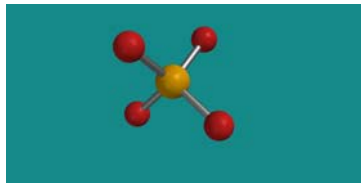


Basepar i DNA.

Innledning

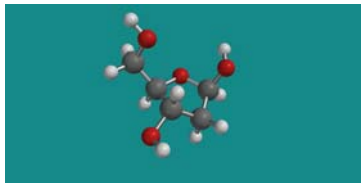
DNA, *deoxyribonucleic acid* er molekylerne som inneholder biologiske cellers genetiske informasjon. DNA er en såkalt biopolymer med byggesteiner (monomere) sammensatt av

a) en fosfatgruppe, PO_4^- (P: orange, O: rød)



Fosfat

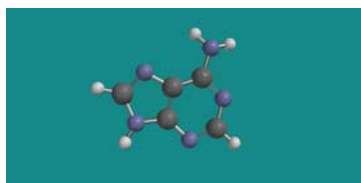
b) et ringformet karbohydrat, sukkeret 2-deoksyribose, $\text{C}_5\text{H}_{10}\text{O}_4$ (C: grå, H: hvit)



2-deoksyribose

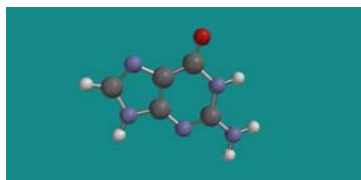
og c) en av fire aminobaser, enten

adenin (A), $\text{C}_5\text{H}_5\text{N}_5$ (N: blå)



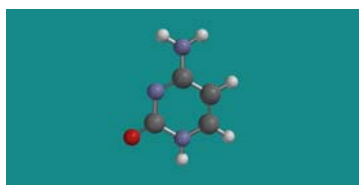
Adenin

guanin (G), $\text{C}_5\text{H}_5\text{N}_5\text{O}$



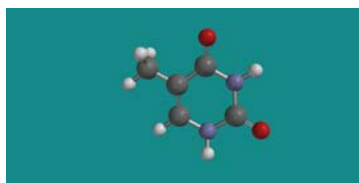
Guanin

cytosin (C), $C_4H_5N_3O$



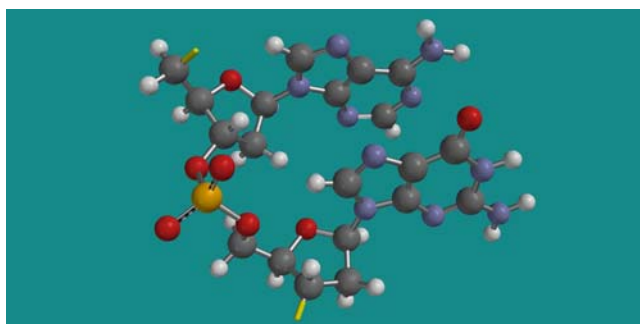
Cytosin

eller thymin (T), $C_5H_6N_2O_2$



Thymin

Aminobasen erstatter hydroksylgruppen (OH) på karbonatomet i posisjon 1 (C1) på karbohydratet (der vi nummererer C-atomene, fra 1 for karbonet umiddelbart til høyre for oksygenatomet i femringen, til 5 for karbonet i CH_2OH -gruppen). Deretter hektes slike "nukleosider" (dvs aminobase + sukker) sammen til lange kjeder ved hjelp av fosfatgrupper mellom C3 i en nukleosid og C5 i en annen. Eksempelvis er en nukleosid med aminobasen guanin (G, nederst) bundet sammen med en nukleosid med aminobasen adenin (A, øverst) på dette viset i figuren nedenfor:



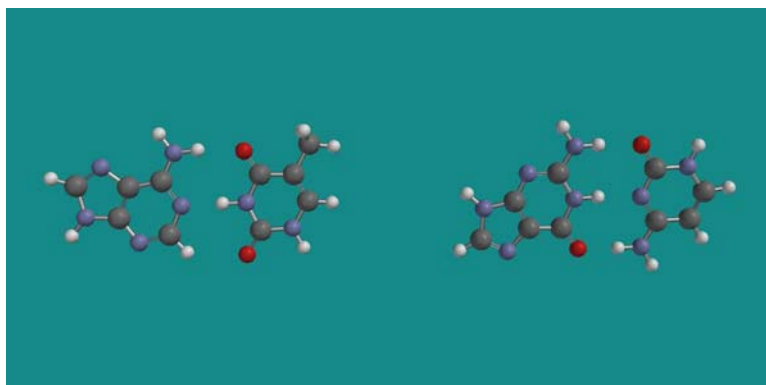
Guanin (nederst) og adenin (øverst)

De gule pinnene (på C3 på sukkeret nederst og C5 på sukkeret øverst) representerer åpne valenser der nye nukleosider kan hektes på, hele tiden med en PO_4 -gruppe som bindeledd. Den repeterende enheten {fosfat + sukker + aminobase} kalles en *nukleotid*. En lang kjede av slike nukleotider kalles en *polynukleotid*.

Man fant tidlig ut at DNA-molekyler inneholder like mye adenin som thymin, og like mye guanin som cytosin, mens innholdet av adenin/thymin i forhold til guanin/cytosin varierer fra art til art. Menneskelig DNA inneholder for eksempel om lag 30% adenin og thymin og om lag 20% guanin og cytosin.

Disse observasjonene var sentrale da James Watson og Francis Crick i 1953 kom opp med sin modell for hvordan DNA er bygget opp, nemlig som en *dobbelspiral* med to polynukleotider tvunnet i hverandre. De to spiralene holdes sammen av såkalte *hydrogenbindinger* mellom en A på den ene spiralene og en T på den andre, eventuelt mellom en G på den ene spiralene og en

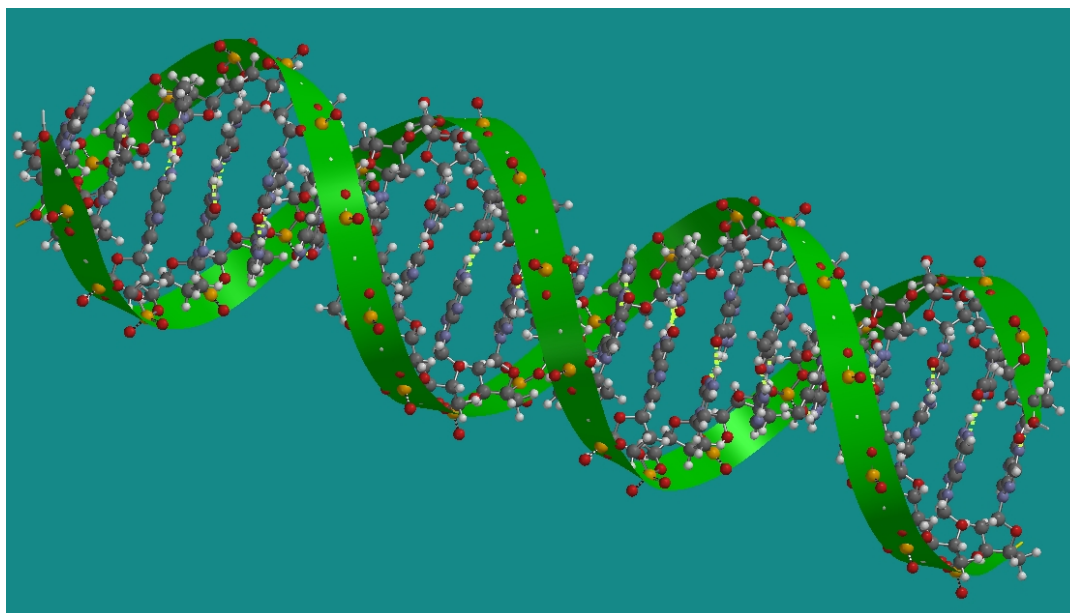
C på den andre. Figuren nedenfor viser de to *Watson – Crick baseparene* A-T (til venstre) og G-C (til høyre):



Watson – Crick basepar: A – T til venstre, G – C til høyre

Hydrogenbindingene som det her er snakk om, er relativt svake bindinger mellom et H-atom på en av basene og enten et O- eller et N-atom på den andre. I baseparet A-T har vi to slike bindinger, en mellom H på adenin og O på thymin (se figuren, øverst), den andre mellom N på adenin og H på thymin ("på midten"). I baseparet G-C har vi tre hydrogenbindinger, mellom H på G og O på C (øverst), mellom H på G og N på C (på midten) og mellom O på G og H på C (nederst). Tilsvarende kjenner du godt fra vann, både i flytende og fast form, hvor vi har svake hydrogenbindinger mellom O-atomet i ett vannmolekyl og et av H-atomene i et annet. Uten slike bindinger ville vann hatt mye lavere smelte- og kokepunkt enn det faktisk har, med dramatiske konsekvenser for både du og jeg og dompapen...

Figuren nedenfor viser en DNA dobbelspiral med 20 monomere i hver av de to polynukleotidene:



DNA dobbelspiral

De grønne båndene går gjennom P-atomene i fosfatgruppene og er kun et visuelt hjelpemiddel for lettere å gjenkjenne de to spiralene. Gule stiplete linjer angir hydrogenbindinger mellom de to basene i hvert basepar.

I denne øvingen skal du bygge og inspisere en DNA dobbelspiral. Du skal også se nærmere på et guanin – cytosin (G – C) basepar: Bindingsenergien knyttet til de tre hydrogenbindingene skal beregnes ved hjelp av Hartree – Fock – beregninger på baseparet G – C og basene G og C hver for seg. Baseparet inneholder så mange atomer at en Hartree – Fock – beregning kanskje vil ta bortimot en halvtimes tid. Derfor setter vi i gang denne beregningen først.

Det er helt essensielt at de to spiralene i DNA ikke er *for* sterkt bundet sammen: Rekkefølgen av baser (for eksempel A-G-G-A-C-C-T-A-G-...) representerer organismens genetiske informasjon, og når organismen vokser, *kopieres* denne informasjonen ved at dobbelspiralen åpnes opp – ved at hydrogenbindingene i baseparet brytes – hvorefter hver av de to spiralene slår seg sammen med nye spiraler. De nye dobbelspiralene blir eksakte kopier av den opprinnelige, i og med at kun A og T, eventuelt G og C kan danne nye basepar.

I en celle finner du 23 par kromosomer, dvs 46 kromosomer i alt. Hvert kromosom er et langt DNA-molekyl, bestående av mange segmenter som kalles *gener*. Det såkalte *genomet* tilsvarer summen av alle genene i en celle. Det er sannsynligvis et sted mellom 20000 og 25000 gener i en menneskelig celle, og til sammen består hele genomet av anslagsvis $3 \cdot 10^9$ basepar. Et kromosom kan være mer enn 10 cm langt.

Gjennomføring og oppgaver

1. Start SPARTAN ved å velge  under Programs.

2. Velg *Nucleotide* byggemeny, trykk "G" en gang, og deretter i arbeidsfeltet. Valget av guanin gir deg automatisk den "komplementære" basen cytosin, og derved baseparet G – C. Fjern atomer inntil du står igjen med G – C baseparet slik som til høyre i figuren øverst på side 3. Pass på at du står igjen med $C_5H_5N_5O$ (guanin) og $C_4H_5N_3O$ (cytosin). Lagre G – C baseparet i en katalog DNA, for eksempel med filnavn *gua_cyt.spartan*.

Baseparet G – C er nå lagret som molekyl nr 1 i filen *gua_cyt* ("gua_cyt:M001", se øverst til venstre i SPARTAN-vinduet). Du skal deretter lagre de to basene G og C hver for seg, i den samme filen *gua_cyt*, som henholdsvis molekyl nr 2 og 3 (jfr øving 2):

Trykk og hold nede venstre og høyre museknapp samtidig. Du er nå i "select"-modus. Velg guanin ved å dra musa over guanin, slik at alle dets atomer havner innenfor det rektangulære området som velges.

Edit – Copy

File – New Molecule

Edit – Paste

Du har nå lagret guanin som molekyl nr 2 i fila *gua_cyt* ("gua_cyt:M002"). Gå tilbake til baseparet (dvs molekyl nr 1) med "tilbake-knappen" i molekylspilleren nederst til venstre i SPARTAN-vinduet. Gjenta prosedyren i de fem punktene, denne gang for cytosin, som dermed blir molekyl nr 3 i *gua_cyt* ("gua_cyt:M003").

Sett i gang en Hartree – Fock – beregning for disse tre molekylene med Setup – Submit. SPARTAN velger automatisk Hartree – Fock med 3-21G(*) basissett. Dessuten velges såkalt "Global calculation" som standard. Det betyr at den valgte beregningsmetoden anvendes på alle de tre molekylene i *gua_cyt*, først baseparet (M001), deretter guanin (M002) og til slutt cytosin (M003). Lukk filen *gua_cyt* med File – Close. Sjekk at jobben har startet som den skal med Option – Monitor.

Gå så videre til de neste oppgavene mens de tre jobbene kjøres i bakgrunnen.

3. Velg *Nucleotide* byggemeny og lag en vilkårlig sekvens av baser (A=adenin, G=guanin, T=thymin, C=cytosin) for den ene spiralen i DNA-molekylet. Lag en sekvens med 20 baser. Velg *Helix B*. Venstreklikk i arbeidsfeltet. Reduser størrelsen (SHIFT+hm) på DNA-strukturen til du har hele strukturen i synsfeltet. Velg *Model-Ribbons* for å gjøre de to spiralene tydeligere. Velg også *Model-Hydrogen Bonds* for å markere hydrogenbindingene mellom de to spiralene.

Oppgaver:

a) Mål "perioden" langs dobbelspiralen. Det enkleste er antagelig å måle avstanden mellom to P-atomer. Hvor mange P-atomer er det pr periode i en gitt spiral?

b) Inspiser DNA-strukturen ved å rotere og skalere molekylet. Veksle gjerne mellom Model – Ball and Spoke og Model – Space Filling. Lokaliser både A – T og G – C basepar og mål avstandene O ... H og N ... H til hydrogenbindingene som holder to og to baser sammen.

Lokaliser karbohydratet 2-deoksyribose og fosfatgruppen. Anslå diameteren på DNA dobbelspiralen. Anslå bredden på "åpningene" inn til sentrum fra siden av DNA-molekylet. (Bruk Model – Space Filling.) Man regner med at både kreftfremmende og krefthindrende molekyler har mulighet til å påvirke DNA ved å smette inn her.

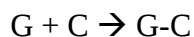
c) DNA-molekyler kan bli *lange*, opptil flere centimeter. Anslå massen til et DNA-molekyl som er 10 cm langt. Atomære masser for atomene som inngår finner du i et periodisk system. Anslå også volumet til et slik DNA-molekyl. Til sammenligning har en celle typisk et volum et sted mellom 10^3 og 10^6 kubikkmikrometer.

Lagre molekylet (i tilfelle du har lyst til å kikke på det flere ganger) som dna og lukk det.

4. Nå har du forhåpentlig brukt så lang tid på punkt 3 at de tre Hartree – Fock – jobbene er ferdige.

Oppgaver:

a) Bestem den beregnede bindingsenergien til G – C baseparet, med andre ord energigevinsten for "reaksjonen"



(Bruk Display – Properties og les av energien E (i atomære enheter, au), eller velg Display – Spreadsheet og legg inn energien (i enheten kcal/mol) med Add – E.)

Kommentar: I virkeligheten er nok bindingsenergien betydelig mindre enn dette, både på grunn av approksimasjonene som er knyttet til selve Hartree – Fock – metoden, og på grunn av at DNA stort sett befinner seg i en løsning (vann), og ikke i gassfase, som vi på sett og vis antar i en slik beregning. Ikke desto mindre: Du ser at bindingsenergien pr hydrogenbinding, av størrelsesorden "noen kcal/mol", er lav sammenlignet med "ordinære" kjemiske bindinger som ionebinding i HCl og kovalent binding i H₂. (Finn disse bindingsenergiene i litteraturen, hvis du da ikke allerede vet omtrent hvor store de er.)

b) Mål avstandene O – N, N – N og N – O (atomet i G nevnt først) for de tre beregnede hydrogenbindingene og sammenlign med de eksperimentelle verdiene, henholdsvis 2.91, 2.95 og 2.86 Å. (Rosenberg et al, *J. Mol. Biol.*, **104**, 109 (1976).) Kommentar: Eksperimentet tilsvarer ikke eksakt samme system, så avvik skyldes ikke nødvendigvis Hartree – Fock – metoden.

5. Alternative basepar?

Bygg en G-A sekvens slik at du får baseparene G – C og A – T etter hverandre. Undersøk (visuelt) om ombytte av C og T (og dermed baseparene G – T og A – C) vil gi lignende gunstige "hydrogenbindingsbetingelser" som i G – C og A – T baseparene.