

COMPUTATIONAL PREDICTION OF PROTEIN-PROTEIN INTERACTIONS IN  
SPHINGOLIPID SIGNALING NETWORK

by

Yasemen Güngörmez

B.S., Chemical Engineering, Yıldız Technical University, 2006

Submitted to the Institute for Graduate Studies in  
Science and Engineering in partial fulfillment of  
the requirements for the degree of  
Master of Science

Graduate Program in Chemical Engineering  
Boğaziçi University  
2010

*To my mother..*

## ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis advisor Prof. Kutlu Özergin Ülgen and co-advisor Assist. Prof. Elif Özkırımlı Ölmez for their support, counsel and motivation.

I am also grateful to the members of the thesis committee Prof. Z. İlsen Önsan, Prof. Türkan Haliloğlu, Assoc. Prof. Özlem Keskin for their time devoted to read and comment on my thesis.

I am also deeply grateful to my best friends Pınar Kanlıkılıçer and Ayşe Ezgi Akkaya for their priceless friendship, everlasting support, encouraging words, helping me out, being the most tolerant and understanding friends during my hardest times. I also thank to Özge Cinel for her never-ending help and friendship.

Heartful thanks are due to Betül Kavun Özbayraktar for her everlasting understanding and willingness to help for not only my studies but also daily life problems. I gratefully thank to Ahmet Özcan for his entertaining friendship and his precious help during my project.

I would like to thank to all my former and present friends in KB 440 and in KB 407: Aysun Eren, Kübra Büyükyanbolu, Saliha Durmuş, Esra Yücel, Elif Dereli, Ceyda Kasavi, and Burcu Özkaral, Celal Ceylan, Aslıgül Doğan, İhsan Ömür Akdağ, Tarık Can, Dilek Eren for sharing all the bad and good moments and making life easier during my work. I wish to thank Hacer Güneş, Arman Bamacıoğlu for their friendship and support.

I gratefully acknowledge, this thesis has been financially supported by YG 2210-TÜBİTAK Graduate Scholarships.

Last but certainly not least, I am forever indebted to my mother. Thank for her endless support, understanding, love and patience throughout my life. I feel lucky for having such a perfect mother. This thesis is dedicated to my mother.

## ABSTRACT

### COMPUTATIONAL PREDICTION OF PROTEIN-PROTEIN INTERACTIONS IN SPHINGOLIPID SIGNALING NETWORK

Proteins carry out most of the work in the cell such as immunological recognition, DNA repair and replication, enzymatic activity, cell signaling by interacting with other proteins. Therefore deciphering protein-protein interactions can give useful information about various mechanisms in the cell and hence provide important clues on disease states such as cancer and apoptosis. In this work, protein-protein interactions involved in sphingolipid metabolism were investigated in an effort to elucidate sphingolipid metabolism. Three categories of proteins were examined; proteins with previously identified partners, proteins with no known partners and clusters (Cluster A, Cluster B and Cluster C). First, the missing interaction partners of six proteins with no known interactions were identified by sequence based prediction methods and then filtered using the GO annotations of protein partners. The putative interaction partners were determined as; YHR135C with YDR294C; YHL020C with YER019W; YKL126W with YGR143W; YPL204W and YHR135C with YGR212W; YAR033W and YGL051W with YJL134W; YAR033W and YGL051W with YKR053C. The structures for these proteins were predicted by homology modeling and the structures of protein complexes were predicted by protein-protein docking. Nearly half of the complexes of the proteins with predicted partners formed biological contacts which mean these model interactions may occur in real systems. Next, the hotspots in every model of interacting protein pairs were identified by KFC. The model with the maximum number of closest hotspots was selected as the putative model structure for that protein complex. The hotspots were then classified according to their chemical features such as acidity, polarity, hydrophobicity and enrichment of certain amino acids. More than 60% of the hotspot residues in all categories of protein complexes are hydrophobic. The most repeated hotspot residue was found to be TYR in Clusters A and B, whereas it was LEU (15.04%) in Cluster C and in protein complexes with known interaction partners.

## ÖZET

### SFİNGOLİPİD SİNYAL AĞINDAKİ PROTEİN-PROTEİN ERKİLEŞİMLERİNİN HESAPSAL TAHMİNİ

Bağışıklık sisteminin sağlanması, DNA onarımı ve kopyalanması, enzimsel faaliyetler, proteinlerin diğer proteinlerle etkileşimi sonucunda oluşan hücre sinyallerinin iletimi gibi hücre içi birçok iş proteinler tarafından yürütülmektedir. Bu nedenle, protein etkileşim ağının aydınlatılması hücre içi çeşitli mekanizmalar hakkında bilgi vererek kanser ve apoptozis gibi hastalıklar hakkında önemli ipuçları sağlayacaktır. Bu çalışmada, sfingolipid metabolizmasını aydınlatmak amacıyla sfingolipid metabolizmasındaki protein-protein etkileşimleri araştırılmıştır. Proteinler üç kategoride incelenmiştir. Bunlar; eşleri daha önce tanımlanmış proteinler, eşleri bilinmeyen proteinler ve kümelerdir (Küme A, Küme B ve Küme C). İlk olarak, altı tane etkileşim eşi bilinmeyen proteinlerin etkileşim eşleri dizi tabanlı tahmin yöntemleri kullanılarak tanımlanmış, daha sonra bu sonuçlar GO açıklamalarından yararlanılarak filtrelenmiştir. Olası etkileşim eşleri; YHR135C ile YDR294C; YHL020C ile YER019W; YKL126W ile YGR143W; YPL204W ve YHR135C ile YGR212W; YAR033W ve YGL051W ile YJL134W; YAR033W ve YGL051W ile YKR053C olarak belirlenmiştir Yapılar bireysel proteinler için homoloji modellemesi ile, protein kompleksleri için ise protein-protein kenetlenmesi ile tahmin edilmiştir. Eşleri tahmin edilen protein komplekslerinin neredeyse yarısı biyolojik etkileşimler kurmakta, bu da bu model etkileşimlerinin gerçek hayatta da olabileceğini göstermektedir. Daha sonra, etkileşim halinde bulunan protein çiftlerinin her modeli için KFC ile işlevsel noktalar belirlenmiştir. Birbirine yakın işlevsel noktaları en fazla bulunan model, o protein kompleksi için olası model olarak belirlenmiştir. İşlevsel noktalar daha sonra kimyasal özelliklerine göre asidik, polar, hidrofobik ve belirli aminoasitlerle zenginleştirilmiş olarak sınıflandırıldı. Bütün kategorilerdeki protein komplekslerinin işlevsel noktalarının %60 dan fazlasının hidrofobik olduğu görüldü. En fazla tekrarlanan aminoasit Küme A ve Küme B de TYR; Küme C de ve etkileşim eşleri bilinen protein komplekslerinde LEU (%15.04) olarak bulunmuştur.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	iv
ABSTRACT.....	v
ÖZET.....	vi
LIST OF FIGURES.....	x
LIST OF TABLES .....	xii
LIST OF SYMBOLS / ABBREVIATIONS. ....	xvi
1. INTRODUCTION.....	1
1.1. Sphingolipids.....	1
1.2. Protein-Protein Interaction .....	1
1.3. Objectives and Plan of the Study .....	2
2. THEORETICAL BACKGROUND .....	4
2.1. Sphingolipids .....	4
2.2. Predicting Protein-Protein Interactions .....	5
2.2.1. Sequence Based Prediction Methods .....	6
2.2.2. Domain Based Prediction Methods .....	8
2.2.3. Structure Based Prediction Methods.....	8
2.3. Modeling Protein Structures .....	10
2.4. Structural Analysis of Interacting Proteins.....	16
2.5. Analysis of Hotspots.....	18
3. METHODS .....	20
3.1. Prediction of protein interaction partners .....	20
3.2. Modeling Individual Protein Structures.....	20
3.3. Modeling Structures of the Interacting Protein Pairs: Docking of Proteins .....	21
3.4. Biological Relevance by NOXclass .....	21
3.5. Identifying Interface Residues and Hotspots .....	22

4. RESULTS AND DISCUSSION .....	24
4.1. Protein-Protein Interactions in the Sphingolipid (SL) Pathway .....	24
4.1.1. Interaction Partners of Sphingolipid Pathway Proteins Identified Using the BIOGRID .....	24
4.1.2. Predictions of Interaction Partners for SL Pathway Proteins with No Known Partners.....	27
4.1.2.1. Interactions with YDR294C .....	28
4.1.2.2. Interactions with YER019W.....	29
4.1.2.3. Interactions with YGR143W .....	30
4.1.2.4. Interactions with YGR212W .....	32
4.1.2.5. Interactions with YJL134W.....	33
4.1.2.6. Interactions with YKR053C .....	34
4.1.3. Clusters / Modules .....	38
4.2. Prediction of Protein Structure .....	39
4.2.1. Proteins with known partners .....	39
4.2.2. Proteins with predicted partners .....	42
4.2.3. Clusters / Modules .....	43
4.3. Prediction of Complex Structures .....	46
4.3.1 Biological relevance of the predicted complex structures.....	46
4.4. Prediction of Hotspots .....	48
4.4.1. Proteins with known interaction partners.....	49
4.4.2. Proteins with predicted partners .....	50
4.4.3. Clusters / Modules .....	54
4.5. Analysis of Hotspots .....	57
4.5.1. Hydrophobicity .....	57
4.5.1.1. Proteins with Known Interaction Partners.....	58
4.5.1.2. Proteins with predicted interaction partners .....	68

4.5.1.3. Clusters/ Modules .....	70
5. CONCLUSIONS AND RECOMMENDATIONS.....	79
5.1. Conclusions.....	79
5.2. Recommendations .....	81
APPENDIX A: HOTSPOT DISTANCES .....	83
A.1. Proteins with Known Interaction Partners .....	83
A.2. Proteins with Predicted Partners.....	95
A.3. Clusters\Modules.....	98
REFERENCES.....	103

## LIST OF FIGURES

Figure 2.1.	Illustration of the four main steps in the PIPE algorithm (Pitre, S., et al., 2006) .....	7
Figure 2.2.	Steps in the comparative protein structure modeling (Eswar et al., 2007).	13
Figure 3.1.	Flow chart of the algorithm used to identified protein interaction network .....	23
Figure 4.1.	PIPE interaction graph for proteins YDR294C-YHR135C .....	29
Figure 4.2.	PIPE interaction graph for proteins YER019W-YHL020C.....	30
Figure 4.3.	PIPE interaction graph for proteins YGR143W-YKL126W .....	31
Figure 4.4.	PIPE interaction graph for proteins YGR212W-YPL204W .....	32
Figure 4.5.	PIPE interaction graph for proteins YGR212W-YHR135C .....	33
Figure 4.6.	PIPE interaction graph for proteins YJL134W-YAR033W.....	34
Figure. 4.7.	PIPE interaction graph for proteins YJL134W-YGL051W.....	34
Figure. 4.8.	PIPE interaction graph for proteins YKR053C-YAR033W .....	35
Figure. 4.9.	PIPE interaction graph for proteins YKR053C-YGL051W .....	35
Figure 4.10.	Complex structure of YDL015C-YBR094W (Model 10) .....	49
Figure 4.11.	Complex structure of YER093C-YKL203C (Model 8) .....	50

Figure 4.12. Complex structure of YGR212W-YPL204W (Model 8) .....	51
Figure 4.13. Complex structure of YDR294C-YHR135C (Model 7).....	51
Figure 4.14. Complex structure of YER019W-YHL020C (Model 10) .....	52
Figure 4.15. Complex structure of YGR143W-YKL126W (Model 3).....	52
Figure 4.16. Complex structure of YJL134W-YAR033W (Model 3) .....	53
Figure 4.17. Complex structure of YKR053C-YAR033W (Model 5).....	53
Figure 4.18. Complex structure of YOR171C-YOR034C (Model 4) in Cluster A .....	54
Figure 4.19. Complex structure of YOR034C-YOR171C (Model 6) in Cluster A .....	55
Figure 4.20. Complex structure of YMR298W-YMR298W (Model 3) in Cluster B....	55
Figure 4.21. Complex structure of YGR060W-YGR060W (Model 4) in Cluster B .....	56
Figure 4.22. Complex structure of YEL027W-YHR026W (Model 1) in Cluster C.....	56
Figure 4.23. Complex structure of YEL027W-YOR016C (Model 2) in Cluster C.....	57

## LIST OF TABLES

Table 2.1.	Qualification of hotspots (Moreira et al., 2007 and Orfan and Rost, 2007).....	17
Table 4.1.	Interaction partners of proteins in the SL pathway identified using the BIOGRID database ( <a href="http://www.thebiogrid.org/">http://www.thebiogrid.org/</a> ).....	24
Table 4.2.	Putative physical interaction partners of YDR294C .....	28
Table 4.3.	Putative physical interaction partners of YER019W .....	29
Table 4.4.	Putative physical interaction partners of YGR143W.....	30
Table 4.5.	Putative physical interaction partners of YGR212W.....	32
Table 4.6.	Putative physical interaction partners of YJL134W .....	33
Table 4.7.	Putative physical interaction partners of YKR053C .....	35
Table 4.8.	GO annotations of Putative interaction partners. IDA: Inferred from Direct Assay, IGI: Inferred from Genetic Interaction, IMP: Inferred from Mutant Phenotype, ISS: Inferred from Sequence or Structural Similarity, IPI: Inferred from Physical Interaction .....	36
Table 4.9.	Proteins in Cluster A and their interaction partners.....	38
Table 4.10.	Proteins in Cluster B and their interaction partners.....	38
Table 4.11.	Proteins in Cluster C and their interaction partners.....	38

Table 4.12.	The templates used in building modeling SL proteins with known partners (Set A). The coverage, sequence, identity and E-value of the alignment are also listed. ....	40
Table 4.13.	The templates used in building models of interaction partners of set A. The coverage, sequence, identity and E-value of the sequence alignment are also listed. ....	40
Table 4.14.	The templates used in building modeling SL proteins with predicted partners (Set B). The coverage, sequence, identity and E-value of the alignment are also listed. ....	42
Table 4.15.	The templates used in building models of interaction partners of set B. The coverage, sequence, identity and E-value of the sequence alignment are also listed. ....	43
Table 4.16.	The templates used in building modeling SL proteins in Cluster A. The coverage, sequence, identity and E-value of the alignment are also listed. ....	44
Table 4.17.	The templates used in building modeling SL proteins in Cluster B. The coverage, sequence, identity and E-value of the alignment are also listed. Template-region-coverage values of Cluster B ....	44
Table 4.18.	The templates used in building modeling SL proteins in Cluster C. The coverage, sequence, identity and E-value of the alignment are also listed. ....	45
Table 4.19.	The templates used in building modeling interaction partners of Cluster C. The coverage, sequence, identity and E-value of the alignment are also listed. ....	45
Table 4.20.	Biological relevance of proteins with predicted partners. ....	47

Table 4.21.	Chemical types of proteins with known interaction partners .....	58
Table 4.22.	Amino acid propensities of hotspots in proteins with known interaction partners.....	67
Table 4.23.	Distribution of the hotspot residues of proteins with known interaction partners according to chemical features .....	68
Table 4.24.	Chemical types of proteins with predicted interaction partners .....	68
Table 4.25.	Amino acid propensities of hotspots in proteins with predicted Interactions .....	70
Table 4.26.	Distribution of hotspot residues of proteins with predicted interactions partners according to chemical types.....	70
Table 4.27.	Chemical types of hotspot residues of proteins in Cluster A .....	71
Table 4.28.	Amino acid propensities of hotspots of proteins in Cluster A.....	72
Table 4.29.	Distribution of hotspot residues of proteins in Cluster A according to the chemical features .....	72
Table 4.30.	Chemical types of hotspots residues of proteins in Cluster B .....	73
Table 4.31.	Amino acid propensities of hotspots of proteins in Cluster B.....	74
Table 4.32.	Distribution of hotspot residues of proteins in Cluster B according to the chemical features .....	74
Table 4.33.	Chemical types of hotspots residues of proteins in Cluster C .....	76

Table 4.34.	Amino acid propensities of hotspots of proteins in Cluster C.....	77
Table 4.35.	Distribution of hotspot residues of proteins in Cluster C according to the chemical features.....	76
Table 4.36.	Percentage of Chemical types of Hotspots.....	78
Table A.1.	The protein pair, hotspots on each protein, and the distances between hotspot residues in proteins with known partners .....	83
Table A.2.	The protein pair, hotspots on each protein, and the distances between hotspot residues in proteins with predicted partners .....	95
Table A.3.	The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster A for 10 models .....	98
Table A.4.	The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster B .....	100
Table A.5.	The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster C .....	101

## LIST OF SYMBOLS / ABBREVIATIONS

3D	Three-dimensional
AvgAF	Average score after filtering
BIND	Biomolecular Interaction Network Database
BIOGRID	Database of Protein and Genetic Interactions
BLAST	Basic Local Alignment Search Tool
DBID	Database of interacting domains
DIP	Database of interacting proteins
DNA	Deoxyribonucleic acid
EDG	Endothelial-differentiating gene
ExpPDB	Expert Protein Data Bank
FASTA	Fast All
FFAS03	Fold & Function Assignment System
GO	Gene Ontology
HOMCOS	Homology Modeling of Complex Structure
InterpreTS	Predicting protein-protein interactions through tertiary structure
IMPALA	Integral Membrane Protein and Lipid Association
IPPRED	Server for proteins interactions inference
K-CON	Knowledge-based Biochemical Contacts analysis
K-FADE	Knowledge-based Fast Atomic Density Evaluation
KFC	Knowledge-based FADE and Contacts
LCB	Long-chain base
MINT	Molecular Interactions Database
MIPS	Munich Information Center for Protein Sequences
MSA	Multiple sequence alignment
NMR	Nuclear magnetic resonance
ORF	Open Reading Frame
PDB	Protein data bank
PIPE	Protein-protein interactions engine
PIPS	Protein-protein Interaction Prediction Server
PPIs	Protein-protein interactions

PreSPI	Prediction system for protein-protein interactions
PRISM	Protein Interactions by Structural Matching
PSI-BLAST	Position-Specific Iterated Basic Local Alignment Search Tool
RasMol	Raster Molecules
RMSD	Root mean square deviation
SL	Sphingolipid
SVM	Support vector machine
TAP	Tandem affinity purification
VMD	Visualization of Molecular Dynamics
Y2H	Yeast-two-hybrid

## **1. INTRODUCTION**

### **1.1. Sphingolipids**

Sphingolipids are a structurally diverse group of molecules based on long-chain sphingoid bases (Worrall et al., 2003). The structural components of sphingolipids are classified into three parts: a sphingoid long-chain base (LCB), which forms the backbone of the sphingolipids, a fatty acid, and a polar head group (Ozbayraktar and Ulgen, 2009).

Sphingolipid signaling was first implicated in human cancers and then their different roles in other human illnesses such as diabetes and heart disease, microbial infections, neurological disorders including Alzheimer's and immune dysfunctions have been realized (Dickson, 2008). To learn how and why the sphingolipid metabolism has evolved and to elucidate mammalian sphingolipid signaling with the information from simple model organisms is very important. Gaining knowledge on sphingolipid metabolism, which induces apoptosis in cancer cells, is expected to improve the efficiency of cancer therapeutics (Ozbayraktar and Ulgen, 2009).

### **1.2. Protein-Protein Interaction**

Proteins carry out the majority of the biological processes in cells. It is generally accepted that protein-protein interactions (PPIs) are responsible for the cell's behavior and its responses to various stimuli (Pitre et al., 2006). Characterizing interacting protein pairs and interaction sites is necessary to fully understand the molecular mechanism of cellular activities (Fukuhara and Kawabata, 2008). The detailed knowledge of the full network of protein-protein interactions, i.e. the distribution and the number of interactions as well as the presence of key nodes in these networks, is expected to provide new insights into the structures and properties of biological systems (Aytuna et al., 2005).

### 1.3. Objectives and Plan of the Study

Proteins carry out most of the work in the cell such as immunological recognition, DNA repair and replication, enzymatic activity, cell signaling by interacting with other proteins. Therefore, predicting PPI network will assist understanding molecular mechanism of cellular activities. Deciphering PPI networks can provide insight into human diseases and ultimately for the intelligent design of therapeutics. To this end, in this study, the network of sphingolipid (SL) signaling proteins was constructed using computational prediction to contribute to the list of missing interactions among the components of sphingolipid PPI network by using new advances in PPI predictions. Protein-protein interactions and interaction regions were predicted by sequence-based (use only sequence information such as PIPE) and structure-based methods (uses structure information such as NOXclass) and the properties of these interactions were analyzed. Calculations of this study were carried out in three groups which are proteins with known partners, proteins with no known partners and proteins in clusters (Cluster A, Cluster B, Cluster C. The new predictions identified by this research can guide the rational design of new experiments.

In this study, protein-protein interactions in sphingolipid signaling pathway of *Saccharomyces cerevisiae* were identified. The yeast *Saccharomyces cerevisiae* is a model system representing a simple eukaryote whose genome can be easily manipulated. Yeast has only a slightly greater genetic complexity than bacteria, and they share many of the technical advantages that permitted rapid progress in the molecular genetics of prokaryotes and their viruses. Some of the properties that make yeast particularly suitable for biological studies include rapid growth, dispersed cells, the ease of replica plating and mutant isolation, a well defined genetic system, and most important, a highly versatile DNA transformation system. Being nonpathogenic, yeast can be handled with little precautions (Sherman, 2002). The reason for choosing *S. cerevisiae* as a model organism is that all the genes encoding the enzymes in sphingolipid metabolism are known in *S. cerevisiae* and most of the enzymes of this organism which are acting in sphingolipid pathway have homologs or orthologs in mammalian species (Ozbayraktar and Ulgen, 2009). This homology is expected to help illuminate the sphingolipid metabolism in humans. It is believed that elucidation of sphingolipid metabolism can provide information towards

understanding apoptosis, cancer and many other critical cellular processes coupled with major diseases.

## 2. THEORETICAL BACKGROUND

### 2.1. Sphingolipids

There are more than 300 sphingolipid molecular species which exist in all eukaryotic cells and which are enriched in plasma membranes, Golgi Apparatus membranes and lysosomes (Ohanian et al., 2001). Lipids provide structure to cell membranes, which produce a long lasting obstacle between extracellular and intracellular compartments through their ability to form a bilayer. It is believed that microdomains called lipid rafts are responsible for membrane-localized signaling by organization of components.. Lipids are essential for signal transduction in response to agonist stimulation as their hydrolysis produces bioactive molecules known to trigger many downstream signaling cascades (Ohanian et al., 2001). At the level of whole multicellular organism, signaling controls growth and development together with the aspects of metabolism and behavior (Durmuş, 2006).

Sphingolipids are characterized by their sphingoid backbone. In yeast and plant cells, phytosphingosine is the most common sphingoid base while in mammalian cells sphingosine is more common (Ohanian et al., 2001). In yeast and mammalian cells, research on sphingolipids and their metabolites, ceramide, sphingosine and sphingosine-1-phosphate is a growing field due to their participation in the control of different cellular phenomena such as growth, cell proliferation, stress responses, differentiation and apoptosis (Worrall et al., 2003). Ceramide, which is a second messenger, is produced by sphingomyelinase-induced hydrolysis of sphingomyelin and by de novo synthesis and its cellular levels are increased with many stimuli such as growth factors, cytokines, G protein-coupled receptor agonists and stress (UV irradiation) (Ogretmen, 2006). Other metabolites of sphingolipids such as sphingosine and sphingosine-1-phosphate which are obtained from ceramide are also signaling molecules. Sphingosine-1-phosphate can act extracellularly through endothelial-differentiating gene (EDG) family G protein-coupled receptors and extracellularly through direct interactions with target proteins and it especially participates in proliferation, differentiation and apoptosis (Ohanian et al., 2001).

## 2.2. Predicting Protein-Protein Interactions

Various experimental methods have been developed to identify protein–protein interactions in various organisms. These involve the traditional top-down proteomic approach where the experiments have been individually designed to identify and validate a small number of specifically targeted interactions or the bottom-up genomic approach, the recently developed high-throughput experiments designed to probe all the potential interactions within entire genome exhaustively (Shen et al., 2006). A number of experimental methodologies are available for solving the three-dimensional (3D) structure of protein interactions; X-ray crystallography; nuclear magnetic resonance (NMR) spectroscopy; electron microscopy and electron tomography, yeast-two-hybrid (Y2H) method, tandem affinity purification (TAP), gene co-expression and synthetic lethality (Fukuhara and Kawabata, 2008; Cockell et al., 2006). Current PPI pairs obtained with experimental methods cover only a fraction of the complete PPI networks therefore computational methods can provide useful information for the prediction of PPIs (Shen et al., 2006).

Recent advances in high-throughput experimental methods for the identification of protein interactions have resulted in a large amount of diverse data that are somewhat incomplete and contradictory (Shoemaker and Panchenko, 2007). The major need now is to explore this disparate data to find biologically relevant interactions and pathways (Tien et al., 2003). Thus in the post-genomic era, a body of research has been allocated to the development of algorithms that can accurately predict novel protein-protein interaction networks *in silico*. Computational methods can address protein-protein interactions at different levels (Aytuna et al., 2005). Genomic and protein sequence analysis can be useful to infer whether proteins do interact, or structural analysis of proteins and their complexes may provide interaction details essential for understanding processes at the microscopic level (Aytuna et al., 2005). Computational methods used for the prediction of protein interaction partners can be classified into two groups based on their functional associations and physical interactions. Gene fusion, gene neighbourhood and phylogenetic profiles predict functional associations between proteins (Cockell et al., 2006). There are several approaches to the computational prediction of physical interactions which are; (1) finding and analyzing subsequences affecting protein-protein interactions from raw protein

sequences, (2) prediction of protein interactions by analyzing the physicochemical properties or tertiary structure of proteins, (3) domain-based protein-protein interaction prediction (Han et al., 2004). Recent advances in the prediction of new physical associations between proteins have adopted a threading based approach (Cockell et al., 2006). *In silico* methods have the advantage of being fast in comparison with the experimental techniques and allow mechanistic details that are sometimes difficult to trace experimentally to be highlighted (Rousseau and Schymkowitz, 2005). Examples of computational protein-protein interaction prediction methods and their web servers are mentioned below.

### 2.2.1. Sequence Based Prediction Methods

Prediction of protein-protein interactions based only on sequence information is an ideal approach for both the computational and experimental senses (Shen et al., 2006). On one hand, they are apt to encounter the problem of overfitting and results that are data-dependent; on the other hand, these methods have not been used to predict PPI networks among many proteins (Shen et al., 2006). Methods using protein genomic and protein sequence data include analysis of presence or absence of genes in related species, conservation of gene neighbourhood, gene fusion events, similarity of phylogenetic trees, correlated mutations on protein surfaces and co-occurrences of sequence domains (Aytuna et al., 2005).

IPPRED is a web based server to infer protein-protein interactions through homology search between candidate proteins and those described as interacting. BLAST program is used for similarity searches and subsequent prediction of homology. A list of inferred interactions between candidate proteins is the output of the IPPRED (Gaffard et al., 2002).

PIPE is capable of predicting protein-protein interactions for any target pair of the yeast, *Saccharomyces cerevisiae*, proteins from primary structure and without the need for any additional information about the proteins. The prediction algorithm relies on previously determined interactions (Pitre, S., et al., 2006). PIPE is able to identify some interactions which cannot be identified by other methods (Pitre, S., et al., 2006). This tool is optimized to predict the likelihood of an interaction between a given pair of proteins.

The PIPE method predicts the likelihood of interaction between two query proteins A and B by measuring how often pairs of subsequences in A and B co-occur in pairs of protein sequences in the dataset that are known to interact. The principle of this method is as follows: assume to have two query proteins A and B, along with the knowledge that certain proteins C and D are interacting. If a region (subsequence)  $a_1$  in A resembles a region in C, and a sequence  $b_1$  in B resembles a region in D, there is a possibility that A and B are also interacting via an interaction between the corresponding  $a_1$  and  $b_1$  sequences, which co-occur in both protein pairs A-B and C-D (Figure 2.1).

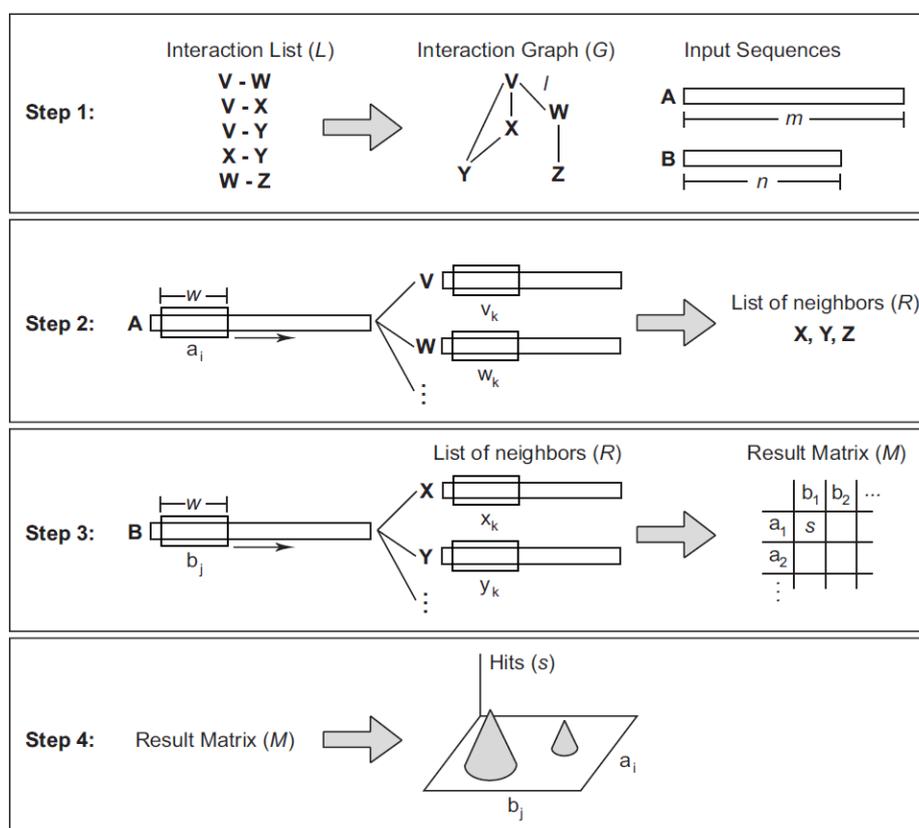


Figure 2.1. Illustration of the four main steps in the PIPE algorithm (Pitre, S., et al., 2006)

PIPE's sensitivity is calculated as  $(TP/(TP+FN))$  [%], its specificity as  $(TN/(TN+FP))$  [%], and its accuracy as  $((TP+TN)/(TP+FN+FP+TN))$  [%] where TP is the number of true positive, FN the number of false negatives, TN the number of true negatives, and FP the number of false positives.

### 2.2.2. Domain Based Prediction Methods

Domains are conserved modular structures important for expressing protein functions (Dohkan et al., 2006). Predicted domain interactions are evaluated using structural data or higher quality interaction sets such as MIPS (Shoemaker and Panchenko, 2007). Accounting for domains in proteins and domain interaction networks can in turn help in predicting protein interactions (Shoemaker and Panchenko, 2007). However, most techniques have many limitations and they usually suffer from low accuracy in prediction and do not provide any interaction possibility ranking method for multiple protein pairs (Han et al., 2004).

PreSPI is a domain combination based prediction system for protein-protein interactions. The validity of the prediction model was evaluated using an interacting set of the protein pairs in yeast and artificially generated non-interacting set of protein pairs. Although the proposed domain combination based prediction method certainly improves the prediction accuracy of the conventional domain based prediction method, it is not without limitations, because domain cannot explain all the details of complex protein-protein interactions and accumulated data are insufficient and erroneous (Han et al., 2004).

Protein-protein Interaction Prediction Server (PIPS) can predict physical protein-protein interactions between yeast, mouse, and human proteins. This support vector machine (SVM)-based method predicts interactions on the existence/absence of domains and amino acid compositions of proteins (Dohkan et al., 2006). This method can be used to extract likely interactions from high-throughput interactions, which is an important problem in obtaining reliable interaction maps (Dohkan et al., 2006).

### 2.2.3. Structure Based Prediction Methods

Through structural analysis of proteins and their complexes, they may provide interaction details, essential for understanding processes at the microscopic level (Aytuna et al., 2005). Methods making use of structural data, usually strive to identify functional protein interfaces and rely on considerations of the solvent accessible surface area buried upon association, free energy changes upon alanine-scanning mutations, *in silico* two-

hybrid systems, scoring functions based on statistical potentials, physicochemical and genomic properties of the surface, such as electrostatics, hydrophobicity, amino acid composition, shape complementarity and planarity and evolutionary conservation (Aytuna et al., 2005). Structure based protein interaction prediction methods can be of three types, i.e. secondary structure based, tertiary structure based and quaternary structure based. Only tertiary structure based web servers will be mentioned below.

InterpreTS is a web-based version of method of predicting protein-protein interactions through tertiary structure. Given a pair of query sequences, homologues in a database of interacting domains (DBID) of known three-dimensional complex structures were searched. The method assesses the fit of two potential interacting partners on a complex of known 3D structure and infers molecular details of how the interaction is likely to occur (i.e. which residues are in contact) (Aloy and Russell, 2002).

The 3D-partner is a web tool to predict interacting partners and binding models of a query protein sequence through structure complexes and a new scoring function. The 3D-partner utilizes IMPALA and PSI-BLAST to identify homologous structures (templates) and interacting partners of a query protein sequence from a 3D-dimer template library and protein sequence databases respectively. These homologous structures and interacting partners were evaluated by a scoring function which considered steric and special-bond matrices (i.e. hydrogen bonds, electrostatic interactions and disulfide bonds) but also the template consensus scores (couple-conserved residue score and template similarity) (Chen et al., 2007).

HOMCOS (Homology Modeling of Complex Structure) is a server which uses homology modeling of complex structures to predict interacting protein pairs and interacting sites. The server has three services which are heterodimers modeling from two query amino acid sequences which are taken from the users, homodimers modeling from one query sequence, and potentially interacting proteins identification from one query sequence respectively. For prediction of new interactions and their interacting sites, homology to a known 3D structure protein complex is known as the most successful tool. Although this tool assumed all the homologues protein pairs interact in a same way there are some exceptions. For instance, proteins belonging to multigene families often show

different interaction specificities, even if their sequence similarity is high and homologous interacting protein pairs sometimes show completely different interacting structural topologies. In spite of having simple concept, HOMCOS can meet a wide range research needs because of the updated dimer database and various output types for model complexes (Fukuhara and Kawabata, 2008).

Prism is a website for protein interface analysis and prediction of putative protein–protein interactions which can also give summary information about related proteins and an interactive protein interface viewer. The algorithm of the server is that, if any two structures contain particular regions on their surfaces that resemble the complementary partners of a known interface, they ‘possibly interact’ through these regions. In order to measure the structural similarity of a target structure to a template, binding site surfaces of target protein are extracted and successive structural alignments between these surfaces and the partner chains of interfaces in the template interface dataset are performed. If the surfaces of two target proteins (A and B) contain regions ‘similar’ to complementary partner chains of a template interface, it is possible to say that A and B may interact through these ‘similar’ regions. Also the presence of the hotspots on the target structure is checked to use the match ratio for the calculation of an ‘evolutionary similarity score’. Combination of these scores contributes to the overall prediction score. In the ‘Predictions’ section of PRISM, users can obtain results in two different ways which are searching for the presence of similarities between a template interface and a target structure, and entering the PDB ID or the sequence of a protein and checking for any predicted protein–protein interactions that the input protein participates in (Ogmen et al., 2005)

### **2.3. Modeling Protein Structures**

Three-dimensional (3D) protein structures provide valuable insights into the molecular basis of protein function, allowing an effective design of experiments, such as site-directed mutagenesis, studies of disease-related mutations or the structure based design of specific inhibitors (Schwede et al 2003). In the absence of an experimentally determined structure, comparative or homology modeling can be used to build a 3D model of the protein if the protein is related to at least one known protein structure. This method predicts the 3D structure of a given protein sequence (target) based primarily on its

alignment to one or more proteins of known structure (templates) (Eswar et al., 2007). Comparative modeling consists of four main steps; (i) fold assignment; which identifies similarity between the target and at least one known template structure (Eswar et al., 2007). This step is facilitated by numerous protein sequence and structure databases, and web available database scanning software. For fold identification there are three main classes of useful protein comparison methods. In the first class, the target sequence is compared with each of the sequences in databases independently, using the pair wise sequence-sequence comparison. FASTA and BLAST are the mostly used programs in this class. In the second class, multiple sequence comparison methods are used to improve the sensitivity of the search, and PSI-BLAST which iteratively expands the set of homologs of the target sequence is the widely employed tool (Marti-Renom et al., 2002). The third class of methods is 3D template matching methods which rely on pair wise comparison of a protein sequence and a protein of known structure. When there are no sequences clearly related to the modeling target, these methods are very useful. (ii) Alignment of the target sequence and the template(s); after the template is selected, alignment method is used to align the target sequence with the template structures. If the identity is higher than 40% for closely related protein sequences, the alignment is almost always true. Regions of low local sequence similarity become common when the overall sequence identity is below 40%. As the sequence similarity decreases, alignments contain an increasingly large number of gaps and alignment errors, regardless of whether they are prepared automatically or manually (Marti-Renom et al., 2002). (iii) Building a model based on the alignment with the chosen template(s); the original and still widely used method to construct a 3D model for the target protein is modeling by rigid body assembly. Modeling by segment matching relies on the approximate positions of conserved atoms in the templates is another method. The third group of methods, modeling by satisfaction of spatial restraints, uses either distance geometry or optimization techniques to satisfy spatial restraints obtained from the alignment. (iv) Predicting model error; the first step in model evaluation is to determine whether the model has the correct fold. If the correct template is picked and if that template is aligned at least approximately correctly with the target sequence, the model has correct fold. Once the fold of a model is accepted, a more detailed evaluation of the overall model accuracy can be obtained based on the similarity between the target and template sequences. A sequence identity above 30% is considered to be relatively good predictor of the expected accuracy. If the target-template sequence identity

falls below 30%, the sequence identity becomes unreliable as a measure of expected accuracy of a single model (Marti-Renom et al., 2002).

MODELLER is one of the programs for modeling protein structure with comparative protein structure modeling. The inputs given to MODELLER are the amino acid sequences of the proteins to be modeled, their alignment with template structures and the atomic coordinates of the templates (Figure 2.2). Once a target- template alignment is constructed, MODELLER calculates automatically a 3-D model containing all non-hydrogen atoms of the target, within minutes with no user invention (Eswar et al., 2007). In the template selection part generally, a sequence identity value above ~ 25% indicates a potential template, unless the alignment is too short (i.e., < 100 residues). If several models are calculated for the same target, the best model can be selected by picking the model with the lowest value of the MODELLER objective function. The value of the objective function in MODELLER is not an absolute measure, in the sense that it can only be used to rank models calculated from the same alignment (Eswar et al., 2007). Once a final model is selected, there are many ways to assess it, such as DOPE program.

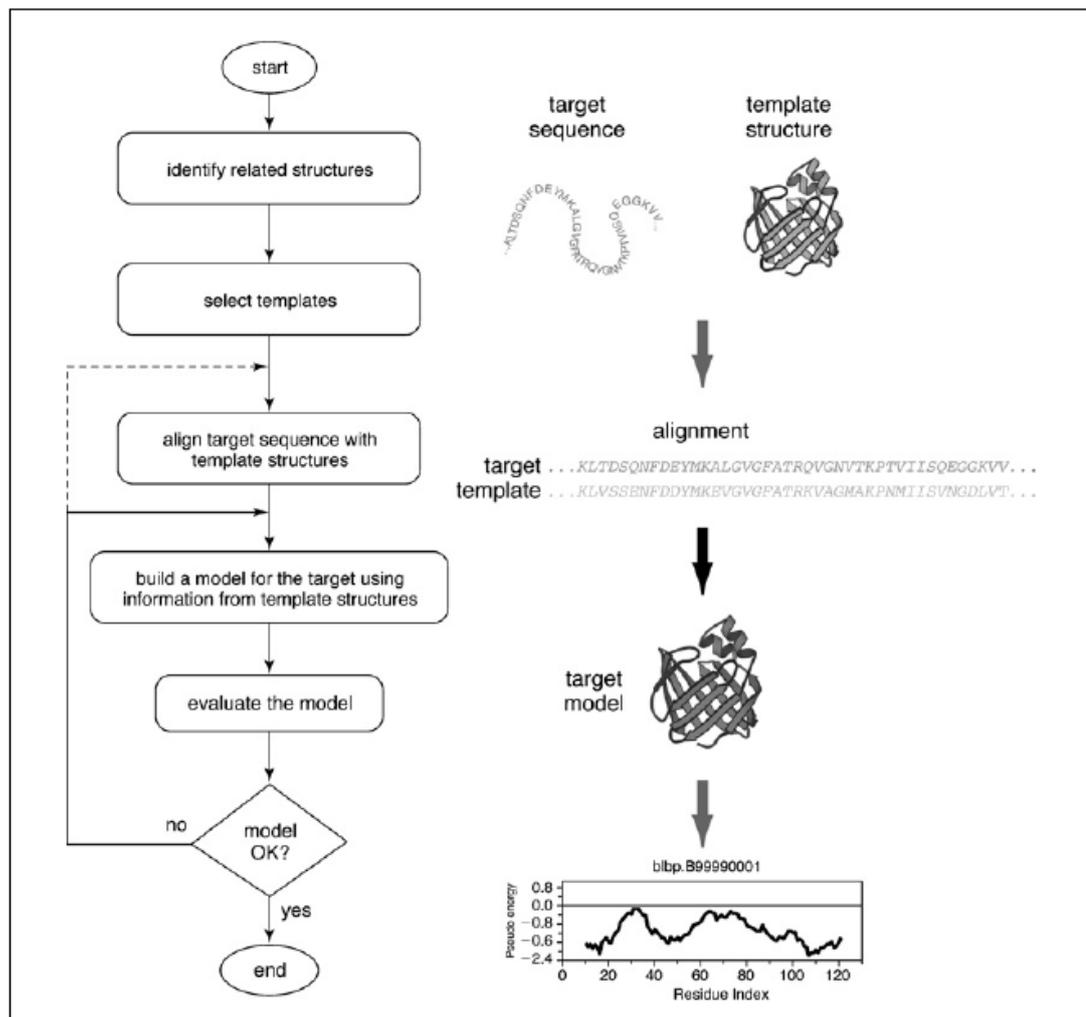


Figure 2.2. Steps in the comparative protein structure modeling (Eswar et al., 2007).

SWISS-MODEL is an automated comparative modeling server to model three dimensional (3D) protein structures (Schwede et al., 2003). SWISS-MODEL provides several levels of modes. In the ‘first approach mode’ only an amino acid sequence of a protein is submitted to build a 3D model. This mode provides a simple interface and requires only an amino acid sequence as input data. The server will automatically select suitable templates. Optionally, the user can specify up to five template structures, either from the ExpDB library or uploaded coordinate files. The automated modeling procedure will start if at least one modeling template is available that has a sequence identity of more than 25% with the submitted target sequence. However, the model reliability decreases as the sequence identity decreases and that target-template pairs sharing less than 50% sequence identity may often require manual adjustment of the alignment (Schwede et al., 2003). In the ‘alignment mode’, the modeling process is based on a user-defined target

template alignment. In this mode the modeling procedure is initiated by submitting a sequence alignment. The user specifies which sequence in the given alignment is the target sequence and which one corresponds to a structurally known protein chain from the ExpDB template library. The server builds the model based on the given alignment. The 'project mode' allows the user to submit a manually optimized modeling request to the SWISS-MODEL server. Complex modeling tasks can be handled with the 'project mode'. The starting point for this mode is a DeepView project file. It contains the superposed template structures, and the alignment between the target and the templates. This mode gives the user control over a wide range of parameters, e.g. template selection or gap placement in the alignment. Furthermore, the project mode can also be used to iteratively improve the output of the 'first approach mode'. The accuracy of a model can vary significantly, even within different regions of the same protein: usually highly-conserved core regions can be modeled much more reliably than variable loop regions or surface residues. SWISS-MODEL result files consist of a C-score, which gives an estimate of the variability of the template structures at this position. Parts of the model where no template information could be used for model building (insertions or deletions) are assigned a C-score of 99. A detailed log file listing all steps performed by the modeling server is provided to the user. This includes force field energy for the overall structure and for each individual residue to identify regions with obvious conformational or electrostatic problems (Schwede et al., 2003).

Robetta is an internet based server that provides automated structure prediction and analysis tools that can be used to infer protein structural information from genomic data (Kim et al., 2004). The ultimate goal of Robetta is to provide structural information of sufficient quality to aid research, infer function and assist drug design. For structure prediction, Robetta parses input sequences into domains and builds structure models in two ways. If a confident match to a protein of known structure can be found with sequence homology it uses comparative modeling; if cannot, it uses the Rosetta *de novo* structure prediction method. Robetta uses a fully automated and slightly modified implementation of the Rosetta software package for protein structure prediction. However, as the original protocol, Robetta generates three- and nine-residue fragment libraries that represent local conformations seen in the PDB, and then assembles models by fragment insertion using a scoring function that favors protein-like features (Kim et al., 2004). In an attempt to predict

structures for full-length protein sequences, Robetta uses a domain prediction method called 'Ginzu' as the initial step for structure prediction. Ginzu is a hierarchical screening procedure that first uses BLAST, PSIBLAST, FFAS03 and 3D-Jury to detect regions in the query sequence that are homologous to experimentally determined structures, and then proceeds with multiple sequence alignment (MSA) based methods to predict putative domains (Kim et al., 2004). If a query is parsed into multiple domains, the final step in this structure prediction procedure is to assemble the domain models into a continuous full-length structure. Robetta uses an iterative domain assembly protocol that starts with the N-terminal domain, and attempts domain association by fragment insertion in the putative linker region assigned by Ginzu using the same scoring method as in de novo protocol. If the chain contains more than two domains, the third domain is added to the previously assembled model, and the procedure continues until the whole chain is assembled. The procedure identifies residues that are involved in the protein-protein interface, and uses a simple free energy function to calculate the changes in the binding free energy upon single substitutions of each side-chain to alanine. Domain predictions and molecular coordinates of models spanning the full length query are given as results. The server can also utilize nuclear magnetic resonance (NMR) constraints data provided by the user to determine protein structures using the RosettaNMR protocol. The de novo protocol is optimized for small single domain proteins (<120 residues) (Kim et al., 2004). Within this limit, models are frequently around 3–7 Å RMSD to more than half of the native structure. Above this limit, models are still likely to have at least 50 residues within 4 Å RMSD. Quality of the model is greatly dependent on the correct selection of the best possible parent template and alignment. Because of these factors, results are highly dependent on the accuracy of the domain assignments.

GRAMM-X is a protein docking web server and its web interface extend the original GRAMM Fast Fourier Transformation methodology by employing smoothed potentials, refinement stage, and knowledge based scoring (Tovchigrechko and Vakser, 2006). Recent progress in docking algorithms and computer hardware makes it possible to implement such procedures as automated web servers. While a number of existing docking methods based on scanning of large database, such as PSI-BLAST searches for evolutionary conserved residues (Tovchigrechko and Vakser, 2006), this method uses a fine-grid

projection of a softened Lennard-Jones potential function calculated for a probe atom (Tovchigrechko and Vakser, 2006):

$$V_{ij}(r) = \frac{1}{\alpha\sigma_{ij}^6 + r^6} \left( \frac{4\varepsilon_{ij}\sigma_{ij}^{12}}{\alpha\sigma_{ij}^6 + r^6} - 4\varepsilon_{ij}\sigma_{ij}^6 \right) \quad (2.1)$$

An important feature of GRAMM is the ability to smooth the protein surface representation to account for possible conformational change upon binding within the rigid body docking approach. The degree of conformational change of the input structures is the main factor that affects the quality of the docking prediction, especially the degree of such change at the interface area is very important (Tovchigrechko and Vakser, 2006). The usage of the server starts with a simple web interface which accepts two PDB protein structures from the user, forms a job request and submits it to the execution queue on the cluster. A temporary page is created by the web server for the future simulation results. The output PDB file contains 10 models ranked as the most probable prediction candidates according to our scoring function.

## 2.4. Structural Analysis of Interacting Proteins

Structural analysis of interacting proteins focus on physicochemical properties such as interface topology, accessible surface area, hydrophobicity, free energy changes for the stability of the complex. There are several methods explained below which are used for structural analysis of interacting proteins.

NOXclass combines six interface properties (interface area, ratio of interface area to protein surface area, amino acid composition of the interface, correlation between amino acid compositions of interface and protein surface, interface shape complementarity, and conservation of the interface) using a support vector machine algorithm, resulting in NOXclass, which is used as a classifier for distinguishing obligate, non-obligate and crystal packing interactions. NOXclass allows the interpretation and analysis of protein quaternary structures. In particular, it generates testable hypotheses regarding the nature of protein-protein interactions, when experimental results are not available (Zhu et al., 2006).

One other analyzing method of protein interactions is finding surface residues and hotspots. The stability of protein complexes is mediated by a collection of biophysical properties therefore hot spot searches typically focus on mutations that disrupt hydrogen bonds, van der Waals contacts and chemical complementarity. Hot spot identification requires the experimental characterization of a mutation's effect on binding affinity. Hotspots can be specified corresponding to their qualifications listed in Table 2.1.

Table 2.1. Qualification of hotspots (Moreira et al., 2007 and Orfan and Rost, 2007)

Area		Energy		Residues	Hydrophobicity	O-ring structure
Low	1150–1200 Å <sup>2</sup>	Low	< 2 kcal/mol	Tryptophan (21%), Arginine (13.3%), Tyrosine (12.3%).	Frequently hydrophobic and buried	Hot spots are usually surrounded by residues not important for binding, whose role would be to shelter the hot spots from the solvent
Standard	1200–2000 Å <sup>2</sup>	Standard	2 kcal/mol		A large extent of nonpolar surface area	
High	2000–4660 Å <sup>2</sup>	High	> 4 kcal/mol			

KFC (Knowledge-based FADE and Contacts) is a machine learning approach for the prediction of binding hot spots, or the subset of residues that account for most of a protein interface's binding free energy. The KFC server is a web-based implementation of this machine learning approach. To predict whether an interface residue is hot spot or not, KFC Server characterizes local structural environment of this residue, compares its environment to the environments of experimentally determined hot spots and this procedure is done for each residue in the interface. The KFC model is comprised of two decision tree-based classifiers: K-FADE and K-CON. K-FADE predicts hot spots using the size of the residue and the radial distribution of shape specificity and interface points. Local shape specificity can be derived from atomic density within the interface (Darnell et al., 2008). The local atomic density is high in a crevice surrounded by nearby atoms and the density is characterized by an atomic density exponent (k), which is higher in a crevice than a protrusion (Darnell et al., 2007). K-CON predicts hot spots in terms of a residue's intermolecular atomic contacts, hydrogen bonds, interface points and chemical type. These models do not analyze site-directed mutations or remote regions of the complex, they analyze the intermolecular geometric contacts within a known protein complex structure and predict binding hot spots within the protein–protein interface (Darnell et al., 2007).

No single method can adequately discover the interactome fully. Converging toward an ideal solution will involve unification of different methods that take up the problem from different, innovative perspectives. This will provide a more complete picture of protein network, leading to a better understanding of biological processes (Shen et al., 2006).

## 2.5. Analysis of Hotspots

Protein–protein interfaces are frequently hydrophobic and bury a large extent of nonpolar surface area. Hence, hydrophobicity is a leading force in protein–protein interactions. Hydrophobic interactions are also a major driving force in protein folding. Hydrophobic interactions are also important in the formation of secondary structure elements such as  $\alpha$ -helices and  $\beta$ -sheets. The peptide backbone is relatively hydrophilic because of the C=O and N-H groups of each peptide bond. Hydrophobic interactions in proteins occur between nonpolar regions of their amino acid residues through van der Waals contacts and are driven by the gain in free energy that results from their movement from polar (aqueous) to nonpolar environment (Moreira et al., 2007). Therefore, hot spots (an area of high energy and binding around an amino acid residue), surrounded by hydrophobic pockets, are found in clusters rather than scattered throughout the interface. They usually do not include hydrogen bonding and electrostatic interactions, because hydrogen bonding between the peptide atoms decreases the hydrophobicity of the backbone and electrostatic complementarity of interacting protein surfaces promotes complex formation and defines the lifetime of complexes (Moreira et al., 2007). A necessary condition for high-affinity binding is the exclusion of bulk solvent from the interacting residues. As the hydrophobic effect has an essential role in stabilizing protein–protein complexes and probably provides the driving force for association in most cases, it is important to establish whether a residue’s hydrophobic contribution to binding depends on its local environment or overall position (central or peripheral) in the interface (Orfan and Rost, 2007). Hydrophobic interactions are essential for stabilizing protein–protein complexes, whose interfaces generally consist of a central cluster of hot spot residues surrounded by less important peripheral residues. (Li et al., 2005). Hot spot residues almost always occur in clusters at the centers of interfaces, with relatively few such residues at the edges. Determination of the magnitude of the hydrophobic effect in macromolecular

recognition in general, and a residue's hydrophobic contribution to binding in particular, is critical to improving methods of designing specific small molecule inhibitors of protein-protein interaction targets. For some amino acids there are significant differences between hotspot and non-hotspot interface residues, while for others there are no substantial differences. Usually, the hot spot of one face packs against the hot spot of the other face establishing a region determinant for complex binding, which may provide sites for drug discovery. At present, protein-protein contact areas are considered to be new prospective drug targets because numerous physiological and pathological cell processes depend on them, and thus can be influenced by external compounds. Most drugs produce their effect by interacting with a biological macromolecule by entirely nonbonded forces or, in some cases, by a covalent interaction. Drugs that interact with proteins present a tight-binding, and often have a high degree of complementarity with the target. The drug often forms hydrogen bonds with the receptor. However, some targets have hydrophobic pockets into which the drug can put perhaps a hydrophobic group of an appropriate size. Therefore, the identification of these critical binding residues on proteins permits a rational design of complexes of high affinity and specificity as well as that of small molecules that can mimic the large interface, which is typical of protein-protein complexes. This is fundamental to the development of small molecule competitive inhibitors of protein-protein interactions, which is crucial in structure based drug design (Moreira et al., 2007).

## **3. METHODS**

### **3.1. Prediction of protein interaction partners**

For the prediction of interaction partners of the proteins in the sphingolipid signaling pathway, 32 proteins known to function in the sphingolipid metabolism were taken into consideration. The interaction information for most of them can be found from databases such as BIOGRID, BIND, DIP, MINT. In this study, BIOGRID was mostly used database because it includes both physical and functional interaction details. However, interaction partners of some proteins in the sphingolipid signaling pathway are unknown. In order to get interaction details of this pathway, the unknown interactions were first identified with the PIPE (Protein-Protein Interaction Engine) server. In this study, the proteins without any interacting partner were matched all 6400 proteins in the yeast with PIPE program and interaction scores were obtained. By looking the average scores after filtering, the most probable interaction partners of each protein were chosen. If the scores are greater than 0.06, it was detected as an interaction by the filter and the corresponding protein was chosen probable interaction partner. The results obtained from the PIPE server was included not only the proteins from the sphingolipid signaling pathway, but also all yeast proteins. Therefore, proteins not related to sphingolipid pathway were eliminated using the Gene Ontology (GO) terms.

### **3.2. Modeling Individual Protein Structures**

Structures and interactions of proteins in a pathway can provide useful information about the signaling mechanism in that pathway. Structure information can be obtained from the PDB (Protein Data Bank) (<http://www.rcsb.org/pdb/>). Unfortunately, none of the protein structures in the sphingolipid signaling pathway could be found in PDB. Therefore comparative modeling was used to get information on structure and interacting regions of the proteins in the sphingolipid pathway.

In the present study, the MODELLER program was used for comparative modeling of protein structures in the sphingolipid signaling pathway and their interaction partners. In

order to start structure modeling, amino acid sequences of the proteins to be modeled, alignment of these proteins with template structures and the atomic coordinates of the templates were given as inputs. For each protein, MODELLER built a few models based on the templates used. The numbers of models could be changed depending on the similarity of that protein with the different templates. The most appropriate model was chosen for each protein according to their E values, the size of the modeled range, and the identity between the template and the model.

### **3.3. Modeling Structures of the Interacting Protein Pairs: Docking of Proteins**

In order to identify the interaction regions of the interacting proteins, structures of the protein complexes should be modeled and examined. To this end, docking was done to the interacting proteins by the GRAMM-X server and complex structures of the protein pairs were obtained, i.e. the MODELLER predicted structures of the proteins in the sphingolipid signaling pathway and their interaction partners were docked by GRAMM-X server.

The usage of the server starts with a simple web interface which accepts two protein structures in PDB format. The output file contains 10 models ranked as the most probable complex structure candidates. With any of the standard structure analysis and visualization tools, such as RasMol, VMD or Swiss PDB Viewer, the user can freely process or view the PDB formatted results. In the present study, VMD (Visualization of Molecular Dynamics) software was used for visualization of the 3D structures of the protein complexes.

### **3.4. Biological Relevance by NOXclass**

Protein complexes obtained from GRAMM-X are 3D protein structures built with the calculations of fine-grid projection of softened Lennard-Jones potential function, however biological relevance of these protein complexes were not known. Therefore, in order to have information whether these complexes' structures are biologically significant or not, NOXclass (Zhu et al., 2006) server was used. By using this server protein complexes were classified as "Biological" "Non-biological" and in addition to this biological protein complexes were grouped as "Obligate" and "Non-obligate".

### 3.5. Identifying Interface Residues and Hotspots

The complex structures of interacting proteins obtained from GRAMM-X were then examined to obtain interface properties. The 10 most probable complex structure models predicted by GRAMM-X were analyzed using the KFC server to identify hotspots and the interface residues (<http://www.mitchell-lab.org/>).

The KFC Server starts to work with the submission of the protein-protein complex structure and the interface to analyze. There are three main sections in the server: the submission page, the queue and the job viewer. KFC model can only analyze structures containing proteins and nucleic acids. After submission of the data, the jobs are processed in a queue. When the job is activated, the server begins to calculate the structural features surrounding each residue in the interface. A hotspot is defined as a residue with at least one atom within 4Å of the opposite binding partner. First, a FADE analysis is performed and the radial distribution of shape specificity markers is calculated about each residue. Next, each residue's intermolecular contacts and hydrogen bonds are tabulated. Finally, the K-FADE (Knowledge-based Fast Atomic Density Evaluation) and K-CON (Knowledge-based Biochemical Contacts analysis) models are applied to the calculated features and the putative hot spots are selected (Darnell et al., 2008). The interactive job viewer enables user to quickly highlight predicted hot spots and surrounding structural features within the protein structure. It displays each interface residue and predicted hot spot, and provides summary information about each residue. After obtaining the hotspot residues, VMD (Visualization of Molecular Dynamics) software was used to calculate the distance between the hotspot residues of each model of protein pairs (complex structures). Among the 10 models of complex structures built by GRAMM-X, the one which has the closest hotspots between the interacting protein pairs was assigned the most appropriate model.

The algorithm which was followed in this study can be seen Figure 3.1. The stages until predicting putative interaction partners were only applied to proteins with unknown interaction partners. However, all the other stages after assigning putative partners were applied to each group of proteins (proteins with known interaction partners, proteins with predicted interaction partners and Clusters).

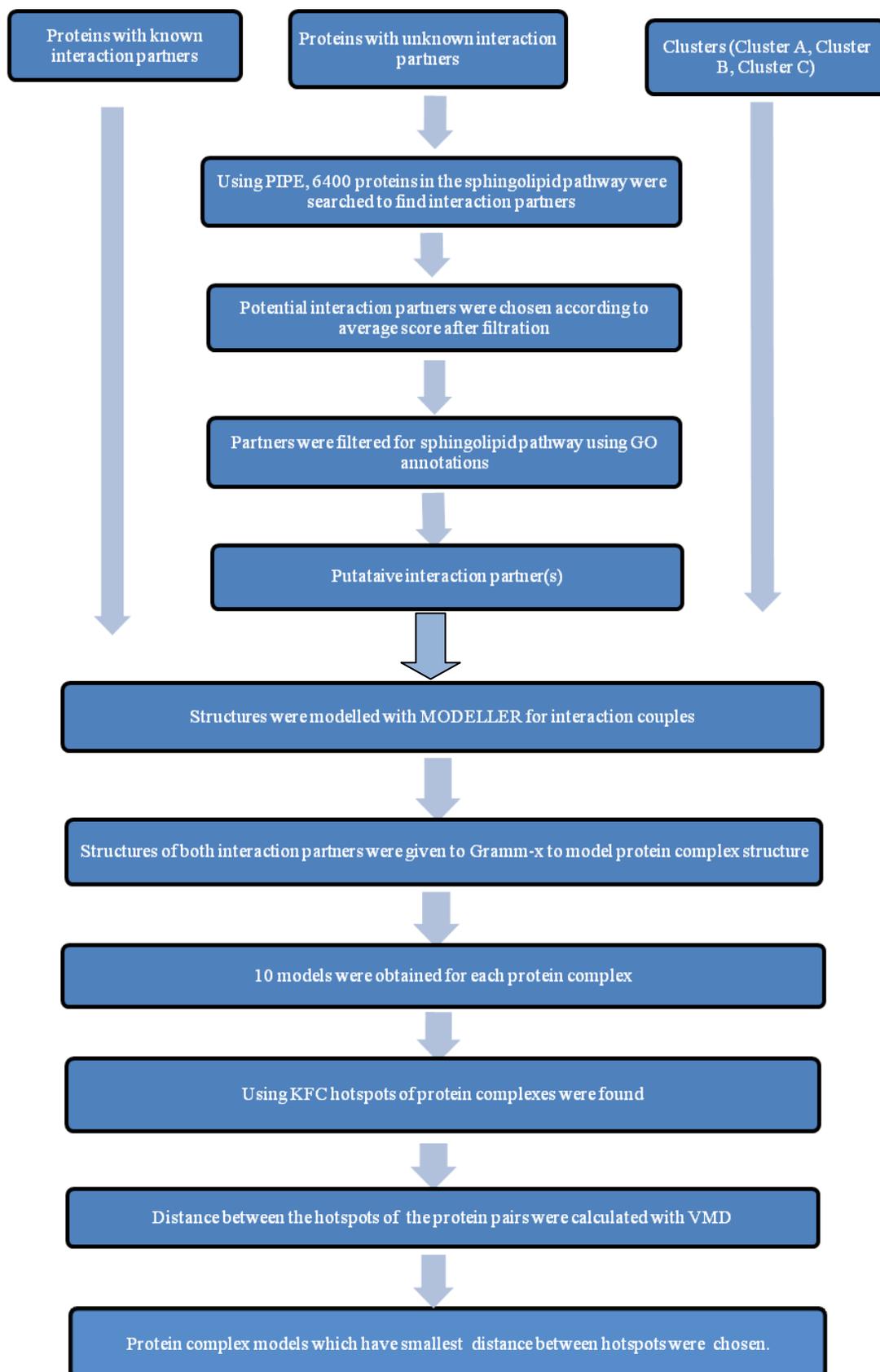


Figure 3.1. Flow chart of the algorithm used to identify protein interaction network

## 4. RESULTS AND DISCUSSION

32 proteins existing in the sphingolipid signaling pathway were investigated in the present study. First, the structures of these proteins and their interaction partners were predicted by MODELLER. Next, the complex structures were built by GRAMM-X. Hotspots were determined by KFC and analyzed for distance values thoroughly. The results were collected under three categories: proteins with known interaction partners, proteins with unknown interaction partners, and proteins in modules / (clusters).

### 4.1. Protein-Protein Interactions in the Sphingolipid (SL) Pathway

#### 4.1.1. Interaction Partners of Sphingolipid Pathway Proteins Identified by BIOGRID

26 proteins known to function in sphingolipid metabolism were searched for their interaction partners in databases. The interaction information of these proteins was obtained from BIOGRID database (<http://www.thebiogrid.org/>). BIOGRID reports both physical and functional interaction details. According to the information obtained from BIOGRID, some proteins had only one interaction partner while some others had many. The proteins in the sphingolipid pathway and their interaction partners were listed in Table 4.1 with their ORF and gene names.

Table 4.1. Interaction partners of proteins in the SL pathway identified using the BIOGRID database (<http://www.thebiogrid.org/>)

ORF name of SL protein	Gene name of SL protein	ORF name of interaction partner	Gene name of interaction partner
YLR242C	ARV1	YOR393W	ERR1
YKL004W	AUR1	YPL057C	SUR1
		YJL196C	ELO1
		YJR105W	ADO1
		YMR010W	NA
		YPL264C	NA
YBR036C	CSG2	YBR036C	CSG2
		YBR161W	CSH1
		YCR034W	FEN1
		YHL003C	LAG1
		YLR372W	SUR4

Table 4.1. Interaction partners of proteins in the SL pathway identified using the BIOGRID database (<http://www.thebiogrid.org/>) (*continued*)

		YDL015C	TSC13
		YJL196C	ELO1
		YOR016C	ERP4
		YMR306W	FKS3
		YML048W	GSF2
		YLR018C	POM34
		YLL048C	YBT1
		YKL065C	YET1
		YHR140W	NA
		YHR110W	ALG1
		YHR026W	PPA1
		YGL051W	MST27
		YEL027W	CUP5
		YBR159W	IFA38
		YBR106W	PHO88
		YAL018C	NA
		YAL007C	ERP2
YBR161W	CSH1	YBR036C	CSG2
YCR034W	FEN1	YDL015C	TSC13
YDR072C	IPT1	YNL307C	MCK1
YKL008C	LAC1	YKL008C	LAC1
		YHL003C	LAG1
		YMR298W	LIP1
		YLR372W	SUR4
		YKL065C	YET1
		YBR159W	IFA38
		YPR048W	TAH18
YHL003C	LAG1	YKL008C	LAC1
		YMR298W	LIP1
YMR296C	LCB1	YDR062W	LCB2
		YBR058C-A	TSC3
YDR062W	LCB2	YBR036C	CSG2
		YMR296C	LCB1
		YNR058C-A	NA
		YGR218W	CRM1
		YLR342W	FKS1
		YKL104C	GFA1
		YBR017C	KAP104
		YER110C	KAP123
		YIL094C	LYS12
YJR077C	MIR1		
YOR171C	LCB4	YOR034C	AKR2
YLR260W	LCB5	YGL137W	SEC27
		YLR213C	CRR1
YMR298W	LIP1	YKL008C	LAC1

Table 4.1. Interaction partners of proteins in the SL pathway identified using the BIOGRID database (<http://www.thebiogrid.org/>) (*continued*)

YMR298W	LIP1	YHL003C	LAG1
		YMR298W	LIP1
YPL006W	NCR1	YGL006W	PMC1
YJL097W	PHS1	YBR159W	IFA38
YMR272C	SCS7	YOR081C	TGL5
YPL057C	SUR1	YBR036C	CSG2
YDR297W	SUR2	YCR034W	FEN1
		YLR372W	SUR4
		YOR016C	ERP4
		YML048W	GSF2
		YKL065C	YET1
		YBR159W	IFA38
		YBR106W	PHO88
		YPL264C	NA
		YGR125W	NA
		YHR007C	ERG11
		YML012W	ERV25
		YKL088W	NA
YLR372W	SUR4	YDL015C	TSC13
YBR265W	TSC10	YJL151C	NA
YER093C		YBR270C	BIT2
		YNL006W	LST8
		YKL203C	TOR2
		YKL003C	MRP17
		YMR068W	AVO2
YBR058C-A	TSC3	YMR296C	LCB1
		YDR062W	LCB2
YPL087W	YDC1	YDL015C	TSC13
		YML012W	ERV25
		YNR065C	NA
YBR183W	YPC1	YCR034W	FEN1
		YLR372W	SUR4
		YOR016C	ERP4
		YML048W	GSF2
		YLR018C	POM34
		YKL065C	YET1
		YHR140W	NA
		YBR159W	IFA38
		YBR106W	PHO88
		YAL007C	ERP2
		YJL117W	PHO86
		YGR060W	ERG25
		YFL041W	FET5
		YDR506C	NA
YKL126W	YPK1	YMR104C	YPK2
		YNL106C	INP52

Table 4.1. Interaction partners of proteins in the SL pathway identified using the BIOGRID database (<http://www.thebiogrid.org/>) (*continued*)

YDL015C	TSC13	YCR034W	FEN1
		YKL008C	LAC1
		YLR372W	SUR4
		YPL087W	YDC1
		YDL015C	TSC13
		YJL196C	ELO1
		YOR016C	ERP4
		YML048W	ERP4
		YKL065C	YET1
		YBR159W	IFA38
		YBR106W	PHO88
		YBR017C	KAP104
		YJL117W	PHO86
		YDR506C	NA
		YPR028W	YOP1
		YBR110W	ALG1
		YBR094W	PBY1
YKL182W	FAS1		

The gene names for some of the proteins were not known; therefore they were defined as “NA” (not available).

#### 4.1.2. Predictions of Interaction Partners of SL Pathway Proteins by PIPE

For six of the proteins functioning in the sphingolipid pathway (YDR294C, YER019W, YGR143W, YGR212W, YJL134W, YKR053C), there have been no interacting partners reported in databases. In this study, the missing interaction partners were identified using the sequence based protein-protein interaction prediction program, PIPE (<http://pipe.cgmlab.org/>). PIPE was used to match each of these six proteins with all 6400 proteins in the yeast proteome using the default values of sensitivity (57%) and specificity (%89). Proteins with average score after filtration (AvgAF) higher than 0.06 were chosen as probable interaction partners. The highest peak on the PIPE graph shows the interaction point with maximum possibility, while a peak with a score higher than 10 (maximum score before filtering) indicates that PIPE is predicting an interaction. However these interacting proteins include not only the sphingolipid signaling pathway proteins but other yeast proteins as well; therefore, the resulting protein list was filtered using GO terms related to the sphingolipid signaling pathway such as “signal”, “endoplasmic”, “golgi”,

“sphingo” and “lipid” and putative interaction partners of sphingolipid signaling pathway proteins were thus selected. GO terms of these proteins were obtained from Saccharomyces Genome Database (<http://www.yeastgenome.org/>). For these six proteins, the keywords used to select the relevant interaction partners in the sphingolipid signaling pathway and the interaction scores (before and after filtering) are listed in Tables 4.2 to 4.7. PIPE reports a contact map between the two interacting proteins and maximum scores for interacting proteins. The maximum scores before filtration are more than 10 and the average scores after filtering (AvgAF) are more than 0.06 for most of the interactions. The interacting residues between the proteins are shown by peaks in the contact map ( Figures 4.1-4.13).

4.1.2.1. Interactions with YDR294C: Table 4.2 lists the putative physical interaction partners of YDR294C protein identified by PIPE. As a result of the elimination process based on the GO annotations, it was found that some of the putative interacting proteins are related to signaling, some of them are related to the endoplasmic reticulum and one of them is related to the Golgi Apparatus. According to the results given in Table 4.2, YHR135C was selected as the most probable interaction partner of the YDR294C. It has the highest peak value (22) in PIPE interaction graph (Figure 4.1). In addition to this, YHR135C may be involved in the sphingolipid pathway due to its location, endoplasmic reticulum. Figure 4.1 shows the interaction map between the YDR294C and YHR135C proteins. The peaks in the graph indicate interacting residues on YDR294C (sequence A) and YHR135C (sequence B) and the maximum interaction score is 22. The highest interaction score was occurred between the residue 113 of YDR294C and the residue 459 of YHR135C.

Table 4.2. Putative physical interaction partners of YDR294C

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Keyword
YHR135C	0.0786	22	Endoplasmic reticulum
YPL203W	0.0662	18	Signal
YPL204W	0.0779	14	Golgi
YBR160W	0.0807	12	Endoplasmic reticulum
YPL031C	0.0768	12	Signal
YGR040W	0.0728	12	Signal
YBL016W	0.0722	12	Signal
YLR113W	0.0625	12	Signal

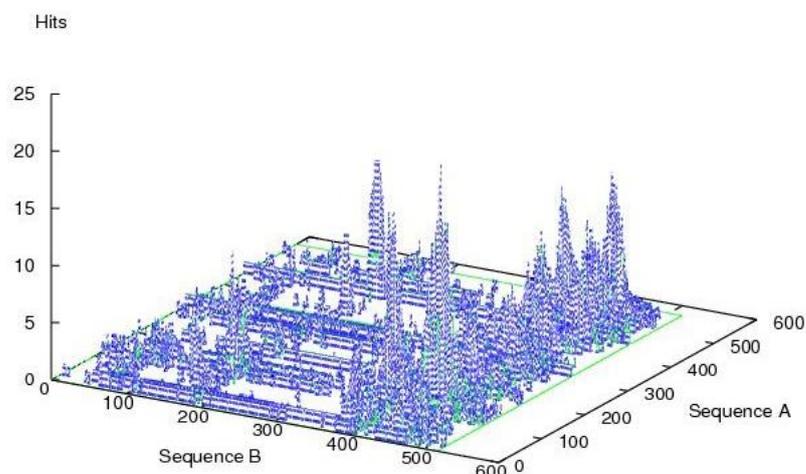


Figure 4.1. PIPE interaction map for proteins YDR294C (Sequence A)-YHR135C (Sequence B)

4.1.2.2. Interactions with YER019W: In Table 4.3, while all PIPE interaction AvgAF scores are higher than 0.06, the scores before filtering vary between 5 and 24. The GO annotations of these putative interaction partners also show a variety in terms of keywords such as lipids, signaling, endoplasmic reticulum and Golgi. The highest maximum PIPE score before filtering, 24, is obtained for the interaction between YER019W and YIL015C, but when selecting the most probable interaction partner of YER019W the values of both AvgAF and maximum PIPE score before filtering and the keywords “lipid” and “endoplasmic” were taken into consideration, and consequently, YHL020C was selected as the most probable interaction partner of YER019W. Figure 4.2 shows the interaction map between YER019W and YHL020C. According to the PIPE results, the maximum peak value in the contact map for these two proteins is 22 between the residue 82 of YER019W and the residues 283 and 284 of YHR020C.

Table 4.3. Putative physical interaction partners of YER019W

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Keyword
YHR030C	0.0894	26	Signal, Endoplasmic
YIL105C	0.0602	24	Signal
YDR099W	0.0953	23	Signal
YHL020C	0.1344	22	Lipid, Endoplasmic
YEL036C	0.0741	22	Golgi
YHR135C	0.1491	21	Endoplasmic

Table 4.3. Putative physical interaction partners of YER019W (*continued*)

YER177W	0.0624	19	Signal
YJR086W	0.1722	15	Signal
YPL203W	0.0628	15	Signal
YPL204W	0.1133	14	Golgi
YKL063C	0.0728	8	Golgi
YNL283C	0.2027	7	Signal
YHL028W	0.1833	7	Endoplasmic
YOL105C	0.1353	7	Signal
YOR008C	0.0821	6	Signal
YJL079C	0.0662	5	Endoplasmic
YNL146W	0.0665	3	Endoplasmic

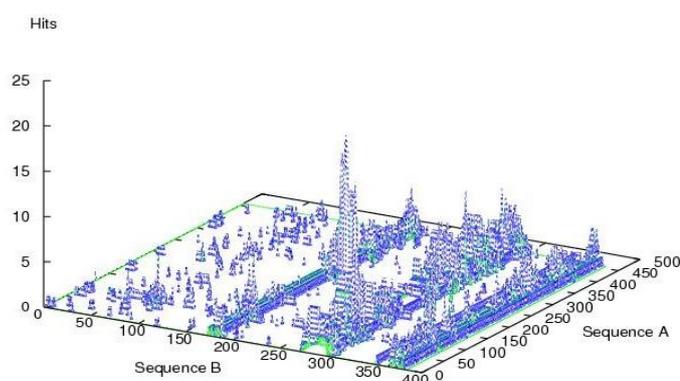


Figure 4.2. PIPE interaction map for proteins YER019W (Sequence A)-YHL020C (Sequence B)

**4.1.2.3. Interactions with YGR143W:** Table 4.4 shows the interaction partners of YGR143W. YKL126W was selected as the most probable interaction partner despite a low PIPE score (maximum PIPE score before filtering: 10) because it has the GO keywords lipid and sphingo. In addition, a maximum PIPE score before filtering of 10 is sufficient to indicate a physical interaction (Pitre et al., 2006). Figure 4.3 shows the interaction map between the YGR143W and YKL126W, the maximum peak value in this graph is 10 between the residue 128 of YGR143W with the residue 458 of YKL126W.

Table 4.4. Putative physical interaction partners of YGR143W

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Key word
YIL105C	0.0769	45	Signal
YHR030C	0.1968	43	Signal, Endoplasmic
YDR099W	0.1009	41	Signal
YHR135C	0.1552	37	Endoplasmic
YNL298W	0.0964	37	Signal

Table 4.4. Putative physical interaction partners of YGR143W (*continued*)

YER177W	0.0781	36	Signal
YHL020C	0.1145	37	Endoplasmic
YJR086W	0.1640	34	Signal
YPL203W	0.2310	32	Signal
YPL204W	0.1272	27	Golgi
YBR288C	0.0692	27	Golgi
YDR477W	0.2127	20	Signal
YDR477W	0.2127	20	Signal
YHL007C	0.0800	14	Signal
YHR082C	0.0616	14	Signal
YJL164C	0.1653	12	Signal
YFR014C	0.1000	12	Signal
YLR362W	0.0646	12	Signal
YKL166C	0.1574	11	Signal
YLR113W	0.1363	11	Signal
YOL016C	0.0899	11	Signal
YBL016W	0.1823	10	Signal
YGR040W	0.1510	10	Signal
YLR248W	0.0923	10	Signal
YKL126W	0.0902	10	Lipid, Spingo
YPL140C	0.0874	10	Signal
YOR231W	0.0821	10	Signal
YKL116C	0.0747	10	Signal
YBL105C	0.0654	10	Signal
YBR160W	0.1494	9	Endoplasmic
YDL214C	0.0676	9	Signal
YHL019C	0.1190	5	Golgi
YBR254C	1.0000	4.5	Golgi
YPL259C	0.7290	4	Golgi
YFR042W	0.7280	3	Endoplasmic
YDR373W	0.1704	3	Signal, Golgi
YDL047W	0.0701	3	Signal
YLL017W	0.0678	3	Signal
YDR246W	0.0617	3	Golgi

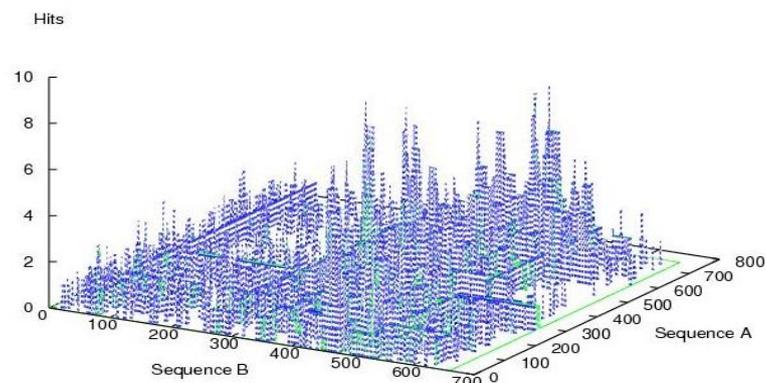


Figure 4.3. PIPE interaction map for proteins YGR143W (Sequence A)-YKL126W (Sequence B)

4.1.2.4. Interactions with YGR212W: YHR135C and YPL204W were chosen as the potential interaction partners of YGR212W due to their average PIPE scores after and before filtering. In addition, their GO terms are related to those of sphingolipid signaling pathway proteins. Figure 4.4 shows the interaction map between YGR212W and YPL204W and Figure 4.5 shows the interaction map between YGR212W and YHR135C. In Figure 4.4, the maximum peak score is 12 between the residue 346 of YGR212W and the residues 116, 118, 119, 120 of YPL204W, and in Figure 4.5, it is 16 between the residue 346 of YGR212W and the residue 458 of YHR135C.

Table 4.5. Putative physical interaction partners of YGR212W

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Keyword
YHR135C	0.0935	16	Endoplasmic
YDR099W	0.0660	15	Signal
YLR423C	0.0610	15	Signal
YHR030C	0.0846	14	Signal, Endoplasmic
YBR160W	0.0744	14	Endoplasmic
YPL203W	0.0703	13	Signal
YPL031C	0.0743	12	Signal
YPL204W	0.0688	12	Golgi
YPL204W	0.0688	12	Golgi
YJR086W	0.0656	10	Signal

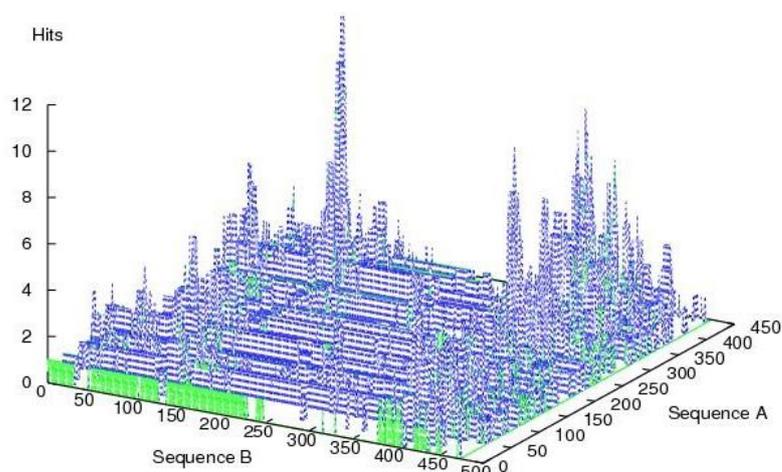


Figure 4.4. PIPE interaction map for proteins YGR212W (Sequence A)-YPL204W (Sequence B)

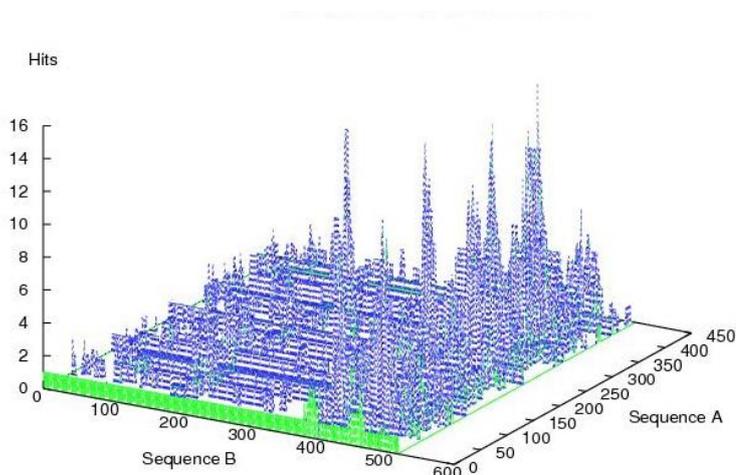


Figure 4.5. PIPE interaction map for proteins YGR212W (Sequence A)-(YHR135C Sequence B)

4.1.2.5. Interactions with YJL134W: Table 4.6 shows putative interaction partners of YJL134W. YAR033W and YGL051W were chosen as the putative interaction partners of YJL134W. Even though PIPE maximum interaction scores before filtering are lower than 10 indicating no interaction, the average scores after filtering are higher than 0.06, suggesting a possible interaction. In addition both proteins are related to endoplasmic reticulum and Golgi apparatus. The maximum interaction scores in the PIPE interaction maps between YJL134W and YAR033W (Figure 4.6) and YJL134W and YGL051W (Figure 4.7) are 3 and there are not any distinctive interaction points. For this reason, it is difficult to indicate distinct interaction regions for these couples of proteins. Interacting residues in pairs are (21-9), (21-10), (49-37), (50-37), (142-138), (142-139), (142-140), (142-141). First residues' numbers are of YJL134W and second ones are of YAR033W. For interactions between YJL134W and YGL051W, the interacting residue pairs are found as follows, respectively; (21-9), (21-10), (49-37), (50-37), (142-37), (142-138), (142-139), (142-140), (142-142).

Table 4.6. Putative physical interaction partners of YJL134W

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Keyword
YGL053W	1.0000	3	Endoplasmic
YGL051W	0.6746	3	Endoplasmic, Golgi
YAR033W	0.6745	3	Endoplasmic, Golgi
YAR031W	0.6059	3	Endoplasmic
YNR075W	0.08995	2	Endoplasmic

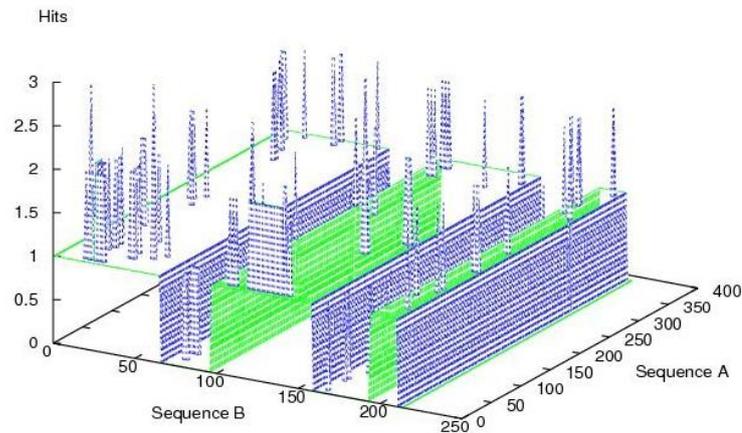


Figure 4.6. PIPE interaction map for proteins YJL134W (Sequence A)-YAR033W (Sequence B)

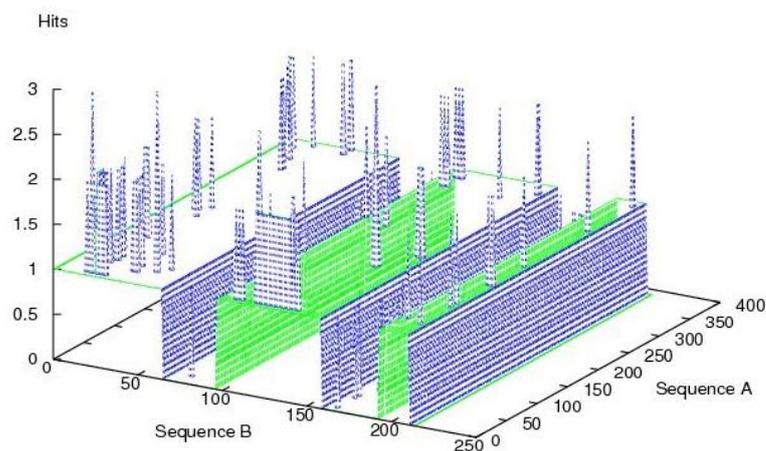


Figure. 4.7. PIPE interaction map for proteins YJL134W (Sequence A)-YGL051W (Sequence B)

4.1.2.6. Interactions with YKR053C: YAR033W and YGL051W were chosen as the potential interaction partners of YKR053C because their AvgAF scores were higher than 0.06. In addition, their GO terms are related to endoplasmic reticulum which was a significant reason for assuming their interaction in the sphingolipid signaling pathway. However, maximum PIPE scores of these interactions were not good enough to define an interaction between any these proteins and YKR053C. The interaction map between YKR053C and YAR033W (Figure 4.8) and between YKR053C and YGL051W (Figure 4.9) show no distinctive interaction. The interaction regions with maximum score between YKR053C and YAR033W protein pair are as follows: (142-140), (108-316), (108-317),

(108-318), (108-319); and between YKR053C with YGL051W protein pair are (142-140), (108-316), (108-317), (108-318), (108-319).

Table 4.7. Putative physical interaction partners of YKR053C

ORF NAME	Average score after filtering	Maximum PIPE score before filtering	Keyword
YAR033W	0.4732	3	Endoplasmic
YGL051W	0.4732	3	Endoplasmic
YGL053W	1.0000	3	Endoplasmic

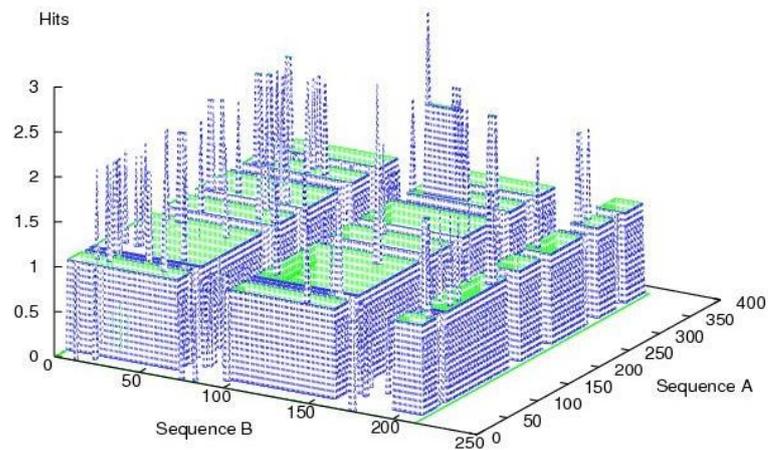


Figure. 4.8. PIPE interaction map for proteins YKR053C (Sequence A)-YAR033W (Sequence B)

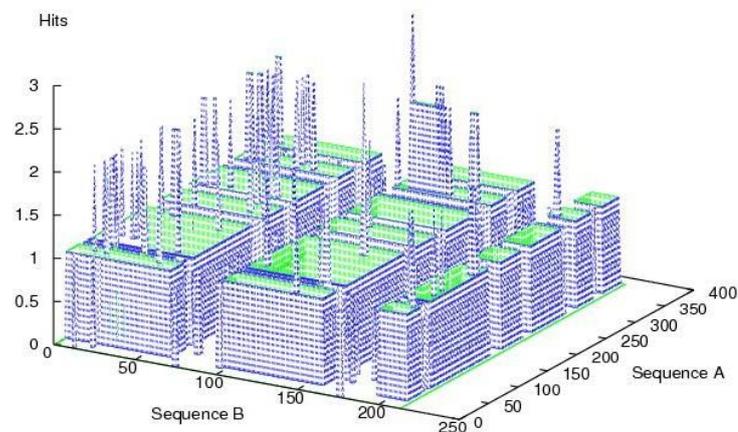


Figure. 4.9. PIPE interaction map for proteins YKR053C (Sequence A)-YGL051W (Sequence B)

Table 4.8 is a summary of the putative interaction partners of the six proteins with previously unknown interaction partners in sphingolipid signaling pathway and their GO annotations. These putative interaction partners were selected based on their PIPE scores as well as their GO annotations. From this table, the relationships of the GO annotations of these putative interaction partners with sphingolipid signaling pathway can be seen.

Table 4.8. GO annotations of Putative interaction partners. IDA: Inferred from Direct Assay, IGI: Inferred from Genetic Interaction, IMP: Inferred from Mutant Phenotype, ISS: Inferred from Sequence or Structural Similarity, IPI: Inferred from Physical Interaction

Protein	Putative partner	Function	Process	Component
YDR294C	YHR135C	Manually curated protein serine/threonine kinase activity (IDA)	Manually curated cell morphogenesis (IGI) cytokinesis (IGI) endocytosis (IGI) protein amino acid phosphorylation (IMP, ISS) response to glucose stimulus (IMP)	Manually curated plasma membrane (IDA) High-throughput endoplasmic reticulum (IDA) mitochondrion (IDA)
YER019W	YHL020C	Manually curated transcription corepressor activity (IPI)	Manually curated endoplasmic reticulum unfolded protein response (IMP) negative regulation of transcription from RNA polymerase II promoter (IMP) phospholipid biosynthetic process (IMP) positive regulation of transcription from RNA polymerase II promoter (IMP)	Manually curated endoplasmic reticulum (IDA) nuclear envelope (IDA) nuclear membrane (IDA) nucleoplasm (IDA) nucleus (IDA)
YGR143W	YKL126W	Manually curated protein serine/threonine kinase activity (IMP, ISS)	Manually curated endocytosis (IMP) protein amino acid phosphorylation (IMP, ISS) sphingolipid metabolic process (IMP)	Manually curated cellular bud neck (IDA) cytosol (IDA) plasma membrane (IDA)
YGR212W	YPL204W	Manually curated protein serine/threonine kinase activity (IDA, IMP)	Manually curated attachment of spindle microtubules to kinetochore during meiosis I (IMP) DNA repair (IMP) ER to Golgi vesicle-mediated transport (IMP) mitosis (IMP) protein amino acid phosphorylation (IMP) ribosomal small subunit biogenesis (IMP)	Manually curated cellular bud neck (IDA) cellular bud tip (IDA) chromosome, centromeric region (IDA) monopolin complex (IDA, IMP, IPI) nucleus (IDA) plasma membrane (IDA)

Table 4.8. GO annotations of Putative interaction partners. IDA: Inferred from Direct Assay, IGI: Inferred from Genetic Interaction, IMP: Inferred from Mutant Phenotype, ISS: Inferred from Sequence or Structural Similarity, IPI: Inferred from Physical Interaction  
(continued)

YGR212W	YHR135C	Manually curated protein serine/threonine kinase activity (IDA)	Manually curated cell morphogenesis (IGI) cytokinesis (IGI) endocytosis (IGI) protein amino acid phosphorylation (IMP, ISS) response to glucose stimulus (IMP)	Manually curated plasma membrane (IDA) High-throughput endoplasmic reticulum (IDA) mitochondrion (IDA)
YJL134W	YAR033W	Manually curated protein binding (IDA)	Manually curated vesicle organization (IGI, IPI)	Manually curated endoplasmic reticulum (IMP, IPI) Golgi apparatus (IPI) integral to membrane (IDA) plasma membrane (IMP)
	YGL051W	Manually curated protein binding (IDA)	Manually curated vesicle organization (IGI, IPI)	Manually curated endoplasmic reticulum (IDA, IMP) Golgi apparatus (IDA) integral to membrane (IDA) plasma membrane (IMP)
YKR053C	YAR033W	Manually curated protein binding (IDA)	Manually curated vesicle organization (IGI, IPI)	Manually curated endoplasmic reticulum (IMP, IPI) Golgi apparatus (IPI) integral to membrane (IDA) plasma membrane (IMP)
	YGL051W	Manually curated protein binding (IDA)	Manually curated vesicle organization (IGI, IPI)	Manually curated endoplasmic reticulum (IDA, IMP) Golgi apparatus (IDA) integral to membrane (IDA) plasma membrane (IMP)

### 4.1.3. Clusters / Modules

Some closely related groups (clusters) in the SL pathway (Tables 4.9, 4.10, 4.11) were previously identified by Betül Kavun Özbayraktar within the framework of her PhD thesis by selecting highly interconnected regions in the sphingolipid pathway, i.e. determining the clusters which consist of only core proteins and their first neighbours among all interaction pairs.

Table 4.9. Proteins in Cluster A and their interaction partners

Proteins in Cluster A	Interaction partners
YOR171C	YOR034C
YOR034C	YOR171C

Table 4.10. Proteins in Cluster B and their interaction partners

Proteins in Cluster B	Interaction partners
YMR298W	YMR298W
	YGR060W
YHL003C	YMR298W
YGR060W	YGR060W
	YHL003C

Table 4.11. Proteins in Cluster C and their interaction partners

Proteins in Cluster C	Interaction partners
YLL006W	YLL006W
	YCR034C
YNL307C	YNL307C
	YAL007C
YDR297W	YOR016C
	YCR034W
YMR058W	YOR016C
	YCR034W
YHR026W	YAL007C
	YHR026W
	YOR016C
YEL027W	YEL027W
	YHR026W
	YOR016C
YHR110W	YAL007C
	YOR016C
	YHR110W
	YCR034W

## 4.2. Prediction of Protein Structure

In the previous section, the interaction partners of 32 proteins involved in the sphingolipid (SL) pathway were identified using the BIOGRID database (<http://www.thebiogrid.org/>) (Table. 4.1) or they were predicted using PIPE (Table 4.8). As a next step, the structures of these proteins were built using MODELLER (<http://www.salilab.org/modeller/>). MODELLER uses homology modeling to build a protein structure based on one or more templates with high sequence identity to the protein of unknown structure. If there are more than one templates, more than one model structure may be built based on them. The best model was selected based on “coverage of sequence alignment”, “sequence identity of alignment”, and “E values” of each model. The most important point considered was the E value when choosing the appropriate model for each protein. If the E value is smaller than or equal to 0.01 that model was accepted. In general, a sequence identity value above ~25% indicates a potential template, unless the alignment is too short (i.e., <100 residues). Therefore, the models with low E value and high identity and maximum coverage were chosen for each protein. However, for some proteins MODELLER could not build models that satisfy all of the above criteria due to the absence of an experimentally verified template structure appropriate for that protein to modeling. In such cases, the best model among the obtained structure models was selected. The most appropriate models chosen for each protein, the templates used for modeling, the region of target sequences, the length of the target sequences, the coverage of the modeled structure for target sequences, the sequence identity of the alignment, and the E values are given in Tables 4.12-4.19. The results were analyzed under three categories; proteins with known partners, proteins with predicted partners, and clusters.

### 4.2.1. Proteins with known partners

Assigning a successful template for modeling the structure is not always possible. Identifying a template which covers several criteria at the same time is very difficult. The coverage of the modelled target structure is sometimes as low as 13% indicating that only 13% of the protein was successfully modelled. Besides, some E values are not low enough. However, many models have sequence identities higher than 25%. Consequently, some

proteins with known interaction partners could not be modelled properly because there was not any suitable template.

Table 4.12. The templates used in building modeling SL proteins with known partners (Set A). The coverage, sequence, identity and E-value of the alignment are also listed.

Protein ORF	Template PDB code	Region of modelled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YLR242C	2ga9D	69-223	321	47.975	16	0
YKL004W	2ckpA	238-314	341	22.287	29	0.064
YBR036C	-	No model	-	-	-	-
YBR161W	-	No model	-	-	-	-
YCR034W	-	No model	-	-	-	-
YDR072C	2bjiA	357-437	527	15.180	34	0.15
YKL008C	2j67A	90-176	418	20.574	24	0.13
YHL003C	1h5uA	190-283	411	22.63	25	0.27
YMR296C	2jg2A	92-421	558	58.960	27	0
YDR062W	1fc4A	158-524	561	65.240	33	0
YOR171C	2qv7A	219-411	624	30.769	29	0.00093
YLR260W	2qv7A	273-398	687	18.195	28	0.00033
YMR298W	1ivyA	43-120	150	51.333	26	0.52
YPL006W	2dhhA	1002-1152	1170	12.820	26	1.9e-07
YJL097W	2g84A	37-111	217	34.101	24	0.2
YMR272C	2ibjA	6-95	384	21.177	40	1.9e-11
YPL057C	-	No model	-	-	-	-
YDR297W	1kcmA	250-309	349	16.905	33	0.081
YLR372W	2uvpA	90-182	345	26.666	28	0
YBR265W	1geeA	5-43	320	11.875	46	0.14
YER093C	1z3hB	7-829	1430	57.482	13	0
YBR058C-A	2gv8A	12-74	80	77.500	33	0.59
YPL087W	3dtuA	159-252	317	29.337	26	0.17
YBR183W	1qleC	199-276	316	24.367	31	0.35
YKL126W	2r5tA	332-624	680	42.941	54	0
YDL015C	1wjnA	1-82	310	26.129	15	0

Table 4.13. The templates used in building models of interaction partners of set A. The coverage, sequence, identity and E-value of the sequence alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YOR393W	2al1A	2-437	437	99.54	68	0
YJL196C	1c17M	197-254	310	18.34	24	0.59

Table 4.13. The templates used in building models of interaction partners of set A. The coverage, sequence, identity and E-value of the sequence alignment are also listed.  
(continued)

YJR105W	1bx4A	8-334	340	95.88	41	0
YMR010W	2bs2C	166-345	405	44.19	24	0.048
YPL264C	2dyrC	13-137	353	35.12	17	0
YHL003C	3emlA	69-310	411	58.63	13	0
YKL065C	2eflA	67-192	206	60.67	10	0
YBR159W	2zatA	63-313	347	72.04	29	2.7e-12
YPR048W	1amoA	1-623	623	99.83	23	0
YDR062W	2jg2A	103-524	561	75.04	27	0
YMR296C	2jg2A	92-421	558	58.96	27	0
YGR218W	2q5dB	18-940	1084	85.05	15	0
YLR342W	1ajkA	8-165	1876	8.36	18	0
YKL104C	2pocA	350-705	717	49.51	81	0
YBR017C	1qbkB	4-917	918	99.45	34	0
YER110C	1ibrB	37-332	1113	26.50	29	6.8e-05
YIL094C	1x0lA	23-371	371	93.80	49	0
YOR034C	2dvwA	14-214	749	26.70	28	4.9e-11
YLR213C	1umzA	149-355	422	48.81	26	0
YNL307C	1j1bA	36-370	375	89.06	43	0
YHL003C	3emlA	69-310	411	58.63	13	0
YBR159W	2zatA	63-313	347	72.04	29	2.7e-12
YOR081C	1ayzA	552-690	749	18.42	19	0
YLR372W	2uvpA	90-182	345	26.66	28	0
YKL065C	2eflA	67-192	206	60.67	10	0
YBR106W	2nvpA	23-110	188	46.27	31	0.18
YGR125W	3cucA	89-294	1036	19.78	17	0
YHR007C	1x8vA	55-525	530	88.67	27	0

Table 4.13. The templates used in building models of interaction partners of set A. The coverage, sequence, identity and E-value of the sequence alignment are also listed.  
(continued)

YML012W	1g3kA	47-138	211	43.12	31	0.51
YKL088W	1e20A	307-497	571	33.27	46	0
YGL137W	2ovrB	66-294	889	25.64	34	0
YGL006W	2o9jA	65-971	1173	77.23	34	0
YMR104C	2r5tA	329-620	677	42.98	54	0
YPL087W	3dtuA	159-252	317	29.33	26	0.17
YBR110W	2bisA	46-443	449	88.41	15	0
YBR094W	2phjA	101-215	753	15.13	41	3.5e-09
YBR270C	1w99A	6-284	545	36.91	15	0
YNL006W	1erjA	26-300	303	90.42	27	0
YKL203C	2qizA	455-1413	2474	38.72	14	0
YMR068W	1n11A	9-213	426	47.88	30	2.7e-12
YNR065C	3f6kA	301-918	1116	55.86	23	0

#### 4.2.2. Proteins with predicted partners

Table 4.14 lists the templates, used in the models of the six proteins with predicted partners. As shown in Table 4.14, the coverage values of the sequence alignment are higher than 40% and E values are smaller than 0.01 for all model structures. Even though, the sequence identities of the alignments are not higher than 25% for any of the proteins the low E values were used to select the template.

Table 4.14. The templates used in building modeling SL proteins with predicted partners (Set B). The coverage, sequence, identity and E-value of the alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YDR294C	3f9Ta	138-532	589	66.893	23	0

Table 4.14. The templates used in building modeling SL proteins with predicted partners (Set B). The coverage, sequence, identity and E-value of the alignment are also listed.  
(continued)

YERO19W	1i9Za	33-379	477	72.536	17	0
YGR143W	2hykA	379-712	771	43.190	23	0
YGR212W	1q9Jb	3-451	468	95.726	15	0
YJL134W	3c08H	24-193	409	41.320	11	0
YKR053C	2ipbA	67-241	404	43.069	16	0

Table 4.15 shows the region and the coverage values of models of the predicted partners of proteins. The coverage and sequence identity of the alignments are high for all the templates, but the E values of YAR033W and YGL051W are very high as 0.81 and 0.82 respectively. Although these two values are not preferable values these models were chosen due to the absence of other appropriate templates.

Table 4.15. The templates used in building models of interaction partners of set B. The coverage, sequence, identity and E-value of the sequence alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YHR135C	1csnA	64-355	538	54.08	67	0
YHL020C	1bg1A	280-404	404	30.69	16	0
YKL126W	2r5tA	332-624	680	42.94	54	0
YPL204W	1ckiA	1-295	494	59.51	66	0
YAR033W	2dhkA	143-219	234	32.47	30	0.81
YGL051W	2dhkA	143-219	234	32.47	33	0.82

#### 4.2.3. Clusters / Modules

The MODELLER results for clusters are listed in Tables 4.16- 4.19. Tables show the some important values of the model structures which built with homology modeling. When the specifications of the structures in three tables examined it is appeared that the sequence identities of the alignments are generally higher than 25% except three proteins, YHL003C, YGR060W, YHR110W. However, the E values of some protein models are generally higher than 0.01 due to the absence of good template structures which are used in

modeling the structures of the proteins of our clusters. Interaction partners of proteins in Cluster A and Cluster B exist in their own clusters, so they are shown in the same table. However, the interaction partners of proteins in Cluster C are different from the proteins in Cluster C, so they are shown in a separate table. Table 4.16 shows the specifications of the modelled structures in Cluster A. In Cluster A, there are two proteins which interact with each other. The results in Table 4.16 show that, the structures which were obtained with homology modeling using MODELLER are quite good to accept them as convenient for these two proteins. The E values of both structures are smaller than 0.01 and the sequence identities are higher than 25%. Moreover the regions of modeled target sequences are longer than 100 residues. Based on these values obtained these two structures are suitable for further use in modeling the complex structures.

Table 4.16. The templates used in building modeling SL proteins in Cluster A. The coverage, sequence, identity and E-value of the alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YOR171C	2qv7A	219-411	624	30.929	29	0.00093
YOR034C	2dvwA	58-265	749	27.777	28	4.9e-11

Table 4.17 gives important information on the model structures of the proteins in Cluster B. The proteins in Cluster B interact with each other; therefore, the proteins in Table 4.17 are both members of Cluster B and their interaction partners. Based on values in Table 4.17, the E values of YHL003C and YGR060W are very good (E values are 0); however, the sequence identity of YMR298W is only higher than 25%. These values show that, none of the proteins in Cluster B meet three criteria at the same time. Nevertheless, the coverage of all three proteins are higher than 50%.

Table 4.17. The templates used in building modeling SL proteins in Cluster B. The coverage, sequence, identity and E-value of the alignment are also listed. Template-region-coverage values of Cluster B

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YMR298W	livyA	43-120	150	52.000	26	0.52

Table 4.17. The templates used in building modeling SL proteins in Cluster B. The coverage, sequence, identity and E-value of the alignment are also listed. Template-region-coverage values of Cluster B (*continued*)

YHL003C	3emlA	69-310	411	62.206	13	0
YGR060W	1fftA	1-282	309	91.262	8	0

Table 4.18 shows the proteins in Cluster C while Table 4.19 shows the interaction partners of the proteins in Cluster C. Both tables give good sequence identities, for example in Table 4.18, the sequence identities of 8 proteins out of 9 proteins are higher than 25%. In Table 4.19 the sequence identities of 7 proteins out of 9 proteins are higher than 25%. Besides, one protein (YAL007C) cannot be modeled due to the absence of any suitable template. E values are not good for proteins in both tables; for instances, in Table 4.18, the proteins with E values are smaller than 0.01 are only YEL027W, YHR110W, YMR058W, YNL307W while in Table 4.19 they are YNL307W, YEL027W, YHR110W.

Table 4.18. The templates used in building modeling SL proteins in Cluster C. The coverage, sequence, identity and E-value of the alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YCR034W	3gjcA	183-250	347	19.596	35	0.67000
YDR297W	1kcmA	250-309	349	17.191	33	0.08500
YEL027W	2bl2A	4-154	160	94.375	25	0
YHR026W	1h2sB	150-206	213	26.760	44	0.81000
YHR110W	2qihA	123-180	212	27.358	12	0
YLL006W	1w0bA	168-249	426	19.248	28	0.83999
YMR058W	1zpuA	22-550	636	83.176	100	0
YNL307C	1j1bA	57-379	375	86.133	43	0
YOR016C	1ve3A	31-107	207	37.198	31	0.20000

Table 4.19. The templates used in building modeling interaction partners of Cluster C. The coverage, sequence, identity and E-value of the alignment are also listed.

Protein ORF	Template PDB code	Region of modeled target sequence	Length of target sequence	Coverage (%)	Sequence identity of the alignment	E value
YLL006W	1w0bA	168-249	426	19.248	28	0.83999
YNL307C	1j1bA	57-379	375	86.133	43	0
YAL007C	No Template					
YOR016C	1ve3A	76-155	207	38.16	31	0.59999

Table 4.19. The templates used in building modeling interaction partners of Cluster C. The coverage, sequence, identity and E-value of the alignment are also listed. (*continued*)

YCR034W	3gjcA	183-250	347	19.30	35	0.67000
YEL027W	2bl2A	4-154	160	94.375	25	0
YCR034W	3gjcA	183-250	347	19.596	35	0.67000
YHR026W	1h2sB	150-206	213	26.760	44	0.81000
YHR110W	2qihA	123-180	212	27.358	12	0

### 4.3. Prediction of Complex Structures

The selected model structures for each protein and their interaction partners in the sphingolipid signaling pathway were used in docking calculations with the GRAMM-X server (<http://vakser.bioinformatics.ku.edu/resources/gramm/grammx>) in order to predict complex structures. GRAMM-X built ten complex structures for each protein pair. These 10 models mean 10 different binding styles between the protein pairs. In order to select the ideal complex model for each protein complex, the determination of the most convenient binding style is necessary. To this end, hotspots were identified and the distance values between them were calculated. After these calculations, the complex model which has hotspots with smallest distance values was selected as the most probable protein complex for interaction. The data related to the selected protein complex models and hotspots of those protein complexes are listed in the Appendix A in Table A.1, Table A.2, Table A.3, Table A.4, Table A.5 for all groups.

#### 4.3.1 Biological relevance of the predicted complex structures

Structural models of the complexes resulting from protein-protein interactions are necessary to understand those processes at the molecular level. To this purpose, in this study, first individual structures and then complex structures were modeled for proteins. In an effort to determine biological relevance of the predicted complex structure models, NOXclass (Zhu et al., 2006) was used (Table 4.46). NOXclass classified the protein pairs as “biological” or “non-biological”. Here, “non-biological” means biologically irrelevant interactions resulting from crystal packing contacts; while “biological” means biologically relevant interactions. Crystal packing contacts are non-specific and have no biological function associated. Biological interactions are then divided into two groups which are “Obligate” and “Non-obligate”. In obligate interactions, interface residues were reported to

be significantly more conserved than those in transient interactions. In addition, the coevolution rate was observed to be lower for obligate interaction partners than for transient interaction partners. In general, obligate and non-obligate proteins have been shown to have distinct interaction preferences. Interface residues in obligate and non-obligate interactions are more highly conserved than those in crystal packing contacts. Obligate interfaces have more contacts than non-obligate interfaces and these contacts are mainly nonpolar. The approaches are based on the assumption that the key residues are involved in biologically relevant interactions.

For proteins predicted partners, the protein complex structures which were selected as ideal ones among the GRAMM-X structure models were examined using NOX-class to understand their biological relevance. As can be seen from the Table 4.20, in consequence of these calculations, protein complexes identified as biological are YGR143W-YKL126W, YGR212W-YHR135C, YJL134W-YAR033W, YJL134W-YGL051W, especially the results of YJL134W-YAR033W and YJL134W-YGL051W are very good such that they are 100% biological. Moreover, the obligate percentages are very high for these protein couples, they are 96.47% and 91.81%, respectively. However, some protein complexes identified as non biological are YDR294C-YHR135C, YER019W-YHL020C, YGR212W-YPL204W, YKR053C-YAR033W, YKR053C-YGL051W. These undesired results may due to the absence of known structures of these proteins in literature so that the model structures obtained by MODELLER and GRAMM-X are not the proper ones. These model structures do not include whole sequences of the proteins, but only some part of them, therefore accepting these models as real structures is not possible. Therefore, obtaining a result such as “non-biological” for these protein complexes is possible.

Table 4.20. Biological relevance of proteins with predicted partners

Protein Complex	Biological (%)	Non-biological (%)	Obligate (%)	Non-obligate (%)
YDR294C-YHR135C	0.00	100.00	0.00	0.00
YER019W-YHL020C	0.00	100.00	0.00	0.00
YGR143W-YKL126W	17.49	82.51	96.47	3.53

Table 4.20. Biological relevance of proteins with predicted partners (*continued*)

YGR212W- YPL204W	0.00	100.00	0.00	0.00
YGR212W- YHR135C	14.43	85.57	91.81	8.19
YJL134W- YAR033W	100.00	0.00	3.82	96.18
YJL134W- YGL051W	100.00	0.00	4.12	95.88
YKR053C- YAR033W	0.00	100.00	0.00	0.00
YKR053C- YGL051W	0.00	100.00	0.00	0.00

#### 4.4. Prediction of Hotspots

Hotspots are small subset of residues that account for a significant part of a protein interface's free energy of binding (Darnell et al., 2008). In this study, hotspots of the above mentioned complex structures were identified by KFC. The server first calculates, the structural features surrounding each residue in the interface and an interface residue was defined as a residue with at least one atom within 4Å of the opposite binding partner. Then, through this server, residues were classified as hot spots if their mutation to alanine resulted in a change of binding energy ( $\Delta\Delta G$ ) greater than 2 kcal/mol. Then, predictive models were created from many different combinations of structurally-derived chemical and physical features that describe the interface residues, and those that best described the hot spot environment were selected as features for the K-FADE (based on shape specificity features calculated by the Fast Atomic Density Evaluator) and K-CON (based on biochemical contact features) models. Finally, the K-FADE and K-CON models are applied to the calculated features and the putative hot spots were selected (Darnell et al., 2008). KFC provides summary information about each residue and chemical properties of hotspots, such as hydrophobicity, polarity, acidity and so on. Hotspots were identified for 10 models of each protein complex and in order to select the ideal model for each protein complex, the distances between all binary combinations of hotspots in protein A and protein B were calculated by VMD (Visual Molecular Dynamics). After this calculation, some

hotspot combinations, distances of which higher than 10 Å were eliminated. The distance threshold between hotspots is generally preferred to be less than 10 Å to obtain accurate results; however in this study it was assigned 10 Å due to low number of the hotspot combinations with a distance value smaller than 10 Å. The residue numbers of hotspots, values of hotspot distances and interaction energies of selected models were listed in Tables A.1, A.2, A.3, A.4, A.5 in Appendix A. The model for each protein complex which had the shortest distances was noted. Finally, according to the distance, the most appropriate complex model which had the shortest hotspot distance was chosen for each protein pair. As a result of these calculations the best models and their hotspots were identified for each protein complex (They can be seen in Appendix A).

#### 4.4.1. Proteins with known interaction partners

Hotspots are identified by KFC for all complex structures of proteins with known interaction partners. Below, two of them, which have better model structures, are shown as typical examples. Figure 4.10 shows the structure of YDL015C-YBR094W protein complex and its hotspots. While PHE89, CYS90, HIS99, LEU101 are hotspots of YDL015C (blue), PRO112, SER148, SER149, GLY150 are hotspots of YBR094W (green). Here, pink spheres are representing the hotspot residues.

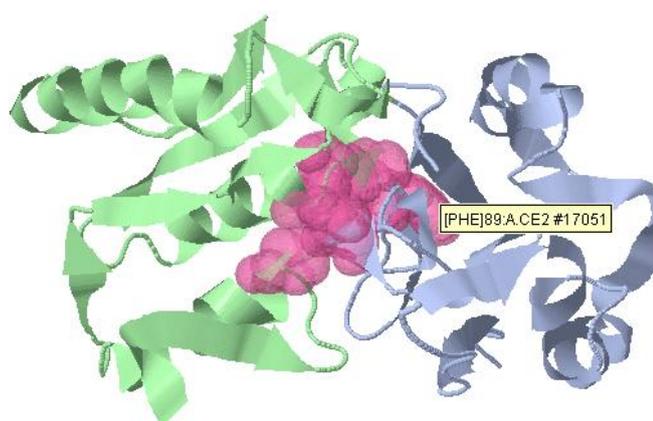


Figure 4.10. Complex structure of YDL015C-YBR094W (Model 10)

Figure 4.11 shows the structure of YER093C-YKL203C protein complex and its hotspots. While SER41, ASN45, LEU68 are hotspots of YER093C (blue), VAL2316, PRO2317, THR2321, MET2323, GLN2238 are hotspots of YKL203C (green).

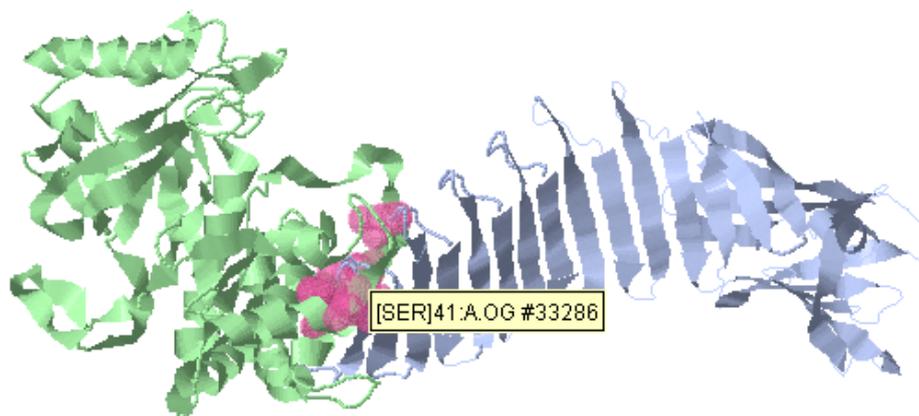


Figure 4.11. Complex structure of YER093C-YKL203C (Model 8)

As mentioned before, the distances between the hotspots were calculated in order to select the best model for each protein complex. Among the distances between the hotspots of proteins with known interaction partners, the smallest distance is 3.09Å. This is quite good value because 4Å is enough as a distance value from one atom to the opposite binding partner. Nevertheless, in this study the distance threshold between the hotspots is taken as 10Å. All these distance values are listed in the Appendix A.

#### 4.4.2. Proteins with predicted partners

Figure 4.12 shows the structure of YGR212W-YPL204W protein complex and its hotspots. While ASN210, ILE214, PHE216 are hotspots of YGR212W (blue); LEU244, PRO, LYS297, LEU280 are hotspots of YPL204W (green).

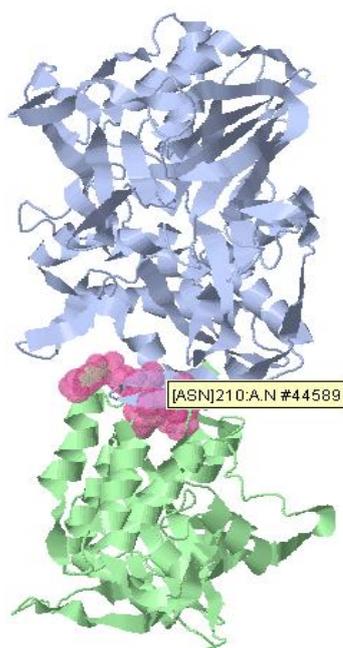


Figure 4.12. Complex structure of YGR212W-YPL204W (Model 8)

Figure 4.13 shows the structure of YDR294C-YHR135C protein complex and its hotspots. VAL429, TRP432, VAL436, ASN437, GLY439, GLU440 are hotspots of YDR294C (blue) and MET89, ILE90, ASN91, GLY92, PRO94, GLN201, ASP203 are hotspots of YHR135C (green).

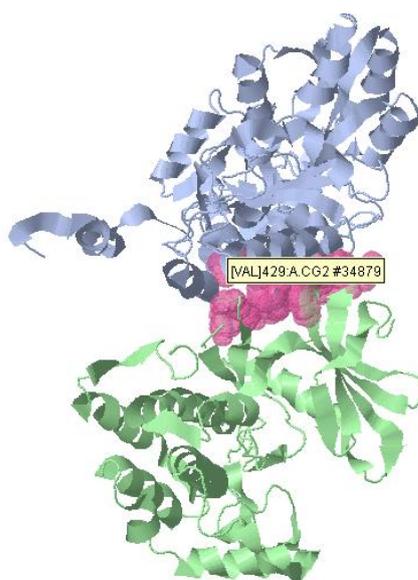


Figure 4.13. Complex structure of YDR294C-YHR135C (Model 7)

Figure 4.14 shows the structure of YER019W-YHL020C protein complex and its hotspots. While ALA158, PHE160, ASP211 are hotspots of YER019W (blue), ALA328, VAL332, LEU333, ALA336, LYS337, LEU340 are hotspots of YHL020C (green).

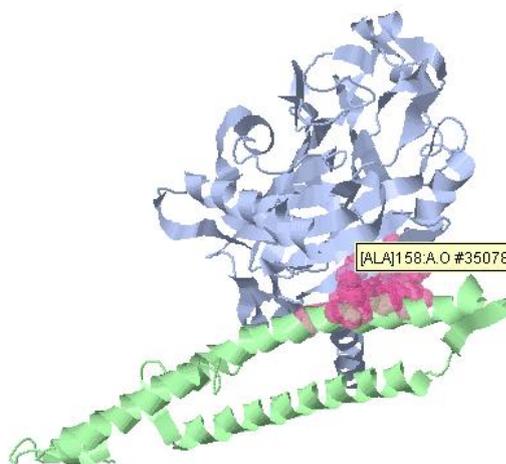


Figure 4.14. Complex structure of YER019W-YHL020C (Model 10)

Figure 4.15 shows the structure of YGR143W-YKL126W protein complex and its hotspots. While PHE585, ILE586, TYR589, ALA625, ILE692 are hotspots of YGR143W (blue), LEU494, GLY521, THR523, LYS524 are hotspots of YKL126W (green).

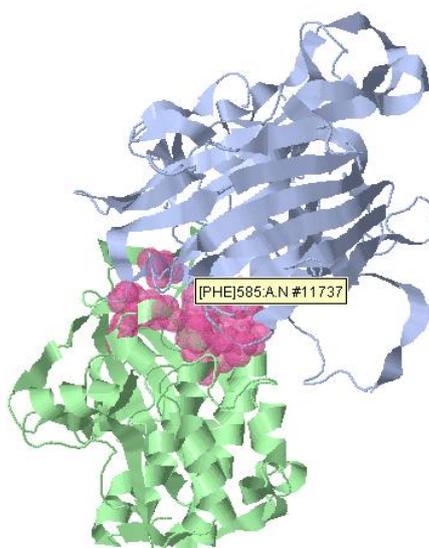


Figure 4.15. Complex structure of YGR143W-YKL126W (Model 3)

Figure 4.16 shows the structure of YJL134W-YAR033W protein complex and its hotspots. While TRP103(H), TYR106(H), PHE107(H), GLU108 are hotspots of YJL134W (blue), ASN178(P), PRO180(H), LEU198 are hotspots of YAR033W (green).

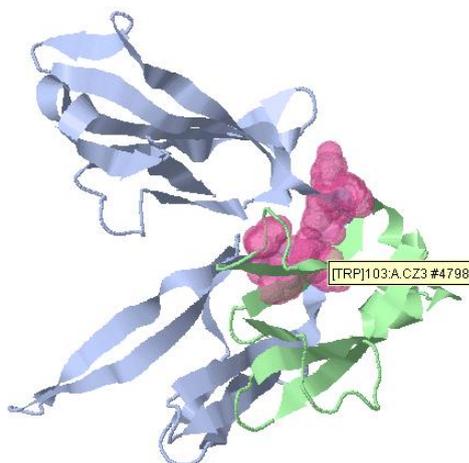


Figure 4.16. Complex structure of YJL134W-YAR033W (Model 3)

Figure 4.17 shows the structure of YKR053C-YAR033W protein complex and its hotspots. While VAL98, PRO102, VAL103, PRO185 are hotspots of YKR053C (blue), TYR148, PHE149, TYR150, GLU164 are hotspots of YAR033W (green).

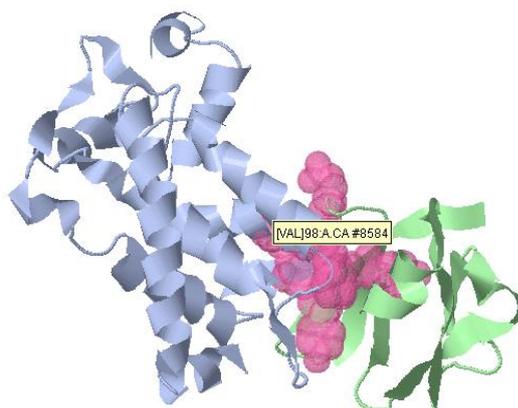


Figure 4.17. Complex structure of YKR053C-YAR033W (Model 5)

The smallest distance between the hotspots of the proteins with predicted partners is 4.26Å. This value is still good as a hotspot distance. However the distances between the hotspots are generally higher than 5Å among the proteins with predicted partners. All these distance values are listed in Table A.2 Appendix A.

#### 4.4.3. Clusters / Modules

Figure 4.18 shows the complex structure of YOR171C (blue)-YOR034C (green) pair. Using KFC, the interface residues and the binding hotspots of this complex structure were identified. The hotspots shown as pink spheres belonging to YOR171C and YOR034C pair are close to each other and are therefore on the binding regions of the proteins. The hotspots of this complex structure are MET 357, TYR 364, TRP 368, PRO 369, ARG 370 on YOR171C (blue) and VAL 232, ARG 237, VAL 238, CYS 240, LEU 244 on YOR034C (green).

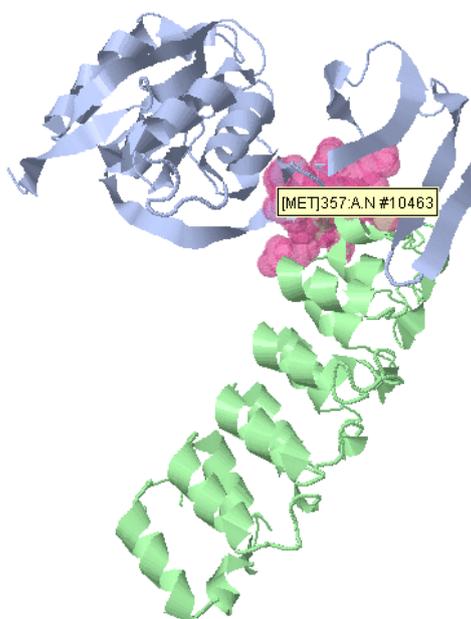


Figure 4.18. Complex structure of YOR171C-YOR034C (Model 4) in Cluster A

Figure 4.19 shows the structure of YOR034C-YOR171C protein complex and its hotspots. Blue chain represents the protein structure of YOR034C (blue) while green chain represents the protein structure of YOR171C. In this complex structure, TYR132, THR145, TYR169, VAL174, VAL175, ASN176 are hotspots of YOR034C; GLN361, PRO362, LEU374, TYR378, ILE381, PHE399, LEU401 are the hotspots of YOR171C (green).

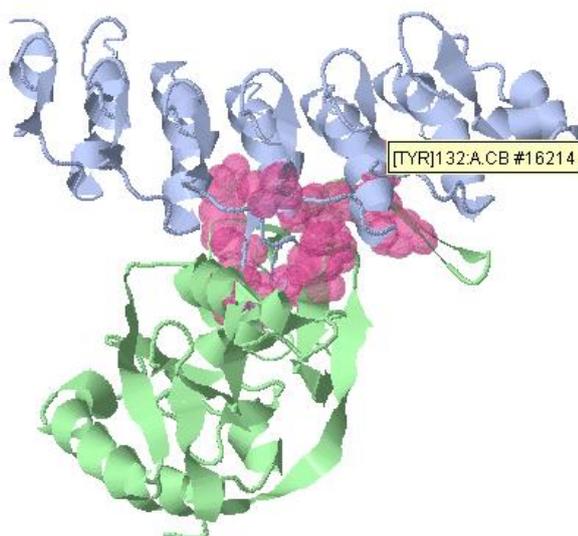


Figure 4.19. Complex structure of YOR034C-YOR171C (Model 6) in Cluster A

The reasons for choosing model 4 for YOR171C and YOR034C complex and model 6 for YOR034C and YOR171C complex are same and they both stem from the distance value between the of hotspots of the protein complex. The models which had closest hotspots were selected as ideal among the 10 models built by GRAMM-X.

Figure 4.20 shows the structure of YMR298W-YMR298W protein complex and its hotspots. THR44, ILE46 are hotspots of YMR298W (blue); SER74, VAL112, TYR114 are the hotspots of YMR298W (green).

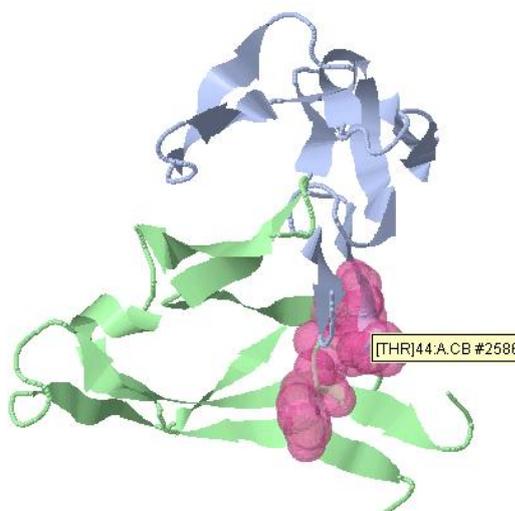


Figure 4.20. Complex structure of YMR298W-YMR298W (Model 3) in Cluster B

Figure 4.21 shows the structure of YGR060W-YGR060W protein complex and its hotspots. While PHE5, GLY12, GLN15 are hotspots of YGR060W (blue); HIS109, GLY146, LEU147 are hotspots of YGR060W (green).

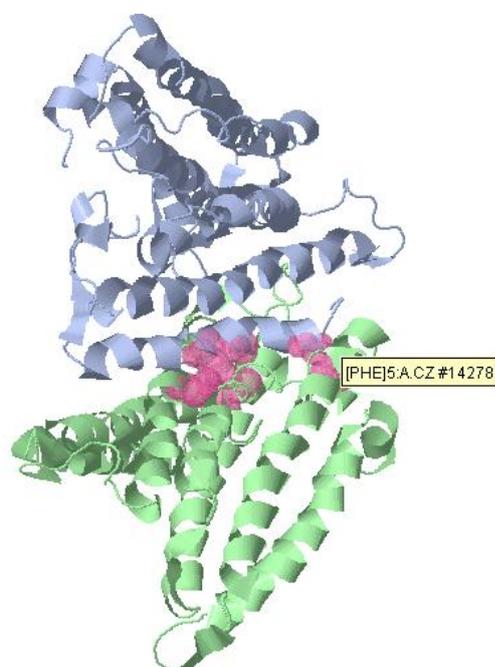


Figure 4.21. Complex structure of YGR060W-YGR060W (Model 4) in Cluster B

Figure 4.22 shows the structure of YEL027W-YHR026W protein complex and its hotspots. While ILE21, LEU26, ALA29, TYR76 are hotspots of YEL027W (blue); VAL163, THR166, GLU188, ILE193, LEU196, LEU197, ILE200 are hotspots of YHR026W (green).

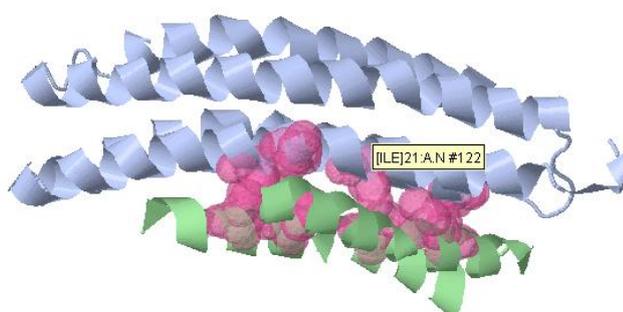


Figure 4.22. Complex structure of YEL027W-YHR026W (Model 1) in Cluster C

Figure 4.23 shows the structure of YEL027W-YOR016C protein complex and its hotspots. While ILE 21, LEU 26, ALA 28, ALA 29 are hotspots of YEL027W (blue); LYS 96, TYR 97, LEU 101, THR 115 are hotspots of YOR016C (green).

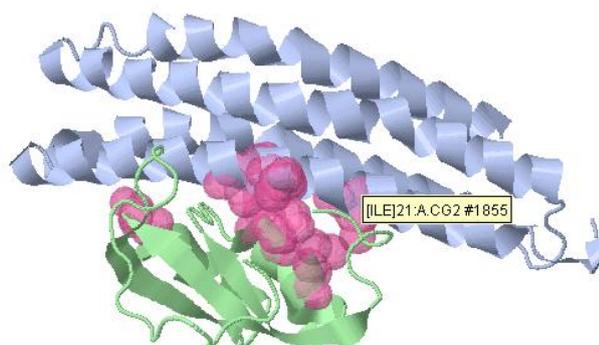


Figure 4.23. Complex structure of YEL027W-YOR016C (Model 2) in Cluster C

Distance values between the hotspots for the proteins in clusters are generally variable, while some hotspots are quite close to the each other some ones are very far away. For instance, the smallest distance between the hotspots of proteins is 4.49Å for Cluster A, 3.73Å for Cluster B, and 3.73Å for Cluster C. However, distances between some hotspots are again very high only if being less than the threshold value (10Å).

## 4.5. Analysis of Hotspots

### 4.5.1. Hydrophobicity

Protein–protein interfaces are frequently hydrophobic and bury a large extent of nonpolar surface area. Hence, hydrophobicity is a leading force in protein–protein interactions. Therefore, hot spots (an area of high energy and binding around an amino acid residue), surrounded by hydrophobic pockets, are found in clusters rather than scattered throughout the interface. They usually do not include hydrogen bonding and electrostatic interactions, because hydrogen bonding between the peptide atoms decreases the hydrophobicity of the backbone and electrostatic complementarity of interacting protein surfaces promotes complex formation and defines the lifetime of complexes (Moreira et al., 2007).

Here, in order to determine the hydrophobicity of the hotspot residues of the proteins of the sphingolipid pathway, chemical characteristics of these hotspots were identified by KFC. According to the results obtained from KFC they were classified as hydrophobic, acidic, polar and basic. The chemical features of hotspots of proteins and their interaction partners in sphingolipid pathway were given in the tables below (Tables 4.21-4.35).

4.5.1.1. Proteins with Known Interaction Partners: Table 4.21 shows the chemical types of the hotspot residues belonging to the proteins with known interaction partners. Based on the results obtained from KFC, the majority of the hotspot residues are hydrophobic. There are some hotspots which are of other chemical types such as polar, acidic and basic but the percentages of them are quite low compared to those of hydrophobic residues.

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners

Protein couples	Protein partners	Hydrophobic	Polar	Acidic	Basic
YLR242C-YOR033W	YLR242C	ILE174 LEU177 ALA180 PHE190	-	-	-
	YOR033W	GLY162 GLY163	ASN208	-	LYS60
YKL004W-YJL196C	YKL004W	ALA241 PRO259 ILE260	SER243	CYS267	ARG268
	YJL196C	ALA203 PHE235 VAL242 TYR243	-	-	-
YKL004W-YJR105W	YKL004W	PRO259 PRO296 LEU297	-	ASP262 GLU293	-
	YJR105W	ALA79 GLY80	-	GLN81	-
YKL004W-YMR010W	YKL004W	PHE266	-	GLU293	LYS254 HIS257
	YMR010W	LEU183 PHE186 GLY329	ASN182	ASP325	-
YKL004W-YPL264C	YKL004W	TYR271	THR272	GLU275	LYS281
	YPL264C	ILE49 ILE120 LEU131	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YKL008C- YKL008C	YKL008C	TYR95 ILE130	ASN131 CYS134	ASP132	-
	YKL008C	ILE130 PHE135 TYR138	-	ASP132	-
YKL008C- YHL003C	YKL008C	ILE130 LEU162 VAL164	ASN131 THR165	-	-
	YHL003C	LEU100 GLY102	ASN103	-	-
YKL008C- YMR298W	YKL008C	VAL153 VAL154 ILE155	-	ASP152	-
	YMR298W	PRO73 PRO111 VAL112	THR83	-	-
YKL008C- YKL065C	YKL008C	LEU91 ALA97	ASN103	-	-
	YKL065C	TYR76	THR120	ASP158	-
YKL008C- YBR159W	YKL008C	ILE92 LEU150 VAL153 PHE158 ILE173	-	-	-
	YBR159W	ILE215 ILE310	GLN231 THR313	GLU312	-
YHL003C- YMR298W	YHL003C	TYR267 VAL268	THR281	-	HIS270
	YMR298W	ILE84	THR83	-	HIS113
YDR062W- YMR296C	YDR062W	GLY372 LEU474 VAL482	-	-	-
	YMR296C	MET371 VAL396	THR194 GLN201	-	HIS374
YDR062W- YGR218W	YDR062W	TYR226 MET307 MET466 PRO467	-	-	-
	YGR218W	VAL1042	THR1031	GLU1041	-
YDR062W- YLR342W	YDR062W	ILE349 PHE350 GLY351	-	GLU348 ASP356	-
	YLR342W	GLY10	ASN72	-	-
YDR062W- YKL104C	YDR062W	MET307 PRO467 VAL482 ALA484 TYR485 PRO486 LEU490	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YDR062W- YKL104C	YKL104C	GLY482 VAL488	-	-	ARG487 HIS490
YDR062W- YBR017C	YDR062W	ILE446 TYR461 PRO463 MET472	-	GLU492	ARG471
	YBR017C	ILE413 PRO415	ASN412	-	HIS416
YDR062W- YER110C	YDR062W	MET466 PRO467 ALA468 VAL481	-	-	ARG471
	YER110C	PHE163 LEU164 LEU166	-	-	-
YDR062W- YJR077W	YDR062W	TYR226 VAL483 TYR485 PRO486	THR255	-	ARG254
	YJR077W	ILE166 PRO175 LEU257	-	GLU164	-
YOR171C- YOR034C	YOR171C	MET357 TYR364 TRP368 PRO369	-	-	-
	YOR034C	VAL232 VAL238 LEU244	CYS240	-	-
YLR260W- YGL137W	YLR260W	TYR383	SER374 THR378 CYS387	-	HIS376
	YGL137W		ASN256	GLU263	LYS264
YLR260W- YLR213C	YLR260W	ILE315 ILE323	-	ASP322	LYS325
	YLR213C	LEU240 GLY311 GLY312	-	-	--
YMR298W- YKL008C	YMR298W	GLY72 PRO73 PRO111	SER74 CYS75 THR83	ASP76	-
	YKL008C	TYR95 PHE99	-	LYS128	
YMR298W- YHL003C	YMR298W	TRP68 GLY72	-	ASP76	ARG78
	YHL003C	TYR95 GLY129	SER96	-	-
YMR298W- YMR298W	YMR298W	TYR48 TRP50 PHE51 VAL64	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YMR298W- YMR298W	YMR298W	TYR81	-	-	ARG90
YJL097W- YBR159W	YJL097W	ILE58 LEU73 VAL103 VAL104	GLN54 CYS55 THR75	-	-
	YBR159W	PRO218 LEU219	-	-	-
YPL006W- YGL006W	YPL006W	LEU1002 LEU1009 ALA1030	THR1034	-	-
	YGL006W	LEU904 ILE904 ILE905 ILE912 LEU969 ILE970	-	-	LYS898
YDR297W- YBR106W	YDR297W	ILE272 ILE279	-	-	-
	YBR106W	PRO26 GLY102	-	GLU67 GLU77	-
YDR297W- YLR372W	YDR297W	TYR252 ALA253 PHE290	CYS250 GLN285 THR289	-	-
	YLR372W	TYR135 GLY138 PHE140 PHE173 LEU174	THR163	-	LYS164
YMR272C- YOR081C	YMR272C	TYR30	GLN31	GLU71	LYS65 HIS70
	YOR081C	ILE564 ALA593 PRO594	-	ASP573	ARG689
YDR297W- YKL065C	TRP261 ALA267 VAL268	-	-	-	-
	TYR106 GLY109 ALA157	-	-	-	-
YDR297W- YBR159W	YDR297W	GLY251 TYR252 ALA253	CYS250	-	-
	YBR159	PHE158 LEU214 ILE215 LEU281	-	-	-
YDR297W- YPL264C	YDR297W	PHE283 ALA284	-	-	-
	YPL264C	ILE21 PHE130	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YDR297W- YGR125W	YDR297W	LEU256 TRP261	-	ASP257	-
	YGR125W	LEU148	ASN154	-	-
YDR297W- YHR007C	YDR297W	LEU254 LEU256 PRO258	-	ASP257	-
	YHR007C	PRO503 PRO504	-	-	-
YDR297W- YML012W	YDR297W	VAL268	GLN275 GLN276	ASP257	-
	YML012W	LEU62 PHE88 PRO91	THR89	-	ARG75 LYS138
YDR297W- YKL088W	YDR297W	GLY251 ALA253 PHE277	THR281 ASN282 GLN285	-	-
	YKL088W	LEU414 TYR463 PHE465	THR457 SER458	-	HIS455
YLR372W- YDL015C	YLR372W	TRP141 TYR161 PHE173 VAL175	THR163	-	-
	YDL015C	PRO56 VAL57 ILE58	-	GLU60	ARG45
YER093C- YNL006W	YER093C	ALA101 LEU102 LEU103 LEU105	SER81 THR97 THR99	ASP124	LYS149
	YNL006W	VAL68	ASN87	-	LYS45
YER093C- YBR270C	YER093C	LEU232 GLY238	THR226 THR239	GLU230	-
	YBR270C	PRO21 LEU45	ASN22 GLN43	ASN39	-
YER093C- YKL203C	YER093C	LEU68	SER41 ASN45	-	-
	YKL203C	VAL2316 PRO2317 MET2323	GLN2238 THR2321	-	-
YER093C- YMR068W	YER093C	PRO88 LEU103 LEU105	ASN65 THR86	-	-
	YMR068W	MET118 TYR121 TYR152	ASN119	GLU15 ASN120	HIS50 LYS85 HIS87
YER093C- YMR068W	YMR068W	MET118 TYR121 TYR152	ASN119	GLU15 ASN120	HIS50 LYS85 HIS87
YBR058C-A- YMR296C	YBR058C-A	LEU46	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YBR058C-A- YMR296C	YMR296C	GLY329 GLY332 MET403	THR331 ASN406 ASN407	ASP404	-
YBR058C-A- YDR062W	YBR058C-A	TYR43 LEU46	-	-	HIS45
	YDR062W	ALA179 LEU180 VAL183 PRO399 PRO401	-	-	-
YPL087W- YDL015C	YPL087W	TYR182 LEU234 LEU240 TRP244	-	-	HIS241
	YDL015C	LEU78 LEU86 LEU101	-	-	HIS99
YBR183W- YLR372W	YBR183W	PHE205 PRO233 LEU237 PRO240	-	-	-
	YLR372W	LEU139 TYR157 TYR158	GLU129 ASN160	GLU128	-
YBR183W- YOR016C	YBR183W	GLY201 LEU204 GLY235	-	GLU239	-
	YOR016C	PHE92 ILE138	SER91	-	-
YBR183W- YLR018C	YBR183W	PHE275	ASN272	-	-
	YKL065C	TYR187 VAL188	SER189	-	-
YBR183W- YKL065C	YBR183W	PHE254 TYR255 ILE256	-	-	-
	YKL065C	-	SER159 THR160	-	-
YBR183W- YHR140W	YBR183W	LEU204 LEU230 LEU234 LEU237 LEU238 PRO240	-	-	HIS241
YBR058C-A- YMR296C	YBR058C-A	LEU46	-	-	-
	YMR296C	GLY329 GLY332 MET403	THR331 ASN406 ASN407	ASP404	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YBR058C-A- YDR062W	YBR058C-A	TYR43 LEU46	-	-	HIS45
	YDR062W	ALA179 LEU180 VAL183 PRO399 PRO401	-	-	-
YPL087W- YDL015C	YPL087W	TYR182 LEU234 LEU240 TRP244	-	-	HIS241
	YDL015C	LEU78 LEU86 LEU101	-	-	HIS99
YBR183W- YLR372W	YBR183W	PHE205 PRO233 LEU237 PRO240	-	-	-
	YLR372W	LEU139 TYR157 TYR158	GLU129 ASN160	GLU128	-
YBR183W- YOR016C	YBR183W	GLY201 LEU204 GLY235	-	GLU239	-
	YOR016C	PHE92 ILE138	SER91	-	-
YBR183W- YKL065C	YBR183W	PHE275	ASN272	-	-
	YKL065C	TYR187 VAL188	SER189	-	-
YBR183W- YKL065C	YBR183W	PHE254 TYR255 ILE256	-	-	-
	YKL065C	-	SER159 THR160	-	-
YBR183W- YHR140W	YBR183W	LEU204 LEU230 LEU234 LEU237 LEU238 PRO240	-	-	HIS241
	YHR140W	ILE65 LEU97	-	GLU98	-
YBR183W- YBR159W	YBR183W	VAL264 ILE265 PHE275	-	-	-
	YBR159W	LEU214 TYR254	THR313	-	-
YBR183W- YBR106W	YBR183W	-	-	-	-
	YBR106W	ILE99	-	-	ARG33

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YBR183W- YGR060W	YBR183W	PHE205 VAL236 PRO240 TRP243 TRP244 LEU247 TYR253 LEU267 PHE275 ILE276	ASN272	-	-
	YGR060W	TRP43 MET47	ASN25 ASN48	-	HIS28
YBR183W- YFL041W	YBR183W	PHE205 TYR209 VAL236 TRP243	-	-	-
	YFL041W	LEU379 GLY382 VAL516	SER381 ASN403 GLN519	ASP341	LYS401
YBR183W- YDR506C	YBR183W	ALA199 ILE203 LEU206 TYR209 ILE210 VAL236	-	-	-
	YDR506C	ILE62	-	-	LYS247
YKL126W- YMR104C	YKL126W	-	ASN495	ASP500	LYS497
	YMR104C	LEU441 ALA442 PRO597 TRP604 PRO615	SER603	-	-
YKL126W- YNL106C	YKL126W	PRO561 LEU562 VAL563 PHE564 PRO565	-	GLU560	-
	YNL106C	VAL574 TYR749 VAL897	THR759	GLU576 ASP764	LYS758 HIS762 HIS763
YDL015C- YKL008C	YDL015C	PRO109	-	ASP69	LYS41 ARG43
	YKL008C	TYR126	ASN131	-	-
YDL015C- YLR372W	YDL015C	LEU86 LEU93 VAL96 LEU97 VAL98	-	-	ARG85
YDL015C- YLR372W	YLR372W	PHE108 MET133 MET134	-	GLU109	-
YDL015C- YPL087W	YDL015C	VAL111 VAL112	THR110	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YDL015C- YPL087W	YPL087W	TYR166 LEU169 VAL206 PHE209	-	-	-
YDL015C- YDL015C	YDL015C	VAL55 PRO56	SER120	-	-
	YDL015C	LEU46 TYR48	SER61	GLU66	-
YDL015C- YJL196C	YDL015C	LEU101 TRP115	CYS90 SER100	-	HIS99
	YJL196C	ALA203 LEU207 PHE213 LEU239 TYR243	-	ASP238	-
YDL015C- YOR016C	YDL015C	TRP84 TYR104	-	-	-
	YOR016C	ILE138	SER91 THR115 ASN136	GLU139	-
YDL015C- YKL065C	YDL015C	PHE63 TRP115	SER117	-	-
	YKL065C	GLY82	SER81	-	ARG98
YDL015C- YBR017C	YDL015C	VAL111	-	ASP113	-
	YBR017C	-	GLN179 THR225	-	-
YDL015C- YBR106W	YDL015C	LEU78 LEU101 TYR103	-	-	-
	YBR106W	ILE92 ILE96	-	ASP93	-
YDL015C- YBR159W	YDL015C	ILE17	-	GLU16 ASP18	HIS37
	YBR159W	TYR254 TYR294 MET297	THR296	GLU292	-
YDL015C- YDR506C	YDL015C	ALA67	THR110	ASP113	LYS10
	YDR506C	PRO392 VAL395 PHE424	-	ASP405	-
YDL015C- YPR028W	YDL015C	VAL55 TRP115 ALA118 TYR122	SER117 SER120	ASP121 ASN123	-
YDL015C- YPR028W	YPR028W	TYR47 LEU48 TYR73 VAL77	-	-	-
YDL015C- YBR110W	YDL015C	TYR92 PRO95 LEU97 VAL98	-	-	-

Table 4.21. Chemical types of hotspot residues in proteins with known interaction partners  
(continued)

YDL015C- YBR110W	YBR110W	ILE168	SER167 GLN170	-	-
YDL015C- YBR094W	YDL015C	PHE89 LEU101	CYS90	-	HIS99
	YBR094W	PRO112 GLY150	SER148 SER149	-	-
YDL015C- YKL182W	YDL015C	LEU78 TRP84 LEU101 TYR103 TYR104 TYR122	-	ASP121	-
	YKL182W	PHE1664 LEU1792 LEU1797 LEU1798	-	-	-

Table 4.22 shows the distribution of the hotspot residues belonging to the proteins with known interaction partners. Among the hotspots of the proteins with known partners, LEU is the most prevalent hotspot residue with the percentage of 13.47. However, CYS is the least appeared hotspot residue with the percentage of 1.68. In addition to this, as shown in the Table 4.23, LEU also has the highest percentage, 21.40%, among the hydrophobic hotspot residues of proteins with known partners while MET has the lowest percentage, 3.74%.

Table 4.22. Amino acid propensities of hotspots in proteins with known interaction partners

Names of the Hotspot Residues	% of the Residues in proteins with known interaction	Name of the Hotspot Residues	% of the Residues in proteins with known interaction
ILE	7.41	PRO	7.07
LEU	13.47	VAL	7.57
ALA	3.37	TYR	8.58
PHE	6.06	ASP	4.71
GLY	4.38	GLU	4.38
ASN	5.05	GLN	2.36
LYS	3.20	HIS	3.53
SER	3.87	THR	5.89
CYS	1.68	MET	2.36
ARG	2.36	TRP	2.69

Table 4.23. Distribution of the hotspot residues of proteins with known interaction partners according to chemical features

Hydrophobic Hotspot Residues	% Presence of Residues in Hydrophobic Hotspots	Polar Hotspot Residues	% Presence of Residues in Polar Hotspots	Acidic Hotspot Residues	% Presence of Residues in Acidic Hotspots	Basic Hotspot Residues	% Presence of Residues in Basic Hotspots
ILE	11.76	ASN	25.0	CYS	1.69	LYS	33.96
LEU	21.40	SER	21.30	ASP	47.46	ARG	26.41
ALA	5.35	THR	32.41	GLU	42.37	HIS	39.62
PHE	9.62	CYS	8.33	GLN	1.69		
GLY	6.95	GLN	12.04	LYS	1.69		
PRO	11.23	GLU	0.92	ASN	5.08		
VAL	12.03						
TYR	13.64						
MET	3.74						
TRP	4.28						

4.5.1.2. Proteins with predicted interaction partners: Table 4.24 shows the chemical features of the hotspot residues of the proteins with predicted interaction partners. The hydrophobic residues are highly dominating in this group compared to polar, acidic or basic residues.

Table 4.24. Chemical types of hotspots residues in proteins with predicted interaction partners

Protein couples	Protein partners	Hydrophobic	Polar	Acidic	Basic
YDR294C-YHR135C	YDR294C	VAL429 TRP432 VAL436 GLY439	ASN437	GLU440	-
	YHR135C	MET89 ILE90 GLY92 PRO94	ASN91 GLN201	ASP203	-
YER019W-YHL020C	YER019W	ALA158 PHE160	-	ASP211	-
	YHL020C	ALA328 VAL332 LEU333 ALA336 LEU340	-	-	LYS337
YJL134W-YAR033C	YJL134W	TRP103 TYR106 PHE107	-	GLU108	-
	YAR033C	PRO180 LEU198	ASN178	-	-

Table 4.24. Chemical types of hotspots residues in proteins with predicted interaction partners (*continued*)

YJL134W- YGL051W	YJL134W	TYR106 PHE107 PRO136	THR109 THR110	-	LYS111
	YGL051W	PHE149 TYR150 TRP219	-	-	-
YGR143W- YKL126W	YGR143W	PHE585 ILE586 TYR589 ALA625 ILE692 TYR696	-	-	-
	YKL126W	LEU494 GLY521	THR523	-	LYS524
YKR053C- YAR033W	YKR053C	VAL98 PRO102 VAL103 PRO185	-	-	-
	YAR033W	TYR148 PHE149 TYR150	-	GLU164	-
YKR053W- YGL051W	YKR053W	ILE115 GLY119 VAL233	ASN239	-	-
	YGL051W	GLY165 VAL166	THR167	ASP154	-
YGR212W- YHR135C	YGR212W	PRO222	ASN384	ASP224	-
	YHR135C	LEU67 GLY92 VAL93 PRO94	ASN91	-	-
YGR212W- YPL204W	YGR212W	ILE214 PHE216	ASN201	-	-
	YPL204W	LEU244 PRO245 LEU280	-	-	LYS297

Table 4.25 shows the participation percentages of amino acids among all the hotspot residues belonging to the proteins with predicted interaction partners. When Table 4.25 is examined it is observed that the mostly appeared amino acids are VAL (10.13 %) and PRO (10.13 %), whereas the rarely appeared amino acids are MET (1.26%) and GLN (1.26%).

Table 4.26 gives the distribution of amino acids based on their chemical types. As we know from Table 4.24, the majority of the hotspot residues are hydrophobic for the proteins with predicted interaction partners and to this end the percentages of the amino acids among hydrophobic hotspots is similar to the results of the whole group. In other

words, the mostly appeared amino acids are again VAL (14.0%) and PRO (14.28%) and the rarely appeared amino acid is MET (1.78%) in hydrophobic case.

Table 4.25. Amino acid propensities of hotspots in proteins with predicted Interactions

Name of the Hotspot Residues	% of the Residues in proteins with predicted interactions	Name of the Hotspot Residues	% of the Residues in cluster proteins with predicted interactions
VAL	10.13	GLN	1.26
ASN	8.86	PRO	10.13
GLU	3.80	ALA	5.06
TRP	3.80	PHE	8.86
GLY	7.59	LYS	5.06
MET	1.26	LEU	8.86
ASP	5.06	TYR	8.86
ILE	6.33	THR	5.06

Table 4.26. Distribution of hotspot residues of proteins with predicted interactions partners according to chemical features

Hydrophobic Hotspot Residues	% Presence of Residues in Hydrophobic Hotspots	Polar Hotspot Residues	% Presence of Residues in Polar Hotspots	Acidic Hotspot Residues	% Presence of Residues in Acidic Hotspots	Basic Hotspot Residues	% Presence of Residues in Basic Hotspots
VAL	14.0	ASN	58.33	GLU	42.86	LYS	100.0
TRP	5.36	GLN	8.33	ASP	57.14		
GLY	10.71	THR	33.33				
MET	1.78						
ILE	8.93						
PRO	14.28						
ALA	7.14						
PHE	12.5						
LEU	12.5						
TYR	12.5						

**4.5.1.3. Clusters/ Modules:** Table 4.27 shows the chemical types of the hotspot residues of Cluster A. These results were obtained from KFC server. As it is seen from Table 4.27, the majority of the hotspots are hydrophobic in Cluster A. Secondly, polar residues are predominant. However, there are no acidic hotspot residues (Aspartic Acid (Asp) and Glutamic acid (Glu)) in Cluster A.

Table 4.27. Chemical types of hotspot residues of proteins in Cluster A

Protein couples	Potein partners	Hydrophobic	Polar	Acidic	Basic
YOR171C- YOR034C	YOR171C	MET357 TYR364 TRP368 PRO369	-	-	PRO370
	YOR034C	VAL232 VAL238	CYS240	-	ARG237
YOR034C- YOR171C	YOR034C	TYR132 TYR169 VAL174 VAL175	THR145 ASN176	-	-
	YOR171C	PRO362 LEU374 ILE381 PHE399 LEU401	GLN361	-	-

Table 4.28 shows the frequency distribution of the amino acids in hotspots of Cluster A. Based on these results in Table 4.28, VAL has the maximum percentage (19.05%) among the hotspot amino acids of Cluster A. Secondly, TYR (14.28%) and PRO (14.28%) are the mostly appearing amino acids. Besides, MET, TRP, CYS, ARG, THR, ASN, GLN, ILE, PHE have the same minimum appearance percentage of 4.76%.

Table 4.28. Amino acid propensities of hotspots of proteins Cluster A

Name of the Hotspot Residues	% of the Residues in Cluster A	Name of the Hotspot Residues	% of the Residues in Cluster A
MET	4.76	THR	4.76
PRO	14.28	ASN	4.76
TYR	14.28	GLN	4.76
TRP	4.76	LEU	9.52
VAL	19.05	ILE	4.76
CYS	4.76	PHE	4.76
ARG	4.76		

The percentage of the amino acids belonging to four main chemical types of hotspots is listed in Table 4.29. According to the values in Table 4.29, VAL has the maximum percentage through hydrophobic hotspot residues (26.67%). However, for other chemical types, the percentages of the amino acids are same.

Table 4.29. Distribution of hotspot residues of proteins in Cluster A according to chemical features

Hydrophobic Hotspot Residues	% Presence of Residues in Hydrophobic Hotspots	Polar Hotspot Residues	%Presence of Residues in Polar Hotspots	Acidic Hotspot Residues	%Presence of Residues in Acidic Hotspots	Basic Hotspot Residues	% Presence of Residues in Basic Hotspots
MET	6.67	CYS	25.0	-	-	PRO	50.0
TYR	20.0	THR	25.0			ARG	50.0
TRP	6.67	ASN	25.0				
PRO	13.33	GLN	25.0				
VAL	26.67						
LEU	13.33						
ILE	6.67						
PHE	6.67						

Table 4.30 shows the chemical types of the hotspots residues in the proteins of Cluster B. According to the data obtained from KFC, the results are nearly same as those of Cluster A concerning the abundances of the hydrophobic residues. Besides, there is not any acidic hotspot residue in Cluster B as well as in the Cluster A.

Table 4.30. Chemical types of hotspot of residues of proteins in Cluster B

Protein couples	Protein partners	Hydrophobic	Polar	Acidic	Basic
YMR298W-YMR298W	YMR298W	ILE46	THR44	-	
	YMR298W	VAL112 TYR114	SER74	-	-
YHL003C-YMR298W	YHL003C	VAL154 GLY187	SER186	-	-
	YMR298W	PHE51 GLY62	CYS53	-	HIS52
YGR060W-YGR060W	YGR060W	PHE5 GLY12	GLN15	-	-
	YGR060W	GLY146 LEU147	-	-	HIS109
YMR298W-YHL003C	YMR298W	TYR81 TYR114 PRO115 ILE116	-	-	ARG90
	YHL003C	PHE230	GLN233 GLN234	-	LYS205
YGR060W-YHL003C	YGR060W	MET65 VAL223	-	-	HIS191

Table 4.30. Chemical types of hotspot of residues of proteins in Cluster B (*continued*)

YGR060W- YHL003C	YHL003C	LEU85 PRO87 PHE307	-	-	-
YMR298W- YHL003C	YMR298W	TYR81 TYR114 PRO115 ILE116	-	-	ARG90
	YHL003C	PHE230	GLN233 GLN234	-	LYS205

Table 4.31 shows the percentages of the hotspot amino acids of proteins in Cluster B. The majority of the hotspot residues are TYR (11.90%), PHE (11.90%), GLN (11.90%) while the minority of them are THR (2.38%), CYS (2.38%), MET (2.38%). When the results of Cluster A and Cluster B are compared it is seen that in both of the clusters TYR has higher percentage values whereas THR, CYS, MET have lower percentage values.

Table 4.31. Amino acid propensities of hotspots of proteins in Cluster B

Name of the Hotspot Residues	% of the Residues in Cluster B	Name of the Hotspot Residues	% of the Residues in Cluster B
ILE	7.14	HIS	7.14
THR	2.38	GLN	11.90
VAL	7.14	LEU	4.76
SER	4.76	ARG	4.76
TYR	11.90	PRO	7.14
GLY	9.52	LYS	4.76
PHE	11.90	MET	2.38
CYS	2.38		

Table 4.32 shows amino acids distribution of hotspot residues of proteins in Cluster B in terms four main chemical types (hydrophobic, polar, acidic, basic). TYR (19.23%) and PHE (19.23%) are the mostly appearing amino acids through hydrophobic hotspot residues. Besides, GLN (55.55%) has the maximum percentage among the polar residues while HIS (42.86%) has it among the basic residues.

Table 4.32. Distribution of hotspot residues of proteins in Cluster B according to chemical features

Hydrophobic Hotspot Residues	% Presence of Residues in Hydrophobic Hotspots	Polar Hotspot Residues	% Presence of Residues in Polar Hotspots	Acidic Hotspot Residues	% Presence of Residues in Acidic Hotspots	Basic Hotspot Residues	% Presence of Residues in Basic Hotspots
ILE	11.54	THR	11.11	-	-	HIS	42.86
VAL	11.54	SER	22.22			ARG	28.57
TYR	19.23	CYS	11.11			LYS	28.57
GLY	15.38	GLN	55.55				
PHE	19.23						
LEU	7.69						
PRO	11.54						
MET	3.85						

Table 4.33 shows the distribution of the hotspots of the proteins in Cluster C in according to chemical types. As in the other clusters, hydrophobic residues have the majority in Cluster C. However, the significant difference here is the existence of acidic hotspot residues. There are five acidic hotspot residues, GLN231, GLU188, ASP283, ASP433, ASP160 belonging to protein complexes in Cluster C.

Table 4.33. Chemical types of hotspot residues of proteins in Cluster C

Protein couples	Protein partners	Hydrophobic	Polar	Acidic	Basic
YLL006W-YLL006W	YLL006W	GLY231 LEU237 LEU241	-	-	-
	YLL006W	GLY231 LEU237	ASN202	-	-
YNL307C-YNL307C	YNL307C	PRO71 GLY213	CYS214 THR215	-	-
	YNL307C	PRO71 PRO195 GLY213	CYS214	-	-
YNL307C-YAL007C	YNL307C	-	-	-	-
	YAL007C	No templates	No templates	No templates	No templates
YDR297W-YOR016C	YDR297W	ILE272 ILE279 PHE287	GLN275	-	-

Table 4.33. Chemical types of hotspot residues of proteins in Cluster C (*continued*)

YDR297W- YOR016C	YDR297W	ILE272 ILE279 PHE287	GLN275	-	-
	YOR016C	-	ASN103 THR107 ASN135 ASN146 LYS147	--	LYS129 LYS147
YMR058W- YOR016C	YMR058W	VAL284	-	-	-
	YOR016C	ILE194	CYS100	-	LYS96 HIS126
YDR297W- YCR034W	YDR297W	GLY278 PHE288 PRO286	GLN285	-	-
	YCR034W	VAL233	-	-	ARG235
YHR026W- YAL007C	YHR026W	-	-	-	-
	YAL007C	No templates	-	-	-
YEL027W- YEL027W	YEL027W	PRO10 PHE11 GLY36 VAL37 ILE39 VAL57	CYS17 CYS40 GLN80	-	LYS81
	YEL027W	ALA33 GLY36 VAL37 ILE39 ILE65 LEU78 GLY79	THR32 CYS40 ASN53 SER71 GLN80	-	LYS81
YLL006W- YCR034C	YLL006W	LEU237	-	-	ARG232 LYS233
	YCR034C	LEU220 ILE238	THR234	GLN231	ARG223
YEL027W- YHR026W	YEL027W	ILE21 LEU26 ALA29 TYR76			
	YHR026W	VAL163 ILE193 LEU196 LEU197 ILE200	THR166	GLU188	
YMR058W- YCR034C	YMR058W	PRO181 VAL284 MET345	THR182 THR410	ASP283 ASP433	
	YCR034C	TRP216 LEU220 GLY224			
YHR110W- YAL007C	YHR110W	-	-	-	-
	YAL007C	No templates	No templates	No templates	No templates

Table 4.33. Chemical types of hotspot residues of proteins in Cluster C (*continued*)

YHR026W- YHR026W	YHR026W	ILE158 ILE165 ILE184	SER155	-	-
	YHR026W	ALA154 ILE158 ILE165 LEU180	SER155 THR169	-	LYS183
YEL027W- YOR016C	YEL027W	ILE21 LEU26 ALA28 ALA29	-	-	-
	YOR016C	TYR97 LEU101	THR115	-	LYS96
YHR026W- YOR016C	YHR026W	GLY195 LEU199 GLY202 LEU203	-	-	-
	YOR016C	ILE176 VAL177 LEU89 GLY93 ILE94	SER91	-	LYS96
YHR110W- YOR016C	YHR110W	LEU141 VAL145 LEU152	-	-	-
	YOR016C	TYR154 LEU155	-	-	LYS82
YHR110W- YHR110W	YHR110W	ALA129 GLY143	-	-	-
	YHR110W	MET138	-	-	LYS135
YHR110W- YCR034W	YHR110W	-	-	ASP160	ARG156 LYS157
	YCR034W	ILE239 LEU243	-	-	-

Table 4.34 shows the percentage of the amino acids belonging to the hotspots of proteins in Cluster C. In this group, ILE (14.28%) and LEU (15.04%) are the residues which have the maximum percentage, and TRP (0.75%), GLU (0.75%), HIS (0.75%), MET (1.5%) have the minimum percentages. As it is understood from these values, MET and TRP again get the minimum percentages.

Table 4.34. Amino acid propensities of hotspots of proteins in Cluster C

Name of the Hotspot Residues	% of the Residues in Cluster C	Name of the Hotspot Residues	% of the Residues in Cluster C
GLY	9.77	VAL	6.77
LEU	15.04	HIS	0.75
ASN	3.76	ARG	3.00

Table 4.34. Amino acid propensities of hotspots of proteins in Cluster C (*continued*)

PRO	4.51	ALA	4.51
CYS	4.51	SER	3.00
THR	6.77	TYR	2.25
ILE	14.28	GLU	0.75
GLN	3.76	ASP	2.25
PHE	2.25	MET	1.50
LYS	9.77	TRP	0.75

Table 4.35 shows the amino acid distribution of the hotspot residues of proteins in Cluster C in terms of four main chemical groups (hydrophobic, polar, acidic, basic). The results show that the distributions of the amino acids in hydrophobic hotspots are consistent with those of the whole Cluster C. In other words, LEU (24.39%) and ILE (23.17%) are the mostly appearing residues again and TRP (1.22%) and MET (2.44%) are the least apparent ones.

Table 4.35. Distribution of hotspot residues of proteins in Cluster C according to chemical features

Name of the Hydrophobic Hotspot Residues	% Presence of Residues in Hydrophobic Hotspots	Name of the Polar Hotspot Residues	% Presence of Residues in Polar Hotspots	Name of the Acidic Hotspot Residues	% Presence of Residues in Acidic Hotspots	Name of the Basic Hotspot Residues	% Presence of Residues in Basic Hotspots
GLY	15.85	ASN	17.24	GLN	20.0	LYS	70.59
LEU	24.39	CYS	20.69	GLU	20.0	HIS	5.88
PRO	7.32	THR	31.03	ASP	60.0	ARG	23.53
ILE	23.17	GLN	13.79				
PHE	3.66	LYS	3.45				
VAL	10.97	SER	13.79				
ALA	7.32						
TYR	3.66						
MET	2.44						
TRP	1.22						

When the amino acids concentrations were examined for all groups, it is seen that some results are quite good and consistent with the literature findings albeit some results in contradiction with the literature. Literature reporting that tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) have very high concentrations as hotspots. As consistent

with the literature, Tyr has very high concentration in Cluster A (14.28%) and in Cluster B (11.90%). However, Trp has very low concentration in proteins with known partners (2.69%) and also in Cluster C (0.75%). For the examples of other contradictory results, Val has very high concentration in proteins with predicted partners (10.13%) and in Cluster A (14.28%); and Leu has also quite high concentration in proteins with known partners (13.47%) and in Cluster C (15.64%) although in the literature it is reported that leucine, serine, threonine, and valine residues are disfavored and essentially absent as hot spots albeit their importance for distinct protein structures.

Table 4.36 shows the distribution percentages of chemical types of the hotspot residues in five groups which are proteins with known interactions, proteins with predicted interactions, Clusters A, B and C. According to the results, the hotspots mostly appeared as hydrophobic at very high percentages for each group changing between 62-75%. The percentages of the hydrophobic residues are higher than half the amount of whole hotspots for each group. These results are consistent with the literature which shows that protein–protein interfaces are frequently hydrophobic. Hence, hydrophobicity is a leading force in protein–protein interactions (Moreira et al., 2007). Hydrophobic interactions are essential for stabilizing protein-protein complexes, whose interfaces generally consist of a central cluster of hot spot residues surrounded by less important peripheral residues (Li et al., 2005). On the other hand, the chemical type of hotspots at minimum appearance is basic for proteins with known interaction partners (8.92%) and for proteins with predicted interaction partners (5.06%), while it is acidic for proteins in Clusters A (0.0%), B (0.0%) and C (3.76%).

Table 4.36. Percentage of Chemical types of Hotspots

	Hydrophobic	Polar	Acidic	Basic
Known Interactions	62.96	18.19	9.93	8.92
Predicted Interactions	70.89	15.19	8.86	5.06
Cluster A	71.43	19.05	0.0	9.52
Cluster B	61.90	21.43	0.0	16.67
Cluster C	61.65	21.80	3.76	12.78

## 5. CONCLUSIONS AND RECOMMENDATIONS

### 5.1. Conclusions

In this study, the protein interaction map of sphingolipid (SL) signaling proteins was constructed using the new advances in computational prediction methods in order to decipher the missing interactions among the components of sphingolipid PPI network and to get an insight about the metabolism. The conclusions obtained from the study are as follows:

- The interaction partners of YDR294C, YER019W, YGR143W, YGR212W, YJL134W, YKR053C were identified by PIPE and the putative interaction partners were determined as; YHR135C with YDR294C; YHL020C with YER019W; YKL126W and YHL020C with YGR143W; YPL204W and YHR135C with YGR212W; YAR033W and YGL051W with YJL134W.
- MODELLER built more than one model for most of the proteins therefore in order to select the best one, filtering was performed based on E values, sequence identities and the sequence coverage of the models. The ideal models were those with parameters such as; E values smaller than 0.01, sequence identity higher than 25% and modeled regions longer than 100 residues.
  - i) For proteins with known partners, the coverage of the modelled target structure is sometimes as low as 13% indicating that only 13% of the protein was successfully modeled. Besides, some E values were higher than 0.01. Nevertheless, many models have sequence identities higher than 25%.
  - ii) When the models of proteins with predicted partners were examined, the sequence coverage of the model structures is more than 40% and E values are smaller than 0.01 for all model structures. It means that first two criteria (E value <0.01 and modeled sequence >100 residue) were met when selecting ideal templates for each target protein. However, the sequence identities of the alignments are lower than 25% for this set of the proteins. Therefore, the model which can ensure maximum criteria at the same time was selected for each protein.
- The calculations were repeated for some interconnected regions (clusters) of the network to obtain more information about sphingolipid signaling pathway. Proteins

of these clusters were defined as follows: YOR171C and YOR034C were the proteins of Cluster A; YMR298W, YGR060W and YHL003C were the proteins of Cluster B; YDR297W, YLL006W, YMR058W, YNL307C, YAL007C, YEL027W, YHR026W, YOR016C, YHR110W and YCR034W were the proteins of Cluster C. For clusters the sequence identities of the alignments were generally higher than 25% except three proteins, YHL003C, YGR060W, YHR110W. However, the E values of some protein models were generally higher than 0.01.

When the biological relevance of the proteins with predicted interactions was examined, it was seen that only some protein complexes formed biological contacts. For example; YGR143W-YKL126W, YGR212W-YHR135C, YJL134W-YAR033W, YJL134W-YGL051W are “Biological” protein complexes, with 100% biological contacts between YJL134W-YAR033W and YJL134W-YGL051W.

- 62-75% of the hotspots (obtained from KFC) of the protein complexes are hydrophobic. (Proteins with known interactions: 62.96%, proteins with predicted interactions: 70.89%, Cluster A: 71.43%, Cluster B: 61.90%, Cluster C: 61.65%). The second highest ratio of hotspots belongs to polar residues. The chemical type of hotspots of minimum percentage is basic both for protein complexes with known interactions (8.92%) and for protein complexes with predicted interactions (5.06%), while it is acidic for protein complexes in Cluster A (0.0%), cluster B (0.0%) and Cluster C (3.76%).
- Among the hotspots of the protein complexes with known partners, LEU is the most observed hotspot residue at a percentage of 13.47 whereas CYS is the least observed hotspot residue (1.68%). LEU also has the highest percentage, 21.40%, among the hydrophobic hotspot residues of proteins with known partners while MET has the lowest percentage, 3.74%. When the proteins with predicted partners are examined it is observed that most of the amino acids are VAL (10.13 %) and PRO (10.13 %), whereas the rarely observed amino acids are MET (1.26%) and GLN (1.26%).
- The most prevalent hotspot residues were reported as tryptophan, arginine and tyrosine by (Moreira et al., 2007). When the amino acid enrichments were analyzed for the Clusters, it is seen that some results are consistent with the literature. For instance, TYR has very high concentration in Cluster A (14.28%) and in Cluster B

(11.90%) in agreement with the literature. However, the maximum repeated hotspot residue in Cluster C was LEU with 15.04%.

## 5.2. Recommendations

In this study, the aim was to elucidate the protein-protein interaction network of sphingolipid signaling pathway and the important regions for interactions. Therefore the most important part of this study was modeling the structures because the models selected were found to affect the rest of the process. However, homology modeling is limited by the sequence alignment. Indeed, for some of the proteins, no structures that could be a template for them either because of low sequence similarity or low structural coverage exist. For this reason, threading might be used in further studies to achieve more successful models than those obtained here. In threading all core elements are aligned with subsequence fragments, and the domains are searched for within each sequence that adopt one of the predefined “folds.” The sequence length determines the number of alternative alignments possible for each core motif, and may exclude domains requiring more residues, but no gap penalties are employed and there is no a priori preference for core motifs of a given size (Madej et al., 1995).

KFC server detects hotspot residues if the residues from protein A and protein B have a distance equal to or smaller than  $4\text{\AA}$ . The server shows hotspot residues individually in two groups which are hotspots in protein A and hotspots in protein B. However from these results it is not possible to understand which residue in protein B is interacting with a hotspot residue from protein A. Therefore in order to find which hotspot residue is interacting with which residue in protein B, the distances between all the combinations of hotspots of protein A and protein B were calculated, and the hotspot combinations which had a distance lower than  $10\text{\AA}$  were selected in the present study. In order to find accurate hotspot combinations, it is still necessary to find all protein B residues which had equal to or less than  $4\text{\AA}$  distance to the hotspots in protein A and this process should be repeated for every hotspot in each protein complex. Due to huge number of proteins and their complex structures in sphingolipid signaling pathway it was very time consuming for this study; so instead of this procedure, only hotspot combinations were taken into consideration here. However, in order to identify the best complex structure, in further studies, all the residues

in each protein in the complex should be tested for distance values and the consistency of the results obtained should be checked.

In order to investigate the interaction regions, protein complexes were built, however there is no information about the energetical stability of these complexes. Calculation of the free energy of a macromolecule based on its high-resolution 3D structure will give information about the energetic contribution to protein stability. Therefore in later studies interaction energies of these protein complexes should be calculated.

## APPENDIX A: HOTSPOT DISTANCES

### A.1. Proteins with Known Interaction Partners

Table A.1. The protein pair, hotspots on each protein, and the distances between hotspot residues in proteins with known interaction partners

Protein complex	Hotspots residues of the identified model	Residue pair in Protein A and Protein B	Distance (Å)	Number of the selected complex model
YLR242C-YOR033W	YLR242C 174, 177, 180	174-162	3.85	model 5
		174-163	6.04	
	YOR033W 162, 163, 208	177-162	6.24	
		177-163	9.71	
		180-162	8.49	
		180-208	4.67	
YKL004W-YJL196C	YKL004W 241, 243, 260, 267, 268	241-203	8.29	model 10
		241-242	7.20	
		241-243	5.55	
		243-242	4.94	
	YJL196C 203, 235, 242, 243	243-243	6.12	
		260-235	8.62	
		267-203	9.58	
		268-203	7.22	
		268-243	8.93	
YKL004W-YJR105W	YKL004W 259, 262, 293, 296, 297	259-79	6.36	model 9
		259-80	7.42	
		259-81	7.05	
		262-79	5.78	
	YJR105W 79, 80, 81	262-80	5.82	
		262-81	7.21	
		293-79	4.94	
		293-80	3.96	
		293-81	6.92	
		296-79	5.60	
		296-80	6.96	
		297-79	7.27	
		297-80	8.57	
YKL004W-YMR010W	YKL004W 254, 266, 293 YMR010W 182, 183, 186, 325, 329	254-325	4.43	model 10
		254-329	6.58	
		266-182	9.97	
		293-182	8.07	
		293-183	6.41	
		293-186	9.43	
YKL004W-YPL264C	YKL004W 271, 272, 281	271-131	5.68	model 9
		272-49	8.95	
		272-131	6.44	

	YPL264C 49, 120, 131	275-49	8.35	
		275-120	7.27	
		281-131	7.24	
YKL008C- YKL008C	YKL008C 95, 130, 131, 132, 134	95-130	9.66	model 4
		130-132	7.98	
		130-138	7.66	
		131-132	4.51	
	YKL008C 130, 132, 135, 138	131-135	9.94	
		131-138	5.09	
		132-130	8.58	
		132-132	4.88	
		132-135	8.12	
		132-138	6.33	
		134-132	9.30	
		134-135	7.93	
		134-138	6.18	
YKL008C- YHL003C	YKL008C 130, 131, 162, 164, 165	130-100	7.86	model 7
		130-102	3.35	
		130-103	4.21	
		131-102	6.28	
	YHL003C 100, 102, 103	131-103	4.64	
		162-100	7.03	
		162-102	6.79	
		164-100	8.15	
		164-102	5.17	
		164-103	7.89	
		165-102	8.90	
YKL008C- YMR298W	YKL008C 152, 153, 154, 155	152-83	8.07	model 1
		152-112	9.43	
		153-83	5.01	
		153-112	8.97	
	YMR298W 83, 111, 112	154-83	5.92	
		154-112	9.35	
		155-83	7.55	
		155-111	9.46	
		155-112	7.50	
YKL008C- YKL065C	YKL008C 91, 97	91-120	8.87	model 1
		91-158	9.34	
	YKL065C 76, 120, 158	97-76	9.70	
		97-120	7.04	
		97-158	9.82	
YKL008C- YBR159W	YKL008C 92, 150, 153, 173	92-215	7.64	model 3
		150-231	8.22	
		150-312	9.03	
	YBR159W 215, 231, 310, 312, 313	150-313	7.39	
		153-312	6.64	
		153-313	4.94	
		173-310	9.42	
		173-312	8.39	
YHL003C- YMR298W	YHL003C 267, 268, 270, 281	267-83	5.95	model 6
		267-84	4.40	
		268-83	5.73	
		268-84	5.29	
	YMR298W 83, 84, 113	270-83	8.61	
		270-113	7.53	

		281-83	7.71	
		281-84	9.68	
		281-113	6.74	
YDR062W- YMR296C	YDR062W 372, 474, 482	372-371	5.41	model 1
		372-374	9.03	
		474-194	8.81	
	YMR296C 194, 201, 371, 374, 396	474-396	9.12	
		482-194	9.66	
		482-201	7.21	
		482-396	9.54	
YDR062W- YGR218W	YDR062W 226, 467, 484, 485	226-1041	7.97	model 5
		226-1042	4.62	
	YGR218W 1031, 1041, 1042	467-1031	9.67	
		484-1031	5.08	
		485-1031	5.26	
YDR062W- YLR342W	YDR062W 348, 349, 350, 351, 356	348-10	9.07	model 3
		349-10	6.95	
		350-10	5.44	
	YLR342W 10,	351-10	3.09	
		356-10	5.68	
YDR062W- YKL104C	YDR062W 307, 467, 482, 484, 485, 486, 490	307-482	8.61	model 6
		307-487	9.75	
		467-488	7.73	
		467-489	8.19	
		467-490	7.20	
	YKL104C 482, 487, 488, 489, 490	482-487	6.47	
		482-488	4.79	
		482-489	7.44	
		482-490	7.98	
		484-482	9.27	
		484-487	4.66	
		484-488	6.24	
		484-489	6.51	
		484-490	5.84	
		485-482	8.59	
		485-487	6.18	
		485-488	8.56	
		485-489	7.71	
		485-490	7.10	
		486-482	6.69	
		486-487	6.42	
		486-488	9.41	
		486-489	8.75	
486-490	9.15			
490-490	9.46			
YDR062W- YBR017C	YDR062W 446, 461, 492	446-412	8.41	model 9
		446-413	6.53	
		461-412	5.63	
	YBR017C 412, 413, 415, 416	461-413	8.26	
		461-415	9.10	
		492-412	7.08	
		492-413	9.08	

		492-415	6.95	
		492-416	9.64	
YDR062W- YER110C	YDR062W 466, 467, 468, 471, 481	466-166	8.85	model 9
		467-163	8.85	
		467-164	8.41	
		467-166	5.18	
	YER110C 163, 164, 166	468-163	9.02	
		468-166	7.11	
		471-163	4.60	
		471-163	6.15	
		471-166	6.25	
		481-163	9.63	
		481-164	6.60	
		481-166	8.40	
YDR062W- YIL094C	YDR062W	No structure	No structure	No structure
YDR062W- YJR077W	YDR062W 226, 254, 255, 483, 485, 486	226-164	9.17	model 2
		226-166	9.85	
		255-166	8.81	
		483-257	5.23	
	YJR077W 164, 166, 175, 257	485-175	7.32	
		485-257	7.38	
		486-175	4.51	
		486-257	7.42	
YOR171C- YOR034C	YOR171C 357, 364, 368, 369,370	357-237	9.11	model 4
		357-238	8.11	
		357-240	8.61	
		364-238	9.35	
	YOR034C 232, 237, 238, 240	364-240	9.38	
		368-232	6.92	
		368-237	6.20	
		368-238	4.49	
		368-240	9.03	
		369-232	5.91	
		369-237	7.22	
		369-238	4.74	
		369-240	7.51	
		370-232	9.32	
		370-237	8.42	
		370-238	4.81	
370-240	5.32			
YLR260W- YGL137W	YLR260W 374, 376, 378, 387	374-256	9.66	model 7
		376-256	9.80	
		376-263	7.90	
		376-264	8.18	
	YGL137W 256, 263, 264	378-264	9.46	
		387-263	8.23	
		387-264	5.78	
YLR260W- YLR213C	YLR260W 315, 322, 323, 325	315-240	4.75	model 3
		322-311	4.72	
		322-312	7.60	
	YLR213C 240, 311, 312	323-312	8.27	
		325-311	4.72	
		325-312	6.67	
YMR298W- YKL008C	YMR298W 72, 73, 74, 75, 76,	72-95	4.74	model 3
		72-99	9.83	

	83, YKL008C 95, 99, 128	73-95 73-99 74-95 74-99 47-128 75-95 75-128 76-128 83-128	3.99 8.28 5.77 9.22 9.12 7.33 7.67 8.18 7.99	
YMR298W- YHL003C	YMR298W 72, 76, 78  YHL003C 95, 96, 129	72-95 72-96 72-129 76-95 76-96 78-95 78-96	5.19 8.48 5.87 5.89 5.25 8.05 6.28	model 1
YMR298W- YMR298W	YMR298W 50, 51, 64  YMR298W 81, 90	50-81 50-90 51-81 51-90 64-90	8.74 8.08 6.98 5.86 8.97	model 8
YJL097W- YBR159W	YJL097W 58, 73, 75, 103, 104  YBR159W 218, 219	58-218 58-219 73-218 73-219 75-218 75-219 103-218 103-219 104-218 104-219	8.65 8.13 8.95 6.07 3.90 4.61 6.97 6.69 4.88 6.30	model 5
YPL006W- YGL006W	YPL006W 1002, 1009, 1030, 1034  YGL006W 898, 904, 905, 970	1002-905 1009-898 1009-905 1030-898 1030-904 1030-905 1030-970 1034-904 1034-905	7.73 8.61 8.33 8.41 9.58 8.44 7.63 7.18 4.68	model 1
YDR297W- YBR106W	YDR297W 272, 279  YBR106W 26, 67, 77, 102	272-26  272-77  279-102	9.83  6.59  3.90	model 1
YDR297W- YLR372W	YDR297W 250, 252, 253, 285, 289  YLR372W 135, 138, 140, 163, 164, 173, 174	250-138 250-140 252-135 252-138 252-140 253-135 253-138 253-140 285-163 285-164	8.73 4.36 6.33 4.40 6.60 7.43 6.50 9.64 5.67 9.02	model 8

		285-173	5.04	
		285-174	7.14	
		289-163	8.52	
		289-173	8.81	
		289-174	8.02	
YMR27C- YOR081C	YMR272W 30, 31, 65, 70, 71	30-594	9.02	model 4
		31-593	9.88	
		31-594	6.67	
	YOR081C 564, 593, 594, 689	31-689	9.30	
		65-594	9.98	
		65-689	8.70	
		70-593	7.05	
		70-594	7.16	
		70-689	9.25	
		71-594	7.38	
YDR297W- YKL065C	YDR297W 261, 267, 268	261-152	9.60	model 10
		267-106	4.28	
	YKL065C 106, 109, 152	267-109	4.00	
		268-106	6.38	
		268-109	3.76	
YDR297W- YBR159W	YDR297W 250, 251, 252, 253	250-281	7.57	model 4
		251-215	8.34	
		251-281	5.17	
	YBR159W 214, 215, 281	252-214	9.07	
		252-215	6.07	
		252-281	4.12	
		253-214	9.49	
		253-215	6.29	
		253-281	6.15	
YDR297W- YPL264C	YDR297W 283, 284	283-21	6.20	model 5
		283-130	6.46	
	YPL264C 21, 130	284-21	3.93	
YDR297W- YGR125W	YDR297W 256, 257, 261	256-154	5.59	model 5
		257-148	9.72	
	YGR125W 148, 154	257-154	4.90	
		261-148	6.92	
YDR297W- YHR007C	YDR297W 256, 257, 258	256-503	5.13	model 2
		256-504	6.08	
	YHR007C 503, 504	257-503	5.63	
		257-504	6.40	
		258-503	5.53	
		258-504	5.45	
YDR297W- YML012W	YDR297W 257, 268, 275, 276	257-89	7.77	model 3
		257-91	5.35	
		257-138	6.00	
		268-62	4.10	
	YML012W 62, 88, 89, 91, 138	268-89	8.39	
		275-88	9.90	
		275-89	8.42	
		275-138	9.35	

		276-89	8.75	
		276-91	9.82	
YDR297W- YKL088W	YDR297W 251, 253, 277, 281, 282, 285	251-457	9.99	model 2
		251-463	5.40	
	251-465	5.89		
	253-463	8.52		
	YKL088W 414, 455, 457, 458, 463, 465	253-465	6.18	
		277-463	9.83	
		277-465	8.90	
		281-414	9.08	
		281-455	8.50	
		281-457	8.08	
		281-458	4.37	
		281-463	6.89	
		281-465	9.61	
		282-455	8.88	
		282-457	9.24	
		282-458	5.84	
		282-463	9.47	
		285-455	6.34	
	285-457	5.83		
	285-458	4.24		
	285-463	6.48		
YLR372W- YDL015C	YLR372W 161, 163, 173, 175	161-45	8.85	model 2
		163-45	9.09	
		163-57	8.49	
		163-58	6.19	
	YDL015C 45, 56, 57, 58, 60	173-45	6.06	
		173-56	8.53	
		173-57	8.21	
		173-58	6.76	
		175-56	5.90	
		175-57	6.02	
		175-58	7.23	
YER093C- YNL006W	YER093C 81, 99, 101, 102, 103, 124	81-68	4.88	model 1
		99-68	6.52	
		101-68	4.74	
		102-68	7.44	
	YNL006W 45, 68, 87	103-45	8.80	
		103-68	5.90	
		103-87	8.89	
		124-87	7.51	
YER093C- YBR270C	YER093C 226, 230, 232, 238, 239	226-21	6.51	model 2
		226-22	5.86	
		226-43	9.94	
		230-21	6.49	
	YBR270C 21, 22, 39, 43, 45	230-22	9.78	
		230-43	6.54	
		230-45	7.24	
		232-43	5.86	
		232-45	5.10	
		238-21	9.30	
		238-43	7.28	
		239-21	7.80	
		238-22	9.48	
		239-39	7.97	

		239-43	7.13	
YER093C- YKL203C	YER093C 41, 45, 68	41-2316	8.96	model 8
		41-2321	3.50	
		41-2323	7.84	
	YKL203C 2238, 2316, 2321, 2323	45-2238	8.54	
		45-2321	3.65	
		45-2323	5.30	
		68-2316	9.98	
YER093C- YMR068W	YER093C 65, 86, 88, 103, 105	65-119	8.12	model 5
		65-120	6.74	
		65-121	6.87	
	YMR068W 15, 85, 118, 119, 120, 121, 152	86-85	6.96	
		86-119	8.11	
		88-118	5.79	
		88-119	4.52	
		88-120	7.67	
		88-152	6.99	
		103-15	5.89	
		105-15	9.06	
YBR058C-A- YMR296C	YBR058C-A 46	46-331	5.83	model 5
		46-332	8.05	
	YMR296C 331, 332, 403, 404, 406, 407	46-403	4.88	
		46-404	3.61	
		46-406	7.59	
		46-407	7.82	
YBR058C-A- YDR062W	YBR058C-A 43, 45, 46	43-179	8.60	model 4
		43-183	9.56	
		43-399	6.38	
	YDR062W 179, 180, 183, 399, 401	43-401	6.33	
		45-179	4.16	
		45-180	5.34	
		45-183	7.27	
		45-399	9.81	
		45-401	5.71	
		46-179	6.49	
		46-180	6.47	
46-401	9.37			
YPL087W- YDL015C	YPL087W 182, 234, 240, 241, 244	182-78	8.32	model 5
		234-86	9.78	
		234-99	8.41	
		234-101	6.38	
	YDL015C 78, 86, 99, 101	240-78	7.14	
		240-101	5.37	
		241-78	9.89	
		241-101	8.45	
		244-78	8.57	
YBR183W- YLR372W	YBR183W 205, 233, 237, 240	205-139	8.57	model 4
		205-158	9.08	
		205-160	7.50	
	YLR372W 128, 129, 139, 157, 158, 160	233-128	9.53	
		233-139	8.04	
		233-158	8.42	
		233-160	8.72	

		237-128	5.69	
		237-129	6.36	
		237-157	9.18	
		237-158	7.44	
		240-128	7.82	
		240-129	8.65	
		240-157	7.08	
		240-158	7.01	
YBR183W- YOR016C	YBR183W 201, 204, 235, 239	201-91	5.52	model 5
		201-92	8.12	
		201-138	7.33	
		204-91	9.02	
	YOR016C 91, 92, 138	204-138	5.54	
		235-91	5.28	
		235-92	5.55	
		239-91	8.84	
		239-92	7.60	
YBR183W- YLR018C	YBR183W 272, 275	272-187	3.86	model 8
		272-188	5.47	
		272-189	5.99	
	YLR018C 187, 188, 189	275-187	7.00	
		275-188	5.59	
		275-189	7.50	
YBR183W- YKL065C	YBR183W 254, 255, 256	254-159	4.38	model 8
		254-160	6.31	
		255-159	5.41	
	YLR018C 159, 160	255-160	6.34	
		256-159	7.60	
		256-160	6.76	
YBR183W- YHR140W	YBR183W 234, 237, 238, 240, 241	234-56	7.96	model 9
		234-98	8.48	
		237-56	8.84	
		237-97	7.58	
	YHR140W 56, 97, 98	237-98	4.41	
		238-56	6.67	
		238-97	8.79	
		238-98	6.87	
		240-97	6.71	
		240-98	6.31	
		241-56	9.15	
		241-97	5.14	
		241-98	5.65	
YBR183W- YBR159W	YBR183W 264, 265, 275	264-313	5.14	model 2
		265-214	8.10	
	YBR159W 214, 254, 313	265-313	3.78	
		275-254	6.95	
YBR183W- YBR106W	YBR183 213, 214, 215, 217	213-67	7.32	model 10
		213-75	9.30	
		214-67	6.43	
		214-75	5.81	
	YBR106W 67, 75	215-67	5.84	
		215-75	5.62	
		217-67	9.92	
		217-75	8.00	
YBR183W-	YBR183W	205-25	6.69	model 4

YGR060W	205, 236, 240, 243, 244, 247, 253, 267, 272, 275, 276  YGR060W 25, 28, 43, 47, 48	205-28	8.12	
		236-25	6.68	
		236-28	4.41	
		240-28	7.52	
		243-47	9.08	
		244-43	8.14	
		244-47	7.64	
		247-43	7.90	
		247-47	7.53	
		247-48	9.31	
		253-43	5.87	
		267-47	9.63	
		267-48	8.66	
		272-48	9.46	
		275-47	8.60	
275-48	5.78			
276-48	8.87			
YBR183W- YFL041W	YBR183W 205, 209, 236  YFL041W 379, 381, 382, 516, 519	205-379	9.02	model 5
		205-381	7.32	
		205-382	5.22	
		209-379	8.46	
		209-381	8.66	
		209-382	8.96	
		236-381	9.75	
		236-382	6.06	
		236-516	7.19	
		236-519	8.34	
YBR183W- YDR506C	YBR183W 199, 203, 206, 209, 210, 236  YDR506C 62, 247	199-247	6.86	model 4
		203-247	8.67	
		206-62	8.14	
		206-247	9.67	
		209-62	8.85	
		210-62	5.88	
		236-247	6.18	
YKL126W- YMR104C	YKL126W 495, 497, 500  YMR104C 441, 442, 597	495-441	9.21	model 10
		495-442	5.58	
		495-615	6.58	
		497-441	5.68	
		497-442	4.22	
		500-597	5.89	
YKL126W- YNL106C	YKL126W 560, 561, 562, 563, 565  YNL106C 574, 576, 897	560-576	7.71	model 5
		561-574	9.57	
		561-576	9.11	
		561-897	8.99	
		562-574	8.92	
		562-576	9.94	
		562-897	8.23	
		563-574	6.10	
		563-576	9.01	
		563-898	5.48	
		564-574	8.28	
		564-897	6.84	
		565-574	7.85	
		565-897	7.74	
YDL015C- YKL008C	YDL015C 41, 43, 69, 109	41-131	4.79	model 6
		43-131	6.58	

	YKL008C 126, 131	69-126 69-131 109-1126 109-131	7.42 9.98 5.43 9.74	
YDL015C- YLR372W	YDL015C 85, 86, 93, 96, 97, 98  YLR372W 108, 109, 133, 134	85-133 85-134 86-109 86-133 86-134 93-108 93-109 96-108 96-109 97-108 97-109 97-134 98-109 98-133 98-134	8.59 6.89 9.42 6.83 5.86 4.43 4.10 7.39 4.80 9.79 6.93 9.58 7.62 7.76 6.97	model 6
YDL015C- YPL087W-	YDL015C 110, 111, 112  YPL087W 166, 169, 206, 209	110-166 110-209 111-166 111-169 111-209 112-166 112-169 112-206 112-209	7.83 6.55 4.53 7.02 5.57 7.53 7.49 9.42 5.81	model 5
YDL015C- YDL015C	YDL015C, 55, 56, 120  YDL015C 46, 48, 61, 66	55-46 56-46 56-48 56-61 120-46 120-61 120-66	7.16 8.21 8.70 7.41 9.52 3.98 6.49	model 1
YDL015C- YJL196C	YDL015C 90, 99, 100, 101, 115  YJL196C 203, 207, 238, 239, 243	90-203 90-207 90-239 90-243 99-203 99-207 99-238 99-239 99-243 100-203 100-207 100-239 100-243 101-203 101-243 115-207	8.69 6.00 7.55 8.81 8.68 7.72 8.43 5.52 5.06 7.87 9.66 8.85 4.98 7.19 7.34 7.92	model 3
YDL015C- YOR016C	YDL015C 84, 104  YOR016C 91, 115, 136, 138,	84-91 84-136 84-138 84-139	6.67 9.38 5.04 6.96	model 8

	139	104-115	8.67	
YDL015C- YKL065C	YDL015C 63, 115, 117	63-81	6.98	model 9
		115-81	7.57	
		115-82	9.02	
	YKL065C 81, 82, 98	115-98	6.92	
		117-81	4.15	
		117-82	4.64	
YDL015C- YBR159W	YDL015C 16, 17, 18, 37	16-254	7.49	model 4
		16-292	8.29	
		16-294	4.67	
	YBR159W 254, 292, 294, 296, 297	16-296	8.07	
		16-297	7.19	
		17-254	7.75	
		17-294	7.58	
		17-297	7.90	
		18-254	6.20	
		37-297	8.48	
YDL015C- YBR017C	YDL015C 111, 113	111-179	5.63	model 6
		111-225	5.75	
	YBR017C 179, 225	113-179	8.11	
		113-225	6.38	
YDL015C- YDR506C	YDL015C 10, 67, 110, 113	10-392	9.00	model 4
		10-395	7.28	
		67-392	9.11	
	YDR506C 392, 395, 424	110-392	3.46	
		110-424	9.65	
		113-424	7.94	
YDL015C- YPR028W	YDL015C 55, 115, 117, 118, 120, 121, 122, 123	55-48	9.20	model 3
		115-77	8.85	
		117-73	8.55	
		117-77	5.43	
	YPR028W 47, 48, 73, 77	118-73	6.57	
		118-77	6.77	
		120-48	9.91	
		120-73	5.22	
		120-77	9.08	
		121-47	9.15	
		121-48	6.80	
		121-73	8.10	
		122-47	6.98	
		122-48	6.32	
		123-47	6.66	
		123-48	7.10	
YDL015C- YBR110W	YDL015C 92, 95, 97, 98	92-167	8.27	model 7
		92-168	9.69	
		95-167	6.93	
	YBR110W 167, 168, 170	95-168	8.87	
		97-167	6.82	
		97-168	6.41	
		97-170	7.39	
		98-167	6.80	
		98-168	5.35	
		98-170	6.01	
YDL015C-	YDL015C	89-148	3.16	model 10

YBR094W	89, 90, 99, 101  YBR094W 112, 148, 149, 150	89-149	6.30	
		89-150	7.14	
		90-148	4.60	
		90-149	7.74	
		90-150	8.34	
		99-112	9.39	
		99-148	6.50	
		99-149	7.60	
		99-150	6.25	
		101-112	6.37	
		101-148	6.78	
		101-149	5.80	
		101-150	3.66	
YDL015C- YKL182W	YDL015C 78, 101, 103  YKL182W 1664, 1792, 1797	78-1664	5.38	model 6
		101-1664	9.00	
		101-1792	7.26	
		101-1797	9.65	
		103-1664	9.05	
		103-1797	9.80	

## A.2. Proteins with Predicted Partners

Table A.2. The protein pair, hotspots on each protein, and the distances between hotspot residues in proteins with predicted partners

Protein complex	Hotspots residues of the identified model	Hotspots at Protein A and Protein B	Distance (Å)	Number of the selected complex model
YDR294C- YHL135C	YDR294C 429, 432, 436, 437, 439, 440	429-201	6.17	model 7
		432-92	9.87	
		432-94	9.70	
		432-201	7.55	
	YHL135C 90, 91, 92, 94, 201, 203	436-91	8.93	
		436-92	8.79	
		436-94	6.47	
		436-201	9.06	
		437-94	9.54	
		437-201	9.42	
		437-203	8.94	
		439-91	7.78	
		439-92	9.52	
		439-94	8.92	

		440-90	8.07	
		440-91	6.55	
		440-92	9.42	
YER019W- YHL020C	YER019W 158, 160, 211	158-332	5.28	model 10
		158-333	5.28	
		158-336	5.19	
	YHL020C 332, 333, 336, 337, 340	158-337	5.92	
		158-340	9.97	
		160-336	7.28	
		160-337	6.79	
		160-340	7.63	
		211-332	6.78	
		211-333	8.87	
		211-336	4.50	
		211-337	8.10	
		211-340	8.29	
YJL134W- YAR033W	YJL134W 103, 106, 107, 108	103-178	8.00	model 3
		103-180	4.83	
		106-180	8.56	
	YAR033W 178, 180, 198	106-198	4.26	
		107-180	9.34	
		108-180	8.43	
		108-198	9.07	
YJL134W- YGL051W	YJL134W 106, 107, 109, 110, 111, 136	106-219	9.42	model 3
		107-149	9.86	
		107-219	8.17	
		109-149	4.82	
	YGL051W 149, 150, 219	109-150	8.21	
		110-149	4.95	
		110-150	8.57	
		111-149	5.72	
		111-150	8.28	
		136-149	5.81	
		136-150	6.34	
YGR143W- YKL126W	YGR143W 585, 586, 589, 625, 692, 696	585-521	5.22	model 3
		585-523	5.60	
		585-524	8.58	
		586-521	6.89	

	YKL126W	586-523	4.77		
	494, 521, 523, 524	586-524	7.31		
		589-523	9.04		
		625-523	10.00		
		625-524	9.99		
		692-494	6.95		
		696-494	8.72		
		696-521	6.95		
YKR053C- YAR033W	YKR053C	98-164	8.55	model 5	
	98, 102, 103, 185	102-148	8.59		
		102-149	6.75		
		102-150	8.84		
	YAR033W 148, 149, 150, 164	102-164	6.8		
		103-149	7.94		
		103-150	8.49		
		103-164	9.94		
		185-148	6.32		
		185-149	4.92		
		185-150	6.28		
		YKR053C- YGL051W	YKR053C		115-167
115, 119, 233, 239			119-154	9.52	
	119-165		9.56		
	119-166		6.91		
YGL051W 154, 165, 166, 167	119-167		4.86		
	233-165		4.95		
	233-166		7.00		
	233-167		8.94		
	239-165		9.44		
	239-166		9.74		
	239-167		7.02		
	YGR212W- YHR135C		YGR212W	222-91	6.63
		222, 224, 384	222-92	8.72	
222-93			6.87		
222-94			9.22		
YHR135C 91, 92, 93, 94		224-91	9.65		
		384-91	5.58		
		384-92	5.77		
		384-93	9.30		

### A.3. Clusters\Modules

Table A.3. The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster A for 10 models

Protein complex	Model	Hotspots residues of the model	Hotspots at Protein A and Protein B	Distance (Å)	
YOR171C-YOR034C	Model 1	YOR171C 243, 267, 381	243-184	9.85	
			243-213	9.35	
			243-214	9.41	
		YOR034C 177, 184, 213, 214	267-184	8.82	
			381-177	9.44	
	Model 2	YOR171C 359	359-250	9.94	
	YOR034C 250				
	Model 3	No Hotspot found	-	-	
	Model 4	YOR171C 357, 364, 368, 369, 370	YOR034C 232, 237, 238, 240,	357-237	9.11
				357-238	8.11
				357-240	8.61
				364-238	9.35
				364-240	9.38
				368-232	6.92
				368-237	6.20
				368-238	4.49
				368-240	9.03
				369-232	5.91
				369-237	7.22
				369-238	4.74
				369-240	7.51
				370-232	9.32
	370-237	8.42			
	370-238	4.81			
	370-240	5.32			
	Model 5	Distance of Hotspots > 10Å	-	-	
	Model 6	YOR171C 333	YOR034C 114	333-114	6.83
Model 7	YOR171C 370, 391				
Model 7	YOR034C 203, 204, 240	370-203	9.27		
		370-204	5.67		
		391-203	8.88		
Model 8	YOR171C 359, 360	391-240	6.09		
		359-240	9.35		

		YOR034C 240	360-240	8.26
	Model 9	YOR171C 236, 379	236-250	9.72
		YOR034C 250, 215	379-215	9.79
	Model 10	Distance of Hotspot > 10Å	-	-
YOR034C- YOR171C	Model 1	YOR034C 232, 236, 238, 244, 259  YOR171C 270, 359, 360, 361	232-360	9.76
			232-361	7.32
			236-361	8.54
			238-270	9.29
			238-360	9.27
			238-361	5.87
			244-359	9.42
			259-360	6.79
			259-361	7.57
	Model 2	No Hotspot found	-	-
	Model 3	YOR034C 175, 177, 212, 213, 214  YOR171C 240, 294	175-294	4.87
			177-294	9.98
			212-240	6.82
			213-240	3.77
			214-240	6.17
			214-294	7.69
	Model 4	YOR034C 250  YOR171C 399	250-399	4.72
	Model 5	Distance of Hotspots > 10Å	-	-
	Model 6	YOR034C 132, 145, 169, 174, 175, 176  YOR171C 361, 362, 374, 381, 399, 401	132-399	4.87
			132-401	8.31
			145-361	9.44
			169-374	7.80
			169-381	9.85
174-362			8.75	
174-381			9.98	
175-362			9.32	
175-381			8.26	
176-362			7.38	
176-374			7.70	
176-381	8.37			
Model 7	YOR034C 62, 63  YOR171C 408	62-408	6.68	
		63-408	5.49	

	Model 8	No Hotspot found	-	-
	Model 9	YOR034C 204, 212, 240, 245	204-373	8.62
			204-387	6.44
			212-408	9.66
		YOR171C 387, 391, 408	240-391	6.09
	Model 10	YOR034C 260	245-408	6.54
			260-234	9.60
		YOR171C 234, 235, 236	260-235	8.79
			260-236	6.90

Table A.4. The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster B

Protein complex	Hotspots residues of the identified model	Hotspots at Protein A and Protein B	Distance (Å)	Number of the selected complex model
YMR298W- YMR298W	YMR298W 44, 46	44-112	6.24	model 3
		44-114	5.52	
		46-74	6.29	
	YMR298W 74, 112, 114	46-112	5.05	
		46-114	9.12	
YHL003C- YMR298W	YHL003C 154, 186, 187	154-51	7.62	model 1
		186-51	5.64	
		186-52	5.15	
	YMR298W 51, 52, 53, 62	186-53	5.71	
		186-62	9.59	
		187-51	6.03	
		187-52	3.73	
187-53	6.20			
YGR060W- YGR060W	YGR060W 212, 216, 223, 269, 274, 275	212-4	4.61	model 8
		216-158	9.51	
		223-4	9.88	
		269-161	9.74	
	YGR060W 4, 158, 161, 252	274-252	9.09	
		275-252	7.21	
YMR298W- YHL003C	YMR298W 81, 90, 114, 115, 116	81-230	7.36	model 6
		90-205	9.06	
		90-230	7.10	
		114-230	9.10	
	YHL003C 205, 230, 233, 234	114-233	4.00	
		114-234	5.01	
		115-233	7.13	
		115-234	8.07	

		116-233	8.44	model 5
		116-234	9.54	
YGR060W- YHL003C	YGR060W 65, 191, 223	65-307	9.67	
		191-85	9.31	
	YHL003C 85, 87, 307	223-85	5.84	
		223-87	9.50	

Table A.5. The protein pair, hotspots on each protein, and the distances between hotspot residues of proteins in Cluster C

Protein complex	Hotspots residues of the identified model	Hotspots at Protein A and Protein B	Distance (Å)	Number of the selected complex model
YMR298W- YMR298W	YMR298W 44, 46	44-112	6.24	model 3
		44-114	5.52	
	YMR298W 74, 112, 114	46-74	6.29	
		46-112	5.05	
		46-114	9.12	
YHL003C- YMR298W	YHL003C 154, 186, 187	154-51	7.62	model 1
		186-51	5.64	
		186-52	5.15	
	YMR298W 51, 52, 53, 62	186-53	5.71	
		186-62	9.59	
		187-51	6.03	
		187-52	3.73	
YGR060W- YGR060W	YGR060W 212, 216, 223, 269, 274, 275	212-4	4.61	model 8
		216-158	9.51	
		223-4	9.88	
	YGR060W 4, 158, 161, 252	269-161	9.74	
		274-252	9.09	
		275-252	7.21	
YMR298W- YHL003C	YMR298W 81, 90, 114, 115, 116	81-230	7.36	model 6
		90-205	9.06	
		90-230	7.10	
		114-230	9.10	
	YHL003C 205, 230, 233, 234	114-233	4.00	
		114-234	5.01	
		115-233	7.13	
		115-234	8.07	
		116-233	8.44	
YGR060W- YHL003C	YGR060W 65, 191, 223	116-234	9.54	model 5
		65-307	9.67	
		191-85	9.31	
		223-85	5.84	

	YHL003C 85, 87, 307	223-87	9.50	
--	------------------------	--------	------	--

## REFERENCES

- Aloy, P. and R.B. Russell., 2002, “*InterPreTS: protein Interaction Prediction through Tertiary Structure*”, *Bioinformatics*, Vol. 19, No. 1, pp. 161-162.
- Aytuna, A.S., et al., 2005, “Prediction of protein- protein interactions by combining structure and sequence conservation in protein interfaces”, *Bioinformatics*, Vol. 21, No. 12, pp. 2850-2855.
- Chen, Y., et al., 2007, “*3D-partner: a web server to infer interacting partners and binding models*”, *Nucleic acid Research*, Vol. 35, Web server issue W561-W567.
- Cockell, S.J., et al., 2006, “Structure-based evaluation of in silico predictions of protein-protein interactions using Comparative Docking”, *Bioinformatics*, Vol. 23, No. 5, pp. 573-581.
- Darnell, S. J., D. Page, and J. C. Mitchell, 2007, “*An automated decision-tree approach to predicting protein interaction hot spots*”, *Proteins* 2007; 68:813–823, Wiley-Liss, Inc., DOI: 10.1002/prot.21474.
- Darnell, S. J., L. LeGault, and J. C. Mitchell, 2008, “*KFC Server: interactive forecasting of protein interaction hot spots*”, *Nucleic Acids Research*, 2008, Vol. 36, W265–W269, Web Server issue doi:10.1093/nar/gkn346.
- Dickson, R.C., 2008, “*New insights into sphingolipid metabolism and function in budding yeast*”, *Journal of Lipid Research*, Vol. 49, pp. 909-921, DOI 10.1194/jlr.R800003-JLR200.
- Dohkan, S., et al., 2006, “Improving the performance of an SVM-based method for predicting protein-protein interactions”, *In Silico Biology* 6, 0048, *Bioinformation Systems* e.V.
- Durmuş, S., 2006, *Constraint-Based Analysis of Signaling Networks: Yeast Signal Transduction and Mammalian Epidermal Growth Factor Receptor Signaling*, M.S. Thesis, Boğaziçi University.

- Eswar, N., et al., 2007, “*Comparative Protein Structure Modeling Using MODELLER*”, Current Protocol in Protein Science 2.9.1-2.9.31, November 2007, DOI: 10.1002/0471140864.ps0209s50.
- Fukuhara, N. and T. Kawabata, 2008, “*HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures*”, Nucleic Acid Research, Vol. 36, Web Server issue W185-W189 doi:10.1093/nar/gkn218.
- Gaffard, N., et al., 2002, “*IPPRED: server for proteins interactions inference*”, Bioinformatics, Vol. 19, No. 7, pp. 903-904.
- Han, D., et al., 2004, “*PreSPI: a domain based prediction system for protein-protein interaction*”, Nucleic Acids Research , Vol. 32, No. 21, pp. 6312-6320.
- Kim, D.E., D. Chivian and D. Baker, 2004, “*Protein structure prediction and analysis using the Robetta server*”, Nucleic Acids Research, Vol. 32, W526–W531, Web Server issue, DOI: 10.1093/nar/gkh468, Oxford University Press.
- K. Ozbayraktar, F. B., and K.O. Ulgen, 2009, “Molecular facets of sphingolipids: Mediators of diseases”, Biotechnol. J., DOI 10.1002/biot.200800322.
- Kortemme, T., D. E. Klm, and D. Baker, 2004, “*Computational Alanine Scanning of Protein-Protein Interfaces*”, p12. [DOI: 10.1126/stke.2192004p12].
- Li, Y., et al., 2005, “Magnitude of the Hydrophobic Effect at Central versus Peripheral Sites in Protein-Protein Interfaces”, Elsevier, Structure, Vol. 13, 297–307, DOI 10.1016/j.str.2004.12.012.
- Madej, T., J.-F. Gibrat, and S.H. Bryant, 1995, “Threading a Database of Protein Cores”, PROTEINS: Structure, Function, and Genetics, 23:356-369.

- Marti-Renom, M.A., B. Yerkovich, A. Sali, 2002, “*Modeling Protein Structure From Its Sequence*”, Laboratories of Molecular Biophysics, Pels Family Center for Biochemistry and Structural Biology, The Rockefeller University, 1230 York Ave, New York, NY 10021, USA.
- Moreira, I. S., P. A. Fernandes, and M. J. Ramos, 2007, “*Hot spots—A review of the protein–protein interface determinant amino-acid residues*”, *Proteins*, 68:803–812, Wiley-Liss, Inc., DOI: 10.1002/prot.21396.
- Ogmen, U., O. Keskin, A.S Aytuna, R. Nussinov and A.Gursoy, 2005, “*PRISM: protein interactions by structural matching*”, *Nucleic Acids Research*, Vol. 33, Web Server issue W331–W336, doi:10.1093/nar/gki585.
- Ogretmen, B., 2006, “*Sphingolipids in cancer: Regulation of pathogenesis and therapy*”, *FEBS Letters* 580, pp. 5467–5476.
- Ohanian, J., and V. Ohanian, 2001, “*Sphingolipids in mammalian cell signalling*”, *Cell. Mol. Life Sci.*, 58, pp. 2053–2068, Birkhäuser Verlag, Basel.
- Orfan, Y., and B. Rost, 2007, “*Protein–Protein Interaction Hotspots Carved into Sequences*”, *PLoS Comput Biol* 3(7): e119. doi:10.1371/journal.pcbi.0030119.
- Pitre, S., et al., 2006, “*PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs*”, *BMC Bioinformatics*, 7:365, doi:10.1186/1471-2105-7-365.
- Rousseau, F. and J. Schymkowitz, 2005, “*A systems biology perspective on protein structural Dynamics and signal transduction*”, *Current Opinion in Structural Biology*, Vol. 15, pp. 23-30.
- Sanchez, I.E., P. Beltrao, F. Schymkowitz, J. Ferkinghoff-Borg, J., Rousseau, F., Serrano, L., 2008, “*Genome-Wide Prediction Prediction of SH2 Domain Targets Using Structural Information and the FoldX Algorithm*”, *PNAS Comput Biol* 4(4): e1000052.

- Schwede, T., J. Kopp, N. Guex., M. C. Peitsch, 2003, “*SWISS-MODEL: an automated protein homology-modeling server*”, *Nucleic Acids Research*, Vol. 31, No. 13 pp. 3381–3385, DOI: 10.1093/nar/gkg520.
- Shen, J., et al., 2006, “Predicting protein-protein interactions based only on sequences information”, *PNAS*, Vol. 104, No. 11, pp. 4337-4341.
- Shoemaker, B.A., and A.R. Panchenko, 2007, “Depichering Protein-Protein Interactions.Part II. Computational methods to predict protein and domain interaction partners”, *Plos Comput Biol* 3(3): e43. doi:10.1371/journal.pcbi.0030043.
- Sherman, F., 2002, “*Getting Started with Yeast*”, *Methods Enzymol.* 350, 3-41.
- Tien, A., et al., 2003, “Identification of the Substrates and Interaction proteins of Aurora Kinases from a Protein Interaction Model”, pp. 93-103, *MCP Papers in Press*.
- Tovchigrechko, A., and I.A. Vakser, 2006, “*GRAMM-X public web server for protein–protein Docking*”, *Nucleic Acids Research*, 2006, Vol. 34, W310–W314, Web Server issue doi:10.1093/nar/gkl206.
- Worrall, D., C. K-Y. Ng, A. M. Hetherington, 2003, “*Sphingolipids, new players in plant signaling*”, *TRENDS in Plant Science*, Vol.8, No.7, pp. 317-320, Elsevier, July.
- Zhu, H. et al. 2006, “*NOXclass: prediction of protein-protein interaction types*”, *BMC Bioinformatics*, 7:27 doi:10.1186/1471-2105-7-27.