

**אוניברסיטת אריאל בשומרון**

הפקולטה למדעי הטבע  
המחלקה לביולוגיה מולקולרית

# **אפיון אזורי קישור שונים של חלבונים בקטריאלים לחלבונים אנושיים**

## **Characterizing Different Linking Regions of Bacteria- Human Protein Complexes**

**שם המנחה: דר' דרור טובי**

**מוגש ע"י:**

**ת.ז. 205836810**

**רועי קישון**

**ת.ז. 310195482**

**אלכס רפופרט**

**תשפ"א**

## תוכן עניינים

1..... עמוד שער

2..... תוכן עניינים

3..... תקציר

4-5..... רקע מדעי

6..... מטרת הפרויקט ושאלות מחקר

7..... שיטות ועיבוד הנתונים

8..... תוצאות

8..... קומלקסים מתאימים

8-9..... סינון חלבונים

9-10..... בחינה ואפיון מבנים

11-12..... מסקנות ודיון

13-14..... מקורות

15..... נספחים

15..... קוד ב-Python

15..... תיעוד האפיון המבני ב-Excel

\*ניתן להחזיק *ctrl* וללחוץ על ראשי פרקים לקפיצה לפרק.

## תקציר

במחקר הבא נבחנים אזורי קישור של מעכבים חיידיים שונים אל חלבונים אנושיים. באמצעים ביואינפורמטיים אנו מנסים לזקק את שידוע לגבי קישורים אלו באופן כמותי. כמה אזורים יעודיים לקישור מעכב יש בחלבונים אנושיים? האם מעכבים שונים עבור אותו חלבון נוטים להקשר באזור שונה? הבנה יותר מעמיקה של תוכנית הפעולה החיידקית לעיכוב חלבונים אנושיים תרחיב את אופקינו, ויכולה לספק תובנות חדשות לדרכי התמודדות עם "אויב" (לא תמיד) קדום זה.

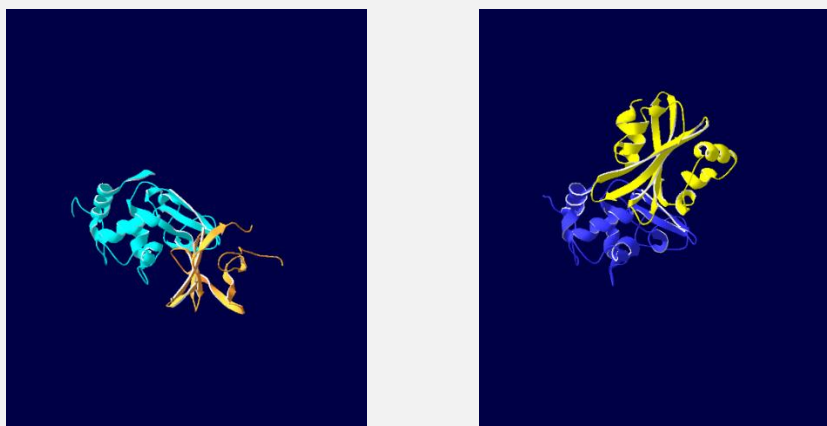
לשם כך נאספו מבנים תלת-מימדים של קומפלקסים רבים ממאגר מבני החלבונים האינטרנטי Protein Data Bank. לאחר סינון וסידור המבנים, בוצעו השוואות מבניות בין אופני קישור של חלבונים חידקיים אל אנושיים. נמצא כי רוב (66% מהנבדקים) חלבוני החידקיים בעלי חפיפה חלקית באזור הקישור לחלבון האנושי. כלומר, רובם נקשרים אל אותו אזור בחלבון האנושי אך הקישור לא חופף במלואו. שאר החלבונים החידקיים שנבדקו (33%), הציגו קישור אל אותו אזור ברמת זהות גבוהה. זאת למרות שרצף חומצות האמינו שלהם לא זהה. בנוסף, לא נמצאו כלל חלבונים אנושיים אליהם נקשרים רצפים חידקיים באזורים שונים לגמרי.

## רקע מדעי

בגוף האדם יש איברים העשויים מרקמות, שמורכבות מתאים המורכבים ממולקולות שונות של שומן, חלבון וכו'. האנטומיה של הגוף היא מורכבת ויש מידע רב עליה במקורות רבים. כיום, מתוך מערכת מורכבת זו, יש תחום מחקר ביולוגי בשם פרוטאומיקה, שמתמקד בחקר חלבונים. חלבונים הם הפועלים וחלק מאבני הבניין של גופינו. הם בעלי חשיבות גבוהה עקב תפקידם הרחב, מבניית שרירים ועד הקניית שלד מבני לתאים. הם גם מהווים את עיקר מערכת החיסון שלנו ועוזרים לנו להלחם בחיידקים.<sup>1,2</sup>

חלבונים חשובים גם מהצד השני של המטבע, עבור חיידקים. לחיידקים לעתים יש פלסמיד, שזהו דנ"א מעגלי דו גדילי נפרד מהדנ"א הכרומוזומלי שיש לכל חיידק. הפלסמידים יכולים לעבור בין חיידקים בדרכים שונות, ולקודד לחלבונים המקנים לחיידק יתרונות השרדותיים. חלבונים שכאלו יכולים למנוע ממערכת החיסון האנושית להרוס את החיידק ואפילו לא לזהות אותו כעויין. הגבלת המערכת החיסונית האנושית על ידי חלבון בקטריאלי, פעמים רבות נעשית על ידי היקשרות החלבון הבקטריאלי לאנושי, ועיכוב פעילותו התקינה. התפיסה הכללית הרווחת היא שהקשרות שכזו נעשית באתר קישור פעיל ספציפי, בודד ויחודי לכל חלבון. הידע שלנו לגבי קישורים אלו אינו מלא, והנושא נחקר לעומק במדע של ימינו.<sup>3-5</sup>

במחקר מקדים שנעשה בקבוצתנו נמצא חיידק עם שני פלסמידים, שמקודדים ל-2 חלבונים חיידקיים, הנקשרים לאתרים שונים על ליזוזים (חלבון המפרק את דופן החיידקים). בכוח כל אחד מרצפים חיידקיים אלו לעכב ולמנוע את פעילות הליזוזים. ההנחה היא שבעזרת 2 רצפים שונים אלו, לחיידק תהיה עמידות גבוהה במיוחד כנגד ליזוזים. במצב רגיל, אם גוף אנושי ימנע ממעכב עיכוב מוצלח על ידי זיהוי והריסתו או אפילו על ידי חסימת אתר הקישור שלו, הגוף ייחשב כמחוסן לאותו חיידק וידע להתמודד איתו. אך במצב מיוחד זה, לאחר התמודדות עם מעכב אחד, החיידק ימשיך לעשות שימוש במעכב השני שברשותו, וימשיך לנצח במערכה אל מול מערכת החיסון.



איור 1 - הדמיה ממוחשבת של קישור חלבונים בקטריאליים שונים (בצהוב וכתום) במקומות שונים בליזוזים (תכלת וכחול). ניתן להבדיל בין אזורי הקישור מכיוון שהליזוזים נמצא בפוזיציה כמעט זהה בשתי התמונות. המעכב מימין – *Ivy family type-c lysozyme inhibitor*. המעכב משמאל – *Membrane-bound lysozyme inhibitor of type-c lysozyme*.

לאור הממצאים, נשאלת השאלה האם קיימים עוד מצבים דומים? האם חלבונים אנושיים נוספים מעוכבים על ידי קישור ב-2 אתרים שונים, על ידי רצפים חידקיים שונים? לאחר בדיקה מעמיקה באינטרנט נמצאו מאמרים רבים העוסקים בדרכי ההשרדות המתפתחות והרבות של חידקים, אך לא נמצא מאמר המדבר על מנגנון חידקי כנ"ל.<sup>6</sup>

צורת המחקר שלנו היא ביואינפורמטית. זהו תת-ענף מתפתח של ביולוגיה יחד עם מדעי המחשב שעושה שימוש בהשגה, אחסון, אנליזה והפצה של מידע ביולוגי. בעזרת כמות המידע העצומה שיש באינטרנט בימינו, אפשר לבדוק התאמה בין רצפים ארוכים על ידי שימוש בכלים כמו Blast. אפשר לסנן, במהירות ועילות, כמות עצומה של מידע על ידי כתיבה ושימוש בקודים בשפות תכנות כמו Python ו-Perl. בנוסף, ניתן לעשות שימוש במאגרי ענק ממוחשבים של חלבונים כמו RCSB PDB (Research Collaboratory for Structural Bioinformatics, Protein Data Bank).<sup>7-9</sup>

למעשה בהקשר הביולוגי, הטכנולוגיה מאפשרת לנו לגשת למספר רב של חלבונים ולסנן מתוכם את אלו אשר נקשרים אליהם חלבונים חידקיים. כך מאופיינים יותר קשרים, העין האנושית מתעסקת במקרים רלוונטים יותר, וגדלים הממצאים. כל זאת לעומת ניסוי רטוב, שמגביל את מספר החלבונים הניתנים לבדיקה על פי תכתיב הזמן והכסף.

על מנת שהמשך המאמר יהיה מובן לכל קורא, כעת ניתן רקע פרטני לגבי הכלים הביואינפורמטיים בהם נעשה שימוש במחקר זה. תחילה, המאגר ממנו נשאב כלל המידע ומבני הקריסטלוגרפיה הוא RCSB PDB.<sup>8</sup> זהו אתר אינטרנט המהווה מאגר עצום של מידע על חלבונים, המכיל את מבנים וקומפלקסים רבים שנשרקו בטכניקות שונות, ונשמרו וירטואלית. כל מבנה ניתן להורדה ולהצגה באתר או בתוכנות שונות, ומספק תצוגה תלת מימדית של המבנה הממשי של החלבון. המידע במאגר מתווסף על ידי ביולגים מבניים רבים, והוא נגיש לכל.<sup>10</sup>

Perl ו-Python הן שפות התכנות בהן השתמשנו. שפות תכנות אלו נחשבות כקלאסיות לעיבוד מידע ונעשה בהן שימוש רב במחקרים ממוחשבים. הן בהתפתחות והתייעלות מתמדת, בין היתר עבור שימושים ביולוגיים. ניתן בעזרתן לעשות אנליזה, ניקוי, סינון, וויזואליזציה לכמויות גדולות של מידע. כלי נחוץ כאשר מתנהלים מול מאגר כה עצום כמו PDB, ומעוניינים לזכך מתוכו את המידע הרצוי.<sup>11,12</sup>

לבסוף, התוכנה בה נבחנו ואופיינו מבנית קומפלקסים שונים היא Swiss PDB Viewer. תוכנה המותאמת במיוחד לפורמט קבצים שהורדו מ-PDB, המאפשרת לבצע אנליזה של מספר חלבונים בו זמנית. ניתן בעזרתה לעשות חפיפות בין קומפלקסים, ולהשוות אזורי ספציפים ואת אופן קישורם לחלבונים אחרים. כלל התצוגות הגרפיות של מבנים חלבוניים במאמר זה לקוחות מאנליזות שנעשו בתוכנה זו.<sup>13,14</sup>

## מטרת הפרויקט ושאלות המחקר

**מטרת על** – אפיון אזורי הקישור של חלבונים חידקיים לאנושיים, אליהם נקשר יותר מחלבון חידקי אחד.

**מטרה 1:** מציאת קומפלקסים של חלבונים אנושיים עם חידקיים ושליפתם ממאגר.

- **מטרת משנה 1.1:** שימוש במאגר ממוחשב המכיל מספר רב של חלבונים.
- **מטרת משנה 1.2:** מציאת דרך לביצוע חיפוש ממוקד שייב קומפלקסים ככתוב ב-"מטרה 1".

**מטרה 2:** סינון חלבונים - השארת אלו אשר אליהם מתחברים מספר חלבונים חידקיים.

- **מטרת משנה 2.1:** ביצוע השוואת רצפים בין הרצפים האנושיים למציאת כפילויות וחלבונים אנושיים דומים (אותם בהמשך נוכל להצליב ולמצוא חלבונים חידקיים שנקשרים באזורים שונים).

- **מטרת משנה 2.2:** כתיבת קוד לסינון החלבונים האנושיים והשארת אלו שבעלי רמת זהות גבוהה (על פי ההשוואה ב-"מטרת משנה 2.1").

- **מטרת משנה 2.3:** הצלבת הרצפים החידקיים שמתחברים לרצפים אנושיים זהים.

**מטרה 3:** בחינה ויזואלית של מבנה כל קומפלקס, תיעוד מיקומי קישור החלבונים החידקיים וסינון כפילויות.

- **מטרת משנה 3.1:** הורדת מבנים המרחביים (קריסטלוגרפיים) של כל קומפלקס רלוונטי ממאגר.
- **מטרת משנה 3.2:** חפיפת החלבונים האנושיים הזהים מכל קומפלקס, אחד עם השני.
- **מטרת משנה 3.3:** השוואה ויזואלית בין אזורי הקישור של החלבונים החידקיים השונים וקביעת הממצאים.

## שיטות ועיבוד הנתונים

כל מהלך העבודה נעשה בצורה ממוחשבת, ועל כן שיטותינו כוללות עיבוד נתונים רב. לצורך המחקר השתמשנו במאגר החלבונים RCSB PDB. ממאגר זה נשלפו כל קומפלקסי החלבונים אשר מורכבים מרצף חיידיקי ורצף אנושי. זאת נעשה על ידי חיפוש מתקדם של קומפלקסים בהם חלבונים בעלי ערך Taxonomy ID של חיידיק(2) ואנושי(9606) יחדיו. התוצאות בהן רצפי החלבונים הרצויים הורדו מהמאגר והמשך העבודה נעשה עליהם.

לאחר מכן, נעשתה השוואת רצפים בין הרצף האנושי מכל קומפלקס, אל הרצפים האנושיים משאר הקומפלקסים. כך נמצאו חלבונים אנושיים דומים ואפילו זהים, שהיו שמורים במאגר כחלק מקומפלקסים שונים של חלבונים. השוואת הרצפים נעשתה באמצעות התאמת רצפים גלובלית, על פי מטריצת הציונים BLOSSUM50. נעשה שימוש בקוד ב-Perl כדי לבצע את ההתאמות למספר רב של זוגות רצפים.

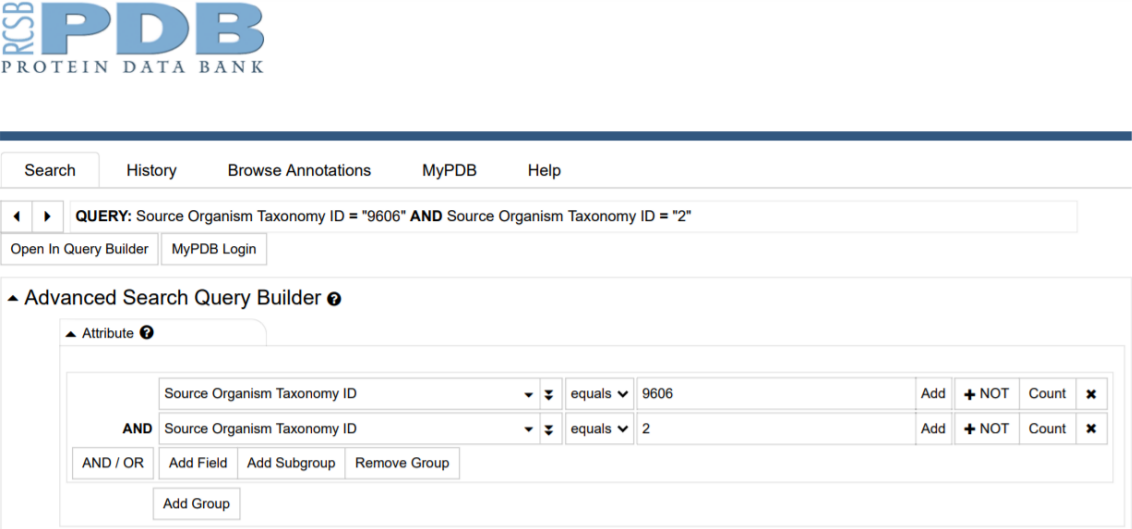
נכתב קטע קוד ב-Python, בעזרתו בוצעו סינון וחלוקה לקבוצות של הרצפים המשווים. תחילה, סוננו קומפלקסים בהם החלבון האנושי והחיידיקי קצרים מ-20 חומצות אמינו (כדי להמנע מ-False Positives). הקומפלקסים חולקו לקבוצות בהן החלבונים האנושיים בעלי רמת זהות גבוהה (>60%), והחלבונים החיידיקיים שלהם בעלי רמות זהות נמוכות (<60%). בסוף פלט קוד זה נוצר מסמך בו יש קומפלקסים מחולקים לקבוצות, בהן החלבונים האנושיים דומים אחד לשני, והחיידיקיים שונים אחד מהשני.

מ-PDB הורד המבנה המרחבי (עד כה דובר על רצפים בלבד) של כל קומפלקס ששרד את הסינון עד כה. הקומפלקס הראשון בכל קבוצה נחשב לחלבון המודל, אליו הושוו שאר חלבוני הקבוצה – מכיוון והרצפים האנושיים באותה הקבוצה זהים, אין ערך להשוות כל אחד מהם ליותר מחלבון אחד אחר בקבוצה. בחינת הקישור (של החלבון החיידיקי) והמבנים נעשתה בתוכנה Swiss-PDB, והחלבונים הוצגו בתצורת Ribbon המדמה את מבנם המרחבי. לבסוף, כאשר המבנים האנושיים מכל קומפלקס סודרו בחפיפה אחד על השני, ניתן היה לבחון את צורות הקישור השונות של החלבונים החיידיקיים (כמודגם באיור 3 בתוצאות). יש לציין שבעת ההשוואות המבניות נפסלו קבוצות False Positives להן לא היתה רלוונטיות ביולוגית לניסוי (הרחבה בדיון).

עבור כל השוואת קומפלקסים תועד: אם החלבונים החיידיקיים נקשרים באזור שונה, באותו אזור, האם יש חפיפה חלקית באזור הקישור, והאם יש ביניהם הפרשי אורך דרסטיים (גדולים מ-100 חומצות אמינו). תיעוד זה נקבע על פי הערכה ויזואלית של המבנים המרחביים של החלבונים, בתוכנה Swiss-PDB.

## תוצאות

1. קומפלקסים בעלי חלבונים אנושיים הקשורים לחלבונים חיידיקים - נמצאו במאגר PDB דרך חיפוש מתקדם על בסיס טקסונומי.



אזור 2 - תיעוד צורת החיפוש המתקדם באתר PDB. נמצאו קומפלקסים המכילים חלבונים חיידיקים ( $Taxonomy\ ID = 2$ ) וגם חלבונים אנושיים ( $Taxonomy\ ID = 9606$ ). סה"כ התקבלו 823 מבנים כתוצאה לחיפוש זה.

2. סינון החלבונים – לאחר השוואה של הרצפים האנושיים בין כל קומפלקס(נעשתה בקוד Perl על ידי המנחה), נעשה סינון התוצאות. הסינון נעשה בקוד Python (נספח 1) ולאחריו נותרו 42 קבוצות חלבונים. בכל קבוצה מספר קומפלקסים בעלי חלבון אנושי דומה/זהה.

pdb1o9a_A.seq	pdb1o9a_B.seq
pdb2rkz_A.seq	pdb2rkz_M.seq
pdb3cal_A.seq	pdb3cal_B.seq
pdb3zrz_A.seq	pdb3zrz_C.seq
pdb4pz5_B.seq	pdb4pz5_B.seq

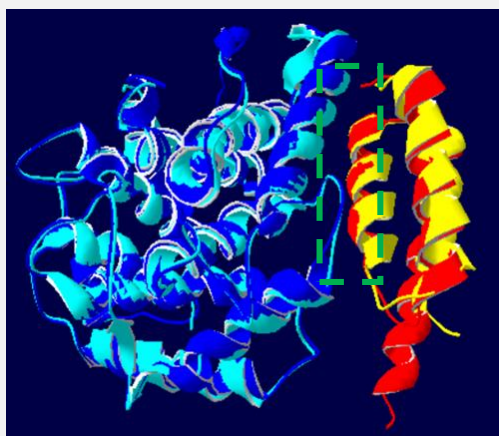
טבלה 1 – דוגמא לקבוצה אחת מתוך התוצאות המסוננות. בכל שורה מוצג רצף אנושי בקומפלקס (משמאל), ורצף חיידיקי הנקשר אליו (מימין). החלבון הראשון מהווה את 'חלבון המודל' עבור הקבוצה, אליו נחפפים שאר המבנים ונבחנות צורות הקישור של הרצף החיידיקי.



Model Prot.	Examined Prot.	Humanoid Identity %	Bacterial Identity %
<b>2gww</b>	2ibf	100	37
	3rf3	100	35
	3tj5	100	19
	3tj6	98	16

טבלה 2 - דוגמא לקבוצת קומפלקסים לאחר השוואה. טורים משמאל לימין: 1: חלבון המודל אליו הושוו שאר החלבונים בקבוצה. 2: שמות שאר החלבונים הקבוצה כפי שהם רשומים ב-PDB. 3: אחוז זהות בין כל חלבון אנושי בקבוצה אל חלבון המודל האנושי. \* 4: אחוז זהות בין כל חלבון חיידיקי בהתאמה, אל חלבון המודל החיידיקי. \* אחוזי זהות אנושיים שווים ל-100 משמעותם שמדובר באותו חלבון אנושי. עם זאת, ניתן לראות כי מדובר ברצפים חיידיקיים שונים הנקשרים אליו על פי השוואת רצפיהם.

**3. בחינה המבנים בצורה ויזואלית והתאמתם** – המבנים המרחביים של קומפלקסים דומים הורדו מ-PDB ונחפפו. תועדו אזורי הקישור השונים של הרצפים החיידיקיים באקסל, וסוכמו הממצאים. מוצגות תוצאות בחינת מבנים של 12 קבוצות (30 קומפלקסים).



איור 3 - דוגמא לאחת מתוצאות החפיפה המבנית. מדובר בחפיפה של 2 קומפלקסים (2gox ו-2noj) אשר החלבון האנושי של כל אחד מהם נחפף עם האחר (בכחול ותכלת). ניתן לראות שבמקרה זה החלבונים החיידיקיים (צהוב ואדום) של כל קומפלקס נקשרים באותו אזור (מסומן בריבוע ירוק מקוקו).

2gox	2noj	Bacterial Identity
TDATIKKEQKLIQAQNLV REFEKTHTVSAHRKAQK AVNLVSFEYKVKKMVLQ ERIDNVLKQGLVR	NKKVVDQAQKAVELFKRTR TVATHRKAQRAVNLIHFQ HSYEKKKLQRQIDLVLKY NTLK	40%

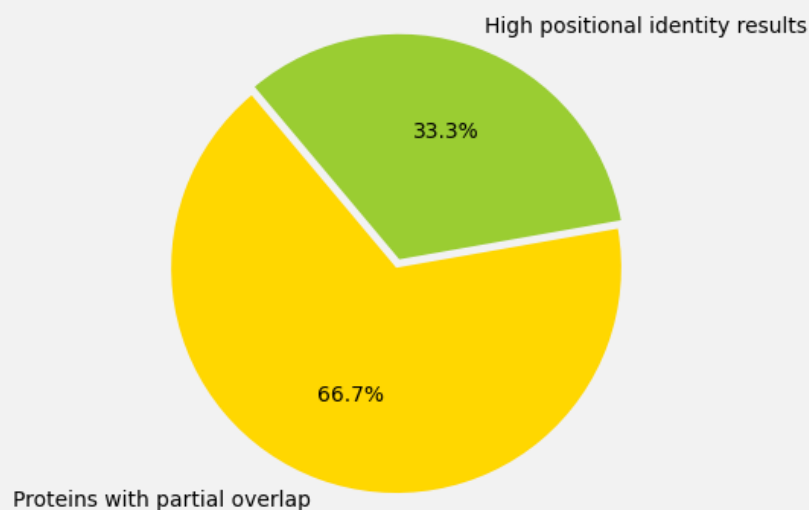
טבלה 3 – המשך אנליזה של החלבונים באיור 3 לעיל (2gox, 2noj). בהשוואת רצפי החלבונים החיידיקיים, ניתן לראות כי יש 40% זהות בלבד. עם זאת, עדיין ההליקס החיידיקי באזור הקישור נשמר מבנית (הליקס אדום-צהוב שבתוך ריבוע ירוק מקוקו באיור 3). רצף חומצות האמינו של אותו ההליקס הודגש בטבלה עבור כל אחד מהחלבונים החיידיקיים.



איור 4 – דוגמא נוספת בין שני קומפלקסים (1g4u, 4did) אשר החלבון האנושי של כל אחד מהם חופף על השני (שניהם בצהוב). ניתן לראות את הרצפים החיידקיים (כחול, תכלת) אשר חופפים באופן חלקי באזור הקישור (ריבוע ירוק מקווקו) אל החלבון האנושי. ניכרים הבדלי אורך וגודל משמעותיים בין רצפים חיידקיים אלו.

	Total amount of examinations	Partial overlap of the connection site	High positional identity (same bacterial chain position)	Completely different bacterial connection site
Proteins	30	20	10	0
>100 Len. Diff.	9	9	0	0
Total %	100%	67%	33%	0%

טבלה 4 - סיכום הממצאים מכלל הבחינות היוזואליות. בטבלה זו מסוכמים טיב הקשרים בין החלבון האנושי והחיידקי של כל קומפלקס. נמצאו 20 קומפלקסים בהם הייתה חפיפה חלקית באזור הקישור החיידקי. 10 קומפלקסים בהם החלבון החיידקי נקשר באותו אזור במדויק. באף השוואה בין הקומפלקסים לא נמצא חלבון חיידקי שנקשר לאזור שונה לגמרי בחלבון האנושי. ב-9 מהמקרים היה הבדלי אורך גדולים מ-100 ח"א בין החלבונים החיידקיים. יש לציין שטיב ההתאמה (חפיפה חלקית או גבוהה) נקבעה בצורה ויזואלית בלבד, ניתן לאמת את הממצאים בבדיקת הרכב חומצות האמינו באזור הקישור במידת הצורך.



איור 5 - גרף עוגה המציג את התפלגות התוצאות. ברוב הקומפלקסים (67%) היתה חפיפה חלקית באתר הקישור (בצהוב). בשאר המקרים (33%) היתה זהות גבוהה באזור הקישור של החלבון החיידקי (ירוק).

## מסקנות ודיון בתוצאות

מתוצאות הניסוי עולה כי למירב חלבוני החיידקים היתה חפיפה חלקית כלשהי באזור הקישור לחלבון האנושי בכל קומפלקס. זהו ממצא שעולה בקנה אחד עם ציפיותנו, מכיוון שאכן לרוב החלבונים האנושיים יש אזורים בודדים מסויימים אליהם נקשרים מעכבים בקטריאלים וליגנדים בכלל. משום שבהרבה קבוצות שנבחנו בניסוי, הרצפים המעכבים דומים אך לא זהים – הגיוני שהחפיפה בין המעכבים היא לרוב חלקית (Partial Overlap) ולא מוחלטת באזור הקישור. יתרה מזאת, רק במקרי החפיפות החלקיות נצפו הבדלי אורכים גדולים בין הרצפים החיידקיים. יתכן שיש קורולציה כלשהי, ושאורך הרצף המעכב מהווה פקטור נוסף המשפיע על אזור ההקשרות שלו לחלבון האנושי.<sup>15</sup>

שתי התוצאות האחרות היו יותר מעניינות ופחות ניתנות לחיזוי: לא היו כלל רצפים חיידקיים שנקשרו באזור שונה לגמרי לאותו חלבון אנושי, ו-33% מרצפי החיידקים נקשרו ברמת דיוק גבוהה לאותו אזור בחלבון האנושי.

איך יתכן ואין מקרים נוספים בהם יש 2 אזורי קישור למעכבים שונים לגמרי? יכול להיות שמקרה הליזוזים לעיל (איור 1) יחיד במינו? אולי קיימים חלבונים עם אזורים שונים בהם יש אפשרות לקישור, אך אחד מהם בעל התאמה גבוהה יותר למעכב, כך שלבסוף לא נעשה שימוש באזורים האחרים. זוהי רק השערה, בפועל לא נמצאו מקרים נוספים במקורות בהם 2 אזורי קישור שונים למעכבים בקטריאלים, אבל הטבע מלמד אף פעם לא להגיד אף פעם.

שליש מהמקרים בהם הרצפים החיידקיים נקשרו באותו אתר (ברמת דיוק גבוהה), גם הם ממצא לא ברור. הרי כל חלבון חיידקי נובע ממקור (חיידק) אחר, והוא שונה ברצפו ממשנהו. עם זאת, הם נקשרים באותו אזור בדיוק (איור 3), מורכבים ממבנים שניוניים דומים ואפילו חופפים אחד על השני באזור הקישור בצורה מושלמת. סברה שעלתה היא שיתכן וחלבונים מעכבים אלו מקורם בחלבון קדמון משותף, אשר אתר הקישור שלו שמור בעוד שאר חומצות האמינו שלו השתנו לאורך האבולוציה. אם סברה זו לא נכונה, אזי החיידקים מפתחים עצמאית חלבונים הנקשרים לאותו אזור בחלבון. אולי קיימים אזורים בחלבונים של גופינו שמהווים מקום קל לקישור, נדרש מחקר בנושא.

יש לציין שלמעשה נבדקו מבנים של 14 קבוצות (בהן מעל 35 קומפלקסים) מתוך 40 קבוצות סופיות שעברו את הסינון עד השלב הסופי. נפסלו ולא נכללו הממצאים מ-2 קבוצות מתוך ה-14, אשר בשתייהן כלל לא היה קישור בין הרצף האנושי והחיידקי. למרות שהיוו חלק מאותו קומפלקס, היה חלבון ביניים מתווך שקישר בין הרצפים הנ"ל והם לא נקשרו ביניהם ישירות. השוואות עבור קבוצות שכאלו שגויות מבחינה ביולוגית ולמרות שעברו את סינון המחשב, נחשבו כ-False Positives. להבנתנו, נוכחות False Positives אלו יכולה לנבוע מ-2 סיבות. יתכן והתיוג של חלבונים אלו במאגר של PDB שגוי, או שנדרשת אופטימיזציה לקוד לביצוע סינון נוסף.

שאר הקבוצות (26 שנותרו מתוך ה-40 הסופיות) לא נבחנו מפאת חוסר זמן. מדובר בקבוצות שבכל אחת מהן בין 2 ל-20 קומפלקסים שכל אחד מהם צריך להבחן ויזואלית בעין אנושית. חלק ניכר מזמן הניסוי עבר בכתיבת הקוד ב-Python, שיפורו והתאמתו לצרכי המחקר.

הימצאות קומפלקסי ה-False Positives לאחר הסינון, הממצאים הנוכחיים (0% קישורים שונים, 33% קישורים זהים) שמעלים שאלות חדשות, 26 הקבוצות הנוספות שלא עברו בחינה מבנית – כולם קוראים למחקר המשך. ניתן לעשות קריאה ושיפור נוספים של הקוד, לבחון את הקבוצות הנותרות ולהסיק מסקנות יותר כוללניות. במידה וימצא חלבון אנושי המעוכב ב-2 אתרים שונים לגמרי, ניתן לבצע בדיקה במעבדה לאימות הממצאים.

1. Gray, H. *Anatomy of the Human Body*. (Lea & Febiger, 1878).
2. Pennington, S. R. & Dunn, M. J. *Proteomics: From Protein Sequence to Function*. (Garland Science, 2001).
3. Eric D. Green, M.D., Ph.D. Plasmid. *Genome.gov*  
<https://www.genome.gov/genetics-glossary/Plasmid>.
4. Gerlach, D. *et al.* Methicillin-resistant *Staphylococcus aureus* alters cell wall glycosylation to evade immunity. *Nature* **563**, 705–709 (2018).
5. Juan Felipe Beltran, Ilana Lauren Brito. Host-microbiome protein-protein interactions reveal mechanisms in human disease. (Meining School of Biomedical Engineering. February, 2020).
6. Reddick, L. E. & Alto, N. M. Bacteria Fighting Back – How Pathogens Target and Subvert the Host Innate Immune System. *Mol Cell* **54**, 321–328 (2014).
7. Bioinformatics. *Genome.gov* <https://www.genome.gov/genetics-glossary/Bioinformatics>.
8. Bank, R. P. D. RCSB PDB: About RCSB PDB: Enabling Breakthroughs in Scientific and Biomedical Research and Education. <https://www.rcsb.org/pages/about-us/index>.
9. Welcome to Python.org. *Python.org* <https://www.python.org/about/>.
10. Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *PubMed*. doi: 10.1093/nar/gkaa1038. PMID: 33211854 (2021).
11. Singh, S., Brueffer, C. & Cock, P. Biopython Project Update 2017. in (2017).  
doi: 10.7490/f1000research.1114502.1.

12. Rodrigues, J. P. G. L. M., Teixeira, J. M. C., Trellet, M. & Bonvin, A. M. J. J. pdb-tools: a swiss army knife for molecular structures. *PubMed Central. PMCID: PMC6343223* (2018) doi: 10.12688/f1000research.17456.1.
13. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *PubMed. PMID: 9504803*. (1997) doi: <https://doi.org/10.1002/elps.1150181505>.
14. Johansson, M. U., Zoete, V., Michielin, O. & Guex, N. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinformatics* (2012) doi: 10.1186/1471-2105-13-173.
15. Alberts, B. *et al.* Protein Function. *Molecular Biology of the Cell. 4th edition* (2002).

## נספחים

1. קטע קצר מהקוד שנכתב ב-Python. מוצג קטע קצר של הפרמטרים (אורך ורמת זהות) על פיהם סוננו החלבונים לאחר השוואת רצפיהם. הקוד המלא (128 שורות) לא מוצג מטעמי סדרנוחות לקורא.

```
...
# Open the compared sequences file, and save certain proteins
# (longer than 20, 60% identity)
if __name__ == '__main__':
    humanListFile = open('C:\\..\\inter_list_seq_i.txt', 'r')
    lines = humanListFile.readlines()

    filtered = []
    for line in lines:
        splitted = line.split(' ')
        if represents_int(splitted[1]):
            firstSize = int(splitted[1])
            secondSize = int(splitted[3])
            match = float(splitted[4])

            if firstSize > 20 and secondSize > 20 and match > 60:
                filtered.append(line)
    ...
```

2. תמונה לדוגמא מקובץ אקסל המקור בו תועדו תוצאות האנליזות. ניתן לראות 3 חלבוני מודל משמאל אשר אל כל אחד מהם מושווים שאר חלבוני הקבוצה שמימנו. שאר העמודות הם הפרמטרים של האנליזה.

Model Prot.	Examined Prot.	Number of Connected Sites	Same Site (as Model)	Different Site	100< Bacterial Length Difference	Partial Overlap
1a2c	1nu9	2	1	1	1	1
1fcc	6fgo	1	1	1	0	1
1g4u	1gzs	1	1	1	1	1
	1he1	1	1	1	0	1
	2h7v	2	1	1	0	1
	3gcg	2	1	1	1	1
	4did	2	1	1	1	1

\*לכלל המעוניין, את הקבצים המלאים ניתן לקבל במייל.