

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 pdbname 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 pdbname. לפני ביצוע התרגיל, מומלץ לבצע פקודת pdbname 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 pdbname (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. קורס מטלב לביולוגים תרגיל מסכם מס' 2 עמ' 1 מתוך 13

(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) להגשה עד סוף הסמסטר:

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

(i) הפונקציה לא מקבלת קלט.

(ii) פלט: מבנה עם שדות המתאימים לפורמט pdb.

(iii) הפונקציה פועלת באופן הבא:

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

(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 לשליפת אטומים השייכים לחומר מסוים לפי מזהה החומר,

מתוך מערך מבנים של אטומים כלשהם. בהמשך תפעילו את הפונקציה כמה פעמים, בחלק מן המקרים על

אטומים של חומצת אמינו ובחלק מן המקרים על אטומים של חומר זר. פעולת הפונקציה זוהי 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 יצינו 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) קלט:

(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) פלט: 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 מערך תאים.

(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 כלל הנתונים.

(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,

קורס מטלב לביולוגים תרגיל מסכם מס' 2 עמ' 2 מתוך 13

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 **לבחירת 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 B(2) 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 **pdb1vvo.ent**. לבדיקת דגם שיש בו יותר מחומר זר אחד, בחרו בקובץ **pdb3t6f.ent**.

(F) בצעו סעיף זה בשיעור תרגול 11. כתבו פונקציה **לשליפת 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 יצינו 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 תיבת 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 שתבצע 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 ליצירת 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.

המשך יבוא...