



מיחשוב ביולוגי . מתוך COM.PIXABAY

חוקרים בטכניון ובמרכז הבינתחומי הרצליה הדגימו שיפור משמעותי ביעילות התהליך הנדרש לאחסונו של מידע דיגיטלי בדי-אן-איי. במאמר שפורסם בכתב העת *Biotechnology Nature* הדגימה הקבוצה אחסון מידע בצפיפות השקולה לאחסון של יותר מ-10 פטה-בייט (מיליון גיגה-בייט) בגרם בודד של די-אן-איי תוך יעול משמעותי של תהליך הכתיבה. לשם המחשה, צפיפות זו מאפשרת, באופן תאורטי, לאחסן בנפח של כפית את כל המידע השמור ב-Youtube.

את המחקר הוביל תלמיד המחקר ליאון ענבי מהפקולטה למדעי המחשב בטכניון בהנחייתו של פרופ' זהר יחיני מהפקולטה למדעי המחשב בטכניון ומבית ספר אפי ארזי למדעי המחשב במרכז הבינתחומי הרצליה. המחקר נערך בשיתוף עם מעבדתו של פרופ' רועי עמית מהפקולטה להנדסת ביוטכנולוגיה ומזון בטכניון.

כמות המידע הדיגיטלי גדלה במהירות עצומה מאז המצאת ההארד-דיסק על ידי IBM בשנות ה-50. אחסונו של מידע זה הפך לאתגר גדול לא רק בהקשר הטכנולוגי אלא גם בהיבט הכלכלי והסביבתי, שכן כיום אחראיות חוות השרתים – מחסני המידע המשרתים את כולנו – לכ-2% מפליטת הפחמן העולמית (שיעור דומה לפליטה המצטברת של כל המטוסים בעולם) ולכ-3% מצריכת החשמל העולמית (יותר מצריכת החשמל של בריטניה כולה). על רקע כל אלה מתפתחת בעשור האחרון גישה טכנולוגית חדשה ומהפכנית: אחסון מידע בדי-אן-איי. טכנולוגיה זו מאפשרת מזעור משמעותי, שמירת המידע לטווח ארוך הרבה יותר (פי אלף) ועלות אנרגטית וכלכלית אפסית.

הרעיון הבסיסי בקידוד מידע על די-אן-איי הוא זה: מולקולת הדי-אן-איי היא שרשרת המורכבת מחוליות הנקראות נוקלאוטידים. הנוקלאוטידים מתחלקים לארבעה סוגים המסומנים באותיות A, C, G ו-T. כדי לאחסן מידע בדי-אן-איי יש לתרגם כל רצף בינארי (המורכב מהסימנים 0 ו-1) לרצף המורכב מאותיות אלה. בשלב הבא מיוצרות, בתהליך הקרוי סינתזה, מולקולות די-אן-איי ממשיות המייצגות את אותם הרצפים. כדי לקרוא את המידע נדרש ריצוף של מולקולות הדי-אן-איי. ריצוף זה מייצר פלט המייצג את רצף הנוקלאוטידים המרכיב כל מולקולה בקלט, ואת הפלט האמור מתרגמים לרצף בינארי המייצג את ההודעה המקורית שקודדנו. הטכנולוגיות המודרניות מאפשרות סינתזה של אלפי סדרות נוקלאוטידים שונות במקביל.

אחסון על די-אן-איי הוא אתגר טכנולוגי מורכב מאוד. בתחום קריאת המידע (ריצוף) התרחשה התקדמות עצומה בעקבות מהפכת הגנום, אולם בכתיבת המידע ישנם עדיין קשיים טכנולוגיים משמעותיים. מכאן חשיבותה של פריצת הדרך שהושגה על ידי חוקרי הטכניון והמרכז הבינתחומי הרצליה ומאפשרת: (1) הגדלה של מספר האותיות המשמשות לקידוד המידע (מעבר ל-4 האותיות המקוריות); (2) הפחתה משמעותית בסבבי הסינתזה הנדרשים לאגירת המידע בדי-אן-איי; (3) שיפור מנגנון תיקון השגיאות בקוד.

הדי-אן-איי הטבעי מורכב כאמור מארבע אבני בניין, הן ארבע האותיות T, G, C, A. צוות החוקרים הגדיל את מספר האותיות לשימוש בפועל, כשכל אות חדשה מהווה צירוף ייחודי של האותיות המקוריות. הרעיון דומה לייצור של צבעים חדשים על ידי ערבוב ייחודי של צבעי בסיס. הגדלת מספר האותיות מאפשר לקודד יותר מידע בכל עמדה ברצף של מולקולות הדי-אן-איי. לדברי פרופ' יכני, "בתהליכי הסינתזה והריצוף הנהוגים כיום מתקיימת יתירות מידע מובנית (redundancy), כיוון שכל מולקולה מיוצרת במספר גדול של עותקים ונקראת במספר גדול של עותקים במהלך הריצוף. הטכנולוגיה שפיתחנו מנצלת את היתירות הזאת להגדלת מספר האותיות האפקטיבי הרבה מעל ל-4 האותיות המקוריות, וכך מאפשרת לנו לקודד כל יחידת מידע בפחות מחזורי סינתזה."

□

החוקרים הצליחו להפחית ב-20% את מספר סבבי הסינתזה הנדרשים ליחידת מידע. יתר על כן, החוקרים הראו כי אפשר יהיה להפחית בעתיד את מספר סבבי הסינתזה ב-75% ללא מאמצי פיתוח משמעותיים. פירוש הדבר הוא שתהליך האחסון יהיה מהיר יותר ויקר פחות. "בעבודה הזאת ישמנו בצורה מעשית קידוד מידע ביעילות סינתזה הגדולה בעשרות אחוזים בהשוואה לקידוד המקובל", מסביר פרופ' עמית. "המחקר כלל יישום בפועל של שיטת הקידוד החדשה לשם אחסון מידע בנפח גדול על מולקולות די-אן-איי ושחזרו לשם בדיקת התהליך." ואכן, על אחד המדפים במעבדתו של פרופ' עמית בטכניון מונחת מבחנה קטנה המכילה בתוכה כ-10 ננוגרם (מיליארדית הגרם) של די-אן-איי, המקודדים אלפי עותקים של התנ"ך בגירסה דו-לשונית.

קבוצת המחקר פיתחה מנגנון מתקדם המאפשר להתגבר על שגיאות שהן חלק בלתי נפרד מתהליך ביולוגי-פיזיקלי כמו זה המתרחש כאן. חלק מרצף הדי-אן-איי של המולקולות המאחסנות את המידע, שתוכנו על ידי ליאון ענבי ופרופ' יכני, משמש לצורך מנגנון תיקון השגיאות האמור. לדברי ליאון ענבי, "בזכות שימוש בקודים לתיקון שגיאות, המותאמים לקידוד הייחודי שיצרנו, יכולנו לבצע קידוד יעיל במיוחד ולשחזר את המידע בהצלחה. כאשר עובדים במערכת המורכבת ממיליוני חלקים (מולקולות), מתרחשים גם אירועים נדירים ביותר (אירועים של אחד למיליון), העלולים לשבש את הקריאה. הקידוד המוקפד איפשר לנו להתגבר על בעיות אלה."

החוקרים מציינים כי "לטכנולוגיה שהוצגה במאמר יש פוטנציאל ליעיל תהליכים נוספים בביולוגיה סינתטית ובביוטכנולוגיה. אנו מאמינים שבשנים הקרובות נראה עלייה משמעותית בשימוש בדי-אן-איי סינתטי במחקר ובתעשייה."

הדי-אן-איי המלאכותי ששימש את החוקרים ותוכנן על ידי הקבוצה יוצר על ידי חברת Bioscience Twist האמריקאית, המעסיקה גם קבוצת פיתוח בתל אביב, ורוצף במרכז הגנומי של הטכניון. המחקר נתמך

חלקית על ידי תוכנית המסגרת 2020 Horizon של האיחוד האירופי. ליאון ענבי נתמך על ידי מלגת אדמס של האקדמיה הישראלית למדעים. במחקר השתתפו גם ד"ר אורנה עטאר ותלמידת המחקר ענבל וקנין.

למאמר המלא ב- Biotechnology Nature

{loadposition content-related}