Sharif University of Technology
Department of Computer Engineering
June 2023

Hadis Ahmadian
400211524
Structrul Bioinformatics Assignment #2

همراه با این گزارش چند فایل ضمیمه شده است که commands.tcl کدهایی است که در ترمینال VMD زده شده و p1.tcl p2.tcl, و p3.tcl درواقع precdureهای خواسته شده هستند.

فایل 1k8h.pdb نیز فایل pdb پروتیین انتخاب شده است.

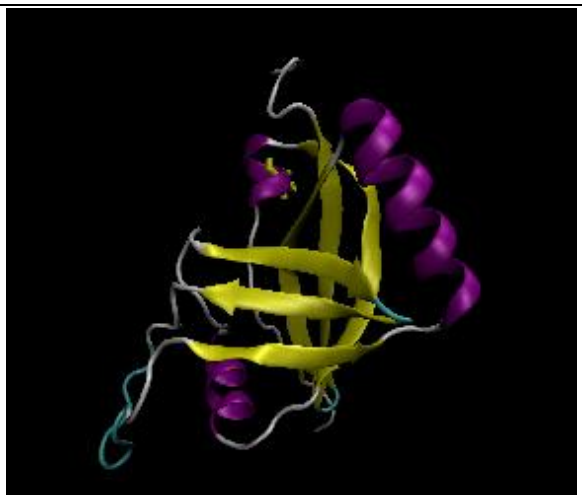
سه فایل txt نیز خروجی precedureها هستند و در ادامه توضیح داده میشوند.

فایل plot.py برای یخش رسم نمودار.

توضیح روند کار و کدهای فایل ها:

در ابتدا با کد زیر ملکول لود شده و نمایش آن به حالت new carton با رنگ آمیزی طبق ساختار در میاید.

```
#change directory
cd prots
#load model
mol load pdb 1k8h.pdb
#for better view
set id 0
mol modstyle 0 $id NewCartoon 0.300000 10.000000 4.100000 0
mol modcolor 0 $id Structure
```



توالی اسید های آمینه و تعداد هر نوع اسید آمینه (۲۰ نوع اسید آمینه) در آن پروتئین محاسبه گردد.

به این منظور قطعه کد زیر زده شده:

```
 #showing length of sequence
[atomselect $id "name CA"] num
 #showing sequence of AAs
[atomselect $id "name CA"] get {resname}
```

خروجی: 133

خروجی :

```
GLY LYS SER ASP LYS ILE ILE PRO ILE ALA GLU ASN LYS GLU ALA LYS ALA
LYS TYR ASP ILE LEU GLU THR TYR GLU ALA GLY ILE VAL LEU LYS GLY SER GLU
VAL LYS SER LEU ARG GLU LYS GLY THR VAL SER PHE LYS ASP SER PHE VAL ARG
ILE GLU ASN GLY GLU ALA TRP LEU TYR ASN LEU TYR ILE ALA PRO TYR LYS HIS
ALA THR ILE GLU ASN HIS ASP PRO LEU ARG LYS ARG LYS LEU LEU LEU HIS LYS
ARG GLU ILE MET ARG LEU TYR GLY LYS VAL GLN GLU LYS GLY TYR THR ILE ILE
PRO LEU LYS LEU TYR TRP LYS ASN ASN LYS VAL LYS VAL LEU ILE ALA LEU ALA
                                         LYS GLY LYS LYS LEU TYR ASP ARG
```

در بخش زیر به ازای هر نام آمینواسید تعداد آن شمرده شده و در خروجی نمایش داده میشود

```
set names { ALA ARG ASN ASP CYS GLU GLN GLY HIS ILE LEU LYS MET PHE PRO
SER THR TRP TYR VAL }
foreach name $names { puts "$name":
puts [[atomselect $id "name CA and resname $name" frame $a] num]}
```

خروجی:

```
ALA : 9 ARG : 7 ASN : 6 ASP : 5 CYS : 0 GLU : 11 GLN : 1 GLY : 8 HIS :
3 ILE : 12 LEU : 15 LYS : 22 MET : 1 PHE : 2 PRO : 4 SER : 5 THR : 4 TRP
                                                        : 2 TYR : 9 VAL : 7
```

یک procedure جدید طراحی شود که توالی ساختار دوم مدل (فریم) های مختلف پروتئین در فایلی به صورت خروجی نوشته شود و درصد انواع ساختارهای دوم مدل ها نیز محاسبه شود.

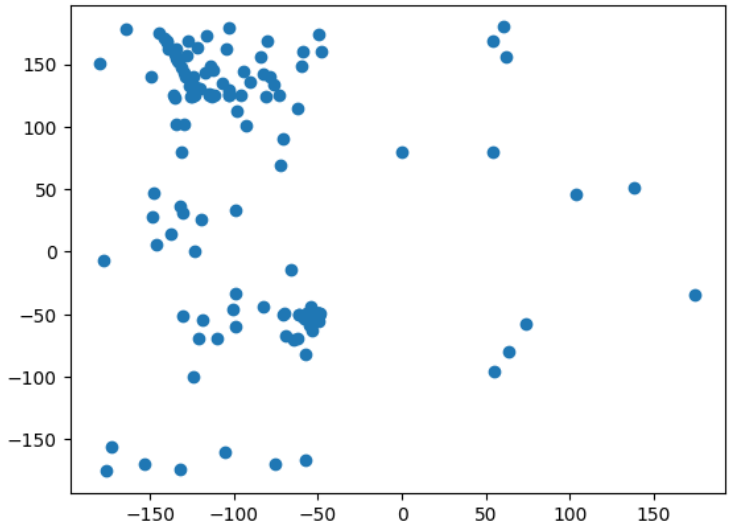
این procedure در فایل p1.tcl نوشته شده است و نام آن frame_struct است که در id یک ملکول لود شده را میگیرد و در یک حلقه به تداد فریم ها (به ازای هر فریم) ابتدا ملکول هایی با نام CA را سلکت میکند (درواقع کربن آلفا رزیدوهارا) و ساختار رزیدوی اتم های سلکت شده را چاپ کرده و سپس به ازای هر نوع ساختار دوم، شمارش میکند چنتا از آن وجود دارد و نسبت آن به کل رزیدو ها را در فایل مینوسید. فایل خروجی این procedure ضمیمه شده و frame_struct_info.txt نام دارد که به دلیل طولانی بودن کد و خروجی در گزارش نوشته نشده و لطفا فایل ها را بررسی کنید.

زوایای phi و psi برای یک مدل پروتئین محاسبه گردد و نمودار راماچاندران را به وسیله هر زبانی که مسلط هستید، رسم کنید.

با کد زیر زوایای فای و سای استخراج شده و سپس خروجی این کد مستقیما به کد پایتون plot.py داده شده

```
[atomselect $id "name CA"] get {phi psi}
```

کد پایتون در فایل plot.py ضمیمه شده که درواقع زوایای فای و سای را دریافت کرده و نمودار را رسم کرداه است. خروجی کد پایتون :



یک procedure جدید طراحی گردد که فایل pdb را load کرده و ماتریس RMSD دو به دو همه- ی فریم های یک پروتئین را حساب کرده و در فایل خروجی بنویسد.

این procedure در فایل p2.tcl نوشته شده و نام آن RMSD-All است که id ملکلول لود شده را از ورودی میگیرد و به کمک حلقه های تو در تو به وسیله ی mesure rmsd فاصله ی rmsd دو به دوی فریم ها را محاسبه میکند. خروجی در فایل RMSD-All-info.txt نوشته شده است. به دلیل طولانی بودن کد و خروجی در گزارش نوشته نشده و لطفا فایل ها را بررسی کنید.

یک procedure جدیدی طراحی نمایید که برای اتم های درگیر در پیوند هیدروژنی، شماره اسید آمینه، نام اسیدآمینه، نوع ساختار دوم آن اسید آمینه، زوایای دو وجهیphi (dihedral angles) ، psi، omegaو مقدار Radius of gyration را بعنوان خروجی بدهد.

این procedure در فایل p3.tcl نوشته شده و نام آن H_bond است و id ملکلول لود شده را به عنوان ورودی میگیرد و ابتدا با measure hbonds ایندکس دونور، اکسپتور و اتم های هیدروژن دخیل در پیوند هیدورژنی را میابد. در ادامه یک بار به ازای دوتور ها یک بار برای اکسپتور ها و یک بار هم برای اتم های هیدروژن نام ، نام و آیدی رزیدو و سایر موارد خواسته شده را در خروجی چاپ میکند. خروجی این percidure در فایل H-bond-info.txt نوشته شده است. به دلیل طولانی بودن کد و خروجی در گزارش نوشته نشده و لطفا فایل ها را بررسی کنید.