

Proteinbretting

F. Knapskog^a, H. Storesund^a

^a*Institutt for fysikk og Institutt for matematiske fag, Norges Teknisk-Naturvitenskapelige Universitet, N-7491 Trondheim, Norway.*

Mål:

1. Visualisering av proteinet

I denne oppgaven skal et protein kunne gjennomføre lovlige rotasjoner og visualiseres i en matrise. Et protein er et polymer bestående av en kjede monomerer, retttere sagt aminosyrer. Mellom påfølgende monomerer er det sterke bindinger som hindrer polymeret i å splitte seg.

Måten proteinet blir visualisert på er å lage en $n \times n$ -matrise hvor alle elementer som ikke er en del av proteinet, er representert ved tallet 0. Proteinets selv er en tallrekke fra 1 til n . En rotasjon kan gå enten med eller mot klokka, om et hvilket som helst monomer. Deretter er det en rekke kontroller som må gjennomføres for å sørge for at en rotasjon er gyldig. For å unngå at proteinet krysser seg selv, sjekkes det at det er samme antall nuller i matrisen som før rotasjonen. For å unngå at proteinet beveger seg ut av matrisen, er midtproteinets fastlåst. Dette er implementert i koden ved at monomerene som endrer posisjon alltid er vekk fra midtmonomeret. Hvis det likevel skulle klare å snike seg ut av matrisens rekkevidde, har det blitt implementert en try except ved indexerror. Dette kan forekomme hvis lengden på proteinet er et partall, siden det ikke er et enkeltprotein i midten av kjeden.

Proteinets

<pre>[[0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [1 2 3 4 5 6 7 8 9 10] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0]]</pre>	<pre>[[0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [1 2 3 4 5 6 7 0 0 0] [0 0 0 0 0 0 8 0 0 0] [0 0 0 0 0 0 9 0 0 0] [0 0 0 0 0 0 10 0 0 0] [0 0 0 0 0 0 0 0 0 0]]</pre>	<pre>[[0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [1 2 3 4 5 6 0 0 0 0] [0 0 10 9 8 7 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0] [0 0 0 0 0 0 0 0 0 0]]</pre>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figur 1: I matrisen til venstre er proteinet opprettet i en 10×10 -matrise. Det består av 10 monomerer i en vannrett kjede med monomernummer 5 fiksert midt i matrisen. Proteinets brettes så med klokken om monomernummer 7 som vist i den midterste matrisen. Videre brettes den en gang til med klokken om monomernummer 6 i matrisen helt til høyre.

2. Energi som funksjon av temperatur

Overgangen videre innebærer å beregne proteinets potensielle energi. Et protein vil alltid prøve å redusere sin potensielle energi så mye som mulig. Ved helt utstrakt posisjon er proteinets potensielle energi satt til null, som er den høyeste mulige potensielle energien proteinet kan ha. Når det bretter seg, vil det oppstå svake bindinger mellom to ikkepåfølgende monomerer hvis de havner rett ved siden av hverandre. Dette medfører en reduksjon i potensiell energi, og derfor vil en slik brett skje naturlig. I det tredje proteinet i figur 1 er det for eksempel svake bindinger mellom monomer nr 5 og 8, 4 og 9, og 3 og 10. Hvis en brett ikke medfører redusert energi, vil den i utgangspunktet ikke gjennomføres.

Dersom den potensielle energien til proteinet er høyere ved den nye tilstanden etter en potensiell brett er det fortsatt en mulighet for at den gjennomføres grunnet termiske fluktuasjoner, altså mange uforutsigbare vibrasjoner i systemet. Her blir en brett ved hjelp av termiske fluktuasjoner gjennomført dersom et tilfeldig tall valgt av pseudotilfeldige algoritmer er større enn

$$e^{\frac{E' - E^0}{k_B T}} \quad (1)$$

der $E' - E^0$ er energidifferansen mellom før og etter bretten utføres. Den potensielle energien regnes ut ved å lage en energimatrise og en nabomatrise. Nabomatrisen setter element ij lik 1 hvis det oppstår en svak binding mellom monomer nr i og j , og null ellers. Energimatrisen er en matrise med tilfeldig genererte tall mellom $-10,4 \cdot 10^{-21}$ og $-3,47 \cdot 10^{-21}$. Den totale potensielle energien finnes da ved å multiplisere matrisene sammen og dele på 2 ettersom både element ij og ji vil bli satt til 1 i nabomatrisen.

2.1. Midlere energi

Den midlere energien, $\langle E \rangle$, beregnes ved hjelp av den kanoniske partisjonsfunksjonen

$$Z = \sum_{ms} e^{-\beta E_{ms}}, \quad (2)$$

der $\beta \equiv \frac{1}{k_B T}$. Den midlere energien blir da

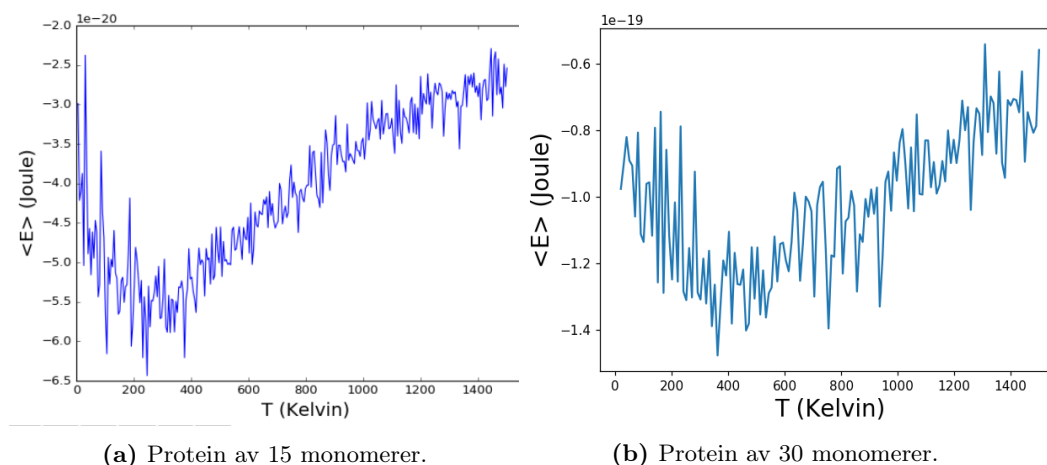
$$\langle E \rangle = \frac{1}{Z} \sum_{ms} E_{ms} e^{-\beta E_{ms}}. \quad (3)$$

Etter hvert forsøk på en gyldig brett, altså en brett som ikke bretter inn i seg selv eller ut av matrisen, vil den gjeldende potensielle energien legges til som energien til mikrotilstanden i summene i (2) og (3). Dette gjennomføres et visst antall ganger per temperatur. Ettersom proteiner ved lave temperaturer krever flere iterasjoner for å oppnå en grei nøyaktighet enn for proteiner ved høye temperaturer varierer dette antallet med temperaturen

$$d(t) = d_{max} e^{-sT}, \quad (4)$$

der d_{max} er satt til 10.000, og s er satt til 0,001. Dermed starter første proteinet ved 0 Kelvin på 10.000 iterasjoner mens ved 1500 Kelvin gjennomføres kun omtrent 2000 iterasjoner. For hver temperatur initialiseres et nytt protein og den midlere energien tilbakestilles til 0. Dette gjennomføres for hver 10. Kelvin fra 0 K til 1500 K og til slutt plottes den midlere energien som funksjon for temperatur. Proteinene har alle 15 monomerer hver.

Midlere energi som funksjon av temperatur



Figur 2: Plottet presenterer midlere energi til proteiner av 15 monomerer målt i Joule gjennomført et antall forsøk på brett bestemt av (4) per temperatur. Temperaturene er fra 0 Kelvin og hver tiende Kelvin frem til og med 1500 Kelvin.

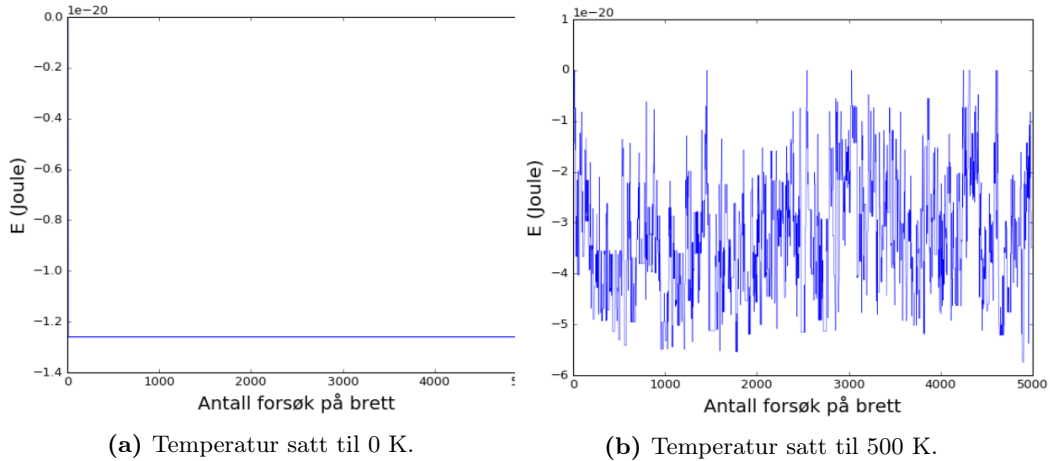
Ettersom proteinet som består av 30 monomerer i 2 er dobbelt så langt blir den numeriske kompleksiteten mye større. Det er dobbelt så mange monomerer proteinet kan brette seg om, mange flere muligheter for svake bindinger og generelt blir alle matriseoperasjonene 4 ganger så store.

2.2. Energi som funksjon av antall brett

Istedenfor å utføre alle iterasjonene for en temperatur og kun plote den midlere energien til slutt skal nå den potensielle energien til proteinene plottes for hver iterasjon. Dette gjennomføres kun for temperaturene 0 Kelvin og 500 Kelvin. Ettersom det fysisk er umulig å sette temperaturen til 0 Kelvin er den her istedenfor satt til 0,001 K da det unngås å dele på null i (1). Proteinene har fortsatt 15 monomerer her.

Figur 3 viser at ved høy temperatur fluktuerte proteinet veldig hyppig, mens ved 0 Kelvin er proteinet fullstendig fastlåst i et lokalt minimumspunkt. Det vil si at uansett hvilken brett som blir forsøkt på vil den potensielle energien til proteinet bli større. I begge tilfeller holder energien seg innenfor et bestemt intervall etter et visst antall forsøk på brettinger. Ettersom det kommer frem hvor stort dette intervallet er på plottene konkluderes det med at det er tilstrekkelig antall forsøk på brettinger.

Potensiell energi som funksjon av antall forsøk på brett



Figur 3: Plottene presenterer den potensielle energien i Joule til hvert sitt protein av 15 monomerer plottet for antall forsøk på gyldige brett utført. Plottet til venstre er utført ved 0 K og plottet til høyre er utført ved 500 K.

2.3. $Temperaturer \geq 300 K$

Ved temperaturer $\geq 300 K$ stiger den midlere energien med temperaturen som vist i figur 2 med en ganske konstant stigning. Den avtar litt mot slutten ettersom den aldri kan overstige 0 J. Proporsjonaliteten mellom midlere energi og temperatur henger sammen med fysikken ettersom temperatur er midlere energi per enhet stoffmengde.

2.4. $Temperaturer < 300 K$

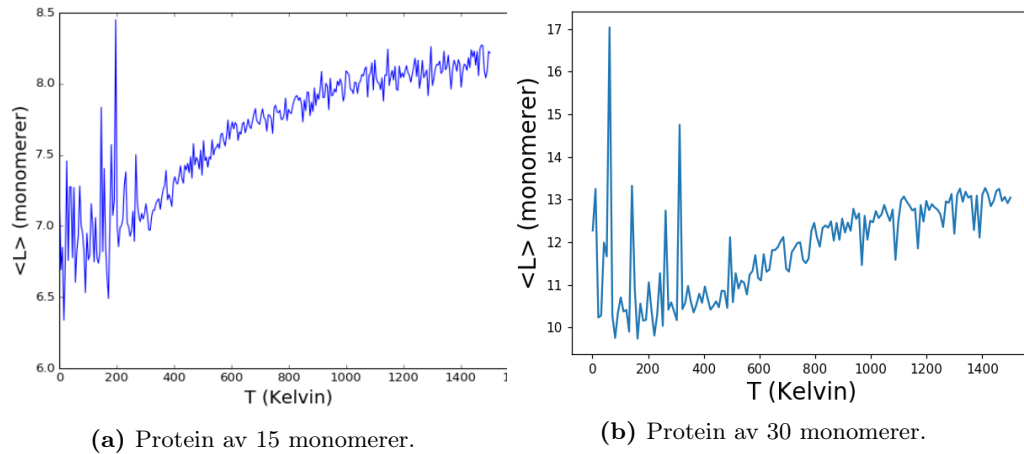
Ved temperaturer $< 300 K$ synker den midlere energien ved stigende temperatur. Det er fordi at dersom temperaturen er for lav vil proteinet låse seg i den første tilstanden den finner. Øker temperaturen litt kan proteinet fluktuere mer og finne seg en tilstand med enda lavere potensiell energi. Slik fortsetter det frem til omtrent 300 K der fluktuasjonen er akkurat stor nok til at proteinet får gravd seg ned i energibrønnen uten at det fluktuerer opp igjen for ofte.

3. Gjennomsnittsdiameter som funksjon av temperatur

Diameteren finnes ved å finne den lengste avstanden mellom 2 monomerer i protein-kjeden og har enheten monomerlengder eller monomerer. For å lage nabomatrisen finnes indeksene til to og to monomerer for å sjekke om de er ved siden av hverandre. Dermed kan avstanden mellom dem beregnes i samme slengen og la den lengste stå igjen. Til slutt beregnes da gjennomsnittet av alle diameterne. Dette itereres et vist antall ganger per temperatur akkurat som i seksjon 2.1. og 2.4., for både 15 monomerer og 30 monomerer.

Figur 4 viser samme trend for temperaturer større enn 300 K som figur 2. Den midlere diameteren stiger jevnt, som henger sammen med at jo større plass proteinet tar, dess færre muligheter for svake bindinger har det. Ved mindre enn 300 K er usikkerheten til

Gjennomsnittsdiameter som funksjon av temperatur



Figur 4: Plottene presenterer gjennomsnittsdiameterne til proteiner iterert et visst antall ganger etter (4), for hver 10. Kelvin fra 0 K til 1500 K. Til venstre er det plottet for 15 monomerer, mens til høyre er det plottet for 30 monomerer.

den midlere diameteren veldig stor. Det kommer av at det utføres mye færre fluktuasjoner som fører til at variasjonen blant diameterne blir liten. Store utslag jevnes dermed ikke ut av alle de andre normale utslagene og står igjen som en stolpe.

4. Gradvis nedkjøling av et protein

4.1. Energi ved nedkjøling

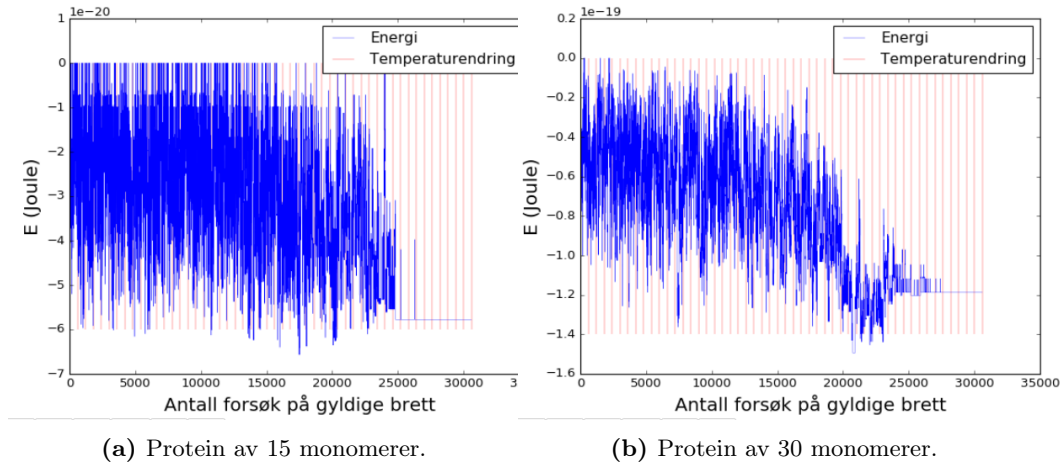
I denne seksjonen kombineres 2.1. og 2.2. ved at den potensielle energien plottes for hvert forsøk på å brette, og etter hver 600. brett synker temperaturen med 30 K. Temperaturen starter på 1500 og slutter på 0 K. Temperaturendringene er markert med vertikale linjer. Proteinene som skal kjøles ned er det samme gjennom hele prosessen og består av 15 monomerer.

Figur 5 viser at begge proteinene fluktuerer like vilt uavhengig av temperaturen den første halvdel. Deretter synker både fluktuasjonen og den potensielle energien. Den potensielle energien til proteinet med 30 monomerer er generelt halvparten så stor som den potensielle energien til proteinet med 15 monomerer. Til slutt ender begge proteinene opp i et lokalt minimum.

4.2. Midlere energi ved nedkjøling

Her kjøles et protein ned akkurat som i seksjon 4.1. med 600 forsøk på brettinger per temperatur med temperaturintervall på 30 K mellom hver i fra 1500 K ned til 0 K. Istedenfor å plote den potensielle energien i hvert steg beregnes nå den midlere energien til én temperatur og plottes før den midlere energien nullstilles og beregnes igjen for neste temperatur. Det er det samme proteinet gjennom hele nedkjølingen og består fortsatt av 15 monomerer.

Potensiell energi som funksjon av antall brett under nedkjøling



Figur 5: Plottet presenterer den potensielle energien for et protein etter antall forsøk på utførelse av gyldige brett. For hver 600. forsøk senkes temperaturen med 30 K. Proteinet starter på 1500 K og ender opp på 0 K. Det er det samme proteinet som kjøles ned i hele prosessen for hver av plottene. Til venstre er proteinet 15 monomerer langt og til høyre er det 30 monomerer langt. Hvert temperaturskifte markeres av en rød vertikal linje.

Figur 6 viser at ved gradvis nedkjøling oppnås et lokalt minimumspunkt mye raskere og ved enda lavere energier enn dersom proteinet starter utstrakt ved en lav temperatur.

4.3. Middeldiameter ved nedkjøling

Akkurat som i seksjon 4.2. gjennomføres her, med unntak av for hver 30. temperatur plottes middeldiameteren istedenfor den midlere energien.

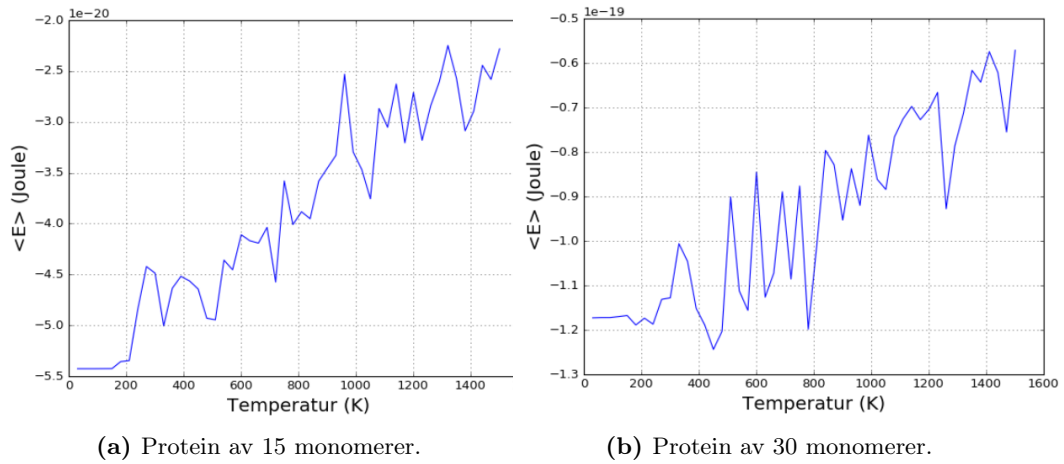
Som i seksjon 3 viser figur 7 at middeldiameteren har en sammenheng med temperaturen, med unntak av at det er det samme proteinet gjennom hele prosessen istedenfor å starte med et nytt utstrakt protein for hver temperatur. Dermed vil proteinet fortsette der det slapp og ikke havne i lokale minimum som er mye høyere enn der det var ved forrige temperatur. De store utslagene fjernes og proteinet låses i et lokale minimumspunkt.

4.4. Visuelle proteiner ved nedkjøling

Etter at nedkjølingen er ferdig havner proteinene i en fastlåst tilstand der den potensielle energien har havnet i et lokalt minimumspunkt. Figur 8 viser tre proteiner i en slik tilstand.

Som figur 8 viser er proteinene i en veldig sammenkveilet tilstand med mange svake bindinger. Dermed blir det spesielt vanskelig for proteinet å fluktuere ut av tilstanden og inn i en ny ved lave temperaturer. Proteinene låses da fast i det nærmeste lokale minimumspunktet for den potensielle energien og forblir der.

Midlere energi som funksjon av temperatur under nedkjøling

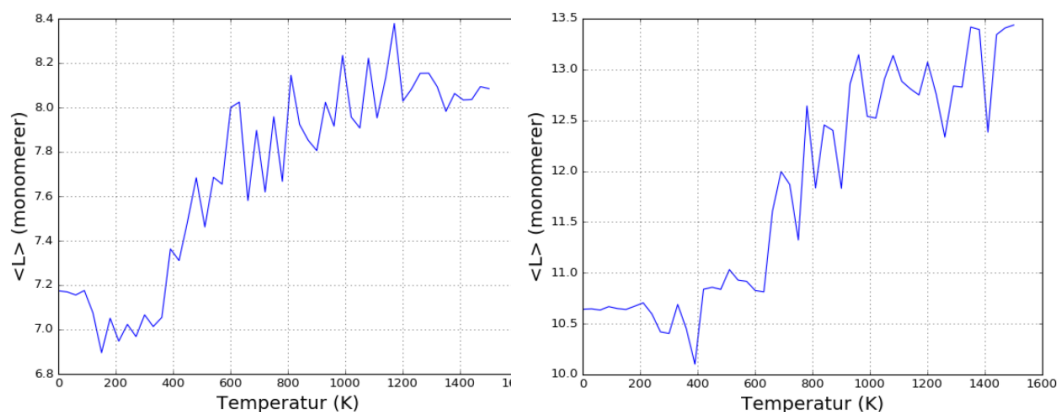


Figur 6: Plottet presenterer to proteiner som kjøles ned fra 1500 K til 0 Kelvin med 30 K per intervall. 600 forsøk på gyldige brett utføres per intervall og den midlere energien til proteinet på det gjeldende intervallet plottes for temperaturen. Det er det samme proteinet som kjøles ned i hele prosessen for hver av plottene. Til venstre er proteinet 15 monomerer langt og til høyre er det 30 monomerer langt.

5. Grundigere nedkjøling av et protein

Til slutt er det i figur 9 lagt ved den midlere energien til ti proteiner som er utsatt for en grundigere nedkjøling. Det vil si antall forsøk på brett per temperatur er i henhold til (4) og temperaturen senkes kun med 5 K av gangen.

Middeldiameter som funksjon av temperatur under nedkjøling

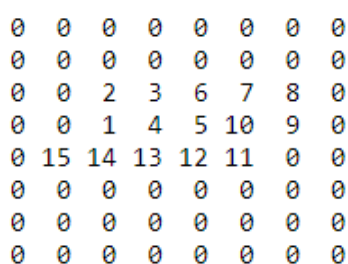


(a) Protein av 15 monomerer.

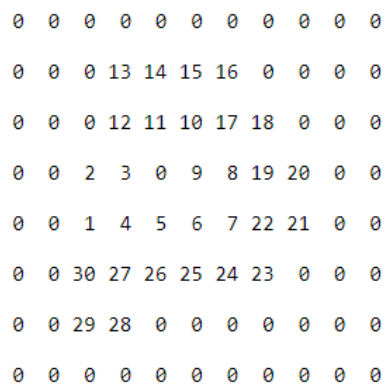
(b) Protein av 30 monomerer.

Figur 7: Plottet presenterer to proteiner som kjøles ned fra 1500 K til 0 Kelvin med 30 K per intervall. 600 forsøk på gyldige brett utføres per intervall og middeldiameteren til proteinet på det gjeldende intervallet plottes for temperaturen. Det er det samme proteinet som kjøles ned i hele prosessen for hver av plottene. Til venstre er proteinet 15 monomerer langt og til høyre er det 30 monomerer langt.

Visuell fremstilling av protein etter nedkjøling



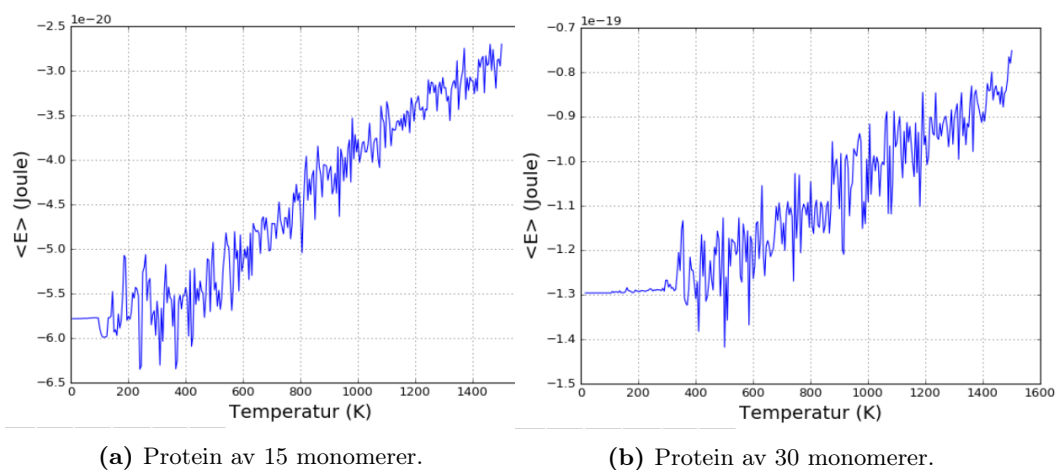
(a) Protein av 15 monomerer.



(b) Protein av 30 monomerer.

Figur 8: Figurene presenterer to proteiner som er nedkjølt i fra 1500 K til 0 K. Temperaturen senkes med 30 K av gangen og det utføres 600 forsøk på gyldige brett hver gang temperaturen senkes. Det er den siste tilstanden til proteinene som vises. Proteinet til venstre er 15 monomerer langt og det til høyre er 30 monomerer langt.

Midlere energi som funksjon av temperatur under nedkjøling



(a) Protein av 15 monomerer.

(b) Protein av 30 monomerer.

Figur 9: Plottet presenterer et protein som kjøles ned fra 1500 K til 0 Kelvin med 5 K per intervall. Antall forsøk på gyldige brett per intervall utføres i henhold til (4) og den midlere energien til proteinet på det gjeldende intervallet plottes for temperaturen. Det er det samme proteinet som kjøles ned i hele prosessen for hver av plottene. Til venstre er proteinet 15 monomerer langt og til høyre er det 30 monomerer langt.