תרגיל מסכם מס' 2: מבנה חלבון

1) **הסבר כללי**:

A) מערכת: קובץ ממאגר pdb כולל נתונים על שרשרת של חומצות אמינו (או כמה שרשרות), וכן על חומרים נוספים שזוהו בתוך הדגם. חומר נוסף כזה יכול להיות הליגנד הקשור לחלבון, או מולקולת מים, או חומרים נוספים. המינוח לאטומים השייכים לחלבון הוא Atom והמינוח לאטומים השייכים לחומרים אחרים הוא Heterogen Atom.

B) שאלת המחקר: מהו אתר ההיקשרות של הליגנד, מהן חומצות האמינו המרכיבות אותו ומה מקומן הן לאורך השרשרת (backbone) והן במרחב. נציג את הנתונים גם בגרף תלת ממדי בו יוצגו במרחב כל האטומים רק של חומצות האמינו שנמצאות בקרבת הליגנד, וגם במפת-צבעים דו-ממדית של טבלת מרחקים הדדיים בין אטומי הליגנד לאטומי החלבון, מסודרים לפי המקום שלהם על השרשרת.

C) מבנה הקובץ: פקודת מטלב pdbread קוראת מידע מקובץ זה ומסדרת אותו במבנה מורכב, שחלק מהשדות שלו הם מערכי-מבנים. מידע חלקי על השדות השונים נמצא בחומר הרצאה 10 וכן בתיעוד מטלב על פקודת pdbread. לפני ביצוע התרגיל, מומלץ לבצע פקודת pdbread על אחד הקבצים, לשמור את התוצאה במשתנה ולצפות בו בעורך המשתנים (ניתן להיכנס לכל אחד מן השדות ואם הוא עצמו כולל מבנה או מערך מבנים, ניתן להיכנס לרמות פנימיות יותר).

D) ארגון נתונים בתוך התוכנית: מידע על החלבון ועל תוצאות עיבוד הנתונים נמצא במבנים הבאים:

i) נתונים על כל החומר נמצאים במבנה מסוג pdb (נוצר ע"י מטלב מקריאת קובץ ה-pdb). שמות השדות מפורטים בתיעוד על פקודת pdbread (אך גם מצוינים בהמשך בהוראות).

ii) מבנה לנתונים של החלקים שנבחרו ע"י המשתמש, עם השדות הבאים:

(1) מזהה לליגנד שנבחר (מערך תווים)

(2) מזהה לשרשרת החלבון שנבחרה ע"י המשתמש (תו בודד)

(3) אטומי הליגנד שנבחר ע"י המשתמש (מערך מבנים עם שדות לפי פורמט pdb)

(4) אטומי שרשרת החלבון שנבחרה ע"י המשתמש (מערך מבנים עם שדות לפי פורמט pdb)

(5) אינדקסים של האטום הראשון מכל חומצת אמינו.

(6) אינדקסים של האטום האחרון מכל חומצת אמינו.

(7) שמות חומצות האמינו (מערך תאים שבכל אחד מהם מערך תווים).

iii) מבנה לערכים שצוינו ע"י המשתמש למרחקים בין אטומים, עם השדות הבאים:

(1) מרחק מינימלי (אטומים במרחקים קטנים ממנו נחשבים כנמצאים במגע/קשר).

(2) מרחק מקסימלי (אטומים במרחקים גדולים ממנו נחשבים כרחוקים מספיק להיות בלתי רלוונטיים).

E) מהלך העבודה: בהמשך קובץ זה תמצאו הוראות לכתיבת פונקציות לביצוע חלקים מסוימים של החישוב. העבודה מחולקת ל-3 חלקים (סעיפים 2-4).

i) כדי לבדוק את הפונקציות שתכתבו בסעיף 2, תפעילו אותן מסקריפט **tst** שתכתבו לצורך כך. את הפונקציות מסעיף 2 וסקריפט הבדיקה יש לשלוח לי במייל עד סוף הסמסטר. אל תמשיכו בכתיבה עד שתקבלו משוב על חלק זה.

ii) מטלות בסעיפים 3)-4) ישתמשו בפונקציות שנכתבו קודם לכן. אין לשכפל קוד או לכתוב מחדש! התוכנית הראשית היא סקריפט **protein** שמפעיל את הפונקציות. תוכלו לבדוק את הפונקציות ע"י הפעלת הסקריפט ומעקב על הריצה ע"י ה-debugger.

iii) לאחר שתסיימו לכתוב את הקוד לסעיפים 3)-4) שלחו לי את הסקריפט protein ואת כל הפונקציות.

iv) מאחר שתשתמשו בפקודות מתוך Statistical and Machine Learning Toolbox ו- Bioinformatics Toolbox, יש לוודא שהכלים מותקנים. בצעו במטלב פקודת ver שמציגה רשימה של הכלים המותקנים. אם הכלים הנ"ל לא מותקנים, הכנסו במטלב ללשונית APPS → get more apps וחפשו את הכלים. בחרו והתקינו.

2) **<u>להגשה עד סוף הסמסטר</u>**:

A) בצעו סעיף זה בשיעור תרגול **11**. כתבו פונקציה ל**בחירה וקריאה של קובץ pdb**. תוכלו להיעזר בקוד מדוגמא 1 בשיעור 8 או דוגמא 5 משיעור 7.

i) הפונקציה לא מקבלת קלט.

ii) <u>פלט</u>: <u>מבנה</u> עם שדות המתאימים לפורמט pdb.

iii) הפונקציה פועלת באופן הבא:

(1) מבקשת מן המשתמש לבחור קובץ (פקודת uigetfile עם פלט של שם הקובץ ושם התיקייה).

(2) מרכיבה את השם המלא של הקובץ (כולל path מקום בדיסק) וקוראת אותו בעזרת פקודת pdbread לתוך ארגומנט הפלט של הפונקציה.

B) בצעו סעיף זה בשיעור תרגול **11**. כתבו פונקציה ל**שליפת אטומים השייכים לחומר מסוים** לפי מזהה החומר, מתוך מערך מבנים של <u>אטומים כלשהם</u>. בהמשך תפעילו את הפונקציה כמה פעמים, בחלק מן המקרים על אטומים של חומצת אמינו ובחלק מן המקרים על אטומים של חומר זר. פעולת הפונקציה <u>זהה לשני סוגי האטומים</u>, כי המידע על שני הסוגים נמצא במבנים עם <u>שמות שדות זהים</u>. לכן יש <u>לכתוב את הפונקציה באופן כללי</u>, וכדאי לתת לארגומנטים ולמשתנים הפנימיים <u>שמות כלליים</u>, שלא יציינו האם אלה אטומים של חומצת אמינו או של חומר זר. לפני כתיבת הקוד, קראו אינטראקטיבית מקובץ מבנה של חלבון וצפו בשדות המצוינים למטה. כדאי גם להיעזר בקובץ ההדגמה d9pdb משיעור 9, כדוגמא לקוד שאוסף ערכים משדה במערך מבנים (בחלק מן המקרים נוצר מערך תאים ובחלק מן המקרים מערך רגיל של תווים). איסוף ערכים משדה במערך מבנים יש גם בדוגמא 3 משיעור 9.

i) <u>קלט</u>:

(1) מערך מבנים המתארים אטומים כלשהם. (הניחו שהתוכנית המפעילה את הפונקציה שלפה את נתוני האטומים מתוך כלל הנתונים על החלבון, ולכן הקלט הוא רק מערך מבנים עם נתוני אטומים).

(2) מזהה חומר (מערך תווים).

ii) <u>פלט</u>: מערך מבנים עם אותם שדות כמו בקלט, הכולל רק אטומים בעלי המזהה המבוקש.

iii) הפונקציה פועלת באופן הבא:

(1) אספו מכל המבנים את תוכן השדה **resName** וצרו <u>מערך תאים</u>.

(2) מחקו מכל השמות במערך התאים בסעיף (1) רווחים בהתחלה ובסוף (פקודת strtrim) ושמרו במשתנה.

(3) השוו בין המזהה לכל אחד מן התאים במערך במשתנה (2).

(4) בחרו רק את המבנים שעבורם התקבל "אמת" בהשוואה, ושמרו בארגומנט הפלט.

C) בצעו סעיף זה בשיעור תרגול **11**. כתבו סקריפט חדש **tst** לבדיקת הפונקציות בסעיף 2).

i) הפעילו את הפונקציה מסעיף A. שמרו את התוצאה במשתנה בשם שמציין שאלה <u>כלל הנתונים</u>.

ii) הציגו בחלון הפקודות את השדה hetID ממערך המבנים שבשדה HeterogenName במשתנה מהסעיף הקודם. (אם יש יותר מחומר זר אחד בדגם, יוצגו מספר מערכי תווים).

iii) הפעילו את הפונקציה מסעיף B על שדה HeterogenAtom שבתוך שדה Model במשתנה מסעיף i),

ועל השדה hetID באיבר הראשון במערך HeterogenName, ושמרו את התוצאה במשתנה.

iv) הריצו את הסקריפט, צפו בעורך המשתנים במשתנה ששמרתם בסעיף הקודם, ובדקו חזותית ששדה resName בכולם זהה לשם שבחרתם בהפעלת הפונקציה.

D) בצעו סעיף זה בשיעור תרגול **11**. כתבו פונקציה ל**בחירת השרשרת והחומר הזר** (ליגנד, Hetero) ע"י המשתמש.

i) קלט: מבנה בפורמט pdb כמצוין בסעיף 1)D(i).

ii) פלט: שני ארגומנטים מסוג מערך תווים: אחד מזהה לשרשרת ואחד מזהה לחומר הזר.

iii) לפני כתיבת הקוד, קראו אינטראקטיבית מקובץ מבנה של חלבון וצפו בשדות המצוינים למטה.

iv) הפונקציה פועלת באופן הבא:

(1) אספו מהשדה ChemName בשדה HeterogenName בכל המבנים במערך הקלט את שמות הליגנדים, וצרו מערך תאים.

(2) מחקו מכל השמות רווחים בהתחלה ובסוף (פקודת strtrim).

(3) בחרו חומר זר:

(a) אם יש רק תא אחד, אז יש רק חומר זר אחד, ותוכנו של התא הוא מזהה החומר הזר.

(b) אם יש יותר מאחד, יש לתת למשתמש לבחור מרשימת השמות ע"י פקודת listdlg (תוכלו להיעזר בדוגמא 9 משיעור 7,דוגמא 4 משיעור 8 או דוגמא 3 משיעור 9). שמרו במשתנה את מספר החומר שנבחר.

(c) שלפו מהאיבר באינדקס שנבחר במערך בשדה HeterogenName את תוכן השדה hetID.

(4) מחקו רווחים בהתחלה ובסוף (פקודת strtrim) ממזהה החומר הזר ושמרו בארגומנט הפלט לחומר הזר.

(5) מאחר שלא בכל השרשרות נמצאים תמיד כל סוגי החומרים הזרים, יש לבדוק באילו שרשרות נמצא החומר הזר שנבחר.

(a) הפעילו את הפונקציה לשליפת אטומים עם מזהה מסוים מסעיף 2)B) על שדה HeterogenAtom שבתוך שדה Model במשתנה הקלט, ועל מזהה החומר הזר שנבחר, ושמרו את התוצאה במשתנה.

(b) כדי לקבל רשימת שמות ממנה יבחר המשתמש, אספו מהשדה ChainID מכל האיברים במשתנה מהסעיף הקודם (5)(a) את מזהי השרשרת, צרו מערך-תאים ושמרו במשתנה.

(c) מאחר שיש יותר מאטום זר אחד בכל שרשרת, יש כפילויות במערך התאים. השתמשו בפקודת unique כדי לבחור רק תא אחד מכל ערך, ושמרו במערך. זוהי הרשימה של מזהי שרשרות ממנה יבחר המשתמש.

(d) אם נשאר רק תא אחד, אז תוכנו הוא מזהה השרשרת שנבחרה.

(e) אם יש יותר מתא אחד, השתמשו בפקודת listdlg כדי לתת למשתמש לבחור את השרשרת. שלפו לפי האינדקס שנבחר את תוכן האיבר המתאים ממערך התאים, ושמרו במשתנה הפלט למזהה השרשרת.

E) בצעו סעיף זה בשיעור תרגול **11**.

i) הוסיפו לסקריפט פקודות לבדיקת הפונקציה מסעיף D. הפעילו את הפונקציה מסעיף D על משתנה כלל הנתונים, כלומר המשתנה ששמרתם בסעיף 2)C(i. שמרו את הפלט במשתנים, אחד לקוד החומר הזר ואחד לקוד השרשרת. כתבו את הפקודה כך שהפלט יוצג בחלון הפקודות.

ii) הריצו את הסקריפט ובדקו ויזואלית שהמשתנים הם מערכי תווים (ולא תאים) ושהם מכילים את השמות שבחרתם כאשר הפעלתם את הפונקציה. הריצו את הסקריפט פעמיים: לבדיקת דגם שיש בו חומר זר אחד, בחרו בקובץ **pdb1vyo.ent**. לבדיקת דגם שיש בו יותר מחומר זר אחד, בחרו בקובץ **pdb3t6f.ent**.

F) בצעו סעיף זה בשיעור תרגול **11**. כתבו פונקציה ל**שליפת נתונים רלוונטיים לשרשרת מסוימת** (בהנחה שיש כמה שרשרות בדגם, הנחה שלרוב נכונה). בהמשך תפעילו את הפונקציה כמה פעמים, בחלק מן המקרים על אטומים של חומצת אמינו ובחלק מן המקרים על אטומים של חומר זר. פעולת הפונקציה זהה לשני סוגי האטומים, כי המידע על שני הסוגים נמצא במבנים עם שמות שדות זהים. לכן יש לכתוב את הפונקציה באופן כללי, וכדאי לתת לארגומנטים ולמשתנים הפנימיים שמות כלליים, שלא יציינו האם אלה אטומים של חומצת אמינו או של חומר זר.

i) קלט:

(1) מערך מבנים (שבכל אחד מהם יש שדה **chainID** שהוא תו בודד).

(2) מזהה של השרשרת (תו אחד).

ii) פלט: מערך מבנים עם אותם שדות כמו בקלט, אבל רק אלה שיש להם chainID מבוקש.

iii) הפונקציה פועלת באופן הבא:

(1) אוספת את תוכן השדה chainID מכל המבנים שבמערך, כך שנוצר מערך עמודה של תווים (**לא** מערך תאים).

(2) משווה את מערך העמודה לתו, כך שנוצר מערך לוגי.

(3) משתמשת במערך הלוגי כדי לשלוף מתוך מערך המבנים רק את אלה שעבורם התנאי הלוגי מתקיים.

G) בצעו סעיף זה בשיעור תרגול **11**. הוסיפו לסקריפט פקודות לבדיקת הפונקציה מסעיף F:

i) הפעילו את הפונקציה מסעיף F על שדה **Atom** שבתוך שדה **Model** במשתנה עם כלל המידע, ועל מזהה השרשרת שנבחר קודם, ושמרו את התוצאה במשתנה (בחרו שם שיסמן שאלה אטומים של החלבון).

ii) הפעילו את הפונקציה מסעיף F על שדה **HeterogenAtom** שבתוך שדה **Model** במשתנה עם כלל המידע, ועל מזהה השרשרת שנבחר קודם, ושמרו את התוצאה במשתנה (בחרו שם שיסמן שאלה אטומים של חומרים זרים).

iii) בדקו אינטראקטיבית את נכונות הפונקציה: צפו במשתנים שנשמרו בעורך המשתנים ובדקו שהשדה **chainID** בכמה מן המבנים במערך הוא מה שבחרתם עבור מזהה שרשרת.

H) בצעו סעיף זה בשיעור תרגול **12**. כתבו פונקציה **לקבל מן המשתמש בתיבת קלט** שני נתונים מספריים, בדומה לשאלה 2 בתרגיל 7. (תוכלו להיעזר גם בדוגמא 5 משיעור 8). הפונקציה מגיבה באופן הבא: בתחילה מוצגת תיבת קלט עם כמה שטחי הקלדה, כאשר בכל שטח מוצגים ערכי ברירת מחדל (ערכים אלה נלקחים מארגומנט הקלט לפונקציה). אם בשדה מסוים הוקלדו נתונים לא מספריים תוצג שוב את תיבת הקלט באופן הבא: בשטחי ההקלדה שבהם הוקלדו נתונים לא מספריים, תוצג הודעת שגיאה (או בקשה להזנת ערך מספרי). בשטחי ההקלדה שבהם הוזנו ערכים מספריים, יופיעו הערכים שהוזנו לאחרונה (ולא ערכי ברירת המחדל שהתקבלו בארגומנט הקלט). באופן זה, אם באחד משטחי ההקלדה הוזן בטעות ערך לא מספרי, המשתמש לא יצטרך להזין מחדש את הערכים המספריים בשטחי ההקלדה האחרים. התהליך נמשך כל עוד באחד משטחי ההקלדה יש נתונים לא מספריים.

i) הנתונים שיתקבלו מן המשתמש הם:

(1) מרחק מינימלי.

(2) מרחק מקסימלי.

ii) קלט: מבנה למרחקים, עם השדות כמתואר בסעיף 1) D) iii).

iii) פלט: מבנה עם שמות שדות זהים (אך עם ערכים שונים בתוך השדות הרלוונטיים, לפי הקלדת המשתמש).

iv) הפונקציה מקבלת מן המשתמש בתיבת קלט אחת את הערכים שצוינו למעלה, ע"י פקודת inputdlg, באופן הבא:

(1) הכינו מערך-תאים (cell array) עם טקסטים לכותרות מעל השטחים בהם יקליד המשתמש.

(2) הכינו מערך-תאים עם טקסטים שיוצגו בשטחי ההקלדה כערכי ברירת-מחדל: קחו את הערכים מהשדות המתאימים במבנה שבארגומנט הקלט. למשל אם שדה ההקלדה הראשון מתייחס לשדה "מרחק מינימלי" במבנה הקלט, אז בשדה ההקלדה הראשון (כלומר בתא הראשון במערך התאים) יוצג הערך שהתקבל בשדה זה בארגומנט הקלט. שימו לב שהערכים בשדות ההקלדה חייבים להיות טקסטים, לכן מספרים ממבנה הקלט יש להמיר לטקסטים.

(3) כתבו לולאת while שתתבצע בלי הפסקה (קריטריון העצירה ייבדק בתוך הלולאה). בגוף הלולאה:

(a) הציגו את תיבת הקלט ע"י פקודת inputdlg ושמרו את התוצאה במשתנה.

(b) המירו כל אחד מהטקסטים שהוקלדו ע"י המשתמש למספר, ושמרו בשדה המתאים במבנה הפלט. אם הוזנו תווים שלא ניתן לתרגם למספר, יתקבל מערך ריק או ערך NaN (תלוי בגרסה).

(c) אם אף אחד מן השדות במבנה הוא לא NaN או ריק, צאו מן הלולאה (פקודת break). בשלב זה, מבנה הפלט יכיל את הערך המבוקש.

(d) השתמשו במערך התאים שקיבלתם בסעיף (a) לערכי ברירת-מחדל שיוצגו בשטחי ההקלדה בפעם הבאה. בשטח ההקלדה שבו הופיעו תווים שלא ניתן לתרגם למספר, רשמו טקסט שמודיע על שגיאה או בקשה להזין מספר. (יש לרשום את הטקסט בשטח ההקלדה ולא ליצור חלונית-שגיאה נפרדת).

(4) ביציאה מן הלולאה, משתנה הפלט יכיל את הערך המבוקש.

I) בצעו סעיף זה בשיעור תרגול **12**. הוסיפו לסקריפט tst פקודות לבדיקת הפונקציה:

i) צרו מבנה (structure) עם שדות כמתואר בסעיף 1) D) iii), כאשר תוכן כל השדות הוא מערך ריק ושמרו במשתנה.

ii) אתחלו את השדה "מרחק מינימלי" למספר 5.

iii) אתחלו את השדה "מרחק מקסימלי" למספר 30.

iv) הפעילו את הפונקציה מסעיף H על המבנה שיצרתם ושמרו את הפלט במשתנה.

v) הריצו את הסקריפט כמה פעמים. בכל פעם הזינו בשטחי ההקלדה ערכים שונים (מספריים או לא מספריים) ובדקו שהתגובה בהרצת הפונקציה היא כמתואר למעלה ושהערכים שמתקבלים בפלט הם מספרים בערכים שהזנתם.

J) בצעו סעיף זה בשיעור תרגול **12**. כתבו פונקציה ל**יצירת מערך-קואורדינטות נומרי מתוך שדות של מערך מבנים** של אטומים.

i) קלט: מערך מבנים שבכל אחד מהם קיימים השדות X,Y,Z, בכל אחד משדות מספר אחד.

ii) פלט: מטריצה עם 3 עמודות.

iii) הפונקציה פועלת באופן הבא:

(1) איספו מן המבנים שבקלט את השדה X וצרפו את המספרים למערך-עמודה. שמרו במשתנה.

(2) חזרו על הפעולה עבור השדות Y,Z.

(3) צרפו את שלושת המשתנים למערך דו-ממדי של 3 עמודות. זה משתנה הפלט.

K) בצעו סעיף זה בשיעור תרגול **12**. הוסיפו לסקריפט tst פקודות לבדיקת הפונקציה:

i) שלפו את 10 האיברים הראשונים ממערך Atom שבתוך שדה Model במשתנה כלל הנתונים ושמרו במשתנה.

ii) הפעילו את הפונקציה מסעיף J על המשתנה ששמרתם ושמרו את הפלט במשתנה. הציגו את המשתנה בחלון הפקודות.

iii) פתחו בעורך המשתנים את מערך המבנים מסעיף i), צפו בשדות X,Y,Z והשוו לפלט בחלון הפקודות.

iv) השתמשו בפלט של הפונקציה כדי לצייר גרף scatter של עשרת האטומים. תנו כותרות לצירים, וסובבו ידנית את התמונה כדי לקבל מושג על מקומות האטומים.

3) **כתיבת התוכנית הראשית**: בסעיף זה תכתבו ותפעילו פונקציות לזיהוי אתר הקשר. הפונקציות יופעלו מתוך התוכנית הסופית **protein** – אפשר להיעזר בקוד שכתבתם בסקריפט tst, אבל הסקריפט tst עצמו איננו חלק מן התוכנית הסופית.

A) ניתן לזהות את האטומים של כל חומצת-אמינו בשרשרת לפי מספר ייחודי שנמצא בשדה **resSeq**. שדה זה מכיל מזהה של כל חומצת אמינו, שהוא חד-ערכי (כלומר גם אם חומצת אמינו מסוימת מופיעה כמה פעמים בחלבון, לכל מופע יש מזהה אחר). לכל האטומים של המופע הספציפי יש אותו מזהה. כתבו פונקציה ל**זיהוי האטומים של התחלת כל חומצת-אמינו** וכן את השמות שלהן.

i) קלט: מערך מבנים עם מידע על אטומים של חומצות-אמינו בשרשרת מסוימת.

ii) פלט:

(1) מערך נומרי עם אינדקסים של האטום הראשון בכל חומצת-אמינו Residue.

(2) מערך נומרי עם אינדקסים של האטום האחרון בכל חומצת-אמינו Residue.

iii) הפונקציה פועלת באופן הבא:

(1) "אוספת" מכל המבנים את השדה resSeq שהוא סקלר נומרי, כך שנוצר מערך עמודה של מספרים (**לא** מערך תאים).

(2) מצאו את כל המקומות שבהם הערך של המזהה משנה ערך (כלומר שונה מן האיבר שאחריו, השתמשו בפקודת find כדי למצוא את האינדקסים של מקום השינוי) ושמרו במשתנה.

(3) מקומות השינוי הם הסוף של חומצות האמינו. צרפו אחרי המערך ששמרתם בסעיף (2) את האינדקס של האטום האחרון של חומצת האמינו האחרונה (שהוא אורך מערך הקלט) ושמרו את התוצאה במשתנה הפלט השני.

(4) המקומות אחרי השינוי הם ההתחלות של חומצות האמינו. הוסיפו 1 לכל איברי המערך ששמרתם בסעיף (2) וצרפו לפניו את האינדקס של האטום הראשון של חומצת האמינו הראשונה (שהוא 1). שמרו את התוצאה במשתנה הפלט הראשון.

B) סעיף זה יש לבצע **אחרי שיעור 13**. במקרים בהם שם החומר הזר ארוך מאוד, הוא מסודר בנתונים כמטריצת תווים (עם רווחים בסוף שורה, כדי להשלים למלבן). כתבו פונקציה **להעביר מטריצה לשורה אחת**. תוכלו להיעזר בקוד מדוגמא 3 בשיעור 7.

i) קלט: מבנה מסוג HeterogenName, שאחד השדות שלו הוא ChemName.

ii) פלט: מבנה עם שמות שדות זהים, שבשדה ChemName יש מערך שורה של תווים.

iii) הפונקציה פועלת באופן הבא:

(1) העתיקו את מבנה הקלט למבנה הפלט.

(2) שלפו את התווים מהשדה ChemName והפכו שורות לעמודות (צרו מטריצה מוחלפת).

(3) הפכו את התוצאה למערך-עמודה.

(4) הפכו את העמודה לשורה.

(5) הסירו מהתוצאה רווחים בהתחלה ובסוף (פקודת strtrim) ושמרו בשדה ChemName במבנה הפלט.

C) סעיף זה יש לבצע **אחרי שיעור 13**. פתחו סקריפט חדש **protein** וכתבו בו, לפי <u>ההוראות הבאות</u>, פקודות להכנת מבנה ל<u>נתונים של החלקים שנבחרו ע"י המשתמש</u>, עם שדות כמתואר בסעיף 1) D) ii).

i) מחקו את כל המשתנים וסגרו את כל החלונות הגרפיים.

ii) כתבו פקודה להפעלת הפונקציה לבחירה וקריאה של קובץ שנכתבה בסעיף 2) A) ושמרו את התוצאה במשתנה. בחרו שם לציין שאלה <u>כלל הנתונים</u>.

iii) הפעילו את הפונקציה להעברת מטריצה לשורה אחת שנכתבה בסעיף 3)B) על מערך המבנים שבשדה HeterogenName ושמרו את התוצאה באותו שדה. (פעולה זו תתקן את שמות החומרים הזרים, במקרה שחולקו לכמה שורות). מאחר שהפונקציה בסעיף 3)B) פועלת על <u>מבנה אחד</u>, השתמשו בפקודה arrayfun כדי להפעיל אותה <u>על כל מבנה לחוד במערך של מבנים</u>. להבנת השימוש בפקודה, היעזרו בדוגמא 2 משיעור 13.

iv) הפעילו על המשתנה מסעיף iii) את הפונקציה מסעיף 2) D) לבחירת השרשרת והחומר הזר.

v) הקצו <u>מבנה חדש</u> לנתונים שנבחרו ע"י המשתמש עם שדות כמתואר בסעיף 1)D))ii) ושמרו את מזהה הליגנד ואת המזהה לשרשרת החלבון שנבחרה בשדות הרלוונטיים.

vi) הפעילו את הפונקציה לשליפת נתונים משרשרת מסוימת שנכתבה בסעיף 2) F) על שדה Atom שבתוך שדה Model במשתנה מסעיף iii) ועל מזהה השרשרת שנבחרה, ושמרו את התוצאה בשדה אטומי שרשרת החלבון במבנה הנתונים.

vii) הפעילו את הפונקציה לשליפת נתונים משרשרת מסוימת שנכתבה בסעיף 2) F) על שדה HeterogenAtom שבתוך שדה Model במשתנה מסעיף iii) ועל מזהה השרשרת שנבחרה, ושמרו את התוצאה במשתנה זמני.

viii) הפעילו את הפונקציה לשליפת אטומי חומר מסוים שנכתבה בסעיף 2) B) על המשתנה הזמני מהסעיף הקודם vii) ועל מזהה הליגנד שנבחר ושמרו את התוצאה בשדה אטומי הליגנד במבנה הנתונים.

ix) הפעילו את הפונקציה לזיהוי התחלת חומצות האמינו שנכתבה בסעיף 3) A) על שדה אטומי שרשרת החלבון במשתנה מסעיף v) ושמרו את התוצאות בשדות: אינדקס אטום ראשון ואינדקס אטום אחרון של חומצות האמינו במשתנה מסעיף v).

x) מצאו את שמות חומצות האמינו:

(1) שלפו משדה אטומי שרשרת החלבון במשתנה מסעיף v) רק את המבנים של האטומים <u>הראשונים</u> בכל חומצת אמינו.

(2) "איספו" מכל המבנים האלה את השדה resName שהוא מערך תווים, צרו מערך-תאים ומחקו מכל השמות רווחים בהתחלה ובסוף (פקודת strtrim). שמרו את התוצאה בשדה שמות חומצות האמינו במשתנה מסעיף v).

xi) מומלץ להריץ את הסקריפט כמה פעמים ולצפות במשתנה מסעיף v) בעורך המשתנים. לבדיקת דגם שיש

בו חומר זר אחד, בחרו בקובץ **pdb5b5f.ent**. לבדיקת דגם שיש בו יותר מחומר זר אחד, בחרו בקובץ **pdb3vgw.ent**.

D) כתבו פונקציה ל**שליפת נתוני קואורדינטות השדרה backbone** של שרשרת.

i) קלט: מבנה לנתונים שנבחרו ע"י המשתמש כמתואר בסעיף 1) D) ii).

ii) פלט: מערך מספרי דו ממדי עם שלוש עמודות (עבור הקואורדינטות X,Y,Z) של אטומי השדרה בלבד: שלושת האטומים הראשונים של כל חומצה אמינית.

iii) הפונקציה פועלת באופן הבא:

(1) שלפו משדה "אינדקס של אטום ראשון" את המערך ושמרו במשתנה. מספר האיברים במערך הוא מספר חומצות האמינו בשרשרת.

(2) צרו מערך של 3 שורות, שאורך כל אחת הוא כמספר חומצות האמינו בחלבון: בשורה הראשונה סדרו את האינדקסים ששמרתם בסעיף (1). בשורה השניה סדרו את האינדקסים, אבל הוסיפו להם 1. בשורה השלישית סדרו את האינדקסים, אבל הוסיפו להם 2. במטריצה שקיבלתם נמצאים כל האינדקסים של אטומי השדרה.

(3) שנו את הממדים של המערך הדו ממדי מסעיף (2) למערך-עמודה חד-ממדי. מערך זה כולל את האינדקסים של אטומי השדרה לפי הסדר הנכון.

(4) שילפו את המבנים השייכים לשדרה מתוך שדה אטומי שרשרת החלבון שבמבנה הקלט.

(5) השתמשו בפונקציה מסעיף 2) J) כדי לשלוף את הקואורדינטות מתוך המבנים במערך שמצאתם בסעיף (4). אלה הקואורדינטות של אטומי השדרה, ארגומנט הפלט.

E) סעיף זה יש לבצע **אחרי שיעור 13**. הוסיפו לסקריפט פקודות למציאת השדרה של השרשרת שנבחרה ולשרטוטה:

i) הפעילו את הפונקציה מסעיף 3) D) לשליפת הקואורדינטות של השדרה על שדה אטומי שרשרת החלבון במבנה מסעיף 3) C), ושמרו את התוצאה במשתנה (בחרו שם שמציין שאלה אטומי השדרה).

ii) פתחו חלון גרפי חדש (מס' 1) ציירו עקומה תלת ממדית של השדרה, במרקרים מחוברים בקו. בחרו צבע בהיר (לדוגמא אפור בהיר או תכלת וכו').

iii) הוסיפו כותרות לצירים וקבעו את הצירים ליחס 1:1:1 (פקודת axis).

iv) הוסיפו לגרף כותרת עם שם הקובץ ומזהה השרשרת, מהצורה הבאה: file *file-name* chain *chain-ID*

F) כתבו **פונקציה לחישוב המרחקים** בין כל האטומים של השרשרת (של חומצות האמינו) לכל האטומים של הליגנד.

i) קלט: מבנה לנתונים שנבחרו, עם שדות כמתואר בסעיף 1) D) ii).

ii) פלט: מערך מספרי דו-ממדי שהוא טבלת המרחקים, כך שהאיבר שהאינדקסים שלו הם (i, j) הוא מרחק בין אטום החלבון מס' j מאטום הליגנד מס' i. (להמחשה, תוכלו להיעזר בשרטוט המרחקים שנמצא בתיאור הכללי של התרגיל המסכם).

iii) הפונקציה פועלת באופן הבא:

(1) הפעילו את הפונקציה ליצירת מערך הקואורדינטות שנכתבה בסעיף 2) J) על שדה אטומי השרשרת ושדה אטומי הליגנד כדי לשלוף מהם את הקואורדינטות, ושמרו את התוצאות במשתנים.

(2) חשבו מן המערכים את המרחקים ההדדיים ושמרו במשתנה הפלט. תוכלו להשתמש בפונקציה d10dist מההרצאה.

G) כתבו פונקציה ל**זיהוי חומצות האמינו שנמצאות בקרבת החומר הזר**.

i) קלט:

(1) מבנה לנתונים שנבחרו, עם שדות כמתואר בסעיף 1) D) ii).

(2) מטריצת מרחקים הדדיים.

(3) מרחק מינימלי (אורך קשר).

ii) פלט: אינדקסים של חומצות האמינו שנמצאות בקרבת החומר הזר.

iii) הפונקציה פועלת באופן הבא:

(1) מצאו במטריצת המרחקים ההדדיים את כל העמודות שיש בהן לפחות מרחק אחד שקטן מהמרחק המינימלי. לשם כך השוו את המרחקים למרחק המינימלי וקבלו מערך לוגי. השתמשו בפקודת any כדי למצוא באילו עמודות יש לפחות איבר אחד שהוא אמת לוגית. יתקבל מערך לוגי חד-ממדי.

(2) השתמשו במערך הלוגי כדי לשלוף ממערך האטומים של השרשרת את המבנים המתאימים לעמודות שנמצאו בסעיף (1). אלה הם האטומים של אתר הקשר.

(3) מצאו את המזהים של חומצות האמינו שהאטומים בסעיף (2) שייכים אליהן. לצורך כך יש לאסוף מכל המבנים של האטומים באתר הקשר את השדה **resSeq** וליצור מערך מספרי (**לא** מערך תאים).

(4) השתמשו בשדה האינדקסים של האטום הראשון מכל חומצת אמינו. שלפו משדה אטומי השרשרת במבנה הקלט הראשון, את המבנה של האטום הראשון מכל חומצת אמינו. יתקבל מערך מבנים.

(5) שלפו ממערך המבנים שנשמר בסעיף (4) את השדה **resSeq** וצרו מערך מספרי (**לא** מערך תאים).

(6) השתמשו בפקודת ismember כדי למצוא אילו מן המזהים בסעיף (5) נמצאים במערך מסעיף (3). יתקבל מערך לוגי: עבור כל מזהה (חומצת אמינו), האם נמצא במערך חומצות האמינו המשתתפות בקשר.

(7) השתמשו בפקודת find כדי למצוא היכן במערך (6) יש ערכים "אמת". האינדקסים שנמצאו הם הפלט.

4) **סיום** (לביצוע אחרי שיעור 13): בסעיף זה תכתבו ותפעילו פונקציות להצגה גרפית של החלבון ואתר הקשר.

A) הוסיפו לסקריפט protein פקודות שמציירות את האטומים הזרים על שרטוט של השדרה (backbone), כמוסבר למטה:

i) הפעילו את הפונקציה ליצירת מערך קואורדינטות שנכתבה בסעיף 2)J) על מערך המבנים של האטומים הזרים, ושמרו במשתנה את הקואורדינטות שלהם.

ii) הוסיפו לגרף של השדרה את האטומים של החומר הזר במרקרים שחורים בלתי מחוברים. השתמשו בפקודת scatter3.

B) כתבו פונקציה ל**הצגה גרפית של המרחקים ההדדיים**, יחד עם אנוטציה של אטומי החומר הזר ושל חומצות האמינו לאורך השרשרת. תוכלו להיעזר בקוד מדוגמא 2 בהרצאה 10, דוגמא 7 בהרצאה 12, ושאלה 2 בתרגיל 10. (להמחשה, תוכלו להיעזר בשרטוט המרחקים שנמצא בתיאור הכללי של התרגיל המסכם).

i) קלט:

(1) מבנה לנתונים שנבחרו ע"י המשתמש, עם שדות כמתואר בסעיף 1)D(ii).

(2) טבלת מרחקים.

(3) מבנה למרחק מינימלי ומקסימלי, כמתואר בסעיף 1)D(iii.

ii) אין פלט.

iii) הפונקציה פועלת באופן הבא:

(1) מייצרת טבלת מרחקים "מורחבת" שהיא הטבלה המקורית, בתוספת שכפול של השורה האחרונה והעמודה האחרונה (כדי לגרום להצגת השורה והעמודה האחרונה). תוכלו להיעזר בשאלה 2 מתרגיל 5.

(2) מציירת משטח צבעוני (pcolor) של טבלת המרחקים ומבטלת את קווי הרשת. בחרו מפת צבעים **שבה הצבע האחרון לבן** (או בהיר מאוד), למשל hot או gray (או יצרו בעצמכם מפה כזאת).

(3) קבעו את מיפוי הצבעים כך שהמינימום והמקסימום (פקודת caxis) הם כמצוין בשדות מינימום ומקסימום בארגומנט הקלט השלישי. באופן זה, חומצות אמינו שנמצאות במרחק גדול יותר מן המקסימום ייצרו תאים בהירים ב"מפת החום", ואילו אזורים הקרובים לחומר הזר מן המינימום יופיעו ככתמים כהים.

(4) הוסיפו מקרא לצבעים colorbar.

(5) תנו לציר y כותרת שהיא השם של החומר הזר. השם מופיע בכל האטומים של החומר הזר בשדה resName.

(6) תנו כותרת גם לציר x שהיא קוד השרשרת שנבחרה: chain *chain-code*.

(7) שלפו מכל האטומים של החומר הזר את השדה AtomName ויצרו מערך-תאים של שמות האטומים של החומר הזר.

(8) הוסיפו את המידע כשנתות לצירים באופן הבא:

(a) שמרו את תכונות מערכת הצירים במשתנה (פקודת gca). לפני כתיבת הפקודות הבאות, הריצו פקודת pcolor על מטריצה מספרית כלשהי וצפו במשתנה שכולל מידע על הצירים. (שימו לב שהקשה כפולה על שם המשתנה פותחת חלונית inspector ולא חלונית רגילה של עורך משתנים).

(b) שנו את תכונות XTick כך שהשנתות ייקבעו במקומות של אחד האטומים הראשונים של כל חומצת אמינו (שדה אינדקסים של האטום הראשון בארגומנט הקלט הראשון).

(c) שנו את תכונת XTickLabel כך שליד השנתות ייכתבו שמות חומצות האמינו. ניתן לסובב את הטקסט (כדי לתת יותר רווח בין הכותרות) ע"י שינוי התכונה XTickLabelRotation.

(d) באופן דומה, שנו את מקום השנתות על ציר y כך שתהיה שנת <u>לכל אטום</u> של החומר הזר, ולידה ייכתב שם האטום הזר. (שמות התכונות דומים, החליפו X ל-Y בכל מקום).

C) הוסיפו לסקריפט פקודות לחישוב המרחקים והצגתם:

i) פתחו חלון גרפי נוסף (מס' 2).

ii) הפעילו את הפונקציה לשליפת נתונים רלוונטיים לשרשרת שנכתבה בסעיף 3) F) על מבנה ה<u>נתונים שנבחרו ע"י המשתמש</u> שנוצר בסעיף 3)C) מסעיף v) ואילך כדי לקבל טבלת מרחקים. שמרו את התוצאה במשתנה.

iii) חשבו את ערך המינימום ואת ערך המקסימום של <u>כלל המרחקים</u> שחושבו, ושמרו במשתנים. (שימו לב שהמרחקים הם מטריצה ולכן במציאת המינימום יש לקבל סקלר ולא מערך).

iv) צרו מבנה למרחקים עם שדות כמתואר בסעיף 1)iii(D), כאשר שדה המרחק המינימלי מאותחל לערך המינימלי של <u>כלל המרחקים</u> ושדה המרחק המקסימלי מאותחל לערך 30.

v) פתחו חלון גרפי חדש (מס' 3).

vi) כתבו לולאה אינסופית (יציאה מן הלולאה כמתואר למטה). בגוף הלולאה:

(1) עברו לחלון גרפי מס' 3 והציגו בו היסטוגרמה באופן הבא. תוכלו להיעזר בדוגמא 5 מהרצאה 12.

(a) חשבו את מערך המרווחים עבור ההיסטוגרמה: מערך במרווחים קבועים של 21 איברים, שהראשון שבהם הוא ערך המינימום שבמבנה למרחקים עם שדות כמתואר בסעיף D(1)iii), והערך האחרון הוא מקסימום המרחקים כפי שחושב בסעיף iii).

(b) הפעילו את פקודת ההיסטוגרמה על כלל המרחקים שחושבו, ועם המרווחים שחושבו בסעיף (a).

(2) השתמשו בפונקציה מסעיף H(2) לקליטה מן המשתמש של ערכי מינימום ומקסימום ושמרו את התוצאה במשתנה בו השתמשתם כקלט לפונקציה.

(3) עברו לחלון הגרפי מס' 2 והפעילו את הפונקציה להצגת המרחקים שנכתבה בסעיף B (4) עם הנתונים מהסעיפים הקודמים כדי להציג את טבלת המרחקים.

(4) שאלו את המשתמש האם לשנות את מרחק המינימום והמקסימום (תוכלו להשתמש בפקודה listdlg מחומר הקורס, או לקרוא בתיעוד על פקודת questdlg).

(5) אם המשתמש בוחר "לא", יש לצאת מן הלולאה (פקודת break). אחרת, הלולאה ממשיכה להתבצע והמשתמש יכול לשנות את ערכי המינימום והמקסימום (להקטין או להגדיל, בהתאם לגודל של הכתמים הכהים במפת המרחקים).

vii) לאחר היציאה מן הלולאה, המבנה למרחקים מכיל את הערכים הסבירים לפי המשתמש.

viii) הוסיפו לגרף כותרת כללית, שהיא קוד הדגם. המידע נמצא במבנה שנקרא מהקובץ, בשדה idCode בתוך שדה Header .

D) הוסיפו לסקריפט פקודות לציור האטומים של חומצות האמינו שמשתתפות באתר הקשר, באופן הבא:

i) הפעילו את הפונקציה לזיהוי חומצות האמינו באתר הקשר שנכתבה בסעיף G(3) על המרחק המינימלי שבמבנה המרחקים, ושמרו במשתנה את הפלט (אינדקסים של חומצות האמינו שמשתתפות בקשר). אורך המערך הוא מספר חומצות-האמינו שנמצאו.

ii) הכינו מפת-צבעים לציור האטומים של חומצות-אמינו אלה, עם מספר צבעים כמספר החומצות שנמצאו. בחרו צבעים שאינם בהירים מדי (למשל צהוב) כך שכל האטומים המשורטטים ייראו היטב על רקע לבן. תוכלו להשתמש באחת ממפות-הצבעים המוכנות של מטלב. (הסבר בשיעור 10).

iii) עברו לחלון הגרפי הראשון.

iv) השתמשו באינדקסים של חומצות האמינו של אתר הקשר כדי לשלוף ממערכי התחלת וסוף חומצת האמינו, רק את האינדקסים של התחלה וסוף של חומצות האמינו שבאתר הקשר. שמרו את התוצאה במשתנים.

v) כתבו לולאה לעבור על כל חומצת-אמינו ולצייר את האטומים בצבע ספציפי לכל חומצת-אמינו. בגוף הלולאה:

(1) שלפו ממערכי האינדקסים שנשמרו בסעיף iv) את אינדקס האטום בו מתחילה החומצה הרלוונטית ואת אינדקס האטום בו היא מסתיימת. צרו מערך של מספרים שלמים עוקבים שמתחיל באינדקס ההתחלה ומסתיים באינדקס הסיום.

(2) שלפו ממערך האטומים של השרשרת את האטומים עם האינדקסים מהסעיף הקודם.

(3) השתמשו בפונקציה מסעיף J(2) כדי לשלוף את קואורדינטות האטומים האלה מן המערך מסעיף (2).

(4) שלפו את הצבע המתאים ממערך מפת הצבעים (שורה עם אינדקס כמספר החומצה).

(5) שרטטו גרף תפזורת של הקואורדינטות (השתמשו בפקודת scatter3). השתמשו בצבע מסעיף (4). מומלץ להוסיף לפקודה ארגומנט "filled" כדי שהמרקרים יהיו מלאים.

vi) צרו מערך-תאים עם טקסטים שהם כותרות לעקומות:

(1) לעקומה הראשונה: תא שמכיל את התווים backbone *chain-ID*

(2) לעקומה השניה: תא שמכיל את התווים של קוד החומר הזר.

(3) לשאר העקומות: תאים שמכילים את שמות חומצות האמינו שמשתתפות בקשר. השתמשו במשתנה האינדקסים מסעיף 4)D(i) כדי לשלוף משדה שמות חומצות האמינו רק את השמות הרלוונטיים: שלפו מכל אטומי השרשרת את האטומים עם האינדקסים שנמצאו בסעיף הקודם, ואספו מכולם את השדה resName למערך-תאים.

(4) שרשרו את התאים משלושת הסעיפים הקודמים למערך-תאים אחד.

vii) השתמשו במערך התאים מסעיף vi) כדי להוסיף מקרא לעקומות.

סוף