

תרגיל 1 – עיבוד מידע תלת מימדי בביולוגיה מבנית

לירון גרשוני (308350503) ודרור בר (203523352)

1 שאלה 1

שלושת המבנים השונים פוענחו באמצעות X-RAY DIFFRACTION. קשה לבצע את הפענוח באמצעות השיטה הזאת כי לעיתים קשה לגבש חלבונים למבנה של קריסטלים. לאחר שמצליחים לגבש את החלבונים התשובה שמתקבלת היא יחסית מהימנה.

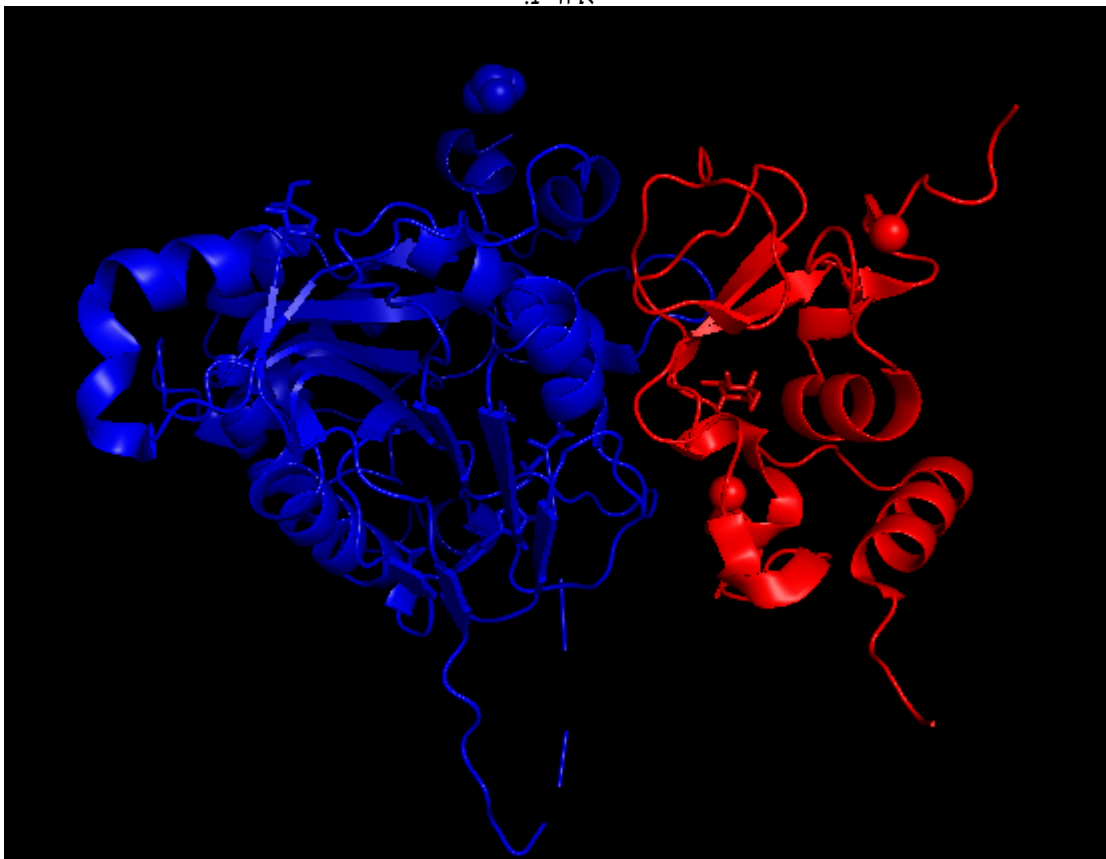
מדובר על השיטה העתיקה ביותר לפיענוח מבנה חלבונים והיא מתבססת על כך שאפשר ליצור גבישים של חלבונים. ניתן לקחת את הגבישים ולהקרין עליהם קרני רנטגן שיוצרים diffraction שבעזרתו אפשר להמיר את המידע לצפיפות של אלקטרונים והמבנה.

הרזולוציה של שלושת המבנים נעה בין $1.8 - 2.0 \text{ \AA}$

2 שאלה 2

הה

איור 1:



3 שאלה 3

בשלושת הקבצים ניתן לראות מספר רב של מבנים שניוניים מהסוג של α -helix, חלקם ארוכים יחסית וחלקם קצרים מאוד (אפילו שלושה ח"א) ולעיתים שני סלילים מופרדים רק ע"י ח"א אחת ולכן ייתכן ומדובר על אותו סליל.

- 6w4h – 16 סלילים
- 6w61 – 17 סלילים
- 6w75 – 28 סלילים

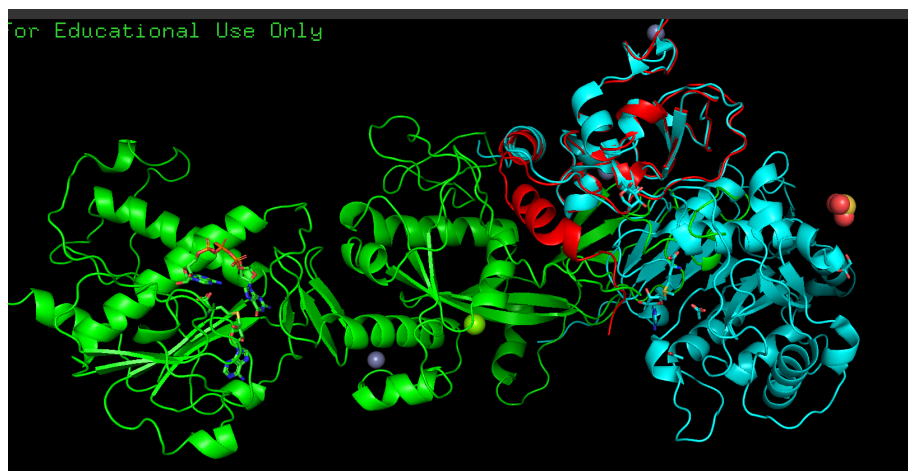
4 שאלה 4

ההומולוג שבו השתמשו הוא 5c8s.1 שהוא בזהות רצפית של 95.5

5 שאלה 5

לא היה לנו ברור האם הכוונה לעמד את ה template אל מול המודל (איור 2) או את המודל אל מול *cov2* (איור 3) ולכן צירפנו את שניהם

איור 2:



איור 3:

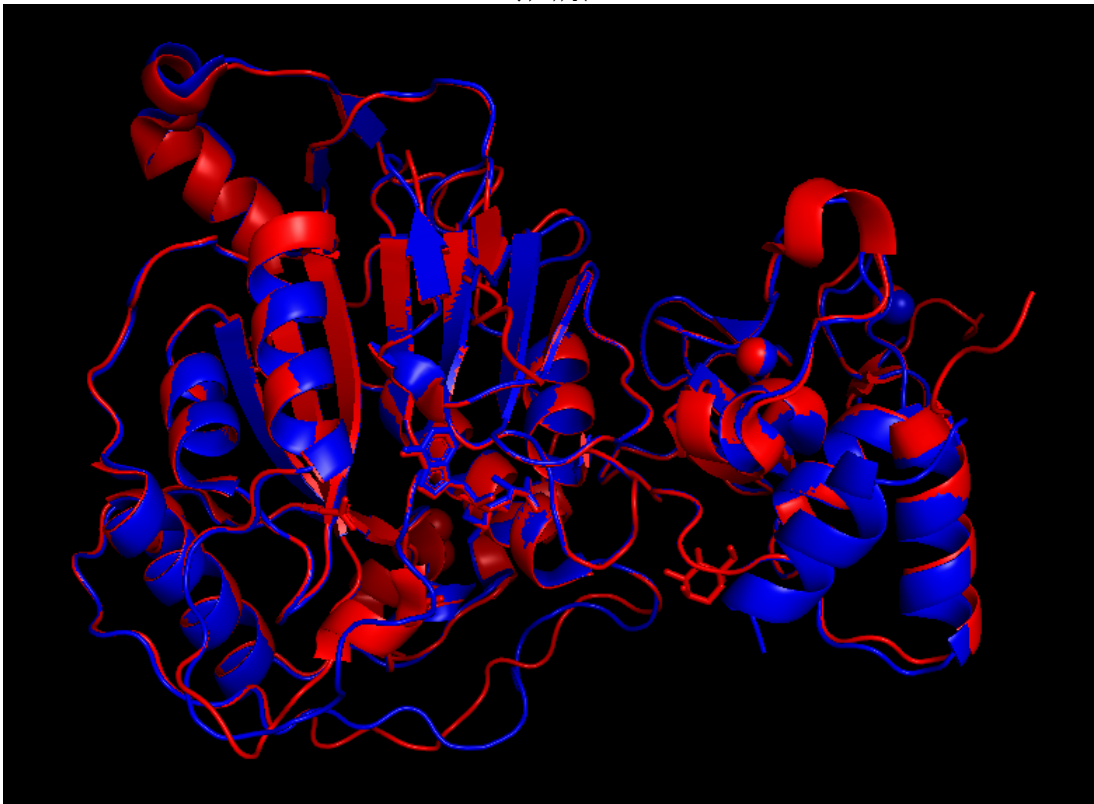
6 שאלה 6

לדעתנו לא תהיה אפשרות ל-*nsp14* ו-*nsp16* להתחבר במקביל ל-*nsp10*:
לפי האתר מדובר על הטרודימר ויש התאמה גדולה בין *nsp14* ל-*nsp16* מה שימנע מצב לחיבור משותף בו זמנית.
בנוסף, ניתן לראות שהחיבור בין *nsp10* לבין *nsp14/16* הוא באותו איזור מה שיימנע חיבור משותף לשניהם בו זמנית.

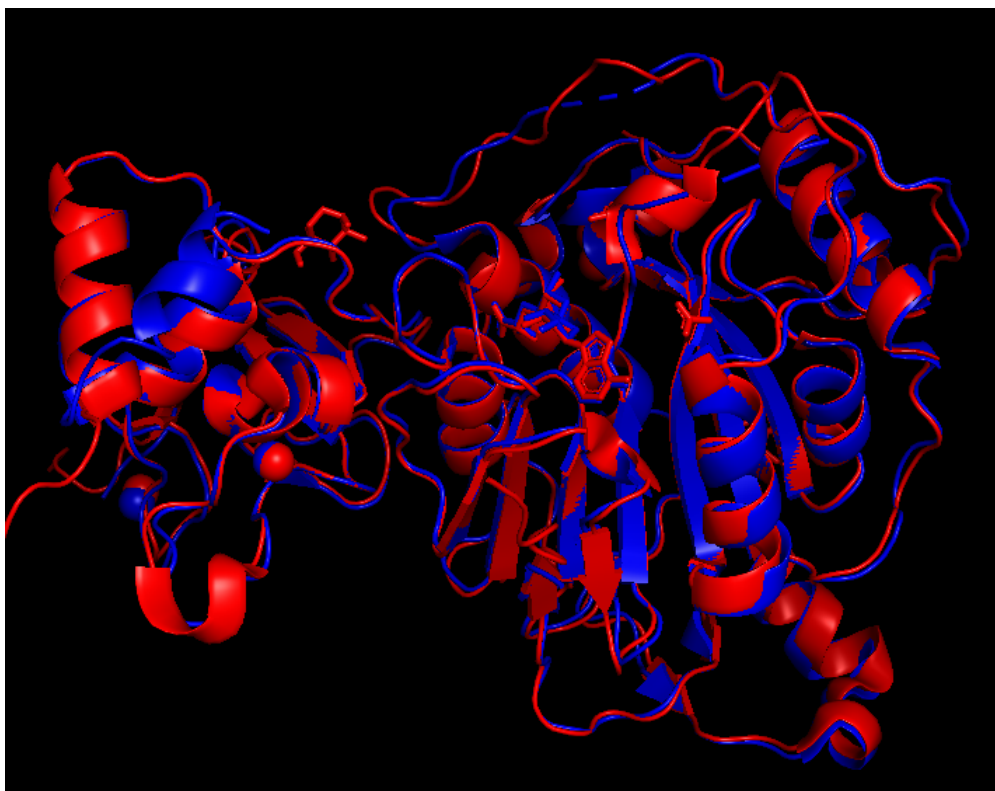
7 שאלה 7

idn של מבנה החלבון של *SARS – CoV* שהשתמשנו בו הוא 3r24, קיימים רזולוציות יותר טובות אך אין בהם את *nsp10, 16* שאותם אנחנו מעוניינים להשוות.
וה-*idn* של *MERS – CoV* שהשתמשנו בו הוא 5yn6.
בשביל עימוד ל-*Cov2* השתמשנו בקובץ 6w4h
עימוד *Cov2* (באדום) מול *SARS – Cov* (בכחול) :

איור 4:



עימוד *Cov2* (אדום) מול *MERS – CoV* (כחול) :

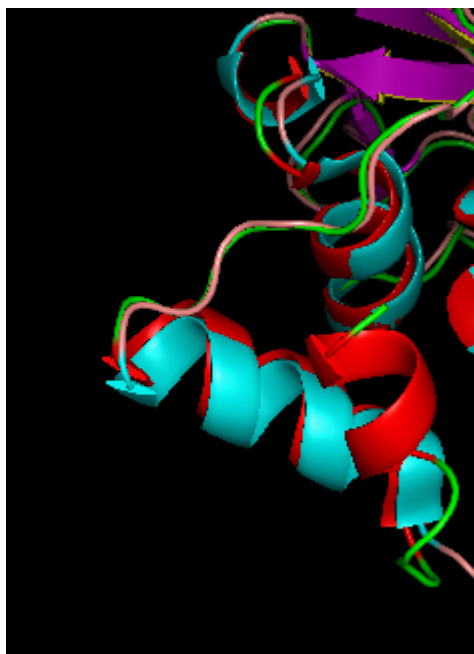


איור 5:

8 שאלה 8

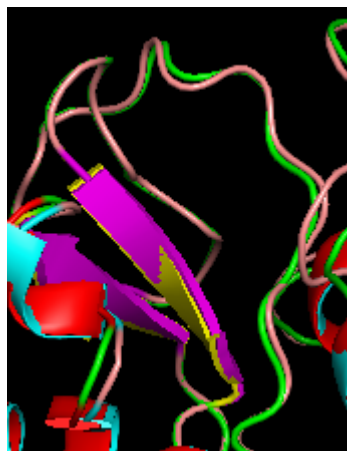
8.1 סעיף א'

נקודת דמיון אחת בין המבנים השונים היא השמירות המבנית בין רצפים ארוכים בעלי מבנה שניוני זהה של α -helix – ניתן לראות בצבע אדום ותכלת



איור 6:

נקודת דמיון שנייה בין המבנים היא חלק מהרצפים של ח"א ללא מבנה שניוני מוגדר שנשארים זהים ומחברים מבנים יותר מורכבים – ניתן לראות בצבע ירוק וצבע גוף



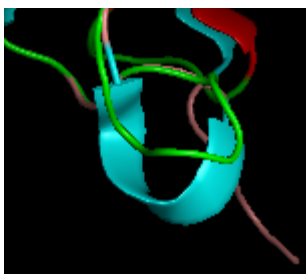
איור 7:

נקודת שוני בין המבנים היא רצף של ח"א בחלק החיצוני של המבנה ללא מבנה שניוני מוגדר שאינם מתלכדים בין המבנים השונים – ניתן לראות את הצבע הירוק מול צבע הגוף ששניהם קטועים אך לא מתלכדים אחרי העימוד



איור 8:

נקודת שוני שנייה היא מבנים שניוניים של a-helix המופיעים במבנה אחד ולא מופיעים מבנה השני



איור 9:

8.2 סעיף ב'

המשוואה לחישוב $RMSD$ בין שני סטים מבניים v, w בעלי n אטומים היא:

$$RMSD(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left((v_{i,x} - w_{i,x})^2 + (v_{i,y} - w_{i,y})^2 + (v_{i,z} - w_{i,z})^2 \right)}$$

כך שאנחנו מבצעים חישוב עבור כל קורדינטה מרחבית

Root-mean-squared deviation of atomic positions בערך

8.3 סעיף ג'

ה $RMSD$ בין $COV2$ לבין $SARS - CoV$ הוא: $0.313A$ באמצעות הפקודה *align* ו- $0.248A$ באמצעות *super*
ה $RMSD$ בין $COV2$ לבין $MERS - CoV$ הוא: $0.490A$ באמצעות הפקודה *align* ו- $0.483A$ באמצעות הפקודה *super*

8.4 סעיף ד'

לחישוב $RMSD$ יש מספר חסרונות

1. שינוי בסוג האטום לא משפיע על הציון שיתקבל – יכול להיות מצב בו אטום מסוים מתחלף באטום אחר (לדוגמה חנקן בפחמן) מה שיוביל לשבירה של מבנה שניוני או שינוי בפונקציה של החלבון אך המיקום המרחבי של כל האטומים יישאר מספיק דומה בשביל לקבל תוצאה נמוכה שתעיד על דמיון.
2. כל אטום שווה ערך בחישוב לאטום אחר למרות היותו במבנה שניוני מסודר לעומת רצף של במבנה שניוני – היינו כנראה רוצים לתת משקל רב יותר להבדל בין אטומים במבנים מורכבים יותר (שניוניים) לעומת המשקל של רצפים שאינם נמצאים במבנים שניוניים.
3. רגישות לנקודות קצה – מספר מועט של אטומים שלא מתאימים אחד לשני בכלל יגרור הגדלה מסיבית של הציון למרות ששאר האטומים יכולים להתאים מאוד.

8.5 סעיף ה'

נציע שיטת מדידה חדשה שתנסה לפתור את הבעיה שהצגנו בסעיף 2, כלומר לתת משקל יותר גדול לאטומים שנמצאים במבנה שניוני מאשר אטומים שאינם מבנה מוגדר: בהינתן שני חלבונים v, u נגדיר פונקציית משקל W ונגדיר נוסחה $WRMSD - Weight RMSD$

$$WRMSD(v, u) = \sqrt{\frac{1}{n} \sum_{i=1}^n W(v_i, u_i) \|v_i - u_i\|^2}$$

$$W(v_i, u_i) = \begin{cases} w_1 & v_i \text{ in secondary structure} \wedge u_i \text{ secondary structure} \\ w_2 & \text{else} \end{cases}$$

כך שניתן לקבוע כי $w_1 > w_2$ ולתת יותר משקל לנקודות הללו.

ייתרון אחד הוא התייחסות למבנה השניוני של החלבון והימצאות של אטום במבנה שניוני או לא – מה שאולי מעיד על פונקציה זהה או שונה.

חסרון אחד הוא העובדה שאנחנו לא נותנים התייחסות לסוג המבנה השניוני $\alpha - helix$ vs $\beta - sheet$ ובנוסף ייתכן שקבוצות מחקר שונות ירצו לתת ציון שונה w_1, w_2 מה שיישבור את האחידות בקבלת תוצאות ממחקרים.

שיטה נוספת שנראית לנו טובה נועדה לפתור את בעיה 3 של נקודות קצה, נריץ את $RMSD$ מספר רב של פעמים (m) כאשר בכל פעם בוחרים באופן אקראי פרקציה (N) מאותם n אטומים, לבסוף ניקח את הממוצע או החציון של כל המדידות הללו.

נקבל את הנוסחה $MRMSD - Multiple RMSD$

$$MRMSD(v, u, N) = median \left[\sum_{j=1}^m \left(\sqrt{\frac{1}{n_j} \sum_{i=1}^{n_j} W(v_i, u_i) \|v_i - u_i\|^2} \right) \right]$$

$n_j = \text{choose random of } N \text{ different atoms from } n$

ייתרון של השיטה הוא שאנחנו מבטלים רגישות לנקודות קצה ורעש

חסרון של השיטה הוא שהיא דורשת הרבה יותר כוח חישובי כי צריך לבצע את פעולת העימוד והחישוב פעמים רבות עם מספר רב של אטומים, בנוסף יהיו חורים בחלבון.