עיבוד תלת ממדי במבנים ביולוגיים 76562

דינה שניידמן וברק רווה

הרצאה 1

הרצאה 2

הרצאה 3

הרצאה 4

הרצאה 5 – **שיטות אופטימיזציה עבור שיטות קיפול חלבונים חלק א׳**

A picture containing graphical user interface

Description automatically generatedקיפול חלבונים הוא התהליך שבו רצף של חומצות אמינו יוצר צורה תלת ממדית. חלבונים מתקמפלים במילי-שניות שזהו פרדוקס בפני עצמו.

**פרדוקס לבנטל**

נתייחס אל חלבון קטן שמורכב מ100 חומצות אמינו. יהיו לנו 99 קשרי פפטידים ו198 זוויות דאדרליות .

נניח ויש לנו 3 מצבים אפשריים עבור כל זווית דאדרלית, יש לנו אפשרויות לקיפול המבנה, אבל חלבונים מתקפלים מהר – פרדוקס בפני עצמו.

נגדיר מספר הגדרות –

* הידרופוביות – לא נוטות ליצור קשרי מימן עם מים.
* פולאריות – טעונות במטען חשמלי, ״אוהבות מים״.

Diagram

Description automatically generatedיש לנו חומצות אמינו הידרופוביות ויש לנו חומצות אמינו פולאריות, והרי חלבונים מורכבים מחומצות אמינו. מה יקרה אם חלבון שמכיל חומצות אמינו הידרופוביות ופולאריות? חלבונים הרי חיים במים, על כן החומצות ההידרופוביות ירצו להיות בפנים והפולאריות ירצו להיות ״בחוץ״, ולכן בהסבר פשוט נאמר שזו הסיבה שחלבונים מתקפלים. בפועל כאשר חומצות הידרופוביות מופנות כלפיי פנים האנרגיה של המבנה יורדת, ובאותה המידה כאשר חומצות פולאריות מופנות כלפי מים האנרגיה של המבנה יורדת. מערכות ביולוגיות אוהבות להיות במצב של אנרגיה מזערית – לכן חלבון מתקפל.

בעיית קיפול חלבונים ניתנת להסתכלות בתור בעיה אלגוריתמית.

**קלט –** רצף של חומצות אמינו.

**פלט –** מבנה תלת ממדי.

יש לנו 2 גישות מרכזיות למידול החלבונים בקיפול –

1. מידול הומולוגי(קומפרטיבי)
2. מידול ab initio (בלטינית: מהתחלה)

בצורת מידול זו נקבל רק את הרצף וממנו נרצה לקבל את המבנה.

נדבר על המידול ab inition. השיטות תחת המידול כנ״ל מנסות למצוא את המבנה הטבעי(הקיפול הטבעי) של חלבון על ידי סימלוץ התהליכים הביולוגים בקיפול החלבון.

+ יתרונות בגישה – יותר כללי, ניתן להרצה גם כאשר אין מודל הומולוגי מוכר לחלבון.

– חסרונות בגישה – הנכונות והסימוכין שלנו על המידול שמסתפק איננו גבוה.

* *Diagram

  Description automatically generated*למדנו על תצוגות שונות אפשריות של חלבונים, ביניהן ייצוג קואורדינטות פנימיות(זוויות דאדרליות).  
  נניח ויש לנו 200 חומצות אמינו, לכל חומצת אמינו זווית , ולכן ממד מרחב הקונפיגורציות הוא 398.

**הגדרה – Configuration space**

מרחב ווקטורי עם סט של קוארדינטת שמתאר את החלבון(או מערכת כלשהיא). ווקטורים ממפה מספר מערכות(רובוט, חלבון, פוזיציה של גוף האדם).   
הקואורדינטות של בד״כ נקראות ״system variables״ כלומר משתני מערכת. ניתן לקרוא להם גם פרמטרים או קואורדינטות כלליות.

חלבון עם 1000 אטומים, ממד מרחב הקונפיגורציות שלו במרחב קרטזי הוא 3000, כי עבור כל אטום יש תצוגה של 3 קואורדינטות. נזכיר כי יש לנו 6 דרגות חופש שנוכל להוריד(כי לא משנה אם מזיזים או מסובבים את החלבון) ולכן נוכל לדייק ולומר 2994 קואורדינטות בלתי תלויות.

**נרצה להגדיר את בעיית ה-ab initio בתור בעיית חיפוש.**

**קלט** – רצף החלבון s, פונקציית אנרגיה ממרחב הקונפיגורציות לממשיים. כל קונפיגורציה תקבל ערך סקלרי.

**פלט** – קונפיגורציה שמספקת לנו מינימום עבור .

למעשה אנחנו רוצים למצוא מינימום לפונקציית האנרגיה הפוטנציאלית של החלבון המקופל.

למה נרצה למזער את האנרגיה של הקונפיגורציה? כי מערכות ביולוגיות אוהבות להיות באנרגיה נמוכה ככל האפשר.

איך נבנה פונקציית אנרגיה?

האטומים מיוצגים בתור ״כדורים״. נרצה שהם לא יתנגשו, נוכל לעשות זאת בעזרת מינימיזציה של סכום פונקציית אינדיקטור עבור כל זוג.

היינו יכולים לעשות זאת בעזרת מרחק אוקלידי: בעזרת העלאת פונקציית האנרגיה תחת מרחק אוקלידי מתאפס של זוג אטומים(הרי כך הם מתנגשים).

נניח ויש לנו 2 מרכזי כדורים –

נרצה פונקציה רציפה . נתבונן על המרחק בין 2 המרכזים –

נוכל להסתכל על הגרף – , גרף רציף ומונוטוני יורד, הרי נרצה לקבל ״מינימום״ לפונקציה כדי שהכדורים לא יתנגשו.   
נניח ורדיוסי הכדורים הם בהתאמה. נרצה להגדיר את כך –

*Diagram

Description automatically generated*

*כאשר במקרה כנ״ל אמנם d קבוע וכך נראת פונקציית האנרגיה תחת קונפיגורציה ספציפית, אבל מיקומי מרכזי הכדורים אינם קבועים כמובן כי יש לנו מצבור הקונפיגורציות.*

*בפועל נבנה את פונקציית האנרגיה בעזרת –*

* *שיקולים פיזיקליים*
* *סטטיסטיקה*
* *למידת מכונה*

*ניתן להראות כי בעיית קיפול החלבונים היא בעיה NP קשה אפילו ב2D, אבל באופן מעשי אנחנו יכולים לפתור את הבעיה באמצעות מרחבים אנרגטיים.*

*שיטות מקובלות הן Molecular Dynamics, Monte Carlo וגם בעזרת למידה עמוקה.*

*אנחנו נדון* ***בדינמיקה מולקולרית*** *בהמשך הקורס, בקצרה אנחנו מנסים לסמלץ את החוקים של ניוטון. אנחנו מנסים לסמלץ על פי החוקים של ניוטון את תזוזת האטומים, אך הבעיה בפועל היא שחלבונים תמיד נתקעים בנקודות מינימום לוקליים. איך נתקעים במינימום לוקליים? מגרילים קונפיגורציה רנדומית ובודקים את האנרגיה שלה. ננסה לשפר את האנרגיה של החלבון עד שנגיע ל-״מבנה הנכון״, אבל במקרה כנ״ל אנחנו עלולים להיתקע במינימום לוקלי.*

*לעומת זאת* ***נדון ב-Monte Carlo*** *בשיעורים הקרובים. באופן מקוצר אנחנו דוגמים קונפיגורציות לא בהכרח בעזרת חוקי הפיזיקה. אנחנו מסמלצים את הקיפול של החלבון אבל לא בצורה שמנסה לחכות את הפיזיקה. השיטה יותר פרקטית, ונפוצה גם במידול הסתברותי, באסיאני ותחומים נוספים. יש לה פחות סיכוי להיתקע במינימום לוקלי אך לא בהכרח שלא תיתקע.*

***שיטות מינימיזציה לוקלית***

* גרדיאנטים ונגזרות

*נרצה שפונקציית האנרגיה שלנו תהיה בקורלציה לפונקציית האנרגיה החופשית. בנוסף נרצה שהפונקציות יחלקו את אותו מינימום גלובלי אנרגטי, מה גם נצטרך גם אסטרטגיית דגימה מבין כל הקונפיגורציות הקיימות.*

*בהרחבת פונקציות לממדים גבוהים, נרצה לסרוק את כל הקונפיגורציות בכמה ממדים שיש לנו ולמצוא את הקונפיגורציה האופטימלית.*

*משטחי אנרגיה הם לא חלקים – הם מפוספסים, אז קשה לנו למצוא את המינימום הגלובלי.*

*Diagram, schematic

Description automatically generated*

*החספוס יגרום לנו להיתקע בנקודות מינימום לוקליים מאוד מהר, בעזרת ההיתקעות שלנו נרצה לשפר את מבנה החלבון ולא רק לקפלו.*

*כזכור גרדיאנטים והסיאנים מכלילים את הנגזרות הראשונה והשנייה עבור פונקציות רב ממדיות.*

*בהינתן פונקציית האנרגיה שלנו . אם פונקציית אנרגיה שלנו רציפה, בהנחה על רציפותה נוכל לגזור אותה ולהוציא ממנה את ההסיאן והגרדיאנט.*

*עבור מה ישמש הגרדיאנט?*

***Text

Description automatically generatedדוגמא –*** *אנרגיית ואן-דר-ואלס בין זוגות אטומים עבור אטומים –*

*מימין חישבנו את הנגזרת של פונקציית האנרגיה של הVDW בין שני אטומים i ,j, כאשר המרחק בין שני אטומים, ו A,B קבועים.*

*מדובר על הוצאת גרדיאנט באופן אנליטי – על פי הכללים, אך נוכל להוציא גם מבחינה נומרית את הגרדיאנט שלנו.*

*נוכל להעריך על פי ההגדרה של נגזרת בהכללה גסה. מבחינה אינטואיטיבית –*

*הגרדיאנט*

*שיטת ה״gradient descent״ הינו האלגוריתם הבסיסי ביותר למציאת מינימום אנרגטי תוך שימוש בחישוב גרדיאנטים.*

*הנחות היסוד שלנו הן שיש לנו –*

1. *מרחב קונפיגורציות*
2. *פונקציית אנרגיה דיפרנציאבילית*
3. *מימוש שמאפשר לגזור את פונקציית האנרגיה בשביל הוצאת גרדיאנט*

*Diagram

Description automatically generatedבמקרה שבו יש לנו את הנ״ל האלגוריתם נראה כך –*

* *איך נבחר את הפרמטר שלנו? נפעיל אסטרטגיית line-search כדי למצוא את גודל הצעד הטוב ביותר כדי להקטין את E למינימום.*
* *Chart

  Description automatically generatedלמה אנחנו ממשיכים כל עוד הגרדיאנט גדול מ ? הגרדיאנט בנקודת מינימום/מקסימום הוא אפס. בסיום האלגוריתם מובטח לנו שהגרדיאנט אפסי(לא מובטח שמדובר במינימה, יכולה להיות גם נקודת פיתול).*

*מה הבעיה בעצם בGradient-Descent? לא רק שאנחנו יכולים להיתקע בנקודת מינימום לוקלי ולא גלובלית, האלגוריתם אמור להיות מאוד איטי. לדוגמא – פונקציית רוזנברוק. הפונקציה תעשה ״זיגזג״ מאוד ארוך וייקח זמן רב להגיע למינימום.*

*פתרון לבעיה – "conjugate gradient descent״. אדפטציה לאלגוריתם שתשמור מידע על הצעדים הקודמים. כאשר המשטח ריבועי תהיה התכנסות טובה מאוד של האדפטציה כנ״ל.*

הרצאה 6+7 – **מציאת שורשים ושיטת מונטה קרלו**

*נרצה למצוא מינימום לוקלי ונוכל לעשות זאת בעזרת מציאת שורשים(למען מציאת נקודות בהן הפונקציה מתאפסת).*

*נוכל לגזור את הפונקציה(לקבל את הגרדיאנט), למצוא שורשים ולמצוא את הנקודות המתאפסות.*

*כאמור לא קל למצוא שורשים לפונקציות, על כן נוכל למצוא קירוב בעזרת טור טיילור מסדר שיספק לנו קירוב בסביבה מסוימת של הפונקציה, ובעזרת קירוב טיילור נוכל למצוא את שורשי הפונקציה תחת סביבה מסוימת.*

*נדבר על שיטות למציאת שורשים בקירוב –*

*שיטת ניוטון-רפסון*

*שיטה למציאת שורשים בחד-ממד. מדובר על אלגוריתם איטרטיבי.*

*באותה המידה כדי לעבור למרחב ביותר מממד אחד, נשתמש בגרידאנט בתור הכללה לנגזרת ראשונה והסיאן בתור הכללה לנגזרת שנייה.*

***שיטות מונטה קרלו***

*נתמקד בשיטות בשם ״שרשראות מרקוב״ במשפחת שיטות מונטה קרלו.*

*שרשרת מרקוב – מצבים עתידיים תלויים אך ורק במצב הנוכחי, ״אין זיכרון״.*

*האלגוריתם שלנו –*

*Graphical user interface

Description automatically generated with medium confidence*

*Graphical user interface, text, application, chat or text message

Description automatically generatedהתכונה המרקובית מתבטאת על ידי כך שאנחנו לא תלויים בצעדים שלנו מהעבר.*

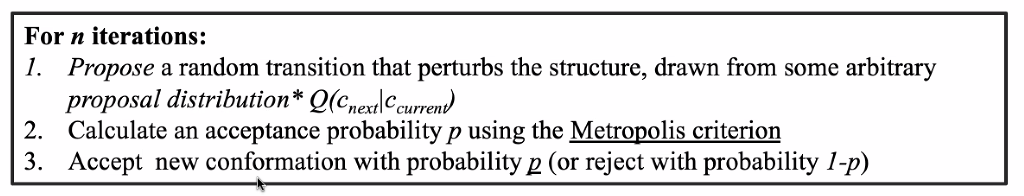
*הצעד הראשון מסתמך על הארגודיות של הגרף.*

*נרצה לדעת איך בוחרים את p כך שp יעזור לנו להתכנס להתפלגות בולצמן, כאשר יש לנו כקלט את פונקציית אנרגיה לכל קונפיגורציה שלנו.*

*נרצה אלגוריתם שיתכנס להתפלגות הזו בלי לחשב את Z שקשה לחשב אותה.*

*קריטריון מטרופוליס-הייסטינגס*

*כעת שינינו את האלגוריתם שלנו להיות –*

**

*שאנו משנים את p על פי קריטריון מטרופוליס שהראנו מלעיל.*

*אבל ההתכנסות תשאף באינסוף להתפלגות בולצמן, אבל אנחנו לא רוצים התכנסות רק באינסוף.  
לא נרצה לסרוק את כל ההתפלגות אלא למצוא דגימות לאנרגיות נמוכות בהסתברות סבירה.*

*אנחנו עובדים בתהליכים מרקובים ולכן אין צורך לזכור את כל הגרף אלא רק את המצב הנוכחי. נצטרך לקבל או לא לקבל לפי קריטריון מטרופוליס.* ***כל השיקולים במונטה קרלו הם שינויים מקומיים****.*

הרצאה 8 – **קיפול חלבוני קומפרטבי**

נזכיר שיש לנו 2 שיטות למציאת קיפולו של חלבון – ab initio, comparative modelling

מידול קומפרטיבי – הסתכלות על מבנים שנפתרו כבר בעלי רצף דומה(טמפלט), הסתמכות על עקרונות פיזיקליים וסטטיסטים.

ab initio – הסתמכות על עקרונות פיזיקלים ועל מבנים קיימים וסטטיסטיקה, כאשר משקל העקרונות מעט אחרת.

* ברוזטה מבצעים clustering למבנים אפשריים שאנחנו מקבלים לתוך ״מחלקות שקילות״.

אז איך מקפלים חלבון? מחפשים אם ישנו מבנה דומה ב**blast**. אם אין נריץ דרך רוזטה.

* בblast מחפשים מבנים דומים לחלבון שאנחנו מספקים.

בדרך כלל דרך blast נקבל תוצאה שאיננה בינארית – נקבל אזורים שכבר ממופים וגם אזורים שאינם ממופים –

*Chart

Description automatically generated with medium confidence*

*הצביעה הינה על פי אופי ההתאמה. נוכל ללחוץ על הצבעים ולקבל מידע על איכות ההתאמה, לדוגמא: באדום חלק שמופפה 100%.*

*חלק שפחות מוגדר יסומן בירוק, ניתן ללחוץ ולראות מה קורה במיפוי.*

*נבחין כי לפני שננסה לקפל את החלבון נצטרך לבדוק אם החלבון בכלל יכול להתקפל, לעיתים יש חלבונים עם מספר צורות ואז הם לא יתקפלו או אין צורך בקיפול לפונקציה שלהם. על חלבון מהצורה כנ״ל נריץ ab-initio על האזור שאיננו מוגדר, והאזורים האחרים (למעט האדום, שממופה כבר כ100%) נריץ קומפרטיב-מודלינג.*

* *דינה חושדת שהאזור באמצע הוא disordered(לא ניתן למיפוי).*

*איך נבצע קומפרטיב-מודלינג?*

*Diagram

Description automatically generated*

*נוציא אינופורמציה ממבנה ידוע ונבנה אתת המבנה לtarget עם פיזיקה וסטטיסטיקה. מה זאת אומרת?*

*נריץ בblast ונמצא alignment. (טמפלט: יש מבנה, טרגט: נרצה למצוא מבנה).*

*נוציא אילוצים מהטמפלט כדי למצוא מבנה לטרגט. נוציא את המרחקים בין חומצות אמינו. לאחר הגדרת אילוצי המרחק נריץ אופטימיזציה למבנה ולאפשרויות עם הgradient decent והשיטות שלמדנו קודם.*

*נריץ סימולציה פיזיקלית ב״מודולר״(נוכל לדמה שם בעזרת אילוצי מרחק דימוי של האטומים תחת כוחות פיזיקלים).*

*Diagram

Description automatically generatedמה זה הוצאת אילוצי המרחק? נדרוש שהמרחק בין שתי חומצות אמינו יהיה ״בערך״. נגדיר סביב המרחק של החומצות בטמפלט פונקציית מרחק סביב המרחק(כמו דלתא כלשהיא שנסבוב סביב נקודה x).*

*המודולר ייקח מרחקי (זוגות סי-אלפא), זוויות דיאדרליות, שרשראות צד ועוד עקרונות פיזיקליים קבועים מראש.*

*נוכל להוציא במודלר אילוצים מכמה טמפלטים (במידה ואחוזי ההתאמה נמוכים) .ניצור ״מולטי טמפלט מודלינג ״. נקבל אילוצים מכמה טמפלטים. נקבל מבנים די דומים אחד לשני, אבל כמובן שיש שגיאות במודלים.*

*נבחין בקשר בין אורך התאמה לאחוזי הדמיון הרצפי:*

*למעשה בחלבונים ארוכים נרשה לעצמנו גם אחוז דמיון רצפי נמוך.*

*שגיאות במבנים*

*עיקר השגיאות בcomparative modeling נובעות מהשלב הראשון – בחירת הטמפלט.*

*עלולה להיות גם שגיאה מהעימוד(alignment). אם תהיה לנו שגיאה בעימוד של ״שיפט אחד ימינה״, לא נוכל לצאת מהבעיה שלנו. בקיצור – garbage in, garbage out.*

*A picture containing tool

Description automatically generatedלפעמים יש לנו גם אזורים ללא טמפלטים – אזורים שלא דומים לשום דבר בטמפלט. על האזור כנ״ל נצטרך לבצע ab initio.*

*בנוסף יכול להיות גם שגיאות בpacking של הside chains – שרשרת צד אחת קופצת להיות שרשרת צד אחרת. במקרה כזה ההיפוך של הside chain יכול להיות מאוד קריטי אם מנסים להבין איך חלבונים מבצעים אינטרקציה בינהם.*

*אז איפה הלולאות הנכונות קריטי לנו? בנוגדנים.*

*נוגדנים הם מה שהגוף מייצר כתגובה לחיסון, נאמר שהם בצורת Y. הם ״החיסוניות האקטיבית שלנו״, והם גם ניתנים בתור טיפול הרבה פעמים, כבר משנות ה80. הנוגדנים נראים אותו הדבר אך מה שמשתנה ביניהם הוא הקצוות ביניהם, והם קובעים אל מה נתקשר ונתחבר.*

*בנוסף יש לנו ננו-נוגדנים שנמצאים בגמלים, אלפקות וכדומה, ואלו קטנים יותר – גם בצורת Y בעלי זרועות קצרות יותר.*

*בנוגדנים יש 6 לולאות צד ובננו-נוגדנים יש 3 לולאות צד. עלות הייצור של ננו-נוגדנים זולה יותר מאשר נוגדנים עצמם.*

*בנוגדנים יש לולאה ארוכה מהיתר(H3) ועל כן יותר דרגות חופש, על כן זה כאב ראש למדל אותה – עלייה נשתמש בab initio, אבל זה כן חשוב שתהיה שיטה חישובית למידול הנוגדנים.*

***תת הבעיה שנתעסק בה כעת הינה מידול הלולאות הללו****. לאחר בנייה של המבנה בcomparative, נרצה לבצע שלב נוסף של אופטימיזציה, ננסה לבנות אותה מחדש בשיטות קרובות לab-initio.*

* *עבור לולאות קצרות נוכל לחפש לולאות דומות בpdb(לולאות עד 4-6 חומצות אמינו).*
* *עבור לולאות ארוכות –   
  למה זה שונה מבעיית קיפול רגילה? יש לנו קצוות קבועים – שזהו אילוץ נוסף. יש פחות דרגות חופש מה שנשמע שעלול להיות קל יותר, אבל מה באמת נעשה?*

*Chart, radar chart

Description automatically generatedנוכל ליצור דאטה-בייס של פרגמנטים ולהריץ מונטה-קרלו. מבנה הלולאה ייפתח מההרצה. למעשה נדאג לפתוח את הלולאה(ולמצוא פתרון) ולאחר מכן לסגור אותה.  
שיטות סגירת הלולאות דווקא קשורות לרובוטיקה (הגעה ליעד ותצורה מסוימת ולחזור בחזרה, בעזרת יד רובוטית לדוגמא).*

*יש לנו 2 שיטות עבור כך –*

* + *Cyclic coordinate descent – הולכים על כל דרגות החופש (כל זרוע) ומחשבים כמה נצטרך לסובב כדי להגיע ליעד. (כחול זה מה שזז, ירוק זו הלולאה והאדום זה לאן נרצה להזיז). נבחין שנרצה לקשר בין איפה שי מעבר בין ירוק לכחול (כדי לסגור את הלולאה).*

*עובד באופן איטרטיבי על ״סיבוב קטן״ על זוויות הלולאה עד שהיא נסגרת. עובד 99% מהזמן.*

*A picture containing calendar

Description automatically generatedהרצאה 10 – קיפול חלבונים בשיטות למידה*

*נזכיר מהו Co-evuluation: קורלציה בהחלפות של חומצות אמינו במהלך האבולוציה.*

*לא תמיד קורלציה בין החלפות בחומצות אמינו גוררות contact בין שתי חומצות אמינו.*

*אנחנו נדון בשימוש בdeep learning בלמידת קיפול חלבונים.*

*במקרה של למידת קיפול החלבונים נקבל הרבה false positive. למה?*

*Diagram

Description automatically generatedנזכיר כי חלבונים יכולים לבצע ״קיפול עצמי״, מה שיכול להטעות את אלגוריתם הקיפול ולהתאים עותק של החלבון אל החלבון עצמו. יש עוד סיבות אבל אנחנו רוצים את סיבת ה״מונומר״, הקיפול העצמי ללא העותק העצמי.*

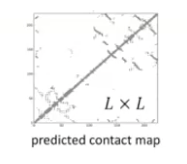
*עבודה עם רשתות נוירונים סיפקו פתרון להיפטרות מFalse Positives.*

*איך נייצג חלבון? בתור וקטור באורך L(אורך החלבון), בכל תא יהיה וקטור באורך 20, כאשר .*

*במקום 1 או 0(בינארי), כי נייצג בכל מקרה ב floating point בגלל רשתות הנוירונים, נחשב את השכיחות של כל חומצת אמינו בכל עמודה.*

* *קלט לרשת –*

1. *שרשור החלבון, ,שכיחות 20 חומצות אמינו(השכיחות של כל עמודה) ועוד 3 פרדיקטורים למבנים שניוניים(הליקסים, בטא סטרנד, אחר) , ועוד 3 קוארדינטות של חשיפה.*
2. *מטריצת קואבולוציה, . מטריצה עם נקודות אדומות כמו שראינו, למה כפול 3? ניקח פשוט כמה שיטות לפרדיקציה של המטריצה.*

* *פלט – contact map. *