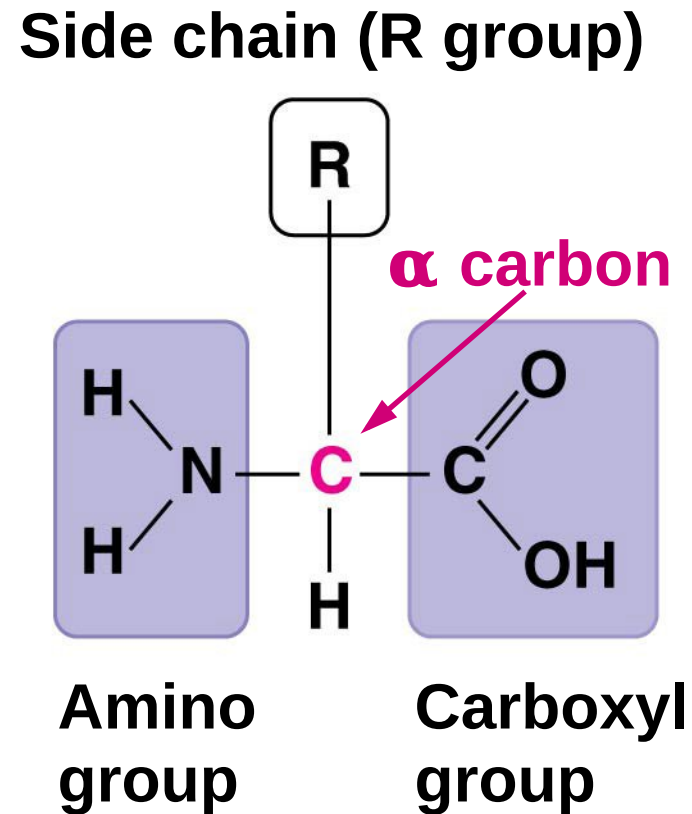


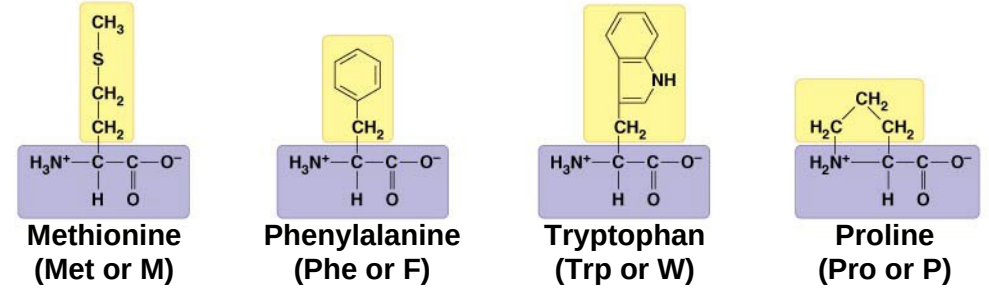
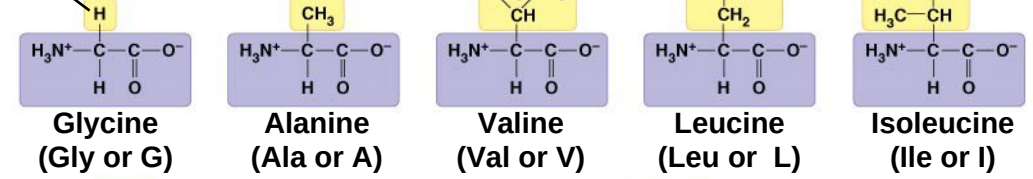
Fig. 5.14, Cambell Biology, 2021 Global edition

AMINOZUREN

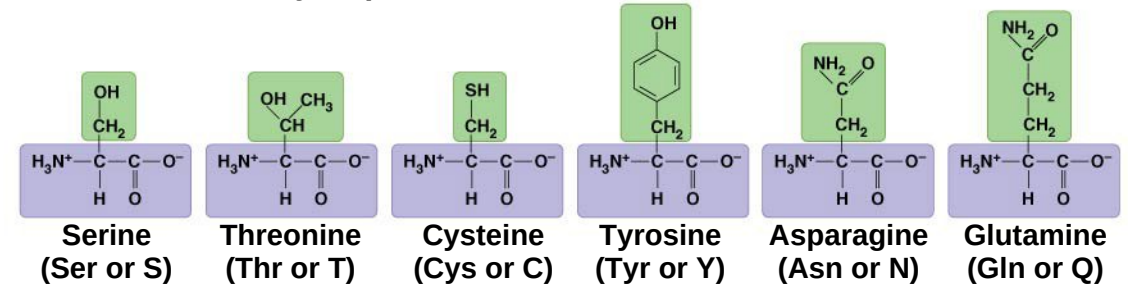
- Karakteriseren:
 - Hydrofobisch
 - Grootte
 - Lading
 - Aromatisch
 - Sulfaat bevattend
- Aparte aminozuren:
 - Gly heeft geen zijketen.
 - Cys maakt zwavelbruggen.
 - Pro betreft amine in zijketen.

Nonpolar side chains; hydrophobic

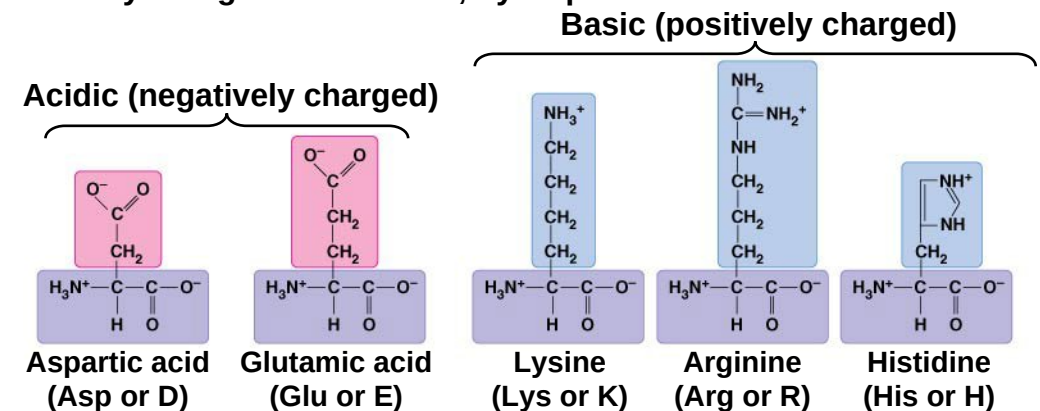
Side chain (R group)



Polar side chains; hydrophilic



Electrically charged side chains; hydrophilic



EIWIT PRIMAIRE STRUCTUU

- Sequentie van aminozuren
- Peptide binding doot condensatie reactie
- N-terminus (aminogroep)
- C-terminus (carboxylgroep)

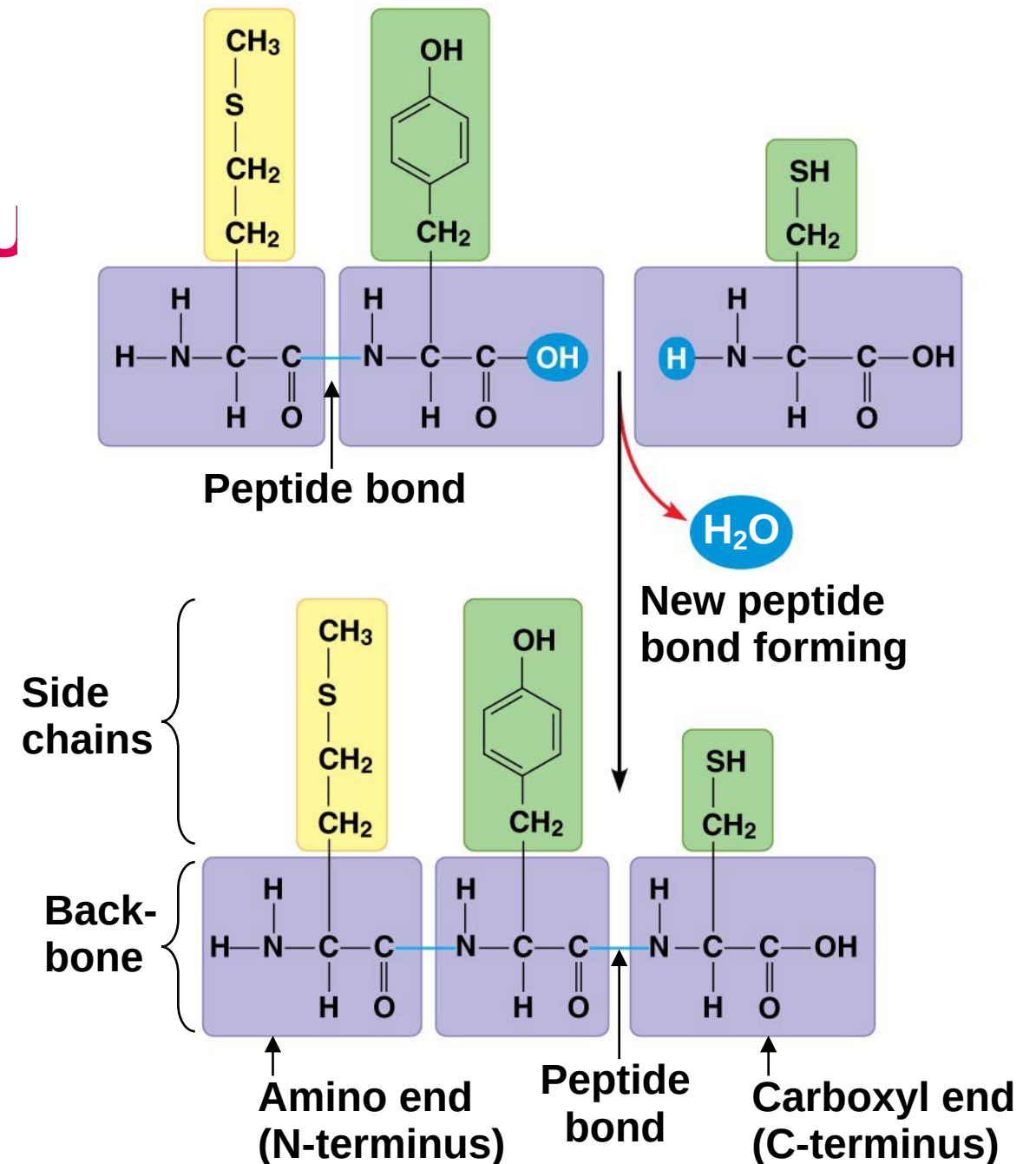
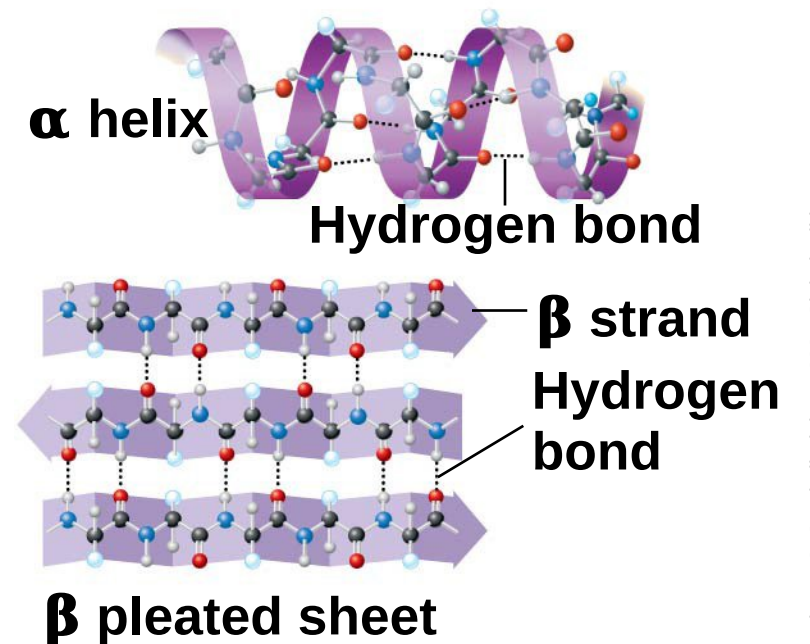


Fig. 5.15, Cambell Biology, 2021 Global edition

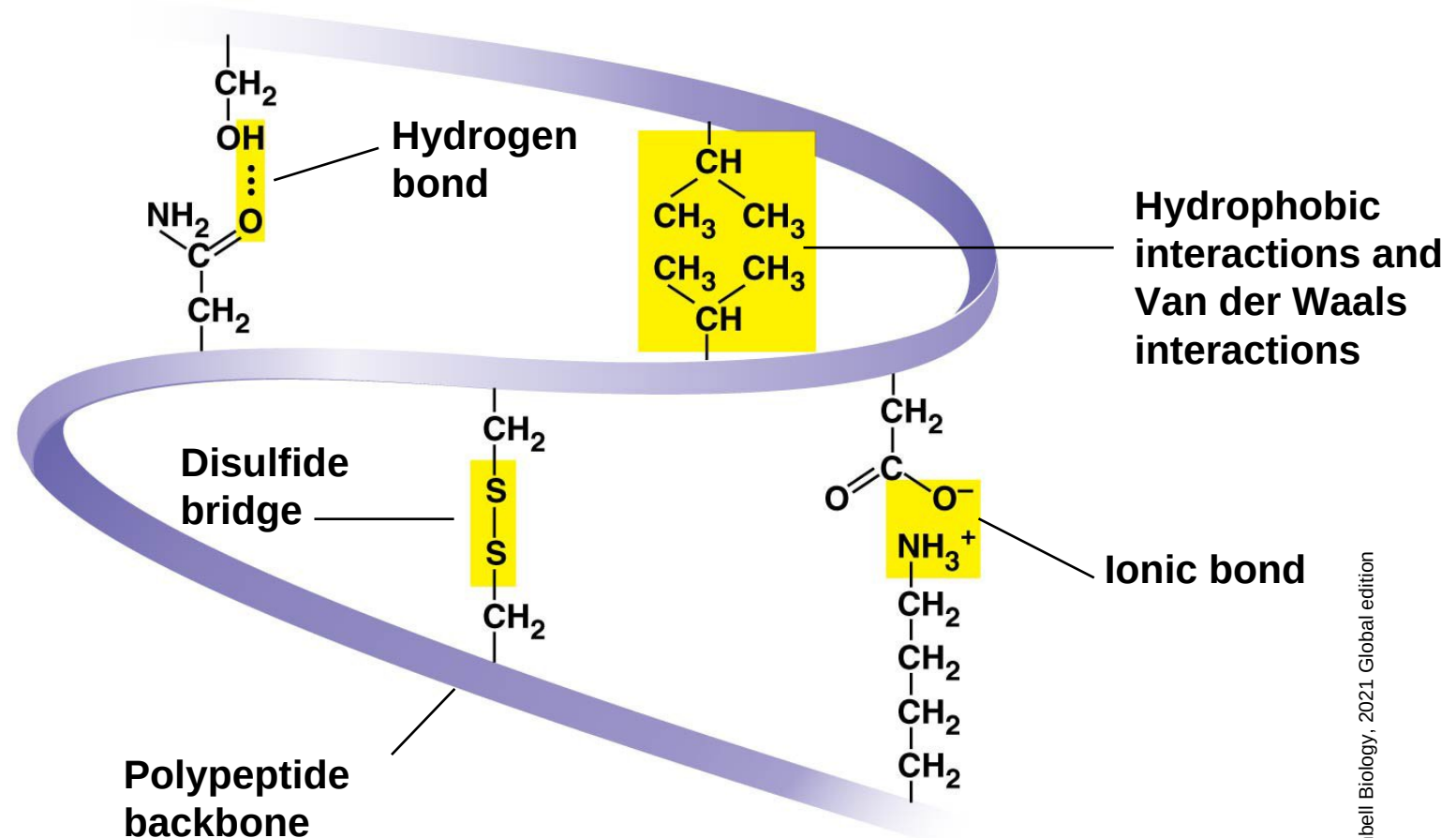
EIWIT SECUNDAIRE STUCTUUR

- Lokale regelmatigigheden in de conformatie door waterstofbruggen tussen carboxylgroep en amine groep.
- Waterstofbruggen tussen H en O.
- Verschillende manieren van vouwing, α -helix en β -sheets.



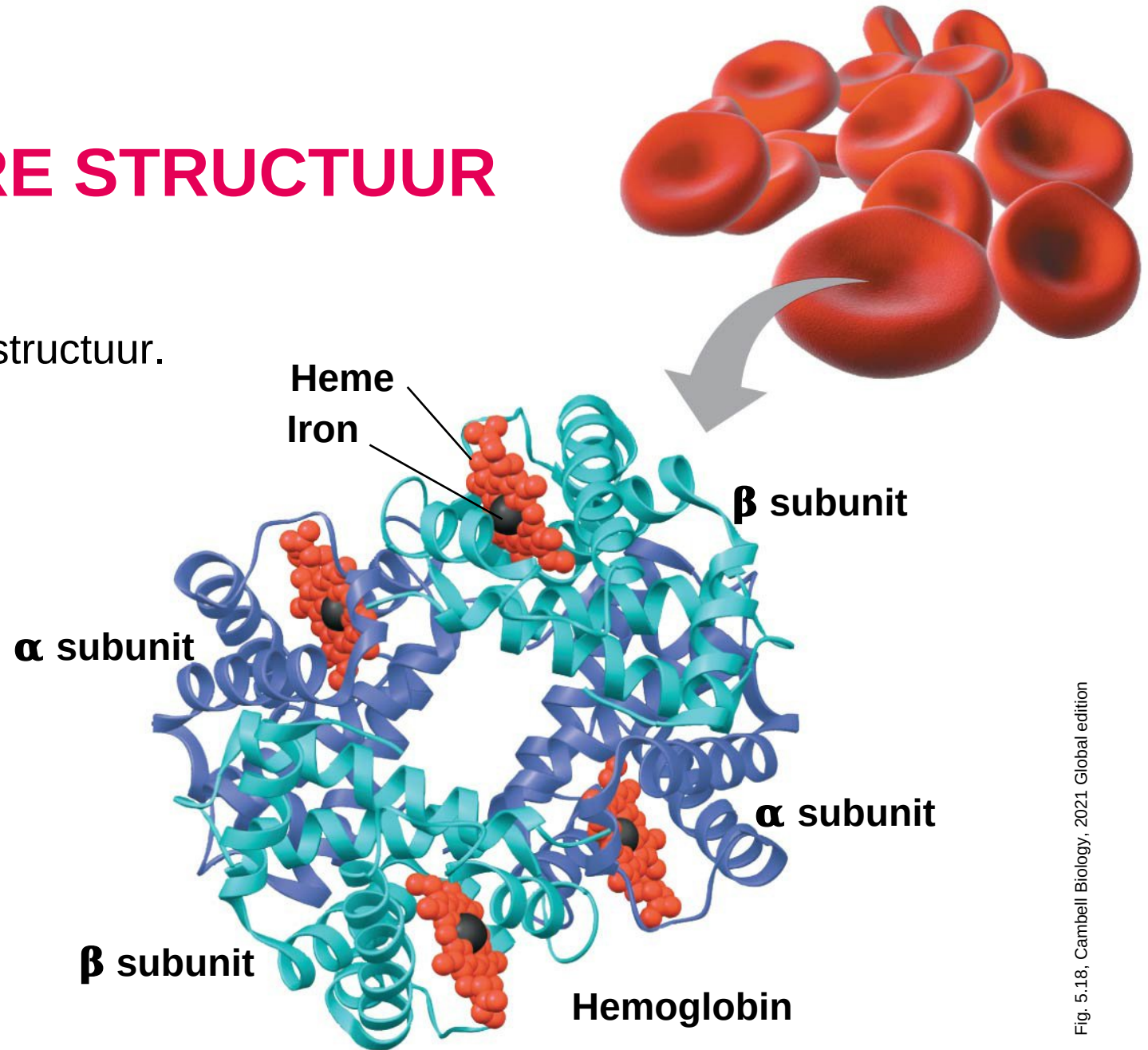
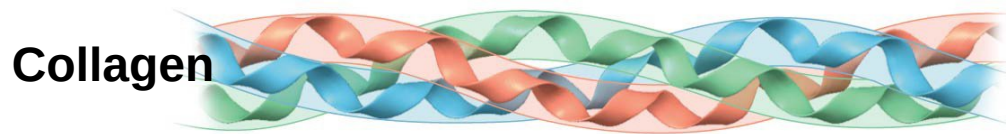
EIWIT TERTIAIRE STRUCTUUR

- Ontstaan door verbindingen tussen restgroepen
 - Waterstofbruggen
 - Ionverbindingen
 - Hydrofobe interacties
 - Zwavelbruggen
 - Van der Waals krachten



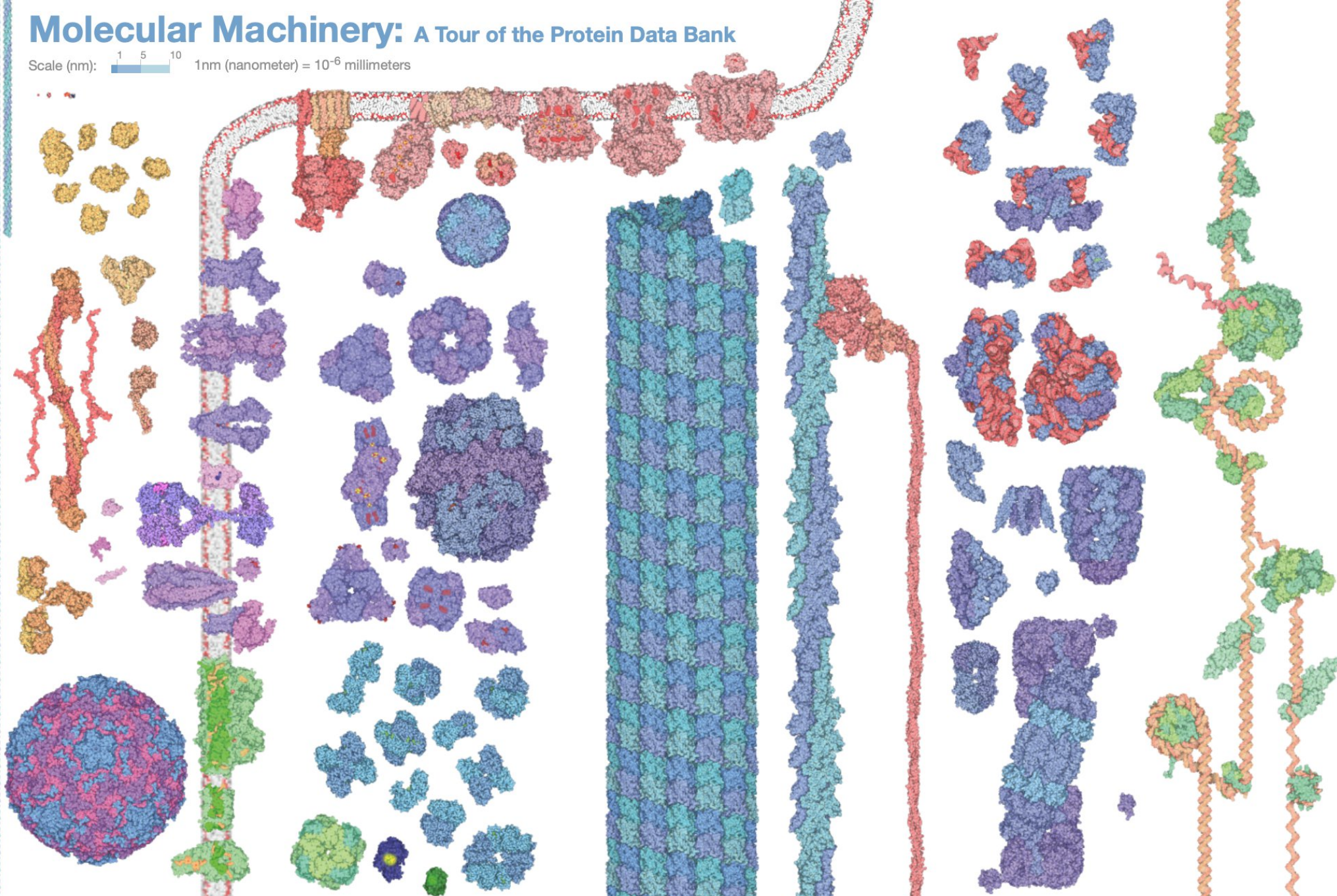
EIWIT QUARTERNAIRE STRUCTUUR

- Meerdere polypeptide ketens.
- Zelfde interacties als in de tertiaire structuur.



Molecular Machinery: A Tour of the Protein Data Bank

Scale (nm): 1 5 10 1nm (nanometer) = 10⁻⁶ millimeters



Structure Function

- Small molecules
- Digestive Enzymes
- Blood Plasma
- Viruses and Antibodies
- Hormones
- Channels, Pumps and Receptors
- Photosynthesis
- Energy Production
- Storage
- Enzymes
- Infrastructure
- Protein Synthesis
- DNA

- Fullscreen
- Auto
- About



Extracellular

Membrane

Intracellular/Cytosol

Intracellular/Nucleus

Cellular Location

<https://cdn.rcsb.org/pdb101/molecular-machinery/>

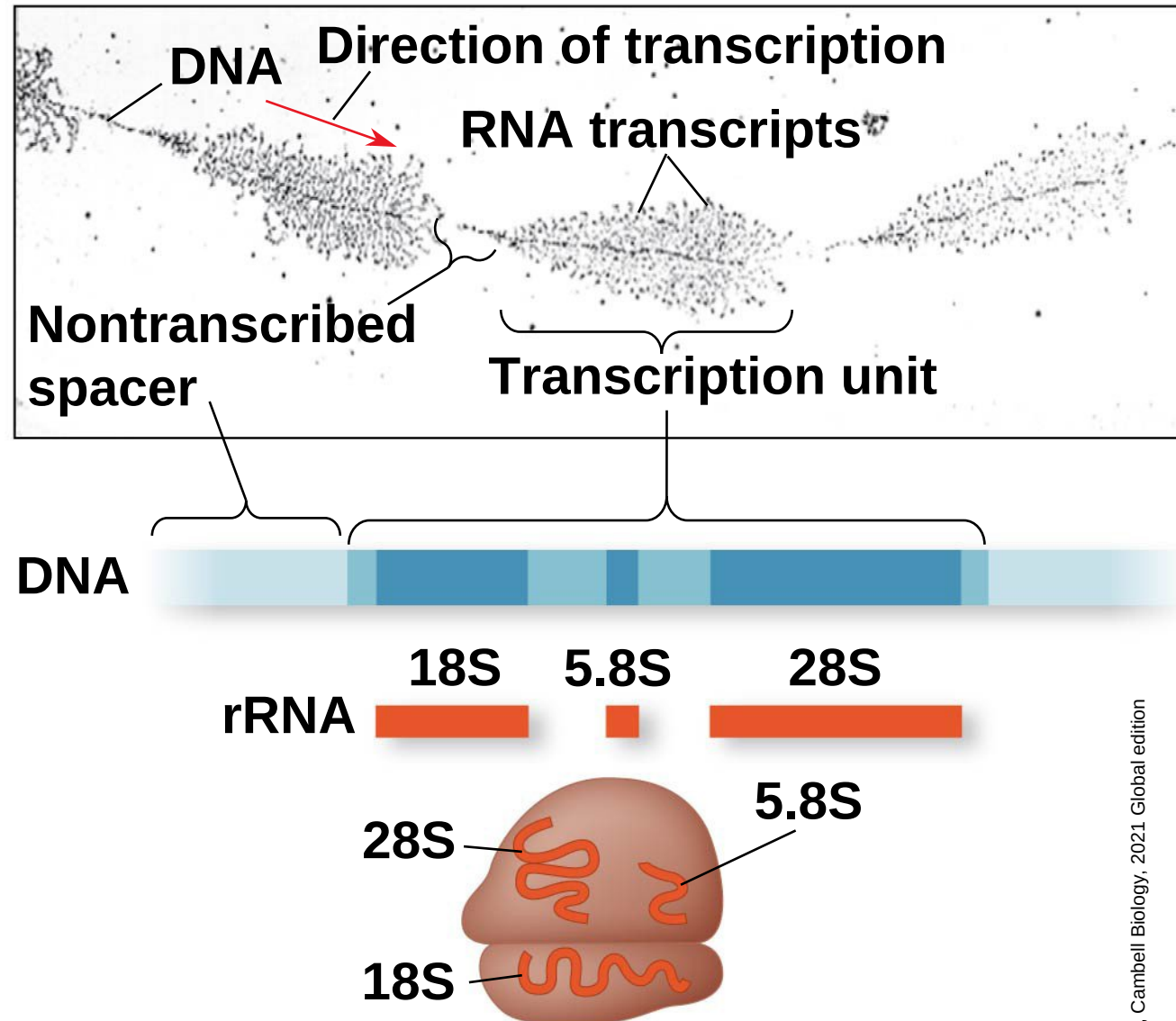
GENOOM EVOLUTIE DRIJVENDE KRACHTEN

- Genoom evolutie wordt teweeg gebracht door:
 - Duplicaties (wat multigen families en super families van genen veroorzaakt).
 - Van genen, chromosomen, exonen.
 - Genoom en gen reorganisatie.
 - Mutaties.

MULTIGEN FAMILIES

- Veel gene in eukaryoten komen voor als 1 kopie per haploide set van chromosomen.
- Families van genen die bestaan uit vergelijkbare sequenties en functies.
- Kunnen vlak bij elkaar liggen op een chromosoom, of op verschillende chromosomen liggen.
- Soms lijken de sequenties zo erg op elkaar dat ze nauwelijks te onderscheiden zijn.
- Omdat ze zo erg op elkaar lijken op sequentie en functioneel niveau, hebben ze vaak gedeelde regulatoire element, waardoor grote hoeveelheden product van deze gen families tegelijk afgeschreven kunnen worden.

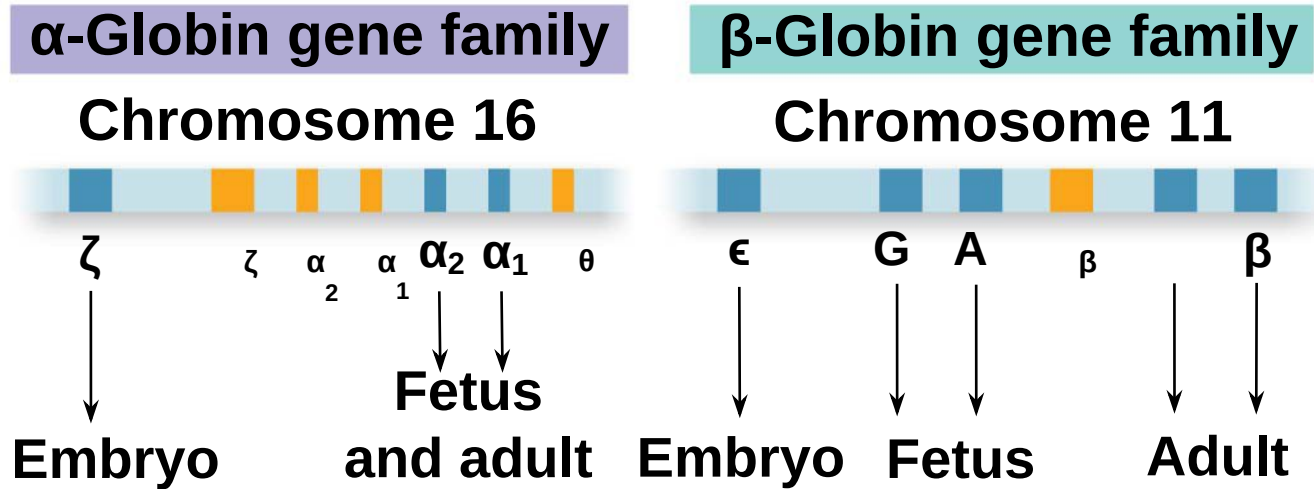
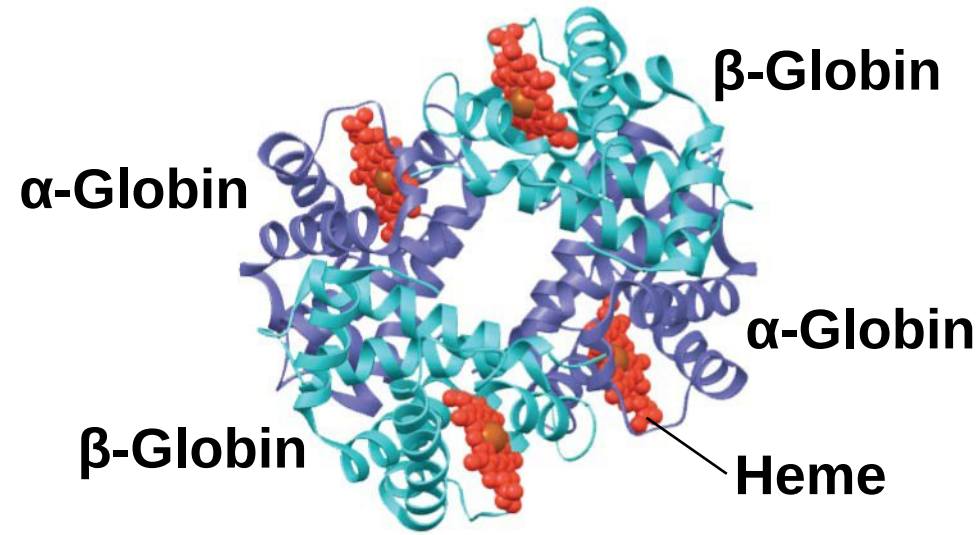
MULTIGEN FAMILIES



(a) Part of the ribosomal RNA gene family

Fig. 20.10, Cambell Biology, 2021 Global edition

MULTIGEN FAMILIES



= pseudogenen

Fig. 20.10, Cambell Biology, 2021 Global edition

(b) The human α -globin and β -globin gene families

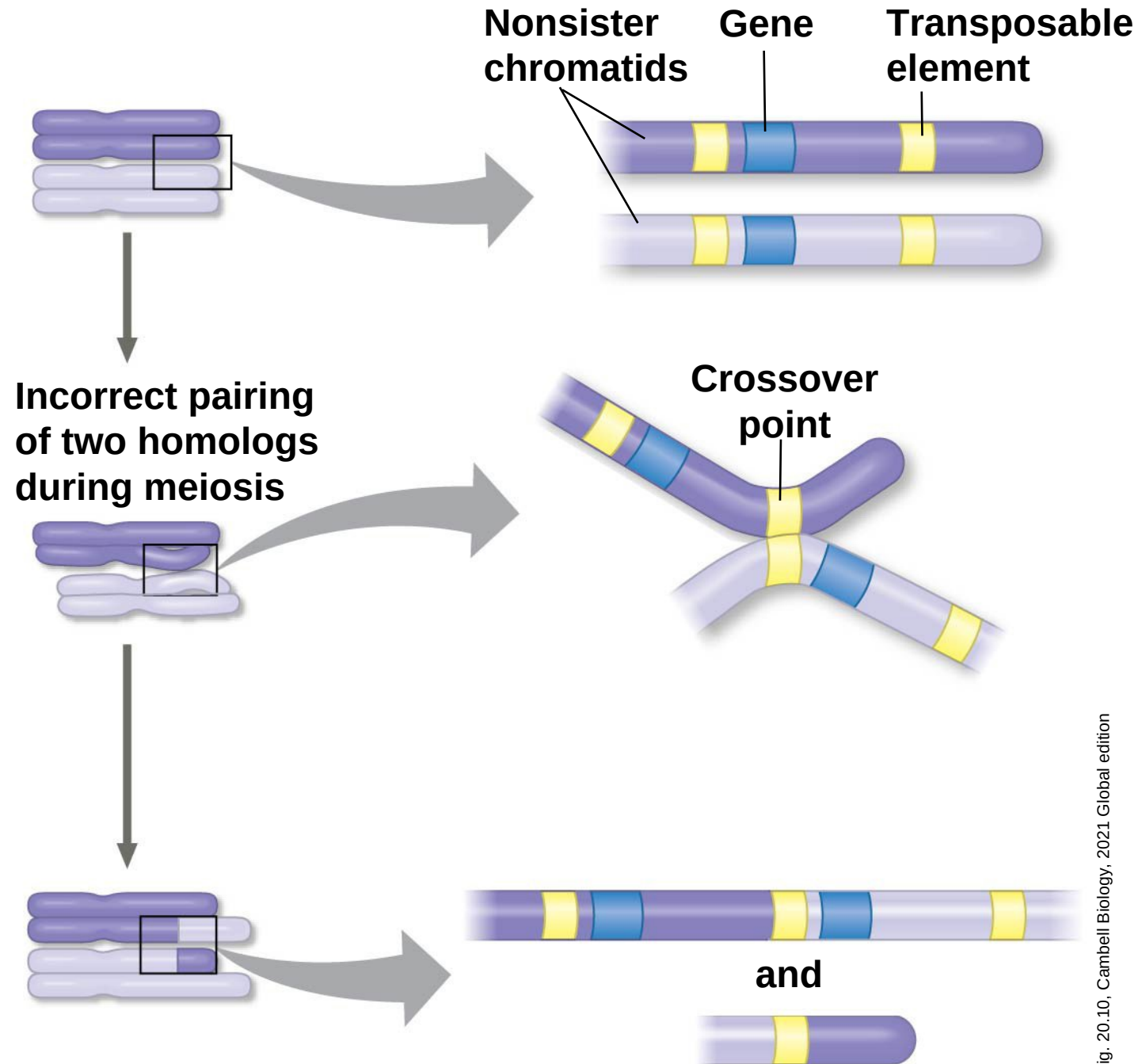
MULTIGEN FAMILIES FORMATIE

Multigen families spelen een hele grote rol in genoom evolutie en diversiteit van multicellulaire organismen.

- Ontstaan door meerde **duplicaties** (kopieën) van (voorouderlijke) genen, gevolgd door mutaties en divergentie (van de sequentie).
- Door **relocatie** van de genen kunnen leden van de families op verschillende chromosomen liggen.
- Door non-synonieme **mutaties** zorgen dat gedupliceerde genen een (net iets andere) functie kunnen krijgen.

DUPLICATIES

- Duplicate gebeurt alleen in ongelijke crossing-over tijdens de meiose wat resulteert in een chromosoom met een deletie en een andere chromosoom met een duplicatie.
- Transposabele elementen zorgen voor recombinatie, of crossing over, tussen chromosomen.



DUPLICATIES

Bij het ontstaan van gen families zijn 4 soorten niveaus van duplicatie:

1. Hele gen duplicatie. Vormen multigen families.
 2. Exon duplicatie en shuffeling. Geeft nieuwe variatie en nieuwe genen.
 3. Multigen familie duplicatie. Geven superfamilies op meerdere chromosomen.
 4. Heel genoom duplicatie. Verdubbeld het aantal chromosomen elke keer.
- > Haploid – diploid – triploid

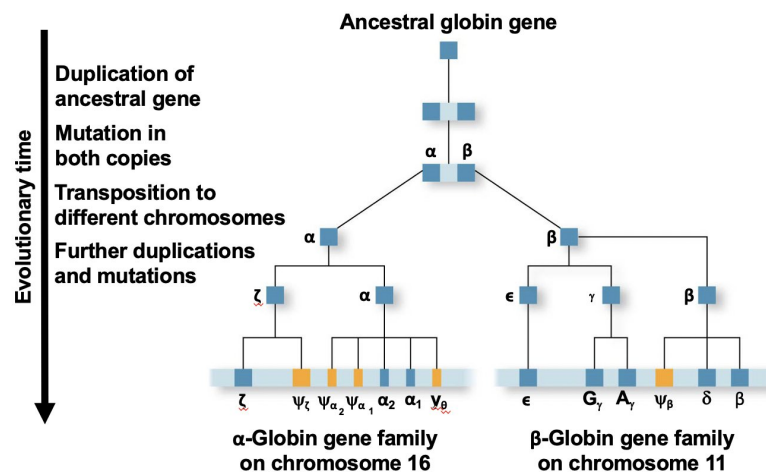


Fig. 20.14, Cambell Biology, 2021 Global edition

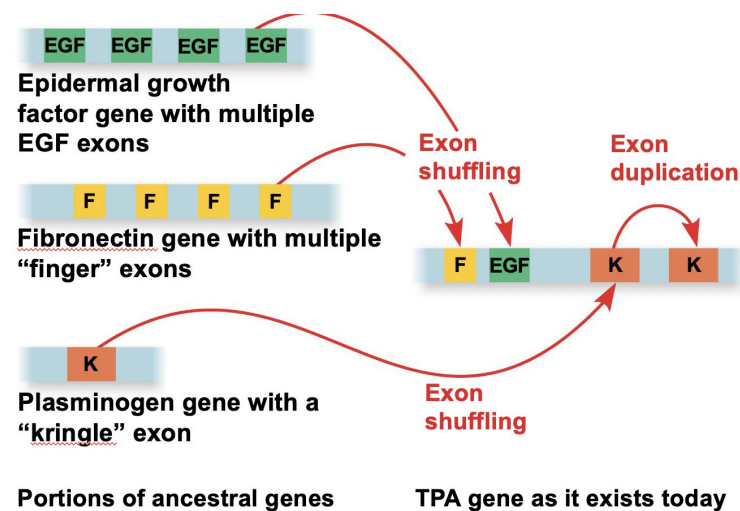


Fig. 20.16, Cambell Biology, 2021 Global edition

RELOCATIE VAN (GEDUPLICEERDE) GENEN

1. Wanneer twee transposons dicht genoeg bij elkaar liggen in een bepaalde regio op een chromosoom, kunnen ze een composiet transposon vormen. Transposase enzym herkent de transposons en knipt ze samen met alle genen die er tussen liggen (cargo DNA) en springen vervolgens naar een andere lokatie op het genoom.
2. Reverse transcriptie kan een stuk RNA terug kopiëren naar DNA en het in een andere chromosoom plakken (retrotransposonen LINEs (L1) and SINEs (Alu)).

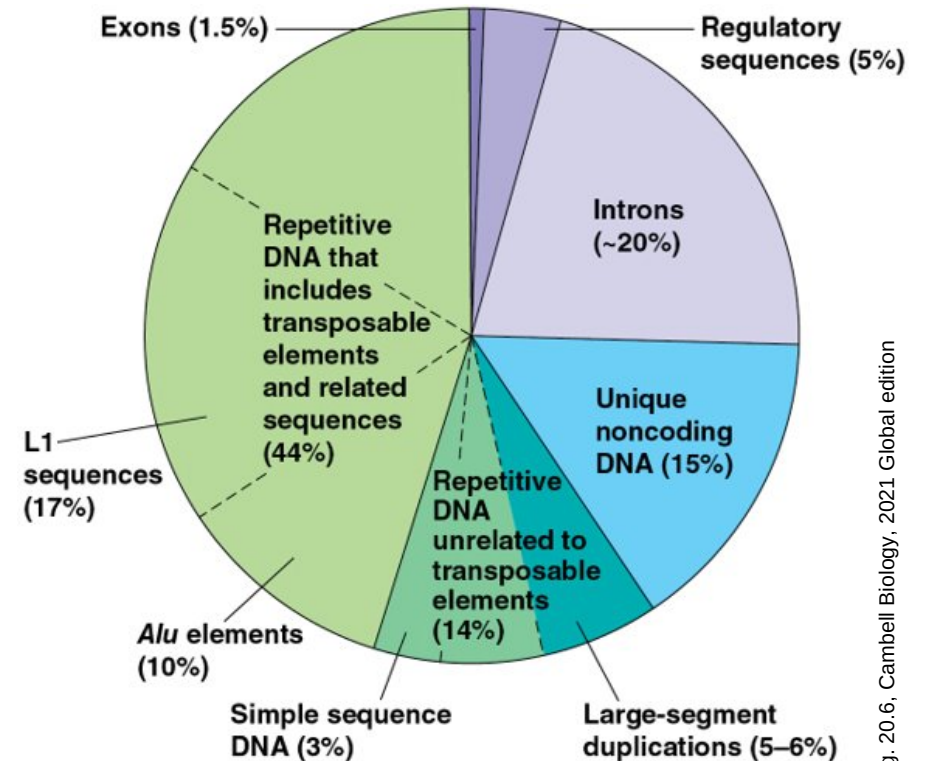
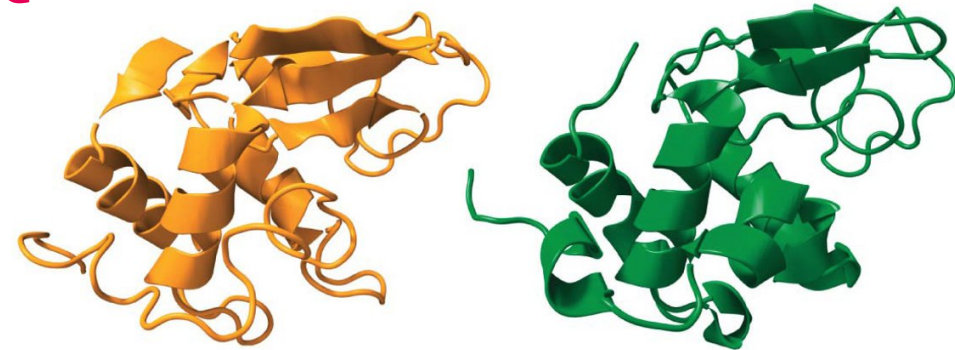


Fig. 20.6, Cambell Biology, 2021 Global edition

DIVERGENTIE VAN DE SEQUENTIES

- Gedurende de evolutie van gedupliceerde genen kunnen mutaties optreden en kunnen de genen zo divergeren dat ze een andere functie krijgen.
- Zoals Lysozyme die gedupliceerd is en ge-evolveerd is tot α -lactalbumin.
 - Lysozyme is een enzym die dieren tegen bacteriële infectie beschermd.
 - α -lactalbumin is een geen enzym die een rol speelt in melkproductie in zoogdieren.



(a) Lysozyme

(b) α -lactalbumin

Lysozyme	1	KVFERCELAR	TLKRLGMDGY	RGISLANWMC	LAKWESGYNT	RATNYNAGDR
α -lactalbumin	1	KQFTKCELSQ	LLK--DIDGY	GGIALPELIC	TMFHTSGYDT	QAIVENN--E
Lysozyme	51	STDYGIFQIN	SRYWCNDGKT	PGAVNACHLS	CSALLQDNIA	DAVACAKRVV
α -lactalbumin	51	STEYGLFQIS	NKLWCKSSQV	PQSRNICDIS	CDKFLDDIT	DDIMCAKKIL
Lysozyme	101	RD PQGIRAWV	AWRNRCQ-NR	DVRQYVQCG	V	
α -lactalbumin	101	D- IKGIDYWL	AHKALCT--E	KLEQWLCEKL	-	

(c) Amino acid sequence alignments of lysozyme and α -lactalbumin

VERANDERING IN GENOOM STRUCTUUR

- Mens heeft 23 paar chromosomen
- Chimpanzee heeft 24 paar chromosomen
- Na de divergentie van de mens van chimpanzee van hun gemeendschappelijke voorouder, zijn twee voorouderlijke chromosomen gefuseerd in mens.
 - Dit weten we door vergelijkende genoom analyse tussen de chromosomen.

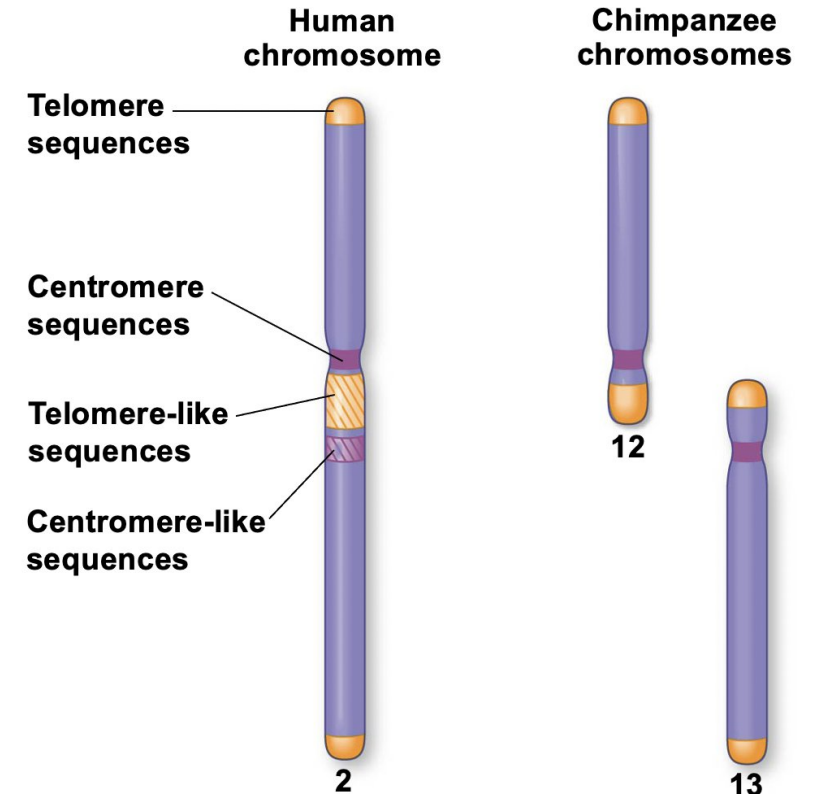


Fig. 20.11, Cambell Biology, 2021 Global edition

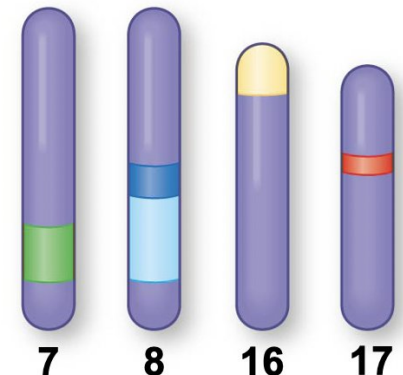
VERANDERING IN GENOOM STRUCTUUR

- Tussen mens en muis is meer tijd verstreken en ligt de gezamenlijke voorouder veel verder terug in de tijd.
- Met vergelijkende genoom analyse krijgen we wel een hypothetische chromosomale evolutie tijdlijn.
- Blokken van genen in mens en muis lijken bij elkaar te blijven liggen.
- Tijdens de evolutie: duplicatie en inversies van grote stukken chromosoom door fouten in mitotische recombinitie waarbij DNA is gebroken en foutief bij elkaar is gezet.

Human chromosome

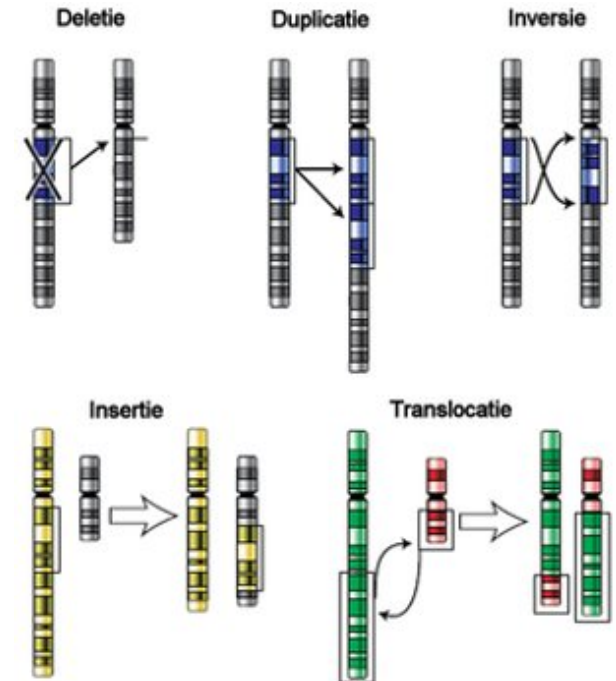


Mouse chromosomes



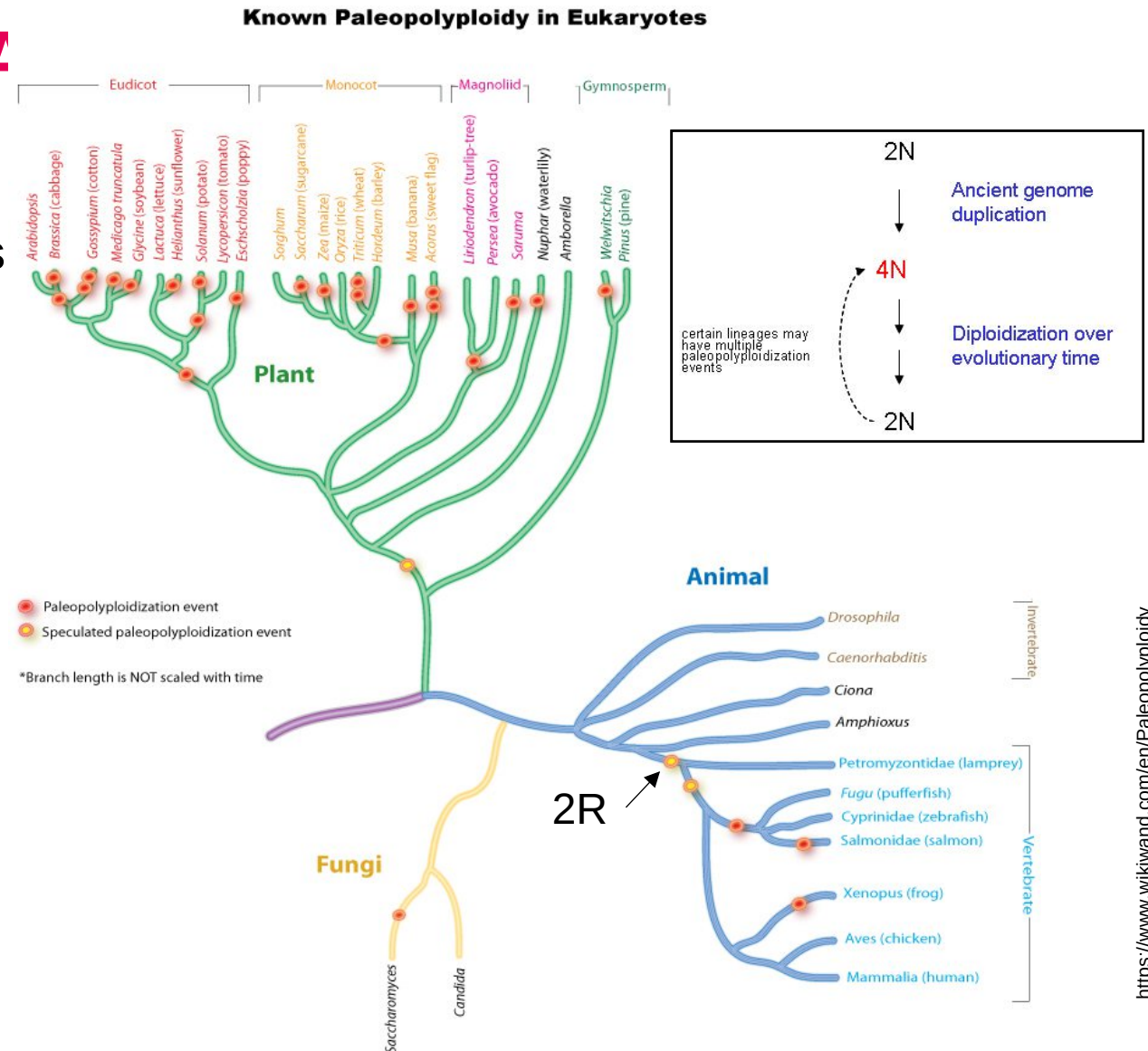
VERANDERING IN GENOOM STRUCTUUR (LAST WEEK)

- Segment (chromosomale) mutaties
- Grote stukken op een chromosoom
- Ploidie mutaties
- Aneuploïdie: te weinig of teveel chromosomen.
- Euploïdie: vermeerdering of vermindering van aantal chromosomen.
- Deleties en duplicaties
- Etc...



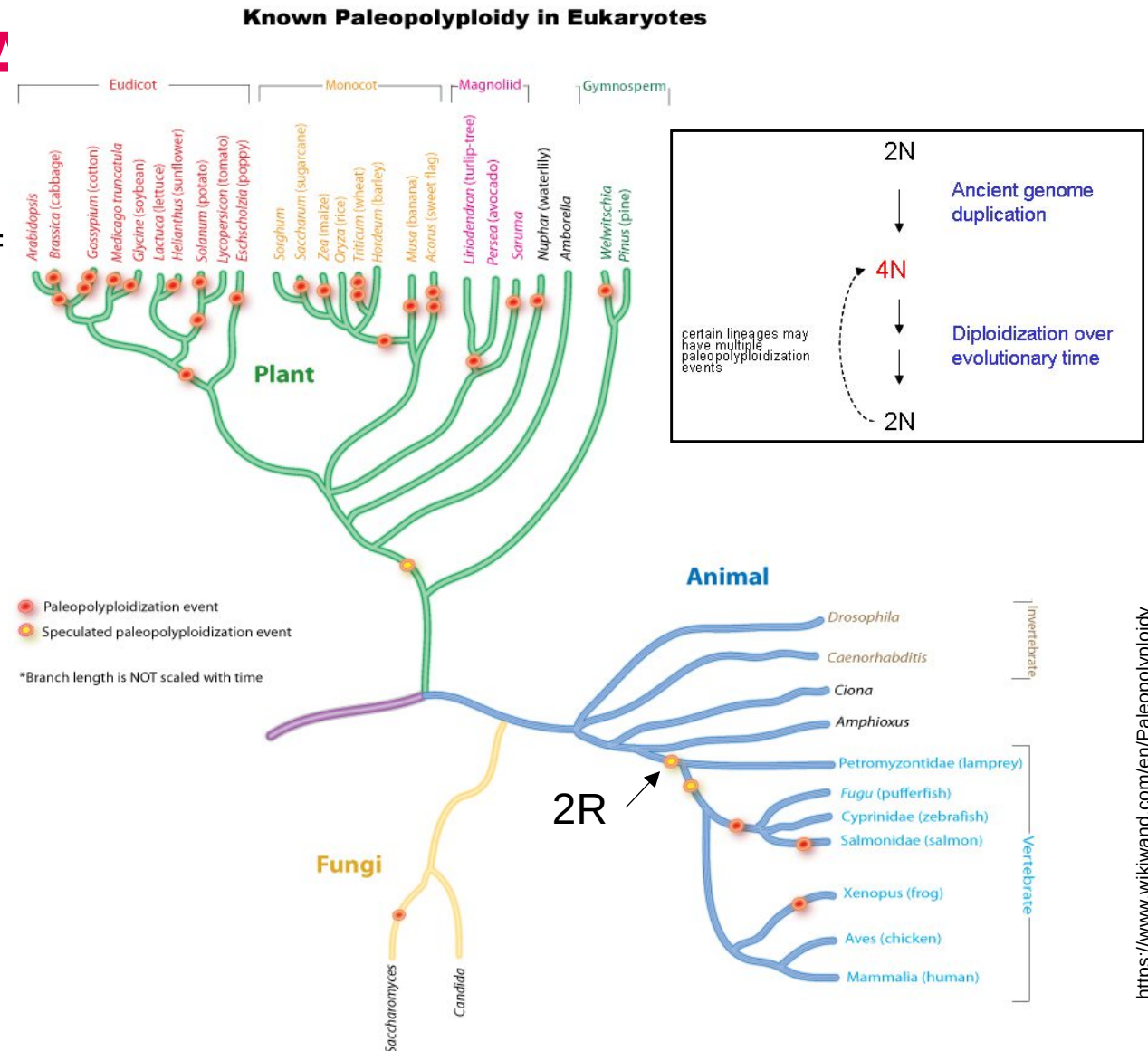
OUDE GENOOM DUPLICATIE

- Paleopolyploidy: ancient polyploïdie.
- Het resultaat van hele genoom duplicaties die tenminste een aantal miljoen jaar geleden (MYA) hebben plaatsgevonden.
- Verdubbeling van genoom in enkel organisme (autopolyploïdie) of gecombineerde genomen van twee organismen (allopolyploïdie).
- Planten zijn bekend om hun vele genoom duplicaties.
- In dieren is het zeldzaam.
 - 2R-hypothese: gedeelde hele genoom duplicatie in vertebraten.



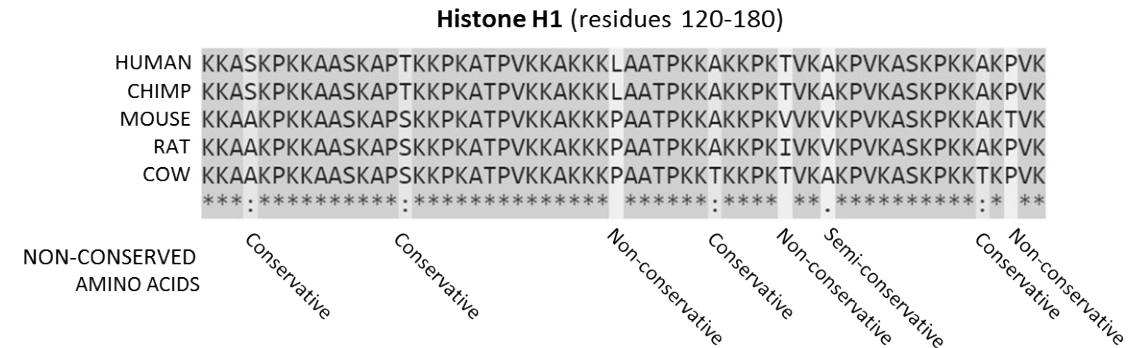
OUDE GENOOM DUPLICATIE

- Detectie is moeilijk maar kan door sequentie homologie te kijken van DNA of eiwitten. Een hele genoom duplicatie als:
 - Gedupliceerde genen in grote gedupliceerde blokken naast elkaar liggen.
 - Deze blokken overlappen niet.
- Door diploidizatie is het moeilijk om hele oude genoom duplicaties terug te vinden.
 - Diploidizatie: het verlies van de polyploïde status door verlies van gedupliceerde genen die overbodig worden.
- Dit alles maakt het opsporen van de oorspronkelijk voorouderlijk genen heel moeilijk



HOMOLOGE SEQUENTIES

- Homologie tussen sequenties duidt op een gemeenschappelijk voorouderlijk gen.
- Homologie wordt gedetecteerd door (multiple) sequence alignments.
- Hierbij wordt gekeken naar **sequentie similariteit**.
 - Percentage identieke base/aminozuren (percent identity).
 - Percentage vergelijkbare aminozuren (percent similarity).

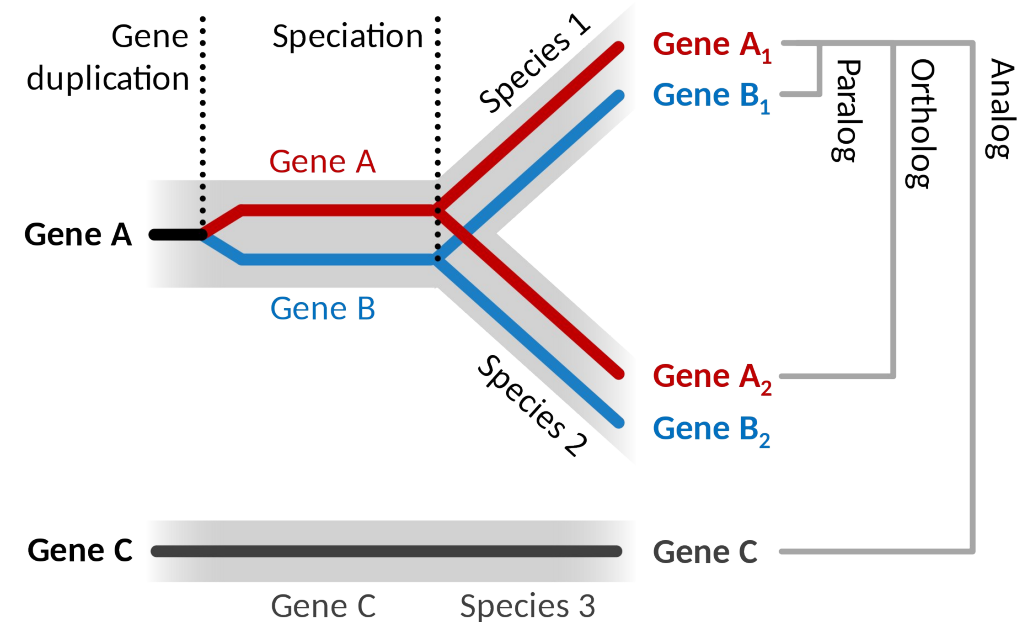


HOMOLOGIE

- Homologie is **transitief**. Als gen (of eiwit) A homolog aan gen B, en gen B homolog is aan gen C, dan zijn gen A en C ook homologen van elkaar.
- Homologie is **qualitatief**. Iets kan homolog zijn of niet. In andere woorden:
 - *Niet*: gen A is meer homolog aan gen B dan gen C.
 - *Ook niet*: gen A is 90% homolog aan gen B en 80% homolog aan gen C.
 - *Wel*: gen A heeft meer percentage sequence identity/similarity met gen B dan gen C.
- Met lage percentage identiteit kan er nog steeds sprake zijn van homologie.
- Waarschijnlijk homolog:
 - > 100 aminozuren > 25% identiteit
 - > 100 nucleotiden > 70% identiteit.
 - Dit kan afwijken van hoe ver je terug in de tijd wil gaan, bv. organismen die ver van elkaar staan (mens en gist) of dicht bij elkaar (mens en aap).

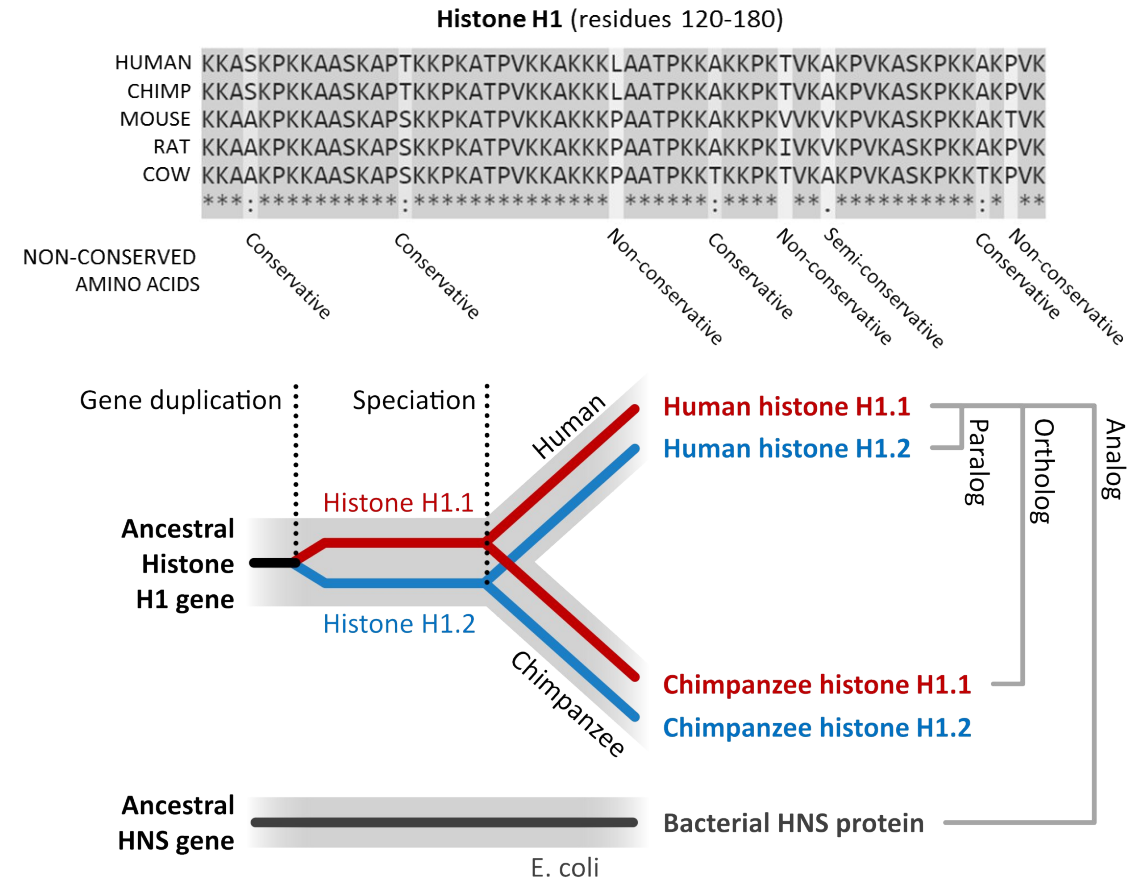
HOMOLOGE SEQUENTIES – VERSCHILLEN IN RELATIES

- Sequenties kunnen op verschillende manieren een gedeeld voorouderlijke sequentie hebben:
 - Door een speciation (soortvorming): een gen in een voorouderlijk organisme bestaat nu in twee aparte organismen. We noemen dit **orthologen**.
 - Door een duplicatie in: een gen in een (voorouderlijk) organisme is gedupliceerd en beide kopieën kunnen zich bevinden in nieuwe organismen. We noemen dit **paralogen**.



HOMOLOGE SEQUENTIES – VERSCHILLEN IN RELATIES

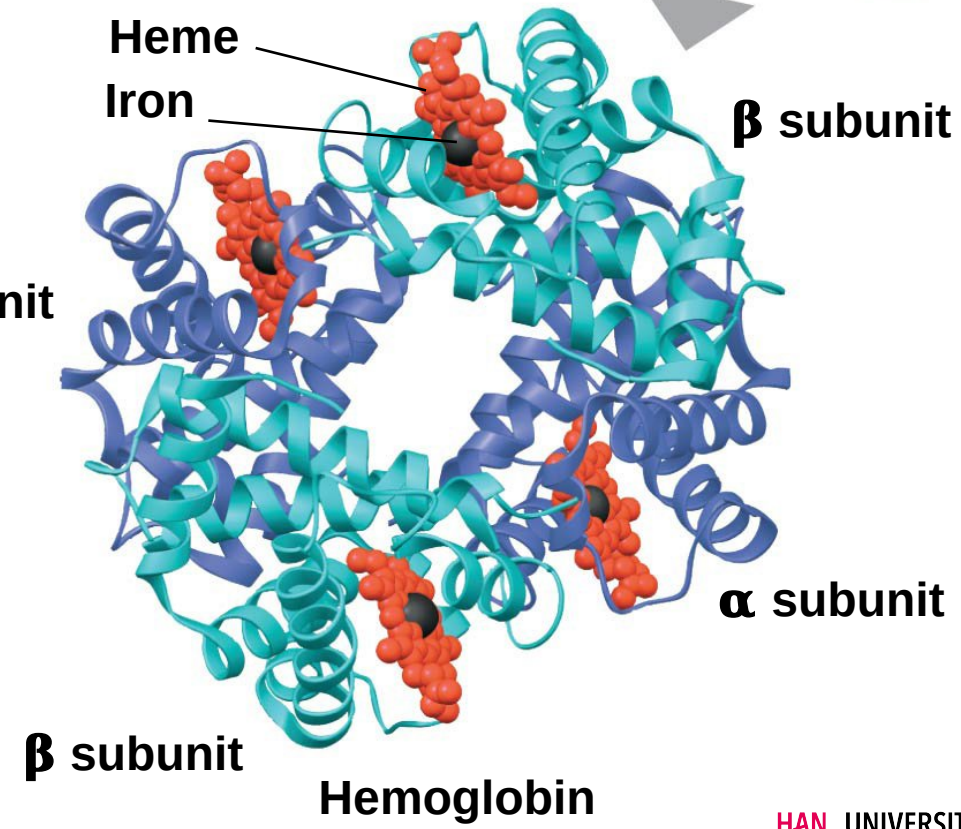
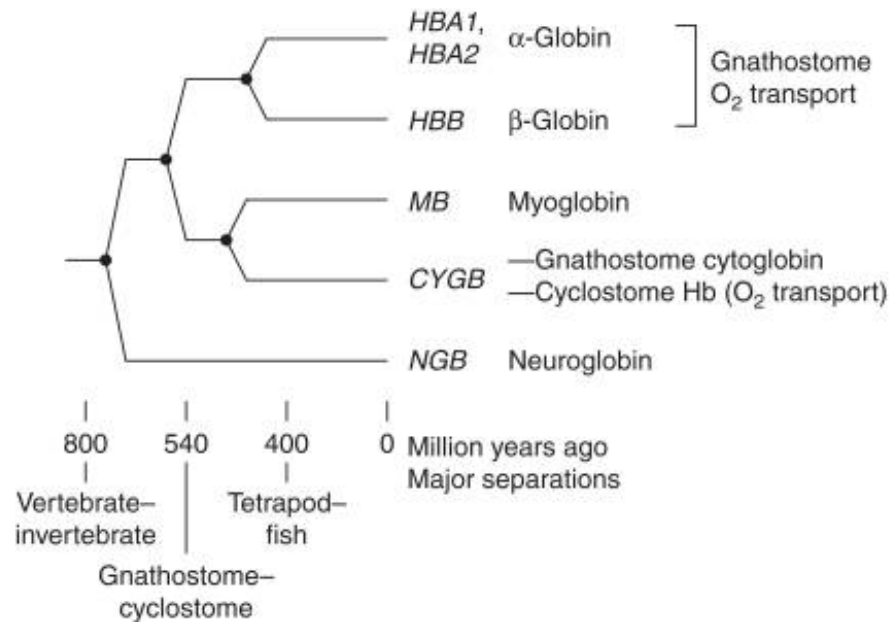
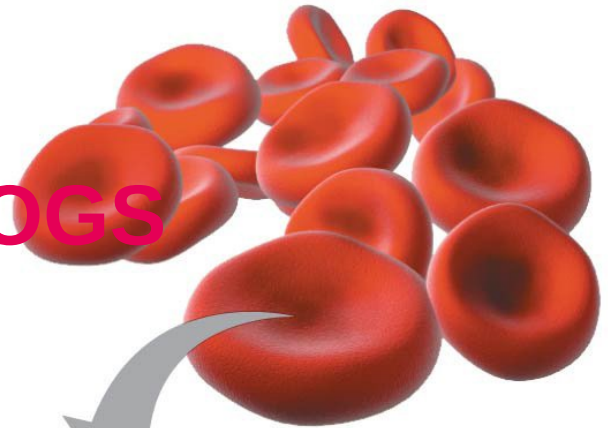
- Sequenties kunnen op verschillende manieren een gedeeld voorouderlijke sequentie hebben:
 - Door een speciation (soortvorming): een gen in een voorouderlijk organisme bestaat nu in twee aparte organismen. We noemen dit **orthologen**.
 - Door een duplicatie in: een gen in een (voorouderlijk) organisme is gedupliceerd en beide kopieën kunnen zich bevinden in nieuwe organismen. We noemen dit **paralogen**.



ORTHOLOGIE VOOR FUNCTIE VAN EIWITTEN

- Orthologen kunnen helpen bij het vergelijken van functies van eiwitten. Als we bijvoorbeeld weten dat in een gist een bepaald eiwit wat doet (door experimenten), kunnen door orthologie afleiden wat de functie is van het orthologe eiwit in mens.
- We noemen dit ook wel de orthology conjecture: orthologen zijn functioneel meer vergelijkbaar dan paralogen.

HEMOGLOBIN ORTHOLOGS AND PARALOGS



GENOOM EVOLUTIE

- Snelheid van duplicaties en inversies lijken de laatste 100 miljoen jaar geleden te zijn versneld.
- Dit komt overeen met het uitsterven van de dinosauriërs en de zoogdieren konden diversifiëren.
- Duplicaties, chromosomale herschikking en hele genoom duplicaties hebben een bijdrage in het ontstaan van nieuwe organismen.
- Voorouderlijke genen zijn gerelateerd door homologie, welke kan worden onderverdeeld in orthologen en paralogen.

NEXT LESSON

- OnderwijsOnline weektaken/theory Aminozuren.doc