תרגיל 3 קורס 76562

רינה קרנאוך ואופק קוה

סמסטר אביב 2021

רוזטה

1. קיבלנו:

```
protocols.antibody.grafting: H1 detected: GYTFTDYHIN (10 residues at positions: 25 to 35)
protocols.antibody.grafting: H3 detected: AGLHPTTTEYYYYGMDV (17 residues at positions: 98 to 115)
protocols.antibody.grafting: H2 detected: WIHPNSGDTNYAQKFQG (17 residues at positions: 49 to 66)
```

על כן נקבל ש –

$$l(H_1) = 10, l(H_2) = 17, l(H_3) = 17$$

$$p(H_1) = [25,35], \qquad p(H_2) = [98,115], \qquad p(H_3) = [49,66]$$

בנוסף למיקומים הנתונים.

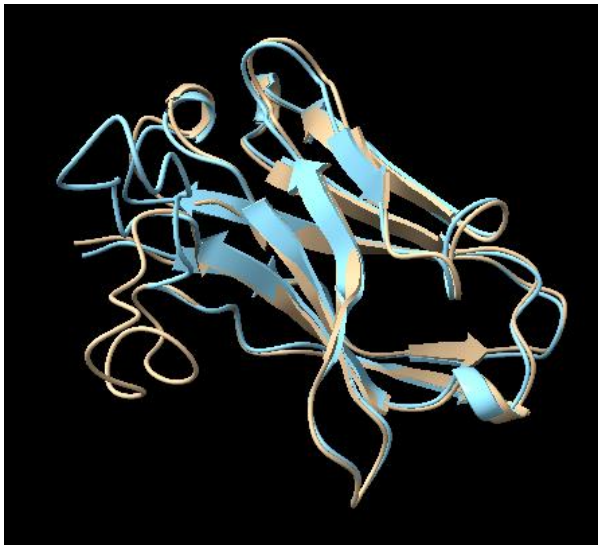
2. נכין טבלה ונדגום בעזרת cat filename.align את מספר הhits שרשום בתחילת ההדפסות:

hits	name
1024	FrH.align
798	H1.align
767	H2.align
16	H3.align

3. הסקור הכולל שקיבלנו הוא 296.536

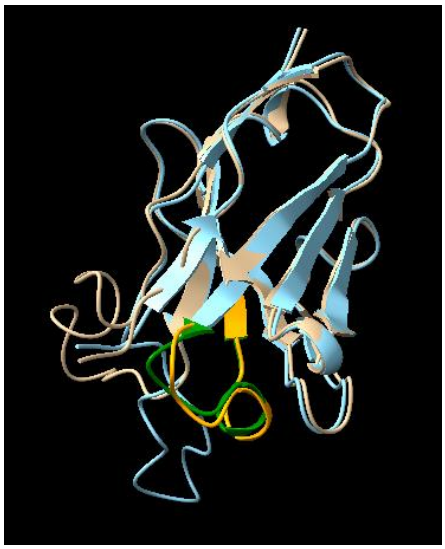
```
total_score
   -296.536
```

4. לאחר $Align$:



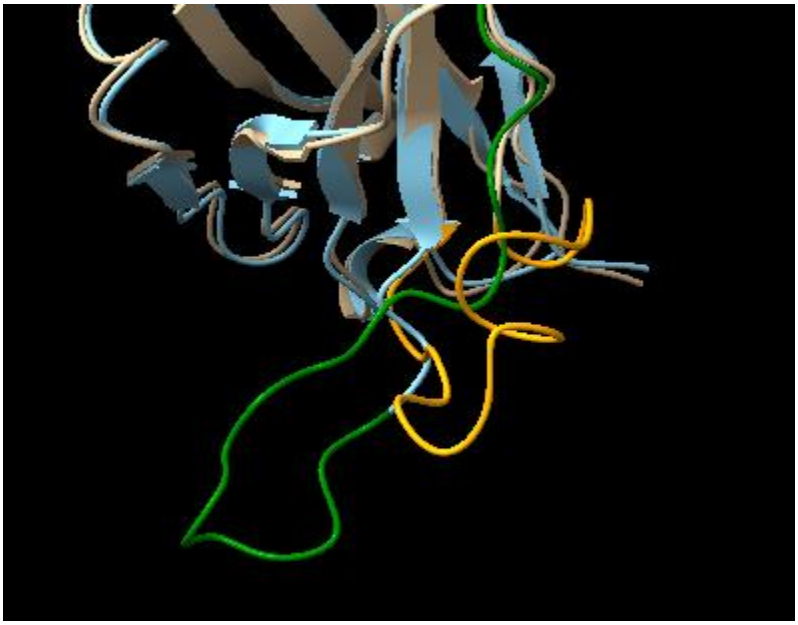
כעת נתייחס לדומה ולשונה בין המודלים בהתאם למיקומים משאלה 1.
נשים לב שכתום צבוע במודל שלנו, וירוק צבוע במודל הref.

עבור $H1$:



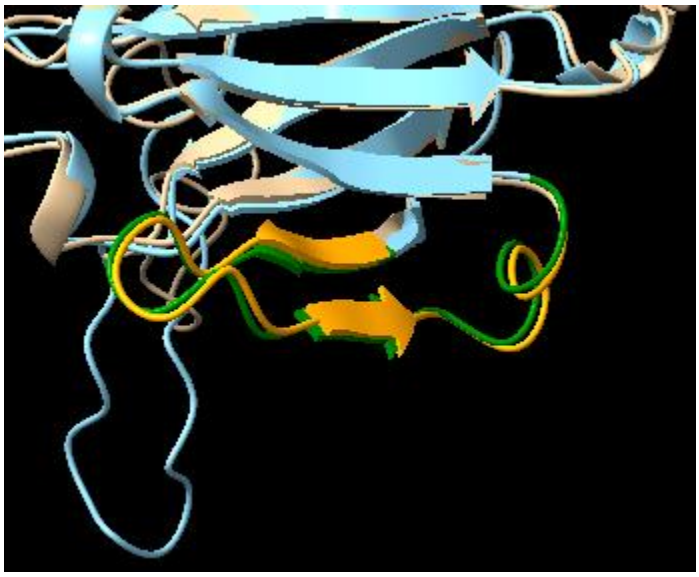
נראית התאמה טובה יחסית אך לא מלאה, כלומר קיימת לולאה ו*beta sheet* שנראות יחסית תואמות אך קיים חלק שונה מאוד.

עבור $H2$:



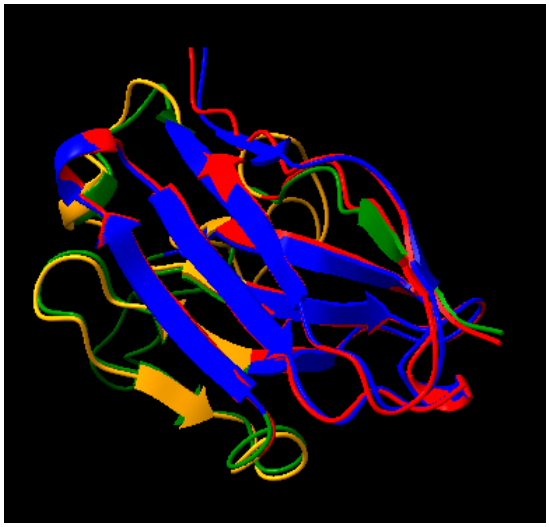
נראית התאמה פחות טובה מהקודמת, כלומר קיימת לולאה ו*beta sheet* שנראות יחסית תואמות אך קיים חלק ארוך יותר ששונה.

עבור $H3$:



נראית כמו ההתאמה הטובה ביותר מבין ההתאמות שכבר עשינו, לא נראה הבדל משמעותי.

עבור Fr:

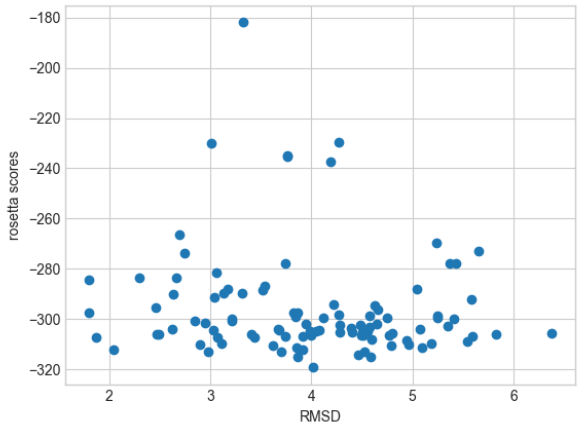


סימנו באדום את הFr במודל שלנו, ובכחול בRef, נין לראות התאמה דיי מושלמת ללא חריגות גדולות בין שני החלקים הנ"ל.

ה$RMSD$ הנ"ל יהיה 0.556:

Matchmaker ref.pdb, chain H (#1) with renumb.pdb, chain H (#2), sequence alignment score = 647.8
RMSD between 109 pruned atom pairs is 0.556 angstroms; (across all 125 pairs: 5.032)

5. קיבלנו את הגרף הבא:



נדמה שערכי פונקציית האנרגיה שנקבל מהרוזטה יקבלו ערכים בסביבות -300, ונצפה לכאלו עם RMSD נמוך כדי שיהיו יציבות.

.6

b. נגרום לcsv להציג לנו את 10 המודלים עם הrmsd הנמוך ביותר:

1	description	total_score	rmsd
2	model-0.relaxed_0031	-319.291	4.016
3	model-0.relaxed_0003	-315.203	3.868
4	model-0.relaxed_0044	-315.103	4.586
5	model-0.relaxed_0046	-314.291	4.465
6	model-0.relaxed_0058	-313.182	2.979
7	model-0.relaxed_0013	-313.104	3.699
8	model-0.relaxed_0002	-312.858	4.529
9	model-0.relaxed_0004	-312.368	2.046
10	model-0.relaxed_0063	-312.242	3.915
11	model-0.relaxed_0048	-311.368	5.098
12	model-0.relaxed_0083	-311.216	3.856
13	model-0.relaxed_0017	-310.534	3.626
14	model-0.relaxed_0047	-310.511	4.788
15	model-0.relaxed_0064	-310.191	4.968
16	model-0.relaxed_0050	-309.956	2.895

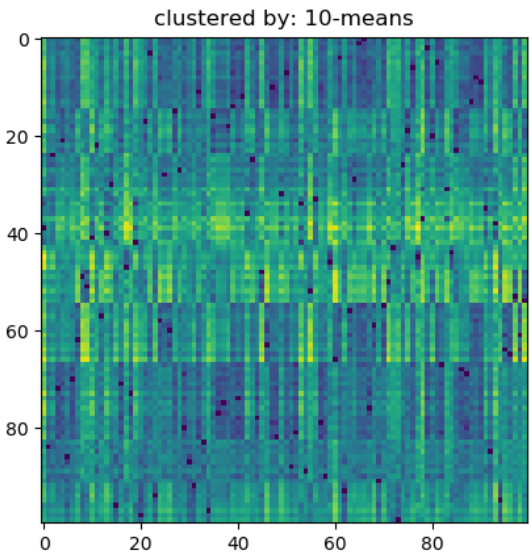
נרשום בטבלה את המידע על המודל עם הrmsd הכי נמוך מבין n המודלים עם האנרגיה המינימלית.

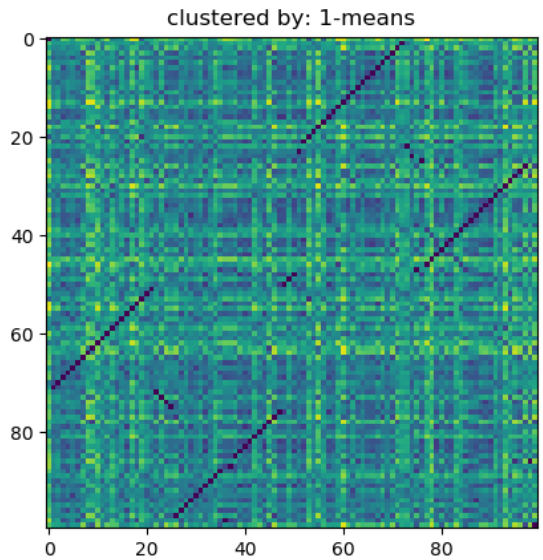
n	$Model$	$Energy$	$RMSD$
1	31	-319.291	4.016
5	58	-313.182	2.979
10	4	-312.368	2.046

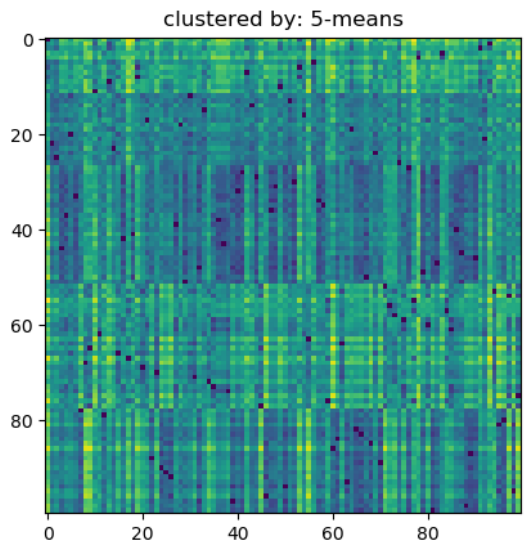
c. יצרנו קוד שיריץ על כל הקבצי pdb ויצור מטריצה עם rmsd, כך ש :

$$matrix[i][j] = rmsd\ of\ model\#i, model\#j$$

7. נרצה ליצור heatmap על פי המטריצה שיצרנו, כאשר אנחנו מבצעים על פי שורות. נקבל:





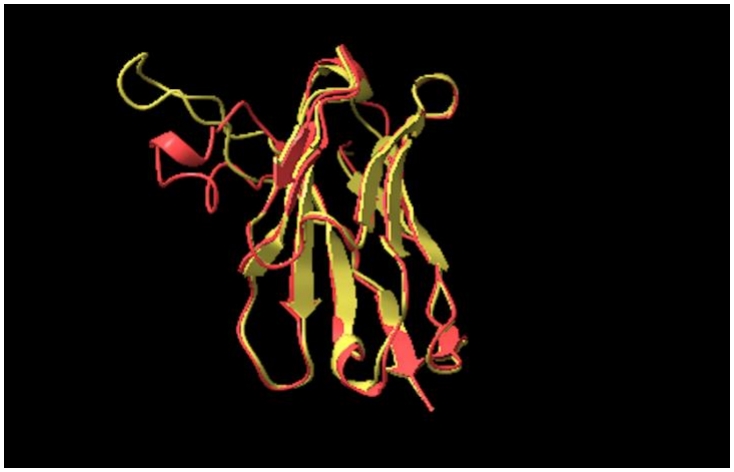


8. לאחר הרצת הקוד קיבלנו:

```
for 1 clusters, minimal is:31 with rmsd, score of:4.016, -319.291
for 5 clusters, minimal is:50 with rmsd, score of:2.895, -309.956
for 10 clusters, minimal is:58 with rmsd, score of:2.979, -313.182
```

9. כעת נבצע align של $ref.pdb$ ושל $model-0.relaxed_0058.pdb$ היות והוא נבחר להיות הטוב ביותר בתנאי 8 עבור n=10, נקבל:

- 58:



ונקבל:

RMSD between 112 pruned atom pairs is 0.559 angstroms; (across all 125 pairs: 3.200)

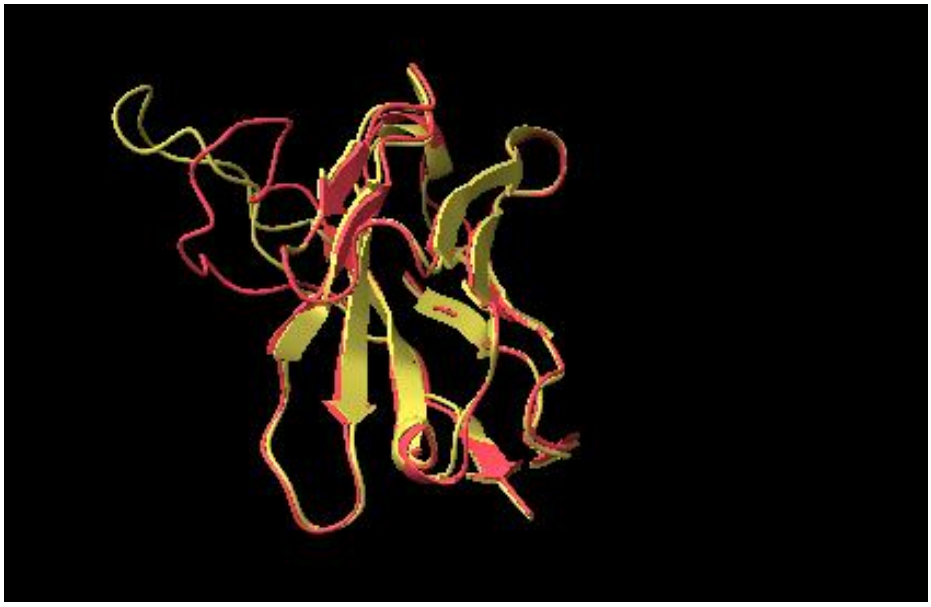
בנוסף נבצע $align$ עם $model-0.relaxed_{0050}.pdb$ ועם $model-0.relaxed_{0004}.pdb$

- 4:

ונקבל:

RMSD between 112 pruned atom pairs is 0.558 angstroms; (across all 125 pairs: 2.287)

- 50:



ונקבל:

RMSD between 111 pruned atom pairs is 0.568 angstroms; (across all 125 pairs: 3.051)

10. עבור n=1 תפסנו את אותם המודלים.
נבחין כי עבור n=5,10 אנחנו תלויים באלגוריתם kmeans שמפלג אותנו לn מרכזים עיקריים(הקלאסטרים שלנו), כאשר אנחנו דוגמים מהם את הטוב ביותר עם score ואז rmsd.
נבחין כי באופן שבו אנחנו בוחרים ערך בעל אנרגיה נמוכה ביותר מקלאסטר מסוים תלוי גם בערכים שנמצאים איתו בקלאסטר, על כן יכול להיות שנבחר ערך טוב יותר מבחינה אנרגטית מקלאסטר יחיד בעל ערך אנרגטי נמוך בקלאסטר זה, אך הוא יגרור בחירה לא טובה מבחינת rmsd בהשוואה עבור n הערכים שבחרנו.

נציג טבלת השוואה לתוצאות בעמוד הבא.

לפני הקליסטור קיבלנו:

n	$Model$	$Energy$	$RMSD$
1	31	-319.291	4.016
5	58	-313.182	2.979
10	4	-312.368	2.046

כאשר איתו קיבלנו:

$clusters = n$	$Model$	$Energy$	$RMSD$
1	31	-319.291	4.016
5	50	-309.956	2.895
10	58	-313.182	2.979

כמובן שיש פה עניין של טריידאוף בין rmsd לenergy, מבחינה אנרגטית נדמה שהקלסטור עזר באופן מינורי, אך לא מבחינת rmsd, אבל כמובן שכמו שראינו בגרף בשאלה 5 קשה לנו למצוא "מינימום מוחלט" עבור שני הערכים.