3D Data Processing in Structural Biology

תרגיל 1

מגישים:

- בר מלינרסקי – ת"ז 318189982
- רחל בן המוזג - ת״ז 300880143

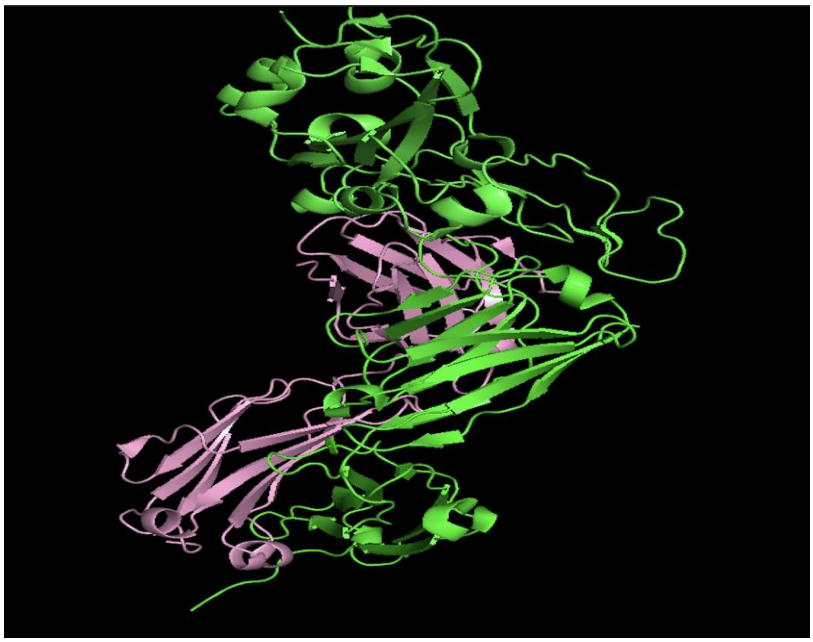
1) בשני המבנים נעשה שימוש בשיטה של : X-RAY DIFFRACTION

הרזולוציה במבנה L5B7 היא 3.18Å ואילו במנה KFX7 היא 2.23Å.

באופן כללי נראה שהרזולוציה שלהם לא גבוהה במיוחד, של הראשון אפילו נמוכה. שכן לפי מה שבדקנו מקובל להחשיב מבנים עם רזולוציה של עד1Å כמבנים עם רזולוציה גבוהה ואילו מבנים עם רזולוציה של 3Å ומעלה כמבנים עם רזולוציה נמוכה.

2)

heavy chain: chain id: H, MOL_ID: 2;
light chain: chain id: L, MOL_ID: 3;

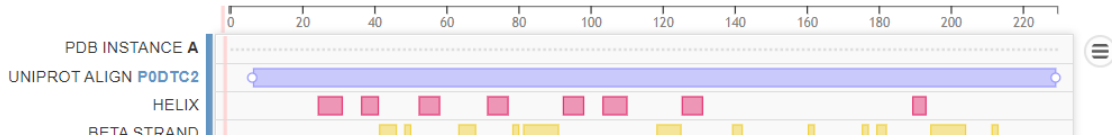


3) בשרשרת A יש 8 alpha-helices

Structural basis for a germline-biased antibody response to SARS-CoV-2 (RBD:C1A-C2 Fab)

INSTANCE

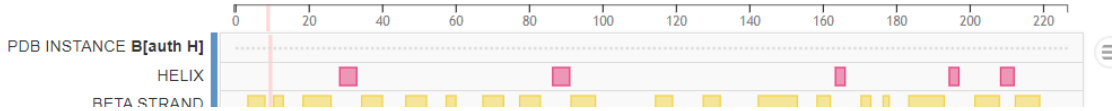
A | Spike glycoprotein - Severe acute respiratory syndrome coronavirus 2 | Help



בשרשרת H יש 5

INSTANCE

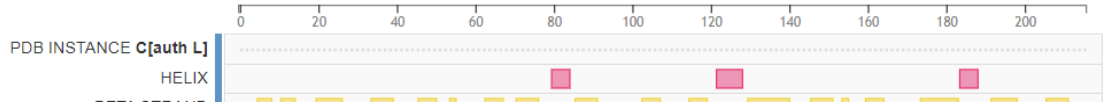
B [auth H] | heavy chain of human antibody C1A-C2 Fab - Homo sapiens | Help



בשרשרת L ש 3

INSTANCE

C [auth L] | light chain of human antibody C1A-C2 Fab - Homo sapiens | Help



4) גילו את המבנה ב-פברואר 2021

7L5B

Display Files | Download Files

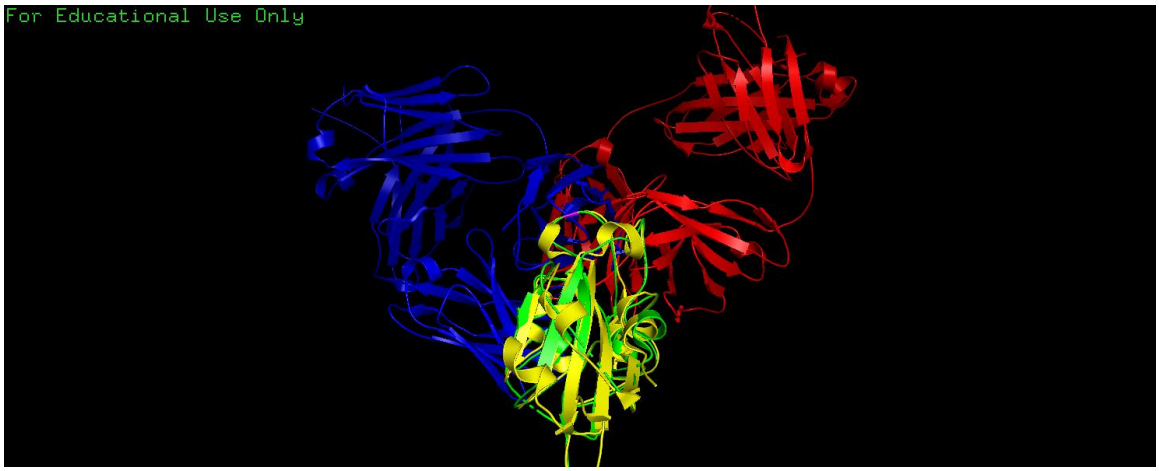
Crystallographic structure of neutralizing antibody 2-15 in complex with SARS-CoV-2 spike receptor-binding Domain (RBD).

Changes made to a PDB entry after its initial release are considered to be either "major" or "minor". The latest minor version of each major version is available as a file download. More information about the PDB versioning is available.

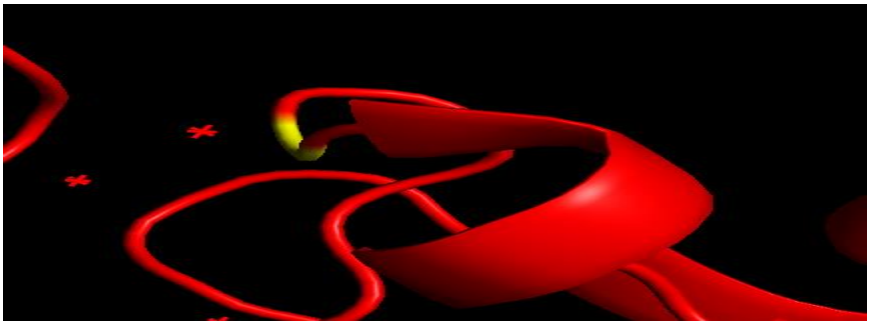
Version Number	Version Date	Version Type/Reason	Version Change	Revised CIF Category	
1.0	2021-02-10	Initial release			Download

5)

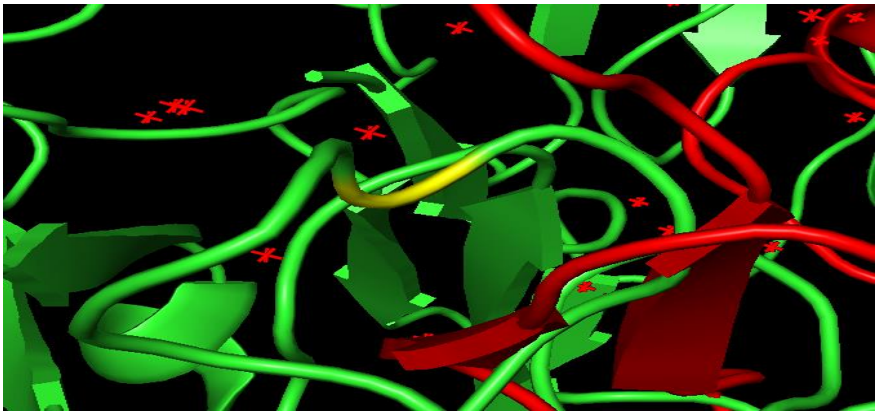
6) לדעתנו הם לא יוכלו להיקשר סימונטלית לאותו RBD שכן ניתן לראות שה-spike שלהם דומה, כלומר הם נלחמים על אותו צורת תקשורת ולכן יפריעו זה לזה.



7) ניתן לראות שב chain A של KFX7, המוטציה נמצאת בחלק המשמעותי, החלק לצידי Ca ולכן ישפיע יותר.



לעומת b715

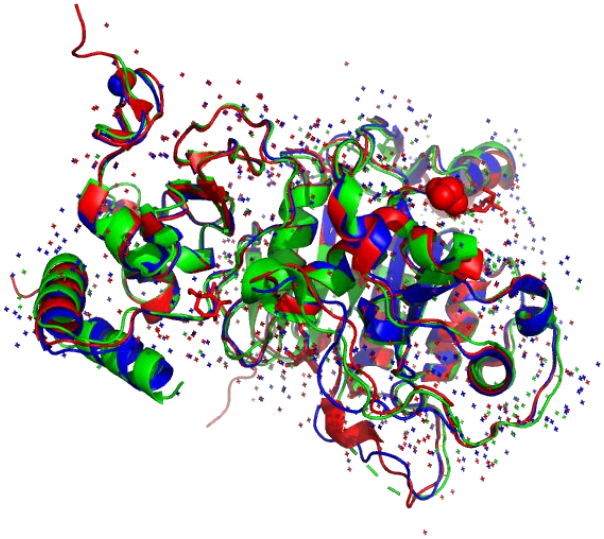


8) ערכנו alignment בין 3 המבנים יחדיו

5YN5 - MERS CoV - Colored green

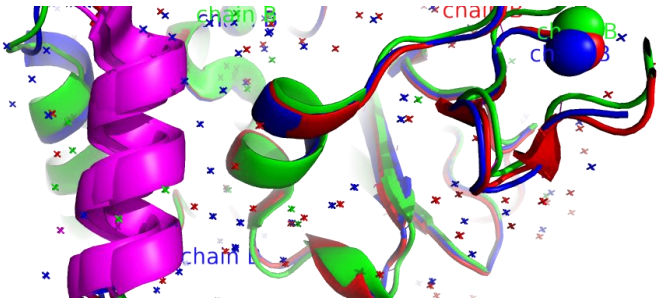
3R24- SARS 2002 - Colored blue

6W4H - SARS CoV 2 - Colored red

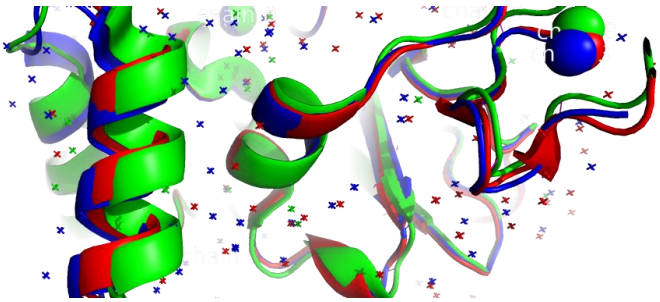


(9<u>דמיון בין 3 המבנים -</u>

1. לשלושתם יש את הסליל הזה שצבענו פה ב-magenta:

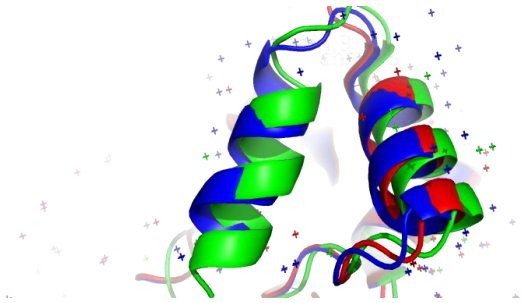


נצרף תמונה נוספת תוך התמקדות על הסליל אך לפני הצביעה שלו, אפשר לראות את 3 המבנים משתלבים יחד בצורה כמעט מושלמת בסליל הזה:

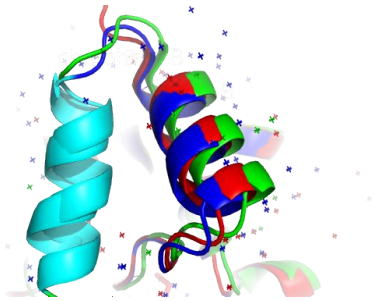


<u>שוני בין המבנים -</u>

1. לצד הסליל המשותף ישנו סליל נוסף שבו ניתן לראות בברור ש-SARS CoV 2 לא לוקח בו חלק אלא רק SARS ו2002 MERS

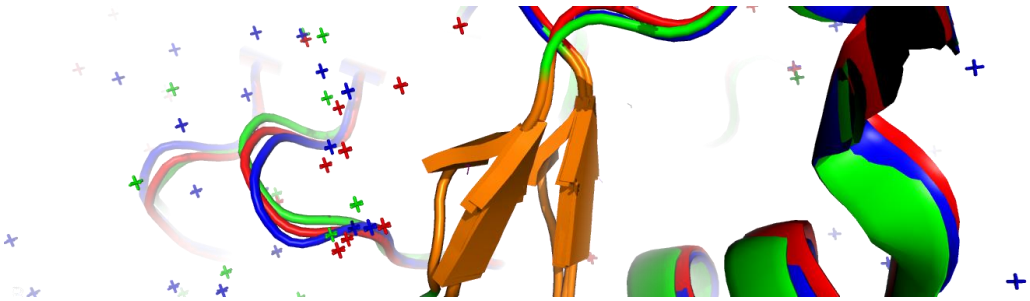


פה צבענו אותו בתכלת לצורך הדגשה :

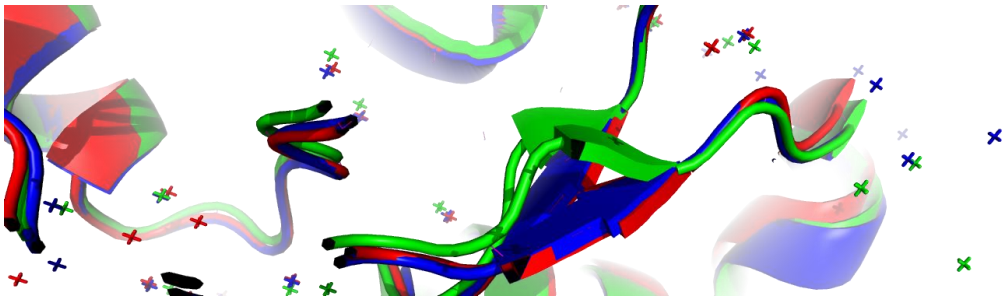


2. ניתן לראות שהמבנים יוצרים יחדיו צורה של חץ כאשר SARS CoV-2 ו-SARS 2002 צמודים זה לזה ואילו MERS נמצא בזווית אליהם

פה סימנו את כל האזור המדובר בכתום -



פה זה בלי הצביעה של האזור כדי שיהיה אפשר לראות את השוני



ב) לכל אטום 3 קורדינטות המייצגות את המיקום שלו על 3 הצירים. כאשר נתון לנו 2 סטים של נקודות המייצגים בהתאמה את האטומים של כל מבנה. אז RMSD שממחושב ע"י הנוסחה הבאה:

Let a molecule be defined by N atoms at positions $A = \{\mathbf{a}_i\}_N$ with coordinates $\mathbf{a}_i = \{x_i, y_i, z_i\}^T$ and associated weights $w = \{w_i\}_N$. Given two sets of N points, A and A', of respective coordinates $\{\mathbf{a}_i\}_N$ and $\{\mathbf{a}'_i\}_N$, describing two conformations of a molecule, we can define the weighted RMSD between them as

$$\mathrm{RMSD}(A, A')^2 = \frac{1}{W} \sum_i w_i |\mathbf{a}_i - \mathbf{a}'_i|^2,$$

ג) ה-Pymol החזיר שה-RMSD של ההתאמה בין שלושת המבנים הוא:

RMSD = 0.240

```
Match: read scoring matrix.
Match: assigning 499 x 443 pairwise scores.
MatchAlign: aligning residues (499 vs 443)...
MatchAlign: score 2208.000
ExecutiveAlign: 441 atoms aligned.
ExecutiveRMS: 12 atoms rejected during cycle 1 (RMSD=1.51).
ExecutiveRMS: 23 atoms rejected during cycle 2 (RMSD=0.60).
ExecutiveRMS: 25 atoms rejected during cycle 3 (RMSD=0.38).
ExecutiveRMS: 17 atoms rejected during cycle 4 (RMSD=0.29).
ExecutiveRMS: 9 atoms rejected during cycle 5 (RMSD=0.25).
Executive: RMSD =    0.240 (318 to 318 atoms)
```

ד) חסרונות בשיטה של RMSD הוא:

1. ככל שיש יותר נקודות פונקציית מרחק תעלה. הפונקציה היא מונוטונית עולה (באופן חלש) לפי הגדרה.
2. מושפע מאוד מהשגיאות הגדולות ביותר. גם אם רוב האטומים מדוייקים יכולה להיות שגיאה גדולה בעקבות כמה הבדלים בודדים. אלגוריתם שמנסה לצמצם את הRMDS יכול לבחור בהתאמה פחות טובה בגלל הבדלים מקומיים, מה שבהסתכלות הכללית תיהיה פחות טובה.
3. לא מושפע מה"תוכן" של האטומים. לכולם משקל זהה בחישוב.
4. צריך אותו אורך שרשרת - עובד הכי טוב שהחלבונים ידועים כדומים.

בונוס - ה) שיטה נוספת שלמדנו שאלטרנטיבית ל- RMSD היא Bottleneck (אותו מימשנו גם בקוד)

כבר משמה ניתן להבין בגדול כיצד הוא מחושב - המרחק המקסימלי בין 2 נקודות (אטומי Ca במקרה שלנו) שהותאמו זה לזה. כפי שניתן לראות בנוסחה מהשיעור:

Bottleneck $\max ||a_{ki} - b_{ti}||$

יתרון:

1) קל ומהיר לחישוב - עשינו את זה ב O(n)
2)

חסרון:

1) יכול להיות מדד לא הכי אמין - יכול להיות זיווג אחד שבאמת רחוק זה מזה בעקבות דגימות לא נכונות לדוגמה של החלבון אבל כל שאר ההתאמה יחסית תקינה - במקרה הזה עדיין נקבל Bottleneck גבוהה.

.10

א.

RMSD(before disordered cleanup) is: 0.24177424792346222

0.194 = (RMSD(after disorder cleanup