1 Johdanto

Työssäni esittelen proteiinin kolmiulotteista rakennetta ja sen erilaisia mallinnustapoja.

**Kun kyseessä on kumminkin tietojenkäsittelytieteen kandityö, kannattaisiko tässä olla rivi kaksi lisää tästä mallinnuksesta, että päästään ns. TKT aiheeseen käsiksi. Nyt alkaa niin, että luulisi olevan kemian työ.**

Proteiini on kemialliselta rakenteeltaan aminohappoketju, joka on laskostunut tietyllä tavalla. Kannattaako sanoa tietyllä tavalla, voisiko sanoa pari sanaa tästä tavasta? Proteiinin tarkka rakenne esitettiin ensimmäisen kerran vuonna 1902. Solu valmistaa aminohapoista proteiineja proteiinisynteesissä. Proteiinin rakenne luokitellaan sen vaiheesta riippuen neljään eri luokkaan. Kannattaako luetella ne neljä luokkaa ja sitten mennä niiden esittelyyn? Primäärirakenne tarkoittaa suoraa aminohappoketjua, sekundäärirakenne kiertynyttä aminohappoketjua. Tertiäärirakenne kuvaa yhden aminohappoketjun tietynlaista kolmiulotteista rakennetta laskostumisen seurauksena. Kvaternäärirakenteesta on kyse, kun useampi aminohappoketju sitoutuu toisiinsa toiminnallisen proteiinin muodostamiseksi.

Proteiinisynteesin (mitä on proteiinisynteesi?) jälkeen proteiini hakee luonnollisen muotonsa laskostumalla natiiviin konformaatioon (mitä tämä tarkoittaa?). Tämä tapahtuu millisekunneissa tai korkeintaan sekunneissa laskettavassa ajassa. Laskostuminen tapahtuu siis huomattavasti konformaatioavaruuden (mikä tämä on?) läpikäyntiä nopeammin, mistä voisi olettaa, että proteiinin laskostuminen ei ole NP-täydellinen (mikä on NP?) ongelma. Laskostumiseen vaikuttaa kuitenkin moni muukin asia kuin laskenta-avaruuden koko, esimerkiksi nimellä kaperonit kutsuttavat proteiinit saattavat toimia muiden proteiinien apuna laskostumisessa esimerkiksi estämällä proteiinien saostumista. Tarvitseeko kummatkin esimerkiksi sanat?

Harkitse sitä, että nämä kappaleet tulisivat ennen kun mennään proteiinien rakenteen yksityiskohtiin. Tai osa niistä.

””

Proteiinin rakenteen ennustaminen on tärkeä laskennallisen biologian sovellutus, sillä proteiinin rakenteen suora tutkinta NMR:llä (Nuclear Magnetic Resonance, ydinmagneettinen resonanssi) tai röntgenkristallograﬁalla on hidasta ja kallista. Ennustamisessa on useita eri tapoja kuten neliöllisten keskiarvojen poikkeama (RMSD), proteiinitietokannan hyödyntäminen (PDB), vertaileva mallinnus, laskosten tunnistaminen ja joukkoistaminen. **Tosi hyvä kappale.**

Proteiinien rakenteen ennustamiseen tarvitaan useita erilaisia laskennallisia menetelmiä, sillä esimerkiksi osa ei toimi jos proteiinien rakenne on liian samanlainen. Toiset taas perustuvat juuri samankaltaisuuteen, sillä evoluution kautta tiedetään (miten evoluutio on opettanut?), että proteiinin kolmiulotteinen rakenne ei ole aina muuttunut, vaikka osa aminohapoista olisi korvautunut toisilla. Keskeisenä työvälineenä on proteiinin rakenteiden tietokanta, jossa on pääasiassa kokeellisilla menetelmillä selvitettyjä rakenteita, tällä hetkellä noin 100 000 proteiinista. Mallinnettavia proteiineja voidaan verrata tietokannan proteiineihin.

Joukkoistamista proteiinin rakenteen ennustamiseksi käytetään esimerkiksi Folditpelissä, jossa pelaajat yrittävät saavuttaa mahdollisimman lähelle natiivikonformaatiota pääsevän rakenteen. Pelaajien tuloksia vertaillaan Rosettan struktuuriin (mikä on Rosettan struktuuri?), joka yrittää ratkaista samat ongelmat kuin pelaajille annetaan. Monissa tapauksissa pelaajat pääsevät Rosettaa parempiin tuloksiin. Parhaan tuloksen antaa ihmisen visuaalisen hahmottamiskyvyn ja strategiataitojen yhdistäminen tietokoneen laskennalliseen kapasiteettiin.

””

Tämä sopii hyvin lopetukseksi.

Proteiinin rakenteen tunteminen on tärkeää, sillä virheellinen proteiinin toiminta tai laskostuminen voi aiheuttaa sairauksia, kuten Alzheimerin tautia. Jotta suoraan proteiineihin vaikuttavia lääkkeitä voitaisiin kehittää, täytyy proteiinin tarkka rakenne tuntea. Ennustamiseen ei ole ainakaan vielä löytynyt aminohappoketjun rakenteeseen liittyvää säännönmukaisuutta esimerkiksi hydrofobisuuteen tai varaukseen liittyen. Tertiäärirakenteen ennustaminen primäärirakenteen perusteella ei siis vielä onnistu kovinkaan hyvin.