

# העתיד כבר כאן: פיתוח חלבונים בעזרת בינה מלאכותית

שירן ברבר-צוקר\*

**זכות המהפכה בחיזוי מבני חלבונים באמצעות בינה מלאכותית ועל ידי שילובה עם שיטות הנדסת חלבונים מתקדמות, נפתח צוהר ללמידה ולשיפור החלבונים המסובכים ביותר בטבע**

חלבונים, מהמולקולות המורכבות ביותר המוכרות למדע, אחראיים על מרבית התהליכים הביולוגים שמאפשרים את החיים. החלבונים מיוצרים כפולימרים ארוכים המורכבים מ-20 סוגים של חומצות אמינו, כאשר רצף חומצות האמינו הוא זה שקובע את המבנה התלת ממדי של החלבון ותפקודו. לחלבונים מגוון רחב של תפקידים. בין היתר, הם אחראיים על בניית שלד התא, על הגנה של התא מפני חומרים זרים, הם מעבירים חומרים נחוצים ממקום למקום ומזרזים תגובות כימיות חיוניות. זירוז תגובות כימיות נעשה על ידי תת קבוצה של חלבונים הנקראים אנזימים, שתפקידם הוא להמיר מולקולה A למולקולה A' או למולקולה B. מאחר שאנזימים טבעיים יכולים לבצע מגוון רחב מאוד של תגובות כימיות בצורה

יפקו בצורה יעילה באורגניזמים פשוטים כאלה. כמו כן, תנאים תעשייתיים בהם האנזימים נדרשים לפעול שונים לחלוטין מהתנאים בתוך התא. למשל, נדרשת טמפרטורה גבוהה או תנאי חומציות קיצוניים - ואלה עלולים לערער את יציבותו ופעילותו של החלבון.

במשך שנים ארוכות מדענים רבים באקדמיה ובתעשייה ניסו לפתח שיטות לשיפור חלבונים - בין אם כדי לשנות או לייעל את פעילותם, ובין אם כדי להגדיל את אפשרויות ההפקה שלהם באורגניזמים פשוטים ואת העמידות שלהם לתנאי סביבה קשים. הדרך הפשוטה ביותר לשינוי פעילות או עמידות של חלבון היא על ידי שינוי הגן המקודד, כלומר הכנסת מוטציות (החלפת חומצה אמינית אחת

באחרת) שישנו את המבנה או התכונות הכימיות של החלבון ובכך יביאו לשינוי הפעילות ו/או העמידות. חלבון ממוצע מורכב ממאות חומצות אמינו, ובתהליך הכנסת המוטציות לרצף החלבון - כל אבן בניין כזו יכולה תיאורטית להיות מוחלפת ב-19 אבני הבניין

האחרות. כך מרחב האפשרויות עבור חלבון שאורכו 300 חומצות אמינו הוא  $20^{300}$  - מספר אסטרונומי שלא ניתן לסרוק אפילו בשיטות חישוביות (לשם השוואה, מספר החלקיקים היסודיים ביקום מוערך בסביבות ה- $10^{80}$ ). השיטה המרכזית והפופולרית ביותר בימינו לשיפור חלבונים נקראת אבולוציה מכוונת (על המצאתה זכתה פרופ' פרנסס ארנולד בפרס נובל לכימיה בשנת 2018) [1]. בשיטה זו נותנים לאורגניזם המבטא את החלבון לערוך מוטציות נקודתיות באופן אקראי ברצף החלבון; לחלבון באורך 300 חומצות אמינו יכולות להיות  $19 \times 300$  מוטציות יחידניות שמשמעותן 5700 חלבונים שונים אשר רק מקצתם באמת יידגמו בשיטה זו. לאחר מכן סורקים את החלבונים שנוצרו ובודקים אילו מהם טובים יותר מהחלבון הטבעי מבחינת התכונה הרצויה (כאשר רובן המוחץ של המוטציות יפגעו בפעילות ובעמידות החלבון). החלבונים

**במשך שנים ארוכות מדענים רבים באקדמיה ובתעשייה ניסו לפתח שיטות לשיפור חלבונים - בין אם כדי לשנות או לייעל את פעילותם, ובין אם כדי להגדיל את אפשרויות ההפקה שלהם באורגניזמים פשוטים ואת העמידות שלהם לתנאי סביבה קשים**

מדויקת וביעילות אנרגטית גבוהה (כלומר בצורה ידידותית לסביבה), השימוש בהם בתעשיות שונות הוא בעל ערך רב. כך למשל, בתעשיות הביוטכנולוגיות הם משמשים ליצירת מולקולות קטנות המשמשות כתרופות או מזרזים תהליכי גיבון של חלב, בתעשיית הכימיה הם מאפשרים יצור מולקולות בסלקטיביות גבוהה שלא ניתן להגיע אליה בשיטות סינתזה כימיות אחרות, ובתעשיית הנייר והחקלאות הם יכולים לפרק פסולת נייר ויבולים כדי לנצל ביעילות או למחזר את החומרים המרכיבים אותם.

אמנם ישנו פוטנציאל גבוה לשימוש באנזימים (וחלבונים בכלל) בשלל תעשיות, אך כל החלבונים נוצרו בתהליך אבולוציוני שבו הם עברו התאמות לסביבת האורגניזם בו הם מצויים. בשביל לעבוד עם חלבון במעבדה או בתעשייה, יש לייצר אותו באורגניזם זול, כחיידק או שמר, אך רוב החלבונים לא

\* ד"ר שירן ברבר-צוקר, ראש מחלקת הנדסת חלבונים בחברת סקאלה ביודיזין

Pos	WT	Full sequence space (20 <sup>358</sup> )	→ Reduced space
1	M	ARNDCQEGHILK <b>F</b> PSTWYV	<b>MIL</b>
2	P	ARNDCQEGHILK <b>M</b> FSTWYV	<b>PA</b>
3	Q	ARNDC <b>E</b> GHILKMF <b>P</b> STWYV	<b>QE</b>
		⋮	
122	S	ARNDCQEGHILK <b>M</b> F <b>P</b> STWYV	<b>ST</b>
123	G	ARNDCQ <b>E</b> HILK <b>M</b> F <b>P</b> STWYV	<b>GANS</b>
124	A	R <b>N</b> DCQEGHILK <b>M</b> F <b>P</b> STWYV	<b>ANST</b>
125	D	ARN <b>C</b> Q <b>E</b> GHILK <b>M</b> F <b>P</b> STWYV	<b>DE</b>
		⋮	
356	L	ARNDCQEGHIL <b>K</b> M <b>F</b> PS <b>T</b> WYV	<b>LIMV</b>
357	V	ARNDCQEGHILK <b>M</b> F <b>P</b> STWY	<b>V</b>
358	A	R <b>N</b> DCQ <b>E</b> GHIL <b>K</b> M <b>F</b> PS <b>T</b> WYV	<b>ANDQEGPST</b>

איור 1. עקרון שיטת הנדסת החלבונים במעבדת פליישמן. במרחב הרצף המלא, עבור כל עמדה בחלבון המורשתית לעבור עיצוב (ממוספר משמאל), ישנן 19 זהויות היכולות להידגם פרט לזהות הטבעית (WT) - זהו מרחב רצף ענק שלא ניתן לדגום באופן יעיל אפילו בצורה חישובית. במעבדת פליישמן בודקים עבור כל עמדה המורשתית לעבור עיצוב, אם כל אחת מ-19 אבני הבניין שאינן זהות הטבעית, נדגמת באותו המקום ברצפים ברצפים דומים (זהויות בעלות רקע כחול). אם הזהות מופיעה ברצפים דומים, בודקים באמצעות חישובים מבניים אם אותה הזהות כמוטציה בודדת מייצבת את מבנה החלבון (זהויות סגולות). אם שני התנאים מתקיימים, תהיה הזהות מותרת להידגם בשיטות חישוביות אשר מאפשרות הכנסת מוטציות רבות בו-זמנית (צד ימין בתמונה). החלת שני התנאים מצמצמת משמעותית את מרחב הרצף המותר, לגודל הניתן לדגום בצורה אפקטיבית בחישובים אטומיים-מבניים.

מתגבשים בקלות, או המבנה שלהם אינו מספיק סדור על מנת ליצור גביש אחיד הדרוש לניתוח מבני. בנוסף גיבוש ופתירת מבני חלבונים הוא תהליך ארוך ויקר שדורש ידע רב והכשרה ארוכה. כך, בעוד יש מאות מיליוני רצפים של חלבונים בבנק הגנים, עד היום נפתר המבנה של פחות מ-200,000 חלבונים. המחסור במבני החלבונים הוא לא רק בעיה לעולם הנדסת החלבונים, אלא גם לכלל עולם הביוכימיה; לרובם המוחץ של החלבונים אין מבנה, ולכן היכולת שלנו להבין את תפקודם לוקה בחסר. וכך במשך עשרות שנים ניסו חוקרים רבים בעולם לפתח שיטות חישוביות לקיפול רצף חומצות אמינו למבנה התלת ממדי כדי לעקוף את הבעיה. בסוף שנת 2018 החברה DeepMind (חברת בת של Alphabet (גוגל)) הציגה לעולם את AlphaFold (AF), אלגוריתם מבוסס בינה מלאכותית לקיפול חלבונים [5]. באמצעות למידת כלל המבנים הקיימים בבנק מבני החלבונים, הראה האלגוריתם שלהם תוצאות טובות משמעותית מכל אלגוריתם קודם בתחום והביא בעצם לפריצת דרך בתחום. שנתיים מאוחר יותר פרסמה החברה את AlphaFold2 (AF2), אלגוריתם משופר משמעותית שמגיע לרמת דיוק דומה לרמת הדיוק של מבנים שנפתרו בצורה ניסויית (איור 2); פרסום זה הביא לפתרון של בעיה מדעית בת עשרות שנים [6]. AF2 פרץ את הסכר בתחום, ומאז קבוצות רבות - החל מקבוצות אקדמיות המתעסקות בנושא זה כבר שנים, ועד קבוצות בחברות ענק שאינן מתמחות בביוכימיה, כגון Meta (פייסבוק) ו-Salesforce - משתמשות באלגוריתמים של בינה מלאכותית בצורות שונות (למשל, על ידי שימוש במודלי שפה

המשופרים יהיו נקודת התחלה לסבב נוסף של מוטציות נקודתיות, וכן הלאה [2]. ברוב המקרים דרושים מספר סיבובים כאלו על מנת להגיע לשיפור מספיק של התכונה הרצויה. על כן זהו תהליך ארוך ויקר שבו נדגמות אלפי וריאציות של החלבון במעבדה. בנוסף מאחר שבתהליך זה המוטציות מצטברות זו אחר זו, התהליך אינו מאפשר שילוב של מוטציות - שכל אחת מהן אינה מועילה או אפילו פוגעת ביעילות החלבון - אך ביחד הן משפרות אותו בצורה משמעותית מאוד.

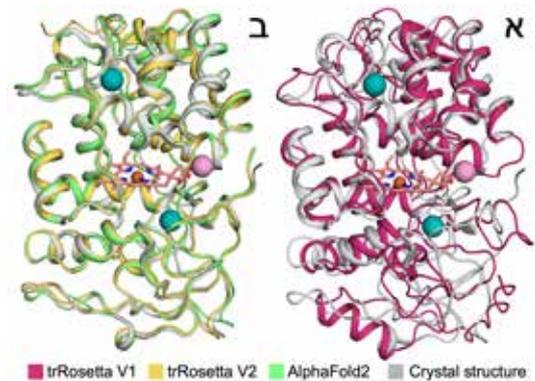
שיטות חישוביות להנדסת חלבונים משקללות את המבנה המרחבי של החלבון ומוצאות דרכים לייצב אותו - זאת על ידי מציאת קומבינציית המוטציות שתביא את החלבון למצב האפשרי הנמוך ביותר באנרגיה. עם זאת, חישובים אלו מביאים לתוצאות מוצלחות בעיקר על חלבונים קטנים ואינם פרקטיים על חלבונים טבעיים, בין היתר משום שמרחב הרצף המלא שלהם אינו ניתן לדגימה.

מעבדתו של פרופ' שראל פליישמן ממכון ויצמן פיתחה בעשור האחרון מספר שיטות חישוביות המשתמשות בשילוב של מידע אבולוציוני וחישובים מבניים (אטומיים) על מנת ליצור חלבונים משופרים הן מבחינת עמידות והן מבחינת פעילות - ללא עבודת מעבדה חזרתית ומאומצת (איור 1) [3,4]. שיטות אלו הן אוטומטיות לחלוטין, מונגשות לכלל האקדמיה כסרברים אינטרנטיים ואינן דורשות מהמשתמש רקע בעיצוב חלבונים. כך, אלפי משתמשים ברחבי העולם כבר השתמשו בשיטות אלו על מנת להנדס חלבונים ממגוון רחב של סוגים ועבור אפליקציות שונות - ועשרות רבות מהם כבר פרסמו את תוצאותיהם בעיתונים מובילים בעולם המדע. השיטה הכללית ביותר ואולי השימושית ביותר שפותחה במעבדה - שנקראת פרוס - מטרתה ייצוב חלבונים, הגברת יכולת הביטוי שלהם באורגניזמים פשוטים והגדלת עמידותם לתנאים קשים, זאת תוך כדי שמירה על פעילותם [3]. פרוס מציעה למשתמשים לבדוק ניסויית מספר קטן של גרסאות עם עשרות מוטציות מהחלבון המקורי, וזאת לעומת עשרות האלפים שנסרקים בפרויקט בודד של אבולוציה מכוונת, למשל.

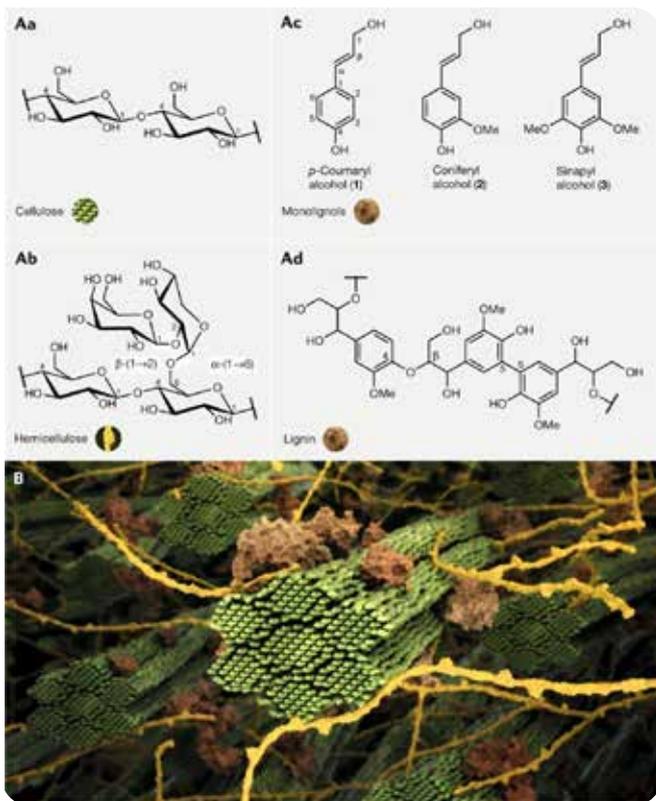
הבעיה המהותית בשימוש בשיטות חישוביות היא שהן דורשות מבנה בעל דיוק אטומי של החלבון שעליו עובדים. השיטה הנפוצה ביותר לפתירת מבני חלבונים היא על ידי גיבושם והקרנתם בגלי X רבי עוצמה. עם זאת, חלבונים רבים אינם

**שיטות חישוביות להנדסת חלבונים משקללות את המבנה המרחבי של החלבון ומוצאות דרכים לייצב אותו - זאת על ידי מציאת קומבינציית המוטציות שתביא את החלבון למצב האפשרי הנמוך ביותר באנרגיה**

צלולוז והמי-צלולוז מרכיב את דופן התא בצמחים, לרבות בקליפת עץ [10] (איור 3). בעוד שכיום ישנן דרכים ידידותיות לפירוק הפולימרים הסוכריים צלולוז והמי-צלולוז לחדי-הסוכרים המרכיבים אותם, סוג הקשרים בין תתי היחידות הארומטיות של ליגנין הם קשים מאוד לפירוק ודורשים תנאים כימיים קשוחים (כמו חום גבוה מאוד או רמת חומציות קיצונית). משום כך ומשום שהליגנין כורך את הצלולוז והמי-צלולוז (איור 3) [11], אי היכולת לפרק אותו מגבילה מאוד את פירוקם של הפולימרים האחרים ואת הפקת האנרגיה מתתי הסוכרים המרכיבים אותם (אופטימלית, אלו היו משמשים לדלקים



איור 2. השוואת מבנים שנפתרו בצורה ניסויית למבנים שחושבו ע"י אלגוריתם בינה מלאכותית בשנים 2020-2021. (א) מודל שחושב על ידי אלגוריתם המבוסס על הפיתוח הראשוני של AF שפורסם בינואר 2020 (ורוד) חפוף על המבנה הניסויי של אותו החלבון (אפור). הדמיון בין המודל החישובי לבין המבנה הניסויי אינו מושלם, אך נחשב באותה העת ליוצא דופן ברמת הדיוק לעומת שיטות חישוביות קודמות. (ב) מודלים של אותו החלבון שחושבו על ידי אלגוריתמים מתקדמים יותר שפורסמו בסוף אותה השנה (צהוב וירוק, האחרון הוא AF2) חפופים על המבנה הניסויי של אותו החלבון (אפור). ניתן לראות כי ההתאמה המבנית בין המודלים לבין עצמם ובין המבנה הניסויי היא גבוהה מאוד. חשוב לציין כי הקלט היחיד של כל האלגוריתמים הללו הוא הרצף הראשוני של החלבון; למרות שהחלבונים קושרים בתוכם מולקולות קטנות החשובות לקיפולו המדויק של החלבון (כפי שניתן לראות במבנה הניסויי בוורוד וטורקיז), האלגוריתמים מחשבים את המבנה בצורה מדויקת כך שנוצרים במודלים אותם אתרי הקישור ריקים. תמונה זו מבוססת על תמונה ממקור (14).

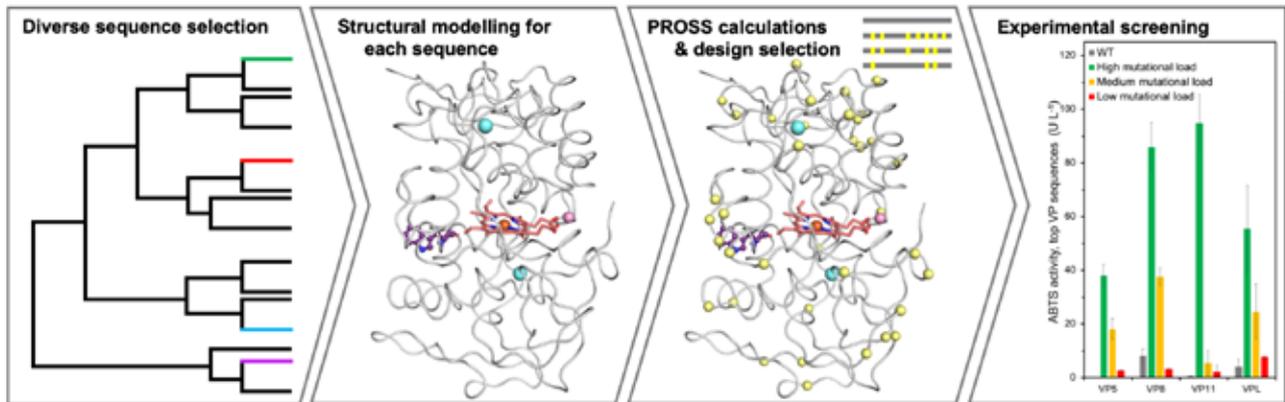


איור 3. הרכב הליגנוצלולוז בדופן התא הצמחי (מתוך מקור 11). (Aa) צלולוז הוא רב-סוכר המורכב מיחידות חוזרות של גלוקוז. יחידות הגלוקוז קשורות זו לזו בקשר גליקוזידי ויוצרות פולימר לינארי ארוך. (Ab) המיצלולוז גם הוא רב-סוכר, אך הוא מורכב מיחידות חוזרות של סוכרים שונים הקשורים ביניהם במספר סוגי קשרים היוצרים גם הסתעפויות. (Ac) קומריל, קוניפריל וסינאפיל אלכוהול הן תתי היחידות העיקריות המרכיבות את הליגנין. (Ad) ליגנין הוא פולימר ארומטי, אמורפי ועמיד מאוד המורכב מתתי היחידות השונות הקשורות ברמות סיעוף שונות. (B) ליגנוצלולוז הוא החומר העיקרי המרכיב את דופן תא הצמח, והוא מורכב מהפולימרים המתוארים ב-(A). באיור זה ניתן לראות כי פולימרי הצלולוז (בירוק) יוצרים מיקרו-סיבים (כאשר שבעה מיקרו-סיבים כאלו יוצרים סיב אחד), וסיבים כרוכים המיצלולוז (צהוב) וליגנין (חום).

ביולוגיים). כמו כן הליגנין עצמו מכיל חומרים ארומטיים רבים בעלי פוטנציאל שימושי בתעשיות כימיות שונות שכיום לא ניתן לנצל כלל [12]. ה-VPs מקורם בפטריות רקב לבנות

מהסוג שבו משתמשים ב-ChatGPT) על מנת לחזות מבנים של חלבונים, לייצר חלבונים חדשים ולשפר חלבונים קיימים [7-9]. למרות ששיטות הבינה המלאכותית עדיין לא הצליחו ליצור חלבונים משופרים משמעותית מחלבונים טבעיים (המירוץ לכך מתקיים בימים אלו, אך כרגע השיטות אינן בשלות לכך) - הן פתחו פתח להרחבת שיטות חישוביות קלאסיות. לדוגמה, בעוד שבעבר דרש האלגוריתם פרוס של קבוצת פליישמן מבנה אטומי ניסויי של חלבון, כיום ניתן לחשב את מבניהם התלת ממדיים של כל החלבונים ולהפעיל את האלגוריתם על המבנים המחושבים. שילוב כזה בין השיטות הוא בעל משמעות אדירה, שכן עד כה פרוס יכול היה לעבוד רק על חלבונים בעלי מבנה ידוע, אך דווקא חלבונים שאין להם מבנה (כלומר, שאינם יציבים על מנת ליצור גביש סדור) - עבורם האלגוריתם יכול להועיל במיוחד (בעיית הביצה והתרנגולת).

רבים מהאנזימים שיכולים להיות מועילים עבור אפליקציות סביבתיות הם בדיוק מהסוג הנידון למעלה - חלבונים אשר מגיעים מאורגניזמים מורכבים ושקשה מאוד לעבוד איתם במעבדה, ועוד יותר בתנאים תעשייתיים. לכן על אנזימים אלו אין הרבה מידע, ובדרך כלל רק נציגים בודדים מכל משפחת אנזימים אופיינו במעבדה, אם בכלל. כך למשל, אנזימים מסוג המסוגלים לחמצן ליגנין, פולימר ארומטי מורכב, שיחד עם



איור 4. עיצוב אנזימי VP. משמאל לימין: בשלב הראשון, נאספו כלל רצפי ה-VPs מבנק הגנים, סווגו לפי רמת הדומות ביניהם ומתוכם נבחרו 12 רצפים מייצגים. בשלב השני חושב המבנה של כל אחד מ-12 הרצפים על ידי אלגוריתם בינה מלאכותית. בשלב השלישי נעשה שימוש במודלים אלו כבסיס לחישובים על ידי האלגוריתם פרוס, אשר הכניס בכל אחד מהמודלים עשרות מוטציות (מיוצגות ככדורים צהובים). שלושה עיצובים של פרוס נבחרו לבחינה ניסויית עבור כל רצף התחלתי. לבסוף כל 48 החלבונים (12 חלבונים טבעיים ו-3 עיצובים לכל חלבון טבעי) נבחנו בצורה ניסויית. בעוד שאף אחד מהחלבונים הטבעיים לא נוצר במעבדה בצורה משמעותית, הרי שהעיצובים שחושבו על ידי פרוס נוצרו בכמות גדולה והראו רמות פעילות גבוהות עבור ארבעה חלבונים. בגרף מוצגת רמת הייצור של חלבונים פעילים עבור ארבעת החלבונים הללו: באפור ניתן לראות את רמת הפעילות של החלבון הטבעי; ובאדום, צהוב וירוק - את רמת הפעילות של אותו חלבון אך עם מוטציות שהוכנסו על ידי פרוס. חשוב לציין כי בכל ארבעת החלבונים, ככל שהוכנסו יותר מוטציות מייצבות, כך עלתה רמת הפעילות של החלבונים (עיצובים עם הכי מעט מוטציות סומנו באדום, עיצובים עם הכי הרבה מוטציות סומנו בירוק). תמונה זו מבוססת על תמונה ממקור (14).

מעטים הופקו בשמר וגם אז רק באופן מינימלי ולא מספק אפילו למחקר בסיסי. לעומת זאת, עבור ארבעה אנזימים - הווריאציות שעוצבו על ידי פרוס הופקו היטב והראו פעילות גבוהה (איור 4). כמו כן הם הראו עמידות גבוהה לחום ולתנאי חומציות קיצוניים ופרופיל פעילות שונה מאוד זה מזה. הצעד הבא יהיה לקחת את האנזימים החדשים הללו ויחד עם אנזימים המפרקים צלולוז והמי-צלולוז להשתמש בהם לפירוק פסולת חקלאית לשם הפקת אנרגיה בצורה ידידותית.

עצם ההצלחה של פרוס מדגימה את דיוקם האטומי של מודלי החלבונים, שכן השיטות החישוביות שעליו הוא מתבסס נכשלות כאשר המבנה האטומי אינו מדויק. מאז העבודה הנ"ל על ה-VPs, פותחו אלגוריתמי חיזוי טובים משמעותית ואף מהירים הרבה יותר (למשל AF2, ראו איור 2; שילובם עם פרוס אף הראה אחוז הצלחה גבוה יותר מאשר במקרה הבוחן של ה-VPs). היכולת להשתמש ברצפים כנקודות פתיחה להנדסה באמצעים חישוביים מגדילה משמעותית את סקאלת העבודה - ולמעשה מאפשרת לסרוק בצורה פשוטה ומיטבית את היכולות של אנזימים טבעיים וכך לגלות פונקציות חדשות או יעילות אנזימטית כאלה שלא ידענו על קיומן לפני כן, ובעצם לנצל את מה שהטבע נותן לנו "בחינם". פרט לכך הרחבת היכולת להנדס אנזימים "קשים" ולאפיין אותם במעבדה מעמיקה את ההבנה של אותם האנזימים ושל עולם החלבונים בכלל. צבירת הידע הזו תאפשר בתורה לשפר שיטות בינה מלאכותית חדשות לאפיין חלבונים, לפתרון מבניהם ולהנדסתם.

וקשה מאוד לבטא אותן במערכות שמריות או חיידקיות, כך שעד היום קיים מבנה ניסויי של VP אחד, ורק אנזים זה אופיין מבנית וביוכימית. יש לציין שאפיונו המקיף התאפשר משום שהונדס באמצעות אבולוציה מכוונת - עבודה שארכה כעשור וכללה סריקה של אלפי וריאציות של אותו אנזים בודד [13]. זאת למרות שישנם מאות רצפים ידועים של VPs בבנק הגנים שעליהם איננו יודעים דבר, והם יכולים תיאורטית להיות יעילים או עמידים הרבה יותר מאותו אנזים שאופיין.

כוחו של השילוב בין חיזוי מבנים מבוסס בינה מלאכותית ואלגוריתמים לשיפור חלבונים הודגם במעבדתו של פליישמן על קבוצת האנזימים הזאת [14]. בחרנו 12 אנזימי VP השונים זה מזה כמה שניתן ברצף ושלא אופיינו ניסויית לפני כן, וחזינו את מבניהם באמצעות אלגוריתם מבוסס בינה-מלאכותית. לאחר מכן עיצבנו את האנזימים בהתבסס על המודל המבני באמצעות אלגוריתם פרוס, ועבור כל אנזים מדדנו במעבדה את יכולת הביטוי של האנזימים המהונדסים בשמרים (3 וריאציות לכל רצף התחלתי). מתוך 12 החלבונים הטבעיים,

**למרות ששיטות הבינה המלאכותית עדיין לא הצליחו ליצור חלבונים משופרים משמעותית מחלבונים טבעיים הן פתחו פתח להרחבת שיטות חישוביות קלאסיות**

1. Gibney E, Van Noorden R, Ledford H, Castelvocchi D & Warren M. 'Test-tube' evolution wins Chemistry Nobel Prize. *Nature* **562**, 176 (2018).
2. Packer MS & Liu DR. Methods for the directed evolution of proteins. *Nature Reviews Genetics* **16**, 379-394 (2015).
3. Goldenzweig A et al. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **63**, 337-346 (2016).
4. Khersonsky O et al. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Molecular Cell* **72**, 178-186 (2018).
5. Senior AW et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706-710 (2020).
6. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
7. Baek M et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
8. Lin Z et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023).
9. Madani A et al. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* (ahead of print) (2023).
10. Ruiz-Dueñas FJ et al. Substrate oxidation sites in versatile peroxidase and other basidiomycete peroxidases. *Journal of Experimental Botany* **60**, 441-452 (2009).
11. Petridis L & Smith JC. Molecular-level driving forces in lignocellulosic biomass deconstruction for bioenergy. *Nature Review Chemistry* **2**, 382-389 (2018).
12. Ragauskas AJ et al. Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843 (2014).
13. Garcia-Ruiz E, Gonzalez-Perez D, Ruiz-Dueñas FJ, Martínez AT & Alcalde M. Directed evolution of a temperature-, peroxide- and alkaline pH-tolerant versatile peroxidase. *Biochemical Journal* **441**, 487-498 (2012).
14. Barber-Zucker et al. Stable and Functionally Diverse Versatile Peroxidases Designed Directly from Sequences. *Journal of the American Chemical Society* **144**, 3564-3571 (2022).