

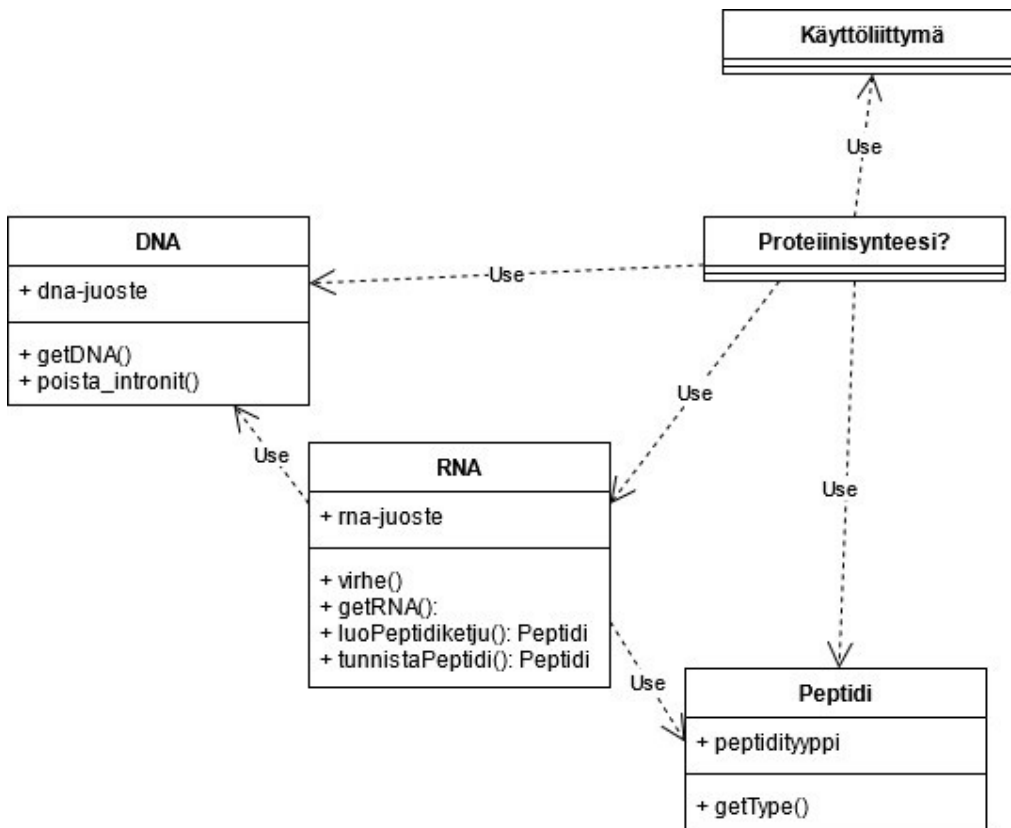
## Jaana Kivistö, 829964, Bioinformaatioteknologia, vuosikurssi: 2020, 26.2.2021

**Proteiinisynteesi:** tarkoituksena tehdä ohjelma, joka mallintaa proteiinisynteesin päätaphtumia. Ohjelma lukee tiedostosta DNA-sekvenssin, lukee toisesta DNA-juovasta RNA:n ja muodostaa peptidiketjun RNA:n emästen mukaan. Lähtökohtaisesti lähdän tavoittelemaan Vaikea-vaikeustasoa eli graafisen käyttöliittymän lisäksi huomioin silmukoinnin ja virhemahdollisuuden RNA:n lukemisessa.

**Käyttötapauskuvaus:** Ohjelma pyytää käyttäjältä tekstitiedoston, muuttaa sen peptidiketjuksi ja näyttää vaiheet ohjelmaruudulla. Esimerkiksi opettaja tarvitsee materiaaliinsa kuvan proteiinisynteesistä. Hän tekee siihen sopivan tiedoston ja syöttää sen ohjelmalle. Ohjelma luo tiedostosta kuvan, jonka opettaja voi liittää materiaaliinsa. Ohjelma lukee ensiksi tiedostosta DNA-sekvenssin talteen erotellen (intronit ja eksonit), lukee siitä RNA-juosteen talteen, josta luetaan peptidit. Lopuksi ohjelma piirtää kaikista kuvan.

**Rakennesuunnitelma:** Proteiinisynteesi käynnistää käyttöliittymän, jonka kautta käyttäjältä pyydetään tiedostopolkua. DNA luokka luodaan antamalla sille parametrina tiedostopolku, jonka se lukee ja jonka perusteella se tallentaa rakenteensa.

RNA luodaan antamalla sille parametrina DNA jonka perusteella se alustaa oman rakenteensa. RNA pystyy luomaan peptidiketjun oman rakenteensa perusteella ja palauttaa sen taulukkona, joka koostuu Peptidi-olioista



**Tietorakenteet:** Olen alustavasti suunnitellut DNA:n ja RNA:n olevan listoissa tai stringeissä, jolloin niitä on helpohko käsitellä pythonin olemassa olevilla metodeilla. Peptidien tiedot tallennan taulukkomuotoon. Yksittäinen peptidi on kolmen emäksen string-rakenne, nimi myös string.

**Tiedostot ja tiedostoformaatit:** Ohjelma lukee tekstitiedostoja, esim. Random DNA sequence generaattorilla luotuja tekstitiedostoja, joissa DNA-juosteen emäket on jaettu useammalle riville. Juosteen pituus on noin 500-1000 emästä (emäsparia). Tiedostossa on vain DNA:n toisen juosteen emäket.

**Algoritmit:** Mielestäni projektissa ei ole varsinaisia loogisia ongelmia. DNA-sekvanssia käydään loopeissa läpi, RNA:ta käydään läpi kolme emästä kerrallaan. Intronien poistaminen tapahtuu etsimällä välit, jotka alkavat emäksillä GU- ja loppuvat AG:hen, ja poistamalla ne.

**Testaussuunnitelma:** Yksikkötestaukset ydinmetodeille, esimerkiksi peptidiketjun luominen oikein. Testaan myös käyttöliittymän manuaalisesti.

**Kirjastot:** PyQt, mutta muista en tiedä, koska kyseiset asiat tulevat vasta kurssin viimeisellä kierroksella. (Onko tähän ohjaajalla vinkkejä?)

**Aikataulu:** Aloitan työn tekemisen käyttöliittymien opiskelun jälkeen. Aloitan luokkien toteuttamisesta, johon arvioisin aikaa kuluvan n. 10h. Tämän jälkeen teen luokille yksikkötestit. Tähän menee n. 4h. Viimeisenä toteutan käyttöliittymän, jonka tekemiseen menee n. 20h, mutta arvio tarkentuu kun olen opiskellut käyttöliittymistä enemmän.

**Kirjallisuusviitteet ja linkit:**

esim. Molecular Cell Biology <https://www.ncbi.nlm.nih.gov/books/NBK21603/>

ja <https://docs.python.org/3/>