

TARTU ÜLIKOOL
Arvutiteaduse instituut
Infotehnoloogia mitteinformaatikutele õppekava

Katrin Ruisu
Qumanize rakenduse analüüs ja sellele
veebiliidese loomine
Magistritöö (15 EAP)

Juhendajad: Ahto Salumets, MSc
Priit Adler, PhD

Tartu 2020

Qumanize rakenduse analüüs ja sellele veebiliidese loomine

Lühikokkuvõte:

Käesolevas magistritöös kirjeldatakse rakendust nimega Qumanize, mis loodi PERMIDES projekti raames mitte-inimese monoklonaalsete antikehade humaniseerimisprotsessi läbiviimiseks. Monoklonaalseid antikehi kasutatakse mitmesuguste vähkkasvajate ja autoimmuunhaiguste ravis. Nende eeliseks teiste ravimite ees on väga kõrge spetsiifilisus ja vähem kõrvaltoimeid. Kuna antikehade avastamine toimub katseloomade immuunsüsteemis, võib nende aminohapete koostis kutsuda inimkehas esile immuunvastuse, millel on rasked tagajärjed. Qumanize on vahend erinevates loomaliikides avastatud antikehade aminohappelise järjestuste humaniseerimiseks. Qumanize loodi käsuarakendusena, kasutamise lihtsustamiseks loodi magistritöö praktilise osana programmile veebipõhine kasutajaliides.

Võtmesõnad:

monoklonaalsed antikehad, humaniseerimine, Flask, Python, bioinformaatika

CERCS: P175

Analysis and web interface generation for the Qumanize application

Abstract:

This master thesis describes an application called Qumanize that was created to carry out a humanization process for non-human monoclonal antibodies. Monoclonal antibodies are used as therapeutics for various cancers and autoimmune diseases. They are credited for their high specificity and reduced side effects compared to other drugs. However, since they are discovered in non-human immune systems, their amino-acid composition may induce an immune response in the human body with severe consequences. Qumanize is a tool for humanizing antibody amino acid sequences discovered in various species. It was created as a command line application, so one part of this thesis project was to create a web interface to this application for better accessibility - this process is also described in detail.

Keywords:

monoclonal antibodies, humanization, Flask, Python, bioinformatics

CERCS: P175

Sisukord

| | |
|--|----|
| 1. Sissejuhatus | 4 |
| 2. Mõisted ja terminid | 5 |
| 3. Kirjanduse ülevaade | 6 |
| 3.1 Antikehade struktuur ja immuunvastus | 6 |
| 3.2 Monoklonaalsed antikehad ravimitena | 7 |
| 3.3 Antikehade humaniseerimine | 8 |
| 3.3.1 CDR-ide identifitseerimine | 9 |
| 3.3.2 Inimese antikeha raamistik | 11 |
| 3.3.3 Humaansuse hindamine | 12 |
| 4. PERMIDES projekt | 13 |
| 4.1 PERMIDES projekti ülevaade | 13 |
| 4.2 Qumanize rakendus | 13 |
| 4.3 PERMIDES projekti tulemused | 14 |
| 5. Kasutajaliidese loomine | 15 |
| 5.1 Qumanize rakenduse tööpõhimõtted | 16 |
| 5.1.1 Qumanize rakenduse jaoks vajalikud Pythoni paketid | 16 |
| 5.1.2 Qumanize rakenduse struktuur | 17 |
| 5.2 Flaski veebiraamistik | 18 |
| 5.2.1 Flask veebiraamistik Qumanize kasutajaliidese loomisel | 19 |
| 5.2.2 Bootstrapi raamistiku kasutamine kujunduses | 21 |
| 5.2.3 Qumanize rakenduse kasutajaliides | 21 |
| 6. Kokkuvõte | 25 |
| 7. Kasutatud kirjandus | 26 |
| Tänuõnad | 32 |
| LISA 1: LIHTLITSENTS MAGISTRITÖÖLE | 33 |
| LISA2: QUMANIZE KÄSUREARAKENDUSE KASUTUSJUHEND | 34 |

1. Sissejuhatus

Ravimite puhul öeldakse tihti, et kui on toime, on ka kõrvaltoime. Siiski üritatakse ravimitööstuses leida preparaate, mille kõrvaltoimed oleksid võimalikult leebed ja ravimi poolt tekitatav kahju ei varjutaks saadavat kasu. Üks viis võimalike kõrvaltoimete vähendamiseks on võimalikult spetsiifiliste ravimite väljatöötamine, mis avaldaksid toimet ainult haiguskolletes ja ei mõjutaks muid keha kudesid või rakke. Monoklonaalsed antikehad (mAB) on ühed sellised terapeutikumid, mida saab arendada väga spetsiifiliste sihtmärkide vastu ning seetõttu vähendada tõsiseid kõrvaltoimeid. mAB-id on leidnud rakendust näiteks erinevate vähkkasvajate ravis ning on patsiendile oluliselt lihtsamad taluda kui kemoteraapia.

Siiski on mAB-idel samuti oht põhjustada patsiendile vaevusi, mille üheks põhjuseks on, et mAB-ide avastamine toimub tavaliselt katseloomas ning seega antikehade aminohappeline järjestus on omane kasutatud loomale ja võib inimesel tekitada immuunvastust (anti-antikeha vastust). Tänapäeva biotehnoloogia võimaldab valke modifitseerida ning rekombinantset (nö katseklaasis) suures mahus toota ja puhastada, seega on võimalik loomas tekitatud mABi aminohappelist järjestust muuta inimesele omasemaks ehk humaniseerida. Antikehade humaniseerimise abil on eesmärgiks toota molekule, mis oleks inimestele minimaalse immunogeensusega e tekitaks võimalikult nõrga immuunvastuse kuid säilitaks samas algse mitte-inimese antikeha spetsiifilisuse ja afiinsuse.

Eesti biotehnoloogiafirma Icosagen Cell Factory (ICF) pakub antikehade avastamise teenust ravimifirmadele ja teadusasutustele ning arendab ka ise uusi terapeutilisi molekule. ICF-is on kasutusel patenteeritud Hybrifree tehnoloogia, kus kasutatakse antikehade avastamiseks katseloomadena kana ja küülikut. Valdav enamus mAB-e on leitud hiires, seega kommertsiaalsed humaniseerimisplatvormid on arendatud ja testitud hiire mudelil. Selleks, et osana antikehade arenduse teenusest pakkuda klientidele ka kanas ja küülikus avastatud antikehade humaniseerimist loodi ICF-is koostöös tervisevaldkonna IT-lahendusi pakkuva ettevõttega Quretec humaniseerimistarkvara Qumanize.

Käesoleva magistritöö keskendub Qumanize rakenduse kirjeldamisele (peatükk 4 ja peatükk 5) avades eelnevalt antikehade avastamise ja humaniseerimise tagamaid (peatükk 3). Antud magistritöö autor töötab ICF-is, kuid rakenduse loomise protsessis ei osalenud, küll aga osales ICF-i poolses rakenduse juurutamise protsessis ja koostas rakendusele kasutusjuhendi. Qumanize rakendus oli algselt loodud töötama käsurealt - programmi lihtsamaks kasutamiseks loodi magistritöö raames rakendusele veebipõhine liides kasutades Flask raamistikku. Seda protsessi on detailselt tutvustatud peatükis 5.

2. Mõisted ja terminid

CDR (*complementarity determining region*) – antikeha afiinsuse eest vastutavad piirkonnad raske ja kerge ahela variaabelregioonidel

Dekoraator – funktsioon, mis võtab argumendiks teise funktsiooni ning väljastab modifitseeritud funktsiooni

Humaniseeritud antikeha – mitte-inimese immuunsüsteemist pärinevad antikehad, mille valgujärjestusi on modifitseeritud nii, et suurendada sarnasust inimese antikeha iduliini järjestustega

IMGT (*international ImMunoGeneTics information system*) – Immuunogeneetika andmebaas, mis pakub avalikku juurdepääsu kvaliteetselt annoteeritud andmetele immuunoglobiinide, koosobivuskomplekside ja T-raku retseptorite genoomi, proteoomi, geneetika ja struktuuri kohta

mAB (*monoclonal antibody*) – monoklonaalne antikeha e antikeha, mis on valmistatud identsete immuunrakkude poolt, mis kõik on unikaalse lähteraku kloonid

Antikehade nummerdamisskeem – antikeha variaabelregioonis olevatele aminohappejääkide positsioonidele numbriliste väärtuste andmine, et erinevaid variabele oleks võimalik omavahel võrrelda

PFM (*position frequency matrix*) – positsioonisagedusmaatriks e maatriks, mis saadakse järjestuste joondamisel ning mis näitab erinevate nukleotiidide või aminohapete sagedusi järjestuse iga positsiooni kohta.

Renderdamine (*rendering*) – käesoleva töö kontekstis HTML koodist visuaalse pildi loomine

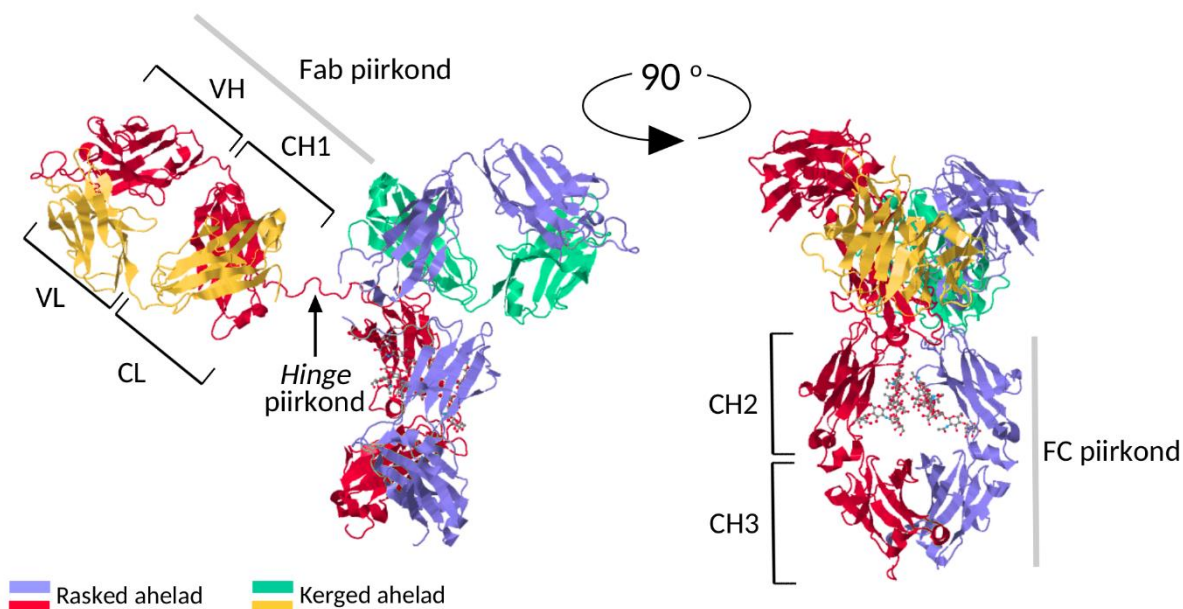
URL (*Uniform Resource Locator*) - Internetis igale dokumendile või muule ressursile vastav unikaalne internetiaadress.

WSGI (*Web Server Gateway Interface*) – veebiserveri võrguvärava kasutajaliides, mis kirjeldab, kuidas veebiserver suhtleb veebirakendusega ja kuidas veebirakendusi saab omavahel päringu protsessimiseks ühendada

3. Kirjanduse ülevaade

3.1 Antikehade struktuur ja immuunvastus

Antikehad ehk immuunoglobuliinid (Ig) on immuunsussüsteemi valgud, mis kaitsevad keha erinevate patogeenide eest. Antikehasid esineb viie erineva klassina ehk isotüübina, mis erinevad raske ahela struktuuri ja funktsiooni spetsiifika poolest: IgG, IgH, IgA, IgM, IgD, IgE. Nendest IgG on kõige levinum moodustades 10-20% vereplasma valkudest. Kõikidel Ig-del on sama põhistruktuur koosnedes kahest kergest ja kahest raskest ahelast. Need omakorda jagunevad konstantseteks ja varieeruvateks piirkondadeks (variaabel-regioon). Konstantsed piirkonnad on oma Ig alatüübi sees üldiselt samad ning varieeruvad piirkonnad sõltuvad konkreetsest antigeenist, mille külge antikeha seondub. Varieeruvad piirkonnad (VH – *variable heavy* ja VL – *variable light*) koos kerge ahela konstandi (CL – *constant light*) ning raske ahela konstandiga CH1 (*Constant Heavy 1*) moodustavad antigeeni siduva osa (Fab – *Fragment antigen binding*) ning raske ahela piirkonnad CH2 ja CH3 moodustavad FC (*Fragment Christallisable*) piirkonna, mis seondub teiste immuunsüsteemi komponentidega nagu fagotsüüdid või komplemendi raja valgud, mis aitavad antigeeni neutraliseerida. Fab ja FC piirkonda ühendab nn *Hinge* piirkond (Joonis 1). [1]



Joonis 1 IgG 3D struktuur [2] kahe erineva nurga alt (Vidarsson et al., 2014 ainetel , struktuur PDB: 1HZH)[3].

Täpsemalt määratlevad antikeha afiinsuse kolm spetsiifilist piirkonda raske ja kerge ahela variaabelregioonis, mida kutsutakse komplementaarsust määravateks regioonideks (CDR e *complementarity determining region*). CDRid vahelduvad nelja raamistikuregiooniga (FR, *framework region*), mille järjestused on antikeha perekonna sees suhteliselt püsivad ja mis ei

ole otseselt seotud antikeha afiinsusega. Raamistikuregioonid paiknevad enamasti üksteise kõrval ning nende otstes paiknevad CDR-id moodustavad silmuseid, see ülesehitus annab variaabelregioonile käesõrmede kuju, kus sõrme otsas on afiinne CDR piirkond. Raske ahela CDR-id koos vastavalt seotud kerge ahela CDR-idega moodustavadki antikehamolekulile antigeeni sidumise saidi. [4]

Antikehi toodetakse plasma B rakkude poolt, mis on tüüp lümfotsüüte mida toodetakse pidevalt luuüdis. Kui B-raku pinnal olev retseptor (BCR – *B-Cell Receptor*) seondub mõnele organismis ringlevale võõrale antigeenile, siis tekib immuunvastus: B-rakk aktiveerub ning diferentseerub plasmarakukus, mis hakkab eritama antikehasid. Seejärel B-rakk jaguneb ja tekivad B-mälurakud, mis on võimelised tulevikus seda sama antigeeni ära tundma moodustades immunoloogilise mälu: sama antigeeni sattumisel organismi hakkavad B-mälurakud jagunema ning diferentseeruvad antikehi tootvateks plasmarakkudeks. Kuna teine aktivatsioon toimub läbi vähem rangema regulatsioonimehhanismi, tekib immuunvastus kiiremini ning antikehad on võimelised haigustekitaja kiiresti neutraliseerima. [1,5]

Selleks, et antikehad ei moodustuks organismi enda antigeenide vastu testitakse ebaküpseid B-rakke enne luuüdist väljumist. Ebaküpsele B-rakule eksponeeritakse organismi enda antigeene ning kui mõni BCR sellele liiga tugevalt seondub, siis vastav B-rakk kas hävitatakse või modifitseeritakse BCR ümber. Seda protsessi nimetatakse negatiivseks selektsiooniks. [1]

3.2 Monoklonaalsed antikehad ravimitena

Monoklonaalsed antikehad (mAB) on võimelised tugevalt ja spetsiifiliselt seonduma oma sihtmärgile (antigeenile) ja seeläbi omavad olulist rolli haiguste diagnostikas ja ravis [6]. Kuna mAB-e saab põhimõtteliselt toota iga antigeeni vastu, on võimalused nende terapeutiliseks kasutamiseks laialdased. Peamiselt on terapeutilisi antikehi rakendatud vähi ja autoimmuunhaiguste puhul, mis on seotud talitlushäirega inimeste enda rakkudes [7]. Läbi spetsiifilise seondumise rakupinna retseptoritele on mABide abil võimalik käivitada rakusurma (apoptoosi) signaaliradasid või blokeerida/muuta raku ainevahetust. mABidele on võimalik külge pookida (konjugeerida) ka erinevaid ballastühendeid nagu toksiine, kemoterapeutikume, radioaktiivseid isotüüpe jne [8]. Seega kui tuumorakk on keha teistest rakkudest eristatav mingi rakupinna retseptori läbi, siis selle vastu saab leida mABi, mis on võimeline toksilise ühendi toimetama otse vähikoldesse vähendades oluliselt ravi kõrvaltoimeid [8,9].

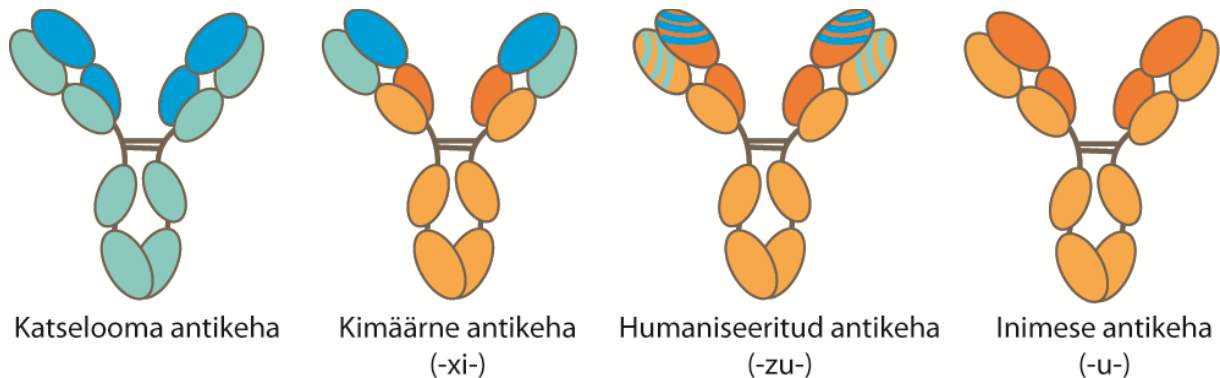
Kuigi tänapäeval on kasutusele võetud ka teisi meetodikaid, on traditsiooniline ja siiani laialdaselt levinud uute mABide avastamine läbi loomade (peamiselt näriliste) immuunvastuse

[10]. Esimene samm antikehade avastamisel on leida sobiv sihtmärk-molekul, mille vastased antikehad võiksid omada terapeutilist potentsiaali. Selle sihtmärk-molekuliga loomi immuniseerides (puhastatud molekul süstitakse looma) tekib neil loomuliku reaktsioonina immuunvastus, mille ühe osana hakkavad B-rakud tootma vajalikke antikehi. Tavapärase hübriidomide tekitamise meetodika puhul fuseeritakse põrna B-rakud müeloomi tüvirakkudega saades hübriidset antikeha tootva rakuliini [11]. Seda meetodikat on peamiselt rakendatud hiires. Icosagen CF poolt välja arendatud Hybrifree tehnoloogia jätab vahele hübriidomide tekitamise sammu. Selle asemel eraldatakse põrna B-rakkudest cDNA, antikeha variaabelregioonid isoleeritakse ning kloonitakse ekspressioonivektorisse [12]. Hybrifree meetodika puhul on kasutatud kanu ning küülikuid. Eelnevalt kirjeldatud meetoditel avastatud antikehi soovitakse lõpuks kasutada peamiselt inimeste ravimiseks, kuid nende aminohappeline järjestus on omane teistele loomadele ja seega inimesele võõras tekitades immuunvastust (anti-antikeha vastus) [13,14], mis võib vähendada ravimi efekti või halvemal juhul tekitada süsteemset põletikuvastust, mis võib lõppeda surmaga. Praegu käibel olevatest mABidest esineb osadel patsientidel anti-antikeha vastust näiteks cetuximab, infliximab ja basiliximab puhul, mis on hiire variaablitega kimäärsed antikehad [15,16].

3.3 Antikehade humaniseerimine

Immuunvastuse vähendamiseks soovitakse hiires või muus loomas avastatud antikeha muuta võimalikult inimese sarnaseks, samas säilitades afiinsuse antigeeni suhtes. Seda protsessi nimetatakse humaniseerimiseks. Antikehade humaniseerimise abil on eesmärgiks toota molekule, mis oleks inimestele minimaalse immunogeensusega e tekitaks võimalikult nõrga immuunvastuse kuid säilitaks samas esialgse antikeha spetsiifilisuse ja afiinsuse. Kõige suurema efekti immuunvastuse vähendamisel annab immunoglobuliini konstantsete osade vahetamine inimese omade vastu [13]. Sellisel kimäärsel antikehal säilitavad ainult variaabelregioonid oma algse kuju ja ülejäänud antikeha vastab inimese immunoglobuliinile. Samas põhjustavad katselooma variaabelregioonid jätkuvalt inimorganismis immuunvastust, mille tõttu identifitseeritakse piirkonnad, mis ei vastuta antikeha-antigeen interaktsiooni eest ning üritatakse neid vahetada rohkem inimestele „sarnaste“ vastu (Joonis 2) [4]. Nende kaotamiseks kasutatakse nn „*CDR crafting*“ meetodit, mille puhul katseloomades (tavaliselt hiires) avastatud antikehade antigeeni siduvad piirkonnad kantakse üle inimese variaabelregioonile [4]. Selle jaoks on vaja leida üksteisele võimalikult sarnased katselooma ja inimese (homoloogilised) variaabelregioonid, et CDR-ide vahelistel aladel kasutada aminohappeid, mis teeksid järjestuse võimalikult „humaanseks“ e inimesele sarnaseks [4]. On

leitud, et ka CDR-i piires ei ole kõik piirkonnad antigeeni sidumiseks vajalikud, seega CDR-i enda järjestust saab samuti humaniseerida jättes muutmata ainult spetsiifilisust määravad piirkonnad (SDR – *Specificity-Determining Residues*)[17].



Joonis 2 Antikeha humaniseerimine. Sinised piirkonnad humaniseeritud antikehal on afiinsust määravad piirkonnad ehk CDR-id.

3.3.1 CDR-ide identifitseerimine

Kuna enamasti antikeha humaniseerimisel CDR piirkondi ei modifitseerita või tehakse seda minimaalselt, siis esmalt on need vaja identifitseerida. CDR-ide identifitseerimiseks on töötatud välja mitmeid skeeme, üks algsemaid ja ka praegu laialdaselt kasutusel olev kannab nime Kabati skeem. Kabat ja kolleegid joondasid erinevaid antikeha variaabelregioone ning eeldusel, et CDR-id on kõige varieeruvama järjestusega määrati neile numbrilised asukohad (rasketel ahelatel positsioonides 31–35b, 50–65, 95–102 ja kergetel ahelatel positsioonides 24–34, 50–56, 89–97) [18,19]. Kabati skeem töötati välja piiratud hulga mABide sekveneerimisandmete põhjal ajal, kui ei olnud veel antikehade struktuuri kohta palju infot ja seega on hilisemad meetodid CDR-ide positsioone täpsustanud.

Veidi hilisem Chothia skeem on pea identne Kabat-i omale, kuid raske ja kerge ahela esimesed CDR-id asuvad struktuurselt õigemal kohas [20]. Hilisemalt on antikeha struktuure arvesse võttes Chothia skeemi täiendatud ning <http://www.bioinf.org.uk/abs/abnum/> lehel on üleval ka tööriist, millega antikehasid nummerdada [21]. Seda tööriista kasutades võrreldi Kabat andmebaasis olevat antikehade nummerdust täiendatud (struktuurselt korrektse) Chothia skeemiga (Martin-i skeem) ning leiti, et Kabat andmebaasis on kuskil 10% antikehadel vale nummerdus [21] hoolimata sellest on see veel praegugi laialdaselt kasutusel.

Välja on töötatud ka antikehade nummerdamissüsteem, mis arvestab paremini antikeha ruumilist struktuuri. WolfGuy nummerdamisskeem määratleb CDRid Kabat ja Chothia

skeemide ühisosa järgi, kuid lisaks sellele määratletakse seal CDRi silmuste tipud ning tipule eelnevad aminohappejäägid nummerdatakse kui „tõusvad“ ning sellele järgnevad kui „laskuvad“ [22,23] (Tabel 1). WolfGuy on konstrueeritud nii, et aminohappe jäägid, mille ruumiline paiknemine on konserveerunud (evolutsiooniliselt stabiilne) nummerdatakse võimalikult suurel määral samaväärsete indeksitega [22,23], mistõttu on võimalik selle kaudu mudeldada antikehade homoloogiat.

Tabel 1 Wolfguy ja Chothia nummerdamisskeemide võrdlus ühe kana antikeha kerge ahela variaabelregiooni näitel. Tõusvad ja laskuvad ahelad identifitseeritud WolfGuy skeemi järgi. CDR piirkonnad Chothia järgi märgitud rohelisega.

| I tõusev | | | I laskuv | | | II tõusev | | | II laskuv | | |
|----------|---------|---------|----------|---------|---------|-----------|---------|---------|-----------|---------|---------|
| | WolfGuy | Chothia | | WolfGuy | Chothia | | WolfGuy | Chothia | | WolfGuy | Chothia |
| A | 503 | 3 | N | 596 | 28 | N | 694 | 51 | G | 751 | 89 |
| L | 504 | 4 | Y | 597 | 29 | D | 695 | 52 | I | 752 | 90 |
| T | 505 | 5 | Y | 598 | 30 | K | 696 | 53 | P | 753 | 91 |
| Q | 506 | 6 | G | 599 | 34 | R | 697 | 54 | D | 754 | 92 |
| P | 507 | 7 | W | 601 | 35 | P | 698 | 55 | N | 755 | 93 |
| S | 509 | 8 | Y | 602 | 36 | S | 699 | 56 | T | 795 | 94 |
| S | 510 | 9 | Q | 603 | 37 | D | 701 | 57 | Y | 796 | 95 |
| V | 511 | 11 | Q | 604 | 38 | I | 702 | 58 | I | 797 | 95 |
| S | 512 | 12 | K | 605 | 39 | P | 703 | 59 | A | 798 | 96 |
| A | 513 | 13 | AS | 605 | 39 | S | 704 | 60 | I | 799 | 97 |
| N | 514 | 14 | P | 606 | 40 | R | 705 | 61 | F | 801 | 98 |
| L | 515 | 15 | G | 607 | 41 | F | 706 | 62 | G | 802 | 99 |
| G | 516 | 16 | S | 608 | 42 | S | 707 | 63 | A | 803 | 100 |
| G | 517 | 17 | A | 609 | 43 | G | 708 | 64 | G | 804 | 101 |
| T | 518 | 18 | P | 610 | 44 | S | 709 | 65 | T | 805 | 102 |
| V | 519 | 19 | V | 611 | 45 | K | 710 | 66 | T | 806 | 103 |
| E | 520 | 20 | T | 612 | 46 | S | 711 | 67 | L | 807 | 104 |
| I | 521 | 21 | L | 613 | 47 | G | 712 | 68 | T | 808 | 105 |
| T | 522 | 22 | I | 614 | 48 | S | 715 | 69 | V | 809 | 106 |
| C | 523 | 23 | Y | 615 | 49 | T | 716 | 70 | L | 810 | 107 |
| S | 551 | 24 | N | 651 | 50 | A | 717 | 71 | | | |
| G | 552 | 25 | | | | T | 718 | 72 | | | |
| G | 553 | 26 | | | | L | 719 | 73 | | | |
| G | 554 | 27 | | | | T | 720 | 74 | | | |
| | | | | | | I | 721 | 75 | | | |
| | | | | | | T | 722 | 76 | | | |
| | | | | | | G | 723 | 77 | | | |
| | | | | | | V | 724 | 78 | | | |
| | | | | | | Q | 725 | 79 | | | |
| | | | | | | V | 726 | 80 | | | |
| | | | | | | E | 727 | 81 | | | |
| | | | | | | D | 728 | 82 | | | |
| | | | | | | E | 729 | 83 | | | |
| | | | | | | A | 730 | 84 | | | |
| | | | | | | V | 731 | 85 | | | |
| | | | | | | Y | 732 | 86 | | | |
| | | | | | | F | 733 | 87 | | | |
| | | | | | | C | 734 | 88 | | | |

IMGT (*international ImMunoGeneTics information system*) on töötanud välja oma nummerdamisskeemi, mis põhineb terve andmebaasi referentsgeenide joendamisel, kus on arvesse võetud kogu immunoglobuliinide superperekonda [24–26]. IMGT standardiseeritud

käsitlus mutatsioonidest, alleelsetest polümorfismidest, 2D ja 3D struktuuridest jne on võimaldanud neil välja arendada mitmeid immunogeneetikas ja immunoinformaatikas kasulikke tööriistu [27] ning IMGT standardeid (sealhulgas nummerdusskeemi) tunnustab ja kasutab ka WHO-IUIS (World Health Organization - International Union of Immunological Societies) [28].

Lisaks eelnevalt käsitletule on mitmeid erinevaid nummerdamisskeeme ja neid on käsitletud paljudes ülevaateartiklites, üks hilisemaid ja põhjalikumaid on Dondelinger et al. [29] <http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/anarci/> lehel asub rakendus, mis võimaldab järjestusi nummerdada erinevate skeemide alusel.

3.3.2 Inimese antikeha raamistik

Antikeha variaabelregiooni CDR piirkondi nimetatakse hüpervarieeruvateks piirkondadeks ning nende külgnevaid alasid „raamistikuks“ (FR - *framework region*). Sellel, millise antikeha variaabeljärjestuse peale CDR-id poogitakse, võib olla suur mõju antikeha omadustele [30]. Üks võimalus, mis on ka laialdaselt kasutusel, on kasutada algseid nõ iduliini immuunglobuliinide variaabelregioone. Need kaasasündinud järjestused on aluseks kõigile organismi poolt toodetud antikehadele ning nende sihtmärgispetsiifilisus saavutatakse läbi genoomsete ümberpaigutuste ja mutatsioonide [31]. Iduliini järjestused, mis on saadud inimese, hiire ja ka teiste loomade sekveneerimisel on koondatud IMGT/GENE-DB andmebaasi [32]. Iduliini raamistiku kasutamise pooldajad väidavad, et kuna iduliini järjestused on vähem immunogeensed, sest „küpsed“ järjestused läbivad afiinsuse saavutamiseks mitmesuguseid ümberkorraldusi, võiks neid immuunreaktsiooni vältimiseks eelistada [33,34]. Kui iduliini raamistikud lähtuvad genoomsest DNA-st ja on arvuliselt piiratud, siis „küpsed“ raamistikud kajastavad bioloogiliselt aktiivseid molekule põhinedes immuunglobuliini aminohappelisel järjestusel [31]. „Küpsed“ järjestused moodustuvad iduliini järjestuse mitmetes kombinatsioonides ümberkorraldamisel ja seega nende arv on ka kordades suurem [31]. Järjestuste rohkust ning vastavust reaalselt afiinsetele immuunglobuliinidele peetakse ka „küpsete“ järjestuste peamiseks eelisteks [35]. Vastavaid järjestusi võib leida näiteks NCBI-GenBank [36], UniProtKB/trEMBL [37] ja Protein DataBank (PDB) [38] andmebaasidest. Selliste andmebaaside kasutamisel tasub aga olla ettevaatlik, sest tihtipeale esineb seal vigu ja vaatamata märksõnade/otsinguparameetrite täpsustamisele võivad valimisse sattuda mitte-inimese järjestused [39]. Raamistiku valikuga võib minna veelgi kitsamaks – näiteks tasulisse andmebaasi TABS on koondatud ainult terapeutiliste antikehade järjestused, mis on kas kliiniliste või prekliiniliste uuringute faasis [40,41]. Tõenäoliselt on seal muidugi päris

humaansete järjestuste arv piiratud ja võib olla raske leida sobivat raamistikku, kuid õnnestumise korral võiks sellise andmebaasiga humaniseerimine anda stabiilseid, hea tootlikkusega ning madalat immuunvastust tekitavaid antikehi.

3.3.3 Humaansuse hindamine

Kuna ravimi manustamisel tekkiv immuunvastus võib vähendada ravimi efekti või halvemal juhul tekitada süsteemset põletikuvastust, mis võib lõppeda surmaga [42], on antikehade humaansuse hindamiseks välja töötatud erinevaid meetodeid. Laialdaselt on kasutusel IMGT/DomainGapAlign, mille antavat humaniseerituse skoori peetakse ka kuldseks standardiks „humaansuse“ hindamisel. Antud tööriist võimaldab aminohappeliste järjestuste ja 2D struktuuride analüüsi. Kasutades IMGT antikehade andmebaasi leitakse sisendjärjestusele kõige sarnasemad järjestused, kõige tõenäolisem päritoluliik ning järjestuste homoloogia protsent [27]. WHO-INN (*International Nonproprietary Names*) ja selle aluseks olev AMA (*American Medical Association*) on kasutanud IMGT skoori uutele MAB-idele universaalse ning standardiseeritud nime andmisel [43]. Tunnustatud mABide nimed koosnevad -mab järelliitest, millele eelneb tähekombinatsioon, mis viitab antikeha tüübile, näiteks kimäärne (-xi-), humaniseeritud (-zu-), või inimese (-u-) antikeha ning millele omakorda eelneb antikeha sihtmärk, näiteks kasvaja(-t(u)-), närvisüsteem (-n(e)-) jne [44]. Nime algus on vabalt valitav. Seega näiteks rinnavähi ravis kasutusel olev tras-tu-zu-mab viitab monoklonaalsele humaniseeritud antikehale, mille sihtmärgiks on kasvaja. Jones ja kolleegid kritiseerisid -zu-liite jaoks vajalikuks piirmääraks seatud IMGT „humaansuse“ skoori $\geq 85\%$, mille järgi ka inimestest avastatud antikehad ei pruugi saada inimese antikeha klassifikatsiooni [43]. Samuti tõid nad välja, et skoor $\geq 85\%$ teeb võimatuks humaniseeritud (-zu-) ja inimese (-u-) järjestuste eristamise ning kirjanduse järgi ei ole antikehadel, mis kvalifitseeruvad inimese järjestuseks immuunvastus oluliselt vähenenud võrreldes nendega, mis ei kvalifitseeru [43]. WHO-INN komisjon vastas kriitikale, et nemad ei kasuta protsentuaalset läveväärtust humaansuse hindamisel ning ei ennusta immuunvastust vaid kasutavad järjestuse andmeid ainult nime panemiseks [45].

Peale IMGT skoori on „humaansuse“ hindamiseks kasutusel ka antikehade arendusteenust pakkuva firma LakePharma poolt arendatud T20 skoor. Nende rakenduses leitakse CDR-id kasutades Kabat nummerdamisskeemi, CDR-id eemaldatakse ning variaabelregiooni või täisjärjestust võrreldakse iga antikehaga andmebaasis kasutades blastp valgu BLASTimis algoritmi. Top 20 sisendjärjestusele kõige lähedasema inimese järjestusega leitud kattuvuse % keskmistatakse T20 skooriks [46]. T20 skoori järgi hinnati täisjärjestusi (konstantsed regioonid

+ variaabelregioon), mille skoor oli üle 80% ja variaabelregioone, mille skoor oli üle 85% inimese sarnaseks [46].

4. PERMIDES projekt

4.1 PERMIDES projekti ülevaade

Projekt sai käivitatud biotehnoloogiafirma Icosagen Cell Factory (ICF) soovist omada humaniseerimisrakendust, mida saaks kasutada nii omaarenduste raames kui pakkuda klientidele osana antikehade arenduse teenusest. ICFis on võimalik teostada mitmeid terapeutiliste antikehade arenduseks vajalikke etappe: prekliinilises osas Hybrifree tehnoloogiaga antikehade avastamine kanadest ja küülikutest [12], rekombinantsete antikehade tootmine transientselt (ajutiselt) imetaja rakuliinis ning 2016-2018 kestnud IcoCell projekti raames arendati välja ka suuremas mahus antikehade tootmine püsiliini abil, mida on vaja kliinilises faasis [47]. Võimaldamaks üleminekut prekliinilisest faasist kliinilisse (st inimorganismi) sooviti arendada koostöös Quretec-iga antikehade humaniseerimist võimaldav tarkvara. Selleks kasutati osaliselt EU projekti PERMIDES (*Personalised Medicine Innovation through Digital Enterprise Solutions*) rahasid. Rahastuse eesmärk on tugevdada personaalmeditsiini kui esilekerkiva tööstusharu konkurentsivõimelisust Euroopas läbi innovaatiliste IT lahenduste.

Kuigi humaniseerimise teenust saab ka sisse osta ning on olemas isegi tasuta rakendused nagu Tabhu [48] sooviti ICFis omada personaalset tarkvara, sest konkurents teiste antikehasid arendavate firmadega on tihe ja et konkurents püsida tuleb pakkuda võimalikult terviklikku teenust. Lisaks on tavaliselt prekliinilises staadiumis olevad antikeha järjestused juba patenteeritud ja konfidentsiaalsuslepinguga kaetud ning nende levitamine ja avalikesse veebirakendustesse laadimine ei ole lubatud.

4.2 Qumanize rakendus

Qureteci arendatud rakendus nimega Qumanize võtab sisendiks antikeha kerge ja raske ahela variaabelregioonid ning väljastab humaniseeritud järjestused. Kuna suur osa antikehade avastustest on toimunud hiires, on ka paljude ettevõtete humaniseerimisteenus hiirekeskne. Qumanize eeliseks on see, et sisendiks võib olla üks kõik mis looma variaabelregioon. Samuti on humaniseerimisteenuse pakkujatel tavaliselt kindel andmebaas, mille hulgast leitakse inimese antikeha variaablite raamistikud, kuhu CDR-id peale pookida. Olenevalt pakkujast

eelistatakse siis vastavalt iduliini või „küpsid“ antikeha raamistikke. Qumanize võimaldab kasutada ükskõik millist andmebaasi, tingimusel, et järjestused on õigele kujule töödeldud.

Projekti käigus kasutati sisendina NCBI geenipangast (NCBI GenBank) kopeeritud järjestusi kasutades otsingus märksõnu „immunoglobulin heavy/light/kappa chain variable region“. Otsingute tulemusena saadi ~80000 raske ja ~29000 kappa (κ) ja lambda (λ) ahela järjestust. Avalikes andmebaasides võib leida eksitavat või puudulikku infot, seetõttu eemaldati kopeeritud andmeid filtreerides mitte-terviklikud ja mitte-inimese järjestused.

Selleks, et tuvastada inimese järjestused, mis on looma sisendjärjestusele kõige sarnasemad, koostab Qumanize programm positsioonisageduse maatriksi (PFM – *Position Frequency Matrix*). Muutmata jäävad CDR piirkonnad, et säiliks antikeha afiinsus. Qumanize tarkvaras ei ole kasutatud enamlevinud nummerdamiskeeme nagu Kabat ja Clothia vaid CDRid tuvastatakse WolfGuy definitsiooni järgi [22]. WolfGuy numeratsiooniskeem omistab variaabli igale konserveerunud positsioonile unikaalse indeksi, identifitseerib CDR-ahelate silmuste tipud ja eristab üksteisest tõusva ja laskuva ahela segmente [22]. Selline indekseerimine annab andmeanalüüsiks lisaparameetri võimaldades säilitada infot variaabelregioonide struktuuri kohta.

Qumanize koostab leitud järjestuste põhjal „humaanse“ konsensusjärjestuse ning hakkab sellest sammhaaval aminohappeid muutes minema tagasi sisendjärjestuse poole. Asendusi tehakse vastavalt PAM120 (*Percent Accepted Mutation*) sarnasusmaatriksile. PAM maatriksid on kalkuleeritud loodusliku valiku teel akumulbeerunud aminohappe mutatsioonide sageduste järgi ning PAM120 vastab ajale, mil 100 aminohappe kohta tekib 120 mutatsiooni [50]. Qumanize rakenduse poolt tekitatud gradiendis vahetatakse PAM120 maatriksi järgi sarnasemad (omavahel lähemas suguluses) olevad aminohapped varem välja kui evolutsiooniliselt kaugel olevad aminohapped.

Rakendus arvutab ka IMGT skoori ning T20 skoori. Eeldatavasti oleks parim järjestus võimalikult algjärjestuse sarnane, kuid saavutaks veel IMGT ja T20 skoori üle 85%.

4.3 PERMIDES projekti tulemused

Rakenduse võimekust hinnati ka eksperimentaalselt – selleks et rakenduse poolt välja pakutud variaabeljärjestused oleksid realselt kasutatavad, peavad väljapakutud antikehad olema hea tootlikkusega, hea afiinsusega antigeenile ning mitte tekitama immuunvastust. Esimest kahte nendest parameetritest oli võimalik ICF-is testida. Katse sisendiks oli üks ICF-is avastatud kana antikeha, mille variaabelregioonid humaniseeriti lisaks Qumanize rakendusele veel kahe teise

pakkuja poolt. Humaniseeritud variaabelregioonid fuseeriti inimese antikeha konstantsete osadega (hIgG1/lambda) ning erinevaid kerge ja raske ahela variaablite kandidaate kombineerides viidi läbi mitu tootmistsükli CHO rakuliinis. Esimese tsükli eesmärk oli elimineerida kerge ja raske ahela kombinatsioonid, mis üldse ei ekspresseeru, ei saa rakust välja või ei tunne ära antigeeni. Selleks teostati väga väikses mahus lühiajaline tootmine kõikide võimalike variaablite kombinatsioonidega ning teostati antigeenile seondumise katse. Kahe teise algoritmiga humaniseeritud antikehade valimist jäi katsesse edasi vastavalt 42% ja 67%, Qumanize algoritmiga 42%. Teise tsükliga toodeti välja valitud antikehi veidi suuremas mahus, et saaks testida nende afiinsust sihtmärkvalgule – igast võrdluses olevast grupist valiti välja üks kõige kõrgema potentsiaaliga antikeha mille afiinsus ei olnud humaniseerimisprotsessi käigus vähenenud ning viidi läbi veel kolmas tootmistsüklil. Sedakorda toodeti antikehi suuremas mahus, puhastati afiinsuskromatograafia abil ning teostati seondumiskineetika analüüs sihtmärkmolekulile, et hinnata väljavalitud antikehade „arendamispotentsiaali“ ehk seda, kui kiiresti ja tugevalt antikehad sihtmärkvalgule seonduvad. Tootlikkuses olid väikesed kõikumised, kuid Qumanize rakendusega humaniseeritud antikeha andis keskmise tootlikkuse ning afiinsuses suuri erinevusi erinevate algoritmide vahel polnud.

Viimane etapp oleks testida immuunvastust inimesel, kuid selleni pole veel jõutud. Praeguste katsete järgi võib öelda, et Qumanize rakendus võimaldab produtseerida sama hea tootlikkuse ja afiinsusega antikehasid, mis teised kommertsiaalselt pakutavad lahendused.

5. Kasutajaliidese loomine

Käesoleva magistr töö praktiliseks eesmärgiks oli luua PERMIDES projekti käigus arendatud antikehade humaniseerimise rakendusele kasutajakeskkond. Rakendus oli kättesaadav serveri kaudu ning funktsioneeris käsura programmina (LISA 1: käsura programmile loodud juhend), sest see lahendus täitis rakenduse loomise hetkel kasutajate vajadusi, kes olid 1-2 üsna heade tehniliste oskustega inimest. Kaugemaks eesmärgiks on siiski rakenduse kasutajate ringi suurendada ning teha rakendus kättesaadavaks ja mõistetavaks ka uutele kasutajatele, seetõttu otsustati alustada ka veebipõhise kasutajaliidese loomist. Käesoleva magistr töö raames valmiski kasutajaliidesele esimene funktsionaalne põhi. Lisaks sujuvamale kasutuskogemusele pakub veebipõhine kasutajaliides võimaluse rakenduse väljundi välimust vastavalt kasutaja soovile modifitseerida (märgistades näiteks hüdrofoobsed ja hüdrofiilsed aminohapped erinevalt või märkides järjestustele CDRid). Kuna antud tööriistale on võimalik lisada väga

paljusid proteoomikast ja immuunoinformaatikast pärit lisafunktsioone, siis klient otsustas, et inimressursi kõige efektiivsem kasutus oleks lisada neid alles siis kui rakenduse kasutamisel ilmneb nende järele vajadus.

Veebiliidese loomise eelduseks on Qumanize rakenduse põhjalik tundmine (vt ptk 5.1), sest käsurea rakenduse muutmine veebirakenduseks eeldab koodi modifitseerimist. Veebiliidese loomiseks kasutati Pythoni põhist Flask raamistikku (vt ptk 5.2) ning kujunduse jaoks saadi abi Bootstrapiga integreeruvast laiendusest Flask-Bootstrap (vt ptk 5.3).

5.1 Qumanize rakenduse tööpõhimõtted

Qumanize rakendus, nagu paljud teised bioinformaatika tarkvarad, on arendatud kasutades Python programmeerimiskeelt. Python on saanud bioteadustes populaarseks programmeerimiskeeleks peamiselt seetõttu, et see on kõrgetasemeline keel, mille loomisel oli eesmärgiks nii hea koodi loetavus, kui ka sobilikkus kiireks prototüüpimiseks. Pythoni paljud saadaolevad raamatukogud ja kolmanda osapoole tööriistakomplektid laiendavad põhikeele funktsionaalsust praktiliselt igasse bioloogia valdkonda (järjestuse ja struktuuri analüüsid, fülogenoomika, töövoo juhtimissüsteemid jne) [51].

5.1.1 Qumanize rakenduse jaoks vajalikud Pythoni paketid

Erinevate bioinformaatikas vajalike moodulite ja skriptide koondamiseks on loodud Biopython Projekt (<http://www.biopython.org>). Biopython-i funktsioonide hulka kuuluvad mitmesuguste bioinformaatikas kasutusel olevate failivormingute parserid (BLAST, Clustalw, FASTA, Genbank jne), juurdepääs veebiteenustele (NCBI, Expasy jne), liideseid levinud ja vähemlevinud programmidele (Clustalw, DSSP, MSMS ...), standardsete järjestuste klass (Seq), erinevad klasterdamise moodulid, KD-puu andmestruktuur jne [52].

Qumanize rakenduses, nagu ka bioinformaatikas üldiselt on keskseks objektiks järjestused. Qumanize saab kasutajalt sisendiks antikeha variaabeljärjestuse FASTA formaadis:

```
>12H3_VL
ALTQPSSVSANLGGTVEITCSGGGNYYGWYQQKSPGSAPVTLIYNNDKRPSDIPSRFSGSKSGSTATLTIITGVQV
EDEAVYFCGIPDNTYIAIFGAGTTLTVL
```

Biopythoni abil tehakse sellest SeqRecord objekt, kust on võimalus tuvastada atribuudid „id“, mis on antud juhul 12H3_VL ja järjestuse atribuudi „seq“, mille kaudu saab kätte Seq objekti Seq („ALTQPSSV...“).

Järjestuste identifitseerimiseks ja joondamiseks tehakse neist edasi ANARCI klassi objektid [53]. ANARCI on vahend antikehade ja T-raku retseptori aminohapete varieeruva domääni

järjestuste klassifitseerimiseks ja nummerdamiseks. ANARCI võimaldab võtta sisendiks ühe või mitu antikeha järjestust ning märkida neid populaarsemate numeratsiooniskeemide alusel: Kabat, Chothia, Enhanced Chothia, IMGT, Wolfguy ja AHO. [53] Programm teeb koostööd HMMER3 paketi ja joondab konkreetse liigi (inimese, hiire, roti, jänese, sea või reesusahvi) domeenitüübi (VH, Vλ või Vκ, Va või Vβ) oletatavaid iduliini järjestusi kasutades tõenäosuslikke mudeleid, mida nimetatakse varjatud Markovi mudeliteks (HMM) [54].

ANARCI töötab ainult Linux platvormil ja seega rakenduse jooksutamiseks Windowsi peal kasutati Dockerit. Docker Desktop on rakendus MacOS- ja Windows-masinatele konteineripõhiste rakenduste ja mikroteenuste loomiseks ja jagamiseks, kus konteineris saab ise määrata Linux põhise operatsioonisüsteemi [55]. Qumanize arendustöö puhul oli kasutusel CentOS.

5.1.2 Qumanize rakenduse struktuur

```
[katrin@DESKTOP-UU4AGLG ~]$ tree /home/katrin/Documents/ico-qumanize
/home/katrin/Documents/ico-qumanize
├── anarci.py
├── annotate.py
├── Documentation
│   ├── README.rst
│   └── requirements.txt
├── home.py
├── human_db_uniq_germline.anarci
├── templates
│   └── template.html
└── utils.py
```

Joonis 3 Qumanize rakenduse failipuu

Rakendus käivitatakse failis `home.py` – läbi selle laetakse üles veebileht renderdades malli `template.html`. Fail `home.py` saab läbi FlaskWTFORMS (vt ptk 5.2) sisendiks kasutaja sisestatud variaabeljärjestused ning kasutaja poolt määratud lisaargumendid (nagu WHO kriteeriumitele vastavus jne) ning suunab need funktsiooni `humanize`. Lisaks võetakse sisendiks inimese antikehade referentsandmebaasi `anarci` formaadis (`human_db_uniq_germline.anarci`), mis on hetkel vaikimisi määratud, kuid teoorias saaks rakenduse juurde lisada ka teiste liikide andmebaase, mille seast kasutaja saab sobiva valida. Funktsioon `humanize` koondab kogu humaniseerimise protsessi ühte tervikusse ning väljastab veebilehele humaniseeritud järjestused. Referentsandmebaasist leitakse kasutaja sisestatud variaabeljärjestustele kindel arv (k) kõige sarnasemaid inimese järjestusi. Need liiguvad edasi faili `anarci.py`, kus leitud

järjestustest omakorda sõelutakse välja omavahel kõige sarnasemad, mille alusel koostatakse positsioonisageduse maatriks (PFM). Järgnevalt leitakse järjestuste CDRid ning vastavalt kasutaja sisestatud järjestusele hakatakse ülejäänud raamistikku järk-järgult muutma, nii et iga järgneva iteratsiooni tulem on uus PFMi konsensusjärjestus. Lõppkokkuvõttes saadakse gradient, mis algab kasutaja sisestatud järjestusega ja aminohappe haaval muundub kokku pandud inimese konsensusjärjestusele sarnasemaks.

Sealt suunatakse järjestused edasi faili `annotate.py`, kus lisatakse järjestustele erinevaid parameetreid (loomaliik IMGT andmebaasi järgi, humaniseeritusse skoorid jne).

Fail `utils.py` sisaldab funktsioone, mille abil luua humaniseeritud järjestusi sisaldavaid fasta faile (esialgu veebirakenduses kasutuses pole). Samuti kontrollitakse seal humaniseeritud järjestuste vastavust WHO kriteeriumile, mis võimaldab kasutaja soovil kriteeriumile mitte vastavad järjestused välja filtreerida. Lisaks on seal funktsioon `add_triplets`, mille abil jagatakse humaniseeritud järjestus kõikideks võimalikeks 3-aminohappelisteks motiivideks ning võrreldakse neid sama moodi jagatud inimese antikehade variaabelregioonidega. Juhul kui humaniseeritud järjestus sisaldab selliseid 3-aminohappelisi kombinatsioone, mida natiivselt inimeses ei leidu (või leidub väga madalas sageduses), siis need võivad olla potentsiaalsed immunogeensuse allikad, seega peaks need ka humaniseeritavas järjestuses ära muutma.

5.2 Flaski veebiraamistik

Flask on Pythoni veebiraamistik. Veebiraamistikud standardiseerivad veebirakenduste loomisest ja juurutamist automatiseerides üldisi ja korduvaid tegevusi veebiarenduses. Näiteks pakuvad paljud veebiraamistikud teeke andmebaasidele juurdepääsu, malliraamide ja seansihalduse jaoks ning soodustavad koodi taaskasutamist. Flask põhineb Werkzeug and Jinja 2 raamatukogudel. Werkzeug on kõikehõlmav veebiserveri lüüsiliidese (WSGI e *Web Server Gateway Interface*) veebirakenduste raamatukogu, mis sisaldab suurt hulka teenuseid ja tööriistu veebirakenduste loomiseks [56]. Jinja 2 on Pythonile loodud keel mallide (*template*) loomiseks e mallimootor, mille aluseks on kasutatud Django sisse-ehitatud malle [57]. Sisuliselt sisaldab Jinja2 malli faili vastust päringule ning vastuse dünaamilised osi muutujatena: `{{muutuja}}` [58].

Flask ning Django on kõige levinumad pythoni põhised veebiraamistikud jättes teised kaugele maha, sealjuures Flask on viimastel aastatel muutunud populaarsemaks kui Django [59]. Nii Flask kui Django on tasuta ja avatud koodiga, kuid Flaskil puuduvad andmebaasi abstraktsioonikiht, vormi kinnitamine ning muud komponendid, kus olemasolevad kolmanda

osapoole teegid pakuvad ühiseid funktsioone – seetõttu nimetatakse Flaski mikrorakendusteks [58]. Seega Django pakub rohkem funktsionaalsust nõ *out-of-the-box* kuid Flask on paindlikum, kompaktsem ning mis antud töö kontekstis oluline – kergem õppida [60].

5.2.1 Flask veebiraamistik Qumanize kasutajaliidese loomisel

```
class SubmitForm(FlaskForm):
    name = TextAreaField('Insert variable sequences in FASTA format', render_kw={"rows": 15, "cols":
7})
    top = IntegerField('How many most similar sequences to the source should be used from the
database?', default=10)
    thr = FloatField('Threshold for germline identity score', default=0.95)
    diss = FloatField('Value 0.1 -1, lower value requires sequences to be more dissimilar to each-
other', default=0.9)
    long = BooleanField('Create more candidate sequences (takes a long time)')
    who = BooleanField('Display only WHO compliant output sequences')
    submit = SubmitField('OK!')

@app.route('/', methods=['GET', 'POST'])
def index():
    hum = None
    out = None
    form = SubmitForm()
    if form.validate_on_submit():
        name = form.name.data
        with open('source.fasta', 'w') as f:
            f.write(name)
        source = ana.run_anarci(utils.read_fain('source.fasta'))
        db = ana.parse_anarci('human_db_uniq_germline.anarci')
        top = form.top.data
        long = form.long.data
        thr = form.thr.data
        diss = form.diss.data
        who = form.who.data
        hum, out = humanize(source, db, top, long, thr, diss, who)
        out = "\n".join(out)
    return render_template('template.html', form=form, hum=hum, out=out)

if __name__ == '__main__':
    app.run(host='0.0.0.0', debug=True)
```

Joonis 4 Flaski komponendid Qumanize rakenduse veebilahenduses failis home.py

Qumanize rakendusele veebiliidese puhul suunatakse brauserist saadatud päring veebiserverisse läbi Flaski klassi objekti, kus on kaardistatud milline Pythoni kood on seotud iga taotletud URL-iga. Seost URL-i ja seda haldava funktsiooni vahel nimetatakse *route* ning selleks kasutatakse `@app.route` dekoraatorit (*decorator* – Pythoni funktsioon, mis võtab argumendiks teise funktsiooni ning väljastab modifitseeritud funktsiooni) (Joonis 4).

Selleks, et käivitada Flaski integreeritud veebiserverit, kasutatakse *run* meetodit:

```
if __name__ == '__main__':
    app.run(host='0.0.0.0', debug=True)
```

Vormide kuvamiseks ning nende kaudu rakendusele sisendi saamiseks kasutati Flask-WTForms laiendust, mis aitab lihtsate Pythoni mudelitega veebivorme luua ja kasutada

pakkudes ka lisa turvakihti, mis on vajalik veebirakendusele Interneti-kasutaja kaudu esitatud andmete kasutamiseks [61]. Iga vormi väli on defineeritud klassi muutujana klassis SubmitForm, kus saab määrata väljade parameetrid, vaikeväärtused, validaatoreid jne (Joonis 4).

Funktsioon `render_template @app.route` dekoraatori juures seob Jinja2 mallimootori rakendusega võttes esimeseks argumendiks faili nime (`template.html`) ning ülejäänud argumendid on võtme/väärtuse paarid, mis annavad mallis kirjeldatud muutujatele väärtused (`form=form`) (Joonis 4).

Joonis 5 välja toodud Jinja2 mallis (`template.html`) on määratud muutujad, mis saavad endale väärtuse programmi poolt tekitatud andmetest siis kui malli sisu teisendatakse (`renderdatakse`) (näiteks `{{out}}` kuvab humaniseerimise käigus loodud gradiendi sisendjärjestusest humaniseeritud järjestuseni). Jinja2 pakub ka kontrollstruktuure, mis võimaldavad kasutada näiteks `if` ja `for` lauseid (Joonis 5).

```
<div class="col-sm-8 text-left">
  <h1>Qumanize</h1>
  {{ wtf.quick_form(form)}}
  <hr>
  <h3>Output</h3>
  {% if form.name.data %}
    <table>
      <thead>
        <tr>
          <th class="col1">ID</th>
          <th class="col2">Humanized sequence</th>
          <th class="col3">Parent organism</th>
          <th class="col4">T20 Score</th>
          <th class="col5">IMGT Score</th>
        </tr>
      </thead><tbody>
        {% for seq in hum %}
          <tr class="trow">
            <td class="col1">{{seq.id}}</td>
            <td class="col2">{{seq.seq}}</td>
            <td class="col3">{{seq.annotations['pr_org']}}</td>
            <td class="col4">{{seq.annotations['t20']}}</td>
            <td class="col5">{{seq.annotations['ident']}}</td>
          {% endfor %}
        </tr>
      </tbody>
    </table>
    <pre>{{out}}</pre>
  {% else %}
    <p>Järjestust pole sisestatud</p>
  {% endif %}
</div>
```

Joonis 5 Qumanize rakenduse kasutajaliidese koodi väljavõte Jinja2 mallis failis `template.html`

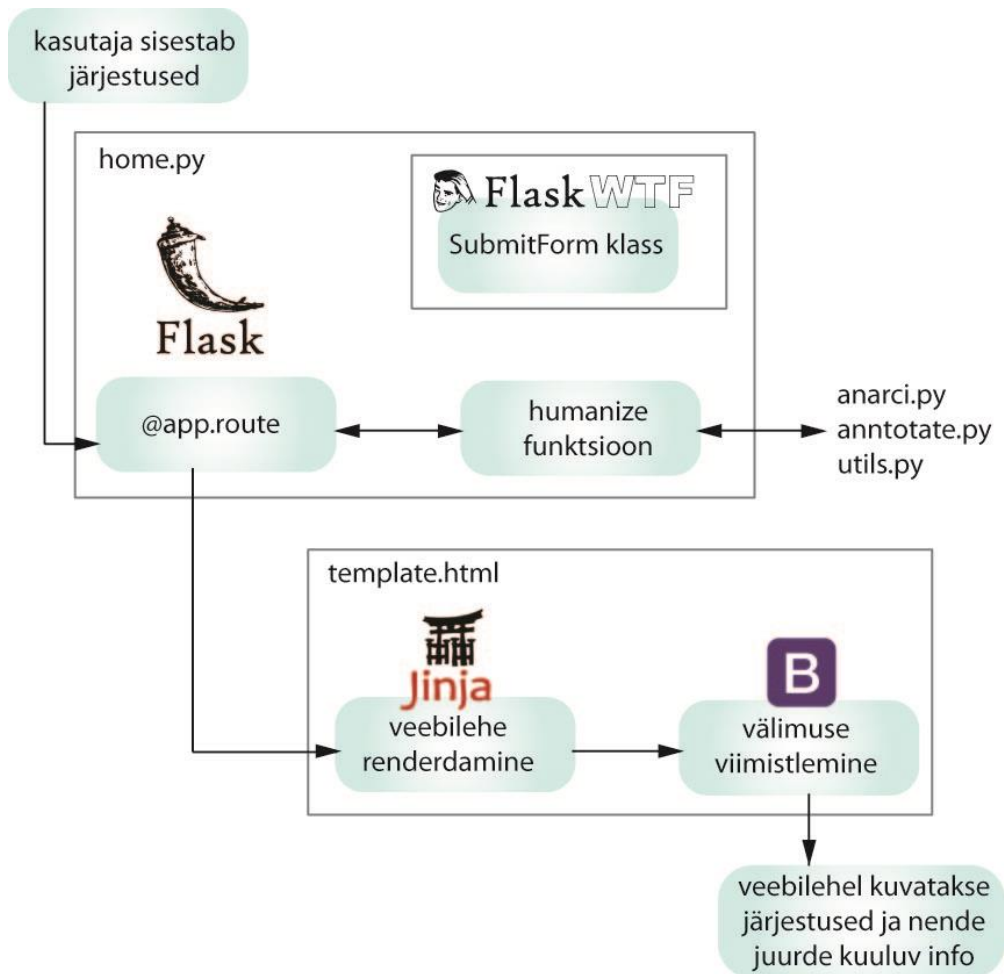
5.2.2 Bootstrapi raamistiku kasutamine kujunduses

WTForms vorme peetakse tänapäeva standardite järgi „koledaks“ - Flask-Bootstrap pakub nendele vormidele professionaalset poleerimiskihti varjutamise, esiletõstmise ja hüpikakendega. Twitteri poolt loodud Bootstrap on avatud lähtekoodiga tööriistakomplekt HTML-i, CSS-i ja JS-iga arendamiseks [62]. Bootstrap võimaldab hoida sama kujundust kõigis veebibrauserites, saab hakkama erinevate ekraanisuurustega, pakub valmis paigutusega lahendusi (*layout-e*), mida saab oma soovile vastavalt ümber kujundada ning stiliseeritud vorme, navigeerimisribasid, nuppe jne. Flask-Bootstrap pakendab Bootstrapi laienduseks, mis pakub kasutamiseks valmis malli, kuhu on installitud Bootstrap-raamistik [63]. Flask-Bootstrap pakub ka meetodit "quick_form", mis käsib Jinja2-l terve veebilehe ühe koodireaga renderdada.

Qumanize rakenduse jaoks kasutati veebis vabalt saada olevat Bootstrapi veebilehe malli <https://maxcdn.bootstrapcdn.com/bootstrap/3.4.0/css/bootstrap.min.css>. Kuigi Bootstrap pakub ilusaid valmislahendusi, pidi siiski nende *stylesheet*'i üsna palju CSS-i abil üle kirjutama. Kohati on modifikatsioone ka keeruline sisse viia, sest *stylesheet*'i CSS failist tuleb üles leida igale muudetavale elemendile vastavad klassinimed.

5.2.3 Qumanize rakenduse kasutajaliides

Valminud Qumanize rakenduse kasutajaliidese arhitektuuri on illustreeritud joonisel 6.



Joonis 6 Flaski veebirakenduse arhitektuur. The veebirakenduse aju on failis home.py, mis sisaldab Flaski @app.route dekoraatorit, kuhu laekub kliendi päring. See suunatakse sealt humanize funktsiooni, mis alustab humaniseerimise protsessi pendeldades järjestusi erinevate funktsioonide ja failide vahel. Kui humaniseerimisprotsess on lõpetatud suunatakse väljund veebilehele tagasi kasutades template.html malli. Kohandatud John Sobanski järgi [64]

Kuna Qumanize rakendus on mõeldud majasiseseks kasutamiseks, siis rõhuti veebiliidese disainis pigem lihtsusele ja arusaadavusele, ainukese värvinüansina kasutati veebiliidese kujunduses tooni #701010, mis on Icosagen Cell Factory „brändi“ üks ametlikke värve, mida leidub nii veebilehel kui reklaammaterjalidel. Vormide kujundus pärineb enamalt jaolt Bootstrap mallilt, mõned komponendid on käsitsi CSS-ga viimistletud. Selleks, et tagada hea jälgitavus sisestab klient järjestused (Joonis 7) ja saab humaniseeritud väljundi (Joonis 8) samal lehel. Tulemuse ootamise ajaks on plaanis lisada laadimisriba.

← → ↻ ⓘ Not secure | 0.0.0.0:5000/#

Icosagen Cell Factory

Qumanize

Insert variable sequences in FASTA format

```
>VH
QEQLLEESGGGLVQPEGSLTLTCKASGFSFSAIAMCWVRQAPGKLEWIGCIATDTGSTYYANWAKGRFTISNPSSTTVTLQMTSLTAADTATYFCARNFYLWGPGLTVVSS
```

How many most similar sequences to the source should be used from the database?

Threshold for germline identity score

Value 0.1 -1, lower value requires sequences to be more dissimilar to each-other

Create more candidate sequences (takes a long time)

Display only WHO compliant output sequences

OK!

Output

Your results will be displayed here

0.0.0.0:5000/#

Joonis 7 Qumanize veebiliidese avaleht koos kliendi sisestatud järjestusega *localhost* testkeskkonnas.

← → ↻ ⓘ Not secure | 0.0.0.0:5000/#

How many most similar sequences to the source should be used from the database?

Threshold for germline identity score

Value 0.1 -1, lower value requires sequences to be more dissimilar to each-other

Create more candidate sequences (takes a long time)
 Display only WHO compliant output sequences

Output

| ID | Humanized sequence | Parent organism | T20 Score | IMGT Score |
|------|---|-----------------|-----------|------------|
| 1VH | QVQLVESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRDNKNTLYLQMNSLRAEDTAVYY CARNFYLGQGTLTVSS | Homo sapiens | 90.76927 | 85.7 |
| 15VH | QVQLEESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRDNKNTLYLQMNSLRAEDTAVYY CARNFYLGQGTLTVSS | Homo sapiens | 90.76927 | 84.7 |
| 16VH | QVQLEESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRDNKNTLYLQMNSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 84.7 |
| 18VH | QVQLEESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRPNKNTLYLQMNSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 83.7 |
| 20VH | QVQLEESGGGLVQPGGSLRSLCKASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRPNKNTLYLQMNSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 82.7 |
| 90VH | QVQLEESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRDNKNTLYLQMNSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 83.7 |
| 91VH | QVQLEESGGGLVQPGGSLRSLCAASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRPNKNTLYLQMNSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 82.7 |
| 95VH | QVQLEESGGGLVQPGGSLRSLCKASGFSFSAIAMCWVRQAPGKGLEWVSCIATDTGSTYYANWAKGRFTISRPNKNTLYLQMTSLRAEDTAVYY CARNFYLGWPGTLTVSS | Homo sapiens | 87.30777 | 81.6 |

0.0.0.0:5000/#

Joonis 8 Qumanize rakenduse väljund, kus kuvatakse humaniseeritud järjestuse liik (algne järjestus oli antud juhul küüliku oma), T20 skoor ja IMGT skoor. WHO kriteeriumitele vastavaks loetakse humaniseeritud järjestust, mille IMGT skoor on üle 85%.

6. Kokkuvõte

Käesolev magistritöö tutvustab ühe biotehnoloogilise tarkvara loomist alustades rakenduse ideest, tööpõhimõttest ning jõudes selle testimiseni laboris. Magistritöö teoreetilise osa eesmärgiks oli tervisevaldkonna IT-lahendusi pakkuva ettevõtte Quretec poolt 2017-2018 aastatel loodud humaniseerimistarkvara Qumanize detailne analüüs. Tausta avamiseks kirjeldati monoklonaalsete antikehade struktuuri, avastamist ja terapeutilist tähtsust. Sellele järgnes humaniseerimise tähtsuse ja meetodika kirjeldus. Järgnevalt kirjeldati nii rakenduse loomist kui tööpõhimõtet. Biotehnoloogiafirmas Icosagen Cell Factory-s eelnevalt tehtud katsete põhjal analüüsiti ka rakenduse poolt tekitatud humaniseeritud antikehajärjestuste kommertsiaalset potentsiaali.

Kuna tegemist oli käsurearakendusega, siis püstitati ka praktiline eesmärk luua kasutajamugavuse suurendamiseks veebipõhine kasutajaliides. Selle jaoks kasutati Pythoni põhist veebiliidese raamistikku Flask. Magistritöös põhjendatakse antud valikut ning tutvustatakse loodud lahendust.

Kuna loodud kasutajaliides on pigem algne funktsioneeriv prototüüp mitte valmis produkt, siis käesolevas magistritöös käsitletud rakenduse edasiarendusena võiks ellu viia järgmised tegevused:

1. Selgitada kliendi tagasiside põhjal välja nõuded rakenduse lõplikule välimusele ning programmi väljundile.
2. Kasutajaliidesele proteoomikast ja immuunoinformaatikast pärit lisafunktsioonide lisamine vastavalt kliendi vajadusele.
3. Jõuda rakenduse poolt tekitatud humaniseeritud järjestuste analüüsiga ka immuunvastuse tekke testimiseni inimorganismis.
4. Rakenduse kommertsiaalne kasutuselevõtt biotehnoloogia firmas Icosagen Cell Factory.

7. Kasutatud kirjandus

- [1] OpenStax. The Adaptive Immune Response: B-lymphocytes and Antibodies. OpenStax CNX 2013. https://cnx.org/contents/AY5_7yUs@4/The-Adaptive-Immune-Response-B-lymphocytes-and-Antibodies (accessed March 1, 2020)
- [2] Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, et al. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science* (80-) 2001;293:1155–9. doi:10.2210/PDB1HZH/PDB.
- [3] Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* 2014;5:520. doi:10.3389/fimmu.2014.00520.
- [4] Safdari Y, Farajnia S, Asgharzadeh M, Khalili M. Antibody humanization methods – a review and update. *Biotechnol Genet Eng Rev* 2013;29:175–86. doi:10.1080/02648725.2013.801235.
- [5] Eales L-J. The Adaptive Immune Response. *Immunol. Life Sci.*, John Wiley & Sons, Ltd; 2005, p. 112–58. doi:10.1002/047086821x.ch4.
- [6] Chames P, Van Regenmortel M, Weiss E, Baty D. Therapeutic antibodies: successes, limitations and hopes for the future. *Br J Pharmacol* 2009;157:220–33. doi:10.1111/j.1476-5381.2009.00190.x.
- [7] Hafeez U, Gan HK, Scott AM. Monoclonal antibodies as immunomodulatory therapy against cancer and autoimmune diseases. *Curr Opin Pharmacol* 2018;41:114–21. doi:10.1016/j.coph.2018.05.010.
- [8] Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. *Nat Rev Cancer* 2012;12:278–87. doi:10.1038/nrc3236.
- [9] Gan HK, van den Bent M, Lassman AB, Reardon DA, Scott AM. Antibody–drug conjugates in glioblastoma therapy: the right drugs to the right cells. *Nat Rev Clin Oncol* 2017;14:695–707. doi:10.1038/nrclinonc.2017.95.
- [10] Lu Z-J, Deng S-J, Huang D-G, He Y, Lei M, Zhou L, et al. Frontier of therapeutic antibody discovery: The challenges and how to face them. *World J Biol Chem* 2012;3:187–96. doi:10.4331/wjbc.v3.i12.187.
- [11] Köhler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975;256:495–7.

- [12] Kivi G, Teesalu K, Parik J, Kontkar E, Ustav M, Noodla L, et al. HybriFree: a robust and rapid method for the development of monoclonal antibodies from different host species. *BMC Biotechnol* 2016;16:2. doi:10.1186/s12896-016-0232-6.
- [13] Hwang WYK, Foote J. Immunogenicity of engineered antibodies. *Methods* 2005;36:3–10. doi:10.1016/J.YMETH.2005.01.001.
- [14] Schellekens H. Immunogenicity of therapeutic proteins: Clinical implications and future prospects. *Clin Ther* 2002;24:1720–40. doi:10.1016/S0149-2918(02)80075-3.
- [15] Descotes J. Immunotoxicity of monoclonal antibodies. *MAbs* 2009;1:104–11. doi:10.4161/mabs.1.2.7909.
- [16] An Z. Therapeutic Monoclonal Antibodies: From Bench to Clinic. *Ther Monoclon Antibodies From Bench to Clin* 2009:1–889. doi:10.1002/9780470485408.
- [17] Kashmiri SVS, De Pascalis R, Gonzales NR, Schlom J. SDR grafting—a new approach to antibody humanization. *Methods* 2005;36:25–34. doi:10.1016/J.YMETH.2005.01.003.
- [18] Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 1970;132:211–50.
- [19] Johnson G, Wu T Te. Kabat Database and its applications: future directions. *Nucleic Acids Res* 2001;29:205–6. doi:10.1093/nar/29.1.205.
- [20] Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 1987;196:901–17. doi:10.1016/0022-2836(87)90412-8.
- [21] Abhinandan KR, Martin ACR. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. *Mol Immunol* 2008;45:3832–9. doi:10.1016/J.MOLIMM.2008.05.022.
- [22] Bujotzek A, Fuchs A, Qu C, Benz J, Klostermann S, Antes I, et al. MoFvAb: Modeling the Fv region of antibodies. *MAbs* 2015;7:838–52. doi:10.1080/19420862.2015.1068492.
- [23] Bujotzek A, Dunbar J, Lipsmeier F, Schäfer W, Antes I, Deane CM, et al. Prediction of VH-VL domain orientation for antibody variable domain modeling. *Proteins Struct*

- Funct Bioinforma 2015;83:681–95. doi:10.1002/prot.24756.
- [24] Giudicelli V, Chaume D, Bodmer J, Müller W, Busin C, Marsh S, et al. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res* 1997;25:206–11.
- [25] Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 2003;27:55–77. doi:10.1016/S0145-305X(02)00039-3.
- [26] Lefranc MP. Unique database numbering system for immunogenetic analysis. *Immunol Today* 1997;18:509.
- [27] Ehrenmann F, Kaas Q, Lefranc M-P. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res* 2010;38:D301–7. doi:10.1093/nar/gkp946.
- [28] WHO-IUS. WHO-IUIS Nomenclature SubCommittee for IG and TR (Report. 2007). http://www.imgt.org/IMGTindex/WHO_IUISncReport.pdf
- [29] Dondelinger M, Filée P, Sauvage E, Quinting B, Muyldermans S, Galleni M, et al. Understanding the Significance and Implications of Antibody Numbering and Antigen-Binding Surface/Residue Definition. *Front Immunol* 2018;9:2278. doi:10.3389/fimmu.2018.02278.
- [30] Li B, Tesar D, Boswell CA, Cahaya HS, Wong A, Zhang J, et al. Framework selection can influence pharmacokinetics of a humanized therapeutic antibody through differences in molecule charge 2014. doi:10.4161/mabs.29809.
- [31] Janeway CA, Travers P, Walport M, Shlomchik MJ. *Immunobiology: The Immune System in Health and Disease*. 5th edition. New York: Garland Science; 2001. The generation of diversity in immunoglobulins. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK27140/>
- [32] Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res* 2005;33:D256-61. doi:10.1093/nar/gki010.
- [33] Pelat T, Bedouelle H, Rees AR, Crennell SJ, Lefranc M-P, Thullier P. Germline

- Humanization of a Non-human Primate Antibody that Neutralizes the Anthrax Toxin, by in Vitro and in Silico Engineering. *J Mol Biol* 2008;384:1400–7.
doi:10.1016/J.JMB.2008.10.033.
- [34] Robert R, Streltsov VA, Newman J, Pearce LA, Wark KL, Dolezal O. Germline humanization of a murine Abeta antibody and crystal structure of the humanized recombinant Fab fragment. *Protein Sci* 2010;19:299–308. doi:10.1002/pro.312.
- [35] Fusion Antibodies. CASE STUDY: Fsn1006 Antibody Humanization, n.d.
<https://www.fusionantibodies.com/news/fsn1006-an-antibody-with-dual-specificity/>
(accessed April 10, 2019).
- [36] Home - Protein - NCBI n.d. <https://www.ncbi.nlm.nih.gov/protein> (accessed April 10, 2019).
- [37] UniProt n.d. <https://www.uniprot.org/> (accessed April 10, 2019).
- [38] RCSB PDB: Homepage n.d. <https://www.rcsb.org/> (accessed April 10, 2019).
- [39] Martin ACR, Rees AR. Extracting human antibody sequences from public databases for antibody humanization: high frequency of species assignment errors (Short Communication). *Protein Engineering, Design and Selection* 2016.;29;10:403–408,
<https://doi.org/10.1093/protein/gzw018>
- [40] TABS. TABS Therapeutic Antibody Database n.d. http://tabs.craic.com/users/sign_in
(accessed April 9, 2019).
- [41] Apgar JR, Mader M, Agostinelli R, Benard S, Bialek P, Johnson M, et al. Beyond CDR-grafting: Structure-guided humanization of framework and CDR regions of an anti-myostatin antibody. *MAbs* 2016;8:1302–18.
doi:10.1080/19420862.2016.1215786.
- [42] Hansel TT, Kropshofer H, Singer T, Mitchell JA, George AJT. The safety and side effects of monoclonal antibodies. *Nat Rev Drug Discov* 2010;9:325–38.
doi:10.1038/nrd3003.
- [43] Jones TD, Carter PJ, Plückthun A, Vásquez M, Holgate RGE, Hötzel I, et al. The INNs and outs of antibody nonproprietary names. *MAbs* 2016;8:1–9.
doi:10.1080/19420862.2015.1114320.

- [44] WHO-INN. International Nonproprietary Names (INN) for biological and biotechnological substances (a review). 2016.
https://www.who.int/medicines/services/inn/inn_bio_inn/en/ (accessed April 9, 2019).
- [45] WHO-INN. Executive Summary Programme on International Nonproprietary Names (INN) Technologies Standards and Norms Regulation of Medicines and other Health Technologies (RHT) Essential Medicines and Health Products (EMP) World Health Organization, Geneva 62 nd Cons. 2016.
https://www.who.int/medicines/services/inn/62nd_Executive_Summary.pdf?ua=1 (accessed April 9, 2019)
- [46] Gao SH, Huang K, Tu H, Adler AS. Monoclonal antibody humanness score and its applications. BMC Biotechnol 2013;13:55. doi:10.1186/1472-6750-13-55.
- [47] IcoCell | Icosagen n.d. <https://www.icosagen.com/icocell> (accessed May 10, 2020).
- [48] Olimpieri PP, Marcatili P, Tramontano A. Tabhu: tools for antibody humanization. Bioinformatics 2015;31:434–5. doi:10.1093/bioinformatics/btu667.
- [49] GenBank Overview n.d. <https://www.ncbi.nlm.nih.gov/genbank/> (accessed February 13, 2019).
- [50] Fitch WM, Dayhoff MO. Atlas of Protein Sequence and Structure, 1972. vol. 22. 1973. doi:10.2307/2412105.
- [51] Ekmekci B, McAnany CE, Mura C. An Introduction to Programming for Bioscientists: A Python-Based Primer. PLoS Comput Biol 2016;12:e1004867. doi:10.1371/journal.pcbi.1004867.
- [52] Chang J, Chapman B, Friedberg I, Hamelryck T, De Hoon M, Cock P, et al. Biopython Tutorial and Cookbook. n.d. <http://biopython.org/DIST/docs/tutorial/Tutorial.html> (accessed April 17, 2020)
- [53] Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. Bioinformatics 2015;btv552. doi:10.1093/bioinformatics/btv552.
- [54] Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform 2009;23:205–11. doi:10.1142/9781848165632_0019.
- [55] Docker Desktop for Mac and Windows | Docker n.d.

- <https://www.docker.com/products/docker-desktop> (accessed May 12, 2020).
- [56] Werkzeug — Werkzeug Documentation (1.0.x) n.d.
<https://werkzeug.palletsprojects.com/en/1.0.x/> (accessed May 4, 2020).
- [57] Jinja | The Pallets Projects n.d. <https://palletsprojects.com/p/jinja/> (accessed April 17, 2020).
- [58] Grinberg M. Flask Web Development: Developing Web Applications with Python. 2018.
- [59] Python Developers Survey 2018 Results n.d.
<https://www.jetbrains.com/research/python-developers-survey-2018/> (accessed May 3, 2020).
- [60] Django vs. Flask in 2019: Which Framework to Choose | TestDriven.io n.d.
<https://testdriven.io/blog/django-vs-flask/> (accessed May 3, 2020).
- [61] Flask-WTF — Flask-WTF 0.14 n.d. <https://flask-wtf.readthedocs.io/en/stable/> (accessed May 4, 2020).
- [62] Bootstrap · The most popular HTML, CSS, and JS library in the world. n.d.
<https://getbootstrap.com/> (accessed March 1, 2020).
- [63] Flask-Bootstrap — Flask-Bootstrap 3.3.7.1 documentation n.d.
<https://pythonhosted.org/Flask-Bootstrap/> (accessed March 1, 2020).
- [64] Pass Bootstrap HTML attributes to Flask-WTForms n.d. <https://john.soban.ski/pass-bootstrap-html-attributes-to-flask-wtforms.html> (accessed May 5, 2020).

Tänuõnad

Tänu Ahtole ja Priidule abi ja kannatlikkuse eest.

LISA 1: LIHTLITSENTS MAGISTRITÖÖLE

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Katrin Ruisu,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Qumanize rakenduse analüüs ja sellele veebiliidese loomine“, mille juhendajad on Priit Adler ja Ahto Salumets, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Katrin Ruisu

14.05.2020

LISA2: QUMANIZE KÄSUREARAKENDUSE KASUTUSJUHEND

Programmi info:

```
Qumanize_wolfguy.py [-h] [-t TOP] [--long]
                    [--ref_germline_thr REF_GERMLINE_THR]
                    [--dissimilarity_index DISSIMILARITY_INDEX]
                    [--who_compliant]
                    source_fn db_fn
```

Python program to humanise foreign antibody variable region sequence.

positional arguments:

```
source_fn          file with source sequence to humanise (in fasta
                    format)
db_fn              human ab sequences in anarci(wolfguy) format to be
                    used for template
```

optional arguments:

```
-h, --help          show this help message and exit
-t TOP, --top TOP   how many most similar sequences to source to be used
                    (default: 10)
--long              more creative way how to combine similar sequences,
                    generates a *lot* more candidate sequences, takes a
                    long time
--ref_germline_thr REF_GERMLINE_THR
                    reference sequences that have germline identity score
                    less that this threshold will be omitted (default:
                    0.95)
--dissimilarity_index DISSIMILARITY_INDEX
                    framework dissimilarity index for reference sequence
                    set, lower value will require sequences in PFM to be
                    more dissimilar to each other (default: 0.9)
--who_compliant     only output sequences that are WHO compliant, i.e. in
                    IMGT domainGapAlign first organism is human and
                    identity score is > 85
```

Uue järjestusefaili lisamine

- Kasutada tekstiredaktorit nano (käsureale kirjutada nano ja vajutada Enter)
- Paste-da fasta formaadis oma järjestused redaktorisse (paremkõps)

Näiteks:

>VH

```
AVTLDES GGG LQTPGGALSLVCKGSGFTFSSYS LGWVRQAPGK GLEYVAAIDNTG SWTNYGAVVKGRATISRDN GQSTVRLQL  
NNLRADDTATYYCAKAAGIGH SYTYGDDIDAWGHGTEVIVSS
```

>VL

```
ALTQPSSV SANLGGTVEITCSGGNYYGWYQQKSPGSAPVTLIYNNDKRPSDIPSRFSGSKSGSTATLTITGVQVEDEAVYFC  
GIPDNTYIAIFGAGTTLTVL
```

- Seejärel väljumiseks vajutada Cntrl + w
- Küsib, kas salvestada? Vajutada y
- Kirjutada faili nimi, näiteks mouse.fa + Enter

Programmi jooksutamine

```
Qumanize_wolfguy.py mouse.fa /opt/icosagen-  
gumanize/data/human_db_uniq_germline.anarci
```

source_fn (tekitatud fasta formaadis fail looma variaablitega)

db_fn (inimese järjestuste fail anarci formaadis – antud aadressil Priidu koostatud andmebaas)

lõppu võib lisada kärke valukust “optional arguments”

Näiteks:

```
Qumanize_wolfguy.py mouse.fa /opt/icosagen-  
gumanize/data/human_db_uniq_germline.anarci -t 10 --who_compliant
```

Kui programm on lõpuni jooksnud, siis ta tekitab eraldi VH ja VL faili sinna samma kausta.

Kausta sisu vaatamiseks võib kasutada käsku ls.

Faili sisu vaatamiseks kasutada käsku cat + faili nimi. Faili nime on käsureal lihtsam kirjutada kasutades Tab klahvi (otsib automaatselt sarnase algusega faile)

Väljund

```
>88VH attr:5.7|7.0|0.0|14|76.84|Homo sapiens(85.6)|0.066  
AVQLDES GGG LVPGGSLRLS CAASGFTFSSYS LGWVRQAPGK GLEYVSAIDNTG SWTNY  
GAVVKGRFTISRDN SKNTLRLQMN SLRAEDTAVYYCAKAAGIGH SYTYGDDIDAWGQGTE
```

T20 skoor (Peaks olema üle 85) ja IMGT skoor (peaks olema üle 85)

Kasulikud käsud käsureal

| | |
|---------------|---|
| ls | näitab kausta sisu |
| cat failinimi | näitab vastava faili sisu |
| nano | tekstiredaktor, kus saab tekitada faile |