**חלק ב – אנליזת חלבונים בעזרת אתר ה UniProt**

**סעיף א**

הסבר: בסעיף זה נדרשנו להצליב בין הנתונים שיש לנו בקובץ ה GeneBank לבין הקובץ שהורדנו מ UniProt. את ההצלבה ביצעתי ע"י המזהה 'GOA'. מזהה זה הינו משותף ל2 המאגרים ומזהה חלבונים באופן ייחודי.  
דרך: על מנת לבצע את ההצלבה באופן האידיאלי מבחינת זמני ריצה, הכנתי 2 מילונים. מיליון עבור הערכים מ UniProt, ועבור הערכים מ GeneBank. בחרתי במילון על מנת לבצע שליפה ב O(1) על מנת לבצע את פעולת ההצלבה המלאה ב O(n). המפתח במילונים אלו היה המזהה 'GOA' ו Name והערכים היו רצף הגנים. כלומר סהכ 4 מילונים.  
תוצאות:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| מקור | ערכים קיימים | ערכים עם מזהה GOA | ערכים עם מזהה Name | הצלבה ע"פ GOA | הצלבה ע"פ Name |
| UniProt | 8541 | 8541 | 6938 | 3381 | 3715 |
| GeneBank | 4237 | 3381 | 4171 | 3381 | 3715 |

הערות: נראה כי הרישומים הקיימים ב GeneBank עבור החלבונים - חסרים ולא מושלמים. כאשר קיימים חלבונים שלא כוללים את המזהה 'GOA'. להערכתי המזהים החסרים נובעים מכך שמאגר ה GeneBank עיקרו הינו גנים ולא חלבונים, לעומת UniProt שזה עיקר ייעודו. בנוסף אף על פי שיש מיפוי רחב יותר של GOA ב UniProt ההצלבה המוצלחת יותר בוצעה ע"י שם החלבון (כנדרש בתרגיל).

צילומי מסך: ראה נספח **חלק ב**

**סעיף ב**

**נספח חלק ב**

להלן פלט של התכנית, הערה חשובה את gn len הינו מספר החלבונים המכילים את המזהה 'GOA'

Text

Description automatically generated