אוניברסיטת תל-אביב

הפקולטה לרפואה ע״ש סאקלר

המדרשה לתארים מתקדמים

התחום: ביואינפורמטיקה

נושא העבודה: **פיתוח כלי ביואינפורמטי למחקר אינטראקטיבי של אינטראקציות בין חלבונים**

העבודה מוגשת על-ידי ניר אורלב, ת.ז. 025729328.

עבודה זו בוצעה כמילוי חלקי של הדרישות לקבלת תואר מוסמך בפקולטה לרפואה ע״ש סאקלר, אוניברסיטת תל-אביב.

בהנחייתם של: פרופ׳ יוסף שילה

פרופ׳ רון שמיר

תאריך: ספטמבר 2002

<u>תודה מקרב לב...</u>

לכל אותם האנשים אשר היו מעורבים בעבודתי, סייעו לי במהלכה, והפכו אותה למהנה יותר.

לחברים, אשר שאלו, התעניינו, הקשיבו לצרותי, וסייעו בכל שיכלו.

לאנשי המעבדה, אשר חיוו את דעתם על התוכנה כביולוגים, פרגנו, ועזרו באפיון התכונות, אשר חשוב היה להוסיף לתוכנה זו.

ובפרט:

לפרופ' יוסי שילה ולפרופ' רון שמיר, אשר הנחו אותי במהלך פיתוח התוכנה ובכתיבת עבודה זו, וסייעו ביצירת הקשרים הדרושים עם חברת Proteome.

להורי, יורם ויונה אורלב, אשר חיזקו אותי בהחלטה ללמוד לתואר שני, ושעודדו אותי בקטעים הקשים, בעיקר במהלך כתיבת עבודה זו.

וליעל הרמן, חברתי הטובה ואשתי לעתיד, אשר הקשיבה, יעצה, דחפה, סייעה, ובעיקר לא התייאשה, לאורך כל השנתיים האחרונות.

תודה רבה לכולכם.

תוכן העניינים

1 תקציר בעברית

עד לפני שנים מעטות, הבעיה המרכזית אשר העסיקה את חוקרי הגנום הייתה זיהוי ופענוח הרצף של גנים הגורמים לפנוטיפ מוגדר. המחקר עסק בעיקר בתכונות מונוגניות, והחיפוש אחר הגן הרלוונטי בוצע בעיקר בשיטת השיבוט האיתורי (positional cloning), או על סמך היכרות עם תפקידו הביוכימי של תוצר הגן. לאחר שנמצא הגן המתאים, נותר לגלות את תפקודו המדויק.

עם התקדמות פענוח הרצף של גנומים מלאים והכניסה לעידן הפוסט-גנומי, הבעיה העיקרית הינה הפוכה – גילוי תפקידו של גן, כשנקודת המוצא היא הרצף שלו. ענף מחקרי זה, המכונה functional genomics, מנסה לקשר בין רצף ופונקציה, במטרה לספק "ניחוש מלומד", אשר יהווה נקודת התחלה למחקר המעבדתי של החלבון. מאמצים רבים נעשים כיום ללמוד על גנים אשר תפקידם אינו ידוע, ועיקרם מתרכז בחיפוש אחר קשר פונקציונאלי בין גנים אלו לגנים מוכרים.

מטרתה העיקרית של הביואינפורמטיקה היא להפוך את הכמויות ההולכות וגדלות של המידע הגולמי הקיים לידע בעל משמעות ביולוגית. אחד האתגרים הכרוכים בכך הוא בניית כלים חישוביים נוחים ופשוטים, אשר יאפשרו לחוקרים גישה מהירה, נוחה ובהירה למידע, באופן שניתן יהיה להפיק ממנו את התועלת המרבית.

בעבודה זו מתואר תהליך הפיתוח של תוכנה ביואינפורמטית בשם PIVOT, אשר תוכל לשמש כלי עזר לחוקר לצורך הצגה נוחה וגמישה של אינטראקציות רבות בין חלבונים שונים. המידע מעובד באופן אוטומטי לגרף, והחוקר יכול באופן אינטראקטיבי לשנות את צורת הגרף, להרחיב או לצמצם את היקפו בהתאם לצרכיו, ולבקש מידע נוסף לגבי החלבונים או האינטראקציות המוצגים. ע"י הצגת המידע בצורה ברורה וקריאה, יכול החוקר להתייחס לכמות גדולה של חלבונים ושל אינטראקציות, לחזות אינטראקציות חדשות באורגניזם מסוים על סמך אלו שזוהו באורגניזם אחר, ולנסות להגיע למסקנות ראשוניות לגבי תפקודם של גנים לא ידועים במערך הגנים עליו הוא מתבונן.

מילות מפתח: ביואינפורמטיקה, ביולוגיה חישובית, אינטראקציות בין חלבונים, עימוד גרפים, גנומיקה פונקציונאלית, חיזוי תפקוד, ניווט במאגרי מידע, שימור אבולוציוני.

2 מבוא וסקירת ספרות

2.1 חשיבותן המחקרית של אינטראקציות חלבונים

המחקר הביולוגי נמצא בעיצומו של שינוי משמעותי הנובע מהגידול הרב ברצפי הדנ״א המפוענחים, ומהטכנולוגיות החדשות המנצלות מידע זה [1]. עד עתה פוענח רצף הגנום של יותר מ-50 אורגניזמים שונים והמידע עומד לרשות הציבור. כמו כן, מתבצעים מאות פרויקטים נוספים העוסקים בקביעת הרצף הגנומי של אורגניזמים אחרים [2-4]. ניסויים רבים מבוצעים כיום תוך התבססות על הידע הגנומי ההולך וגדל, ומאופיינים בהיקפים חסרי תקדים של מידע מצטבר [1].

רצף הדנ״א מהווה תמונה סטטית בלבד של הגנים העומדים לרשותו של התא, בעוד שחייו של תא הם תהליך דינאמי, אשר לכל אורכם התא מגיב לסביבתו, ע״י שינויים בכמויות תוצרי הגנים השונים, בקצב התבטאות הגנים, במודיפיקציות החלות בחלבונים לאחר תרגומם, באינטראקציות בין חלבונים, בשינוי מיקומן של מאקרו-מולקולות בתא, ובאופן בו משפיעים גורמים אלו על ביטויים של גנים נוספים [5].

האתגר העומד בפני החוקרים כיום הוא הבנת תפקידם של הגנים השונים ותוצריהם והבנת מנגנוני הבקרה של ביטוי הגנים [2]. רצף הדנ״א בעצמו אינו מספיק כדי להבין מהו תפקידם של הגנים השונים, כיצד פועלים תוצריהם החלבוניים, כיצד מעוצב מבנה התא, כיצד יוצרים התאים רקמה ואורגניזם, מה משתבש בתא בשעת מחלה, וכיצד ניתן לנצל את הבסיס למחלה לפיתוח תרופה [1].

ההתייחסות הקלאסית לתפקידו של חלבון התרכזה בפעולתה של מולקולת חלבון יחידה, כגון זירוזו של תהליך ביוכימי כלשהו, או קישור אל מולקולה אחרת. כיום, מתייחסים לתפקיד זה כאל ״תפקידו המולקולארי״ של החלבון, בכדי להבדילו מ״תפקידו התאי״ של החלבון, בהיותו חלק ממערכת רחבה של מולקולות המצויות באינטראקציה זו עם זו [6]. מטרת המחקר הגנומי אינה רק יצירת רשימה של גנים קיימים ותפקידיהם, אלא גם הבנת האופן בו הם ותוצריהם מתפקדים יחדיו, ליצירת תאים חיים ואורגניזמים שלמים [1].

עקב מורכבותן הרבה של המערכות הביולוגיות, לא סביר להניח, כי ניתן יהיה בעתיד הקרוב ליצור דיאגראמה מפורטת ומלאה של כל המסלולים התאיים, אף לא עבור יצורים אאוקריוטים חד-תאיים כגון שמר האפייה *S. cerevisiae*. אולם אפילו יצירתה של דיאגראמה כללית ולא

פרטנית יכולה לסייע רבות בארגון המידע הקיים כיום, בבחינתן של השערות חדשות, במתן פרשנות לתצפיות, בתכנון ניסויים, ואף כשלד לבניית מודלים שלמים ומפורטים יותר [1].

2.1.1 <u>השימוש במידע הלקוח מאורגניזמים שונים לחקר חלבוני האדם</u>

הפשטות הביולוגית של אורגניזמים פשוטים, כגון השמר, והיכולת לבצע בהם מניפולציות, הופכים אותם למודלים מצוינים להבנת הפונקציה של גנים רבים [7]. עקב השתמרותם האבולוציונית של חלבונים רבים, הכרת תפקידיו של חלבון באורגניזמים פשוטים וידע על מוטציות הגורמות לפנוטיפים ספציפיים באורגניזם כזה, מאפשרים לעיתים קרובות להסיק על תפקיד החלבונים ההומולוגים באדם, ולהבין מחלות תורשתיות באדם, הפוגעות בתפקוד חלבונים אלה [8].

ראוי לציין בהקשר לכך, כי 37% מהחלבונים בשמר ו- 36% מהחלבונים בנמטודה *C. elegans* הינם בעלי הומולוגיה חזקה לחלבונים באדם [9].

2.1.1.1 השמר – *Saccharomyces cerevisiae*

אחד המודלים העיקריים בו נעשה שימוש לחקר האדם הינו שמר האפייה *S. cerevisiae*. בעבר שימש אורגניזם זה במחקרים גנטיים קלאסיים רבים, וכעת הוא מהווה מודל אידיאלי למחקרים פונקציונאליים רחבי-היקף. השמר הינו אורגניזם חשוב ואינפורמטיבי לצורך חיזוי תפקודי גנים באדם; ל- 50% בקירוב מהגנים האנושיים המעורבים במחלות תורשתיות יש הומולוגים שמריים [2].

יתרונות נוספים של השמר טמונים הן במבנה הגנום שלו, והן בתכונותיו הביולוגיות: הגנום של השמר הינו צפוף בהשוואה לגנומים של אורגניזמים אאוקריוטיים אחרים - הרצפים המקודדים לחלבון מהווים 70% מתוך הרצף הגנומי של השמר (ללא DNA ריבוזומלי). גודלו של הגנום הינו 12 מיליון בסיסים, ומספר הגנים החזויים בו הינו כ-6200, כלומר צפיפות ממוצעת של גן אחד לכל אלפיים זוגות בסיסים [10]. כמו כן, רק 263 מגנים אלו מכילים אינטרונים [11], דבר המפשט מאוד את תהליך זיהוי הגנים הממוחשב.

ככלי מחקרי, קל לגדל את השמר במעבדה, והוא יציב הן במצב ההפלואידי והן במצב הדיפלואידי, נתון המהווה יתרון בחקר מוטציות רצסיביות ובאפיון תפקודם של גנים. יתרון נוסף הוא כי מחדרים של DNA נוטים לעבור בשמר רקומבינציה הומולוגית, דבר המאפשר מניפולציות גנטיות שונות, בייחוד targeted mutagenesis [2].

עד כה אופיינו 3780 מהגנים של השמר בשיטות גנטיות וביוכימיות. ל-560 גנים נוספים ישנם הומולוגים מוכרים בזנים אחרים, אשר מרמזים לגבי תפקודם. תפקודם של כ-1900 גנים נוספים עדיין אינו ידוע [11].

2.2 שיטות לחיזוי פונקציה של חלבון

2.2.1 חיפוש מוטיבים מוכרים ברצף החלבון

מוטיב הינו אזור ברצף החלבון, שיש לו מאפייני רצף מוכרים, ותפקיד מבני או פונקציונאלי, משוער או ידוע, ולפיכך זיהויו יכול להוסיף מידע לגבי תפקיד החלבון בו הוא מצוי. קיימים סוגים רבים של מוטיבים, המספקים סוגים שונים של מידע לגבי החלבון, החל ממידע נקודתי, כגון אתר גליקוזילציה, ועד למידע כללי, המשייך את החלבון למשפחת חלבונים ומסייע בחיזוי תפקידו התאי [12].

דוגמא למוטיבים פשוטים הינם רצפים קצרים, הכוללים חומצות אמיניות מסוימות, ועומדים לרוב בפני עצמם, ללא קשר אל הרצף המקיף אותם. מוטיבים כאלה מציינים, למשל, אתרי פוספורילציה, גליקוזילציה, מיריסטילציה, אתרי קישור לאזורים פונקציונאליים מורכבים יותר (כגון SH2), ועוד.

סוג אחר של מוטיבים, הינם רצפים המשפיעים על מבנהו המרחבי של החלבון, כגון מוטיב של coiled-coil המורכב ממספר מקטעי α-helix אמפיפאטים[‡] הנכרכים זה סביב זה, או מוטיב של helix-loop-helix הקושר יון סידן (Ca^{2+}), המאופיין בשיירים הידרופיליים במקומות קבועים ברצף [13, פרק 3.1].

מוטיבים ספציפיים מאפיינים חלבונים קושרי דנ״א. חלבונים כאלה הינם בעלי חשיבות ביולוגית רבה, מפני שלרוב הם מעורבים בשכפול הדנ״א או בביטויו של מידע גנטי. לכאורה, פניו החיצוניות של סליל הדנ״א כמעט ואינן מושפעות מרצף הנוקליאוטידים, אולם קצותיהם של הנוקליאוטידים הינם גלויים, בעיקר באזור ה- major-groove של הסליל הכפול. כדי שחלבון יוכל לאתר רצף מסוים בדנ״א, על ידי אינטראקציה עם הבסיסים המתאימים החשופים ב-major-groove, עליו להכיל צירוף של חומצות אמיניות בעלות קבוצות פעילות מתאימות, אשר בולטות מפני השטח שלו, והסדר שלהן מתאים לסדר הבסיסים ברצף הדנ״א [14, פרק 8.3.2].

[‡] מולקולה אמפיפאטית - מורכבת מאזורים הידרופיליים ומאזורים הידרופוביים.

המוטיב המבני המוכר ביותר בקבוצת המוטיבים קושרי הדנ״א, הינו ה- helix-turn-helix, אשר נצפה במספר רב של חלבונים, שמלבד זאת אין ביניהם כל דמיון מבני נוסף. חלק מהשיירים במוטיב זה דומים בחלבונים השונים, ומטרתם כנראה לייצב את מבנהו של המוטיב. התבנית הקבועה יחסית של מוטיב זה, מאפשרת לאתרו ולזהות בעזרתו חלבונים העשויים להיות קושרי-דנ״א, וזאת ע״פ הרצף שלהם לבדו [14, פרק 8.3.2].

מבנה קושר דנ״א נוסף המאפשר לרצף חומצות אמיניות לבלוט מספיק אל מחוץ לפני החלבון בכדי להגיב עם נוקליאוטידים ב- major-groove, הינו מבנה אצבעות האבץ (zinc-fingers). מאות גרסאות של מוטיב זה נמצאו בחלבונים העוסקים בבקרת ביטוי של גנים באאוקריוטיים [14, פרק 8.3.2].

קיימים מוטיבים אשר אינם מבוססים על רצף קבוע של חומצות אמיניות, אלא על מאפיינים אחרים לגבי הרצף, כגון תכונותיהן של החומצות האמיניות (הידרופיליות, פולאריות), שכיחותה של חומצה אמינית מסוימת (אזור עשיר בפרולין או בגליצין), וכיו״ב [13, פרק 3.1]. מוטיבים אלו מאפשרים לזהות, למשל, אזורים ממברנליים בחלבון, רצפי איתות (signal sequences), או רצפים מזהים אחרים [12].

אזורי החלבון הנמצאים בתוך ממברנת התא ניתנים לזיהוי ע״פ רצפים הידרופוביים המצויים בתוכם, והדומים לאלו הנמצאים בחלבונים ממברנליים מוכרים. לעיתים, קשה להבדיל בין קטעים אלו, לבין קטעים הידרופוביים הנמצאים בחלקם הפנימי של חלבונים גלובולריים. אזורים ממברנליים מסוימים עשויים לא להתגלות בעזרת השוואת רצפים כזו, אם הם מכילים כמות ניכרת של שיירים פולאריים [14, פרק 7.2.3].

המוטיבים הזוכים כיום לתשומת הלב הרבה ביותר, הנם רצפים אשר מאפשרים להבדיל קבוצה מסוימת של חלבונים מכל האחרים. רצפים אלו משקפים לרוב מאפיינים מבניים ופונקציונאליים של קבוצת חלבונים מסוימת, והם מעידים על מקור משותף לחלבונים הכלולים בקבוצה. מוטיבים כאלו מאפשרים לשייך חלבונים חדשים אל משפחות החלבונים אותן הם מגדירים, ובכך לקבל תחזיות מבניות ופונקציונאליות לגבי חלבונים אלו [12].

קיימים כיום מאגרי מידע רבים, אשר מתבססים על מוטיבים אלו לשם אפיון משפחות החלבונים הנמצאות במאגר, ואשר משמשים לאפיון חלבונים חדשים על פי הרצף שלהם [15]. השימוש במשפחות חלבונים לחיזוי פונקציונאלי, מאפשר לנצל את המידע הקיים לגבי כל אחד מהם למציאת המכנה המשותף המאפיין את הקבוצה ולביצוע חיזוי פונקציונאלי זהיר ומדויק

יותר [16]. כמו כן, ידע לגבי המבנה המרחבי או המנגנון הקטליטי של חלק מחלבוני המשפחה, מאפשר להרחיב את המידע ולחזות מאפיינים אלו גם לגבי חלבונים נוספים בקבוצה [15].

בעבר התמקדה האנליזה מסוג זה במוטיבים קלים לזיהוי, בעיקר מוטיבים הקיימים באנזימים מטבוליים וקשורים לפעילותם הקטליטית. כיום נחקרים גם מוטיבים מגוונים יותר, המאפיינים חלבונים מבניים ורגולטוריים. הגדרת מוטיב יכולה להיעשות ע"י התבוננות בחלבונים בעין, או תוך שימוש בכלים ממוחשבים, כגון תוכנות להשוואת רצפים (sequence alignment). שימוש בעימוד (alignment) של חלבונים אורתולוגיים זה מול זה, מאפשר לזהות מוטיבים פחות בולטים, ובעזרתו ניתן לחזות, מי הם השיירים אשר משמעותיים לתפקודם של החלבונים בקבוצה [12].

2.2.2 <u>חיפוש חלבונים הומולוגיים שתפקידם ידוע</u>

חלבונים הומולוגיים מראים דמיון רב ברמת רצף החומצות האמיניות שלהם. זיהוי חלבונים הומולוגיים לחלבון נתון, כאשר הם עצמם בעלי תפקיד ידוע, יכול לסייע בחיזוי תפקידו של החלבון הנבחן, שכן ההנחה היא, שרצף דומה מרמז על תפקידים דומים ואף על מקור משותף (למשל, היווצרות שני גנים מגן קדמון על ידי דופליקציה) [6].

ההומולוגיה ברמת הרצף עשויה להשתמר במהלך האבולוציה, ועימה תפקידו של החלבון. לפיכך, גם חלבונים הומולוגיים באורגניזמים אחרים שתפקידם ידוע, תורמים מידע על תפקידו האפשרי של החלבון, ומובילים לבחינת תפקידו הביוכימי [17]. לעיתים, קיים הומולוג שתפקידו ידוע באורגניזם ירוד כשמר, אך ניתן להסיק ממנו על תפקידו של חלבון שהתגלה באדם.

בכדי לבדוק האם שני רצפים הינם הומולוגים זה לזה, יש לעמד את רצפיהם זה מול זה (alignment), ובתוך כך לבחון אם קיימים מחדרים (insertions) או השמטות (deletions) של קטעים בחלבון האחד ביחס למשנהו. המטרה באנליזה זו היא למצוא את ההתאמה הטובה ביותר בין הרצפים, כאשר איכות ההתאמה עומדת ביחס הפוך למספר הבדלי הרצף הקיימים בין שני החלבונים [14, פרק 3.2.1].

תוכנות מחשב רבות קיימות למטרה זו, ורובן מבצעות חיפוש של רצפים הומולוגים לרצף נבחן, ע"י השוואתו לכל הרצפים הנמצאים במאגר מידע. לצורך ביצוע משימה זו בזמן סביר, על התוכנות השונות להיות מתוחכמות מאוד. כדי להעריך את הדמיון בין שני רצפים, משווים אותו לדמיון הצפוי להתקבל בהשוואת רצפים אקראיים. חומצות אמיניות בעלות אופי כימי דומה נוטות להחליף זו את זו בחלבונים קרובים, והשימוש במטריצת דמיון (similarity index

matrix), מאפשר להעריך את מידת ההתאמה בין זוגות חומצות אמיניות שונות המוצבות זו מול זו [14, פרק 3.2.1]. סדרת התוכניות המצויה בשימוש הנרחב ביותר כיום, לשם ביצוע חיפוש אחר רצפים הומולוגים מכונה Basic Local Alignment Search Tool - BLAST [4, 18].

כיום ניתן בעזרת זיהוי חלבונים הומולוגים בעלי תפקיד ידוע, לקבל חיזוי פונקציה ברמה כלשהי, לכ- 40-70% של רצפי גנום חדשים [6]. עם זאת, ישנם חלבונים רבים, אשר לגביהם לא ניתן להשתמש בשיטה זו, משום שלא ידוע תפקידו של החלבון ההומולוגי שזוהה, או משום שלא ניתן למצוא כל הומולוג עבורם. במרבית הגנומים שהרצף שלהם פוענח, אין כיום תחזית פונקציונאלית ברורה לכ- 30-35% מהגנים [19]. למשל, בשמר האפייה, מתוך 6217 חלבונים, 2557 חלבונים עדיין לא נחקרו באופן ניסויי, ואין להם הומולוגיה חזקה לחלבון בעל פונקציה ידועה [20].

אחת הבעיות בחיזוי תפקיד ע"פ הומולוגיה, הינה ריבוי הומולוגים. יתכן כי לחלבון יחיד באורגניזם מסוים יהיו באורגניזם אחר מספר הומולוגים המהווים משפחה. תופעה זו דורשת תשומת לב מיוחדת בעת חיזוי הפונקציה, מפני שיתכן ולחלק מההומולוגים נוספו תפקידים חדשים במהלך התפתחותם, אשר לא היו בחלבון הקדמון המשותף [19].

2.2.3 <u>חיזוי מבנה החלבון</u>

המבנה השלישוני של חלבונים השתמר באבולוציה במידה רבה יותר מאשר המבנה הראשוני (רצף החומצות האמיניות). בד"כ, חלבונים הומולוגים, אשר מתוך הדמיון ברצפים שלהם ניתן להניח, שמקורם באותו חלבון קדמון, הינם גם בעלי מבנה דומה מאוד. אך קיימות גם דוגמאות בהן לא נמצאה הומולוגיה ברמת הרצף בין חלבונים, אשר נתגלו כבעלי דמיון מבני, ובעלי פונקציה דומה [14, פרק 6.4].

השימור הקפדני של המבנה המרחבי מוסבר בכך, שלחלבונים הומולוגים בזנים שונים פונקציה דומה, ולכן דרישות המבנה לגבי חלבונים אלו דומות מאוד. יתכן שהמבנה התלת-מימדי של החלבון הוגדר בשלב מוקדם בהתפתחות, והוא חיוני לתפקודו של החלבון. על שינויי הרצף המצטברים לאורך השנים כתוצאה ממוטציות, להיות תואמים את המבנה המרחבי של החלבון, כך שלא יפגעו בתפקודו [14, פרק 6.4].

לפיכך, ידע לגבי המבנה המרחבי של חלבונים הומולוגים, יכול לסייע בזיהוי השיירים התואמים זה לזה בחלבונים השונים, ובעימודם הנכון זה מול זה לצורך השוואת הרצפים שלהם. טכניקות העימוד הרגילות, המתבססות על השוואת רצפים לינאריים המוצבים זה מול זה תוך מזעור מספר השינויים בין הרצפים, נותנות לעיתים קרובות תוצאות, אשר אינן מתיישבות עם ההתאמה המבנית [14, פרק 6.4].

מבנהו המרחבי של חלבון, אשר יש לו הומולוג בעל מבנה ידוע, ניתן לחיזוי בצורה מדוייקת למדי, תוך שימוש בשלד המבני של ההומולוג [21]. הבעיה הופכת למורכבת, כאשר אין לחלבון הומולוג בעל מבנה ידוע, שכן קיימת גמישות רבה במעבר בין מבנהו הראשוני של חלבון למבנהו השלישוני (עליה יעיד מספרם הרב של הרצפים החלבוניים השונים אשר להם מבנה מרחבי זהה). עקב מספרם הגדול של המבנים האפשריים עבור שרשרת פוליפפטידית נתונה, לא ניתן כיום לבדוק בזמן סביר את כולם בזה אחר זה, במטרה לזהות את המבנה בעל האנרגיה החופשית הנמוכה ביותר [14, פרק 6.5].

לעיתים, רואים דמיון במבנה השלישוני של חלבונים גם במקרים בהם אין לכך, למראית עין, כל סיבה אבולוציונית או פונקציונאלית [14, פרק 6.4]. מתצפית בכלל המבנים המרחביים שפוענחו עד היום, עולה ההשערה כי מספר הקונפורמציות הבסיסיות המשמשות לכל החלבונים הינו מוגבל, ונאמד בכמה מאות [14, פרק 6.5]. היות ולעיתים קרובות יש לחלבונים שונים מבנה תלת-מימדי דומה, עשוי כל מבנה חלבוני חדש לשמש כמודל למבנה של מספר חלבונים. אם אכן מספר הקיפולים האפשריים של השרשרות הפוליפפטידיות הוא מוגבל, ניתן לפשט את חיזוי מבנהו של חלבון לבעיה של מציאת הקיפול המתאים ביותר עבור רצף החלבון, מתוך האוסף המוגבל של קיפולים אפשריים מוכרים [21].

מבנהו המרחבי של חלבון מקיים תמיד מספר חוקים בסיסיים: חלקו הפנימי של החלבון ארוז בצפיפות, ללא חללים גדולים, קבוצות טעונות ופולאריות נמצאות על פני החלבון, אלא אם הן קשורות זו לזו בקשרי מימן, וזוויות הפיתול של קשרים מכוונות ברוב המקרים למזער את הלחץ המבני. חוקים אלו אינם מפתיעים, שכן הם עוזרים להקטין את האנרגיה החופשית של המבנה. בעזרתם ניתן לבחון את מידת התאמתו של רצף נתון למבנה שלישוני מסוים [14, פרק 6.5].

חיפוש המבנה התלת-מימדי המתאים לרצף נתון, יכול לגלות קשרים מבניים, אשר אינם נראים בבירור מתוך השוואת רצפים [21]. שיטות שונות להשוואת המבנה של חלבונים משתמשות במאגרי מידע לצורך סיווגם של החלבונים וחלוקתם למשפחות [15].

2.2.4 <u>בדיקת מיקומו התאי של החלבון</u>

זיהוי מיקומו התאי של החלבון עשוי לסייע בהבנת תפקידו. קיימות שיטות שונות לכך, ביניהן immuno-localization, שבה נעשה שימוש בנוגדנים לצביעת החלבון *in situ*, פרקציונציה של תכולת התא וזיהוי המקטע בו מצוי החלבון, ואף חיזוי של מיקום החלבון על פי מוטיבים המצויים ברצף שלו, כגון רצף המצביע על מיקום גרעיני (nuclear localization signal – NLS).

בשיטת ה-immuno-localization, משתמשים בנוגדנים המזהים מולקולות מסוימות לצורך זיהוי מיקומן של מולקולות אלו בתא. הנוגדנים מסומנים בצבע פלואורסצנטי, לשימוש במיקרוסקופ פלואורסצנטי, או בחלקיקים צפופי אלקטרונים כגון colloidal gold sphere, לעבודה ברזולוציה גבוהה עם מיקרוסקופ אלקטרונים [22, פרק 4]. שיטה אחרת לאיתור החלבון בתא, היא לסמנו ע״י epitope-tagging ולאתרו בהמשך בעזרת נוגדן המזהה את האפיטופ שהוצמד לחלבון (ראה להלן) [23, 24].

פרט לשימוש בשיטות אלו לצורך איסוף מידע ראשוני לגבי החלבון, ניתן לבצע בעזרתן co-localization כדי להשוות את מיקומם התאי של חלבונים שונים. השוואה כזו מסייעת באישוש השערות לגבי קיומן של אינטראקציות בין חלבונים *in vivo*, לאחר שזוהו בניסויים *in vitro* [25], כגון co-immunoprecipitation, או two-hybrid screens (ר׳ בהמשך).

Epitope Tagging 2.2.4.1

בשיטה זו מייצרים בתאים חלבון רקומביננטי, שאליו מוצמד פפטיד קצר (אפיטופ), אשר ניתן לזיהוי ע״י נוגדנים מונוקלונליים מסחריים בעלי אפיניות גבוהה לאותו פפטיד, ואשר אינם מגיבים עם חלבוני התא האחרים. בעזרת נוגדנים אלו נוכל לזהות את החלבון שסומן [2], ולבצע ניסויים המבוססים על שימוש בנוגדנים, כגון immunolocalization, ו- immunoprecipitation (סעיפים 2.2.4 ו-2.3.2.1). לרוב האפיטופ מוצמד אל אחד מקצותיו של החלבון, אך ניתן למקמו גם בתוכו, בתנאי שפעילותו של החלבון לא נפגעת כתוצאה מכך. ע״י הוספת מספר עותקים רצופים של האפיטופ, ניתן להגביר את רגישות הזיהוי [26].

היתרון בשיטה זו על פני השימוש בנוגדנים לחלבון עצמו, הוא בכך, שאין צורך לפתח נוגדנים מיוחדים לו, ונעשה שימוש בנוגדנים מאופיינים היטב, אשר תגובתם עם מרכיבי התא האחרים מוכרת. כמו כן, קיימת ביקורת שלילית יעילה לניסוי: תא שאליו לא הוחדר החלבון המסומן [27]. החיסרון בשיטה זו הוא, שחלבון רקומביננטי, אשר לרוב מבוטא ברמה גבוהה לצד החלבון האנדוגני הקיים גם הוא באותו תא, נוטה להתמקם בתא ולעיתים אף לפעול, בדרך שאינה דומה לזו של החלבון האנדוגני. לכן, תוצאות המושגות בשיטה זו, אינן משקפות תמיד סיטואציה ביולוגית נורמאלית.

2.2.5 <u>פגיעה בביטוי החלבון או במבנהו ככלי לאנליזה תפקודית</u>

שיטה רבת-עצמה למציאת תפקידו של גן, היא ניתוח פנוטיפי (phenotypic analysis) של זנים מוטנטים אשר בהם גן זה חסר או פגום [28]. שיטות שונות שמטרתן לקבוע את תפקידם הביולוגי של גנים, מתבססות על פגיעה בגן ובדיקת עמידותו של הזן שהתקבל תחת תנאים פיזיולוגיים שונים [29].

ניתן לבצע מוטגנזה מכוונת (targeted), של גנים ספציפיים במגוון אורגניזמים, החל משמר, המשך בנמטודה, בזבוב התסיסה, ועד ליונקים. פגיעה מכוונת בגנים ספציפיים מבוצעת בד״כ ע״י בניית מחדרים שבקצותיהם אזורים חופפים ברצף שלהם לרצפים מקבילים בגן, והחדרתם לגן בתהליך של רקומבינציה הומולוגית [22, פרק 4]. תהליך זה מאפשר ליצור באופן דקדקני את המוטציות הרצויות, אך הוא מורכב ודורש משאבים רבים. דרך אחרת לפגיעה בביטויו של גן מסוים היא ע״י השתקתו (gene silencing). לשם כך, אפשר להשתמש במולקולות דנ״א או רנ״א, שרצפן משלים לאזור מתוך רצף הגן (antisense), אשר נקשרות ל-mRNA של גן זה ובכך גורמות לחיתוכו ו/או להפרעה בתרגומו [30, 31]. שיטה אחרת, חדשה ויעילה יותר, עושה שימוש ב-siRNA, שהן מולקולות רנ״א דו-גדילי קצרות, לרוב באורך 21-23 נוקליאוטידים. מולקולות אלו גורמות להפעלת מנגנון תאי, אשר מפרק את מולקולות ה-mRNA, שרצפן הומולוגי באופן מושלם לזה של מולקולות ה-siRNA, מבלי לפגוע בביטויים של גנים אחרים. עוצמתה של טכנולוגיה חדשה זו תאפשר בעתיד הקרוב ניתוח פונקציונאלי רחב-היקף של גנים בתרביות תאי יונקים [31-33].

ניתן גם לקבל מוטציות בצורה אקראית בעזרת שימוש בחומרים מוטגניים [22, פרק 4], או ע״י בשימוש בטרנספוזון אשר יחדור לאזורים שונים בגנום [2]. יצירת הזנים המוטנטים בשיטה זו הינה מהירה יחסית, אך בהמשך התהליך, מציאת ההתאמה בין פנוטיפ ספציפי לבין הגן הגורם לו, היא איטית [28]. כמו כן, בעבודה עם טרנספוזון אין לחוקר דרך לכוונו, והיות והטרנספוזון נוטה להעדיף רצפי דנ״א מסוימים על אחרים, כמעט בלתי אפשרי לקבל כיסוי מלא של הגנום ע״י המוטציות הנוצרות, גם כאשר משתמשים במדגם גדול [2, 28].

ניתוח פנוטיפי של זנים בעלי גן פגום (או חסר) הינו משימה קשה, היות ותפקידיהם של רבים מהם יבואו לידי ביטוי רק בתנאי גידול מסוימים, או במצבים פיזיולוגיים ספציפיים, עובדה המסבכת מערכת ניסויית זו. לכן ישנו יתרון חשוב לשיטות, אשר מאפשרות לבדוק במקביל מספר גדול של גנים. אכן קיימות מספר שיטות המאפשרות לבצע בדיקות פנוטיפיות בקנה מידה גנומי [28].

Genetic footprinting 2.2.5.1

בשיטה זו מחדירים טרנספוזונים לתוך אוכלוסייה גדולה של תאים, כדי שיחדרו באקראי לאזורים רבים בגנום של תאים אלו, ובכך יגרמו לפגיעה בפעילותם של גנים שונים. אוכלוסיית התאים מחולקת לדגימות, וכל אחת נחשפת לסוג שונה של סלקציה. שימוש ב-PCR עם פריימרים מתאימים, והשוואת המקטעים שהוגברו בדגימות השונות†, מאפשרים לזהות הבדלים בשרידות התאים בתנאי סלקציה שונים, ולאפיין את הגנים המעורבים בכך [29].

יתרונה של שיטה זו הוא בכך שניתן לבדוק יחסית במהירות את תרומתם של כל הגנים השונים לעמידותו של הזן בתנאי גידול מסוימים [28].

Barcoded deletions 2.2.5.2

בשיטה זו החסרת הגנים השונים מהגנום מבוצעת באופן מדויק, כאשר כל גן מוחסר ע״י רקומבינציה הומולוגית, ובמקומו מוחדר קטע הכולל שני ״barcodes מולקולאריים״ (Uptag ו-Downtag), שהם רצפים בני 20 בסיסים, אשר משמשים לזיהויו של הזן שבו חל האירוע [2, 28]. barcodes אלו מאפשרים לקבל מידע כמותי לגבי שכיחותם היחסית של הזנים בתערובת ע״י היברידיזציה שלהם למערך של אוליגונוקלאוטידים, ומאפשרים להעריך את קצב גדילתם של הזנים השונים [28, 34].

בעזרת שיטה זו ניתן ליצור באופן סיסטמתי, אלפי זנים שונים שכל אחד מהם מכיל חסר של גן שונה. מספר מעבדות באירופה ובארה״ב פועלות בשיתוף, במטרה לסיים את יצירתה של סדרת זנים, המכילים חסרים מסומנים, עבור כל מסגרות הקריאה בשמר *S. cerevisiae*. למרות שיצירת הזנים השונים בשיטה זו דורשת מאמץ ניכר יחסית, הם יכולים לשמש בניסויים חוזרים בהיקף נרחב. בדיקתם של אלפי זנים במקביל ובאופן כמותי, תפחית את כמות העבודה והחומרים הנדרשים, ותעלה את רמת המהימנות של התוצאות ושל עיבודן [28].

---

† על מנת לזהות גן אשר חשוב לשרידות התאים, משתמשים בדנ״א שהופק מהתאים ששרדו בדגימות השונות (כל דגימה בנפרד). מבצעים PCR ע״י שימוש באוסף פריימרים המתאימים לצד אחד בלבד של הגנים השונים המעניינים אותנו, ואילו הפריימר הנגדי יתאים למחדר שהוכנס ע״י הטרנספוזון. כך מתקבלת הגברה רק של אותם המקטעים אשר בהם חדר הטרנספוזון לגנים הנבדקים. משווים את המקטעים שהוגברו בדגימות השונות ע״י הרצתם בג׳ל, במטרה לזהות פסים, אשר אינם מופיעים בדגימה כלשהי. הפסים החסרים מקורם באותם תאים, אשר לא שרדו את הסלקציה שעברה דגימה זו (אך שרדו בדגימות אחרות). הגן הפגום בתאים אלו לא אפשר להם לשרוד את הסלקציה, ולכן חסרים הפסים המייצגים אותו, שמקורם בתאים אלו. לפיכך, הגן המיוצג ע״י פסים אלו חשוב לשרידות תחת הסלקציה שהופעלה. נותר כעת לזהות את הגן המאופיין ע״י הפסים.

2.2.6 שימוש במערכי חלבונים ממוזערים (protein microarrays)

השיטות הביוכימיות המסורתיות מספקות מידע רב ערך לגבי תפקידם של חלבונים, אך חסרונן הוא בכך, שהן מבוצעות עבור כל חלבון בנפרד, והן אינן מעשיות למחקר רחב היקף של מספר רב של חלבונים בבת אחת. כדי להשיג מטרה זו, נעשה שימוש במערכי חלבונים ממוזערים, שמטרתם לאפשר ביצוען של בדיקות שונות לגבי אלפי חלבונים במקביל, ועם זאת להקטין את הכמות הדרושה מכל חלבון לצורך ביצוע הבדיקה [2, 35]. טכנולוגיית מערכי החלבונים הממוזערים נמצאת עדיין בתחילת דרכה. קבוצות חוקרים שונות עוסקות בפיתוחם של דרכים שונות לייצור מערכים אלו, ובהדגמת ישימותם המחקרית [2].

לצורך פיתוחם של protein microarrays, נדרש אוסף חלבונים מנוקים (בד"כ נשאף לקבלת אוסף כל החלבונים של אורגניזם מסוים) [2]. מספר קבוצות מחקר עוסקות כיום בניסיונות להפיק מספר גדול מהחלבונים של השמר *S. cerevisiae*, על ידי הצמדתם לפרומוטור מתאים לצורך הגברת ביטויים, וסימונם בדרכים שונות על מנת לנקותם ולהפרידם משאר חלבוני התא [2, 36]. עבודות אלו יוכלו לסייע ביצירת מערכי חלבונים נרחבים. במקביל לכך, עוסקים החוקרים בפיתוח טכנולוגיות להצמדת החלבונים למשטח [35-37]. השיפור העיקרי הדרוש הוא ברמת הדיוק של פעולה זו ובשיעור החלבונים השומרים על פעילותם לאחר קיבועם למשטח [2].

מערכי החלבונים נמצאים עדיין בשלב האב-טיפוס, ודרושה התקדמות טכנולוגית נוספת לפני שיהפכו לכלי מחקרי נפוץ.

2.2.7 שיוך פונקציונאלי של חלבון נבדק לקבוצת חלבונים ספציפית

שיטות חדשות פותחו בכדי לחזות את תפקידם של חלבונים רבים במקביל. שיטות אלו מבוססות לרוב על מציאת קשר פונקציונאלי בין חלבונים. אם הפונקציה של אחד החלבונים ידועה, ניתן להניח כי החלבונים האחרים משתתפים באותו מסלול פיזיולוגי, או בתהליכים דומים בתא, וייתכן כי קיימת אף אינטראקציה פיזית ביניהם. במידע זה ניתן להשתמש לתכנון ניסויים עתידיים בחלבונים אלו [6].

העבודות המנסות לזהות קשרים פונקציונאליים בין חלבונים, יכולות להתרכז בגן מסוים, או להתבונן במכלול של מספר חלבונים. קיימות לכך שיטות ביולוגיות ניסוייות ושיטות חישוביות. מטרתן המשותפת היא לזהות את רשת האינטראקציות הפיזיות והפונקציונאליות בין חלבוני התא השונים, אשר תאפשר לחזות את תפקידו של חלבון לא מוכר, מתוך יחסיו עם שכניו. שיטות אלו, מתוארות בהרחבה בפרק הבא.

2.3 שיטות לחיפוש אינטראקציות בין חלבונים

שיטות שונות משמשות למציאת קשרים פונקציונאליים בין חלבונים, אשר יוכלו לסייע לחוקרים להשליך מתפקידו הידוע של חלבון מסוים אל תפקידם של חלבונים אחרים אשר פועלים יחד עימו [6]. שיטות אלו אינן מתבססות על דמיון ברצף החלבונים, או במבנה שלהם, ולכן מאפשרות לחזות פונקציה גם לחלבונים להם אין הומולוג ידוע [38].

ניתן לחלק את האינטראקציות בין חלבונים לשלושה תחומים פונקציונאליים עיקריים [39]:

- מסלולים מטבוליים ומסלולי העברת אותות (signaling).
- מסלולים מורפוגניים, שבהם קבוצות חלבונים משתתפות בתפקוד תאי מסוים בשלב התפתחותי כלשהו.
- קומפלקסים מבניים ומכונות מולקולאריות המורכבים ממקרו-מולקולות רבות המצויות באינטראקציה פיזית ופועלות יחדיו.

השיטות השונות לחיפוש אינטראקציות נבדלות בסוג הקשרים אשר מגלים בעזרתן ובמידת הפרטנות שלהם. בעוד ששיטות מעטות בלבד מספקות מידע לגבי השיירים על גבי החלבון אשר משתתפים בתהליך, מרבית השיטות מספקות מידע לגבי עצם קיום האינטראקציה הפיזית (למשל, two-hybrid screens, או שיקוע אימוני משותף). לעיתים, מודלים בבעלי חיים (כגון knockout mice), מספקים מידע על אינטראקציות בין חלבונים, אם כי ברמה כללית ביותר [39].

חשוב להדגיש, כי מידת האמינות של השיטות השונות אינה אחידה, וכי שיטות מסוימות מביאות לתוצאות שגויות, יותר מאחרות. כמו כן, ככל שהיקף הניסוי גדול יותר, רב הסיכוי לתוצאות מוטעות, בין השאר, כי קשה יותר לנתח כל תוצאה באופן פרטני. מידת המהימנות של ממצא לגבי אינטראקציה תעלה, אם נוכל למצוא לאינטראקציה זו חיזוקים במספר שיטות שונות [39].

2.3.1 דליית מידע מן הספרות המקצועית ועיבודו

המידע הביולוגי המופיע בספרות המקצועית, מנוסח בשפה טבעית, תוך שימוש באוצר מילים בלתי מוגבל וללא כללי תחביר מוגדרים, ולפיכך עיבודו ישירות מן הטקסט בעזרת כלים ממוחשבים הוא בעייתי [1, 40]. בעיה מרכזית עמה מתמודדת הביואינפורמטיקה כיום, היא ארגונו של מידע מסוג זה באופן נוח לעיבוד ממוחשב [41].

כיום קיימים מאגרי נתונים רבים, אשר המידע בהם מבוסס על הספרות המקצועית, ואשר מוקדשים לתחומים ביולוגיים שונים, כגון מאגרי מידע של אינטראקציות חלבונים (BIND [42]

ו-DIP [43], ראה סעיף 2.4), מאגרי מידע של מסלולים מטבוליים (KEGG [44] ו-EcoCyc [45]), ועוד.

חשיבותם של מאגרי המידע על אינטראקציות באה לביטוי במספר צורות: בראש ובראשונה, מאגרי מידע אלו מאפשרים לחוקרים לאמת או לשלול את תוצאות ניסוייהם. בנוסף לכך, איסופן וארגונן של האינטראקציות המדווחות בספרות, מאפשרים לחוקר לסייר על פני מפת האינטראקציות המתקבלת, ולגלות מסלולים ומנגנוני בקרה חדשים. ריכוז המידע מאפשר לחקור את תכונותיהן של רשתות האינטראקציות ואת צורת ארגונן [39].

מאמצים להרחבתם של מאגרי מידע אלו מתעכבים בגלל נפחו העצום של המידע הביולוגי בספרות: כיום, קיימות מעל 10 מליון רשומות ב-MedLine [4]. הפיקוח על המידע המוזן אל מאגרי מידע רבים, כגון DIP [43], SwissProt [46], ו- YPD [11], מבוצע באופן ידני כדי להבטיח את אמינותם. תהליך זה הינו "צוואר הבקבוק", הקובע את קצב הגידול של מאגרי המידע [41].

לשם האצת תהליך העיבוד של מידע מן הספרות המקצועית, נעשו ניסיונות שונים לבצע את כולו או את חלקו באופן ממוחשב. עבודות מסוימות מנסות לנתח את המידע ולהמירו באופן ישיר לטבלאות ממוחשבות [40], בעוד שאחרות מנסות לאתר את המאמרים הרלוונטיים ביותר בכדי להביאם בעדיפות ראשונה אל שלב העיבוד הידני [41].

שיטות העיבוד השונות נתקלו בקשיים, עקב המורכבות והגיוון הקיימים בשפה האנגלית (ובשפות אנושיות בכלל). חלקן בחרו לעקוף את המורכבות הרבה של ניתוח שפת דיבור – Natural Language Processing – תוך התבססות על עיסוקו של המידע בתחום מצומצם ומוגדר, ושימוש באוסף כללים פשוט לניתוח המידע [40, 47]. ההתייחסות אל חלבון מסוים בספרות גם היא אינה חד משמעית, ונדרשו מאמצים לשם איסוף כל שמותיהם הנרדפים של החלבונים אל בסיסי נתונים [48].

להלן תיאורן של מספר שיטות שונות בהן נעשה שימוש לצורך עיבוד המידע מן הספרות:

- Sekimizu *et al.* [49] הציעו שיטה לניתוח לשוני של טקסט ע"י איתור שמות העצם המופיעים בו, זיהוי שמות פועל המרבים להופיע בטקסט, וזיהוי הנושא והמושא המתאימים לכל שם פועל מבין שמות העצם שזוהו.
- Ng *et al.* [50, 51] פיתחו מערכת בשם BioNLP אשר מחלקת את הטקסט למשפטים בודדים. כל משפט המכיל מילות מפתח כגון "activate" או "inhibit" מנותח למציאת שמות

החלבונים המופיעים בו. לאחר שימוש במילון של שמות נרדפים להשגת אחידות בשמות החלבונים, זוהו היחסים בין החלבונים ע"י התאמת המשפט לאחת מאוסף תבניות מוגדרות.

- Blaschke *et al.* [47] ניסו לזהות בספרות אינטראקציות בין חלבונים מתוך ניתוח מבני וסטטיסטי של המידע המופיע בתקצירי מאמרים (abstracts). הם השתמשו באוסף מוגדר מראש של שמות חלבונים ושל שמות פועל, שאותם הם חיפשו בתקצירים, זיהו את היחסים בין החלבונים על פי מיקומם במשפט ביחס לשמות הפועל, ובחרו את אותן האינטראקציות אשר זוהו במספר רב של משפטים שונים.
- Ono *et al.* [40] הציעו לבצע ניתוח ראשוני של מבנה המשפטים. הם השתמשו במילון לזיהוי שמות החלבונים ולזיהוי המשפטים המתייחסים ליותר מחלבון אחד. לאחר זיהוי חלקיו התחביריים של כל משפט, הם פיצלו משפטים מורכבים למשפטים הפשוטים המרכיבים אותם. המשפטים הפשוטים נותחו ע"י התאמתם לתבניות משפט מוגדרות, תוך התייחסות מיוחדת למשפטים שליליים המתארים חוסר אינטראקציה.
- מספר עבודות ניסו להשתמש במערכות IE (Information Extraction) קיימות ולהתאימן לניתוח מידע ביולוגי. Thomas *et al.* [52] השתמשו במערכת מסחרית בשם Highlight, אשר פותחה בחברת SRI International. Humphreys *et al.* [53] התבססו על מערכת שעוצבה לצורכי עיבוד הודעות חדשותיות, ע"י ה- US DARPA (Defense Advanced Research Project Agency), וביצעו בה התאמות לעיבוד מידע ביולוגי.
- Marcotte *et al.* [41] הרכיבו באופן ניסויי אוסף של מילים "מפלות" (discriminating) אשר תדירות הופעתן בתקצירי מאמרים העוסקים באינטראקציות חלבונים, גבוהה מאוד, או נמוכה מאוד, בהשוואה לתקצירים אחרים. הם השתמשו במידע זה כדי לדרג את המאמרים ב- MedLine ע"פ הרלוונטיות שלהם לאינטראקציות חלבונים.

חשוב לציין, כי על אף ששיטות אלו יכולות לסרוק את מיליוני המאמרים שנמצאים ב- MedLine, רמת האמינות שלהם פחותה באופן משמעותי מזו המתקבלת בעיבוד אנושי של המידע, והן עלולות להוביל ליותר טעויות מאשר לזיהוי אינטראקציות אמיתיות. מאגרי המידע המפוקחים באופן ידני מועדפים ע"י החוקרים, למרות חסרונם העיקרי שהוא קצב גידולם הנמוך [39].

2.3.2 שיטות מעבדה ניסוייות לזיהוי אינטראקציות בין חלבונים

2.3.2.1 בידוד חלבונים הנמצאים בקומפלקס וזיהוים בעזרת ספקטרומטר מסות

בידודו של קומפלקס חלבוני מתוך תמצית תאים הינו שיטה נפוצה לזיהוי אינטראקציות פיזיות בין חלבונים. כנקודת מוצא משמש חלבון ידוע, ומטרת הניסוי היא לזהות את החלבונים אשר נמצאים בקומפלקס החלבוני המכיל חלבון זה [26].

הקומפלקס מבודד, בד"כ, בעזרת שיטות כרומטוגרפיות ו/או שיקוע אימוני [25, 26]. החלבונים המצויים בקומפלקס מופרדים באלקטרופורזה, מוצאים מן הג'ל ומזוהים בעזרת ספקטרומטר מסות ו-microsequencing. לאחר מכן, חוזרים ומוודאים את קיום האינטראקציה *in vivo* (למשל, בעזרת שיקוע אימוני משותף וצביעות אימוניות), ו- *in vitro* (הוכחת אינטראקציה בין חלבונים רקומביננטיים, שהוכנו ע"י ביטוי בחיידקים ו/או ע"י סינתזה שלהם במבחנה). ראוי לציין כי בד"כ מקובל להוכיח קיום אינטראקציות בין חלבונים בעזרת שתי שיטות שונות לפחות.

חשוב לזכור, כי חלק מהאינטראקציות הינן דינאמיות, ומותנות בתנאים פיזיולוגיים ספציפיים או בתגובת התא לפקטורי גידול או מצבי עקה (stress). תנאים אלה עשויים לגרום למודיפיקציות של החלבון, שבעקבותיהן תחול או תיפסק האינטראקציה.

חשוב לוודא שאינטראקציה שהתגלתה היא אכן אמיתית ומתקיימת בתא, ואינה תוצאה של ארטיפקט שנגרם בעת הכנת תמצית התאים [25, 26]. כמו כן, יש לברר אם החלבונים אשר שוקעו יחד עם החלבון הנחקר, אכן קשורים ישירות לחלבון זה או לחלבונים אחרים הקשורים אליו ומשמשים כמתווכים [25].

מגבלתה של השיטה נעוצה בצורך לבודד את הקומפלקס החלבוני בשיטות פיזיקאליות. חלבונים מסוימים עשויים לא להתגלות, מפני שהאינטראקציה שלהם חלשה מכדי לעמוד בתהליך הבידוד של הקומפלקס [54]. כמו כן, קשירתו של החלבון לנוגדן עשויה לפגוע באינטראקציה ולמנוע את זיהויה [26].

Epitope Tagging: יש המשתמשים בחלבון הנחקר בצורתו הרקומביננטית. החלבון מבוטא בתאים, בד"כ ברמה גבוהה, ולרוב מוצמד אליו אפיטופ מלאכותי המוכר ע"י נוגדן מונוקלונאלי מסחרי. הנוגדן משמש לשיקוע החלבון ועימו חלבונים נוספים הקשורים אליו [26].

GST Pulldown: בשיטה זו מוצמד לחלבון הנחקר פפטיד אשר לו אפיניות גבוהה אל מולקולה קטנה, כגון גלוטטיון או מלטוז, אשר ניתן לקשרה למצע מוצק. למשל, קטע מתוך החלבון של גלותטיון-S-טרנספרז (GST), המוצמד לחלבון הנחקר, מקנה לחלבון אפיניות גבוהה מאוד

לגלותטיון. שיקועו של החלבון יעשה ע״י גרגירים שאליהם קשורות מולקולות גלוטטיון (glutathione affinity matrix) [26], או ע״י glutathione-agarose beads [25]. במקום לשקע את החלבון בעזרת נוגדן, מבוצעת עתה כרומטוגרפיה על עמודה הבנויה מן המצע הנ״ל, והחלבון הנחקר נקשר אליה, ועימו חלבונים נוספים הצמודים אליו.

2.3.2.1.1 *<u>זיהוי חלבונים בעזרת ספקטרומטר מסות</u>*

השיטה העיקרית לזיהוי חלבונים בלתי ידועים, הנספחים לחלבון הנחקר, מבוססת על שימוש בספקטרומטר מסות. לאחר הרצת באלקטרופורזה, מוצא מן הג׳ל פס החלבון, והוא מעוכל לפפטידים קצרים, בעזרת פרוטאז תלוי רצף, כדוגמת טריפסין. מסת הפפטידים הנוצרת נמדדת בספקטרומטר מסות לצורך זיהויים. הפפטידים קלים יותר לחילוץ מהג׳ל מאשר חלבונים שלמים, וקבוצה קטנה מתוכם מספיקה לזיהוי החלבון [55].

תערובות של פחות מ- 100 חלבונים, יופרדו לרוב על ג׳ל חד מימדי, בעוד שלתערובות חלבונים עשירות יותר, כגון תמצית תאים, תשמש אלקטרופורזה דו מימדית. שיטות יעילות לחיתוך החלבונים לפפטידים ולהעברתם מהג׳ל אל ספקטרומטר המסות, מאפשרות כיום לזהות חלבונים שכמותם בדגימה הינה בשיעור של ננוגרמים בודדים. שימוש ברובוטים למיכון התהליך, עשוי לאפשר את ניתוח החלבונים על פני ג׳ל שלם [56].

שתי גישות עיקריות משמשות לזיהוי החלבון מתוך אוסף הפפטידים המתקבלים ממנו. הראשונה נקראת MALDI (Matrix-Assisted Laser Desorption/Ionization), ובעזרתה מתקבל ספקטרום המסות של כל הפפטידים בתערובת, וזו מהווה ״טביעת אצבע״ של החלבון הנחקר. ספקטרום מסות הפפטידים משמש לסריקת מאגרי מידע בחיפוש אחר חלבון בעל ספקטרום מסות תואם (המחושב באופן תיאורטי מתוך הרצף) [56]. ככל שייוצגו יותר חלבונים בבסיסי הנתונים, כן תשתפר מידת ההצלחה בזיהוי [55]. הגישה השנייה משתמשת ב- tandem mass spectrometer, אשר מסוגל להתמקד בסוג אחד של פפטידים בתמיסה, וע״י חיתוכו לקטעים, להפיק מידע לגבי הרצף שלו. מידע זה מתקבל מתוך השוואת דגם החיתוך המתקבל עבור הפפטיד הנבדק אל דגם החיתוך התיאורטי של רצפים במאגר מידע. דגם חיתוך כזה הינו ייחודי לכל פפטיד [56]. שיטה זו מורכבת יותר מקודמתה, אך עושה שימוש במידע מפורט יותר לצורך זיהוי החלבון, ולכן היא מדויקת יותר [55]. צירוף שתי השיטות מאפשר כיום זיהוי וודאי של כל חלבון אנושי. ראוי לציין, כי ניתן בדרך זו לזהות בעת ובעונה אחת מספר חלבונים הנמצאים בתערובת.

כדי להקל את תהליך הזיהוי, ניתן לבצע עיכול של החלבונים בתמיסה, והפרדה חלקית ע״י כרומטוגרפיה של תערובת הפפטידים המתקבלת, לפני הזנתה לספקטרומטר המסות [56].

Phage Display 2.3.2.2

שיטה זו מבוססת על שימוש בבקטריופאג׳ים, אשר ״מציגים״ כלפי הסביבה חלבון או פפטיד שנבחר ע״י החוקר, בהיותו מאוחה עם אחד מחלבוני המעטפת שלהם [55]. הדבר מושג ע״י החדרת קטע דנ״א זר אל גן המקודד לאחד מחלבוני המעטפת של הפאג׳. החלבון המאוחה מופיע במספר עותקים על פני מעטפת הפאג׳, ובכך הופך לנגיש לנוגדנים ולחלבונים אחרים בסביבתו [25].

חלבוני המעטפת המשמשים לכך, הינם בעיקר החלבונים III ו- VIII בפאג׳ M13. החלבון VIII הינו החלבון העיקרי במעטפת הפאג׳, והוא מופיע בכ- 2500 עותקים,ואילו החלבון III הוא חלבון מינורי במעטפת הפאג׳, המופיע בחמישה עותקים בלבד. הצמדת פפטיד או חלבון שלם אל קצהו האמינו טרמינלי, אינה פוגעת בתפקודו [57].

ניתן לסרוק ספריה רבגונית וגדולה של פאג׳ים המציגים קטעי חלבון שונים, במטרה למצוא את הפאג׳ים המציגים קטע אשר מקיים ״affinity selection״ עם חלבון או עם נוגדן נתונים [57]. תהליך הסריקה מתבצע במספר מחזורי העשרה, בתהליך המכונה panning. הפאג׳ים נקשרים אל החלבון המגיב עימם, כשהוא עצמו מחובר למצע מוצק. עודף הפאג׳ים נשטף, והפאג׳ים שנספחו למצע עוברים ריבוי בחיידקים. לאחר מספר מחזורי העשרה, מפענחים את רצף הפאג׳ים, כדי לזהות את הפפטיד שהוחדר אליהם, וגרם לאינטראקציה [25].

שיטה זו משמשת בעיקר לסריקת אוכלוסיית פאג׳ים, המכילים מגוון של פפטידים שונים. ניסוי כזה מסייע, למשל, בזיהוי האפיטופ של נוגדן נתון, או משמש לזיהוי חלבונים המגיבים עם חלבון מסוים ובמיפוי אתרי האינטראקציה בין שני החלבונים [57].

ניתן לבצע בשיטה זו מחקרים רחבי היקף של אינטראקציות חלבונים, ע״י שימוש בפאג׳ים המציגים את כל תוצריהן של ספריות cDNA. ניתן לקבע אל משטח כל חלבון ״פיתיון״ רצוי, וללכוד את כל הפאג׳ים המציגים חלבונים הבאים עמו באינטראקציה [25, 55].

יתרונותיה העיקריים של שיטה זו טמונים ביכולתה לסרוק ספריות גדולות מאוד באופן פשוט, מהיר וזול. חסרונותיה כוללים את הקושי המתעורר בהצגת חלבונים גדולים על פני הפאג׳, וכן בעיות הנובעות מהעובדה, שהחלבונים המוצגים על פני הפאג׳, אינם מצויים במצבם ובסביבתם הטבעיים, ולפיכך עשויים לא להתקפל נכון בבקטריה, או לא לעבור מודיפיקציות-שלאחר-תרגום

הדרושות לקישור יעיל. בהיותם חלבונים מאוחים, הם עשויים לאבד מיכולתם לקשור חלבונים מסוימים. עוד ראוי לציין, כי ריכוזי החלבונים הנבדקים בניסוי אינם משקפים את אלו הקיימים בתא [25].

Two-Hybrid Screens 2.3.2.3

השיטה מבוססת על המבנה המודולארי המאפיין פקטורי שעתוק רבים, אשר מורכבים מחלק הקושר דנ״א, ומחלק אחר, האחראי לשפעול השעתוק. האזור קושר הדנ״א אחראי לספציפיות שלו לרצף מסוים, ובכך מכוון אותו אל הגנים המבוקרים על ידו, ואילו אתר השפעול קושר חלבונים נוספים של מנגנון השעתוק ומאפשר לשעתוק להתרחש. שני אתרים אלו אינם חייבים להיות קשורים זה לזה באופן קוולנטי, ודי בכך שיוצמדו זה לזה ע״י שני חלבונים אחרים, הנקשרים לזה לזה [25].

לפיכך, מפרידים את שני החלקים הפעילים של פקטור שעתוק, ויוצרים שני חלבונים היברידיים ע״י הצמדת כל אחד מהם לחלבון שונה. ה״פיתיון״ (bait) הוא החלבון המאוחה אל האתר קושר הדנ״א, ואילו ה״טרף״ (prey), הוא החלבון המאוחה אל האתר המשפעל את השעתוק. החלבונים ההיברידיים יבוטאו בתא שמר המכיל גנים מדווחים (אחד או יותר) בעלי פרומוטור מתאים. אינטראקציה פיזית בין חלבוני ה״פיתיון״ וה״טרף״, תגרום להתקרבותם של שני האתרים הפעילים זה לזה, וליצירת פקטור שעתוק היברידי פעיל, אשר פעילותו יכולה להיבחן ע״י גנים מדווחים וסלקציה מתאימה [2, 58, 59].

יתרונה של שיטה זו על פני שיטות אחרות טמון ברגישות הגבוהה שלה, וביכולתה לגלות אינטראקציות, אשר אינן נצפות בשיטות אחרות [25, 59]. כמו כן, האינטראקציות מתגלות בתוך סביבה תאית, ואין צורך בביצוע ניקויים ביוכימיים. השימוש בתאי שמר כ-host, מאפשר לבצע סלקציות רבות, לבדוק מגוון רחב של constructs, ואף לבחון את מידת האפיניות של האינטראקציה תוך שימוש במעכבים תחרותיים שונים [25].

לצורך יישומה של שיטה זו בקנה מידה גנומי, מוחדרים פלסמידים המקודדים לחלבוני ״פיתיון״ ו״טרף״, אל זני שמר הפלואידים ממינים הפוכים. לאחר הזיווג, יתקבלו צאצאים דיפלואידיים, המכילים את שני הפלסמידים. אם קיימת אינטראקציה בין החלבונים המקודדים, תתגלה פעילות של פקטור השעתוק, ואז ניתן לזהות את החלבונים המתאימים, ע״י פענוח הרצף בפלסמידים שהם מכילים [2, 55].

שיטה זו ניתנת ליישום בקנה מידה נרחב (high throughput) [60-62]:

Uetz et al. [60] השתמשו בשתי גישות שונות למחקר מקיף של אינטראקציות חלבונים בשמר. הראשונה נקראה "array screening", ובה יצרו 6000 שיבוטים, שכל אחד מהם מכיל פלסמיד "טרף" שונה, ואוספים אלה זווגו עם 192 "פיתיונות" שונים. על פי הגישה השנייה, " library screening", מוחזקים כל אלפי הזנים המבטאים חלבוני "טרף" שונים, יחדיו, כספרייה. כל אחד מן החלבונים מבוטא בנפרד כחלבון "פיתיון", ומזווג אל הספרייה. לאחר ביצוע סלקציה לקבלת התאים הדיפולואידיים, מזהים את אלו שבהם מתקיימת אינטראקציה ומזהים את החלבונים הרלוונטיים.

Ito et al. [62] השתמשו בגישה שלישית, ובה הם שיבטו כל אחת ממסגרות הקריאה הפתוחות בגנום השמר כחלבון "פיתיון" בשמרים מזן MATa, ובנפרד כחלבון "טרף" בשמרים מזן MATα שהנם מהמין הנגדי. זנים אלו הופרדו ל- 62 קבוצות של "פיתיונות" ול- 62 קבוצות של זני "טרף", כאשר בכל קבוצה מיוצגים 96 חלבונים היברידים. קבוצות ה"טרף" השונות זווגו עם קבוצות ה"פיתיון" השונות, ותוצרי הזיווג עברו סלקציה לפעילותם של גנים מדווחים מתאימים. החלבונים ההיברידים אשר הראו אינטראקציה ואפשרו לשמרים לשרוד את הסלקציה, זוהו ע"י פענוח הרצף של הפלסמידים המתאימים.

חסרונותיה של שיטת two-hybrid נובעים בעיקר מכך שהאינטראקציות נבחנות בסביבה ובהקשר שונים מהקיימים בתא האנימלי. השיטה מגלה אינטראקציות אפשריות, אך לא את ההקשר הביולוגי בו הן מתרחשות. חלקן עשויות להתרחש בתא, במצבים פיזיולוגיים מסוימים בלבד, ואילו אחרות עשויות לא להתרחש כלל בתא האנימלי, מפני שהחלבונים ממוקמים למעשה באורגנלות נפרדות בתא [63]. בנוסף לכך, השיטה מוגבלת לבדיקת חלבונים אשר יכולים להימצא בתוך הגרעין ואינם מוצאים אל מחוץ לגרעין באופן אקטיבי [26], ואשר האינטראקציה ביניהם יכולה להתרחש בגרעין ולא רק בציטופלזמה או בתוך ממברנת התא, למשל. זאת משום שאינטראקציה מחוץ לגרעין לא תביא לשעתוק [25, 58]. כמו כן, על החלבונים להתקפל באופן נכון ולהיות יציבים בתא השמר, ולשמור על פעילותם על אף היותם חלבונים מאוחים [2, 25]. אינטראקציות מסוימות עשויות לא להתגלות, משום שאתר האינטראקציה נחסם בעקבות איחוי החלבון [25]. השינויים-שלאחר-התרגום המתרחשים בשמר אינם נשלטים ועשויים להיות שונים מאשר בתא האנימלי (לדוגמא, פוספורילציה וגליקוזילציה). לכן, אינטראקציות שתלויות

בשינויים אלו עשויות לא להתגלות [25, 35, 58]. תוצאות false-negative עשויות להתקבל אפילו בגלל ייצור עודף של החלבון ההיברידי המאוחה לאתר קושר הדנ״א [35].

מגבלה נוספת נגרמת על ידי חלבונים אשר מתפקדים כ״אקטיבטורים״ כאשר הם מאוחים אל האתר קושר הדנ״א, כלומר, הם גורמים לשעתוק הגן המדווח גם ללא אינטראקציה עם החלבון ההיברידי האחר (״self-activator״). תופעה זו מתקיימת בחלבונים רבים, וביניהם גם כאלו אשר בד״כ אינם מעורבים בשעתוק [25], והם אחראים לרבות מתוצאות ה-״false-positive״ של המערכת [35]. לעיתים, ניתן לפתור את הבעיה ע״י הסרתו של קטע מסוים מהחלבון, ובכך לבטל את פעילות השפעול העצמית ללא פגיעה בתכונותיו האחרות של החלבון [25].

כדי להתגבר על חסרונות אלו, פותחו ואריאציות שונות על השיטה הבסיסית, אשר מתבוננות באינטראקציות אשר מתרחשות מחוץ לגרעין, ואשר אינן מבוססות דווקא על מנגנון שפעול השעתוק. שיטה אחת מבוססת על פיצולו של פפטיד ה-ubiquitin לשני קטעים, אשר יוצמדו אל החלבונים הנחקרים. במידה ותתרחש אינטראקציה בין החלבונים, יתקרבו שני הקטעים הנ״ל, ויתרחש חיתוך פרוטאוליטי של חלבון מדווח הצמוד אל הקצה הקרבוקסילי של פפטיד ה-ubiquitin. עקב חיתוך יתקבל חלבון מדווח פעיל, אשר ניתן לגילוי [59]. Stagliar *et al.* [64] השתמשו בשיטה זו, ובפקטור שעתוק בתור החלבון המדווח, כדי לזהות אינטראקציות בין חלבונים השייכים לממברנת ה- endoplasmic reticulum.

שיטה נוספת לגילוי אינטראקציות מחוץ לגרעין, משתמשת בתאי שמר בהם קיימת מוטציה בחלבון cdc25 שהוא פקטור שיחלוף מ-GDP ל-GTP של מולקולת Ras. מוטציה זו אינה מאפשרת לתאים לגדול בטמפ׳ של 37°C, אך החלבון האנושי hSos בהיותו מעוגן לממברנת התא, יכול לעשות קומפלמנטציה לחלבון השמרי הפגום, ולאפשר את גדילת התאים. במערכת זו, מוצמד חלבון ה-hSos אל חלבון ה״פיתיון״, ואילו חלבון ה״טרף״ מסומן ע״י סיגנל מיריסטילציה (myristylation signal), אשר גורם לעיגונו אל ממברנת התא. אם קיימת אינטראקציה בין חלבון ה״טרף״, המעוגן אל הממברנה, וחלבון ה״פיתיון״, ימוקם חלבון ה״פיתיון״ בסמוך אל הממברנה ועמו חלבון ה-hSos, ותתאפשר גדילת התאים בטמפרטורה של 37°C [59, 65]. גרסה שונה מעט למערכת זו, מבוססת על ההכרח ש-Ras יהיה מעוגן בממברנה לצורך תפקודו, ולכן מצמידה את Ras עצמו אל חלבון ה״פיתיון״ (לאחר הסרת הסיגנל אשר אחראי למיקומו הממברנלי), ומסמנת

את חלבון ה"טרף" בסיגנל מיריסטילציה (הגורם לעיגונו לממברנה). האינטראקציה בין החלבונים הנבדקים תביא את Ras אל הממברנה ותאפשר את תפקודו התקין [66].

2.3.3 שימוש בתבניות ביטוי לחיזוי שייכות פונקציונאלית

תבנית ביטוי (expression pattern) מייצגת את רמות הביטוי של הגנים השונים המשועתקים בתא (כלומר, כמויות הרנ"א שליח שלהם), והיא המרכיב העיקרי בפנוטיפ התא ובקביעת תפקודו. בניגוד לגנום, תבנית הביטוי היא מאוד דינאמית ומשתנה במהירות במהלך חיי התא, הן כתוצאה משינויים פיזיולוגיים שגרתיים (למשל, התקדמות מחזור התא), והן בתגובה לגירויים, או לעקה (stress). פרופיל הביטוי של גן מסוים (מיקום וכמות), חשוב להבנת פעילותו ותפקידו הביולוגי של החלבון אותו הוא מקודד [1].

חלבונים בד"כ אינם פועלים לבדם, ולרוב הם והחלבונים עמם הם מתפקדים מבוטאים בתא באותו זמן או מיקום. התבוננות בגנים המתבטאים במספר גדול של תאים שונים, הנבדלים בתנאי הגידול שלהם, ברקמה ממנה הם נלקחו, בהיותם בריאים או חולים, או במצבים פיזיולוגיים שונים, תגלה שונות רבה ברמות הביטוי, ואמורה לאפשר זיהוי של גנים בעלי דגם ביטוי משותף, אשר במקרים רבים הם גם קשורים פונקציונאלית [38]. על מנת לבצע הפרדה מוצלחת של הגנים לקבוצות, בהתאם לרמות הביטוי שלהם, יש צורך לבצע ניסויים היוצרים כמויות מידע גדולות ביותר [38].

קיימות שיטות שונות לבדיקת רמות חלבונים באופן ישיר, כגון זיהוי חלבונים מתוך תערובות בעזרת ספקטרומטר מסות, או הפרדת החלבונים על ג'ל דו-מימדי וזיהויים לאחר מכן בעזרת ספקטרומטר מסות. שיטות אלו הן איטיות ביותר, דורשות עבודה מרובה, וקשה להפיק מהן כמויות מידע מספקות לצורך יצירת קבוצות פונקציונאליות של החלבונים [38]. לכן, מועדפות שיטות הבודקות את רמת ה-mRNA של הגנים, שהן קלות יותר לביצוע, רגישות יותר, ונמצא כי הן אמינות במידה מספקת, שכן עבור רב הגנים קיימת התאמה בין שינויים ברמות ה-mRNA לשינויים ברמות החלבון [1].

תבניות הביטוי של קבוצות גנים עשויות להוות מפתח לזיהוי מנגנוני בקרה בתא, להבנת תפקידם התאי של החלבונים, ולקביעת השתייכותם למסלולים ביוכימיים [1]. בהנחה, שביטוי גנים הפועלים באותו תהליך הינו מתואם ומשתנה בכולם באורח דומה עם שינוי התנאים, השוואת תבנית הביטוי של גן לא מוכר לאלו של גנים מוכרים, תביא אותנו להשערות לגבי שיוכו לקבוצת גנים ידועה ולגבי תפקידו [7, 67].

זיהוי תבניות ביטוי נרחבות התאפשר בעזרת הטכנולוגיה של מערכי דנ״א (DNA arrays), בעזרתה ניתן להתבונן ברמות הביטוי של גנים רבים במקביל. העיקרון עליו מבוססת טכנולוגיה זו הוא ביצוע היברידיזציה בין מולקולות מסומנות של רנ״א או דנ״א המצויות בתמיסה, לבין מולקולות דנ״א המקושרות למשטח מוצק. התפתחותה של טכנולוגיה זו בשנים האחרונות הביאה לשיפור ביעילות הניסויים ובאיכות המידע המתקבל מהם, ומאפשרת כיום לייצר במהירות מערכי דנ״א ממוזערים וצפופים מאוד – DNA microarrays, המכונים גם ״DNA chips״. קיימים כיום מערכים כאלה, המייצגים עשרות אלפי גנים, ובאורגניזמים מסוימים בהם רצף הגנום פוענח במלואו, (למשל, שמר האפייה), הם מכילים את כל הגנים של אותו אורגניזם [1].

ניתוח השינויים החלים המקביל ברמות הביטוי של אלפי גנים, בתאים מסוגים שונים (למשל, תאים מרקמות שונות, תאים ממצבים פיזיולוגיים שונים, תאים סרטניים מול תאים רגילים), מאפשר השגת שתי מטרות: 1. קבלת ״טביעת אצבעות״ מולקולארית של מצב פיזיולוגי ספציפי או סוג תא מסוים. 2. זיהוי קבוצות גנים, שרמת ביטויים מתואמת, והם קשורים לתהליכים ספציפיים. בד״כ קיים קשר פונקציונאלי בין החלבונים המקודדים ע״י גנים אלו [6]. ריכוז התוצאות המתקבלות מסדרות הניסויים, מאפשר להפריד את הגנים לקבוצות (clustering) בהתאם לדמיון בתבניות הביטוי שלהם, ולחזות פונקציה משוערת לאותם חלבונים לא מוכרים, על סמך חלבונים ידועים השייכים לאותה קבוצה [1].

שימוש במערכי דנ״א המכילים עשרות אלפי גנים שונים מאפשר לגלות קשרים פונקציונאליים בין גנים שונים, ללא צורך להעריך מראש מיהם הגנים הרלוונטיים. לכן, שיטה זו מאפשרת לחוקרים למקד את מחקרם על קבוצת גנים המוגדרת ע״י תוצאות הניסוי, אשר יתכן ורבים מהגנים הכלולים בה, לא היו מועמדים למחקר ללא תוצאות אלו [1].

תקפותה של השיטה הודגמה עבור גנים רבים בשמר האפייה, *S. cerevisiae*, אשר הרצף הגנומי שלו פוענח במלואו, ותפקודם של 60% מהגנים שלו ידוע. גנים אשר ידוע היה כי הם פועלים יחדיו, נטו להשתייך לאותה קבוצה בחלוקה ע״פ תבניות הביטוי [1].

לשם פענוח תוצאות המתקבלות מניסויים אלה, נבחנו אלגוריתמים מתמטיים שונים, שהבולטים בהם מיועדים לביצוע תהליך ה-clustering. שימוש בכלים אלה מאפשר לזהות את קבוצות הגנים הפועלים יחדיו, לבחון את מובהקות התוצאות ולהקטין את הפרשנות הסובייקטיבית הניתנת להן [1].

2.3.4 שיטות של גנומיקה השוואתית לחיזוי שייכות פונקציונאלית

פענוח הרצף של גנומים רבים אפשר את פיתוחן של שיטות חישוביות שונות לחיזוי קשר פונקציונאלי בין חלבונים, המבוססות על התבוננות השוואתית בגנומים אלו[§]. שיטות אלו מתבססות על ההכרה, כי רצף הגנום כולל מידע אבולוציוני רב, המעיד על תפקידי גנים וחלבונים ועל היחסים התפקודיים ביניהם. עוצמתן של שיטות אלו טמונה בעובדה, כי הן יוצרות "רשתות" של חלבונים המקושרים פונקציונאלית, שבהם כלולים גם חלבונים אשר מעולם לא אופיינו. תפקידו של החלבון מוגדר ברמה התאית (המסלול התאי, או הקומפלקס החלבוני, אליו הוא שייך), ולא ברמת פעילותו הביוכימית [38]. לכן, גילויים עתידיים הנוגעים לאחד החלבונים ברשת כזו, יוכלו לשמש להסקת מסקנות ולחיזוי תפקוד גם עבור שכניו.

בדיקה של התחזיות שהתקבלו בשיטות אלו הראתה, כי השיטות מספקות מידע ברמת אמינות סבירה באופן כללי, ורמת אמינות טובה מאוד, כאשר הקשר הפונקציונאלי נחזה בעזרת שתיים מהשיטות או יותר [6]. ככל שגדל מספרם של הגנומים המפוענחים, הופכות השיטות למדויקות יותר ומאפשרות השערות מבוססות יותר.

2.3.4.1 פרופילים פילוגנטיים

שיטה זו מתבססת על ההנחה, שחלבונים אשר מתפקדים יחדיו באותו מסלול או כחלק מאותו קומפלקס, התפתחותם האבולוציונית עשויה להיות דומה. זאת, משום שפעילותם של מסלולים ביוכימיים וקומפלקסים חלבוניים עלולה להיפגם עקב אובדן של רכיב בודד בהם [38]. כלומר, במהלך האבולוציה נוטים כל החלבונים הללו להשתמר יחדיו באורגניזמים חדשים, או לא להופיע בהם כלל [68].

פרופיל פילוגנטי של חלבון מסוים הוא תבנית המתארת את קיומו או אי קיומו של חלבון בכל אחד מאוסף אורגניזמים אשר רצף הגנום שלהם פוענח. אם שני חלבונים חולקים אותו פרופיל פילוגנטי, מניחים כי הם קשורים פונקציונאלית, והופעתם או אי הופעתם באורגניזם קשורה לקיומו של אותו מסלול ביוכימי או קומפלקס חלבוני [6, 68].

ההצדקה הסטטיסטית שבבסיסה של שיטה זו נובעת מכך, שמספר הפרופילים הפילוגנטיים האפשריים הוא גדול משמעותית ביחס למספר משפחות החלבונים הקיימות. בהנחה שמתייחסים בתהליך אל 30 גנומים, הרי שישנם 2^{30} אפשרויות שונות של פרופילים פילוגנטיים (שכן בכל אחד

---

§ גנומיקה השוואתית = Comparative Genomics.

מ-30 הגנומים יכול החלבון להופיע או לא להופיע) [38]. מספר זה הוא בסדר גודל של 10^9, בעוד שכמות הגנים בגנומים השונים אינה עולה על 10^5. הסיכוי, שלשני חלבונים יהיה פרופיל פילוגנטי זהה, או דומה מאוד, באופן מקרי, הוא קטן מאוד, וסביר יותר להניח, כי חלבונים כאלו פועלים באותו מנגנון ביוכימי [6]. ככל שמספר הגנומים המפוענחים גדל, כך משתפרת יכולתה של השיטה להפריד את החלבונים לקבוצות פונקציונאליות באופן מדויק ופרטני [68].

בדרך כלל, לא קיים דמיון רב ברצף החומצות האמיניות של חלבונים אשר קשורים פונקציונאלית, ולכן קשרים אשר ניתנים לזיהוי ע״י השוואת פרופילים פילוגנטיים, אינם מתגלים על ידי השיטות המסורתיות המבוססות על השוואת רצפים [38, 68]. בהתאם לכל ההנחות האלה, Pellegrini *et al.* [68] הראו, כי חלבונים אשר להם פרופיל פילוגנטי זהה או דומה מאוד, נוטים ליטול חלק באותו תהליך תאי. כמו כן, הדמיון בין הפרופילים הפילוגנטיים של חלבונים, אשר קשורים פונקציונאלית זה לזה, רב יותר מהדמיון המתקבל עבור קבוצת חלבונים אקראיים. להערכתם של Pellegrini *et al.* [68], ניתן בעזרת שיטה זו לחזות פונקציה כללית נכונה עבור יותר ממחצית החלבונים של *E. coli*.

שימוש אפשרי נוסף בפרופילים פילוגנטיים הוא בחיפוש אחר חלבונים בעלי תפקיד זהה, שאינם הומולוגים. במקרה זה, נצפה לראות פרופילים פילוגנטיים אשר אינם דומים, אלא משלימים זה את זה, כלומר, שני החלבונים מופיעים לסירוגין באורגניזמים שונים. יתכנו גם אורגניזמים אשר בגנום שלהם יקודדו שני החלבונים גם יחד [19].

2.3.4.2 שיטת ״אבן הרוזטה״

נמצאו זוגות חלבונים אשר באורגניזם אחד הם מופיעים כשני חלבונים נפרדים, בעוד שבאחר הם מבוטאים כחלבון מאוחה אחד. חלבון כזה מכונה ״חלבון אבן הרוזטה״, שכן הוא מאפשר את פענוח הקשר הפונקציונאלי בין זוג החלבונים [69].

שני אזורים פונקציונאליים השייכים לחלבון אחד, משתתפים כמעט בוודאות באותה פונקציה, ובמקרים רבים אף באים במגע ישיר זה עם זה. לכן איתורו בגנומים שונים של חלבון יחיד, המורכב מהאיחוי של זוג חלבונים, מהווה הוכחה חזקה לקיומו של קשר פונקציונאלי בין שני חלבונים אלו [6, 38]. קשרים כאלו נפוצים גם בין חלבונים אאוקריוטים נפרדים, אשר מופיעים כחלבון פרוקריוטי יחיד [38].

לפיכך פותחה גישה המניחה, כי איחוי גנים ישרוד אבולוציונית, רק אם ישנו יתרון לצימוד זה. כלומר, הצימוד חל בין שני גנים, שתוצריהם פועלים יחדיו. לכן סביר להניח, כי תוצרי גנים שהינם

מאוחים בגנומים מסוימים, יוצרים אינטראקציה פיזית, או לפחות פונקציונאלית, באורגניזמים בהם הם נפרדים [19, 69].

חשוב להדגיש, כי האינטראקציות המתקבלות בשיטה זו מעידות על קשר פונקציונאלי בין החלבונים, אולם לא בהכרח על מגע פיזי ביניהם. יתכן לדוגמא, שזוג חלבונים עברו איחוי שיתרונו בבקרה על ביטויים המשותף, או בבקרת מעורבותם באיתות תאי [69].

כדי לשפר את אמינות התוצאות המתקבלות בגישה זו, משמיטים במהלך העיבוד אזורים פונקציונאליים "מתירניים", אשר נוטים להתחבר עם מגוון גדול של אזורים פונקציונאליים אחרים, כגון אתרים קושרי דנ"א או אתרים קושרי ATP [69]. כמו כן, חשוב לזהות בצורה נכונה את החלבונים הבודדים המרכיבים את החלבון המאוחה, ולוודא כי הם אורתולוגים (מקורם באותו חלבון קדמון). חיזוי הפונקציה מאבד מאמינותו, כאשר מדובר בחלבונים הומולוגים אחרים [19].

היות ובד"כ אין דמיון רב בין הרצפים של שני החלבונים שעברו איחוי, הרי שקשר פונקציונאלי שנתגלה בשיטת "אבן הרוזטה" קשה לזיהוי ע"י השיטות המסורתיות לחיפוש חלבונים הומולוגים [6].

2.3.4.3 גנים שכנים

קיימת טענה, כי אם הגנים המקודדים שני חלבונים הם שכנים קרובים בכרומוזום במספר גנומים שונים, החלבונים המקודדים על ידם נוטים להיות קשורים פונקציונאלית [6, 38]. מקורה של הנחה זו הוא באנליזה של אופרונים באורגניזמים פרוקריוטים. היות ותוצריהם של גנים המופיעים באופרונים קשורים פונקציונאלית זה לזה [38], קיומו של אופרון בגנום כלשהו, משמעו, כי הגנים הכלולים באופרון, עשויים להוות קבוצה פונקציונאלית גם בגנומים אחרים, בהם הם אינם בהכרח מרוכזים יחדיו באופרון [19, 70].

החיפוש אחר אופרונים לא ידועים הנו תהליך בעייתי, אשר עד היום לא הוגדר באופן מובנה. לכן, שיטה זו מתבוננת בגנים שכנים, כלומר, גנים אשר שומרים על מיקומם היחסי זה לזה במספר אורגניזמים שונים. על הגנים המקבילים באורגניזמים השונים להיות אורתולוגיים, כלומר, מקורם אמור להיות בחלבון קדמון משותף (ולא בשכפול או בהעברה אופקית אשר אינם מעידים על שימור אבולוציוני של מיקום הגנים) [19]. רבים מהגנים הסמוכים זה לזה אינם בעלי תפקוד משותף, אך ניתוח מעמיק של קבוצות גנים אלו עשוי לסייע בחיזויים של קשרים פונקציונאליים חדשים [19].

Overbeek *et al.* [70, 71] חיפשו זוגות של גנים אורתולוגיים, אשר נמצאים בסמיכות זה לזה בגנומים שונים, ודירגו אותם בהתאם למרחק האבולוציוני בין האורגניזמים המתאימים. ככל שהמרחק האבולוציוני בין האורגניזמים גדול יותר, כך קטנה הסבירות, שהגנים שמרו על מיקומם היחסי במקרה, ולכן ניתן להעריך, כי אם סמיכות המקום נשמרת, קיים קשר פונקציונאלי ביניהם.

איכות התחזיות לגבי קשרים פונקציונאליים המתקבלות בשיטה זו היא מצויינת, אך הן נדירות, עקב הדרישה המורכבת למציאת גנים המקיימים שני תנאים: היותם אורתולוגים, ושמירת מיקומם היחסי זה ביחס לזה [38]. Dandekar *et al.* [72] הראו בעבודתם, כי תוצרי 95% מזוגות הגנים שזוהו על ידם בשיטה זו, מתפקדים יחדיו בוודאות או בסבירות גבוהה מאוד.

מידת התאמתה של שיטה זו למחקר גנים אאוקריוטים עדיין אינה ברורה, עקב העדר אופרונים בגנום האאוקריוטי, אם כי קיימות דוגמאות לגנים אאוקריוטים שכנים, אשר קיים ביניהם קשר פונקציונאלי [38].

2.4 מאגרי מידע של אינטראקציות חלבונים

קיימים כיום מאגרי מידע רבים המכילים מידע על אינטראקציות חלבונים. מאגרי מידע אלו נבדלים זה מזה בהיבטים רבים, החל מהיותם פומביים או מסחריים - דבר המשפיע על מידת נגישותם לציבור החוקרים, וכלה בסוג המידע שהם מכילים, הדרך בה הוא התגלה, ומידת הפיקוח על אמינותו [39].

מאגרי מידע מסוימים מכילים אינטראקציות במינים שונים (למשל BIND [42] ו-DIP [43]), בעוד שאחרים מוקדשים לריכוז המידע הנוגע למין מסוים (כדוגמת CYGD [73], FlyNets [74], ו-SPiD [75]). אחדים מאפשרים לכל חוקר להוסיף אליהם מידע במטרה לאפשר את הרחבתם המהירה ועדכונם השוטף (כגון MINT [76]), בעוד שאחרים מעדיפים לשמור על רמת אמינות גבוהה ומאפשרים הזנת מידע חדש רק לאנשי מקצוע שהוסמכו לכך (YPD [77]). חלק ממאגרי המידע מכילים אינטראקציות פחות מהימנות שנתגלו בניסויים רחבי היקף, בעוד שאחרים מקפידים בדבקות על הזנת אינטראקציות אשר אומתו בשיטות מסורתיות ופורסמו בספרות המקצועית בלבד.

להלן מופיע תיאור של כמה ממאגרי המידע של אינטראקציות, אשר קיימים כיום. חשוב לציין, כי מאגרי המידע משתנים מאוד לאורך השנים, כאשר חלקם נסגרים או שאינם מעודכנים באופן

שוטף והופכים להיות מיושנים, חלקם משנים את סוג המידע שהם מכילים ואת צורת ארגונו, ואחרים מתמזגים זה עם זה ליצירת מאגרי מידע חדשים, רחבים ומעודכנים יותר.

2.4.1 YPD - Yeast Proteome Database

מאגר מידע זה [77, 78] פותח על ידי חברת Proteome, ועבר להיות בבעלות חברת Incyte לפני מספר חודשים. מאגר המידע הינו מסחרי, והגישה אליו מותנית בתשלום (גם למוסדות אקדמיים).

המאגר סוקר היבטים ביולוגיים רבים של חלבוני השמר, וביניהם תפקידם התאי, מיקומם, חלבונים דומים במינים אחרים, מודיפיקציות שעוברים החלבונים, בקרת שעתוק הגנים המתאימים, והאינטראקציות בינם לבין חלבונים אחרים בתא. המידע המופיע ב-YPD נאסף מן הספרות המקצועית, ע"י צוות שהוסמך לכך ותחת בקרה מדוקדקת. התיאור של כל אינטראקציה במאגר כולל את השיטות בהן היא אופיינה והפניות למאמרים התומכים בה.

מאגרי המידע אשר כלולים ב- Proteome BioKnowledge Library של חברת Incyte, מכילים מעל 640,000 אנוטציות שמקורן ביותר מ-50,000 מאמרים, ופרט למאגר המידע של השמר *S. cerevisiae*, הם כוללים גם מאגרי מידע עבור *C. elegans*, *S. pombe*, *H. sapiens*, ועוד.

2.4.2 MIPS-CYGD - Comprehensive Yeast Genome Database

MIPS - The Munich Information Center for Protein Sequences [73, 79], מְרָכֵּז מידע גנומי מגוון, הכולל מאגרי מידע גנומי לאורגניזמים שונים, אנוטציות של רצפי חלבונים וכלים לניתוחם, ועוד. המידע מיועד לשימוש אקדמי, לא-מסחרי בלבד.

CYGD הינו מאגר המידע המוקדש לשמר *S. cerevisiae*. המאגר מכיל מידע גנטי, ביוכימי ותאי לגבי רכיבי הגנום השונים - ORFs, גנים של RNA, ומקטעי DNA אחרים. מקורו של רוב המידע המוצג הוא בעיבוד ידני של הספרות המקצועית, בתוצאות של ניתוחים פונקציונאליים רחבי היקף (כדוגמת EUROFAN [80]), וכן בקישורים אל מאגרי מידע נוספים. כמו כן, משולב במאגר זה מידע שמקורו ב-13 מיני שמר אחרים, לצורך ניתוח השוואתי.

המאגר מכיל מידע על אינטראקציות חלבונים וקומפלקסים, המלווה באנוטציות מתאימות. במאגר ישנן כיום כ-8250 אינטראקציות פיזיות ועוד כ-1500 אינטראקציות גנטיות, הכוללות את

תוצאותיהם של ניסויים רחבי היקף לחקירת אינטראקציות פיזיות, גנטיות וקומפלקסים חלבוניים (בשיטות הכוללות Y2H, deletion mutants, mass-spectrometry, ועוד).

2.4.3 BIND - Biomolecular Interaction Network Database

מאגר מידע זה [42, 81] תוכנן כך שיוכל להכיל אינטראקציות בין מגוון רכיבים מולקולאריים שונים כדוגמת חלבון-חלבון, חלבון-רנ״א, חלבון-דנ״א, וחלבון-מולקולה-קטנה.

BIND כולל מידע על אינטראקציות, קומפלקסים חלבוניים, ומסלולים תאיים, אשר התגלה בתוצאות ניסויים. תיאור האינטראקציות מאפשר הזנת מידע מגוון, החל ממיקומה התאי של האינטראקציה, ואתרי הקישור המעורבים בה, ועד לקינטיקה ולתרמודינאמיקה של הריאקציה. כמו כן נשמרים תנאי הניסוי בהם נצפתה כל אינטראקציה. המידע המוזן למאגר אינו מוגבל למין מסוים, ויכול אף להכיל אינטראקציות בין חלבונים במינים שונים (למשל בין וירוס לפונדקאי).

המידע המוזן למאגר לקוח הן מתוך הספרות המקצועית, והן מתוך מקורות מידע כגון מאגרי מידע אחרים ותוצאות ניסויים רחבי-היקף. המידע הספרותי מוזן למאגר ע״י אנשי הצוות של BIND וכן ע״י חוקרים חיצוניים אשר יכולים לשלוח ל-BIND אינטראקציות שפורסמו (ואשר יוזנו למאגר לאחר בדיקתן ע״י מומחים). יבוא של מידע רחב היקף ממקורות אחרים נעשה באופן אוטומטי ע״י כלי תוכנה מתאימים.

מאגר המידע מכיל כיום כ-6150 אינטראקציות, כ-850 קומפלקסים וכ-8 מסלולים. המידע פומבי, נגיש לכל וניתן להורדה מאתר האינטרנט של BIND, וכמוהו גם קוד המקור של התוכנה, בעזרתו ניתן להתקין שרת מקומי של בסיס הנתונים.

2.4.4 DIP - Database of Interacting Proteins

מאגר מידע זה [82, 83] מכיל זוגות חלבונים אשר הוכח באופן ניסויי כי קיימת ביניהם אינטראקציה פיזית, כלומר, שני רצפים בחלבונים אלו נצמדים זה לזה. המידע כולל אינטראקציות אשר פורסמו בספרות המקצועית והוזנו מתוך מאמרים מתאימים, וכן את תוצאותיהם של מספר ניסויי Y2H רחבי היקף.

עבור כל חלבון המופיע באינטראקציה במאגר, נשמר מידע מקומי בסיסי הכולל את שמו, תפקידו, ומיקומו התאי, וכן קישור לפחות לאחד ממאגרי המידע הגדולים לחלבונים (PIR [84], SWISSPROT [85], GENBANK [86]), ולמאגרים נוספים המכילים מידע רלוונטי. המידע הנשמר עבור האינטראקציות המופיעות במאגר כולל את האזור בחלבונים אשר מעורב

באינטראקציה, קבוע הדיסוציאציה, השיטות הניסוייות בעזרתן זוהתה ואופיינה אינטראקציה זו, והפניות למאמרים המתארים אותה.

המאגר מעודכן דרך רשת האינטרנט, הן ע״י אנשי המקצוע של DIP והן ע״י מתנדבים, ובתנאי שביכולתם לספק סימוכין לניסוי בו זוהתה האינטראקציה בין זוג החלבונים. בכדי להאיץ את קצב הזנת המידע אשר מקורו במאמרים, משתמש הצוות של DIP במספר טכניקות לאיתור המאמרים המתאימים, וביניהן חיפושים אוטומטיים על פני MEDLINE בשיטות של כריית מידע (data mining), ואף ניסיונות להרחיב אינטראקציות המוכרות באורגניזם אחד, אל החלבונים ההומולוגים באורגניזם אחר.

מאגר המידע מכיל כיום כ-18000 אינטראקציות בין 6800 חלבונים השייכים ל-110 אורגניזמים שונים, מהם 4700 חלבונים ו-15000 אינטראקציות מתייחסים לשמר *S. cerevisiae*. השימוש במידע הינו חופשי למטרות שאינן מסחריות, והוא ניתן אף להורדה מרשת האינטרנט ע״י משתמשים רשומים (registered).

2.4.5 MINT - Molecular INTeractions database

מטרתו של MINT [76, 87] היא לאחסן אינטראקציות פונקציונאליות בין מולקולות ביולוגיות הכוללות חלבונים, רנ״א, ודנ״א. פרט להיצמדותן של שתי מולקולות זו לזו, מאפשר מאגר המידע לתאר שינויים אנזימטיים באחת המולקולות (פוספורילציה, אצטילציה, יוביקוויטינציה וכיו״ב), ואף אינטראקציות פונקציונאליות לא-ישירות ואינטראקציות גנטיות. במידה וקיים פירוט לגבי הקינטיקה של האינטראקציה, קבועי היצמדות והאתרים המשתתפים באינטראקציה, אף הוא מוזן אל המאגר.

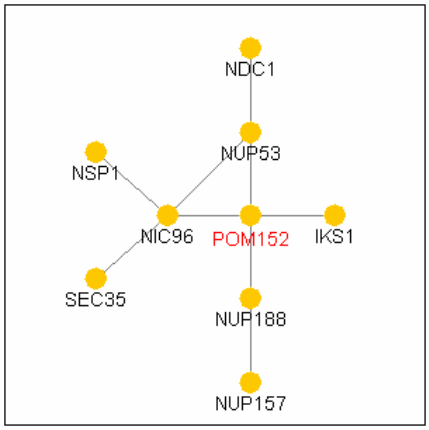
המידע במאגר מתרכז כיום באינטראקציות חלבונים אשר אושרו באופן ניסויי, ואשר הוזנו אל המאגר מתוך הספרות המקצועית ע״י צוות מומחי MINT. להזנת המידע נעזר צוות בסיס הנתונים בכלי תוכנה בשם MINT Assistant אשר מאתר מאמרים המכילים מידע על אינטראקציות ומציג אותם למשתמש במבנה נוח לניתוח. רוב האינטראקציות במאגר המידע מקושרות אל המאמר המתאים דרך PUBMED, אך ניתן להזין אליו גם אינטראקציות אשר לא פורסמו. כמו כן, כולל המאגר אינטראקציות אשר זוהו בניסויי Y2H רחבי היקף, ומידע על אינטראקציות פיזיות וגנטיות בשמר מתוך מאגר המידע MIPS (ראה 2.4.2).

כיום מכיל המאגר מעל 3780 אינטראקציות ישירות (היצמדות מולקולות) ועוד כ-780 אינטראקציות לא-ישירות וגנטיות, בין 3556 חלבונים ב-64 אורגניזמים שונים.

2.4.6 PathCalling Yeast Interaction Database

חברת Curagen [88] פיתחה מערכת תעשייתית בשם PathCalling, המאפשרת ללקוחותיה לבצע סדרת ניסויי Y2H רחבי היקף ברמת אמינות גבוהה, ולהשתמש בכלי ויזואליזציה חזקים בכדי לנתח את התוצאות. כלי הויזואליזציה המוצעים כוללים כלי דו-מימדי המציג את האינטראקציות ומאפשר למשתמש לבנות מהן מסלולים, וכלי נוסף תלת-מימדי, אשר מטרתו לאפשר את השוואת האינטראקציות המוכרות בין אורגניזמים שונים. לכלים אלו ניתנת גישה ללקוחות החברה בלבד.

בעזרת מערכת זו נערך אחד מניסויי ה-Y2H רחבי ההיקף בשמר *S. cerevisiae* [60], ותוצאותיו בליווי אינטראקציות נוספות בשמר שמקורן בספרות המקצועית ובמאגרי מידע נוספים, נאספו למאגר מידע הנגיש לכלל המשתמשים דרך אתר האינטרנט של החברה [89]. דפי המידע מכילים עבור כל חלבון קישורים אל מאגרי מידע אחרים (כגון GenBank ו-SwissProt), את רשימת החלבונים המקיימים עימו אינטראקציה, וכן תמונה סטטית של מפת האינטראקציות במרחק 2 מהחלבון (ראה איור 1). הקשה על אחד החלבונים השכנים בגרף, מעבירה את המשתמש אל דף המידע של חלבון זה (ב-PathCalling), המכיל את מפת האינטראקציות המתאימה לחלבון החדש.



איור 1. תמונת אינטראקציות מתוך Yeast PathCalling

גרף סטטי המופיע בדף המידע של POM152, ומציג את שכניו עד למרחק 2.

2.4.7 Hybrigenics' PIMs - Protein Interaction Maps

חברת Hybrigenics מציעה באתר שלה [90], שלושה מאגרי מידע של אינטראקציות:

- PIMRider HIV literature, המכיל אינטראקציות בין וירוס ה-HIV לבין תאי לימפוציטים מאדם, אשר מקורן בפרסומים בספרות המקצועית.

- PIMRider *H. pylori*, המכיל אינטראקציות בין חלבונים בחיידק זה, אשר נתגלו בניסוי Y2H רחב היקף בעזרת טכנולוגיית ה-PIM [91].
- PIMRider HCV, המכיל אינטראקציות בין חלבוני הוירוס Hepatitis C, אשר נמצאו בניסויי Y2H שבוצעו בחברה בטכנולוגיית ה-PIM.

הגישה לשני מאגרי המידע הראשונים היא חופשית למשתמשים אקדמיים. המידע שהם מציעים פרט לאינטראקציות כולל: אנוטציה פונקציונאלית, רשימת מאמרים רלוונטיים, מידע אודות הרצף ומיקומו בגנום, וקישורים אל מאגרי מידע אחרים של חלבונים (SwissProt).

טכנולוגיית ה-PIM בה משתמשת חברת Hybrigenics כוללת סריקה בשיטת two-hybrid של ספרייה גדולה מאוד של מקטעים אקראיים מתוך הרצף הגנומי, ע"י חלבוני פיתיון. היות ובספרייה ישנם מקטעים רבים בעלי חפיפה, ניתן מתוך התבוננות באינטראקציות המשותפות למקטעים אלו, לאפיין את האזור בחלבון המשתתף באינטראקציה, וכן לתת הערכה למידת המהימנות שלה. מקטעים אשר להם אינטראקציות רבות במיוחד מסומנים, משום שהם פחות אינפורמטיביים, ואף יתכן כי זוהי עדות לתוצאות false-positive.

לאחר קבלת תוצאות הניסוי, הן נבחנות בעזרת אוסף כלים המאפשר התבוננות טקסטואלית וגראפית בחלבונים השונים ובאינטראקציות ביניהם (ראה סעיף 2.5.5).

2.5 כלי ויזואליזציה

בשנים האחרונות התגברה מאוד המגמה של בניית מפות אינטראקציות רחבות היקף לאורגניזמים שונים, ויושמו שיטות ניסוייות לקבלת תוצאות בהיקף גנומי [92-94]. כמויות המידע הרבות אשר נאגרות במאגרי המידע מצריכות כלי ויזואליזציה במטרה להקל על המשתמש בהבנת המידע ובעיבודו. מאגרי מידע רבים מספקים כיום כלי ויזואליזציה המאפשרים התבוננות נוחה במפת האינטראקציות המתוארת על ידם. קיימים גם כלי ויזואליזציה אשר אינם מקושרים למאגר מידע מסוים, ואשר מאפשרים למשתמש להזין מפת אינטראקציות משלו ולהציגה באופן גרפי.

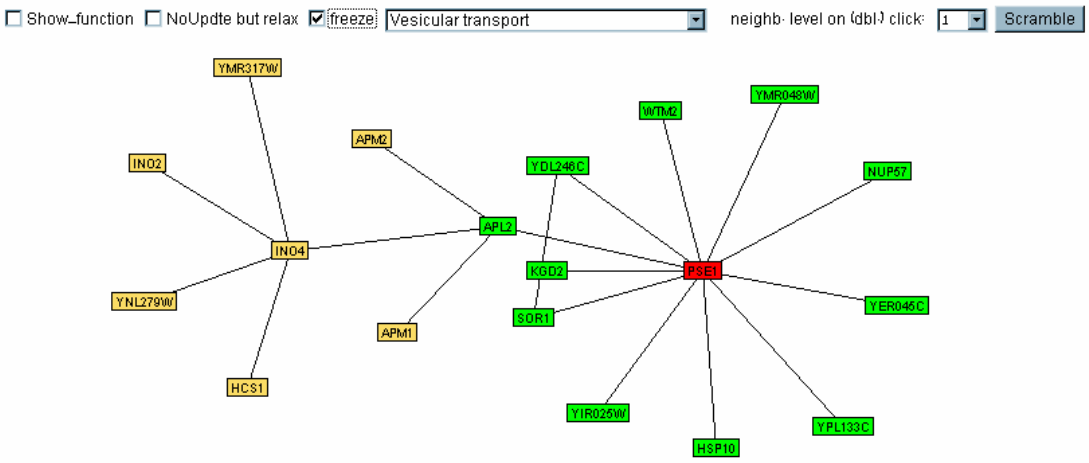
חשוב לציין, כי בזמן תחילת הפיתוח של PIVOT, לא היו קיימים כלים דומים. המגמה של פיתוח כלים גראפיים להצגת מפות האינטראקציות החלה רק לפני כשנתיים. PIVOT כולל מספר תכונות, אשר אינן קיימות בכלים אחרים המוכרים כיום בשוק, וביניהן היכולת למצוא את מסלול

האינטראקציות הקצר ביותר המקשר חלבונים מרוחקים בגרף. להלן סקירה של כמה מכלי הויזואליזציה הנמצאים בשימוש כיום.

2.5.1 Mrowka - Java Applet

כלי זה [95, 96] נכתב כ-Java applet, במטרה לאפשר את הצגתן של אינטראקציות בין חלבונים באופן גרפי. הוא מאפשר למשתמש להציג את החלבונים על פי תפקידם התאי, ולפרוש את שכניו של חלבון בגרף, כאשר עימוד החלבונים על פני הגרף נעשה באופן אוטומטי (ראה איור 2).

הכלי משולב בתוך דף HTML ורץ בתוך דפדפן האינטרנט. הוא אינו מקושר אל מאגר מידע מסוים, אלא המידע הכולל את רשימת החלבונים, תפקידם התאי והאינטראקציות ביניהם, מסופק כחלק מהקוד של דף האינטרנט. על מנת לשנות את המידע ולעדכנו, יש לשנות את הקוד של דף האינטרנט. אין אפשרות בכלי זה לשמור את הגרף עליו עובדים לצורך טעינתו בעתיד.



איור 2. תמונת מסך מתוך ה-Applet של Mrowka

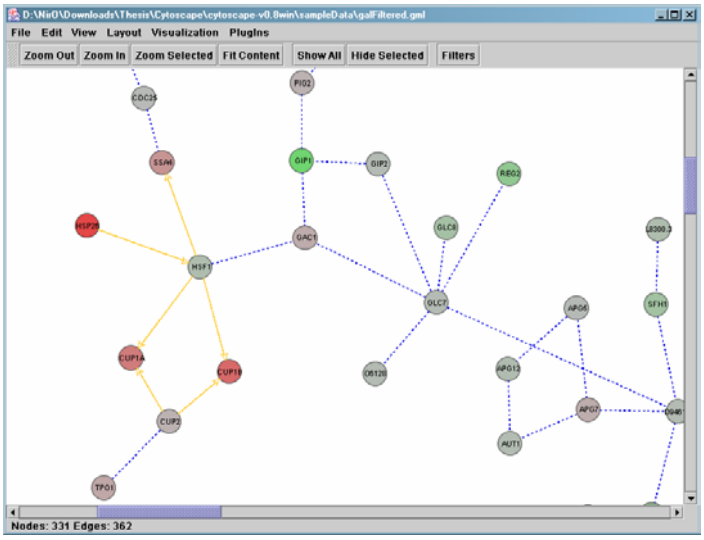
2.5.2 Cytoscape

התוכנה [97] מאפשרת לטעון קובץ אינטראקציות, המכיל רשימה של זוגות חלבונים, להציגם באופן גרפי כאוסף רשתות של אינטראקציות, ולשלב בהן מידע על רמות ביטוי גנים (ראה איור 3). הגנים מקושרים אל אנוטציות פונקציונאליות הלקוחות מתוך מאגר המידע Gene - GO Ontology [98].

לצורך עימוד הגרף, מוצעים לבחירה מספר אלגוריתמים שונים. העימוד לא נעשה באופן דינאמי, אלא עם ההקשה על כפתור מתאים, כאשר המשתמש יכול לבחור האם לעמד את הגרף כולו או רק אזור מסוים ממנו (ראה איור 4).

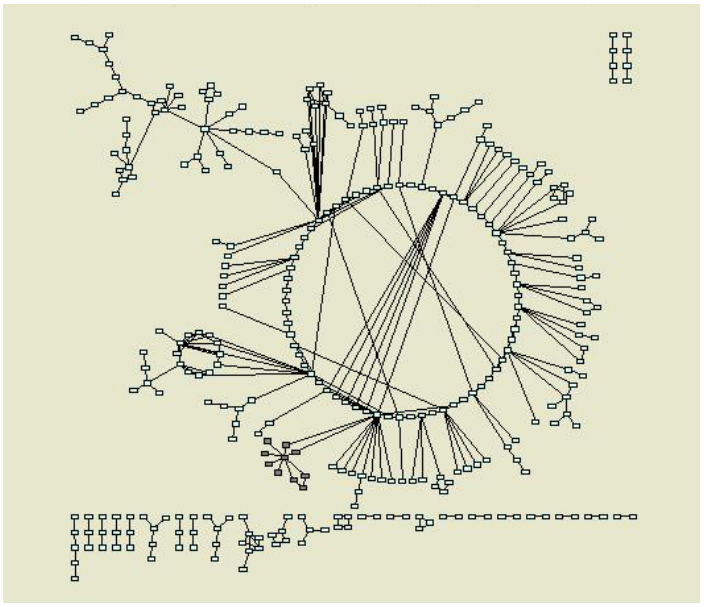
התוכנה מאפשרת למשתמש לשמור את הגרף שנוצר בפורמט GML - Graph Markup Language, שהוא פורמט כללי לתיאור גרפים, הנתמך ע"י כלי תוכנה נוספים.

כמו כן, משולב בכלי אלגוריתם אשר מטרתו לזהות מסלולים ברשת האינטראקציות האחראים לשינויים שנצפו ברמות הביטוי, ע"י חיפוש תת-גרף של גנים אשר מראים במשותף שינוי משמעותי בביטוי על פני מספר תנאים שונים.



איור 3. תמונת מסך מתוך Cytoscape

השינויים ברמות הביטוי מוצגים על פני רשת האינטראקציות.

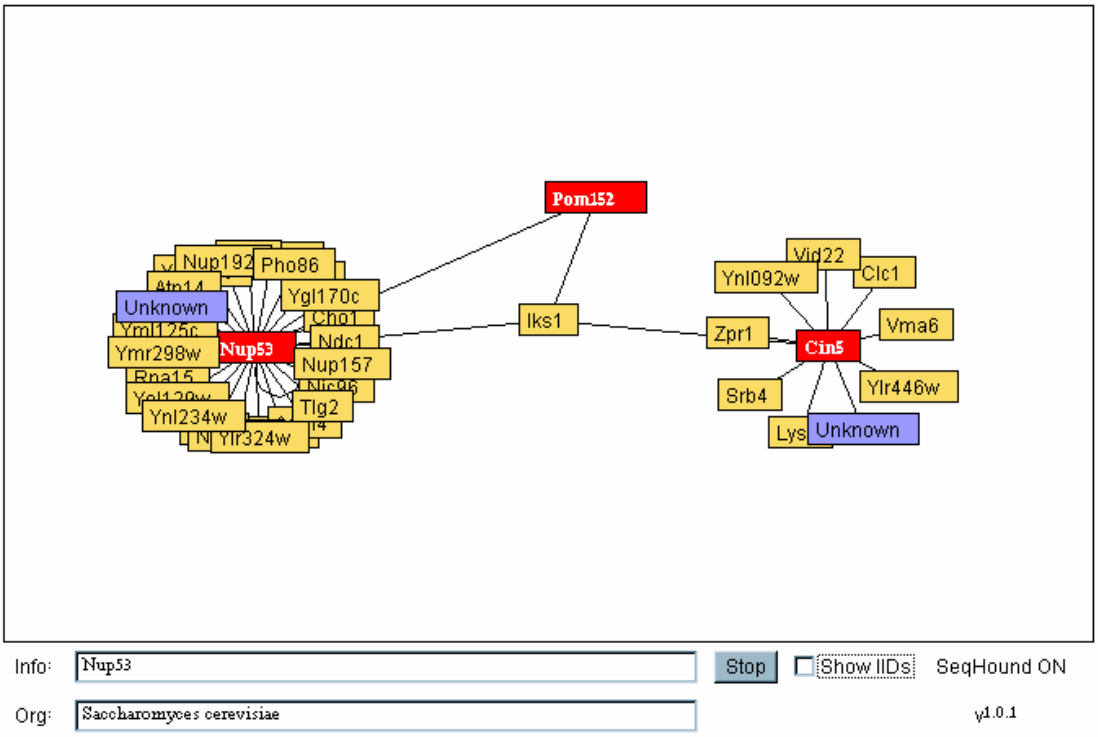


איור 4. תמונת מסך מתוך Cytoscape

עימוד הגרף ע"י אלגוריתם ל- cyclic layout.

2.5.3 BIND Viewer

מאגר המידע BIND מציע למשתמשיו כלי ויזואליזציה [81], המאפשר התבוננות באינטראקציות המופיעות במאגר. הכלי מקושר אל המאגר, וניתן להפעילו ע"י הקשה על קישור המופיע בדפי המידע המתארים אינטראקציות. עם הפעלתו הוא מציג את האינטראקציה המתוארת בדף המידע (ראה איור 5).



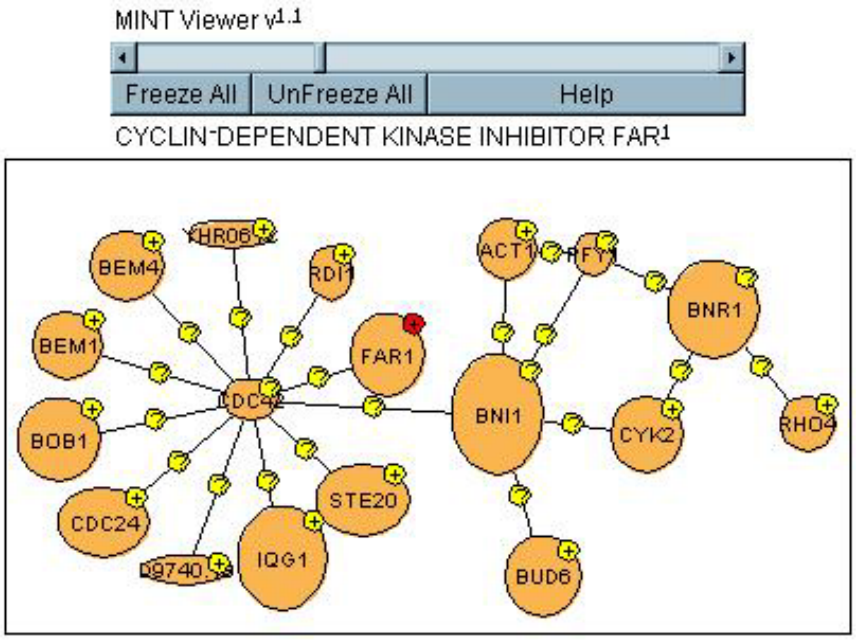
איור 5. תמונת מסך מתוך BIND Viewer

הקודקודים הצבועים באדום מקובעים במקומם, והאחרים מוזזים באופן דינאמי.

הכלי מבצע עימוד דינאמי של הקודקודים בגרף, ומאפשר למשתמש לקבע קודקוד במקום נתון. כמו כן, יכול המשתמש להקיש על קודקוד ובכך להוסיף אל הגרף את שכניו של אותו קודקוד. הקשה על קודקודי הגרף או על הקשתות שבו, יביאו לפתיחת דפי המידע המתאימים מתוך המאגר בדפדפן האינטרנט.

2.5.4 MINT Viewer

מאגר המידע MINT מספק אף הוא כלי ויזואליזציה [76, 87] להצגה גראפית של רשת האינטראקציות, שאותו ניתן להפעיל מתוך דפי המידע של המאגר (ראה איור 6). ע"י הקשה על קודקודים או על קשתות הגרף הוא מאפשר מעבר אל דפי המידע המתאימים. מיקום הקודקודים הוא דינאמי, והקשה על קודקוד גורמת להוספת שכניו אל הגרף. המשתמש יכול להגדיל או להקטין את המרחק בין הקודקודים, ע"י הזזת המחוון המופיע מעל לגרף.



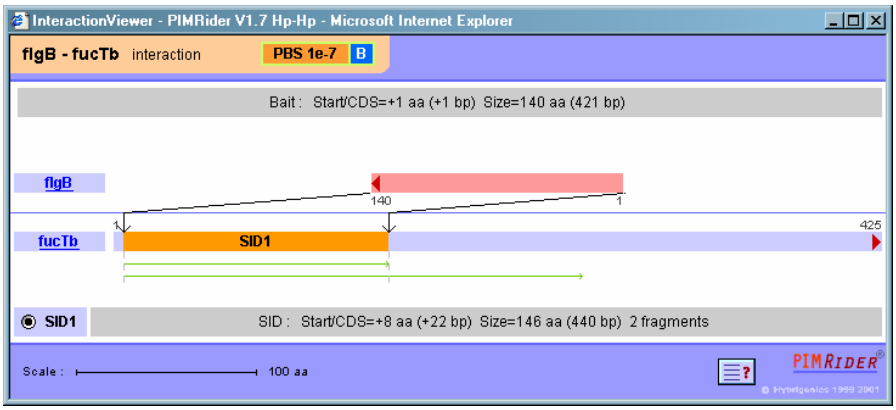
איור 6. תמונת מסך מתוך MINT Viewer

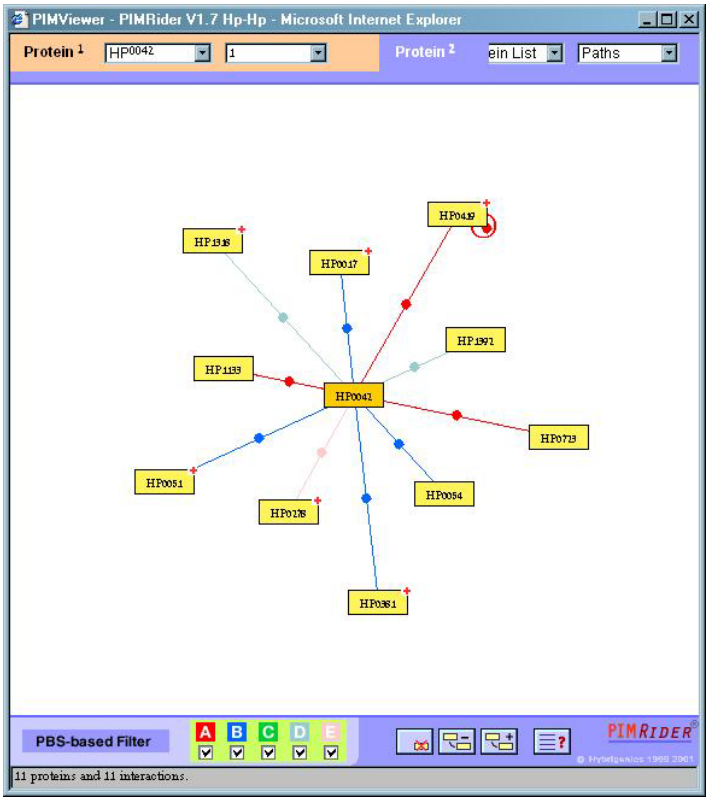
2.5.5 Hybrigenics' PIMRider

טכנולוגיית ה-PIM [90] כוללת בתוכה מספר כלים לתצוגה ויזואלית, המקושרים זה לזה (ראה איור 7). הקשות העכבר של המשתמש בכל אחד מהכלים, יגרמו להצגת המידע המתאים בכלים האחרים. הכלי המציג את רשת אינטראקציות החלבונים הוא ה-PIMViewer. להלן תיאור הכלים:

- ProteinViewer - הוא הרכיב המציג את דפי המידע לגבי החלבונים, הכוללים מידע פונקציונאלי לגבי החלבון, פרטים על האינטראקציות שלו, על ריצפו, ועוד.
- PIMViewer - אחראי על הצגת מפת האינטראקציות של החלבון. עימוד המפה מבוצע באופן דינאמי, והמשתמש יכול להוסיף חלבונים שכנים אל המפה, ולעבור אל דפי המידע של החלבונים עליהם הם מקיש. כמו כן, יכול המשתמש להחליט אילו אינטראקציות הוא מעוניין לראות, בהתאם לציון המהימנות שניתן להם.
- InteractionViewer - מאפשר להתבונן בחלבון פיתיון מסוים, ולהציב מולו את כל המקטעים הגנומיים אשר יצרו עמו אינטראקציה. הרצף המשותף לכל מקטעים אלו מכיל את האתר הדרוש לצורך האינטראקציה (ומכונה SID - Selected Interacting Domain).

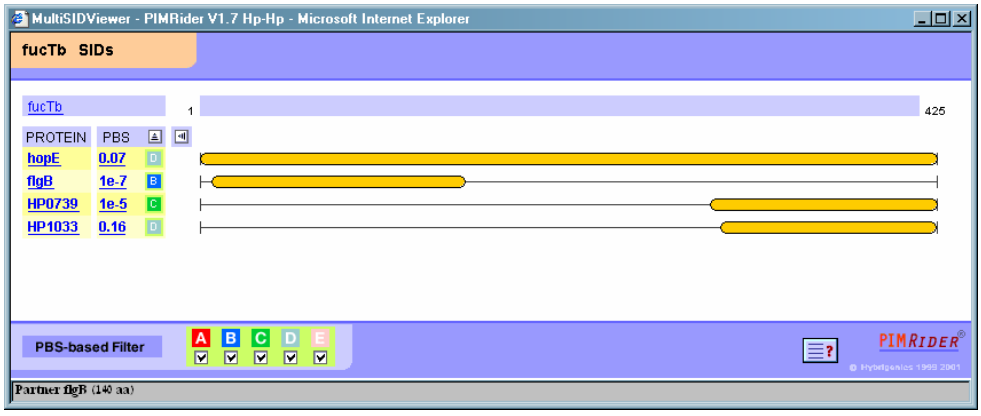
- MultiSIDViewer - כלי זה מאפשר להתבונן במקביל בכל ה-SIDs אשר מקורם באינטראקציות של חלבון מסוים עם כל שכניו. כך ניתן לראות בבירור את האתרים השונים בחלבון המשתתפים באינטראקציות.





איור 7. תמונות מסך מתוך PIMRider

למעלה - InteractionViewer, המאפשר לזהות את האתר המשתתף באינטראקציה.
משמאל - PIMViewer, המציג את מפת האינטראקציות ומאפשר פרישת שכנים נוספים.
למטה - MultiSIDViewer בעזרתו ניתן לראות את האתרים השונים בחלבון המשתתפים באינטראקציות.
האותיות A-E בתחתית שני המסכים האחרונים, מאפשרים לברור את רמת המהימנות של האינטראקציות המוצגות.



3 מטרות העבודה

המחקר הגנומי, אשר בעבר הלא רחוק הסתמך בעיקר על מחקרים נקודתיים המתרכזים בגן בודד, או בפנוטיפ חריג מסוים ומוגדר, עובר לעסוק כיום במחקרים בקנה מידה רחב יותר [1]. מחקרים רבים מתבוננים ברצף המלא של הגנום, ומנסים לגלות בו תופעות לא מוכרות, או להסביר תופעות מוכרות שאינן מובנות לגמרי.

שיטות המחקר בהן משתמשים משתנות אף הן, והופכות להיות מקיפות יותר ופחות ממוקדות. בעוד שבעבר חיפושי האינטראקציות התרכזו בחלבון מסוים אותו חקרו, והעבודה נעשתה פעמים רבות בשיטות ביוכימיות, כיום מנסים להתבונן באינטראקציות רבות, המתרחשות בעת ובעונה אחת בין חלבונים רבים, ומנסים למצוא שיטות חיזוי רחבות היקף לשם כך.

אחד המאפיינים של השיטות החדשות היא יכולתן לבחון כמויות גדולות של גנים וחלבונים באותה עת. לכן, כמויות המידע המתקבלות מכל ניסוי שנערך הן עצומות, ונדרשים גם כלים חדשים על מנת לעבדן. דוגמאות לכך ניתן לראות בניסויים השונים המשתמשים במערכים ממוזערים, בבניית הכלים החישוביים הדרושים לקריאת התוצאות של ניסויים אלו, ובפיתוחי האלגוריתמים השונים הדרושים לצורך עיבוד התוצאות והבנתן, כדוגמת האלגוריתמים השונים ל-clustering.

תפקידה של הביואינפורמטיקה הוא לסייע לחוקרים בהתמודדות עם כמויות המידע העצומות. היא עושה זאת ברמות שונות, החל ממציאת פתרונות לאחסון המידע, עיבודו של המידע, ועד הצגתו לחוקר. מטרתה לסייע לחוקר בהפיכת המידע הגולמי הרב למידע ביולוגי ממוקד ובעל ערך. אחד האתגרים הכרוכים בכך, הוא בניית כלים נוחים ופשוטים, אשר יאפשרו לחוקרים גישה מהירה, נוחה ובהירה למידע, באופן שיוכלו להפיק מהם את התועלת המרבית [99].

היות והניסויים השונים מתבוננים באלפי גנים בעת ובעונה אחת, נאלצים החוקרים לעסוק בתהליכים, ובמנגנונים תאיים אשר הידע לגביהם מועט. לכן מתעורר הצורך במערכות מתוחכמות, אשר יוכלו לרכז, לארגן ולהציג באופן ברור את כל המידע הביולוגי הקיים [1], ואשר יסייעו לחוקר בהתמודדות עם בעיות טכניות, כגון ריבוי הכינויים שיש לכל גן במאגרי המידע השונים, הפורמט השונה של המידע המוצג באתרי הרשת השונים, וההבדלים הקיימים במינוחים המדעיים בהם משתמשים מקורות מידע שונים [100].

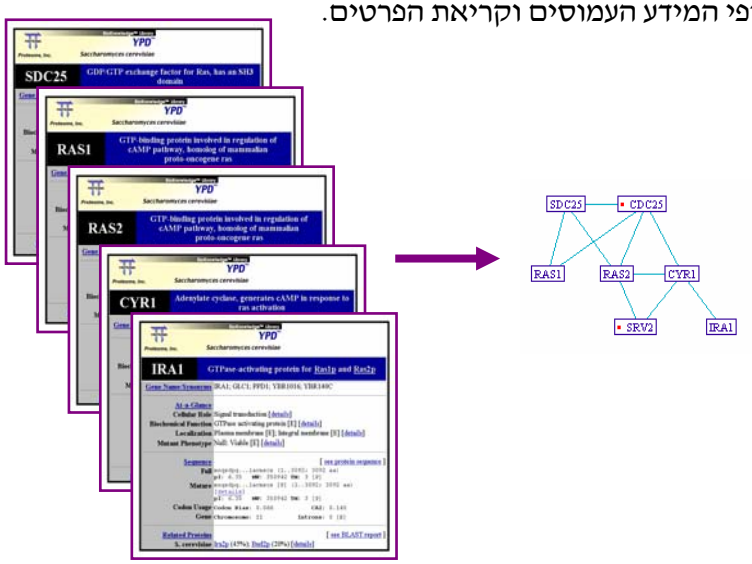
אחד האתגרים העומדים בפני הביואינפורמטיקאים, היא לעזור לחוקר בפענוח מידע בלי לפגוע בכושר השיפוט של החוקר, ובלי לכוונו למסקנות שאינן נובעות באופן ישיר מתוצאות הניסויים.

ברור, כי כלים חישוביים לא יוכלו להחליף את המוח האנושי במתן פרשנות ביולוגית לתוצאות חדשות, אך הכלים המתאימים חשובים כדי לאפשר לו את הגישה הנוחה ביותר אל המידע [1]. החוקר צריך גם להבין את מגבלות הכלים החישוביים, ולא להיקלע למלכוד הפסיכולוגי, ולהניח ש"המחשב תמיד צודק".

קבוצות רבות ברחבי העולם עוסקות בחיפוש אחר האינטראקציות המתרחשות בין החלבונים השונים בתא ובפיתוח שיטות חדשות לגילוי אינטראקציות אלו. קבוצות נוספות עמלות על פיתוח בסיסי נתונים, אשר יכילו את המידע הרב באופן מסודר וזמין לחוקרים. עיקר המידע הזמין בבסיסי נתונים כיום מאורגן בצורת טבלאות טקסטואליות, ולכן הוא מאוד לא מוחשי ולא נוח לעיבוד.

מטרתה של עבודה זו היא לבנות מערכת תוכנה, אשר תסייע לחוקר לבחון באופן גרפי מידע קיים על אינטראקציות בין חלבונים, ולבצע שאילתות ביולוגיות על מידע זה. על התוכנה יהיה להציג את המידע באופן מוחשי יותר, ולאפשר התבוננות נוחה במידע רב יותר ובאינטראקציות מרוחקות. כמו כן, אנו רוצים שניתן יהיה לבצע בעזרתה שאילתות שונות, אשר לא מעשי לבצען באופן ידני. דוגמא לכך הוא חיפוש אחר אינטראקציות בין חלבונים, אשר הקשר ביניהם אינו מוכר, על אף שנראה כי הוא קיים מתוך הדמיון בהתנהגותם הנצפית בניסוי.

התצוגה הגראפית שתציע התוכנה תתרכז במידע לגבי אינטראקציות בין חלבונים, ותאפשר למשתמש לחקור את רשת האינטראקציות הסובבת את החלבון בו הוא מתעניין. התוכנה תציג גם מידע מיידי נוסף לגבי החלבונים המוצגים (כגון ההומולוג האנושי של כל חלבון שמרי), שמטרתו להקל על החוקר בהבנת הגרף ובהתמצאות בו (ראה איור 8). קישור התצוגה הגרפית לדפי המידע המלאים של החלבונים באינטרנט, יאפשר לחוקר לקבל את המידע המלא בעת הצורך. בכך תאפשר התוכנה לחוקר להתרכז באינטראקציות, להתבונן ברבות מהן בעת ובעונה אחת, ולבצע שאילתות לגביהן, לפני פתיחת דפי המידע העמוסים וקריאת הפרטים.



איור 8. הצגת האינטראקציות כמפה גרפית

כדי ליצור באופן ידני את הגרף המתואר באיור זה, על המשתמש לעיין בחמשת דפי המידע המוצגים, ולשלב את מידע האינטראקציות המופיע בהם.

4 תהליך הפיתוח

4.1 בחירת כלי הפיתוח

תוכנה זו פותחה בשפת Java. השיקולים המרכזיים בבחירת Java כשפת הפיתוח היו:

- התכנות ב- Java הוא תכנות מונחה עצמים. היות והפרויקט בו מדובר הוא בעל נפח גדול, ומכיל חלקים עצמאיים רבים המשולבים זה בזה, חשובה ההפרדה המסודרת הנשמרת בגישה מונחית-העצמים (Object Oriented) בין חלקיה השונים של התוכנה.
- Java הנה Platform Independent. פיתוח התוכנה עבור פלטפורמה מסוימת (כנראה Win32 אשר היא העיקרית בשוק), היה מונע מציבור החוקרים העובדים עם פלטפורמות שונות מלהשתמש בה. היות ורבים מן החוקרים עבורם מיועדת התוכנה עובדים בעיקר בסביבות Unix או Mac, כלליות השפה היוותה גורם חשוב.
- Java פותחה במטרה לאפשר עבודה נוחה ברשת. לכן, התמיכה בעבודה מעל רשת מוטמעת בשפה, דבר ההופך את הכתיבה של יישומים כאלו לקלה יותר. כמו כן, לבעיות שונות המתעוררות בעבודה מעל רשת, ישנם פתרונות מובנים בשפה. דוגמא לכך היא היכולת לבצע serialization של אובייקט לצורך שליחתו כרצף של ביטים, וביצוע deserialization בצד המקבל, לקליטת אובייקטים. דוגמא נוספת היא התמודדות השפה עם סכנה של וירוסים ע"י זיהוי קוד שעבר שינוי לאחר צאתו מהקומפיילר התקני. היות וכלי זה מיועד לעבודה מול מאגר מידע מרכזי, הרי שהתמיכה בעבודה מעל רשת חשובה לפיתוחו.
- תמיכה גראפית המתאימה למערכות חלונאיות למיניהן אף היא מעוגנת בשפה. לכן בעבודה עם Java, אין צורך להשתמש בספריות גראפיות נוספות, ישנו פתרון לבעיות אי התאימות בין מערכות חלונאיות שונות (Windows מול Mac), והספריות הקיימות ניתנות להרחבה בקלות. כל אלו חשובים לצורך פיתוח הכלי, אשר הוא בייסודו כלי לוויזואליזציה גראפית.
- תמיכתה של Java בעבודה מול מבני נתונים נעשית דרך ממשק כללי הנקרא JDBC – Java DataBase Connectivity. ממשק זה מאפשר גישה אחידה לבסיסי נתונים שונים, ע"י

החלפת ה- driver בו משתמשים, וחוסך את הצורך בשינויים והתאמות בתוכנה. כמובן שככל שיידרשו פחות שינויים להתאמת הכלי למאגרי המידע השונים, כן ייטב.

4.2 הטיפול בגרף, Swing ו- MVC

חלקה הראשון של העבודה עסק בבניית מערכת לטיפול בגרפים. התוכנה צריכה לבנות גרף, המתאר את החלבונים בבסיס הנתונים כקודקודים בגרף ואת האינטראקציות ביניהם כקשתות. המערכת צריכה ליצור את מבנה הנתונים, אשר יכיל את המידע בגרף, עליה להציג את הגרף למשתמש, וכן לאפשר למשתמש לבצע מניפולציות שונות בו.

שפת Java מכילה כיום שתי ספריות סטנדרטיות עיקריות המטפלות בתצוגה גראפית.

הספרייה הגראפית הסטנדרטית הבסיסית נקראת AWT – Abstract Windowing Toolkit, וליוותה את השפה מראשית דרכה. ספריה זו משתמשת בשירותים של מערכת ההפעלה על מנת לצייר את האובייקטים הגראפיים השונים (חלון, כפתור, תפריט וכו'). בעובדה זו טמונים יתרונותיה של הספרייה, אך גם חסרונותיה. השימוש בשירותים של מערכת ההפעלה מאפשר לספריה לבצע את פעולותיה באופן מהיר ואמין. אך יכולותיה של הספרייה והאובייקטים הגראפיים שהיא מכילה, מוגבלים ע"י האפשרויות המוצעות ע"י מערכת ההפעלה. אם נוסיף לכך את העובדה ששפת Java נכתבה להיות platform independent, הרי שהספרייה הנ"ל מוגבלת רק לאותם כלים המופיעים בכל מערכות ההפעלה השונות. מגבלה נוספת שנובעת משיטת פעולה כזו היא כי אותו אובייקט גרפי (חלון או כפתור) יראה אחרת במערכת Win32 ובמערכת Mac.

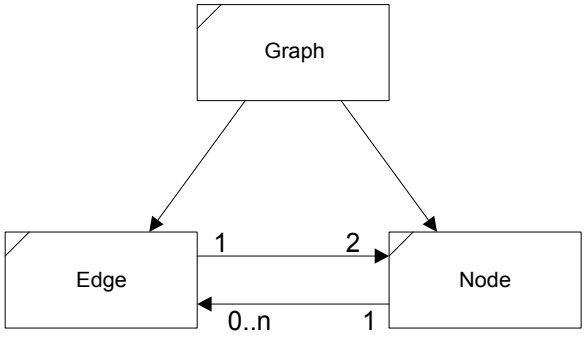
הספרייה הגראפית השנייה נקראת Swing. ספרייה זו שייכת ל-standard extensions, והיא ספריה גראפית, אשר נכתבה בכדי להתגבר על המגבלות של ה-AWT. בספרייה זו נעשה שימוש מינימאלי ביותר באובייקטים גראפיים של מערכת ההפעלה, ועיקר העבודה נעשה ע"י הספרייה עצמה, אשר דואגת לצייר את האובייקטים השונים ולהגיב לפעולות המשתמש. בדרך זו עוקפת הספרייה את הבעיות הנובעות מהצורך להתאים למגוון מערכות הפעלה שונות, ומאפשרת ליצור אובייקטים גראפיים מגוונים ולהציגם באופן אחיד על פני כל המחשבים. חסרונותיה טמונים בהיותה ספרייה חדשה יחסית. לכן עדיין ישנם בה באגים, נעשים לעיתים שינויים בממשקים הדורשים התאמות מצד המשתמשים, וישנם אף חלקים אשר עדיין לא מומשו. כמו כן, מהירות פעולתה איטית יותר מזו המוצעת ע"י אובייקטים של מערכת ההפעלה, אך לא באופן משמעותי מאוד.

בעבודה זו נעשה שימוש בספרייה השנייה – Swing, היות והתעורר הצורך באובייקטים גראפיים המוצעים על ידה בלבד, וכן עקב המלצתה של חברת Sun Microsystems, מפתחת Java, להשתמש בספרייה זו לפיתוח יישומים גראפיים.

לצורך פיתוח החבילה לטיפול בגרף, הוחלט להשתמש בתבנית (pattern) המקובלת של MVC - Model, View, Controller. השימוש בתבנית זו מאפשר הפרדה מסודרת בין מבני הנתונים (model) לבין התצוגה (view) ולבין הטיפול בפעולות המשתמש (controller). כמו כן, ניתן ליצור מספר views שונים המציגים את אותו model וע"י כך לאפשר למשתמש מספר תצוגות שונות של הגרף, תוך שמירה על סנכרון מלא ביניהן ומבלי שהן צריכות להיות מודעות זו לזו. השימוש בתבנית זו נפוץ מאוד לבניית רכיבים גראפיים, וגם ספריית Swing עושה בו שימוש ניכר.

4.2.1 <u>בניית המודל</u>

תפקידו של המודל הוא ארגון ושמירת מבנה הנתונים של הגרף. לשם כך נבנו הקלאסים: Node, Edge, ו- Graph (ראה איור 9). קלאסים אלו שומרים את כל המידע הנוגע למבנה הגרף (topology), וכן גם מידע הנוגע לעימוד הגרף (layout), ומידע לגבי בחירת קודקודים ע"י המשתמש (selection).



איור 9. הקלאסים המרכיבים את מודל הגרף

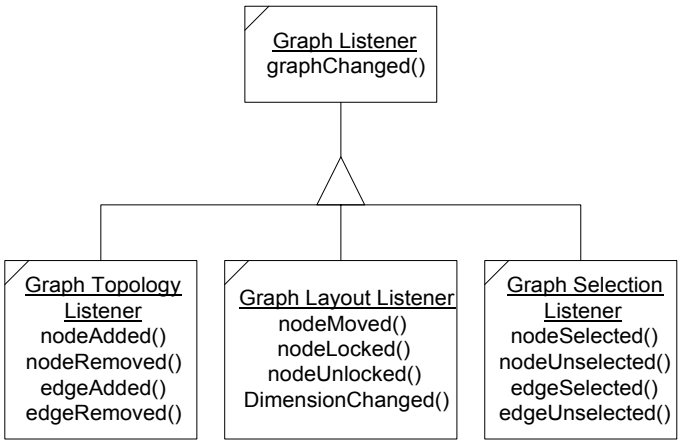
הקלאס Graph מכיל רשימת קודקודים וקשתות המרכיבים אותו. כל קשת מוגדרת ע"י שני הקודקודים, אותם היא מחברת, וכל קודקוד מכיר את הקשתות המקושרות אליו (אפס או יותר).

התוכניתן יכול להצמיד לכל קודקוד בגרף מידע נוסף שישמר עבורו (userdata), ולעשות במידע זה שימוש כרצונו, למשל לצורך ציורו של הקודקוד על המסך באופן מתוחכם (ראה להלן). יש להדגיש כי הגרף תוכנן כרכיב כללי, ללא התייחסות לייעודו הספציפי כגרף של חלבונים. בהמשך ייעשה שימוש ב-userdata, כדי להצמיד לכל קודקוד את החלבון השמרי, אותו הוא מייצג.

רכיבי המודל יכולים לבצע serialization לצורך שמירתם בקובץ, ו-deserialization לשם קריאתם בחזרה. יכולות אלו משמשות את המערכת גם לצורך שכפול הגרף, הדרוש בכדי לאפשר את הדפסתו ברקע.

היות ורכיבי המודל מכילים גם מידע הנוגע לעימוד הגרף, הקלאס Graph מספק פעולות המאפשרות לשנות את עימוד הגרף, לשימושם של מנגנון העימוד האוטומטי (המתואר בהמשך) ושל המשתמש, כגון הזזת הקודקודים או נעילתם במקומם. קלאס זה דואג למנוע התנגשויות בין מנגנון העימוד האוטומטי לבין פעולות המשתמש, וכן מנהל בקרה על גבולות הגרף ודואג למנוע את "בריחת" הגרף מהראשית.

כאשר מתרחשים שינויים במודל, הם מדווחים לכל הרכיבים הרלוונטיים. דבר זה נעשה ע"י מערכת של רישום (registration), בעזרתה כל אחד מהרכיבים המעוניינים במידע על שינויים, נרשמים בקלאס Graph כמאזינים (listeners). בכל פעם שמתרחש שינוי בגרף, נשלחת הודעה המתארת את השינוי לכל המאזינים.



איור 10. הממשקים המאפשרים להאזין לשינויים בגרף.

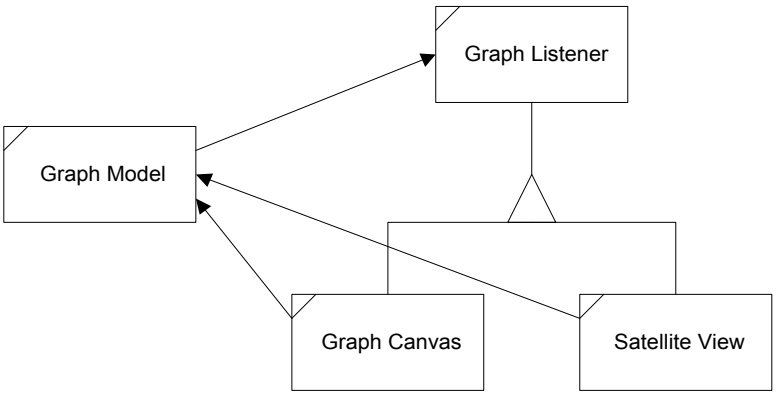
לשם הפחתת עומס ההודעות, והקלה בתכנות של אותם מאזינים, חולקו השינויים למספר תחומים (ראה איור 10) – שינויים במבנה הגרף (topology), שינויים בעימוד הגרף (layout), ושינויים בבחירת הקודקודים (selection). כל מאזין יכול להשתייך לאחד או יותר מהסוגים הבאים – GraphTopologyListener, GraphLayoutListener, GraphSelectionListener, ולקבל את ההודעות הרלוונטיות לגביו.

ההודעות לגבי התרחשויות בגרף מתחלקות לשלושה תחומים. כל אובייקט בתוכנה יכול לממש אחד או יותר מהממשקים, ולהירשם באובייקט Graph כמאזין על מנת לקבל הודעה על כל שינוי רלוונטי.

4.2.2 בניית ה- View

תפקידן של התצוגות (views) השונות של הגרף, הוא לצייר את המידע השמור במודל על המסך. התצוגות יודעות להאזין להודעות המודל (כלומר הן listeners של המודל), ומיד עם יצירתן הן

נרשמות במודל כמאזינות לשינויים (ראה איור 11). בכל פעם שישנו עדכון של המודל, מקבלות התצוגות הודעה מתאימה על כך, ודואגות לעדכן את המידע המוצג בהן. כך, ביצוע של עדכון הנעשה דרך אחת התצוגות, יתבצע במודל עצמו, ויתבטא בכל התצוגות השונות שלו.

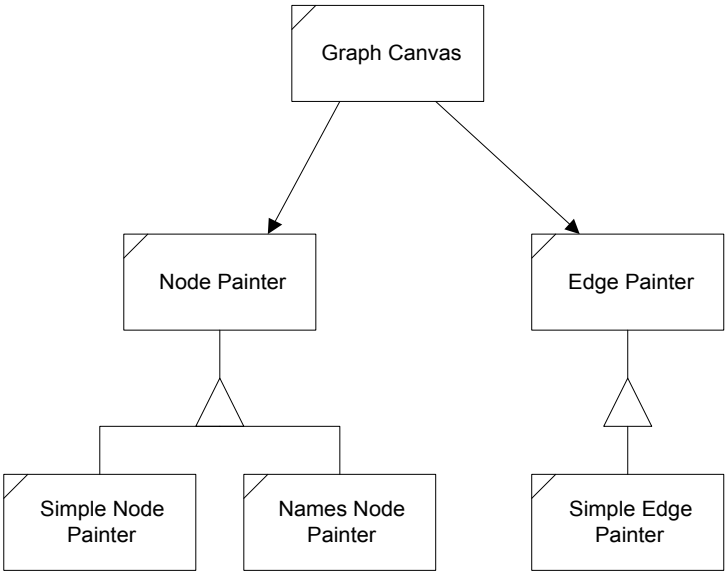


בעבודה זו יצרתי שתי תצוגות שונות לגרף. האחת היא תצוגת הלווין (SatelliteView), אשר מציגה תמיד את הגרף כולו. בכל פעם שהשטח שתופס הגרף גדל, קטן קנה המידה של תצוגה זו בהתאם, כך שתמיד תתקבל תמונה "ממעוף הציפור" של הגרף כולו. בתצוגה זו יכול המשתמש להצביע על אחד הקודקודים (המסומנים כנקודות), ולקבל את שם הקודקוד (דוגמא של תצוגת הלווין מובאת בעמוד 68).

התצוגה השנייה נקראת GraphCanvas, והיא התצוגה המורכבת יותר, אשר דרכה יבצע המשתמש את עיקר עבודתו. בתצוגה זו מוצג אזור חלקי של הגרף ביתר פירוט, וניתן לטייל אל אזורי הגרף הסמוכים בעזרת פסי הגלילה.

הקודקודים והקשתות בתצוגה זו מצוירים ע"י "ציירים" מיוחדים אשר ניתנים להחלפה (NodePainter ו- EdgePainter, ראה איור 12). התוכניתן יכול לספק "ציירים" מתוחכמים אשר יציירו את הקודקודים והקשתות בצורות ובצבעים שונים, יוסיפו להם תוויות טקסט ויציגו כל מידע רצוי לגבי רכיבים אלו. הצייר המתוחכם יוכל להשתמש ב- userdata השמור בקודקוד, אותו הוא מצייר, ואף לבצע "שאילתות חכמות" לרכיבים אחרים במערכת. צייר קודקודים

מתוחכם יכול, למשל, להציג את שמות ההומולוגים האנושיים במקום את שם החלבון השמרי המשויך לקודקוד.



ה- GraphCanvas פונה אל צייר הקודקודים גם בכדי לבקש תווית הסבר (tooltip) על קודקוד, לאחר שהסמן מצביע עליו מספר שניות ברציפות. הצייר יכול להימנע מהחזרת תווית, או להשתמש בתכונה זו להצגת מידע נוסף לגבי הקודקוד.

איור 12. ציירי הקודקודים והקשתות

ה- GraphCanvas נעזר באובייקטים מסוג NodePainter ו-EdgePainter על מנת לצייר את הקודקודים והקשתות, אותם הוא מציג. הקלאס SimpleNodePainter הוא סוג של צייר קודקודים, המציג כל קודקוד כריבוע פשוט, ואילו SimpleEdgePainter הינו צייר קשתות, המצייר את הקשתות כקווים ישרים. החלפת צייר הקודקודים בקלאס NamesNodePainter תגרום לכך, שהקודקודים יצוירו כמלבנים, אשר בתוכם מופיע שם הקודקוד.

4.2.3 <u>בניית ה- Controller ומנגנון ה- Node Events</u>

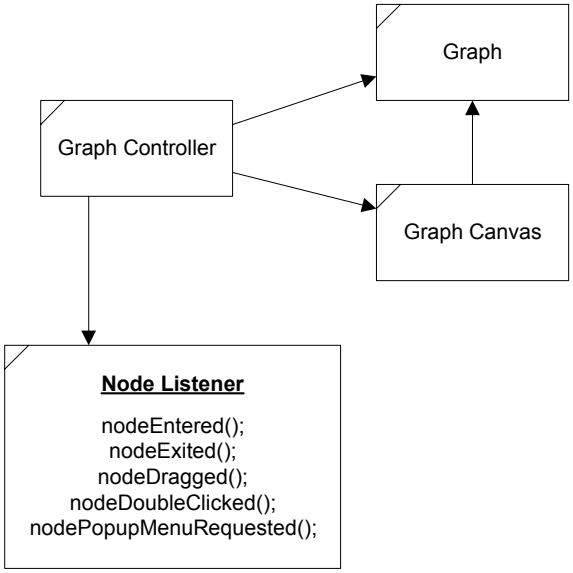
תפקידו של הבקר (Controller) הוא להגיב לפעולות המשתמש על הגרף. לכל תצוגה יתאים בקר שונה, אשר יטפל בהקשות המשתמש על הגרף.

הבקר שבניתי (GraphController), מתייחס לפעולות המשתמש על התצוגה המרכזית (GraphCanvas), ומטרתו להעניק משמעות לוגית להקשות על העכבר הנקלטות באזורים שונים של הגרף. לדוגמא, גרירת העכבר תוך הקשה על קודקוד תגרום לכך, שהקודקוד ייגרר עם העכבר, ומיקומו ישתנה. לעומת זאת, גרירת העכבר בשטח הגרף בלי להצביע על קודקוד מסוים, תביא לסימון מסגרת ולבחירת כל הקודקודים הנמצאים בתוכה.

היות והגרף נכתב בצורה כללית (ללא התייחסות לתוכן המידע המוצג), הרי שהבקר יכול להגיב רק על פעולות של המשתמש, אשר התגובה להן אינה תלויה ביישום בו משולב הגרף. דוגמאות

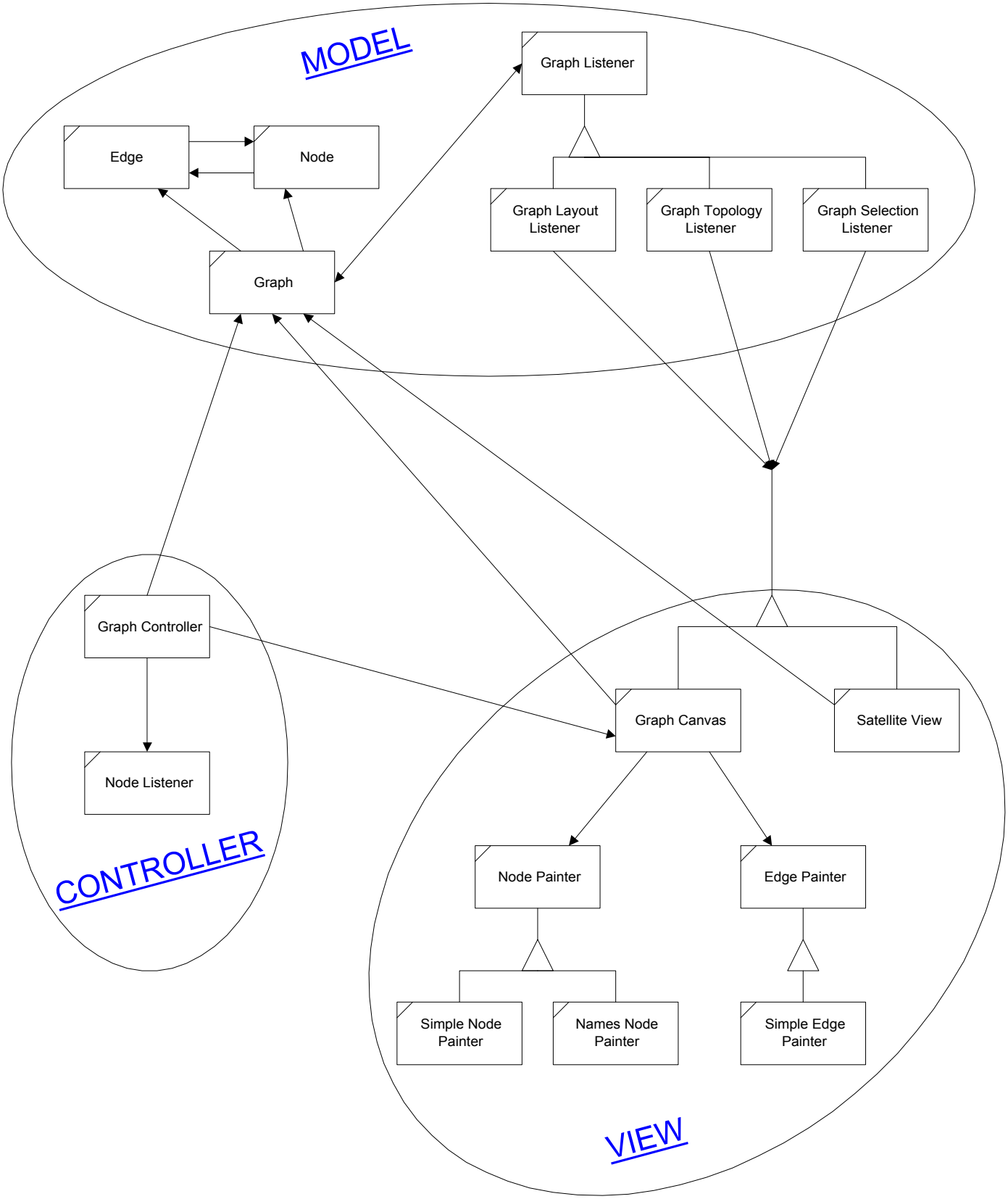
לפעולות "סטנדרטיות" כאלו הן סימון של קודקוד, או גרירתו על פני הגרף. הבקר אינו יכול להגיב לבקשות המשתמש, אשר התגובה להן עשויה להיות שונה מיישום ליישום, כגון הקשה כפולה על קודקוד (double-click). לכן, הבקר מבצע בעצמו שינויים במודל בהתאם לפעולות המשתמש הסטנדרטיות, וכן מאפשר לאובייקטים אחרים במערכת להוסיף תגובה משלהם.

לשם כך, הבקר ממיר את הקשות העכבר הבסיסיות לאירועים בעלי משמעות לוגית, עליהם הוא מדווח לאובייקטים המעוניינים (ראה איור 13). אירועים כאלו הם: גרירת קודקוד, הקשה כפולה על קודקוד, בקשת תפריט-עזר לקודקוד (הקשה על לחצן ימני), וכיו"ב. על מנת לזהות באילו קודקודים מדובר, צריך הבקר להיעזר בתצוגה (GraphCanvas) כדי לתרגם מיקום אבסולוטי על המסך לקודקוד מסוים במודל. גם הפעם נעשה שימוש במנגנון רישום, שבעזרתו נרשמים האובייקטים המעוניינים במידע כמאזינים (NodeListeners), ובכל פעם שהבקר מזהה אירוע לוגי, הוא שולח אליהם הודעת עדכון מתאימה. בעזרת מנגנון זה יכולה התוכנה השלמה PIVOT להגיב, למשל, על הקשה כפולה על קודקוד ע"י הוספת כל החלבונים השכנים אל הגרף.



איור 13. בקר הגרף

הבקר מאזין לפעולות המשתמש ומפרשן בסיועו של ה- GraphCanvas. לאחר זיהוי פעולת המשתמש והקודקודים הרלוונטיים, מבצע הבקר שינויים מתאימים במודל (Graph). רכיבים נוספים המעוניינים להגיב על פעולות המשתמש, יכולים לממש את הממשק NodeListener, ולהירשם בבקר על מנת לקבל הודעות מתאימות.



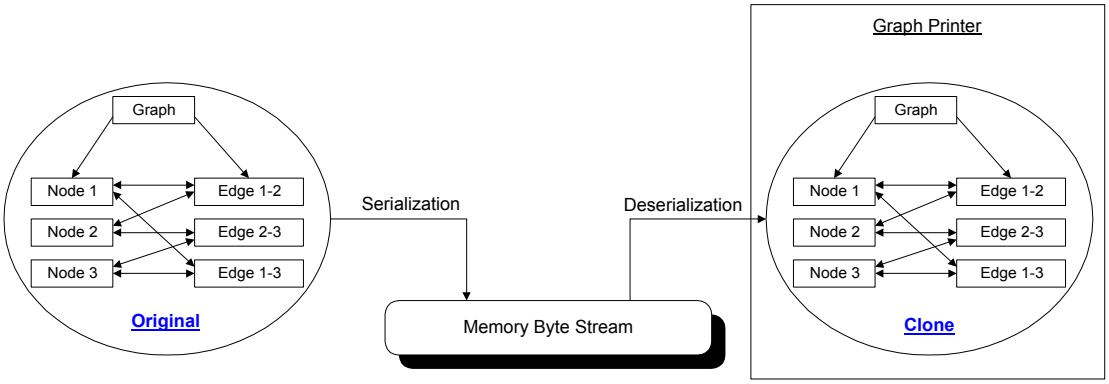
איור 14. מערכת האובייקטים המרכיבים את מודל ה-MVC, בחבילה לטיפול בגרף

4.2.4 הקלאס GraphPrinter

זהו קלאס נוסף אשר שייך לחבילת הגרף (למרות שאינו קשור ישירות למודל ה- MVC). תפקידו של קלאס זה הוא לאפשר את הדפסת הגרף למדפסת. הוא עושה שימוש בציירים לצורך הדפסת הקודקודים והקשתות בגרף, בדומה לקלאס GraphCanvas (ניתן להשתמש באותם הקלאסים של הציירים לשתי המטרות).

הקלאס מקבל כפרמטר נוסף, אובייקט מהקלאס PageFormat, אשר מכיל הגדרות כגון גודל הדף, רוחב השוליים שמחוץ לאזור ההדפסה וכו'. הקלאס מתחשב בכל הפרמטרים בהדפסת הגרף, ובמידת הצורך מודפס הגרף על פני מספר דפים. במקרה זה דואג הקלאס להשאיר אזור חפיפה בגודל אינץ' אחד בין כל שני דפים סמוכים, כך שהמשתמש יוכל לחברם בקלות לדף גדול יחיד.

היות והקלאס שולח את הדפים למדפסת בזה אחר זה, הוא מניח, כי המודל עשוי להשתנות במהלך הדפסה זו (למשל ע"י מנגנון העימוד האוטומטי המופעל ב-thread נפרד – ראה להלן). לכן, עם יצירת אובייקט מהקלאס GraphPrinter, הוא יוצר עותק של מודל הגרף ומשתמש בו לצורך ההדפסה. כדי לשכפל את המודל, מבצעים לו Serialization לתוך אזור פנוי בזיכרון, ומיד מבוצע Deserialization ליצירת עותק חדש של עץ האובייקטים (ראה איור 15).



איור 15. שכפול מודל הגרף לצורך הדפסתו

מודל הגרף משוכפל מיד עם יצירת אובייקט GraphPrinter, בכדי לאפשר המשך עבודה על הגרף המקורי במהלך ההדפסה.

4.3 מנגנון העימוד האוטומטי

אחת הבעיות המרכזיות בהן נתקלתי, היא עימוד הגרף על פני המסך בצורה שתהיה נוחה למשתמש. בנושא זה קיימות עבודות רבות, רחבות היקף, וישנן אף חבילות תוכנה, המבצעות פעולה זו.

החלטתי לנסות וליצור מערכת כזו משלי, אשר יתרונותיה העיקריים על פני מערכות תוכנה מוכנות, הינם:

- האינטגרציה הנוחה שלה למבנה הגרף שיצרתי, ללא צורך בהמרות של מבני הנתונים.
- השליטה המלאה שלי בהתנהגות המערכת, מתוך היכרות עם הקוד ועם שיטת הפעולה.
- היכולת לבצע את תהליך העימוד באופן הדרגתי, כך שתתקבל תזוזה הדרגתית של קודקודי הגרף, שתאפשר למשתמש לעקוב אחר השינויים.

4.3.1 קפיצים וגרביטציה

בבסיס פיתוח מנגנון העימוד האוטומטי עומד הרעיון של ביצוע סימולציה למערכת פיזיקאלית. היות ואנו מעוניינים, שקודקודים המקושרים בקשת ימצאו קרובים זה לזה, הקשתות ישמשו כקפיצים (בעלי אורך אפס). עם זאת, לא נרצה שהקודקודים כולם יתרכזו לאותה נקודה, ולכן לכל אחד מהם ישנם כוחות דחייה כמו למגנטים בעלי אותה קוטביות.

לצורך יישום מנגנון זה, נבנה רכיב בתוכנה, אשר רץ ב-thread עצמאי, בודק באופן מחזורי את הכוחות הפועלים במערכת, ומזיז את הקודקודים בהתאם לכוחות הפועלים עליהם.

הנוסחה הפיזיקאלית לכוח המופעל ע"י קפיץ הנה: $F = k \cdot X$ כלומר הכוח שווה לקבוע k שנקרא קבוע הקפיץ, המוכפל במידת המתיחה של הקפיץ. היות והקפיצים במערכת שלנו נקבעו להיות בעלי אורך אפס, הרי שמידת המתיחה של הקפיץ שווה למרחק בין קצותיו. מרחק זה נמדד בנקודות על המסך (פיקסלים). קבוע הקפיץ, שהנו זהה עבור כל הקפיצים במערכת, יקבע בהמשך בדרך של ניסוי וטעייה.

לחישוב כוח הדחייה הפועל בין הקודקודים, בחרתי להשתמש בנוסחה אנלוגית לנוסחת כוח הגרביטציה, אך לשנות את כיוונו של הכוח המחושב - בניגוד לכוח הגרביטציה שהוא כוח משיכה, הכוח שהגדרתי הוא כוח דחייה. נוסחת כוח הגרביטציה: $F = \frac{G \cdot m_1 \cdot m_2}{d^2}$, קובעת, כי הכוח שווה למכפלת המאסות של הגופים, כפול קבוע G, קבוע הגרביטציה, מחולקת בריבוע המרחק בין הגופים. את קבוע הגרביטציה שוב השארתי להמשך, שכן היחס בינו לבין קבוע הקפיץ יקבע את

היחס בין כוחות הדחייה של הקודקודים לכוחות המשיכה שיוצרות הקשתות. המרחק נמדד גם הפעם בפיקסלים.

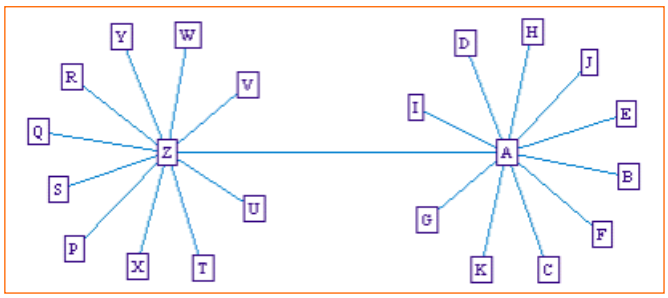
המאסות של הקודקודים העלו שאלה חדשה – האם על הקודקודים להיות שווי מאסה או שמא כדאי להעניק משמעות לגורם זה?!

4.3.2 קביעת מאסה

היות ושאלת המאסה עלתה מתוך נוסחת הגרביטציה, נשאלה השאלה, האם כוחות הדחייה צריכים להשתנות בין קודקודים שונים. היות ואנו מעוניינים לקבל פרישה ברורה של הקודקודים בגרף, הגיוני לצפות כי קודקודים "מרכזיים" בגרף, אשר מוקפים בהרבה קודקודים "קטנים", יהיו מרוחקים יחסית זה מזה, בעוד שהקודקודים ה"קטנים" הקשורים אליהם, יהיו סמוכים זה לזה ככל שניתן. מכאן עלה הרעיון, להגדיר את מאסתו של קודקוד בהתאם לדרגתו. היות וגם לקודקודים בודדים צריכה להיות מאסה (כדי שיהיו גם מולם כוחות דחייה), הוחלט להגדיר את מאסת הקודקוד כ"דרגת הקודקוד + 1".

הגדרה זו השיגה את המטרה הרצויה, ואכן קודקודים "קלים" התרכזו סביב הקודקודים ה"כבדים", בעוד שהאחרונים התרחקו זה מזה ואפשרו תצוגה מרווחת וברורה (ראה איור 16).

מאסת הקודקודים מעורבת גם בחישוב תנועתם, כפי שמתואר בסעיף הבא.



איור 16. השפעת מאסת הקודקודים על העימוד

קביעת מאסת הקודקוד בהתאם לדרגתו מאפשרת פרישה של הגרף בה קודקודים "כבדים" מרוחקים זה מזה, וסביבם מרוכזים הקודקודים ה"קלים" הקשורים אליהם.

4.3.3 חישוב תנועת הקודקודים

כפי שציינתי, קיים במערכת Thread, אשר סורק את הקודקודים באופן מחזורי, מחשב את הכוחות הפועלים על כל קודקוד, ומזיזם בהתאם. כיוון תזוזת הקודקוד, ואף גודל התזוזה, נקבעים בהתאם לסכום הכוחות הפועלים עליו. אך יש לזכור, כי הקודקודים מוזזים בזה אחר זה

ולא במקביל, ולכן איננו מעוניינים בשינויים גדולים מדי בכל שלב. כמו כן, אנו רוצים לקבל תזוזה חלקה והדרגתית של הגרף, ולא "קפיצות" גדולות של קודקודיו. לכן נוספה מגבלה על גודלה המקסימאלי של התזוזה שיבצע קודקוד יחיד בכל מחזור.

כשהזזנו את הקודקודים בהתאם לגודל הכוח הפועל עליהם, קיבלנו את התוצאה שרצינו, אך נתקלנו בבעיה חדשה – הקודקודים המרכזיים בגרף (אלו שלהם שכנים רבים), לא הגיעו למצב יציב אלא "רעדו" מצד לצד. כשחזרנו אל דפי הנוסחאות גילינו, שכוחות הדחייה הפועלים על קודקודים אלו הם גדולים יחסית, שכן המאסה שלהם גדולה. כמו כן הם מוקפים בקודקודים רבים, ולכן פועלים עליהם כוחות רבים אשר גורמים להם לנוע הלוך ושוב.

היות והבעיה אפיינה את הקודקודים בעלי המאסה הגדולה, החלטנו לשלב את מאסת הקודקוד בחישובים לגבי תנועתו. בהתבסס על אינטואיציה פיזיקאלית**, קבענו שתזוזת הקודקוד בכל יחידת זמן תהיה ביחס ישר לגודלו של הכוח הפועל עליו וביחס הפוך למאסת הקודקוד. כשהוספנו את השיקול הנ"ל לחישוב תזוזות הקודקודים, קיבלנו תזוזה חלקה יותר של המערכת, ותופעת

```
For every currentNode ∈ Nodes {
        SpringsForce = 0;
        For every otherNode ∈ AdjacentNodes(currentNode)
                SpringsForce += dist(currentNode, otherNode);   // dist := spring's length

        GravityForce = 0;
        For every otherNode ∈ Nodes
                GravityForce += mass(currentNode) * mass(otherNode) / dist^2;

        Force = SpringForce - SPRING_GRAVITY_RATIO * GravityForce;

        location(currentNode) += FORCE_SIZE_CONST * Force / mass(currentNode);
}
```

איור 17. אלגוריתם עימוד הגרף

אלגוריתם העימוד מחשב עבור כל קודקוד את סך הכוחות המופעלים עליו - כוחות משיכה מצד הקודקודים הסמוכים לו, וכוחות דחייה מצד כל קודקודי הגרף. מיקומו של הקודקוד מעודכן בהתאם לעוצמת הכוח ולכיוונו, ולמאסת הקודקוד.

הרעידות נעלמה לחלוטין. איור 17 מסכם את אלגוריתם העימוד.

4.3.4 נעילת קודקודים (זמנית וקבועה)

כדי לאפשר למשתמש להתערב בפעולת עימוד הגרף מבלי להפסיק לחלוטין את עזרתה של מערכת העימוד האוטומטי, נוספה למערכת זו האופציה לנעילת מיקומם של הקודקודים. קודקוד שמיקומו נעול, לא יוזז ממקומו ע״י מערכת העימוד האוטומטי.

הנעילה הזמנית מאפשרת למשתמש לגרור קודקוד מסוים למקום שונה בגרף, בלי שמערכת העימוד האוטומטי תתערב ותפריע לו בכך. כל עת שהקודקוד ״מטופל״ ע״י המשתמש, הוא נעול זמנית להתערבות המערכת.

המשתמש יכול לנעול קודקודים באופן קבוע, על מנת להשאירם במקומם. בד״כ יעשה שימוש בתכונה זו כדי לקבוע את מיקומי קודקודי הגרף העיקריים, בזמן ששאר קודקודי הגרף יעומדו באופן אוטומטי. אפשרות אחרת היא לקבע קטע מהגרף שרוצים שיישאר ללא שינוי, בזמן שהמשתמש ממשיך לעבוד עם הגרף, ולחקור אזורים אחרים בו.

4.3.5 ״בריחת״ הגרף מהמסך

היות ומנגנון העימוד האוטומטי מניע את הגרף באופן יחסי, ומבלי להתחשב במיקומם האבסולוטי של הקודקודים, התעוררו מצבים בהם ״ברחו״ קודקודי הגרף אל אזורים המרוחקים מהנקודה (0,0). מצב זה אינו רצוי, שכן הוא עלול לגרום לגרף ״לנדוד״ אל קואורדינטות במספרים שחורגים מתחומי ה-integer, ולגרום לסיבוכים מיותרים. יתרון נוסף בשמירת הגרף באזור הנקודה (0,0), הוא בהפחתת החישובים הדרושים לציור הגרף.

בעיה זו נפתרה ע״י הקלאס Graph עצמו, אשר מכיר בכל זמן נתון את גבולותיו. היה ופינתו השמאלית העליונה של הגרף חורגת מהאזור שהוגדר עבורה, מתבצעת הזזה של כל קודקודי הגרף בכיוון המתאים בכדי להחזירו לטווח הקואורדינטות הרצוי. תוך כדי ביצוע ההזזה, נחסמת פעולתו של מנגנון העימוד האוטומטי, כדי לא ליצור התנגשות בין התהליכים.

** הנוסחה הפיזיקאלית המתארת את המרחק שעובר גוף בזמן t (הנע בתנועה שוות תאוצה), היא $X = v_0 t + \frac{1}{2} a t^2$. אם נזניח את מהירותו ההתחלתית של הגוף – V_0, ונתבונן במרחק אשר הגוף עובר ביחידת זמן אחת (t = 1), נקבל כי $X = \frac{1}{2} a$. החוק השני של ניוטון קובע כי $F = m \cdot a$, ועל כן נקבל כי $X = \frac{F}{2m}$.

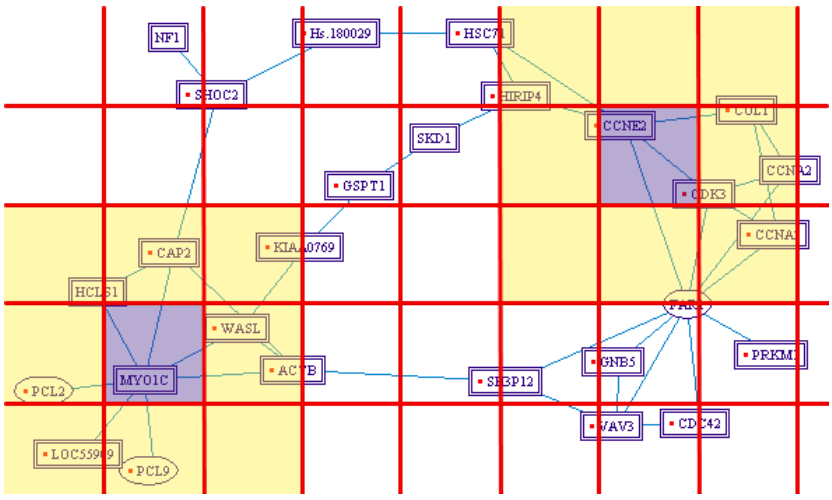
4.3.6 <u>שיפור מהירות העימוד ע"י הקלאס NodesGPS – ניסיון שכשל</u>

הסיבוכיות בה פועל מנגנון העימוד כיום הנה $O(n^2)$ פעולות בכל מחזור עבור גרף המכיל n קודקודים, שכן בכל מחזור ביצוע, עובר המנגנון על כל אחד מקודקודי הגרף, ומחשב את הכוחות הפועלים עליו:

- כוחות המשיכה (הקפיצים) מחושבים רק כנגד הקודקודים השכנים בגרף, בסיבוכיות $O(1)$ לשכן, ועבור גרף שבו m קשתות, הסיבוכיות הכוללת של חישוב כוחות המשיכה היא $O(m)$.
- כוחות הדחייה (המגנטים) מושפעים בעיקר מהקודקודים אשר קרובים פיזית (על פני המסך) זה לזה. לכן קיים הצורך לעבור על <u>כל</u> הקודקודים ולבחון אלו מהם סמוכים ואלו מרוחקים, חישוב הדורש $O(n^2)$ פעולות.

בהנחה שהגרף פרוש ומרווח, היינו מצפים שמספר הקודקודים הסמוכים פיזית לכל קודקוד << n. לכן, אם היינו יודעים בסיבוכיות של $O(1)$ מיהם הקודקודים הסמוכים פיזית לכל קודקוד, היינו יכולים להקטין את סיבוכיות חישוב כוחות הדחייה ל-$O(n)$, ואת סיבוכיות החישוב כולו ל-$O(n+m)$.

לשם כך יצרתי קלאס ששמו NodesGPS, אשר חילק את מפת הגרף לתאים בגדל פיזי קבוע (ראה איור 18). ע"י מעקב אחרי השינויים החלים במודל הגרף, ידע הקלאס בכל שלב אילו קודקודים נמצאים בכל תא במפה. הקודקודים הסמוכים לקודקוד נתון הם אלו הנמצאים עימו באותו תא, או בתאים הסמוכים אליו. כלומר, קלאס זה



איור 18. חישוב מיקום הקודקודים המבוצע ע"י NodesGPS

מפת הגרף מחולקת לתאים בעלי גודל קבוע, ובכל תא נשמרת רשימת הקודקודים המוכלים בו. הקודקודים הסמוכים לקודקוד נתון הם אלו השוכנים עימו באותו תא, או בתאים המקיפים אותו (למשל MYO1C, או CCNE2). כך ניתן לברר מיהם קודקודי הגרף הנמצאים בסמיכות פיזית לקודקוד מסוים, מבלי לחשב את מרחקו מכל קודקודי הגרף האחרים.

שמר את המידע כך שניתן היה לקבל ב-O(1) את רשימת הקודקודים הממוקמים באזור פיזי מבוקש. השימוש בקלאס זה, איפשר להקטין את סיבוכיות מנגנון העימוד האוטומטי, ל-O(n+m).

הבעיה הייתה שהעדכון של המידע באובייקט ה-NodesGPS היה כה אינטנסיבי, שהתוספת הביאה לירידה במהירות התהליך במקום לשיפור לו ציפינו. חשוב לציין, כי הכלי יועד בעיקר לעבודה עם מספר נמוך של קודקודים, ולכן היה חשוב יותר לקבל עבודה מהירה עם מעט קודקודים מאשר לתמוך ביעילות בעבודה עם מספר קודקודים רב.

לפיכך, מערכת ה-NodesGPS בוטלה ונגנזה לטובת המערכת המקורית.

4.4 הטיפול במאגר המידע הביולוגי

4.4.1 חבילות biology, ו-pdb

היות ועל התוכנה לטפל במידע ביולוגי, בניתי מספר קלאסים אשר תפקידם לייצג מידע זה, ולאפשר שאילתות לגביו.

החבילה biology מכילה את הקלאסים:

Species – המייצג אורגניזם מסוים, כגון *H. sapiens*, או *S. cerevisiae*.

Protein – המייצג חלבון ומאופיין ע"י שם החלבון והאורגניזם לו הוא שייך.

Interaction – המייצג אינטראקציה בין שני חלבונים.

החבילה pdb מרכזת קלאסים, המייצגים מאגרי מידע הנוגעים לחלבונים (protein database), ושמטרתם לאפשר ביצוע שאילתות ביולוגיות לגבי חלבונים ואינטראקציות. שני הקלאסים העיקריים בחבילה זו הם Pdb המייצג מידע על חלבונים (protein database), ו-Pidb המייצג מידע על אינטראקציות חלבונים (protein interactions database).

הקלאס Pdb מאפשר למשתמש לבקש מידע הנוגע לחלבון מסוים. מידע כזה הוא, למשל, מחרוזת תווים, המתארת בקצרה את החלבון, את החלבון ההומולוגי לחלבון זה בזן אחר, או המפנה אל כתובת האינטרנט של דף המידע, המתאר את החלבון. מידע כזה אינו ייחודי לחלבון, אלא תלוי בתוכן מאגר המידע בו משתמשים. שאילתה זהה, שתישלח למאגרי מידע שונים, תחזיר מידע שונה, ולכן השאילתות אינן מבוצעות ישירות דרך הקלאס Protein, אלא דרך הקלאס Pdb המייצג מאגר מידע מסוים.

באופן דומה, הקלאס Pidb מייצג מאגר מידע מסוים של אינטראקציות חלבונים, ומאפשר שאילתות לגביהן, כגון קבלת רשימת כל האינטראקציות של חלבון מסוים, או כתובת האינטרנט של דף המידע המתאר אינטראקציה נתונה.

4.4.2 מאגר המידע YPD

במהלך הפיתוח השתמשתי במידע אשר קיבלנו מחברת .Proteome, Inc. מקור המידע הוא במאגר המידע YPD [77], ממנו ניתן לנו ע"י החברה עותק חלקי, המכיל כמות נתונים מוגבלת. היתרונות בקבלת עותק ממאגר המידע טמונים בגישה המהירה המתאפשרת אליו כמידע מקומי, ומכאן גם מהירות התגובה של התוכנה. החסרונות הם בכך, שהמידע אינו מקוון (on-line) ולכן אינו מתעדכן, והוא מידע חלקי בלבד, אשר אינו מכיל את כל הפרטים, בהם רצינו לעשות שימוש. נאלצנו להשתמש בעותק מקומי של המידע גם עקב שיקולים מסחריים של חברת .Proteome Inc, אשר אינה מאפשרת שליפת מידע על מספר חלבונים גדול ממאגר המידע שלה דרך רשת האינטרנט.

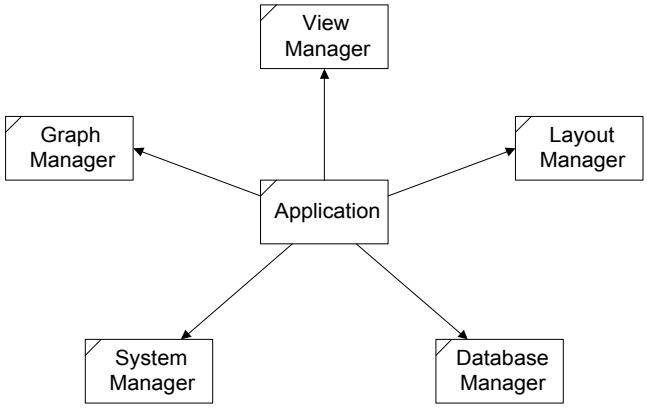
במהלך העבודה הפך הקשר עם חברת .Proteome Inc לבעייתי, עקב שינויים ארגוניים שהתרחשו בחברה, ועל כן לא יכולנו לקבל מידע עדכני יותר, ופרטים נוספים שביקשנו על כל חלבון ואינטראקציה. זו גם הסיבה לכך, שאופציות מסוימות בתוכנה נותרו לא ממומשות (כגון האפשרות לקישור כל אינטראקציה בגרף אל המאמר התומך בה).

לצורך העבודה מול מאגר מידע זה, נכתבו הקלאסים הספציפיים YeastPdb ו-YeastPidb. הקלאס YeastPdb יורש מהקלאס Pdb, ומספק מידע לגבי חלבוני השמר, ואילו הקלאס YeastPidb יורש מ- Pidb, ומטפל במידע הנוגע לאינטראקציות החלבונים.

4.5 בניית הרכיבים לאפליקציה

התוכנה מורכבת מחמש תת-מערכות, אשר מנהלות היבטים שונים של המערכת. תת-מערכות אלו הן: תת-מערכת מאגרי המידע, תת-מערכת הגרף, תת-מערכת התצוגה, תת-מערכת העימוד האוטומטי, ותת-מערכת התקני הקבצים וההדפסה (ראה איור 19). בראש כל אחת מתת-המערכות הנ"ל עומד "מנהל", אשר מסתייע באובייקטים רבים נוספים לצורך עבודתו, ואשר אחראי על יצירתם, ניהולם וסגירתם המסודרת. כאשר אחת מתת-המערכות זקוקה לשירותים מחלק אחר במערכת, היא פונה אל ה"מנהל" המתאים לצורך קבלתם.

הקלאס הראשי בתוכנה הנו הקלאס Application, אשר דואג ליצור את כל תת-המערכות עם הרצת התוכנה (ע״י יצירת המנהלים השונים), להודיע לכל המנהלים על ירידת המערכת ביציאה מהתוכנה, ולאפשר לאובייקטים השונים בתוכנה לאתר את המנהל הדרוש בשעת הצורך.



איור 19. תת המערכות המרכיבות את האפליקציה

בראשה של כל תת-מערכת עומד "מנהל" מתאים, והקלאס Application דואג לאתחול תת-המערכות ולקישור ביניהן.

4.5.1 מנהל מאגרי המידע

תפקידו של מנהל מאגרי המידע הוא לאתחל את מאגר המידע מולו עובדת המערכת עם עלייתה, ולכוון את השאילתות השונות אל מאגר מידע זה.

מנהל מאגרי המידע מספק ממשק, דרכו יכולים חלקי המערכת האחרים לקבל מידע לגבי חלבון נתון, כגון ההומולוג שלו בזנים אחרים, תיאור החלבון, החלבונים אשר נמצאים עמו באינטראקציה, וכו׳. הוא דואג לספק את המידע הנ״ל מתוך מאגרי המידע עמם הוא עובד, ללא צורך שהמשתמש ידע כיצד המידע מאורגן והיכן ניתן למצאו.

בכדי לבצע את מטלותיו עובד הקלאס DatabaseManager מול הקלאסים Pdb ו-Pidb, אשר מגדירים שאילתות כלליות לפנייה למאגרי מידע של חלבונים. בעבודה מול ה-YPD משתמש הקלאס בקלאסים הספציפיים YeastPdb ו-YeastPidb אשר ניגשים לטבלאות הלקוחות מתוך מאגר מידע זה. כדי להחליף את מאגר המידע מולו עובדת האפליקציה, יש לבצע את השינוי בקלאס זה, ולדאוג לאתחלו עם קלאסים ספציפיים אשר קוראים את המידע מטבלאות אחרות.

4.5.2 מנהל הגרף

תפקידו של מנהל הגרף הוא לטפל בחלק המודל של הגרף. טיפול זה כולל:

- אתחול המודל.
- התאמת המודל הכללי לגרף של חלבונים.
- ביצוע Serialization ו- Deserialization של המידע לצורך שמירת הגרף לקובץ או שכפולו

- טיפול בהוספת חלבונים חדשים לגרף תוך שמירה על עקביות המידע המוצג.
- ביצוע שאילתות לסריקה אוטומטית של הגרף (לצורך פרישת כל החלבונים באזור נתון, או לחיפוש מסלולים בין חלבונים מרוחקים).

לצורך התאמת המודל לעבודה עם גרף של חלבונים, נבנה הקלאס PIGraph (Protein Interactions Graph), אשר יורש מהקלאס Graph. קלאס זה דואג להצמיד לשדה userdata של כל קודקוד בגרף את החלבון אותו הוא מייצג (אובייקט מהקלאס Protein). כמו כן, בעת ביצוע Serialization של הגרף ה-userdata אינו נרשם לקובץ, ולכן באחריותו של קלאס זה לעדכן את המידע הנ"ל בעת ביצוע Deserialization.

מנהל הגרף מספק ממשק, המאפשר לכל חלקי המערכת לבצע פעולות שונות על הגרף, ולהתייחס לחלבונים בגרף (Proteins) במקום לקודקודים בו (Nodes). הוא דואג לוודא את תקפות הפעולות שמבצע המשתמש (למשל, אין להוסיף לגרף את אותו החלבון פעמיים), ולבצע את כל השלמות המידע הדרושות לשמירה על עקביות הגרף (למשל, בעקבות הוספת חלבון חדש לגרף יש להוסיף את כל הקשתות בין חלבון זה לבין הקודקודים בגרף, איתם הוא נמצא באינטראקציה).

היות ומנהל הגרף אחראי למבנה המתמטי של הגרף, הוא מטפל גם בשאילתות המורכבות יותר, הדורשות חיפוש אלגוריתמי על פני גרף החלבונים.

4.5.2.1 תצוגת שכנים - NodeExpander

תפקידה של אחת השאילתות הבסיסיות ביותר הוא להציג את שכניו של חלבון כלשהו (ע"י ביצוע double-click על החלבון). הרחבה של שאילתה זו היא להציג את כל שכניו של החלבון עד למרחק נתון. בכדי לענות על שאילתות אלו נוצר הקלאס NodeExpander.

קלאס זה משתמש בשירותים של מנהל מאגרי המידע, ושל מנהל הגרף. הוא פועל ב-thread נפרד, על מנת לאפשר לגרף להציג את הוספת הקודקודים באופן הדרגתי ולעמדם תוך כדי עבודתו, וכדי להציג אינדיקציה למידת התקדמות התהליך (רלוונטי להצגת סביבת השכנים במרחק 2 ומעלה).

הקלאס מופעל ע"י מנהל הגרף, תוך שימוש בפרמטרים קבועים (ברירת המחדל היא להציג את הסביבה במרחק 1 בלבד), או בפרמטרים משתנים לאחר שאלו נקלטים דרך חלון דיאלוג מתאים (NodeExpanderDialog הנתון באחריותו של ה- ViewManager).

חיפוש השכנים (ראה איור 20) נעשה ע"פ אלגוריתם BFS [101], המתחיל מקבוצת קודקודים, ועובר מכל אחד מהם אל כל שכניו שעדיין לא פגש בדרכו. רק לאחר הטיפול בכל השכנים המיידיים, עובר האלגוריתם אל השכנים שלהם. לצורך כך, משתמש הגרף בתור (queue) בשם expansionQueue, המכיל את הקודקודים שיש לפרוש, ועבור כל אחד מהם את המרחק אליו נרצה להמשיך ולפרוש ממנו. כמו כן, נעשה שימוש בקבוצה (set) בשם alreadyExpanded המכילה את רשימת הקודקודים אשר כבר טופלו במהלך ריצת האלגוריתם. בכל פעם שקודקוד נפרש, כל שכניו <u>החדשים</u> נוספים אל סוף התור, עם מרחק פרישה קטן באחד. התור תמיד יכיל בראשו את מרחקי הפרישה הגדולים, ובסופו את הקטנים.

```
While expansionQueue is not empty {
        currentNode, currentDist  = expansionQueue.getHead();
        alreadyExpanded.add(currentNode);

        expandNodeOneLevel(currentNode);    // see below

        If (currentDist > 1)
                Foreach u ∈ adjacentNodes(currentNode)
                        If (u ∉ alreadyExpanded)
                                expansionQueue.addToTail( u, currentDist - 1);
}
```

איור 20. אלגוריתם פרישת השכנים

על מנת לפרוש את כל שכניו של קודקוד עד למרחק נתון, נעשה שימוש באלגוריתם BFS באופן רקורסיבי - אנו פורשים את שכניו המיידיים של הקודקוד, ועוברים לפרוש כל אחד מהשכנים הנ"ל עד למרחק נמוך באחד מהמרחק המקורי.

עלינו לזכור, כי החיפוש אינו נעשה על פני הגרף המוצג, אלא ע"פ הגרף הכללי, כפי שמתואר במאגר המידע מולו עובדים. לכן על קלאס זה לבצע שאילתות על מאגר המידע, אך לא להוסיף לגרף המוצג מידע שכבר מופיע בו בשנית (איור 21).

```
expandNodeOneLevel(Node n) {
        dbProteins = database.getAdjacentProteins(n);
        graphAdjacentProteins = graph.getAdjacentProteins(n);
        newProteins = dbProteins - graphAdjacentProteins;
        graph.add(newProteins);
}
```

איור 21. אלגוריתם לפרישת שכניו המיידיים של קודקוד בודד

בפרישת שכניו המיידיים של קודקוד נתון אנו שולפים את כל שכניו ממאגר המידע, אך דואגים להוסיף אל הגרף רק את השכנים, אשר עדיין אינם מופיעים בו.

4.5.2.2 מציאת מסלול

שאילתה חשובה אחרת מחפשת את המסלולים הקצרים ביותר, המחברים שני חלבונים נתונים על פני הגרף.

שאילתה זו שימושית מאוד במקרים, בהם מסתמן קשר פונקציונאלי בין החלבונים בקבוצה מסוימת (למשל כאשר נצפית עליה ברמת הביטוי שלהם בתנאים מסוימים), בעוד שהקשר הביוכימי ביניהם אינו ידוע.

לצורך ביצועה של שאילתה זו נוצר הקלאס PathFinder, אשר מופעל אף הוא ע"י מנהל הגרף. קלאס זה מקבל חלבון אשר אינו מופיע בגרף, ותת-גרף אליו יש לקשרו (קבוצת קודקודים בת קודקוד אחד או יותר). הקלאס מחפש על פני מאגר המידע את מסלול האינטראקציות הקצר ביותר אשר מאפשר לקשר את החלבון המבוקש אל תת הגרף (מסלול יחיד או מספר מסלולים שווי אורך). יש לשים לב, כי לא מדובר במסלול במובן של pathway, הכולל כיוון פעולה וסדר התרחשות, אלא במסלול לא מכוון בגרף האינטראקציות.

הנתונים הדרושים לקלאס זה נקלטים בעזרת מנהל התצוגה, דרך הדיאלוג PathFinderDialog. לאחר קליטת הנתונים, מפעיל מנהל הגרף את ה-PathFinder, אשר מבצע את החיפוש ב-thread נפרד. למשתמש מוצג אינדיקאטור המעיד על מידת התקדמות החיפוש, והוא יכול לבטלו כל עוד לא נמצא המסלול המבוקש. אורכו המקסימאלי של המסלול נקבע אף הוא ע"י המשתמש.

החיפוש נעשה ע"י BFS, המתחיל מהחלבון החדש (אשר אינו מופיע בגרף), ומסתיים עם מציאת אחד או יותר מחלבוני היעד. משם ממשיכים לשחזור המסלול שאותר ע"י Back Tracking,

ולהוספת מסלול זה אל הגרף המוצג למשתמש. החיפוש נעשה כמובן על פני גרף האינטראקציות המלא המתואר במאגר המידע.

התהליך מוגדר באופן רקורסיבי, כך (ראה איור 22 ואיור 23):

- fromProteins - קבוצת חלבוני המקור המאותחלת לחלבון החדש, אותו אנו מעוניינים לקשר לגרף. קבוצה זו מועברת כפרמטר לפונקציה הרקורסיבית, ומשתנה מקריאה לקריאה.
- toProteins - קבוצת חלבוני היעד, המכילה חלבון אחד או יותר השייכים לגרף המוצג, ואשר אליהם אנו מעוניינים לקשר את החלבון החדש. קבוצה זו נשארת קבועה לאורך הרקורסיה.
- withoutProteins - קבוצת החלבונים, אשר אין לעבור דרכם במסלול המבוקש. קבוצה זו ריקה בתחילת התהליך, ולאורך הרקורסיה מצטרפים אליה החלבונים אשר אותם כבר ביקרנו, כדי למנוע תנועה במעגלים.
- maxMid - המספר המרבי של קודקודי הביניים, בהם ניתן להשתמש בדרך אל חלבוני היעד. מספר זה מהווה פרמטר לפונקציה, וקטן באחד בכל קריאה רקורסיבית.
- תנאי העצירה - maxMid<0 או שקבוצת קודקודי המקור כוללת בתוכה אחד או יותר מקודקודי היעד.
- תהליך הרקורסיה - בכדי לעדכן את הגרף במסלול הקצר ביותר המוביל מחלבוני המקור אל חלבוני היעד, נפעיל רקורסיה, אשר תעדכן את הגרף במסלול הקצר ביותר המוביל משכניהם של חלבוני המקור אל חלבוני היעד ואינו עובר דרך חלבוני המקור, ואשר תחזיר את אוסף השכנים, אשר שייכים למסלול שנמצא. לגרף המתקבל נוסיף את חלבוני המקור הסמוכים לאוסף השכנים שהוחזר (אלו המשתתפים במסלול שזוהה).
- הערך המוחזר - קבוצת קודקודי המקור, אשר משתתפים במסלול שזוהה, ואשר הוספו אל הגרף. אלו הם ה"שכנים" של שלב הרקורסיה הקודם, המשתתפים במסלול.

בכדי לייעל את שלב ה-Back Tracking, נשמר במהלך החיפוש מידע לגבי כיווני התנועה בגרף, בטבלה בשם returnPath. בעזרת מידע זה נשחזר את המסלול לאחר ההגעה אל חלבוני היעד.

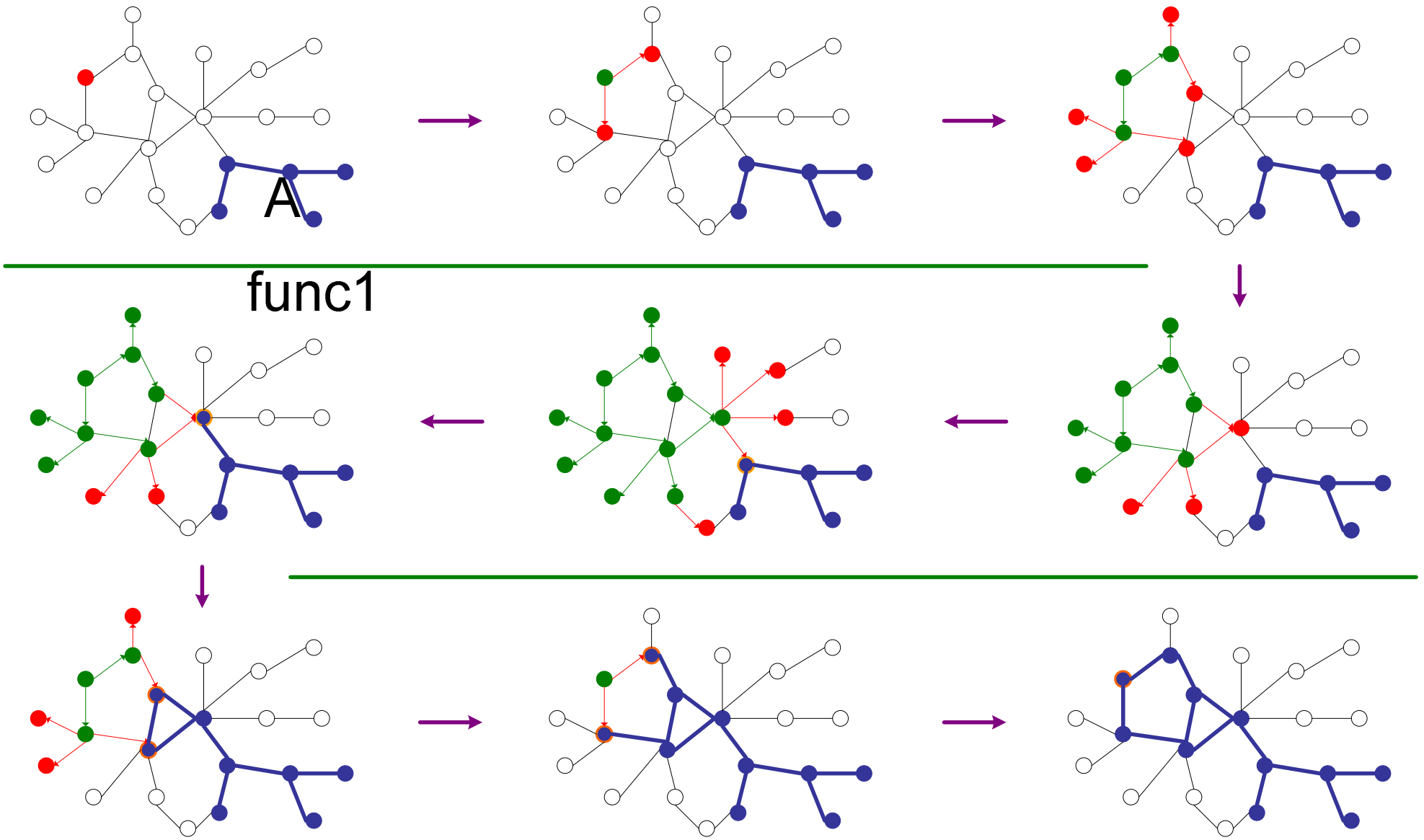
אם במהלך החיפוש נמצא מסלול בעל אורך מתאים, חוזרת הרקורסיה תוך בניית המסלול, החל מקודקודי היעד אליהם הגיע המסלול, אל חלבון המקור, שאותו יש להוסיף לגרף המוצג. בכל שלב בחזרה של הרקורסיה, נוספים לגרף אותם קודקודים מהקבוצה fromProteins, המתחברים אל המסלול, שנבנה עד כה. איור 23 מדגים את פעולת האלגוריתם.

```
connectNewProtein(fromProteins, maxMid, withoutProteins) {
        //  Check recursion exit condition:
        reachedProteins = fromProteins ∩ toProteins;
        if (reachedProteins <> ∅) return reachedProteins;
        If (maxMid < 0) return null;     // Path Not Found!

        //  Prepare parameters for recursive call:
        nextFromProteins = ∅;
        foreach fp ∈fromProteins {
                neighbors = database.getAdjacentProteins(fp);
                neighbors.remove(withoutProteins);
                nextFromProteins.add(neighbors);
                returnPath.add(fp, neighbors);
        }
        nextWithoutProteins = (withoutProteins ∪ nextFromProteins);

        //  Recursive call:
        neighborsOnPath = connectNewProtein(nextFromProteins, maxMid - 1, nextWithoutProteins);

        //  Update graph:
        fromProteinOnPath = returnPath.getOrigin(neighborsOnPath);
        graph.addProteins(fromProteinsOnPath);
        Return fromProteinsOnPath;
}
```

איור 22. אלגוריתם רקורסיבי למציאת המסלול הקצר ביותר המקשר חלבון חדש לגרף

אלגוריתם זה מתחיל מהחלבון החדש, אותו מעוניינים להוסיף לגרף, ונע באופן רקורסיבי מחלבון זה אל שכניו, ומשם אל שכניהם, עד להגעתו אל אחד מקודקודי היעד, או עד לעצירתו לאחר מספר צעדים נתון. כדי למנוע תנועה במעגלים, אנו זוכרים בכל שלב מי הם הקודקודים בהם כבר ביקרנו, ונמנעים מלעבור דרכם שנית.

חיפוש המסלול הקצר ביותר מבוצע ברקורסיה. כדי לאתר את המסלול הקצר ביותר, המתחיל באחד או יותר מקודקודי המקור, אנו מחפשים את המסלול הקצר ביותר, המתחיל באחד או יותר <u>מהשכנים</u> של קבוצת קודקודי המקור, ועם מציאתו מצרפים אליו את קודקודי המקור המתאימים.



איור 23. דוגמא לפעולתו הרקורסיבית של האלגוריתם לקישור חלבון חדש לגרף במסלול הקצר ביותר

החיפוש מתחיל מהחלבון החדש ונע אל שכניו (A-E). במהלך החיפוש מסומנים הקודקודים, אשר אותם ביקרנו בעבר (בירוק). עם ההגעה לאחד מחלבוני היעד (E), מתחילה החזרה מהרקורסיה, ובמהלכה מעודכן הגרף המוצג תוך שימוש במידע שנשמר ב-returnPath (E-I). התווית "#func" המופיעה ליד כל איור, מעידה על מספרה של הקריאה הרקורסיבית אליה התרשים רלוונטי.

4.5.3 מנהל התצוגה

מנהל התצוגה אחראי לכל נושאי התצוגה בתוכנה. נושאים אלו כוללים את בניית החלונות והדיאלוגים השונים בתוכנה והצגתם, בניית התפריטים השונים, קישור כל הרכיבים אל המודלים המתאימים להם, וקבלת קלט מהמשתמש והעברתו לרכיבים המתאימים.

מנהל התצוגה משתמש בקלאסים רבים על מנת לבצע את תפקידיו השונים, אך מציע בעצמו ממשק לכל אותם שירותים שהוא מספק, עבור חלקי התוכנה האחרים.

4.5.3.1 המסך הראשי – AppWindow

המסך הראשי של התוכנה מנוהל ע"י הקלאס AppWindow. קלאס זה יוצר את חלקי המסך השונים ודואג לקשרם אל הגרף הנוכחי ולעדכנם עם פתיחתו של גרף שונה.

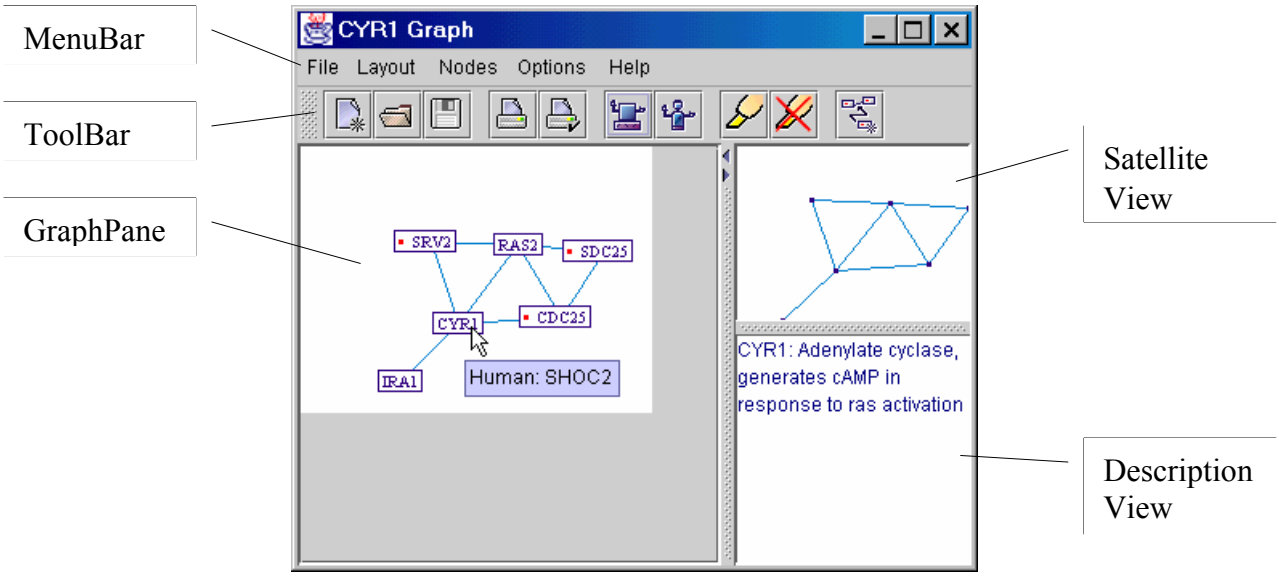
המסך הראשי מחולק לשלושה חלקים עיקריים (ראה איור 24) –

א. אזור הגרף המרכזי, דרכו מבוצעת עיקר העבודה. אזור זה נבנה ומנוהל ע"י הקלאס GraphPane, ומתואר בהמשך.

ב. תמונת הלווין בה נראה תמיד הגרף כולו, המוצג ע"י הקלאס SatelliteView.

ג. אזור בו מופיע מידע לגבי החלבון עליו מצביע סמן העכבר, הנקרא DescriptionView.

בנוסף לרכיבים אלו, מכיל המסך הראשי גם תפריט טקסטואלי – MenuBar ותפריט סמלים – ToolBar:



איור 24. המסך הראשי של התוכנה

4.5.3.2 התפריטים בתוכנה – ToolBar ו- MenuBar

לצורך יצירת התפריטים נעשה שימוש במנגנון של action objects. מטרתם של אובייקטים מסוג זה היא לייצג פעולה מסוימת (כגון "פתיחת קובץ" או "הפעלת מנגנון העימוד האוטומטי"), ולהכיל את כל הדרוש לביצועה. כל אובייקט יודע את שם הפעולה ואת תיאורה, מכיל תמונת icon המייצגת את הפעולה, ויודע לבצע את הפעולה בעת הצורך.

על ידי שימוש באובייקטים אלו מושגות שתי מטרות: ראשית, כל פרטי הפעולה ותכונותיה מרוכזים במקום אחד, וניתן לשנותה בקלות יחסית. שנית, ניתן לאפשר את ביצוע אותן הפעולות, הן מהתפריט הטקסטואלי והן מתפריט הסמלים, ע"י קישורן לאותם אובייקטי הפעולה. בכך נחסך הצורך להגדירן מספר פעמים. כמו כן, ביצוע שינוי באובייקט הפעולה יגרור את עדכונם של שני התפריטים גם יחד. למשל, הפיכתו של אובייקט פעולה מסוים ללא פעיל (disabled), תגרור את סימון הפעולה כמבוטלת, בכל התפריטים בהם הוא מופיע.

הקלאס ActionsHandler, יוצר את אובייקטי הפעולה השונים, ודואג שעבור כל פעולה יועבר אותו אובייקט לשני התפריטים. לקלאסים AppMenuBar ו- AppToolBar לא נותר אלא לפנות ל- ActionsHandler, לבקש את אובייקטי הפעולה הרצויים ולהציבם בסדר הרצוי.

4.5.3.3 אזור הגרף המרכזי – GraphPane

הקלאס GraphPane מנהל את אזור התצוגה המרכזי. הוא דואג ליצור אובייקט מהקלאס GraphCanvas, לספק לו ציירים מתאימים לקודקודים ולקשתות (NodePainter ו-EdgePainter) ולקשרו למודל הגרף, המנוהל ע"י מנהל הגרף. כמו כן, הוא יוצר GraphController אשר מקושר לגרף זה, ודואג להאזין לפעולות המשתמש ולהגיב עליהן.

כפי שתואר לעיל, ה-GraphController מטפל בפעולות המשתמש שהן כלליות לכל גרף (ללא התייחסות ספציפית לגרף של חלבונים), ביניהן גרירת קודקודים, בחירת קודקודים וכו'. כמו כן, הוא יוצר דיווחים לגבי NodeEvents אשר מתבצעים על הגרף, עבור NodeListeners אשר מאזינים לאירועים אלו. NodeListeners כאלו מפורטים בהמשך, וביניהם GraphPaneNodeListener, DescriptionView, ו- NeighborsMarker.

4.5.3.4 GraphPaneNodeListener

זהו קלאס, המופעל ע״י GraphPane, ומאזין ל-NodeEvents. הוא מרחיב את הטיפול בקלט באופן ספציפי לאפליקציה זו ולגרף של חלבונים. קלאס זה דואג לטפל בשני אירועים עיקריים:

א. ביצוע double-click על קודקוד גורם להרצת שאילתה להוספת כל שכניו של הקודקוד לגרף.

ב. פתיחה של תפריט pop-up בעקבות הקשה מתאימה על קודקוד (בקשת תפריט העזר משתנה בין מערכות מחשב שונות, למשל, במערכות Windows משתמשים בלחצן הימני של העכבר).

4.5.3.5 סמן השכנים – NeighborsMarker

קלאס נוסף אשר מטפל ב-NodeEvents הנו NeighborsMarker. קלאס זה מאזין לאירועים של כניסת הסמן לשטחו של קודקוד ועזיבתו את השטח. לאורך כל זמן פעולתו הוא שומר רשימה מעודכנת של קודקודים, אשר נמצאים בשכנות לקודקוד עליו מצביע סמן העכבר (אם הסמן אינו מצביע על קודקוד, הרשימה ריקה). במידע זה נעזרים ציירי הקודקודים והקשתות, כפי שמתואר בסעיף הבא.

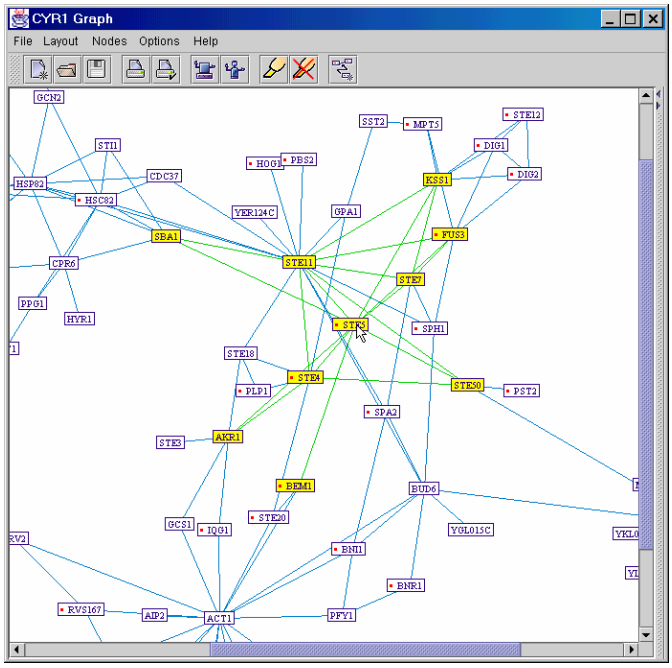
4.5.3.6 ציירי הקשתות והקודקודים

הקלאס GraphCanvas מאפשר לאפליקציה לספק לו NodePainter, ו/או EdgePainter, על מנת לצייר את הגרף באופן שונה. בניגוד לציירים הכלליים שמספקת החבילה לטיפול בגרף, יכולים קלאסים אלו לעשות שימוש במידע ספציפי הקיים באפליקציה זו.

הקלאס NameNodePainter משמש לציור קודקודי הגרף. קלאס זה דואג להציג כל קודקוד תוך שימוש בשם החלבון, אותו הוא מייצג. כמו כן, הוא מבצע שאילתה מול מאגר המידע, על מנת לבדוק, האם ישנם חלבונים שלהם אינטראקציה עם החלבון אותו הוא מצייר ואשר אינם מוצגים בגרף. אם קיימים חלבונים כאלו, יוצג החלבון, כאשר נקודה אדומה לצדו. הנקודה מציינת עבור המשתמש לאילו חלבונים יש אינטראקציות נוספות, אשר אינן מוצגות.

הצייר אחראי לספק גם את הטקסט ל-tooltips אשר מוצגים לכל קודקוד. לשם כך פונה הקלאס NameNodePainter אל מנהל מאגרי המידע ומבקש לדעת את שם ההומולוג האנושי לחלבון השמרי המוצג. כאשר המשתמש מתעכב עם סמן העכבר על חלבון בגרף למשך כשנייה ומעלה, יוצג לו שם החלבון ההומולוגי באדם (ראה איור 24).

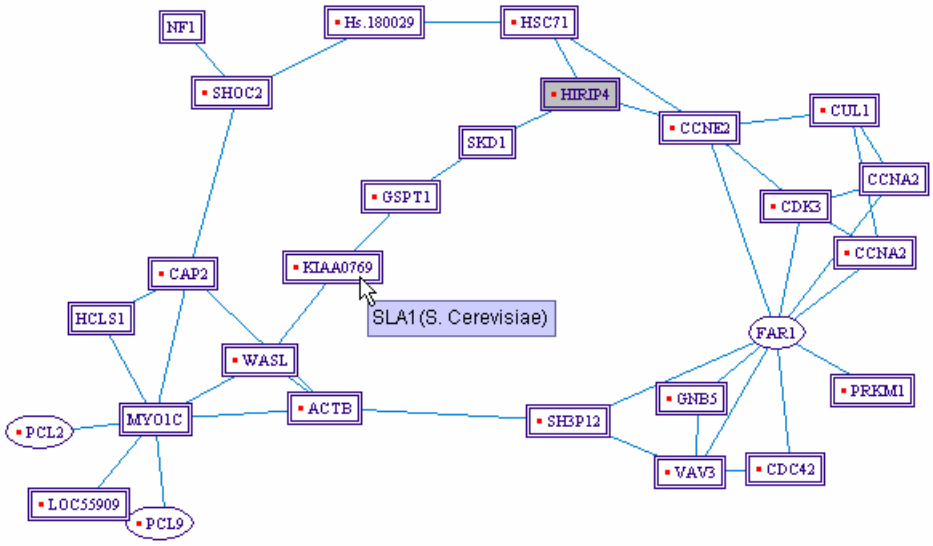
אל צייר הקודקודים NameNodePainter מצטרף גם צייר הקשתות LineEdgePainter, אשר מצייר את הקשתות בין קודקודים סמוכים בצורת קו ישר. שניהם יודעים להתייחס אל סמן השכנים (NeighborsMarker), אשר תואר לעיל, ואשר מנהל רשימה מעודכנת של קודקודים "מסומנים". הציירים יודעים להתייחס לרשימת קודקודים זו, ולצבוע בצבע שונה את הקשתות והקודקודים המסומנים. כך, למשל, יכול המשתמש להצביע על קודקוד מסוים בגרף, ותת-הגרף, המכיל אותו ואת כל שכניו, ייצבע בצבע בהיר (ראה איור 25). כלי זה מקל מאוד את ההתמצאות באזורים צפופים של הגרף.



איור 25. השימוש בסמן השכנים

ציירי הקודקודים והקשתות נעזרים בסמן השכנים כדי להדגיש בצבעים בהירים את סביבתו הקרובה של החלבון, עליו מצביע המשתמש.

צייר קודקודים נוסף, בו נעשה שימוש בתוכנה זו הוא ה- HomologNameNodePainter. צייר זה פונה אף הוא למאגר המידע ומבקש את שם ההומולוג האנושי של החלבון, אותו הוא מצייר, ומשתמש בשם זה על מנת להציג את החלבון. כך מקבל המשתמש גרף, בו מוצגות האינטראקציות הלקוחות מהשמר, אך החלבונים מוצגים בשמותיהם באדם. לשם הבדלה, מוקף כל צומת במסגרת מלבנית כפולה. חלבון, אשר ההומולוג האנושי עבורו אינו מוכר, יוצג בשמו השמרי, אך יסומן במסגרת בצורת אליפסה כדי להבדילו מהחלבונים האחרים. צייר זה מספק tooltips מתאימים, אשר יציגו הפעם את שמו של חלבון השמר המקורי המיוצג ע"י הקודקוד (ראה איור 26).



איור 26. תצוגת החלבונים ההומולוגים באדם

שימוש בצייר קודקודים מתאים מאפשר להחליף את שמות החלבונים המוצגים בשמות החלבונים ההומולוגים להם באדם. חלבונים, אשר ההומולוג האנושי שלהם אינו מוכר, מוצגים בשמם השמרי ומסומנים ע"י אליפסה. האינטראקציות המופיעות בגרף הן האינטראקציות השמריות, אך הן מוצגות כעת כאינטראקציות היפותטיות באדם, אשר אותן יש לאמת במעבדה בשיטות ניסוייות.

4.5.3.7 אזור תמונת הלווין – SatelliteView Panel

אזור זה מנוהל ע״י אובייקט מהקלאס SatelliteView, אשר תואר לעיל. קלאס זה מהווה Listener למודל הגרף הנמצא במנהל הגרף. אזור התצוגה הראשי מקושר אף הוא לאותו מודל, ולכן כל שינוי שיתבצע דרכו במודל, יוצג באופן מיידי גם בתמונת הלווין.

שני אזורי המסך המציגים את הגרף – SatelliteView Panel, ו- GraphPane, אינם מקושרים זה לזה, אלא למודל הגרף בלבד. הקישור אל מודל הגרף הוא באחריות הקלאס AppWindow, אשר דואג גם לשנותו בעת החלפת המודל עליו עובדים (במעבר לעבודה על גרף חדש).

4.5.3.8 DescriptionView

קלאס זה מנהל את חלקו הימני התחתון של המסך. באזור זה מוצג עבור המשתמש תיאור תמציתי של החלבון, על מנת לסייע לו להתמצא בגרף ולהכיר את החלבונים בו.

כדי להציג למשתמש תיאור של החלבון עליו הוא מצביע בכל שלב, נבנה קלאס זה כ-NodeListener, אשר מאזין לקלאס GraphController (המטפל כזכור בהתרחשויות

ב-GraphPane - ראה סעיף 4.2.3). בכל פעם שהקלאס DescriptionView מקבל מאורע מסוג NodeEntered, הוא פונה אל מנהל מאגרי המידע, מבקש לקבל שורת תיאור לגבי החלבון המתאים, ומציג את התיאור הנ"ל למשתמש.

4.5.3.9 מסכים נוספים באפליקציה – DialogsHandler

פרט לחלון הראשי המוצג למשתמש, מציעה התוכנה חלונות עזר רבים נוספים, המשמשים לקבלת קלט מהמשתמש ולהצגת הודעות מיוחדות, ביניהם:

- בחירת שם קובץ – לצורך שמירת הקובץ או פתיחתו של קובץ ישן.
- התראה על שינויים בגרף שלא נשמרו – לפני סגירת התוכנה או פתיחת גרף חדש.
- ביצוע הדפסה, הגדרות הדפסה, התראה על ביטול הדפסה לפני סיומה ביציאה מהתוכנה.
- שאילתות פרישת שכנים, ומציאת מסלול מינימאלי – לקבלת קלט המשתמש, ולהצגת התקדמות התהליך.
- הודעות למשתמש – תקלה בפתיחת קובץ או בשמירתו, אי הצלחה בחיפוש מסלול מינימאלי, וכיו"ב.

הטיפול בכל הדיאלוגים בתוכנה מרוכז בקלאס DialogsHandler, אשר דואג לבנותם, להציגם ולהעביר את הקלט לרכיבי התוכנה המתאימים. בכך הוא מסייע ל-ViewManager, אשר מעביר ישירות אליו את כל הבקשות הכרוכות בהצגת דיאלוגים למשתמש.

4.5.3.10 פתיחת דפי מידע באינטרנט – BrowserHandler

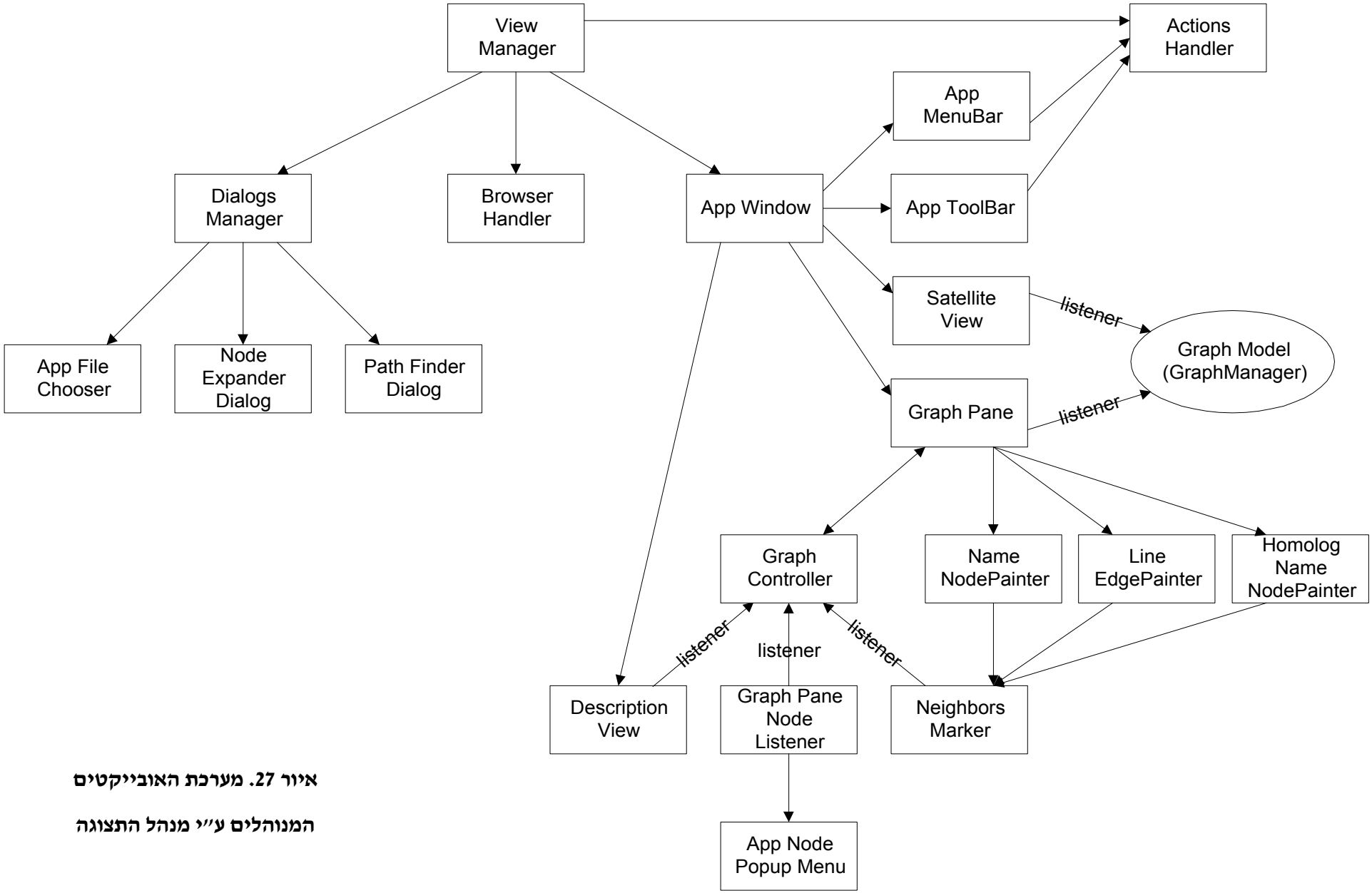
קלאס זה אחראי על הטיפול בקישוריות לדפדפן האינטרנט (Internet browser). כאשר יש צורך בפתיחת דף אינטרנט כלשהו, מקבל קלאס זה את כתובת הדף המתאים, ודואג להעבירה לדפדפן המותקן על המחשב בו משתמשים.

קלאס זה הנו בעייתי מבחינת התפיסה של Java, אשר משתדלת להיות Platform independent. ניתן היה להשתמש בכלים של השפה בלבד, על מנת לבנות קלאס פשוט שיציג דף Html. אך השאיפה שלי הייתה לפתוח את דפי האינטרנט באותו דפדפן אשר בו רגיל המשתמש לעבוד, ואשר דרכו הוא יכול להמשיך ולפנות לאתרים נוספים, לגשת אל ה-bookmarks שלו, וכו'. לשם כך, יש צורך לפנות אל תוכנה חיצונית, המותקנת על מחשב המשתמש, והדרך לבצע פנייה זו תלויה במערכת ההפעלה המותקנת.

לאחר חיפושים רבים באינטרנט, פניה ל- newsgroups בנושא, והתכתבויות עם מספר תוכניתני Java, הבנתי כי תוכניתנים אחרים כבר נתקלו בבעיה לפני, וכי אין לה פתרון "נקי". לכן נאלצתי ליישם את השיטה המוצעת – לזהות את מערכת ההפעלה המותקנת ע"פ משתני סביבה, ולשלוח אליה את הפקודה המתאימה לה לצורך הפעלת הדפדפן.

החיסרון של שיטה זו הוא בכך, שהיא מתייחסת למערכות הפעלה ספציפיות, ולכן יהיה צורך לבצע שינויים בקוד להתאמתו למערכות הפעלה נוספות (למשל, הפקודה הניתנת למערכת Win95 אינה פועלת במערכת Win98). כמו כן, הפקודה אשר יש לשלוח אל מערכת ההפעלה, אינה מתועדת בכל מחשב בצורה ברורה, וקשה לחזות את הצלחתה על מערכות מחשב שונות.

רכיב זה של התוכנה הוא בעייתי, וכרגע אין פתרון סטנדרטי לבעיה. בניתי אותו כקלאס נפרד ובעל ממשק פשוט, כך שכל שינוי עתידי בתחום זה יתרכז כולו במקום יחיד.



איור 27. מערכת האובייקטים

המנוהלים ע"י מנהל התצוגה

4.5.4 <u>מנהל העימוד</u>

מנהל העימוד הוא המנהל הפשוט ביותר. תפקידו לדאוג להפעלת העימוד האוטומטי של הגרף ולהפסקתו בעת הצורך, ולקשר את ה- thread המבצע את העימוד אל מודל הגרף הנמצא במנהל הגרף.

בנייתו של רכיב זה כמנהל עצמאי נועדה לאפשר הרחבות עתידיות הנוגעות לעימוד, כגון בחירת אלגוריתם עימוד אחד מתוך מגוון אפשרויות, טיפול במספר גרפים במקביל, וכדומה.

4.5.5 <u>מנהל התקני מערכת</u>

תפקידו של מנהל זה הוא לספק תמיכה בפעולות שקשורות לרכיבים של מערכת המחשב, ובפרט מערכת הקבצים, והמדפסת.

4.5.5.1 הטיפול במערכת הקבצים - FilesHandler

קלאס זה אחראי על קישור התוכנה אל מערכת הקבצים. הטיפול בגרף, מרגע יצירתו ועד לסגירתו, והקישור בין הגרף לבין הקובץ בו הוא נשמר, מנוטרים ע"י קלאס זה. בין תפקידיו:

- הצגת הדיאלוגים המתאימים וההודעות למשתמש בעת יצירת גרף חדש, פתיחת גרף קיים, שמירת הגרף, וכו' (בסיוע מנהל התצוגה).
- טיפול בגישה אל מערכת הקבצים, ואיתור הקובץ הרצוי לקריאה או לכתיבה.
- ביצוע תהליכי הקריאה מהקובץ או הרישום אליו החל מפתיחתו ועד לסגירתו המסודרת.
- אתחול המערכת לעבודה על גרף חדש, לאחר יצירת גרף חדש או טעינת גרף מקובץ.
- התראה בפני המשתמש על סגירת גרף ללא שמירתו (במידה ונעשו בו שינויים).

קלאס זה נעזר בקלאס GraphChangesMonitor, אשר יודע לרשום את עצמו בגרף הפעיל בתור Listener, מאזין לשינויים בגרף, ומבחין בין שינויים, אשר נרשמים לקובץ (כגון הוספת קודקודים) לבין אלו, אשר אינם נרשמים (כגון בחירה של קודקוד בגרף או נעילת מיקומו). ברגע שקלאס זה מזהה שינוי, הוא מסיר את עצמו מרשימת המאזינים (כדי לחסוך בקריאות מיותרות אליו), ודואג לדווח בעת הצורך ל- FilesHandler, כי הגרף השתנה.

4.5.5.2 הטיפול בעבודות הדפסה

הקלאס PrintHandler דואג לטיפול בכל הקשור להדפסת הגרף. בין תפקידיו:

- הצגת דיאלוג ההדפסה והגדרת העמוד למשתמש.
- יצירת עבודות ההדפסה ושליחתן למדפסת ב-thread-ים נפרדים (הדפסה ברקע).
- ניהול מעקב אחרי עבודות ההדפסה שנשלחו עד לסיומן, או ביטולן עם היציאה מהתוכנית.

לצורך הדפסת הגרף הוא נעזר בקלאס GraphPrinter, אשר תואר לעיל (סעיף 4.2.4). הוא עוטף אותו בקלאס נוסף בשם GraphPageable, אשר מאפשר למשתמש לקבל מידע לגבי מספר העמודים בהדפסה, ולהדפיס רק חלק מעמודי הגרף.

בכדי למנוע את הפסקת העבודה של המשתמש למשך זמן ההדפסה, ה- PrintHandler דואג לבצע את ההדפסות ברקע. מיד עם ההקשה על כפתור ההדפסה, נוצר אובייקט מהקלאס GraphPrinter, אשר משכפל את מודל הגרף במצבו הנוכחי, כך ששינויים שיתרחשו מעתה בגרף לא יבואו לביטוי בהדפסה זו. עם אישורה של עבודת הדפסה, נוצר אובייקט מהקלאס PrintingThread, אשר יוצר thread חדש ממנו תתבצע ההדפסה. thread זה פועל בעדיפות נמוכה, כך שלא יפריע למשתמש בהמשך העבודה עם התוכנה. המשתמש יכול להמשיך ולעבוד ב-thread הראשי, ולשנות את הגרף מבלי לפגוע בהדפסה המתבצעת.

עם יצירתו של thread ההדפסה, הוא נרשם באובייקט מהקלאס PrintingThreadsObserver. קלאס זה מנהל מעקב אחרי ה-thread-ים השונים, ומקבל הודעות על ביטולם או על סיום פעולתם. בעזרת אובייקט זה, יכול ה- PrintHandler לוודא כי לא נותרו עבודות הדפסה פעילות עם היציאה מהתוכנית. אם נותרו כאלו, יתאפשר למשתמש לבטל את כל עבודות ההדפסה שעדיין לא הסתיימו, או לבטל את היציאה מהתוכנית ולהמתין לסיומן.

5 PIVOT ככלי בידי החוקר

מטרתה של התוכנה היא לסייע לחוקר של חלבון חדש, להעלות השערות מושכלות לגבי תפקידו של החלבון בתא ולגבי קשריו עם חלבונים נוספים, מתוך התבוננות באינטראקציות מוכרות בין החלבונים ההומולוגים לחלבון החדש באורגניזמים אחרים. מצב הידע הנוכחי על אינטראקציות בין חלבונים מאפשר טיפול בעזרת PIVOT בעיקר בחלבונים אנושיים או שמריים, אם כי התוכנה היא כללית באופייה. לצורך ההסבר נתרכז בחקירת חלבון אנושי חדש.

בעבר הלא רחוק, התרכז המחקר הגנטי בעיקר בתכונות מונוגניות. תחילת המחקר הייתה בפנוטיפ חריג, ונעשה ניסיון לזהות את הגן האחראי לו בעזרת שיבוט איתורי. זיהוי הגן אפשר להמשיך במחקר לגבי תפקידו של החלבון אותו הוא מקודד.

אל כלי המחקר אשר היו נהוגים בעבר, הצטרף לאחרונה הרצף הגנומי, ועמו אפשרויות מחקר חדשות. כיום רצף הגנום נגיש לחוקרים, וכמוהו תוכנות שונות לחיזוי גנים. כמו כן, רצפי Expressed Sequence Tags (ESTs) רבים הוארכו עד לקבלת הרצף המלא של ה-cDNA המתאים ונוצר מאגר של רצפים מקודדים אשר לא ידוע דבר לגבי תפקידם. ענף המחקר העוסק בהתאמת תפקידו של חלבון לרצף שלו נקרא Functional Genomics.

שיטות שונות מנסות לחזות את תפקידו של החלבון מתוך הרצף שלו, ביניהן חיפוש אזורים פעילים מוכרים בחלבון (Domain Analysis), ניסיונות לחיזוי מבנה החלבון, ועוד (ראה "שיטות לחיזוי פונקציה של חלבון", בסעיף 2.2). אחת הדרכים ה"חזקות" ביותר ללימוד על חלבון חדש, היא מתוך החלבון ההומולוגי שלו באורגניזמים אחרים.

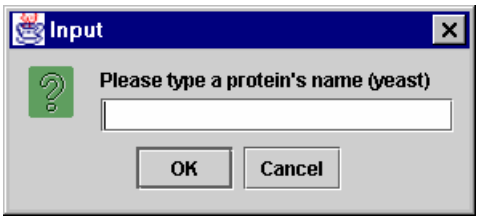
המחקר העוסק באדם מוגבל מאוד, הן מסיבות אתיות, והן מסיבות מעשיות. לאורגניזמים פשוטים (כגון השמר *S. cerevisiae*), יתרונות רבים כאובייקטים למחקר, הנובעים מפשטותם ביולוגית, גודלם הקטן, קצב ההתרבות המהיר שלהם, יכולתו של החוקר לבצע בהם סלקציות, זיווגים רצויים, מניפולציות גנטיות שונות, ועוד. היקף המחקר והידע הקיימים באורגניזמים אלו נרחבים מאוד.

כאשר תפקידו של החלבון ההומולוגי מוכר, נוכל להסיק רבות לגבי החלבון אותו אנו חוקרים. אך גם אם תפקידו של החלבון ההומולוגי אינו מוכר, ניתן לנסות ולהיעזר בידע הקיים לגבי קשריו של חלבון זה עם חלבונים נוספים. שיוך פונקציונאלי של החלבון לקבוצת חלבונים על פי האינטראקציות שלו עמם עשוי לסייע מאוד בזיהוי תפקידו.

פרק זה בעבודה מתאר את התוכנה ואת אופן השימוש בה, תוך שילוב הסברים לגבי הצורך בתכונות השונות שנבנו.

5.1.1 יצירת גרף לעבודה על חלבון חדש

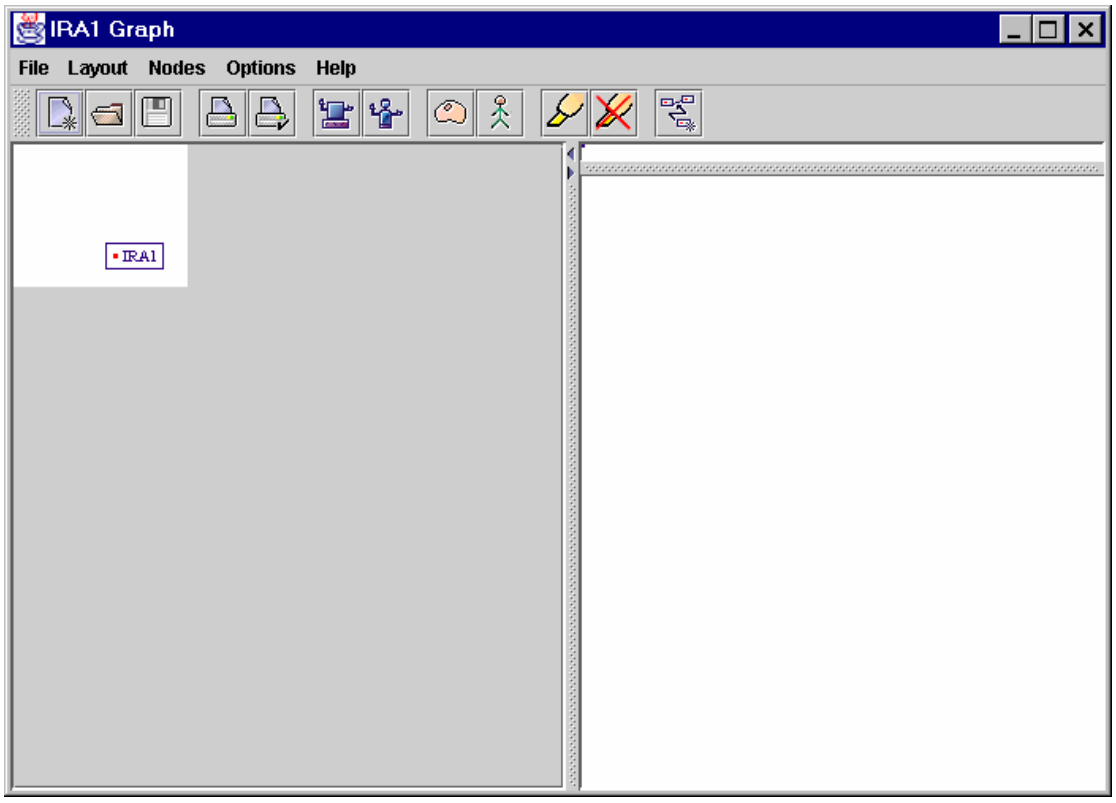
עם הרצת התוכנה, מוצג החלון הראשי של PIVOT כאשר הוא ריק (אינו מכיל גרף). לתחילת העבודה יש לבחור מהתפריט File את האפשרות New, או לחילופין להקיש על הצלמית  המסמלת אותה פעולה. כתוצאה מכך, יוצג החלון המופיע באיור 28, דרכו יש להזין את שם החלבון השמרי שבו נשתמש כנקודת מוצא.



איור 28. קליטת שמו של חלבון חדש

בהמשך, אשתמש בחלבון השמרי IRA1 לצורך ההדגמה.

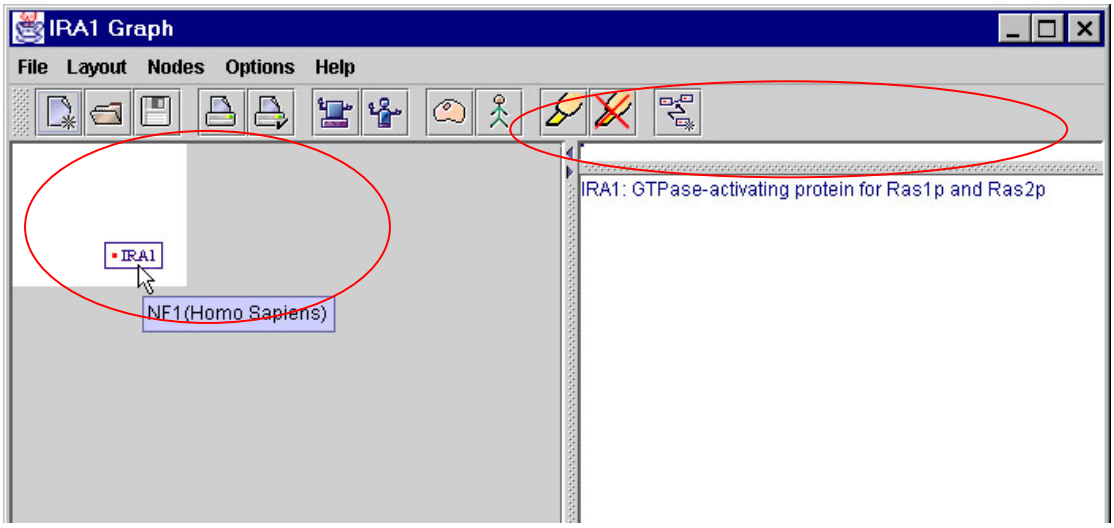
לאחר הזנת שם החלבון השמרי וההקשה על הכפתור OK, ייפתח עבורנו גרף חדש, המכיל את החלבון המבוקש בלבד (ראה איור 29). החלבון מופיע בגרף, כאשר לצידו נקודה אדומה. נקודה זו מציינת, כי יש לחלבון זה אינטראקציות עם חלבונים נוספים, אשר עדיין אינם מופיעים בגרף.



איור 29. מסך העבודה הראשי בתחילת העבודה על החלבון השמרי IRA1

5.1.2 קבלת מידע על החלבון

עוד בטרם הוספת חלבונים נוספים אל הגרף, נוכל לקבל פרטים נוספים על החלבון המשמש כנקודת המוצא. אם נצביע עליו עם סמן העכבר, תופיע בצד ימין של המסך שורת תיאור לגבי חלבון זה. אם נשאיר את הסמן על החלבון למספר שניות, יופיע לידו גם שמו של ההומולוג האנושי שלו (ראה איור 30).



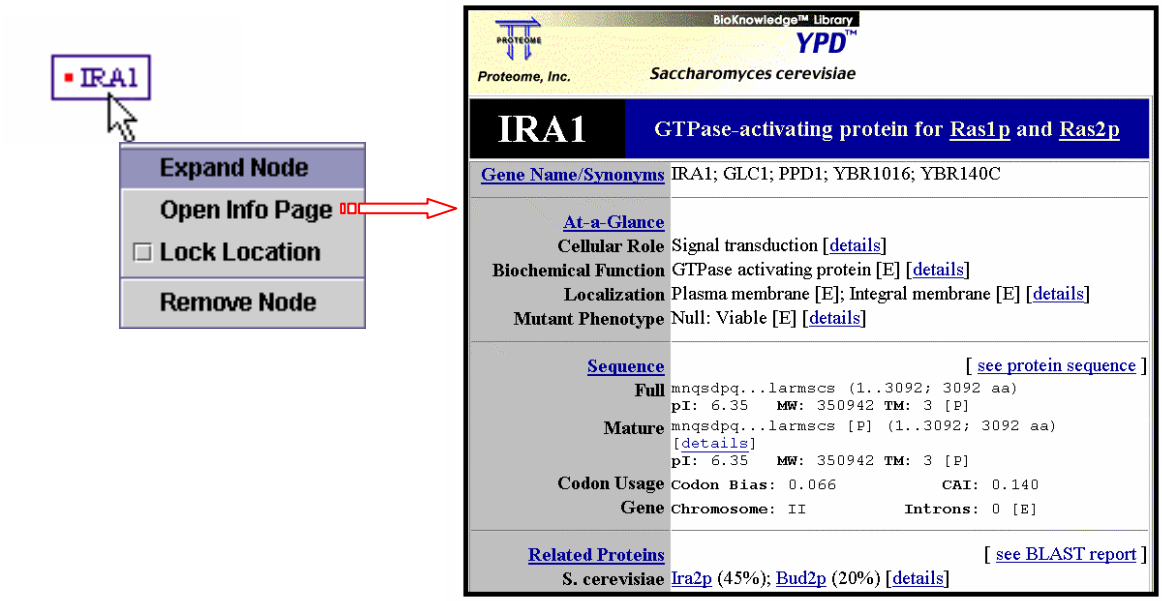
איור 30. מידע בסיסי, המוצג בחלון התוכנה הראשי לכל חלבון

אם נרצה לפתוח את דף המידע המלא לחלבון המוצג בגרף, נציב את סמן העכבר על החלבון, וע"י הקשה על לחצן העכבר הימני, נקבל תפריט עזר, אשר מתוכו נוכל לבחור באפשרות Open Info Page. בחירה זו תביא לפתיחת דף המידע של החלבון מאתר האינטרנט של חברת Proteome[4] Inc., בדפדפן האינטרנט המותקן על המחשב בו אנו משתמשים (ראה איור 31).

הערה: אם בפתיחת הגרף החדש, טעה המשתמש והקליד שם של חלבון, אשר אינו קיים במאגר המידע, עדיין ייפתח גרף, המכיל חלבון יחיד בשם שהוקלד. אך ל"חלבון" זה לא תהיינה אינטראקציות ולכן לא תופיע לידו נקודה אדומה. כמו כן, אם נצביע עליו עם סמן העכבר, לא יופיע תיאור עבורו בצד ימין של המסך, ולא יוצג עבורו שם של הומולוג אנושי ידוע.

---

[4] חברת Proteome נרכשה לאחרונה ע"י חברת Incyte, והגישה אל דפי המידע באתר האינטרנט של החברה נחסמה לשימוש שלא בתשלום. ניסיון מצד המשתמש לפתוח דף מידע מהרשת מתוך PIVOT, יביא להפעלת הדפדפן ולהפנייתו אל כתובת ברשת אשר בעבר סיפקה את המידע הדרוש, אך כיום הינה חסומה.



איור 31. פתיחת דף המידע באינטרנט עבור חלבון בגרף

5.1.3 הוספת השכנים

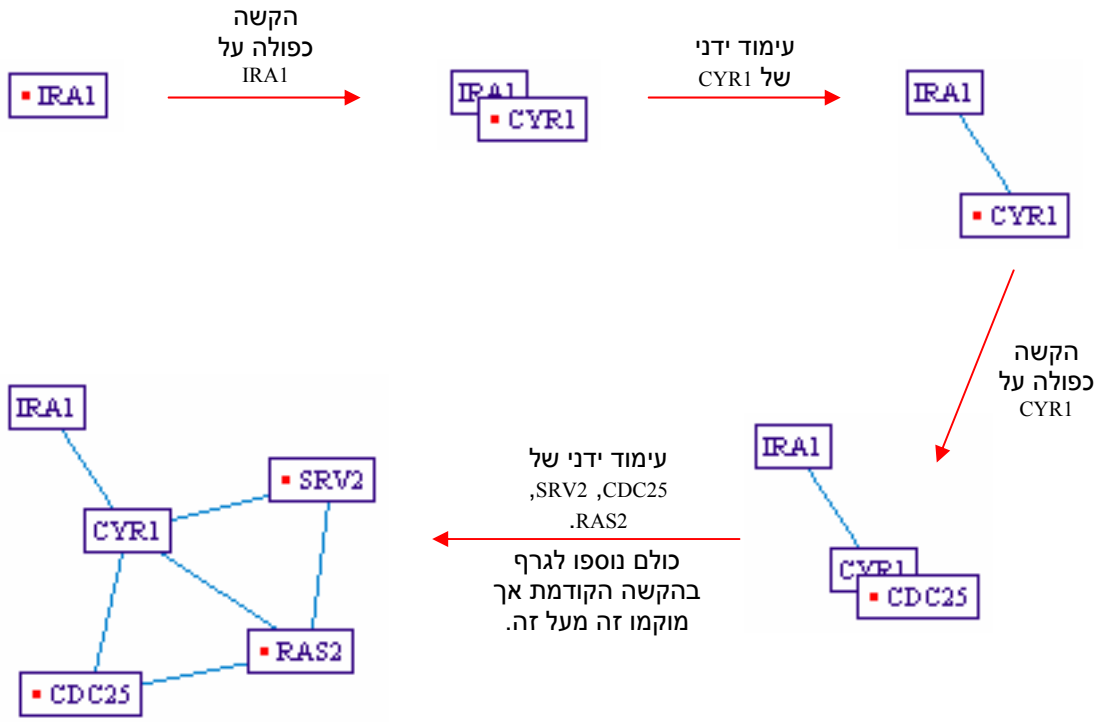
כדי להוסיף אל הגרף את שכניו של החלבון המסומן בנקודה אדומה, יש להצביע על החלבון ולהקיש הקשה כפולה על הלחצן השמאלי של העכבר (double-click). עקב כך, תבוצע שאילתה שתביא לשליפת כל החלבונים, אשר להם אינטראקציה עם החלבון עליו לחצנו, ולהוספתם אל הגרף. הנקודה האדומה שלצד החלבון עליו לחצנו, תיעלם.

כאשר מתווסף חלבון חדש אל הגרף, הוא מקושר בקשתות אל כל החלבונים המופיעים בגרף, אשר איתם הוא מצוי באינטראקציה, ולא רק אל החלבון אשר עליו לחצנו.

חלבוני הגרף נוספים תמיד במיקום סמוך לקודקוד עליו לחצנו. אם לא נעשה שימוש במנגנון העימוד האוטומטי (ראה להלן), יופיעו החלבונים החדשים זה מעל זה, ועל המשתמש יהיה להזיזם בעצמו אל המיקום בגרף אשר הוא בוחר (איור 32).

5.1.4 הפעלת מנגנון העימוד האוטומטי

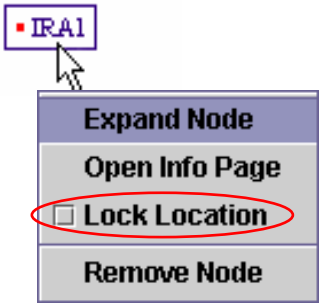
לשם נוחות העבודה עם הגרף, מומלץ להפעיל את מנגנון העימוד האוטומטי. רכיב זה דואג לשנות את מיקומי הקודקודים בהתאם לקשתות המחברות אותם, כדי להשיג פרישה נוחה של הגרף. על מנת להפעילו, יש לבחור מתוך התפריט Layout באפשרות Start Auto Layout, או לחילופין להשתמש בצלמית .



איור 32. פרישת קודקודי הגרף הסמוכים

הקשה כפולה על קודקוד, שבצידו נקודה אדומה, תביא להוספת כל שכניו אל הגרף ולעדכון כל האינטראקציות שלהם עם חלבוני הגרף האחרים. הקודקודים החדשים ממוקמים זה מעל זה, והמשתמש יכול למקם אותם כרצונו ע"י גרירתם על פני הגרף.

המשתמש יכול להתערב בפעולת המנגנון ולגרור קודקוד רצוי בעזרת העכבר לאזור שונה של הגרף (תוך החזקת לחצן העכבר לחוץ). עם שחרור לחצן העכבר, חוזר הקודקוד לאחריות מנגנון העימוד כדי שימקמו באופן אופטימאלי באזור החדש בו הוא הונח, ובהתייחס לקשתות, אליהן הוא קשור. אם נרצה למנוע ממנגנון העימוד מלהתערב בקביעת מיקומו של קודקוד מסוים, נוכל לנעול את הקודקוד במקומו ע"י פתיחת תפריט העזר (לחצן ימני), ובחירה באפשרות Lock Location (איור 33). הקודקוד ייצבע באדום, ומיקומו לא ישתנה, אלא בהתערבות המשתמש. שחרור הנעילה של קודקוד נעשית באותו אופן (הקשה על הלחצן הימני של העכבר לפתיחת תפריט העזר ובחירת האפשרות Lock Location).



איור 33. נעילת מיקומו של קודקוד

אם המשתמש רוצה להפסיק לחלוטין את פעולת העימוד האוטומטי, עליו לבחור מהתפריט Layout באפשרות Stop Auto Layout, או לבחור בצלמית .

5.1.5 מחיקת קודקודים

המשתמש יכול למחוק מהגרף קודקוד, אשר אינו מעניין אותו, ע״י פתיחת תפריט העזר שלו (לחצן ימני של העכבר), ובחירה באופציה Remove Node. הקודקוד יימחק מהגרף עם כל האינטראקציות הקשורות אליו, אך יחזור לגרף בשנית במידה וייקרא ע״י שאילתה כלשהי (למשל double-click על אחד משכניו לגרף).

5.1.6 טיפול במספר קודקודים יחדיו

המשתמש יכול לסמן קבוצה מתוך קודקודי הגרף ולבצע פעולה על כל הקודקודים המסומנים יחדיו. כך ניתן לנעול במקומו אזור שלם בגרף, למחוק מהגרף קבוצה גדולה של קודקודים, או אף לפתוח את שכניהם של כל הקודקודים המסומנים בפעולה אחת.

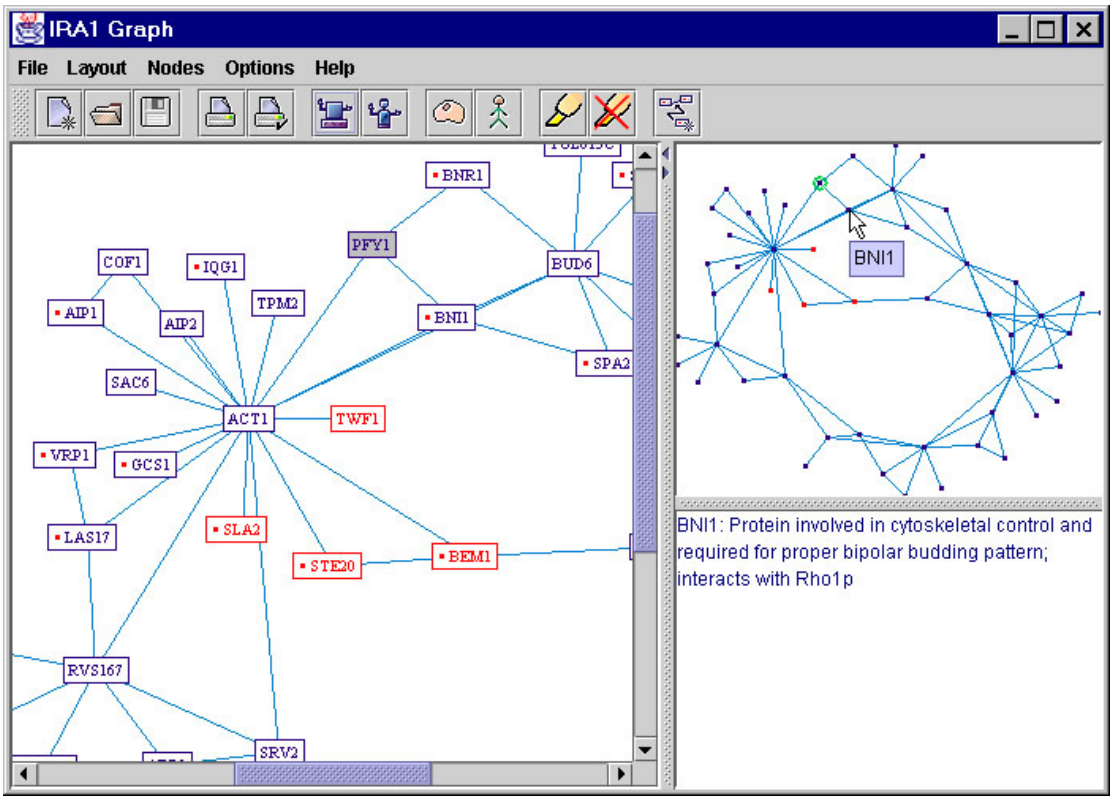
כדי לסמן קבוצת קודקודים יש להשתמש בלחצן העכבר השמאלי על מנת לצייר מסגרת, המקיפה אותה. ניתן להוסיף קודקודים נוספים לאלו המסומנים ע״י החזקת המקש Ctrl לחוץ בעת סימון קודקודים נוספים (כבודדים או כקבוצה ע״י מסגרת). בעת החזקת המקש Ctrl, ניתן גם ללחוץ על קודקודים מסומנים על מנת להסיר מהם את הסימון.

לאחר סימון הקודקודים, ניתן לבחור מהתפריט Nodes מהי הפעולה שברצוננו לבצע על כל הקודקודים בקבוצה – לפרוש את שכניהם (Expand), לפתוח עבורם את דפי המידע שלהם באינטרנט (Open Info Page), לנעול אותם במקומם, לשחרר את נעילתם (Lock/Unlock Location), או למחקם מהגרף (Remove).

5.1.7 שימוש ב״תמונת הלווין״

תמונת הלווין נמצאת בצידו הימני של המסך, מעל לאזור תיאור החלבון. תמונה זו תציג תמיד ״מבט-על״ של הגרף כולו, ובעזרתה קל יותר להתמצא בגרף, כאשר הוא גדל אל מעבר לגבולות המסך. תמונה זו תמלא תמיד את כל השטח המוקצה עבורה, וניתן להגדילה ע״י הגדלת שטח המסגרת שלה (הזזת קווי ההפרדה).

הקודקודים אשר נעולים במקומם, יופיעו בתמונה זו בצבע אדום, ואילו הקודקודים שהמשתמש בחר, יסומנו בעיגול ירוק. על מנת לראות מהו החלבון אותו מייצגת נקודה כלשהי בגרף זה, ניתן להצביע על הנקודה עם סמן העכבר במשך כשניה, ושם החלבון המתאים יופיע לידה (ראה איור 34).

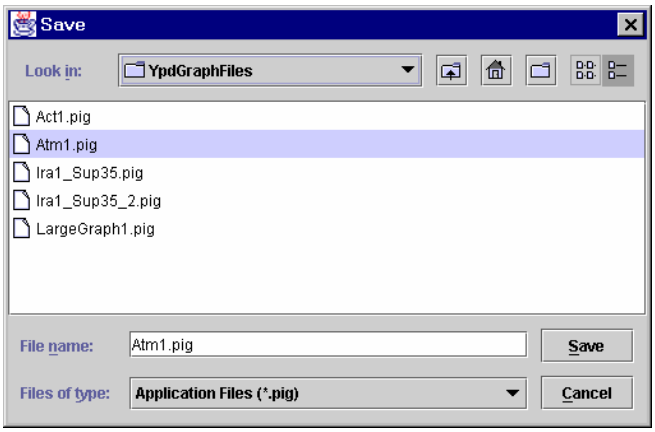


איור 34. תמונת הלווין

תמונת הלווין המופיעה בצידו הימני העליון של המסך, מציגה תמיד את הגרף כולו, ומסייעת למשתמש להתמצא בגרף גדול, אשר חורג מגבולות אזור התצוגה המרכזי.

5.1.8 שמירה וטעינה של גרף

לאחר סיום העבודה על הגרף, ניתן לשמרו לקובץ, ע"י פתיחת התפריט File ובחירה באפשרות Save, או בצלמית  . אם גרף זה נשמר לקובץ בעבר, הוא יישמר תחת אותו שם. אחרת, ייפתח עבורנו דיאלוג, אשר יציג את מערכת הקבצים במחשב ויאפשר למשתמש לבחור היכן ובאיזה שם לשמור את הגרף.



איור 35. דיאלוג שמירת קובץ

כדי לטעון לתוכנה גרף אשר נשמר בעבר, נשתמש באפשרות Open שבתפריט File, או בצלמית  . אם

עבדנו על גרף כלשהו ואנו מנסים לטעון גרף אחר, ייסגר הגרף הישן, והגרף החדש ייטען במקומו. בכל מקרה של טעינת גרף מקובץ, תופסק פעילותו של מנגנון העימוד האוטומטי כדי לקבל את הגרף על המסך, כפי שנשמר ובלי שיתחיל לנוע. מובן, שהמשתמש יכול לבחור להפעיל שוב את מנגנון העימוד.

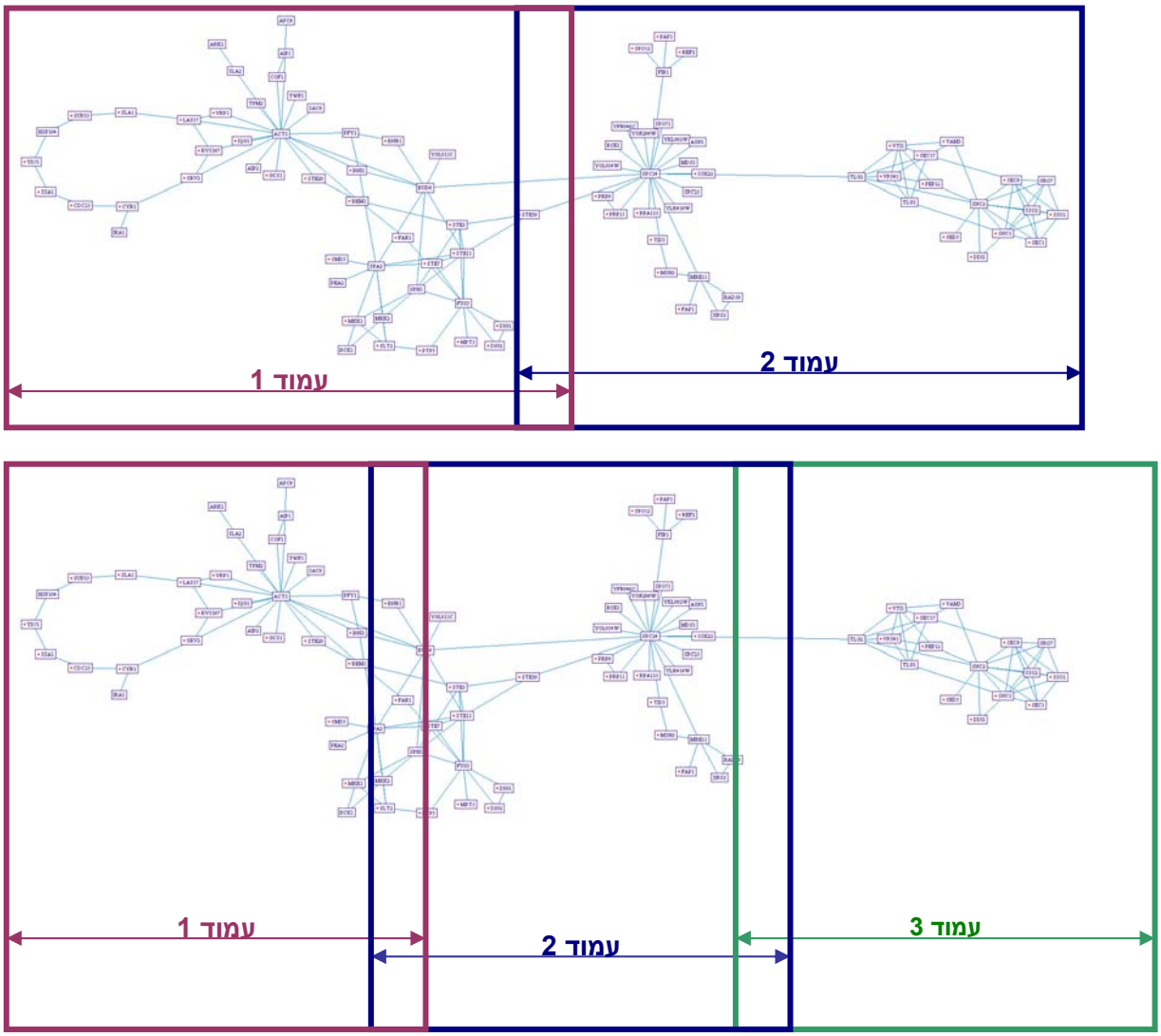
בכל סגירה של הגרף עליו אנו עובדים, אם בגלל היציאה מהתוכנה, ואם בגלל מעבר לעבודה על גרף אחר (פתיחת חדש או טעינה מקובץ), תתריע התוכנה על שינויים שבוצעו בגרף ולא נשמרו. נוכל לשמור את השינויים, להתעלם מהם, או לבטל את הפעולה שבעקבותיה ייסגר הגרף.

ניתן כמובן לשמור גרף גם בשם שונה מזה שבו הוא נשמר בעבר, ע"י שימוש באפשרות Save As שבתפריט File.

5.1.9 <u>הדפסת הגרף</u>

לפני הדפסת הגרף עליו עובדים, יש להגדיר את גודל הדף במדפסת ואת כיוונו. ללא הגדרות אלו יודפס הגרף תוך שימוש בערכי ברירת המחדל. מכיוון שבדרך כלל, הגרף גדול ומודפס על פני מספר עמודים, מומלץ לנסות ולבחור בכיוון דף (Portrait או Landscape), אשר ישתמש בפחות דפים להדפסה (ראה איור 36).

לצורך כיוון הגדרות ההדפסה, יש לבחור מהתפריט File באפשרות Page Setup, או ללחוץ על הצלמית  . להדפסת הגרף, יש לבחור מהתפריט File באפשרות Print, או ללחוץ על הצלמית . דיאלוגים מתאימים ייפתחו ויאפשרו את כיוון הפרמטרים ואת אישור ההדפסה.



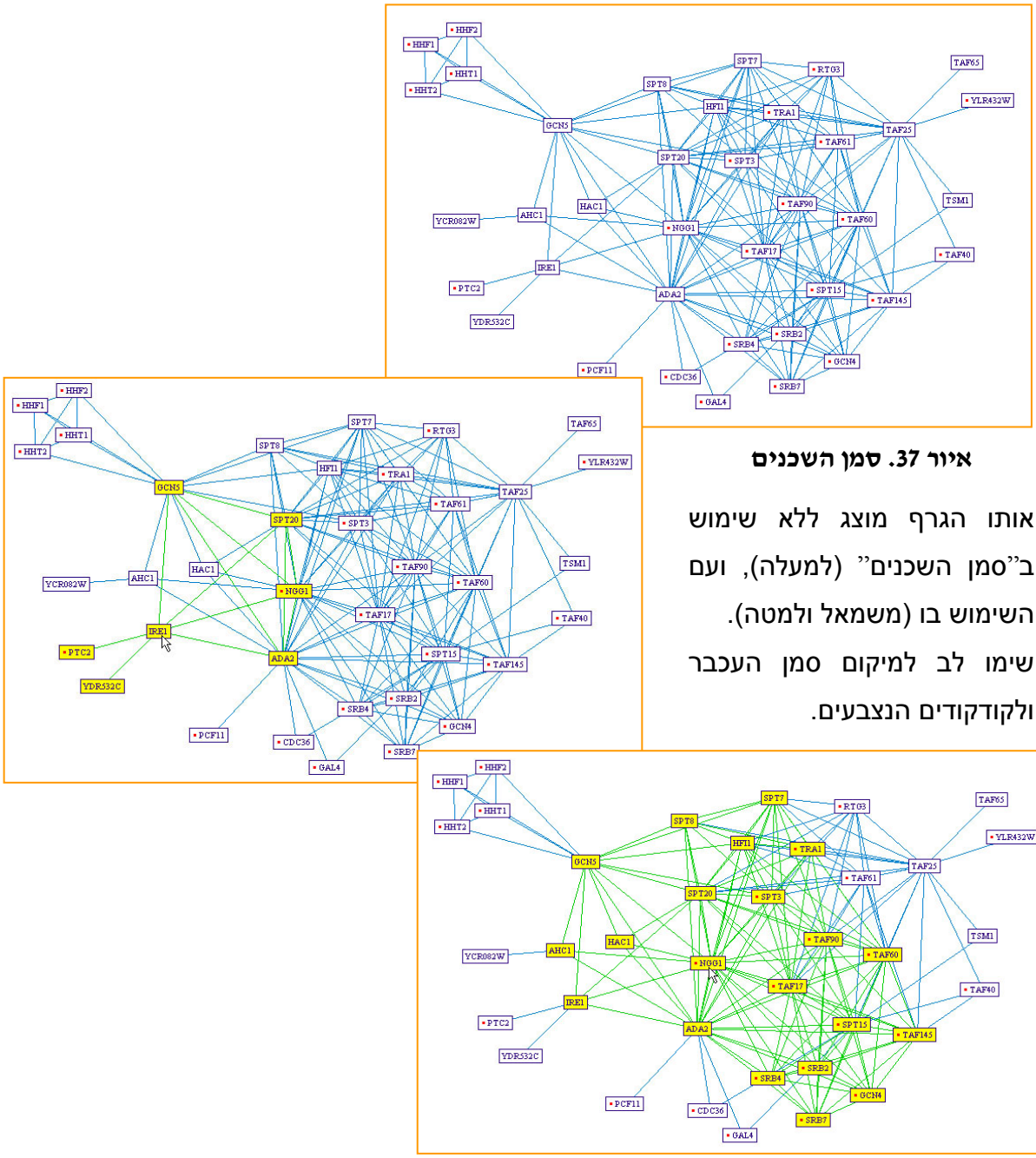
איור 36. חלוקת הגרף לדפים לצורך הדפסתו

הדפסתו של אותו הגרף ב-Landscape (חלק עליון בתמונה), וב-Portrait (חלק תחתון).
אזורי החפיפה בין עמודים סמוכים מקלים על הצמדת הדפים ליריעה גדולה אחת.

5.1.10 <u>השימוש ב"סמן השכנים" לצורך סריקה נוחה של הגרף</u>

מטרתו של "סמן השכנים" היא לסייע לחוקר בהתמצאות באזורים צפופים של הגרף, בהם קשתות רבות חוצות זו את זו, ואף עוברות מתחת לקודקודים, אשר אינם קשורים אליהן. בכל פעם שיצביע המשתמש על קודקוד, יצבע "סמן השכנים" את הקודקוד ואת כל שכניו בצבע צהוב, ובכך יאפשר למשתמש לזהות בקלות, אילו קודקודים קשורים זה לזה (ראה איור 37).

סמן השכנים יופעל עם ההקשה על הצלמית , וביטול פעולתו יעשה ע"י הקשה על .



איור 37. סמן השכנים

אותו הגרף מוצג ללא שימוש ב"סמן השכנים" (למעלה), ועם השימוש בו (משמאל ולמטה).

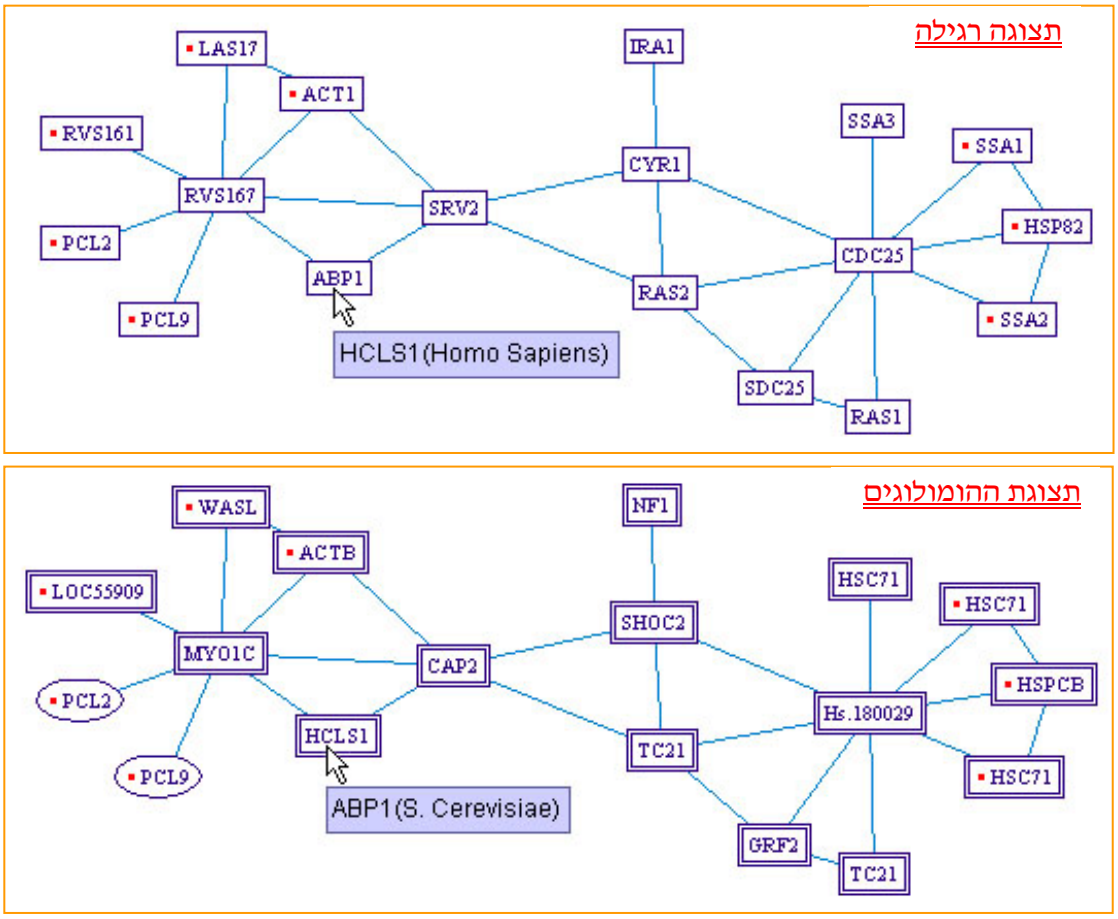
שימו לב למיקום סמן העכבר ולקודקודים הנצבעים.

5.1.11 <u>תצוגת הגרף תוך שימוש בשמות החלבונים ההומולוגים באדם</u>

חוקר העוסק בחלבון אנושי מכיר היטב את שמות החלבונים הקשורים למחקרו באדם. אולם, כאשר אותו חוקר מתבונן באינטראקציות הקיימות באורגניזם אחר, הוא נתקל בשמות חלבונים חדשים, אשר אותם עליו ללמוד כדי להסיק מסקנות מהגרף. לדוגמא, על מנת להשתמש בגרף האינטראקציות של השמר כדי לחזות אינטראקציות אפשריות באדם, צריך החוקר לאתר את החלבון ההומולוגי באדם לכל אחד מחלבוני השמר המוצגים בגרף.

על מנת להקל על החוקר, מאפשרת התוכנה להציג את הגרף עם אותן אינטראקציות (הלקוחות מהשמר), אך להחליף את שמות חלבוני השמר בשמות החלבונים ההומולוגים להם באדם. חלבון, אשר ההומולוג שלו אינו מוכר, יופיע בשמו ה"שמרי", ובתוך מסגרת אליפטית. הביאור (tooltip) המופיע לאחר השארת הסמן על קודקוד מספר שניות ישתנה אף הוא, ויציג עבור חלבון אנושי את ההומולוג השמרי שלו (איור 38).

כדי לעבור לתצוגת ההומולוגים האנושיים, יש ללחוץ על הצלמית , וכדי לחזור אל התצוגה הרגילה יש ללחוץ על הצלמית .



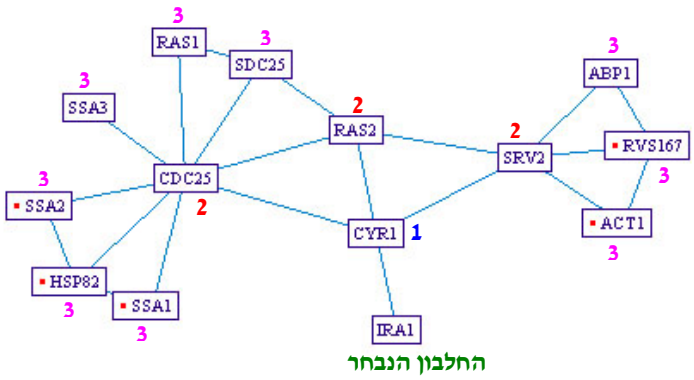
איור 38. תצוגת ההומולוגים

בתצוגה זו מוצגות האינטראקציות המוכרות בשמר כאינטראקציות היפותטיות באדם. שמות חלבוני השמר מוחלפים בשמות החלבונים ההומולוגים להם באדם. חלבון שמרי אשר אין לו הומולוג אנושי מוכר, מוצג בשמו ה"שמרי" כשהוא מוקף באליפסה.

5.1.12 שימוש בשאילתת פרישת השכנים

כאשר חוקר מתחיל לחקור חלבון, הוא ירצה לראותו עם כל החלבונים אשר מקיימים עמו אינטראקציה. הוא גם יהיה מעוניין לראות את החלבונים, אשר אינם באים עמו באינטראקציה ישירה, אלא עקיפה – דרך חלבונים נוספים.

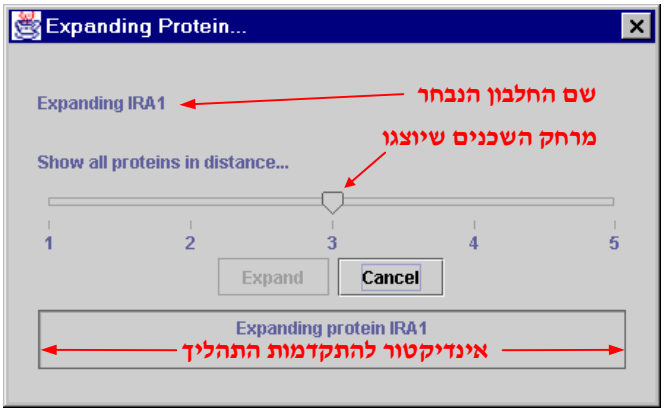
לשם כך, נוצרה שאילתה, המאפשרת למשתמש לפרוש את כל שכניו של קודקוד עד למרחק נתון. שכניו במרחק 1 הם שכניו המיידיים – אלו אשר מצויים באינטראקציה ישירה עימו. ניתן להוסיפם לגרף ע"י הקשה כפולה על החלבון הרצוי. שכניו עד למרחק 2 כוללים גם את החלבונים אשר נמצאים באינטראקציה עם שכניו המיידיים. כלומר, הם חלבונים אשר מקיימים אינטראקציה עם אחד החלבונים הנמצאים במרחק 1, וכן הלאה (איור 39).



איור 39. שימוש בשאילתת פרישת השכנים

כדי להפעיל את השאילתה הזו, יש להצביע עם העכבר על החלבון המבוקש, ולהקיש על פרישת כל שכניו של החלבון IRA1, עד למרחק 3. המספרים המופיעים באיור זה סמוך לקודקודים מעידים על מרחקם של קודקודים אלה מהחלבון המקורי (אינם מוצגים בתוכנה אלא באיור זה בלבד, לצורך הבהרה).

לחצן העכבר הימני לפתיחת תפריט העזר. מתפריט זה יש לבחור באפשרות Expand Node. כתוצאה מכך, ייפתח דיאלוג, אשר יאפשר לנו לבחור עד לאיזה מרחק לפרוש את שכני הקודקוד (ראה איור 40).



איור 40. דיאלוג הקלט לשאילתת פרישת השכנים

עם ההקשה על כפתור האישור, יתחיל תהליך פרישת השכנים, והתקדמותו תוצג עבורנו בתחתית מסך הדיאלוג. נוכל לעצור את התהליך בעזרת לחיצה על הכפתור Cancel. כל המידע שכבר נוסף אל הגרף יישאר בו, אך לא יתווספו

אליו קודקודים נוספים. כמובן שעצירת התהליך אינה פוגעת בתקינות המידע המוצג בגרף (integrity).

5.1.13 שימוש בשאילתת חיפוש מסלולים

שאילתה זו מהווה את אחד הכלים החשובים של התוכנה. השאילתה מקבלת קבוצת חלבונים המופיעים בגרף המוצג, וחלבון נוסף, אשר המשתמש מעוניין למצוא את הקשר בינו לבין חלבונים אלו. היא סורקת את מאגר המידע בחיפוש אחר מסלולי האינטראקציות הקצרים ביותר, אשר מקשרים בין החלבון החדש לאחד החלבונים בקבוצה הנתונה (יתכנו מספר מסלולים שווי אורך). חוקר אשר ביצע ניסוי וגילה מספר חלבונים בעלי התנהגות דומה, יוכל להשתמש בשאילתה זו, על מנת להוסיפם לגרף בזה אחר זה, ולחפש את הקשר ה"קרוב" ביותר ביניהם, כלומר זה המערב מספר נמוך ביותר של חלבונים נוספים. (למשל, קבוצת חלבונים, אשר ביטויים עלה באופן משמעותי לאחר ביצוע מניפולציה כלשהי בתא).

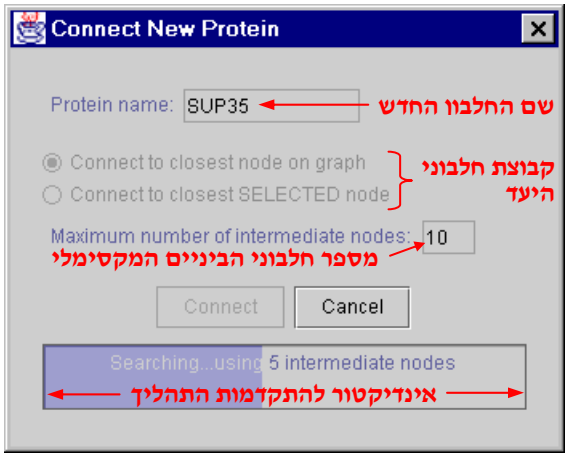
לשם הפעלת השאילתה (איור 41), יש לספק לה שלושה פרמטרים:

1. שם החלבון החדש אשר יש לקשר אל הגרף. חלבון זה עשוי כבר להופיע בגרף. במקרה זה, תנסה התוכנה למצוא מסלול המקשר בינו לבין חלבוני היעד, אשר קצר יותר מזה המופיע כבר בגרף.
2. קבוצת חלבוני היעד, שאל אחד מהם יש לקשר את החלבון החדש. קבוצה זו יכולה להכיל את כל חלבוני הגרף, או רק את חלבוני הגרף שנבחרו ע"י המשתמש (Selected).
3. המספר המקסימאלי של חלבוני הביניים, אשר יכולים להופיע לאורך המסלול. פרמטר זה מגביל את התוכנה לחיפוש מסלולים עד לאורך מקסימאלי נתון. במידה ולא נמצא מסלול שאורכו קצר או שווה למספר הנתון, ייעצר החיפוש, ולמשתמש תוצג הודעה מתאימה.

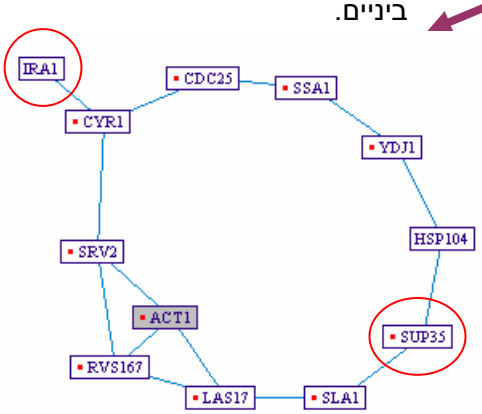
לאורך תהליך החיפוש מוצג למשתמש אינדיקאטור להתקדמות התהליך. כל זמן שלא נמצא המסלול הרצוי, יכול המשתמש להקיש על Cancel לביטול הפעולה.

לצורך הפעלת השאילתה, על המשתמש לבחור בתחילה את קבוצת הקודקודים אליה יש לקשר את החלבון החדש (בחירתם נעשית ע"י העכבר, תוך שימוש במקש Ctrl לבחירת חלבונים נוספים). אין צורך לבחור בקודקודים אם מעוניינים בקישור החלבון החדש אל חלבון כלשהו בגרף כולו, ולא לקבוצת חלבונים מסוימת. לאחר מכן, יש להקיש על הצלמית  , לפתיחת דיאלוג הקלט.

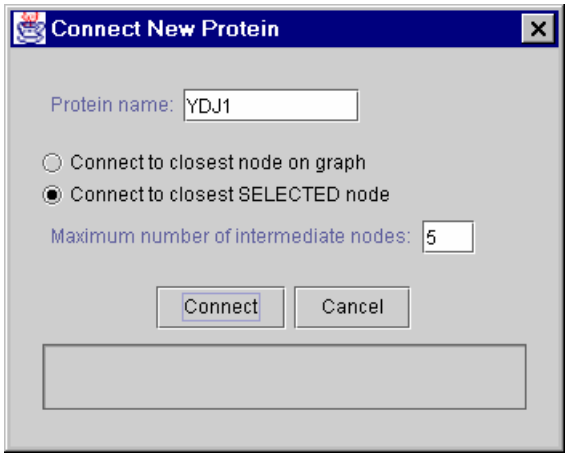
למציאת הקשר בין החלבון IRA1, והחלבון SUP35, נפתח גרף חדש, המכיל את IRA1 בלבד.



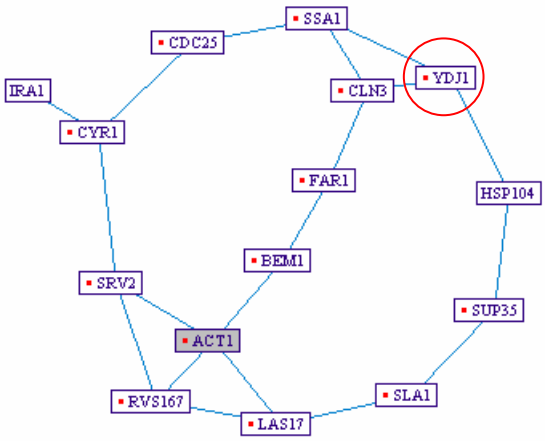
שאילתה זו איתרה שלושה מסלולים מינימליים שווי אורך, בעלי 5 חלבוני ביניים.



כדי לבדוק אם קיים מסלול קצר יותר בין שני חלבוני הגרף ACT1, ו-YDJ1, נבחר את ACT1, ונקיש שוב על הצלמית להצגת הדיאלוג.



השאילתה איתרה מסלול בן 3 חלבוני ביניים. בגרף הקודם, המסלול הקצר ביותר היה בן 4 חלבוני ביניים.



איור 41. שאילתת חיפוש המסלולים

דוגמא לשימוש בשאילתת חיפוש המסלולים, לצורך:

א. קישור חלבון חדש אל הגרף.

ב. חיפוש מסלול קצר יותר בין שני חלבונים המוצגים בגרף.

6 הרחבות עתידיות (הצעות להמשך הפיתוח)

במהלך פיתוחו של כלי זה, הוקדשה מחשבה להרחבות עתידיות אפשריות לגביו. התוכנה פותחה באופן שיאפשר ביצוע הרחבות אלו בקלות, במהירות וללא צורך בביצוע שינויים רחבי היקף בקוד הקיים.

דוגמא לכך היא בנייתם של הציירים עבור הקודקודים והקשתות שבגרף, אשר ע"י החלפתם ניתן לשנות את המידע המוצג בגרף עצמו. המנגנון המאפשר להציג את גרף האינטראקציות תוך שימוש בשמות החלבונים ההומולוגים באדם, נבנה תוך שימוש באפשרות הרחבה זו (הקלאס HomologNameNodePainter).

דוגמא נוספת ניתן לראות בקלאסים Graph, ו-GraphController, אשר מדווחים על אירועים שונים ל-Listeners ומאפשרים להם להגיב לאירועים אלו כרצונם. הקלאס NeighborsMarker משתמש בדיווחים אלו כדי לנהל את רשימת הקודקודים המסומנים בגרף, וציירי הקודקודים והקשתות יודעים להתייחס לרשימה זו, ולצבוע את האזור המתאים בצהוב.

בחלק זה של העבודה אתאר הצעות שונות להמשך פיתוחו של כלי זה. חלק מההצעות מתייחסות להרחבה של תכונות הקיימות כבר בתוכנה, בעוד שאחרות מתייחסות להוספת אפשרויות חדשות לגמרי. הרחבות אלו לא נכללו בכלי הנוכחי, הן בגלל חריגתן מהיקף עבודת גמר זו, והן מפאת מגבלות בגישה אל המידע הביולוגי הדרוש. יש להדגיש, כי פיתוחו של כלי זה נעשה בצמוד למידע שהיה בידינו ואשר התקבל באדיבות חברת Proteome Inc., וכי הדברים שיושמו בו עד כה הושפעו מתוכן מידע זה.

6.1.1 הרחבת השימוש במידע על תכונות החלבונים

המידע אשר היה בידינו לגבי החלבונים במאגר המידע, כלל את שם החלבון בשמר, את שם ההומולוג האנושי שלו, ואת שורת התיאור שלו במאגר המידע YPD. מאגר המידע מכיל נתונים רבים נוספים, אשר חלקם מופיעים כטקסט חופשי, ואחרים מאורגנים בשדות מסודרים, שתוכנם מוגבל לאוסף ערכים ידוע, ואשר נוחים לניתוח ממוחשב (כגון מיקומו התאי של החלבון, התהליך בו הוא משתתף ועוד).

בחלקו הימני התחתון של מסך התוכנה, מופיע אזור, אשר בו מתואר החלבון עליו מצביע המשתמש. ניתן להרחיב את הקלאס המנהל את האזור הזה במסך, כך שיציג כל מידע רצוי הקיים

במאגר המידע לגבי חלבון זה. ניתן גם לאפשר למשתמש לבחור באילו פרטים הוא מעוניין מתוך הקיים במאגר המידע, ולהציגם עבורו באופן מהיר ונוח בחלון זה.

אפשרות נוספת לשימוש במידע הנוסף היא בציור הגרף. על ידי בנייה של צייר קודקודים מתאים ניתן לשנות את צבעם או את צורתם של קודקודי הגרף בהתאם למיקומם התאי או לפונקציה, אשר אליה הם קשורים. ניתן לצבוע את כל קודקודי הגרף בצבעים שונים, או לאפשר למשתמש לבחור את אילו קודקודים לצבוע (למשל, המשתמש יוכל לבקש מהתוכנה לצבוע בצבע שונה את כל החלבונים המופיעים בגרף, אשר מיקומם התאי הוא ממברנת התא).

הרחבה חשובה נוספת היא לאפשר למשתמש לשלב בתצוגת הגרף מידע ניסויי אחר על החלבונים המופיעים בו. ניתן, למשל, להשתמש בתוצאותיו של ניסוי בו נעשה שימוש ב-DNA Chips לבדיקת שינויים ברמת הביטוי של גנים. על ידי צביעת הקודקודים בגרף בצבעים שונים, המעידים על רמת הביטוי של הגנים המתאימים כפי שנמדדה בניסוי, יוכל המשתמש לראות את הקשר בין האינטראקציות שבין החלבונים, לבין השינויים ברמת הביטוי שלהם.

מידע נוסף, אשר ניתן לשלב בתוכנה, מתייחס לחלבונים ההומולוגים באורגניזמים שונים. בעוד שכיום ניתן להתייחס להומולוגים האנושיים של חלבוני השמר, כדאי לאפשר למשתמש את האפשרות לעבוד עם הומולוגים באורגניזמים אחרים.

כמו כן, ניתן בקלות יחסית לאפשר למשתמש לגשת מתוך התוכנה ישירות אל דף המידע באינטרנט של החלבון ההומולוגי לחלבון המוצג באורגניזמים שונים (באופן דומה לאפשרות הקיימת היום לקישור לדף המידע באינטרנט של החלבון השמרי).

6.1.2 <u>הרחבת השימוש במידע על אינטראקציות</u>

המידע אשר היה בידינו ואשר עמו עבדנו, הכיל אינטראקציות פיזיות בלבד. אינטראקציות אלו חסרות כיווניות, ולא מכילות אינטראקציות עצמיות. במאגרי המידע קיים מידע לגבי סוגים נוספים של אינטראקציות, כגון אינטראקציות גנטיות (אינדוקציה, רפרסיה וכו'), אשר בהן הכיווניות חשובה.

ציורן של קשתות הגרף באפליקציה הינו מנוון יחסית. על ידי הרחבתו של צייר הקשתות ניתן יהיה להציג מידע על כיווני האינטראקציות ועל סוגיהן. ניתן לשם כך להשתמש בחיצים, בצבעים שונים, בקווים בעלי עובי שונה או צורה שונה (שלמים, מנוקדים או מקווקווים), ועוד.

האינטראקציות הקיימות ב-YPD, מבוססות כולן על תוצאות ניסויים שפורסמו בספרות המדעית. כדאי לאפשר למשתמש לגשת מתוך התוכנה אל אותו מאמר בו מתוארת האינטראקציה

המעניינת אותו. המידע המקשר בין אינטראקציה לבין המאמרים בהם היא מוזכרת קיים במאגר המידע, ותקצירי המאמרים מצויים ב-PubMed, ולפיכך יישום משימה זו פשוט למדי.

תשומת לב נוספת יכולה להינתן למהימנות הקשת. בעוד שהמידע ב-YPD נאסף באופן פרטני מהעיתונות המקצועית, מצויות במאגרי מידע שונים אינטראקציות, אשר התקבלו מתוך עיבוד טקסט אוטומטי (automatic text-mining), מידע שמקורו בהשערות חישוביות, המבוססות על Comparative Genomics, ומידע אשר נאסף בניסויים רחבי היקף, אשר מהימנותם נמוכה יותר מזו של ניסויים פרטניים בחלבון יחיד. הרחבה אפשרית לתוכנה תתחשב בנתונים אלו כדי להעריך את רמת המהימנות של כל אינטראקציה בגרף. אינטראקציה יכולה לקבל "ציון מהימנות" בהתאם לשיטה האמינה ביותר שבה היא זוהתה, או בהתבסס על שקלול כל השיטות שבהן היא זוהתה, סוגיהן ומספרן. על סמך ציון זה ניתן להציג קשתות מהימנות בלבד, לצבוע את הקשתות בהתאם לרמת המהימנות שלהן, או אף לשלב את המידע באלגוריתם מתוחכם יותר לעימוד הגרף.

6.1.3 <u>הרחבות הקשורות למאגרי המידע</u>

ההרחבות אשר ניתן לבצע בתחום מאגרי המידע הן רבות, ורמת מורכבותן הולכת וגדלה. נתאר אותן, החל מן הפשוטה ביותר ועד למורכבת ביותר.

בתחילה עלינו לזכור, כי התוכנה שלנו פועלת כיום מול בסיס נתונים מקומי, המייצג חלק מן המידע במאגר YPD. הסיבה העיקרית לכך היא שמאגר המידע הנ"ל אינו פומבי, אלא מוצר מסחרי של חברת Proteome Inc., אשר אינה מעוניינת לספק גישה חופשית אליו דרך רשת האינטרנט.

משיקולים אלו נובעות שתי הרחבות חשובות לתוכנה:

א. קישור התוכנה למידע הלקוח ממאגרי מידע אחרים (בעותק מקומי). ישנם ברשת האינטרנט מאגרי מידע רבים העוסקים הן באינטראקציות חלבונים, והן בתיאור החלבונים עצמם (סעיף 2.4). רבים ממאגרי מידע אלו הם נגישים לשימוש חופשי, וניתן אף להוריד עותק מקומי שלהם אל מחשב המשתמש. חשוב לדאוג ולקשר את התוכנה גם למידע הלקוח מהם. יתכן, שכדאי גם לאפשר למשתמש לבנות קבצי מידע משלו, על מנת לשלב בתוכנה את תוצאות ניסוייו.

ב. קישור התוכנה למאגרי מידע דרך רשת האינטרנט. הרחבה זו תגרום להאטה משמעותית בקצב עבודת התוכנה, אך תאפשר לעבוד מול מידע עדכני. את הפגיעה במהירות העבודה ניתן להקטין בשיטות שונות, ביניהן, חיזוי של השאילתות העתידיות וביצוען מראש, או הרצת

תוכנית שירות בצד השרת, אשר תאפשר קשר קבוע ומהיר בין האפליקציה למאגר המידע (פתרון זה תלוי, כמובן, בשיתוף פעולה מצד ספקי המידע). חשוב להדגיש, כי בעבודה ברשת מקומית בעיית המהירות הרבה פחות משמעותית.

השלב הבא, לאחר קישור התוכנה למאגרי נתונים שונים, יעסוק בקישור אל מספר מאגרי מידע במקביל, ובריכוז המידע מכולם. אם נתייחס למידע הנוגע לאורגניזם יחיד בלבד, נצטרך להתמודד עם מספר בעיות, ביניהן:

- כינויו של חלבון בשמות שונים במאגרי מידע שונים.
- האם להציג למשתמש, מאיזה מאגר לקוח כל פרט במידע המוצג, וכיצד?
- האם להציג את המידע הלקוח ממאגרי מידע שונים על פני אותו גרף, או על פני גרפים נפרדים ומקבילים? כיצד יש להתמודד עם עימוד גרף כזה?
- האם לבצע כל שאילתה מול כל מאגרי המידע, או לאפשר למשתמש לעבוד בכל שלב מול מאגר מידע אחר?
- כיצד לשלב בין מאגרי מידע שרמת הפירוט בהם שונה? כיצד להתמודד עם מידע המופיע במאגר מידע אחד ואשר אינו קיים בשני (למשל כיווניות האינטראקציה)?

הרחבת התוכנה כדי לאפשר למשתמש לעבוד מול מידע הלקוח מאורגניזמים שונים במקביל, היא להערכתי בעייתית מדי. היתרון שבעבודה כזו הוא בהתבוננות במידע רב יותר בו זמנית. אך המורכבות של הפתרון ורמת אי הוודאות הטמונה בו, עשויים להקשות על המשתמש, ולהובילו למסקנות שגויות. גם אם יתאפשר אופן זה של עבודה, על המשתמש להתבונן במקביל במספר מצומצם של אורגניזמים בלבד, מהסיבות הבאות:

- עלינו לזכור, כי לחלבונים רבים יש מספר הומולוגים ידועים. תופעה זו נובעת הן משכפול של גנים, הן מהיפרדות של חלבון יחיד למספר חלבונים נפרדים, והן מהשתנות המערכות הביולוגיות עם האבולוציה. לכן על התוכנה לזהות (בעזרת המשתמש) מי הם ההומולוגים הנכונים, ולמצוא דרך לגשר על ההבדלים במספר החלבונים ההומולוגים באורגניזמים השונים.
- יש להציג למשתמש, אילו אינטראקציות בגרף הן אמיתיות לגבי האורגניזם בו הוא מתבונן, ואילו מבוססות על אינטראקציות בין חלבונים הומולוגים באורגניזמים אחרים. יש להקפיד להבחין בין מידע אמיתי לתחזיות של התוכנה.
- אינטראקציות מסוימות עשויות להופיע באורגניזם בודד בלבד, בעוד שאחרות יופיעו באורגניזמים רבים. האם המשמעות היא כי אינטראקציה כזו היא חדשה ולא נחקרה מספיק,

או שמא מדובר במערכת ביולוגית, אשר השתנתה ואינה קיימת עוד? יש צורך להבין את המשמעויות השונות של קשת בגרף, להחליט כיצד להציג מאילו אורגניזמים היא לקוחה, ואולי אף לשלב בשיקולים את המרחק האבולוציוני בין הזנים השונים.

הרחבה אפשרית נוספת היא להשתמש במאגרי מידע, אשר אינם קשורים ישירות לחלבונים או לאינטראקציות, אך מספקים מידע בהקשרים שונים על החלבונים. דוגמא לכך הם מאגרי המידע של מסלולים ביוכימיים. סימון החלבונים בגרף השייכים לאותו מסלול ביוכימי, עשוי לסייע לחוקר בהבנת הגרף והקשר בין החלבונים השונים בו.

7 נספחים

7.1 נספח א' – הוראות התקנה לתוכנה

יש להתקין את התוכנה על מחשב IBM PC, המריץ מערכת הפעלה Windows, היות והגישה אל מאגר המידע נעשית כיום דרך שירות של Windows בשם ODBC (הגישה אל מאגר המידע נעשית באמצעות הדריבר JDBC-ODBC Bridge).

לעבודה זו מצורף תקליטור ועליו מספר קבצים החשובים להרצת התוכנה:

- קבצי התוכנה PIVOT, וקבצי המידע בהם היא משתמשת.
- קבצי התקנה של Java Runtime, המיועדים להתקנה על מערכות Windows.
- סרטון הדגמה, המסביר את אופן השימוש בתוכנה וכיצד להשתמש בתכונותיה השונות (סעיף 7.3).

7.1.1 התקנת Java Runtime

בשלב ראשון יש להתקין על המחשב תוכנה המאפשרת הרצת אפליקציות Java. חשוב לוודא כי מתקינים Java 2 Standard Edition בגרסא 1.4.0 ומעלה.

את קבצי ההתקנה המתאימים ניתן למצוא על התקליטור המצורף בתיקיה Java, או להוריד מאתר האינטרנט http://www.javasoft.com (חינם).

יש להפעיל את קובץ ההתקנה, ולעקוב אחרי ההוראות.

7.1.2 התקנת ODBC data source

יש להתקין מקור מידע ODBC חדש, בשם YeastProteinDB, אשר משתמש ב-driver של MS-Access ומצביע אל הקובץ YeastProteinDB.mdb הנמצא על התקליטור בספריה "pivot\data".

1. במערכת Win98, יש ללכת אל ה-Control Panel, וממנו להפעיל " ODBC Data Sources (32 bit)".

   במערכת Win2000, בוחרים מתוך ה-Control Panel באפשרות " Administrative Tools" ושם ב- "Data Sources (ODBC)".

2. בחלון שנפתח עוברים ל- "User DSN", ושם מקישים על הכפתור "...Add".
3. בוחרים באפשרות "Microsoft Access Driver (*.mdb)", ומקישים על הכפתור "Finish".
4. בשדה "Data Source Name" מקלידים את השם "YeastProteinDB" (בלי רווחים).
5. מקישים על הכפתור "...Select".
6. בוחרים בקובץ "YeastProteinDB.mdb" הנמצא על התקליטור בספריה "pivot\data", ומקישים "OK".
7. מקישים "OK" פעמיים נוספות לסגירת התפריטים.

7.1.3 <u>הפעלת התוכנה</u>

כדי להפעיל את התוכנה נותר לטייל בתקליטור בעזרת הסייר, ולהקיש פעמיים על הקובץ ששמו "pivot.jar" הנמצא בספריה "pivot".

עם תחילת העבודה מוצג למשתמש חלון ריק, המאפשר לו להתחיל בעבודה על גרף חדש, או לטעון גרף ישן, שעליו עבד בעבר. כדי להתחיל בעבודה על גרף חדש יש לבחור באפשרות "...New" מתוך התפריט "File", ולהקליד את שמו של החלבון השמרי, שממנו נרצה להתחיל את העבודה (סעיף 7.2).

לתשומת לב, עם תחילת העבודה על גרף מנגנון העימוד האוטומטי כבוי. ללא הפעלתו, לא יתבצע עימוד אוטומטי של קודקודי הגרף. להפעלת מנגנון העימוד האוטומטי, יש להקיש על הצלמית  - "Activate auto layout", שהיא הצלמית השישית משמאל.

לפני תחילת העבודה, מומלץ לצפות בסרטון להדגמת התוכנה (סעיף 7.3), המציע סקירה מקיפה יותר של השימוש ב-PIVOT.

7.2 נספח ב' - רשימת חלבוני השמר במאגר המידע

רשימת שמות החלבונים אשר להם יש אינטראקציות במאגר המידע של התוכנה, מופיעה בקובץ "proteins_list.txt" בתיקיה "documents", ולהלן:

A1	AAD6	ABD1	ABF1	ABP1	ACE2	ACT1	ADA2
ADE2	ADE6	ADE8	ADH1	ADH2	ADP1	ADR1	ADY1
AFG3	AFR1	AGA1	AGA2	AHC1	AHP1	AIP1	AIP2
AKL1	AKR1	AKR2	ALF1	ALPHA1	ALPHA2	ALR1	AMD1
AME1	ANC1	ANP1	AOS1	APC11	APC2	APC9	APG1
APG12	APG13	APG14	APG16	APG5	APG7	APG9	APL1
APL2	APL4	APL5	APM1	APM2	APS1	APT1	APT2
AQY2	ARC1	ARC18	ARC19	ARC35	ARC40	ARD1	ARG1
ARG3	ARGR1	ARGR2	ARGR3	ARH1	ARK1	ARP1	ARP10
ARP2	ARP3	ARP4	ASK10	ASM4	ASN1	AST1	ATP1
ATP14	ATP15	ATP16	ATP17	ATP18	ATP2	ATP20	ATP3
ATP4	ATP5	ATP6	ATP7	ATP8	ATP9	ATX1	AUT1
AUT2	AUT7	BAS1	BAT2	BBP1	BCK1	BDF1	BDF2
BEM1	BEM4	BET1	BET2	BET3	BET4	BET5	BFR1
BFR2	BIK1	BIM1	BIR1	BMH1	BMH2	BNI1	BNI4
BNR1	BOI1	BOI2	BOP3	BOS1	BRE2	BRF1	BRN1
BRR2	BTT1	BUB1	BUB2	BUB3	BUD6	BUL1	BUL2
BUR6	CAF17	CAF20	CAK1	CAR1	CAR2	CAT2	CBC2
CBF1	CBF2	CBF5	CCC2	CCL1	CCR4	CCT5	CDC10
CDC11	CDC12	CDC123	CDC13	CDC14	CDC16	CDC19	CDC2
CDC20	CDC23	CDC24	CDC25	CDC26	CDC27	CDC28	CDC3
CDC31	CDC33	CDC34	CDC36	CDC37	CDC39	CDC4	CDC40
CDC42	CDC43	CDC45	CDC46	CDC47	CDC48	CDC5	CDC53
CDC54	CDC55	CDC6	CDC7	CDC73	CEF1	CEG1	CEP3
CET1	CFT1	CFT2	CHA4	CHC1	CHK1	CHS3	CIK1
CIN2	CIN4	CIT1	CIT2	CKA1	CKA2	CKB1	CKB2
CKS1	CLA4	CLB1	CLB2	CLB3	CLB4	CLB5	CLC1

CLF1	CLG1	CLN1	CLN2	CLN3	CLP1	CLU1	CMD1
CMK1	CMK2	CMP2	CNA1	CNB1	CNM67	CNS1	COF1
COQ5	COX4	COX5A	CPA1	CPA2	CPR6	CPR7	CRM1
CRN1	CRZ1	CSE1	CSE2	CSE4	CTF13	CTF19	CTF4
CTH1	CTK1	CTK2	CTK3	CTL1	CTR9	CUE1	CUP2
CUS1	CUS2	CYP2	CYR1	CYS4	DAD1	DAL80	DAM1
DBF2	DBF20	DBF4	DBP2	DBP5	DBP7	DCI1	DCP1
DCP2	DDC1	DDI1	DED1	DHH1	DHS1	DIB1	DIG1
DIG2	DIS3	DMC1	DNA2	DNA43	DNL4	DNM1	DOA1
DOC1	DPB11	DPB2	DSS4	DST1	DUN1	DUO1	DUT1
EAP1	EBS1	ECI1	ECM13	ECM19	ECM32	EFB1	EFT2
EGD1	EGD2	ELA1	ELC1	ELP2	ELP3	EMB1	EMP24
END3	ENO1	ENO2	ENT1	ENT3	ERG27	ERO1	ERP1
ERP2	ERV25	ESA1	ESP1	EST1	EXO70	EXO84	FAP1
FAR1	FAR3	FCP1	FCY1	FET3	FET5	FIG1	FIL1
FIN1	FIP1	FIR1	FKS1	FMN1	FOB1	FPR1	FPR4
FRQ1	FTH1	FTR1	FUI1	FUR1	FUR4	FUS2	FUS3
FZF1	GAC1	GAL1	GAL11	GAL3	GAL4	GAL80	GAL83
GAR1	GCD1	GCD10	GCD11	GCD14	GCD2	GCD6	GCD7
GCN1	GCN2	GCN20	GCN3	GCN4	GCN5	GCR1	GCR2
GCS1	GDH1	GDH2	GDI1	GDS1	GFD1	GIC1	GIC2
GIF1	GIM5	GIN4	GIP1	GIP2	GIS4	GLC7	GLC8
GLE1	GLE2	GLG2	GLK1	GLN3	GNA1	GNP1	GOS1
GPA1	GPA2	GPD2	GPR1	GPX2	GRC3	GRD19	GRF10
GRR1	GRX3	GRX5	GSC2	GSG1	GSP1	GSP2	GSY2
GTR1	GYP1	GZF3	HAC1	HAP2	HAP3	HAP4	HAP5
HAT1	HAT2	HCR1	HCS1	HDA1	HDR1	HEM13	HEX3
HFI1	HHF1	HHF2	HHT1	HHT2	HIR1	HIR2	HMO1
HOC1	HOF1	HOG1	HOM3	HOP1	HOR2	HOT1	HPR5
HRB1	HRP1	HRR25	HRT1	HSC82	HSF1	HSH49	HSL1
HSL7	HSP10	HSP104	HSP42	HSP60	HSP82	HST3	HTA1

HTA2	HTA3	HTB1	HTB2	HYR1	HYS2	ICL1	IDH1
IDH2	IKI1	IKI3	IKS1	IME1	IME2	IME4	IMP1
IMP2	IMP3	IMP4	INH1	INO2	INO4	INO80	IPK1
IPL1	IQG1	IRA1	IRE1	IRR1	ISA1	ISC10	IST3
ISU2	ISY1	JNM1	JSN1	KAP104	KAP114	KAP122	KAP123
KAP95	KAR1	KAR2	KAR3	KAR4	KAR9	KCC4	KCS1
KEL1	KEL2	KGD1	KGD2	KIC1	KIN1	KIN28	KIN3
KNS1	KRE11	KRE6	KRR1	KSS1	KTI12	KTR3	LAC1
LAP4	LAS1	LAS17	LAT1	LCB1	LCB2	LCD1	LCP5
LEA1	LEU4	LHP1	LIF1	LOC7	LOS1	LPD1	LPP1
LRS4	LSM1	LSM2	LSM3	LSM4	LSM5	LSM6	LSM7
LSM8	LST7	LUV1	LYS5	LYS9	MAD1	MAD2	MAD3
MAK10	MAK3	MAK31	MAS1	MAS2	MBF1	MBP1	MCD1
MCK1	MCM1	MCM16	MCM2	MCM21	MCM22	MCM3	MCM6
MDH1	MDH3	MDJ1	MDS3	MEC1	MEC3	MED1	MED11
MED2	MED4	MED6	MED7	MED8	MEK1	MEL1	MER1
MES1	MET14	MET17	MET18	MET28	MET30	MET31	MET32
MET4	MEX67	MFALPHA1	MFALPHA2	MGA1	MGE1	MHP1	MHT1
MIF2	MIG1	MIP6	MIR1	MKK1	MKK2	MLH1	MLH2
MLH3	MLP2	MMM1	MMS2	MNN10	MNN11	MNN9	MNS1
MOB1	MOB2	MOG1	MOT1	MPD2	MPP10	MPS1	MPS2
MPT1	MPT5	MRE11	MRP13	MRP8	MRS11	MRS5	MRS6
MRT1	MSB2	MSH2	MSH3	MSH4	MSH5	MSH6	MSI1
MSL1	MSL5	MSN5	MSO1	MSR1	MSS1	MTF1	MTH1
MTO1	MTR10	MTR2	MTR3	MTW1	MUD2	MUM2	MUS81
MVP1	MYO2	MYO3	MYO4	MYO5	NAB2	NAM7	NAM9
NAN1	NAP1	NAT1	NBP1	NBP2	NBP35	NCB2	NCE103
NDC1	NDD1	NEM1	NET1	NFI1	NGG1	NHP10	NHP2
NIC96	NIF3	NIP1	NIP100	NIP29	NIP7	NMD2	NMD3
NMD4	NMD5	NOP1	NOP5	NOT3	NOT5	NPI46	NPL3
NPL4	NPL6	NPR2	NRD1	NRG1	NSP1	NTC20	NTC40

NTF2	NTH1	NUD1	NUF1	NUF2	NUM1	NUP1	NUP100
NUP116	NUP120	NUP133	NUP145	NUP157	NUP159	NUP170	NUP188
NUP192	NUP2	NUP42	NUP49	NUP53	NUP57	NUP82	NUP84
NUP85	NUT2	NVJ1	NYV1	OAF1	ORC1	ORC2	ORC3
ORC4	ORC5	ORC6	OST1	OST2	OST3	OST4	OST5
PAB1	PAC1	PAC2	PAF1	PAN1	PAN2	PAN3	PAP1
PAT1	PBI2	PBN1	PBP1	PBS2	PCF11	PCL1	PCL10
PCL2	PCL5	PCL6	PCL7	PCL8	PCL9	PDB1	PDI1
PDR1	PDR11	PDS1	PDX1	PEA2	PEP12	PEP3	PEP5
PEP7	PET122	PET191	PET494	PET54	PEX13	PEX14	PEX17
PEX18	PEX19	PEX21	PEX3	PEX5	PEX7	PEX8	PFK1
PFK2	PFY1	PGD1	PGI1	PHB1	PHB2	PHM2	PHO13
PHO4	PHO80	PHO81	PHO85	PIB1	PIG1	PIG2	PIK1
PIP2	PKA3	PKC1	PKH1	PLC1	PLP1	PLP2	PMA1
PMD1	PMP1	PMP2	PMS1	PMT1	PMT2	POB3	POL1
POL2	POL30	POL32	POM152	POP2	POT1	PPA2	PPE1
PPG1	PPH21	PPH22	PPT1	PPZ1	PRB1	PRC1	PRE1
PRE10	PRE2	PRE3	PRE5	PRI1	PRI2	PRP11	PRP19
PRP2	PRP21	PRP3	PRP38	PRP39	PRP4	PRP40	PRP45
PRP46	PRP5	PRP6	PRP8	PRP9	PRS1	PRS2	PRS3
PRS4	PRS5	PRT1	PSE1	PST2	PTA1	PTC1	PTC2
PTC4	PTM1	PTP2	PTP3	PUP1	PUP2	PUT3	PUT4
PWP2	PXA1	PXA2	QCR6	RAD1	RAD10	RAD14	RAD16
RAD17	RAD18	RAD2	RAD23	RAD24	RAD27	RAD3	RAD4
RAD5	RAD50	RAD51	RAD52	RAD53	RAD54	RAD55	RAD57
RAD6	RAD7	RAD9	RAI1	RAM1	RAM2	RAP1	RAS1
RAS2	RAT1	RAV1	RBL2	RCK1	RCK2	RCL1	RCS1
RDH54	RDI1	RED1	REF2	REG1	REG2	REP1	REP2
RET1	REV1	REV3	REV7	RF3	RFA1	RFA2	RFA3
RFC1	RFC2	RFC3	RFC4	RFC5	RFX1	RGA1	RGD1
RGR1	RGS2	RGT2	RHO1	RHO2	RHO3	RHO4	RIB4

RIF1	RIF2	RIM101	RIM11	RIM15	RIS1	RLF2	RLM1
RNA1	RNA14	RNA15	RNP1	RNR1	RNR2	RNR3	RNR4
ROM2	ROX3	RPA12	RPA135	RPA190	RPB10	RPB11	RPB2
RPB3	RPB4	RPB5	RPB7	RPB8	RPC10	RPC11	RPC19
RPC25	RPC31	RPC34	RPC37	RPC40	RPC53	RPC82	RPD3
RPG1	RPL10	RPL11A	RPL11B	RPL12A	RPL12B	RPL25	RPL30
RPL31A	RPL31B	RPL42B	RPL5	RPN1	RPN10	RPN11	RPN12
RPN2	RPN3	RPN4	RPN5	RPN6	RPN8	RPN9	RPO21
RPO26	RPO31	RPO41	RPP0	RPP1A	RPP1B	RPP2A	RPP2B
RPS22A	RPS22B	RPS26A	RPS26B	RPS28A	RPS28B	RPS31	RPS8B
RPT1	RPT3	RPT4	RPT5	RPT6	RRN10	RRN11	RRN6
RRN7	RRN9	RRP1	RRP4	RRP42	RRP43	RRP6	RSC8
RSE1	RSG1	RSP5	RSR1	RTA1	RTG1	RTG3	RTS1
RTT101	RUB1	RVS161	RVS167	SAC2	SAC3	SAC6	SAC7
SAE2	SAG1	SAP1	SAP155	SAP185	SAP190	SAP4	SAR1
SAS3	SBA1	SBH1	SBH2	SCC2	SCD5	SCJ1	SCM2
SCM4	SCS3	SCW11	SDC25	SDH1	SDH2	SDH3	SDS22
SDS3	SEC1	SEC10	SEC11	SEC13	SEC14	SEC15	SEC16
SEC17	SEC18	SEC2	SEC20	SEC21	SEC22	SEC23	SEC24
SEC31	SEC34	SEC35	SEC4	SEC5	SEC6	SEC61	SEC62
SEC63	SEC66	SEC72	SEC8	SEC9	SED1	SED4	SED5
SEH1	SEN15	SEN2	SEN34	SEN54	SER3	SER33	SET1
SET2	SFB2	SFB3	SFH1	SFL1	SFP1	SFT2	SGN1
SGS1	SGT1	SGT2	SHE2	SHE3	SHO1	SHR3	SIC1
SIG1	SIK1	SIM1	SIN3	SIN4	SIP1	SIP2	SIP3
SIP4	SIR1	SIR2	SIR3	SIR4	SIS2	SIT4	SIW14
SKI3	SKI6	SKI8	SKN7	SKP1	SKT5	SLA1	SLA2
SLD2	SLI15	SLN1	SLT2	SLU7	SLY1	SMB1	SMC1
SMC2	SMC3	SMC4	SMD1	SMD2	SMD3	SME1	SMI1
SMK1	SML1	SMP2	SMT3	SMX2	SMX3	SMY1	SMY2
SNC1	SNC2	SNF1	SNF11	SNF12	SNF2	SNF3	SNF4

SNF5	SNF6	SNF7	SNF8	SNO1	SNO3	SNP1	SNT309
SNU114	SNU23	SNX4	SNZ1	SNZ2	SNZ3	SOF1	SOH1
SOM1	SOR1	SPA2	SPB1	SPC1	SPC19	SPC2	SPC24
SPC25	SPC3	SPC34	SPC42	SPC72	SPC97	SPC98	SPH1
SPO1	SPO11	SPO12	SPO13	SPO21	SPO7	SPO71	SPP2
SPP381	SPP41	SPR28	SPT14	SPT15	SPT16	SPT2	SPT20
SPT23	SPT3	SPT4	SPT5	SPT6	SPT7	SPT8	SQT1
SRA1	SRB2	SRB4	SRB5	SRB6	SRB7	SRB8	SRB9
SRL2	SRM1	SRN2	SRO7	SRO77	SRP1	SRP101	SRP102
SRP68	SRP72	SRV2	SSA1	SSA2	SSA3	SSB1	SSC1
SSK1	SSK2	SSK22	SSL1	SSL2	SSN3	SSN6	SSN8
SSO1	SSO2	SSP1	SSS1	SST2	SSU72	STB1	STB2
STB3	STB4	STB5	STB6	STD1	STE11	STE12	STE13
STE18	STE20	STE23	STE24	STE3	STE4	STE5	STE50
STE7	STH1	STI1	STM1	STN1	STO1	STP1	STS1
STT3	STT4	STU2	SUA7	SUI1	SUI2	SUI3	SUP35
SUP45	SWE1	SWI3	SWI4	SWI5	SWI6	SWP1	SXM1
SYF1	SYF2	SYG1	TAD2	TAD3	TAF145	TAF17	TAF19
TAF25	TAF40	TAF47	TAF60	TAF61	TAF65	TAF67	TAF90
TAH18	TAP42	TBF1	TCM62	TDH2	TEF1	TEF2	TEF4
TEL2	TEM1	TFA1	TFA2	TFB1	TFB2	TFB3	TFB4
TFC1	TFC4	TFC5	TFC7	TFG1	TFG2	THI6	TID3
TIF1	TIF2	TIF3	TIF34	TIF35	TIF4631	TIF4632	TIF5
TIM11	TIM17	TIM18	TIM22	TIM23	TIM44	TIM54	TIM9
TIP20	TLG1	TLG2	TOA1	TOA2	TOF1	TOF2	TOM20
TOM22	TOM37	TOM40	TOM5	TOM6	TOM7	TOM70	TOM72
TOP1	TOP2	TOP3	TOR2	TPD3	TPK3	TPM2	TPO1
TPS1	TPS2	TPS3	TRA1	TRF4	TRP2	TRP3	TRP5
TRR2	TRS120	TRS130	TRS20	TRS23	TRS31	TRS33	TRX1
TRX2	TSA1	TSC3	TSL1	TSM1	TSP1	TUB1	TUB2
TUB3	TUB4	TUP1	TWF1	UBA2	UBA3	UBC13	UBC6

UBC7 UBC9 UBI4 UBP10 UBP3 UBP5 UBP8 UBR1

UBR2 UFD1 UFD2 UFD4 UFE1 UGA4 UGA5 ULA1

UME6 UPF3 URE2 URH1 URK1 VAB2 VAC8 VAM3

VAM6 VAM7 VAN1 VIK1 VMA1 VMA10 VMA13 VMA2

VMA22 VMA4 VMA6 VMA7 VMA8 VPH1 VPS15 VPS16

VPS17 VPS21 VPS27 VPS29 VPS30 VPS33 VPS34 VPS35

VPS36 VPS4 VPS41 VPS45 VPS5 VPS53 VPS9 VRP1

VTI1 WBP1 WSC3 WTM1 WTM2 XPT1 XRS2 YAK1

YAL014C YAL028W YAL036C YAL049C YAP1 YAP1801 YAP1802 YAP3

YAP5 YAR003W YAR014C YAR031W YAR033W YAR066W YBL010C YBL051C

YBR052C YBR077C YBR094W YBR108W YBR134W YBR141C YBR175W YBR190W

YBR194W YBR228W YBR250W YBR270C YBR281C YBR284W YCG1 YCK1

YCK2 YCL046W YCP4 YCR022C YCR030C YCR045C YCR050C YCR059C

YCR063W YCR082W YCR086W YCR087W YCR106W YDJ1 YDL001W YDL011C

YDL012C YDL063C YDL071C YDL076C YDL089W YDL110C YDL113C YDL117W

YDL133W YDL144C YDL146W YDL203C YDL213C YDL214C YDL216C YDL225W

YDL239C YDL246C YDR013W YDR020C YDR026C YDR061W YDR063W YDR070C

YDR071C YDR078C YDR084C YDR115W YDR128W YDR132C YDR140W YDR152W

YDR179C YDR200C YDR214W YDR215C YDR255C YDR267C YDR273W YDR279W

YDR326C YDR348C YDR357C YDR372C YDR383C YDR398W YDR412W YDR469W

YDR482C YDR485C YDR489W YDR532C YEF3 YEL015W YEL023C YEL041W

YEL068C YER007C-A YER045C YER079W YER092W YER116C YER124C

YER126C YER147C YFH1 YFL023W YFL042C YFL061W YFR003C YFR008W

YFR011C YFR032C YFR042W YFR043C YFR046C YFR047C YFR057W YGL015C

YGL024W YGL051W YGL096W YGL104C YGL161C YGL170C YGL174W YGL198W

YGL214W YGL230C YGL239C YGL242C YGL245W YGR003W YGR010W YGR017W

YGR024C YGR046W YGR058W YGR068C YGR117C YGR122W YGR136W YGR153W

YGR154C YGR160W YGR173W YGR179C YGR232W YGR250C YGR268C YGR269W

YGR278W YGR294W YHB1 YHL002W YHL006C YHL018W YHL042W YHL046C

YHR022C YHR032W YHR035W YHR039C YHR073W YHR083W YHR100C YHR105W

YHR111W YHR115C YHR122W YHR134W YHR140W YHR145C YHR188C YHR197W

YHR198C YHR204W YHR207C YHR216W YIF1 YIL001W YIL007C YIL008W

YIL011W YIL028W YIL065C YIL105C YIL112W YIL113W YIL132C YIL151C

YIL163C YIL172C YIP1 YIR025W YIR044C YJL011C YJL019W YJL048C

YJL058C YJL064W YJL065C YJL075C YJL084C YJL112W YJL149W YJL151C

YJL160C YJL178C YJL184W YJL185C YJL200C YJL211C YJL218W YJR015W

YJR056C YJR072C YJR083C YJR136C YJU2 YKE2 YKL002W YKL023W

YKL033W YKL052C YKL061W YKL070W YKL075C YKL090W YKL116C YKL155C

YKL161C YKL183W YKR007W YKR011C YKR022C YKR030W YKR060W YKR083C

YKR096W YKR104W YKT6 YKU70 YKU80 YLL032C YLL049W YLR016C

YLR046C YLR049C YLR065C YLR072W YLR097C YLR128W YLR132C YLR135W

YLR154C YLR190W YLR224W YLR238W YLR243W YLR254C YLR266C YLR269C

YLR270W YLR294C YLR312C YLR315W YLR322W YLR323C YLR324W YLR328W

YLR345W YLR352W YLR368W YLR376C YLR386W YLR392C YLR419W YLR423C

YLR424W YLR432W YLR435W YLR456W YLR465C YMD8 YML053C YML068W

YML088W YML119W YMR009W YMR025W YMR048W YMR057C YMR068W YMR071C

YMR077C YMR087W YMR093W YMR102C YMR140W YMR163C YMR181C YMR210W

YMR211W YMR226C YMR233W YMR263W YMR269W YMR288W YMR291W YMR295C

YMR312W YMR317W YMR322C YNG2 YNL047C YNL056W YNL078W YNL086W

YNL091W YNL092W YNL094W YNL099C YNL116W YNL122C YNL127W YNL146W

YNL155W YNL157W YNL164C YNL171C YNL201C YNL218W YNL234W YNL279W

YNL288W YNL311C YNL335W YNR004W YNR005C YNR022C YNR025C YNR029C

YNR048W YNR054C YNR068C YNR069C YOL034W YOL036W YOL070C YOL082W

YOL083W YOL101C YOL106W YOL111C YOL129W YOL144W YOR006C YOR056C

YOR062C YOR078W YOR105W YOR111W YOR138C YOR161C YOR164C YOR206W

YOR215C YOR220W YOR262W YOR264W YOR275C YOR284W YOR292C YOR315W

YOR331C YOR353C YOR359W YPD1 YPL025C YPL070W YPL088W YPL105C

YPL110C YPL133C YPL157W YPL180W YPL192C YPL201C YPL222W YPL229W

YPL238C YPL246C YPL260W YPR008W YPR011C YPR040W YPR049C YPR078C

YPR093C YPR105C YPR115W YPR118W YPR152C YPT1 YPT31 YPT53

YPT7 YRA1 YRB1 YRB2 YRF1-4 YSC84 YSH1 YTA12

YTA6 YTH1 ZAP1 ZDS1 ZDS2 ZIP2 ZRC1

7.3 נספח ג' - סרטון להדגמת התוכנה

כדי לצפות בסרטון, המדגים את העבודה ב-PIVOT יש להשתמש בסייר ולאתר את הקובץ ששמו "pivot-demo.avi" בספריה "movie". הקשה כפולה על הקובץ תגרום להפעלתו. מומלץ לצפות בסרטון בגודלו המקורי (100%).

אורך הסרטון כ-9 דקות, והוא מציג את העבודה ב-PIVOT, החל מן היסוד, ועד השימוש בתכונותיה המתקדמות של התוכנה.

הסרטון דובר אנגלית (במבטא ישראלי ועם הצופים הסליחה).

8 רשימת ספרות

1. Lockhart, D.J. and E.A. Winzeler, *Genomics, gene expression and DNA arrays.* Nature, 2000. **405**(6788): p. 827-36.

2. Kumar, A. and M. Snyder, *Emerging technologies in yeast genomics.* Nat Rev Genet, 2001. **2**(4): p. 302-12.

3. TIGR website, at "http://www.tigr.org"/.

4. NCBI website, at "http://www.ncbi.nlm.nih.gov/Genomes"/.

5. *Proteomics.* Nat Biotechnol, 2000. **18 Suppl**: p. IT45-6.

6. Eisenberg, D., E.M. Marcotte, I. Xenarios, and T.O. Yeates, *Protein function in the post-genomic era.* Nature, 2000. **405**(678 :(8p. 823-6.

7. Harris, T., *Genetics, genomics, and drug discovery.* Med Res Rev, 2000. **20**(3): p. 203-11.

8. McKusick, V.A., *Genomics: structural and functional studies of genomes.* Genomics, 1997. **45**(2): p. 244-9.

9. *Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.* Science, 1998. **282**(5396): p. 2012-8.

10. Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S.G. Oliver, *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.

11. Costanzo, M.C., J.D. Hogan, M.E. Cusick, B.P. Davis, A.M. Fancher, P.E. Hodges, P. Kondu, C. Lengieza, J.E. Lew-Smith, C. Lingner, K.J. Roberg-Perez, M. Tillberg, J.E. Brooks, and J.I. Garrels, *The yeast proteome database (YPD) and Caenorhabditis elegans proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information.* Nucleic Acids Res, 2000. **28**(1): p. 73-6.

12. Bork, P. and E.V. Koonin, *Protein sequence motifs.* Curr Opin Struct Biol, 1996. **6**(3): p. 366-76.

13. Lodish, H.F., *Molecular cell biology*. 2000, W.H. Freeman: New York.

14. Creighton, T.E., *Proteins* : *structures and molecular properties*. 2nd ed. 1992, New York: W.H. Freeman. xiii, 507.

15. Koonin, E.V., R.L. Tatusov, and M.Y. Galperin, *Beyond complete genomes: from sequence to structure and function.* Curr Opin Struct Biol, 1998. **8**(3): p. 355-63.

16. Kriventseva, E.V., M. Biswas, and R. Apweiler, *Clustering and analysis of protein families.* Curr Opin Struct Biol, 2001. **11**(3): p. 334-9.

17. Bentley, D.R., *The Human Genome Project--an overview.* Med Res Rev, 2000. **20**(3): p. 189-96.

18. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

19. Galperin, M.Y. and E.V. Koonin, *Who's your neighbor? New computational approaches for functional genomics.* Nat Biotechnol, 2000. **18**(6): p. 609-13.

20. Marcotte, E.M., M. Pellegrini, M.J. Thompson, T.O. Yeates, and D. Eisenberg, *A combined algorithm for genome-wide prediction of protein function.* Nature :(6757)402 .1999 ,p. 83-6.

21. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure.* Science, 1991. **253**(5016): p. 164-70.

22. Alberts, B., *Molecular biology of the cell*. 2nd ed ,1989 .New York: Garland Pub. xxxv, 1219, 44.

23. Burns, N., B. Grimwade, P.B. Ross-Macdonald, E.Y. Choi, K. Finberg, G.S. Roeder, and M. Snyder, *Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae.* Genes Dev, 1994. **8**(9): p. 1087-105.

24. Kumar, A., S.A. des Etages, P.S. Coelho, G.S. Roeder, and M. Snyder, *High-throughput methods for the large-scale analysis of gene function by transposon tagging.* Methods Enzymol, 2000. **328**: p. 550-74.

25. Phizicky, E.M. and S. Fields, *Protein-protein interactions: methods for detection and analysis.* Microbiol Rev, 1995. **59**(1): p. 94-123.

26. Coligan, J.E., *Current protocols in protein science*. 1999, Wiley: Brooklyn, N.Y. Chapter 19.

27. Kolodziej, P.A. and R.A .Young, *Epitope tagging and protein surveillance.* Methods Enzymol, 1991. **194**: p. 508-19.

28. Winzeler, E.A., D.D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J.D. Boeke, H. Bussey, A.M. Chu, C. Connelly, K. Davis, F. Dietrich, S.W. Dow, M. El Bakkoury, F. Foury, S.H. Friend, E. Gentalen, G. Giaever, J.H. Hegemann, T. Jones, M. Laub, H. Liao, R.W. Davis, and et al., *Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis.* Science, 1 :(5429)285 .999p. 901-6.

29. Smith, V., D. Botstein, and P.O. Brown, *Genetic footprinting: a genomic strategy for determining a gene's function given its sequence.* Proc Natl Acad Sci U S A, 1995. **92**(14): p. 6479-83.

30. Crooke, S.T., *Molecular mechanisms of action of antisense drugs.* Biochim Biophys Acta, 1999. **1489**(1): p. 31-44.

31. Caplen, N.J., S. Parrish, F. Imani, A. Fire, and R.A. Morgan, *Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems.* Proc Natl Acad Sci U S A, 2001. **98**(17): p. 9742-7.

32. Elbashir, S.M., J. Harborth, K. Weber, and T. Tuschl, *Analysis of gene function in somatic mammalian cells using small interfering RNAs.* Methods, 2002. **26**(2): p. 199-213.

33. Elbashir, S.M., J. Harborth ,W. Lendeckel, A. Yalcin, K. Weber, and T. Tuschl, *Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.* Nature, 2001. **411**(6836): p. 494-8.

34. Shoemaker, D.D., D.A. Lashkari, D. Morris, M. Mittmann, and R.W. Davis, *Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.* Nat Genet, 1996. **14**(4): p. 450-6.

35. MacBeath, G. and S.L. Schreiber, *Printing proteins as microarrays for high-throughput function determination* .Science, 2000. **289**(5485): p. 1760-3.

36. Zhu, H., J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M. Gerstein, M.A. Reed, and M. Snyder, *Analysis of yeast protein kinases using protein chips.* Nat Genet, 2000. **26**(3): p. 283-9.

37. Arenkov, P., A. Kukhtin, A. Gemmell, S. Voloshchuk, V. Chupeeva, and A. Mirzabekov, *Protein microchips: use for immunoassay and enzymatic reactions.* Anal Biochem, 2000. **278**(2): p. 123-31.

38. Marcotte, E.M., *Computational genetics: finding protein function by nonhomology methods.* Curr Opin Struct Biol, 2000. **10**(3): p. 359-65.

39. Xenarios, I. and D. Eisenberg, *Protein interaction databases.* Curr Opin Biotechnol, 2001. **12**(4): p. 334-9.

40. Ono, T., H. Hishigaki, A. Tanigami, and T. Takagi, *Automated extraction of information on protein-protein interactions from the biological literature.* Bioinformatics, 2001. **17**(2): p. 155-61.

41. Marcotte, E.M., I. Xenarios, and D. Eisenberg, *Mining literature for protein-protein interactions.* Bioinformatics, 2001. **17**(4 :(p. 359-63.

42. Bader, G.D., I. Donaldson, C. Wolting, B.F. Ouellette, T. Pawson, and C.W. Hogue, *BIND--The Biomolecular Interaction Network Database.* Nucleic Acids Res, 2001. **29**(1): p. 242-5.

43. Xenarios, I., E. Fernandez, L. Salwinski, X.J. Duan, M.J. Thompson, E.M. Marcotte, and D. Eisenberg, *DIP: The Database of Interacting Proteins: 2001 update.* Nucleic Acids Res, 2001. **29**(1): p. 239-41.

44. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27.30-

45. Karp, P.D., M. Riley, M. Saier, I.T. Paulsen, S.M. Paley, and A. Pellegrini-Toole, *The EcoCyc and MetaCyc databases.* Nucleic Acids Res, 2000. **28**(1): p. 56-9.

46. Bairoch, A. and R. Apweiler, *The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000.* Nucleic Acids Res, 2000. **28**(1): p. 45-8.

47. Blaschke, C., M.A. Andrade, C. Ouzounis, and A. Valencia, *Automatic extraction of biological information from scientific text: protein-protein interactions.* Proc Int Conf Intell Syst Mol Biol, 1999: p. 60-7.

48. Yoshida, M., K. Fukuda, and T. Takagi, *PNAD-CSS: a workbench for constructing a protein name abbreviation dictionary.* Bioinformatics, 2000. **16**(2): p. 169-75.

49. Sekimizu, T., H.S. Park, and J. Tsujii, *Identifying the Interaction between Genes and Gene Products Based on Frequently Seen Verbs in Medline Abstracts.* Genome Inform Ser Workshop Genome Inform, 1998. **9**: p. 62-71.

50. Ng, S.K. and M. Wong, *Toward Routine Automatic Pathway Discovery from On-line Scientific Text Abstracts.* Genome Inform Ser Workshop Genome Inform, 1999. **10**: p. 104-112.

51. Wong, L., *PIES, a protein interaction extraction system.* Pac Symp Biocomput, 2001: p. 520-31.

52. Thomas, J., D. Milward, C. Ouzounis, S. Pulman, and M. Carroll, *Automatic extraction of protein interactions from scientific abstracts.* Pac Symp Biocomput, 2000: p. 541-52.

53. Humphreys, K., G. Demetriou, and R. Gaizauskas, *Two applications of information extraction to biological science journal articles: enzyme interactions and protein structures.* Pac Symp Biocomput, 2000: p. 505-16.

54. Mendelsohn, A.R. and R. Brent, *Protein interaction methods--toward an endgame.* Science, 1999. **284**(5422): p. 1948-50.

55. Pandey, A. and M. Mann, *Proteomics to study genes and genomes.* Nature, 2000. **405**(6788 :(p. 837-46.

56. Yates, J.R., 3rd, *Mass spectrometry. From genomics to proteomics.* Trends Genet, 2000. **16**(1): p. 5-8.

57. Kay, B.K., J. Kasanov, and M. Yamabhai, *Screening phage-displayed combinatorial peptide libraries.* Methods, 2001. **24**(3): p. 240-6.

58. Walhout, A.J., S.J. Boulton, and M. Vidal, *Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm.* Yeast, 2000. **17**(2): p. 88-94.

59. McAlister-Henn, L., N. Gibson, and E. Panisko, *Applications of the yeast two-hybrid system* .Methods, 1999. **19**(2): p. 330-7.

60. Uetz, P., L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M .Johnston, S. Fields, and J.M. Rothberg, *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. **403**(6770): p. 623-7.

61. Walhout, A.J., R. Sordella, X. Lu, J.L. Hartley, G.F. Temple, M.A. Brasch, N. Thierry-Mieg ,and M. Vidal, *Protein interaction mapping in C. elegans using proteins involved in vulval development.* Science, 2000. **287**(5450): p. 116-22.

62. Ito, T., K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, and Y. Sakaki, *Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1143-7.

63. Oliver, S., *Guilt-by-association goes global.* Nature, 2000. **403**(6770): p. 601-3.

64. Stagljar, I., C. Korostensky, N. Johnsson, and S. te Heesen, *A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo.* Proc Natl Acad Sci U S A, 19 :(9)95 .98p. 5187-92.

65. Aronheim, A., E. Zandi, H. Hennemann, S.J. Elledge, and M. Karin, *Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions.* Mol Cell Biol, 1997. **17**(6): p. 3094-102.

66. Broder, Y.C., S. Katz, and A. Aronheim, *The ras recruitment system, a novel approach to the study of protein-protein interactions.* Curr Biol, 1998. **8**(20): p. 1121-4.

67. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein, *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.

68. Pellegrini, M., E.M. Marcotte, M.J. Thompson, D. Eisenberg, and T.O. Yeates, *Assigning protein functions by comparative genome analysis:*

protein phylogenetic profiles. Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4285-8.

69. Marcotte, E.M., M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates, and D. Eisenberg, *Detecting protein function and protein-protein interactions from genome sequences.* Science, 1999. **285**(5428): p. 751-3.

70. Overbeek, R ,.M. Fonstein, M. D'Souza, G.D. Pusch, and N. Maltsev, *Use of contiguity on the chromosome to predict functional coupling.* In Silico Biol, 1999. **1**(2): p. 93-108.

71. Overbeek, R., M. Fonstein, M. D'Souza, G.D. Pusch, and N. Maltsev, *The use of gene clusters to infer functional coupling.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2896-901.

72. Dandekar, T., B. Snel, M. Huynen, and P. Bork, *Conservation of gene order: a fingerprint of proteins that physically interact.* Trends Biochem Sci, 1998. **23**(9): p. 324-8.

73. Mewes, H.W., D. Frishman, U. Guldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Munsterkotter, S. Rudd, and B. Weil, *MIPS: a database for genomes and protein sequences.* Nucleic Acids Res, 2002. **30**(1): p. 31-4.

74. Sanchez, C., C. Lachaize, F. Janody, B. Bellon, L. Roder, J. Euzenat, F. Rechenmann, and B. Jacq, *Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database.* Nucleic Acids Res, 1999. **27**(1): p. 89-94.

75. Hoebeke, M ,.H. Chiapello, P. Noirot, and P. Bessieres, *SPiD: a subtilis protein interaction database.* Bioinformatics, 2001. **17**(12): p. 1209-12.

76. Zanzoni, A., L. Montecchi-Palazzi, M. Quondam, G. Ausiello, M. Helmer-Citterich, and G. Cesareni, *MINT: a Molecular INTeraction database.* FEBS Lett, 2002. **513**(1): p. 135-40.

77. Costanzo, M.C., M.E. Crawford, J.E. Hirschman, J.E. Kranz, P. Olsen, L.S. Robertson, M.S. Skrzypek, B.R. Braun, K.L. Hopkins, P. Kondu, C. Lengieza, J.E. Lew-Smith, M. Tillberg, and J.I. Garrels ,*YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information.* Nucleic Acids Res, 2001. **29**(1): p. 75-9.

78. Proteome BioKnowledge Library website, at "http://www.incyte.com/sequence/proteome/index.shtml".

79. CYGD-MIPS website, at "http://mips.gsf.de/proj/yeast/CYGD/interaction/main.html".

80. Dujon, B., *European Functional Analysis Network (EUROFAN) and the functional analysis of the Saccharomyces cerevisiae genome.* Electrophoresis, 1998. **19** :(4)p. 617-24.

81. BIND website, at "http://www.bind.ca/index.phtml".

82. Xenarios, I., L. Salwinski, X.J. Duan, P. Higney, S.M. Kim, and D. Eisenberg, *DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions.* Nucleic Acids Res, 2002. **30**(1): p. 303-5.

83. DIP website, at "http://dip.doe-mbi.ucla.edu"/.

84. PIR - Protein Information Resource website, at "http://pir.georgetown.edu"/.

85. SWISS-PROT website, at "http://www.ebi.ac.uk/swissprot"/.

86. GenBank, NCBI website, at "http://www.ncbi.nlm.nih.gov/Genbank/index.html".

87. MINT website, at "http://cbm.bio.uniroma2.it/mint"/.

88. CuraGen website, at "http://portal.curagen.com"/.

89. PathCalling Yeast Interaction Database website, at "http://portal.curagen.com/extpc/com.curagen.portal.servlet.PortalYeastList".

90. Hybrigenics' PIM website, at "http://pim.hybrigenics.com/pimriderlobby/current/PimRiderLobby.htm".

91. Rain, J.C., L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F .Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, and P. Legrain, *The protein-protein interaction map of Helicobacter pylori.* Nature, 2001. **409**(6817): p. 211-5.

92. von Mering, C., R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork ,*Comparative assessment of large-scale data sets of protein-protein interactions.* Nature, 2002. **417**(6887): p. 399-403.

93. Legrain, P., J. Wojcik, and J.M. Gauthier, *Protein--protein interaction maps: a lead towards cellular functions.* Trends Genet, 2001 :(6)17 .p. 346-52.

94. Marcotte, E. and S. Date, *Exploiting big biology: integrating large-scale biological data for function inference.* Brief Bioinform, 2001. **2**(4): p. 363-74.

95. Visualizing protein-protein interaction Java applet's website, at "http//:www.charite.de/bioinformatics/interaction/index.html".

96. Mrowka, R., *A Java applet for visualizing protein-protein interaction.* Bioinformatics, 2001. **17**(7): p. 669-71.

97. Cytoscape website, at "http://www.cytoscape.org"/.

98. *Creating the gene ontology resource: design and implementation.* Genome Res, 2001. **11**(8): p. 1425-33.

99. Searls, D.B., *Bioinformatics tools for whole genomes.* Annu Rev Genomics Hum Genet, 2000. **1**: p. 251-79.

100. Gelbart, W.M., *Databases in genomic research.* Science, 1998. **282**(538 :(9p. 659-61.

101. Cormen, T.H., *Introduction to algorithms*. 2nd ed. 2001, Cambridge, Mass.: MIT Press. xxi, 1180 cm.

proteins and interactions at the same time, predict unknown interactions based on known interactions in other species, and come up with preliminary predictions about the function of unknown genes.

Keywords: bioinformatics, computational biology, protein-protein interaction, graph layout, functional genomics, function prediction, database navigation, evolutionary conservation.

Abstract 9

Until a few years ago, the main challenges facing geneticists were the identification and sequencing of genes that were responsible for pathogenic phenotypes. Genetic research was focused primarily on monogenic traits, and the search for the relevant gene was carried out mainly by positional cloning, or by using available knowledge of the biochemical function of the gene's product. Once the gene is identified, the detailed biochemical activity of its product has to be determined.

Advances in sequencing technology that enable the sequencing of complete genomes brought us to the post-genomic era in which genetic research takes a new course - discovering a gene's function from its raw sequence. The new emerging field of Functional Genomics attempts to assign a function to a given sequence, supplying a starting point for further biochemical study of the encoded protein. Efforts are made today to annotate genes of unknown function mainly on the basis of their functional relationship to already known genes.

The main objective of bioinformatics is to transform the vast amounts of raw data available today into comprehensible biological knowledge. One of the major challenges is to build user-friendly tools that give researchers fast and clear access to the information in the simplest, yet productive and unrestricting fashion.

This work describes the development of a bioinformatics software tool called PIVOT that was designed to help researchers view multiple protein-protein interactions simultaneously via a clear and flexible graphical presentation. The information is automatically processed to a graphical map, which the user can interactively manipulate, add or remove proteins, and request further information about particular proteins. The clear presentation of the data allows the researcher to observe many

Tel Aviv University

The Sackler Faculty of Medicine

Graduate School

Field: Bioinformatics

Title: **Development of a bioinformatic tool for interactive study of protein-protein interactions**

Submitted by Nir Orlev, ID. 025729328.

This work was carried out in partial fulfillment of the requirements for an M. Sc. Degree, in the Sackler Faculty of Medicine, Tel Aviv University.

Advisors: Prof. Yosef Shiloh

Prof. Ron Shamir

September 2002